RNA Modification

International Edition: DOI: 10.1002/anie.201708276
German Edition: DOI: 10.1002/ange.201708276

A Radiolabeling-Free, qPCR-Based Method for Locus-Specific Pseudouridine Detection

*Zhixin Lei and Chengqi Yi**

Abstract: *Pseudouridine (Ψ) is the most abundant post-transcriptional RNA modification. Methods have been developed for locus-specific Ψ detection; however, they often involve radiolabeling of RNA, require advanced experimental skills, and can be time-consuming. Herein we report a radiolabeling-free, qPCR-based method to rapidly detect locus-specific Ψ. Pseudouridine residues were labeled chemically, and the resulting adducts induced mutation/deletion during reverse transcription (RT) to generate qPCR products with different melting temperatures and hence altered melting curves. We validated our method on known Ψ sites in rRNA and then used it to sensitively detect Ψ residues in lncRNA and mRNA of low abundance. Finally, we applied our method to pseudouridine synthase identification and showed that Ψ616 in PSME2 mRNA is dependent on PUS7. Our facile and cost-effective method takes only 1.5 days to complete, and with slight adjustment it can be applied to the detection of other epitranscriptomic marks.*

Over 100 different types of post-transcriptional RNA modification have been characterized to date.[1] Among them, pseudouridine (Ψ), also known as the "fifth nucleoside" of RNA, is overall the most abundant.[2] Pseudouridine, which has a carbon–carbon bond linking C1′ to uracil C5, is a linkage isomer of uridine and shares an identical molecular weight and a Watson–Crick interface. The extra NH group has endowed Ψ with unique physical/chemical properties, thus allowing Ψ to improve base stacking and stabilize RNA structure.[2] Pseudouridine synthesis can be catalyzed by two types of pseudouridine synthases (PUSs): 1) the RNA-dependent PUSs, which require box H/ACA RNA to act as a guide for substrate recognition; and 2) the RNA-independent PUSs, which require no cofactor to catalyze Ψ formation.

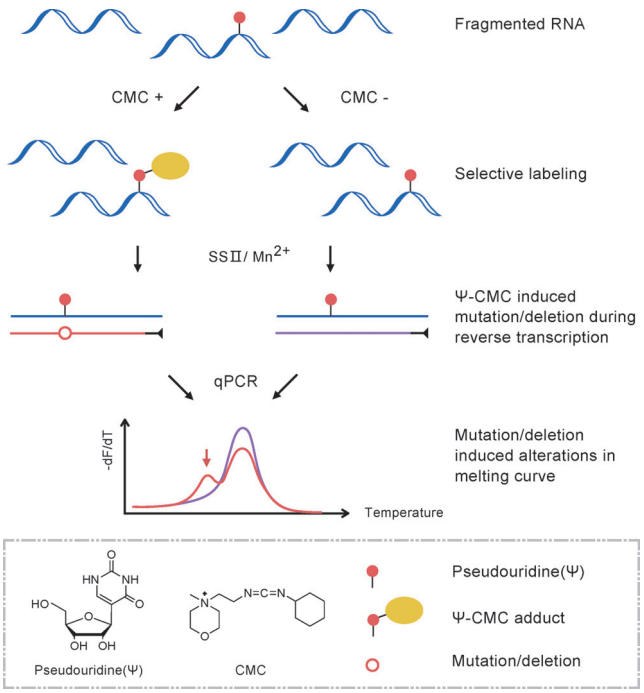
[*] Prof. C. Yi
State Key Laboratory of Protein and Plant Gene Research
School of Life Sciences, Peking-Tsinghua Center for Life Sciences
Department of Chemical Biology and
Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University
Beijing 100871 (China)
E-mail: chengqi.yi@pku.edu.cn
Homepage: http://www.yi-lab.org/
Z. Lei
Peking-Tsinghua Center for Life Sciences
Academy for Advanced Interdisciplinary Studies, Peking University
Beijing 100871 (China)
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.201708276.

Besides noncoding RNA, Ψ is also present in mRNA. By the use of high-throughput sequencing, four recently developed technologies, termed Ψ-seq, Pseudo-seq, CeU-Seq, and PSI-seq, have enabled transcriptome-wide Ψ detection.[3] All four methods are based on the selective chemical labeling of Ψ by *N*-cyclohexyl-*N*′-(2-morpholinoethyl)carbodiimide (CMC).[4] The Ψ–CMC adduct causes the reverse transcriptase (RTase) to stop one nucleotide from the 3′ end of RNA, thereby achieving Ψ detection at single-nucleotide resolution. These studies revealed the dynamic nature of mRNA pseudouridylation in yeast, mouse, and human cells, thus expanding our understanding of the epitranscriptome-mediated regulation of gene expression.

Although these high-throughput methods enabled Ψ detection on a transcriptome-wide scale, they can be costly and require sophisticated bioinformatics analysis. For studies aiming to investigate specific Ψ sites of interest, locus-specific Ψ-detection methods are highly desirable. In fact, the aforementioned CMC-based selective labeling was first utilized in a primer-extension assay to detect locus-specific Ψ sites in rRNA.[5] Such chemistry has also been combined with mass spectrometry to detect Ψ sites, which are otherwise mass-silent.[6] Furthermore, a ligation-based method, which makes use of the different ligation efficiency of T4 DNA ligase in the presence of RNA modifications, has been developed to detect Ψ in rRNA.[7] More recently, a quantitative method termed "SCARLET",[8] which is derived from an RNase H based method previously reported,[9] has been developed. However, these methods have several limitations: the majority of them require radioactive labeling and advanced experimental skills and are often time-consuming, whereas the mass-spectrometry-based method involves a sophisticated instrument.

Herein, we present a facile method for the locus-specific detection of Ψ sites in noncoding RNA and mRNA. This method is based on Ψ–CMC-induced mutation/deletion in cDNA synthesis, thus leading to qPCR products with different melting temperatures (Scheme 1). Unlike the existing methods that rely on RT stops caused by Ψ–CMC, we utilized new conditions of RT that allow read-through of Ψ–CMC, simultaneously generating mutation and/or deletion in cDNA. Such mutation/deletion will result in altered melting curves of the qPCR products, thus enabling the locus-specific detection of Ψ modification.

We first sought to identify RT conditions that allow read-through of the Ψ–CMC adduct. We tested different RTases and found that Superscript II is capable of reading through the Ψ–CMC adduct in the presence of Mn^{2+} but not Mg^{2+}, consistent with the SHAPE-Map results (see Figure S1 a,b in the Supporting Information).[10] Efficient read-through events were observed for a range of Mn^{2+} concentrations (see



Scheme 1. Workflow of the method. Ψ-containing RNA is specifically labeled by CMC, and then reverse transcribed by Superscript II (SSII) with Mn^{2+} buffer. The Ψ–CMC adducts cause SSII to introduce a mutation/deletion at or around the Ψ site in the synthesized cDNA, thus giving rise to a new peak (indicated by the red arrow) in the melting curve of qPCR products.

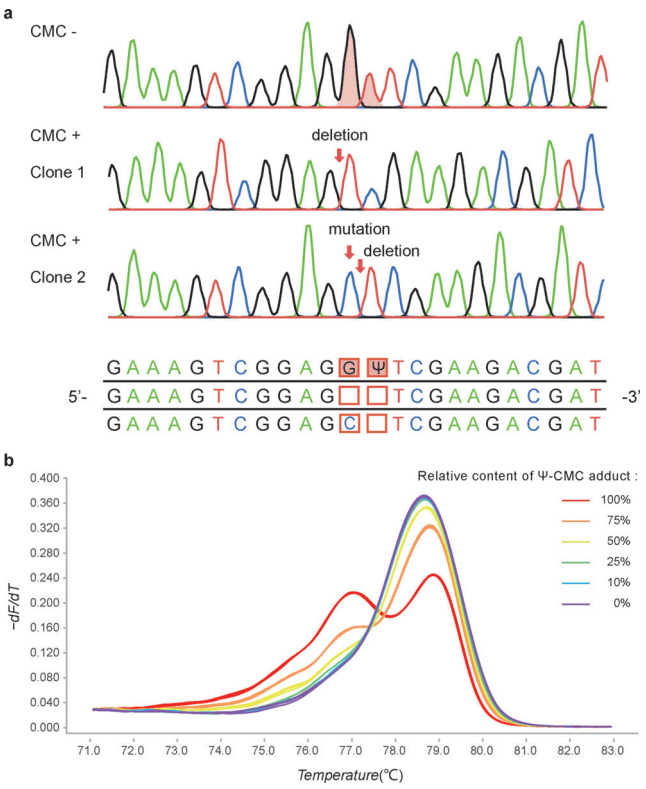


Figure 1. Optimized RT conditions allow misincorporation in cDNA and Ψ detection in rRNA. a) Sanger sequencing results of SSII reading through the CMC adduct at Ψ1045 in 18S rRNA. b) Melting curves of qPCR products containing Ψ1045 in 18S rRNA. The curves were obtained by high-resolution melting analysis. The relative content of the Ψ–CMC adduct refers to the relative ratio of mixed RNA samples. Ψ sites in rRNA are assumed to be 100% modified.

Figure S1c); we decided to use the conditions with 6 mM Mn^{2+} on the basis of the overall consideration of read-through efficiency and fidelity. We then examined the read-through products for Ψ1045 in 18S rRNA and Ψ1582 in 28S rRNA. As expected, Sanger sequencing results showed that our optimized RT conditions gave rise to different types of misincorporation surrounding the modification sites (Figure 1a; see also Figure S1d).

We then analyzed the melting curves of the qPCR products. A shift in the melting temperature could be found in the melting curves (see Figure S2); such a shift was only observed when the Ψ–CMC adduct, Superscript II, and Mn^{2+} were all present, thus suggesting that the shift is specifically induced by the misincorporation in cDNA. To improve sensitivity, we turned to high-resolution melting analysis (HRM), which is commonly used for polymorphism detection and genotyping.[11] Evident changes were observed in the melting curves with HRM (Figure 1b; red vs. purple curves); we think such alterations are due to the presence of mixed cDNA amplicons showing different types of misincorporation and hence lower melting temperatures than the unlabeled control sample. By systematically varying the percentage of modification, we showed that the degree of melting-curve alteration quantitatively correlated with the modification level (Figure 1b). Importantly, when rRNA regions devoid of Ψ modification were similarly analyzed by HRM, the melting curves of CMC-treated samples were almost the same as those of the untreated samples (see Figure S3), thus suggesting our approach is capable of detecting Ψ sites in rRNA.

Because the length of the amplicons can influence the results of HRM analysis, we next optimized the length of the amplicons to further improve the sensitivity of our method. We compared different length of amplicons (ca. 60–70 bp, ca. 80–90 bp, and ca. 140–150 bp) for four known Ψ sites in rRNA and observed greater alteration of melting curves as the amplicons became shorter (see Figure S4). Assuming a 100% modification level for Ψ sites in rRNA, Ψ sites with an approximately 10% modification level (sample created by mixing the modified with unmodified RNAs) could still be distinguished from the untreated samples when amplicons were shorter than 90 bp (see Figure S4). Although the choice of amplicon length for specific Ψ sites of interest could also be affected by their sequence context, our results indicated that shorter amplicons can in general allow improved detection sensitivity.

Besides detection of the aforementioned single Ψ sites in rRNA, our method also allowed detection of Ψ sites that are adjacent to each other. Previous methods relying on RT stops favor detection of Ψ sites that are located at the 3′ end of RNA, thus limiting the detection of neighboring Ψ sites in RNA. We designed primers to cover Ψ1045 and Ψ1056 in 18S rRNA, and observed the presence of an additional peak in the melting curves as compared to that of Ψ1045 alone (Figure 1b; see also Figure S5a). We further subjected the qPCR

amplicons to high-throughput sequencing, and we were able to simultaneously capture the two Ψ sites at positions 1045 and 1056 (see Figure S5b). Similarly, when we examined Ψ1781, Ψ1782, and Ψ1792 in 28S rRNA simultaneously, we found an additional alteration as compared to that of Ψ1792 alone (see Figure S5c,d).

To demonstrate that read-through-induced misincorporation is a general feature of Ψ sites located in different sequence contexts, we utilized high-throughput sequencing to comprehensively examine the pattern of mutation/deletion for all Ψ sites in rRNA. First, an average stop rate of approximately 7.7% was observed for the 98 Ψ sites in rRNA (as compared to the average stop rate of approximately 33.1% in our original CeU-seq data),[3d] thus demonstrating that read-through is a dominant event under our optimized RT conditions (see Figure S6). Second, misincorporation events, particularly mutation events, are specific and precise for Ψ sites but not other nucleotides in RNA (Figure 2a; see also Figure S7a). Third, although the exact misincorporation rate varies for each site, an average misincorporation rate of 33.9% was observed for the Ψ sites, which reflects the overall effect of the modification level, CMC-labeling yield, and read-through efficiency. Mutation and deletion account for approximately 14.3 and 18.3% of the total 33.9% misincorporation events, respectively (Figure 2b,d). Fourth, we analyzed the pattern of mutation/deletion for all Ψ sites. Similar frequencies of Ψ mutation to A, C, or G were observed, and the mutation signature of Ψ was slightly dependent on the identity of its 3′ nucleotide, but not the 5′ nucleotide in RNA (Figure 2c; see also Figure S8). For deletion events, 1 bp deletion appears to be the most frequent case, followed by 2 bp and 3 bp deletion (Figure 2e; see also Figure S7b). Taken together, these observations suggest that misincorporation induced by the Ψ–CMC adduct is a common phenomenon and can be used to identify Ψ sites embedded in different sequence contexts.

We then aimed to apply our method to detect Ψ sites in mRNA and long noncoding RNA (lncRNA). We chose the human *EEF1A1* mRNA, which contains two Ψ sites at position 519 and 875; U519 was shown to be modified by approximately 56% by SCARLET.[3d] Alterations to the melting curves were observed for both sites (Figure 3a); the degree of alteration is not as strong as in rRNA, probably owing to the lower modification level at these mRNA sites. We also validated Ψ519 by Sanger sequencing and high-throughput sequencing (see Figures S9a and S10a). Importantly, negligible alteration was observed for the control regions that are devoid of Ψ modification (Figure 3a). We also showed that even without polyA + selection (using total RNA isolated from cells directly), our method can also readily detect Ψ sites in mRNA (Figure 3a; see also Figure S11). For lncRNA, we applied our method to position 5590 of MALAT1, a metastasis-associated lung adenocarcinoma transcript, and found a clear shift (see Figure S12). This Ψ site was also validated by both Sanger sequencing and high-throughput sequencing (see Figures S9b and S10b). By comparing the melting curves derived from the endogenous *MALAT1* RNA to those derived from model RNAs containing a predetermined level of modification at position 5590, we further estimated that Ψ5590 of *MALAT1* is modified to a level of about 75% in vivo (see Figure S12).

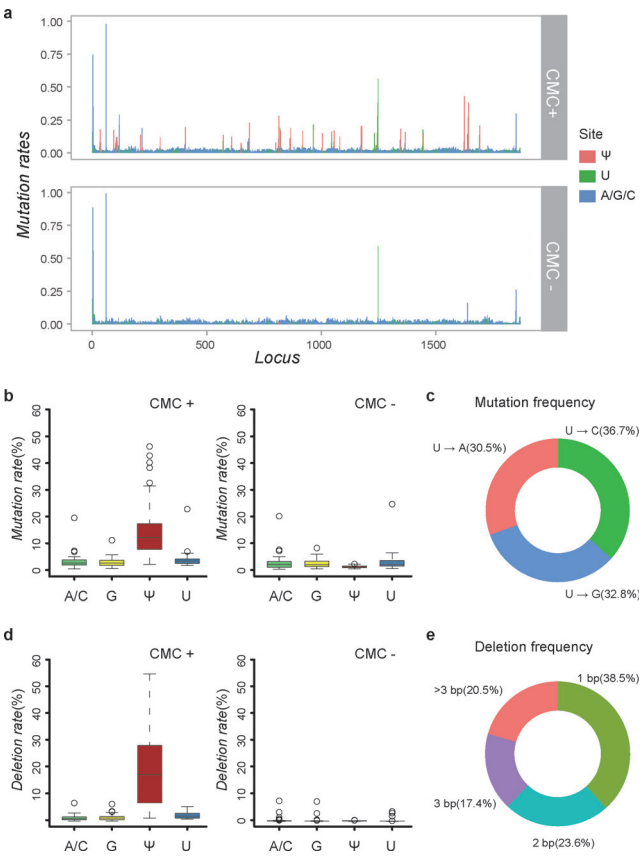


Figure 2. Mutation/deletion is a general feature of Ψ–CMC adducts under improved RT conditions. a) Read-through events generated high mutation rates precisely at the sites of Ψ modification. Mutation rates of 18S rRNA are shown, with identified Ψ sites marked with red bars, unmodified U sites with green, and A/G/C sites with blue. The sites with a high mutation rate in the CMC(−) sample (CMC-independent) are single-nucleotide polymorphisms (SNPs) and other RNA modifications. For example, the green bar at position 1248 corresponds to a 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine residue, whereas the blue bar near the 3′ end (position 1851) corresponds to a $N^{6,6}$-dimethyladenosine residue. b) Mutation rates were significantly increased at the known Ψ sites. c) Pie plot of the mutation frequency of U to A, G, or C. d) Deletion rates were significantly increased at the known Ψ sites. e) Pie plot of the frequency of 1 bp, 2 bp, 3 bp, or >3 bp deletion.

We further applied our method to detect Ψ sites in mRNAs of lower abundance. We chose the mRNA of *RPL18A* (reads per kilobase per million mapped reads, or rpkm ≈ 1600) and *HPRT1* (rpkm ≈ 110), which are expressed at a level of approximately 80% and 5.5% of *GAPDH* from RNA-seq data of HEK293T cells. We found that as long as the mRNA had enough abundance to produce a reliable melting curve in qPCR, a reproducible shift in the melting curve could be observed (see Figure S13). Thus, our qPCR-based method is very sensitive for Ψ detection.

Finally, we used our locus-specific approach to identify Ψ synthase responsible for mRNA pseudouridylation. We focused on *PSME2* (rpkm ≈ 30), whose expression level is only about 1.5% that of *GAPDH* and is the lowest of all the

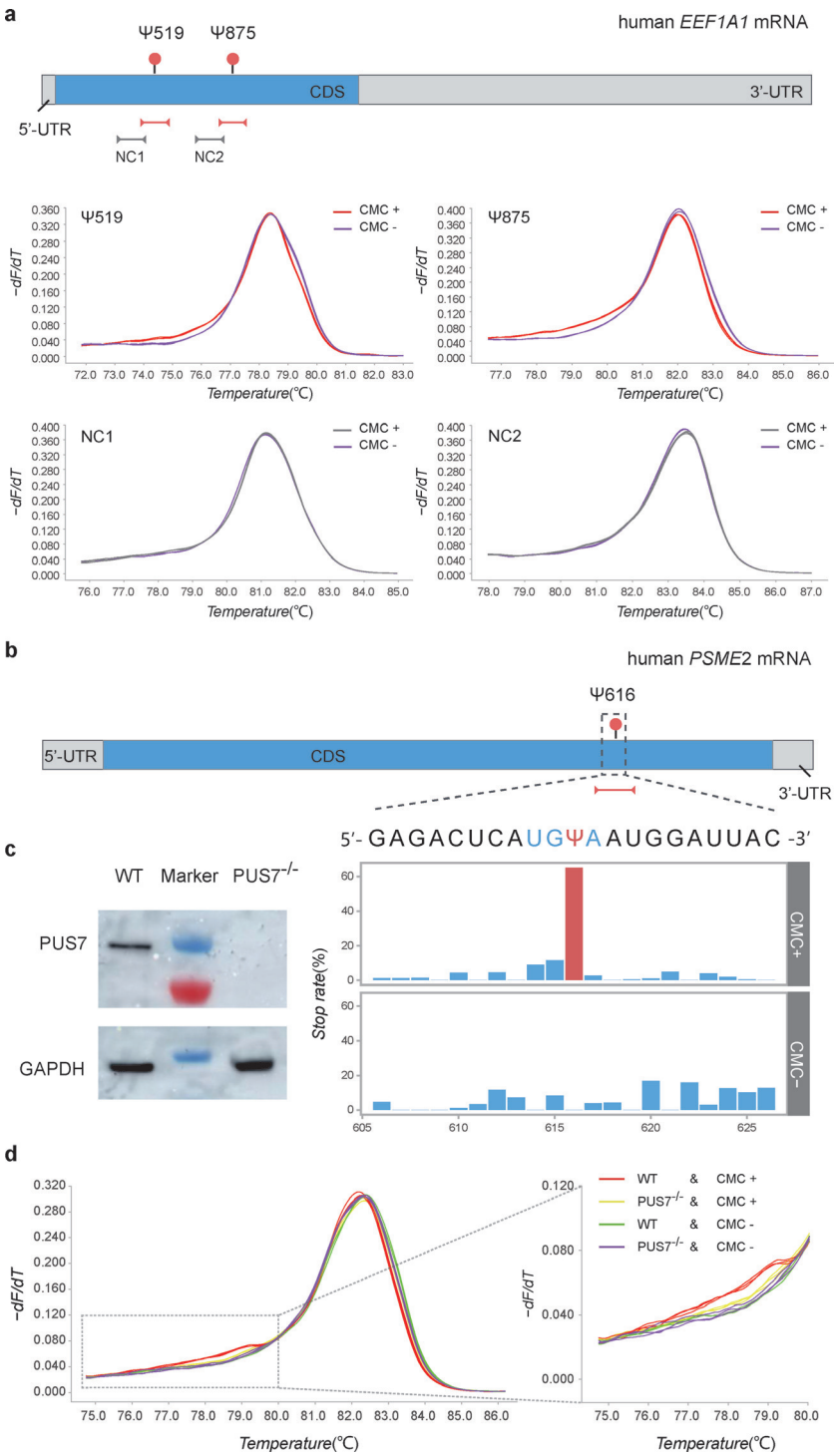


Figure 3. Our method sensitively detects Ψ sites in mRNA and lncRNA. a) HRM results for the two Ψ sites in human *EEF1A1* mRNA, determined directly from total RNA. b) Stop-rate-difference map of human *PSME2* mRNA, based on the published data of CeU-Seq. The UGΨA motif is highlighted. c) Western blot result of the PUS7 knock-out cell line. d) HRM results for position 616 in human *PSME2* mRNA, from the wild-type (WT) and PUS7$^{-/-}$ cell lines.

mRNAs examined in this study. *PSME2* mRNA contains a Ψ site at position 616; however, the enzyme responsible for this modification is unknown. Ψ616 is located in a UGUA context (Figure 3b), which is thought to be a consensus motif for the Ψ synthase PUS7.[12] To test this hypothesis, we generated PUS7$^{-/-}$ knock-out (KO) cells by using the CRISPR/Cas9 system and validated the absence of PUS7 protein by western blot analysis (Figure 3c). We then applied our method to RNA isolated from the KO cells and observed the disappearance of the melting-curve alteration (Figure 3d). This result provides the first experimental evidence that Ψ616 in *PSME2* mRNA is dependent on PUS7.

In conclusion, we have described a cost-effective, radiolabeling-free, qPCR-based method for the rapid detection of Ψ sites in a locus-specific manner. For any given Ψ candidates, our method only requires a commercially available reagent for Ψ labeling and primers for qPCR, and can be completed within 1.5 days. We demonstrated the utility of our method in detecting Ψ sites not only in abundant rRNA but also in mRNA of low abundance. We also showed that our method is capable of identifying enzymes responsible for Ψ sites of particular interest. Although our method is not quantitative, the modification level of a given site could be estimated with proper standardization. Furthermore, our method can be readily combined with sequencing to obtain the modification site at single-base resolution. Finally, because read-through-induced misincorporation is also known for many other types of RNA modifications,[13] particularly those with an altered Watson–Crick face (such as N^1-methyladenosine, N^1-methylguanosine, $N^{2,2}$-dimethylguanosine, N^3-methylcytosine, and N^3-methyluridine), with slight adjustment our approach could also be readily applied to detect more RNA modifications in the transcriptome.

Experimental Section

See the Supporting Information for experimental details. Sequencing data have been deposited with the Gene Expression Omnibus (GEO) under the accession number GSE102476.

Acknowledgements

We thank Xushen Xiong for sharing the demultiplexing script, Shiqing Ma for assistance with reverse-transcription optimization, Lixia Wang for generating the PU7$^{-/-}$ cell line, and the Protein Core at the School of Life Sciences. This research was supported by

Beijing Natural Science Foundation (5162012), the National Natural Science Foundation of China (21522201 and 21472009), and the National Basic Research Foundation of China (2016YFC0900301).

Conflict of interest

The authors declare no conflict of interest.

Keywords: epitranscriptomics · gene expression · high-resolution melting analysis · nucleosides · RNA modification

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 14878–14882
Angew. Chem. **2017**, *129*, 15074–15078

[1] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262 – D267.

[2] a) X. Li, S. Ma, C. Yi, *Curr. Opin. Chem. Biol.* **2016**, *33*, 108 – 116; b) J. Karijolich, C. Yi, Y. T. Yu, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 581 – 585; c) F. Spenkuch, Y. Motorin, M. Helm, *RNA Biol.* **2014**, *11*, 1540 – 1554; d) M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341 – 351.

[3] a) T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert, *Nature* **2014**, *515*, 143 – 146; b) A. F. Lovejoy, D. P. Riordan, P. O. Brown, *PloS one* **2014**, *9*, e110799; c) S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, A. Regev, *Cell* **2014**, *159*, 148 – 162; d) X. Li, P. Zhu, S. Ma, J. Song, J. Bai, F. Sun, C. Yi, *Nat. Chem. Biol.* **2015**, *11*, 592 – 597.

[4] N. W. Ho, P. T. Gilham, *Biochemistry* **1967**, *6*, 3632 – 3639.

[5] A. Bakin, J. Ofengand, *Biochemistry* **1993**, *32*, 9754 – 9762.

[6] B. Addepalli, P. A. Limbach, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1363 – 1372.

[7] a) M. Saikia, Q. Dai, W. A. Decatur, M. J. Fournier, J. A. Piccirilli, T. Pan, *RNA* **2006**, *12*, 2025 – 2033; b) Q. Dai, R. Fong, M. Saikia, D. Stephenson, Y. T. Yu, T. Pan, J. A. Piccirilli, *Nucleic Acids Res.* **2007**, *35*, 6322 – 6329.

[8] N. Liu, M. Parisien, Q. Dai, G. Q. Zheng, C. He, T. Pan, *RNA* **2013**, *19*, 1848 – 1856.

[9] Y. T. Yu, M. D. Shu, J. A. Steitz, *RNA* **1997**, *3*, 324 – 331.

[10] N. A. Siegfried, S. Busan, G. M. Rice, J. A. E. Nelson, K. M. Weeks, *Nat. Methods* **2014**, *11*, 959 – 965.

[11] a) I. Simko, *Trends Plant Sci.* **2016**, *21*, 528 – 537; b) M. Tamburro, G. Ripabelli, *Ann. Ig.: medicina preventiva e di comunita* **2017**, *29*, 293 – 316; c) C. T. Wittwer, G. H. Reed, C. N. Gundry, J. G. Vandersteen, R. J. Pryor, *Clin. Chem.* **2003**, *49*, 853 – 860.

[12] I. Behm-Ansmant, A. Urban, X. J. Ma, Y. T. Yu, Y. Motorin, C. Branlant, *RNA* **2003**, *9*, 1371 – 1382.

[13] a) W. V. Gilbert, T. A. Bell, C. Schaening, *Science* **2016**, *352*, 1408 – 1412; b) M. Helm, Y. Motorin, *Nat. Rev. Genet.* **2017**, *18*, 275 – 291; c) I. A. Roundtree, M. E. Evans, T. Pan, C. He, *Cell* **2017**, *169*, 1187 – 1200; d) M. Schaefer, U. Kapoor, M. F. Jantsch, *Open Biol.* **2017**, *7*, 170077.

Manuscript received: August 11, 2017
Accepted manuscript online: September 28, 2017
Version of record online: October 19, 2017

