Check for updates

Epigenetics

International Edition: DOI: 10.1002/anie.202000414
German Edition: DOI: 10.1002/ange.202000414

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**

Dedicated to Dr. Klaus Römer on the occasion of his 80th birthday

Abstract: *The removal of 5-methyl-deoxycytidine (mdC) from promoter elements is associated with reactivation of the silenced corresponding genes. It takes place through an active demethylation process involving 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 next step can occur through the action of a glycosylase (TDG), which cleaves fdC out of the genome for replacement by dC. A second pathway is proposed to involve C–C bond cleavage that converts fdC directly into dC. A 6-aza-5-formyl-deoxycytidine (a-fdC) probe molecule was synthesized and fed to various somatic cell lines and induced mouse embryonic stem cells, together with a 2′-fluorinated fdC analogue (F-fdC). While deformylation of F-fdC was clearly observed in vivo, it did not occur with a-fdC, thus suggesting that the C–C bond-cleaving deformylation is initiated by nucleophilic activation.*

The nucleobase modification 5-formyl-deoxycytidine (fdC, **1**) is found in stem cells during early development and in the brain.[1–5] These tissues are particularly rich in 5-hydroxymethyl-deoxycytidine (hmdC) from which fdC (**1**) is produced.[6,7] The 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 intermediates that can cause genome instability, it was suggested early on 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) are indeed possible.[15] Nevertheless, it requires activation of the nucleobase by a nucleophilic addition to the C6 position. For fdC, an additional hydrate formation on the formyl group seems to be necessary, as depicted in Figure 1 A. Although activation with a helper nucleophile is well known as the central mechanistic process during the methylation of dC to mdC by DNA methyltransferases (Dnmts),[18,19] it remains to be confirmed whether such activation occurs in vivo as well.

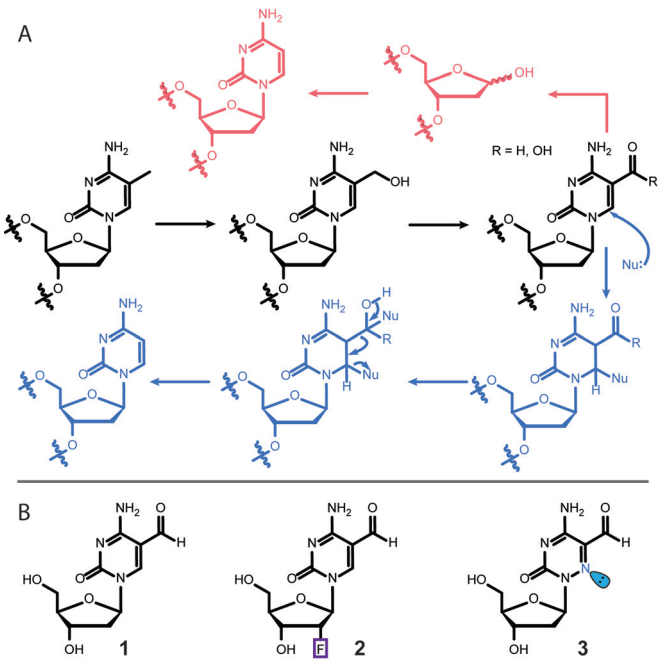


Figure 1. A) The mdC removal pathways that involve oxidation to hmdC, fdC, and cadC followed by either base-excision repair (magenta) or C–C bond cleavage (blue). B) Structures of fdC (**1**) and the two probe molecules **2** and **3** used for this study.

[*] M. Sc. A. Schön,[+] E. Kaminska,[+] M. Sc. F. Schelter,[+] M. Sc. E. Ponkkonen, M. Sc. E. Korytiaková, Dr. S. Schiffers, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians Universität München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de
Homepage: https://www.carellgroup.de

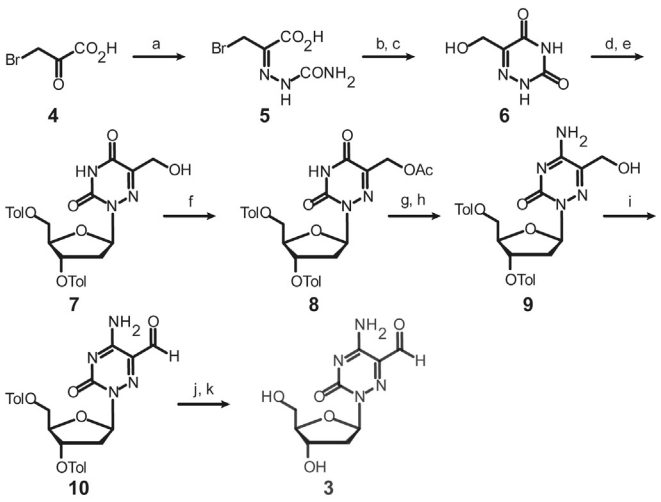
[+] These authors contributed equally to this work.

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.202000414.

In this work, we investigated this hypothesis with two probe molecules, 2′-fluorinated-fdC (F-fdC, **2**) and 6-aza-fdC (a-fdC, **3**). The two compounds were simultaneously fed to different cell types, including primed stem cells. This led to random incorporation of these bases at the "C" sites in 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 the two nucleobases is the presence of an in-ring nitrogen atom (6-aza atom), which features a lone pair that prohibits nucleophilic addition. These results thus provide strong evidence that nucleophilic activation is the central governing mechanistic event that is required for C–C bond cleavage in vivo.

The fluorinated nucleoside F-fdC (**2**) was recently introduced by us as a deformylation probe.[17] Compound **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 conversion into the acid chloride, subsequent cyclization, and hydrolysis to give hydroxymethylated 6-aza-uracil (**6**).[20] Vorbrüggen nucleosidation with Hoffers' chlorosugar subsequently provided the nucleoside **7** as a mixture of the α- and β-anomers, which could be separated by recrystallization. Next, we acetyl-protected the hydroxymethyl group to give **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 removal of the toluoyl groups furnished the 6-aza-5-formyl-deoxycytidine nucleoside (a-fdC) **3** in a good total yield of 22% with respect to **6** (Supporting Information).



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%.

Compound **3** features a nitrogen atom instead of a carbon atom at the 6-position, which possesses a lone pair that blocks any nucleophilic addition to this position. Compound **3** is consequently a perfect model system to investigate whether such a nucleophilic activation is required for the deformylation, as mechanistically postulated (Figure 1A).

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 the Supporting Information). 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 neither compound decreases cell viability up to 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 (mESCs) 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 (see the Supporting Information). After three days, the cells were harvested and lysed, and the genomic DNA was extracted using an optimised protocol (see the Supporting Information). This was followed by an enzymatic digestion of gDNA to single nucleosides and analysed according to a method 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 downstream 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 (see the Supporting Information). To enable exact quantification, calibration curves using these standards were determined (see the Supporting Information). Quantification was performed in the linear region.

During the analysis, we noted that an unusually low amount of a-fdC (**3**) was detected because it showed a broad elution profile with very low intensity (Figure 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 (see the Supporting Information). Due to the additional electron-withdrawing in-ring nitrogen atom, compound **3** exists partially as its hydrate in aqueous solution (20%, see the Supporting Information).

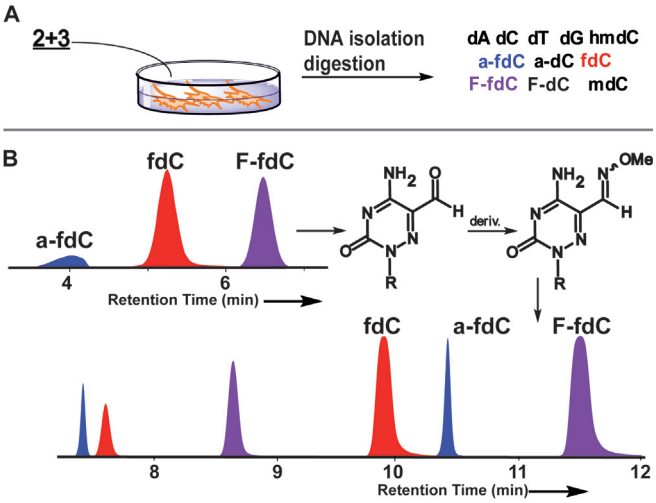


Figure 2. A) Overview of the experimental steps with the feeding and analysis. B) Analysis scheme and the reaction of a-fdC with methoxyamine to block hydrate formation and of a typical UHPL-chromatogram before (C-8 column) and after derivatization (C-18 column) for exact quantification. Peak splitting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

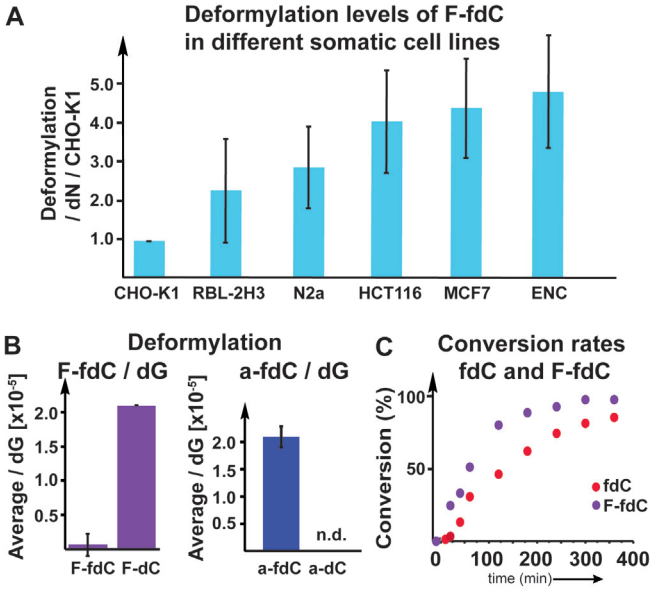


Figure 3. A) Deformylation data for F-fdC in different cell types, showing that F-fdC is deformylated in very different cells. Deformylation rate was calculated by the F-dC + F-mdC/dN per F-fdC/dN, then the values were normalized to the cell line with the lowest deformylation level (CHO-K1 = 1). B) The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

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 analysis with methoxyamine. Addition of CH_3ONH_2 (150 mM) to the digestion solution indeed provided the methoxyoxime of a-fdC in quantitative yield after just 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 of the derivatized a-fdC, we constructed an external calibration curve (see the Supporting Information).

With this method in hand, we next quantified all nucleosides present in the genome of the cells treated with a mixture of **2** + **3**.

Figure 3 A shows that we indeed detected the fluorinated F-dC (**2**), thus confirming very efficient deformylation activity. We tested different cell types and found different levels of deformylation activity. But in all cases, the conversion of F-fdC into F-dC was clearly detectable. Most interesting is that we observed the highest deformylation activity in cells associated with neuronal properties. This is in line with neurons featuring the highest levels of hmdC and fdC. In contrast, Figure 3 B shows that for a-fdC (**3**), 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 the 6-position is also required in vivo for efficient deformylation.

In order to substantiate this result, we next performed in vitro studies with bisulfite. Bisulfite is a strong nucleophile that has been 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, which 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 the Supporting Information). Indeed, when we reacted fdC with bisulfite, we observed efficient deformylation and deamination to dU. We then studied to what extent the reaction is influenced by the 2′-F atom present in F-fdC, in order to estimate whether the in vivo deformylation could be just the result of the 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 Figure 3 C). 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 occurs 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 in fdC compared to the TDG +/−, whereas a-fdC and F-fdC stay constant, thus showing that these compounds are indeed not repaired by the

TDG protein (see the Supporting Information). The bisulfite studies, however, show that the F-fdC compound is not a perfect but sufficient reporter of this C–C bond cleavage. Treatment of a-fdC (**3**) with bisulfite did not provide the deformylated product a-dC under any circumstances, showing that the inability to react with a nucleophile at the 6-position totally blocks the C–C bond cleavage. 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 introduced by the C6-carbon-to-nitrogen exchange. While the chemistry that allows the transformation of fdC into dC is now elucidated, we next need to find the nucleophiles that perform the reaction in vivo.

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).

Conflict of interest

The authors declare no conflict of interest.

Keywords: demethylation · DNA modifications · epigenetics · formylcytidine

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 5591 – 5594
Angew. Chem. **2020**, *132*, 5639 – 5643

[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; *Angew. Chem.* **2015**, *127*, 12691 – 12695.

[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; *Angew. Chem.* **2011**, *123*, 7146 – 7150.

[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; *Angew. Chem.* **2011**, *123*, 6588 – 6596.

[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; *Angew. Chem.* **2016**, *128*, 11974 – 11978.

[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; *Angew. Chem.* **2012**, *124*, 6622 – 6626.

[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. (Ukr. Ed.)* **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. Kriukienė, Z. Liutkevičiūtėa, S. Klimašauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916 – 6930.

Manuscript received: January 9, 2020
Accepted manuscript online: January 30, 2020
Version of record online: February 25, 2020

