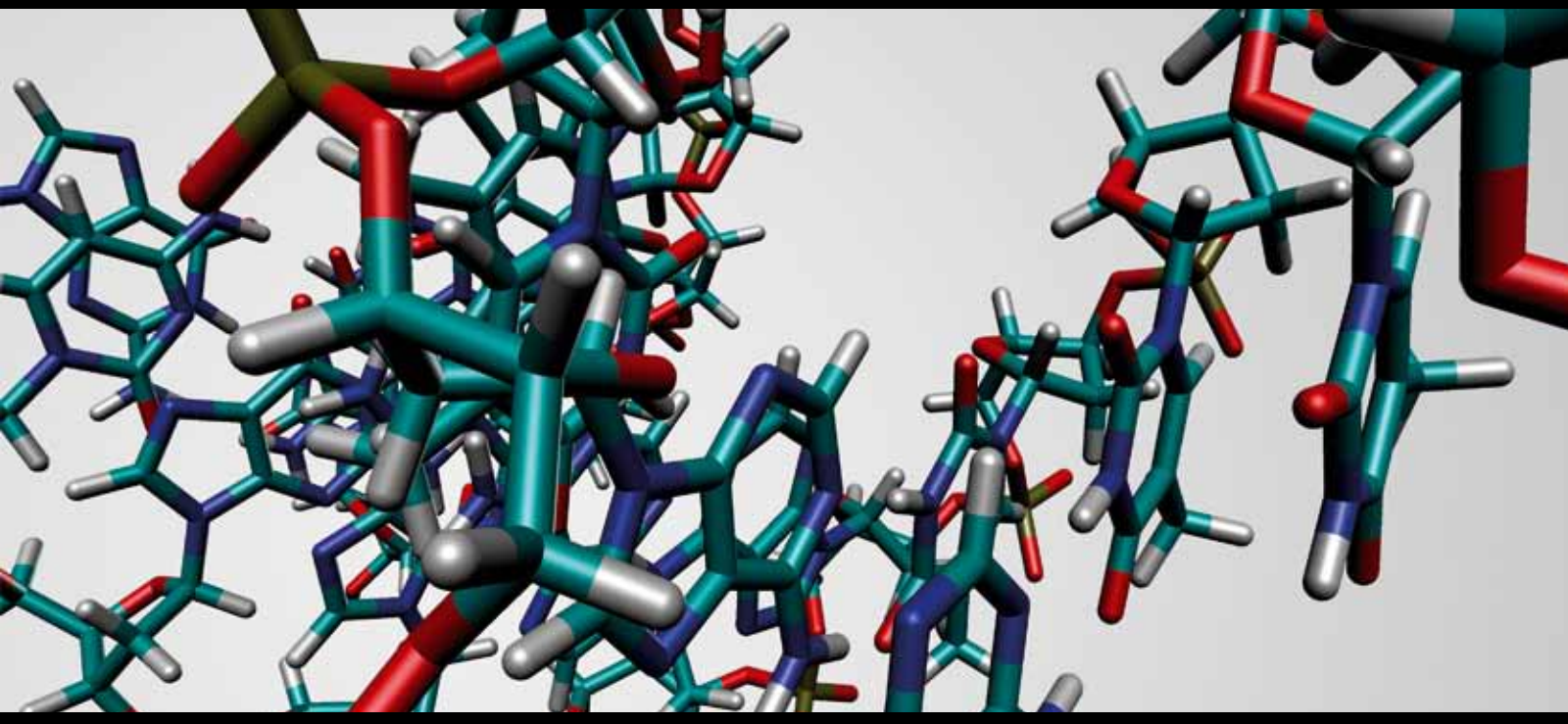
Journal of Nucleic Acids

# Expansion of the Genetic Alphabet: Unnatural Nucleobases and Their Applications

Guest Editors: Subhendu Sekhar Bag, Jennifer M. Heemstra, Yoshio Saito, and David M. Chenoweth



# Expansion of the Genetic Alphabet: Unnatural Nucleobases and Their Applications

Journal of Nucleic Acids

# Expansion of the Genetic Alphabet: Unnatural Nucleobases and Their Applications

Guest Editors: Subhendu Sekhar Bag, Jennifer M. Heemstra, Yoshio Saito, and David M.Chenoweth

# Editorial Board

# Contents

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 718582, 2 pages
doi:10.1155/2012/718582

*Editorial*

# Expansion of the Genetic Alphabet: Unnatural Nucleobases and Their Applications

**Subhendu Sekhar Bag,[1] Jennifer M. Heemstra,[2] Yoshio Saito,[3] and David M. Chenoweth[4]**

[1] *Department of Chemistry, Indian Institute of Technology Guwahati, Gumahati 781039, India*
[2] *Department of Chemistry and Center for Cell & Genome Science, The University of Utah, Salt Lake City, UT 84112, USA*
[3] *Department of Chemical Biology and Applied Chemistry, School of Engineering, Nihon University, Koriyama, Fukushima 963-8642, Japan*
[4] *Department of Chemistry, Roy and Diana Vagelos Laboratories (IAST), University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104, USA*

Correspondence should be addressed to Subhendu Sekhar Bag, ssbag75@iitg.ernet.in

Received 22 November 2012; Accepted 22 November 2012



Nucleic acids are essential biomolecules that encode all of the information necessary for life. Specific pairing of A with T (or U) and C with G during replication, transcription, and translation is the key to effective transmission of genetic information between generations, as well as accurate conversion of genetic information into protein sequence. Given the magnitude of the tasks orchestrated by the Watson-Crick base pairing, it is striking to consider that biological systems accomplish these tasks using only four nucleobases. Realizing the powerful nature of base-pair recognition, researchers have been inspired to ask the question of whether the genetic code can be artificially expanded to generate biological systems having novel functions. It was this question that led Alex Rich in 1962 to propose the concept of orthogonal base pairing between *iso*-G and *iso*-C and inspired Professor Steven A. Benner in the late 1980s to expand the genetic alphabet from four to six letters. Benner's early research focused on the development of new base pairs having hydrogen bonding patterns orthogonal to those in the canonical Watson-Crick base pairs. In 1994, Professor Eric T. Kool opened a new functional dimension with the creation of nonhydrogen bonding unnatural nucleobase surrogates.

Expansion of the genetic alphabet has dramatically increased the functional potential of DNA, for example, by enabling site-directed oligonucleotide labeling and *in vitro* selections with oligonucleotides having increased chemical diversity. Translation of an expanded DNA alphabet into RNA is a challenging task, but one which has potential to give rise to semisynthetic organisms with increased biodiversity. This special issue highlights recent accomplishments at the interface of organic chemistry and molecular biology which hold promise to further expand the potential of nucleic acids having unnatural nucleobases. Specifically, the reports in this special issue focus on the synthesis of unnatural nucleobases and nucleic acid backbones, the exploration of their structure and duplex stabilizing ability, and the polymerase mediated replication and transcription of DNA containing unnatural nucleobases.

T. Lönnberg and a coworker report the synthesis and study of a bis(pyrazolyl)purine ribonucleoside having increased hydrophobic surface area and the ability to form complex with metal ions. The hydrogen bonding pattern of this nucleoside makes it complementary to thymine and uridine. The authors demonstrate that the bis(pyrazolyl) nucleobase is capable of forming a $Pd^{2+}$-mediated base pair with uridine in the monomeric state. When incorporated into an oligonucleotide, the bis(pyrazolyl) nucleobase stabilizes DNA duplexes when paired with thymine, but this stabilization appears to result from increased $\pi$-stacking interactions rather than metal complexation. These studies open the door to applications using unnatural nucleobases to increase the binding affinity of probes and therapeutics targeted at native DNA and RNA.

Much effort has focused on the incorporation of unnatural nucleobases into native DNA and RNA, and the availability of nonnative backbones such as LNA, PNA, and

GNA opens the door for further expansion of nucleic acid structure and function. C. Förster and coworkers evaluate the structural properties of "all LNA" duplexes and demonstrate that these duplexes have a relaxed helical structure unique from that of DNA and RNA. J. M. Heemstra and coworkers report an expedient synthetic route to the Fmoc-protected PNA backbone, which serves as a key intermediate in the synthesis of PNA monomers having unnatural nucleobases. Together, these studies facilitate the exploration of nucleic acids having both nonnative backbones and unnatural nucleobases, which is in turn anticipated to provide molecules having novel structural and functional properties.

The ability to use native or engineered replication and transcription machinery with nucleic acids containing unnatural nucleobases is critical to many *in vitro* and *in vivo* applications. I. Hirao and coworkers describe a two-unnatural-base-pair system capable of incorporating unnatural nucleobases into DNA via polymerase chain reaction, then into RNA via T7 transcription. Using this system, they demonstrate sequence-specific incorporation of a biotinylated nucleobase into a 260-mer RNA sequence. Generating modified RNA of this length is not feasible using purely synthetic methods; thus this work significantly improves access to site-specifically labeled RNA sequences. Additionally, P. J. Beuning and a coworker review the use of native A and Y family DNA polymerases for replication of DNA containing unnatural nucleobases. This review serves as an excellent resource for researchers seeking to utilize unnatural nucleobases and provides significant insight into how polymerases deal with both synthetic and damaged DNA nucleobases.

Collectively, these reports demonstrate the potential of unnatural nucleobases for applications in chemistry and biology and offer valuable tools for further exploring nucleic acid structural diversity.

## Acknowledgments

We would like to thank the authors and reviewers for their contributions to this special issue.

*Subhendu Sekhar Bag*
*Jennifer M. Heemstra*
*Yoshio Saito*
*David M. Chenoweth*

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 354549, 5 pages
doi:10.1155/2012/354549

*Research Article*

# Convenient and Scalable Synthesis of Fmoc-Protected Peptide Nucleic Acid Backbone

**Trevor A. Feagin, Nirmal I. Shah, and Jennifer M. Heemstra**

*Department of Chemistry and the Center for Cell & Genome Science, University of Utah, Salt Lake City, UT 84112, USA*

Correspondence should be addressed to Jennifer M. Heemstra, heemstra@chem.utah.edu

Received 1 May 2012; Accepted 11 June 2012

Academic Editor: Subhendu Sekhar Bag



The peptide nucleic acid backbone Fmoc-AEG-OBn has been synthesized via a scalable and cost-effective route. Ethylenediamine is mono-Boc protected, then alkylated with benzyl bromoacetate. The Boc group is removed and replaced with an Fmoc group. The synthesis was performed starting with 50 g of Boc anhydride to give 31 g of product in 32% overall yield. The Fmoc-protected PNA backbone is a key intermediate in the synthesis of nucleobase-modified PNA monomers. Thus, improved access to this molecule is anticipated to facilitate future investigations into the chemical properties and applications of nucleobase-modified PNA.

## 1. Introduction

Peptide nucleic acid (PNA) [1] has recently emerged as a promising alternative to the native nucleic acids DNA and RNA (Figure 1) for a wide variety of applications including antisense therapy [2] and gene diagnostics [3]. The key advantages of PNA over DNA and RNA are its resistance to degradation by cellular nucleases [4] and its relatively higher binding affinity and mismatch selectivity in duplex formation [5]. PNA can be generated by Fmoc- or Boc-solid phase peptide synthesis [6], and Fmoc-protected monomers bearing each of the four canonical nucleobases are commercially available. Recently, the incorporation of modified nucleobases into PNA has been shown to enable synthesis of nucleic acids having unique physicochemical properties [7]. However, PNA monomers bearing modified nucleobases are not commercially available, and must instead be synthesized in the laboratory. Suitable reactions have been reported for preparation of modified nucleobases and coupling of these nucleobase acetic acids to the PNA backbone (Figure 2) [7–9]. However, to our knowledge, a scalable and cost-effective synthesis for the protected *N*-[2-(Fmoc)aminoethyl]glycine benzyl ester (Fmoc-AEG-OBn) backbone **1** has yet to be reported. Synthesis of the Fmoc-protected carboxylic acid backbone Fmoc-AEG-OH has been reported [10], and coupling of nucleobase acetic acids with Fmoc-AEG-OH has been described in the patent literature [11]. However, this coupling reaction provides moderate-to-low yields of PNA monomer [12, 13]. Here, we describe a synthesis of **1** that proceeds in four steps with an overall yield of 32%, utilizes inexpensive reagents, and can be scaled to produce large quantities of final product in a single batch with only minimal purification.

## 2. Materials and Methods

*2.1. General Methods.* Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification. Flash column chromatography was carried out using silica gel 60 (230–400 mesh). $^1$H and $^{13}$C NMR chemical shifts are expressed in parts per million ($\delta$) using residual solvent protons as internal standard ($\delta$ 7.26 ppm ($^1$H) and 77.16 ppm ($^{13}$C) for $CHCl_3$). Coupling constants, $J$, are reported in Hertz (Hz), and splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br (broad), and app (apparent). Mass spectra were obtained through the Mass Spectrometry Facility, University of Utah.

*2.2. tert-Butyl(2-aminoethyl)carbamate (6).* A 2 L round bottom flask was charged with ethylenediamine (306.5 mL,

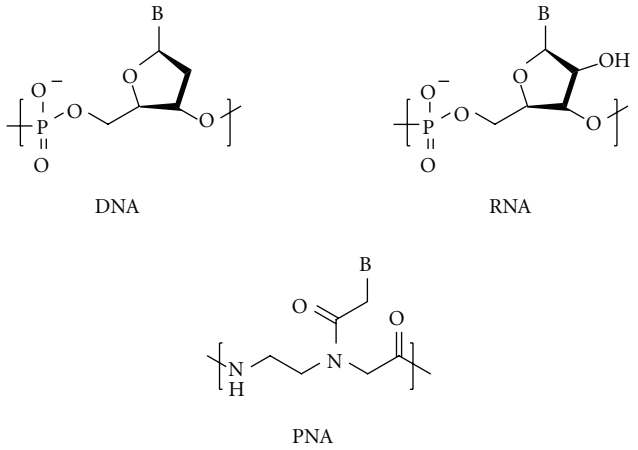



Figure 1: Chemical structure of DNA, RNA, and PNA.

4.58 mol) and tetrahydrofuran (600 mL). Boc anhydride (50.0 g, 229 mmol) was dissolved in 400 mL tetrahydrofuran and added to the solution of ethylenediamine via addition funnel over 45 min with vigorous stirring. After 18 h, the reaction was quenched by addition of 500 mL $H_2O$. The aqueous phase was saturated with solid $K_2CO_3$, then the phases were separated and the organic phase was dried over $Na_2SO_4$, filtered, and concentrated to give a pale yellow oil. The oil was dissolved in 700 mL toluene and concentrated to azeotrope remaining ethylenediamine, yielding 29.24 g of pale yellow oil (80%). $^1$H NMR (300 MHz, $CDCl_3$) $\delta$ 4.90 (br s, 1H), 3.16 (q, $J$ = 5.8 Hz, 2H), 2.78 (t, $J$ = 5.9 Hz, 2H), 1.43 (s, 9H), 1.09 (br s, 2H). $^{13}$C NMR (75 MHz, $CDCl_3$) $\delta$ 156.3, 79.1, 43.3, 41.8, 28.4. HRMS (ESI) $m/z$ 161.1295 (calcd $[M + H]^+$ = 161.1290).

*2.3. Benzyl 2-((2-((tert-butoxycarbonyl)amino)ethyl)amino) acetate (9).* A 2 L round bottom flask was charged with **6** (29.24 g, 183 mmol), triethylamine (25.5 mL, 183 mmol), and acetonitrile (450 mL). The reaction mixture was stirred and ethyl bromoacetate (28.9 mL, 182 mmol) added via syringe over 2 min. After 100 min, the reaction mixture was diluted with 500 mL EtOAc and washed with 500 mL 2 M $K_2CO_3$ (aq.), then 500 mL brine. The organic phase was dried over $Na_2SO_4$, filtered, and concentrated to a pale yellow oil, which was purified by flash column chromatography, (70 mm diameter column, 285 g silica gel, 1 : 1 hexanes : EtOAc, 99 : 1 EtOAc : $Et_3N$) to give 40.56 g of pale yellow oil (72%). $^1$H NMR (300 MHz, $CDCl_3$) $\delta$ 7.38–7.33 (m, 5H), 5.17 (s, 2H), 4.99 (br s, 1H), 3.45 (s, 2H), 3.20 (q, $J$ = 5.7 Hz, 2H), 2.74 (t, $J$ = 5.8 Hz, 2H), 1.59 (br s, 1H), 1.44 (s, 9H). $^{13}$C NMR (75 MHz, $CDCl_3$) $\delta$ 172.3, 156.1, 135.5, 128.6, 128.4, 128.3, 79.0, 66.5, 50.4, 48.7, 40.1, 28.4. HRMS (ESI) $m/z$ 331.1640 (calcd $[M + Na]^+$ = 331.1634).

*2.4. Benzyl 2-((2-((((9H-fluoren-9-yl)methoxy)carbonyl) amino)ethyl)amino)acetate (Fmoc-AEG-OBn 1).* To a 2 L round bottom flask was added **9** (40.46 g, 131 mmol) and dichloromethane (200 mL). The reaction mixture was stirred in an ice bath and 200 mL trifluoroacetic acid added. The ice bath was removed and the reaction mixture stirred for 20 min, then concentrated to a yellow oil. The oil was dissolved in 400 mL toluene and concentrated to azeotrope remaining trifluoroacetic acid. The resulting yellow oil was dissolved in 700 mL dichloromethane and stirred under $N_2$ in an ice bath. Fmoc-OSu (44.26 g, 131 mmol) was added all at once, then triethylamine (54.8 mL, 393 mmol) was added dropwise via addition funnel over 5 min. The ice bath was removed and the reaction mixture was stirred for 2 h, then washed with 500 mL 1 M $K_2CO_3$ (aq.), dried over $Na_2SO_4$, filtered, and concentrated to a yellow oil. The oil was purified by flash column chromatography, (70 mm diameter column, 250 g silica gel, 1 : 1 hexanes : EtOAc, 98 : 2 EtOAc : MeOH) to give 31.04 g of pale yellow oil that crystallized into a white solid upon standing (55%). $^1$H NMR (300 MHz, $CDCl_3$) $\delta$ 7.76 (d, $J$ = 7.4 Hz, 2H), 7.61 (d, $J$ = 7.3 Hz, 2H), 7.42–7.28 (m, 9H), 5.29 (br s, 1H), 5.18 (s, 2H), 4.40 (d, $J$ = 7.0 Hz, 2H), 4.22 (t, $J$ = 6.8 Hz, 1H), 3.47 (s, 2H), 3.29 (q, $J$ = 5.4 Hz, 2H), 2.78 (t, $J$ = 5.6 Hz, 2H), 1.58 (br s, 1H). $^{13}$C NMR (75 MHz, $CDCl_3$) $\delta$ 172.3, 156.6, 144.0, 141.3, 135.5, 128.6, 128.5, 128.4, 127.6, 127.0, 125.1, 120.0, 66.6, 66.4, 50.4, 48.6, 47.3, 40.6. HRMS (ESI) $m/z$ 431.1972 (calcd $[M + H]^+$ = 431.1971).

## 3. Results

An extensive literature search reveals only one published route to **1** as shown in Figure 3(a). This route, reported by Hudson and coworkers, proceeds in three steps with an overall yield of 81%, but requires the use of costly *N*-(2-aminoethyl)glycine as the starting material [8]. Furthermore, in our hands, this synthetic route has proven difficult to reproduce. Alternatively, Porcheddu and coworkers have described a synthesis of Fmoc-AEG-OMe **2** that proceeds in three steps with an overall yield of 66% (Figure 3(b)), but requires three equivalents of IBX to accomplish the oxidation step [9]. Thomson and coworkers have described a synthesis of Fmoc-AEG-O*t*-Bu·HCl **3·HCl** that proceeds in two steps with 46% overall yield from inexpensive starting materials (Figure 3(c)), but does not produce analytically pure material [14]. Furthermore, the exocyclic amines of the PNA nucleobases are typically protected with acid-labile Boc or Bhoc protecting groups. Deprotection of the *tert*-butyl ester requires strongly acidic conditions, making **3** unsuitable for use with Boc- or Bhoc-protected nucleobases.

Inspired by the ease and cost-effectiveness of the Thomson route, we first envisioned synthesis of **1** by alkylation of ethylenediamine with benzyl bromoacetate, followed by reaction with Fmoc *N*-hydroxysuccinimide ester (Fmoc-OSu). Unfortunately, after alkylation and aqueous workup, only benzyl alcohol was recovered. We hypothesize that alkylation product **4** rapidly cyclizes to give benzyl alcohol and water soluble cyclic piperazinone **5** (Figure 4(a)), and that this undesired cyclization can only be suppressed by the use of a bulky ester such as a *tert*-butyl ester. In fact, this cyclization has been demonstrated previously using the analogous methyl ester [15]. We next considered reversing the order of the two reactions such that ethylenediamine

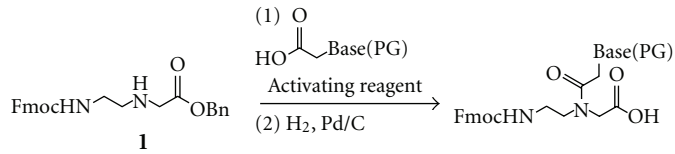



Figure 2: Synthesis of Fmoc-protected PNA monomers.

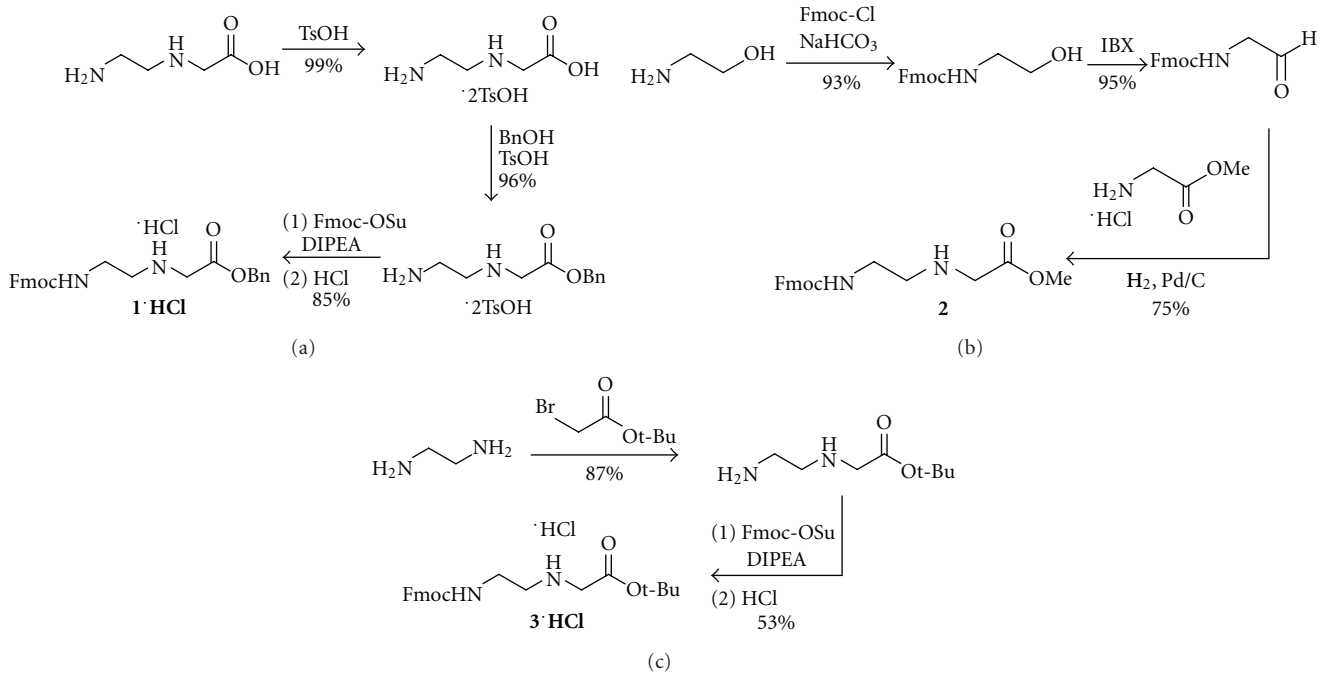



Figure 3: Reported synthetic routes to the Fmoc-AEG-OR backbone.

would first be monoprotected with Fmoc-OSu, then alkylated with benzyl bromoacetate to give **1**. However, Fmoc-ethylenediamine cannot be directly prepared by the reaction of ethylenediamine with Fmoc-Cl or Fmoc-OSu. Rather, a three-step process is required in which ethylenediamine is mono-Boc protected (**6**), then Fmoc protected (**7**), and finally the Boc group is removed under acidic conditions to give **8** as the TFA salt [16]. Unfortunately, our attempts to alkylate **8**·**TFA** with benzyl bromoacetate failed to yield the desired product **1**, likely due to the instability of the free base of **8** (Figure 4(b)).

Fortunately, we were able to obtain Boc-ethylenediamine **6** in 80% yield from ethylenediamine and Boc anhydride using a modified version of a reported procedure [17], and this was successfully alkylated with benzyl bromoacetate to give **9** in 72% yield. We then deprotected the Boc group using trifluoroacetic acid (TFA) to give a quantitative yield of free amine, which was importantly found to be stable to cyclization when isolated as the TFA salt. In the final step, we combined the amine TFA salt with Fmoc-OSu prior to adding base, so that protection of the primary amine could compete with cyclization to give the desired product **1** in 55% yield. Starting with 50 g of Boc anhydride, we were able to generate 31 g of analytically pure **1** in a single batch using inexpensive reagents (Figure 4(c)) [18].

A key to the scalability of our synthetic route is the relatively facile purification of the synthetic intermediates and final product. The Boc protection step to give **6** requires only aqueous workup, and the deprotection step requires simple concentration and removal of TFA via formation of an azeotrope with toluene. The alkylation to produce **9** and the Fmoc protection to give **1** require flash column chromatography, but a large difference in $R_f$ between the products and impurities makes purification possible using only a silica plug.

## 4. Discussion

Fmoc-protected PNA backbone **1** is a key intermediate in the synthesis of Fmoc-protected PNA monomers having modified nucleobases. However, to date, a scalable and cost-effective synthetic route to this molecule has yet to be reported in the literature. An efficient synthesis of the Boc-protected backbone has been reported, but our attempts to

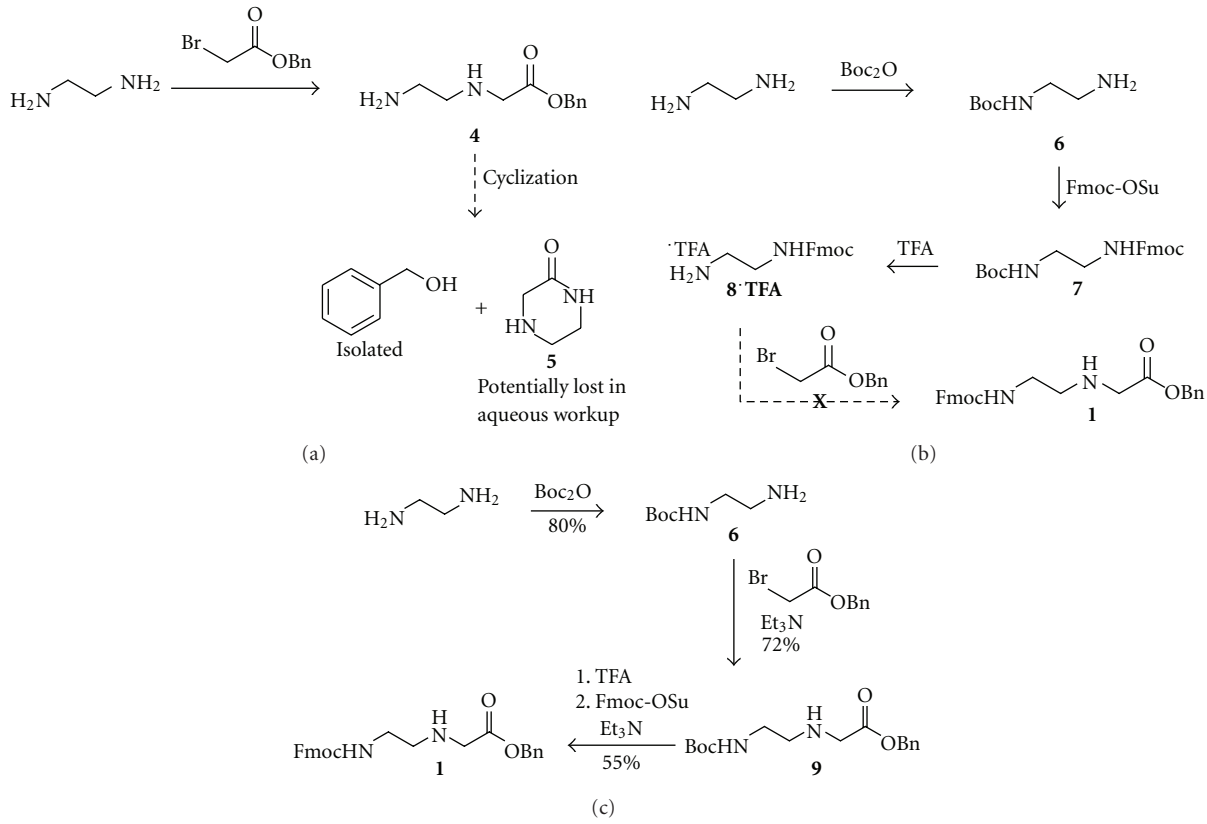



Figure 4: Synthetic route to Fmoc-AEG-OBn 1.

utilize this synthetic route with Fmoc in place of Boc failed to give product, likely due to the instability of synthetic intermediate **8**. Rather, synthesis of **1** can be initiated using a Boc protecting group, followed by a protecting group swap to provide the Fmoc-protected product. The first two steps of our synthetic route mirror those of the published synthesis for the Boc-protected monomer [19]. However, replacement of the Boc group with Fmoc poses a significant challenge, as this step proceeds through unstable intermediate **4**. We were able to perform this transformation by generating the free base of **4** at reduced temperature and in the presence of Fmoc-OSu, enabling Fmoc protection to effectively compete with cyclization, providing **1** in moderate yield.

In summary, we describe here a novel route to the PNA backbone Fmoc-AEG-OBn **1**. Using this route, we have rapidly synthesized 31 g of **1** using inexpensive starting materials and only minimal purification. The overall yield for our synthetic route is modest at 32%; however, the low cost of starting materials and ease of purification enable this synthesis to be tractable on a large scale. Having a convenient route to access **1** is anticipated to ease the synthesis of new Fmoc-protected PNA monomers, presumably furthering the exploration of PNA having unique modified nucleobases.

## Acknowledgment

The authors gratefully acknowledge financial support from the University of Utah.

## References

[1] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt, "Sequence-selective recognition of DNA by strand displacement with thymine-substituted polyamide," *Science*, vol. 254, no. 5037, pp. 1497–1500, 1991.

[2] L. Good and P. E. Nielsen, "Antisense inhibition of gene expression in bacteria by PNA targeted to mRNA," *Nature Biotechnology*, vol. 16, no. 4, pp. 355–358, 1998.

[3] P. E. Nielsen, "Peptide nucleic acid: a versatile tool in genetic diagnostics and molecular biology," *Current Opinion in Biotechnology*, vol. 12, no. 1, pp. 16–20, 2001.

[4] C. Gambacorti-Passerini, L. Mologni, C. Bertazzoli et al., "In vitro transcription and translation inhibition by anti-promyelocytic leukemia (PML)/retinoic acid receptor $\alpha$ and anti-PML peptide nucleic acid," *Blood*, vol. 88, no. 4, pp. 1411–1417, 1996.

[5] M. Egholm, O. Buchardt, L. Christensen et al., "PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules," *Nature*, vol. 365, no. 6446, pp. 566–568, 1993.

[6] D. A. Braasch, C. J. Nulf, and D. A. Corey, "Synthesis and purification of peptide nucleic acids," in *Current Protocols in Nucleic Acid Chemistry*, chapter 4, pp. 4.11.1–4.11.18, 2002.

[7] F. Wojciechowski and R. H. E. Hudson, "Nucleobase modifications in peptide nucleic acids," *Current Topics in Medicinal Chemistry*, vol. 7, no. 7, pp. 667–679, 2007.

[8] F. Wojciechowski and R. H. E. Hudson, "A convenient route to N-[2-(Fmoc)aminoethyl]glycine esters and PNA oligomerization using a Bis-N-Boc nucleobase protecting group strategy," *Journal of Organic Chemistry*, vol. 73, no. 10, pp. 3807–3816.

[9] A. Porcheddu, G. Giacomelli, I. Piredda, M. Carta, and G. Nieddu, "A practical and efficient approach to PNA monomers compatible with Fmoc-mediated solid-phase synthesis protocols," *European Journal of Organic Chemistry*, no. 34, pp. 5786–5797, 2008.

[10] E. Lioy, J. Suarez, F. Guzmàn, S. Siegrist, G. Pluschke, and M. E. Patarroyo, "Synthesis, biological, and immunological properties of cyclic peptides from *Plasmodium falciparum* merozoite surface protein-1," *Angewandte Chemie International Edition*, vol. 40, no. 14, pp. 2631–2635, 2001.

[11] J. M. Coull, M. Egholm, R. P. Hodge, M. Ismail, and S. B. Rajur, "Synthons for the synthesis and deprotection of peptide nucleic acids under mild conditions," U.S. Patent 6, 172, 226, 2001.

[12] F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran, and N. Winssinger, "Expanding the scope of PNA-encoded libraries: divergent synthesis of libraries targeting cysteine, serine and metalloproteases as well as tyrosine phosphatases," *Tetrahedron*, vol. 63, no. 28, pp. 6577–6586, 2007.

[13] F. Altenbrunn and O. Seitz, "*O*-Allyl protection in the Fmoc-based synthesis of difficult PNA," *Organic & Biomolecular Chemistry*, vol. 6, no. 14, pp. 2493–2498, 2008.

[14] S. A. Thomson, J. A. Josey, R. Cadilla et al., "Fmoc mediated synthesis of peptide nucleic acids," *Tetrahedron*, vol. 51, no. 22, pp. 6179–6194, 1995.

[15] R. W. Hay, A. K. Basak, and M. P. Pujari, "Kinetics and mechanism of the copper(II) promoted hydrolysis of the methyl ester of ethylenediaminemonoacetate," *Transition Metal Chemistry*, vol. 11, no. 1, pp. 27–30, 1986.

[16] P. Kocis, O. Issakova, N. F. Sepetov, and M. Lebl, "Kemp's triacid scaffolding for synthesis of combinatorial nonpeptide uncoded libraries," *Tetrahedron Letters*, vol. 36, no. 37, pp. 6623–6626, 1995.

[17] P. Xu, T. Zhang, W. Wang, X. Zou, X. Zhang, and Y. Fu, "Synthesis of PNA monomers and dimers by Ugi four-component reaction," *Synthesis*, no. 8, pp. 1171–1176, 2003.

[18] The free base of 1 is stable for several days at $-20^{\circ}$C, and can be easily converted to the HCl salt if long-term storage is desired.

[19] L. D. Fader, M. Boyd, and Y. S. Tsantrizos, "Backbone modifications of aromatic peptide nucleic acid (APNA) monomers and their hybridization properties with DNA and RNA," *Journal of Organic Chemistry*, vol. 66, no. 10, pp. 3372–3379, 2001.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 230943, 8 pages
doi:10.1155/2012/230943

*Research Article*

# PCR Amplification and Transcription for Site-Specific Labeling of Large RNA Molecules by a Two-Unnatural-Base-Pair System

**Michiko Kimoto,[1,2] Rie Yamashige,[1] Shigeyuki Yokoyama,[1,3] and Ichiro Hirao[1,2]**

[1] *RIKEN Systems and Structural Biology Center (SSBC), 1-7-22 Suehiro-cho, Tsurumi-ku, Kanagawa, Yokohama, 230-0045, Japan*
[2] *TAGCyx Biotechnologies, 1-6-126 Suehiro-cho, Tsurumi-ku, Kanagawa, Yokohama, 230-0045, Japan*
[3] *Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

Correspondence should be addressed to Ichiro Hirao, ihirao@riken.jp

Received 1 March 2012; Accepted 16 April 2012

Academic Editor: Subhendu Sekhar Bag



For the site-specific labeling and modification of RNA by genetic alphabet expansion, we developed a PCR and transcription system using two hydrophobic unnatural base pairs: 7-(2-thienyl)-imidazo[4,5-*b*]pyridine (**Ds**) and 2-nitro-4-propynylpyrrole (**Px**) as a third pair for PCR amplification and **Ds** and pyrrole-2-carbaldehyde (**Pa**) for the incorporation of functional components as modified **Pa** bases into RNA by T7 transcription. To prepare **Ds**-containing DNA templates with long chains, the **Ds**-**Px** pair was utilized in a fusion PCR method, by which we demonstrated the synthesis of 282-bp DNA templates containing **Ds** at specific positions. Using these **Ds**-containing DNA templates and a biotin-linked **Pa** substrate (Biotin-**Pa**TP) as a modified **Pa** base, 260-mer RNA transcripts containing Biotin-**Pa** at a specific position were generated by T7 RNA polymerase. This two-unnatural-base-pair system, combining the **Ds**-**Px** and **Ds**-**Pa** pairs with modified **Pa** substrates, provides a powerful tool for the site-specific labeling and modification of desired positions in large RNA molecules.

## 1. Introduction

Site-specific labeling and modification of large RNA molecules provide a wide variety of applications in many areas, such as biochemical and biophysical studies, synthetic biology, *in vitro* evolution, generation of functional nucleic acids, and construction of nanomaterials and biosensors. The site-specific incorporation of functional nucleotide analogs into RNA molecules is performed by chemical synthesis, post-transcriptional modification, and enzymatic incorporation of nucleotide analogs as substrates. Among them, chemical synthesis is a commonly employed method. However, it is only capable of synthesizing small RNA molecules. Other site-specific modifications of RNA, by posttranscriptional modification [1, 2] and enzymatic incorporation using cap or triphosphate analogs [3–6], are limited to terminal modifications of RNA. In contrast to these RNA labeling methods, introducing an artificial, extra base pair (unnatural base pair), as a third base pair, to *in vitro* transcription systems allows the site-specific incorporation of an unnatural base linked with functional groups into desired positions of RNA during transcription mediated by the unnatural base pair. Thus, several unnatural base pairs that function in polymerase reactions have rapidly been developed for site-specific labeling of RNA molecules [7–18].

Here, for the site-specific incorporation of functional components into large RNA molecules, we report a fusion PCR and transcription system that employs two unnatural base pairs of 7-(2-thienyl)-imidazo [4,5-*b*]pyridine (**Ds**) and 2-nitro-4-propynylpyrrole (**Px**) [19, 20] and **Ds** and pyrrole-2-carbaldehyde (**Pa**) [17] (Figure 1). The **Ds**-**Px** pair exhibits high efficiency and selectivity in PCR amplification as a third base pair. Under optimized conditions, more than 97% of the **Ds**-**Px** pair survives in the $10^{28}$-fold amplified DNAs through exponential 100-cycle PCR (10 cycles repeated 10 times). In particular, a modified **Px** base, 4-(4,5-dihydroxypent-1-yn-1-yl)-2-nitropyrrole (Diol1-**Px**, Figure 1), has extremely high specificity as a pairing partner of **Ds**, and thus the misincorporation rates of Diol1-d**Px**TP and d**Ds**TP opposite the natural bases in templates during PCR amplification

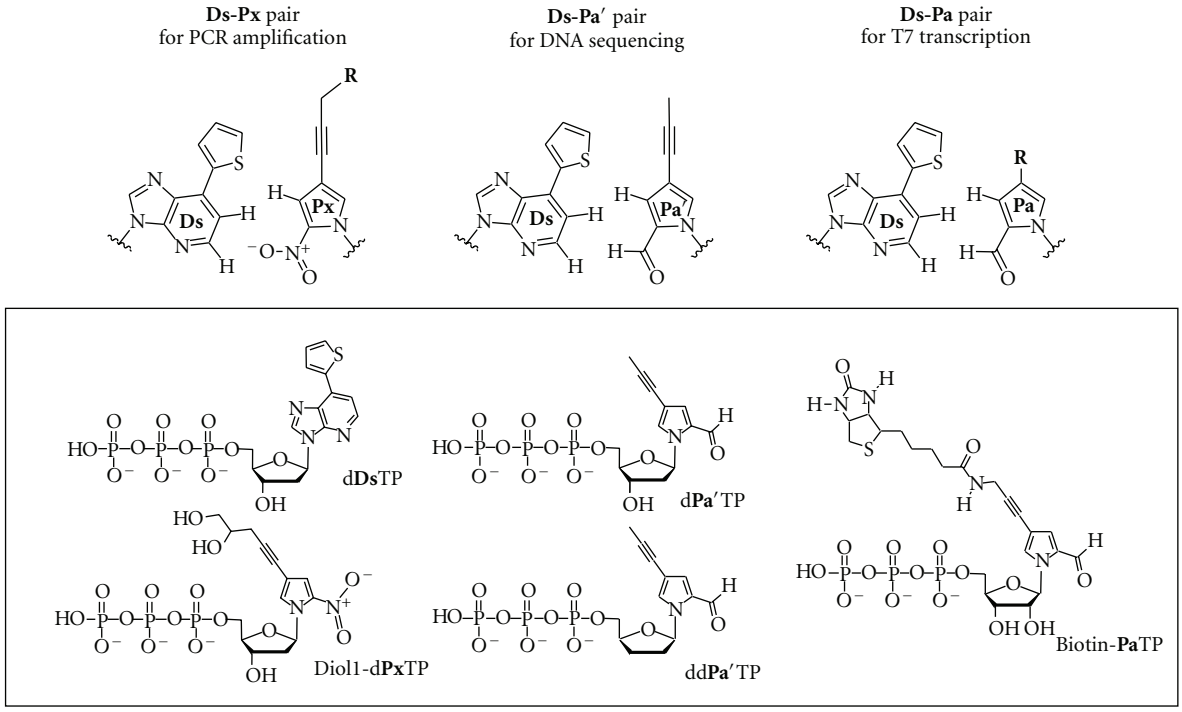



Figure 1: Chemical structures of the **Ds**-**Px**, **Ds**-**Pa**′, and **Ds**-**Pa** pairs. The **Ds**-**Px** pair is employed for PCR amplification to prepare the **Ds**-containing DNA templates. The **Ds**-**Pa**′ pair is employed for DNA sequencing, to determine the **Ds** position in the DNA templates. The **Ds**-**Pa** pair is employed for T7 transcription, to incorporate Biotin-**Pa** into RNA opposite **Ds** in the templates. **Px** and **Pa** bases modified with various functional groups (R) can also be used as substrates for PCR and transcription, respectively. The unnatural substrates used in this study are summarized in the box.

are as low as 0.005% per base pair per replication [20]. Here, we applied the **Ds**-**Px** pair to fusion PCR [21], using small DNA templates containing **Ds** with Diol1-d**Px**TP and d**Ds**TP, to prepare large **Ds**-containing DNA templates for a genetic-expansion transcription system, and generated 282-bp double-stranded DNA fragments containing the **Ds** and Diol1-**Px** pair. Using the **Ds**-containing DNA templates, we performed the site-specific incorporation of Biotin-**Pa**TP, as a functional unnatural base substrate, into 260-mer transcripts by T7 RNA polymerase. In transcription, the **Pa** base is superior to the **Px** base in terms of both the incorporation selectivity opposite **Ds** and the chemical stability of the nucleotide. However, in PCR amplification, the **Ds**-**Px** pair is more selective and efficient than the **Ds**-**Pa** pair, and thus to utilize both pairs' advantages, we developed a two-unnatural-base-pair system, the **Ds**-**Px** pair for fusion PCR, and the **Ds**-**Pa** pair for T7 transcription (Figure 2). By attaching functional groups of interest to the **Pa** base, this fusion PCR and transcription system could be a powerful tool for the site-specific labeling and functionalization of large RNA molecules.

## 2. Materials and Methods

*2.1. Materials.* Oligonucleotides containing **Ds** were synthesized with an Applied Biosystems 392 DNA synthesizer, using CE phosphoramidite reagents for the natural and **Ds** bases (Glen Research), and were purified by gel electrophoresis. Oligonucleotides comprising natural bases only were synthesized as described above or purchased from Invitrogen. AccuPrime *Pfx* DNA polymerase, SYBR Gold nucleic acid gel stain, and 10x PBS were purchased from Invitrogen. Streptavidin and silica-membrane columns for PCR product purification (Wizard SV Gel and PCR Clean-Up System) were purchased from Promega. T7 RNA polymerase was purchased from Takara. Toluidine Blue O was purchased from Chroma Gesellschaft Schmidt & Co. The RNA ladder marker (DynaMarker, RNA Low II) was purchased from BioDynamics Laboratory, Inc. The DNA ladder marker (2-Log DNA Ladder) was purchased from New England Biolabs. Unnatural nucleoside triphosphates (d**Ds**TP, Diol1-d**Px**TP, Biotin-**Pa**TP, dd**Pa**′TP, and d**Pa**′TP) were synthesized as described previously [17, 19, 20]. The BigDye Terminator v1.1 Cycle Sequencing Kit was purchased from Applied Biosystems. Centri-Sep spin columns were purchased from Princeton Separations. Natural NTP and dNTP Sets (100 mM solutions: ATP, CTP, GTP, and UTP, and dATP, dCTP, dGTP and dTTP, resp.) were purchased from GE Healthcare. Gel images were analyzed with a bioimaging analyzer, LAS4000 (Fuji Film). The plasmid DNA used for PCR amplification was provided by Dr. Tsutomu Kishi (RIKEN).

*2.2. Preparation of DNA Templates for T7 Transcription.* To prepare the 282-bp double-stranded DNA fragments containing the **Ds**-**Px** pair as templates for T7 transcription,

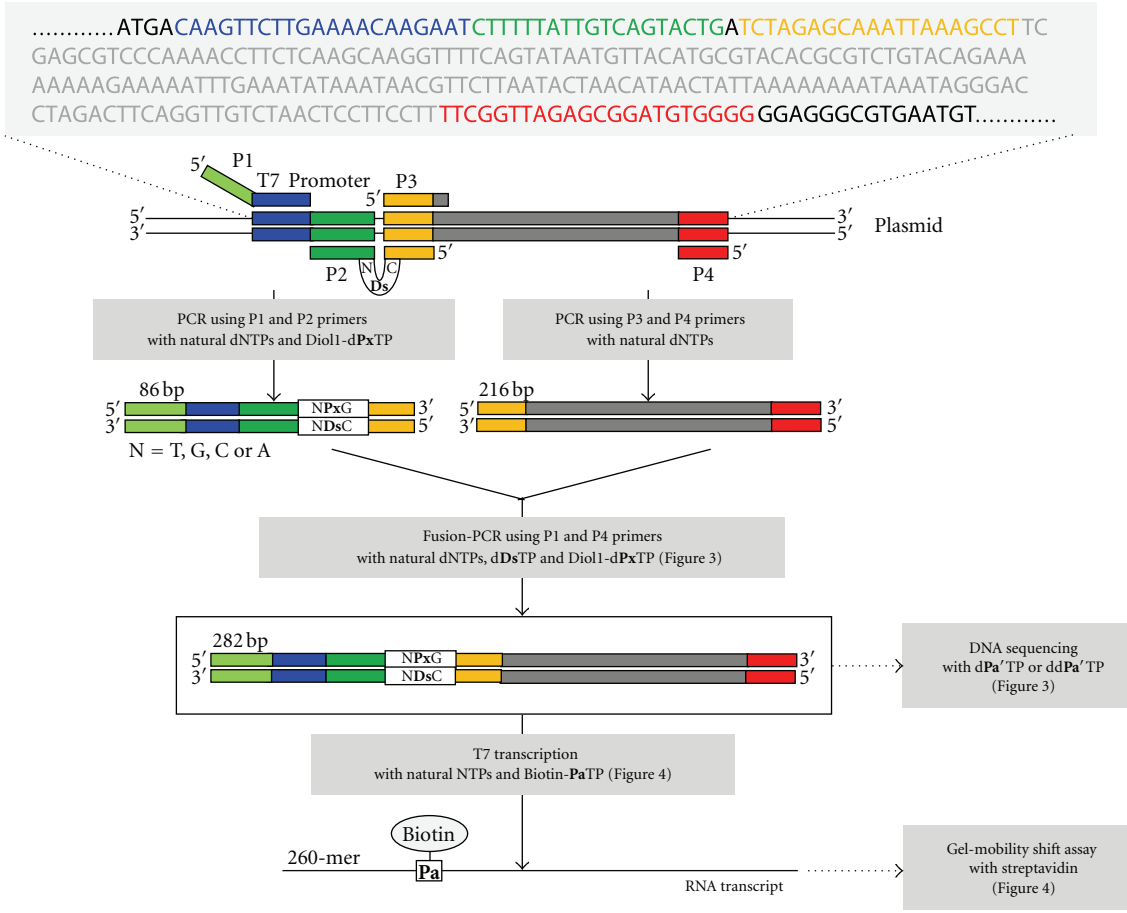



Figure 2: Experimental scheme of PCR amplification and transcription by a two-unnatural-base-pair system for site-specific biotin labeling of RNA molecules. DNA fragments containing the **Ds**-**Px** pair at the internal positions were prepared by a fusion PCR method. The original plasmid sequence, used as the PCR template, is shown at the top. The C**Ds**N sequence (N = T, G, A, or C) in the P2 primer is integrated into the 86-bp DNA fragments. The 282-bp products generated by fusion PCR were used as templates for T7 transcription with Biotin-**Pa**TP.

two steps of PCR using AccuPrime *Pfx* DNA polymerase were performed as follows. First, 86-bp DNA fragments containing the **Ds**-Diol1-**Px** pair were amplified through 25-cycle PCR (50 $\mu$L) from 1 ng/$\mu$L plasmid DNA template, by using 1 $\mu$M 5′-primer (P1: 5′-AAGCT*TAATACGACT CACTATAG*CAAGTTCTTGAAAACAAGAAT-3′, the T7 promoter region is italicized, and the region complementary to the plasmid is underlined) and 3′-primer (P2: 5′-AGGCTT TAATTTGCTCTAGAC**Ds**NCAGTACTGACAATAAAAAG-3′, N = T, G, A, and C; the region complementary to the plasmid is underlined) and 0.02 U/$\mu$L AccuPrime *Pfx* DNA polymerase, in the manufacturer's reaction buffer with 600 $\mu$M each natural dNTP, 50 $\mu$M Diol1-d**Px**TP, and 2 mM $MgSO_4$. The final concentrations of dNTPs and $MgSO_4$ were adjusted by adding 300 $\mu$M each natural dNTP, 50 $\mu$M Diol1-d**Px**TP, and 1 mM $MgSO_4$ to the 1xAccuPrime *Pfx* reaction mix, which originally contained 300 $\mu$M each natural dNTP and 1 mM $MgSO_4$ [20]. PCR conditions were 30 sec at 94°C, 30 sec at 55°C, and 2 min at 65°C per cycle, and the amplified PCR products were purified by gel electrophoresis. As control experiments, we also performed PCR amplification using the P1 and P2 primers without **Ds** (5′-AGGCTT TAATTTGCTCTAGACATCAGTACTGACAATAAAAAG-3′), in the presence and absence of Diol1-d**Px**TP. A 216-bp double-stranded DNA fragment was amplified by 25-cycle PCR using 1 $\mu$M P3 primer (5′-TCTAGAGCAAAT TAAAGCCTTCG-3′, the sequence complementary to the P2 primer is underlined) and P4 primer (5′-CCCCACATCCGC TCTAACCGAA-3′) from 1 ng/$\mu$L plasmid DNA template in the reaction buffer, containing 600 $\mu$M each natural dNTP and 2 mM $MgSO_4$, followed by gel purification. Next, we performed fusion PCR (100 $\mu$L) with AccuPrime *Pfx* DNA polymerase, using 1 $\mu$M each of the P1 and P4 primers, and the purified 86-bp (25 nM) and 216-bp (1 nM) DNA fragments, in the reaction buffer with 600 $\mu$M each natural dNTP and 2 mM $MgSO_4$, in the presence or absence of 50 $\mu$M each of d**Ds**TP and Diol1-d**Px**TP. PCR conditions were 30 sec at 94°C and 3 min at 65°C per cycle. After 25-cycle PCR amplification, the products were purified by passage through silica-membrane columns, according to the manufacturer's instructions. The concentrations of the purified products were determined from their UV

absorbance. The purified products were used for DNA sequencing and transcription experiments.

*2.3. DNA Sequencing.* The cycle sequencing reaction (10 $\mu$L) was performed with the Cycle Sequencing mix (4 $\mu$L) from the BigDye Terminator v1.1. Cycle Sequencing Kit, 1 $\mu$L of 2 $\mu$M Sequencing primer (5′-TGACAAGTTCTT GAAAACAAGAAT-3′), 2 $\mu$L of 250 $\mu$M d**Pa**′TP or dd**Pa**′TP, and 3 $\mu$L of 50 nM DNA fragments [20]. The sequencing cycle parameters were 25 cycles of 10 sec at 96°C, 5 sec at 50°C, and 4 min at 60°C. The residual dye terminators were removed with Centri-Sep columns, and the solutions were dried with a centrifugal vacuum evaporator. The residues were resuspended in 3 $\mu$L of a formamide/BlueDextran/EDTA loading buffer and analyzed with an ABI 377 DNA sequencer, using a 6% polyacrylamide-6 M urea gel. The sequence peak patterns were analyzed with the Applied Biosystems PRISM sequencing analysis software, v3.2.

*2.4. Transcription.* Transcription (20 $\mu$L) was performed in a buffer containing 40 mM Tris-HCl (pH 8.0), 24 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, and 0.01% Triton X-100, in the presence of 2 mM each NTP, 0 or 2 mM Biotin-**Pa**TP, 100 nM 282-bp DNA template, and 2.5 U/$\mu$L T7 RNA polymerase. After an incubation at 37°C for 3 h, the reaction was quenched by adding an equivalent volume of a denaturing solution, containing 10 M urea in 1xTBE. The reaction mixtures were heated at 75°C for 3 min, and 7.5 $\mu$L aliquots were fractionated on a 5% denaturing polyacrylamide-7 M urea gel. After electrophoresis, the transcribed products on the gel were stained with toluidine blue and detected by a LAS 4000 imager. For gel-mobility shift assays, the full-length products were purified on a 5% denaturing polyacrylamide-7 M urea gel.

*2.5. Gel-Mobility Shift Assay.* We detected the biotinylated RNA transcripts (260-mer) by gel-mobility shift assays, using streptavidin. We incubated the mixture (10 $\mu$L) of 0.5 pmol transcripts and excess amounts of streptavidin (800 ng) for 1 h at 25°C, in 1x PBS containing 5% glycerol. The biotinylated RNA-streptavidin complexes were separated from the free RNAs on a non-denaturing 8% polyacrylamide gel, and the RNAs on the gel were stained with SYBR Gold and detected by an LAS 4000 imager.

## 3. Results and Discussion

*3.1. Fusion-PCR Mediated by the Ds-Px Pair for Preparing Long DNA Templates Containing Ds at Desired Positions.* To examine fusion PCR [21] using the **Ds**-**Px** pair, we prepared the 282-bp double-stranded DNA (dsDNA) templates containing the **Ds** and Diol1-**Px** pair at internal positions by two PCR steps, using four PCR primers (P1 to P4) and a plasmid DNA, as shown in Figure 2. The P2 primer contains **Ds** and the P3 region. In the first PCR, we used the plasmid as an initial PCR template to introduce the **Ds**-**Px** pair, and prepared two dsDNA fragments (86- and 216-bp) by using AccuPrime *Pfx* DNA polymerase [20]. The 86-bp DNA fragment was amplified by the P1 and P2 primers and corresponds to the 5′ region of the 282-bp template, which contains the T7 promoter sequence in its 5′ region and the **Ds**-**Px** pair in its 3′ region. The 216-bp DNA fragment was amplified by the P3 and P4 primers and corresponds to the 3′ region of the 282-bp template, which comprises natural bases only. Since the P2 and P3 primers share a 20-mer complementary sequence, the 282-bp templates can be prepared by the second PCR step (fusion PCR) using the 86-bp and 216-bp DNA fragments and the P1 and P4 primers.

For preparing the **Ds**-containing 86-bp DNA fragments, we performed 25 cycles of PCR using each of the four P2 primers encoding a 5′-C**Ds**N-3′ sequence (N = T, G, C, or A), in the presence of Diol1-d**Px**TP, as well as the natural dNTPs as substrates and the P1 primer. Each amplified PCR product was purified by gel electrophoresis, to completely remove the original plasmid template. As control experiments, we also performed PCR amplification using a non-**Ds**-containing P2 primer with a 5′-CAT-3′ sequence, in the presence and absence of the unnatural base substrate, Diol1-d**Px**TP. The end-point analysis of the PCR products on the agarose gel revealed no differences among their PCR amplification efficiencies (data not shown). Thus, under the conditions employed here by using AccuPrime *Pfx* DNA polymerase, the nature of the natural base nucleotide at the 5′-neighboring position of **Ds** does not affect the amplification efficiency.

To prepare the full-length 282-bp templates by fusion PCR, we performed 25-cycle PCR using the P1 and P4 primers and the gel-purified 86-bp and 216-bp DNA fragments as templates, in the presence of Diol1-d**Px**TP and d**Ds**TP together with the natural dNTPs as substrates. Figure 3 shows the agarose gel of the PCR-amplified products. The fusion PCR involving the **Ds**-**Px** pair with all four sequence contexts generated the full-length products with high efficiency, similar to that of the control experiment using the DNA fragments with only the natural base sequence context.

The site-specific incorporation of the **Ds**-**Px** pair into each of the 282-bp DNA fragments was confirmed by dideoxy dye terminator sequencing of the **Ds**-containing DNA strands in the presence of the dideoxyribonucleoside or deoxyribonucleoside triphosphate of 4-propynylpyrrole-2-carbaldehyde (d**Pa**′TP or dd**Pa**′TP, Figure 1), another pairing partner of **Ds** (Figure 3), according to our previously reported method [19, 20]. In the presence of dd**Pa**′TP, the sequencing reaction terminated at the unnatural base position because of the incorporation of dd**Pa**′TP opposite **Ds** in the DNA templates, and thus the following base peaks disappeared. In the sequencing in the presence of d**Pa**′TP, the d**Pa**′TP was incorporated opposite **Ds** in the DNA templates, and the following base peaks appeared, but there is no peak at the unnatural base position because no dye terminator corresponding to the unnatural base was present in the sequencing reaction. By comparing the sequence patterns of the DNA fragment containing 5′-A**Px**G-3′ with that containing 5′-ATG-3′, the differences are clearly observed. These results indicated that the **Ds**-**Px** pair functions well in fusion PCR using AccuPrime *Pfx* DNA polymerase, and from the sequencing analysis, the retention rate of the unnatural base pair in the amplified DNA fragments was more than 97%.

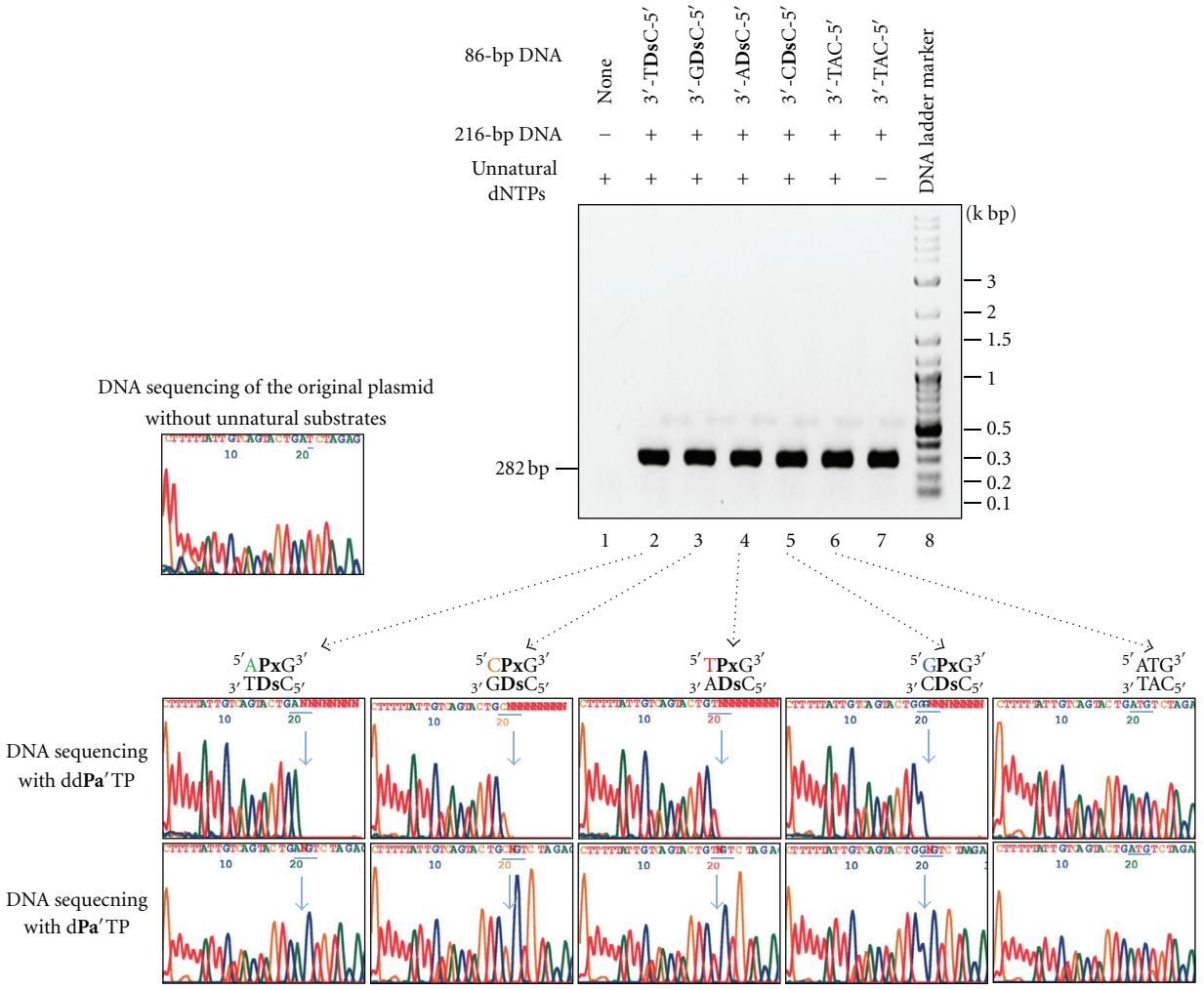



Figure 3: PCR amplification and DNA sequencing of 282-bp DNA templates containing **Ds** at specific positions. After 25-cycle fusion PCR using the P1 and P4 primers and the 86-bp and 216-bp DNA fragments as templates in the presence or absence of d**Ds**TP and Diol1-d**Px**TP, the amplified PCR products were analyzed on a 1% agarose gel. Using the purified PCR products, the sequencing reactions were performed in the presence of dd**Pa**′TP (upper sequencing panels) or d**Pa**′TP (lower sequencing panels). The light blue arrow indicates the unnatural base position. The light blue bar indicates the inserted sequence.

*3.2. T7 Transcription, Mediated by the Ds-Pa Pair for the Site-Specific Biotinylation of Large RNA Molecules Using a Biotin-Pa Substrate.* Next, we examined the transcription by T7 RNA polymerase, using the 282-bp DNA templates containing **Ds** and Biotin-**Pa**TP as a functional component. Transcription was performed in the presence of 2 mM Biotin-**Pa**TP, as well as 2 mM natural NTPs, at 37°C for 3 h, and the full-length transcripts (260-mer) were analyzed by denaturing polyacrylamide gel electrophoresis (Figure 4). As control experiments, the non-**Ds**-containing DNA templates with the 5′-CAT-3′ sequence, which were PCR amplified with or without the unnatural base substrates, were transcribed with or without Biotin-**Pa**TP. All of the transcription reactions produced similar yields of the 260-mer products.

To characterize the selectivity of the Biotin-**Pa** incorporation into RNA fragments by transcription, we performed gel-mobility shift assays of the biotinylated transcripts in the presence of streptavidin (Figure 4). As for the incorporation site, we already confirmed the site specificity of the Biotin-**Pa** incorporation opposite **Ds** in templates with various sequence contexts [17, 18]. From the gel shift assays, we estimated the biotinylation yields of each transcript (92% of the transcript was biotinylated for the 3′-C**Ds**C-5′ template sequence, 84% for 3′-T**Ds**C-5′, 75% for 3′-G**Ds**C-5′, and 72% for 3′-A**Ds**C-5′). Thus, in contrast to fusion PCR involving the **Ds**-**Px** pair, the biotinylation selectivity depended to some extent on the neighboring bases of **Ds**, and the order of the effective sequence contexts of the 282-bp DNA templates was the same as that of the previously determined effective sequence context of the 35-mer synthetic DNA template for a 17-mer RNA transcript with **Pa** [18]. The biotinylation yields resulted from the sum of the 25-cycle PCR and T7 transcription selectivities. However, the sequencing analysis (Figure 3) indicated that the retention of the unnatural-base-pair in the amplified DNA was more than 97% after 25-cycle PCR, and thus the biotinylation yields

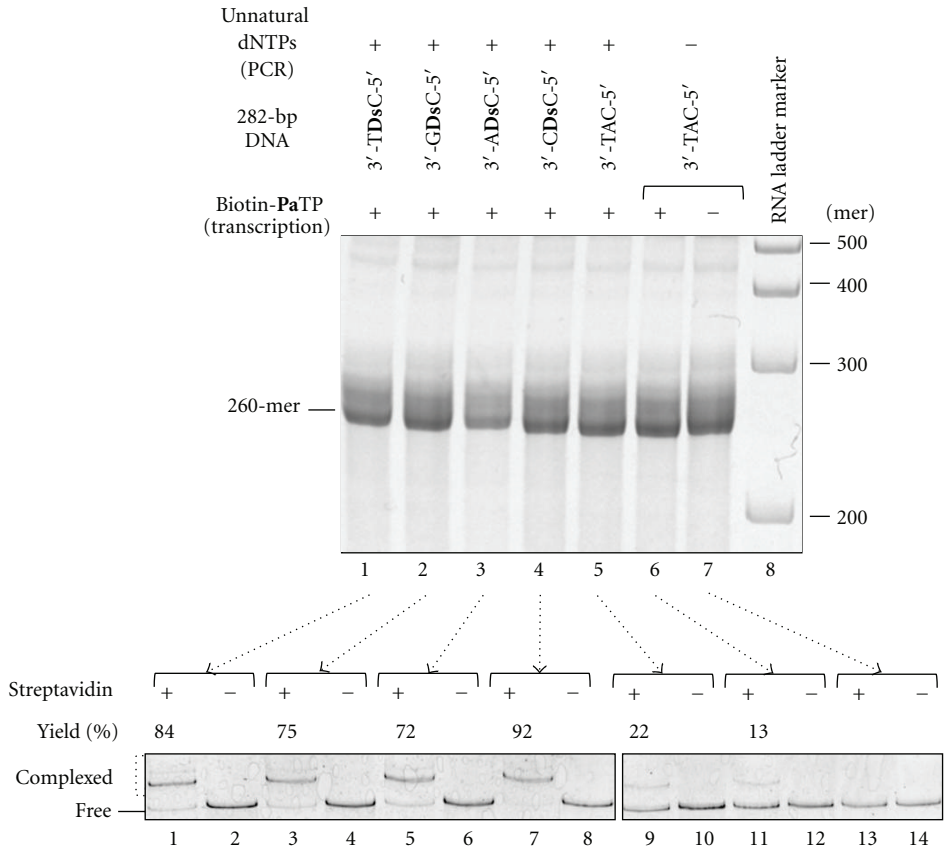



Figure 4: T7 transcription for site-specific biotin labeling of 260-mer transcripts. Transcripts from each 282-bp DNA template containing 3′-T**Ds**C-5′, 3′-G**Ds**C-5′, 3′-A**Ds**C-5′, or 3′-C**Ds**C-5′, in the presence of the natural NTPs (2 mM) and Biotin-**Pa**TP (2 mM), were analyzed on a 5% denaturing polyacrylamide gel (upper gel image). After the purification of each 260-mer transcript, gel-mobility shift assays were performed using streptavidin. Biotinylated RNA-streptavidin complexes were separated from free RNAs on an 8% nondenaturing polyacrylamide gel, and the amounts of the complexes (yields) were determined from the band intensities. The experiments were independently repeated twice, and representative data are shown. The gel-mobility shift values were averaged from two data sets.

actually depended on the transcription selectivities of the **Ds**-**Pa** pair when using each DNA template.

We also assessed the misincorporation of Biotin-**Pa**TP opposite the natural bases in the two-unnatural base pair system. The misincorporation rates of Biotin-**Pa**TP into the transcript using the control 3′-TAC-5′ templates, which were prepared by fusion PCR in the presence and absence of d**Ds**TP and Diol1-d**Px**TP, were 22% and 13%, respectively. Thus, the misincorporations correspond to only 0.096% (for 22% total) and 0.054% (for 13% total) per position in the 260-mer transcript, as calculated by the following equation: $y = (1 - x/100)^{260}$, where $x$% is the Biotin-**Pa** misincorporation rate opposite the natural bases per position and $y$ is the ratio of nonbiotinylated transcript. These misincorporation rates were lower than those of the noncognate natural base pairings, as we previously reported that the misincorporation of biotin-linked UTP opposite G, C, and T in the T7 transcription of a non-A-containing template was around 0.16% per base [17]. In addition, the value of 0.054% is quite consistent with the value (0.06%) that we previously determined by T7 transcription of 152-mer transcripts in the presence of 2 mM Biotin-**Pa**TP and 2 mM natural NTPs [17]. This misincorporation rate also corresponds to the sum of the 25-cycle PCR and T7 transcription selectivities. Thus, the difference between 0.096% and 0.054% was caused by the **Ds** misincorporation into the template strand during 25-cycle PCR amplification with d**Ds**TP and Diol1-d**Px**TP. In the 25-cycle fusion PCR, the PCR primers were mostly consumed, and the theoretical amplification cycle was estimated as ten ($2^{10} \approx 1000$ = [primer]/[216-bp DNA fragments]). Therefore, the **Ds** misincorporation opposite a natural base per replication was calculated as approximately 0.004%($\approx$(0.096% − 0.054%)/10). Consequently, the **Ds** misincorporation into the nontemplate strand would occur, and thus the misincorporation rates of Diol1-d**Px**TP and d**Ds**TP opposite the natural bases in templates during PCR amplification would be 0.008%, which is as high as that of the noncognate natural base pairings.

## 4. Conclusion

In this study, we demonstrated the two-unnatural-base-pair system for the site-specific labeling of large RNA molecules by fusion PCR and T7 transcription. In fusion PCR, more

than 97% of the **Ds**-**Px** pair was retained in the amplified 282-bp DNA fragments, and the misincorporation rate of the unnatural bases opposite the natural bases was 0.008% per base per replication. We employed the natural base substrates at relatively high concentrations (600 $\mu$M each), as compared to the unnatural base substrates (d**Ds**TP and Diol1-d**Px**TP, 50 $\mu$M each) in the fusion PCR, to reduce the misincorporation rate. By using this method, the biotin-linked unnatural base was site-specifically incorporated at predetermined positions of RNA transcripts with selectivities ranging from 72% to 92%, depending on the sequence contexts around the unnatural base. In addition, the misincorporation rate of the biotin-unnatural base opposite the natural bases was around 0.05% per natural base. These selectivity and misincorporation rates were obtained using 2 mM unnatural and natural base substrates in T7 transcription. The incorporation selectivity (90−96%) and the misincorporation rate of Biotin-**Pa** depend on the ratio of Biotin-**Pa**TP to the natural base NTPs in transcription [17]. Thus, the incorporation efficiency and misincorporation rates can be adjusted by changing the concentration ratios between the unnatural and natural base substrates. To reduce the misincorporation of the unnatural base substrate opposite the natural bases, with the sacrifice of the incorporation efficiency of the unnatural base substrate at the desired positions, transcription should be performed with 1 mM unnatural base substrate and 2 mM natural base substrates [12, 17]. Since several functional groups can be attached to the **Pa** base, this method could be applied to a wide range of site-specific labeling and functionalization of large RNA molecules.

## Acknowledgments

The authors thank Tsutomu Kishi for preparing the plasmid and helpful discussions and Akira Sato for preparing the unnatural substrates. This work was supported by the Targeted Proteins Research Program and the RIKEN Structural Genomics/Proteomics Initiative, the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan, and by Grants-in-Aid for Scientific Research (KAKENHI 19201046 to I. H., 20710176 to M. K.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

[1] P. Z. Qin and A. M. Pyle, "Site-specific labeling of RNA with fluorophores and other structural probes," *Methods*, vol. 18, no. 1, pp. 60–70, 1999.

[2] D. Chatterji and V. Gopal, "Fluorescence spectroscopy analysis of active and regulatory sites of RNA polymerase," *Methods in Enzymology*, vol. 274, pp. 456–478, 1996.

[3] F. Huang, G. Wang, T. Coleman, and N. Li, "Synthesis of adenosine derivatives as transcription initiators and preparation of 5′ fluorescein- and biotin-labeled RNA through one-step in vitro transcription," *RNA*, vol. 9, no. 12, pp. 1562–1570, 2003.

[4] A. Draganescu, S. C. Hodawadekar, K. R. Gee, and C. Brenner, "Fhit-nucleotide specificity probed with novel fluorescent and fluorogenic substrates," *Journal of Biological Chemistry*, vol. 275, no. 7, pp. 4555–4560, 2000.

[5] S. Seetharaman, M. Zivarts, N. Sudarsan, and R. R. Breaker, "Immobilized RNA switches for the analysis of complex chemical and biological mixtures," *Nature Biotechnology*, vol. 19, no. 4, pp. 336–341, 2001.

[6] L. Zhang, L. Sun, Z. Cui, R. L. Gottlieb, and B. Zhang, "5′-sulfhydryl-modified RNA: initiator synthesis, in vitro transcription, and enzymatic incorporation," *Bioconjugate Chemistry*, vol. 12, no. 6, pp. 939–948, 2001.

[7] Y. Tor and P. B. Dervan, "Site-specific enzymatic incorporation of an unnatural base, N6-(6-ammohexyl)isoguanosine, into RNA," *Journal of the American Chemical Society*, vol. 115, no. 11, pp. 4461–4467, 1993.

[8] I. Hirao, T. Ohtsuki, T. Fujiwara et al., "An unnatural base pair for incorporating amino acid analogs into proteins," *Nature Biotechnology*, vol. 20, no. 2, pp. 177–182, 2002.

[9] M. Kimoto, M. Endo, T. Mitsui, T. Okuni, I. Hirao, and S. Yokoyama, "Site-specific incorporation of a photo-crosslinking component into RNA by T7 transcription mediated by unnatural base pairs," *Chemistry and Biology*, vol. 11, no. 1, pp. 47–55, 2004.

[10] R. Kawai, M. Kimoto, S. Ikeda et al., "Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs," *Journal of the American Chemical Society*, vol. 127, no. 49, pp. 17286–17295, 2005.

[11] K. Moriyama, M. Kimoto, T. Mitsui, S. Yokoyama, and I. Hirao, "Site-specific biotinylation of RNA molecules by transcription using unnatural base pairs," *Nucleic Acids Research*, vol. 33, no. 15, article e129, 2005.

[12] M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama, and I. Hirao, "Fluorescent probing for RNA molecules by an unnatural base-pair system," *Nucleic Acids Research*, vol. 35, no. 16, pp. 5360–5369, 2007.

[13] Y. Hikida, M. Kimoto, S. Yokoyama, and I. Hirao, "Site-specific fluorescent probing of RNA molecules by unnatural base-pair transcription for local structural conformation analysis," *Nature Protocols*, vol. 5, no. 7, pp. 1312–1323, 2010.

[14] D. A. Malyshev, J. S. Young, P. Ordoukhanian, and F. E. Romesberg, "PCR with an expanded genetic alphabet," *Journal of the American Chemical Society*, vol. 131, no. 41, pp. 14620–14621, 2009.

[15] Y. J. Seo, S. Matsuda, and F. E. Romesberg, "Transcription of an expanded genetic alphabet," *Journal of the American Chemical Society*, vol. 131, no. 14, pp. 5046–5047, 2009.

[16] Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian, and F. E. Romesberg, "Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs," *Journal of the American Chemical Society*, vol. 133, no. 49, pp. 19878–19888, 2011.

[17] I. Hirao, M. Kimoto, T. Mitsui et al., "An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA," *Nature Methods*, vol. 3, no. 9, pp. 729–735, 2006.

[18] M. Morohashi, M. Kimoto, A. Sato, R. Kawai, and I. Hirao, "Site-specific incorporation of functional components into RNA by an unnatural base pair transcription system," *Molecules*, vol. 17, no. 3, pp. 2855–2876, 2012.

[19] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, and I. Hirao, "An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules," *Nucleic Acids Research*, vol. 37, no. 2, article e14, 2009.

[20] R. Yamashige, M. Kimoto, Y. Takezawa et al., "Highly specific unnatural base pair systems as a third base pair for PCR amplification," *Nucleic Acids Research*, vol. 40, no. 6, pp. 2793–2806, 2012.

[21] E. Burke and S. Barik, "Megaprimer PCR: application in mutagenesis and gene fusion," *Methods in Molecular Biology*, vol. 226, pp. 525–532, 2003.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 530963, 17 pages
doi:10.1155/2012/530963

*Review Article*

# Synthetic Nucleotides as Probes of DNA Polymerase Specificity

**Jason M. Walsh[1] and Penny J. Beuning[1,2]**

[1] *Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, 102 Hurtig Hall, Boston, MA 02115, USA*
[2] *Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, MA 02115, USA*

Correspondence should be addressed to Penny J. Beuning, beuning@neu.edu

Received 23 January 2012; Accepted 21 March 2012

Academic Editor: Yoshio Saito



The genetic code is continuously expanding with new nucleobases designed to suit specific research needs. These synthetic nucleotides are used to study DNA polymerase dynamics and specificity and may even inhibit DNA polymerase activity. The availability of an increasing chemical diversity of nucleotides allows questions of utilization by different DNA polymerases to be addressed. Much of the work in this area deals with the A family DNA polymerases, for example, *Escherichia coli* DNA polymerase I, which are DNA polymerases involved in replication and whose fidelity is relatively high, but more recent work includes other families of polymerases, including the Y family, whose members are known to be error prone. This paper focuses on the ability of DNA polymerases to utilize nonnatural nucleotides in DNA templates or as the incoming nucleoside triphosphates. Beyond the utility of nonnatural nucleotides as probes of DNA polymerase specificity, such entities can also provide insight into the functions of DNA polymerases when encountering DNA that is damaged by natural agents. Thus, synthetic nucleotides provide insight into how polymerases deal with nonnatural nucleotides as well as into the mutagenic potential of nonnatural nucleotides.

## 1. Introduction

Since the structure of DNA was determined [1, 2], biochemists have sought more detailed ways to study DNA and the proteins that interact with it [3, 4]. Solid phase nucleic acid synthesis of DNA molecules facilitates the site-specific incorporation of a wide range of chemically modified bases and sugar-phosphate backbones, allowing the roles of specific atoms in DNA function and recognition to be probed. Synthetic nonnatural nucleobases are useful for a variety of studies of DNA polymerase function, such as studies of DNA polymerase specificity, mutagenesis, and dynamics, as well as fluorescence resonance energy transfer (FRET) analysis of DNA polymerase interactions with DNA. The study of mutagenesis facilitated by DNA polymerases has attracted increasing interest because replication defects can lead to certain human diseases like the cancer-prone syndrome xeroderma pigmentosum variant (XPV) [5, 6] and other diseases [7–9], as well as potentially contribute to antibiotic resistance [10, 11]. Moreover, specialized damage-bypass DNA polymerases are implicated in conferring cellular tolerance to cancer chemotherapy agents that act via DNA damage, thereby decreasing their effectiveness [12–16]. This paper will focus on the ability of DNA polymerases to recognize and accept nonnatural bases either on the template strand or as the incoming triphosphate nucleotide. Much of the work discussed here will deal with A family polymerases (e.g., Klenow fragment (KF) of pol I and Taq DNA polymerase), but more recent work with Y family polymerases [17] and their ability to utilize certain nucleotide analogs will also be discussed.

DNA polymerases generally adopt a right-hand fold, in which the thumb and fingers bind DNA and nucleotide (Figure 1) [18, 19]. DNA polymerases add nucleotides to the growing DNA strand via nucleophilic attack of the free 3′ hydroxyl group of the DNA primer on the alpha phosphate of the incoming deoxynucleotide with release of pyrophosphate. DNA polymerase active site residues, which are usually glutamate or aspartate and are located in the palm domain, coordinate divalent magnesium ions that serve to activate the 3′-OH nucleophile (Figure 1) [20–25]. The catalytic cycle is generally accompanied by conformational changes in the

fingers domain. In replicating DNA, DNA polymerases have to be able to form all four base pair combinations specifically and efficiently in order to maintain the integrity of the genome (Figure 2); however, when replication errors occur, the mismatched bases can be removed by the exonucleolytic proofreading function of DNA polymerases [26]. Replicative DNA polymerases possess an exonuclease domain that may be part of the same or a separate polypeptide that utilizes a metal-dependent mechanism to excise mismatched bases [26, 27]. The proofreading process involves translocation of the primer terminus from the polymerase active site to the exonuclease active site; after the phosphodiester bond is hydrolyzed to remove the mismatched base, the primer strand reanneals to the template so that polymerization can continue [27, 28]. Replication errors that escape proofreading can be repaired by the mismatch repair system [29].

Based on sequence conservation, DNA polymerases are divided into A, B, C, D, X, and Y families. The A and B family DNA polymerases can be involved in replication or repair, whereas members of the C family are involved in DNA replication [33]. X family DNA polymerases are involved in repair, and Y family DNA polymerases are specialized for copying damaged DNA [33] in a process known as translesion synthesis (TLS). In general, replicative DNA polymerases cannot copy damaged DNA; rather, a specialized TLS polymerase must be recruited to extend primers a sufficient distance past distortions in DNA templates to allow replicative DNA polymerases to recover synthesis [34–37]. DNA replication past damage or unusual DNA structures requires the ability to both insert a nucleotide opposite a modification in the template as well as to extend the newly generated primer beyond that position. Some polymerases may be able to insert a nucleotide opposite nonstandard bases but be unable to extend the resulting primer terminus, as discussed below.

The four canonical bases vary in their chemical and geometric properties, but the C1′–C1′ distance of the standard Watson-Crick base pairs and the backbone C–O–P–O–C bonds remain constant regardless of the particular base pair [38]. Expansion of the nucleobase alphabet must take some of these structural considerations into account; usually nonnatural bases need to have similar geometries as the natural bases, usually but not necessarily retain some level of hydrogen bonding capabilities, and usually have $\pi$ electron systems in order to retain the stability provided by base stacking. Hydrophobicity and base stacking interactions are also important for DNA structure [38].

## 2. Abasic Sites and Small Molecule Substitutions

Stable, synthetic abasic sites were first introduced into DNA in 1987 [39]. As it is estimated that 10,000 abasic sites form in each human cell per day [29], it was important to develop a stable, synthetic abasic site in order to facilitate the study of DNA polymerase interactions. Furthermore, it is informative to determine the activity of DNA polymerases in the absence of an instructional base. Takeshita et al. introduced 3-hydroxy-2-hydroxymethyl-tetrahydrofuran into DNA, which is a model for the predominant cyclic version of 2′-deoxyribose. Therefore, this analog serves as the sugar lacking the base, or an AP (apurinic/apyrimidinic) site [39]. It was shown that KF of *Escherichia coli* pol I, as well as calf-thymus DNA polymerase $\alpha$, add dATP opposite synthetic abasic sites most frequently [39], leading to the proposal that DNA polymerases generally follow the "A-rule," inserting A in the absence of specific coding information [40]. The cocrystal structure of KlenTaq DNA polymerase with the furan synthetic abasic site in the templating position suggests a mechanism for this, as a protein Tyr side chain fills the space left vacant by the missing base and acts as a pyrimidine base mimic [41]. Some C family DNA polymerases that are error prone and/or involved in mutagenesis can also bypass synthetic abasic sites, in the case of *Streptococcus pyogenes* by incorporating dA, dG, or to a lesser extent dC, and in the case of *Bacillus subtilis* by weakly incorporating dG and generating one-nucleotide deletions via a misalignment mechanism [42–44]. *Saccharomyces cerevisiae* B-family member DNA pol zeta only weakly bypasses abasic sites [45]. Strikingly, African Swine Fever Virus (ASFV) DNA pol X is a highly error-prone DNA polymerase but is unable to copy DNA containing an abasic site [46]. Pol X cannot insert a nucleotide opposite an abasic site, nor can it extend a primer terminus containing an abasic site [46].

Because Y family DNA polymerases are known to copy noncanonical DNA structures, their proficiency at copying synthetic abasic sites has been examined in some detail. The model Y family DNA polymerase *Sulfolobus solfataricus* Dpo4 copies synthetic abasic sites mainly by incorporating dA but also by generating small deletions [47]. *E. coli* DinB (DNA pol IV) efficiently copies DNA containing a synthetic abasic site, but primarily by generating (−2) deletions [48]. *E. coli* DNA pol V (UmuD′$_2$C) bypasses synthetic abasic sites by inserting primarily dA (~70%) or dG (~30%) opposite the modification [49, 50]. Even though both *E. coli* Y family DNA polymerases can copy DNA containing abasic sites, Pol V is used to bypass abasic sites *in vivo*, probably because base substitutions are generally less harmful than frameshift mutations [48]. Human DNA pol iota, which, like ASFV pol X, is highly inaccurate when replicating undamaged DNA [51], can efficiently incorporate dG opposite an abasic site but is unable to extend from primer termini containing abasic sites [52]. Human DNA pol eta copies abasic sites by incorporating predominantly dA but also dG [53, 54], whereas human pol kappa incorporates predominantly dA but also generates one nucleotide deletions [55, 56]. Y family member Rev1 from yeast incorporates dC opposite abasic sites, which have been suggested to be the cognate lesion of Rev1 [57–59]. Yeast pol alpha, replicative DNA polymerase pol epsilon, and Y family pol eta are all capable of bypassing abasic sites, whereas replicative DNA polymerase pol delta is less efficient [60, 61]. Intriguingly, yeast pol eta and KF add a pyrene nucleotide opposite the template abasic site more efficiently than adding A, likely in part because pyrene is approximately the same size as a base pair and can engage in base stacking interactions [62, 63]. Bacteriophage T4 DNA polymerase incorporates nucleotide triphosphate versions

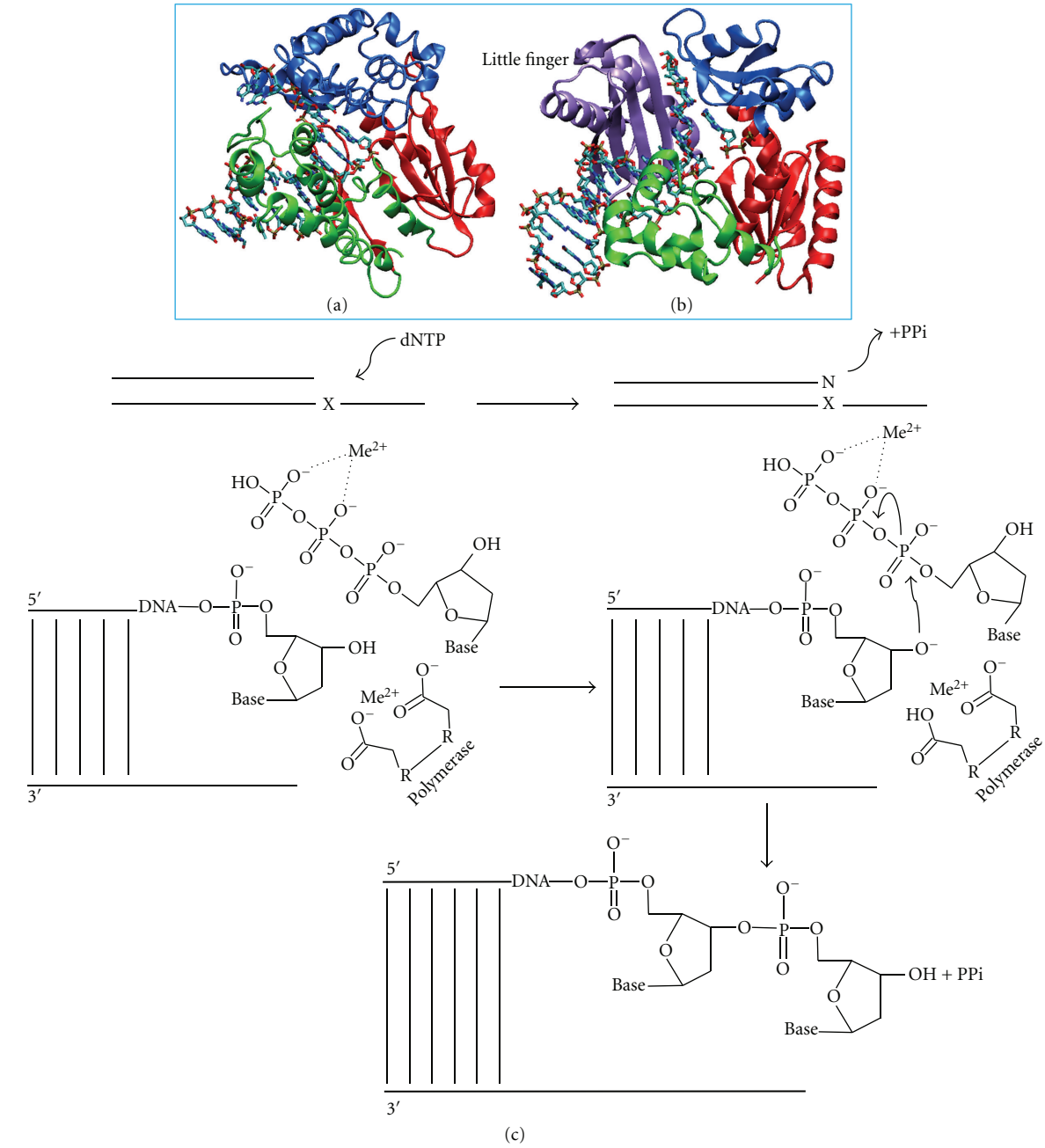



Figure 1: Comparison of the overall folds of (a) a replicative DNA polymerase, *Bacillus stearothermophilus* DNA pol I [30], and (b) a Y family DNA polymerase, *Sulfolobus solfataricus* Dpo4 [31]. The respective thumb domains are shown in green, palm domains are in red, and fingers domains are in blue. The little finger domain unique to the Y family DNA polymerases is in purple [31]. The "vestigial" exonuclease domain of *Bs* pol I has been omitted for clarity [32]. (c) Polymerase catalyzed DNA replication (phosphoryl transfer) reaction. Polymerization of DNA occurs at a free 3′ hydroxyl group of the deoxyribose. Polymerases use a divalent magnesium ion ($Me^{2+}$) to coordinate the negative charges of both the phosphate groups and the aspartic acid or glutamic acid in the active site of the polymerase [26].

of 5-nitroindolyl, 5-cyclohexyl-indole, and 5-cyclohexenyl-indole opposite abasic sites more efficiently than it incorporates dAMP [64, 65]. Due to the complicated responses of even the relatively forgiving Y family DNA polymerases to the synthetic model abasic site, it has been demonstrated that multiple DNA polymerases may be used to bypass DNA damage efficiently while minimizing mutations [66, 67].

Y family DNA polymerases are able to copy DNA containing noncanonical structures ranging from abasic sites to bulky DNA adducts [68–73]. Therefore, it was of interest to determine the minimal features of DNA required for

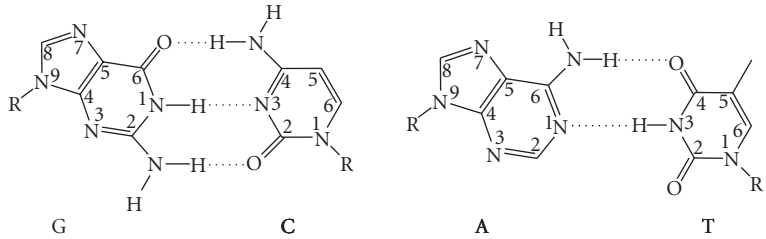



Figure 2: The canonical Watson-Crick base pairs. Standard numbering is indicated. Unless otherwise noted, R indicates the position of the deoxyribose in all figures.

replication. Short (three or 12) chains of methylene ($CH_2$) residues in the middle of canonical DNA templates were used to probe tolerance for minimal DNA backbones. *E. coli* pols I, II, and III were unable to replicate either DNA structure. On the other hand, both pols IV and V could replicate the three- or 12-methylene linker-containing DNA *in vitro*, although, in an analogous situation to abasic sites, only pol V is observed to replicate these unusual structures *in vivo* [74]. Human DNA polymerases showed more subtle differences, in that pols eta, kappa, and iota could replicate a three-methylene linker by inserting nucleotides opposite the noninstructional segment, but only pols eta and kappa could fully bypass the modified gap [75]. Pols eta and iota could insert nucleotides opposite the 12-methylene linker, whereas pol kappa had little to no activity, and none of these three polymerases could completely bypass the 12-methylene linker [75]. Clearly, at least some Y family DNA polymerases are capable of replicating non-DNA segments.

In order to probe the size tolerance for bases in the active site, a series of dG analogs with increasingly large substituents at the $N^2$ position in the minor groove were constructed and used as the template base with a range of DNA polymerases. The $N^2$ modifications included methyl, ethyl, isobutyl, benzyl, $CH_2$-napthyl, $CH_2$-anthracenyl, and, in some cases, $CH_2$-benzo[a]pyrenyl derivatives [76–80]. Bacteriophage T7 DNA polymerase (exonuclease$^-$) and HIV-1 reverse transcriptase are both able to bypass the $N^2$-methyl derivative efficiently, although significantly less efficiently than unmodified DNA, but are not able to bypass any of the larger adducts [80]. Moreover, even the methyl substituent caused a high frequency of misincorporation [80]. On the other hand, each of the human Y family DNA polymerases is more tolerant of the size-expanded bases [76–79]. Rev1 is the most tolerant of $N^2$-dG-substitutions, followed by pol iota and pol kappa, whereas pol eta is the least tolerant, showing a decrease in activity of approximately two orders of magnitude between the $CH_2$-napthyl and $CH_2$-anthracenyl substituents [79]. A similar analysis of $O^6$-substituted bases showed that only Rev1 and pol iota could tolerate size-expanded substituents up to the benzyl substitution, but pol eta and pol kappa showed decreased activity even with an $O^6$-methyl substitution [79, 81]. The use of a series of well-defined synthetic base modifications provides insights into the steric constraints of DNA polymerase active sites and allows detailed comparisons to be made between replicative and damage-bypass polymerases.

## 3. Methyl-Substituted Phenyl Analogs

Efforts have been made to examine how DNA polymerases recognize methyl-substituted phenyl-based analogs that do not appear to be large enough to perturb DNA structure (Figure 3) [82]. There was significant self-base pairing of these substituted phenyl nucleobase analogs, which was not observed with the benzene analog [82]. Generally, in incorporation opposite these analogs, the Klenow fragment discriminates most against dCTP and dGTP, which tend to be the most hydrophilic nucleotides, while dTTP incorporation varies with the extent of methyl substitution, and dATP is added to these bases most efficiently [82]. The 2-substituted methyl-bearing phenyl groups generally favored dATP addition, but with the 3-substituted benzene rings, KF discriminated against dATP [82]. Interestingly, KF inserts dATP opposite MM1, DM2, DM5, and TMB (Figure 3) with a rate comparable to that of template dT [82]. This is hypothesized to be related not just to shape mimicry of dT in the template but to the placement of the specific substituents on the phenyl ring, which when appropriately oriented, can foster hydrophobic packing with the incoming dATP [82]. The most efficiently extended of these small substituted benzene derivatives are the ones that contain a methyl group at the 4-position [82]. Subsequent work using methoxy substituents, which unlike the methyl-substituted phenyl rings can form hydrogen bonds, suggests that positioning a hydrogen bond acceptor in the minor groove enhances both selectivity and efficiency of DNA synthesis by KF [83].

## 4. Hydrophobic Base Analogs

The use of hydrophobic and van der Waals interactions have been the driving force behind the development of a variety of unnatural nucleobases as possible base pairing partners and to assess polymerase utilization (Figure 4) [84]. The first of these is a self-pairing base known as 7-propynyl isocarbostyril nucleoside (PICS), which stabilizes the DNA helix when paired with itself but is destabilizing when paired with dA, dC, dG, or dT [84]. The PICS base does not demonstrate structural similarity to the natural bases, but the incorporation of dPICSTP opposite PICS in the template strand by KF is more efficient than the natural bases, ranging from 20-fold more efficient than dTTP insertion opposite dPICS to ~140-fold more efficient than dGTP insertion opposite PICS [84]. KF does not extend beyond the

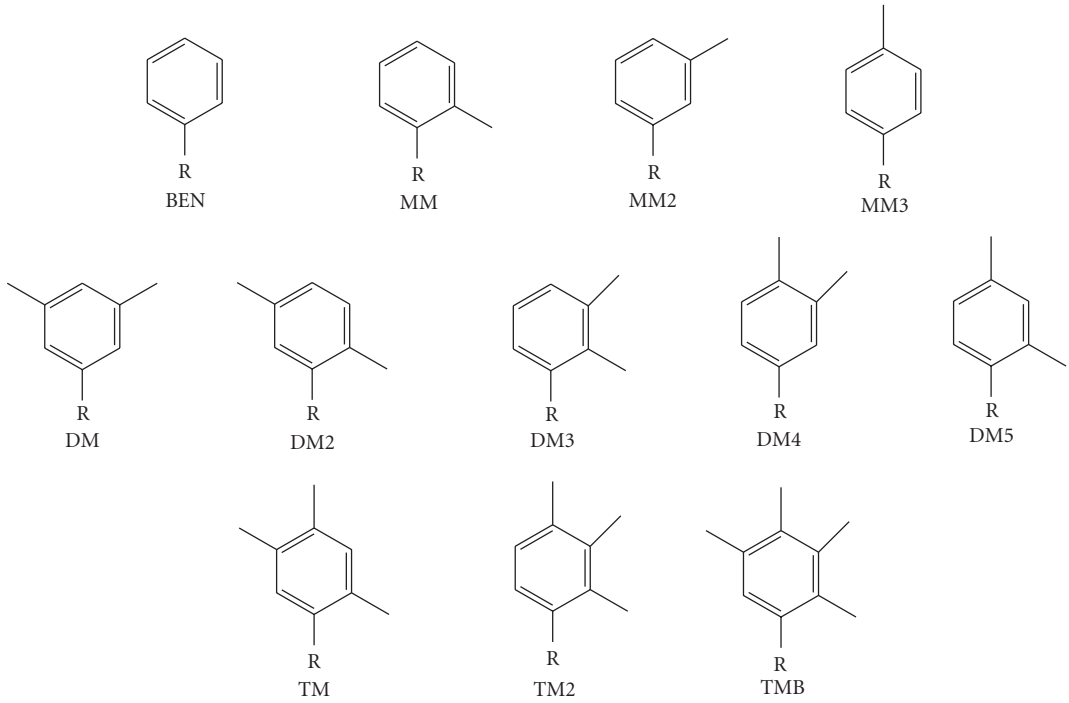



Figure 3: Small molecule analogs based on the benzene parent. BEN is benzene; the MM series is monomethylated at the 2, 3, and 4 positions, respectively; the DM series is dimethylated; the TM series is trimethylated; TMB is the tetramethylated benzene analog [82].

PICS : PICS base pair, however, which is postulated to be due to a perturbation in the position of the 3′-OH of the growing primer strand caused by the nonnatural base pair [84].

Other hydrophobic nonnatural nucleobases are based on either the naphthalene system (Figure 4), nitrogenous base-like skeleton (Figure 4), or the skeleton of benzene (Figure 4) substituted with methyl, halide, or cyano groups [85]. While dATP is the nucleoside triphosphate most generally inserted opposite these analogs, the bromo and cyano adducts show interesting differences in KF discrimination, in that dG, dT, and dC are incorporated across from the 2-bromo derivative within threefold of the catalytic efficiency of dATP incorporation [85]. The cyano derivative at the same position leads to incorporation of dGTP ∼sevenfold less efficiently than dATP, incorporation of dTTP even less efficiently, with no dCTP incorporation detected [85]. Relative to the benzene parent, only incorporation of dCTP from the 4-bromo derivative and dGTP paired opposite the 2-cyano derivative were more efficient [85]. KF primer extension after unnatural base pairing is more intriguing; specifically, 3-position substituted benzenes showed no detectable extension, with the exception of the 3-fluoro derivative which may be too small to inhibit extension due to steric hindrance (Figure 4) [85]. The base pair 4Br : 2CN was the most efficiently extended, most likely because the CN can act as a hydrogen bond acceptor and can be a driving force in primer extension [85].

## 5. Purine/Pyrimidine Mimics

Pyrimidine nucleotide analogs can affect polymerase activity in different ways. On one hand, pyrimidine nucleotide analogs lacking the 2-keto group can inhibit DNA polymerase activity [87]. Specifically, 2-amino-5-(2′-deoxy-$\beta$-D-ribofuranosyl)pyridine-5′-triphosphate (d*CTP), a cytosine analog, and 5-(2′-deoxy-$\beta$-D-ribofuranosyl)-3-methyl-2-pyridone-5′-triphosphate (d*TTP), a thymine analog, completely block Taq DNA polymerase from inserting them along a growing DNA strand (Figure 5) [87]. In these two analogs, in addition to the keto deletion, the C–N glycosidic bond functionality is removed and replaced with a slightly longer C–C bond, which may alter steric and electronic complementarity between the nucleotides and the polymerase [87]. These modified triphosphates, however, are tolerated by T7 RNA polymerase [88]; thus, it was concluded that the lack of the carbonyl functionality of these analogs is more responsible for the inhibition of Taq DNA polymerase than that of the longer C–C bond [87].

An effort to probe recognition of purines by *Bacillus stearothermophilus* DNA pol I utilized a number of azapurine derivatives and found that substitutions of carbon at N-1 or N-3 caused the most severe defects in efficiency, whereas alterations at N-1 or $N^6$ resulted in loss of fidelity [89]. A similar type of analysis found that removal of the exocyclic 2-amino group of G had little effect on the efficiency of either T7 DNA polymerase or Dpo4 [90]. However, replacement of the 2-amino group by progressively larger and less electronegative substituents, F, O, and Br, led to decreasing activity by both T7 DNA pol and Dpo4 [90]. This observation led to the suggestion that the trend was due to both the size and charge of the C-2 substituent [90].

Azole heterocyclic carboxamides can act as nucleobase mimics and, in fact, structurally can take on the appearance

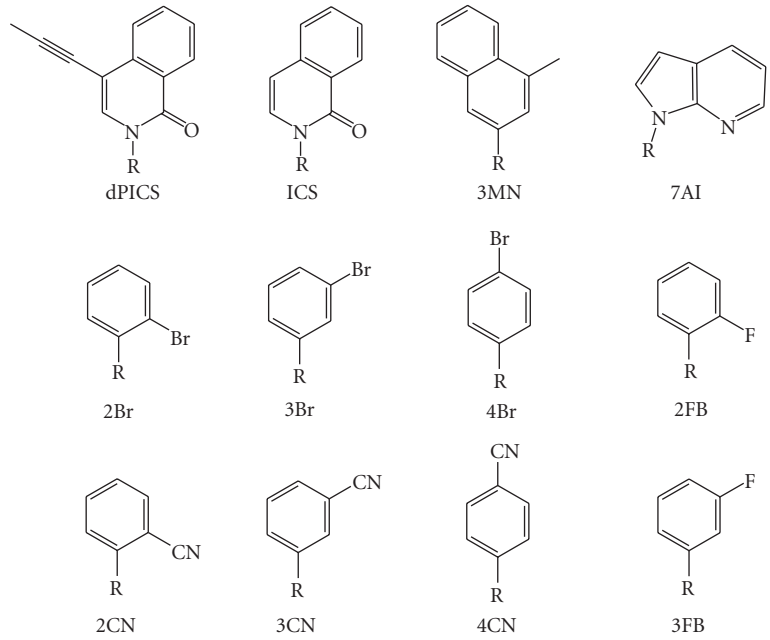
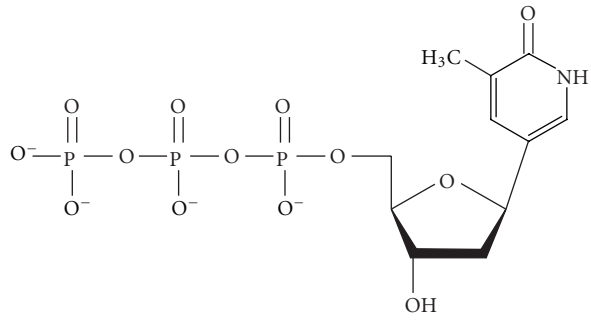



Figure 4: Hydrophobic nucleobase analogs: 7-propynyl isocarbostyril nucleoside (dPICS) [84]; isocarbostyril nucleoside (ICS); 3-methylnaphthalene (3MN); azaindole (7AI) [85]; bromo phenyl derivatives at positions 2, 3, and 4 (2Br, 3Br, and 4Br, resp.); cyano derivatives at positions 2, 3, and 4 (2CN, 3CN, and 4CN, resp.), fluoro derivatives at positions 2 and 3 (2FB, 3FB, resp.) [86].

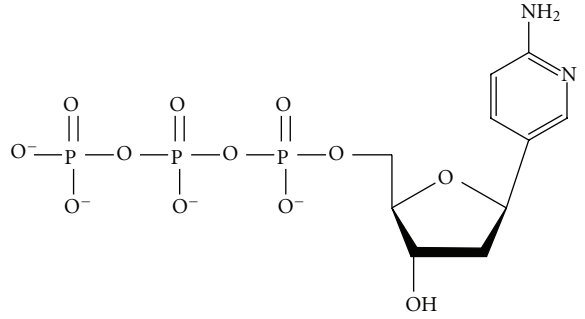
of either purines or pyrimidines (Figure 6) [91]. Because these analogs are small, they have some molecular mobility and can shift in order to adjust the hydrogen bonding patterns and electronic interactions to allow pairing with different bases [91]. Each of these azole heterocyclic carboxamides show some preference for pairing with specific incoming dNTPs, based on the position of the hydrogen bond donors and acceptors (Figure 6) [91]. For example, (1H)-1,2,3-Triazole-4-carboxamide directs the insertion of dGTP, but others do not [91]. The modified bases 1,2,4-triazole-3-carboxamide and 1,2,3-triazole-4-carboxamide, as well as 1,2-pyrazole-3-carboxamide orient in a way to promote hydrogen bonding to dC [91]. Taq DNA polymerase can utilize these analogs in PCR reactions but has different incorporation efficiencies for the different analog-dNTP pairs [91]. The presence of an azole analog in a DNA template reduces the catalytic efficiency for matched versus mismatched base pairs from 1000-fold discrepancy for natural base pairs to ~50-fold difference for base pairs involving azole analogs [91]. Therefore, these analogs are treated less stringently, but also incorporated less efficiently than natural bases by Taq DNA pol I, and demonstrate the complexity of the process of nucleotide addition, which involves electrostatic interactions, hydrogen bonding, and shape recognition [91].

Other scaffolds for unnatural self-pairing heteroatom-containing purine mimics have been developed, known as furo or thieno pyridinones (furo[2,3-c]pyridin-7(6H)-one: 7OFP, thieno[2,3-c]pyridin-7(6H)-one: 7OTP, furo[2,3-c]pyridin-7-thiol: 7TFP, furo[3,2-c]pyridin-4(5H)-one: 4OFP, thieno[3,2-c]pyridin-4(5H)-one: 4OTP, furo[3,2-c]pyridin-4-thiol: 4TFP) (Figure 6) [86]. The goal of using these analogs is to increase the ability of the DNA polymerase to continue to extend after the analog is bypassed, which is an important step in DNA polymerization, especially for DNA damage tolerance [34–37]. The most stable base pairing of these analogs is self-pairing followed by dA, dG, dC in that order, with the sulfur moiety providing more stabilization than that of oxygen [86]. KF does not discriminate strongly when synthesizing the furo versus the thieno pyridinones as self-pairs but does exhibit differences when extending beyond the unnatural bases when they are self-paired [86]. Most of these analogs disrupted the addition of dCTP to dG at the next nucleotide position after the pyridinone self-pair, with the exception of 4TFP [86]. No natural nucleotide triphosphate is found to be inserted by KF opposite 7TFP making it the most selective. The pyridinone 4OTP is the second most selective for its self-pairing, with only dTTP a modest 1.7-fold more efficiently incorporated, and selectivity drops in the following order: 7OTP, 4TFP, 4OFP, with 7OFP being the least selective [86]. Each of these analogs, with the exception of 4OFP nucleotide triphosphate, is efficiently incorporated by KF opposite dG, with the other templating bases having lower incorporation efficiencies but that are within 20-fold of the natural DNA pairs being synthesized [86]. The extension beyond these analogs by KF polymerase increases by at least fivefold over the PICS-type analogs [86]. The purine mimic 5-nitro-indolyl-2′-deoxyribose-5′-triphosphate is known to block *E. coli* DNA replication, not by inhibiting the polymerase directly but by inhibiting the ability of the clamp loader to assemble the entire replisome by blocking ATP binding and hydrolysis [93]. However, in the Taq system, a directed evolution experiment led to the identification of



5-(2′-deoxy-$\beta$-D-ribofuranosyl)-3-methyl-2-pyridone-5′-triphosphate (d*TTP)

(a)



2-amino-5-(2′-deoxy-$\beta$-D-ribofuranosyl) pyridine-5′-triphosphate (d*CTP)

(b)

Figure 5: Taq polymerase activity is severely inhibited by (a) 5-(2′-Deoxy-$\beta$-D-ribofuranosyl)-3-methyl-2-pyridone-5′-triphosphate d*TTP and (b) 2-Amino-5-(2′-deoxy-$\beta$-D-ribofuranosyl)pyridine-5′-triphosphate (d*CTP), dC, and dT analogs, respectively, lacking the 2-keto groups [87].

a DNA polymerase variant containing multiple mutations that facilitates bypass of the 5-nitro-indole analog, while polymerization by wild-type Taq was strongly blocked [94]. The mutations were concentrated in and near the active site but were also found throughout the DNA polymerase, indicative of the multiple mechanisms by which this Taq DNA polymerase variant is able to copy unusual DNA structures [94].

There is evidence that some unnatural dA mimics paired with abasic sites are proofread. Purines are generally added opposite abasic sites; unnatural nucleotides based on the indole scaffold substituted at the five position (Figure 7) were used to probe insertion by T4 DNA polymerase [95]. Despite the difference in size and shape, both 5-phenyl-indolyl-2′deoxyriboside triphosphate (5-PhITP) and 5-nitro-indolyl-2′-deoxyriboside triphosphate (5-NITP) are rapidly incorporated opposite an abasic site, whereas the 5-fluoro (dFITP) and 5-amino (dAITP) analogs have a very low efficiency of incorporation; the increase in $\pi$ electrons of the former is apparently a key contributor to catalytic efficiency [95]. Two of these analogs, dNITP and dPhITP, are used as chain terminators (Figure 7) [96] but are excised more efficiently when inserted opposite a natural nucleoside as opposed to an abasic site [96]. Evidence also exists for structural changes to allow these chain terminators to be readily incorporated. Furthermore, KF proofreads bases paired with the template purine analog 4-methylbenzimidazole as efficiently as it proofreads natural mismatches; however, it is less efficient at removing 4-methylbenzimidazole from a primer terminus, suggesting that natural bases may be specifically recognized by the exonuclease active site [97].

Modified bases 6H,8H-3,4-dihydropropyrimido[4,5-c] oxazin-7-one (P) and $N^6$-methoxy-2,6-diaminopurine (K) are generic pyrimidine and purine mimics, respectively (Figure 7) [99]. Taq DNA polymerase copies each of these as expected: P is treated generically as a pyrimidine in the template strand, pairing with either dG or dA, and K is treated by Taq as a general templating purine, pairing with either dC or dT [99]. Taq shows a preference to use P as dT in PCR reactions, giving a dT : dC ratio of 3 : 2, while preferring to use K as dA, with an dA : dG ratio of 7 : 1 [99]. These analogs are effective as universal bases due to the prevalence of tautomeric forms, observed in nuclear magnetic resonance (NMR) experiments, that allow base pairing to multiple partners [98–100].

## 6. isoC and isoG

isoC and isoG were recognized as forming base pairs in DNA and RNA in the late eighties and early nineties (Figure 8) [101, 102] and then were accepted as a third base pair of DNA in 2003 [103]. The isoC : isoG base pair is different from its natural counterparts in the transposition of the amine and carbonyl groups on both dG and dC; however, standard Watson-Crick hydrogen-bonding is still present [104]. These analogs were first demonstrated to be useful in improving PCR efficiency [105]. isoG can take on the enol form, which base pairs with T, but it can also adopt the keto form, which base pairs readily with the thymine analog 5-methylisocytosine (MiC) [106]. The recombination protein RecA can mediate strand exchange with DNA containing iG and MiC base pairs at rates comparable to those of the natural bases, which expands the range of recombination-competent genetic material [104].

## 7. Thymidine Analogs

Thymidine analogs have been particularly useful for probing DNA replication. Difluorotoluene, for example, is a synthetic dT analog, in which the hydrogen bonding capabilities seen in dA : dT base pairing are reduced or eliminated (Figure 9) [107–109]. Nevertheless, this analog can serve as a very good templating base for KF [107]. Difluorotoluene promotes efficiency of insertion as the incoming nucleotide only about fourfold less than that of natural dTTP [107]. When dA at the primer end is paired with difluorotoluene as the template base, dA is removed by KF exonucleolytic proofreading as efficiently as a natural base mismatch [97]. A similar effect was observed with human mitochondrial DNA polymerase gamma [110]. On the other hand, when difluorotoluene is

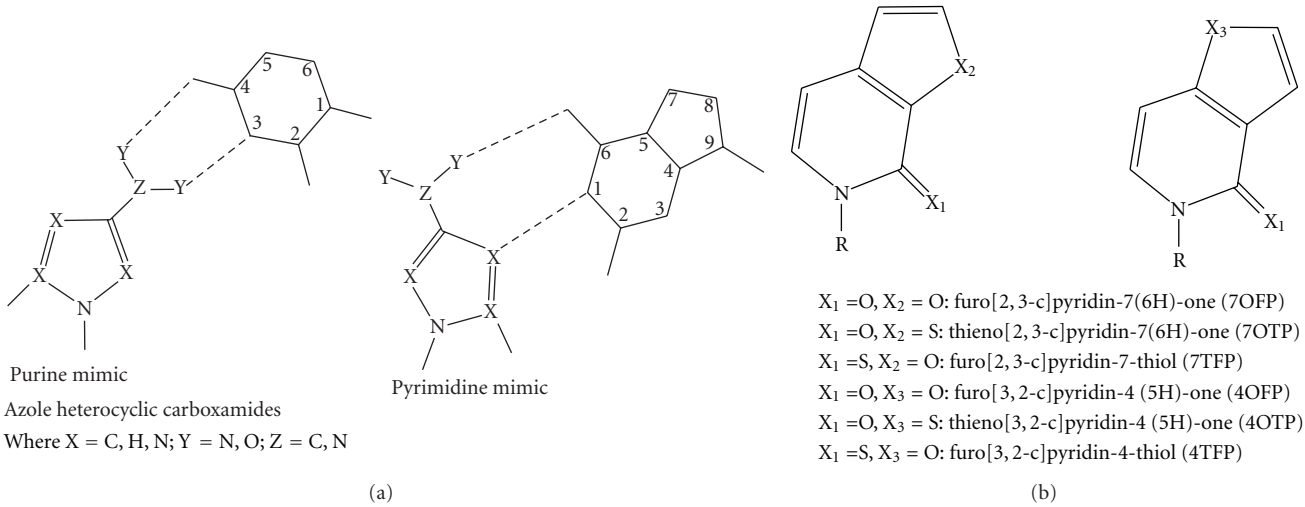



Figure 6: Classes of purine/pyrimidine mimics. (a) Skeleton structure of azole heterocyclic carboxamides. Depending on the position of the heteroatoms and their ability to donate or accept hydrogen bonds, these analogs can form base pairs with either a purine or a pyrimidine [91]. (b) General structure of the furo/thieno pyridinones; different heteroatoms in each position give the following compounds: furo[2,3-c]pyridin-7(6H)-one: 7OFP, thieno[2,3-c]pyridin-7(6H)-one: 7OTP, furo[2,3-c]pyridin-7-thiol: 7TFP, furo[3,2-c]pyridin-4(5H)-one: 4OFP, thieno[3,2-c]pyridin-4(5H)-one: 4OTP, furo[3,2-c]pyridin-4-thiol: 4TFP [86, 92].

at the primer terminus, the relative efficiency of removal is approximately 40-fold lower than that of natural base mismatches, again suggesting that specific interactions with natural bases govern removal by the exonuclease domain [97]. Difluorotoluene is an efficient template base for KF [97]. In contrast, difluorotoluene is poorly replicated by yeast pol eta and human pol kappa [112, 113], while *S. solfataricus* Dpo4 exhibits low activity but is able to carry out primer extension on templates containing difluorotoluene [114].

Hydrogen bonding capacities can be retained in a structure such as 2-thioTTP, in order to improve fidelity of PCR, which can be decreased by the tautomerization of dG to form the isoG minor tautomer [105, 115]. Use of 2-thioTTP increases fidelity of those PCR reactions that include isoC and isoG [92, 105, 116] by 5% using KlenTaq DNA polymerase [115]. This is due to introducing a specific steric interaction that prevents pairing between isoG and 2-thioTTP [115]. The yellow-colored 4-Se-T is also capable of hydrogen bonding with dA and is efficiently incorporated as 4-SeTTP into DNA by KF [117].

Thymidine analogs have also been used to study the steric interactions that govern nucleotide additions. Incrementally increasing the size of the substituent in place of the carbonyl oxygen on thymidine with a series of halide substitutions (F, Cl, Br, I) demonstrates that the replicative polymerase KF has a specific "tightness" that allows for only some substitutions to be incorporated. The highest efficiency of incorporation by KF was with base pairs that are larger than natural base pairs [111]. In contrast, T7 DNA polymerase is more stringent and has an optimum that is closer to the size of natural base pairs [111, 118]. Moving the substituents around the thymidine ring and probing the activity of KF led to the conclusion that KF is remarkably sensitive to the overall shape of the template base and incoming nucleotide [119]. KF achieves maximal fidelity of incorporation with the chlorosubstituted analog 2,4-dichloro-5-toluene-1-$\beta$-D-deoxyriboside (Figure 9) [111]. The catalytic efficiency of KF with these analogs showed that with the increase in size by 0.66 Å (H → Cl), KF was more efficient by a factor of ~180 [111]. This trend of increasing steric hindrance with these thymidine analogs utilized by KF is consistent with the steric hindrance seen with 4′ substituted dTTP analogs noted previously [120]. In contrast, the presence of 4′ substituted T analogs in the template are well tolerated by KF [121]. The model Y family DNA polymerase, *S. solfataricus* Dbh, incorporates 4′-modified dTTP analogs relatively efficiently and binds to the analogs nearly as well as binding to unmodified dTTP [122]. Similarly, Y family DNA polymerase Dpo4 exhibits much less size selectivity than KF, as determined with halogen-substituted thymine analogs [123]. Thus, although some Y family DNA polymerases require hydrogen bonding for efficient replication, these studies confirm their generally accommodating active sites.

## 8. Fluorescent Base Analogs

*8.1. 2-Aminopurine.* The most common fluorescent base analog in use today is 2-aminopurine (2AP), which can form hydrogen bonds and base pair with either of the pyrimidines thymine or cytosine (Figure 10) [124]. A recent crystal structure of DNA containing a 2AP : dC base pair in the active site of the Y567A variant of RB69 DNA polymerase suggests that the 2AP : dC pair may contain a bifurcated hydrogen bond between $N^2$-H of 2AP and N3 and O2 of dC [125]. In this example, the Y567A active site mutation in the nascent base-pair-binding pocket is both less discriminating in the formation of mismatched base pairs and is better able to extend mismatched primer termini [125]. The modified base

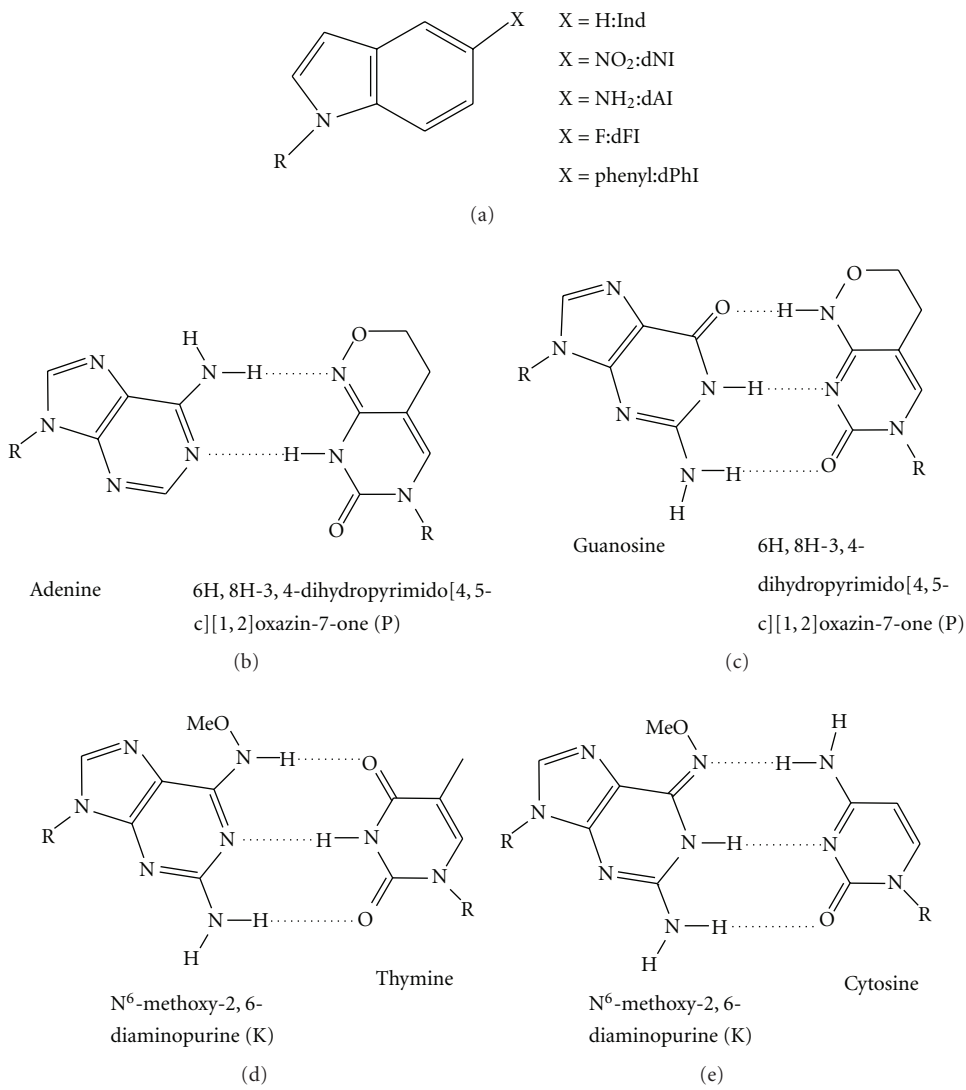



Figure 7: Purine and pyrimidine mimics. (a) Purine analogs based on the indole scaffold: 5-substituted indolyl-2′deoxyriboside triphosphates where X = H, indole (Ind); X = $NO_2$, 5-nitro-1-indole (dNI); X = $NH_2$, 5-amino-1-indole (dAI); X = F, 5-fluoro-1-indole (dFI); X = phenyl, 5-phenyl-1-indole (dPhI) [95, 96]. ((b)–(e)) Pyrimidine mimic 6H,8H-3,4-dihydropyrimido[4,5-c][1,2]oxazin-7-one (P), and purine mimic $N^6$-methoxy-2,6-diaminopurine (K) base pairing partners. (b) Adenine : P. (c) Guanosine : P. (d) Thymine : K. (e) Cytosine : K. The ability of these analogs to form different tautomers can result in mutations [98].

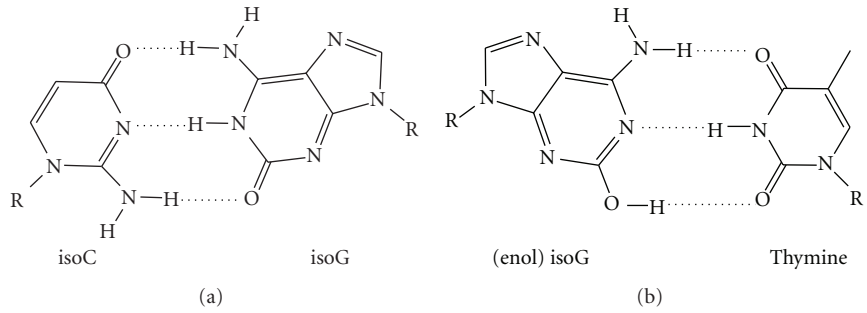



Figure 8: isoC, isoG, and their base pairs (a) isocytosine : isoguanosine and (b) (enol) isoG : thymine [101–106].

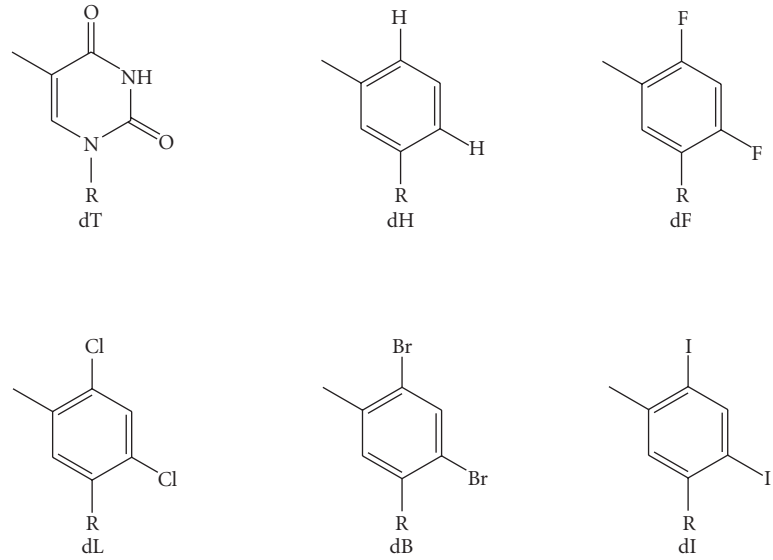



Figure 9: Small molecule thymidine (dT) analogs: 3-toluene-1-$\beta$-D-deoxyriboside (dH); 2,4-difluoro-5-toluene-1-$\beta$-D-deoxyriboside (dF); 2,4-dichloro-5-toluene-1-$\beta$-D-deoxyriboside (dL); 2,4-dibromo-5-toluene-1-$\beta$-D-deoxyriboside (dB); 2,4-diiodo-5-toluene-1-$\beta$-D-deoxyriboside (dI) [111]. The deoxyribose groups are indicated by R.

2AP is commercially available and has been used to study a number of DNA-binding protein interactions including KF [126], EcoRI, DNA methyltransferase [127], endonuclease [128, 129], and uracil DNA glycosylase [130]. The fluorescence of this analog is sequence-context dependent, with the most pronounced effect occurring when the base is surrounded by other purines; much like other fluorescent nucleobases, its fluorescence is quenched when it is within DNA [131]. KF has been shown to utilize 2AP, and the fluorescence has been used to give insights into the dynamics of this protein as it synthesizes DNA [126, 132, 133]. For example, in one FRET experiment with a labeled KF, the mechanism of the fingers closing conformational change was studied [133] and was found to be influenced by the added nucleotide. Specifically, mismatched nucleotides are detected before the polymerase "closes" on the DNA suggesting that the mismatched nucleotide itself may destabilize the "open" polymerase conformation [133]. The role in the conformational change of the divalent cation (usually $Mg^{2+}$ or $Ca^{2+}$ but, in this case, an "exchange inert" Rh(III)) was also probed using 2AP [134], and it was found that dNTP binding in the absence of the correct ion can induce the conformational shift [134]. The ability of the ion to diffuse to the proper position before the nucleophilic attack can occur may influence the reverse conformational shift observed in the presence of the incorrect nucleotide [134].

Fluorescence spectroscopy with 2AP can be used to study DNA polymerization on a millisecond time scale, and probe single events like nucleotide addition, base pairing interactions, and subsequent excision via nuclease activity [126, 132]. Insertion kinetics have been measured for the monophosphate version of 2AP (dAPMP versus dAMP); dAPMP is found to be misincorporated at similar rates to the incorporation of the natural triphosphate dATP opposite dT by KF [126]. This makes 2AP useful in studying polymerase activity as it is misincorporated about as frequently as dA is incorporated. However, this incorporation is influenced by the sequence surrounding the primer terminus, with double the rate of misincorporation of 2AP triphosphate if the nearest neighbor to the nascent base pair is dG, dC, or dA, as compared to dT [126].

Y family polymerases also have been studied using 2AP. Dbh adds dTTP correctly opposite 2AP in the template strand and binds various DNA substrates containing 2AP with $K_D$ values similar to those of natural DNA substrates [135]. Use of 2AP to monitor conformation changes during the base-skipping phenomenon, which can generate frameshift mutations as seen with Y family polymerases, provides evidence that the misincorporation pathway is distinct from the correct dNTP incorporation process [135]. Fluorescence from 2AP has been also used to probe the proofreading mechanism by which bases are excised via nuclease activity of phage T4 polymerase [136].

The analog 2AP has been used together with the base analog pyrrolo-dC as a FRET pair as the excitation and emission wavelengths of these two nucleotide probes are compatible [137], though this pair has not yet been utilized to study DNA polymerases. Pyrrolo-dC alone has been used to study DNA/RNA hybrids [138], single-stranded DNA hairpins [139], and base pair flipping [140]. Two potential drawbacks of using 2AP are the sequence dependence of its fluorescence and that it can perturb the DNA structure or be mutagenic if it forms a wobble pair with dT [124]. A 2AP : dT base pair destabilizes duplex DNA by ~8°C relative to a dA : dT base pair [141].

*8.2. tC: 1,3-Diaza-2-oxophenothiazine.* The synthetic cytosine analog tC was developed first by Lin et al. [142] but

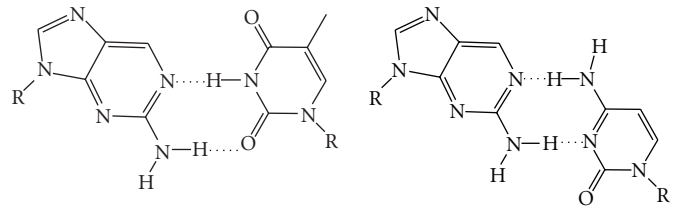



Figure 10: 2-Aminopurine (2AP) and its base pairs (a) 2AP base paired with thymine and (b) 2AP base paired with cytosine [124].

then used as a probe of DNA polymerases by Wilhelmsson and coworkers [143–145]. The fluorescence quantum yield of this nucleotide analog, unlike 2AP, is not sensitive to the surrounding environment [144, 146]. This base also is incorporated into DNA, shows canonical base pairing with guanosine (Figure 11), and does not perturb the B-form structure of DNA. In fact, a dG : tC base pair stabilized DNA by 3°C [124]. Different DNA polymerases have different efficiencies in utilizing tC in template DNA and in incorporating tC into the growing DNA primer strand. For example, KF utilizes template tC in preference to a template C, as KF apparently has a flexible enough active site to accommodate the extra cyclic ring system. Klenow also preferentially incorporates the tC nucleotide triphosphate in the growing DNA strand. *E. coli* DinB (pol IV), which is a Y family DNA polymerase [17], also utilizes the tC nucleotide triphosphate more efficiently than dCTP, similar to Klenow [147]. DinB also can extend from tC at the primer terminus [147]. However, DinB shows a 12-fold decrease in the catalytic efficiency of incorporation of dGTP opposite template tC as compared to the natural dC in the template strand and is unable to extend from the newly generated primer terminus [147]. Primer extension by DinB is inhibited unless the primer terminus is at least 3-4 nucleotides beyond the tC analog, which suggests that the "TLS patch" of nucleotides required beyond noncognate bases for DNA polymerases to resume efficient synthesis is shorter for a Y family DNA polymerase than for replicative polymerases. Moreover, the striking asymmetry of the DinB active site has also been observed in the case of B family DNA polymerases human polymerase alpha and herpes simplex virus I DNA polymerase when probed with nonnatural nucleotide analogs [148].

*8.3. tC°: 1,3-Diaza-2-oxophenoxazine.* The oxo-analog of tC is tC°, 1,3-diaza-2-oxophenoxazine (Figure 11) [142], which has several similar properties to that of tC in that it stabilizes B-form DNA by 3°C and it base pairs with G in a standard Watson-Crick configuration [149]. It is exceptionally bright, on average 10–50 times brighter than 2AP, 3-MI, and 6-MAP [149]. The tC° analog, like tC, can be utilized by KF and by human DNA primase [146, 150, 151]. This analog has proven useful in high-density labeling of PCR products using a deep vent DNA polymerase and therefore should be useful in biotechnology applications [151].

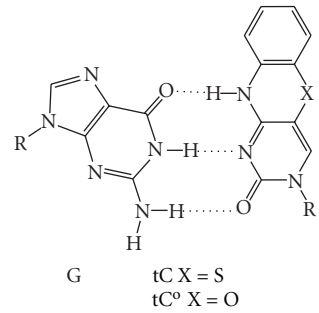



Figure 11: Fluorescent cytosine analogs tC and tC° form canonical base pairs with dG, do not perturb B-form DNA structure, and can pair with a FRET donor to probe DNA polymerase dynamics [144–146, 149, 150].

## 9. Conclusions

Nonnatural nucleotides continue to provide an important tool for the study of DNA and its interacting protein partners. In particular, DNA polymerases that are responsible for the systematic replication of DNA, whether accurate or mutagenic, are required to specifically recognize and efficiently base pair with a large number of noncanonical DNA structures. An increasingly expanding genetic alphabet of nonnatural nucleobases provides the ability to obtain an unparalleled level of detail about how DNA polymerases discriminate among many different DNA structures. From the first introduction of artificial abasic sites [39] to the use of bright nonperturbing fluorescent analogs that are used to probe polymerase opening and closing dynamics on a nascent base pair [144, 146], nonnatural nucleotides are now fully integrated into DNA polymerase research. There remains however a need for novel DNA bases that have specific properties in order to better study the interactions of DNA polymerases with DNA. In particular, efficiently generating both phosphoramidite monomers and triphosphate versions of a given modified base can be a significant synthetic challenge. The understanding of DNA polymerase specificity for synthetic nucleobases, discussed in this paper and elsewhere [28, 38, 124, 152, 153], is increasing; in the future, synthetic bases will continue to be used for a variety of purposes including probing proteins and small molecules that bind to DNA, optimizing unnatural bases for coding as a synthetic genetic code [154], synthesizing unnatural

biopolymers [155], and improving the prospects of DNA as a nanomaterial and a drug target [156].

## Acknowledgments

This work was supported by the National Science Foundation CAREER Award, Grant no. MCB-0845033, and the Northeastern University Office of the Provost. P. J. Beuning is a Cottrell Scholar of the Research Corporation for Science Advancement. The authors thank Jana Sefcikova and Lisa Hawver for helpful suggestions on the paper.

## References

[1] R. E. Franklin and R. G. Gosling, "Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate," *Nature*, vol. 172, no. 4369, pp. 156–157, 1953.

[2] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[3] A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid," *Harvey Lectures*, vol. 53, pp. 83–112, 1957.

[4] K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, and H. G. Khorana, "Studies on polynucleotides. XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases," *Journal of Molecular Biology*, vol. 56, no. 2, pp. 341–361, 1971.

[5] A. M. Cordonnier and R. P. Fuchs, "Replication of damaged DNA: molecular defect in Xeroderma pigmentosum variant cells," *Mutation Research*, vol. 435, no. 2, pp. 111–119, 1999.

[6] A. Bresson and R. P. Fuchs, "Lesion bypass in yeast cells: pol eta participates in a multi-DNA polymerase process," *EMBO Journal*, vol. 21, no. 14, pp. 3881–3887, 2002.

[7] V. Pages and R. P. Fuchs, "How DNA lesions are turned into mutations within cells?" *Oncogene*, vol. 21, no. 58, pp. 8957–8966, 2002.

[8] S. S. Lange, K. Takata, and R. D. Wood, "DNA polymerases and cancer," *Nature Reviews Cancer*, vol. 11, no. 2, pp. 96–110, 2011.

[9] L. A. Loeb and R. J. Monnat Jr., "DNA polymerases and human disease," *Nature Reviews Genetics*, vol. 9, no. 8, pp. 594–604, 2008.

[10] C. Miller, L. E. Thomsen, C. Gaggero, R. Mosseri, H. Ingmer, and S. N. Cohen, "SOS response induction by beta-lactams and bacterial defense against antibiotic lethality," *Science*, vol. 305, no. 5690, pp. 1629–1631, 2004.

[11] R. Napolitano, R. Janel-Bintz, J. Wagner, and R. P. Fuchs, "All three SOS-inducible DNA polymerases (pol II, pol IV and pol V) are involved in induced mutagenesis," *EMBO Journal*, vol. 19, no. 22, pp. 6259–6265, 2000.

[12] M. R. Albertella, C. M. Green, A. R. Lehmann, and M. J. O'Connor, "A role for polymerase eta in the cellular tolerance to cisplatin-induced damage," *Cancer Research*, vol. 65, no. 21, pp. 9799–9806, 2005.

[13] M. R. Albertella, A. Lau, and M. J. O'Connor, "The overexpression of specialized DNA polymerases in cancer," *DNA Repair*, vol. 4, no. 5, pp. 583–593, 2005.

[14] E. Bassett, N. M. King, M. F. Bryant et al., "The role of DNA polymerase eta in translesion synthesis past platinum-DNA adducts in human fibroblasts," *Cancer Research*, vol. 64, no. 18, pp. 6469–6475, 2004.

[15] Y. W. Chen, J. E. Cleaver, F. Hanaoka, C. F. Chang, and K. M. Chou, "A novel role of DNA polymerase eta in modulating cellular sensitivity to chemotherapeutic agents," *Molecular Cancer Research*, vol. 4, no. 4, pp. 257–265, 2006.

[16] T. V. Ho, A. Guainazzi, S. B. Derkunt, M. Enoiu, and O. D. Scharer, "Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases," *Nucleic Acids Research*, vol. 39, pp. 7455–7464, 2011.

[17] H. Ohmori, E. C. Friedberg, R. P. Fuchs et al., "The Y-family of DNA polymerases," *Molecular Cell*, vol. 8, no. 1, pp. 7–8, 2001.

[18] P. H. Patel, M. Suzuki, E. Adman, A. Shinkai, and L. A. Loeb, "Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection," *Journal of Molecular Biology*, vol. 308, no. 5, pp. 823–837, 2001.

[19] R. G. Federley and L. J. Romano, "DNA polymerase: structural homology, conformational dynamics, and the effects of carcinogenic DNA adducts," *Journal of Nucleic Acids*, vol. 2010, Article ID 457176, 15 pages, 2010.

[20] P. M. Burgers and F. Eckstein, "A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate," *The Journal of Biological Chemistry*, vol. 254, no. 15, pp. 6889–6893, 1979.

[21] J. Adler, I. R. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. IV. linkage of single deoxynucleotides to the deoxynucleoside ends of deoxyribonucleic acid," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 44, pp. 641–647, 1958.

[22] M. J. Bessman, I. R. Lehman, J. Adler, S. B. Zimmerman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. III. The incorporation of pyrimidine and purine analogues into deoxyribonucleic acid," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 44, pp. 633–640, 1958.

[23] M. J. Bessman, I. R. Lehman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction," *The Journal of Biological Chemistry*, vol. 233, no. 1, pp. 171–177, 1958.

[24] S. R. Kornberg, I. R. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, "Enzymatic cleavage of deoxyguanosine triphosphate to deoxyguanosine and tripolyphosphate," *The Journal of Biological Chemistry*, vol. 233, no. 1, pp. 159–162, 1958.

[25] I. R. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*," *The Journal of Biological Chemistry*, vol. 233, no. 1, pp. 163–170, 1958.

[26] A. J. Berdis, "Mechanisms of DNA polymerases," *Chemical Reviews*, vol. 109, no. 7, pp. 2862–2879, 2009.

[27] L. J. Reha-Krantz, "DNA polymerase proofreading: multiple roles maintain genome stability," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1049–1063, 2010.

[28] I. Lee and A. J. Berdis, "Non-natural nucleotides as probes for the mechanism and fidelity of DNA polymerases," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1064–1080, 2010.

[29] E. Friedberg, G. C. Walker, W. W. Siede, R. D. Wood, R. Schultz, and T. Ellenberger, *DNA Repair and Mutagenesis*, ASM Press, 2nd edition, 2006.

[30] S. J. Johnson, J. S. Taylor, and L. S. Beese, "Processive DNA synthesis observed in a polymerase crystal suggests

a mechanism for the prevention of frameshift mutations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 7, pp. 3895–3900, 2003.

[31] H. Ling, F. Boudsocq, R. Woodgate, and W. Yang, "Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication," *Cell*, vol. 107, no. 1, pp. 91–102, 2001.

[32] J. R. Kiefer, C. Mao, C. J. Hansen et al., "Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution," *Structure*, vol. 5, no. 1, pp. 95–108, 1997.

[33] P. J. Rothwell and G. Waksman, "Structure and mechanism of DNA polymerases," *Advances in Protein Chemistry*, vol. 71, pp. 401–440, 2005.

[34] R. P. Fuchs and S. Fujii, "Translesion synthesis in *Escherichia coli*: lessons from the NarI mutation hot spot," *DNA Repair*, vol. 6, no. 7, pp. 1032–1041, 2007.

[35] S. Fujii and R. P. Fuchs, "Defining the position of the switches between replicative and bypass DNA polymerases," *EMBO Journal*, vol. 23, no. 21, pp. 4342–4352, 2004.

[36] D. F. Jarosz, S. E. Cohen, J. C. Delaney, J. M. Essigmann, and G. C. Walker, "A DinB variant reveals diverse physiological consequences of incomplete TLS extension by a Y-family DNA polymerase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 50, pp. 21137–21142, 2009.

[37] V. Pages, R. Janel-Bintz, and R. P. Fuchs, "Pol III proofreading activity prevents lesion bypass as evidenced by its molecular signature within *E. coli* cells," *Journal of Molecular Biology*, vol. 352, no. 3, pp. 501–509, 2005.

[38] E. T. Kool, "Replacing the nucleobases in DNA with designer molecules," *Accounts of Chemical Research*, vol. 35, no. 11, pp. 936–943, 2002.

[39] M. Takeshita, C. N. Chang, F. Johnson, S. Will, and A. P. Grollman, "Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases," *The Journal of Biological Chemistry*, vol. 262, no. 21, pp. 10171–10179, 1987.

[40] B. S. Strauss, "The "A rule" of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions?" *BioEssays*, vol. 13, no. 2, pp. 79–84, 1991.

[41] S. Obeid, N. Blatter, R. Kranaster et al., "Replication through an abasic DNA lesion: structural basis for adenine selectivity," *EMBO Journal*, vol. 29, no. 10, pp. 1738–1747, 2010.

[42] I. Bruck, M. F. Goodman, and M. O'Donnell, "The essential C family DnaE polymerase is error-prone and efficient at lesion bypass," *The Journal of Biological Chemistry*, vol. 278, no. 45, pp. 44361–44368, 2003.

[43] E. Le Chatelier, O. J. Bécherel, E. d'Alençon et al., "Involvement of DnaE, the second replicative DNA polymerase from *Bacillus subtilis*, in DNA mutagenesis," *The Journal of Biological Chemistry*, vol. 279, no. 3, pp. 1757–1767, 2004.

[44] C. S. McHenry, "Breaking the rules: bacteria that use several DNA polymerase IIIs," *EMBO Reports*, vol. 12, no. 5, pp. 408–414, 2011.

[45] J. E. Stone, D. Kumar, S. K. Binz et al., "Lesion bypass by *S. cerevisiae* Pol zeta alone," *DNA Repair*, vol. 10, no. 8, pp. 826–834, 2011.

[46] S. Kumar, B. J. Lamarche, and M. D. Tsai, "Use of damaged DNA and dNTP substrates by the error-prone DNA polymerase X from African swine fever virus," *Biochemistry*, vol. 46, no. 12, pp. 3814–3825, 2007.

[47] K. A. Fiala, C. D. Hypes, and Z. Suo, "Mechanism of abasic lesion bypass catalyzed by a Y-family DNA polymerase," *The Journal of Biological Chemistry*, vol. 282, no. 11, pp. 8188–8198, 2007.

[48] A. Maor-Shoshani, K. Hayashi, H. Ohmori, and Z. Livneh, "Analysis of translesion replication across an abasic site by DNA polymerase IV of *Escherichia coli*," *DNA Repair*, vol. 2, no. 11, pp. 1227–1238, 2003.

[49] N. B. Reuven, G. Arad, A. Maor-Shoshani, and Z. Livneh, "The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication," *The Journal of Biological Chemistry*, vol. 274, no. 45, pp. 31763–31766, 1999.

[50] M. Tang, P. Pham, X. Shen et al., "Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis," *Nature*, vol. 404, no. 6781, pp. 1014–1018, 2000.

[51] K. Bebenek, A. Tissier, E. G. Frank et al., "5′-deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro," *Science*, vol. 291, no. 5511, pp. 2156–2159, 2001.

[52] Y. Zhang, F. Yuan, X. Wu, J. S. Taylor, and Z. Wang, "Response of human DNA polymerase iota to DNA lesions," *Nucleic Acids Research*, vol. 29, no. 4, pp. 928–935, 2001.

[53] C. Masutani, R. Kusumoto, S. Iwai, and F. Hanaoka, "Mechanisms of accurate translesion synthesis by human DNA polymerase eta," *EMBO Journal*, vol. 19, no. 12, pp. 3100–3109, 2000.

[54] R. J. Kokoska, S. D. McCulloch, and T. A. Kunkel, "The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and *Sulfolobus solfataricus* Dpo4," *The Journal of Biological Chemistry*, vol. 278, no. 50, pp. 50537–50545, 2003.

[55] S. M. Sherrer, K. A. Fiala, J. D. Fowler, S. A. Newmister, J. M. Pryor, and Z. Suo, "Quantitative analysis of the efficiency and mutagenic spectra of abasic lesion bypass catalyzed by human Y-family DNA polymerases," *Nucleic Acids Research*, vol. 39, no. 2, pp. 609–622, 2011.

[56] Y. Zhang, F. Yuan, X. Wu et al., "Error-free and error-prone lesion bypass by human DNA polymerase kappa *in vitro*," *Nucleic Acids Research*, vol. 28, no. 21, pp. 4138–4146, 2000.

[57] N. Kim, S. V. Mudrak, and S. Jinks-Robertson, "The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites," *DNA Repair*, vol. 10, pp. 1262–1271, 2011.

[58] J. M. Pryor and M. T. Washington, "Pre-steady state kinetic studies show that an abasic site is a cognate lesion for the yeast Rev1 protein," *DNA Repair*, vol. 10, pp. 1138–1144, 2011.

[59] C. Otsuka, D. Loakes, and K. Negishi, "The role of deoxycytidyl transferase activity of yeast Rev1 protein in the bypass of abasic sites," *Nucleic Acids Research*, no. 2, pp. 87–88, 2002.

[60] N. Sabouri and E. Johansson, "Translesion synthesis of abasic sites by yeast DNA polymerase epsilon," *The Journal of Biological Chemistry*, vol. 284, no. 46, pp. 31555–31563, 2009.

[61] B. Zhao, Z. Xie, H. Shen, and Z. Wang, "Role of DNA polymerase eta in the bypass of abasic sites in yeast cells," *Nucleic Acids Research*, vol. 32, no. 13, pp. 3984–3994, 2004.

[62] T. J. Matray and E. T. Kool, "A specific partner for abasic damage in DNA," *Nature*, vol. 399, no. 6737, pp. 704–708, 1999.

[63] L. Sun, K. Zhang, L. Zhou et al., "Yeast pol eta holds a Cis-Syn thymine dimer loosely in the active site during elongation opposite the 3′-T of the dimer, but tightly opposite the 5′-T," *Biochemistry*, vol. 42, no. 31, pp. 9431–9437, 2003.

[64] X. Zhang, I. Lee, X. Zhou, and A. J. Berdis, "Hydrophobicity, shape, and pi-electron contributions during translesion DNA

synthesis," *Journal of the American Chemical Society*, vol. 128, no. 1, pp. 143–149, 2006.

[65] E. Z. Reineks and A. J. Berdis, "Evaluating the contribution of base stacking during translesion DNA replication," *Biochemistry*, vol. 43, no. 2, pp. 393–404, 2004.

[66] R. E. Johnson, M. T. Washington, L. Haracska, S. Prakash, and L. Prakash, "Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions," *Nature*, vol. 406, no. 6799, pp. 1015–1019, 2000.

[67] S. Shachar, O. Ziv, S. Avkin et al., "Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals," *EMBO Journal*, vol. 28, no. 4, pp. 383–393, 2009.

[68] D. Jarosz, P. Beuning, S. Cohen, and G. C. Walker, "Y-family DNA polymerases in *Escherichia coli*," *Trends in Microbiology*, vol. 15, no. 2, pp. 70–77, 2007.

[69] J. M. Walsh, L. A. Hawver, and P. J. Beuning, "*Escherichia coli* Y family DNA polymerases," *Frontiers in Bioscience*, vol. 16, no. 7, pp. 3164–3182, 2011.

[70] L. S. Waters, B. K. Minesinger, M. E. Wiltrout, S. D'Souza, R. V. Woodruff, and G. C. Walker, "Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance," *Microbiology and Molecular Biology Reviews*, vol. 73, no. 1, pp. 134–154, 2009.

[71] M. T. Washington, K. D. Carlson, B. D. Freudenthal, and J. M. Pryor, "Variations on a theme: Eukaryotic Y-family DNA polymerases," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1113–1123, 2010.

[72] W. Yang, "Damage repair DNA polymerases Y," *Current Opinion in Structural Biology*, vol. 13, no. 1, pp. 23–30, 2003.

[73] W. Yang and R. Woodgate, "What a difference a decade makes: insights into translesion DNA synthesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 40, pp. 15591–15598, 2007.

[74] A. Maor-Shoshani, V. Ben-Ari, and Z. Livneh, "Lesion bypass DNA polymerases replicate across non-DNA segments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 14760–14765, 2003.

[75] S. Adar and Z. Livneh, "Translesion DNA synthesis across non-DNA segments in cultured human cells," *DNA Repair*, vol. 5, no. 4, pp. 479–490, 2006.

[76] J. Y. Choi, K. C. Angel, and F. P. Guengerich, "Translesion synthesis across bulky $N^2$-alkyl guanine DNA adducts by human DNA polymerase kappa," *The Journal of Biological Chemistry*, vol. 281, no. 30, pp. 21062–21072, 2006.

[77] J. Y. Choi and F. P. Guengerich, "Adduct size limits efficient and error-free bypass across bulky $N^2$-guanine DNA lesions by human DNA polymerase eta," *Journal of Molecular Biology*, vol. 352, no. 1, pp. 72–90, 2005.

[78] J. Y. Choi and F. P. Guengerich, "Kinetic evidence for inefficient and error-prone bypass across bulky $N^2$-guanine DNA adducts by human DNA polymerase," *The Journal of Biological Chemistry*, vol. 281, no. 18, pp. 12315–12324, 2006.

[79] J. Y. Choi and F. P. Guengerich, "Kinetic analysis of translesion synthesis opposite bulky $N^2$- and $O^6$-alkylguanine DNA adducts by human DNA polymerase REV1," *The Journal of Biological Chemistry*, vol. 283, no. 35, pp. 23645–23655, 2008.

[80] J. Y. Choi and F. P. Guengerich, "Analysis of the effect of bulk at $N^2$-alkylguanine DNA adducts on catalytic efficiency and fidelity of the Processive DNA polymerases bacteriophage $T_7$ exonuclease- and HIV-1 reverse transcriptase," *The Journal of Biological Chemistry*, vol. 279, no. 18, pp. 19217–19229, 2004.

[81] J. Y. Choi, G. Chowdhury, H. Zang et al., "Translesion synthesis across $O^6$-alkylguanine DNA adducts by recombinant human DNA polymerases," *The Journal of Biological Chemistry*, vol. 281, no. 50, pp. 38244–38256, 2006.

[82] S. Matsuda, A. A. Henry, and F. E. Romesberg, "Optimization of unnatural base pair packing for polymerase recognition," *Journal of the American Chemical Society*, vol. 128, no. 19, pp. 6369–6375, 2006.

[83] S. Matsuda, A. M. Leconte, and F. E. Romesberg, "Minor groove hydrogen bonds and the replication of unnatural base pairs," *Journal of the American Chemical Society*, vol. 129, no. 17, pp. 5551–5557, 2007.

[84] D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, and F. E. Romesberg, "Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base," *Journal of the American Chemical Society*, vol. 121, no. 49, pp. 11585–11586, 1999.

[85] G. T. Hwang and F. E. Romesberg, "Substituent effects on the pairing and polymerase recognition of simple unnatural base pairs," *Nucleic Acids Research*, vol. 34, no. 7, pp. 2037–2045, 2006.

[86] A. A. Henry, C. Yu, and F. E. Romesberg, "Determinants of unnatural nucleobase stability and polymerase recognition," *Journal of the American Chemical Society*, vol. 125, no. 32, pp. 9638–9646, 2003.

[87] M. J. Guo, S. Hildbrand, C. J. Leumann, L. W. McLaughlin, and M. J. Waring, "Inhibition of DNA polymerase reactions by pyrimidine nucleotide analogues lacking the 2-keto group," *Nucleic Acids Research*, vol. 26, no. 8, pp. 1863–1869, 1998.

[88] J. A. Piccirilli, S. E. Moroney, and S. A. Benner, "A C-nucleotide base pair: methylpseudouridine-directed incorporation of formycin triphosphate into RNA catalyzed by T7 RNA polymerase," *Biochemistry*, vol. 30, no. 42, pp. 10350–10356, 1991.

[89] M. Trostler, A. Delier, J. Beckman et al., "Discrimination between right and wrong purine dNTPs by DNA polymerase I from *Bacillus stearothermophilus*," *Biochemistry*, vol. 48, no. 21, pp. 4633–4641, 2009.

[90] H. Zhang, U. Bren, I. D. Kozekov, C. J. Rizzo, D. F. Stec, and F. P. Guengerich, "Steric and electrostatic effects at the C2 atom substituent influence replication and miscoding of the DNA deamination product deoxyxanthosine and analogs by DNA polymerases," *Journal of Molecular Biology*, vol. 392, no. 2, pp. 251–269, 2009.

[91] N. Paul, V. C. Nashine, G. Hoops et al., "DNA polymerase template interactions probed by degenerate isosteric nucleobase analogs," *Chemistry and Biology*, vol. 10, no. 9, pp. 815–825, 2003.

[92] A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, and F. E. Romesberg, "Efforts to expand the genetic alphabet: identification of a replicable unnatural DNA self-pair," *Journal of the American Chemical Society*, vol. 126, no. 22, pp. 6923–6931, 2004.

[93] K. Eng, S. K. Scouten-Ponticelli, M. Sutton, and A. Berdis, "Selective inhibition of DNA replicase assembly by a non-natural nucleotide: exploiting the structural diversity of ATP-binding sites," *ACS Chemical Biology*, vol. 5, no. 2, pp. 183–194, 2010.

[94] F. J. Ghadessy, N. Ramsay, F. Boudsocq et al., "Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution," *Nature Biotechnology*, vol. 22, no. 6, pp. 755–759, 2004.

[95] X. Zhang, I. Lee, and A. J. Berdis, "The use of non-natural nucleotides to probe the contributions of shape complementarity and pi-electron surface area during DNA

polymerization," *Biochemistry*, vol. 44, no. 39, pp. 13101–13110, 2005.

[96] X. Zhang, I. Lee, and A. J. Berdis, "A potential chemotherapeutic strategy for the selective inhibition of promutagenic DNA synthesis by nonnatural nucleotides," *Biochemistry*, vol. 44, no. 39, pp. 13111–13121, 2005.

[97] J. C. Morales and E. T. Kool, "Importance of terminal base pair hydrogen-bonding in 3'-end proofreading by the Klenow fragment of DNA polymerase I," *Biochemistry*, vol. 39, no. 10, pp. 2626–2632, 2000.

[98] F. Hill, D. M. Williams, D. Loakes, and D. M. Brown, "Comparative mutagenicities of $N^6$-methoxy-2,6-diaminopurine and $N^6$-methoxyaminopurine 2′-deoxyribonucleosides and their 5′-triphosphates," *Nucleic Acids Research*, vol. 26, no. 5, pp. 1144–1149, 1998.

[99] F. Hill, D. Loakes, and D. M. Brown, "Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 8, pp. 4258–4263, 1998.

[100] M. Zaccolo, D. M. Williams, D. M. Brown, and E. Gherardi, "An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues," *Journal of Molecular Biology*, vol. 255, no. 4, pp. 589–603, 1996.

[101] C. Y. Switzer, S. E. Moroney, and S. A. Benner, "Enzymatic recognition of the base pair between isocytidine and isoguanosine," *Biochemistry*, vol. 32, no. 39, pp. 10489–10496, 1993.

[102] J. A. Piccirilli, T. Krauch, S. E. Moroney, and S. A. Benner, "Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet," *Nature*, vol. 343, no. 6253, pp. 33–37, 1990.

[103] M. J. Moser, D. J. Marshall, J. K. Grenier et al., "Exploiting the enzymatic recognition of an unnatural base pair to develop a universal genetic analysis system," *Clinical Chemistry*, vol. 49, no. 3, pp. 407–414, 2003.

[104] K. P. Rice, J. C. Chaput, M. M. Cox, and C. Switzer, "RecA protein promotes strand exchange with DNA substrates containing isoguanine and 5-methyl isocytosine," *Biochemistry*, vol. 39, no. 33, pp. 10177–10188, 2000.

[105] S. C. Johnson, C. B. Sherrill, D. J. Marshall, M. J. Moser, and J. R. Prudent, "A third base pair for the polymerase chain reaction: inserting isoC and isoG," *Nucleic Acids Research*, vol. 32, no. 6, pp. 1937–1941, 2004.

[106] A. M. Maciejewska, K. D. Lichota, and J. T. Kuśmierek, "Neighbouring bases in template influence base-pairing of isoguanine," *The Biochemical Journal*, vol. 369, no. 3, pp. 611–618, 2003.

[107] S. Moran, R. X. Ren, S. Rumney, and E. T. Kool, "Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication," *Journal of the American Chemical Society*, vol. 119, no. 8, pp. 2056–2057, 1997.

[108] S. Xia, S. H. Eom, W. H. Konigsberg, and J. Wang, "Structural basis for differential insertion kinetics of dNMPs opposite a difluorotoluene nucleotide residue," *Biochemistry*, vol. 51, pp. 1476–1485, 2012.

[109] S. Xia, W. H. Konigsberg, and J. Wang, "Hydrogen-bonding capability of a templating difluorotoluene nucleotide residue in an RB69 DNA polymerase ternary complex," *Journal of the American Chemical Society*, vol. 133, no. 26, pp. 10003–10005, 2011.

[110] H. R. Lee, S. A. Helquist, E. T. Kool, and K. A. Johnson, "Base pair hydrogen bonds are essential for proofreading selectivity by the human mitochondrial DNA polymerase," *The Journal of Biological Chemistry*, vol. 283, no. 21, pp. 14411–14416, 2008.

[111] T. W. Kim, J. C. Delaney, J. M. Essigmann, and E. T. Kool, "Probing the active site tightness of DNA polymerase in subangstrom increments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 44, pp. 15803–15808, 2005.

[112] W. T. Wolfle, M. T. Washington, E. T. Kool et al., "Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase kappa," *Molecular and Cellular Biology*, vol. 25, no. 16, pp. 7137–7143, 2005.

[113] M. T. Washington, S. A. Helquist, E. T. Kool, L. Prakash, and S. Prakash, "Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase eta," *Molecular and Cellular Biology*, vol. 23, no. 14, pp. 5107–5112, 2003.

[114] A. Irimia, R. L. Eoff, P. S. Pallan, F. P. Guengerich, and M. Egli, "Structure and activity of Y-class DNA polymerase DPO4 from *Sulfolobus solfataricus* with templates containing the hydrophobic thymine analog 2,4-difluorotoluene," *The Journal of Biological Chemistry*, vol. 282, no. 50, pp. 36421–36433, 2007.

[115] A. M. Sismour and S. A. Benner, "The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system," *Nucleic Acids Research*, vol. 33, no. 17, pp. 5640–5646, 2005.

[116] A. M. Sismour, S. Lutz, J. H. Park et al., "PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from human immunodeficiency virus-1," *Nucleic Acids Research*, vol. 32, no. 2, pp. 728–735, 2004.

[117] J. Caton-Williams and Z. Huang, "Synthesis and DNA-polymerase incorporation of colored 4-selenothymidine triphosphate for polymerase recognition and DNA visualization," *Angewandte Chemie*, vol. 47, no. 9, pp. 1723–1725, 2008.

[118] T. W. Kim, L. G. Brieba, T. Ellenberger, and E. T. Kool, "Functional evidence for a small and rigid active site in a high fidelity DNA polymerase: probing $T_7$ DNA polymerase with variably sized base pairs," *The Journal of Biological Chemistry*, vol. 281, no. 4, pp. 2289–2295, 2006.

[119] H. O. Sintim and E. T. Kool, "Remarkable sensitivity to DNA base shape in the DNA polymerase active site," *Angewandte Chemie*, vol. 45, no. 12, pp. 1974–1979, 2006.

[120] M. Strerath, J. Cramer, T. Restle, and A. Marx, "Implications of active site constraints on varied DNA polymerase selectivity," *Journal of the American Chemical Society*, vol. 124, no. 38, pp. 11230–11231, 2002.

[121] M. Strerath, D. Summerer, and A. Marx, "Varied DNA polymerase-substrate interactions in the nucleotide binding pocket," *ChemBioChem*, vol. 3, pp. 578–580, 2002.

[122] J. Cramer, G. Rangam, A. Marx, and T. Restle, "Varied active-site constraints in the Klenow fragment of *E. coli* DNA polymerase I and the lesion-bypass Dbh DNA polymerase," *ChemBioChem*, vol. 9, no. 8, pp. 1243–1250, 2008.

[123] S. Mizukami, T. W. Kim, S. A. Helquist, and E. T. Kool, "Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase," *Biochemistry*, vol. 45, no. 9, pp. 2772–2778, 2006.

[124] L. M. Wilhelmsson, "Fluorescent nucleic acid base analogues," *Quarterly Reviews of Biophysics*, vol. 43, no. 2, pp. 159–183, 2010.

[125] L. J. Reha-Krantz, C. Hariharan, U. Subuddhi et al., "Structure of the 2-aminopurine-cytosine base pair formed in the polymerase active site of the RB69 Y567A-DNA polymerase," *Biochemistry*, vol. 50, pp. 10136–10149, 2011.

[126] L. B. Bloom, M. R. Otto, J. M. Beechem, and M. F. Goodman, "Influence of 5'-nearest neighbors on the insertion kinetics of the fluorescent nucleotide analog 2-aminopurine by Klenow fragment," *Biochemistry*, vol. 32, no. 41, pp. 11247–11258, 1993.

[127] B. W. Allan and N. O. Reich, "Targeted base stacking disruption by the EcoRI DNA methyltransferase," *Biochemistry*, vol. 35, no. 47, pp. 14757–14762, 1996.

[128] P. O. Lycksell, A. Graslund, F. Claesens, L. W. McLaughlin, U. Larsson, and R. Rigler, "Base pair opening dynamics of a 2-aminopurine substituted EcoRI restriction sequence and its unsubstituted counterpart in oligonucleotides," *Nucleic Acids Research*, vol. 15, pp. 9011–9025, 1987.

[129] T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graslund, and L. W. McLaughlin, "Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies," *Biochemistry*, vol. 28, no. 23, pp. 9095–9103, 1989.

[130] J. T. Stivers, K. W. Pankiewicz, and K. A. Watanabe, "Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase," *Biochemistry*, vol. 38, no. 3, pp. 952–963, 1999.

[131] D. C. Ward, E. Reich, and L. Stryer, "Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives," *The Journal of Biological Chemistry*, vol. 244, no. 5, pp. 1228–1237, 1969.

[132] R. A. Hochstrasser, T. E. Carver, L. C. Sowers, and D. P. Millar, "Melting of a DNA helix terminus within the active site of a DNA polymerase," *Biochemistry*, vol. 33, pp. 11971–11979, 1994.

[133] C. M. Joyce, O. Potapova, A. M. DeLucia, X. Huang, V. P. Basu, and N. D. Grindley, "Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity," *Biochemistry*, vol. 47, no. 23, pp. 6103–6116, 2008.

[134] H. R. Lee, M. Wang, and W. Konigsberg, "The reopening rate of the fingers domain is a determinant of base selectivity for RB69 DNA polymerase," *Biochemistry*, vol. 48, no. 10, pp. 2087–2098, 2009.

[135] A. M. DeLucia, N. D. Grindley, and C. M. Joyce, "Conformational changes during normal and error-prone incorporation of nucleotides by a Y-family DNA polymerase detected by 2-aminopurine fluorescence," *Biochemistry*, vol. 46, no. 38, pp. 10790–10803, 2007.

[136] U. Subuddhi, M. Hogg, and L. J. Reha-Krantz, "Use of 2-aminopurine fluorescence to study the role of the beta hairpin in the proofreading pathway catalyzed by the phage T4 and RB69 DNA polymerases," *Biochemistry*, vol. 47, no. 23, pp. 6130–6137, 2008.

[137] A. A. Marti, S. Jockusch, Z. Li, J. Ju, and N. J. Turro, "Molecular beacons with intrinsically fluorescent nucleotides," *Nucleic Acids Research*, vol. 34, no. 6, article e50, 2006.

[138] C. Dash, J. W. Rausch, and S. F. J. Le Grice, "Using pyrrolodeoxycytosine to probe RNA/DNA hybrids containing the human immunodeficiency virus type-1 3′ polypurine tract," *Nucleic Acids Research*, vol. 32, no. 4, pp. 1539–1547, 2004.

[139] X. Zhang and R. M. Wadkins, "DNA hairpins containing the cytidine analog pyrrolo-dC: structural, thermodynamic, and spectroscopic studies," *Biophysical Journal*, vol. 96, no. 5, pp. 1884–1891, 2009.

[140] K. Yang and R. J. Stanley, "The extent of DNA deformation in DNA photolyase-substrate complexes: a solution state fluorescence study," *Photochemistry and Photobiology*, vol. 84, no. 3, pp. 741–749, 2008.

[141] J. Petruska and M. F. Goodman, "Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity," *The Journal of Biological Chemistry*, vol. 260, no. 12, pp. 7533–7539, 1985.

[142] K. Y. Lin, R. J. Jones, and M. Matteucci, "Tricyclic 2′-deoxycytidine analogs: syntheses and incorporation into oligodeoxynucleotides which have enhanced binding to complementary RNA," *Journal of the American Chemical Society*, vol. 117, no. 13, pp. 3873–3874, 1995.

[143] L. M. Wilhelmsson, A. Holmen, P. Lincoln, P. E. Nielsen, and B. Norden, "A highly fluorescent DNA base analogue that forms Watson-Crick base pairs with guanine," *Journal of the American Chemical Society*, vol. 123, pp. 2434–2435, 2001.

[144] G. Stengel, J. P. Gill, P. Sandin et al., "Conformational dynamics of DNA polymerase probed with a novel fluorescent DNA base analogue," *Biochemistry*, vol. 46, no. 43, pp. 12289–12297, 2007.

[145] P. Sandin, P. Lincoln, T. Brown, and L. M. Wilhelmsson, "Synthesis and oligonucleotide incorporation of fluorescent cytosine analogue tC: a promising nucleic acid probe," *Nature Protocols*, vol. 2, no. 3, pp. 615–623, 2007.

[146] P. Sandin, G. Stengel, T. Ljungdahl, K. Borjesson, B. Macao, and L. M. Wilhelmsson, "Highly efficient incorporation of the fluorescent nucleotide analogs tC and tCo by Klenow fragment," *Nucleic Acids Research*, vol. 37, no. 12, pp. 3924–3933, 2009.

[147] J. M. Walsh, I. Bouamaied, T. Brown, L. M. Wilhelmsson, and P. J. Beuning, "Discrimination against the cytosine analog tC by *Escherichia coli* DNA polymerase IV DinB," *Journal of Molecular Biology*, vol. 409, no. 2, pp. 89–100, 2011.

[148] T. J. Lund, N. A. Cavanaugh, N. Joubert et al., "B family DNA polymerases asymmetrically recognize pyrimidines and purines," *Biochemistry*, vol. 50, pp. 7243–7250, 2011.

[149] P. Sandin, K. Borjesson, H. Li et al., "Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue," *Nucleic Acids Research*, vol. 36, pp. 157–167, 2008.

[150] G. Stengel, B. W. Purse, L. M. Wilhelmsson, M. Urban, and R. D. Kuchta, "Ambivalent incorporation of the fluorescent cytosine analogues tC and tCo by human DNA polymerase alpha and Klenow fragment," *Biochemistry*, vol. 48, no. 31, pp. 7547–7555, 2009.

[151] G. Stengel, M. Urban, B. W. Purse, and R. D. Kuchta, "High density labeling of polymerase chain reaction products with the fluorescent base analogue tCo," *Analytical Chemistry*, vol. 81, no. 21, pp. 9079–9085, 2009.

[152] E. T. Kool, J. C. Morales, and K. M. Guckian, "Mimicking the structure and function of DNA: insights into DNA stability and replication," *Angewandte Chemie*, vol. 39, pp. 990–1009, 2000.

[153] A. A. Henry and F. E. Romesberg, "Beyond A, C, G and T: augmenting nature's alphabet," *Current Opinion in Chemical Biology*, vol. 7, no. 6, pp. 727–733, 2003.

[154] K. Seo, J. Yin, P. Donthamsetti, S. Chandani, C. Lee, and E. Loechler, "Amino acid architecture that influences dNTP

insertion efficiency in Y-family DNA polymerase V of *E. coli*," *Journal of Molecular Biology*, vol. 392, no. 2, pp. 270–282, 2009.

[155] D. Loakes and P. Holliger, "Polymerase engineering: towards the encoded synthesis of unnatural biopolymers," *Chemical Communications*, no. 31, pp. 4619–4631, 2009.

[156] C. Bailly and M. J. Waring, "Use of DNA molecules substituted with unnatural nucleotides to probe specific drug-DNA interactions," *Methods in Enzymology*, vol. 340, pp. 485–502, 2001.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 156035, 8 pages
doi:10.1155/2012/156035

*Research Article*

# Features of "All LNA" Duplexes Showing a New Type of Nucleic Acid Geometry

**Charlotte Förster,[1,2] André Eichert,[1,3] Dominik Oberthür,[4] Christian Betzel,[4] Reinhard Geßner,[2] Andreas Nitsche,[5] and Jens P. Fürste[1]**

[1] *Institut für Chemie und Biochemie, Freie Universität Berlin, 14195 Berlin, Germany*
[2] *Chirurgische klinik II—Visceral, Transplantations-, Thorax- und Gefäßchirurgie, Universitätsklinik Leipzig, Liebigstraße 20, 04103 Leipzig, Germany*
[3] *Laboratory of Cell Biology and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA*
[4] *Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, c/o DESY, 22603 Hamburg, Germany*
[5] *Zentrum für Biologische Sicherheit 1, Robert Koch Institut, Nordufer 20, 13353 Berlin, Germany*

Correspondence should be addressed to Jens P. Fürste, fuerste@zedat.fu-berlin.de

Received 13 January 2012; Accepted 1 March 2012

Academic Editor: Subhendu Sekhar Bag



"Locked nucleic acids" (LNAs) belong to the backbone-modified nucleic acid family. The 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose nucleotides are used for single or multiple substitutions in RNA molecules and thereby introduce enhanced bio- and thermostability. This renders LNAs powerful tools for diagnostic and therapeutic applications. RNA molecules maintain the overall canonical A-type conformation upon substitution of single or multiple residues/nucleotides by LNA monomers. The structures of "all" LNA homoduplexes, however, exhibit significant differences in their overall geometry, in particular a decreased twist, roll and propeller twist. This results in a widening of the major groove, a decrease in helical winding, and an enlarged helical pitch. Therefore, the LNA duplex structure can no longer be described as a canonical A-type RNA geometry but can rather be brought into proximity to other backbone-modified nucleic acids, like glycol nucleic acids or peptide nucleic acids. LNA-modified nucleic acids provide thus structural and functional features that may be successfully exploited for future application in biotechnology and drug discovery.

## 1. Introduction

Modified nucleic acids have great potential for applications in oligonucleotide-based drug design. As natural RNA and DNA molecules are highly sensitive towards nuclease digestion and often possess low thermal stability, great effort has been made to design nucleic acid modifications that stabilize RNA or DNA while simultaneously maintaining the overall Watson-Crick base pairing ability. Modified nucleic acids are indispensable for future applications comprising diagnostic and clinical approaches like the use of aptamers or the siRNA technology.

Extensive and challenging experiments and investigations have been undertaken to develop nucleotide analogues that maintain the overall A-RNA-type conformation and N-type sugar puckering, as such modifications are likely to allow the substitution of RNA without large changes in functionality. Considerable effort has been made in the synthesis and characterization of 2′-*O*-methyl-RNAs [1], 2′-F-RNAs [2], phosphoramidate-RNAs [3], and the "locked" nucleic acid family [4]. By using locked nucleotide building blocks containing the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose (LNA) modification, a significant increase in thermostability can be observed in accordingly substituted

RNAs. For example, the melting temperature of modified RNA helices can be increased by +2 to +10°C per LNA monomer substitution.

To understand the stabilizing effects of LNA-substituted RNAs, numerous structural investigations have been performed during the past years to investigate their conformation in detail. These studies provided insights in the local geometric parameters of mix-mer LNA-RNA helices and of LNA-RNA heteroduplexes. The 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose LNA-RNA mix-mer duplexes maintain mainly the overall A-type nucleic acid conformation [5]. On the other hand, the 2′-*O*,4′-*C*-*α*-L-ribofuranose LNA modification is used in DNA substitution, as this modification preserves the overall B-type nucleic acid geometry of DNA [6]. Thus, there are two powerful nucleotide modifications with great potential in drug design, the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose nucleotides (LNA) for RNA substitution and the 2′-*O*,4′-*C*-methylene-*α*-L-ribofuranose nucleotides to modify DNA.

The structure of heteroduplexes, consisting of one fully modified LNA strand hybridized to either RNA or to DNA, revealed the following: the RNA conformation is maintained upon hybridizing a 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose LNA strand to RNA, whereas a mixed N- and S-type sugar puckering is induced by hybridizing a 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose LNA to DNA [7]. The B-type conformation is maintained by using a 2′-*O*,4′-*C*-methylene-*α*-L-ribofuranose LNA strand targeted to DNA [6]. It is generally accepted that the *2′-O*,4′-*C*-methylene-*β*-D-ribofuranose "locks" the LNA in the C3′-endo conformation. This approach is used to direct the geometry of the phosphate backbone in a manner to orient the duplex towards a more efficient base stacking.

Even though the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose LNA-RNA mix-mer duplexes maintain the overall A-type nucleic acid conformation, molecular dynamics simulations [8] and a crystal structure [9] of "all" LNA duplexes, consisting exclusively of 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose building blocks, yielded insights into a novel nucleic acid geometry. An "all LNA" duplex shows alterations in the local and overall helical parameters as compared to natural RNA and can rather be compared to other modified nucleic acids, like glycol nucleic acids (GNAs) [10], peptide nucleic acids (PNAs) [11], or homo-DNA [12]. An LNA duplex appears as a right-handed, antiparallel helix that maintains the canonical Watson-Crick base pairing and the 2′-exo conformation for all nucleotides. Nevertheless, the LNA duplex shows a considerable decrease in the helical twist, roll and propeller twist, which facilitates a widening of the major groove and a decrease of the minor groove dimensions. These alterations induce a large hollow cave in the middle of the duplex that is obvious in a projection perpendicular to the helical axis. Due to an enlarged helical rise and the unwinding of the helix, which results from the decrease in the twist angle parameters, the LNA duplex possesses an increased helical pitch. The unique nucleic acid geometry of "all LNA" helices apparently induces a more efficient and stable base stacking, which contributes to the higher thermostability of LNAs and LNA-modified nucleic acids.

Table 1: Data and refinement statistics of $tRNA^{Ser}$ microhelix and LNA helix [9, 15].

| | $tRNA^{Ser}$ microhelix | LNA helix |
|---|---|---|
| *Data acquisition* | | |
| Space group | *C*2 | *C*2 |
| Cell constants | | |
| *a*, *b*, *c* (Å) | 35, 79, 39.13, 31.37 | 77.91, 40.74, 30.06 |
| *α*, *β*, *γ* (°) | 90.00, 111.1, 90.00 | 90.00, 91.02, 90.00 |
| Resolution (Å) | 120–1.20 (1.22–1.20) | 80.00–1.90 (1.93–1.90) |
| $R_{merge}$ | 7.4 (15.4) | 7.3 (21.7) |
| $I/\sigma I$ | 18.7 (1.8) | 19.7 (1.0) |
| Completeness (%) | 99.2 (99.1) | 98.0 (97.2) |
| Redundancy | 7.1 (8.6) | 4.8 (3.8) |
| *Refinement* | | |
| No. of reflections | 12,806 | 7,382 |
| $R_{work}$ / $R_{free}$ | 19.0 (20.1) | 22.9 (28.8) |
| Atoms | | |
| Nucleic acid | 293 | 314 |
| Magnesium | 2 | 1 |
| Cobalt hexamine | — | 3 |
| Water oxygens | 97 (2 mol/au) | 44 |

Values in parentheses are given for the highest-resolution shell.

Interestingly, the structure of an RNA/LNA heteroduplex [13] is a geometric intermediate between the RNA and the "all LNA" conformation.

## 2. Material and Methods

*2.1. Crystallization of the LNA Helices.* The 7mer LNA helix was derived from the *E. coli* $tRNA^{Ser}$ isoacceptor with the data base "Compilation of tRNA sequences and sequences of tRNA genes" ID RS 1661 [14] and represents the sequence of the $tRNA^{Ser}$ aminoacyl stem microhelix that has been crystallized previously possessing the sequence 5′-(G-G-U-G-A-G-G-)-3′ and 5′-(C-C-U-C-A-C-C-)-3′ [15]. The LNA helix contained exclusively 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose building blocks. The base sequence of the RNA was maintained for further comparative studies, except for the U to T and the C to $m^5C$ exchange used in standard LNA synthesis. The chemically synthesized single strands with the sequences 5′-$(G\text{-}G\text{-}T\text{-}G\text{-}A\text{-}G\text{-}G)^L$-3′ and 5′-$(m^5C\text{-}m^5C\text{-}T\text{-}m^5C\text{-}A\text{-}m^5C\text{-}m^5C)^L$-3′ were purchased from IBA (Göttingen, Germany) with HPLC purification grade. Crystals were grown within 3-4 days using 40 mM sodium cacodylate, pH 5.5, 20 mM cobalt hexamine, 80 mM sodium chloride, 20 mM magnesium chloride, and 10% (v/v) MPD with equilibration against 1 mL 33–41% (v/v) MPD at 21°C using the hanging drop vapour diffusion technique [16].

*2.2. Diffraction Data Collection and Structure Determination and Refinement.* Data collection of the LNA crystals was performed at the ELETTRA synchrotron (Trieste, Italy)

Table 2: Selected overall helical parameters of the $tRNA^{Ser}$ microhelix structure (PDB ID: 3GVN) compared to the LNA-RNA hybrid (PDB ID: 1H0Q) and the LNA helix (PDB ID: 2X2Y). The two LNA molecules correspond to two LNA helices located in the asymmetric unit of the crystal structure.

| | Twist (°) | Rise (Å) | Slide (°) | Roll (°) | $\chi$-displacement (Å) | Propeller twist (°) |
|---|---|---|---|---|---|---|
| RNA $tRNA^{Ser}$ microhelix | 32.46 | 2.64 | −1.68 | 6.61 | −4.25 | −10.46 |
| LNA-RNA hybrid | 29.22 | 2.65 | −2.24 | 6.07 | −5.40 | −12.84 |
| LNA (molecule A) $tRNA^{Ser}$ microhelix | 25.97 | 2.81 | −2.49 | 4.08 | −6.60 | −6.65 |
| LNA (molecule B) $tRNA^{Ser}$ microhelix | 26.13 | 2.84 | −2.47 | 4.15 | −6.47 | −7.45 |

Table 3: Overall helical parameters for natural (RNA and DNA) and modified nucleic acids (LNA, GNA, and PNA).

| | Base pairs/helical turn | Twist (°) | Rise (Å) | P-P distance (Å) | Pitch (Å) |
|---|---|---|---|---|---|
| RNA | 11 | 32 | 2.6 | 6.0 | 30 |
| DNA | 10 | 36 | 3.4 | 7.0 | 34 |
| LNA | 14 | 26 | 2.8 | 5.6 | 39 |
| GNA | 16 | 22.9 | 3.8 | 5.4 | 60 |
| PNA | 18 | 19 | 3.2 | 5.4 | 58 |

beam line XRD-1 at a wavelength of 1.0 Å and a temperature of 100 K. The crystal diffracted up to 1.9 Å resolution [16]. The corresponding $tRNA^{Ser}$-microhelix was measured at the DESY synchrotron (Hamburg, Germany) at a wavelength of 0.8123 Å, 100 K temperature, and diffracted up to 1.2 Å [16]. All data were analyzed and processed using the programs from the HKL-2000 suite [17]. Molecular replacement calculations were performed using the program PHASER [18] within the CCP4i program suite [19]. The RNA structure was solved by molecular replacement using an artificially constructed RNA. The LNA structure was solved by using a model built from the previously solved $tRNA^{Ser}$ microhelix structure but exchanging the riboses by *2′-O,4′-C*-methylene-*β*-D-ribofuranose residues [9]. Standard LNA nucleotides were used for model building, which comprises the standard U to T and C to $m^5C$ substitutions in LNA as compared to RNA. Refinement calculations were done applying the program REFMAC [20], and electron density maps were calculated using FFT [21], as implemented in the CCP4i package [19]. Data and refinement statistics are shown in Table 1. The program X3DNA [22] was used to calculate the local and overall geometrical parameters. Structure representations and graphical analysis of helices were performed with the programs COOT [23] and PYMOL [24].

## 3. Results and Discussion

We analysed the crystal structure of a "locked" nucleic acid duplex [9], which contains exclusively *2′-O,4′-C*-methylene-*β*-D-ribofuranose nucleotides (Figure 1), in comparison to the structures of the naturally occurring RNA as well as to other backbone-modified nucleic acids like glycol nucleic acids (GNAs) or peptide nucleic acids (PNAs).

The LNA helix structure reveals a nucleic acid duplex geometry that significantly differs from the canonical A-type RNA structure (Figure 2, Tables 2 and 3). The structure of the LNA duplex appears as a stretched helical ladder with altered local and overall geometric parameters. The observed geometry can be rather compared to that of glycol nucleic acids (GNAs) [25], peptide nucleic acids (PNAs) [26, 27], or homo-DNAs [12]. We detected a notable decrease in several local and overall helical parameters in the LNA helix, like the twist, roll and propeller twist, as compared to a corresponding RNA molecule (Table 2). This results in a widening of the major groove, a decrease in helical winding and an increased helical pitch. The major groove dimensions in the LNA duplex showed values of around 24-25 Å in diameter, as compared to 16 Å observed for the canonical A-RNA duplex. Concomitantly, the minor groove of LNA duplexes is narrower (about 15 Å) than that of standard RNA helices (19 Å). On the other hand, the slide and rise values are slightly increased in the LNA helices. Moreover, the shift of the base pairs in the LNA duplex results in an empty tunnel running through the center of the helix.

In the LNA helix, the low twist angle of 26° and the large pitch of 14 base pairs per turn lead to an unwinding of the duplex, as compared to RNA, which possesses a twist of 32° and a pitch of 11 base pairs per turn. The helical rise of LNA falls into a range of 2.8–3.0 Å, whereas the helical rise in RNA is 2.6 Å. Due to the increased rise and

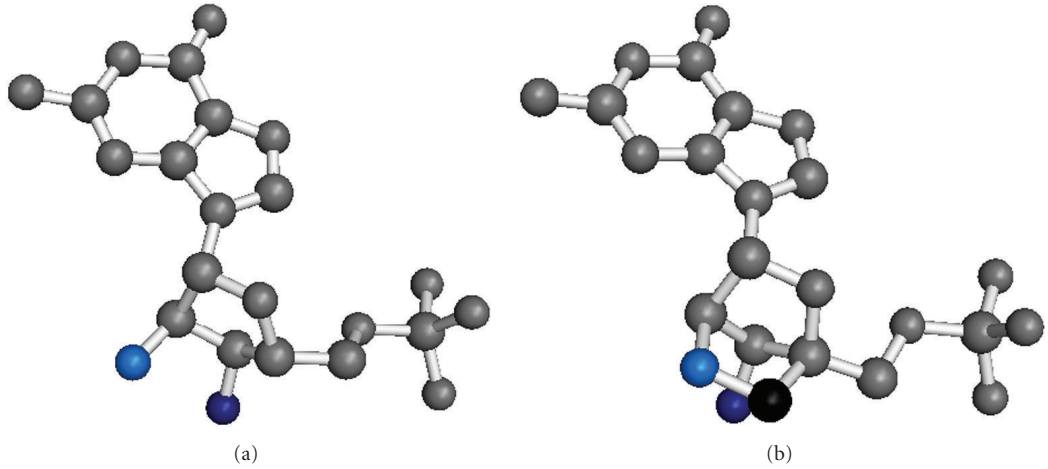



Figure 1: Guanosine monophosphate shown as RNA nucleotide (a) and as LNA nucleotide with the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose modification (b). Oxygen atoms are coloured in light blue (2′-oxygens) and dark blue (3′-oxygens), respectively. The additional carbon atom from the methylene group in the LNA is shown in black.

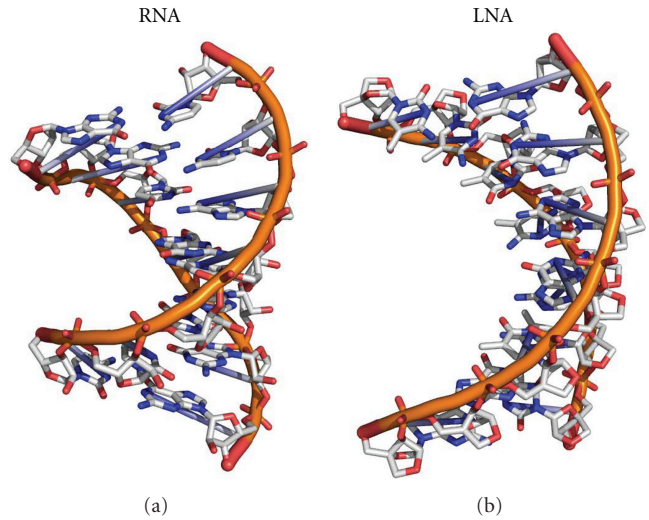



Figure 2: Crystal structure of the RNA duplex r[GGUGAGG]·r[CCUCACC] (PDB ID: 3GVN) as compared to the corresponding LNA helix (PDB ID: 2X2Y).

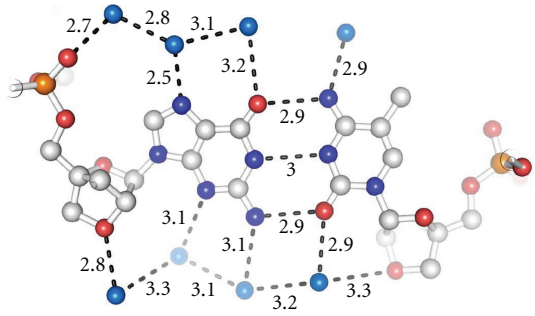



Figure 3: Hydration pattern within the LNA duplex (PDB ID: 2X2): a region of the LNA helix showing the base pair (G4-m$^5$C69)$^L$. The hydration pattern resembles that known for RNA, as the bridged 2′-oxygen atom in LNA acts as a hydrogen bond acceptor similar to the 2′-oxygen in the hydroxyl group of RNA.

the unwinding of the helix, the LNA possessed an enlarged helical pitch of 39 Å, as compared to 29 Å in RNA helices. The backbone torsion angles resembled the sc$^-$, ap$^+$, sc$^+$, sc$^+$ ap, sc$^-$, and ap$^+$ conformation for the $\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, $\zeta$, and $\chi$ angles with the sugar puckering being in the 2′-exo conformation. The phosphate-phosphate distances are in the region of 5.6 Å as compared to 6.0 Å for RNA. It is conceivable that the altered helical parameters in LNA duplexes provide an enhancement in nucleotide stacking, leading towards stronger Π-Π interactions of the base pairs.

It is well accepted that the extensive hydration of the RNA minor groove plays an important role in the structure/function relationship [28]. As the specific hydration pattern of RNA is governed by the 2′-hydroxyl group, it has been questioned whether the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose in LNA allows a comparable hydration as

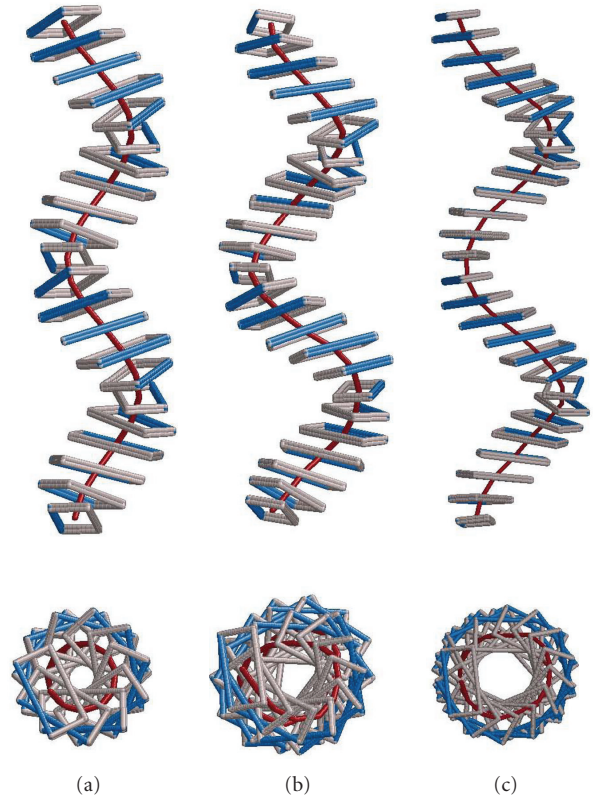



Figure 4: Structure of an idealized RNA helix (a) compared to the RNA/LNA hybrid helix derived from the structure with PDB code 1H0Q (b) and compared to an "all LNA" helix derived from the structure with PDB code: 2X2Q (c) as Calladine-Drew plot. To visualize the overall geometry, the helices were extended to 22 base pairs for RNA, to 25 base pairs for the RNA/LNA, and to 28 base pairs for LNA, respectively [9].

described for RNAs. Therefore, we focussed our investigations on analyzing the arrangement of the solvent molecules surrounding the LNA duplex. We observed that the distribution of water molecules in the LNA minor groove follows the general pattern known for RNA hydration, as the bridged 2′ oxygen atoms in the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose moieties serve as hydrogen bond acceptors similar to the 2′-hydroxyl residues in RNA. An example for the LNA hydration is shown in Figure 3.

Interestingly, the structure of an RNA/LNA hybrid helix represents a geometric intermediate between RNA and "all LNA" helices. In Figure 4, we present the structure of idealized RNA as compared to the RNA/LNA hybrid (PDB ID: 1H0Q) and to the "all LNA" duplex (PDB ID: 2X2Q) as Calladine-Drew plot. To better visualize the overall geometry in this figure, the helices were extended to 22 base pairs for RNA, to 25 base pairs for RNA/LNA, and to 28 base pairs for LNA showing two full helical turns each [9]. For an overall comparison of A-RNA and B-DNA helices to the conformation of different backbone-modified nucleic acid types, like GNA and PNA (Figure 5), we displayed the selected nucleic acid helices with a total length of 46 base pairs (Figure 6). We illustrate the natural DNA and RNA and the synthetic GNA (PDB ID: 2JJA) and PNA (PDB ID: 1PUP) as compared to the LNA (PDB ID: 2X2Q) duplex structure. The standard A- and B-type nucleic acid conformations are paraphrased by the RNA and DNA helices. The GNA shows the structure of a helical ribbon with only one large minor groove and completely lacks the major groove, which is instead a convex surface [25]. The PNA resembles the helix with a wide and deep major groove concomitant with a narrow and shallow minor groove [26, 27]. The weakly twisted right-handed homo-DNA structure has been described to explain the inability of allo-, altro-, and glucanosyl-nucleotides to form stable base pairing systems (picture not shown) [12]. In the middle of Figure 6 we present the extended structure of the LNA duplex, which represents the unusual geometry, which can rather be brought into vicinity of GNA, PNA, and homo-DNA than to the natural nucleic acid duplexes DNA and RNA. The LNA helix, however, possesses a pitch of 39 Å with 14 base pairs per helical turn and an average rise of 2.8 Å. DNA and RNA show an average pitch of 30 Å and 34 Å, respectively, as compared to mean values of 60 Å, 58 Å, and 53 Å found for GNA, PNA, and homo-DNA. (Figure 6, Table 3, no data shown for homo-DNA). Conclusively, regarding the geometry of the LNA duplex in total, this helix resembles a more natural nucleic acid, when compared to the other backbone-modified duplexes.

As has been reviewed [4], an increase of the melting temperature between +2° to +10°C can be observed per LNA building block added in strands hybridized to RNA.

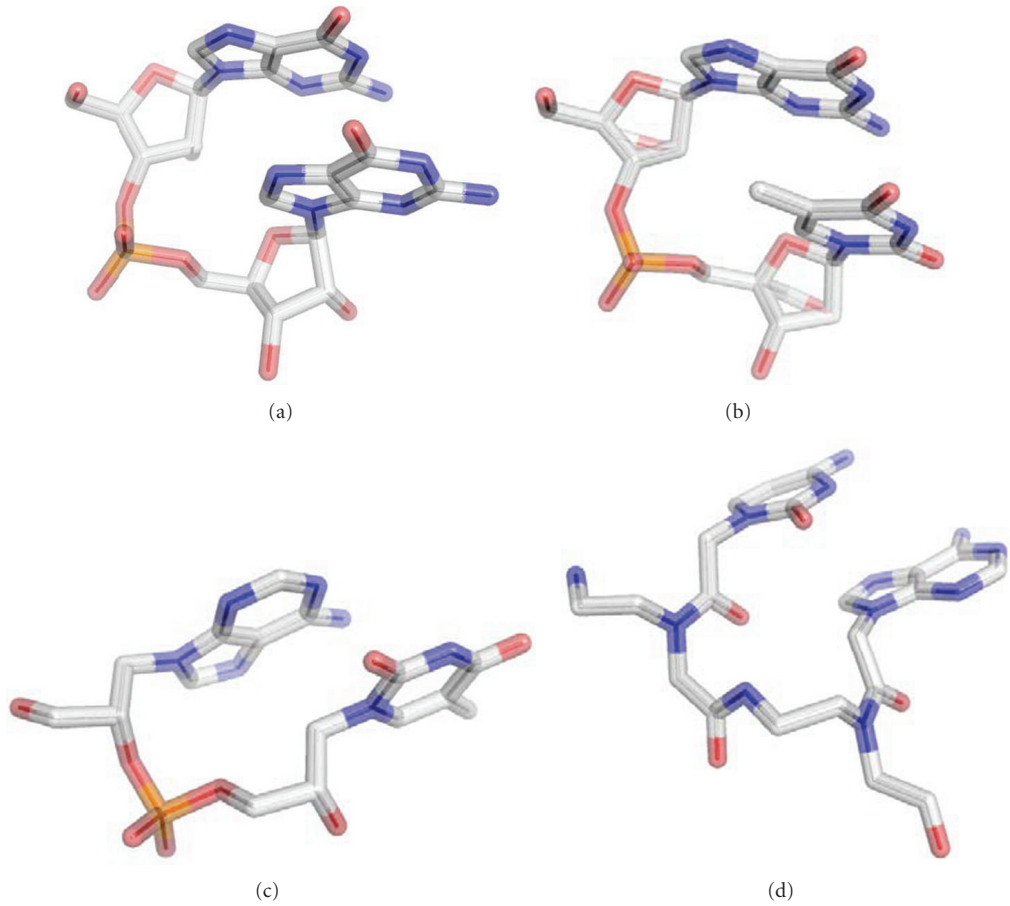



Figure 5: Dinucleotide conformations as observed in RNA (a), LNA with the 2′-*O*,4′-*C*-methylene-*β*-D-ribofuranose modification (b), GNA (c), and PNA (d).

The short 7 bp LNA duplex, derived from the $tRNA^{Ser}$ microhelix for this study, exhibits a melting temperature of above 90°C, whereas the corresponding RNA has a $T_m$ value of 45.0°C [9]. We have previously investigated the melting temperature of another LNA 7mer helix in comparison to its natural RNA counterpart [29]. Similarly, the LNA duplex possesses a $T_m$ value of 84.3°C, whereas its corresponding RNA helix melts at 22.4°C. The drastic shift in thermostability in both LNAs as compared to the RNAs corresponds to an average of 4.5°C per nucleotide building block, which is consistent with the reviewed observations [4]. Thus, the thermostability data and the structure properties of LNAs provide new perspectives for future nucleic acid drug applications, which is an encouraging outlook.

Considering that the increase in $T_m$ values by substituting natural nucleic acids with single or multiple nucleotides by LNA residues seems to be a summative property, the challenge of using LNAs as tools in nucleic acid stabilization becomes obvious. Nearly any natural nucleic acid can be modified by introducing single and multiple LNA building blocks or even complete LNA duplexes, thereby stepwise increasing the thermostability depending on the number of introduced LNA residues. Depending on the particular requirements, any nucleic acid can thus be stabilized at will with little or no loss of function. In this respect, LNA substitution may serve as a reliable method to stabilize nucleic acids, in particular aptamers, for clinical applications.

An upcoming challenge is to stabilize aptamer stem regions by introducing LNA portions without affecting the loop regions that are usually essential for target binding and specificity. Several reports in the literature highlight a forthcoming application of LNA-substituted aptamers with retained or even improved ligand-binding capacity. An exciting example is the use of LNA modifications within hammerhead ribozymes that improve the overall cleaving capacity [30]. In addition, LNA modifications have successfully been introduced into antisense oligonucleotides and DNAzymes that were targeted to functionally selected binding sites and inhibited HIV-1 expression [31]. A third example describes a G-quadruplex thrombin aptamer, which retained the biological activity to a varying extent depending on the nucleotide positions that were LNA modified [32]. These selected reports are a snapshot of numerous studies demonstrating the great potential of LNA substitutions in functional nucleic acids and possible therapeutic applications. The crystal structure of the "all locked" nucleic acid helix contributes to the understanding of the structure/function relationship and the high thermostability of

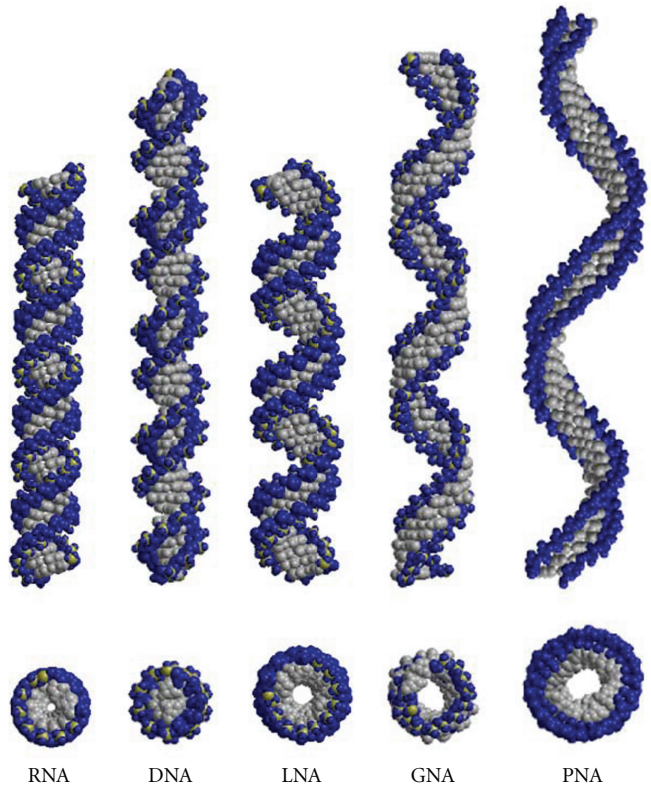



Figure 6: Overall helical structures of natural nucleic acids (RNA and DNA) and of synthetic, modified nucleic acids (LNA, GNA, and PNA). All helices were extended to a total of 46 base pairs. The helices were constructed by using the following structures: DNA (idealized), RNA (idealized), LNA (PDB ID: 2X2Y), GNA (PDB ID: 2JJA), and PNA (PDB ID: 1PUP). Phosphate oxygen atoms are shown in blue, phosphates in yellow, and all other atoms are presented in grey. Top picture shows the side view of the duplexes and the bottom picture presents a projection along the helical axis.

these molecules. In summary LNAs possess an encouraging potential for the development of new stabilized nucleic acids and will promote future applications in diagnostics, drug discovery, and clinical therapy [4].

## Acknowledgments

This work was supported by the BiGRUDI network of the Robert Koch Institute (Berlin) financed by the Bundesministerium für Bildung und Forschung. A. Eichert was member of the Dahlem Research School of the Free University Berlin and funded by the Friedrich-Ebert-Stiftung, Germany, and is now supported by a DFG fellowship. We gratefully acknowledge the ELETTRA synchrotron, Trieste, Italy, and the DESY synchrotron, Hamburg, Germany for providing beam time and Ursula Erikli for copyediting the paper.

## References

[1] M. Manoharan, "2'-Carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation," *Biochimica et Biophysica Acta*, vol. 1489, no. 1, pp. 117–130, 1999.

[2] A. M. Kawasaki, M. D. Casper, S. M. Freier et al., "Uniformly modified 2′-deoxy-2′-fluoro phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for RNA targets," *Journal of Medicinal Chemistry*, vol. 36, no. 7, pp. 831–841, 1993.

[3] S. M. Gryaznov, "Oligonucleotide N3' → P5'-phosphoramidates as potential therapuetic agents," *Biochimica et Biophysica Acta*, vol. 1489, no. 1, pp. 131–140, 1999.

[4] M. Petersen and J. Wengel, "LNA: a versatile tool for therapeutics and genomics," *Trends in Biotechnology*, vol. 21, no. 2, pp. 74–81, 2003.

[5] M. Petersen, K. Bondensgaard, J. Wengel, and J. Peter Jacobsen, "Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA: RNA hybrids," *Journal of the American Chemical Society*, vol. 124, no. 21, pp. 5974–5982, 2002.

[6] K. M. Nielsen, M. Petersen, A. E. Håkansson, J. Wengel, and J. P. Jacobsen, "alpha-L-LNA (alpha-L-ribo configured locked nucleic acid) recognition of DNA: an NMR spectroscopic study.," *Chemistry*, vol. 8, no. 13, pp. 3001–3009, 2002.

[7] K. E. Nielsen, J. Rasmussen, R. Kumar, J. Wengel, J. P. Jacobsen, and M. Petersen, "NMR studies of fully modified locked nucleic acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid," *Bioconjugate Chemistry*, vol. 15, no. 3, pp. 449–457, 2004.

[8] V. Pande and L. Nilsson, "Insights into structure, dynamics and hydration of locked nucleic acid (LNA) strand-based duplexes from molecular dynamics simulations," *Nucleic Acids Research*, vol. 36, no. 5, pp. 1508–1516, 2008.

[9] A. Eichert, K. Behling, C. Betzel, V. A. Erdmann, J. P. Fürste, and C. Förster, "The crystal structure of an "All Locked" nucleic acid duplex," *Nucleic Acids Research*, vol. 38, no. 19, pp. 6729–6736, 2010.

[10] M. K. Schlegel, L. O. Essen, and E. Meggers, "Duplex structure of a minimal nucleic acid," *Journal of the American Chemical Society*, vol. 130, no. 26, pp. 8158–8159, 2008.

[11] H. Rasmussen, S. J. Kastrup, J. N. Nielsen, J. M. Nielsen, and P. E. Nielsen, "Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 Å resolution," *Nature Structural Biology*, vol. 4, no. 2, pp. 98–101, 1997.

[12] M. Egli, P. S. Pallan, R. Pattanayek et al., "Crystal structure of homo-DNA and nature's choice of pentose over hexose in the genetic system," *Journal of the American Chemical Society*, vol. 128, no. 33, pp. 10847–10856, 2006.

[13] M. Petersen, K. Bondensgaard, J. Wengel, and J. Peter Jacobsen, "Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids," *Journal of the American Chemical Society*, vol. 124, no. 21, pp. 5974–5982, 2002.

[14] M. Sprinzl and K. S. Vassilenko, "Compilation of tRNA sequences and sequences of tRNA genes," *Nucleic Acids Research*, vol. 33, pp. D139–D140, 2005.

[15] A. Eichert, J. P. Fürste, A. Schreiber et al., "The 1.2 Å crystal structure of an *E. coli* tRNASer acceptor stem microhelix reveals two magnesium binding sites," *Biochemical and Biophysical Research Communications*, vol. 386, no. 2, pp. 368–373, 2009.

[16] K. Behling, A. Eichert, J. P. Fürste, C. Betzel, V. A. Erdmann, and C. Förster, "Crystallization and X-ray diffraction analysis of an all-locked nucleic acid duplex derived from a tRNASer microhelix," *Acta Crystallographica Section F*, vol. 65, no. 8, pp. 809–812, 2009.

[17] Z. Otwinowski and W. Minor, "Processing of X-ray diffraction data collected in oscillation mode," *Methods in Enzymology*, vol. 276, pp. 307–326, 1997.

[18] A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read, "Likelihood-enhanced fast translation functions," *Acta Crystallographica Section D*, vol. 61, no. 4, pp. 458–464, 2005.

[19] Collaborative Computational Project, Number 4, "The CCP4 suite: programs for protein crystallography," *Acta Crystallographica*, vol. D50, pp. 760–763, 1994.

[20] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method," *Acta Crystallographica Section D*, vol. 53, no. 3, pp. 240–255, 1997.

[21] R. J. Read and A. J. Schierbeek, "A phased translation function," *Journal of Applied Crystallography*, vol. 21, pp. 490–495, 1988.

[22] X. J. Lu and W. K. Olson, "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures," *Nucleic Acids Research*, vol. 31, no. 17, pp. 5108–5121, 2003.

[23] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *Acta Crystallographica Section D*, vol. 60, no. 12, pp. 2126–2132, 2004.

[24] W. L. DeLano, *The PyMOL Molecular Graphics System*, De Lano Scientific, Palo Alto, Calif, USA, 2002.

[25] M. K. Schlegel, L. O. Essen, and E. Meggers, "Duplex structure of a minimal nucleic acid," *Journal of the American Chemical Society*, vol. 130, no. 26, pp. 8158–8159, 2008.

[26] W. He, E. Hatcher, A. Balaeff et al., "Solution structure of a peptide nucleic acid duplex from NMR data: features and limitations," *Journal of the American Chemical Society*, vol. 130, no. 40, pp. 13264–13273, 2008.

[27] H. Rasmussen, S. J. Kastrup, J. N. Nielsen, J. M. Nielsen, and P. E. Nielsen, "Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 Å resolution," *Nature Structural Biology*, vol. 4, no. 2, pp. 98–101, 1997.

[28] P. Auffinger and E. Westhof, "Hydration of RNA base pairs," *Journal of Biomolecular Structure and Dynamics*, vol. 16, pp. 693–707, 1998.

[29] C. Förster, D. Oberthuer, J. Gao et al., "Crystallization and preliminary X-ray diffraction data of an LNA 7-mer duplex derived from a ricin aptamer," *Acta Crystallographica Section F*, vol. 65, no. 9, pp. 881–885, 2009.

[30] J. K. Christiansen, S. Lobedanz, K. Arar, J. Wengel, and B. Vester, "LNA nucleotides improve cleavage efficiency of singular and binary hammerhead ribozymes," *Bioorganic and Medicinal Chemistry*, vol. 15, no. 18, pp. 6135–6143, 2007.

[31] M. R. Jakobsen, J. Haasnoot, J. Wengel, B. Berkhout, and J. Kjems, "Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites," *Retrovirology*, vol. 4, article 29, 2007.

[32] L. Bonifacio, F. C. Church, and M. B. Jarstfer, "Effect of locked-nucleic acid on a biologically active G-quadruplex. A structure-activity relationship of the thrombin aptamer," *International Journal of Molecular Sciences*, vol. 9, no. 3, pp. 422–433, 2008.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 196845, 11 pages
doi:10.1155/2012/196845

*Research Article*

# Metal Ion Chelates as Surrogates of Nucleobases for the Recognition of Nucleic Acid Sequences: The $Pd^{2+}$ Complex of 2,6-Bis(3,5-dimethylpyrazol-1-yl)purine Riboside

**Sharmin Taherpour and Tuomas Lönnberg**

*Department of Chemistry, University of Turku, Vatselankatu 2, 20014 Turku, Finland*

Correspondence should be addressed to Tuomas Lönnberg, tuanlo@utu.fi

Received 2 January 2012; Accepted 22 February 2012

Academic Editor: David M. Chenoweth



A 2,6-bis(3,5-dimethylpyrazol-1-yl)purine ribonucleoside has been prepared and incorporated as a conventionally protected phosphoramidite into a 9-mer 2′-*O*-methyl oligoribonucleotide. According to 1H NMR spectroscopic studies, this nucleoside forms with $Pd^{2+}$ and uridine a ternary complex that is stable at a micromolar concentration range. CD spectroscopic studies on oligonucleotide hybridization, in turn, suggest that the $Pd^{2+}$ chelate of this artificial nucleoside, when incorporated in a 2′-*O*-methyl-RNA oligomer, is able to recognize thymine within an otherwise complementary DNA strand. The duplex containing thymidine opposite to the artificial nucleoside turned out to be somewhat more resistant to heating than its counterpart containing 2′-deoxycytidine in place of thymidine, but only in the presence of $Pd^{2+}$. According to UV-melting measurements, replacement of 2′-*O*-methyladenosine with the artificial nucleoside markedly enhances hybridization with a DNA target, irrespective of the identity of the opposite base and the presence of $Pd^{2+}$. With the thymidine containing DNA target, the $T_m$ value is 2–4°C higher than with targets containing any other nucleoside opposite to the artificial nucleoside, but the dependence on $Pd^{2+}$ is much less clear than in the case of the CD studies.

## 1. Introduction

In nature, recognition of nucleic acid sequences is based on Watson-Crick base pairing, while vertical stacking of base pairs accounts for most of the duplex stability. At room temperature, 10 base pairs are typically required for a stable duplex, meaning that targeting sequences shorter than this by the conventional strategy, that is, using complementary oligonucleotides, is problematic. Sequences within a double-stranded region are also difficult to target because thermodynamics favor formation of the longer duplex. The first problem has partly been addressed by using small molecules exhibiting high affinity towards certain specific short sequences [1]. No clear-cut relationship exists between the target sequence and the structure of the small molecule, however, making the design and preparation of such molecules a demanding task. PNA oligonucleotides offer a solution to the second problem, as PNA/DNA heteroduplexes are more stable than native double-stranded DNA, but even in the best cases a substantial excess of the PNA is still needed for efficient invasion [2, 3].

Oligonucleotides composed of unnatural monomers with an enhanced affinity towards their complement nucleobases could potentially form stable duplexes even with short target sequences as well as invade double-stranded DNA even when present in only stoichiometric amounts. One way to achieve the desired high affinity is to exploit the coordination of a ring nitrogen (N1 of purines or N3 of pyrimidines) of the natural nucleobase to a soft metal ion, such as $Pd^{2+}$, carried by an artificial nucleobase [4]. Discrimination between the four natural nucleobases, in turn, could be achieved through a combination of additional destabilizing (steric) and stabilizing (hydrogen bonding) interactions. A number of studies on metal-ion-mediated base pairs have been published [5, 6], but the focus has generally been on

expanding the genetic code by introducing a completely new artificial base pair [7, 8] or impregnating DNA with metal ions for nanotechnological applications [9–11], rather than developing high-affinity complements for the natural nucleobases. In the former, both of the ligands of the metal-ion-mediated base pair may be freely designed for maximum performance (in many cases, the same ligand is used in both strands), whereas in the latter the sole artificial nucleobase must be able to not only coordinate the metal ion but also accommodate the steric and hydrogen bonding requirements of the natural nucleobase, presenting a unique challenge.

In the present study, the $Pd^{2+}$complex of 2,6-bis-(3,5-dimethylpyrazol-1-yl)purine ribonucleoside (**1**) is used as an artificial nucleoside. Nucleoside **1**, with no potential for stabilizing hydrogen bond interactions, may be considered as a reference structure for future studies. On a monomeric level, the binding affinity and selectivity of this artificial nucleoside were studied by NMR titrations. For a more realistic model, nucleoside **1** was additionally converted to a phosphoramidite building block (**2**) and incorporated into a 2′-*O*-methyl-RNA oligonucleotide. The effect of $Pd^{2+}$ on the secondary structure and thermal stability of duplexes between this modified oligonucleotide and complementary 2′-*O*-methyl-RNA and DNA oligonucleotides was investigated by melting temperature and CD spectrometric studies. For comparison, the same experiments were also performed on a 2′-*O*-methyl-RNA oligonucleotide having adenosine in the place of the artificial nucleoside.

## 2. Results

*2.1. Synthesis of 2,6-bis(3,5-dimethylpyrazol-1-yl)purine Ribonucleoside (1) and Its Conversion to a Phosphoramidite Building Block (2).* Synthesis strategy for the phosphoramidite building block **2** is presented in Scheme 1. First, the 2,6-dihydrazinopurine riboside **3** was prepared by treating 2′,3′,5′-tri-*O*-acetyl-6-chloro-2-iodopurine riboside (**4**) with hydrazine hydrate at room temperature [12]. After that, the hydrazino groups were converted to 3,5-dimethylpyrazol-1-yl groups with 2,4-pentanedione [13] and the 5′-OH was protected as a 4,4′-dimethoxytrityl ether. Finally, the 2′-OH was silylated with TBDMSCl and the 3′-OH phosphitylated by the conventional methods.

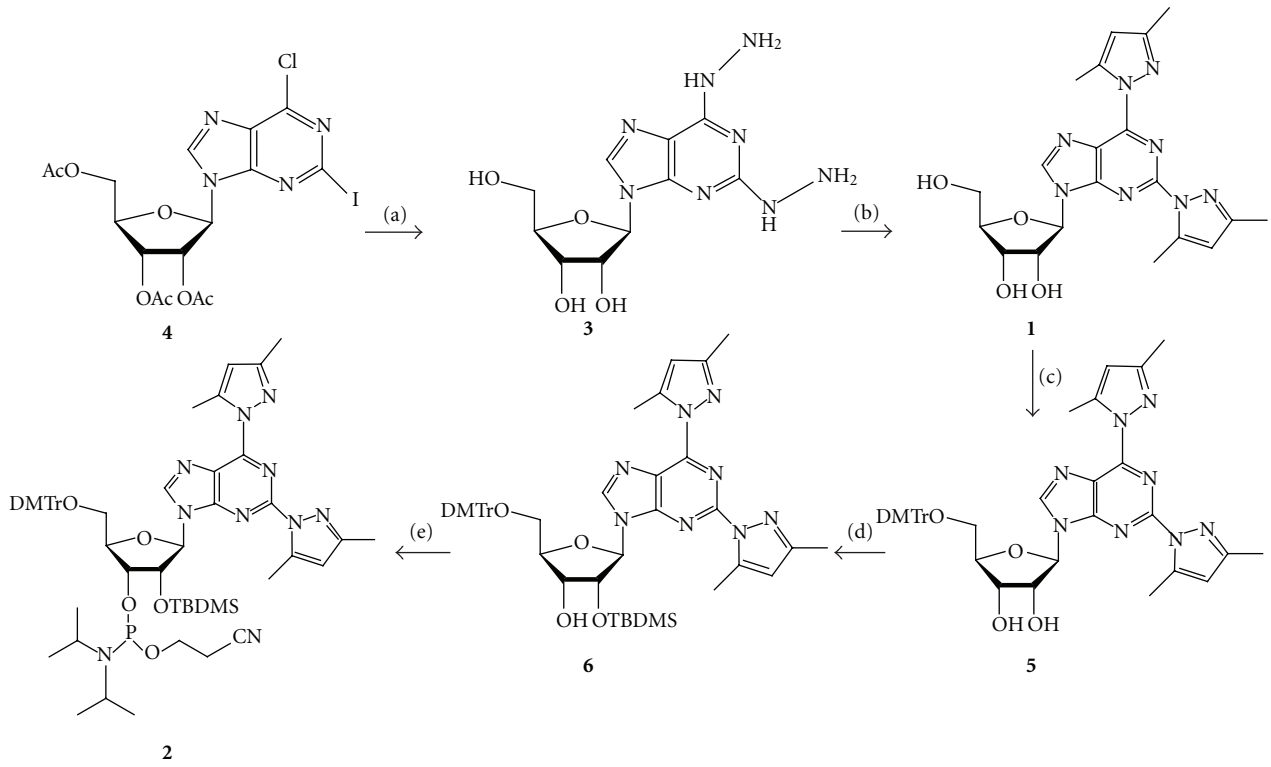
*2.2. Synthesis of Oligonucleotides.* 2′-*O*-Methyl-RNA oligonucleotides **7A, 8U**, **8A**, **8G**, **8C**, **9U**, **9A**, **9G,** and **9C**, as well as the DNA oligonucleotides **10T**, **10A**, **10C,** and **10G**, were synthesized from commercial phosphoramidite building blocks by conventional phosphoramidite strategy on an automated synthesizer. The modified phosphoramidite monomer (**2**) was incorporated in the middle of a 9-mer 2′-*O*-Me-RNA sequence (**7X**) manually using elongated coupling time (60 min). The coupling yield for building block **2** was 36% and the subsequent couplings proceeded with normal (approximately 99%) efficiency. The crude oligonucleotides were purified by reversed-phase high-performance liquid chromatography (RP-HPLC) and characterized by electrospray ionization mass spectrometry (ESI-MS). The concentrations of the purified oligonucleotides were determined UV-spectrophotometrically using molar absorption coefficients calculated by an implementation of the nearest-neighbors method. The sequences of the oligonucleotides synthesized, as well as their RP-HPLC retention times, wavelengths of UV absorption maxima and observed and calculated molecular weights are summarized in Table 1.

*2.3. NMR Spectrometric Titrations.* The formation and stability of the complexes between the artificial nucleoside **1** and uridine in the presence and absence of $Pd^{2+}$ were studied by $^1$H NMR spectrometric titration. The NMR spectra were recorded in a 120 mM deuterated phosphate buffer at pH 7.3. The initial concentration of the nucleoside components of the putative ternary complex, *namely,* **1** and uridine, was 3.6 mmol $L^{-1}$, while a deficit of $Pd^{2+}$ (0.7 eq., 2.5 mmol $L^{-1}$ as $K_2PdCl_4$) was employed to ensure exclusive formation of the desired ternary complex. The titration was carried out by diluting the sample with buffer until the uridine concentration was 0.061 mmol $L^{-1}$.

Over the entire concentration range studied, two distinct sets of signals were observed for uridine: one referring to the free nucleoside and the other one (presumably) to the ternary complex (Figures 1(a)–1(d)). The two sets of signals were particularly well resolved in the case of H5 and H1′ resonance of uridine and the integrals of these signals were used to determine the mole fraction of the complex at each concentration (Figure 2). To exclude the possibility that the observed changes in the chemical shifts of the H5 and H1′ of uridine were caused by $Pd^{2+}$ alone, a corresponding experiment with uridine and $Pd^{2+}$ but without the artificial nucleoside **1** was carried out (Figures 1(f)–1(i)). In this case, an upfield change in the chemical shifts of the H5 and H1′ of uridine was observed, in striking contrast with the downfield shift associated with formation of the putative ternary complex. In other words, the H5 and H1′ signals assigned to uridine in its ternary complex with $Pd^{2+}$ and the artificial nucleoside **1** were not observed in the absence of the latter. No effort was made to characterize the complexes between uridine and $Pd^{2+}$ but the NMR spectra suggest that a number of different relatively weak complexes prevail, again in sharp contrast with the single strong complex formed in the presence of the artificial nucleoside **1**.

Upon introduction of $Pd^{2+}$, the signals for the hydrogens on the pyrazole rings of **1** shift downfield by approximately 0.4 ppm. The signals of two of the methyl substituents of the pyrazole rings exhibit an even greater downfield shift (approximately 0.6 and 0.8 ppm, resp.), whereas the other two are shifted slightly upfield. Adding uridine to the system hardly affects the chemical shifts of these protons (data not shown).

As expected, dilution of the sample was accompanied by decrease of the proportion of signals arising from the ternary complex (Figures 1(a)–1(d)). However, even at the lowest concentration employed, that is, 61 $\mu$mol $L^{-1}$, the mole fraction of the complex was still more than 80% of the saturation level observed at high concentrations. In the absence of $Pd^{2+}$, no complex was formed even though a high concentration (13.84 mmol $L^{-1}$) of the nucleosides was




Scheme 1: Preparation of the phosphoramidite building block **2**. Reagents and conditions: (a) $H_2NNH_2 \cdot H_2O$; (b) acetylacetone, THF, TFA; (c) DMTrCl, pyridine; (d) TBDMSCl, imidazole, DMF; (e) 2-cyanoethyl-*N*, *N*-diisopropylchlorophosphoramidite, triethylamine, $CH_2Cl_2$.

Table 1: The sequences, RP-HPLC retention times, wavelengths of UV absorption maxima, and observed and calculated molecular weights of the oligonucleotides prepared.

| Sequence | | $t_R$ /minute[a] | $\lambda$/nm | *m/z* obsd. | *m/z* calcd. |
|---|---|---|---|---|---|
| **7X** | 5′-GCGC**X**CCGG-3′[b] | 16.0 | 267.8 | 3152.8 | 3152.7 |
| **7A** | 5′-GCGC**A**CCGG-3′ | 16.1 | 257.6 | 2993.6 | 2993.8 |
| **8U** | 3′-CGCG**U**GGCC-5′ | 17.2 | 258.4 | 2970.6 | 2970.8 |
| **8A** | 3′-CGCG**A**GGCC-5′ | 17.5 | 257.6 | 2993.6 | 2993.8 |
| **8G** | 3′-CGCG**G**GGCC-5′ | 14.1 | 257.2 | 3009.6 | 3009.8 |
| **8C** | 3′-CGCG**C**GGCC-5′ | 16.6 | 258.8 | 2969.6 | 2969.8 |
| **9U** | 3′-CGC**AUA**GCC-5′ | 11.8 | 260.2 | 2938.6 | 2938.8 |
| **9A** | 3′-CGC**AAA**GCC-5′ | 12.3 | 259.1 | 2961.6 | 2961.8 |
| **9G** | 3′-CGC**AGA**GCC-5′ | 11.9 | 261.0 | 2977.6 | 2977.8 |
| **9C** | 3′-CGC**ACA**GCC-5′ | 12.4 | 267.0 | 2937.6 | 2937.8 |
| **10T** | 3′-d(CGCG**T**GGCC)-5′ | 10.8 | 258.0 | 2714.3 | 2714.8 |
| **10A** | 3′-d(CGCG**A**GGCC)-5′ | 10.0 | 257.2 | 2723.4 | 2723.8 |
| **10C** | 3′-d(CGCG**C**GGCC)-5′ | 11.0 | 271.2 | 2699.3 | 2699.8 |
| **10G** | 3′-d(CGCG**G**GGCC)-5′ | 10.5 | 256.0 | 2739.4 | 2739.8 |

[a] See Section 5 for the HPLC conditions. [b] X refers to the artificial nucleoside **1**.

used (Figure 1(e)). In all likelihood, $Pd^{2+}$ forms a tridentate chelate with the artificial nucleoside **1** by binding to N1 of the purine ring and N2 of the pyrazolyl moieties, the fourth coordination site being filled by the N3 of uridine (Figure 3) [14]. The ability of thymine to form base pairs mediated by N3-coordinated soft metal ions has previously been demonstrated and uracil most likely exhibits similar behavior [15–19]. While no spectra could be obtained at sufficiently low concentrations to allow for reliable determination of the stability of the complex, the data at hand suggest a dissociation constant in the low micromolar range.

*2.4. CD Spectropolarimetric Measurements.* To obtain information about the impact of the modified nucleoside and

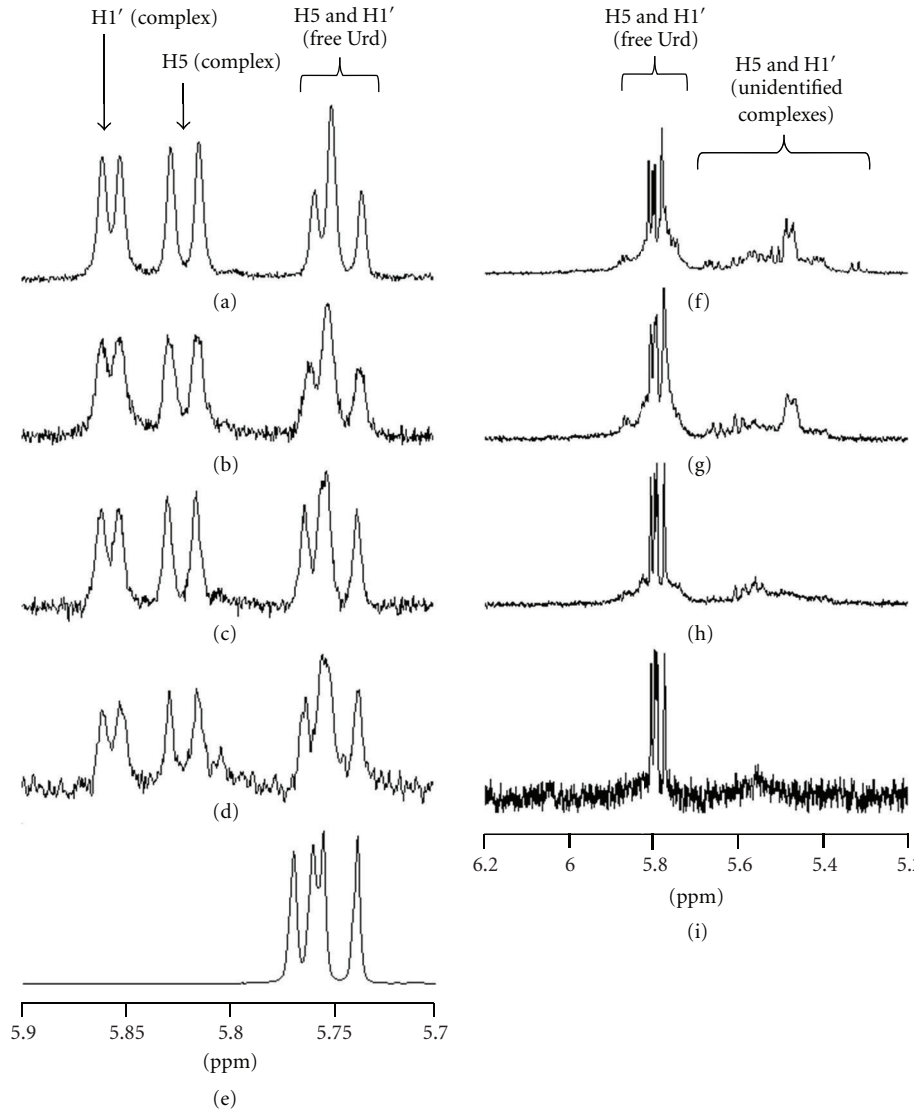



Figure 1: $^1$H NMR spectra of uridine (H5 and H1′) at various concentrations of uridine, **1** and $Pd^{2+}$ at pH 7.3. (a) [Urd] = [**1**] = 3.6 mmol $L^{-1}$, [$Pd^{2+}$] = 2.52 mmol $L^{-1}$; (b) [Urd] = [**1**] = 1.49 mmol $L^{-1}$, [$Pd^{2+}$] = 1.04 mmol $L^{-1}$; (c) [Urd] = [**1**] = 0.57 mmol $L^{-1}$, [$Pd^{2+}$] = 0.40 mmol $L^{-1}$; (d): [Urd] = [**1**] = 0.061 mmol $L^{-1}$, [$Pd^{2+}$] = 0.042 mmol $L^{-1}$; (e) [Urd] = [**1**] = 13.84 mmol $L^{-1}$, [$Pd^{2+}$] = 0; (f) [Urd] = 3.6 mmol $L^{-1}$, [$Pd^{2+}$] = 2.52 mmol $L^{-1}$; (g) [Urd] = 1.49 mmol $L^{-1}$, [$Pd^{2+}$] = 1.04 mmol $L^{-1}$; (h) [Urd] = 0.57 mmol $L^{-1}$, [$Pd^{2+}$] = 0.40 mmol $L^{-1}$; (i) [Urd] = 0.061 mmol $L^{-1}$.

$Pd^{2+}$ on the secondary structure of oligonucleotides, CD spectra of the duplexes of the modified (**7X**) and unmodified (**7A**) 2′-*O*-methyl oligoribonucleotides with complementary DNA or 2′-*O*-methyl-RNA oligonucleotides were measured over a wide temperature range (6–94°C) in the presence and absence of $Pd^{2+}$. The experiments were carried out at pH 7.4 (20 mmol $L^{-1}$ cacodylate buffer), the ionic strength being adjusted to 0.1 mol $L^{-1}$ with $NaClO_4$. The concentration of the oligonucleotides was 3.0 $\mu$mol $L^{-1}$ in each experiment and the concentration of $Pd^{2+}$ (added as $K_2PdCl_4$) either 0 or 3.0 $\mu$mol $L^{-1}$. In all cases, at low temperatures the CD spectra are characteristic of an A-type duplex and almost all of these duplexes denature on increasing temperature, as evidenced by the gradual decrease of the CD signals [20–23]. The sole exception to this behavior is the duplex **7X**:**10T**, having thymidine opposite to the artificial nucleoside. This duplex seemed to be somewhat resistant to heating in the presence, but not in the absence, of $Pd^{2+}$. Illustrative examples of the CD spectra obtained are presented in Figure 4 (see Supplementary Material available online at doi:10.1155/2012/196845).

The thermal loss of ellipticity for selected duplexes in the presence and absence of $Pd^{2+}$ is presented in Figures 5–7. In the case of the 2′-*O*-methyl-RNA/DNA heteroduplexes, the unmodified full-match duplex **7A**:**10T** exhibits a sigmoid curve with an inflection point at approximately 50°C regardless of the presence of $Pd^{2+}$ (Figure 5(a)). The curves for a corresponding mismatched duplex, **7A**:**10C**, also essentially overlap but are much more linear with no clear inflection point. With the modified oligonucleotide

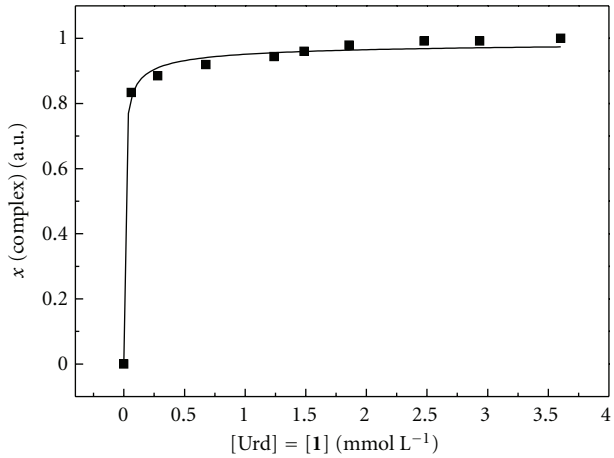



Figure 2: Mole fraction of uridine in complex as a function of the concentration of uridine and **1**; pH = 7.3.

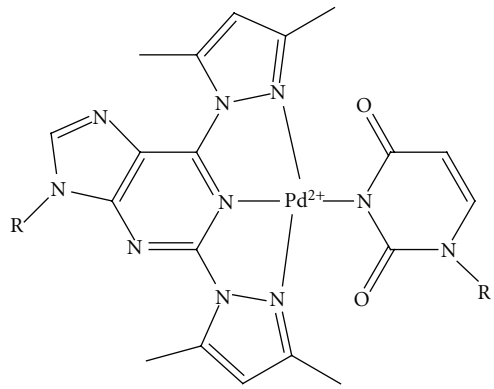



Figure 3: Proposed ternary complex between **1**, $Pd^{2+}$, and uridine.

**7X**, the situation is quite different: the curves for the "mismatched" duplex **7X**:**10C** in the presence and absence of $Pd^{2+}$ overlap with each other and also with the curve for the "full-match" duplex **7X**:**10T** in the absence of $Pd^{2+}$ (Figure 5(b)). In the presence of $Pd^{2+}$, however, duplex **7X**:**10T** is remarkably resistant to heating, losing significantly less of its ellipticity than the other combinations. In other words, the artificial 2,6-bis(3,5-dimethyl-1*H*-pyrazol-1-yl)purine nucleobase seems to be able to recognize thymine, but only in the presence of $Pd^{2+}$.

With the 2′-*O*-methyl-RNA homoduplexes between unmodified oligonucleotides, the ellipticity of the matched duplex **7A**:**8U** exhibits a strongly sigmoid curve with an inflection point at approximately 75°C and the mismatched duplex **7A**:**8C** shows a similar curve with an inflection point at approximately 60°C (Figure 6(a)). The inflection points are essentially independent on the presence of $Pd^{2+}$, but in the case of the mismatched duplex **7A**:**8C**, the loss of ellipticity at high temperatures is somewhat lower in the presence of $Pd^{2+}$. In contrast, no discrimination between uridine and cytidine is observed with the modified oligonucleotide **7X**: both duplexes **7X**:**8U** and **7X**:**8C** exhibit largely similar sigmoid curves with inflection points at approximately 60°C (Figure 6(b)). In both cases, the maximum loss of ellipticity is somewhat lower in the presence of $Pd^{2+}$. The plots of the duplexes containing multiple mismatches are essentially linear and independent on both the sequence and the presence of $Pd^{2+}$ (Figure 7).

*2.5. Melting Temperature Measurements.* The melting temperatures of the duplexes formed between the modified oligonucleotide **7X** and complementary 2′-*O*-methyl-RNA or DNA sequences bearing A, G, U (or T), or C nucleoside opposite to the artificial nucleoside **1** were measured in the presence and absence of $Pd^{2+}$. For comparison, the respective duplexes formed by the unmodified oligonucleotide **7A** were also studied. The $T_m$ values were measured in a 20 mmol $L^{-1}$ cacodylate buffer at pH 7.4, the ionic strength being adjusted to 0.1 mol $L^{-1}$ with $NaClO_4$. The concentration of the oligonucleotides was 3.0 $\mu$mol $L^{-1}$ in each experiment and the concentration of $Pd^{2+}$ (added as $K_2PdCl_4$) either 0 or 3.0 $\mu$mol $L^{-1}$. The results of the $T_m$ measurements are summarized in Table 2.

As expected, the unmodified oligonucleotide **7A** forms the most stable 2′-*O*-methyl-RNA/DNA heteroduplex with the fully complementary DNA oligonucleotide **10T**, but the discrimination is not very strict. For example, the full-match duplex **7A**:**10T** exhibits a $T_m$ value of only 5°C higher than the mismatched duplex **7A**:**10A** (50.0 and 45.0°C, resp.). Adding 1 eq. of $Pd^{2+}$ to the oligonucleotide mixtures has little effect on the stability of any of the duplexes studied. Unexpectedly, all of the respective heteroduplexes formed by the modified oligonucleotide **7X** turned out to be significantly more stable than the ones formed by **7A**, with $T_m$ values ranging from 58.6 to 62.6°C. The most stable duplex was the one between **7X** and **10T**, although the selectivity was even lower than in the case of **7A**. Furthermore, the selectivity is only marginally improved in the presence of $Pd^{2+}$.

In the case of the 2′-*O*-methyl-RNA homoduplexes, the unmodified oligonucleotide **7A** expectedly forms the most stable duplex with the fully oligonucleotide **8U** placing a uridine opposite to the central adenosine of **7A** ($T_m$ = 75°C). Introducing a mismatch on both sides of this central base pair results in $T_m$ a drop of 46°C with a matched central base pair (**7A**:**8U** → **7A**:**9U**) and a drop of 12–17°C with a mismatched central base pair (**7A**:**8A** → **7A**:**9A**, **7A**:**8G** → **7A**:**9G**, and **7A**:**8C** → **7A**:**9C**). When the central adenosine of **7A** is replaced with the artificial nucleoside **1** (resulting in the modified oligonucleotide **7X**), $T_m$ values of approximately 60°C are observed with all the homoduplexes studied, both in the presence and in the absence of $Pd^{2+}$. In other words, the modification is strongly destabilizing compared to the full-match 2′-*O*-methyl-RNA homoduplex **7A**:**8U**, neither stabilizing nor destabilizing in the 2′-*O*-methyl-RNA homoduplexes containing a single mismatch and, unexpectedly, strongly stabilizing in the 2′-*O*-methyl-RNA homoduplexes containing multiple mismatches (as well as in the 2′-*O*-methyl-RNA/DNA heteroduplexes).

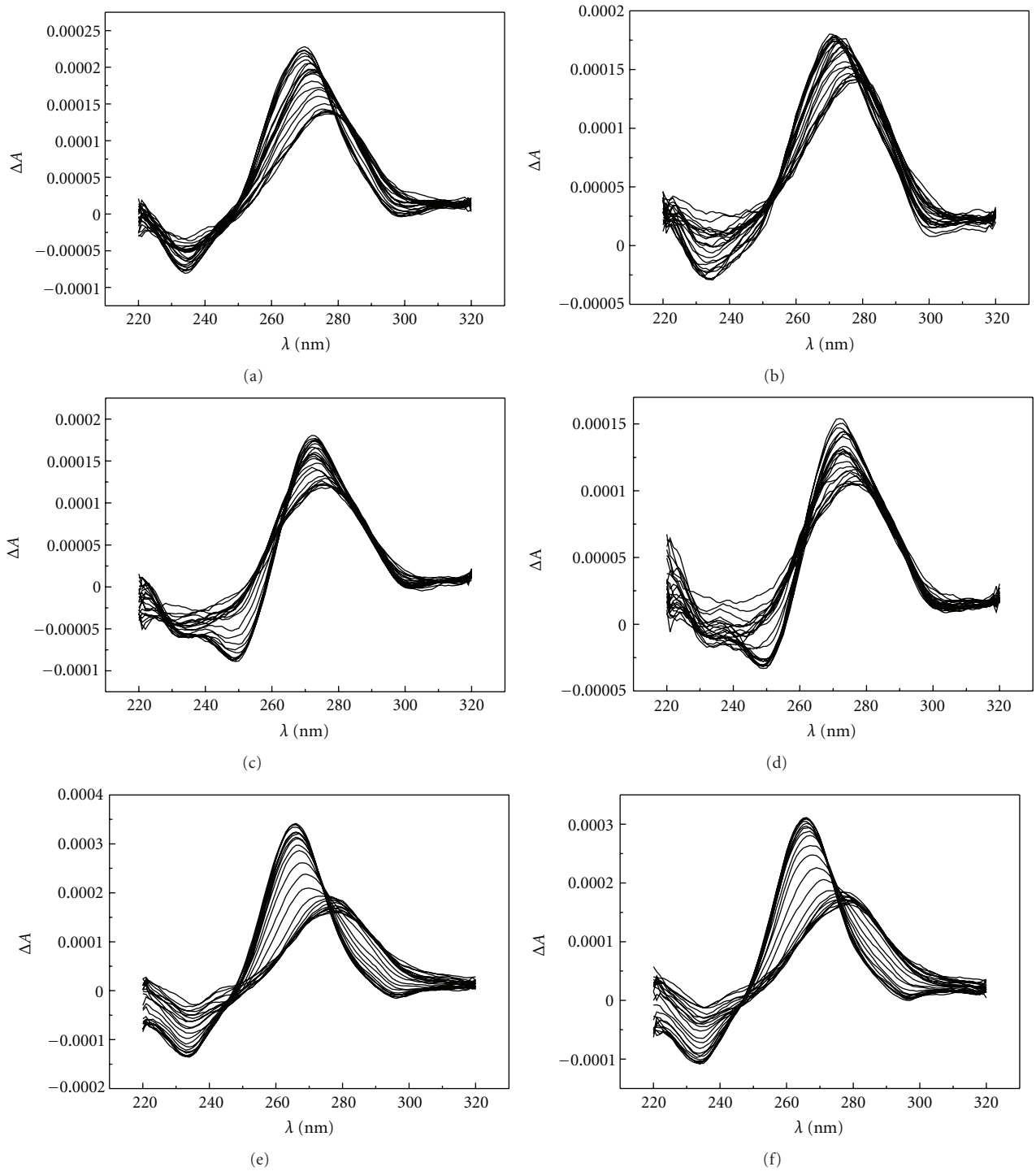



Figure 4: CD spectra of the duplexes formed between the modified (**7X**) or unmodified (**7A**) oligonucleotide and complementary DNA oligonucleotides having thymidine (**10T**) or 2′-deoxycytidine (**10C**) opposite to the artificial monomer (or adenosine) in the presence and absence of $Pd^{2+}$; [oligonucleotides] = 3.0 $\mu$M, $[K_2PdCl_4]$ = 0/3.0 $\mu$M, $I(NaClO_4)$ = 0.1 mol $L^{-1}$, pH = 7.4. (a) **7X:10T**, (b) **7X:10T** + $Pd^{2+}$, (c) **7X:10C**, (d) **7X:10C** + $Pd^{2+}$, (e) **7A:10T**, (f) **7A:10T** + $Pd^{2+}$.

## 3. Discussion

The $Pd^{2+}$ chelate of the 2,6-bis-(3,5-dimethylpyrazol-1-yl) purine nucleobase presented in this study has no potential for hydrogen bonding interactions with the canonical nucleobases—in other words, its base pairing relies solely on metal ion coordination. On the other hand, the modified nuclobase presents a much greater surface for $\pi$-$\pi$ stacking interactions and the observed $Pd^{2+}$-independent stabilization of the mismatched and 2′-*O*-methyl-RNA/DNA

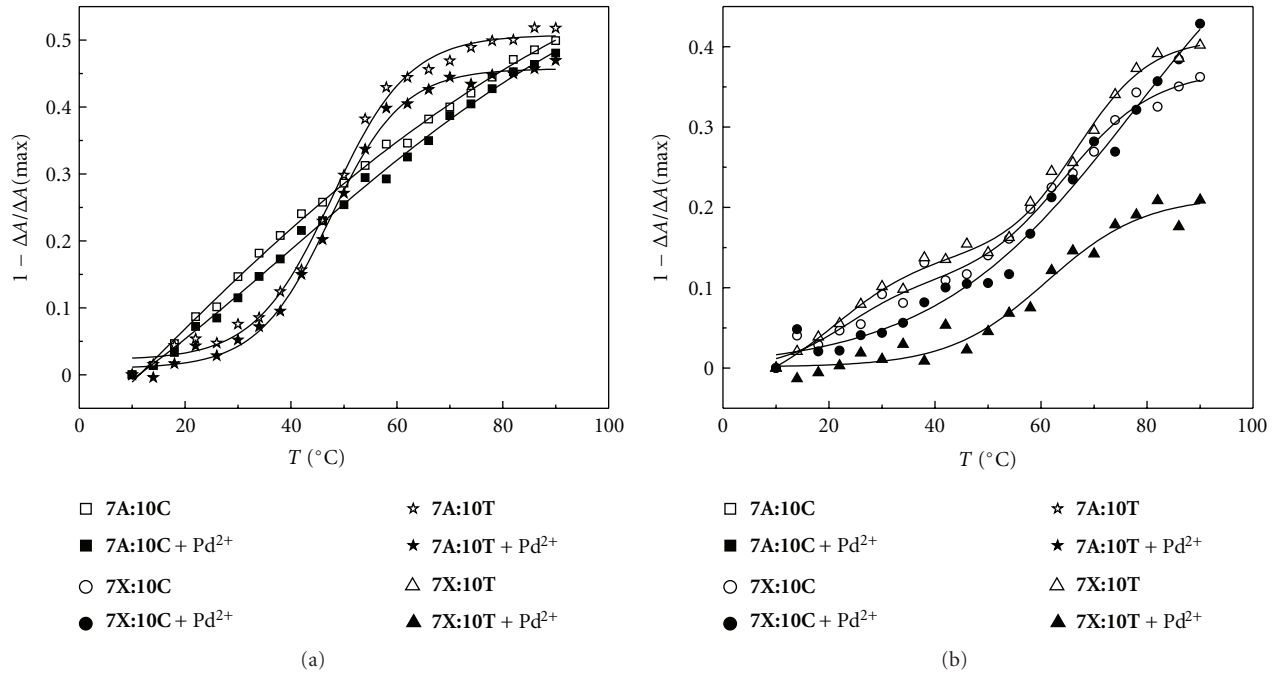



Figure 5: Temperature versus $1 - \Delta A/\Delta A(\max)$ profiles of the duplexes formed between the modified (**7X**) and unmodified (**7A**) oligonucleotides with the DNA oligonucleotides **10C** and **10T** in the presence and absence of $Pd^{2+}$; [oligonucleotides] = 3.0 $\mu$M, [$K_2PdCl_4$] = 0/3.0 $\mu$M, scan range 220–320 nm; $I(NaClO_4)$ = 0.1 mol $L^{-1}$; pH = 7.4.

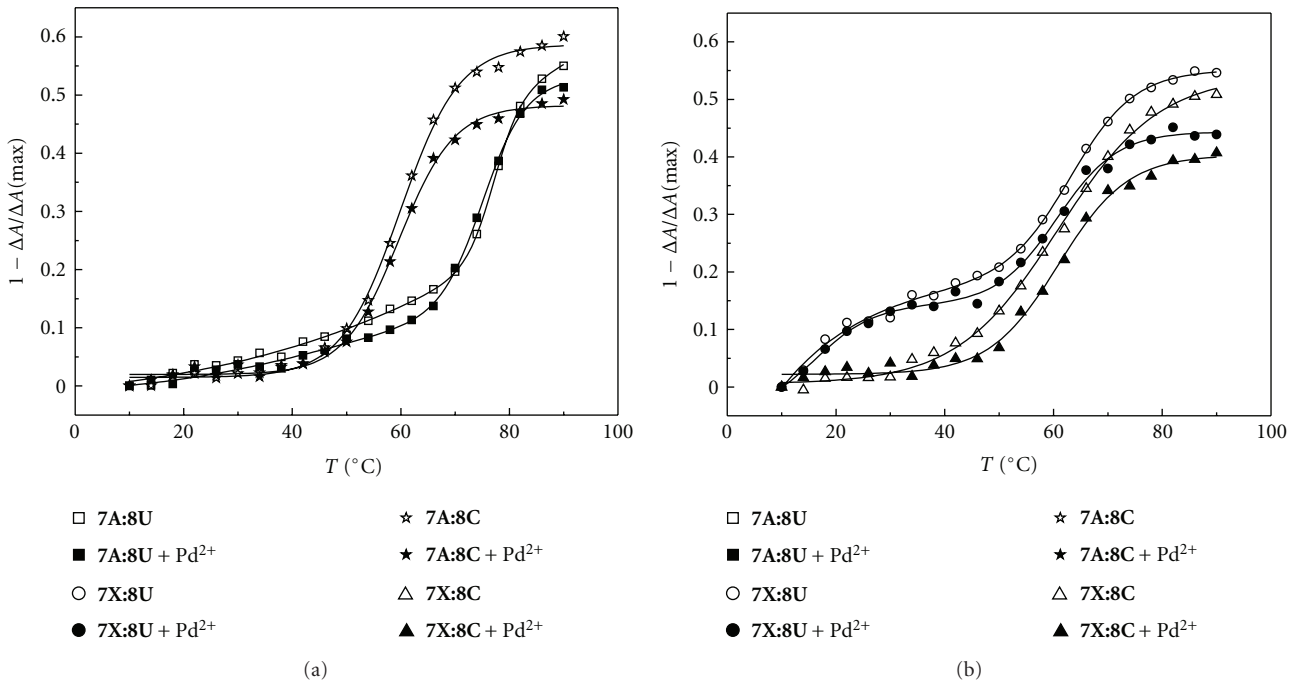



Figure 6: Temperature versus $1 - \Delta A/\Delta A(\max)$ profiles of the duplexes of the modified (**7X**) and unmodified (**7A**) oligonucleotides with the 2′-*O*-methyl-RNA oligoribonucleotides **8U** and **8C** in the presence and absence of $Pd^{2+}$; [oligonucleotides] = 3.0 $\mu$M, [$K_2PdCl_4$] = 0/3.0 $\mu$M, scan range 220–320 nm; $I(NaClO_4)$ = 0.1 mol $L^{-1}$; pH = 7.4.

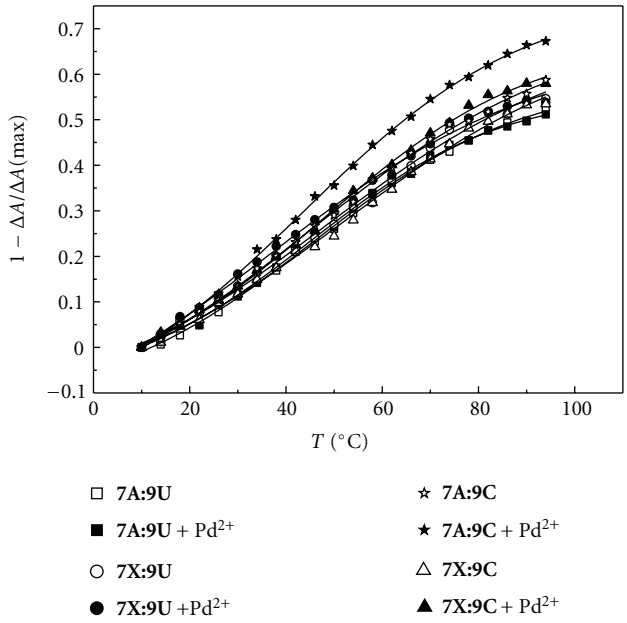



Figure 7: Temperature versus $1 - \Delta A/\Delta A(\max)$ profiles of the duplexes of the modified oligonucleotide **7X** or **7A** with oligoribonucleotide **9U** and **9C** in the presence and absence of $Pd^{2+}$; [oligonucleotides] = 3.0 $\mu$M, $[K_2PdCl_4]$ = 0/3.0 $\mu$M, scan range 220–320 nm; $I(NaClO_4)$ = 0.1 mol $L^{-1}$; pH = 7.4.

heteroduplexes is probably a result of this enhanced base-stacking. The $Pd^{2+}$-dependent selective recognition of a thymine base within a complementary DNA oligonucleotide, in turn, is probably attributable to two factors, *namely,* the relatively small size of the thymine base, as well as its potential to act as an anionic nitrogen ligand. Of the four native nucleobases, only the pyrimidines may form a base pair of Watson-Crick geometry with the artificial metallo-nucleobase (and even these base pairs would be approximately 1.3 Å longer than the natural ones). Of the two pyrimidine nucleobases, only thymine (or uracil) is readily deprotonated to yield a negatively charged nitrogen ligand with a significantly greater affinity for metal ions. Indeed, $Hg^{II}$ has been shown to selectively bind to a T-T mispair within a double-stranded oligonucleotide by coordinating to the N3 of both thymine residues [16–19]. The fact that no selectivity for uracil is observed with the 2′-*O*-methyl-RNA homoduplexes may be reasoned if the bulky dimethylpyrazole moieties cause a large enough steric hindrance to prevent the metallo-base pair from assuming planar geometry. Presumably, a nonplanar base pair should be more easily accommodated within the more flexible 2′-*O*-methyl-RNA/DNA heteroduplex than the relatively rigid 2′-*O*-methyl-RNA homoduplex.

## 4. Conclusions

A 2′-*O*-methyl-RNA oligonucleotide incorporating an artificial nucleobase that serves as a tridentate chelate for soft metal ions has been synthesized. When hybridized with a complementary DNA oligonucleotide, this modified oligonucleotide selectively recognizes a thymine residue opposite to the artificial nucleobase. The discrimination is somewhat more pronounced in the presence of 1 eq. of $Pd^{2+}$, but the difference is too small to be taken as compelling evidence for the formation of a metal-ion-mediated base pair between thymidine and **1**. Within fully matched 2′-*O*-methyl-RNA homoduplexes, the modification is destabilizing regardless of the presence of $Pd^{2+}$, probably because the proposed metallo-base pair cannot assume a planar geometry needed to fit within the relatively tight base stack of these oligonucleotides. Within mismatched duplexes, as well as 2′-*O*-methyl-RNA/DNA heteroduplexes, the modification is stabilizing but the effect is rather sequence and $Pd^{2+}$ independent, suggesting that the origin of this stabilization is the large $\pi$-$\pi$ stacking surface of the artificial nucleobase.

The metal-ion-carrying nucleobase presented in this study represents a reference structure with no potential of specific interactions with the natural nucleobases and the preference for thymine is probably based on its small size and ability to serve as an anionic nitrogen ligand. The fact that the strong base pairing observed at the monomeric level is only modestly reflected in the thermal stability of the respective double-stranded oligonucleotide is most likely due to difficulties in accommodating this (presumably) nonplanar base pair within an oligonucleotide environment. It should be noted that the CD and $T_m$ studies do not provide first-hand information on the interactions (if any) between thymine (or uracil) and the $Pd^{2+}$ complex of the artificial nucleoside **1**. In other words, the possibility that within a double-stranded oligonucleotide, a complex of a different geometry prevails cannot be completely ruled out. More elaborate structures featuring additional hydrogen bonding interactions as well as carefully placed steric constraints will, in all likelihood, exhibit greatly enhanced affinity and selectivity. For example, replacing the dimethylpyrazole moieties with monosubstituted hydrazines would result in a structure not only sterically less hindered than **1** but also capable of donating two hydrogen bonds to the oxo substituents of thymine or uracil. On the other hand, using a pyrimidine, rather than a purine, nucleoside as the parent compound would alleviate some of the steric crowding caused by the relatively bulky $Pd^{2+}$ ion. Finally, other soft metal ions could be screened to find the optimal combination of affinity and selectivity.

## 5. Experimental

*5.1. General.* 2′,3′,5′-Tri-*O*-acetyl-6-chloro-2-iodopurine riboside was a commercial product that was used as received. For the triethylammonium acetate buffer used in HPLC chromatography, triethylamine was distilled before using. The NMR spectra were recorded with a Bruker Avance 500 NMR spectrometer and the chemical shifts are given in ppm. The mass spectra were recorded with a Bruker micrOTOF-Q ESI-MS spectrometer and the CD spectra with an Applied Photophysics Chirascan spectropolarimeter. The

Table 2: The $T_m$ of the duplexes formed between modified (**7X**) or unmodified (**7A**) oligonucleotide and complementary 2′-*O*-methyl-RNA (or DNA) sequences bearing A, G, U (or T), or C nucleoside opposite to the artificial nucleoside **1** (in the case of oligonucleotide **7X**) or adenosine (in the case of oligonucleotide **7A**) in the presence and absence of $Pd^{2+}$; [oligonucleotides] = 3.0 $\mu$M, [$K_2PdCl_4$] = 0/3.0 $\mu$M, $I$($NaClO_4$) = 0.1 M; pH = 7.4.

| Complementary oligonucleotide | Sequence | $T_m$/°C, duplex with **7X** (5′-GCGCXCCGG-3′) $Pd^{2+}$ | | $T_m$/°C, duplex with **7A** (5′-GCGCACCGG-3′) $Pd^{2+}$ | |
|---|---|---|---|---|---|
| | | + | − | + | − |
| **8U** | 3′-CGCG**U**GGCC-5′ | 60.6 ± 0.6 | 61.8 ± 0.4 | 73.7 ± 0.1 | 75.1 ± 0.3 |
| **8A** | 3′-CGCG**A**GGCC-5′ | 61.6 ± 0.1 | 61.4 ± 0.4 | 56.3 ± 0.1 | 56.3 ± 0.2 |
| **8G** | 3′-CGCG**G**GGCC-5′ | 60.1 ± 0.6 | 61.6 ± 0.3 | 60.4 ± 0.1 | 61.2 ± 0.3 |
| **8C** | 3′-CGCG**C**GGCC-5′ | 59.9 ± 0.4 | 60.7 ± 0.2 | 58.8 ± 0.2 | 59.0 ± 0.2 |
| **9U** | 3′-CGC**AUA**GCC-5′ | 60.4 ± 0.1 | 60.4 ± 0.3 | 27.6 ± 0.1 | 26.9 ± 0.3 |
| **9A** | 3′-CGC**AAA**GCC-5′ | 61.7 ± 0.9 | 61.0 ± 0.9 | 44.3 ± 0.6 | 43.6 ± 0.7 |
| **9G** | 3′-CGC**AGA**GCC-5′ | 60.1 ± 0.2 | 61.8 ± 0.4 | 43.5 ± 0.9 | 44.3 ± 0.3 |
| **9C** | 3′-CGC**ACA**GCC-5′ | 61.1 ± 0.3 | 60.3 ± 0.5 | 43.9 ± 0.7 | 45.7 ± 0.8 |
| **10T** | 3′-d(CGCG**T**GGCC)-5′ | 62.6 ± 1.3 | 60.8 ± 0.2 | 49.5 ± 0.5 | 50.0 ± 0.1 |
| **10A** | 3′-d(CGCG**A**GGCC)-5′ | 59.1 ± 0.9 | 60.5 ± 0.5 | 43.5 ± 0.4 | 45.0 ± 0.3 |
| **10C** | 3′-d(CGCG**C**GGCC)-5′ | 60.7 ± 0.2 | 61.6 ± 0.1 | 45.5 ± 0.1 | 50.3 ± 0.2 |
| **10G** | 3′-d(CGCG**G**GGCC)-5′ | 58.6 ± 0.4 | 60.2 ± 0.2 | 40.8 ± 0.9 | 43.9 ± 0.5 |

oligonucleotides were assembled on an Applied Biosystems 3400 DNA/RNA synthesizer.

*5.2. 9-(β-D-Ribofuranosyl)-2,6-dihydrazinopurine (**3**).* To neat hydrazine hydrate (6.5 mL, 134 mmol) was added 2′, 3′, 5′-tri-*O*-acetyl-6-chloro-2-iodopurine riboside (1.03 g, 1.91 mmol) and the resulting mixture was stirred for 14 d at room temperature. Over the course of the first day, the appearance of the reaction mixture changed from yellowish and cloudy to clear and colorless to light red, followed by formation of a white precipitate. The reaction mixture was diluted with 2-propanol (15 mL) and stirred for 15 min, after which the precipitate was collected by filtration and washed with cold 2-propanol (2 × 100 mL). Finally, the precipitate was dried under vacuum to yield 0.534 g (89%) of **2** as a white powder. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta_H$ 7.92 (s, 1H, H8), 5.76 (d, 1H, H1′, $J$ = 6.3 Hz), 4.56 (dd, 1H, H2′, $J_1$ = 5.35, $J_2$ = 5.50 Hz), 4.11 (dd, 1H, H3′, $J_1$ = 3.1 Hz, $J_2$ = 4.95), 3.89 (m, 1H, H4′), 3.61 (dd, 1H, H5′ $J_1$ = 3.9 Hz, $J_2$ =11.9 Hz), 3.52 (dd, 1H, H5′′ $J_1$ = 3.9 Hz, $J_2$ =11.9 Hz). $^{13}$C NMR (125 MHz, DMSO) $\delta_C$ 162.2, 155.9, 136.8, 136.7, 113.5, 87.4, 85.9, 73.6, 71.2, 62.2. HRMS (ESI$^+$): *m/z* calcd 313.1367 obsd 313.1378 [M+H]$^+$.

*5.3. 9-(β-D-Ribofuranosyl)-2, 6-bis(3,5-dimethylpyrazol-1-yl) purine (**1**).* To dry compound **3** (0.337 g, 1.08 mmol) was added 6.4 eq. of dry 2,4-pentanedione (0.70 mL, 6.912 mmol) and 0.1 eq. of trifluoroacetic acid (8.27 $\mu$L, 0.108 mmol). The reaction mixture was stirred at room temperature for 16 h, during which time the product crystallized. The volatiles were removed under reduced pressure to yield crude **1** that was used in the next step without further purification. For the NMR titrations, a small sample was purified by silica gel chromatography eluting with a mixture of MeOH, $CH_2Cl_2$ and $Et_3N$ (9 : 90 : 1, *v*/*v*). $^1$H NMR (500 MHz, $D_2O$) $\delta_H$ 8.46 (s, 1H, H8), 6.02 (d, 1H, H1′, $J$ = 5.0 Hz), 6.0 (s, 1H, pyrazole), 5.88 (s, 1H, pyrazole), 4.61 (dd, 1H, H2′, $J_1$ = $J_2$ = 5.1 Hz), 4.27 (dd, 1H, H3′, $J_1$ = 5.1 Hz, $J_2$ = 5.0 Hz), 4.1 (m, 1H, H4′), 3.78 (dd, 1H, H5′, $J_1$ = 2.7 Hz, $J_2$ =12.8 Hz), 3.7 (dd, 1H, H5′′, $J_1$ = 2.7 Hz, $J_2$ = 12.8 Hz), 2.30 (s, 3H, $CH_3$), 2.28 (s, 3H, $CH_3$), 2.12 (s, 3H, $CH_3$), 2.04 (s, 3H, $CH_3$). $^{13}$C NMR (125 MHz, $D_2O$) $\delta_C$ 153.9, 153.8, 152.1, 150.5, 149.6, 146.8, 143.8, 143.5, 107.5, 106.4, 70.5, 68.0, 67.8, 28.9, 27.9, 23.1, 13.8, 13.0, 12.5, 12.3. HRMS (ESI$^+$): *m/z* calcd 441.1993 obsd 441.2036 [M + H]$^+$.

*5.4. 9-[5-O-(4,4′-Dimethoxytrityl)-β-D-ribofuranosyl]-2,6-bis(3,5-dimethylpyrazol-1-yl)purine (**5**).* To a solution of crude compound **1** (0.477 g, 1.08 mmol) in dry pyridine (10 mL) was added 1.6 eq. of DMTrCl (0.589 g, 1.738 mmol). The reaction mixture was stirred for 4 d at room temperature, after which it was concentrated under reduced pressure. The residue was dissolved in $CH_2Cl_2$ (40 mL) and washed with saturated aq. $NaHCO_3$ (50 mL). The organic phase was dried with $Na_2SO_4$ and evaporated to dryness. The residue was purified by silica gel chromatography eluting with a mixture of EtOAc, hexane, and $Et_3N$ (stepwise gradient from 60 : 39 : 1 to 99 : 0 : 1, *v*/*v*). Yield 0.765 g (82.7%). $^1$H NMR (500 MHz, $CDCl_3$) $\delta_H$ 8.36 (s, 1H, H8), 7.24 (m, 9H, Ar), 6.77 (m, 4H, Ar), 6.43 (d, 1H, H1′, $J$ = 6.4 Hz), 6.11 (s, 1H, pyrazole), 6.05 (s, 1H, pyrazole), 4.74 (dd, 1H, H2′, $J_1$ = 5.4 Hz, $J_2$ = 5.5 Hz), 4.42 (m, 2H, H3′ & H4′), 3.76 (s, 6H, $OCH_3$), 3.44 (dd, 1H, H5′, $J_1$ = 3.4 Hz, $J_2$ =10.5 Hz), 3.27 (dd, 1H, H5′′, $J_1$= 3.4 Hz, $J_2$ = 10.5 Hz), 2.66 (s, 3H, $CH_3$), 2.63 (s, 3H, $CH_3$), 2.39 (s, 3H, $CH_3$), 2.27 (s, 3H, $CH_3$). $^{13}$C NMR (125 MHz, $CDCl_3$) $\delta_C$ 158.5, 153.1, 151.2, 150.7, 150.4, 149.8, 149.5, 144.4, 143.2, 142.8, 142.7, 135.5, 130.1, 128.1, 127.9, 126.9, 123.8, 122.9, 113.3, 110.5, 110.1, 88.9,

86.6, 85.6, 76.1, 72.2, 63.5, 55.2, 14.8, 14.1, 13.9, 13.8. HRMS ($ESI^+$): *m/z* calcd 765.3227 obsd 765.3125 $[M + Na]^+$.

*5.5. 9-[5-O-(4,4′-Dimethoxytrityl)-2-O-tert-butyldimethylsilyl-β-D-ribofuranosyl]-2,6-bis(3,5-dimethylpyrazol-1-yl)purine (**6**).* To a solution of dry compound **5** (0.543 g, 0.731 mmol) in dry DMF (10 mL) was added a solution of dry imidazole (0.5 g, 7.34 mmol) in DMF (15 mL), followed by 2.63 eq. of TBDMSCl (0.29 g, 1.92 mmol). After being stirred for 4 days, the reaction was quenched by adding MeOH (3 mL). Stirring was continued for 7 min, after which EtOAc (40 mL) and water (50 mL) were added and the organic and aqueous phases separated. The organic phase was dried with $Na_2SO_4$ and evaporated to dryness. The residue was purified by silica gel chromatography eluting with a mixture of EtOAc, $CH_2Cl_2$, and $Et_3N$ (stepwise gradient from 10 : 89 : 1 to 20 : 79 : 1, *v*/*v*). Yield 0.527 g (84%). $^1H$ NMR (500 MHz, $CDCl_3$) $\delta_H$ 8.37 (s, 1H, H8), 7.31 (m, 9H, Ar), 6.82 (m, 4H, Ar), 6.23 (d, 1H, H1′, $J$ = 5.9 Hz), 6.10 (s, 1H, pyrazole), 6.04 (s, 1H, pyrazole), 4.79 (dd, 1H, H2′, $J_1$ = 5.5 Hz, $J_2$ = 5.6 Hz), 4.32 (m, 1H, H3′), 4.27 (m, 1H, H4′), 3.77 (s, 6H, $OCH_3$), 3.52 (dd, 1H, H5′, $J_1$ = 3.5 Hz, $J_2$ = 9.2 Hz), 3.38 (dd, 1H, H5′′, $J_1$ = 2.4 Hz, $J_2$ = 9.2 Hz), 2.82 (s, 3H, $CH_3$), 2.65 (s, 3H, $CH_3$), 2.40 (s, 3H, $CH_3$), 2.33 (s, 3H, $CH_3$), 0.82 (s, 9H, Si–$CCH_3$), 0.00 (s, 3H, $SiCH_3$), −0.17 (s, 3H, $SiCH_3$). $^{13}C$ NMR (125 MHz, $CDCl_3$) $\delta_C$ 158.6, 154.6, 153.4, 151.5, 150.8, 150.1, 144.4, 144.2, 142.5, 142.2, 135.4, 130.0, 128.0, 127.0, 121.8, 113.3, 110.7, 109.8, 87.3, 86.8, 84.2, 76.7, 71.7, 63.6, 55.2, 25.7, 25.5, 17.9, 15.0, 14.3, 13.9, −4.8, −5.1. HRMS ($ESI^+$): *m/z* calcd 879.4092 obsd 879.4038 $[M + Na]^+$.

*5.6. 9-{5-O-(4,4′-Dimethoxytrityl)-2-O-tert-butyldimethylsilyl-3-O-[(2-cyanoethoxy)-(N,N-di-isopropylamino)phosphinyl]-β-D-ribofuranosyl}-2,6-bis(3,5-dimethylpyrazol1yl)purine (**2**).* To a solution of dry compound **6** (114 mg, 0.133 mmol) in dry DCM (4 mL) was added 9 eq. of dry $Et_3N$ (0.167 mL, 1.197 mmol) and 2.5 eq. of 2-cyanoethyl-*N*,*N*-di-isopropylchlorophosphoramidite (75 μL, 0.333 mmol). The mixture was stirred under nitrogen atmosphere for 3 days, after which the reaction was quenched with MeOH (200 μL). Stirring was continued for 10 min, after which $CH_2Cl_2$ (40 mL) was added and the resulting solution washed with saturated aq. $NaHCO_3$ (40 mL). The organic phase was dried with $Na_2SO_4$ and evaporated to dryness. The residue was purified by silica gel chromatography eluting with a mixture of EtOAc, $CH_2Cl_2$, and $Et_3N$ (10 : 89 : 1, *v*/*v*). Yield 0.111 g (79%). $^1H$ NMR (500 MHz, $CDCl_3$) $\delta_H$ 8.41 (s, 1H, H8), 7.33 (m, 9H, Ar), 6.83 (m, 4H, Ar), 6.29 (m, 1H, H1′), 6.10 (s, 1H, pyrazole), 6.04 (s, 1H, pyrazole), 4.82 (dd, 1H, H2′, $J_1$ = $J_2$ = 5.50 Hz), 4.35 (m, 2H, H3′ & H4′), 3.78 (s, 6H, $OCH_3$), 3.62 (m, 2H, $CH_2$), 3.60 (m, 1H, H5′), 3.50 (m, 1H, H5′′), 2.86 (s, 3H, $CH_3$), 2.67 (s, 3H, $CH_3$), 2.64 (s, 2H, CH), 2.40 (s, 3H, $CH_3$), 2.33 (s, 3H, $CH_3$), 2.27 (m, 2H, $CH_2$), 1.19 (d, 6H, $CH_3$, $J$ = 7.5 Hz), 1.05 (d, 6H, $CH_3$, $J$ = 8.3 Hz), 0.75 (s, 9H, Si–$CCH_3$), 0.00 (s, 3H, $SiCH_3$), −0.16 (s, 3H, $SiCH_3$). $^{13}C$ NMR (125 MHz, $CDCl_3$) $\delta_C$ 158.6, 154.9, 153.2, 151.5, 150.8, 150.1, 144.3, 144.2, 142.5, 142.4, 142.1, 135.3, 130.1, 130.0, 128.1, 128.0, 127.1, 121.8, 117.5, 113.3, 110.6, 109.8, 86.9, 86.8, 84.7, 84.1, 72.9, 63.5, 58.9, 58.8, 55.2, 43.0, 25.6, 24.6, 20.5, 18.0, 15.0, 14.9, 14.3, 13.9, −4.6, −5.0, −5.1. $^{31}P$ NMR (162 MHz, $CDCl_3$) $\delta_P$ 152.1, 148.9. HRMS ($ESI^+$): *m/z* calcd 1057.5218 obsd 1057.5188 $[M + H]^+$.

*5.7. Oligonucleotide Synthesis.* The oligonucleotides were assembled on a CPG-supported succinyl linker at a loading of 27 mmol $g^{-1}$. Standard phosphoramidite strategies for RNA (600 s coupling time) or DNA (20 s coupling time) were used throughout the sequences, except for the modified phosphoramidite building block **2**, which was coupled manually using an increased coupling time (60 min). Based on trityl response, coupling yield for building block **2** was 36% and the other couplings proceeded with normal efficiency. The products were released from support and deprotected by conventional treatment with 33% aq. $NH_3$ (5 hours at 55°C). The crude oligonucleotides were purified by RP-HPLC on a Hypersil ODS column C18 (250 × 4.6 mm, 5 μm) eluting with a mixture of 0.10 mol $L^{-1}$ aq. triethylammonium acetate and MeCN, the flow rate being 1.0 mL $min^{-1}$. The amount of MeCN was increased linearly from 10 to 40% during 25 min for the 2′-*O*-methyl-RNA oligonucleotides and from 10 to 30% during 25 min for the DNA oligonucleotides. The purified products were characterized by ESI-MS analysis.

## Acknowledgments

The kind help of Dr. Jorma Arpalahti with the coordination chemistry of $Pd^{2+}$ and financial support from the Academy of Finland are gratefully recognized.

## References

[1] J. R. Thomas and P. J. Hergenrother, "Targeting RNA with small molecules," *Chemical Reviews*, vol. 108, no. 4, pp. 1171–1224, 2008.

[2] P. E. Nielsen and M. Egholm, *Peptide Nucleic Acids*, Horizon Bioscience, Wymondham, UK, 2nd edition, 2004.

[3] T. Ishizuka, J. Yoshida, Y. Yamamoto et al., "Chiral introduction of positive charges to pna for double-duplex invasion to versatile sequences," *Nucleic Acids Research*, vol. 36, no. 5, pp. 1464–1471, 2008.

[4] H. Lönnberg, "Proton and metal ion interaction with nucleic acid bases, nucleosides, and nucleoside monophosphates," in *Chemistry—Coordination Equilibria in Biologically Active Systems*, K. Burger, Ed., pp. 284–346, Ellis Horwood Ltd., Chichester, UK, 1990.

[5] J. Müller, "Metal-ion-mediated base pairs in nucleic acids," *European Journal of Inorganic Chemistry*, no. 24, pp. 3749–3763, 2008.

[6] G. H. Clever, C. Kaul, and T. Carell, "DNA-metal base pairs," *Angewandte Chemie, International Edition*, vol. 46, no. 33, pp. 6226–6236, 2007.

[7] E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg, and P. G. Schultz, "A novel copper-mediated DNA base pair," *Journal of the American Chemical Society*, vol. 122, no. 43, pp. 10714–10715, 2000.

[8] B. D. Heuberger, D. Shin, and C. Switzer, "Two Watson-Crick-like metallo base-pairs," *Organic Letters*, vol. 10, no. 6, pp. 1091–1094, 2008.

[9] M. Shionoya and K. Tanaka, "Artificial metallo-DNA: a bio-inspired approach to metal array programming," *Current Opinion in Chemical Biology*, vol. 8, no. 6, pp. 592–597, 2004.
[10] K. Tanaka, Y. Yamada, and M. Shionoya, "Formation of silver(I)-mediated dna duplex and triplex through an alternative base pair of pyridine nucleobases," *Journal of the American Chemical Society*, vol. 124, no. 30, pp. 8802–8803, 2002.
[11] K. Tanaka, A. Tengeiji, T. Kato, N. Toyama, and M. Shionoya, "A discrete self-assembled metal array in artificial DNA," *Science*, vol. 299, no. 5610, pp. 1212–1213, 2003.
[12] K. Niiya, R. A. Olsson, R. D. Thompson, S. K. Silvia, and M. Ueeda, "2-(N′-alkylidenehydrazino)adenosines: potent and selective coronary vasodilators," *Journal of Medicinal Chemistry*, vol. 35, no. 24, pp. 4557–4561, 1992.
[13] K. A. Brien, C. M. Garner, and K. G. Pinney, "Synthesis and characterization of 2,6-bis-hydrazinopyridine, and its conversion to 2,6-bis-pyrazolylpyridines," *Tetrahedron*, vol. 62, no. 15, pp. 3663–3666, 2006.
[14] M. A. Halcrow, "The synthesis and coordination chemistry of 2,6-bis(pyrazolyl)pyridines and related ligands—versatile terpyridine analogues," *Coordination Chemistry Reviews*, vol. 249, no. 24, pp. 2880–2908, 2005.
[15] L. D. Kosturko, C. Folzer, and R. F. Stewart, "The crystal and molecular structure of a 2 : 1 complex of 1-methylthymine-mercury(II)," *Biochemistry*, vol. 13, no. 19, pp. 3949–3952, 1974.
[16] Z. Kuklenyik and L. G. Marzilli, "Mercury(II) site-selective binding to a DNA hairpin. relationship of sequence-dependent intra- and interstrand cross-linking to the hairpin—duplex conformational transition," *Inorganic Chemistry*, vol. 35, no. 19, pp. 5654–5662, 1996.
[17] A. Ono and H. Togashi, "Highly selective oligonucleotide-based sensor for mercury(II) in aqueous solutions," *Angewandte Chemie, International Edition*, vol. 43, no. 33, pp. 4300–4302, 2004.
[18] Y. Miyake, H. Togashi, M. Tashiro et al., "Mercury$^{II}$-mediated formation of thymine–Hg$^{II}$–thymine base pairs in DNA duplexes," *Journal of the American Chemical Society*, vol. 128, no. 7, pp. 2172–2173, 2006.
[19] Y. Tanaka, S. Oda, H. Yamaguchi, Y. Kondo, C. Kojima, and A. Ono, "$^{15}$N-$^{15}$N J-coupling across Hg$^{II}$: direct observation of Hg$^{II}$-mediated T-T base pairs in a DNA duplex," *Journal of the American Chemical Society*, vol. 129, no. 2, pp. 244–245, 2007.
[20] J. Brahms, A. M. Michelson, and K. E. van Holde, "Adenylate oligomers in single- and double-strand conformation," *Journal of Molecular Biology*, vol. 15, no. 2, pp. 467–488, 1966.
[21] J. Brahms, J. C. Maurizot, and A. M. Michelson, "Conformation and thermodynamic properties of oligocytidylic acids," *Journal of Molecular Biology*, vol. 25, no. 3, pp. 465–480, 1967.
[22] C. A. Busch and H. A. Scheraga, "Optical activity of single-stranded polydeoxyadenylic and polyriboadenylic acids; dependence of adenine chromophore cotton effects on polymer conformation," *Biopolymers*, vol. 7, no. 3, pp. 395–409, 1969.
[23] H. T. Steely Jr., D. M. Gray, and R. L. Ratliff, "CD of homopolymer DNA-RNA hybrid duplexesand triplexes containing A.T or A.U base pairs," *Nucleic Acids Research*, vol. 14, no. 24, pp. 10071–10090, 1986.