Published online March 29, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 6 ***1937–1941***
DOI: 10.1093/nar/gkh522

A third base pair for the polymerase chain reaction: inserting isoC and isoG

Scott C. Johnson, Christopher B. Sherrill, David J. Marshall, Michael J. Moser and James R. Prudent*

Research Department, EraGen Biosciences Inc., 918 Deming Way, Madison, WI 53717, USA

Received January 22, 2004; Revised and Accepted March 10, 2004

ABSTRACT

Two additional bases (isoguanosine and isocytosine), generating a third base pair, have been implemented in PCR. Enzyme fidelity for the third base pair is demonstrated using molecular thermodynamic melting, chemical cleavage and molecular beacons. When amplifying as few as 15 targets containing multiple non-natural base pairs with 40 cycles of amplification, our results confirm sequence conservation. The additional sequence space provided by three base pairs allows for the construction of molecular tools that achieve higher complexity and better discrimination than those possible with natural DNA alone.

INTRODUCTION

The rule-based molecular recognition displayed by DNA makes it ideal throughout biotechnology, where molecules binding molecules are needed for such applications as localizing DNA (1), assembling nanostructures (2), building antibody-like molecules called 'aptamers' (3,4) and performing tasks by which information on a molecular structure is obtained. Expanding DNA chemistry to include additional base pairs would enhance the capabilities of this powerful molecular recognition system. This in turn has lead chemists to develop new ways of increasing the number of DNA building blocks. The first experimental data suggesting that new base pairs could be used in replication, transcription and translation was demonstrated by Benner and colleagues using shuffled hydrogen bonding schemes (5,6).

More recently, Romesberg and colleagues used the idea of hydrophobic interactions to create base pairs that did not rely on hydrogen bonding (7,8). With a specific mixture of two polymerases, the authors demonstrated incorporation followed by extension. Going one step further, Yokoyama and colleagues combined the concepts of shuffled hydrogen bonds and van der Waals interactions to develop a base pair that could be polymerized into the RNA transcript site specifically opposite a non-natural counterpart (9,10). More recently, our group demonstrated that consecutive non-natural bases could be incorporated specifically opposite their non-natural counterparts and that other replication-dependent enzymes can efficiently recognize a third base pair (11). Yet the ability to place a third base pair into a commonly used replication system such as PCR has not been demonstrated.

PCR is the most widely used method in molecular biology with applications that include: amplification, detection, quantitation, cloning, genotyping of genetic material, selection and screening of antibody- and/or enzyme-like nucleic acids and taggant analysis (12). Until now, PCR applications have been limited to placing non-natural base pairs into the priming regions and thus not requiring high fidelity incorporation of the correct complements (13). The addition of a new base pair between the priming sites of an exponential amplification system like PCR requires that the new bases have a low rate of misincorporation. Here, we present data demonstrating that chemistry. We show that the isoguanosine:isocytosine (iG:iC) base pair can act as a third base pair for extension, replication and PCR amplification. Using primer extension melt analysis, we first show that the major percentage of incorporated nucleotide opposite iG is iC. We then confirm high polymerase fidelity, resulting in maintenance of the iC:iG base pair following PCR using a method of acid cleavage partial sequencing. Finally, an iC:iG-containing molecular beacon was constructed and used to demonstrate applicability. The additional base pair will be useful in numerous other applications such as aptamer production or multiplexed diagnostics.

MATERIALS AND METHODS

Primer extension and melt analysis

A 5′-labeled template DNA oligonucleotide TET-d(T-iC-CGT-iG-CCGTCTCCGTCGTCAGCCGTCA) at 200 nM was combined with a 2-fold excess of each control d(TGAC-GGCTGACGACGGAGACGGGACG), d(TGACGGCTGA-CGACGGAGACGGAACG), d(TGACGGCTGACGACGG-AGACGGTACG), d(TGACGGCTGACGACGGAGACGG-CACG), d(TGACGGCTGACGACGGAGACGG-iC-ACG′) or experimental DNA oligonucleotide d(TGACGG-CTGACGACGGAG). Extension reactions were conducted in a 10 µl reaction volume containing 10 mM bis-tris-propane–HCl pH 9.1, 40 mM potassium acetate, 2 mM magnesium chloride, 0.1 mg/ml bovine serum albumin, 25 µM

*To whom correspondence should be addressed. Tel: +1 608 662 9000; Fax: +1 608 662 9004; Email: jprudent@eragen.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

deoxynucleoside triphosphates (dGTP, dATP, dTTP, dCTP and diCTP), 10 μM deoxyisoguanosine triphosphate–Dabcyl (iGTP-Dabcyl) and 1× TiTaq (Clontech, Palo Alto, CA). Extension was performed using the following profile: 30 s at 95°C, 30 s at 60°C and 1 min at 65°C. Reactions were terminated by the addition of EDTA at 10 mM. Thermal melt analysis from 60 to 90°C was performed on an iCycler (Bio-Rad, Hercules, CA) using the following profile: 30 s at 60°C, 60 cycles of 10 s at 60°C with an increase of 0.5°C each cycle.

Acid cleavage analysis

Primer extension was conducted by combining SCJ1244 [Cy3-d(CCAATGTACGGGGTAAACTCT)] at 300 nM with either SCJ1247 [d(GAAGTCCAGCAATCAGAACTATG-GCGACTCTCTACCTCCTGCAGGCCCTACCACTTCCC-AATAGCTAAGAGTTTACCCCGTACATTGG)] or SCJ1248 [d(GAAGTCCAGCAATCAGAACTATGGCGACTCTCTAC-CTCCTGC-iG-GGCCC-iC-ACCACTTCCCAATAGCTAA-GAGTTTACCCCGTACATTGG)] at 200 nM in a 25 μl reaction volume using the same buffer conditions listed previously except for 25 μM natural deoxynucleoside triphosphates (GATC) and 50 μM non-natural deoxynucleoside triphosphates (iG and iC). Extension was performed using the following profile: 1 min at 95°C, 10 s at 55°C and 30 s at 72°C. PCR amplicons were generated using the DNA oligonucleotides SCJ1244 and SCJ1243 [d(GAAGTCCAGCAA-TCAGAACTATG)] as primers at 200 nM. Ten-fold serial dilutions of SCJ1247 and SCJ1248 were made and 2.5 μl was used in a 10 μl reaction mix as described for the extension. Reactions were cycled according to the following profile: 1 min at 95°C, 40 cycles of 10 s at 95°C, 10 s at 55°C and 30 s at 72°C.

Acid cleavage of the extension products was performed by mixing equal volumes of the extension reaction with 100 mM glacial acetic acid. Reactions were incubated for 30 min at 95°C. Reaction vessels were opened and the acid was allowed to evaporate at 95°C. Two volumes of 100 mM ammonium hydroxide were added and the reactions were incubated for 5 min at 95°C. Reaction vessels were opened and the base was allowed to evaporate at 95°C. Cleavage products were resuspended in formamide, heated for 1 min at 95°C and analyzed by 7 M urea–10% polyacrylamide gel electrophoresis and fluorescence imaging. Cleavage products were quantitated using ImageQuant software (Amersham, Piscataway, NJ).

Molecular beacon analysis

PCR amplicons were generated as described in the acid cleavage analysis section except that the DNA oligonucleotide primers used were SCJ1243 and d(CCAATGTAC-GGGGTAAACTCT). The dual-labeled DNA oligonucleotide containing non-natural bases FAM-d(GTGCCGGT-iG-GGGCC-iC-GCAGGAGGGGCAC)-Dabcyl was used as a molecular beacon at 200 nM. DNA oligonucleotide templates for the PCR reactions were the same 10-fold serial dilutions of SCJ1247 and SCJ1248. Aliquots of 2.5 μl of template from the dilution series was added to each reaction. The reactions were performed using an ABI 7700 instrument (Applied Biosystems, Foster City, CA) according to the following profile: 1 min at 95°C, 40 cycles of 10 s at 95°C, 20 s at 61°C and 1 min at 65°C. Real-time data was acquired at 61°C.

RESULTS

Non-natural base incorporation and melt analysis

We constructed a primer extension system to determine the identity of bases incorporated opposite iG under a variety of conditions. Parameters evaluated included nucleoside concentrations, buffering conditions and polymerase. In this system we incorporated a fluorescence-quencher-modified diGTP on the opposite DNA strands from a fluorescein label. Thermal melt analysis following primer extension separates the two strands, yielding the melting temperature (T_m). Comparison of the product T_m to those of control constructs where the base identity is known identifies the string of nucleotide incorporated, most importantly the nucleotide across from the non-natural base. These experiments led us to the discovery that increasing the ratio of non-natural triphosphates to natural triphosphates and using TiTaq (a nuclease-deficient, N-terminal truncated mutant of *Thermus aquaticus* DNA polymerase) significantly reduced misincorporation opposite iG. These experiments also confirmed previous reports showing that thymidine was the major nucleoside incorporated opposite the iG when iCTP is withheld from the reaction, supporting the iG:T tautomer theory (Fig. 1).

Confirmation of high polymerase fidelity

We next used PCR and an acid cleavage partial sequencing method to determine polymerase fidelity. It is well known that iC is more labile than naturally occurring bases under acid conditions (14). At low pH, nucleophilic substitution of the nucleobases followed by β elimination leads to strand cleavage (15). At positions where the nucleobase is iC, this rate is faster than that of the naturally occurring bases. We used this information to monitor percentage of iC maintained in the final amplicon. Two 88 nt templates (1 and 2) were synthesized containing either two non-natural bases or none. The iC and iG in Template 1 were changed to T and A, respectively, in Template 2 since our primer extension experiments suggested that this would be the resolution of the iC:iG base pair if misincorporation were to occur. All reactions were cycled 40 times and included the six base nucleotide triphosphate mix (2-fold higher concentration of diCTP and diGTP over A, G, C and T), TiTaq and a forward primer fluorescently labeled with Cy3. Products were subjected to acidic and then basic conditions to cleave the strands primarily at positions containing the iC. Cleavage products were resolved using PAGE and percent cleaved was calculated by densitometry (Fig. 2). For reactions containing the non-natural bases, percent cleavage was plotted versus input template. The slope of the line (3.86) is the decrease in percent cleaved after each cycle, which indicates that the fidelity of the iC:iG pair is ~96%.

Molecular beacon analysis of PCR amplified products

To demonstrate commercial applicability, we used a popular methodology termed 'molecular beacons' to identify the build-up of iC:iG-containing amplicons during PCR (16). Molecular beacons are biosensors for the detection of specific DNA sequences. Specifically, molecular beacons are single-stranded DNA structures that contain a pair of interacting

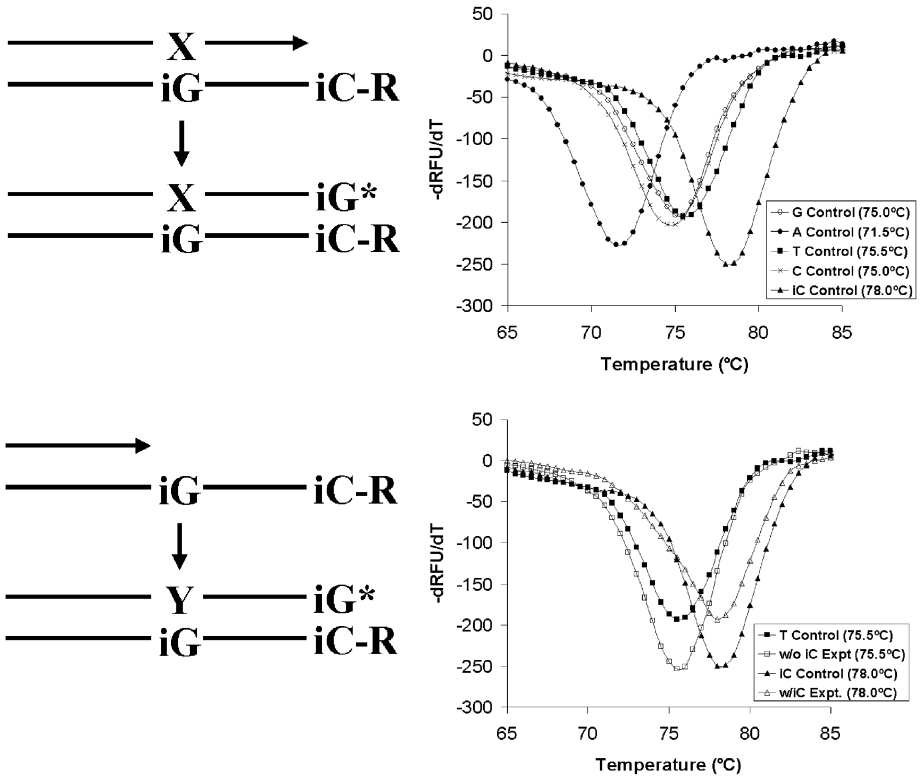

Figure 1. Primer extension and melt analysis. On the left, the diagrams show where the 3′-end of primer hybridization to the template ended (after the iG in controls or before the iG for experimental). On the right, melt analysis of primer extension products where bases opposite iG were controlled (**X** = A,G,C,T,iC) or experimentally determined (**Y**) is shown. All reactions were treated with TiTaq, dG,dA,dT,dCTP, quencher-modified diGTP and diCTP when designated. R, tetrachlorofluorescein; *, Dabcyl.

fluorescent molecules that change fluorescence intensity once hybridized to their specific targets. Molecular beacons bind with high specificity and single base discrimination. These attributes appear ideal to discriminate between templates constructed from three base pairs and related templates constructed from two. In order to compare PCR efficiency, both templates were amplified using the same reaction mix which included the six base nucleotide triphosphate mix, TiTaq, a molecular beacon specific for Template 1 (non-natural base-containing target) and the identical primer set. The fluorescence output from the experiments was monitored and linear quantitation curves were constructed from the data (Fig. 3). PCR is an exponential amplification technology described by the formula $N = N_o(e)^{(K\ t)}$, where N is the of molecules, t is the time (rounds of PCR) and K is a constant, and thus a 10-fold increase in amplicon requires ~3.32 replication cycles for perfect doubling. Since we plotted the log of copy number to cycle threshold, a perfect doubling of each amplification round would provide a slope of ~0.301 [$(2^{1/0.301}) = 10$]. The slope of the line for Template 1 (slope = –0.28 ± 0.01) indicates that amplification was not perfect and 3.6 cycles were required to yield a 10-fold increase, representing an efficiency of ~93% (efficiency = $t_{perfect}/t_2 = 1/t_2 = -m/\log 2$, where m is the slope]. Comparing peak fluorescence output from the two reactions sets (1 versu 2), there appears to be a 3-fold increase when Template 1 was used, indicating that the beacon was specific for the correct target.

DISCUSSION

We have presented data demonstrating that an additional base pair can be implemented in the most widely used method in molecular biology, PCR. Earlier reports using the base pair (iC:iG) suggested that iG would not be useful since thymidine would misincorporate opposite iG due to a minor tautomeric form of iG that forms a mispair with T (17). From the data reported here, it appears that the efficiency of the six base PCR system is high. Both the acid cleavage partial sequencing and the beacon experiments point to an overall efficiency of ~96 ± 3%. We believe previous attempts fell short due to reasons other than tautomerization of the iG nucleotide. From our experience, other reasons seem more plausible and include: inappropriately low triphosphate concentrations; choice of polymerase; poor quality of iC/iG-containing molecules; reaction pH values that were too destabilizing to iC at high cycling temperatures; purification methods for templates and primers that were destabilizing to iC. We believe that we have now built a robust and reproducible system for expanded base pair chemistry using iC/iG.

So how will the expanded base chemistry be useful? In many commercial applications, DNA is involved in two contexts. In the first, DNA-like binding is used to perform tasks by which information on a molecular structure is obtained. In the second, DNA is the ‘analyte’, the target of the analysis, present in the biological sample in unknown amounts, sometimes having unknown sequence. DNA of the

second type can obstruct the performance of DNA of the first type. Therefore, a molecular system that behaves like natural DNA but does not cross-react with natural DNA would be useful. Paradoxically, if the new system shares structural similarity to DNA like the one presented here, the rich enzymology and technology already developed for DNA can be further exploited.

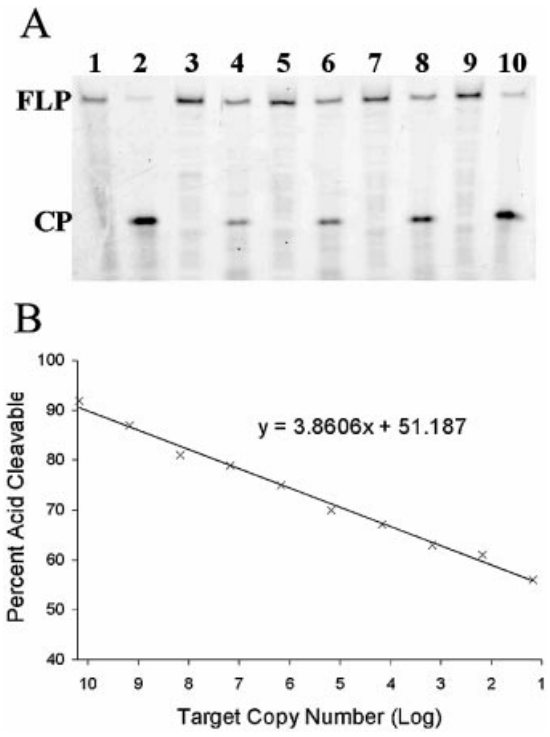

Figure 2. Acid cleavage analysis. (**A**) Acid cleavage products using natural (odd lanes) and non-natural (even lanes) DNA targets. Reactions were resolved by PAGE and detected by fluorescent imaging. Lanes 1 and 2 were loaded with extension reactions that were initiated with 10^{10} copies of input target. Lanes 3 and 4, 5 and 6, 7 and 8 and 9 and 10 were loaded with PCR reactions initiated with 1.5×10^{1}–10^{10} input targets in 1000-fold increments and amplified for 40 cycles. Bands that appear in the scan that were used to determine percent cleaved are labeled (FLP, full length product; CP, cleavage product). (**B**) Cleavage analysis. Using a series of 10-fold dilutions of the non-natural template and 40 cycles of PCR, the percent of cleaved product was determined as above and plotted against input copy number.

Consequently, the ability to place additional base pairs into DNA has several possible applications. Urdea and colleagues were the first to demonstrate diagnostic utility using branched DNA technology (2). Our group has recently demonstrated methods for simplifying real-time PCR using non-natural bases. Another application would be for molecular coding, where numerous technologies use DNA. Coding can be implemented in diagnostics to capture reaction products on solid phases (18–20), in tagging, where objects are marked with identifiable DNA segments called taggants, and in DNA computing (21). All would benefit from having extra bases orthogonal to naturally occurring DNA. Yet another application is in the field of aptamers, single-stranded and highly folded DNA and RNA molecules that, like antibodies, can bind target molecules with high affinity and specificity. Typically, aptamers are selected from large libraries of chemically synthesized oligonucleotides after multiple rounds of target binding and PCR amplification. (22) Catalytically active aptamers have been isolated with rate enhancement and turnover activities. (3) However, aptamers made of standard nucleic acids are somewhat limited in enzymatic capability and therefore modified nucleic acids have been explored (23,24). To expand on this, the use of multiple functional groups may be possible. Traditionally a single functionality is attached to uracil. This may be for enzymatic reasons, where placing chemical substituents on additional bases leads to premature termination. By increasing the number of building blocks, chemical substituents could be spaced further apart, thus allowing the polymerases to better process full-length sequences. The high fidelity of iC/iG incorporation demonstrated here is well within the range required for these applications since coding sequences used to capture iC/iG-containing products should not cross-react with misincorporated natural sequences and aptamer selection schemes would eliminate unfit sequences.

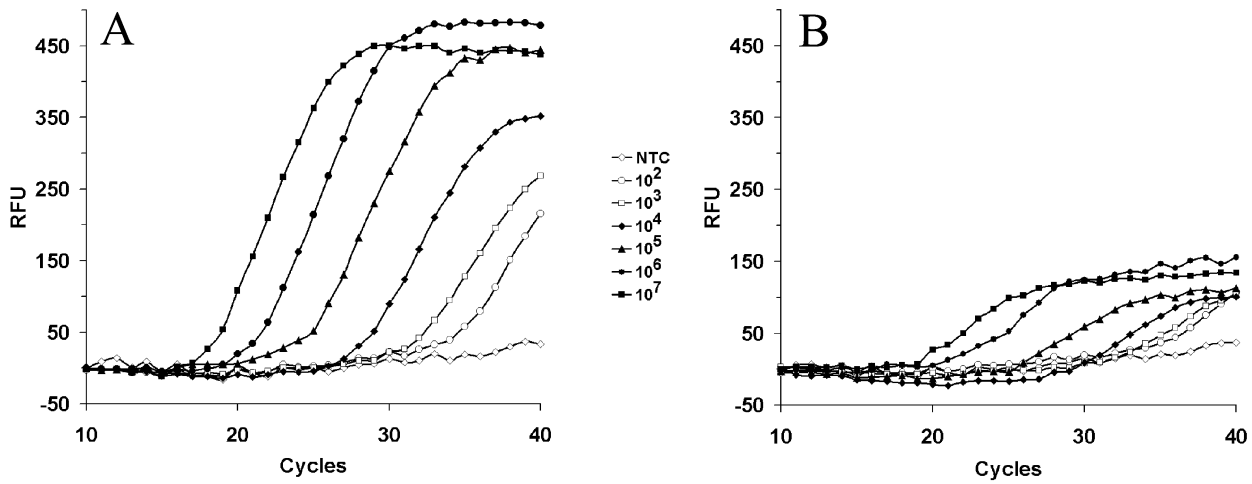

Figure 3. Molecular beacon analysis. Amplification analysis using reactions containing Template 1 (**A**) or Template 2 (**B**) and a molecular beacon specific for Template 1. Ten-fold dilutions of each target from 10^2 to 10^7 were constructed and added to the reaction series. The reactions were fluorescently monitored through 40 cycles of PCR.

The data presented should generate interest in not only creating new technologies using additional base pairs, but also in creating new organisms that rely on high fidelity enzymatic incorporation and specific enzyme recognition. Additional work to better understand the conditions necessary to use expanded genetic alphabets may also allow us to better understand how nature's alphabet evolved.

ACKNOWLEDGEMENTS

We would like to thank Dr Floyd Romesburg and Dr Steven Benner for helpful comments in drafting this article. We would also like to thank the NIH and the National Institute of General Medical Sciences for providing project support with STTR funding grant 5 R42 GM063368-03.

REFERENCES

1. Pease,A.C., Solas,D., Sullivan,E.J., Cronin,M.T., Holmes,C.P. and Fodor,S.P. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA*, **91**, 5022–5026.
2. Collins,M.L., Irvine,B., Tyner,D., Fine,E., Zayati,C., Chang,C., Horn,T., Ahle,D., Detmer,J., Shen,L.P. *et al.* (1997) A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res.*, **25**, 2979–2984.
3. Prudent,J.R., Uno,T. and Schultz,P.G. (1994) Expanding the scope of RNA catalysis. *Science*, **264**, 1924–1927.
4. Hermann,T. and Patel,D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
5. Piccirilli,J.A., Krauch,T., Moroney,S.E. and Benner,S.A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, **343**, 33–37.
6. Switzer,C.Y., Moroney,S.E. and Benner,S.A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.*, **111**, 8322–8323.
7. McMinn,D.L., Ogawa,A.K., Wu,Y., Liu,J. and Romesberg,F.E. (1999) Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. *J. Am. Chem. Soc.*, **121**, 11585–11586.
8. Tae,E.L., Wu,Y., Xia,G., Schultz,P.G. and Romesberg,F.E. (2001) Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *J. Am. Chem. Soc.*, **123**, 7439–7440.
9. Ohtsuki,T., Kimoto,M., Ishikawa,M., Mitsui,T., Hirao,I. and Yokoyama,S. (2001) Unnatural base pairs for specific transcription. *Proc. Natl Acad. Sci. USA*, **98**, 4922–4925.
10. Mitsui,T., Kitamura,A., Kimoto,M., To,T., Sato,A., Hirao,I. and Yokoyama,S. (2003) An unnatural hydrophobic base pair with shape complementarity between pyrrole-2-carbaldehyde and 9-methylimidazo[(4,5)-b]pyridine. *J. Am. Chem. Soc.*, **125**, 5298–5307.
11. Moser,M.J. and Prudent,J.R. (2003) Enzymatic repair of an expanded genetic information system. *Nucleic Acids Res.*, **31**, 5048–5053.
12. Clelland,C.T., Risca,V. and Bancroft,C. (1999) Hiding messages in DNA microdots. *Nature*, **399**, 533–534.
13. Moser,M.J., Marshall,D.J., Grenier,J.K., Kieffer,C.D., Killeen,A.A., Ptacin,J.L., Richmond,C.S., Roesch,E.B., Scherrer,C.W., Sherrill,C.B. *et al.* (2003) Exploiting the enzymatic recognition of an unnatural base pair to develop a universal genetic analysis system. *Clin. Chem.*, **49**, 407–414.
14. Voegel,J.J. and Benner,S.A. (1996) Synthesis, molecular recognition, and enzymology of oligonucleotides containing the non-standard base pair between 5-aza-7-deazaisoguanine and 6-amino-3-methylpyrazin-2(1H)-one, a donor-acceptor-acceptor purine analog and an acceptor-donor-donor pyrimidine analog. *Helv. Chim. Acta*, **76**, 1881–1898.
15. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
16. Tyagi,S. and Kramer,F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, **14**, 303–308.
17. Switzer,C.Y., Moroney,S.E. and Benner,S.A. (1993) Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry*, **32**, 10489–10496.
18. Landegren,U., Kaiser,R., Sanders,J. and Hood,L. (1988) A ligase-mediated gene detection technique. *Science*, **241**, 1077–1180.
19. Oliphant,A., Barker,D.L., Stuelpnagel,J.R. and Chee,M.S. (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques*, Suppl., 56–58 and 60–61.
20. Chen,J., Iannone,M.A., Li,M.S., Taylor,J.D., Rivers,P., Nelsen,A.J., Slentz-Kesler,K.A., Roses,A. and Weiner,M.P. (2000) A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.*, **10**, 549–557.
21. Gibbons,A., Amos,M. and Hodgson,D. (1997) DNA computing. *Curr. Opin. Biotechnol.*, **8**, 103–106.
22. Ellington,A.D. and Szostak,J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
23. Battersby,T.R., Ang,D.N., Burgstaller,P., Jurczyk,S.C., Bowser,M.T., Buchanan,D.D., Kennedy,R.T. and Benner,S.A. (1999) Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog. *J. Am. Chem. Soc.*, **121**, 9781–9789.
24. Santoro,S.W., Joyce,G.F., Sakthivel,K., Gramatikova,S. and Barbas,C.F.,III (2000) RNA cleavage by a DNA enzyme with extended chemical functionality. *J. Am. Chem. Soc.*, **122**, 2433–2439.