Influencing epigenetic information with a hydrolytically stable carbocyclic 5-aza-2'-deoxycytidine

Thomas M. Wildenhof,[a] Sarah Schiffers,[a] Franziska R. Traube,[a] Peter Mayer[a] and Thomas Carell*[a]

[a] Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-13, Munich, Germany
* address correspondence to: Thomas.carell@lmu.de

This article was first published online on 17.06.2019 in Angewandte Chemie International Edition under DOI 10.1002/anie.201904794

Dedicated to Prof. J. Rebek, Jr. on the occasion of his 75th birthday

Abstract: *5-Aza-2'-deoxycytidine (AzadC) is an antimetabolite in clinical use, which reduces the level of the epigenetic modification 5-methyl-2'-deoxycytidine (mdC). AzadC is incorporated into the genome of proliferating cells, where it inhibits the DNA methyltransferases (DNMTs) in a suicide process leading to a reduction of mdC. The loss of mdC, which is a transcriptional silencer in promoters, leads to the reactivation of genes including tumor suppressor genes, which elicits a beneficial effect. The problem associated with AzadC is that the compound is hydrolytically unstable. It decomposes during treatment to a variety of poorly characterized hydrolysis products. After its incorporation into the genome, this hydrolytic instability generates abasic sites. It is consequently difficult to dissect if the activity of the compound is caused by DNMT inhibition or more generally by DNA lesion formation. We now discovered that a disarmed version of AzadC, in which the ribose oxygen was replaced by a CH_2-group, is surprisingly stable under a variety of pH values while keeping the epigenetic activity against the DNMTs.*

5-Aza-2'-deoxycytidine (decitabine, AzadC) is a nucleoside analogue that is able to manipulate epigenetic information.[1-5] Epigenetic information in DNA is associated with the formation of 5-methyl-2'-deoxycytidine (mdC) from 2'-deoxycytidine (dC) with the help of DNA methyltransferases (DNMTs) and *S*-adenosylmethionine (SAM) as the methylating cofactor.[6-7, 4] Methylation of dC to mdC in promoter regions is typically associated with transcriptional silencing of genes.[8-9] AzadC is a prodrug that is inside cells converted into the corresponding active triphosphate and subsequently incorporated into the genome during cell division. The mode of action of AzadC involves reaction of its electrophilic C6 positions with a DNMT active site thiol nucleophile (Fig. 1a).[10-11] This generates a covalent intermediate that is methylated by the SAM cofactor as depicted in Fig 1a. Due to the N-atom at position 5 of the triazine heterocycle, the final β-elimination reaction, which would usually release mdC from the DNMT enzyme, is not possible anymore. The consequence is the formation of a covalent DNA-DNMT crosslink. As a result of administering AzadC, a large drop of the mdC levels (hypomethylating effect) is observed, which leads to the reactivation of silenced tumor suppressor genes in cancer cells.[1] This epigenetic effect is hoped to re-differentiate cancer cells back into normally proliferating cells. AzadC is currently in use as one of the first pharmaceuticals that operates at the epigenetic level for the treatment of myelodysplastic syndromes (MDS)[2] and for acute myeloid leukemia (AML)[4]. Clinically, it is administered in several cycles, with each cycle involving one week of treatment and three weeks of pausing.

The problem associated with AzadC is that the compound hydrolyses in aqueous solution following the path depicted in Fig. 1b. This hydrolysis compromises the activity of AzadC, particularly over the long treatment times. In order to circumvent this problem, it is necessary to generate an AzadC compound that can demethylate (and hence react with an S-nucleophile), while hydrolysis (reaction with an O-nucleophile) should be blocked. Such a compound may allow to dissect how demethylation and lesion

formation contribute to the anti-cancer activity, which is an information needed for the design of new epigenetically acting antimetabolites.

Here, we report that replacing the oxygen of the ribose by a CH_2-group has a surprisingly large remote effect on the reactivity of the heterocycle. The created carbocyclic version of AzadC (cAzadC, **1**) still inhibits DNMTs but is hydrolytically stable (Fig. 1c).

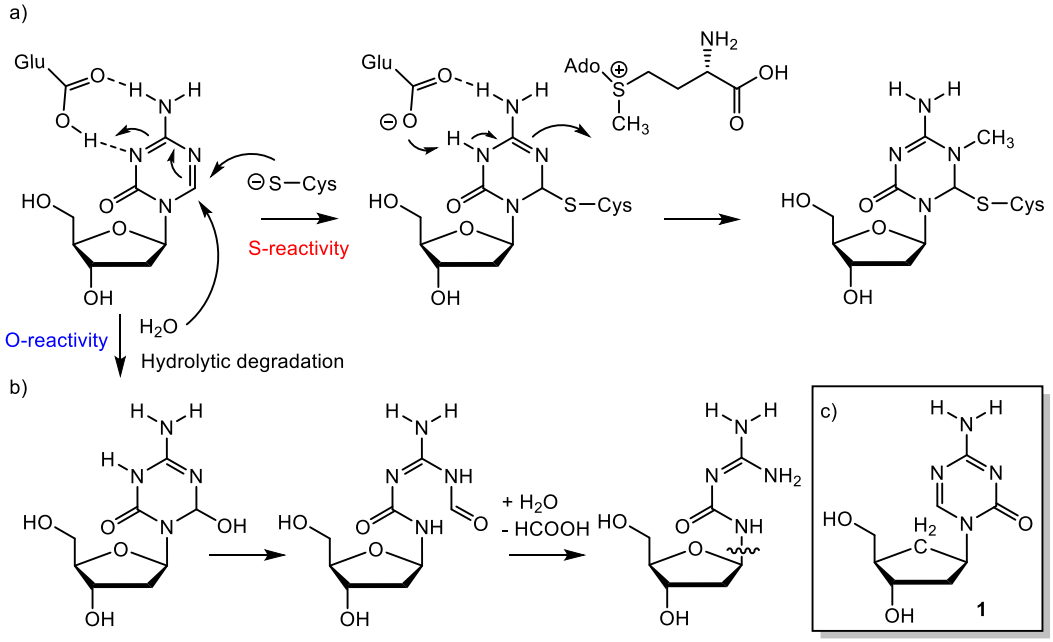
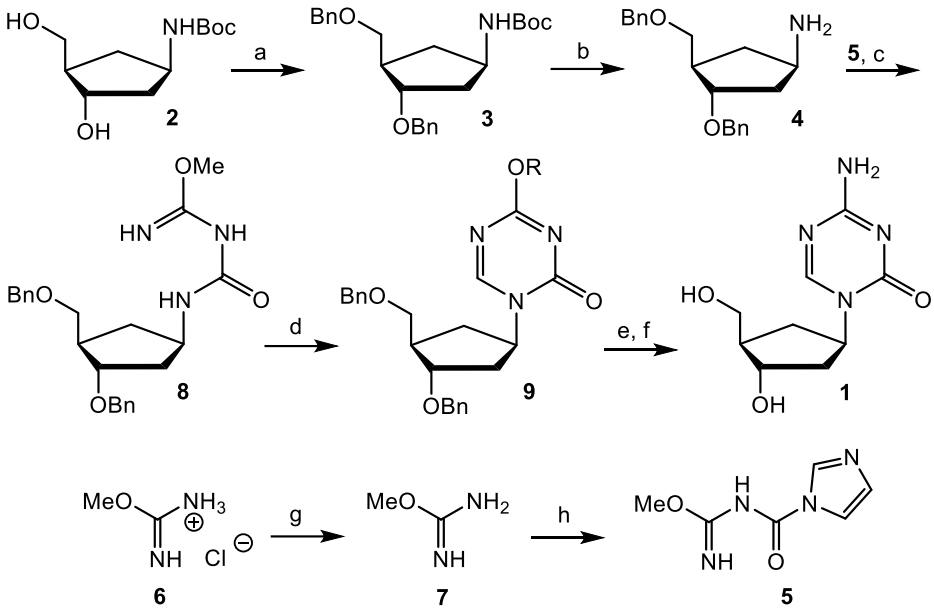


Figure 1. Depiction of 5-aza-2'-deoxycytidine (decitabine, AzadC) together with its mode of action. a) Active site thiol reacts with the C6-position of AzadC. b) Hydrolytic degradation pathway that goes in hand with reaction of a water molecule with the C6-position (O-reactivity) of AzadC. This leads to a final base loss and formation of an abasic site. c) Boxed, depiction of the carbocyclic version cAzadC **1**.

The synthesis of cAzadC **1** is depicted in Scheme 1. It starts with the Boc-protected aminocyclopentane derivative **2** that we used previously to synthesize DNA lesion analogues.[12-15] Compound **2** was first benzyl-protected to **3**, Boc-deprotected to **4**, and then reacted with carbimidazole **5**, which was prepared in two steps from isomethylurea **6** after generation of the free base **7** with potassium hydroxide and reaction of **7** with carbonyldiimidazole. This provides the carbamoylurea-cyclopentane nucleoside analogue **8**. Cyclization to the triazine base **9** was subsequently performed with triethylorthoformate. Reaction of **9** with NH_3 in methanol and deprotection of the benzyl groups with BCl_3 in dichloromethane furnished the final compound cAzadC **1** as the free nucleoside.



Scheme 1. Synthesis of the carbocyclic 5-aza-2'-deoxycytidine (cAzadC, **1**). a) NaH, BnBr, DMF, 0 °C, 1.5 h and stirred for additional 2 h at r.t.; b) TFA (30%), CH_2Cl_2 then Na_2CO_3, 10 min r.t.; c) CH_3CN, reflux, 2 h d) $HC(OEt)_3$, TFA cat., reflux, 3 h; e) NH_3 (7 N, MeOH), 3 h, r.t., then H_2O ;f) CH_2Cl_2, -78 °C, BCl_3, 1 h, then → r.t., 2 h, MeOH, 20 min. g) KOH, Et_2O:H_2O (39:1), -15 °C, 30 min, h) carbonyldiimidazole, THF, r.t., 3 h; R = Me or Et.

Recrystallization of compound cAzadC **1** from hot methanol gave colourless needles, which allowed us to solve the crystal structure that is depicted in Fig. 2. Interesting is the observation that cAzadC **1** exists with two different cyclopentane conformations in the crystal (Fig. 2; Fig.SI 1). One conformer adopts a C6'-endo (P = 88.2 °, ν_{max} = 47.8 °) conformation (Fig. 2a), while the second exists as the C2'-endo-C3'-exo (South, P = 150.8 °, ν_{max} = 45.4 °) conformer (Fig. 2b). The latter conformation is typical for 2'-deoxynucleosides in DNA. This shows that the cAzadC **1** nucleoside can adopt the correct DNA-type conformation, fueling hope that the analogue has the potential to get phosphorylated and integrated into the genome.

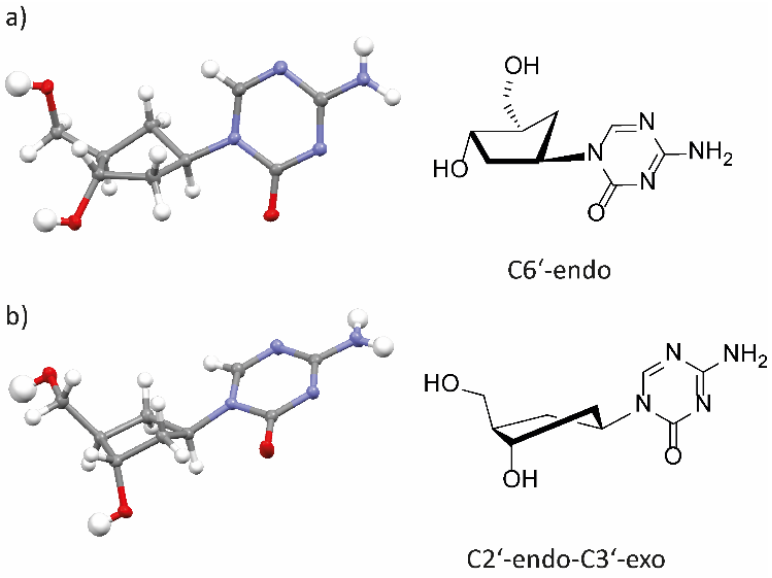


Figure 2. Crystal structure of carbocyclic 5-aza-2'-deoxycytidine (cAzadC **1)** showing the molecule in the observed C6'-endo conformation (a) and the C2'-endo-C3'-exo conformation ($^2T_1$) (b).

We next investigated the stability of cAzadC **1** in direct comparison to the pharmaceutical AzadC (Fig. 3). Since one treatment cycle goes over four weeks we decided to measure the stability at a time point related to a half cycle (14 d). We dissolved AzadC and cAzadC **1** at a concentration of 100 mM in a phosphate buffer (100 mM) at three different pH values (7.4, 5.5 and 8.5) and measured NMR spectra after keeping the solutions at r.t. Since tumour cells often provide an acidic micro-environment,[16] the stability under slightly acidic pH is particularly informative. As evident from the data shown in Fig. 3, the pharmaceutical AzadC strongly degraded within these 14 d. Importantly, at pH = 5.5 and at pH = 8.5, intact AzadC was only hardly detectable anymore. At physiological pH (7.4), AzadC was still present after 14 d but the level of degradation is dramatic. In contrast to these results, we observed for cAzadC **1** surprisingly no degradation at all tested pH values, including pH = 5.5. This result led to the surprising discovery that the simple O → CH_2 exchange causes a strong remote disarming effect that seems to change the properties of the triazine ring so that reaction with water is stopped.

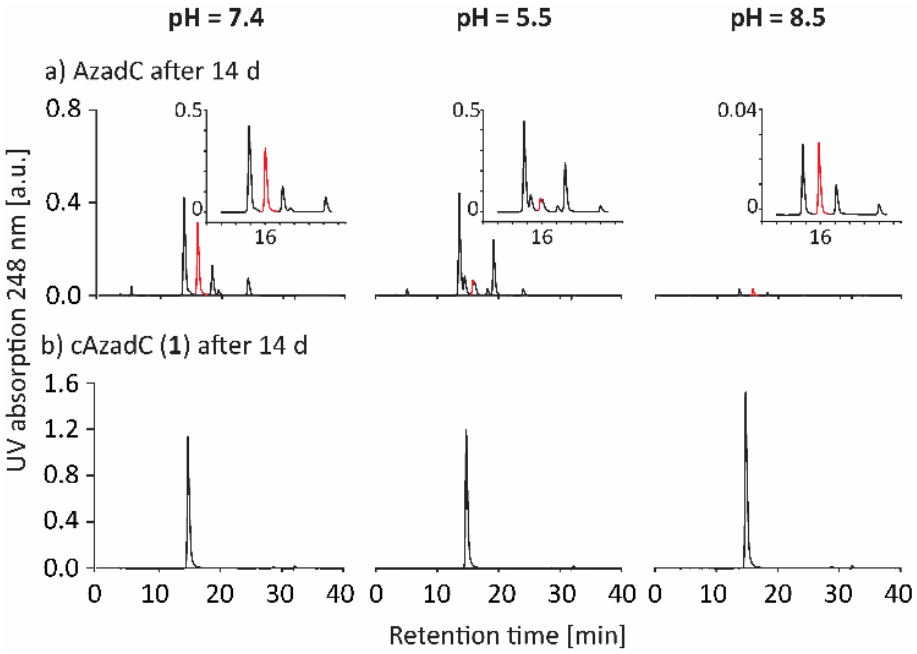


Figure 3. HPLC-based stability measurements showing (a) severe hydrolytic decomposition of 5-aza-2'-deoxycytidine (AzadC) solutions at different pH values, while (b) the carbocyclic compound cAzadC **1** was stable at all three pH values. The inset table in (a) shows the chromatogram between t1 = 10 min and t2 = 20 min for AzadC. The AzadC signal is depicted in red.

We next investigated if this disarming effect would influence the biological functions. We used for this purpose mouse embryonic stem cells (mESC) that were primed in serum/LIF as a model system, since mdC levels increase from naïve to primed state.[17] We added cAzadC **1** in two different concentrations (1 µM and 5 µM) to mESC that have been primed for 48 h and allowed the cells to further proliferate under priming conditions in the presence of cAzadC **1** for additional 72 h. After the 72 h, we harvested the cells, isolated the DNA and digested the DNA down to the nucleoside level using our described protocol.[18] The levels of mdC were finally precisely quantified using isotope dilution UHPLC-MS^2. To this end, isotopically labelled standards of the nucleosides were spiked in for exact quantification.[19, 18] In addition to mdC, we quantified the levels of 5-hydroxymethyl-2'-deoxycytidine (hmdC), which is formed from mdC by the action of TET enzymes.[20-21] The absolute levels of hmdC are in mESC more than ten times lower than the mdC levels[20, 22]. The consequence is that even after a substantial reduction of mdC, there should be sufficient mdC to keep the hmdC levels constant. The question if and by how much the

hmdC level is affected can therefore inform us about how epigenetic reprogramming is organized. Parallel to the quantification of mdC and hmdC we also quantified to which extent cAzadC **1** itself was incorporated into the genome of the mESC. Detection of AzadC in the genome of treated cells is only possible after treatment of the DNA with $NaBH_4$. Application of $NaBH_4$ reduces the C(5)=C(6) double bond, which stabilizes the compound so that its quantification becomes possible.[23, 19] To our delight, we noted that the stability of cAzadC **1** allowed its quantification without this pre-treatment. We also noted that the applied enzymatic digestion protocol allowed to digest genomic DNA (gDNA) even in the presence of large amounts of cAzadC **1**. Taken together, quantification of cAzadC **1** by UHPLC-MS^2 using an external calibration curve (Fig. SI2) was possible in parallel to quantification of canonical and epigenetic bases.

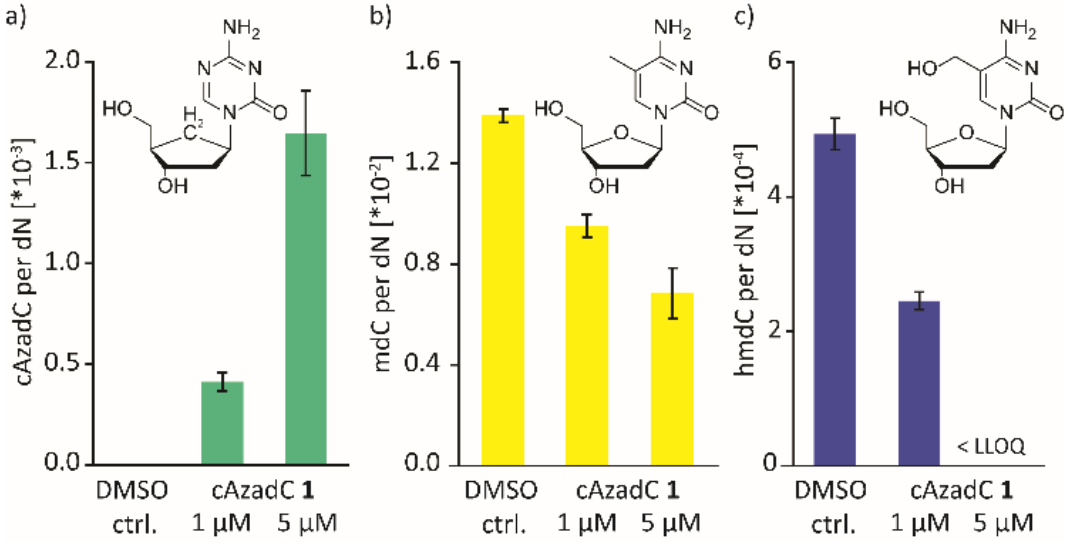


Figure 4. Depiction of the quantification data of DNA modifications of carbocyclic 5-aza-2'-deoxycytidine-treated (cAzadC **1**) mouse embryonic stem cells (mESC) obtained by UHPLC-MS^2. For each condition, three biological replicates were measured in technical triplicates. For each technical replicate, 0.5 µg of DNA were digested. Bar graphs represent mean, error bars represent standard deviation. LLOQ indicates the lower limit of quantification.

At 1 µM cAzadC **1** concentration, we detected a cAzadC **1** level of $5x10^{-4}$ cAzadC per dN (Fig. 4a). This amounts to almost 3 million cAzadC nucleotides integrated into the genome. At the higher concentration of 5 µM cAzadC **1**, the level increased 3-fold to $1.7x10^{-3}$ cAzadC per dN and consequently to more than 8 million integrated cAzadCs per genome. Compared to the incorporation of AzadC, which reaches $1.2x10^{-3}$ AzadC per dN, when applied with 1 µM[19], the levels of cAzadC **1** reaches about a third of this level. The data clearly show that the carbocyclic version of AzadC (cAzadC **1)** is incorporated and that it reaches in the genome finally comparable levels at 5 µM concentration.

Importantly, after exposing the mESC for 72 h at 1 µM cAzadC in the medium, we detected a reduction of the mdC values by almost 30% (Fig. 4b). At 5 µM concentration in the medium, the mdC levels dropped even to about 50% of the original value. A decrease to 50% is observed for AzadC as well. Here, however, the 50%-reduction is reached faster (24 h) and already with lower AzadC concentration (1 µM).[19] The data show that the carbocyclic version cAzadC **1** needs simply more time to affect the mdC levels by the same amount. We believe that this effect is caused by a potentially slower conversion of cAzadC **1** into the triphosphate. The slower kinetics of cAzadC **1**, however, is not necessarily a disadvantage given the long treatment times that are applied in the clinic.

Very interesting is also the discovery that the hmdC levels were reduced to about 50% already in the 1 µM experiment. At 5 µM, we were even unable to detect hmdC above background levels using 0.5 µg

of genomic DNA. The result shows that the hmdC level dropped even faster than the mdC levels, although hmdC is more than ten times less abundant in the genome. This result is interesting. It indicates that hmdC might be potentially predominantly generated in the mdC maintenance process during cell division. We see here that compound cAzadC **1** is a perfect tool molecule that now allows to gain further insight into the interplay between methylation of dC to mdC and oxidation of mdC to hmdC. With the new compound cAzadC **1** in hand we can now begin to clearly correlate demethylation of the genome with the corresponding cellular effects without compromising DNA damaging effects. Finally, cAzadC **1** may not only be a valuable tool compound but potentially even a next generation epigenetic pharmaceutical. In summary, we show that the replacement of the in-ring O-atom by a CH_2-unit stabilizes the pharmaceutical so that its nucleophilic reaction with water is stopped. The new nucleoside cAzadC **1** is accepted by the phosphorylating enzymes in cells and the corresponding cAzadC-triphosphates are efficiently incorporated into the genome. cAzadC **1** is incorporated in the genome with several million nucleotides and it causes the mdC level to decrease to 70% relative to the control levels.

Experimental Section

Synthesis
All synthetic procedures are described in detail in the supplementary material.

Cell culture of mESC for cAzadC treatment
Feeder independent wt J1 (strain 129S4/SvJae)[24] cells were cultured in the presence of serum and LIF as previously described[25]. They were routinely maintained on gelatinized plates in 2i/L medium. For priming experiments, 2i cultures were passaged when applicable in DMEM supplemented with FBS and LIF as above but lacking the inhibitors. For drug treatment, cells were moved into the primed state by removing 2i from the medium. Cells were incubated 2 d in DMEM supplemented with FBS and LIF in 6-well plates (*VWR*). After splitting, $2x10^5$ cells were transferred into a 6-well plate culture dish and supplemented with either 1 µM (in 0.01% DMSO) or 5 µM cAzadC (in 0.05% DMSO) and treated for 72 h. After removal of the medium and washing the cells with DPBS, they were directly lysed with RLT^+ buffer as described in a previous publication[18]

gDNA isolation, total enzymatic digest and UHPLC-MS^2
The gDNA was isolated as described previously[18]. Due to the higher stability of cAzadC **1**, a hydrogenation procedure was not necessary and the gDNA was directly subjected to a total enzymatic digest and analyzed using UHPLC-MS^2 as described in a previous publication.[19]

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via the programs SFB1309 (TP-A4), SFB1361, SPP1784, and GRK2338/1 (P12). 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). F.R.T. thanks the Boehringer Ingelheim Fonds for a PhD fellowship.

Keywords: epigenetics • 5-methyl-2'-deoxycytidine • DNA methylation • antimetabolite • decitabine

References

[1] M. Daskalakis, T. T. Nguyen, C. Nguyen, P. Guldberg, G. Kohler, P. Wijermans, P. A. Jones, M. Lubbert, *Blood* **2002**, *100*, 2957-2964.

[2] H. Kantarjian, J. P. Issa, C. S. Rosenfeld, J. M. Bennett, M. Albitar, J. DiPersio, V. Klimek, J. Slack, C. de Castro, F. Ravandi, R. Helmer, 3rd, L. Shen, S. D. Nimer, R. Leavitt, A. Raza, H. Saba, *Cancer* **2006**, *106*, 1794-1803.

[3] C. Stresemann, F. Lyko, *Int. J. Cancer* **2008**, *123*, 8-13.

[4] Y. Koh, Y. A. Kim, K. Kim, J.-A. Sim, S.-S. Yoon, S. M. Park, Y. H. Yun, *Blood* **2016**, *128*, 2381-2381.

[5] M. Nieto, P. Demolis, E. Béhanzin, A. Moreau, I. Hudson, B. Flores, H. Stemplewski, T. Salmonson, C. Gisselbrecht, D. Bowen, F. Pignatti, *Oncologist* **2016**, *21*, 692-700.

[6] S. S. Smith, B. E. Kaplan, L. C. Sowers, E. M. Newman, *Proc. Natl. Acad. Sci.* **1992**, *89*, 4744-4748.

[7] G. G. Wilson, R. J. Roberts, S. Kumar, J. Posfai, M. Sha, S. Klimasauskas, X. Cheng, *Nucleic Acids Res.* **1994**, *22*, 1-10.

[8] A. Bird, *Genes Dev.* **2002**, *16*, 6-21.

[9] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89-97.

[10] R. Jüttermann, E. Li, R. Jaenisch, *Proc. Natl. Acad. Sci.* **1994**, *91*, 11797-11801.

[11] S. Gabbara, A. S. Bhagwat, *Biochem. J.* **1995**, *307*, 87-92.

[12] M. Ober, H. Müller, C. Pieck, J. Gierlich, T. Carell, *J. Am. Chem. Soc.* **2005**, *127*, 18143-18149.

[13] H. Müller, T. Carell, *Eur. J. Org. Chem.* **2007**, *2007*, 1438-1445.

[14] F. Büsch, J. C. Pieck, M. Ober, J. Gierlich, G. W. Hsu, L. S. Beese, T. Carell, *Chem. Eur. J.* **2008**, *14*, 2125-2132.

[15] T. H. Gehrke, U. Lischke, K. L. Gasteiger, S. Schneider, S. Arnold, H. C. Müller, D. S. Stephenson, H. Zipse, T. Carell, *Nat. Chem. Bio.* **2013**, *9*, 455.

[16] Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, Y. Baba, *Cancer Cell Int.* **2013**, *13*, 89.

[17] S. Takahashi, S. Kobayashi, I. Hiratani, *Cell. Mol. Life Sci.* **2018**, *75*, 1191-1203.

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

[19] S. Schiffers, T. M. Wildenhof, K. Iwan, M. Stadlmeier, M. Müller, T. Carell, *Helv. Chim. Acta* **2019**, *102*, e1800229.

[20] 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.

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

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

[23] A. Unnikrishnan, A. N. Q. Vo, R. Pickford, M. J. Raftery, A. C. Nunez, A. Verma, L. B. Hesson, J. E. Pimanda, *Leukemia* **2018**, *32*, 900-910.

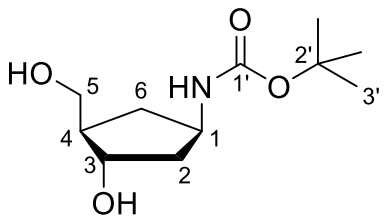
[24] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.

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

Supplementary Material

Synthetic procedures

***tert*-Butyl-N-[(1*R*,3*S*,4*R*)-3-hydroxy-4-(hydroxymethyl)cyclopentyl]-carbamate (2)**



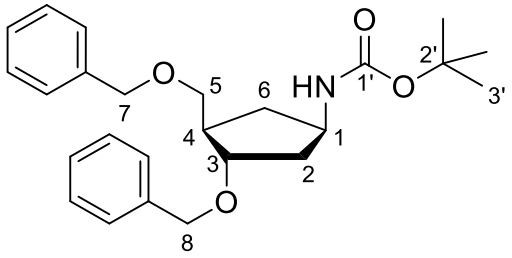
2

$C_{11}H_{21}NO_4$

231.29 g/mol

2 was synthesized from *tert*-Butyl-*N*-((1*S*,2*R*,4*R*,5*R*)-4-((*tert*-butyldimethylsilyloxy)methyl)-6-oxa-bicyclo-[3.1.0]-hex-2-yl)-carbamate as described previously.[1]

***tert*-Butyl-N-[(1*R*,3*S*,4*R*)-3-benzyloxy-4-(benzyloxymethyl)-cyclopentyl]-carbamate (3)**



3

$C_{25}H_{23}NO_4$

411.54 g/mol

Boc-protected **2** (3.107 g, 13.43 mmol, 1.0 eq.) was dissolved in dry DMF (75 mL). The solution was cooled to 0 °C and NaH (60% dispersion in mineral oil, 1.182 g, 29.55 mmol, 2.2 eq.) was added in three portions. After 15 min at 0 °C, benzyl bromide (4.00 mL, 33.58 mmol, 2.5 eq.) was added dropwise. The reaction mixture was stirred for 1.5 h at 0 °C and for additional 2 h at rt. The reaction mixture was diluted using CH_2Cl_2 (30 mL) and the reaction was quenched with sat. $NaHCO_3$ (30 mL). The mixture was extracted with CH_2Cl_2 (500 mL). The organic phase was washed three times with sat. $NaHCO_3$ (1 × 400 mL, 2 × 100 mL) and dried over Na_2SO_4. The solvent was removed *in vacuo*. The crude product was purified by column

chromatography on silica gel with a stepwise gradient *i*Hex/EtOAc (9:1→7:1→4:1) to afford product **3** as a colorless solid (3.39 g, 8.24 mmol, 61%).

R_f = 0.67 (iHex/EtOAc = 2:1)

Mp = 74 - 76 °C

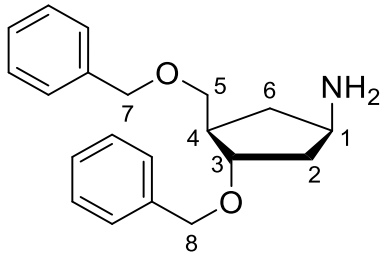
HRMS [ESI+]: $[C_{25}H_{33}NO_4Na]^+$, $[M+Na^+]^+$ calculated.: 434.2292, found: 434.2299

^{1}H-NMR (599 MHz, $CDCl_3$): *δ*/ppm = 7.39 – 7.26 (m, 10H, Ar-H), 4.76 (s, 1H, NH), 4.53 – 4.37 (m, 4H, C_7-H and C_8-H), 4.18 – 4.08 (m, 1H, C_1-H), 3.96 – 3.87 (m, 1H, C_3-H), 3.53 – 3.41 (m, 2H, C_5-H), 2.36 – 2.25 (m, 2H, C_6-H_a, C_4-H), 2.11 – 2.05 (m, 1H, C_2-H_a), 1.77 – 1.68 (m, 1H, C_2-H_b), 1.43 (s, 9H, $C_{2'}$-H), 1.29 – 1.20 (m, 1H, C_6-H_b).

^{13}C-NMR (151 MHz, $CDCl_3$): *δ*/ppm = 155.5 ($C_{1'}$), 138.60, 138.23, 128.36, 128.28, 127.59, 127.56, 127.42 (C-Ar), 81.0 (C_3), 79.2 ($C_{2'}$), 73.3 (C_8), 71.9 (C_5), 71.3 (C_7), 50.4 (C_1), 44.9 (C_4), 39.6 (C_2), 35.0 (C_6), 28.6 ($C_{2'}$).

FTIR (ATR): $\tilde{\nu}$/cm^{-1} = 3339 (*w*), 2973 (*w*), 2929 (*w*), 2858 (*w*), 1692 (*m*), 1496 (*m*), 1453 (*m*), 1390 (*w*), 1364 (*m*), 1274 (*m*), 1247 (*m*), 1166 (*s*), 1091 (*s*), 1067 (*s*), 1027 (*m*), 1012 (*m*).

(1R, 3S, 4R)-1-Amino-3-benzyloxy-4-(benzyloxymethyl)cyclopentane (4)



4

$C_{20}H_{25}NO_4$

311,42 g/mol

A solution of Boc-protected amine **3** (3.351 g, 8.14 mmol, 1.0 eq.) and trifluoroacetic acid (12 mL, 30% (v/v)) in dry CH_2Cl_2 (28 mL) was stirred for 1 h at rt. After removing the solvent *in vacuo*, the residue was resuspended in sat. Na_2CO_3 (100 mL) and stirred for 10 min. The mixture was extracted with EtOAc (3 × 150 mL) and the combined organic phases were dried over Na_2SO_4. Removing the solvent *in vacuo* resulted in the free amine **4** as viscous brown oil, which was used without further purification due to its instability towards O_2.

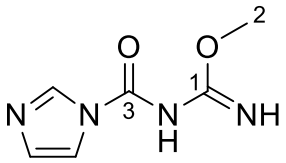
Analytical data of purified **4**:

HRMS [ESI$^+$]: $[C_{23}H_{29}N_4O_4Na]^+$, $[M+Na^+]^+$ calculated: 312.1958, found: 312.1955

^{1}H-NMR (400 MHz, $CDCl_3$): *δ*/ppm = 7.38 – 7.20 (m, 10H, C-Ar), 4.57 – 4.36 (m, 4H, C_7, C_8), 3.95 – 3.85 (m, 1H, C_1), 3.62 – 3.50 (m, 1H, C_3), 3.50 – 3.39 (m, 2H, C_5), 2.41 – 2.28 (m, 1H, C_4), 2.28 – 2.13 (m, 1H, C_6-H_a), 2.13 – 1.96 (m, 3H, NH_2, C_2-H_a), 1.65 – 1.50 (m, 1H, C_2-H_b), 1.19 – 1,.08 (m, 1H, C_6-H_a),

^{13}C-NMR (101 MHz, $CDCl_3$): *δ*/ppm = 138.8, 138.4, 128.5, 128.5, 127.79, 127.75, 127.7, 127.6 (C-Ar), 81.82 (C_3), 73.23 (C_7), 72.58 (C_5), 71.10 (C_8), 51.40 (C_1), 45.63 (C_4), 42.07 (C_2), 38.08 (C_6).

2-Methyl-1-(1-imidazoylcarbonyl)-isourea (5)



5

$C_6H_8N_4O_4$

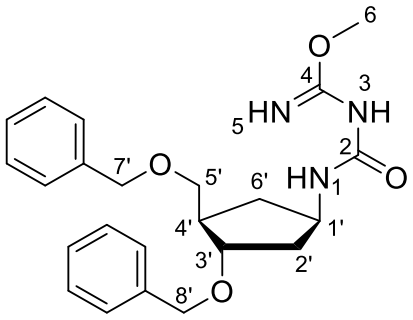
168.16 g/mol

To a solution of *O*-methylisourea hydrochloride **6** (5.00 g, 45.65 mmol, 1.0 eq.) in Et_2O:H_2O (39:1) at – 15 °C, KOH powder (51.22 g, 912.96 mmol, 20.0 eq.) was added portion wise under constant stirring. After 30 min, the mixture was filtrated, and the residue was washed with ice-cold Et_2O (3 × 50 mL). The combined organic phases were concentrated to 20 mL *in vacuo*. Cooling to – 20 °C resulted in precipitation. Vacuum filtration under N_2 flow resulted in *O*-methylisourea **7** as a colorless wax (1.759 g, 23.74 mmol, 52%). A solution of **7** (1.24 g, 16.74 mmol, 1.0 eq.) and carbonyldiimidazole (2.90 g, 17.91 mmol, 1.07 eq.) in dry THF (35 mL) was stirred for 3 h at rt. After 1 h, a colorless precipitate was observed. The reaction mixture was concentrated to 3 mL *in vacuo* and filtrated. The residue was washed with ice-cold THF (2 × 3 mL) and the combined organic phases were lyophilized, which resulted in colorless solid **5** (1.946 g, 11.57 mmol, 69%).

^{1}H-NMR (400 MHz, $CDCl_3$): *δ*/ppm = 8.60 (br s, 1H, N-H), 8.35 (s, 1H, Ar-H), 7.58 (s, 1H, Ar-H), 7.04 (s, 1H, Ar-H), 6.04 (br s, 1H, N-H), 3.96 (s, 3H, C_2-H).

^{13}C-NMR (101 MHz, $CDCl_3$): *δ*/ppm = 165.7 (C_1), 157.3 (C3), 137.5(Ar), 130.0 (Ar), 117.2 (Ar), 55.3 (C_2).

1-[(1'*R*,3'*S*,4'*R*)-3'-Benzyloxy-4'-(benzyloxymethyl)-cyclopentyl]-methylisobiuret (8)



8

$C_{23}H_{29}N_4O_4$

411.50 g/mol

The crude product **4** was dissolved in dry acetonitrile (75 mL) and the carbonyl imidazole **5** (1.506 g, 8.96 mmol, 1.1 eq.) was added. The mixture was refluxed for 2 h. The solvent was removed *in vacuo*. The crude product was purified by column chromatography on silica gel with a stepwise gradient of *i*Hex/EtoAc (2:1→1:1) to obtain the methylisobiurete **8** as a yellow oil (2.613 g, 6.35 mmol, 78%).

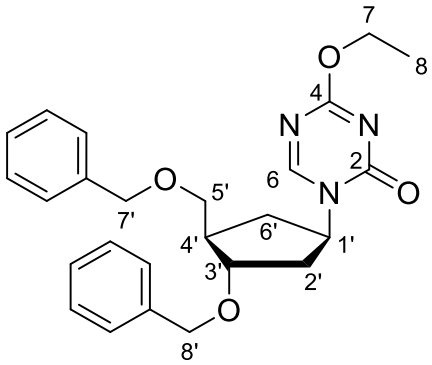
R_f = 0.51 (DCM/MeOH = 19:1)

HRMS [ESI$^+$]: $[C_{23}H_{29}N_4O_4Na]^+$, $[M+Na^+]^+$ calculated: 434.2050, found: 434.2049

^{1}H-NMR (599 MHz, $CDCl_3$): *δ*/ppm = 7.35 – 7.16 (m, 10H, Ar-H), 5.42 (d,$^3J_{1',N1}$ = 7.8 Hz, 1H, N_1-H), 4.50 – 4.34 (m, 4H, $C_{7'}$-$H_{a,b}$, $C_{8'}$-$H_{a,b}$), 4.33 – 4.16 (m, 1H, $C_{1'}$-H), 3.93 – 3.81 (m, 1H, $C_{3'}$-H), 3.62 (s, 3H, C_6-H), 3.43 – 3.31 (m, 2H,$C_{5'}$-H), 2.34 – 2.18 (m, 2H, $C_{6'}$-H_a, $C_{4'}$-H), 2.06 (ddd, $^2J_{2'a,2'b}$ = 13,0 Hz, $^3J_{2'a,1'/3'}$ = 6,9 Hz, $^3J_{2'a,1'/3'}$ = 4,6 Hz, 1H, $C_{2'}$-H_a), 1.73 (ddd, $^2J_{2'a,2b'}$ = 13.0 Hz, $^3J_{2'b,1'/3'}$ = 7.0 Hz, $^3J_{2'b,1'/4'}$ = 7.0 Hz, 1H, $C_{2'}$-H_b), 1.28 – 1.11 (m, 1H, $C_{6'}$-H_b).

^{13}C-NMR (151 MHz, $CDCl_3$): *δ*/ppm = 163.8 (C_2), 162.3 (C_4), 138.9 (C_{Ar}), 138.5 (C_{Ar}), 128.51 (C_{Ar}), 128.41 (C_{Ar}), 127.76 (C_{Ar}), 127.74 (C_{Ar}), 127.70 (C_{Ar}), 127.56 (C_{Ar}), 81.1 ($C_{3'}$), 73.3 ($C_{7'}$),71.9 ($C_{5'}$), 71.2 ($C_{8'}$), 53.8 (C_6), 49.5 ($C_{1'}$), 45.0 ($C_{4'}$), 39.5 ($C_{2'}$), 35.0 ($C_{6'}$).

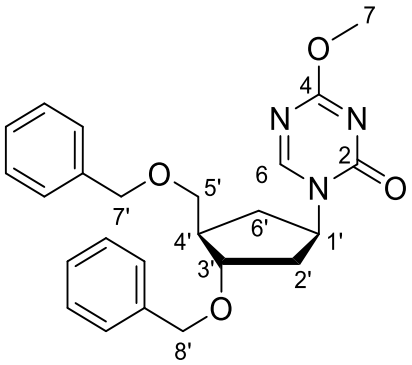
4-Ethoxy-1-[(1'*R*,3'*S*,4'*R*)-3'-benzyloxy-4'-(benzyloxymethyl)-cyclopentyl]-1H-[1,3,5]-triazin-2-one (9a) and **4-methoxy-1-[(1'*R*,3'*S*,4'*R*)-3'-benzyloxy-4'-(benzyloxymethyl)-cyclopentyl]-1H-[1,3,5]-triazin-2-one (9b)**



9a

$C_{25}H_{29}N_3O_4$

435.52 g/mol



9b

$C_{24}H_{27}N_3O_4$

421.50 g/mol

Trifluoroacetic acid (75 µL) was added to a solution of **8** (2.613 g, 6.35 mmol, 1.0 eq) in triethyl ortho-formiate (60 mL). The mixture was refluxed for 3 h. The solvent was removed *in vacuo*. The residue was co-evaporated once with MeOH and the crude product was purified by column chromatography on silica gel with a stepwise gradient of *i*Hex/EtOAc (2:1→1:1→1:2) to obtain **9a** (1.774 g, 4.07 mmol, 64%, colorless solid) and **9b** (268.0 mg, 0,64 mmol, 10%, colorless oil).

9a:

R_f = 0.71 (iHex/EtOAc, 1:3)

HRMS [ESI$^+$]: $[C_{25}H_{30}N_3O_4H]^+$, $[M+H^+]^+$ calculated: 436.2232, found.: 436.2231.

^{1}H-NMR (400 MHz, $CDCl_3$): *δ*/ppm = 8.23 (s, 1H, C_6-H), 7.38 – 7.24 (m, 10H, Ar-H), 5.07 – 4.89 (m, 1H, $C_{1'}$-H), 4.56 – 4.39 (m, 6H, $C_{7'}$-$H_{a,b}$, $C_{8'}$-$H_{a,b}$, C_7-$H_{a,b}$), 4.07 – 3.08 (m, 1H, $C_{3'}$-H), 3.61 – 3.46 (m, 2H, $C_{5'}$-$H_{a,b}$), 2.52 – 2.36 (m, 2H, $C_{6'}$-H_a, $C_{4'}$-H), 2.28 (ddd, 1H, $^2J_{2'a,2'b}$ = 13,3 Hz, ^3J = 7,6 Hz, ^3J = 2,7 Hz, $C_{2'}$-H_a), 2.11 (ddd, 1H, $^2J_{2'a,2'b}$ = 13.3 Hz, ^3J = 10.0 Hz, ^3J = 6.3 Hz, $C_{2'}$-H_b), 1.84 – 1.68 (m, 1H, $C_{6'}$-H_b), 1.40 (t, 3H, $^3J_{8,7}$ = 7,1 Hz, C_8-H),

^{13}C-NMR (101 MHz, $CDCl_3$): *δ*/ppm = 169.3 (C_4), 158.4 (C_6), 155.0 (C_2), 138.2 (C_{Ar}), 138.1 (C_{Ar}), 128.6 (C_{Ar}), 128.6 (C_{Ar}), 127.93 (C_{Ar}), 127.85 (C_{Ar}), 127.82 (C_{Ar}), 80.4 ($C_{3'}$), 73.4 ($C_{7'/8'}$), 71.26 ($C_{7'/8'}$), 71.25 ($C_{5'}$), 65.1 (C_7), 56.8 ($C_{1'}$), 44.9 ($C_{4'}$), 36.9 ($C_{2'}$), 32.6 ($C_{6'}$), 14.2 (C_8).

9b:

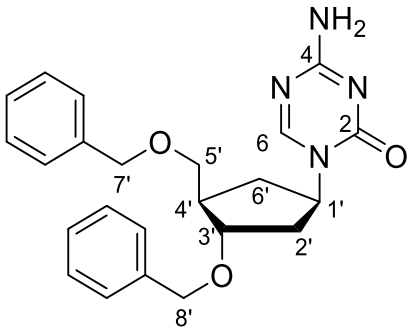
R_f = 0.63 (iHex/EtOAc, 1:3)

HRMS [ESI$^+$]: $[C_{24}H_{28}N_3O_4]^+$, $[M+H^+]^+$ calculated: 422.2074 , found: 422.2075.

^{1}H-NMR (599 MHz, $CDCl_3$): *δ*/ppm = = 8.23 (s, 1H, C_6-H), 7.38 – 7.25 (m, 10H, Ar-H), 5.04 – 4.93 (m, 1H, $C_{1'}$-H), 4.57 – 4.43 (m, 6H, $C_{7'}$-$H_{a,b}$, $C_{8'}$-$H_{a,b}$), 4.09 – 3.98 (m, 4H, $C_{3'}$-H, C_7-H), 3.60 – 3.50 (m, 2H, $C_{5'}$-$H_{a,b}$), 2.50 – 2.39 (m, 2H, $C_{6'}$-H_a, $C_{4'}$-H), 2.28 (ddd, 1H, $^2J_{2'a,2'b}$ = 13,1 Hz, ^3J = 7,4 Hz, ^3J = 2,7 Hz, $C_{2'}$-H_a), 2.11 (ddd, 1H, $^2J_{2'a,2'b}$ = 13.1 Hz, ^3J = 10.1 Hz, ^3J = 6.3 Hz, $C_{2'}$-H_b), 1.80 – 1.68 (m, 1H, $C_{6'}$-H_b).

^{13}C-NMR (101 MHz, $CDCl_3$): *δ*/ppm = 169.9 (C_4), 158.5 (C_6), 155.0 (C_2), 138.3 (C_{Ar}), 138,0 (C_{Ar}), 128.63 (C_{Ar}), 128.57 (C_{Ar}), 127.93 (C_{Ar}), 127.84 (C_{Ar}), 127.82 (C_{Ar}), 80.4 ($C_{3'}$), 73.5 ($C_{7'/8'}$), 71.3 ($C_{7'/8'}$), 56.9 ($C_{1'}$), 56.0 (C_7), 44.9 ($C_{4'}$), 36.9 ($C_{2'}$), 32.6 ($C_{6'}$).

4-Amino-1-[(1'*R*,3'*S*,4'*R*)-3'-benzyloxy-4'-(benzyloxymethyl)-cyclopentyl]-1H-[1,3,5]triazin-2-one (10)



10

$C_{23}H_{26}N_4O_3$

406.47 g/mol

Ethoxytriazine **9a** (1.630 g, 3.74 mmol, 1.0 eq.) was dissolved in methanolic NH_3 (7 N in MeOH, 60 mL) and stirred for 3 h at rt. The mixture was diluted with H_2O (340 mL), resulting in a colorless precipitate, which was extracted wit EtOAc (3 × 200 mL). The organic phases were combined, dried and the solvent was removed *in vacuo*, which is resulted in the benzyl-protected cAzadC **10** (1.478 g, 3.63 mmol, 97%) as a colorless foam. Synthesis of **10** was also successful starting from the methoxytriazine **9b** and following the same procedure.

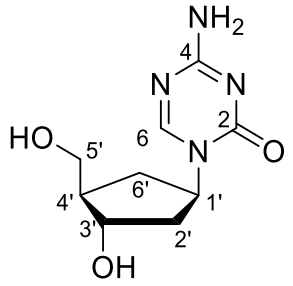
R_f = 0.50 (CH_2Cl_2/MeOH, 9:1)

HRMS [ESI$^+$]: $[C_{23}H_{27}N_4O_3]^+$, $[M+H^+]^+$ calculated: 407.2078, found.: 407.2081.

^{1}H-NMR (599 MHz, $CDCl_3$): *δ*/ppm = 8.01 (s, 1H, C_6-H), 7.35 – 7.23 (m, 10H, Ar-H), 5.78 (s, 2H, NH), 4.92 – 4.76 (m, 4H, $C_{7'}$-H and $C_{8'}$-H), 4.59 – 4.46 (m, 1H, $C_{1'}$-H), 4.07 – 3.95 (m, 1H, $C_{3'}$-H), 3.55 – 3.45 (m, 2H, $C_{5'}$-H), 2.44 – 2.28 (m, 2H, $C_{6'}$-H_a, $C_{4'}$-H), 2.24 – 2.16 (m, 1H, $C_{2'}$-H_a), 2.14 – 2.05 (m, 1H, $C_{2'}$-H_b), 1.78 – 1.67 (m, 1H, $C_{6'}$-H_b).

^{13}C-NMR (151 MHz, $CDCl_3$): *δ*/ppm = 165.7 (C_4), 157.1 (C_6), 154.2 (C_2), 138.3, 138.3, 128.6, 128.5, 127.84, 127.78 (C-Ar), 80.4 ($C_{3'}$), 73.4 ($C_{8'}$), 71.4 ($C_{5'}$), 71.2 ($C_{8'}$), 56.7 ($C_{1'}$), 44.9 ($C_{4'}$), 36.8 ($C_{2'}$), 32.5 ($C_{6'}$).

4-Amino-1-[(1'*R*,3'*S*,4'*R*)-3'-hydroxy-4'-(hydroxymethyl)-cyclopentyl]-1H-[1,3,5]triazin-2-on (cAzadC, 1)



1

$C_9H_{14}N_4O_3$

226.24 g/mol

A solution of benzyl-protected **10** (1.478 g, 3.63 mmol, 1.0 eq) in CH_2Cl_2 (100 mL) was cooled to -78 °C and BCl_3 (1 M in DCM, 12.7 mL, 12.7 mmol, 3.5 eq.) was added dropwise. The mixture was stirred for 1 h at -78 °C and stirred for additional 2 h at rt. The reaction was quenched by addition of MeOH (85 mL) under constant stirring. The solvent was removed *in vacuo*. The crude product was purified by column chromatography on silica gel with a stepwise gradient CH_2Cl_2/MeOH (9:1→7:3) to obtain cAzadC (**1**, 0.740g, 3.27 mmol, 90%). Recrystallization from hot MeOH resulted in 57% cAzadC (**1**, 465.3 mg, 2.06 mmol) as colorless acicular monocrystalls.

R_f = 0.1 (CH_2Cl_2/MeOH, 9:1)

Mp.: 198-200 °C

HRMS [ESI$^+$]: $[C_9H_{15}N_4O_3]^+$, $[M+H^+]^+$ calculated: 227,1138, found.: 227,1139.

^{1}H-NMR (400 MHz, D_2O): *δ*/ppm = 8.29 (s, 1H, C_6-H), 4.90 – 4.62 (m, 1H, $C_{1'}$-H, Overlap with D_2O signal), 4.29 – 4.16 (m, 1H, $C_{3'}$-H), 3.73 (dd, $^2J_{5'a,\ 5'b}$ = 11.2 Hz, $^3J_{5'a,\ 4'}$ = 5.7 Hz, 1H, $C_{5'a}$-

H), 3.62 (dd, $^{2}J_{5'a,\,5'b}$ = 11.2 Hz, $^{3}J_{5'b,\,4'}$ = 6.8 Hz, 1H, $C_{5'b}$-H), 2.41 – 2.29 (m, 1H, $C_{6'a}$-H), 2.28 – 2.18 (m, 1H, $C_{2'a}$-H), 2.17 – 2.02 (m, 2H, $C_{2'b}$-H, $C_{4'b}$-H), 1.77 – 1.55 (m, 1H, $C_{6'b}$-H).

^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 165.5 (C_4), 158.5(C_6), 156.4(C_2), 72.2 ($C_{3'}$), 62.8 ($C_{5'}$), 56.2 ($C_{1'}$), 48.11 ($C_{4'}$), 38.0 ($C_{2'}$), 32.0 ($C_{6'}$).

FTIR (ATR): $\tilde{\nu}$/cm^{-1} = 3194 (*m*), 1653 (*m*), 1540 (*s*), 1437 (*s*), 1278 (*s*), 1036 (*s*).

X-ray crystal structure analysis of cAzadC 1

The cAzadC **1** monocrystals were analysed by X-ray crystallography (Table SI 1). The solved structure (Figure SI 1) shows the cAzadC **1** in two different Pucker conformations. Further figures were generated with the program *Mercury 3.5.1.* (*Cambridge Crystallographic Data Center*). The data are deposited under CCDC 1910952.

Table SI 1: Crystallographic data for cAzadC (**9**)

net formula	$C_9H_{14}N_4O_3$	transmission factor range	0.9087–0.9580
M_r/g mol^{-1}	226.24	refls. measured	34579
crystal size/mm	0.100 × 0.080 × 0.030	R_{int}	0.0373
T/K	100(2)	mean σ(*I*)/*I*	0.0197
radiation	MoKα	θ range	3.031–25.40
diffractometer	'Bruker D8Venture'	observed refls.	3474
crystal system	orthorhombic	*x, y* (weighting scheme)	0.0372, 0.4868
space group	'P 21 21 21'	hydrogen refinement	mixed
a/Å	8.4451(3)	Flack parameter	0.3(3)
b/Å	11.0958(4)	refls in refinement	3787
c/Å	21.9635(8)	parameters	321
α/°	90	restraints	0
β/°	90	$R(F_{obs})$	0.0284
γ/°	90	$R_w(F^2)$	0.0699
V/Å^3	2058.09(13)	*S*	1.054
Z	8	shift/error$_{max}$	0.001
calc. density/g cm^{-3}	1.460	max electron density/e Å^{-3}	0.180
μ/mm^{-1}	0.112	min electron density/e Å^{-3}	−0.173
absorption correction	multi-scan		

C-H: constr, O-H and N-H: refall

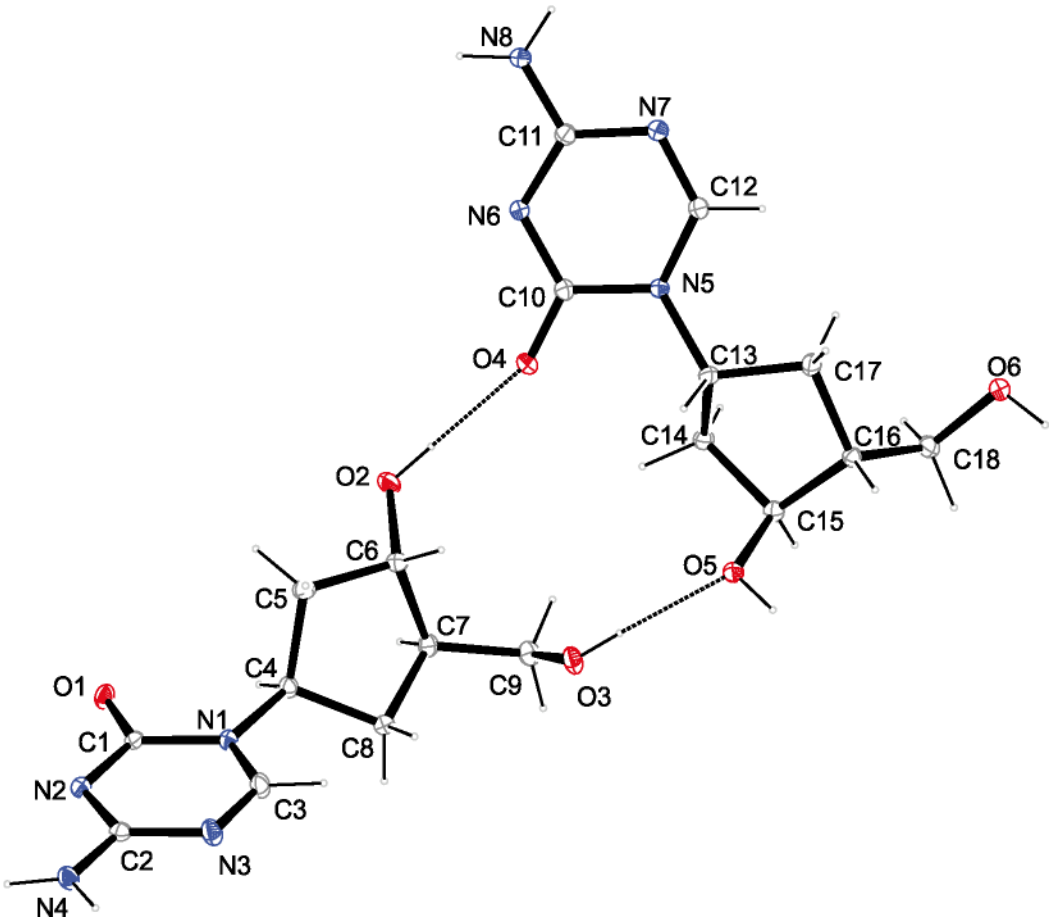


Figure SI 1: Monocrystal X-ray structure analysis of cAzadC (**9**) with elipsoids for the probability of presence of 50%.

HPLC stability analysis of cAzadC (1)

To determine the hydrolytic stability of the carba-cyclic nucleoside cAzadC (**1**) in comparison with AzadC, both nucleosides were dissolved in aqueous KH_2PO_4 buffer (100 mM, pH = 7.4, 5.5 and 8.5) in 100 mM concentration. The solutions were immediately analysed *via* HPLC (t=0 h) with subsequent HPLC analyses every 1.5 h or at indicated time points. As stationary phase an EC 250/4 NUCLEOSIL 120-3 C18 (*Macherey-Nagel*) chromatography column was used. The mobile phase consisted of water (buffer A) and acetonitrile (buffer B) at a flow-rate of 0.5 mL/min as follows: 0→25 min, 0→5% buffer B; 25 min→28 min, 5%→80%; 28 min→38 min, 80%; 38 min→43 min 80%→0%; 43 min→45 min, 0%.

UHPLC-MS^2 analysis of cAzadC (1) and calibration curves for cAzadC (1)

UHPLC-MS^2 for quantification of DNA modifications was performed as describes previously.[2] The external calibration curve was generated by serially diluting pure cAzadC [20 pmol, 10 pmol, 5 pmol, 2.5 pmol, 1.25 pmol, 0.625 pmol, 0.3125 pmol] and measuring it in technical triplicates prior to each measurement. Linear regression was done by *OriginPro 2016G*.

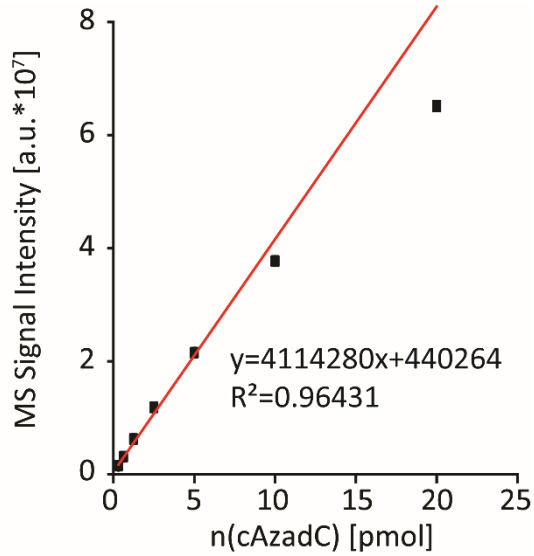


Figure SI 2: Calibration curve for carbacyclic 5-aza-2'-deoxycytidine **1** (cAzadC) quantification in genomic DNA by tandem mass spectrometry (MS^2).

[1] B. M. Domínguez, P. M. Cullis, *Tetrahedron Lett.* **1999**, *40*, 5783-5786.

[2] S. Schiffers, T. M. Wildenhof, K. Iwan, M. Stadlmeier, M. Müller, T. Carell, *Helv. Chim. Acta* **2019**, *102*, e1800229.