Variants of a *Thermus aquaticus* DNA Polymerase with Increased Selectivity for Applications in Allele- and Methylation-Specific Amplification

Matthias Drum¹, Ramon Kranaster¹,², Christina Ewald¹, Rainer Blasczyk³, Andreas Marx¹*

¹ Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany, ² myPOLS Biotec, University of Konstanz, Konstanz, Germany, ³ Institute for Transfusion Medicine, Hannover Medical School, Hannover, Germany

Abstract

The selectivity of DNA polymerases is crucial for many applications. For example, high discrimination between the extension of matched versus mismatched primer termini is desired for the detection of a single nucleotide variation at a particular locus within the genome. Here we describe the generation of thermostable mutants of the large fragment of *Thermus aquaticus* DNA polymerase (*KlenTaq*) with increased mismatch extension selectivity. In contrast to previously reported much less active *KlenTaq* mutants with mismatch discrimination abilities, many of the herein discovered mutants show conserved wild-type-like high activities. We demonstrate for one mutant containing the single amino acid exchange R660V the suitability for application in allele-specific amplifications directly from whole blood without prior sample purification. Furthermore, the identified mutant supersedes other commercially available enzymes in human leukocyte antigen (HLA) analysis by sequence-specific primed polymerase chain reactions (PCRs).

Introduction

Personalized medicine providing therapies, adapted to each patient’s genetic predisposition[1][2][3][4][5], is mainly supported by the analysis of single nucleotide polymorphisms (SNPs)[1]. SNPs are single base variations and besides copy number variations the most abundant type of genetic variation found between members of one species[6][7][8][9]. SNPs located in coding sequences can lead to structural and functional changes in the affected proteins, enzymes or receptors. For example the prothrombin G20210A mutation is one of the most common genetic polymorphisms known to predispose to a first episode of venous thromboembolism[10]. Most SNPs, however, are found in non-coding intergenic regions and often show no phenotypic effect. Intergenic SNPs present interesting markers for the determination of parentage[11][12], anthropology[13][14] or forensic tasks e.g. genetic fingerprinting. Many of these variations can affect predispositions for diseases or responses to drugs, chemicals and vaccines[1][15], which makes them especially interesting for pharmacogenomics[16][17]. The human genome project greatly contributed to personalized medicine by identifying more than 2.4 million SNPs in 2001[18][9]. This created a basis for the first human haplotype map (HapMap) project with more than one million SNPs for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations[19]. In a second step an additional 2.1 million SNPs were added[20]. The next phase could further improve the quality with an extended set of 1,184 samples from 11 populations[21]. With the 1000 Genomes Project, a validated haplotype map of 38 million SNPs was published in 2012[22]. Genomes of 1,092 individuals sampled from 14 populations drawn from Europe, East Asia, sub-Saharan Africa and the Americas were analysed through a combination of low-coverage whole-genome sequence data, targeted deep exome sequence data and dense SNP genotype data[22]. It is most likely that SNP genotyping will be one of the future key technologies to diagnose these genetic variations among whole populations as well as in single patients. Different techniques can be used for the analysis of SNPs such as selective primer extension, e.g. minisequencing[23][24][25], pyrosequencing[26] or allele specific amplification (ASA)[27][23][28]. ASA and selective primer extension[23] depend on the inefficient extension of a mismatched primer/template complex. Therefore, highly selective DNA polymerases are urgently needed. Allele specific amplification through real-time PCR (ASA) allows detection of SNPs in a very efficient way. ASA, unlike most other methods for SNP detection, does not require preliminary amplification of the target genetic material[29][30]. It combines amplification and detection in a single reaction, based on the discrimination of matched and mismatched primer/template complexes. The increase of amplified DNA during the reaction can be monitored in real-time through the increase of a fluorescence signal caused by a dye such as SYBR Green I[31] emitting upon binding to double-stranded DNA. The match case comprises a correct Watson-Crick base pair at the 3’-primer end,
whereas the mismatch case features a non-canonical base pair. The mismatch should result in a less efficient or, ideally, no product amplification\[^{30}\]\[^{32}\]\[^{33}\]\[^{34}\]\[^{35}\]\[^{36}\]\[^{37}\]. In real-time ASA this is reflected with a delayed or absent fluorescence signal product amplification\[^{30}\]\[^{32}\]\[^{33}\]\[^{34}\]\[^{35}\]\[^{36}\]\[^{37}\]. In real-time The mismatch should result in a less efficient or, ideally, no replication fidelity are known\[^{54}\], e.g. for the Klenow fragment of E. coli DNA polymerase I and the thermostable Thermus aquaticus (Taq) DNA polymerase\[^{55}\] or the Pyrococcus furiosus (Pfu) DNA polymerase\[^{56}\]. It is known that the exchange of amino acids, affecting the interaction between polymerase and primer/template complex or the binding pocket’s geometry, can lead to a change in selectivity of DNA polymerases\[^{57}\]. In previous studies, an increase in Taq DNA polymerase selectivity was achieved by substituting the polar amino acid side chains in Q and H (Gln, and His) of motif C for leucine bearing a non-polar side chain. Motif C is highly conserved in the palm domain within DNA polymerase families A, B, X and Y and plays a role in the identification of mismatched bases in the primer/template complex\[^{58}\]\[^{59}\]. While discovered with Taq DNA polymerase, a member of the DNA polymerase sequence family A, in further studies\[^{60}\] this concept to increase mismatch extension selectivity could be transferred to the B family Pfu DNA polymerase. The respective amino acids are found in the highly conserved motifs YGEDTD and KXY in eukaryotes, bacteria, archaea, and viruses\[^{61}\]\[^{62}\].

Here we present a systematic study of the influence on the mismatch extension selectivity of all basic amino acids (arginine and lysine), which are in direct contact with the phosphate backbone of the primer strand in the closed conformation of the Klenow fragment of Taq DNA polymerase (KlenTaq). The chosen mutation sites vary in distance from the enzyme’s active center. We identified several discriminative active mutants that were functionally characterized and evaluated in ASA, HLA typing and MSP.

### Materials and Methods

**Reagents and Instruments**

Oligonucleotides were purchased from Biomers or Metabion, HeLa genomic DNA and Taq 2x master mix was bought from New England Biolabs, dNTPs were either from Roche or Fermentas, Phusion DNA polymerase was purchased from Thermo Scientific, Platinum Taq and AmpliTaq Gold DNA polymerases from Life Technologies, DNase I, SphI, and HindIII from Fermentas, the Gel Extraction and EpiTect MSP kit from Qiagen and used according to their manuals. KlenTaq and its respective mutants were recombinantly expressed in E. coli BL21 (DE3) and purified with Ni-IDA as previously described\[^{63}\].

Enzyme purity and quantity were determined by SDS-PAGE using an albumin standard dilution curve. KlenTaq variants were stored in 50 mM Tris-HCl (pH 9.2), 16 mM (NH\(_4\))\(_2\)SO\(_4\), 0.1% Tween20, 2.5 mM MgCl\(_2\), 50% glycerol at \(-20^\circ\)C. For real-time PCR a Chromo4 instrument from Bio-Rad or a Roche LightCycler 96 or 480 system was used. SYBR green I was purchased from Molecular Probes or Roche and used according to their manuals.

**KlenTaq** and its wild-type DNA polymers were amplified by PCR using Phusion DNA polymerase in the supplied buffer according to the manufacturer’s specifications. For saturation mutagenesis either NNK (forward) or MNN (reverse) primers were used. One of the following randomization primers was used together with the respective KlenTaq primers to create a mismatched primer/template complex: SphI-for [200 nM, 34 nt, 5'-d(CAT ACG GAT CCG CAT GCA-3')] and SphI-rev [200 nM, 36 nt, 5'-d(GCT CAG CTA ATT AAG CTT TCT CCT TGG CGG AGA GCC)-3'].

### Saturation mutagenesis and library construction

**KlenTaq** wild-type DNA genes were amplified by PCR using Phusion DNA polymerase in the supplied buffer according to the manufacturer’s specifications. For saturation mutagenesis either NNK (forward) or MNN (reverse) primers were used. One of the following randomization primers was used together with the respective KlenTaq primers to create a mismatched primer/template complex: SphI-for [200 nM, 34 nt, 5'-d(CAT ACG GAT CCG CAT GCA-3')] and SphI-rev [200 nM, 36 nt, 5'-d(GCT CAG CTA ATT AAG CTT TCT CCT TGG CGG AGA GCC)-3'].
CCT GCC CGG CCA CCC CTT CAA CCT CCA CTC CNN KGA CCA CTT GGA AAG GGT CCT C) -3', K508deg [200 nM, 45 nt, 5'-d(CCG CCA TCG GCA AGA AGN NKA CGG CCA AGC CCT CCA CCG GCG) -3'], R536deg [200 nM, 50 nt, 5'-d(GTG GAG AAT CAG AGC ACC TA)-3'], R387deg [200 nM, 50 nt, 5'-d(CCC AAT CTC GAG AGT CTA GGC ATG ACT GCA GAA AAT CAA GCC GT)-3'] and F20-rev [750 nM, 20 nt, 5'-d(CGCG GCA CCG AGC CGC CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA GCC GT)-3'].

The target PCR product was purified by gel electrophoresis and used as a primer in a second PCR with either the SphII or HindIII-rev primer to generate the KlenTag full-length product. The amplified products were purified by gel electrophoresis, digested with the respective restriction enzymes SphII and HindIII and cloned into a suitable expression vector. The ligation reaction was transformed into E. coli BL21 (DE3) cells and colonies were picked randomly.

Screening and real-time ASA assay with KlenTag wild-type and mutants (lysate or purified enzyme)

Reaction mixtures contained 50 mM Tris-HCl (pH 9.2), 16 mM (NH₄)₂SO₄, 0.1% Tween20, 2.5 mM MgCl₂, 250 μM of each dNTP and 0.6x SYBR Green I. As primer F20-for CCT CGC CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA GCC GT-3' and F20-rev [750 nM, 20 nt, 5'-d(CCG CCA TCG GCA AGA AGT CTA GGC ATG ACT GCA GAA AAT CAA GCC GT)-3'] were used. As templates, either F90A [60 pM, 90 nt, 5'-d(CCG GCA GCA CGG GCC GCC GCT)-3'] or F90G [60 pM, 90 nt, 5'-d(CGCG GCA GCA CGG GCC GCC GCT)-3'] were used. As templates, either F90A [60 pM, 90 nt, 5'-d(CCG GCA GCA CGG GCC GCC GCT)-3'] or F90G [60 pM, 90 nt, 5'-d(CGCG GCA GCA CGG GCC GCC GCT)-3'] were used. Both templates only differ by the SNP C was applied and the reaction was started by the addition of 200 nM dATP. After incubation for specified times the reactions were either stopped manually or with a KinTek Rapid Quench-Flow instrument by the addition of 500 mM EDTA. Product mixtures were separated by 12% denaturing PAGE and visualised using phosphorimaging. Quantification was done using Quantity One software from Biorad.

Reactions kinetics of KlenTag wild-type and mutants

Reaction mixtures contained 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.0 mM MgCl₂, 1 mM DNA polymerase wild-type or mutant, DNA primer [100 nM, 20 nt, 5'-d(CGT TGG TCC TGA AGG AGG AT)-3'], template F33A [130 nM, 33 nt, 5'-d(AAA TCA AGC TAT CCT CCT TGA GCA CCA AGC TAG)-3'] or F33G [130 nM, 33 nt, 5'-d(AAA TCA AGC TAT CCT CCT TGA GCA CCA AGC TAG)-3'] for the match or mismatch case, respectively. After an initial denaturation and annealing step (95°C for 2 min, 0.5°C/s cooling to 40°C for 30 sec), a temperature of 37.0°C was applied and the reaction was started by the addition of 600 μM dATP. After incubation for specified times the reactions were either stopped manually or with a KinTek Rapid Quench-Flow instrument by the addition of 500 mM EDTA. Product mixtures were separated by 12% denaturing PAGE and visualised using phosphorimaging.

ASA assay with KlenTag variants using human genomic DNA (gDNA)

Reaction mixtures (20 μL) contained 50 mM Tris-HCl (pH 9.2), 16 mM (NH₄)₂SO₄, 0.1% Tween20, 2.5 mM MgCl₂, 220 μM of each dNTP, 0.6x SYBR green I, 5% DMSO, 20 ng HeLa gDNA, and 100 nM of the respective KlenTag variant. As a forward primer either F2forG [100 nM, 23 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] or F2forA [100 nM, 23 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] was used in conjunction with primer F2rev [100 nM, 20 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] or F2rev [100 nM, 20 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] was used in conjunction with primer F2rev [100 nM, 20 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3']. Real-time PCR was conducted by an initial denaturation cycle (95°C for 3 min), followed by 50 PCR cycles (95°C for 20 s, 57°C for 10 s and 72°C for 30 s) and analysis of an amplicon size of 64 base-pairs by melting curve measurement.

ASA assay with KlenTag R660V mutant and commercially available DNA polymerases using human gDNA allele-standards

Reaction mixtures (10 μL) contained 50 mM Tris-HCl (pH 9.2), 16 mM (NH₄)₂SO₄, 0.1% Tween20, 2.0 mM MgCl₂, 200 μM of each dNTP, 1x SYBR green I, 10 ng gDNA allele-standard and 100 nM KlenTag R660V. As forward primer F2forG [100 nM, 23 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] or F2forA [100 nM, 23 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] was used in conjunction with reverse primer F2rev [100 nM, 20 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3'] or F2rev [100 nM, 20 nt, 5'-d(CGCG CAA CTT TGG TCC TGA AGG AGG AT)-3']. Taq 2X Master Mix (New England Biolabs) and Platinum Taq (Life Technologies) were used as described by the manufacturer. In case of the Platinum Taq, the reaction contained the 0.5 U enzyme and 1.5 mM MgCl₂. Templates for homozygote, heterozygote and wild type were used. After an initial denaturation cycle (95°C for 3 min), the product was amplified by 60 PCR cycles (95°C for 20 s, 57°C for 10 s and 72°C for 30 s) and analyzed by melting curve measurement. For endpoint PCRs, the cycling was stopped after 50 cycles and the PCR product immediately analysed by 2.5% agarose gel electrophoresis.

Multiplexing ASA assay with R660V from gDNA allele-standards and blood

SNP Factor II G20210A (respective amplicon size is 64 bp for the A-allele or 94 bp for the G-allele) was detected by direct amplification from 10 ng of genomic reference DNA using both allele-specific primers in the same reaction: F2forG [100 nM,

PLOS ONE | www.plosone.org 3 May 2014 | Volume 9 | Issue 5 | e96640
55 nt, 5′-d(ATC CAA CTC TCT ACG CAA TGG CAC TAG AGA CCC AAT AAA AGT GAC TCT CAG G\(-3\))', F2forA [125 mM, 23 nt, 5′-d(CCC AAT AAT AGG AAC GCC GAG TGG GGG G)\(-3\)] and reverse primer F2rev [100 mM, 20 nt, 5′-d(GGA GAG AGC TGG CCA TGA AT)\(-3\)].

Results and Discussion

For our studies we employed *KlenTaq* [65]. Compared to Taq DNA polymerase, *KlenTaq* is approximately twice as thermostable and only displays half the error rate of *Taq* DNA polymerase[65][66][67]. Structural data of *KlenTaq* shows that in the closed conformation five basic amino acids make contact with the primer strand[68]. Those are K508, R487, R356, R587 and R660 (see Figure 1). In order to study the impact of these basic amino acids on selectivity and activity we used saturation mutagenesis at these positions. Using degenerate NKX-primers for each amino acid position, a library consisting of at least 320 mutants was compiled and screened, guaranteeing a 99.9% coverage of all 20 possible amino acids under ideal mutagenesis conditions[69].

Screening was conducted as described previously in a high-throughput format using heat-treated *E. coli* lysates[70]. Our screening strategy uses 384 well plate libraries and monitors DNA polymerase activity by *SYBR green* I mediated quantification of synthesized double stranded DNA and its subsequent melting temperature assessment. Each mutant was hence tested in parallel and real-time for: a) thermostability for sufficient PCR activity, b) specificity by producing the correct amplicon length and c) their 3′ end mismatch extension selectivity. One reaction contained a matched primer template duplex whereas in another reaction a single mismatch at the 3′-terminus of the primer template complex was induced through exchange of one nucleotide in the template strand (A into G). Entities with increased extension selectivity were identified as those variants causing amplification curves with lower threshold crossing (c(t)) cycle numbers with matched DNA substrates than with mismatched DNA substrates. Depending on the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants.

Real-time MSP assay with *KlenTaq* R660V mutant

Reaction mixtures contained 50 mM Tris-HCl (pH 9.2), 16 mM (NH₄)₂SO₄, 0.1% Tween20, 2.0 mM MgCl₂, 200 μM of each dNTP, 1x SYBR green I and 50 nM *KlenTaq* R660V. As forward primer MSP-Forw [100 mM, 30 nt, 5′-d(GCC CGA TTC GTT GTT TAT TAG TTA TTA TGT)-3′] and reverse primer MSP-RevG [100 mM, 31 nt, 5′-d(TGG AAC TCC GAA ATA ATC GCA CCG ACC AAT CAC GAC G)-3′] were used in conjunction with reverse primers MSP-Rev-G [100 mM, 31 nt, 5′-d(TCA AAA TCC GAA ATA ATC GCA CCG ACC AAT CAC GAC G)-3′].

Selective *KlenTaq* Mutants

For our studies we employed *KlenTaq* [65]. Compared to Taq DNA polymerase, *KlenTaq* is approximately twice as thermostable and only displays half the error rate of *Taq* DNA polymerase[65][66][67]. Structural data of *KlenTaq* shows that in the closed conformation five basic amino acids make contact with the primer strand[68]. Those are K508, R487, R356, R587 and R660 (see Figure 1). In order to study the impact of these basic amino acids on selectivity and activity we used saturation mutagenesis at these positions. Using degenerate NKX-primers for each amino acid position, a library consisting of at least 320 mutants was compiled and screened, guaranteeing a 99.9% coverage of all 20 possible amino acids under ideal mutagenesis conditions[69].

Screening was conducted as described previously in a high-throughput format using heat-treated *E. coli* lysates[70]. Our screening strategy uses 384 well plate libraries and monitors DNA polymerase activity by *SYBR green* I mediated quantification of synthesized double stranded DNA and its subsequent melting temperature assessment. Each mutant was hence tested in parallel and real-time for: a) thermostability for sufficient PCR activity, b) specificity by producing the correct amplicon length and c) their 3′ end mismatch extension selectivity. One reaction contained a matched primer template duplex whereas in another reaction a single mismatch at the 3′-terminus of the primer template complex was induced through exchange of one nucleotide in the template strand (A into G). Entities with increased extension selectivity were identified as those variants causing amplification curves with lower threshold crossing (c(t)) cycle numbers with matched DNA substrates than with mismatched DNA substrates. Depending on the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants. When mutating the position of the respective arginine or lysine residue, mutagenesis led to varying amounts of PCR inactive mutants.
conditions used in the screening, employment of all selected mutants leads to significantly increased mismatch extension selectivity. The most promising examples for each position were sequenced. In only three cases among the identified hits was the basic amino acid arginine exchanged to another basic amino acid such as histidine (R487H) or lysine (R536K and R587K). Two variants with mutations to tyrosine and threonine were found (K508Y and R660T). In several cases substitutions towards hydrophobic amino acids were identified: R487V, K508W, R587I, and R660V. In fact, these results are in agreement with our earlier findings showing that mutations towards hydrophobic amino acids can increase mismatch extension selectivity [60][65][71].

In order to directly compare activity and mismatch extension selectivity of the selected mutants we performed primer extension experiments next. Therefore we used a 20 nucleotides (nt) radioactively labelled primer and two 33 nt templates having the same sequence except for one specific position (Figure 2A). Depending on the template, the 3'-end of the primer was either matched or mismatched. Reactions were incubated isothermally at 55°C and started by addition of dNTPs and quenched after 10 min, followed by PAGE analysis. As depicted in Figure 2 the KlenTaq wild-type shows full-length product formation and, as reported before for 3'-5' exonuclease-deficient polymerases, a non-templated additional nucleotide incorporation for the match and the mismatch case [72][73][74]. Notably not all primer is converted for the mismatch case indicating the intrinsic mismatch extension discrimination. None of the mutants show product formation for the mismatch case after 10 min, but yield full-length product for the match case. Whereas some mutants like R487V or K508Y clearly show reduced activity, which is detected by shorter reaction products, other mutants such as R587K or R660V show clean full-length product formation in the match case indicating a conserved high activity comparable to the wild-type enzyme.

As mutations at position R487 and K508 reduced the activity significantly (see Figure 2), they were excluded from further analyses. To characterise the identified mutants in more detail, we investigated single nucleotide incorporation turnover rates (under...
pre-steady state conditions) for extension from matched and mismatched primer strands. High concentrations of enzyme were used to saturate binding of the DNA substrate. At high dNTP concentrations (600 \text{ M}) the first order rate reflects $k_{\text{cat}}$\[^{75}\]. The results are shown in Figure 3 and Table 1. Notably, reaction times for the mismatch cases were up to 600 s whereas for the match cases full conversion was detected after 1 s for all polymerases.

A measure for the selectivity is $k_A/k_G$ as shown in Figure 3 and Table 1. This ratio is significantly increased in the selected mutants, compared to the wild-type enzyme, given as $(k_A/k_G)/\text{wt}$. For instance, the discrimination of the matched vs. mismatched primer was increased by up to 36-fold in mutant R487V compared to KlenTaq wild-type.

Attempting to further improve the mismatch discrimination selectivity, we applied a shuffling approach, as this has been reported to be promising for other directed evolution approaches in order to optimize KlenTaq properties\[^{76}\]. Therefore, we shuffled the two most interesting mutant genes for each position and the wild-type gene and created a new library consisting of over 2300 members. Under ideal shuffling conditions, this covers all 243 possible mutant combinations with over 98% completeness\[^{77}\]. This library was screened as described above; however except for some single mutants, the most interesting hit we could isolate contained the following mutations K508Y, R587K, and R660T.

Experiments with the purified enzyme showed similar discrimination as the best single mutants, but lower PCR activity. Sequencing of inactive variants or variants with poor activity revealed that the combination of some of the best single point mutations leads to a drop in activity. Since some of these variants showed a further loss in activity during purification and storage, we assume that the combination of these mutations leads to a decrease in protein stability. Noteworthy, none of the above-described single mutants showed any loss in PCR activity, during purification or long-term storage over six months and more at $20^\circ\text{C}$.

Next, we investigated the performance of the most active single mutants on human HeLa genomic DNA in the context of the Factor II prothrombin SNP. The prothrombin SNP is connected to an increased thrombosis risk and therefore of medical interest\[^{10}\]. All mutants show increased $D_c(t)$ values on genomic DNA compared to KlenTaq wild-type. Most mutants show wild-type like high activity with $c(t)$ values of 30 for the match case compared to KlenTaq wild-type with a $c(t)$ value of 28 (Table 2). The mutant R660V shows the highest mismatch extension discrimination of 18 cycles compared to 8.5 for the KlenTaq wild-type as well as wild-type like high activity. Therefore, further experiments were carried out with this enzyme.
First, we evaluated how mutant R660V performs in comparison with commercially available polymerases. Thus, we tested Platinum Taq DNA polymerase, a commercially available wild-type Taq DNA polymerase in comparison to mutant R660V (see Figure 4, A-C) in their capability to amplify from a homozygote (A/A) human gDNA. The gDNA was obtained from the National Institute for Biological Standards and Control (NIBSC). We found, that the use of mutant R660V is resulting in almost doubled Δc(t) values compared to the commercially available enzymes. Encouraged by this result, we also tested the remaining naturally occurring cases: wild-type (genotype G/G) and heterozygote (genotype G/A) with the mutant R660V (see Figure 4, D-F).

As shown in Figure 4 C-D, for both homozygote cases product formation was only detected for the corresponding match primer. The signal of the mismatched primers is shifted to higher c(t) values with the start of the exponential phase after 60 cycles. In the case of the heterozygote template product formation was detected for both allele specific primers, as expected. Therefore, fast and reliable SNP genotyping is possible and with a higher reliability compared to the commercially available polymerases standrdly used. Detection is also possible by standard PCR and subsequent analyses by agarose gel electrophoresis (Figure 4 F). After 50 PCR cycles clear product bands with the expected migration in the gel are visible. As shown all three possible SNP genotypes can be clearly distinguished.

Next we established a multiplexing ASA assay for the Factor II prothrombin SNP by incorporating a 30-nucleotide overhanging sequence at the 5′ end of one of the allele-specific primers. Thus, both alleles can be detected in the same reaction according to their respective melting temperatures[78]. We successfully tested our assay on all three gDNA references, displaying the wild-type, homozygote or heterozygote genotype (see Figure 5A). Notably, all experiments can be conducted in any suitable real-time PCR machine able to perform melting point analysis with a suitable dye such as SYBR Green I.

As reported previously, Taq DNA polymerase activity is significantly inhibited in the presence of less than 0.2% whole blood[79][80]. In contrast, KlenTaq exhibits an increased resistance[80]. Thus, we studied PCR performance of mutant R660V in the presence of whole blood. Due to fluorescence quenching effects of blood on SYBR Green I[80], a 30-fold increased dye concentration was required for real-time PCR. Additionally we used a 5-times higher DNA polymerase concentration in comparison to ASA using purified gDNA samples. The genotype of the blood sample was successfully identified as wild-type (wt; G/ G), as a melting peak was observed only for the G-allelic primer, indicating that the direct detection of SNPs from whole blood samples is achievable (Figure 5B). This may pave the way for clinical diagnostics without the need of time and cost intensive sample preparations.

Table 1: Single turnover rates with std. error (pre-steady-state) of KlenTaq wild-type (wt) and mutants.

|            | kA (s⁻¹)ᵃ | kG (s⁻¹)b | Ka/KG   | (Ka/KG)/(Ka/KG)wt |
|------------|-----------|-----------|---------|-------------------|
| wild-type  | 7.00±0.59 | 0.17±0.02 | 41.2    | 1.00              |
| R487H      | 4.74±0.26 | 0.0062±0.0003 | 766    | 18.6              |
| R487V      | 3.06±0.29 | 0.0021±0.0001 | 1485   | 36.0              |
| K508W      | 3.65±0.22 | 0.0042±0.0003 | 878    | 21.3              |
| K508Y      | 5.01±0.37 | 0.0084±0.0004 | 595    | 14.4              |
| R536K      | 5.65±0.28 | 0.016±0.001 | 355    | 8.61               |
| R536L      | 5.69±0.69 | 0.047±0.004 | 121    | 2.94               |
| R587I      | 4.66±0.32 | 0.023±0.002 | 205    | 4.98               |
| R587K      | 5.29±0.48 | 0.053±0.003 | 100    | 2.44               |
| R660T      | 4.90±0.55 | 0.0088±0.0003 | 559    | 13.6               |
| R660V      | 2.72±0.25 | 0.0093±0.0004 | 294    | 7.13               |

ᵃ kA = kcat match primer in conversions per second (s⁻¹)
ᵇ kG = kcat mismatch primer in conversions per second (s⁻¹)

doi:10.1371/journal.pone.0096640.t001

Subsequently we examined the performance of KlenTaq R660V in HLA typing by sequence-specific primed PCR (PCR-SSP). A primer mix specific for a limited set of HLA alleles from the HLA-DRB1 allelic groups HLA-DRB1*03, *11, *13, *14 has been used to compare AmpliTaq Gold DNA polymerase and KlenTaq R660V with regard to sensitivity and specificity under the PCR conditions previously described[81]. The specificity of the primer mix for the HLA-DRB1 allelic groups HLA-DRB1*03, *11, *13, *14 relies on a two base pair difference at the 3′ end of the sense primer in intron 1 (I1-RB9). The target sequence of the antisense primer in intron 2 (I2-RB28) is identical in all HLA-DRB1 alleles tested. The reaction is designed in a way that the employed primer mix should amplify a 465 bp product of the DNAs in the setups analyzed in rows 1, 2 and 3, but not in rows 4, 5 and 6 of Figure 6. We found, that while KlenTaq R660V gave clear results in the sensitivity as well as the specificity controls, AmpliTaq Gold DNA polymerase also led to weak amplifications in rows 4 and 5 demonstrating less specificity. The designation of the genes and alleles listed in Figure 6 follow the WHO Nomenclature for Factors of the HLA system (http://hla.alleles.org/nomenclature/index.html). After the HLA-prefix followed by the name of the gene (e.g. DRB1) an asterix is inserted to separate the gene’s name from the allele designation. Each HLA allele name has a unique number corresponding to up to four sets of digits separated by colons. The digits before the first colon describe the type, which often corresponds to the serological antigen carried by an allelotype. The next set of digits is used to list the subtypes, numbers being assigned in the order in which DNA sequences have been determined. Alleles whose numbers differ in the two sets of digits must differ in one or more nucleotide substitutions that change the amino acid sequence of the encoded protein. Alleles that differ only by synonymous nucleotide substitutions (also called silent or non-coding substitutions) within the coding sequence are distinguished by the use of the third set of digits. Alleles that only differ

PLOS ONE | www.plosone.org 7 May 2014 | Volume 9 | Issue 5 | e96640
Selective KlenTaq Mutants

by sequence polymorphisms in the introns or in the 5' or 3' untranslated regions that flank the exons and introns are distinguished by the use of a fourth set of digits. The primer mix used in figure 6 is specific for all alleles of the HLA-DRB1 types (digits before the first colon) 03, 11, 13 and 14. All other alleles should not be amplified.

Table 2. Threshold crossing points (c(t)) of real-time PCR experiments, detecting SNP Factor II on HeLa genomic DNA.

| Allele   | c(t) match | c(t) mismatch | Δc(t) |
|----------|------------|---------------|-------|
| wild-type| 28         | 30            | 8.5   |
| R536L    | 37         | 46            | 9     |
| R536K    | 33         | 47            | 14    |
| R537L    | 32         | 46            | 14    |
| R537K    | 31         | 47            | 16    |

Figure 4. ASA assay of Factor II prothrombin, Human gDNA (WHO International Standards) with R660V mutant. Allele-specific primers A and G result in a 3' end match or mismatch depending on the chosen template. In red primer A is shown, in blue primer G. RFU = relative fluorescence units. A) ASA real time PCR curves of Factor II homozygote (genotype A/A) using Taq DNA polymerase, B) Platinum Taq DNA polymerase, or C) mutant R660V. D) ASA real time PCR curves of Factor II wild-type (genotype G/G), E) ASA homozygote (genotype G/A), F) Agarose gel electrophoresis analysis of ASA PCR. M = Marker, A = Primer A, G = Primer G, wild-type (WT) indicates the used wild-type template, homo, the homozygote and hetero, the heterozygote template.

doi:10.1371/journal.pone.0096640.g004

Figure 5. Multiplexing ASA melting curve assay of Factor II prothrombin. A) Melting curves of the multiplexing ASA products resulting from three different reference gDNA samples are shown (wild-type in blue (genotype G/G), heterozygote in black (genotype G/A), homozygote in red (genotype A/A)). B) Resulting melting curve from a whole blood sample, identifying the wild-type genotype.

doi:10.1371/journal.pone.0096640.g005
We next used a PCR-based assay to determine the error rate and the error spectrum of KlenTaq R660V. A 1657 base pairs DNA fragment was amplified by PCR and cloned into a suitable expression vector, transformed into E. coli cells and several single clones analysed by sequencing. With an error rate of 2.8 × 10^{-5} mutations per base per duplication KlenTaq R660V exhibits somewhat improved fidelity than KlenTaq wild-type with 8.8 × 10^{-5} mutations per base per division (Table 3). Notably, mainly the number of transitions dropped whereas the number of transversions is unaffected.

Finally, we explored the suitability of KlenTaq R660V for 5 mC detection and performed MSP within the genomic sequence context of Septin 9. Hypermethylation in the promoter region of the Septin 9 gene is a known colon cancer marker. In these experiments we employed two templates that only differ at the position that is opposite the 3’-primer end. The primer terminates either opposite 5 mC or dU, the product after bisulfite treatment (Figure 7A). To benchmark the performance of KlenTaq R660V we also used a commercially available kit with the same templates and primers. As shown in Figure 7 KlenTaq R660V exhibits clear discrimination for both used primers with the respective mismatch template as well as high activity (Figure 7). Compared with the commercial kit, mutant KlenTaq R660V shows higher activity and discrimination. Thus, robust and fast detection of the epigenetic methylation status is possible using KlenTaq R660V.

Conclusions

Here we demonstrated that the selectivity of a KlenTaq DNA polymerase can be altered by substituting a polar amino acid residue that interacts with the backbone of the primer strand. We successfully identified mutants with increased mismatch selectivity for each examined amino acid position. These findings emphasize the power of combining an initial rational design approach with the rigorous use of a screening based combinatorial enzyme design. However, by gene shuffling the best performing single mutants, we were not able to further improve the desired properties, as combinations of advantageous single point mutations resulted in decreased protein stability and activity. The most promising mutant was thoroughly characterized. We selected the

**Table 3. Error rate of KlenTaq wild-type and KlenTaq R660V.**

| Enzyme                  | No. of clones | Error rate[^b] [10^{-5}] | Average no. of mutations per clone[^a] | No. of insertions | No. of deletions | No. of transversions |
|-------------------------|---------------|--------------------------|----------------------------------------|-------------------|------------------|---------------------|
| KlenTaq wild-type       | 32            | 8.8                      | 0.4                                    | 5                 | 0                | 0                   |
| KlenTaq R660V           | 26            | 2.8                      | 0.2                                    | 0                 | 2                | 0                   |

[^a]: Number of mutations per 650 bases (KlenTaq wild-type) or 726 bases (R660V) sequenced per clone.

[^b]: Error rate equals number of mutations per base per division.
mutant R660V for investigation and found that the enzyme has increased mismatch selectivity and could be used even in multiplexing assays using genomic DNA templates demonstrating its suitability for SNP detection. Additionally, \textit{KlenTaq} R660V is able to perform ASA from DNA in the presence of whole blood with no previous DNA purification. We could also show that \textit{KlenTaq} R660V is suitable for application in MSP to detect the methylation status at a single site. To investigate the impact of the single mutation on overall DNA polymerase selectivity we determined the error rate and spectra of \textit{KlenTaq} R660V and found it to be somewhat increased compared to the wild-type enzyme. Furthermore, the proficiency in HLA typing of \textit{KlenTaq} R660V was investigated. In HLA typing by sequence-specific primed PCR (PCR-SSP) we compared \textit{KlenTaq} R660V with the commercially available AmpliTaq Gold which is widely used in this field. The specificity of the primer mix for the HLA-DRB1 allelic groups HLA-DRB1*03, *11, *13, *14 relies on a two base pair difference at the 3' end of the sense primer in intron 1 (I1-RB). The target sequence of the antisense primer in intron 2 (I2-RB20) is identical in all HLA-DRB1 alleles tested. Whereas the sensitivity of the Taq DNA polymerases was equivalent the specificity of the \textit{KlenTaq} R660V was superior.

**Author Contributions**
Conceived and designed the experiments: MD RK RB AM. Performed the experiments: MD RK CE RB. Analyzed the data: MD RK CE RB AM. Wrote the paper: MD RK CE RB AM.

**References**

1. Roses AD (2000) Pharmacogenetics and the practice of medicine. Nature 405: 857–865.
2. Roses AD (2002) Genome-based pharmacogenetics and the pharmaceutical industry. Nat. Rev. Drug Disc. 1: 541–549.
3. Roses AD (2002) Pharmacogenetics place in modern medical science and practice. Life Sciences 70: 1471–1480.
4. Ginsburg GS, Willard HF (2012) Genomic and Personalized Medicine, Second Edition: V1-2. 2nd ed. Ginsburg GS, Willard HF, editors Academic Press. 1350 p.
5. Lutz MW, Crenshaw DG, Saunders AM, Roses AD (2011) The importance of being connected. J. Alzheimers Dis. 24: 247–251.
6. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. (2004) Large-scale copy number polymorphism in the human genome. Science 305: 525–528.
7. Shi MM (2001) Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clin Chem 47: 164–172.
8. Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2: 930–942.
9. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304–1351.
10. Marchiori A, Mosena L, Prins MH, Prandoni P (2007) The risk of recurrent venous thromboembolism among heterozygous carriers of factor V Leiden or prothrombin G20210A mutation. A systematic review of prospective studies. Haematologica 92: 1107–1114.
11. Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, et al. (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22: 164–167.
12. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, et al. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am. J. Hum. Genet. 66: 979–988.
13. Groth J, Johnson T, Bly R, Rutnik Z, Boyko AR, et al. (2008) Genes mirror geography within Europe. Nature 456: 90–101.
14. Schlebusch CM, Skoglund P, Spodin P, Gattepaille LM, Hernandez D, et al. (2012) Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. Science 338: 374–379.
15. Kathiresan S, Voight BF, Musunuru K, Ardissino D, Mammarci PM, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 41: 334–341.
16. McCarthy JJ, Hilfiker R (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. Nat Biotechnol 18: 505–508.
17. Relling MV, Devriux T (2001) Pharmacogenetics and cancer therapy. Nat Rev Cancer 1: 99–108.
18. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.
19. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299–1320.
20. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–861.
21. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, PeLtonen L, Altshuler DM, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467: 52–59.
22. Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 490: 56–65.
52. Strerath M, Marx A (2005) Genotyping—from genomic DNA to genotype in a
51. Strerath M, Cramer J, Restle T, Marx A (2002) Implications of active site
49. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, et al. (1992) A
47. Deng D, Liu Z, Du Y (2010) Chapter 5 - Epigenetic Alterations as Cancer
46. Sandoval J, Esteller M (2012) Cancer epigenomics: beyond genomics. Current
44. Heichman KA, Warren JD (2012) DNA methylation biomarkers and their utility
43. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM (2006) DNA motifs
39. Bunce M, O’Neill CM, Barnardo MC, Krausa P, Browning MJ, et al. (1995)
37. Wilhelm J, Reuter H, Tews B, Pingoud A, Hahn M (2002) Detection and
36. Shively L, Chang L, LeBon JM, Liu Q, Riggs AD, et al. (2003) Real-time PCR
35. Ishikawa Y, Tokunaga K, Kashiwase K, Akaza T, Tadokoro K, et al. (1995)
32. Germer S, Higuchi R (1999) Single-tube genotyping without oligonucleotide
31. Wilhelm J, Pingoud A (2003) Real-Time Polymerase Chain Reaction.
ChemBioChem 4: 1120–1128.
30. Germer S, Higuchi R (1999) Single-tube genotyping without oligonucleotide
32. Germer S, Holland MJ, Higuchi R (2000) High-throughput SNP allele-
frequency determination in pooled DNA samples by kinetic PCR. Genome Res
10: 258–266.
28. Shi MM, Bleavins MR, la Iglesia de FA (1999) Technologies for detecting
genetic polymorphisms in pharmacogenomics. Mol Diagn 4: 343–351.
27. Myakishev MV, Khripin Y, Hu S, Hamer DH (2001) High-throughput SNP
analysis and validation using real-time PCR. Nucleic Acids Res 29: 2503–2506.
26. Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989) Allele-specific enzymatic
amplification by quantitative guanine genomic DFX for diagnosis of sickle cell anemia.
Proc Natl Acad Sci USA 86: 2757–2760.
25. Wartiovaara A, Syvanen AC (2002) Analysis of nucleotide sequence variations
and polymorphisms by artificial mismatch hybridization. Biochemistry 9: 2858–2865.
24: 2858–2865.
23. Pastinen T, Raitio M, Lindroos K, Tainio P, Peltonen L, et al. (2000) A system
for specific, high-throughput genotyping by allele-specific primer extension on
microarrays. Genome Res 10: 1031–1042.
22. Svytchen AC (1999) From gels to chips: “minisequencing” primer extension for
analysis of point mutations and single nucleotide polymorphisms. Hum Mutat
3: 1–10.
21. Wartiovaara A, Svytchen AC (2002) Analysis of nucleotide sequence variations
by solid-phase minisequencing. Methods Mol Biol 187: 57–63.
20. Ronaghi M (2003) Pyrosequencing sheds light on DNA sequencing. Genome
Res 13: 3–11.
19. Myakishiev MV, Kliripin Y, Hu S, Hamer DH (2001) High-throughput SNP
genotyping by allele-specific PCR with universal energy-transfer-labeled
terminators. Genome Res 11: 163–169.
18. Shi MM, Bleavins MR, the iglesia de FA (1999) Technologies for detecting
genetic polymorphisms in pharmacogenomics. Mol Diagn 4: 343–351.
17. Newton CR, Graham A, Hepinnsill LE, Powell MJ, Summers C, et al. (1989)
Analysis of any point mutation in DNA. The amplification refractory mutation
system (ARMS). Nucleic Acids Res 17: 2503–2516.
16. Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989) Allele-specific enzymatic
amplification by quantitative guanine genomic DFX for diagnosis of sickle cell anemia.
Proc Natl Acad Sci USA 86: 2757–2760.
15. Wilhelm J, Pingoud A (2003) Real-Time Polymerase Chain Reaction.
ChemBioChem 4: 1120–1128.
14. Germer S, Higuchi R (1999) Single-tube genotyping without oligonucleotide
probes. Genome Res 9: 72–76.
13. Germer S, Holland MJ, Higuchi R (2000) High-throughput SNP allele-
frequency determination in pooled DNA samples by kinetic PCR. Genome Res
10: 258–266.
12. Guo Z, Liu Q, Smith LM (1997) Enhanced discrimination of single nucleotide
polymorphisms by artificial mismatch hybridization. Nat Biotechnol 15: 331–
335.
11. Ishikawa Y, Tokunaga K, Kashiwase K, Akaza T, Tadokoro K, et al. (1995)
Sequence-based typing of HLA-A2 alleles using an primer with an extra base
mismatch. Hum Immunol 42: 315–318.
10. Shively L, Chang L, LeBon JM, Liu Q, Riggs AD, et al. (2003) Real-time PCR
assay for quantitative mismatch detection. Biotechniques 34: 490–502.
9. Wilhelm J, Reuter H, Tews B, Pingoud A, Hahn M (2002) Detection and
quantification of insertion/deletion variations by allele-specific real-time PCR:
application for genotyping and chimerism analysis. Biol Chem 303: 1423–1433.
8. Ehrlich H (2012) HLA DNA typing: past, present, and future. Tissue Antigens
80: 1–11.
7. Bunce M, O'Neil CM, Barnardo MC, Krausa P, Brownning MJ, et al. (1995)
Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3,
DRB4, DQB1 by PCR with 144 primer mixes utilizing sequence-
specific primers (PCR-SSP). Tissue Antigens 46: 355–367.
6. Ehrlich H, Wang R (1981) 3-Methylcytosine in eukaryotic DNA. Science 212:
1350–1357.
5. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies
and beyond. Nat Rev Genet 13: 484–492.
4. Dawson MA, Kouzidzis T (2012) Cancer Epigenetics: From Mechanism to
Therapy. Cell 150: 12–27.
3. Felton FA, Lee EE, Costello JF, Plass C, Vertino PM (2006) DNA motifs
associated with aberrant CpG island methylation. Genomics 87: 572–579.
2. Heichman KA, Warren JD (2012) DNA methylation biomarkers and their utility
for solid cancer diagnostics. Clin. Chem. Lab. Med. 50: 1707–1711.
1. Hehn H, Esteller M (2012) DNA methylation profiling in the clinic: applications
and challenges. Nat Rev Genet 13: 679–692.
2001) Enhanced Fidelity in Mutant DNA polymerases. DNA Repair (Amst.) 4: 1390–1398.
2007) Mutations useful in error-prone PCR. Nucleic Acids Res 32: e55–e55.
2005) Enhanced Fidelity in Mutant DNA polymerases. DNA Repair (Amst.) 4: 1390–1398.
2009) Single nucleotide polymorphism genotyping using allele-specific PCR and
fluorescence melting curves. Biotechniques 34: 1068–1072.
2008) Single nucleotide polymorphism genotyping using allele-specific PCR and
fluorescence melting curves. Biotechniques 34: 1068–1072.
2005) Mutations useful in error-prone PCR. Nucleic Acids Res 32: e55–e55.
2000) Lockd nucleic acid (DNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 32: e55-e55.
2008) Lockd nucleic acid (DNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 32: e55-e55.
2000) Lockd nucleic acid (DNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 32: e55-e55.
2000) Lockd nucleic acid (DNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 32: e55-e55.
2000) Lockd nucleic acid (DNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 32: e55-e55.