JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Identification of a Selective Polymerase Enables Detection of N^6-Methyladenosine in RNA

Emily M. Harcourt,[†] Thomas Ehrenschwender,[†] Pedro J. Batista,[‡] Howard Y. Chang,[‡] and Eric T. Kool*[,†]

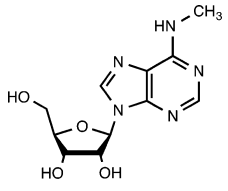
[†]Department of Chemistry, Stanford University, Stanford, California 94305, United States

[‡]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, United States

Supporting Information

ABSTRACT: N^6-methyladenosine (m^{6}A) is the most abundant mRNA modification and has important links to human health. While recent studies have successfully identified thousands of mammalian RNA transcripts containing the modification, it is extremely difficult to identify the exact location of any specific m^{6}A. Here we have identified a polymerase with reverse transcriptase activity (from *Thermus thermophilus*) that is selective by up to 18-fold for incorporation of thymidine opposite unmodified A over m^{6}A. We show that the enzyme can be used to locate and quantify m^{6}A in synthetic RNAs by analysis of pausing bands, and have used the enzyme in tandem with a nonselective polymerase to locate the presence and position of m^{6}A in high-abundance cellular RNAs. By this approach we demonstrate that the long-undetermined position of m^{6}A in mammalian 28S rRNA is nucleotide 4190.

In the quest to understand cellular function at the molecular level, the study of post-transcriptional modification of RNA is of vital interest. In particular, N^6-methyladenosine (m^{6}A, **1**) is a relatively abundant modification in the mRNA of higher eukaryotes and some viruses.[1] Although its discovery in mRNA occurred decades ago,[2] there has been renewed interest in m^{6}A due to the finding that it acts as a substrate for fat mass and obesity-associated protein (FTO),[3] an oxidative demethylase which has been linked to obesity and the regulation of homeostasis;[4−6] and for AlkBH5, an oxidative demethylase in the same family.[7]



1

Identifying the function of m^{6}A modifications has proved challenging.[8−12] Early work using enzymatic digestion and radiolabeling led to the discovery of the consensus sequence RAC (R = A or G) for m^{6}A[13−18] and the identification of specific modified sites in Rous sarcoma virus RNA[19−21] and in bovine prolactin mRNA.[22,23] This research also showed that methylation at any particular site can be incomplete, with a methylation extent of 20−90% at one site. Modern RNA sequencing techniques have advanced our ability to identify transcripts modified by m^{6}A. Using massively parallel sequencing and m^{6}A-selective antibodies, two groups recently identified thousands of modified mRNAs and ncRNAs from mice and humans and confirmed the consensus sequence of RRACU for the general location of methylated adenines.[24,25] After completion of the present work, Liu et al. reported a new method for detecting and quantifying m^{6}A at a specific site using multiple enzymatic steps.[26] However, the ability to interrogate the methylation status of any specific adenine at nucleotide resolution, without painstaking digestion analysis, has remained elusive.

The ability to locate m^{6}A modifications in RNAs at nucleotide resolution will no doubt aid in understanding their function. Polymerase enzymes offer a possible mechanism for locating modifications due to their sterically sensitive active sites. Notably, polymerase selectivity has previously been harnessed to detect m^{6}A in DNA via single-molecule sequencing.[27] An early attempt at a related single-molecule sequencing technique for RNA has also been described,[28] but employed an enzyme with low selectivity (HIV-RT; see below), and will need further development before it is practical. Another class of DNA-processing enzymes, ligases, can also be sensitive to structure, and Dai et al. describe a technique in which ligation of complementary DNAs occurs more favorably in the presence of A than m^{6}A.[29] However, the conditions were tuned carefully for the specific reaction, and no attempt was made to detect m^{6}A in an actual sample of cellular RNA.

We postulated that there might exist a polymerase enzyme with substantial selectivity against m^{6}A, and that such an enzyme might be harnessed for site-specific detection of the modification. We carried out a screen of enzymes with reverse transcriptase activity, monitoring their ability to extend a radiolabeled DNA primer by incorporating thymidine triphosphate (dTTP) opposite either A or m^{6}A in an RNA template (Figure 1). The data showed that only recombinant *Thermus thermophilus* DNA polymerase I (*Tth* DNA pol) showed strong selectivity (61% vs 15% primer extension) among the enzymes tested. This DNA polymerase is known to act as a reverse transcriptase in the presence of Mn^{2+}.[30]

Since *Tth* DNA pol showed selectivity in the context of one specific template sequence under one set of conditions, we

Received: October 16, 2013
Published: December 11, 2013

dx.doi.org/10.1021/ja4105792 | J. Am. Chem. Soc. 2013, 135, 19079−19082

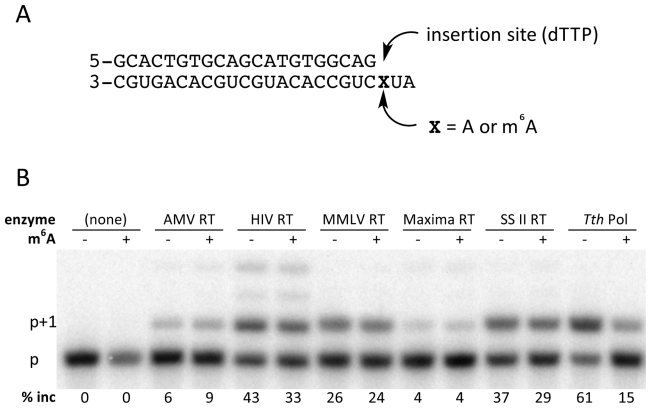


Figure 1. Screen of polymerase selectivity for incorporation of dTTP opposite A or m^{6}A in an RNA template. (A) Sequences of RNA template/DNA primer used in screen. (B) Autoradiogram showing primer extension (p+1 band) in the presence of A or m^{6}A. Products were resolved on a 20% polyacrylamide denaturing gel using a ^{32}P-5′-labeled primer.

tested whether variations in temperature, time, and buffer composition might enhance selectivity (see Supporting Information (SI)). In particular, Mn^{2+} is known to decrease enzyme selectivity;[31] however, we found that Mn^{2+} was required for reverse transcriptase activity in *Tth* DNA pol.

We next investigated whether this selectivity would extend to other sequence contexts. A 24mer template sequence was chosen from the 3′-untranslated region (3′-UTR) of the *eef2* gene, which was found to be highly expressed and highly modified in mouse tissue and mouse embryonic stem cells.[25,32] The template was synthesized containing either A or m^{6}A, and the bases on either side of the A/m^{6}A were varied systematically to allow a comparison of sequence context effects (see SI for details). In addition, several of the native sequence contexts in which m^{6}A has been reported to occur were also synthesized. Single nucleotide incorporation kinetics were determined for each sequence containing A and m^{6}A using steady-state methods[33] (Table 1).

Table 1. Steady-State Incorporation Efficiency for Insertion of dTTP Opposite A or m^{6}A in Synthetic RNAs by *Tth* DNA pol in Varied Sequence Contexts

sequence context	efficiency, X = A (V_{max}/K_m)	efficiency, X = m^{6}A (V_{max}/K_m)	A/m^{6}A ratio
UXC	0.86 ± 0.17	0.10 ± 0.04	8.6
GXC	0.58 ± 0.15	0.14 ± 0.03	4.0
(G)GXC	1.22 ± 0.17	0.19 ± 0.04	6.4
GXU	0.95 ± 0.61	0.10 ± 0.10	9.1
CXU	1.13 ± 0.77	0.14 ± 0.14	8.2
CXG	1.39 ± 0.26	0.13 ± 0.03	10.4
AXG	1.51 ± 0.61	0.10 ± 0.02	15.6
AXA	1.87 ± 1.25	0.11 ± 0.06	17.5
(G)AXC	1.24 ± 0.32	0.10 ± 0.05	13.1
UXA	3.52 ± 0.94	0.33 ± 0.22	10.6

The RNA templates containing A show 4- to 18-fold better enzyme efficiency with *Tth* pol than templates with the corresponding sequence containing m^{6}A. Overall, then, differences in context produce moderate to negligible differences in selectivity. The UAA sequence context is processed with higher efficiency than other sequences, whether or not adenine is modified. Additionally, 5′ G and 3′ C both appear to decrease enzyme efficiency to a small degree. Most notably, the selectivity of the *Tth* polymerase in the consensus methylation site context (GAC) is 4−6.4-fold, thus supporting the notion that the enzyme's selectivity may be useful for identifying the most common occurrences of m^{6}A in naturally occurring RNAs.

Next we asked whether the *Tth* enzyme could be used in a quantitative sense to evaluate the degree of methylation at a specific site. To test this, we mixed known ratios of m^{6}A-containing RNAs with their A-containing counterparts and measured the yield of dTTP incorporation at a fixed time point. The percent extension of the primer in this RNA context was linearly proportional to the amount of m^{6}A present (Figure 2), suggesting that the polymerase can be used in quantitative evaluations of the extent of methylation at one position.

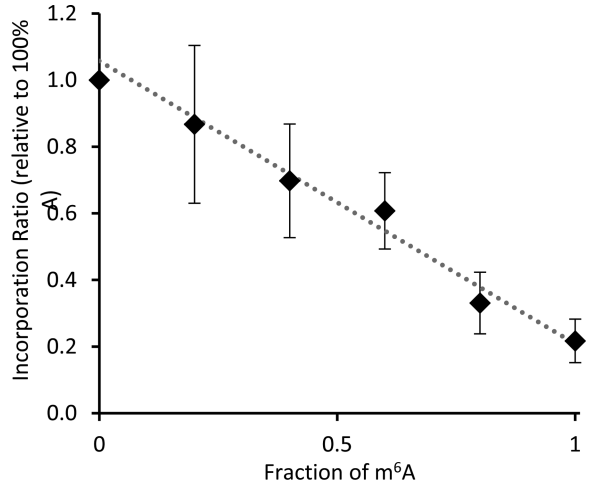


Figure 2. T insertion is correlated to the relative amount of m^{6}A at target position. 5'-AG**X**CUGCCACAUGCUGCACAGUGC-3' was used as the template RNA at 1 μM concentration with varied ratios of m^{6}A:A at the target position. Error bars show standard deviations from five trials.

We proceeded to carry out experiments to test whether this enzyme could be employed in probing methylation in RNAs extracted from mammalian cells. Since the amount of RNA and its secondary structure are likely to affect primer extension efficiencies, we introduced two control strategies. The first of these involves the use of two simultaneous primers (one adjacent to the probed methylation site and one nearby but adjacent to a nucleotide with known methylation status). Comparison of these two should account for the amount of a given RNA present in a sample. The second control makes use of the same primers with a nonselective enzyme, avian myeloblastosis virus reverse transcriptase (AMV RT). AMV RT was selected because it showed consistent ratios relative to *Tth* DNA pol across a range of RNA concentrations. Insertion of T at a site of interest could then be compared both to the control site and to incorporation by a control nonselective enzyme.

Human rRNA is known to contain two m^{6}A modifications, one at position 1832 in the 18S subunit, and one at position 4189 or 4190 in the 28S subunit.[34] On the basis of ratios to nearby unmodified bases after enzymatic digestion, these sites are reported to consist only of the modified base (100% m^{6}A).[35,36] Primers were designed to interrogate these three

dx.doi.org/10.1021/ja4105792 | J. Am. Chem. Soc. 2013, 135, 19079−19082

sites, as well as two other sites known to contain A (1781 in 18S; 4984 in 28S). Total RNA was extracted from 293T cells, and the RNAs were probed with primer sets (see SI). As seen in Figure 3, results showed less than 20% incorporation of T by

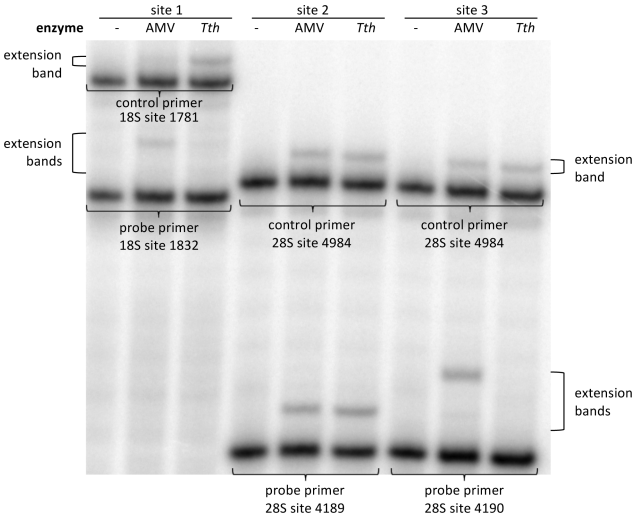


Figure 3. Identification of methylated adenine sites in human rRNA. Primers were designed to probe for m^6A at three sites in rRNA. Incorporation at these sites by *Tth* pol was compared to incorporation by *Tth* pol at known nonmethylated control sites and to incorporation by nonselective AMV RT.

Tth DNA pol at the 1832 m^6A site and the 4190 m^6A site relative to controls, consistent with complete methylation in our quantitative experiments with synthetic RNAs. In contrast, we observed a high degree of incorporation (~120%) by *Tth* pol relative to AMV RT at the 4189 site. As a result, we can assign the previously undetermined site of methylation in human 28S RNA as 4190 and not at the neighboring adenine at 4189.

Finally, we attempted to detect m^6A in a cellularly expressed mRNA (Figure 4). We chose a known site in the 3′-UTR of the bovine prolactin (bPRL) transcript.[23] This is the only precisely mapped m^6A site in a mammalian mRNA; native levels of modification are estimated to be ~20%, and this presents a challenging case for detection.[23] The 3′-UTR was cloned into a

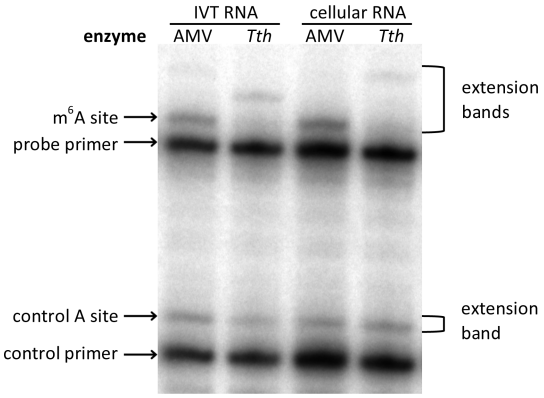


Figure 4. Identification of the m^6A site in the *bPRL* 3′-UTR. Incorporation of dTTP with *in vitro* transcribed (IVT) RNA was compared to incorporation templated by cellular *bPRL* RNA to differentiate the site of methylation. Incorporation with nonselective AMV RT is shown as an additional control.

plasmid and overexpressed in 293T cells. Two primers were designed, one to detect the known modification site and one to detect a nearby adenosine (see SI). The putative m^6A site showed significantly lower incorporation than the nearby control A site. As an additional control, we performed *in vitro* transcription of the *bPRL* transcript in the absence of m^6A, allowing us to compare the primer incorporation at the site of interest when only A was present. At the high enzyme concentrations initially used, no difference was seen between the total cellular RNA and the *in vitro* transcribed RNA. Use of a lower enzyme concentration, however, did reveal a significant difference in the m^6A:A ratio (0.75 ± 0.09 for IVT RNA vs 0.43 ± 0.04 for total cellular RNA; all extension products were included in quantification), confirming the presence of m^6A in the mRNA expressed in 293T cells.

In summary, we have characterized a commercially available polymerase that discriminates m^6A from A in all tested sequence contexts. We have used it to detect m^6A in abundant cellular RNAs. While further development is needed before this method is robust enough for detection and quantification of m^6A in lower abundance mRNAs, it seems likely that the inherent selectivity of this enzyme will prove useful in the development of future m^6A analysis techniques.

■ ASSOCIATED CONTENT

Supporting Information

Experimental methods and materials, additional data, and characterization of synthetic intermediates. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

kool@stanford.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the National Institutes of Health (GM068122) for support. E.M.H. is supported by a Graduate Research Fellowship (DGE-1147470) from the NSF. T.E. is supported by a Postdoctoral Research Fellowship from the DAAD. P.J.B. is the Kenneth G. and Elaine A. Langone Fellow of the Damon Runyon Cancer Research Foundation; H.Y.C. is an Early Career Scientist of Howard Hughes Medical Institute.

■ REFERENCES

(1) Bokar, J. A. The Biosynthesis and Functional Roles of Methylated Nucleosides in Eukaryotic mRNA. In *Fine-tuning of RNA Functions by Modification and Editing*; Grosjean, H., Ed.; Topics in Current Genetics *12*; Springer-Verlag: Berlin Heidelberg, 2005; pp 141−177.

(2) Desrosiers, R.; Friderici, K.; Rottman, F. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3971−3975.

(3) Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.-G.; He, C. *Nat. Chem. Biol.* **2011**, *7*, 885−887.

(4) Frayling, T. M.; Timpson, N. J.; Weedon, M. N.; Zeggini, E.; Freathy, R. M.; Lindgren, C. M.; Perry, J. R. B.; Elliott, K. S.; Lango, H.; Rayner, N. W.; Shields, B.; Harries, L. W.; Barrett, J. C.; Ellard, S.; Groves, C. J.; Knight, B.; Patch, A.-M.; Ness, A. R.; Ebrahim, S.; Lawlor, D. A.; Ring, S. M.; Ben-Shlomo, Y.; Jarvelin, M.-R.; Sovio, U.; Bennett, A. J.; Melzer, D.; Ferrucci, L.; Loos, R. J. F.; Barroso, I.; Wareham, N. J.; Karpe, F.; Owen, K. R.; Cardon, L. R.; Walker, M.; Hitman, G. A.; Palmer, C. N. A; Doney, A. S. F.; Morris, A. D.; Smith, G. D.; Hattersley, A. T.; McCarthy, M. I. *Science* **2007**, *316*, 889−894.

(5) Gerken, T.; Girard, C. A.; Tung, Y.-C. L.; Webby, C. J.; Saudek, V.; Hewitson, K. S.; Yeo, G. S. H.; McDonough, M. A.; Cunliffe, S.; McNeill, L. A; Galvanovskis, J.; Rorsman, P.; Robins, P.; Prieur, X.; Coll, A. P.; Ma, M.; Jovanovic, Z.; Farooqi, I. S.; Sedgwick, B.; Barroso, I.; Lindahl, T.; Ponting, C. P.; Ashcroft, F. M.; O'Rahilly, S.; Schofield, C. J. *Science* **2007**, *318*, 1469−1472.
(6) Fischer, J.; Koch, L.; Emmerling, C.; Vierkotten, J.; Peters, T.; Brüning, J. C.; Rüther, U. *Nature* **2009**, *458*, 894−898.
(7) Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C.-M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W.-L.; Song, S.-H.; Lu, Z.; Bosmans, R. P. G.; Dai, Q.; Hao, Y.-J.; Yang, X.; Zhao, W.-M.; Tong, W.-M.; Wang, X.-J.; Bogdan, F.; Furu, K.; Fu, Y.; Jia, G.; Zhao, X.; Liu, J.; Krokan, H. E.; Klungland, A.; Yang, Y.-G.; He, C. *Mol. Cell* **2013**, *49*, 18−29.
(8) Bachellerie, J.; Amalric, F.; Caboche, M. *Nucleic Acids Res.* **1978**, *5*, 2927−2944.
(9) Stoltzfus, C. M.; Dane, R. W. *J. Virol.* **1982**, *32*, 918−931.
(10) Finkel, D.; Groner, Y. *Virology* **1983**, *131*, 409−425.
(11) Camper, S. A.; Albers, R. J.; Coward, J. K.; Rottman, F. M. *Mol. Cell. Biol.* **1984**, *4*, 538−543.
(12) Carroll, S. M.; Narayan, P.; Rottman, F. M. *Mol. Cell. Biol.* **1990**, *10*, 4456−4465.
(13) Schibler, U.; Kelley, D. E.; Perry, R. P. *J. Mol. Biol.* **1977**, *115*, 695−714.
(14) Wei, C.; Moss, B. *Biochemistry* **1977**, *16*, 1672−1676.
(15) Dimock, K.; Stoltzfus, C. M. *Biochemistry* **1977**, *16*, 471−478.
(16) Canaani, D.; Kahana, C.; Lavi, S.; Groner, Y. *Nucleic Acids Res.* **1979**, *6*, 2879−2899.
(17) Nichols, J. L.; Welder, L. *Plant Sci. Lett.* **1981**, *21*, 75−81.
(18) Narayan, P.; Ludwiczak, R. L.; Goodwin, E. C.; Rottman, F. M. *Nucleic Acids Res.* **1994**, *22*, 419−426.
(19) Beemon, K.; Keith, J. *J. Mol. Biol.* **1977**, *113*, 165−179.
(20) Kane, S. E.; Beemon, K. *Mol. Cell. Biol.* **1985**, *5*, 2298−2306.
(21) Csepany, T.; Lin, A.; Baldick, C. J.; Beemon, K. *J. Biol. Chem.* **1990**, *265*, 20117−20122.
(22) Horowitz, S.; Horowitz, A.; Nilsen, T. W.; Munns, T. W.; Rottman, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5667−5671.
(23) Narayan, P.; Rottman, F. M. *Science* **1988**, *242*, 1159−1162.
(24) Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G. *Nature* **2012**, *485*, 201−206.
(25) Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R. *Cell* **2012**, *149*, 1635−1646.
(26) Liu, N.; Parisien, M.; Dai, Q.; Zheng, G.; He, C.; Pan, T. *RNA* **2013**.
(27) Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. *Nat. Methods* **2010**, *7*, 461−465.
(28) Vilfan, I. D.; Tsai, Y.-C.; Clark, T. A.; Wegener, J.; Dai, Q.; Yi, C.; Pan, T.; Turner, S. W.; Korlach, J. *J. Nanobiotechnology* **2013**, *11*, 8.
(29) Dai, Q.; Fong, R.; Saikia, M.; Stephenson, D.; Yu, Y.; Pan, T.; Piccirilli, J. A. *Nucleic Acids Res.* **2007**, *35*, 6322−6329.
(30) Myers, T. W.; Gelfand, D. H. *Biochemistry* **1991**, *30*, 7661−7666.
(31) Hillebrand, G. G.; Beattie, K. L. *Nucleic Acids Res.* **1984**, *12*, 3173−3184.
(32) Batista, P. J.; Chang, H. Unpublished data.
(33) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83−126.
(34) Maden, B. E. H. *Prog. Nucleic Acid Res. Mol. Biol.* **1990**, *39*, 241−303.
(35) Maden, B. E. H. *J. Mol. Biol.* **1986**, *189*, 681−699.
(36) Maden, B. E. H. *J. Mol. Biol.* **1988**, *201*, 289−314.