How to cite: *Angew. Chem. Int. Ed.* **2024**, *63*, e202413304
doi.org/10.1002/anie.202413304

DNA Polymerase | Hot Paper

A DNA Polymerase Variant Senses the Epigenetic Marker 5-Methylcytosine by Increased Misincorporation

*Melanie Henkel, Alexander Fillbrunn, Virginie Marchand, Govindan Raghunathan, Michael R. Berthold, Yuri Motorin, and Andreas Marx**

Abstract: Dysregulation of DNA methylation is associated with human disease, particularly cancer, and the assessment of aberrant methylation patterns holds great promise for clinical diagnostics. However, DNA polymerases do not effectively discriminate between processing 5-methylcytosine (5 mC) and unmethylated cytosine, resulting in the silencing of methylation information during amplification or sequencing. As a result, current detection methods require multi-step DNA conversion treatments or careful analysis of sequencing data to decipher individual 5 mC bases. To overcome these challenges, we propose a novel DNA polymerase-mediated 5 mC detection approach. Here, we describe the engineering of a thermostable DNA polymerase variant derived from *Thermus aquaticus* with altered fidelity towards 5 mC. Using a screening-based evolutionary approach, we have identified a DNA polymerase that exhibits increased misincorporation towards 5 mC during DNA synthesis. This DNA polymerase generates mutation signatures at methylated CpG sites, allowing direct detection of 5 mC by reading an increased error rate after sequencing without prior treatment of the sample DNA.

The most abundant epigenetic modification in the human genome is 5-methylcytosine (5 mC), accounting for 4 % of all cytosine (C) nucleobases.[1] The postreplicative enzymatic addition of a methyl group to the 5-carbon atom of C is described as the process of DNA methylation[2] and occurs mainly in a symmetrical pattern at 5′-cytosine-phosphate-guanine-3′ (CpG) dinucleotides.[3] DNA methylation is involved in the regulation of gene expression[4] and has a crucial impact on processes such as X-inactivation,[5] genomic imprinting,[6] cellular development and differentiation.[7] Dysfunction and alterations in methylation patterns can be linked to the onset of various human diseases, especially malignancies.[8] Hence, abnormal hyper- and hypomethylation are considered promising biomarkers for cancer diagnosis and prognosis.[9]

DNA polymerases are necessary for the replication and maintenance of genetic information encoded in DNA. They are widely used in biotechnological applications for nucleic acid analysis through amplification or sequencing.[10] However, DNA polymerases do not distinguish between processing 5 mC and unmodified C,[11] resulting in the loss of methylation information during DNA synthesis. The base pairing properties of 5 mC remain unaffected because the methyl group of 5 mC does not interfere with the Watson-Crick hydrogen bonding[12] and is well-accommodated in the major groove of the DNA double helix, where significantly bulkier modifications, up to several orders of magnitude larger than the natural substrate, are accepted by DNA polymerases.[13]

Consequently, short-read sequencing methods commonly used for single-base resolution methylation profiling require a chemical and/or enzymatic pre-treatment of the DNA sample to discriminate 5 mC sites from unmodified Cs. Bisulfite sequencing is the gold standard for 5 mC

[*] M. Henkel, G. Raghunathan, Prof. Dr. A. Marx
Department of Chemistry, Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstraße 10, 78464 Konstanz (Germany)
E-mail: andreas.marx@uni-konstanz.de

A. Fillbrunn, Prof. Dr. M. R. Berthold
Department of Computer Science, Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstraße 10, 78464 Konstanz (Germany)

V. Marchand, Prof. Dr. Y. Motorin
Epitranscriptomics and Sequencing (EpiRNA-Seq) Core Facility, UAR2008/US40 Ingénierie Biologie Santé en Lorraine (IBSLor), CNRS-UL-INSERM
Université de Lorraine
9 Avenue de la Forêt de Haye, BP 20199, 54505 Vandoeuvre-les-Nancy (France)

Prof. Dr. M. R. Berthold
KNIME AG
Talacker 50, 8001 Zurich (Switzerland)

Prof. Dr. Y. Motorin
Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), UMR7365 CNRS-Université de Lorraine
Université de Lorraine
9 Avenue de la Forêt de Haye, BP 20199, 54505 Vandoeuvre-les-Nancy (France)

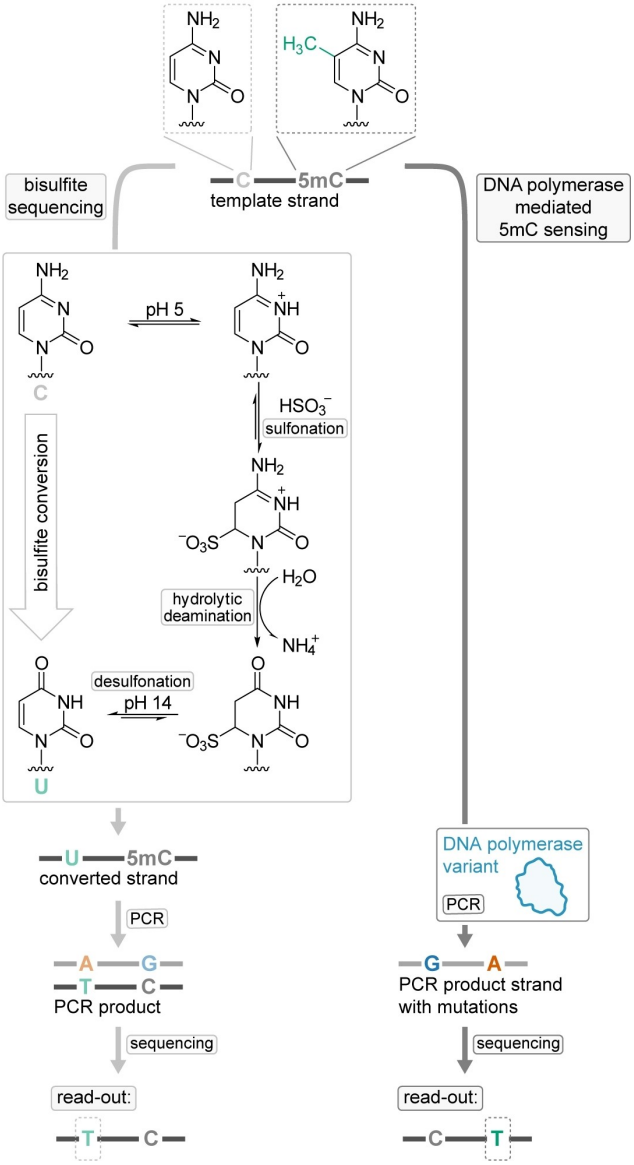
detection and relies on bisulfite treatment of the template DNA, which converts C to uracil (U) while leaving 5 mC unchanged.[14] The conversion is followed by amplification and sequencing, resulting in the reading of 5 mC as remaining C, while unmethylated Cs are read as thymine (T) bases.[15] Although this method is well established, it has significant drawbacks.[16] The conversion of unmodified C to U reduces the complexity of DNA sequences and impairs downstream processes and sequencing data analysis.[17] Furthermore, harsh reaction conditions lead to fragmentation and degradation of up to 99 % of the sample DNA.[18] Enzymatic bisulfite-free conversion techniques use considerably milder reaction conditions and the sample DNA stays largely intact after treatment.[19] Still, conversion-based methods generally suffer from multiple processing steps, adenine (A)T-rich sequence products, and incomplete or incorrect conversion.[20] In addition to sequencing-by-synthesis approaches, Oxford Nanopore long-read sequencing, which features direct 5 mC detection, represents a promising alternative.[21] However, the successful detection of 5 mC requires large amounts of unamplified template DNA and extensive algorithm training.[22] Furthermore, the variability of signal changes leads to differing as well as limited accuracy of nanopore methylation calling tools.[23]

To enable direct detection of 5 mC at single-base resolution without the need for sample conversion before usage, we report herein on an engineered thermostable enzyme based on the large fragment of the *Thermus aquaticus* DNA polymerase (KlenTaq DNA polymerase,[24] hereafter referred to as KTq) with altered fidelity opposite 5 mC. The modified base is detected from its unmodified counterpart by increased misincorporation during polymerase chain reaction (PCR). Subsequently, the resulting mutation signatures at methylated CpG sites are detected by reading elevated error rates (T base calls) in the native sequence context after next-generation sequencing (NGS) (Scheme 1).

First, a screening-based engineering approach was developed to discover a KTq variant with increased misincorporation opposite methylated bases from a mutant library. The DNA polymerase variants applied for screening included single amino acid substitutions with a broad spectrum of target mutation sites.[25] Additionally, functional promising mutation sites were rationally combined to generate double mutation variants, resulting in a total of 970 focused KTq variants. Furthermore, we tested over 2100 KTq variants that contained multiple mutations, established by combinatorial shuffling of active mutants using the random chimeragenesis on a transient template (RACHITT)[26] method followed by preselection for PCR activity.[25h]

The libraries were expressed in *Escherichia coli* (*E. coli*) and the DNA polymerase variants were evaluated directly from cell lysates in single-nucleotide incorporation experiments. The screening reactions were performed in parallel with oligonucleotides that had the same sequence and either C or 5 mC at the site of first incorporation as templates. The 5′-labeling of the primers with two different fluorescent dyes, 6-carboxyfluorescein (FAM) and hexachlorofluorescein (HEX), and the varying 5′-overhangs allowed the pooling of primer extension reactions for the multiplexed analysis of several KTq variants by capillary electrophoresis (CE) (Figure 1a). As substrate for single-nucleotide incorporation either the complementary 2′-deoxyguanosine-5′-triphosphate (dGTP) (match) or the non-complementary 2′-deoxyadenosine-5′-triphosphate (dATP) (mismatch) was applied. We selected dATP to identify KTq variants with increased misincorporation activity because the KTq wild-type enzyme misincorporates 2′-deoxyadenosine-5′-monophosphate (dAMP) more efficiently than the other mismatching nucleotides opposite C and 5 mC (Figure S1).

To identify a DNA polymerase with an increased error rate opposite 5 mC, we screened for KTq variants that discriminate against 5 mC by reduced 2′-deoxyguanosine-5′-monophosphate (dGMP) (match) incorporation opposite



Scheme 1. Comparison of bisulfite sequencing (left) and DNA polymerase mediated 5 mC sensing (right) strategies for the detection of 5 mC.

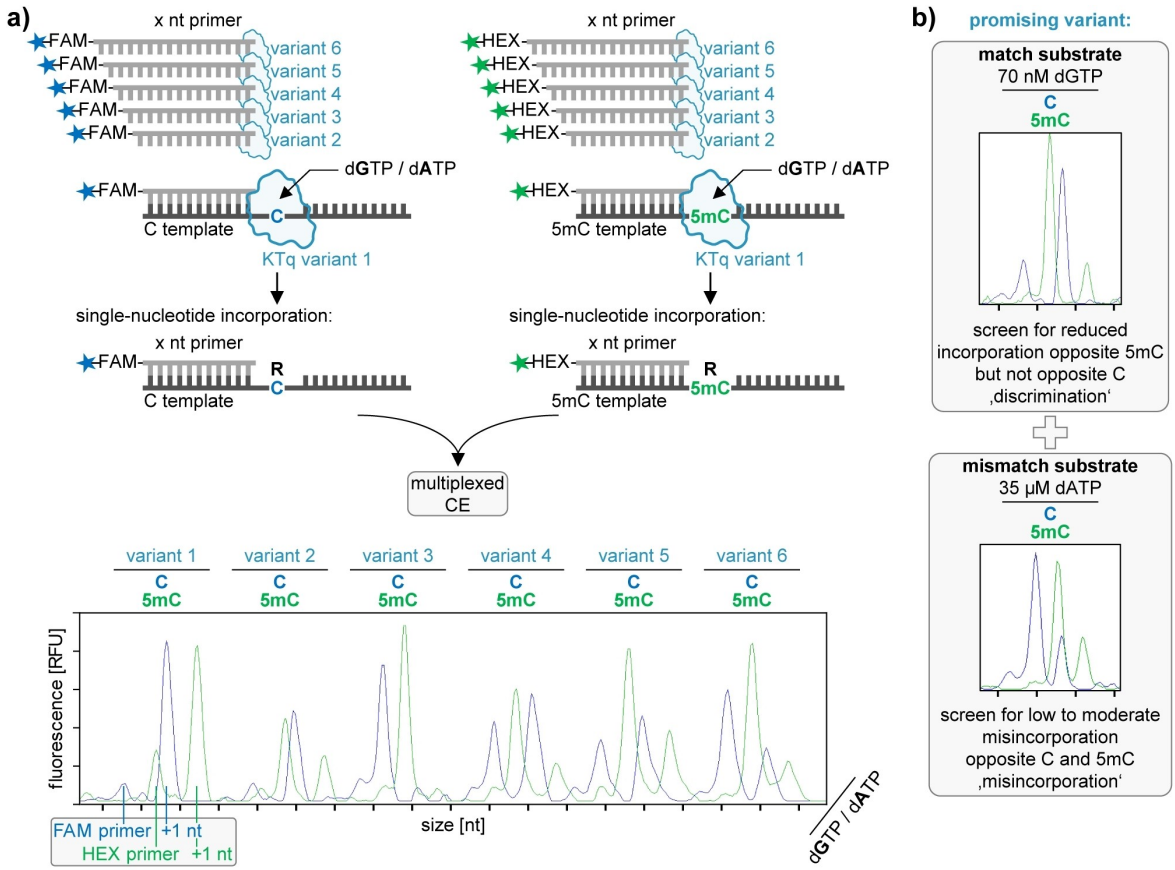


Figure 1. Screening for KTq variants with increased misincorporation opposite 5 mC. a) DNA polymerase library expression lysates were used for single-nucleotide incorporation of dGMP or dAMP opposite C and 5 mC. Utilization of primers with different length and 5′-fluorescent labeling, FAM for reactions with C and HEX for reactions with 5 mC, enabled the multiplexed analysis of 12 primer extension reactions in one capillary by CE. Primer size correlates with oligonucleotide migration time and differences for FAM- and HEX- labeled primers of same length derive from differential electrophoretic mobility of the fluorophores. Fluorescent signal shift to the right in the electropherograms corresponds to nucleotide incorporation (R = purine base) and the comparison of extension peak intensity to the intensity of non-elongated primer allows a qualitative evaluation of the primer extension reaction efficiencies. b) Screened characteristics of promising KTq variants: high 5 mC discrimination for dGMP incorporation and low to moderate efficiency for dAMP misincorporation opposite C and 5 mC. Displayed electropherograms were obtained during screening experiments and are exemplarily shown for anticipated results.

the modified template base. Guided by previous studies on DNA polymerase fidelity,[27] we reasoned that this decreased efficiency of the KTq variants to incorporate the matching nucleotide opposite 5 mC would instead increase the ratio of mismatch nucleotide incorporation. To promote this misincorporation, KTq variants were additionally screened for low to moderately increased dAMP misincorporation, but without discrimination, opposite C and 5 mC (Figure 1b). This should result in a catalytically active DNA polymerase variant that generates mutation signatures opposite 5 mC but processes canonical nucleotides without increased error rates. Thus, only KTq variants that misincorporated dAMP with similar efficiency opposite C and 5 mC, as well as elongated no more than 50 % of the primer by dAMP incorporation were selected, resulting in approximately 3 % of the initially applied KTq variants. Furthermore, only those KTq variants with the same or higher 5 mC discrimination efficiency compared to the KTq wild-type were selected in a next screening. In this final screening round, 12 promising KTq variants were identified as the most promising hits, representing 0.4 % of the mutant library (Figures S2 and S3). These KTq variants were derived from the combinatorial RACHITT mutant library and named according to their location in the library (Table S1). After gene expression and purification of the 12 most promising enzymes, the screened properties were verified in further single-nucleotide incorporation experiments (Figure S4).

In addition, we conducted multiple-nucleotide incorporation experiments to gain insight into the DNA polymerase elongation capability by adding 2′-deoxycytidine-5′-triphosphate (dCTP) as the second nucleotide for incorporation. The efficiency of the DNA polymerases to process correctly incorporated nucleotides was evaluated by primer elongation after dGMP incorporation and showed that the KTq variants discriminated against 5 mC by reduced elongation after dGMP incorporation opposite the methylated template base (Figure S5). However, the processing and elongation of an incorrectly incorporated nucleotide poses a challenge to DNA polymerases, leading to mismatch-induced stalling and contributing to overall replication fidelity.[25g,28] Therefore,

we focused on selecting DNA polymerase mutants that efficiently elongate mismatches. Monitoring primer elongation after dAMP misincorporation revealed that only the KTq variants RII G7, RIII H20, RIV A8 and RIV D15 were able to efficiently extend a mismatch (Figure S6). Testing the PCR efficiency and robustness of the DNA polymerases while having less dGTP (match nucleotide) available for the amplification reaction, confirmed that all KTq variants were PCR active and amplified the correct PCR product (Figure S7). Among the variants, RIII H20, RIV A8 and RIV D15 showed the highest PCR efficiency (Figure S7c). Considering this, only the KTq variants RIII H20, RIV A8 and RIV D15 combined the desired incorporation characteristics, namely 5 mC discrimination, increased misincorporation, sufficient mismatch extension capability and activity in DNA synthesis.

Next, the KTq variants RIII H20, RIV A8 and RIV D15 were evaluated for the generation of 5 mC-dependent mutation signatures during DNA synthesis (Figure 2a). To enhance the 5 mC-dependent signatures, linear amplification of an unmodified C and a CpG modified 5 mC template was performed in the presence of a reduced concentration of the matching nucleotide first (Figure S8). Reducing the dGTP concentration would increase the formation of mismatches opposite Cs since less of the matching nucleotide is available for incorporation.[29] In this case, we reasoned that error formation at methylated sites might be favored due to the reduced efficiency of the KTq variants to incorporate dGMP opposite 5 mC. Thus, an unbalanced dNTP pool promotes both dAMP misincorporation and 5 mC discrimination, and thereby facilitates specific 5 mC sensing. Subsequently, the obtained PCR products served as templates for NGS library preparation and were subjected to sequencing. A self-scripted KNIME[30] workflow was used for data analysis and unique molecular identifier (UMI)-based error calculation (Figure 2b).

Indeed, the KTq variant RIV A8 exhibited an up to twofold increased error rate at multiple methylated CpG sites, namely C24, C32 and C72 in the 5 mC template compared to the C template (Figure 3a, black arrows). Detailed analysis of the error rates for amplification from both templates revealed that the DNA polymerase preferentially incorporated mismatches opposite templating C bases, with an average error of 3.2 % opposite C compared to 0.27 % opposite all non-C bases. Furthermore, different sequence positions, including non-C bases, showed variable error rates. DNA polymerase fidelity heavily depends on the DNA sequence context and secondary structures.[31] Consequently, the characteristics of the engineered mismatch formation would be equally affected. Here, it is particularly important that RIV A8 processed both templates with comparable fidelity at similar positions. Therefore, it is even more striking that a significant difference in error was exclusively detected by comparing methylated and unmethylated CpG sites (Figure 3b left, black arrows). At multiple methylated CpG sites C24, C32 and C72, the RIV A8 variant featured an average 16.5-fold increase in error rate difference (Δerror rate) compared to the KTq wild-type enzyme, which featured a marginally increased Δerror rate opposite 5 mC only at position C24 and C32 (with a mean Δerror rate of 5.9 % at methylated CpG sites for RIV A8 and a mean Δerror rate of 0.36 % for the KTq wild-type at methylated CpG sites) (Figure S9). Remarkably, initial characterization by primer extension reactions revealed no significant differences in 5 mC discrimination capacity between the KTq wild-type and RIV A8 (Figure S10 and Table S2). The mutation signature analysis showed that RIV A8 is more error-prone in general but confirms that an increased dAMP misincorporation (detected as T base calls) led to the enhanced mismatch formation opposite 5 mC by RIV A8 (Figures 3b right and S11). Moreover, RIV A8 was able to reproduce the distinctive 5 mC-dependent error signature with a similar outcome in a repetition experiment, thus making it a suitable candidate for detecting 5 mC by increased misincorporation (Figure S12). Also the mutants RIII H20 and RIV D15 exhibited increased dAMP misincorporation opposite 5 mC, although with lower efficiency compared to RIV A8 (Figures S13 and S14).

We were delighted to find that the RIV A8 variant also senses methylated CpG sites in human genomic DNA (gDNA) by generating site-specific 5 mC-dependent signatures (Figure 3d and 3e, black arrows). Due to the inherently low copy number of gDNA, only small amounts of linear PCR product were obtained after the reaction with the RIV A8 variant. Therefore, an exponential amplification by a high-fidelity DNA polymerase followed to generate the concentration required for NGS library preparation. This resulted in a mixture of starting material and RIV A8 product (Figure S15), which required further data analysis. Based on the previous finding that RIV A8 processed identical positions in each template with similar fidelity, and that a significant error difference resulted only from methylation (Figure S14c), we standardized the absolute error rates into customized z-scores (Figure S16). Using the z-score values, calculation of the differences between the modified 5 mC and unmodified C template (Δz-score) confirmed the ability of RIV A8 to detect 5 mC in the artificial training template, displayed by an average 11.4-fold increase in Δz-score at the methylated CpG sites C24, C32 and C72 compared to unmethylated C positions in the 5 mC template (Figures 3c and S17a). Despite the challenging and complex nature of amplification from gDNA, RIV A8 was able to detect methylation levels greater than 50 % at CpG sites C24 and C32 in the natively methylated gDNA (gDNA native) (Figure S8). This was indicated by a 9.5-fold increase of average Δz-score opposite these CpG sites in comparison to C bases (Figures 3d and S17b). Furthermore, by reading an increased misincorporation at the various CpG sites C24, C32 and C72 in the enzymatically CpG methylated gDNA template (gDNA mCpG) RIV A8 confirmed DNA methylation sensing with a 10.7-fold increase in average Δz-score opposite the CpG sites compared to C bases (Figures 3e, S8 and S17c).

In recent years, evidence has emerged that the 5 mC oxidation product, 5-hydroxymethylcytosine (5 hmC), serves not only as an intermediate in the 5 mC demethylation pathway, but also as a stable epigenetic mark in its own

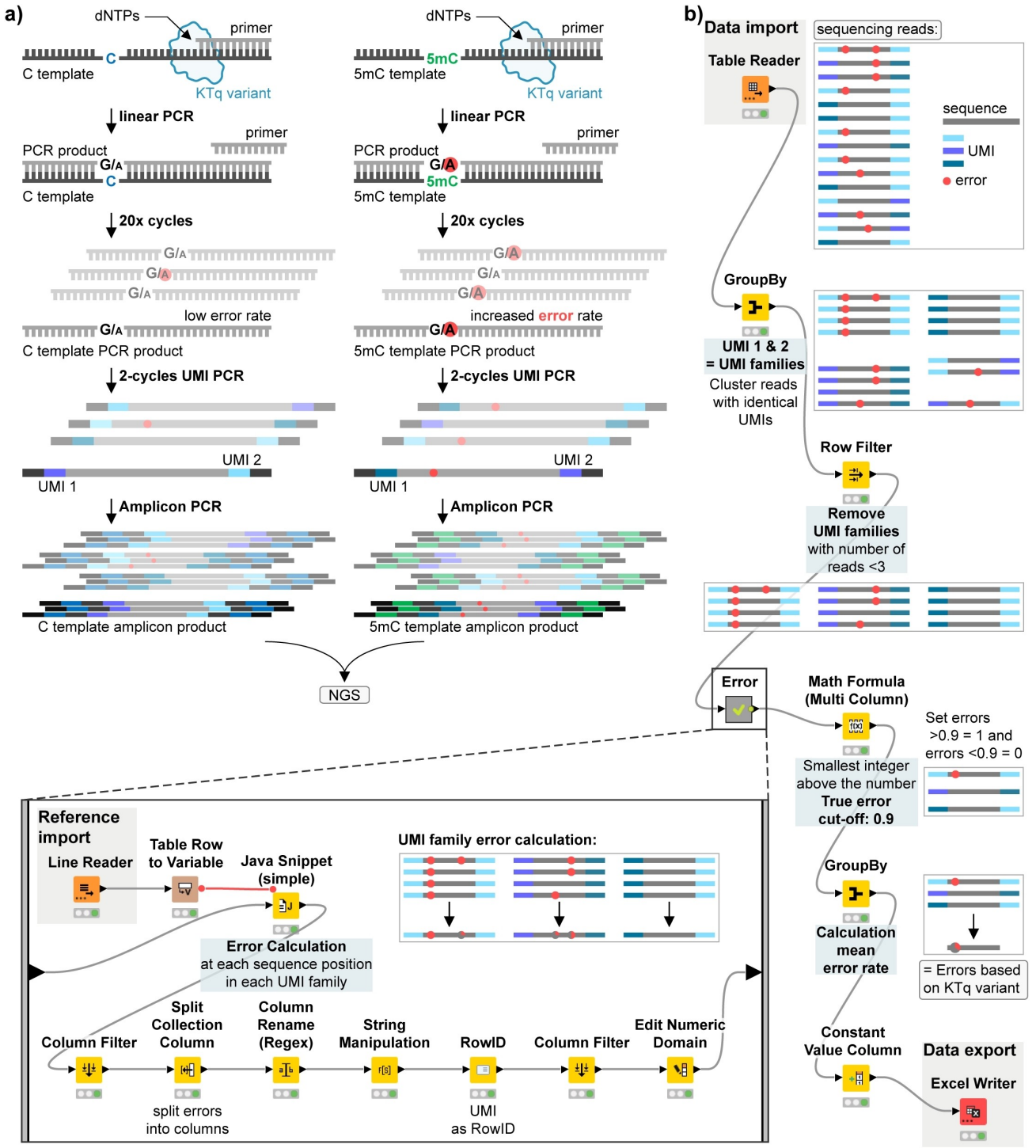


Figure 2. Strategy for the detection of 5 mC by reading an increased error rate. a) KTq variants were used for primer elongation opposite C and 5 mC in a linear PCR. The increased error rate opposite 5 mC (indicated by dots) derives from a selective mismatch formation during DNA synthesis by dAMP misincorporation. Methylation information is represented as a mutation, that is a mispaired A base, in the product DNA. The PCR product functions as template for the labeling with unique molecular identifier (UMI) sequences (sequencing primer binding site in dark grey and one color per UMI). Errors are preserved during amplification in the amplicon PCR (sequencing P5 and P7 adapter in black, indexes in blue or green) and libraries were analyzed by NGS. b) KNIME data analysis workflow used for error calculation. Errors based on the misincorporation by the KTq variant can be distinguished from sequencing errors by using the depicted UMI strategy. Reads were sorted by identical UMI sequences into UMI family groups (one UMI family derived from one linear PCR product) and UMI families with a minimum of three reads were further processed. Reads were aligned to the reference sequence and the error rate at each position was calculated within each family. Next, only errors which are present in 90% of all reads within one family (true error cut-off: 0.9), were considered for calculating the mean error rate over all UMI families. For this, the true UMI family errors were set to 1 and the calculated mean error rate represents the error derived from the KTq variant. 5 mC readout is facilitated by reading an increased error rate opposite 5 mC positions.

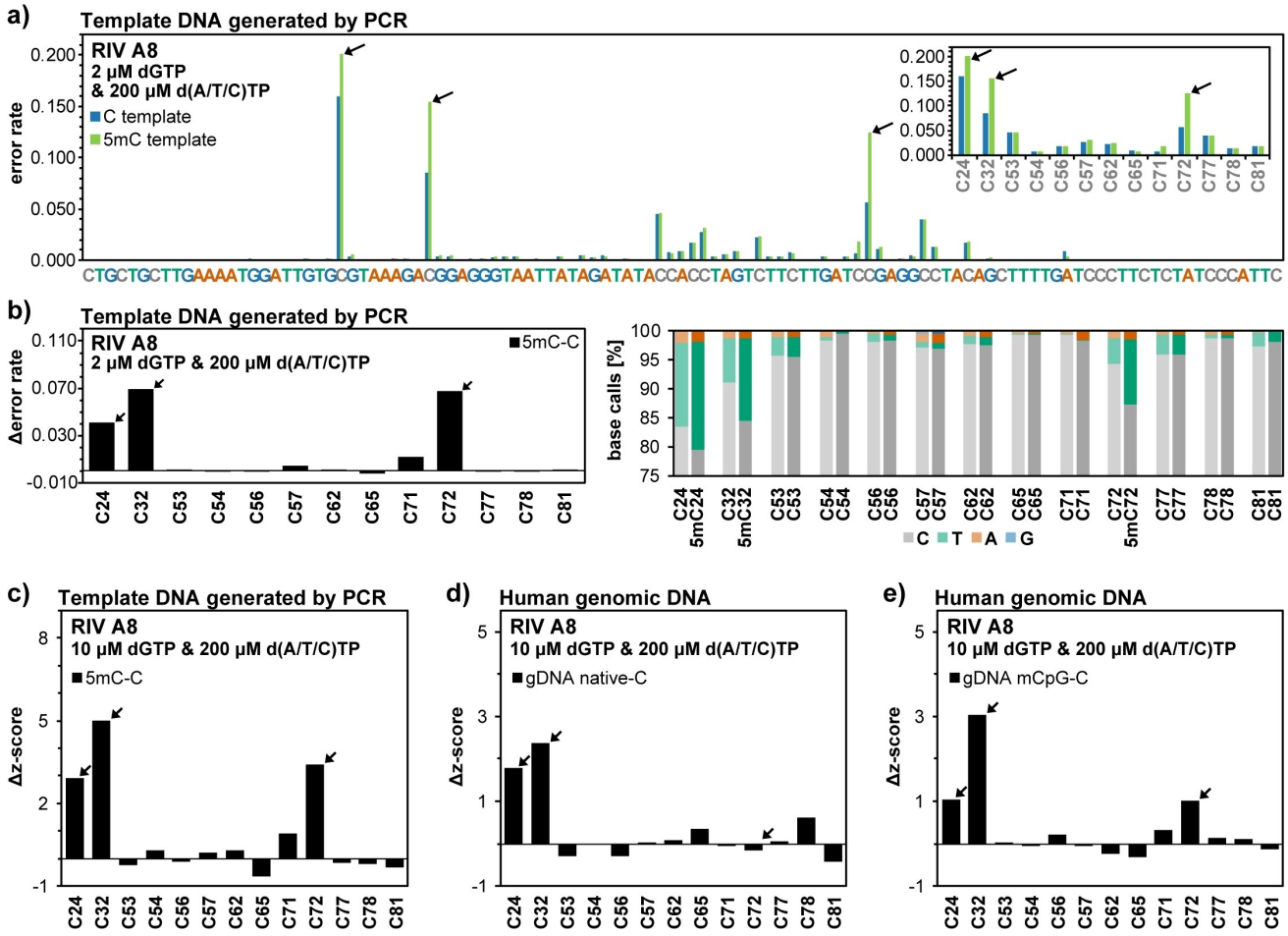


Figure 3. KTq variant RIV A8 senses 5 mC by increased misincorporation. a) Error rates of linear PCR products generated by the RIV A8 variant in presence of an unbalanced dNTP pool with 2 µM dGTP (match) and 200 µM d(A/T/C)TP (each) by amplifying either the unmodified C or modified 5 mC template (50 pM). b) Error rate difference (Δerror rate) (left) and mutation signature (right) at C and 5 mC positions of RIV A8. Error rate differences were calculated by subtracting error rates obtained by amplifying the C template from error rates obtained by amplifying the 5 mC template. c) Z-score difference (Δz-score) at C and CpG positions calculated by subtracting z-scores of standardized error rates of the C template from z-scores of standardized error rates of the 5 mC template (3 fM). Z-scores were calculated by dividing the difference between error rate and the mean error rate of exclusively C base positions by the standard deviation of the latter. Error rates are from linear PCR products received from RIV A8 variant in presence of an unbalanced dNTP pool with 10 µM dGTP (match) and 200 µM d(A/T/C)TP (each). d) Z-score difference at C and CpG positions calculated by subtracting z-scores of standardized error rates of the C template from z-scores of standardized error rates of the natively methylated gDNA (gDNA native) (62.5 ng gDNA). RIV A8 detects CpG methylation at C24 and C32 by increased misincorporation opposite 5 mC in native gDNA. At CpG site C72 (only minor methylation) no increased misincorporation can be detected. e) Z-score difference at C and CpG positions calculated by subtracting z-scores of standardized error rates of the C template from z-scores of standardized error rates of the CpG methylated gDNA (gDNA mCpG). RIV A8 detects methylation at CpG sites C24, C32 and C72 by increased misincorporation opposite 5 mC in mCpG gDNA. All CpG sites are indicated by black arrows. Error rates derived from NGS libraries which were prepared and sequenced once using the Illumina NextSeq 2000 system.

right.[32] The evaluation of RIV A8 for sensing DNA hydroxymethylation demonstrated that the variant discriminates 5hmC from unmodified C bases by exhibiting similar altered incorporation characteristics to those shown for 5 mC (Figures S18–S20).

The mechanism by which the multi-mutant variant RIV A8 discriminates 5 mC and exhibits increased dAMP misincorporation is currently unclear. However, as no single or rationally designed variant has the required characteristics, we speculate that the individual effects of the distal mutations contribute to a synergistic alteration of DNA polymerase fidelity (Figure S21). To design a DNA polymerase that efficiently detects 5 mC and still retains catalytic activity, it was necessary to overcome critical features that contribute to replication fidelity, such as reliable 5 mC processing, substrate selectivity and mismatch discrimination. The mutations can act at different mechanistic levels. Substitutions close to the active site, such as E507A, A570K and I614M, may discriminate against correct incorporation opposite methylated template bases,[25c,33] and these mutations may promote misincorporation by maintaining the correct orientation of residues when encountering mismatches, thus facilitating incorporation and elongation of incorrect nucleotides.[25d,34] Combined, this could result in

reduced catalytic efficiency for dGMP incorporation opposite 5 mC and thus enhance tolerance for mismatch formation at this site. Additionally, mutated residues N483K, S515K, K540N and V586G, which contact the primer and/or template strand, could enhance DNA binding and stabilize transition states, promote mismatch elongation and improve DNA synthesis.[25d,35]

In conclusion, we report here a new approach for 5 mC detection in highly methylated DNA by exploiting the properties of the discovered KTq variant RIV A8.

Supporting Information

The authors have cited additional references within the Supporting Information.[36,37]

Acknowledgements

MH and AF gratefully acknowledge support by the Konstanz Research School Chemical Biology. We thank Martina Adam-Wels for technical assistance in cloning and screening of the mutant libraries. This work has been partly funded by the Deutsche Forschungsgemeinschaft MA 2288/22-1. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

MH and AM are inventors of a PTC application made by the University of Konstanz based on the most relevant discoveries of this research work.

Data Availability Statement

High-throughput sequencing data are available in the NCBI GEO database, record GSE233599 and GSE270850. The KNIME workflow group created for processing of the sequencing data is available at https://hub.knime.com/-/spaces/-/~qkjAV-djKlP3LTTv/. All data needed to evaluate the conclusions in the paper are present in the paper and/or in the Supporting Information. Crystal structure information is available at the RCSB Protein Data Bank under the accession number 1QSS. Additional data may be requested from the authors.

Keywords: 5-methylcytosine · DNA polymerase · protein engineering · DNA methylation · next-generation sequencing

[1] A. Breiling, F. Lyko, *Epigenet. Chromatin* **2015**, *8*, 24.
[2] a) H. Leonhardt, A. W. Page, H.-U. Weier, T. H. Bestor, *Cell* **1992**, *71*, 865–873; b) M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247–257; c) E. Li, Y. Zhang, *Cold Spring Harbor Perspect. Biol.* **2014**, *6*.
[3] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315–322.
[4] C. Luo, P. Hajkova, J. R. Ecker, *Science* **2018**, *361*, 1336–1340.
[5] G. Csankovszki, A. Nagy, R. Jaenisch, *J. Cell Biol.* **2001**, *153*, 773–784.
[6] E. Li, C. Beard, R. Jaenisch, *Nature* **1993**, *366*, 362–365.
[7] Y. Atlasi, H. G. Stunnenberg, *Nat. Rev. Genet.* **2017**, *18*, 643–658.
[8] a) M. V. C. Greenberg, D. Bourc'his, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590–607; b) M. Ehrlich, *Epigenetics* **2019**, *14*, 1141–1163.
[9] a) M. Esteller, *Oncogene* **2002**, *21*, 5427–5440; b) A. S. Wilson, B. E. Power, P. L. Molloy, *Biochim. Biophys. Acta Rev. Cancer* **2007**, *1775*, 138–162; c) W. J. Locke, D. Guanzon, C. Ma, Y. J. Liew, K. R. Duesing, K. Y. C. Fung, J. P. Ross, *Front. Genet.* **2019**, *10*; d) A. Nishiyama, M. Nakanishi, *Trends Genet.* **2021**, *37*, 1012–1027.
[10] a) K. Bebenek, T. A. Kunkel, in *Adv. Protein Chem., Vol. 69*, Academic Press, **2004**, pp. 137–165; b) J. Aschenbrenner, A. Marx, *Curr. Opin. Biotechnol.* **2017**, *48*, 187–195.
[11] a) J.-C. Shen, S. Creighton, P. A. Jones, M. F. Goodman, *Nucleic Acids Res.* **1992**, *20*, 5119–5125; b) M. J. Howard, K. G. Foley, D. D. Shock, V. K. Batra, S. H. Wilson, *J. Biol. Chem.* **2019**, *294*, 7194–7201.
[12] a) A. C. Dantas Machado, T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, H. J. Bussemaker, R. Rohs, *Briefings Funct. Genomics* **2015**, *14*, 61–73; b) E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Nat. Chem. Rev.* **2017**, *1*, 0069.
[13] a) M. Welter, D. Verga, A. Marx, *Angew. Chem. Int. Ed.* **2016**, *55*, 10131–10135; b) J. Balintová, M. Welter, A. Marx, *Chem. Sci.* **2018**, *9*, 7122–7125.
[14] a) R. Shapiro, V. DeFate, M. Welcher, *J. Am. Chem. Soc.* **1974**, *96*, 906–912; b) 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; c) H. Hayatsu, *Mutat. Res. Rev. Mutat. Res.* **2008**, *659*, 77–82.
[15] R. P. Darst, C. E. Pardo, L. Ai, K. D. Brown, M. P. Kladde, *Curr. Protoc. Mol. Biol.* **2010**, *91*, 7.9.1–7.9.17.
[16] a) C. Grunau, S. J. Clark, A. Rosenthal, *Nucleic Acids Res.* **2001**, *29*, e65–e65; b) D. P. Genereux, W. C. Johnson, A. F. Burden, R. Stöger, C. D. Laird, *Nucleic Acids Res.* **2008**, *36*, e150–e150; c) N. Olova, F. Krueger, S. Andrews, D. Oxley, R. V. Berrens, M. R. Branco, W. Reik, *Genome Biol.* **2018**, *19*, 33.
[17] a) Y. Xi, W. Li, *BMC Bioinf.* **2009**, *10*, 232; b) L. Ji, T. Sasaki, X. Sun, P. Ma, Z. A. Lewis, R. J. Schmitz, *Front. Genet.* **2014**, *5*.
[18] K. Tanaka, A. Okamoto, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912–1915.
[19] a) Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C.-X. Song, *Nat. Biotechnol.* **2019**, *37*, 424–429; b) R. Vaisvila, V. K. C. Ponnaluri, Z. Sun, B. W. Langhorst, L. Saleh, S. Guan, N. Dai, M. A. Campbell, B. S. Sexton, K. Marks, M. Samaranayake, J. C. Samuelson, H. E. Church, E. Tamanaha, I. R. Corrêa, S. Pradhan, E. T. Dimalanta, T. C. Evans, L. Williams, T. B. Davis, *Genome Res.* **2021**, *31*, 1280–1289.
[20] a) L.-Y. Zhao, J. Song, Y. Liu, C.-X. Song, C. Yi, *Protein Cell* **2020**, *11*, 792–808; b) Y. Kong, E. A. Mead, G. Fang, *Nat. Rev. Genet.* **2023**, *24*, 363–381.
[21] A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, J. H. Gundlach, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 18904–18909.

[22] a) J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, W. Timp, *Nat. Methods* **2017**, *14*, 407–410; b) Y. Wang, Y. Zhao, A. Bollas, Y. Wang, K. F. Au, *Nat. Biotechnol.* **2021**, *39*, 1348–1365; c) M. C. Lucas, E. M. Novoa, *Nat. Methods* **2023**, *20*, 25–29.

[23] a) S. L. Amarasinghe, S. Su, X. Dong, L. Zappia, M. E. Ritchie, Q. Gouil, *Genome Biol.* **2020**, *21*, 30; b) Z. W.-S. Yuen, A. Srivastava, R. Daniel, D. McNevin, C. Jack, E. Eyras, *Nat. Commun.* **2021**, *12*, 3438; c) Y. Liu, W. Rosikiewicz, Z. Pan, N. Jillette, P. Wang, A. Taghbalout, J. Foox, C. Mason, M. Carroll, A. Cheng, S. Li, *Genome Biol.* **2021**, *22*, 295.

[24] W. M. Barnes, *Gene* **1992**, *112*, 29–35.

[25] a) M. Suzuki, A. K. Avicola, L. Hood, L. A. Loeb, *J. Biol. Chem.* **1997**, *272*, 11228–11235; b) M. Suzuki, S. Yoshida, E. T. Adman, A. Blank, L. A. Loeb, *J. Biol. Chem.* **2000**, *275*, 32728–32735; c) P. H. Patel, L. A. Loeb, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5095–5100; d) E. Loh, L. A. Loeb, *DNA Repair* **2005**, *4*, 1390–1398; e) E. Loh, J. Choe, L. A. Loeb, *J. Biol. Chem.* **2007**, *282*, 12201–12209; f) W. Zheng, B. R. Brooks, S. Doniach, D. Thirumalai, *Structure* **2005**, *13*, 565–577; g) J. Aschenbrenner, S. Werner, V. Marchand, M. Adam, Y. Motorin, M. Helm, A. Marx, *Angew. Chem. Int. Ed.* **2018**, *57*, 417–421; h) G. Raghunathan, A. Marx, *Sci. Rep.* **2019**, *9*, 590.

[26] W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires, D. J. Monticello, *Nat. Biotechnol.* **2001**, *19*, 354–359.

[27] a) W. A. Beard, D. D. Shock, B. J. Vande Berg, S. H. Wilson, *J. Biol. Chem.* **2002**, *277*, 47393–47398; b) W. A. Beard, S. H. Wilson, *Structure* **2003**, *11*, 489–496.

[28] a) M.-M. Huang, N. Arnheim, M. F. Goodman, *Nucleic Acids Res.* **1992**, *20*, 4567–4573; b) S. Ayyadevara, J. J. Thaden, R. J. Shmookler Reis, *Anal. Biochem.* **2000**, *284*, 11–18; c) L. B. Huber, N. Kaur, M. Henkel, V. Marchand, Y. Motorin, A. E. Ehrenhofer-Murray, A. Marx, *Nucleic Acids Res.* **2023**, *51*, 3971–3987.

[29] A. M. de Paz, T. R. Cybulski, A. H. Marblestone, B. M. Zamft, G. M. Church, E. S. Boyden, K. P. Kording, K. E. J. Tyo, *Nucleic Acids Res.* **2018**, *46*, e78–e78.

[30] A. Fillbrunn, C. Dietz, J. Pfeuffer, R. Rahn, G. A. Landrum, M. R. Berthold, *J. Biotechnol.* **2017**, *261*, 149–156.

[31] a) G. G. Hillebrand, K. L. Beattie, *J. Biol. Chem.* **1985**, *260*, 3116–3125; b) P. C. Loewen, J. Switala, *Gene* **1995**, *164*, 59–63.

[32] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935; c) X. Wu, Y. Zhang, *Nat. Rev. Genet.* **2017**, *18*, 517–534; d) D.-Q. Shi, I. Ali, J. Tang, W.-C. Yang, *Front. Genet.* **2017**, *8*.

[33] a) B. Arezi, N. McKinney, C. Hansen, M. Cayouette, J. Fox, K. Chen, J. Lapira, S. Hamilton, H. Hogrefe, *Front. Microbiol.* **2014**, *5*; b) Y. Lim, I.-H. Park, H.-H. Lee, K. Baek, B.-C. Lee, G. Cho, *J. Mol. Diagn.* **2022**, *24*, 1128–1142.

[34] a) P. H. Patel, L. A. Loeb, *J. Biol. Chem.* **2000**, *275*, 40266–40272; b) Y.-C. Tsai, K. A. Johnson, *Biochemistry* **2006**, *45*, 9675–9687; c) J. L. Ong, D. Loakes, S. Jaroslawski, K. Too, P. Holliger, *J. Mol. Biol.* **2006**, *361*, 537–550.

[35] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4552–4557.

[36] a) Y. Li, V. Mitaxov, G. Waksman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9491–9496; b) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225–1233.

[37] N. Eriksson, S. Wu, C. B. Do, A. K. Kiefer, J. Y. Tung, J. L. Mountain, D. A. Hinds, U. Francke, *Flavour* **2012**, *1*, 22.

Manuscript received: July 15, 2024
Version of record online: October 24, 2024