Published online 27 March 2023

Nucleic Acids Research, 2023, Vol. 51, No. 8 ***3971–3987***
https://doi.org/10.1093/nar/gkad177

A dual-purpose polymerase engineered for direct sequencing of pseudouridine and queuosine

Luisa B. Huber[1,†], Navpreet Kaur[2,†], Melanie Henkel[1], Virginie Marchand[3], Yuri Motorin[4], Ann E. Ehrenhofer-Murray[2,*] and Andreas Marx[1,*]

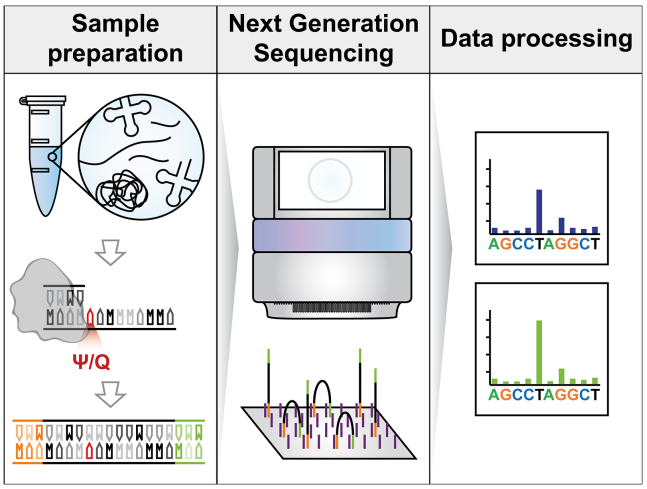
[1]Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstraße 10, 78464Konstanz, Germany, [2]Institut für Biologie, Humboldt-Universität zu Berlin, Philippstraße 13, Rhoda-Erdmann-Haus, 10099 Berlin, Germany, [3]Epitranscriptomics and RNA Sequencing Core Facility, Université de Lorraine, CNRS, INSERM, IBSLor (UAR2008/US40), F-54000Nancy, France and [4]Université de Lorraine, CNRS, IMoPA (UMR7365), F-54000 Nancy, France

Received November 09, 2022; Revised February 17, 2023; Editorial Decision February 22, 2023; Accepted February 24, 2023

ABSTRACT

More than 170 posttranscriptional RNA modifications are so far known on both coding and noncoding RNA species. Within this group, pseudouridine (Ψ) and queuosine (Q) represent conserved RNA modifications with fundamental functional roles in regulating translation. Current detection methods of these modifications, which both are reverse transcription (RT)-silent, are mostly based on the chemical treatment of RNA prior to analysis. To overcome the drawbacks associated with indirect detection strategies, we have engineered an RT-active DNA polymerase variant called RT-KTq I614Y that produces error RT signatures specific for Ψ or Q without prior chemical treatment of the RNA samples. Combining this polymerase with next-generation sequencing techniques allows the direct identification of Ψ and Q sites of untreated RNA samples using a single enzymatic tool.

GRAPHICAL ABSTRACT



INTRODUCTION

Posttranscriptional modifications on RNA affect nearly all aspects of the RNA life cycle in the cell, generating an additional layer in the regulation of gene expression (1). Although more and more RNA modifications are being discovered (2,3), many detection methodologies still bear limitations and thus hamper fast and straightforward monitoring (4,5). In addition to detection approaches based on specific nucleotide labeling (6,7), induced cleavage of ribose–phosphate backbone (8,9) or liquid chromatography coupled to mass spectrometry (LC–MS) (10), next-generation sequencing (NGS) has proven to be a powerful and widely applicable tool (11). Many direct detection methods rely on the location of modifying groups at the Watson–Crick face of the nucleotide (12). While processing these noncanonical nucleotides, increased misincorporation and/or elevated

*To whom correspondence should be addressed. Tel: +49 30 2093 49630; Fax: +49 30 2093 49641; Email: ann.ehrenhofer-murray@hu-berlin.de
Correspondence may also be addressed to Andreas Marx. Tel: +49 7531 885139; Fax: +49 7531 885140; Email: andreas.marx@uni-konstanz.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

reverse transcription (RT) stops are observed, which lead to unique signatures at modification sites (13–16). This strategy is, however, restricted to those modifications that alter Watson–Crick base pairing during nucleotide incorporation. RNA modifications that do not affect RT are erased during amplification from the RNA source and are therefore more difficult to map. In this case, most NGS-based approaches take advantage of specific antibodies or chemicals for RNA recognition and modification in order to introduce bulky moieties at the nucleotide of interest (17,18). Therefore, RT signatures are generated during reverse transcription.

Two examples of RT-silent RNA modifications are pseudouridine (Ψ) and queuosine (Q). Ψ was the first RNA modification to be discovered and is also the most abundant internal modification in the cellular transcriptome (19,20). To detect Ψ, target RNA molecules are treated with *N*-cyclohexyl-*N*′-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT) prior to the RT reaction (21–23). CMCT reacts with residues of uridine, guanine and Ψ. After alkaline treatment at pH 10.4, the adduct remains specifically at the *N*3 position of Ψ, and thus, stable Ψ adducts are formed. This bulky moiety is not RT-silent and leads to increased misincorporation and/or higher RT abortion at Ψ sites (24). Moreover, treatment of RNA with hydrazine/aniline or bisulfite was recently coupled to NGS Ψ-profiling methods (5,25,26).

Queuosine [the respective base is termed queuine (q)] is a hypermodified 7-deazaguanine derivative with an aminomethyl side chain and an attached cyclopentenediol moiety. It is located at the wobble position 34 of tRNAs with a GUN anticodon, namely $tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{His}$ and $tRNA^{Tyr}$ (27,28). Q modification of these tRNAs helps the cells decode the respective C- and U-ending Q codons, which ensures smooth translation across the translatome and prevents mistranslation, and the absence of Q modification in eukaryotes causes protein misfolding and endoplasmic reticulum stress (29,30). Of note, while Q is present in both prokaryotes and eukaryotes, it can only be biosynthesized by eubacteria (27). Eukaryotes rely on external sources (food, gut microbiome) to uptake and metabolize Q and q (31), thus providing a means for a nutritional factor to control translation in these organisms (28–30). Current detection methods for Q are mostly based on radioactive guanine exchange (32), LC–MS (33) or boronic acid-based gel electrophoresis methods (34,35). Since Q-modified nucleotides are processed by conventional DNA polymerases such as guanine during RT, the profiling of Q with sequencing strategies is only possible with prior periodate treatment (36). Oxidation of the q base results in short deletions in the cDNA directly opposite the modification site and thus can be detected by NGS.

Indirect mapping strategies have several general drawbacks such as false negative and false positive discovery rates due to incomplete chemical conversion or partial CMC cleavage of uridine or guanine residues (37). Furthermore, harsh reaction conditions are required in order to achieve reaction turnover, which can affect the integrity of the RNA sample (38).

In a previous study, we identified RT-active KlenTaq DNA polymerase (RT-KTq) mutants that are able to discriminate methylated nucleotides such as 2′-*O*-methyladenosine, 2′-*O*-methylcytidine and N^6-methyladenosine (m^6A) (39,40). In this study, we present the development of a novel enzymatic tool, a new variant of the KlenTaq DNA polymerase, for the detection of both Ψ and Q. This enzyme causes increased misincorporation of G opposite Ψ and of T opposite Q. Error rates are further enhanced by lowering the respective dNTP levels. We used this enzyme to identify Ψ on synthetic RNA molecules and on tRNAs *ex cellulo* in the yeast *Saccharomyces cerevisiae*. We furthermore identified Q modification on *in vitro*-transcribed and Q-modified tRNAs as well as on Q-tRNAs *ex cellulo* from *Schizosaccharomyces pombe* and *Shigella flexneri*. Altogether, this enzyme variant constitutes a dual-purpose enzymatic tool that can variably be employed to measure pseudouridylation or queuosinylation using a single molecular tool.

MATERIALS AND METHODS

Oligonucleotides

DNA oligonucleotides were ordered from Biomers in HPLC grade and directly used in screening assays and primer extension reactions that were analyzed by capillary electrophoresis (CE). For primer extension reactions analyzed by denaturing PAGE, oligonucleotides were purified by preparative PAGE prior to usage. Thereafter, radioactive labeling with [γ-^{32}P]-ATP and T4 polynucleotide kinase (NEB) was performed following the manufacturer's protocol. To detect Q modification, DNA oligonucleotides purchased from Metabion were used for reverse transcription and amplification of the respective tRNAs. RNA oligonucleotides were purchased PAGE purified from Purimex/IDT and were directly used. The concentration of the used oligonucleotides was determined by measuring the absorbance at 260 nm and applying the Lambert–Beer law. DNA or RNA sequences of all oligonucleotides are listed in Supplementary Table S1.

Saccharomyces cerevisiae growth condition

The *S. cerevisiae* strains used in this study were purchased from the EUROSCARF collection (Germany) and are listed in Supplementary Table S2. They were grown in standard YPD medium (20 g/l peptone, 20 g/l dextrose, 10 g/l yeast extract) at 30°C to exponential phase (0.6–0.7 OD_{600}).

Schizosaccharomyces pombe growth condition

The *S. pombe* strain used in this study is listed in Supplementary Table S3. Cells were cultured at 30°C in YES (5 g/l yeast extract, 30 g/l glucose, 250 mg/l adenine, 250 mg/l histidine, 250 mg/l leucine, 250 mg/l uracil, 250 mg/l lysine), which does not contain queuine. Synthetic queuine (kindly provided by Hans-Dieter Gerber and Gerhard Klebe (Universität Marburg)) was added to 0.1 μM to the culture (41).

Shigella flexneri growth condition

S. flexneri serotype 5a strains used in this study are listed in Supplementary Table S4. Cells were cultured in TSB (Tryptone Soya Broth) medium supplemented with 100 μg/ml

ampicillin or 50 μg/ml kanamycin for selection. Liquid cultures and plates were incubated at 37°C overnight, unless indicated otherwise.

Generation of *Escherichia coli* cell lysates in 96-well plates

One milliliter of LB medium with 100 μg/ml carbenicillin disodium salt was inoculated with glycerol stocks of *E. coli* and cells were grown at 37°C in a shaker. Induction was performed by the addition of IPTG (0.4 mM final concentration) after reaching an optical density at 600 nm (OD_{600}) of 0.6. Cells were incubated at 37°C in a shaker for 3 h and then harvested by centrifugation at 4°C for 30 min. Pellets were resuspended and lysed as described previously (40).

Screening of RT-KTq library with fluorescently labeled primer

The RT-KTq mutant library including all possible single mutants at positions I614, E615, L616, F638, R660, A661, T664, G668, V669, L670, Y671, G672, M673, R746, K747, F749 and N750 was applied in the screening assay. Plasmids are listed in Supplementary Table S5. The library was created by the site-directed mutagenesis strategy as described (39,40) and glycerol stocks were directly used to express proteins in 96-deep-well plates. Cell lysates containing RT-KTq variants were prepared and 4 μl was transferred in a 96-well reaction plate. Screening was performed with either dATP or dGTP. Prior to reaction, sufficient amounts of primer and template (unmodified or modified) in a ratio of 1:2 were annealed beginning with 95°C for 2 min and then cooled down stepwise to 4°C. After prewarming reaction plates at 55°C on a thermocycler, 16 μl reaction mix containing RT-KTq reaction buffer [50 mM Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v) Tween 20], dATP or dGTP, annealed primer–template and Milli-Q water was added to reach the final concentrations of 100 nM RNA template, 50 nM primer, 0.5 μM dATP or 20 μM dGTP in 1× RT-KTq reaction buffer. Different fluorescently labeled primers that vary in their size were used for each plate column enabling a concurrent analyzation of 12 reactions by CE: (i) 20 nt 6-carboxyfluorescein (FAM), (ii) 25 nt FAM, (iii) 30 nt FAM, (iv) 35 nt FAM, (v) 40 nt FAM, (vi) 45 nt FAM, (vii) 20 nt hexachlorofluorescein (HEX), (viii) 25 nt HEX, (ix) 30 nt HEX, (x) 35 nt HEX, (xi) 40 nt HEX and (xii) 45 nt HEX. In this way, the two different emission spectra measured for either FAM- and HEX-labeled primers could be used to assign which template (unmodified or modified, respectively) was present in the reaction. Primer extension was performed at 55°C for 10 min and then stopped by adding 20 μl CE stop solution [80% (v/v) formamide, 20 mM EDTA]. Reaction mixtures were analyzed by CE. The nucleotide giving the best mismatch incorporation efficiency was determined in preceding experiments with the RT-KTq DNA polymerase as parental enzyme of the mutant library (Supplementary Figure S1). Suitable nucleotide concentrations were determined beforehand in primer extension reactions with diluted dATP and dGTP stock solutions. Screening of *E. coli* cell lysates was performed once.

Capillary electrophoresis

The analysis of one reaction plate by CE was performed as previously described (40).

Large-scale protein production and purification

The gene of the respective DNA polymerase variant was expressed in *E. coli* BL21-Gold DE3 (Novagen) cells. Four hundred milliliters of LB medium with carbenicillin disodium salt (final concentration 100 μg/ml) was inoculated with 4 ml of the respective liquid overnight culture and incubated at 37°C on a shaker. When the OD_{600} reached 0.6–0.8, gene expression was induced by adding 1 mM IPTG. The culture was incubated at 37°C on a shaker for another 4 h. Then, cells were harvested by centrifugation at 4°C and 4400 rpm for 30 min and pellets were stored at −20°C. Cell pellets were resuspended and lysed in 10 ml lysis buffer [10 mM Tris–HCl (pH 9.2), 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1% (v/v) Triton X-100] containing 1 mg/ml lysozyme at 37°C for 20 min. After denaturing of *E. coli* host proteins at 75°C for 45 min, lysates were ultracentrifuged at 20 000 rpm for 1 h to remove cell debris. Imidazole (5 mM) was added to the supernatant and metal ion-based affinity purification was performed employing the cOmplete™ His-tag purification resin (Roche). For this step, 3 ml nickel beads were washed four times with 5 ml calibration buffer [10 mM Tris–HCl (pH 9.2), 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1% (v/v) Triton X-100, 5 mM imidazole] by an alternating process of mixing and centrifugation at 1000 rpm and 4°C for 5 min. The cell lysate was combined with the calibrated beads and shaken at 4°C in an overhead shaker overnight. The mixture was applied onto a 15 ml chromatography column with moistened frit and the flow-through was collected. Afterward, nickel beads were washed four times with 5 ml washing buffer [10 mM Tris–HCl (pH 9.2), 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1% (v/v) Triton X-100, 20 mM imidazole]. Protein elution was achieved by loading six times 1.6 ml elution buffer [100 mM Tris–HCl (pH 9.2), 5 mM $MgCl_2$, 200 mM imidazole] onto the column and collecting the flow-through separately at 4°C. The success of the protein purification was monitored by SDS–PAGE. Pure enzyme fractions were combined and concentrated by using a ViVAspin (ViVAspin 6/20, 30 kDa MWCO, Sartorius) at 4°C and 3500 rpm. Additionally, remaining imidazole was removed by washing the protein solution four times with 20 ml elution buffer II [100 mM Tris–HCl (pH 9.2), 5 mM $MgCl_2$] followed by a final concentration step to obtain an end volume of 0.2–0.4 ml. For storage, 1/9 volume of 20× storage buffer [100 mM Tris–HCl (pH 9.2), 320 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 2% (v/v) Tween 20] was added to the protein solution and additionally autoclaved glycerol was poured to a final ratio of 1:1. RT-KTq DNA polymerase variants were stored in aliquots at −20°C.

Primer extension with fluorescently labeled primers and purified enzymes

All reaction mixtures contained finally 50 nM fluorescently labeled DNA primer, 100 nM RNA template (modified or unmodified), 28 μM dATP or 200 μM dGTP, and 2 or 5 nM RT-KTq variant in 1× RT-KTq reaction buffer [50 mM

Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v) Tween 20] to obtain an end volume of 20 μl. Annealing was performed similarly as in the screening before. Then, the annealed primer–template duplex was combined with either dATP or dGTP in 1× RT-KTq reaction buffer. After heating the mixture to 55°C, 8 μl DNA polymerase in 1× RT-KTq reaction buffer was added to start the reaction. Primer extension was conducted at 55°C for 10 min and quenched by the addition of 20 μl CE stop solution [80% (v/v) formamide, 20 mM EDTA]. Reactions were directly analyzed by CE. Suitable concentrations of dATP (28 μM) and dGTP (200 μM) were determined in preliminary experiments using purified RT-KTq DNA polymerase. The assay was repeated eight times in total, out of which three assays were performed with 2 nM enzyme, two with different primer combinations for different variants and three experiments were conducted with 5 nM enzyme.

Primer extension with radioactively labeled primer

For multiple incorporation experiments, all four nucleotides were used. The reaction mixture contained 225 nM RNA template (unmodified or modified), 150 nM radioactively labeled DNA primer, 100 or 200 μM of each nucleotide (as indicated) and 2 nM RT-KTq variant in 1× RT-KTq reaction buffer [50 mM Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v) Tween 20]. Reactions were carried out in 10 μl end volume. First, template and primer were annealed by heating the mixture to 95°C for 2 min and then cooling it down stepwise to 4°C. Afterward, nucleotides in 1× RT-KTq reaction buffer were added and prewarmed to 55°C. Primer extension was started by the addition of 4 μl RT-KTq variant in 1× RT-KTq reaction buffer. Reactions were allowed to proceed for indicated reaction times and then were quenched by the addition of 20 μl stop solution [80% formamide, 20 mM EDTA, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanole]. After heating to 95°C for 5 min, reactions were analyzed by 12% denaturing PAGE and visualized by phosphor imaging conducted on a Typhoon™ FLA 9500 (GE Healthcare Life Sciences). Reactions with RT-KTq variants processing pseudouridinylated RNA were performed three times, and reactions in the context of Q were performed twice.

Generation of *S. flexneri tgtΔ*

To generate an *S. flexneri* strain lacking TGT (*tgtΔ*), the λ-Red-mediated homologous recombination was used to replace *tgt* with a kanamycin resistance gene (KAN^R) (42,43). Correct integration was verified by PCR analysis.

Isolation of total RNA from *S. pombe* and *S. flexneri*

S. pombe and *S. flexneri* cells were grown to an OD_{600} of 0.8–1 in 50 ml cultures. Fifty OD of cells were harvested and stored at −80°C until further processing. To isolate total RNA from *S. pombe* or *S. flexneri*, 50 OD of cells were harvested and 1 ml of phenol and glass beads were added. After vigorous shaking for 5 min, samples were centrifuged at 20 000 × *g* for 5 min to clear the cell debris. Equal volume of a phenol/chloroform/isoamyl alcohol mix (25:24:1) was added to the aqueous phase and centrifuged at 20 000 × *g* for 5 min. After mixing the upper phase with an equal volume of chloroform followed by centrifugation at 20 000 × *g* for 5 min, the RNA was precipitated at −80°C for 1 h using 0.7 volume of isopropyl alcohol. Following precipitation, total RNA was washed with 70% ethanol and dissolved in DEPC-treated water. Verification of Q modification in *S. pombe* and *S. flexneri* was performed by separating RNAs in 10% polyacrylamide gels [acrylamide/bisacrylamide (19:1), urea 8 M] supplemented with 5 mg/ml 3-(acrylamido)phenylboronic acid (APB) as described previously (44) followed by northern blotting as previously described (45).

Preparation of RNA extracts from *S. cerevisiae*

Total RNA from yeast cells was isolated using hot acid phenol. RNA concentration was measured on NanoDrop One and RNA quality was assessed by CE using a PicoRNA Chip on Bioanalyzer 2100 (Agilent Technologies). Yeast total tRNA fraction was enriched using NucleoBond RNA 80 kit (Macherey-Nagel). Briefly, 50–80 μg of total RNA was loaded onto AXR80 column previously equilibrated with RO buffer. The column was washed four times with R1 buffer and tRNA fraction was eluted using R2 buffer. The eluate was then precipitated with an equal volume of isopropanol and centrifuged. The pellet was washed with 80% ethanol and dissolved with RNase-free water.

tRNA substrates for *in vitro* modification

The *S. pombe* $tRNA^{Asp}$ and $tRNA^{Asn}$ substrates were prepared as previously described (46).

Recombinant expression and purification of hTGT and Q modification of tRNAs *in vitro*

The pCDF-Duet1 vector coexpressing the human TGT (hTGT) heterodimer QTRT1 and QTRT2 with a cleavable N-terminal 6xHis tag to QTRT1 was kindly provided by Prof. Dr Ralf Ficner (GZMB, Göttingen). Expression and purification of hTGT and Q modification of *in vitro*-transcribed tRNAs were carried out as previously described (45). Q modification was verified by separating RNAs in polyacrylamide gels containing APB (see above).

Reverse transcription and amplification of tRNAs for analysis of Q using NGS

Reverse transcription on *in vitro*-transcribed *S. pombe* tRNAs and tRNAs isolated from *S. pombe* and *S. flexneri* for the detection of Q was performed using the RT-KTq I614Y and RNA-specific primers (Supplementary Table S1). Five hundred nanograms of RNA was hybridized to 500 nM of the respective stem–loop primer with the following thermocycling program: 95°C for 2 min, 65°C for 30 s, ramp down to 4°C with 5°C/30 s. After hybridization and incubation on ice for 1 min, 100 μM of each dNTP and 100 nM RT-KTq variant were added. The reverse transcription was carried out in RT-KTq reaction buffer [50 mM Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v)

Tween 20] in a total volume of 20 μl. For experiments using decreased dCTP concentrations, 25, 12.5 and 6.25 μM dCTP were added to the RT reaction. Following incubation at 55°C for 1 h and incubation on ice for 10 min, the cDNA was amplified using *Thermus aquaticus* (Taq) polymerase. Amplification was performed with 5 μl cDNA in reaction buffer containing 250 μM dNTPs, 2 mM $MgCl_2$, 1× Taq buffer + KCl and 1.25 U Taq polymerase (Thermo Fisher) in a total volume of 200 μl. An initial denaturation was performed at 96°C for 3 min, followed by 40 cycles of 96°C for 30 s, 50°C for 45 s and 72°C for 45 s, and a final elongation step at 72°C for 5 min. Primers binding at the 3′ end contained a barcoded sequence for multiplexing the PCR products (Supplementary Table S1). Samples were separated on a 2% agarose gel, and the respective PCR products were gel purified using the QIAquick Gel Extraktion Kit (Qiagen) according to the manufacturer's protocol.

Library preparation of PCR products for NGS was performed with the NEXTflex™ qRNA-Seq™ Kit v2—Set C (Bioo Scientific) according to the manufacturer's instructions. Briefly, 100 ng of multiplexed PCR product was incubated with the NEXTflex™ End Repair Enzyme Mix in a total volume of 50 μl for 30 min at 22°C. After purifying the DNA via Agencourt AMPure XP Magnetic Beads (Beckman) according to the manufacturer's protocol, adenylation of the DNA was performed by addition of NEXTflex™ Adenylation Mix. After incubation at 37°C for 30 min, followed by incubation for 5 min at 70°C, the DNA was adapter ligated with the Molecular Index Adaptor. After purification using Agencourt AMPure XP Magnetic Beads, the DNA was amplified with 12 cycles of PCR reaction according to the manufacturer's instructions using qRNA-Seq™ Barcoded Primers. Amplicons were purified using Agencourt AMPure XP Magnetic Beads and resuspended in 20 μl nuclease-free water. The library was subjected to high-throughput sequencing. Amplicons were sequenced on an Illumina MiSeq platform using paired-end mode (2 × 150 bp) sequencing. Each DNA library was sequenced once and the resulting coverage was used for error calculations (Supplementary Table S6). Error rates at each position of the templates and Δerror rates of two bases assigned to each other were calculated and illustrated using CoverageAnalyzer (47).

DNA amplicon library preparation from Ψ-modified oligonucleotides for NGS analysis

Reverse transcription reactions were conducted similarly to multiple primer extension reactions with radioactively labeled primers. Reaction mix contained 225 nM RNA template (unmodified or modified oligonucleotide), 150 nM DNA reverse primer with overhang, 100 μM of each nucleotide [or 100 μM d(G/C/T)TP and 2 μM dATP, as indicated] and 2 nM RT-KTq variant in 1× RT-KTq reaction buffer [50 mM Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v) Tween 20]. After annealing of template/primer and heating the mixture to 55°C, 4 μl DNA polymerase in 1× RT-KTq reaction buffer was added to start reverse transcription (10 μl reaction volume). After 10 min reaction time, mixture was heated to 95°C for 5 min and annealing was allowed once again by cooling stepwise to 4°C. The reaction was heated to 55°C and 10 μl SuperScript IV reaction mix (Invitrogen) was added to reach a final volume of 20 μl. Reactions contained finally 10 U/μl SuperScript IV (Invitrogen), 300 μM dNTP each, 5 mM DTT and 1× SuperScript IV RT buffer (Invitrogen). After primer elongation for 10 min at 55°C, the reaction was put on ice. For each DNA library, the described reverse transcription was performed three times and combined after reaction. Sixty microliters of combined reaction mix was mixed with 240 μl buffer NTC (Macherey-Nagel) and clean-up was conducted with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's specifications. Elution was achieved by adding 15 μl Milli-Q water and centrifugation for 1 min at 13 000 rpm. Then, RNA template was digested by addition of 0.05 U/μl RNase H (NEB) in 1× RNase H reaction buffer at 37°C for 20 min. Enzyme was inactivated by heating at 65°C for 20 min. Purification of DNA was performed by agarose gel electrophoresis (2.5% agarose gel in 1× TAE buffer), excision of the respective bands and applying the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). To achieve DNA binding to the columns, buffer NTC was added to the gel pieces (400 μl buffer per 100 mg gel) and incubated at 50°C until gel was dissolved. Elution was conducted in the same way as described above. Afterward, DNA was repaired by using 0.1 μl PreCR Repair Mix (NEB) with 100 μM of each dNTP, 1× NAD^+ (provided by the manufacturer) and 1× ThermoPol reaction buffer (NEB) to reach 20 μl reaction volume. DNA repair was performed at 37°C for 20 min. Without any purification step, DNA was used as template for primer extension reaction with 0.02 U/μl Q5 Hot Start High-Fidelity DNA Polymerase (NEB), 200 μM of each dNTP, 300 nM DNA forward primer with overhang and 0.2× Q5 reaction buffer in 25 μl reaction volume. A one-step cycling protocol was chosen where denaturation at 98°C for 1 min, annealing at 62°C for 45 s and final extension at 72°C for 2 min were conducted. Reaction was analyzed and purified by agarose gel electrophoresis as described above. This time, NTI buffer (Macherey-Nagel) was used for dissolving the gel pieces at 55°C. A second DNA repair with the PreCR mix (NEB) was conducted in the similar way as described above. Unique molecular identifier (UMI) introduction was obtained by applying Q5 Hot Start High-Fidelity DNA Polymerase (NEB) in a three-step PCR reaction. The reaction mixture contained 20 μl repair mix, 0.02 U/μl DNA polymerase, 200 nM forward and reverse primers, and finally 0.2× Q5 reaction buffer to reach 25 μl reaction volume. After initial denaturation at 98°C for 2 min, two cycles with 98°C for 10 s, 62°C for 30 s and 72°C for 30 s were carried out followed by a final elongation step at 72°C for 2 min. Reaction was analyzed and purified by agarose gel electrophoresis as described above. Here, buffer NTI was diluted with Milli-Q water in a ratio of 2:1 and afterward used for dissolving gel pieces at 50°C. Concentration of the DNA was determined with a qPCR approach, where an oligonucleotide with the same length was applied as reference in different template concentrations. 1× NEBNext Ultra II Q5 Master Mix (NEB) was used for PCR with 400 nM forward and reverse primers and 1× SYBR Green I in 5 μl reaction volume. After an initial denaturation step at 98°C for 1 min, a three-step cycling protocol

with 35 PCR cycles followed with 98°C for 10 s, 60°C for 30 s and 72°C for 30 s, and finally an elongation step at 72°C for 2 min was conducted. Fluorescent intensity was measured at the end of each cycle and melting curves were determined at the end of the reaction. For analysis, Cq values of reactions with reference DNA were blotted against the decadic logarithm of the template concentrations and linear regression was created. DNA concentrations were calculated by using the linear predictor function. Results were multiplied by 2 since reference DNA was applied double stranded and target DNA in the analyzed samples was only present in the single-stranded format. DNA was repaired once again by using 0.05 μl PreCR mix (NEB), 1× NAD^+, 100 μM of each dNTP, 59.0 fM DNA and 1× ThermoPol reaction buffer (NEB) to reach 11.25 μl end volume. After reaction for 20 min at 37°C, DNA mix was directly used for amplicon PCR. Here, 400 nM forward and reverse primers (equipped with Illumina Adapter Sequence and Indices), the entire repair mix (final DNA concentration 26.6 fM) and 1× NEBNext Ultra II Q5 Master Mix (NEB) were used for reaction in 25 μl end volume. After an initial denaturation step at 98°C for 5 min, a three-step cycling protocol with 35 PCR cycles followed with 98°C for 10 s, 70°C for 30 s and 72°C for 30 s, and finally an elongation step at 72°C for 2 min was conducted. Reaction was analyzed and purified by agarose gel electrophoresis as described above. To dissolve gel pieces, buffer NTI (Macherey-Nagel) was diluted with Milli-Q water in a ratio of 2:1 and 450 μl NTI mix was added to the gel pieces. After complete dissolving at 50°C, mixture was diluted once again with Milli-Q water to reach a final volume of 1750 μl. DNA was loaded on the column and eluted with 20 μl Milli-Q water. After performing the repair reaction on the whole eluate as described above, DNA was purified by using the QIAEX II System (Qiagen). Two microliters of particle slurry was used for each sample and elution of DNA was achieved by adding 18 μl Milli-Q water to the dried beads, incubation for 5 min and taking 16 μl after centrifugation. The concentration of the final DNA library was determined by Quantus™ Fluorometer (Promega) and usually ranged between 20 and 85 nM. After quality control with Bioanalyzer 2100 (Agilent), DNA libraries were pooled and sequenced in paired-end mode (2 × 75 bp) on an IIlumina NextSeq2000 instrument. Each indicated reaction condition was applied to both the unmodified RNA and its modified version; thus, two DNA libraries were finally obtained. Each DNA library was sequenced once and the resulting coverage was used for error calculation (Supplementary Table S6). Error rates at each position of the templates and Δerror rates of two bases assigned to each other were calculated using KNIME.

DNA amplicon library preparation from tRNA extracts for Ψ analysis using NGS

Three hundred nanograms of RNA (isolated from *S. cerevisiae*) was annealed to 300 nM reverse primer (containing UMI sequence) in the similar way as described above. Hybridized template/primer was added to the reaction mix containing 100 μM of d(G/C/T)TP (each) and 2 μM dATP in 1× RT-KTq reaction buffer [50 mM Tris–HCl (pH 9.2), 16 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1% (v/v) Tween 20]. After heating to 55°C, RT was started by adding 100 nM RT-KTq I614Y in 1× RT-KTq reaction buffer to reach a final reaction volume of 10 μl. After 1 h reaction time, mixture was heated to 95°C for 5 min and annealing was allowed once again by cooling to 4°C stepwise. The reaction was heated to 55°C and 10 μl SuperScript IV reaction mix was added to reach a final volume of 20 μl. Reactions contained finally 10 U/μl SuperScript IV (Invitrogen), 300 μM dNTP each, 5 mM DTT and 1× SuperScript IV RT buffer (Invitrogen). After primer elongation for 10 min at 55°C, the reaction was put on ice. Reaction mix was mixed with 160 μl buffer NTC (Macherey-Nagel) and clean-up was conducted with NucleoSpin Gel and PCR Clean–up XS Kit (Macherey-Nagel) according to the manufacturer's specifications. Elution was achieved by adding 16 μl elution buffer (Macherey-Nagel). DNA was used for second UMI introduction with 0.02 U/μl Q5 Hot Start High-Fidelity DNA Polymerase (NEB), 200 μM of each dNTP, 200 nM DNA forward primer and 1× Q5 reaction buffer in 25 μl reaction volume. A one-step cycling protocol was chosen where denaturation at 98°C for 1 min, annealing at 62°C for 45 s and final extension at 72°C for 2 min were conducted. Buffer NTI (Macherey-Nagel) was mixed with water in a ratio of 3:1 and 50 μl NTI mix was added to the PCR reaction mix. The NucleoSpin Gel and PCR Clean–up XS Kit (Macherey-Nagel) was used for purification and 16 μl elution buffer (Macherey-Nagel) was added for elution. Concentration of the DNA was determined with a qPCR approach described above. DNA was repaired by using 0.05 μl PreCR mix (NEB), 1× NAD^+, 100 μM of each dNTP, 48.9 fM DNA and 1× ThermoPol reaction buffer (NEB) to reach 11.25 μl end volume. After reaction for 20 min at 37°C, DNA mix was directly used for amplicon PCR with NEBNext Ultra II Q5 Master Mix (NEB) as described above (final DNA concentration 22.0 fM). Purification of DNA was performed by agarose gel electrophoresis (2.5% agarose gel in 1× TAE buffer), excision of the respective bands and applying the NucleoSpin Gel and PCR Clean-up XS Kit (Macherey-Nagel). Here, buffer NTI (Macherey-Nagel) was diluted with Milli-Q water in a ratio of 2:1 and 450 μl NTI mix was added to the gel pieces. After complete dissolving at 50°C, mixture was diluted once again with Milli-Q water to reach a final volume of 1750 μl. DNA was loaded on the column and eluted with 20 μl elution buffer (Macherey-Nagel). DNA repair on the whole eluate and purification with the QIAEX II system (Qiagen) were performed as described above. The concentration of the final DNA library was determined by Quantus™ Fluorometer (Promega) and usually ranged between 6 and 46 nM. After quality control with Bioanalyzer 2100 (Agilent), DNA libraries were pooled and sequenced in paired-end mode (2 × 75 bp) on an IIlumina NextSeq2000 instrument. The reaction condition above was applied to both $tRNA^{Gly}$ (GCC) from wt and *pus4Δ*; thus, two DNA libraries were finally obtained. Each DNA library was sequenced once and the resulting coverage was used for error calculation (Supplementary Table S6). Error rates at each position of the templates and Δerror rates of two bases on the same position were calculated using KNIME. We confirmed our findings for Ψ55 by analyzing the same tRNA from S. cerevisiae wild type (wt) and pus4Δ as used in HydraPsiSeq (Supple-

mentary Figure S10) as described in (5). The results show that tRNA from wt displays almost complete Ψ modification and tRNA from pus4Δ does not carry Ψ55.

NGS data processing of amplicon libraries for analysis of Q modification

Reads obtained for tRNAs from *S. pombe* and *S. flexneri* for the detection of Q were processed using in-house R scripting and the Bioconductor package ShortRead (48). Following the processing, which included trimming of PCR primers, selection of high-quality reads and sorting of the reads based on the sequence in the degenerate region of the RT primer, reads were trimmed to the expected size of the transcript using Cutadapt (49). Subsequently, alignment to RNA sequences from *S. pombe* (https://www.pombase.org/) and *S. flexneri* (http://www.ensembl.org/) was performed using SAMtools and Bowtie2 (50,51). Settings were chosen to allow up to one mismatch in seed with a seed length of 6. Misincorporation was determined using CoverageAnalyzer (47). The *P*-values of error rates were determined using a Wilcoxon–Mann–Whitney (WMW) test (GraphPad Prism, version 6.00).

NGS data processing of amplicon libraries for the analysis of Ψ

Sequencing data were processed by using the software KNIME (Supplementary Table S8). Quality and sequence are read from the FastQ files. Read1 and Read2 were merged giving the expected size of transcript. High-quality data were filtered and aligned to the RNA reference of the oligonucleotide or tRNA sequences (http://gtrnadb.ucsc.edu). Reads were sorted into UMI families and family read numbers were counted. Error calculation are performed in each UMI family having >10 or >15 (tRNA data) reads. If 90% of the reads within each family carry the misincorporation, the error is set to 1 (otherwise 0). Afterward, the error is calculated over all UMI families at each position of the transcript and plotted using Microsoft Excel software. The coverage and the number of UMIs used for error calculation are listed individually for each sequencing library in Supplementary Table S6. *P*-values for the statistical analysis in Supplementary Figure S11 were determined using the WMW test (GraphPad Prism, version 6.00).

RESULTS

Strategy for the detection of Ψ and Q in RNA

Here, we report the development of a novel Ψ/Q detection approach that makes use of an RT-active DNA polymerase variant that displays an increased error rate opposite Ψ and Q and operates without prior chemical treatment of the RNA (Figure 1). Such RT signals are amplified during subsequent PCR reactions and are detected as mutation hot spots after performing NGS and data processing.

Screening for a DNA polymerase variant that discriminates Ψ from U

First, we set out to develop a Ψ-responsive polymerase. The enzyme was evolved from the thermostable and RT-active Taq variant termed RT-KTq that contains the following mutations: L459M S515R I638F M747K (52). In order to identify an RT-active DNA polymerase that can be used for Ψ detection, a screen with 5′-fluorophore-labeled DNA primers duplexed with either unmodified or modified RNA templates was performed in parallel (Figure 2A) (40). This screen was applied to test 312 RT-KTq single mutants that were generated as described (39,40). Mutations within the polymerase were chosen based on their close proximity to the nascent base pair based on the crystal structure of the RT-KTq DNA polymerase complexed with a DNA/RNA primer/template complex and the incoming nucleoside triphosphate (Figure 2B) (52). Genes of the DNA polymerase mutants were expressed in *E. coli*, and respective cell lysates were directly used to perform single-nucleotide incorporation experiments after heat treatment. Two RNA oligonucleotides were used as templates that exhibited either U or Ψ on the same position in the RNA sequence. DNA primers for reverse transcription were designed with either FAM or HEX fluorescent tags and with DNA overhangs of distinct lengths. After performing single primer extension, reactions were pooled and analyzed by CE. Since the DNA primers were distinguished by both size and fluorescent signals, the analysis of extension peaks in the electropherogram allowed a qualitative assessment of the individual RT reactions (Figure 2A).

Guided by our previous studies where m^6A was investigated (40), we searched for DNA polymerase variants whose dAMP incorporation (matching nucleotide) opposite Ψ is decreased, which demonstrates a lower efficiency of the enzyme in processing the modified nucleotide (Figure 2A, right). Such a DNA polymerase may also tend to incorporate the incorrect nucleotide to a higher level opposite Ψ compared to U. This would result in an enzyme generating an RT signature at modification sites while reverse transcribing canonical nucleotides without increased error rates. To achieve this, both dAMP incorporation and dGMP misincorporation were monitored opposite U and Ψ within the screening procedure. dGMP was chosen as the mismatching nucleotide, because it was incorporated with the highest efficiency by the parental RT-KTq (Supplementary Figure S1).

The screening revealed that decreased incorporation of dAMP opposite Ψ was already observed for the parental RT-KTq, but some variants with single mutations at positions I614, F638, R660, A661, T664, G668, G672, K747 and F749 showed higher discrimination between modified and unmodified templating nucleotides (examples in Supplementary Figures S2 and S3). Furthermore, DNA polymerases with mutations at I614, F638, A661 and T664 showed an increased dGMP misincorporation compared to RT-KTq. The most promising DNA polymerase variants exhibiting a combination of high Ψ discrimination and/or high misincorporation rate were selected for further analysis. Single-incorporation studies with purified enzymes were performed using similar reaction conditions as the above screening assay. RT-KTq and variants with mutations I614A/D/Y, F638H/L/Q, R660D/K/N, A661E/N/R, T664A/K/N/P, G668K/M, G672S, K747G/M/R/Y and F749C were purified and screened with the same

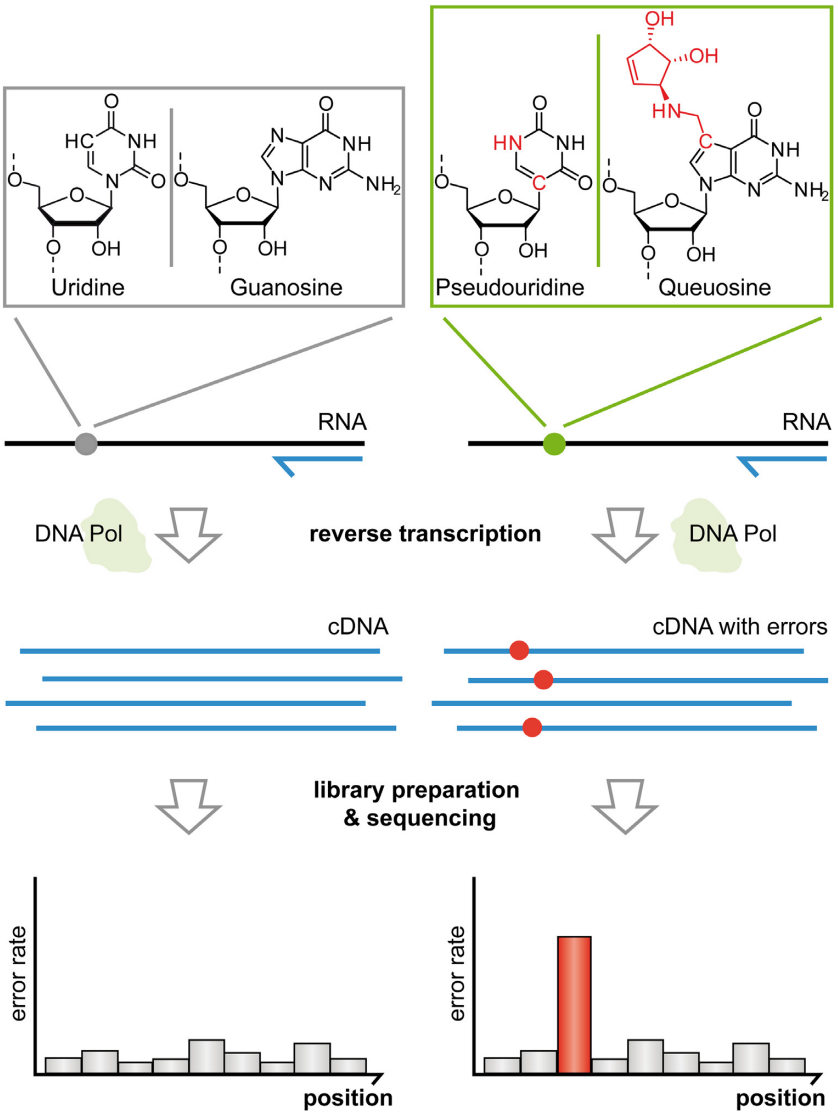


Figure 1. Strategy for the detection of the RNA modifications Ψ and Q by increased misincorporation. During the RT step, the DNA polymerase enzyme incorporates higher levels of mismatched nucleotides opposite modified nucleotides (right side, mismatch depicted as circle on the cDNA). Resulting errors within the cDNA strand are retained during the following amplification steps and detected by NGS. Sites of higher error rates (error hot spots) are used to distinguish modified (right) from unmodified positions (left).

fluorescently labeled primers and RNA templates as above. Analysis of all single-nucleotide incorporation reactions by CE and data processing verified 16 out of 24 enzymes to feature decreased dAMP incorporation only opposite Ψ (namely RT-KTq I614A/D/Y, F638Q, R660D/K, A661N/R, T664P, G668M/K, G672S, K747G/M/Y and F749C; Supplementary Figure S4). Furthermore, variants with mutations at I614A/D/Y, F638L/Q, A661E/R and T664K/P featured similar or better misincorporation of dGMP compared to the parental enzyme (Supplementary Figure S5).

To gain further insights, nucleotide incorporation experiments were performed using radioactively labeled primers. Primers were hybridized to RNA templates carrying either Ψ or U at the position of first incorporation. Reactions were performed in the presence of all dNTPs and stopped after different time intervals. Reactions using RT-KTq variants with mutations I614A, I614Y, G672S, K747G, K747M, K747Y or F749C were subsequently analyzed by PAGE and phosphor imaging (Supplementary Figure S6). While most of the variants exhibited little Ψ discrimination, the most prominent discrimination effect was visible in the band pattern from reactions with RT-KTq I614Y (Figure 2C, right). This variant shows increased stalling after processing Ψ, resulting in a prominent band directly after the first incorporation (Figure 2C, black arrow). We surmise that the difficulties of the polymerase in extending the primer might come from incorrect Watson–Crick base pairing between the 3′ end of the primer and the modified templating nucleotide. Hence, processing Ψ by RT-KTq I614Y causes partial misincorporation directly at the first position of the primer, and its extension is consequently hampered. Of note, the overall activity of RT-KTq I614Y during reverse transcription was higher than that of the parental RT-KTq

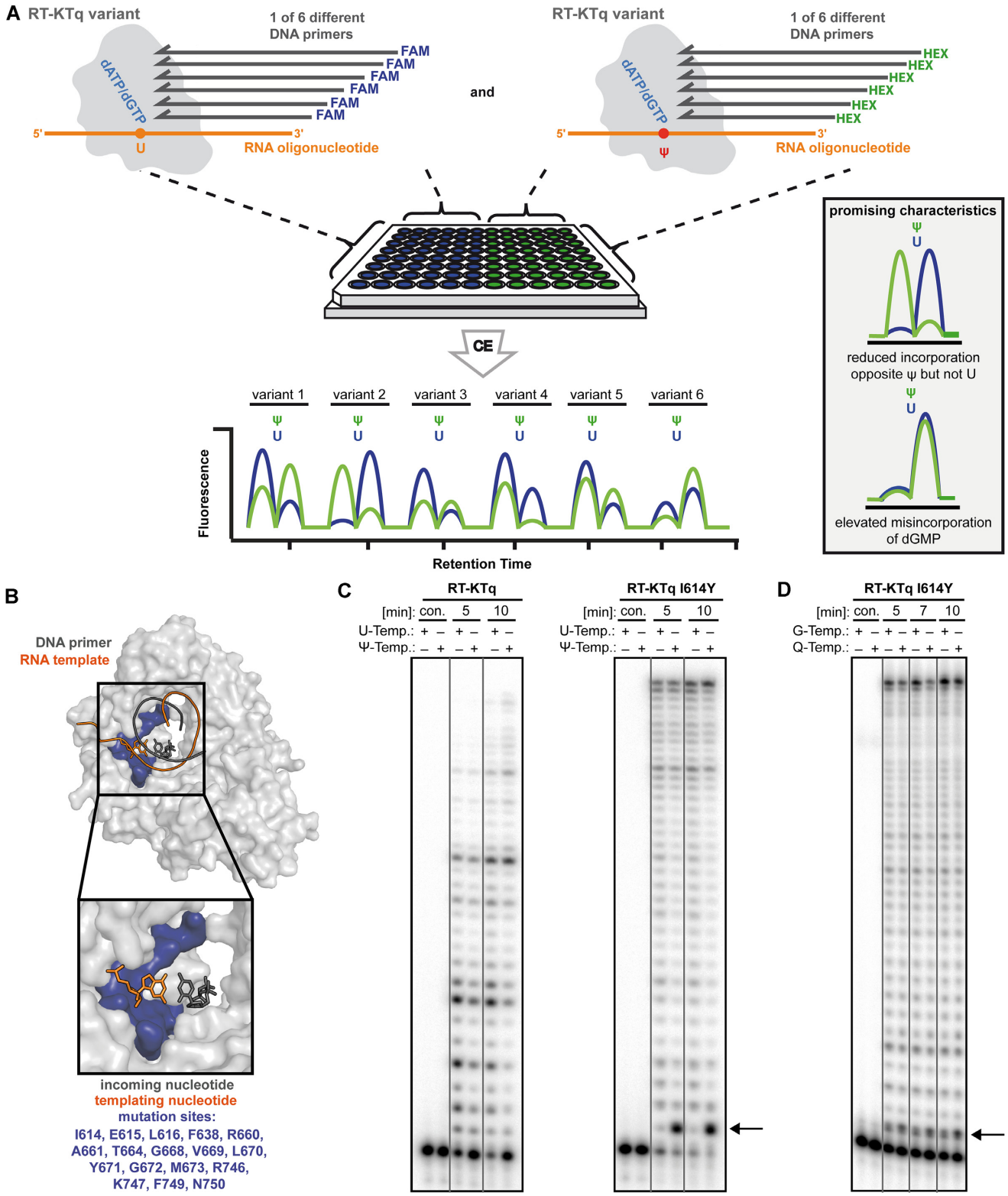


Figure 2. Screening for DNA polymerase variants with the ability to discriminate Ψ from U and validation of enzyme performance. (**A**) DNA polymerase expression lysates or purified enzymes were employed in single primer extension reactions with fluorescently labeled primers of different lengths (FAM = 6-carboxyfluorescein; HEX = hexachlorofluorescein). Incorporation of dAMP or dGMP was monitored opposite Ψ or U, respectively, and reactions were analyzed by CE. (**B**) Crystal structure of RT-KTq with amino acids used for enzyme library preparation highlighted. Adapted from PDB ID: 4BWM using PyMOL (Schrödinger, LLC, New York, NY). (**C**), (**D**) Multiple primer extension reactions with radioactively labeled primer and modified or unmodified template RNA; 2 nM RT-KTq variant and 100 μM dNTP were applied for either 5, 7 or 10 min reaction time. Reaction was analyzed by denaturating PAGE and phosphor imaging. (**C**) RNA oligonucleotide with either U or Ψ was used as template. Experiments were repeated twice. (**D**) RNA oligonucleotide with either G or Q was employed. Experiments were performed twice.

enzyme, since more full-length product was obtained in the same reaction time.

Increased misincorporation opposite Ψ sites by RT-KTq I614Y

In order to gain insights into nucleotide incorporation of RT-KTq I614Y at Ψ sites, the enzyme was used to reverse transcribe from RNA templates, cDNA was amplified and the amplicons were subsequently subjected to NGS (Figure 2C). Since we speculated that misincorporation opposite Ψ leads to hampered extension (53,54), a commercial reverse transcriptase was applied in a follow-up reaction to promote RT. NGS reads were processed with KNIME (55), aligned to the reference sequence and error rates were calculated for each position.

Interestingly, we observed different error rates across the RNA sequence also at positions that do not carry Ψ (Figure 3A). This indicates that the polymerase as such is more error-prone, and that the errors also depend on other features such as sequence or secondary structure of the RNA (56,57). Importantly, however, we observed an increased error rate at the Ψ site compared to the unmodified template, and the error rate was higher for RT-KTq I614Y than for the parental enzyme. Other sites of the RNA showed largely equal error rates (Figure 3A, left). Plotting the difference in error rates (with Ψ − without Ψ) for the two enzymes showed that 5% error increase is detected when Ψ is processed by RT-KTq (Supplementary Figure S7A, right, black arrow) and 16% error increase when RT-KTq I614Y was used for catalysis (Figure 3A, right, black arrow). Thus, the mutation I614Y increased the difference of error on the modification site by a factor of 3.

In the next step, we aimed to enhance the discrimination effect of RT-KTq I614Y at Ψ sites by employing an imbalance of dNTPs. To this end, we decreased the dATP concentration from 100 to 2 μM in the first RT step while leaving other nucleotides at 100 μM. With this strategy, misincorporation was expected to increase since less of the matching nucleotide for Ψ/U is available. Indeed, the error difference was increased to 32% on modified versus unmodified template (Figure 3B, black arrow). Comparing the performance of RT-KTq I614Y with that of the parental RT-KTq at an imbalanced nucleotide pool, the error difference at Ψ sites was increased by a factor of ∼5 (RT-KTq showed 7% error increase; Supplementary Figure S7B, right, black arrow). Inspection of the sequence reads at the Ψ site for RT-KTq I614Y showed ∼50% T reads (dAMP incorporation during RT), 30% C reads (dGMP misincorporation), 13% G reads (dCMP misincorporation) and 6% A reads (dTMP misincorporation) (Figure 3B, bottom, and Supplementary Table S7). Thus, most misincorporation opposite Ψ was G. At the unmodified U site, most misincorporation was dCMP (9%), followed by dGMP misincorporation with 7%. Remarkably, the reduction of dATP in the reaction mix had little to no effect on the fidelity of RT-KTq or RT-KTq I614Y on positions other than U (Supplementary Figure S8). The mutation I614Y slightly increased the errors at A and G sites, but had the biggest impact on the fidelity at U bases. The highest increase in error rates was observed at Ψ sites for both RT-KTq and RT-KTq I614Y (Supplementary Figure S8, red dots), but the mutation I614Y and reduction of dATP exacerbated the error susceptibility of the DNA polymerase to the modification.

Next, RT-KTq I614Y was used to monitor a known Ψ site in the sequence context of human 18S rRNA. Ψ1445 is an interesting target regarding the multisystem disorder X-linked dyskeratosis congenita since the level of Ψ was reduced at position 1445 in patient cells with DKC1 gene mutations (58). We employed two 40-nt unmodified or modified oligonucleotides corresponding to the rRNA sequence around position 1445. This time, RT-KTq I614Y produced ∼4% more errors opposite Ψ than U (Figure 3C, top, black arrow). All other positions did not display a fidelity difference >0.5%. Also, as for the synthetic RNA template, C reads accounted for the largest proportion of the Ψ signature in the sequence context of human 18S rRNA (Figure 3C, bottom, and Supplementary Table S7). This Ψ mark was not observed with the parental RT-KTq enzyme. At the modification site, no discrimination of Ψ and U was observed, but a slight increase of errors was detected at the adjacent position (Supplementary Figure S9A, right, black arrow). Altogether, this demonstrates that RT-KTq I614Y is able to discriminate between Ψ and U independently of the RNA template.

In order to investigate the performance of the enzyme on RNA extracts, RT-KTq I614Y was further employed for the analysis of Ψ55 in *S. cerevisiae* $tRNA^{Gly}$(GCC). Purified total tRNA fractions were prepared from *S. cerevisiae* wt and *pus4Δ* deletion strain and employed in RT with RT-KTq I614Y and imbalanced nucleotide concentrations as described above. The level of Ψ55 in these samples was furthermore determined by HydraPsiSeq (Supplementary Figure S10) as described in (5). Of note, the absence of Pus4 in yeast cells results in the absence of Ψ55 in the tRNA sequence, allowing the latter to be used as a natural control template (59). After library preparation, NGS and data processing, an elevated error difference of ∼6% was observed for tRNA from the *pus4Δ* strain compared to tRNA from wt (Figure 3D, top, black arrow). Interestingly, the 5-methyluridine (m^5U) adjacent to Ψ55 showed the second highest error difference. Other Ψ sites did not show an error signature since pseudouridylation was equal on both tRNA species.

Like the results received from experiments with synthetic RNA oligonucleotides, the proportion of C reads (misincorporation of dGMP) increased opposite Ψ compared to the corresponding U (Figure 3D, bottom, and Supplementary Table S7). Of note, the DNA polymerase tends to stall after processing Ψ; therefore, the RT detection primer should bind directly next to the modification site. The difference in error rates at non-Ψ bases is significantly smaller than that at the Ψ site (Supplementary Figure S11), rendering RT-KTq I614Y a useful enzymatic tool for the detection of Ψ.

RT-KTq I614Y shows increased misincorporation of dTMP opposite Q in tRNAs

Having established that RT-KTq I614Y can faithfully detect Ψ modification in RNAs, we investigated whether other modifications are detectable using this approach. Our data showed slightly increased error rates at positions harbor-

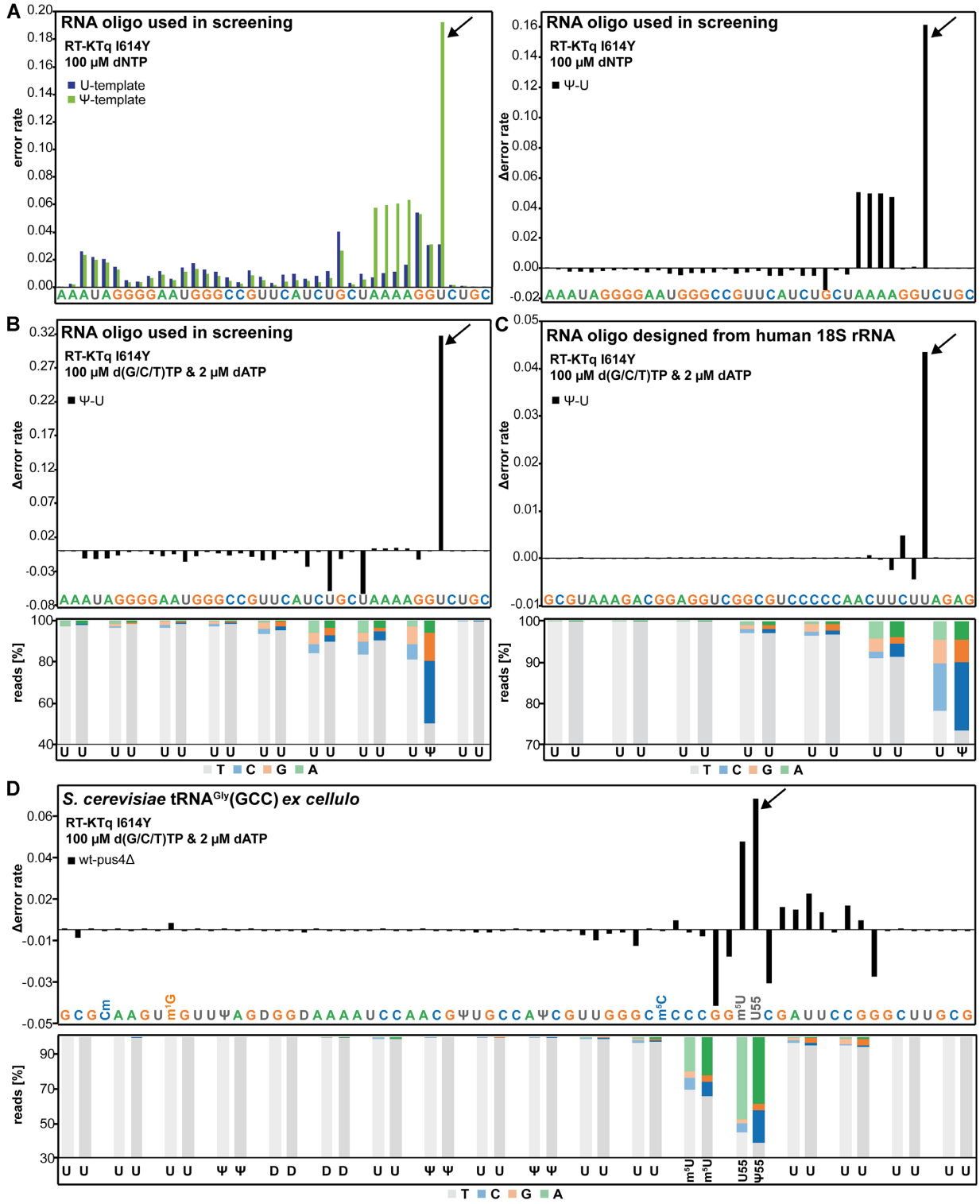


Figure 3. RT-KTq I614Y features elevated error rates opposite Ψ. (**A**) Left: Error rates from RT-KTq I614Y processing the unmodified (blue bars) and modified RNA template (green bars) used in the screening; 100 µM of each dNTP was present in the RT reaction and DNA products were used for amplicon library preparation and NGS. Right: Error rates from RT-KTq I614Y determined from unmodified template are subtracted from error rates with modified template. (**B**) Top: Difference of error rates from RT-KTq I614Y processing either unmodified or modified template used in the screening; 100 µM d(G/C/T)TP and 2 µM dATP were applied in RT. Bottom: Mismatch signature of RT-KTq I614Y opposite Ψ and unmodified Us. (**C**) Top: Difference of error rates from RT-KTq I614Y processing either unmodified or modified RNA template designed from human 18S rRNA around position 1445. RT was performed with 100 µM d(G/C/T)TP and 2 µM dATP. Bottom: Mismatch signature of RT-KTq I614Y opposite Ψ and unmodified Us. (**D**) Top: Difference of error rates from RT-KTq I614Y, processing either tRNA isolated from *S. cerevisiae* wt or *pus4Δ*; 100 µM d(G/C/T)TP and 2 µM dATP were used for reaction (Cm = 2′-*O*-methylcytidine; m^1G = 1-methylguanosine; D = dihydrouridine; m^5C = 5-methylcytidine; m^5U = 5-methyluridine). Bottom: Mismatch signature of RT-KTq I614Y opposite Ψ and Us (or D and m^5U). Each DNA library was sequenced once.

ing Gs adjacent to Ψ in the *S. cerevisiae* tRNA sequencing data (Supplementary Figure S9C). Similar observations were made for the RNA oligonucleotide used in the screening (Supplementary Figure S7C). In the context of tRNA modifications, Q represents an important G modification in RNA (27,28). We next sought to determine whether RT-KTq I614Y is also able to discriminate between Q and G and thus could be used to detect Q modification in tRNAs. For this purpose, a synthetic RNA oligonucleotide was Q-modified *in vitro* using recombinant human tRNA guanine transglycosylase (hTGT), and modification was verified using APB gels (Supplementary Figure S12A). RT-KTq I614Y was used in a similar primer extension experiment with radioactively labeled primer as described above. The first extension opposite G or Q leads to a double band with different intensities (Figure 2D, black arrow). When Q is present in the template, the intensity of the upper band increases, while the lower band is less prominent. Given the result of the previous experiments with Ψ, we suspected that misincorporation opposite Q is probably favored and error signatures will arise during RT. Of note, the band intensity of the first incorporation opposite Q is less intense than seen for Ψ, indicating that RT-KTq I614Y might extend the mismatched primer opposite Q more efficiently.

Hence, RT-KTq I614Y was used for RT of *in vitro*-transcribed and enzymatically Q-modified $tRNA^{Asn}$ from *S. pombe* (Supplementary Figure S12B) (31) using a stem–loop primer specific to the 3′ end of this tRNA and 100 μM of all dNTPs, followed by PCR amplification and NGS. Interestingly, RT-KTq I614Y exhibits increased error rates, particularly opposite G positions, at many sites along the tRNA template sequence (Figure 4A), and the highest error rates were found at the 5′ end of the tRNA, i.e. at the 3′ end of the reverse transcription. Importantly, while the error on G34- and Q34-$tRNA^{Asn}$ was similar at many sites, we observed a pronounced increase of the error rate on Q-modified $tRNA^{Asn}$ position 34 (G34: 4.2%; Q34: 10.6%), thus showing that Q modification enhanced misincorporation. At G34-$tRNA^{Asn}$, we observed 96% G reads (due to dCMP incorporation opposite G34) and 3% A reads (dTMP incorporation) (Figure 4B, top, and Supplementary Table S7). Importantly, A reads increased to 9.5% (due to dTMP incorporation) at position 34 in Q-modified Q34-$tRNA^{Asn}$ (Figure 4B and Supplementary Table S7). This demonstrated that Q modification alters the pairing properties of guanine to allow stronger wobble base pairing with thymidine.

We also investigated error rates opposite Q in *in vitro*-transcribed $tRNA^{Asp}$ (Figure 4C). Here, Q modification increased misincorporation at position 34 from 0.3% (G34) to 0.9%.

Detection of Q modification on tRNAs *ex cellulo* in *S. pombe* and *S. flexneri*

RT-KTq I614Y was next tested for the analysis of Q-modified sites in tRNAs isolated from *S. pombe* cells. To ensure Q modification in *S. pombe*, the growth medium was supplemented with queuine, since this type of yeast, like other eukaryotes, lacks the enzymes of the queuosine biosynthetic pathway and therefore relies on external sources for Q (30,46). Feeding yeast cells with queuine induces Q modification of the four tRNAs with a GUN anticodon to >95% (45). Total RNA was isolated, and $tRNA^{Asp}$ was reverse transcribed with RT-KTq I614Y as above. Importantly, Q modification was detectable on $tRNA^{Asp}$ from *S. pombe*, since a higher error rate was observed at position 34 from queuine-supplemented versus unsupplemented cells (Figure 5A; 1.2% for Q34 versus 0.1% for G34). However, the error rate was low compared to the error rate on *in vitro*-transcribed tRNAs. Therefore, to enhance error rates opposite Q, in an equivalent approach to the one employed above for Ψ, we performed RT reactions on $tRNA^{Asp}$ *ex cellulo* under conditions of reduced dCTP levels, the matching base incorporated opposite Q/G. Indeed, while this increased the error rates at many G positions in $tRNA^{Asp}$ (Supplementary Figure S13), it strongly enhanced the discrimination of G versus Q at position 34, whereas the error rate at other G positions that are not Q-modified was roughly equal (Figure 5B).

At the lowest concentration (6.25 μM, 1/16th the concentration of the other dNTPs), the error rates at Q34 and G34 were 23.4% and 4.4%, respectively. Further reduction of dCTP levels prevented the amplification of the tRNA from the complex sample. It was further of interest to determine the lower limit of detection of Q modification. Mixing of *in vitro*-transcribed and Q-modified with unmodified $tRNA^{Asp}$, reverse transcription (6.25 μM dCTP) and amplification/NGS as above revealed increased error rates down to a level of ∼10–15% Q modification (Figure 5C and Supplementary Figure S14).

We next proceeded to measure Q modification on the full set of Q-tRNAs *ex cellulo* from a prokaryotic source, *S. flexneri*. This bacterium contains all enzymes responsible for the *de novo* biosynthesis of Q-containing tRNAs from GTP.

To obtain an unmodified control, tRNAs were analyzed from a strain carrying a deletion of the enzyme encoding tRNA guanine transglycosylase (*tgtΔ*), which catalyzes the incorporation of the precursor $preQ_1$ into bacterial tRNAs (Supplementary Figure S12C). The isolated RNA was subsequently used for RT with RT-KTq I614Y using primers specific to the Q-tRNAs, followed by PCR amplification and NGS. dCTP was used at 25 μM, because PCR amplification failed at lower concentrations. This resulted in increased error rates opposite Q compared to G at all four Q-tRNAs from *S. flexneri* (Supplementary Figure S15, black arrow). Except for $tRNA^{Asp}$, the difference amounted to ∼10–20%, and the two $tRNA^{Tyr}$ isoacceptors showed similar error rates (Figure 6).

To determine the difference in error rates at Q versus non-Q position, the error rates from all experiments were compiled and subjected to statistical analysis (Supplementary Figure S16). Indeed, the error rates at Q positions were significantly higher than those at non-Q positions (*P*-value <0.0001) showing that RT-KTq I614Y can detect Q modification.

Taken together, this demonstrates that reverse transcription with RT-KTq I614Y combined with NGS is a powerful method to identify Q modification sites on tRNAs. Of note, as above for Ψ, the error rate of the polymerase alone is not sufficient to detect the modification, because some non-

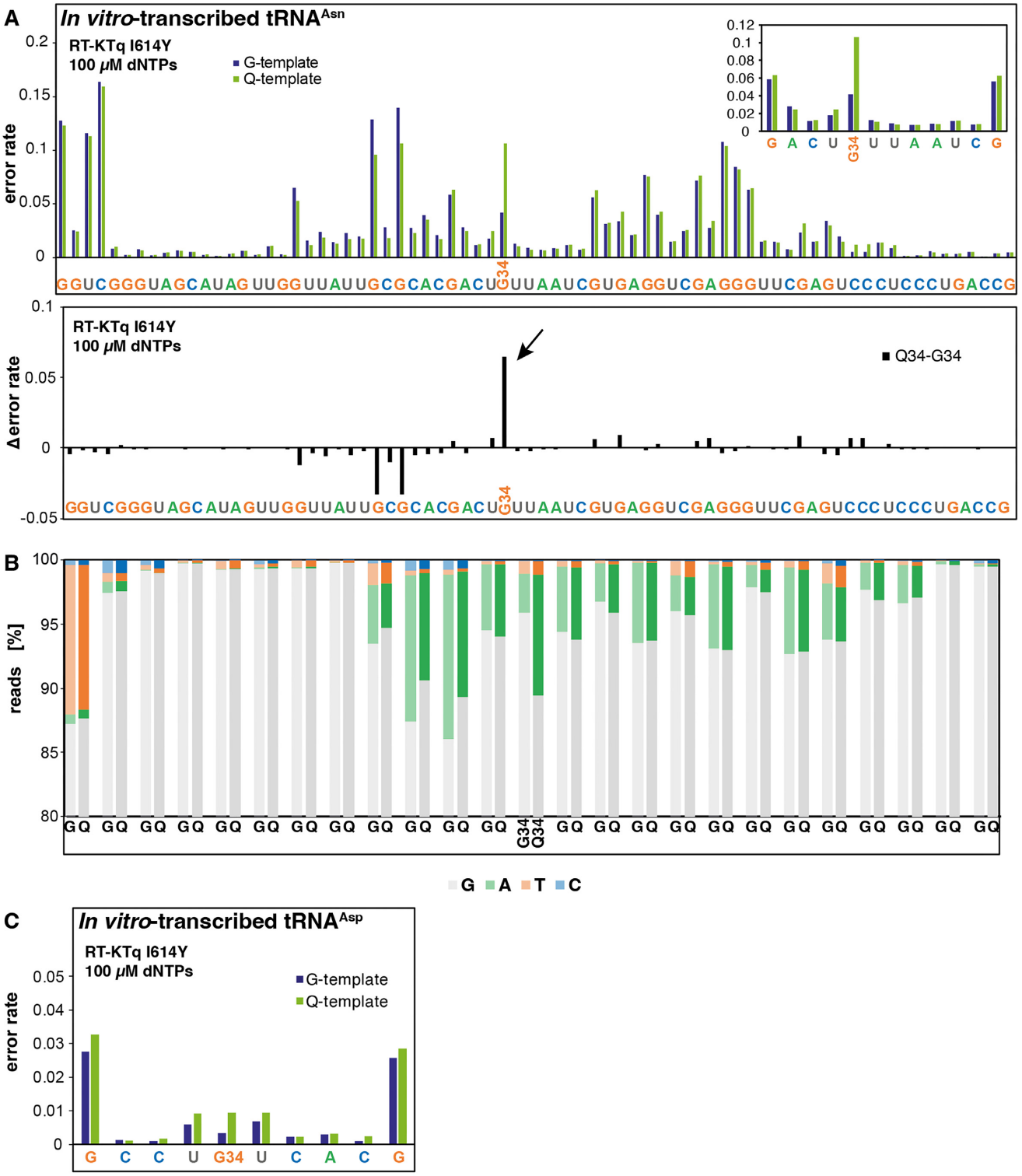


Figure 4. RT-KTq I614Y causes enhanced error rates opposite Q. (**A**) *In vitro*-transcribed tRNA^{Asn} (*S. pombe*) that was enzymatically Q-modified (green) or not (blue) was reverse transcribed, amplified and amplicons were subjected to NGS. The plot shows the error rate of each position along the tRNA (top) and the difference in error rates between Q-modified and unmodified tRNA^{Asn} (bottom). (**B**) Mismatch signature of RT-KTq I614Y opposite Q and all unmodified Gs present in tRNA^{Asn}. (**C**) Error profile of *in vitro*-transcribed tRNA^{Asp} (*S. pombe*). Analysis and representation as in panel (A). Each sample was sequenced once.

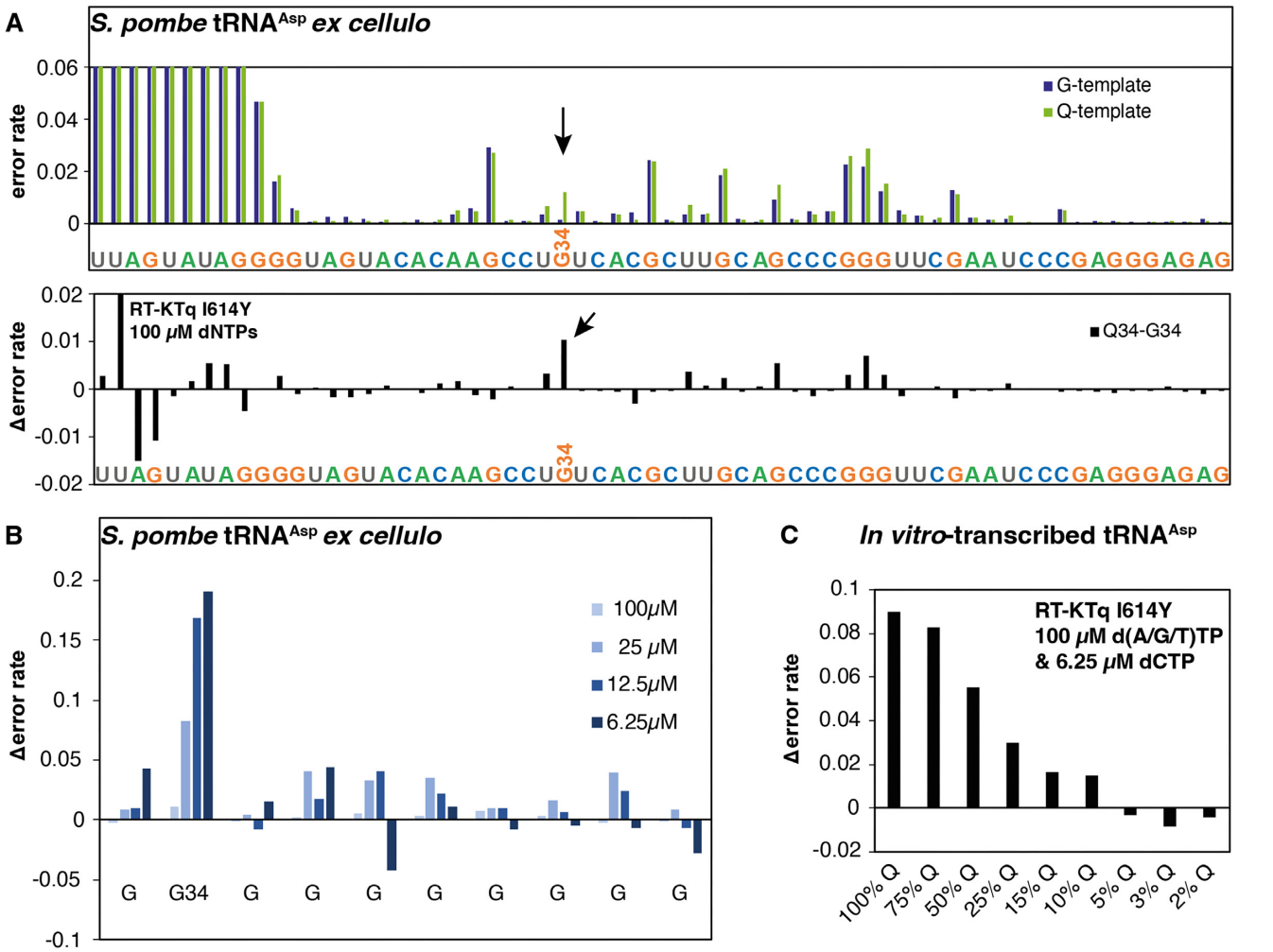


Figure 5. Detection of Q signature in tRNA isolated from *S. pombe*. (**A**) Top: Mutational profile of $tRNA^{Asp}$ isolated from wt *S. pombe* cells that were cultured in the presence (Q34, green) or absence (blue) of queuine. Bottom: Difference in error rates between Q-modified and unmodified $tRNA^{Asp}$. Reverse transcription was performed with RT-KTq I614Y and 100 μM dNTPs. (**B**) Decreasing the dCTP concentration enhances discrimination of RT-KTq I614Y between Q and G in $tRNA^{Asp}$ from *S. pombe*. tRNA samples were obtained as in panel (A) and were reverse transcribed with the indicated dCTP concentrations and 100 μM d(A/G/T)TP. Representation as in panel (A) (bottom). (**C**) The limit for Q detection with RT-KTq I614Y is ∼10–15%. *In vitro*-transcribed $tRNA^{Asp}$ (*S. pombe*) that was enzymatically Q-modified *in vitro* was mixed with unmodified tRNA to the indicated levels, reverse transcribed with RT-KTq I614Y (6.25 μM dCTP) and processed as in panel (A). Representation as in panel (A) (bottom). Each sample was sequenced once.

Q-modified G sites also show high misincorporation. However, the comparison of error rates between unmodified and Q-modified templates reliably identifies Q sites.

DISCUSSION

In this work, we describe the engineering of an RT-active DNA polymerase, RT-KTq I614Y, that is capable of discriminating Ψ and Q from their unmodified counterparts by generating misincorporation signatures during cDNA formation that can be detected by NGS. With this tool, direct Ψ and Q detection is possible without prior chemical modification of the RNA, and it thus presents a significant improvement over existing methods for Ψ and Q detection (36). The enzyme was discovered in a screening approach in which we screened an RT-KTq mutant library and monitored single-nucleotide incorporation opposite U or Ψ. We identified several mutation sites in RT-KTq that reduced dAMP incorporation efficiency opposite Ψ compared to the parental enzyme. Our further analysis and characterization lead to the identification of RT-KTq I614Y. We also observed that this enzyme displays increased misincorporation opposite Q modification sites, so it is possible that it can also detect other RNA modifications. We note that the increased error rate *per se* is insufficient to identify a given modification, but that the comparison of error rates of unmodified versus modified template reliably indicates the modified site. It is also interesting to note that the most frequent misincorporation opposite Q was dTTP, which for the first time demonstrates that Q modification enhances wobble base pairing (28).

The mechanism by which this polymerase acquires its properties is currently unclear. It is conceivable that chemical differences in Ψ and Q compared to their unmodified counterparts influence the transition states leading to mismatch insertion and thereby confer the observed char-

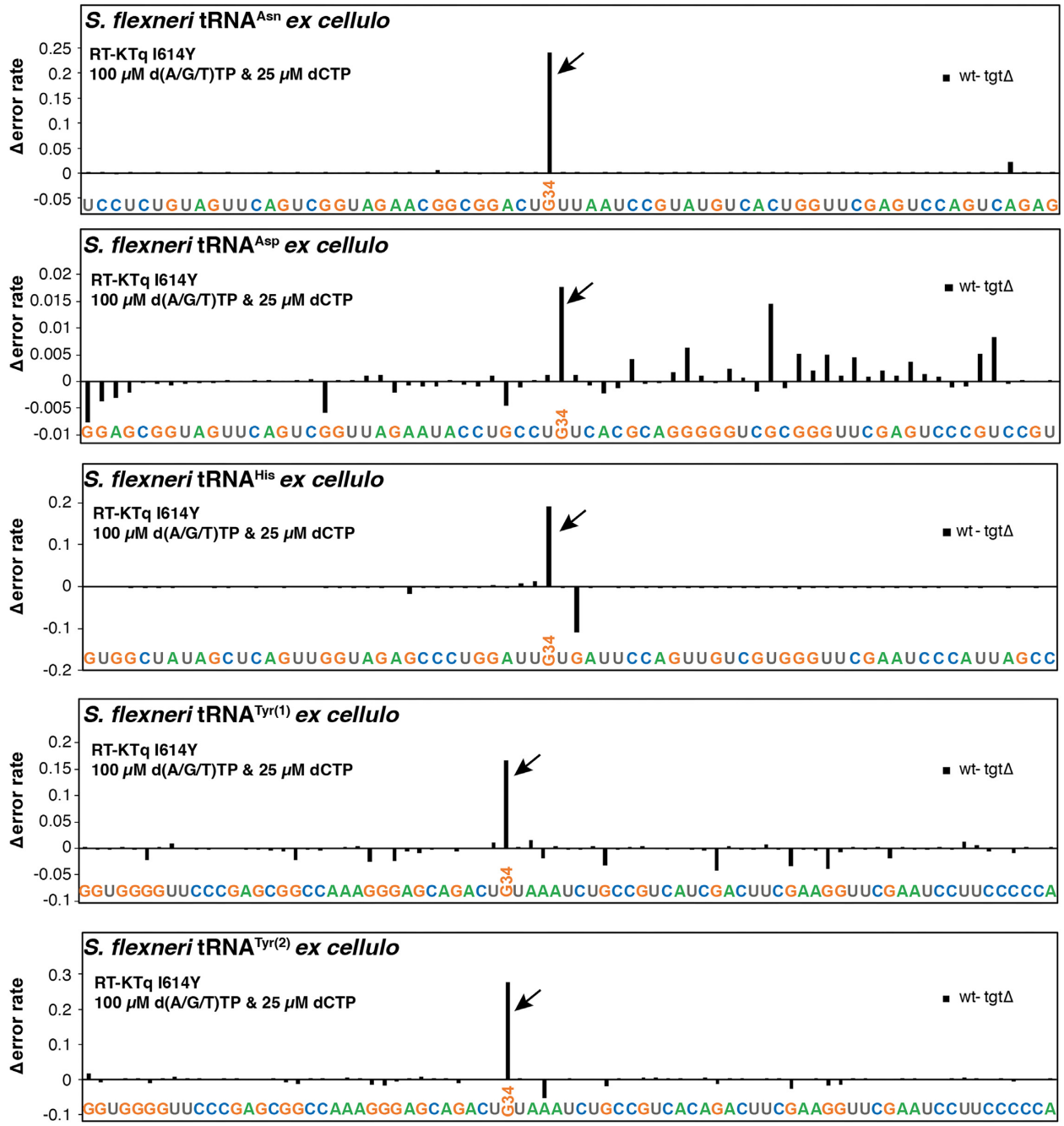


Figure 6. Detection of Q signature in tRNAs from *S. flexneri*. tRNAs were amplified from wt and *tgtΔ* cells of *S. flexneri*, reverse transcribed with RT-KTq I614Y (25 μM dCTP) and processed as in Figure 5. Plots show the difference in misincorporation (wt − *tgtΔ*). Each sample was sequenced once.

acteristics. Indeed, amino acid I614 forms part of the hydrophobic pocket that binds the nucleobase and deoxyribose of the incoming dNTP and mutations at this site were identified as being able to alter nucleobase pairing fidelity (60). Our multiple nucleotide incorporation studies indicate that RT-KTq I614Y stalls after nucleotide misincorporation opposite Ψ, which hampers some applications requiring efficient mismatch extension. Further enzyme engineering might overcome this obstacle for Ψ sequencing. Nevertheless, our results add a new DNA polymerase to the tool box of enzymes to be employed in epitranscriptome analysis and further demonstrate that DNA polymerase engineering is a viable strategy to obtain enzymes with suitable properties.

DATA AVAILABILITY

High-throughput screening data are available in the NCBI GEO database: records GSE217198 and GSE216921. All data needed to evaluate the conclusions in the paper are present in the paper and/or in the Supplementary Data. Crystal structure information is available at the RCSB Pro-

tein Data Bank under the accession number 4BWM. Additional data may be requested from the authors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

L.B.H. and M.H. gratefully acknowledge support by the Konstanz Research School Chemical Biology. The authors thank Ralf Ficner for reagents and Bärbel Raupach and Arturo Zychlinsky for help with *S. flexneri* work.

FUNDING

Deutsche Forschungsgemeinschaft [EH237/13-1, EH237/13-2, MA 2288/17-1 and MA 2288/17-2]. Funding for open access charge: University of Konstanz.
Conflict of interest statement. None declared.

REFERENCES

1. Roundtree,I.A., Evans,M.E., Pan,T. and He,C. (2017) Dynamic RNA modifications in gene expression regulation. *Cell*, **169**, 1187–1200.
2. Barbieri,I. and Kouzarides,T. (2020) Role of RNA modifications in cancer. *Nat. Rev. Cancer*, **20**, 303–322.
3. Wilkinson,E., Cui,Y.H. and He,Y.Y. (2022) Roles of RNA modifications in diverse cellular functions. *Front. Cell Dev. Biol.*, **10**, 828683.
4. Motorin,Y. and Helm,M. (2019) Methods for RNA modification mapping using deep sequencing: established and new emerging technologies. *Genes (Basel)*, **10**, 35.
5. Marchand,V., Bourguignon-Igel,V., Helm,M. and Motorin,Y. (2022) Analysis of pseudouridines and other RNA modifications using HydraPsiSeq protocol. *Methods*, **203**, 383–391.
6. Ovcharenko,A. and Rentmeister,A. (2018) Emerging approaches for detection of methylation sites in RNA. *Open Biol.*, **8**, 180121.
7. Hartstock,K. and Rentmeister,A. (2019) Mapping N^6-methyladenosine (m^6A) in RNA: established methods, remaining challenges, and emerging approaches. *Chemistry*, **25**, 3455–3464.
8. Marchand,V., Blanloeil-Oillo,F., Helm,M. and Motorin,Y. (2016) Illumina-based RiboMethSeq approach for mapping of 2′-*O*-Me residues in RNA. *Nucleic Acids Res.*, **44**, e135.
9. Lin,S., Liu,Q., Jiang,Y.Z. and Gregory,R.I. (2019) Nucleotide resolution profiling of m^7G tRNA modification by TRAC-Seq. *Nat. Protoc.*, **14**, 3220–3242.
10. Yoluç,Y., Ammann,G., Barraud,P., Jora,M., Limbach,P.A., Motorin,Y., Marchand,V., Tisné,C., Borland,K. and Kellner,S. (2021) Instrumental analysis of RNA modifications. *Crit. Rev. Biochem. Mol. Biol.*, **56**, 178–204.
11. Schwartz,S. and Motorin,Y. (2017) Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biol.*, **14**, 1124–1137.
12. Helm,M. and Motorin,Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.
13. Tserovski,L., Marchand,V., Hauenschild,R., Blanloeil-Oillo,F., Helm,M. and Motorin,Y. (2016) High-throughput sequencing for 1-methyladenosine (m^1A) mapping in RNA. *Methods*, **107**, 110–121.
14. Kietrys,A.M., Velema,W.A. and Kool,E.T. (2017) Fingerprints of modified RNA bases from deep sequencing profiles. *J. Am. Chem. Soc.*, **139**, 17074–17081.
15. Vandivier,L.E., Anderson,Z.D. and Gregory,B.D. (2019) HAMR: high-throughput annotation of modified ribonucleotides. *Methods Mol. Biol.*, **1870**, 51–67.
16. Wang,J., Toffano-Nioche,C., Lorieux,F., Gautheret,D. and Lehmann,J. (2021) Accurate characterization of *Escherichia coli* tRNA modifications with a simple method of deep-sequencing library preparation. *RNA Biol.*, **18**, 33–46.
17. Behm-Ansmant,I., Helm,M. and Motorin,Y. (2011) Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids*, **2011**, 408053.
18. Helm,M., Schmidt-Dengler,M.C., Weber,M. and Motorin,Y. (2021) General principles for the detection of modified nucleotides in RNA by specific reagents. *Adv. Biol. (Weinh.)*, **5**, e2100866.
19. Cohn,W.E. and Volkin,E. (1951) Nucleoside-5′-phosphates from ribonucleic acid. *Nature*, **167**, 483–484.
20. Charette,M. and Gray,M.W. (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*, **49**, 341–351.
21. Carlile,T.M., Rojas-Duran,M.F., Zinshteyn,B., Shin,H., Bartoli,K.M. and Gilbert,W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
22. Schwartz,S., Bernstein,D.A., Mumbach,M.R., Jovanovic,M., Herbst,R.H., León-Ricardo,B.X., Engreitz,J.M., Guttman,M., Satija,R., Lander,E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
23. Lovejoy,A.F., Riordan,D.P. and Brown,P.O. (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One*, **9**, e110799.
24. Zhou,K.I., Clark,W.C., Pan,D.W., Eckwahl,M.J., Dai,Q. and Pan,T. (2018) Pseudouridines have context-dependent mutation and stop rates in high-throughput sequencing. *RNA Biol.*, **15**, 892–900.
25. Khoddami,V., Yerra,A., Mosbruger,T.L., Fleming,A.M., Burrows,C.J. and Cairns,B.R. (2019) Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl Acad. Sci. U.S.A.*, **116**, 6784–6789.
26. Marchand,V., Pichot,F., Neybecker,P., Ayadi,L., Bourguignon-Igel,V., Wacheul,L., Lafontaine,D.L.J., Pinzano,A., Helm,M. and Motorin,Y. (2020) HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res.*, **48**, e110.
27. Fergus,C., Barnes,D., Alqasem,M.A. and Kelly,V.P. (2015) The queuine micronutrient: charting a course from microbe to man. *Nutrients*, **7**, 2897–2929.
28. Ehrenhofer-Murray,A.E. (2017) Cross-talk between Dnmt2-dependent tRNA methylation and queuosine modification. *Biomolecules*, **7**, 14.
29. Tuorto,F., Legrand,C., Cirzi,C., Federico,G., Liebers,R., Müller,M., Ehrenhofer-Murray,A.E., Dittmar,G., Gröne,H.J. and Lyko,F. (2018) Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.*, **37**, e99777.
30. Müller,M., Legrand,C., Tuorto,F., Kelly,V.P., Atlasi,Y., Lyko,F. and Ehrenhofer-Murray,A.E. (2019) Queuine links translational control in eukaryotes to a micronutrient from bacteria. *Nucleic Acids Res.*, **47**, 3711–3727.
31. Patel,B.I., Heiss,M., Samel-Pommerencke,A., Carell,T. and Ehrenhofer-Murray,A.E. (2022) Queuosine salvage in fission yeast by Qng1-mediated hydrolysis to queuine. *Biochem. Biophys. Res. Commun.*, **624**, 146–150.
32. Okada,N., Yasuda,T. and Nishimura,S. (1977) Detection of nucleoside Q precursor in methyl-deficient *E. coli* tRNA. *Nucleic Acids Res.*, **4**, 4063–4075.
33. Costa,A., Païs de Barros,J.P., Keith,G., Baranowski,W. and Desgres,J. (2004) Determination of queuosine derivatives by reverse-phase liquid chromatography for the hypomodification study of Q-bearing tRNAs from various mammal liver cells. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, **801**, 237–247.
34. Igloi,G.L. and Kössel,H. (1985) Affinity electrophoresis for monitoring terminal phosphorylation and the presence of queuosine in RNA. Application of polyacrylamide containing a covalently bound boronic acid. *Nucleic Acids Res.*, **13**, 6881–6898.
35. Zhang,W., Xu,R., Matuszek,Ż., Cai,Z. and Pan,T. (2020) Detection and quantification of glycosylated queuosine modified tRNAs by acid denaturing and APB gels. *RNA*, **26**, 1291–1298.
36. Katanski,C.D., Watkins,C.P., Zhang,W., Reyer,M., Miller,S. and Pan,T. (2022) Analysis of queuosine and 2-thio tRNA modifications by high throughput sequencing. *Nucleic Acids Res.*, **50**, e99.
37. Zaringhalam,M. and Papavasiliou,F.N. (2016) Pseudouridylation meets next-generation sequencing. *Methods*, **107**, 63–72.

38. Zhao,L.Y., Song,J., Liu,Y., Song,C.X. and Yi,C. (2020) Mapping the epigenetic modifications of DNA and RNA. *Protein Cell*, **11**, 792–808.
39. Aschenbrenner,J. and Marx,A. (2016) Direct and site-specific quantification of RNA 2′-*O*-methylation by PCR with an engineered DNA polymerase. *Nucleic Acids Res.*, **44**, 3495–3502.
40. Aschenbrenner,J., Werner,S., Marchand,V., Adam,M., Motorin,Y., Helm,M. and Marx,A. (2018) Engineering of a DNA polymerase for direct m^6A sequencing. *Angew. Chem. Int. Ed. Engl.*, **57**, 417–421.
41. Müller,M., Hartmann,M., Schuster,I., Bender,S., Thüring,K.L., Helm,M., Katze,J.R., Nellen,W., Lyko,F. and Ehrenhofer-Murray,A.E. (2015) Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine. *Nucleic Acids Res.*, **43**, 10952–10962.
42. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 6640–6645.
43. Murphy,K.C. and Campellone,K.G. (2003) Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.*, **4**, 11.
44. Yuan,Y., Hutinet,G., Valera,J.G., Hu,J., Hillebrand,R., Gustafson,A., Iwata-Reuyl,D., Dedon,P.C. and de Crecy-Lagard,V. (2018) Identification of the minimal bacterial 2′-deoxy-7-amido-7-deazaguanine synthesis machinery. *Mol. Microbiol.*, **110**, 469–483.
45. Bessler,L., Kaur,N., Vogt,L.M., Flemmich,L., Siebenaller,C., Winz,M.L., Tuorto,F., Micura,R., Ehrenhofer-Murray,A.E. and Helm,M. (2022) Functional integration of a semi-synthetic azido-queuosine derivative into translation and a tRNA modification circuit. *Nucleic Acids Res.*, **50**, 10785–10800.
46. Becker,M., Müller,S., Nellen,W., Jurkowski,T.P., Jeltsch,A. and Ehrenhofer-Murray,A.E. (2012) Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Res.*, **40**, 11648–11658.
47. Hauenschild,R., Werner,S., Tserovski,L., Hildebrandt,A., Motorin,Y. and Helm,M. (2016) CoverageAnalyzer (CAn): a tool for inspection of modification signatures in RNA sequencing profiles. *Biomolecules*, **6**, 42.
48. Morgan,M., Anders,S., Lawrence,M., Aboyoun,P., Pagès,H. and Gentleman,R. (2009) ShortRead: a Bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, **25**, 2607–2608.
49. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10–12.
50. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
51. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
52. Blatter,N., Bergen,K., Nolte,O., Welte,W., Diederichs,K., Mayer,J., Wieland,M. and Marx,A. (2013) Structure and function of an RNA-reading thermostable DNA polymerase. *Angew. Chem. Int. Ed. Engl.*, **52**, 11935–11939.
53. Huang,M.M., Arnheim,N. and Goodman,M.F. (1992) Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.*, **20**, 4567–4573.
54. Ayyadevara,S., Thaden,J.J. and Shmookler Reis,R.J. (2000) Discrimination of primer 3′-nucleotide mismatch by Taq DNA polymerase during polymerase chain reaction. *Anal. Biochem.*, **284**, 11–18.
55. Jagla,B., Wiswedel,B. and Coppee,J.Y. (2011) Extending KNIME for next-generation sequencing data analysis. *Bioinformatics*, **27**, 2907–2909.
56. Hillebrand,G.G. and Beattie,K.L. (1985) Influence of template primary and secondary structure on the rate and fidelity of DNA synthesis. *J. Biol. Chem.*, **260**, 3116–3125.
57. Loewen,P.C. and Switala,J. (1995) Template secondary structure can increase the error frequency of the DNA polymerase from *Thermus aquaticus*. *Gene*, **164**, 59–63.
58. Bellodi,C., McMahon,M., Contreras,A., Juliano,D., Kopmar,N., Nakamura,T., Maltby,D., Burlingame,A., Savage,S.A., Shimamura,A. *et al.* (2013) H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell Rep.*, **3**, 1493–1502.
59. Becker,H.F., Motorin,Y., Planta,R.J. and Grosjean,H. (1997) The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res.*, **25**, 4493–4499.
60. Patel,P.H., Kawate,H., Adman,E., Ashbach,M. and Loeb,L.A. (2001) A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.*, **276**, 5044–5051.