A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition GDCh Chemie

www.angewandte.org

Accepted Article

Title: Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells

Authors: Alexander Schön, Ewelina Kaminska, Florian Schelter, Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, and Thomas Carell

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.202000414
Angew. Chem. 10.1002/ange.202000414

Link to VoR: http://dx.doi.org/10.1002/anie.202000414
http://dx.doi.org/10.1002/ange.202000414

WILEY-VCH

10.1002/anie.202000414

WILEY-VCH

RESEARCH ARTICLE

Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells

Alexander Schön,[#] Ewelina Kaminska,[#] Florian Schelter,[#] Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers and Thomas Carell*

Abstract: *5-Methyl-deoxycytidine (mdC) is considered the fifth base of the genetic system. Its presence in promoter elements takes part in silencing the corresponding gene. Its removal, associated with the reactivation of silenced genes, is possible via an enigmatic active demethylation process that requires the oxidation of mdC to 5-hydroxymethyl-deoxycytidine (hmdC) and further on to 5-formyl-deoxycytidine (fdC) and 5-carboxy-deoxycytidine (cadC) with the help of α-ketoglutarate dependent Tet-oxygenases. The removal of mdC occurs at the level of the fdC and cadC oxidation state. One pathway involves the action of a dedicated glycosylase (TDG), which cleaves fdC out of the genome, allowing its replacement by dC. A second is proposed to involve C-C bond cleavage that converts fdC directly into dC. While the TDG-induced repair-based pathway is well characterized, information about the C-C bond cleaving process is very limited. Here, we report the synthesis of a novel 6-aza-5-formyl-deoxycytidine (a-fdC) probe molecule, which was fed to various somatic cell lines and induced mESCs, together with a 2'-fluorinated fdC analog (F-fdC). While deformylation of F-fdC was clearly observed in vivo, it did not occur with a-fdC, suggesting that the C-C bond cleaving deformylation reaction must be initiated by a nucleophilic activation.*

5-Formyl-deoxycytidine (fdC) **1**, is a dC-derived nucleobase that is found in stem cells during early development and in brain.[1-5] These tissues are particularly rich in 5-hydroxymethyl-deoxycytidine (hmdC) from which fdC **1** is produced.[6, 7] Formation of hmdC and fdC requires oxidation reactions that are performed by α-ketoglutarate-dependent Tet enzymes with 5-methyl-deoxycytidine (mdC) being the initial starting molecule.[8-10] This cascade of oxidation reactions is a part of an active demethylation process, in which mdC as a silencer of transcription is replaced by unmodified dC.[11] The central molecule that is removed seems to be fdC.[12, 13] It can be cleaved out of the genome by a dedicated DNA glycosylase, which creates an abasic site, that is further processed leading to the insertion of an unmodified dC.[14] Because abasic sites are harmful DNA repair intermediate can cause genome instability. It was early on suggested that fdC might be directly deformylated to dC by C-C bond cleavage.[15, 16] Evidence for the existence of such a direct deformylation process was recently reported.[17] Model studies showed that direct deformylation of fdC and potentially also decarboxylation of 5-carboxy-deoxycytidine (cadC) is indeed possible.[15] Nevertheless, it requires activation of the nucleobases by a nucleophilic addition to the C6 position. For fdC, an additional hydrate formation step of the formyl group seems to be necessary as depicted in Fig. 1A. Although the activation with a helper nucleophile is well known as the central mechanistic process during methylation of dC to mdC by DNA-methyltransferases (Dnmts),[18, 19] it remains yet to be confirmed if such activation occurs *in vivo* as well. Here, we investigated this hypothesis with the two probe molecules, 2'-fluorinated-fdC **2** (F-fdC) and 6-aza-fdC **3** (a-fdC). Both compounds were simultaneously fed to different cell types, including primed stem cells. This led to random incorporation of these bases at the "C" sites into the respective genomes. Furthermore, it led to the presence of F-fdC and a-fdC not only at CpG sites. Ultrasensitive UHPLC-QQQ-MS^2 was subsequently used to interrogate the chemical processes that occur at F-fdC and a-fdC in the genomes. The data show that while F-fdC is efficiently deformylated, this does not occur for a-fdC. The only difference between both nucleobases is the presence of an in-ring nitrogen atom (6-aza atom), which features a lone pair that prohibits a nucleophilic addition. The experiment provides consequently strong evidence that the nucleophilic activation *in vivo* is the central governing mechanistic event that is required for C-C bond cleavage.

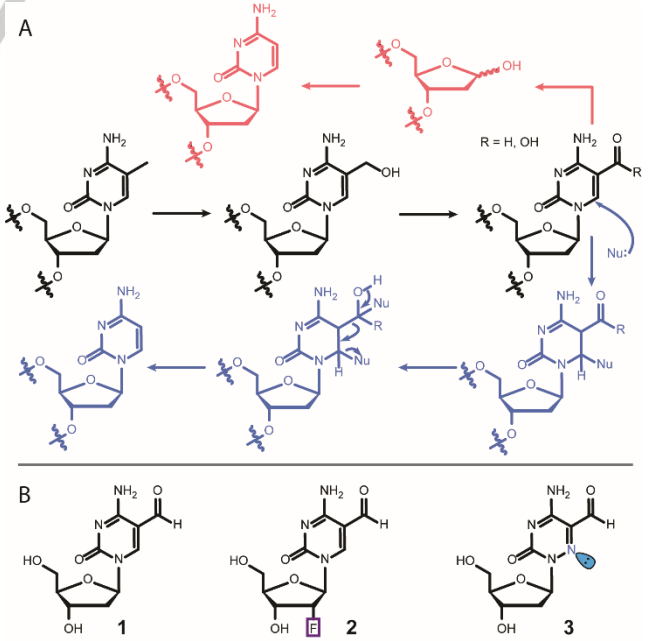


Figure 1. **(A)** Depiction of the mdC removal pathways that involve oxidation to hmdC, fdC and cadC followed by either base excision repair (magenta) or C-C bond cleavage (blue). **(B)** Representation of fdC **1** and the two probe molecules **2** and **3** used for this study.

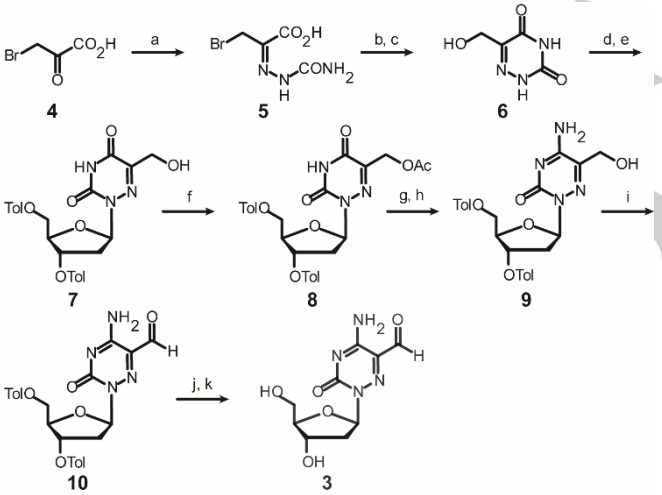
[a] M. Sc. A. Schön, B. Sc. E. Kaminska, M. Sc. F. Schelter, M. Sc. E. Ponkkonen, M. Sc. E. Korytiaková, Dr. S. Schiffers and Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians Universität München
Butenandtstr. 5-13, 81377 München
Fax: (+) 49 2180 77756
E-mail: thomas.carell@lmu.de
Homepage: www.carellgroup.de
[#] these authors contributed equally.
Supporting information for this article is given via a link at the end of the document.

10.1002/anie.202000414

The fluorinated nucleoside F-fdC **2** was recently introduced by us as a deformylation probe.[17] **2** is an antimetabolite that is effectively incorporated into the genomes of growing cells. The 2'-Fluoro group is required to block all types of glycosylases, so that base excision repair is efficiently inhibited. This ensures high levels of F-fdC **2** in the genome, as required to observe potential deformylation processes.

The synthesis of the novel nucleoside a-fdC **3** is depicted in Scheme 1. The synthesis was started with bromo pyruvic acid **4**, which we first converted into the semicarbazone **5** followed by the conversion into its acid chloride, subsequent cyclization and hydrolysis to give hydroxymethylated 6-aza-uracil **6**.[20] Vorbrüggen nucleosidation with Hoffers' chlorosugar provided subsequently the nucleoside **7** as a mixture of the α- and β-anomers, which could be separated by recrystallization. Next, we acetyl-protected the hydroxymethylgroup to **8** and then used a standard procedure to convert the U-base **8** into the C-derived base **9** by amination of the 4-triazole intermediate with ammonium hydroxide. This led to the concomitant cleavage of the acetyl-protecting group. Dess-Martin oxidation of **9** to **10** and final deprotection of the toluoyl-groups furnished the 6-aza-5-formyl-deoxycytidine nucleoside (a-fdC) **3** in a good total yield of 22% in respect to **6** (SI).

Compound **3** features a nitrogen atom instead of a carbon atom at C6, which possesses a lone pair that blocks any nucleophilic addition to this position. Compound **3** is consequently a perfect model system to investigate, if such a nucleophilic activation is required for the deformylation as mechanistically postulated (Fig. 1A).



Scheme 1. Synthesis of the probe molecule a-fdC **3**. **a.** semicarbazide·HCl, NaOAc, HOAc, H_2O, 0 °C to r.t., 2.5 h, 49%. **b.** pyridine, $SOCl_2$, 80 °C, 75 min. **c.** H_2O, 110 °C., 17 h, 74% over 2 steps. **d.** TMSCl, HMDS, 135 °C, 75 min, *then* **e.** Hoffer´s chlorosugar, $CHCl_3$, r.t.,17 h, 56% over 2 steps. **f.** Ac_2O, pyridine, r.t., 22 h, 96%. **g.** *1,2,4*-triazole, $POCl_3$, NEt_3, MeCN 0 °C to r.t., 18 h, *then* **h.** NH_4OH, *1,4*-dioxane, 40 °C, 5 h, 84%. **i.** *Dess-Martin* periodinane, CH_2Cl_2, -15 °C to r.t., 1 h, 89%. **j.** NaOMe, MeOH, benzene, r.t., 1.5 h, *then* **k.** *reversed-phase* HPLC, 54%.

The nucleosides **2** and **3** were subsequently added at a concentration of 350 µM to the media of Neuro-2a, RBL-2H3, CHO-K1 cells for 72 hours (see SI). During this time, the nucleosides are converted *in vivo* into the corresponding triphosphates and then incorporated into the genome of the dividing cells. Initial studies, in which we fed the nucleosides individually allowed us to determine that both compounds do not decrease cell viability until a concentration of 400 µM, thus the experiments were conducted below the toxicity level. In addition, we tested **2** and **3** at 350 µM on E14 TDG +/- and -/- mouse embryonic stem cells (mESC) under a three-day priming process with C/R media. This system allowed us to exclude the BER pathway, leading to a detectable and quantifiable accumulation of natural fdC (SI). After three days, the cells were harvested and lysed, the genomic DNA was extracted using an optimised protocol (SI). This was followed by an enzymatic digestion of gDNA to single nucleosides and analysed according to a protocol that we reported recently in detail.[21] The obtained nucleoside mixture containing mostly the canonical nucleosides dA, dC, dG and dT, plus the non-canonical nucleosides mdC, hmdC and fdC, as well as the incorporated molecules F-fdC and a-fdC and their potential follow-up products (F-dC, F-mdC, a-dC, a-mdC). Nucleosides were separated by ultra-HPL-chromatography and characterized by coupling of the UHPLC system to a triple quadrupole mass spectrometer. For exact quantification of the nucleosides by isotope dilution, isotopically labelled standards of F-fdC and of the product F-dC were spiked into the analysis mixture as internal standards (SI). To enable exact quantification, calibration curves using these standards were determined (SI). Quantification was performed in the linear region.

During the analysis, we noted an unusually low amount of detected a-fdC **3** because it showed a broad elution profile with very low intensity (Fig. 2B). All attempts to sharpen the elution profile in order to gain sensitivity failed. NMR analysis of compound **3** showed the reason for broad elution profile (SI). Due to the additional electron withdrawing in-ring nitrogen atom, compound **3** exists partially as its hydrate in aqueous solution (20%, see SI). Although the ease of hydrate formation may foster deformylation, the hydrate/carbonyl equilibrium makes efficient detection of compound **3** basically impossible. In order to circumvent the problem, we started to derivatize a-fdC **3** before the analysis with methoxyamine. Addition of CH_3ONH_2 (150 mM) to the digestion solution provided indeed the methoxyoxime of a-fdC in quantitative yield already after 15 min at 25 °C and pH 10. The naturally present fdC **1** and the probe molecule F-fdC **2**, however, react as well, but unfortunately not quantitatively. To reduce impurities during MS measurements we decided against using a catalyst for oxime formation. We therefore decided to analyse the digested DNA in two batches. The first one contained the digested untreated DNA to quantify all bases other than a-fdC. In the second batch, we treated the digested DNA with methoxyamine for a-fdC quantification. For quantification purposes of derivatized a-fdC we performed an external calibration curve (SI).

10.1002/anie.202000414
WILEY-VCH

RESEARCH ARTICLE

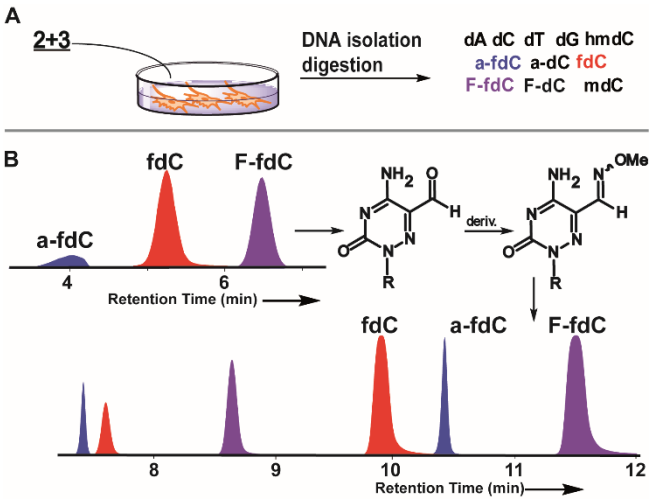


Figure 2. (A) Overview of the experimental steps with the feeding and analysis. **(B)** Analysis scheme and depiction of the reaction of a-fdC with methoxyamine to block hydrate formation and of a typical UHPL-chromatogram before (C-8 column) and after derivatization (C-18 column) for exact quantification. Peak splitting is due isomerization (blue peaks: a-fdC, red peaks: fdC and purple peaks: F-fdC).

With this method in hand, we next quantified all nucleosides present in the genome of the cells treated with a mixture of (**2+3**). Fig. 3A shows that we indeed detected the fluorinated F-dC compound **2,** confirming very efficient deformylation activities. We tested different cell types and found different levels of deformylation activities. But in all cases, the conversion of F-fdC into F-dC was clearly detectable. Most interesting is that we observed the largest deformylation activites in cells associated with neuronal properties. It goes in line with neurons featuring the highest levels of hmdC and fdC. In contrast, Fig. 3B shows for the a-fdC compound **3** that we were unable to detect any formation of the deformylated compound a-dC despite the high propensity of **3** to exist in the hydrated form which is one prerequisite for efficient C-C bond cleavage. This result suggests that the ability to react with a nucleophile at position C6 is also *in vivo* required for efficient deformylation.

In order to substantiate this result, we next performed *in vitro* studies with bisulfite. Bisulfite is a strong nucleophile that was reported to cause deformylation of fdC by first attacking the C6 position, followed by conversion of the C5-C6 saturated fdC adduct into the bisulfite adduct that then undergoes deformylation.[22] The deformylated product dC is then further converted into dU by the well-known bisulfite induced deamination reaction of dC (see SI). Indeed, when we reacted fdC with bisulfite, we observed efficient deformylation and deamination to dU. We then studied to which extent the reaction is influenced by the C2'-F atom present in F-fdC, in order to estimate if the *in vivo* deformylation could be just the result of the present 2'-F atom. Treatment of F-fdC with bisulfite also led to deformylation and deamination to F-dU and indeed the reaction is a little faster compared to fdC (see Fig. 3C). Although the difference is measurable, it is in total rather small. With these data in hand, we can conclude that we may overestimate the amount of deformylation that can occur with fdC lacking the 2'-F atom. We can certainly exclude that deformylation *in vivo* is occurring only with F-fdC. It is unfortunate that we are unable to measure the direct deformylation of fdC because of the presence of efficient BER processes. A TDG -/- cell line showed a huge increase of fdC compared to the TDG +/-, whereas a-fdC and F-fdC stay constant showing that these compounds are indeed not repaired by the TDG protein (see SI). The bisulfite studies however show that the F-fdC compound is not a perfect but sufficient reporter of this C-C bond cleavage reaction. Treatment of the a-fdC compound **3** with bisulfite provided under no circumstances the deformylated product a-dC, showing that the inability to react with a nucleophile at position C6 totally blocks the C-C bond cleavage reaction. We can therefore conclude that the deformylation of fdC during active demethylation requires oxidation of mdC to fdC. fdC can undergo a direct C-C bond cleavage to dC, but this reaction requires a helper nucleophile to attack the C6-position, which is blocked in the case of a-fdC by the lone pair of the C6 carbon to nitrogen exchange. While the chemistry that allows the transformation of fdC to dC is now elucidated, we next need to find the nucleophiles that perform the reaction *in vivo*.

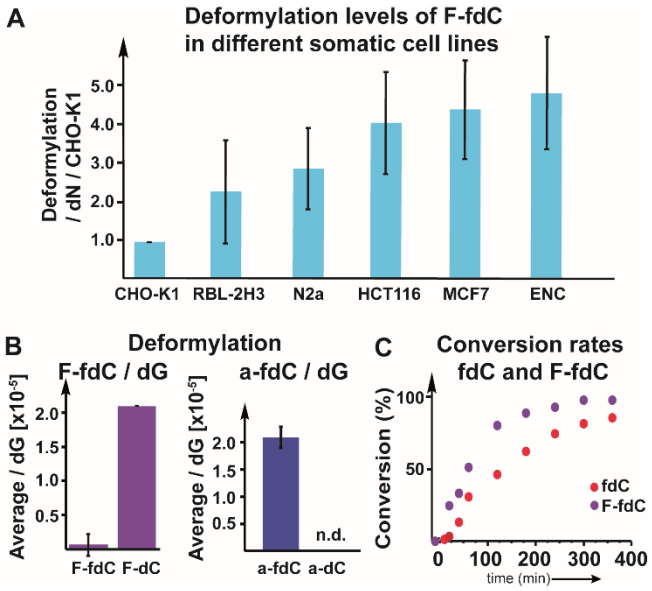


Figure 3. (A) Depiction of the deformylation data of F-fdC in different cell types showing that F-fdC is deformylated in very different cells. Deformylation rate was calculated by the F-dC + F-mdC/dN per F-fdC/dN, then the values were normalized to the cell line with the lowest deformylation level (CHO-K1=1). **(B)** Depiction of the deformylation of F-fdC/dG and a-fdC/dG showing the induced differences within the C6 carbon to nitrogen exchange. **(C)** Depiction of the bisulfite data showing that the deformylation of fdC and F-fdC is comparable proving that the 2'-F substitution has only a small accelerating effect. Whereas the reaction of a-fdC could not be detected.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) GRK2338 (Project ID 321812289), SFB1309 (PID 325871075), SFB1361 (PID 393547839) and SPP1784 (PID 255344185). This project has received additional

funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866).

Keywords: epigenetics • formylcytidine • deformylation • C-C bond cleavage • active demethylation

[1] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511-12514.

[2] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**, *41*, 6421-6429.

[3] 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.

[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.

[5] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555-557.

[6] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.

[7] D. Q. Shi, I. Ali, J. Tang, W. C. Yang, *Front.Genet.* **2017**, *8*, 100.

[8] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50* 6460-6468.

[9] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. Van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159.

[10] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303.

[11] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. DalBello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674-683.

[12] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797-11800.

[13] T. Fu, L. Liu, Q. L. Yang, Y. Wang, P. Xu, L. Zhang, S. Liu, Q. Dai, Q. Ji, G. L. Xu, C. He, C. Luo, L. Zhang, *Chem. Sci.* **2019**, *10*, 7407-7417.

[14] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.

[15] 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.

[16] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516 - 6520.

[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] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274-293.

[19] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2916-2921.

[20] I. V. Alekseeva, A. S. Shalamai, V. S. Shalamai, V. P. Chernetski., *Ukr. Khim. Zh.* **1976**, *42*, 398-401.

[21] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14,* 283 - 312.

[22] E. Kriukiene, Z. Liutkevičiute, S. Klimašauskas, *Chem. Soc. Rev.* **2012**, 41, 6916-6930

10.1002/anie.202000414

WILEY-VCH

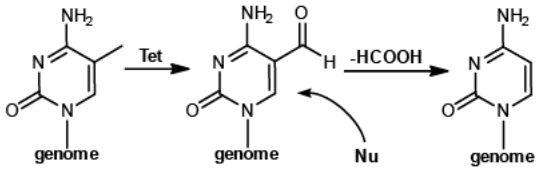
RESEARCH ARTICLE

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

In the course of active demethylation in stem cells, mdC is oxidized to fdC which is deformylated potentially after addition of a nucleophile at the C6-position.



*Alexander Schön, Ewelina Kaminska, Florian Schelter, Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, Thomas Carell**

Page No. – Page No.

Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells