CHEMBIOCHEM

Supporting Information

Structures of KOD and 9°N DNA Polymerases Complexed with Primer Template Duplex

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

cbic_201300175_sm_miscellaneous_information.pdf

Materials and Methods

Enzymes, oligodeoxynucleotides, nucleotides

Primer and templates were purchased from *ThermoFisher* (HPLC purified), ddNTPs from *Jena-BioScience*, natural dNTPs from *Fermentas.*

Purification

9°N: For crystallization experiments with *Thermococcus* sp. 9°N-7 DNA polymerase the polymerase mutant (D141A, E143A) with reduced exonuclease activity was used.[1] The exoenzyme is named 9°N throughout the paper. 9°N was cloned into a pET21b vector and overexpressed in *E. coli* BL21(DE3). Expression was induced by addition of 1 mM IPTG final concentration at OD_{600} = 0.6 – 0.8. Cells were harvested after 4 h, resuspended in lysis buffer (50 mM Tris pH 8.55, 10 mM $MgCl_2$, 16 mM $(NH_4)_2(SO_4)$, 0.1% Triton X-100 and 0.1% Thesit) and lysed for 1 h at 37 °C by adding Lysozyme to 0.5 mg mL^{-1} final concentration. After heat denaturation (10 min at 75 °C) 1 mM PMSF was added and denatured proteins and cell debris were pelleted by ultracentrifugation (1 h, 35 000 *g*). The supernatant was loaded onto a Hep-arin column with a low salt buffer (100 mM NaCl, 20 mM Tris pH 7.5, 0.1 mM EDTA, 1 mM DTT, 10% glycerol) and proteins were eluted by increasing concentration of NaCl using a step gradient. Purest fractions (as determined by SDS-PAGE) were pooled and concentrated. The protein was further purified by gel filtration chromatography using a Superdex-200 column and the following buffer: 200 mM NaCl, 20 mM Tris pH 7.5, 0.1 mM EDTA, 1 mM DTT, 10% glycerol. After purification the protein was concentrated to 7 mg mL^{-1} and stored at 4 °C.

KOD: For the crystallization experiments the exonuclease deficient mutant D141A, E143A was used. The protein was expressed in *E. coli* BL21(DE3) using a codon-optimized sequence in a pET24 as expression system (GeneArt). Cells were induced at OD_{600} ~ 0.6 using 1 mM IPTG final concentration in LB medium. Purification was performed as previously described,[2] with the addition of an anion exchange step, using the flowthrough for further purification. This was followed by a step using Superdex 200 SEC. The eluate was concentrated to ~10 mg mL^{-1} and stored at 4°C.

Crystallization

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

enzymes (DNA bound polymerase structures) additionally false positive coloured crystals were found. In these crystals the colouring resulted most probably from unspecifically bound DNA in the crystal lattice.

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

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

KOD: Protein was mixed with a primer/template (primer: 5'-CGC GAA TTG CG-3'; template: 5‘-(Cy5) AAA TTC GCA GTT CGC G-3') complex in a 1:1.5 molar ratio in presence of 20 mM $MgCl_2$ and incubated for 30 min at 37 °C. A five molar excess of ddATP was added and again incubated for 30 min at 37 °C, leading to a final protein concentration of about 7 mg mL^{-1}. Screening setups were performed using a Gryphon robot (ARI Robots) and various commercially available screens. In order to separate the numerous apo crystals forms from DNA complexed ones, a method described by Jiang & Egli[3] was used. Identified conditions were refined in a grid screening varying pH and precipitant and reproduced without the dye. Diffracting crystals were obtained from a condition containing 50 mM sodium cacodylate pH 8, 10% *v/v* 2-propanol, 20 mM $MgCl_2$, 1 mM cobalt(III)-hexamine chloride, 1 mM spermine (derived from condition G3 Natrix Screen, Hampton Research). Crystals were cryoprotected by soaking in the mother liquor containing 20% ethylene glycol and flash-frozen in liquid nitrogen.

Data collection and processing

Datasets were collected at the beamline X06SA-PX of the Swiss light Source (SLS) at the Paul Scherrer Institut in Villigen, Switzerland. Data reduction was done using the XDS package.[4] Statistics of data collection and refinement for both structures are given in Table S1. Simulated annealing electron density omit maps for the DNA are shown in Figure S3.

9°N: Data reduction was done in space group P22121 with the cell dimension a = 112.2 Å, b = 142.6 Å, c = 66.7 Å and α, β, γ = 90.0°. The structure was solved by molecular replacement using the 9°N apo structure (PDB ID: 1QHT[5]) as search model. The structure was improved by

alternated refinement with PHENIX[6] and model building in COOT.[7] Refinement was accompanied by validation using the Molprobity server.[8] The restraint file for the Cy5 dye was created using the Grade Web Server.[11]

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

KOD: Data reduction was done in space group P22121 with the cell dimensions a = 144.9 Å, b = 111.8 Å, c = 67.1 Å and $\alpha, \beta, \gamma = 90°$. In this case three datasets were merged by using XSCALE. The structure was solved by molecular replacement using the KOD apo structure (PDB ID: 1WNS)[2] as search model. Difference density guided modelling of the DNA was performed using COOT.[7] The structure was improved by alternated refinement with PHE-NIX[6] and model building in COOT;[7] structure quality was surveyed using the Molprobity server.[8] For KOD, the $I/\sigma = 2$ resolution is ~2.4 Å.

Structural details 9°N

In case of the 9°N structure the Cy5 dye is well ordered and thereby visible in the electron density. The four single stranded nucleotides at the 5' end of the template show an interesting arrangement reminiscent of the observation in previously reported KlenTaq structures with a 3'-dT-dA-dA-dA-5' and 3'-dNaM-dA-dA-dA-3' single stranded template overhang.[10] In the present case the templating nucleotide dGT4 (numbering of nucleotides see Figure 2) is flipped out of the stacking arrangement and the subsequent nucleobases dAT3 and dAT2 stack on top of the last formed base pair. The last nucleotide dAT1 and the attached Cy5 dye interact with the exonuclease domain. The dye points towards a symmetry mate where it is stabilized be-tween Arg99, His103 and Pro104. Due to this arrangement the 5'-overhanging template was stabilized but in a conformation probably not occurring in solution though favourable for crys-tal formation. In proximity to the nucleotide dAT2 positive difference density was observed. Alternative conformations of residues dAT2 and Cy5 can explain this density. We modelled a second conformation of dAT2 by rotating the nucleotide approximately 150° around the C3' sugar atom and a second conformation of Cy5. The occupancy of the alternate conformations is below 50% for both residues, therefore in some cases the position is not occupied. In all figures we only show the position of the dAT2 and Cy5 conformation with the higher occu-pancy.

Compared to the apo 9°N structure published in Beese et al. in 2000,[5] we could model some additional regions of the thumb domain. These regions are P569–L574, E648–L659, T667–K674, G706–I710, D721–A729 and A751–Y757. Our 9°N model is complete up to the last 23 C-terminal residues which were not resolved in the electron density.

Structural details KOD:

Due to the stabilizing effects of the DNA binding, electron density allowed a rebuilding of the thumb 1 and 2 domains in KOD, including regions T667–G677, R688–G696, K705–D712 and A747–F748, which were missing in the apo structure. The binary KOD model could be built up to residue 756, leaving the last 18 amino acids unresolved. The binary KOD model could be built up to residue 756, leaving the last 18 amino acids unresolved. The 5' overhanging tem-plate nucleobase in the KOD structure stacks above the last basepair and the binding pocket for the incoming nucleotide is empty. The template base dAT3 is extending to the T-cleft and the subsequent nucleotides dAT2 and dAT1 are not resolved.

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

All figures showing structures in this paper were prepared using PyMOL.[11]

Table S1: Summary of data collection and refinement

	9°N	KOD
PDB ID	**4K8X**	**4K8Z**
Data collection		
Wavelength	0.97793Å	1.00002Å
Spacegroup	18 ($P2_12_12$)	18 ($P2_12_12$)
Cell dimensions		
a, b, c (Å)	112.21, 142.62, 66.70	144.87, 111.78, 67.10
α,β,γ (°)	90.0	90.0
Resolution I/σ=2 (Å)	2.6	2.4
Resolution (Å)	48.7-2.28 (2.42-2.28)	49.3-2.29 (2.35-2.29)
Total reflections*	330620 (54378)	803790 (46136)
Unique reflections*	49600 (7914)	49755 (3506)
R_{meas} (%)* +	17.2 (282.1)	13.5 (196.5)
I/σI *	9.07 (0.82)	16.65 (1.81)
Completeness (%)*	99.9 (99.9)	99.7 (96.7)
Refinement		
Resolution (Å)[$]*	48.7-2.28 (2.32-2.28)	45.06-2.29(2.33-2.29)
No. unique reflections*	49542	49743
R_{work}/R_{free} *[+]	19.5/23.8 (36.6/42.2)	18.1/21 .8 (31.3-36.2)
Coordinate error (Å)[a]	0.38	0.26
B-factors		
DNA	73.8	51.9
Protein	60.8	46.4
R.m.s deviations		
Bond lengths (Å)	0.008	0.003
Bond angles (°)	1.16	0.72
Ramachandran[#] (%)		
Favored	97.24	97.75
Allowed	2.62	2.25
Outlier	0.13	-

*Numbers in brackets refer to the highest resolution shell.

+ for definition of R_{meas}, see [12]

#as determined by MolProbity[8]

[$] Data used in refinement up to a CC(1/2) of around 50%, for details see above and[9].

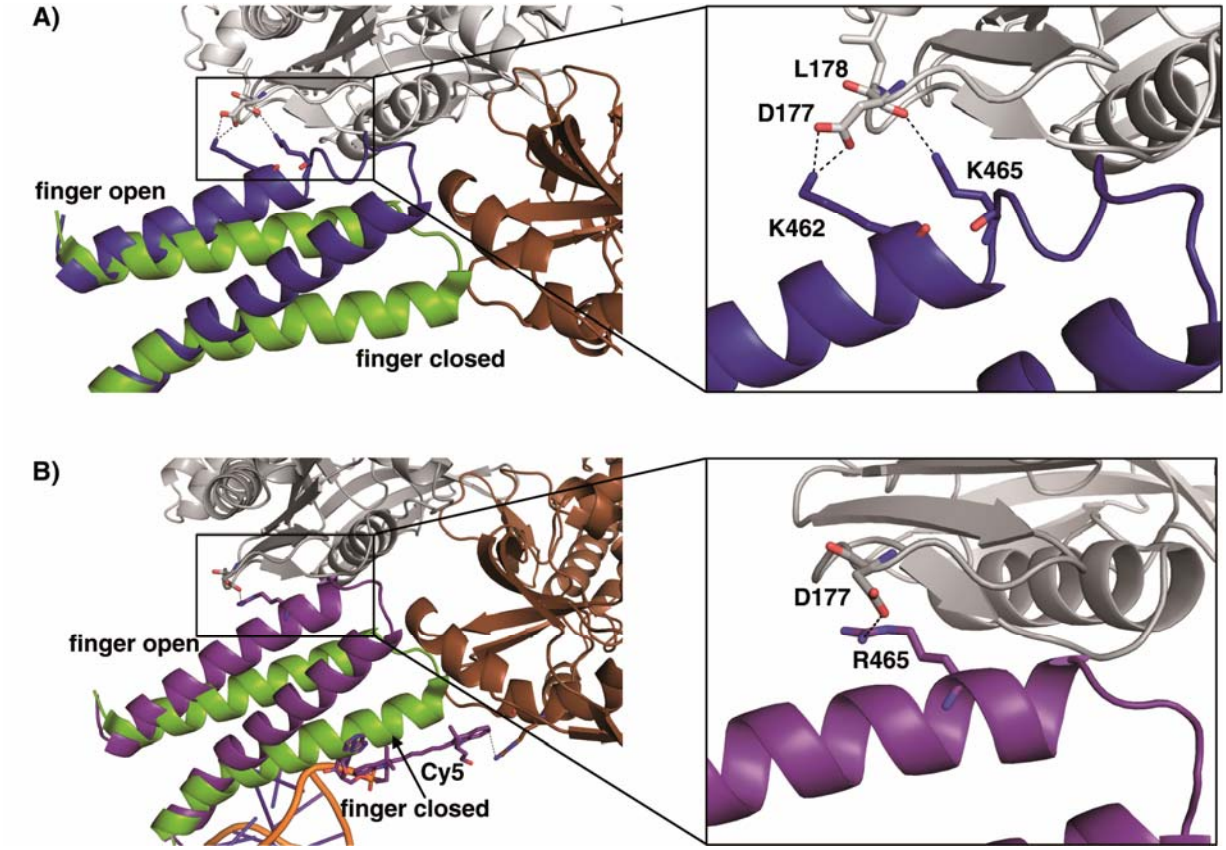

Figure S1: Crystal contacts of the finger domain in KOD and 9°N binary complexes which hinder finger domain closure. Finger domains are shown as cartoon in A) KOD (blue) and B) 9°N (purple). The position of the closed finger domain from the superposed ternary structure of DNA polymerase delta (PDB ID: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

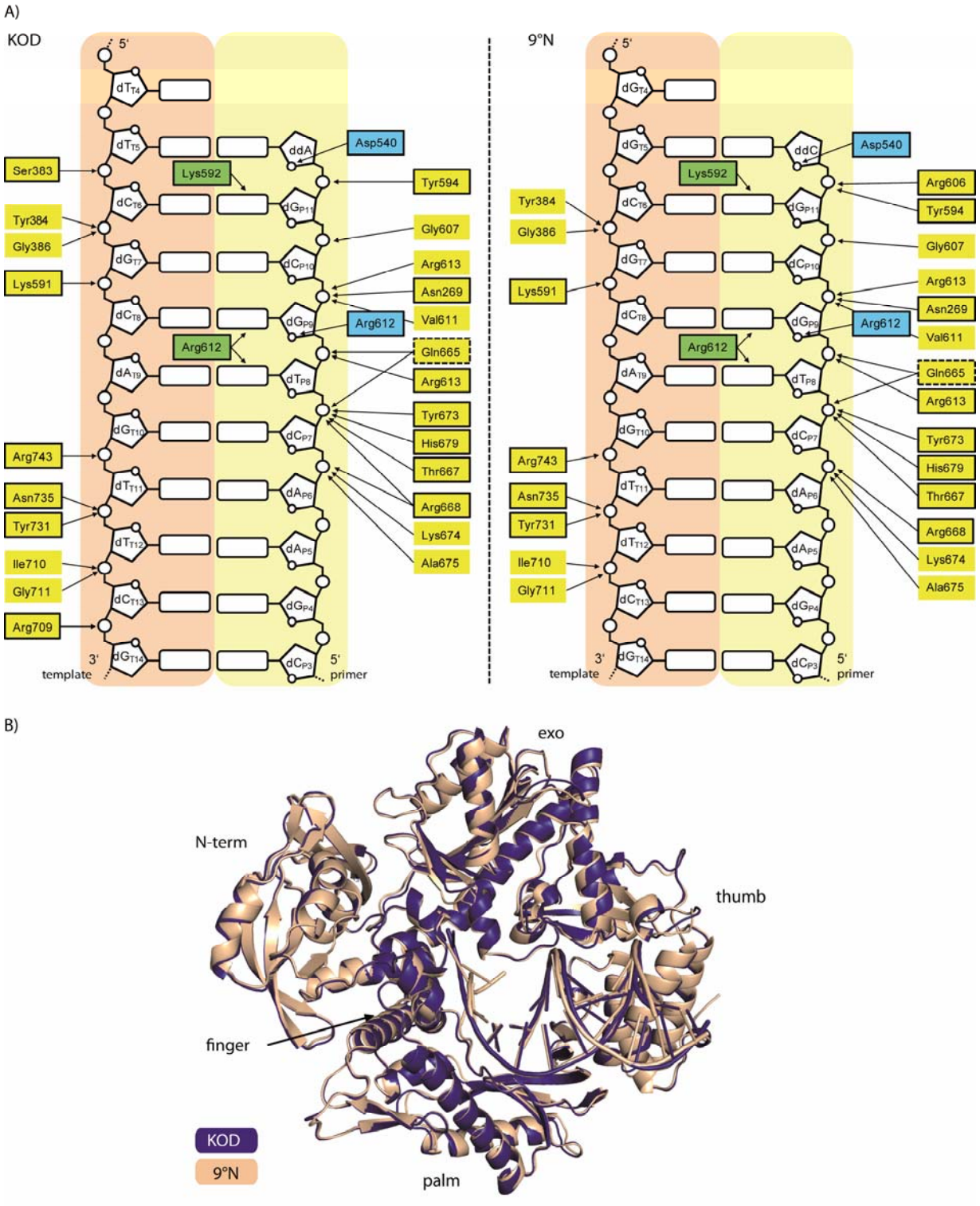

Figure S2: A) Comparison of protein-DNA interactions in KOD and 9°N binary structures. Only direct contacts up to a distance of 3.6Å are shown. Side chain interactions are marked with a solid lining, contacts with the protein backbone are shown without lining and residues where both interactions are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

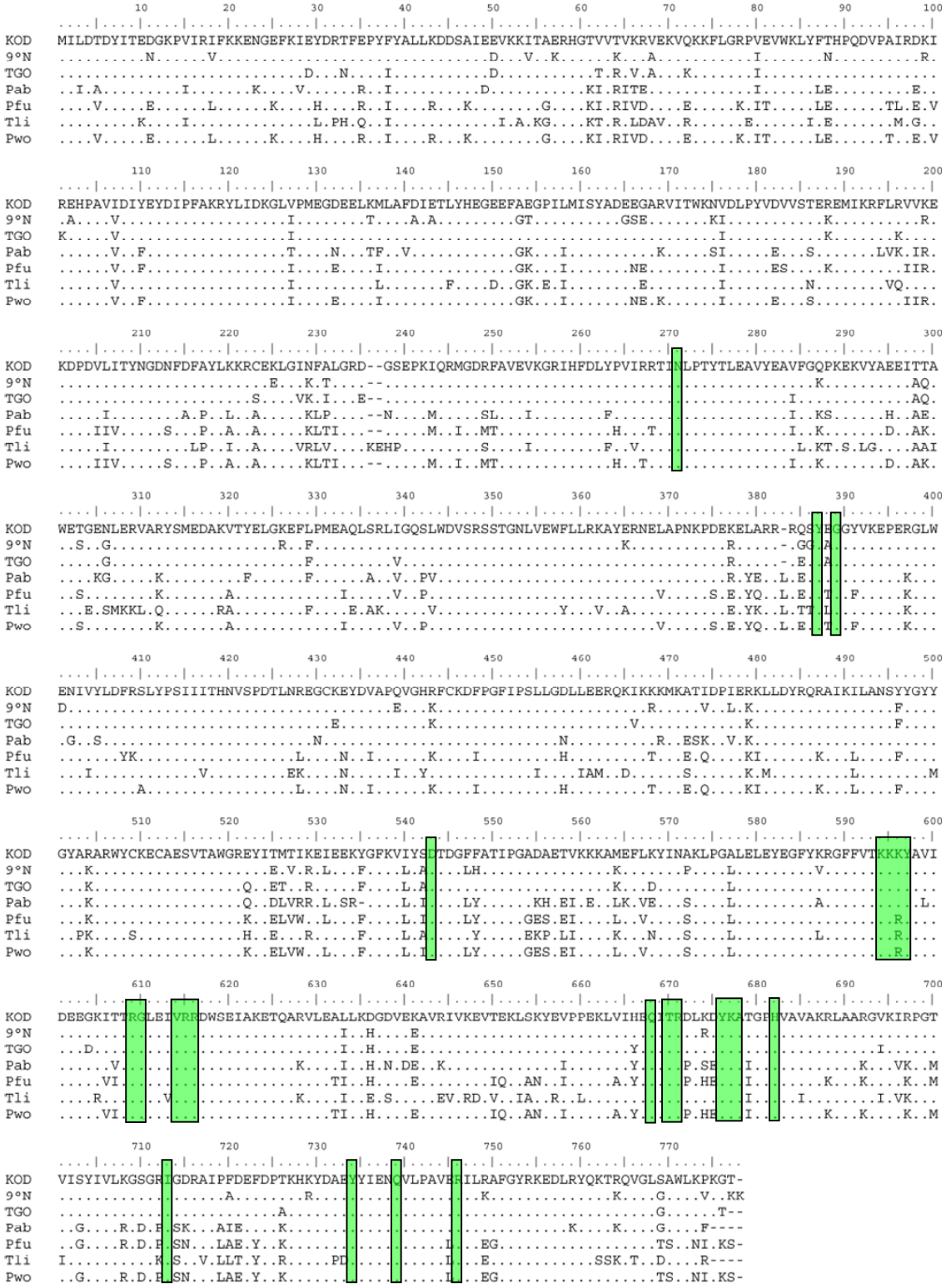

Figure S3: Sequence alignment of archaeal family B DNA polymerases. Highlighted residues show amino acids in involved in DNA binding in the binary complexes of 9°N and KOD that are conserved in all aligned archaeal polymerases. Points indicate sequence identity. The graphic was made using BioEdit[13]

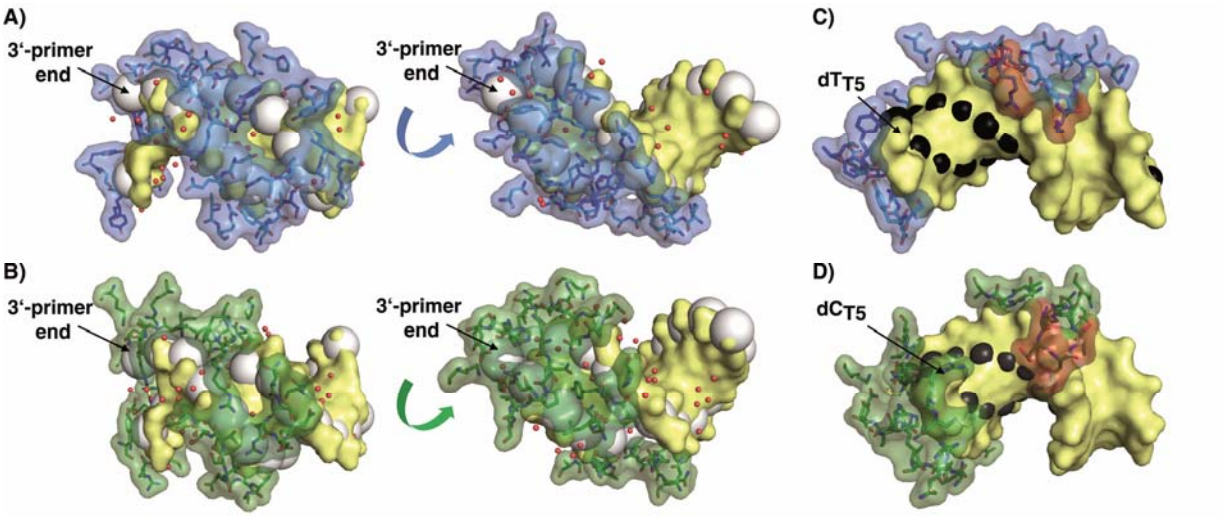

Figure S4: Comparison of minor and major groove of KOD (blue, A) and C)) and KlenTaq (green, B) and D)) DNA polymerases. Residues in a distance up to 4.5Å around the DNA duplex are shown as sticks and surface. The DNA duplex is shown as yellow surface; the 5' single stranded template nucleotides are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

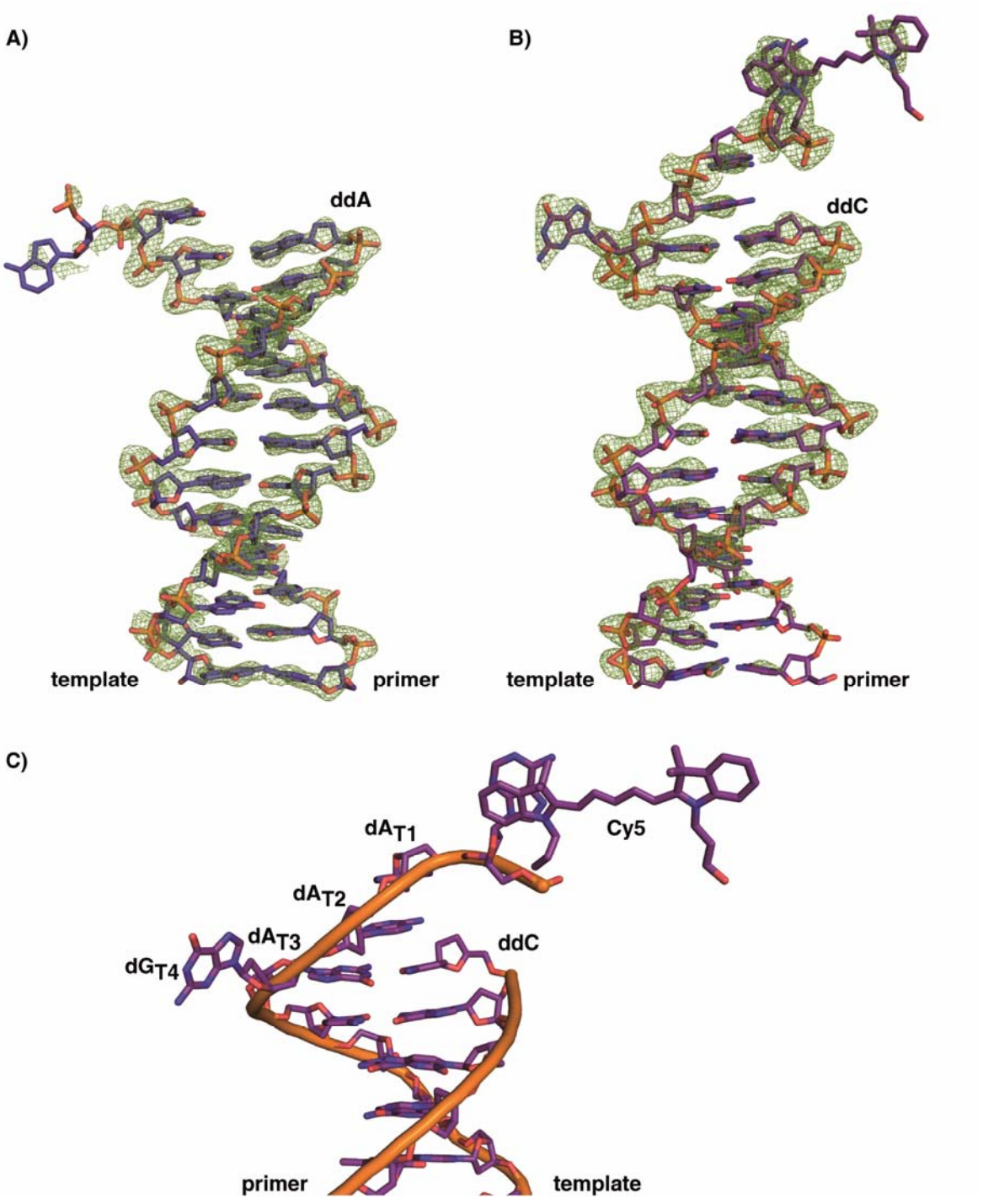

Figure S5: A,B) Simulated annealing omit maps of DNA primer and template in KOD (blue) and 9°N (purple) are shown at 3σ. C) Arrangement of the single stranded DNA template in the 9°N binary structure shown as sticks. The first single stranded nucleobase (G_{T4}) is flipped out of the stacking arrangement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

[1] M. W. Southworth, H. Kong, R. B. Kucera, J. Ware, H. W. Jannasch, F. B. Perler, *Proc Natl Acad Sci U S A* **1996**, *93*, 5281-5285.
[2] H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai, *J. Mol. Biol.* **2001**, *306*, 469-477.
[3] X. Jiang, M. Egli, *Curr. Protoc. Nucleic Acid. Chem.* **2011**, *Chapter 7*, Unit 7 15 11-18.
[4] a) W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132; b) W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 133-144.
[5] A. C. Rodriguez, H. W. Park, C. Mao, L. S. Beese, *J. Mol. Biol.* **2000**, *299*, 447-462.
[6] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, *et al.*, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213-221.
[7] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486-501.
[8] a) I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res* **2007**, *35*, W375-383; b) V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 12-21.
[9] P. A. Karplus, K. Diederichs, *Science* **2012**, *336*, 1030-1033.
[10] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, A. Marx, *Nat Chem Biol* **2012**, *8*, 612-614.
[11] Schrodinger, LLC, **2010**.
[12] K. Diederichs, P. A. Karplus, *Nat Struct Biol* **1997**, *4*, 269-275.
[13] T. A. Hall, *Nucleic Acids Symposium Series* **1999**, *41*, 95-98.