RNA Methylation

Engineering of a DNA Polymerase for Direct m6A Sequencing

Joos Aschenbrenner*, Stephan Werner*, Virginie Marchand, Martina Adam, Yuri Motorin, Mark Helm, and Andreas Marx*

Abstract: Methods for the detection of RNA modifications are of fundamental importance for advancing epitranscriptomics. N6-methyladenosine (m6A) is the most abundant RNA modification in mammalian mRNA and is involved in the regulation of gene expression. Current detection techniques are laborious and rely on antibody-based enrichment of m6A-containing RNA prior to sequencing, since m6A modifications are generally “erased” during reverse transcription (RT). To overcome the drawbacks associated with indirect detection, we aimed to generate novel DNA polymerase variants for direct m6A sequencing. Therefore, we developed a screen to evolve an RT-active KlenTaq DNA polymerase variant that sets a mark for N6-methylation. We identified a mutant that exhibits increased misincorporation opposite m6A compared to unmodified A. Application of the generated DNA polymerase in next-generation sequencing allowed the identification of m6A sites directly from the sequencing data of untreated RNA samples.

Cellular RNAs are posttranscriptionally modified through the enzymatic introduction of more than 150 modifications.[6] The research field of epitranscriptomics aims to investigate the role of these modifications, which possess functional importance but do not alter the RNA sequence itself.[2] Therefore, reliable and straightforward methods to detect modifications in a transcriptome-wide manner are required. However, while nucleic acid analysis in general has profited tremendously from the rise of next-generation sequencing (NGS) technologies,[3], the enormous potential of these techniques has so far only rarely been adapted for the direct analysis of modified nucleotides. This is because modifications in an RNA template strand that do not alter the sequence are “erased” during reverse transcription (RT). Modifications located at the Watson–Crick face of the nucleobase constitute an exception to this rule since they affect RT, resulting in the appearance of “RT signatures” at modification sites.[4] These signatures arise from increased incorporation of mismatched nucleotides and/or accumulated rates of RT abortion at modification sites. On this basis, direct prediction of N6-methyladenosine (m6A) sites from NGS sequencing data has been realized.[5] This approach is, however, restricted to modifications that interfere with correct Watson–Crick base pairing. To overcome this limitation, we aimed to evolve a novel RT system that introduces signatures opposite a normally RT-silent modification.

N6-methyladenosine (m6A) was chosen as target modification because it is a reversible[6] and highly abundant[7] modification in mammalian mRNA. m6A modification of cellular RNA has been demonstrated to affect mRNA splicing,[8] nuclear export,[9, 10] translation,[10] and degradation.[11] Proposed functions include the generation of “translational pulses”,[12] the control of the circadian clock,[9] the initiation of the DNA damage response,[13] and the clearance of maternal RNA[14] and pluripotency factors.[15] Furthermore, m6A modification can also be found in other cellular RNAs, including rRNA, tRNA, and IncRNA.[16, 17] Current methods to map m6A typically employ immunoprecipitation of m6A-specific antibodies and covalent crosslinking of the antibody to the RNA molecule prior to analysis by NGS.[16] These methods are complex and laborious and the results may suffer from artifacts deriving from poor antibody specificity and cross-reactivity.[17] Therefore, novel m6A detection systems might benefit from reverse transcriptases (RTases) that sense m6A during cDNA synthesis and create a signal that can be passed on during PCR. The fact that certain RT-active DNA polymerases are capable of discriminating m6A from unmodified A has been shown by a previous study.[18] We were able to advance this feature by engineering an RT-active DNA polymerase that features significantly increased error rates opposite m6A but not unmodified A. The enzyme was evolved from a thermostable KlenTaq DNA polymerase variant with RT activity (KlenTaq L459M S515R I638F M747K, henceforth referred to as RT-KTQ).[19]

As a first step, the incorporation of complementary and non-complementary nucleotides opposite m6A and unmodified A by RT-KTQ was investigated. Single-nucleotide incorporation was performed with each of the four dNTPs, employing a primer hybridized to two different RNA oligonucleotides of the same sequence carrying either A or m6A at the site of first incorporation (Figure S1 in the

[6] J. Aschenbrenner, A. Adam, Prof. Dr. A. Marx
Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz
Universitätstraße 10, 78457 Konstanz (Germany)
E-mail: andreas.marx@uni-konstanz.de
S. Werner, Prof. Dr. M. Helm
Institute of Pharmacy and Biochemistry
Johannes Gutenberg University Mainz
Staudingerweg 5, 55128 Mainz (Germany)
Dr. V. Marchand, Prof. Dr. Y. Motorin
Laboratoire Ingénierie Moléculaire et Physiopathologie Articulaire
IMoPA, UMR7365 CNRS-UL, Biopôle de L’Université de Lorraine
9, Avenue de la Forêt de Haye, 54505 Vandoeuvre-les-Nancy (France)

[6a] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201710209.

© 2017 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 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, and is not used for commercial purposes.
As expected, misincorporation of dAMP, dCMP, and dGMP was considerably less efficient than incorporation of the complementary dTMP for both templates. Moreover, dTMP and dAMP incorporation differed only slightly between the A and m6A templates, whereas dCMP and dGMP incorporation was significantly hampered opposite m6A. Inspired by previous studies that utilized capillary electrophoresis (CE) to monitor the activity of DNA polymerases and other enzymes, we conceived an assay to screen for DNA polymerase variants with increased misincorporation opposite m6A. The screen involved extension of 5'-fluorophore-labeled primer strands through single-nucleotide incorporation, followed by CE. Multiplexed analysis of several primer extension reactions could be achieved by employing primers of different lengths labeled with different fluorophores (Figure 1a). More specifically, six primers were designed that possess the same 3'-terminal 20 nucleotides complementary to the RNA template but differ in their length due to varying 5'-overhangs (Table S1 in the Supporting Information). Additionally, primers were employed as 5'-FAM- and 5'-HEX-labeled variants. The devised assay was applied to screen a library composed of RT-KTQ single mutants created by site-directed mutagenesis. Mutation sites were selected based on their vicinity to the nascent base pair in a crystal structure of an RT-KTQ closed complex (Figure 1b). For each site, all 19 mutants were generated and expressed in E.coli in 96-well plates. Single-nucleotide incorporation was then performed with the RT-KTQ expression lysates after heat-denaturation of the E.coli host proteins.

It was reported that low-fidelity DNA polymerases differ from high-fidelity DNA polymerases mainly in the efficiency of correct nucleotide incorporation, whereas the incorporation of incorrect nucleotides is comparable. Thus, we reasoned that enhanced error rates opposite m6A would probably derive from decreased dTMP incorporation rather than from increased misincorporation. For this reason, the developed screening assay was employed to monitor dTMP incorporation opposite m6A and unmodified A. Here, we looked for variants with considerably decreased incorporation of dTMP opposite m6A but not A (Figure S2). Furthermore, to ensure that only incorporation of the correct nucleotide was reduced and not overall activity, dAMP incorporation was monitored in an additional screening (Figure S3). Evaluation of the screening data was performed by qualitative assessment of extension peaks in the electro-

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Screening for DNA polymerase variants with increased misincorporation rates opposite m6A. a) DNA polymerase expression lysates were applied to catalyze the incorporation of dTMP or dAMP opposite A and m6A. Utilization of primers with different length and fluorophores (FAM = 6-carboxyfluorescein; HEX = hexachlorofluorescein) enabled the joint analysis of 12 reaction mixtures in one capillary. b) Amino acids in proximity to the nascent base pair were chosen for saturation mutagenesis. Adapted from PDB ID: 4BWMI using PyMOL (Schrödinger, LLC; New York, NY). c) Anticipated outcome for promising RT-KTQ variants: high m6A discrimination for dTMP incorporation and high efficiency for dAMP misincorporation.
As at emplate, we employed the another modification that affected RT 2018. We tRNA extracts as at emplate for RT-KTQ G668Y Y671A features elevated error rates opposite another modified nucleotide (5-methyluridine, T) or another sites with error rates of more than 10% were located at the 5-end of the RNA molecule, where rates are inaccurate due to low coverage. The lower coverage derives from reduced activity of the enzyme (Table S3) and synthesis arrest on account of tRNA secondary structure and modifications, and it cannot be resolved decisively by altered reaction conditions. Another modification that affected RT by this enzyme was uridine-5-oxyacetic acid (V), which triggered high rates of RT arrest. In contrast, almost all unmodified A sites exhibited error rates below 5%. Only A5 constituted an exception, with an error rate of 8.5%. We assume that the increased error rate at this position might arise from the fact that this nucleotide is located directly adjacent to the RT-affecting uridine-5-oxyacetic acid. The m'A signature was not observed when unmodified RT-KTQ was applied (Figure S6). When comparing the two analyzed m'A sites, misincorporation and arrest patterns vary due to different sequence contexts, as has been observed previously for m'A signatures.

We aimed to investigate why the engineered RT-KTQ G668Y Y671A double mutant exhibited an elevated error rate opposite m'A (and an increased amount of T-reads in particular), whereas other mutations identified by the initial screening did not show this effect. Therefore, we determined...
the incorporation rates of dAMP and dTMP opposite A and m^A^ at a given dNTP concentration of 100 μM (Table S3, Figure S7,S8). For unmodified RT-KTQ, the ratio of dTMP to dAMP incorporation rate was similar for A and m^A^.

For RT-KTQ Y671A, dAMP misincorporation rates were comparable to the unmodified RT-KTQ. However, dTMP incorporation was significantly reduced opposite m^A^, whereas it decreased only slightly opposite A. A similar effect was observed for the G668Y mutation. It was necessary to combine both mutations in an RT-KTQ double mutant to attain an increased amount of T-reads at the m^A^ site. Correspondingly, RT-KTQ G668Y Y671A featured an even further reduced incorporation rate of dTMP opposite m^A^.

For this enzyme, dTMP incorporation opposite m^A^ was only 1.6 times faster than dAMP misincorporation. In contrast, mutation of residues I614 and G672 did not result in elevated error rates opposite m^A^.

Whereas the G672H mutation delivered the most prominent discrimination of m^A^ during dTMP incorporation, it also hampered the misincorporation of dAMP tremendously. RT-KTQ I614A featured significantly increased rates of dAMP misincorporation but lost m^A^ discrimination.

In this study we provide a novel engineering strategy to create reverse transcriptase variants exhibiting RT signatures as a response to encountering a specific RNA modification. The strategy to evolve an "m^A^-sensing" RT-active DNA polymerase involved the generation of DNA polymerase libraries in combination with a primer-extension-based screening assay. Notably, the assay should also be suitable for other modifications and the throughput of the assay could be increased for future projects by employing more primers of different length, a greater variety of 5'-fluorophores, and/or several orthogonal primer/template sequences. Qualitative examination of the screening data for variants with increased m^A^ discrimination during dTMP incorporation but with unaffected dAMP misincorporation delivered promising mutants. Interestingly, the identified Y671 residue is located directly at the C-terminal end of the O-helix and is known to undergo substantial conformational changes upon dNTP binding, thereby playing an important role in the selectivity of KlenTaq DNA polymerase and the homologous KF DNA polymerase.[24]

We have been able to show that the engineered RT-KTQ G668Y Y671A delivers prominent RT signatures at m^A^ sites in different sequence contexts, without exerting elevated error rates opposite unmodified nucleotides and the majority of the other modified nucleotides present in the E.coli tRNA Val. Only uridine-5-oxyacetic acid and 5-methyluridine resulted in the emergence of high arrest rates and increased dGMP misincorporation, respectively. However, these RT signatures are highly characteristic, which might enable their distinction from (and detection simultaneously to) m^A^. RT-KTQ G668Y Y671A could contribute to the development of new sequencing approaches to map m^A^ sites in cellular RNA in the future. Here, a technology that is orthogonal to the present antibody-enrichment methodologies (MeRIP)[7a,b,16] is desperately needed to validate candidate sites[17] and to simplify detection procedures. The development of such assays necessitates algorithms to identify m^A^-specific RT signatures and to distinguish them from signals deriving from other sources. As already implemented for the detection of m^A^, machine-learning-based algorithms can be trained to predict modification sites when fed with sufficient data for modification-specific RT signatures.[5] For this purpose, sequencing data from modified RNA oligonucleotides and/or validated m^A^ sites in tRNA, mRNA, or lncRNA could be utilized to generate training data sets.[18,25]

Acknowledgements

We acknowledge support by the Deutsche Forschungsgemeinschaft (SPP 1784), the European Research Council (ERC Advanced Grant 339834), the Carl Zeiss Stiftung (stipend to J.A.), the ANR HTRNAmod (ANR-13-ISV8- 0001/HE 3397/8-1; grant to Y.M.), and the Konstanz Research School Chemical Biology. We further thank the Next-Generation Sequencing Core Facility, FR3209 BMCT CNRS-UL, Lorraine University, Vandoeuvre-les-Nancy (France) for sequencing services.

Conflict of interest

The authors declare no conflict of interest.
