DNA Synthesis from Aberrant Substrates by *KlenTaq* DNA Polymerase

A Functional and Structural Analysis

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

an der Universität Konstanz
Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Chemie

vorgelegt von
Nina Blatter

Konstanz, 2013

Tag der mündlichen Prüfung: 7. März 2014

1. Referent: Prof. Dr. Andreas Marx
2. Referent: Prof. Dr. Martin Scheffner

Parts of this work have been published in:

ChemBioChem. **2014,** in press DOI: 10.1002/cbic.201400051	**N. Blatter**, A. Prokup, A. Deiters, A. Marx "Modulating the pK_a of a Tyrosine in *KlenTaq* DNA Polymerase that is crucial for Abasic Site Bypass by In Vivo Incorporation of a Non-canonical Amino Acid"
Angew. Chem. Int. Ed. **2013**, 52, 11935 *Angew. Chem.* **2013**, 125, 12154	**N. Blatter**, K. Bergen, O. Nolte, W. Welte, K. Diederichs, J. Mayer, M. Wieland, A. Marx "Structure and Function of an RNA-Reading Thermostable DNA Polymerase"
EMBO J. **2010**, 29, 1738	S. Obeid[a], **N. Blatter[a]**, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, A. Marx "Replication through an Abasic DNA Lesion: Structural Basis for Adenine Selectivity" [a] contributed equally
Patent	WO2014023318 A1 (2014) „New DNA Polymerases with Increased Substrate Scope"

Other publications

ChemBioChem. **2011**, 12, 1574	S. Obeid, A. Schnur, C. Glöckner**, N. Blatter**, W. Welte, K. Diederichs, A. Marx "Learning from Directed Evolution: *Thermus Aquaticus* DNA Polymerase Mutants with Translesion Synthesis Activity"

Danksagung

In erster Linie möchte ich mich bei Prof. Dr. A. Marx für die Aufnahme in seiner Arbeitsgruppe, der Vergabe eines sehr interessanten Promotionsthemas und für die hervorragenden Arbeitsbedingungen bedanken. Darüber hinaus ein großes Danke für das in mich gesetzte Vertrauen, die fachliche sowie auch die menschliche Unterstützung. Vor allem wenn es auch mal nicht so gut lief, hattest Du immer ein offenes Ohr.

Ich möchte mich bei Prof. Dr. Martin Scheffner für die Übernahme des Zweitgutachtens sowie bei Prof. Dr. Jörg S. Hartig für die Übernahme des Prüfungsvorsitzes bedanken.

Der ganzen jetzigen und ehemaligen AG Marx danke ich für das super Arbeitsklima, die wissenschaftliche Diskussionsbereitschaft und eine sehr schöne Zeit in und außerhalb des Labors in den letzten Jahren.

Allen voran danke ich meinen Laborkollegen, den alten sowie den neuen, die jederzeit sehr hilfsbereit waren und mit denen man immer viel Spaß haben konnte ☺.

Des Weiteren danke ich….

unter anderem Samra Obeid, Ramon Kranaster, Christian Glöckner, Konrad Bergen, Jutta Mayer und Markus Wieland für die Zusammenarbeit an verschiedenen Publikationen. Auch Prof. Dr. W. Welte und Prof. Dr. K. Diederichs danke ich für die gute Zusammenarbeit.

meinen Mitarbeiterpraktikanten Simon Geigges, Irene Griesser und Marius Hausberger, die an Teilen dieser Arbeit mitgewirkt haben.

vielen ehemaligen Kollegen für die schöne Zeit in und außerhalb des Labors. Darunter auch Christian Glöckner, meinem ersten Betreuer, der mich in die Welt der DNA-Polymerasen eingeführt hat.

Karin Betz, Samra Obeid, Daniel Schneider, Holger Bußkamp und Janina Watzdorf für das Korrekturlesen meiner Arbeit.

vor allem auch Daniel Schneider und Holgi (den Nicht-Laborkollegen ☺) für die Diskussions- und Hilfsbereitschaft bei diversen wissenschaftlichen Fragen.

allen Freunden, die mich während des Studiums begleitet haben. Ohne Euch und ohne Charlottes Tafel Schokolade wäre ich heute bestimmt nicht hier!

und allen anderen, die hier leider keinen Platz gefunden haben.

Meiner Familie und meinem Freund Michi möchte ich für ihre andauernde Unterstützung und manchmal auch in stressigen Zeiten für ihre Geduld danken. Ihr habt immer an mich geglaubt, ohne Euch wäre das hier nicht möglich gewesen!

Table of Content

I. General Introduction

1. DNA and RNA: Structure and Function

The discovery of DNA and RNA, and the elucidation of their role and function, spans several decades. It first began in 1869 when Friedrich Miescher isolated a substance he called "nuclein"[1] and progressed in the following years with the discovery of the DNA and RNA components.

The question of the DNA function remained until 1944, when O. Avery, C. McLeod and M. McCarty identified DNA as the fundamental unit of the *transforming principle* (genetic material).[2] While the base composition of the DNA was already determined in 1950 by E. Chargaff, who stated that 'the ratio of Adenine to Thymine and of Guanine to Cytosine were nearly 1.0 in all species studied'[3], the definite molecular structure of DNA was postulated in 1953 by J. Watson and F. Crick;[4, 5] also based on the work done by Rosalind Franklin (**Figure 1**).[6]

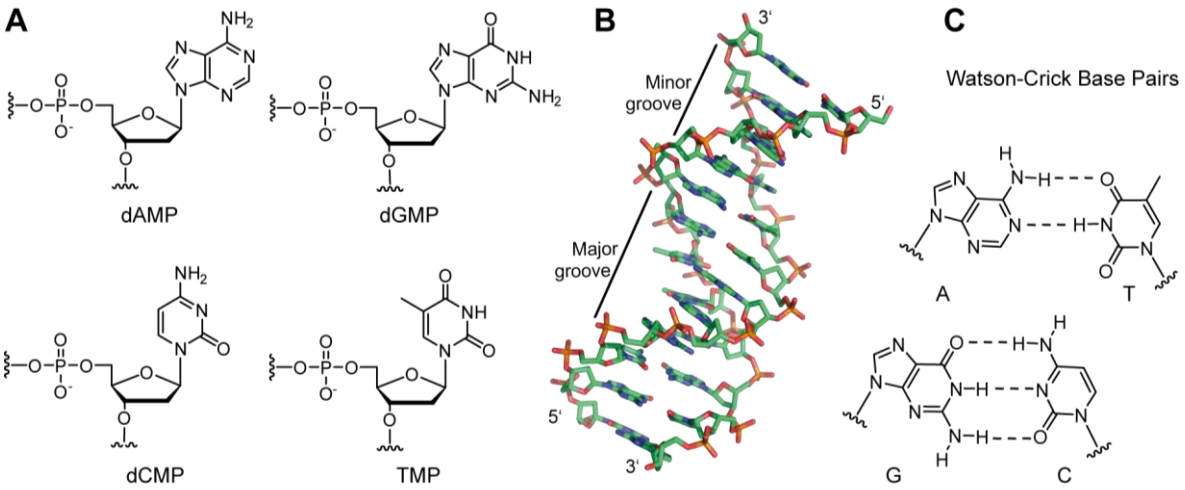


Figure 1. DNA structure and building blocks.
A) Chemical structures of 2'-deoxyadenosine monophosphate (dAMP), 2'-deoxyguanosine monophosphate (dGMP), 2'-deoxycytidine monophosphate (dCMP) and 2'-deoxythymidine monophosphate (TMP). B) Helical structure of DNA (PDB ID 1BNA[7]). Major and minor groove are depicted. C) Watson-Crick base pairing of adenine with thymine and guanine with cytosine. Hydrogen bonds are indicated as dashed lines.

Nowadays, it is fundamental knowledge that DNA is the carrier of genetic information and consists of four different units: the nucleotides (**Figure 1A**). Every nucleotide possesses a phosphate, a 2'-deoxyribose and a nucleobase moiety. Four different nucleobases are available: the pyrimidines thymine and cytosine, and the purines adenine and guanine. The nucleobases are linked to the sugar moiety via an N-glycosidic bond. Phosphodiester bonds further connect the sugar and the phosphate moiety and give rise to a DNA strand with a 5'

and 3' end. Two strands align to each other in an antiparallel fashion forming a double helix with a minor and a major groove (**Figure 1B**). DNA can adopt various conformations depending on the base sequence and the environment.[8, 9] The conformations can differ in the number of base pairs per helical turn, the geometries of the minor and the major groove and the sugar conformation. The B form is most common under physiological conditions with the sugar puckers exhibiting a C2'-endo conformation (**Figure 1B, Figure 2C**).[9] The bases point towards the core with their plane perpendicular to the helix axes forming hydrogen bonds with each other. Another conformation is the A form which can occur in dehydrated DNA samples, in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes.[10, 11] The sugar puckers in the A form exhibit a C3'-endo conformation. A fundamental principle, however, is that adenine always pairs with thymine and guanine with cytosine, thus the sequence of one strand always determines the sequence of the second strand (**Figure 1C**).
In contrast to the double-stranded DNA, RNA molecules are single-stranded and can form highly structured molecules. The components of both, DNA and RNA, are similar, with the exception that RNA possesses a ribose instead of a 2'-deoxyribose moiety and that uracil replaces thymine as a nucleobase (**Figure 2A, B**).

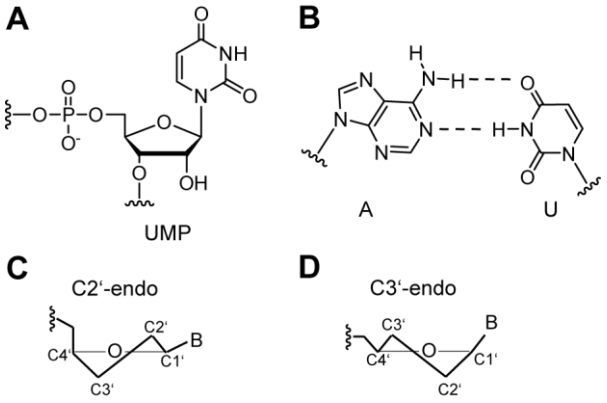


***Figure 2*.** Structural features of RNA. A-B) Chemical structure of uridine monophosphate (UMP) and base pairing with adenine are depicted in A) and B), respectively. C-D) Scheme of C2'-endo and C3'-endo conformations of the sugar puckers are shown in C) and D), respectively.

Due to the additional 2'-OH group at the sugar, RNA is more prone to hydrolysis than and not as stable as DNA. The typical C3'-endo conformation of the sugar moiety gives rise to the characteristic A form of the RNA (**Figure 2D**). Whereas DNA codes the genetic material, RNA directs for the synthesis of proteins (mRNA),[12] but also makes up the genome of various viruses.[13, 14]

2. DNA Polymerases

2.1 Structure and Function of DNA Polymerases

In 1953, when Watson and Crick published the three dimensional DNA structure, they further stated, that 'It has not escaped our notice that the specific pairing we have postulated suggests a possible copying mechanism for the genetic material'.[4] During 1955 to 1957, Arthur Kornberg identified the enzyme responsible for this copying mechanism and named it 'DNA polymerase' (**Figure 3**).[15, 16]

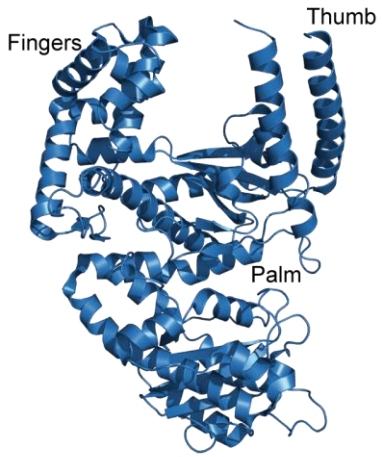


Figure 3. Tertiary structure of *E. coli* DNA polymerase I (PDB ID 1KFD).[17] The structure resembles a right hand with the subdomains: fingers, thumb and palm.

DNA polymerases are involved in DNA replication, DNA repair and recombination.[18, 19] In order to catalyse DNA synthesis, these enzymes require a DNA primer with a free hydroxyl group at the 3' end hybridized to a DNA template. Nucleoside triphosphates (dNTPs) are the substrates of a DNA polymerase which incorporates the monophosphate at the 3' OH end of the primer strand, whereby the template directs for correct nucleotide incorporation. A phosphodiester bond is formed and pyrophosphate is released. Thus, DNA synthesis proceeds from the 5' to the 3' end. To date, seven different DNA polymerase families are known (families A, B, C, D, X, Y, RT), which differ in the primary structure of their catalytic subunits. However, they all resemble a right hand and have three common subdomains: palm, fingers and thumb. These domains form a cleft in which the DNA is bound and elongated.[18, 19]

2.2 DNA- vs RNA-dependent DNA Polymerases

DNA- and RNA-dependent DNA polymerases differ in their use of the templating nucleic acid, with DNA-dependent DNA polymerases replicating from DNA and RNA-dependent DNA polymerases from RNA as template. Both classes of enzymes use dNTPs as substrate and catalyse DNA synthesis. DNA-dependent DNA polymerases can belong to sequence families A, B, C, D, X or Y, whereas RNA-dependent DNA polymerases, also termed reverse transcriptases, comprise a separate sequence family.[19]

DNA-dependent DNA polymerases of sequence families A, B and X are generally assumed to be involved in replication or repair depending on the organism the DNA polymerase originates from. So far, DNA-dependent polymerases belonging to sequence families C and D were observed only in bacteria or archaea, respectively, and are assumed to take part in the replication process. Members of family Y, however, are involved in translesion synthesis in all organisms of life.[19]

H. Temin and D. Baltimore discovered reverse transcriptases in 1970 and could demonstrate for the first time that part of the 'central dogma'[20] of biology, defining the flow of information from DNA to RNA to proteins, can be partly reversed.[21, 22] In nature, reverse transcriptases are essential in the retroviral life-cycle, as they can reverse transcribe the viral RNA genome into DNA to be integrated into the host's genome. Although, reverse transcriptases structurally resemble other DNA polymerases with the shape of a right hand, the subdomains are more functionally than structurally homologous.[23] So far, structural and functional studies have been focused on the reverse transcriptase of HIV-1 retrovirus (HIV-RT), as it poses a major target in the battle against AIDS. Crystal structure analysis of this enzyme yielded significant insights into how a DNA polymerase can accept both DNA and RNA as template, due to the intriguing ability of HIV-RT, as well as other reverse transcriptases, to act as a DNA- and RNA-dependent DNA polymerase.[24-26]

2.3 DNA Catalysis

2.3.1 DNA Catalysis - Mechanism

The mechanism of DNA synthesis is similar in all DNA polymerases and can be divided into five different steps, depicted in the scheme below (**Figure 4**).[23, 27-29] As *KlenTaq* DNA polymerase, a member of sequence family A, is in the focus of this work, the steps are mainly discussed with regard to the changes occurring in this enzyme.

The DNA is bound in a crevice formed by the thumb, palm and fingers subdomains of the DNA polymerase. For members of DNA polymerase family A, a conformational change of the thumb domain is observed upon binding of the primer/template duplex (step 1, E:P/T), which brings the tip of the thumb (a helix-loop-helix motif) in proximity to the DNA duplex.[30] Simultaneously, the 3' terminus of the primer is aligned in the active site and the single-

stranded overhang of the template is flipped out involving a sharp twist in the sugar-phosphate backbone.[30]

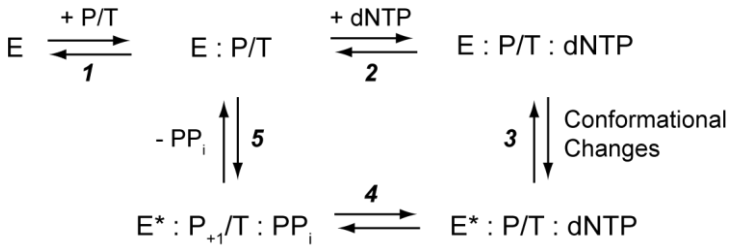


Figure 4. Kinetic pathway of nucleotide incorporation. Depicted complexes as well as the kinetic steps are described in the text. Scheme was adapted from Rothwell and Waksman.[28, 29]

The next step involves weak binding of the incoming dNTP by the finger subdomain and thereby progression from a binary to an open ternary complex (step 2, E:P/T:dNTP). The following step 3 corresponds to a conversion from an inactivated (E:P/T:dNTP) to an activated complex (E*:P/T:dNTP) in which all components are efficiently aligned in the active site, thus facilitating the chemical reaction.[29, 30] The formation of the active complex was suggested to be the rate-limiting step in DNA catalysis and was thought to be connected to a conformational change of the protein finger domain. Crystal structure analysis of *KlenTaq* DNA polymerase revealed a reorientation of the tip of the finger domain by 46° rotation towards the active site from an open to a closed conformation. This conformational change affects the orientation of the O helix located in the finger domain. Whereas the orientation of the O helix in the open form resembles its conformation in a binary *KlenTaq* complex, it packs against the templating nucleobase and the incoming nucleotide in the closed state. This conformational change of the O helix releases a tyrosine residue at position 671 (Tyr671) from its stacking arrangement on top of the first base pair duplex and makes room for the templating base opposite the incoming dNTP. Recent studies, however, negate this conformational change of the fingers domain as the rate-limiting step, and suggest that conformational changes in the active site, such as the arrangement of side chains or the binding of metal ions, are responsible.[28] With every component poised for catalysis, the chemical reaction (S_N2) can occur (step 4, E*:P_{+1}/T:PP_i). Two metal ions, present in the active site, promote DNA catalysis and stabilize the trigonal bipyrimidal transition state of the reaction (**Figure 5**).

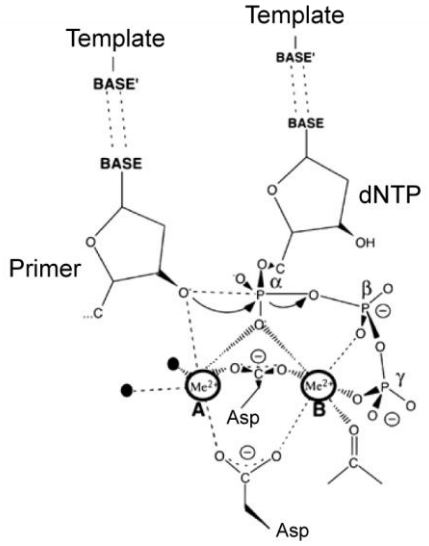


Figure 5. The 'two-metal ion' mechanism of DNA polymerases. Two metal ions facilitate DNA synthesis and stabilize the trigonal bipyrimidal transition state of the reaction. Scheme was adapted from Brautigam and Steitz[27] showing the active site of *E. coli* DNA polymerase I. Water molecules are depicted as black spheres.

Metal ion A supports deprotonation of the 3' OH group of the primer and facilitates the nucleophilic attack of the primer on the α-phosphate of the incoming nucleotide. Metal ion B interacts with the triphosphate moiety of the bound nucleotide and facilitates the release of the pyrophosphate (PP_i). Both metal ions are further coordinated by catalytically active aspartic acid residues in the palm domain, which further stabilize the transition state. This 'two-metal ion' mechanism is conserved in every DNA polymerase sequence family (**Figure 5**).[23, 27, 31, 32]

The reaction is completed with the formation of the phosphodiester bond and the release of pyrophosphate (step 5, E:P/T). Next, the DNA polymerase either translocates for the next round of incorporation or dissociates from the primer/template duplex and DNA synthesis ends. Recently, it was shown that general acid catalysis is also part of the nucleotidyl transfer mechanism.[33] An active site amino acid residue protonates the pyrophosphate leaving group, thus neutralizing the negative charge which is formed during the transition state and contributing to the release of pyrophosphate.

2.3.2 DNA Catalysis – Kinetic Analysis

The reaction pathway of an enzyme, in this study that of a DNA polymerase, can be investigated by single nucleotide incorporation experiments using either steady[34, 35] or pre-steady[36] state kinetic conditions. Steady state kinetic measurements provide information about K_M and v_{max} of an enzymatic reaction, whereas K_M is defined as the Michaelis-Menten constant and v_{max} represents the maximal velocity of the reaction. However, these paramters are determined based on the complete reaction pathway (see **Figure 4**) with the system in an equilibrium state. Steady state kinetics are conducted using 'single completed hit' conditions facilitated by an excess of primer/template complex compared to the amount of enzyme used and a maximal primer conversion of 20 %.[34, 35] These conditions ensure that every DNA polymerase binds to a primer/template complex maximal once.

Pre-steady state kinetics allow the analysis of individual steps in the reaction pathway and can provide the affinity K_d and incorporation rate constant k_{pol}. These constants are independent from association or dissociation events of the polymerase to the primer/template complex

due to 'single turnover' conditions used in the experiment.[36] The excess of enzyme compared to the amount of primer/template complex used ensures that every primer/template complex is already occupied by a DNA polymerase, thus facilitating immediate nucleotide binding and incorporation after dNTP addition. In this case, K_d is defined as the affinity of the dNTP substrate to the enzyme and k_{pol} designates the maximal incorporation rate of the enzyme. The incorporation efficiency of the DNA polymerase is defined as k_{pol}/K_d.

Stopped- and quench-flow techniques are commonly used to measure substrate conversion in such a set-up.[37, 38] Stopped-flow methods require an optical signal for the reaction of interest to be followed over time, such as fluorescently labelled oligonucleotides or intrinsic protein fluorescence.[39, 40] However, not every system can be followed using an optical signal and it is often difficult to asign the signal to a particular step in the reaction pathway.[37]

Quench-flow experiments are based on a quenching agent which is used to terminate the reaction after short reaction times which allows immediate quantification of substrate conversion. In general, radioactively labelled substrates are used to visualize the reaction products for subsequent quantification. Furthermore, rapid chemical quench flow instruments facilitate the measurement of substrate conversion after short time periods, which are not accessible in a manual set-up (see chapter **VI 2.5.5**). In this work, pre-steady state kinetic measurements were performed using a quench-flow approach.

2.4 Selectivity and Fidelity of DNA Polymerases

DNA polymerases have to recognize their substrate with high specificity in order to maintain the stability of the genome. The fidelity of these enzymes varies, from translesion bypass DNA polymerases with a correct incorporation rate from 10:1 to high-fidelity polymerases with a rate from 100 000:1.[19] In nature, fidelity is further increased by the mismatch repair system,[41] DNA polymerase auxiliary proteins (PCNA,[42] RF-C,[43] RP-A[44]) or the 3'-5'-exonuclease activity (proofreading domain) which many polymerases possess.[45]

How DNA polymerases select from the pool of four nucleotides, differing only in the nucleobase moiety, is still a question of ongoing investigations. First, Watson-Crick hydrogen bonding between the complementary bases was thought to be the main factor employed by DNA polymerases to select the right nucleotide.[4, 5] But it was realized that hydrogen bonding cannot be the only selection factor, as the free energy barrier between matched and mismatched base pairs is too low to account for the fidelity many polymerases possess.[46] In fact, nowadays, steric effects are suggested as one of the most important elements for the selection process, supported by base stacking, water solvation and hydrogen bond interactions between the protein and the minor groove of the DNA (for more detailed information see Kunkel et al.[46]). New classes of molecules were designed and their acceptance by several DNA polymerases tested in order to investigate the influence of steric restraints. The lab of Kool developed isosteric nonpolar molecules which lack hydrogen

bonding ability but resemble the natural nucleobases in shape and size. For example, the incorporation of dAMP opposite difluorotoluene (a shape analogue of T, **Figure 6A**) in the template strand was highly efficient in case of the Klenow Fragment of DNA polymerase I from *E. coli,* bacteriophage T7 DNA polymerase and HIV-RT,[47-50] supporting the hypothesis that steric constraints are one major factor in nucleotide selection.

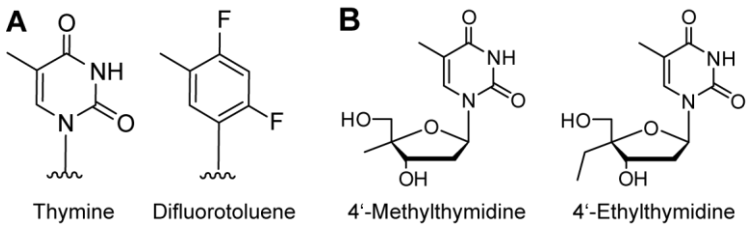


Figure 6. Modified nucleotides designed to investigate the influence of hydrogen bonding on nucleotide selection and the model of 'active site tightness'. A) Comparison of natural thymine and its isosteric, nonpolar analogue difluorotoluene. B) Nucleoside analogues with a methyl or ethyl substitution at the C4' position.

In this context, the ideas of a 'size exclusion' model and 'active site tightness' were proposed.[50-52] The nucleotide binding pocket was suggested to be limited in its size and shape, thus accommodating the natural base pairs, but rejecting nucleobases that violate these geometric restraints. Furthermore, the model suggests, that the active site tightness depends on the DNA polymerase studied, with a low-fidelity polymerase exhibiting a more flexible and a high-fidelity polymerase a more rigid active site. Studies investigating the acceptance of nucleotide analogues with a methyl or ethyl substitution at the C4' position provided direct evidence for the involvement of considerably varied steric effects on fidelity among different DNA polymerases (**Figure 6B**).[53] Nowadays, the studies investigating the acceptance of difluorotoluene as a nonpolar shape mimic are challenged due to the fact that difluorotoluene can form hydrogen bonds under specific conditions.[54] However, both hydrogen bonding and steric constraints are still considered to play a major role in nucleotide selection.[55]

As mentioned in the previous chapter **I 2.3.1**, conformational changes within the enzyme are also suggested to play a crucial role in the selection process. For *KlenTaq* DNA polymerase different models exist on how conformational changes influence the selection process. On the one hand, it was proposed that every nucleotide is non-specifically bound to the O helix in the open form and delivered to the active site during a fast conformational change of the fingers subdomain. There, the discrimination is thought to occur via slow conformational changes such as the arrangement of side chains or the assembly of the two catalytically active ions.[28, 29, 56] A second model supports the existence of an early intermediate state which enables the

templating base to 'preview' the incoming dNTP in the open ternary complex.[57-59] Only correctly paired dNTPs can bypass this intermediate state, whereas mismatched nucleotides are rejected and released from the enzyme-DNA-complex. New data from crystal structure analysis of DNA polymerase I from *Bacillus stearothermophilus* bound to a mismatched nucleotide revealed an ajar conformation of the enzyme which was suggested to act as checkpoint in the selection pathway.[60] Wu et al. proposed that the templating base is used to align or misalign the incoming nucleotide via hydrogen bond formation in a partially closed conformation. In case of a mismatched base pair (dG:dTTP), the incoming nucleotide is misplaced in the active site, with the remaining space filled with water. Thus, due to the still present solvation and the misalignment of the nucleotide, the chemical reaction is hampered and provides a selection mechanism against mismatched nucleotides. Further studies of the selection pathway of DNA polymerases will show if these intermediate recognition states are common in all DNA polymerase sequence families.

2.5 *KlenTaq* DNA Polymerase – Structure and Function

The large fragment of DNA polymerase I from *Thermus Aquaticus* (*Taq*) is termed Klenow fragment of *Taq* DNA polymerase (in short *KlenTaq*, **Figure 7**).

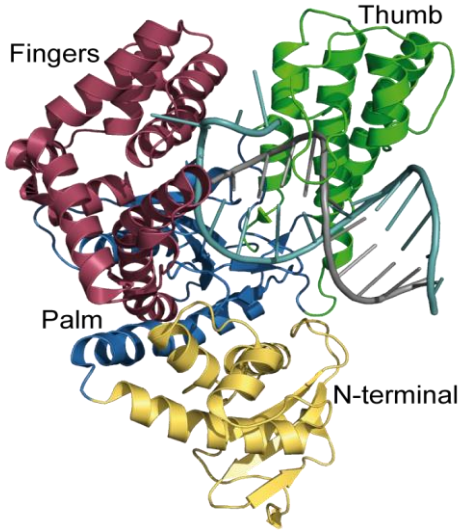


Figure 7. Crystal structure of *KlenTaq* DNA polymerase in complex with a DNA duplex (PDB ID 3RTV[61]). The fingers, thumb, palm and N-terminal subdomains are indicated in red, green, blue and yellow, respectively. DNA primer and template are coloured in grey and cyan.

This terminology is based on the enzyme's sequence homology to *E. coli* DNA polymerase I[62] which comprises an N-terminal 5'-3' endonuclease, a central 3'-5' exonuclease (proofreading) and a C-terminal polymerase domain (for review see Patel et al.[63]). The Klenow fragment of *E. coli* DNA polymerase I is devoid of the N-terminally located 5'-3' endonuclease domain, by definition, but harbours the remaining two.[64, 65] Thus, the *KlenTaq* DNA polymerase only consists of the C-terminal domain associated with the polymerase activity, as the Klenow fragment excludes the 5'-3' endonuclease domain and the fact that *Taq* DNA polymerase is devoid of the 3'-5' exonuclease domain per se.

In our studies, the N-terminal truncated form of *Taq* DNA polymerase I comprises 540 amino acids (amino acids 293-832 of the *Taq* sequence).[66] It is a thermostable enzyme and thus, together with its full-length variant, often employed in various biotechnological applications such as PCR-based reactions. *KlenTaq* DNA polymerase belongs to the family of DNA-dependent DNA polymerases, more specifically to sequence family A, whose members are involved in DNA repair and replication. Structurally, the enzyme exhibits the shape of a right hand, typical of DNA polymerases, with the subdomains: palm, fingers and thumb (**Figure 7**).[62, 67] Structure analysis of *KlenTaq* DNA polymerase has added significant contributions to the understanding of how DNA polymerases recognize the cognate substrate,[30, 61, 68-71] process abasic sites[72-74] and accept modified or unnatural nucleotides.[61, 69-71]

3. Abasic Site Bypass - 'A-rule'

Cellular DNA is continuously damaged by both endogenous and exogenous agents. Although exogenous agents like oxidative stress, exposure to ionizing radiation and alkylating agents contribute to DNA damage, the majority of the mutations are caused in an endogenous manner.[75] The most common damage under physiological conditions are abasic sites resulting from the spontaneous cleavage of the bond connecting the sugar and the nucleobase in DNA (**Figure 8**).[76] Approximately 10,000 bases are spontaneously lost in a cell each day leaving behind apurinic and apyrimidinic lesions; thereby purines are more susceptible to hydrolysis than pyrimidines.[76-78] Additionally, these lesions are generated enzymatically as an intermediate in base excision repair.[79]

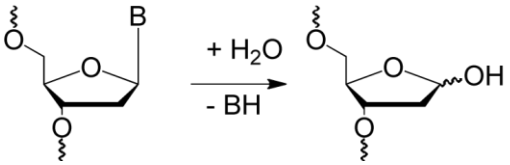


Figure 8. Formation and structure of a natural occurring abasic site.

Numerous cell repair mechanisms and check points exist, which are able to remove abasic lesions. However, those damages that remain undetected pose a challenge for DNA polymerases involved in replication.[80, 81] As these lesions are devoid of genetic information, and therefore non-instructional, they give rise to the formation of mutations.[78] These mutations are the result of DNA synthesis past a lesion by either replicative or translesion synthesis

(TLS) DNA polymerases.[80] Whereas high-fidelity replicative DNA polymerases often stall at sites of DNA damage during replication and repair, the cell can employ specialized enzymes capable of bypassing DNA damage. These TLS DNA polymerases possess a low processivity and exhibit, in terms of nucleotide incorporation, a low fidelity.[80] Nevertheless they are employed by the cell to continue DNA synthesis past a DNA lesion, due to the fact that the acceptance of an increase in mutational load is in some cases preferable to cell death. DNA polymerases belonging to the Y family are generally assumed to be the major group of TLS enzymes inherently possessing a more open and less constrained active site allowing the accommodation of lesions perturbing the DNA structure.[80, 82] Various structural and functional studies were conducted to investigate the mechanism DNA polymerases of both families, replicative and TLS DNA polymerases, possess to overcome abasic sites.[83-92]
In case of *E. coli* DNA polymerase II[90] and Dpo4 from *S. Solfataricus*[85, 88], the enzymes can facilitate abasic site bypass via a looping-out mechanism. Crystal structures of Dpo4 in complex with various abasic site containing templates show the lesion looped out in an extrahelical position with the incoming nucleotide placed against the base 5' to the lesion.[85] Another member of the Y family Rev1 employs an 'amino acid templating' mechanism with an arginine residue directing for dCMP incorporation, whereas the lesion is driven to an extrahelical position. As a considerably amount of abasic sites stems from the loss of guanine, the enzyme adapted in a unique way to always incorporate a dCMP opposite this type of lesion to increase the probability to bypass thus 'error-free'.[92] Additionally, crystallographic data from human DNA polymerase ι, also a member of the Y family, show a unique adaption to abasic site bypass. The abasic lesion and the incoming dNTP are located intrahelically with both moieties coming in close proximity to each other due to the enzyme's constricted active site cleft. Thus, the incoming dNTP is stabilized. This alignment may underlie DNA polymerase ι's ability to insert all four nucleotides opposite an abasic site. The small preference for dGMP is derived from a combination of increased base stacking and specific interactions with the polymerase by the nucleotide.[89]

Although the described enzymes show unique mutation spectra when bypassing an abasic site, several DNA polymerases like T4 DNA polymerase,[93] DNA polymerase α,[35] β,[94] δ,[95] HIV-RT,[96] Klenow fragment of *E. coli* DNA polymerase I[84] and RB69 DNA polymerase[86, 87] were reported, in either *in vivo* or *in vitro* studies, to incorporate dAMP and to a lesser extent dGMP opposite an abasic site. A phenomenon which was found mainly for DNA polymerases from sequence families A (including human DNA polymerases γ and θ) and B (including human DNA polymerases α, ε, and δ) and was termed 'A-rule'.[78, 83, 84, 97-102] As Watson-Crick base pairing cannot be the reason for the selectivity of dAMP incorporation, due to the lack of genetic information inherent in an abasic site, the mechanistic basis of the 'A-rule' is still under debate.[35, 52, 82, 86-88, 94, 101, 103-106]

The fact that most DNA polymerases following the 'A-rule' also facilitate template-independent incorporation of dAMP at blunt-ended DNA[107] implies that the incorporation selectivity is not guided by the structure of the abasic site. Several properties like the superior base stacking or solvation of adenine were discussed as the major determinants of the 'A-rule'.[108, 109] This would be in good concordance with the fact that adenine has the greatest π-π-stacking energy of all natural occurring bases, followed by G, T and C.[110] The model of stacking interactions as the major determinant of the 'A-rule' was discussed to be further supported by the fact that an artificial 5-nitro-1-indolyl nucleotide (dNITP, **Figure 9**) is incorporated with higher efficiency than dAMP opposite an abasic site analogue; as was reported for B family DNA polymerase members, such as RB69 or T4.[104, 111] This nucleotide analogue is unable to undergo hydrogen bonding but possesses higher stacking abilities.

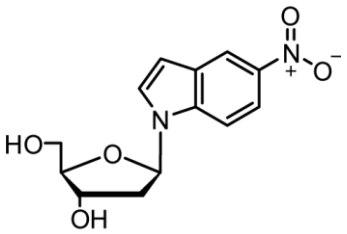


Figure 9. Structure of the 5-nitro-1-indolyl nucleoside.

Structural data of an exonuclease deficient variant of RB69 DNA polymerase in complex with an abasic site analogue containing template and dNITP suggest that dipole-induced dipole stacking interactions of the nitro moiety with the n-1 template base 3' to the lesion are the main factors for the enhanced incorporation. Additionally, the orientation of the indole ring resembles that of the purine in a regular dTMP-dAMP base pair, which leaves the DNA phosphate backbone significantly unperturbed.[111]

Several crystal structures of RB69 DNA polymerase, a member from sequence family B, in complex with an abasic site analogue containing template are available.[86, 87, 109, 111] But only recently, Xia et al. were able to solve the structure of the enzyme in a ternary complex with dATP as the incoming nucleotide. They proposed that stacking interactions and partial charge interactions between the incoming nucleotide and the penultimate base pair adjacent to the abasic site contribute to the different incorporation efficiencies of dNTPs opposite an abasic site.[109]

So far, structural data of *KlenTaq* DNA polymerase, a member of sequence family A, is the only other available structure for a DNA polymerase from sequence families A and B, which follow the 'A-rule' when bypassing abasic sites. Obeid et al. were able to solve the structure of *KlenTaq* DNA polymerase in complex with an abasic site analogue containing template and an incoming 2',3'-dideoxyadenosine-5'-triphosphate (ddATP).[72] The structural data suggests that *KlenTaq* DNA polymerase follows the 'A-rule' by utilizing an 'amino acid templating' mechanism. Thereby, interactions with a tyrosine residue at position 671 might account for the preference of purines over pyrimidines. Tyr671 was proposed to assume the role of the templating nucleobase as its shape and size resembles a six membered pyrimidine nucleobase in the template strand. Thus, it might direct for preferential purine incorporation to maintain

an enhanced geometric fit to the active site.[72] Further structural and functional data, giving insights into the nature of the abasic site bypass mechanism, however, are of great relevance; especially in view of the fact that abasic sites result in transversion mutations commonly found in human cancers.[112]

4. Application of DNA Polymerases in Molecular Biology and Diagnostics

To date, DNA polymerases are employed in various applications spanning the fields of molecular biology, biotechnology and diagnostics. Many different kinds of enzymes exist with properties ranging from high substrate specificity (e.g. Phusion™ DNA polymerase) to an increased substrate spectrum,[113] or even optimized characteristics such as inhibitor resistance.[114] The use of DNA polymerases in the polymerase chain reaction (PCR)[115] has revolutionized molecular biology and clinical diagnostics. Key in these processes is the ability of DNA polymerases to recognize a primer/template complex and promote DNA synthesis. Thus, DNA can be exponentially amplified during repeated cycles of heating and cooling and can be analysed subsequently. Due to the growing field of chemical fluorescence probes, amplification can be monitored even in real time. DNA polymerases are further applied in diverse sequencing methods,[68, 116-118] for the detection of single nucleotide polymorphisms (SNP) in the field of personalized medicine[119, 120] or in the directed evolution of enzymes via error-prone PCR.[121, 122] The evolution of DNA polymerases towards the acceptance of non-natural substrates opens further applications in DNA- or RNA-labelling,[123] in *in vitro* selection methods such as SELEX[124] or in general the development of XNA replication systems.[125] Also belonging to the family of DNA polymerases, reverse transcriptases, such as the enzymes from Moloney murine leukemia virus (MoMLV) and Avian myeloblastosis virus (AMV), are used in reverse transcription PCR, a crucial method for RNA detection employed in molecular biology or clinical diagnostics.[14] The applications of DNA polymerases in general are diverse with a permanent growing number of optimized enzymes tailored for a specific application.

5. Directed Evolution of DNA Polymerases

5.1 Overview

The process of *in vitro* evolution, especially directed enzyme evolution,[126, 127] has proven to be a powerful method to generate these enzyme variants with improved or new properties tailored for specific applications. In contrast to a rational design of mutants, the directed evolution of proteins requires no structural information of the protein, as mutations are introduced randomly. The method comprises an iterative process of three different steps:

Random mutagenesis in order to generate a library of enzyme variants, the expression of the enzymes and a subsequent screening or selection step. During this process mutations accumulate until a desired level of improvement is achieved, with the gene of the most promising variant selected after every round and employed as template in the next cycle.[128]
Mutations can be either introduced on the entire target gene coding for the respective protein or on selected amino acid positions. The introduction can be facilitated by various techniques such as saturation mutagenesis, DNA shuffling, StEP (Staggered Extension Process) or error-prone PCR.[121, 129-131] Subsequent transformation into a host organism, e.g. *E. coli,* generates the library. However, it is crucial that the phenotype and genotype are 'connected' in the library due to the following selection or screening step. High-throughput screening strategies achieve separation from other variants by conventional compartmentalization based on multi-well plates. One method for high-throughput screening was established in our lab which employs the fluorescent dye SYBRGreen I to identify active polymerase mutants.[132] The dye exhibits an increased fluorescence signal upon binding to the minor groove of double-stranded DNA (emission at 520 nm). Therefore, the amplification of DNA in PCR by active DNA polymerase mutants can be visualized either in real time or through end-point determination.
Connecting the phenotype to the genotype in selection based strategies can be facilitated e.g. in phage display, ribosome display, mRNA display or water-oil emulsions.[133-135] The method of compartmentalized self-replication (CSR)[136] also relies on the formation of water and oil emulsions and provides a powerful tool for the evolution of DNA polymerases. It is based on a simple feedback loop with active polymerases replicating their own gene. Thus, adaptive gain is directly translated into genetic amplification of the encoding gene.

5.2 DNA Shuffling

The next chapter will focus on DNA shuffling, as it was the method of choice in this work. DNA shuffling is defined as the *in vitro* recombination of selected genes by random fragmentation and PCR reassembly.[130, 137] This method is based on four different steps consisting of gene preparation, DNA fragmentation, reassembly of these fragments in a self-priming polymerase reaction and an amplification of the recombined fragments in PCR (**Figure 10**).
DNA fragmentation can be achieved via DNase I digestion of the parental DNA[130, 137] or via short randomly designed primers which anneal to the parental DNA and are extended by a DNA polymerase at or below room temperature.[138] In the following reassembly step, the fragmented genes are reassembled in a 'reverse' PCR without using primers. The fragments replace the primer, as homologous stretches anneal and form a primer/template complex elongated by a thermostable DNA polymerase. Consequently, the number of DNA molecules decreases during DNA reassembly, whereas in standard PCR the number of DNA molecules exponentially increases.[137] This step also offers the possibility to either introduce mutations

by including synthetic oligonucleotides with 3' and 5' ends that are homologous to the genes or to eliminate redundant mutations by backcrossing with the wild-type sequence.[130, 137]

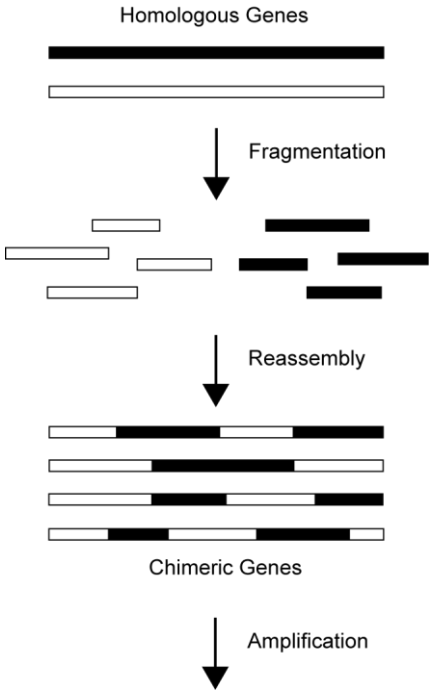


Figure 10. Principle of DNA shuffling. Depicted are the fragmentation of the parental DNA (homologous genes), the recombination, reassembly and amplification in PCR.

DNA shuffling was first reported to be successfully applied in a β-lactamase model system resulting in enzyme mutants with an increased antibiotic resistance against cefotaxime.[130] A high point mutation rate of 0.7 % was observed which is comparable to the rate in error-prone PCR. Whereas a high-error rate is desired for gaining diversity in *in vitro* evolution applications, the opposite is true for studies focusing on the structure-function relationship between homologous genes or, studies in which beneficial mutations were already identified and the respective mutants are to be recombined without gaining new mutations. Consequently, protocols were developed in which each step was optimized to yield a low error-rate.[139] Thus, one protocol reported an error-rate as low as 0.05 %, which was mainly achieved by including high-fidelity DNA polymerases during gene preparation, reassembly and in the post-amplification PCR step.[139]

5.3 Thermostable DNA Polymerases with Reverse Transcriptase Activity

The evolution of DNA polymerases towards the acceptance of non-natural substrates facilitates a variety of applications in both molecular biology and diagnostics, as described before (chapter **I 5.1**). The acceptance of non-cognate substrates by DNA-directed DNA polymerases also includes the usage of RNA as template for DNA synthesis in a process called reverse transcription. How DNA- or RNA- dependent DNA polymerases discriminate between the natural templates (RNA vs DNA) and maintain their substrate specificity is still a subject of ongoing investigations.[140, 141] Structural studies yielding insights into this process are lacking and thus designing DNA polymerases in a rational fashion to accept both, DNA and RNA, as template remains a challenge. However, thermostable DNA polymerases accepting both substrates would provide a crucial tool for the so-called reverse transcription PCR (RT-PCR), a fundamental technique utilized in many applications in molecular biology and clinical diagnostics such as transcriptome analysis, pathogen detection as well as disease-specific marker recognition.[14, 142]

The detection and quantification of RNA in RT-PCR is generally based on the enzyme-mediated reverse transcription of RNA to its complementary DNA (cDNA) by a reverse transcriptase and a subsequent amplification of the resulting DNA by a DNA-dependent DNA polymerase in PCR. The detection can be even monitored in real time. The reverse transcriptase and the DNA-dependent DNA polymerase can be applied either in separate (two-enzymes/two tubes) or single (two-enzymes/one tube) reactions. One tube reactions having the reverse transcription prior to PCR amplification, termed one-step RT-PCR, are time- and work-saving. Additionally, the risk of contamination is reduced as, in general, an RNA digestion step or the addition of different buffers can be omitted.[14, 142]

Although two enzyme mixtures are state of the art, several drawbacks arise from the heat-instability of commonly used retroviral mesophilic reverse-transcriptases[143] such as MoMLV and AMV. Performing the reverse transcription step within a one-step RT-PCR set-up requires low temperatures (i.e. 45 °C) to allow activity of the reverse transcriptase, which facilitates unspecific priming, low yield on complex targets e.g. from secondary structure formation of the mRNA template and premature reaction termination.[142] Furthermore, the reverse transcription step results in a time addition to the PCR protocol, a disadvantage especially in the field of point of care testing or outbreak situations when hundreds of swabs need to be analysed in a short period of time. Therefore, the development or discovery of heat-stabile reverse transcriptases would be desirable but was shown to have its limitations.[143-146] So far an increase in thermostability was gained by eliminating the RNase H activity,[143] by site-directed mutagenesis[145, 146] or random mutations,[144] but the achieved thermostability was insufficient for the use of these enzymes in PCR.

Consequently, strong efforts have been undertaken to evolve thermostable DNA-dependent DNA polymerases with reverse transcriptase activity applicable in RT-PCR.[147-151] These enzymes offer the possibility to perform one step RT-PCR at high temperatures minimizing

secondary structure formation of RNA, enhance specificity and, in general, render the use of two enzymes unnecessary, thereby eliminating problems such as inhibitory effects between a reverse transcriptase and a DNA polymerase.[152-154] In addition, time consuming optimizations for two enzymes could be omitted, further providing a work and time reduction.

Most efforts to develop DNA-dependent DNA polymerases with increased reverse transcriptase activity focused on enzymes from sequence family A.[147, 148, 150, 155-157] Particularly, *Taq* DNA polymerase was the object of various studies due to some intrinsic reverse transcriptase activity exhibited by the enzyme.[158] However, few enzymes were shown to be applicable in RT-PCR[147, 148, 150, 151] and to the best of my knowledge only two of these enzymes are currently commercially available, one belonging to sequence family A and one isolated from a viral metagenomic library.[147, 151] Thus, the demand for DNA polymerases with increased reverse transcriptase activity persists.

6. Aim of this Work

The aim of this work was to investigate the ability of *KlenTaq* DNA polymerase to adapt to aberrant structures and even to accept those as substrates. In detail, the incorporation mechanism opposite an abasic site containing template and the processing of RNA as a non-cognate template were to be studied in a functional and structural analysis of the wild-type and mutant enzymes.

Abasic sites are the most common damage under physiological conditions and can arise spontaneously or enzymatically through hydrolysis of the sugar-nucleobase bond, leaving behind a non-coding lesion.[76] DNA polymerases from sequence families A and B preferentially incorporate an adenine opposite this lesion; a phenomenon termed the 'A-rule'.[78, 83, 84, 97-102] But what are the determinants of this process of the selection of purines over the pyrimidines? Furthermore, this preference for purines is also observed for the template-independent addition of nucleotides at blunt-ended DNA. Does a general mechanism exist, which can be applied in both cases? These issues should be investigated focussing on *KlenTaq* DNA polymerase as a model system for members of sequence family A. Structural analysis of *KlenTaq* DNA polymerase in complex with an abasic site containing template performed by Dr. Samra Obeid suggested that an 'amino acid templating' mechanism might facilitate abasic site bypass.[72] A tyrosine protein side chain was proposed to fill the space of the absent template nucleobase and mimic the shape and size of a pyrimidine, and in consequence, direct for purine incorporation. However, crystal structure analysis only provides a static view of the incorporation event. Thus, the aim of this work was to corroborate the structural results by a site-directed mutagenesis approach. The tyrosine residue should be mutated to various amino acids and the resulting variants should be investigated in primer extension as well as pre-steady state kinetic experiments regarding their lesion bypass activity.

The second part of this work should focus on the ability of *KlenTaq* DNA polymerase, a DNA – dependent DNA polymerase, to process RNA as a templating nucleic acid. Two important aspects should be addressed in this project. First, the idea was to develop a thermostable DNA-dependent DNA polymerase which is capable of reverse transcription and PCR, thus providing a crucial tool applicable in reverse transcription PCR (RT-PCR).[14, 142] To date, two enzymes, a reverse transcriptase and a PCR-competent DNA polymerase, are standard in RT-PCR. However, several drawbacks arise from the use of two enzymes[152-154] and the heat-instability of commonly used retroviral mesophilic reverse transcriptases.[143] For this purpose, two previously reported thermostable *KlenTaq* variants[148, 159] already exhibiting an increased tendency to accept aberrant substrates should be recombined via DNA shuffling: A combination of both was promising to yield mutants with even more pronounced properties.

Variants with a reverse transcriptase activity exceeding the parental enzymes should be selected for further characterization. Additionally, the most promising variant should be applied in RT-PCR experiments, testing its potential as a crucial tool for molecular biology and clinical diagnostics.[14]

In a second approach, we want to get insights into the fundamental biological question of how a DNA-dependent DNA polymerase can discriminate between DNA and RNA as a templating nucleic acid. Accurate replication of DNA is crucial for the maintenance of a functional genome,[160] thus, it is essential in cells that a DNA-dependent DNA polymerase can discriminate between DNA and RNA. But what are the structural features responsible for this selection process? To address this question, *KlenTaq* wild-type or variants capable of processing RNA more efficient should be crystallized in complex with DNA and RNA as template. Thereby, we hope to gain further insights into the structural features implemented by a DNA-dependent DNA polymerase to discriminate between the different substrates.

II. Results and Discussion – Abasic Site Bypass and Template-Independent Nucleotide Addition at Blunt-Ended DNA

1. Abasic Site Bypass

1.1 Introduction

Structural data of *KlenTaq* DNA polymerase in complex with an abasic site analogue containing template and an incoming ddATP (*KlenTaq*$_{AP}$) suggest that the DNA polymerase utilizes an 'amino acid templating' mechanism to bypass an abasic site.[72]
In detail, the obtained crystal structure revealed a conformation of the enzyme very similar to already described structures of the enzyme with a cognate substrate (PDB ID 1QSY),[68] but with major changes in the fingers domain. More specifically, the O helix located in the fingers domain adopts a state that leaves the active site more open compared to a closed ternary state already described for the enzyme in complex with a natural DNA duplex (**Figure 11A,** for further details see chapter **I 2.3.1**).

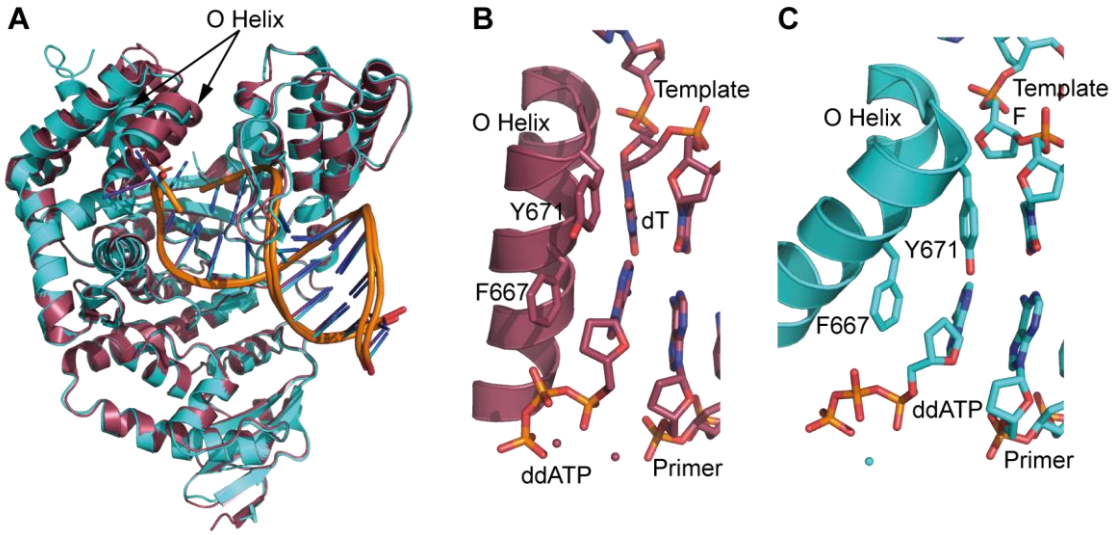


Figure 11. Comparison of *KlenTaq*$_{AP}$ structure and *KlenTaq* bound to non-damaged template.
A) Overlay of *KlenTaq*$_{AP}$ structure (cyan) and *KlenTaq* bound to non-damaged template (red, PDB ID 1QSY). The different conformations of the O helix are indicated with an arrow. B)-C) Active site arrangement of *KlenTaq* bound to non-damaged template and *KlenTaq*$_{AP}$ structure. Depicted are residues Y671, F667, the O helix and the incoming ddATP opposite dT (B) or abasic site analogue F (C). Water molecules are shown as spheres. Graphic was adapted from Obeid, Blatter et al.[72]

Furthermore, only one Mg^{2+}-ion was observed in the active site. The metal ion coordinates to the triphosphate moiety of the incoming nucleotide and forms interactions with two surrounding water molecules. The abasic site analogue is intrahelically placed in the DNA substrate and the tetrahydrofuran moiety of the analogue is rotated by about 90° in comparison to the respective sugar conformation in the natural case. This conformational

change allows a tyrosine side chain at position 671 to flip in and fill the void left by the absent nucleobase (**Figure 11B, C**). Additionally, a hydrogen bond interaction network is established between the incoming ddATP, Tyr671 and surrounding residues. The distance between the hydroxyl group of Tyr671 and N3 of the incoming ddATP indicates a further interaction and a possible hydrogen bond formation (**Figure 12**).

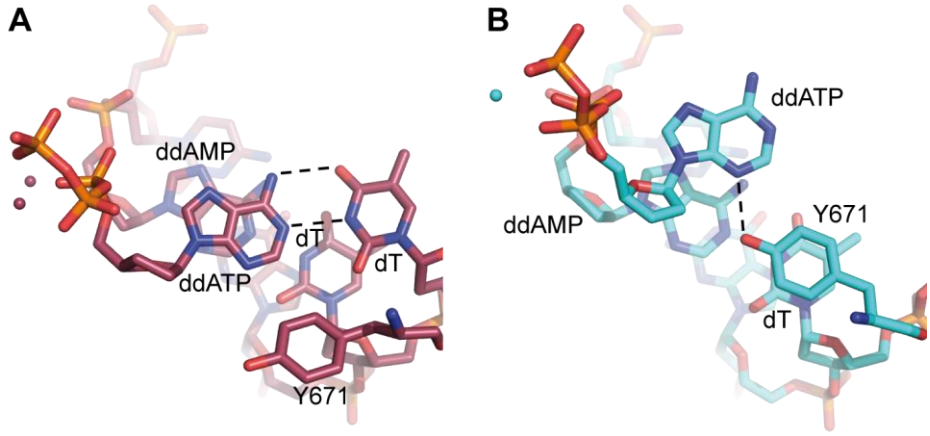


Figure 12. Hydrogen bond formation and stacking interactions. A) *KlenTaq* (PDB ID 1QSY, red) processing ddATP opposite template nucleobase dT. Incoming ddATP stacks to the nucleobase at the primer end (ddAMP). B) *KlenTaq*$_{AP}$ (cyan) processing ddATP opposite F. Y671 stacks to the templating nucleobase (dT) 3' to the abasic site analogue F. Hydrogen bonding is indicated with a dashed line; the primer and the template terminus are shown in transparent in both cases. Graphic was adapted from Obeid, Blatter et al.[72]

Furthermore, a different alignment of the ddATP was observed in the active site compared to the structure with a non-damaged template. The orientation of ddATP in the *KlenTaq*$_{AP}$ structure negates stacking interactions with the primer strand and increases the distance from the primer 3'-terminus to the α-phosphate of the ddATP by 2.3 Å. Although stacking interactions are lost between the ddATP and the primer strand, new interactions are formed. Thus, Tyr671 stacks to the template nucleobase positioned 3' of the abasic site, further stabilizing the active site arrangement (**Figure 12**).

The arrangement in the active site suggests that the tyrosine residue at position 671 might assume the role of the templating nucleobase. It fills the void left by the abasic site and resembles in its shape and size a six membered pyrimidine nucleobase in the template strand. Consequently, it might direct for purine incorporation opposite the abasic site, thus providing an optimal geometric fit to the active site.[51, 52] Additionally, the amino acid side chain forms a distinct hydrogen bond network with the incoming nucleotide as described above.

1.2 Results

1.2.1 *KlenTaq* Follows the 'A-rule'

The structure impressively shows that the Tyr671 fills the vacant space of the missing nucleobase. However, as crystal structure analysis only provides a static view of nucleotide incorporation, one can only assume that the tyrosine also adopts the function of the nucleobase, namely to direct for nucleotide incorporation. Thus, in order to corroborate the proposed model, various functional studies are required and were conducted as part of my PhD thesis. First, selected mutations at position 671 were introduced and primer extension studies as well as pre-steady state kinetics were performed with the obtained variants (**Figure 13, Table 1**). To further investigate the role of hydrogen bonding in abasic site bypass, we also implemented changes in the nucleobase of the substrate and studied the incorporation efficiency of the resulting modified ddATP analogue (d3DATP), synthesized by Dr. Samra Obeid. Pre-steady state kinetic measurements of dAMP and dGMP incorporation opposite an abasic site, as well as the incorporation of dAMP opposite dT, were performed by Dr. Christian Glöckner in preliminary work. Dr. Ramon Kranaster contributed to the site-directed mutagenesis and purification of *KlenTaq* variants. Part of this work is published in the EMBO Journal.[72]

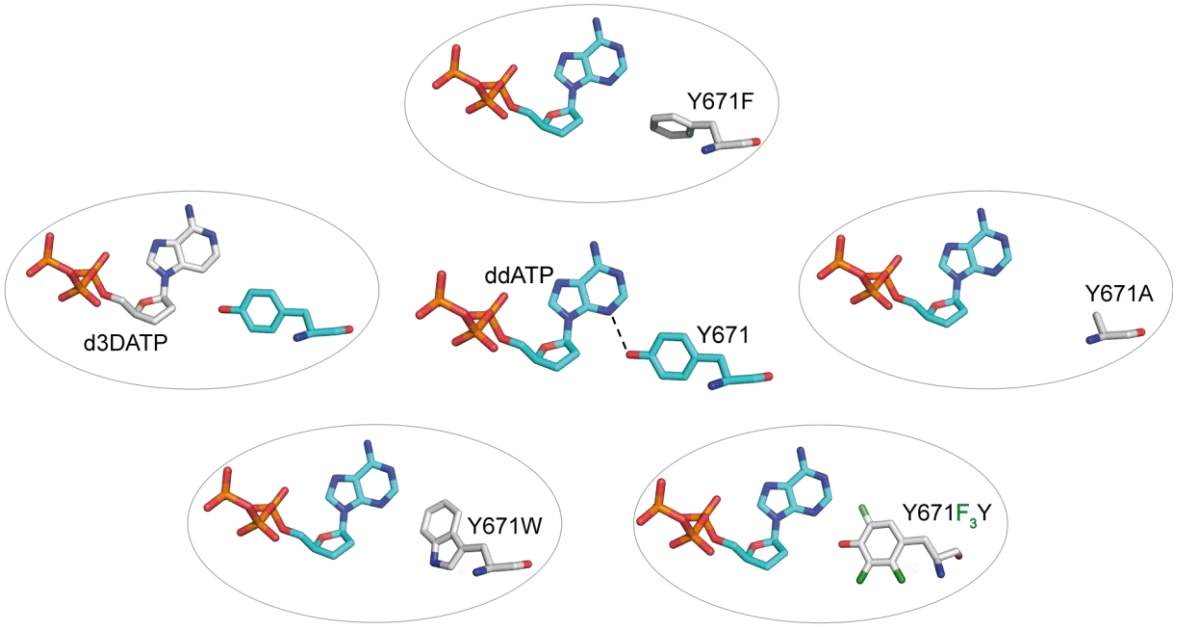


Figure 13. Schematic representation of the mutated residues in the protein or at the nucleobase moiety. Interactions partners are shown in the sticks model, whereas the mutated partner is shown in grey.

First, single nucleotide incorporation studies opposite an abasic site analogue were conducted showing that the enzyme indeed follows the 'A-rule' (**Figure 14**). Tetrahydrofuran, an isosteric and isoelectronic analogue of deoxyribose, was elected as a stable model for an abasic site, termed henceforth F (**Figure 14A**).

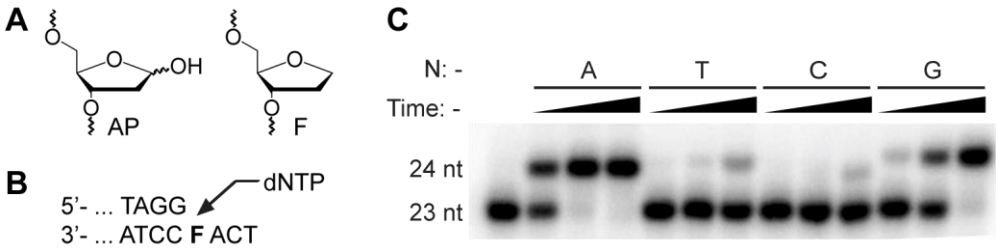


Figure 14. Nucleotide incorporation opposite an abasic site analogue F. A) Chemical structure of natural abasic site (left) and abasic site analogue F (tetrahydrofuran, right). B) Partial primer/template sequence used in primer extension experiments. C) Single nucleotide incorporation opposite F with 500 nM *KlenTaq* wild-type. Reaction times were 2, 10 and 60 min, respectively. The respective dNTP (N) is indicated. Primer is depicted in the first lane; nt: nucleotide. Graphic was adapted from Obeid, Blatter et al.[72]

Using a 5'-[^{32}P]-radioactively labelled DNA primer annealed to an F-containing template, preferential dAMP incorporation was observed for *KlenTaq* wild-type (**Figure 14B, C**). Although the incorporation of pyrimidines was also visible after prolonged incubation times, they are clearly less efficiently processed opposite a lesion than purines, consistent with previous reports in the literature.[35, 78, 84, 96, 99]

1.2.2 Tyr671 Mimics a Pyrimidine Nucleobase

Next, the tyrosine was exchanged for an alanine residue to investigate the role of the aromatic ring system. Indeed, an activity loss was observed for the Y671A variant in case of the natural template as well as a further reduction in case of the abasic site F containing template; thus indicating a general importance of the aromatic residue at this position (**Table 1, Figure 15A**). The efficiency of dAMP incorporation opposite canonical template dT is more than 5350-fold reduced compared to the wild-type enzyme and, because of the low overall incorporation efficiency, incorporation of TMP and dCMP opposite F could not be determined at all (**Table 1**). Interestingly, the mutant incorporates dGMP four times more efficient than dAMP opposite F. The reduced activity of this variant in case of dAMP incorporation opposite natural dT is in good concordance with previous findings in the literature which implicate that Tyr671 may act as a positioning device for the DNA during DNA catalysis.[30]

Table 1. Transient kinetic analysis of nucleotide incorporation opposite F by *KlenTaq* and mutants.

Enzyme	template	dNTP	k_{pol} [s^{-1} x 10^{-2}]	K_d [μM]	k_{pol} / K_d [μM^{-1} s^{-1} x 10^{-4}]
wild-type	dT	A	516 ± 44	15.3 ± 4.7	3373
wild-type	F	A	2.73 ± 0.23	149 ± 35	1.83
wild-type	F	G	0.43 ± 0.01	65.5 ± 5.2	0.66
wild-type	F	T	0.03 ± 0.006	299 ± 122	0.01
wild-type	F	C	0.02 ± 0.001	294 ± 17	0.007
Y671A	dT	A	1.06 ± 0.06	168 ± 26	0.63
Y671A	F	A	0.02 ± 0.002	304 ± 68	0.007
Y671A	F	G	0.05 ± 0.002	182 ± 23	0.03
Y671A	F	T	n.a.	n.a.	n.a.
Y671A	F	C	n.a.	n.a.	n.a.
Y671F	dT	A	2265 ± 161	54.1 ± 12.0	4187
Y671F	F	A	1.16 ± 0.06	233 ± 30	0.50
Y671F	F	G	0.33 ± 0.03	294 ± 54	0.11
Y671F	F	T	0.02 ± 0.002	245 ± 64	0.008
Y671F	F	C	0.01 ± 0.001	262 ± 65	0.004
wild-type	dT	d3A	59.6 ± 3.3	95.4 ± 13.4	62.5
wild-type	F	d3A	0.12 ± 0.01	572 ± 67	0.02
Y671W	dT	A	91.6 ± 15.2	363 ± 104	25.2
Y671W	F	A	0.10 ± 0.02	514 ± 182	0.02
Y671W	F	G	n.a.	n.a.	n.a.
Y671W	F	T	0.35 ± 0.03	382 ± 76	0.09
Y671W	F	C	0.92 ± 0.12	845 ± 161	0.11

n.a.: not accessible since the turnover after 1h using 600 μM dNTP was less than 20%.

Next, a Y671F mutant was constructed to probe the importance of the hydroxyl group for the enzymatic activity. Although, the incorporation efficiency of dAMP opposite a natural template was comparable to *KlenTaq* wild-type, a four-fold reduced efficiency of dAMP incorporation opposite F was observed (**Figure 15B, Table 1**). Kinetic analysis revealed that the reduced activity resulted from an increased K_d and a reduced k_{pol}, indicating the requirement of the hydroxyl group for efficient catalysis. The reduced incorporation efficiency opposite F corroborates the results of the structural data, which indicate hydrogen bond formation between the hydroxyl group of Tyr671 and the N3 of the incoming ddATP. However, the variant still follows the 'A-rule' (**Figure 15B**).

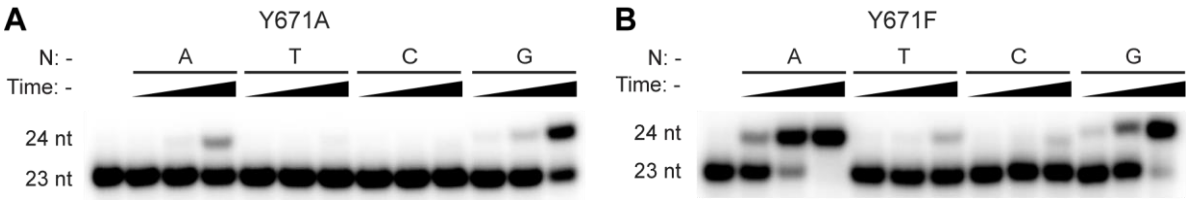


Figure 15. Primer extension experiments. A-B) Single nucleotide incorporation opposite F with 500 nM *KlenTaq* Y671A (A) and Y671F (B). Reaction times were 2, 10 and 60 min, respectively. The respective dNTP (N) is indicated. Primer is depicted in the first lane; nt: nucleotide.

The hydrogen bond formation between these two moieties was further investigated using 3-deaza-2'-deoxyadenosine-5'-triphosphate (d3DATP), an adenine analogue in which the N3 is substituted by a nonpolar CH. As already described in earlier reports,[161] the incorporation efficiency of this analogue opposite dT decreased by 54-fold compared to the natural substrate. In case of the abasic site template, a further decline in incorporation efficiency was observed with a more than 90-fold reduction compared to dATP. The same interaction was studied by either employing the Y671F mutant or the modified d3DATP. Kinetic data obtained from both studies demonstrate that an interaction between Tyr671 and N3 of the incoming adenine occurs which contributes to stabilizing the active site arrangement in the abasic site bypass (**Table 1**).

Finally, a Y671W mutant was constructed with the idea in mind that a bicyclic indole consisting of a six-membered ring fused to a five-membered ring would resemble in its shape and size a purine. Thereby, by transforming the aromatic ring of the tyrosine to a tryptophane, we hoped to switch the mechanism and direct for preferential incorporation of pyrimidines. And indeed, in single nucleotide primer extension reactions preferred incorporation of dCMP and TMP opposite the abasic site F was observed for Y671W (**Figure 16A, B**), thus providing an optimal geometric fit to the active site. Pre-steady state kinetic analysis with this mutant confirmed the results of the primer extension studies (**Figure 16C, D, Table 1**). Whereas the wild-type enzyme incorporates TMP and dCMP with 183- and 261-fold reduced efficiency opposite the abasic site compared to dAMP, respectively, this nucleotide selectivity is reversed in Y671W (**Table 1**). The mutant enzyme lost the preference for dATP and instead incorporates the pyrimidines TMP and dCMP about five-fold more efficient than dAMP.

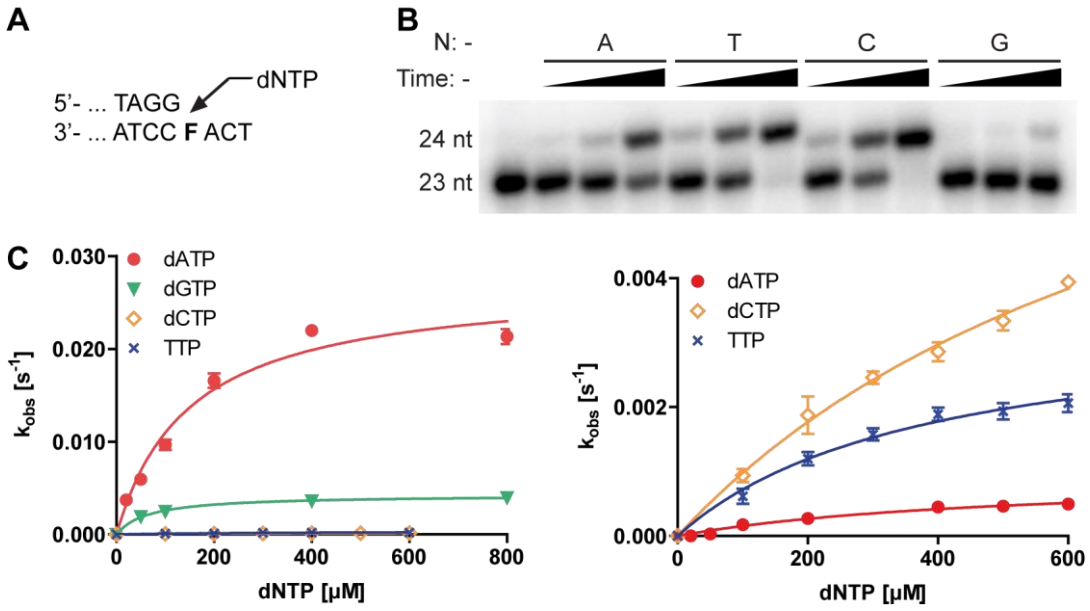


Figure 16. Nucleotide incorporation opposite F. A) Partial primer/template sequence used in primer extension experiments and kinetic analysis. B) Single nucleotide incorporation opposite F with 500 nM *KlenTaq* Y671W. Reaction times were 2, 10 and 60 min, respectively. The respective dNTP (N) is indicated. Primer is depicted in the first lane; nt: nucleotide. C) Pre-steady state kinetic analysis of single nucleotide incorporation opposite F with *KlenTaq* wild-type (left) and Y671W (right), respectively. Observed pre-steady state rates (k_{obs}) were plotted against the respective dNTP concentration and fitted to a hyperbolic equation. Graphic was adapted from Obeid, Blatter et al.[72]

In summary, based on the results of the crystal structure[72] and the mutagenesis studies performed in this work, we propose an 'amino acid templating' mechanism. A tyrosine residue at position 671 replaces the absent templating nucleobase in the abasic site containing substrate and mimics in size and shape a pyrimidine, thus directing for preferential purine incorporation. Thereby, an optimal geometric fit at the active site is maintained.

A collaboration with the group of Prof. A. Deiters from the Pittsburgh University, Pennsylvania, allowed us to further study the interactions involved in abasic site bypass.[162] They were able to site-specifically incorporate a fluorinated tyrosine analogue at position 671 in *KlenTaq* DNA polymerase. Fluorine atoms have similar steric properties as hydrogen atoms, but they exhibit a much higher electronegativity, thus allowing fluorotyrosines to be utilized as probes for hydrogen bonding interactions in proteins.[163] Depending on the positioning of fluorine atoms at the phenyl ring, the pk_a values of fluorotyrosine probes can be manipulated and the acidity of the tyrosine's hydroxyl group can be increased.[164-166] Furthermore, the selection of an appropriate fluorinated tyrosine analogue and a reaction pH below or above the analogue's pk_a value facilitates the modulation of the protonation state of the tyrosine's

hydroxyl group and thus provides a tool to study hydrogen bond formation in proteins. In this study, hydrogen bonding between the hydroxyl group of Tyr671 and the N3 of the incoming dATP was investigated by incorporating 2,3,5-fluorotyrosine (**Figure 17A**) at position 671 in *KlenTaq* DNA polymerase (*KlenTaq* F_3Y). The tri-substituted fluorotyrosine (pk_a: 6.4) was selected due to its highly decreased pk_a value compared to natural tyrosine (pk_a: 9.9).[165]

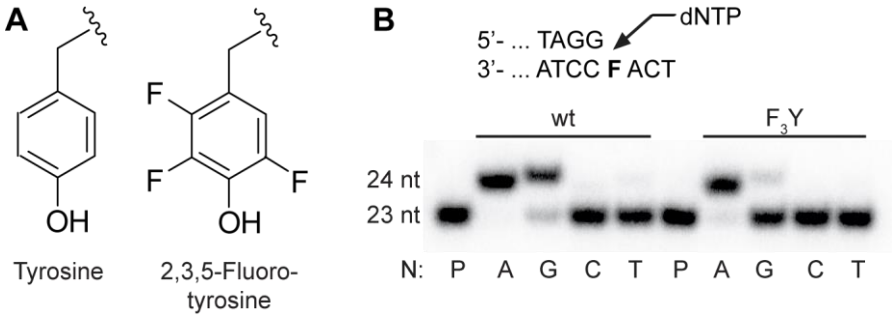


Figure 17. Studies with a *KlenTaq* variant containing the non-natural amino acid 2,3,5-fluorotyrosine at position 671 (*KlenTaq* F_3Y). A) Chemical structure of the cognate (left) and the non-natural amino acid (right). B) Single nucleotide incorporation opposite F for 25 nM *KlenTaq* wild-type (wt) and *KlenTaq* F_3Y, respectively. The reactions were incubated for 120 min. Partial primer and template sequences are shown. The respective dNTP is indicated (N). P: Primer; nt: nucleotide. Graphic was adapted from Blatter et al.[162]

Incorporation of the fluorinated analogue was analysed by SDS-PAGE and MALDI-MS/MS (**Figure S1**, chapter **VII 1.1**). SDS-PAGE analysis, performed by A. Prokup from the group of Prof. Deiters, demonstrated that protein expression was only observed in the presence of 2,3,5-fluorotyrosine. Furthermore, the expressed protein variant was digested with trypsin and the fluorinated tyrosine containing peptide confirmed by MALDI-MS/MS analysis.

Initially, single nucleotide incorporation experiments opposite F were conducted to investigate if the variant still follows the 'A-rule'. Buffer conditions with a pH of 7.5, employed in previous experiments, were maintained and indeed, preferential dAMP incorporation was observed for *KlenTaq* F_3Y. Reactions containing pyrimidines as substrate yielded no visible primer elongation (**Figure 17B**). Next, the enzymatic activities of *KlenTaq* wild-type and the F_3Y variant were compared incorporating dAMP opposite natural dT or F in the template. Polymerase amounts were diluted in a step-wise manner and showed slightly reduced incorporation efficiency for *KlenTaq* F_3Y compared to *KlenTaq* wild-type opposite dT, but a more pronounced activity loss when dAMP incorporation opposite F was measured (**Figure 18**). The results indicate a reduced incorporation rate for dAMP opposite F for *KlenTaq* F_3Y compared to *KlenTaq* wild-type.

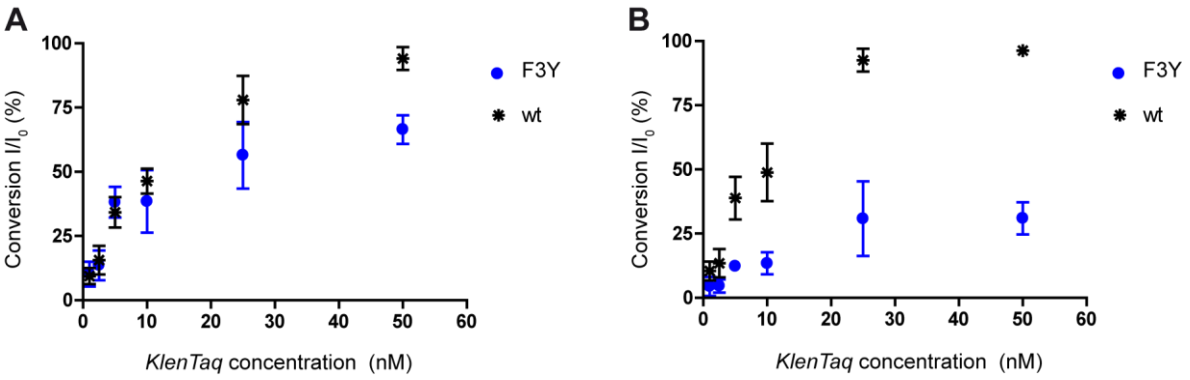


Figure 18. Incorporation of dAMP opposite natural template (dT) or abasic site analogue F shown in A) and B), respectively. Polymerase amounts were diluted in a step-wise manner. Primer conversion was plotted against the enzyme concentrations used. The incubation time was adjusted to 30 min for the abasic site containing template as opposed to 5 s for the natural template. Graphic was adapted from Blatter et al.[162]

We assume that the natural tyrosine's hydroxyl group in *KlenTaq* wild-type exhibiting a pk_a value of 9.9 is protonated at the reaction pH (7.5) employed. Thus, hydrogen bonding between Tyr671 and the N3 of the incoming dATP can occur. However, due to the lower pk_a value of 2,3,5-fluorotyrosine (pk_a: 6.4), reaction conditions presumably favour the deprotonated state of the hydroxyl group which disrupts the formation of the hydrogen bond preventing stabilization of the incoming nucleotide. To further corroborate these observations, single nucleotide incorporation studies opposite F were conducted in a time-dependent manner (**Figure 19**). An excess of enzyme in relation to the primer/template complex should exclude potential differences in the binding efficiencies of both enzymes.

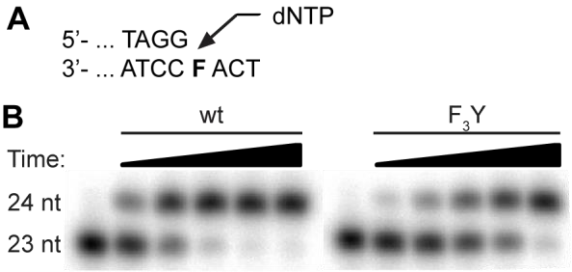


Figure 19. Time course experiment of dAMP incorporation opposite abasic site F with 30 s, 2 min, 5 min, 10 min and 30 min reaction time. DNA polymerase concentration employed was 1 µM. Primer is depicted in the first lane; nt: nucleotide. Graphic was adapted from Blatter et al.[162]

As observed in the previous experiment, the incorporation efficiency of the *KlenTaq* F_3Y variant was significantly lower opposite F compared to the wild-type enzyme, with 50 % primer conversion after 5 min instead of approximately 1 min for *KlenTaq* wild-type (**Figure 19**). The results corroborate previous findings of a Y671F mutant showing reduced dAMP incorporation efficiency opposite F compared to the wild-type and a lower incorporation rate for the adenine analogue d3DATP compared to natural dAMP. Thus, various functional studies indicate that the formation of a hydrogen bond between the hydroxyl group of Tyr671 and the N3 of the incoming adenine plays an important role in abasic site bypass. Furthermore, the experiments highlight the utility of fluorinated tyrosine analogues for investigating hydrogen bond formation at protein active sites.

1.2.3 Nucleotide Incorporation Opposite a Natural Abasic Site in the Template

Due to the high reactivity of natural abasic sites, a respective analogue is used in most studies. The tetrahydrofuran moiety implemented in this study, termed abasic site analogue F, was already described in the literature as a stable alternative to the natural lesion.[35, 167] Nevertheless, the following experiments should shine light on whether the tetrahydrofuran moiety provides the same results as a natural abasic site in case of *KlenTaq* DNA polymerase bypassing this lesion.

Single nucleotide primer extension reactions were set up using a template with either F or dUMP at the respective position in the DNA strand. In short, the respective primer/template complex was annealed and incubated with uracil-DNA glycosylase (UDG) for 18 h at 37 °C. The glycosylase cleaves the uracil moiety from the respective template strand resulting in a natural abasic site at this position.[168] The F-containing template was treated in the same manner to achieve identical reaction conditions. Without further purification, *KlenTaq* wild-type was added and the reactions were started by addition of the respective nucleotide (**Figure 20A, B**).

No difference was noted for the reaction with the natural abasic site compared to the tetrahydrofuran analogue F. The preference for purines was maintained and the incorporation efficiency of every nucleotide was comparable in both cases.

To confirm the presence of the natural abasic site, the template was radioactively labelled at the 5'-end, treated with UDG and subsequently analysed on a denaturing PAGE gel (**Figure 20C**). Analysis confirmed the reactivity of the natural abasic site, as degradation products were visible. However, the majority of the template was still intact. NaOH treatment of the UDG incubated template resulted in cleavage of the DNA, consistent with earlier findings that the abasic site is labile to basic conditions,[169, 170] subsequently confirming the presence of the natural abasic site.

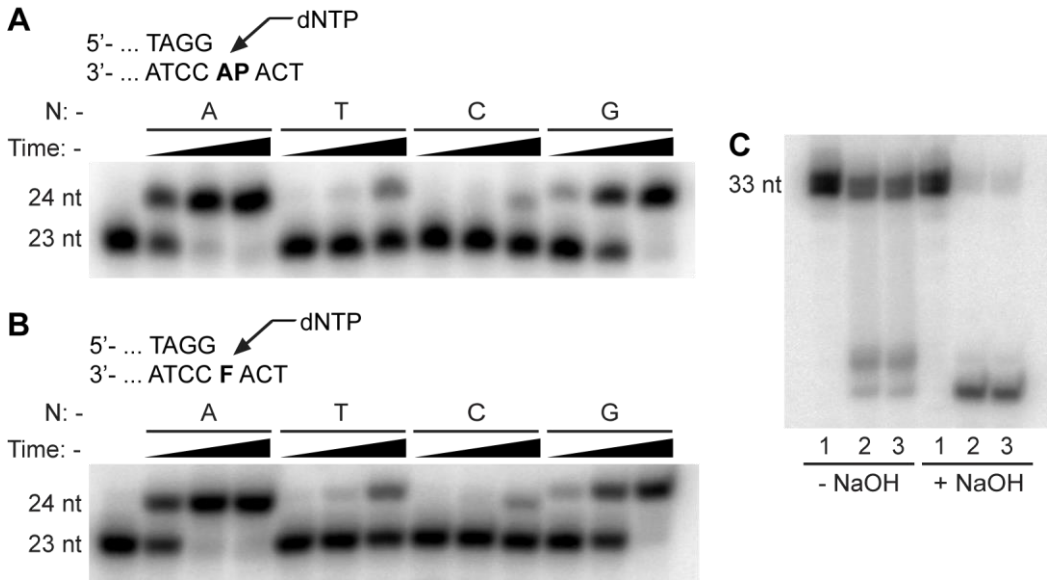


Figure 20. Nucleotide incorporation opposite natural abasic site. A-B) Single nucleotide incorporation of *KlenTaq* wild-type (500 nM) opposite natural abasic site (AP) or abasic site analogue F in A) and B), respectively. Partial primer/template sequence is shown. The respective dNTP is indicated (N). Primer is depicted in the first lane; nt: nucleotide. Reaction times were 2, 10 and 60 min. C) Detail of PAGE gel depicting the presence of AP. The 5'-[P^{32}]-radioactively labelled template (containing dUMP instead of F) was incubated without UDG (1), with UDG for 18 h (2) or with UDG for 18 h + the reaction time for nucleotide incorporation at 37 °C (3). Performed NaOH treatment of the DNA after the reaction with UDG is indicated.

1.3 Discussion

DNA polymerases from sequence families A and B follow the 'A-rule' when bypassing abasic sites with the mechanism underlying this bypass still under debate. From DNA polymerases belonging to these sequence families A and B, only crystal structures of RB69 DNA polymerase, a member from sequence family B, were available in complex with an abasic site containing template.[86, 87, 109, 111] However, recently Dr. Samra Obeid was able to obtain the first structure of a sequence family A member, *KlenTaq* DNA polymerase, bypassing an abasic site analogue.[72] *KlenTaq* DNA polymerase, the model system used in this study, follows the 'A-rule' and exhibits, similar to other DNA polymerases of this sequence family, an incorporation efficiency following the order of A>G>T≥C.

The structural data suggested a possible bypass mechanism as well as the mechanistic origin for the purine selectivity of this DNA polymerase. The vacant space created from the non-coding properties of the abasic site is filled by a tyrosine side chain (Tyr671) which may

assume the function of the templating nucleobase. Tyr671 was postulated to mimic in shape and size a pyrimidine nucleobase and thus direct for purine incorporation. Additionally, a hydrogen bond interaction is observed between the incoming ddATP and Tyr671, further stabilized by a hydrogen bond network formed between Arg587, the primer strand and the incoming nucleotide.[72]

Functional studies performed in this work provided evidence for the proposed mechanism and thus corroborated the structural data. Various *KlenTaq* variants, exhibiting a mutation at position 671, were constructed and analysed in primer extension as well as pre-steady state kinetic experiments regarding their incorporation efficiency opposite the abasic site. The most impressing experiment, investigating the properties of a tryptophane mutant (Y671W), demonstrated that the purine selectivity could be switched to preferential pyrimidine incorporation. This was allocated in primer extension as well as pre-steady state kinetic experiments. It is assumed that the bicyclic indole consisting of a six-membered ring fused to a five-membered ring mimics the approximate size and shape of a purine and in consequence directs for pyrimidine incorporation to conserve the geometric fit at the active site. The importance of the hydrogen bond interaction between the N3 of the incoming purine nucleotide and the hydroxyl group of Tyr671 was also highlighted by further kinetic data based on mutagenesis studies. A Y671F mutant showed a reduced incorporation efficiency opposite F compared to the wild-type, whereas the incorporation of d3DAMP opposite F also resulted in a decrease in incorporation efficiency compared to natural dAMP. Although these observations explain the preference of purines over pyrimidines, they cannot account for the preference of adenine over guanine, as the interaction observed for adenine between the N3 and the hydroxyl group of Tyr671 is also able to form in case of guanine. However, recently Obeid at al. successfully solved structures of *KlenTaq* DNA polymerase in complex with an abasic site analogue containing template and ddGTP in the active site which provided insights into the preferential incorporation of adenine over guanine.[74] The exocyclic C_2-NH_2 group of guanine seems to prevent an arrangement as observed for ddATP due to the geometric constraints present in the active site. Thus, a realignment of ddGTP had to occur which was observed in two different structures of ddGTP opposite an abasic site. The structures vary in the positioning of the incoming ddGTP in the active site and in the interaction pattern with Tyr671. Yet, in both structures the different interaction with Tyr671 compared to adenine results in an enlarged distance of the α-phosphate and the 3'-primer terminus. The degree of misalignment, measured in the distance between the α-phosphate and the 3'-primer terminus, was observed to equal the order of incorporation efficiency for the different nucleotides opposite an abasic site.

The overall decline in incorporation efficiency opposite an abasic site compared to dT observed for *KlenTaq* wild-type might also be explained by the unfavourable distance between the α-phosphate of the incoming nucleotide and the 3'-hydroxyl end of the primer as

well as the presence of only one Mg^{2+} ion observed in the crystal structure.[72] The distance between the α-phosphate and the 3'-primer terminus is shorter for ddATP than for ddGTP,[74] but still enlarged compared to the alignment in the natural case. This arrangement indicates that further changes as well as binding of a second metal ion in the active site have to occur in order to facilitate phosphodiester bond formation. The lower incorporation efficiency of replicative DNA polymerases in lesion bypass and consequently its pausing at sites of DNA damage may allow the cell machinery to act and provide a TLS enzyme which can overcome the lesion. Interestingly, another member of sequence family A, DNA polymerase θ, was found to bypass abasic sites with high activity preferentially incorporating dAMP opposite the lesion.[105] DNA polymerase θ is involved in the diversification of immunoglobulin (Ig) genes during somatic hypermutation[171] and thus, the error-prone bypass of abasic sites might be an evolutionary advantage, as it provides the possibility for physiological mutation in somatic hypermutation of Ig genes.

Homo sapiens (pol theta)	2513	RQQAKQICYGIIYG	2526
Mus musculus (pol theta)	2376	RQHAKQICYGIIYG	2389
Rattus norvegicus (pol theta)	2336	RQQAKQICYGIIYG	2349
Gallus gallus (pol theta)	2409	RQQAKQICYGIIYG	2422
Drosophila melanogaster(mus308)	1874	RNSTKQVCYGIVYG	1887
C. elegans (pol theta)	1489	RDAVKQLCYGLIYG	1502
Thermus aquaticus	659	RRAAKTINFGVLYG	672
Thermus thermophilus	661	RRAAKTVNFGVLYG	674
Thermus filiformis	659	RRAAKTVNFGVLYG	672
Deinococcus radiodurans	784	RRAAKTVNFGVLYG	797
Escherichia coli (K12)	754	RRSAKAINFGLIYG	767
Haemophilus influenzae	757	RRNAKAINFGLIYG	770
Streptococcus pneumoniae	715	RRNAKAVNFGVVYG	728
Mycobacterium tuberculosis	729	RRRVKAMSYGLAYG	742
Mycobacterium leprae	736	RRRVKAMSYGLAYG	749
Treponema pallidum	823	RRIAKTINFGIVYG	836
Chlamydia trachomatis	694	RYQAKAVNFGLVYG	707
Borrelia burgdorferi	734	RRIAKSINFGIIYR	746
Helicobacter pylori	711	RSIAKSINFGLVYG	724
Lactococcus lactis	703	RRNAKAVNFGVVYG	716
Rickettsia typhi	700	RRKAKAINFGIIYG	713
Streptomyces coelicolor	732	RRKIKAMSYGLAYG	745
Synechocystis sp.	796	RNLGKTINFGVIYG	809
Aquifex aeolicus	402	RQLAKAINFGLIYG	415
T3 DNA polymerase	518	RDNAKTFIYGFLYG	531
T5 DNA polymerase	562	RQAAKAITFGILYG	575
T7 DNA polymerase	518	RDNAKTFIYGFLYG	531

Figure 21. Amino acid sequence alignment of DNA polymerases highlighting the conserved position equivalent to Tyr671 in *KlenTaq* DNA polymerase. Graphic was adapted from Obeid, Blatter et al.[72]

So far the role of Tyr671, as a positioning device for the incoming dNTP in translesion synthesis, has not been shown before. Previous reports only demonstrated a potential role of Tyr671 and the corresponding Tyr766 residue in DNA polymerase I from *E. coli* in discriminating ribonucleotides and non-canonical nucleotides.[172-174] A sequence alignment of A family DNA polymerases showed a conservation of this residue throughout the evolution from prokaryotic to eukaryotic organisms, suggesting that this 'amino acid templating' mechanism in translesion bypass might be a general feature applied by other DNA polymerases of this sequence family as well (**Figure 21**).

Structural data for RB69 DNA polymerase in complex with an abasic site containing template suggests stacking interactions and partial charge interactions between the incoming nucleotide and the penultimate base pair adjacent to the abasic site as the basis for the 'A-rule'.[86, 87, 109] Our results reveal an 'amino acid templating' mechanism with a geometric fit to the active site as the major determinant for purine incorporation, further

supported by hydrogen bond formation. No stacking interactions are observed between the incoming dNTP and the primer strand. Admittedly, it cannot be excluded that stacking interactions play a role at states prior or later to the one resolved in the crystal structure.[72]
The importance of individual amino acid residues in abasic site bypass was already reported for polymerases of other sequence families, namely DNA polymerase β and DNA polymerase from yeast Rev1. Polymerase β employs an arginine residue at position 238 for adenine incorporation via hydrogen bond formation.[94] In Rev1 DNA polymerase, an arginine residue (Arg324) is again responsible for the nucleotide incorporation opposite an abasic site. The amino acid residue forms specific hydrogen bonds with the Watson-Crick edge of an incoming dCTP, thereby facilitating preferential dCMP incorporation opposite an abasic site. As purines, especially dGTP, are prone to depurination, dCTP incorporation was probably evolved as the best 'guess' when bypassing abasic sites.[92] Consequently, the role of side chain residues in lesion bypass was demonstrated before, but an 'amino acid templating' mechanism, where an amino acid residue adopts the shape and size of a nucleobase directing for preferred purine incorporation opposite an abasic site, has not been described so far. However, the importance of geometric constraints in abasic site bypass was already reported in literature and is in good concordance with our observations. In detail, Matray and Kool identified a pyrene nucleoside triphosphate as an optimal partner for nucleotide incorporation opposite an abasic site.[103] The fluorescent 'base' possesses the approximate size of one Watson-Crick base pair and is efficiently and specifically inserted by DNA polymerases opposite an abasic site, thus indicating geometric complementary as one of the major determinants for nucleotide incorporation.
In addition to the kinetic studies performed in this work, functional studies with a natural abasic site demonstrated the utility of the tetrahydrofuran analogue in studies of abasic site bypass, as it provided the same results as a natural abasic site employed in primer extension reactions. The allocation of a *KlenTaq* variant, featuring a fluorotyrosine analogue at position 671, from the group of Prof. A. Deiters further provided us with the possibility to investigate the role of hydrogen bonding in abasic site bypass and demonstrated the potential of fluorotyrosine analogues as a crucial tool to investigate hydrogen bonding in proteins in general.

2. Template-Independent Nucleotide Addition at Blunt-Ended DNA

2.1 Introduction

Numerous DNA polymerases are capable of catalysing template-independent nucleotide addition at blunt-ended DNA resulting in a single-stranded overhang at the 3' end of the primer.[107, 175, 176]

This process was suggested to play a role in the replication of telomeres of eukaryotic chromosomes,[177] during rearrangement of immunoglobulines[178] and was exploited in biotechnological applications e.g. in the ligation step during cloning (TOPO TA PCR cloning system, Invitrogen) by *Taq* DNA polymerase. Additionally, it was thought to contribute to the genomic hypermutability of retroviruses, as studies of HIV-RT showed, that the enzyme adds additional bases beyond the 5' end of the RNA template, resulting in misincorporation events in the replication cycle of the HIV genome.[175]

Although template-independent nucleotide addition at blunt-ended DNA was described for only a few DNA polymerases in the beginning, it soon become apparent that various enzymes of archaeal, bacterial, eukaryotic and viral origin are able to add nucleotides in a non-templated manner.[107, 176] Comparable to the bypass of a non-instructional lesion like an abasic site, many of these DNA polymerases showed preferential incorporation of dAMP at blunt-ended DNA. This was found for different sequence families as well (family A: *Taq* DNA polymerase; family B: polymerase α; family X: polymerase β; RT-family: AMV).[107] Functional and structural studies with DNA polymerase Dpo4, a member of the Y family, indicated base stacking interactions between the incoming nucleotide and the terminal bases of the duplex DNA as the major driving force for preferential dAMP incorporation.[176] A higher incorporation efficiency of a pyrene nucleotide with four aromatic rings supported these results, as the analogue possesses increased stacking abilities without the possibility of hydrogen bond formation. Other driving forces for the adenine selectivity, however, cannot be excluded, as further data for other DNA polymerase sequence families is still lacking.

2.2 Results

The mechanistic basis of template-independent nucleotide addition to the 3'-termini of blunt-ended DNA was discussed to be similar to abasic site bypass with the superior stacking ability of adenine as the major driving force. Thus, we wondered if the 'amino acid templating' mechanism we found for *KlenTaq* DNA polymerase in case of abasic site bypass may also apply here. First, single nucleotide incorporation studies of *KlenTaq* wild-type and blunt-ended DNA were conducted and a preferred incorporation of dAMP was observed (**Figure 22A, B**).

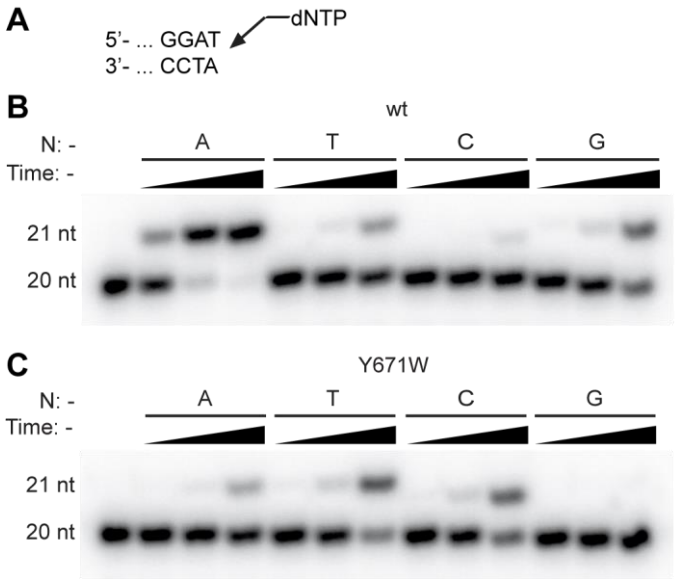


Figure 22. Primer extension experiments at blunt-ended DNA. A) Partial primer/template sequence is shown. B-C) Single nucleotide incorporation at blunt-ended DNA with *KlenTaq* wild-type (wt) and Y671W are shown in B and C, respectively. Reaction times were 2, 10 and 60 min. The respective dNTP (N) is indicated. Primer is depicted in the first lane. Graphic was adapted from Obeid, Blatter et al.[72]

Kinetic data confirmed these findings, revealing that dAMP incorporation is about 19-fold more efficiently incorporated compared to dGMP. Furthermore, pyrimidine incorporation was even less efficient with a 38-fold and 280-fold reduced efficiency for TMP and dCMP compared to dAMP, respectively (**Table 2**).

Table 2. Transient kinetic analysis of nucleotide incorporation at blunt-ended DNA by *KlenTaq* wild-type and Y671W.

Enzyme	dNTP	k_{pol} [s^{-1} x 10^{-2}]	K_d [μM]	k_{pol} / K_d [μM^{-1} s^{-1} x 10^{-4}]
wild-type	A	3.11 ± 0.08	275 ± 13	1.13
wild-type	G	0.21 ± 0.01	371 ± 39	0.06
wild-type	T	0.13 ± 0.01	503 ± 98	0.03
wild-type	C	0.07 ± 0.02	1729 ± 620	0.004
Y671W	A	0.12 ± 0.01	450 ± 42	0.03
Y671W	G	n.a.	n.a.	n.a.
Y671W	T	0.52 ± 0.03	747 ± 69	0.07
Y671W	C	0.44 ± 0.04	746 ± 108	0.06

n.a.: not accessible since the turnover after 1h using 600 μM dNTP was less than 20%.

If template-independent incorporation at blunt-ended DNA also relies on the proposed 'amino acid templating' mechanism, it should be possible to switch the preferred purine selectivity for pyrimidine incorporation employing the Y671W mutant. Indeed, primer extension experiments and pre-steady state kinetics corroborated these observations (**Figure 22C, Table 2**)**.** TMP and dCMP are about two-fold more efficiently incorporated than dAMP by the Y671W mutant, whereas an incorporation of dGMP could not be observed even after prolonged incubation times.

Although, the kinetic data already hints at the important role of Tyr671 in template-independent addition of nucleotides at blunt-ended DNA, crystallisation trials were set-up by Dr. Samra Obeid using *KlenTaq* wild-type and a blunt-ended DNA duplex. Crystals were obtained and the structure was solved to a resolution of 2.2 Å (*KlenTaq*$_{BE}$).

The overall structure was very similar to structures with a cognate template, but the O helix in the fingers domain was again, as observed for *KlenTaq* in complex with a F-containing template, in a half-open conformation. This conformation allows the Tyr671 to flip in and assume the role of a positioning device for the incoming ddATP, mimicking a pyrimidine (**Figure 23**). The same 'amino acid templating' mechanism observed for a non-instructive lesion, therefore mediates template-independent nucleotide incorporation. A mechanism corroborated both by kinetic and crystal structure analysis.

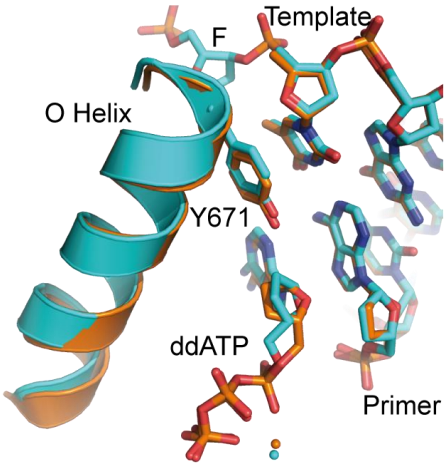


Figure 23. Close-up view of the active site in *KlenTaq*$_{BE.}$ Overlay of *KlenTaq*$_{AP}$ (cyan) and *KlenTaq*$_{BE}$ (orange) structure reveals similar arrangements of residues. Depicted are residues Y671, the O helix and the incoming ddATP opposite abasic site analogue F or on top of the primer/template duplex. Water molecules are shown as spheres. Graphic was adapted from Obeid, Blatter et al.[72]

2.3 Discussion

Many DNA polymerases catalyse template-independent nucleotide addition at blunt-ended DNA with a strong preference for incorporation of adenine. Structural data from the Dpo4 DNA polymerase indicate that the superior base stacking properties of adenine are the major reason for the preferred dAMP incorporation at blunt-ended DNA.[176] One can only assume that this mechanism is also the driving force for dAMP incorporation opposite an abasic site by this DNA polymerase. So far, crystal structures of Dpo4 bypassing an abasic site are only available for the looping out mechanism[85], which the DNA polymerase can employ, and not for the second competing pathway in which dATP is preferentially inserted opposite an abasic site. Crystal structure analysis of *KlenTaq* DNA polymerase in complex with blunt-ended DNA revealed no stacking interactions between the incoming dNTP and the DNA duplex.[72] Moreover, crystal structure and kinetic analysis strongly support the notion that the mechanism underlying nucleotide incorporation at blunt-ended DNA and abasic site bypass are the same for *KlenTaq* DNA polymerase. In both cases, the amino acid residue Tyr671 flips in and assumes the role of the templating nucleobase. It directs for preferential purine incorporation, thereby providing an optimal geometric fit for the active site. This conclusion is corroborated by the findings that mutation of the templating tyrosine to tryptophan switches the selectivity for purines to pyrimidines. However, we cannot exclude that stacking interactions play a role at states prior or later to the one resolved in the crystal structure. To elucidate if abasic site bypass and template-independent nucleotide addition are correlated in other DNA polymerases as well, further structural data has to be available. But the fact that an abasic site as well as blunt-ended DNA exhibit non-coding properties supports this model from a structural point of view.

III. Results and Discussion – Directed Evolution of *KlenTaq* Variants Capable of Reverse Transcription

1. Generation and Characterisation of *KlenTaq* Variants with Reverse Transcriptase Activity

1.1 Introduction

The polymerase chain reaction (PCR) is a key process in many biotechnological applications and relies on the ability of DNA polymerases to specifically recognize and accept its nucleic acid substrate and promote DNA synthesis.[160] However, the number of applications that requires DNA polymerases to accept non-cognate substrates continuously increases. Thus, DNA polymerases with loosened specificity and an expanded substrate spectrum are in demand and are largely accessible by means of directed evolution (see chapter **I 4**). The acceptance of non-cognate substrates by DNA-dependent DNA polymerases also includes the usage of RNA as template for DNA synthesis. Thermostable DNA polymerases capable of reverse transcription and PCR would provide a crucial tool for RT-PCR, a fundamental technique applied in molecular biology and clinical diagnostics (see chapter **I 5.3**).[14, 142] This part of my work focused on the development of such thermostable DNA polymerase variants exhibiting reverse transcriptase and PCR activity in one protein scaffold.

Our group previously introduced an evolved *KlenTaq* variant, termed KTq M1 (L322M, L459M, S515R, I638F, S739G, E773G), exhibiting limited reverse transcriptase activity, while simultaneously maintaining its DNA polymerase properties.[148] The enzyme, generated by error-prone PCR, comprises six mutations distributed over the whole protein scaffold. The contributions of the individual mutations regarding the increased reverse transcriptase activity were not elucidated. Furthermore, the application of the respective full-length *Taq* enzyme in one-step detection of pathogenic RNA using *Taq*Man probes was reported in the literature.[150]

However, the demand for DNA-dependent DNA polymerases with increased reverse transcriptase activity persists. Therefore, our aim was to further evolve the M1 variant towards an enzyme exhibiting superior reverse transcriptase activity compared to the parental enzyme, while still retaining its DNA polymerase activity. Consequently, *KlenTaq* variant M1 was recombined by DNA shuffling with a second variant *KlenTaq* M747K (KTq M747K), already described in the literature to possess an increased substrate spectrum.[159] Substitution of the methionine with a positively charged lysine resulted in an increase of the positively charged surface potential in close proximity to the negatively charged backbone of the DNA. It was suggested that the enhanced electrostatic interaction between the enzyme and the nucleic acid backbone promotes the acceptance of aberrant substrates and results in the increased lesion bypass activity of this variant.[73, 159] Furthermore, the mutant was also reported to exhibit some reverse transcriptase activity.[156] However, its potential application in RT-PCR had never been investigated.

Thus, a combination of the two variants, M1 and M747K, was promising to yield mutants with new properties. As structural data regarding the role and function of the M1 mutations contributing to the increased reverse transcriptase activity was lacking, DNA shuffling was the method of choice for recombining both enzymes in order to get all possible mutation combinations between the M747K mutation and the six mutations from KTq M1.

1.2 Results

1.2.1 Library Generation via DNA Shuffling

KTq M1 comprises six mutations distributed over the entire protein scaffold. Together with KTq M747K, this results in seven positions being either mutated or displaying the wild-type amino acid sequence. Mutation sites of both variants are depicted in **Figure 24**.

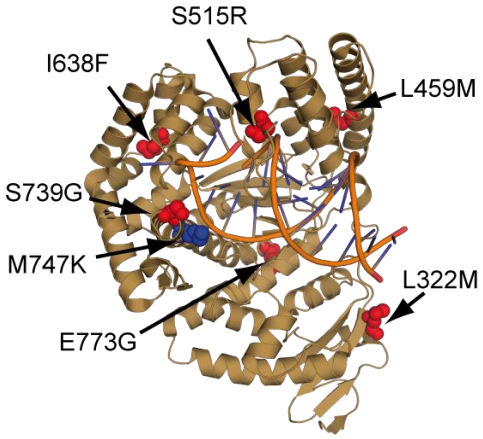


Figure 24. Structure of *KlenTaq* DNA polymerase with mutation sites of M1 and M747K highlighted in red and blue, respectively. Mutation sites are shown as spheres. Graphic was adapted from Blatter et al.[179]

Consequently, 2^7 = 128 mutation combination possibilities exist in total. A library size of approximately 1200 clones is required to have a 99 % probability to sample all possible variants, presuming that every mutation combination is equiprobable.[180] Taking this into consideration, a library of 1570 clones was generated in this study to obtain high library coverage. Calculations were based on simple statistical analysis provided by Patrick et al.[180] Different established protocols from the literature had to be combined and optimized to achieve efficient shuffling of the three genes (KTq wt, KTq M1 and KTq M747K).[139, 181, 182]
First, parental DNA was amplified in a nested PCR using a high-fidelity enzyme (Phusion™ DNA polymerase). Primers were designed to anneal approximately 160 bp up- and downstream to

the KTq gene. Amplified parental DNA was then digested with DNase I in the presence of $MgCl_2$. After 1 min incubation time at 15 °C, the reaction was stopped by the addition of EDTA and simultaneous heat-inactivation of DNase I at 65 °C for 10 min (**Figure 25A**). The digestion step yielded DNA fragments between 50 bp and 200 bp. DNA in the negative control was mixed with DNase I and EDTA simultaneously with subsequent heat-inactivation. The fragmentation visible in the negative control is probably due to the high activity of the enzyme even present during the heat-inactivation step (**Figure 25A**). The use of Mg^{2+} instead of Mn^{2+} during the digestion step seemed to promote recombination, presumably due to the fact that DNase I cleaves double-stranded DNA in a statistically random fashion in the presence of Mg^{2+} and forms no blunt-ended DNA.[183]

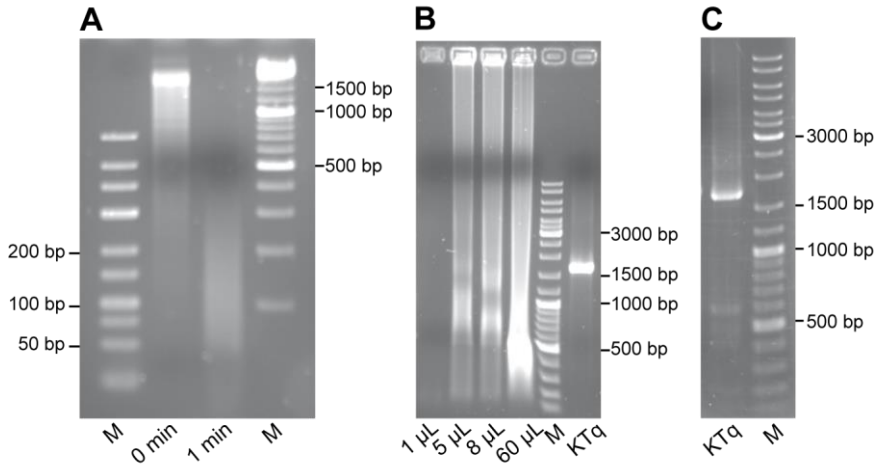


Figure 25. Recombination of KTq wild-type, KTq M1 and KTq M747K genes. A) Parental DNA after digestion with 0.004 u/µL DNase I for 0 (negative control) and 1 min at 15 °C. B) Reassembly of 1, 5, 8 or 60 µL of digested fragments in a PCR-like reaction without primers (total reaction volume: 50 or 100 µL). C) Product formation after amplification of the reassembled product by standard PCR (optimized conditions). M: Marker; KTq: Amplified *KlenTaq* gene (1645 bp).

Without further purification,[182] DNA fragments were reassembled in an Assembly PCR using no primers, 60 cycles and a low annealing temperature of 45 °C to promote recombination.[182] Agarose gel analysis depicted the formation of longer extension products (**Figure 25B**). Modification of this step revealed less point mutations when using Phusion™ instead of *Taq* DNA polymerase. Due to the fact that we were particularly interested in the effect of the M1 mutations regarding the reverse transcriptase activity, new mutations resulting from misincorporation events caused by *Taq* DNA polymerase were to be avoided. Therefore, the use of the Phusion™ DNA polymerase exhibiting a high accuracy in DNA synthesis offered the

opportunity for reassembling the fragments without introducing additional mutations. Different volumes (1, 5, 8, 60 µL) of fragmented genes were utilized in the Assembly PCR with 5 µL in a total reaction volume of 50 µL yielding the best results after amplification.
A final PCR was performed using again Phusion™ DNA polymerase and primers with restriction sites for cleavage with *SphI* and *HindIII*. Conditions were optimized to minimize the amplification of by-products (**Figure 25C**). The amplified product containing shuffled full-length *KlenTaq* genes was gel extracted and digested with the respective restriction enzymes *SphI* and *HindIII*. After cloning the purified product into pGDR11[184] vector, the ligation reaction was transformed into *E. coli* BL21 (DE3) cells and colonies were picked randomly from selective agar plates to generate the library.

1.2.2 Library Screening and Identification of *KlenTaq* Variants

For initial screening, *KlenTaq* variants of the generated library were expressed in 96-deep well plates at 37 °C for 4 h and harvested afterwards. Pellets were lysed using lysozyme and digested with DNase I at 37 °C for 15 min. Addition of DNase I during the lysis step should ensure the removal of bacterial DNA which might later on influence the screening process. After heat-denaturation of host proteins as well as heat-inactivation of DNase I at 75 °C for 45 min, centrifugation was conducted to remove bacterial cell debris. Afterwards, the lysates were directly tested in an activity screening, performed in 384-well plates, using a real-time PCR activity screen with DNA as template. A 90 nt template was employed in PCR and amplified by the enzyme present in the respective lysate based on an already established protocol in our group.[132] DNA amplification was visualized by binding of SYBRGreen I to double-stranded DNA. Approximately 80 % of the expressed enzymes were found to be PCR active, thus, a selection step was omitted and subsequently all clones were screened for reverse transcriptase activity based on a method described earlier.[148] In short, MS2 bacteriophage RNA (50 pg/µL) was used as template in real-time RT-PCR performed in 96-well plates. First, reverse transcription was conducted using an initial denaturation step of 30 s at 95 °C, an annealing step at 55 °C for 35 s and elongation for 15 min at 72 °C. Additionally, 50 PCR cycles were performed with 30 s at 95 °C, 35 s at 55 °C and 40 s at 72 °C. The variants, capable of transcribing the RNA into DNA and subsequent amplification of the DNA in PCR, were identified by the fluorescence signal of SYBRGreen I binding to double-stranded DNA (**Figure 26**).

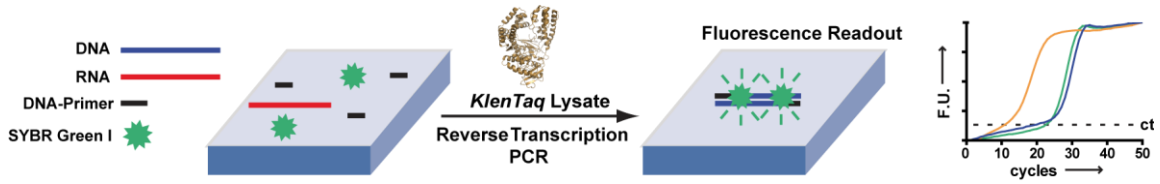


Figure 26. Scheme of real-time RT-PCR screening strategy. RNA template is reverse transcribed and the resulting DNA amplified by a RT-PCR active *KlenTaq* variant. The reaction is followed in real-time based on SYBRGreen I fluorescence upon binding to double-stranded DNA. Hits (orange) were defined as *KlenTaq* variants with a crossing threshold (ct)-value lower than the ct-value of the parental enzymes M1 (green) and M747K (blue). F.U.: fluorescence unit.

30 % of the expressed enzymes exhibited reverse transcriptase activity comparable or increased to the parental enzymes M1 and M747K and were subjected to another round of screening. The second round was performed under the same conditions in order to corroborate the results from the first run. However, only mutants with a crossing threshold (ct-) value lower than the parental enzymes were selected in this screening round. These variants were combined on one 96-well plate in order to provide normalised reaction conditions and further screened for reverse transcriptase activity under more challenging conditions, including a reduction of the reverse transcription time (down to 7.5 min) in a first screening and a reduction of RNA concentration (down to 5 pg/µL, 15 min) in a second screening. These experiments were performed twice with lysates resulting from two different expression cultures. Variants exhibiting the best results, namely the lowest ct-value, in all experiments (initial conditions, reverse transcription time of 7.5 min and 5 pg/µL RNA) were selected and sequenced at GATC-Biotech (Konstanz, Germany). Mutation combinations and occurrence of those are depicted in **Figure 27.**

The mutation combination L459M, S515R, I638F and M747K was found to be present in eight out of nine selected variants. In four out of these eight variants this mutation combination was present without additional mutations. The remaining four variants exhibited the mutation combination in addition to L322M or L322M and E773G. Only one mutant, the least active one of all, showed three mutations (S515R, I638F and M747K) with the L459M missing. Interestingly, the S739G mutation was not present in all variants identified. The different variants were selected for further studies and termed RT-KTq 1-4 (see **Figure 27**). One mutant that exhibited the mutation combination present in RT-KTq 4 additionally comprised the point mutation A798D. The variant was also investigated in the subsequent studies, but no differences were recognizable compared to RT-KTq 4. Thus, the corresponding data will be neglected in the following chapters. For further comparison with the identified variants, all seven mutations, six mutations from M1 and the M747K mutation, were combined in one protein scaffold (KTq M1/M747K) by site-directed mutagenesis.

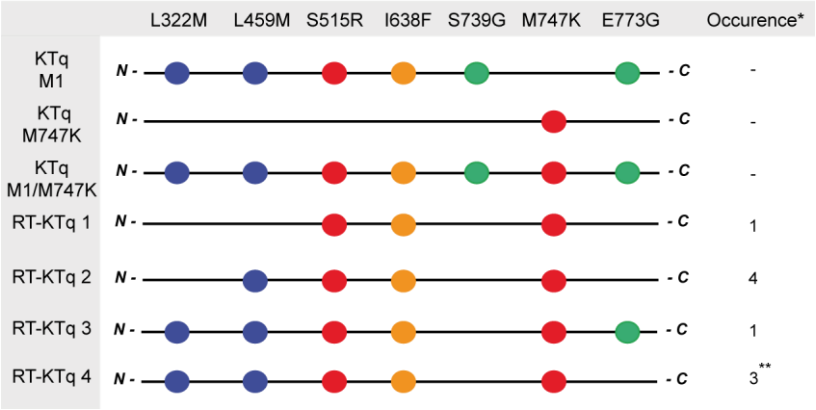


Figure 27. Mutation combinations of the parental enzymes KTq M1 and M747K, KTq M1/M747K and all sequenced variants are depicted. Mutations are colour-coded with a mutation to methionine in blue, positively charged mutations in red, mutations to glycine in green and the hydrophobic mutation to phenylalanine in orange.
* The occurrence defines the number of clones (out of nine sequenced hits) exhibiting the same mutation combination.
** Point mutation A798D was additionally present in two out of three clones.

KTq wild-type, parental enzymes M1 and M747K, KTq M1/M747K and the identified *KlenTaq* variants were expressed, purified via His-tag affinity chromatography, adjusted to equal protein concentrations (**Figure 28A**) and characterized regarding their properties.

1.2.3 Characterisation of Reverse Transcriptase Activity

In order to gain more insights into the reverse transcriptase activity of the protein variants, all enzymes were examined in reverse transcription primer extension experiments. In detail, a 5'-[^{32}P]-radioactively labelled DNA primer was annealed to an RNA template and the complex subsequently employed as substrate. Reactions were incubated at 72 °C and analysed on a denaturing PAGE-gel (**Figure 28B**).
In the control experiment, KTq wild-type and all variants catalysed full extension of the primer when a DNA template was used. Template-independent addition of one nucleotide, as described in the literature for DNA polymerases,[72, 107] was perceived for all enzymes in case of full-length product formation. Using RNA as template, KTq M747K incorporates up to three nucleotides, whereas minimal full-length product formation was observed for KTq M1 after 5 min incubation time (**Figure 28C**). However, KTq M1/M747K and all variants identified showed almost quantitative formation of full-length product, thus exhibiting an efficiency by

far exceeding the parental enzymes. Primer elongation for the wild-type enzyme using RNA as template was not observed (**Figure 28C**).

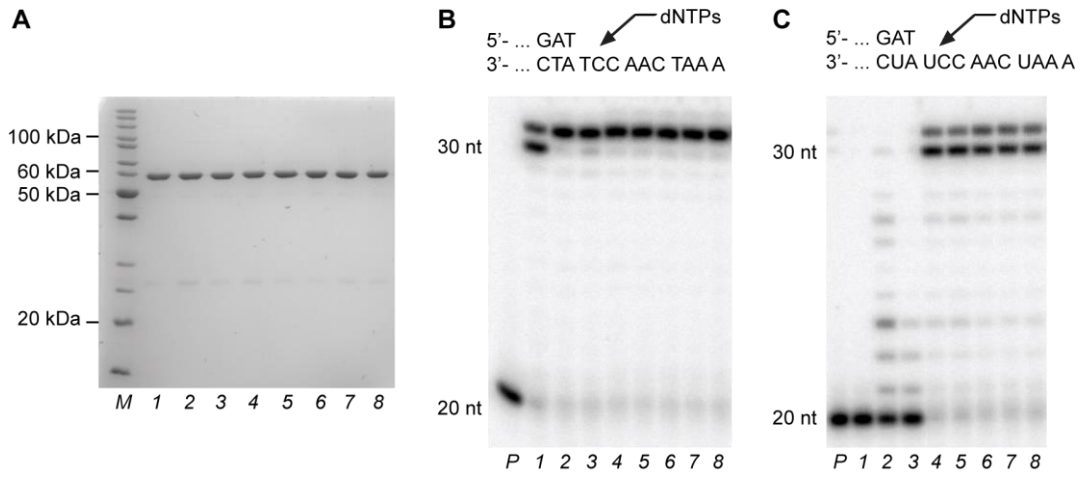


Figure 28. Primer extension experiments with KTq wt (1), M1 (2), M747K (3), M1/M747K (4), RT-KTq 1 (5), RT-KTq 2 (6), RT-KTq 3 (7) and RT-KTq 4 (8). A) SDS-PAGE of purified KTq wt and variants. M: Marker B) Primer extension reactions with 5'-[^{32}P]-labelled 20 nt primer annealed to a DNA (left) or an RNA (right) template. Reactions were catalysed by KTq wt and variants. Reactions with DNA or RNA were incubated for 1 or 5 min, respectively. Partial primer/template sequence is depicted on top.

Quantification of the respective activities[185] on a DNA- and RNA-template confirmed the superiority of the identified *KlenTaq* variants and KTq M1/M747K compared to the parental enzymes (**Table 3**). For activity measurements, primer extensions were performed and dNTP conversion per min determined and plotted against polymerase amounts used; the resulting slope yielding the specific activity of the respective enzyme (**Figure S2** in chapter **VII 1.2**). The specific activities for DNA as template showed a slight increase in activity for all identified mutants except RT-KTq 1 compared to KTq wild-type, but were in the same order of magnitude as observed for the parental enzymes. However, the efficiency of accepting RNA as template was increased manifold (at least by a factor of ten) in the new variants, demonstrating superior reverse transcriptase activities compared to the previously evolved enzymes M1 and M747K. Interestingly, the specific activities of the identified variants and KTq M1/M747K are in the same range when using RNA as template demonstrating that the additional mutations present in M1/M747K, RT-KTq 3 and 4 showed no further effect on reverse transcriptase activity. Only RT-KTq 1 exhibits a slightly reduced overall activity compared to the other identified variants when using either DNA or RNA as template.

Table 3. Specific activities of KTq variants on DNA- or RNA-template.

Variant	Specific Activity[a] [min^{-1}] DNA template	Specific Activity [min^{-1}] RNA template
KTq wt	145 ± 14	n.a.[b]
KTq M1	252 ± 21	1.67 ± 0.05
KTq M747K	279 ± 18	0.31 ± 0.01
KTq M1/M747K	342 ± 11	32.2 ± 1.8
RT-KTq 1	181 ± 12	15.7 ± 0.6
RT-KTq 2	349 ± 13	34.9 ± 1.3
RT-KTq 3	310 ± 17	28.6 ± 1.4
RT-KTq 4	322 ± 15	25.4 ± 1.2

[a] Quantified conversion of dNTPs per minute was plotted against varying polymerase amounts used; the resulting slope yielding the specific activity of the respective enzyme. [b] n.a.: not accessible, as no significant activity on an RNA template was observed.

In summary, by the recombination of two *KlenTaq* mutants, a new generation of *KlenTaq* variants was obtained that feature an up to 20-fold gain in reverse transcriptase activity compared to the parental variant KTq M1 and an up to 100-fold increase compared to KTq M747K. If an additive effect between the mutations of the parental enzymes would be in effect, a two-fold increase was to be expected. Interestingly, the results suggest an interaction between the individual mutations in a non-additive manner, resulting in a synergistic effect of the mutations. A phenomenon gaining increasing awareness, recently (see Discussion, chapter **III 1.3** for detailed information).[186]

1.2.4 RT-PCR

The increased reverse transcriptase activity of the identified variants compared to the parental enzymes was confirmed by applying the purified enzymes in real-time RT-PCR experiments using MS2 bacteriophage RNA as template (**Figure 29A**). The experimental set-up was based on the protocol employed in the screening (chapter **III 1.2.2**). With as low as 5 nM polymerase concentration, the identified *KlenTaq* variants and KTq M1/M747K showed product formation approximately 10 cycles earlier than the parental enzymes KTq M1 and M747K. Expected product formation was analysed on an agarose gel (**Figure 29**). As expected, the wild-type enzyme did not show any product formation. By-product formation was also visible, presumably related to the employed sequence context, as the by-product was also formed in the positive controls using DNA (**Figure 29B**). A rise in fluorescence was also noticed in case of negative controls using water instead of RNA. However, the ct values were

much higher and probably due to unspecific product formation as observed via agarose gel analysis (**Figure 29C**).

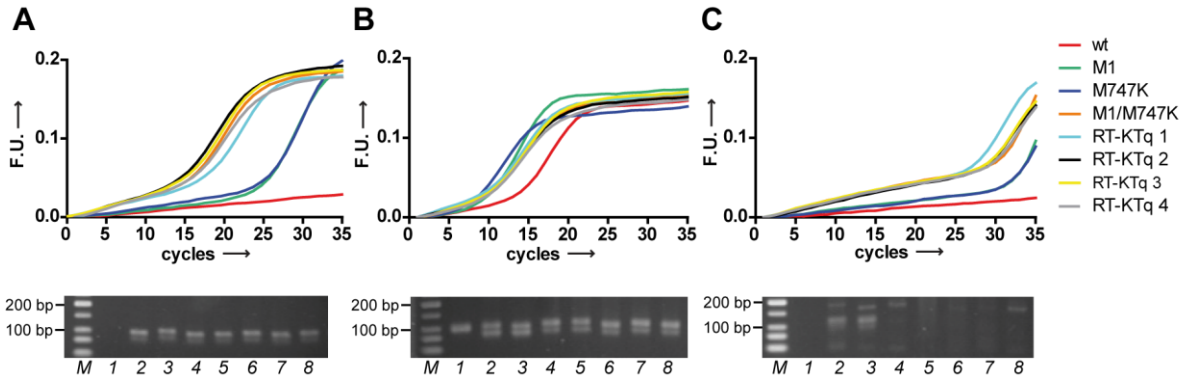


Figure 29. Real-time RT-PCR performed with KTq wild-type (wt) and variants (as indicated); F.U. = fluorescence units. The fluorescence readout was based on the binding of SYBRGreen I to double-stranded DNA. A) Amplification of a 100 bp target sequence from MS2 bacteriophage RNA. B) Amplification of a 100 bp target sequence from the respective DNA template. C) Negative control using water instead of RNA.
Product formation was analysed on a 2.5 % agarose gel as depicted below, with wt in lane 1, M1 in lane 2, M747K in lane 3, M1/M747K in lane 4, RT-KTq 1 in lane 5, RT-KTq 2 in lane 6, RT-KTq 3 in lane 7 and RT-KTq 4 in lane 8. M: Marker.

To further investigate the potential of the identified variants in RT-PCR, 16S- and 23S-ribosomal RNA was used as template in RT-PCR to test another sequence context. Again, the identified *KlenTaq* variants and KTq M1/M747K showed product formation at a ct value more than 10 cycles lower than the parental enzymes (**Figure 30A**). Interestingly, correct product formation was also visible in the negative control, implicating that 16S- and 23S-ribosomal RNA from *E. coli* was not completely removed during purification of the enzymes, and thus was reverse transcribed by the enzymes during the reaction (**Figure 30B**).

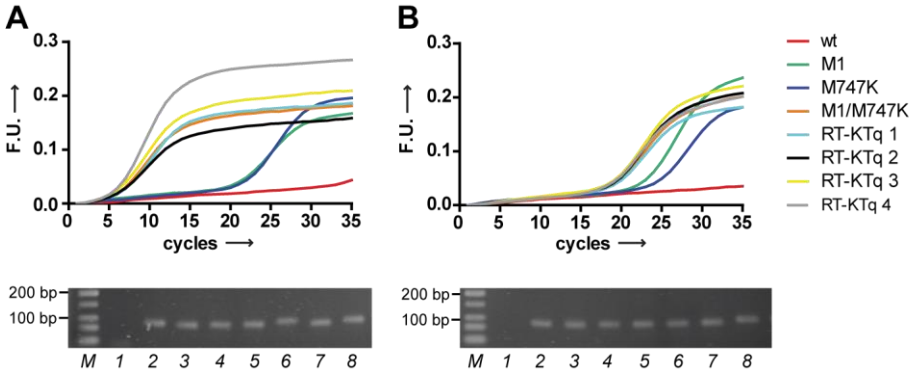


Figure 30. Real-time RT-PCR performed with KTq wild-type (wt) and variants (as indicated); F.U. = fluorescence units. The fluorescence readout was based on the binding of SYBRGreen I to double-stranded DNA. A) Amplification of a 100 bp target sequence from 16S- and 23S-ribosomal RNA from *E. coli*. B) Negative control using water instead of RNA.
Product formation was analysed on a 2.5 % agarose gel as depicted below, with wt in lane 1, M1 in lane 2, M747K in lane 3, M1/M747K in lane 4, RT-KTq 1 in lane 5, RT-KTq 2 in lane 6, RT-KTq 3 in lane 7 and RT-KTq 4 in lane 8. M: Marker.

In both set-ups (**Figure 29** and **30**), no difference in reverse transcriptase activity was observed between the identified mutants themselves and KTq M1/M747K.
Next, the amplicon size which can be generated in RT-PCR by the new variants was studied and compared to the parental enzymes (**Figure 31**).

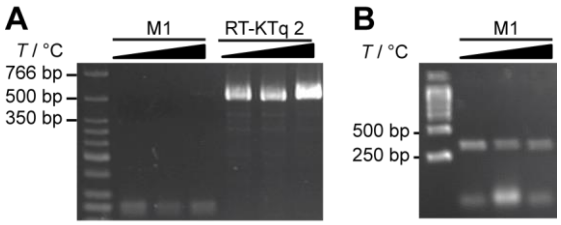


Figure 31. Amplicon size obtained in RT-PCR. A) KTq M1 and RT-KTq 2 amplifying a 510 bp fragment from MS2 bacteriophage RNA. B) KTq M1 amplifying a 321 bp fragment from MS2 bacteriophage RNA. The annealing step was performed at 65, 68 and 70 °C, respectively (T/°C).

RT-KTq 2, exhibiting the four mutations deemed important for the reverse transcriptase activity, and KTq M1 were selected for this experiment. As shown in **Figure 31A**, RT-KTq 2 is able to generate significantly longer products than the parental enzyme KTq M1. Due to the lower reverse transcriptase activity of KTq M1 compared to RT-KTq 2, the reverse transcription step was extended to 30 min reaction time. But product formation was only observed for RT-KTq 2 in case of an amplicon with 510 bp (**Figure 31A**). Product formation by KTq M1 failed and was only noticed for an amplicon reduced to 321 bp (**Figure 31B**).

1.2.5 Selection of RT-KTq 2

RT-KTq 2 was selected for further in depth-analysis due to its increased reverse transcriptase activity compared to the parental enzymes and the minimal number of mutations compared to the other identified variants. Indeed, RT-KTq 1 has one mutation (L459M) less, however, this mutant exhibited a two-fold reduced overall activity (**Table 3,** chapter **III 1.2.3**). Furthermore, the addition of more mutations, featured in RT-KTq 3 and 4, does not seem to result in another gain in reverse transcriptase activity, further emphasizing the importance of the four mutations present in RT-KTq 2. Additionally, RT-KTq 2 exhibits a higher stability compared to KTq M1/M747K, the mutant comprising all mutations. Thermal denaturation was performed and followed by CD spectroscopy resulting in melting temperatures of 100.1 ± 0.1 °C for RT-KTq 2 and 97.9 ± 0.1 °C for KTq M1/M747K (**Figure 32**).

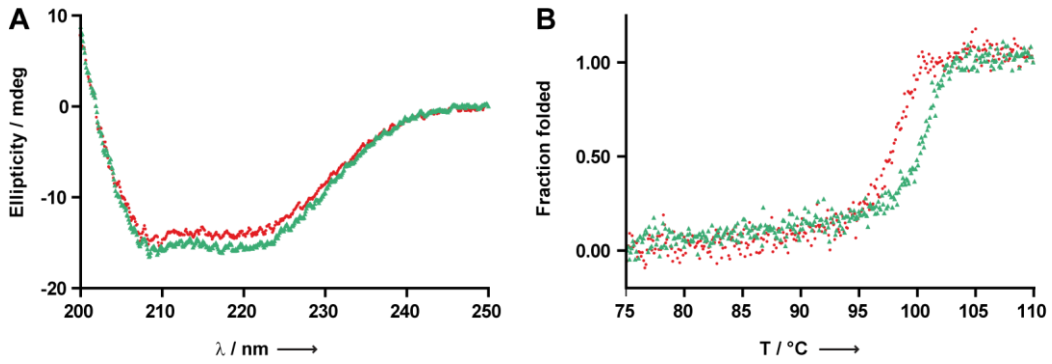


Figure 32. Stability of KTq M1/M747K and RT-KTq 2. a) CD-spectra of KTq M1/M747K and RT-KTq 2 are shown in red and green, respectively. b) Thermal denaturation of KTq M1/M747K and RT-KTq 2 was measured following the ellipticity at 209 nm. Melting temperatures were determined in two separate experiments. Graphic was adapted from Blatter et al.[179]

1.2.6 Influenza A and B Detection: Multiplex RT-PCR

Influenza virus was targeted in RT-PCR, to evaluate the application of RT-KTq 2 in clinical diagnostics. As a genetically highly dynamic virus known for considerable antigenic variation,[187] influenza viruses have been the cause for several pandemics.[188] Consequently, rapid detection methods are required for identifying new outbreaks in time and subsequent medical intervention. Along these lines, RT-PCR methods were established for the detection of influenza viruses, as they provide the ability to distinguish between influenza A and B, enable high sample throughput and are easily adapted to novel targets.[189] In quantitative real-time PCR, *Taq*Man probes are commonly used as they ensure high reaction specificity.[14] The *Taq*Man Assay relies on the 5´–3´ nuclease activity of a DNA polymerase to cleave a dual-labelled *Taq*Man probe during hybridization to the complementary target sequence and fluorophore-based detection (for further details see chapter **III 3.1**.). *Taq*Man-based multiplex real-time RT-PCR was selected in this experiment to detect influenza viruses A and B simultaneously. As 5'-nuclease activity is required for the hydrolysis of dual-labelled *Taq*Man probes, an enzyme blend of *Taq* DNA polymerase wild-type and RT-KTq 2 was employed in this assay.

Extracted RNA was obtained from respiratory swab samples from patients known to be either influenza A or B positive. These were kindly provided by Labor Dr. Brunner, Konstanz (Germany). The samples were analysed in the FAM and HEX channel using specific *Taq*Man probes consisting of an oligonucleotide conjugated to a 5' reporter dye 6-carboxyfluorescein (FAM) and a 3' minor groove binder (MGB) for influenza A detection and an oligonucleotide with a 5' reporter dye hexachlorofluorescein (HEX) and a 3' black hole quencher (BHQ-1) for influenza B detection (**Figure 33**). *Taq*Man probes conjugated to minor groove binders,[190] as well as optimized primers recommended for influenza A detection, were employed to further increase the specificity of the reaction. Each reaction set-up contained both primer pairs targeting influenza A and B, respectively, both dual-labelled probes and either extracted RNA from influenza A or B positive samples. Cycling was performed with 60 s initial denaturation at 95 °C and 50 cycles of two-step cycling with denaturation for 15 s at 95 °C and combined annealing/extension for 60 s at 60 °C. An additional reverse transcription step was omitted, further promoting time reduction. As shown in **Figure 33,** detection of both viruses, influenza A and B, was highly specific since only a signal increase in the expected fluorescence channel was observed (detection of influenza A and B in the FAM and HEX channel, respectively). Furthermore, template dilutions ranging from undiluted to 100-fold diluted extracts resulted in a positive fluorescence signal, although the signal for influenza B detection decreased from higher to lower amounts of RNA (**Figure 33**).

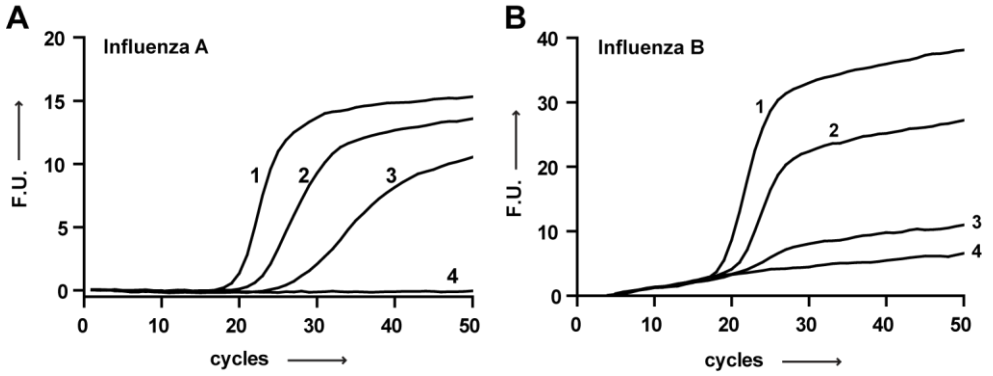


***Figure 33*.** Detection of influenza viruses A and B from RNA extracts in multiplex RT-PCR. A) Detection of influenza A in the FAM channel. B) Detection of influenza B in the HEX channel. A template dilution series was employed in both cases: undiluted (1), 1:10 (2), 1:100 (3) and no RNA as control (4).

The observed decrease in total fluorescence for influenza B detection indicates unspecific priming. However, further optimization of primers used for influenza B detection might enhance the reaction specificity. The differentiation between influenza A and B was still highly specific as described above. The experiment demonstrates the potential of RT-KTq 2 for diagnostic applications, as it enables rapid and specific detection of influenza A and B detection in multiplex RT-PCR.

1.3 Discussion

In summary, thermostable DNA polymerases combining reverse transcriptase and PCR activity in one protein scaffold were developed, thus facilitating their use in various applications in molecular biology or clinical diagnostics. The new DNA polymerase variants were discovered by DNA shuffling of a previously reported variant showing some reverse transcriptase activity with a variant that has an enhanced substrate spectrum.

The recombination of both enzymes via DNA shuffling was based on already existing protocols reported in the literature.[137, 139, 181, 182] As additional point mutations were not desired, the protocol was modified in order to reduce the error-rate during recombination. Based on reports,[139] we employed a DNA polymerase (Phusion™ DNA polymerase) with high fidelity in the gene preparation, reassembly and post-amplification step. In addition, the purification step after DNase I digestion was omitted, due to the small size of the fragments and resulting difficulties during purification with commercially available kits. An amplification based on undigested parental DNA, resulting from omitting the purification step after digestion as suggested in the literature, was not observed. Rather, a wide range of mutation combinations was found after sequencing various clones of the existing library. A recombination rate was

not determined, but a total number of one to four mutations were detected when sequencing various clones from the library. *KlenTaq* expression, lysis and heat-denaturation of the host proteins as well as the screening system were already established in our group and were adopted.[132, 148] DNA amplification was followed in real time. As a large number of variants showed a reverse transcriptase activity superior to the parental enzymes, we subsequently increased the selection pressure in the screening system, such as in reducing the reverse transcription time or the RNA concentration. Sequencing of the nine most promising variants revealed four different mutation combinations. Although a preceding sequencing of the original library revealed mainly clones with a lower total number of mutations (maximum four), the identified hits comprised at least four mutations, with the exception of one variant. On the one hand this shows high library coverage of all mutation combinations and on the other hand it suggests a general importance of combining a higher number of mutations for efficient reverse transcriptase activity. The mutations L459M, S515R, I638F and M747K were present in all clones, except RT-KTq 1, indicating a general significance of these mutations. Additionally, this variant was found in four out of nine clones sequenced. One could argue that this could be due to a bias in the recombination step, but additional mutations as shown in RT-KTq 3 and 4 result in no gain in reverse transcriptase activity, whereas the removal of L459M as depicted in RT-KTq 1 seems to lead to an overall loss in activity; thus supporting the notion that the before mentioned mutation combination is the minimal number of mutations required for the superior reverse transcriptase activity inherent in all variants identified. Interestingly, the S739G mutation was not present in any of the identified clones. A comparison with KTq M1/M747K, the mutant generated by site-directed mutagenesis and the only enzyme exhibiting all possible mutations, including S739G, showed no change in reverse transcriptase activity. But the absence of this mutation in the identified hits might be due to a loss in thermostability as seen in thermal denaturation experiments conducted with the mutants KTq M1/M747K and RT-KTq 2.

Furthermore, the combination of mutations has synergistic effects in the resulting mutant enzymes compared to the parental enzymes M1 and M747K. The mutations seem to interact with each other not in an additive but rather in a non-additive manner.[186] Recently, the phenomenon of non-additive interactions between mutations has gained increasing awareness in evolutionary biology.[191, 192] So far, adaptation which underlies C. Darwin's theory of natural selection was considered as the major determinant for evolution. An organism adapts to the environment and ecology and thus facilitates evolution. The process of adaptation was always thought to depend on the context-independent (additive) genetic effects that exist in a population. Epistasis as context-dependent (non-additive) genetic effects was only recently suggested as a key player in evolution besides adaptation.[192] Indeed, a statistical analysis of proteins performed by Breen et al. suggests that functional interactions between amino acids are a primary factor of protein-sequence evolution as well.[191]

All identified and purified variants were applied in RT-PCR exhibiting a rise in fluorescence approximately 10 cycles earlier than the parental enzymes, corroborating the results obtained from the screening. The enzymes reveal a superior reverse transcriptase activity compared to the parental enzymes. Thus, they offer the opportunity for application in a rapid RT-PCR method omitting even the need for the reverse transcription step. Even more, due to the possibility of hot-start RT-PCR, higher specificity of RNA detection is possible. Focusing on RT-KTq 2 and its application in clinical diagnostics, the enzyme was successfully applied in rapid RT-PCR for the simultaneous detection of influenza A and B in multiplex RT-PCR.

Most DNA polymerases evolved towards reverse transcriptase and PCR activity in one enzyme scaffold belong to the family A DNA polymerases.[147-150, 155, 157] Only mutants of Tgo DNA polymerase, belonging to the family B archaeal DNA polymerases, were reported to possess the ability to accept both DNA and RNA as substrate.[193] Although tremendous efforts have been invested in the evolution of DNA polymerases with increased reverse transcriptase activity, few of them were shown to be applicable in RT-PCR and few are commercially available. The existing one-step RT-PCR kits generally contain an enzyme blend comprising a reverse transcriptase and a DNA polymerase. Thus, the properties of both enzymes have to be considered when optimizing analysis of new RNA targets for clinical diagnostics. But few enzymes are presently on the market, which are capable of both reverse transcription and subsequent amplification. Recently, Moser et al. published a thermostable DNA polymerase isolated from a viral metagenome library exhibiting innate reverse transcriptase activity; commercially available under the brand name PyroPhage.[151] Alternatively, the use of *Tth* polymerase, isolated from the thermophilic eubacterium *Thermus thermophilus*,[147] in one-step RT-PCR can be considered, although several drawbacks, such as lower sensitivity compared to two-enzyme mixtures, arise from the fact that Mn^{2+} is required as divalent cation.[142]

The identified enzymes represent a promising alternative to two-enzyme mixtures by minimizing work and time costs, avoiding time consuming optimization and inhibitory effects present between a reverse transcriptase and a PCR-competent DNA polymerase.[152-154] Particularly, in certain important fields of clinical diagnostics such as point of care testing minimized time consumption and reliable detection is highly important. In addition, performing the reverse transcription step at high temperatures prevents unspecific priming and biased RNA detection based on secondary structure formation of specific RNA molecules.[142] In summary, the variants exhibit very promising properties, which allow applications such as fast detection and quantification of RNA in RT-PCR, thus facilitating their use in transcriptome analysis, pathogen as well as disease-specific marker detection.

2. Crystallization Studies with RT-KTq 2

2.1 Introduction

The next part of this work addressed two related aspects. First, the contributions of the RT-KTq 2 variant's individual mutations to the observed superior reverse transcriptase activity were investigated. Additionally, we were also interested in the fundamental question how DNA-dependent DNA polymerases can discriminate between the templating nucleic acids (DNA vs RNA). Especially *in vivo,* the discrimination between NTPs and dNTPs and between RNA and DNA is essential, as the amount of ribo- exceeds that of desoxyribosubstrates by far. Thus, the mechanism of discrimination is a subject of ongoing investigations. But to date, the vast majority of functional and structural data existing for DNA and RNA polymerases mainly focus on the discrimination mechanisms between the incorporation of either NTPs or dNTPs;[155, 194-199] much less is known on how DNA-dependent DNA polymerases recognize or discriminate between the templating nucleic acids DNA and RNA.[140] To the best of my knowledge only crystal structures of reverse transcriptases exist which give insights into the acceptance of both, RNA and DNA, as template. HIV-1 RT is probably the best characterized reverse transcriptase, with structural data available for the enzyme in complex with an all DNA and a DNA/RNA hybrid duplex.[24-26, 200] Another characterized reverse transcriptase is the enzyme from the gammaretrovirus Moloney murine leukaemia virus (Mo-MLV). But so far, structural data of this enzyme only yielded an intermediate translocation state with the DNA duplex lacking the contact to critical active site residues.[201] Recently, the first crystal structure of the reverse transcriptase from xenotropic murine leukaemia virus-related virus, a close relative of Mo-MLV, in complex with a DNA/RNA hybrid duplex was reported.[202] These structures of reverse transcriptases processing RNA or DNA as template yielded insights in the contacts formed between the protein and the substrate. It further provided information about the geometry of the hybrid duplex and demonstrated how the enzyme achieves to accommodate the geometric changes from an all DNA to a hybrid duplex in its active site. Structural data, however, of a 'pure' DNA-dependent DNA polymerase, whose wild-type parental ancestors show no significant reverse transcriptase activity, is lacking. Thus, we set out to crystallise *KlenTaq* DNA polymerase in complex with RNA as template.

Part of this work is published in Angewandte Chemie[179] and was conducted with the support of B. Sc. Simon Geigges and Irene Griesser as part of the course 'Gene Expression' under my supervision. Dipl.-Biol. Konrad Bergen continuously supported this study with his expert knowledge and performed the data recording and reduction, structure solution and subsequent refinement.

2.2 Results

2.2.1 Crystallization Trials of KTq Wild-type in Complex with a DNA/RNA Hybrid

Crystal structure analysis of *KlenTaq* DNA polymerase provided significant insights into how a DNA polymerase recognizes the cognate substrate[30, 68, 203] or accepts various modified substrates.[61, 69-71] Thus, *KlenTaq* DNA polymerase represented an ideal model system to study the mechanism the enzyme employs to discriminate between the templating nucleic acids. Crystallization trials were set-up containing *KlenTaq* DNA polymerase wild-type and an 11mer DNA primer and a 16mer RNA template. However, in spite of intensive studies to crystallize *KlenTaq* DNA polymerase wild-type in complex with a DNA primer/RNA template duplex, no crystals were obtained. Different crystallization strategies, implemented by Irene Griesser, as well as various conditions were utilized, but eventually failed. An initial strategy employed dideoxynucleotides to terminate elongation of the primer and catch the enzyme in complex with a second ddNTP in the active site (**Figure 34A**).

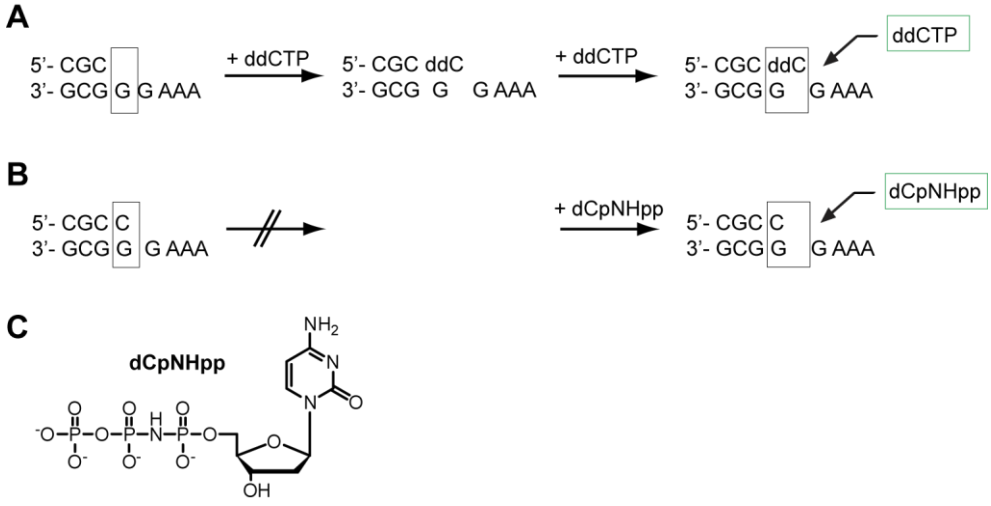


Figure 34. Crystallization strategies. A) An initial strategy employed a ddCTP to terminate primer elongation and catch a second ddCTP in the active site. B) A second approach employed a +1 nucleotide elongated primer and a non-hydrolysable nucleotide (dCpNHpp) to forgo nucleotide incorporation. Partial primer/template sequences are shown. Grey boxes mark the different lengths of the primer. Green boxes mark the nucleotide analogue caught in the active site. C) Chemical structure of dCpNHpp.

This strategy was already successfully employed to crystallize *KlenTaq* DNA polymerase in complex with various substrates.[30, 61, 72] However, crystallization set-ups employing different

commercially available screens as well as optimized conditions for *KlenTaq* yielded no crystals using RNA as template.

KlenTaq wild-type was already shown to exhibit a low incorporation rate opposite RNA, thus a second approach tried to forgo nucleotide incorporation opposite RNA which was a requirement in the first strategy to obtain crystals. A primer with an additional nucleotide at the 3' end together with a non-hydrolysable nucleotide, 2'-deoxy-cytidine-5'-[(α,β)-imido]triphosphate (dCpNHpp), was employed in this set-up (**Figure 34B**). But again, crystallization trials yielded no protein crystals.

Presumably, the crystallization process is hampered by the poor propensity of the enzyme to process or bind RNA which leads to the formation of unstable complexes resulting in structural heterogeneity. If this is the case, the enhanced reverse transcriptase activity of the RT-KTq 2 variant should overcome this limitation. With this idea in mind, crystallization trials with the RT-KTq 2 variant in complex with two different substrates, an all DNA duplex and a DNA/RNA hybrid duplex, were conducted. The crystallisation studies should provide insights into the restraints present in the wild-type enzyme which prevents its efficient processing of RNA as template and which are reversed in the mutant enzyme.

2.2.2 RT-KTq 2 in Complex with a DNA Duplex

B. Sc. Simon Geigges introduced the mutations of RT-KTq 2 in a codon-optimized *KlenTaq* gene by site-directed mutagenesis, purified the enzyme based on established protocols in the literature[204] (see chapter **VI 2.4.1.2**) and contributed to the crystallization studies.

Crystallization trials were set up with an 11mer DNA primer and a 16mer RNA template and for structural comparison also with a primer/template DNA duplex with the respective sequence. The structure of RT-KTq 2 bound to an all DNA duplex and an incoming 2',3'-dideoxycytidine-5'-triphosphate (ddCTP) was solved at high resolution of 1.55 Å (termed RT-KTq 2_{DNA}, PDB ID 4BWJ). Again, a 2'-3'-dideoxynucleotide (ddCTP) was employed to terminate elongation of the primer and trap the enzyme in complex with a second ddCTP in the active site. Data collection and refinement statistics are listed in **Table 4**.

Data was processed and reduced using XDS.[205] The structure was solved using difference Fourier methods and refined using the PHENIX suite.[206] Manual refinement and model rebuilding was performed using Coot.[207] The enzyme in complex with a DNA/DNA duplex crystallized in the same space group 152 ($P3_121$) already described for ternary structures of the wild-type enzyme.[30, 61] Ramachandran statistics for RT-KTq 2_{DNA} show that 97.61 % of the residues are in favoured regions, 2.20 % in allowed regions and 0.18 % are outlier.

Table 4. Data collection and refinement statistics.

	RT-KTq 2_{DNA} PDB ID 4BWJ	RT-KTq 2_{RNA} PDB ID 4BWM
Data collection		
Space group	152 ($P3_121$)	18 ($P2_12_12$)
Cell dimensions		
a, *b*, *c* (Å)	a=b=108,.43 c=89.81	143.19 86.28 63.09
α, β, γ (°)	90, 90, 120	90, 90, 90
Resolution (Å)	46.83 (1.64) -1.55	47.985 (1.85) -1.749
R_{meas}	6.3 (112.5)	8.3 (111.5)
$I/\sigma I$	15.56 (1.64)	13.73 (1.70)
Completeness (%)	99.7 (98.0)	99.9 (99.4)
Redundancy	9.6 (8.6)	6.59 (6.62)
Refinement		
Resolution (Å)	47.99-1.55	46.83-1.75
No. reflections	87950	79632
R_{work}/ R_{free}	15.9/20.1	16.8/20.8
No. atoms (H/noH)		
Protein	8765/4345	-/4345
Ligand/ion	906/569	-/548
Water	476/476	-/624
B-factors		
Protein	43.80	33.97
Nucleic acids	36.52	71.52
Water	42.29	43.59
R.m.s deviations		
Bond lengths (Å)	0.011	0.007
Bond angles (°)	1.504	1.089

*Highest resolution shell is shown in parenthesis.

The overall structure of RT-KTq 2_{DNA} is nearly identical to closed ternary structures already described for *KlenTaq* wild-type with a DNA-duplex (e.g. PDB ID 3RTV[61]) and thus shows a low root mean square deviation (rmsd) of 0.33 Å for Cα (**Figure 35**). The active site arrangement is identical in both structures with a ddCTP and two Mg^{2+}-ions in the active site poised for catalysis. The O helix (**Figure 35A**) adopts a closed state already described for ternary complexes of *KlenTaq* wild-type.[30]

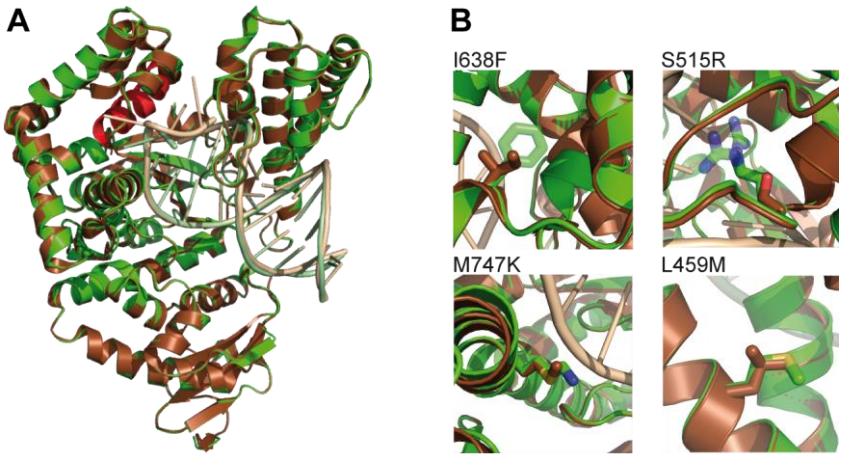


Figure 35. Structural comparison of KTq wild-type and RT-KTq 2_{DNA}. A) Structure of KTq wild-type (sand, PDB ID 3RTV) superposed with RT-KTq 2_{DNA} (green) bound to DNA-duplex. The O helix in both structures is depicted in red. B) Mutation sites are highlighted. Graphic was adapted from Blatter et al.[179]

2.2.3 RT-KTq 2 in Complex with a DNA/RNA Hybrid

The crystal structure of the enzyme variant in complex with a DNA/RNA hybrid and an incoming ddCTP was also solved at a resolution of 1.75 Å (termed RT-KTq 2_{RNA}, PDB ID 4BWM). Data processing and refinement was performed as described for RT-KTq 2_{DNA}, but in case of RT-KTq 2_{RNA} molecular replacement (PHASER) was used to solve the phase problem.[208] Part of the protein had to be retraced, as well as the hybrid duplex, neglecting the two outermost base pairs, which were not resolved. Data collection and refinement statistics are listed in **Table 4**. Ramachandran statistics for RT-KTq 2_{RNA} show that 98.53 % of the residues are in favoured regions, 1.47 % in allowed regions and 0 % are outlier.

An overlay of RT-KTq 2_{DNA} and RT-KTq 2_{RNA} revealed changes in the fingers and thumb domain of the RT-KTq 2_{RNA} structure, presumably due to a shift of the hybrid duplex and its altered geometry (**Figure 36A**). RT-KTq 2_{RNA} also crystallized in a different space group 18 ($P2_12_12$) than RT-KTq 2_{DNA} and is in an ajar state compared to RT-KTq 2_{DNA} with a partially-closed conformation of the O helix located in the fingers domain (**Figure 36**). Additionally, only one Mg^{2+}-ion was observed in the enzyme's active site. A partially-closed conformation of the O helix is characteristic for *KlenTaq* in complex with an abasic site analogue containing template (PDB ID 3LWL)[72] (**Figure 36B**) and was also described for the large fragment of DNA polymerase I from *Bacillus stearothermophilus* bound to DNA and a mismatched nucleotide in the active site cleft.[60]

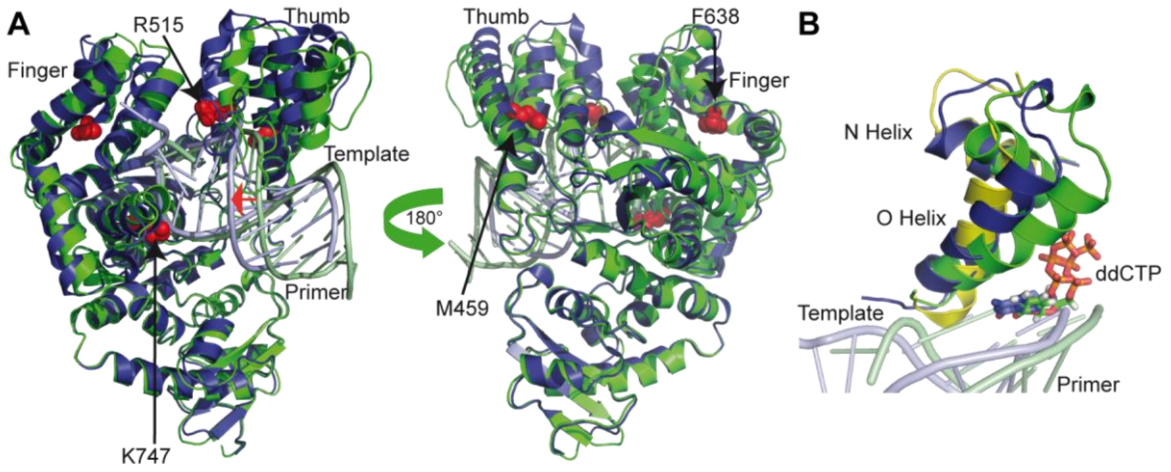


Figure 36. Superposition of crystal structures RT-KTq 2_{DNA} and RT-KTq 2_{RNA}. A) Protein scaffold of RT-KTq 2_{DNA} (green) superposed with RT-KTq 2_{RNA} (blue). DNA-duplex and DNA/RNA hybrid are coloured in pale green and pale blue, respectively. Mutation sites are highlighted as red spheres. Shift of the hybrid duplex is indicated with a red arrow. B) Close-up view of the helices O and N in RT-KTq 2_{DNA} and RT-KTq 2_{RNA}. Superposition reveals the partially-closed conformation of the O helix in RT-KTq 2_{RNA}. The O helix conformation of *KlenTaq* in complex with an abasic site analogue containing template is depicted in yellow (PDB ID 3LWL). Graphic was adapted from Blatter et al.[179]

The partially-closed conformation of the O helix, however, might result from new crystal contacts formed in RT-KTq 2_{RNA} due to the different space group the enzyme crystallized in. Thus, we cannot make definite predictions about the origin of the structural changes observed in this part of the protein. Therefore, we focused on the changes in the thumb domain, presumably triggered by the altered geometry of the hybrid duplex. The RNA template adopts a sugar puckering typical of A form conformation in the hybrid duplex, which was expected. Interestingly, the DNA primer in the hybrid duplex reveals a variety of puckers which deviate from the typical B form (**Table S1**, see chapter **VII 1.3**).[209] The sugar puckers in the primer nucleotides free from enzyme interactions show C3'-endo conformations, whereas the other sugar moieties adopt various conformations such as C2'-endo, C3'-endo, C1'-exo or O4'-endo (**Table S1**).

In case of *KlenTaq* wild-type in complex with an all DNA duplex, the three base pairs at the end of the duplex adjacent to the O helix already adopt A form.[30] Thus, only structural changes of the protein scaffold in regions neighbouring the nucleic acids downstream of these base pairs are necessary to accommodate the changed geometry of the hybrid duplex. An overlay of RT-KTq 2_{DNA} and RT-KTq 2_{RNA} additionally revealed that part of the thumb domain would have clashed with the hybrid duplex, therefore further structural changes of this part of the enzyme are required to accommodate the hybrid duplex (**Figure 37**).

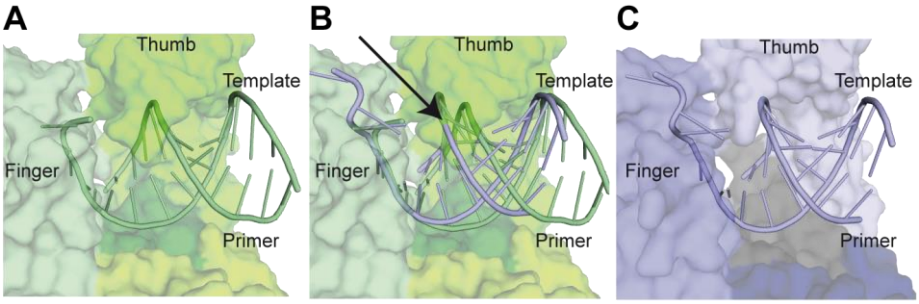


Figure 37. Accommodation of the nucleic acid substrate in the enzyme's binding cleft. Close-up view shows the DNA/DNA duplex in RT-KTq 2_{DNA} (A), overlay of DNA and hybrid duplex in RT-KTq 2_{DNA} (B) and DNA/RNA hybrid duplex in RT-KTq 2_{RNA} (C), respectively. The overlay of the DNA/RNA and the all DNA duplex in RT-KTq 2_{DNA} highlights the need for a rearrangement of the thumb domain due to the altered geometry of the DNA/RNA hybrid that would clash with the enzyme (highlighted with the arrow). Graphic was adapted from Blatter et al.[179]

Consequently, several changes in the interaction pattern are observed and contacts between the enzyme and the hybrid duplex are lost in RT-KTq 2_{RNA}, mainly in the 3' half of the template (**Figure 38**).

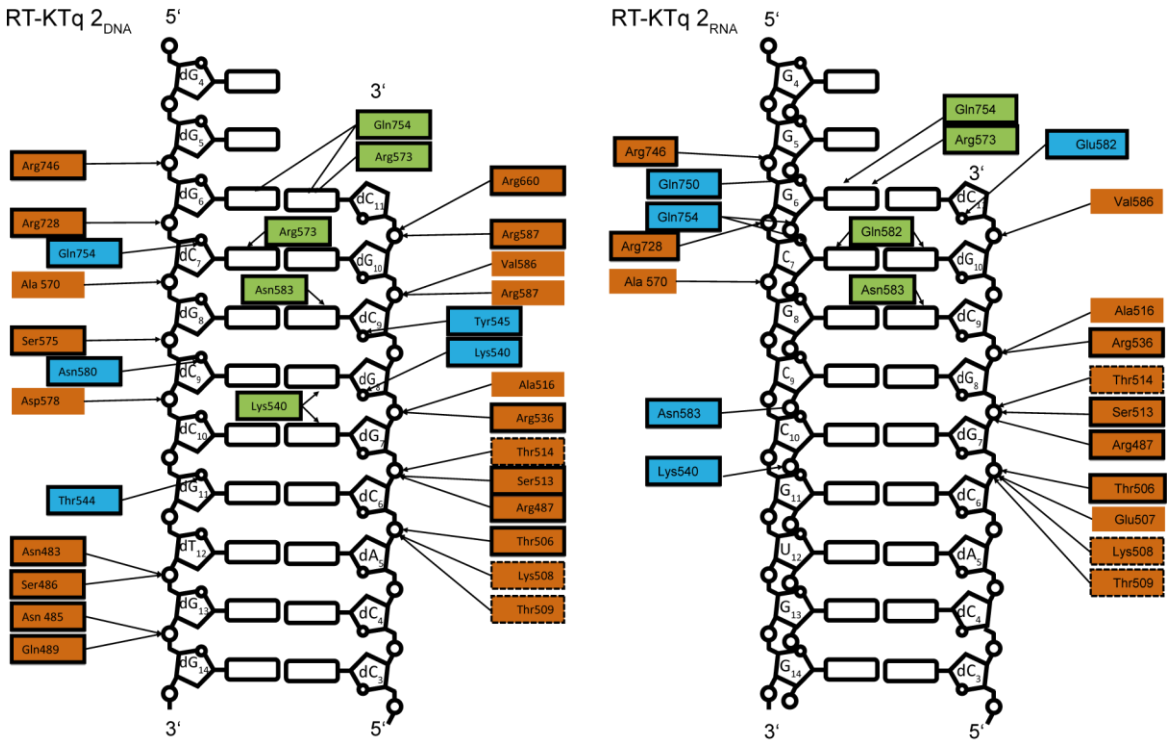


Figure 38. Interaction map of RT-KTq 2_{DNA} (left) and RT-KTq 2_{RNA} (right). Only direct contacts up to a distance of 3.6 Å are shown. Side chain interactions are marked with a solid lining, contacts with the protein backbone are shown without lining and residues where both interactions are found are shown with dashed lining. Interactions to the phosphate backbone, sugar oxygen or nucleobase are shown in orange, blue and green, respectively. Graphic was adapted from Blatter et al.[179]

Remarkably, the major part of the reordered protein scaffold originates in the thumb domain, downstream of L459M in the H helix (**Figure 39A**), which is consistent with previous findings that mutations in this helix influence substrate acceptance.[210] We suggest that the structural changes in this subdomain result from the mutation of L459 to the sterically less demanding and more flexible side chain of methionine. The mutated amino acid facilitates a bending of helix H which in turn causes a movement of helices H1 and H2 as well as loop H1H2 and brings them in close proximity to the hybrid duplex. The loop H1H2 in the thumb domain was already described as part of a nucleic acid binding motif of *KlenTaq* wild-type in complex with a DNA duplex.[30] Thus, the nucleic acid binding characteristics of this motif[30] are maintained in case of the hybrid duplex due to structural changes originating from the L459M mutation (**Figure 39A**). To further corroborate this finding, the nucleic acid binding motif was superposed alone, independently from the rest of the structure, and indeed it was found that the deviation between the wild-type and the RT-KTq 2_{RNA} structure in the nucleic acid binding motif starts in the N-terminal portion of the H helix containing the L459M mutation (**Figure 39B**), eventually leading to a shifted location of the motif.

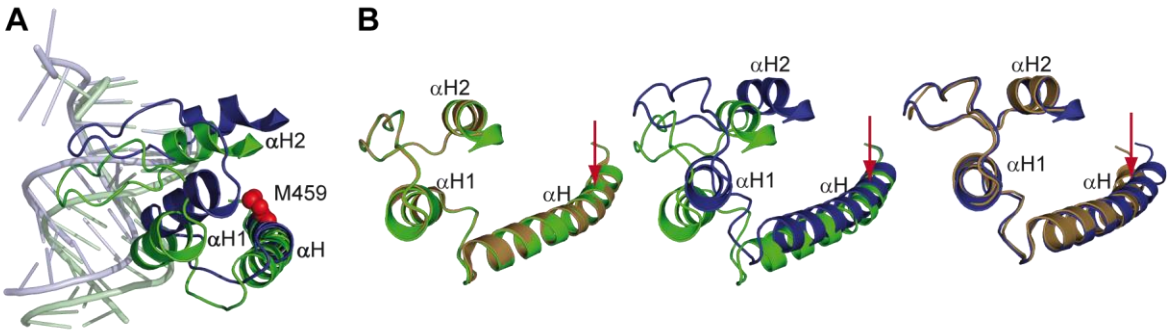


Figure 39. Influence of L459M mutation on accommodating the hybrid duplex. RT-KTq $2_{DNA,}$ RT-KTq 2_{RNA} and wild-type are depicted in green, blue and sand, respectively. A) Close-up view highlights the position of the M459 mutation within H helices in RT-KTq 2_{DNA} and RT-KTq 2_{RNA}. M459 is depicted as red spheres. B) Structural relocation of helices H, H1 and H2 in the thumb domain. Overlay of protein structure of KTq wild-type (PDB ID 3RTV) and RT-KTq 2_{DNA} shows no relocation of the motif (left). Superposition of protein structure of RT-KTq 2_{DNA} and RT-KTq 2_{RNA} reveals shift in helices H, H1 and H2 (middle). Superposition of the motif alone independent from the overall structure reveals the origin of the shift (right). Deviation starts in the N-terminal portion of helix H containing the L459M mutation. The red arrow highlights the position of L459M within H helices. Graphic was adapted from Blatter et al.[179]

The second mutation S515R present in the variant is located in helix H2, with the arginine residue displaying two different conformations in RT-KTq 2_{RNA}. Both conformations allow

interactions with the hydroxyl group of S513 and the peptide backbone of K505, thus stabilizing the previously mentioned nucleic acid binding motif in the thumb domain (**Figure 40A**). The role of I638F is less obvious, though it could allow a tighter packing of the fingers domain due to hydrophobic and stacking interaction with neighbouring amino acids, thereby resulting in a compaction of this region. In contrast to the RT-KTq 2_{DNA} structure, R515 and F638 are resolved in simulated annealing omit maps, indicating a stabilization of the corresponding region in RT-KTq 2_{RNA} and further supporting our observations of a stabilizing effect resulting from these mutations (**Figure 40B**).

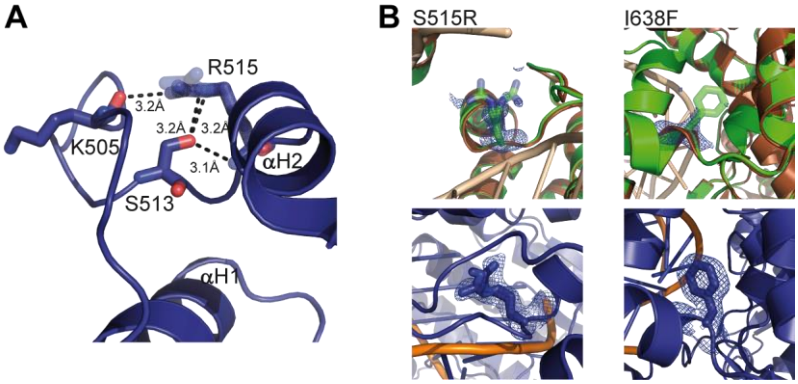


Figure 40. Stabilizing effects of S515R and I638F. A) Close-up view highlights the interaction network of mutation S515R. The R515 mutation is depicted as blue sticks. An alternative conformation of its side chain found in the crystal structure is transparent. B) Simulated annealing omit maps (σ level of 3) showing mutations S515R and I638F in RT-KTq 2_{DNA} (green) superposed with KTq wild-type (PDB 3RTV, sand, top) and in RT-KTq 2_{RNA} (blue, bottom). Graphic was adapted from Blatter et al.[179]

The substitution of the hydrophobic amino acid methionine with a positively charged lysine at position 747 resulted in an increased positively charged surface potential in close proximity to the negatively charged backbone of the DNA (**Figure 41**), as reported previously in the literature.[73] This might promote binding of the nucleic acid backbone and thus, foster the ability to accept aberrant substrates, such as RNA or abasic site containing templates[73], by electrostatic interactions.

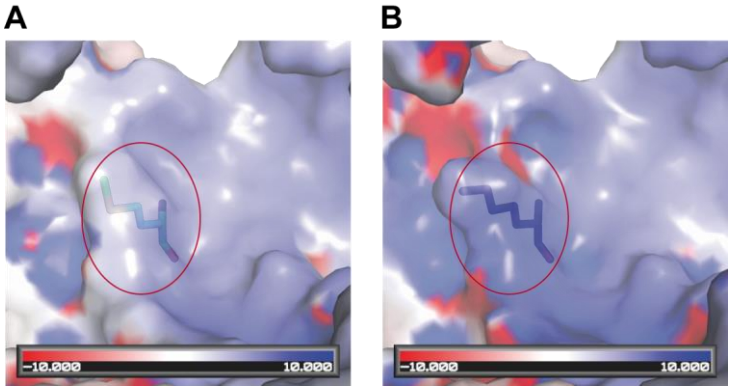


Figure 41. Electrostatic surface potentials at the mutation site M747K. A) KTq wild-type (PDB ID 3RTV) with M747. B) RT-KTq 2_{RNA} with K747. The respective residue is highlighted. Blue and red represent the positive and the negative charge potentials, respectively (kTe^{-1}).

2.3 Discussion

In summary, a DNA-dependent DNA polymerase (RT-KTq 2) in complex with an all DNA and a DNA/RNA hybrid duplex were successfully crystallized and the structures solved to a resolution of 1.55 and 1.75 Å, respectively. Furthermore, the molecular role of each mutation present in RT-KTq 2 was elucidated with regards to its contribution to accommodate the distinct geometry of the hybrid duplex and to process RNA efficiently as template.

The RT-KTq 2_{DNA} structure shows a ternary complex of the protein almost identical to already published structures of the wild-type enzyme.[30, 61] The protein in the RT-KTq 2_{RNA} structure, however, adopts an ajar state with a partially-closed conformation of the O helix in the fingers domain and only one Mg^{2+}-ion present in the active site. The partially-closed conformation might be due to the different space group the protein crystallized in with distinct residues forming new crystallization contacts, thus presumably preventing a closure of the O helix.

A comparison of the RT-KTq 2_{RNA} and RT-KTq 2_{DNA} structure revealed major changes both in the fingers and thumb domain. We focused on the changes in the thumb domain, as we could not exclude that the structural alterations in the fingers domain result from the newly formed crystal contacts.

The major part of the reordered protein scaffold in the thumb domain originates downstream of L459M in the H helix, which is consistent with previous findings that mutations in helix H influence substrate acceptance.[210] In detail, Kranaster *et al.* identified several *KlenTaq* mutant enzymes comprising various mutations located in helix H, which result in a different processing pattern of various substrates such as an abasic site template or a primer/template mismatch compared to the wild-type enzyme. We suggest that in this case the mutation

L459M contributes to preventing the clash of the thumb domain with the hybrid duplex and enables helices H1 and H2 to come in close contact to the hybrid duplex, thereby maintaining the characteristics of the nucleotide binding motif encompassed by these helices. Furthermore, the S515R mutation stabilizes this motif by interacting with surrounding amino acid residues. The M747K mutation is proposed to increase the positively charged surface potential in proximity to the negatively charged substrate backbone and thus enhances the acceptance of non-cognate substrates by electrostatic interactions.[73] As previously described, the mutation of the methionine at position 747 to a positively charged lysine resulted in an increased acceptance of an abasic site containing template and might explain the enhanced lesion bypass activity of this variant.[73, 159] Furthermore, the efficient processing of a non-cognate RNA primer/DNA template duplex by incorporation of ribonucleotides was reported when a positively charged residue was introduced in the thumb domain of Tgo DNA polymerase,[211] thus further supporting the notion that a positively charged residue in close proximity to negatively charged nucleic acids might foster the acceptance of aberrant substrates. For I638F a role is difficult to predict. However, the resolution of F638 in RT-KTq 2_{RNA} indicates a stabilisation of this region, presumably due to possible hydrophobic and stacking interaction of F638 with neighbouring amino acids (F632, H639, F667, F700).

To further investigate the changes of the protein which are necessary to adapt to the changed geometry of the hybrid duplex, the observed interactions in RT-KTq 2_{RNA} were compared with the ones present in reverse transcriptases such as HIV-1 RT. In both enzymes, the terminal base pairs of a DNA/DNA duplex in the active site always adopt an A form conformation. It was suggested for reverse transcriptases, that an A form conformation at the active site allows the polymerase to use either DNA or RNA as template, as both templates can adopt an A form, but RNA cannot adopt B form conformation.[23] The acceptance of both templates is necessary in the retroviral life-cycle. **Figure 42** depicts the three terminal base pairs in the active site of HIV-1 RT showing the A form characteristic C3'-endo conformation of the sugar puckers.

However, HIV-1 RT exhibits a higher activity processing RNA than DNA as template, which was found to be a result from more extensive contacts formed between the enzyme and nucleotides of the RNA template downstream of the terminal base pairs in the active site.[26] In analogy, we found less contacts between the 3' half of the RNA template and the residues of *KlenTaq* DNA polymerase, further elucidating the loss in activity for *KlenTaq* DNA polymerase when using RNA instead of DNA as template. Although the sugar puckers of the primer strand adopt various conformations in the RT-KTq 2_{RNA} structure, several interactions are maintained due to the presence of the L459M mutation. The methionine residue facilitates a shift of helices H1 and H2, thus maintaining interactions between the protein and the nucleic acid backbone; albeit with a one base pair shift in the 3' direction of the primer. Interestingly, most interactions are between the sugar-phosphate backbone and enzyme residues, unspecific interactions which might alleviate the acceptance of the two different templates.

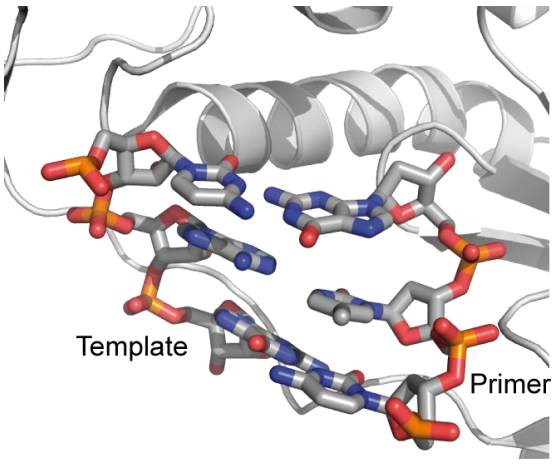


Figure 42. Active site of HIV-1 RT (PDB ID 1 HYS).[26] The three terminal base pairs in the active site are shown. Sugar puckers of the primer and template exhibit a C3'-endo conformation.

Insights into the role of the mutations present in the enzyme variant which allowed an efficient use of RNA as template also indirectly provided information about the restraints present in the wild-type enzyme, which prevent efficient RNA processing. Thus, one can assume that the distinct geometry of the hybrid duplex and a subsequent clash with the thumb domain are one major determinant for the wild-type enzyme's discrimination between RNA and DNA.

To further investigate the changes in the fingers domain, additional mutations could be introduced in the RT-KTq 2 variant in order to facilitate the closure of the O helix. Based on the crystal structure of RT-KTq 2 in complex with the hybrid duplex, K. Bergen identified one amino acid residue W417, which might sterically prevent the enzyme in the new space group to adopt a closed conformation. Thus, mutating this residue to alanine might result in an active ternary state of the enzyme and might provide further insights into structural alignments of the primer/template duplex as well as the incoming dNTP in the enzyme's active site.

3. Generation of Full-Length *Taq* DNA Polymerase Variants

3.1 Introduction

So far, the described experiments were conducted with *KlenTaq* variants, but in many applications, such as *Taq*Man based real-time RT-PCR, the full-length *Taq* enzyme is preferred. The detection and quantification of RNA in real time is a crucial tool in molecular biology and molecular diagnostics. Monitoring the abundance of RNA molecules can be achieved either with fluorescent dyes or with probe based chemistries. Fluorescent dyes such as SYBRGreen I or Eva Green bind to double-stranded DNA emitting a higher fluorescence signal than in the unbound state.[142] Thus, a signal increase is observed as polymerization proceeds with a correlation between the signal strength and the amount of double-stranded DNA formed. The use of probe based chemistry is often preferred, as a higher level of specificity is achieved. Molecular beacons, hybridisation or hydrolysis probes are currently employed in real-time RT-PCR,[14, 142] whereas the use of hydrolysis probes in the so-called *Taq*Man assay is particular known. It utilizes the 5'-3' nuclease activity of a DNA polymerase to hydrolyse a dual-labelled *Taq*Man probe which is hybridized to a target sequence. The *Taq*Man probe consists of an oligonucleotide complementary to the target amplicon, which adds further specificity to the reaction, conjugated with a 5' fluorescent donor and a 3' acceptor (quencher). Hydrolysis of the probe sets the fluorophore and the quencher free. Consequently, a fluorescence signal can be detected. Thus, *Taq*Man based real-time RT-PCR combines high specificity and flexibility, as it is easy to adjust to new targets.
However, *KlenTaq* DNA polymerase is devoid of a 5'-3' endonuclease activity, which is only present in the full-length enzyme. Detection of influenza A and B virus in multiplex *Taq*Man based RT-PCR, described in chapter **III 1.2.6**, was solely achieved by employing a mixture of RT-KTq 2 and *Taq* DNA polymerase wild-type. Although, the mixture yielded promising results, the use of the full-length enzymes was further investigated. *Taq* wt, M1, M747K, M1/M747K and RT-Taq 2 were generated (**Figure 43A**) and tested regarding their reverse transcriptase activity.

3.2 Results

3.2.1 Generation and Purification

Genes for *KlenTaq* wild-type, the parental enzymes M1 and M747K, KTq M1/M747K and the identified variant RT-KTq 2 encoded on the expression vector pGDR11 were amplified in PCR, in addition to the endonuclease domain of *Taq* wild-type encoded on the pASK-IBA 37+ vector. The endonuclease domain and the respective *KlenTaq* domain were then ligated in a blunt-ended manner and cloned back into the expression vector pGDR11 (**Figure 43A**).[184]

PGDR11 vector was preferred to the pASK-IBA 37+ vector due to higher expression levels. After expression for 4.5 h, the cells were lysed at 37 °C and host proteins removed via heat-denaturation at 75 °C for 45 min. A centrifugation step removed the bacterial cell debris. Next, the proteins were further purified via His-tag affinity chromatography based on the procedure used for *KlenTaq* DNA polymerase. Due to many termination products (**Figure 43B**), gel filtration using a Sephadex-75 column was applied for further purification (**Figure 43C**).

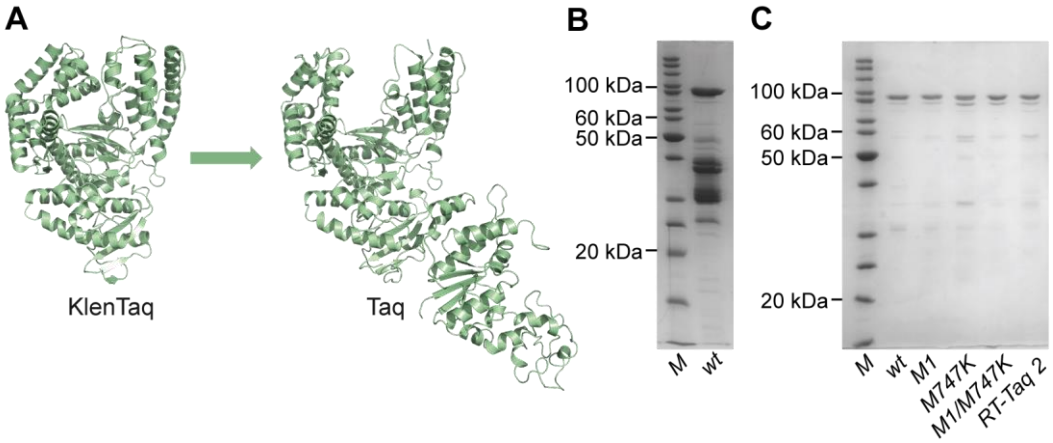


Figure 43. Generation of full-length enzyme *Taq*. A) Structure of *KlenTaq* (PDB ID 3RTV) and *Taq* (PDB ID 1TAU[212]) DNA polymerases. B) Purification grade of *Taq* wild-type (wt) after His-tag affinity chromatography analysed on SDS-PAGE gel. C) SDS-PAGE of purified *Taq* wt, *Taq* M1, *Taq* M747K, *Taq* M1/M747K and RT-*Taq* 2 as indicated.

The purity after size exclusion chromatography was sufficient for performing preliminary biochemical studies. The proteins were adjusted to equal target-protein concentrations (**Figure 43C**) and tested regarding their intrinsic endonuclease activity and their evolved reverse transcriptase activity.

3.2.2 Endonuclease Activity of *Taq* Wild-type and Variants

One important feature for the application in *Taq*Man based real-time RT-PCR is the intrinsic 5'-3' endonuclease activity of the respective enzyme. To investigate a possible influence of the mutations on the enzyme`s nuclease activity, the wild-type and all variants were examined in an already established nuclease activity assay.[150] In short, a 5'-[^{32}P]-radioactively labelled oligonucleotide was annealed to a stable DNA hairpin resulting in a structure exhibiting a displaced 5' end and a frayed 3' primer terminus (**Figure 44A**). Such a construct was shown to be a preferred cleavage substrate for *E. coli* DNA polymerase I and *Taq* DNA polymerase.[213]

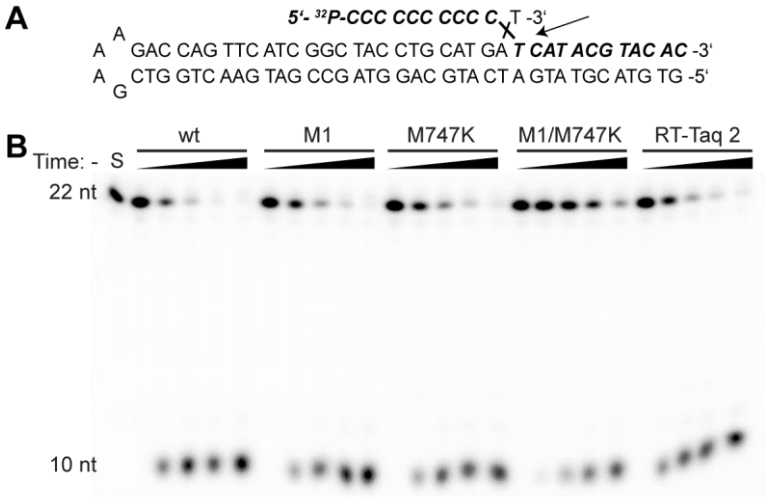


Figure 44. Nuclease activity. A) Hairpin structure of template and 22-nt substrate (bold). The arrow indicates the expected cleavage position. B) Nuclease activities of *Taq* DNA polymerases wild-type (wt), M1, M747K, M1/M747K and RT-*Taq* 2 as indicated. Substrate cleavage was performed at 30 °C and determined at different time points (0, 5, 15, 30, 60 min). Reaction products were separated by denaturing PAGE. S: 22-nt substrate.

Reactions were incubated for different lengths of time up to 60 min and subsequently analysed on a denaturing PAGE-gel (**Figure 44B**). RT-*Taq* 2 exhibits an endonuclease activity comparable to the parental enzymes M1 and M747K with almost quantitative cleavage of the displaced 5' end after 30 min. Only the *Taq* M1/M747K mutant showed a slightly reduced activity of the endonuclease domain, with uncut substrate still visible after 60 min incubation time.

3.2.3 Insights into the Reverse Transcriptase Activity

Next, the reverse transcriptase activities of the full-length enzymes were investigated in primer extension experiments. The set-up employed for the N-terminally truncated variants was also applied here. In short, a 5'-[^{32}P]-radioactively labelled DNA primer was annealed to an RNA template and elongated via the respective enzyme. Reactions were incubated at 72 °C for 30 sec, 1 min and 5 min and subsequently analysed on a denaturing PAGE-gel (**Figure 45**).

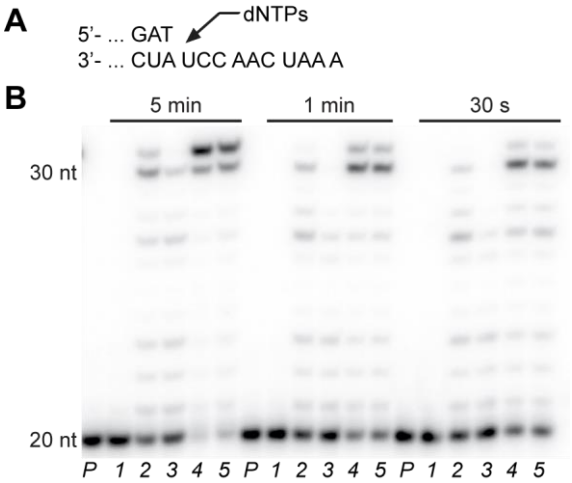


***Figure 45*.** Primer extension experiments with *Taq* wild-type (wt) and variants. A) Partial primer/template sequence. B) Primer extension reactions with 5'-[^{32}P]-labelled 20 nt primer annealed to an RNA template. Reactions were catalysed by *Taq* wt (1), M1 (2), M747K (3), M1/M747K (4) and RT-Taq 2 (5). Reactions were incubated for 30 s, 1 or 5 min, as indicated.

The experiments corroborated the initial findings of the respective *KlenTaq* variants and showed increased reverse transcriptase activity for RT-*Taq* 2 and *Taq* M1/M747K compared to the parental enzymes *Taq* M1 and M747K. Again template-independent nucleotide addition at blunt-ended DNA was observed in case of full-length product formation.[72, 107]

3.2.4 RT-PCR

The reverse transcriptase activity is essential for applications like *Taq*Man based real-time RT-PCR. Thus, the potential of the full-length variants in real-time RT-PCR was investigated using an MS2 bacteriophage RNA as template (**Figure 46**). Under the conditions studied, the generated *Taq* variants RT-*Taq* 2 and M1/M747K showed product formation approximately eight cycles earlier than the parental enzymes M1 and M747K. Expected product formation was determined via agarose gel analysis (**Figure 46**). Remarkably, the wild-type enzyme also exhibited a distinct reverse transcriptase activity with product formation visible after approximately 30 cycles. This is in good accordance with reported previous studies showing some reverse transcriptase activity for the *Taq* wild-type enzyme.[158]

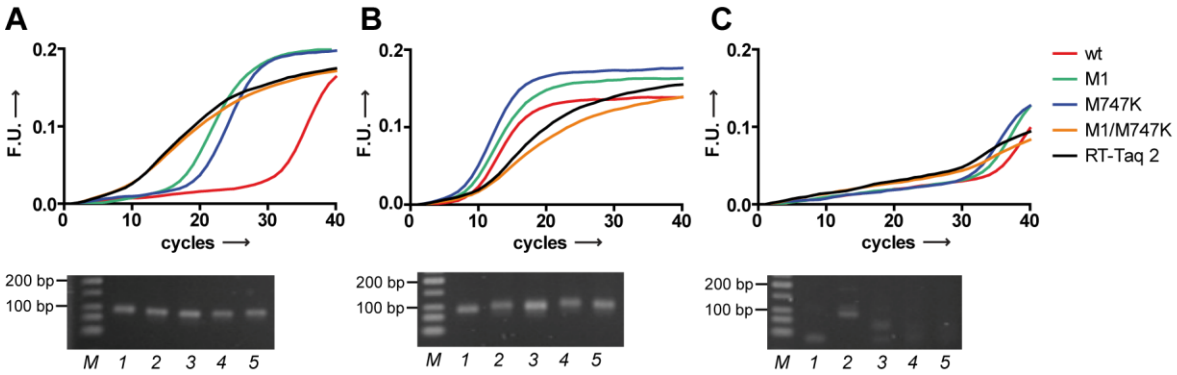


Figure 46. Real-time RT-PCR performed with *Taq* wild-type (wt) and variants (as indicated); F.U. = fluorescence units. The fluorescence readout was based on the binding of SYBRGreen I to double-stranded DNA. A) Amplification of a 100 bp target sequence from MS2 bacteriophage RNA. B) Amplification of a 100 bp target sequence from the respective DNA template. C) Negative control using water instead of RNA.
Product formation was analysed on a 2.5 % agarose gel as depicted below, with wt in lane 1, M1 in lane 2, M747K in lane 3, M1/M747K in lane 4, and RT-*Taq* 2 in lane 5.

As preceding experiments confirmed the endonuclease and reverse transcriptase activity of RT-*Taq* 2, the enzyme was subsequently employed in *Taq*Man based real-time RT-PCR (**Figure 47**). A 90 nt RNA stretch of human β-actin mRNA was targeted and amplified from total RNA isolated from Jurkat cells. Samples were analysed in the FAM channel using a *Taq*Man probe consisting of an oligonucleotide conjugated to FAM at the 5' end and a black hole quencher (BHQ-1) at the 3' end.

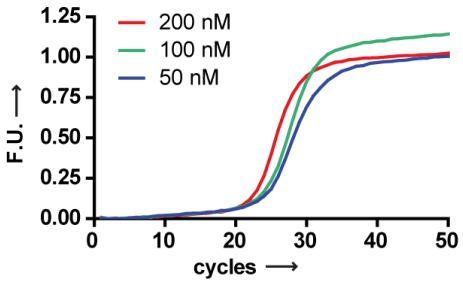


Figure 47. Detection of a 90 nt β-actin transcript from extracted human total RNA (Jurkat cells) using *Taq*Man based real-time RT-PCR. A polymerase dilution series was employed as indicated; F.U. = fluorescence units.

A fluorescence signal was observed in case of three different polymerase concentrations (50, 100 and 200 nM). Thus, the variant was successfully applied in RNA detection and shows potential for further applications in *Taq*Man based real-time RT-PCR.

3.3 Discussion

Although, a mixture of RT-KTq 2 and full-length *Taq* DNA polymerase wild-type was successfully applied in *Taq*Man based RT-PCR, the first experiments implementing the full-length variant RT-*Taq* 2 were promising regarding its nuclease and reverse transcriptase activity. The mutations present in the new variants can obviously influence the endonuclease activity, as seen for *Taq* M1/M747K. Nevertheless, the nuclease activity of RT-*Taq* 2 was comparable to the activity of the parental enzymes and is sufficient for its application in *Taq*Man based RT-PCR. Interestingly, an overall higher reverse transcriptase activity was observed for the full-length enzymes compared to the N-terminally truncated versions in primer extension experiments. However, variants *Taq* M1/M747K and RT-*Taq* 2 still exhibit enhanced reverse transcriptase activity compared to the parental enzymes. Furthermore, the RT-*Taq* 2 was successfully applied in RNA detection via *Taq*Man based real-time RT-PCR highlighting its potential for applications in diagnostics and molecular biology. RNA detection with other sequence contexts has to be explored. Additional experiments investigating the stability or other intrinsic properties of the enzymes are also of interest for their future application.

4. Substrate Spectrum Analysis of *KlenTaq* Variants

4.1 Introduction

Most DNA polymerases recognize their substrates with exceptional high specificity limiting their use in biotechnology. DNA polymerases with increased substrate spectra are therefore of great interest and a focus of directed evolution. A major goal of directed evolution is the production of enzymes with properties tailored for specific applications (see chapter **I 4, 5**). Especially, the evolution of *Taq* DNA polymerase represented a paramount aim in research, as the enzyme is employed in various biotechnological applications. Through various techniques *Taq* or *KlenTaq* mutants were evolved towards the ability to incorporate NTPs with higher efficiency than the wild-type,[195, 196, 199] to accept sugar-modified dNTPs or NTPs[214, 215] or to replicate from lesion containing templates.[72, 159, 215]. Even the acceptance of unnatural nucleotides was achieved, paving the road for the development of artificial genetic systems; thereby increasing the information potential of RNA and DNA.[216]
As both parental enzymes KTq M1 and M747K possess an increased substrate spectrum, the mutation combination of both was promising to yield variants with even more advanced properties. Therefore, KTq wild-type, parental enzymes M1 and M747K as well as the N-terminally truncated (*KlenTaq*) variants RT-KTq 1-4 were analysed regarding their substrate specificity.

4.2 Results

4.2.1 Acceptance of Damaged DNA

DNA polymerases capable of lesion bypass synthesis provide a crucial tool for paleontological, archaeological or forensic studies in which damaged DNA has to be processed in PCR.[217] However, most DNA damages pose a significant barrier for DNA polymerases and even prevent DNA synthesis.[81] Strong efforts were undertaken to evolve thermostable DNA polymerases with improved lesion tolerance property, respectively.[73, 159, 218, 219] As *KlenTaq* variant M747K is known to possess an increased lesion bypass activity that might contribute to a higher propensity to amplify from damaged DNA,[159] all identified *KlenTaq* variants were tested regarding their ability to accept and bypass DNA damage.
The wild-type, the parental enzymes and all new variants were employed in primer extension experiments using a DNA template containing a tetrahydrofuran abasic site analogue F. Reactions were incubated at 72 °C for 5 min. Product formation was analysed via denaturing PAGE (**Figure 48A**). We mainly observed stalling of KTq wild-type prior to the abasic site, whereas the parental enzymes M1 and M747K showed higher incorporation efficiency

opposite the lesion. The identified *KlenTaq* variants and KTq M1/M747K, on the contrary, exceeded the parental enzymes by far and showed increased full-length product formation.

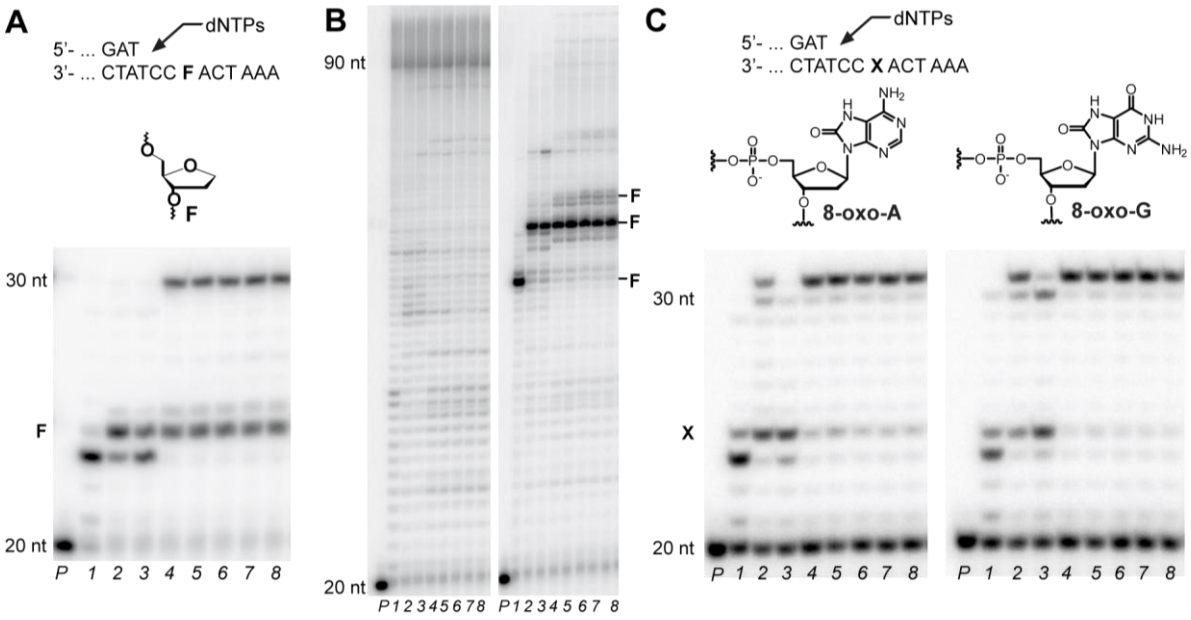


Figure 48. Lesion bypass ability of KTq wild-type (wt, 1), M1 (2), M747K (3), M1/M747K (4), RT-KTq 1 (5), RT-KTq 2 (6), RT-KTq 3 (7) and RT-KTq 4 (8). Primer extension reactions with different lesions are shown. A) Abasic site analogue F containing template (33 nt). Reactions were incubated for 5 min at 72 °C. B) Undamaged template (left) or three abasic site analogues F containing template (right, 90 nt). Reactions were incubated for 120 or 60 min at 72 °C, respectively. C) 8-oxo-dA (left) or 8-oxo-dG (right) containing templates. Reactions were incubated for 30 sec at 72 °C. X: respective lesion.
Partial primer and template sequences as well as the lesion structures are shown. Location of the lesion in the template strand is indicated. P: Primer.

Next, the ability of the identified variants to bypass three abasic sites in one template-strand was investigated (**Figure 48B**). The incubation time was increased to 60 min, as three lesions, distributed over the entire template, present a major challenge for DNA polymerase synthesis. Full-length product formation was observed for all enzymes in case of the respective non-damaged DNA template. Using the F-containing template, parental enzymes and selected variants extended the primer to full-length, however, minimal product formation was visible. Whereas the wild-type enzyme stalled in front of the first lesion in the template, the second abasic site presented the major challenge for all variants. High amounts of termination products were visible. However, the selected variants synthesized higher amounts of full-length product compared to the parental enzymes M1 and M747K and bypassed the third lesion with higher efficiency. Thus, increased lesion bypass efficiency was maintained in case of the newly identified variants. The acceptance of two additional endogenous lesions, 2'-

deoxy-7,8-dihydro-8-oxoadenosine (8-oxo-dA) and –8-oxoguanosine (8-oxo-dG), generated in nature by reactive oxygen species, was investigated in further experiments (**Figure 48C**). 8-oxo-dG is one of the most prevalent lesions found in DNA, besides abasic sites, and often the cause of transversion mutations, as dAMP is preferentially incorporated opposite this lesion.[87] Primer extension experiments were conducted and analysed via denaturing PAGE. All variants showed a higher bypass efficiency for these two lesions compared to the abasic site analogue F. Full-length product formation, as well as template-independent nucleotide addition, was observed for all variants. However, whereas the parental enzymes stalled in front of the lesion, almost no reaction termination products were visible for the selected variants, indicating higher lesion bypass activity for the identified variants.

Next, the performance of the newly evolved variants was tested on a highly damaged DNA template (**Figure 49**).

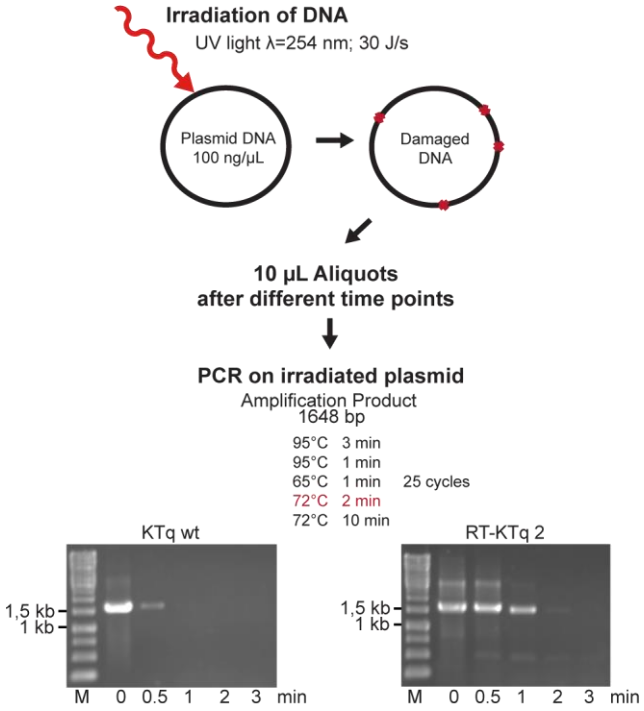


Figure 49. PCR amplification of a 1648 bp DNA fragment using UV-irradiated (λ = 254 nm) plasmid DNA as template. Plasmids were irradiated for 0, 0.5, 1, 2 and 3 min. M: Marker.

RT-KTq 2 was selected for this experiment, as the activity of the new variants in lesion bypass DNA synthesis was comparable and the variant RT-KTq 2 was already chosen in previous reverse transcription experiments for in-depth analysis.
To generate a highly damaged template, we introduced lesions on plasmid DNA by UV-light irradiation at λ=254 nm, took samples after various time points and subsequently used this DNA as template in PCR. While the wild-type enzyme failed to produce amplification from DNA that was irradiated for longer than 30 s, the reaction with RT-KTq 2 yielded products from DNA templates being irradiated for as long as 2 min (**Figure 49**).

4.2.2 Incorporation of Ribonucleotides

The high substrate specificity of DNA-dependent DNA polymerases is fundamental for genome stability.[160] Especially in cells, the discrimination between ribo- and desoxyribonucleotides is essential, as the amounts of NTPs exceed dNTPs by far. Most DNA-dependent DNA polymerases have evolved nucleotide selection mechanisms to restrict the misincorporation of NTPs into DNA. NTP exclusion is thought to occur via a steric exclusion model with an amino acid side chain acting as a 'steric gate'.[155, 195, 196, 198, 199, 220] It prevents NTP misincorporation due to a steric clash with the 2'-OH group of the ribose sugar. Although mutations of the 'steric gate' generally promote NTP incorporation by DNA polymerases, they do not by themselves enable the extension of a primer by NTP incorporation beyond short termination products. Geometric constraints in the binding cleft are suggested as one limiting factor, due to the fact that the DNA polymerase has to accommodate the distinct geometry of an RNA/DNA duplex rather than the B form helix of a cognate DNA duplex. Thus, the evolution of a DNA polymerase into an RNA polymerase can provide insights how polymerases discriminate between the nucleic acid substrates dNTP and NTP. Consequently, the ability of the *KlenTaq* variants to incorporate NTPs instead of dNTPs in a growing DNA primer strand was investigated.
Primer extension experiments were performed at 72 °C and the reactions analysed on a denaturing PAGE-gel (**Figure 50**). For KTq M1 and M747K only minor primer elongation was observed with a main product of four nucleotides incorporation formed for M1. In contrast, the newly identified variants showed high conversion of the primer after 30 min and up to eight nucleotides incorporation. Noticeably, the primer elongation is hindered by the incorporation of UMP as seen by the inability of the enzymes to produce full-length product (30 nt) after 30 min reaction time.

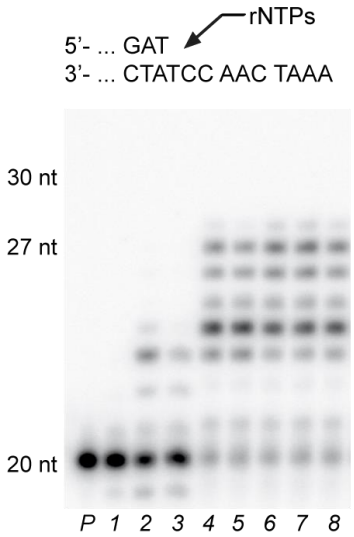


Figure 50. Incorporation of ribonucleotides.
Primer extension experiments with KTq wild-type (wt, 1), M1 (2), M747K (3), M1/M747K (4), RT-KTq 1 (5), RT-KTq 2 (6), RT-KTq 3 (7) and RT-KTq 4 (8). A 5'-[^{32}P]-labelled 20 nt primer was annealed to a DNA (33 nt) template. Reactions were incubated for 30 min. Partial primer/template sequence is depicted on top. P: Primer.

4.2.3 Error-Spectrum Analysis

A higher acceptance of modified substrates is often associated with a higher propensity of the enzyme to incorporate wrong nucleotides in basic DNA synthesis leading to a high error-rate. In order to gain insights into the fidelity of the identified DNA polymerases, a reported PCR-based assay was employed.[221] In short, a 1,648 bp fragment encoding for KTq wild-type was amplified in PCR followed by subsequent cloning of the PCR product and sequencing of several clones. Error-rates were determined for KTq wild-type, the parental enzymes, RT-KTq 1 and KTq M1/M747K the variants with the lowest and highest number of mutations, respectively, and RT-KTq 2 exhibiting the four most promising mutations (**Table 5**). In comparison, the error rate of the identified hits was approximately 10- to 16- fold increased to that found for KTq wild-type. In contrast to KTq M747K with an error-rate of about two-fold higher compared to the wild-type, the identified variants are in the same range as the parental enzyme M1. As expected, transitions are more abundant than transversions in all enzymes.

Table 5. Error rate and error-spectrum of KTq wild-type (wt) and variants.

Enzyme	No. of clones	Average no. of mutations per clone[a]	Error rate[b]	No. of individual substitutions						No. of deletions	No. of insertions
				Transitions		Transversions					
				AT → GC	GC → AT	AT → TA	AT → CG	GC → TA	GC → CG		
KTq wt	26	0.2	4.2×10^{-5}	6	0	0	0	0	0	0	0
KTq M1	23	2.4	48×10^{-5}	27	15	7	0	1	4	1	0
KTq M747K	26	0.5	9.1×10^{-5}	9	4	1	0	0	0	0	0
KTq M1/M747K	22	2.9	53×10^{-5}	33	7	15	2	4	0	2	1
RT-KTq 1	25	2.3	42×10^{-5}	27	12	9	1	2	2	4	0
RT-KTq 2	21	3.6	68×10^{-5}	29	14	14	2	7	5	5	0

[a] Number of mutations per 650 bases sequenced per clone.

[b] Error rate equals number of mutations per base per division.

4.3 Discussion

The combination of two variants M1 and M747K already known to possess an increased substrate spectrum yielded mutants with enhanced properties.

The ability to process DNA templates with different types of lesions was observed for all variants, with an efficiency exceeding the parental enzymes by far. All variants exhibited a higher efficiency in bypassing oxidative damage compared to an abasic site analogue F. A possible explanation might be the complete loss of genetic information in the abasic site, in contrast to residual information in the oxidative lesions, which might inhibit DNA synthesis stronger past the lesion.

Even with an irradiated, highly-damaged template, the RT-KTq 2 variant was more efficient than the wild-type. Regarding this experiment, one has to point out, that the mutant also possesses a higher sensitivity (data not shown) for the templating substrate in PCR than the wild-type enzyme. Therefore, the higher efficiency of the variant could be due to its higher lesion bypass activity, but also from the ability to amplify from a lower number of undamaged DNA copies still present in the reaction mixture. Both properties, however, would be an asset in paleontological, archaeological or forensic studies, in which the existent DNA is not only damaged, but often also low in numbers.[217]

The incorporation of ribonucleotides was observed for all enzymes, again with the identified variants outperforming the parental enzymes by far, but no full-length product was formed. This may be due to the nucleotide sequence of the template which directs for incorporation of three final UMPs. UMP incorporation was already described in the literature to severely hamper the elongation of the primer in RNA synthesis.[195] The selection against UTP by *Taq* DNA polymerase is 100-fold more stringent compared to the three other nucleotides.[195]

Furthermore, the rejection of NTPs by DNA polymerases, such as *Taq* DNA polymerase, was proposed to occur via an exclusion model: A single amino acid residue prevents the incorporation of NTPs by simply acting as a steric gate colliding with the 2'-OH group of the incoming ribonucleotide.[195, 198, 222] In *Taq* DNA polymerase E615 was reported to fulfil this role.[195] But a mutation at this position which might circumvent the discrimination against NTPs is not present in either of the identified variants. Therefore another mechanism has to apply in order to result in the increased ribonucleotide incorporation efficiency displayed by the mutant enzymes. It was previously suggested[155] that termination after six events of ribonucleotide incorporation results from the formation of an altered RNA/DNA duplex geometry upon proceeding RNA synthesis. RNA prefers a C3'-endo conformation, which poses no problem in the enzyme's active site, where the terminal three base pairs normally adopt this conformation.[30] However, downstream of these base pairs the nucleosides of a DNA duplex normally adopt a typical C2'-endo sugar pucker conformation. Thus, with ongoing primer elongation and ribonucleotide incorporation, the developing RNA/DNA hybrid duplex deviates from the preferred conformation and disrupts the enzyme-duplex interactions. The mutations L459M and M747K were already shown in this work to play an important role in adapting to the changed geometry of a DNA primer/RNA template duplex in reverse transcription. The same mechanism could be in effect, with the M747K mutation increasing the positively charged surface potential in close proximity to a negatively charged nucleic acid substrate and thus, foster the ability to accept aberrant substrates by electrostatic interactions. A higher flexibility of the thumb domain due to the L459M mutation may also contribute to some extent to avoid a clash with the duplex whose geometry changes with proceeding ribonucleotide incorporation.

The high error of the recombined variants was to be expected, as an increased substrate spectrum often correlates with lower fidelity. But for the applications in mind, such as the use of RT-KTq 2 in RT-PCR, forensic or archaeological studies, the error-rate plays a minor role. Especially in RT-PCR, the focus is on RNA detection and quantification instead of generating cDNA libraries. In fact, the high error propensity of the enzyme combined with the ability to perform PCR would allow the generation of mutated genes that are useful e.g. for directed enzyme evolution.[122]

IV. Summary

The high substrate specificity of replicative DNA polymerases is essential in order to maintain the stability of a cell's genome.[160] But furthermore, it is fundamental to understand how a DNA polymerase can adapt to or discriminate between aberrant or non-cognate substrates present in various biological processes. The evolution of DNA polymerases towards an enhanced acceptance of non-cognate substrates is also a paramount goal of directed enzyme evolution, as these enzymes pave the way for various biotechnological applications. In this study, the focus is on *KlenTaq* DNA polymerase adapting to abasic sites present in DNA and on the processing of RNA as a non-cognate substrate.

In the first part of this work, the enzyme's incorporation mechanism opposite an abasic site was investigated. Abasic sites are the result of a spontaneous or enzymatic loss of the nucleobase and are the most common damage in DNA observed under physiological conditions.[76] Although they are non-instructive, DNA polymerases from family A and B preferentially incorporate dAMP opposite the lesion, a phenomenon termed 'A-rule'.[78, 83, 84, 97-100, 102] This preference results in transversion mutations often found in human cancers.[112] Thus, it is vital to understand how DNA polymerases process these lesions and why dAMP is preferred. Superior desolvation and stacking interactions of adenine were always thought to be the major determinants for the 'A-rule'.[108, 109] However, Dr. Samra Obeid was able to obtain crystal structures of *KlenTaq* in complex with an abasic site analogue containing template, which suggested a possible bypass mechanism as well as the mechanistic origin for the purine selectivity of this DNA polymerase.[72] A highly conserved amino acid side chain (Tyr671) was suggested to replace the missing nucleobase in the template strand and mimic a six membered pyrimidine nucleobase in shape and size, thus directing for purine incorporation in order to maintain an optimal geometric fit at the active site. But, crystal structure analysis only provides a static view of nucleotide incorporation. For further evidence that an 'amino acid templating' mechanism is facilitating the incorporation of purines, functional studies were required and performed in this work. By mutating the tyrosine at position 671 via site-directed mutagenesis, variants were generated and subsequently analysed in primer extension experiments as well as kinetic studies to further corroborate the important role of this side chain in abasic site bypass.

Analysis of a Y671A variant demonstrated the significance of the aromatic moiety at this position in general, whereas a Y671F mutant highlighted the importance of hydrogen bond formation between the cognate tyrosine's hydroxyl group and the N3 of the incoming adenine. Incorporation studies of a modified nucleotide, namely 3-deaza-2'-deoxyadenosine-5'-triphosphate (d3DATP), opposite an abasic site, as well as primer extension experiments with a variant containing 2,3,5-fluorotyrosine at position 671, further confirmed the stabilizing effect of the hydrogen bond formed.

Analysis of a Y671W mutant provided the most compelling evidence that indeed an 'amino acid templating' mechanism is at work with an optimal geometric fit at the active site as the major determinant. In detail, primer extension experiments and kinetic analysis of Y671W

revealed that the purine selectivity could be switched to preferential pyrimidine incorporation. It is assumed that the bicyclic indole, consisting of a six-membered ring fused to a five-membered ring, mimics the approximate size and shape of a purine and in consequence directs for pyrimidine incorporation to conserve the geometric fit at the active site. Thus, the functional studies corroborated the lesion bypass mechanism derived from the crystal structure analysis.
In a second aspect of this work, template-independent addition of nucleotides at blunt-ended DNA was investigated.[107] The high preference for dAMP incorporation at blunt-ended DNA resembles the incorporation opposite an abasic site. Crystal structure analysis performed by Dr. Samra Obeid revealed the same arrangement in the active site of the enzyme with Tyr671 directing for preferential purine incorporation. Functional studies performed in this work with the tryptophane mutant Y671W corroborated these findings in primer extension and kinetic studies.

In the second project of this work, the acceptance of another non-cognate substrate of *KlenTaq* DNA polymerase was investigated. In general, the substrate selection of a DNA polymerase is a very stringent process;[160] the acceptance of non-cognate substrates, however, facilitates new applications in biotechnology, molecular biology and diagnostics. This part of my work focused on the ability of the DNA-dependent DNA polymerase *KlenTaq* to accept or exclude the non-cognate template RNA regarding two different aspects. First, the idea was to develop a *KlenTaq* variant which is able to accept DNA and RNA as the templating nucleic acid, thus facilitating the enzyme's application in reverse transcription PCR (RT-PCR), resulting in it being a valuable tool in transcriptome analysis, in pathogen detection as well as in disease-specific marker recognition.[14] Second, we wanted to obtain insights into the fundamental biological question of how a DNA-dependent DNA polymerase can discriminate between DNA and RNA as a templating nucleic acid.
The detection and quantification of RNA in RT-PCR is generally mediated by two enzymes: a reverse transcriptase and a PCR-competent DNA polymerase. However, several drawbacks arise from the heat-instability of retroviral reverse transcriptases.[143] Thus, the use of one enzyme, a thermostable DNA-dependent DNA polymerase capable of reverse transcription, would offer the possibility to perform one step RT-PCR at high temperatures minimizing secondary structure formation of RNA,[142] enhance specificity[142] and, in general, supersede the use of two enzymes, thereby eliminating problems such as inhibitory effects between a reverse transcriptase and a DNA polymerase.[152-154] An additional reverse transcription step, necessary for the reverse transcriptase, could be omitted, further providing a work and time reduction.
For this purpose, two *KlenTaq* mutants, M1 (L322M, L459M, S515R, I638F, S739G, E773G) and M747K, exhibiting some reverse transcriptase[148] and lesion bypass activity,[159] were recombined by DNA shuffling. Screening of the generated library for PCR and reverse

transcriptase activity yielded enzyme mutants with properties by far exceeding the parental enzymes. Sequencing and functional characterization of the most promising variants indicated that four mutations (L459M, S515R, I638F and M747K) are important for the observed increase in reverse transcriptase activity. One mutant which featured one mutation less showed an overall loss in enzyme activity, whereas variants with more mutations resulted in no gain in reverse transcriptase activity.

In summary, we gained a new generation of *KlenTaq* variants which exhibit an up to 20-fold gain in reverse transcriptase activity compared to the parental variant KTq M1, and a more than 100-fold increase compared to KTq M747K. The results suggest a surprising synergistic effect of the mutations. Further RT-PCR experiments with the purified enzymes confirmed the superior reverse transcriptase activity of the identified variants. RT-KTq 2, which possesses the most promising mutations L459M, S515R, I638F and M747K, was selected for in depth-analysis. The enzyme was successfully applied in real-time multiplex RT-PCR, facilitating rapid detection of influenza viruses A and B. Thus, it provides a valuable tool for rapid RT-PCR, crucial e.g. in important fields of clinical diagnostics such as point of care testing, as a reverse transcription step can be omitted and preceding reaction condition optimizations for a reverse transcriptase and a DNA polymerase are redundant. The thermostable enzyme further offers the possibility to perform RT-PCR at high temperatures, preventing unspecific priming and secondary structure formation of mRNA. The full-length enzyme of the variant showed further promising properties which facilitates its future application in *Taq*Man based RT-PCR.

Few thermostable enzymes capable of both reverse transcription and PCR are commercially available,[147, 151] probably also due to the lack of knowledge regarding the mechanism of how DNA-dependent DNA polymerases discriminate between the templating nucleic acids DNA and RNA. Therefore, it is difficult to rationally design such DNA polymerases. The lack of knowledge is due to the fact that structural data of DNA-dependent DNA polymerases, whose wild-type parental ancestors show no significant reverse transcriptase activity, processing RNA is not available. Thus, we set out to crystallize *KlenTaq* DNA polymerase in complex with RNA as substrate to get insights into the fundamental question how DNA-dependent DNA polymerases discriminate between the templating nucleic acids DNA and RNA.

Although crystallization trials of *KlenTaq* wild-type in complex with a DNA/RNA hybrid failed, we were able to obtain crystals of the RT-KTq 2 variant in complex with an all DNA as well as a DNA/RNA hybrid duplex and could solve the structures. For the first time, we received insights into how a DNA-dependent DNA polymerase processes RNA. Together with Dipl.-Biol. Konrad Bergen, we identified the crucial role of the mutations present in the enzyme variant which enabled the mutant to accommodate the distinct geometry of the hybrid duplex, allowing an efficient processing of RNA as template. Consequently, we also gained insights into the restraints present in the wild-type enzyme which prevent efficient RNA processing.

The L459M mutation contributes to prevent the clash of the thumb domain with the hybrid duplex and facilitates the nucleotide binding motif in the thumb domain to maintain its interactions with the nucleic acid duplex. Furthermore, the S515R mutation stabilizes this motif by interacting with surrounding amino acid residues. Whereas the role of I638F is difficult to predict, we suggest that the M747K mutation increases the positively charged surface potential in the proximity of the negatively charged substrate backbone and thus enhances the acceptance of non-cognate substrates by electrostatic interactions.
A last aspect of this work focused on the increased substrate spectra of the new *KlenTaq* variants. With these enzymes in hand, many applications come to mind. The increased lesion bypass activity makes them valuable tools for paleontological, archaeological or forensic studies in which damaged DNA has to be processed in PCR reactions.[217] The high error-rate combined with the ability to perform PCR would allow the generation of mutated genes that are useful e.g., for directed enzyme evolution.[122] Further applications were already investigated in our lab such as the use of RT-KTq 2 in allele-specific PCR, direct amplification from blood samples and the acceptance of commercially available modified nucleotides, but are not part of this work. But it reveals the high potential of these enzymes facilitating many PCR-based applications.

This work consisted of two major studies which both investigated DNA synthesis by *KlenTaq* DNA polymerase from two aberrant substrates, abasic site containing templates and RNA. Both studies give insights into the nature of template processing by DNA polymerases: In general, one can conclude that these enzymes possess a certain flexibility to adapt to non-cognate substrates. This flexibility can be altered or enhanced by directed enzyme evolution. In detail, an amino acid residue assumes the role of the templating nucleobase in abasic site bypass and directs for preferential purine incorporation. However, mutation of this residue still allowed the enzyme to bypass the lesion, but altered the enzyme's bypass behaviour towards preferential pyrimidine incorporation. Additionally, the poor processing of RNA as template by *KlenTaq* DNA polymerase was enhanced by introducing specific mutations into the enzyme. Thus, the intrinsic poor binding and processing of the non-cognate substrate RNA was efficiently increased by directed enzyme evolution and resulted in the enzyme's high reverse transcriptase activity.
In this work, directed enzyme evolution also facilitated the investigation of biological processes and mechanisms. Mutating the tyrosine residue 671 to tryptophane approved a proposed model for the 'A-rule'. Furthermore, evolution of *KlenTaq* DNA polymerase wild-type towards a mutant capable of processing RNA allowed insights into the restrictions present in the wild-type enzyme, which might be responsible for discriminating between the two templates DNA and RNA. Thus, by directed enzyme evolution, insights were gained into fundamental principles of DNA synthesis.

V. Zusammenfassung

Die hohe Substratspezifität von DNA Polymerasen ist essentiell, um die Stabilität des Genoms einer Zelle aufrechtzuerhalten.[160] Aber darüber hinaus ist es wichtig zu verstehen, wie DNA-Polymerasen in der Lage sind, sich an nicht-natürliche Substrate, die in diversen biologischen Prozessen vorkommen, anzupassen oder gegen sie zu diskriminieren.
Des weiteren ist die Evolution von DNA-Polymerasen, hin zu einer erweiterten Akzeptanz von nicht-natürlichen Substraten, auch ein vorrangiges Ziel der gerichteten Enzymevolution, da diese Enzyme in diversen Bereichen der Biotechnologie Anwendungen finden würden. Im Fokus dieser Arbeit steht die Anpassung an, in der DNA vorkommende, abasische Stellen und die Prozessierung von RNA als nicht-natürliches Substrat durch die *KlenTaq* DNA-Polymerase.

Der erste Teil dieser vorliegenden Arbeit befasst sich mit der Frage, welcher Mechanismus dem Nukleotideinbau gegenüber abasischen Stellen zu Grunde liegt. Abasische Stellen sind das Resultat eines spontanen oder enzym-katalysierten Verlustes der Nukleobase und stellen die häufigste Art von DNA-Schäden unter physiologischen Bedingungen dar.[76] Obwohl abasische Stellen die genetische Information, die in der Nukleobase gespeichert war, verloren haben, bauen DNA-Polymerasen der Familien A und B bevorzugt dAMP gegenüber einer solchen DNA-Läsion ein.[78, 83, 84, 97-100, 102] Ein Phänomen das als „A-rule“ bezeichnet wird. Diese Präferenz für dAMP führt jedoch zu Transversionsmutationen, die oft mit Krebs in Verbindung gebracht werden.[112] Daher ist es essentiell zu verstehen, wie DNA-Polymerasen diese Läsionen prozessieren und warum dAMP bevorzugt wird. Die besseren Desolvatisierungseigenschaften sowie die besseren π-π Wechselwirkungen von Adenin wurden bisher als die bestimmenden Faktoren für die „A-rule“ angesehen.[108, 109] Dr. Samra Obeid war jedoch in der Lage Kristallstrukturen der *KlenTaq* DNA-Polymerase im Komplex mit einem Templat, das eine abasische Stelle enthält, zu erhalten.[72] Diese Kristallstruktur deutet auf einen möglichen Mechanismus hin, durch welchen erklärt werden kann, wie das Enzym die DNA-Läsion überliest und worin die Präferenz für Purine begründet liegt: Eine hoch konservierte Seitenkette (Tyr671) soll die fehlende Nukleobase im Templatstrang ersetzen.[72] Durch ihre Ähnlichkeit mit einer Pyrimidinnukleobase bezüglich ihrer Form und Größe wird dadurch der Einbau von Purinen bevorzugt, da dadurch die geometrische Form des aktiven Zentrums nicht beeinträchtigt wird. Die Kristallstrukturanalyse bietet jedoch nur eine Momentaufnahme des Einbaus von Nukleotiden. In diesem Teil der Arbeit wurden funktionelle Studien durchgeführt, um weitere Anhaltspunkte zu finden, die einen Protein-Templat-Mechanismus als zugrunde liegenden Faktor für die Präferenz des Purineinbaus bestätigen. Durch die Mutation des Tyrosins an der Stelle 671 über zielgerichtete Mutagenese, wurden Polymerase-Varianten generiert, die anschließend in Primerverlängerungsreaktionen und kinetischen Studien untersucht wurden. Diese Studien belegten die wichtige Rolle der Seitenkette im Überlesen von abasischen Stellen.
Die Analyse einer Y671A-Variante demonstrierte die Bedeutung des aromatischen Restes an dieser Position im Allgemeinen, während eine Y671F-Variante die Wichtigkeit einer

bestehenden Wasserstoffbrücke zwischen der Hydroxylgruppe des natürlichen Tyrosins und dem N3 des eintretenden dAMP verdeutlichte. Sowohl Einbaustudien des modifizierten Nukleotids 3-Deaza-2'-desoxyadenosin-5'-triphosphat (d3DATP) gegenüber einer abasischen Stelle, als auch Primerverlängerungsreaktionen mit einer Variante, bei welcher das Tyrosin an der Stelle 671 mit einem 2,3,5-Fluorotyrosin ersetzt wurde, belegten das Vorhandensein und die stabilisierende Wirkung der gebildeten Wasserstoffbrücke.
Die Analyse einer Y671W-Variante lieferte jedoch den überzeugendsten Beweis dafür, dass ein Protein-Templat-Mechanismus, mit geometrischen Beschränkungen als wichtigster Faktor, dem Ganzen zu Grunde liegt. Im Einzelnen zeigte die Analyse der Primerverlängerungsreaktionen und der kinetischen Studien der Y671W-Variante, dass der bevorzugte Einbau von Purinen umgekehrt werden kann, sodass Pyrimidine bevorzugt als Substrat akzeptiert werden. Es wird vermutet, dass das Indolsystem des Tryptophans, welches aus einem mit einem Fünfring gekoppeltem Sechsring besteht, der Größe und Form eines Purins ähnelt und daher bevorzugt Pyrimidine eingebaut werden, um eine optimale Anpassung innerhalb des aktiven Zentrums zu gewährleisten. Dadurch konnte der in der Kristallstruktur angedeutete Mechanismus über in dieser Arbeit durchgeführte funktionelle Studien bestätigt werden.
Ein anderer Gesichtspunkt dieser Arbeit bestand in der Eigenschaft von DNA-Polymerasen einen templatunabhängigen 3'-Überhang zu generieren.[107] Der templatunabhängige, bevorzugte Einbau von dAMP am 3'-Ende der DNA erinnert an den Einbau gegenüber einer abasischen Stelle. Die von Dr. Samara Obeid durchgeführte Kristallstrukturanalyse offenbarte dieselbe Anordnung im aktiven Zentrum des Enzyms, die wiederum nahelegte, dass die Tyrosinseitenkette an Position 671 verantwortlich für den bevorzugten Einbau von Purinen war. Funktionelle Studien der Y671W-Variante, im speziellen Primerverlängerungsreaktionen und kinetische Studien, die in dieser Arbeit durchgeführt wurden, belegten die aus der Kristallstruktur gewonnenen Erkenntnisse.

Der zweite Teil dieser Arbeit befasst sich mit der Akzeptanz eines anderen nicht-natürlichen Substrates durch die *KlenTaq* DNA-Polymerase. Im Allgemeinen ist die Substratselektion einer DNA-Polymerase ein sehr stringenter Prozess.[160] Jedoch eröffnet die Akzeptanz von nicht-natürlichen Substraten neue Anwendungen in der Biotechnologie, der Molekularbiologie und der Diagnostik. Dieser Teil meiner Arbeit befasst sich mit der Fähigkeit der DNA-abhängigen DNA-Polymerase *KlenTaq* das nicht-natürliche Templat RNA zu akzeptieren bzw. auszuschließen. Dies wurde hinsichtlich zwei verschiedener Gesichtspunkte untersucht. Zum einen war es das Ziel eine *KlenTaq*-Variante zu entwickeln, die in der Lage ist, sowohl DNA als auch RNA als Templat zu akzeptieren. Dies würde die Anwendung des Enzyms in reverser Transkriptions PCR (RT-PCR) ermöglichen und es zu einem wichtigen Werkzeug für die Transkriptomanalyse oder für die Erkennung von Pathogenen und krankheitsspezifischen Markern werden lassen.[14] Zum anderen wollten wir Einblicke in die grundlegende

biologische Frage, wie eine DNA-abhängige DNA-Polymerase zwischen den Nukleinsäuretemplaten DNA und RNA unterscheiden kann, erhalten.

Die Detektion und Quantifizierung von RNA wird in RT-PCR grundsätzlich durch zwei verschiedene Enzyme realisiert: einer reversen Transkriptase und einer PCR-fähigen DNA-Polymerase. Allerdings ist die Hitzeinstabilität von retroviralen reversen Transkriptasen ein großer Nachteil.[143] Der Einsatz einer thermostabilen DNA–abhängigen DNA-Polymerase mit der Fähigkeit zur reversen Transkription würde daher die Anwendung in RT-PCR bei hohen Temperaturen ermöglichen. Hohe Temperaturen verringern die Ausbildung von Sekundärstrukturen der RNA,[142] fördern die Spezifität[142] und würden im Allgemeinen die Verwendung von zwei Enzymen unnötig machen. Dadurch würden Probleme, wie der inhibitorische Effekt zwischen einer reversen Transkriptase und einer DNA-Polymerase eliminiert werden.[152-154] Darüber hinaus könnte ein zusätzlicher RT-Schritt, notwendig für die reverse Transkriptase, ausgelassen werden, was eine Arbeits- und Zeitersparnis zur Folge hätte.

Zu diesem Zweck wurden zwei *KlenTaq* (KTq)-Varianten M1 (L322M, L459M, S515R, I638F, S739G, E773G) und M747K über DNA-Shuffling kombiniert, um eine Enzymbibliothek mit allen möglichen Mutationskombinationen zu erhalten. Der Literatur zufolge besitzen beide Enzyme eine gewisse reverse Transkriptase-Aktivität,[148] wie auch die Fähigkeit DNA-Schäden zu überlesen[159]. Die Durchmusterung der entstandenen Varianten-Bibliothek, bezüglich PCR- und reverser Transkriptase-Aktivität, ergab Enzymmutanten mit Eigenschaften, die die Aktivitäten der parentalen Enzyme bei weitem übertrafen. Die Sequenzierung und funktionelle Charakterisierung der vielversprechendsten Varianten deutete darauf hin, dass vier Mutationen (L459M, S515R, I638F, M747K) für den beobachteten Anstieg der reversen Transkriptase-Aktivität wichtig sind. Während das Fehlen der L459M-Mutation in einer Variante einen Verlust an Gesamtaktivität zur Folge hatte, führten zusätzliche Mutationen zu keinem Gewinn an reverser Transkriptase-Aktivität. Zusammenfassend kann gesagt werden, dass eine neue Generation von *KlenTaq*-Varianten entwickelt werden konnte, die gegenüber der parentalen Variante KTq M1 eine bis zu 20-fach, und gegenüber KTq M747K eine bis zu 100-fach, erhöhte reverse Transkriptase-Aktivität aufweist. Die Ergebnisse deuten auf einen unerwarteten synergistischen Effekt der Mutationen hin. Weitere RT-PCR Experimente mit den gereinigten Enzymen bestätigten die überlegene reverse Transkriptase-Aktivität der gefundenen Varianten. RT-KTq 2, die die vier vielversprechendsten Mutationen (L459M, S515R, I638F, M747K) aufweist, wurde in weiteren Experimenten untersucht. Das Enzym wurde erfolgreich in *Taq*Man basierter Echtzeit-RT-PCR eingesetzt und ermöglichte die schnelle Detektion von Influenza Viren A und B.

Das Enzym ermöglicht den Einsatz in schneller RT-PCR, da ein RT-Schritt nicht notwendig ist und vorhergehende Reaktionsoptimierungen für eine reverse Transkriptase und DNA-Polymerase überflüssig sind. Die RT-PCR wiederum ist essentiell für z.B. wichtige Bereiche

der klinischen Diagnostik, wie der patientennahen Labordiagnostik (POCT). Das thermostabile Enzym erlaubt weiterhin die RT-PCR bei hohen Temperaturen durchzuführen, die die unspezifische Bindung von Primern und die Bildung von Sekundärstrukturen der RNA verhindern. Die Volllängenenzyme der Varianten zeigten darüber hinaus vielversprechende Eigenschaften, die den Einsatz in *Taq*Man basierter RT-PCR realisieren.

Momentan sind nur wenige thermostabile Enzyme kommerziell erhältlich,[147, 151] die in der Lage sind, sowohl revers zu transkribieren als auch DNA zu amplifizieren. Dies liegt wahrscheinlich auch an dem Mangel an Kenntnissen bezüglich des Mechanismus den DNA-Polymerasen nutzen, um zwischen den Nukleinsäuretemplaten DNA und RNA zu diskriminieren. Daher ist es schwierig solche DNA-Polymerasen rational zu entwickeln. Dieser Mangel an Wissen resultiert vermutlich daraus, dass von DNA-abhängigen DNA-Polymerasen, deren Wildtyp-Vorläufer keine signifikante reverse Transkriptase-Aktivität aufweisen, keine strukturellen Daten vorliegen, die zeigen wie diese Enzyme RNA prozessieren. Daher setzten wir es uns als Ziel die *KlenTaq* DNA-Polymerase im Komplex mit RNA als Templat zu kristallisieren, um Einblicke zu gewinnen, wie DNA-abhängige DNA-Polymerasen zwischen den verschiedenen Nukleinsäuretemplaten unterscheiden können.

Obwohl es uns nicht gelang *KlenTaq*-Wildtyp in Komplex mit einem DNA/RNA-Hybriden zu kristallisieren, konnten wir Kristalle der RT-KTq 2 Variante, sowohl in Komplex mit einem DNA/DNA-Duplex als auch mit einem DNA/RNA-Hybrid-Duplex, erhalten und deren Struktur lösen. Die Kristallstruktur erlaubte zum ersten Mal Einblicke, wie eine DNA-abhängige DNA-Polymerase RNA prozessiert. Zusammen mit Dipl.-Biol. Konrad Bergen gelang es, die wichtige Rolle der Enzymmutationen zu identifizieren, die es der Mutante erlauben die unterschiedliche Geometrie des Hybrid-Duplexes zu akzeptieren und dadurch RNA als Templat effizient zu prozessieren.

Die L459M Mutation trägt dazu bei, dass eine Kollision der Daumendomäne mit dem Hybrid-Duplex verhindert wird und ermöglicht gleichzeitig dass ein Nukleotidbindemotiv in der Daumendomäne ihre Interaktionen mit dem Nukeinsäureduplex beibehält. Darüber hinaus stabilisiert die S515R Mutation dieses Motiv durch Interaktionen mit umgebenden Aminosäureresten. Während die Rolle der I638F Mutation schwer vorherzusagen ist, vermuten wir, dass die M747K Mutation die positive Oberflächenladung in der Nähe des negativ geladenen Substratrückgrats erhöht. Eine daraus folgende, erhöhte elektrostatische Interaktion könnte die Bindung des Nukleinsäurerückgrats erleichtern und dadurch die Fähigkeit, ungewöhnliche Substrate zu akzeptieren, fördern.

Ein letzter Gesichtspunkt der vorliegenden Arbeit befasst sich mit dem erweiterten Substratspektrum der neuen *KlenTaq*-Varianten. Die Anwendungsmöglichkeiten scheinen vielseitig zu sein. Mit ihrer erhöhten Eigenschaft DNA-Läsionen zu überlesen, wären sie nützliche Werkzeuge in paläontologischen, archäologischen oder forensischen Studien in denen geschädigte DNA in PCR-Reaktionen prozessiert werden muss.[217] Die hohe Fehlerrate, kombiniert mit der Fähigkeit DNA in PCR zu amplifizieren, würde es erlauben Mutationen in

Gene einzuführen, was z.B. Anwendung in der gerichteten Enzymevolution finden würde.[122] Weitere Anwendungen wurden bereits untersucht, wie z.B. der Einsatz von RT-KTq 2 in allel-spezifischer PCR, der direkten Amplifikation von DNA aus Blutproben und die Akzeptanz von kommerziell verfügbaren, modifizierten Nukleotiden. Diese Studien sind aber nicht Teil dieser Arbeit. Sie zeigen jedoch das große Potential dieser Enzyme in PCR-basierten Anwendungen.

Diese Arbeit bestand aus zwei großen Studien, die beide das Ziel hatten, die DNA-Synthese, katalysiert durch die *KlenTaq* DNA Polymerase, ausgehend von zwei ungewöhnlichen Substraten zu untersuchen: die DNA-Synthese ausgehend von abasischen Stellen enthaltenden Templaten und RNA.
Beide Studien geben Einblicke in die Natur der Templatprozessierung durch DNA-Polymerasen: Allgemein ist zu sagen, dass diese Enzyme eine gewisse Flexibilität besitzen, um sich an ungewöhnliche Substrate anzupassen. Diese Flexibilität kann durch gerichtete Enzymevolution verändert oder sogar erhöht werden. Detailliert betrachtet, übernimmt eine Aminosäure im Überlesen abasischer Stellen die Rolle der Nukleobase im Templat, wodurch bevorzugt Purine eingebaut werden. Die Mutation dieser Aminosäure verhinderte nicht den Bypass der abasischen Stelle durch das Enzym, sondern veränderte nur das Bypass-Verhalten hin zu einer Präferenz von Pyrimidinen. Zusätzlich konnten Mutationen in das Wildtyp-Enzym eingebracht werden, die die Prozessierung von RNA als Templat erhöhen. Die schwache intrinsische Bindung und Prozessierung des nicht-natürlichen Substrats RNA wurden durch gerichtete Enzymevolution verbessert und hatte eine höhere reverse Transkriptase-Aktivität des Enzyms zur Folge.
In dieser Arbeit ermöglichte die gerichtete Enzymevolution auch das Untersuchen von biologischen Prozessen und Mechanismen. Die Mutation der Tyrosinseitenkette zu Tryptophan bestätigte das beschriebene Modell der „A-rule“. Darüber hinaus erlaubte die Evolution einer *KlenTaq* DNA-Polymerase Mutante, mit der Fähigkeit RNA-Template zu prozessieren, Einblicke in die Einschränkungen, die der Wildtyp besitzt, und die vermutlich verantwortlich für die Diskriminierung zwischen RNA und DNA sind.
Durch gerichtete Enzymevolution konnten daher Einblicke in die der DNA-Synthese zu Grunde liegende Prinzipien erhalten werden.

VI. Material and Methods

1. Material

1.1 General

Chemicals were used in p.a. or molecular biology grade quality. Water used for buffers was obtained from a combined reverse osmosis/ultrapure water system (Milli-Q, Sartorius). Water used for reactions containing RNA and DNA was from a bidistillation apparatus.

1.2 Chemicals

[γ-32P]-ATP	Hartmann Analytik
Acetic Acid	Roth
Acrylamide-bisacrylamide (25 %), Rotiphorese sequencing gel concentrate (19:1)	Roth
Acrylamide-bisacrylamide (30 %), Rotiphorese gel 30 (37.5:1)	Roth
Agar-agar	Roth
Agarose (LE)	Roche
Ammonium sulfate	Roth
Ammoniumperoxodisulfate (APS)	Roth
β-Mercaptoethanol	Merck
Benzamidine hydrochloride hydrate	Fluka
Bromophenol blue	Roth
Calcium chloride	Merck
Carbenicillin disodium salt	Roth
Chloroform-Isoamyl alcohol (24:1)	Roth
Coomassie Brilliant Blue G 250	Roth
Coomassie Roti-Blue (5x)	Roth
Coomassie Roti-Quant (5x)	Roth
DMSO	New England Biolabs
1,4-Dithiothreitol (DTT)	Roth
Ethanol (100 %)	Sigma-Aldrich
Ethidium bromide (0.1 %)	Roth
Ethylenediaminetetraacetic acid (EDTA) disodium salt	Roth
Formamide	Roth
Glucose	Sigma-Aldrich
Glycerol	VWR
Hydrochloric acid (37 %)	Merck
Imidazole	Merck
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth
LB-broth	Roth

Magnesium chloride hexahydrate	Acros Organics
Magnesium sulfate	Merck
Manganese chloride	Riedel de Haën
Methanol	Sigma-Aldrich
N, N, N', N'-Tetramethylethylenediamine (TEMED)	Roth
Ni-IDA sepharose (Chelating Sepharose Fast Flow)	GE Healthcare
Phenol-Chloroform-Isoamyl alcohol (25:24:1)	Roth
Phenylmethylsulfonyl fluoride (PMSF)	Roth
Polyethylenimine (PEI)	Sigma-Aldrich
Potassium chloride	Roth
Potassium phosphate dibasic K_2HPO_4	Sigma-Aldrich
Potassium phosphate monobasic KH_2PO_4	Roth
2-Propanol	Sigma-Aldrich
Q sepharose	GE Healthcare
Sephadex G-25 Superfine	GE Healthcare
SOB-medium	Roth
Sodium acetate	Merck
Sodium chloride	Roth
Sodium dodecyl sulfate (SDS)	Roth
Sodium hydroxide	Riedel de Haën
SYBRGreenI	Fluka
Tris-(hydroxymethyl)-aminomethane (TrizmaBase)	Sigma Aldrich
TritonX-100	Roth
Trypton	Roth
Tween20	Roth
Urea	Roth
Xylene cyanol	Roth

1.3 Nucleotide Triphosphates and Radiochemicals

ATP, GTP, CTP, UTP	Fermentas, Roche
dATP, dGTP, dCTP, TTP	Fermentas, Roche
[γ-^{32}P]-ATP	Hartmann Analytik

1.4 Oligonucleotides

Oligonucleotides were obtained from Purimex, Metabion or ThermoScientific. Cloning primers were used desalted, template oligonucleotides HPLC-purified. Primers used for primer extension studies were purified by preparative PAGE. Primers for site-directed mutagenesis and crystallization studies were also ordered HPLC-purified.

1.5 DNA and Protein Standards

Bovine Serum Albumine Standard (BSA, 2 mg/mL)	Thermo Scientific
Gene Ruler DNA ladder mix	Fermentas
Gene Ruler 1kb DNA ladder	Fermentas
Gene Ruler High Range DNA Ladder	Fermentas
Low Molecular Weight DNA Ladder	New England Biolabs
Page Ruler Unstained Protein Ladder	Fermentas

1.6 Enzymes and Proteins

Antarctic phosphatase	New England Biolabs
DNase I, RNase-free	Roche, Fermentas
DpnI	New England Biolabs
HindIII	Fermentas
Lysozyme	Roth
Phusion™ High Fidelity DNA polymerase	Finnzymes
Shrimp alkaline phosphatase (SAP)	Fermentas
SphI (PaeI)	Fermentas
Taq DNA polymerase	R. Kranaster; Fermentas
T4 DNA ligase	Fermentas, Roche
T4 polynucleotidekinase (PNK)	Fermentas

1.7 Kits

JBScreen Nuc-Pro HTS	Jena Bioscience
MinElute Reaction Cleanup Kit	Qiagen
peqGOLD Gel Extraction Kit	PEQLAB
peqGOLD Plasmid Miniprep Kit	PEQLAB
QIAquick Gel Extraction Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
QuikChange Multi Site Directed Mutagenesis Kit	Stratagene
Rapid DNA ligation Kit	Fermentas

1.8 Bacterial Strains and Plasmids

E. coli XL10-Gold	Stratagene *Genotype: Tetr Δ (mcrA)183* *Δ (mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lacHte [F¢ proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr]*
E. coli BL21-Gold(DE3)	Stratagene *Genotype: E. coli B F- ompT hsdS(r_B- m_B-) dcm+ Tetr gal λ(DE3) endA Hte*
pGDR11[184]	Ampicilin resistance; derivative of pEQ31 (Qiagen) containing lacIq gene
pASK-IBA 37+	Ampicilin resistance; IBA

1.9 Media

LB-medium	2 % (w/v) LB-Broth
LB-agar plates	2 % (w/v) LB-Broth 2 % (w/v) Agar
SOC-medium	2 % (w/v) Tryptone 0,5 % (w/v) Yeast Extract 0,05 % (w/v) NaCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM Glucose

1.10 Buffers and Solutions

1.10.1 Buffers and Solutions for Electrophoresis

1x Agarose Running Buffer (TAE)	40 mM Tris·HCl pH 7.5 40 mM Acetic acid 1 mM EDTA

0.5x Agarose Running Buffer (TBE)	45 mM Trizma Base 45 mM Boric acid 1 mM EDTA pH 8.0
Agarose gel staining solution	0.5x TBE/1x TAE supplemented with ethidium bromide
6x Agarose DNA Loading Dye	Fermentas: 10 mM Tris-HCl pH 7.6 0.03 % Bromophenol blue 0.03 % Xylene cyanol FF 60 % Glycerol 60 mM EDTA
Coomassie colloidal staining solution	20 % (v/v) Methanol (tech.) 1x Roti-Blue
Coomassie colloidal destaining solution	25 % (v/v) Methanol (tech.)
Coomassie Roti-Blue in methanol	0.115 % Coomassie Roti-Blue 10 % Acetic acid 50 % Methanol stirred overnight and filtrated
Coomassie methanol staining solution	50 % (v/v) Coomassie Roti-Blue in methanol 10 % (v/v) Acetic acid (tech.)
Coomassie methanol destaining solution	50 % (v/v) Methanol (tech.) 10 % (v/v) Acetic acid (tech.)
1x Denaturing PAGE Loading Dye (stop solution)	80 % (v/v) Formamide 20 mM EDTA 0.25 % (w/v) Bromophenol blue 0.25 % (w/v) Xylene cyanol
1x Denaturing PAGE Gel Solution I	25 % Acrylamide in 8.3 M Urea 2 % N,N´-Methylenbisacrylamide
1x Denaturing PAGE Gel Solution II	8,3 M Urea in 10x TBE

1x Denaturing PAGE Gel Solution III	8,3 M Urea
Denaturing PAGE Running buffer	1x TBE
6x SDS-PAGE Loading Dye	225 mM Tris-HCl pH 6.8 50 % (v/v) Glycerol 5 % (w/v) SDS 0.05 % (w/v) Bromophenol blue 12.5 % (v/v) β-Mercaptoethanol
10x SDS-PAGE Running Buffer	250 mM Tris-HCl pH 8.9 2 M Glycine 1 % (w/v) SDS
SDS-PAGE Separating Gel Solution (12 %)	375 mM Tris-HCl pH 8.8 0.1 % (w/v) SDS 12 % Acrylamide/ Bisacrylamide (37.5:1) 0.1 % (w/v) APS 0.1 % (w/v) TEMED
SDS-PAGE Stacking Gel Solution (4 %)	250 mM Tris-HCl pH 6.8 0.05 % (w/v) SDS 4 % Acrylamide/ Bisacrylamide (37.5:1) 0.1 % (w/v) APS 0.25 % (w/v) TEMED
10x TBE	900 mM Trizma Base 900 mM Boric Acid 20 mM EDTA pH 8.0

1.10.2 Buffers for Enzymatic Reactions

1x *KlenTaq* Reaction Buffer	50 mM Tris·HCl pH 9.2 16 mM $(NH_4)_2SO_4$ 2.5 mM $MgCl_2$ 0.1 % Tween 20

1x RQF Reaction Buffer	20 mM Tris·HCl pH 7.5 50 mM NaCl 2 mM $MgCl_2$

1.10.3 Buffers for Protein Purification

1x *KlenTaq* Basic Buffer	300 mM NaCl 2.5 mM $MgCl_2$ 10 mM Tris·HCl pH 9.0 0.1 % Triton X-100
1x *KlenTaq* Elution Buffer	100 mM Tris·HCl pH 9.2 5 mM $MgCl_2$ 200 mM Imidazole
1x *KlenTaq* Lysis Buffer	1x *KlenTaq* Basic Buffer 1 mg/ml Lysozyme
1x *KlenTaq* Lysis Buffer II (Crystallization)	50 mM Tris·HCl pH 8.55 10 mM $MgCl_2$ 16 mM NH_4SO_4 0.1 % Thesit 0.1 % Triton X 1 mM PMSF
1x *KlenTaq* Storage Buffer	1x *KlenTaq* Reaction Buffer 50 % Glycerol
1x *KlenTaq* Washing Buffer	1x *KlenTaq* Basic Buffer 20 mM Imidazol
1x *KlenTaq* Ion Exchange Buffer I	20 mM Tris·HCl pH 8.55
1x *KlenTaq* Ion Exchange Buffer II	20 mM Tris·HCl pH 8.55 2 M NaCl
1x *KlenTaq* Gelfiltration Buffer	20 mM Tris·HCl pH 7.5 1 mM EDTA 1 mM β-Mercaptoethanol 150 mM NaCl

1x *Taq* Gelfiltration Buffer	100 mM Tris HCl pH 9.2 5 mM $MgCl_2$ 150 mM NaCl 10 % Glycerol

1.11 Instruments

Agarose Gel System	Fisher Scientific
ÄKTA Purifier (Unicorn 5.20)	GE Healthcare
Autoclave	Systec 3150 ELV
Block heater	Stuart, Grant Instruments
CD spectrometer, J-815	Jasco
Centrifuge (Multifuge 4KR/Biofuge Primo R)	Heraeus
Centrifuge (5810R)	Eppendorf
Centrifuge (MiniSpin)	Eppendorf
Electroporator (GenePulser Xcell)	BIORAD
Freezer -20°C (profi line)	Liebherr
Freezer -80°C (-86 ULT Freezer)	Thermo Scientific Forma
Gel Dryer (Model 583)	BIORAD
Incubator	Memmert
Incubator shaker Innova 4430	New Brunswick Scientific
Magnetic stirrer (MR 3000 D)	Heidolph
Molecular Imager ChemiDoc XRS	BIORAD
Multidrop (Combi)	ThermoScientific
Nanodrop Spectrophotometer (ND-1000)	PEQLAB
Overhead Shaker (Rax 2)	Heidolph
PAGE System (Sequi Gen GT)	BIORAD
PCR-Thermocycler	Biometra
pH Meter (Seven Easy)	Mettler Toledo
Phosphorimager (Molecular Imager FX)	BIORAD
Phosphorimager Screens and Cassettes	Fuji
Photometer (Biophotometer)	Eppendorf
Photometer (Carry 100 Bio)	Varian
Pipettes	Eppendorf
Pipettes Multichannel	Brand, Süd-Laborbedarf
Pipettor	Hirschmann Laborgeräte
Pipetting robot (Microlab STAR)	Hamilton Robotics
Power Pack 3000	BIORAD
Radioactive Shields	Roth
Real-Time PCR Cycler (icycler)	BIORAD
Real-Time PCR Cycler (LightCycler 480)	Roche

RQF-3 (Rapid Quench Flow)	KinTeq Corporation
Scale (PJ3000, PG403S)	Mettler
SDS-PAGE System	BIORAD
Speed-Vac (Concentrator 5301)	Eppendorf
Sterile Hood	HERA
Thermomixer	Eppendorf
Ultracentrifuge (L-60)	Beckmann Coulter
Ultrapure water system (arium 611)	Sartorius
Vortexer (7-2020)	Heidolph
Water Bath	Memmert

1.12 Disposables

Columns (Empty Reservoir 3 mL, 15 mL)	Biotage
Cuvettes (Plastibrand 1.5 mL)	Brand
96-Deepwell Plates	TreffLab
384-Deepwell Plates	ThermoScientific
Electroporation Cuvettes (Gene Pulser)	BIORAD
Falcon (15 mL and 50 mL)	Sarstedt, Peske
Gas Permeable Adhesive Seals	ABgene
Glass Wool	Serva
Gloves	MaiMed medical, VWR
Petri Dishes	Peske
Sealing Foils	ABgene, 4titude
Sephadex G25 Columns	GE Healthcare
Sterile Filter Paper	Milipore
Syringes	Brand, Henke Sass Wolf
Syringes Needles	Braun, Mormject
Syringe Sterile Filters (0.2 µm, 0.45 µm)	Peske
Tips	Axygen
Tips Pipeting Robot	Hamilton Robotics
TLC Plates (silica gel 60 F254)	Merck
PCR Strips (0.2 mL)	ThermoScientific
PCR-Tubes (1.5 and 2 mL)	Sarstedt, Brand
PCR-Tubes (0. 2 mL)	TreffLab
PCR-96-well Plates	ABgene, Thermo Scientific
PCR 384-well Plates	Roche
Sephadex A-25 DEAE Column Material	Sigma
Whatman Chromatography Paper (3mm)	GE Healthcare

2. Methods

2.1 Methods of Molecular Biology

2.1.1 PCR

Reaction mixtures contained 200 or 500 nM of the respective primers, 7.5-10 ng of the respective template/plasmid, 200 µM dNTPs, optional 3 % DMSO and 1 u Phusion DNA polymerase in 1x HF- or GC-buffer in a total volume of 50 µL. After initial denaturation for 30 or 60 s at 98 °C, 25-30 PCR cycles were performed with 10 s at 98 °C, 20-30 s at the respective hybridisation temperature, 15-30 s/kbp for elongation at 72 °C and one final elongation for 5-10 min at 72 °C.

2.1.2 Colony PCR

Colony PCR was employed to identify *E. coli* single colonies carrying a desired cloning product. Thus, single colonies grown on LB agar plates over-night were picked and suspended in 10 µL water. The following PCR was performed with *Taq* DNA polymerase, kindly provided by Dr. Ramon Kranaster. In short, 1 µL of bacterial suspension, 400 nM primers forward and reverse, 200 µM dNTPs and approximately 100 nM *Taq* DNA polymerase in 1x *KlenTaq* Reaction Buffer were mixed in a total volume of 20 µL. Amplification was performed with one initial denaturation step for 6 min at 95 °C, 30 PCR cycles with 30 s at 98 °C, 60 s at the respective hybridisation temperature, 2 min at 72 °C and alternatively with one final elongation step for 5 min at 72 °C. Product formation was analysed on a 0.8 % TBE agarose gel.

2.1.3 Site-directed Mutagenesis

Reactions were carried out according to the QuikChange Site-Directed Mutagenesis Kit protocol from Stratagene. In short, reaction mixtures contained 200 or 500 nM primers forward and reverse, containing the desired mutations, 25-50 ng of the respective template, 200 µM dNTPs, 3 % DMSO and 1 u Phusion DNA polymerase in 1x HF- or GC-buffer in a total volume of 50 µL. After initial denaturation for 60 s at 98 °C, 18 PCR cycles were performed with 10 s at 98 °C, 30 s at the respective hybridisation temperature, 3.5 min at 72 °C (30 s/kbp) and one final elongation for 10 min at 72 °C. Product formation was analysed on a 0.8 % TBE agarose gel. Next, methylated parental plasmid DNA was digested by adding 10-20 u DpnI and incubation for 3 h at 37 °C. Heat-inactivation of the enzyme was conducted at 80 °C for 20 min. Reactions were purified via a MinElute Reaction Cleanup-Kit (Qiagen) and subsequently transformed into *E.coli* BL21 (DE3) or XL10 Gold.

2.1.4 Restriction Digest of double-stranded DNA

Plasmid DNA and DNA PCR products were digested with the respective restriction enzymes according to the manufacturer's protocol in the supplied buffers. Digestion with two different enzymes was performed subsequently in separate reactions after purification with the MinElute Reaction Cleanup-Kit (Qiagen) in between. 3 µg of plasmid DNA were digested with 2 µL (total volume 50 µL) of the respective enzyme (*Pae* I 10 u/µL; *HindIII* 20 u/µL) for 3 h at 37 °C in total. 1 out of 2 µL was added after 1.5 h incubation time. Heat-inactivation of the enzyme was performed at 65 °C or 80 °C for 20 min. Afterwards plasmid DNA was directly dephosphorylated and purified by preparative agarose gel electrophoresis using 0.8 % TAE agarose gels. Digestion of various amounts of PCR products followed the same procedure with the exception that the DNA was not dephosphorylated and solely purified after digestion via the MinElute Reaction Cleanup-Kit (Qiagen).

2.1.5 Dephosphorylation of double-stranded DNA

For 5'-dephosphorylation of digested plasmid DNA, 2 µL of Antarctic or Shrimp Alkaline Phosphatase and the respective buffer were added directly or alternatively after purification with the MinElute Reaction Cleanup-Kit (Qiagen) to the digested plasmid DNA (total volume 50 µL). After incubation for 2-3 h at 37 °C, the enzyme was heat-inactivated for 20 min at 80 °C. Afterwards, plasmid DNA was purified by preparative agarose gel electrophoresis using 0.8 % TAE agarose.

2.1.6 Ligation of DNA

For ligation of double-stranded DNA, 25 or 50 ng of pre-cut, dephosphorylated plasmid DNA were mixed with a 3- or 5- fold excess of DNA insert and 1 µL of the respective T4 DNA ligase in the supplied buffer. Ligation was either performed for 30 min at room temperature employing the Rapid Ligation Kit (Fermentas) or for 16 h at 16 °C using T4 DNA ligase from Roche. Insert DNA was omitted in the negative controls to test for religation of the vector. All reactions were purified via the MinElute Reaction Cleanup-Kit (Qiagen) preceding transformation into *E. coli* cells.

2.1.7 Isolation of Plasmid DNA

Plasmid DNA was isolated from 4 mL *E. coli* liquid cultures (chapter **VI 2.3.4**) grown over-night. Cells were harvested via centrifugation at 4 °C for 15 min and DNA isolated employing a plasmid Miniprep Kit. For isolation of higher amounts of DNA a plasmid Midiprep Kit from Qiagen was used according to the manufacturer's instructions.

2.1.8 DNA Sequencing

Plasmid DNA was isolated and sent to GATC Biotech, Germany for sequencing. DNA concentrations varied in a range from 30-100 ng/µL. DNA sequences were analyzed using the ChromasLite software and SDSC Biology Workbench.

2.1.9 Analytical Agarose Gel Electrophoresis

DNA containing samples were mixed with 6x Agarose Loading Dye and separated on a 0.8 or 2.5 % agarose gel in 0.5x TBE buffer by applying 100-150 Volt. Agarose gels were selected according to the fragments size. After staining with ethidium bromide in 0.5x TBE buffer for 15-30 min and destaining in 0.5x TBE buffer for 15 min, the DNA was visualized by UV light using a Molecular Imager ChemiDoc XRS system (BIORAD).

2.1.10 Preparative Agarose Gel Electrophoresis

To purify DNA fragments, samples were mixed with 6x Agarose Loading Dye and separated on a 0.8 or 2.5 % agarose gel in 1x TAE buffer by applying 100 Volt. Agarose gels were selected according to the fragments size. After staining with ethidium bromide in 1x TAE buffer for 15-30 min and destaining in 1x TAE buffer for 15 min, the DNA was visualized by preparative UV light and excised with a scalpel. DNA was isolated from the agarose gel by employing a QIAquick Gel Extraction Kit (Qiagen).

2.1.11 Analytical Denaturing PAGE

Analytical denaturing PAGE was employed to separate and analyze DNA oligonucleotides. First, a 12 % PAGE-gel was prepared by mixing 48 mL 1x Denaturing PAGE Gel Solution I, 42 mL 1x Denaturing PAGE Gel Solution II, and 10 mL 1x Denaturing PAGE Gel Solution III with 800 µL 10 % APS and 40 µL TEMED. The solution was then applied to the gel chamber. Final gel thickness was 0.4 mm. DNA samples were supplied with 1x Denaturing PAGE Loading Dye (stop solution) in a 2.25 fold excess and 1.5 µL (per well) of the sample was loaded onto the gel. DNA fragments were then separated in 1x TBE buffer by applying 100 W and 3000 V at up to 50 °C. Subsequently, the gel was transferred to whatman paper, dried *in vacuo* at 80 °C for at least 45 min and exposed to a phosphor imager screen over-night. Readout was facilitated using the Molecular Imager FX (BIORAD).

2.1.12 Preparative Denaturing PAGE

Preparative denaturing PAGE was employed to purify DNA oligonucleotides. PAGE-gels were prepared (8-12 %) according to the DNA fragment's size. First, respective amounts of 1x

Denaturing PAGE Gel Solution I to III were mixed with 1800 µL 10 % APS and 90 µL TEMED to a total volume of 250 mL. The solution was then applied to the gel chamber. Final gel thickness was 1.5 mm. DNA samples were supplied with 1x Denaturing PAGE Loading Dye in a 2.25 fold excess and the total volume loaded onto the gel. DNA fragments were then separated in 1x TBE buffer by applying 100 W and 800 V at up to 50 °C. Subsequently, the desired DNA fragments were visualized by UV light (254 nm) using TLC plates coated with a fluorescence indicator and excised employing a scalpel. The gel was then shredded with a syringe, water was added and the suspension incubated over-night at 50 °C. Finally, the DNA-containing solution was separated from the remaining gel pieces by filtrating them through a syringe filled with silanised glass-fibres wool. Oligonucleotides were then ethanol precipitated, dissolved in ddH_2O and concentrations determined by absorption measurements at 260 nm.

2.2 Oligonucleotide Based Methods

2.2.1 5'-Phosphorylation of Oligonucleotides Using [γ-^{32}P]-ATP

DNA or RNA oligonucleotides were radioactively labelled using [γ-^{32}P]-ATP. Reaction mixtures (total volume 50 µL) contained the respective oligonucleotide (400 nM), 400 nCi/µL [γ-^{32}P]-ATP and 0.4 U/µL T4 polynucleotide kinase (PNK) in the supplied 1x PNK reaction buffer A (forward reaction) and were incubated for 60 min at 37 °C. Reaction termination occurred at 95 °C for 2 min. Subsequent purification of the labelled oligonucleotide was achieved by gel filtration (Sephadex G25). For primer extension reactions, 20 µL of unlabelled primer were added to obtain a final primer concentration of 3 µM.
For kinetic experiments reaction mixtures (50 µL) contained the respective oligonucleotide (10 µM), 1 µCi/µL [γ-^{32}P]-ATP (5 µL) and 1 U/µL PNK (5 µL) in the supplied 1x PNK reaction buffer A. After the reaction, 20 µL of water was added and final DNA primer concentration determined via TLC and subsequent phosphoimaging.

2.2.2 Ethanol Precipitation

For DNA precipitation, samples were mixed with 1/10 volume of 3 M NaOAc pH 5.2 and 2.5 volumes of ice-cold 100 % ethanol. Mixtures were incubated at -20 °C over-night and afterwards, the solutions were centrifuged for 30 min at 4 °C (20.000 × g). The supernatant was discarded and the resulting pellet washed with 70 % ice-cold ethanol. After centrifugation at 4 °C for 30 min (20.000 × g) the supernatant was again discarded and the washing/ centrifugation step repeated. The resulting DNA pellet was dried *in vacuo* or at room temperature, resolved in ddH_2O and stored at -20 °C.

2.2.3 DNA or RNA Concentration Determination

DNA or RNA concentration was determined by employing the Nanodrop ND1000. 1.5 or 2 µL were loaded onto the pedestals of the instrument and the absorption of the DNA or RNA sample was measured at 260 nm. The online tool *Oligo Calc: Oligonucleotide Properties Calculator* facilitated the calculation of the approximate DNA concentration using the molar extinction coefficient. Plasmid DNA concentration determination was based solely on an internal estimation provided by the Nanodrop system. As a reference, either water or the respective buffer was employed.

2.3 Microbiological Methods

2.3.1 Preparation of Electrocompetent *E. coli* Cells

500 mL pre-warmed LB media without antibiotic was inoculated with 2 mL of an over-night *E. coli* cell culture in LB media. Cells were grown to an OD_{600} of 0.5-0.6 at 25 °C with 200 rpm shaking. Afterwards, cells were incubated on ice for 15 min and then harvested by centrifugation (4 °C, 30 min, 4000 x g). All further steps were performed on ice. Supernatant was discarded and the pellet resuspended in 500 mL sterile, ice-cold water. After centrifugation at 4000 x g for 30 min at 4 °C, the supernatant was discarded and the washing step repeated once. Cells were then resuspended in the remaining water and transferred to a 50 mL falcon tube. After centrifugation for 10 min at 4000 x g at 4 °C, the supernatant was discarded and the pellet resuspended in 25 mL 10 % ice-cold glycerol. After a final harvesting step, the supernatant was discarded and the cells resuspended in 1 mL 10 % ice-cold glycerol. 80 µL aliquots were generated and shock frozen in liquid nitrogen for long term storage at - 80 °C.

2.3.2 Transformation in Electrocompetent or Chemically Competent *E. coli* Cells

2-3 µL of purified ligation mixture were mixed with 50-100 µL electrocompetent *E. coli* cells in a pre-cooled Gene Pulser electroporation cuvette (1 mm) and incubated on ice for several minutes. After pulsing (1.8 kV, 25 µF and 200 Ω, Gene Pulser Xcell) the cells were mixed with 1 mL pre-warmed SOC medium, transferred to a 2 mL reaction tube and incubated at 37 °C for 45-60 min shaking. After incubation, 50-100 µL cell suspension was plated on LB agar.

2 µL of purified ligation mixture were mixed with 50 µL chemically competent *E. coli* cells (kindly provided by Dr. Marina Rubini) and incubated on ice for 5 min. After incubation for 30 min at 37 °C, the cells were mixed with 1 mL pre-warmed SOC medium and incubated at 37 °C for 45-60 min shaking. After incubation, the cells were centrifuged (MiniSpin, 3.3 rpm,

10 min) and 900 µL of the supernatant were discarded. The pellet was resuspended in the remaining 100 µL and plated on LB agar.

2.3.3 Plate Culture

After the transformation, up to 100 µL cells suspension were transferred to LB agar plates supplemented with 100 mg/L carbenicillin and grown over night at 37 °C. Single colonies were then used to inoculate a respective liquid culture.

2.3.4 Liquid Culture

Single colonies or glycerol stocks were used to inoculate LB media supplemented with 100 mg/L carbenicillin. Cultures were grown over-night or at least for 6 h at 37 °C and 220 rpm shaking.

2.3.5 *E. coli* Glycerol Stock Preparation

For storage of *E. coli* cells, liquid cultures were grown over-night, supplemented with glycerol in a 1:1 ratio and shock frozen in liquid nitrogen. Storage occurred at -80 °C.
Glycerol stocks of *E. coli* BL21 (DE3) cells containing the library of recombined *KlenTaq* variants were prepared in 384-deepwell plates. In detail, single colonies were picked and grown in 150 µL LB media per well supplemented with 100 mg/L carbenicillin at 30 °C shaking, over-night. Afterwards, 100 µL 50 % glycerol per well were added to a final concentration of 20 % (v/v). After shaking for 30 min, cells were stored at -80 °C.

2.4 Biochemical Methods

2.4.1 Gene Expression and Protein Purification

2.4.1.1. Expression and Purification of KTq wild-type and variants

PGDR11 expression vector harbouring the respective *KlenTaq* gene was transformed into *E. coli* BL21 (DE3). For expression, a 5 mL over-night culture of these cells was used to inoculate 500 mL LB media supplemented with 100 mg/L carbenicillin. Cells were grown at 37 °C at 200 rpm shaking to an OD of 0.6-0.8. Expression was induced by addition of IPTG to a final concentration of 1 mM. After 4 h expression, cells were harvested by centrifugation (4 °C, 4000 x g, 30 min) and stored over-night at -20 °C. Cells were lysed in 20 mL 1x *KlenTaq* Lysis Buffer supplemented with 1 mM Benzamidine at 37 °C for 15 min, followed by heat denaturation of host proteins at 75 °C for 45 min. Ultra-centrifugation (Ultracentrifuge, L-60) at 40 000 rpm for 60 min at 4 °C was used to remove bacterial cell debris.

Supernatants were incubated with 50 % Ni-IDA sepharose slurry in the presence of 5 mM imidazole for 1.5 h at 4 °C gently shaking. After washing with 30 mL 1x *KlenTaq* Washing Buffer containing 20 mM imidazol, elution was carried out using 6 mL 1x *KlenTaq* Elution Buffer. The imidazole was removed using VivaSpin columns and 1x *KlenTaq* Elution Buffer without imidazole. For storage, glycerol was added to a final concentration of 50 %, $(NH_4)_2SO_4$ to a final concentration of 16 mM, Tween20 to a final concentration of 0.1 %, Tris·HCl pH 9.2 to a final concentration of 50 mM and $MgCl_2$ to a final concentration of 2.5 mM (1x *KlenTaq* Storage Buffer). Purified enzymes were stored at -20 °C. Concentration was determined by Bradford Assay or SDS-PAGE analysis.

2.4.1.2. Expression and Purification of KTq wild-type and RT-KTq 2 for crystallization

PGDR11 expression vector harbouring the respective *KlenTaq* gene (wild-type, RT-KTq 2 (codon optimized) without His-tag in pGDR11*, chapter **VII 2.3.4**.) was transformed into *E. coli* BL21 (DE3). Based on this vector, the enzyme was expressed without an N-terminal His-Tag.

For expression, 10 mL over-night culture of these cells was used to inoculate 1000 mL LB media supplemented with 100 mg/L carbenicillin. Cells were grown at 37 °C at 200 rpm shaking to an OD of 0.6-0.8. Expression was induced by addition of IPTG to a final concentration of 1 mM. After 4.5 h expression, cells were harvested by centrifugation (4 °C, 4000 x g, 30 min) and stored over-night at -20 °C. Cells were lysed in 20 mL 1x *KlenTaq* Lysis Buffer II containing 0.7 mg/mL lysozyme at room temperature for 60 min, followed by heat denaturation of host proteins at 75 °C for 45 min. Ultra-centrifugation (Ultracentrifuge, L-60) at 40 000 rpm for 60 min at 4 °C was used to remove bacterial cell debris.

Afterwards, supernatants were treated stepwise with 5 % polyethylenimine (PEI) to remove bacterial DNA in the supernatant. After PEI was added, the suspension was shaken for 30 min at 4 °C. Then, centrifugation (4000 x g, 30 min, 4 °C) was carried out before adding additional PEI. These steps were repeated three times. The supernatant was then filtrated using a syringe sterile filter (0.45 µm) and afterwards loaded onto a Q sepharose column (kindly provided by Karin Betz). The protein was purified via anion exchange chromatography on an ÄKTApurifier. The column was equilibrated with 2 column volumes (CV) of 1x *KlenTaq* Ion Exchange Buffer I, followed by 2 CV of 1x *KlenTaq* Ion Exchange Buffer II and again 2 CV of buffer I. Elution was carried out at 4 °C by applying a gradient of 0-1 M NaCl (0-50 % buffer II, pH 8.55) with a flow rate of 1 mL/min.

In detail:

150 mL 100 % buffer I (Washing step)

250 mL 0-8 % buffer II (Elution)

100 mL 8-20 % buffer II (Elution)

100 mL 20-50 % buffer II (Elution)
30 mL 100 % buffer II (Elution)

For column purification 2 CV 100 % buffer II were applied. Fractions were controlled by SDS-PAGE and pooled. The buffer was exchanged to 1x *KlenTaq* Gel Filtration Buffer and the protein concentrated using VivaSpin columns.
Size exclusion chromatography was carried out using a superdex 75 column and an ÄKTApurifier. For equilibration, 2 CV of 1x *KlenTaq* Gel Filtration Buffer were used. The protein was eluted at 4 °C with a flow rate of 0.5 mL/min in 1 CV (peak between 55-65 mL). Fractions were analysed by SDS-PAGE and pooled. Protein was stored at 4 °C. Expression and purification yielded 26 mg pure protein of KTq wild-type and 10 mg of RT-KTq 2 from 1 L expression culture.

2.4.1.3. Expression and Purification of Taq wild-type and variants

PGDR11 expression vector harbouring the respective *Taq* gene was transformed into *E. coli* BL21 (DE3). For expression, an over-night culture (5 mL each) of these cells was used to inoculate 4 x 500 mL LB media supplemented with 100 mg/L carbenicillin. Cells were grown at 37 °C at 200 rpm shaking to an OD of 0.6-0.8. Expression was induced by addition of IPTG to a final concentration of 0.5 mM. After 4.5 h expression, cells were harvested by centrifugation (4 °C, 4000 rpm, 30 min) and stored over-night at -20 °C. Cells were lysed in 20 mL 1x *KlenTaq* Lysis Buffer supplemented with 1 mM benzamidine at 37 °C for 15 min, followed by heat denaturation of host proteins at 75 °C for 45 min. Ultra-centrifugation (Ultracentrifuge, L-60) at 40 000 rpm for 60 min at 4 °C was used to remove bacterial cell debris.
Supernatants were incubated with 50 % Ni-IDA sepharose slurry in the presence of 5 mM imidazole for 1.5 h at 4 °C gently shaking. After washing with 30 mL 1x *KlenTaq* Washing Buffer containing 20 mM imidazol, elution was carried out using 6 mL 1x *KlenTaq* Elution Buffer. The imidazole was removed using VivaSpin columns and 1x *KlenTaq* Elution Buffer without imidazole. Size exclusion chromatography was carried out using a superdex 75 column and an ÄKTApurifier. For equilibration, 2 CV of 1x *Taq* Gelfiltration Buffer were used. The protein was eluted at 4 °C with a flow rate of 0.5 mL/min in 1 CV (peak between 48-68 mL). Protein was sampled in 1 mL fractions. Fractions were checked by SDS-PAGE and pooled. The 1x *Taq* Gelfiltration Buffer was exchanged using VivaSpin columns and 1x *KlenTaq* elution buffer without imidazole. For storage, glycerol was added to a final concentration of 50 %, $(NH_4)_2SO_4$ to a final concentration of 16 mM, Tween20 to a final concentration of 0.1 %, Tris·HCl pH 9.2 to a final concentration of 50 mM and $MgCl_2$ to a final concentration of 2.5 mM (1x *KlenTaq* Storage Buffer). Purified enzymes were stored at -20 °C. Concentration was determined by Bradford Assay or SDS-PAGE.

2.4.2 SDS-PAGE

For SDS-PAGE analysis of proteins, a 4 % stacking and a 12 % separating gel were used. Protein samples were mixed with 6x SDS-PAGE Loading Dye and heat-denatured at 95 °C for 5 min. Afterwards, 5-20 µL of the protein sample were loaded on the gel and the proteins separated by applying 25-30 mA in 1x SDS-PAGE Running Buffer. Proteins were stained for 30 min using methanol staining solution and destained with coomassie methanol destaining solution for at least 30 min. Alternatively, proteins were stained with coomassie colloidal over-night and destained with coomassie colloidal destaining solution for at least 30 min. As a marker, Unstained Protein Ladder (Fermentas) was used.

2.4.3 Protein Concentration Determination

The Bradford-Assay was employed to determine the protein concentration. Roti-Quant was diluted with water in a 1:5 ratio and immediately filtrated. 980 µL suspension was then added either to 20 µL of the respective *KlenTaq* DNA polymerase stored in 1x *KlenTaq* Storage Buffer or to 20 µL of 0, 0.1, 0.2, 0.3, 0.4, 0.6 and 0.8 mg/mL BSA, respectively. The BSA dilution series was used as standard and was prepared in 1x *KlenTaq* Storage Buffer. After incubation for 5 min, the absorption was measured at 595 nm using the Cary 100 Bio photometer. Protein concentrations were determined by use of the BSA standard curve and verified via SDS-PAGE analysis.

2.5 Methods and Assays for Chapter II

2.5.1 Site-directed Mutagenesis of Y671A, Y671F and Y671W

Genes encoding for *KlenTaq* mutants Y671A and Y671W were constructed by Dr. Ramon Kranaster (Dissertation, 2010, Universität Konstanz).

To introduce the Y671F mutation in *KlenTaq* wild-type gene encoded on the pGDR11 vector,[184] whole plasmid PCR was performed according to the QuickChange® Site-Directed Mutagenesis Kit Protocol by Stratagene, described in chapter **VI 2.1.3**. For PCR, Phusion DNA polymerase and supplied GC-buffer were used in combination with the following primers P-Y671F-2 fwd and P-Y671F-2 rev. Transformation was conducted in *E. coli* XL10. Sequencing of the open-reading frame was performed by GATC Biotech (Germany) using the following primers: pQE-FP, pQE-RP and pQE-KTQ-mid. Purification of the mutant enzymes followed the protocol described in chapter **VI 2.4.1.1**. Purification of *KlenTaq* wild-type, employed for studies described in **chapter II**, was performed by Dr. Christian Glöckner (Dissertation, 2008, Universität Konstanz) in preliminary work.

2.5.2 Primer Extension Experiments opposite F with *KlenTaq* Wild-type, Y671A, Y671F and Y671W

Reaction mixtures (20 µL) contained 100 nM primer P-F23, 130 nM F-containing template F33XA, 100 µM dNTPs and 500 nM of the respective *KlenTaq* DNA polymerase in 1x RQF Reaction Buffer. Reaction mixtures were incubated at 37 °C. Incubation times are provided in the respective figure legends. Primer was labelled using [γ-^{32}P]-ATP described in chapter **VI 2.2.1**. Reactions were stopped by addition of 45 µL 1x Denaturing PAGE Loading Dye, denatured at 95 °C for 5 min and analysed by 12 % denaturing PAGE. Visualization was performed by phosphoimaging. Blunt-end extension experiments were conducted as described above using DNA primer (P-BluntEnd) and template (T-BluntEnd) that form a blunt-end only at one terminus.

2.5.3 Primer Extension Experiments with *KlenTaq* F_3Y

Incorporation opposite F with *KlenTaq* wild-type and F_3Y and all four nucleotides, respectively: Extension experiments were conducted as described in chapter **VI 2.5.2** using 100 nM primer P-F23, 130 nM F-containing template F33XA, 100 µM of the respective dNTP and 25 nM of *KlenTaq* DNA polymerase in 1x RQF Reaction Buffer. Incubation times are provided in the respective figure legends.

Incorporation opposite F or dT with a polymerase dilution series: Experiments were performed as described in chapter **VI 2.5.2** using 100 nM primer P-F23, 130 nM F-containing template F33XA or the respective natural DNA template F33-7T, 100 µM of the respective dNTP as well as the respective polymerase concentration (50, 25, 10, 5, 2.5, 1 nM) in 1x RQF Reaction Buffer. Reaction mixtures were incubated at 37 °C for 5 s in case of the natural template and 30 min for the F-containing template.

Incorporation opposite F in a time-dependent manner: Experiments were performed as described in chapter **VI 2.5.2** using 100 nM primer P-F23, 130 nM F-containing template F33XA, 100 µM of the respective dNTP and 1 µM polymerase of the respective *KlenTaq* DNA polymerase in 1x RQF Reaction Buffer. Incubation times are provided in the figure legends.

2.5.4 Primer Extension Experiments with Natural Abasic Site

Reaction conditions were employed as described in chapter **VI 2.5.2**. 20 µL of the reactions contained 100 nM primer P-F23, 130 nM F-containing template F33XA or dU-containing template F33dU, respectively, 100 µM dNTPs in 1x RQF Reaction Buffer, 0.16 µL UDG ($5*10^{-4}$ u/µL) and 500 nM of the respective *KlenTaq* DNA polymerase. Reaction mixtures were annealed at 95 °C for 2 min without UDG, *KlenTaq* DNA polymerase and dNTPs present. After

cooling to 0 °C, UDG was added and the reaction mixture incubated for 18 h at 37 °C. Next, *KlenTaq* DNA polymerase was added on ice and the reaction started by addition of dNTPs at 37 °C (reaction temperature). Incubation times are provided in the respective figure legends. Primer was labelled using [γ^{32}P]-ATP as described in chapter **VI 2.2.1**. Reactions were stopped by addition of 45 µL 1x Denaturing PAGE Loading Dye, denatured at 95 °C for 5 min and analysed by 12 % denaturing PAGE. Visualization was performed by phosphoimaging.
To check for conversion of dU to a natural abasic site, both templates were labelled with [γ-^{32}P]-ATP and submitted to the same conditions described above. After incubation with UDG for 18 h at 37 °C, 1x Denaturing PAGE Loading Dye and NaOH were added. Heat-denaturation for 10 min at 95 °C in presence of NaOH should cleave a natural abasic site, but not the abasic site analogue F. Conversion was analysed by 12 % denaturing PAGE. Visualization was performed by phosphoimaging.

2.5.5 Pre-steady State Kinetic Analysis

The rate of single turnover in single nucleotide incorporation experiments was determined by rapid quench flow kinetics using a chemical quench flow apparatus (RQF-3, KinTek Corp., University Park, PA) as described earlier (Dr. Christian Glöckner, Dissertation, 2008, Universität Konstanz). In short, 15 µL of an enzyme/primer/template mix and 15 µL dNTP solutions were filled manually in two sample loops. A three-way valve was used to put the loaded sample loops in line with syringes containing the reaction buffer (1x RQF Reaction Buffer). A drive motor was used to apply pressure to the syringes which pushed the reaction buffer into the sample loops. Thus, both solutions were mixed in a first mixing chamber and passed into a reaction delay line where the reaction occured. Reaction times from 2 ms to 100 ms were varied by the length (volume) of the reaction delay line and the flow rate through the reaction delay line. For longer reaction times, the drive motor paused after the first push to mix the reactants together and then after a specific time period performed a second push which added the quenching solution to the reaction in a second mixing chamber. After termination of the reaction by adding the quenching solution, the sample was collected and product formation was determined by denaturing PAGE. For reaction times longer than 5 s, a manual quench was performed.
In detail, 15 µL of radioactively labelled primer P-F23 (200 nM), annealed to the F-containing template F33XA (260 nM) or natural DNA template F33-7T (260 nM), and *KlenTaq* DNA polymerase (2 µM) in 1x RQF Reaction Buffer were rapidly mixed with 15 µL of a dNTP solution in buffer at 37 °C. Quenching was achieved by adding 0.3 M EDTA solution at defined time intervals before mixing with 1x Denaturing PAGE Loading Dye. Due to the fact that the volume of the quenched samples collected from the quench flow apparatus after different time points varied, different volumes of 0.3 M EDTA were added to adapt all samples to one total volume. Next, quenched samples were analysed by 12 % denaturing PAGE. Visualization

was performed by phosphoimaging. For kinetic analysis, experimental data were fit by nonlinear regression using the program GraphPad Prism 4. The data were fit to a single exponential equation: [conversion] = A * (1 - exp(-k_{obs} * t)). The observed catalytic rates (k_{obs}) were then plotted against the dNTP concentrations (x) used and the data were fitted to a hyperbolic equation (k_{obs} = (K_{pol} * x)/(k_d + x)) to determine the K_d of the incoming nucleotide. The incorporation efficiency is given by k_{pol}/K_d. Kinetic experiments for blunt-end measurements were performed under the same conditions using DNA primer (P-BluntEnd) and template (T-BluntEnd) as describe above.

2.5.6 MALDI/MS-MS Analysis of *KlenTaq* F_3Y

The expressed protein *KlenTaq* F_3Y was in-gel digested with trypsin and the resulting fragments were analyzed by MALDI-MS/MS at the 'Functional Genomics Center Zurich'.

2.6 Methods and Assays for Chapter III

2.6.1 Generation and Characterization of *KlenTaq* Variants with Increased Reverse Transcriptase Activity

2.6.1.1. Cloning of KTq M1 gene in pGDR11

KTq M1 gene encoded on the pASK-IBA 37+ vector (see chapter **VII 2.3.5** for sequence) was amplified in PCR using the following primers P-KTQ-*SphI* and P-KTQ-*HindIII* according to chapter **VI 2.1.1**. In short, reaction mixtures (50 µL) contained 200 nM of the respective primers, 7.5 ng of the respective plasmid (*KlenTaq* M1, pASK-IBA 37+), 200 µM dNTPs, 3 % DMSO and 1 u Phusion DNA polymerase in 1x HF-buffer. Amplified DNA was gel-extracted, digested with *SphI* and *HindIII* (see chapter **VI 2.1.4**) and ligated in pGDR11[184] vector (see chapter **VI 2.1.6**). The ligation reaction was transformed into chemically competent *E. coli* XL10 cells. Plasmid DNA was then isolated and afterwards transformed into *E. coli* BL21 (DE3) cells. Sequencing of the open-reading frame was performed by GATC Biotech (Germany) using the following primers: pQE-FP, pQE-RP and pQE-KTQ-mid (see chapter **VII 2.3.3** for sequence).

2.6.1.2. Site-directed mutagenesis of KTq M747K and M1/M747K

To introduce the M747K mutation in *KlenTaq* wild-type or the M1 gene encoded on the pGDR11 vector,[184] whole plasmid PCR was performed according to the QuickChange® Site-Directed Mutagenesis Kit Protocol by Stratagene, described in chapter **VI 2.1.3**. For PCR, Phusion DNA polymerase and supplied GC-buffer was used in combination with the following primers forward (P-M747K fwd) and reverse (P-M747K rev). Transformation was conducted

in *E. coli* BL21 (DE3) cells. Sequencing of the open-reading frame was performed by GATC Biotech (Germany) using the following primers: pQE-FP, pQE-RP and pQE-KTQ-mid (see chapter **VII 2.3.2** for sequence. Purification of the mutant enzymes followed the protocol described in chapter **VI 2.4.1.1**.

2.6.1.3. Library Generation via DNA shuffling[130, 139, 181, 182]

Parental DNA (pGDR11[184] *KlenTaq* genes encoding wt, M1 and M747K) was amplified in a nested PCR. Reaction mixtures (100 µL) contained 0,03 u/µL Phusion DNA polymerase (Thermo Scientific), P-DNA Shuffling primer fwd and rev (200 nM each), 200 µM of each dNTP, 3 % (v/v) DMSO and 12 fmol of the respective *KlenTaq* gene. After initial denaturation for 60 s at 98 °C, 25 PCR cycles were performed with 10 s at 98 °C, 30 s at 70 °C, 60 s at 72 °C and one final elongation for 10 min at 72 °C. As high amounts of template were needed for the DNase digestion step, PCR products were gel extracted and reused as template in PCR under the same conditions (43.5 fmol template in a 100 µL reaction volume).

3 µg amplified DNA (1 µg of each amplified *KlenTaq* gene) were digested in a total reaction volume of 80 µL with 0.33 u DNase I (Fermentas) in 10 mM Tris HCl pH 7.5, 0.1 mM $CaCl_2$ in the presence of 2.5 mM $MgCl_2$ for 1 min at 15 °C. The reaction was terminated by addition of EDTA (9 µL, 25 mM) and subsequent incubation at 65 °C for 10 min. DNA fragments in the size range of 50-200 bp were obtained.

Without further purification, 1, 5, 8 or 60 µL of DNA fragments were utilized in an Assembly PCR (50 µL total volume in case of 1, 5 and 8 µL; 100 µL total volume in case of 60 µL) using no primers, 200 µM of each dNTP, 3 % (v/v) DMSO and 1 u Phusion Polymerase in 1x HF-buffer. The fragments were reassembled using 60 cycles and an annealing temperature of 45 °C to promote recombination. After initial denaturation for 60 s at 98 °C, 60 PCR cycles were performed with 10 s at 98 °C, 30 s at 45 °C, 60 s at 72 °C and one final elongation for 10 min at 72 °C. The reaction mixture containing 5 µL DNA fragments was used for the following reactions.

A final PCR was performed with a 1:2 dilution of the Assembly PCR. 4 µL of this dilution was used as template in PCR (200 µL) using 200 nM primers with restriction sites for cleavage with *SphI* (P-KTQ-*SphI*) and *HindIII* (P-KTQ-*HindIII*), 200 µM of each dNTP, 3 % (v/v) DMSO and 0.03 u/µL Phusion DNA polymerase in 1x HF-buffer. After initial denaturation for 60 s at 98 °C, 20 PCR cycles were performed with 10 s at 98 °C, 30 s at 72 °C, 60 s at 72 °C and one final elongation for 10 min at 72 °C. The amplified products containing recombined full-length *KlenTaq* genes were gel extracted, digested with restriction enzymes *SphI* and *HindIII* (see chapter **VI 2.1.4.**) and cloned into pGDR11 vector. The ligation reaction was transformed into *E. coli* BL21 (DE3) cells and colonies were picked randomly to generate a library containing 1,570 shuffled *KlenTaq* variants. Glycerol stocks containing the library were prepared as described in chapter **VI 2.3.5.**

2.6.1.4. Library Expression and Screening

KlenTaq variants of the generated library were expressed in 96-well plates. Cells were grown in 1 mL LB-media (per well) supplemented with 100 mg/mL carbenicillin by inoculation with 5 µL of the 384-deepwell library stock (chapter **VI 2.3.5**.) using a pipetting robot. Cells were grown to an OD_{600} of 0.6 - 0.8 and protein expression induced by addition of IPTG (1 mM final concentration). After protein expression at 37 °C for 4.5 h, cells were harvested by centrifugation at 4000 x g, for 45 min at 4 °C. Pellets were lysed using lysozyme at a concentration of 0.1 mg/mL and digested with DNase I (0.01 mg/mL) at 37 °C for 15 min in 1x *KlenTaq* Reaction Buffer (800 µL per well). After heat-denaturation at 75 °C for 45 min and centrifugation (4000 x g, 45 min at 4 °C) to remove bacterial cell debris, the lysates were directly used in an activity screen performed in 384-well plates using real-time PCR. Reaction mixtures (20 µL) contained 60 pM template (F90A), 250 µM of each dNTP, 750 nM of each primer (P-F20, P-F20$^-$), 0.6x SYBRGreen I and 10 µL of the respective lysate in 1x *KlenTaq* Reaction Buffer. Initial denaturation at 95 °C for 1 min was conducted, 30 PCR cycles were performed with 10 s at 95 °C, 20 s at 55 °C and 30 s at 72 °C. *KlenTaq* variants were further screened for reverse transcriptase activity in real-time RT-PCR, carried out in 96-well plates using MS2 RNA based on a method published earlier by Sauter *et al*.[148] Visualization was based on SYBRGreen I binding to double-stranded DNA. Reaction mixtures (20 µL) contained 50 pg/µL MS2 RNA (Roche) as template, 200 µM of each dNTP, 100 nM of each primer (P-MS2-5'; P-MS2-3'), 0.6x SYBRGreen I and 5 µL of the respective lysate in 1x *KlenTaq* Reaction Buffer. First, reverse transcription was conducted using an initial denaturation step of 30 s at 95 °C, an annealing step at 55 °C for 35 s and elongation for 15 min at 72 °C. Additionally, 50 PCR cycles were performed with 30 s at 95 °C, 35 s at 55 °C and 40 s at 72 °C.

KTq mutants with a ct-value lower or in the range of the ct value of the parental mutants, M1 and M747K, were selected for a second round of screening. The second round was performed under the same conditions as the first run in order to confirm the results from the first run. However, only mutants with a crossing threshold (ct-) value lower than the parental enzymes were selected in this screening round. These variants were combined on one 96-well plate in order to provide normalised reaction conditions and further screened for reverse transcriptase activity under more challenging conditions, including a reduction of the reverse transcription time (down to 7.5 min) in a first screening and a reduction of RNA concentration (down to 5 pg/µL, 15 min) in a second screening. These experiments were performed twice with lysates resulting from two different expression cultures. Variants exhibiting the best results, namely the lowest ct-value, in all experiments (inital conditions, 7.5 min and 5 pg/µL) were selected and their genes sequenced at GATC Biotech, Germany. Sequencing of the open-reading frame was performed using the following primers: pQE-FP, pQE-RP and pQE-KTQ-mid. Purification of the wild-type and identified variants followed the protocol described in chapter **VI 2.4.1.1**.

2.6.1.5. Primer Extension Reactions with an RNA template

Reaction mixtures (20 µL) contained 150 nM radioactively labelled primer (P-F20), 225 nM RNA template (F30RNA) or the respective DNA template (F33A), 200 µM of each dNTP and 25 nM of the respective *KlenTaq* DNA polymerase in 1x *KlenTaq* Reaction Buffer. Reaction mixtures were incubated at 72 °C and terminated after 5 min by addition of 45 µL 1x Denaturing PAGE Loading Dye. After denaturation at 95 °C for 5 min, reaction mixtures were separated using a 12 % denaturing PAGE gel. Visualization was performed by phosphoimaging.

2.6.1.6. DNA Polymerase Specific Activity Determination[185]

Primer extension reactions were performed at 72 °C (described in chapter **VI 2.6.1.5**). Reaction mixtures (20 µL) contained 150 nM radioactively labelled primer (P-F23), 225 nM RNA template (F30RNA) or the respective DNA template (F33A), 200 µM of each dNTP and various amounts of the respective *KlenTaq* DNA polymerase in 1x *KlenTaq* Reaction Buffer. Reactions with DNA or RNA as template were incubated for 10 min or 30 min, respectively. Polymerase amounts present in reactions with DNA or RNA as template were 2, 1, 0.75, 0.5, 0.25, 0.125 fmol or 400, 300, 200, 150, 100, 50, 30, 20, 15, 10, 5, 3, 2, 1, 0.5, 0.3 fmol, respectively. The observed intensities of each band yielded the conversion of dNTPs in each reaction. DNTP conversion per min was then plotted against the amount of enzyme. The linear range was analysed using linear regression with the slopes yielding the specific activity of the respective enzyme (**VII 1.2**, Figure S2).

2.6.1.7. Real-time RT-PCR

RT-PCR with MS2 Bacteriophage RNA.[148] Reaction mixtures (20 µL) for real-time RT-PCR contained 50 pg/µL MS2 RNA (Roche) or 40 pM MS2DNA, 100 nM of each primer (P-MS2-5'; P-MS2-3'), 200 µM of each dNTP, 0.6x SYBRGreen I and 5 nM of the respective, purified DNA polymerase in 1x *KlenTaq* Reaction Buffer. First, reverse transcription was conducted using an initial denaturation step of 30 s at 95 °C, an annealing step at 55 °C for 35 s and elongation for 7.5 min at 72 °C. After 1 min at 95 °C, 50 PCR cycles were performed with 30 s at 95 °C, 35 s at 55 °C and 40 s at 72 °C. Formation of double-stranded DNA was visualized by SYBRGreen I binding. Correct product formation was confirmed by agarose gel analysis.

RT-PCR with 16S- and 23S - rRNA.[148] Reaction mixtures (20 µL) for real-time RT-PCR contained 40 pg/µL 16S- and 23S- rRNA from *E. coli* (Roche), 100 nM of each primer (P-S16-5'; P-S16-3'), 200 µM of each dNTP, 0.6x SYBRGreen I and 5 nM of the respective, purified DNA polymerase in 1x *KlenTaq* Reaction Buffer. First, reverse transcription was conducted using an initial denaturation step of 30 s at 95 °C, an annealing step at 66 °C for 35 s and elongation for

7.5 min at 72 °C. After 1 min at 95 °C, 40 PCR cycles were performed with 30 s at 95 °C, 35 s at 66 °C and 40 s at 72 °C. Formation of double stranded DNA was detected by binding of SYBRGreen I. Correct product formation was confirmed by agarose gel analysis.

2.6.1.8. RT-PCR: Longer Amplicons

510 bp fragment: RT-PCR reaction mixtures contained 350 pg/µL MS2 RNA (Roche), 200 nM of each primer (P-MS2-5'-ES 490; P-MS2-3'-ES), 250 µM of each dNTP and 80 nM of the respective DNA polymerase in 1x *KlenTaq* Reaction Buffer. First, reverse transcription was conducted using an initial denaturation step of 60 s at 95 °C, an annealing step of 60 s at 65 °C and an elongation step of 30 min at 72 °C. 30 PCR cycles were subsequently performed with 60 s at 95 °C, 60 s at the respective annealing temperature indicated in the provided figure legends and 90 s at 72 °C for elongation. Correct product formation was confirmed by agarose gel analysis.
321 bp fragment: Reactions were conducted as described above with the exception of 400 nM of each primer, 7.5 min reverse transcription and the following primer sequences: (P-MS2-5'-ES 301; P-MS2-3'-ES). Primers used in this experiment exhibit an overhang with cleavage sites for *SphI* and *HindIII*, respectively, as they were also used for cloning in studies omitted in this work.

2.6.1.9. CD-spectra Measurement and Thermal Denaturation[223]

1x KlenTaq Storage Buffer was exchanged via dialysis over night at 4 °C with buffer containing 137 mM NaCl, 2.7 mM KCl, 10.2 mM Ka_2HPO_4, 1.8 mM KH_2PO_4 (pH 7.4) employing Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific). CD spectra measurements were conducted at 20 °C using quartz cuvettes (light path 1 mm) and 250 µL protein sample (2.2 µM). CD spectra were determined from 200 to 250 nm (50 nm/min) with 0.1 nm data intervals and were averaged from 6 scans (CD spectrometer J815).
Thermal denaturation was performed based on CD spectroscopy by following the ellipticity at the two local minima 209 and 220 nm. Data collection was carried out at every 0.1 °C with a temperature slope of 0.2 min^{-1}. The reaction was irreversible as precipitation was observed under the experimental conditions.

2.6.1.10. Multiplex RT-PCR

RNA extracts of respiratory swab samples were from patients known to be either influenza A or B positive (kindly provided by Labor Dr. Brunner, Konstanz, Germany). All samples were analysed with the commercial RIDA®GENE Flu assay (R-Biopharm AG, Darmstadt, Germany) which detects influenza A incl. H1N1 variant and influenza B strains in different fluorescence channels. Purified RNA was kindly provided by ProlagoBiotec. RNA was extracted using the

QIAamp Viral RNA Mini Kit (Qiagen) and digested with DNase I (ThermoScientific) according to the manufacturer's protocol. Primers for influenza A virus detection (recommended by the WHO in "WHO information for molecular diagnosis of influenza virus in humans", August 2011) target the RNA coding for the matrix protein (P-Influenza A-fwd, P-Influenza A-rev), whereas primers for influenza B virus detection target the RNA of the hemagglutinin gene (P-Influenza B-fwd; P-Influenza B-rev). *Taq*Man probes for influenza A detection consisted of an oligonucleotide with a 5' reporter dye 6-carboxyfluorescein (FAM) and a 3' minor groove binder (MGB) (TaqMan Probe-Influenza A), whereas the *Taq*Man probe for influenza B detection consisted of an oligonucleotide with a 5' reporter dye hexachlorofluorescein (HEX) and a 3' black hole quencher (BHQ-1) (TaqMan Probe-Influenza B) (for sequences see chapter **VII 2.1.1**).

*Taq*Man based multiplex real-time RT-PCR for the simultaneous detection of influenza A and B was performed using the above mentioned primers and *Taq*Man probes. Reaction mixtures (10 µL) contained 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 7.5 mM $MgCl_2$, 500 µM of each dNTP, 0.8 M betaine, 1 µL of the respective template dilution of extracted RNA of influenza A or B positive samples, 50 nM of RT-KTq 2 and 10 nM *Taq* wild-type (combined with aptamer). Primers for influenza A and B detection were added in concentrations of 600 nM and 400 nM, respectively. *Taq*Man probes for influenza A and B detection were applied in a concentration of 100 nM and 400 nM, respectively. To increase reaction specificity an aptamer (TaqMan-Aptamer, for sequence see chapter **VII 2.1.1.**) with the ability to inhibit *Taq* polymerase at low temperatures was added to the reaction.[224] Cycling was performed with 60 s initial denaturation at 95 °C and 50 cycles of two-step cycling with denaturation for 15 s at 95 °C and combined annealing/extension for 60 s at 60 °C.

2.6.2 Crystallization Studies with RT-KTq 2

2.6.2.1. Site-directed Mutagenesis of RT-KTq 2

RT-KTq 2 mutations were introduced in the *KlenTaq* wild-type gene (KTq wild-type (codon optimized) without His-tag in pGDR11*, see chapter **VII 2.3.4**) by employing the QuikChange Multi Site Directed Mutagenesis Kit from Stratagene. Reactions were carried out according to the manufacturer's specifications. In short, reaction mixtures (25 µL) contained 50 ng of each primer, 100 ng of plasmid, 1 µL supplied dNTP mix, and 1 µL supplied multi-enzyme blend. Cycling was performed with 60 s initial denaturation at 95 °C and 30 cycles of a denaturation step at 95 °C for 60 s, an annealing step for 60 s at 55 °C and an extension step for 13 min at 65 °C. PGDR11 expression vector harbouring the respective *KlenTaq* gene (RT-KTq 2 (codon optimized) without His-tag in pGDR11*, chapter **VII 2.3.4**) was transformed into *E. coli* BL21 (DE3). The following primers were used to introduce the mutations: P-NB_SD_L459M, P-

NB_SD_S515R, P-NB_SD_I638F, P-NB_SD_M747K. Expression and purification of the enzymes followed the protocol described in chapter **VI 2.4.1.2**.

2.6.2.2. Crystallization Trials with KTq wild-type in Complex with a DNA/RNA Hybrid Duplex

Crystallization solutions with KTq wild-type in complex with a DNA/RNA hybrid duplex contained 6.6 mg/mL protein in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol were mixed in a 1:2 or 1:3 ration with the DNA/RNA duplex (P-GC-Crystallisation/T-RNA-Crystallisation), 1 mM ddCTP and 19 mM $MgCl_2$. Solutions were incubated at room temperature or 30 °C for 60 min and subsequently mixed in a 2:1, 1:1 or 1:2 ratio with the reservoir solution. Different commercially available screens as well as a screen (kindly provided by Dr. Samra Obeid) with conditions optimized for *KlenTaq* DNA polymerase were utilized.
In a second approach the same conditions were applied with the following changes: A different primer (P-GCC-Crystallisation) as well as 2'-deoxy-cytidine-5'-[(α,β)-imido]triphosphate (dCpNHpp) instead of ddCTP was employed in the set-up.

2.6.2.3. Crystallization of RT-KTq 2 in Complex with an all DNA Duplex

Crystallization solutions with RT-KTq 2 in complex with an all DNA duplex contained 6.4 mg/mL protein in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, DNA template (T-DNA-Crystallisation) (158 µM), DNA primer (P-GC-Crystallisation) (158 µM), 1 mM ddCTP and 19 mM $MgCl_2$. Solutions were incubated at room temperature for 60 min and subsequently mixed in 1:1 ratio with the reservoir solution (100 mM Tris HCl, pH 7.5, 0.2 M magnesium formate, and 15 % PEG 8000). Crystals were produced by the hanging drop vapour diffusion method by equilibrating against 0.5 mL of the reservoir solution at 18 °C (Qiagen, EasyXtal 15-well tools). Crystal formation was observed after 1 day. Crystals were flash-frozen in liquid nitrogen and were measured at 100 K with a wavelength of 1.00000 Å. Datasets were recorded on beamlines PXI and PXIII on a Pilatus 6M and 2M, respectively, at the SwissLightSource (SLS), Paul-Scherrer-Institut, Villigen, Switzerland. Data was processed and reduced using XDS.[205] The structure of RT-KTq 2_{DNA} was solved using difference Fourier methods. The structure was refined using the PHENIX suite.[206] Manual refining and model rebuilding was performed using Coot.[207] Model quality was determined by the MolProbity web server.[225] Molecular graphics were drawn with Pymol.[226]

2.6.2.4. Crystallization of RT-KTq 2 in Complex with a DNA/RNA Hybrid Duplex

Crystallization solutions containing RT-KTq 2 (6.6 mg/mL) in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, RNA template (T-RNA-Crystallisation) (217 µM), DNA primer (P-GC-Crystallisation) (217 µM), 1 mM ddCTP, 19 mM $MgCl_2$ were incubated at 30 °C for 60 min and subsequently mixed in 1:1 ratio with the reservoir solution (100 mM Tris HCl, pH 8.5, 0.2 M magnesium formate, and 20 % PEG 8000). Crystals were produced by the hanging drop vapour diffusion method by equilibrating against 0.5 mL of the reservoir solution at 18 °C (Qiagen, EasyXtal 15-well tools). Crystal formation was observed after 3 days for the RNA template. Streak seeding led to the formation of crystals with improved diffraction characteristics. Crystals were flash-frozen in liquid nitrogen and were measured at 100 K with a wavelength of 1.00000 Å. Datasets were recorded on beamlines PXI and PXIII on a Pilatus 6M and 2M, respectively, at the SwissLightSource (SLS), Paul-Scherrer-Institut, Villigen, Switzerland. Data was processed and reduced using XDS.[205] Molecular replacement (PHASER) was used to solve the phase problem.[208] Manual refining, model rebuilding and subsequent analysis were performed as described for the RT-KTq 2_{DNA} structure. In case of RT-KTq 2_{RNA} parts of the protein had to be retraced, as well as the hybrid duplex, neglecting the 2 outermost base pairs, which were not resolved.

2.6.3 Generation of Full-length *Taq* DNA Polymerase Variants

2.6.3.1. Cloning of Full-Length Taq DNA Polymerase Variants

Genes for *KlenTaq* wild-type, the parental enzymes M1 and M747K, KTq M1/M747K and the identified variant RT-KTq 2 encoded on the expression vector pGDR11 were amplified in PCR in addition to the endonuclease domain of *Taq* wild-type encoded on the pASK-IBA 37+ vector (see chapter **VII 2.3.6** for sequence). The endonuclease domain and the respective *KlenTaq* domain were then ligated in a blunt-ended manner and cloned back into the expression vector pGDR11.

In detail, reaction mixtures for amplification of the endonuclease domain of *Taq* wild-type, encoded on pASK-IBA 37+, contained 200 nM of primers forward (P-Exo-BE-fwd-*SphI*) and reverse (P-Exo-BE-rev), 40 ng of the respective plasmid, 200 µM dNTPs and 4 u Phusion DNA polymerase in 1x HF-buffer in a total volume of 200 µL. After initial denaturation for 30 s at 98 °C, 30 PCR cycles were performed with 10 s at 98 °C, 20 s at 70 °C, 20 s at 72 °C and one final elongation for 5 min at 72 °C.

PCR reactions for amplification of the respective *KlenTaq* domain, encoded on pGDR11, were performed under the same conditions with the following primers forward (P-KTQ-BE-fwd) and reverse (P-KTQ-*HindIII*), as well as a prolonged elongation time of 30 s. Further purification, digestion and ligation steps were conducted as described in chapters **VI 2.1.4–2.1.6** with *SphI* digestion of the endonuclease domain and HindIII digestion of the *KlenTaq*

domain. However, due to blunt-end ligation, endonuclease and *KlenTaq* domains had to be phosphorylated prior to the ligation step. Thus, the respective amplified products were purified with the MinElute Reaction Cleanup-Kit (Qiagen) after digestion and directly applied in the phosphorylation reaction. DATP was added to a final concentration of 800 µM as well as 12.5 u of PNK in the supplied buffer A (total volume 25 µL). After incubation for 30 min at 37 °C, the enzyme was heat-inactivated at 75 °C for 10 min. Transformation was conducted in *E. coli* BL21 (DE3) (see chapter **VII 2.3.7** for sequence).

2.6.3.2. Nuclease Activity Assay

Reaction mixtures (60 µL) contained 150 nM 22-nt substrate (Substrate Exonuc), 225 nM template (Template Exonuc), 50 nM of each dNTP and 150 nM of the respective *Taq* DNA polymerase in 1x *KlenTaq* Reaction Buffer. Cleavage of this substrate was determined at different time points (0, 5, 15, 30, 60 min) at a reaction temperature of 30 °C. Primer was labelled using [γ-^{32}P]-ATP described in chapter **VI 2.2.1**. Reactions were stopped by addition of 1x Denaturing PAGE Loading Dye, denatured at 95 °C for 5 min and analysed by 12 % denaturing PAGE. Visualization was performed by phosphoimaging.

2.6.3.3. Primer Extension Reactions with an RNA template

Reactions were performed as described in chapter **VI 2.6.1.5.** Incubation times are provided in the respective figure legends.

2.6.3.4. Real-time RT-PCR

Reactions were performed as described in chapter **VI 2.6.1.7** with MS2RNA employed as template in RT-PCR.

2.6.3.5. TaqMan based real-time RT-PCR

Total RNA was extracted from Jurkat cells using the RNeasy Mini Plus Kit (Qiagen) and digested with DNase I (Thermo Scientific) according to the manufacturer's protocol. The RNA was kindly provided by ProLago Biotech (Konstanz, Germany). Reaction mixtures (20 µL) contained 0.1 ng/µL RNA, 300 nM of each primer forward (P-TotalRNA-fwd) and reverse (P-TotalRNA-rev), 200 µM of each dNTP, 50 nM of the *Taq*Man probe and the respective concentration of RT-*Taq* 2 in 1x *KlenTaq* Reaction Buffer. The *Taq*Man probe consisted of an oligonucleotide with a 5' reporter dye (FAM) and a 3' BHQ-1 quencher (TaqMan-Probe-TotalRNA). Cycling was performed with 120 s initial denaturation at 95 °C and 50 cycles of

two-step cycling with denaturation for 15 s at 95 °C and combined annealing/extension for 30 s at 60 °C.

2.6.4 The Increased Substrate Spectrum of *KlenTaq* Variants

2.6.4.1. Primer Extension Reactions with Lesions-containing Templates

Reaction mixtures (20 µL) contained 150 nM radioactively labelled primer (P-F20), 225 nM F-containing template (F33XA), 200 µM of each dNTP and 25 nM of the respective *KlenTaq* DNA polymerase in 1x *KlenTaq* Reaction Buffer. Reaction mixtures were incubated at 72 °C and terminated after 5 min by addition of 45 µL 1x Denaturing PAGE Loading Dye. After denaturation at 95 °C for 5 min, reaction mixtures were separated using a 12 % denaturing PAGE gel. Visualization was performed by phosphoimaging.

For lesion bypass experiments with a three abasic site analogues containing template, the same conditions were applied with the exception of 100 nM polymerase concentration and primer P-F20, the natural (F90A) as well as the F-containing (F903X) template. Incubation times are provided in the respective figure legend.

Lesion bypass experiments with an 8-oxo-A/G (X) lesion were performed as described above, using DNA primer (P-F20), lesion-containing template (F33-8-oxoA/G) and 10 nM of the respective *KlenTaq* DNA polymerase. Incubation times are provided in the respective figure legend.

2.6.4.2. PCR Amplification from Damaged DNA

Plasmid DNA (pGDR11-codon optimized gene encoding *KlenTaq* wild-type, 100 ng/µL) was irradiated with UV light (λ=254 nm; 0.09 J $s^{-1}cm^{-2}$) on ice, 10 µL aliquots were taken after different incubation times and subsequently used as template in PCR. PCR reaction mixtures contained 125 pM irradiated plasmid, 200 nM primer forward (P-KTQ-*SphI* codon opt) and reverse (P-KTQ-*HindIII* codon opt), 200 µM of each dNTP and 50 nM polymerase in 1x *KlenTaq* Reaction Buffer. Initial denaturation at 95 °C for 3 min was conducted, 25 PCR cycles were performed with 1 min at 95 °C, 1 min at 65.1 °C and 2 min at 72 °C and one final elongation step for 10 min at 72 °C. Product formation (1648 bp) was analyzed on a 0.8 % TBE agarose gel.

2.6.4.3. Primer Extension Reactions with NTPs as Substrate

Primer extensions with NTPs as substrates were performed as described in chapter **VI 2.6.4.1**, using DNA primer (P-F20), DNA template (F33A) and 100 nM of the respective *KlenTaq* DNA polymerase. Reaction mixtures were incubated for 30 min at 72 °C.

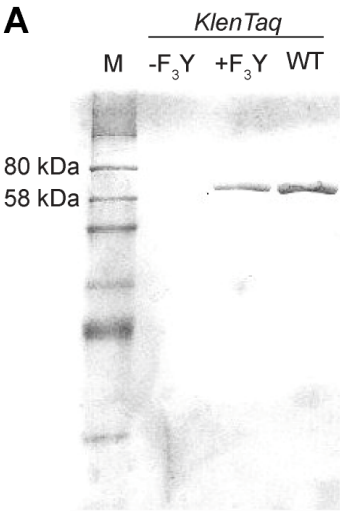
2.6.4.4. Error rate and Error-Spectrum

A PCR-based assay, reported by Patel *et al*.,[221] was used to determine the error rate and the error spectrum. PCR reactions (100 µL) contained 30 fmol template (pGDR11-codon optimized gene encoding *KlenTaq* wild-type), 250 µM of each dNTP, 400 nM of each primer (P-KTQ-*SphI* codon opt) and reverse (P-KTQ-*HindIII* codon opt) and 80 nM of DNA polymerase in 1x *KlenTaq* Reaction Buffer. Cycling was performed with 60 s initial denaturation at 95 °C and 30 PCR cycles with denaturation for 60 s at 95 °C, annealing for 60 s at 65.1 °C and extension for 2 min at 72 °C. PCR yield was determined via agarose gel analysis to calculate the number of replication cycles. PCR products were subsequently purified using preparative gel electrophoresis, digested with *SphI* and *HindIII* and cloned into the pGDR11 vector. Colonies were randomly picked and 650 bp from each clone were sequenced. Number of mutations per bp per clone yielded the mutation frequency. The error-rate was determined from the mutation frequency divided by the number of replication cycles.

VII. Appendix

1. Supplementary Information

1.1 MS/MS Analysis



B

Calculated Masses:
TINFGVLYGMSAHR 8: Trifluoro

N-Term.	Ion	a	b	i	y	C-Term.	Ion
1	T	74.060	102.055	74.060	175.119	14	R
2	I	187.144	215.139	86.096	312.178	13	H
3	N	301.187	329.182	87.055	383.215	12	A
4	F	448.255	476.250	120.081	470.247	11	S
5	G	505.277	533.272	30.034	601.287	10	M
6	V	604.345	632.340	72.081	658.309	9	G
7	L	717.429	745.424	86.096	875.344	8	Y*
8	Y*	934.464	962.459	190.047	988.428	7	L
9	G	991.486	1019.481	30.034	1087.496	6	V
10	M	1122.526	1150.521	104.053	1144.518	5	G
11	S	1209.558	1237.553	60.044	1291.586	4	F
12	A	1280.596	1308.590	44.049	1405.629	3	N
13	H	1417.654	1445.649	110.071	1518.713	2	I
14	R	1573.756	1601.750	129.113	1619.761	1	T

y_{13} y_{12} y_{11} y_{10} y_9 y_8 y_7 y_6 y_5 y_4 y_3 y_2
T I N F G V L Y* G M S A H R
b_2 b_3 b_4 b_5 b_6 b_7 b_9 b_{13}

Y* = 2,3,5-Trifluorotyrosine

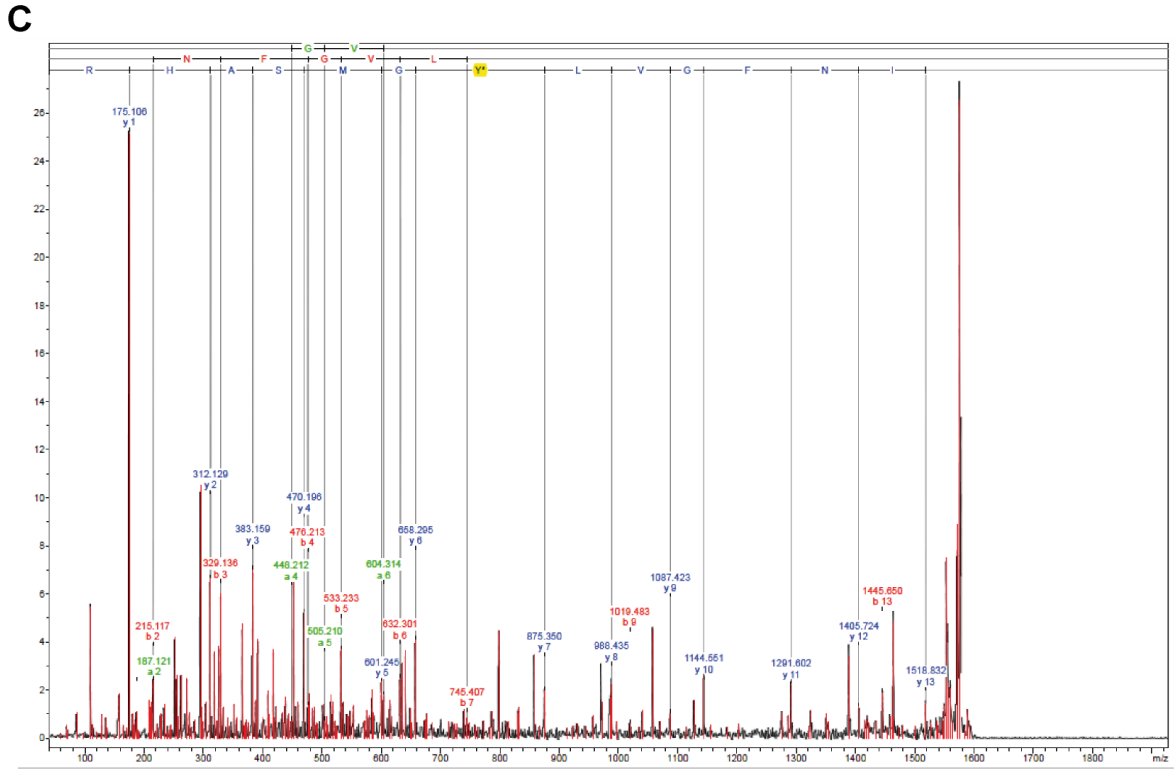


Figure S1. *KlenTaq* F_3Y analysis. A) SDS-PAGE showing the incorporation of F_3Y into *KlenTaq*. M: Marker, $-F_3Y$: *KlenTaq* expressed without F_3Y, $+F_3Y$: *KlenTaq* expressed with F_3Y, WT: Wild-type *KlenTaq*. SDS-PAGE shows heat-treated, purified protein. B) Calculated masses of a-, b-, i- and y-ions of the fluorinated tyrosine containing peptide. All b- and y-ions marked in the peptide sequence were found in the corresponding spectrum. C) Corresponding Maldi-MS/MS spectrum of the peptide in B).

1.2 Specific Activity Measurements

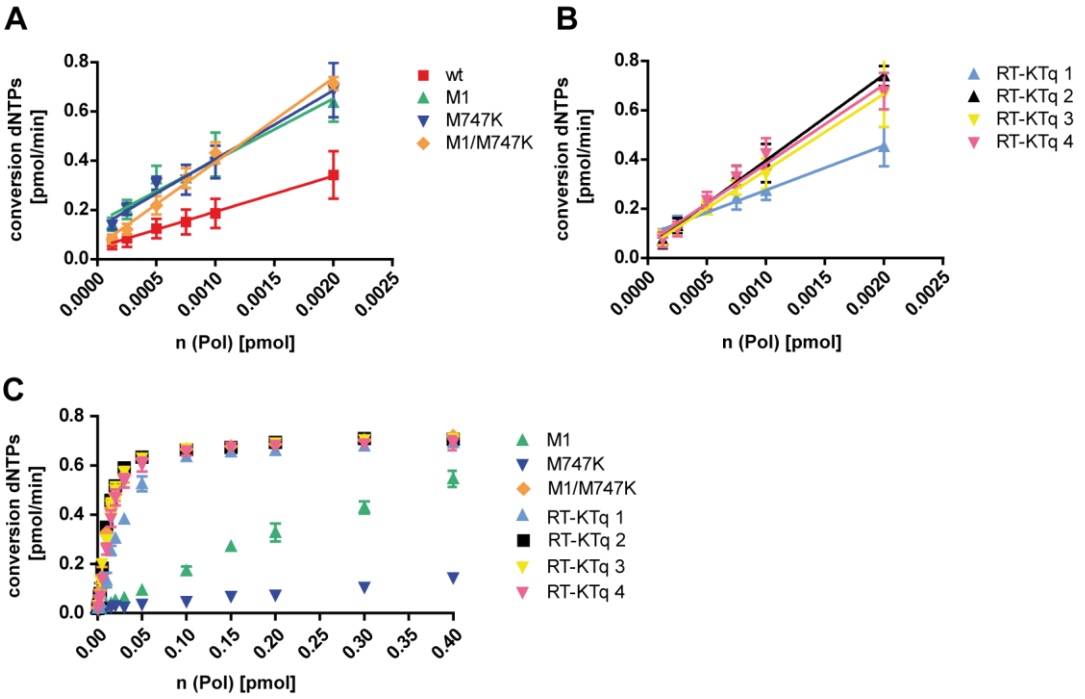


Figure S2. Specific activities on DNA template (A-B) and RNA template (C). Quantified conversion of dNTPs per minute was plotted against varying polymerase amounts used; the resulting slope in the linear range yielding the specific activity of the respective enzyme.

1.3 Sugar Pucker Conformations

Table S1. Sugar puckers for the primer and template nucleotides in the DNA/RNA hybrid duplex.

Primer

Base	v0	v1	v2	v3	v4	P	vmax	χ	γ	type
dC	11.21	-23.74	26.61	-20.82	6.21	**354.47**	26.73	-163.74	160.71	C2'-exo
dC	-2.24	-21.63	35.56	-37.7	25.45	**21.8**	38.3	-161.6	45.25	C3'-endo
dA	-1.58	-23.29	37.72	-39.59	26.13	**20.76**	40.34	-166.82	54.64	C3'-endo
dC	1.63	-25.09	37.5	-37.48	22.83	**16.23**	39.05	-160.48	50.12	C3'-endo
dG	-18.04	9.46	2.04	-12.48	19.07	**83.93**	19.29	-139.13	51.44	O4'-endo
dG	-19.66	34.04	-34.85	24.39	-3.18	**166.31**	35.87	-112.52	50.37	C2'-endo
dC	-43.81	38.72	-19.84	-4.78	30.39	**117.42**	43.08	-125.91	48.32	C1'-exo
dG	-32.22	46.09	-42.2	25.4	4	**156.34**	46.07	-113.74	28.87	C2'-endo
dC	1.61	-26.4	40.84	-41.7	24.75	**17**	42.71	-156.56	174.77	C3'-endo
ddC	1.31	-19.96	30.01	-29.15	17.36	**15.28**	31.11	-149.56	51.87	C3'-endo

Template

Base	v0	v1	v2	v3	v4	P	vmax	χ	γ	type
A	-19.93	35.34	-36.3	25.8	-3.91	167.11*	37.24	13.35	57.75	C2'-endo
A	-22.15	37.44	-37.52	25.76	-2.48	164.81*	38.88	-142.24	-172.22	C2'-endo
A	-26.09	40.25	-38.17	24.12	1.08	159.76*	40.68	123.09	-147.65	C2'-endo
G	7.21	-28	36.98	-33.97	16.96	**7.86**	37.33	-176.53	-168.16	C3'-endo
G	6.08	-27.45	37.18	-34.67	18.05	**9.52**	37.7	-162.19	50.47	C3'-endo
G	4.9	-26.87	37.58	-35.77	19.45	**11.46**	38.35	-156.88	51.82	C3'-endo
C	1.7	-25.66	38.57	-38.73	23.39	**16.32**	40.19	-158.8	57.18	C3'-endo
G	10.49	-27.7	33.52	-28.57	11.46	**1.02**	33.53	-173.83	-167.56	C3'-endo
C	2.5	-24.5	35.91	-35.48	20.85	**14.86**	37.16	-159.5	50.84	C3'-endo
C	1.33	-23.15	34.88	-35.07	21.3	**16.55**	36.39	-154.96	56.9	C3'-endo
G	2.3	-24.54	36.27	-35.9	21.2	**15.16**	37.58	-159.16	50.24	C3'-endo
U	1.44	-23.81	35.94	-36.13	21.87	**16.49**	37.48	-160.42	48.39	C3'-endo
G	2.15	-23.4	34.51	-34.25	20.3	**15.27**	35.78	-179.1	174.01	C3'-endo
G	4.94	-26.31	36.5	-34.66	18.77	**11.17**	37.2	179.37	159.34	C3'-endo

* The first three nucleotides in the template show a C2'-endo conformation, but due to low resolution of these residues we cannot make any predictions about the sugar conformations of these residues. P-values for the first three bases of the template are therefore not highlighted in bold. The two outermost base pairs are not listed, as they were not resolved in RT-KTq 2_{RNA}.

2. Sequences

2.1 Oligonucleotides

2.1.1 Primers and Templates for Primer Extension, PCR and RT-PCR

Primer Extension

P-F20 5'-d(CGT TGG TCC TGA AGG AGG AT)-3'
P-F20- 5'-d(CGC GCA GCA CGC GCC GCC GT)-3'
P-F23 5'-d(CGT TGG TCC TGA AGG AGG ATA GG)-3'

F33A 5'-d(AAA TCA ACC TAT CCT CCT TCA GGA CCA ACG TAC)-3'
F33-7T 5'-d(AAA TCA TCC TAT CCT CCT TCA GGA CCA ACG TAC)-3'

F33XA 5'-d(AAA TCA FCC TAT CCT CCT TCA GGA CCA ACG TAC)-3'
F: abasic site analogue

F33-8-oxoA/G 5'-d(AAA TCA XCC TAT CCT CCT TCA GGA CCA ACG TAC)-3'
X: lesion

F33dU	5’-d(AAA TCA dUCC TAT CCT CCT TCA GGA CCA ACG TAC)-3’
F30RNA	5’-AAA UCA ACC UAU CCU CCU UCA GGA CCA ACG-3’
F90A	5’-d(CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC CGC CGT GGA CAG AGG ACT GCA GAA AAT CAA CCT ATC CTC CTT CAG GAC CAA CGT ACA GAG)-3’
F903X	5’-d(CCG TCA GCT GTG CCG TCG CGC AGC ACG CGC FGC CFT GGA CAF AGG ACT GCA GAA AAT CAA CCT ATC CTC CTT CAG GAC CAA CGT ACA GAG)-3’ F: abasic site analogue
P-Blunt End	5’-d(CGT TGG TCC TGA AGG AGG AT)-3’
T-Blunt End	5’-d(ATC CTC CTT CAG GAC CAA CGA AA)-3’
Substrate Exonuc	5’-d(CCC CCC CCC CTC ATA CGT ACA C)-3’
Template Exonuc	5’-d(GTG TAC GTA TGA TCA TGC AGG TAG CCG ATG AAC TGG TCG AAA GAC CAG TTC ATC GGC TAC CTG CAT GAT)-3’

PCR

P-DNA Shuffling primer fwd	5’-d(CGA GGC CCT TTC GTC TTC AC)-3’
P-DNA Shuffling primer rev	5’-d(CTT AGC TCC TGA AAA TCT CGC C)-3’
P-MS2-5’	5’-d(ATC GCT CGA GAA CGC AAG TT)-3’
P-MS2-3’	5’-d(CG GAC TTC ATG CTG TCG GTG)-3’
MS2DNA	5’-d(ATC GCT CGA GAA CGC AAG TTC TTC AGC GAA AAG CAC GAC AGT GGT CGC TAC ATA GCG TGG TTC CAT ACT GGA GGT GAA ATC ACC GAC AGC ATG AAG TTC G)-3’
P-KTQ-*SphI* codon opt	5'-d(GGA TCC *GCA TGC* AGC ACT GGA AGA AGC ACC TTG GCC TCC G)-3'
P-KTQ-*HindIII* codon opt	5'-d(CTA ATT *AAG CTT* TTA TTC TTT TGC AGA CAG CC)-3'

RT-PCR

P-MS2-5’	5’-d(ATC GCT CGA GAA CGC AAG TT)-3’
P-MS2-3’	5’-d(CG GAC TTC ATG CTG TCG GTG)-3’
P-MS2-5’-ES 490	5’-d(GAT CGC ATG CCT AGA GGC ACT TGC CTA CTA CG)-3’
P-MS2-5’-ES 301	5’-d(GAT CGC ATG CGT CGG AGG TAT GTC AGA TCC AC)-3’
P-MS2-3’-ES	5’-d(GCT AAA GCT TCG GAC TTC ATG CTG TCG GTG ATT TC)-3’
P-S16-5’	5’-d(CTG GCG GCA GGC CTA ACA CA)-3’
P-S16-3’	5’-d(GCA GTT TCC CAG ACA TTA CT)-3’
P-Influenza A-fwd	5'-d(CCM AGG TCG AAA CGT AYG TTC TCT CTA TC)-3’
P-Influenza A-rev	5'-d(TGA CAG RAT YGG TCT TGT CTT TAG CCA YTC CA)-3’
P-Influenza B-fwd	5'-d(TGG AAC CAA ATA TAG ACC TCC TG)-3’
P-Influenza B-rev	5'-d(GTG TAT CCG TGC CAA CCT G)-3’
P-TotalRNA-fwd	5'-d(CAC TCT TCC AGC CTT CCT TC)-3’
P-TotalRNA-rev	5'-d(GGA TGT CCA CGT CAC ACT TC)-3’
TaqMan Probe-Influenza A	5'-FAM-d(ATY TCG GCT TTG AGG GGG CCT G)-MGB-3’
TaqMan Probe-Influenza B	5'-HEX-d(TTC GGA GCT ATT GCT GGT TTC)-BHQ-1-3’
TaqMan Probe-TotalRNA	5'-FAM-d(TGC CAC AGG ACT CCA TGC CC)-BHQ-1-3’
TaqMan-Aptamer	5'-d(TTC TCG GTT GGT CTC TGG CGG AGC AAG ACC AGA CAA TGT ACA GTA TT G GCC TGA TCT TGT GTA TGA TTC GCT TTT CCC)-3’

2.1.2 Primers for Cloning

P-KTQ-*SphI*	5’-d(CAT ACG GAT CC*G CAT GC*A GCC CTG GAG GAG GCC C)-3’
P-KTQ-*HindIII*	5’-d(GCT CAG CTA ATT *AAG CTT* TCT CCT TGG CGG AGA GCC)-3’
P-Exo-BE-fwd *SphI*	5’-d(CGG ATC C*GC ATG C*AA TGA GGG GGA TGC TGC) -3’
P-Exo-BE-rev	5’-d(CTT GGG GCT TTC CAG AAG GCC GAA CTC G) -3’

P-KTQ-BE-fwd	5’-d(GCC CTG GAG GAG GCC CCC TGG CCC CCG)-3’
P-KTQ-*SphI* codon opt	5'-d(GGA TCC *GCA TGC* AGC ACT GGA AGA AGC ACC TTG GCC TCC G)-3'
P-KTQ-*HindIII* codon opt	5'-d(CTA ATT *AAG CTT* TTA TTC TTT TGC AGA CAG CC)-3'

2.1.3 Primers for Site-directed Mutagenesis

P-Y671F-2 fwd	5’-d(CAT CAA CTT CGG GGT CCT CTT CGG CAT GTC GG)-3’
P-Y671F-2 rev	5’-d(CCG ACA TGC CGA AGA GGA CCC CGA AGT TGA TG)-3’
P-M747K fwd	5’-d(GGC CGA GCG CAA GGC CTT CAA CAT G)-3’
P-M747K rev	5’-d(CAT GTT GAA GGC CTT GCG CTC GGC C)-3’
P-NB_SD_L459M	5'-d(GTT GCA TAT CTG CGT GCA ATG AGC CTG GAA GTT GCA GAA G)-3’
P-NB_SD_S515R	5'-d(CGG TAA ACG TAG CAC CCG CGC AGC AGT TCT GGA AGC C)-3’
P-NB_SD_I638F	5'-d(GTT TCA GGA AGG TCG CGA TTT TCA TAC CGA AAC CGC AAG C)-3’
P-NB_SD_M747K	5'-d(CGT GAA GCA GCA GAA CGT AAA GCC TTT AAT ATG CCG G)-3’

2.1.4 Primers and Templates for Crystallization

P-GC-Crystallisation	5’-d(GAC CAC GGC GC)-3’
P-GCC-Crystallisation	5’-d(GAC CAC GGC GCC)-3’
T-DNA-Crystallisation	5’-d(AAA GGG CGC CGT GGT C)-3’
T-RNA-Crystallisation	5’-AAA GGG CGC CGU GGU C-3’

2.1.5 Primers for Sequencing

pQE-FP	5´-d(CGG ATA ACA ATT TCA CAC AG)-3´
pQE-RP	5´-d(GTTCTG AGG TCA TTA CTGG)-3´
pQE-KTQ-mid	5'-d(CGT AAG GGA TGG CTA GCTCC)-3'

2.2 Plasmids

2.2.1 pGDR11

PGDR11 is a derivative of pQE31 vector.[184]

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATT
CATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCATGCGAGCTCGGTACCCCGGGTCGACCTGC
AGCCAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTG
CCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGA
TATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCA
GCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATG
CTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACT
GAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAA
CCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGG
CCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTT
CATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTT
TTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCTCTAGATTACCGTGCAGTCGATGATAAGCTG
TCAAACATGAGAATTGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGC
CAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCA
ACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATG
GTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGT
AATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTT
GAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGC
AGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACC
GTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAG
CAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCG
ACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTG
CAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCG
CCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCA
TACTCTGCGACATCGTATAACGTTACTGGTTTCATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTT
TGCACCATTCGATGGTGTCGGAATTTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCTAGAGCTGCCTCGCGCGTTTCGGTGATG
ACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCG
TCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAG
CAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCTGTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGG
ATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAG
CTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTA
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGCTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAAT
AAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG
TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCA
CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT
GTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA

AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

2.2.2 pASK-IBA 37+

CCATCGAATGGCCAGATGATTAATTCCTAATTTTTGTTGACACTCTATCATTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAA
AGTGAAATGAATAGTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAAATGGCTAGCAGAGGATCGCATCACC
ATCACCATCACATCGAAGGGCGCCGAGACCGCGGTCCCGAATTCGAGCTCGGTACCCGGGGATCCCTCGAGGTCGACCTGCAGGGGGACCAT
GGTCTCTGATATCTAACTAAGCTTGACCTGTGAAGTGAAAAATGGCGCACATTGTGCGACATTTTTTTTGTCTGCCGTTTACCGCTACTGCG
TCACGGATCTCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTT
CCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTC
GCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGAT
TTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCT
TACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA
GACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCG
GCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGA
ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG
CGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACA
GAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGAC
AACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATG
AAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTA
GCTTCCCGGCAACAATTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGC
TGATAAATCTGGAGCCGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAGGAATTAATGATG
TCTCGTTTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCCA
GAAGCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACGCCTTAGCCATTGAGATGTTAGATAGGC
ACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCAT
CGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAACAAGG
TTTTTCACTAGAGAATGCATTATATGCACTCAGCGCAGTGGGGCATTTTACTTTAGGTTGCGTATTGGAAGATCAAGAGCATCAAGTCGCTA
AAGAAGAAAGGGAAACACCTACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGATCACCAAGGTGCAGAGCCAGCC
TTCTTATTCGGCCTTGAATTGATCATATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAAAAGCAGCATAACCTTTTTCCGTG
ATGGTAACTTCACTAGTTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGT
CCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCG
TGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTC
CTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCC
TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGACCCGACA

2.3 Expression Vectors

The initial ATG codon is shown in green, the coding sequence for the His-Tag is depicted in orange, the respective coding *KlenTaq* sequence (based on the amino acids 293-832 of *Taq* gene) or *Taq* sequence is italicized in blue. Enzymes expressed based on the pGDR11 vector contain an N-terminal (MRGSHHHHHHTDPHA) and C-terminal (KA) amino acid additions (shown underlined).

2.3.1 *KlenTaq* Wild-type in pGDR11

KlenTaq wild-type in pGDR11 was used to express *KlenTaq* wild-type with an N-terminal His-tag for the studies in **chapter II** and the characterization experiments in **chapter III**.

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATT
CATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCATGCAGCCCTGGAGGAGGCCCCCTGGCCCC
CGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGCCGG
GTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAG
GGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGC
GCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAG
GAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGC
CTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCA
ACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGC
GCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACAT
TGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCT
CCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACAT
CCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGG
TCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGC
TTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGT
GCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCA
TGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCC
CCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGG
GATAGGGGAGGACTGGCTCTCCGCCAAGGAGAAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCC
ATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGA
AGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTG
CTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTT
ATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCC
TTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATT
CGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTG
AGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAA
GGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCG
ATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAAT
AAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCTCT
AGATTACCGTGCAGTCGATGATAAGCTGTCAAACATGAGAATTGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTG
GTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG
CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATAT
CCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATG
CCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGT
GAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGA
CCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAG
ATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAA
TCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCC
ACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCG
GGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCATTCACCACCCTGAATTGACTCTCTTCCGGG
CGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCGGAATTTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATC
TAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGT
ATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAT
ACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCTGTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC

GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACC
CCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
CATAGCTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCAC
GCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAG
TCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTA
GCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCC
GGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
GAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

2.3.2 *KlenTaq* M747K in pGDR11

KlenTaq M747K in pGDR11 was used to express *KlenTaq* M747K with an N-terminal His-tag and was characterised in later experiments in **chapter III**. Mutation is depicted in red.

…CATGCA*GCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATC*
TTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTT
CTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCC
TTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCT
TCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCAC
ATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGT
CTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCA
AGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTAC
CGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCA
GACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGG
CCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAAC
CTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGAT
GCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGG
CCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTAC
*GTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGC**AAG**GCCTTCAA*
CATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCC
TTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTAT
*CCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG*AAAGCTTAATTAGCT…

2.3.3 *KlenTaq* M1 in pGDR11

KlenTaq M1 in pGDR11 was used to express *KlenTaq* M1 (L322M, L459M, S515R, I638F, S739G, E773G) with an N-terminal His-tag and was characterised in later experiments in **chapter III**. Mutations are depicted in red.

…CATGCA*GCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATC*
*TT**ATG**GCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTT*
CTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCC
TTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCT
TCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCAC
*ATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCC**ATG**TCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGT*
CTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCA
*AGACGGAGAAGACCGGCAAGCGCTCCACC**AGA**GCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTAC*
CGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCA
GACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGG
CCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAAC
*CTGATCCGGGTCTTCCAGGAGGGGCGGGAC**TTC**CACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGAT*
GCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGG
CCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTAC
*GTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAG**GGC**GTGCGGGAGGCGGCCGAGCGCATGGCCTTCAA*
*CATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTG**GGG**GAAATGGGGGCCAGGATGCTCC*
TTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTAT
*CCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG*AAAGCTTAATTAGCT…

2.3.4 RT-KTq 2 without His-tag in pGDR11*

RT-KTq 2 (codon optimized) without His-tag in pGDR11* was used to express RT-KTq 2 for crystallization. It was generated by site-directed mutagenesis employing the plasmid encoding *KlenTaq* wild-type without His-tag in pGDR11* kindly provided by Dr. Bastian Holzberger (Dissertation, 2012, Universität Konstanz; pGDR11**KlenTaq* WTΔHis6).

…ATACAT*ATGGCACTGGAAGAAGCACCTTGGCCTCCGCCTGAAGGTGCATTTGTTGGTTTTGTTCTGAGCCGTAAAGAACCGATGTGGGCAG*
ATCTGCTGGCACTGGCAGCAGCACGTGGTGGTCGTGTTCATCGTGCACCGGAACCGTATAAAGCTCTGCGCGATCTGAAAGAAGCACGCGGT
CTGCTGGCAAAAGATCTGAGCGTTCTGGCACTGCGTGAAGGTCTGGGACTGCCTCCGGGTGATGATCCGATGCTGCTGGCATATCTGCTGGA
TCCGAGCAATACCACACCGGAAGGTGTTGCACGTCGTTATGGTGGTGAATGGACCGAAGAAGCAGGCGAACGCGCAGCACTGAGCGAACGTC
TGTTTGCAAATCTGTGGGGTCGTCTGGAAGGTGAAGAACGTCTGCTGTGGCTGTATCGTGAAGTTGAACGTCCGCTGTCTGCAGTTCTGGCA
*CACATGGAAGCAACCGGTGTTCGTCTGGATGTTGCATATCTGCGTGCA**ATG**AGCCTGGAAGTTGCAGAAGAAATTGCACGTCTGGAAGCAGA*
AGTTTTTCGTCTGGCAGGCCATCCGTTTAATCTGAATAGCCGTGATCAGCTGGAACGTGTTCTGTTTGATGAACTGGGTCTGCCAGCAATTG
*GTAAAACCGAAAAAACCGGTAAACGTAGCACC**CGC**GCAGCAGTTCTGGAAGCCCTGCGTGAAGCACATCCGATTGTGGAAAAAATTCTGCAG*
TATCGCGAACTGACCAAACTGAAAAGCACCTATATCGATCCGCTGCCGGATCTGATTCATCCGCGTACCGGTCGTCTGCATACCCGTTTTAA
TCAGACCGCAACCGCAACCGGTCGCCTGAGCAGCAGCGATCCGAATCTGCAGAATATTCCGGTTCGTACACCGCTGGGTCAGCGTATTCGTC
GTGCATTTATTGCAGAAGAAGGTTGGCTGCTGGTTGCACTGGATTATAGCCAGATAGAACTGCGTGTTCTGGCCCATCTGAGCGGTGATGAA
*AATCTGATTCGCGTGTTTCAGGAAGGTCGCGAT**TTT**CATACCGAAACCGCAAGCTGGATGTTTGGTGTTCCGCGTGAAGCAGTTGATCCGCT*
GATGCGTCGTGCAGCAAAAACCATTAATTTTGGCGTGCTGTATGGTATGAGCGCACATCGTCTGAGCCAGGAACTGGCAATTCCGTACGAAG
AAGCCCAGGCATTTATCGAACGTTATTTTCAGAGCTTTCCGAAAGTTCGTGCCTGGATTGAAAAAACCCTGGAAGAAGGTCGTCGTCGCGGT
*TATGTTGAAACCCTGTTTGGTCGTCGTCGTTATGTTCCGGATCTGGAAGCACGTGTTAAAAGCGTTCGTGAAGCAGCAGAACGT**AAA**GCCTT*
TAATATGCCGGTTCAGGGCACCGCAGCAGATCTGATGAAACTGGCCATGGTTAAACTGTTTCCGCGTCTGGAAGAAATGGGTGCACGTATGC
TGCTGCAGGTTCATGATGAACTGGTGCTGGAAGCACCGAAAGAACGTGCAGAAGCAGTTGCCCGTCTGGCAAAAGAAGTTATGGAAGGCGTT
*TATCCGCTGGCAGTTCCGCTGGAAGTTGAAGTTGGTATTGGTGAAGATTGGCTGTCTGCAAAAGAA*TAAAAGCTTAAT…

2.3.5 *KlenTaq* M1 in pASK-IBA 37+

KlenTaq M1 in pASK-IBA 37+ was used to generate *KlenTaq* M1 (L322M, L459M, S515R, I638F, S739G, E773G) in pGDR11 by cloning the *KlenTaq* domain into the pGDR11 vector.

CCATCGAATGGCCAGATGATTAATTCCTAATTTTTGTTGACACTCTATCATTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAA
AGTGAAATGAATAGTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAA**ATG**GCTAGCAGAGGATCG**CATCACC**
ATCACCATCACATCGAAGGGCGC*GCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAG*
*CCCATGTGGGCCGATCTT**ATG**GCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAA*
GGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCG
CCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCC
CTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTC
*CGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCC**ATG**TCCCTGGAGGTGGCCGAGGAGATCGCCC*
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGG
*CTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACC**AGA**GCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGA*
GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCC
ACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGG
CAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCT
*CTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGAC**TTC**CACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGG*
CCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCC
ATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGG
*CAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAG**GGC**GTGCGGGAGGCGGCCG*
*AGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTG**GGG**GAAATG*
GGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGT
*CATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG*TGATATCTAACTA
AGCTTGACCTGTGAAGTGAAAAATGGCGCACATTGTGCGACATTTTTTTTGTCTGCCGTTTACCGCTACTGCGTCACGGATCTCCACGCGCC
CTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTT
TCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTC
CACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGA
TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTCAGGTGGCACT
TTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTT
TTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
AAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGA
CGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATG
GCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGA
GCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTGA
TAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGT
GAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAGGAATTAATGATGTCTCGTTTAGATAAAAGTA
AAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTAGAGCAG
CCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACGCCTTAGCCATTGAGATGTTAGATAGGCACCATACTCACTTTTGCCC
TTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGTAC
ATTTAGGTACACGGCCTACAGAAAAACAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTTATGCCAACAAGGTTTTTCACTAGAGAATGCA
TTATATGCACTCAGCGCAGTGGGGCATTTTACTTTAGGTTGCGTATTGGAAGATCAAGAGCATCAAGTCGCTAAAGAAGAAAGGGAAACACC
TACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGATCACCAAGGTGCAGAGCCAGCCTTCTTATTCGGCCTTGAAT
TGATCATATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAAAAGCAGCATAACCTTTTTCCGTGATGGTAACTTCACTAGTTT
AAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGA
AAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT
TGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG

GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCT
CTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT
TTTGCTGGCCTTTTGCTCACATGACCCGACA

2.3.6 *Taq* Wild-type in pASK-IBA 37+

Taq wild-type in pASK-IBA 37+ was used to generate *Taq* wild-type and all variants in pGDR11 by cloning the endonuclease domain into the pGDR11 vector.

TCTGCCGTTTACCGCTACTGCGTCACGGATCTCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGAC
CGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAA
ATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCA
TCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCC
TATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA
ATTTTAACAAAATATTAACGCTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACAT
TCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACT
TTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTT
GGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACA
CTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGAT
CGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTGATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCC
TTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC
TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAA
GCATTGGTAGGAATTAATGATGTCTCGTTTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTT
TAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACGCCTTA
GCCATTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTT
TAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACAGTATGAAACTCTCGAAAATCAAT
TAGCCTTTTTATGCCAACAAGGTTTTTCACTAGAGAATGCATTATATGCACTCAGCGCAGTGGGGCATTTTACTTTAGGTTGCGTATTGGAA
GATCAAGAGCATCAAGTCGCTAAAGAAGAAAGGGAAACACCTACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGA
TCACCAAGGTGCAGAGCCAGCCTTCTTATTCGGCCTTGAATTGATCATATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAAA
AGCAGCATAACCTTTTTCCGTGATGGTAACTTCACTAGTTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG
CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA
GAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG
CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCG
GTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGA
AACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATG
GAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGACCCGACACCATCGAATGGCCAGATGAT
TAATTCCTAATTTTTGTTGACACTCTATCATTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAATGAATAGTTCGAC
AAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAAGCTAGCAGAGGATCGATCGAAGGG
CGC*ATGAGGGGGATGCTGCCCCTCTTTGAGCCCAAGGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCCACGCCCT*
GAAGGGCCTCACCACCAGCCGGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACG
CGGTGATCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACGGGGGGTACAAGGCGGGCCGGGCCCCCACGCCGGAGGAC
TTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCT
GGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCA
TCCACGTCCTCCACCCCGAGGGGTACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGCCGACTACCGG
GCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGAGAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCT
GGAAGCCCTCCTCAAGAACCTGGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTGGGACC
TGGCCAAGGTGCGCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGG
CTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTT
CGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCG

AGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTC
CCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTG
GACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGC
TTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTG
TCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCT
GGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGG
CCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGAC
CTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCC
AGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCC
AGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTC
GGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGG
CCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCC
CGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGT
GAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGG
AGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGG
*CTCTCCGCCAAGGAG*TGATATCTAACTAAGCTTGACCTGTGAAGTGAAAAATGGCGCACATTGTGCGACATTTTTTTTG

2.3.7 *Taq* Wild-type in pGDR11

Taq wild-type in pGDR11 was used used to express *Taq* wild-type with an N-terminal His-tag employed in experiments in **chapter III**.

CTCGAGAAATCATAAAAAAT*TTATTTGCTTTGTGAGCGG*ATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATT
CATTAAAGAGGAGAAATTAACTATGAGAGGATCTCACCATCACCATCACCATACGGATCCGCATGCA*ATGAGGGGGATGCTGCCCCTCTTTG*
AGCCCAAGGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGGAG
CCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGCCAAGGC
CCCCTCCTTCCGCCACGAGGCCTACGGGGGGTACAAGGCGGGCCGGGCCCCCACGCCGGAGGACTTTCCCCGGCAACTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGCTGGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGCCAAGAAGGCGGAAAAG
GAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCT
CATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGCCGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACC
TTCCCGGGGTCAAGGGCATCGGGGAGAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCGG
CTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACCTGCCCCT
GGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGT
TCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAG
CCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAA
GGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCG
CCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCC
CTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTC
CGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGG
CTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGA
GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCC
ACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGG
CAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCT
CTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGG
CCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCC
ATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGG
CAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCG
AGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATG
GGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGT
*CATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG*AAAGCTTAATTAG
CTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTT
ATT*GGTGAGAATCCAAGCTAG*CTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATA

TATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCC
TTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCG
TATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGC
TCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCT
AAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTT
CTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTG
ATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGG
TGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCT
GTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCTCTAGATTACCGTGCAGTCGATGATAAGCTGTCAAACATGAGAATTG
TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTT
CACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA
TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCG
CCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGC
ACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAA
ATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTC
ATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTA
CCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAG
GTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGT
ATAACGTTACTGGTTTCATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGT
GTCGGAATTTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTG
ACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCG
GGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAG
TGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCTGTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAA
CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCT
CCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
GAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAG
AGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATG
CTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGCTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGG
AGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGT
TAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAAC
GATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAAC
CAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACA
TTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

2.4 Protein Sequences

2.4.1 Sequence Alignment of *KlenTaq* Wild-type and Variants

The initial methionine is shown in green, the His-Tag is depicted in orange, the respective *KlenTaq* domain (based on the amino acids 293-832 of *Taq* gene) is shown in black. Additional amino acids resulting from the expression vector pGDR11 are shown in blue. Mutations are highlighted in red.

```
KTq_M1/M747K_pGDR11      MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLMALAAA
KTq_M1_pGDR11            MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLMALAAA
RT-KTq3_pGDR11           MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLMALAAA
RT-KTq4_pGDR11           MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLMALAAA
RT-KTq2_pGDR11           MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAA
KTq_M747K_pGDR11         MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAA
KTq_wt_pGDR11            MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAA
RT-KTq1_pGDR11           MRGSHHHHHHTDPHAALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAA

KTq_M1/M747K_pGDR11      RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
KTq_M1_pGDR11            RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
RT-KTq3_pGDR11           RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
RT-KTq4_pGDR11           RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
RT-KTq2_pGDR11           RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
KTq_M747K_pGDR11         RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
KTq_wt_pGDR11            RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
RT-KTq1_pGDR11           RGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA

KTq_M1/M747K_pGDR11      YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
KTq_M1_pGDR11            YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
RT-KTq3_pGDR11           YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
RT-KTq4_pGDR11           YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
RT-KTq2_pGDR11           YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
KTq_M747K_pGDR11         YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
KTq_wt_pGDR11            YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL
RT-KTq1_pGDR11           YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLL

KTq_M1/M747K_pGDR11      WLYREVERPLSAVLAHMEATGVRLDVAYLRAMSLEVAEEIARLEAEVFRL
KTq_M1_pGDR11            WLYREVERPLSAVLAHMEATGVRLDVAYLRAMSLEVAEEIARLEAEVFRL
RT-KTq3_pGDR11           WLYREVERPLSAVLAHMEATGVRLDVAYLRAMSLEVAEEIARLEAEVFRL
RT-KTq4_pGDR11           WLYREVERPLSAVLAHMEATGVRLDVAYLRAMSLEVAEEIARLEAEVFRL
RT-KTq2_pGDR11           WLYREVERPLSAVLAHMEATGVRLDVAYLRAMSLEVAEEIARLEAEVFRL
KTq_M747K_pGDR11         WLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRL
KTq_wt_pGDR11            WLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRL
RT-KTq1_pGDR11           WLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRL

KTq_M1/M747K_pGDR11      AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
KTq_M1_pGDR11            AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
RT-KTq3_pGDR11           AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
RT-KTq4_pGDR11           AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
RT-KTq2_pGDR11           AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
KTq_M747K_pGDR11         AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHP
KTq_wt_pGDR11            AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHP
RT-KTq1_pGDR11           AGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTRAAVLEALREAHP
```

```
KTq_M1/M747K_pGDR11      IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
KTq_M1_pGDR11            IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
RT-KTq3_pGDR11           IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
RT-KTq4_pGDR11           IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
RT-KTq2_pGDR11           IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
KTq_M747K_pGDR11         IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
KTq_wt_pGDR11            IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS
RT-KTq1_pGDR11           IVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSS

KTq_M1/M747K_pGDR11      DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
KTq_M1_pGDR11            DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
RT-KTq3_pGDR11           DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
RT-KTq4_pGDR11           DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
RT-KTq2_pGDR11           DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
KTq_M747K_pGDR11         DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
KTq_wt_pGDR11            DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN
RT-KTq1_pGDR11           DPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDEN

KTq_M1/M747K_pGDR11      LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
KTq_M1_pGDR11            LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
RT-KTq3_pGDR11           LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
RT-KTq4_pGDR11           LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
RT-KTq2_pGDR11           LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
KTq_M747K_pGDR11         LIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
KTq_wt_pGDR11            LIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR
RT-KTq1_pGDR11           LIRVFQEGRDFHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHR

KTq_M1/M747K_pGDR11      LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
KTq_M1_pGDR11            LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RT-KTq3_pGDR11           LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RT-KTq4_pGDR11           LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RT-KTq2_pGDR11           LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
KTq_M747K_pGDR11         LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
KTq_wt_pGDR11            LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RT-KTq1_pGDR11           LSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR

KTq_M1/M747K_pGDR11      RYVPDLEARVKGVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLGEMGA
KTq_M1_pGDR11            RYVPDLEARVKGVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLGEMGA
RT-KTq3_pGDR11           RYVPDLEARVKSVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLGEMGA
RT-KTq4_pGDR11           RYVPDLEARVKSVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLEEMGA
RT-KTq2_pGDR11           RYVPDLEARVKSVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLEEMGA
KTq_M747K_pGDR11         RYVPDLEARVKSVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLEEMGA
KTq_wt_pGDR11            RYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGA
RT-KTq1_pGDR11           RYVPDLEARVKSVREAAERKAFNMPVQGTAADLMKLAMVKLFPRLEEMGA

KTq_M1/M747K_pGDR11      RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
KTq_M1_pGDR11            RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
RT-KTq3_pGDR11           RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
RT-KTq4_pGDR11           RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
RT-KTq2_pGDR11           RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
KTq_M747K_pGDR11         RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
KTq_wt_pGDR11            RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
RT-KTq1_pGDR11           RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
```

```
KTq_M1/M747K_pGDR11      LSAKEKA
KTq_M1_pGDR11            LSAKEKA
RT-KTq3_pGDR11           LSAKEKA
RT-KTq4_pGDR11           LSAKEKA
RT-KTq2_pGDR11           LSAKEKA
KTq_M747K_pGDR11         LSAKEKA
KTq_wt_pGDR11            LSAKEKA
RT-KTq1_pGDR11           LSAKEKA
```

2.4.2 Sequence of *Taq* Wild-type

The initial methionine is shown in green, the His-Tag is depicted in orange, the *Taq* domain is shown in black. Additional amino acids resulting from the expression vector pGDR11 are shown in blue.

```
MRGSHHHHHHTDPHAMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRHEA
YGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGYLITPAWLW
EKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKR
REPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLA
KDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHME
ATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRE
LTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLI
RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVE
TLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPL
AVPLEVEVGIGEDWLSAKEKA
```

3. Abbreviations

Amino acid nomenclature

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartate	P	Pro	Proline
E	Glu	Glutamate	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophane
L	Leu	Leucine	Y	Tyr	Tyrosine

^{32}P	Phosphorus isotope ^{32}P
A	Adenine
AMV	Avian myeloblastosis virus
APS	Ammonium persulfate
ATP	Adenosine 5'-triphosphate
bp	base pairs
BSA	Bovine serum albumin
C	Cytosine
CTP	Cytidine 5'-triphosphate
dATP	2'-Desoxyadenosine 5'-triphosphate
dCTP	2'-Desoxycytidine 5'-triphosphate
dGTP	2'-Desoxyguanosine 5'-triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	2'-Desoxyribonucleoside 5'-triphosphate
DTT	1,4-Dithiotreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
GTP	Guanosine 5'-triphosphate
h	hour(s)
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	kilobases
kDa	kilodaltons
KlenTaq	Klenow fragment of *Taq* DNA polymerase
KTq	*KlenTaq*
LB	Lysogeny Broth
M	Molar [mol/L]
min	minute(s)
MoMLV	Moloney murine leukemia virus
MS	Mass spectrometry
Ni-IDA	Nickel iminodiacetic acid
nt	nucleotide(s)
NTP	Ribonucleoside 5'-triphosphate

OD_{600}	Optical density (λ = 600 nm)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
RT	Reverse transcription
RNA	Ribonucleic acid
RQF	Rapid Quench Flow
rpm	Rotations per minute
s	second(s)
SDS	Sodium dodecyl sulfate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
t	time
T	Thymine
TAE	Tris-acetate-EDTA
Taq	Thermus aquaticus
TBE	Tris-borate-EDTA
TEMED	N, N, N', N'-Tetramethylethylenediamine
TLC	Thin layer chromatography
TLS	Translesion synthesis
Tris	Tris(hydroxymethyl)aminomethane
TTP	Thymidine 5'-triphosphate
U	Uracil, units
UDG	Uracil-DNA glycosylase
UTP	Uridine 5'-triphosphate
Wt	Wildtype
UV	Ultraviolet

VIII. References

1. Dahm R (2005) Friedrich Miescher and the discovery of DNA. *Dev. Biol.* **278**, 274.
2. Avery OT, Macleod CM and McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J. Exp. Med.* **79**, 137.
3. Chargaff E et al. (1950) Nucleotide composition of pentose nucleic acids from yeast and mammalian tissues. *J. Biol. Chem.* **186**, 51.
4. Watson JD and Crick FH (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964.
5. Watson JD and Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737.
6. Franklin RE and Gosling RG (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172**, 156.
7. Drew HR et al. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2179.
8. Arnott S et al. (1980) Left-handed DNA helices. *Nature* **283**, 743.
9. Ghosh A and Bansal M (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 620.
10. Basham B, Schroth GP and Ho PS (1995) An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6464.
11. Lu XJ, Shakked Z and Olson WK (2000) A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819.
12. Lodish H, Berk A, Zipursky SL, Molceular Cell Biology, Edn. 4th. (W. H. Freeman, New York, 2000).
13. Fiers W et al. (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**, 500.
14. Bustin SA and Mueller R (2005) Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin. Sci. (Lond.)* **109**, 365.
15. Lehman IR, Bessman MJ, Simms ES and Kornberg A (1958) Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. *J. Biol. Chem.* **233**, 163.
16. Bessman MJ, Lehman IR, Simms ES and Kornberg A (1958) Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction. *J. Biol. Chem.* **233**, 171.
17. Beese LS, Friedman JM and Steitz TA (1993) Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32**, 14095.
18. Bebenek K and Kunkel TA (2004) Functions of DNA polymerases. *Adv. Protein Chem.* **69**, 137.
19. Huebscher US, Spadari S, Villani G, Maga G, DNA polymerases: Discovery, Characterization and Functions in Cellular DNA Transactions. (World Scientific Publishing Co. Pte. Ltd., 2010).
20. Crick FH (1958) On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138.
21. Temin HM and Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211.

22. Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209.

23. Joyce CM and Steitz TA (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63**, 777.

24. Huang H, Chopra R, Verdine GL and Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669.

25. Ding J et al. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 A resolution. *J. Mol. Biol.* **284**, 1095.

26. Sarafianos SG et al. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449.

27. Brautigam CA and Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54.

28. Rothwell PJ, Mitaksov V and Waksman G (2005) Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: implications for the molecular basis of fidelity in DNA polymerases. *Mol. Cell* **19**, 345.

29. Rothwell PJ and Waksman G (2005) Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71**, 401.

30. Li Y, Korolev S and Waksman G (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514.

31. Steitz TA, Smerdon SJ, Jager J and Joyce CM (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022.

32. Steitz TA (1998) A mechanism for all polymerases. *Nature* **391**, 231.

33. Castro C et al. (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212.

34. Boosalis MS, Petruska J and Goodman MF (1987) DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *J. Biol. Chem.* **262**, 14689.

35. Randall SK et al. (1987) Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. Biol. Chem.* **262**, 6864.

36. Johnson KA (1993) Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685.

37. Barman TE et al. (2006) The identification of chemical intermediates in enzyme catalysis by the rapid quench-flow technique. *Cell. Mol. Life Sci.* **63**, 2571.

38. Johnson KA (1998) Advances in transient-state kinetics. *Curr. Opin. Biotechnol.* **9**, 87.

39. Divita G et al. (1993) Kinetics of interaction of HIV reverse transcriptase with primer/template. *Biochemistry* **32**, 7966.

40. Rittinger K, Divita G and Goody RS (1995) Human immunodeficiency virus reverse transcriptase substrate-induced conformational changes and the mechanism of inhibition by nonnucleoside inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8046.

41. Jiricny J (2006) The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7**, 335.

42. Maga G and Hubscher U (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J. Cell Sci.* **116**, 3051.

43. Davey MJ, Jeruzalmi D, Kuriyan J and O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. *Nat. Rev. Mol. Cell Biol.* **3**, 826.

44. Fanning E, Klimovich V and Nager AR (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* **34**, 4126.

45. Shevelev IV and Hubscher U (2002) The 3' 5' exonucleases. *Nat. Rev. Mol. Cell Biol.* **3**, 364.

46. Kunkel TA (2004) DNA replication fidelity. *J. Biol. Chem.* **279**, 16895.

47. Moran S, Ren RX and Kool ET (1997) A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10506.

48. Moran S, Ren RX, Rumney S and Kool ET (1997) Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.* **119**, 2056.

49. Liu D, Moran S and Kool ET (1997) Bi-stranded, multisite replication of a base pair between difluorotoluene and adenine: confirmation by 'inverse' sequencing. *Chem. Biol.* **4**, 919.

50. Kool ET, Morales JC and Guckian KM (2000) Mimicking the Structure and Function of DNA: Insights into DNA Stability and Replication. *Angew. Chem. Int. Ed. Engl.* **39**, 990.

51. Goodman MF (1997) Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10493.

52. Kool ET (2002) Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191.

53. Summerer D and Marx A (2001) DNA Polymerase Selectivity: Sugar Interactions Monitored with High-Fidelity Nucleotides. *Angew. Chem. Int. Ed. Engl.* **40**, 3693.

54. Xia S, Eom SH, Konigsberg WH and Wang J (2012) Structural basis for differential insertion kinetics of dNMPs opposite a difluorotoluene nucleotide residue. *Biochemistry* **51**, 1476.

55. Kool ET and Sintim HO (2006) The difluorotoluene debate--a decade later. *Chem. Commun. (Camb.)*, 3665.

56. Joyce CM and Benkovic SJ (2004) DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43**, 14317.

57. Rothwell PJ and Waksman G (2007) A pre-equilibrium before nucleotide binding limits fingers subdomain closure by Klentaq1. *J. Biol. Chem.* **282**, 28884.

58. Joyce CM et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* **47**, 6103.

59. Rothwell PJ et al. (2013) dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J. Biol. Chem.* **288**, 13575.

60. Wu EY and Beese LS (2011) The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286**, 19758.

61. Betz K et al. (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat. Chem. Biol.* **8**, 612.

62. Korolev S et al. (1995) Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-A resolution: structural basis for thermostability. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9264.

63. Patel PH et al. (2001) Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.* **308**, 823.

64. Klenow H, Overgaard-Hansen K and Patkar SA (1971) Proteolytic cleavage fo native DNA polymerase into two different catalytic fragments. Influence of assay condtions on the change of exonuclease activity and polymerase activity accompanying cleavage. *Eur. J. Biochem.* **22**, 371.

65. Setlow P, Brutlag D and Kornberg A (1972) Deoxyribonucleic acid polymerase: two distinct enzymes in one polypeptide. I. A proteolytic fragment containing the polymerase and 3' leads to 5' exonuclease functions. *J. Biol. Chem.* **247**, 224.

66. Barnes WM (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**, 29.

67. Li Y, Kong Y, Korolev S and Waksman G (1998) Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci.* **7**, 1116.

68. Li Y, Mitaxov V and Waksman G (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9491.

69. Obeid S et al. (2010) Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21327.

70. Bergen K et al. (2012) Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates. *J. Am. Chem. Soc.* **134**, 11840.

71. Obeid S et al. (2012) Interactions of non-polar and "Click-able" nucleotides in the confines of a DNA polymerase active site. *Chem. Commun. (Camb.)* **48**, 8320.

72. Obeid S et al. (2010) Replication through an abasic DNA lesion: structural basis for adenine selectivity. *EMBO J.* **29**, 1738.

73. Obeid S et al. (2011) Learning from directed evolution: Thermus aquaticus DNA polymerase mutants with translesion synthesis activity. *ChemBioChem* **12**, 1574.

74. Obeid S, Welte W, Diederichs K and Marx A (2012) Amino acid templating mechanisms in selection of nucleotides opposite abasic sites by a family a DNA polymerase. *J. Biol. Chem.* **287**, 14099.

75. De Bont R and van Larebeke N (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169.

76. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709.

77. Lindahl T and Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610.

78. Loeb LA and Preston BD (1986) Mutagenesis by apurinic/apyrimidinic sites. *Annu. Rev. Genet.* **20**, 201.

79. Lindahl T (1982) DNA repair enzymes. *Annu. Rev. Biochem.* **51**, 61.

80. Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71**, 17.

81. Hubscher U, Maga G and Spadari S (2002) Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71**, 133.

82. Taylor JS (2002) New structural and mechanistic insight into the A-rule and the instructional and non-instructional behavior of DNA photoproducts and other lesions. *Mutat. Res.* **510**, 55.

83. Goodman MF, Cai H, Bloom LB and Eritja R (1994) Nucleotide insertion and primer extension at abasic template sites in different sequence contexts. *Ann. N. Y. Acad. Sci.* **726**, 132.

84. Shibutani S, Takeshita M and Grollman AP (1997) Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J. Biol. Chem.* **272**, 13916.

85. Ling H, Boudsocq F, Woodgate R and Yang W (2004) Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts. *Mol. Cell* **13**, 751.

86. Hogg M, Wallace SS and Doublie S (2004) Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J.* **23**, 1483.

87. Freisinger E, Grollman AP, Miller H and Kisker C (2004) Lesion (in)tolerance reveals insights into DNA replication fidelity. *EMBO J.* **23**, 1494.

88. Fiala KA and Suo Z (2007) Sloppy bypass of an abasic lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem.* **282**, 8199.

89. Nair DT et al. (2009) DNA synthesis across an abasic lesion by human DNA polymerase iota. *Structure* **17**, 530.

90. Wang F and Yang W (2009) Structural insight into translesion synthesis by DNA Pol II. *Cell* **139**, 1279.

91. Choi JY et al. (2010) Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha, delta, eta, iota, kappa, and REV1. *J. Mol. Biol.* **404**, 34.

92. Nair DT et al. (2011) DNA synthesis across an abasic lesion by yeast REV1 DNA polymerase. *J. Mol. Biol.* **406**, 18.

93. Berdis AJ (2001) Dynamics of translesion DNA synthesis catalyzed by the bacteriophage T4 exonuclease-deficient DNA polymerase. *Biochemistry* **40**, 7180.

94. Beard WA et al. (2009) DNA polymerase beta substrate specificity: side chain modulation of the "A-rule". *J. Biol. Chem.* **284**, 31680.

95. Mozzherin DJ et al. (1997) Proliferating cell nuclear antigen promotes DNA synthesis past template lesions by mammalian DNA polymerase delta. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6126.

96. Cai H, Bloom LB, Eritja R and Goodman MF (1993) Kinetics of deoxyribonucleotide insertion and extension at abasic template lesions in different sequence contexts using HIV-1 reverse transcriptase. *J. Biol. Chem.* **268**, 23567.

97. Sagher D and Strauss B (1983) Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: uniqueness of adenine nucleotides. *Biochemistry* **22**, 4518.

98. Schaaper RM, Kunkel TA and Loeb LA (1983) Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 487.

99. Lawrence CW, Borden A, Banerjee SK and LeClerc JE (1990) Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res.* **18**, 2153.

100. Avkin S, Adar S, Blander G and Livneh Z (2002) Quantitative measurement of translesion replication in human cells: evidence for bypass of abasic sites by a replicative DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3764.

101. Strauss BS (2002) The "A" rule revisited: polymerases as determinants of mutational specificity. *DNA Repair (Amst)* **1**, 125.

102. Pages V, Johnson RE, Prakash L and Prakash S (2008) Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1170.

103. Matray TJ and Kool ET (1999) A specific partner for abasic damage in DNA. *Nature* **399**, 704.

104. Reineks EZ and Berdis AJ (2004) Evaluating the contribution of base stacking during translesion DNA replication. *Biochemistry* **43**, 393.

105. Seki M et al. (2004) High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J.* **23**, 4484.

106. Zahn KE, Wallace SS and Doublie S (2011) DNA polymerases provide a canon of strategies for translesion synthesis past oxidatively generated lesions. *Curr. Opin. Struct. Biol.* **21**, 358.

107. Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**, 9677.

108. Zhang X, Lee I and Berdis AJ (2004) Evaluating the contributions of desolvation and base-stacking during translesion DNA synthesis. *Org. Biomol. Chem.* **2**, 1703.

109. Xia S et al. (2012) Contribution of partial charge interactions and base stacking to the efficiency of primer extension at and beyond abasic sites in DNA. *Biochemistry* **51**, 4922.

110. Guckian KM et al. (2000) Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J. Am. Chem. Soc.* **122**, 2213.

111. Zahn KE, Belrhali H, Wallace SS and Doublie S (2007) Caught bending the A-rule: crystal structures of translesion DNA synthesis with a non-natural nucleotide. *Biochemistry* **46**, 10551.

112. Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366.

113. Eid J et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133.

114. Kermekchiev MB, Kirilova LI, Vail EE and Barnes WM (2009) Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res.* **37**, e40.

115. Saiki RK et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487.

116. Sanger F et al. (1974) Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **90**, 315.

117. Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463.

118. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3.

119. Shi MM (2001) Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* **47**, 164.

120. Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* **2**, 930.

121. Ling MM and Robinson BH (1997) Approaches to DNA mutagenesis: an overview. *Anal. Biochem.* **254**, 157.

122. Labrou NE (2010) Random mutagenesis methods for in vitro directed enzyme evolution. *Curr. Protein Pept. Sci.* **11**, 91.

123. Seo YJ et al. (2011) Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. *J. Am. Chem. Soc.* **133**, 19878.

124. Ellington AD and Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818.

125. Pinheiro VB et al. (2012) Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341.

126. Powell KA et al. (2001) Directed Evolution and Biocatalysis. *Angew. Chem. Int. Ed. Engl.* **40**, 3948.

127. Davids T, Schmidt M, Bottcher D and Bornscheuer UT (2013) Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* **17**, 215.

128. Tracewell CA and Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13**, 3.

129. Cadwell RC and Joyce GF (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28.

130. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389.

131. Zhao H et al. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258.

132. Gloeckner C, Kranaster R and Marx A (2010) Directed evolution of DNA polymerases: construction and screening of DNA polymerase mutant libraries. *Curr. Protoc. ChemBiol* **2**, 89.

133. Tawfik DS and Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652.

134. Jermutus L et al. (2001) Tailoring in vitro evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 75.

135. Lin H and Cornish VW (2002) Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed. Engl.* **41**, 4402.

136. Ghadessy FJ, Ong JL and Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4552.

137. Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10747.

138. Shao Z, Zhao H, Giver L and Arnold FH (1998) Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res.* **26**, 681.

139. Zhao H and Arnold FH (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307.

140. Ricchetti M and Buc H (1993) E. coli DNA polymerase I as a reverse transcriptase. *EMBO J.* **12**, 387.

141. Wohrl BM, Krebs R, Goody RS and Restle T (1999) Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J. Mol. Biol.* **292**, 333.

142. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25**, 169.

143. Gerard GF et al. (2002) The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res.* **30**, 3118.

144. Arezi B and Hogrefe H (2009) Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* **37**, 473.

145. Yasukawa K, Mizuno M, Konishi A and Inouye K (2010) Increase in thermal stability of Moloney murine leukaemia virus reverse transcriptase by site-directed mutagenesis. *J. Biotechnol.* **150**, 299.

146. Konishi A, Yasukawa K and Inouye K (2012) Improving the thermal stability of avian myeloblastosis virus reverse transcriptase alpha-subunit by site-directed mutagenesis. *Biotechnol. Lett.* **34**, 1209.

147. Myers TW and Gelfand DH (1991) Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase. *Biochemistry* **30**, 7661.

148. Sauter KB and Marx A (2006) Evolving thermostable reverse transcriptase activity in a DNA polymerase scaffold. *Angew. Chem. Int. Ed. Engl.* **45**, 7633.

149. Schonbrunner NJ et al. (2006) Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'-5' exonuclease activity. *Biochemistry* **45**, 12786.

150. Kranaster R et al. (2010) One-step RNA pathogen detection with reverse transcriptase activity of a mutated thermostable Thermus aquaticus DNA polymerase. *Biotechnol. J.* **5**, 224.

151. Moser MJ et al. (2012) Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One* **7**, e38371.

152. Sellner LN, Coelen RJ and Mackenzie JS (1992) Reverse transcriptase inhibits Taq polymerase activity. *Nucleic Acids Res.* **20**, 1487.

153. Fehlmann C, Krapf R and Solioz M (1993) Reverse transcriptase can block polymerase chain reaction. *Clin. Chem.* **39**, 368.

154. Chumakov KM (1994) Reverse transcriptase can inhibit PCR and stimulate primer-dimer formation. *PCR Methods Appl.* **4**, 62.

155. Ong JL et al. (2006) Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* **361**, 537.

156. Vichier-Guerre S et al. (2006) A population of thermostable reverse transcriptases evolved from Thermus aquaticus DNA polymerase I by phage display. *Angew. Chem. Int. Ed. Engl.* **45**, 6133.

157. Shandilya H et al. (2004) Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles* **8**, 243.

158. Jones MD and Foulkes NS (1989) Reverse transcription of mRNA by Thermus aquaticus DNA polymerase. *Nucleic Acids Res.* **17**, 8387.

159. Gloeckner C, Sauter KB and Marx A (2007) Evolving a thermostable DNA polymerase that amplifies from highly damaged templates. *Angew. Chem. Int. Ed. Engl.* **46**, 3115.

160. Kunkel TA and Bebenek K (2000) DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497.

161. Trostler M et al. (2009) Discrimination between right and wrong purine dNTPs by DNA polymerase I from Bacillus stearothermophilus. *Biochemistry* **48**, 4633.

162. Blatter N, Prokup A, Deiters A and Marx A (2014) Modulating the pK of a Tyrosine in KlenTaq DNA Polymerase that Is Crucial for Abasic Site Bypass by in Vivo Incorporation of a Non-canonical Amino Acid. *ChemBioChem*. doi: 10.1002/ cbic.201400051

163. Thorson JS, Chapman E, Murphy EC and Schultz PG (1995) Linear Free energy Analysis of Hydrogen Bonding in Proteins. *J. Am. Chem. Soc.* **117**, 1157.

164. Brooks B, Phillips RS and Benisek WF (1998) High-efficiency incorporation in vivo of tyrosine analogues with altered hydroxyl acidity in place of the catalytic tyrosine-14 of Delta 5-3-ketosteroid isomerase of Comamonas (Pseudomonas) testosteroni: effects of the modifications on isomerase kinetics. *Biochemistry* **37**, 9738.

165. Seyedsayamdost MR, Reece SY, Nocera DG and Stubbe J (2006) Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* **128**, 1569.

166. Seyedsayamdost MR et al. (2006) pH Rate profiles of F_nY_{356}-R2s (n = 2, 3, 4) in Escherichia coli ribonucleotide reductase: evidence that Y356 is a redox-active amino acid along the radical propagation pathway. *J. Am. Chem. Soc.* **128**, 1562.

167. Takeshita M et al. (1987) Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* **262**, 10171.

168. Stuart GR and Chambers RW (1987) Synthesis and properties of oligodeoxynucleotides with an AP site at a preselected position. *Nucleic Acids Res.* **15**, 7451.

169. Bailly V and Verly WG (1988) Possible roles of beta-elimination and delta-elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells. *Biochem. J.* **253**, 553.

170. Bailly V, Derydt M and Verly WG (1989) Delta-elimination in the repair of AP (apurinic/apyrimidinic) sites in DNA. *Biochem. J.* **261**, 707.

171. Masuda K et al. (2007) DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J. Biol. Chem.* **282**, 17387.

172. Suzuki M, Baskin D, Hood L and Loeb LA (1996) Random mutagenesis of Thermus aquaticus DNA polymerase I: concordance of immutable sites in vivo with the crystal structure. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9670.

173. Bell JB, Eckert KA, Joyce CM and Kunkel TA (1997) Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272**, 7345.

174. Minnick DT et al. (1999) Side chains that influence fidelity at the polymerase active site of Escherichia coli DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067.

175. Peliska JA and Benkovic SJ (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science* **258**, 1112.

176. Fiala KA et al. (2007) Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase. *J. Mol. Biol.* **365**, 590.

177. Blackburn EH (1984) The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* **53**, 163.

178. Alt FW and Baltimore D (1982) Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4118.

179. Blatter N et al. (2013) Structure and Function of an RNA-Reading Thermostable DNA Polymerase. *Angew. Chem. Int. Ed. Engl.* **52**, 11935.

180. Patrick WM, Firth AE and Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.* **16**, 451.

181. Joern J, Directed Evolution Library Creation, Vol. 231, Page 85 (eds. Arnold F & Georgiou G, Humana Press, 2003).

182. Schlee S et al. (2009) Activation of anthranilate phosphoribosyltransferase from Sulfolobus solfataricus by removal of magnesium inhibition and acceleration of product release. *Biochemistry* **48**, 5199.

183. Sambrook JR, Russell DW, Molecular Cloning: A Laboratory Manual. The Third edition. (Cold Spring Harbor Laboratory Press, New York, 2001).

184. Peist R et al. (1997) Characterization of the aes gene of Escherichia coli encoding an enzyme with esterase activity. *J. Bacteriol.* **179**, 7679.

185. Rudinger NZ, Kranaster R and Marx A (2007) Hydrophobic amino acid and single-atom substitutions increase DNA polymerase selectivity. *Chem. Biol.* **14**, 185.

186. Reetz MT (2013) The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed. Engl.* **52**, 2658.

187. Ghedin E et al. (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162.

188. Palese P (2004) Influenza: old and new threats. *Nat. Med.* **10**, S82.

189. Kumar S and Henrickson KJ (2012) Update on influenza diagnostics: lessons from the novel H1N1 influenza A pandemic. *Clin. Microbiol. Rev.* **25**, 344.

190. Kutyavin IV et al. (2000) 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res.* **28**, 655.

191. Breen MS et al. (2012) Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535.

192. Wagner GP (2012) Genetics: The inner life of proteins. *Nature* **490**, 493.

193. Jozwiakowski SK and Connolly BA (2011) A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *ChemBioChem* **12**, 35.

194. Joyce CM (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1619.

195. Patel PH and Loeb LA (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* **275**, 40266.

196. Ogawa M et al. (2001) Enhanced ribonucleotide incorporation by an O-helix mutant of Thermus aquaticus DNA polymerase I. *Mutat. Res.* **485**, 197.

197. Nick McElhinny SA et al. (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **6**, 774.

198. Brown JA and Suo Z (2011) Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry* **50**, 1135.

199. Xia G et al. (2002) Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6597.

200. Jacobo-Molina A et al. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6320.

201. Das D and Georgiadis MM (2004) The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure* **12**, 819.

202. Nowak E et al. (2013) Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res.* **41**, 3874.

203. Li Y and Waksman G (2001) Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP- trapped ternary complex of Klentaq1: insights into nucleotide incorporation and selectivity. *Protein Sci.* **10**, 1225.

204. Holzberger B, Rubini M, Moller HM and Marx A (2010) A highly active DNA polymerase with a fluorous core. *Angew. Chem. Int. Ed. Engl.* **49**, 1324.

205. Kabsch W (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125.

206. Adams PD et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213.

207. Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486.

208. McCoy AJ et al. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658.

209. Fedoroff O, Salazar M and Reid BR (1993) Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J. Mol. Biol.* **233**, 509.

210. Kranaster R and Marx A (2009) Taking fingerprints of DNA polymerases: multiplex enzyme profiling on DNA arrays. *Angew. Chem. Int. Ed. Engl.* **48**, 4625.

211. Cozens C et al. (2012) A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8067.

212. Eom SH, Wang J and Steitz TA (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **382**, 278.

213. Lyamichev V, Brow MA, Varvel VE and Dahlberg JE (1999) Comparison of the 5' nuclease activities of Taq DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6143.

214. Fa M, Radeghieri A, Henry AA and Romesberg FE (2004) Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* **126**, 1748.

215. Ghadessy FJ et al. (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* **22**, 755.

216. Henry AA and Romesberg FE (2005) The evolution of DNA polymerases with novel activities. *Curr. Opin. Biotechnol.* **16**, 370.

217. Paabo S et al. (2004) Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645.

218. McDonald JP et al. (2006) Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res.* **34**, 1102.

219. d'Abbadie M et al. (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25**, 939.

220. Wang W, Wu EY, Hellinga HW and Beese LS (2012) Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.* **287**, 28215.

221. Patel PH et al. (2001) A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.* **276**, 5044.

222. Astatke M, Ng K, Grindley ND and Joyce CM (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3402.

223. Holzberger B et al. (2012) Structural insights into the potential of 4-fluoroproline to modulate biophysical properties of proteins. *Chem. Sci.* **3**, 2924.

224. Dang C and Jayasena SD (1996) Oligonucleotide inhibitors of Taq DNA polymerase facilitate detection of low copy number targets by PCR. *J. Mol. Biol.* **264**, 268.

225. Chen VB et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12.

226. Schrodinger (2010) LLC.

Eidesstaatliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, im Dezember 2013

Nina Blatter