ChemPubSoc Europe

DOI: 10.1002/cbic.201300175

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

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

DNA polymerases are the key enzymes in many biotechnological applications, such as genome sequencing, molecular diagnostics, DNA conjugation and selection of aptamers by systematic enrichment of ligands by exponential amplification (SELEX).[1] For these methods the ability of DNA polymerases to process modified 2′-deoxynucleoside-5′-*O*-triphosphates (dNTPs) is often essential.[1a,2]

Functional studies with nucleotides that are modified at the sugar residue[3] or the nucleobase[4] have shown many times that archaeal DNA polymerases belonging to sequence family B (e.g., *Thermococcus* sp. 9°N-7 (9°N) DNA polymerase, *Pyrococcus furiosus* (Pfu) DNA polymerase and *Thermococcus kodakaraensis* (KOD) DNA polymerase) are more efficient at utilising modified nucleotides than are DNA polymerases from sequence family A, such as *Thermus aquaticus* (Taq) DNA polymerase. The reason for this remains elusive, mainly because of paucity of structural data. Although structures of family A DNA polymerases are available (even in complex with modified nucleotides),[5] family B DNA polymerases from archaea are only very poorly defined: only the apo-structures and those of enzymes with the DNA at the exonuclease subunit of the enzyme (editing mode) have been characterised.[6]

Here we report two structures of family B DNA polymerases, KOD and 9°N. They were obtained in binary (enzyme bound to primer/template duplex) complexes, with primer and template strand positioned in the polymerase cleft in the replicative state. We were able to solve the crystal structures of binary KOD (PDB ID: 4K8Z) and 9°N (PDB ID: 4K8X) complexes at resolutions of 2.4 and 2.6 Å, respectively.

An important step towards obtaining the binary structures was the identification of crystals that bear complexed DNA. It turned out that both enzymes readily crystallise in apo form, even though DNA was present. Therefore, DNA binding had to be identified by an approach in which Cy-labelled DNA assists visualisation and identification of DNA-containing crystals.[7] Coloured crystals of both polymerases were obtained under different crystallisation conditions and were reproduced with unlabelled DNA. In the case of 9°N, the resolution decreased slightly with the unlabelled template. As the primer/template (p/t) duplex parts were identical in labelled- and unlabelled-template structures, we use the Cy5-containing 9°N structure for analysis in this study. In contrast, refined crystals of KOD delivered better diffraction characteristics without the dye; thus, the structure with the unlabelled template was used.

Despite the presence of magnesium ions and ddNTP in the crystallisation trials, no electron density for bound triphosphate and/or coordinated magnesium ions was found in the active sites of either enzyme. The most probable reason for this is the fact that in both crystal structures the tip of the finger domain is in contact with the exonuclease domain of a symmetry-related molecule. Furthermore, a second symmetry mate could sterically hinder finger domain closure (see Figure S1 in the Supporting Information).

The overall structures of KOD and 9°N are very similar (Figure S2). They show the typical domain composition of DNA polymerases: N-terminal (N-term, 1–130 and 338–372), exonuclease (exo, 131–338), finger (448–499), thumb (591–774) and palm (374–447 and 500–590)[8] (Figure 1; depictions of 9°N can be found in the Supporting Information).

Although the thumb domain was further partitioned into two subdomains (thumb-1 and thumb-2) in the publication of the KOD apo structure,[6b] we do not discriminate these two subdomains here. When comparing our binary structures with the KOD apo structure (PDB ID: 1WNS),[6b] it is evident that there is some rearrangement in the palm and thumb domains of KOD and 9°N upon DNA binding (Figure 1). The KOD binary and apo structures align with a Cα RMSD of 1.31 Å. The most pronounced movement of structural elements upon DNA binding are within the thumb domain, which rotates towards the DNA (Figure 1). In the following discussion of the structures we employ the nomenclature introduced for the 9°N apo structure.[6a] Reordered and additionally built structural elements within the thumb domain include helices R, S, T, U and V+W, β-sheets 26 to 28 and the connecting loops. The biggest movement is observed for helix U, which is shifted (~22 Å) and rotated (~27°) towards the DNA. In both structures this inward movement generates a distinct contact region between thumb and DNA, and the otherwise only partly resolved thumb domain is more structured (Figure 1). After the thumb domain, the biggest rearrangements are observed for helix L (palm domain, residue 374–379, located at the interface to the finger domain) and the loop connecting it to β-sheet 18 (residues 397–404). Starting at Val 389, the loop together with the helix moves by up to 4 Å outwards, thereby opening the template cleft (dark green in Figure 1).

In both structures the duplex part of the DNA is bound, almost identically, in a groove formed by the thumb and palm domains. For instance, KOD contacts the double-stranded DNA (dsDNA) between nucleotide dG_{T14} and dC_{T6} on the template

[a] *K. Bergen,[+] K. Betz,[+] Prof. W. Welte, Prof. K. Diederichs, Prof. A. Marx*
Departments of Chemistry and Biology
Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)
E-mail: Andreas.Marx@uni-konstanz.de

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201300175.

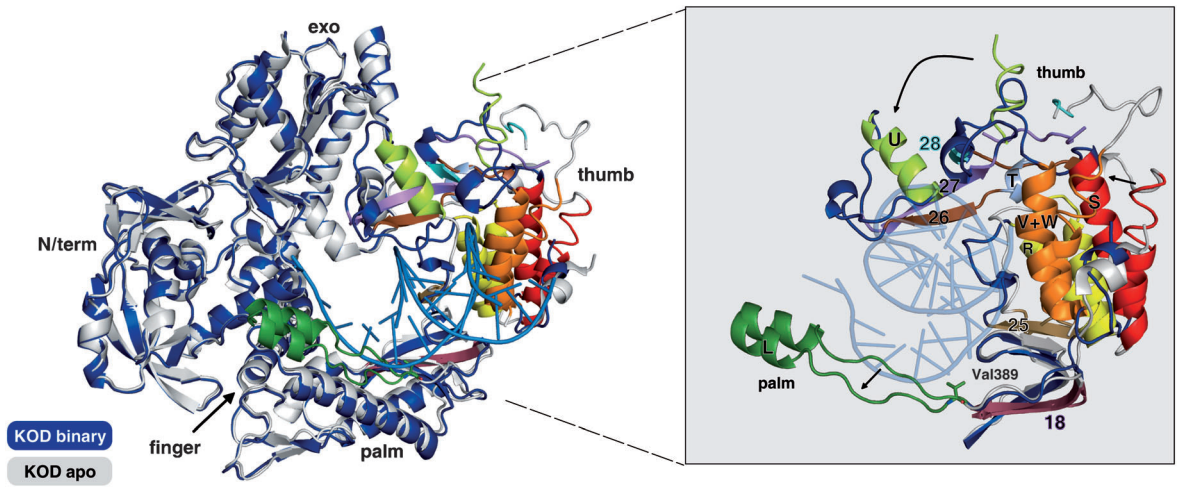

Figure 1. Overlay of KOD DNA polymerase in the binary complex (violet) with the KOD apo structure (PDB ID: 1WNS, grey). Polymerase domains are labelled. Detail: altered elements of thumb and palm domains. Corresponding secondary structure elements of apo and binary structures are shown in the same colours to highlight movement of the thumb domain upon DNA binding and visualise differing parts of the enzyme.

strand and dC_{P7} and ddA on the primer strand (for p/t numbering see Figures 2 and S2). 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. The residues directly interacting with the DNA are identical in both enzymes, with one exception; this is reflected in the sequence differences between KOD and 9°N (Figure S2 A; overlay in Figure S2 B). Based on the high structural and sequence similarity

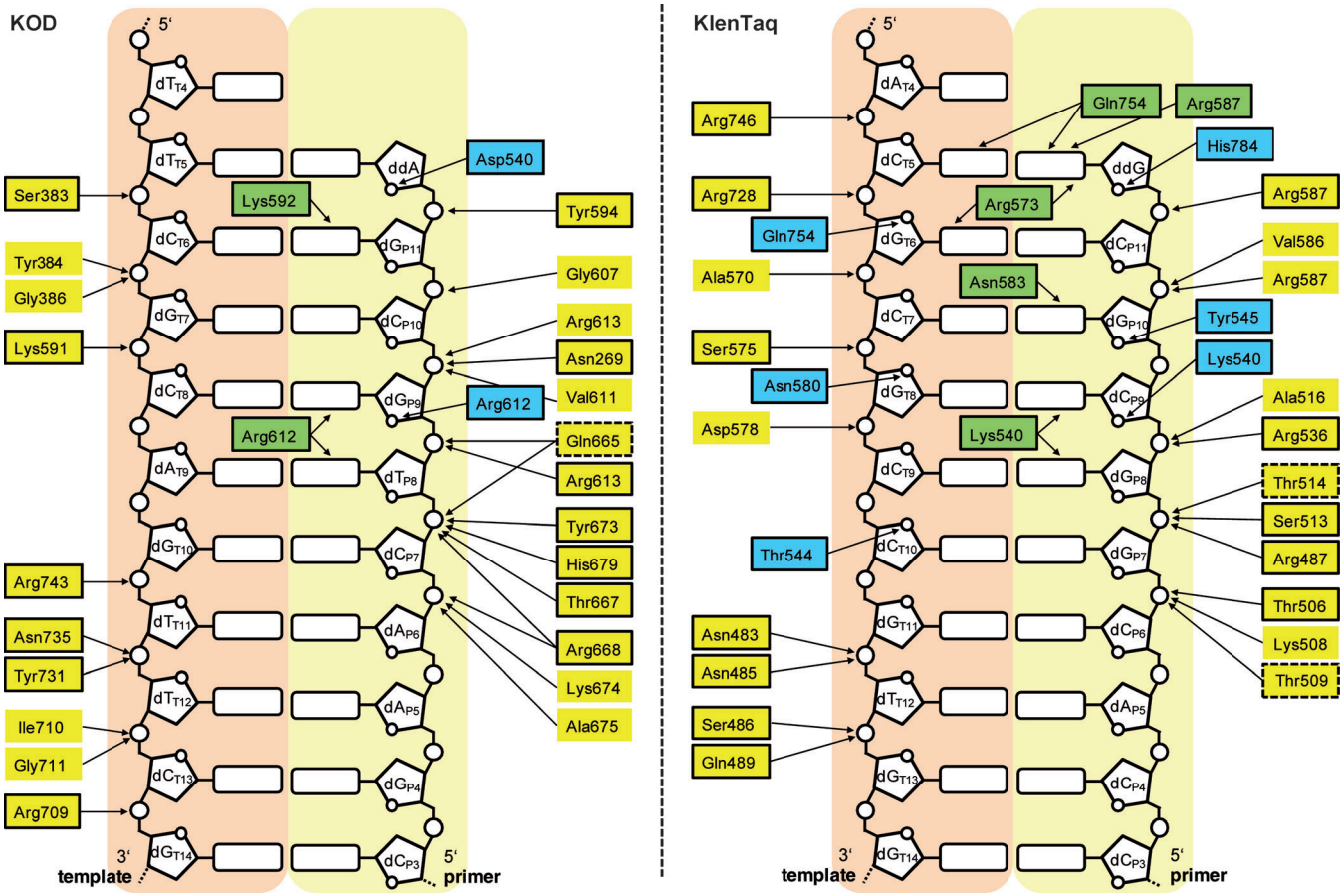

Figure 2. Comparison of protein–DNA interactions in KOD (B family) and KlenTaq (A family, PDB ID: 3SZ2) polymerases. Only direct contacts up to 3.6 Å are shown: side-chain interactions (solid outline), contacts with protein backbone (no outline) and residues with both interactions (dashed outline). Interactions to the phosphate backbone, sugar oxygen and nucleobase are shown in yellow, blue and green, respectively.

between these enzymes (Cα RMSD: 0.76 Å, sequence identity: 91%) the structure of KOD was chosen for a more detailed description in the following analysis; it is representative of both polymerases.

As observed in the extensively studied B-family polymerase RB69, the p/t duplex in KOD and 9°N binary structures 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).[9] The duplex also follows other known geometric characteristics for B-DNA. As already stated for other binary or ternary B-family structures, this is in contrast to the A-form DNA observed at the primer 3′-end in A-family polymerases, as reported for T7, BF (large fragment of *Bacillus stearothermophilus* polymerase) and KlenTaq (large fragment of the polymerase I from *Thermus aquaticus)* DNA polymerases.[10]

Similarly to other B-family polymerases,[11] KOD forms direct contacts with the DNA duplex, primarily with the DNA phosphate backbone. Only very few direct interactions with the nucleobases or sugar oxygens are observed (Figure 2). Residues in the loop connecting β-sheet 26 and helix U (666–677) in the thumb domain establish the first contacts to the primer strand. A second loop between β-sheet 25 and helix R (606–616) interacts with the primer at the phosphate backbone and the minor groove. The 3′ part of the template is contacted by residues in the connecting loop between β-sheets 27 and 28 (709–711) and by residues in helix V+W (729–745). Some residues known to be important in DNA binding are situated in motifs conserved throughout family B polymerases; these include, for example, the sequence motif KKRY (KKKY in KOD, residues 591–594), which is unique to family B DNA polymerases[12] and mediates contacts to the p/t duplex near the active site. A second conserved region (I/YxGG/A motif) is situated in the previously described loop in the palm domain and stabilises the template strand at the 5′-end of the duplex (Figure 1, dark green). Apart from those in conserved motifs, residues involved in DNA binding are not conserved throughout the B family.[11] However, we found that the great majority of residues that directly contact the DNA duplex are indeed well conserved in archaeal members (Figure S3).

The obtained structural insights in this study might aid rationalisation of the observed differences in the capability of family B and A DNA polymerases in polymerising chemically modified nucleotides. Thus, we compared the data reported here with a representative family A polymerase, the structurally well-characterised KlenTaq DNA polymerase, by employing the KlenTaq binary structure with the PDB ID: 3SZ2 (Figure 2 and 3).[5b–d, 10c, 13] In both enzymes the majority of direct protein side- and main-chain contacts to dsDNA are at the phosphate backbone. The most striking difference, however, is the number of direct contacts to the nucleobases and sugar oxygens. Whereas in KlenTaq six nucleobases are contacted by five protein side chains, in the KOD structure only three nucleobases are contacted by just two residues. Interestingly, in both enzymes, interactions with the nucleobases of two primer nucleotides (8 and 9) by one protein side chain were found (Lys540 in KlenTaq, Arg612 in KOD). Furthermore, remarkably fewer interactions with the ribose moiety were found in KOD compared to KlenTaq. Whereas in KlenTaq six interactions of sugar oxygens with protein side chains were observed, only two residues in KOD (Asp540 and Arg612) interact with a sugar.

As all sugar and nucleobase contacts are through the minor groove, one would assume a sterically less-hindered minor groove in the case of B-family polymerases. This has been probed by modifying nucleotides, and might explain why B-family polymerases show better performance in utilising nucleotides that, for example, bear small C4′ modifications (i.e., methyl, ethyl, vinyl, methylene, etc.).[3] Nevertheless, bulkier modifications might not be tolerated by either enzyme as these would interfere with residues (of both proteins) that are further from the interior of the minor groove (Figure S4). In general, for complexes with both enzymes, the DNA minor groove is better covered by protein residues than is the major groove.

Modifications at C5 in pyrimidines and C7 of 7-deazapurines (which have been shown to be more efficiently polymerised by family B DNA polymerases)[4] point towards the major groove. Thus, we next inspected the major groove of DNA in the KOD and KlenTaq binary structures. The number of residues interacting with the phosphate backbone is similar, and, in the 3D model, the major grooves seem to be well accessible in both enzymes (Figures 3 and S4C, D). However, one major difference is the tip of the thumb domain. In both enzymes the tip of the thumb domain (residues 506–509 in KlenTaq, 668–675 in KOD; red in Figures 3A, B and S4C, D) makes contacts with the primer strand. The architectures of the contacting loops differ greatly. Whereas the loop in the KOD structure hovers above the minor groove without extending deeply into it (Figures 3A and S4C), the corresponding loop in KlenTaq extends over the phosphate backbone towards the major groove (Figures 3B and S4D). The better accessibility of the major groove might explain the observed superiority of family B DNA polymerases in polymerising nucleotides that bear bulky modifications at C5 in pyrimidines and C7 of 7-deazapurines.[4]

In summary, we describe the crystal structures of archaeal family B DNA polymerases, namely KOD and 9°N, in binary complexes with DNA in the replicative mode. As, up to now, only structures of bacterial or eukaryotic members of the family in a binary or ternary replicative complex have been characterised, these structures contribute to a better mechanistic understanding of thermophilic B-family polymerases of Archaea, which are extensively used when modified nucleotides are used as substrates. With these new models at hand, the generation of optimised enzymes and substrates is facilitated. Despite this progress, one remaining challenge is the crystallisation of ternary complexes of archaeal B-family polymerases with dsDNA and bound triphosphate in the active site. 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.

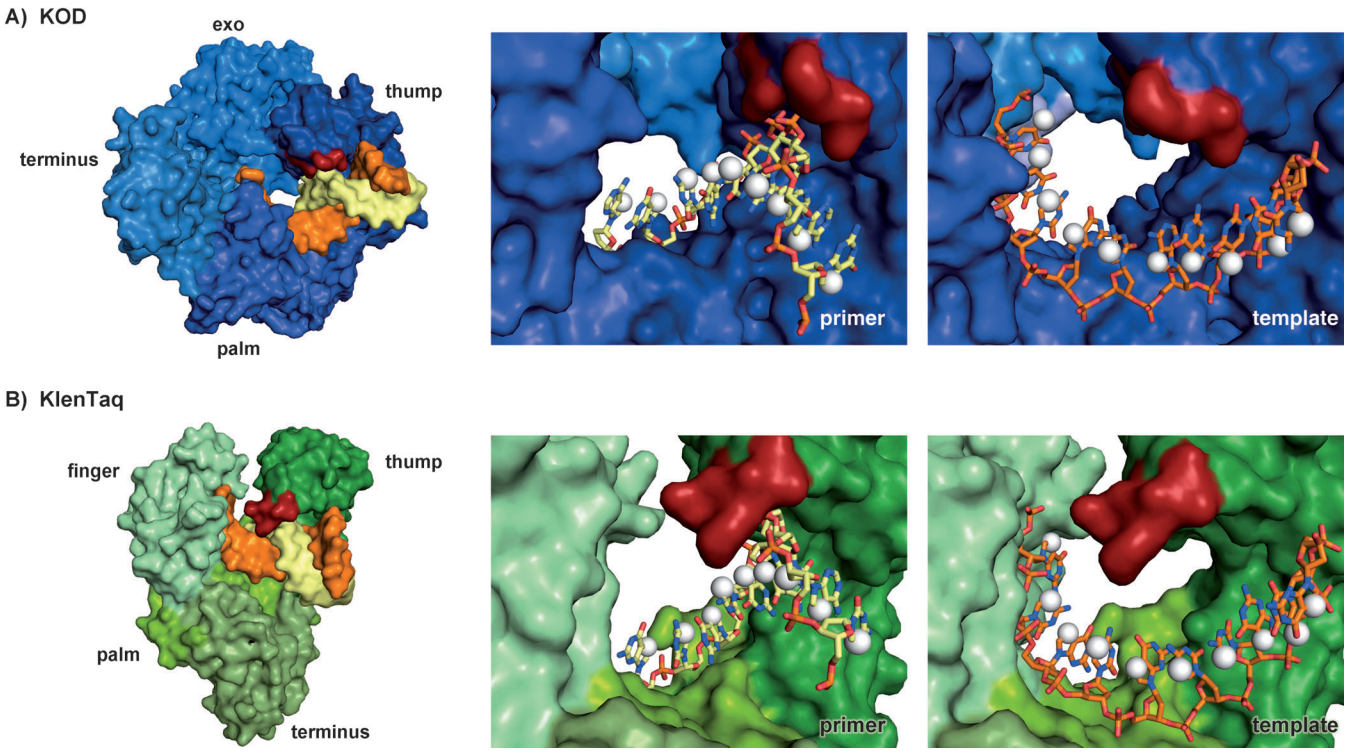

Figure 3. Comparison of DNA environment in family B and family A DNA polymerases. A) KOD and B) KlenTaq (PDB ID: 3SZ2) binary complexes are shown, with the tip of the thumb domain in red in both structures. (The finger domain of KOD is not visible in this orientation.) DNA primer and template are shown as yellow and orange surface or sticks. Middle and right panels show details of the primer and template binding clefts. The C5 in pyrimidines and C7 of 7-deazapurines nucleobases, which are mainly used for attachment of modifications, are marked as white spheres.

Acknowledgements

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

Keywords: DNA replication · family B · modified nucleotide · polymerase chain reaction · X-ray structure

[1] a) R. *Kranaster*, A. *Marx*, *ChemBioChem* **2010**, *11*, 2077; b) G. Mayer, *Angew. Chem.* **2009**, *121*, 2710 – 2727; *Angew. Chem. Int. Ed.* **2009**, *48*, 2672 – 2689; c) S. H. Weisbrod, A. Marx, *Chem. Commun.* **2008**, 5675 – 5685.

[2] a) D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. K. Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, et al., *Nature* **2008**, *456*, 53 – 59; b) T. D. Harris, P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J. W. Efcavitch, F. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S. R. Quake, K. Steinmann, E. Thayer, A. Tyurina, et al., *Science* **2008**, *320*, 106 – 109; c) T. Ohbayashi, M. Kuwahara, M. Hasegawa, T. Kasamatsu, T. Tamura, H. Sawai, *Org. Biomol. Chem.* **2005**, *3*, 2463 – 2468; d) T. S. Seo, X. Bai, D. H. Kim, Q. Meng, S. Shi, H. Ruparel, Z. Li, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5926 – 5931.

[3] a) R. N. Veedu, B. Vester, J. Wengel, *Org. Biomol. Chem.* **2009**, *7*, 1404 – 1409; b) R. N. Veedu, B. Vester, J. Wengel, *J. Am. Chem. Soc.* **2008**, *130*, 8124 – 8125; c) R. N. Veedu, B. Vester, J. Wengel, *ChemBioChem* **2007**, *8*, 490 – 492; d) F. Di Pasquale, D. Fischer, D. Grohmann, T. Restle, A. Geyer, A. Marx, *J. Am. Chem. Soc.* **2008**, *130*, 10748 – 10757; e) M. Strerath, J. Gaster, D. Summerer, A. Marx, *ChemBioChem* **2004**, *5*, 333 – 339; f) M. Strerath, A. Marx, *Angew. Chem.* **2002**, *114*, 4961 – 4963; *Angew. Chem. Int. Ed.* **2002**, *41*, 4766 – 4769.

[4] a) A. Baccaro, A.-L. Steck, A. Marx, *Angew. Chem.* **2012**, *124*, 260 – 263; *Angew. Chem. Int. Ed.* **2012**, *51*, 254 – 257; b) O. Thum, S. Jäger, M. Famulok, *Angew. Chem.* **2001**, *113*, 4112 – 4115; *Angew. Chem. Int. Ed.* **2001**, *40*, 3990 – 3993; c) 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, K. D. Janda, *Angew. Chem.* **2005**, *117*, 2182 – 2186; *Angew. Chem. Int. Ed.* **2005**, *44*, 2144 – 2148; d) M. Hocek, M. Fojta, *Org. Biomol. Chem.* **2008**, *6*, 2233 – 2241; e) S. Ikonen, H. Macíčková-Cahová, R. Pohl, M. Sanda, M. Hocek, *Org. Biomol. Chem.* **2010**, *8*, 1194 – 1201; f) H. Weizman, Y. Tor, *J. Am. Chem. Soc.* **2002**, *124*, 1568 – 1569; g) G. Giller, T. Tasara, B. Angerer, K. Mühlegger, M. Amacker, H. Winter, *Nucleic Acids Res.* **2003**, *31*, 2630 – 2635; h) J. P. Anderson, B. Angerer, L. A. Loeb, *Biotechniques* **2005**, *38*, 257 – 264; i) P. M. E. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem.* **2008**, *120*, 8478 – 8487; *Angew. Chem. Int. Ed.* **2008**, *47*, 8350; j) S. Kumar, A. Sood, S. Rao, J. Nelson, US 7109316, **2006**; k) S. Jäger, M. Famulok, *Angew. Chem.* **2004**, *116*, 3399 – 3403; *Angew. Chem. Int. Ed.* **2004**, *43*, 3337 – 3340; l) P. Brázdilová, M. Vrábel, R. Pohl, H. Pivonková, L. Havran, M. Hocek, M. Fojta, *Chem. Eur. J.* **2007**, *13*, 9527 – 9533; m) F. Seela, Y. M. Chen, *Chem. Commun.* **1996**, 2263 – 2264; n) G. A. Burley, J. Gierlich, M. R. Mofid, H. Nir, S. Tal, Y. Eichen, T. Carell, *J. Am. Chem. Soc.* **2006**, *128*, 1398; o) P. Kielkowski, H. Macíčková-Cahová, R. Pohl, M. Hocek, *Angew. Chem.* **2011**, *123*, 8886 – 8889; *Angew. Chem. Int. Ed.* **2011**, *50*, 8727 – 8730; p) J. Gierlich, K. Gutsmiedl, P. M. Gramlich, A. Schmidt, G. A. Burley, T. Carell, *Chem. Eur. J.* **2007**, *13*, 9486 – 9494; q) S. Obeid, M. Yulikov, G. Jeschke, A. Marx, *Angew. Chem.* **2008**, *120*, 6886 – 6890; *Angew. Chem. Int. Ed.* **2008**, *47*, 6782 – 6785.

[5] a) K. Betz, F. Streckenbach, A. Schnur, T. Exner, W. Welte, K. Diederichs, A. Marx, *Angew. Chem.* **2010**, *122*, 5308 – 5311; *Angew. Chem. Int. Ed.* **2010**, *49*, 5181 – 5184; b) S. Obeid, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21327 – 21331; c) K. Bergen, A. L. Steck, S. Strütt, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* **2012**, *134*, 11840 – 11843; d) S. Obeid, H. Bus-

skamp, W. Welte, K. Diederichs, A. Marx, *Chem. Commun.* **2012**, *48*, 8320 – 8322.

[6] a) A. C. Rodriguez, H. W. Park, C. Mao, L. S. Beese, *J. Mol. Biol.* **2000**, *299*, 447 – 462; b) H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai, *J. Mol. Biol.* **2001**, *306*, 469 – 477; c) T. Kuroita, H. Matsumura, N. Yokota, M. Kitabayashi, H. Hashimoto, T. Inoue, T. Imanaka, Y. Kai, *J. Mol. Biol.* **2005**, *351*, 291 – 298; d) K. P. Hopfner, A. Eichinger, R. A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3600 – 3605; e) S. J. Firbank, J. Wardle, P. Heslop, R. J. Lewis, B. A. Connolly, *J. Mol. Biol.* **2008**, *381*, 529 – 539; f) T. Killelea, S. Ghosh, S. S. Tan, P. Heslop, S. J. Firbank, E. T. Kool, B. A. Connolly, *Biochemistry* **2010**, *49*, 5772 – 5781; g) S. W. Kim, D.-U. Kim, J. K. Kim, L.-W. Kang, H.-S. Cho, *Int. J. Biol. Macromol.* **2008**, *42*, 356 – 361; h) Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, J. Kuriyan, *Structure* **1999**, *7*, 1189 – 1199; i) J. Gouge, C. Ralec, G. Henneke, M. Delarue, *J. Mol. Biol.* **2012**, *423*, 315 – 336.

[7] X. Jiang, M. Egli, *Current Protocols in Nucleic Acid Chemistry* **2011**, *46*, 7.15.1 – 7.15.8.

[8] P. J. Rothwell, G. Waksman, *Adv. Protein Chem.* **2005**, *71*, 401 – 440.

[9] R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, M. L. Kopka, *Science* **1982**, *216*, 475 – 485.

[10] a) S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *391*, 251 – 258; b) J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304 – 307; c) Y. Li, S. Korolev, G. Waksman, *EMBO J.* **1998**, *17*, 7514 – 7525.

[11] M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **2001**, *105*, 657 – 667.

[12] D. K. Braithwaite, J. Ito, *Nucleic Acids Res.* **1993**, *21*, 787 – 802.

[13] a) 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; b) Y. Li, Y. Kong, S. Korolev, G. Waksman, *Protein Sci.* **1998**, *7*, 1116 – 1123; c) Y. Li, V. Mitaxov, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9491 – 9496; d) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225 – 1233; e) S. Obeid, N. Blatter, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, A. Marx, *EMBO J.* **2010**, *29*, 1738 – 1747.

Received: March 25, 2013
Published online on June 3, 2013