Sequencing Methods

International Edition: DOI: 10.1002/anie.201700537
German Edition: DOI: 10.1002/ange.201700537

Selective Enzymatic Demethylation of N^2,N^2-Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing

*Qing Dai[+], Guanqun Zheng[+], Michael H. Schwartz[+], Wesley C. Clark, and Tao Pan**

Abstract: *The abundant Watson–Crick face methylations in biological RNAs such as N^1-methyladenosine (m^1A), N^1-methylguanosine (m^1G), N^3-methylcytosine (m^3C), and N^2,N^2-dimethylguanosine (m^2_2G) cause significant obstacles for high-throughput RNA sequencing by impairing cDNA synthesis. One strategy to overcome this obstacle is to remove the methyl group on these modified bases prior to cDNA synthesis using enzymes. The wild-type E. coli AlkB and its D135S mutant can remove most of m^1A, m^1G, m^3C modifications in transfer RNA (tRNA), but they work poorly on m^2_2G. Here we report the design and evaluation of a series of AlkB mutants against m^2_2G-containing model RNA substrates that we synthesize using an improved synthetic method. We show that the AlkB D135S/L118V mutant efficiently and selectively converts m^2_2G modification to N^2-methylguanosine (m^2G). We also show that this new enzyme improves the efficiency of tRNA sequencing.*

tRNAs decode mRNA codons and are essential for cells. In humans, tRNA abundance is tissue dependent[1] and tRNA expression and mutations are known to be associated with neurological pathologies and cancer development and proliferation.[2] Efficient and quantitative tRNA sequencing methods are crucial for biological studies of tRNA. However, standard sequencing methods are very ineffective for tRNA-seq, primarily due to the high levels of four Watson–Crick face methylations, m^1A, m^1G, m^3C and m^2_2G that impair cDNA synthesis. Recently, we reported an efficient and quantitative high-throughput tRNA sequencing method (DM-tRNA-seq) by using a highly processive thermostable group II intron reverse transcriptase, and importantly, two demethylases that remove most base methylations and thus significantly improved the tRNA sequencing efficiency: for example, the amount of longer and full-length cDNA was substantially increased after the treatment of tRNA with demethylases.[3] Despite this success, our sequencing data showed that the m^2_2G at position 26 of tRNAs was not effectively demethylated and caused a significant stop in cDNA synthesis. To alleviate this problem, it would be highly desirable to identify a third demethylase with enhanced activity toward m^2_2G in RNA and apply it for tRNA sequencing.

Previously, Zhu et al. showed that mutating active-site residues of an RNA/DNA demethylase switched its demethylation activity from m^1A to m^6A modification.[4] We have also shown that an AlkB D135S mutant demethylates m^1G much more efficiently than the wild-type enzyme.[3] Therefore, we reasoned that further mutagenesis of residues around the active site of AlkB might produce a demethylase with higher activity for our target m^2_2G modification.

Unlike m^1A, m^1G and m^3C which only have one methyl group in their structures, m^2_2G has two methyl groups at the exocyclic amine (Figure 1 A). For sequencing purposes, removal of one methyl group to N^2-methylguanosine (m^2G) should be sufficient for highly efficient cDNA synthesis by the thermophilic reverse transcriptase.[3] Indeed, we found that m^2_2G blocks the reverse transcriptase (RT) reaction of thermostable group II intron reverse transcriptase while it fully reads through m^2G modification (Figure S1 in the Supporting Information). m^2_2G is present in about half of human tRNAs at position 26 at the junction between the D-arm and the anticodon arm,[5] as indicated in the tertiary structures of tRNA (Figure 1 B). m^2_2G26 can form non-canonical base pairing with A44, which has a destabilizing effect on RNA duplexes, as revealed by the crystal structure

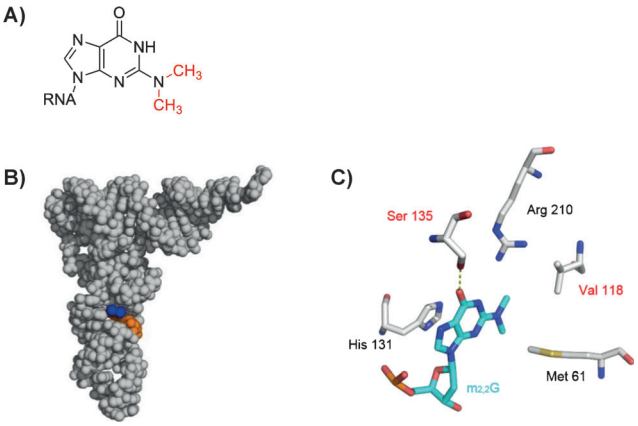


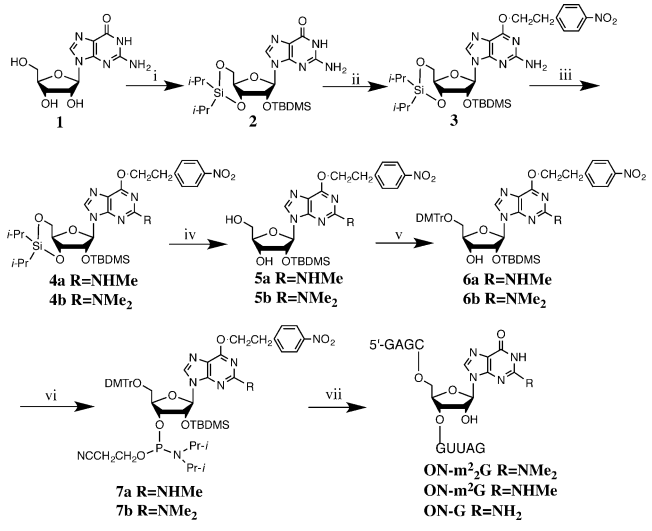
Figure 1. AlkB mutant design for m^2_2G demethylation. A) Chemical structure of m^2_2G. B) Location of m^2_2G modification in tRNA tertiary structure (PDB 4tna). The m^2_2G residue is in orange, and the two methyl groups are in blue. C) View of AlkB active site stereochemistry model with m^2_2G coordination based on protein data bank (PDB) ID 3BIE. m^2_2G is labeled in cyan.

[*] Dr. Q. Dai[+]
Department of Chemistry, The University of Chicago
Chicago, IL 60637 (USA)
Dr. G. Zheng,[+] Dr. M. H. Schwartz,[+] Dr. W. C. Clark, Prof. T. Pan
Department of Biochemistry & Molecular Biology
The University of Chicago
Chicago, IL 60637 (USA)
E-mail: taopan@uchicago.edu
[+] These authors contributed equally to this work.

Supporting information for this article can be found under: http://dx.doi.org/10.1002/anie.201700537.

of an RNA duplex with m^2_2G:A pairs.[6] The m^2_2G26 residue is buried in the structure, making it particularly difficult to be demethylated. The co-crystal structure of *E. coli* AlkB with a nucleic acid substrate[7] was used to model the m^2_2G base into the active site (Figure 1C). AlkB D135S mutant we used previously for DM-tRNA-seq works efficiently toward m^1G, in which the shorter side chain of S135 not only creates more room to accommodate guanosine but also forms the crucial hydrogen bond to stabilize guanosine in the active site. We reasoned that shortening the side chain of amino acids within the van der Waals ratios of the methyl groups on m^2_2G would allow better accommodation for m^2_2G, while leaving the rest of the active site intact would potentially enhance the activity towards m^2_2G. Therefore, three mutants, Arg210-to-Lys, Leu118-to-Val, and Met61-to-Ser were prepared.

To test the activity of these potential m^2_2G demethylases, we synthesized three 9mer RNA oligos as demethylase substrates. The oligo sequence mimics the human $tRNA^{Phe}$ sequence around the position of 26 (5′-GAGCXUUAG, X = m^2_2G, m^2G or G) (Scheme 1). It was generally believed that



Scheme 1. Improved syntheses of m^2G and m^2_2G phosphoramidites and their incorporation into RNA to prepare demethylase substrates. i) a) $(t\text{-Bu})_2Si(OTf)_2$, DMF; b) imidazole, TBDMS-Cl, 90 °C, 75 %. ii) *p*-nitrophenylethanol, Ph_3P, DIAD, 1,4-dioxane 100 °C, 92 %. iii) $(HCHO)_x$, $NaBH_3CN$, AcOH, 40 °C. iv) HF·Py/THF. v) DMTr-Cl/Py. vi) $(i\text{-Pr})_2NP(Cl)OCH_2CH_2CN$, CH_2Cl_2, $(i\text{-Pr})_2NEt$, 1-methyl-imidazole. vii) RNA synthesis and deprotection.

the protection of O^6 of guanosine with *p*-nitrophenylethyl (NPE) group would not only lower the polarity and increase the solubility of guanosine derivatives to facilitate their purification, but also enhance the nucleophilicity of the exocyclic amine.[8] Subsequent conversion of 2-amine to 2-methylamine and 2-dimethylamine were accomplished by converting 2-NH_2 to 2-F, followed by substitution of 2-F with methylamine and dimethylamine. The use of fluoride reagent to introduce 2-F precluded the application of Beigelman's optimal silyl protecting strategy for phosphoramidite synthesis[9] and therefore several extra steps were needed for protecting group exchanges.

Several methods have been reported for preparing m^2G and m^2_2G derivatives[10] but the most straightforward one used a "one-step" reductive amination reaction to methylate the exocyclic amine directly.[10b] We attempted to combine the reductive amination with Beigelman's protecting strategy to synthesize both phosphoramidites **7a** and **7b** at the same time. We started our synthesis from guanosine **1** by selective protecting 3′,5′-hydroxyls with di-*tert*-butylsilyl group and 2′-hydroxy with TBDMS in a "one-pot" reaction to give intermediate **2** in 75 % yield.[9] Protection of O^6 using *p*-nitrophenylethyl group by the treatment of **2** with *p*-nitrophenylethanol in the presence of triphenylphosphine (Ph_3P) and diisopropyl azodicarboxylate (DIAD) gave **3** in 92 % yield.[11] The reductive amination reaction between the exocyclic amine in **3** with paraformaldehyde and sodium cyanoborohydride ($NaBH_3CN$) in acetic acid at 40 °C resulted in sequential methylation of exocyclic amine to give a mixture of **4a** and **4b**. Depending on the amount of paraformaldehyde and $NaBH_3CN$ used, either **4a** or **4b** could be obtained as major products. All the protecting groups in intermediate **3** remained stable under the acidic conditions. Treatment of **4a**/**4b** with hydrogen fluoride pyridine removed the 3′,5′-di-*tert*-butylsilyl group quantitatively while keeping 2′-TBDMS intact to give intermediates **5a**/**5b**. 5′-Tritylation gave the mixture of **6a**/**6b**. Although two types of methylated derivatives could be readily separated by column chromatography at **4a**/**4b** or **5a**/**5b** stage, it is more efficient to separate **6a** and **6b** by column chromatography. Subsequent 3′-phosphitylation **6a** and **6b** under standard conditions provided m^2_2G phosphoramidite **7a** and m^2G phosphoramidite **7b** in 88 % and 85 % yield, respectively.

Phosphoramidite **7a** or **7b** was then incorporated into RNA oligos with similar efficiency as the commercial unmodified guanosine phosphoramidite to prepare the oligos **ON-m^2_2G** and **ON-m^2G**. As a control, we also synthesized an unmodified oligo **ON-G** with the same sequence. After deprotection and HPLC purification, the structures of three RNA oligos were confirmed by Maldi-TOF MS, which showed 14 Dalton difference in mass between **ON-m^2_2G** and **ON-m^2G** and between **ON-m^2G** and **ON-G** (Figure 2B).

We tested the demethylation activities of the AlkB mutant proteins using these RNA oligos as substrates. Oligo **ON-m^2_2G** (1 nmol) was incubated with each enzyme in the reaction buffer containing 300 mM KCl, 2 mM $MgCl_2$, 50 μM of $(NH_4)_2Fe(SO_4)_2\cdot 6H_2O$, 300 μM α-ketoglutarate (α-KG), 2 mM L-ascorbic acid, 50 μg mL^{-1} BSA and 50 mM MES buffer (pH 5.0) for 2 h at room temperature, and then quenched by the addition of 5 mM EDTA. RNA oligos in the reaction were recovered by ethanol precipitation and dissolved in water. Part of samples were digested to free nucleosides by treatment of nuclease P1 and alkaline phosphatase and then analyzed by HPLC (Figure 2A), while the rest of samples were directly analyzed by Maldi-TOF MS (Figure 2B). As shown in Figure 2A, the D135S/L118V mutant gave the best results among the five candidates. The m^2_2G peak disappeared completely with the appearance of m^2G accordingly; suggesting that one methyl group from m^2_2G was fully removed. Other enzymes such as the wild-type AlkB,

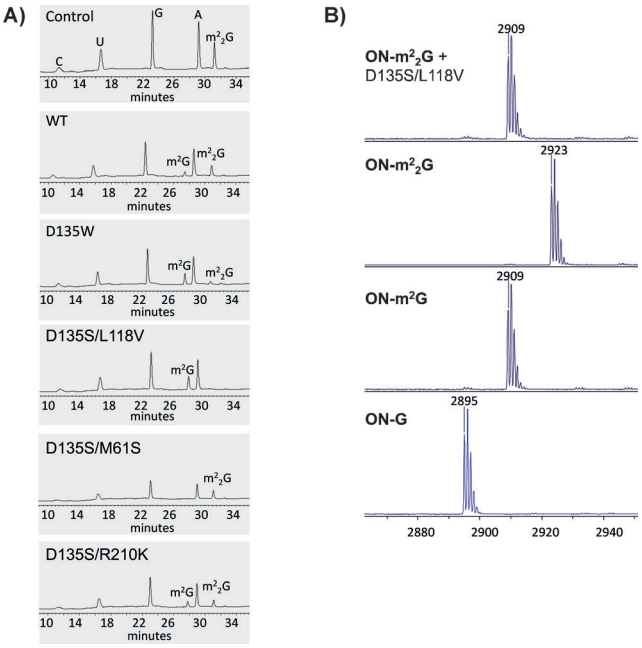


Figure 2. Evaluations of a series of AlkB mutants for efficient $m^2{}_2G$ 5′-GAGCXUUAG, X = $m^2{}_2G$ (**ON-$m^2{}_2G$**), m^2G (**ON-m^2G**) or G (**ON-G**). A) HPLC traces of the digestion products after **ON-$m^2{}_2G$** was treated with demethylases showed that AlkB D135S/L118V mutant quantitatively converts $m^2{}_2G$ to m^2G. B) Comparison of MALDI-TOF MS of unmodified RNA oligo **ON-G**, m^2G-containing RNA oligo **ON-m^2G**, $m^2{}_2G$-containing RNA oligo **ON-$m^2{}_2G$**, and **ON-$m^2{}_2G$** which was treated with AlkB D135S/L118V mutant confirmed that the demethylation product of **ON-$m^2{}_2G$** is **ON-m^2G** instead of **ON-G**.

D135S or D135S/R210K mutant showed limited activities while D135S/M61S mutant showed no activity at all toward $m^2{}_2G$. The Maldi-TOF MS of the sample treated with D135S/L118V mutant also showed complete disappearance of the peak at *m/z* 2922 and appearance of a new peak at 2908 which is identical to that of oligo **ON-m^2G**, indicating that one of two methyl groups in the $m^2{}_2G$ oligo **ON-$m^2{}_2G$** was removed completely. No peak identical to that of **ON-G** was observed (*m/z* = 2894), again indicating that only one methyl was removed. Indeed, when we treated **ON-m^2G** with AlkB D135S/L118V mutant, both HPLC and Maldi-TOF showed no reaction, further confirming that AlkB mutants can only remove one methyl group from $m^2{}_2G$. The shorter side chain of Val118 compared with original Leu118, which sits directly opposite to methyl groups on $m^2{}_2G$, results in more space for $m^2{}_2G$ to fit in the active site to be demethylated.

We also used a 9mer oligo containing an m^1G modification[3] to test whether the AlkB D135S/L118V mutant can also demethylate m^1G efficiently. Unlike AlkB D135S mutant which demethylated m^1G quantitatively, we observed that AlkB D135S/L118V mutant had only limited activity toward m^1G, suggesting that AlkB D135S and D135S/L118V mutants can have distinct substrate selectivity (Figure S2). To obtain the optimal result in tRNA sequencing, it is therefore best to use a three-demethylase mixture: wild-type AlkB (to remove m^1A and m^3C), D135S (to remove m^1G) and D135S/L118V (to remove $m^2{}_2G$).

We then tested whether AlkB D135S/L118V mutant improves $m^2{}_2G$ demethylation in tRNA. Like our previous DM-tRNA-seq method,[3] we used the combination of three demethylases, wild-type, D135S and D135S/L118V to treat tRNA samples from HEK293T cells and constructed sequencing libraries to evaluate the effect of AlkB D135S/L118V mutant on high-throughput tRNA sequencing. As controls, we also sequenced the same samples without demethylase treatment or treated with just two demethylases, wild-type and D135S. To evaluate the effect of the D135S/L118V demethylase on $m^2{}_2G$, we applied a quantitative measure, modification index (MI) which is the sum of the fraction of sequencing reads derived from RT mutations (from G to A/C/T in this case) and stop at the $m^2{}_2G26$ in each tRNA[5b] (Figure 3A). Removal of the $m^2{}_2G$ methyl group

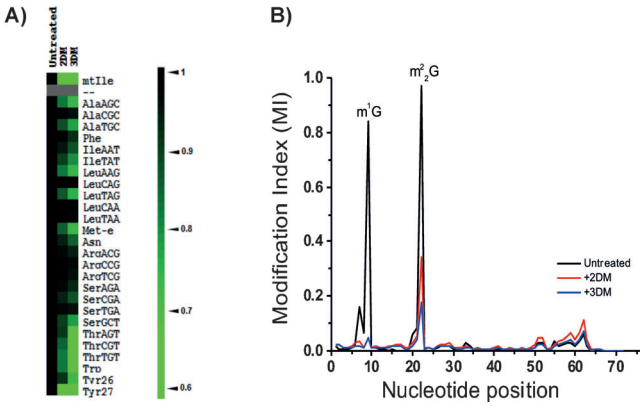


Figure 3. Application of AlkB D135S/L118V to high-throughput tRNA-seq. A) tRNA-seq heatmap result showing the modification index (MI) change at position 26 in tRNAs with known $m^2{}_2G$ sites. 2DM = wild-type AlkB + D135S; 3DM = wild-type AlkB + D135S + D135S/L118V. The higher the MI value, the higher the fraction of modifications. Therefore, decreased modification fraction corresponds to lower MI value or more intense green color for the corresponding tRNA isoacceptor. B) Full MI plot for the full-length mitochondrial $tRNA^{Ile}$. MI at the $m^2{}_2G$ nucleotide is 0.972, 0.343, 0.178 for untreated (black), +2DM (red), and +3DM (blue), respectively.

would result in increased cDNA synthesis through this nucleotide and incorporation of the correct nucleotide. Therefore, reduced level of $m^2{}_2G26$ would decrease the MI values at this position. Indeed, treatment using the three-demethylase mixture (+3DM) reduced MI values for the majority of all known tRNAs with known $m^2{}_2G26$ modifications at greater extent compared to using the two-demethylase mixture (+2DM). Because $m^2{}_2G26$ is located in a very structured region of tRNA (Figure 1B), the absolute demethylation fraction is still substantially lower than the m^1A, m^1G and m^3C modifications shown in the previous study.[3] Among the cytosolic tRNAs, $tRNA^{Thr}$s, $tRNA^{Trp}$ and $tRNA^{Tyr}$ show a greater extent of MI reduction than other tRNAs (Figure 3A); this result may be attributed to the variable stabilities of cytosolic tRNAs that carry many modifications of distinct types at distinct sites. Consistent with the idea that the low absolute $m^2{}_2G$ demethylation fraction is related to tRNA structure, the mitochondrial $tRNA^{Ile}$ has the highest level of MI value reduction (Figure 3B), because mitochon-

drial tRNAs are markedly less structured than the cytosolic tRNAs.[5b]

In summary, we have identified an AlkB mutant (D135S/L118V) that can efficiently remove one methyl group from the $m^2{}_2G$ to convert it to m^2G. Unlike $m^2{}_2G$, m^2G does not block RT reaction under our sequencing conditions, resulting in improved tRNA sequencing efficiency. Different AlkB mutants showed demethylation selectivity on different base methylations, suggesting a role of the active-site residues in substrate recognition and the possibility to find other AlkB mutants to selectively demethylate other base modifications such as dimethyl adenosine. Besides the application of these demethylases in tRNA sequencing, these selective demethylases may also serve as useful tools for developing single-base resolution sequencing methods for base methylations in other RNA species such as mRNA.[12]

Acknowledgements

This work was supported by National Institute of Health grants in USA (K01 HG006699 to Q.D., and RM1HG008935 to T.P.). We thank Dr. Matthew Eckwahl-Sanna for helping edit the figures.

Conflict of interest

The authors declare no conflict of interest.

Keywords: AlkB mutant · demethylase · methylguanosine · reverse transcriptase reaction · tRNA sequencing

[1] K. A. Dittmar, J. M. Goodenbour, T. Pan, *PLoS Genet.* **2006**, *2*, e221.
[2] a) S. Kirchner, Z. Ignatova, *Nat. Rev. Genet.* **2015**, *16*, 98 – 112; b) J. A. Abbott, C. S. Francklyn, S. M. Robey-Bond, *Front. Genet.* **2014**, *5*, 158.
[3] G. Zheng, Y. Qin, W. C. Clark, Q. Dai, C. Yi, C. He, A. M. Lambowitz, T. Pan, *Nat. Methods* **2015**, *12*, 835 – 837.
[4] C. Zhu, C. Yi, *Angew. Chem. Int. Ed.* **2014**, *53*, 3659 – 3662; *Angew. Chem.* **2014**, *126*, 3733 – 3736.
[5] a) J. Edqvist, K. B. Straby, H. Grosjean, *Biochimie* **1995**, *77*, 54 – 61; b) W. C. Clark, M. E. Evans, D. Dominissini, G. Zheng, T. Pan, *RNA* **2016**, *22*, 1771 – 1784.
[6] P. S. Pallan, C. Kreutz, S. Bosio, R. Micura, M. Egli, *RNA* **2008**, *14*, 2125 – 2135.
[7] C. G. Yang, C. Yi, E. M. Duguid, C. T. Sullivan, X. Jian, P. A. Rice, C. He, *Nature* **2008**, *452*, 961 – 965.
[8] P. Rife, C. S. Cheng, P. B. Moore, S. A. Strobel, *Nucleosides Nucleotides* **1998**, *17*, 2281 – 2288.
[9] V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983 – 1985.
[10] a) K. Piecyk, R. E. Davis, M. Jankowska-Anyszka, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 4391 – 4395; b) M. Sekine, M. Aoyagi, M. Ushioda, A. Ohkubo, K. Seio, *J. Org. Chem.* **2005**, *70*, 8400 – 8408; c) N. D. Abeydeera, C. S. Chow, *Bioorg. Med. Chem.* **2009**, *17*, 5887 – 5893; d) M. Sekine, T. Satoh, *J. Org. Chem.* **1991**, *56*, 1224 – 1227.
[11] K. Kamaike, K. Kinoshita, K. Niwa, K. Hirose, K. Suzuki, Y. Ishido, *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 59 – 75.
[12] a) D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Dore, N. Amariglio, G. Rechavi, C. He, *Nature* **2016**, *530*, 441 – 446; b) X. Li, X. Xiong, K. Wang, L. Wang, X. Shu, S. Ma, C. Yi, *Nat. Chem. Biol.* **2016**, *12*, 311 – 316.

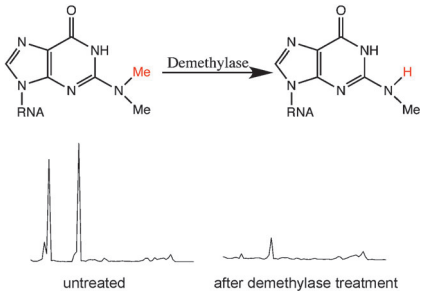
Manuscript received: January 16, 2017
Revised: February 10, 2017
Final Article published: ■■ ■■, ■■■■

Communications

Sequencing Methods

Q. Dai, G. Zheng, M. H. Schwartz, W. C. Clark, T. Pan* ——— ■■■■–■■■■

Selective Enzymatic Demethylation of N^2,N^2-Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing



An AlkB mutant has been identified that efficiently removes one methyl group from N^2,N^2-dimethylguanosine ($m^2{}_2G$) to convert into N^2-methylguanosine (m^2G). Unlike $m^2{}_2G$, m^2G does not block reverse transcription reaction under sequencing conditions, resulting in improved tRNA sequencing efficiency.