Non-natural Amino Acids

DOI: 10.1002/anie.200905978

A Highly Active DNA Polymerase with a Fluorous Core**

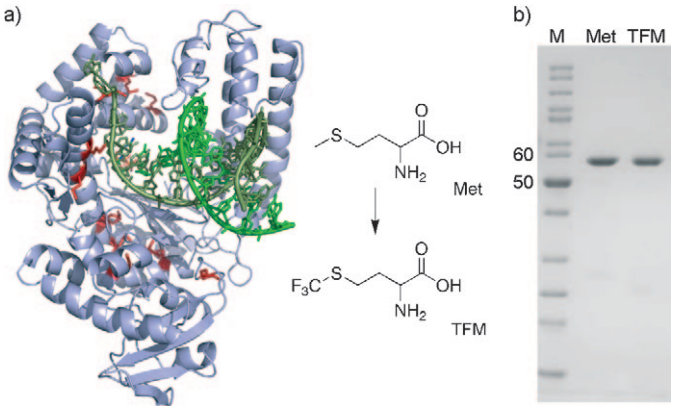
*Bastian Holzberger, Marina Rubini, Heiko M. Möller, and Andreas Marx**

DNA polymerases catalyze all DNA synthesis in the cell[1] and are key tools in important molecular biological core technologies.[2] Apart from naturally available DNA polymerases, several modified DNA polymerases with new characteristics have been developed. To date, directed evolution using the 20 natural amino acids is a promising method for the creation of nucleic acid polymerases with modified properties.[3] Yet, the incorporation of non-natural amino acids may lead to enhanced chemical and biological diversity of protein structures and properties by introduction of functional groups that are not represented by the natural amino acids.[4]

The use of auxotrophic strains, which lack the ability to biosynthesize one specific natural amino acid, offers the possibility to replace one amino acid by a non-natural analogue. After depletion of the natural amino acid in a defined medium, inducing the expression of the target protein with concomitant addition of the non-natural analogue leads to the incorporation of the non-natural amino acid by selective pressure incorporation.[5]

Fluorinated-protein synthesis has been extensively studied in the past.[6] It has also been shown that trifluorinated amino acid analogues such as trifluoromethionine (TFM) with drastically enhanced hydrophobicity[7] may lead to proteins with novel characteristics.[4] However, efforts to engineer new enzyme functions and properties by use of TFM instead of methionine are hampered, as the replacement of natural amino acids by trifluorinated analogues can lead to unfavorable interactions. Thus, the global substitution of natural amino acids by trifluorinated analogues may cause the naturally evolved protein scaffold to lose correct folding, sufficient stability, or enzymatic activity.[4a] Hence, only a few examples of proteins are known in which methionine was globally replaced by TFM,[8] including one single example of an enzyme (phage lysozyme) that is 18 kDa in size, in which about 70% of the three methionine residues are replaced by TFM.[8b]

Herein, we present the generation of a multifluorinated DNA polymerase. The N-terminally truncated version of DNA polymerase I from *Thermus aquaticus* (*KlenTaq*) is a thermophilic DNA polymerase composed of 540 amino acids (63 kDa), including 13 methionine (Met) residues[9] that were globally replaced by TFM with a substitution level of approximately 82% (Scheme 1). The multifluorinated *KlenTaq* was highly active and exhibited a similar selectivity as the wild-type (wt) enzyme. Moreover, the introduction of the NMR-active nucleus ^{19}F offers the possibility to study DNA polymerase dynamics by ^{19}F NMR spectroscopy. Despite its large size of 63 kDa, at least nine individual ^{19}F resonances are observed, which allow us to distinguish different states of the DNA polymerase on the way to incorporating a canonical or a noncanonical nucleotide. To our knowledge, this is by far the largest enzymatically active protein with Met globally replaced by TFM.



Scheme 1. Multi-TFM-labeled *KlenTaq* DNA polymerase. a) Methionine residues (Met) in *KlenTaq* DNA polymerase (PDB code: 3 KTQ) are highlighted in red. b) SDS-PAGE gel of purified *KlenTaq* wt (Met) and *KlenTaq* TFM (TFM). M: Marker [kDa].

First, we incorporated TFM into the *KlenTaq* DNA polymerase using the Met-auxotrophic *E. coli* strain B834-(DE3) in a defined minimal medium. Thus, after depletion of Met, we incubated the cells for several hours in stationary phase, washed and resuspended them, added D,L-TFM, and induced protein expression. We used the racemic mixture of TFM, as it is assumed that *E. coli* incorporates only the L-form into proteins. *KlenTaq* wt was expressed using the same *E. coli* strain in standard LB medium. Both enzymes were purified to homogeneity (Scheme 1b). The substitution level of *KlenTaq* TFM was analyzed by electrospray-ionization mass spectrometry (ESI-MS; Figure 1 and the Supporting Information). Comparison of the molecular weights of *KlenTaq* TFM and the wild type shows that on average approximately 82% of the 14 Met amino acids (13 from *KlenTaq* gene, 1 initial Met) were replaced by TFM. By

[*] B. Holzberger, Dr. M. Rubini, Prof. Dr. H. M. Möller, Prof. Dr. A. Marx
Department of Chemistry and Konstanz Research School Chemical Biology, University of Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)
Fax: (+49) 7531-88-5140
E-mail: andreas.marx@uni-konstanz.de
Homepage: www.chemie.uni-konstanz.de/~agmarx/

[**] We gratefully acknowledge Dr. A. Marquardt (Proteomics Facility, University of Konstanz) for the support with the MS analysis and financial support by the Konstanz Research School Chemical Biology and the DFG.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200905978.

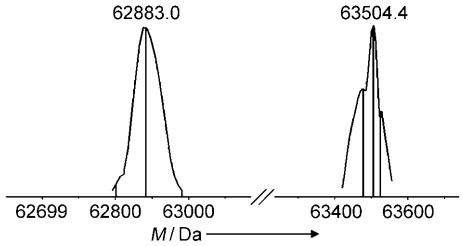


Figure 1. Deconvoluted ESI mass spectra profiles of *KlenTaq* wild type (62883.0 Da, calcd: 62830 Da) and *KlenTaq* TFM (63504.4 Da). M: molecular mass. TFM: trifluoromethionine. See also the Supporting Information.

performing a tryptic digest and analyzing the fragments with a combination of liquid chromatography (LC) and ESI-MS, we were able to identify seven of the 14 Met positions (see the Supporting Information). At position M673 we detected only TFM, while at the other six positions we found the Met-bearing peptide fragment as well as the TFM-modified one. This result indicates that every Met is fluorinated to some extent.

Next, we investigated whether the highly modified *KlenTaq* TFM is still enzymatically active. We carried out primer extension reactions at temperatures between 37 and 70 °C (Figure 2a). Both enzymes, *KlenTaq* TFM and wild type, were able to extend the primer strand to full length in every case (30 nucleotides (nt)). The incorporation of an additional nucleotide in a nontemplated manner led to 31 nt long products. This result has been observed before using 3′→5′ exonuclease-deficient DNA polymerases.[10] By performing primer extension reactions at 72 °C, we determined the specific activities of the DNA polymerases and observed that *KlenTaq* TFM retains more than one-third of the activity of the wt enzyme (see the Supporting Information). Hence, the multifluorinated *KlenTaq* TFM is still a highly active DNA polymerase, even at temperatures above 70 °C.

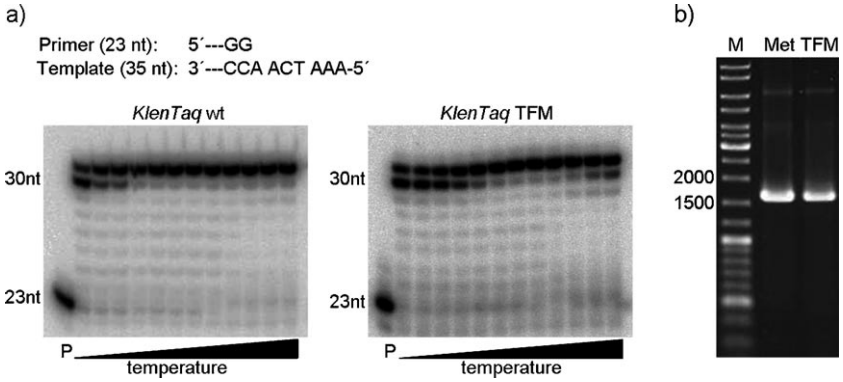


Figure 2. Enzymatic activity of *KlenTaq* TFM. a) Primer extension reactions catalyzed by *KlenTaq* wt and *KlenTaq* TFM. Partial primer template sequences employed are depicted on top. P: primer only; temperature: primer extension at 37.0, 37.8, 40.1, 43.7, 47.6, 51.5, 55.4, 59.4, 63.3, 66.9, 69.2, and 70.0 °C. b) PCR amplification of a 1655 bp fragment by *KlenTaq* wt (Met) and *KlenTaq* TFM (TFM). PCR program: initial 30 s at 95 °C and 35 cycles with 20 s at 95 °C, 60 s at 65 °C, and 120 s at 72 °C. M: Marker [bp], wt: wild type.

To investigate whether the high Met→TFM substitution level affects the stability of the protein, we studied the thermostability of *KlenTaq* TFM in comparison to the wt. The enzymes were incubated for different periods of time at 95 °C. Subsequently, their activity was examined in primer extension reactions. We observed that *KlenTaq* TFM lost stability and exhibited less than 50% primer extension activity after incubation at 95 °C for more than 30 min, whereas even after five hours at 95 °C the wt showed an activity of 90% (see the Supporting Information). Nevertheless, it turned out that *KlenTaq* TFM is still PCR-active and is still able to amplify even rather long PCR products (Figure 2b). Furthermore, in real-time PCR experiments the activity of *KlenTaq* TFM was similar to the wt activity (see the Supporting Information).

To quantify the selectivity of *KlenTaq* TFM, we determined the error spectra of the enzyme employing a reported PCR-based assay (see the Supporting Information).[11] The error rate of 3.2×10^{-5} is quite similar to that found for wild-type *KlenTaq* (8.8×10^{-5}).[12] In summary, our studies show that the global substitution of Met by TFM had most impact on the stability of the enzyme. Although the $CH_3\rightarrow CF_3$ substitution definitely increases the steric bulk of Met,[4,7] and despite the fact that the Met residues are mainly located in the core of *KlenTaq* DNA polymerase, activity and fidelity are almost unchanged. Thus, the use of a highly stable protein enables the incorporation of trifluorinated amino acids with high substitution levels even in the core of an enzyme without disturbing enzymatic activity significantly.

Several structural and functional studies of *KlenTaq* indicate that the enzyme undergoes significant conformational changes upon DNA and nucleotide binding.[13] The introduction of ^{19}F by TFM substitution of methionine allows the dynamics of the enzyme to be studied during substrate recognition and DNA synthesis. Thus, we investigated the enzyme by ^{19}F NMR spectroscopy. Free *KlenTaq* TFM without any substrates displays resonances grouped around the ^{19}F chemical shift of the TFM amino acid at $\delta=-41$ ppm. From the expected 14 resonances corresponding to the 14 sequence positions of TFM, at least nine distinct resonances (at $\delta=-38.31$, -39.20, -39.42, -39.68, -40.12, -40.52, -40.95, -41.03, and -41.39 ppm) can be distinguished (Figure 3A). These resonances show significant differential line broadening ranging from 12 Hz of the sharpest signal at $\delta=-40.95$ ppm to approximately 80 Hz of the broad signals at $\delta=-38.31$ and -41.39 ppm.

These differences are most likely caused by different relaxation times of TFM residues at different positions in the 3D structure of *KlenTaq* TFM. It can be expected that CF_3 groups located at surface-exposed, highly flexible positions exhibit much smaller line widths than residues in the hydrophobic core of the protein, owing to shorter local rotational correlation times and the lower number of protons as sources

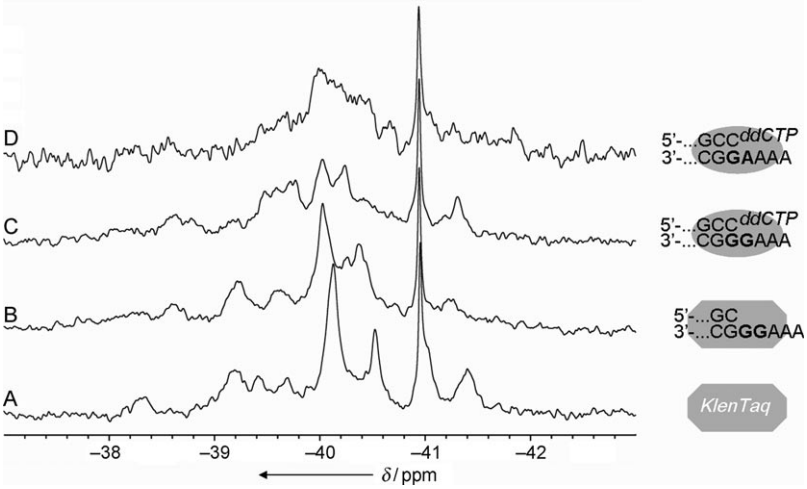


Figure 3. ^{19}F NMR spectra of *KlenTaq* TFM A) without nucleic acid, B) with 2 equiv DNA substrate (**GG**), C) with 2 equiv DNA substrate (**GG**) and after addition of 10 equiv ddCTP, and D) with 2 equiv of a different DNA substrate (**GA**) and after addition of over 100 equiv ddCTP. Enzyme concentrations for (A–C) were 42 μM and 18 μM for (D); intensity was adjusted to (A–C).

of fast dipolar relaxation. Heterogeneity caused by the incomplete substitution of Met by TFM may broaden the peaks as well. We therefore conclude that the sharp resonance at $\delta = -40.95$ ppm, observed in all spectra irrespective of the presence of substrates, originates from the initial TFM residue located in the presumably highly flexible N terminus. All other resonances are more or less strongly affected by adding substrates. Upon addition of a DNA primer/template complex leading to the formation of a binary protein–DNA complex, several sharp resonances (full width at half maximum less than 40 Hz) show up at new chemical shifts ($\delta =$ −40.03, −40.37, and −41.24 ppm), and the appearance of the other, broader signals also changes, for example, around $\delta =$ −38.31 or −39.40 ppm (Figure 3B). Subsequent addition of 2′,3′-dideoxycytidine-5′-triphosphate (ddCTP) should lead to phosphodiester formation with one ddCTP and formation of the ternary complex with a second ddCTP bound in the active site.[13] Indeed, the spectrum of the ternary complex (Figure 3C) is markedly different from those of both the free DNA polymerase and the binary complex, now with relatively sharp signals at $\delta = -40.02$, −40.23, and −41.30 ppm. Again, the broader resonances also change their appearance, for example, around $\delta = -38.69$ or −39.59 ppm. Taken together, this evidence is indicative of another conformational change of *KlenTaq* TFM upon interaction with ddCTP.

Performing the same experiments with a DNA template that contains adenine (A) instead of guanine (G) at the appropriate site in the template leads to the formation of a mismatched C–A pair upon addition of ddCTP. We found that the ^{19}F NMR spectrum differs from the match case. Almost all CF_3 resonances of the mismatch complex (Figure 3D) suffer from severe line broadening. Apart from the signal at $\delta =$ −40.95 ppm that we attribute to the N-terminal TFM residue (see above), there are no resonances with sharp lines. Instead, a rather broad signal hump is present around $\delta = -40.2$ ppm. This finding suggests major changes of protein dynamics or the presence of structural heterogeneity in the mismatch case at the relatively high concentration of ddCTP used herein to ensure saturation of the nucleoside triphosphate binding site. These structural changes might contribute to the selectivity of *KlenTaq* TFM by reducing the efficiency of noncanonical nucleotide incorporation. Thus, using ^{19}F NMR spectroscopy we were able to clearly distinguish four different states of the DNA polymerase on the way to nucleotide incorporation.

In summary, we report the generation of a multifluorinated DNA polymerase that exhibits nearly the same enzymatic activity and fidelity as the parental enzyme. Owing to the high Met→TFM substitution level and the high signal intensity of the CF_3 resonance even when incorporated into this comparatively large protein, the *KlenTaq* TFM serves as an optimal ^{19}F NMR probe to elucidate alterations of conformation and dynamics during enzyme catalysis. These findings open up new opportunities for further investigations, such as the directed evolution of DNA polymerases with new characteristics for future applications owing to the expanded amino acid repertoire[3,14] or comparative studies of DNA polymerase dynamics by ^{19}F NMR spectroscopy, to gain further insights into DNA synthesis fidelity and substrate spectra.

Received: October 23, 2009
Published online: January 13, 2010

Keywords: DNA polymerases · NMR spectroscopy · non-natural amino acids · protein engineering · trifluoromethionine

[1] A. Kornberg, T. A. Baker, *DNA Replication*, 2nd ed., Freeman, New York, **1991**.
[2] J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, **2001**.
[3] a) D. Loakes, P. Holliger, *Chem. Commun.* **2009**, 4619–4631; b) A. A. Henry, F. E. Romesberg, *Curr. Opin. Biotechnol.* **2005**, *16*, 370–377; c) S. Brakmann, *Cell. Mol. Life Sci.* **2005**, *62*, 2634–2646.
[4] a) N. Budisa, *Engineering the Genetic Code*, Wiley-VCH, Weinheim, **2006**; b) Y. Ryu, P. G. Schultz, *Nat. Methods* **2006**, *3*, 263–265; c) L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34–68; *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66; d) A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.
[5] C. Minks, S. Alefelder, L. Moroder, R. Huber, N. Budisa, *Tetrahedron* **2000**, *56*, 9431–9442.
[6] See, for example: a) J. C. Jackson, J. T. Hammill, R. A. Mehl, *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166; b) J. T. Wong, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6303–6306; c) E. A. Pratt, C. Ho, *Biochemistry* **1975**, *14*, 3035–3040; d) O. M. Rennert, H. S. Anker, *Biochemistry* **1963**, *2*, 471–476.
[7] E. N. G. Marsh, *Chem. Biol.* **2000**, *7*, R153–R157.

[8] a) N. Budisa, O. Pipitone, I. Siwanowicz, M. Rubini, P. P. Pal, T. A. Holak, M. L. Gelmi, *Chem. Biodiversity* **2004**, *1*, 1465 – 1475; b) H. Duewel, E. Daub, V. Robinson, J. F. Honek, *Biochemistry* **1997**, *36*, 3404 – 3416.
[9] W. M. Barnes, *Gene* **1992**, *112*, 29 – 35.
[10] J. M. Clark, *Nucleic Acids Res.* **1988**, *16*, 9677 – 9686.
[11] P. H. Patel, H. Kawate, E. Adman, M. Ashbach, L. A. Loeb, *J. Biol. Chem.* **2001**, *276*, 5044 – 5051.
[12] C. Gloeckner, K. B. M. Sauter, A. Marx, *Angew. Chem.* **2007**, *119*, 3175 – 3178; *Angew. Chem. Int. Ed.* **2007**, *46*, 3115 – 3117.
[13] a) W. J. Allen, P. J. Rothwell, G. Waksman, *Protein Sci.* **2008**, *17*, 401 – 408; b) P. J. Rothwell, G. Waksman, *J. Biol. Chem.* **2007**, *282*, 28884 – 28892; c) P. J. Rothwell, V. Mitaksov, G. Waksman, *Mol. Cell* **2005**, *19*, 345 – 355.
[14] a) C. C. Liu, A. V. Mack, M.-L. Tsao, J. H. Mills, H. S. Lee, H. Choe, M. Farzan, P. G. Schultz, V. V. Smider, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17688 – 17693; b) T. H. Yoo, A. J. Link, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887 – 13890; c) J. K. Montclare, D. A. Tirrell, *Angew. Chem.* **2006**, *118*, 4630 – 4633; *Angew. Chem. Int. Ed.* **2006**, *45*, 4518 – 4521.

