

RNA Pol II as a sensor of 5caC

Cell Research (2015) 25:1089-1090. doi:10.1038/cr.2015.103; published online 28 August 2015

Cytosines in genomic DNA come in different flavors. A recent paper published in *Nature* reveals that an oxidized variety, 5-carboxylcytosine, is recognized by the elongating RNA polymerase, shedding light on the role of DNA oxidation in transcription regulation.

Methylation of cytosine bases (5mC) occurs in the DNA of many organisms including vertebrates. In addition to 5mC, three new varieties of modifications have been discovered in the recent years: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosines (5caC). They are formed through the sequential oxidation of the methyl group in 5mC by the family of ten-eleven translocation (TET) methylcytosine dioxygenases [1-3].

The role and fate of these oxidized 5mCs (oxi-mCs) have remained largely unexplored. In theory, they can either act independently as epigenetic marks or serve as intermediates toward DNA demethylation. The highly oxidized products 5fC and 5caC can be specifically recognized and excised by the thymine DNA glycosylase TDG, which triggers base excision repair, leading to the conversion of an originally modified cytosine into an unmodified one [2, 4]. The removal of 5fC and 5caC by TDG at least partially contributes to their rarity in embryonic stem cell genomic DNA compared with the 5hmC level. Unlike 5mC and 5hmC, no specific DNA binding proteins have been characterized to selectively recognize 5fC/5caC, except TDG. As such, 5fC/5caC are perceived as transient intermediates generated in the process of active DNA demethylation [5]. A previous study

by Kellinger *et al.* demonstrated that 5fC/5caC induce a transient pause of RNA polymerase II elongation complex (Pol II EC) *in vitro* [6], implying that these modifications may alter gene expression, but the biochemical mechanisms and biological relevance behind this observation remain elusive.

In a recent study published in *Nature*, Wang *et al.* [7] reconstituted yeast elongating RNA polymerase II (Pol II) complexed with a DNA template containing 5caC. In the structure determined at 3.3 Å resolution, 5caC adopts two configurations: a midway conformer and an insert conformer. In the midway conformer, the 5-carboxyl moiety is contacted by a specific hydrogen bond to Q531 from the fork region of pol II subunit Rpb2, a region of the polymerase the authors therefore designate as the 'epi-DNA recognition loop' (Figure 1). This interaction causes a rotation of the side chain of Q531 and a shift of 5caC, which in turn results in dislocation of the incoming GTP nucleotide, impeding its addition to the growing 3' end of the RNA strand. Unlike the midway conformer, the insert conformer does not form a hydrogen bond between the 5-carboxyl moiety and Q531, thus confers no negative effect on GTP addition. Consistent with these findings, *in vitro* nucleotide incorporation assay showed significantly decreased RNA extension at the position opposite to the 5caC in the template, as reported earlier by the same team [6]. To further validate the interaction between 5caC and Q531, the team tested two mutants (Rbp2 Q531H and Q531A) of yeast Pol II and found that the Q531H mutant behaves similarly to the wild-type Pol II

because of the hydrogen bond formation between the His residue (as appears in mammalian Rbp2) and 5caC. However, the Q531A mutation abolished the interaction, suppressing the negative effect of 5caC on GTP incorporation in Pol II transcription.

The structural and biochemical observations above have biological implications *in vivo*. Through a global nuclear run-on assay coupled with deep sequencing (GRO-Seq), the team demonstrated that in mouse ES cells, 5caC and 5fC accumulated in gene bodies by cellular depletion of the TDG glycosylase could significantly reduce the rate of Pol II transcription elongation. They suggest that an array of 5caCs within a transcribed genomic region may serve as speed bumps for the transcriptional machinery to enable fine-tuning of the elongation rate (Figure 1) and propose that the pausing effect might be most significant for the transcription of long genes, such as those most commonly associated with crucial neuronal functions in the brain which is relatively rich in oxi-mCs.

The Q531-5caC interaction appears in the interior of the Pol II-DNA complex, therefore it is less likely to be a crystallization artifact. Indeed, similar interactions may arise in the presence of structural alteration in the DNA template, as found in the cases of elongation complexes arrested by lesions in the DNA template. Moreover, the relatively low resolution of the crystal structure achieved at 3.3 Å is compounded by the conformational heterogeneity of 5caC (midway conformer vs insert conformer), casting concerns on the accuracy of modeling such interactions.

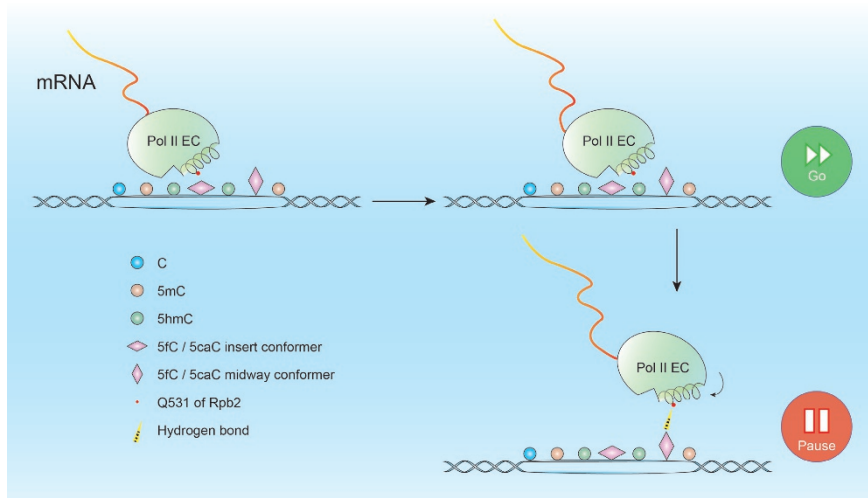


Figure 1 RNA polymerase II engaged in elongation moves along DNA. When it encounters a 5caC in the template strand, the Q531 residue from the conserved epigenetic DNA recognition loop of the Rpb2 subunit forms a hydrogen bond with the 5-carboxyl moiety of 5caC, which causes a 90° rotation of the side chain of Q531, resulting in a shift of 5caC into a so-called 'midway conformer position'. As a consequence, the incoming GTP shifts to a different position, inducing a transient pausing of Pol II. According to this model, 5fC behaves similarly to 5caC. Unmodified cytosine (C), 5mC and 5hmC, like the insert conformers of 5fC/5caC, do not form specific contacts with Q531, thus have no pausing effect on Pol II.

Although the applicability of the 5caC interaction model based on Pol II from yeast which has no modified cytosine in its genome is another concern, the replacement of Q531 by a histidine, the mammalian variant, appears to maintain the suppressive effect on GTP addition and modeling suggests the presence of the hydrogen bond.

The fate of 5caCs that attenuate Pol II progression is unknown. In an interesting scenario, the pausing reaction

of the elongating Pol II could signal the removal of 5caCs from an actively transcribed DNA template by recruiting chromatin and DNA-modifying factors. Notably, the 5fC/5caC glycosylase TDG is involved in transcription through its association with the co-activators CBP/p300 in euchromatin [8]. Transcription-coupled processing of 5fC/5caC thus remains to be a conspicuous mechanism to consider.

Despite the aforementioned caveats

inherent in the study, the work has presented clear evidence that sheds light on how enzymatic DNA oxidation regulates gene expression. The revelation of 5caC's direct impact on the RNA pol II activity is a far-reaching finding. 5caC is obviously different from 5mC as the latter exerts its function on transcription by modulating the affinity of DNA binding proteins and chromatin structure [9]. One next important issue concerns how transient pausing of Pol II by 5caC is released.

Jian-Huang Xue¹, Guo-Liang Xu¹

¹Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Correspondence: Guo-Liang Xu

E-mail: glxu@sibs.ac.cn

References

- 1 Tahiliani M, Koh KP, Shen Y, *et al.* *Science* 2009; **324**:930-935.
- 2 He YF, Li BZ, Li Z, *et al.* *Science* 2011; **333**:1303-1307.
- 3 Ito S, Shen L, Dai Q, *et al.* *Science* 2011; **333**:1300-1303.
- 4 Maiti A, Drohat AC. *J Biol Chem* 2011; **286**:35334-35338.
- 5 Kohli RM, Zhang Y. *Nature* 2013; **502**:472-479.
- 6 Kellinger MW, Song CX, Chong J, *et al.* *Nat Struct Mol Biol* 2012; **19**:831-833.
- 7 Wang L, Zhou Y, Xu L, *et al.* *Nature* 2015; **523**:621-625.
- 8 Tini M, Benecke A, Um SJ, *et al.* *Mol Cell* 2002; **9**:265-277.
- 9 Klose RJ, Bird AP. *Trends Biochem Sci* 2006; **31**:89-97.