X-ray analysis of KlenTaq and 9°N DNA polymerases

- structural insights into the replication of an artificial base pair and DNA binding

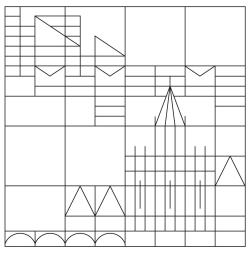
Dissertation

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

vorgelegt von

Karin Betz

an der
Universität Konstanz



Mathematisch - Naturwissenschaftliche Sektion
Fachbereich Chemie

Konstanz, 2014

Tag der mündlichen Prüfung:	04.07.2014
1. Referent	Prof. Dr. Andreas Marx
2. Referent	Prof. Dr. Kay Diederichs
3. Referent	Prof. Dr. Jörg Hartig
Prüfungsvorsitzender	Prof. Dr. Valentin Wittmann

Die vorliegende Arbeit entstand in der Zeit von Februar 2010 bis April 2014 in der Arbeitsgruppe von Prof. Dr. Andreas Marx am Lehrstuhl für Organische und Zelluläre Chemie im Fachbereich Chemie sowie den Arbeitsgruppen von Prof. Dr. Kay Diederichs und Prof. Dr. Wolfram Welte am Lehrstuhl für Proteinkristallografie und Molekulare Bioinformatik im Fachbereich Biologie an der Universität Konstanz. Die Arbeit wurde durch ein Stipendium der Graduiertenschule Chemische Biologie der Universität Konstanz gefördert.

Teile dieser Arbeit wurden veröffentlicht in:

Nat. Chem. Biol., vol. 8, no. 7, pp. 612–614, **2012**	K. Betz*, D. A. Malyshev*, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, and A. Marx, "Klentaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry."
J. Am. Chem. Soc., vol 135, no. 49, pp. 18637-18643, **2013**	K. Betz*, D. A. Malyshev*, T. Lavergne, W. Welte, K. Diederichs, F. E. Romesberg, and A. Marx, "Structural insights into DNA replication without hydrogen bonds."
ChemBioChem, vol. 14, no. 9, pp. 1058–1062, **2013**	K. Bergen*, K. Betz*, W. Welte, K. Diederichs, and A. Marx, "Structures of KOD and 9°N DNA polymerases complexed with primer template duplex."

Weitere Publikation:

Angew. Chem., Int. Ed. Engl., vol. 49, no. 30, pp. 5181-5184, **2010**	K. Betz*, F. Streckenbach*, A. Schnur, T. Exner, W. Welte, K. Diederichs, and A. Marx; "Structures of DNA polymerases caught processing size-augmented nucleotide probes."

* authors contributed equally to the work

Contents

1 Introduction

The determination of molecular structures of proteins and enzymes using X-ray crystallography tremendously contributed to our understanding of processes going on in the cells of our body and other organisms. In this work we aim to increase our knowledge about one specific class of enzymes: DNA polymerases. DNA polymerases are not only responsible for duplication of the genome in a cell prior to cell division and also mediate DNA repair and recombination but they also play a key role in many molecular biological methods routinely used in today's research and diagnostics. Although many DNA polymerases are already well studied biochemically and structurally, there are still open questions to answer.

In the first project of this work the well characterized Klenow Fragment of DNA polymerase I from *Thermus aquaticus* was used to study the incorporation of a hydrophobic artificial base pair into DNA and its continued elongation. Since several years researchers try to increase the coding potential of DNA by searching for a third (artificial) base pair which could complement the two existing ones. To facilitate optimization of promising candidates, molecular insights into the acceptance of such a pair by DNA polymerases are of great importance and should be provided through this work.

In a second project the B-family DNA polymerase 9°N from the strain *Thermococcus* species 9°N-7 is aimed to be crystallized in complex with DNA and triphosphate substrates. The structures should lead to further insights into the high fidelity replication mechanism of this enzyme and related archaeal B-family members.

1.1 Structure of nucleic acids

1.1.1 Structure of DNA

Deoxyribonucleic acid, or short DNA, is the basis of life as it carries and stores the genetic information of most living organisms. It is a biopolymer in which one strand consists

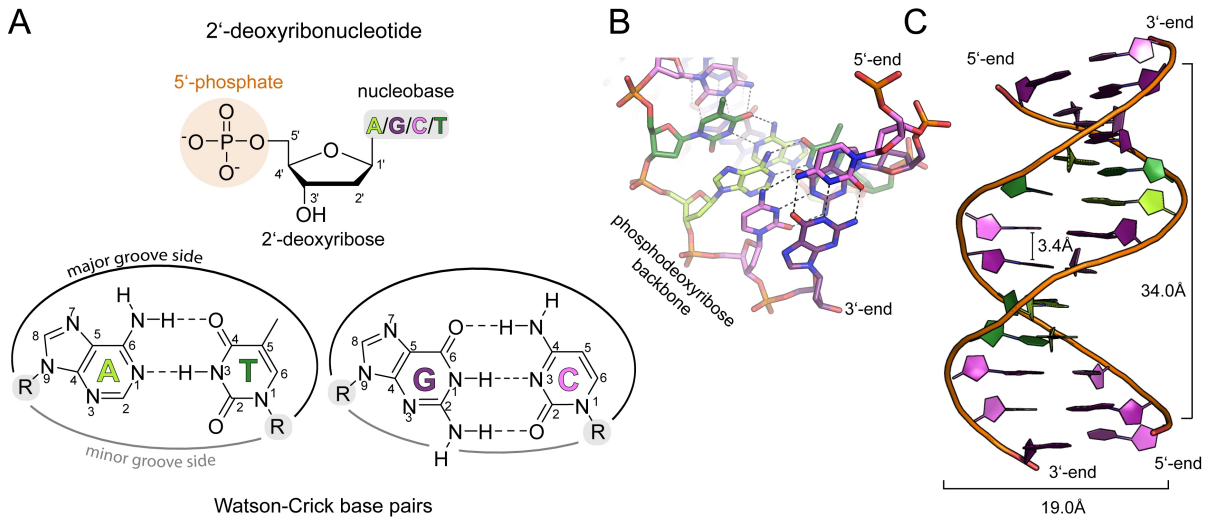


Figure 1.1: A: Chemical structure of 2'-deoxyribonucleotides. One nucleotide consists of the 2'-deoxyribose moiety, a 5'-monophosphate moiety and a nucleobase. The position of the nucleobase is indicated by the letters A, G, C and T and the structures of the four nucleobases forming Watson-Crick base pairs are shown beneath. Hydrogen bonds are indicated by dashed lines. Atom numbering for the purines and pyrimidines as well as the 2'-deoxyribose is given. Unless noted otherwise, R indicates the position of the deoxyribose in all figures. **B, C**: Structure of the DNA double helix in B-form (PDB ID: 2L8Q).

of a chain of monomers called 2'-deoxyribonucleotides. Those building blocks, in turn, can be divided into three parts: the nucleobase moiety, the sugar (2'-deoxyribose) moiety and the phosphate moiety (**Figure 1.1 A**). In DNA four different nucleotides are found which only vary in their nucleobase moiety. The nucleobase can either be a purine base: adenine (A) or guanine (G) or a pyrimidine base: thymine (T) or cytosine (C). Together with the invariable 2'-deoxyribose and phosphate moieties they form the nucleotides 2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxycytidine and 2'-deoxythymidine. The nucleobase moieties are connected to the sugar C1' atom via an N-glycosidic bond. Repeating sugars and phosphates are connected via phosphodiester bonds and form the backbone of the DNA (**Figure 1.1 B**). The four nucleotides interact via their respective donor and acceptor functional groups along the Watson-Crick edges of the nucleobase and form the two specific Watson-Crick base pairs guanine-cytosine (G-C; via three hydrogen bonds) and adenine-thymine (A-T; via two hydrogen bonds, **Figure 1.1 A**). Due to the specific pairing a given DNA strand can direct the synthesis of a complementary DNA

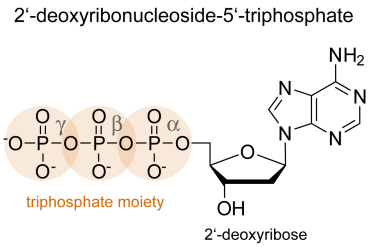


Figure 1.2: Structure of the DNA polymerase substrate dATP.

strand by DNA polymerases (described in detail in the next sections). These enzymes use 2'-deoxyribonucleoside-5'-triphosphates (short dNTPs) as substrates (**Figure 1.2**). The four natural substrates deoxyadenosin triphosphate, deoxycytidin triphosphate, deoxyguanosin triphosphate and deoxythymidin triphosphate are abbreviated with dATP, dCTP, dGTP and dTTP, respectively. Triphosphates lacking both, the 2' OH and the 3' OH group of the ribose moiety (2',3'-dideoxyribonucleoside triphosphates) which are often used for the crystallization of DNA polymerases in ternary complexes (see later) are named ddNTPs.

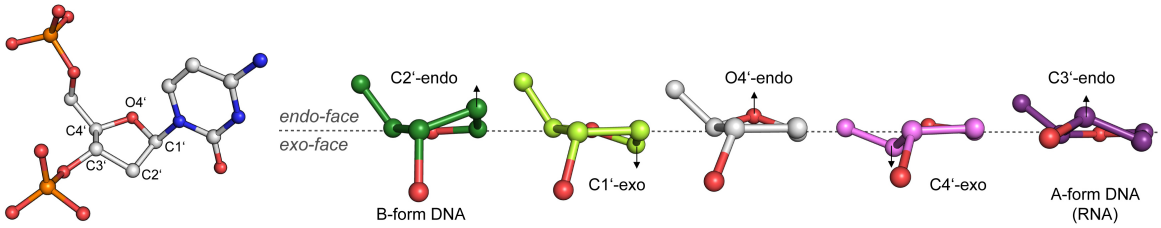


Figure 1.3: Sugar pucker of 2'-deoxyribose in DNA. Transition of sugar conformation from C2'-endo to C3'-endo is shown. Arrows indicate the atom that is situated above (endo) or below (exo) the reference plane. The figure was created in analogy to Figure 5 in [1]. Therefore structures of dsDNA (PDB IDs: 1IH6 and 1IH1) were used and the sugar pucker was determined using the PROSIT server[2].

DNA mostly appears as a double-stranded molecule. Two complementary strands pair with each other in an anti-parallel fashion creating a 3'- and a 5'-end and form a double helical structure with the two strands coiling around a common axis (**Figure 1.1C**). The two strands are held together by base pairing, base stacking and solvent interactions and are simultaneously destabilized by the negative charges of the phosphate backbone (for details see [1]). Apart from the most commonly known B-form, DNA can adopt many different structures depending on the DNA sequence and different environmental factors (see review of Gosh and Bansal[3]). One factor which influences the shape of DNA is the conformation of the 2'-deoxyribose sugar (called "sugar pucker") that can be described by the pseudorotation angle psi (ψ). Commonly two kinds of sugar conformations occur in DNA: the C2'-endo and the C3'-endo puckers, but other variations are also observed. The sugar puckers which occur in the transition from a C2'-endo puckering to a C3'-endo puckering (transformation from B-form DNA to A-form) are shown in **Figure 1.3**[1]. Endo is defined such that the respective atom is located above the plane (on the same side as the nucleobase) formed by the other four atoms of the furanose ring. If an atom is

placed below this plane it is in its exo form[1]. In DNA and RNA different sugar puckers entail different phosphate-to-phosphate distances along one strand, e.g. ∼7Å distance for the C2'-endo conformation and ∼ 6Å distance for C3'-endo[1]. Thus, if C3'-endo sugars are present, higher salt concentrations are needed to counter balance the shorter distance between the negative charges of the phosphates[1].

B-DNA. The B-form DNA structure first described by Watson & Crick, is the most common structural form of DNA occurring in the cell and is observed under conditions of high relative humidity. The first crystal structure analysis of more than a complete turn of right-handed B-DNA was published in 1980 by Wing *et al.*[4]. The complementary DNA strands form a right-handed, antiparallel double helix with approximately 10.4 residues per turn[5] and a distance of 3.4Å along the helical axis between two nucleotides[6] (**Figure 1.1 C**). The planes of the nucleobases are almost perpendicular to the helix axis (1° kinked). The double helix exhibits a minor and a major groove which are 6.0 and 11.6Å wide, respectively[4] (**Figure 1.4 A**). The widths are measured as the phosphate-to-phosphate distance across the two strands in a direction perpendicular to the trajectory of the strands[1]. Functional groups of the Watson-Crick base pairs which point into the minor groove are the N3 atom of purines and the amino group of guanine as well as the 2-carboxy groups of pyrimidines (see **Figure 1.1 A**). Positions 6-8 of purines and 4-6 of pyrimidines point into the major grooves. For that reason bulky DNA modifications are often attached to position C7 of 7-deazapurines and to the C5 position of pyrimidines[7–17]. The predominant 2'-deoxyribose conformation in B-DNA is the C2'-endo form, although the sugar puckers are variable to some extent[18].

A-DNA. In A-form DNA, which occurs under low humidity conditions, the double helix is also right-handed and antiparallel, but it is compressed along the helix (helix height per complete turn: 26.7Å vs. 34.0Å in B-DNA) resulting in an increased number of base pairs per turn and a sliding of base pairs making the duplex broader overall across the helix (diameter of 23Å in A-DNA vs. 19Å in B-DNA[19], **Figure 1.4 B**). The 2'-deoxyribose conformation is C3'-endo, leading to a tilting of the nucleobases with respect to the helix axis by approximately 19°. The A-helix exhibits a shallow, wide minor groove and a deep, narrow major groove with widths of 11.1Å and 2.2Å, respectively (**Figure 1.4 C**). A-form DNA plays a role in DNA synthesis by A-family DNA polymerases as described in section 1.2.3.1.

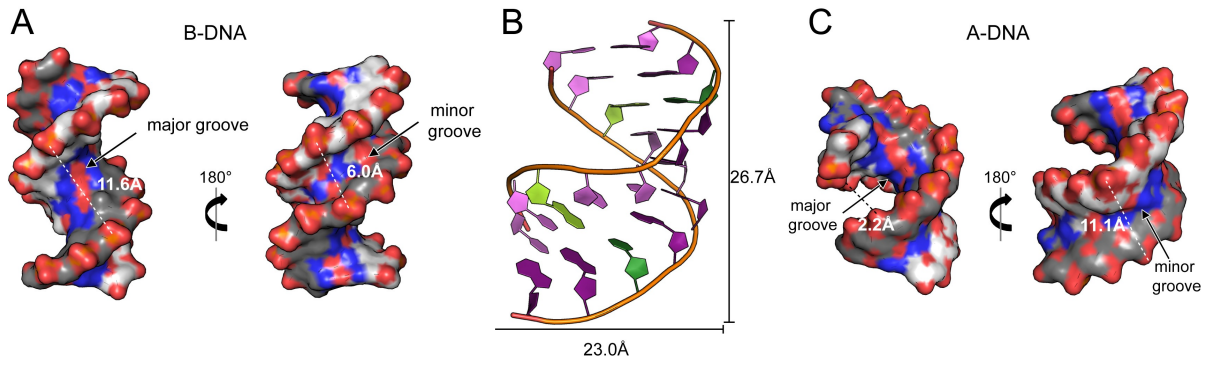


Figure 1.4: A: Surface representation of B-DNA (PDB ID: 2L8Q[20]) showing the major and minor groove. The groove width is indicated by a white dashed line. **B:** Structure of one helix turn of A-DNA (PDB ID: 1ZF8[21]). **C**: Major and minor groove in A-DNA. The groove width is again indicated. Data for helix diameter, helix pitch and minor and major groove widths are from [19].

1.1.2 Structure of RNA

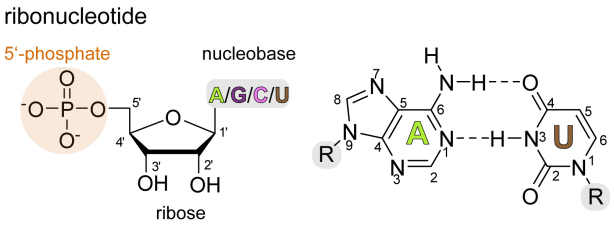


Figure 1.5: Structure of ribonucleotides - building blocks of RNA.

RNA (ribonucleic acid) is built up by similar building blocks as DNA. The main difference is the sugar moiety which is ribose instead of 2'-deoxyribose, carrying a hydroxyl group at the 2' position (**Figure 1.5**). Furthermore the nucleobase thymine is replaced by uracil (U) which also pairs with adenine. RNA is a mainly single-stranded molecule but has the ability to fold into a multitude of three dimensional structures as e.g. in tRNA. RNA is synthesized by RNA polymerases (short RNAP) from a DNA template (transcription) using ribonucleotide triphosphates (rNTPs or NTPs) as substrates. The major role of RNA is to direct the synthesis of proteins (with mRNA, tRNA and rRNA participating in the process) but it is also involved in regulatory processes (e.g. snRNA, siRNA, miRNA[22]). Furthermore, RNA is the coding material in the genome of various viruses.

1.2 DNA polymerases

1.2.1 History of DNA polymerases

The enzyme "DNA polymerase" was first discovered in 1955 by Arthur Kornberg and his colleagues Robert Lehman, Maurice Bessman and Ernie Simms, in the department of Microbiology at the Washington University School of Medicine[23]. Their basic experiment to find the enzyme responsible for DNA synthesis relied on the conversion of [^{14}C]thymidine and later α-[^{32}P]dTTP into an acid-insoluble, DNaseI-sensitive product by *E. coli* extracts in the presence of ATP and $MgCl_2$. In numerous experiments in which they fractionated the *E. coli* extracts, the researchers found that at least three different fractions were needed for the reaction. One fraction contained the enzyme catalysing phosphodiester bond formation and which they named "DNA polymerase". A second fraction contained DNA and a third contained a mixture of deoxynucleotide kinases and nucleoside-diphosphate kinase which produced the substrates dCTP, dATP and dGTP. For further *in vitro* experiments Kornberg and his colleagues established a method to purify the polymerase enzyme from a large amount of *E. coli* cells[24]. In addition they prepared all four ^{32}P-labeled deoxyribonucleotides by isolation of DNA from *E. coli* cells grown on ^{32}P-orthophosphate, degradation of the DNA with pancreatic DNase and snake venom phosphodiesterase to dNMPs, and purification/separation via anion exchange chromatograpy. dNTPs were prepared using kinases partially purified from *E. coli* in the presence of ATP. Initially Kornberg assumed that in their experiment the insoluble product is synthesized by addition of dTMP to pre-existing DNA chains[25]. In the course of their experiments, however, the group found that the reaction is based on a DNA template and that all four dNTPs as well as a primer are required for a proper reaction. Therefore, with their studies until 1985, Kornberg *et al.* could prove the template directed synthesis of DNA that was proposed by Watson and Crick in 1953[6].
The DNA polymerase first discovered was later named DNA polymerase I. In the following years four additional DNA polymerases (DNA polymerase II, III, IV and V) were discovered in *E. coli* and even more different DNA polymerases were found in eukaryotes. It was realized later that DNA polymerases can also exhibit endonuclease activity, excising nucleotides in 5'-3' direction and - depending on the polymerase - can possess an additional 3'-5' exonuclease domain.

In 1985, 30 years after the discovery of DNA polymerase I, the first crystal structure of a DNA polymerase was solved. In the group of Steitz, Ollis *et al.* were able to solve

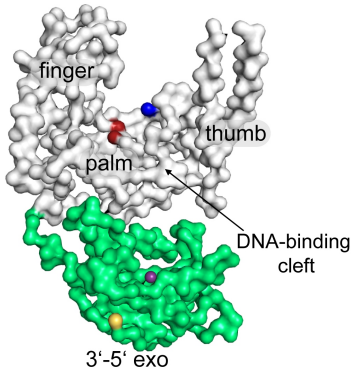


Figure 1.6: Crystal structure of the *E. coli* Klenow Fragment (PDB ID: 1DPI[26]) with Cα atoms shown in surface representation. The polymerase domain and the 3'-5' exonuclease domain are shown in grey and green, respectively. The N-terminus is indicated as blue sphere and the C-terminus as yellow sphere. The red spheres show the polymerase active site residues Asp705 and Asp882 and the purple sphere shows a Zn ion in the exonuclease active site.

the crystal structure of the C-terminal fragment (Klenow Fragment, see section 1.2.2.1) of *E. coli* DNA polymerase I in complex with a substrate dTMP but without DNA to a resolution of 3.3Å[26] (**Figure 1.6**). The structure revealed that the enzyme consists of a smaller N-terminal domain (amino acid 324-517, green in **Figure 1.6**) with the bound dTMP and a larger C-terminal domain (521-928, grey in **Figure 1.6**) which adopts a structure *"similar to the shape of a right hand holding a rod"*[26]. This larger domain in turn was subdivided into a finger, a thumb and a palm domain. The binding site for dTMP was supposed to be the exonuclease active site (indicated by a purple sphere in **Figure 1.6**) and the large cleft built up by the finger and thumb subdomain with a positive electrostatic potential was assumed to bind the DNA duplex. Furthermore, it was suggested that a small, flexible subdomain might close after DNA binding and therefore support processive nucleotide addition[26].

Over the years several crystal structures of DNA polymerases from different families either without DNA (apo), with a bound DNA substrate (binary complex) or an additionally bound triphosphate substrate (ternary complex) have been solved. These molecular models immensely contribute to our understanding of DNA polymerase mechanisms (see next section).

1.2.2 DNA polymerases - overview

DNA polymerases adopt different functions in a cell: replication of the genome, DNA repair and recombination. During replication the whole genome is duplicated and each

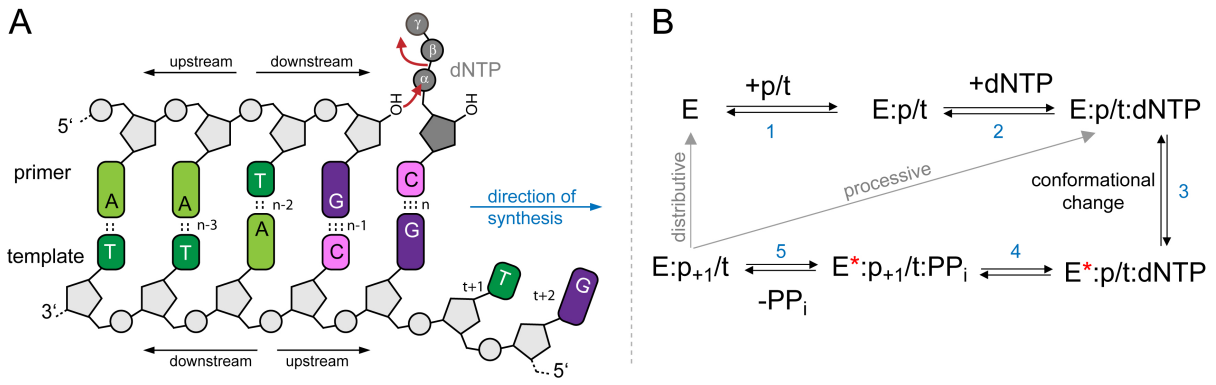


Figure 1.7: **A**: Scheme of nucleotide incorporation by a DNA polymerase. The catalyzed attack of the primer 3' OH group on the α phosphate of the substrate dNTP and the release of pyrophosphate is indicated by red arrows. DNA is synthesized in 5'-3' direction. Numbering of nucleotides in a ternary complex is shown. The base pair in the active site is termed n and base pairs in downstream template direction are numbered n-1, n-2, etc. Upstream template nucleotides are denoted by increasing numbers. Primer and template nucleotides are specified with p and t, respectively. **B**: Kinetic model of nucleotide incorporation. E = enzyme; PPi = pyrophosphate.

parental strand is complemented by a daughter strand which leads to two copies of the same DNA macromolecule. Replicative DNA polymerases need an existing DNA or RNA primer oligonucleotide annealed to the template to be able to perform DNA synthesis. The template-dependent elongation of the primer strand in 5'-3' direction is catalyzed by the enzyme, incorporating the appropriate nucleotides according to the Watson & Crick base pairing rules. A basic mechanism for the DNA polymerase reaction has been defined by kinetic studies. A scheme of the nucleotide insertion reaction and the different steps in the reaction cycle of most DNA polymerases are illustrated in **Figure 1.7 A** and **B**. After binding of the substrate DNA primer/template (p/t) to the enzyme E (step 1) the dNTP is associated and the loose complex E:p/t:dNTP is formed (step 2). In this step selection for a correct substrate already takes place to some degree (see below). After binding of the triphosphate, conformational changes are triggered (including open-to-closed transition of the finger domain in many polymerases, creating a tight binding pocket) and all components involved in the reaction are arranged for catalysis (step 3, activated enzyme indicated by a red asterisk). One of these conformational changes is rate-limiting. As soon as the triphosphate, the active site residues and the metal ions are properly aligned, the phosphodiester bond can be formed (step 4). Thereby, a nucleophilic attack of the free primer terminal 3' OH on the α phosphate of the substrate leads to the formation of a phosphodiester bond, following a two-metal-ion mechanism[27]. The two metal ions (typically Mg^{2+}) are coordinated by two strictly conserved carboxylate (aspartate and/or

glutamate) residues within the active site. A more detailed description of the chemical reaction mechanism is given in [28]. After the reaction the polymerase relaxes (step 5) and the pyrophosphate is released from the polymerase[28]. Along with pyrophosphate dissociation the enzyme presumably translocates to the next templating position. After one incorporation event the polymerase either binds a new substrate dNTP and the cycle is started again at point 2 (processive synthesis) or the enzyme dissociates from the substrate DNA (distributive synthesis).
DNA polymerases guide the repeating steps of nucleotide binding, base pairing, phosphodiester bond formation, pyrophosphate release and movement to the next templating position with very high rates and accuracy[28]. The repetitive cycle is performed with a fast rate of 600 nt/s in *E. coli*[29] and 30 nt/s in humans[30]. Besides DNA replication, DNA repair is essential, as replicative DNA polymerases also incorporate wrong nucleotides and additionally the DNA is damaged from environmental stress causing various lesions in the DNA. Different DNA repair mechanism exist and are performed by repair polymerases, specialized for their specific task (for more information about repair polymerases see [31]).

To maintain the genetic information it is crucial that only correct nucleotides are incorporated during DNA replication. Therefore DNA polymerases need to effectively select the correct binding partner from a pool of structurally similar molecules present in the cellular environment. This selection is accomplished by a combination of different selectivity factors (described in section 1.2.4). Polymerases differ greatly in their fidelity to incorporate nucleotides. Generally replicative polymerases show high accuracy generating approximately one mistake in 1000-1.000.000 incorporation events (fidelity: 10^{-3} to 10^{-6}), depending on the polymerase[32, 33]. In contrast, repair enzymes have fidelities in the range of 10^{-2}-10^{-4} but for individual enzymes fidelity can be even lower[34].
Based on structural and sequence similarity DNA polymerases are classified into seven different families (A, B, C, D, X, Y, and RT)[32, 35]. As mentioned above different DNA polymerases in a cell are mostly specialized on their specific functions in replication or repair of DNA. Polymerases undertaking special tasks are also often related and belong to the same family. The main replicative polymerases of archaea and eukaryotes belong to family B whereas the eubacterial replicative polymerases (pol III in *E. coli*) belong to family C. The X and Y family predominantly contain repair polymerases with X-family members being mainly involved in gap-filling DNA repair[36] and Y-family members being specified to recognize and bypass different kinds of DNA lesions in the process

called translesion DNA synthesis (TLS)[37]. Examples for family Y are polymerase IV (Dpo4) and V from bacteria and archaea, as well as eukaryotic polymerases η, κ, ι and Rev1[32]. Examples for family X are the eukaryotic polymerases β, σ, μ and λ[32]. Reverse transcriptases (RTs) make up an own family. These enzymes are mainly found in retroviruses in which they are responsible for the conversion of RNA via an RNA/DNA duplex into double-stranded DNA. Therefore, retroviral RTs possess three activities: a RNA-dependend RNA polymerase, a ribonuclease H to digest the ssRNA and a DNA-dependend DNA polymerase[38]. A prominent member of the family is the RT from the human immunodeficiency virus 1 (HIV1-RT)[32]. This enzyme (besides others) acts as a dimer whereas other members of the family are monomeric.
Family A and B DNA polymerases, which are the focus of this study, are described in more detail in the next sections.

1.2.2.1 Family A DNA polymerases

Family A polymerases can be subdivided into replicative and repair enzymes[32]. The replicative members are polymerases from the bacteriophages T3, T5 and T7 as well as the mitochondrial DNA polymerase γ. To accomplish accurate DNA replication these enzymes can interact with other proteins. T7 polymerase for example interacts with thioredoxin which functions as processivity factor hindering polymerase dissociation from the substrate[32]. Most promiment members of family A DNA polymerases are the repair DNA polymerases I from *E. coli*, short pol I (the first DNA polymerase discovered[39]), *Thermus aquaticus* (Taq) and *Bacillus stearothermophilus* (Bst). These enzymes are involved in nucleotide excision repair and in the processing of Okazaki fragments that are generated during lagging strand synthesis in replication[32]. DNA polymerase I exhibits three functional activities which are localized on separate domains: the N-terminal domain contains the 5'-3' endonuclease activity, the central domain bears the 3'-5' exonuclease activity (not functional in all members) and the C-terminal domain (comprising finger, thumb and palm) is responsible for the polymerase activity. During replication RNA primers from Okazaki fragments are excised by the 5'-3' endonuclease activity and the gaps are filled using the polymerase activity[32].
In 1970 it could be shown that a fragment of *E. coli* DNA polymerase I (**Figure 1.8 A**) obtained by proteolytic cleavage with Subtilisin keeps polymerase as well as 3'-5' exonuclease activity but loses 5'-3' endonuclease activity[40]). According to its discoverer this prominent fragment is named Klenow Fragment (short KF)[40] and has been mainly

used for the polymerase chain reaction (PCR) before the discovery of thermostable DNA polymerases[41].

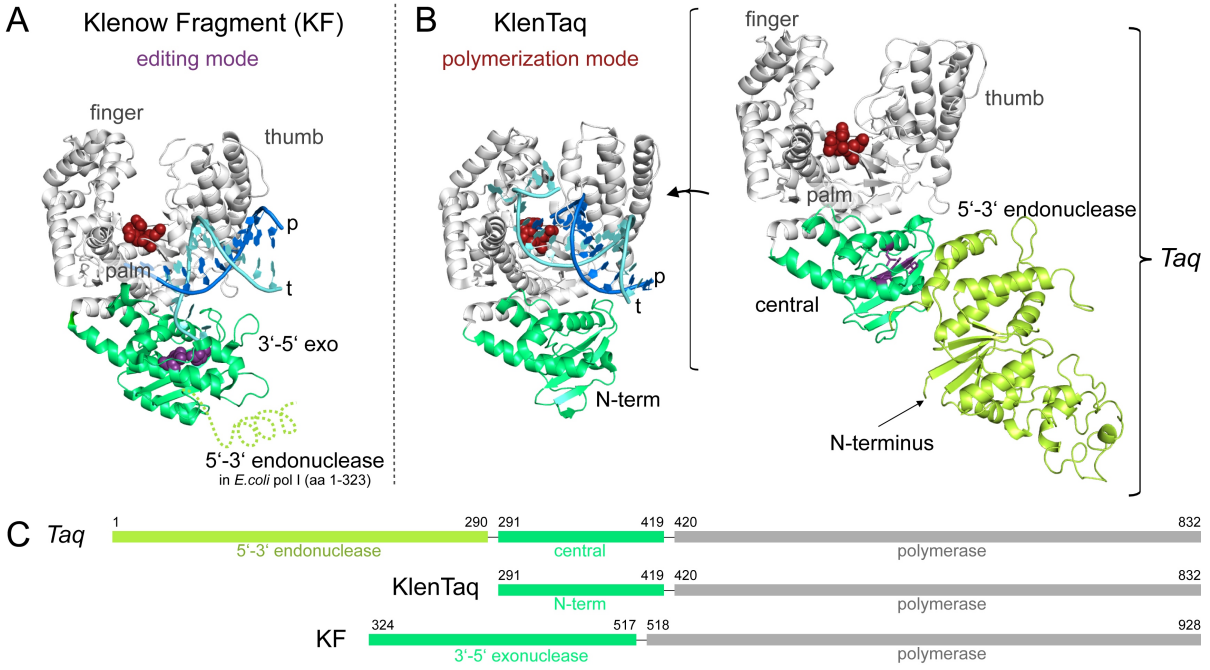


Figure 1.8: Crystal structures of **A**: the *E. coli* Klenow Fragment (KF, PDB ID: 1KLN[42]) and **B**: *Taq* (PDB ID: 1TAQ[43]) and KlenTaq (PDB ID: 3SV3[44]). The overall structures are shown as cartoon with the domains colored as indicated in panel C. The position of the 5'-3' endonuclease domain in *E. coli* pol I is indicated by dashed lines. Active site residues are shown as spheres colored red for the polymerase domain and purple for the 3'-5' exonuclease domain, respectively. For *Taq* (no 3'-5' exo activity) the residues corresponding to active site residues of the 3'-5' exo domain in KF are shown as purple sticks. The DNA in the KF and KlenTaq structures, bound in the editing or polymerization mode, respectively, is shown in blue (primer) and cyan (template). **C**: Domain architecture in *Taq*, KlenTaq and KF.

1.2.2.2 KlenTaq DNA polymerase

The KlenTaq DNA polymerase is the large fragment of DNA polymerase I from the extreme thermophile *Thermus aquaticus*, a gram-negative bacterium which was isolated from thermal springs or man-made thermal habitats and has an optimum growth temperature of 70°C [45]. The full-lengh enzyme is called *Taq* DNA polymerase and was first purified and characterized in 1976 by Trela *et al.*[46]. The *Taq* DNA poylmerase is thermostable with a temperature optimum of 76-80°C and a half-life of 9 min at 97.5°C[47]. It is homologous to the *E.coli* DNA polymerase I (sequence identity: 35.1%). A structure of full length *Taq* is shown in **Figure 1.8** B (right side). The enzyme is divided into two major domains, the N-terminal domain or nuclease domain (residues 1-290), hosting the

5'-3' endonuclease activity and the bigger C-terminal domain (291-832) which includes the polymerase domain with finger, thumb and palm subdomains involved in polymerase activity and the central domain. This central domain is equivalent to the 3'-5' exonuclease proofreading domain in *E. coli* pol I but *Taq* completely lacks this activity as it harbours only hydrophobic regions instead of catalytic residues in this section[48]. Analogously to the Klenow Fragment of *E. coli* (described above) a similar fragment was created for *Taq* DNA polymerase I and named KlenTaq[48] (**Figure 1.8 B**, left side). As now the central domain is located at the N-terminus it is called N-terminal domain (short N-term). Due to thermostability *Taq* and KlenTaq are widely used in PCR[41].
The large fragment of *Taq* DNA polymerase used in this study entails amino acids 293-832 of the full length *Taq* and is referred to as KlenTaq or KTQ throughout the work. Also for *Bacillus stearothermophilus* pol I a similar fragment (*Bacillus* fragment, short BF[49]) has been created and is widely used for structural and mechanistical studies of DNA polymerases.

1.2.2.3 Family B DNA polymerases

The B-family of DNA polymerases comprises the *E. coli* pol II, eukaryotic replicative polymerases α, δ and ε as well as DNA polymerases encoded on plasmids of mitochondria, various fungi and plants. Furthermore viral DNA polymerases, some polymerases of bacteriophages (such as T4, phi29 and RB69) and archaebacterial DNA polymerases belong to the B-family[32]. The members are mainly involved in DNA replication, carrying out processive DNA synthesis and are able to correct errors by their strong 3'-5' exonuclease activity, which is over a 1000 times higher than that of *E. coli* pol I[32].
In *in vitro* applications the archaeal B-family polymerases are especially useful for the incorporation of modified nucleotides. In several cases it has been shown that the insertion of sugar[50–52] or nucleobase[7, 14, 16, 53–56] modified nucleotides is accomplished more efficiently by archaeal B-family members than by polymerases from the family A such as *Taq* or KlenTaq DNA polymerase. The reasons for this preferential acceptance of modified substrates has not yet been elucidated, mainly due to missing structural data.
Some prominent archaeal B-family polymerases that are used in biochemical and/or diagnostic applications, are *Thermococcus* species 9°N-7 (9°N)[57], *Thermococcus kodakarensis* (KOD)[58, 59], *Pyrococcus furiosus* (Pfu), Vent DNA polymerase from *Thermococcus litoralis* and Deep Vent polymerase from *Pyrococcus* species strain GB-D and all of them are commercially available.

1.2.3 Structure of DNA polymerases

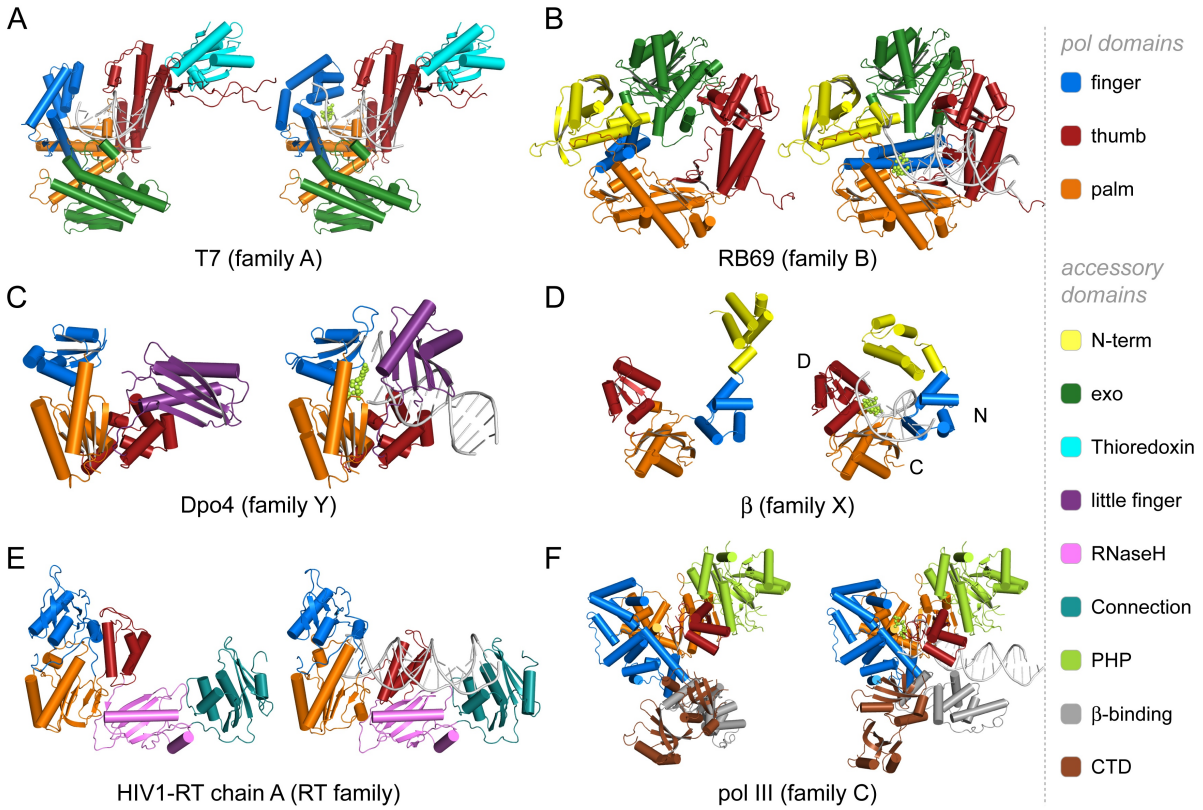


Figure 1.9: Apo and binary or ternary structures of representatives of different DNA polymerase families. The different domains are colored with the color code indicated on the right. The DNA is shown in grey and the bound triphosphate substrate is shown as spheres (if available) in light green. In each panel the apo structures are shown on the left side and the binary or ternary structures are shown on the right side. **A**: Family A T7 DNA polymerase binary structure (PDB ID: 1SKS) and ternary structure with bound ddATP (PDB ID: 1SKR)[60]. **B**: Family B apo RB69 structure (PDB ID: 1IH7) and replicating complex with bound dTTP (PDB ID: 1IG9)[61] of the exonuclease deficient mutant (D222A/D327A). **C**: Family Y Dpo4 apo (PDB ID: 2RDI)[62] and ternary complex with bound ddGTP (PDB ID: 1JXL)[63]. **D**: Family X DNA polymerase β apo form (PDB ID: 1BPD)[64] and ternary complex (PDB ID: 2BPG)[65]. **E**: RT family apo HIV1-RT chain A (p66 residue 1-556) of the heterodimer (chain B: p51 residue 1001-1472 is not shown) (PDB ID: 1HMV) and in complex with DNA (PDB ID: 2HMI)[66]. **F**: Family C polymerase III from *thermus aquaticus* in apo form (PDB ID: 2HPI)[67] and ternary complex with bound dATP (PDB ID: 3E0D)[68].

After the first crystal structure of a DNA polymerase was solved in 1985[26] crystal structures of DNA polymerases from all different DNA polymerase families were reported up to date. The structurally least characterized family is family D for which only the structure of the N-terminal domain of *Pyrococcus horikoshii* polD exists (PDB ID: 3O59)[69]. The initial comparison of the polymerase architecture to a "right hand" observed in the first structure of KF DNA polymerase[26] was transferred to various subsequently solved

structures of DNA polymerases from different families. Hence, the polymerase subdomains are generally subdivided into a finger, thumb and palm domain. To get a rough impression of the differences and similarities of polymerase structures either apo, binary and/or ternary complexes of DNA polymerases from different families are shown in **Figure 1.9** with the respective domains in the same color. Generally, the DNA is bound in an U-shaped crevice between the finger (blue) and the thumb (red) domain which is often equipped with basic protein side chains that can interact with the negatively charged DNA backbone. The palm subdomain (orange) forms the floor of this cleft. Upon DNA binding the thumb domain usually undergoes a conformational change to make contact to the DNA duplex (shown for family A, B, C and Y members[68, 70]). The catalytic site is located in the palm domain which includes the catalytically essential amino acids (aspartate and/or glutamate) that coordinate the two metal ions (usually Mg^{2+}) necessary for the phosphoryl transfer reaction. The finger domain is responsible for binding of the substrate triphosphate. In some polymerases it has been shown that the finger domain changes its conformation from an open to a closed state in order to generate an active enzyme-substrate complex. This motion was first visualized in the open binary and closed ternary crystal structures of KlenTaq DNA polymerase published by Waksman *et al.*[71] and has also been observed in structures of other DNA polymerases afterwards e.g. in B-family[61, 72], X-family[65, 73] and RT-family members[74–76]. In addition to the polymerase domain some polymerases possess extra accessory domains that help to fulfill their respective biological function. Examples are N-terminal, 3'-5'-exonuclease, lyase (pol β, family X), little finger (Dpo4, family Y), processivity factor (A-family, e.g. thioredoxin in T7) or RNAseH (RT family) domains[32]. Further information to structure and role of these additional domains is found in the publications corresponding to the respective structures (references see caption of **Figure 1.9** or in current reviews: X-family[36], Y-family[77].

Although DNA polymerases of the different families bear little sequence identity their polymerase active sites are similar[49]. A common catalytic core with two key active site metal ions coordinated by two invariant aspartate residues and one additional variable aspartate or glutamate residue is conserved in all families. These residues are positioned in two common motifs, motifs A and C, shared by all polymerases[80] and exhibiting very similar folding patterns. The sequences of these motifs are well conserved within a family but only partially conserved between families A and B[81]. An additional motif found in family A and B polymerases is motif B (KxxNFGxxYG in polymerases I of the

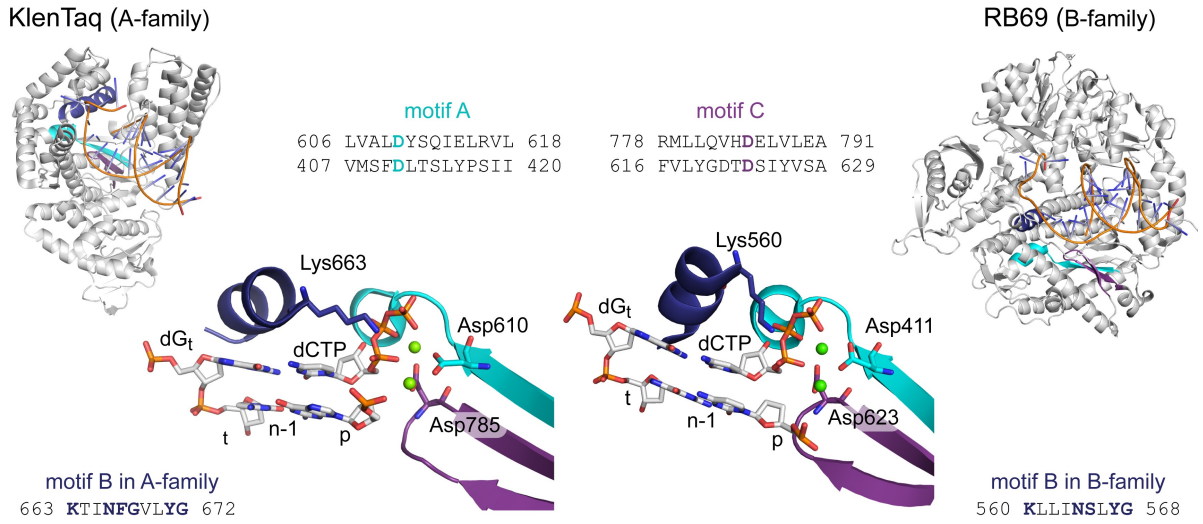


Figure 1.10: Similar folding patterns of motifs A (cyan), B (blue) and C (purple) in polymerase A- and B-family members. The overall structures as well as a zoom into the motifs and the n and n-1 base pairs are shown for KlenTaq (PDB ID: 3RTV[44]) and RB69 (PDB ID: 3NCI[78]) ternary complexes. Highly conserved aspartate and lysine residues interacting with the triphosphate substrate are shown as sticks. Metal ions are shown as spheres. Alignment of motif A and C as well as the patterns for motif B conserved within polymerase A- and B-families are taken from [79].

A-family and KxxxNSxYG in the B-family[82–84]) which is located in the finger domain and residues within are interacting with the triphosphate substrate and take part in the reaction mechanism. As an example, motifs A, B and C are visualized for KlenTaq (family A) and RB69 (family B) DNA polymerases in **Figure 1.10**. A recent study analyzing 15 polymerase ternary substrate complex crystal structures reveals that these critical elements as well as the incoming nucleotide are highly conserved in structural space between polymerase families[85].

It is still not fully understood where and how dNTPs first bind in polymerases of different families although for some polymerases assumptions have been made based on structural or kinetic experiments. In A-family members it is assumed that a dNTP binds to the finger domain in the open state of the enzyme[86] and the substrate is delivered into the active site by finger domain closure. Furthermore it is supposed that within this closing step a selection of the correct nucleotide already takes place[86]. In contrast to this first contact via the finger domain it was recently suggested that dNTPs get access to the catalytic core of the polymerase via diffusion through a channel in A, B and X families[85]. These somehow contradictory results indicate that there is still necessity for further investigation. In the B-family polymerase RB69 the insertion site is already defined in the open state of the enzyme[61] (in contrast to A-family polymerases, see next

section) and a kinetic study of active site mutants shows that the dNTP can directly bind to it without foregoing conformational changes[82]. As in the present work representatives of family A and B polymerases were investigated their structure will be described in more detail in the following.

1.2.3.1 Structure of A-family DNA polymerases

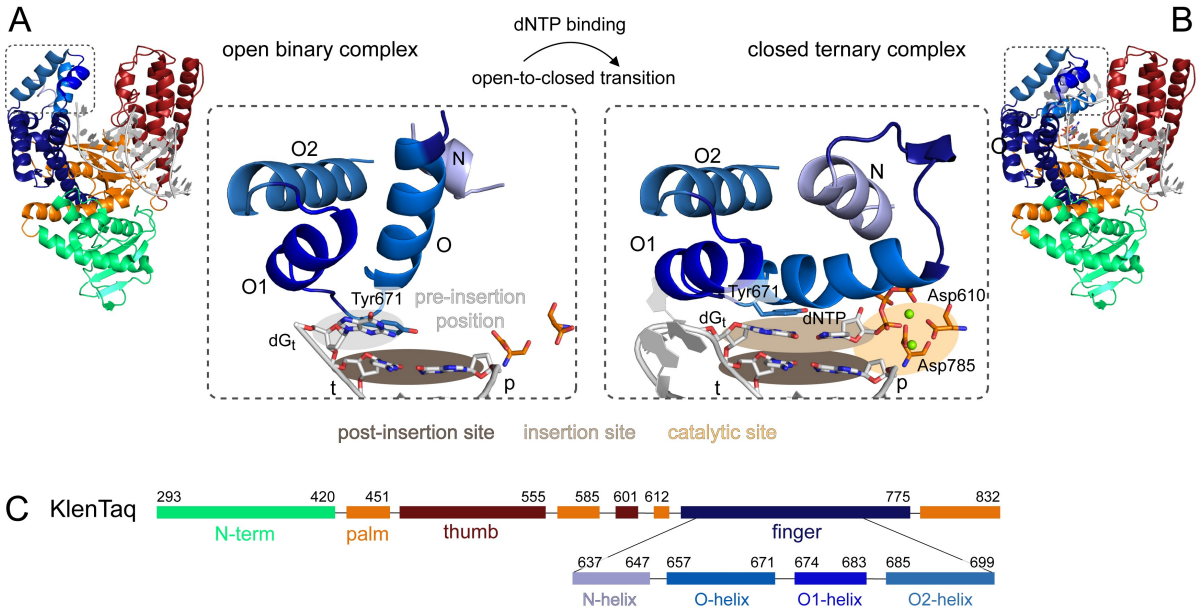


Figure 1.11: KlenTaq polymerase transition from an open to a closed state upon dNTP binding. **A**: Open binary complex of KlenTaq (PDB ID: 3SZ2[44]). **B**: Closed ternary complex of KlenTaq (PDB ID: 3RTV[87]). Polymerase domains and helices of the finger domain are colored as shown in panel C. Defined sites involved in catalysis and finger domain helices N, O, O1 and O2 are shown in the zoom in views. Sites are indicated by colored ovals: post-insertion site (n-1 position) in brown, pre-insertion position (n in binary complex) in grey, insertion site (n in ternary complex) in light brown and catalytic site in orange. **C**: Scheme of KlenTaq domain architecture with different subdomains shown in different colors and numbered accordingly. Relevant finger domain helices are shown in different shades of blue.

Numerous structures of viral (T7) and bacterial (KF, *Taq*, KlenTaq, and BF) A-family DNA polymerases (wild type (WT) and mutants) in apo forms or in binary or ternary complexes have been solved. These include complexes with different natural or modified substrates, DNA lesions, ions, etc. The available structures greatly contribute to the understanding of DNA polymerase mechanisms in the incorporation of natural or modified substrates or the processing of DNA lesions. Recently, the structure of a human member of the family, the mitochondrial polymerase γ has been solved[88], completing the structural data of A-family polymerases by a eukaryotic representative. As in this work

KlenTaq was used for structural studies its structure will be described exemplarily for A-family polymerases in more detail in the following.
The overall structure of KlenTaq is similar to a right hand as has already been mentioned before. The polymerase domain is subdivided into the finger, thumb and palm domain, followed by the N-terminal domain (**Figure 1.11 C**). The N-terminal domain has no similarity to the N-terminal domain found in other polymerases (e.g. B-family polymerases). Upon DNA binding the thumb domain closes and together with the finger domain interacts with the DNA. In this binary state the templating nucleotide adopts an extrahelical position (referred to as pre-insertion position), realized by a rotation around the phosphodiester bond in the template strand (**Figure 1.11 A**, zoom in) This position of the templating nucleotide allows no direct interaction with the incoming dNTP in an open finger conformation. Instead of the templating nucleotide a tyrosine residue (Tyr671) of the O-helix stacks on top of the terminal base pair. A large conformational change of the finger domain is observed after triphosphate binding (**Figure 1.11 B**). Within this rearrangement the tyrosine is moved away and the templating nucleotide can be rotated back, adopting a position suitable to pair with the incoming substrate within the insertion site. Simultaneously, the O-helix residues come closer to the p/t duplex defining a tightly restrained insertion site on top of the nascent base pair (**Figure 1.11 A** vs. **B**). In this closed complex the substrate triphosphate is located in the insertion site where it establishes hydrogen bonds with the templating nucleotide. The conformational change(s) after triphosphate binding also position the catalytic residues and metal ions in the catalytic site, properly oriented for the coming reaction. It is still a matter of debate which of the conformational changes occuring prior to the chemistry step is rate limiting in the reaction[28]. However, several studies exclude that the large open-to-closed movement of the finger domain determines the overall speed of the reaction[89]. Therefore, a smaller rearrangement following this step might be rate-limiting, instead. In KlenTaq the two catalytic Mg^{2+} ions are coordinated by the triphosphate moiety of the substrate, the two conserved aspartate side chains in the active site (Asp611 and Asp785) as well as two water molecules[44]. After a successful incorporation event the polymerase moves to the next templating position and the newly formed base pair is handed on to the post-insertion site (n-1).
As mentioned above, structural data indicate that in case of A-family polymerases dNTPs bind to the open state of the enzyme. First interactions are thereby mediated by basic residues of the finger domain with the triphospahte moiety of the substrate. As the tem-

plating nucleotide is flipped away in the open state is is unable to form hydrogen bonds with an incoming nucleotide without a conformational change. However, the conformational closing movement is not needed to be accomplished completely for the substrate to be able to interact with the templating nucleotide. In a recent study the large fragment of DNA polymerase I from *Bacillus stearothermophilus* (BF) was trapped in a partially closed state (also termed "ajar") with a mismatched pair[90]. In this ajar state the O-helix is moved somewhat towards the insertion site, whereby the templating nucleotide is flipped into the insertion site and forms hydrogen bonds with the incoming triphosphate. It is suggested that this intermediate, partially closed state is one of the checkpoints in which selection for correct nucleotides takes place. Thereby, matched or mismatched nucleotides induce different conformational changes, leading to a tight, catalytic complex in case of a correct substrate or release of the dNTP in case of a mispair[86].

1.2.3.2 Structure of B-family DNA polymerases

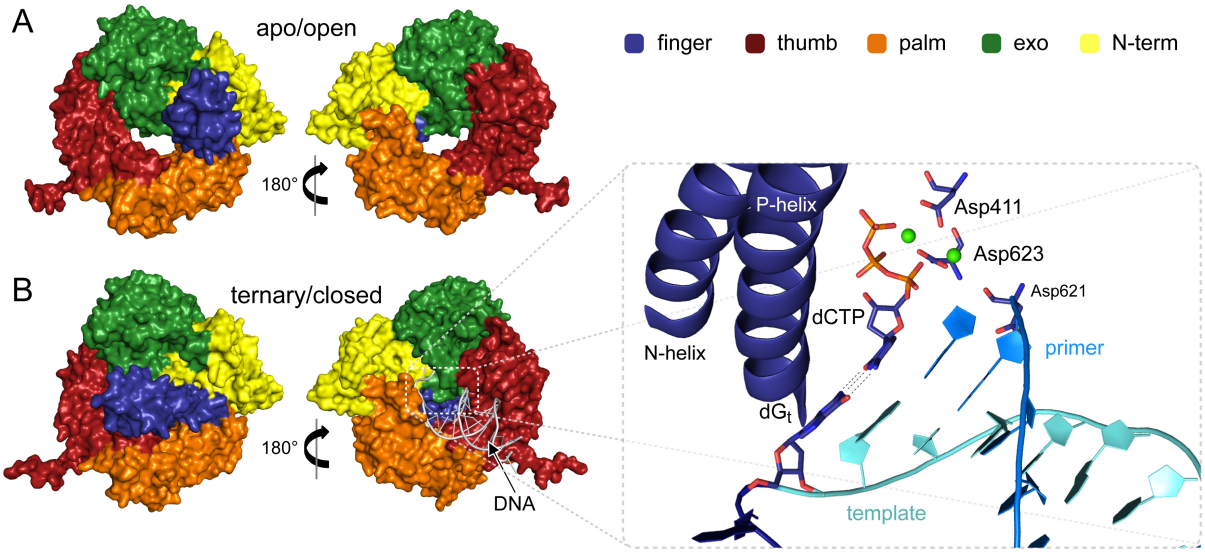


Figure 1.12: Structure of RB69 DNA polymerase **A:** in its apo form (open structure, PDB ID: 1IH7[61]) or **B:** as a ternary complex (closed structure, PDB ID: 3NCI[78]). The enzyme is shown as surface with the domains colored as indicated in the figure. The DNA is shown as grey cartoon. A detailed view of the active site is given on the right side. Here, the two finger domain helices as well as the p/t complex is shown as cartoon. The incoming and templating nucleotide and the protein residues involved in catalysis are shown as sticks. Ca^{2+} ions are shown as green spheres.

Besides the great amount of A-family polymerase crystal structures, a reasonable number of crystal structures of polymerases belonging to family B have also been solved in the last decade. The first B-family polymerase structure solved was the apo structure of bac-

teriophage RB69 gp43 replicative DNA polymerase (short RB69) published in 1997 by Wang *et al.*[72]. Still, this enzyme is the structurally best characterized B-family member with numerous WT and mutant crystal structures of binary and ternary complexes solved. Examples are structures with natural substrates[61, 91], DNA lesions (8-oxodG, abasic site analogue[92, 93] and thymine glycol[94, 95]), mismatched nucleotides[96] and nucleotide analogs (2,4-difluorotoluene[97] and 3-deaza-2'-deoxyadenosine[98]). Ternary complexes of the polymerase domains of eukaryotic pol α[99] and δ[100] were also reported and recently a structure of the catalytic core of yeast DNA polymerase ε has been solved in a ternary complex[101]. From the archaeal members of the family B, however, no binary or ternary complex structure was solved up to the beginning of this work. Due to the relatively low sequence similarity of the bacteriophage RB69 DNA polymerase to archaeal B-family polymerases (only 18% sequence identity[61]) it is difficult to draw conclusions from the RB69 structures on the mechanisms of the archaeal polymerases. Only several apo DNA polymerase structures (e.g. from *Thermococcus gorgonarius* (Tgo) [102], 9°N[57], KOD[58], Pfu[103], *Sulfolobus solfataricus* (Sso)[104], *Desulfurococcus* strain Tok (D.Tok)[105]) and structures with the DNA bound in the editing mode[106–108] were available when this project was started. Ternary structures of archaeal representatives of the family are missing completely. An overview of B-family DNA polymerases structures in an apo form or in complex with natural substrates is given in **Table 7.3** in the Appendix on page 178.

In analogy to the structure of the Klenow Fragment, the polymerase domain of B-family polymerases is also divided into finger, thumb and palm domains (**Figure 1.12**). Together with the 3'-5' exonuclease and the N-terminal domain the overall structure resembles a circular molecule with a central hole (in the open state). As no binary or ternary struture of 9°N, the B-family polymerase in focus of this work, is present, substrate binding and active site arrangements of the B-family polymerases are shown using the crystallographic data of RB69 DNA polymerase instead. In the apo form the enzyme adopts an open conformation characterized by an open thumb and finger domain (**Figure 1.12 A**). The fingers domain consists of two long alpha helices, named N- and P-helix. Similar to A-family polymerases the finger domain closes upon dNTP binding generating a closed active site (**Figure 1.12 A** vs. **B**). The palm subdomain contains the catalytically important aspartate residues (Asp411 and Asp623 in RB69, **Figure 1.12**, zoom in and Asp404 and Asp542 in 9°N) that coordinate the metal ions involved in phosphodiester bond formation.

The 3'-5' exonuclease domain is homologous to the one in the Klenow Fragment[32] and harbours the two catalytic residues which are conserved among the 3'-5' proofreading domains of many DNA polymerases[109]. For crystallographic studies focussing on the replicative mode, these residues (Asp141 and Glu143 in 9°N) are often mutated to alanine to avoid degradation of the added primer or template oligonucleotide. Polymerases lacking 3'-5' exonuclease activity are ususally termed exo-. A comparison of archaeal enzymes with the related DNA polymerase of bacteriophage RB69 shows that the thermostable enzymes have a more compact structure with fewer and shorter loops and at least one disulfide bridge. These differences might be an effect of their adaption to the growth at extreme temperatures[110].

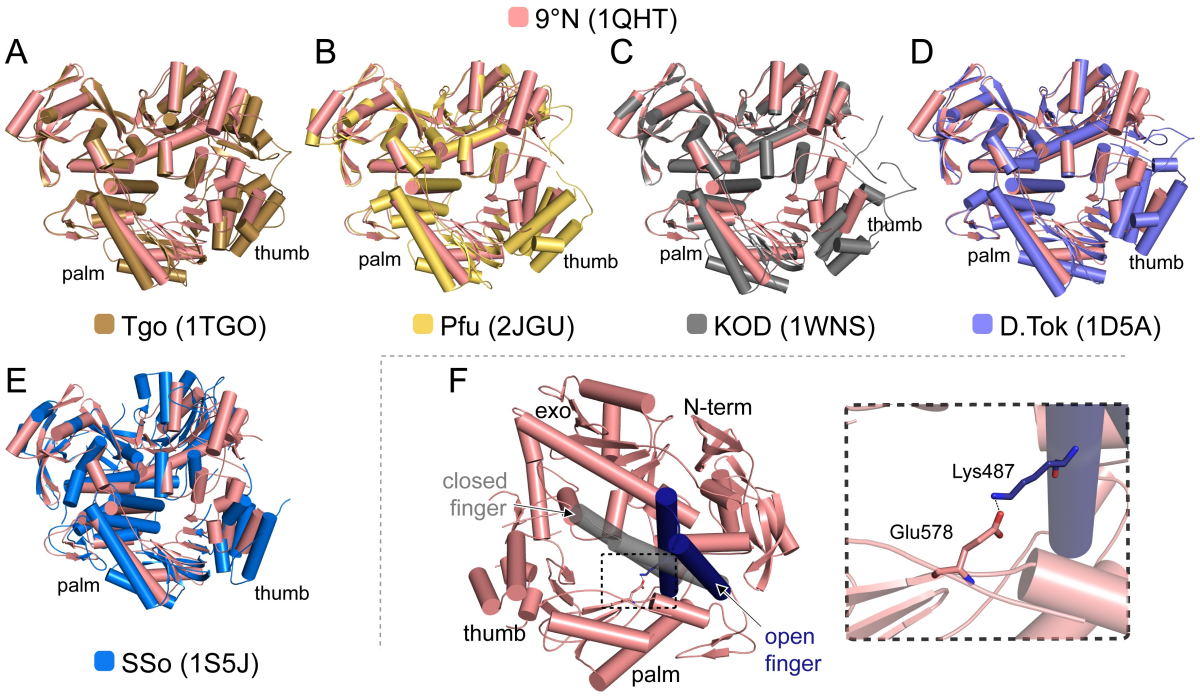


Figure 1.13: **A-E**: Overlay of the 9°N apo structure 1QHT[57] with archaeal B-family apo structures from **A**: *Thermococcus gorgonarius* (Tgo)[102], **B**: *Pyrococcus furiosus* (Pfu)[103], **C**: *Thermococcus kodakarensis* (KOD)[58][59], **D**: *Desulfurococcus* strain Tok (D.Tok)[105]) and **E**: *Sulfolobus solfataricus* (Sso)[104]. PDB IDs of the respective structures are given in the figure. **F**: Structure of 1QHT is shown in an orientation ~180°C rotated around y compared to A-E. The salt bridge in 1QHT which keeps the palm domain in a closed conformation[61] is shown enlarged in the zoom in box. The finger domain is shown in blue and adopts an open conformation. The closed finger domain of the superposed ternary structure of yeast pol δ (PDB ID: 3IAY[100]) is shown in grey.

Apo structure of the DNA polymerase 9°N.

The apo structure of 9°N reported in the year 2000 by Rodriguez *et al.* shows some major differences to related B-family DNA polymerase apo structures, e.g. KOD[58], Tgo[102], D.Tok[105], SSo[104] or Pfu[103]. Although DNA polymerases usually adopt an open conformation if not bound to DNA[72, 102] the apo structure of 9°N is in a closed conformation[57]. Actually the finger domain (which is usually meant referring to an open or closed state) is in a closed state only with respect to the palm domain but if the overall enzyme structure is superposed with other B-family polymerase structures rather the finger is in an open state. The palm and the thumb, however, are in a closed state. This different situation is visualized in **Figure 1.13** where the apo structure of 9°N - named 1QHT throughout this work - is shown in an overlay with other binary archaeal B-family polymerase structures. Although the palm and thumb domain positions vary somehow throughout the shown structures, indicating their flexibility in the uncomplexed state, it is obvious that only in case of 1QHT these two domains are in a closed state. The fact that the finger domain is positioned in an open state when regarding the overall structure is visualized in panel **F** of **Figure 1.13**. The difference between a closed and the open finger domain is indicated by a superposition of the ternary structure of polymerase δ (PDB ID: 3IAY) and showing its closed finger domain (grey) compared to that of 9°N (blue). Rodriguez *et al.* state that the closed conformation of apo 9°N (of the palm and thumb domain) is stabilized by a salt bridge between the conserved residues Glu578 in the palm and Lys487 of the fingers domain (**Figure 1.13 F**). Indeed this salt bridge seems to keep the palm domain in a "artificially closed" conformation that is not observed in other B-family structures neither in an apo form, nor in a DNA or nucleotide complexed form. Therefore, this structure seems to be biased by crystal contacts and a binary structure of 9°N in a fully open complex would be required to be better able to describe movements in the enzyme upon DNA or substrate binding as soon as a binary or ternary structure is available.

1.2.4 Selectivity of DNA polymerases

Accurate replication of DNA is essential for the viability of all living organisms. This need calls for the error-free performance of DNA polymerases which should duplicate DNA by incorporating only complementary nucleotides according to the Watson-Crick base pairing rule and avoid any mispairs. Thereby the polymerase has to select the one correct pairing

partner for a given templating nucleotide out of the cellular nucleotide pool. Despite many biochemical and structural studies elucidate the selectivity of DNA polymerases, this process is still not fully understood. The propensity by which a polymerase inserts a correct or incorrect nucleotide is described by the terms "efficiency" or "specificity", and is usually measured by insertion kinetics.

DNA polymerase fidelity is defined by the catalytic efficiency for the insertion of a correct nucleotide divided by the efficiency of incorrect nucleotide insertion. It was shown that the difference between low and high fidelity DNA polymerases is mainly related on the efficiency of correct nucleotide insertion whereas insertion of incorrect nucleotides is similar[111]. Correct nucleotides are bound with similar affinities by polymerases from different families (2-55μM)[112] but the intrinsic rate constant for insertion varies over 10^5-fold[111]. This means that the decreased incorporation efficiency for correct nucleotides of low-fidelity polymerases is attributed to the slow insertion rather than to binding affinity. The fidelity of nucleotide incorporation relies on three main steps: (1) The selection of the correct nucleotide during the polymerization reaction (2) the excision of misinserted nucleotides by an associated 3'-5' exonuclease activity and (3) postreplicative excision of an incorrectly inserted nucleotide by mismatch repair mechanisms. A reduced rate of extension beyond a mismatch by DNA polymerases provides time for the exonuclease activity to process the primer-terminus or leads to dissociation of the polymerase after misinsertion so that the mismatch can be removed by various cellular nucleases[28]. Replicative DNA polymerases performing step 1, the insertion reaction, are indeed very selective enzymes with fidelities up to 10^{-6}. Selectivity varies depending on the polymerase and the sequence environment[113]. Repair polymerases in contrast exhibit lower fidelity at undamaged templates[114]. The second and third step: proofreading ability through 3'-5' exonucleases as well as other mismatch repair mechanisms, add to the accuracy of DNA polymerases by a factor of 100-1000 so that a final fidelity of up to 10^{-10} for DNA replication can be reached[28].

The factors contributing to the selectivity of nucleotide incorporation by DNA polymerases are diverse and discrimination is gained at different steps of the nucleotide insertion pathway. Furthermore, factors contributing to fidelity and the extent of their influence differ among DNA polymerases and therefore it is not possible to define one selectivity mechanism that holds true for all of them[115]. Also, the bypass of DNA lesions such as e.g. a non-instructional abasic site and the incorporation of mispaired nucleotides again show different mechanisms and follow different kinetics than the incorporation of

canonical nucleotides[115, 116]. Therefore for each polymerase and each situation (with either a correct or incorrect nucleotide, incorporation of modified/artificial nucleotides or in bypass of a lesion, ...) different effects may play a role to promote or hinder the incorporation reaction. Some factors that are involved in nucleotide insertion by polymerases will be described below but detailed information is found in several reviews[28, 116].

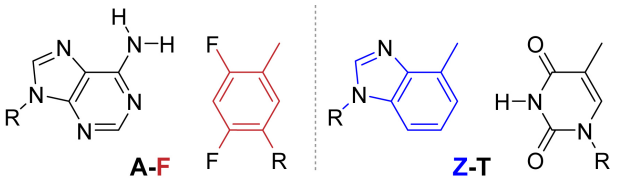


Figure 1.14: Structures of the isosteric nucleotide analogs 2,4-difluorotoluene (**F**) and 4-methylbenzimidazole (**Z**) forming base pairs with the natural nucleobases A and T, respectively.

Hydrogen bonds between the pairing partners and the geometric fit of the base pair into the active site of the enzyme play major roles in discrimination of correct and mismatched nucleotides. Additional contributions to efficient and selective replication of the natural nucleotides are made by minor groove hydrogen bonding, solvent effects and π-π stacking interactions[115, 117]. The specific hydrogen bonds formed by the Watson-Crick base pairs have long been regarded as the main selectivity determinants[6, 118]. However, the small free energy barriers between matched and mismatched base pairs in solution[113, 119, 120] would account for error rates of 1:100 and therefore revealed that the selectivity of DNA polymerases can not depend primarily on hydrogen bonding capability[33]. Later studies with nucleotide analogs showed that shape complementarity of a formed base pair and its steric fit into the enzyme's active site must also contribute to the fidelity of DNA polymerases[116, 121]. Kool *et al.* showed that nucleotide analogs which are similar in shape to the natural ones but unable to form hydrogen bonds (e.g. 2,4-difuorotoluene (**F**) and 4-methylbenzimidazole (**Z**), **Figure 1.14**) are selectively incorporated by DNA polymerases[116]. However, the reliability of this study is questioned, as difluorotoluene exhibits H-bonding capabilities under specific conditions[122]. Nevertheless the importance of this study was and is enormous as it can be defined as the starting point for researchers to investigate selectivity factors of DNA polymerases other than hydrogen bonds. It is still a matter of debate to what extent hydrogen bonds and shape complementarity contribute to the selectivity of DNA polymerases but it is clear that both play an important role[123]. Therefore, all the above-mentioned factors and probably some

more contribute to polymerase selectivity and, depending on the polymerase and the kind of nucleotide incorporated, they do so with different extent. Some selectivity factors are described briefly in the following.

1.2.4.1 Steric effects

The steric effects that play a role in nucleotide insertion are based on the shape and size complementarity of the template base and the base of the incoming nucleotide on the one hand, and on the tightness of the enzyme's active site around the nascent base pair on the other[116]. Kool and his group could first show that 2,4-difluorotoluene (shape analog of dT) is both incorporated opposite a templating dA with remarkable efficiency[124] and specifically codes for the incorporation of dA using several polymerases (KF, T7 and HIV-RT)[124–126]. This study revealed that DNA polymerization can occur in the absence of hydrogen-bonding between the pairing nucleobases, as mentioned above. Other polymerases, however, do not process difluorotoluene as a substrate (e.g. pol α and β)[127] again indicating that selection mechanisms are different among polymerases. Replication without hydrogen bonds was afterwards shown for many other hydrophobic analogs, either isosteric to the natural ones or with altered size[116]. The finding that H-bonds are not absolutely required to form a stable base pair was the basis for the consideration of hydrophobic nucleobases in the process of developing a third base pair that would expand the genetic code (see section 1.3). A prerequisite for DNA polymerases to simultaneously recognize the four natural pairs (dA-dT, dT-dA, dG-dC and dC-dG) but to exclude any kind of mispairs via steric constraints is that the correct pairs have similar shape or possess specific characteristic functional groups located at similar positions in all pairs that can be contacted by the polymerase. Indeed, the four natural base pairs are similar in size with a thickness of around 3.4Å (thickness of aromatic ring) and a constant distance between the glycosidic bond positions of the two pairing partners (C1'-C1'-distance = 10.7-11.0Å)[116]. In an overlay of all four pairs Kool *et al.* visualized that they only differ in size and shape in the major and minor groove dimensions but are similar in overall length and a consensus base-pair shape can be described (**Figure 1.15 B**). This consensus pocket is thought to define the base pair dimension which a base pair should adopt to be successfully processed by a polymerase in the insertion step. A nucleotide forming a mispair that would not fit into the pocket and cause steric clashes in turn will either not bind or if it binds the triphosphate moiety will not be properly aligned for a subsequent

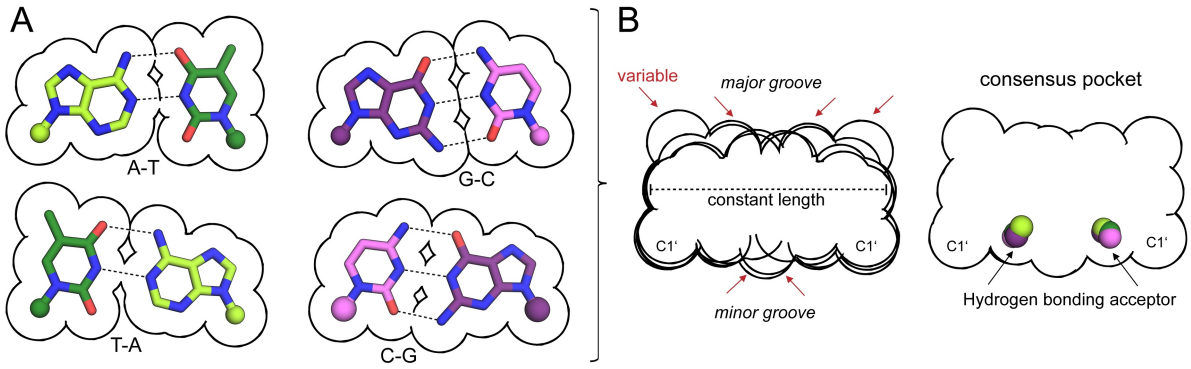


Figure 1.15: Space-filling shapes and consensus pocket of the four natural base pairs. **A**: The base pairs shown are from the insertion sites of BF crystal structures: dG-ddCTP, dC-ddGTP, dA-ddTTP and dT-ddATP with PDB IDs: 4DQP[128], 3TI0, 3PV8 and 3THV[78], respectively. Only the nucleobase moieties (as sticks) and the C1' atoms (as spheres) are shown. **B**: The consensus pocket was generated by an overlay of all pairs in PyMOL. Variable positions are indicated by red arrows. The figure was created in analogy to Figure 7 and 8 in [116].

insertion reaction. Kool *et al.* describe this situation in their "size exclusion hypothesis". As the pocket is only tightly constrained by the enzyme in the closed state the closing movement of the finger domain has to take place before the reaction. It is likely that full closure can only take place in case of a correctly shaped pair whereas mispairs prevent closure.

In an extension to the steric fit model the rigidity of the enzyme's active site surrounding the binding pocket is taken into account in the so called "active site tightness hypothesis". Since the four natural base pairs are similar in length and thickness but vary in the minor and major groove the enzyme needs to be flexible in some parts and rigid in others. Tight interactions of the protein and the base pair is present in high-fidelity polymerases so that already small variations from a natural pair causes steric problems. In low-fidelity polymerases, in contrast, the pocket is less rigid and can adapt to changes in size and/or shape of the base pair. For detailed information about the active site tightness and substrate fit hypothesis see [116].

1.2.4.2 Minor groove hydrogen bonding

In addition to the interbase Watson & Crick hydrogen bonds also hydrogen bonds between the minor groove edge of the nucleobases in the DNA duplex and the DNA polymerase

contribute to selectivity. These minor groove H-bonds are not only formed with the newly established base pair in the insertion site, but correct base pairs are also probed via the minor groove in the post-insertion site and several positions downstream of the DNA template[129]. Ternary complexes of DNA polymerases with DNA and dNTP substrate showed that functional groups at the minor groove side of the nucleobases (O2 atoms of pyrimidines and N3 atoms of purines) specifically interact with the side chains of amino acids in proteins either directly or via water molecules[61, 64, 71, 74, 98, 129–132]. As all four nucleotides carry these groups donating lone-pair electrons at similar relative positions from the nucleic acid backbone (see **Figure 1.15**, indicated by colored spheres) polymerases can recognize all four base pairs with the same interaction mechanism[133]. Disruption of the interaction patterns in the minor groove can lead to misalignment of the components involved in catalysis in the active site and therefore hinder insertion or elongation of non-canonical nucleotides[134].

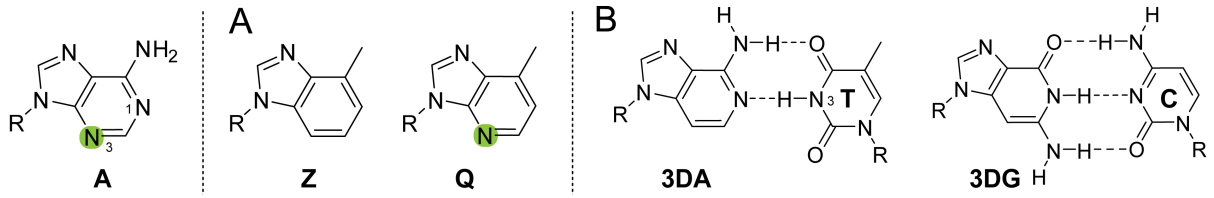


Figure 1.16: A: Nucleobase analogs of adenine which are unable to form hydrogen bonds at the Watson-Crick edge and/or on the minor groove side. B: Analogs of adenine or guanine which lack minor groove H-bonding acceptors but keep Watson-Crick H-bonds. Minor groove H-bonding acceptors are shown in green.

The role of these specific anchor points in nucleotide insertion was first investigated in the Kool lab using two non-polar nucleoside isosteres of dA that either possess or lack Watson-Crick H-bonding donors pointing to the minor groove. The first is d**Z** that was already used to evaluate steric effects in nucleotide selection (section 1.2.4.1) and the second is d**Q** (9-methyl-1-H-imidazo[(4,5)-b]pyridine) which also lacks all Watson-Crick H-bonding groups but compared to d**Z** retains the minor groove acceptor nitrogen N3 of adenine (**Figure 1.16 A**). Single nucleotide insertion experiments with dA, d**Z** and d**Q** at the templating position and dATP, d**Z**TP and d**Q**TP substrates, respectively, revealed that the presence or absence of a minor groove H-bond acceptor within the newly formed base pair has little influence on the insertion reaction[127, 135] in KF, T7, *Taq* and HIV1-RT. In other polymerases, in contrast, improved incorporation is observed in presence of a single minor groove interaction at either the template base or the incoming

triphosphate or at both positions[127]. Elongation studies implied that a hydrogen bond between the primer-terminal nucleobase (position n-1) and the polymerase is essential for the elongation reaction in KF, KlenTaq and HIV1-RT[135, 136] but that a H-bond acceptor at the n-1 template nucleotide is not markedly promoting the reaction. In the last years, the suitability of d**Q** and d**Z** as purine analogues was questioned as their C1-H groups can clash with the N3-H of a pairing dT, resulting in a tilt of the dT nucleobase by around 25°, as was shown in a ternary structure of RB69 DNA polymerase with a dT-d**Q**TP pair in the insertion site[98]. Therefore, effects observed with the d**Z** analog might also result from the different geometry of the base pairs formed and not only from its missing minor groove H-bonding capability. As revised analogues, 3-deazaadenine (**3DA**) or 3-deazaguanine (**3DG**) (**Figure 1.16 B**) were used to structurally and functionally investigate the role of minor groove H-bonds in DNA synthesis by polymerases[90, 134, 137]. These studies further reveal that minor groove H-bonding interactions can be critical for efficient nucleotide incorporation[90, 134] but the interaction patterns are different among polymerases, so that their role in gaining selectivity has to be regarded individually for each enzyme.

1.2.4.3 Solvent effects

Besides hydrogen bonds and steric constraints the solvation of nucleobases plays a role in creating selectivity. In polar solvents the functional groups of DNA bases are hydrogen bonded to water or other solvent molecules which markedly increases their size. If hydrogen bonds to water molecules are broken upon formation of the base pair and no new hydrogen bonds are build, insertion is unfavourable. Moreover, the additional steric demand of nucleobases due to the attached waters might prevent insertion of nucleotides bearing small base moieties opposite nucleotides of the same kind[138].

1.2.4.4 Base stacking

Besides hydrogen bonding between base pairs, base stacking plays an important role in dNTP binding and especially in stabilizing the whole DNA duplex. The importance of base stacking becomes apparent when nucleotides are incorporated opposite an abasic site lesion or at the blunt end of a DNA duplex. In both cases there is no templating nucleobase

available which could direct the incorporation. The natural nucleobases show decreasing stacking ability with A > G > C,T[139, 140]. Π stacking also plays an important role in the insertion of hydrophobic nucleotide analogs[115] (see section 1.3).

1.2.4.5 Discrimination against ribonucleotides

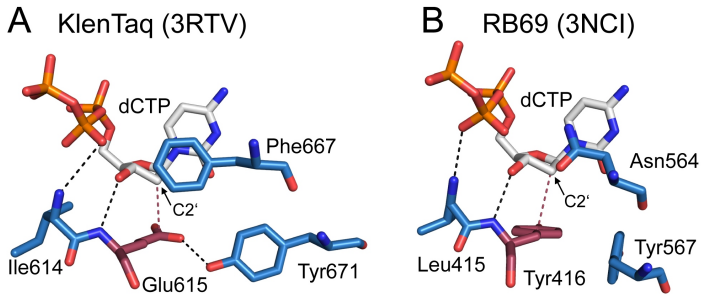


Figure 1.17: Structural basis for ribonucleotide exclusion in KlenTaq[44] and RB69[78]. Steric gate and other nearby active site residues are shown as raspberry and blue sticks, respectively. The substrate dCTP is shown in grey. Hydrogen bonding interactions are shown as black dashed lines. The shortest distance between the steric gate residue and the C2' atom is shown by raspberry dashed lines. Figure was created in analogy to Figure 1 in [141]

DNA polymerases need to actively discriminate against ribonucleotide incorporation as cellular concentrations of rNTPs are about 10-2000 times the concentration of dNTPs depending on the rNTP and dNTP pair and organism[142–144]. Indeed, DNA polymerases reach high discrimination rates. Brown *et al.* summarized sugar selectity of different polymerases and reported values ranging from 280 up to 4.440.000 fold discrimination[141]. This discrimination is accomplished by a bulky active site residue, called steric gate, present in most DNA polymerases. The side chains of these residues cause a clash with the 2' OH group of a ribonucleotide if it binds in the active site[145]. The steric gate residues characterized to date are Glutamate for A-family polymerases (Glu615 in KlenTaq, Glu658 in BF, Glu710 in KF[146, 147]) and Tyrosine or Phenylalanine in members of B, X, Y and RT families[141] (Tyr416 in RB69, Tyr409 in 9°N). Besides these bulky steric gate residues a protein backbone segment might be used to exclude rNTPs in some polymerases (e.g. polymerase β and λ). In addition to the steric gate residue also other nearby active site residues are involved in sugar discrimination. These residues arrange the incoming substrate and the steric gate residues properly to guarantee high selectivity. For example in KlenTaq the triphosphate moiety of an incoming substrate is contacted by the backbone NH group of Ile614 whereas its nucleobase engages stacking interactions

with Phe667 (**Figure 1.17 A**). In addition, Try671 forms a H-bond with the steric gate residue Glu615[141], further stabilizing the arrangement. Similar interaction networks are found in other DNA polymerases[141], e.g. in Family B (**Figure 1.17 B**).

1.3 Artificial Base Pairs

As DNA and RNA are built up by only 4 building blocks (A, G, C, T(U)) their coding potential but also other biotechnological and biomedical applications based on nucleic acids are limited. A third base pair which is readily and selectively incorporated into DNA by DNA polymerases and transcribed by RNA polymerases along with the natural base pairs would expand the genetic alphabet and enhance the great potential of DNA or RNA in many areas. An enlarged genetic code would e.g. provide a new method for site-specific incorporation of novel functionalities into nucleic acids and proteins[133]. An artificial base pair should fulfill mainly two prerequisites. First, the base pair building blocks need to specifically interact with each other and maintain DNA duplex stability. Furthermore, the pair should be fully orthogonal to the natural base pairs and needs to be efficiently and selectively inserted by DNA and RNA polymerases. Since the discovery of the molecular structure of DNA in 1953[6] researchers tried to deeply understand this structure and the corresponding function of DNA. Chemists started to manipulate the DNA structure by synthesizing modified building blocks and testing them for their ability to be inserted into a DNA strand and further functional properties. The nucleotides have been modified at all three moieties, the sugar, the phosphate and the nucleobase (reviewed in [148]).

Since the late 1980s researches aim at developing an artificial base pair (with natural sugar and phosphate moieties and modified nucleobase) which could efficiently complement the two Watson-Crick base pairs in DNA and RNA. Different strategies have been followed by either using similar structures with orthogonal hydrogen-bonding patterns compared to the natural bases or completely different structures which pair without hydrogen bonds but via hydrophobic interaction. The first artificial base pair was developed by Benner and his colleagues in 1989[149]. One of their motivations to prepare modified mononucleotide building blocks was to investigate the question if an early form of life relied on RNA catalysis rather than using enzymes (before proteins emerged). One of the reasons why proteins might have turned out to be better catalysts could be the fact that they are composed of a greater number of building blocks with a wider range of different chemi-

cal functionalities[150]. Therefore Benner and coworkers aimed to synthesize nucleotides bearing diverse catalytic groups. Soon, other groups joined, following the aim to design and continuously improve artificial base pairs. Thereby they relied on the different concepts: nonstandard hydrogen bonding patterns, shape complementarity, hydrophobicity, electrostatic repulsion and combinations thereof[151]. In the following, different starting points of different groups working on artificial base pairs will be shown and the way towards their latest, best replicated and transcribed candidate will be briefly described.

1.3.1 Hydrogen-bonding base pairs

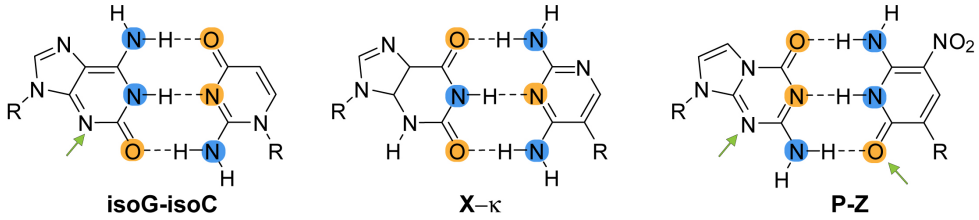


Figure 1.18: Hydrogen-bonded, unnatural base pairs developed in the Benner group. H-bonding donors and acceptors at the Watson-Crick edge of the nucleobases are colored blue and orange, respectively. H-bonding acceptors pointing into the minor groove are indicated by green arrows.

The Benner group first introduced the base pairs formed by isoguanine and isocytosine (**isoG-isoC**) or xanthosine and 2,6-diaminopyrimidine **X**-κ[149][150] (**Figure 1.18**). For the first time they could show that a base pair with an altered hydrogen bonding pattern can be inserted into an oligonucleotide by a DNA polymerase (KF) and an RNA polymerase (T7-RNAP)[149]. However, in case of **isoC-isoG** the incorporation turned out to be little selective. Problems are that **isoC** hydrolyses slowly to U and **isoG** can tautomerize to a form which is complementary to U[150]. The nucleotide κ was able to direct the incorporation of xanthosine into DNA and RNA products but still with moderate selectivity. In this work it was proposed that "*the additional letters in the nucleoside alphabet might eventually be used to expand the genetic code, increasing the number of amino acids that can be incorporated translationally into proteins*"[150]. Shortly after this proposal the non-natural amino acid 3-iodotyrosine was successfully incorporated into a polypeptide by translation of an mRNA containing the (**isoC**)AG codon[152]. To overcome the limitations of the above-mentioned base pairs new hydrogen-bonding base pairs have been developed. The base pair between 6-amino-5-nitro-2(1H)-pyridone (d**Z**) and 2-amino-8-imidazo[1,2-a]-1,3,5-triazin-4(8H)-one (d**P**) shows a pyDDA-puAAD hydrogen bonding

scheme[153] (**Figure 1.18**, right panel). In this nomenclature py stands for pyrimidine, pu for purine and D and A for hydrogen-bonding donor and acceptor, respectively. The electron withdrawing nitro group at the 5 position of d**Z** lowers epimerization of the C-glycoside[154]. Furthermore, the pair was designed to present electron lone pairs to the minor groove, hypothesized to improve specificity for polymerase recognition. The pair was shown to be incorporated by A- and B-family polymerases in PCR with fidelities per round of 94.4% for *Taq*, 97.5% for Vent (exo-) and 97.5% for Deep Vent (exo-)[155]. Still, the relatively low selectivity limits the applicability of these hydrogen-bonding artificial base pairs.

1.3.2 Base pairs combining hydrogen-bonding and shape complementarity

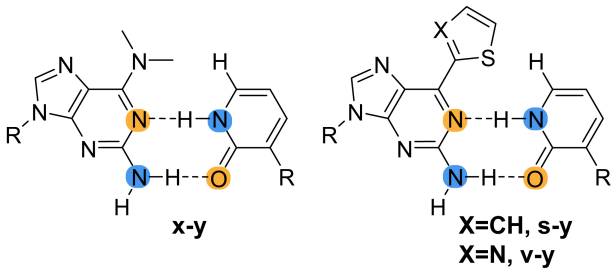


Figure 1.19: Unnatural base pairs relying on hydrogen bonds and shape complementarity developed in the Hirao group.

In designing artificial base pairs Ichiro Hirao and coworkers used shapes similar to natural bases and added steric groups to avoid mispairs with the natural counterparts[156]. In their first artificial pair 2-amino-6-(N,N-dimethyl-amino)purine (**x**) and pyridine-2-one (**y**) the team combined the concept of nonstandard hydrogen bonding pattern with shape complementarity[157]. To avoid mispairing with T the **x** bears a bulky dimethylamino group that would clash with the 4-keto group of T upon formation of the **x**-T mispair. Although the pair showed 90% selectivity in transcription by T7 RNA polymerase[158], the replication selectivity was poor and therefore the pair requested further optimization. By replacing the dimethylamino group with a thienyl or thiazolyl group in **s**[159] and **v**[160], respectively, an additional electrostatic repulsion prevents mispairing with T on the one hand, and, on the other hand DNA duplexes containing the optimized pairs gained in stability. Although the **s**-**y** and **v**-**y** pairs could be used successfully in applications such as the incorporation of unnatural amino acids into proteins[161] or the site

specific labelling of RNA aptamers[162], the pair was not sufficiently orthogonal to the natural pairs in replication. After some optimization steps the group switched from their hydrogen-bonding candidates to base pairs with hydrophobic and stacking interactions and developed their to date most promising base pair **Ds-Px** (see section 1.3.3).

1.3.3 Non-hydrogen-bonding base pairs

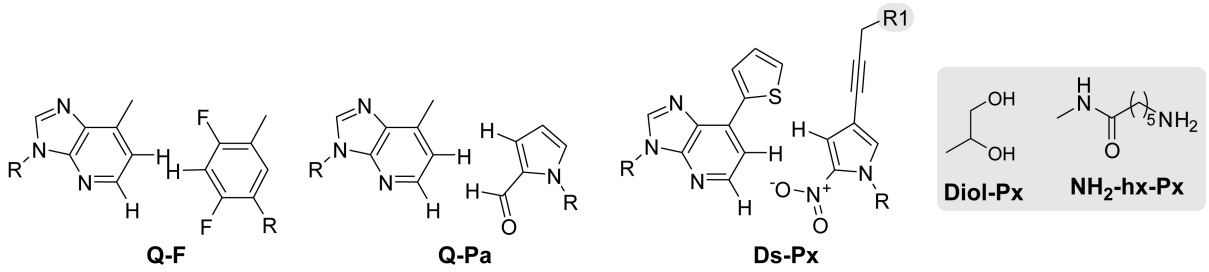


Figure 1.20: Unnatural base pairs relying on hydrophobic forces and shape complementarity developed in the Hirao group.

The development of artificial base pairs without hydrogen bonds began with the work of Kool *et al.* in the mid-1990's (see section 1.2.4.1. The results of their studies suggested that hydrogen bonds are not necessarily needed for efficient and selective replication of DNA. Inspired by these findings researchers started to develop base pairs without hydrogen bonds. Note that the hydrophobic analogs described in this work are not actually basic, but are termed "bases" or "nucleobases" by analogy to their natural counterparts. Based on the the hydrophobic A-T pair analogues **Z-F** (**Figure 1.14**) and **Q-F** (**Figure 1.20**), Hirao *et al.* improved their initially hydrogen bonding pairs (**Figure 1.19**) by switching towards a hydrophobic pair with enhanced stacking interactions. The group first developed an improved binding partner for Kool's **Q** nucleotide, the five-membered ring base analog pyrrole-2-carbaldehyde (**Pa**). An NMR structure of free duplex DNA containing **Q-Pa** indicated that the **Q-Pa** pair adopts a similar structure as the natural pairs[163]. However, this pair showed poor orthogonality to the natural pairs in replication. The best pair developed in the Hirao group to date consists of 7-(2-thienyl)imidazol[4,5-b]-pyridine (**Ds**) (derived from **s**, see **Figure1.19**) and 2-nitro-4-propynylpyrrole (**Px**, derived from **Pa**). Using Deep Vent DNA polymerase with 3'-5' exonuclease activity the pair of **Ds** and a dihydroxyethyl derivative of **Px**, namely Diol-**Px** showed a selectivity of 99.77-99.92% (depending on sequence context) per replication under optimized conditions and the misincorporation rate of $5x10^{-5}$ errors/bp approaches the intrinsic mispairing error

rate among the natural bases ($2x10^{-5}$ for Deep Vent DNA polymerase (exo+)[164]). Furthermore it could be shown that after 100 cycles of PCR (10 x 10 cycles using a d**Ds**TP and NH_2-hx-d**Px**TP) more than 97% of the unnatural **Ds-Px** pair was retained in the amplified DNA fragments[164]. As an additional advantage of the **Px** nucleotide functional groups can be easily atttached to the amino group of **Px** and thereby site-specifically introduced into DNA[151].

The Romesberg group at the Scripps Research Institute in California also works on searching for stable and replicable artificial base pairs. The group started with the investigation of predominantly hydrophobic self-pairs such as 7-propynyl isocarbostyril (**PICS**)[165] in 1999 (**Figure 1.21**). Although the **PICS** substrate could efficiently be incorporated opposite a templating **PICS** by KF the homo-pair could not be further extended. This can be explained by the assumption that the large hydrophobic **PICS** molecules stack on each other at the p/t end and are thereby not properly aligned for primer extension[166]. By derivatization of simple molecules like benzene and naphthalene the group defined their prerequisites for a good pair. Among them is the presence of a H-bond acceptor in *ortho* position for proper extension of the pair (e.g. a methoxy group)[167] which has already been shown by Kool *et al.* to be important[135]. By screening a library of 3600 base pairs towards efficient and accurate polymerase insertion[168] and further optimization of the hits the base pair d**MMO2**-d**5SICS** was found to be the most promising. An additional modification of d**MMO2**: a ring fusion at the meta and para position, yielded the naphthalene derivative d**NaM** which showed improved efficiency and selectivity in replication opposite d**5SICS**[168]. The d**NaM**-d**5SICS** pair is the artificial base pair investigated in this work and its incorporation properties will therefore be described in detail in the next section.

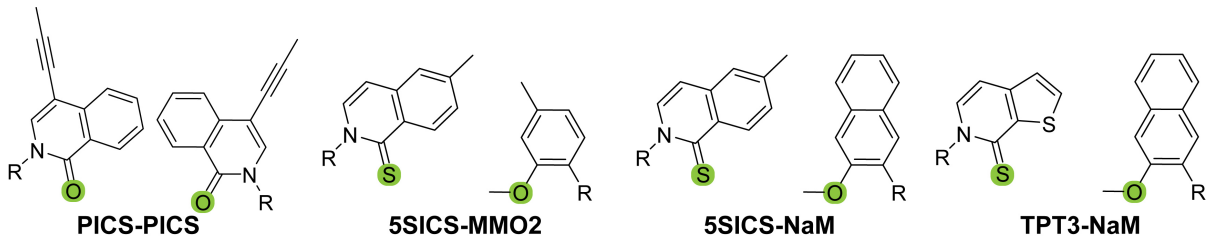


Figure 1.21: Unnatural base pairs relying on hydrophobic and packing forces developed in the Romesberg group. Hydrogen bonding acceptors pointing into the minor groove are shown in green.

Recently the Romesberg group managed to further optimize the d**NaM**-d**5SICS** pair

and published their to date most efficiently replicated pair d**NaM**-d**TPT3**[169] (**Figure 1.21**). DNA containing d**TPT3**-d**NaM** is amplified by OneTaq (New England Biolabs, mixture of exo- *Taq* and exo+ DeepVent polymerases) *"with an efficiency that is only 4-fold lower than that of DNA containing just the natural base pairs, and with a fidelity in excess of 99.98%"*[169]. The error rate of the reaction is 10^{-4} per nucleotide and overlaps with the 10^{-4} to 10^{-7} error rate of fully natural DNA with commonly used PCR systems[169].

1.3.3.1 The base pair dNaM-d5SICS

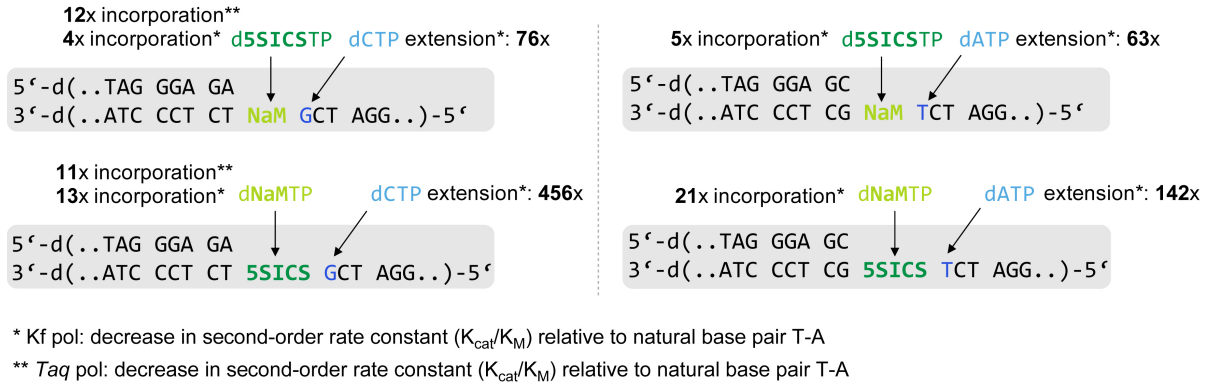


Figure 1.22: Summary of results from kinetik studies for d**NaM**-d**5SICS** incorporation and extension. Insertion kinetic results are taken from [170], elongation kinetic results are from [171].

The d**NaM**-d**5SICS** pair developed in the Romesberg group showed to be efficiently accepted by DNA and RNA polymerases in replication[170, 171] and transcription[172] processes *in vitro*. The Romesberg group determined kinetic data for the incorporation of d**NaM** and d**5SICS** with either KF or *Taq*[170] (**Figure 1.22**). They could show that with *Taq* the incorporation of d**NaM** opposite d**5SICS** is only 11 times less efficient than incorporation of dA opposite dT and d**5SICS** insertion opposite d**NaM** is similar (12 times impaired). In KF there is a difference between incorporation of d**5SICS** or d**NaM** opposite the respective partner, however the values are still in the same range. Whereas insertion of d**5SICS** is only 3.7 times less efficient compared to dA opposite dT, d**NaM** insertion is 13.3 times decreased. Thereby, the K_M in d**5SICS** insertion opposite d**NaM** is 6.4 times lower (0.039µM vs. 0.25µM) than the K_M in d**NaM** insertion opposite d**5SICS** whereas k_{cat} is 1.8 times lower. In *Taq* the K_M is similar for both nucleotides. With these and additional kinetics, measuring misincorporation opposite the artificial pair, Professor Romesberg and his group could show that the synthesis of d**NaM**-d**5SICS** by KF

and *Taq* is not only very efficient but also selective[170]. Moreover, the synthesis is not strongly dependent on strand or sequence context[170].
Elongation of the artificial pair is still the rate-limiting step in d**NaM**-d**5SICS** replication. Elongation kinetics with KF have been measured[171]. Depending on the base pair flanking the artificial pair on the 3' template side extension efficiencies range from 142–456 fold decrease compared to extension of a dA-dT pair with d**NaM** in the primer and d**5SICS** in the template and from 63–76 fold decrease with d**5SICS** in the primer and d**NaM** in the template (**Figure 1.22**).

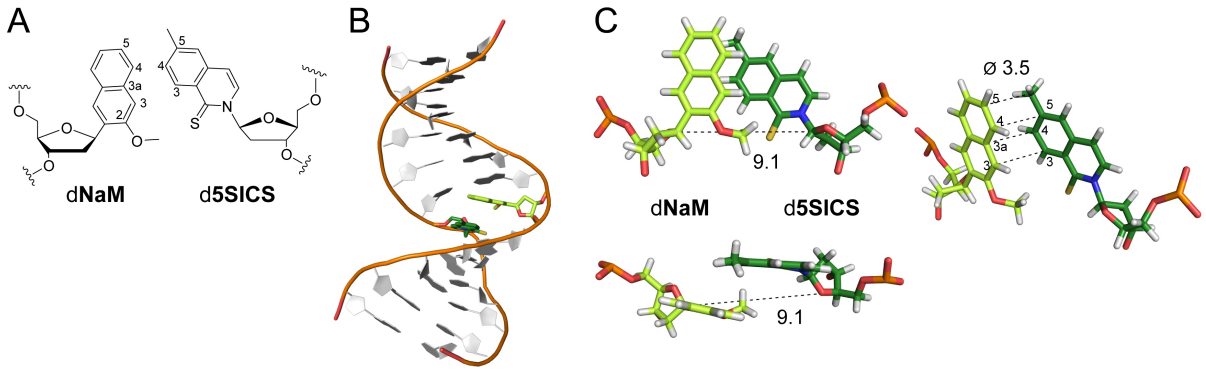


Figure 1.23: Structure of d**NaM**-d**5SICS** in free duplex DNA. **A**: Chemical structure of the d**NaM**-d**5SICS** pair. **B**: Model of double-stranded DNA containing the d**NaM**-d**5SICS** pair in the middle of the duplex. D**NaM** and d**5SICS** are shown in light and dark green sticks, respectively. The rest of the duplex is shown as cartoon. **C**: Intercalated pairing of d**NaM**-d**5SICS** from three views, with a C1'-C1' distance of 9.1Å. The nucleobase atoms positioned above each other are indicated by black dashed lines.

The d**NaM**-d**5SICS** base pair is mainly hydrophobic and pairs based on hydrophobic and packing forces. An NMR study of a similar pair (d**MMO2**-d**5SICS**) revealed that the nucleotides pair via partial interstrand intercalation in free duplex DNA[173]. Thereby d**5SICS** stacks well with its adjacent dT on the one side and partially intercalates to the opposite strand between d**MMO2** and its 5' dA, on the other. Correspondingly d**MMO2** packs with d**5SICS** from the opposite strand[173]. The base moieties of the d**5SICS**-d**MMO2** pair are positioned in an arrangement such that the C4 of d**5SICS** is positioned almost directly over (or under, depending on the point of view) the C3 of d**MMO2**[173]. The same partially intercalated structure is adopted by d**NaM**-d**5SICS**[44] (**Figure 1.23**). Here the two nucleobases run almost parallel and pair such that their edges are located above each other with the C3, C3a, C4 and C5 atoms of d**NaM** situated above the C3, C4, C5 atoms and the methyl group attached to C5 of d**5SICS**, respectively, with

an average distance of 3.5Å. The internucleotide distance between the C1' atoms of the 2'-deoxyribose moieties is 9.1Å.

1.3.4 Inorganic artificial base pair

In more recent years a new generation of artificial base pairs have been reported which exhibit a metal-mediated base pairing instead of hydrogen-bonding interactions[174–179]. Chelator nucleotides together with metals ions form planar complexes which can stack properly within a DNA duplex and thereby strongly enhance the thermal stability of the modified DNA relative to a duplex with normal H-bond interactions[180]. Carell *et al.* have developed an inorganic metal base pair which is bridged by a copper ion (dS-Cu-dS)[181] (**Figure 1.24 A**). They introduced a salen containing nucleobase that can form a self pair via a reversible crosslink with a ethylene diamine and chelation of a metal ion. This pair forms highly stable duplex structures via coordination forces instead of hydrogen bonds[181–184]. The group could show that the **dS-dS** self pair is fully orthogonal to the natural dG-dC and dT-dA pairs and can be replicated and amplified via PCR by DNA polymerases[185]. Three crystal structures of BF in complex with either p/t containing a **dS** at the templating position or a **dS**-Cu-**dS** and **dS-dS** at the n-5 position, respectively, have been solved (**Figure 1.24 B**). The structures reveal that the base pair is geometrically similar to a natural pair and does not perturb the structure of the enzyme or the DNA duplex. A limitation, however, is the speed of the amplification. The PCR reaction is much slower than with only canonical components because a lot of time is required to disassemble and reassemble the salen complex during replication[185]. For further information about the special applications of metal-mediated artificial base pairs see [186, 187].

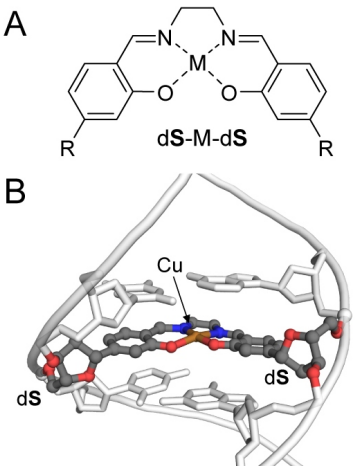


Figure 1.24: Inorganic unnatural base pair developed in the Carell lab.

1.4 Aim of this work

Non-hydrogen-bonding artificial base pairs are promising candidates on the way to develop a third base pair which is replicated and transcribed with same efficiencies and fidelities as the natural dA-dT and dG-dC pairs and therefore to expand the genetic code. Instead of hydrogen bonds these pairs form mainly via hydrophobic and packing forces. One promising base pair candidate to expand the genetic code is the d**NaM**-d**5SICS** pair developed in the Romesberg group. The pair is efficiently and selectively replicated and transcribed by DNA and RNA polymerases[170, 172]. To elucidate this efficient reaction the Dwyer group (in cooperation with the Romesberg group) determined a solution structure of a similar pair (d**NaM**-d**MMO2**) in free duplex DNA. In contrast to the planar edge to edge pairing of natural nucleobases the hydrophobic pair was shown to adopt an interstrand intercalating arrangement[173]. This arrangement was also confirmed for the d**NaM**-d**5SICS** pair[44]. The intercalating pairing of the promising base pair candidates raised several questions. First: how can DNA polymerases be so efficient in incorporating such a hydrophobic unnatural pair although it pairs via intercalation. Second: Does the pair also intercalate in the active site of a polymerase or does it adopt a different arrangement? Third: What happens after the insertion reaction, meaning: what is the basis for the lower elongation efficiency compared to the incorporation efficiency of the artificial pair? No other hydrophobic artificial base pair with shapes different from the natural pairs has yet been investigated structurally within the active site of a DNA polymerase from which conclusions could be made. Therefore, the aim of this study was to answer the raised questions and elucidate the incorporation and elongation mechanism of d**NaM**-d**5SICS** by DNA polymerases. The process should be investigated via X-ray crystallography using the well studied model enzyme KlenTaq DNA polymerase.
Investigation of the replication process was divided into two parts. In the first part of the work, the incorporation of the unnatural nucleotides opposite their respective partners by KlenTaq DNA polymerase should be studied. In the second part, it should be elucidated how the DNA strand containing the non-natural pair is elongated further by a natural nucleotide. The overall goal of the project thereby was to get structural insights into the incorporation and elongation mechanism of the unnatural base pair d**NaM**-d**5SICS** in particular, but also of hydrophobic artificial base pairs in general. Understanding the mechanism by which DNA polymerases replicate predominantly hydrophobic artificial base pairs is of great interest for the design of new base pairs and optimization of existing

ones. In addition, these studies can help understanding the DNA polymerases' acceptance of modified substrates. To realize that aim it was intended to crystallize KlenTaq DNA polymerase trapped in different states: prior to- and after incorporation of the artificial pair and in the course of elongating it with a natural substrate. The states along the reaction that can be trapped crystallographically are visualized in **Figure 1.25**. Using different p/t designs and approaches to trap the enzyme in the desired state it is aimed to obtain structures of KlenTaq (1) in binary and ternary pre-insertion complexes, (2) in a post-insertion complex with the d**NaM**-d**5SICS** pair positioned at the p/t junction in the post-insertion site and (3) in a pre-elongation step with the next natural substrate triphosphate bound. Finally, analyzing the crystal structures, a model for the replication of hydrophobic artificial base pairs should be developed.

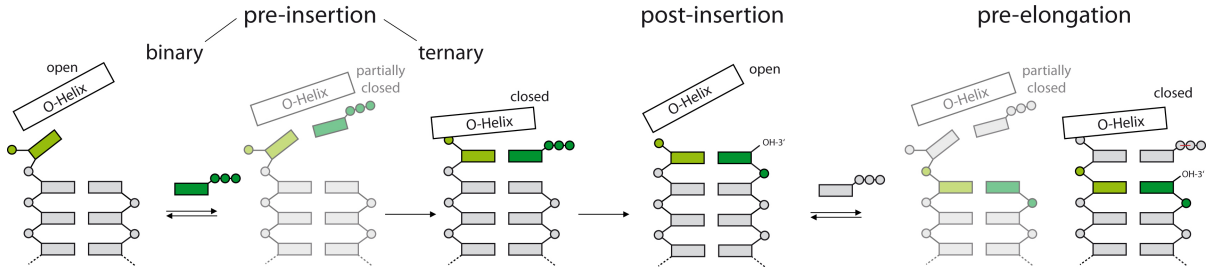


Figure 1.25: The DNA duplex near the active site of a DNA polymerase is shown schematically with the phosphate backbone indicated by circles and the sugar and base moieties by rectangles. Light and dark green color indicate hydrophobic artificial nucleotides, respectively. The O-helix is shown in its different states: open, partially closed and closed. Intermediate states in the reaction cycle (partially closed) are shown in light grey.

A second goal of this work was the crystallization of an archaeal B-family DNA polymerase in complex with DNA. Therefore the thermostable DNA polymerase of the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (short: 9°N) was used. For this enzyme only an apo structure has been solved to date and structural data of binary and/or ternary complexes of it and related archaeal B-family polymerases lack completely. Obtaining a binary and/or ternary complex would deliver structural knowledge about incorporation efficiency and fidelity of this specific enzyme and also archaeal B-family DNA polymerases in general. This is of special interest as thermostable B-family DNA polymerases together with A-family members are widely used in standard biochemical methods e.g. in PCR reactions with different purposes but also in biotechnological applications such as the selection of aptamers by SELEX (systematic evolution of ligands by exponential enrichment[188, 189]), genome sequencing[190], molecular diagnostics[191] (e.g. SNP detection) and DNA functionalization via post-synthetic conjugation[192]. While many

crystal structures of A-family DNA polymerases in binary and ternary complexes exist, only few B-family DNA polymerases have been crystallized in complex with DNA or with DNA and triphosphate. The archaeal B-family polymerases are especially useful for the incorporation of modified nucleotides. In several cases it has been shown that the insertion of sugar[50, 52] or nucleobase [7][14, 16, 53–56] modified nucleotides is accomplished more efficiently by archaeal B-family members than polymerases from the family A such as *Taq* or KlenTaq DNA polymerase. The reasons for this preferential acceptance of modified substrates has not yet been elucidated, mainly due to missing structural data. Hence, the aim of this study was to structurally investigate the DNA polymerases KOD from *Thermococcus kodakaraensis* and 9°N simultaneously. In the present work the results for the 9°N are described whereas the structure of KOD is described in the PhD thesis of my colleague Konrad Bergen and results of both studies are published together[193].

2 Results and Discussion part I: Structural snapshots of KlenTaq DNA polymerase processing the artificial base pair dNaM-d5SICS

The work described in this chapter was performed in collaboration with Prof. Romesberg and his group, in particular Denis Malyshev, from the Scripps Research Insitute in California. Parts of the results are published in [44, 194].

2.1 Incorporation of dNaM and d5SICS

As described in chapter 1.3, "Artificial Base Pairs", the development of a third base pair which can be efficiently replicated by DNA polymerases is of great interest and stimulates several groups of researchers in continuously improving their unnatural base pair candidates towards enhanced orthogonality to the natural pairs and trim them for the use in different applications. Although kinetic data indicate that hydrophobic artificial base pairs can be inserted with high efficiency by DNA polymerases and are also moderately well elongated[164, 170, 171], it is not clear how DNA polymerases handle these structures on a molecular level. NMR studies revealed that in free duplex DNA the artificial base pair d**NaM**-d**5SICS** pairs in an intercalating manner[44, 173]. Such a structure seems difficult to fit properly into the active site of a DNA polymerase, especially in case of high fidelity DNA polymerases which characteristically have tight binding pockets. To explore this contradictory observations we worked together with the Romesberg goup at the Scripps Research Institue in California and investigated the incorporation and elongation process for their base pair d**NaM**-d**5SICS**. This unnatural pair is a promising representative for hydrophobic base pairs with shapes that are very different from the natural purines and pyrimidines and which pair without hydrogen bonding but via hy-

drophobic and packing forces.
In a first step the insertion reaction of d**NaM** and d**5SICS** opposite their respective partners was investigated. Therefore, crystal structures of binary DNA polymerase complexes with a p/t duplex carrying each of the unnatural nucleotides in the templating position and the corresponding ternary complexes with the pairing triphosphate bound in the catalytically active state should be solved. In our studies the well characterized KlenTaq DNA polymerase (see section 1.2.2.1) was used and trapped in the desired state utilizing a 2'-3'-dideoxynucleotide at the 3'-end of the primer. This strategy to trap the enzyme prior to the insertion reaction was first used by Pelletier *et al.* to obtain a ternary strcuture of DNA polymerase β in 1994[65]. It became a standard method afterwards and was successfully used to crystallize many different DNA polymerases in complexes with substrates. Templates bearing d**NaM** and d**5SICS** at the directing position were synthesized by Denis Malyshev from the Romesberg group and annealed to an unmodified primer. In analogy to previous crystallization studies in our and other groups[71, 87, 195] we used a 16mer template and a 11mer primer. The nucleic acids were annealed and subsequently the p/t duplex was mixed with the purified KlenTaq DNA polymerase and ddCTP was added in the presence of 20mM $MgCl_2$ to passivate the primer strand for further elongation (**Figure 2.1 A**). After incorporation of ddCMP a 12mer DNA duplex is formed which exhibits a 5'-d(AAA**NaM**)-3', a 5'-d(AAA**5SICS**)-3' or a 5'-d(AAC**5SICS**)-3' single-stranded (ss) template overhang, respectively (see **Figure 2.1**). The KlenTaq gene was expressed and the protein was purified as described in the methods, section 6.2.2. For the ternary complexes the corresponding artificial triphosphate was added in excess and should bind in the insertion site (**Figure 2.1 B**). The artificial triphosphates have also been synthesized by Denis Malyshev from the Romesberg group. Crystallization setups were prepared as described in the methods section 6.3.2.

2.1.1 Data collection, structure solution and refinement

A detailed description of the structure solution and refinement process for all structures presented in this work is given in the methods section. Briefly, diffraction images were recorded at the Swiss light source (SLS) at the Paul Scherrer Institute in Villigen, Switzerland and the data was subsequently processed using XDS[196]. Structures were either solved by direct refinement against previously solved structures with the same space group

and cell dimensions, or by molecular replacement (MR). Refinement was performed using Phenix[197] and model building was done in Coot[198]. For refinement, data up to a resolution at which the $CC_{1/2}$ value[199] is still significant was used. For structures obtained before the introduction of the $CC_{1/2}$ value in 2012 by Prof. Diederichs and Prof. Karplus, data was used to a resolution at which I/σ reached a value around 1. In the past it has been a common rule in protein crystallography to set the resolution cutoff of a crystallographic dataset at the point where the I/σ value drops below 2. Therefore, to facilitate comparison with other deposited structures, the resolution which corresponds to an I/σ value of around 2 (named $I/\sigma\sim$ 2 resolution) is specified for all structures reported in this work.

2.1.2 Binary pre-insertion complexes

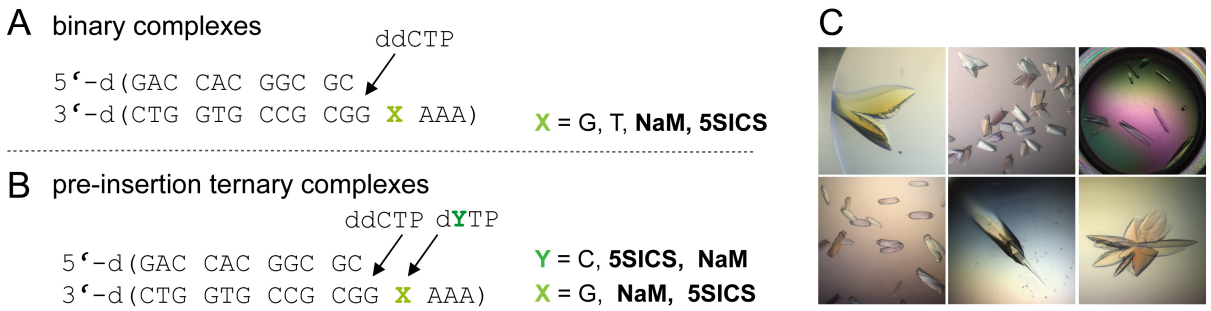


Figure 2.1: P/t design used to crystallize KlenTaq in binary or ternary pre-insertion complexes. **A:** P/t design for KlenTaq binary complexes with templating dG, dT, d**NaM** or d**5SICS**, respectively. **B:** P/t design for KlenTaq ternary complexes prior to incorporation of d**NaM**MP, d**5SICS**MP or dCMP, respectively. For the unnatural templating nucleotides additionally templates with a 3'-CAA-5' template overhang were used. **C:** Example images of crystals obtained from these setups.

First, the structures of the binary pre-insertion complexes of KlenTaq bound to a p/t with d**NaM** and d**5SICS** at the templating position were solved and termed $KTQ_{d\mathbf{NaM}}$ and $KTQ_{d\mathbf{5SICS}}$, respectively. For comparison with the natural situation, structures of KlenTaq in the binary state with a p/t containing dG or dT at the templating position but otherwise similar primer and template sequences (varying only in the +1 template nucleotide compared to the described $KTQ_{d\mathbf{5SICS}}$ structure) were solved and termed KTQ_{dG} and KTQ_{dT}, respectively. All crystals of the binary complexes were in the same space group: $P3_121$ (no. 152), with similar cell dimensions. Structures were solved by difference Fourier techniques after rigid body refinement of a previously determined KlenTaq structure (PDB ID: 3M8S)[87]. The statistics of data collection and refinement are given

in **Table 2.1** for the structures including artificial nucleotides and **Table 2.2** for the structures with natural DNA. The binary structures $KTQ_{d\mathbf{NaM}}$, $KTQ_{d\mathbf{5SICS}}$, KTQ_{dG} and KTQ_{dT} have I/σ=2 resolutions of 1.95Å, 2.9Å, 2.15Å and 2.0Å, respectively. From the binary structures only $KTQ_{d\mathbf{NaM}}$ was modelled completely. In the other three structures several residues of the finger domain, especially residues of the loop between the N- and O-helix were not modelled due to disorder. Missing residues are 645-654 in KTQ_{dG}, 647-653 in KTQ_{dT} and 648-658 (including part of the O-helix) in $KTQ_{d\mathbf{5SICS}}$.

Overall structure. The overall structures of $KTQ_{d\mathbf{NaM}}$ and $KTQ_{d\mathbf{5SICS}}$ are similar to those of the binary complexes KTQ_{dG} and KTQ_{dT}. Thereby, the structure with templating d**NaM** is more similar to KTQ_{dT} (rmsd: 0.352Å vs. 0.827Å to KTQ_{dG}) and the structure with the templating d**5SICS** is more similar to KTQ_{dG} (rmsd: 0.648Å vs. 0.814Å to KTQ_{dT}). Furthermore, the structures $KTQ_{d\mathbf{NaM}}$ and $KTQ_{d\mathbf{5SICS}}$ are similar to the previously solved binary complex of KlenTaq from the Waksman group[71] (PDB ID: 4KTQ; 0.964Å and 0.545Å, **Figure 2.2 A**). The natural complexes KTQ_{dG} and KTQ_{dT} align to 4KTQ with a rmsd of 0.650Å and 0.814Å, respectively. In all structures the enzyme is in an open state, characterized by an open finger domain. In $KTQ_{d\mathbf{NaM}}$ and KTQ_{dT}, however, the finger domain helices N, O, O1 and O2 and connecting loops adopt different positions (**Figure 2.2** zoom-in view) due to a special arrangement of the 5' ss template end (described below). Investigating the B-factors of the enzyme revealed, that the finger domain (especially helices N, O, O1 and O2 including residues 639-700) is the most flexible domain in all structures (**Figure 7.2**, Appendix page 181). Due to the pairwise similarity in the overall structures of KlenTaq either containing natural or artificial templating nucleotides it can be concluded that d**NaM** and d**5SICS** do not influence the overall enzyme or duplex DNA structure in a binary complex.

Templating nucleotide and ss template arrangement. In $KTQ_{d\mathbf{5SICS}}$ and KTQ_{dG} the templating nucleotides are rotated away from the insertion site and point towards the solvent. The upstream 5' ss template nucleotides could not be modelled due to disorder in both structures. Also in the structure 4KTQ published by Li *et al.* the templating nucleotide adopts an extra-helical position, however, the nucleobase is oriented differently (**Figure 2.3 A**). As has been shown before by Li *et al.*[71] the template base of the nascent base pair (position n-1) in KTQ_{dG} and $KTQ_{d\mathbf{5SICS}}$ is covered by the C-terminal O-helix residue Tyr671 (**Figure 2.4 A** and **B**, lower right panel) which is highly conserved in

	KTQ_{dNaM}	KTQ_{d5SICS}
PDB ID	**3SYZ**	**4CCH**
Data collection		
Date of data collection	2011/02/03	2013/04/09
Wavelength (Å)	1.0015	1.0000
Space group	$P312_1$	$P312_1$
Cell dimensions		
a, b, c (Å)	108.5, 108.5, 90.5	114.3, 114.3, 91.60
α, β, γ (°C)	90, 90, 120	90, 90, 120
Resolution (Å)	47.0-1.95 (2.07-1.95)*	49.5-2.55 (2.70-2.55)
No. of total reflections	491656 (72624)	205511 (25923)
No. of unique reflections	45007 (7124)	22906 (3650)
R_{meas} (%)⋆	13.6 (102.4)	20.4 (144.2)
I/σ	11.23 (2.21)	9.16 (0.65)
Completeness (%)	99.8 (98.6)	99.9 (99.7)
Redundancy	10.9 (10.2)	9.0 (7.1)
$CC_{1/2}$ (%)[199]	-	99.2 (52.3)
Refinement		
Resolution (Å)	46.6-1.95	43.5-2.55
No. of reflections	86484	43559
R_{work}/R_{free}	16.0/20.0	20.7/26.1
Coordinate error (Å)[a]	0.47	0.42
No. of atoms		
Protein	8711	8458
DNA/Triphosphate	583/-	795/-
Water	314	116
Average B-factors (Å²)		
Protein	42.0	66.8
DNA/Triphosphate	42.1/-	52.0/-
Water	42.2	38.3
R.m.s deviations		
Bond lengths (Å)	0.006	0.005
Bond angles (°)	1.077	0.770
Ramachandran (%)[b]		
Favored	95.99	95.04
Allowed	2.44	4.58
Outlier	0.36	0.38

Table 2.1: Summary of the data collection and refinement statistics for KlenTaq binary and ternary pre-insertion complexes containing the artificial nucleotides d**NaM** and d**5SICS**. Data reduction and refinement data is stored in the following folders:
KTQ_{dNaM}: /home/kbetz/SLS2011/feb03/NaM-d5-3SYZ
KTQ_{d5SICS}: /home/kbetz/SLS2013/apr09/13-5SICS-bin-8/phenix/4CCH
$KTQ_{dNaM\text{-}d5SICSTP}$: /home/kbetz/SLS2010/dec12/nam-tp-d24
$KTQ_{d5SICS\text{-}dNaMTP}$: /home/kbetz/SLS2012/oct22/NaMTP-8/phenix
* Values in parentheses correspond to those in the outer resolution shell.
⋆ For definition of R_{meas} see [200]
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

	$KTQ_{dNaM\text{-}d5SICSTP}$	$KTQ_{d5SICS\text{-}dNaMTP}$
PDB ID	**3SV3**	**4C8K**
Data collection		
Date of data collection	2010/12/12	2012/08/22
Wavelength (Å)	0.9999	0.99987
Space group	$P312_1$	$P312_1$
Cell dimensions		
a, b, c (Å)	108.5, 108.5, 91.0	115.0, 115.0, 91.0
α, β, γ (°C)	90, 90, 120	90, 90, 120
Resolution (Å)	47.0-2.10 (2.23-2.10)*	48.6-2.16 (2.29-2.16)
No. of total reflections	364815 (56777)	227942 (14630)
No. of unique reflections	36558 (5822)	34375 (3094)
R_{meas} (%)*	15.7 (136.3)	9.5 (217.4)
I/σ	12.41 (2.33)	14.44 (0.47)
Completeness (%)	99.9 (99.5)	91.1 (51.2)
Redundancy	10.0 (9.8)	6.6 (4.7)
$CC_{1/2}$ (%)[199]	-	99.8 (47.8)
Refinement		
Resolution (Å)	47.0-2.10	48.6-2.17
No. of reflections	69989	33357
R_{work}/R_{free}	16.7/21.5	18.1/22.1
Coordinate error (Å)[a]	0.55	0.25
No. of atoms		
Protein	8692	8650
DNA/Triphosphate	511/64	826/49
Water	232	138
Average B-factors (Å^2)		
Protein	48.8	56.6
DNA/Triphosphate	41.3/48.7	50.3/77.4
Water	48.7	45.6
R.m.s deviations		
Bond lengths (Å)	0.008	0.007
Bond angles (°)	1.144	0.870
Ramachandran (%)[b]		
Favored	97.05	97.77
Allowed	2.77	2.04
Outlier	0.18	0.19

Table 2.1: continued

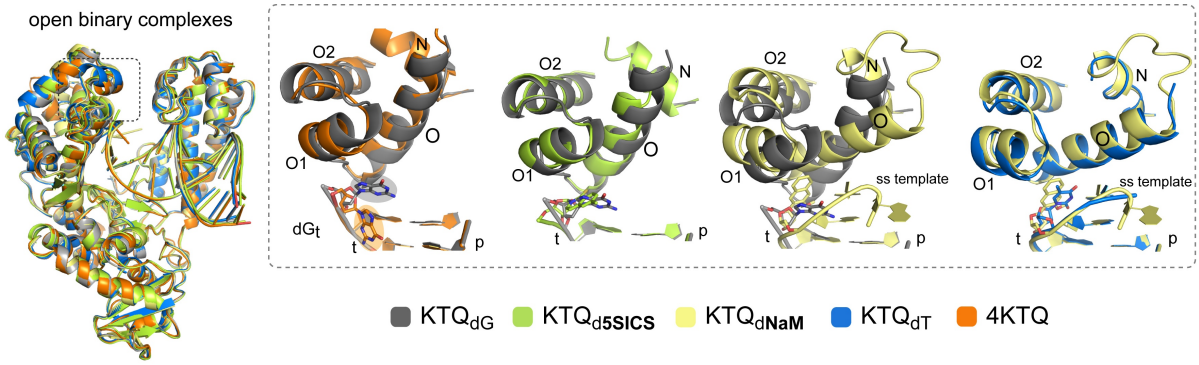


Figure 2.2: Open binary complexes of KlenTaq with templating nucleotides d**NaM**, d**5SICS**, dG or dT. Overlay of $KTQ_{d\mathbf{NaM}}$, $KTQ_{d\mathbf{5SICS}}$, KTQ_{dG}, KTQ_{dT} and 4KTQ. Zoom into a part of the finger domain (visualizing helix N, O, O1 and O2) shows pairwise superimpositions.

family A DNA polymerases[83]. It has been suggested that Tyr671 may act as a positioning device for the DNA when it binds to the enzyme, such that the first base pair at the p/t junction can stack against the aromatic amino acid side chain and finds its proper position in the active site[71]. Furthermore, Tyr671 plays a role in maintaining selectivity in the polymerase reaction (shown for the related KF polymerase[202]). Due to the different arrangements observed in KTQ_{dG} and 4KTQ despite identical 5' ss template sequences, it is assumed that the ss portion of the template is flexible in the binary structures of KlenTaq, and that the differences observed are not functionally relevant. Flexibility of the templating nucleobases is also indicated by high B-factors and poorly defined electron density maps (especially in case of d**5SICS**) as compared to the double-stranded part of the DNA. In contrast to the situation in KlenTaq the templating nucleotide in open BF structures is positioned in a defined pre-insertion site and is not as flexible as in KlenTaq (**Figure 2.3 A** vs. **B**).

In $KTQ_{d\mathbf{NaM}}$ and KTQ_{dT} the templating nucleotides as well as the residual ss template behave different than in $KTQ_{d\mathbf{5SICS}}$ and KTQ_{dG}. Similarly, the templating nucleotides are flipped away from the developing DNA duplex, however, to a different position and the ss template rotates back towards the duplex (**Figure 2.4 C** and **D**). Here, the two upstream nucleotides, dA_{t+1} and dA_{t+2}, stack between the exposed nascent base pair in the post-insertion site and Phe667 of the finger domain O-helix (**Figure 2.4 C** and **D** lower right panel). The hydrophobic nucleobase of d**NaM** and the nucleobase of dT are positioned in a "pocket" comprised of O-helix residues Met673, Tyr671, Phe667 and Thr664 and the template nucleotides dA_{t+1} and dA_{t+2} (**Figure 2.4 C** and **D**, upper right

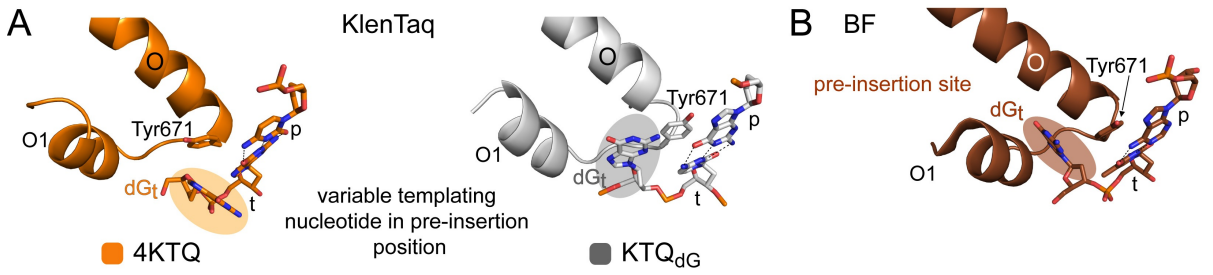


Figure 2.3: Position of the templating nucleotide in KlenTaq and BF. **A**: Different position of the templating nucleotide in 4KTQ and KTQ_{dG}. The position of dG_t is visualized by a colored oval. **B**: Templating nucleotide bound in the pre-insertion site formed by O- and O1-helix in a binary BF structure (PDB ID: 1L3U[203]).

panel). In this arrangement the templating nucleotides as well as dA_{t+1} and dA_{t+2} are well ordered.

The back-rotation of the template seems to be dependent on the templating nucleotide as we observe it for a templating dT and d**NaM** but not for dG or d**5SICS**. It is assumed that the nucleobases of dG or d**5SICS** due to their shape might not allow a similar stabilization of the nucleobase by the O-helix residues in the shallow "pocket" as observed for dT or d**NaM**. This in turn could lead to a situation that renders the backward rotation of the template towards the nascent base pair unfavorable and instead the template points towards the solvent. The backward rotation allows the the template dA nucleotides to form stacking interactions with Phe667 and the nascent base pair. This arrangement might be especially fostered with 5' ss dA residues as adenine shows superior base stacking ability over the other three nucleobases[139, 140]. Changing the template overhang from 5'-d(AAA**NaM**)-3' to 5'-d(AAC**NaM**)-3' leads to a structure in which d**NaM** plus the upstream dC_{t+1} are flipped out of the stacking arrangement and the two consecutive dA residues (dA_{t+2} and dA_{t+3}) stack on the nascent base pair (data not shown). This observation implies that a ss overhang consisting of only dC or dT might not adopt the same stacking arrangement as observed for dA due to their lower stacking abilities. However, further experiments are required to verify this assumption. With the templating d**5SICS** no difference in the structure was observed with the ss template overhang being either 5'-d(AAC)-3' (reported in this work) or 5'-d(AAA)-3' (data not shown). In both complexes the 5' ss template was not sufficiently ordered to be modelled.

Overall, the structures of the binary complexes with a templating dT, dG, d**NaM** or d**5SICS** are all very similar except from the discussed ss template arrangement which is

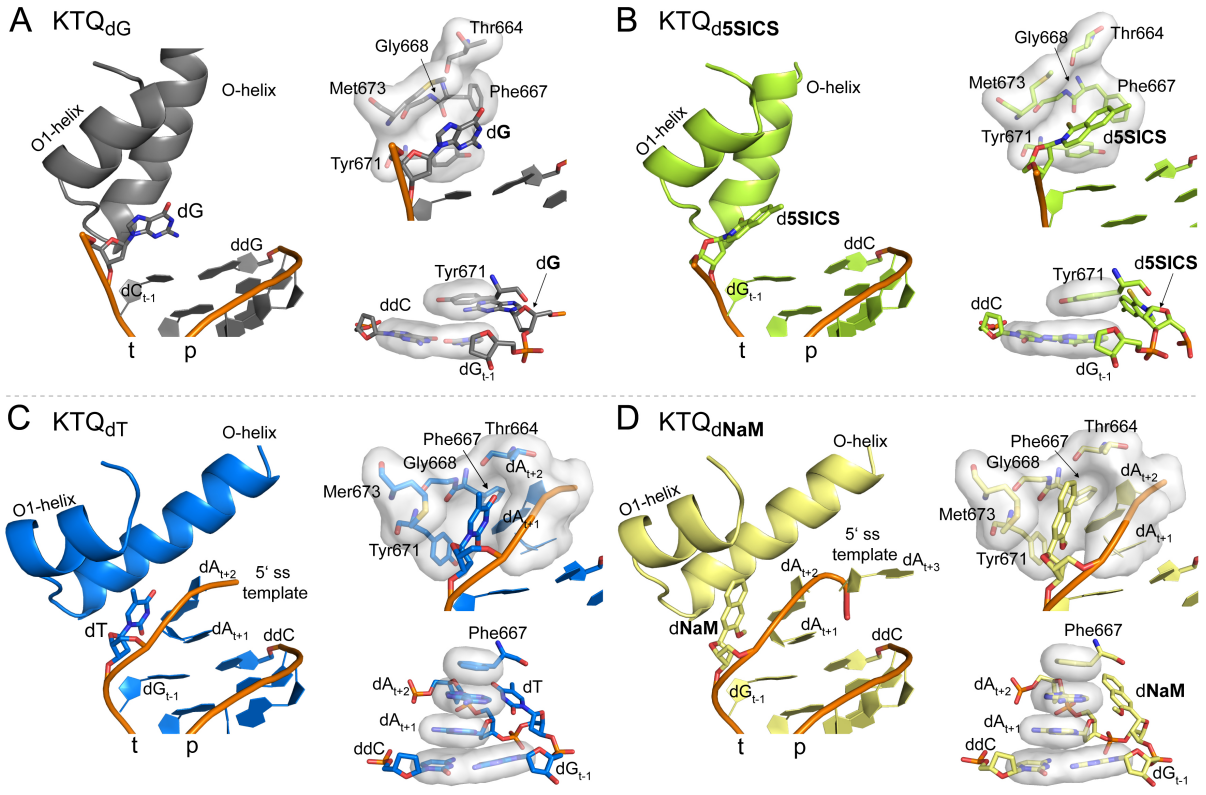


Figure 2.4: **A:** KTQ_{dG} (gray), **B:** KTQ_{d5SICS} (limon), **C:** KTQ_{dT} (blue) and **D:** KTQ_{dNaM} (paleyellow) are shown in different arrangements. At the left side of each panel the O- and O1-helices and the p/t end are shown as cartoon and the templating nucleotide is displayed in stick representation. Residues and their surfaces that accommodate the templating nucleobase (in **C** and **D**) are shown in the upper right corner. The same residues are also shown in **A** and **B** for comparison. The stacking arrangement of the nascent base pair with either Tyr671 (KTQ_{d5SICS} and KTQ_{dG}) or downstream nucleobases of dA_{t+1}, dA_{t+2} and Phe667 (KTQ_{dT} and KTQ_{dNaM}) is shown in the bottom right corner. To better visualize the stacking, the surfaces of the nucleobase moieties of the n-1 pair, dA_{t+1}, dA_{t+2} and the tyrosine or phenylalanine side chains are shown.

different in KTQ_{dT} and KTQ_{dNaM} compared to KTQ_{dG} and KTQ_{d5SICS}. The backward rotation of the template is most likely not relevant in solution. Especially for a longer template overhang this arrangement seems unfavoured as large movements of the template in an open-to-closed transition of the enzyme would be necessary. In all binary complexes the finger domain is the most flexible domain suggesting that it is relatively dynamic in the open conformation. Finally, the unnatural nucleotides d**NaM** and d**5SICS** don't seem to perturb the structure of the enzyme in the open state. In a reverse conclusion the flexibility of the enzyme in the open state with no special constraints acting on the templating nucleotide makes it tolerating the unnatural nucleotides at that position without any changes in its own structure. From that point of view one might expect that the KlenTaq

DNA polymerase and related enzymes are able to well tolerate templating nucleotides carrying even larger nucleobase moieties in an open state.

	KTQ_{dG}	KTQ_{dT}	$KTQ_{dG\text{-}dCTP}$
PDB ID	**3SZ2**	**3SV4**	**3RTV**
Data collection			
Date of data collection	2011/03/27	2011/05/12	2010/10/15
Wavelength (Å)	1.2300	0.9000	1.0000
Space group	$P312_1$	$P312_1$	$P312_1$
Cell dimensions			
a, b, c (Å)	108.4, 108.4, 89.8	108.7, 108.7, 90.2	108.3, 108.3, 90.4
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	46.9-2.15 (2.28-2.15)*	47.1-2.00 (2.12-2.00)	46.9-2.15 (2.28-2.15)
No. of total reflections	371812 (58733)	420008 (64393)	487017 (77362)
No. of unique reflections	33563 (5352)	42985 (6685)	49071 (7765)
R_{meas} (%)*	16.6 (142.4)	11.7 (102.0)	8.9 (95.0)
I/σ	14.77 (1.90)	13.79 (2.06)	17.27 (2.33)
Completeness (%)	100.0 (99.9)	99.9 (99.6)	99.7 (98.5)
Redundancy	11.1 (11.0)	10.0 (9.6)	9.9 (10.0)
Refinement			
Resolution (Å)	46.9-2.15	46.9-1.99	46.9-1.90
No. of reflections	64185	80646	93428
R_{work}/R_{free}	19.0/23.7	17.9/21.6	16.2/20.2
Coordinate error (Å)[a]	0.64	0.56	0.48
No. of atoms			
Protein	8474	8626	4374
DNA/Triphosphate	510/-	562/-	570/28
Water	290	303	331
Average B-factors (Å^2)			
Protein	44.9	44.0	35.1
DNA	40.2	41.9	39.2
Water	39.2	43.5	40.1
R.m.s deviations			
Bond lengths (Å)	0.003	0.003	0.007
Bond angles (°)	0.758	0.770	1.102
Ramachandran (%)[b]			
Favored	94.35	95.74	97.46
Allowed	5.27	3.89	2.36
Outlier	0.38	0.37	0.18

Table 2.2: Summary of the data collection and refinement statistics for KlenTaq binary and ternary pre-insertion complexes with natural substrates. Data processing and refinement results are stored in the following folders:
KTQ_{dG}: /home/kbetz/SLS2011/mar27/tempg-e5
KTQ_{dT}: /home/kbetz/SLS2011/may12/wt-t-nitro6
$KTQ_{dG\text{-}dCTP}$: /home/kbetz/SLS2010/oct15/TempG-o11
* Values in parentheses correspond to those in the outer resolution shell.
* For definition of R_{meas} see [200]
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

2.1.3 Ternary pre-insertion complexes

After investigating binary complexes of KlenTaq with templates carrying d**NaM** or d**5SICS** the next step was to solve and analyze ternary complexes of KlenTaq with d**NaM**TP and d**5SICS**TP bound in the active site. Therefore, cocrystalliztaion experiments as shown in **Figure 2.1 B** were made. For both strand contexts (X=d**NaM** and X=d**5SICS**) cocrystallization with the respective triphosphates (Y=d**5SICS**TP and Y=d**NaM**TP) failed in several attempts. Alternatively soaking of the available binary complex crystals was tried and the two structures $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ and $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ could be solved. For more detailed comparison with the enzyme's conformation in complex with the "standard substrates" also the fully natural complex with a templating dG and a pairing dCTP in the active site was solved and termed $KTQ_{dG\text{-}dCTP}$. Although a ternary complex of KlenTaq with natural substrate (structure with ddC at the primer end and ddCTP in the active site, PDB ID: 3KTQ, solved in 1998) already exists[71] using a structure determined under similar conditions for comparison was preferred, especially due to the reason that methods of data collection, data reduction and refinement continuously improve. The two natural ternary complexes $KTQ_{dG\text{-}dCTP}$ and 3KTQ are very similar with a rmsd of 0.298Å. All three ternary structures solved were in the same space group and similar cell dimensions as the binary structures reported in the last section. The statistics of data collection and refinement for $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ and $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ are given in **Table 2.1** and for $KTQ_{dG\text{-}dCTP}$ statistics are given in **Table 2.2**. $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$, $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ and $KTQ_{dG\text{-}dCTP}$ have I/σ∼ 2 resolutions of 2.1Å, 2.3Å and 1.9Å, respectively.

2.1.3.1 Structure of $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$

Overall structure. The soaking experiments with d**5SICS**TP and crystals of the binary complex of KlenTaq with d**NaM** in the templating position resulted in a closed and active complex with the d**5SICS**TP bound and stabilized in the insertion site. The addition of the triphosphate not only induced the transition from the open to the closed state of the polymerase by closing of the finger domain but also led to a dramatic movement in the ss portion of the template, flipping d**NaM** back into the insertion site where the two unnatural nucleotides pair (**Figure 2.5**). The d**5SICS**TP is buried between the terminal primer nucleotide (ddC) and the O-helix of KlenTaq.

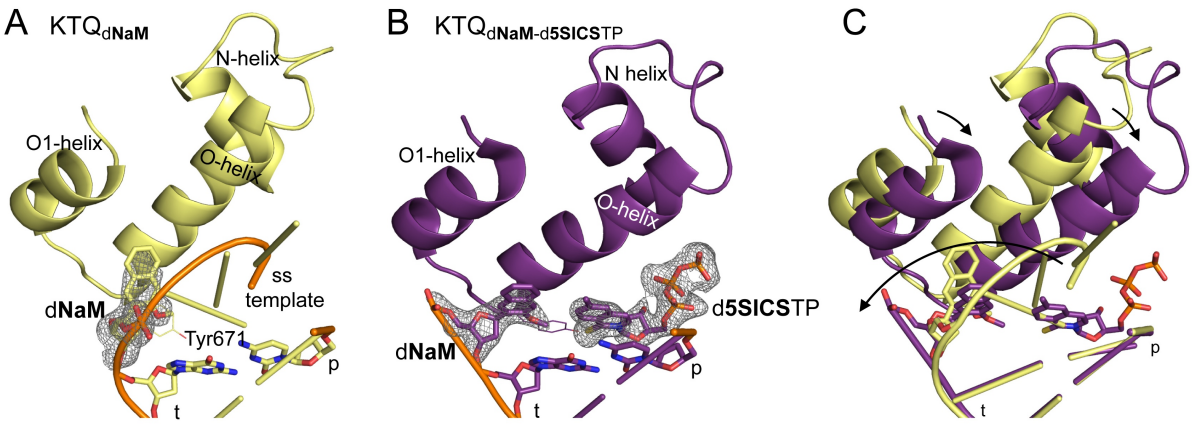


Figure 2.5: Conformational change of the ss template and the finger domain in the transition from KTQ_{dNaM} (**A**) to $KTQ_{dNaM\text{-}d5SICSTP}$ (**B**) upon d**5SICS**TP binding. In **C** an overlay of both structures is shown. O-, O1- and N-helices of the finger domain and the p/t in the active site are shown as cartoon. The n-1 base pair as well as the templating nucleotide and the incoming triphosphate are shown as sticks. D**NaM** and d**5SICS**TP are surrounded by simulated annealing omit maps contoured at 3σ. Arrows in **C** indicate the movements.

The overall structure of $KTQ_{dNaM\text{-}d5SICSTP}$ is similar to the one of $KTQ_{dG\text{-}dCTP}$ (rmsd: 0.434Å, **Figure 2.6 A**). The finger domains of both ternary structures superpose well but the O1- and O-helix are slightly shifted away from the newly formed base pair in the insertion site. Comparing the B-factors of $KTQ_{dNaM\text{-}d5SICSTP}$ and $KTQ_{dG\text{-}dCTP}$ relatively larger B-factors within the finger domain are observed for the structure harbouring the artificial pair, suggesting that the finger domain is slightly more flexible here (**Figure 7.3**, Appendix page 182). This observation may explain the slightly lower insertion efficiency of d**5SICS**MP compared to natural nucleotides[170], as in a more flexible closed complex components involved in the chemical reaction are probably not as rigidly arranged. The DNA duplex near the insertion site adopts an A-form, with a C3'-endo conformation of the 2'-deoxyribose moieties. This A-form DNA near the active site was described before in other A-family DNA polymerase structures[71, 129]. The A-form is characterized by a wide and shallow minor groove which allows easy access of the protein side chains to the edges of the nucleobases. Nucleic acid properties of the DNA duplex in $KTQ_{dNaM\text{-}d5SICSTP}$, were determined using Curves+[204]. The sugar pucker of the p/t complex is shown in the Appendix, **Figure 7.1** on page 181.

Insertion site. The insertion site accommodating the newly formed base pair is confined on one side by the O- and O1-helix residues and on the other by the base pair in the post-insertion site. The nucleobases of the artificial base pair have a different shape

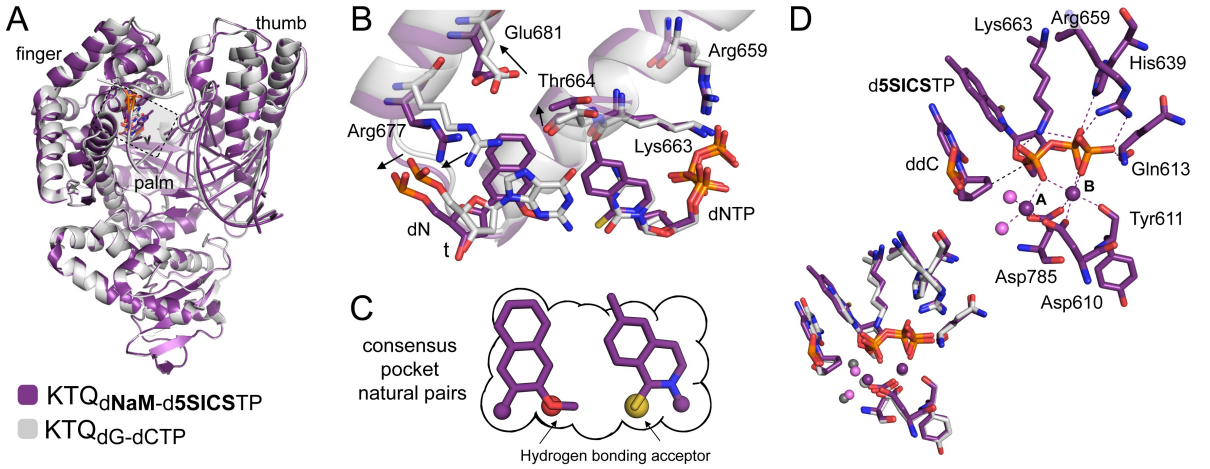


Figure 2.6: **A**: Superposition of the overall ternary structures of KlenTaq containing dG-dCTP (gray) or d**NaM**-d**5SICS**TP (purple). The finger domain O-helices adopt similar positions in both structures (**A**) but the active site is slightly enlarged (**B**). The relevant residues are shown as sticks and shifts are indicated by arrows. **C**: The consensus pocket of the four natural nucleobases is shown as black line. The d**NaM**-d**5SICS**TP pair of $KTQ_{dNaM\text{-}d5SICSTP}$ is placed such that it is properly oriented on the right edge. At the positions where the pair differs from the natural shape the enzyme adapts by small shifts in surrounding side chains, shown in panel **B**. **D**: Residues of the active site as well as d**5SICS**TP and the primer terminal ddC are shown as sticks. Coordination of the catalytic Mg^{2+} ions and stabilization of the substrate triphosphate is indicated by dashed lines. Active site residues are located similar in $KTQ_{dNaM\text{-}d5SICSTP}$ and $KTQ_{dG\text{-}dCTP}$ as visualized in the overlay at the lower left edge of the panel. Water molecules and magnesium ions are shown as light pink and purple spheres, respectively, for $KTQ_{dNaM\text{-}d5SICSTP}$, and as dark gray and light gray spheres, respectively, for $KTQ_{dG\text{-}dCTP}$.

than the natural nucleobases and hence the pair they form shows a different structure. The differences in the d**NaM**-d**5SICS**TP pair compared to the natural pairs is visualized in an overlay of the unnatural pair with the consensus pocket of the four natural pairs defined by Kool *et al.*[116] (section 1.2.4.1; **Figure 2.6 C**). Due to the different shape of the nucleobases of the artificial base pair compared to the natural ones, the insertion site in the $KTQ_{dNaM\text{-}d5SICSTP}$ structure is slightly enlarged relative to $KTQ_{dG\text{-}dCTP}$. This enlargement is indicated by a small shift of the O- and O1-helix on the one hand and by a somewhat different position of the templating nucleotide on the other (**Figure 2.6 B**). The enlarged nucleobase parts point towards the developing major groove leaving the minor groove site unperturbed. More specifically, the side chain of the residue Arg677 which interacts with the phosphate group of the templating nucleotide would clash with the d**NaM** nucleobase on the major groove side and is thereby shifted away from insertion site (**Figure 2.6 B**). Along with it the phosphate moiety of d**NaM** as well as the whole O1-helix is shifted compared to their position in $KTQ_{dG\text{-}dCTP}$. Thereby the interaction of Arg677 and the templating d**NaM** is maintained. Within the O-helix mainly Thr664

shifts appreciably to make space for the methyl substituent of d**5SICS** whereas the other side chains keep their position. Overall, movements of O- and O1-helix side chains adjusted to the shape of the d**NaM**-d**5SICS** base pair were observed. This means that protein side chains shift where necessary, leading to an enlarged insertion site, whereas the remaining proximate side chains are hardly affected.

The ortho substituents of unnatural nucleobases have been proven to be essential for their replication by structure-activity relationship data[168]. It was shown that efficient synthesis requires hydrophobic substituents for both, the primer and template analogue pointing into the developing minor groove whereas extension is benefited with a H-bond acceptor in the primer nucleobase and a hydrophobic substituent in the template nucleobase[168]. Our structure reveals that the ortho substituents of both unnatural nucleobases are oriented into the developing minor groove in a fashion analogous to that of the hydrogen-bond acceptors of the natural nucleobases (N3 of purines and O2 and pyrimidines)[205]. The d**5SICS**TP sulfur atom can participate in a water-mediated hydrogen-bonding network with Glu615, Gln754 and Asn750 wereas the d**NaM** methoxy group packs with the sulfur atom of d**5SICS**TP, the backbone of Tyr667 and Gly668 on one side and the guanine of the upstream template nucleotide (dG_{t-1}) on the other (**Figure 2.7**). The overlay of the nucleobases of d**NaM**-d**5SICS**TP with the consensus pocket of natural pairs in **Figure 2.6 C** visualizes the similar position of the H-bonding acceptors pointing to the minor groove. The arrows in the figure indicate the positions of the H-bonding donors in the natural nucleotides and the red and yellow spheres show the H-bonding acceptors of the artificial nucleotides, respectively.

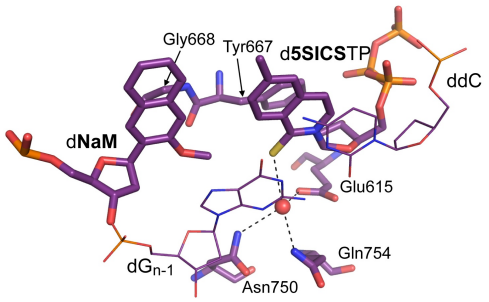


Figure 2.7: Positioning of the ortho substituents of d**NaM**-d**5SICS** in $KTQ_{dNaM\text{-}d5SICSTP}$

Catalytic site. Interestingly, apart from the slight movement of the mentioned O- and O1-helix residues the active site of $KTQ_{dNaM\text{-}d5SICSTP}$ and $KTQ_{dG\text{-}dCTP}$ looks very similar. In the catalytic site the triphosphate moiety engages in electrostatic interactions with the basic side chains of His639, Arg659 and Lys663 and the amide backbone of Gln613, comparable to the natural complex (**Figure 2.6 C**). Moreover, just like in the pair dG-

dCTP the ribose moieties of d**NaM** and d**5SICS**TP adopt the C3'-endo conformation. Together with the side chains of Asp785, Asp610, the backbone of Tyr611 and two water molecules, the phosphate groups of the incoming triphosphate coordinate the two catalytically essential magnesium ions (A and B). The presence of the two coordinated metal ions indicates that indeed a complex prior to the chemical reaction was trapped. The virtually identical distance between the C3' of d**5SICS**TP and the α phosphate in $KTQ_{\text{dNaM-d5SICSTP}}$ (3.8Å) and that observed for the natural triphosphate in $KTQ_{\text{dG-dCTP}}$ (3.9Å) indicates that the artificial pair is properly positioned for an imminent insertion reaction.

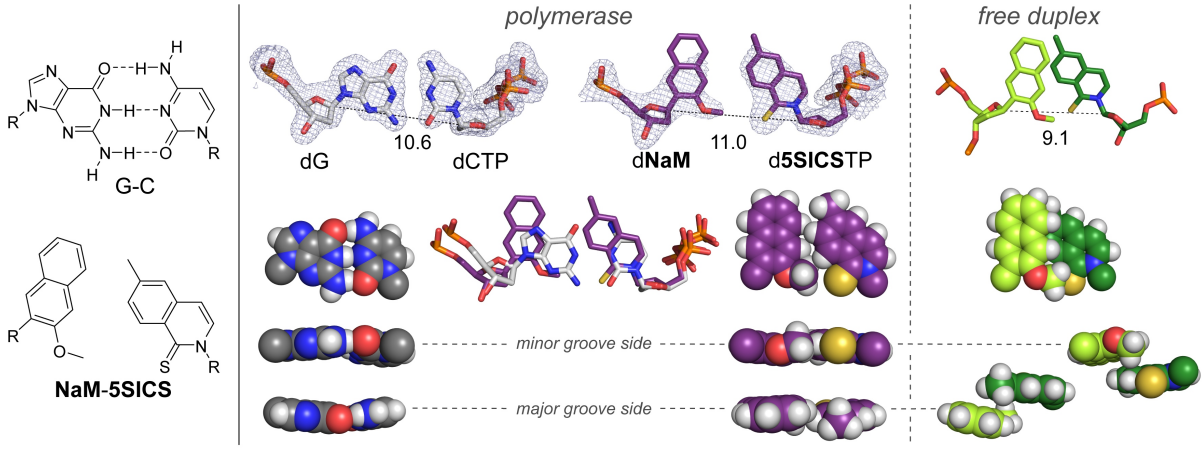


Figure 2.8: Structures of the natural dG-dCTP base pair (gray) and the unnatural d**NaM**-d**5SICS**TP pair (violet) in the active site of KlenTaq and of d**NaM**-d**5SICS** in free duplex DNA (green). Chemical structures are shown in the left panel. The base pairs in the active site of KlenTaq are shown in the middle, surrounded by a simulated annealing omit map contoured at 3σ. Below, an overlay of both pairs in stick representation is shown and space-filling models with views from the top, the minor- and the major groove side are visualized. The same arrangement for d**NaM**-d**5SICS** in free duplex DNA[44] is shown in the right panel. C1'-C1' distances are given in Å.

Base pair geometry. The most interesting finding in the structure of $KTQ_{\text{dNaM-d5SICSTP}}$ was the arrangement of the two pairing nucleotides with respect to each other. In great contrast to the intercalated structure d**NaM**-d**5SICS** forms in free duplex DNA (see section 1.3.3.1 and **Figure 2.8**, right panel) the nucleobases of the d**NaM**-d**5SICS**TP pair adopt a planar structure within the active site of KlenTaq (**Figure 2.8**, middle). The nucleobases align with a nearly optimal edge-to-edge packing just as a natural pair. The average distance between the edges of the hydrophobic bases is 4.2Å and the C1'-C1' internucleotide distance is in the same range as that of a natural base pair (11.0Å versus

10.6Å, respectively, compared to 9.1Å for d**NaM**-d**5SICS** in the free DNA duplex[173]). In an overlay of dG-dCTP and d**NaM**-d**5SICS**TP the ribose moieties of d**5SICS**TP and dCTP superimpose perfectly, whereby the complete d**NaM** nucleotide is shifted slighly away due to the larger nucleobase, as was mentioned before. Thus, although unable to form hydrogen bonds and its preference to maximize packing interactions by intercalation, the d**NaM**-d**5SICS**TP pair adopts a Watson-Crick like geometry in the active site of KlenTaq. The reduced intercalation might in part develop due to the A-form structure of the p/t terminus in the polymerase active site, which is wider than the B-form structure of the free duplex. However, this circumstance alone can hardly explain the behaviour of the pair. As the triphosphate is not covalently bound greater intercalation with the templating nucleotide should be possible if forces coming from the enzyme are not considered. Therefore, as a counterpart to the forces which lead to d**NaM**-d**5SICS** intercalation in free duplex, impacts from surrounding protein residues and nucleotides have to be taken into account in the active site of the enzyme. Within the closed active site of DNA polymerases usually many forces act which are responsible to test the geometry of a base pair and properly orient all residues involved in catalysis. The fact that in the closed enzyme state a planar pair instead of intercalation is observed for d**NaM**-d**5SICS**TP suggests that the sum of the interactions between the developing artificial base pair and the polymerase active site is stronger than forces leading to intercalation of the pair in a free duplex. As a result the d**NaM**-d**5SICS**TP pair adopts a planar geometry.

2.1.3.2 Structure of KTQ$_{\text{d5SICS-dNaMTP}}$

Motivated by the results obtained for d**5SICS**TP paired opposite d**NaM** in a pre-insertion complex of KlenTaq the next step was to also investigate the opposite strand context with d**5SICS** in the template and d**NaM**TP as incoming nucleotide. Therefore, binary KlenTaq crystals with d**5SICS** in the templating position were soaked by adding d**NaM**TP in excess and the structure KTQ$_{\text{dNaM-d5SICSTP}}$ was solved. Although soaking with d**NaM**TP has led to the formation of a ternary complex, it exhibits some major differences from the one observed in KTQ$_{\text{dNaM-d5SICSTP}}$. In contrast to a closed catalytic complex we obtained a complex in which the finger domain is only partially closed and the d**NaM**TP is bound to the finger domain instead of the insertion site. Thus, the structure found for KTQ$_{\text{d5SICS-dNaMTP}}$ can be regarded as a partially closed complex which is

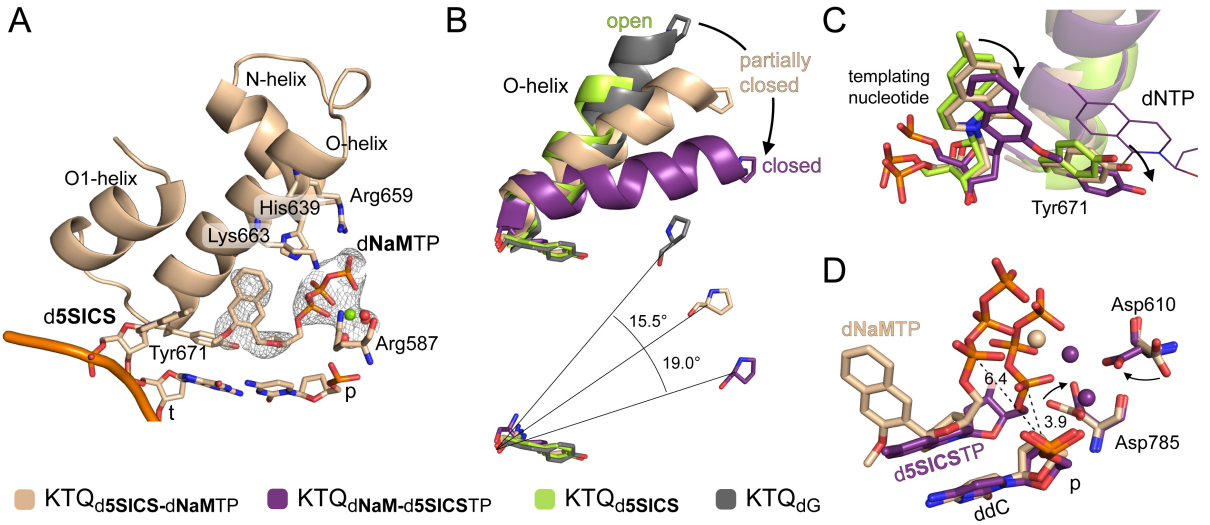


Figure 2.9: **A**: The pre-catalytic ternary complex KTQ$_{\mathbf{d5SICS}\text{-dNaMTP}}$ is shown in the same arrangement as structures in Figure 2.5. **B-D**: Superposition of KTQ$_{\mathbf{d5SICS}\text{-dNaMTP}}$ (sand), KTQ$_{\mathbf{d5SICS}}$ (limon), KTQ$_{\mathrm{dG}}$ (gray) and KTQ$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}\text{TP}}$ (purple). The differences in the O-helix in the open, partially closed and closed state of the enzyme (**B**), the templating nucleotides and Tyr671 (**C**) and the active site (**D**) are visualized. Movements from an open to a closed enzyme state are indicated by arrows and the distances between C1' and the α phosphates in KTQ$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}\text{TP}}$ and KTQ$_{\mathbf{d5SICS}\text{-dNaMTP}}$ are shown by black dashed lines.

similar to previously described "ajar" or half-open complexes in presence of a mismatched base pair (in BF) or an abasic site (in KlenTaq)[90, 206].

In KTQ$_{\mathbf{d5SICS}\text{-dNaMTP}}$ the d**NaM**TP is bound near the O-helix, stabilized by ionic interactions with O-helix residues Arg659 and Lys663 as well as His639 of the N-helix and Arg587 from the N-terminal end of the K-helix in the thumb domain (**Figure 2.9 A**). The triphosphate moiety, together with three water molecules, coordinates a single magnesium ion. However, the waters are not well resolved and might adopt slightly different positions within the molecules forming the crystal. The distance between the ribose C3' and the α phosphate atom indicates that a pre-catalytic complex was trapped. In the complex with d**NaM**TP bound near the active site the distance is 6.4Å whereas in the fully closed active complexes KTQ$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}\text{TP}}$ or KTQ$_{\mathrm{dG}\text{-dCTP}}$ it is significantly shorter: 3.8Å or 3.9Å, respectively (**Figure 2.9 D**). Moreover, the catalytic aspartate residues are not yet positioned properly to coordinate two metal ions as it would be necessary to guarantee an efficient insertion reaction.

The N- and O-helix of the finger domain adopt a conformation intermediate between the open and closed states with slightly more similarity to the open state as indicated by the rmsd values from superposition of the backbone atoms (N, Cα, C + O) of residues

637-699 in PyMOL. The rmsd of residues 637-699 is 0.800Å and 1.687Å relative to the open $KTQ_{d\mathbf{5SICS}}$ and the closed $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ complex, respectively. A comparison between the open, partially closed and closed state is visible in **Figure 2.9 B-D**. The rotation angle between the open, half-open and closed state of the O-helix (measured from the mean Tyr671 backbone oxygen to the Cα atom of Pro656) is approximately 15.5 and 19.0°. The O-helix of KTQ_{dG} was used for comparison as $KTQ_{d\mathbf{5SICS}}$ lacks several C-termianl O-helix residues. Although the O-helix is shifted towards the insertion site, Tyr671 still stacks on the nascent base pair. But, along with the O-helix rotation and the slight movement of the templating d**5SICS** from its extrahelical position towards the insertion site, Tyr671 is also slightly displaced (**Figure 2.9 C**). As the sugar and base moieties of the triphosphate are not specifically stabilized by direct interactions in the partially closed state it can be assumed that in a pre-catalytic complex first interactions of the enzyme with the substrate are solely made via the triphosphate moiety in case of a hydrophobic artificial nucleotide whereas in case of natural nucleotides it has been shown that H-bonds with the templating nucleotide can already form in the "ajar" state[90].

2.1.4 Discussion and Conclusion

In summary we have solved binary and ternary complexes with both d**NaM** in the templating position and its partner d**5SICS**TP bound in the active site ($KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$) and with d**5SICS** in the templating position and a d**NaM**TP substrate bound near the active site in a pre-catalytic state ($KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$). For comparison we also solved natural binary complexes with templating dG and dT nucleotides (KTQ_{dG} and KTQ_{dT}) as well as the ternary complex with a templating dG and dCTP bound in the insertion site ($KTQ_{dG\text{-}dCTP}$)

2.1.4.1 Unnatural base pair replication by inducing a Watson-Crick geometry

For $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ we obtained an active closed ternary complex with all components (apart from the missing 3' OH group of the 2'-3' dideoxy primer terminus) arranged for effective catalysis just as it is found in the fully natural complex. Furthermore, the soaking experiment showed that d**5SICS**TP addition induces full closure of the finger domain and the d**NaM**-d**5SICS**TP pair adopts a planar structure in the polymerase insertion site, comparable to a natural pair. This result was surprising as in free duplex

DNA the hydrophobic nucleotides adopt an intercalating structure. However, the fact that the hydrophobic artificial pair is able to behave just as a Watson-Crick base pair could well explain its efficient incorporation kinetics[170].
Therefore, our structural data provide an explanation for the empirical observation that complementary H-bond formation is not required for the efficient and selective replication of DNA[125, 165, 171, 207] even when the nucleobase shape is different compared to the natural substrates. Previously it has been shown for isosteric nucleotide analogs that H-bonds are not necessary for efficient replication as long as the shape of the formed base pair is within the constraints of the enzyme's active site[116]. The pairing of d**NaM**-d**5SICS** in the active site of KlenTaq is able to induce the open to closed transition of the finger domain and resultant the nucleotides adopt a planar structure. This shows that the pair has enough plasticity to adopt its preferred intercalated structure in free duplex DNA but can also form a planar pair if distinct forces are applied. Correspondingly, the data reveal that the polymerase active site is not only capable of selecting for a correct relative arrangement among the pairing nucleotides but also, at least in some cases, capable of enforcing it. Furthermore, it is apparent that the insertion site of KlenTaq is sufficiently flexible to accommodate slightly enlarged nucleobases if the additional space required extends towards the developing major groove. Moreover, the efficient replication of d**NaM**-d**5SICS** by a variety of other polymerases, including polymerases from different families[170, 208], suggests that these observations with KlenTaq can be applied to other polymerases as well. Polymerases might have evolved to favor a coplanar geometry over intercalation to prevent natural nucleotide mispairing via cross-strand intercalation and instead allow only the more specific, edge-to-edge hydrogen-bonding interactions. Finally, the data of our insertion study reveal that structural mimicry of a natural base pair is not required for unnatural base pair design, but base pairs relying on hydrophobic and packing forces are well suited to expand the genetic alphabet. The main reason for that is the fact that hydrophobic and packing forces are strong on the one hand but relative plastic on the other, enabling the pairs to adopt different structures depending on their environment.

2.1.4.2 Pre-catalytic complex for insertion of dNaM

Unlike with the addition of d**5SICS**TP to KTQ_{dNaM} crystals the addition of d**NaM**TP to crystals of KTQ_{d5SICS} did not fully induce the open-to-closed conformational change of the finger domain. Rather, soaking of binary crystals with d**NaM**TP resulted in the

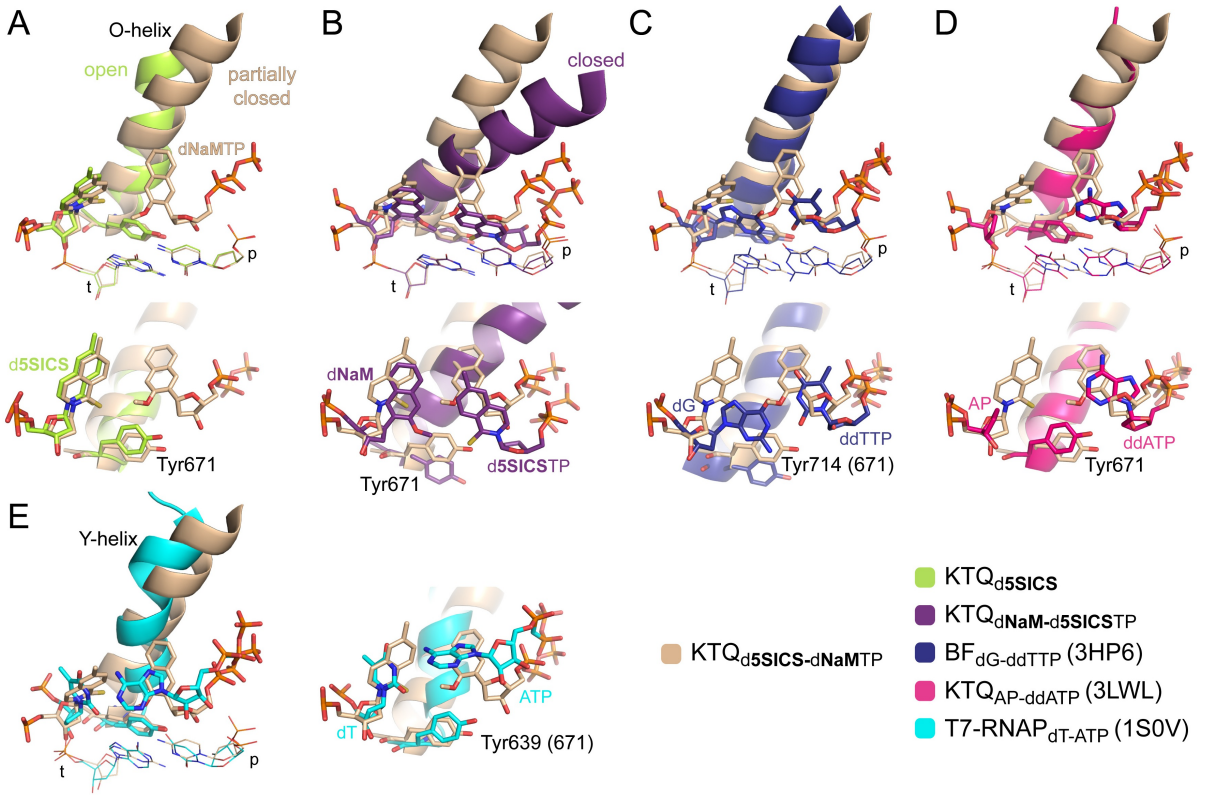


Figure 2.10: Comparison of KTQ$_{\text{d5SICS-dNaMTP}}$ with open, closed, and partially closed polymerase structures. The templating nucleotide, bound triphosphate, and Tyr671 are shown as sticks, the O-helix is shown as cartoon, and the base pair in the post-insertion site is shown as lines. **A**: Superposition of KTQ$_{\text{d5SICS-dNaMTP}}$ with KTQ$_{\text{d5SICS}}$. Tyr671 stacks on top of the last base pair in both structures. **B**: Superposition of KTQ$_{\text{d5SICS-dNaMTP}}$ with KTQ$_{\text{dNaM-d5SICSTP}}$ (PDB ID: 3SV3). In the closed structure the templating nucleotide is moved inwards and displaces Tyr671. **C**: Superposition of KTQ$_{\text{d5SICS-dNaMTP}}$ with the mismatch dG-ddTTP complex of BF (PDB ID: 3HP6)[90]. Both complexes are partially closed but in BF$_{\text{dG-ddTTP}}$ Tyr671 is displaced by the templating dG. **D**: Superposition of KTQ$_{\text{d5SICS-dNaMTP}}$ with KlenTaq in complex with an AP site in the template and ddATP bound at the O-helix (PDB ID: 3LWL)[209]. Both complexes are partially closed and Tyr671 is located in the templating position. **E**: Superposition of KTQ$_{\text{d5SICS-dNaMTP}}$ with a T7 RNA polymerase DNA/RNA complex in a partially closed state (PDB ID: 1S0V)[210]

formation of a structure wherein the DNA polymerase finger domain closes only partially and the d**NaM**TP is bound via its triphosphate moiety to the O-helix in a pre-catalytic state (**Figure 2.10**, compared to open (**A**) and closed states (**B**)). The difference in the ternary complexes KTQ$_{\text{dNaM-d5SICSTP}}$ and KTQ$_{\text{d5SICS-dNaMTP}}$ may indicate lower incorporation efficiency of d**NaM** opposite d**5SICS** compared to d**5SICS** opposite d**NaM** in KlenTaq, as it is observed in KF (see section 1.3.3.1).

2.1.4.3 Pre-catalytic DNA polymerase complexes and their potential role

Similar pre-catalytic complexes as the one found in the $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ structure have been described before (**Figure 2.10**). Beese and Wu trapped the BF polymerase in an "ajar" state in complex with a mismatched base pair (dG-ddTTP) in the inesertion site[90] (**Figure 2.10 C**). In this partially closed structure the residue Tyr714 (Tyr671 in KlenTaq) is displaced from its position in the insertion site and the templating nucleotide adopts its position. The polymerase is in the half-open state and the triphohsphate substrate is bound by O-helix residues. It has been suggested that this ajar conformation is adopted to probe complementarity of incoming substrates and templating nucleotides before the enzyme is triggered to close and adopts the catalytically competent state[90]. In our structure the location of the triphosphate is very similar. However, compared to the partially open ternary complex described by Beese *et al.*, in our case the templating d**5SICS** is still in its pre-insertion position and Tyr671 is situated at the templating position, as in the open binary structure, although slightly displaced. Thus, our structure appears in an intermediate state between the open binary complex and the ajar state observed with BF.

Additional structures in half-open ternary complexes described are those of KlenTaq with substrates bound opposite an abasic site analog in the templating position[206, 209]. Members of family A and B polymerases preferentially incorporate dAMP opposite an abasic site. The mechanism underlying this so-called "A-rule" was studied by Obeid *et al.* in the Marx group. Partially open ternary structures were reported with ddATP, ddGTP (two different states) and ddTTP bound to the O-helix in the finger domain. In all structures Tyr671 mimics the templating base and is situated in the insertion site and the O-helix is in a half-open conformation similar to the one observed in this work for $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ (**Figure 2.10 D**). Also in T7 RNA polymerase (T7 RNAP) which is related to A-family DNA polymerases (11.5% sequence identity with *Taq*), rNTP binding in a half-open conformation was observed (PDB ID: 1S0V, **Figure 2.10 E**). In this structure ATP is bound to a partially-closed helix Y (analog to O-helix) without forming a distinct Watson-Crick base pair with the templating nucleotide but in an orientation which is stated to be appropriate to achieve base-specific recognition upon subtle movements of the templating base and/or the substrate[210]. In the complex the insertion site is blocked by a conserved tyrosine (Tyr639 in T7 RNAP). A mechanism was suggested, in which the substrate selection occurs prior to the transition into the catalytically active

form in RNA polymerases, similar as has been shown later for DNA polymerases[90]. Apart from these structural results a partially closed or ajar state has also been observed for the homologous *E. coli* DNA polymerase I via biophysical studies[211, 212]. It has been shown that the ajar state is especially favored in case of an incorrect nucleotide substrate bound to the enzyme. This intermediate state within the three-state mechanism for nucleotide substrate selection, was proposed to be a fidelity checkpoint for nucleotides before the enzyme adopts the closed state[212]. Thereby, natural templating nucleotides may already establish H-bonds with the incoming nucleotide as is shown in the structures of BF[90] and discrimination against mispaired nucleotides may occur. In case of a correct nucleotide the ajar state is not resolved in FRET studies as the ternary complex seamlessly progresses to the long-lived closed ternary complex[212].
Consequently it can be concluded for our pair that d**NaM**TP opposite d**5SICS** rather behaves like a mispair than like a canonical pair, as in $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ the half-open state was observed in contrast to a closed state. Our data further reveal that the ajar state observed in the FRET studies seems not only to be favored in case of natural mispairs but also in case of difficult to incorporate substrates or special template situations as it was observed for an abasic site template[206, 209] and here, for the studied hydrophobic artificial pair in the strand context which shows lower incorporation efficiency in KF (d**5SICS** in the template and d**NaM**TP as substrate). In contrast, if d**5SICS**TP is positioned opposite d**NaM**, the pair behaves like a natural base pair in which the ajar state resides only very transiently and the long-lived closed state is observed. In the end, it is clear that incorporation of d**NaM**TP would require significant rearrangement of the polymerase to reach the catalytically competent closed state, while d**5SICS**TP spontaneously induces formation of the catalytically competent closed complex explaining its superior incorporation behaviour in KF[170]. However, the fact that a pre-catalytic complex is observed for one of the nucleotides indicates that similar to natural substrates also in case of hydrophobic unnatural base pairs the insertion mechanism includes a distinct intermediate step. To what extend discrimination of the artificial pairing partner from natural substrates takes place in the partially-closed state with a templating d**NaM** or d**5SICS** can not be deduced from the obtained structures and requires further investigation .

2.1.4.4 Model of a closed KlenTaq complex containing a d5SICS-dNaMTP pair

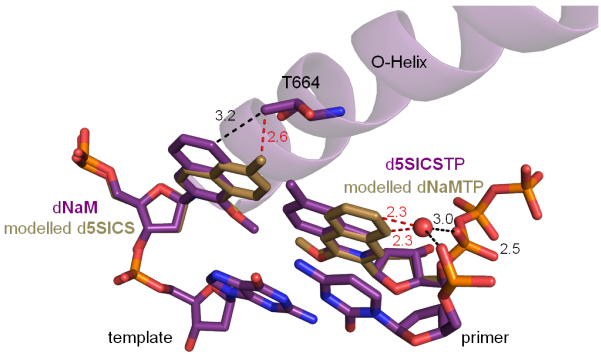


Figure 2.11: Potential clash of Thr664 with d**5SICS** in a closed KlenTaq structure harbouring a d**5SICS**-d**NaM**TP pair. Model of d**5SICS**-d**NaM**TP pair (brown) superposed on the d**NaM**-d**5SICS**TP pair in the closed $KTQ_{dNaM\text{-}d5SICSTP}$ structure (purple). Potential clashes are indicated by red dashes. Distances are given in Å.

As for d**5SICS**TP opposite d**NaM** the state of the closed enzyme prior to the reaction could be trapped but for the other strand context this was impossible, it seems as if the closed arrangement is less favourable in case of d**5SICS**-d**NaM**TP. To investigate this circumstance the d**5SICS**-d**NaM**TP pair was modelled into the structure of $KTQ_{dNaM\text{-}d5SICSTP}$ in the program COOT by superposing the sugar moiety of the two nucleotide triphosphates and rotating the base moiety of d**NaM**TP into the same plane as the one of d**5SICS**TP. The same was done for the templating nucleotide. Investigating the surrounding residues a possible clash between d**5SICS** and Thr664 is apparent (2.6Å, **Figure 2.11**). However, the base moiety still has enough freedom to rotate around the bond connecting it to the sugar C1' and the polymerase has enough flexibility to adapt to the different nucleobase shape to avoid the clash. Therefore, the clash might not be the only or main reason for the impaired incorporation efficiency. Furthermore, mutating Thr664 to a serine to avoid the clash but retain a potential function of the OH group, did also not lead to a closed ternary complex with d**NaM**TP in the active site. But rather in the structure of the mutant the serine side chain points into the same direction as the methyl group of threonine in the wild type (WT) structure (data not shown) retaining the possible clash. Furthermore, also by soaking of crystals of the Thr664Ser mutant with d**NaM**TP we could only obtain a partially closed ternary complex as for the WT enzyme indicating that this small change within the protein might probably not alter d**NaM**TP binding or incorporation. In addition to the possible clash with the threonine side chain

a water molecule bridging the primer terminus and the α phosphate of the d**5SICS**TP would need to be displaced in case of a bound d**NaM**TP. However, it can only be speculated about any role of this water molecule e.g. in stabilizing the triphosphate in its position, and therefore it is not clear if loosing it could have an influence on inducing a stable closed complex. But, as this water molecule is also found in natural complexes of e.g. KlenTaq and BF it might indeed play a role. It is concluded that the different shape of the d**5SICS**-d**NaM**TP base pair cannot be the only explanation for the fact that a closed ternary complex was not obtained, but there might be several factors playing together hindering the formation of a stable active state. One might need to create a Thr664Ala mutant to fully exclude the effect of a possible steric hindrance of Thr664 and a templating d**5SICS** on the closing of the polymerase. Furthermore, kinetic data for d**NaM**MP and d**5SICS**MP incorporation of KlenTaq and several active site single mutants would help to be better able to explain the observed structural data.

2.2 Elongation of the dNaM-d5SICS pair

2.2.1 Preliminary studies

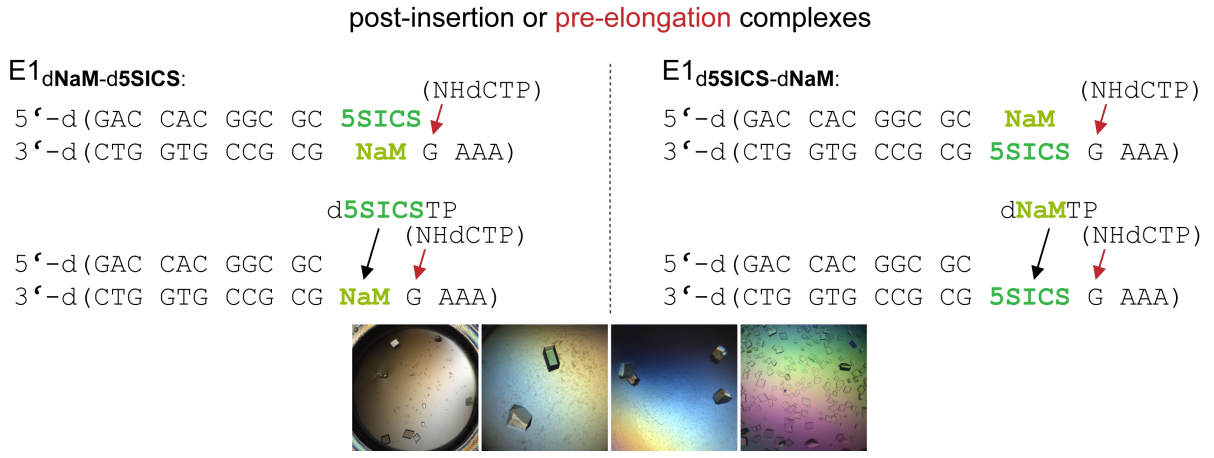


Figure 2.12: P/t complexes $E1_{dNaM\text{-}d5SICS}$ (left) and $E1_{d5SICS\text{-}dNaM}$ (right) used to crystallize KlenTaq elongating a d**NaM**-d**5SICS** pair. To obtain post-insertion binary complexes no non-hydrolyzable triphosphate was added. To screen for pre-elongation ternary complexes non-hydrolyzable dCTP was added, indicated by red arrows. Example images of hits obtained from these setups are shown below the sequences.

Motivated by the insights obtained from the incorporation studies of the artificial nucleotides d**NaM** and d**5SICS** we next sought to investigate the structures of the un-

natural pair in the post-insertion site after a successful incorporation event and before its elongation by a natural substrate. Therefore, we aimed to crystallize KlenTaq with first, the unnatural pair at the post-insertion site (binary complex) and second, the same complex but with the subsequent natural substrate bound opposite the templating nucleotide (ternary complex). For this purpose initially a p/t complex with the same length as in the incorporation studies was used but with the d**NaM** or d**5SICS** shifted one position downstream (**Figure 2.12**). Furthermore, a dG was chosen to be located at the templating position to direct the binding of a non-hydrolyzable dCTP derivative to trap the enzyme in a closed state. Crystallization trials were performed, following two approaches in parallel. Either a primer already carrying the artificial nucleotide at its 3'-end was used (upper panel in **Figure 2.12**) or it was incorporated by KlenTaq just prior to screening for crystals (lower panel in **Figure 2.12**). Crystallization setups were made for both strand contexts, with either d**NaM** in the template and d**5SICS** in the primer (p/t sequence $E1_{\mathbf{dNaM\text{-}d5SICS}}$) or with d**5SICS** in the template and d**NaM** in the primer (p/t sequence $E1_{\mathbf{d5SICS\text{-}dNaM}}$). Furthermore, setups were made in presence of different non-hydrolyzable dCTP analogs (either CH_2, CF_2 or NH bridging the α and β phosphate). Templates for these experiments have also been synthesized by Denis Malyshev from the Romesberg group. The CF_2 bridged non-hydrolyzable dCTP was kindly provided by Frank Streckenbach.

For all different setups with these p/t complexes nicely defined three dimensional crystals were obtained (example images in **Figure 2.12**). The crystals diffracted to around 2-3Å exhibiting all the same space group (no. 18, $P22_12_1$) with similar cell parameters and harbouring two molecules in the asymmetric unit. In these structures, however, the DNA was always bound with the 5' overhanging template nucleotides positioned in the active site (in contrast to the usual DNA binding with the p/t junction positioned in the active site) and the whole DNA duplex thereby shifted away from the enzyme's palm domain. Two examples are shown in **Figure 2.13 A-C** and **Figure 2.14 A-C** in comparison to the natural binary complex KTQ_{dG} and the shift of the DNA is indicated by a red arrow. Due to the positioning of the DNA duplex further away from the active site it is less well stabilized by the enzyme as in binary complexes where the p/t junction resides in the post-insertion site. It is assumed that the DNA is located at a position that is only favored due to crystal packing and is not relevant in solution for the elongation process. Nevertheless, the structures that were obtained were analyzed and gave a hint on how to

proceed in designing new templates for further experiments.
Although in most structures the d**NaM**-d**5SICS** pair and also the ss template region were found disordered, some structures show defined electron density for the unnatural pair and in some of these cases an intercalated structure (**Figure 2.13 D**) is observed. Two of those structures obtained at the beginning of the performed elongation studies will be described in more detail in the following two sections. Although crystallization screening of KlenTaq with the p/t complexes $E1_{dNaM\text{-}d5SICS}$ and $E1_{d5SICS\text{-}dNaM}$ was performed with an excess of non-hydrozyzable dCTP the enzyme crystallized only in the just described binary state. Therefore, I continued optimizing the crystals without addition of non-hydrozyzable dCTP to get binary structures of higher quality for better analysis of the unusual template arrangement (see next sections).

2.2.1.1 Structure of $KTQ(E1)_{dNaM\text{-}d5SICS}$

The first structure, named $KTQ(E1)_{dNaM\text{-}d5SICS}$ was obtained from a setup with d**NaM** in the template, a natural 11mer primer and d**5SICS**TP added as a substrate (sequence context $E1_{dNaM\text{-}d5SICS}$ see **Figure 2.12** lower left panel). The data was reduced in space group $P22_12_1$ (no. 18) with the cell parameters a, b, c = 139.0Å, 166.4Å, 60.0Å and α, β, γ = 90° with two molecules in the asymmetric unit. The structure with an I/σ=2 resolution of 2.65Å was solved by molecular replacement (MR) using $KTQ(E1)_{d5SICS\text{-}dNaM}$ (next section) as search model. Data collection and refinement statistics are given in **Table 2.3** on page 71. DNA is ordered in both molecules but only in one of the molecules the d**NaM**-d**5SICS** arrangement is properly defined in the active site (molecule D). In molecule A the ss template overhang was not modelled due to insufficiently defined electron density. Therefore molecule D is described in the following.
The DNA is bound with its 5' ss template overhang in the active site where the nucleobase of dA_{t+3} packs against the aromatic side chain of Phe667 (**Figure 2.13 E**). The stacking continues with the downstream nucleotides dA_{t+2}, dA_{t+1} and dG_t. The primer terminal d**5SICS** stacks on its upstream primer nucleotide $dC_{p\text{-}2}$ with a distance between the center of the aromatic rings of 3.6Å. The d**NaM** in the template intercalates between its pairing partner d**5SICS** and the templating dG_t (**Figure 2.13 D,E**). The intercalation behaviour of d**5SICS**-d**NaM** is different here than was observed in free duplex DNA[44, 173]. In contrast to the nucleobase edges positioned above each other in the d**NaM**-d**5SICS** pair in free duplex DNA, the nucleobases in $KTQ(E1)_{dNaM\text{-}d5SICS}$

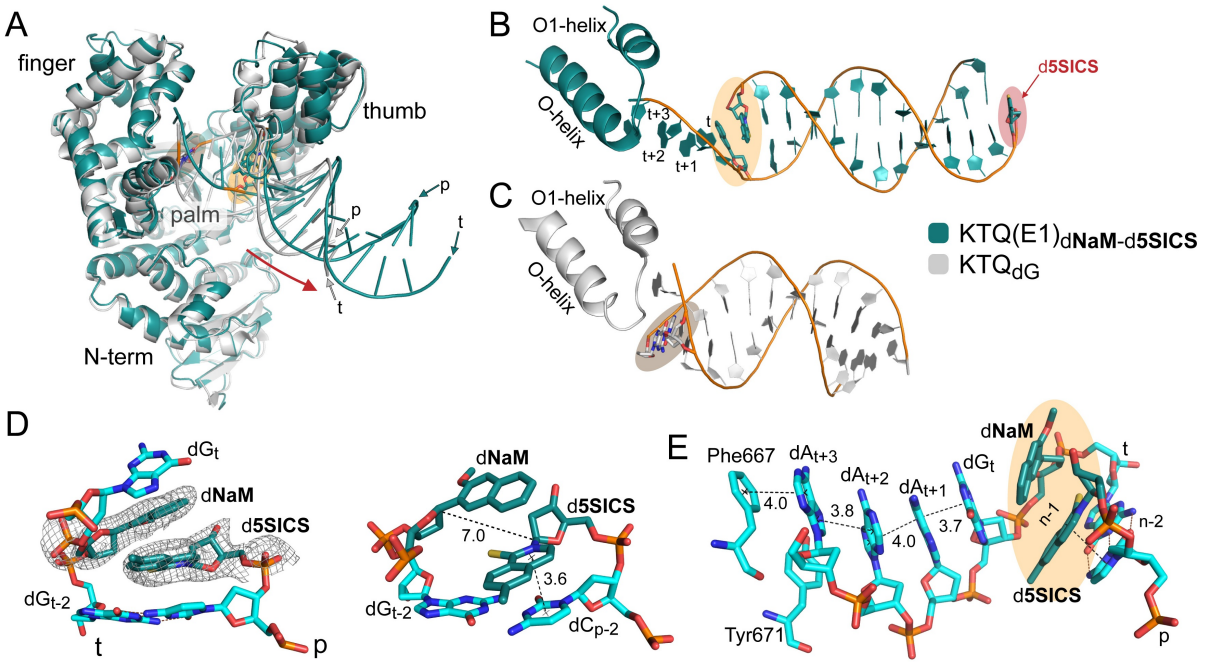


Figure 2.13: Structure of KTQ(E1)$_{\text{dNaM-d5SICS}}$ with the 5' ss template overhang bound in the active site. **A**: The overall structure of KTQ(E1)$_{\text{dNaM-d5SICS}}$ (deepteal) is shown in an overlay with KTQ$_{\text{dG}}$ (gray). The shift of the p/t is indicated by a red arrow and the respective primer and template ends are labelled with p and t. The position of the artificial pair is emphasized by an orange oval and that of the natural pair by a brown oval. **B**: The complete p/t complex of KTQ(E1)$_{\text{dNaM-d5SICS}}$ is shown as cartoon together with O- and O1-helices of the finger domain. The artificial pair is shown as sticks. The additional d**5SICS**MP incorporated at the 3' end of the template is indicated in red. **C**: Same arrangement as in B but for KTQ$_{\text{dG}}$ **D**: Zoom into the position of the p/t junction in KTQ(E1)$_{\text{dNaM-d5SICS}}$. The artificial pair is shown as sticks surrounded by a simulated annealing omit map contoured at 3σ (left panel). The stacking interaction of the d**5SICS** and the dC$_{\text{p-2}}$ nucleobase as well as the C1'-C1' distance is indicated by a dashed line (right panel). **E**: Enlarged view of the DNA upstream from the n-2 position. The stacking arrangement of the ss template and Phe667 as well as d**5SICS** and dC$_{\text{p-2}}$ is indicated by dashed lines. Distances are given in Å.

are situated above each other such that the phenyl ring of d**NaM** is located on top of the pyridine ring of d**5SICS** (**Figure 2.13 D**). The nucleobases do not run parallel to each other but the d**NaM** base is kinked by approximately 22°. In this arrangement the closest distance between the artificial nucleobases is 3.6Å. The internucleotide distance between the C1' atoms of the ribose moieties (C1'-C1'-distance) of d**NaM** and d**5SICS** comprises 7.0Å which is shorter than the distance of the two nucleotides in free duplex DNA (9.1Å[44, 173]) indicating a higher degree of intercalation. This discrepancy may be explained by the different positions of the pair. Whereas in the solution structure of DNA alone the pair is positioned in the middle of the double helix and is covalently constrained on both sides, it is located at the end of a p/t complex in KTQ(E1)$_{\text{dNaM-d5SICS}}$, where less restraints acting on the primer terminus may allow the increased intercalation.

Indeed, the flexibility of d**5SICS** enables its nucleobase to be positioned parallel to the plane of dC_{t-2} resulting in a proper stacking position.

Side note: An additional d**5SICS** nucleotide has been incorporated at the template 3'-end of the blunt-end DNA without a directing templating nucleotide (**Figure 2.13 B**, red). Such template-independent incorporation events are often observed for A-family DNA polymerases[213, 214] as well as members of other polymerase families and is therefore not unusual. In the present case the additional overhanging nucleotide at the 3' template end is involved in crystal contacts promoting stabilization but has no apparent influence on the overall structure. The distinct stacking property of the aromatic d**5SICS** nucleobase most likely favours the observed blunt-end elongation reaction.

2.2.1.2 Structure of $KTQ(E1)_{d5SICS\text{-}dNaM}$

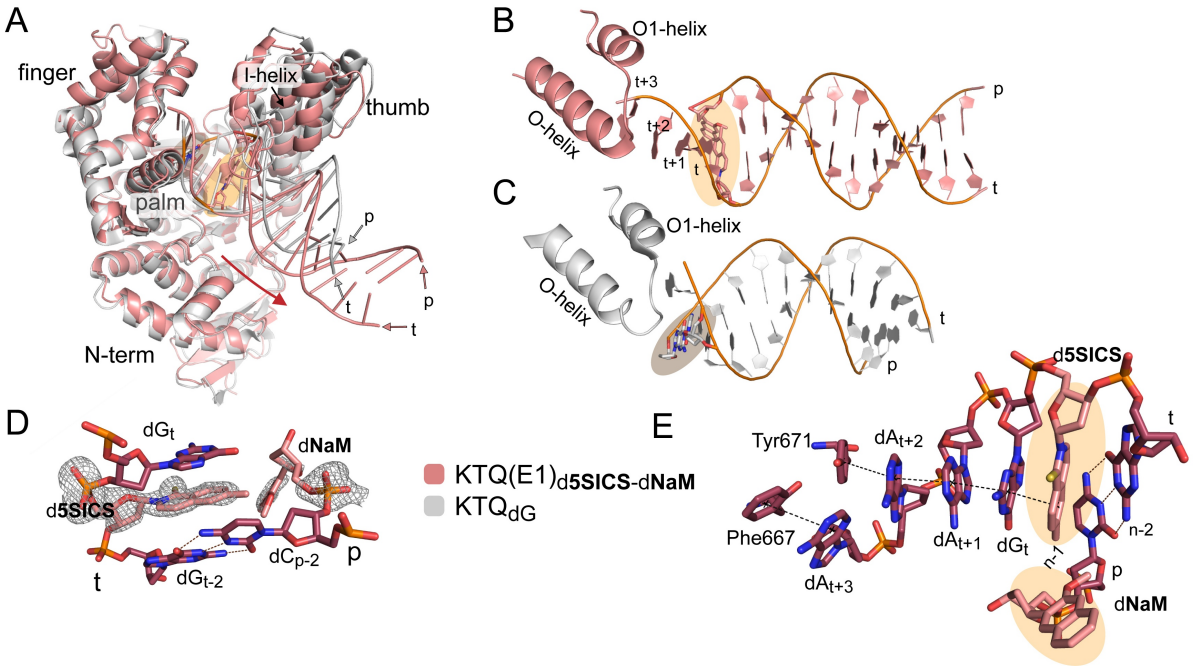


Figure 2.14: Structure of $KTQ(E1)_{d5SICS\text{-}dNaM}$ with the 5' ss template overhang bound in the active site. **A**: The overall structure of $KTQ(E1)_{d5SICS\text{-}dNaM}$ (salmon) is shown in an overlay with KTQ_{dG} (gray). The shift of the p/t is indicated by a red arrow and the respective primer and template ends are labelled with p and t. The position of the artificial pair is emphasized by an orange oval and that of the natural n-1 pair is indicated by a brown oval. **B**: The complete p/t complex of $KTQ(E1)_{d5SICS\text{-}dNaM}$ is shown as cartoon together with O- and O1-helices of the finger domain. The artificial pair is shown as sticks. **C**: Same arrangement as in B but for KTQ_{dG}. **D**: Zoom into the position of the p/t junction in $KTQ(E1)_{d5SICS\text{-}dNaM}$. The artificial pair is shown as sticks surrounded by a simulated annealing omit map contoured at 3σ. **E**: Enlarged view of the DNA upstream from the n-2 position. The stacking arrangement of the ss template and Tyr671 on the one side and d**5SICS** on the other is indicated by dashed lines.

A second structure, named KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$ was solved from successful setups of KlenTaq with a template carrying d**5SICS** at the templating position and a primer already containing d**NaM** (context E1$_{\mathbf{d5SICS\text{-}dNaM}}$, **Figure 2.12** upper right panel). The crystals of KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$ grew in space group $P22_12_1$ (no. 18) with cell dimensions a, b, c = 127.5Å, 172.1Å, 57.0Å and angles α, β, γ = 90° and contain two protein/DNA complexes per asymmetric unit. The structure was solved by MR using the KTQ binary complex KTQ$_{dG}$ as a starting model. The I/σ=2 resolution of the structure is 2.0Å. Data collection and refinement statistics are given in **Table 2.3**. From the two molecules in the asymmetric unit only one shows well ordered DNA (molecule D). In the second molecule (molecule A) the DNA as well as the tip of the thumb domain (protein residues 505-514) which usually interacts with the primer strand, were not modelled. In molecule D only the first five protein residues were not modelled in the final structure due to lack of continuous electron density. Molecule D will be described in the following.
Again the DNA is bound with the 5' overhanging template nucleotides in the active site similar to the situation in KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$ (**Figure 2.14 A**). The nucleobase of dA_{t+3} packs against the aromatic side chain of Phe667 and dA_{t+2} is positioned below the side chain of Tyr671 enabling stacking interactions (**Figure 2.14 D**). This arrangement is different from the one in KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$ where Tyr671 is rotated and does not stack on dA_{t+2} (**Figure 2.13 E**). The nucleotide dA_{t+3} shows higher B-factors than dA_{t+2} suggesting that its stacking with Phe667 has no or only weak stabilizing effects. Furthermore, for the d**5SICS**-d**NaM** pair a different arrangement at the p/t junction is observed here as compared to the previous structure. The d**5SICS** in the template is positioned between its neighbouring nucleotides (dG_{t-2} and dG_t). The nucleobase of d**5SICS** runs parallel to the one of dG_t and the two stack on each other with a distance between the center of the aromatic rings of 3.3Å. A continuous stacking line is formed by d**5SICS**, its upstream nucleotides dG_t, dA_{t+1} and dA_{t+2} as well as Tyr671 (**Figure 2.14 E**). The d**NaM** in the primer is rotated away from its pairing partner and the base moiety points towards helix I of the thumb domain (**Figure 2.14**). Due to the fact that in KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$ the artificial nucleotides do not intercalate but d**NaM** points away from the insertion site, the whole extended p/t complex (from the template 5'- to 3'-end) is shorter by 3.4Å which corresponds to the height of one nucleotide.

Both structures (KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$ and KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$) are in an open conformation and show similar overall protein structures as KTQ$_{dG}$ (**Figure 2.13 A** and

Figure 2.14 A) however, with a completely different DNA binding resulting in a large shift of the DNA out of the enzyme, as described above. In both structures stacking of the ss template overhang with aromatic residues near the insertion site is observed. Whereas in KTQ(E1)$_{\mathrm{dNaM\text{-}d5SICS}}$ the terminal dA_{t+3} nucleobase stacks against Phe667, dA_{t+3} is less well resolved in KTQ(E1)$_{\mathrm{d5SICS\text{-}dNaM}}$ and instead dA_{t+2} stacks with Tyr671. These interactions, together with others between the DNA and the polymerase as well as the crystal contacts in which the DNA is involved, seem to result in a sufficiently stable complex so that well diffracting crystals can grow. Thereby, the differences in the arrangement of the artificial pair may be dependent on the strand context.
As in the structures obtained KlenTaq was not trapped in the desired pre-elongation arrangement, we designed new templates for further crystallization trials. Therefore the knowledge about the intercalating arrangement of the artificial pair in KTQ(E1)$_{\mathrm{dNaM\text{-}d5SICS}}$ at the p/t junction was helpful (see section 2.2.2, "Design of new templates - part I").

2.2.1.3 Post-elongation complexes

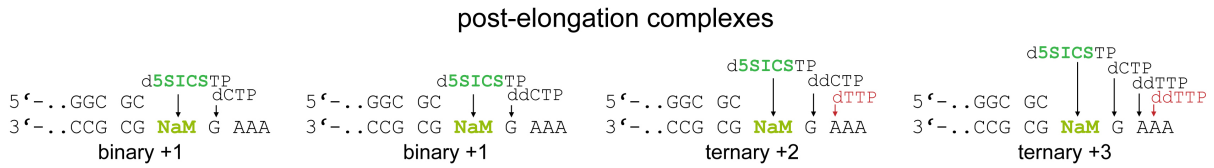


Figure 2.15: P/t design used to crystallize KlenTaq in post-elongation complexes. Nucleotides added to the reaction mixture are shown above the p/t. Black arrows indicate nucleotides which should be incorporated by the polymerase. Red arrows indicate nucleotides which should bind in the active site of the enzyme and are not inserted, leading to ternary complexes.

Besides designing new templates several attempts were made to use the existing templates (E1$_{\mathrm{dNaM}}$ and E1$_{\mathrm{d5SICS}}$, **Figure 2.12**) to obtain crystals with an elongated d**NaM**-d**5SICS** pair (post-elongation complexes) by adding either dCTP, ddCTP, ddCTP and dTTP or dCTP and ddTTP to the enzyme/p/t mixture before crystallization (**Figure 2.15**). However, only the already described crystal form was obtained and the structures always revealed that no additional natural triphosphate has been incorporated or bound stably. Reasons might be that either the elongation process is too inefficient under the conditions used, leading to only partial conversion to elongated complexes or that the elongated complexes were unable to form crystals in the setups we tried, whereby different reasons are conceivable.

	KTQ(E1)$_{\text{dNaM-d5SICS}}$	KTQ(E1)$_{\text{d5SICS-dNaM}}$	KTQ(E7)$_{\text{d5SICS-dNaM}}$
PDB ID	not published	not published	not published
Data collection			
Date of data collection	2011/09/05	2011/06/09	2012/11/23
Wavelength (Å)	1.00001	0.99987	0.97858
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	139.0 166.4 60.0	139.1, 166.4, 60.0	65.0, 129.7, 188.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Molecules/ASU	2	2	2
Resolution (Å)	48.7-2.29 (2.43-2.29)	47.5-1.89 (2.01-1.89)	47.2-2.89 (3.06-2.89)
No. of total reflections	372404 (57946)	637532 (83982)	249618 (37933)
No. of unique reflections	62856 (9594)	100444 (15446)	36329 (5614)
R_{meas} (%)*	7.5 (236.4)	14.7(172.6)	15.1 (190.7)
I/σ	11.42 (0.56)	9.67(1.21)	10.35 (1.03)
Completeness (%)	99.0 (94.0)	99.3 (95.7)	99.0 (96.4)
Redundancy	5.9 (6.0)	6.3 (5.4)	6.8 (6.8)
$CC_{1/2}$ (%)[199]	99.7 (33.6)	99.8(36.8)	99.8 (29.4)
Refinement			
Resolution (Å)	45.5-2.30	47.5-1.90	45.9-2.90
No. of reflections	62726	189400	36067
R_{work}/R_{free}	22.1/24.2	20.9/24.9	20.1/26.0
Coordinate error (Å)[a]	0.42	0.28	0.43
No. of atoms			
Protein (A/C)	4271/4263	8460/8568	4278/3658
DNA (A/C)	470/598	899/-	822/388
Water	23	659	2
Average B-factors (Å^2)			
Protein	105.8/92.8	45.8/49.9	87.2/134.3
DNA	112.9/90.7	94.2	149.5/171.8
Water	84.4	42.6	70.7
R.m.s deviations			
Bond lengths (Å)	0.003	0.004	0.004
Bond angles (°)	0.683	0.895	0.798
Ramachandran (%)[b]			
Favored	95.70	98.3	95.5
Allowed	4.20	1.6	3.90
Outlier	0.09	0.09	0.60

Table 2.3: Summary of the data collection and refinement statistics for KTQ(E1)$_{\text{dNaM-d5SICS}}$, KTQ(E1)$_{\text{d5SICS-dNaM}}$ and KTQ(E7)$_{\text{d5SICS-dNaM}}$. Data processing and refinement files are stored in the following folders:
KTQ(E1)$_{\text{dNaM-d5SICS}}$: /home/kbetz/SLS2011/sep05/nam-5sics2-s34
KTQ(E1)$_{\text{d5SICS-dNaM}}$: /home/kbetz/SLS2011/jun09/5sics-nam2-j5
KTQ(E7)$_{\text{d5SICS-dNaM}}$: /home/kbetz/SLS2012/nov23/E7-5SICS-2
* Values in parentheses correspond to those in the outer resolution shell.
* For definition of R_{meas} see [200].
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

2.2.2 Design of new templates - part I

Based on the insights gained from the structures described in the last sections, new templates were designed to further screen for crystals of KlenTaq in a ternary pre-elongation complex with d**NaM**-d**5SICS** in the post-insertion site. Thereby, in the first instance two different approaches of template variation were followed. The first approach was based on the assumption that the d**NaM**-d**5SICS** pair might also intercalate at the p/t end if it is properly positioned within the active site similar to its arrangement in KTQ(E1)$_{\text{dNaM-d5SICS}}$. As the d**NaM**-d**5SICS** pair in KTQ(E1)$_{\text{dNaM-d5SICS}}$ captures roughly the height of two natural base pairs I decided to shorten the p/t duplex by one base pair for subsequent experiments (templates E2, E3 and E4 in **Table 2.4**). In the second approach the length of the p/t duplex was kept the same, which should be appropriate in case of a planar d**NaM**-d**5SICS** pair in the active site after the incorporation reaction (template E5 and E6 in **Table 2.4**). However, the ss template overhang was chopped by one or more nucleotides with the intention to avoid binding of the overhang into the enzyme's active site. Furthermore, the dA nucleotides were exchanged to dT which has a decreased stacking potential compared to dA[139, 140], rendering it less accessible for interactions with the aromatic side chain of Phe667 or Tyr671. Thus, a structure similar to KTQ(E1)$_{\text{dNaM-d5SICS}}$ or KTQ(E1)$_{\text{d5SICS-dNaM}}$ should be less favored. Different templates combining these two approaches were designed and the templates shown in **Table 2.4** were used for crystallization trials with KlenTaq. All templates were annealed to primers generating a blunt end on the 5' primer side and a template overhang on the other. Non-hydrolyzable nucleotides (NHdCTP or NHdGTP) were used to trap the enzyme in a closed active state.

Shortening the 5' ss template overhang without shortening the duplex (templates E5 and E6) did not lead to diffracting crystals. However, the p/t duplexes which have been shortened by one nucleotide at the blunt end (templates E2, E3 and E4) yielded crystals in complex with KlenTaq from which structures could be solved. Analyzing the first structures with these p/t complexes revealed that again only binary complexes could be trapped, but now with the artificial pair properly positioned within the insertion site (structures are described below). Thereupon it was decided to investigate the binary complexes after efficient incorporation of an artificial nucleotide (post-insertion complexes) in

template name	5'-d(sequence)-3'							structure name	PDB ID
$E1_{dNaM}$	AAAG	**NaM**	GCG	CCG	TGG	TC		$KTQ(E1)_{dNaM\text{-}d5SICS}$	-
$E1_{d5SICS}$	AAAG	**5SICS**	GCG	CCG	TGG	TC		$KTQ(E1)_{d5SICS\text{-}dNaM}$	-
$E2_{dNaM}$	AG	**NaM**	GCG	CCG	TGG	T		$KTQ(E2)_{dNaM\text{-}d5SICS}$	4C8L
$E3_{dNaM}$	TTC	**NaM**	GCG	CCG	TGG	T		$KTQ(E3)_{dNaM\text{-}d5SICS}$	4C8O
$E3_{d5SICS}$	TTC	**5SICS**	GCG	CCG	TGG	T		$KTQ(E3)_{d5SICS\text{-}dNaM}$	4C8M
$E4_{dNaM}$	TTG	**NaM**	GCG	CCG	TGG	C		$KTQ(E4)_{dNaM\text{-}d5SICS}$	4C8N
$E5_{dNaM}$	G	**NaM**	GCG	CCG	TGG	TC		-	-
$E6_{dNaM}$	TG	**NaM**	GCG	CCG	TGG	TC		-	-

Table 2.4: Part I of newly designed sequences used to screen for KlenTaq pre-elongation complexes. Previously used template sequences E1 are shown in gray.

more detail. This state can also deliver important insights into the elongation process, as the arrangement after insertion represents at the same time the starting point for subsequent elongation. In this work post-insertion complexes are defined as the state after insertion reaction took place and before a new natural template, that would initiate the elongation process, is bound. The templates E2, E3 and E4 were again used in binary setups to reproduce the crystallization conditions found in the screening in presence of non-hydrolyzable dNTP derivatives. Structures with the artificial pair in both strand contexts were solved: with d**NaM** at the primer terminus and d**5SICS** at the template position n-1 and vice versa. With these post-insertion complexes in hand, we were able to get insights into the behaviour of the artificial pair after a successful incorporation cycle. These complexes are described in detail in the following section.

2.2.3 Structures of post-insertion complexes

As crystallization of ternary pre-elongation complexes so far turned out to be unsuccessful within the tested conditions we focussed on the step just after incorporation of an artificial nucleotide which is at the same time also the first step before elongation. From these complexes conclusions concerning the elongation reaction should be drawn as this step is still the rate-limiting factor in polymerases processing d**NaM**-d**5SICS**[171].
In total, four post-insertion complexes with the artificial pair bound in the post-insertion site were solved. Three of those are in the same strand context with d**5SICS** positioned at the primer terminus and d**NaM** in the template just varying in template sequence and

length (see $E2_{\mathbf{dNaM}}$, $E3_{\mathbf{dNaM}}$ and $E4_{\mathbf{dNaM}}$ in **Table 2.4**). The fourth complex is with the opposite strand context ($E3_{\mathbf{d5SICS}}$ in **Table 2.4**) viz. d**5SICS** in the template and d**NaM** at the primer terminus. The structures are referred to as KTQ$(E2)_{\mathbf{dNaM\text{-}d5SICS}}$, KTQ$(E3)_{\mathbf{dNaM\text{-}d5SICS}}$, KTQ$(E4)_{\mathbf{dNaM\text{-}d5SICS}}$, and KTQ$(E3)_{\mathbf{d5SICS\text{-}dNaM}}$, respectively. All complexes crystallized in the same space group $C222_1$ (no. 20) with similar cell parameters. Data collection and refinement statistics are summarized in **Table 2.6** on page 88. In the four binary complexes, the polymerase adopts the expected open conformation and the overall enzyme structure is similar to that observed in $KTQ_{\mathbf{dNaM}}$, $KTQ_{\mathbf{d5SICS}}$ or binary complexes with natural p/t (see section 2.1.2). However, the presence of the unnatural pair in the post-insertion site has a significant effect on the p/t structure. Especially the arrangement of the unnatural pair itself is highly different from its structure observed in the pre-insertion complex $KTQ_{\mathbf{dNaM\text{-}d5SICS}TP}$. Instead of pairing edge to edge the nucleobases stack on each other as was observed in the complex KTQ$(E1)_{\mathbf{dNaM\text{-}d5SICS}}$ (section 2.2.1.1) but again with a different arrangement.

2.2.3.1 Structure of KTQ$(E2)_{\mathbf{dNaM\text{-}d5SICS}}$

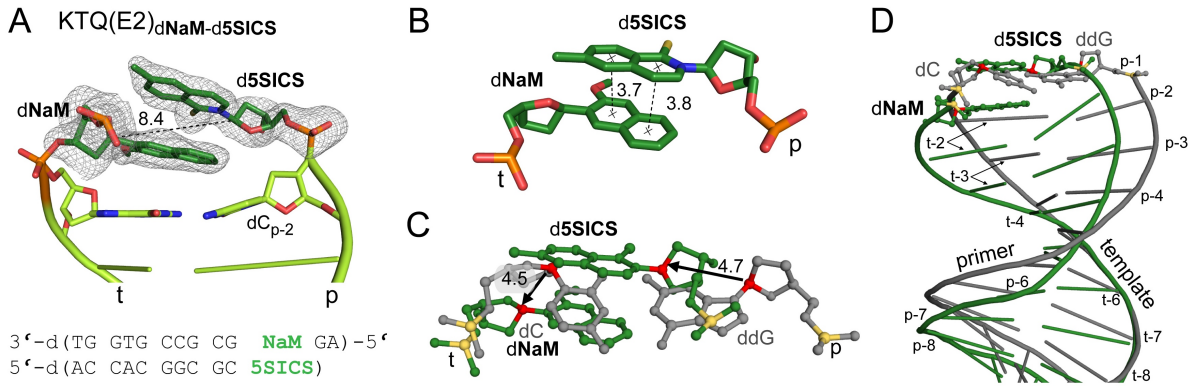


Figure 2.16: Intercalating structure of the nascent base pair in KTQ$(E2)_{\mathbf{dNaM\text{-}d5SICS}}$ and its impact on the surrounding nucleic acid structure. **A**: The unnatural base pair (dark green) and n-2 and n-3 base pairs (light green) are shown as stick or cartoon, respectively with d**NaM**-d**5SICS** surrounded by a simulated annealing omit map contoured at 3σ. Below, the sequence of the p/t is given. **B**: The π-π stacking interaction of the unnatural nucleobases is indicated by dashed lines. **C** and **D**: Comparison of KTQ$(E2)_{\mathbf{dNaM\text{-}d5SICS}}$ (green) and KTQ_{dG} with a natural dC-ddG base pair at the post-insertion site (gray). For this illustration the complete enzyme/DNA complexes were superposed in PyMOL. **C**: Arrangements of the n-1 base pair of both structures is shown in ball and stick representation with C1' atoms colored red and phosphate atoms colored yellow. Position differences of C1' atoms from the planar natural to the intercalated unnatural pair are indicated by arrows. **D**: The duplex portions of the primer/template complexes are shown as cartoon visualizing their different arrangement in the two structures. Distances are given in Å.

In the first post-insertion structure KTQ(E2)$_{\text{dNaM-d5SICS}}$ the template d**NaM** nucleobase cross strand intercalates into the primer strand between d**5SICS** and its 5' dC (dC_{p-2}) (**Figure 2.16 A**). The artificial nucleobases run almost parallel and stack nearly perfectly on each other with distances between the centers of the respective aromatic rings of 3.7 and 3.8Å (**Figure 2.16 B**). To enable the intercalative pairing the artificial nucleotides shift greatly relative to the positions a corresponding natural base pair adopts. The movements needed are shown in **Figure 2.16 C** and **D**. The C1' of the primer d**5SICS** moves 4.7Å towards the template and the C1' of the d**NaM** in the template shifts by 4.5Å into the downstream template direction. The C1'-C1' distance serves as a measure for the extend of intercalation and is 8.4Å in the present case, 2.1Å shorter compared to the 10.5Å distance that is typical for a natural Watson-Crick base pair in the post-insertion site. The extent of intercalation is even greater than in free duplex DNA (C1'-C1' distance: 9.1Å). A possible explanation might be a decreased level of structural restraints acting on the pair when d**NaM**-d**5SCIS** is positioned at the end of a duplex, as opposed to the middle, leading to increased intercalation.

Along with the observed intercalating arrangement the 5' ss template nucleotides adopt an unusual position. The templating nucleobase (dG_n) is positioned near the primer terminus forming H-bonds with the phosphate backbone of the primer d**5SICS** nucleotide via its N1 atom and its C2 amino group, respecitvely (**Figure 2.17 A**). The upstream template nucleotide dA_{n+1} extends parallel to the O-helix and its nucleobase interacts with Arg587 of the palm domain.

The unusual structure of the intercalated pair also affects the position of the remaining p/t duplex within the enzyme. In an overlay of the complete complexes KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ_{dG} the differences in the p/t duplex become apparent (**Figure 2.16 D**). The t-2 and t-3 template nucleotides are perturbed to a high extent whereby the rest of the upstream template nucleotides again adopt similar positions as in KTQ_{dG}. Thereby mainly the phosphate moieties of nucleotides t-4 to t-8 superpose with the ones of the phosphates in KTQ_{dG} but the nucleobases point to different directions slightly shifted away from the active site. The primer is shifted throughout by approximately the same extent. If not the complete enzyme/DNA complex but only the p/t duplexes of KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ_{dG} are superposed, the shape of the duplex starting from n-4 nucleotides is roughly the same in both structures. This indicates that the shape of the DNA duplex starting from three nucleotides away from the artificial pair is not affected by the intercalation but only its position within the enzyme is altered.

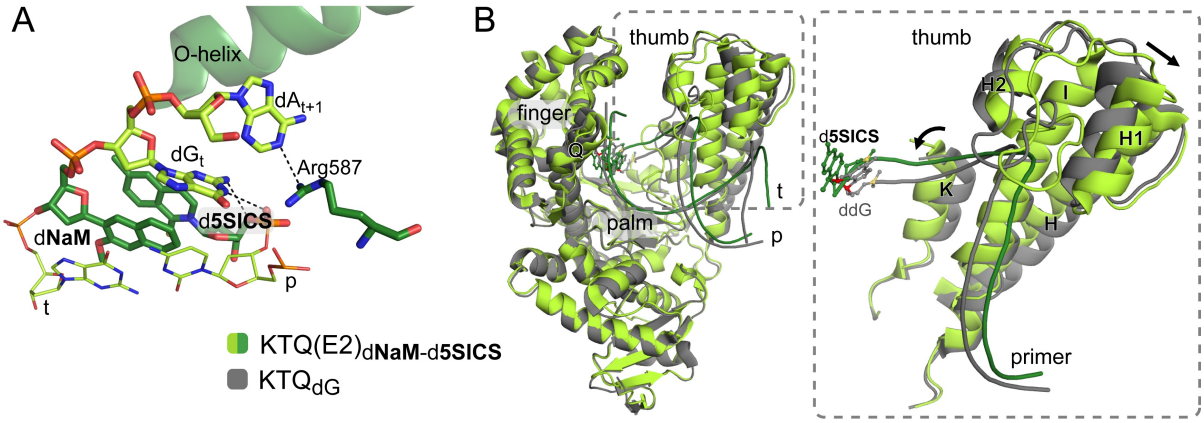


Figure 2.17: **A**: Arrangement of the ss template overhang in KTQ(E2)$_{dNaM-d5SICS}$. The templating dG interacts with the phosphate moiety of the 3'-terminal primer nucleotide d**5SICS** and dA$_{t+1}$ can form a hydrogen bond with Arg587. **B**: Changes in overall structure of KTQ(E2)$_{dNaM-d5SICS}$ due to the intercalated n-1 base pair. Overlay of the overall structures of KTQ(E2)$_{dNaM-d5SICS}$ (green) and KTQ$_{dG}$ (gray) with a zoom into the thumb domain and the primer strand. Helices discussed in the text are labelled. The movements within the thumb domain are indicated by black arrows.

The enzyme adjusts to the changed DNA position by movement of the domains interacting with the DNA (**Figure 2.17 B**). The thumb domain is rotated resulting in a movement of helices H, I and K, which interact with the 3' side of the primer towards the active site. Helices H1 and H2 and its connecting loop, which interact with the 5' side of the primer are moved further away following the repositioned primer strand. Within the finger domain only a small movement of the N-terminal part of the Q-helix is visible whereas the rest of the finger as well as the palm and N-terminal domain are unchanged.
The intercalated state of d**NaM**-d**5SICS** appears to be accommodated by a network of protein residues of the finger domain, including Asn750, Tyr671, Gln754, and Glu615, which pack on the free 3' face of the d**5SICS** nucleobase (**Figure 2.18A**). The primer terminus appears further stabilized by H-bonds between the 3' OH and the phosphate backbone with His784 and Arg587, respectively (**Figure 2.18 D**). The interaction of Arg587 is also observed with a natural substrate (**Figure 2.18 F**), whereas no interaction of His784 with the 3'-OH is present in case of a natural primer 3'-end. However, His784 might interact with the oxygen of the ribose moiety if the side chain has the proper rotamer, which in case of KTQ$_{dG}$ was not specifiable from the electron density. Instead, the 3' OH of a fully natural substrate forms a H-bond with Asp785 as is shown in an analogous structure of BF[129] (**Figure 2.18 I**). For KlenTaq no structure is published with a 3' OH at the primer terminus (all structures have dideoxy terminated primer ends). In an

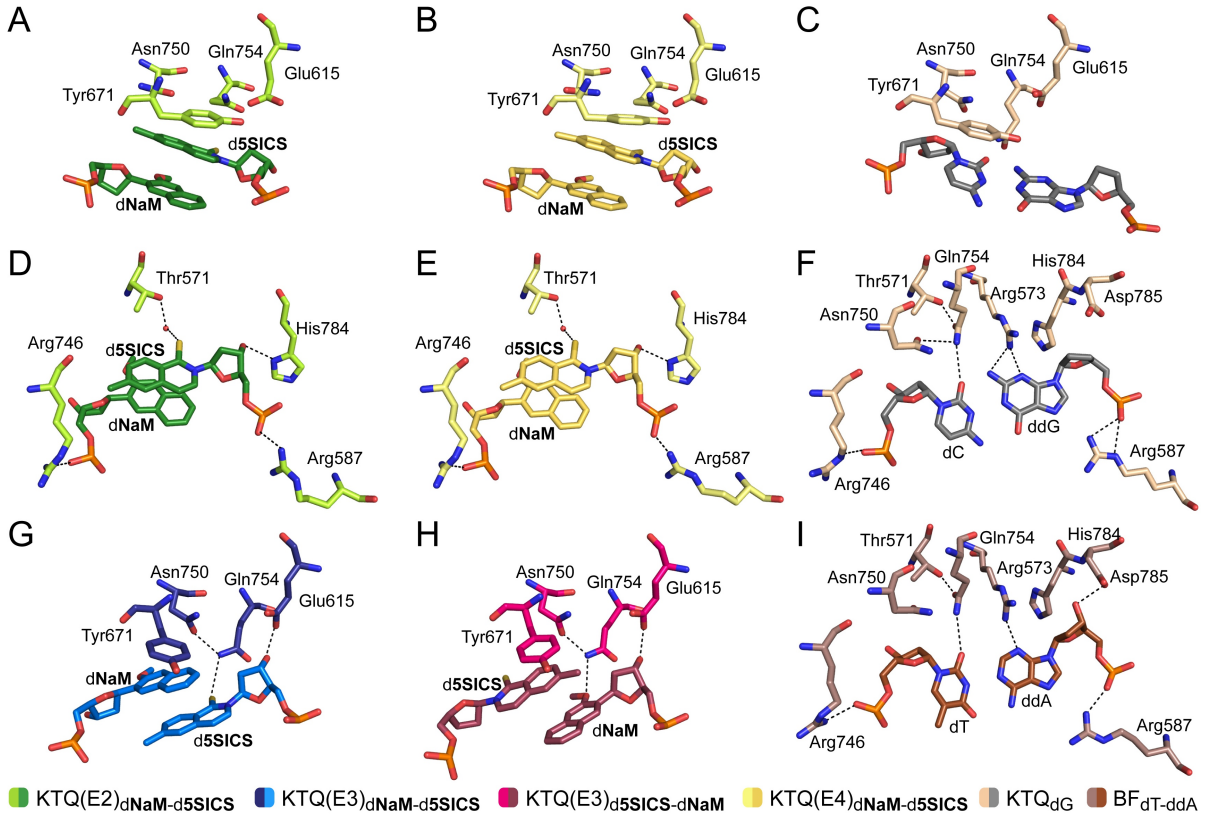


Figure 2.18: Interactions of protein residues with the artificial base pair in the post-insertion complexes KTQ(E2)$_{\text{dNaM-d5SICS}}$ (green), KTQ(E3)$_{\text{dNaM-d5SICS}}$ (blue), KTQ(E4)$_{\text{dNaM-d5SICS}}$ (yellow) and KTQ(E3)$_{\text{d5SICS-dNaM}}$ (raspberry) or natural counterparts in KTQ$_{\text{dG}}$ (gray) or a binary BF structure with a 3' OH group at the primer terminus (PDB ID: 4BDP[129], brown). Hydrogen bonding interactions are shown as dashed lines.

unpublished KlenTaq structure containing a 3' OH at the primer terminus, however, the same interaction as in BF is observed. Additional stabilization of the unnatural primer nucleotide is conceivable by a water mediated H-bond of the sulfur atom of d**5SICS** with Thr571. Furthermore, the phosphate of the template d**NaM** interacts with Arg746 for which an analogous interaction is observed with a fully natural substrate.

2.2.3.2 Structure of KTQ(E3)$_{\text{dNaM-d5SICS}}$

In the second binary post-insertion complex that was solved, a different mode of intercalation is observed, although the template only varies by one nucleotide exchange and an additional nucleotide at the 5' ss template overhang. In the structure KTQ(E3)$_{\text{dNaM-d5SICS}}$ the unnatural base pair also pairs in an intercalated mode but instead of the templat-

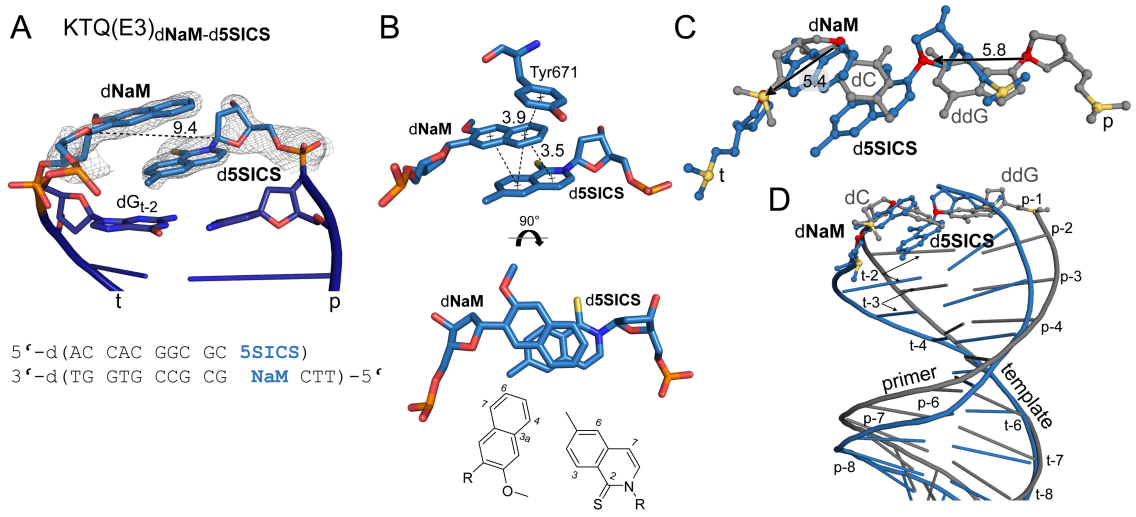


Figure 2.19: **A**: Intercalation pattern of KTQ(E3)$_{\text{dNaM-d5SICS}}$. The artificial pair is surrounded by a simulated annealing mFo-DFc omit map contoured at 3σ and the C1'-C1' distance given in Å. Below, the sequence of the p/t is given. **B**: The π-π stacking interactions of the unnatural nucleobases and Tyr671 are indicated by dashed lines. **C** and **D**: Same arrangement as in Figure 2.16 but for KTQ(E3)$_{\text{dNaM-d5SICS}}$ is shown.

ing base the primer d**5SICS** inserts between the template d**NaM** and its 3' dG (dG$_{t-2}$) (**Figure 2.19 A**). The extent of intercalation appears less than in KTQ(E2)$_{\text{dNaM-d5SICS}}$, with a C1'-C1' distance of 9.4Å (compared to 8.4Å). This reduced intercalation is connected with the positioning of the nucleobases above each other. In contrast to the arrangement in KTQ(E2)$_{\text{dNaM-d5SICS}}$, where the complete isoquinoline and naphthalene moieties of d**5SICS** and d**NaM**, respectively, run parallel and stack almost perfectly on each other, the nucleobases stack only partially in KTQ(E3)$_{\text{dNaM-d5SICS}}$. In the present case the nucleobase of d**NaM** is kinked by approximately 8° compared to the d**5SICS** nucleobase. Furthermore, the C3a, C4, C6 and C7 atoms of d**NaM** are located roughly above C4, C2, C7 and C6 of d**5SICS**, respectively and the two nucleobases are separated by an average distance of 3.5Å. Despite the different intercalation manner in KTQ(E3)$_{\text{dNaM-d5SICS}}$ and KTQ(E2)$_{\text{dNaM-d5SICS}}$, the overall structure of both complexes is very similar (rmsd: 0.503Å). Furthermore, also the influence of the intercalating artificial pair on the residual p/t residues is similar, so that the nucleotides from position n-2 to n-10 superpose well (**Figure 7.4**, Appendix page 182). Compared to the natural binary complex KTQ$_{\text{dG}}$ intercalation is reached by a movement of the C1' from the primer

d**5SICS** by 5.8Å towards the template and a simultaneous shift of the C1' from the template d**NaM** in direction of the downstream template nucleotide (**Figure 2.19 C**). A stabilization of the ss template by flipping towards the primer terminus is not observed in KTQ(E3)$_{\text{dNaM-d5SICS}}$. The d**NaM** shields the templating nucleobase from interacting with the primer strand, and possibly as a result, neither the downstream nucleotides nor Arg587 are well resolved in the KTQ(E3)$_{\text{dNaM-d5SICS}}$ structure.
While the overall structure of the polymerase is similar in KTQ(E3)$_{\text{dNaM-d5SICS}}$ and KTQ(E2)$_{\text{dNaM-d5SICS}}$, there are significant differences in the interactions of protein residues with the unnatural base pair (**Figure 2.18 A** and **D** vs. **G**). Generally, the whole region within the active site is more disordered, characterized by a less well defined electron density and higher B-factors especially in the region of the 5' template and surrounding protein residues. In KTQ(E3)$_{\text{dNaM-d5SICS}}$, the O-helix residue Tyr671 covers the template d**NaM** nucleobase on the template 5' side (distance between aromatic rings: 3.9Å) while Gln754 can form a H-bond with the sulfur of d**5SICS**. In the natural complex Gln754 interacts with the minor groove H-bonding acceptor of the template base rather than the primer nucleotide as observed in the present intercalated state (**Figure 2.18 G** vs. **F**). Comparable to KTQ(E2)$_{\text{dNaM-d5SICS}}$ and the natural complex, the orientation of Gln754 is fixed by interaction with a second protein side chain which is Asn750 in KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ(E3)$_{\text{dNaM-d5SICS}}$ and Thr571 and Asp750 in the natural complex. The 3' OH group at the primer end of d**5SICS** points to a different direction as observed for d**NaM** in KTQ(E2)$_{\text{dNaM-d5SICS}}$ and H-bonds with Glu615. Other than in KTQ(E2)$_{\text{dNaM-d5SICS}}$ or the natural complex KTQ$_{\text{dG}}$ no specific interactions are observed between KlenTaq and the phosphates or ribose moiety of the unnatural nucleotides in neither the primer nor template strands here.

2.2.3.3 Structure of KTQ(E3)$_{\text{d5SICS-dNaM}}$

To examine the effect of strand context the structure of KTQ(E3)$_{\text{d5SICS-dNaM}}$ with d**5SICS** in the template paired opposite d**NaM** at the primer terminus was solved. Again an open binary structure was obtained with the artificial nucleotides pairing via intercalation (**Figure 2.20 A**). In this case the nucleotides pair in a manner similar to the pair in the structure KTQ(E3)$_{\text{dNaM-d5SICS}}$ with the same primer and template sequence but opposite strand context. The intercalation extent is slightly reduced with a C1'-C1' distance of 10.0Å (compared to 9.4Å). The two base moieties of d**NaM** and d**5SICS** run almost parallel but stack on each other in a way that allows only partial π-π stacking interaction

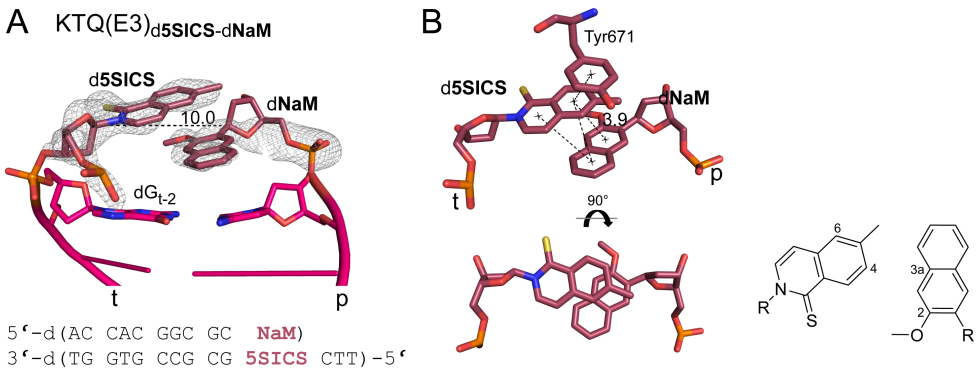


Figure 2.20: A: Intercalation pattern of KTQ(E3)$_{\mathbf{d5SICS\text{-}dNaM}}$. The artificial pair is surrounded by a simulated annealing mFo-DFc omit map contoured at 3σ and the C1'-C1' distance given in Å. Below, the sequence of the p/t complex is given. **B:** The π-π stacking interactions of the unnatural nucleobases and Tyr671 are indicated by dashed lines. Distances are given in Å.

(**Figure 2.20 B**). If regarded from the top, the C4, and C6 atoms of d**5SICS** are located above C2 and C3a of d**NaM** with a distance of 3.9Å between the center of the closest aromatic rings. This positioning is slightly different than observed with the opposite strand context in KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$. However, the protein-DNA interactions, including those involving the unnatural base pair, are conserved in KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$ and KTQ(E3)$_{\mathbf{d5SICS\text{-}dNaM}}$. Thereby Gln754 forms a H-bond with the unnatural nucleotide at the primer terminus via their H-bond acceptors that point towards the minor groove. With KTQ(E3)$_{\mathbf{d5SICS\text{-}dNaM}}$ this is the methoxy group of d**NaM** and with KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$ it is the sulfur atom of d**5SICS** (**Figure 2.18 H**).

2.2.3.4 Structure of KTQ(E4)$_{\mathbf{dNaM\text{-}d5SICS}}$

As two different modes of intercalation with templates either having a 5'-d(AG)-3' or a 5'-d(TTC)-3' template overhang were observed (in KTQ(E2)$_{\mathbf{dNaM\text{-}d5SICS}}$ compared to KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$), the reason for this behaviour was further investigated. Therefore a template with a 5'-d(TTG)-3' single-stranded portion (named E4$_{\mathbf{dNaM}}$) exhibiting the same templating nucleotide as E2$_{\mathbf{dNaM}}$ and the same length as E3$_{\mathbf{dNaM}}$ was used. The structure of KTQ(E4)$_{\mathbf{dNaM\text{-}d5SICS}}$ was solved and an open binary complex structure similar to the ones solved before was obtained. The structures KTQ(E4)$_{\mathbf{dNaM\text{-}d5SICS}}$ and KTQ(E2)$_{\mathbf{dNaM\text{-}d5SICS}}$ or KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$ align with a rmsd of 0.322Å and 0.520Å, respectively. The unnatural base pair is again intercalated and in a manner identical to that observed with KTQ(E2)$_{\mathbf{dNaM\text{-}d5SICS}}$ (**Figure 2.21 A**). The templating d**NaM** nucleotide is again positioned between the primer d**5SICS** and dC$_{p\text{-}2}$ with an almost iden-

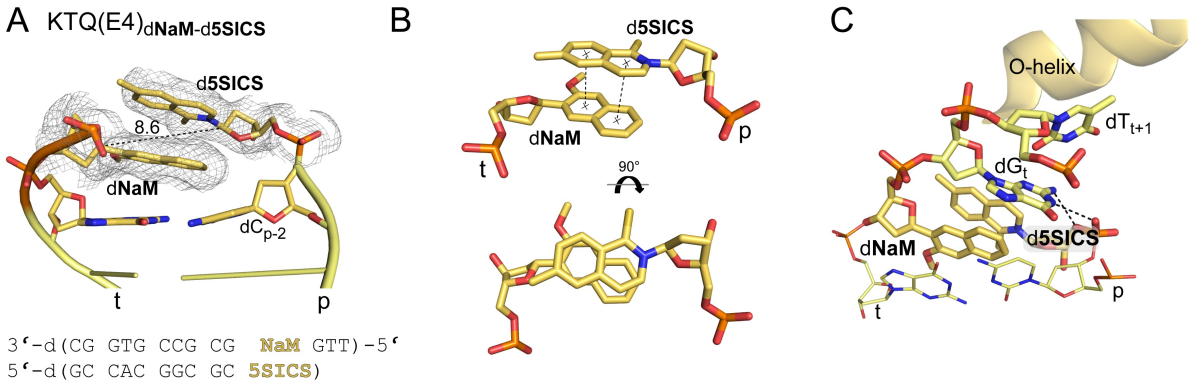


Figure 2.21: Intercalation pattern of KTQ(E4)$_{\text{dNaM-d5SICS}}$. **A**: The artificial pair is surrounded by a simulated annealing mFo-DFc omit map contoured at 3σ and the C1'-C1' distance given in Å. Below, the sequence of the p/t complex is shown. **B:** Side and top view on the artificial pair. The π-π stacking interaction of the unnatural nucleobases is indicated by dashed lines. **C**: Arrangement of the ss template overhang in KTQ(E4)$_{\text{dNaM-d5SICS}}$. The templating dG interacts with the phosphate moiety of the 3'-terminal primer nucleotide d**5SICS**.

tical C1'-C1' distance (8.6Å vs. 8.4Å). The unnatural nucleobases stack on each other in a distance of 3.7Å (**Figure 2.21 B**). Investigating the 5' ss template overhang revealed that dG_t interacts with the phosphate moiety of the primer terminus and the dT_{t+1} residue adopts the same position as the dA_{t+1} in KTQ(E2)$_{\text{dNaM-d5SICS}}$ (**Figure 2.21 C**), however Arg587 does not seem to stabilize the overhanging dT. The last 5' ss template dT is not resolved in the structure. Furthermore, the protein residues interacting with the unnatural pair are the same as described before for KTQ(E2)$_{\text{dNaM-d5SICS}}$ (**Figure 2.18 B** and **E**).

2.2.4 Design of new templates - part II

As the varied templates (E2-E6 in **Table 2.5**) did not yield ternary elongation complexes, two additional p/t constructs were designed and applied to crystallization screening. In the first template the duplex is shortened by an additional base pair at the blunt end side and the template further has a four nucleotide ss overhang (template E7$_{\text{dNaM}}$ and E7$_{\text{d5SICS}}$ in **Table 2.5**). This template was designed following two intentions. First, the shorter length on the 3' side should result in a different crystal lattice (explanation see section 2.2.5) and therefore probably yield ternary crystals. Furthermore, a four nucleotide 5' overhang was chosen to enable simultaneous screening for post-elongation complexes similar as described for E1$_{\text{dNaM-d5SICS}}$ p/t complexes in section 2.2.1.3. Crystals were obtained of KlenTaq in complex with the described template carrying d**5SICS** in the

template name	5'-d(sequence)-3'							structure name	PDB ID
$E1_{dNaM}$	AAAG	**NaM**	GCG	CCG	TGG	TC		$KTQ(E1)_{dNaM\text{-}d5SICS}$	-
$E1_{d5SICS}$	AAAG	**5SICS**	GCG	CCG	TGG	TC		$KTQ(E1)_{d5SICS\text{-}dNaM}$	-
$E2_{dNaM}$	AG	**NaM**	GCG	CCG	TGG	T		$KTQ(E2)_{dNaM\text{-}d5SICS}$	4C8L
$E3_{dNaM}$	TTC	**NaM**	GCG	CCG	TGG	T		$KTQ(E3)_{dNaM\text{-}d5SICS}$	4C8O
$E3_{d5SICS}$	TTC	**5SICS**	GCG	CCG	TGG	T		$KTQ(E3)_{d5SICS\text{-}dNaM}$	4C8M
$E4_{dNaM}$	TTG	**NaM**	GCG	CCG	TGG	C		$KTQ(E4)_{dNaM\text{-}d5SICS}$	4C8N
$E5_{dNaM}$	G	**NaM**	GCG	CCG	TGG	TC		-	-
$E6_{dNaM}$	TG	**NaM**	GCG	CCG	TGG	TC		-	-
$E7_{dNaM}$	TTTC	**NaM**	GCG	CCG	TGC			-	-
$E7_{d5SICS}$	TTTC	**5SICS**	GCG	CCG	TGC			$KTQ(E7)_{d5SICS\text{-}dNaM}$	-
$E8_{dNaM}$	G	**NaM**	GCG	CCG	TGG	C		-	-

Table 2.5: Part II of newly designed sequences used to screen for KlenTaq pre-elongation complexes. Previously tested sequences are shown in gray.

templating position and a primer containing d**NaM** and the solved structure is named $KTQ(E7)_{d5SICS\text{-}dNaM}$. However, with this p/t design a similar structure as with the E1 templates was obtained which means that again the ss template overhang is positioned near the insertion site. The structure is described and discussed in section 2.2.4.1, "Structure of $KTQ(E7)_{d5SICS\text{-}dNaM}$".

In the second template ($E8_{dNaM}$ in **Table 2.5**) the p/t duplex length is unchanged compared to the templates that yielded binary structures (E2, E3 and E4) but it has no 5' ss template overhang in addition to the templating nucleotide. Therefore a blunt end would result in case of dNTP binding. This template was designed to test if a blunt end in the active site (after dNTP binding) leads to a ternary complex, as the templating nucleotide has more freedom and could probably more easily rotate into the insertion site from it's extrahelical position. However, also with this template no ternary structure could be obtained under the tried conditions.

2.2.4.1 Structure of $KTQ(E7)_{d5SICS\text{-}dNaM}$

For template $E7_{d5SICS}$ (**Table 2.5**) we were able to solve a structure in space group $P2_12_12_1$ (no. 19) and cell parameters a, b, c = 110.1Å, 111.9Å, 169.1Å and $\alpha, \beta, \gamma = 90°$. Data collection and refinement statistics are given in **Table 2.3** on page 71. From the

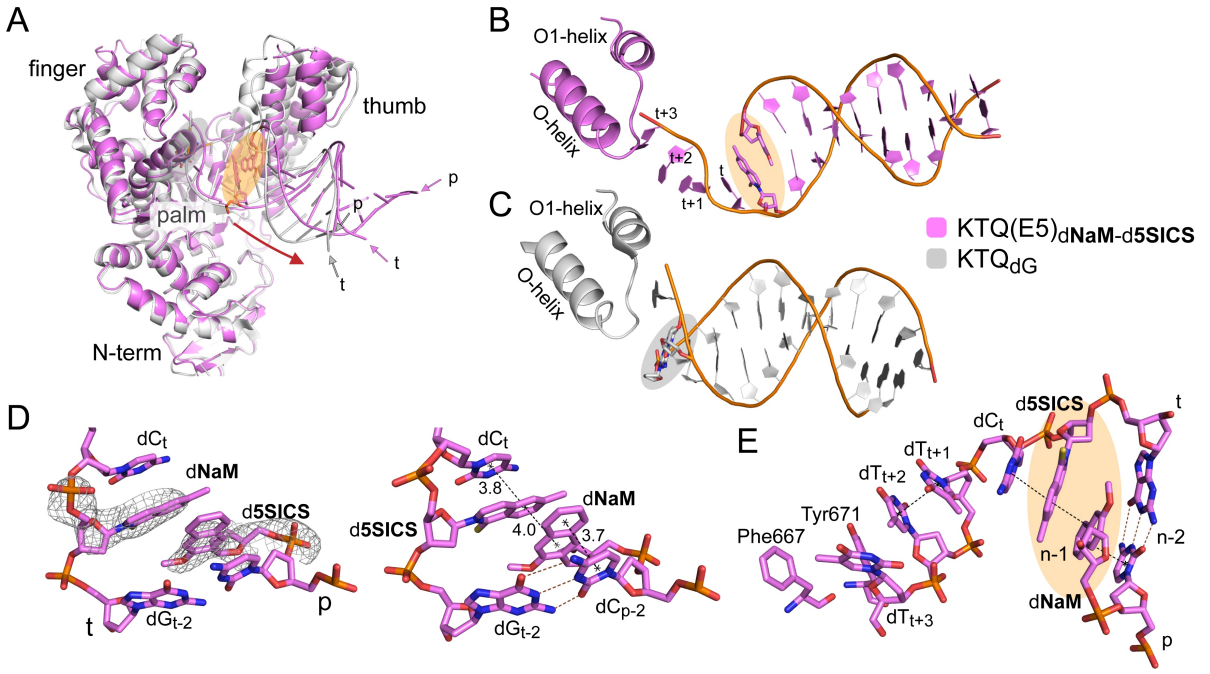


Figure 2.22: Structure of KTQ(E7)$_{\mathbf{d5SICS\text{-}dNaM}}$ with the 5' ss template overhang bound in the active site. **A:** The overall structure of KTQ(E7)$_{\mathbf{d5SICS\text{-}dNaM}}$ (violet) is shown in an overlay with KTQ$_{dG}$ (gray). The shift of the p/t is indicated by a red arrow and the respective primer and template ends are labelled with p and t. The position of the artificial pair is emphasized by an orange oval and that of the natural n-1 pair is indicated by a brown oval. **B:** The complete p/t complex of KTQ(E7)$_{\mathbf{d5SICS\text{-}dNaM}}$ is shown as cartoon together with O- and O1-helices of the finger domain. The artificial pair is shown as sticks. **C:** The same arrangement as in B but for KTQ$_{dG}$. **D:** Zoom into the position of the p/t junction. The artificial pair is shown as sticks surrounded by a simulated annealing omit map contoured at 3σ. and its stacking interactions are indicated by dashes. **E:** Enlarged view of the DNA upstream from the n-2 position. The stacking arrangement of part of the ss template and dC$_t$, d**5SICS** and d**NaM** is indicated by dashed lines. Distances are given in Å.

two molecules in the asymmetric unit molecule A is more ordered and will be described in the following. Molecule D shows high B-factors over the whole enzyme and the DNA duplex. Furthermore, the ss template DNA bound in the active site was not modelled due to disorder. Also in molecule A the DNA strand shows high B-factors but all nucleotides could be modelled.

The overall structure of the protein superposes well with to the open structure KTQ$_{dG}$ (**Figure 2.22 A**). Similar to KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$ and KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$ which were described in section 2.2.1, "Preliminary studies", the template overhang is bound in the active site shifting the DNA duplex part out of the enzyme. The last 5' ss nucleotide (dT$_{t+3}$) is most probably positioned near Phe667 but the nucleobase as well as the Phe667 side chain show high flexibility. The nucleobase of dT$_{t+2}$, packs against Tyr671, how-

ever, the orientation of the aromatic systems do not indicate a π-π stacking interaction. The d**5SICS**-d**NaM** pair intercalates, similar as in KTQ(E1)$_{\text{dNaM-d5SICS}}$. The template d**5SICS** stacks between the d**NaM** at the primer 3' terminus and the templating dC_t nucleotide, resulting in a C1'-C1' distance between the unnatural pair of 8.4Å. Between dC_t and dT_{t+1} the template kinks so that the nucleobases do not stack on each other. However, dT_{t+1} again stacks with dT_{t+2}.
In the section 2.2.2, "Design of new templates - part I" it was described that terminal dA residues were exchanged to dT to decrease stacking properties and avoid binding of the ss template overhang in the active site. With the just described structure it is apparent that this strategy did not lead to success, as we observe a similar arrangement with ss dT nucleotides. Instead, it is assumed that a four nucleotide template overhang in combination with the artificial pair at the p/t junction favors complexes as observed in KTQ(E7)$_{\text{d5SICS-dNaM}}$, KTQ(E1)$_{\text{dNaM-d5SICS}}$ and KTQ(E1)$_{\text{d5SICS-dNaM}}$. One factor that could play a role in formation of this arrangement of the p/t within the enzyme could be favourable crystal contacts of the DNA duplex that promote it. Furthermore, it seems as if the d**5SICS** nucleotide rather than d**NaM** directs the arrangement of the ss template overhang. If d**5SICS** is in the template the upstream nucleotide is oriented in a stacking arrangement with it which is continued differently in upstream direction depending on the ss template sequence being 5'-d(AAAG)-3' or 5'd(TTTC)-3' in KTQ(E1)$_{\text{d5SICS-dNaM}}$ or KTQ(E7)$_{\text{d5SICS-dNaM}}$, respectively. If d**5SICS** is in the primer it stacks with the upstream primer nucleotide and does not directly interact with the ss template overhang. In this case the templating dG as well as thee upstream template nucleotides are positioned to form a stacking cascade with Phe667. To avoid this kind of arrangement in future crystallization trials with the artificial pair, either a longer or shorter template overhang should be used.

2.2.5 Obstacles in obtaining a pre-elongation complex

As already stated above multiple attempts were made to obtain suitable crystals to solve structures of a ternary complexes of KlenTaq with the artificial pair at the post-insertion site and a subsequently bound correct natural triphosphate (dCTP or dGTP). To avoid elongation a non-hydrolyzable nucleotide analog was used for these experiments (NHdCTP or NHdGTP). First, templates were designed in length similar to the ones used for the incorporation situation but with the artificial nucleotide shifted by one position

in 3' template direction. However, these did not lead to the desired crystals (see section 2.2.1). Therefore various shorter templates have been designed and screened. The non-hydrolyzable triphosphate was always added in molar excess (10-100x related to the enzyme). This corresponds to a 1.15-11.5mM concentration of non-hydrolyzable dNTP compared to 0.12mM enzyme, 0.14mM p/t and 0.35mM ddCTP in a 100µl crystallization solution (enzyme:p/t:ddCTP ratio 1:1.2:3) and a final protein concentration of 7.0mg/ml. In each case either a binary structure identical to the post-insertion complexes KTQ(E2), KTQ(E3) or KTQ(E4) described in section 2.2.3 was observed, with no electron density associated with the triphosphate apparent or structures of apo KlenTaq or KlenTaq with uninterpretable electron density in the active site were obtained. One problem which might be a reason for the difficulty of trapping a pre-elongation complex is the high amount of natural triphosphate needed for efficient elongation of the unnatural pair[171]. The K_M values determined by Seo *et al.*[171] for elongation of the d**NaM**-d**5SICS** pair by insertion of dCMP or dAMP are in the range of 2-4µM in both strand contexts. In practice, as a rule of thumb for co-crystallization or soaking, the ligand concentration should be around 10 times the K_d, or even greater if possible, and the total ligand added should be in excess of protein[215]. These concentrations are reached in our experiments. However, in the kinetic studies a molar excess of triphosphate compared to enzyme of minimal 1000x is used (1-2000µM triphosphate and 0.1-1.2nM Kf)[171]. To realize that situation in our experiments, soaking of binary crystals with a high excess of non-hydrolyzable triphosphate analog would be an appropriate approach, as on the one hand for an excessive screening with co-crystallization a high amount of the costly compound is needed and on the other hand extremely high concentrations of ligand might also hamper crystallization. In addition, soaking is the method of choice if no crystals of ternary complexes are obtained, which could be due to the low binding affinity of the ligand, resulting in an inhomogeneous crystallization solution. In this project, however, soaking experiments with the binary crystals that were obtained with the artificial pair located at the p/t junction in the active site of KlenTaq were not possible since the finger domain of the enzyme is trapped in an open conformation by crystal contacts. In the crystal, specific interactions of the finger domain with a symmetry mate are observed by the O1-helix residue Ser644 interacting with Arg716 of a symmetry neighbour and of the backbone oxygen of Trp645 interacting with Arg717 (**Figure 2.23**). Furthermore, the location of the N-terminus of a second symmetry mate might block closure of the finger domain by a clash with the O-helix. Therefore, to further pursue the aim of this particular project, additional attempts

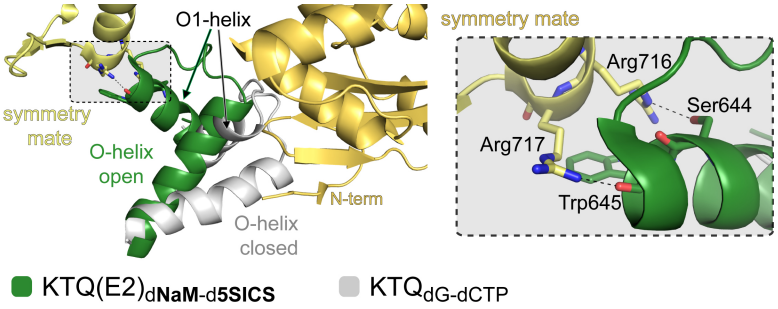


Figure 2.23: Involvement of the finger domain O1-helix in crystal contacts of KTQ(E2)$_{\text{dNaM-d5SICS}}$ and potential clash of a closed O-helix with a symmetry mate. A detailed view of the H-bonding interactions of O1-helix residues is shown on the right side.

need to be made in future to successfully crystallize the desired ternary complexes directly or to obtain binary crystals which could be used for soaking experiments. As the crystals in which the finger domain is stabilized in the open conformation are well ordered and the same crystal form was found in different conditions mutation of the enzyme might be needed to avoid this crystal packing. Furthermore, template and primer lengths could be additionally varied to increase the chance to find new crystal forms. Residues which seem well suited to mutate in this context are Arg716 and/or Arg717 as they - to my knowledge - have not been shown to be crucial for the enzyme's activity.

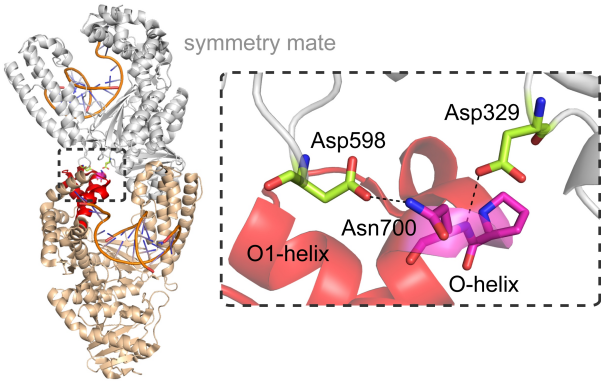


Figure 2.24: Crystal contact within the finger domain of a BF binary complex structure (PDB ID: 1L3S).

Similar to the obtained binary structures, crystal contacts in BF structures of Beese *et al.* trapped the finger domain in an open conformation and hindered ternary complex formation[203]. Here, H-bonds between two aspartate residues (Asp598 and Asp329) to the side chain and main chain nitrogen of asparagine 700 at the N-terminal end of the O-helix stabilized the open conformation (**Figure 2.24**). Those contacts were removed

by mutating aspartate 598 or 329 to alanine singly and in combination and two different crystal forms were obtained in which the finger can move freely[128, 203]. This success delivers additional motivation to create finger domain mutants of KlenTaq in future experiments to yield ternary pre-elongation complexes of the artificial pair dNaM-d5SICS.

2.2.6 Discussion and Conclusion

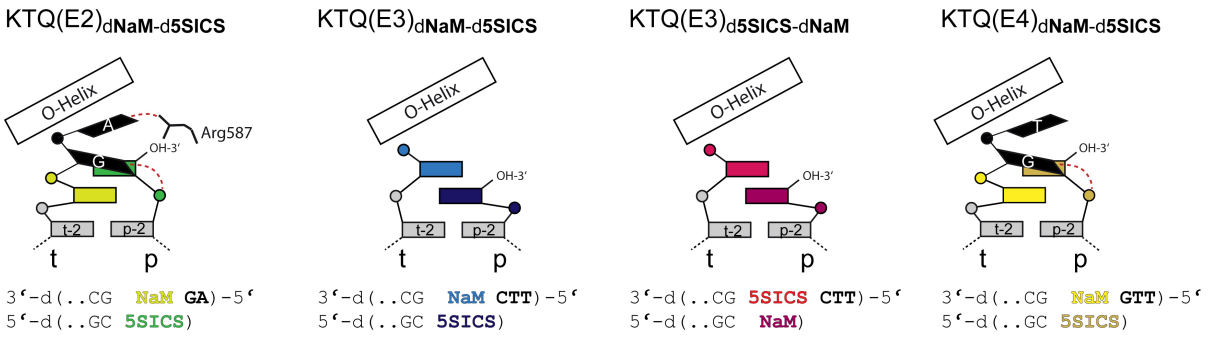


Figure 2.25: Overview of post-insertion complexes with dNaM-d5SICS in the post-insertion site. The scheme shows the arrangement of the artificial base pair and the ss template overhang, as well as part of its sequence. Interactions of the p/t overhang with the phosphate moiety of the primer terminus (in $KTQ(E2)_{dNaM\text{-}d5SICS}$ and $KTQ(E4)_{dNaM\text{-}d5SICS}$) and additionally Arg587 in $KTQ(E2)_{dNaM\text{-}d5SICS}$ are shown as red dashed lines. With a dC at the templating position ($KTQ(E3)_{dNaM\text{-}d5SICS}$ and $KTQ(E3)_{d5SICS\text{-}dNaM}$) the ss template overhang does not interact with the primer terminus and is flexible instead.

In the present study four post-insertion complexes were solved, with the artificial base pair dNaM-d5SICS located at the n-1 position at the p/t junction followed by a templating dG or dC nucleotide. Furthermore, the three complexes in which dNaM is located in the template and d5SICS is positioned at the primer terminus have a slightly different sequence and/or lenght of the 5' ss template nucleotides. In all four structures the unnatural nucleobases pair in an intercalated manner, similar as it was observed in NMR studies of free duplex DNA harbouring this unnatural base pair[173, 194] but with different stacking arrangements. For the four structures solved, two different modes of intercalation are observed. Thereby a common mode of intercalation is found with the templates E2 and E4 (5'-d(AG)-3' ss overhang and 5'-d(TTG)-3' ss overhang) which demonstrates that the mode of pairing is unlikely to depend on the length of the ss template. In these two structures the dNaM inserts between its pairing partner d5SICS and the flanking $dC_{p\text{-}2}$ allowing the templating dG to move towards the primer terminus

	KTQ(E2)$_{\text{dNaM-d5SICS}}$	KTQ(E3)$_{\text{dNaM-d5SICS}}$
PDB ID	**4C8L**	**4C8O**
Data collection		
Date of data collection	2013/06/21	2013/02/05
Wavelength (Å)	1.0000	1.00001
Space group	$C222_1$	$C222_1$
Cell dimensions		
a, b, c (Å)	65.6, 100.5, 204.3	64.8, 99.2, 203.6
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	48.8-1.70 (1.80-1.70)*	48.2-1.75 (1.86-1.75)*
No. of total reflections	492762 (76284)	439052 (70744)
No. of unique reflections	74923 (1184)	66214 (10539)
R_{meas} (%)*	10.6 (275.8)	6.5 (216.4)
I/σ	11.57 (0.85)	17.73 (0.99)
Completeness (%)	99.7 (98.6)	99.6 (99.2)
Redundancy	6.6 (6.4)	6.6 (6.7)
$CC_{1/2}$ (%)	99.9 (49.2)	100.0 (52.7)
Refinement		
Resolution (Å)	48.4-1.70	47.9-1.75
No. of reflections	74501	66155
R_{work}/R_{free}	18.9/21.4	18.5/21.2
Coordinate error (Å)[a]	0.37	0.28
No. of atoms		
Protein	8698	8680
DNA	772	739
Water	274	230
Average B-factors (Å^2)		
Protein	51.0	57.7
DNA	50.6	71.5
Water	43.4	49.3
R.m.s deviations		
Bond lengths (Å)	0.011	0.005
Bond angles (°)	1.239	0.964
Ramachandran (%)[b]		
Favored	97.23	97.97
Allowed	2.77	1.84
Outlier	0.00	0.18

Table 2.6: Summary of the data collection and refinement statistics for KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ(E3)$_{\text{dNaM-d5SICS}}$. Data processing and refinement files are stored in the following folders:
KTQ(E2)$_{\text{dNaM-d5SICS}}$: /home/kbetz/SLS2013/jun21/jun21-E3-NaM-6/phenix/4C8L
KTQ(E3)$_{\text{dNaM-d5SICS}}$: /home/kbetz/SLS2013/feb05/13-E6-NaM-2/phenix
KTQ(E3)$_{\text{d5SICS-dNaM}}$: /home/kbetz/SLS2013/jun21/jun21-E6-5SICS-2/phenix
KTQ(E4)$_{\text{dNaM-d5SICS}}$: /home/kbetz/SLS2012/jul05/E2-NaM-plus2-3/phenix/4C8N
* Values in parentheses correspond to those in the outer resolution shell.
* For definition of R_{meas} see [200].
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

	KTQ(E3)$_{\text{d5SICS-dNaM}}$	KTQ(E4)$_{\text{dNaM-d5SICS}}$
PDB ID	**4C8M**	**4C8N**
Data collection		
Date of data collection	2013/06/21	2012/07/05
Wavelength (Å)	1.0000	0.97793
Space group	$C222_1$	$C222_1$
Cell dimensions		
a, b, c (Å)	65.4, 101.5, 204.5	65.6, 101.1, 204.3
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	49.2-1.57 (1.66-1.57)*	49.1-1.88 (1.99-1.88)*
No. of total reflections	577762 (53959)	361542 (48980)
No. of unique reflections	94841 (14820)	55373 (8568)
R_{meas} (%)*	6.2 (155.1)	13.1 (200.5)
I/σ	13.48 (0.78)	9.1 (0.80)
Completeness (%)	99.5 (97.1)	99.4 (96.6)
Redundancy	6.1 (3.6)	6.5 (5.7)
$CC_{1/2}$ (%)	99.9 (53.1)	99.8 (47.6)
Refinement		
Resolution (Å)	49.3-1.57	49.1-1.88
No. of reflections	94768	55305
R_{work}/R_{free}	17.9/20.7	19.6/23.2
Coordinate error (Å)[a]	0.28	0.34
No. of atoms		
Protein	8728	8657
DNA	770	774
Water	338	203
Average B-factors (Å^2)		
Protein	49.0	51.7
DNA	64.9	50.0
Water	46.2	43.5
R.m.s deviations		
Bond lengths (Å)	0.011	0.007
Bond angles (°)	1.283	1.062
Ramachandran (%)[b]		
Favored	98.35	98.14
Allowed	1.28	1.48
Outlier	0.37	0.37

Table 2.6: continued

and form stabilizing interactions with it (**Figure 2.17** and **2.25**). In contrast a rather inverse intercalation pattern is observed for the pair in KTQ(E3) stuctures (with a 5'-d(TTC)-3' template overhang) for both situations: d**NaM** in the template or d**5SICS** in the template. Here the nucleotide in the primer is not located above the one in the template but stacks between its partner and the downstream $dG_{t\text{-}2}$. With this mode of intercalation the templating dC nucleotide is unable to mediate interactions with the primer terminus and in turn is rotated away from the duplex. The smaller dC might not

be able to interact with the primer terminus similar to dG and therefore not promote the same stabilization. These observations suggest that the intercalation pattern depends mainly on sequence-specific interactions of the flanking nucleotides (only shown for the templating nucleotide) with the specific packing interactions between the intercalating nucleobases being of secondary importance. Interestingly, the polymerase appears to be able to provide unique stabilizing interactions to the two types of intercalated structures at the primer terminus. In KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ(E4)$_{\text{dNaM-d5SICS}}$ one face of the primer terminus is unpacked by a flanking nucleobase, and its position is stabilized by packing interactions with Tyr671, Asn750, Gln754 and Glu615. Moreover, the primer terminus may be stabilized by ionic interaction between its phosphate moiety and Arg587, a water mediated H-bond between its sulfur and Thr571, and a H-bond between its 3' OH group and His784. On the template side the position of d**NaM** can be stabilized by an ionic interaction between its phosphate moiety and Arg746. In KTQ(E3)$_{\text{dNaM-d5SICS}}$ and KTQ(E3)$_{\text{d5SICS-dNaM}}$ the intercalated state adopted leaves one face of the unnatural nucleobase in the template (d**NaM** and d**5SICS**, respectively) unpacked by a flanking nucleobase, and its position seems stabilized by packing interactions with O-helix residue Tyr671. In this case, the position of the primer terminus can be stabilized by a H-bond between its 3' OH and Gln615 and by a H-bond between Gln754 and the H-bond acceptor ortho to the glycosidic linkage (methoxy in d**NaM** and sulfur in d**5SICS**). In contrast to the situation in KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ(E4)$_{\text{dNaM-d5SICS}}$ neither the template nor the primer strand is stabilized via interactions with the phosphate backbones of the artificial nucleotides here. Overall, some of the interactions observed with both intercalation modes are also found in a fully natural complex as is shown for a dC-ddG pair at position n-1 in **Figure 2.4** whereas some interactions occurring in the natural complex are not found for the artificial pair (e.g. with Arg573 or Asp785). The rather different interactions by which the two intercalated structures are accommodated in the post-insertion site reveals that the polymerase is surprisingly plastic in adapting to this non-natural arrangement. However, both arrangements observed require deintercalation and significant remodeling of the polymerase active site for incorporation of the next dNTP. This circumstance likely explains why structures with the next correct natural triphosphate could not be obtained and also why extension of the unnatural base pair is less efficient than incorporation of the unnatural nucleotides. Besides the accommodation of an intercalated d**NaM**-d**5SICS** pair it is assumed that the polymerase is also able to adapt to a planar structure of d**NaM**-d**5SICS** in the post-insertion site. To visualize that, the

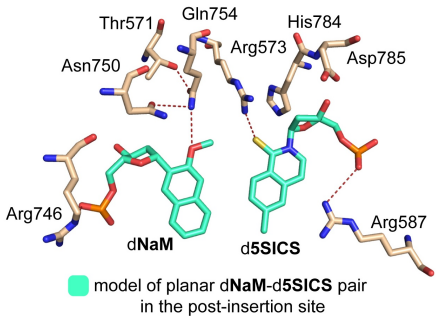


Figure 2.26: Model of d**5SICS**-d**NaM** pairing edge to edge in the post-insertion site. Description see text.

planar pair from the $KTQ_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}TP}$ structure was modelled into the post-insertion site of KTQ_{dG} by superposing the pairs in the program Coot. The model is shown in **Figure 2.26**. Possible interactions of the pair with the enzyme are shown by brown dashes. Just as the minor groove H-bonding acceptors of a natural pair the methoxy and sulfur of d**NaM** and d**5SICS**, respectively, could interact with Gln754 and Arg573. Furthermore, the interactions of the phosphate backbone with Arg746 in the template and Arg587 in the primer as well as Asp785 with the 3' OH group would most likely be possible. Due to the slightly larger internucleotide distance of the unnatural pair compared to a natural pair (as observed in the insertion complex $KTQ_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}TP}$) the enzyme side chains surrounding the pair might need to shift to some extent. In turn, the arrangement of components involved in catalysis might be positioned differently compared to the natural case and furthermore can explain the lower elongation efficiency. Still, it is assumed that a planar arrangement of the pair is indeed possible at the post-insertion site, whereby the fact that an intercalated state could exclusively be crystallized is a strong evidence for it to be more stable and therefore strongly favoured.

2.3 Mechanism of replication for hydrophobic artificial base pairs

Based on all pre-insertion and post-insertion structures reported in this work, as well as previously reported structural data [32, 90, 120, 173, 216], the following mechanism of replication is proposed for a hydrophobic artificial base pair (**Figure 2.27**). The unnatural triphosphate initially binds to the open O-helix via electrostatic interactions which results in a flexible complex that samples different conformations. As soon as sufficiently

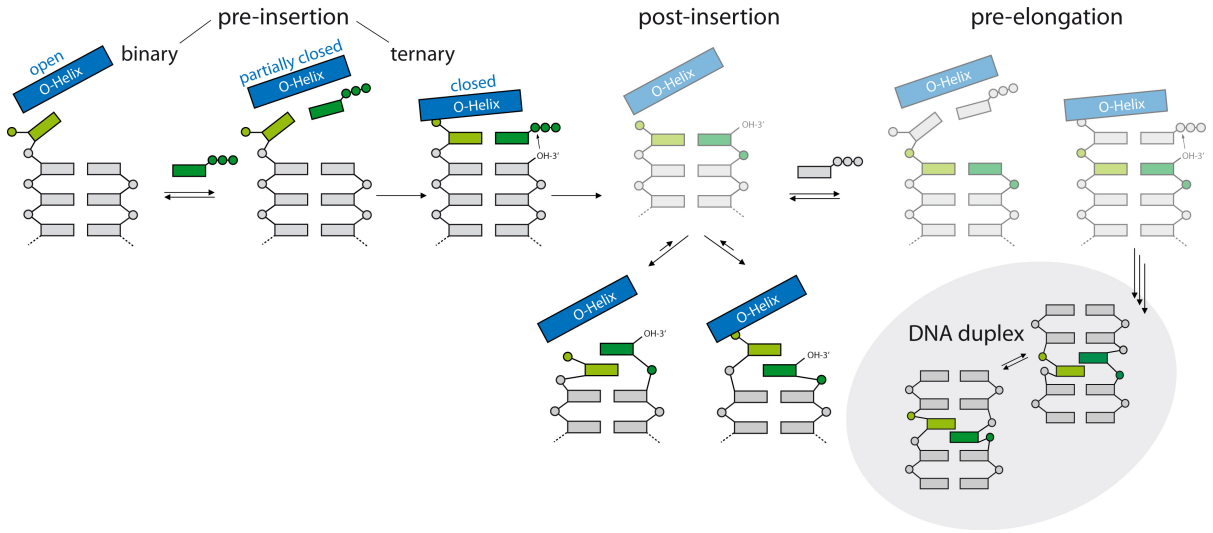


Figure 2.27: Scheme of the proposed mechanism of replication for hydrophobic artificial base pairs. Intermediates not yet validated by structural studies (i.e., elongation complexes) are shown in lighter color. The steps corresponding to incorporation of the unnatural monophosphate and subsequent extension of the nascent unnatural base pair are shown. Thereby the O-helix of the protein is shown as blue rectangle, phosphates are indicated with circles, natural nucleosides are indicated with gray rectangles, and the unnatural nucleosides are indicated with dark and light green rectangles. The structure of the unnatural pair in free duplex DNA[44, 173] after several rounds of extension and enzyme dissociation is shown in the gray oval.

stabilizing hydrophobic and packing interactions are made between the templating nucleotide and the newly bound triphosphate, the open-to-closed transition is induced. The closure in turn induces the unnatural base pair to adopt a planar, Watson-Crick-like pairing which fits into the constraints of the active site. Subsequently the incorporation reaction takes place. In case of d**5SICS** incorporation the intermediate states seem to be only transient and by using a dideoxy primer terminus an active stable complex with all components (including the catalytic magnesium ions) positioned for the imminent reaction could be trapped. For the insertion of d**NaM**, however, a closed state was not obtained but the enzyme was trapped in between, at an ajar-like state with the unnatural triphosphate remaining bound to the partially-closed O-helix. Therefore, either this half-open complex is relatively stable in the special case of d**NaM**TP or the corresponding closed state is unstable. This means that based on the observed structure further thermal fluctuations are necessary to reach a catalytically active state and to proceed with the chemical step. After incorporation of either d**5SICS**MP or d**NaM**MP, the polymerase returns to the open conformation and pyrophosphate is released[217]. In a natural complex the newly incorporated nucleotide would now pair with its partner in

the post-insertion site and the templating nucleotide would be rotated away from the p/t duplex adopting different positions. Tyr671 instead would stack on the exposed nascent base pair. In case of a hydrophobic unnatural base pair, however, the nucleotides adopt a cross-strand intercalated structure similar as in free duplex DNA but with a different arrangement, rendering continued primer elongation difficult. In this state depending on the nucleotides flanking the artificial pair, different intercalation modes are possible which the DNA polymerase is able to stabilize by specific interactions. Before elongation, additional thermal fluctuations are needed to both deintercalate the unnatural base pair and reorganize the polymerase active site. If the unnatural pair has to be reorganized into a fully planar Watson-Crick-like arrangement or partial intercalation is allowed for further elongation is still unclear. However, as the chemistry step requires a very specific arrangement of the catalytic residues and Mg^{2+} ions in the active site, as well as precise positioning of the 3' OH group on the primer for the nucleophilic attack on the incoming triphosphate mimicking the natural structure as good as possible might indeed be necessary. Furthermore, we do not know to what extent deintercalation is influenced or forced by binding of the new substrate or subsequent finger domain closure. Anyhow, the extension reaction can take place as was shown by kinetic studies albeit still with low efficiencies[170, 171, 173], rendering the extension the bottleneck in the replication of DNA containing the unnatural base pair. In any case our model predicts that an optimization of the artificial pair might be possible by making changes to the nucleobase analogs that would decrease the stability of the intercalated structures e.g. by reducing the aromatic surface area of the nucleobase. A generalized model of unnatural base pair replication has already been proposed before we obtained our structures[171, 218]. Romesberg and co-workers stated that "*efficient replication of predominantly hydrophobic unnatural base pairs requires nucleobase intercalation that is sufficient to facilitate synthesis but not so stabilizing as to inhibit deintercalation and continued primer extension*" [171]. Indeed it could be shown recently in the Romesberg group that reducing hydrophobicity of d**5SICS** and therefore favouring deintercalation yielded an artificial pair with improved incorporation and elongation properties[169]. This success strongly approves the importance of understanding the mechanisms of insertion and extension of unnatural hydrophobic nucleotides, and thereby the present work, to be able to further improve promising third base pair candidates.

3 Results and Discussion part II: Structural characterization of the B-family DNA polymerase 9°N

A part of the work described in this chapter is published in [193].

3.1 First attempts to crystallize 9°N in complex with DNA

When our group started the structural work on B-family DNA polymerases, an apo structure of 9°N (short for the DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7)[57] had already been available for ten years and no binary or ternary structure was subsequently solved by the same or other groups to complete the study. We aimed to crystallize 9°N in presence of DNA and substrates to get more insights into the incorporation mechanism of this high fidelity enzyme and to lay the foundation for a future study to characterize the ability of 9°N and other archaeal family B polymerases in handling modified substrates. Functional studies with nucleotides that are modified at the sugar residue[50–52] or the nucleobase[7, 14, 16, 53–56] have shown many times that archaeal DNA polymerases belonging to sequence family B e.g. 9°N, *Pyrococcus furiosus* (Pfu) DNA polymerase and *Thermococcus kodakaraensis* (KOD) DNA polymerase are more efficient at utilising modified nucleotides than are DNA polymerases from sequence family A, such as *Taq* DNA polymerase.

In the present crystallographic study a 9°N variant containing the exonuclease active site mutations Asp141Ala, Glu143Ala was used to diminish exonuclease activity. This double mutant is named 9°N throughout this work. The 9°N gene was cloned into the expression vector pET21b and was overexpressed in *E. coli* without an affinity tag for purification (see section 6.2.4 for details). The enzyme was purified from *E. coli* proteins

and from a truncated 9°N version via heat denaturation followed by a Heparin column and size-exclusion chromatography. For crystallization setups the purified enzyme was mixed with a p/t complex and ddNTP in order to trap closed active ternary complexes. Crystal screening was first performed with different p/t constructs that have already been successfully used in crystallization experiments with other family B DNA polymerases, e.g. from Enterobacteria phage RB69 (RB69)[93], *Thermococcus gorgonarius* (Tgo)[106] or yeast DNA polymerase δ (pol δ)[100]. Furthermore, templates of different length (11 to 16 nucleotides long) were used in crystal screening. These templates were annealed to appropriate primers resulting in oligonucleotide complexes forming a blunt end on one side and leaving a two or three nucleotide 5' template overhang on the other side.

C-terminally truncated 9°N: 9°N△25. To ease crystallization a construct of 9°N lacking 25 residues at the C-terminus was created and used in crystal screening besides the full length 9°N. This region was not resolved in the available 9°N apo structure 1QHT and also in structures of KOD obtained in our group (data not shown). Also in the published apo KOD structure (1WNS) 16 residues of the C-terminus were not modelled. Therefore this region was assumed to be flexible and removing it could make the molecule more rigid, favouring crystallization. The 9°N△25 gene was expressed (see section 6.2.5) and the purified protein was shown to be PCR active, indicating that the deletion did not significantly affect the enzyme's activity.

A large amount of crystallization conditions were screened with 9°N and 9°N△25 and the above-mentioned different p/t complexes and crystals grew under various conditions. Most of the crystals had a needle-like shape, but also some three dimensional crystals were found in the tested conditions. However, although DNA was present in all setups only apo structures could be obtained. Different reasons are imaginable for the fact that only apo structures instead of structures containing DNA were solved. Either the DNA that was used is not bound stably enough by the protein and thus growth of crystals containing DNA was hindered due to an inhomogeneous solution. Or, crystals growing with the protein alone are more ordered, resulting in well defined crystals which were preferentially used for further optimization and diffraction experiments. In this scenario crystals containing DNA could just have been missed to get optimized by chance. To overcome this problem two different approaches were considered. The first plan was to covalently link the DNA and the polymerase via a disulfide bond. This approach where

a polymerase is cross-linked to a primer/template complex by a disulfide bond involving an introduced cysteine residue and an alkanthiol-modified dG nucleotide has been developed by Harrison and Verdine laboratories and was successfully used for crystallization of the HIV-1 Reverse Transcriptase in complex with DNA[74]. In a second approach dye labelled templates were used to visually distinguish between DNA-containing and apo crystals[219]. As the second approach was easier to realize I stopped working on the cross-linking project which was started first and concentrated on the second approach. Its concept and results are described in detail in section 3.3, "Crystallization of 9°N in complex with DNA".

3.2 Apo structures of 9°N

In the course of trying to crystallize 9°N in complex with DNA several structures of apo 9°N in different space groups were solved. Thereby also a fully open complex was obtained in contrast to the available closed 9°N apo structure 1QHT (described in section 1.2.3.2). Although DNA was present in the setups no electron density for DNA was observable. In most structures also the possibility that DNA is bound but disordered could be excluded as either there is no space for DNA binding due to symmetry related molecules occupying the DNA-binding cleft and/or the enzyme itself is stabilized in a conformation in the crystal that would not be expected to allow stable DNA binding. From apo and DNA-complexed structures of RB69[61] and also from results shown later (see section 3.4) and recently published studies[193, 220] it is known that in B-family polymerases the thumb domain moves upon DNA binding from an open to a closed state.
Three of the 9°N apo structures solved in this work are described in the following, whereby especially the fully open structure (9°N-apo1) is of importance and described in more detail. The Data collection and refinement statistics for the 9°N apo structures are shown in **Table 3.1**.
Side note: In their publication of the apo 9°N structure 1QHT Rodriguez *et al.* introduced a nomenclature for secondary structure elements of the enzyme [57]. In the following description of 9°N structures this nomenclature is used.

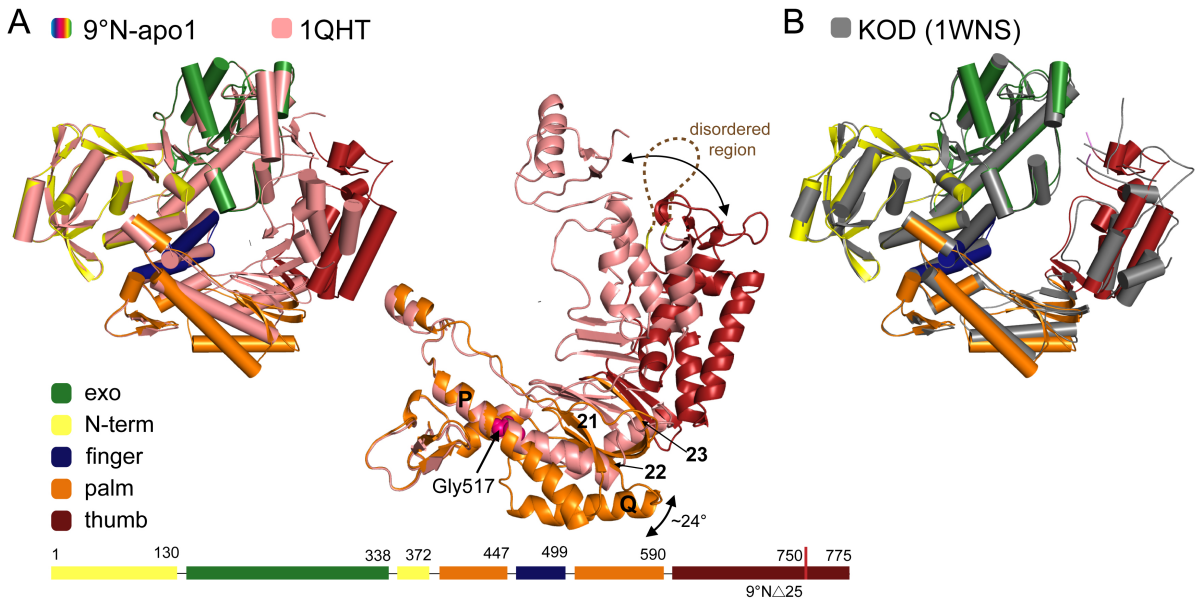


Figure 3.1: **A**: Overall structure of 9°N-apo1 with its different domains colored as indicated, superposed on the previously published apo structure 1QHT[61] (salmon). An enlarged view of the thumb and palm domain which differ most among the two structures is shown and the differences are indicated by arrows. The location of the unresolved region in 9°N-apo1 is indicated by a brown dashed line. **B**: Overlay of 9°N-apo1 (colored by domains) with the open apo structure of KOD (PDB ID: 1WNS)[58] shown in gray.

3.2.1 Structure of 9°N-apo1

The first 9°N apo structure described in this work is termed 9°N-apo1 and adopts a fully open conformation like the apo structures of closely related B-family members KOD (PDB ID: 1WNS)[58], Pfu (PDB ID: 2JGU) or Tgo (PDB ID: 1TGO)[102]. The crystallization setups were made with the C-terminally truncated protein 9°N△25 described before. 9°N-apo1 was solved in space group $P2_12_12_1$ (no. 19) with cell parameters a, b, c = 59.3Å, 111.0Å, 149.5Å and α, β, γ = 90°. The protein was modelled from residue 1 to 751 lacking the region in the thumb domain between residue 667 and 696 which is fully disordered and includes Helix U and the loops connecting it to beta sheets 25 and 26 (**Figure 3.1 A**). Disorder within the thumb domain is frequent in apo B-family polymerase structures[57, 58, 102, 105] and indicates the domain's flexibility if no DNA is bound. The overall structure shows the typical DNA polymerase domain architecture with N-terminal (N-term, 1–130 and 339–372), exonuclease (exo, 131–338), finger (448–499), thumb (591–775, 591-750 in 9°N△25) and palm domain (373–447 and 500–590) (**Figure 3.1 A**). Compared to the existing 9°N apo structure 1QHT the N-term, finger and exo domains are almost indistinguishable but the arrangement of the thumb and palm domain differ greatly. Compared to 1QHT the overall structure has an rmsd of 1.890Å with rmsds

of 0.434Å, 0.389Å, 0.406Å, 2.919Å, 1.827Å for the N-term, exo, finger, palm and thumb domains if the domains are aligned separately in Coot. In 9°N-apo1 the palm domain is rotated outwards by approximately 24° compared to its position in 1QHT. The rotation starts from a hinge point in helix P at residue Gly517 (pink in **Figure 3.1 A**). Along with helix P the C-terminal part of the palm domain up to residue 590 including β sheets 21 and 22, helix Q and β sheet 23 as well as the complete thumb domain are also positioned farther away from the N-term, exo and finger domain, generating an open DNA-binding cleft. Compared to 1WNS (apo structure of KOD) (rmsd 1WNS vs. 9°N-bin1: 0.948Å) the overall structures are more similar with well overlapping exo, N-term, and finger domains and small differences in the palm and thumb (**Figure 3.1 B**). As 9°N-bin1 shows a more open structure compared to 1QHT and this conformation is probably closer to the one occurring in solution, it was used for comparison with binary 9°N structures solved in this work.

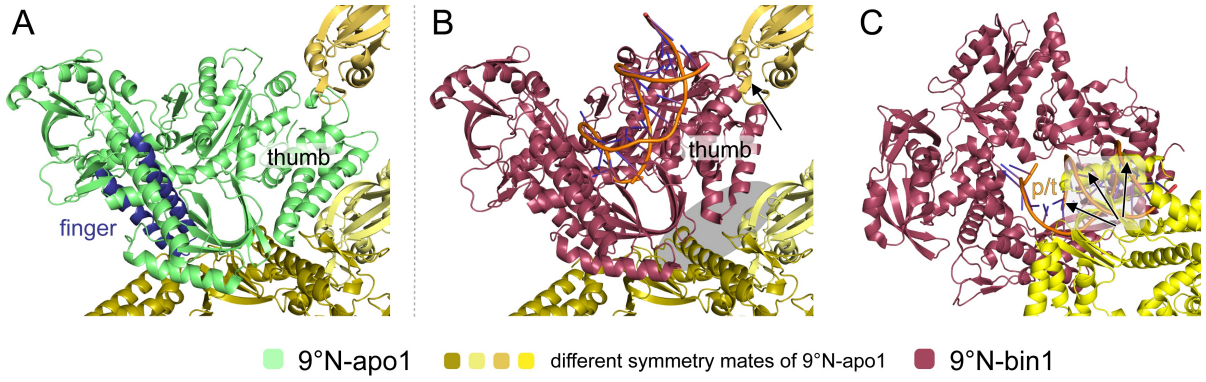


Figure 3.2: **A**: Crystal contacts of 9°N-apo1 (green). Symmetry related molecules interacting with the thumb domain are shown in different shades of yellow. The finger domain is shown in blue. **B** and **C**: The structure of a binary 9°N structure (raspberry) is shown in the same orientation as 9°N-apo1. The clashes that would occur between symmetry related molecules and 9°N in a DNA bound state are shown by arrows and the region in which crystal contacts would be released due to thumb domain movement is shaded in gray.

A possible explanation why 9°N-apo1 crystallized in an apo form and not in a binary complex although DNA was present was found when analyzing the crystal contacts. To investigate if these crystals would have formed with a bound DNA molecule the apo structure was superposed with a 9°N structure in a binary complex (described later in this work) and it was analyzed if crystal contacts are lost or if clashes of the binary structure with the symmetry mates of the apo structure would occur. The open thumb domain of 9°N-bin1 is involved in crystal contacts with three symmetry related molecules

(different shades of yellow in **Figure 3.2 A**). In a DNA bound form these crystal contacts would not be possible as the thumb domain adopts a different orientation. In this orientation it would on the one hand clash with a symmetry mate (arrow in **Figure 3.2 B**) and on the other hand would be too far away to mediate a contact formed in the apo structure (shaded area in **Figure 3.2 B**). Furthermore, an additional symmetry mate of 9°N-apo1 extends into the groove between the thumb and palm domain which is occupied by the p/t duplex in a binary complex (indicated by arrows in **Figure 3.2 C**). As the crystal form obtained seems to be well stabilized with the open apo enzyme only protein-alone molecules were included in the packing during crystal formation. If the DNA was per se only bound in some molecules in solution, although it was added in excess to the crystallization setup or if DNA dissociated from the protein so that it can fit into the crystal packing is not clear. Generally, the assumption that the apo enzyme is able to form more stable crystal contacts then the DNA-bound 9°N could explain why only apo crystals were obtained (or at least identified) in the crystallization screens prepared.

3.2.2 Structures of 9°N-apo2 and 9°N-apo3

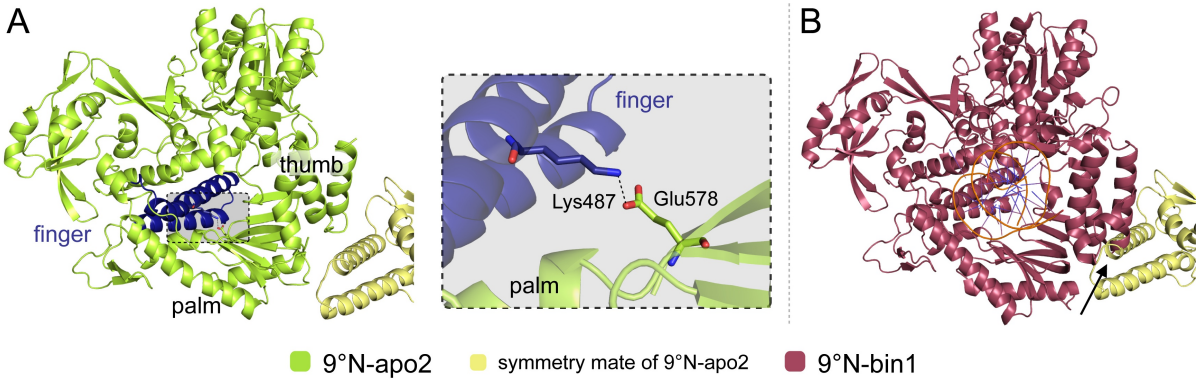


Figure 3.3: **A**: Overall structure of 9°N-apo2 is shown in limon and its finger domain is colored blue. The interaction between Lys487 and Glu578 is shown in a zoom-in view. Part of a symmetry related molecule near the thumb and palm domains is shown in paleyellow. The salt bridge keeping the palm domain in a closed state is shown in the zoom-in box. **B**: The thumb domain of a binary structure of 9°N (structure 9°N-bin1) aligned with respect to 9°N-apo2 would clash with the shown symmetry mate, explaining why no DNA is bound in this crystal form and why the enzyme shows a closed palm and thumb domain.

In addition to the just described open apo structure several additional structures have been solved with resolutions ranging from 2.5Å to less than 4Å. Two of the structures with the best resolution will be shortly described focussing on possible explanations why

	9°N-apo1	9°N-apo2	9°N-apo3
PDB ID	not published	not published	not published
Data collection			
Date of data collection	2012/04/27	2012/03/23	2012/02/03
Wavelength (Å)	1.00003	1.0000	1.0000
Space group	$P2_12_12_1$	$P2_12_12_1$	$P6_3$
Cell dimensions			
a, b, c (Å)	59.3, 111.0, 149.5	94.0, 98.1, 112.5	130.7, 130.7, 121.0
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 120
Molecules/ASU	1	1	1
Resolution (Å)	46.5-2.05 (2.17-2.05)*	48.8-1.90 (2.01-1.90)	44.4-2.80 (22.97-2.80)
I/σ=2 (Å)	2.2	2.3	3.25
No. of total reflections	414960 (66931)	549422 (84860)	296443 (47692)
No. of unique reflections	62715 (9982)	82640 (13155)	28918 (4652)
R_{meas} (%)⋆	8.4 (194.0)	15.5 (539.6	16.4 (532.0)
I/σ	14.86 (1.10)	9.00 (0.33)	12.42 (0.45)
Completeness (%)	99.9 (99.8)	99.9 (99.6)	99.9 (99.8)
Redundancy	6.6 (6.7)	6.5 (6.5)	10.3 (10.3)
$CC_{1/2}$ (%)	99.9 (55.8)	99.9 (17.2)	99.9 (12.4)
Refinement			
Resolution (Å)	46.5-2.05	48.8-1.90	44.4-2.8
No. of reflections	42668	82271	28917
R_{work}/R_{free}	20.1/23.5	21.8/25.9	25.6/31.0
Coordinate error (Å)[a]	0.29	0.45	0.65
No. of atoms			
Protein	6756	11775 (with H)	5646
Water	243	200	0
B-factors			
Protein	74.7	73.6	125.1
Water	60.4	52.3	
R.m.s deviations			
Bond lengths (Å)	0.003	0.007	0.003
Bond angles (°)	0.774	1.005	0.801
Ramachandran (%)[b]			
Favored	97.4	97.04	90.9
Allowed	2.36	2.68	7.78
Outlier	0.00	0.28	1.32

Table 3.1: Summary of the data collection and refinement statistics for 9°N apo structures. Data processing and refinement files are stored in the following folders:
9GN-apo1: /home/kbetz/SLS2012/apr27/9GN-14mer-5_2/
9GN-apo2: /home/kbetz/SLS2012/mar23/9gn-2_2
9GN-apo3: /home/kbetz/SLS-2012/feb03/9GN-11mer-t6
* Values in parentheses correspond to those in the outer resolution shell.
⋆ For definition of R_{meas} see [200].
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

no DNA was found in the crystals although it was present in the setup. The first of these structures (termed 9°N-apo2) was obtained with the 9°N△25 protein. It resulted from

crystals with space group $P2_12_12_1$ (no. 19) and the following cell parameters: a, b, c = 94.0Å, 98.1Å, 112.5Å and α, β, γ = 90°. This is the same crystal form as the one of the previously published structure 1QHT. 1QHT and 9°N-apo2 are very similar with a rmsd of only 0.428Å (**Figure 3.3**). The same salt bridge as in 1QHT (section 1.2.3.2) traps the enzyme in the observed conformation with a closed thumb and palm domain. In 9°N-apo2 a more open conformation of the thumb domain (as observed in 9°N-bin1) would not be possible as a crystal contact with the finger domain of a neighbouring molecule would be disturbed (**Figure 3.3 B**). Besides the salt bridge, this is another probable explanation why in this crystal form the enzyme appears in a more closed palm and thumb conformation although no DNA or triphosphate is bound.

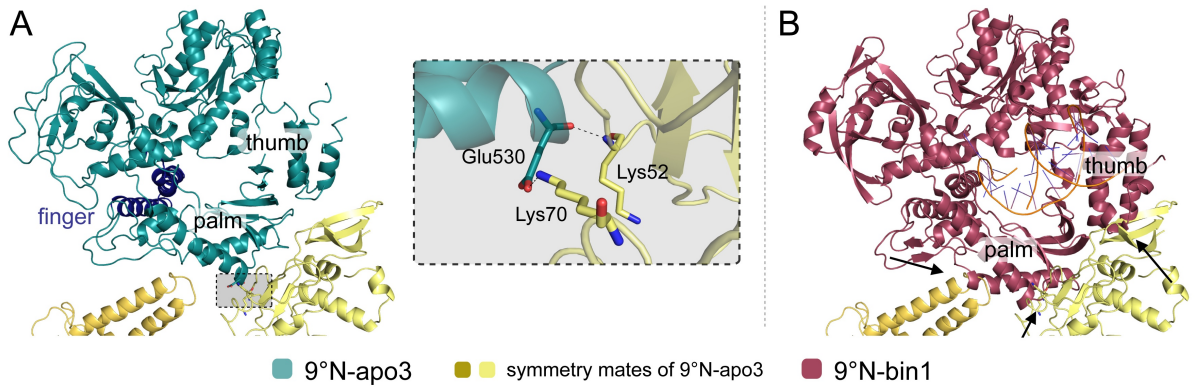


Figure 3.4: A: Overall structure of 9°N-apo3 is shown in deepteal and its finger domain is colored blue. Parts of two symmetry related molecules are shown in gold and paleyellow, respectively. The crystal contact described in the text is shown in a zoom-in view. **B**: The thumb and palm domains of a binary structure of 9°N aligned with respect to 9°N-apo3 would clash with the shown symmetry mates explaining why no DNA is bound in this crystal form. The clashes are indicated by black arrows.

The next structure (termed 9°N-apo3) was solved in space group $P6_3$ (no. 169) with cell axis a, b, c = 130.7Å, 130.7Å, 121.0Å and angles α, β, γ = 90°, 90°, 120°. Similar to 1QHT and 9°N-apo2 the enzyme adopts a closed state, with respect to the palm and thumb domain (rmsd of 9°N-apo3 and 1QHT is 1.526Å). This conformation is again stabilized by the crystal lattice. In a crystal contact of the palm domain the carbonyl of Glu530 can form a H-bond to the backbone of Lys52 whereas its side chain can form a H-bond to the backbone of Lys70 (**Figure 3.4 A**). A fully open conformation as observed in 9°N-apo1 would not be possible in that crystal form. Furthermore, molecules with bound DNA could not be included into the lattice, as the positions of thumb and palm domains as adopted in the binary structure described later would clash with parts of the N-terminal

domain of one symmetry mate and the finger domain of another (indicated by black arrows in **Figure 3.4 B**).

3.2.3 Discussion and Conclusion

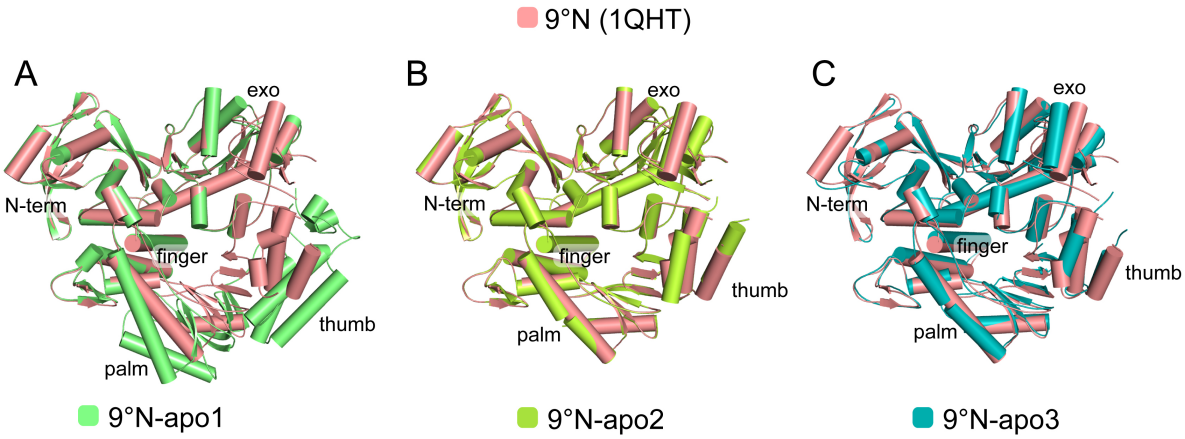


Figure 3.5: Structures of the three reported apo 9°N structures in overlays with the previously published closed apo 9°N structure 1QHT[57].

Although it was aimed to solve binary or ternary 9°N structures in this work, also the obtained apo structures were analyzed. Thereby the structure 9°N-apo1 showed a fully open enzyme state (**Figure 3.5 A**). In contrast to the available 9°N structure 1QHT, 9°N-apo1 is similar to other apo B-family polymerase structures and therefore provides a much better model for the enzyme in an uncomplexed state and can be used in future comparison with binary or ternary structures. Besides 9°N-apo1, two structures in which the thumb and palm domain adopt a closed conformation were obtained (**Figure 3.5 B** and **C**). This circumstance indicates that the two domains are flexible in the apo enzyme and crystal contacts can lead them to be trapped in different conformations. Furthermore, the situation also recalls that crystal contacts can lead to conformations within proteins that do not occur in solution and therefore the analysis of their influence on the structure is an important task. In all three apo 9°N structures described, enzyme molecules bound to DNA would not have been able to be included in the crystal lattice as their different conformation would have led to clashes. Compared to the fully open 9°N-apo1 structure the binary structure (described later in this work) shows a more closed thumb domain and the bound DNA would interfere with a neighbouring molecule. Compared to the 9°N-apo2 and 9°N-apo3 structures with closed palm and thumb domains the binary enzyme shows

a more open palm and thumb domain which would also clash with respective symmetry mates.

3.3 Crystallization of 9°N in complex with DNA

As 9°N readily crystallized as apo enzyme in many different conditions although DNA was present it was searched for a possibility to label DNA containing crystals allowing to visually distinguish between crystals containing DNA and apo 9°N crystals. In a method described by Jiang & Egli[219] Cy-labelled nucleic acid is used to discriminate between apo and protein-DNA co-crystals. Therefore I decided on using 5' Cy3- or Cy5-labelled templates in complex with unmodified primers in the crystallization experiments with 9°N. The cyanine dyes Cy3 and Cy5 were chosen to be attached to the 5' template position as the overhanging template is pointing into the solvent and is flexible, whereas the opposite side of the p/t duplex is often involved in crystal contacts. Crystal screening was started using templates of two different length: a 12- and a 16mer oligonucleotide, annealed to a 7- or 11mer primer, respectively (**Figure 3.6**) together with the purified full length 9°N and the addition of ddCTP. In subsequent experiments I also screened for crystals using a 17mer/12mer template/primer complex and the previously used 16mer with primers shortened by one or two nucleotides resulting in a 3' template overhang of a single dG nucleotide or a dGdC dinucleotide. The sequences of the p/t complexes used are shown in **Figure 3.6** and are named according to their template and primer length separated by a slash plus the attached dye in subscript letters.

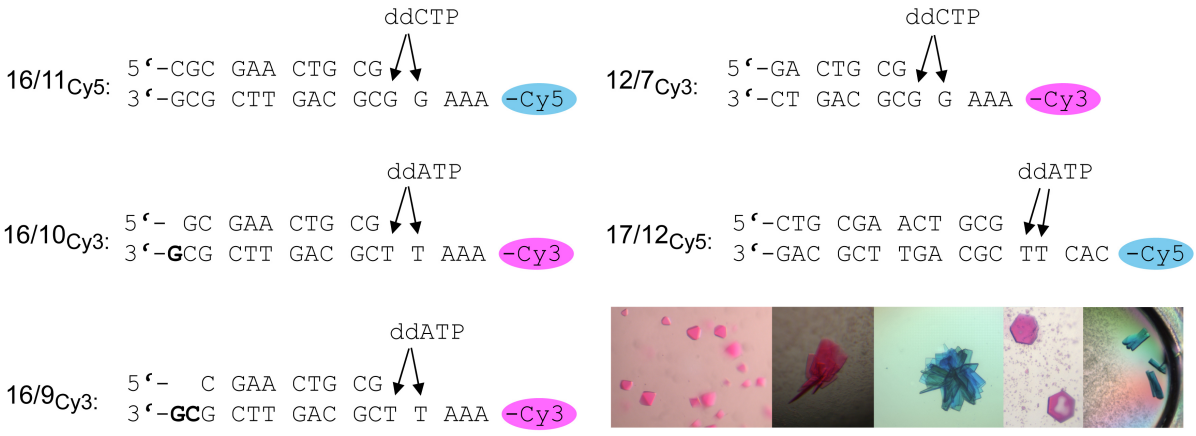


Figure 3.6: Cy-labelled p/t constructs used to screen for binary 9°N crystals and pictures of some hits obtained with the labelled templates.

Using the above-mentioned approach colored crystals were obtained under different crystallization conditions and with all different p/t constructs. Images of example hits are shown in **Figure 3.6**. Crystals grown in the screening plates were usually directly tested for diffraction if they were large enough to harvest them and crystallization hits were then tried to be reproduced and optimized towards better diffraction. Several datasets have been recorded and structures of 9°N could be solved by MR using the apo structure 1QHT. Most of the structures obtained from the colored crystals also revealed the presence of DNA in the electron density, however not always completely resolved. In contrast, in some structures DNA was not visible and the color in the crystals most probably resulted from unspecific binding of DNA somewhere in the crystal lattice. Although most of the crystallization setups were made with the aim to yield a ternary complex, the enzyme only crystallized in a binary form. Despite the presence of magnesium ions and ddNTP the structures never revealed a bound substrate triphosphate and/or bound metal ions near the active site. Though unable to obtain a ternary structure, yielding a binary complex of 9°N still was a great success as no binary structure of an archaeal B-family DNA polymerase was published to that date. The first binary structure obtained in this study that contained a throughout well ordered p/t complex will be described in more detail in the next chapter.

3.4 Crystal structures of 9°N in complex with DNA

3.4.1 Structure of 9°N-bin1

The first well resolved binary 9°N structure obtained (termed 9°N-bin1) was solved simultaneously with a binary structure of KOD in our group. The work on KOD polymerase was done by my colleague Konrad Bergen and the two binary structures were published together[193]. The crystals of 9°N-bin1 grew in space group $P22_12$ (no. 18) with cell dimensions a, b, c = 112.2Å, 142.6Å, 66.7Å and $\alpha, \beta, \gamma = 90°$. Data collection and refinement statistics are shown in **Table 3.2**. The setups for these crystals were made with the $16/11_{Cy5}$ oligonucleotide duplex shown in **Figure 3.6** and the addition of ddCTP. The I/σ=2 resolution of the structure is 2.6Å.

Protein and DNA overall structure. The overall structure of 9°N-bin1 shows the typical domain composition of DNA polymerases, including finger, thumb and palm domain.

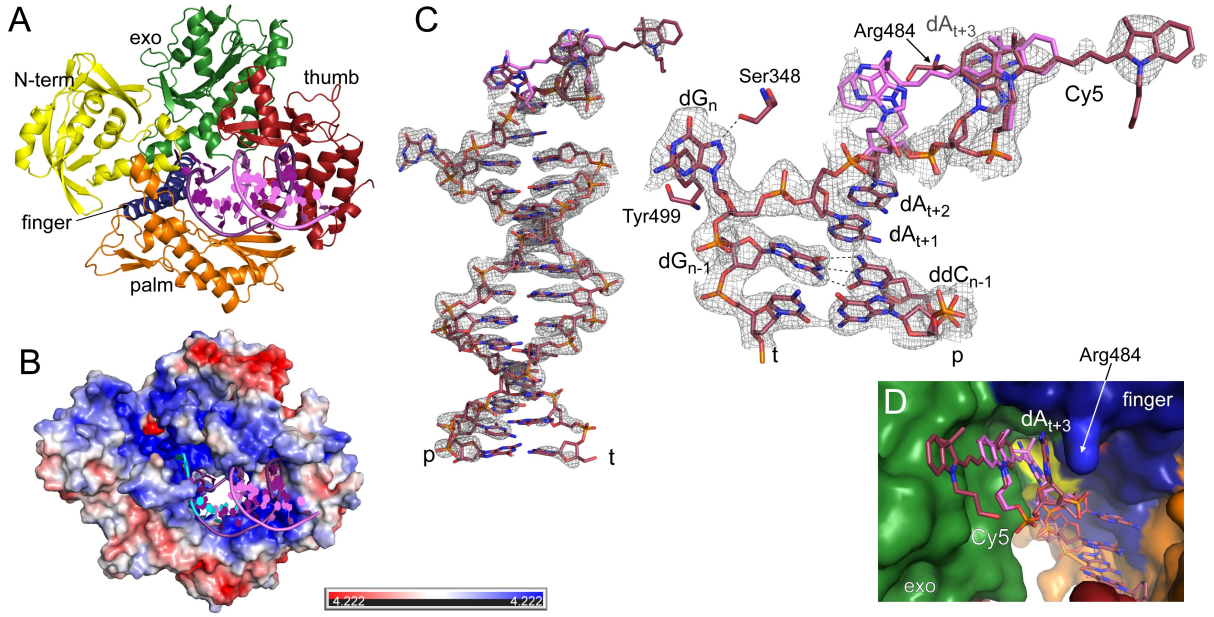


Figure 3.7: **A**: Overall structure of 9°N-bin1 with domains colored as in Figure 3.1. The primer is shown in violet and the template is shown in purple. **B**: Same orientation as in A but the electrostatic surface of 9°N is shown with a gradient from positive (red) to negative (blue) potential. Electrostatic surface was created using the APBS tool in PyMOL. **C**: The Cy5 labelled p/t complex bound by 9°N-bin1 (raspberry) is shown surrounded by a mFo-DFc simulated annealing omit map contoured at 3σ. In the zoom-in view (right) the omit map is contoured at 2σ. The alternative conformations of Cy5 and dA_{n+2} with the lower occupancies are colored violet. **D**: Position of the Cy5 dye surrounded by the protein shown as surface from a view 180° rotated with respect to the orientation in panel A.

More specifically, 9°N-bin1 consists of the following parts: N-terminal (N-term, 1–130 and 339–372), exonuclease (exo, 131–338), finger (448–499), thumb (591–775) and palm domain (373–447 and 500–590). The protein was modelled from residue 1 to 759 lacking the last 23 C-terminal residues which were already known to be relatively flexible[57]. The enzyme adopts an open conformation with respect to the finger domain. The DNA is bound in a groove formed by the thumb and palm domains which is equipped with charged residues generating a positive potential responsible to bind the negatively charged DNA backbone (**Figure 3.7 B**). The Cy5 dye attached to the 5' template end is visible in the electron density. A simulated annealing omit map of the complete DNA is shown in **Figure 3.7 C**. The four ss nucleotides at the 5'-end of the template show an interesting arrangement. The templating nucleotide dG_t is flipped out of the stacking arrangement with the nascent base pair and now packs with Tyr499 at the very N-terminal end of the finger domain and is hydrogen bonded by Ser348 of the N-terminal domain (**Figure 3.7 C**, zoom-in view). The upstream nucleobases dA_{t+1} and dA_{t+2} are flipped back towards the DNA duplex and stack on top of the nascent base pair. The last nucleotide dA_{t+3}

	9°N-bin1	9°N-bin2 (R465A)	9°N-bin3 (R465E)
PDB ID	**4K8X**	not published	not published
Data collection			
Date of data collection	2012/07/05	2012/12/07	2013/02/05
Wavelength (Å)	0.97793	1.00000	1.00001
Space group	$P2_12_12$	$P6_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	112.21, 142.62, 66.70	185.8, 185.8, 122.3	110.1, 111.9, 169.1
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Molecules/ASU	1	2	2
Resolution (Å)	48.7-2.28 (2.42-2.28)*	49.1-2.69 (2.86-2.69)	47.8-2.60 (2.75-2.60)
I/σ=2 (Å)	2.6	3.0	2.9
No. of total reflections	330620 (54378)	684270 (103421)	403136 (44457)
No. of unique reflections	49600 (7914)	66252 (10454)	63896 (9116)
R_{meas} (%)⋆	17.2 (282.1)	17.4 (548.4)	8.8 (237.1)
I/σ	9.07 (0.82)	11.62 (0.48)	13.78 (0.58)
Completeness (%)	99.9 (99.9)	99.6 (97.9)	98.1 (88.1)
Redundancy	6.7 (6.9)	10.3 (9.9)	6.3 (4.9)
$CC_{1/2}$ (%)	99.9 (49.2)	99.9 (29.8)	99.9 (25.8)
Refinement			
Resolution (Å)	48.7-2.28	29.1-2.70	47.8-3.6
No. of reflections	49542	45928	63802
R_{work}/R_{free}	19.5/23.8	18.4/24.7	20.9/25.5
Coordinate error (Å)[a]	0.38	0.48	0.47
No. of atoms			
Protein (A/B)	6252	6172/6123	6175/5645
DNA (A/B)	594	509/488	589/487
Water	134	175	270
B-factors			
Protein (A/B)	60.8	86.4/96.2	98.5/117.9
DNA (A/B)	73.8	83.7/88.0	109.7/142.1
Water	52.0	71.1	85.8
R.m.s deviations			
Bond lengths (Å)	0.008	0.003	0.003
Bond angles (°)	1.16	0.707	0.739
Ramachandran (%)[b]			
Favored	97.24	96.07	96.70
Allowed	2.63	3.79	2.89
Outlier	0.13	0.13	0.41

Table 3.2: Summary of the data collection and refinement statistics for 9°N binary structures. Data reduction and refinement files are stored in the following folders:
9GN-bin1: /home/kbetz/SLS2012/jul05/9GN-Cy5-soak-3_run1/2-3/2-1/4K8X
9GN-bin2: /home/kbetz/SLS2012/dec07/9GN-2nt-dec-2
9GN-bin3: /home/kbetz/SLS2013/feb05/13-Lena-2
* Values in parentheses correspond to those in the outer resolution shell.
⋆ For definition of R_{meas} see [200].
[a] maximum likelihood based (as determined by PHENIX[197])
[b] as determined by MolProbity[201]

stacks against Arg484 of the finger domain whereby the attached Cy5 dye interacts with the exonuclease domain (**Figure 3.7 D**). The dye points towards a symmetry mate where it is placed between Arg99, His103 and Pro104 (described in detail later). Due to this arrangement the 5'-overhanging template seems to be stabilized but in a conformation probably not relevant in solution for longer templates though favourable for crystal formation in our setup. In proximity to the nucleotide dA_{t+2} positive difference density was observed which could be explained by alternative conformations of residues dA_{t+2} and Cy5.

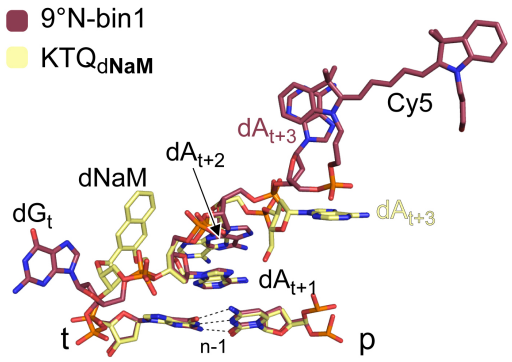


Figure 3.8: Comparison of the ss template overhang of 9°N-bin1 and $KTQ_{\mathbf{dNaM}}$. Structures were aligned by superposition of the nascent base pair. T+1 and t+2 template residues are arranged similarly.

A second conformation of dA_{t+2} was modelled by rotating the nucleotide approximately 150° around the C3' sugar atom and also a second conformation of Cy5 was modelled rotating it by approximately 180° around the aliphatic chain connecting it to dA_{t+3}. The alternative conformations are shown in violet in **Figure 3.7 C**. The arrangement of the 5' ss template part observed in 9°N-bin1 is similar as was found for KlenTaq binary structures $KTQ_{\mathbf{dNaM}}$ and KTQ_{dT} with a 5'-d(AAA**NaM**)-3' or 5'-d(AAAT)-3' template overhang, respectively, described in chapter 2.1.2. An overlay of the template overhang of 9°N-bin1 and $KTQ_{\mathbf{dNaM}}$ is shown in **Figure 3.8**. As was already stated for the two mentioned KlenTaq structures this arrangement is believed to be only relevant in the crystal structure and is not standard in solution, especially if the ss template overhang is longer.

As was observed in the extensively studied B-family polymerase RB69, the p/t duplex of 9°N binary structure maintains a B-form DNA conformation, with most ribose moieties showing the ideal puckering for B-DNA (C2'-endo) or conformations that are also found in more flexible B-DNA (C1'-exo, C3'-exo, O4'-endo and C4'-exo conformations)[18]. The DNA colored by sugar conformation is shown in **Figure 7.1 B** in the Appendix on page

181. The DNA duplex also follows other known geometric characteristics for B-DNA. As already stated for other binary or ternary B-family polymerase structures, this is in contrast to the A-form DNA observed near the insertion site in A-family polymerases, as reported for T7, BF and KlenTaq DNA polymerases[71, 129, 130].

Conformational changes upon DNA binding. To evaluate changes within the enzyme upon DNA binding 9°N-bin1 is compared to the open apo structure 9°N-apo1 solved during this work. Thereby several rearrangements of the enzyme induced by the binding to DNA became apparent. The most pronounced movements are accomplished within the thumb and palm domain which is also indicated by the rmsd values for the superposition of the individual domains of the apo and the binary structure. Whereas 9°N-bin1 and 9°N-apo1 align with a rmsd of 1.461Å for the complete enzyme the rmsd is only 0.397Å, 0.368Å and 0.344Å for the N-term, finger and exo domain, respectively and explicitly higher for the thumb (1.743Å) and palm domain (1.252Å), respectively.

Compared to 9°N-apo1 the biggest movements of the enzyme while binding the p/t

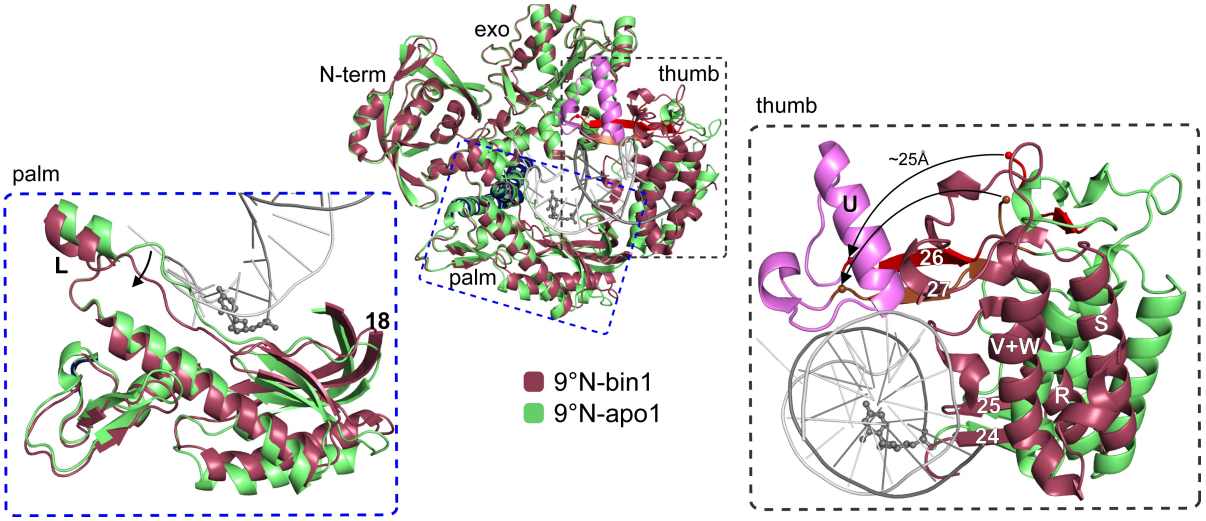


Figure 3.9: Overlay of 9°N-bin1 (raspberry) with 9°N-apo1 (green). The palm, thumb, exo and N-terminal domains are labelled and the finger domain is located in the back, shown in blue. The U-helix and the loops connecting it to β sheets 26 and 27 are shown in violet for 9°N-bin1 and are almost completely disordered in 9°N-apo1. Zoom-in views for the thumb and palm domains are given. The movements within the two domains upon transition from the apo polymerase to the binary complex are indicated by black arrows. The primer-terminal nucleotide ddC is shown in ball-and-stick representation.

duplex are made by the thumb domain (**Figure 3.9**). Upon closing, the thumb domain moves in one entity and the previously unresolved helix U (violet in **Figure 3.9**) gets ordered together with the loops connecting it to beta sheet 26 and 27. Moreover, the

beta sheets 26 and 27 themselves get better structured (red and brown in **Figure 3.9**). Residues within the additionally resolved region contact the primer strand. The Cα atoms of the amino acids at the end of β sheets 26 and 27 are approximately 25Å away from each other in the open and closed state. Also helices R, S and V+W are moved towards the p/t duplex. Thereby residues of helix V+W and beta sheets 24, 25 and the loop connecting beta sheet 25 and helix R engage contacts with the p/t duplex. The DNA duplex is bound in the range of primer residues p-7 and ddC and template residues t-2 and t-9 (**Figure 3.10**). The unbound DNA duplex has higher B-factors than the duplex DNA bound by the enzyme with average B-factors above 100($Å^2$) for primer nucleotides p-8 to p-12 and template nucleotides t-11 and t-12.
Besides the thumb domain also rearrangements are observed within the palm domain. Here, especially helix L (residue 374–379), located at the interface to the finger domain) and the loop connecting it to β-sheet 18 (residues 397–404) is affected by DNA binding. Starting at Val 389, the loop together with the helix moves by up to 4Å from its former position, thereby opening the cleft into which the DNA binds.
Side note: The helix V+W appears as one helix here, whereas it is separated in two parts in 1QHT and was named V and W, respectively by Rodriguez *et al.*

Protein-DNA contacts. The DNA duplex is mainly contacted by residues of the thumb domain, as well as some residues of the palm domain and one of the exonuclease domain. In **Figure 3.10** the main chain (solid rectangles) and side chain interactions (rectangle without lining) of 9°N with the DNA duplex are visualized. Residues interacting with both, main and side chain are shown in dashed rectangles. The 5' ss nucleotides were not included in the analysis. Direct contacts via H-bonds up to 3.2Å between donor and acceptor are shown with black arrows and contacts via a water molecule are indicated by red arrows. Furthermore, interactions with the nucleobases, ribose moiety and phosphate backbone are shown in brown, gray and green, respectively. 9°N, similar to other B-family polymerases, contacts the DNA duplex mainly via the phosphate backbone whereby only very few interactions with the nucleobase or sugar moieties are observed (**Figure 3.10**). From the 5' to the 3' side, the primer is first contacted by residues in the loop connecting β sheet 26 and helix U (residues 666–677) followed by phosphate and minor groove interactions with residues of a second loop between β sheet 25 and helix R (606–616). The template is contacted at its 3' side by residues of the connecting loop between β sheets 27 and 28 (709–711) and by residues in helix V+W (729–745). Some residues known to

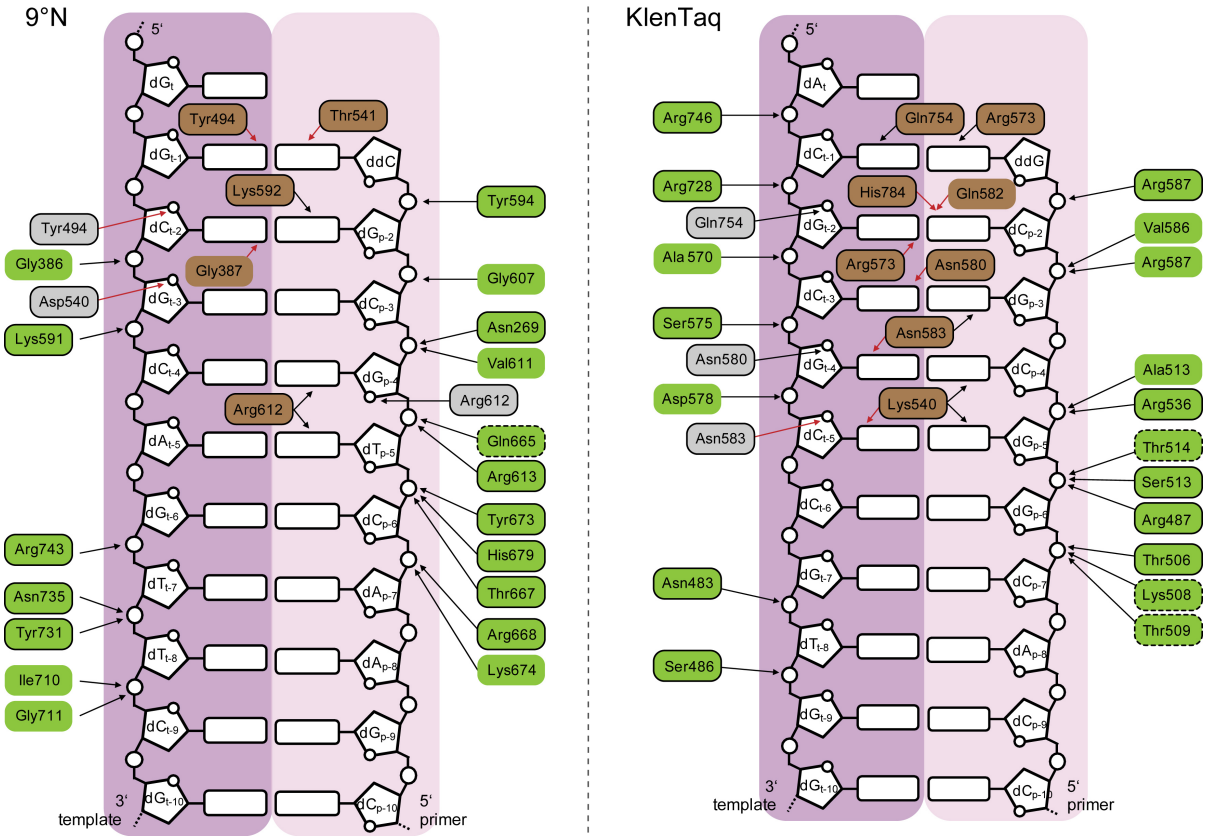


Figure 3.10: Comparison of protein-DNA interactions in 9°N-bin1 and KTQ_{dG} (PDB ID: 3SZ2[44]). Description see text.

be important in DNA binding are situated in motifs conserved throughout family B DNA polymerases. One example is the sequence motif KKRY (KKKY in 9°N, residues 591–594, **Figure 3.11 A**), which is unique to family B DNA polymerases[35] and mediates contacts to the p/t duplex near the active site. It is located in the loop between β sheet 23 and 24 including part of β sheet 24. In 9°N the side chain of Lys591 within the motif directly contacts the phosphate of $dC_{t\text{-}4}$ in the template, Lys592 can form a H-bond with the nucleobase of $dG_{p\text{-}2}$ of the primer and Tyr594 interacts with the phosphate moiety of the primer terminus (**Figure 3.11 A**). In the related RB69 DNA polymerase similar interactions of the residues in the KKRY motif with the DNA are observed[91]. A second conserved region (I/YxGG/A motif[35], residues 384-387, **Figure 3.11 B**) is situated in the extended loop of the palm domain connecting helix L and beta sheet 19 and interacts with the template strand at the 5'-end of the duplex . In 9°N the motif is YAGG whereas in RB69 it is YPGA. The backbone nitrogen of Gly386 interacts with the phosphate of $dG_{t\text{-}3}$ and Gly387 interacts via a water molecule with the minor groove side of the template

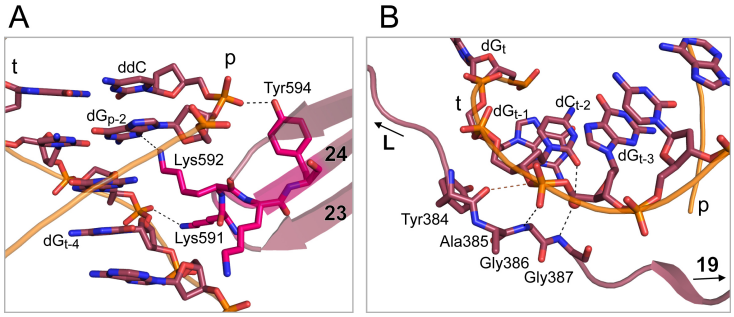


Figure 3.11: Interactions of 9°N-bin1 with the bound p/t duplex. **A:** Interactions within the KKRY motif described in the text. **B:** Interactions of the I/YxGG/A motif described in the text.

nucleobase t-2 (**Figure 3.11 B**). Apart from those in conserved motifs residues involved in DNA binding are not conserved throughout the B-family DNA polymerases[61]. However, we found that the great majority of residues that directly contact the DNA duplex are indeed conserved well in Euryarchaeota and in part also in other phyla of the Archaea. To visualize that, a sequence alignment was made using six *Thermococcus* (9°N, KOD, Tgo, Tfu, Thy, Tli), four *Pyrococcus* (Pfu, Pab, Pwo, Deep Vent), one *Desulfurococcus* (D.Tok) and two *Sulfolobus* (Sac, Sso) members. The *Sulfolobus* polymerases have longer sequences and are more distantly related to the rest of the aligned archaeal B-family polymerases. The sequence identities of the respective archaeal B-family members and 9°N is given in **Table 7.5** in the Appendix on page 180. The following residues interacting with the DNA in 9°N-bin1 are identical in the aligned family members or the difference is given in brackets: Asn269 (Lys in *sulfolobus* members), Tyr494, Asp540, Thr541, Lys591, Lys592, Tyr594, Gly607, Val611 (Lys in *sulfolobus*), Arg612 (Lys in *sulfolobus*), Arg613, Gln665 (Met in *sulfolobus*), Thr667 (Ser in *sulfolobus*), Arg668 (Lys in *sulfolobus*), Tyr673, Lys674 (can be Arg or Thr), His679, Ile710 (Val in D.Tok, absent in *sulfolobus*), Gly711 (can be Ser), Tyr731 (Lys in *sulfolobus*), Asn735 (Thr or Ala in *sulfolobus*) and Arg743 (Gln in *sulfolobus*). Gly386 and 387 which interact via their backbone are variable. The alignment is shown in **Figure 7.5** on page 183).

Mismatch sensing in the minor groove. Some interactions of 9°N with the minor groove of the newly formed product DNA have already been described above. An additional interaction pattern involving water molecules and responsible for mismatch sensing is described in detail here. For RB69 DNA polymerase (1.8Å ternary structure, PDB ID: 3NCI) it has been shown that a network of five water molecules which are stabilized by the surrounding protein side chains, interact with the N3 and O2 atoms in the minor groove of the DNA duplex with all four nucleobases of the two base pairs at the n-1 and n-2 position[78] (**Figure 3.12 B**). Sensing of the minor groove N3 and O2 atoms

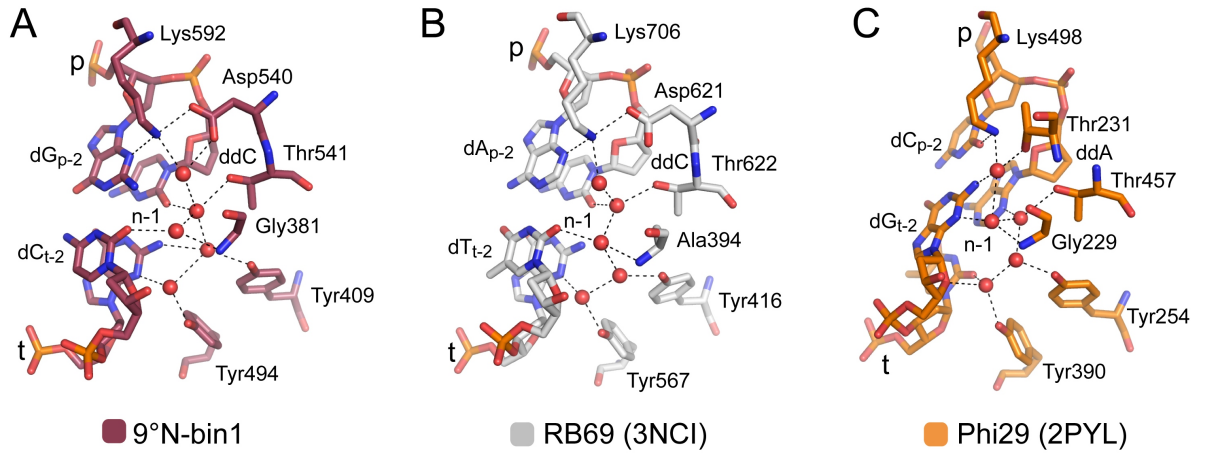


Figure 3.12: Water mediated network of protein residues interacting with the minor groove side of the n-1 and n-2 base pairs in **A**: 9°N-bin1, **B**: RB69 ternary complex and **C**: Phi29 ternary complex.

of purines and pyrimidines, respectively, is a sequence independent way to discriminate between correct and incorrect Watson-Crick base pairs. Wang *et al.* state that this interaction pattern is disturbed to some minor or major extent in case of a misincorporation, decreasing the chance of elongation and promoting the polymerase to initiate repair mechanisms. As analogous interaction patterns are observed in polymerase δ and phi29 ternary structures[100, 221] (**Figure 3.12 C**) is likely to be general in B-family DNA polymerases. In 9°N-bin1, indeed, also a similar pattern is observed and the analogous protein residues as in RB69 are involved in the network. The different connections between nucleobases, waters and amino acids will be shortly described. The five water molecules form a network with highly conserved protein residues Asp540, Lys592, Thr541, Gly381 and Tyr409. The first water (numbered from up to down in **Figure 3.12**) bridges the conserved Asp540 (Asp621 in RB69) and Lys592 (Lys706 in RB69) which in turn directly H-bonds to the p-2 nucleobase. The second water interacts with the terminal primer nucleotide (ddC, n-1) and is stabilized by Thr541 (Thr622 in RB69). The third water connects the t-2 nucleobase and the amide backbone of Gly381 (Ala394 in RB69) while the fourth water interacts with the amino group of guanine at position t-2 and is itself bound by Tyr409 (Tyr416 in RB69). The fifth water lies between and connects the hydrogen bonding acceptor of the nucleobase of t-2, its ribose oxygen and a second tyrosine residue, Tyr494 (Tyr567 in RB69). In RB69 the main difference is that all waters are connected in a row whereas in 9°N water 2 and 4 are connected and 3 is only hydrogen bonded to 2. The fact that not only in RB69, phi29 and pol δ the same interaction patterns of protein residues with nucleotides near the p/t junction are observed, but also in the archaeal polymerase

9°N, indicates that the mechanisms contributing to selectivity in B-family DNA polymerases are conserved throughout the family.

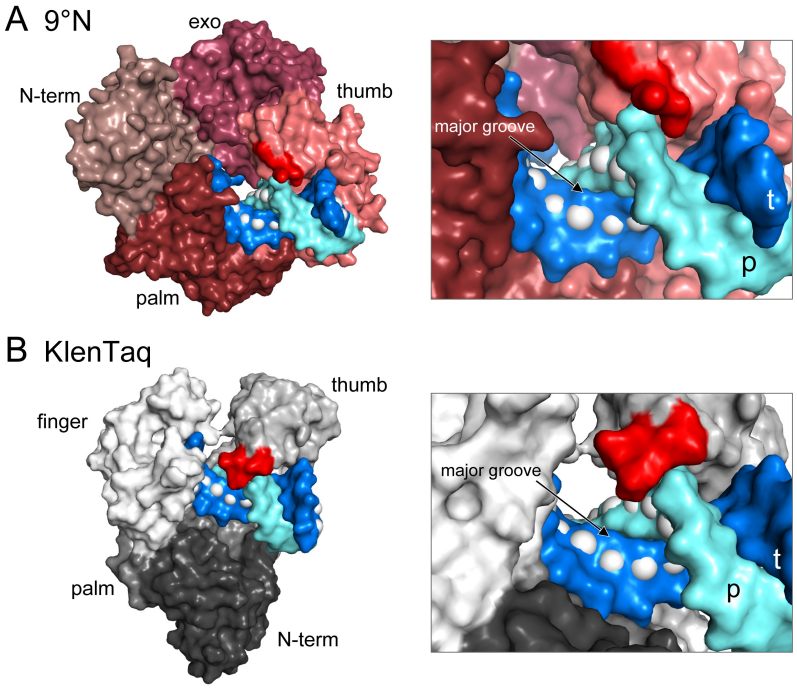


Figure 3.13: Comparison of DNA environment in family B and family A DNA polymerases. **A:** 9°N and **B:** KlenTaq (PDB ID: 3SZ2) binary complexes are shown as surface with the different domains colored in different shades of red/brown and gray, respectively. The primer and template is shown as marine and cyan surface, respectively, and the tip of the thumb domain is pointed out in red. Right panels show details of the primer and template binding clefts. The C5 of pyrimidine and C7 of 7-deazapurine nucleobases which are mainly used for the attachment of modifications are marked as white spheres.

Comparison of 9°N and KlenTaq. As already mentioned in section 1.2.2.3 several archaeal B-family polymerases show improved incorporation properties for chemically modified nucleotides over A-family polymerases [7, 14, 16, 50–56]. With the structural data obtained in this work it is tried to elucidate these observed differences. Therefore, the binary structure of 9°N is compared to KlenTaq as a representative family A polymerase using the KlenTaq binary structure (KTQ_{dG}) described in section 2.1.2 (PDB ID: 3SZ2)[44]. In both enzymes the majority of direct protein DNA duplex contacts are mediated via the phosphate backbone (**Figure 3.10**). However, the number of contacts (direct or water-mediated) to the nucleobases and sugar oxygens vary considerably. In KlenTaq ten nucleobases are contacted by seven different protein side chains, whereas in the 9°N structure only six nucleobases are contacted by just five residues. Interestingly, in both

enzymes a single protein side chain simultaneously contacts the two nucleobase moieties of the p-4 and p-5 primer nucleotides (Lys540 in KlenTaq, Arg612 in 9°N). In addition to the lower amount of nucleobase contacts also fewer direct interactions with the ribose moiety were found in 9°N compared to KlenTaq. In KlenTaq two direct interactions of sugar oxygens with protein side chains are observed (Gln754 and Asn580), for 9°N only one residue (Arg612) interacts with a ribose moiety. However, this is rationalized by the interactions via water molecules in both structures leading to a sum of three ribose interactions in each structure. Note that counting the number of interactions always depends on the cut-off one chooses for an interaction event. In contrast to the analysis here, in the publication of the 9°N and KOD binary structures[9] all residues within 3.6Å that could form a H-bond were included for comparison and under this assumption also a significantly higher amount of direct ribose interacting residues is found in KlenTaq compared to 9°N.

Based on the just described observations one would expect a sterically less hindered minor groove in the case of B-family polymerases, as all sugar and nucleobase contacts are located in the minor groove. The steric space of the minor groove has experimentally been probed by modifying nucleotides at the C4' position of the ribose moiety. Indeed it was shown that B-family polymerases show better performance than an A-family polymerase in utilising nucleotides that bear small C4' modifications (i.e., methyl, ethyl, vinyl, methylene, etc.) which is in accordance to the presented structural results[52]. Nevertheless, bulkier modifications might not be tolerated by either enzyme as these would interfere with residues (of both proteins) that are further away from the interior of the minor groove. In general, in polymerase/DNA complexes the DNA minor groove is not as well accessible as is the major groove. Therefore bulky modifications are preferentially attached to the C5 position in pyrimidines and C7 of 7-deazapurines which point into the major groove[7–17] and increase the probability that they are well accepted by polymerases. Kinetic studies revealed that nucleotides bearing modifications pointing to the major groove are as well more efficiently polymerized by family B DNA polymerases compared to A-family members[7][14, 16, 53–56]. For the 9°N and KlenTaq binary structures the major groove of the DNA seems to be well accessible in both enzymes (**Figure 3.13**). As mentioned above the number of residues interacting with the phosphate backbone which can extend into the minor and major groove is similar in 9°N and KlenTaq. However, one major difference is the structure of the tip of the thumb domain (red in

Figure 3.13). In both enzymes the tip of the thumb domain (residues 506–509 in KlenTaq, 668–675 in 9°N) makes contacts with the primer strand but the architecture of the contacting loops differs greatly. Whereas the loop in the 9°N structure is positioned above the minor groove without extending deeply into it (**Figure 3.13 A**), the corresponding loop in KlenTaq extends over the phosphate backbone towards the major groove (**Figure 3.13 B**). The better accessibility of the major groove, at least in the duplex region close to the thumb tip, might explain the observed superiority of family B DNA polymerases in polymerizing nucleotides that bear bulky modifications at C5 in pyrimidines and at C7 in 7-deazapurines.

Comparison with Pfu binary complex. Recently the structure of a mutant of the hyperthermophilic archaeon *Pyrococcus furiosus* (Pfu) replicative DNA polymerase has been solved in a binary complex with DNA (PDB ID: 4AIL)[220]. The Pfu variant carrying eight mutations is named Pfu-E10 and is capable of efficiently incorporating cyanine dye-labelled dCTP. The binary structure is very similar to our binary structure 9°N-bin1 with a rmsd of 1.088Å. The interactions of the DNA duplex with the respective enzyme are similar in 9°N and Pfu-E10 binary complexes. Especially for the residues directly contacting the minor groove (Arg612 and Lys592 in 9°N and Arg613 and Lys593 in Pfu). The residues corresponding to the ones interacting with the phosphate backbone in 9°N are also present in Pfu, however some of them show slightly larger interaction distances. Furthermore, Pfu-E10 shows an additional H-bond with the phosphate group of the p-2 nucleotide via Ser384 where 9°N, in contrast to many other archaeal B-family members, has a glycine in the sequence and can not form that contact. The water-mediated minor groove H-bond of the t-1 nucleotide with Tyr494 observed in 9°N is also present in Pfu-E10 whereas the p-1 interaction with Thr541 via a bridging water molecule is not observed. This might however be just due to the lower resolution of the structure in which the water molecule is not sufficiently ordered. Overall, the similarities in 9°N and Pfu-E10 binary structures indicate common mechanisms in archaeal family B polymerases and approve that transfer of information gained for one enzyme to other archaeal members from which no binary structures exist is well possible.

DNTP binding and catalysis. As no ternary structure of archaeal B-family DNA polymerases exist our binary structure is compared with the available ternary structures, especially of the related bacteriophage polymerases RB69 and phi29 to disclose similari-

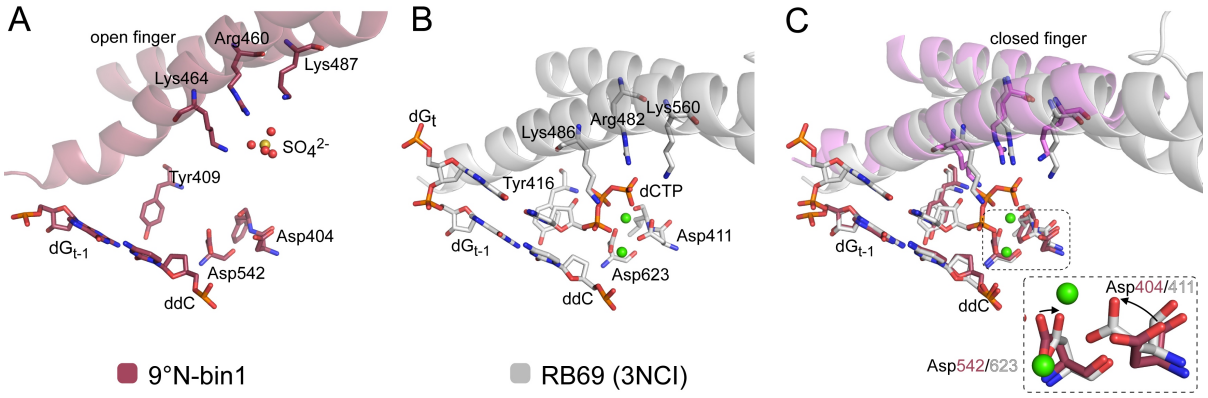


Figure 3.14: Residues involved in dNTP binding and catalysis in 9°N (raspberry) and RB69 (gray). The finger domain is shown as cartoon and residues discussed in the text as well as the p/t terminus and the newly formed base pair in the insertion site (for RB69) is shown as sitcks. **A**: A sulfate ion is bound by the basic residues of the finger domain. Asp404 is not properly aligned for catalysis in the open enzyme state. **B**: Active site arrangement in RB69 (PDB ID: 3NCI) **C**: Model of a closed 9°N structure. The C-terminal half part of the finger domain was superposed on the corresponding region of the closed RB69 finger and is shown in violet. Side chains were not altered. Residues contacting the triphosphate are situated in similar positions in RB69 and 9°N in a closed state. The movements of the two aspartate residues necessary to coordinate the two metal ion are indicated by black arrows.

ties. Several motifs in DNA polymerase sequences are conserved within a family but only weekly conserved between different families. Exceptions are motif A (DxxSLYPS) and C (YGDTDS) which are found in many DNA and RNA polymerases[83] and are directly involved in catalysis and fidelity mechanisms. In 9°N the sequences of motif A and C are **D**FRSLYPS and YADT**D**G. The absolutely conserved aspartate residues within these motifs (indicated in bold) usually coordinate the two divalent metal ions together with the triphosphate moiety of the substrate. In our binary structure, Asp404 and Asp542 are slightly shifted from their target position in an active ternary complex and need to rearrange before catalysis (**Figure 3.14**). The binding of the triphosphate is likely to be very similar as in other DNA polymerases, via charged residues of the finger domain. In 9°N a dNTP would most probably interact with the residues Lys487, Arg460 and Lys464 (Lys560, Arg482 and Lys486 in RB69) after closure of the finger domain in analogy to their interaction with the triphosphate in the closed structure of RB69. In our binary structure electron density for a sulfate ion instead of a triphosphate was present near these residues in the finger domain open state (**Figure 3.14 A**). For A-family polymerases it has been shown that a triphosphate substrate is initially bound by charged residues of the open finger and then delivered into the active site by a closing movement[71]. Recently also half-open or "ajar" states have been trapped crystallographically which support this

theory[90]. In B-family polymerases, however, a study suggested that here a dNTP directly binds into the active site of the enzyme and the finger domain closes afterwards generating a tight binding pocket in which the substrate is stabilized.
In RB69 the ribose moiety of the substrate triphosphate stacks on top of the phenyl ring of Tyr416 (of motif A) which is responsible for discrimination against ribonucleotides (section 1.2.4.5). In 9°N analogously the triphosphate is expected to stack on Tyr409 which adopts the same role as steric gate residue in archaeal as in other B-family polymerases[145, 222, 223].

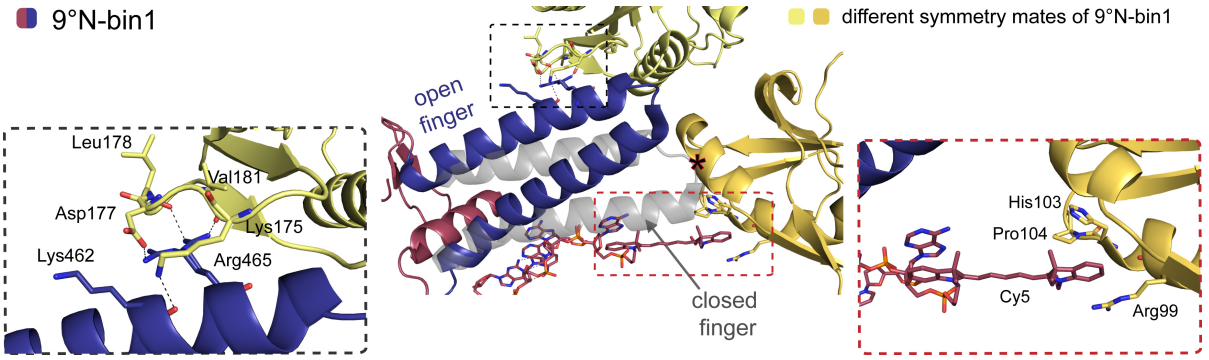


Figure 3.15: Crystal contacts of the finger domain in 9°N-bin1 which hinder finger domain closure. The closed finger domain of the superposed DNA polymerase δ (PDB ID: 3IAY) is shown in gray. Two different symmetry mates near the finger domain are shown in paleyellow and gold, respectively. A detailed view on specific H-bonds between the open finger domain and a neighbour molecule is shown. Residues taking part in H-bonding are shown as sticks and bonds are indicated as dashed lines. The Cy5 dye and the 5' ss template nucleotides are shown as sticks and the Cy5 interaction to a symmetry neighbor is visible.

Crystal contact. As mentioned above the main goal of the present project was to obtain a structure of 9°N in a ternary complex. After changing the crystallization strategy using dye-labelled templates I was able to select and optimize crystals of 9°N in complex with DNA and solve the just described first binary structure of 9°N (9°N-bin1) containing well ordered DNA. Due to the fact that cocrystallization did not lead to ternary complexes soaking of binary crystals with triphosphate substrate was the next obvious step. However, before starting soaking experiments the crystal contacts of the binary crystals in hand were inspected. It was found that the finger domain is stabilized in an open conformation and therefore soaking might not be successful. As shown in **Figure 3.15** the finger domain is in close proximity to the exonuclease domain of a symmetry-related molecule and Arg465 (with two alternative conformations) interacts with the side chain of Asp177 and the carbonyl backbone of Leu178 and Val181. Furthermore, the backbone oxygen of Lys462 can hydrogen bond with the backbone of Lys175. A second symmetry mate

extends to the space a closed finger is supposed to occupy and therefore finger domain closure is additionally hindered. Furthermore, one of the two alternative conformations of Cy5 stacks between Pro104 and His103 of this second neighbouring molecule on one side and Arg99 on the other side, additionally stabilizing the structure. Because of these contacts the crystals are not useful for soaking experiments as they are expected to get damaged upon finger domain movement. Also Pfu-E10 (described above) shows a crystal packing in which the finger domain is involved in lattice contacts and thus these crystals are also not suitable for soaking[220]. This leads to the assumption that the finger domain is well suitable to make crystal contacts under different conditions and therefore many crystals may have to be screened to find a crystal form in which the finger domain can move freely. Hence, we decided to mutate the arginine at position 465 to disturb the unwanted crystal contact and enhance the chance to reach our goal by further screening. We mutated the arginine to alanine first, to disrupt the H-bond and later to a glutamate to additionally introduce a repulsive charge to eventually disturb the interaction. With the two mutants I screened for ternary crystals using again the Cy-labelled p/t constructs introduced before. Thereby either ternary complexes should result directly or binary crystals with a different packing should be obtained, so that soaking can be performed. Indeed several crystallization hits were found and binary (but no ternary) structures with reasonable resolution for each of the two mutants could be solved. Both structures showed crystal forms different from 9°N-bin1. The structures are named 9°N-bin2 and 9°N-bin3 and will be described in the next section.

3.4.2 Structure of 9°N-bin2 - Arg465Ala mutant

With the Arg465Ala mutant a crystal structure of 9°N with an I/σ=2 resolution of 3.0Å was solved (9°N-bin2). Crystals grew in space group $P6_1$ (no. 169) with the cell axis a, b, c = 185.8Å, 185.8Å, 122.3Å and angles α, β, γ = 90°, 90°, 120°. The structure was solved by MR using 9°N-bin1 as a model. Two molecules were found in the asymmetric unit. The setups were made with a Cy3-labelled 16mer template annealed to a 9mer primer resulting in a 3'-d(GC) template overhang on one side and a 5'-Cy3-d(AAATT) template overhang on the other (construct $16/9_{Cy3}$ in **Figure 3.6**). DdATP was added to the protein/DNA mixture to trap the enzyme in a closed ternary state. The dCdG template overhang was supposed to be involved in crystal contacts hybridizing with the dCdG overhang of a neighbouring molecule. Indeed these crystal contacts were observed

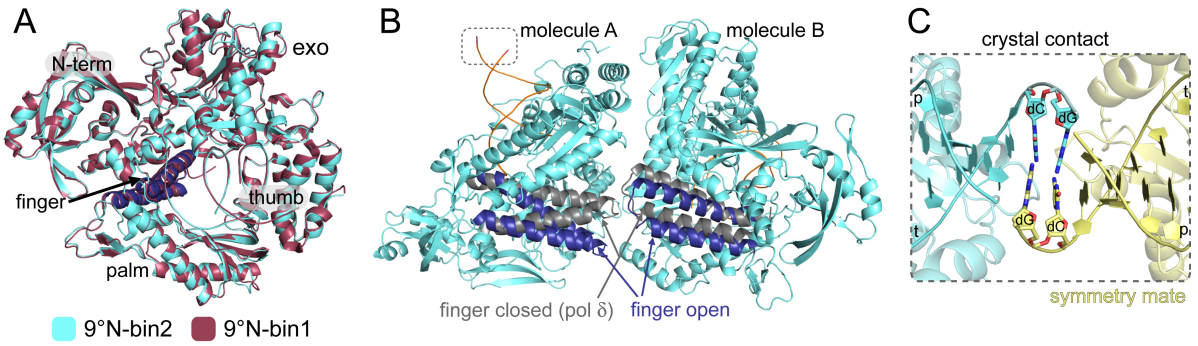


Figure 3.16: A: Overlay of 9°N-bin2 and 9°N-bin1 binary complexes. **B:** Molecule A and B of 9°N-bin2 facing each other with the finger domain. The position of the finger domains of a superposed closed ternary complex of DNA polymerase δ (PDB ID: 3IAY) on each molecule is shown in gray. **C:** Crystal contact via the 3' overhanging template nucleotides of neighbouring molecules.

(**Figure 3.16 C**). The two nucleotide overhangs of neighbouring molecules pair. However, in one of the two base pairs the nucleobases are slightly further away from each other than in an ideal Watson-Crick pair and only one out of three H-bonds is within the distance of 3.2Å. Both molecules contained DNA but no triphosphate substrate and showed an open binary conformation. The protein was modelled from residue 1-751 lacking the 24 C-terminal residues. Molecule A and B have an almost identical overall structure with an rmsd of 0.472Å. The structure 9°N-bin2 is very similar to the previously described structure 9°N-bin1 as indicated by an rmsd of 0.713Å for molecule A and 0.774Å for molecule B. In the overlay all protein domains as well as the DNA duplex superpose well (**Figure 3.16 A**). The mutation from arginine at position 465 to alanine therefore seems not to disturb the overall enzyme or DNA structure.

Template arrangement. The 5' ss template is arranged differently in 9°N-bin2 than in 9°N-bin1. The Cy-dye is not visible in the electron density and furthermore the t+2 and t+3 nucleotides are not well resolved. The templating nucleotide adopts the same position as the templating nucleotides in other binary family B polymerase structures (e.g. KOD (PDB ID: 4K8Z)[193] and phi29 (PDB ID: 2PZS)[224]). Accordingly the templating nucleotide is positioned on top of the nascent base pair, slightly shifted towards the template side and the nucleobase is kinked by approximately 20° compared to the position it adopts after pairing with its partner (**Figure 3.17 A**). This arrangement is different compared to A-family polymerases where the templating nucleobase is displaced to a much higher degree from its target position in the closed form and is situated in a pre-insertion site found in BF structures[129] or just flipped away from the DNA duplex and

pointing towards the solvent as shown for KlenTaq[44, 71]. The upstream 5' ss nucleotides point towards a groove between the N-term and exo domain which is realized by an approximately 130° rotation around the ribose-phosphate backbone between nucleotides t and t+1 (**Figure 3.17 B**). This arrangement is similar as the one found in the binary KOD structure[193]. Furthermore, in phi29[224] the 5' ss template points into the same direction but is lead through a complete tunnel and the protein is interacting with the t+1, t+2 and t+3 nucleotides.

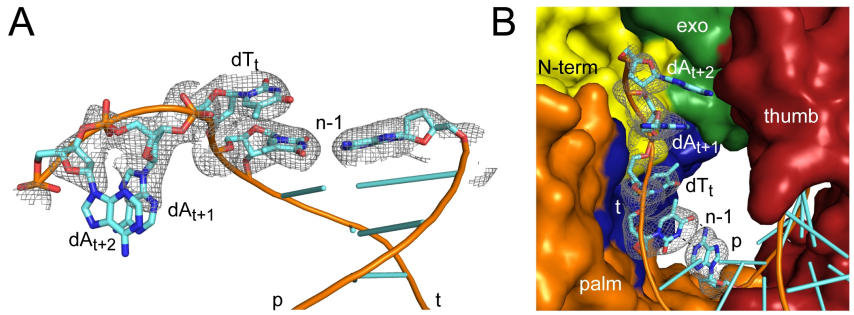


Figure 3.17: Arrangement of the templating dT, dA$_{t+1}$ and dA$_{t+2}$ in 9°N-bin2. **A**: The nascent base pair and the ss template overhang are shown surrounded by a mFo-DFc simulated annealing omit map contoured at 3σ. **B**: The protein is shown as surface with the domains color coded as in Figure 3.1.

Crystal contacts. Compared to the previously described binary structure the finger domain is not involved in crystal contacts in 9°N-bin2 (**Figure 3.16 B**). The finger domains of the two molecules in the asymmetric unit point towards each other and would come close to each other if both domains are closed. This was visualized by superposing the ternary pol δ structure on both molecules of 9°N-bin2 in PyMOL and subsequently aligning the finger domain of 9°N-bin2 with the closed finger domain of pol δ. The closest distance between Cα atoms of the two finger domains would be approximately 4.5Å. Thereby a probable clash could result but the domains could also get stabilized by binding to each other. Soaking experiments with these binary crystals have not yet been tested but are planned in future. To further optimize the crystals the crystal contact mediated by the overlapping 3' template ends might be further stabilized by increasing the pairing bases to three (in this case a mix of two different templates would be necessary) or by using a long p/t duplex at which two polymerase molecules can bind, one at each of the the two ends.

3.4.3 Structure of 9°N-bin3 - Arg465Glu mutant

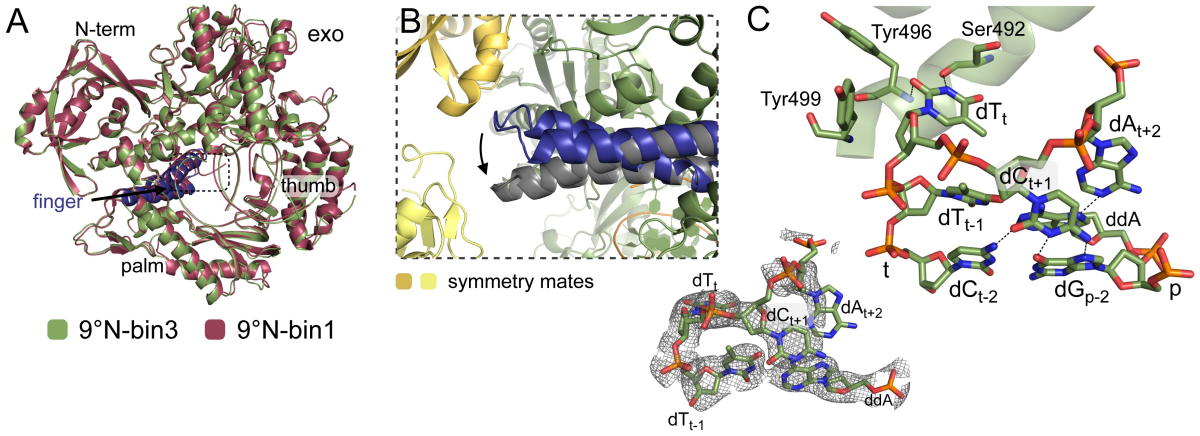


Figure 3.18: **A**: Overlay of 9°N-bin3 with 9°N-bin1. The finger domain is shown in blue. **B**: Parts of two symmetry mates which come closest to the finger domain are shown in gold and paleyellow, respectively. The position of the closed finger domain of the superposed pol δ is shown in gray. The finger domain is not involved in crystal contacts and a closure is not sterically hindered by the symmetry mates. **C:** Arrangement of the ss template overhang. Possible hydrogen bonding interactions are indicated by dashed lines. Bottom left edge: the nascent base pair (n-1) and the t, t+1 and t+2 nucleotides are shown surrounded by their simulated annealing omit map contoured at 3σ.

A third binary crystal structure was solved with the 9°N Arg465Glu mutant. The mutant was crystallized with a $17/12_{Cy5}$ p/t construct (**Figure 3.6**) and in presence of ddATP generating a 5'-Cy5-d(CACT) template overhang. The crystals grew in space group $P2_12_12_1$ (no. 19) with cell parameters a, b, c = 110.1Å, 111.9Å, 169.1Å and α, β, γ = 90°. The structure with an I/σ=2 resolution of 2.9Å was solved by MR using 9°N-bin1 as a model and harbours two molecules in the asymmetric unit. Both molecules contain DNA and have almost identical structures with an rmsd of 0.656Å. However, molecule A is more ordered and will therefore be used for description.

Overall structure and template arrangement. The overall structure shows an open enzyme state similar to 9°N-bin1 (**Figure 3.18 A**). In contrast to 9°N-bin1 and 9°N-bin2 the ss 5' template overhang adopts yet another structure in 9°N-bin3 (**Figure 3.18 C**). The templating dT_t is flipped out of the stacking arrangement with the p/t duplex and hydrogen bonds with the backbone of Tyr496 and Ser492 of the finger domain. Its sugar moiety is oriented roughly parallel to Tyr499. The t+1 dC flips back into the major groove side of the DNA duplex formed by the n-1 to n-3 base pairs. The position of the dC may be stabilized by H-bonds of dC to the hoogsteen edge of the dG_{p-2} and the 4-amino group

of dC_{t-2}. DA_{t+2} in turn interacts with the primer terminal nucleotide whereas the last template nucleotide dC_{t+3} and the Cy5 dye are not resolved.

Crystal contacts. Investigating the crystal packing revealed that the finger domain is not involved in crystal contacts and also the symmetry related molecules do not occupy the space the closed finger domain is supposed to adopt (**Figure 3.18 B**). Therefore finger domain closure would theoretically be possible and also (in addition to 9°N-bin2) this crystal form could be useful for soaking experiments with triphosphate substrates.

3.5 Discussion and Conclusion

In summary, several apo and binary structures of 9°N were solved during this work and three of each were characterized in more detail. Especially a fully open apo structure of 9°N and the binary structure of 9°N WT supply important models for future studies on 9°N and related enzymes.

3.5.1 Apo 9°N structures

The previously published 9°N apo structure was trapped in an unusual state with a closed thumb and palm domain by crystal contacts. One of the here described apo structures adopts a fully open conformation similar to other apo B-family polymerases crystallized so far (e.g. KOD, Tgo, RB69 [58, 61, 102]). Thereby it is confirmed that the closed apo conformation observed by Rodriguez *et al.* is only a crystal artefact and 9°N similar to other DNA polymerases appears with an open thumb and palm domain if no DNA is bound or at least is flexible to adopt different positions. Furthermore, this open apo structure in contrast to the structure reported by Rodriguez *et al.* can directly be used for comparison with 9°N binary complexes to elucidate the movements the enzyme undergoes upon DNA binding.

3.5.2 Features of the 9°N binary structures

After extensively trying to crystallize 9°N in complex with DNA and only obtaining apo structures the use of Cy-labelled templates finally lead to success and a 9°N binary structure was solved. In addition to the WT structure also structures of two single mutants

(Arg465Ala and Arg465Glu) were solved. These mutants were created to disturb crystal contacts formed by the finger domain in 9°N-bin1 and with the intention to obtain binary crystals suitable for soaking. The 9°N WT binary structure revealed that the DNA binding and also the DNA structure is similar to other available binary or ternary B-family polymerase complexes such as those from RB69[91], KOD[193], phi29[221] or the Pfu mutant E10[220]. In B-family polymerases the DNA adopts a B-form in the active site whereas in A-family members it is widened near the active site adopting an A-form. The fact that in our 9°N binary structure the DNA also adopts a B-form shows that also in archaeal family B polymerases this characteristic is conserved. The interaction pattern with the DNA is similar in archaeal and the bacterial members of the family, however, the residues interacting are not conserved throughout the family, but well within Euryarchaeota.

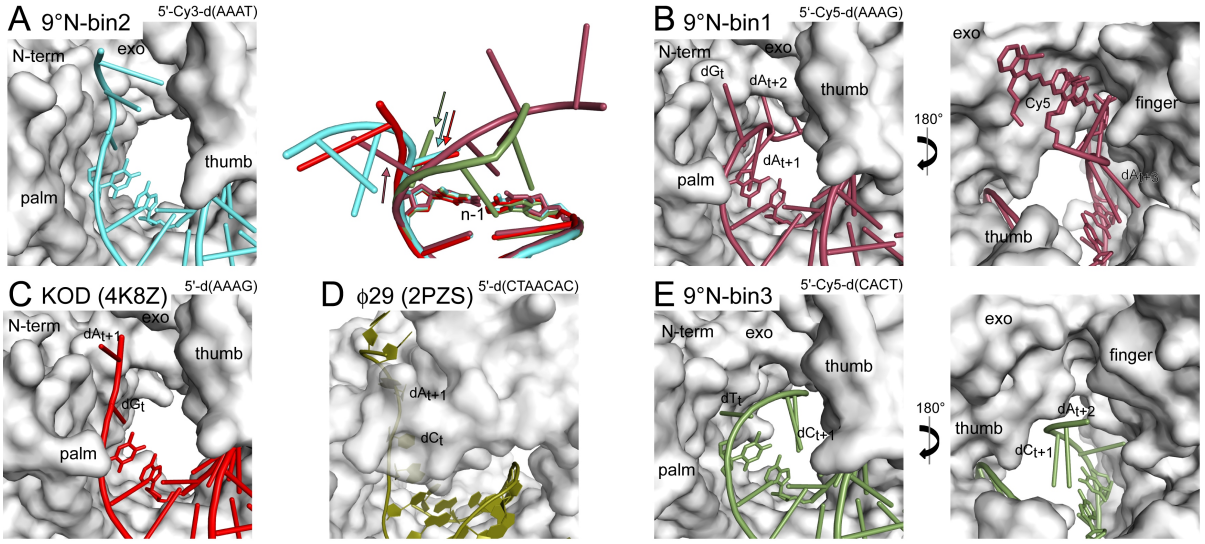


Figure 3.19: Arrangement of the 5' ss template overhang in 9°N-bin1 (raspberry), 9°N-bin2 (cyan), 9°N-bin3 (smudge) and related KOD (red) and phi29 (yellow) structures. The protein is shown as surface and domains are labelled. In 9°N-bin2, and KOD (PDB ID: 4K8Z[193]) structures the ss template overhang points towards a channel between the N-term and the exo domain. In phi29 (PDB ID: 2PZS) a similar arrangement is observed but here the protein forms a complete tunnel around the ss template. In 9°N-bin1 and 9°N-bin3 the template points towards the finger domain. To visualize that, these complexes are also shown from a 180° rotated view. In the overlay (right side in panel A) the templating nucleotides are indicated by arrows in the respective colors.

It is interesting that we find three different ss template overhang arrangements in our three binary 9°N structures. The differences we observe with either a 5'-Cy5-d(AAAG) overhang (9°N-bin1), a 5'-Cy3-d(AAAT) overhang (9°N-bin2) or a 5'-Cy5-d(CACT) overhang

(9°N-bin3) reveal that the 5' ss template is flexible and may adopt different arrangements in a crystal structure most probably depending on its sequence and length (**Figure 3.19 A**, **B** and **E**). The arrangement we observe in 9°N-bin2 is most similar to the ones found in the binary structures of KOD and phi29 (5'-d(AAAG) and 5'-d(CTAACAC) overhang, respectively; **Figure 3.19 C** and **D**). Here, the templating nucleobase is positioned on top of the t-1 nucleobase but slightly kinked and the remaining ss template part is rotated away and points into a groove between the N-term and exo domain. This arrangement is the one occurring most likely also in solution as here the usually much longer templates are directed through a groove between the palm and exo domain towards the solvent. The other two arrangements (observed in 9°N-bin1 and 9°N-bin3) seem to be stabilized only with that specific sequence and length and it is not clear whether they also occur in solution or even have an influence on the incorporation process. In these two arrangements the template end connected to the Cy-dye (resolved in 9°N-bin1 and not resolved in 9°N-bin3) points in a direction parallel to the finger domain. If the finger domain closes these arrangements can not be adopted any more and also for that reason seem disadvantageous as the template would need to move dramatically during each open and close cycle. Therefore, it is assumed that these back-flipping arrangements are just stable if (I) the template is as short as it is in our setups and (II) the enzyme is in an open state. In contrast it should be rather irrelevant with longer templates and in solution where the enzyme is dynamic and opens and closes alternately.

3.5.3 Importance of the 9°N binary structures

Archaeal family B DNA polymerases are used in various biotechnological applications such as genome sequencing, molecular diagnostics, DNA conjugation, synthesis of artificial genetic polymers (XNAs) and selection of aptamers by systematic enrichment of ligands by exponential amplification (SELEX)[189, 191, 192, 225]. In many of these methods the incorporation of specifically modified dNTPs is essential. To enhance incorporation efficiency of these substrates by DNA polymerases either the dNTP can be optimized by synthesizing different variants and subsequent testing or the enzyme can be evolved or rationally engineered via mutation to better accept a specific substrate. Rational optimization of either substrates or enzyme requires a high quality structural model, ideally of the enzyme in catalytically competent complexes with substrates. Our 9°N structure,

together with the binary KOD structure simultaneously solved in our group[193] and the structure of a mutant Pfu polymerase from the Leslie group[220] now serve as new models, that facilitate the generation of optimized enzymes and substrates for the archaeal family B polymerases.

3.5.4 9°N ternary structure - a remaining challenge

Despite the progress made by the determination of the 9°N binary complex structure, one remaining challenge is the crystallization of active ternary complexes of archaeal B-family polymerases with a p/t complex and a bound triphosphate in the replicative state. Ternary structures would help elucidate the mechanisms of selectivity, processivity and fidelity of these enzymes, which are widely used in biotechnology and molecular diagnostics, and elaborate their incorporation properties with specific dNTP modifications.

Probable reason for not obtaining ternary structures. Although we tried to crystallize 9°N in a ternary complex we never observed electron density for a bound triphosphate or metal ions near the active site. One reason for that could be that although dNTP is present in excess no stable population of closed active 9°N complexes are formed. A recent single-molecule study of the archaeal family B DNA polymerase B1 form *Sulfolobus solfataricus* revealed that only 66% of molecules adopt the closed conformation in the presence of a correct dNTP[226]. In KF in contrast, 90% of the enzymes are in the closed state with a correct dNTP[212]. This observation would explain the lack of ternary complex crystal structures of archaeal B-family polymerases as lacking homogeneity of closed complexes in solution hampers crystal formation. The circumstance that the closed conformation is not stabilized enough in case of triphosphate binding (although the enzyme should be stuck in that position as no 3' OH group for reaction is present) also explains why only binary complexes resulted from the setups, although triphosphate was present. If the population of closed enzyme molecules in solution is not homogeneous enough to lead to crystal formation soaking of binary crystals is the method of choice. Therefore binary crystals are required, in which the conformational changes of the enzyme upon substrate binding can be accomplished without disturbing the crystal lattice. In 9°N-bin1, however, the finger domain is trapped in an open state by crystal contacts making them unusable for soaking experiments. By generation of two single mutants at position

Arg465 which is involved in the mentioned crystal contact and screening for new crystal forms two additional binary structures could be solved in which the finger domain is not contacting symmetry related molecules and which could therefore be used for future soaking experiments. It remains to be tested if also the WT enzyme would crystallize in the same crystal forms as found for Arg465Ala in 9°N-bin2 and Arg465Glu in 9°N-bin3 in the respective conditions.

In a future study it is aimed to obtain a ternary structure based on the knowledge gained during this work. If crystallization with natural substrates can be established, lesion bypass and incorporation of modified nucleotides by the 9°N DNA polymerase will be investigated in detail.

4 Summary

DNA is composed of four nucleotides which specifically form two Watson-Crick base pairs. A third base pair that can be used orthogonally to the two natural base pairs would increase the coding potential of DNA and revolutionize its applicability in many research and application areas. Prerequisites for such an artificial base pair are the specific pairing of the two partners, a resulting stable DNA duplex structure and the acceptance of either nucleotide by DNA and RNA polymerases. The to date most promising unnatural base pair candidates are mainly hydrophobic artificial nucleobases which do not form hydrogen bonds but pair via hydrophobic and packing forces. One representative is the pair d**NaM**-d**5SICS**, developed in the group of Prof. Romesberg at the Scripps Research Institute in California. This pair is efficiently and selectively replicated by DNA polymerases and transcribed by RNA polymerases. In free duplex DNA the nucleobases show a cross-strand intercalated pairing which is in great contrast to the edge-to-edge pairing of correct natural base pairs. This finding raised the question about how the nucleotides can be so efficiently incorporated by DNA polymerases although their pairing behaviour is different. In this context it is questioned if the intercalated arrangement is retained in the active site of the enzyme and a different incorporation or elongation mechanism as for natural substrates takes place, or, if the structure of the artificial pair changes significantly to fit into the active site.

In the first part of this work these questions were addressed in a cooperation with Prof. Romesberg and members of his group. Using X-ray crystallography the acceptance of the unnatural base pair d**NaM**-d**5SICS** by KlenTaq DNA polymerase was investigated on a molecular level. KlenTaq binary complexes were solved with DNA containing the artificial nucleotides at the templating position and corresponding ternary structures with the respective partner nucleoside triphosphates bound in the insertion site were obtained by soaking of binary crystals. The binary complexes revealed that in the open enzyme state the unnatural nucleotides do not influence the enzyme or DNA structure. For the ternary complexes, two different reaction states between nucleotide binding and the in-

sertion reaction were trapped. Whereas the structure with the d**NaM**-d**5SICS**TP pair shows a catalytically active closed enzyme state, a bound d**NaM**TP opposite d**5SICS** resulted in an only partially closed polymerase finger domain. Interestingly, the base pair d**NaM**-d**5SICS**TP adopts a planar structure in the active site of KlenTaq, similar to a natural Watson-Crick pair. This structure shows that the artificial pair is able to change its intercalating arrangement towards an edge-to-edge arrangement and it is assumed that the DNA polymerase itself induces this transition. By closure of the finger domain a tight binding pocket is formed in which only planar base pairs fit properly. Furthermore, it was shown that the polymerase is able to adapt to small differences in size of the artificial pair compared to the natural ones as long as the differing moieties extend towards the developing major groove. The partially closed complex in presence of d**NaM**TP shows a pre-catalytic step before complete closure of the finger domain. This complex indicates that also in case of a hydrophobic artificial base pair the enzyme traverses specific steps before full closure which could coincide with the kinetic checkpoints observed for natural nucleotide insertion. However, the reasons for observing these two different complexes in case of d**NaM**TP or d**5SICS**TP binding are still unclear.

In addition to the insertion reaction also the elongation of the artificial base pair was investigated. The elongation reaction is still the bottleneck in replication of hydrophobic artificial base pairs. As I was unable to obtain ternary complexes with the artificial base pair in the post-insertion position n-1, followed by a natural triphosphate bound in the insertion site, I concentrated on the characterization of binary complexes after insertion of the unnatural nucleotide. Thereby, three complexes with d**NaM** in the primer and d**5SICS** in the template with slightly varying sequence and length of the ss template overhangs were solved. Additionally, a structure with d**5SICS** in the template and d**NaM** in the primer was obtained. All structures show that the artificial pair adopts an intercalated structure after formation of the phosphodiester bond and the transition from a closed into an open enzyme state. Due to the opening of the enzyme the pair is not constricted any more and an intercalated structure in which the hydrophobic and packing interactions are maximized is preferred. This arrangement can be stabilized by the enzyme in its open conformation. Furthermore, the enzyme is able to adapt to the changes in the p/t duplex structure resulting from the intercalating artificial pair. The intercalating structure of the nascent d**NaM**-d**5SICS** pair explains their relatively weak elongation. Prior to continued DNA synthesis an enormous rearrangement of the p/t complex is required. Thus, to optimize elongation, a reduction of packing interactions seems

promising. Indeed, the Romesberg group recently optimized the d**NaM**-d**5SICS** pair by reducing the aromatic surface area of the d**5SICS** scaffold, resulting in a new pair with decreased packing interactions and increased elongation efficiency. This success confirms the importance to study and understand the basis of artificial base pair replication on a molecular level.
In summary, it was shown that hydrophobic artificial nucleotides are well suited to form a third base pair and expand the genetic code. These pairs are flexible enough to intercalate on the one hand but can form a planar structure, if necessary, on the other. DNA polymerases are able to accept the pairs by inducing them to adopt a Watson-Crick like conformation. At the same time they are flexible enough to accommodate slightly enlarged nucleobases in a closed enzyme state or adapt to a variable p/t duplex structure in an open complex.

The second part of this work deals with the archaeal B-family DNA polymerase 9°N. Thermostable polymerases are important tools in molecular biology, e.g. to introduce modified substrates into DNA. Thereby, archaeal polymerases of the B-family showed superior insertion properties compared to thermostable A-family DNA polymerases such as *Taq*. Structural studies of DNA polymerases greatly contributed to a better understanding of those specific enzymes. Thereby, especially structures of DNA polymerases in complex with DNA and substrate are of great importance. Although bacteriophage B-family polymerases, particularly DNA polymerase RB69, are structurally well characterized, binary or ternary complexes of archaeal B-family polymerases lack completely. To better understand their properties and to generate a model that can be used in future design of modified enzymes and substrates it was aimed to crystallize a binary and/or ternary complex of 9°N and to analyze the structure(s).
After some preliminary studies a 9°N structure in complex with DNA was solved. As expected, the structure shows that the DNA is bound in a crevice formed by the finger and thumb subdomains. Upon binding of the DNA, the thumb domain closes and engages interactions with it. Furthermore, the DNA forms a B-form double helix as in other polymerase/DNA complexes of the B-family. Interactions with the DNA are mainly mediated via the phosphate backbone. Comparing DNA interactions in 9°N and the A-family polymerase KlenTaq revealed a probable reason for the better acceptance of modified nucleotides by archaeal B-family polymerases. In the major groove the tip of the thumb domain of KlenTaq might clash with modifications attached to the frequently

used positions C5 of pyrimidines and C7 of 7-deazapurines. In 9°N in contrast, the tip of the thumb domain is smaller at that position and does not extend into the major groove. Finally, the now available binary model of an archaeal B-family DNA polymerase is a good basis to rationally design enzymes and substrates for specific purposes in future. Despite this success a ternary complex is lacking for further characterization of this class of enzymes. As it seems difficult to obtain such a complex via cocrystallization, soaking of binary crystals might be the method of choice to proceed in this study. As the finger domain closes upon substrate binding it should be able to move freely in the crystals used for soaking. In the mentioned 9°N binary crystals, however, this is not the case as the finger domain is involved in crystal contacts. By mutation of an amino acid mediating such a contact two additional binary 9°N single mutant structures in different crystal forms could be solved. These structures fulfill the prerequisite of a flexible finger domain and can be used for soaking studies. However, if soaking of these structures is successful remains to be tested. In case of success further studies with various modified substrates are planned in future.

5 Zusammenfassung

DNA ist aus vier unterschiedlichen Nukleotiden aufgebaut, die spezifisch zwei Watson-Crick Basenpaare bilden. Ein drittes, nicht natürliches Basenpaar, das neben den beiden natürlichen verwendet werden kann, würde die Anwendbarkeit des Makromoleküls DNA in vielen Forschungs- und Anwendungsbereichen revolutionieren. Solch ein unnatürliches Basenpaar sollte stabile DNA Duplexe bilden, nur spezifisch miteinander paaren und selektiv von DNA- und RNA-Polymerasen akzeptiert werden. Die derzeit vielversprechendsten Kandidaten sind hydrophobe "Nukleobasen", die sich in ihrer Struktur von den natürlichen Purinen und Pyrimidinen unterscheiden. Sie bilden keine Wasserstoffbrücken aus und paaren stattdessen über hydrophobe und Stapelwechselwirkungen. Eines davon ist das Paar d**NaM**-d**5SICS**, welches in der Gruppe von Prof. Romesberg am Scripps Research Institute in Kalifornien entwickelt wurde. Die Gruppe konnte zeigen, dass das Paar selektiv und effizient von DNA-Polymerasen repliziert und auch von RNA-Polymerasen transkribiert wird. In freier doppelsträngiger DNA interkalieren die hydrophoben Basen und paaren nicht in einer Ebene, wie es bei den natürlichen Basenpaaren der Fall ist. Diese nicht natürliche, interkalierende Anordnung warf die Frage auf, wie diese Nukleotide dennoch so effizient von DNA-Polymerasen eingebaut werden können. Dabei galt zu klären, ob diese interkalierende Anordnung der Basen zueinander auch im aktiven Zentrum von DNA-Polymerasen anhält und dadurch der Einbau- oder Verlängerungsmechanismus ein anderer ist, oder ob sich die Struktur des artifiziellen Paars hin zu einer natürlichen Anordnung verändert, um in das aktive Zentrum des Enzyms zu passen.
In einer Kooperation mit der Romesberg Gruppe wurde im ersten Teil dieser Arbeit die Akzeptanz des unnatürlichen d**NaM**-d**5SICS** Paares durch die KlenTaq DNA-Polymerase mittels Proteinkristallographie und Röntgenstrukturanalyse auf molekularer Ebene untersucht. Dabei zeigten binäre Komplexe der KlenTaq mit den artifiziellen Nukleotiden an der Templatposition, dass diese sich nicht wesentlich von den Komplexen mit natürlicher DNA unterscheiden. Das heisst, dass in der offenen Konformation die artifiziellen Nukleotide die Enzymstruktur nicht verändern. Weiterhin wurden zwei unter-

schiedliche ternäre Komplexe mit dem nicht natürlichen Basenpaar in der Einbauposition erhalten. Der ternäre Komplex mit dem eintretenden Nukleotid d**5SICS**TP zeigte eine aktive, geschlossene Enzymstruktur, während für ein eintretendes d**NaM**TP eine halb geschlossene Konformation des Enzyms erhalten wurde. Interessanterweise paart das Basenpaar d**NaM**-d**5SICS**TP im aktiven Zentrum von KlenTaq ähnlich wie ein natürliches Paar in einer Watson-Crick Geometrie. Dadurch wurde gezeigt, dass das artifizielle Paar durchaus in der Lage ist seine Struktur anzupassen und es wird vermutet, dass dafür die DNA-Polymerase selbst verantwortlich ist. Durch das Schließen der Fingerdomäne entsteht eine enge Bindungstasche in die nur planare Basenpaare gut hineinpassen. Weiterhin wurde gezeigt, dass die Polymerase in der Lage ist, kleine Größenunterschiede des Basenpaares zu tolerieren, solange diese sich in die große Furche erstrecken. Der halb geschlossene Komplex in Anwesenheit von d**NaM**TP stellt einen Schritt vor dem vollständigen Schließen der Fingerdomäne dar und zeigt damit eine prä-katalytische Situation. Daraus wird geschlossen, dass die Polymerase auch beim Einbau hydrophober Nukleotide verschiedene Zwischenschritte durchläuft, bevor die Reaktion stattfindet. Der halb geschlossene Komplex könnte dabei mit einem der für natürliche Substrate beobachteten prä-katalytischen Komplexe übereinstimmen. Warum jedoch zwei unterschiedliche Komplexe im Falle von d**NaM**TP- oder d**5SICS**TP-Bindung gefunden wurden, ist noch unklar.

Zur Vervollständigung der Einbaustudien wurde im Rahmen dieser Arbeit auch die Verlängerung des artifiziellen Paares strukturell untersucht. Da es nicht gelang ternäre Komplexe mit dem artifiziellen Basenpaar in der n-1 Position und dem nächsten natürlichen Triphosphat in der Bindungstasche zu kristallisieren, konzentrierte ich mich auf die Charakterisierung von binären Komplexen nach dem Einbau des nicht natürlichen Nukleotids. Dabei wurden drei Komplexe mit d**NaM** im Primer und d**5SICS** im Templat mit leicht abweichender Sequenz und Länge des einzelsträngigen Templatüberhangs, sowie eine Struktur mit d**5SICS** im Templat und d**NaM** im Primer gelöst. Alle Strukturen zeigen, dass das artifizielle Paar nach der Ausbildung der Phosphodiesterbindung und dem Öffnen der Fingerdomäne, eine interkalierende Struktur annimmt. Durch das Öffnen des Enzyms wird das Paar nicht mehr in die Watson-Crick Geometrie gezwungen und eine interkalierende Struktur, in der die hydrophoben und Stapelwechselwirkungen maximiert sind, wird bevorzugt. In der offenen Konformation ist das Enzym in der Lage dieses Arrangement zu stabilisieren und sich an die daraus resultierenden Änderungen im p/t Komplex anzupassen. Die Tatsache einer interkalierenden Struktur am p/t Ende

erklärt, warum die hydrophoben, artifiziellen Paare relativ schlecht verlängert werden. Für den Einbau des nächsten Nukleotids ist eine enorme Umlagerung des p/t Komplexes notwendig. Es liegt daher nahe, dass eine Reduzierung der Stapeleigenschaften des artifiziellen Paares hilfreich wäre um die Verlängerungsreaktion zu optimieren. In der Tat wurde in der Romesberg Gruppe das d**NaM**-d**5SICS** Paar durch Verminderung der aromatische Oberfläche des d**5SICS** Gerüsts verbessert und ein neues Basenpaar mit schlechteren Stapeleigenschaften und besserer Verlängerungseffizienz erhalten. Dieser Erfolg zeigt deutlich, wie wichtig es ist die Mechanismen für den Einbau von artifiziellen Basenpaaren auf molekularer Ebene zu verstehen, um darauf basierend Entscheidungen für weitere Optimierung treffen zu können.
Zusammenfassend wurde gezeigt, dass hydrophobe artifizielle Nukleotide gut geeignet sind, um ein drittes Basenpaar zu erzeugen. Die Flexibilität, mit der diese Basenpaare einerseits interkalieren, anderseits aber auch eine planare Geometrie annehmen können, zeichnet sie dafür besonders aus. DNA-Polymerasen sind in der Lage diese Basenpaare zu akzeptieren, indem sie sie in eine Watson-Crick Konformation zwingen und gleichzeitig selbst flexibel genug sind, um sich an leicht veränderte Strukturen der Nukleobasen oder eine veränderte p/t Duplex Struktur im offenen Komplex anzupassen.

Der zweite Teil meiner Arbeit beschäftigt sich mit der archaealen DNA-Polymerase 9°N, die zu den B-Familie Polymerasen gehört. Thermostabile DNA-Polymerasen sind wichtige Werkzeuge in der Molekularbiologie, z.B. um modifizierte Substrate in DNA einzubringen. Dabei zeigten archaeale B-Familie DNA-Polymerasen bessere Einbaueigenschaften im Vergleich zu thermostabilen Polymerasen der A-Famile, wie z.B. *Taq*. Strukturelle Studien von DNA-Polymerasen haben deutlich zum besseren Verständnis dieser spezifischen Enzyme beigetragen. Dabei sind v.a. Strukturen von DNA-Polymerasen im Komplex mit DNA und Substraten von großer Bedeutung. Während Vertreter der B-Familie Polymerasen aus Bakteriophagen, v.a. die Polymerase RB69, gut charakterisiert sind, fehlten bis zum Beginn dieser Arbeit binäre oder ternäre Komplexe von archaealen Polymerasen der B-Familie gänzlich. Um deren Eigenschaften besser zu verstehen, aber auch um ein Modell zu erhalten, das in Zukunft das Design von modifizierten Enzymen und Substraten erleichtert, sollte im Rahmen dieser Arbeit ein binärer und/oder ternärer Komplex der 9°N DNA-Polymerase kristallisiert und die Struktur(en) analysiert werden.
Nach diversen Vorarbeiten konnte eine 9°N Struktur im Komplex mit DNA gelöst werden. Diese Struktur zeigt wie erwartet, dass die DNA in einem Spalt zwischen der Finger- und

Daumendomäne bindet. Dabei schliesst sich die Daumendomäne und interagiert mit der DNA. Weiter bildet die DNA - wie in anderen Polymerasestrukturen der B-Familie - eine B-Form aus und wird vom Enzym hauptsächlich durch sequenzunabhängige Interaktionen über das Phosphatrückgrat stabilisiert. Durch den Vergleich von DNA-Interaktionen in 9°N mit Protein-DNA Interaktionen der A-Familie Polymerase KlenTaq wurden die Gründe für die bessere Akzeptanz modifizierter Nukleotide von archaealen B-Familien Polymerasen untersucht. Ein Vergleich der Strukturen zeigt, dass in KlenTaq die Spitze der Daumendomäne in die große Furche hineinragt und dort Modifikationen, die an den gängigen Positionen C5 von Pyrimidinen und C7 von 7-Deaza-Purinen sitzen, sterisch stören kann. Bei der 9°N hingegen ist die Daumendomäne in diesem Bereich weniger ausladend und zeigt eher in die kleine Furche, was wiederum die bessere Akzeptanz von Modifikationen, die in die große Furche ragen, durch archaeale B-Familie Polymerasen erklärt. Letztendlich ist das nun zugängliche binäre Modell einer archaealen B-Familie Polymerase ein guter Ausgangspunkt, um in Zukunft Enzyme und Substrate gezielt für bestimmte Zwecke zu konstruieren. Trotz dieser Fortschritte fehlt für eine weitere Charakterisierung ein ternärer Komplex der 9°N. Da ein solcher Komplex über Kokristallisation bisher nicht erfolgreich war, sollte die Zugabe von Substrat zu fertigen binären Kristallen (*soaking*) zum Ziel führen. Da die Fingerdomäne nach Substratbindung von einer offenen in eine geschlossene Konformation übergeht, sollte sie in den Kristallen, die für das *soaking* benutzt werden, flexibel sein. In der beschriebenen 9°N Struktur ist die Fingerdomäne jedoch an Kristallkontakten beteiligt, die kein Schließen der Domäne zulassen. Durch die Mutation einer Aminosäure, die an einem der Kristallkontakte der Fingerdomäne beteiligt ist, wurden zwei zusätzliche binäre 9°N Strukturen in anderen Kristallformen erhalten. In diesen Kristallen ist die Fingerdomäne frei beweglich, sodass nun binäre Kristalle zweier 9°N Mutanten für zukünftige *soaking*-Versuche zur Verfügung stehen. Im Falle des Erfolgs sind weitere Studien mit modifizierten Substraten geplant.

6 Materials and Methods

6.1 General

6.1.1 Chemicals

Chemicals used in this work were in p.a. grade and are listed in the following table. Deionized and afterwards distilled water (in a water distillation apparatus: Aquatron, Dunn Labortechnik, Asbach) was used for the preparation of all buffers and in molecular biological experiments.

Substance	**Supplier**
Acetic acid	Roth
Acrylamide (30% Solution)	Roth
Agar	Roth
Agarose	Invitrogen
Ammonium acetate (NH_4CH_3COO)	Merck
Ammonium iodide (NH_4I)	Fluka
Ammonium sulfate ($(NH_4)_2SO_4$)	Roth
Ammonium phosphate ($(NH_4)_3PO_4$)	Merck
Ampicillin sodium salt	Roth
APS (Ammonium persulfate)	Fluka
β-Mercaptoethanol	Merck
Bromphenol Blue	Merck
Coomassie Brilliant Blue R250	Merck
DMSO	Merck
DTT	Roth
EDTA (Ethylendiaminetetraacetate)	Merck
Ethanol (100%)	Roth
Ethylene glycol	Merck

Substance	Supplier
Gel Red	VWR
Glycerol	Roth
Glycine	Roth
HEPES (2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure)	Fluka
Hydrochloric acid (37%)	VWR
IPTG (Isopropyl-beta-D-thiogalactopyranoside)	AppliChem
Kanamycin sulfate	Roth
Lithium chloride	Acros Organics
Lithium sulfate	Fluka
Lysozyme	Roth
Magnesium acetate ($Mg(CH_3COO)_2$)	Merck
Magnesium chloride ($MgCl_2$)	Acros Organics
Magnesium formate ($Mg(HCOO)_2$)	Fluka
Magnesium sulfate ($MgSO_4$)	Merck
MES (2-(N-morpholino)ethanesulfonic acid)	Roth
MPD (2-Methyl-2,4-pentanediol)	Sigma
PMSF (Phenylmethylsulfonyl fluoride)	Fluka
Polyethylene glycole 1000 (PEG 1000)	Fluka
Polyethylene glycole 3350 (PEG 3350)	Sigma
Polyethylene glycole 4000 (PEG 4000)	Fluka
Polyethylene glycole 5000MME (PEG 5000MME)	Fluka
Polyethylene glycole 8000 (PEG 8000)	Sigma
Polyethyleneimine solution (PEI)	Sigma
Potassium chloride (KCl)	Riedel-de Haën
SDS (Sodium dodecyl sulfate)	Roth
Sodium acetate ($NaCH_3COO$)	Roth
Sodium cacodylate trihydrate	Fluka
Sodium chloride (NaCl)	Roth
Sodium formiate (NaHCOO)	Fluka
Sodium hydroxide (NaOH)	Riedel-de Haën
Spermine	Fluka
TEMED (N,N,N',N'-Tetramethylethylenediamine)	Fluka

Substance	Supplier
Thesit	Fluka
Triton X-100	Merck
Tris (Tris(hydroxymethyl)-aminomehtane)	Sigma
Tryptone	BD, Le Pont de Claix, F
Xylen cyanol FF	Fluka
Yeast Extract	Roth

6.1.2 Oligonucleotides und triphosphates

Primers and templates for crystallization were purchased HPLC purified from Metabion, MWG Eurofins or Thermo Fischer Scientific. Modified oligonucleotides were synthesized by Denis Malyshev in the Romesberg Group at the Scripps Research Institute in La Jolla, California and purified as described in [44, 87]. Non-hydrolyzable triphosphates were purchased from Jena Bioscience and stored in aliquots at -80°. Primer for site-directed mutagenesis or cloning were purchased desalted from Metabion and used without further purification. For PCR reactions a dNTP mix (10mM each) from Thermo Scientific was used. Unmodified triphosphates for crystallization experiments (ddNTPs or dNTPs) were from Jena Bioscience, Fermentas or Roche.

6.1.3 Media, buffers and solution

Medium/buffer/solutions	components
LB-Medium	10g/l Trypton 5g/l Yeast Extract 10g/l NaCl
LB agar	LB-Medium 16g/l agar
Coomassie staining solution	30% (v/v) Ethanol 10% (v/v) Acetic acid 1.5g/l Coomassie Brilliant Blue R250

Destaining solution	30% (v/v) Ethanol 10% (v/v) Acetic acid
SDS-PAGE Tris Buffer (Upper Tris)	0.5M Tris-HCl pH 8.8 0.4% (w/v) SDS
SDS-PAGE Tris Buffer (Lower Tris)	0.5M Tris-HCl pH 6.8 0.4% (w/v) SDS
SDS-Page 2x loading buffer	50 mM Tris (pH 7.1) 2% (w/v) SDS 50% (v/v) Glycerol 5% (v/v) β-Mercaptoethanol 0.1% w/v Bromphenol Blue
SDS-running-buffer	25mM TRIS 192mM Glycin 0.1% SDS
50x-TAE-Buffer pH 8.3	2M Tris HCl Acetic acid (pH adjustment) 50mM EDTA
6x Agarose gel loading buffer	30% (w/v) Glycerol 0.025 % (w/v) Bromphenol Blue 0.025 % (w/v) Xylen cyanol FF 60mM EDTA
Oligonucleotide Annealing Buffer	10mM Tris pH 7.5 50mM NaCl 1mM EDTA

6.1.4 Plasmids and bacterial strains

plasmid	antibiotic resistance	size	comment
pGdR11	ampicillin	4753bp	derivative of pQE-31 (Qiagen) T5 promotor, lac operator
pET21b	ampicillin	5443bp	
pET24a	kanamycin	5310bp	

strain	properties
E. coli BL21 (DE3)	T7 expression strain
E. coli DH5α	cloning strain
E. coli Mach1-T1	cloning strain

For genotypes of respective strains see http://openwetware.org[227]

6.2 Molecular biological methods

6.2.1 General procedures

6.2.1.1 SDS-PAGE

In this work the discontinuous SDS-PAGE method according to Lämmli 1970[228] was used to separate proteins. Its characteristic is the difference in pH and acrylamide concentration in the stacking and resolving gel. All gels used in this work consisted of a 4% stacking gel and a 12.5% resolving gel (see **Table 6.3**). The protein samples were mixed

Substance	stacking gel (4%)	resolving gel (12,5%)
30% acrylamide	0.65ml	4.3ml
Upper-Tris (pH 8.8)	1.25ml	-
Lower-Tris (pH 6.8)	-	2.5ml
ddH2O	3.1ml	3.3ml
APS	35 µl	75 µl
TEMED	5 µl	10 µl

Table 6.3: Composition of a 12.5% SDS-PAGE

1:1 with SDS-PAGE 2x loading buffer and loaded directly or were boiled for 5min at 99°C prior to loading on the gel. Electrophoresis was conducted in SDS-running-buffer at a voltage of 160V for 10min followed by 250V for 35min. Gels were stained with Coomassie staining solution by boiling the solution in a microwave and keeping the gel for additional

15min on a shaker. Subsequently gels were destained in destaining solution until the bands became clear against the background. As molecular weight markers Mark12™ Unstained Standard, Invitrogen or PageRuler™ Prestained Protein Ladder, Thermo Scientific were used.

6.2.1.2 Agarose gel electrophoresis

Agarose gel electrophoresis (1% (w/v) agarose in 1x TAE-Buffer) was used for separation of DNA after PCR, digestion or ligation. The gel was prepared by dissolving 0.5g agarose in 50ml water and boiling in a microwave. 1ml 50x TAE-Buffer and 1µl staining reagent Gel Red was added and the mixture was poured into the gel chamber. Before loading, samples were mixed with 6x agarose gel loading buffer. Gel electrophoresis was conducted in 1x TAE at 100V. Results were documented under UV ligt at 254nm. As size marker the 1kb+ DNA Ladder from Fermentas was used.

6.2.1.3 Concentration of protein solutions

Protein solutions were concentrated before size exclusion chromatography and after the whole purification procedure by VivaSpin concentrators (Vivascience) capturing volumes of 15 or 6ml. Concentration was performed using a cooled centrifuge (4°C) at different speeds. By means of the centrifugal acceleration the molecules are filtered through the membrane. Depending on the pore size of the membrane proteins of particular size are retarded whereas smaller proteins, molecules and solvent can pass the membrane. For the KlenTaq DNA polymerase membranes with a pore size of 50kDa and 30kDa were used, for 9°N a pore size of 50kDa was used.

6.2.1.4 Concentration measurement via UV-spectroscopy

Concentration of DNA or protein solutions were measured using a NanoDrop device. The protein concentration was calculated using the Lambert-Beer law:

$$E_{280} = \varepsilon_{280} \cdot c[Protein] \cdot d$$

with
E_{280} = absorption at 280nm
ε_{280} = molar absorption coefficient at 280nm in $l/(mol \cdot cm)$

$c[Protein]$ = protein concentration in mol/l
d = cuvette width in cm (path length), defined as 1cm in NanoDrop

The theoretical extinction coefficients of KlenTaq and 9°N were determined using the ProtParam tool[229] of the ExPASy Bioinformatics Resource Portal (http://web.expasy.org). For KlenTaq the theoretical extinction coefficient is 70360 $l/(mol \cdot cm)$ at 280nm. For 9°N the theoretical extinction coefficient is 115100 $l/(mol \cdot cm)$ at 280nm. The described method for the determination of protein concentrations is only useful for already purified proteins, as contamination by additional proteins or other substances can influence the absorption measurement.

6.2.1.5 Site-directed mutagenesis

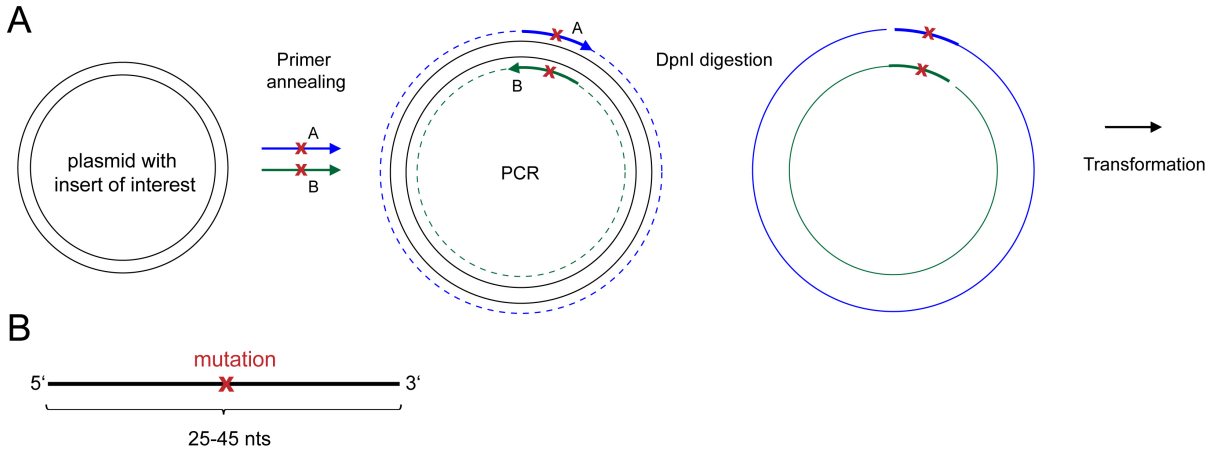


Figure 6.1: **A**: Schematic representation of the SDM method used. **B**: Design of a primer for the creation of a point mutation using the method shown in A. Scheme was created according to the QuickChange® site-directed Mutagenesis manual from Stratagene.

Site-directed mutageneis (SDM) was used to introduce point mutations into KlenTaq and 9°N for different purposes during this work. For the experiments a strategy following the QuickChange® site-directed mutagenesis system from Stratagene was used. The principle of the method relies on two primers (a forward and a reverse primer) for each mutation site which anneal at the plasmid and are elongated by PCR (see **Figure 6.1**). The primer thereby contains the mutation which should be introduced into the vector in the middle. Depending on the primer design and the sequence context at the location of mutation one or several nucleotides can be exchanged in one step. For succesful mutage-

nesis the primes have to conform to some general criteria which are described in detail in the aforementioned manual of the Stratagene Kit. Briefly, these criteria are: length 25-45bp, T_m ≥78°C, mutation in the middle of the primer and around 10-15bp on each side of the mutation, GC-content 40% and preferential primer endings C or G. For design of primers used in this work the mentioned criteria were tried to be followed, however melting temperatures of primers were often lower than 78°C. After PCR amplification and DpnI digestion of the parental DNA, circular vectors containing the introduced mutation are obtained and can be used for transformation of bacterial cells. The sequences of the primers used for the different SDM experiments are shown in the respective sections.

PCR amplification

Step	temperature	time	cycles
1 Initial denaturation	98°C	1min	1
2 Denaturation	98°C	10sec	
3 Annealing	gradient	30sec	20-25
4 Extension	72°C	4min	
5 Final Extension	72°C	10min	1

Table 6.4: PCR-cycling steps for SDM. Annealing temperatures varied depending on the T_m of the used primers. Usually a gradient was used to determine the optimal annealing temperature. Steps 2-4 were repeated 20-25 times.

If not stated otherwise all PCR reactions (polymerase chain reaction [230]) were performed using the Phusion DNA polymerase which was purified by Annemarie Weber in the group of Prof. Welte and kindly provided for this work. PCR was conducted similar as described in the protocol for Phusion high-fidelity DNA polymerase from Finnzymes. A standard SDM PCR cycling program used and a reaction mixture (total volume 20µl) is shown in **Tables 6.4** and **6.5**, respectively. For the PCR in SDM the annealing temperature was varied in a gradient around the melting temperature of the two primers. Generally an annealing temperatures ranging from 5-10°C above to 5-10°C below the higher melting temperature of the primers was used. For the elongation time approximately 30s/1000 nucleotides was used in 20-25 elongation cycles. The reaction was hold at 4°C until analysis by agarose gel electrophoresis. Control reactions were performed without template and without polymerase, respectively. If amplification failed the reaction was optimized by addition of DMSO (Finnzymes) in concentration steps between 2% and 10%.
PCR reactions were controlled on an agarose gel (see section 6.2.1.2). Positive fractions of different annealing temperatures were pooled together and parental methylated

Component	end concentration	reaction [µl]
Primer fwd (10µM)	200µM	0.4
Primer rev (10µM)	200µM	0.4
template (X/µl)	∽1pg-10ng/20µl	x
dNTPs (10µM)	200µM	0.4
5x Phusion HF reaction buffer (Finnzymes)	1x	4
Phusion Polymerase		0.3
DMSO, optional	2-10%	x
ddH_2O		add to 20µl

Table 6.5: PCR pipetting scheme for site-directed mutagenesis in a reaction volume of 20µl. Depending on the template concentration the amount of template added was adjusted or template was diluted.

E. coli plamid DNA was digested by addition of 1µl DpnI (New England Biolabs) and incubation at 37°C for 1h. DpnI was heat denatured at 80°C for 20min. Usually 5µl of the amplified DNA was used to transform chemically competent *E. coli* BL21 or *E. coli* DH5α cells.

6.2.1.6 Transformation

Generally 5µl of DNA sample from PCR or ligation reactions, or 1µl of purified plasmid were mixed on ice with 50µl of chemically competent *E. coli* cells (thawed on ice). After incubation on ice for 30min the mixture was heated for 90 seconds at 42°C in a water bath and subsequently cooled on ice for 5min. The cells were mixed with 250µl LB-Medium and incubated for 60min at 37°C while shaking. After incubation, cells were spread on LB-agar plates containing the respective antibiotic (100µg/ml ampicillin or 50µg/ml kanamycin) and incubated at 37°C over night.

6.2.1.7 DNA Sequencing

After DNA preparation from liquid cultures using the NucleoSpin Plasmid Kit from Sartorius, 30µl of plasmid DNA with a minimal concentration of 30ng/µl were sent for sequencing to GATC Biotech AG (Konstanz or Köln, Germany) and results were analysed afterwards.

6.2.1.8 Test expression

To test the expression of genes after cloning or site-directed mutagenesis a test expression from colonies or glycerol-stocks of cells was performed. Therefore, 5-10ml LB-medium

containing the respective antibiotic were inoculated with a colony from an LB-agar plate using a pipette tip or ∼10µl of a glycerol-stock from cells containing an expression plasmid with the gene of interest. The culture was grown at 37°C while shaking over night. 50-100µl of the overnight culture were transferred to a fresh 5-10ml LB-medium culture and grown for 1-2h until the OD_{600} reached 0.6-0.8. Before induction samples were taken for each clone. Expression was induced with 0.1-1mM IPTG and samples were taken after specific time points. The OD_{600} of each sample was measured in a cuvette and according to the OD the volume used for preparing the SDS-gel samples was calculated, so that comparable amounts of cells were used. Samples were centrifuged in a tabletop centrifuge at 3500 x g for 5min and the supernatant was removed. The cell pellet was re-suspended in 50µl 2X SDS sample buffer. The sample was heated for 10min at 70°C and afterwards again centrifuged at full-speed for 5min in order to pellet the genomic DNA. The supernatant of the control reaction (before induction) and the samples after specific expression times were applied to SDS-PAGE.

6.2.1.9 Activity test of polymerases

For activity testing of purified polymerases a PCR reaction to amplify the polymerase β gene (1010bp) in the pGdR11 vector was performed. The vector and the appropriate primers for amplification were kindly provided by my colleague Konrad Bergen. For the expression test, polymerases were diluted to 0.1mg/ml. Reactions were performed in either 5x Phusion HF buffer (New England Biolabs) or 10x ThermoPol buffer (New England Biolabs). The pipetting scheme and the PCR cycling program are shown in the following tables. An example result is shown in **Figure 6.2**.

Step	temperature	time	cycles
1 Initial denaturation	98°C	30s	1
2 Denaturation	98°C	10sec	
3 Annealing	62°C	30sec	30
4 Extension	72°C	30s	
5 Final Extension	72°C	10min	1

Table 6.6: PCR cycling steps for polymerase activity test. Cycles 2-4 were repeated 30 times.

Component	reaction [µl]
Primer fwd [10µM]	0.5
Primer rev [10µM]	0.5
template	1
dNTPs [10µM]	0.5
5x Phusion HF reaction buffer	5
polymerase to test	0.25
ddH_2O	17.25µl

Table 6.7: PCR pipetting scheme for polymerase activity test in 25µl total volume.

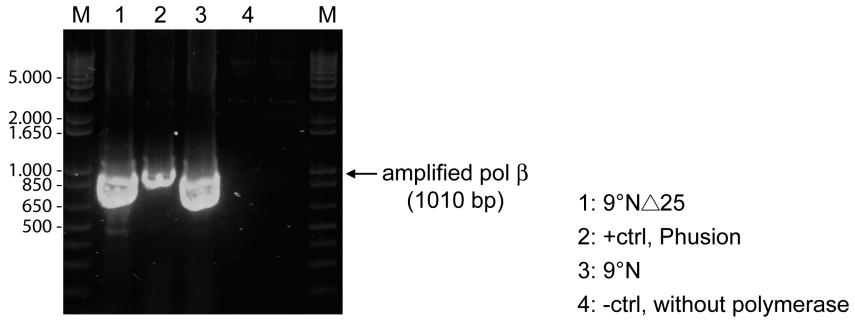


Figure 6.2: Result of activity test of 9°N and the truncated version 9°N△25. Both enzymes were able to amplify the polymerase β gene. As a control reaction Phusion DNA polymerase was used. The negative control without polymerase shows no product.

6.2.2 Expression and Purification of KlenTaq

In this work the large fragment (residues 291-832) of DNA polymerase I from *Thermus aquaticus* was used. The protein contains 540 amino acids and has a molecular weight of 60854 Da and an isoelectric point of 5.89 (determined by the ExPASy ProtParam tool, http://web.expasy.org/protparam/). For the overexpression of the KlenTaq gene in *E. coli*, a *E. coli* codonoptimized construct was used. During my master thesis[231], in collaboration with Dr. Samra Obeid[232], the construct of a KlenTaq double mutant (M747K I614K) from GENEART was cloned from the purchased vector pMK-RQ into the vector pET-21b using the restriction enzymes NotI (Fermentas) and NdeI (New England Biolabs). The WT plasmid was obtained by Site-directed mutagenesis. During his PhD work[233] Dr. Bastian Holzberger from the Marx group cloned the KlenTaq WT gene into the vector pGdR11, a derivative of pQE-31 (Qiagen) which contains an additional $laqI^q$ for enhanced expression control[234]. In my work both expression vectors for KlenTaq were used in *E. coli* BL21 (DE3) cells. Both vectors contain an ampicillin resistance cassette. Expression of the inserted gene is regulated by a Lac-Operon, inducible by addition

of IPTG. Stocks of cells were stored at -80°C in 50% glycerol.

buffer	components
Lysis buffer	50mM Tris, pH 8.55 10mM $MgCl_2$ 16mM $(NH_4)_2SO_4$ 0.1% Triton X100 0.1% Thesit
SQ1	20mM Tris, pH 8.55 1mM EDTA 1mM β-Mercaptoethanol
SQ2	SQ1 buffer 1M NaCl
Gelfiltration buffer (KlenTaq)	20mM Tris, pH 7.5 0.15M NaCl 1mM EDTA 1mM β-Mercaptoethanol

Table 6.8: Buffers for purification of KlenTaq.

Large scale expression of KlenTaq.

The expression and purification protocol used for the KlenTaq DNA polymerase was optimized towards good yield of protein in a short time together with my colleague Dr. Samra Obeid. For expression an *E. coli* codonoptimized construct without a His-tag was used. *E. coli* BL21 (DE3) Gold cells with the protein construct either in the pET-21b or the pGdR11 vector were cultivated in LB-medium containing 100µg/ml ampicillin at 37°C and 200rpm. A 80ml overnight culture was inoculated with 50µl of a glycerol stock and 10ml thereof were inoculated into the main culture the next day. After cells reached an optical density (OD_{600}) of 0.6-0.8, protein expression was induced by addition of 1mM IPTG. After 4h of expression, cells were harvested at 7000g and 4°C for 7min. Subsequently the cell pellets were resuspended in lysis buffer (∼ 4ml buffer/g cell pellet). Freshly dissolved Lysozyme (Roth, Karlsruhe; stock solution 10mg/ml) was added to an end concentration of 0.5mg/ml. Lysis was performed at 37°C for 1h.

Purification of KlenTaq.

Heat-lable proteins were heat denatured at 80°C for 20min. Cell debris and denatured

proteins were pelleted by ultracentrifugation at 4°C and 35000g for 1h. To get rid of the bacterial DNA a PEI precipitation was performed. Usually this precipitation is conducted under high salt concentration to avoid precipitation of the protein together with the DNA. However, if this method is used, additional efforts have to be made to get rid of the salt before applying the protein to an anion exchange column. As a compromise the precipitation was performed at low salt concentration with only a small amount of PEI to precipitate only part of the genomic DNA. Therefore, a 5% PEI solution was added tropwise. Between the addition of each portion of 100µl of PEI the sample was shaken. Approximately 100µl of PEI solution were added to 10ml of lysate. Subsequently the lysate was incubated on ice for 10min and afterwards centrifuged for 10min at 4000g. Anion exchange chromatography was performed for further purification.

Anion exchange chromatography. For anion exchange chromatography the material Q Sepharose (GE Healthcare) was used, packed into a XK16/20 column (GE Healthcare). The anion exchange column was stored in 20% ethanol and was equilibrated before use. Equilibration was carried out using 2 column volumes (CV) of ion exchange buffer 1 (SQ1), followed by 2CV SQ2 buffer for full regeneration and again SQ1 buffer. The protein solution was loaded on the column with the help of a superloop with a constant flow rate of 1ml/min. The unbound proteins were washed out with 200ml SQ1. Protein was eluted with a gradient raising from 0 to 8% buffer SQ2 in 300 min; 8 to 20% in 100 min and 20 to 50% in 100 min with a constant flow rate of 1ml/min. Afterwards the column was washed with 100% SQ2. Protein fractions were verified by SDS gel electrophoresis, pooled and concentrated prior to size exclusion chromatography.

Size exclusion chromatography. For further purification and buffer exchange, size exclusion chromatography was performed. The used column was 50cm in length with a diameter of 1.6cm. The column was stored in 20% ethanol and equilibrated with 2CV of KlenTaq gelfiltration buffer before use. 1ml of the concentrated protein solution was applied to the column. Protein was eluted with a flow rate of 1ml/min and 1.5 column volumes of gelfiltration buffer. Pure KlenTaq protein usually eluted as a single peak from the size-exclusion column. The eluted fractions were tested by SDS gel electrophoresis and pure protein fractions were pooled. The protein was concentrated to around 10-30mg/ml and stored at 4°C.

6.2.3 Creation of KlenTaq mutant T664S

To create the KlenTaq T664S mutant the codon ACC was exchanged to AGC. The SDM strategy described in section 6.2.1.5 was not successful in this case despite trying three different primer pairs. Although a PCR product could be amplified sequencing revealed that only the WT or a construct containing the sequence of the used primer multiple times, was obtained. Reasons for that might be the high AT content of the region in which the mutation should be introduced. To overcome that problem a different SDM strategy was used[235] which is shown in **Figure 6.3**. In this method two primers (A and B) are used which are designed tail-to-tail at their 5' ends in order to obtain inverse PCR products. The mutation in primer A is placed 9 nts away from the 5' end and 18-27 nts from the 3' end. The match region at the 3' side is used to determine the T_m for PCR.

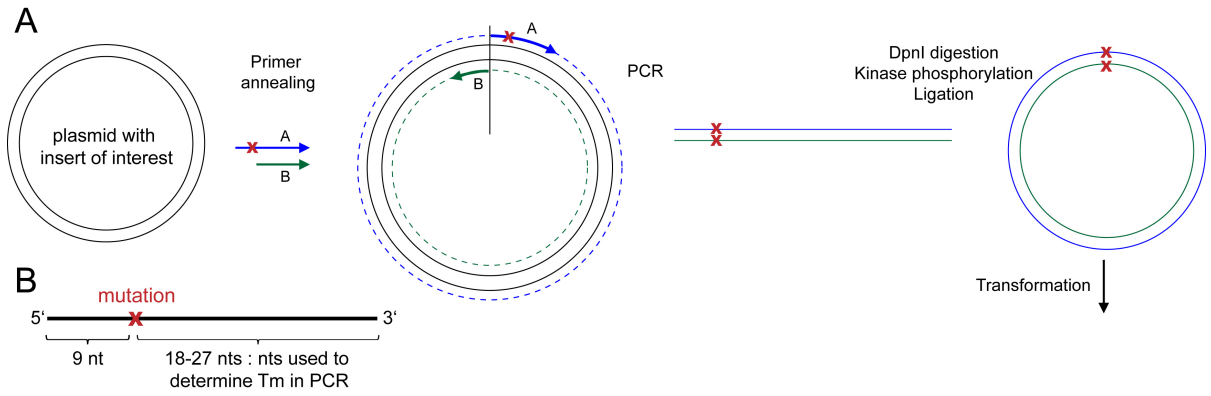


Figure 6.3: **A**: Schematic representation of an alternative SDM method. **B**: Design of a primer for the creation of a point mutation using the method shown in A. Scheme was adapted from [235] and slightly changed.

The following primers were used to successfully create the T664S mutant with the exchanged codon shown in red.

p-KTQ-T664S-for-4: 5′-d(GCAGCAAAAAGCATTAATTTTGGCGTGCTGTATGGTATGAGC)-3′
p-KTQ-T664S-rev-4: 5′-d(ACGACGCATCAGCGGATCAACTGCTTC)-3′

The PCR product was treated with DpnI to digest the circular plasmid template. Therefore 1µl of DpnI was added to the PCR product and incubated for 30min at 70°C. Subsequently the plasmids were phosphorylated by addition T4 polynucleotide kinase and gaps were ligated by T4 DNA Ligase. Therefore, 40.5µl water, 5µl 10x T4 DNA ligase buffer, 3µl PCR product, 1µl T4 polynucleotid kinase and 0.5µl T4 DNA ligase were mixed in a

reaction tube and incubated at room temperature for 30min. Afterwards 5µl of the reaction mixture were used to transform *E. coli* Mach1-T1 cells. The plasmid of an overnight culture was prepared and verified by sequencing using the GATC standard primers pBR-3 and pET-RP. The plasmid was used to transform *E. coli* BL21 (DE3) cells. Expression and purification of the mutant was conducted as described for the WT KlenTaq. The activity of the mutant was tested by using it in a PCR reaction as described in section 6.2.1.8.

6.2.4 Expression and purification of 9°N

For crystallization experiments with *Thermococcus* species 9°N-7 DNA polymerase a mutant carrying the two mutations D141A and E143A was used. This mutant shows reduced exonuclease activity[236]. The exo- enzyme is named 9°N throughout this work. The 9°N gene was cloned into a pET21b vector and overexpressed in E.coli BL21 (DE3). The protein 9°N has 775 amino acids, a molecular weight of 89677 Da and an isoelectric point (pI) of 8.83.

Preparation of expression constructs of 9°N and 9°N-pol.
The 9°N was cloned from a expression construct containing a His-Tag in pGdR11 vector into the vector pET24a with the purpose to remove the His-Tag. The full length 9°N gene was amplified via PCR with the primers p-9°N-full-for and p-9°N-full-rev. The annealing temperature was 60°C and the elongation time was 90s. Initially PCR was tried with shorter primers (three nucleotides shorter at the 5'-end) but the product could not be amplified. Simultaneously to the full length construct also only the polymerase domain of 9°N, named 9°N-pol, was amplified (primer p-9°N-pol-for and elongation time: 80s) with the purpose to investigate its stability and functionality in order to use it in crystallization experiments.

Primer sequences for amplification of the full length 9°N or 9°N-pol:

p-9°N-full-for: 5′-d(GAGCGAGCGCATATGATTCTGGATACCG)-3′
p-9°N-full-rev: 5′-d(CGACGACTCGCGGCCGCTTATTTTTTGCCTTTAACTTTC)-3′
p-9°N-pol-for: 5′-d(CGAGCGCATATGAGCAGCACCGGCAAC)-3′

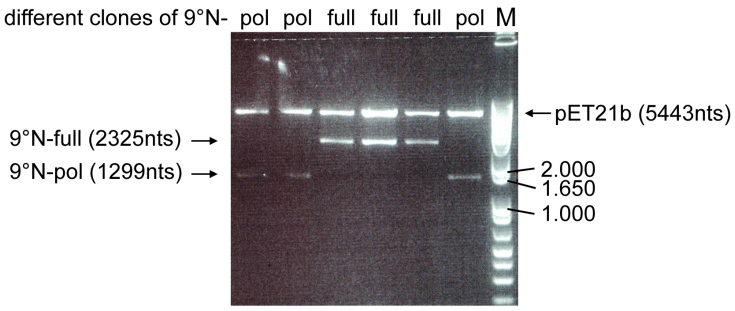


Figure 6.4: Digestion test of plasmids from different clones of 9°N and 9°N-pol using the restriction enzymes NdeI and NotI. The presence of the inserts could be confirmed in all 6 clones shown.

PCR products were analysed by agarose gel electrophoresis and purified via gel-extraction with the QIAquick Gel Extraction Kit. The vector pET24a was prepared from E.coli XL1 blue via NucleoSpin Plasmid Kit (Sartorius). 1µg of the two inserts were digested in a 30µl reaction using 0.5µl NotI and NdeI (Fermentas) restriction enzymes with 3µl of the appropriate supplied buffer (10x Buffer O). The vector pET24a was successfully digested in a 25µl reaction using 0.5µl of NotI and NdeI, respectively. The samples were digested at 37°C for 1h. The vector was dephosphorylated by treatment with antarctic phosphatase. For the reaction 20µl of the vector were treated with 2µl antarctic phosphatase and 2.4µl 10x antarctic phosphatase buffer and incubated for 15min at 37°C. The phosphatase was inactivated by incubation at 65°C for 5min. The digested vector was purified by gel-extraction. For the ligation, a 3:1 molar ratio of insert:vector was used. The vector has a size of 5400 nucleotides, whereas 9°N-pol comprises 1299 nucleotides and 9°N-full contains 2325 nucleotides. The reaction was performed in 20µl using 2µl T4 ligase and incubation for 15min at 25°C. The reactions were kept on ice until transformation. 5µl of the ligation reactions were used to transform chemically competent *E.coli* DH5α cells. For each construct 3 clones were cultivated, plasmids were purified and tested by a digestion reaction to confirm the presence of an insert with the correct size (**Figure 6.4**). The correct sequence was confirmed by sequencing. Plasmids were used to transform chemically competent *E.coli* BL21 (DE3) cells and a 10ml test expression was performed. The full length 9°N as well as the polymerase domain could be successfully expressed by induction with IPTG (see **Figure 6.5**). Stocks of cells were stored at -80°C in 50% glycerol.

Polymerase Domain of 9°N.

Before the test expression samples of 9°N and 9°N-pol were loaded onto a SDS gel, a heat denaturation step was performed for 20min at 75°C to test thermostability of the

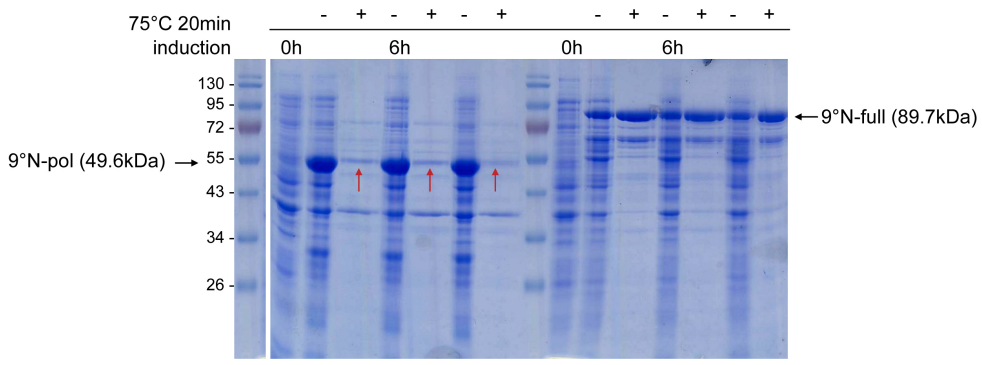


Figure 6.5: Expression test of 9°N full length (89.7kDa) and polymerase domain (49.6kDa) with three clones each before induction of expression (0h) and after 6h expression. A sample of each clone was heat denatured for 20min at 75°C. Whereas the full length 9°N is heat stable, the polymerase domain alone is heat labile (indicated by red arrows).

enzymes. It turned out that the polymerase domain alone is not heat stable and therefore was not used for further experiments (see **Figure 6.5**). The polymerase domain has a size of 49571 Da.

Large scale expression and purification of 9°N.

buffer	components	buffer	components
Lysis buffer	50mM Tris, pH 8.55 10mM $MgCl_2$ 16mM $(NH_4)_2SO_4$ 0.1% Triton X100 0.1% Thesit	Gelfiltration buffer (9°N)	200mM NaCl 20mM TRIS pH 7.5 0.1mM EDTA 1mM DTT 10% Glycerol
Hep1	100mM NaCl 20mM TRIS pH 7.5 0.1mM EDTA 1mM DTT 10% Glycerol	Hep2	1M NaCl 20mM TRIS pH 7.5 0.1mM EDTA 1mM DTT 10% Glycerol

Table 6.9: Buffers for purification of 9°N

E. coli BL21 cells carrying the expression vector were grown in LB medium at 37°C in presence of 50µg/ml kanamycin and shaking at 150-200rpm. Usually 2-6l of culture were grown simultaneously in 1l portions and 2l flasks. Cultures were inoculated by adding 10ml of a pre-culture grown over night. Expression was induced by addition of 0.5mM IPTG final concentration at OD_{600}=0.6-0.8. Cells were harvested after 4h, resuspended

in lysis buffer and either stored at -20°C or lysed directly. Lysis was performed for 1h at 37°C by adding freshly dissolved Lysozyme (10mg/ml stock) to 0.5mg/ml end concentration. After heat denaturation (10min at 75°C) 1mM PMSF was added and denatured proteins and cell debris were pelleted by ultracentrifugation (1h, 35000g).

Heparin column. The supernatant was loaded onto a 5ml HiTrap™ Heparin column (GE Healthcare) using a superloop and a flow rate of 1ml/min with the low salt buffer Hep1. The Heparin column was stored in 20% ethanol and was equilibrated before loading with Hep1, Hep2 and again Hep1 buffer. Unbound proteins were eluted by washing the column with 200ml of Hep1. Proteins were eluted by increasing concentration of NaCl using a step gradient and Hep2 buffer (**Figure 6.6**). Purest fractions (as determined by SDS-PAGE) were pooled and concentrated. The column was washed with 100% Hep2 buffer after use.

Size-exclusion chromatography. The protein was further purified by gel filtration chromatography using a XK 16/70 column containing Superdex 200 material (GE Healthcare). The column was stored in 20% ethanol and equilibrated with water and 9°N gelfiltration buffer before use. 1ml of the concentrated sample was applied to the column and the column was washed with 2CV of 9°N gelfiltration buffer. The protein eluted in one single peak and the retention time indicated a monomeric solution. After purification the protein was concentrated to around 7mg/ml and stored at 4°C. When concentrating beyond 7mg/ml or concentrating in lower salt buffer, precipitates of 9°N were observed.

6.2.5 Creation, expression and purification of 9°N△25 C-terminal deletion mutant

The C-terminal deletion mutant of 9°N, lacking 25 residues, was created to test if it shows improved crystallization properties compared to the full length enzyme. As the C-terminal part in the known 9°N apo structure 1QHT[57] was not resolved and as well in other archeal B-family polymerase apo structures the C-terminal end was disordered[58], deleting the region was assumed to yield more rigid molecules, favouring crystallization. The construct was created by amplification of the 9°N gene using the primers p-9°N-full-for (see section 6.2.4) and p-9°N-△25-rev. For amplification of the 2250 base pairs an elongation time of 70sec was used. The amplified band was extracted from the agarose gel and digested with the restriction enzymes NotI and NdeI. Simultaneously the pET24a vector was digested. Both digestion products were purified via gel extraction and ligated

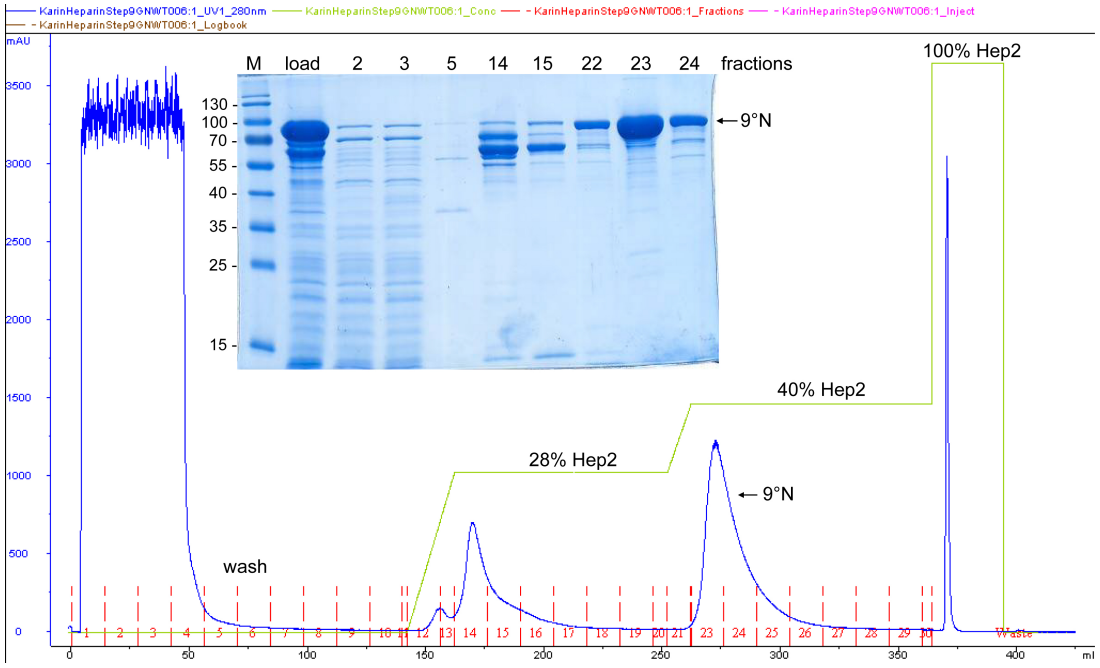


Figure 6.6: Chromatogram and SDS-gel of fractions from the 9°N heparin column. Elution gradient with raising high salt buffer (Hep2) is shown in green.

as shown in **Table 6.10**. 5µl of the ligation solution was used to transform *E. coli* DH5α cells (control with 1µl unligated vector) and the presence of the insert was tested via colony PCR. Plasmids of positive clones were prepared and sequenced and one positive construct was used for transformation of *E. coli* BL21 (DE3) cells. 9°N△25 could be expressed and its activity was confirmed by using it in a PCR reaction to amplify the gene of polymerase β (see section 6.2.1.8). 9°N△25 has 750 amino acids, a molecular weight of 86739 Da, a theoretical pI of 8.43 and an extionction coefficient of 115310 $l/(mol \cdot cm)$ (determined using the ProtParam tool[229]). The protein was purified similar to the full length 9°N. Due to the slightly different pI 9°N△25 eluted earlier from the Heparin column (already at 28% Hep2) and the elution gradient was changed to 22% in the first step, followed by 9°N△25 elution at 30% Hep2.

The sequence of the reverse primer used in PCR is:
p-9°N-△25-rev: `5′-d(CGACTCGCGGCCGCTTAATAGCCAAACGCTTTCAGAATACG)-3′`

Substance	[µl]
10x Ligase buffer	2
Ligase	1
Insert (33.9ng/µl)	10.5
Vector (7.9ng/µl)	6.5

Table 6.10: Ligation pipetting scheme. A control was performed with 10.5µl ddH_2O instead of the insert. Reaction was incubated for 11h at 18°C.

6.2.6 Creation of 9°N mutants R465A and R465E

To create the R465A mutant the codon CGC was changed to GCC. Therefore two nucleotides were exchanged simultaneously. The R465E mutant was created based on the R465A mutant and the codon GCC was changed to GAA. Site-directed mutatgenesis was performed as described in section 6.2.1.5 with the following forward and reverse primers (changed codons are shown in red):

p-9°N-R465A-for: 5′-d(CGCCAGAAAATCAAAGCCAAAATGAAAGCGACCG)-3′
p-9°N-R465A-rev: 5′-d(CGGTCGCTTTCATTTTGGCTTTGATTTTCTGGCG)-3′
p-9°N-R465E-for: 5′-d(CGCCAGAAAATCAAAGAAAAAATGAAAGCGACCG)-3′
p-9°N-R465E-rev: 5′-d(CGGTCGCTTTCATTTTTTCTTTGATTTTCTGGCG)-3′

An example PCR result of R465E is shown in **Figure 6.7**. A gradient of the annealing temperature from 55 to 72 °C as well as the addition of different concentration of DMSO was tested to find the best PCR condition. In the R465E case lower annealing temperatures and the presence of 10% DMSO yielded good results.

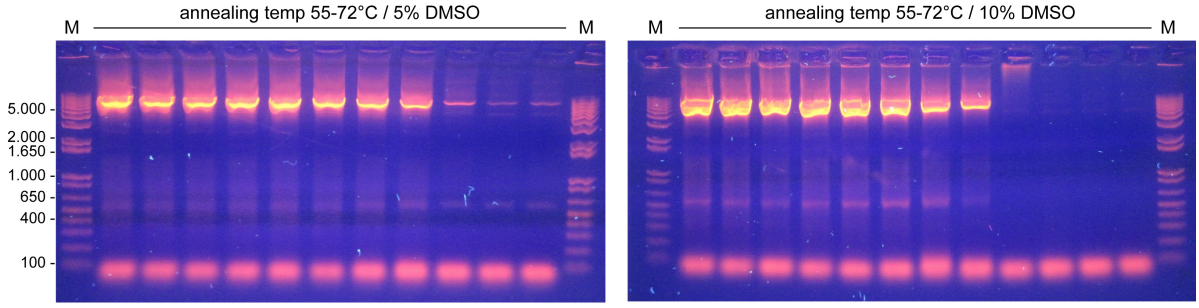


Figure 6.7: SDM PCR to create the 9°N R465E mutant. Lower annealing temperatures and addition of 10% DMSO yielded best amplification results.

Both mutants could be expressed and purified as the 9°N WT. The activity of the mutants was verified by using them in a PCR reaction to amplify DNA polymerase β (see section

6.2.1.8). Results of R465A overexpression and the activity test of R465E are shown exemplarily in **Figure 6.8 A** and **B**.

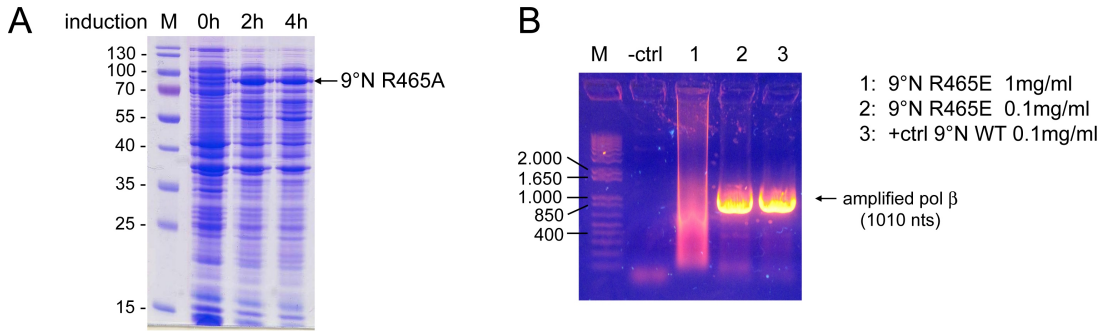


Figure 6.8: **A**: Overexpression of 9°N R465A. **B**: Activity test of 9°N R465E. The control reaction (-ctrl) without polymerase as well as two samples containing 9°N R465E in different concentrations and a positive control (+ctrl) with 9°N WT are shown.

6.3 Crystallization and X-ray analysis

6.3.1 General procedures

6.3.1.1 Protein crystallization

Protein crystals are built up by a regular arrangement of molecules in three dimensional space. Thereby the so called asymmetric unit (ASU) that can contain one or more macromolecules, builds up the unit cell as described by the symmetry operators of the space group. Unit cells translated in direction a, b and c make up the crystal. A basic requirement for the crystallization of proteins and other macromolecules is a highly pure (monodisperse) and highly concentrated protein solution. This solution has to be brought to oversaturation, which can be achieved by different methods. In an oversaturated solution the crystallization proceeds in two steps: nucleation and crystal growth. For nucleation several protein molecules have to align in an appropriate orientation to each other to form a stable nucleus in the solution. Afterwards single protein molecules can attach themselves to the nucleus resulting in crystal growth.

6.3.1.2 Vapor diffusion method

One method to realize oversaturation during a crystallization approach is vapor diffusion. Two varieties of this method were used in this work: the sitting drop and the hanging

drop method. Using the vapor diffusion method the pure protein solution (in this work: solution of protein, p/t, ddNTP or dNTP) is mixed with a specific volume of a reservoir solution containing precipitant and other additives in a droplet. The droplet is spatially separated from the reservoir solution in a sealed well. According to Raoult's law the vapor pressure above a solution is inversely proportional to the concentration of precipitant. Due to the higher partial pressure of the protein/reservoir droplet, water and volatile substances diffuse through the vapor phase out of the droplet into the reservoir solution, driving the system towards equilibrium. Simultaneously the protein- and precipitant concentrations rise and slowly an oversaturation is reached. Diffusion stops when vapor pressure equilibrium is reached. The slow generation of oversaturation favours nucleation against precipitation. The formation of nuclei and following crystal growth lowers the protein concentration and oversaturation drops into the metastable region. The slowly raising precipitant concentration in the droplet can compensate this shifting for a while enabling the crystals to reach bigger sizes. Schemes of the sitting and the hanging drop method, are shown in **Figure 6.9**.

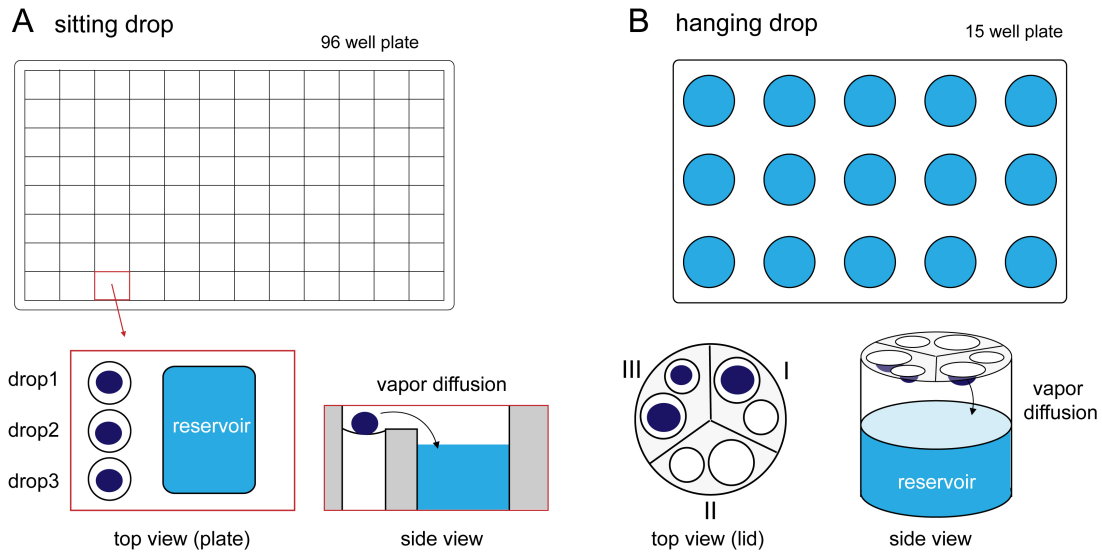


Figure 6.9: Sitting and hanging drop vapor diffusion crystallization method **A**: Schematic representation of the sitting drop method in 96-well format. Drops 1-3 were usually used to set up three different protein:reservoir ratios. **B**: Schematic representation of the hanging drop method in 15-well format. Crystallization solutions are mixed with the reservoir in the lid and the well containing reservoir alone is closed afterwards. Arrows indicate the transport of water and volatile substances from the crystallization drop to the reservoir solution containing a higher precipitant concentration

Sitting drop method. In this work the sitting drop method was mainly used to find initial crystallization conditions by the help of a pipetting robot. Each well of a 96 well

sitting drop plate was filled with 80-100µl of one condition of a commercially available crystallization screen. Subsequently a small volume of the protein solution (100-400nl) and different amounts of reservoir solution (ratios protein:reservoir 1:1, 1:2 or 2:1), respectively, were pipetted in a depression in each well by a Cartesian pipetting robot or a Gryphon robot (Art Robbins Instruments, ARI; since 03/2012). Thus, simultaneously 96 different crystallization conditions with three different protein:reservoir ratios could be tested. The plates were sealed and kept vibration protected at 18°C.

Hanging drop method. In the hanging drop method protein and reservoir solutions are mixed on a cover which afterwards is used to close the well containing the reservoir solution. In this work the hanging drop method was used for reproducing and optimizing initially found- and testing of published crystallization conditions in 15 well plates. Therefore 0.8ml of a self-made screen varying in pH and precipitant concentration with respect to the previously found condition was provided as reservoir. 1-2µl of protein solution and 0.5-2µl of reservoir were mixed in a gap in the cover with which the well was closed.

6.3.1.3 Crystallization of DNA polymerases

For successful crystallization of a protein or enzyme, a monodisperse solution of building blocks is required. To realize that prerequisite in case of DNA polymerases two different approaches were used during this work. To trap the polymerase in an active ternary state a dideoxy terminated primer is used to avoid an insertion reaction upon binding of the substrate triphosphate. This approach was first used by Pelletier *et al.*[65] to crystallize polymerase β and became a standard method to trap polymerases in ternary complexes. The second approach is based on the use of a α-β non-hydrolyzable substrate triphosphate to inhibit an insertion reaction and also enables trapping of the enzyme in an active ternary complex.
If not stated otherwise the primers and templates were dissolved in Oligonucleotide Annealing Buffer to an end concentration of 6mM. For annealing, primers and templates were mixed in a 1:1 ratio and heated to 95°C for 5min followed by stepwise cooling to 8°C. To form the desired binary or ternary enzyme/DNA complex, the DNA was mixed with the purified protein and - depending on the experiment - with additional ddNTP and/or dNTP in the presence of 20mM $MgCl_2$. After mixing DNA and protein the solution was incubated for 30min at 30°C if not stated otherwise and incubation was repeated for each addition of triphosphates.

Before using the solution to set up the crystallization experiment it was filtered through a 0.22µm sterile filter to get rid of possible precipitates or dust. Crystals were grown by vapor diffusion using the sitting or hanging drop method (see section 6.3.1.2) against different reservoir solutions. To find initial crystallization conditions commercially available Screens from Qiagen (Classics, PEGI, PEGII, Anions, Cations, MPD, pHClear, Cryos, PACT, $AmSO_4$), Hampton Research (Natrix), Molecular Dimensions (MIDAS) or Jena Bioscience (NucPro) were used in a sitting drop approach with the aid of a crystallization robot (Gryphon, ARI; since 03/2012). In our screening method 96 conditions and 3 different protein:reservoir ratios (1:2, 1:1 and 2:1) for each condition were tested on one plate. The drop sizes for screening varied between 0.25 and 0.4µl. Crystallization plates were sealed and kept vibration protected at 18°C. In case of promising hits further crystallization setups for improvement or preservation of larger crystals in sitting or hanging drop plates were prepared with a larger drop size. To optimize crystals, screens varying in precipitant concentration and/or pH were prepared in a deep well block in 2ml scale to be used in sitting-drop experiments in 96-well format or in 5ml scale to be used in hanging-drop experiments in 15-well format. If optimization was performed in sitting drop plates the drop size was increased to up to 1µl. In some cases also salts were varied or an additive screening was performed. Also in some optimization steps it was tried to crystallize directly in the presence of cryo-protectants like glycerol or ethylene glycol. Crystallization plates were usually inspected after 2 days, 1 week and 1 month. Crystal usually grew within this time period but plates were stored up to 2 years. Crystallization plates were inspected by eye using a binocular from Olympus and with the help of the Rock Imager 2 from Formulatrix.

6.3.1.4 Data collection

All crystals were tested for diffraction at the Swiss Light Source (SLS) of the Paus Scherrer Institute (PSI) in Villigen (CH), either on the beamlines PXI (X06SA - PILATUS 6M Detector) or PXIII (X06DA - PILATUS 2M Detector, since November 2011). Datasets were measured if test images showed sufficient diffraction properties. Prior to the transport to the synchrotron crystals were frozen in liquid nitrogen. To avoid water crystallization which can lead to ice rings and damage of the crystal lattice most of the crystals were cryo-protected in the reservoir solutions substituted with up to 20% (v/v) ethylene glycol or glycerol before freezing. For dataset acquisition with the PILATUS detectors generally 180° of the crystal were recorded by rotating the crystal in 0.1° steps with 0.1s exposure

time/frame. This fine φ-slicing strategy was shown to be optimal for measuring crystals with a single-photon-counting pixel detector[237]. At the beamline the intensity of the beam was adjusted such that diffraction spots are well visible in the diffraction viewer and the dose per dataset was not significantly higher than 30MGy to avoid radiation damage. For datasets recorded with a MarCCD Detector crystals were rotated in 1° or 0.5° steps and exposed for about 1s/frame.

6.3.1.5 Data processing and structure determination

Data processing was done using the XDS package[196]. Data statistics were inspected using XDSSTAT (provided by Prof. Kay Diederichs) or the XDS graphical user interface (XDSGUI). Structures were solved by difference Fourier techniques after rigid body refinement of a previously solved structure in the same space group and cell dimensions using phenix.refine or by molecular replacement (MR). For MR the Phenix GUI was used applying starting models lacking water molecules or ligands. Depending on the expected number of molecules/ASU (determined by Matthews coefficient) the search was performed for one or two molecules.

6.3.1.6 Refinement and model building

Refinement of the structures was performed using the program phenix.refine of the Phenix suite[197]. For the first structure of a specific space group, solved by Molecular Replacement, R-free values were generated using the uniqueify command of the CCP4 suite[238]. These R-free values were used for refinement of all subsequent structures solved by difference Fourier techniques after rigid body refinement of the first structure. For structures with better resolution R-free flags were extended using the program cad of the CCP4 suite[238]. Model building was done in Coot[198]. Model quality was evaluated using the Molprobity server[201] (http://molprobity.biochem.duke.edu/).
Side chains with disordered electron density were not deleted but modelled and high B-factors indicate their variance. Disordered regions were deleted and refinement including a simulated annealing step was performed. Deleted regions were tried to be placed into the omit map in Coot. If the backbone could not be traced properly the missing region was not modelled. If difference density appeared after refinement of ions such as iodide, sulphate or phosphate their occupancy was refined.

A typical refinement procedure (which varied slightly in each case) is shown in the following. Between each step the model was inspected and changed/improved based on difference density maps. Model building and refinement steps after step 4 were repeated several times until protein geometry and data statistics were satisfactory and could not be further improved.

1 rigid body refinement against starting model or MR
2 refinement using individual sites and individual ADP
3 automatic addition of water molecules (ordered solvent)
4 addition of TLS groups in refinement
5 addition of NCS refinement if NCS was present, geometry validation by Molprobity
6 addition of hydrogen atoms (for high resolution structures), geometry validation by Molprobity
7 optimization of target weights

In case of bad geometry of the starting model (high amount of Ramachandran and rotamer outliers) the geometry was optimized by a refinement step in phenix.refine in which the experimental data was excluded, resulting in an idealization of the stereochemistry (wxc_scale = 0). Before running this refinement step hydrogen atoms were added and removed again afterwards. Thereafter, refinement was run with Ramachandran restraints turned on for several macro cycles before turning them off again. This method remarkably improved quality of previously "bad" models.
To generate geometry restraint files (Crystallographic Information File = CIF file) for novel ligands, phenix.elbow or the grade web server[239] was used. PDB files for the unnatural nucleotides and Cy5 were created using sketcher of the CCP4 suite. Simulated annealing omit maps were created by setting the occupancy of the residues of interest to zero and running a refinement step in Phenix including simulated annealing.

6.3.1.7 Structure superposition and creation of figures

Overall structures were superposed in Coot[198] by the tool "secondary structure superpose" and the resulting core rmsd values are given for each comparison. Overlays of structures shown in the figures were generated in PyMol via the "align" command including all atoms. If not stated otherwise the whole enzyme-p/t-substrate complex was

aligned. Otherwise the parts which were aligned or the atoms included are stated. All figures were created using PyMol and the final figures were arranged using Adobe Illustrator.

6.3.2 Crystallization of KlenTaq in binary or ternary pre-insertion complexes

KlenTaq was purified as described in section 6.2.2 and crystallization setups were made as described in section 6.3.1.3. Details about protein concentrations used to obtain the different KlenTaq crystals and respective crystallization conditions are given below.
The following primer and templates were used to crystallize KlenTaq in binary or ternary pre-insertion complexes. The respective templating nucleotide is shown in bold.

p-11mer:	5′-d(GAC CAC GGC GC)-3′
t-G:	5′-d(AAA **G** CGC GCC GTG GTC)-3′
t-T:	5′-d(AAA **T** GGC GCC GTG GTC)-3′
p-ddC:	5′-d(GAC CAC GGC GCddC)-3′
t-I1$_{\mathbf{dNaM}}$:	5′-d(AAA **NaM** GGC GCC GTG GTC)-3′
t-I1$_{\mathbf{d5SICS}}$:	5′-d(AAC **5SICS** GGC GCC GTG GTC)-3′

KTQ$_{\mathbf{dT}}$ (PDB ID: 3SV4):
The open binary complex KTQ$_{\mathrm{dT}}$ was obtained by incubating KlenTaq (end concentration 7.0mg/ml) with a p-11mer/t-T complex and ddCTP in a molar ration of 1:1.2:5 in the presence of 20mM $MgCl_2$. Crystals grew under the following conditions: 18% PEG 8000, 0.2M $Mg(HCOO)_2$ and 0.1M Tris HCl (pH 8.5). Prior to flash-cooling the crystal was cryoprotected by soaking it consecutively in reservoir solutions containing 10% and 20% glycerol. In the KTQ$_{\mathrm{dT}}$ structure residues 647-653 of the finger domain as well as the last ss 5' template nucleotide were not modelled.

KTQ$_{\mathbf{dG}}$ (PDB ID: 3SZ2):
The open binary complex of KTQ$_{\mathrm{dG}}$ was obtained by incubating KlenTaq (final concentration 6.55mg/ml) with a p-11mer/t-G complex and ddGTP in a molar ratio of 1:1.5:5 in the presence of 20mM $MgCl_2$. Crystals grew under the following conditions: 15% PEG 8000, 0.2M $Mg(HCOO)_2$ and 0.1M Tris HCl (pH 8.0). Prior to flash-cooling the crystal was cryoprotected by soaking it consecutively in reservoir solutions containing 10% and 20% glycerol. In the KTQ$_{\mathrm{dG}}$ structure the complete polymerase was modelled. In the

DNA duplex the three 5' ss template residues 5'-d(AAA)-3' are not resolved.

$KTQ_{dG\text{-}dCTP}$ (PDB ID: 3RTV):
The closed ternary complex of $KTQ_{dG\text{-}dCTP}$ was obtained by incubating KlenTaq (final concentration 6.8mg/ml) with a p-11mer/t-G complex, ddGTP and dCTP in a molar ratio of 1:2:10:10 in the presence of 20mM $MgCl_2$. After incorporation of ddGMP the polymerase is trapped in a catalytically active complex with dCTP bound in the insertion site and two coordinated magnesium ions. Crystals grew under the following conditions: 30% PEG 4000, 0.2M $NH_4(CH_3COO)$, 0.01M $Mg(CH_3COO)_2$ and 0.05M sodium cacodylate (pH 6.5). Prior to flash-cooling the crystal was cryoprotected by soaking it consecutively in reservoir solutions containing 10% and 20% glycerol. In the ternary $KTQ_{dG\text{-}dCTP}$ structure all protein and DNA residues were modelled.

KTQ_{dNaM} (PDB ID: 3SYZ):
For crystallization of the binary complex of KTQ_{dNaM} KlenTaq (final concentration 7.1mg/ml) was incubated with a p-ddC/t-$I1_{dNaM}$ complex in a molar ratio of 1:1.5 in the presence of 20mM $MgCl_2$. Crystals grew under the following conditions: 11% PEG 8000, 20mM $MgCl_2$, 0.2M $Mg(HCOO)_2$, 0.1M Tris HCl (pH 8.0). Prior to flash-cooling the crystal was cryoprotected by soaking it in reservoir solution containing 10% glycerol. In the binary KTQ_{dNaM} structure all protein and DNA residues were modelled.

$KTQ_{dNaM\text{-}d5SICSTP}$ (PDB ID: 3SV3):
Crystals of the ternary complex were obtained by soaking of crystals of the binary KlenTaq complex with d**NaM** in the templating position for 20h with 6.7mM d**5SICS**TP. Crystals used for soaking with d**5SICS**TP grew in 13% PEG 8000, 0.1M NH_4Cl, 0.1M $Mg(HCOO)_2$, 0.1M Tris HCl (pH 8.0) and the final protein concentration was 7.4mg/ml. During soaking, two d**5SICS**TP molecules bound to the polymerase. One molecule is located in the active site as expected and a second molecule packs between R801 of one polymerase and Q633 and P816 of a symmetry related polymerase molecule. The second triphosphate seems well stabilized through a π-cation interaction with R801 as well as several hydrogen-bonds. In the ternary $KTQ_{dNaM\text{-}d5SICSTP}$ DNA structure the three 5' ss template residues 5'-d(AAA)-3' are not resolved.

KTQ_{d5SICS} (PDB ID: 4CCH):

The binary complex KTQ_{d5SICS} was obtained by incubating purified KlenTaq with p-11mer/t-$I1_{d5SICS}$ and ddCTP in a molar ratio of 1:1.2:5 in the presence of 20mM $MgCl_2$. The final concentration of KlenTaq was 5mg/ml. Setups were made using the sitting drop vapor diffusion method by mixing protein/p/t/ddCTP and reservoir solution in a 1:1 ratio (0.5µl + 0.5µl) and incubation over 70µl reservoir. Crystals were obtained in the following condition: 20% PEG 8000, 0.1M Tris (pH 8.0), 0.2M $Mg(HCOO)_2$, 20% Glycerol. The N-terminal amino acid 293 and the loop between residues 647 and 659 were not modelled in the structure due to disorder. Furthermore, the three 5' ss template residues 5'-d(AAC)-3' are not resolved.

$KTQ_{d5SICS\text{-}dNaMTP}$ (PDB ID: 4C8K):

The ternary complex $KTQ_{d5SICS\text{-}dNaMTP}$ was obtained by incubating KlenTaq with p-11mer/t-$I1_{d5SICS}$ and ddCTP in a molar ration of 1:1.2:2 in the presence of 20mM $MgCl_2$. The final concentration of KlenTaq was 6.5mg/ml. Setups were made using the sitting drop vapor diffusion method by mixing protein/p/t/ddCTP and reservoir solution in a 1:1 ratio. Crystals were obtained with the following reservoir solution: 15% PEG 8000, 0.1M Tris pH 8.0, 0.2M $Mg(HCOO)_2$. For soaking, crystals were transferred into a drop containing the same amounts of protein/p/t/ddCTP and reservoir solution and 2mM d**NaM**TP. Crystals were soaked for 10 days. Before cryo-cooling, crystals were consecutively soaked in reservoir solutions containing 10% and 20% glycerol.

6.3.3 Crystallization of KlenTaq with dNaM-d5SICS in post-insertion complexes

The following primers and templates were used to crystallize KlenTaq in post-insertion complexes.

p-11mer	5′-d(GAC CAC GGC GC)-3′
p-E1$_{\mathbf{dNaM}}$	5′-d(GAC CAC GGC GC**NaM**)-3′
t-E1$_{\mathbf{dNaM}}$	5′-d(A AAG **NaM**GC GCC GTG GTC)-3′
t-E1$_{\mathbf{d5SICS}}$	5′-d(A AAG **5SICS**GC GCC GTG GTC)-3′
p-E2/3	5′-d(ACC ACG GCG C)-3′
t-E2$_{\mathbf{dNaM}}$	5′-d(A G**NaM**G CGC CGT GGT)-3′
p-E3$_{\mathbf{d5SICS}}$	5′-d(ACC ACG GCG C**5SICS**)-3′
t-E3$_{\mathbf{dNaM}}$	5′-d(TT C**NaM**G CGC CGT GGT)-3′
t-E3$_{\mathbf{d5SICS}}$	5′-d(TT C**5SICS**G CGC CGT GGT)-3′
t-E4$_{\mathbf{dNaM}}$	5′-d(TT G**NaM**G CGC CGT GGC)-3′
p-E7$_{\mathbf{dNaM}}$	5′-d(GCA CGG CGC **NaM**)-3′
t-E7$_{\mathbf{d5SICS}}$	5′-d(TT TC**5SICS** GCG CCG TGC)-3′

KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$:
The binary complex KTQ(E1)$_{\mathbf{dNaM\text{-}d5SICS}}$ was obtained by incubating KlenTaq with p-11mer/t-E1$_{\mathbf{dNaM}}$, d**5SICS**TP and NHdCTP in a molar ratio of 1:1.5:5:20 in the presence of 20mM $MgCl_2$. Setups were made using the hanging drop vapor diffusion method by mixing KlenTaq/p/t/d**5SICS**TP and reservoir solution in a 1:1 ratio. The final concentration of KlenTaq was 6.1mg/ml. Crystals were obtained with the following reservoir solution: 24% PEG 6000, 0.05M NaHCOO, 0.1M MES (pH 6.5). The initial hit was found in condition C9 of the Qiagen PEG screen and the condition was varied by addition of different salts.

KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$:
The binary complex KTQ(E1)$_{\mathbf{d5SICS\text{-}dNaM}}$ was obtained by incubating KlenTaq with p-E1$_{\mathrm{NaM}}$/t-E1$_{\mathbf{d5SICS}}$ in a molar ratio of 1:1.5 in the presence of 20mM $MgCl_2$. Setups were made using the hanging drop vapor diffusion method by mixing KlenTaq/p/t/d**5SICS**TP and reservoir solution in a 1:1 ratio. The initial hit was found in condition E12 of the Qiagen PEG screen. The final concentration of KlenTaq was 6.55mg/ml. Crystals grew in

the following reservoir solution: 20% PEG 3350 0.2M NH_4I. Electron density for several iodine ions was found and ions were modelled. Occupancies of the ions varied and were refined in Phenix.

KTQ(E2)$_{\mathbf{dNaM\text{-}d5SICS}}$ (PDB ID: 4C8L):
The binary complex KTQ(E2)$_{\mathbf{dNaM\text{-}d5SICS}}$ was obtained by incubating KlenTaq with p-E2/t-E2$_{\mathbf{dNaM}}$ and d**5SICS**TP in a molar ratio of 1:1.2:5 in the presence of 20mM $MgCl_2$. Setups were made using the sitting drop vapor diffusion method by mixing KlenTaq/p/t/d**5SICS**TP and reservoir solution in a 1:1 ratio. The final concentration of KlenTaq was 6.2mg/ml. Crystals were obtained with the following reservoir solution: 0.2M $(NH_4)_2SO_4$, 0.1M MES (pH 6.5), 28% PEG 5000MME. Crystals were cryoprotected in reservoir solution containing 20% glycerol before freezing.

KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$ (PDB ID: 4C8O):
The binary complex KTQ(E3)$_{\mathbf{dNaM\text{-}d5SICS}}$ was obtained by incubating KlenTaq with p-E3$_{\mathbf{d5SICS}}$/t-E3$_{\mathbf{dNaM}}$ in a molar ratio of 1:1.2 in the presence of 20mM $MgCl_2$. Setups were made using the sitting drop vapor diffusion method by mixing KlenTaq/p/t and reservoir solution in a 1:1 ratio. The final concentration of KlenTaq was 6.2mg/ml. Crystals were obtained with the following reservoir solution: 0.2M $(NH_4)_2SO_4$, 0.1M MES (pH 6.5), 30% PEG 5000MME. Crystals were frozen without cryoprotection.

KTQ(E3)$_{\mathbf{d5SICS\text{-}dNaM}}$ (PDB ID: 4C8M):
The binary complex KTQ(E3)$_{\mathbf{d5SICS\text{-}dNaM}}$ was obtained by incubating KlenTaq with p-E3/t-E3$_{\mathbf{d5SICS}}$ and d**NaM**TP in a molar ratio of 1:1.2:5 in the presence of 20mM $MgCl_2$. Setups were made using the sitting drop vapor diffusion method mixing KlenTaq/p/t/d**NaM**TP and reservoir in a 1:1. The final concentration of KlenTaq was 6.2mg/ml. Crystals were obtained with the following reservoir solution: 0.2M $(NH_4)_2SO_4$, 0.1M MES (pH 6.5), 30% PEG 5000MME and 20% Glycerol.

KTQ(E4)$_{\mathbf{dNaM\text{-}d5SICS}}$ (PDB ID: 4C8N):
The binary complex KTQ(E4)$_{\mathbf{dNaM\text{-}d5SICS}}$ was obtained by incubating KlenTaq with p-E2/t-E4$_{\mathbf{dNaM}}$ and d**5SICS**TP in a molar ratio of 1:1.5:8 in the presence of 20mM $MgCl_2$. Setups were made using the hanging drop vapor diffusion method mixing KlenTaq/primer-template/d**5SICS**TP and reservoir in a 1:1 ratio. The final concentration of KlenTaq was

6.2mg/ml. Crystals were obtained with the following reservoir solution: 1.2M Li_2SO_4, 2% PEG 1000, 50mM HEPES (pH 7.5). Crystal were frozen without cryoprotection.

KTQ(E7)$_{\mathbf{d5SICS\text{-}dNaM}}$:
The binary complex KTQ(E7)$_{\mathbf{d5SICS}\text{-}\mathrm{dNaM}}$ was obtained by incubating KlenTaq with p-E7$_{\mathrm{dNaM}}$/t-E7$_{\mathbf{d5SICS}}$ in a molar ratio of 1:1.2 in the presence of 20mM $MgCl_2$. The final concentration of KlenTaq was 6.0mg/ml. Crystals were obtained in a sitting drop experiment with the following reservoir solution (condition G3 of the Qiagen Anions screen): 0.1M sodium tartrate pH 5.6, 1.2M potassium sodium tartrate. Crystal were frozen without cryoprotection.

	protein [mg/ml]	primer	template	(d)dNTP-1	dNTP-2	molar ratio
KTQ$_{\mathrm{dT}}$	7.0	p-11mer	t-T	ddCTP	-	1:1.2:5
KTQ$_{\mathrm{dG}}$	6.55	p-11mer	t-G	ddGTP	-	1:1.5:5
KTQ$_{\mathrm{dG\text{-}dCTP}}$	6.8	p-11mer	t-G	ddGTP	dCTP	1:2:10:10
KTQ$_{\mathbf{dNaM}}$	7.1	p-ddC	t-I1$_{\mathbf{dNaM}}$	-	-	1:1.5
KTQ$_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}\mathrm{TP}}$	7.4	p-ddC	t-I1$_{\mathbf{dNaM}}$	-	soak	1:1.5
KTQ$_{\mathbf{d5SICS}}$	5.0	p-11mer	t-I1$_{\mathbf{5SICS}}$	ddCTP	-	1:1.2:5
KTQ$_{\mathbf{d5SICS}\text{-}\mathrm{dNaMTP}}$	6.5	p-11mer	t-I1$_{\mathbf{5SICS}}$	ddCTP	soak	1:1.2:2
KTQ(E1)$_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}}$	6.1	p-11mer	t-E1$_{\mathbf{dNaM}}$	d**5SICS**TP	NHdCTP	1:1.5:5:20
KTQ(E1)$_{\mathbf{d5SICS}\text{-}\mathrm{dNaM}}$	6.55	p-E1$_{\mathrm{NaM}}$	t-E1$_{\mathbf{d5SICS}}$	-	-	1:1.5
KTQ(E2)$_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}}$	6.2	p-E2	t-E2$_{\mathbf{dNaM}}$	d**5SICS**TP	-	1:1.2:5
KTQ(E3)$_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}}$	6.2	p-E3$_{\mathbf{d5SICS}}$	t-E3$_{\mathbf{dNaM}}$	-	-	1:1.2
KTQ(E3)$_{\mathbf{d5SICS}\text{-}\mathrm{dNaM}}$	6.2	p-E3	t-E3$_{\mathbf{d5SICS}}$	d**NaM**TP	-	1:1.2:5
KTQ(E4)$_{\mathbf{dNaM}\text{-}\mathrm{d}\mathbf{5SICS}}$	6.2	p-E2	t-E4$_{\mathbf{dNaM}}$	d**5SICS**TP	-	1:1.5:8
KTQ(E7)$_{\mathbf{d5SICS}\text{-}\mathrm{dNaM}}$	6.0	p-E7$_{\mathrm{dNaM}}$	t-E7$_{\mathbf{d5SICS}}$	-	-	1:1.2

Table 6.11: Summary of components used in crystallization setups for KlenTaq structures.

	crystallitation condition	cryo
KTQ_{dT}	18% PEG 8000, 0.2M $Mg(HCOO)_2$, 0.1M Tris HCl (pH 8.5)	10+20% GOL
KTQ_{dG}	15% PEG 8000, 0.2M $Mg(HCOO)_2$, 0.1M Tris HCl (pH 8.0)	10+20% GOL
$KTQ_{dG\text{-}dCTP}$	30% PEG 4000, 0.2M $NH_4(CH_3COO)$,	10+20% GOL
	0.01M $Mg(CH_3COO)_2$, 0.05M sodium cacodylate (pH 6.5)	
KTQ_{dNaM}	11% PEG 8000, 20 mM MgCl2, 0.2M $Mg(HCOO)_2$,	10% GOL
	0.1M Tris HCl (pH 8.0)	
$KTQ_{dNaM\text{-}d5SICSTP}$	20% PEG 8000, 0.2M $Mg(HCOO)_2$, 0.1M Tris (pH 8.0)	20% GOL,
KTQ_{d5SICS}	20% PEG 8000, 0.2M $Mg(HCOO)_2$, 0.1M Tris (pH 8.0),	-
	20% GOL.	
$KTQ_{d5SICS\text{-}dNaMTP}$	15% PEG 8000, 0.2M $Mg(HCOO)_2$, 0.1M Tris pH 8.0	10+20% GOL
$KTQ(E1)_{dNaM\text{-}d5SICS}$	24% PEG 6000, 0.05M HCOONa 0.1M MES (pH 6.5)	-
$KTQ(E1)_{d5SICS\text{-}dNaM}$	20% PEG 3350 0.2M NH_4I	-
$KTQ(E2)_{dNaM\text{-}d5SICS}$	0.2M $(NH_4)_2SO_4$, 28% PEG 5000MME, 0.1M MES (pH 6.5)	20% GOL
$KTQ(E3)_{dNaM\text{-}d5SICS}$	0.2M $(NH_4)_2SO_4$, 30% PEG 5000MME, 0.1M MES (pH 6.5)	-
$KTQ(E3)_{d5SICS\text{-}dNaM}$	0.2M $(NH_4)_2SO_4$, 30% PEG 5000MME, 0.1M MES (pH 6.5),	-
	20% GOL	
$KTQ(E4)_{dNaM\text{-}d5SICS}$	1.2M Li_2SO_4, 2% PEG 1000, 50mM HEPES (pH 7.5)	-
$KTQ(E7)_{d5SICS\text{-}dNaM}$	0.1M Na tartrate pH 5.6, 1.2M K/Na tartrate	-

Table 6.12: Summary of crystallization and cryo protection conditions for KlenTaq structures. GOL = glycerol

6.3.4 Crystallization of apo 9°N

9°N-apo1:

For crystallization setups 9°N△25 was mixed with annealed p/t DNA (p-14: 5'-d(CG GAC TGC G)-3'; t-14mer: 5'-d(AAA GGC GCA GTC CG)-3') and ddCTP in a molar ratio of 1:1.2:5 and incubated for 30min at 30°C in presence of 20mM $MgCl_2$. The final protein concentration after adding all components was 7.0mg/ml. Diffracting crystals grew in the Natrix Screen condition A6 containing 0.1M $(NH_4)_2SO_4$, 0.01M $MgCl_2$, 0.05M MES (pH 5.6) and 20% PEG 8000. Crystals were cryoprotected consecutively in reservoir solution containing 10 and 20% ethylene glycol. In the structure amino acids 1-750 were modelled, with a gap between residues 667 and 692 in the thumb domain.

9°N-apo2:

For crystallization setups 9°N△25 was mixed with annealed p/t DNA (p-12: 5'-d(GAC TGC G)-3'; t-12mer: 5'-d(AAA GGC GCA GTC)-3') and ddCTP in a molar ratio of 1:1.2:5 and incubated for 30min at 30°C in presence of 20mM $MgCl_2$. The final protein concentration after adding all components was 7.0mg/ml. Diffracting crystals grew in the Natrix Screen condition A10 containing 0.005M $MgSO_4$, 0.05M MES (pH 6.0) and 5% PEG 4000. Crystals were cryoprotected consecutively in reservoir solution containing 10 and 20% ethylene glycol. In the structure of 9°N-apo2 several loops were disordered and could not be modelled. The final model lacks residues 570-574, 653-658, 667-674, 706-710 and 721-727.

9°N-apo3:

For crystallization setups 9°N△25 was mixed with annealed p/t DNA (p-11: 5'-d(GCT GCG)-3'; t-11mer: 5'-d(AAA GGC GCA GC)-3') and ddCTP in a molar ratio of 1:1.2:5 and incubated for 30min at 30°C in presence of 20mM $MgCl_2$. The final protein concentration after adding all components was 6.8mg/ml. Diffracting crystals grew in the Qiagen Nucleix Screen condition G8 containing 1.5mM Spermine, 10mM $MgCl_2$, 0.05M Na Cacodylat (pH 7.0), 2.2M $(NH_4)_2SO_4$. The crystal was frozen without cryoprotection. In the structure 9°N-apo3 several loops were disordered and could not be modelled. The final model lacks residues 609-612, 649-674, 693-699 and 707-728.

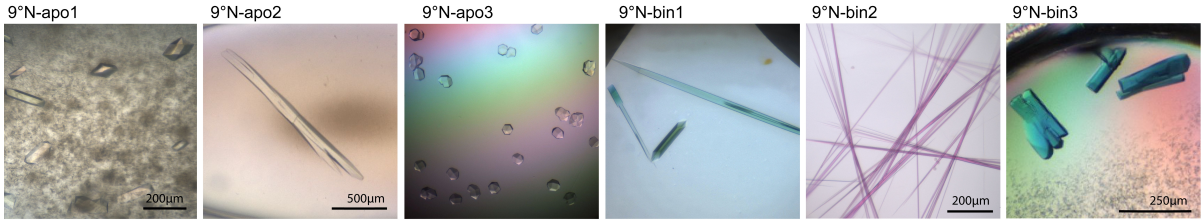


Figure 6.10: Images of crystals which resulted in 9°N apo and binary structures.

6.3.5 Crystallization of 9°N in a binary complex

The following primer and template sequences were used to successfully crystallize 9°N WT and two single mutants (R465A and R465E) in complex with DNA.

p-16	5′-d(CGC GAA CTG CG)-3′
t-16$_{Cy5}$	5′-Cy5-d(A AAG GCG CAG TTC GCG)-3′
p-2nt	5′-d(CGA ACT GCG)-3′
t-16$_{Cy3}$	5′-Cy3-d(A AAT TCG CAG TTC GCG)-3′
p-17	5′-d(CTG CGA ACT GCG)-3′
t-17$_{Cy5}$	5′-Cy5-d(CA CTT CGC AGT TCG CAG)-3′

9°N-bin1:
For crystallization setups 9°N was mixed with annealed p/t DNA (p-16/t-16$_{Cy5}$) and ddCTP in a molar ratio of 1:1.5:10 and incubated for 30min at 30°C in presence of 20mM $MgCl_2$. The final protein concentration after adding all components was 4.5mg/ml. Diffracting crystals grew in a drop containing 0.2M LiCl and 2.2M $(NH_4)_2SO_4$ (Qiagen Ammoniumsulfate screen condition B6). For cryoprotection crystals were soaked in the mother liquor containing 20% ethylene glycol.
Compared to the apo 9°N structure published in Beese *et al.* in 2000[57] 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. The 9°N-bin1 model is complete up to the last 23 C-terminal residues which were not resolved in the electron density. Conditions which yielded well-diffracting crystals were tried to be reproduced with the analogous unlabelled template. Thereby the resolution of the diffraction pattern slightly decreased. As the primer/template duplex parts were identical in labelled- and unlabelled template structures, only the Cy5-containing 9°N structure is described in this work.

9°N-bin2

For crystallization setups 9°N R465A was mixed with annealed p/t DNA (p-2nt/t-16_{Cy3}) and ddATP in a molar ratio of 1:1.5:8 and incubated for 30min at 30°C in presence of 20mM $MgCl_2$. The final protein concentration after adding all components was 4.3mg/ml. Diffracting crystals grew in a drop containing 50% MPD, 0.2M $(NH_4)_3PO_4$ and 0.1M Tris (pH 8.5) (Qiagen Classics screen condition B9). In both molecules in the ASU DNA is resolved, however in molecule A, the last 5' ss adenosine as well as Cy3 is not resolved and in molecule B two adenosine nucleotides as well as Cy3 is not resolved.

9°N-bin3

For crystallization setups 9°N R465E was mixed with annealed p/t DNA (p-17/t-17_{Cy5}) and ddATP in a molar ratio of 1:1.2:6 and incubated for 30min at 30°C. $MgCl_2$ was added to a final concentration of 20mM. The final protein concentration after adding all components was 4.55mg/ml. Diffracting crystals grew in a hanging drop vapor diffusion experiment in a drop containing 0.16M $(NH_4)_2SO_4$, 80mM $Na(CH_3COO)$ (pH 4.6), 22% PEG 4000 and 20% Glycerol. In this plate the initial hit condition H1 from the Qiagen Cryos screen was refined by varying the PEG 4000 concentration and pH. The structure contains two molecules in the ASU. In molecule a the last 19 C-terminal residues are disordered. In molecule B the model lacks residues 649-658 and 702-775. The DNA bound to molecule B is less well ordered than the one bound to molecule A.

	protein (mg/ml)	primer	template	ddNTP	ratio	Cryo
9°N-apo1	9°N△25 (7.0)	p-14mer	t-14mer	ddCTP	1:1.2:5	10 + 20% EG
9°N-apo2	9°N△25 (7.0)	p-12mer	t-12mer	ddCTP	1:1.2:5	10 + 20% EG
9°N-apo3	9°N△25 (6.8)	p-11mer	t-11mer	ddCTP	1:1.2:5	-
9°N-bin1	9°N (4.5)	p-16mer	t-16mer-Cy5	ddCTP	1:1.5:10	20% EG
9°N-bin2	9°N R465A (4.3)	p-2nt	t-16mer-Cy3	ddATP	1:1.5:8	-
9°N-bin3	9°N R465E (4.55)	p-17mer	t-17mer-Cy5	ddATP	1:1.2:6	20% Glycerol in reservoir

	crystallization condition
9°N-apo1	0.1M $(NH_4)_2SO_4$, 0.01M $MgCl_2$, 0.05M MES (pH 5.6), 20% PEG 8000
9°N-apo2	0.005M $MgSO_4$, 0.05M MES (pH 6.0), 5% PEG 4000
9°N-apo3	1.5mM Spermine, 10mM $MgCl_2$, 0.05M Na Cacodylat (pH 7.0), 2.2M $(NH_4)_2SO_4$
9°N-bin1	0.2M LiCl and 2.2M $(NH_4)_2SO_4$
9°N-bin2	50% MPD, 0.2M $(NH_4)_3PO_4$ and 0.1M Tris (pH 8.5)
9°N-bin3	0.16M $(NH_4)_2SO_4$, 80mM $Na(CH_3COO)$ (pH 4.6), 22% PEG 4000 and 20% Glycerol

Table 6.13: Summary of components used in crystallization setups for 9°N structures. Cryoprotection and crystallization conditions are also given. EG = ethylene glycol. Final protein concentrations after mixing with DNA and Triphosphate(s) are given in brackets.

7 Appendix

7.1 Abbreviations

Following abbreviations were used during this work:

A	Adenine
ADP	Atomic displacement parameters
APS	Ammoniumperoxidsulfate
ASU	Asymmetric Unit
BF	*Bacillus* fragment; Large fragment of DNA polymerase I from *Bacillus stearothermophilus*
pb	Base pair
C	Cytosine
CV	Column volume
ddNTP	2',3'-dideoxynucleoside-5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside-5'-triphosphate
dpo1	B-family DNA polymerase gene B1
Dpo4	Y-family DNA polymerase from *Sulfolobus solfataricus*
ds	Double-stranded
DTT	1,4-Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EG	Ethylene glycol
G	Guanine
GOL	Glycerol
GUI	Graphical user interface
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC	High pressure liquid chromatography
IPTG	Isopropyl-beta-D-thiogalactopyranoside
K_d	Dissociation constant
KF-	Klenow fragment of *E. coli* DNA polymerase I (3'-5'-exo-)
KTQ	= KlenTaq = Klenow fragment of DNA polymerase I from *Thermus aquaticus*
KOD	B-family DNA polymerase from *Thermococcus kodakaraensis*
K_{pol}	Rate constant of dNTP-incorporation
LB	Luria-Bertani
M	Molar [mol/L]
MR	Molecular replacement
min	minute
NCS	Non crystallographic symmetry
NHdCTP	2'-Deoxy-cytidine-5'-[(α,β)-imido]triphosphate
NMR	Nuclear magnetic resonance
nt(s)	Nucleotide(s)
OD	Optical density
p	Primer
PAGE	Polyacrylamide-gelelectrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDB ID	PDB identification = PDB code
PEG	Polyethylene glycole
PEI	Polyethyleneimine
pI	Isoelectric point
PMSF	Phenylmethylsulfonyl fluoride
Pol	Polymerase
p/t	Primer/template
RB69	DNA polymerases I from Enterobacteria phage RB69
Rmsd	Root mean square deviation
RNA	Ribonucleic acid
rNTP	= NTP = ribonucleoside-5'-triphosphate
rpm	Rotations per minute
RT	Room temperature, reverse transcriptase

s	Seconds
SDS	Sodium dodecyl sulfate
ss	Single-stranded
t	Template
T	Thymine
Taq	*Thermus aquaticus*
TEMED	N,N,N',N'-tetramethylethylenediamine
Tgo	DNA polymerase I from *Thermococcus gorgonarius*
TLS	Translation/Libration/Screw - refinement parameter
Tris	Tris-(hydroxymethyl)-aminoethane
UV	Ultraviolet
WT	Wild type

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

Table 7.2: Amino acid nomenclature, one and three letter codes

7.2 Supplementary tables

Name		apo	binary	ternary
9°N	archaeal	1QHT (2.1Å)[57]	4K8X (2.6Å)[193]	-
KOD	archaeal	1WNS (3.0Å)[58]	4K8Z (2.4Å)[193]	-
TGO	archaeal	1TGO (2.5Å)[102]	(2XHB⋆ (2.7Å)[107])	-
			(2VWJ* (2.8Å)[106])	-
D.Tok	archaeal	1D5A (2.4Å)[105]	-	-
SSo	archaeal	1S5J (2.4Å)[104]	-	-
Pfu	archaeal	2JGU (2.6Å)[103]	-	-
		4AHC (E10, 2.4Å)[220]	4AIL (E10, 2.9Å)[220]	-
RB69	viral/phage	1IH7 (2.2Å)[61]	-	3NCI (1.8Å)[91]
T4	viral/phage	1NOZ* (2.2Å)[240]	1NOY* (2.2Å ss DNA[240])	-
phi29	viral/phage	1XHX (2.2Å)[224]	2PZS (2.6Å)[221]	2PYJ (2.0Å, ddGTP)[221]
				2PYL (2.2Å, dTTP)[221]
HHVI	viral/phage	2GV9 (2.7Å)[241]	-	-
polII	bacterial	3K5O (2.2Å)[242]	-	3K59 (1.9Å, dCTP)[242]
		1Q8I (2.0Å) (n.p.)	-	3K58 (2.1Å, dTTP)[242]
		-	-	3K57 (2.1Å, dATP)[242]
		-	-	3MAQ (2.4Å, dGTP)[242]
δ	yeast	-	-	3IAY (2.6Å, dCTP)[100]
ε	yeast			4M8O (2.2Å, dATP)[101]
α	yeast	4FVM (2.3Å)[99]	4FXD (3.0Å, DNA/RNA)	4FYD (3.1Å, DNA/ RNA + dGTP)[99]

Table 7.3: Crystal structures of B-family DNA polymerases in apo, binary or ternary forms listed with their PDB IDs, resolution and reference. For ternary complexes also the bound triphosphate is given. ⋆hypoxanthin in ss temp, *uracil in ss temp, *fragment structure, n.p. = not published.

Name	PDB ID	storage location
KTQ$_{dG}$	3SZ2	/home/kbetz/SLS2011/mar27/tempg-e5
KTQ$_{dT}$	3SV4	/home/kbetz/SLS2011/may12/wt-t-nitro6
KTQ$_{dG\text{-}dCTP}$	3RTV	/home/kbetz/SLS2010/oct15/TempG-o11/
KTQ$_{\mathbf{dNaM}}$	3SYZ	/home/kbetz/SLS2011/feb03/NaM-d5-3SYZ/
KTQ$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}TP}$	3SV3	/home/kbetz/SLS2010/dec12/nam-tp-d24/
KTQ$_{\mathbf{d5SICS}}$	4CCH	/home/kbetz/SLS2013/apr09/13-5SICS-bin-8/phenix/4CCH
KTQ$_{\mathbf{d5SICS}\text{-}\mathbf{dNaM}TP}$	4C8K	/home/kbetz/SLS2012/oct22/NaMTP-8/phenix
KTQ(E1)$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}}$	-	/home/kbetz/SLS2011/sep05/nam-5sics2-s34
KTQ(E1)$_{\mathbf{d5SICS}\text{-}\mathbf{dNaM}}$	-	/home/kbetz/SLS2011/jun09/5sics-nam2-j5
KTQ(E2)$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}}$	4C8L	/home/kbetz/SLS2013/jun21/jun21-E3-NaM-6/phenix/4C8L
KTQ(E3)$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}}$	4C8O	/home/kbetz/SLS2013/feb05/13-E6-NaM-2/phenix
KTQ(E3)$_{\mathbf{d5SICS}\text{-}\mathbf{dNaM}}$	4C8M	/home/kbetz/SLS2013/jun21/jun21-E6-5SICS-2/phenix
KTQ(E4)$_{\mathbf{dNaM}\text{-}\mathbf{d5SICS}}$	4C8N	/home/kbetz/SLS2012/jul05/E2-NaM-plus2-3/phenix/4C8N
KTQ(E7)$_{\mathbf{d5SICS}\text{-}\mathbf{dNaM}}$	-	/home/kbetz/SLS2012/nov23/E7-5SICS-2
9GN-apo1	-	/home/kbetz/SLS2012/apr27/9GN-14mer-5_2/
9GN-apo2	-	/home/kbetz/SLS2012/mar23/9gn-2_2
9GN-apo3	-	/home/kbetz/SLS-2012/feb03/9GN-11mer-t6
9GN-bin1	4K8X	/home/kbetz/SLS2012/jul05/9GN-Cy5-soak-3_run1/2-3/2-1/4K8X
9GN-bin2	-	/home/kbetz/SLS2012/dec07/9GN-2nt-dec-2
9GN-bin3	-	/home/kbetz/SLS2013/feb05/13-Lena-2

Table 7.4: Storage location for processing and refinement results of structures discussed in this work. Diffraction images are stored on /nfs/loop1/synchrotron/ in folders of the appropriate date (SLS20XX/month/kbe) followed by the same primary folder names as indicated here.

acronym	source	uniprot sequence code	Sequence identity to 9°N
9°N	*Thermococcus sp.* (strain 9°N-7)	Q56366	-
KOD	*Thermococcus kodakaraensis*	P77933	90.7%
Tgo	*Thermococcus gorgonarius*	P56689	91.7%
Tli	*Thermococcus litoralis*	P30317	76.8%
Tfu	*Thermococcus fumicolans*	P74918	90.6%
Thy	*Thermococcus hydrothermalis*	Q9HH05	91.5%
Pfu	*Pyrococcus furiosus*	P61875	80.2%
Pab	*Pyrococcus abyssi*	P0CL76	81.8%
Pwo	*Pyrococcus woesei*	P61876	80.2%
Deep Vent	*Pyrococcus sp.* (strain GB-D)	Q51334	83.2%
DTok	*Desulfurococcus sp.* (strain Tok)	Q7SIG7	88.6%
Sac	*Sulfolobus acidocaldarius*	P95690	26.4%
Sso dpo1	*Sulfolobus solfataricus*	P26811	29.3%

Table 7.5: B-family DNA polymerases used in the sequence alignment shown in Figure 7.5. Sequence identities were determined by pairwise alignment with the 9°N sequence using the ClustalOmega web server[243].

7.3 Supplementary figures

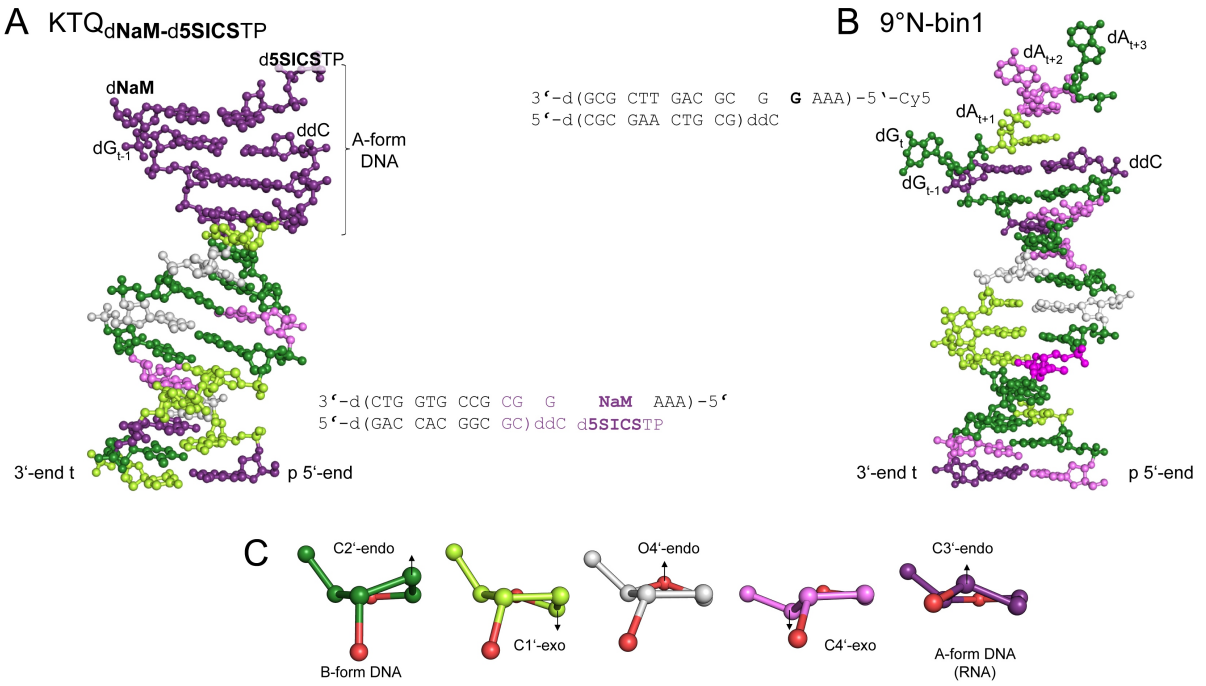


Figure 7.1: Sugar pucker of the p/t complex in **A:** $KTQ_{dNaM\text{-}d5SICSTP}$ and **B:** 9°N-bin1. The p/t is shown in a ball-and-stick model and nucleotides are coloured according to their sugar conformation as is shown in panel C. The sequences of the p/t duplexes are shown. The sugar conformation of polymerase complexed DNA was determined using Curves+[204]. **C**: Caption see **Figure 1.3**.

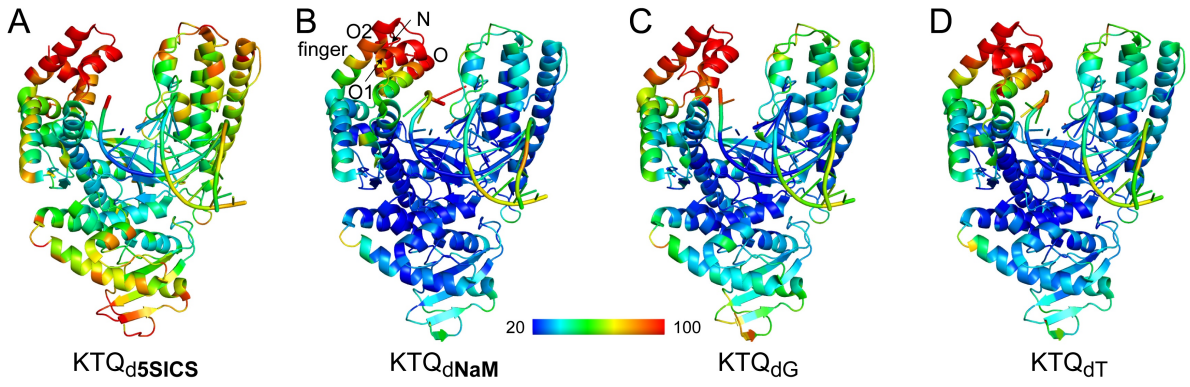


Figure 7.2: Binary KlenTaq structures with templating d**5SICS**, d**NaM**, dG and dT nucleotides coloured according to the temperature factor. The gradient is running from B-factor=20 (blue) to B-factor=100 (red). In all structures the finger domain, including helices N, O, O1 and O2 is the most flexible part of the enzyme. The flexibility of the p/t complex not bound by the enzyme is also visible.

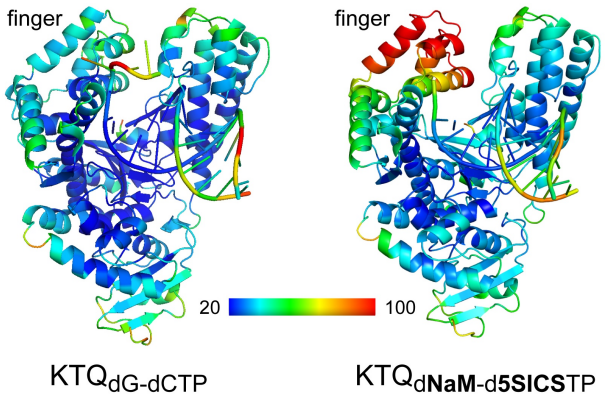


Figure 7.3: Ternary structures of KlenTaq with a natural dG-dCTP pair or with the unnatural pair d**NaM**-d**5SICS**TP in the insertion site. The structures are coloured according to the temperature factor. The higher flexibility of the finger domain in the closed state in presence of the unnatural base pair is obvious.

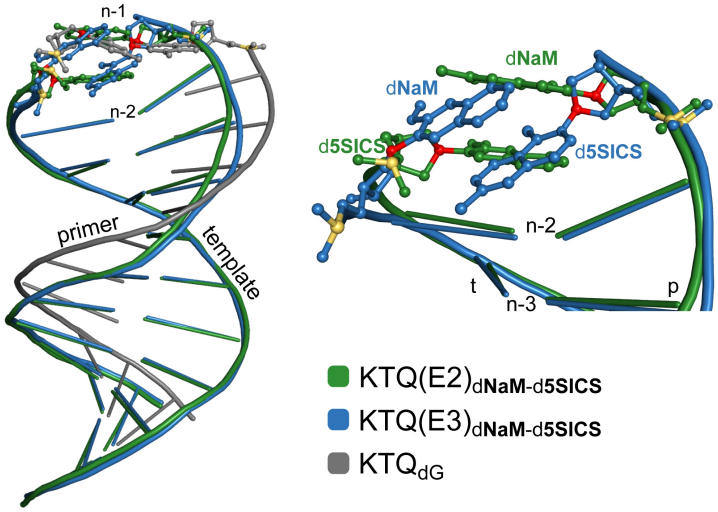


Figure 7.4: Overlay of the p/t duplex of structures KTQ(E2)$_{\text{dNaM-d5SICS}}$ and KTQ(E3)$_{\text{dNaM-d5SICS}}$. Although the artificial pairs show a different intercalating arrangement the influence on the rest of the p/t duplex is similar in both structures.

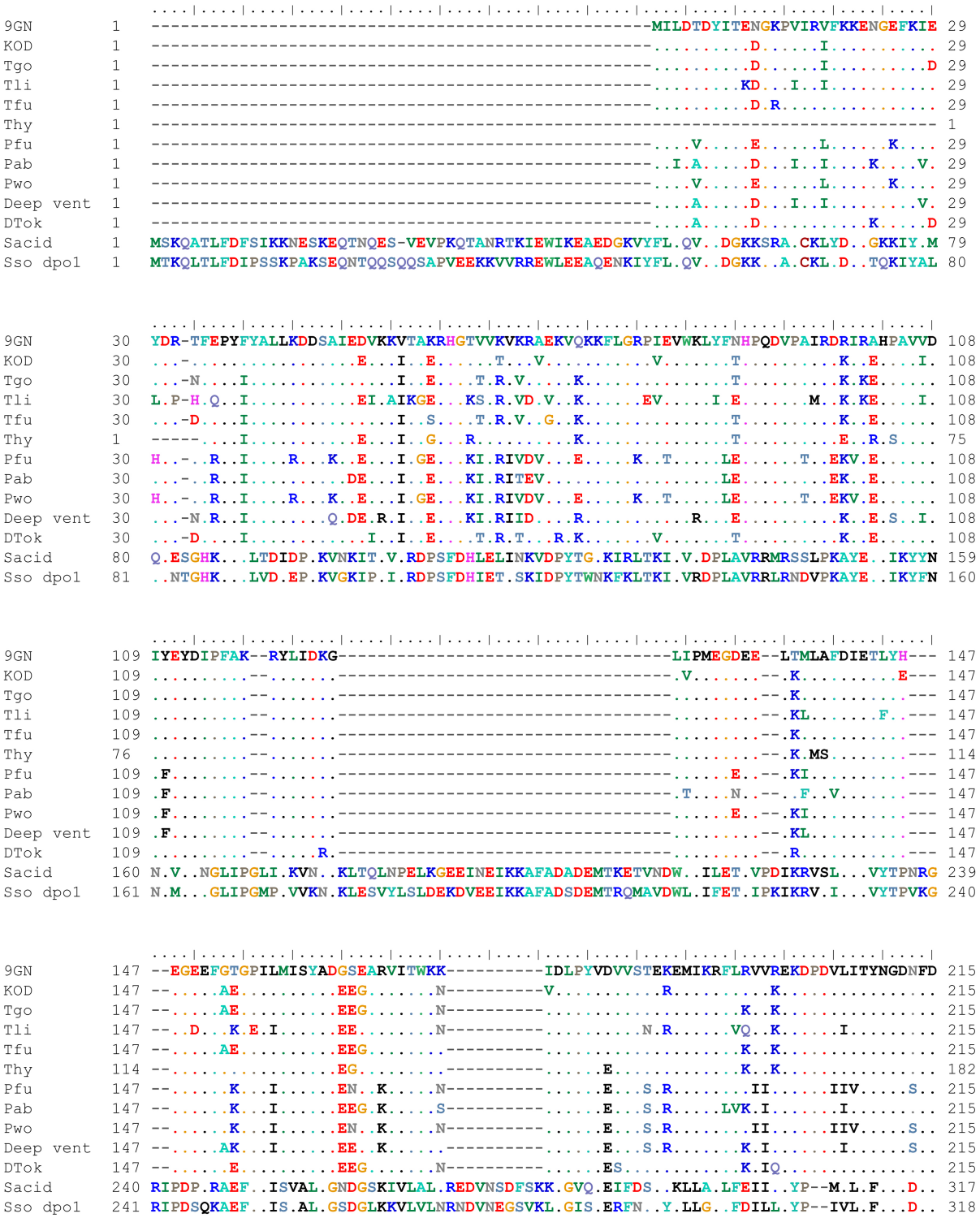


Figure 7.5: Multiple sequence alignment was created using ClustalW and visualized with the software BioEdit. Acronyms of sequences and Uniprot codes are given in Table 7.5. Identical amino acids are indicated by dots and amino acids involved in DNA interaction are shaded by yellow rectangles. *Thermococcus* members: 9°N, KOD, Tgo, Tfu, Thy, Tli; *Pyrococcus* members: Pfu, Pab, Pwo, Deep Vent; *Desulfurococcus* member: D.Tok and *Sulfolobus* members: Sac, Sso dpo1

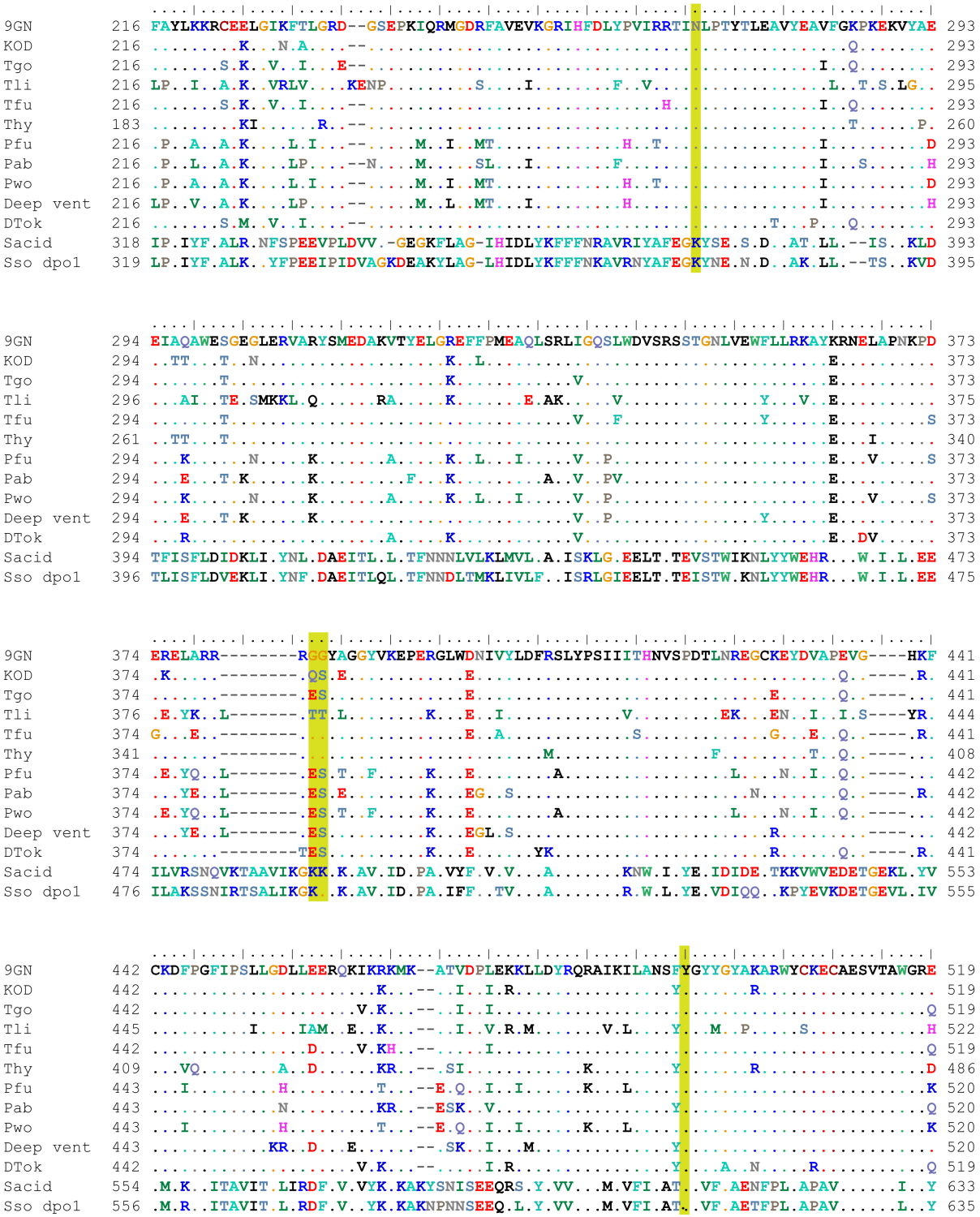


Figure 7.5 (continued)

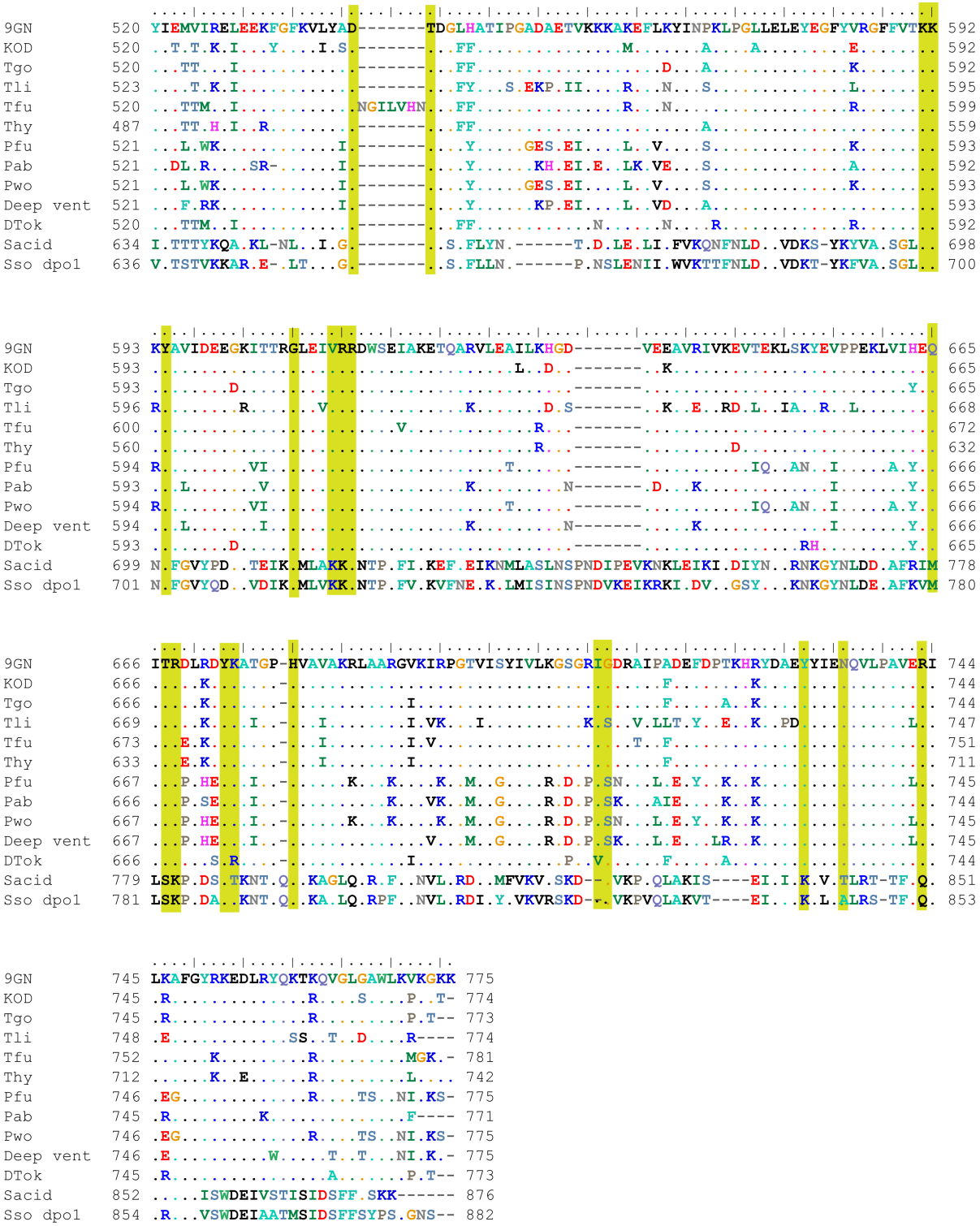


Figure 7.5 (continued)

7.4 Sequences

Nucleic acid and amino acid sequence of KlenTaq (residues 293-832 of the full length *Taq* DNA polymerase, without the starting amino acid methionine). The mutation site of the KlenTaq T664S single mutant is indicated by a yellow rectangle.

```
     ATGGCACTGGAAGAAGCACCTTGGCCTCCGCCTGAAGGTGCATTTGTTGGTTTTGTTCTGAGCCGTAAAG
293   M  A  L  E  E  A  P  W  P  P  P  E  G  A  F  V  G  F  V  L  S  R  K   314

     AACCGATGTGGGCAGATCTGCTGGCACTGGCAGCAGCACGTGGTGGTCGTGTTCATCGTGCACCGGAACC
315  E  P  M  W  A  D  L  L  A  L  A  A  A  R  G  G  R  V  H  R  A  P  E  P 338

     GTATAAAGCTCTGCGCGATCTGAAAGAAGCACGCGGTCTGCTGGCAAAAGATCTGAGCGTTCTGGCACTG
339    Y  K  A  L  R  D  L  K  E  A  R  G  L  L  A  K  D  L  S  V  L  A  L   361

     CGTGAAGGTCTGGGACTGCCTCCGGGTGATGATCCGATGCTGCTGGCATATCTGCTGGATCCGAGCAATA
362   R  E  G  L  G  L  P  P  G  D  D  P  M  L  L  A  Y  L  L  D  P  S  N   384

     CCACACCGGAAGGTGTTGCACGTCGTTATGGTGGTGAATGGACCGAAGAAGCAGGCGAACGCGCAGCACT
385  T  T  P  E  G  V  A  R  R  Y  G  G  E  W  T  E  E  A  G  E  R  A  A  L 408

     GAGCGAACGTCTGTTTGCAAATCTGTGGGGTCGTCTGGAAGGTGAAGAACGTCTGCTGTGGCTGTATCGT
409    S  E  R  L  F  A  N  L  W  G  R  L  E  G  E  E  R  L  L  W  L  Y  R   431

     GAAGTTGAACGTCCGCTGTCTGCAGTTCTGGCACACATGGAAGCAACCGGTGTTCGTCTGGATGTTGCAT
432   E  V  E  R  P  L  S  A  V  L  A  H  M  E  A  T  G  V  R  L  D  V  A   454

     ATCTGCGTGCACTGAGCCTGGAAGTTGCAGAAGAAATTGCACGTCTGGAAGCAGAAGTTTTTCGTCTGGC
455  Y  L  R  A  L  S  L  E  V  A  E  E  I  A  R  L  E  A  E  V  F  R  L  A 478

     AGGCCATCCGTTTAATCTGAATAGCCGTGATCAGCTGGAACGTGTTCTGTTTGATGAACTGGGTCTGCCA
479    G  H  P  F  N  L  N  S  R  D  Q  L  E  R  V  L  F  D  E  L  G  L  P   501

     GCAATTGGTAAAACCGAAAAAACCGGTAAACGTAGCACCAGCGCAGCAGTTCTGGAAGCCCTGCGTGAAG
502   A  I  G  K  T  E  K  T  G  K  R  S  T  S  A  A  V  L  E  A  L  R  E   524

     CACATCCGATTGTGGAAAAAATTCTGCAGTATCGCGAACTGACCAAACTGAAAAGCACCTATATCGATCC
525  A  H  P  I  V  E  K  I  L  Q  Y  R  E  L  T  K  L  K  S  T  Y  I  D  P 548

     GCTGCCGGATCTGATTCATCCGCGTACCGGTCGTCTGCATACCCGTTTTAATCAGACCGCAACCGCAACC
549    L  P  D  L  I  H  P  R  T  G  R  L  H  T  R  F  N  Q  T  A  T  A  T   571

     GGTCGCCTGAGCAGCAGCGATCCGAATCTGCAGAATATTCCGGTTCGTACACCGCTGGGTCAGCGTATTC
572   G  R  L  S  S  S  S  D  P  N  L  Q  N  I  P  V  R  T  P  L  G  Q  R  I   594

     GTCGTGCATTTATTGCAGAAGAAGGTTGGCTGCTGGTTGCACTGGATTATAGCCAGATAGAACTGCGTGT
595  R  R  A  F  I  A  E  E  G  W  L  L  V  A  L  D  Y  S  Q  I  E  L  R  V 618

     TCTGGCCCATCTGAGCGGTGATGAAAATCTGATTCGCGTGTTTCAGGAAGGTCGCGATATTCATACCGAA
619    L  A  H  L  S  G  D  E  N  L  I  R  V  F  Q  E  G  R  D  I  H  T  E   641

     ACCGCAAGCTGGATGTTTGGTGTTCCGCGTGAAGCAGTTGATCCGCTGATGCGTCGTGCAGCAAAAACCA
642   T  A  S  W  M  F  G  V  P  R  E  A  V  D  P  L  M  R  R  A  A  K  T   664
```

```
      TTAATTTTGGCGTGCTGTATGGTATGAGCGCACATCGTCTGAGCCAGGAACTGGCAATTCCGTACGAAGA
665   I  N  F  G  V  L  Y  G  M  S  A  H  R  L  S  Q  E  L  A  I  P  Y  E  E  688

      AGCCCAGGCATTTATCGAACGTTATTTTCAGAGCTTTCCGAAAGTTCGTGCCTGGATTGAAAAAACCCTG
689     A  Q  A  F  I  E  R  Y  F  Q  S  F  P  K  V  R  A  W  I  E  K  T  L  711

      GAAGAAGGTCGTCGTCGCGGTTATGTTGAAACCCTGTTTGGTCGTCGTCGTTATGTTCCGGATCTGGAAG
712   E  E  G  R  R  R  G  Y  V  E  T  L  F  G  R  R  R  Y  V  P  D  L  E    734

      CACGTGTTAAAAGCGTTCGTGAAGCAGCAGAACGTATGGCCTTTAATATGCCGGTTCAGGGCACCGCAGC
735   A  R  V  K  S  V  R  E  A  A  E  R  M  A  F  N  M  P  V  Q  G  T  A  A  758

      AGATCTGATGAAACTGGCCATGGTTAAACTGTTTCCGCGTCTGGAAGAAATGGGTGCACGTATGCTGCTG
759     D  L  M  K  L  A  M  V  K  L  F  P  R  L  E  E  M  G  A  R  M  L  L  781

      CAGGTTCATGATGAACTGGTGCTGGAAGCACCGAAAGAACGTGCAGAAGCAGTTGCCCGTCTGGCAAAAG
782   Q  V  H  D  E  L  V  L  E  A  P  K  E  R  A  E  A  V  A  R  L  A  K    804

      AAGTTATGGAAGGCGTTTATCCGCTGGCAGTTCCGCTGGAAGTTGAAGTTGGTATTGGTGAAGATTGGCT
805   E  V  M  E  G  V  Y  P  L  A  V  P  L  E  V  E  V  G  I  G  E  D  W  L  828

      GTCTGCAAAAGAATAA  832
829     S  A  K  E  *
```

Nucleic acid and amino acid sequence of 9°N full length (amino acids 1-775) and 9°N△25 (amino acids 1-750). Exo- and R465 mutation sites are indicated.

```
     ATGATTCTGGATACCGATTACATTACCGAAAACGGCAAACCGGTGATTCGTGTGTTCAAAAAAGAAAACG
1     M  I  L  D  T  D  Y  I  T  E  N  G  K  P  V  I  R  V  F  K  K  E  N   23

     GCGAATTCAAAATCGAATACGATCGTACCTTTGAACCGTATTTTTACGCGCTGCTGAAAGATGATAGCGC
24   G  E  F  K  I  E  Y  D  R  T  F  E  P  Y  F  Y  A  L  L  K  D  D  S  A 47

     GATCGAAGATGTGAAAAAAGTGACCGCGAAACGTCATGGCACCGTGGTGAAAGTGAAACGTGCGGAAAAA
48     I  E  D  V  K  K  V  T  A  K  R  H  G  T  V  V  K  V  K  R  A  E  K  70

     GTGCAGAAAAAATTTCTGGGCCGTCCGATTGAAGTGTGGAAACTGTATTTCAACCATCCGCAGGATGTGC
71    V  Q  K  K  F  L  G  R  P  I  E  V  W  K  L  Y  F  N  H  P  Q  D  V   93

     CGGCGATTCGTGATCGTATTCGTGCGCATCCGGCGGTGGTGGATATTTATGAATATGATATCCCGTTCGC
94   P  A  I  R  D  R  I  R  A  H  P  A  V  V  D  I  Y  E  Y  D  I  P  F  A 117

     GAAACGTTATCTGATTGATAAAGGCCTGATTCCGATGGAAGGCGATGAAGAACTGACCATGCTGGCCTTT
118    K  R  Y  L  I  D  K  G  L  I  P  M  E  G  D  E  E  L  T  M  L  A  F  140

     GCGATTGCGACCCTGTATCACGAAGGCGAAGAATTTGGCACCGGCCCGATTCTGATGATTAGCTATGCGG
141   A  I  A  T  L  Y  H  E  G  E  E  F  G  T  G  P  I  L  M  I  S  Y  A  163

     ATGGCAGCGAAGCGCGTGTGATTACCTGGAAAAAAATCGATCTGCCGTATGTGGATGTGGTGAGCACCGA
164  D  G  S  E  A  R  V  I  T  W  K  K  I  D  L  P  Y  V  D  V  V  S  T  E 187

     AAAAGAAATGATCAAACGCTTTCTGCGTGTGGTGCGTGAAAAAGATCCGGATGTGCTGATTACCTATAAC
188    K  E  M  I  K  R  F  L  R  V  V  R  E  K  D  P  D  V  L  I  T  Y  N  210

     GGCGATAACTTTGATTTCGCGTATCTGAAAAAACGTTGCGAAGAACTGGGCATCAAATTTACCCTGGGCC
211   G  D  N  F  D  F  A  Y  L  K  K  R  C  E  E  L  G  I  K  F  T  L  G  233

     GTGATGGTAGCGAACCGAAAATTCAGCGTATGGGCGATCGTTTTGCGGTGGAAGTGAAAGGCCGTATTCA
234  R  D  G  S  E  P  K  I  Q  R  M  G  D  R  F  A  V  E  V  K  G  R  I  H 257

     TTTTGATCTGTATCCGGTGATTCGCCGTACCATTAACCTGCCGACCTATACCCTGGAAGCGGTGTATGAA
258    F  D  L  Y  P  V  I  R  R  T  I  N  L  P  T  Y  T  L  E  A  V  Y  E  280

     GCGGTGTTTGGCAAACCGAAAGAAAAAGTGTACGCGGAAGAAATTGCGCAGGCGTGGGAAAGCGGCGAAG
281   A  V  F  G  K  P  K  E  K  V  Y  A  E  E  I  A  Q  A  W  E  S  G  E   303

     GCCTGGAACGTGTGGCGCGTTATAGCATGGAAGATGCGAAAGTGACCTATGAACTGGGCCGTGAATTTTT
304  G  L  E  R  V  A  R  Y  S  M  E  D  A  K  V  T  Y  E  L  G  R  E  F  F 327

     CCCGATGGAAGCGCAGCTGTCTCGTCTGATTGGCCAGAGCCTGTGGGATGTGAGCCGTAGCAGCACCGGC
328    P  M  E  A  Q  L  S  R  L  I  G  Q  S  L  W  D  V  S  R  S  S  T  G  350

     AACCTGGTGGAATGGTTTCTGCTGCGTAAAGCGTATAAACGTAACGAACTGGCCCCGAACAAACCGGATG
351   N  L  V  E  W  F  L  L  R  K  A  Y  K  R  N  E  L  A  P  N  K  P  D  373
```

exo- mutation sites D141A, E143A

```
     AACGTGAACTGGCCCGTCGTCGTGGCGGTTATGCGGGCGGTTATGTGAAAGAACCGGAACGTGGCCTGTG
374  E  R  E  L  A  R  R  R  G  G  Y  A  G  G  Y  V  K  E  P  E  R  G  L  W  397

     GGATAACATTGTGTATCTGGATTTTCGTAGCCTGTATCCGAGCATTATTATCACCCATAACGTGAGCCCG
398    D  N  I  V  Y  L  D  F  R  S  L  Y  P  S  I  I  I  T  H  N  V  S  P  420

     GATACCCTGAACCGTGAAGGCTGCAAAGAATATGATGTGGCGCCGGAAGTGGGCCATAAATTCTGCAAAG
421    D  T  L  N  R  E  G  C  K  E  Y  D  V  A  P  E  V  G  H  K  F  C  K   443

     ATTTCCCGGGCTTTATTCCGAGCCTGCTGGGCGATCTGCTGGAAGAACGCCAGAAAATCAAACGCAAAAT
444  D  F  P  G  F  I  P  S  L  L  G  D  L  L  E  E  R  Q  K  I  K  R  K  M  467

     GAAAGCGACCGTTGATCCGCTGGAAAAAAAACTGCTGGATTATCGTCAGCGCGCTATTAAAATTCTGGCC
468    K  A  T  V  D  P  L  E  K  K  L  L  D  Y  R  Q  R  A  I  K  I  L  A  490

     AACAGCTTCTATGGCTATTATGGTTATGCGAAAGCGCGTTGGTATTGCAAAGAATGCGCGGAAAGCGTGA
491    N  S  F  Y  G  Y  Y  G  Y  A  K  A  R  W  Y  C  K  E  C  A  E  S  V   513

     CCGCGTGGGGCCGTGAATATATCGAAATGGTGATCCGCGAACTGGAAGAAAAATTCGGCTTCAAAGTGCT
514  T  A  W  G  R  E  Y  I  E  M  V  I  R  E  L  E  E  K  F  G  F  K  V  L  537

     GTATGCGGATACCGATGGCCTGCATGCGACCATTCCGGGTGCGGATGCGGAAACCGTGAAAAAAAAAGCG
538    Y  A  D  T  D  G  L  H  A  T  I  P  G  A  D  A  E  T  V  K  K  K  A  560

     AAAGAATTCCTGAAATACATCAATCCGAAACTGCCGGGCCTGCTGGAACTGGAATATGAAGGCTTTTATG
561    K  E  F  L  K  Y  I  N  P  K  L  P  G  L  L  E  L  E  Y  E  G  F  Y   583

     TGCGTGGCTTTTTCGTGACCAAAAAAAAATACGCGGTGATCGATGAAGAAGGCAAAATTACCACCCGTGG
584  V  R  G  F  F  V  T  K  K  K  Y  A  V  I  D  E  E  G  K  I  T  T  R  G  607

     CCTGGAAATTGTGCGTCGTGATTGGAGCGAAATTGCGAAAGAAACCCAGGCGCGTGTGCTGGAAGCGATT
608    L  E  I  V  R  R  D  W  S  E  I  A  K  E  T  Q  A  R  V  L  E  A  I  630

     CTGAAACATGGCGATGTGGAAGAAGCGGTGCGTATTGTTAAAGAAGTGACCGAAAAACTGAGCAAATATG
631    L  K  H  G  D  V  E  E  A  V  R  I  V  K  E  V  T  E  K  L  S  K  Y   653

     AAGTTCCGCCGGAAAAACTGGTGATTCATGAACAAATTACCCGTGATCTGCGTGATTATAAAGCGACCGG
654  E  V  P  P  E  K  L  V  I  H  E  Q  I  T  R  D  L  R  D  Y  K  A  T  G  677

     TCCGCATGTGGCGGTGGCAAAACGTCTGGCAGCGCGTGGCGTGAAAATTCGTCCGGGCACCGTGATTAGC
678    P  H  V  A  V  A  K  R  L  A  A  R  G  V  K  I  R  P  G  T  V  I  S  700

     TATATTGTGCTGAAAGGCAGCGGCCGTATTGGCGATCGTGCGATTCCGGCGGATGAATTTGATCCGACCA
701    Y  I  V  L  K  G  S  G  R  I  G  D  R  A  I  P  A  D  E  F  D  P  T   723

     AACATCGTTATGATGCGGAATATTATATCGAAAACCAGGTGCTGCCGGCGGTGGAACGTATTCTGAAAGC
724  K  H  R  Y  D  A  E  Y  Y  I  E  N  Q  V  L  P  A  V  E  R  I  L  K  A  747

     GTTTGGCTATCGTAAAGAAGATCTGCGCTATCAGAAAACCAAACAGGTGGGCCTGGGCGCGTGGCTGAAA
748    F  G  Y  R  K  E  D  L  R  Y  Q  K  T  K  Q  V  G  L  G  A  W  L  K  770

     GTTAAAGGCAAAAAA 775
771    V  K  G  K  K
```

end of 9°N△25 sequence

R465 - site of mutations R465A and R465E

References

[1] P. S. Ho and M. Carter, *DNA structure: alphabet soup for the cellular soul.* InTech, 2011.

[2] G. Sun, J. H. Voigt, I. V. Filippov, V. E. Marquez, and M. C. Nicklaus, "PROSIT: Pseudorotational online service and interactive tool, applied to a conformational survey of nucleosides and nucleotides," *J Chem Inf Comput Sci*, vol. 44, no. 5, pp. 1752–1762, 2004.

[3] A. Ghosh and M. Bansal, "A glossary of DNA structures from A to Z," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 59, pp. 620–626, Apr 2003.

[4] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, "Crystal structure analysis of a complete turn of B-DNA," *Nature*, vol. 287, pp. 755–758, Oct. 1980.

[5] J. C. Wang, "Helical repeat of DNA in solution," *Proc Natl Acad Sci U S A*, vol. 76, no. 1, pp. 200–203, 1979.

[6] J. D. Watson, F. H. Crick, *et al.*, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[7] A. Baccaro, A.-L. Steck, and A. Marx, "Barcoded nucleotides," *Angew Chem Int Ed*, vol. 51, no. 1, pp. 254–257, 2012.

[8] S. Obeid, A. Baccaro, W. Welte, K. Diederichs, and A. Marx, "Structural basis for the synthesis of nucleobase modified DNA by *Thermus aquaticus* DNA polymerase," *Proc Natl Acad Sci U S A*, vol. 107, no. 50, pp. 21327–21331, 2010.

[9] K. Bergen, A.-L. Steck, S. Strütt, A. Baccaro, W. Welte, K. Diederichs, and A. Marx, "Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates," *J Am Chem Soc*, vol. 134, no. 29, pp. 11840–11843, 2012.

[10] V. Borsenberger, M. Kukwikila, and S. Howorka, "Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates," *Org Biomol Chem*, vol. 7, no. 18, pp. 3826–3835, 2009.

[11] A. R. Kore, "Solid-phase synthesis of new ribo and deoxyribo BrdU probes for labeling and detection of nucleic acids," *Tetrahedron Lett*, vol. 50, no. 7, pp. 793 – 795, 2009.

[12] P. M. Gramlich, C. T. Wirges, A. Manetto, and T. Carell, "Postsynthetic DNA modification through the copper-catalyzed azide–alkyne cycloaddition reaction," *Angew Chem Int Ed*, vol. 47, no. 44, pp. 8350–8358, 2008.

[13] M. Hocek and M. Fojta, "Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. construction and applications of base-functionalized nucleic acids," *Org*

Biomol Chem, vol. 6, no. 13, pp. 2233–2241, 2008.

[14] G. F. Kaufmann, M. M. Meijler, C. Sun, D.-W. Chen, D. P. Kujawa, J. M. Mee, T. Z. Hoffman, P. Wirsching, R. A. Lerner, and K. D. Janda, "Enzymatic incorporation of an antibody-activated blue fluorophore into DNA," *Angew Chem Int Ed*, vol. 44, no. 14, pp. 2144–2148, 2005.

[15] J. P. Anderson, B. Angerer, and L. A. Loeb, "Incorporation of reporter-labeled nucleotides by DNA polymerases," *Biotechniques*, vol. 38, no. 2, pp. 257–264, 2005.

[16] O. Thum, S. Jäger, and M. Famulok, "Functionalized DNA: a new replicable biopolymer," *Angew Chem Int Ed*, vol. 40, no. 21, pp. 3990–3993, 2001.

[17] F. Seela and M. Zulauf, "7-deazaadenine-DNA: Bulky 7-iodo substituents or hydrophobic 7-hexynyl chains are well accommodated in the major groove of oligonucleotide duplexes," *Chem Eur J*, vol. 4, no. 9, pp. 1781–1790, 1998.

[18] R. Dickerson, H. Drew, B. Conner, R. Wing, A. Fratini, and M. Kopka, "The anatomy of A-, B-, and Z-DNA," *Science*, vol. 216, no. 4545, pp. 475–485, 1982.

[19] S. Neidle, *Principles of nucleic acid structure*. Academic Press, 2010.

[20] O. Julien, J. R. Beadle, W. C. Magee, S. Chatterjee, K. Y. Hostetler, D. H. Evans, and B. D. Sykes, "Solution structure of a dna duplex containing the potent anti-poxvirus agent cidofovir," *Journal of the American Chemical Society*, vol. 133, no. 7, pp. 2264–2274, 2011. PMID: 21280608.

[21] F. A. Hays, A. Teegarden, Z. J. R. Jones, M. Harms, D. Raup, J. Watson, E. Cavaliere, and P. S. Ho, "How sequence defines structure: A crystallographic map of DNA structure and conformation," *Proc Natl Acad Sci U S A*, vol. 102, no. 20, pp. 7157–7162, 2005.

[22] S. Clancy, "RNA splicing: introns, exons and spliceosome," *Nature Education*, vol. 1, no. 1, p. 31, 2008.

[23] I. Lehman, "Discovery of DNA polymerase," *J Biol Chem*, vol. 278, no. 37, pp. 34733–34738, 2003.

[24] I. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid," *J biol Chem*, vol. 233, pp. 163–170, 1958.

[25] I. Lehman, S. B. Zimmerman, J. Adler, M. J. Bessman, E. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. V. Chemical composition of enzymatically synthesized deoxyribonucleic acid," *Proc Natl Acad Sci U S A*, vol. 44, no. 12, p. 1191, 1958.

[26] D. Ollis, P. Brick, R. Hamlin, N. Xuong, and T. Steitz, "Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP," *Nature*, vol. 313, pp. 762–766, 1985.

[27] T. A. Steitz, "Structural biology: A mechanism for all polymerases," *Nature*, vol. 391, no. 6664, pp. 231–232, 1998.

[28] A. J. Berdis, "Mechanisms of DNA polymerases," *Chem Rev*, vol. 109, no. 7, pp. 2862–2879, 2009.

[29] A. M. Breier, H.-U. G. Weier, and N. R. Cozzarelli, "Independence of replisomes in *Escherichia coli* chromosomal replication," *Proc Natl Acad Sci U S A*, vol. 102, no. 11, pp. 3942–3947, 2005.

[30] W. L. Fangman and B. J. Brewer, "A question of time: replication origins of eukaryotic chromo-

somes," *Cell*, vol. 71, no. 3, pp. 363–366, 1992.

[31] U. Huebscher and G. Maga, "DNA replication and repair bypass machines," *Curr Opin Chem Biol*, vol. 15, pp. 627–635, Oct. 2011.

[32] P. J. Rothwell and G. Waksman, "Structure and mechanism of DNA polymerases," *Adv Protein Chem*, vol. 71, pp. 401–440, 2005.

[33] S. D. McCulloch and T. A. Kunkel, "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases," *Cell Res*, vol. 18, no. 1, pp. 148–161, 2008.

[34] K. Bebenek, A. Tissier, E. G. Frank, J. P. McDonald, R. Prasad, S. H. Wilson, R. Woodgate, and T. A. Kunkel, "5'-deoxyribose phosphate lyase activity of human DNA polymerase ι in vitro," *Science*, vol. 291, no. 5511, pp. 2156–2159, 2001.

[35] D. K. Braithwaite and J. Ito, "Compilation, alignment, and phylogenetic relationships of DNA polymerases.," *Nucleic Acids Res*, vol. 21, no. 4, p. 787, 1993.

[36] J. Yamtich and J. B. Sweasy, "DNA polymerase family X: Function, structure, and cellular roles," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, no. 5, pp. 1136 – 1150, 2010.

[37] A. R. Lehmann, A. Niimi, T. Ogi, S. Brown, S. Sabbioneda, J. F. Wing, P. L. Kannouche, and C. M. Green, "Translesion synthesis: Y-family polymerases and the polymerase switch," *DNA Repair*, vol. 6, no. 7, pp. 891 – 899, 2007.

[38] T. Matamoros, M. Alvarez, V. Barrioluengo, G. Betancor, and L. Menéndez-Arias, *Reverse Transcriptase and Retroviral Replication, DNA Replication and Related Cellular Processes.* Dr. Jelena Kusic-Tisma (Ed.), ISBN: 978-953-307-775-8, InTech, DOI: 10.5772/21660, 2011.

[39] M. J. Bessman, A. Kornberg, I. Lehman, and E. Simms, "Enzymatic synthesis of deoxyribonucleic acid," *Biochim Biophys Acta*, vol. 21, pp. 197–198, 1956.

[40] H. Klenow and I. Henningsen, "Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis," *Proc Natl Acad Sci U S A*, vol. 65, no. 1, pp. 168–175, 1970.

[41] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich, "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase," *Science*, vol. 239, no. 4839, pp. 487–491, 1988.

[42] L. Beese, V. Derbyshire, and T. Steitz, "Structure of dna polymerase i klenow fragment bound to duplex dna," *Science*, vol. 260, no. 5106, pp. 352–355, 1993.

[43] Y. Kim, S. H. Eom, J. Wang, D.-S. Lee, S. W. Suh, and T. A. Steitz, "Crystal structure of *Thermus aquaticus* DNA polymerase," *Nature*, vol. 376, pp. 612–616, 1995.

[44] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, and A. Marx, "KlenTaq polymerase replicates unnatural base pairs by inducing a watson-crick geometry," *Nat Chem Biol*, vol. 8, no. 7, pp. 612–614, 2012.

[45] T. D. Brock and H. Freeze, "*Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile," *J Bacteriol*, vol. 98, no. 1, pp. 289–297, 1969.

[46] A. Chien, D. B. Edgar, and J. M. Trela, "Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*.," *J Bacteriol*, vol. 127, no. 3, pp. 1550–1557, 1976.

[47] F. C. Lawyer, S. Stoffel, R. Saiki, S. Chang, P. Landre, R. Abramson, and D. Gelfand, "High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity.," *Genome Res*, vol. 2, no. 4, pp. 275–287, 1993.

[48] S. Korolev, M. Nayal, W. M. Barnes, E. Di Cera, and G. Waksman, "Crystal structure of the large fragment of thermus aquaticus DNA polymerase I at 2.5Å resolution: structural basis for thermostability," *Proc Natl Acad Sci U S A*, vol. 92, no. 20, pp. 9264–9268, 1995.

[49] J. R. Kiefer, C. Mao, C. J. Hansen, S. L. Basehore, H. H. Hogrefe, J. C. Braman, and L. S. Beese, "Crystal structure of a thermostable *Bacillus* DNA polymerase l large fragment at 2.1Å resolution," *Structure*, vol. 5, pp. 95–108, Jan. 1997.

[50] R. N. Veedu, B. Vester, and J. Wengel, "Enzymatic incorporation of LNA nucleotides into DNA strands," *Chembiochem*, vol. 8, no. 5, pp. 490–492, 2007.

[51] R. N. Veedu, B. Vester, and J. Wengel, "Polymerase chain reaction and transcription using locked nucleic acid nucleotide triphosphates," *J Am Chem Soc*, vol. 130, no. 26, pp. 8124–8125, 2008.

[52] M. Strerath, J. Gaster, D. Summerer, and A. Marx, "Increased single-nucleotide discrimination of PCR by primer probes bearing hydrophobic 4'C modifications," *Chembiochem*, vol. 5, no. 3, pp. 333–339, 2004.

[53] H. Sawai, A. N. Ozaki, F. Satoh, T. Ohbayashi, M. M. Masud, and H. Ozaki, "Expansion of structural and functional diversities of DNA using new 5-substituted deoxyuridine derivatives by PCR with superthermophilic KOD Dash dna polymerase," *Chem Commun*, no. 24, pp. 2604–2605, 2001.

[54] H. A. Held and S. A. Benner, "Challenging artificial genetic systems: thymidine analogs with 5-position sulfur functionality," *Nucleic Acids Res*, vol. 30, no. 17, pp. 3857–3869, 2002.

[55] G. A. Burley, J. Gierlich, M. R. Mofid, H. Nir, S. Tal, Y. Eichen, and T. Carell, "Directed DNA metallization," *J Am Chem Soc*, vol. 128, no. 5, pp. 1398–1399, 2006.

[56] J. Gierlich, K. Gutsmiedl, P. M. Gramlich, A. Schmidt, G. A. Burley, and T. Carell, "Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry," *Chem Eur J*, vol. 13, no. 34, pp. 9486–9494, 2007.

[57] A. C. Rodriguez, H.-W. Park, C. Mao, and L. S. Beese, "Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9°N n-7," *J Mol Biol*, vol. 299, no. 2, pp. 447–462, 2000.

[58] H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, and Y. Kai, "Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1,"

J Mol Biol, vol. 306, no. 3, pp. 469–477, 2001.

[59] T. Kuroita, H. Matsumura, N. Yokota, M. Kitabayashi, H. Hashimoto, T. Inoue, T. Imanaka, and Y. Kai, "Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases," *J Mol Biol*, vol. 351, no. 2, pp. 291–298, 2005.

[60] Y. Li, S. Dutta, S. Doublié, H. Moh'd Bdour, J.-S. Taylor, and T. Ellenberger, "Nucleotide insertion opposite a cis-syn thymine dimer by a replicative DNA polymerase from bacteriophage T7," *Nat Struct Mol Biol*, vol. 11, no. 8, pp. 784–790, 2004.

[61] M. C. Franklin, J. Wang, and T. A. Steitz, "Structure of the replicating complex of a pol α family DNA polymerase," *Cell*, vol. 105, no. 5, pp. 657–667, 2001.

[62] J. H. Wong, K. A. Fiala, Z. Suo, and H. Ling, "Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4," *J Mol Biol*, vol. 379, no. 2, pp. 317–330, 2008.

[63] H. Ling, F. Boudsocq, R. Woodgate, and W. Yang, "Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication," *Cell*, vol. 107, no. 1, pp. 91 – 102, 2001.

[64] M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, and J. Kraut, "Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism," *Science*, vol. 264, no. 5167, pp. 1930–1935, 1994.

[65] H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut, "Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP," *Science*, vol. 264, no. 5167, pp. 1891–1903, 1994.

[66] J. Ding, K. Das, Y. Hsiou, S. G. Sarafianos, A. D. Clark Jr, A. Jacobo-Molina, C. Tantillo, S. H. Hughes, and E. Arnold, "Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution," *J Mol Biol*, vol. 284, no. 4, pp. 1095–1111, 1998.

[67] S. Bailey, R. A. Wing, and T. A. Steitz, "The structure of T. aquaticus DNA polymerase III is distinct from eukaryotic replicative DNA polymerases," *Cell*, vol. 126, no. 5, pp. 893 – 904, 2006.

[68] R. A. Wing, S. Bailey, and T. A. Steitz, "Insights into the replisome from the structure of a ternary complex of the DNA polymerase III subunit," *J Mol Biol*, vol. 382, pp. 859–869, Oct. 2008.

[69] I. Matsui, Y. Urushibata, Y. Shen, E. Matsui, and H. Yokoyama, "Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase d large subunit from *Pyrococcus horikoshii*," *FEBS Lett*, vol. 585, no. 3, pp. 452–458, 2011.

[70] B. A. Maxwell, C. Xu, and Z. Suo, "Conformational dynamics of a Y-family DNA polymerase during substrate binding and catalysis as revealed by interdomain Förster resonance energy transfer," *Biochemistry*, vol. 53, no. 11, pp. 1768–1778, 2014.

[71] Y. Li, S. Korolev, and G. Waksman, "Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis

for nucleotide incorporation," *EMBO J*, vol. 17, no. 24, pp. 7514–7525, 1998.

[72] J. Wang, A. Sattar, C. Wang, J. Karam, W. Konigsberg, and T. Steitz, "Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69," *Cell*, vol. 89, no. 7, pp. 1087 – 1099, 1997.

[73] M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut, and H. Pelletier, "Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism," *Biochemistry*, vol. 36, no. 37, pp. 11205–11215, 1997.

[74] H. Huang, R. Chopra, G. L. Verdine, and S. C. Harrison, "Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance," *Science*, vol. 282, no. 5394, pp. 1669–1675, 1998.

[75] L. Kohlstaedt, J. Wang, J. Friedman, P. Rice, and T. Steitz, "Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor," *Science*, vol. 256, no. 5065, pp. 1783–1790, 1992.

[76] A. Jacobo-Molina, J. Ding, R. G. Nanni, A. D. Clark, X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L. Ferris, and P. Clark, "Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0Å resolution shows bent DNA," *Proc Natl Acad Sci U S A*, vol. 90, no. 13, pp. 6320–6324, 1993.

[77] J. D. Pata, "Structural diversity of the Y-family DNA polymerases," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, no. 5, pp. 1124 – 1135, 2010.

[78] W. Wang, H. W. Hellinga, and L. S. Beese, "Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis," *Proc Natl Acad Sci U S A*, vol. 108, no. 43, pp. 17644–17648, 2011.

[79] C.-C. Wang, A. Pavlov, and J. D. Karam, "Evolution of RNA-binding specificity in T4 DNA polymerase," *J Biol Chem*, vol. 272, no. 28, pp. 17703–17710, 1997.

[80] C. A. Brautigam and T. A. Steitz, "Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes," *Curr Opin Struc Biol*, vol. 8, no. 1, pp. 54 – 63, 1998.

[81] P. H. Patel and L. A. Loeb, "Getting a grip on how DNA polymerases function," *Nat Struct Mol Biol*, vol. 8, pp. 656–659, Aug. 2001.

[82] G. Yang, M. Franklin, J. Li, T.-C. Lin, and W. Konigsberg, "Correlation of the kinetics of finger domain mutants in RB69 DNA polymerase with its structure," *Biochemistry*, vol. 41, no. 8, pp. 2526–2534, 2002. PMID: 11851399.

[83] M. Delarue, O. Poch, N. Tordo, D. Moras, and P. Argos, "An attempt to unify the structure of polymerases," *Protein Eng*, vol. 3, no. 6, pp. 461–467, 1990.

[84] P. Palanivelu, "DNA polymerases: An insight into their active sites and catalytic mechanism," *International Journal of Biochemistry Research & Review*, vol. 3, no. 3, 2013.

[85] S. Wu, W. A. Beard, L. G. Pedersen, and S. H. Wilson, "Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site," *Chem Rev*, vol. 114, no. 5,

pp. 2759–2774, 2014.

[86] K. A. Johnson, "The kinetic and chemical mechanism of high-fidelity DNA polymerases," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, pp. 1041–1048, May 2010.

[87] K. Betz, F. Streckenbach, A. Schnur, T. Exner, W. Welte, K. Diederichs, and A. Marx, "Structures of DNA polymerases caught processing size-augmented nucleotide probes," *Angew Chem Int Ed*, vol. 49, no. 30, pp. 5181–5184, 2010.

[88] Y.-S. Lee, W. D. Kennedy, and Y. W. Yin, "Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations," *Cell*, vol. 139, no. 2, pp. 312 – 324, 2009.

[89] P. J. Rothwell, V. Mitaksov, and G. Waksman, "Motions of the fingers subdomain of Klentaq1 are fast and not rate limiting: Implications for the molecular basis of fidelity in DNA polymerases," *Molecular Cell*, vol. 19, no. 3, pp. 345 – 355, 2005.

[90] E. Y. Wu and L. S. Beese, "The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an 'ajar' intermediate conformation in the nucleotide selection mechanism," *J Biol Chem*, vol. 286, no. 22, pp. 19758–19767, 2011.

[91] M. Wang, S. Xia, G. Blaha, T. A. Steitz, W. H. Konigsberg, and J. Wang, "Insights into base selectivity from the 1.8Å resolution structure of an RB69 DNA polymerase ternary complex," *Biochemistry*, vol. 50, no. 4, pp. 581–590, 2011.

[92] E. Freisinger, A. P. Grollman, H. Miller, and C. Kisker, "Lesion (in) tolerance reveals insights into DNA replication fidelity," *EMBO J*, vol. 23, no. 7, pp. 1494–1505, 2004.

[93] K. E. Zahn, H. Belrhali, S. S. Wallace, and S. Doublié, "Caught bending the A-rule: crystal structures of translesion DNA synthesis with a non-natural nucleotide," *Biochemistry*, vol. 46, no. 37, pp. 10551–10561, 2007.

[94] P. Aller, M. A. Rould, M. Hogg, S. S. Wallace, and S. Doublie, "A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol," *Proc Natl Acad Sci U S A*, vol. 104, no. 3, pp. 814–818, 2007.

[95] P. Aller, S. Duclos, S. S. Wallace, and S. Doublié, "A crystallographic study of the role of sequence context in thymine glycol bypass by a replicative DNA polymerase serendipitously sheds light on the exonuclease complex," *J Mol Biol*, vol. 412, no. 1, pp. 22–34, 2011.

[96] S. Xia, J. Wang, and W. H. Konigsberg, "DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase," *J Am Chem Soc*, vol. 135, no. 1, pp. 193–202, 2012.

[97] S. Xia, W. H. Konigsberg, and J. Wang, "Hydrogen-bonding capability of a templating difluorotoluene nucleotide residue in an RB69 DNA polymerase ternary complex," *J Am Chem Soc*, vol. 133, no. 26, pp. 10003–10005, 2011.

[98] S. Xia, T. D. Christian, J. Wang, and W. H. Konigsberg, "Probing minor groove hydrogen bonding interactions between RB69 DNA polymerase and DNA," *Biochemistry*, vol. 51, no. 21, pp. 4343–4353, 2012.

[99] R. L. Perera, R. Torella, S. Klinge, M. L. Kilkenny, J. D. Maman, and L. Pellegrini, "Mechanism for priming DNA synthesis by yeast DNA polymerase α," *eLife*, vol. 2, 2013.

[100] M. K. Swan, R. E. Johnson, L. Prakash, S. Prakash, and A. K. Aggarwal, "Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ," *Nat Struct Mol Biol*, vol. 16, pp. 979–986, Sept. 2009.

[101] M. Hogg, P. Osterman, G. O. Bylund, R. A. Ganai, E.-B. Lundström, A. E. Sauer-Eriksson, and E. Johansson, "Structural basis for processive dna synthesis by yeast DNA polymerase ε," *Nat Struct Mol Biol*, vol. 21, pp. 49–55, Jan. 2014.

[102] K.-P. Hopfner, A. Eichinger, R. A. Engh, F. Laue, W. Ankenbauer, R. Huber, and B. Angerer, "Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*," *Proc Natl Acad Sci U S A*, vol. 96, no. 7, pp. 3600–3605, 1999.

[103] S. W. Kim, D.-U. Kim, J. K. Kim, L.-W. Kang, and H.-S. Cho, "Crystal structure of Pfu, the high fidelity DNA polymerase from *Pyrococcus furiosus*," *Int J Biol Macromol*, vol. 42, no. 4, pp. 356–361, 2008.

[104] C. Savino, L. Federici, K. A. Johnson, B. Vallone, V. Nastopoulos, M. Rossi, F. M. Pisani, and D. Tsernoglou, "Insights into DNA replication: The crystal structure of DNA polymerase B1 from the archaeon *Sulfolobus solfataricus*," *Structure*, vol. 12, no. 11, pp. 2001 – 2008, 2004.

[105] Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, and J. Kuriyan, "Crystal structure of an archaebacterial DNA polymerase," *Structure*, vol. 7, no. 10, pp. 1189–1199, 1999.

[106] S. J. Firbank, J. Wardle, P. Heslop, R. J. Lewis, and B. A. Connolly, "Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography," *J Mol Biol*, vol. 381, no. 3, pp. 529–539, 2008.

[107] T. Killelea, S. Ghosh, S. S. Tan, P. Heslop, S. J. Firbank, E. T. Kool, and B. A. Connolly, "Probing the interaction of archaeal DNA polymerases with deaminated bases using X-ray crystallography and non-hydrogen bonding isosteric base analogues," *Biochemistry*, vol. 49, no. 27, pp. 5772–5781, 2010.

[108] J. Gouge, C. Ralec, G. Henneke, and M. Delarue, "Molecular recognition of canonical and deaminated bases by *p. abyssi* family b dna polymerase," *J Mol Biol*, 2012.

[109] L. Blanco, A. Bernad, and M. Salas, "Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases," *Gene*, vol. 112, no. 1, pp. 139 – 144, 1992.

[110] *DNA polymerases*. World Scientific Publishing Co. Pte. Ltd., 2010.

[111] W. A. Beard, D. D. Shock, B. J. Vande Berg, and S. H. Wilson, "Efficiency of correct nucleotide insertion governs DNA polymerase fidelity," *J Biol Chem*, vol. 277, no. 49, pp. 47393–47398, 2002.

[112] S.-J. Kim, W. A. Beard, J. Harvey, D. D. Shock, J. R. Knutson, and S. H. Wilson, "Rapid segmental and subdomain motions of DNA polymerase β," *J Biol Chem*, vol. 278, no. 7, pp. 5072–5081, 2003.

[113] T. A. Kunkel and K. Bebenek, "DNA replication fidelity 1," *Annu Rev Biochem*, vol. 69, no. 1,

pp. 497–529, 2000.

[114] U. Hübscher, G. Maga, and S. Spadari, "Eukaryotic DNA polymerases," *Annu Rev Biochem*, vol. 71, no. 1, pp. 133–163, 2002.

[115] E. A. Motea, I. Lee, and A. J. Berdis, "Insights into the roles of desolvation and π-electron interactions during DNA polymerization," *Chembiochem*, vol. 14, no. 4, pp. 489–498, 2013.

[116] E. T. Kool, "Active site tightness and substrate fit in DNA replication," *Annu Rev Biochem*, vol. 71, no. 1, pp. 191–219, 2002.

[117] J. Poater, M. Swart, C. F. Guerra, and F. M. Bickelhaupt, "Selectivity in DNA replication. Interplay of steric shape, hydrogen bonds, π-stacking and solvent effects," *Chem Commun*, vol. 47, pp. 7326–7328, 2011.

[118] J. D. WATSON and F. H. C. Crick, "Genetical implications of the structure of deoxyribonucleic acid," *Nature*, vol. 171, pp. 964–967, May 1953.

[119] J. Petruska, M. F. Goodman, M. S. Boosalis, L. C. Sowers, C. Cheong, and I. Tinoco, "Comparison between DNA melting thermodynamics and DNA polymerase fidelity," *Proc Natl Acad Sci U S A*, vol. 85, no. 17, pp. 6252–6256, 1988.

[120] H. Echols and M. F. Goodman, "Fidelity mechanisms in DNA replication," *Annu Rev Biochem*, vol. 60, no. 1, pp. 477–511, 1991.

[121] T. J. Matray and E. T. Kool, "A specific partner for abasic damage in DNA," *Nature*, vol. 399, pp. 704–708, June 1999.

[122] S. Xia, S. H. Eom, W. H. Konigsberg, and J. Wang, "Structural basis for differential insertion kinetics of dNMPs opposite a difluorotoluene nucleotide residue," *Biochemistry*, vol. 51, no. 7, pp. 1476–1485, 2012.

[123] E. T. Kool and H. O. Sintim, "The difluorotoluene debate - a decade later," *Chem Commun*, no. 35, pp. 3665–3675, 2006.

[124] S. Moran, R. X.-F. Ren, and E. T. Kool, "A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity," *Proc Natl Acad Sci U S A*, vol. 94, no. 20, pp. 10506–10511, 1997.

[125] S. Moran, R. X.-F. Ren, S. Rumney IV, and E. T. Kool, "Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication," *J Am Chem Soc*, vol. 119, no. 8, pp. 2056–2057, 1997.

[126] D. Liu, S. Moran, and E. T. Kool, "Bi-stranded, multisite replication of a base pair between difluorotoluene and adenine: confirmation by inverse sequencing," *Chemistry & Biology*, vol. 4, no. 12, pp. 919 – 926, 1997.

[127] J. C. Morales and E. T. Kool, "Varied molecular interactions at the active sites of several DNA polymerases: Nonpolar nucleoside isosteres as probes," *J Am Chem Soc*, vol. 122, pp. 1001–1007, Feb. 2000.

[128] W. Wang, E. Y. Wu, H. W. Hellinga, and L. S. Beese, "Structural factors that determine selectivity of a high fidelity dna polymerase for deoxy-, dideoxy-, and ribonucleotides," *J Biol Chem*, vol. 287, no. 34, pp. 28215–28226, 2012.

[129] J. R. Kiefer, C. Mao, J. C. Braman, and L. S. Beese, "Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal," *Nature*, vol. 391, pp. 304–307, Jan. 1998.

[130] S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger, "Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution," *Nature*, vol. 391, no. 6664, pp. 251–258, 1998.

[131] T. E. Spratt, "Identification of hydrogen bonds between *Escherichia coli* DNA polymerase I (Klenow fragment) and the minor groove of DNA by amino acid substitution of the polymerase and atomic substitution of the DNA," *Biochemistry*, vol. 40, no. 9, pp. 2647–2652, 2001. PMID: 11258875.

[132] K. Singh and M. J. Modak, "Presence of 18-Å long hydrogen bond track in the active site of *Escherichia coli* DNA polymerase I (klenow fragment). Its requirement in the stabilization of enzyme-template-primer complex," *J Biol Chem*, vol. 278, no. 13, pp. 11289–11302, 2003.

[133] I. Hirao and M. Kimoto, "Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma," *Proc Jpn Acad, Ser B*, vol. 88, no. 7, p. 345, 2012.

[134] M. D. McCain, A. S. Meyer, S. S. Schultz, A. Glekas, and T. E. Spratt, "Fidelity of mispair formation and mispair extension is dependent on the interaction between the minor groove of the primer terminus and Arg668 of DNA polymerase I of *Escherichia coli*," *Biochemistry*, vol. 44, no. 15, pp. 5647–5659, 2005. PMID: 15823023.

[135] J. C. Morales and E. T. Kool, "Minor groove interactions between polymerase and DNA: More essential to replication than Watson-Crick hydrogen bonds?," *J Am Chem Soc*, vol. 121, no. 10, pp. 2323–2324, 1999.

[136] J. C. Morales and E. T. Kool, "Functional hydrogen-bonding map of the minor groove binding tracks of six DNA polymerases," *Biochemistry*, vol. 39, no. 42, pp. 12979–12988, 2000.

[137] C. L. Hendrickson, K. G. Devine, and S. A. Benner, "Probing minor groove recognition contacts by DNA polymerases and reverse transcriptases using 3-deaza-2'-deoxyadenosine," *Nucleic Acids Res*, vol. 32, no. 7, pp. 2241–2250, 2004.

[138] E. T. Kool, "Replication of non-hydrogen bonded bases by DNA polymerases: a mechanism for steric matching," *Biopolymers*, vol. 48, no. 1, pp. 3–17, 1998.

[139] S. Bommarito, N. Peyret, and J. S. Jr, "Thermodynamic parameters for DNA sequences with dangling ends," *Nucleic Acids Res*, vol. 28, no. 9, pp. 1929–1934, 2000.

[140] K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi, and E. T. Kool, "Factors contributing to aromatic stacking in water: Evaluation in the context of DNA," *J Am Chem Soc*, vol. 122, no. 10, pp. 2213–2222, 2000.

[141] J. A. Brown and Z. Suo, "Unlocking the sugar steric gate of DNA polymerases," *Biochemistry*, vol. 50, no. 7, pp. 1135–1142, 2011.

[142] T. Traut, "Physiological concentrations of purines and pyrimidines," *Mol Cell Biochem*, vol. 140, no. 1, pp. 1–22, 1994.

[143] P. Ferraro, E. Franzolin, G. Pontarin, P. Reichard, and V. Bianchi, "Quantitation of cellular deoxynucleoside triphosphates," *Nucleic Acids Res*, vol. 38, no. 6, p. e85, 2010.

[144] S. A. Nick McElhinny, B. E. Watts, D. Kumar, D. L. Watt, E.-B. Lundström, P. M. J. Burgers, E. Johansson, A. Chabes, and T. A. Kunkel, "Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases," *Proc Natl Acad Sci U S A*, vol. 107, no. 11, pp. 4949–4954, 2010.

[145] C. M. Joyce, "Choosing the right sugar: How polymerases select a nucleotidesubstrate," *Proc Natl Acad Sci U S A*, vol. 94, no. 5, pp. 1619–1622, 1997.

[146] M. Astatke, K. Ng, N. D. F. Grindley, and C. M. Joyce, "A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides," *Proc Natl Acad Sci U S A*, vol. 95, no. 7, pp. 3402–3407, 1998.

[147] P. H. Patel and L. A. Loeb, "Multiple amino acid substitutions allow DNA polymerases to synthesize RNA," *J Biol Chem*, vol. 275, no. 51, pp. 40266–40272, 2000.

[148] F. Wojciechowski and C. J. Leumann, "Alternative DNA base-pairs: from efforts to expand the genetic code to potential material applications," *Chem Soc Rev*, vol. 40, no. 12, pp. 5669–5679, 2011.

[149] C. Switzer, S. E. Moroney, and S. A. Benner, "Enzymatic incorporation of a new base pair into DNA and RNA," *J Am Chem Soc*, vol. 111, no. 21, pp. 8322–8323, 1989.

[150] J. A. Piccirilli, T. Krauch, S. E. Moroney, and S. A. Benner, "Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet.," *Nature*, vol. 343, no. 6253, pp. 33–37, 1990.

[151] I. Hirao, M. Kimoto, and R. Yamashige, "Natural versus artificial creation of base pairs in DNA: origin of nucleobases from the perspectives of unnatural base pair studies," *Acc Chem Res*, vol. 45, no. 12, pp. 2055–2065, 2012.

[152] J. D. Bain, C. Switzer, R. Chamberlin, and S. A. Bennert, "Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code," *Nature*, vol. 356, pp. 537–539, Apr. 1992.

[153] Z. Yang, D. Hutter, P. Sheng, A. M. Sismour, and S. A. Benner, "Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern," *Nucleic Acids Res*, vol. 34, no. 21, pp. 6095–6101, 2006.

[154] D. Hutter and S. A. Benner, "Expanding the genetic alphabet: non-epimerizing nucleoside with the py DDA hydrogen-bonding pattern," *J Org Chem*, vol. 68, no. 25, pp. 9839–9842, 2003.

[155] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, and S. A. Benner, "Enzymatic incorporation of a third nucleobase pair," *Nucleic Acids Res*, vol. 35, no. 13, pp. 4238–4249, 2007.

[156] R. Kwok *et al.*, "Chemical biology: DNA's new alphabet.," *Nature*, vol. 491, no. 7425, pp. 516–518, 2012.

[157] M. Ishikawa, I. Hirao, and S. Yokoyama, "Synthesis of 3-(2-deoxy-β-d-ribofuranosyl) pyridin-2-one and 2-amino-6-(N, N-dimethylamino)-9-(2-deoxy-β-d-ribofuranosyl) purine derivatives for an unnatural base pair," *Tetrahedron Lett*, vol. 41, no. 20, pp. 3931–3934, 2000.

[158] T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, and S. Yokoyama, "Unnatural base pairs for specific transcription," *Proc Natl Acad Sci U S A*, vol. 98, no. 9, pp. 4922–4925, 2001.

[159] T. Fujiwara, M. Kimoto, H. Sugiyama, I. Hirao, and S. Yokoyama, "Synthesis of 6-(2-thienyl) purine nucleoside derivatives that form unnatural base pairs with pyridin-2-one nucleosides," *Bioorganic & medicinal chemistry letters*, vol. 11, no. 16, pp. 2221–2223, 2001.

[160] T. Mitsui, M. Kimoto, Y. Harada, S. Yokoyama, and I. Hirao, "An efficient unnatural base pair for a base-pair-expanded transcription system," *J Am Chem Soc*, vol. 127, no. 24, pp. 8652–8658, 2005.

[161] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, *et al.*, "An unnatural base pair for incorporating amino acid analogs into proteins," *Nat Biotechnol*, vol. 20, no. 2, pp. 177–182, 2002.

[162] R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama, and I. Hirao, "Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs," *J Am Chem Soc*, vol. 127, no. 49, pp. 17286–17295, 2005.

[163] T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, and S. Yokoyama, "An unnatural hydrophobic base pair with shape complementarity between pyrrole-2-carbaldehyde and 9-methylimidazo [(4, 5)-b] pyridine," *J Am Chem Soc*, vol. 125, no. 18, pp. 5298–5307, 2003.

[164] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, and I. Hirao, "Highly specific unnatural base pair systems as a third base pair for PCR amplification," *Nucleic Acids Res*, vol. 40, no. 6, pp. 2793–2806, 2012.

[165] D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, and F. E. Romesberg, "Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base," *J Am Chem Soc*, vol. 121, no. 49, pp. 11585–11586, 1999.

[166] A. A. Henry and F. E. Romesberg, "Beyond A, C, G and T: augmenting nature's alphabet," *Curr Opin Chem Biol*, vol. 7, no. 6, pp. 727–733, 2003.

[167] S. Matsuda, A. M. Leconte, and F. E. Romesberg, "Minor groove hydrogen bonds and the replication of unnatural base pairs," *J Am Chem Soc*, vol. 129, no. 17, pp. 5551–5557, 2007.

[168] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, and F. E. Romesberg, "Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet," *J Am Chem Soc*, vol. 130, no. 7, pp. 2336–2343, 2008.

[169] L. Li, M. l. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, and F. E. Romesberg, "Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications," *J Am Chem Soc*, 2013.

[170] T. Lavergne, D. A. Malyshev, and F. E. Romesberg, "Major groove substituents and polymerase

recognition of a class of predominantly hydrophobic unnatural base pairs," *Chem Eur J*, vol. 18, no. 4, pp. 1231–1239, 2012.

[171] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, and F. E. Romesberg, "Optimization of an unnatural base pair toward natural-like replication," *J Am Chem Soc*, vol. 131, no. 9, pp. 3246–3252, 2009.

[172] Y. J. Seo, S. Matsuda, and F. E. Romesberg, "Transcription of an expanded genetic alphabet," *J Am Chem Soc*, vol. 131, no. 14, pp. 5046–5047, 2009.

[173] D. A. Malyshev, D. A. Pfaff, S. I. Ippoliti, G. T. Hwang, T. J. Dwyer, and F. E. Romesberg, "Solution structure, mechanism of replication, and optimization of an unnatural base pair," *Chem Eur J*, vol. 16, no. 42, pp. 12650–12659, 2010.

[174] H. Weizman and Y. Tor, "2, 2'-bipyridine ligandoside: a novel building block for modifying DNA with intra-duplex metal complexes," *J Am Chem Soc*, vol. 123, no. 14, pp. 3375–3376, 2001.

[175] E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg, and P. G. Schultz, "A novel copper-mediated DNA base pair," *J Am Chem Soc*, vol. 122, no. 43, pp. 10714–10715, 2000.

[176] S. Atwell, E. Meggers, G. Spraggon, and P. G. Schultz, "Structure of a copper-mediated base pair in DNA," *J Am Chem Soc*, vol. 123, no. 49, pp. 12364–12367, 2001.

[177] N. Zimmermann, E. Meggers, and P. G. Schultz, "A novel silver (I)-mediated DNA base pair," *J Am Chem Soc*, vol. 124, no. 46, pp. 13684–13685, 2002.

[178] M. Tasaka, K. Tanaka, M. Shiro, and M. Shionoya, "A palladium-mediated DNA base pair of a β-C-nucleoside possessing a 2-aminophenol as the nucleobase," *Supramol Chem*, vol. 13, no. 6, pp. 671–675, 2001.

[179] K. Tanaka, Y. Yamada, and M. Shionoya, "Formation of silver(I)-mediated DNA duplex and triplex through an alternative base pair of pyridine nucleobases," *J Am Chem Soc*, vol. 124, no. 30, pp. 8802–8803, 2002.

[180] H.-A. Wagenknecht, "Metal-mediated DNA base pairing and metal arrays in artificial DNA: Towards new nanodevices," *Angew Chem Int Ed*, vol. 42, no. 28, pp. 3204–3206, 2003.

[181] G. H. Clever, K. Polborn, and T. Carell, "A highly DNA-duplex-stabilizing metal-salen base pair," *Angew Chem Int Ed*, vol. 44, no. 44, pp. 7204–7208, 2005.

[182] G. H. Clever, Y. Soeltl, H. Burks, W. Spahl, and T. Carell, "Metal-salen-base-pair complexes inside DNA: Complexation overrides sequence information," *Chem Eur J*, vol. 12, no. 34, pp. 8708–8718, 2006.

[183] G. H. Clever and T. Carell, "Controlled stacking of 10 transition-metal ions inside a dna duplex," *Angew Chem Int Ed*, vol. 46, no. 1-2, pp. 250–253, 2007.

[184] G. H. Clever, S. J. Reitmeier, T. Carell, and O. Schiemann, "Antiferromagnetic coupling of stacked CuII-salen complexes in DNA," *Angew Chem*, vol. 122, no. 29, pp. 5047–5049, 2010.

[185] C. Kaul, M. Müller, M. Wagner, S. Schneider, and T. Carell, "Reversible bond formation enables the replication and amplification of a crosslinking salen complex as an orthogonal base pair," *Nature*

chemistry, vol. 3, no. 10, pp. 794–800, 2011.

[186] G. H. Clever, C. Kaul, and T. Carell, "DNA–metal base pairs," *Angew Chem Int Ed*, vol. 46, no. 33, pp. 6226–6236, 2007.

[187] P. Scharf and J. Müller, "Nucleic acids with metal-mediated base pairs and their applications," *ChemPlusChem*, vol. 78, no. 1, pp. 20–34, 2013.

[188] R. Stoltenburg, C. Reinemann, and B. Strehlitz, "SELEX - a (r)evolutionary method to generate high-affinity nucleic acid ligands," *Biomol Eng*, vol. 24, no. 4, pp. 381–403, 2007.

[189] G. Mayer, "The chemical biology of aptamers," *Angew Chem Int Ed*, vol. 48, no. 15, pp. 2672–2689, 2009.

[190] E. R. Mardis, "Next-generation DNA sequencing methods," *Annu Rev Genomics Hum Genet*, vol. 9, pp. 387–402, 2008.

[191] R. Kranaster and A. Marx, "Engineered DNA polymerases in biotechnology," *Chembiochem*, vol. 11, no. 15, pp. 2077–2084, 2010.

[192] S. H. Weisbrod and A. Marx, "Novel strategies for the site-specific covalent labelling of nucleic acids," *Chem Commun*, no. 44, pp. 5675–5685, 2008.

[193] K. Bergen, K. Betz, W. Welte, K. Diederichs, and A. Marx, "Structures of KOD and 9°N DNA polymerases complexed with primer template duplex," *Chembiochem*, vol. 14, no. 9, pp. 1058–1062, 2013.

[194] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, F. E. Romesberg, and A. Marx, "Structural insights into DNA replication without hydrogen bonds," *J Am Chem Soc*, vol. 135, no. 49, pp. 18637–18643, 2013.

[195] Y. Li, Y. Kong, S. Korolev, and G. Waksman, "Crystal structures of the Klenow fragment of *Thermus aquaticus* DNA polymerase I complexed with deoxyribonucleoside triphosphates," *Protein Sci*, vol. 7, no. 5, pp. 1116–1123, 1998.

[196] W. Kabsch, "XDS," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 66, no. 2, pp. 125–132, 2010.

[197] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, *et al.*, "Phenix: a comprehensive python-based system for macromolecular structure solution," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 66, no. 2, pp. 213–221, 2010.

[198] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 60, no. 12, pp. 2126–2132, 2004.

[199] P. A. Karplus and K. Diederichs, "Linking crystallographic model and data quality," *Science*, vol. 336, no. 6084, pp. 1030–1033, 2012.

[200] K. Diederichs and P. A. Karplus, "Improved R-factors for diffraction data analysis in macromolecular crystallography," *Nat Struct Biol*, vol. 4, no. 4, pp. 269–275, 1997.

[201] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, *et al.*, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids," *Nucleic Acids Res*, vol. 35, no. suppl 2, pp. W375–W383, 2007.

[202] O. Bermek, N. D. F. Grindley, and C. M. Joyce, "Prechemistry nucleotide selection checkpoints in the reaction pathway of DNA polymerase I and roles of Glu710 and Tyr766," *Biochemistry*, vol. 52, no. 36, pp. 6258–6274, 2013.

[203] S. J. Johnson, J. S. Taylor, and L. S. Beese, "Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations," *Proc Natl Acad Sci U S A*, vol. 100, no. 7, pp. 3895–3900, 2003.

[204] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, "Conformational analysis of nucleic acids revisited: Curves+," *Nucleic Acids Res*, vol. 37, no. 17, pp. 5917–5929, 2009.

[205] A. S. Meyer, M. Blandino, and T. E. Spratt, "Escherichia coli DNA polymerase I (Klenow fragment) uses a hydrogen-bonding fork from Arg668 to the primer terminus and incoming deoxynucleotide triphosphate to catalyze DNA replication," *J Biol Chem*, vol. 279, no. 32, pp. 33043–33046, 2004.

[206] S. Obeid, W. Welte, K. Diederichs, and A. Marx, "Amino acid templating mechanisms in selection of nucleotides opposite abasic sites by a family A DNA polymerase," *J Biol Chem*, vol. 287, no. 17, pp. 14099–14108, 2012.

[207] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, and S. Yokoyama, "An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA," *Nat Methods*, vol. 3, no. 9, pp. 729–735, 2006.

[208] D. A. Malyshev, Y. J. Seo, P. Ordoukhanian, and F. E. Romesberg, "PCR with an expanded genetic alphabet," *J Am Chem Soc*, vol. 131, no. 41, pp. 14620–14621, 2009.

[209] S. Obeid, N. Blatter, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, and A. Marx, "Replication through an abasic DNA lesion: structural basis for adenine selectivity," *EMBO J*, vol. 29, no. 10, pp. 1738–1747, 2010.

[210] D. Temiakov, V. Patlan, M. Anikin, W. T. McAllister, S. Yokoyama, and D. G. Vassylyev, "Structural basis for substrate selection by T7 RNA polymerase," *Cell*, vol. 116, no. 3, pp. 381 – 391, 2004.

[211] Y. Santoso, C. M. Joyce, O. Potapova, L. Le Reste, J. Hohlbein, J. P. Torella, N. D. F. Grindley, and A. N. Kapanidis, "Conformational transitions in dna polymerase I revealed by single-molecule FRET," *Proc Natl Acad Sci U S A*, vol. 107, no. 2, pp. 715–720, 2010.

[212] S. Y. Berezhna, J. P. Gill, R. Lamichhane, and D. P. Millar, "Single-molecule förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I," *J Am Chem Soc*, vol. 134, no. 27, pp. 11261–11268, 2012.

[213] J. M. Clark, C. M. Joyce, and G. Beardsley, "Novel blunt-end addition reactions catalyzed by DNA

polymerase I of *Escherichia coli*," *J Mol Biol*, vol. 198, no. 1, pp. 123 – 127, 1987.

[214] K. A. Fiala, J. A. Brown, H. Ling, A. K. Kshetry, J. Zhang, J.-S. Taylor, W. Yang, and Z. Suo, "Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase," *J Mol Biol*, vol. 365, no. 3, pp. 590–602, 2007.

[215] D. E. Danley, "Crystallization to obtain protein–ligand complexes for structure-aided drug design," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 62, pp. 569–575, Jun 2006.

[216] T. A. Kunkel, "DNA replication fidelity," *J Biol Chem*, vol. 279, no. 17, pp. 16895–16898, 2004.

[217] A. A. Golosov, J. J. Warren, L. S. Beese, and M. Karplus, "The mechanism of the translocation step in DNA replication by DNA polymerase I: A computer simulation analysis," *Structure*, vol. 18, no. 1, pp. 83 – 93, 2010.

[218] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, and F. E. Romesberg, "Efforts toward expansion of the genetic alphabet: Structure and replication of unnatural base pairs," *J Am Chem Soc*, vol. 129, no. 34, pp. 10466–10473, 2007. PMID: 17685517.

[219] X. Jiang and M. Egli, "Use of chromophoric ligands to visually screen co-crystals of putative protein-nucleic acid complexes," *Curr Protoc Nucleic Acid Chem*, pp. 7–15, 2011.

[220] S. A. Wynne, V. B. Pinheiro, P. Holliger, and A. G. Leslie, "Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA," *PloS one*, vol. 8, no. 8, p. e70892, 2013.

[221] A. J. Berman, S. Kamtekar, J. L. Goodman, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, and T. A. Steitz, "Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases," *EMBO J*, vol. 26, no. 14, pp. 3494–3505, 2007.

[222] A. F. Gardner and W. E. Jack, "Determinants of nucleotide sugar recognition in an archaeon DNA polymerase," *Nucleic Acids Res*, vol. 27, no. 12, pp. 2545–2553, 1999.

[223] E. M. Kennedy, C. Hergott, S. Dewhurst, and B. Kim, "The mechanistic architecture of thermostable *Pyrococcus furiosus* family B DNA polymerase motif a and its interaction with the dNTP substrate," *Biochemistry*, vol. 48, no. 47, pp. 11161–11168, 2009.

[224] S. Kamtekar, A. J. Berman, J. Wang, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, and T. A. Steitz, "Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29," *Molecular Cell*, vol. 16, no. 4, pp. 609 – 618, 2004.

[225] V. B. Pinheiro and P. Holliger, "The XNA world: progress towards replication and evolution of synthetic genetic polymers," *Curr Opin Chem Biol*, vol. 16, no. 34, pp. 245 – 252, 2012. Synthetic biology / Analytical techniques.

[226] B. A. Maxwell and Z. Suo, "Single-molecule investigation of substrate binding kinetics and protein conformational dynamics of a B-family replicative DNA polymerase," *J Biol Chem*, vol. 288, no. 16, pp. 11590–11600, 2013.

[227] OpenWetWare, "E. coli genotypes — openwetware," 2013. "[Online; accessed 14-April-2014]", Available from: $http://openwetware.org/index.php?title = E._coli_genotypes\&oldid = 682826$.

[228] U. K. Laemmli *et al.*, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, 1970.

[229] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, *et al.*, "Protein identification and analysis tools on the ExPASy server," in *The proteomics protocols handbook*, pp. 571–607, Springer, 2005.

[230] R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim, "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia," *Science*, vol. 230, no. 4732, pp. 1350–1354, 1985.

[231] K. Betz, "Structural snapshots of KlenTaq DNA polymerase processing 4'-alkylated substrates," Master thesis, Universität Konstanz, 2009.

[232] S. Obeid, *Snapshots of DNA polymerase processing aberrant substrates: Structural insights into abasic site bypass & polymerization of 5-alkynylated nucleotide analogs.* Dissertation, Universität Konstanz, 2011.

[233] B. Holzberger, *From Multifluorinated DNA Polymerases to Insights into DNA Synthesis by NMR Spectroscopy.* Dissertation, Universität Konstanz, 2012.

[234] R. Peist, A. Koch, P. Bolek, S. Sewitz, T. Kolbus, and W. Boos, "Characterization of the aes gene of *Escherichia coli* encoding an enzyme with esterase activity.," *J Bacteriol*, vol. 179, no. 24, pp. 7679–86, 1997.

[235] R. V. Pei-Chung Hsieh, *Enzyme Engineering: Methods and Protocols.* (C) Springer Science+Business Media New York, 2013.

[236] M. Southworth, H. Kong, R. Kucera, J. Ware, H. Jannasch, and F. Perler, "Cloning of thermostable DNA polymerases from hyperthermophilic marine archaea with emphasis on *thermococcus* sp. 9 degrees n-7 and mutations affecting 3'-5'exonuclease activity," *Proc Natl Acad Sci U S A*, vol. 93, no. 11, pp. 5281–5285, 1996.

[237] M. Mueller, M. Wang, and C. Schulze-Briese, "Optimal fine-slicing for single-photon-counting pixel detectors," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 68, no. 1, pp. 42–56, 2011.

[238] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, *et al.*, "Overview of the CCP4 suite and current developments," *Acta Crystallogr Sect D: Biological Crystallography*, vol. 67, no. 4, pp. 235–242, 2011.

[239] O. S. Smart, T. O. Womack, A. Sharff, C. Flensburg, P. Keller, W. Paciorek, C. Vonrhein, and G. Bricogne, "grade, version 1.1.1." http://www.globalphasing.com, 2011.

[240] J. Wang, P. Yu, T. C. Lin, W. H. Konigsberg, and T. A. Steitz, "Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions," *Biochemistry*, vol. 35, no. 25, pp. 8110–8119, 1996.

[241] S. Liu, J. D. Knafels, J. S. Chang, G. A. Waszak, E. T. Baldwin, M. R. Deibel, D. R. Thomsen, F. L. Homa, P. A. Wells, M. C. Tory, R. A. Poorman, H. Gao, X. Qiu, and A. P. Seddon, "Crystal structure of the herpes simplex virus 1 DNA polymerase," *J Biol Chem*, vol. 281, no. 26, pp. 18193–18200, 2006.

[242] F. Wang and W. Yang, "Structural insight into translesion synthesis by DNA pol II," *Cell*, vol. 139, no. 7, pp. 1279 – 1289, 2009.

[243] F. Sievers, A. Wilm, D. Dineen, T. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. Thompson, and D. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.," *Mol Syst Biol*, vol. 7, pp. 539–, 2011.

Danksagung

Am Ende meiner Doktorarbeit möchte ich mich gerne bei einigen Personen bedanken, die direkt oder indirekt zum Gelingen dieser Arbeit beigetragen haben.

Der erste Dank gilt meinem Doktorvater Andreas Marx. Danke für die Aufnahme in deine Arbeitsgruppe schon zur Bachelor- und Masterarbeit und schließlich für das interessante und abwechslungsreiche Promotionsthema. Danke für deine stete Unterstützung, Förderung und Motivation und v.a. auch dein Vertrauen in mich während der gesamten gemeinsamen Jahre.

Gleichzeitig danke ich Kay Diederichs und Wolfram Welte, in deren Arbeitsgruppen ich den Großteil meiner praktischen Arbeit durchgeführt habe. Danke für die Aufnahme in eure Arbeitsgruppen und für die gute Zusammenarbeit. Danke Kay für die Übernahme des Zweitgutachtens und deine Betreuung und Hilfe in jeglichen Fragen, die den strukturellen Teil meiner Arbeit betreffen.

Ausserdem danke ich meinen Kooperationspartnern Prof. Dr. Floyd Romesberg, Denis Malyshev und Thomas Lavergne vom Scripps Research Institute in Kalifornien für die gute und erfolgreiche Zusammenarbeit, die trotz der Entfernung hervorragend funktioniert hat.

Danke an Prof. Dr. Valentin Wittmann für die Übernahme des Prüfungsvorsitzes und Prof. Dr. Jörg Hartig für die Anfertigung des Drittgutachtens.

Danke Hans-Jürgen Apell für deine Teilnahme in meinem Thesis Komitee der Graduiertenschule Chemische Biologie und für deine Unterstützung.

Den gesamten ehemaligen und jetzigen Mitarbeitern der AG Marx, Welte, Diederichs und Apell danke ich für die schöne Zeit im und ums Labor auf M12, L9 und neuerdings ML5. Danke für eure Hilfe bei diversen Problemen und Fragen und die tolle Arbeitsatmosphäre. Danke auch für die vielen Ablenkungen zwischendurch (Kuchen, Kaffe, Schokolade, Grillen, Joga, ...), die immer wieder neue Arbeitsmotivation lieferten. Besonders danken möchte ich außerdem meinen Polymerasestruktur-Kollegen Samra und Konni. Danke für die unzähligen Diskussionen, die gute Zusammenarbeit und v.a. die immer abwechslungsreichen und spannenden SLS-Trips.

Danke an dich Benni und an alle meine Kollegen und Freunde für die gemeinsamen Jahre hier in Konstanz und anderswo. Ich danke euch für die vielen freizeitlichen Unternehmungen außerhalb des Labors; danke fürs Klettern, Fußballspielen, Laufen, Weggehen, Zusammenwohnen, Wandern, Eis essen, Brettspielen, uvm. Auch wenn ich euch hier nicht persönlich erwähne, hoffe ich doch, dass ihr wisst, wer gemeint ist.

Zum Schluss danke ich ganz besonders meiner Familie, v.a. meinen Eltern. Danke für eure Liebe, eure Unterstützung und euer Vertrauen in mich während meines gesamten bisherigen Lebenswegs. Ohne euch wäre diese Promotion nicht möglich gewesen. Danke!