5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription

Matthew W Kellinger1, Chun-Xiao Song2, Jenny Chong1, Xing-Yu Lu2, Chuan He2 & Dong Wang1

Although the roles of 5-methylcytosine and 5-hydroxymethylcytosine in epigenetic regulation of gene expression are well established, the functional effects of 5-formylcytosine and 5-carboxylcytosine on the process of transcription are not clear. Here we report a systematic study of the effects of five different forms of cytosine in DNA on mammalian and yeast RNA polymerase II transcription, providing new insights into potential functional interplay between cytosine methylation status and transcription.

At least five forms of cytosine have been identified in cells: cytosine, 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Fig. 1a). The 5mC and 5hmC modifications play critical roles in epigenetic regulation of gene expression and maintenance of cellular identity1,2. The presence of cytosine, 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) is thought to occur at a frequency of between 104 and 105 per genome (Online Methods)3–5. This level is comparable to the frequency of DNA damage caused by oxidative stress, which is commonly investigated in the context of transcription and DNA-damage repair5. Their presence in the gene bodies raises an intriguing question of how these cytosine derivatives may affect RNA polymerase II (Pol II) transcription activity and transcription dynamics. To address this question, here we have systematically investigated the rate of Pol II elongation and substrate specificity (kcat/Km or kpol/Kq) for DNA templates containing a site-specific C, 5mC, 5hmC, 5fC or 5caC.

We first measured in vitro Pol II elongation efficiency using purified mammalian Pol II with DNA templates containing a site-specific C, 5mC, 5hmC, 5fC or 5caC (Fig. 1b). Our results show a striking difference in mammalian Pol II transcription efficiency among these five forms of cytosine. GTP incorporation efficiencies following a 15-s incubation for 5fC- and 5caC-containing templates were significantly lower than those for the C template (40 ± 5 and 39 ± 3 percent (mean ± s.d.; n = 3; P = 0.0002; unpaired two-tailed t-test), respectively), whereas no noticeable difference was observed among C, 5mC and 5hmC templates (Fig. 1c). We also observed similar pausing effects of 5fC and 5caC on transcription elongation efficiency when using purified yeast Pol II (45 ± 5 (n = 4; P < 0.0001) and 52 ± 12 (n = 4; P = 0.0004) percent, respectively; Fig. 1d and Supplementary Fig. 1). Pol II has strictly conserved (almost identical) active site residues (Supplementary Fig. 2) and shares common enzymatic mechanisms for nucleotide incorporation in eukaryotes from yeast to human10,11. Thus, it is not surprising that similar effects were observed for both mammalian and yeast Pol II, and yeast Pol II can be used as a model enzyme to study the effects of 5fC and 5caC templates on mammalian Pol II transcription elongation. It should be noted that we observed a similar trend of significant reduction of Pol II nucleotide incorporation and elongation efficiency on 5fC and 5caC templates from Pol II elongation complexes with different scaffolds (either with a full transcription bubble or downstream edge of the transcription bubble; Fig. 1b–e and Supplementary Fig. 3), indicating that the observed reduction in transcription elongation efficiency is not due to specific sequence context or scaffold setting. Taken together, these results reveal that Pol II elongation efficiencies for 5fC and 5caC templates are significantly reduced relative to those for the C template.

To further investigate the Pol II transcription elongation kinetics and quantitatively measure the effect of cytosine modifications on Pol II transcription elongation, we then determined the full pre–steady state single-turnover NTP incorporation kinetics for templates C, 5hmC, 5fC and 5caC. These single-turnover experiments allow us to directly determine the key kinetic parameters kpol of GTP incorporation and Kapp (apparent substrate dissociation constant) of nucleotide incorporation at a single site. Furthermore, we can determine the ratio of kpol to Kapp (substrate specificity), an important kinetic parameter for measuring Pol II enzymatic efficiency. The comparison of substrate specificity among these different forms of cytosine residues gives a quantitative measurement of their effects on Pol II transcription elongation. Yeast Pol II was used to overcome the technical difficulty of purifying a sufficient amount of mammalian Pol II for extensive enzymology study. Notably, Pol II polymerization rates (kpol) for GTP incorporation against 5fC and 5caC were significantly reduced, to 2.0% and 1.3% of kpol of GTP incorporation against C template, respectively, whereas no noticeable changes were observed in the case of 5hmC compared with unmethylated C template (Fig. 2a, Supplementary Fig. 4 and Supplementary Table 1). The Pol II specificity constants for GTP incorporation against 5fC and 5caC templates decreased significantly, by factors of ~30.0 and 4.2-fold, respectively (Fig. 2b), whereas the specificity constant was reduced only slightly.

1Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA. 2Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois, USA. Correspondence should be addressed to D.W. (dongwang@ucsd.edu).

Received 21 February; accepted 19 June; published online 22 July 2012; doi:10.1038/nsmb.2346
in the case of 5hmC compared with unmodified C template (Fig. 2b). Differences in both $k_{\text{pol}}$ and apparent $K_{\text{dapp}}$ produced the altered GTP specificity (Supplementary Table 1). In addition, product formation was significantly more biphasic for 5fC and 5caC templates than for C and 5hmC templates (Supplementary Table 2). The observation of distinct phases suggests that multiple populations of Pol II elongation complexes exist (Supplementary Note): one population is poised for rapid GTP incorporation (observed as the fast phase of product formation, such as for Pol II in the post-translocation state), whereas the second, ‘paused’ population of Pol II complexes requires a longer time for GTP incorporation (observed as the slower phase, such as for Pol II in the pretranslocation, frayed or backtracked states). This slower phase may reflect a rate-limiting isomerization that must occur before GTP incorporation.

To investigate whether the presence of 5fC and 5caC promotes the backtracking population of Pol II complex, we measured the rate of TFIIH-mediated backtracked RNA cleavage for all five scaffolds. The TFIIH-mediated cleavage rates for scaffolds containing 5fC and 5caC increased by 5.0- and 2.6-fold, respectively, whereas no noticeable changes were observed in the cases of 5hmC and 5mC, as compared with unmodified C template (Fig. 2c). Taken together, these results indicate that the presence of 5fC and 5caC greatly shifts Pol II from an active population (poised for elongation) to a paused population and promotes Pol II stalling and RNA backtracking.

Finally, we investigated whether these cytosine derivatives cause significant changes in transcriptional fidelity. We first tested the incorporation efficiency of individual NTP over templates with five different cytosine derivatives. We found that the misincorporation efficiency of ATP was significantly higher than that of UTP and CTP across all five templates (Fig. 1d and Supplementary Fig. 5). We then examined full pre–steady state single-turnover ATP misincorporation kinetics for C and 5fC templates. Notably, specificities for ATP misincorporation were significantly reduced, by a factor of $10^{-3} - 10^{-4}$, compared to those for correct GTP incorporation, suggesting that ATP misincorporation is relatively inefficient. In contrast to GTP incorporation, there were no noticeable changes in specificities for ATP misincorporation between C and 5fC templates (Fig. 3a and Supplementary Table 3). To quantitatively
evaluate the effect of 5fC on fidelity of nucleotide incorporation, we calculated the discrimination constants for GTP over ATP incorporation (defined as \(k_{\text{pol}}/K_{\text{d,app}} \) for GTP and \(k_{\text{pol}}/K_{\text{d,app}} \) for ATP) for C and 5fC templates (Fig. 3b and Supplementary Table 3). The discrimination constants of NTP incorporation for C and 5fC templates were 2.8 × 10^4 and 900, respectively. Therefore, substitution of C with 5fC in the templates on Pol II transcriptional efficiency and fidelity and to reveal the molecular basis for the effects of 5fC or 5caC modifications on Pol II transcription may add another layer of transcriptional regulation that result from these templates induce Pol II transient pausing in regulating Pol II transcription. Decreased rates of Pol II transcription elongation in a transcription-coupled manner through Pol II pausing. Future investigations into the effects of 5fC and 5caC on Pol II transcription may provide new insights into the functional interplay between cytosine methylation status and transcriptional regulation.

In summary, we have systematically investigated the effects of all known cytosine modifications on the rate of Pol II nucleotide incorporation and substrate specificity for DNA templates containing a site-specific C, 5mC, 5hmC, 5fC or 5caC. We show that Pol II polymerization rates \(k_{\text{pol}} \) and specificity constants \(k_{\text{pol}}/K_{\text{d}} \) for GTP incorporation against 5fC and 5caC are reduced significantly in comparison with those for C template, whereas essentially no changes are observed for 5mC and 5hmC templates. We further reveal that the discrimination of GTP over ATP is reduced by a factor of ~30 for 5fC template in comparison with C template. Our results indicate that Pol II can read and distinguish subtle changes in cytosine modifications at the 5 position and process them differently. These findings provide an important new perspective on the potential functional interplay between modification status of cytosine and transcription.

**METHODS**

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

**ACKNOWLEDGMENTS**

D.W. acknowledges support from the US National Institutes of Health (GM085136), start-up funds from the Skaggs School of Pharmacy and Pharmaceutical Sciences at the University of California, San Diego, and the Kimmel Scholar Award from the Sidney Kimmel Foundation for Cancer Research. C.H. acknowledges the National Institutes of Health for support (GM071440). We thank J. Conaway and R. Conaway (Stowers Institute for Medical Research) for the generous gift of purified mammalian Pol II. We also thank E.P. Geiduschek, J.R. Halpert, X-D. Fu, B. Ren, Y. Zhang and K. Zhang for their insightful comments and critical reading of the manuscript.

**AUTHOR CONTRIBUTIONS**

D.W. conceived the original idea. M.W.K., C.H. and D.W. designed the experiments. C.S. and X.L. carried out synthesis of DNA templates. J.C. and D.W. purified Pol II. M.W.K. and D.W. performed transcription assays. M.W.K. and D.W. carried out data analysis. M.W.K., J.C., C.H. and D.W. wrote the paper.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/mbio.2346. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jaenisch, R. & Bird, A. Nat. Genet. 33 (suppl.), 245–254 (2003).
2. Law, J.A. & Jacobsen, S.E. Nat. Rev. Genet. 11, 204–220 (2010).
3. Itô, S. et al. Science 333, 1300–1303 (2011).
4. Gu, Y.F. et al. Science 333, 1303–1307 (2011).
5. Pfaffendorfer, T. et al. Angew. Chem. Int. Edn Engl. 50, 7008–7012 (2011).
6. Wu, H. & Zhang, Y. Genes Dev. 25, 2436–2452 (2011).
7. Pastor, W.A. et al. Nature 473, 394–397 (2011).
8. Tahiliani, M. et al. Science 324, 930–935 (2009).
9. Maiti, A. & Drohat, A.C. Nat. Genet. 43, 902–911 (2011).
10. Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D. & Kornberg, R.D. Science 941–954 (2006).
11. Shukla, S. et al. Nature 479, 74–79 (2011).
12. Pinto, P.A. et al. EMBO J. 30, 2431–2444 (2011).
13. Core, L.J. & Lis, J.T. Science 319, 1791–1792 (2008).
14. de la Mata, M. et al. Mol. Cell 12, 525–532 (2003).
15. McDowell, J.C., Roberts, J.W., Jin, D.J. & Gross, C. Science 266, 822–825 (1994).
16. Maiti, A. & Drohat, A.C. J. Biol. Chem. 286, 35334–35338 (2011).
17. Cortellino, S. et al. Cell 146, 67–79 (2011).
18. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Science 941–954 (2006).
19. Corliss, D. et al. Nature 470, 419–423 (2011).
20. Tini, M. et al. Mol. Cell 9, 265–277 (2002).
ONLINE METHODS
These studies used both mammalian and yeast Pol II10,11,21–24. Mammalian Pol II was purified from rat liver as described25. Saccharomyces cerevisiae Pol II was purified using established methods11. The RNA primer and DNA nontemplate oligonucleotides were purchased from IDT. Template DNA oligonucleotides containing site-specific DNA modification were synthesized and purified as described26. These modifications were confirmed by mass spectrometry. RNA primers were radiolabeled using [y-32P]ATP and T4 Polynucleotide Kinase (NEB). The Pol II elongation complexes for transcription assays were assembled as described31.

Transcription elongation reactions with mammalian Pol II. Aliquots of mammalian Pol II purified from rat liver were preincubated with 1.25 µM DNA–RNA scaffold (RNA, 5′-GAGCGUGAGCC-3′; nontemplate strand, 5′-CCGGGCTTCCG-3′; template strand, 5′-TTTCGACCTGCGTCCACGTC-3′; X = C, 5mC, 5fC, 5caC) and 0.83 mM [y-32P]CTP (3,000 Ci/mmol) in reaction buffer (50 mM KCl, 25 mM Tris, pH 7.5, 5 mM MgCl2, 12.5 mM DTT) at room temperature for 20 min to form elongation complexes and 3′-end-labeled RNA. The Pol II complex solution was then mixed with equal volumes of nucleotide solution, containing 1 mM GTP in reaction buffer, for 15 s. The reactions were then quenched with one volume of 0.5 M EDTA solution (pH 8.0). The products were analyzed by denaturing PAGE and then visualized using a storage phosphor screen and Pharos FX imager (Bio-Rad). Substrate and product bands were quantified using Image Lab 3.0.

Transcription elongation reactions with Saccharomyces cerevisiae Pol II. Transcription elongation reactions were performed by preincubating 100 nM scaffold (RNA, 5′-GAGCGUGAGCC-3′; nontemplate strand, 5′-CCGGGCTTCCG-3′; template strand, 5′-TTTCGACCTGCGTCCACGTC-3′; X = C, 5mC, 5fC, 5caC) with 400 nM Pol II in elongation buffer (40 mM KCl, 20 mM Tris, pH 7.5, 10 mM DTT) at room temperature for 20 min before mixing with equal volumes of nucleotide solution containing various concentrations of NTP and 10 mM MgCl2 in elongation buffer. The final reaction conditions after mixing were 50 nM transcription scaffold, 200 nM enzyme, 5 mM MgCl2 and various final NTP concentrations in elongation buffer. The reactions were allowed to proceed for various times before being quenched with one volume of 0.5 M EDTA solution (pH 8.0). The products were analyzed by denaturing PAGE and then visualized using a storage phosphor screen and PhosFlair FX imager (Bio-Rad). Substrate and product bands were quantified using Image Lab 3.0.

Transcription elongation reactions with scaffold containing a full transcription bubble. The transcription elongation complex containing a full transcription bubble (scaffolds B or C in Fig. 1b) was assembled according to the protocol developed previously27,28, with minor modifications described below. Briefly, 1 pmol of Pol II was incubated with 0.3 pmol RNA–template DNA hybrid (see scaffolds B or C in Fig. 1b) for 10 min at room temperature. An aliquot of 1.6 pmol of nontemplate oligonucleotides was added, and the incubation was allowed to proceed for another 10 min at room temperature to produce RNA32P-labeled at the 3′ end. The preassembled transcription elongation complexes were mixed with equal volumes of nucleotide solution containing various concentrations of individual or mixed NTP (1 µM (Fig. 1e) or 10 µM (Supplementary Fig. 3)) and 10 mM MgCl2 in elongation buffer. The final reaction conditions after mixing were 75 nM transcription scaffold, 250 nM enzyme, 5 mM MgCl2 and various final NTP concentrations in elongation buffer. The reactions were allowed to proceed for various times before being quenched with one volume of 0.5 M EDTA solution (pH 8.0). The products were analyzed by denaturing PAGE and then visualized using a storage phosphor screen and PhosFlair FX imager (Bio-Rad). Substrate and product bands were quantified using Image Lab 3.0.

Single-turnover nucleotide incorporation assays. Specificity constants (kcat/Km) governing GTP incorporation against C, 5mC, 5fC or 5caC templates were determined by rapid-quench techniques. Yeast Pol II (200 nM) was preincubated with 50 nM scaffold in elongation buffer at room temperature for 20 min. The enzyme–scaffold complex was then mixed with increasing concentrations of GTP (C and ShmC template: 10, 25, 50, 100, 250 and 1,000 µM; 5fC and 5caC template: 5, 10, 25, 50, 100, 250, 500 and 2,500 µM) and 5 mM MgCl2 in elongation buffer (all concentrations are final after mixing). Reactions were quenched at various time points by addition of one volume of 0.5 M EDTA (pH 8.0). Reactions requiring quench times faster than 5 s were performed using a RQF-3 Rapid Quench Flow (KinTek Corp.).

TFIIH cleavage assays. TFIIH cleavage assays were performed using methods similar to those described for single-turnover nucleotide incorporation assays. The enzyme–scaffold complex was rapidly mixed with a solution of elongation buffer, 10 mM DTT, 5 mM MgCl2, and 1.5 µM TFIIH (all concentrations final after mixing). Reactions were quenched from 0.25 to 120 min by addition of 0.5 M EDTA (pH 8.0). Cleavage products were separated by denaturing PAGE.

Data fitting. Nonlinear-regression data fitting was performed using GraFit 5.0. The time dependence of product formation for various concentrations of GTP was fit to a double exponential equation (1). This fitting yielded two phases corresponding to fast and slow observed rates of product formation. The GTP concentration dependence of the fast phase was fit to a hyperbolic equation (2) to obtain values for kpol and kapp (Supplementary Fig. 4). Alternatively, the data were fit globally using Kinetic Global Explorer (KinTek Corp.)30,31. The specificity constant was determined by kcat/Km = kpol/Kapp.

\[
\text{Product} = A_1e^{-k_1t} + A_2e^{-k_2t} + C
\]

\[
k_{obs} = -\frac{k_{pol}[S]}{k_{app} + [S]}
\]

Occurrences of 5fC and 5caC per genome. Calculation of the number of occurrences of 5fC and 5caC per genome was performed by determining the total number of cytosine residues in the human genome. The diploid genome contains 1.2 × 1010 bases (2 for sister chromosomes) × 2 (for two DNA strands) × 2.9 × 109 bases31 and a GC content of 46%32, resulting in 2.8 × 109 cytosine residues per genome. The occurrence of 5fC and 5caC per genome was performed by determining the total number of 5fC and 5caC occurrences to be 6 × 104 per genome. The occurrence of 5fC is estimated to be 1 × 109 to 9 × 109 per genome3. Similarly, levels of 5fC and 5caC were calculated based on recent reports. One study reported that there are about 20 5fC and 3 5caC in every 109 cytosine3. Thus, total 5fC and 5caC are estimated as ~5.6 × 1010 (2(20 × 109) × 2.8 × 109) and 8.4 × 109 (3(20 × 109) × 2.8 × 109) per genome, respectively. Another study reported that the highest 5fC level is up to 0.02% of total cytosine in mouse ES cells, which leads to an even higher level of 5fC (5.6 × 109 × 0.02% × 2.8 × 109 cytosine per genome)3. Conversely, oxidative DNA damage has been reported to occur transiently at a rate of about 104 per day33.

21. Werner, F. & Grohmann, D. Nat. Rev. Microbiol. 9, 85–98 (2011).
22. Nudler, E. Annu. Rev. Biochem. 78, 335–361 (2009).
23. Conaway, J.W., Shilatifard, A., Dvir, A. & Conaway, R.C. Trends Biochem. Sci. 25, 375–380 (2000).
24. Shilatifard, A., Conaway, J.W. & Conaway, R.C. Curr. Opin. Genet. Dev. 7, 199–204 (1997).
25. Serizawa, H., Conaway, R.C. & Conaway, J.W. J. Biol. Chem. 268, 17300–17308 (1993).
26. Dai, Q. & He, C. Org. Lett. 13, 3446–3449 (2011).
27. Sidorenkov, I., Komissarova, N. & Kashlev, M. Mol. Cell 2, 55–64 (1998).
28. Kireeva, M.L., Komissarova, N., Waugh, D.S. & Kashlev, M. J. Biol. Chem. 275, 6530–6536 (2000).
29. Johnson, K.A., Simpson, Z.B. & Blom, T. Anal. Biochem. 387, 20–29 (2009).
30. Johnson, K.A., Simpson, Z.B. & Blom, T. Anal. Biochem. 387, 30–41 (2009).
31. Venter, J.C. et al. Science 291, 1304–1305 (2001).
32. Romiguer, J., Ranwez, V., Douzy, E.J. & Gallier, N. Genome Res. 20, 1001–1009 (2010).
33. Sentürker, S. et al. FEBS Lett. 416, 286–290 (1997).