Unnatural DNA Base Pairing

Deutsche Ausgabe: DOI: 10.1002/ange.201807845
Internationale Ausgabe: DOI: 10.1002/anie.201807845

Unnatural Cytosine Bases Recognized as Thymines by DNA Polymerases by the Formation of the Watson–Crick Geometry

Hu Zeng[+], Manas Mondal[+], Ruyi Song[+], Jun Zhang, Bo Xia, Menghao Liu, Chenxu Zhu, Bo He, Yi Qin Gao, and Chengqi Yi**

Abstract: *The emergence of unnatural DNA bases provides opportunities to demystify the mechanisms by which DNA polymerases faithfully decode chemical information on the template. It was previously shown that two unnatural cytosine bases (termed "M-fC" and "I-fC"), which are chemical labeling adducts of the epigenetic base 5-formylcytosine, can induce C-to-T transition during DNA amplification. However, how DNA polymerases recognize such unnatural cytosine bases remains enigmatic. Herein, crystal structures of unnatural cytosine bases pairing to dA/dG in the KlenTaq polymerase-host–guest complex system and pairing to dATP in the KlenTaq polymerase active site were determined. Both M-fC and I-fC base pair with dA/dATP, but not with dG, in a Watson–Crick geometry. This study reveals that the formation of the Watson–Crick geometry, which may be enabled by the A-rule, is important for the recognition of unnatural cytosines.*

DNA polymerases are exquisite machines that incorporate the cognate nucleotides into a growing DNA strand. During DNA replication, DNA polymerases undergo complex conformational changes from a binary "open" complex to a catalytically competent "closed" ternary complex.[1] Molecular recognition events are encoded at different sites on the polymerase surface.[1a,2] Recognition events at the insertion site, where the incoming nucleotide triphosphate pairs with the template, are the most critical for replication fidelity.[1c,3] Many structural and biochemical studies have revealed that the Watson–Crick geometry is essential to the recognition of the cognate nucleotides in the active site of DNA polymerases.[1b,3] In fact, Watson–Crick geometry could also be formed by mismatched base pairs via high-energy tautomeric forms, leading to spontaneous mutagenesis in the rare tautomer hypothesis.[4]

To study the nucleotide recognition mechanism of DNA polymerases, many unnatural base pairs have been designed and synthesized.[1c,5] Besides these artificially modified bases, chemical modifications also occur naturally to DNA bases. For instance, 5-methylcytosine (5mC), also known as the fifth base in DNA, is a well-documented epigenetic marker and plays critical roles in gene expression in higher eukaryotes.[6] 5mC can be sequentially oxidized by the TET family proteins to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC); the latter two are excised by the DNA glycosylase TDG, thereby achieving active DNA demethylation in mammals.[7] Because these modifications do not interfere with base-pairing (Figure 1a and Supporting Information, Figure S1), labeling strategies that alter the chemical structures of these epigenetic markers have been developed to enable their detection. Using 5fC as an example, bisulfite-dependent technologies that elegantly couple *O*-ethylhydroxylamine labeling of 5fC ("fCAB-seq"),[8] sodium

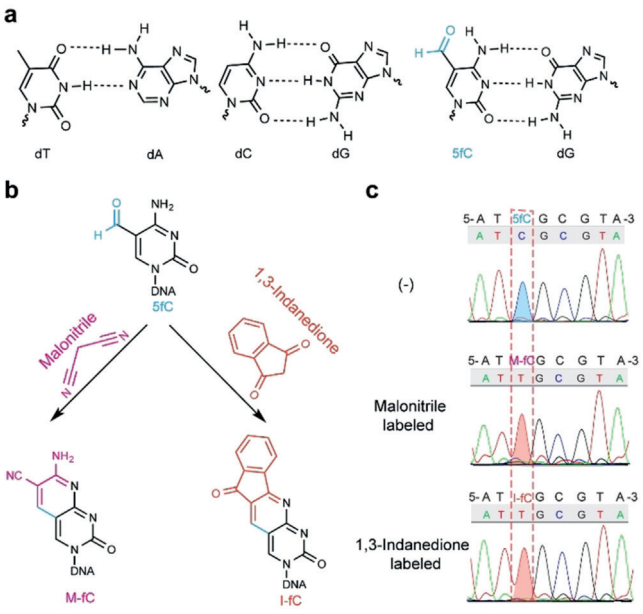

Figure 1. Two unnatural cytosine bases ("M-fC" and "I-fC"), which are chemical labeling adducts of the epigenetic base 5fC, can cause C-to-T transition during DNA amplification. a) Chemical structures of dT:dA, dC:dG and 5fC:dG base pairs. b) Chemical structures of "M-fC" and "I-fC". c) M-fC and I-fC result in C-to-T transition during DNA amplification.

[*] H. Zeng,[+] B. Xia, M. Liu, Dr. C. Zhu, B. He, Prof. C. Yi
School of Life Sciences, Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering and Peking-Tsinghua Center for Life Sciences, Peking University
Beijing 100871 (China)
E-mail: chengqi.yi@pku.edu.cn
Homepage: http://www.yi-lab.org/
Dr. M. Mondal,[+] R. Song,[+] Dr. J. Zhang, Prof. Y. Q. Gao
Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering and Biomedical Pioneering Innovation Center, Peking University
Beijing 100871 (China)
E-mail: gaoyq@pku.edu.cn
Homepage: http://www.chem.pku.edu.cn/gaoyq
[+] These authors contributed equally to this work.
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.201807845.

borohydride reduction of 5fC ("redBS-seq")[9] or M.SssI methyltransferase-mediated discrimination of cytosine bases ("MAB-seq")[10] have been reported to detect 5fC at single-base resolution. Recently, two bisulfite-free 5fC detection methods have been developed[11] that utilize selective chemical labeling of 5fC via malononitrile and 1,3-indanedione, respectively (Figure 1b). The labeling adducts (termed "M-fC" and "I-fC") are read as thymine instead of cytosine by DNA polymerases, causing C-to-T transitions during DNA amplification, thereby enabling 5fC detection (Figure 1c). However, it is difficult to reconcile the C-to-T transition with the accepted mechanism of DNA polymerase recognition.

To characterize the base pairing mode of the two unnatural cytosine bases, we used the host–guest complex (HGC) system in which the M-fC/I-fC:dA base pairs and a normal dT:dA pair can be directly visualized[12] (Figure 2a). Because this C-to-T transition is a general feature of DNA polymerases from different families[11a] (Supporting Information, Figure S2), we chose KlenTaq polymerase (an N-terminally truncated form of Taq polymerase) as a representative DNA polymerase, which is well characterized at the structural and functional level.[13] We found that both the M-fC:dA and I-fC:dA pairs form a Watson–Crick-like geometry within the duplex DNA, without causing noticeable distortion to the DNA backbone (Figure 2b). As a comparison, when the same site of the DNA duplex was replaced by a 5fC:dA pair, a wobble geometry was observed for the 5fC:dA pair (Supporting Information, Figure S3), consistent with the previous finding that 5fC is only marginally mutagenic.[14] An overlay of the M-fC:dA and I-fC:dA with dT:dA pair further showed that these unnatural base pairs mimic the geometry of the natural pair, despite the fact that M-fC and I-fC bases are larger in size than dT. More specifically, the C1′–C1′ distance and λ angles of the M-fC:dA and I-fC:dA pairs are very similar to that of a cognate base pair (Figure 2c). Hence, these crystallographic observations suggest that both M-fC and I-fC can pair with an opposing dA via Watson–Crick geometry, in a position of DNA duplex devoid of contact with polymerase.

To investigate how the unnatural cytosine could pair with an incoming dATP during DNA replication, we solved the structures of the ternary complex in which M-fC base pairs with dATP at the incorporation site of KlenTaq polymerase (Figure 3a). We also solved the structure of a cognate dT:dATP pair in the active site. A comparison of the M-fC:dATP structure and dT:dATP structure showed that the formation of the M-fC:dATP pair is able to induce O helix closure, a signature of the "closed" ternary structure (Figure 3b). Most importantly, the template M-fC also formed a Watson–Crick-like geometry with the incoming dATP, similar to the geometry observed in the host–guest system. Both the C1′–C1′ internucleotide distance and λ angles are roughly the same as that of the cognate dT:dATP base pair (Figure 3c). Thus, the presence of the unnatural cytosine in the template strand is capable of inducing the formation of a Watson–Crick-like pair with an incoming dATP in the incorporation site of the DNA polymerase.

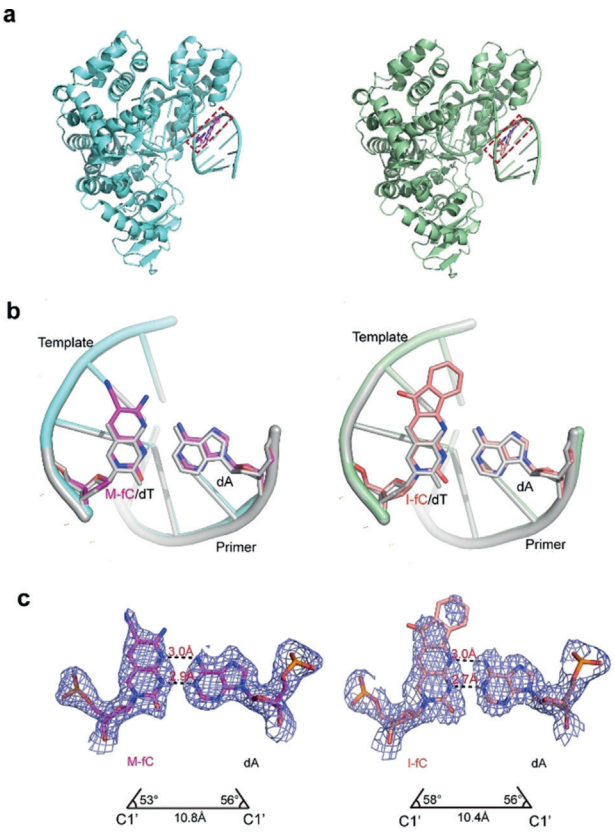

Figure 2. M-fC and I-fC pair with an opposing dA in the Watson–Crick geometry. a) Overall structures of M-fC:dA and I-fC:dA containing DNA in the HGC system. b) Views of base pairing for M-fC:dA and I-fC:dA, respectively. A matched structure with a dT:dA pair (shown in gray) was overlaid. c) M-fC:dA and I-fC:dA 2F_o−F_c electron density maps contoured at 1σ, respectively.

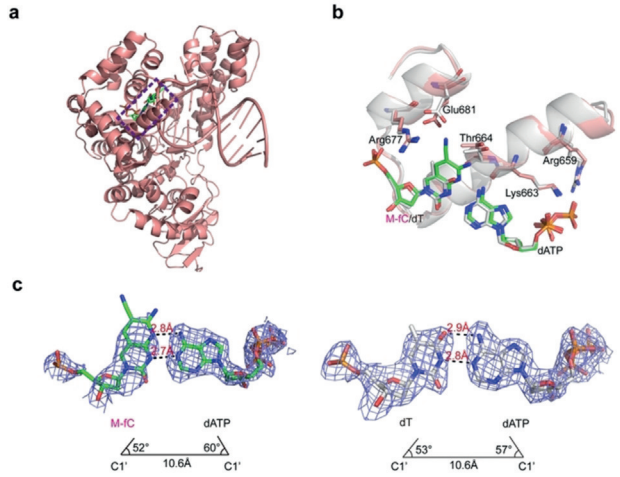

Figure 3. The Watson–Crick geometry of M-fC:dATP pair in the active site of KlenTaq polymerase. a) Overall structure of an M-fC:dATP pair at the incorporation site of KlenTaq polymerase. b) Superposition of M-fC:dATP structure (pink) and dT:dATP structure (gray), indicating the formation of a closed ternary complex. Helices O and O1 and the base pairs in the active site are shown. c) The base pair geometry including the C1′–C1′ distances and λ angles of M-fC:dATP (left) and dT:dATP (right) is shown. The 2F_o−F_c electron density maps contoured at 0.9σ around the base pairs in the active site of KlenTaq polymerase.

Because M-fC and I-fC specifically pair with dATP, but not dGTP, we then investigated the discrimination mechanism against dGTP. As expected, we were not able to obtain a ternary complex with a dGTP opposite the unnatural cytosine, even at higher concentrations of dGTP, using longer soaking times than required for dATP and in the presence of $MnCl_2$. We then performed MD simulations with a dGTP in the active site. It became apparent from our computational analysis that dGTP shears significantly along the major groove (1.5 Å) and forms a stable wobble base pair with M-fC (Supporting Information, Figure S4a,b). A similar pattern was also found in the optimized geometry of the M-fC:dG pair using DFT (Supporting Information, Figure S5). An important feature that distinguishes Watson–Crick pairs from wobble pairs is the symmetry of the λ angle.[15] In the case of the M-fC:dATP pair, the λ angles are largely symmetrically distributed, which resembles the symmetric λ angles of a cognate dT:dATP pair. However, in the case of the sheared M-fC:dGTP pair, λ angle distribution becomes asymmetric (Supporting Information, Figure S4c,d). We also obtained the crystal structure of M-fC:dG pair in the HGC system, and found that the M-fC:dG pair is indeed in the wobble geometry (Supporting Information, Figure S6). Collectively, our analyses suggest that dGTP cannot form a cognate base pair with the unnatural cytosines in the active site and hence may be regarded as a mismatched nucleotide for incorporation by DNA polymerases.

To further investigate the C-to-T transition mechanism caused by the two unnatural cytosines, primer extension experiments of single nucleotide incorporation opposite a templating M-fC/dT with dATP were performed in a time-dependent manner. We found that KlenTaq polymerase extended M-fC much more slowly than dT (Supporting Information, Figure S7a). Quantification of the primer extension results of M-fC incorporation by pre-steady-state kinetics confirmed that KlenTaq polymerase very slowly extended the unnatural cytosine (Supporting Information, Figure S7b,c and Table S3). The low incorporation efficiency indicates that dATP is not the ideal substrate to be incorporated opposite the unnatural cytosines during DNA synthesis; however, given enough time, dA can be faithfully incorporated to near completion. Preferential incorporation of dA opposite a modified or lesioned base by DNA polymerases has been referred to as the A-rule.[16] Indeed, previous studies showed that KlenTaq polymerase followed the A-rule and incorporated dATP opposite an abasic site with much less efficiency than opposite dT.[17] For abasic sites or lesioned DNA bases, the incorporation of dA is preferential, but not specific; yet we found that dA is incorporated in a high-fidelity fashion for M-fC and I-fC. These results imply that, although further studies are needed to explore the atomic reasons for the A-rule, it is a potential explanation of these crystallographic observations that DNA polymerases specifically insert dATP to form the Watson–Crick geometry with unnatural cytosines.

In summary, the unnatural bases M-fC and I-fC can only pair with dA/dATP, but not dG, in the Watson–Crick geometry. It is the geometry-selectivity of DNA polymerase that enabled the C-to-T transition of the unnatural cytosines, and the A-rule may be involved in the formation of the Watson–Crick geometry. The complete transition of M-fC/I-fC to dT by DNA polymerases may have multiple implications. First, this could aid the rational design of novel labeling strategies, and hence sequencing technologies, for more epigenetic markers (for instance, 5hmC and 5caC). Second, the base-pairing specificity of DNA has been exploited to direct the assembly of DNA origami and programmable nanoscale materials.[18] Hence, the chemical labeling-induced base-pairing transition of 5fC could find applications in the development of controllable DNA-based nanomechanical devices. Third, DNA has been utilized as a robust and efficient storage architecture.[19] The "convertible" nature of 5fC and its labeling adducts could encode an additional layer of information during sequencing-based retrieval of digital information.

Acknowledgements

We thank the staff of the BL17U1/BL18U/BL19U1 beamlines of the National Facility for Protein Science Shanghai (NFPS) at the Shanghai Synchrotron Radiation Facility for assistance during data collection. We also thank Dr. Yafen Wang at Wuhan University for experimental advice and Xiaoxue Zhang for technical assistance. Part of the analysis was performed on the High Performance Computing Platform of the Center for Life Science at Peking University. This work is supported by grants from the National Natural Science Foundation of China (nos. 2182570 and 21472009 to C.Y. and 21573006 and 21233002 to Y.Q.G.), SLS-Qidong Innovation Fund (to C.Y.), and the Fok Ying Tung Education Foundation (no. 161018 to C.Y.).

Conflict of interest

The authors declare no conflict of interest.

Keywords: A-rule · base pairing · DNA polymerase · unnatural bases · Watson–Crick geometry

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 130–133
Angew. Chem. **2019**, *131*, 136–139

[1] a) P. J. Rothwell, G. Waksman, *Adv. Protein Chem.* **2005**, *71*, 401–440; b) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219; c) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529.

[2] a) K. A. Johnson, *Annu. Rev. Biochem.* **1993**, *62*, 685–713; b) T. A. Steitz, Y. W. Yin, *Philos. Trans. R. Soc. London Ser. B* **2004**, *359*, 17–23; c) C. M. Joyce, S. J. Benkovic, *Biochemistry* **2004**, *43*, 14317–14324.

[3] H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477–511.

[4] a) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964–967; b) M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285–289; c) V. H. Harris, C. L. Smith, J. W. Cummins, A. L. Hamilton, H. Adams, M. Dickman, D. P. Hornby, D. M. Williams, *J. Mol. Biol.* **2003**, *326*, 1389–1401; d) W. Wang, H. W. Hellinga, L. S. Beese, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17644–17648.

[5] a) M. Kimoto, R. S. Cox 3rd, I. Hirao, *Expert Rev. Mol. Diagn.* **2011**, *11*, 321 – 331; b) A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727 – 733; c) S. Ludmann, A. Marx, *Chimia* **2016**, *70*, 203 – 206.
[6] a) R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89 – 97; b) H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129 – 140; c) Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204 – 220.
[7] a) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935; b) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930; c) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129 – 1133; d) A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334 – 35338; e) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303; f) Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303 – 1307; g) T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012; *Angew. Chem.* **2011**, *123*, 7146 – 7150.
[8] a) C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678 – 691; b) X. Lu, D. Han, B. S. Zhao, C. X. Song, L. S. Zhang, L. C. Dore, C. He, *Cell Res.* **2015**, *25*, 386 – 389.
[9] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435 – 440.
[10] a) F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, G. L. Xu, *Cell Stem Cell* **2014**, *15*, 447 – 459; b) X. Hu, L. Zhang, S. Q. Mao, Z. Li, J. Chen, R. R. Zhang, H. P. Wu, J. Gao, F. Guo, W. Liu, G. F. Xu, H. Q. Dai, Y. G. Shi, X. Li, B. Hu, F. Tang, D. Pei, G. L. Xu, *Cell Stem Cell* **2014**, *14*, 512 – 522; c) H. Wu, X. Wu, L. Shen, Y. Zhang, *Nat. Biotechnol.* **2014**, *32*, 1231 – 1240; d) F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674; e) X. Wu, A. Inoue, T. Suzuki, Y. Zhang, *Genes Dev.* **2017**, *31*, 511 – 523.
[11] a) B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, *Nat. Methods* **2015**, *12*, 1047 – 1050; b) C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720 – 731 e725.
[12] a) K. D. Goodwin, M. A. Lewis, F. A. Tanious, R. R. Tidwell, W. D. Wilson, M. M. Georgiadis, E. C. Long, *J. Am. Chem. Soc.* **2006**, *128*, 7846 – 7854; b) S. Lee, B. R. Bowman, Y. Ueno, S. Wang, G. L. Verdine, *J. Am. Chem. Soc.* **2008**, *130*, 11570 – 11571.
[13] A. Hottin, A. Marx, *Acc. Chem. Res.* **2016**, *49*, 418 – 427.
[14] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782 – 13788.
[15] W. N. Hunter, T. Brown, N. N. Anand, O. Kennard, *Nature* **1986**, *320*, 552 – 555.
[16] D. Sagher, B. Strauss, *Biochemistry* **1983**, *22*, 4518 – 4526.
[17] S. Obeid, N. Blatter, R. Kranaster, A. Schnur, K. Diederichs, W. Welte, A. Marx, *EMBO J.* **2010**, *29*, 1738 – 1747.
[18] a) M. R. Jones, N. C. Seeman, C. A. Mirkin, *Science* **2015**, *347*, 1260901; b) F. Hong, F. Zhang, Y. Liu, H. Yan, *Chem. Rev.* **2017**, *117*, 12584 – 12640.
[19] Y. Erlich, D. Zielinski, *Science* **2017**, *355*, 950 – 954.

Manuscript received: July 9, 2018
Revised manuscript received: September 25, 2018
Accepted manuscript online: November 8, 2018
Version of record online: December 5, 2018

