

The Structure of an Archaeal B-Family DNA Polymerase in Complex with a Chemically Modified Nucleotide

Heike M. Kropp, Kay Diederichs, and Andreas Marx*

Dedicated to Professor Ronald T. Raines on the occasion of his 60th birthday

Abstract: Archaeal B-family DNA polymerases (DNA pols) are the driving force of cutting-edge biotechnological applications like next-generation sequencing. The acceptance of chemically modified nucleotides by DNA pols is key to these technologies. Until now, no structural data have been available for these DNA pols in complex with modified substrates, which could build the basis for understanding interactions between the enzyme and the chemically modified nucleotide and for the further development of next-generation nucleotides. For the first time, we crystallized an exonuclease-deficient variant of the wild-type B-family KOD DNA pol with a modified nucleotide in a closed, ternary complex. We also crystallized the A-family DNA pol KlenTaq with the same nucleotide. The reported structural data reveal how the protein and the DNA modulate two distinct conformations of the appended moiety in the A- and B-family DNA pols and how these influence the processing of the modified nucleotide. Overall, this study provides first insight into the interplay between B-family DNA pols and relevant modified substrates.

DNA polymerases (DNA pols) are the driving force of numerous biotechnological applications such as DNA and RNA sequencing,^[1] SELEX^[2] and microarray analysis.^[3] All of these applications rely on a DNA polymerase and its capability to accept chemically modified nucleotides. Surprisingly, even though DNA pols have evolved to utilize the canonical nucleotides, they have been shown to also process diverse modified analogues.^[4]

The acceptance of a modified nucleotide depends on the modification as well as the utilized DNA pol. Among the thermostable DNA pols like KlenTaq DNA pol (Klenow fragment of *Thermus aquaticus* DNA pol I, A-family) and KOD DNA pol (*Thermococcus kodakaraensis*, B-family), the B-family DNA pols often exhibit a broader substrate tolerance for C5-modified pyrimidines and 7-deaza-modified

purines,^[5] which are the modifications that are most important for present-day applications.

Recently, crystal structures of closed ternary complexes of KOD DNA pol^[6] and a mutant^[7] were published, giving first insights into the coordination of the triphosphate and the metal ions in the active site. However, structural data for DNA pols in complex with the above-mentioned nucleobase-modified nucleotides is so far only limited to the A-family KlenTaq DNA pol.^[8] However, enzymes from this enzyme family are less often suitable for the above-mentioned techniques. These studies indicate that hydrogen bonding^[8d,e] and electrostatic interactions^[8f] between the appended moiety and the enzyme support the incorporation of the modified substrate. Due to the lack of structural data for archaeal B-family DNA pols in complex with modified nucleotides, the underlying mechanisms that enhance the acceptance of unnatural substrates remain elusive for B-family DNA pols that are most important for numerous applications.

Herein, we addressed this issue and successfully crystallized an exonuclease deficient variant of wild-type KOD DNA pol in complex with a nucleotide bearing the modification that is employed in next-generation sequencing.^[1a,b] For direct comparison, we also crystallized KlenTaq DNA pol with the same modified nucleotide. Interestingly, the appended moiety adopts two distinct conformations for KOD and KlenTaq DNA pol. The obtained data illustrate the different interaction pattern the investigated appended moiety undergoes in A- and B-family DNA pols. These results reveal the mechanisms by which the modified nucleotide is processed by the archaeal B-family DNA pols that are crucial for many cutting edge applications.

Since unmodified dATP crystallized in complex with KOD DNA pol in our recent studies,^[6] we focused on modifying this nucleotide and synthesized the 7-deaza-modified adenosine derivative dA*TP (Figure 1, for details see Scheme S1 in the Supporting Information). This modification was chosen due to its importance in DNA sequencing.^[1a,b]

The obtained dA*TP was utilized in co-crystallization trials with KOD DNA pol. By following the approach as for KOD DNA pol in complex with unmodified dATP (KOD-dATP),^[6] we obtained a closed ternary structure of KOD DNA pol in complex with dA*TP (KOD-dA*TP). The structure was solved using difference Fourier techniques at a resolution of 2.0 Å (PDB ID: 5OMF).^[6] The structure shows a complex poised for catalysis, with a closed finger domain packing against the nascent dA*TP-dTMP base pair. The active site exhibits two metal ions (Mg²⁺ (A) and Mn²⁺). The

[*] M. Sc. H. M. Kropp, Prof. Dr. A. Marx
Department of Chemistry and
Konstanz Research School Chemical Biology, University of Konstanz
Universitätsstrasse 10, 7857 Konstanz (Germany)
E-mail: andreas.marx@uni-konstanz.de
Prof. Dr. K. Diederichs
Department of Biology and
Konstanz Research School Chemical Biology, University of Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)

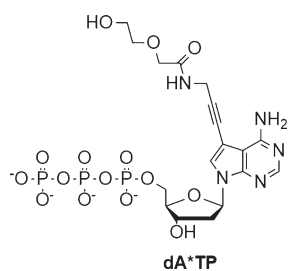


Figure 1. Chemical structure of 7-deaza-modified adenosine triphosphate (dA*TP).

Mn²⁺ in the B-site was determined by anomalous signal. The third metal ion (C), which is so far unique to DNA pols δ ,^[9] 9^oN, and KOD when bound to the natural substrate,^[6] is missing in the KOD-dA*TP structure (Figure 2D and Figure S2B). It is usually coordinated by the γ -phosphate, E580, D404, and three water molecules. One of these water molecules is coordinated by E578. Whereas all residues and water molecules are positioned as observed in KOD-dATP, E580 adopts an “open” conformation, which impairs a coordination of a third metal ion (Figure 2D and Figure S2B). A similar conformation for E580 was also observed for 9^oN DNA pol, which was previously crystallized with two metal

ions in the active site.^[6] So far, the role of the third metal ion in 9^oN and KOD DNA pol is unknown. In DNA pol δ , it is believed to modulate the catalytic efficiency by affecting the incorporation of the either right or wrong nucleotide.^[9] This may also apply for the archaeal B-family DNA pols. As observed for 9^oN DNA pol, the third metal ion seems to have a generally weaker binding affinity than the other two.^[6] However, overall the KOD-dA*TP structure exhibits high similarity compared to the recently published KOD-dATP structure (rmsd value of 0.264 for 697 Ca atoms; Figure S1A).^[6]

In the active site, dA*TP forms a canonical Watson–Crick base pair with the templating thymidine. The appended moiety points into the crevice between finger and palm domain (Figure 2C) and is structurally well resolved until the amide bond (Figure 2A). The ethylene glycol chain lacks electron density and was therefore modelled with zero occupancy. The carbonyl group of the appended moiety forms hydrogen bonds to K487 (2.9 Å), which in turn interacts with the α - (2.9 Å) and β -phosphate (3.1 Å) of dA*TP and is crucial for the nucleotidyl-transfer reaction.^[10] Additionally, R484 is positioned on the opposite site (5.0 Å) of the carbonyl group, leaving it to be sandwiched between two positively charged amino acid side chains (Figure 2B). These hydrogen bonding and electrostatic interactions seem to well stabilize

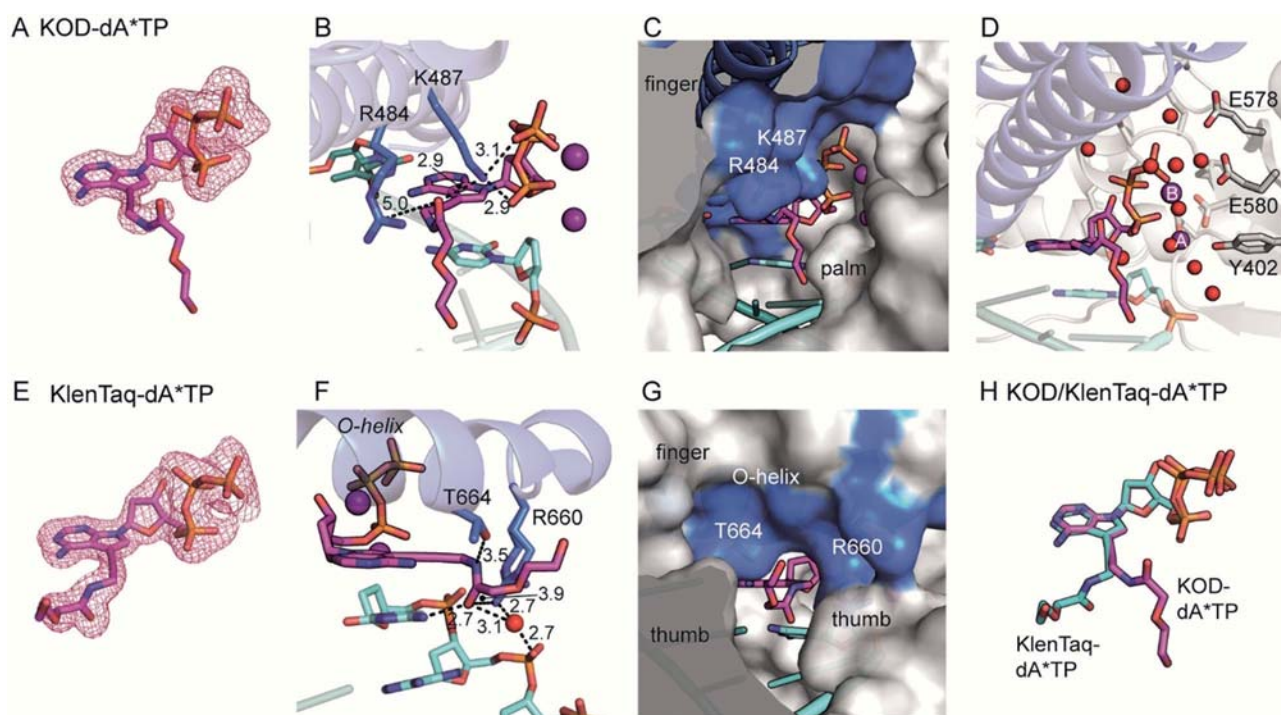


Figure 2. Structural data of KOD-dA*TP and KlenTaq-dA*TP. The protein is shown in grey with the finger domain of KOD DNA pol and the O-helix of KlenTaq DNA pol in blue. The dA*TP and the corresponding metal ions are shown in purple. Water molecules are shown as red spheres. Distances are shown as black dashes and are given in Å. A–D) KOD-dA*TP. A) Omit map of dA*TP at 3 σ . B) The appended moiety is sandwiched between the two positively charged amino acid side chains of K487 (2.9 Å) and R484 (5.0 Å). C) The 7-deaza moiety points into the crevice between the finger and palm domain. D) Active site with the two metal ions and E580 in the “open” conformation. E–G) KlenTaq-dA*TP. E) Omit map of dA*TP at 3 σ . F) The 7-deaza moiety is positioned underneath the O-helix, forming hydrogen bonds to the N6 of the 3' terminal dCMP (2.7 Å) via a water molecule (3.1 Å) to the primer backbone (2.7 Å) and to T664 (3.5 Å). G) The 7-deaza moiety is sandwiched between T664 and R660 of the O-helix and points into the crevice between finger and thumb domain. H) Superimposition of dA*TP of KOD-dA*TP (purple) and KlenTaq-dA*TP (blue).

the dA*TP in the active site. Since R484 is in close proximity to the appended moiety, it, and other amino acid side chains of the finger domain like I488, may be a target for mutations to improve the acceptance of modified nucleotides by KOD DNA pol.

In addition to E580, Y402 adopts a different conformation in KOD-dA*TP (Figure 2D and Figure S2B) compared to KOD-dATP. The side chain of Y402 moves towards E580 by 1.2 Å, which may be due to the different conformation of E580, opening space at that position. Through this movement, Y402 loses its hydrogen bonds to a water molecule coordinating metal ion A (3.7 Å) and a water molecule coordinating the primer phosphate backbone (3.7 Å). Upon binding of the third metal ion and the associated conformational change of E580, Y402 may move closer to the active site. Thereby, the stabilizing interactions of the primer strand and the active-site water molecules will be restored.

For comparison, we also crystallized KlenTaq DNA pol in complex with dA*TP (KlenTaq-dA*TP). We set up crystallization trials employing the same primer/template sequence as well as the same amount of MgCl₂ and MnCl₂ as used for KOD DNA pol. To compare KlenTaq-dA*TP to a natural structure crystallized using the same primer/template and metal ion composition, we also crystallized KlenTaq DNA pol in complex with dATP (KlenTaq-dATP). We solved the KlenTaq-dA*TP and KlenTaq-dATP structures using difference Fourier techniques at resolutions of 2.0 Å (PDB ID: 6FBI^[8f]). Overall, KlenTaq-dATP and KlenTaq-dA*TP adopt a very similar conformation with an rmsd value of 0.108 for 468 Ca atoms (Figure S1B). Both structures show complexes poised for catalysis, each with two metal ions coordinated in the active site. In both cases, the metal ions are Mg²⁺ and Mn²⁺ for the A and B metal ion, respectively, as observed for KOD-dA*TP. The O-helix is fully closed and packs against the nascent dATP-dTMP and dA*TP-dTMP pair, with K663 forming hydrogen bonds to the α- [2.7 Å (dATP), 3.2 Å (dA*TP)] and β- [2.8 Å (dATP), 2.8 Å (dA*TP)] phosphate (Figure S3).

In KlenTaq-dA*TP, the appended moiety is situated between the O-helix residues R660 and T664 (3.5 Å) and points towards the thumb domain (Figure 2G). R660, as observed for R484 in KOD-dA*TP, is in close proximity (3.9 Å) to the carbonyl group of the modified nucleotide and was modelled in two conformations (Figure 2F, Figure S3). Both conformations of R660 are within hydrogen-bonding distance to the primer backbone (2.9 Å) (Figure S3). Compared to KOD-dA*TP, the appended moiety in KlenTaq-dA*TP is more rigid, since electron density for the ethylene glycol chain is observed (Figure 2E). The carbonyl group, compared to KOD-dA*TP, is turned by 180° (Figure 2H) and points towards the N6 of the 3'-terminal cytidine, hydrogen bonding with the exocyclic amine (2.7 Å; Figure 2F). Since the DNA duplex close to the active site of A-family DNA polymerases adopts the A-form (Table S2), the duplex has a narrower major groove and a closer nucleobase distance at the 3'-primer terminus in KlenTaq DNA pol compared to KOD DNA pol. Hydrogen-bonding of the carbonyl group to the N6 of the cytidine may thereby be facilitated. Additionally, the carbonyl group forms a hydrogen bond via a water

molecule (3.1 Å) to the primer backbone (2.7 Å; Figure 2F). This additional interaction with the primer backbone is not observed in KlenTaq-dATP.

In order to evaluate the efficiency of dA*TP processing in comparison to its natural counterpart by KOD and KlenTaq DNA pol, we performed single-nucleotide-incorporation experiments in which the modified dA*TP directly competed with natural dATP (Figure 3), as described before.^[8d,e] We

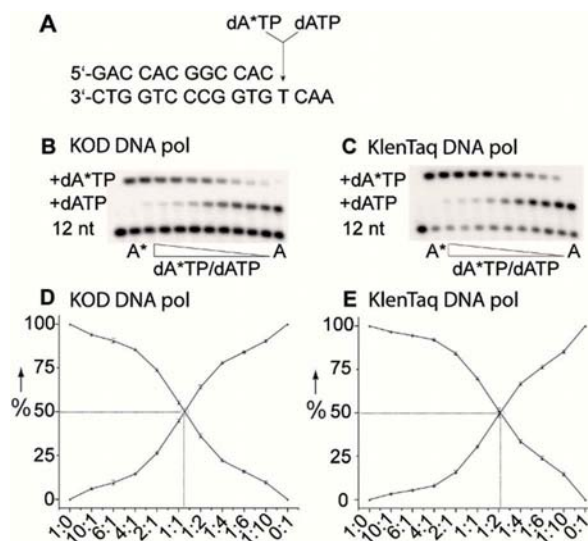


Figure 3. Competitive primer-extension experiments. A) Sequence of the primer/template complex. B) PAGE analysis of the competition experiment using KOD DNA pol. C) PAGE analysis of the competition experiment using KlenTaq DNA pol. D) Graphical readout of the PAGE analysis using KOD DNA pol. The point of 50% incorporation is indicated with a dashed line. E) Same as for (D) but using KlenTaq DNA pol. Data from triplicates.

used the 5'-³²P-labelled 12-nt primer and 16-nt template with the same sequence as that used for crystallization. The reactions were analyzed and the ratio of dATP versus dA*TP was determined by 16% denaturing polyacrylamide gel electrophoresis (PAGE) and autoradiography by exploiting the significantly different migration of the primer elongated by either processing of dA*TP or dATP due to the additional modification of dA*TP. For KOD DNA pol, the appended moiety of dA*TP has only minor effects on the enzyme, since KOD DNA pol incorporates the modified nucleotide with a 1.2 fold higher efficiency compared to dATP (Figure 3). Interestingly, KlenTaq DNA pol processes dA*TP with a 2-fold higher efficiency compared to dATP. Similar superior processing was reported for other dATP analogues as well.^[8e,11] This may be due to the interactions formed by the carbonyl group of the modified nucleotide and the primer strand, which might lead to an additional stabilization of the ternary complex. Since the hydrogen bond of the carbonyl to the N6 of the 3'-terminal cytosine is sequence specific, we wondered whether the enhanced incorporation efficiency by KlenTaq DNA pol when using dA*TP persists if a thymidine instead of a cytidine is located at the 3' primer terminus, thereby preventing hydrogen-bond formation between the

carbonyl group and the 3' terminal primer nucleobase. By employing the 3'-terminal dTMP primer in the single-nucleotide-incorporation competition experiments, we indeed observed a drop of the incorporation efficiency to 1.5-fold versus the 3'-terminal dCMP primer (Figure S4).

Comparing KOD and KlenTaq DNA pol now leads us to two main differences between the enzymes that result in the two distinct conformations of dA*TP observed here: the channel volume that is accessible within the enzyme to accommodate modifications, and the DNA conformation close to the active site. KOD DNA pol has a larger channel volume within its ternary closed complex compared to KlenTaq DNA pol^[6] as well as a larger crevice between its finger and thumb/palm domain. Therefore, KOD DNA pol has more space to accommodate modifications, which results in a highly flexible ethylene glycol chain of dA*TP in KOD-dA*TP. Additionally, KOD DNA pol exhibits B-form DNA throughout, which results in larger internucleobase distances and a wider major groove compared to the A-form DNA of the first three primer nucleotides in KlenTaq DNA pol (Table S2). The B-form DNA in KOD DNA pol offers more space to accommodate major-groove modifications such as those at C5 of pyrimidines and 7-deaza of purines. Along with the protein environment, the B-form seems to prevent sequence-specific interactions between modifications attached at incoming dNTPs and downstream nucleotides compared to KlenTaq-dA*TP. Even though the more compact active complex and the A-form DNA in KlenTaq DNA pol seem to promote the processing of a single dA*TP, these properties may reduce the incorporation of multiple modified nucleotides in a consecutive manner. Here, a DNA pol with more space and fewer interactions, like KOD DNA pol, would be advantageous.

In summary, we have gained insight into the processing of a modified nucleotide by KOD DNA pol, a member of the DNA pol family B, which is crucial for many current applications. The obtained crystal structures allow a direct comparison of interactions formed by modified nucleotides in a B- and A-family DNA pol. They further reveal how the protein and DNA environment modulate the conformation of the appended moiety in each family, leading to two distinct conformations. This study broadens our so-far limited knowledge about the interplay between biotechnologically important archaeal B-family DNA pols and relevant chemically modified substrates and builds the basis to understand their capability to accept a wide range of modified nucleotides. Future studies aim at elucidating how nucleotides with more than one modification are processed within the confines of DNA pols.

Acknowledgements

We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG), and the Swiss Light Source of the Paul Scherrer Institute for access to the beamline PXI—X06SA and great support.

Conflict of interest

The authors declare no conflict of interest.

Keywords: B-family polymerases · DNA polymerase · modified nucleotides · next-generation sequencing · X-ray crystallography

- [1] a) D. R. Bentley, et al., *Nature* **2008**, *456*, 53–59; b) M. L. Metzker, *Nat. Rev. Genet.* **2010**, *11*, 31–46; c) C. W. Fuller, S. Kumar, M. Porel, M. Chien, A. Bibillo, P. B. Stranges, M. Dorwart, C. Tao, Z. Li, W. Guo, S. Shi, D. Korenblum, A. Trans, A. Aguirre, E. Liu, E. T. Harada, J. Pollard, A. Bhat, C. Cech, A. Yang, C. Arnold, M. Palla, J. Hovis, R. Chen, I. Morozova, S. Kalachikov, J. J. Russo, J. J. Kasianowicz, R. Davis, S. Roever, G. M. Church, J. Ju, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5233–5238; d) S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333–351; e) J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, M. Sano Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9145–9150.
- [2] a) K. Sefah, Z. Yang, K. M. Bradley, S. Hoshika, E. Jiménez, L. Zhang, G. Zhu, S. Shanker, F. Yu, D. Turek, W. Tan, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1449–1454; b) G. Mayer, *Angew. Chem. Int. Ed.* **2009**, *48*, 2672–2689; *Angew. Chem.* **2009**, *121*, 2710–2727.
- [3] a) P. S. Mischel, T. F. Cloughesy, S. F. Nelson, *Nat. Rev. Neurosci.* **2004**, *5*, 782–792; b) C. C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A.-L. Borresen-Dale, P. O. Brown, D. Botstein, *Nature* **2000**, *406*, 747–752; c) J. R. Pollack, C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein, P. O. Brown, *Nat. Genet.* **1999**, *23*, 41–46.
- [4] A. Hottin, A. Marx, *Acc. Chem. Res.* **2016**, *49*, 418–427.
- [5] a) M. Mehedi Masud, A. Ozaki-Nakamura, M. Kuwahara, H. Ozaki, H. Sawai, *ChemBioChem* **2003**, *4*, 584–588; b) M. Kuwahara, J.-i. Nagashima, M. Hasegawa, T. Tamura, R. Kitagata, K. Hanawa, S.-i. Hososhima, T. Kasamatsu, H. Ozaki, H. Sawai, *Nucleic Acids Res.* **2006**, *34*, 5383–5394; c) H. A. Held, S. A. Benner, *Nucleic Acids Res.* **2002**, *30*, 3857–3869; d) S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.* **2005**, *127*, 15071–15082; e) M. Mačková, S. Boháčová, P. Perlíková, L. Poštová Slavětínská, M. Hocek, *ChemBioChem* **2015**, *16*, 2225–2236; f) 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. Int. Ed.* **2005**, *44*, 2144–2148; *Angew. Chem.* **2005**, *117*, 2182–2186; g) T. Tasara, B. Angerer, M. Diamond, H. Winter, S. Dörhöfer, U. Hübscher, M. Amacker, *Nucleic Acids Res.* **2003**, *31*, 2636–2646.
- [6] H. M. Kropp, K. Betz, J. Wirth, K. Diederichs, A. Marx, *PLoS One* **2017**, *12*, e0188005.
- [7] N. Chim, C. Shi, S. P. Sau, A. Nikoomezar, J. C. Chaput, *Nat. Commun.* **2017**, *8*, 1810.
- [8] a) S. Obeid, H. Bußkamp, W. Welte, K. Diederichs, A. Marx, *Chem. Commun.* **2012**, *48*, 8320–8322; b) S. Obeid, H. Bußkamp, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* **2013**, *135*, 15667–15669; c) S. Obeid, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21327–21331; d) K. Bergen, A.-L. Steck, S. Strütt, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* **2012**, *134*,

- 11840–11843; e) A. Hottin, K. Betz, K. Diederichs, A. Marx, *Chem. Eur. J.* **2017**, *23*, 2109–2118; f) H. M. Kropp, S. L. Dürr, C. Peter, K. Diederichs, A. Marx, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9992–9997.
- [9] M. K. Swan, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, *Nat. Struct. Mol. Biol.* **2009**, *16*, 979–986.
- [10] C. Castro, E. D. Smidansky, J. J. Arnold, K. R. Maksimchuk, I. Moustafa, A. Uchida, M. Gotte, W. Konigsberg, C. E. Cameron, *Nat. Struct. Mol. Biol.* **2009**, *16*, 212–218.
- [11] P. Kielkowski, J. Fanfrlík, M. Hocek, *Angew. Chem. Int. Ed.* **2014**, *53*, 7552–7555; *Angew. Chem.* **2014**, *126*, 7682–7685.