Check for updates

A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition GDCh Chemie

www.angewandte.org

## Accepted Article

**Title:** Intragenomic decarboxylation of 5-carboxy-2'-deoxycytidine

**Authors:** Ewelina Kamińska, Eva Korytiaková, Andreas Reichl, Markus Müller, and Thomas Carell

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *Angew. Chem. Int. Ed.* 10.1002/anie.202109995

**Link to VoR:** https://doi.org/10.1002/anie.202109995

WILEY-VCH

10.1002/anie.202109995

WILEY-VCH

COMMUNICATION

# Intragenomic decarboxylation of 5-carboxy-2'-deoxycytidine

Ewelina Kamińska,[#[a]] Eva Korytiaková,[#[a]] Andreas Reichl[[a]], Markus Müller[[a]] and Thomas Carell*[[a]]

**Abstract:** *Cellular DNA is composed of four canonical nucleosides (dA, dC, dG and T), which form two Watson-Crick base pairs. In addition, 5-methylcytosine (mdC) may be present. The methylation of dC to mdC is known to regulate transcriptional activity. Next to these five nucleotides, the genome, particularly of stem cells, contains three further dC derivatives, which are formed by stepwise oxidation of the methyl group of mdC with the help of Tet enzymes. These are 5-hydroxymethyl-dC, (hmdC), 5-formyl-dC (fdC) and finally 5-carboxy-dC (cadC). It is believed that fdC and cadC are converted back into dC, which establishes an epigenetic control cycle that starts with methylation of dC to mdC, followed by oxidation and removal of fdC and cadC. While fdC was shown to undergo intra-genomic deformylation to give dC directly, a similar decarboxylation of cadC was postulated but not yet observed on the genomic level. Here, we show, using metabolic labelling in several cell types, that cadC decarboxylates, which confirms that both fdC and cadC are nucleosides that are directly converted back to dC within the genome by C-C bond cleavage.*

5-Formyl-dC (fdC) and 5-carboxyl-dC (cadC) are nucleosides that are found in significant amounts in neurons and stem cells.[1] They are formed by oxidation of 5-methyl-dC (mdC) by the action of Tet enzymes via 5-hydroxymethyl-dC (hmdC).[2, 3] HmdC is found in these genomes in large quantities. The initial methylation of dC to mdC is performed by the dedicated methyltransferases Dnmt1, 3a and 3b.[4-7] The higher oxidized mdC derivatives, fdC and cadC are known to be removed by the repair glycosylase Tdg, which cleaves the glycosidic bond between the sugar and the corresponding base.[8, 9] In result, there is formation of abasic sites processed by AP endonuclease, and finally replaced by an unmodified dC (Fig. 1a).[10]

Since the discovery that methylation of dC is followed by oxidation chemistry, it was postulated that fdC and cadC might directly deformylate or decarboxylate to give dC. Chemically, these C-C bond cleavage reactions have the advantage that potentially harmful abasic site intermediates, formed during Tdg-mediated active demethylation, will not occur.[11] While for fdC, deformylation was shown to occur *in vivo*, for cadC just a putative decarboxylation mechanism was postulated *in vitro* so far.[12, 13] It is still unknown whether decarboxylation of cadC occurs in stem cells.[13-16] Here, we use our previously described metabolic labelling approach to prove that cadC, if present in the genome of stem- and somatic cells, does decarboxylate.[17, 18]

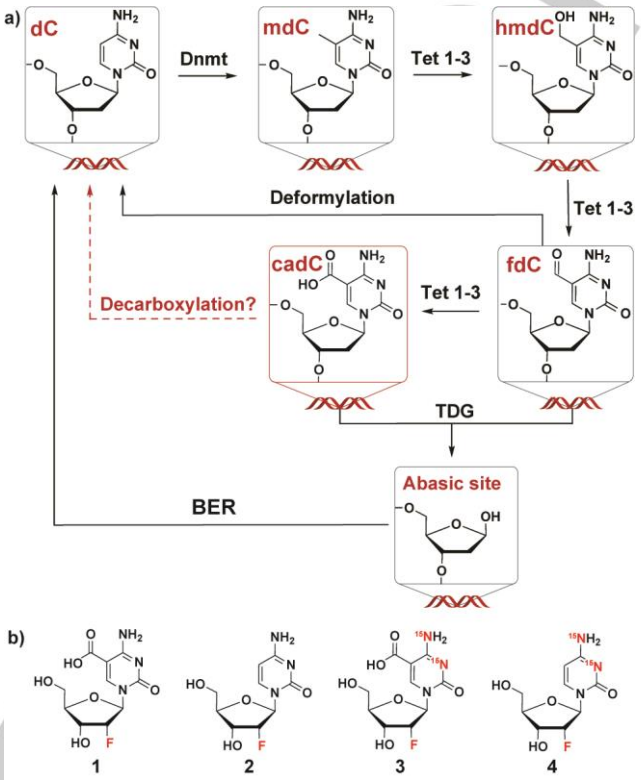


**Figure 1**. **a**) Depiction of the active demethylation pathways via Tdg-mediated excision or direct deformylation and decarboxylation. **b**) Depiction of the metabolically fed nucleosides 2'-F-cadC **1** and of the product nucleoside 2-'F-dC **2** formed after decarboxylation. MS reference compound $[^{15}N]_2$ 2'F-cadC **3** and $[^{15}N]_2$ 2'F-dC **4**.[17]

A direct demethylation reaction in DNA by C-C bond cleavage will give a product that is identical with the natural dC. Therefore, it is important to incorporate a reporter nucleoside (cadC*) into the genome of the cells, which generates a decarboxylated dC*. This product must be detectable with high accuracy in the presence of an overwhelming amount of natural dC. In the past, we successfully experimented with 2'fluoro-labelled nucleosides. Thus, for this study we decided to use 2'F-cadC (**1**) as cadC* (Fig. 1b).[17] The synthesis of **1** and its triphosphate was reported by us previously.[19, 20] The 2'F-atom ensures that the formed decarboxylated product 2'F-dC **2** is readily detectable by UHPLC-MS/MS due to the *m/z* = +19 Da mass shift relative to dC. The specific shift in retention time for fluorinated compounds also allows to distinguish them from canonical dC. This reduces an overlap of 2'F-dC with dC, which avoids ion suppression that makes quantification of even very small quantities of product possible. The 2'F-atom served also a second purpose. Natural cadC is barely detectable in wildtype cells.[8] The levels, however, increase by almost two orders of magnitude, when the base excision repair (BER) pathway is interrupted by knocking out the *TDG* gene (Fig. SI-6).[21] The 2'F substitution has the same effect. It blocks the BER process, which leads to higher amounts of detectable, incorporated 2'F-cadC (**1**).[19]

[a] B. Sc. E. Kamińska, M. Sc. E. Korytiaková, M. Sc. Andreas Reichl, Dr. M. Müller, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 Munich, Germany
http://www.carellgroup.de
* Corresponding author: Thomas.Carell@lmu.de
# these authors contributed equally

Supporting information for this article is given via a link at the end of the document.

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/anie.202109995




As a reference compound for exact quantification of **1,** we synthesized the isotopologue [$^{15}$N]$_2$ 2'F-cadC **3**, as an internal standard. Quantification of **2** required the standard [$^{15}$N]$_2$ 2'F-dC **4** (Fig. 1b).

A disadvantage of using 2'F-cadC for studying natural decarboxylation is of course the unnatural character of the nucleoside. In order to investigate if the 2'F-atom influences the decarboxylation reaction, we saturated the C5-C6 bond of the cadC - and 2'F-cadC-methylesters and studied the spontaneous decarboxylation behavior after ester cleavage.[13] We found that both compounds decarboxylate in a similar manner, reassuring us that the effect of the 2'F-atom on the studied decarboxylation reaction is negligible (Fig. SI-8). When we started to experiment with 2'F-cadC **1** *in cellulo***,** we learned that it is problematic to label the cellular genome with this compound. The negatively charged 2'F-cadC was only to a small extent taken up by cells and in addition, its intracellular conversion to the triphosphate, as needed for incorporation into the genome, happened to be inefficient as well. It is known, that the phosphorylation of cadC by kinases to the corresponding nucleotides is inefficient.[22]

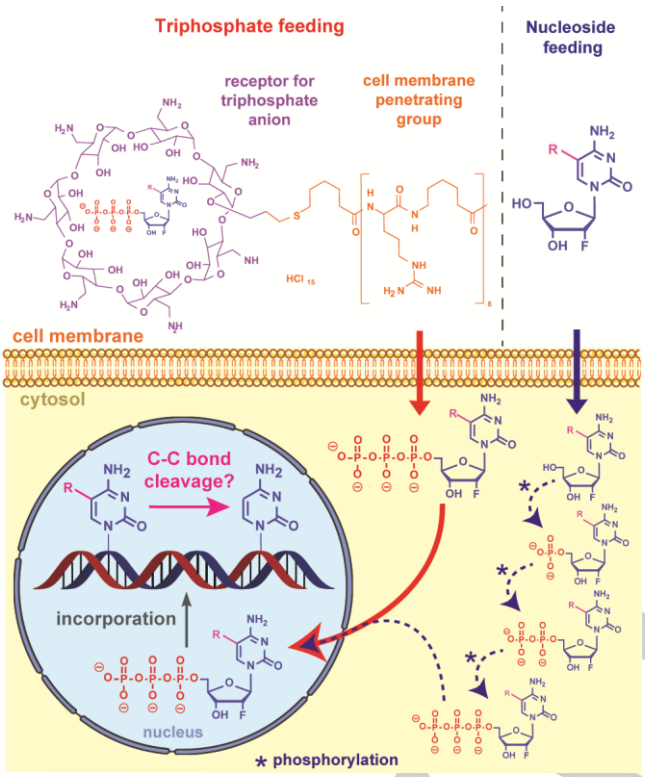


**Figure 2**. Depiction of the experimental setup for the inefficient delivery of **1** (blue path) and the accelerated transfer of **1**-TP into cells with the help of the calixarene transporter (red path).

Although we detected 2'F-cadC (**1**) in the cytosol, we were unable to detect incorporated **1** in the genome. We had to continuously feed the cells for 3 days with **1** to reach detectable, but unquantifiable levels of 2'F-cadC **1**. After intensive experimentation with different 2'F-cadC delivery methods, we finally succeeded with the help of a cyclodextrin transporter, which is modified with a cell-penetrating peptide derivative (Fig. 2).[23] This transporter encapsulates nucleoside triphosphates and allows them to be transferred across the cell membrane. Application of this transporter was indeed successful. It allowed us to deliver the **1**-TP with only one 30 min feeding pulse in tricine buffer. After the feeding, the cells were washed, and fresh medium was applied (Fig. 2).

In the first experiment, we fed Neuro-2a cells for 30 min with the **1**-TP loaded transporter. The cells were harvested after 24 hours. We then isolated the free nucleotide pool and investigated it regarding the presence of 2'F-cadC **1** and 2'F-dC **2**. In the cytosol, we detected to the presence of delivered **1**. In addition, we saw no **2** (Fig. 3a). Together, the data show that 2'F-cadC **1** is a stable compound that, based on our data, does not spontaneously decarboxylate during delivery or under physiological conditions in cells.

Next, we harvested the cells using RLT buffer (Qiagen) supplemented with 400 µM of 2,6-di-*t*Butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM) as well as β-mercaptoethanol (1:100). The genomic DNA was isolated using a spin column kit (Zymo Research) and digested according to our established method.[18] In brief, the isolated DNA was incubated for 5 min at 95 °C, cooled down on ice and incubated for 4 h at 37°C with the Degradase digestion mixture (Zymo Research) with the addition of the isotope-labelled internal standards. Next, the samples were diluted with 450 µL of water and were extracted with chloroform. After lyophilization of the aqueous phase, the digested samples were resuspended in water, filtered and analyzed by UHPLC-MS/MS.[18]

In order to prove that this digestion method is efficient for 2'F-cadC-containing DNA, we digested a short ssDNA containing a synthetically embedded 2'F-cadC **1** (Fig. SI-1). Best results were obtained with the Degradase mix (Fig. SI-1).[24] We next tested the digestion using a 147-base long dsDNA (Widom 601)[25] containing **1**. Using LC-MS, we detected all canonical DNA bases (dA, dC, dG, T) and in addition, 2'F-cadC **1** at the expected level, confirming that efficient digestion is possible using the Degradase method. In contrast to this, however we noticed that when we added the same amount of a 2'F-cadC-containing DNA strand to normal genomic DNA, we obtained an astonishingly small signal for 2'F-cadC (**1**). Indeed, only 10% of the expected signal was detected, arguing that the detectability of 2'F-cadC (**1**) is strongly reduced in a complex environment (Fig. SI-2). Due to this strong signal suppression, we therefore abstained from exact quantification of **1,** using internal standard **3** in all further experiments.

To investigate the decarboxylation process, we isolated and digested genomic DNA from Neuro-2a cells fed with **1**-TP. In the above-described method, indeed we clearly detected a signal for 2'F-cadC **1** in the genome of cells fed with **1**-TP based on its retention time, which was identical with the reference compound **3** and its fragmentation pattern, which was indistinguishable from **3** (Fig. 3b). To our delight, we also detected the decarboxylated product **2**, which was not seen in the cytosolic fraction. Compound **2** was detected at levels of 2.4x10$^{-5}$ per dN, 24 h post feeding. The result shows that while the decarboxylated product **2** does not form in the cytosol by spontaneous decarboxylation, it is present in the genome, arguing that decarboxylation takes place when 2'F-cadC **1** is incorporated into genomic DNA.

Next, we investigated if the decarboxylation signals forms in a dose dependent manner (Fig. 3c). Feeding Neuro-2a cells with final concentrations of 100 µM and 300 µM of the **1**-TP loaded transporter gives indeed an increasing signal for 2'F-dC **2** (Fig. 3d). We also see an increase in the MS signal intensity of 2'F-cadC **1** (Fig. 3c), as expected.

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/anie.202109995




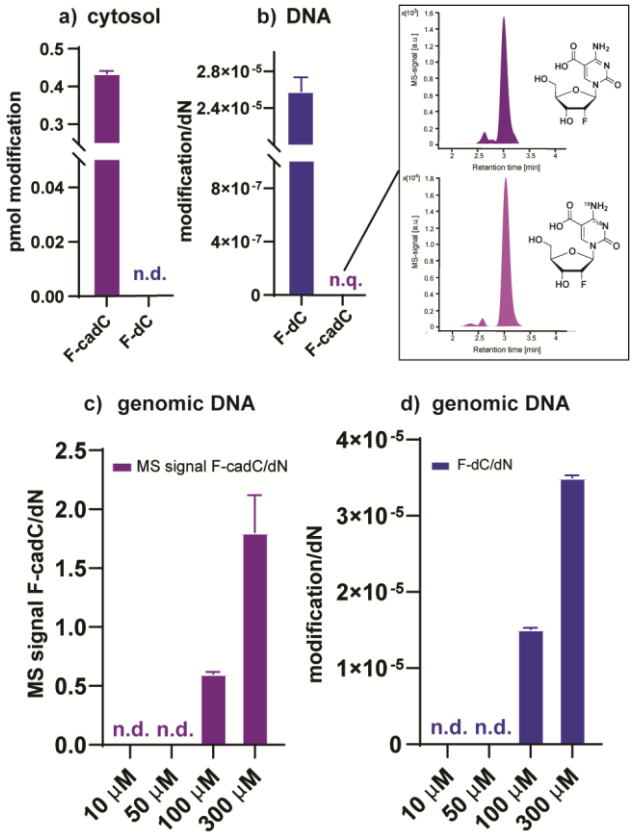


**Figure 3**. Metabolic feeding of 2'-F-cadCTP to Neuro-2a cells. Investigation of the decarboxylation reaction in the cytosol (a) and in genomic DNA (b). Dose response data showing increased MS signal of 2'-F-cadC **1** (c) and levels of 2'-F-dC **2**, (d) after 24h with increasing feeding concentration.

Then, we performed a time-course experiment to investigate at which timepoint decarboxylation becomes detectable (Fig. SI-5). We discovered that 2'F-dC **2** is detectable only after about 8 h at the earliest, which shows that the C-C bond cleavage has a late onset (Fig. SI-5a). As a control experiment, to further exclude that we detect the incorporation of an impurity, we co-fed a 100 µM solution of **1**-TP with a 1% impurity of 2'F-dCTP **2**-TP (Fig. SI-5b). In this experiment, we see again genome incorporated **1** and 2'F-dC **2**, but now, **2** appeared already after 30 min.[17] Together, the data show that the detected 2'F-dC (**2**) is formed from 2'F-cadC (**1**) by C-C bond cleavage.

In the next experiment, we analyzed deformylation of 2'F-fdC (**5**) and decarboxylation of 2'F-cadC (**1**) side by side (Fig. 4a). We fed Neuro-2a cells with 2'F-fdC-TP in one experiment and with 2'F-cadC-TP in the second. We used the same amount (400 µM) of both materials and of the transporter. We then exchanged the medium and allowed the cells to recover for 24 h. Afterwards, the cells were harvested, the genomic DNA was isolated, digested and analyzed by UHPLC-MS/MS using the synthetic internal standards **4** as well as $[^{15}N]_2$ 2'F-fdC (**6**) for quantification (Fig. 4c).

We detected clearly the fed starting material 2'F-fdC **5** and again only just traces (due to ion suppression and/or low incorporation) of 2'F-cadC **1**. Despite this, the amount of the C-C bond cleavage product 2'F-dC **2** was higher when we fed **1**-TP compared to **5**-TP (Fig. 4a). The exact quantification of **2** in both experiments, using the isotopically labelled material **4**, clearly showed more 2'F-dC **2** derived from decarboxylation than deformylation. This interesting result shows that we either dramatically underestimate the genomic incorporation of 2'F-cadC **1** due to the large ion suppression, or that 2'F-cadC **1** decarboxylates to 2'F-dC **2** more efficiently than 2'F-fdC **5** deformylates. Based on earlier chemical studies and our current knowledge of both processes, we speculate that chemically-favoured decarboxylation is the contributing factor.[26]

We next asked the question if the product nucleoside 2'F-dC (**2**) is remethylated. To this end, we quantified the levels of 2'F-mdC (**7**) in both experiments using the internal standard $[D_3]$ 2'F-mdC (**8**). We clearly detected **7** in both experiments (Fig. 4a). Again, higher amounts were detected from 2'-F-cadCTP feeding. This finding confirms that after deformylation or decarboxylation, the formed product **2** is remethylated, which suggests the presence of a putative regulatory chemical cycle that starts with methylation of dC by Dnmt enzymes to mdC, followed by oxidation of mdC to hmdC, fdC and cadC and then deformylation and decarboxylation of fdC and cadC to dC, which could start a new circle.[27] Interestingly, when we measured the levels of **7** formed by methylation of **2** 24 h post **1**-TP feeding, we found only 0.2% methylation. When the same measurement was performed after 72 h, we measured 2.1%, very close to the natural methylation levels of Neuro-2a cells of about 2.5% (Fig. SI-7). This result shows that while an early harvest allows us to detect higher amounts of 2'F-dC, the levels of 2'F-mdC are underestimated, potentially because the methyltransferases do not have sufficient time to achieve remethylation.

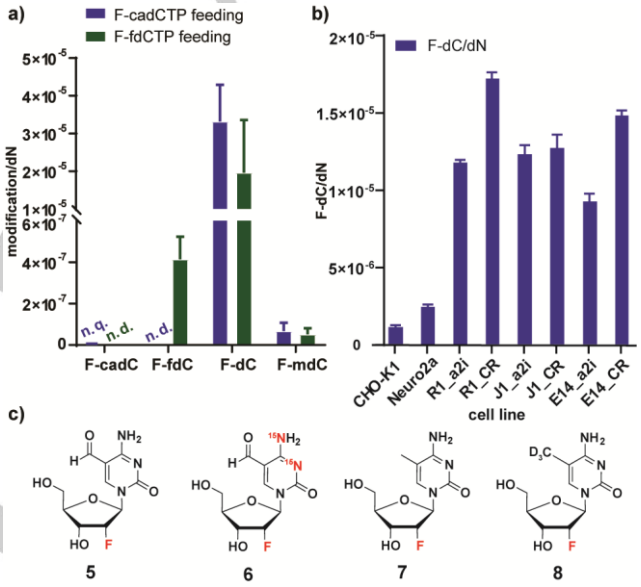


**Figure 4. a**) Comparison of deformylation and decarboxylation by feeding of **1**-triphosphate and **5**-triphosphate as well as investigation of the remethylation of the product **2**. **b**) Decarboxylation in different cells (J1, R1, E14). **c**) Depiction of 2'F-fdC (**5**) and 2'F-mdC (**7**) the reference compounds $[^{15}N_2]$ 2'F-fdC (**6**) and $[D_3]$ 2'F-mdC (**8**) needed for exact quantification.[17]

Finally, we studied how much decarboxylation is taking place in different cells (Fig. 4b). We investigated the process in somatic (Neuro-2a, CHO-K1) and in mouse embryonic stem cells (J1, E14, R1). The stem cells were investigated at the pre-implantation and post-implantation stages, obtained by culturing the cells in either a2i or CR media as described in the SI. The data depicted in Fig. 4b show that upon feeding of **1**-TP at 100 µM for 24 h, **2** is clearly detectable in all cases but the levels vary between the different cell types. The highest levels of **2** were formed in stem cells cultured under CR conditions. These are the cells representing post-implantation embryos, that naturally have the highest methylation levels due to epigenetic reprogramming during cell lineage differentiation.[28] In

Accepted Manuscript

This article is protected by copyright. All rights reserved.

general, stem cells show an about ten times higher decarboxylation activity than somatic cells, which underpins the potential epigenetic importance of the process.

In summary, the presented data show that next to deformylation of fdC, we also need to consider decarboxylation of cadC as a mechanism for active demethylation. All further efforts now need to be concentrated at finding the cellular entities or circumstances that enable these C-C bond cleavage reactions. Although the here reported data clearly point to the existence of decarboxylation, we need to emphasize that cells are complex systems and we feed an unnatural compound. Without clear identification of the biological entity responsible for the process, we cannot completely rule out that unknown processes other than intragenomic decarboxylation are responsible for the measured data. During the review process of this manuscript Feng and coworkers showed an incorporation of the F-carboxylcytidine as a nucleoside and interestingly managed to detect the decarboxylation of cadC to dC much earlier on.[29]

## Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for financial support via SFB1309 (PID 325871075) and SFB1361 (PID 393547839). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912). Additional funding was provided by the Volkswagen Foundation (EvoRib) and the DFG priority program SPP1784 (PID 277203618).

**Keywords**: epigenetics; carboxylcytidine; decarboxylation; C-C bond cleavage; active demethylation

[1] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
[2] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.
[3] 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.
[4] Z. Li, H. Dai, S. N. Martos, B. Xu, Y. Gao, T. Li, G. Zhu, D. E. Schones, Z. Wang, *Genome Biol.* **2015**, *16*.
[5] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem* **2013**, *288*, 9084-9091.
[6] C. Luo, P. Hajkova, J. R. Ecker, *Science* **2018**, *361*, 1336-1340.
[7] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257.
[8] 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.
[9] L. S. Pidugu, Q. Dai, S. S. Malik, E. Pozharski, A. C. Drohat, *J. Am. Chem. Soc.* **2019**, *141*, 18851-18861.
[10] H. E. Krokan, M. Bjoras, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012583-a012583.
[11] A. J. Wossidlo M, Sebastiano V, Lepikhov K, Boiani M, Reinhardt R, Schöler H, Walter J., *EMBO J.* **2010**, *29*, 1877-1888.
[12] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.
[13] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
[14] G. Song, G. Wang, X. Luo, Y. Cheng, Q. Song, J. Wan, C. Moore, H. Song, P. Jin, J. Qian, H. Zhu, *Nat. Commun.* **2021**, *12*, 795.
[15] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887.
[16] Y. Feng, N.-B. Xie, W.-B. Tao, J.-H. Ding, X.-J. You, C.-J. Ma, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *CCS Chemistry* **2021**, *3*, 994-1008.
[17] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78.
[18] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283-312.
[19] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368-4371.
[20] A. S. Schröder, E. Parsa, K. Iwan, M. Wallner, S. Serdjukow, T. Carell, *Chem. Commun.* **2016**, *52*, 14361-14364.
[21] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.
[22] M. Zauri, G. Berridge, M.-L. Thézénas, K. M. Pugh, R. Goldin, B. M. Kessler, S. Kriaucionis, *Nature* **2015**, *524*, 114-118.
[23] Z. Zawada, A. Tatar, P. Mocilac, M. Buděšínský, T. Kraus, *Angew. Chem. Int. Ed.* **2018**, *57*, 9891-9895.
[24] F. Yuan, Y. Bi, J.-Y. Zhang, Y.-L. Zhou, X.-X. Zhang, C.-X. Song, *RSC Adv.* **2019**, *9*, 29010-29014.
[25] W. J. Lowary PT, *J Mol. Biol.* **1998**, *276*, 19-42.
[26] E. Korytiaková, E. Kamińska, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2021**, *60*, 16869-16873.
[27] A. Parry, S. Rulands, W. Reik, *Nat. Rev. Genet.* **2021**, *22*, 59-66.
[28] S. Takahashi, Kobayashi, S., Hiratani, I., *CMLS* **2018**, *75*, 1191-1203.
[29] Y. Feng, J.-J. Chen, N.-B. Xie, J.-H. Ding, X.-J. You, W.-B. Tao, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *Chem. Sci.* **2021**.

Accepted Manuscript

This article is protected by copyright. All rights reserved.