Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Structures of DNA Polymerases Caught Processing Size-Augmented Nucleotide Probes**

*Karin Betz, Frank Streckenbach, Andreas Schnur, Thomas Exner, Wolfram Welte, Kay Diederichs, and Andreas Marx**

anie_200905724_sm_miscellaneous_information.pdf

Supporting Information

Structures of DNA polymerases caught processing size-augmented nucleotide probes

Karin Betz, Frank Streckenbach, Andreas Schnur, Thomas Exner, Wolfram Welte, Kay Diederichs, and Andreas Marx*

Departments of Chemistry and Biology, Konstanz Research School Chemical Biology, Universität Konstanz, Universitätsstrasse 10, 78457 Konstanz (Germany), Fax: (+49)7531-88-5140, andreas.marx@uni-konstanz.de.

Proteins, nucleotides and oligonucleotides

Protein expression and purification were conducted as described.[1] Modified thymidine-5'-triphosphates dT^RTPs were synthesized as described previously.[2] Unmodified dNTPs are commercially available (*Roche*).

DNA oligonucleotides for enzyme kinetics were synthesized on an Applied-Biosystems 392 DNA/RNA synthesizer and purified by reversed phase HPLC (DMT-ON) and afterwards by preparative PAGE on a 12% polyacrylamide gel containing 8M urea (DMT-OFF). DNA primer oligonucleotides were labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase (*Fermentas*) according to the procedure recommended by the manufacturer.

Pre-steady state enzyme kinetics

The rate of single turnover, single nucleotide incorporation was determined using rapid quench flow kinetics on a chemical quench flow apparatus (RQF-3, KinTek Corp., University Park, PA) as previously described.[3] In brief, 15 µl of radiolabelled primer/template complex (50 nM) and DNA polymerase (500 nM) in reaction buffer (RQF buffer, Tris*HCl pH 7.5 20mM, NaCl 50mM, $MgCl_2$ 2mM)[4] were rapidly mixed with 15 µl of a dNTP solution in reaction buffer at 37°C. Quenching was achieved by 0.3M EDTA solution (pH 7.0) at defined time intervals. For the investigation of dT^RTP incorporation a 24nt primer (5'-GTG GTG CGA AAT TTC TGA CAG ACA) and a 36nt template (5'-GTG CGT CTG TC**A** TGT CTG TCA GAA ATT TCG CAC CAC) were applied. Quenched samples were analyzed on a 12% denaturing PAGE followed by phosphor imaging. For kinetic analysis experimental data were fit by nonlinear regression using the program GraphPad Prism 4. The data were fit to a single exponential equation: [conversion]=A*(1-exp(-k_{obs} t)). The observed catalytic rates (k_{obs}) were then plotted against the dNTP concentration used and the data were fitted to a hyperbolic equation [k_{obs}]=k_{cat}*[dNTP]/(K_D + [dNTP]) to determine the K_D and k_{cat} of the incoming nucleotide. The incorporation efficiency is given by k_{pol}/K_D. The depicted data derived from double repeated experiments.

Crystallization and structure determination

The closed ternary complex of the *KlenTaq* bound to DNA primer (5'-d(GAC CAC GGC GC)-3') and template (5'-d(AAA AGG CGC CGT GGT C)-3'), an incorporated ddCMP and the steric probes $dT^{Me}TP$ and $dT^{Et}TP$ crystallized in condition C3 of the commercially available Nucleix Screen from Qiagen (0.2M ammonium acetate, 0.01M magnesium acetate 0.5M sodium cacodylate, pH 6.5 and 30% (w/v) PEG 8000) at 18°C. Purified *KlenTaq* DNA polymerase (14mg/ml), the primer/template complex, ddCTP and the respective steric probe were mixed in a molar ratio of 1:3:10:20 in 20mM Mg_2Cl and crystallized using the vapour diffusion sitting drop method. Datasets from the ternary complexes were collected at the beamline X06SA-PX of the Swiss Light Source (SLS) at the Paul Scherrer

Institut in Villigen, Switzerland, using the wavelength 1.000Å and a PILATUS 6M detector. Data reduction was done with the XDS package[5] in space group $P3_121$ with cell dimensions of a,b=108.7Å, c=90.7Å and a,b=108.8Å, c=90.5Å and to resolutions of 2.2 and 2.0Å for the $dT^{Me}TP$ and $dT^{Et}TP$ trapped structures, respectively. The two structures were solved by difference Fourier techniques. Refinement was performed with PHENIX[6] and model rebuilding was done with COOT.[7] Figures of the structures were created in PyMOL.[8]

Accession numbers

The coordinates and structure factors have been deposited in the Protein Data Bank with the accession numbers 3M8S and 3M8R.

***KlenTaq* DNA Polymerase mutation, expression, and purification**

Mutant I614A was introduced into the *KlenTaq* DNA polymerase ORF via site-directed mutagenesis with the following primers: forward: 5'-d(CCTGGACTATAGCCAGGCAGAGCTCAGGGTGCTG)-3'; reverse: 5'-d(CAGCACCCTGAGCTCTGCCTGGCTATAGTCCAGG)-3'. The sequence was determined by DNA sequencing. Expression of the *KlenTaq* DNA polymerases wt and I614A (in pGDR11 vector) was conducted in *E. coli* BL21 (DE3) Gold. Expression cultures were inoculated with 2% of an overnight culture grown at 30°C, and expression was induced at OD600 = 0.5 with IPTG at a final concentration of 1 mM. After 4 hr of expression, cultures were centrifuged at 4500xg at 4°C for 10 min. All cultures were grown in LB medium (Roth) with 100 mg/l carbenicillin. Collected cell pellets were resuspended in lysis-buffer (10mM Tris-HCl [pH 9.2], 300mM NaCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, 1 mg/ml lysozyme, 5 ml lysis buffer/50 ml culture) and lysed at 37°C for 10 min. *E. coli* proteins were denaturated through heating to 75°C for 45 min with a subsequent centrifugation at 15,000xg for 30 min at 4°C. Centrifuged expression lysates were incubated with pre-equilibrated (in 10 mM Tris-HCl [pH 9.2], 0.3 M NaCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100) NTA-matrix (QIAGEN, 2–3 ml slurry/20 ml lysate) for 60 min at 4°C in an overhead shaker. The NTA matrix was then washed twice with NTA-wash-buffer (10 mM Tris-HCl [pH 9.2], 0.3 M NaCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, 20 mM imidazole), and the protein was eluted twice with NTA-elution-buffer (10 mM Tris-HCl [pH 9.2], 0.3 M NaCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, 200 mM imidazole). The elution buffer was exchanged by storage buffer (elution buffer without imidazole) using VivaSpins. For storage, glycerol was added to 50%, $(NH_4)_2SO_4$ to a final concentration of 16 mM and Tween 20 to 0.1%. Enzyme purity was controlled by SDS-PAGE. Protein concentrations were determined using Bradford assay. Purified enzymes were stored at -20°C.

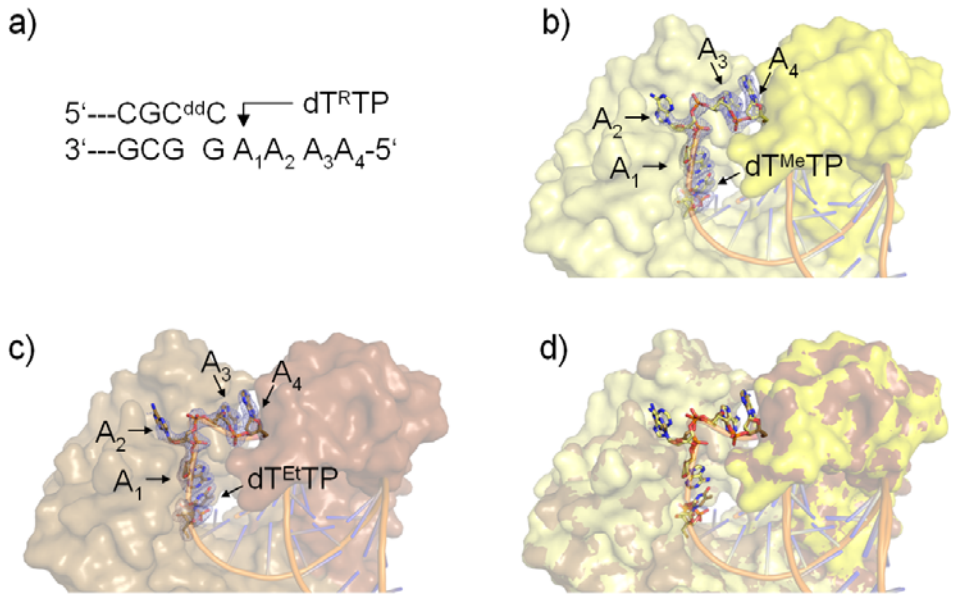

Figure S1. Orientation of single-stranded template strand. a) Numbering template nucleotides. b) Model of structure derived from ternary complex of *KlenTaq*, primer, template, and dT^{Me}TP. The respective nucleotides are shown as stick model with carbon atoms in yellow, oxygen atoms in red, nitrogen atoms in blue and phosphorous atoms in orange, and are superimposed with the final refined mFo-DFc simulated annealing omit map. The protein is shown as Connolly surface. c) Same as in b) for complex containing dT^{Et}TP. d) Superposition of structures depicted in b) and c) using the same colour code.

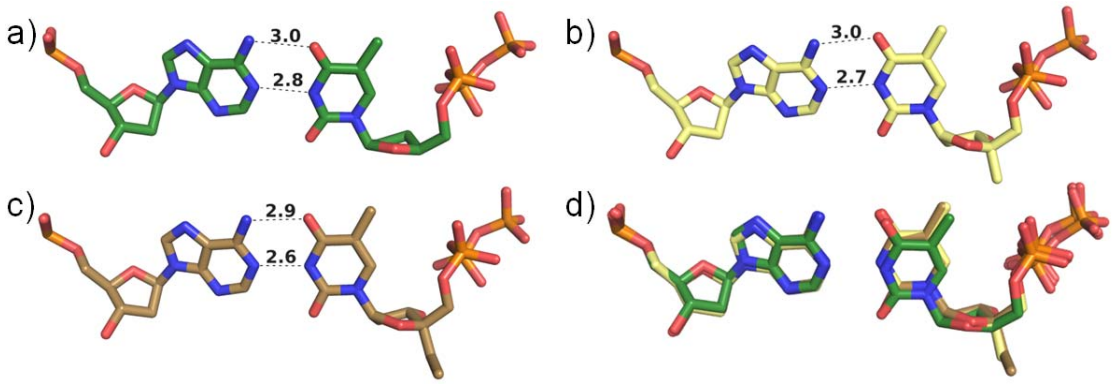

Figure S2. Close-up view on coding adenosine and incoming thymidine analogue. a) Stick model of structure derived from ternary complex of *KlenTaq*, primer, template, and ddT^{H}TP (pdb: 1QTM). The indicated distances were determined by PyMOL and are shown in Å. b) Same as in a) for complex containing dT^{Me}TP. c) Same as in a) for complex containing dT^{Et}TP. d) Superposition of structures depicted in a)-c) using the colour code of Figure 1.

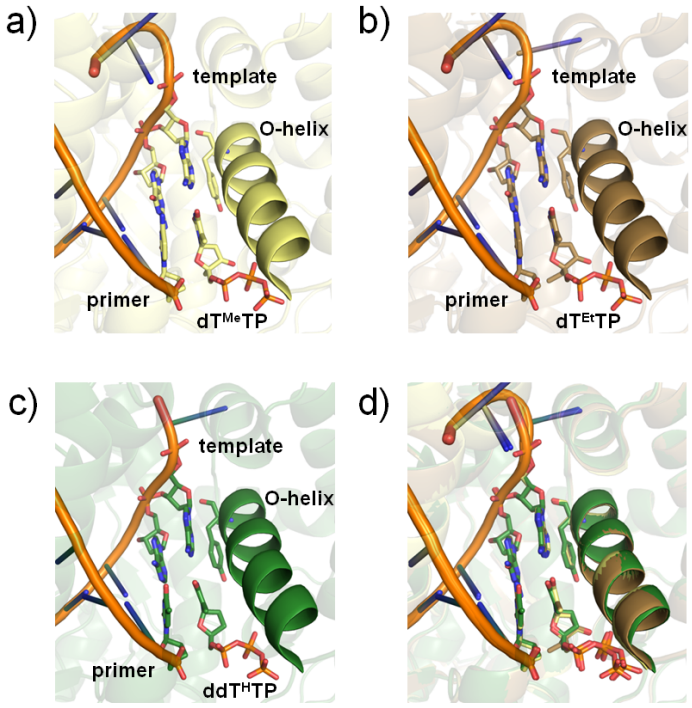

Figure S3. Close-up views highlighting the conformation of the O-helix. a) Ternary complex of *KlenTaq*, primer, template, and $dT^{Me}TP$. The primer terminus and the nascent nucleotide pair are depicted in sticks. b) Same as in a) for complex containing $dT^{Et}TP$. c) Same as in a) using the reported data containing $ddT^{H}TP$ (pdb: 1QTM). d) Superposition of structures depicted in a)-c) using the colour code of Figure 1.

Table S1. Summary of crystallographic data

Crystal	***KlenTaq*** **$dT^{Me}TP$**	***KlenTaq*** **$dT^{Et}TP$**
Data collection		
Wavelength (Å)	1.000	1.000
Resolution (Å) [a]	47.1-2.20 (2.33-2.20)	47.1-2.00 (2.12-2.00)
Total reflections [a]	311920 (46718)	420922 (67045)
Unique reflections [a]	46718 (5046)	42216 (6678)
Completeness (%)[a]	99.7 (98.3)	99.8 (98.6)
Mean I/σ[a]	12.9 (1.9)	22.1 (2.6)
R_{meas} (%)[a]	11.5 (105.2)	7.5 (87.9)
Refinement		
Resolution range (Å) [a]	47.1-2.20 (2.27-2.20)	47.1-2.00 (2.05-2.00)
R_{cryst} (%)[a]	17.0 (25.8)	18.0 (21.6)
R_{free} (%)[a]	22.4 (36.2)	22.1 (28.0)
Protein residues	539	539
DNA (primer/template)	12/16	12/16
nucleotides	$dT^{Me}TP$	$dT^{Et}TP$
Water molecules	123	304
Mg^{2+} ions	2	2
Average B factors (Å^2)		
Protein	44.4	47.9
DNA	40.6	39.8
$dT^{R}TP$	33.4	39.7
r.m.s. deviations		
Bond length (Å)	0.012	0.011
Bond angle (°)	0.97	0.83
Max. likelihood estimate for coord. error (Å)	0.27	0.26
Ramachandran outliers	1	1

R_{meas} = redundancy independent R-factor (intensities)
(For definition of R_{meas} see Diederichs, K. & Karplus, P. A. *Nature Struct. Biol.* **4**, 269-275 (1997))
[a]Values in parentheses correspond to those in the outer resolution shell

Table S2. P-values of sugar moieties in primer, template and nucleotides of the structures containing $dT^{Me}TP$ and $dT^{Et}TP$, respectively determined using the PROSIT server.[9] It should be noted that the P-value (column 7) was not specifically restrained during refinement in phenix.refine.

dT^{Me}-Template

No.	v_0	v_1	v_2	v_3	v_4	***P***	vmax	χ	γ	TYPE		
A	-13,4	25,0	-26,6	19,4	-3,9	**169,5**	27,0	-127,9	50,8	C2'-endo	B-Form	Southern
A	-31,8	45,2	-40,7	23,8	4,8	**155,2**	44,8	99,3	43,2	C2'-endo	B-Form	Southern
A	-6,5	-17,4	33,2	-37,8	28,0	**28,2**	37,6	4,6	60,3	C3'-endo	A-Form	Northern
A	7,5	-25,1	32,2	-29,0	13,6	**5,7**	32,4	-171,5	177,9	C3'-endo	A-Form	Northern
G	2,7	-27,6	40,0	-39,6	23,5	**14,9**	41,4	-160,0	52,8	C3'-endo	A-Form	Northern
G	-4,1	-22,4	38,7	-42,2	29,1	**24,0**	42,3	-148,7	43,7	C3'-endo	A-Form	Northern
C	-2,2	-17,4	28,9	-30,9	21,1	**22,5**	31,3	-169,5	177,8	C3'-endo	A-Form	Northern
G	-14,1	29,2	-32,3	25,2	-7,2	**173,8**	32,5	-105,6	50,3	C2'-endo	B-Form	Southern
C	-14,0	29,0	-32,2	25,0	-7,1	**173,8**	32,4	-98,7	39,5	C2'-endo	B-Form	Southern
C	-9,1	-9,7	23,5	-29,4	24,3	**36,3**	29,1	-141,9	45,4	C4'-exo	A-Form	Northern
G	-13,8	25,7	-27,4	20,0	-4,1	**169,6**	27,8	-110,6	51,1	C2'-endo	B-Form	Southern
T	-29,6	36,6	-29,9	13,6	9,9	**145,8**	36,1	-110,9	47,5	C1'-exo	B-Form	Southern
G	-34,1	37,7	-27,1	8,3	16,0	**136,3**	37,4	-106,3	43,8	C1'-exo	B-Form	Southern
G	-37,9	34,4	-18,5	-2,8	25,3	**119,5**	37,5	-114,2	36,2	C1'-exo	B-Form	Southern
T	-37,8	38,2	-24,5	3,4	21,5	**128,7**	39,2	-131,2	57,8	C1'-exo	B-Form	Southern
C	-40,4	44,1	-31,3	9,0	19,5	**135,4**	44,0	-117,0	46,9	C1'-exo	B-Form	Southern

dT^{Me}-Primer

No.	v_0	v_1	v_2	v_3	v_4	***P***	vmax	χ	γ	TYPE		
G	-14,8	-11,9	32,1	-41,5	35,5	**39,0**	41,2	-166,7	-10,6	C4'-exo	A-Form	Northern
A	-8,0	-1,6	9,9	-14,8	14,4	**49,4**	15,3	-137,1	46,0	C4'-exo	A-Form	Northern
C	23,2	-30,3	25,7	-13,1	-6,2	**329,5**	29,8	-157,0	-144,9	C2'-exo	A-Form	Northern
C	-21,8	10,9	3,1	-15,7	23,7	**82,5**	23,6	-138,1	38,9	O4'-endo	A-Form	Northern
A	-9,5	28,4	-35,8	31,1	-13,8	**183,7**	35,8	-102,8	46,7	C3'-exo	B-Form	Southern
C	-34,8	23,0	-3,5	-16,6	32,5	**95,8**	34,9	-118,7	43,8	O4'-endo	B-Form	Southern
G	-30,8	36,1	-28,0	11,1	12,2	**141,7**	35,6	-106,3	49,1	C1'-exo	B-Form	Southern
G	-38,7	28,0	-7,4	-14,8	33,6	**101,1**	38,1	-120,9	44,2	O4'-endo	B-Form	Southern
C	-34,7	44,5	-37,0	18,1	10,2	**147,9**	43,6	-96,0	50,5	C2'-endo	B-Form	Southern
G	3,1	-27,3	39,4	-38,7	22,8	**14,4**	40,7	-149,6	38,0	C3'-endo	A-Form	Northern
C	-3,8	-18,7	32,6	-35,8	25,1	**24,7**	35,9	-151,6	49,2	C3'-endo	A-Form	Northern
ddC	8,5	-21,3	25,3	-21,2	8,1	**359,6**	25,3	-149,5	73,0	C3'-endo	A-Form	Northern
$dT^{Me}TP$	21,7	-37,2	38,0	-24,7	1,2	**344,3**	39,5	-138,0	63,4	C2'-exo	A-Form	Northern

dT^{Et}-Template

No.	ν_0	ν_1	ν_2	ν_3	ν_4	***P***	vmax	χ	γ	TYPE		
A	-19,8	33,0	-33,0	22,5	-1,9	**164,4**	34,3	-114,9	49,3	C2'-endo	B-Form	Southern
A	-33,6	47,3	-42,4	24,5	5,5	**154,6**	47,0	90,3	45,5	C2'-endo	B-Form	Southern
A	26,9	-31,6	24,5	-9,6	-10,7	**321,7**	31,2	171,1	-40,7	C2'-exo	A-Form	Northern
A	4,4	-23,4	32,3	-30,7	16,7	**11,1**	32,9	-174,6	179,2	C3'-endo	A-Form	Northern
G	0,3	-25,0	38,8	-40,0	25,0	**18,4**	40,9	-161,2	53,2	C3'-endo	A-Form	Northern
G	-3,1	-22,2	37,3	-40,1	27,3	**22,8**	40,4	-149,7	45,1	C3'-endo	A-Form	Northern
C	-7,5	-13,7	28,1	-33,2	26,0	**31,5**	33,0	-173,6	178,5	C3'-endo	A-Form	Northern
G	-14,9	28,8	-30,9	23,3	-5,6	**171,2**	31,3	-104,9	51,9	C2'-endo	B-Form	Southern
C	-19,6	29,9	-28,3	17,8	1,0	**159,4**	30,2	-106,9	44,2	C2'-endo	B-Form	Southern
C	-12,4	-5,9	20,5	-28,3	25,7	**43,8**	28,4	-142,8	42,2	C4'-exo	A-Form	Northern
G	-12,7	23,1	-24,4	17,8	-3,3	**168,8**	24,9	-107,6	47,8	C2'-endo	B-Form	Southern
T	-33,2	35,1	-24,4	6,0	17,0	**133,4**	35,5	-118,5	53,6	C1'-exo	B-Form	Southern
G	-34,5	32,1	-18,1	-1,3	22,4	**121,7**	34,4	-113,4	41,6	C1'-exo	B-Form	Southern
G	-35,5	34,3	-20,4	0,5	21,9	**124,6**	36,0	-117,9	49,7	C1'-exo	B-Form	Southern
T	-43,8	35,7	-14,9	-10,0	33,6	**110,4**	42,7	-136,4	58,9	O4'-endo	B-Form	Southern
C	-37,1	46,7	-38,0	17,9	11,8	**146,4**	45,6	-108,9	46,3	C2'-endo	B-Form	Southern

dT^{Et}-Primer

No.	ν_0	ν_1	ν_2	ν_3	ν_4	***P***	vmax	χ	γ	TYPE		
G	-11,8	-13,4	31,6	-39,5	32,2	**35,8**	39,0	-161,0	-15,3	C3'-endo	A-Form	Northern
A	-1,0	-16,6	26,6	-27,9	18,3	**20,4**	28,4	-148,0	52,8	C3'-endo	A-Form	Northern
C	25,1	-29,9	23,2	-9,4	-9,8	**322,2**	29,4	-146,5	-103,5	C2'-exo	A-Form	Northern
C	-27,4	27,5	-17,5	2,2	15,7	**128,2**	28,3	-127,5	46,4	C1'-exo	B-Form	Southern
A	-13,6	29,5	-33,4	26,4	-8,2	**175,3**	33,5	-104,3	37,3	C2'-endo	B-Form	Southern
C	-37,6	24,5	-3,6	-18,0	35,1	**95,4**	37,6	-121,9	45,7	O4'-endo	B-Form	Southern
G	-30,8	35,7	-27,1	10,1	12,9	**140,3**	35,3	-107,3	52,0	C1'-exo	B-Form	Southern
G	-42,6	32,7	-10,9	-13,4	35,0	**105,2**	41,6	-120,8	41,8	O4'-endo	B-Form	Southern
C	-34,5	45,8	-39,1	20,6	8,5	**150,5**	44,9	-91,4	46,8	C2'-endo	B-Form	Southern
G	2,0	-27,2	40,2	-40,0	24,4	**15,9**	41,8	-151,4	37,4	C3'-endo	A-Form	Northern
C	-7,5	-16,5	32,5	-37,8	28,7	**29,9**	37,5	-149,9	54,2	C3'-endo	A-Form	Northern
ddC	10,4	-25,6	29,4	-25,1	9,3	**359,0**	29,4	-149,2	56,2	C2'-exo	A-Form	Northern
dT^{Et}TP	20,3	-38,2	40,9	-28,2	4,5	**348,4**	41,7	-139,2	66,3	C2'-exo	A-Form	Northern

Table S3. Motif A consensus sequence alignement

```
Thermus aquaticus                  DYSQIELR
Thermus thermophilus               DYSQIELR
Thermus caldophilus                DYSQIELR
Thermus filiformis                 DYSQIELR
Rickettsia felis                   DYSQIELR
Mycobacterium tuberculosis         DYSQIEMR
Homo sapiens (pol theta)           DYSQIELR
Escherichia coli (pol I)           DYSQIELR
Chlamydia trachomatis (pol I)      DYSQIELR
Chlamydia pneumoniae (pol I)       DYSQIELR
Treponema pallidum (pol I)         DYSQIELV
Methylobacterium (pol I)           DYSQIELR
Deinococcus radiodurans (pol I)    DYSQIELR
Helicobacter pylori (pol I)        DYSQIELR
Borrelia burgdorferi (pol I)       DYSQIELA
Bacillus subtilis (pol I)          DYSQIELR
```

References

[1] a) Y. Li, Y. Kong, S. Korolev, G. Waksman, Protein **1998**, *7*, 1116-23. b) Korolev, M. Nayal, W. M. Barnes, E. Di Cera, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9264-8. c) Y. Li, S. Korolev, G. Waksman, *EMBO J.* **1998**, *17*, 7514-25. d) Y. Li, V. Mitaxov, G.Waksman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9491-6. e) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225-33.

[2] Streckenbach, G. Rangam, H. M. Möller, A. Marx, *ChemBioChem* **2009**, *10*, 1630-1633.

[3] F. Di Pasquale, D. Fischer, D. Grohmann, T. Restle, A. Geyer, A. Marx, *J. Am. Chem. Soc.* **2008**, *130*, 10748-10757.

[4] P. J. Rothwell, V. Mitaksov, G. Walksman, *Mol. Cell* **2005**, 19, 345-355.

[5] W. Kabsch, *J. Appl. Cryst.* **1993**, *26*, 795-800.

[6] P. D. Adams, R. W. Grosse-Kunstleve, L.-W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, T. C. Terwilliger, *Acta Cryst.* **2002**, *D58*, 1948-1954.

[7] P. Emsley, K. Cowtan, *Acta Cryst*. **2004**, *D60*, 2126-2132.

[8] W. L. DeLano, DeLano Scientific, Palo Alto, CA, USA **2002**.

[9] G. Sun, J.H. Voigt, I.V. Filippov, V.E. Marquez, M.C. Nicklaus, *J Chem Inf Comput Sci*, **2004**, 44, 1752-1762.