The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain

Skirmantas Kriaucionis and Nathaniel Heintz[*]

Laboratory of Molecular Biology, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA.

*To whom correspondence should be addressed. E-mail: heintz@mail.rockefeller.edu

In spite of the importance of epigenetic regulation in neurological disorders, little is known about neuronal chromatin. Cerebellar Purkinje neurons have large and euchromatic nuclei, whilst granule cell nuclei are small and have a more typical heterochromatin distribution. While comparing the abundance of 5-methylcytosine (mC) in Purkinje and granule cell nuclei, we detected the presence of an unusual DNA nucleotide. Using thin layer chromatography (TLC), high pressure liquid chromatography (HPLC) and mass spectrometry (MS) we have identified the nucleotide as 5-hydroxymethyl-2'-deoxycytidine (hmdC). hmdC comprises 0.6% of total nucleotides in Purkinje cells, 0.2% in granule cells, and is not present in cancer cell lines. hmdC is a constituent of nuclear DNA that is enriched in the brain, suggesting a role in epigenetic control of neuronal function.

To enable investigation of Purkinje and granule cell nuclei, we took advantage of the fact that ribosomal proteins are assembled in the nucleolus of all cells. Consequently, transgenic mice containing ribosomal protein L10a fusion with EGFP, for example the *Pcp2 bacTRAP* and *NeuroD1 bacTRAP* lines that express in cerebellar Purkinje and granule neurons, respectively, contain fluorescent nucleoli (*1*, *2*) (fig. S1D, E, fig. S2, D). We utilized this fact, and the fluorescence activated cell sorter (FACS), to obtain ~ 95% pure preparations of Purkinje or granule cell nuclei. While determining the total amount of cytosine methylation in CpG content in Purkinje and granule cell genomic DNA, we consistently observed a significantly lower amount of mC in Purkinje DNA (Fig. 1C) and the presence of a unidentified spot "x" on TLC plates (Fig. 1A). Treating the DNA samples overnight with a mix of RNases A and T1, followed by a one hour with RNase H and V1 (hydrolyzing ssRNA, RNA-DNA hybrids, and dsRNA respectively) did not remove spot "x" (fig. S3A). When the sample was treated with RNase free DNase, spot "x" disappeared, together with the majority of the signal from other spots (fig. S3A). Using the same labeling procedure on total cerebellar RNA, we did not observe spot "x". Spot "x" was not 2'-deoxyuridine monophosphate, which could appear after cytosine

deamination, since a hydrolysate of uracil containing DNA generated a spot migrating significantly below spot "x" (fig. S3B). Fragmenting DNA with *Fok*I restriction endonuclease and filling in with α-P^{32}dATP did not yield spot "x" on the resulting TLC plates, demonstrating that "x" is preferentially found in the context of xpG dinucleotides (fig. S3C). From total nucleosides neighboring G, we estimated that "x" constitutes 0.59%±0.05 in Purkinje cell DNA and 0.23%±0.01 in granule cell DNA. We noticed that the actual increase of xpG is proportional to the decrease of mCpG (Fig. 1C). Considering the quantitative "x" relation with mC, and the fact that the abundance of the other nucleosides was not different between these cell types (Fig. 1C and Fig. 1 legend), we tested the possibility that "x" is 5-hydroxymethyl-2'-deoxycytidine, which is found in T-even bacteriophage DNA (*3*). The results demonstrated that hmdC monophosphate generated a spot that comigrated with the "x" spot on a TLC plate (Fig. 2A). When a synthetic DNA hydrolysate was separated with reverse phase HPLC, hmdC eluted just after the cytosine peak, consistent with the published observations (*4*) (Fig. 2B). HPLC analysis of cerebellar genomic DNA resulted in a small, but reproducible, peak on the HPLC chromatogram in the same position (Fig. 2B). To provide definitive proof that "x" is hmC, the corresponding fraction was analyzed with high precision mass spectrometry. Mass spectra identified the presence of two ions, m/z 142.06 ± 0.01 and 280.11 ± 0.02, which matched the theoretical isotopic molecular weights of ions derived from hmdC - m/z 142.06 and 280.09 respectively (Fig. 3C). MS collision induced fragmentation of the hmdC corresponding fraction from synthetic DNA produced the same ions (fig. S4). Taken together, these data demonstrate the presence of hmC in mouse cerebellar DNA. We were unable to detect hmC in four different cell lines of mouse and human origin (fig. S5A). The distribution of hmC in mouse tissues displays the enrichment exclusively in brain, with higher abundance in cortex and brainstem (fig. S5B).

It is unlikely that the hmC we observe *in vivo* is a product of DNA damage (*5*, *6*). We do not observe any other DNA damage products such as 8-oxoguanine, a preferential target

for oxidants (*7*) or thymidine glycol, which is produced *in vitro* by the oxidation of mC (*6*). In addition, hmC is enriched in brain, but not in other metabolically active non-proliferating tissues (fig. S5B). Finally, contrary to what one would expect for oxidative DNA damage, we find that there is no correlation between the age of adult mice and the amount of hmC in Purkinje and granule cells (fig. S6).

An early publication suggesting the presence of hmC in mammalian genomes (*8*) has not been reproduced by others (*9, 10*). If treated with bisulfite, hmC has been suggested to produce cytosine 5-methylsulfonate, which would be deaminated even at a slower rate than mC (*11*) leading to the interpretation of hmC as mC after bisulfite sequencing. Although active DNA demethylation is considered to occur, no enzyme has been found that can remove the methyl group from 5-methylcytosine (*12*). The presence of hydroxylated methyl group could either indicate an intermediate for oxidative demethylation or a stable end-product, which eliminates the need of removal of the methyl group, by modulating affinity of proteins which read the mC signal in non-dividing neuronal cell types. The result that methyl-CpG binding domain of MeCP2 protein has a lower affinity towards sequences containing hmC supports this notion (*13*). It is notable that hmC is nearly 40% as abundant as mC in Purkinje cell DNA. Given the critical role of mC in epigenetic regulation of the genome, we believe that hmC may also have an important biological role *in vivo*.

References and Notes

1. J. P. Doyle *et al.*, *Cell* **135**, 749 (2008).
2. M. Heiman *et al.*, *Cell* **135**, 738 (2008).
3. G. R. Wyatt, S. S. Cohen, *Biochem. J.* **55**, 774 (1953).
4. S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **25**, 553 (1997).
5. A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **30**, 4068 (2002).
6. S. Zuo, R. J. Boorstein, G. W. Teebor, *Nucleic Acids Res.* **23**, 3239 (1995).
7. J. Cadet, T. Douki, J. L. Ravanat, *Nat. Chem. Biol.* **2**, 348 (2006).
8. N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **126**, 781 (1972).
9. R. M. Kothari, V. Shankar, *J. Mol. Evol.* **7**, 325 (1976).
10. J. H. Gommers-Ampt, P. Borst, *FASEB J.* **9**, 1034 (1995).
11. H. Hayatsu, M. Shiragami, *Biochemistry* **18**, 632 (1979).
12. S. K. Ooi, T. H. Bestor, *Cell* **133**, 1145 (2008).
13. V. Valinluck *et al.*, *Nucleic Acids Res.* **32**, 4100 (2004).
14. B. H. Ramsahoye, *Methods Mol. Biol.* **200**, 9 (2002).
15. We thank Betsy Gauthier for the excellent technical assistance, Svetlana Mazel, Christopher Bare, Xiaoxuan Fan for flow cytometry advice and nuclei sorts, Haiteng Deng and Joseph Fernandez for acquirement of MS data and help with HPLC. We are grateful to Heintz lab

Supporting Online Material

Fig. 1. Quantification of mC and "x" abundance in Purkinje and granule neurons. **A**, 2D TLC separation of nucleoside monophosphates from genomic DNA in Purkinje and granule cells. **B**, reference map to the TLC spots (A is dAMP, etc) (*14*), with added "x" position. **C**, percentage shows the abundance of a nucleotide neighboring G. Error bars show SEM (n=11), p values were derived from Matt-Whitney statistics. Neither abundance of dTMP or dAMP were different between the samples (p=0.743 and p=0.793 respectively).

Fig. 2. 2D TLC, HPLC and MS identification of hmC. **A**, 2D TLC analysis of synthetic DNA templates indicates that hmC comigrates with the "x" spot (Fig 1). **B**, HPLC chromatograms (A, 254 nm) of the nucleosides derived from synthetic and cerebellum DNA. The peaks were identified using MS. The black arrow points to the peak, which elutes at the same time as hmdC. **C**, MS of the fraction corresponding to the HPLC peak indicated above. Closed black arrows indicate the masses of 5-hydroxymethylcytosine and 5-hydroxymethyl-2'-deoxylcytidine sodium ions (structures are shown on the insets). Open arrows indicate the ions generated by 2'-deoxycytidine, which elutes in a large nearby peak and spills over into the analyzed fraction.

A Purkinje cells | Granule cells

B

C

A Synthetic DNA C   Synthetic DNA C+hmC

B

dC hmdC   mdC   dG   dT   dA

Synthetic DNA
C+hmC

Synthetic DNA
C

Cerebellum
DNA

A,
mAU   80
60
40
20
0

16   18   20   22   24   26   28   30   t, min

C

Intensity, counts

$[$ structure $] + H^+$

$[$ structure $] + Na^+$

2.3e4
2.1e4
1.9e4
1.7e4
1.5e4
1.3e4
1.1e4
9000.0
7000.0
5000.0
3000.0
1000.0
0.0

280.1103

149.0332
157.0939

142.0701

202.1946

244.2810 250.0972
266.0723
261.1485
273.1870

130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280
m/z, amu