# Recognition of Remote Mismatches by DNA Polymerases

Michael Strerath,[a, b] Jens Gaster,[a, b] and Andreas Marx*[a, b]

The ability of DNA polymerases to selectively copy a template strand according to the Watson–Crick rule is crucial for the survival of any species. Besides editing the insertion of a canonical nucleotide opposite the corresponding template nucleobase, most DNA polymerases proceed with extension from mismatched primer termini with significantly diminished efficiency compared to matched substrates.[1] This is a crucial parameter for the prevention of inadvertent sealing of mutations and might allow other cellular processes to come into play to repair the nascent DNA strand.[1] Thus, mismatch extension discrimination significantly contributes to overall DNA polymerase selectivity.[1] Several crystal structures of DNA polymerases bound to their substrates indicate that the enzymes make complex interactions with the primer–template complex and nucleotide substrates during catalysis of DNA polymerization.[2] Contact points with the primer–template complex are manifold and reach upto several nucleotide pairs beyond the catalytic center. These points of enzyme contact with the primer–template complex might be responsible for the observed discrimination of mismatch extension through auditing of Watson–Crick hydrogen bonding patterns as well as shape complementary.[1–3] Interestingly, it has been shown for *Escherichia coli* DNA polymerase I, that it transfers the primer strand to its 3′-5′-exonuclease site even when mismatches are located at distal positions beyond the catalytic center.[4] Here we show that the auditing of canonical base pairing within the primer–template duplex applies to DNA polymerases that are deficient of 3′-5′-exonuclease functionality and also originate from different DNA polymerase families. Interestingly, DNA polymerase discrimination of canonical over noncanonical duplexes containing single mismatches up to four nucleotide positions beyond the catalytic center, is further increased through chemical modification of the primer strands. We show that these features can be exploited for efficient DNA diagnosis by employing real-time polymerase chain reaction (PCR).

To test the action of distal mismatches in the primer–template complex on DNA polymerase function, we designed primer template complexes in which mismatches were moved from the 3′-terminal position up to five positions away from the catalytic center (Figure 1 A). These DNA duplexes served as substrates for single nucleotide (i.e., dAMP) insertion opposite

[a] *Dipl.-Chem. M. Strerath, Dipl.-Chem. J. Gaster, Prof. Dr. A. Marx*
*Kekulé-Institut für Organische Chemie und Biochemie, Universität Bonn*
*Gerhard-Domagk-Straße 1, 53121 Bonn (Germany)*
*E-mail: Andreas.Marx@uni-konstanz.de*

[b] *Dipl.-Chem. M. Strerath, Dipl.-Chem. J. Gaster, Prof. Dr. A. Marx*
*Current address:*
*Fachbereich Chemie, Universität Konstanz*
*Universitätsstraße 10, 78457 Konstanz (Germany)*
*Fax: (+49) 7531-88-5140*

 **DOI: 10.1002/cbic.200400107**  

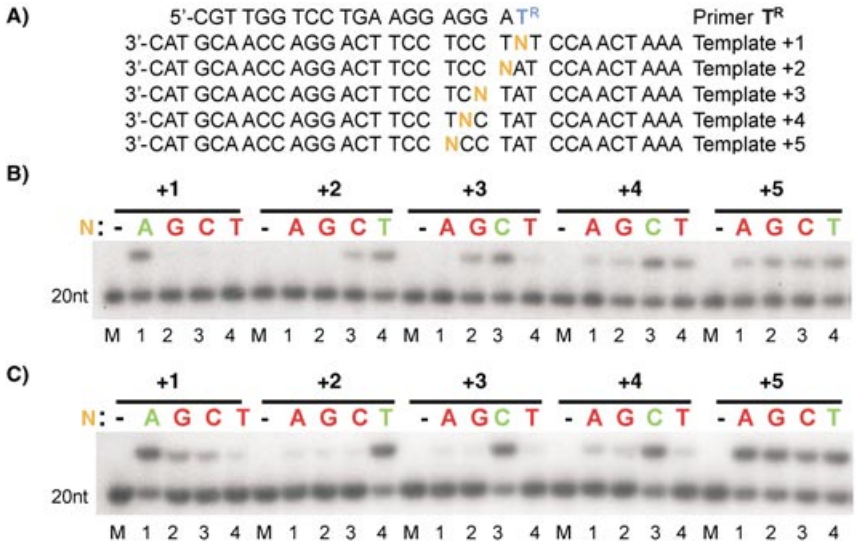


***Figure 1.*** *A) DNA substrates, B) single nucleotide insertion catalyzed by Taq DNA polymerase, C) single nucleotide insertion catalyzed by Vent exo⁻ DNA polymerase. All reactions contained equal amounts of the respective enzyme, dATP, and unmodified primer template complex. M: marker. Sequences of primer and templates are indicated in the figure. Green indicates matched cases, red indicates mismatched cases. N = A, G, C, or T, as indicated in panels B and C.*

the canonical T residue in the template strand. We investigated two DNA polymerases deficient of 3′-5′-exonuclease activity: *Thermus aquaticus* (Taq) DNA polymerase, a member of DNA polymerase family A and an exonuclease deficient mutant DNA polymerase of the archaea *Thermococcus litoralis* (Vent exo⁻ DNA polymerase), an enzyme from DNA polymerase family B.[5]

First, we evaluated the effects of distal mismatches within the primer template duplexes on both enzymes. As apparent in Figure 1B, most mismatches within the primer template complex significantly diminished nucleotide insertion by Taq DNA polymerase, whereas under identical conditions matched complexes were efficiently extended.

Mismatch discrimination seems to be most significant at position +1, but remained significant for most of the mismatches even when they were located distal to the catalytic center at positions +2 to +4. As of position +5, no significant discrimination could be detected. Vent exo⁻ DNA polymerase was also able to detect single nucleotide mismatches remote from the catalytic center (Figure 1C). In these cases no strong preference for the location of a respective mismatch in between position +1 to +4 was observed. Interestingly, although both enzymes are members of different DNA polymerase families with little sequence homology,[5] they continue with DNA synthesis at lower efficiency even when mismatches are located up to four nucleotide moieties distal to the 3′-primer terminus. The ability of DNA polymerases to distinguish between canonically and noncanonically paired primer template duplexes further adds to overall enzyme fidelity.[1] This additional control might be caused by intensive contact of the enzymes with the primer–template stem, which are indicated by several crystal structures of DNA polymerases bound to the DNA substrate.[2]

The feature of DNA polymerases to transmit the presence of a single mismatch in the primer template into significantly less efficient DNA synthesis might be carried forward to new or improved methods for DNA diagnosis. Most of the known procedures for the detection of single nucleotide variations within genes are applied after PCR amplification of the target.[6] Reliable direct single nucleotide analysis (e.g., through allele-specific PCR) should supersede methods that depend on post-PCR analysis. We have recently shown that through employment of 4′-vinylated primer probes and Vent exo⁻ DNA polymerase, differentiation between a match versus single nucleotide mismatches located at the 3′-terminal primer position can be achieved.[7,8] Thus, matched primer–template duplexes are amplified with significantly higher efficiency compared to mismatched ones; this can lead to a conclusion about the sequence opposite the 3′-primer terminus. Based on our findings described herein, we asked whether remote mismatches within the primer–template complex are also discriminated with PCR. Such a desired ability would significantly extend the approach of allele-specific PCR. First, we conducted primer extension reactions under similar conditions applied in PCR reactions and compared unmodified with 4′-vinylated primer probes in their specificity. As shown in Figure 2, under the conditions applied, unmodified probes are extended without any significant discrimination between matched and single mismatched complexes while the introduction of a single 4′-vinyl group at the 3′-primer terminus causes a significant increase in discrimination.

Thus, even when a single mismatch is located at remote positions within the primer–template complex, extension selectivity was significantly enhanced through 4′-vinylated probes. These effects persist until position +4. Interestingly, the observed effects are independent of the mismatch sequence. Measured melting curves of unmodified and 4′-vinylated complexes containing mismatches at distal positions indicate that the modification has no significant impact on thermal stability (results not shown). Thus, the observed effects are likely to be caused by differential enzyme–substrate interactions. These results support the notion that enzyme conformational changes occur globally, as does checking for correct Watson–Crick base pairing in the primer–template complex.[1] We believe that the observed effects of the modified substrates originate from these global enzyme transitions, which are triggered by mismatches at remote positions. These changes are monitored by the added bulk at the 4′-vinylated primer terminus, which decreases the propensity of extension of the geometrically altered, that is, mismatched primer–template, by the enzyme.

Next, we investigated whether we are able to exploit the observed features in an allele-specific PCR system.[9] We studied

  

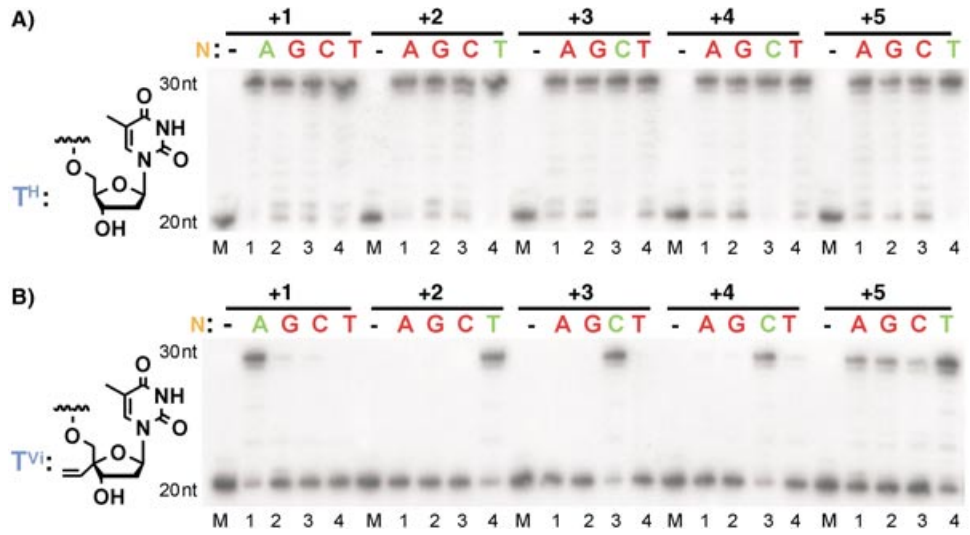


***Figure 2.*** *Primer extension by using A) unmodified or B) 4′-vinylated thymidine moieties at the 3′-terminal position of the primer. Each reaction contained equal amounts of Vent exo⁻ DNA polymerase and the respective primer–template as indicated. Sequences are shown in Figure 1, N = A, G, C, or T, as indicated. Conditions: All four dNTPs [0.2 mM each], and Vent exo⁻ DNA polymerase [0.8 units]. M = marker. Green indicates matched cases, red indicates mismatched cases.*

PCR product formation in real-time by SYBR Green I detection of double-stranded DNA.[10] In cases were unmodified primers were used no significant discrimination was observed and the formation of nearly congruent amplification curves were detected (Figure 3).

However, under the same reaction conditions significant differences between the amplification of matched and mismatched complexes were observed when 4′-vinylated primers were used. In these cases, any mismatch located at positions +1 to +4 resulted in an amplification curve with a threshold crossing cycle number significantly higher than observed for matched complexes. Thus, the system composed out of 4′-vinylated primers and Vent *exo*-DNA polymerase is able to faithfully detect single nucleotide variations even when they are located distal to the 3′-primer terminus. This is a hitherto unrecognized DNA polymerase ability with significant impact on the development of highly allele-specific PCR systems needed for the direct diagnosis of single nucleotide variations within genes, like mutations or polymorphisms.

Taken together, our results show that DNA polymerases detect mismatches within the primer–template complex up to four positions distal to the catalytic center. This process is believed to significantly add to the overall selectivity of enzymatic DNA synthesis. The global enzyme transitions, which are thought to be responsible for the observed effects, can be monitored by 4′-modified primer strands that act as steric probes. Based on these findings we were able to significantly extend methodology that allows direct analysis of single nucleotide variations within genes through allele-specific PCR.

## Experimental Section

**DNA substrates:** Unmodified double-grade HPLC purified DNA substrates, which were used as templates, were purchased from IBA GmbH, Göttingen, Germany. The synthesis of modified primer strands was carried out as described.[7b] Primer DNA strands were subsequently purified by preparative electrophoresis on a 12% polyacrylamide gel that contained 8 M urea. The integrity of all modified oligonucleotides was confirmed by matrix-assisted laser

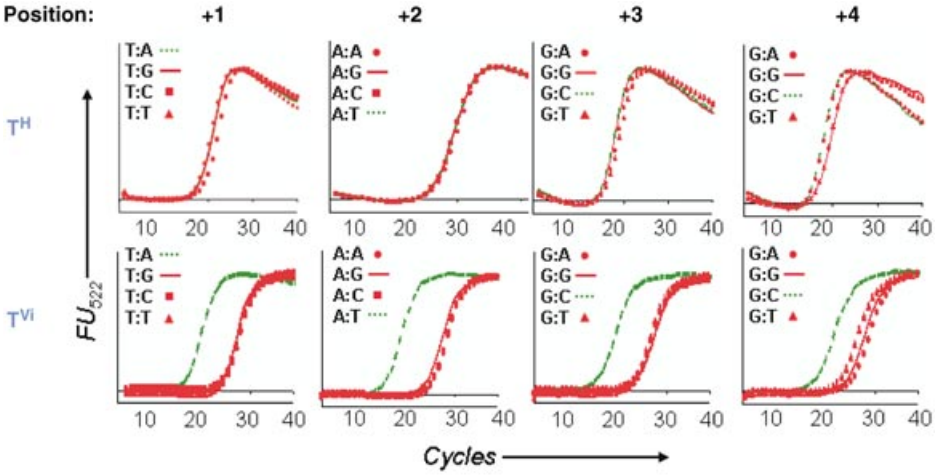


***Figure 3.*** *Real-time PCR experiments derived from unmodified ($T^H$) and 4′-vinylated ($T^{Vi}$) primers and mismatches located at several positions within the primer template complex (+1, +2, +3 or +4 nucleotides away from the reaction center).[11] Green indicates matched cases, red indicates mismatched cases. All experiments were conducted under identical reaction conditions and contained equal amounts of dNTPs, DNA substrate, and Vent exo⁻ DNA polymerase.*

  

desorption ionization-time of flight mass spectrometry (MALDI-TOF MS).

**DNA oligonucleotide sequences:** Reverse primer for PCR experiments: 5′-CG CGC AGC ACG CGC CGC CGT. Templates for PCR investigations: 90+1: 5′-CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA CCT *N*TC CTC CTT CAG CAA CGT ACA GAG; 90+2: 5′-CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA CCT A*N*C CTC CTT GAC CAA CGT ACA GAG; 90+3: 5′-CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA CCT AT*N* CTC CTT CAG GAC CAA CGT ACA GAG; 90+4: 5′-CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA CCT ATC *N*TC CTT CAG GAC CAA CGT ACA GAG, *N* = A, G, C, or T.

**Primer extension assays:** Primer extension assays were conducted as described before.[7] Reactions were initiated by the addition of a polymerase solution in 1x reaction buffer and heating to 72 °C. Reactions promoted by the Vent DNA polymerase (*exo*- mutant; New England Biolabs) were performed in New England Biolabs "ThermoPol" buffer (10 mM KCl, 20 mM Tris-HCl (pH 8.8 at 25 °C), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100). Reactions that were carried out with Taq DNA polymerase (Amplitaq, Applied Biosystems) were performed in Applied Biosystems "10x PCR Buffer" buffer (10 mM Tris-HCl (pH 8.3 at 25 °C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (*w*/*v*) gelatin). Assays included primer–template complex (150 nM) and the respective enzyme (0.25 units; units defined by the supplier). After incubation for 1 min the reaction mixtures were cooled to 0 °C and the reaction was quenched by the addition of PAGE-gel loading buffer (60 μL, 80% formamide, 20 mM EDTA) and subsequently heated to 95 °C for 10 min and analyzed as described.[7]

**Real-time PCR experiments:** Real-time PCR was performed by using ABI PRISM 7700 or iCycler (BIORAD) systems together with the DNA primer–templates mentioned before. The presented results are from at least two independently performed measurements that originated from one master-mix as described.[7]

## *Acknowledgements*

*We kindly acknowledge the financial support provided by the DFG, VolkswagenStiftung, Roche Diagnostics, and Fonds der Chemischen Industrie.*

**Keywords:** DNA · DNA polymerase · DNA replication · oligonucleotides · polymerase chain reaction

[1] Reviews: a) H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477; b) M. F. Goodman, S. Creighton, L. B. Bloom, J. Petruska, *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83; c) K. A. Johnson, *Annu. Rev. Biochem.* **1993**, *62*, 685; d) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497; e) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191.

[2] a) S. J. Johnson, J. S. Taylor, L. S. Beese, *Proc. Natl. Acad. USA* **2003**, *100*, 3895; b) H. Ling, F. Boudsocq, R. Woodgate, W. Yang, *Cell* **2001**, *107*, 91; c) M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **2001**, *98*, 413; d) S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *391*, 251; e) Y. Li, S. Korolev, G. Waksman, *EMBO J.* **1998**, *17*, 7514; f) J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304; g) H. F. Huang, R. Chopra, G. L. Verdine, S. C. Harrison, *Science* **1998**, *282*, 1669.

[3] a) J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323; b) J. C. Morales, E. T. Kool, *Biochemistry* **2000**, *39*, 12979; c) T. E. Spratt, *Biochemistry*, **2001**, *40*, 2647; d) M. T. Washington, W. T. Wolfle, T. E. Spratt, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5113.

[4] a) T. E. Carver Jr., R. A. Hochstrasser, D. P. Millar, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10670; b) E. H. Z. Thompson, M. F. Bailey, E. J. C. van der Schans, C. M. Joyce, D. P. Millar, *Biochemistry* **2002**, *41*, 713.

[5] a) M. Delarue, O. Poch, N. Tordo, D. Moras, P. Argos, *Prot. Eng.* **1990**, *3*, 461; b) H. Kong, R. B. Kucera, W. E. Jack, *J. Biol. Chem.* **1993**, *268*, 1965.

[6] Recent reviews: a) P. Y. Kwok, *Annu. Rev. Genom. Hum. Genet.* **2001**, *2*, 235; b) B. W. Kirk, M. Feinsod, R. Favis, R. M. Kliman, F. Barany, *Nucleic Acids Res.* **2002**, *30*, 3295.

[7] a) M. Strerath, A. Marx, *Angew. Chem.* **2002**, *114*, 4961; *Angew. Chem. Int. Ed.* **2002**, *41*, 4766; b) M. Strerath, J. Gaster, D. Summerer, A. Marx, *ChemBioChem* **2004**, *5*, 333.

[8] Further modifications used: a) D. Latorra, K. Campbell, A. Wolter, J. M. Hurley, *Hum. Mutat.* **2003**, *22*, 79; b) B. Tews, J. Wilhelm, D. Summerer, M. Strerath, A. Marx, P. Friedhoff, A. Pingoud, M. Hahn, *Biol. Chem.* **2003**, *384*, 1533.

[9] C. R. Newton, A. Graham, L. E. Heptinsall, S. J. Powell, K. Kalsheker, J. C. Smith, A. F. Markham, *Nucleic Acids Res.* **1989**, *17*, 2503.

[10] Review: J. Wilhelm, A. Pingoud, *ChemBioChem* **2003**, *4*, 1120.

[11] Primer and template +1 sequences are in the context of human acid ceramidase comprising the mutation A107. See: J. Bär, T. Linke, K. Ferlinz, U. Neumann, E. H. Schuchman, K. Sandhoff, *Hum. Mut.* **2001**, *17*, 199.

Received: April 15, 2004

  