Structural insights into DNA polymerases encountering aberrant substrates

Dissertation

Zur Erlangung des Doktorgrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Sektion

der

Universität Konstanz

Fachbereich Biologie

Vorgelegt von

Dipl. Biol. Konrad Bergen

aus Stuttgart

Juli 2014

Prüfungsvorsitz: Prof. Dr. Hans-Jürgen Appell

1. Referent: Prof. Dr. Wolfram Welte

2. Referent: Prof. Dr. Andreas Marx

Tag der mündlichen Prüfung: 05.12.2014

Table of Contents

Zusammenfassung

Der Reproduktionszyklus allen Lebens beruht auf der zuverlässigen und genauen Verdopplung des genetischen Materials; dies wird von Enzymen bewerkstelligt, die als DNA abhängige DNA Polymerasen bezeichnet werden. Diese Enzyme katalysieren den Einbau von 2'-deoxy-ribonukleosid Triphosphaten (dNTP) komplementär zum parentalen Einzelstrang der Desoxyribonukleinsäure (DNS), indem das 5' Phosphat des einzubauenden dNTPs unter Abspaltung von Pyrophosphat kovalent mit der 3' Hydroxygruppe des wachsenden Primerstrangs verbunden wird. *In vivo* ist dieser Prozess nahezu fehlerfrei (eine Fehleinbaute auf 10^9-10^{10} Basenpaare). Auch *in vitro* können sehr geringe Fehlerraten (im Bereich von 10^{-6} -10^{-7}) beobachtet werden, abhängig vom jeweiligen Enzym.

Diese Eigenschaften erlauben den Einsatz in biotechnischen Prozessen, die auf exakten Kopien der Ursprungsmaterials beruhen. Zusätzlich sind einige dieser Enzyme in der Lage modifizierte und unnatürliche dNTPs zu prozessieren.

Die hier vorgestellten Teile der Arbeit wurden im Laufe der Dissertation in drei Publikationen veröffentlicht. Die Arbeit "KlenTaq caught incorporating C5 and 7-deaza modified nucleotides" konzentriert sich auf strukturelle Daten, die aus der Kristallisation und Röntgenstrukturanalyse ternärer Komplexe des Klenow-/Stoffel-Fragments der DNA Polymerase I aus *Thermus aquaticus* (*KlenTaq*) mit modifizierten dNTPS gewonnen wurden. Die verwendeten dNTP Analoga (von A. Baccaro und A. Steck zur Barcodierung von Oligonukleotiden entwickelt) zeigten eine herausragende Akzeptanz durch die Polymerase, daher wurden kristallographische Studien durchgeführt, um die strukturelle Basis aufzuklären. Es konnte gezeigt werden, dass Typ und Länge des Linkers an der Nukleobase durch Wechselwirkungen zwischen Modifikation und Enzym einen entscheidenden Einfluss auf die Akzeptanz durch die Polymerase haben. In einem weiteren Manuskript ("Structures of KOD and 9° North Polymerases Complexed with Primer Template Duplex") werden binäre Strukturen der B-Familien Polymerasen KOD und 9°N im replikativen Modus beschrieben; sowie deren Interaktion zwischen Nukleinsäure und den Enzymen. Dies erlaubte Rückschlüsse auf die ausgeprägte Fähigkeit dieser Proteine hochmodifizierte DNS zu synthetisieren.

Den letzten Teil der Arbeit stellt eine kristallographische Studie einer *KlenTaq* Mehrfachmutante. Die von N. Blatter gefundene Mutante weist eine starke reverse

Transkriptions-Aktivität auf. Über einen strukturellen Vergleich der Polymerase im Komplex mit doppelsträngiger DNA sowie einem DNA-RNA-Hybriden konnte der Einfluss der Mutationen auf die notwendigen strukturellen Anpassungen im Manuskript "Structure and Function of a thermostable DNA polymerase reading RNA" aufgezeigt werden.

Zusammenfassend ermöglichen strukturelle Einblicke in die Prozessierung natürlicher und anomaler Substrate durch DNA Polymerasen eine rationale Herangehensweise in der Wahl des Enzyms (und seiner Mutanten) sowie dem Design der Modifikation des Substrates, um wertvolle Werkzeuge für die Molekularbiologie und die klinische Dignostik zu entwickeln.

Summary

The reproduction cycle of all life relies on faithful and accurate reproduction of the genome. This task is performed by the enzyme class of DNA-dependant DNA polymerases. These enzymes catalyze the incorporation of 2'-deoxy-ribonucleoside-triphosphates (dNTP) in complementary fashion to the parental single-stranded deoxyribo-nucleic acid (DNA) by linking the 5' phosphate of the incoming dNTP to the 3' hydroxy function of the growing oligomer, energized by the cleavage of pyrophosphate. This process is *in vivo* quite immaculate (1 error in 10^9-10^{10} base pairs). *In vitro*, error rates as low as 10^{-6} -10^{-7} are observed, depending on the polymerase

The very same abilities allow the application in biotechnology, relying on exact copies of parental material. Additionally, some of the enzymes are also capable of processing modified and non-natural dNTPs.

The studies presented in this thesis are described in three papers which were published during the time of the dissertation. The publication "KlenTaq caught incorporating C5 and 7-deaza modified nucleotides" focuses on structural data obtained from ternary complexes of the Klenow/Stoffel fragment of DNA Polymerase (DNA pol) I of *Thermus aquaticus* (*KlenTaq*) with modified dNTPs. The dNTP analogues used in this study (generated for DNA barcoding purposes as published by Baccaro and Steck) displayed excellent incorporation characteristics, so crystallographic studies were performed to elucidate the structural basis. We found that the type and length of the linker attached to the nucleobase moiety play a major role in the acceptance by DNA polymerase due to the interaction between enzyme and the analogue. A second manuscript ("Structures of KOD and 9° North Polymerases Complexed with Primer Template Duplex") describes structures of B-family DNA pols in a binary replicating complex and the interactions between enzyme and DNA, allowing suggestions on the distinct generation of highly modified DNA by these enzymes. Finally we performed a crystallographic study on a *KlenTaq* mutant capable of reverse transcription (generated by N. Blatter). In this manuscript ("Structure and Function of a thermostable DNA polymerase reading RNA"), we describe the adaptation of a DNA dependent DNA polymerase mutant binding an RNA/DNA hybrid.

Overall, structural insights on processing of natural and aberrant substrates by DNA polymerase allow a more rational approach in the choice of the enzyme (and its mutants) and the design of substrate modifications to generate valuable tools for molecular biology and clinical diagnostics.

1 Introduction

1.1 Macromolecular Crystallography

1.1.1 A short history of macromolecular crystallography

The in-depth visualization of biological macromolecules as well as other small molecules requires techniques that are able to resolve distances in the range of (the) covalently bonded atoms.

The ascension of crystallographic methods in this field began with experiments in the early days of the twentieth century. Back then, the newly discovered X-ray radiation enabled von Laue and his assistants, Knipping and Friedrich, to determine the wave character of the radiation by the interference with cristalline matter. This result was awarded with a Nobel prize in 1914. These findings allowed the formulation of "Bragg's Law" and guided the way to the new scientific fields of X-ray spectroscopy and X-ray crystallography, that allow insights in the spectral characteristics and the structural composition of (simple) compounds and resulted in another Nobel Prize for father and son Bragg.[1]

Though the successful crystallisation of biological macromolecules has been reported earlier[2], the first fruitful application of the basic idea on protein crystals could be achieved in the 1950s by Kendrew and Perutz, using the phase difference of native and heavy-metal substituted crystals, resulting in the structure determination of myoglobin/hemoglobin.[3,4]

This breakthrough enabled the scientific community to increase the count of solved structure. With a rising number of available structures, molecular replacement was developed as an easy and pure computational method to solve the phase problem also at lower resolutions.[5] Increasing computing power and access to particle accelerators and thereby tunable radiation sources enabled a fast growing number of solved protein structures, which are now made available for the public in the PDB.[6]

Today, the PDB contains around 90000 solved structures based on X-ray scattering.

1.2 Biologic polymers

The living cell contains three forms of informationally connected macromolecules, (deoxy-) ribonucleic acids (DNA and RNA) and proteins. The "central dogma of molecular biology", as hypothesised by Crick, describes the informational flow and its directions in the cellular enviroment. The scheme of the possible transfers is depicted in Figure 1, and divided into general and special transfers. In the general case, the sequential information in DNA can be duplicated and transcribed into RNA. The informational room of RNA covers the translation into proteins. However, sequential information contained in a protein can not be translated back into nucleic acids nor transferred in a horizontal way (protein self-replication).

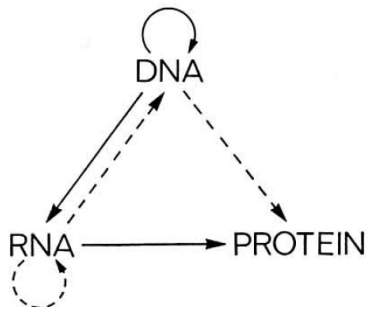


Figure 1 General and special transfers according to the central dogma of molecular biology. General transfers are marked by solid arrows, special transfers by dotted ones. Graphic adopted from [7]

The special transfers encompass the reverse transcription (RNA to DNA), RNA duplication and cell-free systems employing direct translation.[7]

1.2.1 DNA and its structure

Though the first mention of the deoxyribonucleic acid (DNA) appeared in the second half of the 19th century, the speculations on its role as hereditary material[8] were not proven before the mid of the next century by *Avery*.[9] A few years later, in 1953, Watson and Crick were able to propose a three-dimensional model of the DNA by combining the x-ray data of *Wilkins et al.* [10] and *Franklin et al.*[11] with Chargaffs finding of the ratio of nucleobases[12]. This discovery and its inherent properties laid the foundation of modern molecular biology.

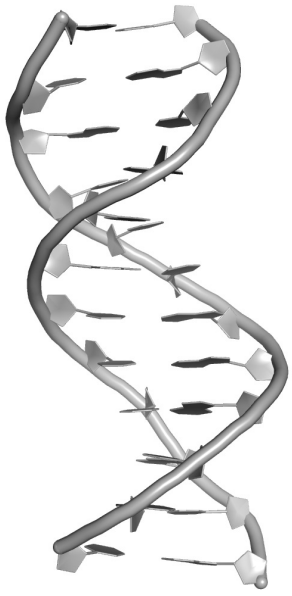


Figure 2 X-Ray structure of a B-DNA Dodecamer (PDB ID 1BNA)

The postulated model showed a right-handed double-helical arrangement (see Figure 2)[13,14] of a macromolecule composed of 4 building blocks (nucleoside monophosphates or nucleotides). In detail, the building blocks consist of a phosphate, a 2'-deoxy-ribose and a nucleobase linked by a N-glycosidic bond to the sugar. The bases can be further grouped into

pyrimidines (cytosine [C] and thymine [T]) and purines (adenine [A] and guanine [G], Figure 3A). The polymer is assembled by the formation of a phosphoester between the phosphate moiety and the 3' hydroxy function of the sugar. The interstrand interactions are the nucleobase-specific hydrogenbonding patterns between purines and pyrimidines, namely the bonding between G and C and A and T. As already concluded by Watson and Crick, if the sequence of one chain is known, the other one is automatically determined. And as Watson and Crick worded it: "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.*"[13]

It took another 3 decades to proof the model which was based on the fibre diffraction data, when *Wing et al.* were able to crystallise and solve the structures of short DNA sequences.[15] Depending on humidity and ionic strength of the environment, DNA can exist in distinct forms, exemplarily A, B and Z.

The most frequent species under physiological conditions is the B-form and in its structure closest to the Watson-Crick model, characterized by a right-handed double helix with ten nucleotides per turn and a pitch of 34 Å. The two grooves twisted around the surface of the molecule are shaped by the deoxyribose-phosphate backbone and display a width of 22 Å for the major and 12 Å for the minor groove, respectively.[15] The bases are located in the "center" of the helix and show no or little inclination. In this conformation the nucleobases are easily accessible.[16]

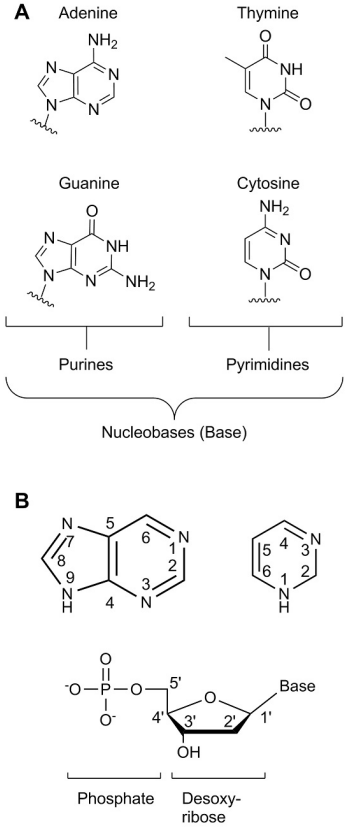


Figure 3 A: Chemical structures of the nucleobases. B: Numbering of the (hetero-) cycles for purines and pyrimidines and composition of a nucleoside monophosphate

The A form is a more condensed and broader (11bp per turn) right-handed helical arrangement with a deeper and less accessible major groove; decentered and more inclined bases are typical.

Z DNA is a special case, confined to a dinucleotide repeat of an alternating purine/pyrimidine sequence (typically G/C). Opposed to A- and B-forms it forms a left-handed helix with 6 dinucleotides per turn and alternating *sys*/*anti* conformation of nucleobase at the glycosyl bond. This results in the typical zig-zag pattern

of the backbone.In detail, the pitch difference between A- and B-form can be explained by the puckering of the ribofuranose moiety.[16]

1.2.2 Sugar puckering in DNA and RNA

Due to steric restraints the five-membered ring can not be planar, leading to an extrusion of 1 or 2 atoms out of the plane formed by C1', O1' and C4' (see Figure 3B and 4). The thereby formed puckering of the ribofuranose ring influences the interphosphate distance (see Figure 3). B-DNA is usually characterized by C2' endo conformations while A-DNA displays a C3' endo conformation (see Figure 4). The other highly abundant nucleic acid, RNA (ribonucleic acid), is differing from DNA in the presence of a hydroxyl group on C2' atom of the sugar and exchange from thymidine to uracil. This base differs from thymine in a non-methylated C5 (see figure 3A). Due to the steric restraints of the C2' hydroxy-function, RNAs are confined to the 3' endo puckering and thereby to an A-form only while DNA displays a higher flexibility in terms of the sugar puckering.[16] For further nomenclature and abbreviations concerning the conformational description of polynucleotides see[17] and references therein.

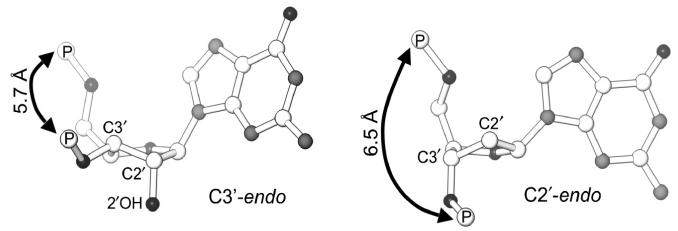


Figure 4 Left panel: C3'-endo sugar puckering (here RNA), typical for A-form nucleic acids. Right panel: C2' endo puckering. Graphic is taken from [18]

1.2.3 Proteins

Proteins consist of the 22 proteinogenic L-α amino acids (the 20 common ones plus the rare species selenocysteine and pyrrolysine), which are linked by peptide bonds as a linear polymer. Due to the planar character of the peptide bond, angle ω is restricted to 0° (cis) or 180° (trans) (see Figure 5). The secondary structures formed by the sequence of aminoacids are defined by the torsion angles around the C_α bonds to the amino function (φ) and to the carbonyl group (ψ) and thereby the possible hydrogen bridging. On an empirical base, the energetically favoured combinations can be plotted by the geometrical requirements of the different folds (Ramachandran plot, Figure 6).[19]

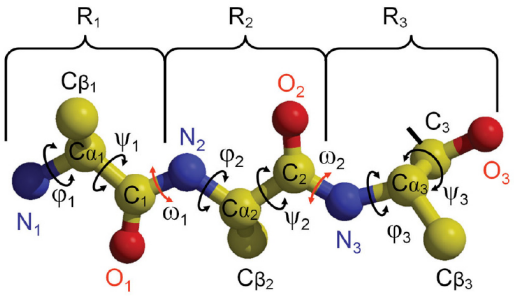


Figure 5 peptide backbone torsion angles, ψ and φ (black arrows) are restrained from van-der-Waals repulsion, defining by combination the secondary structure (see **Figure 6**). ω (red arrow) is restrained to values of 0° or 180° by the planar peptide bond. Graphic taken from[19]. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, C 2009-2014 Garland Science/Taylor & Francis LLC.

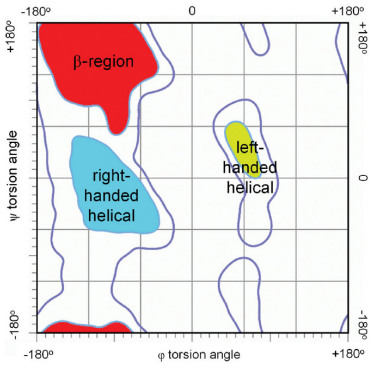


Figure 6 Ramachandran plot, energetically favored combinations of ψ and φ for secondary are plotted. Graphic adapted from[19]. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, C 2009-2014 Garland Science/Taylor & Francis LLC.

1.3 DNA Polymerases

1.3.1 DNA Polymerase I

The question "whether enzyme or selfreplication" put up by Watson&Crick[13,14], was answered in a series of three publications 3 years later. Kornberg and Coworkers described the purification of an enzyme capable to fullfil this task and the (bio-) synthesis of the substrate, deoxyribonucleoside triphosphates (dNTPs). Experimental data showed the generation of radioactively labelled DNA, the necessity of all 4 nucleotides and the presence of DNA, serving as template and primer in the polymerisation reaction.[20,21]. They also further characterized the "polymerase" with brominated, deaminated and methylated analogues of the natural dNTPs.[22] The identified enzyme showcased distinct 5'-3' and 3'-5' exonuclease activities besides the polymerase function. The 5'-3' exonuclease can be separated by a limited proteolytic digest, leaving a 68kDa C-terminal fragment containing the polymerase and 3'-5' exonuclease function, named Klenow-fragment (KF), in reference to its discoverer.[23,24] A further truncation of about 200 N-terminal residues diminishes the proofreading activity of the enzyme.[25]

1.3.2 DNA Polymerases: Families

Since Kornbergs discovery, a multitude of different polymerases have been identified over all realms. A first compilation and subsequent sequence analysis by *Delarue et al.* and *Braithwaite and Ito* lead to a classification along the homology to the respective *E. coli* genes. A comparison allowed the identification of three functional motifs which are conserved in all DNA-dependent DNA Polymerases (see Figure 7).[26-29]

Family A is subdivided in replicative and repair enzymes. The former contains proteins of bacteriophageal origin like T3, T5, T7 as well as the eukaryotic mitochondrial DNA polymerase γ, while the latter encompasses DNA polymerases I from e.g *E. coli* or *T. aquaticus*. Most members of the DNA polymerase I group exhibit 3'-5' proofreading and 5'-3 exonuclease activity, which is essential for the degradation of Okazaki fragments.[30]

Family B enzymes encompass the main eukaryotic replication enzymes α, δ, ε, archeal DNA polymerases, plasmid-encoded mitochondrial enzymes from fungi and

plants as well as polymerases from viruses and bacteriophages. A common feature is a very strong 3'-5' proofreading activity.

Family C polymerases are active as multisubunit complexes (10+) in the replication of the bacterial chromosomal DNA.[30]

Family D polymerases are restricted to the euryarchaeota and emergent subphyla. The active polymerase is thought to be heterodimeric with a slight homology of the smaller subunit to the exonuclease domain of family B enzymes. [31,32]

Family X members were identified by their homology to (human) DNA polymerase β. Enzymes of this class are involved in base excision repair, non-homologous end joining (NHEJ) and VDJ recombination. Additionally to the polymerase domain, a variety of accessory functions like protein-protein interaction domains or a dRP lyase activity are attributed to these proteins. [33]

Polymerases involved in translesion synthesis are grouped in Family Y. In consequence to the ability to bypass DNA damage, they show a low fidelity, a distributive synthesis behaviour and lack proofreading functions. This family contains besides *E. coli* polymerases IV and V enzymes from eubacterial, archeal and eukaryotic origin.[34,35]

While Family A, B, C, D, X, and Y DNA polymerases are solely DNA-dependant enzymes, reverse transcriptase (RT) are able to process DNA and RNA templates. A common feature among the retroviral reverse transcriptases is the RNAseH domain, a necessity for the cleavage of viral RNA during DNA synthesis.[30]

Enzymes used in this study belonged to Family A and B.

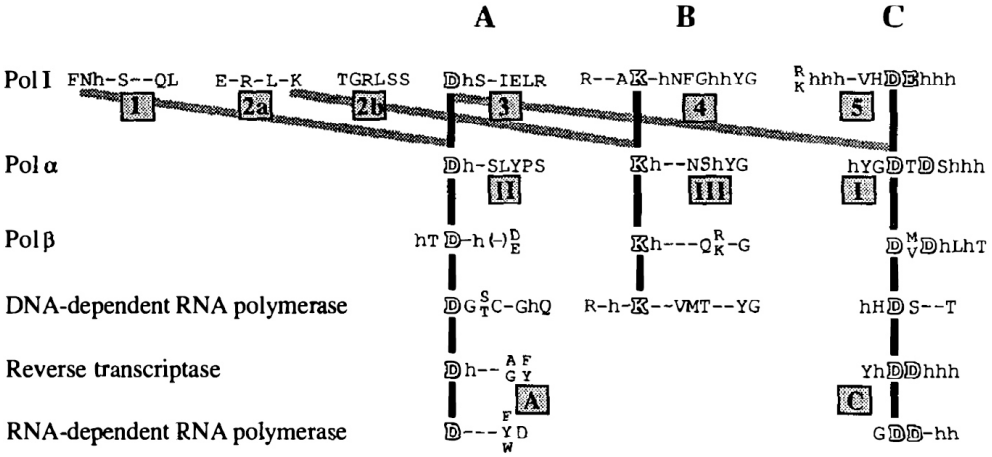


Figure 7 Alignment of conserved motifs throughout different polymerase families, conserved hydrophobic aminoacids are depicted as h. The graphic is adapted from ref [29].

1.3.3 Structural characterisation of DNA Polymerases

High yield expression systems of the so called Klenow fragment[36] (see Chapter 1.3.1) allowed crystallographic studies on this enzyme. A 3.3 Å structure revealed the expected fold in the two functional domains; exonuclease and polymerase.

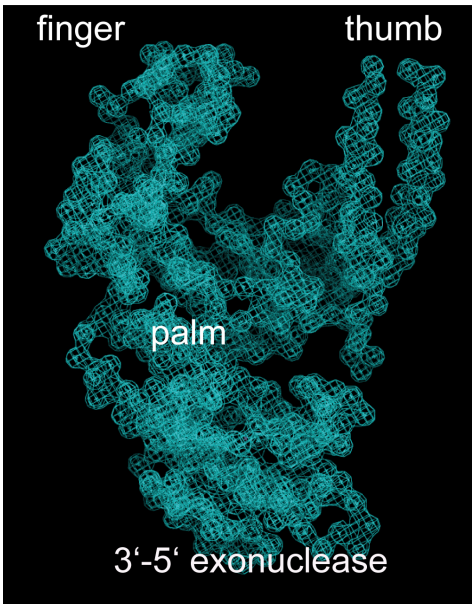


Figure 8 crystal structure of the Klenow fragment of E. coli DNA polymerase I. (PDB ID 1DPI)

Since the overall shape of the polymerase domain is reminiscent of a right hand; the domains were named analogously as thumb, palm and finger; the groove formed in between already pointed out to be the DNA binding site (see Fig. 8).[37] Binary and ternary structures of A-family polymerases revealed the structural movement upon binding of DNA and function of the subdomains.[38-40]

The right hand analogy holds true for A, B, (D), Y and RT families as they share common fold for the palm subdomain, consisting of 4 antiparallel β strands[37,41-43] while members of the C and X display a 5 stranded mixed β-sheet[44-46], giving the impression of a left-handed topology. A convergent development can be assumed since the orientation of the catalytic residues is on an opposed "handedness" in the palm superfamilies.[47]

1.3.4 Common Motifs in DNA Polymerases

Ternary structures of A-family polymerases allowed an interaction mapping of the previously mentioned motifs. Motif A is located in the palm domain and contains one of the active site aspartate residues. It forms contacts to the DNA primer and the Mg^{2+} ion.

Motif B is part of the finger/nucleotide selection domain and forms a helix which contacts the nascent basepair as well as hydrogen bonds to the incoming nucleoside triphosphate.

Motif C contains the second catalytic aspartate and is also located in the palm domain.

Motifs A and C are also conserved in DNA dependent RNA polymerases, RNA dependent RNA polymerases and DNA/RNA dependent DNA polymerases (reverse transcriptases).[50] The structural localisation of the motifs is shown in examples from families A, B and X in Figure 9.

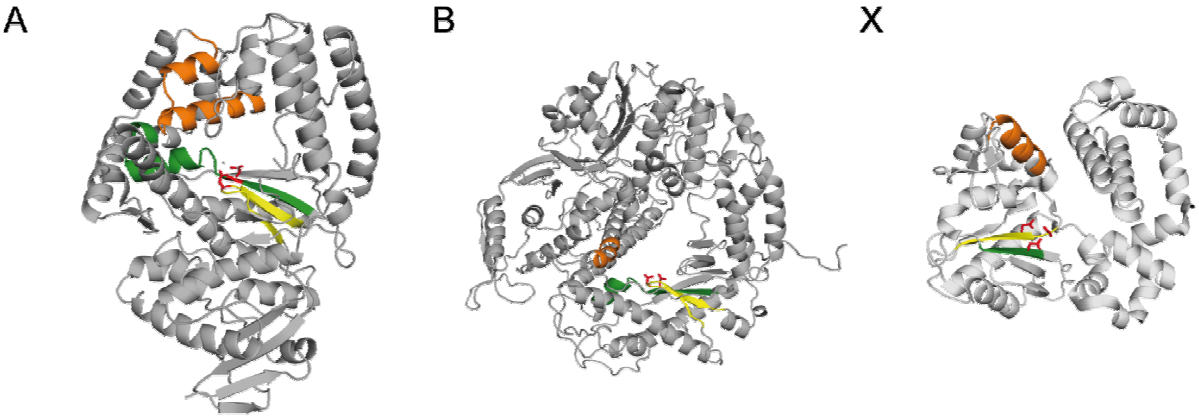


Figure 9 exemplary structural depiction for motifs A (green), B (orange) and C (yellow) in polymerases from family A (KlenTaq, PDB accession code 3KTQ [38]), B (rb69, PDB ID 1WAJ [48]) and X (Pol β, accession code 2FMP [49]). Active site aspartates are highlighted in red.

1.3.5 DNA Polymerases: Mechanism

The minimal model of enzymatic DNA synthesis is a multi-staged process which can be divided into 7 steps (see Figure 10), beginning with the binding of the enzyme (E) to the primer/template complex (p/t) (Figure 10, [1]). During this step a conformational change in the thumb domain occurs. With the entrance of the dNTP, an open ternary complex is formed. The following step leads to activated ternary complex E':p/t:dNTP, including a large conformational change of the finger domain. The transition into the activated state controls the incorporation rate of the whole cycle, though its detailed nature remains unclear, it is to be considered independent of the fingerdomain movement velocitiy.[51] The activated complex can now undergo a reaction step, leading to the formation of a phosphodiester bond. The chemical step completes the elongation of the primer strand (Figure 10, step [4]). A further conformational change facilitates the release of the pyrophosphate (PPi) (step [5]). Thereafter, the polymerase will either translocate the 3' terminus for a new round of incorporation (processive synthesis, step [6]) or dissociate (distributive synthesis, step [7]).[30]

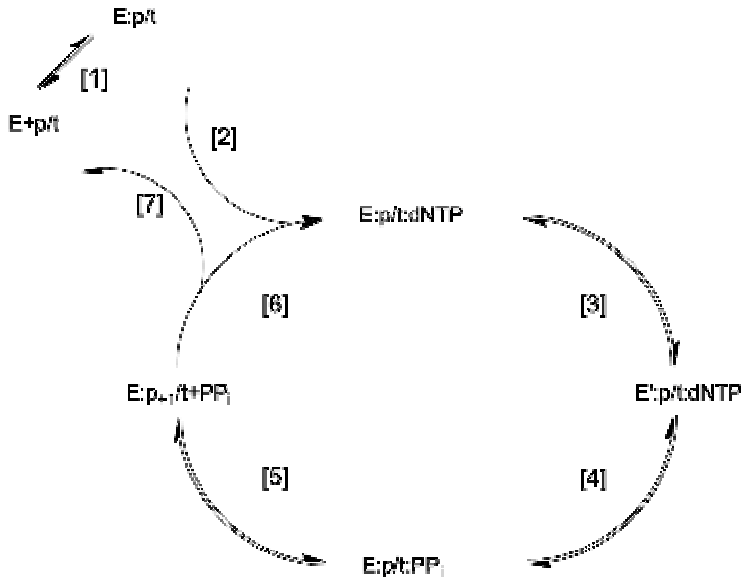


Figure 10 Kinetic mechanism of DNA polymerisation. Graphic adapted and modified from [30].

As an addition to steps 2 and 3 of the shown minimal model, *Rothwell et al.* recently proposed a free energy driven model of the nucleotide binding state based on single molecule FRET measurements using *KlenTaq* DNA polymerase. It shows a non-triggered oscillating motion of the finger domain, cycling through 3 different states; open, closed and nucleotide binding. Since this motion occurs in a substrate independent manner, therefore, it has to be recognized as an intrinsic feature of the protein fold. The equilibrium between the states is redistributed by the nature of the bound substrate.[52]

1.3.6 DNA Polymerases: two metal-ion mechanism

The actual transfer of a new nucleotide on the 3' end of the primer, catalysed by DNA polymerases, is driven by the hydrolysis of the dNTP and the release of pyrophosphate. Despite the multitude of DNA polymerases originating from different families, the nucleotide transfer reaction is based on a common principle, the 2 metal-ion mechanism.[53]

At least two aspartic acid residues, located in motifs A and C (see Figure 7), coordinate two divalent cations. Metal ion A is octahedrally coordinated by the two aspartates, the 3' OH moiety of the primer, the α-phosphate and two water molecules while metal ion B is complexed by the triphosphate, the aspartate of motif A and as well as two additional water molecules. (see Figure 11).

In catalysis metal ion A thereby increases the nucleophilic character of the primer hydroxyl group by contributing to the deprotonation. The proton acceptor can be a water molecule [54] or an active site amino acid.[55] This allows a nucleophilic attack on the α–phosphate of the dNTP. Metal ion B guides the triphosphate in the active site. Both metal ions stabilise the pentavalent transition state of the α-phosphate.

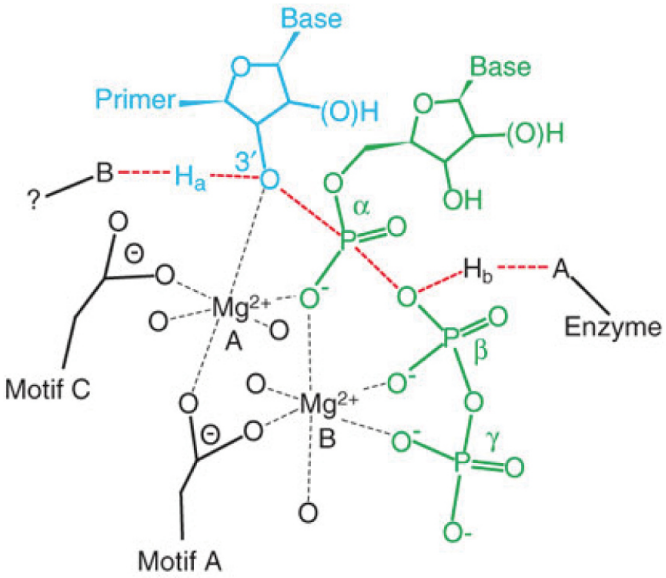


Figure 11 Extended 2 metal ion mechanism with a general acid mechanism, modified from ref[56]

Nearby lysines (as well as arginines and histidines)[56] or ordered water molecules[55] can serve as a general acid in the protonation of the pyrophosphate. Recent time-resolved crystallographic studies on DNA polymerase η and β reported a third metal-ion, appearing either during[54] or after[55] the bond formation, replacing the general acid. The additional positive charge could play a role in pyrophosphorolysis[55].

1.4 Biotechnical application of DNA Polymerases

1.4.1 Overview

The base for biotechnical applications of DNA polymerases is the amplification of nucleic acids in the polymerase chain reaction (PCR).[57]

This technique can used for a multitude of different purposes exemplarily the simple amplification of a defined DNA sequence,[57] quantitative (Q-) PCR for diagnostic purposes[58] or next generation sequencing (NGS) methods[59].

A number of modern biotechnological applications use modified nucleotides for structural characterization, immobilization, DNA conjugation or for the selection of aptamers by systematic enrichment of ligands by exponential amplification (SELEX).[60-62] Besides modified unnatural base pairs, modified analogs of the natural (deoxy-) ribonucleotides are used for that purpose.The first category engulfs a variety of artificial base-pairs as reviewed in Hirao et al.[63].

The second group covers modified analogs of the natural nucleotides. In most cases the nucleobase bears the added functionality.[64-73] The incorporation of such a building block allows the labeling, tagging and immobilization of the polymer as well as an additional modification after the incorporation e.g. click reaction.[74,75] Most recently, Hollenstein et al. proposed the usage of SELEX with nucleotides bearing functions found in the active site of serine protease to generate DNAzymes with a high substrate specificity.[76]

Building highly functionalized DNA can be achieved in two different ways (i) solid support synthesis and (ii) enzymatic synthesis. While (i) is limited in terms of oligomer length and coupling efficiency, (ii) relies on the ability of DNA polymerases to accept the modifications during the steps of incorporation and elongation.

In a series of trial and error, the C5 position in pyrimidines and a deaza-modification on the N7 position in purines have been found to be most efficient positions for modifications at the nucleobase.[64-73]

1.4.2 Structural insights

Though literature shows a diversity of modified dNTPs, there are only a few reports focusing on the structural exploration in the context of the processing enzyme.

Linear alkane modifications at the C4' position of the sugar have been shown to be of use as steric probes to examine interactions of polymerases with the minor grove.[77] Structural studies using KlenTaq polymerase displayed a highly conserved residue in motif A, isoleucine 614, providing steric hindrance towards a bulky modification on the 4'-position. I614A mutants displayed the expected better incorporation tolerance.[78]

Modifications at the nucleobase are, though often used, still unpredictable in their incorporation acceptance. *Obeid et. al.* solved structures of KlenTaq polymerase cocrystallized with two previously published modified pyrimidines triphosphates [60,79]. The results of this structural study point towards importance of the linker-length and its ability to form hydrogenbonds with the enzyme.[64]

In further studies, incorporation and elongation of alkin-modified pyrimidines have been shown, exploiting pi-cation interaction of a rigidly attached aromatic moiety, allowing the further modification of the polymer using click chemistry.[74,75]

2 Publications

The following chapter contains the publications which emerged from experiments and cooperations during the timeline of this work. Each subchapter contains a short introduction, objectives and intentions of the project as well as the published manuscript and the Supporting Information. Numbering of figures and tables is kept in the order appearing in the respective publication. Contributions to the work are listed in chapter 2.4. Final remarks and an outlook towards the future of the projects are given in chapter 3. The data sets used for the studies are listed in the Appendix (Chapter 5) and refer to the location in the data storage used in the AG Welte/Diederichs.

2.1 "KlenTaq DNA polymerase caught incorporating C5 and 7-deaza modified nucleotides"

2.1.1 Introduction to "KlenTaq DNA polymerase caught incorporating C5 and 7-deaza modified nucleotides"

The thermophilic eubacterium *Thermus aquaticus* (Taq) has been first described in 1969 by *Brock and Freeze*[80]- 9 Years later, Chien et al. were able to purify and characterize a thermostable DNA polymerase from the very same organism called Taq DNA polymerase.[81] Further characterization over the years showed a high degree of homology on sequential as well as structural level with polymerases from family A.[26]

Analogously to the *E. coli* polymerase I, Taq polymerase I shows a fragmentation of the full-length protein into a small fragment, containing the 5'-3' exonuclease domain and a large one containing the polymerase function as the Taq polymerase shows no 3'-5' proofreading activity. The large fragment, KlenTaq or Stoffel-fragment, has been therefore extensively used for structural studies to elucidate the the mechanistical base for incorporation of natural triphosphates[38,39,82-84], lesion bypass[85], modified[64,74,75,78,86] and non-natural nucleotides[87].

2.1.2 Goals of the project

Chemically altered substrates (dNTPs) can be used to label DNA for multi-purpose applications (see chapter 1.4). Most of these approaches modify the N7 positions of

purines and the C5 positions of pyrimidines using a multitude of linkers and attached functions (see Chapter 1.3). In order to gain insights into the mechanistic base of the incorporation of modified dNTPs (further referred to as dN*TP and dN**TP), synthesized and biochemically characterized by Baccaro and Steck[66], structural investigations were undertaken. Though being key to a rational approach in the development of enzymatically added augmentations to DNA, structural data of base-modified nucleoside triphosphates has been limited to modified thymidines[64] at the begin of this study.Therefore the structural basis for the superior incorporation rates of the dN*(*)TPs should be elucidated.

2.1.3 KlenTaq DNA polymerase caught incorporating C5 and 7-deaza modified nucleotides

Reproduced with permission from
"Structures of KlenTaq DNA Polymerase Caught While Incorporating C5-Modified Pyrimidine and C7-Modified 7-Deazapurine Nucleoside Triphosphates"
Konrad Bergen, Anna-Lena Steck, Stefan Strütt, Anna Baccaro, Wolfram Welte, Kay Diederichs, and Andreas Marx
Journal of the American Chemical Society **2012** *134* (29), 11840-11843

Structures of *KlenTaq* DNA Polymerase Caught While Incorporating C5-Modified Pyrimidine and C7-Modified 7-Deazapurine Nucleoside Triphosphates

Konrad Bergen,‡ Anna-Lena Steck,‡ Stefan Strütt, Anna Baccaro, Wolfram Welte, Kay Diederichs, and Andreas Marx*

Departments of Chemistry and Biology, Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstr. 10, 78457 Konstanz (Germany)

ABSTRACT: *The capability of DNA polymerases to accept chemically modified nucleotides is of paramount importance for many biotechnological applications. Although these analogues are widely used, the structural basis for the acceptance of the unnatural nucleotide surrogates is only sparsely explored. Here, we present in total six crystal structures of modified 2'-deoxynucleoside-5'-O -triphosphates (dNTPs) carrying modifications at C5 position of pyrimidines and C7 of 7-deazapurines in complex with a DNA polymerase and primer/template complex. The respective modified dNTPs are in positions poised for catalysis leading to incorporation. This structural data provide insight into the mechanism of incorporation and acceptance of modified dNTPs. Our results open the door for rationally designing of modified nucleotides which should offer great opportunities for future applications.*

The ability of DNA polymerases to process nucleobase-modified 2'-deoxynucleoside-5'-o-triphosphates (dNTPs) is often the essential step in many biotechnological applications.[60,88-91] Modified nucleotides are used for structural characterization, immobilization, DNA conjugation or for the selection of aptamers by systematic enrichment of ligands by exponential amplification (SELEX).[60-62] For instance, dye-labeled nucleotides are of outstanding importance in DNA sequencing approaches.[89,91-93] In most cases, the modification is linked to the nucleobase moiety.[64,65,67-73,94] Thereby the C5 position of pyrimidines and the C7 of 7-deazapurines were identified as best suited for the introduction of modifications without compromising DNA polymerase activity.[64,65,67-73,94] Although modified nucleotides are widely employed, the mechanisms by which they are accepted as substrate and incorporated by DNA polymerases are still unclear and until now, the acceptance of modified nucleotides by a DNA polymerase is often not predictable.[95] Hitherto, structural data of DNA polymerases in complex to modified nucleotides are limited to a single report on thymidine analogs.[64]

Here, we present several crystal structures, a set of modified 2'-deoxynucleoside-5'-O-triphosphates (dN*TPs) carrying the same aminopentinyl modification[94,96,97] bound to DNA polymerase and primer/template complex. The modifications are linked either to the C5 position of pyrimidines or to the 7- deaza position of purines (see Figure 1 for dC*TP, dT*TP, dG*TP, dA*TP). The amine modified nucleotides are well suited for further functionalization via amide bond formation and are therefore of great interest.[66,95] Additionally, these nucleotides are compared to their natural counterparts by structural and functional means. For our studies, we chose the N-terminally truncated form of the DNA polymerase I from Thermus aquaticus (KlenTaq) DNA polymerase due to its well known characteristics on the structural and functional level. [39,51,64,78,82-84,98,99]

We first investigated the efficiency of nucleotide incorporation of the nucleobase-modified nucleotides dN*TP in comparison to their natural counterparts. We performed single nucleotide incorporation experiments in which the modified nucleotides directly compete for incorporation with their natural counterparts. This experimental setup was previously used for the same purpose[64,94] as well as to study DNA polymerase selectivity.[29] In Figure 1B-D an exemplary study employing dA*TP and dATP is depicted [see Figure S1in the Supporting Information(SI) for results for the other dN*TPs]. We used a 24-nucleotide (nt) primer with a 32P-label at the 5'-end

and four different 36-nt templates, which code for the extension of the primer by a single complementary nucleotide (the DNA sequences are listed in the SI). The ratio of unmodified versus modified nucleotide incorporation is easily accessible via denaturating polyacrylamide gel electrophoresis (PAGE) analysis and phosphorimaging because of the significantly different retention times resulting from the additional modification at the modified nucleotide (Figure 1C). Similar observations of lower mobility for modified reaction products have been reported before.[79,100] Interestingly, we found that KlenTaq DNA polymerase incorporates the purine analogues with approximately the same efficiency as the natural counterparts, whereas the pyrimidine analogues were incorporated with about 12.3 to 34-fold lower efficiency than their natural counterparts (Figure 1E).

Knowing that KlenTaq DNA polymerase accepts the dN*TPs, we aimed at solving crystal structures of KlenTaq DNA polymerase in complex with the modified nucleotides. To obtain crystals suitable for structure elucidation of the KlenTaq DNA polymerase in complex with DNA primer/template and the dN*TPs, we employed different crystallization strategies[78,83,98,101] and obtained best results with a method similar to the one reported by Beard et al.[101]. We crystallized binary complexes of KlenTaq DNA polymerase in complex with primer and template first and than soaked the crystals with the respective dN*TP (Figure 1A, Supporting Information). Details on structure solution and refinement can be found in the Supporting Information. All structures of KlenTaq DNA polymerase in complex with the four modified dN*TPs were obtained in high resolution (1.9-2.0 Å) and adopt conformations similar to the unmodified cases (rmsd. Cα atoms 0.33-0.38 Å). The respective modified dN*TPs are bound in positions poised for catalysis and undergo canonical Watson-Crick nucleobase pairing to the templating nucleobases (Figure 2, Figure 3C). As well, the O-helix of the finger-domain is packed tightly to the nascent base pair thereby forming a closed, active complex, comparable to the structures observed when natural substrates were used. The distances of the 3' end of the primer to the alpha-phosphates of the modified nucleotide are slightly higher than in the natural case and range between 3.8 and 3.9 Å (Figure 2). In previously reported structures containing C5 modified dTTPs, Arg660[39,99], that is interacting with the primer strand when an unmodified nucleotide is bound, is displaced substantially due to the steric hindrance of the bulky C5 modification.[64] The structures presented here show only a smaller reorientation of Arg660 compared to the previously reported structure (pdb ID 3OJU,

Figure S2, Supporting Information) with the exception for the case when dG*TP is bound (Figure 2).[39] Based on amino acid alignment of several A-family DNA polymerases it is known that this Arg660, in particular, is located within the motif B[27,64] and is conserved in bacteria[64]. Hence, it is likely that the stabilizing effect caused by Arg660 applies to other DNA polymerases in this sequence family as well.

The direct comparison of the modified dN*TPs revealed an unexpected feature concerning the orientation of the aminopentinyl modifications. Interestingly, as the modifications of dC*TP, dG*TP and dT*TP point towards the base, only varying the plane due to electrostatic reasons (Figure S2, Supporting Information), the modification of dA*TP is pointing towards the phosphate of the primer terminus (Figure 2C). To verify whether this is a single observation specific for this modification we studied extended modifications as those in 7-(N-(10-hydroxydecanoyl)-aminopentinyl)-7-deaza-dATP (dA**TP) and 5-(N-(10-hydroxydecanoyl)-aminopentinyl)-dUTP (dT**TP) (Figure 3A). The compounds were synthesized and crystallized in the same manner as described above (for details see SI). Notably, both analogues are accepted by KlenTaq DNA polymerase. In competition experiments, dA**TP was incorporated with approximately the same efficiency and dT**TP with ~12-fold lower efficiency than their natural counterparts (Figure S3). The structures derived from this data showed similar overall properties as unmodified cases (Table S1 in the SI, rmsd. Cα atoms = 0.30-0.54 Å). Intriguingly, the orientations as well as the stabilization sites differ for the dA**TP and the dT**TP, consequently the modifications are extending from the enzyme active site through different cavities (Figure 3B, C). In detail, dA**TP is stabilized mainly by Lys663 and most likely by Arg660 (Figure 3C, left panel; Lys663 in the background). Lys663 is within interaction distance to the amide bond. In regard to dT**TP, hydrogen bonds to the amide are formed by residues Thr664 and Arg660, leading the modification towards the cavity mainly formed by residues of the O-helix.

In comparison to the previous structures of nucleobase-modified nucleotides,[64] the modifications used in this study are more flexible and lead to only small disorder in the active site. The relatively low disturbing effects on enzyme conformations can be explained by the interaction patterns of the modifications either with the base (in case of dT*TP and dG*TP), the phosphate of the primer terminus (dA*TP), or the residues of O-helix (in case of dA**TP and dT**TP). The amide bond in the modification leads to further hydrogen-bonding of the modification to residues in the O-helix. These

properties might also explain the high incorporation efficiencies of the modified nucleoside triphosphates proven by the functional studies (Figure 1 and S3). In direct comparison with the previously solved structure of modified dTTP with a spin-label modification,[64] the smaller displacement of R660 as well as the additional hydrogen-bonding capacity of the modified triphosphate can explain the higher incorporation rates and acceptance of dT*TP and dT**TP (Figure 2A-C). As a consequence of these observations, the combination of modification length and the positioning of the amide bond between the subsets of the modification seem to be an important factor for efficiency of the incorporation of the modified building block.

This structural data provide insight into the mechanism of acceptance and incorporation of modified dNTPs by a DNA polymerase that is widely used in biotechnological applications and should open the door for rationally designing of modified nucleotides. The beneficial combination of rational designed modified nucleotides and directed evolution of DNA polymerases[102-106] offers great opportunities for future applications.

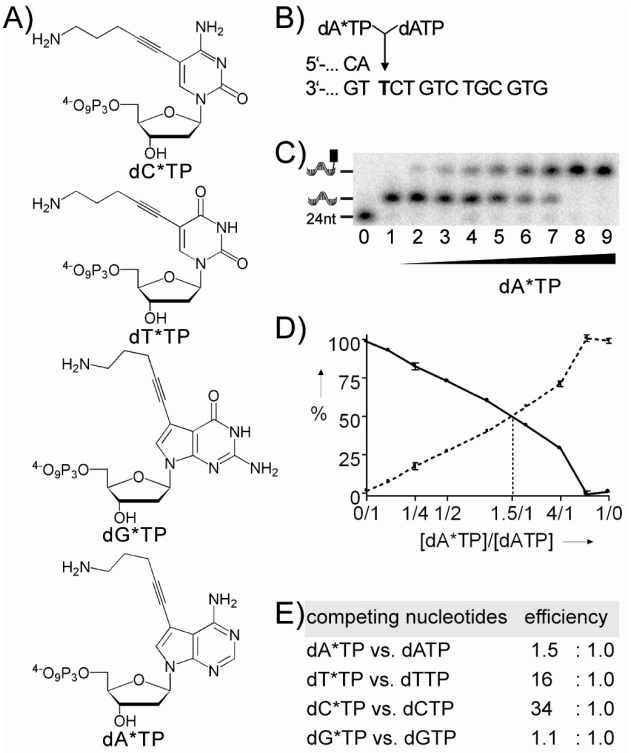


Figure 1. (A) Structures of the aminopentinyl-modified nucleoside triphosphates dN*TP. (B) Exemplary partial DNA sequences of primer and template for the reactions employing dA*TP and dATP. C) PAGE analysis of an exemplary competition experiments employing KlenTaq DNA polymerase. The ratio of dA*TP/dATP was varied. Lane 0: 5'-32P-labeled primer; lane 1: ratio: 0/1; lane 2: ratio: 1/10; lane 3: ratio: 1/4; lane 4: ratio: 1/2; lane 5: ratio: 1/1; lane 6: ratio: 2/1; lane 7: ratio: 4/1; lane 8: ratio: 10/1; lane 9: ratio: 1/0. D) Evaluation of the incorporation efficiency using mixtures with varied composition of dA*TP (■, dashed line) and dATP (•, solid line) and KlenTaq DNA polymerase. The % conversion is plotted versus the dA*TP /dATP ratio. The dotted line marks the approximate ratio where both nucleotides are equally incorporated. E) Overview of the efficiencies of the presented modified nucleotides in competition with their natural counterparts (see Figure S1).

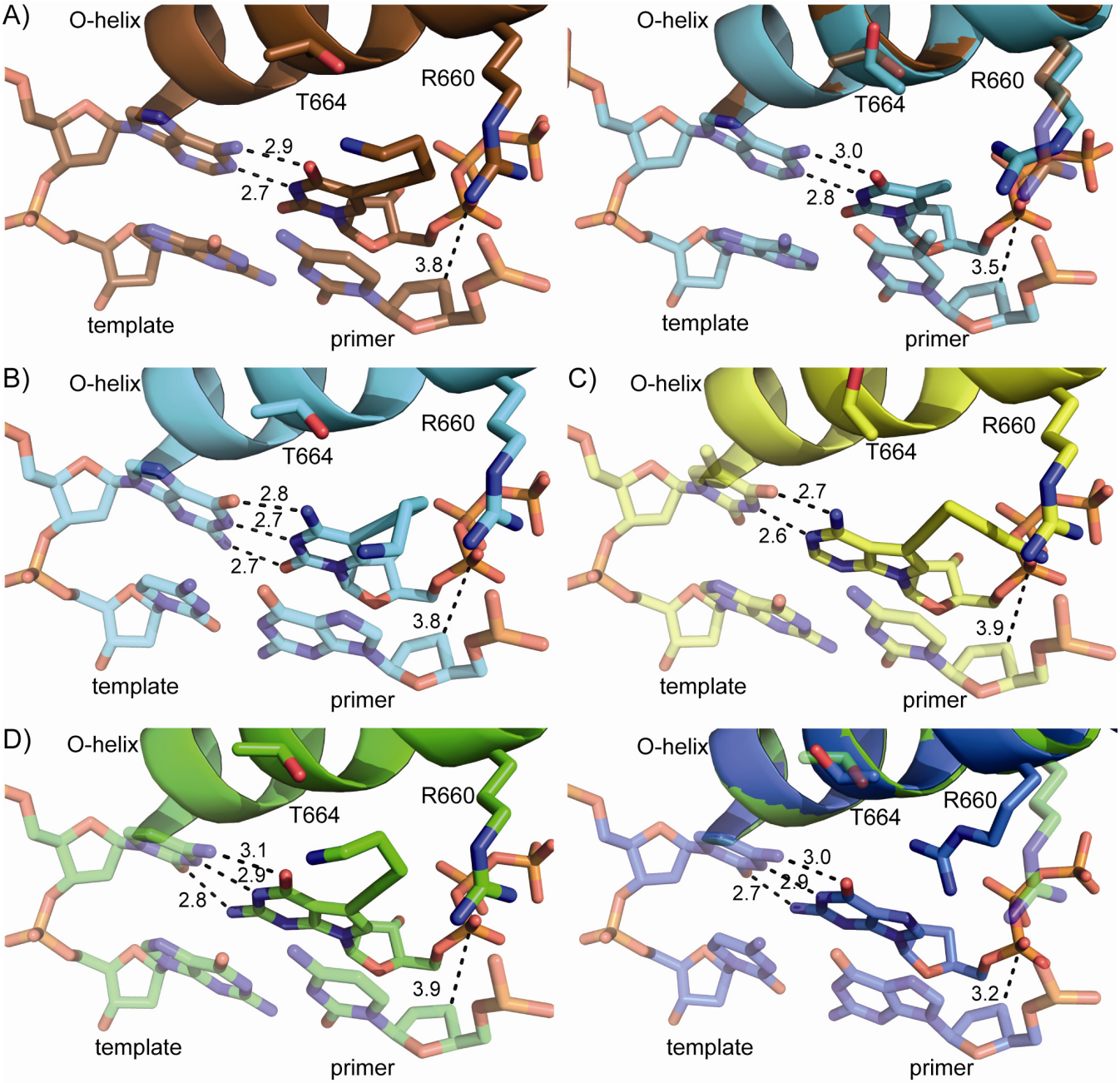


Figure 2. Close-up views of KlenTaq DNA polymerase showing the incoming nucleoside triphosphate and the O-helix. The dashed lines highlight the Watson-Crick base pairing interactions and the distance of the □-phosphate to the primer 3'-terminus. All distances are in Å. (A) Left panel: interaction distances and orientation of dT*TP and the position of Arg660. Right panel: interaction distances of the natural ddTTP (PDB ID 1QTM) and an overlay of the Arg660 residues. (B) Same as (A) left panel for the structure containing dC*TP. (C) Same as (A) left panel for the structure containing dA*TP.(D) Left panel: same as (A) left panel for the structure containing dG*TP. Right panel: structure containing ddGTP (PDB ID 1QSS) and an overlayof Arg660 as in (A).

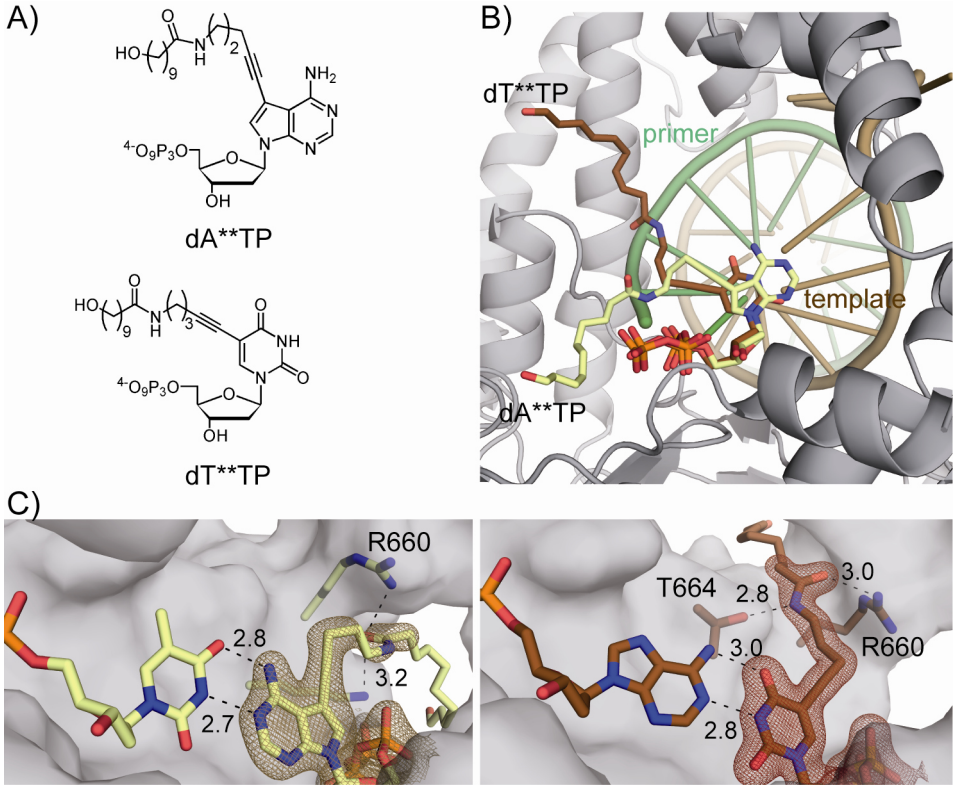


Figure 3. A) Depiction of the used compounds (top dA**TP, below dT**TP). B) Close-up view of the active center. The orientations of the modifications are shown in sand for the dA**TP and in brown for the dT**TP. Parts of the enzyme, including the O-helix are removed for better visibility. C) Close-up views of the nascent base pairs and the orientation of the attached modifications. Depicted is the model density at 1 σ. The distances in Å are indicated in dashed lines. The left panel shows dA**TP with the stabilizing residue Lys663 (K663, in the background) and Arg660 (R660), the active site is shown as Connolly surface. The right panel shows dT**TP with Arg660 and Thr664 forming hydrogen bonds to the amide of the modification. Both cavities are lined by Arg660.

ASSOCIATED CONTENT

Supporting Information

Synthetic procedures for the used compounds dA**TP and dT**TP; the reactions in solution, the preparations of the crystals; additional figures of the crystal structures and the crystallographic tables; complete references 1a), 1c) and 3a). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Andreas.Marx@uni-konstanz.de

Author Contributions

‡ These authors contributed equally.

ACKNOWLEDGMENTS

We gratefully acknowledge funding by the Konstanz Research School Chemical Biology and the Ministerium für Wissenschaft, Forschung und Kunst, Baden-Württemberg for funding within the programme Bionik, and support and access to beamlines PXI and III at the Swiss Light Source (SLS) at the Paul-Scherrer Institute (PSI), Villigen, Switzerland.

2.1.4 Supporting Information:

Crystallization of *KlenTaq* DNA polymerase in complex with C5 and 7-deaza modified nucleotides

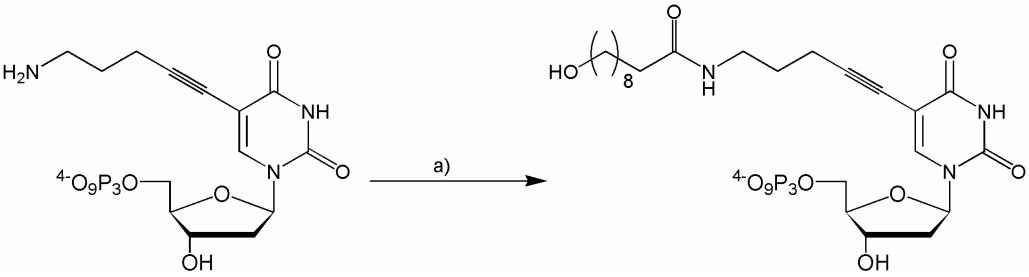
Konrad Bergen,[†,‡] Anna-Lena Steck,[#,‡] Stefan Strütt,[†] Anna Baccaro, [#] Wolfram Welte,[†] Kay Diederichs,[†] and Andreas Marx [#,*]

Table of contents

2',3'-Dideoxy-cytidine-5'-triphosphate was purchased from *JenaBioscience*. Succinimidyl 10-hydroxydecanoate[107] were prepared according to literature. 5-(aminopentynyl)-2'-deoxyuridinetriphosphate **dT*TP**, 5-(aminopentynyl)-2'-deoxycytidinetriphosphate **dC*TP**, 7-(aminopentynyl)-7-deaza-2'-deoxyadenosinetriphosphate **dA*TP** and 7-(aminopentynyl)-7-deaza-2'-deoxyguanosinetriphosphate **dG*TP** were synthesized according to known procedures.[108-112] All reagents are commercially available and were used without further purification. Solvents were stored over molecular sieves (*Fluka*) and used directly without further purification, unless otherwise noted. All synthetic reactions were performed under an inert atmosphere. Flash chromatography was done using *Merck* silica gel G60 (230–400 mesh) and *Merck* precoated plates (silica gel 60 F254) were used for TLC. NMR spectra were recorded on *Bruker* Avance 400 (^{1}H: 400 MHz, ^{13}C: 101 MHz, ^{32}P: 162 MHz) spectrometer and *Bruker* 600 (^{1}H: 600 MHz). The solvent signals were used as references and the chemical shifts converted to the

TMS scale and are given in ppm (δ). HRMS spectra were recorded on a *Bruker* mircOTOF II in the negative mode.

Synthesis of 2'-deoxythymidine analogue dTTP**



dT*TP **dT**TP**

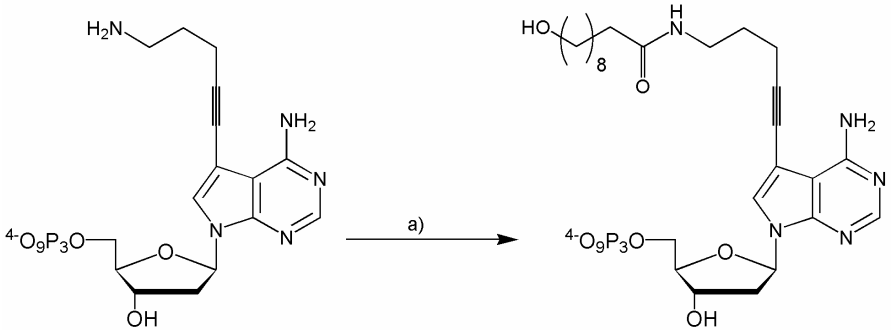
Scheme S1 Synthesis of modified 2'-deoxythymidine analogue for crystallization a) succinimidyl 10-hydroxydecanoate, DMSO, 5 h, rt.

5-(N-(10-hydroxydecanoyl)-aminopentynyl)- 2'-deoxyuridinetriphosphate ***dT**TP***

5-(aminopentynyl)-2'-deoxyuridinetriphosphate **dT*TP** (548µg, 1µmol) and succinimidyl 10-hydroxydecanoate (2.85 mg, 10 µmol) were dissolved in DMSO (0.5 ml). After shaking at room temperature for five hours, the solvent was removed *in vacuo*. The residue was suspended in water and filtered. The solution was purified by RP-HPLC (Nucleosil 100-5 C18 PPN, 5-100 % acetonitrile/ 0.1 M TEAA buffer (pH 7.0)) to give triphosphate **dT**TP** in quantitative yield. ^{1}H-NMR (600 MHz, MeOD) δ = 8.02 (s, 1H, H-6), 6.26 (t, ^{3}J = 6.8 Hz, 1H, H-1'), 4.66 – 4.62 (m, 1H, H-3'), 4.36 – 4.31 (m, 1H, H-5'a), 4.23 – 4.18 (m, 1H, H-5'b), 4.10 – 4.07 (m, 1H, H-4), 3.55 (t, ^{3}J = 6.8 Hz, 2H, -CH_2CH_2OH), 3.35 – 3.31 (m, 2H, -CH_2NH-, superimposed by MeOH), 3.22 – 3.14 (m, 24H, Et_3N), 2.46 (t, ^{3}J = 6.8 Hz, 2H, -C≡CCH_2-), 2.32 – 2.26 (m, 2H, H-2'a/b), 2.24 – 2.20 (m, 2H, -COCH_2-), 1.80 (p, ^{3}J = 6.8 Hz, 2H, -$CH_2CH_2CH_2$NH-), 1.64 – 1.58 (m, 2H, -$COCH_2CH_2$-), 1.57 – 1.51 (m, 2H, -CH_2CH_2OH), 1.38 – 1.28 ppm (m, 48H, Et_3N, -CH_2-); ^{31}P-NMR (162 MHz, MeOD) δ -10.20 (d, ^{2}J = 21.3 Hz, 1P,

P_γ), -11.23 (d, ^2J = 21.6 Hz, 1P, P_α), -23.55 ppm (t, ^2J = 22.9 Hz, 1P, P_β).; HRMS (negative mode): *m/z*: calcd for $[C_{24}H_{39}N_3O_{16}P_3]^-$: 718.1549; found: 718.1545.

Synthesis of 2'-deoxyadenosine analogue dA**TP



dA*TP **dA**TP**

Scheme S2 ***Synthesis of modified 7-deaza-2'-deoxyadenosine analogue a) succinimidyl 10-hydroxydecanoate, DMSO, 5 h, rt.***

7-(N-(10-hydroxydecanoyl)-aminopentynyl)-7-deaza-2'-deoxyadenosinetriphosphate

dA**TP

7-(aminopentynyl)-7-deaza-2'-deoxyadenosinetriphosphate **dA*TP** (570 µg, 1µmol) and succinimidyl 10-hydroxydecanoate (2.85 mg, 10 µmol) were dissolved in DMSO (0.5 ml). After shaking at room temperature for five hours, the reaction mixture was freeze-dried. The residue was suspended in water and filtered. The solution was purified by RP-HPLC (Nucleosil 100-5 C18 PPN, 5-100 % acetonitrile/0.1 M TEAA buffer (pH 7.0)) to give triphosphate **dA**TP** in quantitative yield. ^{1}H-NMR (400 MHz, MeOD): δ = 8.20 (br, 1H, H-2), 7.72 (s, 1H, H-8), 6.64 (t, ^{3}J = 6.2 Hz, 1H, H-1'), 4.72 (br, 1H, H-3'), 4.32 – 4.20 (m, 2H, H-5'a/b), 4.14 (br, 1H, H-4'), 3.58 – 3.53 (m, 2H, -CH_2CH_2OH), 3.35 (m, 2H, -CH_2NH-, superimposed by MeOH), 3.24 – 3.19 (m, 13H, Et_3N), 2.61 – 2.50 (m, 3H, H-2'a, -C≡CCH_2-), 2.38 – 2.32 (m, 1H, H-2'b), 2.24 – 2.20 (m, 2H, -COCH_2-), 1.87 – 1.79 (m, 2H, -$CH_2CH_2CH_2$NH-), 1.65 – 1.63 (m, 2H, -$COCH_2CH_2$-), 1.58 – 1.52 (m, 2H, -CH_2CH_2OH), 1.35 (m, 36H, Et_3N, -CH_2-); ^{31}P-

NMR (162 MHz, MeOD): δ -10.09 (d, J = 20.0 Hz, 1P, P_γ), -10.99 (d, J = 21.3 Hz, 1P, P_α), -22.99 - -23.58 (m, 1P, P_β). HRMS (negative mode): *m/z*: cald for $[C_{26}H_{41}N_5O_{14}P_3]^-$: 740.1868, found: 740.1891.

Enzymes, oligodeoxynucleotides, nucleotides

KlenTaq DNA polymerase was expressed and purified as described before.[113,114] T4 polynucleotide kinase PNK was purchased from *Fermentas*. Primer and templates were purchased from *Metabion* and *ThermoFisher*. [γ-^{32}P]ATP was purchased from *Hartmann Analytics* and natural dNTPs from *Roche*.

Buffers and solutions

- 1 M TEAA buffer (1 M acetic acid, 1 M triethylamine, (pH 7))
- 10 x *KlenTaq* reaction buffer (500 mM Tris HCl (pH 9.2), 160 mM $(NH_4)_2SO_4$, 25 mM $MgCl_2$, 1% Tween 20)
- PAGE gel loading buffer (80% formamide, 20 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanole FF)

5'-Radioactive labeling of ODNs

DNA oligonucleotide primers were radioactively labeled at the 5' terminus by a ^{32}P containing phosphate group using T4 PNK which transfers the γ-phosphate group from [γ-^{32}P]ATP to the 5' hydroxyl group. The reactions contained primer (0.4 µM), PNK reaction buffer (1 ×), [γ-^{32}P]ATP (0.8 µCi/µl) and T4 PNK (0.4 U/µl) in a total volume of 50 µl and were incubated for 1 h at 37 °C. The reaction was stopped by denaturing the T4 PNK for 2 min at 95 °C and buffers and excess [γ-^{32}P]ATP were removed by gel filtration (MicroSpin Sephadex G-25). Addition of unlabeled primer (20 µl, 10 µM) led to a final concentration of 3 µM of diluted radioactive labeled primer.

Gel electrophoresis

Denaturing polyacrylamide gels (12 %) were prepared by polymerization of a solution of urea (8.3 M) and bisacrylamide/acrylamide (12 %) in TBE buffer using ammonium peroxodisulfate (APS, 0.08 %) and N,N,N',N'-tetramethylethylene-diamine (TEMED, 0.04 %). Immediately after addition of APS and TEMED the solution was filled in a sequencing gel chamber (*Bio-Rad*) and left for polymerization for at least 45 min. After addition of TBE buffer (1 ×) to the electrophoresis unit, gels were prewarmed by electrophoresis at 100 W for 30 min and samples were added and separated during

electrophoresis (100 W) for approx. 1.5 h. The gel was transfered to *Whatman* filter paper, dried at 80 °C, *in vacuo*, using a gel dryer (model 583, *Bio-Rad*) and exposed to a imager screen. Readout was performed with a molecular imager (FX, *Bio-Rad*).

DNA sequences

radioactive-labeled primer: 5' d(GTG GTG CGA AAT TTC TGA CAG ACA)

template (incorporation of dTMP): 5' d(GTG CGT CTG TC**A** TGT CTG TCA GAA ATT TCG CAC CAC)

template (incorporation of dAMP): 5' d(GTG CGT CTG TC**T** TGT CTG TCA GAA ATT TCG CAC CAC)

template (incorporation of dCMP): 5' d(GTG CGT CTG TC**G** TGT CTG TCA GAA ATT TCG CAC CAC)

template (incorporation of dGMP): 5' d(GTG CGT CTG TA**C** TGT CTG TCA GAA ATT TCG CAC CAC)

Primer extension reaction

A typical primer extension reaction (10 μL) employing *KlenTaq* DNA polymerase contained 1 x *KlenTaq* reaction buffer, 50 nM ^{32}P-labeled primer, 75 nM template, 200 μM deoxynuclesidtriphosphate mixture, and 100 nM *KlenTaq* DNA polymerase. First primer and template were annealed. Afterwards the primer template complex, nucleotides and DNA polymerase were incubated (60°C; 10 sec). The reactions were quenched by addition of 30 μL PAGE gel loading buffer and the product mixtures were analyzed by 12% denaturing polyacrylamide gel and subjected to autoradiography. Quantification was done by using the Bio-Rad Quantity One software. The conversion in % was plotted *versus* the concentration using the program GraphPad Prism4. All reactions were done in duplicates.

dNTP mixture: 0/1, 1/1, 2/1, 4/1, 10/1, 20/1, 50/1, 100/1, 1/0 (dT*TP,dT**TP, dC*TP)

dNTP mixture: 0/1, 1/10, 1/4, 1/2, 1/1, 2/1, 4/1, 10/1, 1/0 (dA*TP, dA**TP, dG*TP)

Crystallization, data collection and analysis

The protein was overexpressed and purified as described earlier[114] and concentrated to 18 mg/ml for storage. For crystallization, a 11-nt primer and a 16-nt template were stepwise annealed and added to the polymerase (protein to DNA ratio 1:1.2, final protein concentration of ~6.2 mg/ml). Dideoxy-terminated primer ends were created by the addition of a 5 molar excess of 2',3'-Dideoxy-cytidine-5'-triphosphate. The

solution was set to final concentration of 20 mM $MgCl_2$ and incubated for 1 hour at 30°C.

Binary crystals forms of the *KlenTaq* DNA polymerase with dideoxy-terminated primer strands were grown in hanging drop plates (*Qiagen*) against 1 ml reservoir (Crystallization condition derived from NucPro HTS Screen, *JenaBioScience* (100 mM Tris·HCl pH 8, 200mM Mg-formate, 18% PEG 8000)) in a 1:1 ratio protein/DNA to reservoir. Fully grown crystals were harvested after 6 days of growth and transferred in a 2 µl drops of stabilizing solution consisting of the crystallization reservoir with additional 20% ethylene glycol, which was pre-equilibrated for 2 h. Soaking was performed by diluting the modified dNTPs in the stabilizing solution before equilibration. Best results were found at a final concentration of 1mM (dA*TP,) and 2mM (dT*TP), followed by 30 minutes incubation. After soaking, crystals were transferred in liquid nitrogen.

Data were collected at beamlines PXI and PXIII at the Swiss Light Source (SLS), Paul-Scherrer Institute, Villigen, Switzerland. Data integration and reduction was performed using XDS[115]. Structure solution was done using difference Fourier-Methods with the PHENIX suite[116].

Generation of library and geometry files was performed using the program Sketcher in the CCP4i suite[117]. Subsequent refinement of the data was done employing Coot[118] and ML refinement methods of the PHENIX suite (see Table S2).

Table S1 Data collection and refinement

PDB ID	**4DFM**	**4DFP**	**4DFJ**	**4DF8**	**4DFK**	**4DF4**
Data collection	**KlenTaq dC***	**KlenTaq dG***	**KlenTaq dT***	**KlenTaq dA***	**KlenTaq dT****	**KlenTaq dA****
Spacegroup	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Cell dimensions						
***a, b, c* (Å)**	a,b=108.2 c=90.1	a,b=108.3 c=90.3	a,b=108.9 c=90.4	a,b=108.6 c=90.5	a,b=107.8 c=89.7	a,b=107.8 c=89.7
α, β, γ (°)	α,β=90 γ=120	α,β=90 γ=120	α,β=90 γ=120	α,β=90 γ=120	α,β=90 γ=120	α,β=90 γ=120
Resolution*	47.15-(1.94) 1.89	46.96-(2.12) 2.00	46.96-(1.95) 1.90	47.03-(2.12) 2.00	46.86-(1.75) 1.65	46.20-(2.22) 2.09
R_{meas}* +	19.3 (95.4)	13.5 (129.2)	14.4 (88.5)	14.2 (153.4)	7.3 (134.5)	14.1 (108.1)
***I/σI* ***	12.77 (1.83)	12.36 (1.69)	14.86 (1.61)	14.43 (1.78)	16,14 (1.65)	11,33 (1.73)
Completeness (%)*	98.6 (82.0)	99.8 (98.9)	94.1 (62.1)	99.9 (99.4)	99.8 (98.5)	99.7 (98.0)
Refinement						
Resolution (Å)*	47.15-1.89	46.96-2.00	46.96-2.00	47.03-2.00	46.86-1.65	46.20-2.20
No.unique reflections*	48759	41788	46509	41984	73657	30932
R_{work}/R_{free} *	8.6/21.5	18.5/22.4	18.0/20.4	19.5/23.4	17.0 /19.5	17.7/22.7
B-factors	35.0	38.2	32.9	39.6	34.8	27.6
DNA	34.8	38.7	34.9	36.6	37.9	29.4
Protein						
R.m.s deviations	0.008	0.007	0.007	0.007	0.009	0.007
Bond lengths (Å)	1.42	1.30	1.39	1.38	1.55	1.33
Bond angles (°)						
Ramachandran#						
Favored	97.7	97.4	97.4	97.6	97.8	97.2
Allowed	2.1	2.4	2.4	2.2	1.8	2.6
outlier	0.2	0.2	0.2	0.2	0.4	0.2

*Numbers in brackets refer to the highest resolution shell
+ for definition of R_{meas}, see [119]
#as determined by MolProbity

R.m.s deviations Cα-atoms	0.34 (3KTQ)	0.34 (1QSS)	0.33 (1QTM)	0.38 (1QSY)	0.30 (1QTM)	0.54 (1QSY)

Deviation of Cα-atoms to the corresponding unmodified structures (the related PDB IDs are noted in brackets)

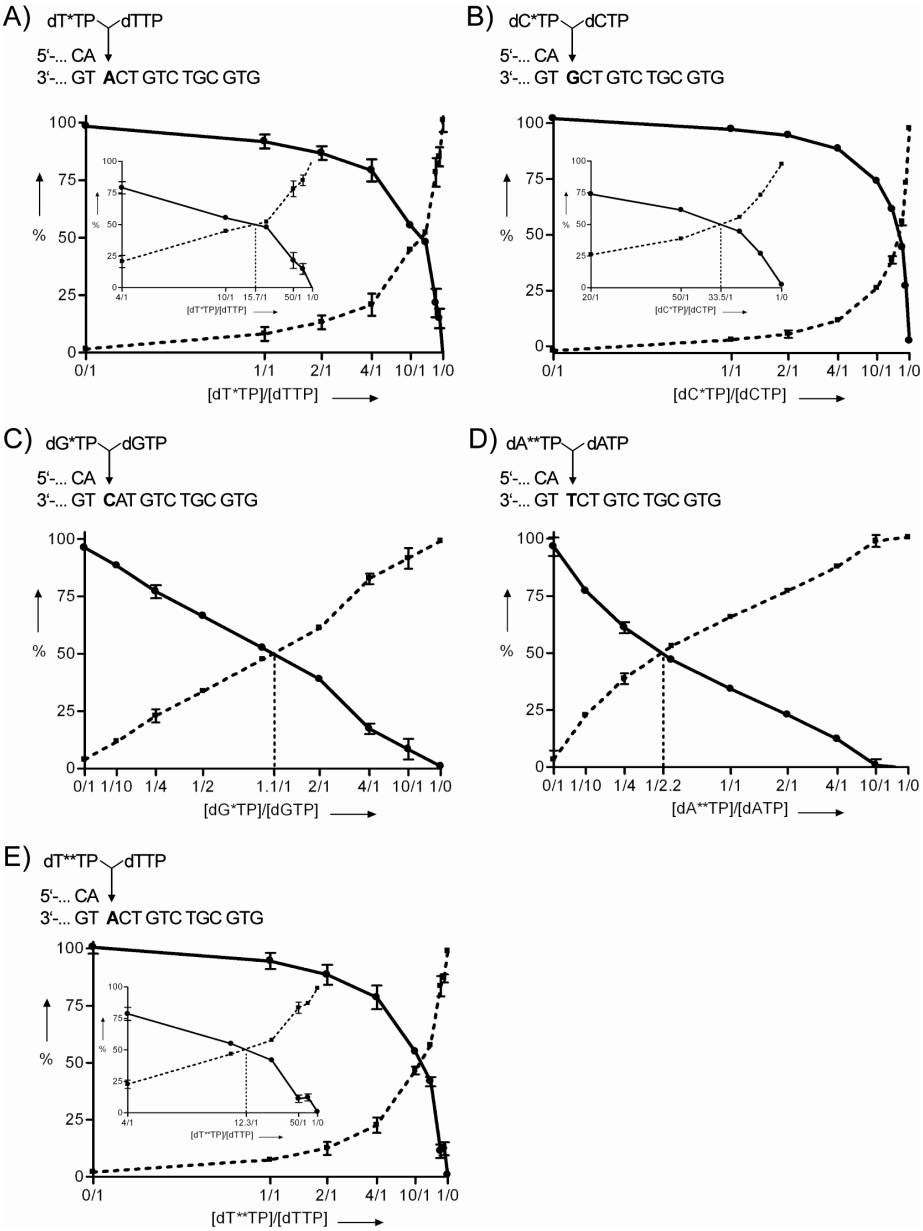


Figure S1: Competition experiments of dN*TP *versus* dNTP. The conversion in % was plotted *versus* the concentration using the program GraphPad Prism4. The dotted line marks the approximate ratio where both nucleotides are equally incorporated. A) Partial DNA sequences of primer and template for the incorporation of dT*TP and dTTP. Evaluation of the incorporation efficiency using dT*TP (■, dashed line)/dTTP (•, solid line) mixtures and *KlenTaq* DNA polymerase. B) Partial DNA sequences of primer and template for the incorporation of dC*TP and dCTP. Evaluation of the incorporation efficiency using dC*TP (■, dashed line)/dCTP (•, solid line) mixtures and *KlenTaq* DNA polymerase. C) Partial DNA sequences of primer and template for the incorporation of dG*TP and dGTP. Evaluation of the incorporation efficiency using dG*TP (■, dashed line)/dGTP (•, solid line) mixtures and *KlenTaq* DNA polymerase. D) Partial DNA sequences of primer and template for the incorporation of dA**TP and dATP. Evaluation of the incorporation efficiency using dA**TP (■, dashed line)/dATP (•, solid line) mixtures and *KlenTaq* DNA polymerase. E) Partial DNA sequences of primer and template for the incorporation of dT**TP and dTTP. Evaluation of the incorporation efficiency using dT**TP (■, dashed line)/dTTP (•, solid line) mixtures and *KlenTaq* DNA polymerase.

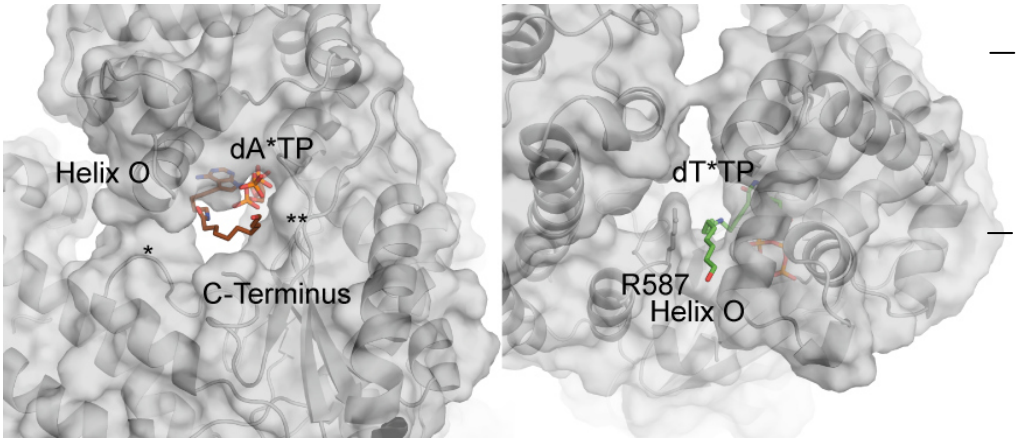


Figure S2 Detailed perspective of the cavities of KlenTaq$_A$ (left) and KlenTaq$_T$ (right). In case of the dA*TP, elements of different secondary structures form the hollow, including Helix O and loops 582-587 (marked *) 819-822 (**) as well as the C-terminal portion. Concerning the dT*TP, Helix O and R587 shaping the breach for the linker exit.

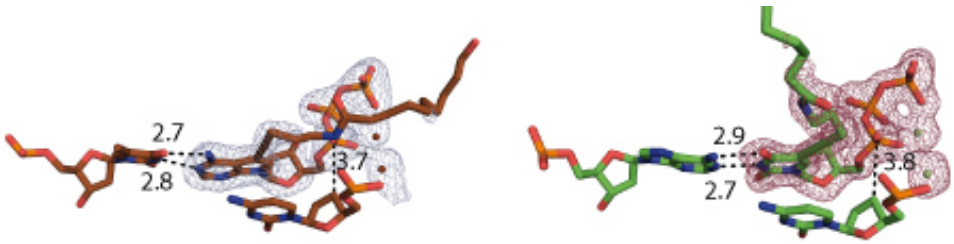


Figure S3 Snapshots of the active center. Distances from the 3' carbon of the DOC to the α-phosphate (left dA**TP, right dT**TP) as well as hydrogen-bonding with the template, omit density for the triphosphates (left dA**TP, right dT**TP) at 3σ

2.2 " Structures of KOD and 9° North Polymerases Complexed with Primer Template Duplex"

2.2.1 Introduction to " Structures of KOD and 9° North Polymerases Complexed with Primer Template Duplex"

Belonging to the largest family of DNA polymerases, α-like or B-family polymerases can be divided into 2 subgroups: protein- and RNA primed enzymes.[26,34,120] Protein-primed polymerases are active in the replication of linear genomes and use the hydroxyl function of Ser, Thr or Tyr residues as starting point for the DNA synthesis. The protein primer remains covalently bound to the 5' end, the so-called terminal protein (TP).[120-123]

The second phylogenetic group contains the major eukaryotic replicative enzymes (Polymerase α,δ,ε), archeal and bacteriophageal enzymes as well as Pol II from γ-proteobacteria.[30,34]

A structural characterisation of the nucleic acid primed enzymes showed a similar overall disk/donut shape. which is further subdivided into a N-terminal domain, a 3'-5' exonuclease activity and the three subdomains of the polymerase function (thumb, palm, finger).[48,124-129]

Hyperthermophile archeal member of this family are of special interest, due to their inherent thermostability and the high incorporation rates (e.g. KOD1 polymerase 100-130 bases per second).[130] These enzymes display a strong 3'-5' proofreading ability[30] and some show the ability to detect deaminated bases and consequently stall replication. The underlying structural principles for this detection have been shown recently, identifying a pocket in the N-terminal domain.[131,132]

2.2.2 Goals of the project

Due to their accuracy, high incorporations rates and tolerance towards modified substrates, archeal B-family polymerases are mainly used in biotechnical application. Despite this general interest, structural data on replicating archeal polymerases is rare. At the beginning of the project, all available structures of (nucleic-acid primed) B-Family polymerases in replication mode were of mesophilic origin.[124,125,133] Prior to be able studying the substrate tolerance *in cristallo*, the crystallization setup to generate DNA complexed crystals has to be established. The presented study is

concentrated on polymerases of *Thermococcales* (namely strains *T.9°N-7* and *T.kodakaraensis*) and was performed in collaboration with K. Betz.

2.2.3 Structures of KOD and 9° North Polymerases Complexed with Primer Template Duplex

Reproduced with permission from
"Bergen, K., Betz, K., Welte, W., Diederichs, K. and Marx, A. (2013), Structures of KOD and 9°N DNA Polymerases Complexed with Primer Template Duplex.
ChemBioChem, 14:1058–1062. doi:10.1002/cbic.201300175"

Structures of KOD and 9°N DNA polymerases complexed with primer template duplex

Konrad Bergen‡, Karin Betz,‡ Wolfram Welte, Kay Diederichs and Andreas Marx*

[a] *K. Bergen, K. Betz, Prof. W. Welte, Prof. K. Diederichs, Prof. A. Marx*
‡ Authors contributed equally
Departments of Chemistry and Biology, Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany
Fax: (+) +49 7531 885139

DNA polymerases are the key enzymes in many biotechnological applications such as genome sequencing, molecular diagnostics, DNA conjugation or selection of aptamers by systematic enrichment of ligands by exponential amplification (SELEX)[60-62]. Within these methods the ability of DNA polymerases to process modified 2'-deoxynucleoside-5'-O-triphosphates (dNTPs) is often the essential step[60,88-91]

Functional studies using nucleotides that are modified at the sugar residue[134] or the nucleobase[73,135,136] have shown many times that DNA polymerases from archaea belonging to the sequence family B (e.g. *Thermococcus* sp. 9°N-7 (9°N) DNA polymerase, *Pyrococcus furiosus* (Pfu) DNA polymerase, and *Thermococcus kodakaraensis* (KOD) DNA polymerase) are more efficient in utilizing modified nucleotides than the DNA polymerases from sequence family A like *Thermus aquaticus* (Taq) DNA polymerase. The origin of this feature remains widely elusive mainly due to the lack of structural data. While structures from family A DNA polymerases are available even in complex with modified nucleotides[64,78,137,138], family B DNA polymerases from archaea are only very poorly defined and only the apo-structures and those of the enzymes with the DNA placed towards the exonuclease subunit of the enzyme (editing mode) have been described so far.[128,131,139-145]

Here we report on two structures of family B DNA polymerases, KOD and 9°N that were obtained in the binary complex with primer and template strand positioned in the polymerase cleft in the replicative state. We were able to solve the crystal structures of binary KOD and 9°N complexes with resolutions of 2.4 Å and 2.6 Å, respectively.

An important step towards obtaining the binary structures was the identification of crystals that bear complexed DNA. It turned out that both enzymes readily crystallize as apo forms even though DNA is present. Utilizing a reported approach[146] in which Cy-labelled DNA assists to visualize and identify DNA containing crystals, led to success. Coloured crystals of both polymerases were obtained in different crystallization conditions and were reproduced with unlabelled DNA. In case of 9°N, resolution decreased slightly with the unlabelled template. As the primer/template (p/t) duplex part is identical in both structures (with labelled and unlabelled template) we use the Cy5 containing 9°N structure for analysis in this paper. In contrast, refined crystals of KOD delivered better diffraction characteristics without the dye, thus, the structure with the unlabelled template was used for further discussion.

Despite that magnesium ions and ddNTP were present in the crystallization trials, no electron density for a bound triphosphate and/or coordinated magnesium ions were found in the active sites of both enzymes. The most probable reason for this is the fact that in both structures the tip of the finger domain is involved in a crystal contact with the exonuclease domain of a symmetry related molecule. Furthermore, a second symmetry mate could sterically hinder finger domain closure (Figure S1).The overall structures of KOD and 9°N are very similar (see Figure S2).They show the typical domain composition of DNA polymerases, comprising N-terminal (N-term, 1-130 and 338-372), exonuclease (exo, 131-338), finger (448-499), thumb (591-774) and palm (374-447 and 500-590) domains.[147] (Figure 1A, depictions of 9°N are in the supporting information).

While in the description of the KOD apo structure the thumb domain has been further divided into the two subdomains thumb-1 and thumb-2[140] we will not discriminate between these two subdomains in our description. When comparing both binary structures with the KOD apo structure (PDB ID: 1WNS[140]) some rearrangements in the palm and thumb domains of KOD and 9°N are observed upon DNA binding. The KOD binary and apo structures align with a Cα RMSD of 1.31 Å. An overlay of the structures is shown in Figure 1A. The most pronounced movement of structural

elements upon DNA binding are performed within the thumb domain which rotates towards the DNA (Figure 1). In the following discussion of the structures we draw on the nomenclature introduced for the 9°N apo structure.[126] Reordered and additionally built structural elements within the thumb domain include the helices R, S, T, U, V+W, the β-sheets 26-28 and connecting loops. The biggest movement is performed by helix U which shifts around 22 Å and ~27° towards the DNA. In both structures this inward movement generates a distinct contact region between thumb and DNA whereby the otherwise only partly resolved thumb domain becomes more structured (Figure 1). The biggest rearrangements besides the thumb domain are made by helix L in the palm domain (residue 374-379) that is located at the interface to the finger domain and the loop connecting it to β-sheet 18 (residue 397-404). Starting at Val 389, the loop together with the helix move by up to 4 Å outwards, thereby opening the template cleft (Figure 1, in dark green).

In both structures the duplex part of the DNA is bound almost identically in a groove formed by the thumb and palm domain. The protein contacts the double stranded DNA (dsDNA) between nucleotide dG_{T13} and dC_{T6} on the template strand and dG_{P7} and ddA on the primer strand (numbering of p/t see Figure 2 and S2). The DNA duplex is mainly contacted by residues of the thumb domain as well as some residues of the palm and one residue of the exonuclease domain. The protein residues directly interacting with the DNA are identical in both enzymes with only one exception which is caused by the sequence differences between KOD and 9°N (Figure S2 A). An overlay of the two binary structures is shown in Figure S2 B. Based on the high structural and sequence similarity of both enzymes (Cα RMSD: 0.76 Å, sequence identity: 91%) the structure of KOD was chosen for a more detailed description in the following analysis as representative of both polymerases.

As in the extensively studied B-family polymerase RB69 the p/t duplex in KOD and 9°N binary structures maintain a B-form conformation with most ribose moieties showing the ideal puckering for B-DNA (C2‘-endo) or conformations which are as well found in more flexible B-DNA (C1‘-exo, C3‘-exo, O4‘-endo and C4‘exo conformations)[16]. The duplex also follows other known geometric characteristics for B-DNA. As already stated for other binary or ternary B-family structures, this is in contrast to the A-form DNA observed at the primer 3‘ end in A-family polymerases as reported for T7, BF *(*large fragment of *Bacillus stearothermophilus* polymerase) and

KlenTaq (large fragment of the polymerase I from *Thermus aquaticus)* DNA polymerases.[40,99,148]

Similar to other B-family polymerases[133] KOD forms direct contacts with the DNA duplex primarily via the DNA phosphate backbone. Only very few direct interactions with the nucleobases or sugar oxygens are observed (Figure 2). Residues in the loop connecting β-sheet 26 and helix U (666-677) in the thumb domain establish the first contacts to the primer strand. A second loop (residues 606-616) between β-sheet 25 and helix R interacts with the primer via the phosphate backbone and the minor groove. The 3' part of the template is contacted by residues (709-711) in the connecting loop between β-sheet 27 and 28 and by residues in helix V+W. Some residues known to be important in DNA binding are situated in motifs conserved throughout family B polymerases, this includes for example the sequence motif KKRY (in KOD residues 591-594, KKKY), which is unique to family B DNA polymerases[149] and mediates contacts to the p/t duplex near the active site. A second conserved region (I/YxGG/A motif) is situated in the previously described loop in the palm domain and stabilizes the template strand at the 5'-end of the duplex (Figure 2, dark green). Apart from the residues located in conserved motifs most of the other residues involved in DNA binding are not conserved throughout the whole B-family[133]. However, we found that a great majority of residues that directly contact the DNA duplex are indeed well conserved in archeal members of the family (Figure S3).

The obtained structural insights in this study may aid in rationalizing the observed differences in the capability of family B and A DNA polymerases in polymerizing chemically modified nucleotides. Thus, we compared the herein reported data with a representative of the family A polymerases. We use the structurally well characterized KlenTaq DNA polymerase[39,64,83,84,87,98,99,137,138] as model (Figure 2 and 3). In both enzymes the majority of direct protein side- or main chain contacts to the dsDNA are mediated via the phosphate backbone. The most striking difference, however, is the number of direct contacts to the nucleobases and sugar oxygens. Whereas in KlenTaq six nucleobases are contacted by four protein side chains in the KOD structure only three nucleobases are contacted by 2 residues. Interestingly, in both enzymes interactions with the nucleobase of the two primer nucleotides 9 and 8 are observed by one protein side chain (Lys 540 in KlenTaq and Arg 612 in KOD). Furthermore, remarkably fewer interactions with the ribose moiety are found in KOD

compared to KlenTaq. Whereas in KlenTaq six interactions of sugar oxygens with protein side chains are observed only two specific residues in KOD interact with the sugar (Asp 540 and Arg 612).

As sugar and nucleobase contacts are all localized in the minor groove one would assume a sterically less hindered minor groove in case of B family polymerases. This has been probed by modified nucleotides and might explain why B-family polymerases show better performance in utilizing nucleotides that e.g. bear small C4'-modifications (i.e., methyl, ethyl, vinyl, methylene, etc.).[150-156] Nevertheless, bulkier modifications might not be tolerated by both enzymes since they would interfere with residues of both proteins that are in farer distance to the interior of the minor groove (Figure S4). In general the minor grooves of the DNA complexed by both enzymes are better covered by protein residues than the major grooves.

Modifications at C5 in pyrimidines and C7 of 7-deazapurines, that have been shown to be more efficiently polymerized by family B DNA polymerases[60,65,66,68-70,73,94,150-152,157-166] point towards the major groove. Thus, we next inspected at the confines of the major groove of the DNA in the KOD and KlenTaq binary structures. The number of residues interacting with the phosphate backbone is similar and also in the 3D model the major grooves seem to be well accessible in both enzymes (Figure 3 and S4 C,D). However, one major difference is the tip of the thumb domain. In both enzymes the tip of the thumb domain (residues 506-509 in KlenTaq and residues 668-675 in KOD, marked in red in Figure 3A,B and S4 C,D) undergoes contacts with the primer strand. The architectures of the contacting loops differ greatly. While the loop in the KOD structure hovers above the minor groove without extending deeply into it (Figure 3A, S4 C), the corresponding loop in the KlenTaq extends over the phosphate backbone towards the major groove (Figure 3B, S4 D). The better accessibility of the major groove might by a causative for the observed superiority of family B DNA polymerases in polymerizing nucleotides that bear bulky modifications at C5 in pyrimidines and C7 of 7-deazapurines.[55,67,76,79-83,140-142,147-156]

In summary, we describe the first two crystal structures of archeal family B DNA polymerases in a binary complex bearing the DNA in the replicative mode. Since up to now only structures of bacterial or eukaryotic members of the family in a binary or ternary replicative complex are described these structures contribute to a better mechanistic understanding of thermophilic B family polymerases of archaea that are

extensively used when modified nucleotides are used as substrates. With these new models at hand the generation of optimized enzymes and substrates will be facilitated. Despite this progress, one remaining challenge is the crystallization of ternary complexes of archeal B family polymerases with dsDNA and bound triphosphate in the active site. Ternary structures would help to elucidate the mechanisms of selectivity, processivity and fidelity of these enzymes which are widely used in biotechnology and molecular diagnostics and to elaborate their incorporation properties with specific dNTP modifications.

Acknowledgements

We thank the staff at the beamlines PXI and PXIII at the SLS for their support. We also acknowledge support by the Konstanz Research School Chemical Biology and DFG

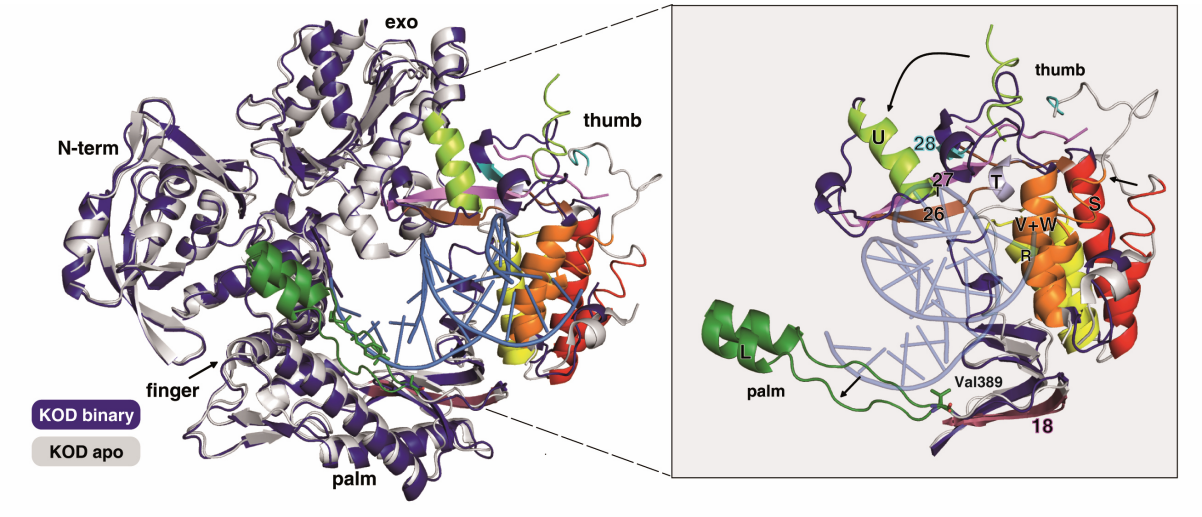


Figure 1. Overlay of KOD DNA polymerase in the binary complex (violet) with the KOD apo structure (PDB ID: 1WNS, grey) shown as cartoon. Polymerase domains are labelled. Zoom in: detailed view on the altered elements of thumb and palm domain. Corresponding secondary structure elements of apo and binary structures are shown in the same colours to indicate the movement of the thumb domain upon DNA binding and visualize remodelled parts of the enzyme.

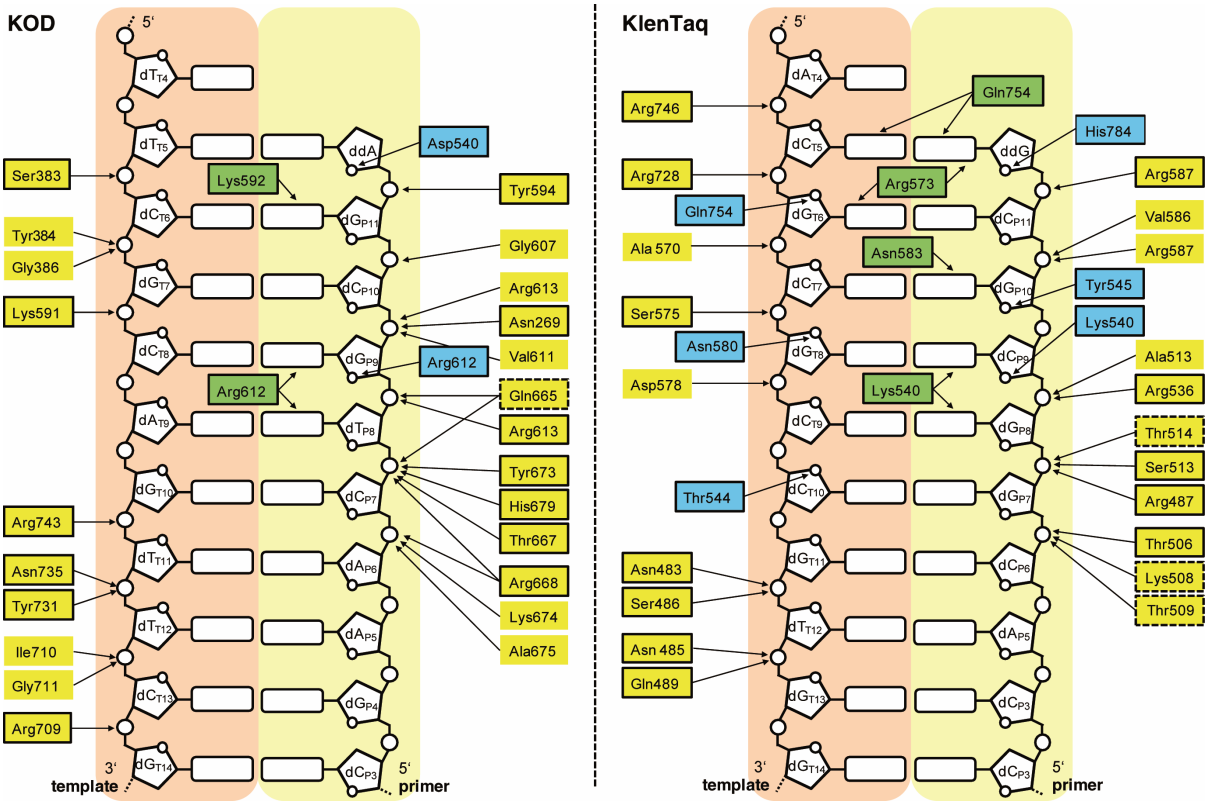


Figure 2. Comparison of protein-DNA interactions in KOD (B-family) and KlenTaq (A-Family) polymerases. Only direct contacts up to a distance of 3.6Å are shown. Side chain interactions are marked with a solid lining, contacts with the protein backbone are shown without lining and residues where both interactions are found are shown with dashed lining. Interactions to the phosphate backbone, sugar oxygen or nucleobase are shown in yellow, blue and green, respectively.

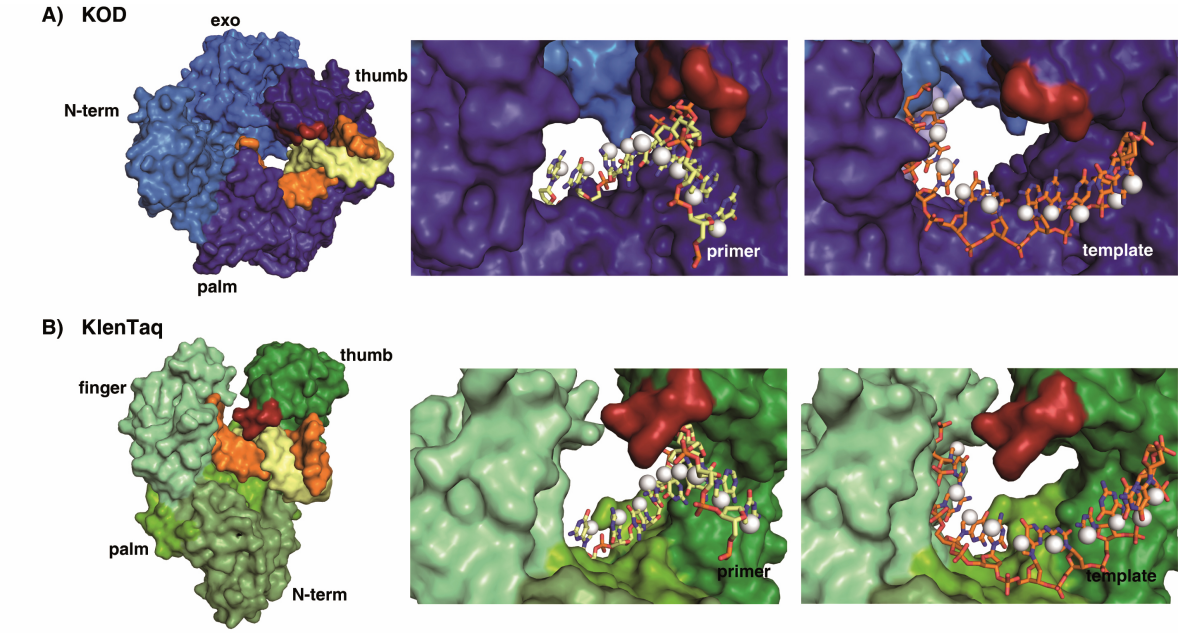


Figure 3. Comparison of DNA environment in family B DNA and family A polymerases. KOD A) and KlenTaq (PDB ID: 3SZ2) B) binary complexes are shown in surface representation with domains coloured in different shades of green and blue, respectively. The finger domain of KOD is in the back and not visible in this representation. DNA primer and template are shown as yellow and orange surface or sticks. Middle and right panel show a zoom into the primer and template binding cleft. The C5 in pyrimidines and C7 of 7-deazapurines nucleobases, which are mainly used for attachment of modifications, are marked as white spheres.

2.2.4 Supporting Information

Structures of KOD and 9°N DNA polymerases complexed with primer template duplex

Konrad Bergen, Karin Betz, Wolfram Welte, Kay Diederichs, and Andreas Marx *

Enzymes, oligodeoxynucleotides, nucleotides

Primer and templates were purchased from *ThermoFisher* (HPLC purified), ddNTPs from *JenaBioScience*, natural dNTPs from *Fermentas*

Purification

9°N: For crystallization experiments with *Thermococcus* sp. 9°N-7 DNA polymerase the polymerase mutant (D141A, E143A) with reduced exonuclease activity was used.[167] The exo- enzyme is named 9°N throughout the paper. 9°N was cloned into a pET21b vector and overexpressed in E.coli BL21(DE3). Expression was induced by addition of 1mM IPTG final concentration at OD_{600}=0.6-0.8. Cells were harvested after 4h, resuspended in lysis buffer (50mM Tris pH 8.55, 10mM $MgCl_2$, 16mM $(NH_4)_2(SO_4)$, 0.1% Triton X-100 and 0.1% Thesit) and lysed for 1h at 37°C by adding Lysozyme to 0.5mg/ml final concentration. After heat denaturation (10min at 75°C) 1mM PMSF was added and denatured proteins and cell debris were pelleted by ultracentrifugation (1h, 35000g). The supernatant was loaded onto a Heparin column with a low salt buffer (100mM NaCl, 20mM Tris pH 7.5, 0.1mM EDTA, 1mM DTT, 10% Glycerol) and proteins were eluted by increasing concentration of NaCl using a step gradient. Purest fractions (as determined by SDS PAGE) were pooled and concentrated. The protein was further purified by gel filtration chromatography using a Superdex-200 column and the following buffer: 200mM NaCl, 20mM Tris pH 7.5, 0.1mM EDTA, 1mM DTT, 10% Glycerol. After purification the protein was concentrated to 7mg/ml and stored at 4°C.

KOD: For the crystallization experiments the exonuclease deficient mutant D141A, E143A was used. The protein was expressed in E.coli BL21(DE3) using a codon-optimized sequence in a pET24 as expression system (GeneArt). Cells were induced

at OD_{600} ~ 0.6 using 1mM IPTG final concentration in LB medium. Purification was performed as previously described[140], with the addition of an anion exchange step, using the flowthrough for further purification. This was followed by a step using Superdex 200 SEC. The eluate was concentrated to ~10mg/ml and stored at 4°C.

Crystallization

For the identification of DNA complexed crystals a method described by Jiang&Egli[146] was successfully used for KOD and 9°N. In this approach dye-labelled DNA helps to visually distinguish between DNA bound and apo crystals. In our crystallization setups we used a DNA duplex consisting of a 16mer template and an 11mer primer with a Cy5 dye label attached to the single stranded 5'-end of the template strand. The primer end was enzymatically terminated by adding a 2',3'-dideoxynucleoside-5'-O-triphosphate to the mixture of polymerase and the primer/template (p/t) complex. Though the described approach delivered positive hits for both enzymes (DNA bound polymerase structures) additionally false positive coloured crystals were found. In these crystals the colouring resulted most probably from unspecifically bound DNA in the crystal lattice.

9°N: For crystallization setups 9°N was mixed with annealed p/t DNA (primer: 5'-CGC GAA CTG CG-3'; template: 5'-(Cy5) AAA GGC GCA GTT CGC G-3') and ddCTP in a molar ratio of 1:1.5:10 and incubated for 30min at 37°C. $MgCl_2$ was added to a final concentration of 20mM. The final protein concentration after adding all components was 4.5mg/ml. Crystallization setups were made using the vapour diffusion sitting drop method with the help of a Gryphon robot (ARI robots) and commercially available screens. Diffracting crystals grew in a drop containing 0.2M Lithium chloride and 2.2M ammonium sulfate. For cryoprotection crystals were soaked in the mother liquor containing 20% ethylene glycol.

Crystals were reproducible with an unlabelled p/t using the same sequence as above but those crystals showed lower diffraction quality. Therefore we used the Cy5-containing structure in this paper.

KOD: Protein was mixed with a primer/template (primer: 5'-CGC GAA TTG CG-3'; template: 5'-(Cy5) AAA TTC GCA GTT CGC G-3') complex in a 1:1.5 molar ratio in presence of 20mM $MgCl_2$ and incubated for 30 minutes at 37°C. A five molar excess

of ddATP was added and again incubated for 30 minutes at 37°C, leading to a final protein concentration of about 7 mg/ml. Screening setups were performed using a Gryphon robot (ARI Robots) and various commercially available screens. In order to separate the numerous apo crystals forms from DNA complexed ones, a method described by Jiang&Egli[146] was used. Identified conditions were refined in a grid screening varying pH and precipitant and reproduced without the dye. Diffracting crystals were obtained from a condition containing 50mM sodium cacodylate pH 8, 10% v/v 2-propanol, 20mM $MgCl_2$, 1 mM cobalt(III)-hexamine chloride, 1mM spermine (derived from condition G3 Natrix Screen, Hampton Research). Crystals were cryoprotected by soaking in the mother liquor containing 20% ethylene glycol and flashfrozen in liquid nitrogen.

Data collection and processing

Datasets were collected at the beamline X06SA-PX of the Swiss light Source (SLS) at the Paul Scherrer Institut in Villigen, Switzerland. Data reduction was done using the XDS package[168,169]. Statistics of data collection and refinement for both structures are given in Table S1. Simulated annealing electron density omit maps for the DNA are shown in Figure S3.

9°N: Data reduction was done in space group $P22_12_1$ with the cell dimension a=112.2 Å, b=142.6 Å, c=66.7 Å and α,β,γ=90.0°. The structure was solved by molecular replacement using the 9°N apo structure (PDB ID: 1QHT[139]) as search model. The structure was improved by alternated refinement with PHENIX[170] and model building in COOT[171]. Refinement was accompanied by validation using the Molprobity server[172,173]. The restraint file for the Cy5 dye was created using the Grade Web Server.[11]

Data was used in refinement up to 2.28Å, where the CC1/2 value[174] approaches 50%. In the highest resolution shell (2.42Å - 2.28Å) the I/σ is 0.82 and R-meas is 282.1%. To facilitate comparison with other deposited structures, we also give account of the resolution which corresponds to an I/σ value of around 2. This resolution, which we call the I/σ=2 resolution, is 2.6Å for the 9°N binary structure.

KOD: Data reduction was done in space group $P22_12_1$ with the cell dimensions a=144.9 Å, b=111.8 Å, c=67.1 Å and α,β,γ=90°. In this case three datasets were merged using XSCALE. The structure was solved by molecular replacement using the KOD apo structure (PDB ID: 1WNS)[140] as search model. Difference density guided modelling of the DNA was performed using COOT[171]. The structure was improved by alternated refinement with PHENIX[170] and model building in COOT[171]; structure quality was surveyed using the Molprobity server[172,173]. For KOD, the I/σ=2 resolution is ~2.4Å.

Structural details 9°N

In case of the 9°N structure the Cy5 dye is well ordered and thereby visible in the electron-density. The four single stranded nucleotides at the 5' end of the template show an interesting arrangement reminiscent of the observation in previously reported KlenTaq structures with a 3'-dT-dA-dA-dA-5' and 3'-dNaM-dA-dA-dA-3' single stranded template overhang.[87] In the present case the templating nucleotide dG_{T4} (numbering of nucleotides see Figure 2) is flipped out of the stacking arrangement and the subsequent nucleobases dA_{T3} and dA_{T2} stack on top of the last formed basepair. The last nucleotide dA_{T1} and the attached Cy5 dye interact with the exonuclease domain. The dye points towards a symmetry mate where it is stabilized between Arg 99, His 103 and Pro 104. Due to this arrangement the 5'-overhanging template was stabilized but in a conformation probably not occurring in solution though favourable for crystal formation. In proximity to the nucleotide dA_{T2} positive difference density was observed. Alternative conformations of residues dA_{T2} and Cy5 can explain this density. We modelled a second conformation of dA_{T2} by rotating the nucleotide approximately 150° around the C3' sugar atom and a second conformation of Cy5. The occupancy of the alternate conformations is below 50% for both residues, therefore in some cases the position is not occupied. In all figures we only show the position of the dA_{T2} and Cy5 conformation with the higher occupancy.

Compared to the apo 9°N structure published in Beese et al. in 2000[139], we could model some additional regions of the thumb domain. These regions are P569-L574, E648-L659, T667-K674, G706-I710, D721-A729 and A751-Y757. Our 9°N model is complete up to the last 23 C-terminal residues which were not resolved in the electron density.

Structural details KOD:

Due to the stabilizing effects of the DNA binding, electron density allowed a rebuilding of the thumb 1 and 2 domains in KOD, including regions T667-G677, R688-G696, K705-D712 and A747-F748, which were missing in the apo structure. The binary KOD model could be built up to residue 756, leaving the last 18 amino acids unresolved. The binary KOD model could be built up to residue 756, leaving the last 18 amino acids unresolved. The 5' overhanging template nucleobase in the KOD structure stacks above the last basepair and the binding pocket for the incoming nucleotide is empty. The template base dA_{T3} is extending to the T-cleft and the subsequent nucleotides dA_{T2} and dA_{T1} are not resolved.

When comparing the two enzymes concerning the interaction pattern, diverging residues in S2 are with the exception of S383 in KOD (G383 in 9°N) due to the 3.6Å cut off used for mapping.

All figures showing structures in this paper were prepared using PyMOL[175].

Table S1: Summary of data collection and refinement

	9°N	**KOD**
PDB ID		
Data collection		
Wavelength	0.97793Å	1.00002Å
Spacegroup	18 ($P22_12_1$)	18 ($P22_12_1$)
Cell dimensions		
a, b, c (Å)	112.21, 142.62, 66.70	144.87, 111.78, 67.10
α,β,γ (°)	90.0	90.0
Resolution I/σ-2 (Å)	2.6	2.4
Total reflections*	330620 (54378)	803790 (46136)
Unique reflections*	49600 (7014)	49755 (3506)
R_{meas} (%)* +	17.2 (282.1)	13.5 (196.5)
I/σI *	9.07 (0.82)	16.65 (1.81)
Completeness (%)*	99.9 (99.9)	99.7 (96.7)
Refinement		
Resolution (Å)[$]*	2.28	2.29
No. unique reflections*	49542	49743
R_{work}/R_{free} *[+]	19.6/24.2	18.1/21.8
B-factors		
DNA	78.8	51.9
Protein	62.9	46.4
R.m.s deviations		
Bond lengths (Å)	0.006	0.003
Bond angles (°)	1.09	0.72
Ramachandran[#] (%)		
Favored	95.91	97.75
Allowed	3.69	2.25
Outlier	0.4	-

*Numbers in brackets refer to the highest resolution shell.
+ for definition of R_{meas}, see [119]
#as determined by MolProbity[172,173]
[$] Data used in refinement up to a CC(1/2) of around 50%, for details see above and[174].

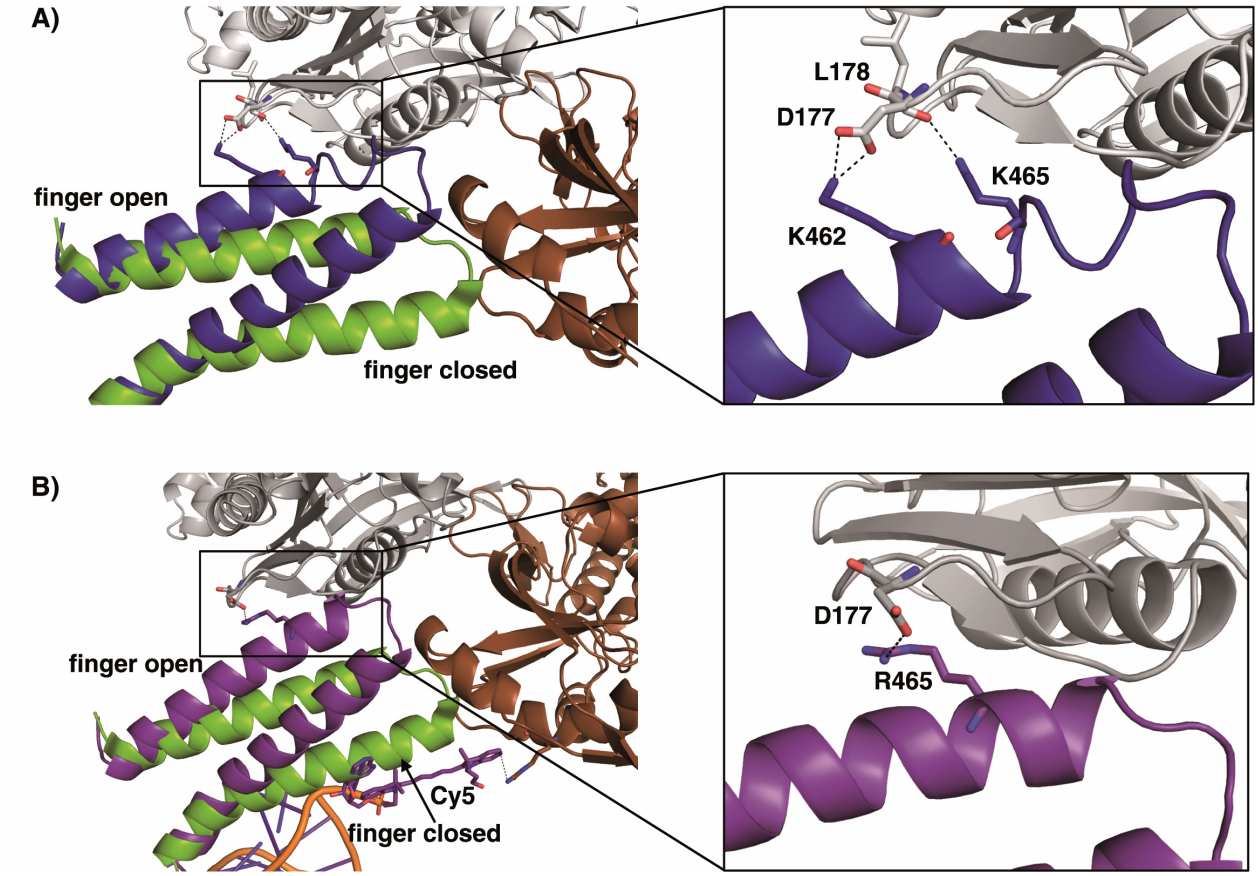


Figure S1: Crystal contacts of the finger domain in KOD and 9°N binary complexes which hinder finger domain closure. Finger domains are shown as cartoon in A) KOD (blue) and B) 9°N (purple). The position of the closed finger domain from the superposed ternary structure of DNA polymerase delta (PDB ID: 3IAY) is shown in green. Two different symmetry mates near the finger domain are shown in brown and grey, respectively. A detailed view on specific hydrogen bonds between the open finger domain and a neighbor molecule is shown for both structures. Residues taking part in hydrogen bonding are shown as sticks and bonds are indicated as dashed lines. B) The Cy5 dye in the 9°N binary structure is shown as sticks and its interaction to a symmetry neighbor is visible.

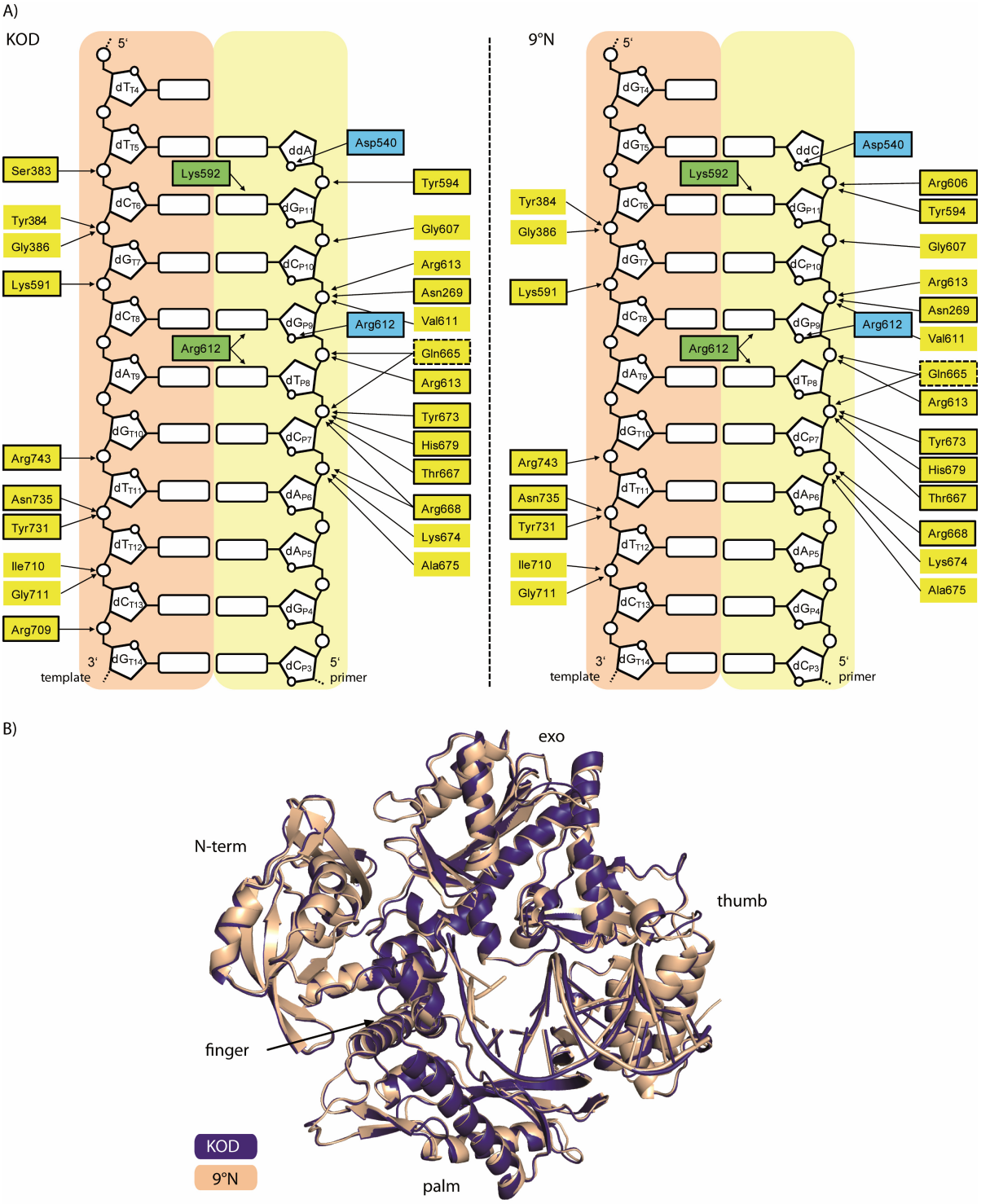


Figure S2: A) Comparison of protein-DNA interactions in KOD and 9°N binary structures. Only direct contacts up to a distance of 3.6Å are shown. Side chain interactions are marked with a solid lining, contacts with the protein backbone are shown without lining and residues where both interactions are found are shown with dashed lining. Interactions to the phosphate backbone, sugar oxygen or nucleobase are coloured yellow, blue and green, respectively. B) Overlay of the binary structures of KOD (blue) and 9°N (sand).

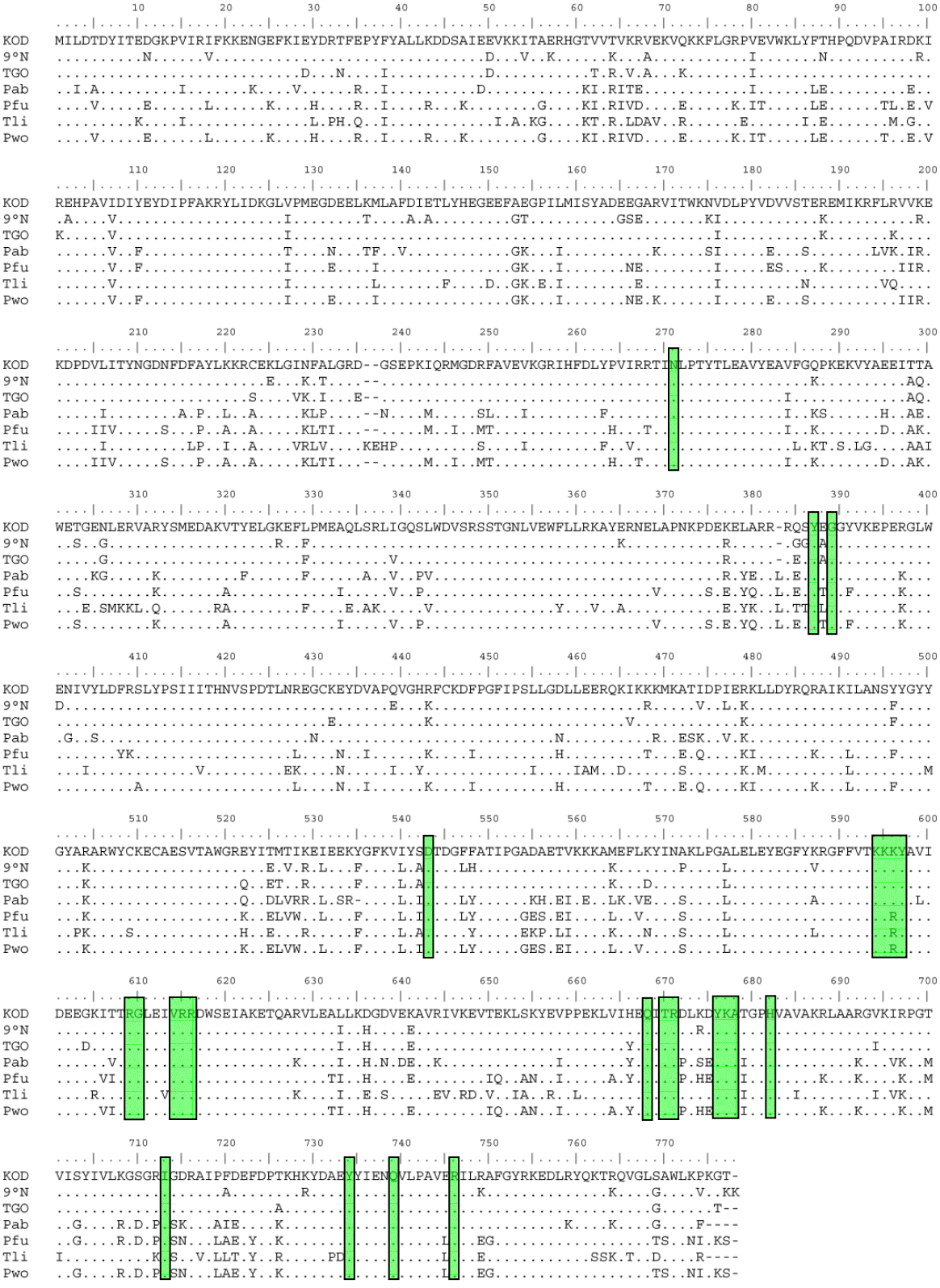


Figure S3: Sequence alignment of archeal family B DNA polymerases. Highlighted residues show amino acids in involved in DNA binding in the binary complexes of 9°N and KOD that are conserved in all aligned archeal polymerases. Points indicate sequence identity. The graphic was made using BioEdit[176]

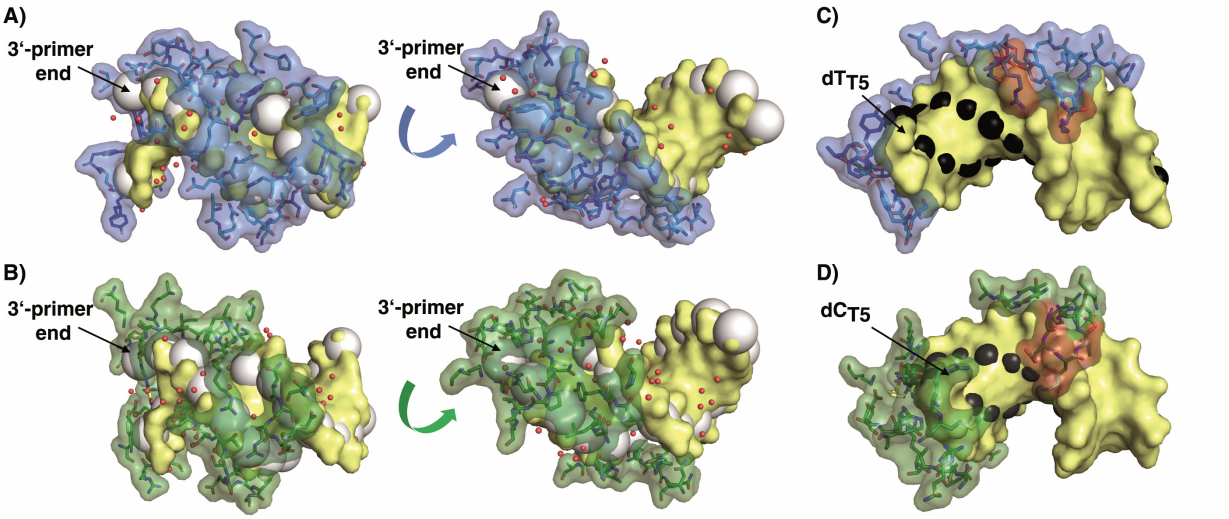


Figure S4: Comparison of minor and major groove of KOD (blue, A) and C)) and KlenTaq (green, B) and D)) DNA polymerases. Residues in a distance up to 4.5Å around the DNA duplex are shown as sticks and surface. The DNA duplex is shown as yellow surface; the 5‘ single stranded template nucleotides are not shown. A) and B) show the minor groove from 2 different orientations. Water molecules are shown as small red spheres. The sugar 4‘atoms are marked as white spheres with a radius of 2 Å. Protein side chain which do not directly interact with the dsDNA but cover the minor groove are: Tyr 384, Val 389, Tyr 494, Thr 541, Thr 590, Lys 591, Lys 593, Asp 614, Trp 615, Glu 609, Glu 664, Ala 675, Thr 676, Pro 678. C) and D) show the major groove of the dsDNA. Nucleobase atoms C5 for pyrimidines and N7 for purines are indicated by black spheres. The residues located at the tip of the thumb domain and interacting with the phosphate backbone near the 5‘end of the primer are colored in red.

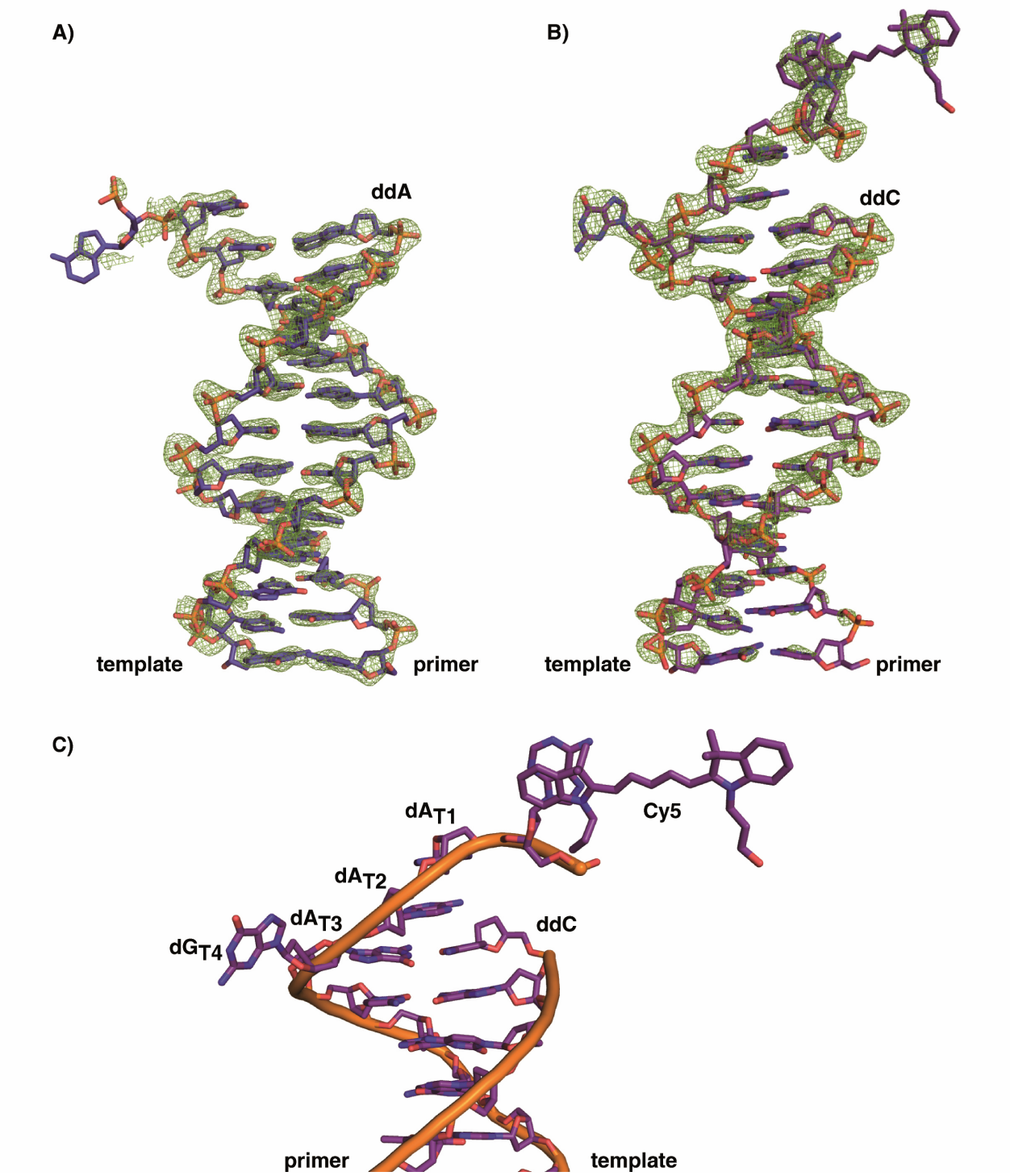


Figure S5: A,B) Simulated annealing omit maps of DNA primer and template in KOD (blue) and 9°N (purple) are shown at 3σ. C) Arrangement of the single stranded DNA template in the 9°N binary structure shown as sticks. The first single stranded nucleobase (G_{T4}) is flipped out of the stacking arrangement of the duplex. The following two bases (A_{T3} and A_{T2}) are flipped back and stack on the last base pair. The last template nucleotide as well as the attached Cy5 dye point towards the exonuclease domain and the Cy5 dye is stabilized additionally by a symmetry related molecule where the dye is positioned between Arg 99 and Pro 104. For Cy5 and dA_{T2} in the 9°N structure only the conformations with the higher occupancy are shown.

2.3 *"Structure and Function of an RNA-reading thermostable DNA polymerase "*

2.3.1 Introduction to "Structure and Function of an RNA-reading thermostable DNA polymerase"

Reverse transcription is the essential step of the retroviral reproduction cycle. In this process the RNA contained in the viral capsoid is transcribed into doublestranded DNA while the RNA portion is digested. This task is performed by a class of DNA polymerase which are characterized as reverse transcriptases (RT)[30].

This process can also be used to detect and quantitate transcripts and thereby monitor altered patterns of gene expression[177,178].

2.3.2 Goals of the Project

Gain of function by mutation experiments often reveal a widened substrate spectrum of the enzymes. As *KlenTaq* is a known target for the above mentioned studies[179-182], the reshuffling of two mutants by N. Blatter identified a mutant capable of elevated reverse transcriptase activity.[183] Our aim was a further characterization of the altered enzyme on a structural level. Structural snapshots of evolved enzymes allow tracking of the original fidelity as well as insights on the newly aquired capability and therefore the possibility to a more rationalized approach in a structure guided design of further improved enzymes.

2.3.3 Structure and Function of an RNA-reading thermostable DNA polymerase

Reproduced with permission from
„Structure and Function of an RNA-Reading Thermostable DNA Polymerase“
Blatter, N., Bergen, K., Nolte, O., Welte, W., Diederichs, K., Mayer, J., Wieland, M., Marx, A.
Angewandte Chemie International Edition, Volume 52, Issue 45, pages 11935-11939, November 4, 2013

Structure and Function of an RNA-reading thermostable DNA polymerase

Nina Blatter[1], Konrad Bergen[1], Wolfram Welte[2], Kay Diederichs[2], Jutta Mayer[3], Markus Wieland[3] & Andreas Marx[1,3]

[1]Department of Chemistry, [2]Department of Biology, and [1,3]Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstrasse 10, D 78457 Konstanz, Germany

[3]Prolago Biotec, Universitätsstrasse 10, 78457 Konstanz, Germany

To whom correspondence should be addressed:

Prof. Dr. Andreas Marx
Department of Chemistry and
Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstrasse 10
D 78457 Konstanz, Germany

Tel.: +49 7531 885139
Fax: +49 7531 885140
Email: andreas.marx@uni-konstanz.de

Keywords: directed evolution · DNA polymerase · polymerase chain reaction · reverse transcription · structural biology

The high substrate specificity of DNA-dependent DNA polymerases is essential for genome stability as well as many biotechnological applications.[184] Especially in cells, the discrimination between ribo- and deoxyribonucleotides and between RNA and DNA is fundamental since the concentration of ribo- exceeds the one of desoxyribo-analogues by far. Although the selection mechanisms for nucleotide incorporation have been intensively investigated for DNA and RNA polymerases[185-193], much less is known on how DNA-dependent DNA polymerases discriminate between the different nucleic acid templates (DNA vs RNA). Some viral DNA polymerases (i.e., reverse transcriptases) are able to utilize both DNA and RNA as a template for nucleic acid synthesis. Crystal structure analysis of these enzymes, complexed either to a RNA or DNA template, have contributed significantly to our understanding of this process.[193-198] However, similar structural data of DNA-dependent DNA polymerases that have poor propensity to process RNA templates are lacking, presumably due to hampered crystallisation by the formation of unstable complexes leading to structural heterogeneity. Structure analysis of *KlenTaq* DNA polymerase, a shortened form of *Thermus aquaticus* DNA polymerase, has added significant contributions to the understanding of how DNA polymerases recognize the cognate substrate[38,39,64,74,84,86,87] process abasic sites[85,194,195] non-natural nucleotides.[74,84,86,87]

Since our attempts to obtain suitable crystals of KlenTaq complexed to RNA failed, we set out to engineer the enzyme in a way that it is capable of processing an RNA template more efficiently, which might result in improved crystallisation properties. Indeed, we were able to obtain a significantly improved KlenTaq variant from which we obtained structural insights of a DNA-dependent DNA polymerase while processing RNA as template for the first time. Furthermore, the generated KlenTaq variant turned out to be a thermostable DNA polymerase with significant reverse transcriptase activity resulting in a valuable tool for crucial applications in clinical diagnostics and molecular biology such as transcriptome analysis and pathogen as well as disease-specific marker detection. For the generation of an improved enzyme variant the mutation sites of two KlenTaq (KTq) variants M1 (L322M, L459M, S515R, I638F, S739G, E773G)182 and M747K179 were recombined by DNA shuffling. Both variants, M1 and M747K, were reported to possess either some reverse transcriptase activity or an expanded substrate spectrum. DNA shuffling was employed since KTq M1, previously evolved in error-prone PCR, comprises six mutations distributed over the enzyme scaffold, but the individual contributions of the mutations are unknown.

We generated a library of 1,570 clones in order to obtain high library coverage (as described in the Supporting Information). Proteins were expressed in 96-well plates and, after lysis and heat-denaturation of host proteins, used in a real-time PCR activity screen with DNA as template. As approximately 80 % of the expressed enzymes were found to be PCR active, we subsequently screened for reverse transcriptase activity based using MS2 RNA as template as described earlier.182 In this screening, the two variants RT-KTq 1 and RT-KTq 2 were identified and selected for further characterization as they showed significantly higher reverse transcriptase activity compared to the parental enzymes as well as a minimal number of mutations. RT-KTq 1 possesses three mutations (S515R, I638F and M747K), whereas RT-KTq 2 displays the three mutations of RT-KTq 1 and an additional one (L459M, Figure 1a). For comparison, we also combined all seven mutations in one protein scaffold (KTq M1/M747K) by site-directed mutagenesis. KTq wild-type, parental enzymes M1 and M747K, KTq M1/M747K and RT-KTq 1 and 2 were expressed, purified, adjusted to equal protein concentrations (Figure S1), and further characterized.

First, reverse transcriptase activity of the wild-type and all variants were examined in primer extension experiments with a 5'-[^{32}P]-radioactively labeled DNA primer annealed to an RNA template. Reactions were incubated at 72 °C for 1 min and subsequently analyzed on a denaturing PAGE (Figure 1b). Using RNA as template, we observed no primer elongation for the wild-type enzyme. KTq M747K incorporates up to three nucleotides, whereas KTq M1 was able to synthesize longer extension products up to seven nucleotides elongation. KTq M1/M747K and RT-KTq 1 and 2, however, showed almost quantitative formation of full-length product after 1 min, thus exceeding the parental enzymes by far (Figure 1b). Template-independent addition of one nucleotide, as described for DNA polymerases[85,194,196], was observed in case of full length product formation. Quantification of the respective activities[197] on a DNA- and RNA-template confirmed the superiority of the identified variants M1/M747K, RT-KTq 1 and 2 compared to the parental enzymes (Table 1). We found that the recombination of two KTq mutants resulted in a new generation of KTq variants which 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 (Table 1). If the effect of combining the mutations was additive, a two-fold increase in reverse transcriptase activity would be expected. However, our results suggest an interaction between the individual mutations in a non-additive

manner resulting in a synergistic effect of the mutations.[198] Due to the higher overall activity of RT-KTq 2 compared to RT-KTq 1 (Table 1) and a higher stability compared to KTq M1/M747K (Figure S2), we focused on RT-KTq 2 for further in-depth analysis. We next aimed at obtaining crystals of RT-KTq 2 for structural insights. Therefore, we crystallized RT-KTq 2 with a 11mer DNA primer and a 16mer RNA template and, for comparison, with the respective primer/template DNA duplex. Details on sample preparation, crystallization and data statistics are listed in the Supporting Information. We successfully solved structures of RT-KTq 2 bound to a DNA primer/template complex and an incoming 2',3'-dideoxycytosine-5'-triphosphate (ddCTP) at a resolution of 1.55 Å (termed RT-KTq 2_{DNA}, PDB ID 4BWJ).

The overall structure of RT-KTq 2_{DNA} is almost identical to closed ternary structures already described for KTq wild-type with a DNA-duplex[38,39,64,74,84,86,87] (e.g. PDB ID 3RTV[87]) and shows a low Cα rmsd of 0.33 Å (Figure S3a,b, in the Supporting definition). We were also able to obtain crystals of the enzyme variant in complex with a DNA/RNA hybrid at a resolution of 1.75 Å (termed RT-KTq 2_{RNA}, PDB ID 4BWM). This structure shows a halfopen state, presumably caused by crystal contacts. Interestingly, a comparison of RT-KTq 2_{RNA} with RT-KTq 2_{DNA} revealed major changes in the thumb domain due to the altered geometry of the hybrid duplex (Figure 2a). The RNA template thereby adopts a sugar puckering typical of A-form conformation, whereas the DNA primer reveals a variety of puckers deviating from the typical B-form (see Table S1 in the Supporting Information).[199] The terminal three base pairs of an DNA/DNA duplex, which are adjacent to the O-helix of the DNA polymerase, already adopt an A-form.[38] Therefore, structural changes in the protein scaffold were only required in regions neighbouring the nucleic acids downstream of these three terminal base pairs in order to adopt to the changed geometry of a hybrid DNA/RNA duplex (Figure 2b). The overlay of the all DNA and the hybrid duplex in RT-KTq 2_{DNA} demonstrates that part of the thumb domain would have clashed with the hybrid DNA/RNA duplex, which required further structural adaption of the protein scaffold. Consequently, we observe a reduced number of contacts compared to RT-KTq 2_{DNA} between the enzyme and the hybrid duplex, mainly in the 3' half of the template, and several changes in the interaction pattern (see Figure S4 in the Supporting Information). Remarkably, the major part of the reordered protein scaffold originates in the thumb domain, downstream of L459M in the αH (Figure 2c, see also

Figure S3c in the Supporting Information). This shift is probably caused by the mutation of L459 to the sterically less demanding and more flexible side chain of methionine that facilitates the bending of αH and the subsequent movement of αH1 and αH2 to come in proximity of the hybrid duplex, thereby maintaining the nucleic acid binding characteristics of this motif (Figure 2c).[38] This motif is stabilized by the S515R mutation located in αH2. It displays two conformations of the arginine residue in the RT-KTq 2_{RNA} structure, thus allowing interactions with the hydroxyl group of S513 and the peptide backbone of K505 (Figure 2d). The role of I638F is less obvious, though it probably accounts for a tighter folding of the finger domain due to hydrophobic and stacking interaction with neighbouring amino acids thereby resulting in a compaction of this region. These observations are further verified by an overall stabilization of the protein region compared to the RT-KTq 2_{DNA} structure as seen by the resolution of R515 and F638 in simulated annealing omit maps (Figure S3d,e). Finally, the mutation M747K enhances the positively charged surface potential in close proximity to the negatively charged backbone of the DNA.[194] This might promote binding of the nucleic acid backbone and thus, foster the ability to accept aberrant substrates. Interestingly, there is analogy to a finding, where a positive charge introduced into the thumb subdomain of a thermostable DNA polymerase allows more efficient processing of a non-cognate RNA primer/DNA template duplex by incorporation of ribonucleotide substrates.[200]

Next, the potential of RT-KTq 2 to be applied in molecular biology or clinical diagnostics was investigated by performing RT-PCR experiments. RNA is the template for transcriptome analysis, for pathogen detection as well as for disease-specific marker recognition in the reverse transcription polymerase chain reaction (RT-PCR).[177,178] This method is based on the detection and quantification of RNA by the enzyme-mediated reverse transcription of RNA to its complementary DNA and a subsequent amplification by PCR, thereby allowing monitoring the abundance of specific RNA molecules even in real-time fashion. Since the DNA polymerases employed in standard PCR are devoid of reverse transcriptase activity, two enzymes are required for RT-PCR: a reverse transcriptase and a PCR-competent DNA polymerase. Although these two enzyme mixtures are state of the art, several drawbacks arise from the heat-instability of commonly used retroviral mesophilic reverse transcriptases[201] such as unspecific priming, low yield on complex targets for example from secondary structureformation of the mRNA template and premature

reaction termination.[178] In addition, the reverse transcription step usually adds up to 30 min to the PCR protocol, a time addition which during outbreak situations – when hundreds of swabs need to be analyzed – provides a critical bottleneck. Thus, thermostable DNA polymerases capable of reverse transcription and PCR would supersede the use of two enzymes and provide time and work reduction. First, the purified enzymes were used in real-time RT-PCR experiments using MS2 bacteriophage RNA as template.[182] Under the conditions studied, RT-KTq 2 exhibits product formation approximately 10 cycles earlier than the parental enzyme KTq M1, whereas the wild-type enzyme has no detectable activity (Figure 3a). Formation of the expected product was confirmed by agarose gel analysis. Additionally, we demonstrated that RT-KTq 2 is able to generate significantly longer products than the parental enzyme KTq M1 (see Figure S5 in the Supporting Information). To investigate the application of RT-KTq 2 for the detection of various mRNA transcripts from extracted human total RNA isolated from Jurkat cells, products ranging from 71 to 111 base pairs (bp) were amplified by RT-PCR (Figure 3b). DNA contamination-based amplification could be excluded by designing intron-spanning products. Ultra-fast two-step cycling was employed in this experiment, with only 30 s initial denaturation at 98 °C and 30 cycles of two-step cycling with denaturation for 2 s and 15 s for combined annealing/extension at the temperature indicated. Also, an additional reverse transcription step was omitted,thus saving further time. Thus, RNA detection was achieved in approximately 20 min. Expected product formation was observed for all targets (Figure 3b). Additionally, we investigated RNA detection sensitivity of RT-KTq 2 performing real-time RT-PCR experiments. A 90 nucleotide (nt) RNA stretch of human β-actin mRNA was targeted and amplified from total RNA isolated from Jurkat cells (Figure 3c). Again, two-step cycling was used for PCR and product formation was visualized by binding of SYBRGreen I. We demonstrated that RNA detection was linear over a broad range of concentrations from 100 ng to as low as 1 pg of RNA, competing with commercially available enzymes (see Figure S6 in the Supporting Information).

Next, we set out to detect influenza virus A in TaqMan-based real-time RT-PCR using RT-KTq 2. Influenza viruses have been the cause for several pandemics[202], therefore rapid detection methods are required for identifying new outbreaks in time and subsequent medical intervention. As 5'-nuclease activity is needed for the hydrolysis of dual-labeled TaqMan probes, we used an enzyme blend of Taq DNA

polymerase wild-type and RT-KTq 2 in our assay. RNA extracts were from respiratory swab samples of patients known to be influenza A positive. A 141 nt RNA stretch coding for the influenza A matrix protein was targeted in real-time RT-PCR using two-step cycling. Furthermore, a TaqMan probe conjugated with minor groove binders[203] was used to increase reaction specificity and optimized primers recommended for the detection influenza A. An increase in the fluorescence signal was observed in all cases for seven different specimens; correct product formation was visualized on an agarose gel (Figure 3d).

In summary, we have report the first structure of a DNA-dependent DNA polymerase, whose wild-type ancestor has no significant reverse transcriptase activity, incorporating a nucleotide using an RNA target. The results highlight the altered structure of the DNA/RNA duplex in comparison to the DNA/DNA duplex. Additionally, crucial amino acid sites are identified that prevent the wild-type enzyme from using RNA as template. While the mutation M747K might increase binding affinity to any negatively charged nucleic acid, S515R stabilizes the nucleic acid binding motif in the thumb domain. L459M results in an increased flexibility of the thumb domain, which contributes to prevent the clash with a DNA/RNA hybrid template. Furthermore, we have reported some very promising properties of the variant that allow applications like fast detection and quantification of RNA by RT-PCR in approximately 20 min. The enzyme facilitates multiple applications like rapid transcription analysis as well as the detection RNA and DNA pathogens. Since RT-KTq 2 exhibits reverse transcriptase and PCR activity in a single enzyme, it has great potential for minimizing labour and time consuming optimizations without compromising reliable analysis.

Figures

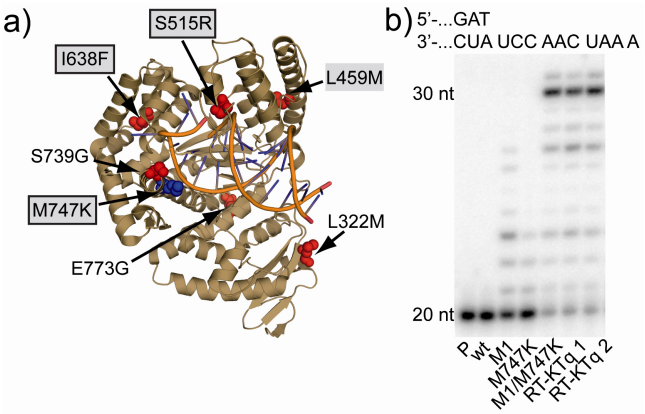


Figure 1. Characterization of KTq wild-type and mutants. a) Structure of KTq DNA polymerase (PDB 3RTV) highlighting the mutation sites of KTq M1 and M747K in red and blue, respectively. Mutations of RT-KTq 2 are highlighted in gray, whereas mutations of RT-KTq 1 are highlighted with black lining. b) Primer extension reactions with a [^{32}P]-labeled 20 nt primer annealed to RNA template. Reaction mixtures were incubated for 1 min at 72 °C. Partial primer and template sequences are shown. P = primer only; wt = KTq wild-type

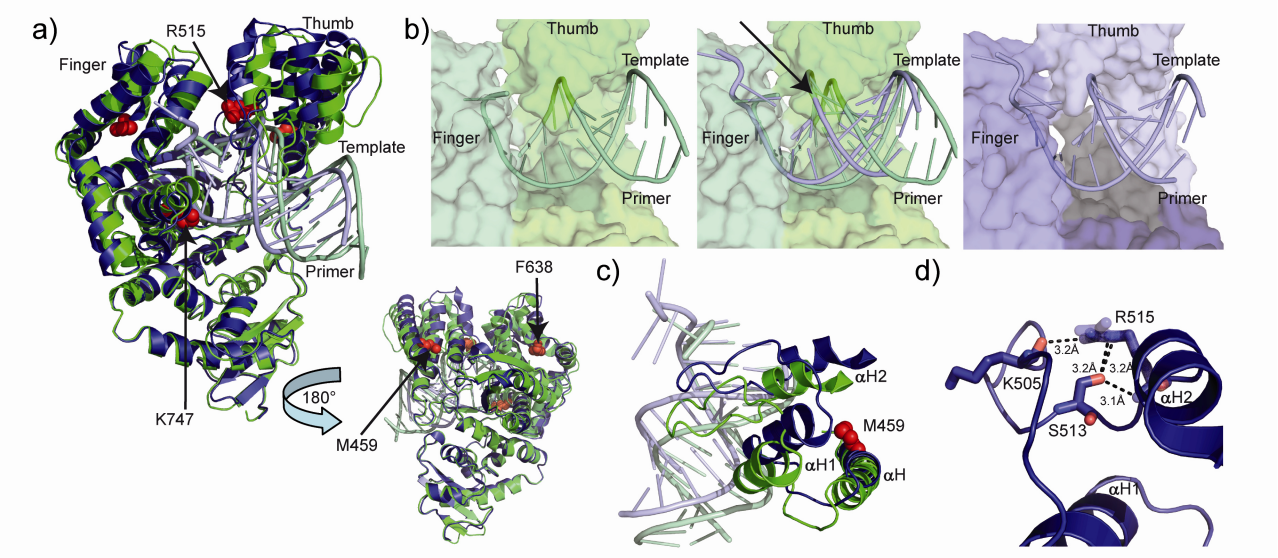


Figure 2. Crystal structures of RT-KTq 2_{DNA} and RT-KTq 2_{RNA}. a) Structure of RT-KTq 2_{DNA} (green) superimposed with RT-KTq 2_{RNA} (blue). DNA-duplex and DNA/RNA hybrid are coloured in pale green and pale blue, respectively. Mutation sites are highlighted as red spheres. b) Close-up view shows DNA duplex in RT-KTq 2_{DNA} (left), overlay of DNA and hybrid duplex in RT-KTq 2_{DNA} (middle) and DNA/RNA hybrid duplex in RT-KTq 2_{RNA} (right), respectively. The overlay of the hybrid and the DNA duplex 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). c) Close-up view highlights the position of the M459 mutation on H-helices. M459 is depicted as red sphere. d) Close-up view highlights the interaction network of mutation S515R. The R515 is depicted as blue sticks. An alternative conformation is transparent

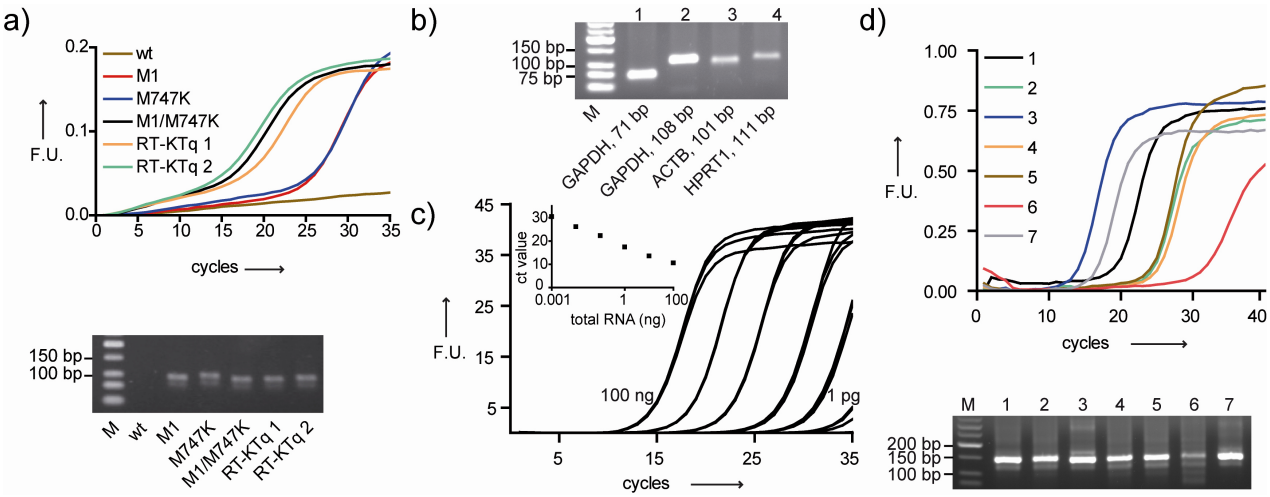


Figure3. RT-PCR using RT-KTq 2. a) Real-time RT-PCR performed with wild-type (wt), M1, M747K, M1/M747K, RT-KTq 1 and 2 (as indicated) amplifying a 100 bp target sequence from MS2 bacteriophage RNA. Fluorescence readout was based on SYBRGreen I binding to double-stranded DNA. Product formation was analyzed on a 2.5 % agarose gel. M: Marker. b) Detection of mRNA transcripts from human total RNA (Jurkat cells) in about 20 min using ultra-fast PCR. Target amplicons: GAPDH (71 bp; 108 bp, annealing/extension at 61°C), actin (ACTB; 101 bp, annealing/extension at 72°C) and Hypoxanthine phosphoribosyltransferase 1 (HPRT1; 111 bp, annealing/extension at 63°C) in lane 1, 2, 3 and 4, respectively. M: Marker. c) Real-time RT-PCR employing a ten-fold dilution series of total RNA (Jurkat cells) amplifying a 90 bp fragment from human beta actin. Dilution series ranged from 100 ng to 1 pg total RNA. Reactions were performed in triplicates. Inset: Ct (cycle threshold) values were plotted against the amount of total RNA. d) Detection of influenza virus A from RNA extracts of seven different specimens. Product formation was analyzed on a 2.5 % agarose gel. M: Marker.

Table 1. Specific activities of *KlenTaq* variants on DNA- or RNA-template.

Variant	Specific Activity [min^{-1}] DNA template	Specific Activity [min^{-1}] RNA template
KlenTaq wt	145 ± 14.2	n.a.
KlenTaq M1	252 ± 20.6	1.67 ± 0.05
KlenTaq M747K	279 ± 17.6	0.31 ± 0.01
KlenTaq M1/M747K	342 ± 11.0	32.2 ± 1.79
KlenTaq Hit 1	181 ± 11.9	15.7 ± 0.57
KlenTaq Hit 2	349 ± 12.7	34.9 ± 1.26

2.3.4 Supporting Information

Structure and Function of a thermostable DNA polymerase reading RNA

Nina Blatter, Konrad Bergen, Wolfram Welte, Kay Diederichs, Jutta Mayer, Markus Wieland & Andreas Marx

Table of Contents:

Methods

Library Generation.199,200 Parental DNA (pGDR11[204] *KlenTaq* genes encoding wt, M1 and M747K) was amplified in a nested PCR (100 µl) using 0,03 u/µl Phusion polymerase (Thermo Scientific), DNA Shuffling primer forward 5'-d(CGA GGC CCT TTC GTC TTC AC)-3' and reverse 5'-d(CTT AGC TCC TGA AAA TCT CGC C)-3' (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 was 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 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 (80 µl total volume). The reaction was terminated by addition of EDTA and incubation at 65 °C for 10 min. DNA fragments in the size range of 70-150 bp were obtained.

Without further purification, 5 µl of DNA fragments were utilized in an Assembly PCR (50 µl total volume) using no primers, 200 µM of each dNTP, 3 % (v/v) DMSO and 1 u Phusion Polymerase. 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. Optimization of this step revealed less point mutations when using the Phusion polymerase instead of Taq polymerase.

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 primers with restriction sites for

cleavage with SphI 5'-d(CAT ACG GAT CCG CAT GCA GCC CTG GAG GAG GCC C)-3' and HindIII 5'-d(GCT CAG CTA ATT AAG CTT TCT CCT TGG CGG AGA GCC)-3' (200 nM each), 200 µM of each dNTP, 3 % (v/v) DMSO and 0.03 u/µl Phusion polymerase. 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 the respective restriction enzymes SphI and HindIII (Fermentas) 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. Theoretically, recombination of M1 and M747K via DNA-Shuffling comprises seven positions being either mutated or displaying the wild-type amino acid sequence, thus resulting in 2^7 = 128 mutation combination possibilities in total. Therefore, a library size of approximately 1,200 clones is required to have a 99 % probability to sample all possible variants, presuming that every mutation combination is equiprobable.[205]

Screening and Identification of Hits. *KlenTaq* variants of the generated library were expressed in 96-well plates. Cells were grown in 1 ml LB-media (per well) and protein expression was induced by addition of IPTG (1 mM final concentration) at an OD of 0.6 - 0.8. Protein expression was conducted at 37 °C for 4.5 h in LB media (carbenicillin 100 mg/l) and cells harvested afterwards. 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 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$. After heat-denaturation at 75 °C for 45 min and centrifugation to remove bacterial cell debris, the lysates were directly used in an activity screening performed in 384-well plates using real-time PCR. Reaction mixtures (20 µl)

contained 60 pM template (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'), 250 µM of each dNTP, 750 nM of each primer (5'-d(CGT TGG TCC TGA AGG AGG AT)-3'; 5'-d(CGC GCA GCA CGC GCC GCC GT)-3'), 0.6x SYBRGreen I and 10 µl of the respective lysate in 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$. 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. 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.*[7] DNA formation was visualized by binding of SYBRGreen I. In short, reaction mixtures (20 µl) contained 50 pg/µl MS2 RNA (Roche) as template, 200 µM of each dNTP, 100 nM of each primer (5'-d(ATC GCT CGA GAA CGC AAG TT)-3'; 5'-d(CG GAC TTC ATG CTG TCG GTG)-3'), 0.6x SYBRGreen I and 5 µl of the respective lysate in 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$. 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.

KlenTaq mutants with a ct-value lower than the ct value of the parental mutants M1 and M747K were selected 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) in a second screening. Variants with the most promising results in all experiments were selected and their genes sequenced at GATC-Biotech.

Proteins and Oligonucleotides. Proteins for the characterization experiments[206]

and crystallization trials were purified[20] as described, respectively. Protein expression was conducted in *E. coli* BL21 (DE3) cells. Oligonucleotides were purchased from Metabion or Biomers (Germany).

Primer Extension Assay. Reaction mixtures (20 µl) contained 150 nM radioactively labeled primer (5'-d(CGT TGG TCC TGA AGG AGG AT)-3'), 225 nM RNA template (5'-AAA UCA ACC UAU CCU CCU UCA GGA CCA ACG-3') or the respective DNA template (5'-d(AAA TCA ACC TAT CCT CCT TCA GGA CCA ACG TAC)-3'), 200 µM of each dNTP and 25 nM of the respective *KlenTaq* DNA polymerase in 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20 and 2.5 mM $MgCl_2$. Reaction mixtures were incubated at 72 °C and terminated after 1 min by addition of 45 µl stop solution (80 % [v/v] formamide, 20 mM EDTA, 0.25 % [w/v] bromophenol blue, 0.25 % [w/v] xylenecyanol). After denaturation at 95 °C for 5 min, reaction mixtures were separated using a 12 % denaturing PAGE gel. Visualization was performed by phosphoimaging.

DNA Polymerase Activity Determination.[10] Primer extension reactions were performed at 72 °C (described as above) with an incubation time of 10 min for DNA as template and 30 min for RNA as template. Reaction mixtures (20 µl) contained 150 nM radioactively labeled primer (5'-d(CGT TGG TCC TGA AGG AGG ATA GG)-3'), 225 nM RNA template (5'-AAA UCA ACC UAU CCU CCU UCA GGA CCA ACG-3') or the respective DNA template (5'-d(AAA TCA ACC TAT CCT CCT TCA GGA CCA ACG TAC)-3'), 200 µM of each dNTP and various amounts of the respective *KlenTaq* DNA polymerase in 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20 and 2.5 mM $MgCl_2$. 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 analyzed and slopes were obtained using linear regression yielding the specific activity of the respective enzyme.

CD-spectra Measurement and Thermal Denaturation.[207] *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 Na_2HPO_4, 1.8 mM KH_2PO_4 (pH 7.4) using Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific). CD spectra measurements were performed at 20 °C using quartz cuvettes (light path 1 mm) and 250 µl protein sample (2.2 µM). CD spectra were measured 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 conducted 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.

Crystallization and Structure Determination. Crystallization solutions containing *KlenTaq* (6.6 mg/ml) in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, RNA template (5'-d(AAA GGG CGC CGU GGU C)-3') (217 µM), DNA primer 5'-d(GAC CAC GGC GC)-3') (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).

Crystallization solutions with *KlenTaq* 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 (5'-d(AAA GGG CGC CGT GGT C)-3') (158 µM), DNA primer 5'-d(GAC CAC GGC GC)-3') (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 for DNA as template and after 3 days for the RNA template. Streak seeding led to the formation of crystals with improved diffraction characteristics in case of RNA as template. 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.[168,169] The structure of $KlenTaq_{DNA}$ was solved using difference Fourier methods. In case of $KlenTaq_{RNA}$ molecular replacement (PHASER) was used to solve the phase problem.[170] Both structures were solved and refined using the PHENIX suite.[170] Manual refining and model rebuilding was performed using Coot.205 Model quality was determined by the MolProbity web server.[173] Ramachandran statistics for $KlenTaq_{DNA}$: favored 97.61%, allowed 2.20 % and 0.18 % outlier. Statistics for $KlenTaq_{RNA}$: 98.53/1.47/0. Molecular graphics were drawn with Pymol.[175] Data collection and refinement statistics can be found in Table S2. In case of $KlenTaq_{RNA}$ parts of the protein had to be retraced, as well as the hybrid duplex, neglecting the 2 outermost base pairs, which were not resolved.

RT-PCR with MS2 Bacteriophage RNA.[7] Reaction mixtures (20 µl) for real-time RT-PCR contained 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1% Tween 20, 2.5 mM $MgCl_2$, 200 µM of each dNTP, 100 nM of each primer (5'-d(ATC GCT CGA GAA CGC AAG TT)-3'; 5'-d(CG GAC TTC ATG CTG TCG GTG)-3'), 0.6x SYBRGreen I (Sigma-Aldrich), 5 nM of the respective purified DNA polymerase and 50 pg/µl MS2 RNA (Roche). 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 binding of SYBRGreen I. Correct product formation was confirmed by agarose gel analysis.

Amplification of a 510 bp Fragment from MS2 RNA in RT-PCR. RT-PCR reactions contained 50 mM Tris HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 200 nM of each primer (5'-d(GAT CGC ATG CC TAG AGG CAC TTG CCT ACT ACG)-3'; 5'-d(GCT AAA GCT TCG GAC TTC ATG CTG TCG GTG ATT TC)-3'), 80 nM of the respective DNA polymerase and 350 pg/µl MS2 RNA (Roche). 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 and 90 s at 72 °C. Correct product formation was confirmed by agarose gel analysis.

Ultra-fast Two-step Cycling 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. Reaction mixtures (10 µL)

contained 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$, 200 µM of each dNTP, 100 nM of each primer, 50 nM of DNA polymerase and 1 ng/µl total RNA. Cycling was performed with 30 s initial denaturation at 98 °C and 30 cycles of two step cycling with denaturation for 1 s at 92 °C and a combined annealing/extension for 10 s at the respective temperature. GAPDH (71 bp; 108 bp) transcripts were amplified at an annealing/extension temperature of 61 °C using primer forward (5'-d(GAA GGT GAA GGT CGG AGT CAA C)-3'; 5'-d(CGT CAA GGC TGA GAA CGG GA)-3') and reverse (5'-d(CAG AGT TAA AAG CAG CCC TGG T)-3'; 5'-d(ACG TAC TCA GCG CCA GCA TC)-3'), respectively. Beta actin (101 bp) and HPRT1 (111 bp) transcripts were amplified at an annealing/extension temperature of 72 °C and 63 °C respectively, using primer forward (5'-d(TGC CCT GGC ACC CAG CAC AA)-3') and reverse (5'-d(AGG TGG ACA GCG AGG CCA GGA)-3') for actin and primer forward (5'-d(TGC TGA GGA TTT GGA AAG GGT GTT)-3') and reverse (5'-d(AGG GCT ACA ATG TGA TGG CCT CC)-3') for HPRT1.

Real-time RT-PCR, Template Dilution Series. 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. Reaction mixtures of real-time RT-PCR experiments (20 µl) contained 50 mM Tris-HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 0.1 % Tween 20, 2.5 mM $MgCl_2$, 200 µM of each dNTP, 100 nM of each primer forward (5'-d (CAC TCT TCC AGC CTT CCT TC)-3') and reverse (5'-d(GGA TGT CCA CGT CAC ACT TC)-3'), 1x SYBRGreen I, 50 nM of DNA polymerase and the respective concentration of total RNA. Template concentrations used in this experiment were in a range from 100 ng to 1 pg, ten-fold diluted in a step-wise manner. Cycling was performed with 60 s initial denaturation at 95 °C and 37 cycles of two-step cycling with denaturation for 15 s at 95 °C and combined

annealing/extension for 30 s at 72 °C. Formation of double stranded DNA was visualized by binding of SYBRGreen I.

Supplementary Figures

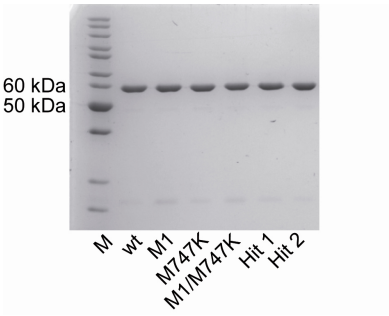


Figure S1. SDS-PAGE analysis of purified *KlenTaq* wild-type (wt) and variants.

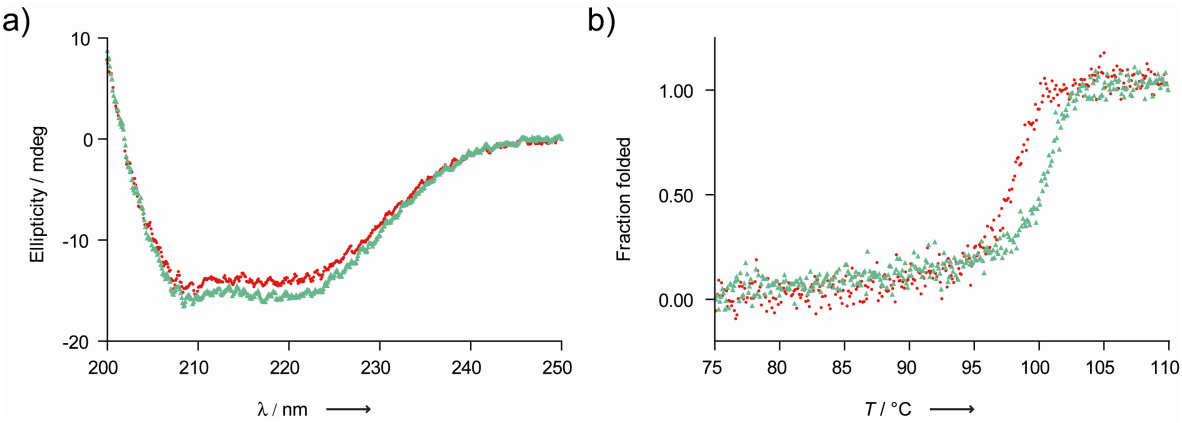


Figure S2. Stability of *KlenTaq* M1/M747K and *KlenTaq* Hit 2. a) CD-spectra of *KlenTaq* M1/M747K and *KlenTaq* Hit 2 are shown in red and green, respectively. b) Thermal denaturation of *KlenTaq* M1/M747K and *KlenTaq* Hit 2 was measured following the ellipticity at 209 nm. Melting temperatures were determined in two separate experiments for *KlenTaq* M1/M747K (T_M 97.9 ± 0.1 °C) and *KlenTaq* Hit 2 (T_M 100.1 ± 0.1 °C).

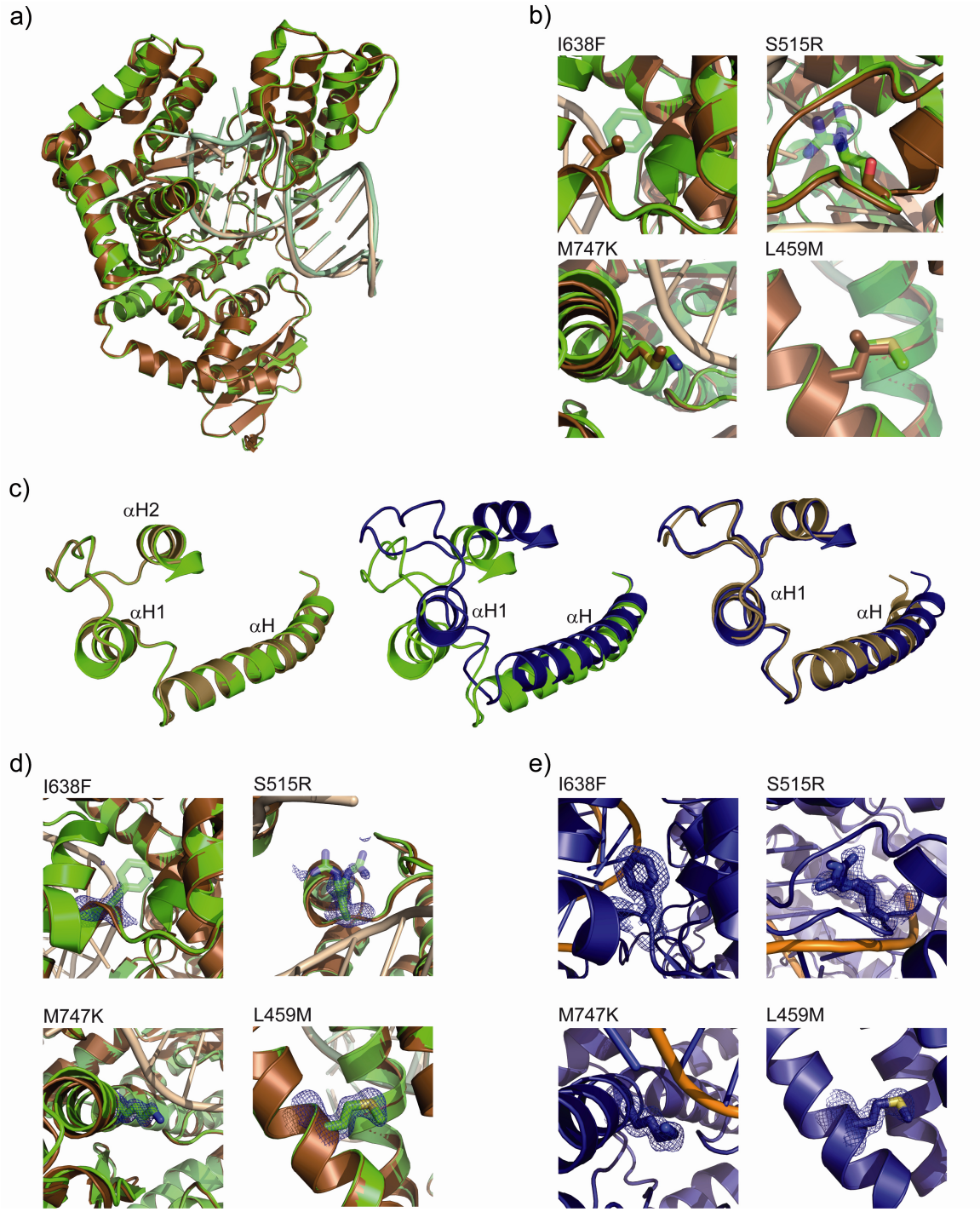


Figure S3. Structural details of *KlenTaq* wild-type, *KlenTaq* Hit 2_{DNA} and *KlenTaq* Hit $2_{RNA.}$ a) Structure of *KlenTaq* wild-type (sand, PDB 3RTV) superimposed with *KlenTaq* Hit 2_{DNA} (green) bound to DNA-duplex. Mutation sites are highlighted in b). c) Structural relocation of αH, αH1 and αH2 in the thumb domain. Overlay of protein structure of *KlenTaq* wild-type (PDB 3RTV, sand) and *KlenTaq* Hit 2_{DNA} (green) shows no relocation of the motif (left). Superimposition of protein structure of *KlenTaq* Hit 2_{DNA} (green) and *KlenTaq* Hit 2_{RNA} (blue) reveals shift in αH, αH1 and αH2 (middle). Superimposition of the motif alone reveals the origin of the shift. Deviation starts in the N-terminal portion of αH containing the L459M mutation (right). d)-e) Simulated annealing omit maps (σ level of 3) showing mutation sites of *KlenTaq* Hit 2. d) Mutation sites in *KlenTaq* Hit 2_{DNA} (green) superimposed with *KlenTaq* wild-type (PDB 3RTV, sand). e) Mutation sites in *KlenTaq* Hit 2_{RNA}.

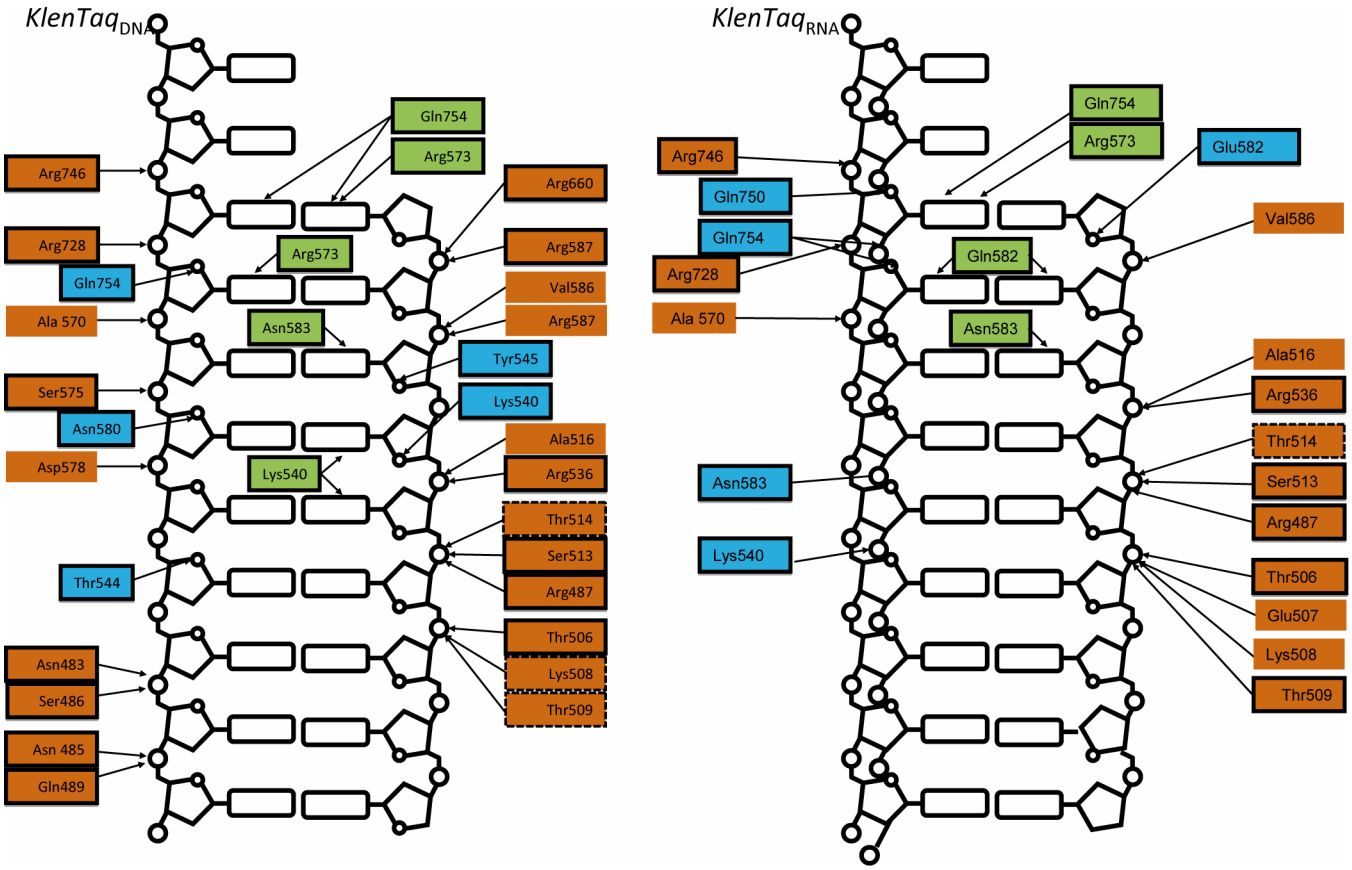


Figure S4. Interaction map of *KlenTaq* Hit 2 (closed ternary complex) with DNA duplex (left) and *KlenTaq* Hit 2 with DNA/RNA hybrid (right). Only direct contacts up to a distance of 3.6 Å are shown. Side chain interactions are marked with a solid lining, contacts with the protein backbone are shown 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.

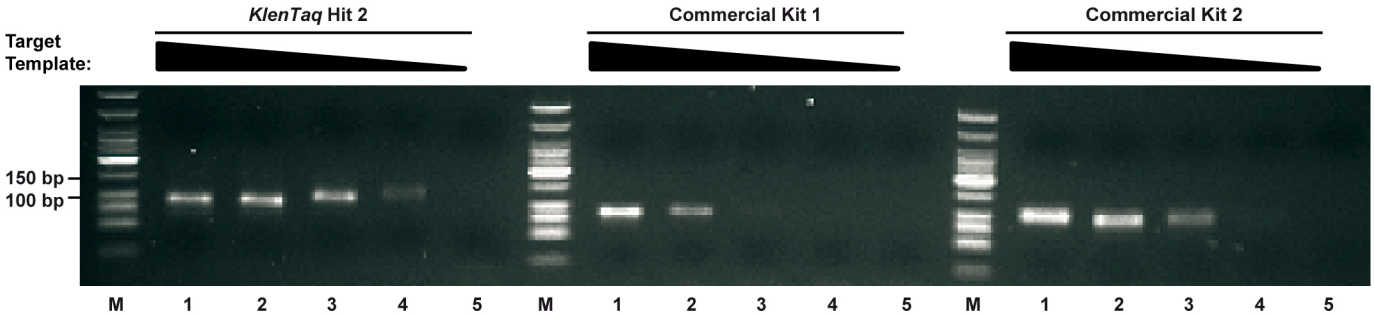


Figure S5. Comparison of sensitivity of *KlenTaq* Hit 2 with commercially available Superscript III One-Step RT-PCR System (Invitrogen) (Commercial Kit 1) and Titan One Kit (Roche; Commercial Kit 2). A 90 bp fragment from human beta actin was amplified from total RNA (Jurkat cells) with decreasing template amounts of 1000 pg (lane 1), 100 pg (lane 2), 10 pg (lane 3), 1 pg (lane 4) and without RNA as control (lane 5). Product formation was analyzed on a 2.5 % agarose gel. M: Marker.

Supplementary Tables

Table S1. Sugar pucker conformations of primer and template nucleotides.

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

Table S1. The sugar puckers in the primer nucleotides show several different conformations ranging from C3'-endo conformations for sugar residues free from enzyme interactions to C2'-endo, C1'-exo and O4'-endo. The sugar puckers in the template nucleotides adopt RNA-typical A-form conformation. 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 ***KlenTaq*** **Hit 2**$_{RNA}$.

Table S2. Data collection and refinement statistics (Molecular replacement)

	*KlenTaq*DNA PDB ID 4BWJ	*KlenTaq*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)	47.99-175	46.83-1.55
No. reflections	87950	79632
R_{work}/ R_{free}	16.3/19.6	17.3/20.3
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 (º)	1.117	1.090

*Highest resolution shell is shown in parenthesis.

2.4 Declaration of contributions

Chapter 2.1

"Structures of *KlenTaq* DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates"

Synthesis and characterization of dN*TP and dN**TPs were performed by A. Steck and A. Baccaro. Crystallization, soaking experiments, data recording and reduction, structure solution, refinement and deposition of dA*TP, dA**TP, dT*TP and dT**TP containing structures were realized by K. Bergen. dC*TP and dG*TP containing structures were crystallized, solved and refined by S. Strütt during his bachelor thesis, supervised by K. Bergen. The manuscript was written by A. Steck, K. Bergen and A. Marx.

Chapter 2.2

"Structures of KOD and 9°N DNA polymerases complexed with primer template duplex"

Protein purification, crystallization, data recording and reduction, structure solution refinement and deposition of the KOD/DNA structure were done by K. Bergen.
K. Betz, K. Bergen and A. Marx wrote the manuscript.

Chapter 2.3

"Structure and Function of a thermostable DNA polymerase reading RNA"

K. Bergen contributed to this work with aid and expertise in crystallization using different techniques, data recording and reduction, structure solution, subsequent refinement, deposition of structural data and wrote parts of the manuscript.

3 Discussion

3.1 Concluding Remarks and Outlook

In nature, all DNA catalysis is performed by DNA polymerases and they are therefore key in many cellular processes like replication, recombination and lesion repair.[30]
The ability to replicate a nucleic acid polymer has been exploited for numerous applications in modern molecular biology and diagnostic purposes.
The efficient enzymatic generation of functionalized DNA is essential for numerous biotechnological applications like NGS, SELEX (see Chapter 1.4)
Therefore, an in-depth understanding of the interaction between the enzymes, DNA polymerases, and the substrate, a modified dNTP, is crucial for the rational combination of enzyme, the modification type and its attachment. This goal can be achieved in a structure based attempt.
In the present study we analysed DNA-dependent DNA Polymerases from families A and B regarding the acceptance of modified nucleotides and a mutant member of an A-family polymerase employing RNA as templating nucleic acid.

So far structural data of DNA polymerases processing base-modified dNTPs was limited to C5-modified pyrimidines.[64,74,75] We were able to solve structures of a complete set of modified dNTPs, bearing an amino-pentinyl linker at the C5 (pyrimidines) and the 7-deaza position (purines) respectively. Additionally, structures containing dUTP and dATP, modified with a 5- respectively 7-(N-(10-hydroxydecanoyl)-aminopentinyl) modification, previously used by Baccaro et al as "barcoded nucleotides",[94] have been solved. We observed a similar arrangement for the amino-pentinyl moiety in the dG*TP, dT*TP and dC*TP containing structures and a unique one in case of the modified dA*TP. To scrutinise if this deviation also occurs on further modified substrate, crystallisation attempts using the above mentioned linkers resulted in the visualization of two different loopholes in *KlenTaq* polymerase. Furthermore, the amide bond connecting the linker subsets is involved in a stabilisation via hydrogen-bonding to K663 in case of the dA**TP; dT**TP displays interactions to R660 and T664. These findings, in combination with recent publications[64,74,75], can aid in the customisation of linker-payload combination used in the enzyme-mediated generation of functionalised DNA employing family A enzymes. Further research on the elongation of the incorporated modification as well as

consecutive incorporation and elongation of multiple modifications will be performed in the collaboration of the Marx and Welte groups.

Though thermostable B-family enzymes from archeal origin are often used in biotechnological applications due to their processivity, fidelity and the ability to incorporate (highly) modified dNTPs[73,90,103,104,150,208,209] structural data on this subject is lacking. Almost all of the recent structural research is focused on the enzymes in the editing mode and the recognition of deaminated bases[131,208].

To understand the extended capabilities to cope with (highly) modified substrates structural data of enzymes in the replication mode are needed. This part of the study was focused on the setup of a stable crystallization system, applicable to similar experiments as in Chapter 2.1.

Though the KOD polymerase crystallised readily in a vast amount of conditions, but nearly all diffraction grade crystals turned out to be consisting of an uncomplexed, open conformation. In consequence, different approaches to overcome this obstacle had to be undertaken. The final successful strategy, as described in Chapter 2.2, was based on using fluorophoric labelled DNA in the crystallisation trials to discern the DNA bound crystals. The project delievered additional models for archeal B-family polymerases besides the most recently described Pfu DNA polymerase mutant E10.[209]

Future work on this subject should be focused on the generation of ternary complexes of the enzymes based on the visual tagging technique used in this study (see above). Addition of a further fluorophor to the incoming nucleotide, employing a FRET based approach, could help in the identification process of ternary complexes. The structural insights derived form such an approach will allow insights in the binding, incorporation and a possible customisation of modified substrates (see chapter 2.1) as well as rational mutation approaches to promote an increased acceptance.[78]

The previous paragraphs covered the acceptance and incorporation of modified nucleoside triphosphate as aberrant substrates. With the exception of RT enzymes, DNA dependent and RNA dependent polymerases are fixed to their respective template. The reverse transcription reaction is an invaluable tool which is used in clinical diagnostics for the detection of pathogen or disease specific markers as well as a tool for molecular biology (e.g. cDNA libraries) or in transcriptomic based

experiments and monitoring. Therefore, time and cost saving applications are of high importance. The mutant described by Nina Blatter[210] can be used in a single enzyme solution, reducing the investement in an additional enzyme and reaction step.

In a structure based attempt we were able show the impact of two of the mutational sites on the ability to cope with an altered geometry of the nucleic acid. The enhanced flexibility gained by a decrease in the steric demand of an aminoacid side chain as well as the stabilization of the loop region between the H1-H2 helices.

Further research on this topic could allow the transfer of the enhanced substrate spectrum to other polymerases by rationally mutating key residues to improve the domain flexibility in order to achieve an adaptation to the altered substrate geometry.

4 References

1. Eckert, M. (2012) *Annalen der Physik* **524**, A83-A85
2. McPherson, A. (1991) *J. Cryst. Growth* **110**, 1-10
3. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) *Nature* **181**, 662-666
4. Bragg, L. P., M.F. (1954) *Proc. R. Soc. Lond. A* **225** 315-329
5. Evans, P., and McCoy, A. (2008) *Acta Crystallographica Section D* **64**, 1-10
6. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535-542
7. Crick, F. (1970) *Nature* **227**, 561-563
8. Dahm, R. (2010) *EMBO Rep.* **11**, 153-160
9. Avery, O. T., Macleod, C. M., and McCarty, M. (1944) *J. Exp. Med.* **79**, 137-158
10. Wilkins, M. H., Stokes, A. R., and Wilson, H. R. (1953) *Nature* **171**, 738-740
11. Franklin, R. E., and Gosling, R. G. (1953) *Nature* **171**, 740-741
12. Chargaff, E., Magasanik, B., Vischer, E., Green, C., Doniger, R., and Elson, D. (1950) *J. Biol. Chem.* **186**, 51-67
13. Watson, J. D., and Crick, F. H. (1953) *Nature* **171**, 737-738
14. Watson, J. D., and Crick, F. H. (1953) *Nature* **171**, 964-967
15. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) *Nature* **287**, 755-758
16. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. L. (1982) *Science* **216**, 475-485
17. (2001) *Curr Protoc Nucleic Acid Chem* **Appendix 1**, Appendix 1C
18. Rice, P. A., and Correll, C. C. (2008) *Protein-nucleic acid interactions : structural biology*, Royal Society of Chemistry, Cambridge
19. Rupp, B. (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York
20. Bessman, M. J., Lehman, I. R., Simms, E. S., and Kornberg, A. (1958) *J. Biol. Chem.* **233**, 171-177
21. Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958) *J. Biol. Chem.* **233**, 163-170
22. Bessman, M. J., Lehman, I. R., Adler, J., Zimmerman, S. B., Simms, E. S., and Kornberg, A. (1958) *Proc. Natl. Acad. Sci. U. S. A.* **44**, 633-640
23. Brutlag, D., Atkinson, M. R., Setlow, P., and Kornberg, A. (1969) *Biochem. Biophys. Res. Commun.* **37**, 982-989
24. Klenow, H., and Henningsen, I. (1970) *Proc. Natl. Acad. Sci. U. S. A.* **65**, 168-175
25. Freemont, P. S., Ollis, D. L., Steitz, T. A., and Joyce, C. M. (1986) *Proteins* **1**, 66-73
26. Braithwaite, D. K., and Ito, J. (1993) *Nucleic Acids Res.* **21**, 787-802
27. Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990) *Protein Eng.* **3**, 461-467
28. Ito, J., and Braithwaite, D. K. (1991) *Nucleic Acids Res.* **19**, 4045-4057
29. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem.* **63**, 777-822
30. Rothwell, P. J., and Waksman, G. (2005) *Adv. Protein Chem.* **71**, 401-440
31. Ishino, Y., Komori, K., Cann, I. K., and Koga, Y. (1998) *J. Bacteriol.* **180**, 2232-2236
32. Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997) *Genes Cells* **2**, 499-512
33. Yamtich, J., and Sweasy, J. B. (2010) *Biochim. Biophys. Acta* **1804**, 1136-1150
34. Filee, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002) *J. Mol. Evol.* **54**, 763-773

35. Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) *Mol. Cell* **8**, 7-8
36. Joyce, C. M., and Grindley, N. D. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1830-1834
37. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature* **313**, 762-766
38. Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17**, 7514-7525
39. Li, Y., and Waksman, G. (2001) *Protein Sci.* **10**, 1225-1233
40. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* **391**, 251-258
41. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) *Cell* **107**, 91-102
42. Zhou, B. L., Pata, J. D., and Steitz, T. A. (2001) *Mol. Cell* **8**, 427-437
43. Wing, R. A., Bailey, S., and Steitz, T. A. (2008) *J. Mol. Biol.* **382**, 859-869
44. Davies, J. F., 2nd, Almassy, R. J., Hostomska, Z., Ferre, R. A., and Hostomsky, Z. (1994) *Cell* **76**, 1123-1133
45. Lamers, M. H., Georgescu, R. E., Lee, S. G., O'Donnell, M., and Kuriyan, J. (2006) *Cell* **126**, 881-892
46. Bailey, S., Wing, R. A., and Steitz, T. A. (2006) *Cell* **126**, 893-904
47. Kunkel, T. A. (2009) *Cold Spring Harbor Symp. Quant. Biol.* **74**, 91-101
48. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H., and Steitz, T. A. (1997) *Cell* **89**, 1087-1099
49. Batra, V. K., Beard, W. A., Shock, D. D., Krahn, J. M., Pedersen, L. C., and Wilson, S. H. (2006) *Structure* **14**, 757-766
50. Loh, E., and Loeb, L. A. (2005) *DNA Repair (Amst)* **4**, 1390-1398
51. Rothwell, P. J., Mitaksov, V., and Waksman, G. (2005) *Mol. Cell* **19**, 345-355
52. Rothwell, P. J., Allen, W. J., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J., Waksman, G., and Seidel, C. A. (2013) *J. Biol. Chem.* **288**, 13575-13591
53. Steitz, T. A. (1998) *Nature* **391**, 231-232
54. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y. J., and Yang, W. (2012) *Nature* **487**, 196-201
55. Freudenthal, B. D., Beard, W. A., Shock, D. D., and Wilson, S. H. (2013) *Cell* **154**, 157-168
56. Castro, C., Smidansky, E. D., Arnold, J. J., Maksimchuk, K. R., Moustafa, I., Uchida, A., Gotte, M., Konigsberg, W., and Cameron, C. E. (2009) *Nat. Struct. Mol. Biol.* **16**, 212-218
57. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51 Pt 1**, 263-273
58. Joyce, C. (2002) *Methods Mol. Biol.* **193**, 83-92
59. Metzker, M. L. (2010) *Nat. Rev. Genet.* **11**, 31-46
60. Obeid, S., Yulikov, M., Jeschke, G., and Marx, A. (2008) *Angew. Chem., Int. Ed.* **47**, 6782-6785
61. Mayer, G. (2009) *Angew. Chem., Int. Ed.* **48**, 2672-2689
62. Weisbrod, S. H., and Marx, A. (2007) *Chem. Commun.*, 1828-1830
63. Hirao, I., and Kimoto, M. (2012) *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **88**, 345-367
64. Obeid, S., Baccaro, A., Welte, W., Diederichs, K., and Marx, A. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21327-21331
65. Anderson, J. P., Angerer, B., and Loeb, L. A. (2005) *BioTechniques* **38**, 257-264
66. Baccaro, A., Steck, A. L., and Marx, A. (2012) *Angew. Chem. Int. Ed. Engl.* **51**, 254-257

67. Borsenberger, V., Kukwikila, M., and Howorka, S. (2009) *Org. Biomol. Chem.* **7**, 3826-3835
68. Gramlich, P. M., Wirges, C. T., Manetto, A., and Carell, T. (2008) *Angew. Chem., Int. Ed.* **47**, 8350-8358
69. Hocek, M., and Fojta, M. (2008) *Org. Biomol. Chem.* **6**, 2233-2241
70. Kaufmann, G. F., Meijler, M. M., Sun, C., Chen, D. W., Kujawa, D. P., Mee, J. M., Hoffman, T. Z., Wirsching, P., Lerner, R. A., and Janda, K. D. (2005) *Angew. Chem. Int., Ed. Engl.* **44**, 2144-2148
71. Kore, A. R. (2009) *Tetrahedron Lett.* **50**, 793
72. Seela, F., and Zulauf, M. (1998) *Chem-Eur. J.* **4**, 1781-1790
73. Thum, O., Jager, S., and Famulok, M. (2001) *Angew. Chem., Int. Ed.* **40**, 3990-3993
74. Obeid, S., Busskamp, H., Welte, W., Diederichs, K., and Marx, A. (2012) *Chem. Commun. (Camb.)* **48**, 8320-8322
75. Obeid, S., Busskamp, H., Welte, W., Diederichs, K., and Marx, A. (2013) *J. Am. Chem. Soc.* **135**, 15667-15669
76. Hollenstein, M. (2013) *Org. Biomol. Chem.* **11**, 5162-5172
77. Summerer, D., and Marx, A. (2002) *J. Am. Chem. Soc.* **124**, 910-911
78. Betz, K., Streckenbach, F., Schnur, A., Exner, T., Welte, W., Diederichs, K., and Marx, A. (2010) *Angew. Chem., Int. Ed.* **49**, 5181-5184
79. Baccaro, A., and Marx, A. (2010) *Chemistry* **16**, 218-226
80. Brock, T. D., and Freeze, H. (1969) *J. Bacteriol.* **98**, 289-297
81. Chien, A., Edgar, D. B., and Trela, J. M. (1976) *J. Bacteriol.* **127**, 1550-1557
82. Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E., and Waksman, G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9264-9268
83. Li, Y., Kong, Y., Korolev, S., and Waksman, G. (1998) *Protein Sci.* **7**, 1116-1123
84. Li, Y., Mitaxov, V., and Waksman, G. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9491-9496
85. Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W., and Marx, A. (2010) *EMBO J.* **29**, 1738-1747
86. Bergen, K., Steck, A. L., Strutt, S., Baccaro, A., Welte, W., Diederichs, K., and Marx, A. (2012) *J. Am. Chem. Soc.* **134**, 11840-11843
87. Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Dwyer, T. J., Ordoukhanian, P., Romesberg, F. E., and Marx, A. (2012) *Nat. Chem. Biol.* **8**, 612-614
88. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James,

T., Huw Jones, T. A., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., and Smith, A. J. (2008) *Nature* **456**, 53-59

89. Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J. W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S. R., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H., and Xie, Z. (2008) *Science* **320**, 106-109
90. Ohbayashi, T., Kuwahara, M., Hasegawa, M., Kasamatsu, T., Tamura, T., and Sawai, H. (2005) *Org. Biomol. Chem.* **3**, 2463-2468
91. Seo, T. S., Bai, X., Kim, D. H., Meng, Q., Shi, S., Ruparel, H., Li, Z., Turro, N. J., and Ju, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5926-5931
92. Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., Sano Marma, M., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., and Ju, J. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9145-9150
93. Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D. H., Turro, N. J., and Ju, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5932-5937
94. Baccaro, A., Steck, A.-L., and Marx, A. (2011) *Angew. Chem., Int. Ed.* **51**, 254-257
95. Weisbrod, S. H., and Marx, A. (2008) *Chem. Commun.*, 5675-5685
96. Augustin, M. A., Ankenbauer, W., and Angerer, B. (2001) *J. Biotechnol.* **86**, 289-301
97. Seela, F., Feiling, E., Gross, J., Hillenkamp, F., Ramzaeva, N., Rosemeyer, H., and Zulauf, M. (2001) *J. Biotechnol.* **86**, 269-279
98. Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W., and Marx, A. (2010) *EMBO J.* **29**, 1738-1747
99. Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17**, 7514-7525
100. Jager, S., Rasched, G., Kornreich-Leshem, H., Engeser, M., Thum, O., and Famulok, M. (2005) *J. Am. Chem. Soc.* **127**, 15071-15082
101. Upton, T. G., Kashemirov, B. A., McKenna, C. E., Goodman, M. F., Prakash, G. K., Kultyshev, R., Batra, V. K., Shock, D. D., Pedersen, L. C., Beard, W. A., and Wilson, S. H. (2009) *Org. Lett.* **11**, 1883-1886
102. Leconte, A. M., Patel, M. P., Sass, L. E., McInerney, P., Jarosz, M., Kung, L., Bowers, J. L., Buzby, P. R., Efcavitch, J. W., and Romesberg, F. E. (2010) *Angew. Chem. Int., Ed. Engl.* **49**, 5921-5924
103. Staiger, N., and Marx, A. (2010) *ChemBioChem* **11**, 1963-1966
104. Ramsay, N., Jemth, A. S., Brown, A., Crampton, N., Dear, P., and Holliger, P. (2010) *J. Am. Chem. Soc.* **132**, 5096-5104
105. Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., and Holliger, P. (2004) *Nat. Biotechnol.* **22**, 755-759

106. Fa, M., Radeghieri, A., Henry, A. A., and Romesberg, F. E. (2004) *J. Am. Chem. Soc.* **126**, 1748-1754
107. Lebedev, A. V., Combs, D., and Hogrefe, R. I. (2007) *Bioconjug. Chem.* **18**, 1530-1536
108. Seela, F., and Zulauf, M. (1999) *Helv Chim Acta* **82**, 1878-1898
109. Muehlegger, K., Angerer, B., Seela, F., Ankenbauer, W., Augustin, M., Gumbiowski, K., and Zulauf, M. (2000) *WO 0068422,*
110. Ramzaeva, N., Mittelbach, C., and Seela, F. (1997) *Helv Chim Acta* **80**, 1809-1822
111. Seela, F., Feiling, E., Gross, J., Hillenkamp, F., Ramzaeva, N., Rosemeyer, H., and Zulauf, M. (2001) *J. Biotechnology* **86**, 269
112. Yasuo, K., Eiko, O., and Kazuko, M. (2003) *JP 2003116581,*
113. Summerer, D., Rudinger, N. Z., Detmer, I., and Marx, A. (2005) *Angew. Chem. Int. Ed. Engl.* **44**, 4712-4715
114. Betz, K., Streckenbach, F., Schnur, A., Exner, T., Welte, W., Diederichs, K., and Marx, A. (2010) *Angew. Chem. Int. Ed. Engl.* **49**, 5181-5184
115. Kabsch, W. (2010) *Acta Crystallogr.* **D 66**, 125-132
116. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr.* **D 66**, 213-221
117. (1994) *Acta Crystallogr.* **D 50**, 760-763
118. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr.* **D 60**, 2126-2132
119. Diederichs, K., and Karplus, P. A. (1997) *Nat Struct Biol* **4**, 269-275
120. DePamphilis, M. L. (1996) *DNA replication in eukaryotic cells*, Cold Spring Harbor Laboratory Press, Plainview, N.Y.
121. Mendez, J., Blanco, L., and Salas, M. (1997) *EMBO J.* **16**, 2519-2527
122. Esteban, J. A., Salas, M., and Blanco, L. (1993) *J. Biol. Chem.* **268**, 2719-2726
123. Longas, E., de Vega, M., Lazaro, J. M., and Salas, M. (2006) *Nucleic Acids Res.* **34**, 6051-6063
124. Wang, F., and Yang, W. (2009) *Cell* **139**, 1279-1289
125. Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009) *Nat. Struct. Mol. Biol.* **16**, 979-986
126. Rodriguez, A. C., Park, H. W., Mao, C., and Beese, L. S. (2000) *J. Mol. Biol.* **299**, 447-462
127. Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3600-3605
128. Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R., and Kuriyan, J. (1999) *Structure* **7**, 1189-1199
129. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001) *J. Mol. Biol.* **306**, 469-477
130. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., and Imanaka, T. (1997) *Appl. Environ. Microbiol.* **63**, 4504-4510
131. Killelea, T., Ghosh, S., Tan, S. S., Heslop, P., Firbank, S. J., Kool, E. T., and Connolly, B. A. (2010) *Biochemistry* **49**, 5772-5781
132. Gouge, J., Ralec, C., Henneke, G., and Delarue, M. (2012) *J. Mol. Biol.* **423**, 315-336
133. Franklin, M. C., Wang, J., and Steitz, T. A. (2001) *Cell* **105**, 657-667
134. Veedu, R. N., Vester, B., and Wengel, J. (2009) *Org. Biomol. Chem.* **7**, 1404-1409
135. Jager, S., and Famulok, M. (2004) *Angew. Chem., Int. Ed.* **43**, 3337-3340
136. Jager, S., Rasched, G., Kornreich-Leshem, H., Engeser, M., Thum, O., and Famulok, M. (2005) *J. Am. Chem. Soc.* **127**, 15071-15082

137. Bergen, K., Steck, A. L., Strutt, S., Baccaro, A., Welte, W., Diederichs, K., and Marx, A. (2012) *J. Am. Chem. Soc.* **134**, 11840-11843
138. Obeid, S., Busskamp, H., Welte, W., Diederichs, K., and Marx, A. (2012) *Chem. Commun.* **48**, 8320-8322
139. Rodriguez, A. C., Park, H. W., Mao, C., and Beese, L. S. (2000) *J. Mol. Biol.* **299**, 447-462
140. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001) *J. Mol. Biol.* **306**, 469-477
141. Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., Imanaka, T., and Kai, Y. (2005) *J. Mol. Biol.* **351**, 291-298
142. Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3600-3605
143. Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., and Connolly, B. A. (2008) *J. Mol. Biol.* **381**, 529-539
144. Kim, S. W., Kim, D. U., Kim, J. K., Kang, L. W., and Cho, H. S. (2008) *Int. J. Biol. Macromol.* **42**, 356-361
145. Gouge, J., Ralec, C., Henneke, G., and Delarue, M. (2012) *J. Mol. Biol.* **423**, 315-336
146. Jiang, X., and Egli, M. (2011) *Curr. Protoc. Nucleic Acid. Chem.* **Chapter 7**, Unit 7 15 11-18
147. Rothwell, P. J., and Waksman, G. (2005) *Adv. Protein Chem.* **71**, 401-440
148. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature* **391**, 304-307
149. Braithwaite, D. K., and Ito, J. (1993) *Nucleic Acids Res.* **21**, 787-802
150. Veedu, R. N., Vester, B., and Wengel, J. (2009) *Org. Biomol. Chem.* **7**, 1404-1409
151. Veedu, R. N., Vester, B., and Wengel, J. (2008) *J. Am. Chem. Soc.* **130**, 8124-8125
152. Veedu, R. N., Vester, B., and Wengel, J. (2007) *ChemBioChem* **8**, 490-492
153. Di Pasquale, F., Fischer, D., Grohmann, D., Restle, T., Geyer, A., and Marx, A. (2008) *J. Am. Chem. Soc.* **130**, 10748-10757
154. Strerath, M., Gaster, J., Summerer, D., and Marx, A. (2004) *ChemBioChem* **5**, 333-339
155. Strerath, M., and Marx, A. (2002) *Angew. Chem.* **114**, 4961-4963
156. Strerath, M., and Marx, A. (2002) *Angewandte Chemie International Edition* **41**, 4766-4769
157. Ikonen, S., Macickova-Cahova, H., Pohl, R., Sanda, M., and Hocek, M. (2010) *Org. Biomol. Chem.* **8**, 1194-1201
158. Brázdilová, P., Vrábel, M., Pohl, R., Pivoňková, H., Havran, L., Hocek, M., and Fojta, M. (2007) *Chemistry – A European Journal* **13**, 9527-9533
159. Burley, G. A., Gierlich, J., Mofid, M. R., Nir, H., Tal, S., Eichen, Y., and Carell, T. (2006) *J. Am. Chem. Soc.* **128**, 1398-1399
160. Giller, G., Tasara, T., Angerer, B., Muhlegger, K., Amacker, M., and Winter, H. (2003) *Nucleic Acids Res.* **31**, 2630-2635
161. Jager, S., and Famulok, M. (2004) *Angew. Chem. Int. Ed. Engl.* **43**, 3337-3340
162. Kumar, S. B. M., NJ, US), Sood, Anup (Flemington, NJ, US), Rao, Sudhakar (Belle Mead, NJ, US), Nelson, John (Hillsborough, NJ, US) (2006) Oligonucleotide tagged nucleoside triphosphates (OTNTPs) for genetic analysis. (E Healthcare Bio-Sciences Corp. (Piscataway, N., US)7109316 ed., United States
163. Seela, F., and Chen, Y. (1996) *Chem. Commun.*, 2263-2264
164. Weizman, H., and Tor, Y. (2002) *J. Am. Chem. Soc.* **124**, 1568-1569
165. Kielkowski, P., Macíčková-Cahová, H., Pohl, R., and Hocek, M. (2011) *Angew. Chem.* **123**, 8886-8889
166. Gierlich, J., Gutsmiedl, K., Gramlich, P. M. E., Schmidt, A., Burley, G. A., and Carell, T. (2007) *Chemistry – A European Journal* **13**, 9486-9494

167. Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W., and Perler, F. B. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5281-5285
168. Kabsch, W. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132
169. Kabsch, W. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133-144
170. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
171. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501
172. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35**, W375-383
173. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21
174. Karplus, P. A., and Diederichs, K. (2012) *Science* **336**, 1030-1033
175. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
176. Hall, T. A. (1999) *Nucleic Acids Symposium Series* **41**, 95-98
177. Bustin, S. A., and Mueller, R. (2005) *Clin. Sci. (Lond.)* **109**, 365-379
178. Bustin, S. A. (2000) *J. Mol. Endocrinol.* **25**, 169-193
179. Gloeckner, C., Sauter, K. B., and Marx, A. (2007) *Angew. Chem. Int. Ed. Engl.* **46**, 3115-3117
180. Kranaster, R., Drum, M., Engel, N., Weidmann, M., Hufert, F. T., and Marx, A. (2010) *Biotechnol. J.* **5**, 224-231
181. Vichier-Guerre, S., Ferris, S., Auberger, N., Mahiddine, K., and Jestin, J. L. (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 6133-6137
182. Sauter, K. B., and Marx, A. (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 7633-7635
183. Blatter, N., Bergen, K., Nolte, O., Welte, W., Diederichs, K., Mayer, J., Wieland, M., and Marx, A. (2013) *Angew. Chem. Int. Ed. Engl.*
184. Kunkel, T. A., and Bebenek, K. (2000) *Annu. Rev. Biochem.* **69**, 497-529
185. Ricchetti, M., and Buc, H. (1993) *EMBO J.* **12**, 387-396
186. Joyce, C. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1619-1622
187. Patel, P. H., and Loeb, L. A. (2000) *J. Biol. Chem.* **275**, 40266-40272
188. Ogawa, M., Tosaka, A., Ito, Y., Yoshida, S., and Suzuki, M. (2001) *Mutat. Res.* **485**, 197-207
189. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G., and Romesberg, F. E. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6597-6602
190. Ong, J. L., Loakes, D., Jaroslawski, S., Too, K., and Holliger, P. (2006) *J. Mol. Biol.* **361**, 537-550
191. Nick McElhinny, S. A., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundstrom, E. B., Johansson, E., Chabes, A., and Kunkel, T. A. (2010) *Nat. Chem. Biol.* **6**, 774-781
192. Brown, J. A., and Suo, Z. (2011) *Biochemistry* **50**, 1135-1142
193. Wang, W., Wu, E. Y., Hellinga, H. W., and Beese, L. S. (2012) *J. Biol. Chem.* **287**, 28215-28226
194. Obeid, S., Schnur, A., Gloeckner, C., Blatter, N., Welte, W., Diederichs, K., and Marx, A. (2011) *ChemBioChem* **12**, 1574-1580
195. Obeid, S., Welte, W., Diederichs, K., and Marx, A. (2012) *J. Biol. Chem.* **287**, 14099-14108

196. Clark, J. M. (1988) *Nucleic Acids Res.* **16**, 9677-9686
197. Rudinger, N. Z., Kranaster, R., and Marx, A. (2007) *Chem. Biol.* **14**, 185-194
198. Reetz, M. T. (2013) *Angew. Chem. Int. Ed. Engl.* **52**, 2658-2666
199. Fedoroff, O., Salazar, M., and Reid, B. R. (1993) *J. Mol. Biol.* **233**, 509-523
200. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R., and Holliger, P. (2012) *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8067-8072
201. Gerard, G. F., Potter, R. J., Smith, M. D., Rosenthal, K., Dhariwal, G., Lee, J., and Chatterjee, D. K. (2002) *Nucleic Acids Res.* **30**, 3118-3129
202. Palese, P. (2004) *Nat. Med.* **10**, S82-87
203. Kutyavin, I. V., Afonina, I. A., Mills, A., Gorn, V. V., Lukhtanov, E. A., Belousov, E. S., Singer, M. J., Walburger, D. K., Lokhov, S. G., Gall, A. A., Dempcy, R., Reed, M. W., Meyer, R. B., and Hedgpeth, J. (2000) *Nucleic Acids Res.* **28**, 655-661
204. Peist, R., Koch, A., Bolek, P., Sewitz, S., Kolbus, T., and Boos, W. (1997) *J. Bacteriol.* **179**, 7679-7686
205. Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) *Protein Eng.* **16**, 451-457
206. Holzberger, B., Rubini, M., Moller, H. M., and Marx, A. (2010) *Angew. Chem. Int. Ed. Engl.* **49**, 1324-1327
207. Holzberger, B., Obeid, S., Welte, W., Diederichs, K., and Marx, A. (2012) *Chem. Sci.* **3**, 2924
208. Gill, S., O'Neill, R., Lewis, R. J., and Connolly, B. A. (2007) *J. Mol. Biol.* **372**, 855-863
209. Wynne, S. A., Pinheiro, V. B., Holliger, P., and Leslie, A. G. (2013) *PLoS One* **8**, e70892
210. Bebenek, K., and Kunkel, T. A. (2000) *Cold Spring Harbor Symp. Quant. Biol.* **65**, 81-91

5 Appendix

5.1 List of used data sets

The given path and name of image series refers to the storage system used in the AG Welte/Diederichs, linking the PDB ID to the respective data

5.1.1 Data sets used for the structure solution in Chapter 2.1

"KlenTaq caught incorporating C5 and 7-deaza modified nucleotides"

PDB ID **4DF8**
contains the dA*TP
/nfs/loop1/mnt/synchrotron/SLS-2011/mar27/klentaq-annalena-soak1_1_00???.img

PDB ID **4DF4**
contains the dA**TP
/nfs/loop1/mnt/synchrotron /SLS-2011/may12/kb/tempt-1mM-3_1_00???.img.

PDB ID **4DFJ**
contains the dT*TP
merged from
/nfs/loop1/mnt/synchrotron/SLS-2011/apr08/V-5_2_0????.cbf.bz2
/nfs/loop1/mnt/synchrotron/SLS-2011/apr08/VIII-2_2_0????.cbf.bz2
/nfs/loop1/mnt/synchrotron/SLS-2011/apr08/k-1_2_0????.cbf.bz2

PDB ID **4DFK**
contains the dT**TP
/nfs/loop1/mnt/synchrotron/SLS-2011/jun09/kb/dTx-lang1mM-3_1_0????.cbf

PDB ID **4DFM**
contains the dC*TP
merged from
/nfs/loop1/mnt/synchrotron/SLS-2011/jun09/ss/TempG-tern-5mM-6_1_00???.cbf
/nfs/loop1/mnt/synchrotron/SLS-2011/jun09/ss/TempG-tern-25mM-3_1_00???.cbf
/nfs/loop1/mnt/synchrotron/SLS-2011/jun09/ss/TempG-1mMsoak-KBII-3_1_00???.cbf

PDB ID **4DFP**
contains the dG*TP
/nfs/loop1/mnt/synchrotron /SLS-2011/may12/ss/tempC-05mm-2_1_0????.cbf

Crystallization, data collection and reduction, structure solution and refinement for PDB IDs 4DFM and 4DFP were performed by Stefan Strütt.

5.1.2 Data sets used in Chapter 2.2

"Structures of KOD and 9°N DNA polymerases complexed with primer template duplex"

PDB ID **4K8Z**
merged from
/nfs/loop1/synchrotron/SLS-2012/jun28/kb/Kod-G3-16mer-A5-1_2_0????.cbf
/nfs/loop1/synchrotron/SLS-2012/jun28/kb/Kod-G3-16mer-A5-3_1_0????.cbf
/nfs/loop1/synchrotron/SLS-2012/jun28/kb/Kod-G3-16mer-A5-3_2_0????.cbf

Data sets used for PDB ID 4K8X, see PhD thesis K. Betz

5.1.3 Data sets used in Chapter 2.3

"Structure and Function of a thermostable DNA polymerase reading RNA"

PDB ID **4BWJ**
/nfs/loop1/SLS-2012/feb29/nb/A4-ohnecryo-D10-NucPro_1_0????.cbf

PDB ID **4BWM**
/nfs/loop1/SLS-2012/apr27/nb/KTQ-D9-RNA_4_1_0????.cbf

5.2 List of abbreviations

Å	Angstroem
bp	Base pairs
c	Concentration
cDNA	complementary DNA
CV	Column volume
ds	Double stranded
dsDNA	Double stranded DNA
E.coli	*Escherichia coli*
nt	Nucleotide
OD	Optical density
PAGE	Polyacrylamide-gelelectrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Database
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonylfluorid
pol	Polymerase
PP	Pyrophosphate
r.m.s.d.	root mean square deviation
RNA	ribonucleic acid
rpm	rotations per minute
RT	Reverse transcription
SDS	Sodium dodecyl sulfate
SNP	single nucleotide polymorphism
ss	single stranded
ssDNA	single stranded DNA
Taq	*Thermus aquaticus*
UV	Ultraviolet
v	Volume

Nucleotide nomenclature

(d)ATP	(2'-deoxy-)adenosine-5'-triphosphate
(d)CTP	(2'-deoxy-)cytidine-5'-triphosphate
(d)GTP	(2'-deoxy-)guanosine-5'-triphosphate
(d)TTP	2'-deoxythymidine-5'-triphosphate
UTP	uridine-5'-triphosphate
(d)NTP	2'-deoxynucleoside-5'-triphosphate
ddATP	2',3'-dideoxyadenosine-5'-triphosphate
ddCTP	2',3'-dideoxycytidine-5'-triphosphate
ddTTP	2',3'-dideoxythimidine-5'-triphosphate
ddNTP	2',3'-didesoxynucleoside-5'-triphosphate

Mono- and diphosphates are abbreviated with (d)NMP and (d)NDP respectively.

5.3 Amino Acid nomenclature

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

5.4 Danksagungen

Zuerst möchte ich mich bei meinen beiden Professoren Prof. Dr. Wolfram Welte und Prof. Dr. Andreas Marx für die Möglichkeit zur Promotion bedanken, wie auch für ihre Unterstützung und die herausragenden Arbeitsbedingungen. Ebenso möchte ich mich bei Pr. Dr. Kay Diederichs für die Hilfestellung bei kristallographischen Frage(n)(stellungen) bedanken (auch wenn die eine oder andere "Konni-Frage" sein mußte). Prof. Dr. Hans-Jürgen Appell möchte ich unter anderem für die Übernahme des Prüfungsvorsitz danken.

Im besonderen gilt Dank:

Anna-Lena und Nina für die produktive und angenehme Zusammenarbeit
Karin als verbleibendem Rest der Polymerasen Fraktion für die geteilte Denkerzelle und gute Zusammenarbeit (tja, jetzt bist du wohl die Entenmama...)
Allen Synchrotronreisenden beider Arbeitsgruppen für die entspannte Arbeitsatmosphäre der (teils stressigen) Nachtschichten (ein Hoch auf Gummibärchen und Kaffee)
Annemarie für ihre Hilfsbereitschaft, die vielen Diskussionen und den Enthusiasmus und die Kongenialität beim Aufbau alten Laborgeräts :-).
Andy für viele Diskussionen und die Hilfe während der Diplomarbeit und zu Beginn der Dissertation (Danke Papa Ente).
Michi für Korrekturen und interessante Diskussionen.
Karsten für Hilfe im kristallographischen Bereich (und den vielen Diskussionen rund ums Angeln und die Welt im Allgemeinen).
Stefan, für seine zuverlässige Mitarbeit während seiner Bachelorarbeit
Peter für seine stete Hilfe mit dem technischen Gerät und den Kaffee
Den jetzigen (und in der Zwischenzeit hauptsächlich) ehemaligen Mitgliedern der Arbeitsgruppen Welte und Marx für das angenehme Arbeitsumfeld und die stete Hilfsbereitschaft.
Allen meinen Freunden, die mich während der (allzulangen) Schreibezeit ertragen und unterstützt haben. @Flo: Der Meterfisch kommt auch noch (möge Dich der Angelwahnsinn nie verlassen)

Und nicht zuletzt meiner Familie für den Rückhalt in den teilweise schweren Zeiten. Ohne Eure Unterstützung wäre diese Arbeit nicht fertiggeworden. Helmut, ein letzter Dank geht an Dich, auch wenn Du das Ende dieser Unternehmung nicht mehr mitbekommen hast. Auch wenn nun manches zu kurz kam, mehr Unterstützung und Akzeptanz konnte sich ein Sohn nicht wünschen!

Anne, Du bist das Beste was mir passieren konnte.