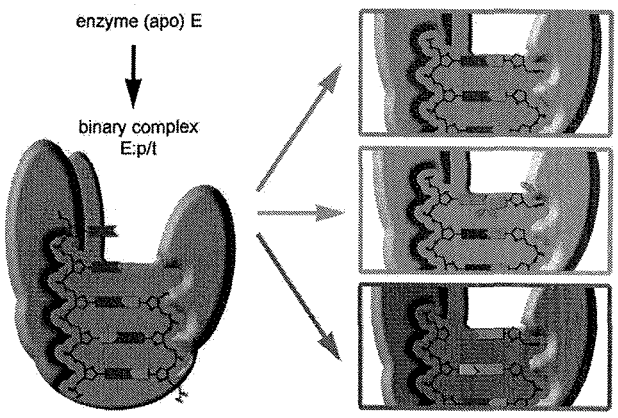
Zuerst ersch. in : ChemBioChem ; 13 (2012), 5. - pp. 635-639
DOI : 10.1002/cbic.201290014

KlenTaq DNA Polymerase Adopts Unique Recognition States when Encountering Matched, Mismatched, and Abasic Template Sites: An NMR Study

Bastian Holzberger, M. Gabriele Pszolla, Andreas Marx, and Heiko M. Möller*[a]

The overall fidelity of DNA synthesis, and thus the accuracy of genome replication in general, is governed by the fundamental intrinsic fidelity of DNA polymerases.[1] These enzymes catalyze the template-directed addition of nucleotides to the 3′-end of a DNA primer strand with high fidelity.[2] In the process, DNA polymerases contribute actively to nucleotide selection by introducing only correct nucleotides into the growing primer strand with high efficiencies. A currently used model for DNA polymerase selectivity is based on geometric complementarity of the nascent nucleobase pair. Correctly matching nucleotides are readily accepted and arranged for phosphodiester bond formation whereas incorrect 2′-deoxynucleoside triphosphates (dNTPs) are processed with lower efficiency. The discrimination in the early steps of nucleotide incorporation is thus governed by a series of conformational alterations leading from a conformation with an open, solvent-exposed active site to a closed conformation.[3] DNA polymerases are also characterized by their varying abilities to tolerate DNA template modifications and lesions.[4] So far, X-ray crystallography and fluorescence-based studies have predominantly been used to gain insight into these processes.[3] Although these methods have been very valuable for our current understanding of DNA polymerase mechanisms, there are still some drawbacks. Fluorescence-based studies, for instance, require the incorporation of bulky fluorescent probes that are susceptible to hydrophobic or stacking interactions, potentially introducing non-natural effects and masking natural interactions. Additionally, with each pair of fluorescent probes only a restricted subset of dynamic properties out of a highly complex overall enzymatic process can be studied. Crystallization trials, on the other hand, are performed under non-natural conditions and crystal packing effects can contribute to the nature of the 3D structure. Additionally, structural heterogeneity might prevent crystallization of certain states completely. In conclusion, for our understanding of DNA polymerase mechanisms and to minimize the risk of misinterpretation it is absolutely crucial to combine a whole set of methods. Interestingly, NMR has been exploited only sparsely for studying nucleotide polymerase mechanisms.[5] However, residue-specific isotopic labeling retains an enzyme in its native, unperturbed state with multiple probes for investigating global and local dynamics simultaneously. Furthermore, solution NMR can be performed under near physiological conditions to study dynamic states in equilibrium.

Here we have used NMR to investigate the thermophilic KlenTaq DNA polymerase, a member of DNA polymerase family A frequently used as a model system for mechanistic studies.[4c,6] By introducing carbon-13 labels at the ε-methyl groups of KlenTaq's methionine (Met) residues we obtained spin probes distributed all over the protein, which allowed us to monitor conformational changes and substrate binding of the DNA polymerase during catalysis. Through the use of 2′,3′-dideoxynucleoside triphosphates (ddNTPs), primers lacking their 3′-terminal hydroxy groups were generated. The DNA polymerase is thus caught in a stationary—that is, non-propagating—ternary complex comprising a further ddNTP bound to the active site of the enzyme opposite to the next template site (Scheme 1).[6f,h] When comparing matched with mismatched complexes we found unique conformations for incorrect base pair formation that are clearly distinct from the match case but rather are related to the binary, open complex. Similar results were obtained when we investigated nucleotide recognition opposite an abasic site template. In all investigated noncanonical cases, the transitions from an open complex to a closed, productive complex appears to be hampered, explaining the significantly reduced catalytic efficiencies for formation of noncanonical nucleobase pairs.



Scheme 1. Addition of a DNA primer/template complex (p/t) to the apo form of the DNA polymerase E leads to formation of the binary complex E:p/t (red). Subsequent addition of 2′,3′-dideoxynucleoside triphosphates (ddNTPs) leads to ternary state formation. Depending on the template sequence, different ternary states are accessible. These include a match case with ddCTP opposite G (green), mismatch cases with, for example, ddCTP opposite A (orange), and abasic site states with, for example, an incoming ddATP opposite to the abasic site analogue F (blue).

[a] *B. Holzberger, M. G. Pszolla, Prof. Dr. A. Marx, Dr. H. M. Möller*
Department of Chemistry and Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)
E-mail: heiko.moeller@uni-konstanz.de

The KlenTaq DNA polymerase is very large by NMR standards (63 kDa) and so conventional $^{1}H,^{15}N$ correlations suffer from severe line broadening and thus strongly decreased sensitivity when recorded in a protonated background. Perdeuteration of all side chains is a prerequisite for obtaining high-quality amide correlations by exploiting the TROSY effect.[7] Furthermore, the sequential assignment of the KlenTaq DNA polymerase, which is indispensable for gaining insights into local dynamics and conformational changes, would be a very challenging task. We therefore exploited the much narrower linewidths of Met methyl groups, the signals of which can easily be recorded even in the case of a fully protonated enzyme, demonstrating a much faster, less costly, and virtually label-free method[8] to provide insight into DNA polymerase dynamics and substrate selection. In detail, we used [methyl-^{13}C]Met to introduce carbon-13 spin probes and to monitor primer/template complex and nucleotide binding at 13 sites simultaneously. To this end, we collected $^{1}H,^{13}C$ HSQC spectra of the enzyme when encountering matched, mismatched, and abasic template sites (Scheme 1). We worked with comparatively low protein concentrations (45–70 μM) at 60 °C, to be as close as possible to native conditions of the thermophilic bacterium *Thermus aquaticus*. As a welcome side effect, elevated temperatures are highly desirable for NMR spectroscopy of large proteins because shorter correlation times, both of the whole enzyme and of its Met side chains, further increase sensitivity. At room temperature a number of Met methyl crosspeaks were broadened beyond detection due to excessively fast dipolar relaxation (or intermediate exchange), similarly to what has been observed in other cases.[8d] Perdeuteration and exploitation of the methyl-TROSY[7] effect will therefore be mandatory in most cases for obtaining methyl correlations of similar sized proteins that are not stable at elevated temperature.

The [methyl-^{13}C]Met-labeled KlenTaq DNA polymerase was obtained by using a Met auxotrophic *E. coli* strain and replacing natural Met with [methyl-^{13}C]Met in the growth medium.[9] The enzyme was expressed and purified by standard procedures (for experimental details see the Supporting Information). It is interesting to note that HMQC experiments turned out to be slightly less sensitive than HSQCs under all conditions tested (data not shown). This is not completely unexpected because the methyl-TROSY effect operational in HMQC depends strongly on the presence of external relaxation sources. Apparently, the dramatic sensitivity gain that can be achieved through the methyl-TROSY effect in highly deuterated proteins through cancellation of intra-methyl relaxation pathways is completely destroyed by external relaxation in a fully protonated KlenTaq DNA polymerase. Furthermore, fast relaxation and extensive spin diffusion in conjunction with the low concentrations of KlenTaq DNA polymerase in our experiments precluded the detection of nuclear Overhauser enhancements (NOEs) originating from the Met methyl groups. Nevertheless, ^{13}C-labeled KlenTaq DNA polymerase displayed distinct Met ε-methyl crosspeaks in the $^{1}H,^{13}C$ HSQC between $\delta(^{13}C) = 10.5$–18.5 and $\delta(^{1}H) = 0.9$–2.6 ppm (Figure S1 in the Supporting Information).

To assign the individual peaks, we created all 13 single mutants in which one Met was mutated to alanine (Ala) and generated an additional KlenTaq construct possessing a protease cleavage site to remove the N-terminal Met (for experimental details see the Supporting Information). Comparison of the $^{1}H,^{13}C$ HSQC spectra of these Met→Ala single-site mutants with the HSQC of the wild-type enzyme allowed us to assign all signals to internal Met residues and the conformationally heterogeneous N-terminal Met and its sulfoxide. The crosspeaks of the N-terminal Met residue showing conformational heterogeneity and of its sulfoxide form were established by comparison of the wild type with the N-terminally truncated form. Superpositions of spectra of the wild type and of the mutants used for assignment are shown in Figure S2. The vast majority of assignments were obvious because one crosspeak was clearly missing in the spectrum of each single-site mutant whereas all other crosspeaks remained unaffected. In some cases the Met→Ala mutation appears to induce structural changes that lead to slightly shifted signals of other Met methyl crosspeaks (e.g., M444A and M765A). We repeated this assignment procedure for all relevant complexes (Figures S3–S5).

Upon addition of a DNA primer/template complex leading to the formation of the binary state (Scheme 1), the majority of Met ε-methyl groups are significantly affected (Figure 1 A). These chemical shift perturbations are brought about by local structural changes leading to altered conformations of Met side chains, by interactions with the chemical shift anisotropy of the nucleobases, and by electrostatic interactions with the phosphate backbone. It is important to note that chemical shift changes are observed for many residues at distances of more than 10 Å from the DNA binding site. This is indicative of a network of long-range conformational changes, as has also been observed in a poliovirus RNA polymerase.[5d] The most pronounced chemical shift changes occur at residues M317, M374, M444, M747, M751, M761, M765, and M807. This information was correlated to the known 3D structures of KlenTaq DNA polymerase.[6f–h] Most of these residues are located in the rigid palm domain of the enzyme that is involved in DNA p/t binding, but M747 and M751 are located next to the incoming DNA in the fingers domain. In contrast, residues M646, M658, and M673, located either at the tip of the fingers domain or directly at the hinge of the O helix (M673), as well as M775 and the N-terminal Met resonances, remain largely unperturbed upon DNA binding. Subsequent addition of ddCTP leads to primer elongation through the incorporation of one equivalent ddCMP followed by formation of the ternary G·ddCTP complex with ddCTP opposite G in the template strand (Scheme 1). Indeed, residues that experienced strong chemical shift perturbation upon DNA binding are now only slightly affected (Figure 1 B). However, as well as M747, M751, and M761, located in proximity to the active site, the Met residues in the flexible fingers domain—M646, M658, and M673—show especially significant shifts upon ddCTP binding. M673 is located at the hinge of the O helix and next to the catalytically important tyrosine Y671; M646 and M658 are positioned in the N and O helixes of the fingers domain, respectively. We attribute the observed

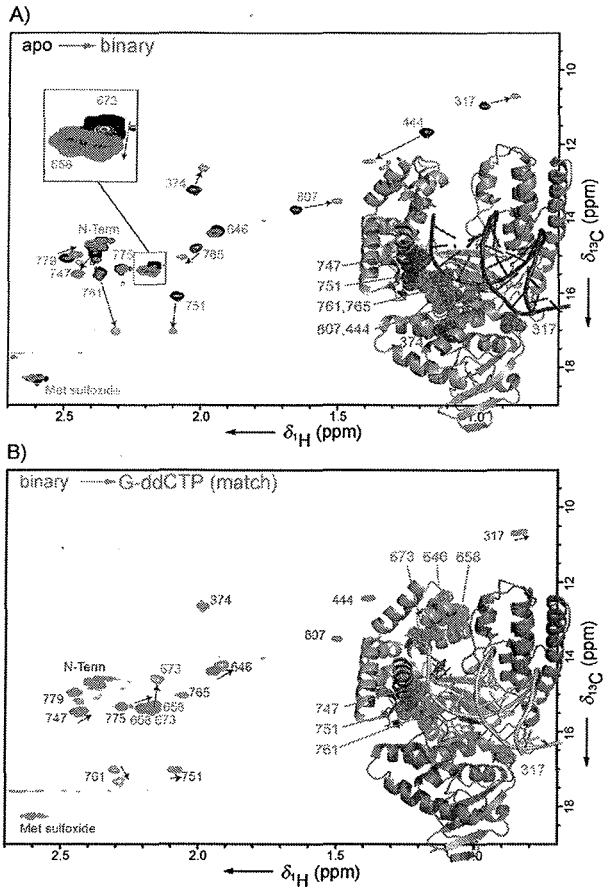


Figure 1. Substrate-dependent chemical shift perturbations of Met ε-methyl groups in $^{1}H,^{13}C$ HSQC spectra of KlenTaq DNA polymerase. The residue numbering of full-length Taq DNA polymerase has been used. The fingers domain, which undergoes large conformational changes upon nucleotide binding, is shown in gray. A) Chemical shift changes upon formation of the binary E:p/t complex (red) by addition of DNA p/t complex (2 equiv) to the apo form (black). Met residues are assigned and highlighted in red (PDB ID: 4KTQ). B) Chemical shift changes upon formation of the ternary G·ddCTP complex (green) by addition of ddCTP (10 equiv). Met residues are light green (PDB ID: 3KTQ).

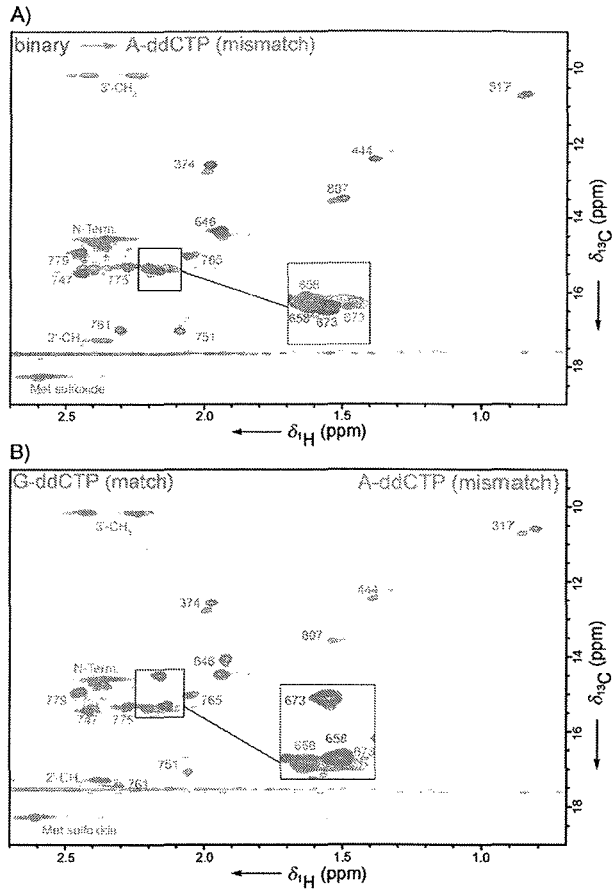


Figure 2. Comparison of match and mismatch ternary states. A) Chemical shift changes upon formation of the ternary complex A·ddCTP (orange) after addition of ddCTP (100 equiv). B) Comparison of the ternary complexes A·ddCTP (orange) and G·ddCTP (green) after addition of ddCTP (100 equiv). Because we used a large excess of ddCTP to ensure complete complex formation, the 2′- and 3′-CH_2 crosspeaks of ddCTP are visible.

chemical shift perturbations to structural differences induced by the large conformational changes of the fingers domain upon ddCTP binding.[6f] Notably, the observed shift alterations are quite large in comparison with related perturbations reported in the literature.[5d]

Also of note is the fact that structural data for A-family DNA polymerases reporting on the recognition and processing of mismatched nucleotides in the active site have up to now been rare,[10] and in the case of KlenTaq DNA polymerase still lacking. To gain further insight into these processes we performed the same experiments depicted above with a DNA template containing instead of 3′-GG the nucleotides 3′-GA, to induce (after incorporation of ddCMP opposite G) mismatch formation with an incorrect ddCTP opposite to the templating A. The resulting spectrum of A·ddCTP (Figure 2A) differs significantly from that of the match case G·ddCTP (Figure 2B). In the mismatch case the crosspeak positions of M658 and M673 are, in part, reminiscent of the binary state. Nonetheless, the mismatch situation can be clearly distinguished from the binary state and the match case. In addition to signal shifting and severe broadening of M751 and M761, located in the vicinity of the active site, the Met residues of the fingers domain—M646, M658, and M673—in particular show different behavior from the match case. M673's resonance broadens in the proton dimension instead of showing the sizeable upfield shift of the carbon resonance upon correct nucleotide binding whereas M751 and M761 shift in completely different directions compared to the match case. We want to emphasize that match and mismatch scenarios are clearly distinguishable at all nucleotide concentrations tested here; see NMR titration (Figures S7 and S8) and the comparison at 100 equivalents ddNTP (Figure S11). Exchange signal broadening thus indicates enhanced local dynamics or heterogeneity. It is noteworthy that the only difference between match and mismatch formation is located in the DNA template strand. The addition of the same nucleotide thus apparently resulted in different conformations although the spectra of the initial binary states were similar for

all investigated scenarios (Figure S6). This demonstrates the supreme capability of solution NMR clearly to distinguish different chemical environments arising from varied substrate combinations. Along with the fact that the remaining mismatch cases with ddCTP opposite C and T show signal patterns quite similar to that of A·ddCTP (Figure S7), this is strong evidence that structurally distinct ternary complexes are indeed formed.

Interestingly, the characteristic chemical shift perturbations upon addition of ddCTP appear at different nucleotide concentrations. Whereas 4 equivalents or less are sufficient in the match case, in the mismatch scenarios 20–40 equivalents of ddCTP are required to induce significant signal perturbations (Figure S7 and S8). Although there are a number of perturbations induced by correct and incorrect nucleotide binding, the signal of M646 was used exclusively for comparison of extents of complex formation because it remains visible in all scenarios. In accordance with kinetic measurements that report on different affinities towards nucleotides prior to the formation of the closed state,[11] this indicates interactions between the incoming nucleotide and the templating base in a complex related rather to an open complex. The concentration dependence seen here by NMR—that is, approximately 200 μM of matching ddNTP versus 1–2 mM of mismatching ddNTP leading to significant formation of the ternary complexes—is qualitatively compatible with affinities determined by fluorescence studies (28 vs 200 μM for the initial binding reactions of matching and mismatching dNTPs, respectively).[11] It should be mentioned that match and mismatch scenarios remain clearly distinguishable even at the highest nucleotide concentrations (4–7 mM) and that the match case does not display any further perturbations that could be characteristics of the mismatched situation. This suggests that there are no additional interactions with the DNA polymerase at high ddNTP concentrations. This is in good agreement with crystallographic studies carried out at similarly high nucleotide concentrations that do not report additional nucleotide binding sites at all (e.g., see refs. [4c], [10], and [6f]).

With use of a DNA template displaying the abasic site analogue F (Figure S9) the addition of ddATP induced the formation of a ternary state with ddATP bound opposite F. Aside from severe signal shifting and broadening of M444, M673, M751, and M761, the resonances of M646 and M747 also show up at different crosspeak positions in relation to the binary state (Figure 3). These alterations are also present after addition of ddTTP to a binary complex displaying a 3′-AF template (Figure S10). However, the shift alterations that characterize the closure of the fingers domain in the match case are also not observed in the presence of templating abasic sites here. This is in agreement with a crystal structure displaying ddATP opposite to F,[4c] in which the O helix is positioned in a rather open conformation. Although the spectra of mismatched and abasic template sites are closely related, it is possible to discriminate between them clearly through the unique crosspeak positions of, for example, M646, M673, or M747 (Figure S11).

In summary, we have investigated DNA binding, nucleotide recognition, and the conformational changes associated with active-site closure by tracking the crosspeaks of [methyl-

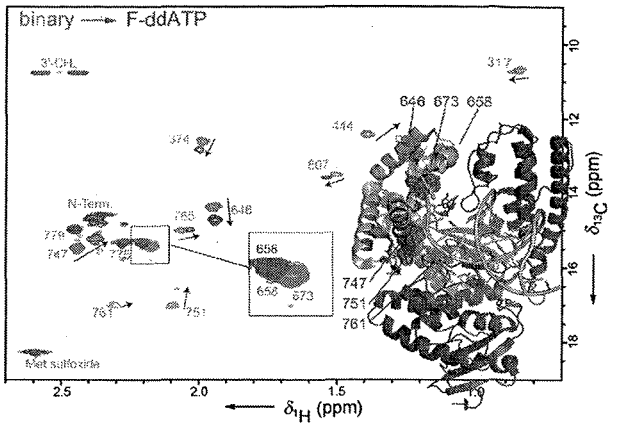


Figure 3. Chemical shift changes upon formation of the ternary complex F·ddATP (blue) after addition of ddATP (100 equiv). Met residues are pink (PDB ID: 3LWL).

^{13}C]Met ε-methyl groups. In this approach, chemical shift differences encountered in mismatched or abasic template sites relative to binary and matched states indicate that KlenTaq DNA polymerase cycles through distinct paths in canonical and in noncanonical cases. So far, structural data for A-family DNA polymerases processing mismatched nucleotides have only been reported for BF DNA polymerase,[10] and in the case of KlenTaq DNA polymerase are still unavailable. We were thus able to support a model based on unique recognition states[11,12] by use of label-free NMR techniques. However, our studies also revealed that differences in local dynamics or conformational heterogeneity might contribute to selectivity of DNA polymerases by reducing the efficiency of incorporation and promoting substrate release in cases of incorrect base pairing or DNA lesions such as abasic sites.

Experimental Section

For NMR experiments the protein storage buffer was exchanged into NMR buffer [Tris·HCl (pH 9.2, 50 mM), $(NH_4)_2SO_4$ (16 mM), $MgCl_2$ (2.5 mM) in D_2O]. NMR spectra were measured at protein concentrations between 18 μM and 70 μM at 333 K in MATCH NMR tubes (3 mm). All NMR experiments were performed with a Bruker Avance III 600 MHz spectrometer equipped with an inverse H/C/N-TCI-cryoprobe (5 mm) with actively shielded *z*-gradient. ^{1}H,^{13}C HSQC spectra were generally acquired with 574×256 complex points in the direct and indirect dimension, respectively; 8–80 scans with a recovery delay of 1.5 s were accumulated for each t1-increment. All NMR spectra were processed and analyzed with Bruker's TopSpin software (v2.1 and v3.0). Overlapped peaks were deconvoluted with assumption of a Lorentzian line shape. DNA primer [5′-d(GAC CAC GGC GC)-3′] and templates [5′-d(AAA YXG CGC CGT GGT C)-3′ with YX = GG, AG, TG, CG, FA, FT] were synthesized by Metabion or Purimex (HPLC-purified). Annealing was performed by heating for 5 min to 95 °C and cooling to 23 °C. ddNTPs were from Jena Bioscience. ddNTPs were added from concentrated stock solutions to minimize dilution effects. Up to 20 equiv, 10 mM stock solutions were used. For subsequent titration points ddNTPs were added from 100 mM stock solutions.

Acknowledgements

Financial Support through the Konstanz Research School Chemical Biology and the DFG is gratefully acknowledged. B.H. thanks the Carl Zeiss Foundation for a Ph.D. scholarship. We thank Prof. Dr. Ulrich Steiner for helpful discussion.

Keywords: abasic sites · DNA polymerases · DNA replication · mismatches · NMR spectroscopy

[1] a) U. Hübscher, G. Maga, S. Spadari, *Annu. Rev. Biochem.* **2002**, *71*, 133; b) C. M. Joyce, T. A. Steitz, *Annu. Rev. Biochem.* **1994**, *63*, 777; c) L. A. Loeb, R. J. Monnat, Jr., *Nat. Rev. Genet.* **2008**, *9*, 594; d) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497.

[2] a) H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477; b) M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493; c) E. T. Kool, J. C. Morales, K. M. Guckian, *Angew. Chem.* **2000**, *112*, 1046; *Angew. Chem. Int. Ed.* **2000**, *39*, 990; d) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191; e) A. J. Berdis, *Chem. Rev.* **2009**, *109*, 2862.

[3] a) C. M. Joyce, S. J. Benkovic, *Biochemistry* **2004**, *43*, 14317; b) C. M. Joyce, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 1032; c) K. A. Johnson, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 1041.

[4] a) S. Prakash, R. E. Johnson, L. Prakash, *Annu. Rev. Biochem.* **2005**, *74*, 317; b) W. Yang, R. Woodgate, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15591; c) S. Obeid, N. Blatter, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, A. Marx, *EMBO J.* **2010**, *29*, 1738.

[5] a) Z. Ren, H. Wang, R. Ghose, *Nucleic Acids Res.* **2010**, *38*, 5105; b) B. Bose-Basu, E. F. DeRose, T. W. Kirby, G. A. Mueller, W. A. Beard, S. H. Wilson, R. E. London, *Biochemistry* **2004**, *43*, 8911; c) T. W. Kirby, E. F. DeRose, W. A. Beard, S. H. Wilson, R. E. London, *Biochemistry* **2005**, *44*, 15230; d) X. Yang, J. L. Welch, J. J. Arnold, D. D. Boehr, *Biochemistry* **2010**, *49*, 9361; e) B. Holzberger, M. Rubini, H. M. Möller, A. Marx, *Angew. Chem.* **2010**, *122*, 1346; *Angew. Chem. Int. Ed.* **2010**, *49*, 1324.

[6] a) P. J. Rothwell, V. Mitaksov, G. Waksman, *Mol. Cell* **2005**, *19*, 345; b) P. J. Rothwell, G. Waksman, *J. Biol. Chem.* **2007**, *282*, 28884; c) W. J. Allen, P. J. Rothwell, G. Waksman, *Protein Sci.* **2008**, *17*, 401; d) K. Datta, V. J. LiCata, *Nucleic Acids Res.* **2003**, *31*, 5590; e) S. Korolev, M. Nayal, W. M. Barnes, E. Di Cera, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9264; f) Y. Li, S. Korolev, G. Waksman, *EMBO J.* **1998**, *17*, 7514; g) Y. Li, Y. Kong, S. Korolev, G. Waksman, *Protein Sci.* **1998**, *7*, 1116; h) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225.

[7] a) K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366; b) V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, *J. Am. Chem. Soc.* **2003**, *125*, 10420.

[8] For example, see a) W. C. Jones, T. M. Rothgeb, F. R. Gurd, *J. Biol. Chem.* **1976**, *251*, 7452; b) P. R. Rosevear, *Biochemistry* **1988**, *27*, 7931; c) M. J. DellaVecchia, W. K. Merritt, Y. Peng, T. W. Kirby, E. F. DeRose, G. A. Mueller, B. Van Houten, R. E. London, *J. Mol. Biol.* **2007**, *373*, 282; d) X. Zheng, G. A. Mueller, E. F. DeRose, R. E. London, *Antiviral Res.* **2009**, *84*, 205.

[9] A. J. Link, D. A. Tirrell, *Methods* **2005**, *36*, 291.

[10] E. Y. Wu, L. S. Beese, *J. Biol. Chem.* **2011**, *286*, 19758.

[11] Y.-C. Tsai, K. A. Johnson, *Biochemistry* **2006**, *45*, 9675.

[12] Y. Santoso, C. M. Joyce, O. Potapova, L. Le Reste, J. Hohlbein, J. P. Torella, N. D. F. Grindley, A. N. Kapanidis, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 715.