

DNA Methylation

Direct Sensing of 5-Methylcytosine by Polymerase Chain Reaction**

Joos Aschenbrenner, Matthias Drum, Hüsnü Topal, Markus Wieland, and Andreas Marx*

Abstract: The epigenetic control of genes by the methylation of cytosine resulting in 5-methylcytosine (5mC) has fundamental implications for human development and disease. Analysis of alterations in DNA methylation patterns is an emerging tool for cancer diagnostics and prognostics. Here we report that two thermostable DNA polymerases, namely the DNA polymerase KlenTaq derived from *Thermus aquaticus* and the KOD DNA polymerase from *Thermococcus kodakaraensis*, are able to extend 3'-mismatched primer strands more efficiently from 5mC than from unmethylated C. This feature was advanced by generating a DNA polymerase mutant with further improved 5mC/C discrimination properties and its successful application in a novel methylation-specific PCR approach directly from untreated human genomic DNA.

Methylation of cytosines is a major epigenetic mark and the most abundant DNA modification in vertebrates.^[1] Methylated cytosines are found as symmetrical 5-methylcytosine (5mC) of the dinucleotide CpG within promoter regions,^[2] in which 75% are methylated throughout the mammalian genome. In recent years, it has become evident that promoter methylation is crucial for activating and silencing gene expression.^[3] Moreover, dynamic changes of methylation patterns are important for the development of mammals.^[4] They control, for example, X-inactivation,^[5] genomic imprinting,^[5] and the development of primordial germ cells.^[6] Furthermore, alterations in DNA methylation can be an integral event in the onset of diseases like cancer.^[7] The discovery that particular hypo- or hypermethylation events are unique for human malignancy^[8] renders 5mC a promising biomarker for cancer diagnosis. In reality, many DNA-methylation-based biomarkers have been evaluated in cancer research.^[9] Therefore, several methods for the detection of 5mC in the human genome are employed.^[10] The main

approaches rely on endonuclease digestion,^[11] affinity enrichment,^[12] and bisulfite conversion.^[13] The most common method to obtain single-nucleotide resolution is based on treatment of the sample with bisulfite, resulting in the conversion of cytosine to uracil whereas 5mC remains unchanged;^[14] this is followed by DNA sequencing or PCR amplification. All available methods, however, have significant drawbacks. While affinity enrichment falls short in yielding information about individual CpG dinucleotides, endonuclease- and bisulfite-based methods are prone to false-positive results due to incomplete conversion.^[15] The necessity of implementing pretreatment and purification steps results in a higher risk of contamination^[16] as well as loss of DNA sample.^[17] Taken together, the laborious, time-consuming,^[18] and error-prone nature of DNA methylation assays is a significant barrier for their advent in practical clinical diagnostics. Here, we demonstrate the feasibility of a superior DNA methylation profiling approach directly from genomic samples.

In order to examine the behavior of two thermostable DNA polymerases when encountering a primer bearing

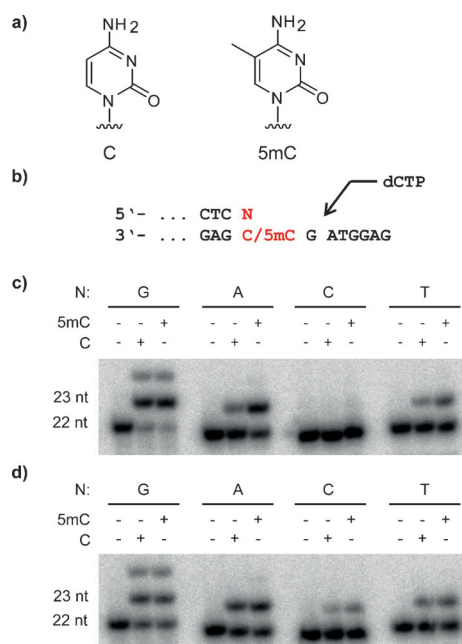


Figure 1. Comparison of KlenTaq and KOD exo- in single-nucleotide incorporation with different primers and either methylated or non-methylated template. a) Chemical structure of C (left) and 5mC (right). b) Partial primer template sequences used in primer extension experiments (N = G/A/T/C). c,d) Single-nucleotide incorporation catalyzed by KlenTaq (c) and KOD exo- (d) using primers bearing one of the four different nucleotides at the 3'-terminus opposite either C or 5mC. 100 μ M dCTP and 25 nM of the respective DNA polymerase were applied. The reaction time was 60 s.

[*] J. Aschenbrenner, M. Drum, Dr. H. Topal, Dr. M. Wieland, Prof. Dr. A. Marx
 Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz
 Universitätsstrasse 10, 78457 Konstanz (Germany)
 E-mail: andreas.marx@uni-konstanz.de

[**] We acknowledge support by the European Research Council (ERC Advanced Grant 339834), the Bundesministerium für Wirtschaft und Technologie, and the Konstanz Research School Chemical Biology.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201403745>.

© 2014 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

a mismatch at its 3'-terminus opposite either C or 5mC, we investigated the extension from four primers differing in their 3'-terminal nucleotide (A-, C-, G-, or T-primer) paired with two different oligonucleotide templates that carry either a C or a 5mC opposite the primer end (Figure 1b). Therefore, we performed single-nucleotide-incorporation experiments followed by analysis through denaturing polyacrylamide gel electrophoresis (PAGE) and visualization using autoradiography. Figure 1c shows extensions of the four different primers by the sequence family A DNA polymerase KlenTaq.^[19] Whereas there is no difference in extending from C or 5mC when the matched G-primer is used, the A- and T-primers are more efficiently extended in the case of the 5mC template. Under the chosen conditions, the extension for the A-primer was 56% when paired with the 5mC template in comparison to 18% when paired with the C template. The primer bearing C at its 3'-terminus was not extended for either template. Along these lines, we also tested an exonuclease-deficient variant of a DNA polymerase descending from another sequence family, the family B KOD DNA polymerase (KOD exo-)^[20] originating from *Thermococcus kodakaraensis*. We found that discrimination between the mispaired C and 5mC templates is also observed with KOD exo-, albeit to a lower extent than with KlenTaq (47% extension of the A-primer when paired with 5mC template as compared to 34% with C template; Figure 1d).

Next, we aimed at generating DNA polymerase mutants with enhanced discrimination. We inspected previously published KlenTaq and KOD exo- crystal structures with bound primer template complex,^[21] looking for amino acids that might be able to interfere with the mismatched cytosine with the intention to exchange those by sterically more demanding residues. We came to the conclusion that the template binding cleft of KlenTaq is already packed with sterically demanding amino acids. For KOD exo-, however, we identified a glycine (G498) in immediate proximity to the mismatched cytosine (Figure 2a). Noticeably, G498 is located

in the template binding site where it is able to interact with the phosphate backbone of the nucleotide paired to the primer's 3'-terminal nucleobase (Figure 2a). We reasoned that mutating G498 into a sterically more demanding amino acid would result in a more crowded template binding site, which might effect discrimination between C and 5mC upon extension from mismatched primer strands. In fact, we found that mutating G498 into methionine results in a KOD exo-variant that exhibits the desired properties. The KOD exo-G498M mutant was obtained by site-directed mutagenesis followed by expression in *E. coli* BL21 as previously described.^[21b,22] The purified KOD exo- wild-type and KOD G498M enzymes were analyzed by SDS PAGE (Figure 2b) and compared using the very same reaction conditions and enzyme concentrations. First, we applied KOD G498M in primer-extension experiments as described above. The engineered KOD G498M features considerably increased discrimination between the C and 5mC template as compared to that of the wild-type enzyme when the mismatched A-primer was used (44% extension rate for 5mC as compared to 20% for C) (Figure 3a). This effect is even more pronounced for full-length primer extension. Here, 58% of the applied primer was extended to the full-length product on the 5mC template as compared to 12% on the C template (Figure 3b). Interestingly, when the mismatched reaction was performed with template C a significant pausing of the enzyme was observed after incorporation of one nucleotide (Figure 3b), a result that was not observed to this extent with the 5mC template. We determined steady-state kinetics^[23] for primer extension by the incorporation of a single nucleotide (Table S1). All three enzymes were found to be equally efficient for either template when the matched G-primer was used (Table S1, Figure S1, and Figure S2). When comparing extension from the mismatched C template to the mismatched 5mC template for KOD exo- wild-type, we found that the discrimination is mainly based on differences of K_M . In contrast, for KOD G498M as well as for KlenTaq wild-type

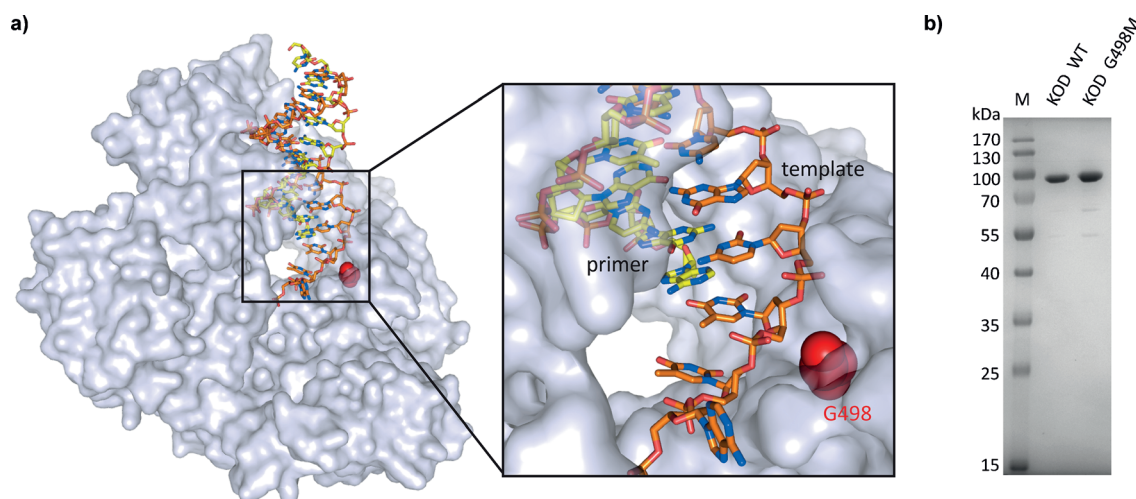


Figure 2. Rational design and preparation of a KOD mutant with enhanced 5mC discrimination. a) Crystal structure of KOD with bound primer template complex. Primer and template are shown as yellow and orange rods. G498 is highlighted as red spheres. Adapted from PDB 4K8Z^[21b] using PyMOL (Schrödinger, LLC; New York, NY). b) Coomassie-stained SDS PAGE of Ni-NTA-purified 6×His-tagged KOD wild-type (middle) and KOD G498M (right).

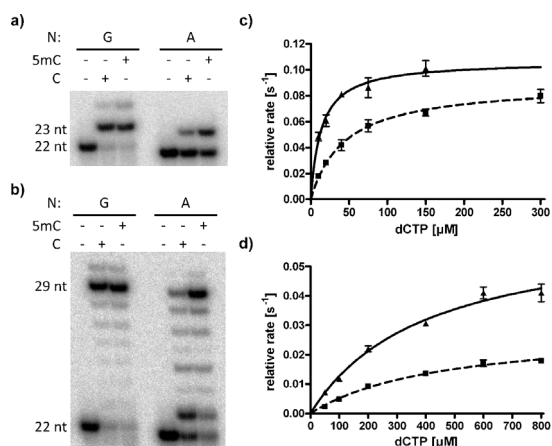


Figure 3. KOD G498M features enhanced 5mC discrimination.

a) Single-nucleotide incorporation with KOD G498M using primers bearing G or A at the 3'-terminus opposite either C or 5mC. 100 μ M dCTP and 25 nM KOD G498M were applied. The reaction time was 5 min. b) Full-length primer extension experiment with KOD G498M using primers bearing G or A at the 3'-terminus opposite either C or 5mC. 100 μ M dCTP and 25 nM KOD G498M were applied. The reaction time was 30 min. c,d) Steady-state kinetics of single-nucleotide incorporation next to A-mismatched C (dashed) and 5mC (solid) catalyzed by KOD exo- wild-type (c) and KOD G498M (d). Data points derive from triplicates. Error bars represent the standard deviation.

we found that the differences are mainly based on k_{cat} (Table S1, Figure S1, and Figure 3c,d).

We examined the potential to exploit these differences in the catalytic efficiency of processing 5mC and C in PCR experiments on human genomic DNA (gDNA). Therefore, we analyzed a particular CpG site in the promoter region of NANOG in HeLa gDNA. NANOG is an epigenetically regulated gene which is associated with pluripotency of cells^[24] and was found to be hypomethylated in its promoter region in metastatic human liver cancer cells.^[25] Previous evaluation of the analyzed methylation site by bisulfite sequencing characterized it as unmethylated in HeLa cells.^[25] As a control for full methylation we employed HeLa genomic DNA that was enzymatically methylated in vitro with a CpG methylase. We designed forward primers bearing either G or A at their 3'-termini opposite the cytosine of interest and a reverse primer binding 43 nt downstream so that PCR amplification should deliver a 86 bp amplicate (Figure S3). PCR employing the wild-type KOD exo- or KlenTaq enzymes did not show any differentiation in PCR efficiency. However, the use of KOD G498M DNA polymerase instead revealed that this variant is indeed capable of discriminating between methylated and unmethylated cytosine (Figure 4). While real-time PCR and melting curves are similar for untreated and CpG-methylated DNA when matched primers are used (i.e., terminating with G), employment of the mismatched primer (i.e., terminating with A) leads to delayed amplification as well as reduced endpoint fluorescence when the untreated DNA is compared to the CpG-methylated

DNA (Figure 4a). Melting curves of the amplified DNA confirm this decrease in PCR efficiency for the non-methylated template (Figure 4b). Finally, quantitative analysis of the reaction mixtures by agarose gel electrophoresis substantiates these findings. As seen in Figure 4c, the yield of a specific amplicate for the methylated template is reasonable considering we employed a mismatched primer. For the unmethylated template, however, the yield is significantly reduced. Moreover, KOD G498M is highly selective for the desired amplicate when the methylated template is used, whereas a slight band of byproduct appears when the unmethylated template is used. Primer dimers are formed in PCR without HeLa genomic DNA but do not emerge in the presence of template.

So far, detection of 5mC with single-base resolution has been restricted to the troublesome conversion of DNA by bisulfite or any other manipulation prior to analysis. Here we demonstrate the feasibility of a PCR system sensing 5mC directly from genomic DNA. The approach is based on the differential extension of mismatched primer strands by two well-studied DNA polymerases depending on whether the primer terminates opposite a template 5mC or C. KlenTaq was investigated as a prominent member of sequence family A and an exonuclease-deficient variant of KOD DNA polymerase as a representative of the sequence family B.^[26] Both DNA polymerases are thermostable and commonly used in many core biotechnological applications. The discrimination between C and 5mC was greater with KlenTaq than with KOD exo-. Interestingly, KlenTaq wild-type exhibits a more sterically crowded environment close to the relevant template site in comparison to KOD exo-. This might well explain the enhanced discriminating properties of KlenTaq wild-type compared to KOD exo- wild-type. However, we were able to enhance the discrimination of

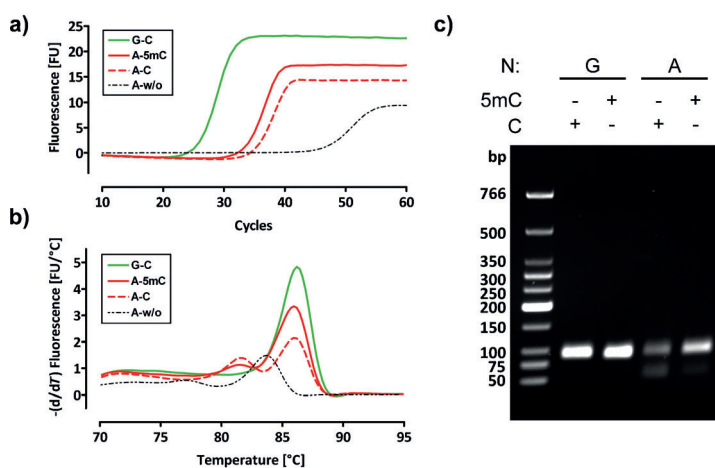


Figure 4. KOD G498M differentiates between methylated and non-methylated cytosine in PCR experiments from HeLa genomic DNA. a) Real-time PCR of the NANOG promoter region catalyzed by KOD G498M with matched (green) and mismatched primer (red) from HeLa gDNA (dashed) or CpG methylated HeLa gDNA (solid). b) Melting curves of amplicates deriving from (a). c) Agarose gel from PCR amplicates deriving from HeLa gDNA (C) or CpG methylated HeLa gDNA (5mC) with matched (lane G) or mismatched primer (lane A). Curves and gels are representative for numerous experiments.

KOD exo- by increasing the steric crowding of the template binding site in close vicinity to the residues bearing the C or 5mC of interest. Primer extensions in the presence of all four dNTPs indicate that KOD exo- G498M not only discriminates for C over 5mC upon extension by one nucleotide but also when the mismatch is already one position distal from the primer terminus (Figure 3b). Furthermore, KOD G498M is capable of discriminating between C and 5mC in PCR from a genomic DNA target. On this basis, the methylation state of a single nucleotide in the entire human genome can be ascertained by a single PCR step.

As crystal structures of mismatched primer template complexes bound to DNA polymerases have not yet been described, one can only speculate about the origin of the 5mC/C discrimination. It has been proposed that when a mismatch is encountered, active misalignment of catalytic residues in the DNA polymerase results in reduced incorporation rate and therefore higher specificity of the enzyme.^[27] Thus, discrimination could derive from differences in the gained substrate binding energy by misalignment of catalytic residues. Thereby, the enzyme would be kept in an inactive conformation more efficiently when bound to mismatched C rather than mismatched 5mC, resulting in a favored extension of 5mC. The differential states might be due to steric interference between the enzyme and the additional 5-methyl group in 5mC.

In the past, the possibility of different extension efficiencies was discussed as a mechanism for the increase of the mutation rates of 5mC in vivo. In 1992, Shen et al. investigated three DNA polymerases in their properties of extending mismatched primer strands at the C and 5mC position.^[28] They found significant differences in mismatch extension only when using AMV reverse transcriptase. However, no significant discriminations were observed for a sequence family A DNA polymerase (i.e., the exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I) and a sequence family B enzyme (i.e., *Drosophila* DNA polymerase α). As mentioned, the herein investigated enzymes belong to sequence family A (KlenTaq) and B (KOD exo-) and show discrimination. These findings might hint at the fact that subtle changes of the enzyme scaffold might cause altered effects of C/5mC discrimination. Thus, future investigations will aim at evolving new DNA polymerase mutants with even more enhanced discrimination for application in methylation specific PCR approaches.^[29]

Received: March 27, 2014

Published online: June 12, 2014

Keywords: DNA methylation · DNA polymerases · enzyme engineering · polymerase chain reaction

- [1] A. P. Bird, *Nucleic Acids Res.* **1980**, *8*, 1499–1504.
 [2] M. Ehrlich, R. Wang, *Science* **1981**, *212*, 1350–1357.
 [3] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068–1070.
 [4] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915–926.
 [5] E. Li, *Nat. Rev. Genet.* **2002**, *3*, 662–673.

- [6] a) S. Ooi, D. Wolf, O. Hartung, S. Agarwal, G. Daley, S. Goff, T. Bestor, *Epigenet. Chromatin* **2010**, *3*, 17; b) M. Weber, D. Schübeler, *Curr. Opin. Cell Biol.* **2007**, *19*, 273–280.
 [7] a) M. R. H. Estéicio, J. Gallegos, C. Vallot, R. J. Castoro, W. Chung, S. Maegawa, Y. Oki, Y. Kondo, J. Jelinek, L. Shen, H. Hartung, P. D. Aplan, B. A. Czerniak, S. Liang, J.-P. J. Issa, *Genome Res.* **2010**, *20*, 1369–1382; b) P. Jones, S. Baylin, *Nat. Rev. Genet.* **2002**, *3*, 415–428.
 [8] a) M. Esteller, J. Garcia-Foncillas, E. Andion, S. N. Goodman, O. F. Hidalgo, V. Vanaclocha, S. B. Baylin, J. G. Herman, *N. Engl. J. Med.* **2000**, *343*, 1350–1354; b) M. Esteller, J. M. Silva, G. Dominguez, F. Bonilla, X. Matias-Guiu, E. Lerma, E. Bussaglia, J. Prat, I. C. Harkes, E. A. Repasky, E. Gabrielson, M. Schutte, S. B. Baylin, J. G. Herman, *J. Natl. Cancer Inst.* **2000**, *92*, 564–569.
 [9] a) H. Heyn, M. Esteller, *Nat. Rev. Genet.* **2012**, *13*, 679–692; b) P. Laird, *Nat. Rev. Cancer* **2003**, *3*, 253–266.
 [10] P. Laird, *Nat. Rev. Cancer* **2010**, *11*, 191–203.
 [11] J. Kaput, T. W. Sneider, *Nucleic Acids Res.* **1979**, *7*, 2303–2322.
 [12] S. H. Cross, J. A. Charlton, X. Nan, A. P. Bird, *Nat. Genet.* **1994**, *6*, 236–244.
 [13] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831.
 [14] H. Hayatsu, Y. Wataya, K. Kai, S. Iida, *Biochemistry* **1970**, *9*, 2858–2865.
 [15] J. Harrison, C. Stirzaker, S. J. Clark, *Anal. Biochem.* **1998**, *264*, 129–132.
 [16] D. P. Genereux, W. C. Johnson, A. F. Burden, R. Stöger, C. D. Laird, *Nucleic Acids Res.* **2008**, *36*, e150.
 [17] C. Grunau, S. J. Clark, A. Rosenthal, *Nucleic Acids Res.* **2001**, *29*, e65.
 [18] S. J. Clark, A. Statham, C. Stirzaker, P. L. Molloy, M. Frommer, *Nat. Protoc.* **2006**, *1*, 2353–2364.
 [19] a) W. M. Barnes, *Gene* **1992**, *112*, 29–35; b) S. Korolev, M. Nayal, W. M. Barnes, E. Di Cera, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9264–9268.
 [20] a) H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai, *J. Mol. Biol.* **2001**, *306*, 469–477; b) M. Takagi, M. Nishioka, H. Kahirahara, M. Kitabayashi, H. Inoue, B. Kawakami, M. Oka, T. Imanaka, *Appl. Environ. Microbiol.* **1997**, *63*, 4504–4510; c) M. Nishioka, H. Mizuguchi, S. Fujiwara, S. Komatsubara, M. Kitabayashi, H. Uemura, M. Takagi, T. Imanaka, *J. Biotechnol.* **2001**, *88*, 141–149.
 [21] a) K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, O. Phillip, F. E. Romesberg, A. Marx, *Nat. Chem. Biol.* **2012**, *8*, 612–614; b) K. Bergen, K. Betz, W. Welte, K. Diederichs, A. Marx, *ChemBioChem* **2013**, *14*, 1058–1062.
 [22] a) S. Obeid, N. Blatter, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, A. Marx, *EMBO J.* **2010**, *29*, 1738–1747; b) 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.
 [23] a) M. S. Boosalis, J. Petruska, M. F. Goodman, *J. Biol. Chem.* **1987**, *262*, 14689–14696; b) S. Creighton, M. M. Huang, H. Cai, N. Arnheim, M. F. Goodman, *J. Biol. Chem.* **1992**, *267*, 2633–2639; c) J. Petruska, M. F. Goodman, M. S. Boosalis, L. C. Sowers, C. Cheong, I. Tinoco, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6252–6256.
 [24] a) J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, J. A. Thomson, *Science* **2007**, *318*, 1917–1920; b) Y. Tay, J. Zhang, A. M. Thomson, B. Lim, I. Rigoutsos, *Nature* **2008**, *455*, 1124.
 [25] X. Q. Wang, R. K. Ng, X. Ming, W. Zhang, L. Chen, A. C. Y. Chu, R. Pang, C. M. Lo, S. W. Tsao, X. Liu, R. T. P. Poon, S. T. Fan, *PLoS One* **2013**, *8*, e72435.

- [26] D. K. Braithwaite, J. Ito, *Nucleic Acids Res.* **1993**, *21*, 787–802.
[27] Y.-C. Tsai, K. A. Johnson, *Biochemistry* **2006**, *45*, 9675–9687.
[28] J.-C. Shen, S. Creighton, P. A. Jones, M. F. Goodman, *Nucleic Acids Res.* **1992**, *20*, 5119–5125.
[29] a) C. Gloeckner, R. Kranaster, A. Marx, *Curr. Protoc. Chem. Biol.* **2010**, *2*, 89–109; b) C. Gloeckner, K. B. M. Sauter, A. Marx, *Angew. Chem.* **2007**, *119*, 3175–3178; *Angew. Chem. Int. Ed.* **2007**, *46*, 3115–3117; c) S. Giesecking, K. Bergen, F. Di Pasquale, K. Diederichs, W. Welte, A. Marx, *J. Biol. Chem.* **2011**, *286*, 4011–4020; d) T. Chen, F. E. Romesberg, *FEBS Lett.* **2014**, *588*, 219–229.
-