

Enhanced Fidelity in Mismatch Extension by DNA Polymerase through Directed Combinatorial Enzyme Design**

Daniel Summerer, Nicolas Z. Rudinger, Ilka Detmer, and Andreas Marx*

Dedicated to Professor Bernd Giese on the occasion of his 65th birthday

The fidelity of DNA polymerase activity is of central importance for numerous biotechnological applications.^[1,2] The imperfect fidelity of DNA polymerases under the unnatural conditions of several techniques such as those of the polymerase chain reaction (PCR) either restricts the application of these enzymes or demands their tedious optimization. Thus, a prime target for the design of DNA polymerases with altered functions is high fidelity in the formation of Watson–Crick base pairs during DNA synthesis. Progress in this area generates valuable tools for many biological applications like PCR, sequencing protocols, mutagenesis techniques, and genotyping. Through the engineering of DNA and RNA polymerases, it has been possible to design

enzymes that accept dideoxynucleoside triphosphates,^[3] exhibit an altered substrate spectrum and lower fidelity, higher thermostability, decreased activity at low temperatures, and enhanced resistance to inhibitors.^[4]

Herein, we describe an efficient automated high-throughput setup for the rapid parallel screening of DNA polymerase mutant libraries. The readout is based exclusively on enzyme activity through detection of the reaction product by fluorescence. By the randomization of a gene cassette of the Klenow fragment of *E. coli* DNA polymerase I (3'→5'-exonuclease deficient, KF⁻) and subsequent comparative automated screening, several active variants with significantly higher extension fidelity than the wild-type enzyme were identified. The new properties of mutated KF⁻ forms were transferred to the thermostable *Thermus aquaticus* (Taq) DNA polymerase to provide tools that supersede the wild-type enzyme in highly accurate PCR-based genotyping techniques.

Motif C was recently suggested to be involved in a reaction mechanism shared by members of the A and B families of DNA polymerases, in which mismatches in the primer–template substrate are recognized through indirect H-bonding between the minor groove and a histidine side chain (Figure 1).^[5] This β-strand–turn–β-strand structure harbors acidic side chains that bind catalytically essential magnesium ions, reflected in the high conservation among family A DNA polymerases like KF⁻, and also among sequence families B, RT, X, single subunit RNA polymerases, and the lesion bypass DNA polymerases of family Y.^[6] Therefore, motif C seemed a promising target for directed polymerase engineering through focused randomization and subsequent screening.

To modify motif C, we constructed a library of 1316 KF⁻ mutants randomized at the consensus residues 879–881 (Q879, V880, and H881).^[7] This QVH consensus sequence is directly adjacent to the essential catalytic carboxylate group of D882 and forms the main part of a loop that connects the two β strands of motif C. It closely interacts with the deoxyribose moiety of the 3'-terminal primer nucleotide (Figure 1b).

Protein expression was conducted in 96-well plates. Enzymes were screened directly after lysis and dilution, and further purification steps were not required. DNA polymerase activity was monitored after reaction termination by quantitation of synthesized double stranded DNA through staining with Sybr green I (Figure 1c). This setup links enzyme activity to a signal without the need for artificial substrates, which could interfere with the enzymatic reaction. To determine if DNA binding and modifying agents or bacterial DNA in the crude lysate perturbs signal generation, we assayed lysates of *E. coli* expression cultures containing the KF⁻ wild-type coding vector against those without the KF⁻ coding vector. Fluorescence measurements revealed low background activity of the negative control with a considerable signal-to-noise ratio of ≈5:1. This enabled the assessment of a large dynamic range in enzyme activity.

We assayed single clones in 384-well plates for the ability to extend primer–template complexes with either a T_{primer}/A_{template} match or a T/G mismatch at the 3' end of the primer (Figure 1). An automated pipetting device and the fluorescent readout allowed the screening and evaluation of 384

[*] Dr. D. Summerer, Dipl.-Biol. N. Z. Rudinger, Dipl.-Chem. I. Detmer, Prof. Dr. A. Marx
Fachbereich Chemie, Universität Konstanz
Universitätsstrasse 10, M 726, 78457 Konstanz (Germany)
Fax: (+49) 7531-88-5140
E-mail: andreas.marx@uni-konstanz.de

[**] We thank the Volkswagen Foundation for financial support and M. Strerath for his support in the preparation of the manuscript.

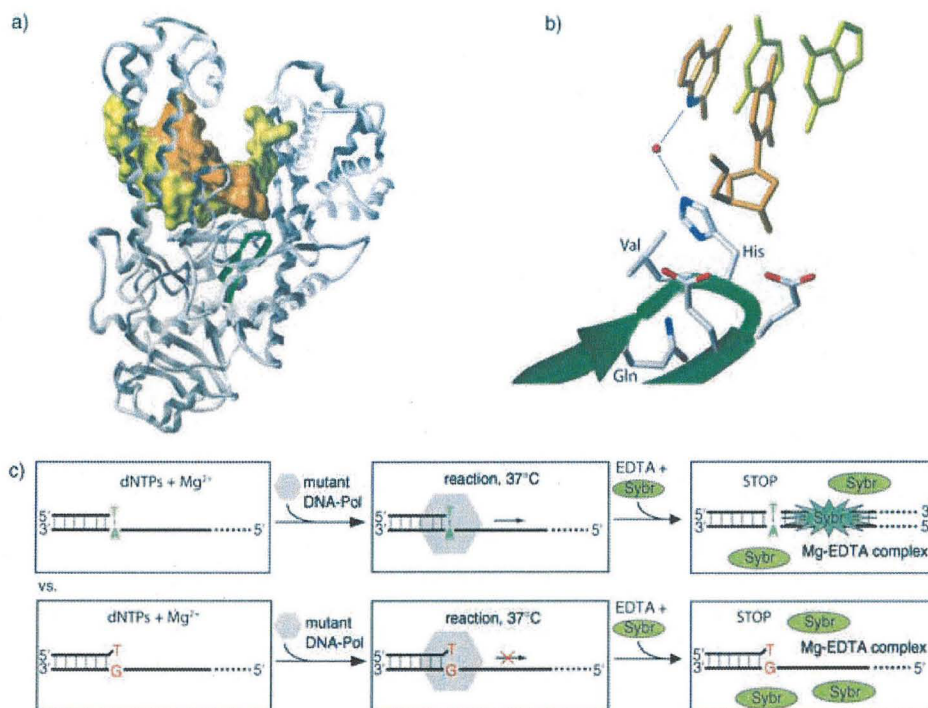


Figure 1. Structural properties of motif C and screening for DNA polymerase variants with increased selectivity. a) Overall view of *Bst* DNA polymerase I (PDB entry 2BDP), which shares high homology with KF^- .^[16] The enzyme is shown in gray, motif C in green. The substrate DNA is depicted as a Connolly surface with primer in orange, and template strand in yellow. b) Detailed view of the turn of motif C and the two nucleotide pairs at the 3' terminus of the primer. For clarity, the sugar-phosphate groups are not depicted for each nucleotide. Primer nucleotides are in orange, template nucleotides in yellow. The indirect hydrogen bond between N3 of histidine (blue) and the 3'-penultimate primer nucleotide is shown in black and is mediated by a water molecule (red sphere). c) General scheme of the screening approach for DNA polymerases with increased extension fidelity. Enzyme variants that extend the canonical primer-template duplex but fail to extend the noncanonical duplex are identified with the DNA duplex-specific dye Sybr green I in a subsequent fluorescence read-out.

reactions in approximately 40 minutes. Extension fidelity was measured by determining the ratios of fluorescence of primer extensions from matched versus mismatched substrates. The template sequences were derived from the human single nucleotide polymorphism (SNP) G1691A of the Factor V Leiden (FVL) gene.^[8] The screen revealed considerable mutability of the targeted QVH sequence. Despite the fact that this motif is highly conserved among several DNA and RNA polymerase families,^[6] 47% of the mutants exhibited measurable primer extension activity.

The three most selective mutants, PLQ, LVG, and LVL, were chosen for further characterization. First, we assayed the purified enzymes in radiometric primer extensions (Figure 2). All possible base-pair combinations at the primer 3' terminus were tested under conditions that promote mismatch extension: excess enzyme over primer-template complex and high dNTP concentrations.^[7] Measurements revealed that wild-type KF^- is capable of extending almost all mismatches under the chosen conditions, albeit to varying extents consistent with previous studies (Figure 2b).^[9] In contrast, all three selected mutants show a marked decrease in the efficiency of mismatch extension. Particularly, mutant LVL fails to fully extend mismatched primer termini in most cases. Steady-state kinetics measurements of single-nucleotide extensions show that in all cases of processing a properly matched primer-

template complex, mutants display similar steady-state k_{cat} values as the wild-type enzyme.^[7] K_M values of the mutants were generally higher than those for wild-type KF^- , reflected by a slight decrease in DNA synthesis efficiency by the mutant forms. Remarkably, no significant elongation of mismatched primer termini by the mutants was detected under steady-state and single-turnover conditions, whereas the results obtained for wild-type KF^- are consistent with previous reports.^[9] This indicates that the k_{cat} values for all base-pair combinations assayed were drastically lower for the KF^- variants than for the wild-type polymerase. Binding studies were performed to determine whether the effects on mismatch extension fidelity are caused by a decrease in substrate-binding affinity in the mutants.^[7] Interestingly, only mutant PLQ exhibits a significant decrease in binding affinity to a mismatched primer terminus. For this reason, altered binding affinity does not generally appear to be mandatory for an increase in mismatch extension fidelity.

We next investigated whether the observed effects could be transferred to wild-type *Taq* DNA polymerase (*Taq* wt). Though *Taq* wt and KF^- are both members of the DNA polymerase family A, they differ in several properties such as thermostability and fidelity.^[10] The *Taq* QVH consensus sequence was mutated into LVL and the extension fidelity of the resulting mutant polymerase was investigated. We

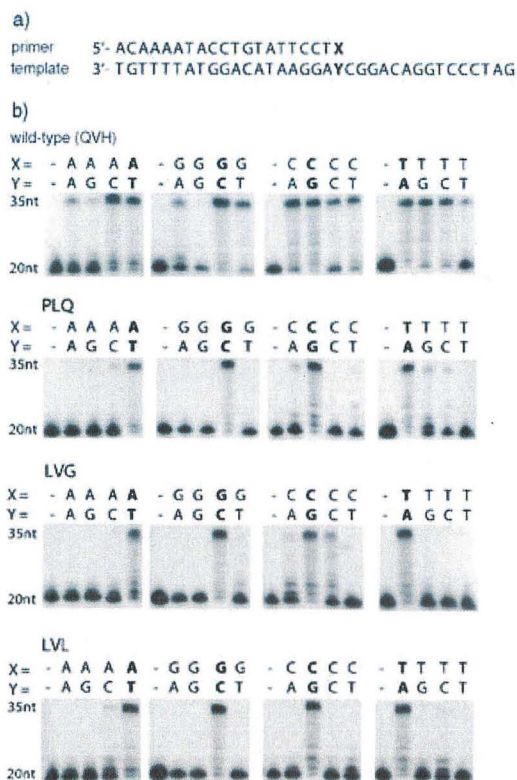


Figure 2. Increased primer extension fidelity of the KF⁻ mutants PLQ, LVL, and LVG (positions 879–881) in comparison with the wild-type enzyme in different sequence contexts. The conditions were chosen to promote mismatch extension. All reactions contained equal amounts of primer–template complex, enzyme, and dNTPs.^[7] a) The sequence of the primer–template duplex employed is derived from the human FVL SNP G1691A.^[7] b) Primer extensions catalyzed by the depicted enzyme variants. The first lane of each gel represents a control reaction without enzyme. Nucleotide sequences at the 3' end of the primer are shown on top of each gel image. Unextended primer (20 nt) and full-length product (35 nt) are indicated at the left side of each image. X = 3'-terminal primer nucleotide, Y = template nucleotide pairing with X leading to matched (bold) or mismatched primer termini.^[7]

conducted primer extension reactions in three sequence contexts that contain prominent SNPs.^[11,12] These experiments show that there are only subtle differences between the extension of matches and mismatches with *Taq* wt under the chosen conditions (Figure 3).

In contrast, the mutant *Taq* DNA polymerase (*Taq* LVL) is clearly capable of discriminating transversion and transition SNPs (Figure 3). *Taq* LVL might greatly expand the technical scope and allow an improvement of PCR-based techniques like allele-specific PCR (asPCR).^[13] To test the applicability of *Taq* LVL for asPCR, we performed real-time PCR reactions with the substrate sequence contexts described above. We measured the difference in the threshold-crossing cycle number (ΔC_t) between a matched and a mismatched primer–template complex (Figure 3). *Taq* wt displays weak or no discrimination, whereas *Taq* LVL leads to ΔC_t values of 10 for all sequence contexts. Therefore, the mutant *Taq* LVL shows greater discrimination than *Taq* wt in asPCR, regardless of the sequence context applied or enzyme and template

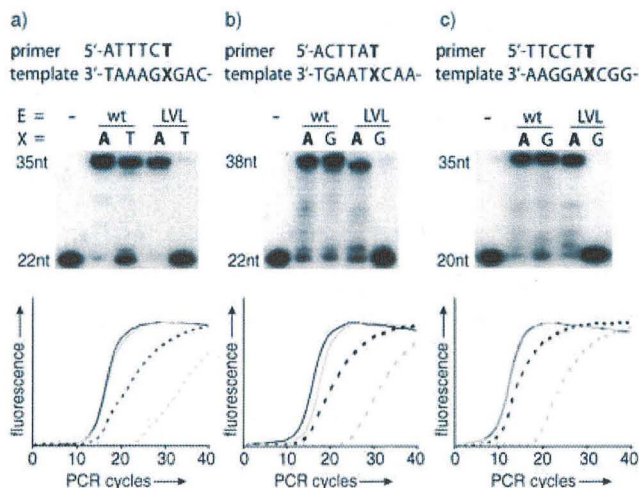


Figure 3. The impact of LVL mutations on the fidelity of *Taq* DNA polymerase within three sequence contexts. Top panel: partial nucleotide sequences of primer–template complexes used for primer extensions. The first lane of each gel represents a control reaction without enzyme. The lengths (nt) of primer and full-length product are shown on the left side of each gel; E = enzyme, X = template nucleotide. Lower panel: real-time allele-specific PCR experiments with either wild-type or mutant LVL *Taq* DNA polymerase in three sequence contexts.^[7] Solid lines: wild-type *Taq* DNA polymerase; dashed lines: LVL *Taq* DNA polymerase; black: matched primer–template substrates; gray: mismatched primer–template substrates. Experiments conducted in a) the human BRAF somatic SNP T1796A, b) in the human DPYD SNP G735A, and c) in the FVL SNP G1691A sequence context.^[7]

concentrations tested, as shown in additional experiments (Supporting Information).^[7]

In summary, the identification of several active DNA polymerase variants that display increased primer extension fidelity shows that none of the chosen amino acid positions is essential for catalysis, and that all positions tolerated substitutions. However, the nature of the side chain of V880 seems to be most conserved, as all identified active and more selective mutants bear a nonpolar amino acid side chain at this position. Furthermore, the absence of any charged residues within the selected active mutants indicates that the introduction of ionic residues in this region might inactivate the enzyme by interference with D882, the magnesium ions, or the incoming dNTP substrate.

The fact that the apparent increased fidelity is not limited to the sequence context employed during screening, is suggestive of a common mechanism for error sensing. It was recently proposed that the histidine group of QVH might be involved in DNA polymerase extension fidelity mechanisms by editing H-bonding patterns in the minor groove of the primer–template duplex (Figure 1b).^[5] One of the mutants that exhibits increased extension fidelity has three amino acid substitutions at the targeted site (QVH to PLQ), in which a glutamine side chain replaces histidine at position 881. As mentioned above, glutamine can often replace the H-bond donating ability of the imidazole ring of histidine.^[14] Thus, improved editing of H-bond patterns by Q881 in the context of the PLQ sequence might be the cause for the increased fidelity in this variant. However, we identified mutants LVG

and LVL, both of which lack hydrogen bonding capability within the wild-type QVH region, yet have significantly higher extension fidelity. Thus, abolishing hydrogen bonding to the minor groove results in enhanced polymerase fidelity. Similar results for another KF⁻ variant (H881A) were recently reported.^[15] The origins of the observed selectivity currently remain elusive. It could be that a loss of hydrogen bonding in a complex that is already destabilized (upon mismatch extension catalysis) actually facilitates the editing capacity over that of the wild-type polymerase, in which hydrogen bonding is present.

The approach of rapid DNA polymerase screening discussed herein, which has led to highly valuable enzyme variants, can be used to further tailor DNA polymerase activities for obtaining insight into biological processes and new tools for biotechnological applications.

Keywords: DNA polymerase · DNA recognition · molecular evolution · polymerase chain reaction

- [1] a) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529; b) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219.
- [2] J. Sambrook, D. W. Russell, *Molecular cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA **2001**.
- [3] Y. Li, V. Mitaxov, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9491–9496.
- [4] a) F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4552–4557; b) S. Brakmann, S. Grzeszik, *ChemBioChem* **2001**, *2*, 212–219; c) G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6597–6602; d) P. H. Patel, L. A. Loeb, *J. Biol. Chem.* **2000**, *275*, 40266–40272; e) M. B. Kermekchiev, A. Tzekov, W. M. Barnes, *Nucleic Acids Res.* **2003**, *31*, 6139–6147; f) M. Fa, A. Radeghieri, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 1748–1754; g) F. J. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate, P. Holliger, *Nat. Biotechnol.* **2004**, *22*, 755–759.
- [5] M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **2001**, *105*, 657–667.
- [6] M. Delarue, O. Poch, N. Tordo, D. Moras, P. Argos, *Protein Eng.* **1990**, *3*, 461–467.
- [7] Detailed experimental procedures are provided in the Supporting Information.
- [8] H. Engel, L. Zwang, H. van Vliet, J. J. Michiels, J. Stibbe, J. Lindemans, *Thromb. Haemostasis* **1996**, *75*, 267–269.
- [9] a) C. M. Joyce, X. C. Sun, N. D. Grindley, *J. Biol. Chem.* **1992**, *267*, 24485–24500; b) D. T. Minnick, K. Bebenek, W. P. Osheroff, R. M. Turner, Jr., M. Astatke, L. Liu, T. A. Kunkel, C. M. Joyce, *J. Biol. Chem.* **1999**, *274*, 3067–3075.
- [10] M. Suzuki, A. K. Avicola, L. Hood, L. A. Loeb, *J. Biol. Chem.* **1997**, *272*, 11228–11235.
- [11] H. Davies, G. R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M. J. Garnett, W. Bottomley, N. Davis, N. Dicks, R. Ewing, Y. Floyd, K. Gray, S. Hall, R. Hawes, J. Hughes, V. Kosmidou, A. Menzies, C. Mould, A. Parker, C. Stevens, S. Watt, S. Hooper, R. Wilson, H. Jayatilake, B. A. Gusterson, C. Cooper, J. Shipley, D. Hargrave, K. Pritchard-Jones, N. Maitland, G. Chenevix-Trench, G. J. Riggins, D. D. Bigner, G. Palmieri, A. Cossu, A. Flanagan, A. Nicholson,
- J. W. C. Ho, S. Y. Leung, S. T. Yuen, B. L. Weber, H. F. Siegler, T. L. Darrow, H. Paterson, R. Marais, C. J. Marshall, R. Wooster, M. R. Stratton, P. A. Futreal, *Nature* **2002**, *417*, 949–954.
- [12] X. X. Wei, H. L. McLeod, J. McMurrugh, F. J. Gonzalez, P. Fernandez Salguero, *J. Clin. Invest.* **1996**, *98*, 610–615.
- [13] a) A. C. Syvänen, *Nat. Rev. Genet.* **2001**, *2*, 930–942; b) M. M. Shi, *Clin. Chem.* **2001**, *47*, 164–172; c) M. V. Myakishev, Y. Khripin, S. Hu, D. H. Hamer, *Genome Res.* **2001**, *11*, 163–169.
- [14] A. Fersht in *Structure and Mechanism in Protein Science*, Freeman, New York, **1999**, p. 425.
- [15] D. T. Minnick, K. Bebenek, W. P. Osheroff, R. M. Turner, M. Astatke, L. Liu, T. A. Kunkel, C. M. Joyce, *J. Biol. Chem.* **1999**, *274*, 3067–3075.
- [16] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304–307.