Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilian-Universität München

Quantification of epigenetic bases and oxidative lesions in lung tissues

Florian Christian Schelter

aus:

Trostberg, Germany

2021

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstaatliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 22. November 2021

_____
Florian Schelter

Dissertation eingereicht am: 25.11.2021

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Dr. Pavel Kielkowski

Mündliche Prüfung am: 13.01.2022

Für meine Familie,

Großeltern

und Freunde

Happiness is the key to success.
If you love what you are doing, you will be successful.
It always seems impossible until it's done.
Nothing will come of nothing.

Albert Schweitzer

Table of content

List of abbreviations

-/-	knockout of both alleles of a gene
+/+	fully functional gene
°	degree
8-oxo-dG	8-8xo-2’-deoxyguanosine
a-dC	6-aza-2’deoxycytidine
a-fdC	6-aza-5-formyl-2´-deoxycytidine
APE1	apurinic/apyrimidinic endonuclease 1
ATP	adenosine triphosphate
BER	base excision repair
cadC	5-carboxyl-2’-deoxycytidine
CE	collision energy
CpG	dC-dG-dinucleotide
cm	catalytic mutant
COPD	chronic obstructive pulmonary disease
CSE	cigarette smoke extract
dA	2’-deoxyadenosine
dC	2’-deoxycytidine
DNA	deoxyribonucleic acid
DNMT	DNA-methyltransferase
dG	2’-deoxyguanosine
ESI	electrospray ionization
fdC	5-formyl-2’-deoxycytidine
F-dC	2'-fluoro-2’-deoxycytidine
F-fdC	2'-fluoro-5’-formyl-2’-deoxycytidine
HEK293T	human embryoid kidney
HESI	heated electrospray ionization
hmdC	5-hydroxymethyl-2’-deoxycytidine
HPLC	high performance liquid chromatography
IPF	idiopathic pulmonary fibrosis
LC	liquid chromatography
LSD1	Lys-demethylases 1
MBDs	methyl-CpG-binding domains

mdC	5-methyl-2'-deoxycytidine
MeCN	acetonitrile
MeOH	methanol
mESC	murine embryonic stem cells
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NF-κB	nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
NSCLC	non-small cell lung cancers
Ogg1	8-oxoguanine glycosylase
RNA	ribonucleic acid
RNOS	reactive nitrogen and oxygen species
RNS	reactive nitrogen species
ROS	reactive oxygen species
SAM	*S*-adenosyl-methionine
T	2'-deoxythymidine
Tdg/TDG	thymine-DNA-glycosylase
TET	ten-eleven translocation enzyme
UHPLC	ultrahigh performance liquid chromatography
UV	ultraviolet light
vs.	versus

List of publications

Schelter, F.[#], Kirchner, A.[#], Traube, F.R., Müller, M., Steglich, W. and Carell, T., 5-Hydroxymethyl-, 5-Formyl- and 5-Carboxydeoxycytidines as Oxidative Lesions and Epigenetic Marks, *Chem. Eur. J.* **2021**, 27(31):8100-8104. DOI: 10.1002/chem.202100551

F. Schelter[#], A. Schön[#], E. Kaminska[#], E. Ponkkonen, E. Korytiaková, Sarah Schiffers, T. Carell*; Analysis of an active deformylation mechanism of 5-formyldeoxycytidine (fdC) in stem cells, *Angew. Chem. Int. Ed.* **2020**, *59*(14), 5591-5594; DOI: 10.1002/anie.202000414.

 # these authors contributed equally to this publication

Conference papers

GRK2338 Meeting, Herrsching (2019) - poster presentation *“New Synthesis road for isotopically labelled inosine modifications”*.

RNA epigenetics in human disease, Cambridge (2019) - poster presentation *“Synthesis of new isotopically labelled Inosine modifications”*.

Summary

Nowadays, lung diseases like chronic obstructive pulmonary disease (COPD), idiopathic pulmonary fibrosis (IPF) and lung cancer, play a crucial role in the lives of many people. The numbers of deaths related to these diseases are increasing steadily. Unfortunately, there is still no effective treatment available and the underlying mechanism not fully elucidated. However, a connection between smoking and the development of these diseases is obvious. Through the consumption of tobacco products, various toxic substances and reactive oxygen species (ROS) are inhaled into the lungs leading to damages in the cellular compartments. While most of the organic compounds inhaled by mammals become active carcinogens in the attempt to eliminate them, ROS are free oxygen radicals, like superoxides or hydrogen peroxides, that can directly react with cellular biomolecules, leading ultimately to diseases such as COPD, IPF or cancer. Besides tobacco smoke as the major external source, ROS are also generated in cellular processes, such as the electron transport chain in mitochondria or NADPH oxidases. Here, they also play a role as signaling agents. Therefore, a healthy balance of the generation and disposal of these ROS is necessary. An important marker of the influence of ROS on DNA, is 8-oxo-2'-deoxyguanosine (8-oxo-dG), the oxidation product of 2'-deoxyguanosine (dG), which is formed by the reaction of hydroxyl radicals with DNA. Besides the direct oxidation of DNA strands, ROS also have an influence on other processes such as the methylation of dC. Here it might affect a methyl group transfer of the cofactor *S*-adenosyl-methionine (SAM) on to 2'-deoxycytidine (dC) to form 5-methyl-2'-deoxycytidine (mdC), one of the most important regulatory nucleotides in epigenetics. mdC can then further be oxidized to 5-hydroxymethyl-2'-deoxycytidine (hmdC), 5-formyl-2'-deoxycytidine (fdC) and 5-carboxyl-2'-deoxycytidine (cadC) by either ten-eleven translocation enzyme (Tet) or ROS. Knowing about the global amounts of these modified nucleosides from several studies, it is still questionable to what extent epigenetic processes and oxidative damages contribute to their origin. This question was addressed by investigating the nature grown fungus *Amanita muscaria*. Judging from its fully sequenced genome, *Amanita muscaria* does not show any Tet or Tet-like enzymes, making this mushroom a good example for the detection of the basal levels of oxidatively damaged nucleosides by ultra-high-performance liquid chromatography coupled to a triple quadrupole mass spectrometer (UHPLC-MS/MS). A comparison to cell lines like HEK293T and iNGN cells, as well characterized epigenetic model systems grown under controlled conditions, was performed, to gain information about the basal levels of modified dC-derived DNA products.

These obtained data were then, in a yet unpublished work, put into relation with the development of human lung tumors. In a first step, the methylation levels of cytosine and of its oxidative derivatives, as well as of 8-oxo-dG, were examined in detail by UHPLC-MS/MS, to compare the epigenetic amount of these modifications in healthy and tumor-like tissues. This allowed us to unravel the differences that occur under those two conditions. In an attempt, to learn more about the major causes that leads to the development of lung cancer, we studied in a second experiment how cigarette smoke leads to changes of the nucleosides mentioned above. To this end, we treated two different cell lines, A549 and EKVX, with cigarette smoke and extracted the DNA at different time points over a long period and compared the results to unexposed control cells using UHPLC-MS/MS. In a further experiment, short-term hyperoxia treated mice were examined for the effects on biological functions of the lung. This is a quite interesting question since early born infants sometimes need to be artificially ventilated after birth over a short time. Unfortunately, little is known about the long-term effects of this treatment. Hereby, we looked for the modification levels, especially of 8-oxo-dG, as the main marker for oxidative stress, to measure the potential side effect of short-time ventilated mice. Also, the effect on repair processes was evaluated within this study, since little is known about the regulating system.

In another part of this thesis, we focused on the active deformylation of fdC to dC. An earlier publication already suggested a direct deformylation of fdC to dC, but the underlying mechanism is still not fully elucidated. Therefore, mechanistic conclusions of this demethylation could be obtained by means of a chemically modified fdC analogue in *in vivo* studies, using various analytical methods including UHPLC-MS/MS. Here, we proofed that a nucleophilic attack on the C6-position of fdC is necessary for deformylation to dC.

1. Introduction

1.1 Lung diseases

A human breathes up to 15 times a minute and inhales and exhales every time up to 700 mL of air. Summing this up for a day, up to 15.000 liters of air are exchanged. With this, the lung delivers fresh oxygen to and removes CO_2 from the blood. But not only the exchange of oxygen and CO_2 takes place during breathing, also small, airborne particles and gases can enter and harm the lungs[1]. Besides dust, viruses and exhaust gases, smoking plays a crucial role for the development of mild or severe lung damage and lung diseases[2-3].

The most common lung diseases are asthma, idiopathic pulmonary fibrosis (IPF), pneumonia, and chronic obstructive pulmonary disease (COPD). Asthma and pneumonia are well known diseases caused by an allergic inflammation[2] or bacterial infection of the airways, which can successfully be treated with medication nowadays. IPF is a rare type of pulmonary fibrosis leading to scarred lung tissues. It is a deadly disease with a survival time of only 2-3 years[2]. COPD is a progressive degeneration of the airways and one of the leading causes of death from smoking[2, 4-5]. The Federal Statistical Office of Germany provided data for the years 1998 to 2019[6], which show the numbers of deaths caused by lung diseases (Figure **1-1**).

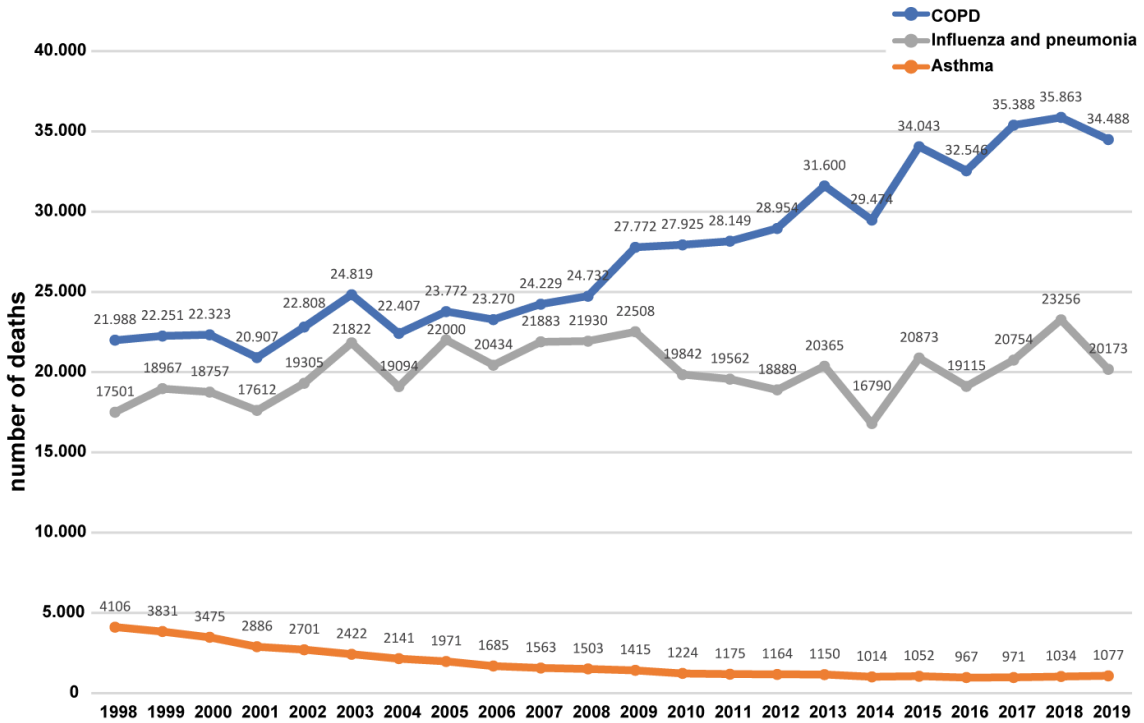


Figure 1-1: Deaths caused by different lung diseases in Germany from 1998 to 2019.

Figure **1-1** shows that the mortality rate for asthmatic diseases have declined gradually from 1998 until 2019. The fatality rates caused by pneumonia have remained almost constant at a number of around 20.000 cases. In contrast, the number of COPD cases has been rising steadily over the years and almost doubled the number of deaths from pneumonia in 2019. The different progression of the curves shown above already indicates the major problem with COPD. There is still no efficient medical treatment available. A study of IPF-related deaths in England and Wales attributed 82.702 deaths to IPF from 1979 until 2016[7], but this information is not presented here.

1.2 Influence of oxidative stress on lung diseases and epigenetics

1.2.1 Definition of reactive oxygen and reactive nitrogen species

Reactive oxygen species (ROS) are oxygen containing radicals with one or more unpaired electrons[8]. The term ROS is frequently used in biology and medicine and comprises for example superoxides ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($OH^{\bullet}$)[8]. Similar to ROS, reactive nitrogen species (RNS), which include nitric oxide ($NO^{\bullet}$) or peroxy-nitrite and nitrogen dioxide radical ($NO_2^{\bullet}$) are generated in the cells or are available in nature[8].

1.2.2 Cigarette smoke induces oxidative stress as a trigger for COPD and lung cancer

All the diseases mentioned in chapter **1.1** are predominantly related to exogeneous and endogenous oxidative stress. The lung has an surface area of ~ 140 m^2, making it vulnerable to exogeneous ROS[9]. These ROS harm the bronchial structure by inflicting oxidative stress leading to inflammatory processes[10]. The most common exogenous factors of ROS and RNS are ionizing radiation[11] and smoking, whereby smoking is one of the most important exogenous sources. Every puff of cigarette smoke, consisting of 95% gases such as nitrogen, oxygen, carbondioxide or nitric oxide, contains up to 1 x 10^{15} free radicals[12-15]. For example, the reaction product of nitric oxide and superoxide, peroxynitrite, as well as the formation of an quinone-hydroquinone radical complex, generates superoxide anions that can cause single-strand breakage in DNA[15]. *In vitro*

experiments demonstrated also lipid peroxidation in human blood plasma caused by tobacco smoke[15]. The last 5% of inhaled tobacco smoke contains for example benzo[*a*]pyrene (BaP) one of the most extensively studied compounds and inducer of lung cancer[15]. Compounds like BaP are not carcinogenic themselves, but their metabolites, like BaP-epoxides, are (see Figure **1-2**). These metabolites result from enzymatic processing by cytochrome P450 and other enzymes, to excrete them easily[15].

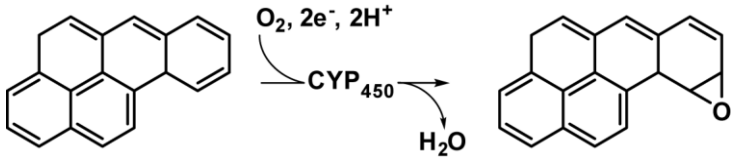


Figure 1-2: Depiction of the CYP450 induced generation of an epoxide as a degradation product of BaP.

Further intermediates are formed by glutathione *S*-transferases, sulfatases or uridine-5'-diphosphate-glucuronosyltransferases[15]. The formed products, which are mostly epoxides of BaP, may react with DNA forming DNA adducts, other macromolecules[16] or harm the cellular structure and hereby cause inflammations. This inflammatory process is triggered by components of the immune system like granulocytes, macrophages and lymphocytes to eliminate the pathogens and to restore a healthy condition[2, 17]. It is known that COPD patients have elevated numbers of activated alveolar macrophages and neutrophils in areas of inflammation[18-20]. This inflammation can damage the epithelial cells of the respiratory tract, which leads to destruction of the original lung structure and function[2, 17], resulting in a fragmentation of the reticular basement membrane and an extended inflammation caused by apoptosis[21]. Macrophages generate further intracellular ROS for the phagocytic process for the uptake and clearance of apoptotic cells[18-20]. Detrimental effects include mucous glands, hyperplastic obstruction of the airways, alveolar destruction and metaplasia[22]. If the cellular repair mechanisms are not working properly, an inflamed cell can further develop to a cancerous cell, based on processes that lead to genomic instability[11]. There are different lung cancer variants such as small cell lung cancer (SCLC) or non-small cell lung cancers (NSCLC, 85% of lung cancer cases)[23-25]. The most common cancer types[26] for NSCLC are adenocarcinoma or squamous cell cancer[2, 25], as shown by studies of *Taghizadeh et al.*[27] and *Agudo et al.*[28]. A link between smoking and the development of COPD or lung cancer through oxidative stress can be suspected[29].

1.2.3 Impact of oxidative stress on cells

While tobacco smoke is the major external source, reactive nitrogen and oxygen radicals (RNOS) are also generated in cellular processes which can harm the cells, too[30]. Most intracellular RNOS are produced as by-products in mitochondria during the electron transport chain, where 1-2% of electrons react with oxygen[31-33]. Hereby, superoxides ($O_2^{\bullet-}$) are formed by the incomplete conversion of oxygen to water during the oxidative phosphorylation producing adenosine triphosphate (ATP). These superoxides have a short half-life and are catalytically degraded by superoxide dismutases[34] to hydrogen peroxide (H_2O_2) with a longer half-life[35]. This H_2O_2 can pass the mitochondrial membrane and enter the cytosol causing an activation of pro-inflammatory cytokines[33], or it can be disproportioned to water and oxygen by a catalase[35-36]. RNOS are further generated by peroxisomes (containing H_2O_2 producing enzymes)[37], NADPH oxidases and the endoplasmic reticulum, where H_2O_2 is generated as a by-product during oxidative protein folding[11, 38]. RNOS also have an impact on cellular repair mechanisms or damage cellular protective mechanisms[12], like the "nuclear shield"[39], consisting of free radical scavengers such as glutathione transferases[39], vitamins C and E, and antioxidant enzymes like catalases, superoxidases or dismutases[9, 14, 30, 40-43]. In Figure **1-3** (modified according to *Durham et al.*[12]) different biological processes are shown where oxidant/antioxidant imbalances of RNOS cause injuries by oxidizing cellular constituents like proteins, lipids or nucleic acids[9, 33].

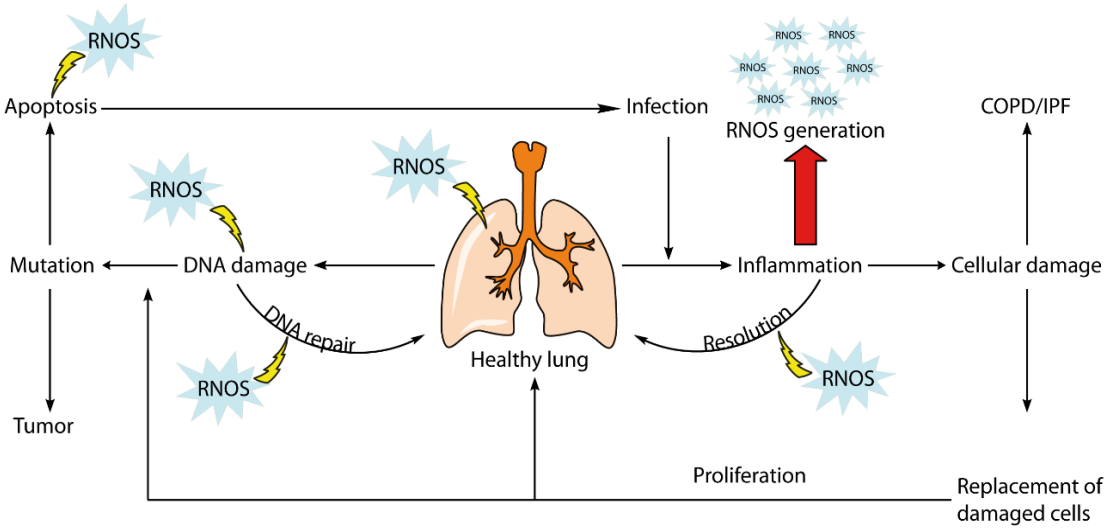


Figure 1-3: Depiction of different cellular processes as a response to internal or external oxidative lesions. It further shows the places of action of RNOS[12].

If not destroyed, RNOS react with multiple cellular components, eventually causing DNA damage or inflammations. RNOS can interact with DNA by oxidizing the canonical nucleosides dG, dA, T or dC, thereby generating the most abundant DNA oxidation product 8-oxo-2'-deoxyguanosine (8-oxo-dG). The ROS-caused oxidation mechanism of 2'-deoxyguanosine to 8-oxo-2'-deoxyguanosine (8-oxo-dG) was described by different laboratories[44-47] and is shown in Figure **1-4** (modified according to *Fleming et al.*[36]).

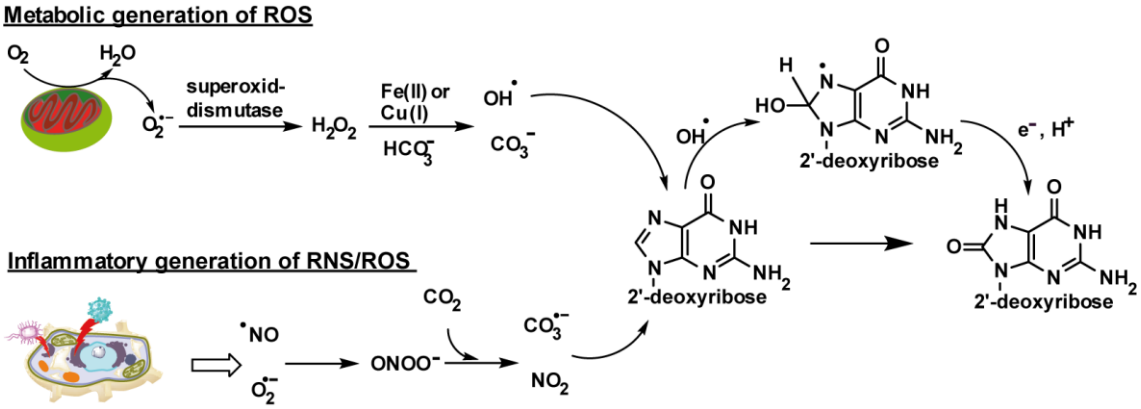


Figure 1-4: Oxidative modification of dG induced by ROS generated during metabolism and inflammation.

RNOS can also affect mismatches or chromosomal changes by attacking phosphates and base sites. Furthermore, they are also responsible for single strand breaks or double strand breaks (DSB)[9, 36, 48]. RNOS-induced dangerous alterations need to be corrected by repair mechanisms, which ironically might itself be damaged by ROS as mentioned above. Thereby, DSB can be mutagenic causing chromosomal rearrangements, or loss of genetic information due to dysfunctional DNA repair mechanisms[11]. Furthermore, RNOS can cause proliferation of tumorigenic cells through inactivation of protein tyrosine phosphatase, which belongs to a few regulator enzymes of cell growth, proliferation and signal transduction by oxidizing the enzyme's catalytic thiolate group, leading to irreversible inactivation[9, 49-51]. In general, an excessive and uncontrolled release of RNOS can cause an overstrained repair system and results in cellular defects[36, 52]. RNOS are also involved in beneficial processes. For example, during inflammations caused by infections with microorganisms, toxins or radiation, the immune system sends macrophages to the inflamed area to eliminate the pathogens by the production of RNOS[33, 36, 52-54]. These RNOS can inhibit the resolution of an inflammation, causing further spreading of the inflammation and in turn an extended RNOS production[12] or can help the cellular repair[33]. Thereby, the time and intensity

of an inflammation plays a crucial role in the development of certain diseases like COPD and IPF[9, 54] where the oxidation of lipids causes a permanent damage of cell membranes[9]. Furthermore, RNOS have an important regulatory function as second messengers. They act as modulators of signalling mechanisms like proliferation, apoptosis, cytokines or transcription[9]. These cellular mechanisms are mostly triggered by certain kinases of mitogen-activated protein kinases (MAPK) family due to a relay of signals from the outside of a cell to the inside by phosphorylation of proteins[9]. Several studies suggested that an activation of extracellular signal-regulated kinase (ERK1/2) in lung cells, has a protective effect towards cell survival[55]. Furthermore, an induction of the stress-activated protein kinases C-Jun-terminal protein kinase (JNK1/2) and p38 kinase, by MAPK kinases, induce apoptosis[55]. First studies revealed possible approaches for therapy of certain lung diseases through modulation of the components of the MAPK signalling pathways[55], but this needs to be further investigated.

Thus, RNOS play a crucial role in different cellular processes. On the one hand, they can harm cells in different processes, inducing DNA damage or inflammation, as illustrated in Figure **1-3**. But on the other hand, RNOS function as signalling agents, like protein kinases and phosphorylation[56], for immune cell activation[30, 57] or vascular remodelling and coordinate cellular processes[58], like maintaining a healthy cellular state[35-36, 54, 59-60]. Summing it up, RNOS play both, positive and negative, roles in the cellular metabolism. Thereby, the reaction with DNA, leading to the formation of oxidized dC or dG derivatives plays a major role. These oxidation products might, when forming epigenetically relevant bases as oxidative lesions, have an influence on epigenetic processes.

1.2.4 Detection of oxidative lesions in urine

RNOS, like superoxides or hydrogen peroxide, cannot be quantified in body fluids, since they are highly reactive and have short half-lives[61]. But their stable and excreted oxidized biomolecules can be measured for example by mass spectrometry[61]. Different matrices like blood, salvia, urine and tissues are usable for the detection of these biomolecules[61], but urine is the preferred medium, since it has an easy, non-invasive sampling method[61]. Besides 8-oxo-dG, the tissue membrane phospholipid 8-isoprostane, an RNOS caused oxidation product of arachidonic acid, as

well as neopterin synthesized by the macrophages, an catabolic product of guanosine triphosphate working as an inflammatory indicator, are biomolecules of interest for the inflammatory measurements[61-62]. These biomolecules are excreted into the urine and can be detected after several extraction steps by UHPLC-MS/MS[61]. First analysis associated a high excretion of urinary 8-oxo-dG with lung and breast cancer, whereas 8-isoprostane is correlated with cardiovascular diseases[63-65].

1.3 Epigenetics

The term epigenetics has evolved over the years from the word “epigenesis”, which was coined by William Harvey a physician and physiologist around 1650[66]. The original idea was proposed by Aristotle who described with “epigenesis” a process of increasing complexity arisen from initially homogenous material[66]. The Greek term epigenesis can be translated as a state arising from a preformed origin (gr. *epi* after, *genesis origin)*[66]. In 1942 Conrad Waddington defined it as the entire complex of developmental processes between genotype and phenotype[66-67]. Today, the term epigenetics is used to describe molecular processes regulating and altering the gene expression in a level beyond the actual DNA sequence[68-69]. But how can the typically more than 200 different cell types of a living organism have the same origin[70]? The main reason for the development of various cell types is gene regulation via orchestrated transcriptional programs, which can then be firmly established for example by the methylation of DNA[71-72].

1.3.1 DNA methylation in human diseases and cancer

DNA methylation is the addition of a methyl group (-CH_3) to a carbon atom of a base in DNA, usually the C-5 carbon of cytidine, linearly linked to a guanosine via a phosphate group (CpG)[30]. DNA methylation, is one of the key players in epigenetic silencing of transcription to balance and regulate cell biological processes[73] and thus, has a critical role on gene expression control[73-74]. Furthermore, it can affect cell behaviour and contribute to the aging process due to changes of the methylation pattern over time caused by internal, like maintenance and cell divisions, or external factors like ROS[30, 75]. It is catalysed in mammals by the DNA methyltransferases Dnmt1, Dnmt3A and Dnmt3B. Dnmt1 is abundant during the entrance of the S-phase, whereas Dnmt3A and 3B expression increase in the G0 or G1 to S-phase[73, 76-77]. Hereby, maintenance of existing

DNA methylation (see Figure **1-5**) is subject to Dnmt1, while Dnmt3A and Dnmt3B establish dynamic methylation patterns in early embryonic stages[73, 78] which was proven in mouse studies[78]. Dnmts catalyse the transfer of a methyl group from the cofactor *S*-adenosyl-methionine (SAM) on to the C5-position of dC (Figure **1-5**)[79] generating 5-methyl-2'-deoxycytidine (mdC).

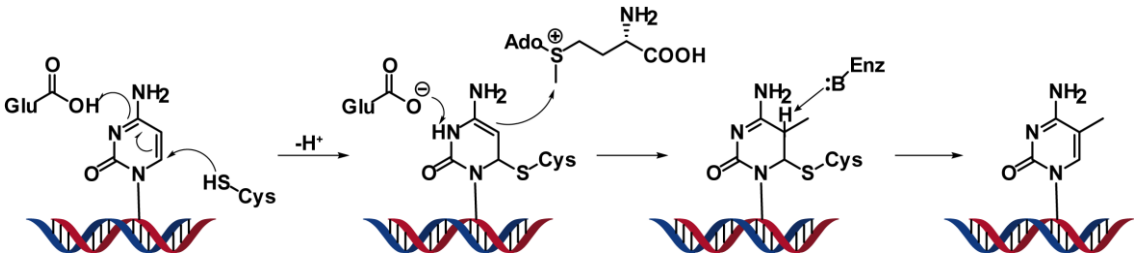


Figure 1-5: Depiction of the methylation mechanism of dC by Dnmt enzymes to mdC.

Inappropriate maintenance of DNA methylation, like hypermethylation in promoter regions, can result in an inactivation of tumor suppressor genes. Whereas, hypomethylation can induce genomic instability, which might result in development of cancer[80]. In somatic cells, 98% of DNA methylation occurs in CpG dinucleotides, where a dC is located at the 5'end of dG, whereas in embryonic stem cells (ESCs) 25% of the methylation occurs in non-CpG context[81-82]. Thus, mdC is the most common DNA methylated nucleoside with an abundance of 4 % of dC in the human genome[82-84]. Furthermore, mdC is considered to be a reversible epigenetic mark being essential for cellular differentiation, genome stability and development[85]. Thus, human diseases can occur when the epigenetic information mentioned above is not properly established or maintained.

1.3.2 Role of Tet enzymes in DNA methylation

The main function of ten-eleven translocation enzyme (*Tet*1 and *Tet*3), member of the 2-oxoglutarate (2OG)- and Fe(II)- dependent dioxygenases, is the oxidation of the methyl group[86-89] of mdC[89-93]. This well-known oxidation cascade of mdC to 5-hydroxymethyl-2'-deoxycytosine (hmdC), 5-formyl-2'-deoxycytosine (fdC), and 5-carboxyl-2'-deoxycytosine (cadC) is depicted in Figure **1-6**[94].

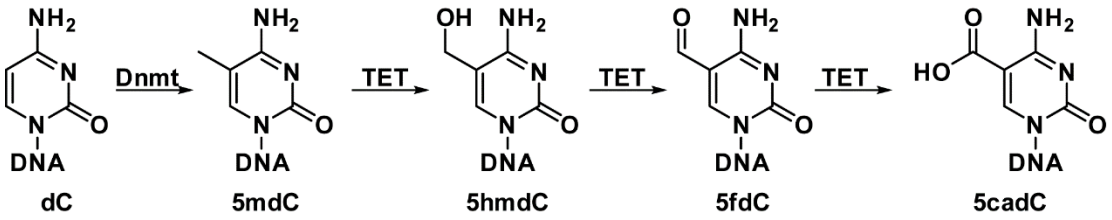


Figure 1-6: Depiction of the dC, mdC and its Tet induced oxidized derivatives by the active demethylation.

The relative amounts of these oxidized bases in stem cell DNA is decreasing from mdC (~4-6% of dC)[82-84] over hmdC (1-10% of mdC value, and up to 1.3% per dG in brain)[89, 95-98] to fdC (one hundredth of hmdC)[84, 98], and finally cadC, being about 1000-fold less abundant compared to hmdC[98]. Despite its low abundance, fdC has been shown to have a functional role in DNA beyond being a demethylation intermediate of mdC[99].

Since the discovery of these oxidized dC derivatives, a possible link between active demethylation of mdC and a possible reactivation of silenced genes has been investigated. In this context, mdC is removed directly and selectively from the genome via the two key molecules fdC and cadC[100]. This is done by thymine-DNA-glycosylase (Tdg), which cleaves the glycosidic bond and leaves an abasic site in the DNA strand[90, 101] (see Figure **1-7**). This process is described as the base excision repair (BER)[100-102]. The remaining 2'-deoxyribose unit in the strand is subsequently cut out via β-elimination causing a single-strand break[103], before the 5'end is cut by the human AP endonuclease (*APE1*). This step is followed by a closing of the gap with a new nucleoside at this position via polymerase and ligase[104].

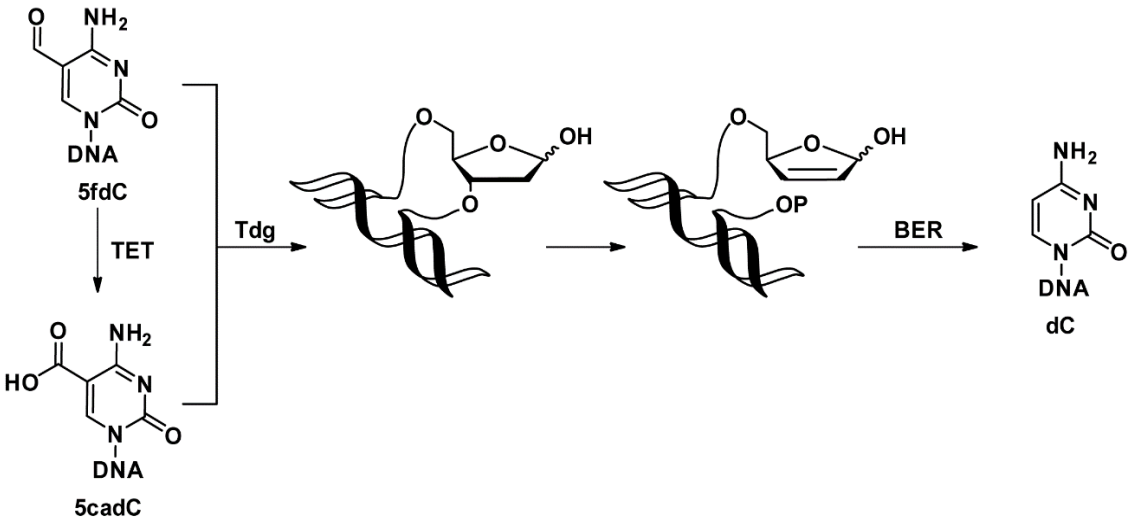


Figure 1-7: Depiction of the cleavage by Tdg and BER.

Next to the active demethylation the passive demethylation plays a crucial role (shown in Figure **1-8**). Here, inactivated methyltransferases in combination with several cell divisions cause the loss of mdC as an epigenetic information[105].

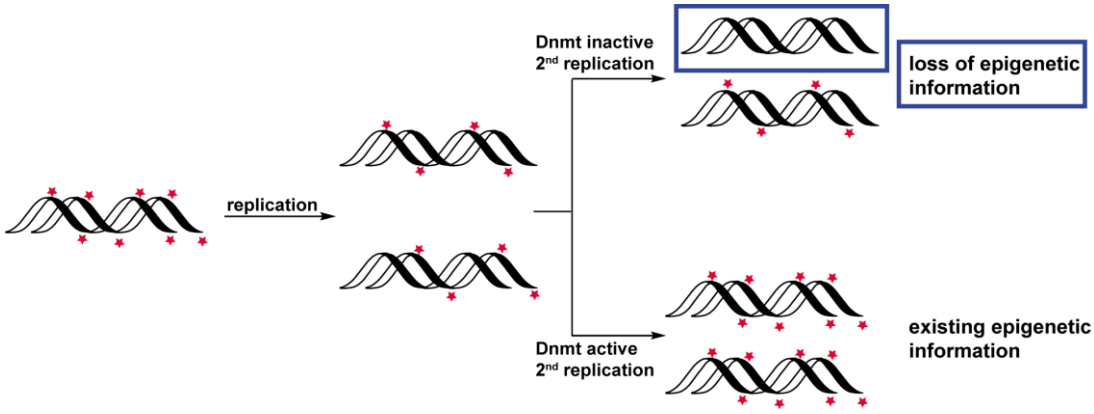


Figure 1-8: Depiction of the passive demethylation pathway. The red stars simulate the position of the mdC nucleosides.

An erroneous methylation, as mentioned above, by Dnmts causes a disruption of the epigenetic landscapes. Tet enzymes, removing these erroneous epigenetic marks, have emerged as an important tumor suppressor mechanism in cancer[106]. *Tet*2 is one of the most frequently mutated genes in cancerous tissues and cells[106] and its inactivation was shown to be relevant for disease development or acceleration and worsening of an existing malignancy or tumor development[106]. Further, missense mutations of isocitrate dehydrogenase (IDH1/2) play a crucial role in this process. They are responsible for the conversion of isocitrate into 2-hydroxy-glutarate (2-HG). 2-HG, in turn, is a competitive inhibitor of all α-KG-dependent dioxygenases, including TET enzymes, and consequently leads to TET inhibition and thus reduced hmdC levels[98, 107]. Since tumours are hypoxic and lack oxygen, the oxidation of mdC to hmdC, by Tet is reduced[98]. Thus, a sub-expression of TET proteins and lower hmdC levels are a general hallmark of many cancer types, like lung cancer[108-110].

1.3.3 Influence of ROS on DNA methylation

Beside its effects on genes and enzymes, ROS might also have a direct effect on the DNA methylation process. *Afanas'ev et al.*[111] postulated a possible mechanism of ROS induced DNA methylation. They described an attack of a superoxide molecule on the C-5 atom of cytidine resulting

in a deprotonation (see Figure **1-9**). This could lead to a nucleophilic attack on the positive charged *S*-adenosyl-methionine (SAM), generating the methylated cytidine[111-112].

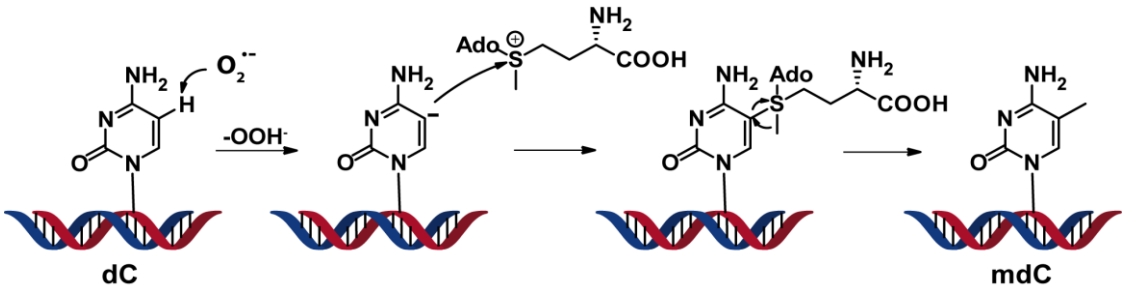


Figure 1-9: Proposal for a superoxide-dependent formation of mdC.

O'Hagan et al.[113] described a large complex containing Sirtuin1, Dnmts and polycomb repressive complex 4 (PRC4) increasing DNA methylation[30, 113] in areas with high levels of transcription and CpG. They postulated that 8-oxoguanine DNA glycosylase (*Ogg1*), the enzyme responsible for the excision of 8-oxoguanine[36], is responsible for the recruitment of this complex[30]. Since Ogg1 can be highly enriched in CpG rich areas due to ROS caused 8-oxo-dG formation, this might explain the changes of methylation patterns of hypermethylation in CpG rich areas and hypo-methylation in CpG poor areas[30]. Furthermore, 5-Hydroxymethylcytidine (hmdC) (see Figure **1-10**), the oxidation product of mdC, might also play a role here, since mdC could have been directly oxidated by ROS[30, 98], detected as incorrectly incorporated and thus excised by repair mechanisms[114-115].

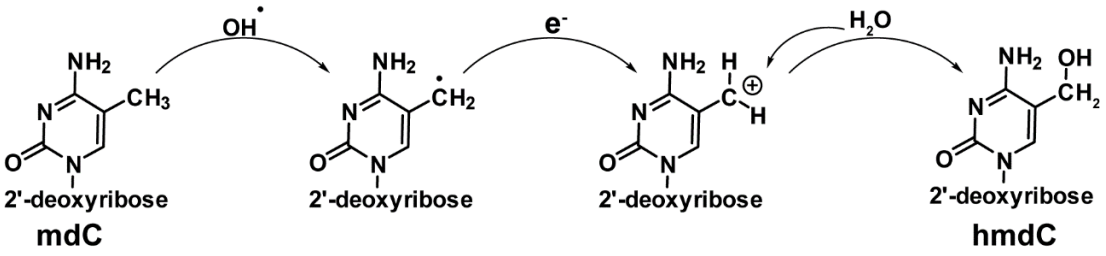


Figure 1-10: ROS induced hmdC formation.

It is also possible that the fdC levels in mammals can be ROS-caused, beside been enzymatically generated by Tet enzymes, which was proposed by *Carell et al.*[98] (see Figure **1-11**).

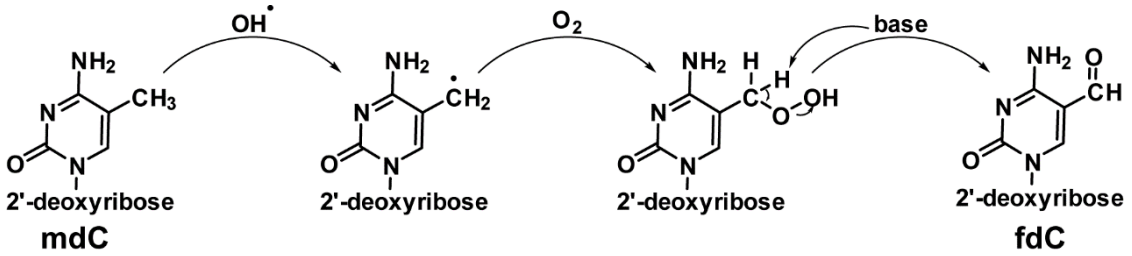


Figure 1-11: Proposed mechanism of ROS induced fdC formation beginning with mdC.

Thus, an interplay of ROS induced oxidation of mdC to hmdC or fdC and the Tet induced formation of these two molecules could overlap here, causing changes in the methylation pattern. It was shown, that anti-oxidant treatment of sperm of infertile men positively influenced the methylation pattern and reduced DNA damage[30, 116]. The oxidation product of dG affects the methylation of cytosine by Dnmts, when it is located close to cytosine moieties[30, 117-118]. An oxidized dG, on the other hand, does not influence the copying of methylation patterns of hemimethylated strands[117-118]. This might result in a hypomethylation in subsequent rounds of cell divisions, as it was depicted in Figure **1-8**[30].

1.3.4 Interplay of epigenetics and oxidative damage

Of the four canonical nucleotides, the guanine heterocycle is the electron richest, making dG the most suitable nucleotide for an oxidative modification[36, 119-120]. Further oxidative lesions of DNA bases are e.g. 8-oxo-deoxyadenosine (8-oxodA), N^6-hydroxymethyl-2'-deoxyadenosine (hm^6dA) and 5-formyl-2'-deoxyuridine (fdU) (Figure **1-12**).

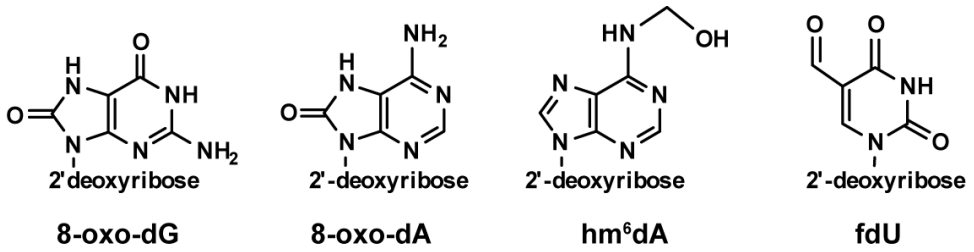


Figure 1-12: Known oxidative lesions of DNA.

8-oxo-dG is found at a constant level of approximately ~ 10^3 - 10^6 molecules per dG in the mammalian genome. The level, however, can vary upon oxidative or inflammatory stress, representing an indicator for generation and repair of DNA damage[121-126]. Whether or not 8-oxo-dG has in

addition to being a lesion also beneficial regulatory functions is currently investigated[36]. Therefore, a comparison with 5-methyl-2'-deoxycytidine (mdC), the most important epigenetic marker, is warranted. Both molecules are chemically stable and have strong Watson-Crick base pairing[36, 127]. 8-oxo-dG is favourably formed within GpG rich areas[36], while mdC is favourably formed in CpG rich areas by Dnmts[128-130]. 5-methyl-2'-deoxycytidine can be oxidized by Tet or ROS to hmdC and fdC, as shown in Figure **1-10** and **1-11**. 8-oxo-dG is the oxidation product of dG by a ROS induced process shown earlier in Figure **1-4**. Both nucleosides can cause base-mismatches and are removed from the genome by a well-known repair mechanism[36, 131-134]. After deamination of genomic mdC, mdC is read as thymidine (T) creating a T-dG mismatch. In contrast, oxidation of the dG in dG-dC base pairs, can result in dG to T transversion[135] based on an incorrect installation of a 2'-deoxythymidine:2'-deoxyadenosine opposite 8-oxo-dG[135-137]. This is caused by a rotation of the deoxyribose unit in the "anti" position by the C8 oxygen of 8-oxo-dG. Therefore, the "syn" conformation is populated, which in turn mimics T[136]. Removal of mdC and 8-oxo-dG requires the BER pathway[36, 131, 138]. While mdC is oxidized to 5-formyl-2'-deoxycytidine (fdC) or 5-carboxy2'-deoxycytidine (cadC) and then excised by thymine DNA glycosylase (*Tdg*)[98, 100, 103, 122], 8-oxo-guanine can be directly excised by O*gg1*[36] (see Figure **1-13**) before being catalytically replaced by the apurinic/apyrimidinic endonuclease (APE1) to return to the unmodified nucleobase dC or dG[48, 59, 139] (see Figure **1-13**).

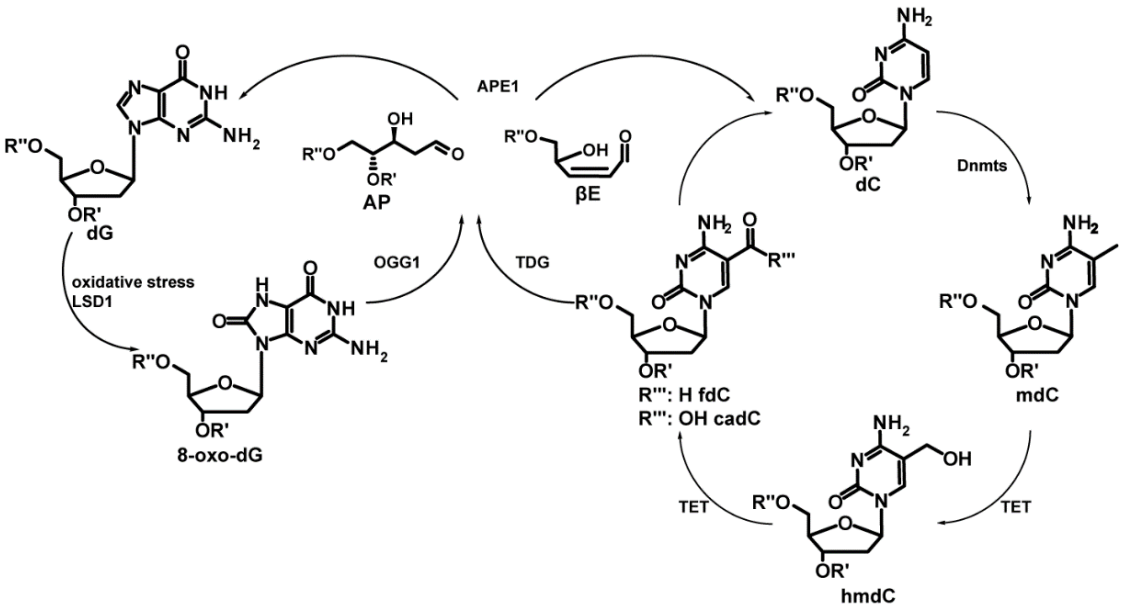


Figure 1-13: Depiction of the closely related oxidative cycles of 8-oxo-dG and mdC.

There are regions of putative dG-quadruplex sequences, which feature a high dG-density [36, 140], as depicted in Figure **1-14**[141].

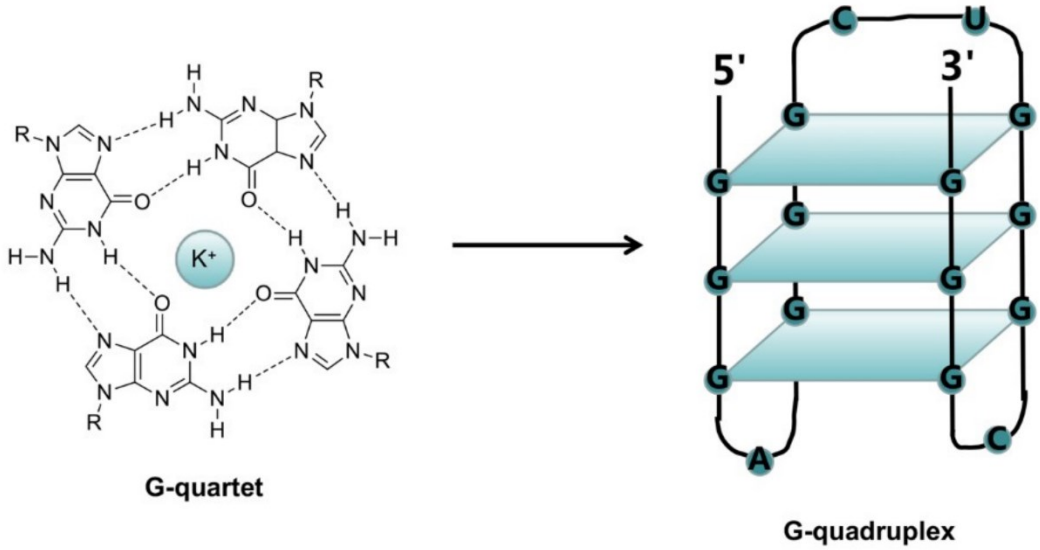


Figure 1-14: Depiction of the formation of G4 folds (quartet) and dG-quadruplexes in areas of high dG levels[141].

Nearly 10,000 of these dG quadruplexes were found in the human chromatin[140] and up to 700,000 in the human genome[142]. These quadruplexes can be targeted by different regulatory DNA-binding proteins, to guide and regulate processes like the synthesis of messenger RNA (mRNA)[36, 143-144] or were biased next to gene promoters to up and down regulate transcription[36]. Further, they are associated with hindered mdC writing into DNA, since Dnmt1 shows a higher binding affinity to these G4 folds compared to duplex- or single stranded DNA, resulting in a protective role of these G4 folds towards DNA methylation[36, 145]. It was shown by *Fleming et al.*[146] in case of an oxidative lesion in the G-quadruplex forming sequence of a promoter, that the presence of 8-oxo-dG can lead to an increase of ~300% in gene expression. *Bielskutė et al.*[147] suggested an stabilizing effect of 8-oxo-dG for G quadruplexes and a potential novel role of oxidative stress in gene transcription. Further, a study of *Ramon et al.[148]* showed a correlation of 8-oxo-dG induced expression of genes regulated by the transcription factor SP1 which is involved in cellular processes, like cell differentiation, cell growth, apoptosis, immune responses or carcinogenesis[149]. SP1 binds to GC-rich areas. They showed that an exchange of dG to 8-oxo-dG alters the expression of the genes regulated by SP1[148].

*Ogg*1, the repair enzyme of 8-oxodG, also seems to have a potential epigenetic function. Firstly, *Ogg*1 affects methylation levels*[113]*. Furthermore, *Ogg*1, bound to its substrate, can be coupled

to DNA allocation of NF-κB[150]. *Pan et al.*[150] described a model of an induced architectural change in DNA, caused by *Ogg*1 coupled to 8-oxo-dG, which modulates the NF-κB interface and therefore function in epigenetic regulation[150-151]. This is comparable to methyl-CpG-binding domains (MBDs) in the presence of mdC whose proteins are primary candidates for methylation readout, recruitment of chromatin remodellers, histone deacetylases and methylases[152].

Secondly, *Ogg*1$^{-/-}$ mice showed improved resistance to gastrointestinal infections[153], glucose intolerance[154] and lipogenesis in the liver[136]. On the contrary, an overexpression of *Ogg*1 in fibroblasts and pulmonary epithelial cells protects the cells of toxic effects. *Ogg*1 also shows to be relevant for certain diseases like Huntington's disease[155]. Here, experiments indicated that an inhibition of *Ogg*1 resulted in a decreased expansion of this disease. This might be caused by an accumulation of unprocessed repair intermediates in the brain[155].

An inhibition of *Ogg*1 and increased 8-oxo-dG levels could have beneficial effects, whereas an induction of *Ogg*1 forms AP-sites showing a toxic effect. These findings need to be further investigated. Still, little is known about the consequences of an up- or downregulation of *Ogg*1. Furthermore, 8-oxo-dG is a suitable and potentially very well measurable marker for oxidative damage, caused by internal and external stress, of all kinds. Yet, the basal levels of DNA modifications caused by oxidative stress, is still unknown and will be investigated in this thesis in a comparison with lung tissues and lung cancer.

1.3.5 Basics of mass spectrometry

Over the years clinical laboratories recognized mass spectrometry as a useful tool for the characterisation of many analytes in different matrices like blood, urine and tissues.[156] The technique is used for toxicology purposes, screening of hormones, proteins, enzymes and nucleosides in new-born or elderly patients.[156-157] Over the past few years the technique of mass spectrometry measurements changed quite a bit to a very sensitive instrument.[158] As mass spectrometry just allows the detection of the masses of the analytes, the analytes need to be separated before by gas chromatography (GC) or liquid chromatography (LC). The instrument of choice usually LC requires the molecules to be soluble in water or in special cases in organic solvents. For nucleosides the separation method of choice is LC. By using a reverse phase column with a gradient of water and acetonitrile with additives like formic acid or ammonia formiate[159] the nucleosides can be separated chromatographically. After separation, the molecules need to be transferred from

the liquid phase to the gas phase. This is done at the ion source, usually by an electrospray ionisation (ESI), in which the solvent is gently evaporated and the molecules are transferred as charged ions to the mass spectrometer (see Figure **1-15**)[160-161].

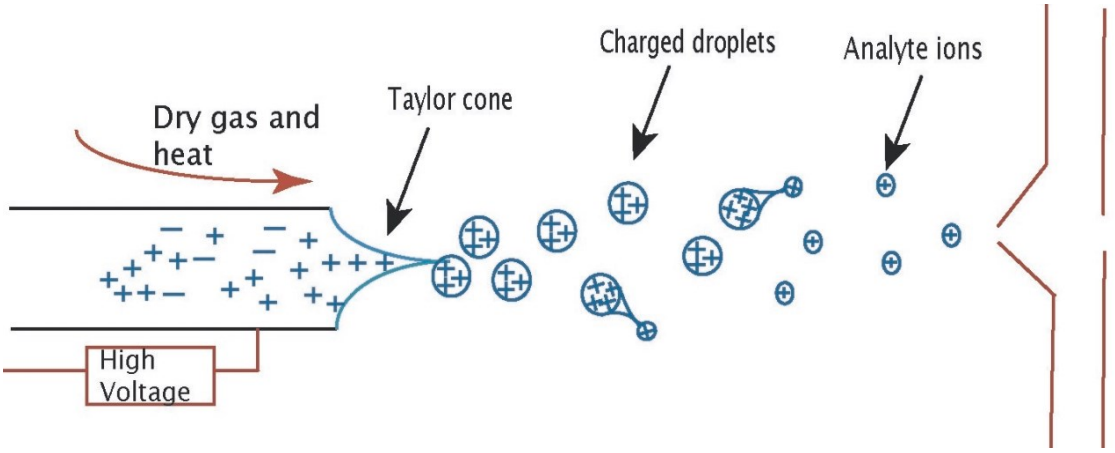


Figure 1-15: Depiction of an electrospray ionisation (ESI) source. By applying high voltage and drying gas the liquid droplets loaded with the analytes are evaporated. If the surface of the droplets is decreasing and the ions are getting closer, they repel each other (coulomb explosion), creating smaller droplets until they are present as individual ions.

In the mass spectrometer the masses of the charged ions are analysed by the mass analyzer. This mass analyzer mostly consists of a quadrupole. By changing the voltage between two opposite poles, the analytes oscillate with different amplitudes through the mass analyzer depending on their masses. Analytes with a divergent mass over charge ratio (m/z) collide with the quadrupoles and thus will not be detected[162]. In contrast, molecules with the correct m/z ratio can pass through and be detected.

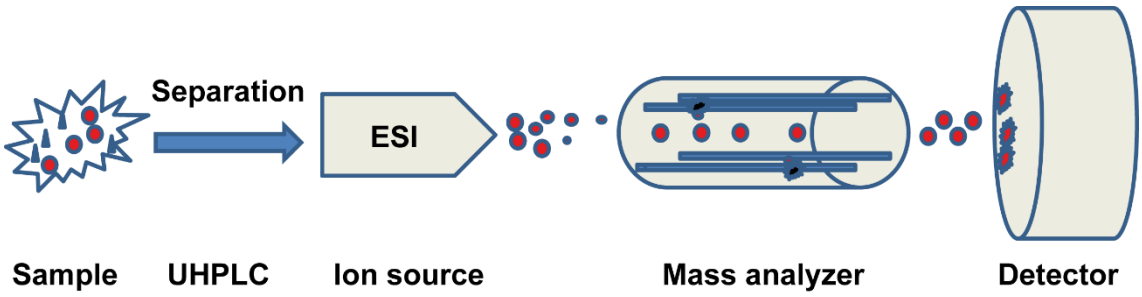


Figure 1-16: Schematic illustration of a simple mass spectrometer construction.

Figure **1-16** shows a very simple example of a mass spectrometer. In this study we used a triple quadrupole (TQ) system which has two mass analyzers connected by a collision cell (see Figure **1-17**[163]).

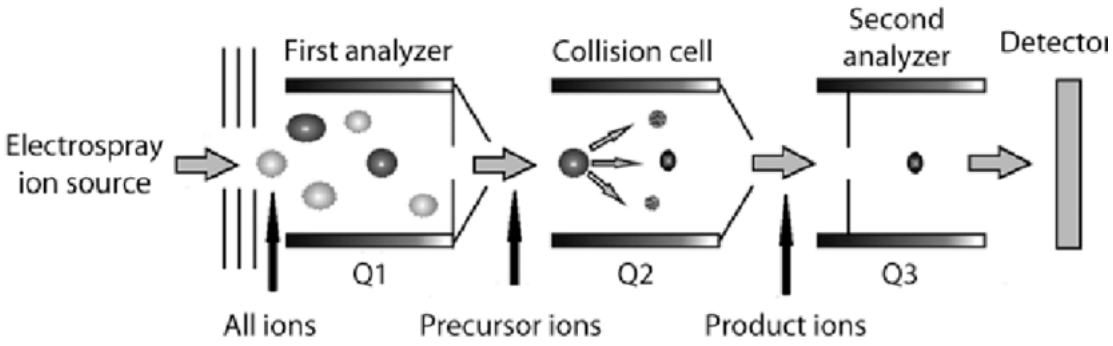


Figure 1-17: Depiction of a triple quadrupole system.

In a TQ system, so-called parent/precursor ions (molecular ions) are selected in a first analyzer and fragmented into daughter/product ions in the collision cell, by collision of the parent ions with nitrogen gas. For nucleosides, these are usually the two fragments of the base and the sugar residue. The mass transfer from parent to product ion is characteristic.[161] These product ions are then detected by a detector and converted into an electronic signal, which is converted to a chromatogram in the software of the computer.

2. Aim of this work

In the context of this work, several questions regarding DNA modifications and oxidative damage and their effects were addressed by mass spectrometry measurements. Therefore, this work is divided into different parts.

The first part of this work targeted the question about the origin of the different DNA modification levels. Until today, all collected data about the modification levels of mdC, hmdC, fdC and cadC, show the sum of epigenetically developed and potentially basal existing modifications. Therefore, one could ask whether basal levels of these known modifications exists. To investigate this question, an eukaryotic organism was needed, showing no relevant epigenetically active enzymes producing these modifications. This question was addressed in chapter **3.1** using the fly agaric, *Amanita muscaria*, an in nature grown and a phylogenetic distant eukaryotic organism.

Another part of this work addressed the deformylation pathway of fdC to dC. Since a direct C-C bond breakage of genomic fdC has been shown earlier[164], the question of the mechanisms behind it and the enzymes involved remains unanswered. It was planned to study these questions in cell culture experiments, in Chapter **3.2**, with the help of a new, synthetically produced 6-aza-2'-deoxycytidine analogue.

Another question that was addressed, focused at changes of epigenetic base levels between healthy and tumor tissue. Therefore, patient material from lung tumor patients was used and analysed by mass spectrometry, regarding the content of oxidative damage and the derivatives of dC (Chapter **4.1**).

Chapter **4-3** addressed the question how the direct impact of smoking like conditions influences the levels of DNA modifications. Since smoking is still one of the main causes of developing lung diseases, its impact on the different DNA modifications requires evaluation. Therefore, cell culture studies and mass spectrometry studies with cigarette smoke extracts were done.

Finally, the question of how artificial respiration influences oxidative DNA damage was investigated. Here, 8-oxo-dG plays a role as a good marker (Chapter **4.4**). It is already known that long-term ventilation of patients has strong side effects. The question if this happens also in the case of short-term ventilation, such as those performed with neonatal infant was investigated.

3. Published work

3.1 5-hydroxymethyl-, 5-formyl- and 5-carboxydeoxycytidines as oxidative lesions and epigenetic marks

Florian Schelter, [+,a] Angie Kirchner, [+,a,b] Franziska R. Traube,[a] Markus Müller,[a] Wolfgang Steglich,[a] and Thomas Carell* [a]

(+ shared first authors)

Prologue

With respect to the quantities of the four oxidized nucleotides 5-hydroxymethyl-2'-deoxycytidine (hmdC), 5-formyl-2'-deoxycytidine (fdC), 5-carboxy-2'-deoxycytidine (cadC) and 8-oxo-dG, little is known about their origin. It is unknown whether they are caused by ROS or epigenetic activities. So far, experimentally determined values are generated by oxidative damage and epigenetics. It is known that mdC is oxidized by ten-eleven translocation (Tet) enzymes with molecular oxygen, Fe (II) and α-ketoglutarate as co-substrates to hmdC, fdC and cadC. It would therefore be interesting to find an organism naturally lacking Tet enzymes.

The naturally grown mushroom *Amanita muscaria*, with its established fully sequenced genome[165], showed in data base searches no Tet homologues when compared to higher eukaryotes or other mushrooms like *Coprinopsis cinerea*[90]. It presents therefore a perfect organism to study the oxidative back genome levels of hmdC, fdC, cadC and 8-oxo-dG. Therefore, we performed an analysis of the modified nucleoside levels of the mushroom and compared it with HEK293T cells, exhibiting low Tet activity[166] and iNGN cells with high Tet activities[167-168].

Contributions of the authors

My part of this project involved the analysis of the biological samples by ultra-high performance-liquid chromatography in combination with mass spectrometry (UHPLC-MS/MS) and the evaluation of the resulting data. Therefore, I prepared the mushroom samples by washing, drying, cutting and grinding it. I further did the isolation and digestion of DNA and RNA of this mushroom (*Amanita muscaria*) and the UHPLC-MS/MS-based quantification of the nucleosides of interest. To exclude, that the levels of the non-canonical DNA nucleosides were biased by bacterial DNA contaminations, I performed control experiments to check for bacterial RNA modifications like i^6A in RNA extracted from the same samples. The derivatization of the abasic sites for the investigation of the Tdg repair mechanism was also done by me. The final quantification of the abasic sites, as well as the graphical presentation of the results were done by me, too. *Dr. Angie Kirchner* conducted the iNGN cell culture experiments and the analysis of the Tdg-studies in mESCs. *Dr. Franziska Traube* grew the HEK293T cells. *Dr. Markus Müller* did the database searches for the *Amanita muscaria* genome. *Prof. Dr. Thomas Carell* thankfully had the idea for this project and supervised it.

Article:

https://doi.org/10.1002/chem.202100551

Chemistry Europe
European Chemical Societies Publishing

5-Hydroxymethyl-, 5-Formyl- and 5-Carboxydeoxycytidines as Oxidative Lesions and Epigenetic Marks

Florian Schelter[+],[a] Angie Kirchner[+],[a, b] Franziska R. Traube,[a] Markus Müller,[a] Wolfgang Steglich,[a] and Thomas Carell*[a]

In memory of François Diederich

Abstract: The four non-canonical nucleotides in the human genome 5-methyl-, 5-hydroxymethyl-, 5-formyl- and 5-carboxydeoxycytidine (mdC, hmdC, fdC and cadC) form a second layer of epigenetic information that contributes to the regulation of gene expression. Formation of the oxidized nucleotides hmdC, fdC and cadC requires oxidation of mdC by ten-eleven translocation (Tet) enzymes that require oxygen, Fe(II) and α-ketoglutarate as cosubstrates. Although these oxidized forms of mdC are widespread in mammalian genomes, experimental evidence for their presence in fungi and plants is ambiguous. This vagueness is caused by the fact that these oxidized mdC derivatives are also formed as oxidative lesions, resulting in unclear basal levels that are likely to have no epigenetic function. Here, we report the xdC levels in the fungus *Amanita muscaria* in comparison to murine embryonic stem cells (mESCs), HEK cells and induced pluripotent stem cells (iPSCs), to obtain information about the basal levels of hmdC, fdC and cadC as DNA lesions in the genome.

The genetic system consists of four deoxyribonucleotides (dA, dC, dG and T), which create the sequence information that encodes the construction of proteins from amino acids. In higher organisms, four additional nucleotides are present that are all derived from 5-methyl-deoxycytidine (mdC). These four epigenetic nucleotides are mdC itself and the three derivatives 5-hydroxymethyl-, 5-formyl-, and 5-carboxydeoxycytidine (hmdC, fdC and cadC = xdC) formed by oxidation of mdC with the help of α-ketoglutarate and oxygen requiring Tet-enzymes.[1–4] These epigenetic nucleotides control the transcriptional activity of the genes in which they are embedded in a poorly understood way. It is however known, that the nucleotides fdC and cadC are substrates for the DNA repair glycosylase Tdg.[5–8] This enzyme is able to hydrolyse the glycosidic bond of fdC and cadC, thereby generating abasic (AP) sites. These AP sites may subsequently be processed to β-elimination sites (βE). Finally, AP and βE-sites are replaced by unmodified dC,[9] which ultimately leads to the conversion of the xdCs back to dCs (Figure 1).[10]

This whole process of dC methylation by dedicated DNA methyltransferases (Dnmt enzymes), Tet-induced oxidation to xdC and Tdg-initiated conversion of fdC and cadC to dC sites, creates a dynamic active methylation and demethylation circle depicted in Figure 1.[11–14]

Although this circle is now well established in mammalian genomes, experimental data about the levels of xdC bases in other higher eukaryotes such as plants and fungi are ambiguous. This suggests that at least in some of these species the detected xdC levels might be caused by just oxidative stress

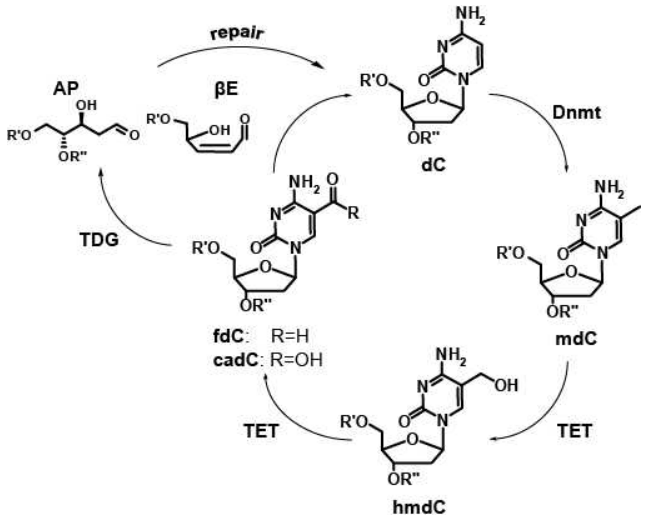


Figure 1. Depiction of the mdC oxidation cycle. Dnmt enzymes methylate dC to mdC, which is oxidized by Tet enzymes to give xdCs (hmdC + fdC + cadC). Tdg cleaves the glycosidic bond of fdC and cadC to create abasic sites (AP). Other bifunctional repair enzymes produce βE-sites. Both BER intermediates are further processed leading to an insertion of dC. (R′ and R″ = DNA).

[a] F. Schelter,[+] Dr. A. Kirchner,[+] Dr. F. R. Traube, Dr. M. Müller, Prof. Dr. W. Steglich, Prof. Dr. T. Carell
Ludwigs-Maximilian-Universität München
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.lmu.de

[b] Dr. A. Kirchner[+]
Cancer Research UK Cambridge Institute, Li Ka Shing Centre
University of Cambridge
Cambridge CB2 0RE (UK)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202100551

(ROS). In *Coprinopsis cinerea*,[15] Tet homologues were identified, but data about the genomic xdC levels were not reported. It is a general problem that the xdC levels formed by oxidative stress are not known. In this context, it is important that repair enzymes such as the thymidine-DNA-glycosylase (Tdg) are able to remove fdC and cadC from the genome. The basal oxidative stress levels depend consequently not only on the amount of ROS, but also on the repair activity.[16]

The availability of new ultra-high performance liquid chromatography (UHPLC)-MS^2 procedures in combination with synthetic isotope standards and reagents enable the quantification of AP and βE-sites,[9,17] and the parallel determination of the xdC levels and BER repair events. This now allows us to gain information about the ROS induced xdC levels.[18–19] We selected as a model organism the fungus *Amanita muscaria*,[20] which was collected from its natural environment. The methylation levels in fungi of the basidiomycota are diverse between 2–5% per dC.[21–22] Here, we measured for *Amanita muscaria* an mdC level of 3.3%, which is quite comparable to the levels in mammalian cell cultures[23] and tissues.[24–25] Furthermore, analysis of the annotated *Amanita* genome,[20] regarding the content of Tet enzymes also in comparison with *Coprinopsis cinerea*, for which a Tet enzyme was recently reported,[26] revealed no obvious Tet homologues making *Amanita muscaria* an ideal species to study the ROS induced levels of hmdC, fdC and cadC.

For the MS-based quantification of the non-canonical nucleotide levels in *Amanita muscaria*, we removed pieces from the stem and the cap and isolated the DNA material after bead milling of the cellular material using a reported procedure (Supporting Information).[27] The DNA was enzymatically digested into single nucleosides and analysed according to a protocol that we reported recently in detail (Supporting Information).[27–28] For exact quantification of the mdC and xdC levels, we added synthetic isotope standards of all bases as shown in Figure 2A. These function as internal references. Exact quantification was performed as recently described.[27] In brief, the obtained nucleoside mixture was separated by UHPLC and characterised by coupling to a triple quadrupole mass spectrometer Figure 2B. We quantified the amounts of the xdC derivatives and for assessment of oxidative stress the oxidative lesion 8oxodG, again with the synthetically prepared standards depicted in Figure 2A. We then performed the same experiment with human iPSCs (neurogenin1/2-inducible iPSCs, so called iNGNs)[29] and HEK 293T cells.

The obtained data depicted in Figure 2C show that the mdC levels measured in *Amanita muscaria* (cap and stem) are indeed comparable to those detected in HEK 293T and to data measured earlier in human and mouse tissues.[25]

In contrast to the rather similar mdC levels, however, we detected as expected for an organism with putatively no Tet expression very little hmdC. The exact quantification revealed an extremely low level of hmdC with 2.5×10^{-3} per dG. This is ten-fold lower compared to HEK 293T cells, which have measurable but very low Tet expression levels.[30] This strong reduction of the hmdC levels is interesting particularly because cell culture conditions tend to reduce mitochondrial content and oxidative phosphorylation. The hmdC value is furthermore

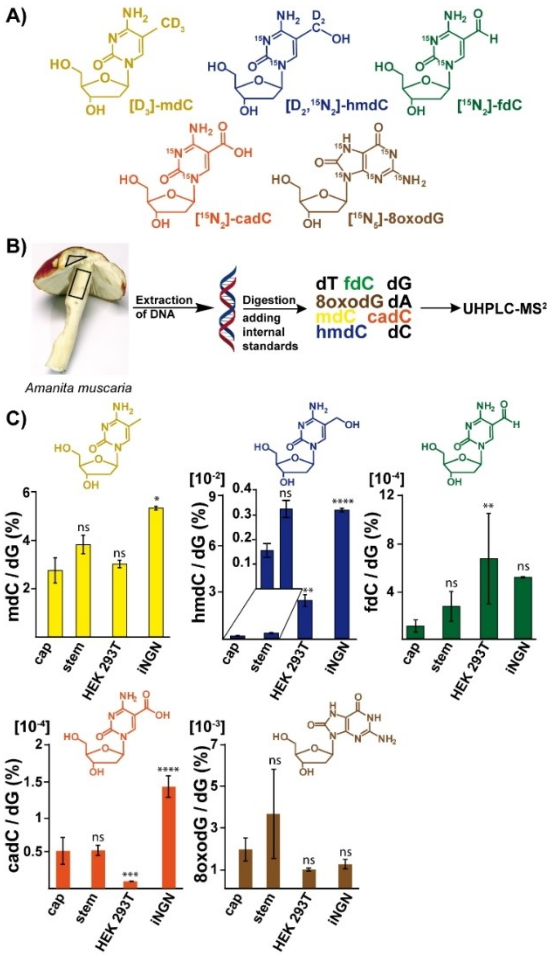


Figure 2. A) Depiction of the synthetic isotopically labelled nucleosides used as internal standards. B) Experimental procedure. C) Data about the absolute levels of mdC, xdC bases and the main oxidative lesion 8oxodG from six biological replicates. Bars show mean, error bars show standard deviation (s.d.). (Ordinary one way ANOVA with Tukey's multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

40-times lower compared to induced pluripotent stem cells (iNGN), also kept in culture, where TET-mediated epigenetic processes play a vital regulatory role.[31–32] We concluded that the low hmdC level in *Amanita muscaria* represents a level that we would consider to be a basal level of hmdC that is formed by oxidative stress under natural (wild life) conditions. In order to gain support for this idea, we next measured the 8oxodG levels. 8oxodG is an oxidative lesion derived from dG. Indeed, we found that the 8oxodG levels are similar to those of hmdC. Interestingly, we detected slightly more 8oxodG in the *Amanita muscaria* stem, where the hmdC was higher as well, showing a potential correlation between the 8oxodG, the hmdC levels and hence oxidative stress.

The comparison of the fdC and cadC levels in the fungus *Amanita muscaria*, in HEK 293T- and in iNGN cells resulted in a wide range of values. The fdC levels are lowest in *Amanita muscaria*, again in agreement with lacking epigenetic relevance. Regarding cadC, we detected the highest values in iNGNs, again supported by epigenetic activity. In general, however, the

differences are surprisingly small given that we compare data from an organism collected from nature with cells kept under controlled cell culture conditions, with likely reduced oxidative phosphorylation.

We next turned our attention to mESCs, where epigenetic oxidation of mdC is a well-studied process.[33] Here, we investigated three different cell lines (Figure 3). The first is a wildtype cell line containing biallelic copies of Tdg ($Tdg^{+/+}$). The second cell line is deficient in Tdg ($Tdg^{-/-}$) and the third mESC line contains a point mutation in the catalytic centre that encodes for a catalytically mutant Tdg (Tdg^{CM}) with impaired glycosylase activity.[8,34] This is a very important cell line, because a complete Tdg knockout could also affect protein complexes that contain Tdg as a component,[8,35–36] which could potentially influence the experimental outcome. These cell lines were generated by CRISPR-Cas9-mediated genome editing. A single-stranded repair template containing either mutations in exon 2 ($Tdg^{-/-}$) or exon 4 (Tdg^{CM}), respectively, was transiently transfected together with a plasmid containing the endonuclease Cas9 and the respective guide RNA. Earlier work could demonstrate that a D151A point mutation in the catalytic centre of Tdg preserves fdC binding by the glycosylase, but does not lead to base excision.[37] For the knockout, mutations causing a shift in the reading frame were predicted. The genomic sequence of the repair template was altered accordingly, and a restriction site was introduced or removed to identify targeted clonal populations (Supporting Information). Tdg knockout and catalytic mutant candidates were validated by Sanger sequencing and Western blotting (Supporting Information).

The exact quantification data of all mESC cell lines are depicted in Figure 3. The first observation is that the hmdC levels are similar in all three cell lines, arguing that all three cell lines have a comparable Tet activity, as expected. The second observation is that the hmdC levels are two orders of magnitude higher compared to *Amanita muscaria*. These high Tet-generated hmdC values support the epigenetic relevance of this non-canonical base in mESCs. In agreement with the idea that the glycosylase Tdg removes fdC and cadC, we see interesting differences in deficient and mutant Tdg mESCs. In fully repair competent $Tdg^{+/+}$ cells we see low fdC and cadC levels, but the fdC value is still ten times higher compared to *Amanita* showing that the repair process is unable to remove all fdC formed in mESCs. This finding is in full agreement with previous reports showing that fdC is a permanent or semi-permanent base that has a long life-time in specific genome regions.[5,38–39] Interestingly, we see that the $Tdg^{-/-}$ and Tdg^{CM} cells behave similar, arguing that disruption of Tdg-containing protein complexes has only a minor impact on the xdC levels. Both cell lines ($Tdg^{-/-}$ and Tdg^{CM}) have the expected elevated fdC and cadC levels due to lack of repair. The levels of oxidative damage, as measured by the 8oxodG levels, were in the mESC cultures lower than *Amanita muscaria* but comparable to the HEK 293T and iNGN data. This can be explained with the fact that *Amanita muscaria* is an organism that was collected from nature, while in culture, cells are kept under controlled laboratory conditions. The low 8oxodG levels but more strikingly the 10-times higher cadC levels in the $Tdg^{+/+}$ mESCs, compared to *Amanita muscaria*, let us conclude that the vast majority of the xdCs in mESCs are epigenetically formed.

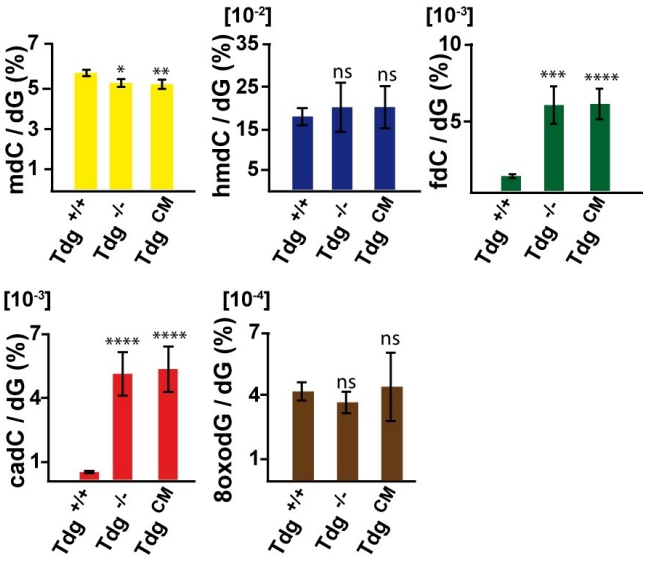


Figure 3. Depiction of the modified xdC levels in Tdg-proficient ($Tdg^{+/+}$), knockout ($Tdg^{-/-}$) and catalytic mutant (Tdg^{CM}, N151A) mESCs. Graphs show data from four biological replicates, bars represent mean, error bars represent s.d. (Ordinary one-way ANOVA with Dunnett's multiple comparison test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

Our data led us to assume that the xdC levels detected in *Amanita muscaria* and potentially also those seen in HEK 293T cells represent the levels that are derived exclusively from oxidative stress ($hmdC = 5 \times 10^{-6}$, $fdC = 5 \times 10^{-7}$ and $cadC = 1 \times 10^{-7}$). This leads to the conclusion, that in mESCs, the levels of hmdC are 100-times above the ROS damage level. The fdC value is about 10-times above the ROS damage level in mESCs. Interestingly, in mESCs ($Tdg^{+/+}$) the cadC level is reduced to the ROS level. The elevated levels of fdC and cadC in BER-deficient mESC cells ($Tdg^{-/-}$ and Tdg^{CM}) support the idea of a quick fdC and cadC repair by Tdg.

Interesting is the observation that in mESCs BER reduces cadC to the ROS level arguing that, if at all, only very few cadCs might have a permanent or at semi-permanent character as it was seen for fdC.[5,39]

In order to finally investigate if *Amanita muscaria* has a normal base excision repair process, we next quantified the AP and βE-site levels. To this end the isolated DNA was treated with the recently introduced AP- and βE-site detection reagent.[9] The hydroxylamine reagent reacts efficiently with the aldehyde functional group to give stable adducts, which after full digestion of the DNA, generate AP- and βE-adducts that can be quantified by UHPLC-MS^2.

For exact quantification we synthesized isotopically modified versions of the AP- and βE-site adducts (shown in Figure 4A), which were again used as internal standards. The reagent allows extremely sensitive detection, because it eliminates N_2 in the gas phase, which leads to a defined MS transition, which can be easily detected in MS/MS experiments

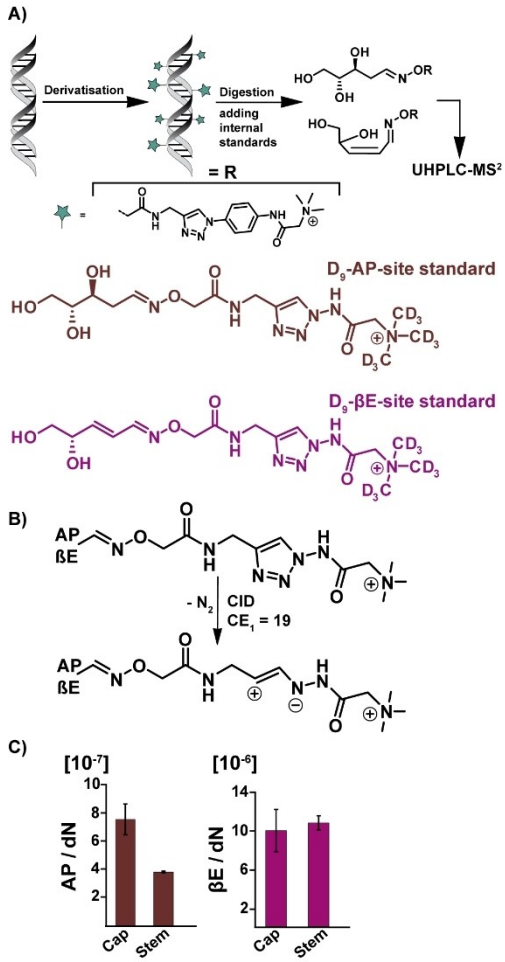


Figure 4. A) Workflow of the experiment used to quantify the AP- and βE-sites as BER intermediates. B) Depiction of fragmentation pattern, which gives a defined daughter ion after loss of N_2.) C) Depiction of absolute values of AP- and βE-sites in stem and cap of *Amanita muscaria*. Three biological replicates each measured as technical duplicate.

(Figure 4B).[9] The experimental workflow and the data are shown in Figure 4. With this method we quantified the AP-sites to be 4–8×10^{-7} and βE-sites to be around 1×10^{-5} per dN. These values are very comparable to previously published data from HEK 293T cells (8.8×10^{-7} AP- and 2.0×10^{-6} βE sites per dN),[9] supporting functional BER in *Amanita muscaria*. The measured basal ROS levels of the xdCs are consequently generated in a repair competent system.[9]

In summary, we used ultrasensitive UHPLC-MS^2 methods with synthetic internal isotope standards to quantify the levels of mdC, hmdC, fdC, cadC, 8oxodG, AP- and βE-sites in two systems that have no or low Tet activity (*Amanita muscaria* and HEK 293T cells) and compared the levels with data from cells that are epigenetically active at low (iNGN) and high (mESC) levels. We could deduce that the levels of hmdC = 5×10^{-6}, fdC = 5×10^{-7} and cadC = 1×10^{-7} are levels that can be considered to be the basal, steady state levels generated by oxidative stress and base excision repair. Epigenetic activity, created by Tet induced oxidation, increases the hmdC levels by a factor of 100 and the fdC value by a factor of 10. If repair is switched off as in $Tdg^{-/-}$ or Tdg^{CM} cells, the levels of fdC and cadC increase as expected by a factor of 5–10. Our data suggest that, while some of the fdCs have a permanent or semi-permanent character, because in mESCs even with full BER ($Tdg^{+/+}$) the levels never drop to the ROS level, this is questionable for cadCs. Removing fdC at these (semi) permanent sites then requires Tdg-independent pathways that were recently investigated in detail.[19]

Acknowledgements

We thank Kerstin Kurz and Claudia Scherübl for experimental help. Further we would like to thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via GRK2338 (Project ID 321812289), via CRC1309 (Project ID 325871075) and CRC1361 (Project ID 893547839). This project has received additional funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 741912 (EpiR) and Marie Skłodowska-Curie grant agreement No 765266 (LightDyNAmics). We are grateful for funding from the Volkswagen Foundation. Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: DNA repair · DNA modification · epigenetics · mass spectrometry · oxidative lesion

[1] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468; *Angew. Chem.* **2011**, *123*, 6588–6596.
[2] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[3] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. 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.
[4] 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.
[5] 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.
[6] 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.
[7] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.
[8] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67–79.
[9] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359–10364.
[10] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674–683.

[11] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594; *Angew. Chem.* **2020**, *132*, 5639–5643.
[12] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581.
[13] A. Jeltsch, *ChemBioChem* **2002**, *3*, 382–382.
[14] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2916–2921.
[15] L. M. Iyer, D. Zhang, R. F. de Souza, P. J. Pukkila, A. Rao, L. Aravind, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1676–1683.
[16] S. Ito, I. Kuraoka, *DNA Repair* **2015**, *32*, 52–57.
[17] A. C. Drohat, C. T. Coey, *Chem. Rev.* **2016**, *116*, 12711–12729.
[18] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312; *Angew. Chem.* **2018**, *130*, 4377–4394.
[19] F. Spada, S. Schiffers, A. Kirchner, Y. Zhang, G. Arista, O. Kosmatchev, E. Korytiakova, R. Rahimoff, C. Ebert, T. Carell, *Nat. Chem. Biol.* **2020**, *16*, 1411–1419.
[20] A. Kohler, A. Kuo, L. G. Nagy, E. Morin, K. W. Barry, F. Buscot, B. Canbäck, C. Choi, N. Cichocki, A. Clum, J. Colpaert, A. Copeland, M. D. Costa, J. Doré, D. Floudas, G. Gay, M. Girlanda, B. Henrissat, S. Herrmann, J. Hess, N. Högberg, T. Johansson, H.-R. Khouja, K. LaButti, U. Lahrmann, A. Levasseur, E. A. Lindquist, A. Lipzen, R. Marmeisse, E. Martino, C. Murat, C. Y. Ngan, U. Nehls, J. M. Plett, A. Pringle, R. A. Ohm, S. Perotto, M. Peter, R. Riley, F. Rineau, J. Ruytinx, A. Salamov, F. Shah, H. Sun, M. Tarkka, A. Tritt, C. Veneault-Fourrey, A. Zuccaro, A. Tunlid, I. V. Grigoriev, D. S. Hibbett, F. Martin, C. Mycorrhizal Genomics Initiative, *Nat. Genet.* **2015**, *47*, 410–415.
[21] T. Binz, N. D'Mello, P. A. Horgen, *Mycologia* **1998**, *90*, 785–790.
[22] A. Zemach, I. E. McDaniel, P. Silva, D. Zilberman, *Science* **2010**, *328*, 916.
[23] H. G. Leitch, K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, P. Hajkova, *Nat. Struct. Mol. Biol.* **2013**, *20*, 311–316.
[24] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
[25] 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.
[26] 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.
[27] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283–312.
[28] F. Yuan, Y. Bi, J.-Y. Zhang, Y.-L. Zhou, X.-X. Zhang, C.-X. Song, *RSC Adv.* **2019**, *9*, 29010–29014.
[29] V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, *Mol. Syst. Biol.* **2014**, *10*, 760.
[30] M. Y. Liu, J. E. DeNizio, R. M. Kohli, *Methods in Enzymology, Vol. 573*, Ed.: R. Marmorstein, Academic Press, **2016**, pp. 365–385.
[31] K. Hochedlinger, R. Jaenisch, *Cold Spring Harbor Perspect. Biol.* **2015**, *7*.
[32] Y. Gao, J. Chen, K. Li, T. Wu, B. Huang, W. Liu, X. Kou, Y. Zhang, H. Huang, Y. Jiang, C. Yao, X. Liu, Z. Lu, Z. Xu, L. Kang, J. Chen, H. Wang, T. Cai, S. Gao, *Cell Stem Cell* **2013**, *12*, 453–469.
[33] Y. Atlasi, H. G. Stunnenberg, *Nat. Rev. Genet.* **2017**, *18*, 643–658.
[34] 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.
[35] D. Cortázar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, P. Schär, *Nature* **2011**, *470*, 419–423.
[36] H. M. Hassan, B. Kolendowski, M. Isovic, K. Bose, H. J. Dranse, A. V. Sampaio, T. M. Underhill, J. Torchia, *Cell Rep.* **2017**, *19*, 1685–1697.
[37] L. S. Pidugu, Q. Dai, S. S. Malik, E. Pozharski, A. C. Drohat, *J. Am. Chem. Soc.* **2019**, *141*, 18851–18861.
[38] E.-A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.
[39] 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.

Manuscript received: February 12, 2021
Accepted manuscript online: March 26, 2021
Version of record online: May 1, 2021

3.2 Analysis of an active deformylation mechanism of 5-formyldeoxycytidine (fdC) in stem cells

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

([#] shared first authors)

Prologue

Although earlier studies by *Iwan et al.*[164] suggested a direct deformylation of fdC to dC, the underlying deformylation mechanism *in vivo* was still unknown. It was shown in *in vitro* studies by *Schiesser et al.*[102] for cadC containing oligonucleotides, that an interaction of cysteine and alkaline amino acids like histidine or arginine decarboxylates cadC in synthetic oligonucleotides. They furthermore showed that saturating the C5-C6 double bond greatly accelerates this process[169]. Therefore, a nucleophilic attack on the C6-position of the pyrimidine ring of fdC was claimed to be essential for active deformylation.

To prove this hypothesis, an fdC analogon was synthesized. A carbon to nitrogen exchange at the C6 position yielded 6-aza-5-formyl-2´-deoxycytidine (a-fdC), whose nitrogen atom, with its free electron pair, blocks a nucleophilic attack at the 6-position of the pyrimidine ring. A parallel cell culture feeding experiment with 2'-fluorinated fdC analogue (F-fdC*)* and a-fdC showed that F-fdC has strong deformylation activities *in vivo* to form 2'-fluoro-deoxycytidine (F-dC) whereas a-fdC was not deformylated to 6-aza-deoxycytidine (a-dC) *in vivo*. Hence a nucleophilic attack on the 6-position of fdC appears necessary for the *in vivo* deformylation to deoxycytidine (dC).

Contributions of the authors

My part of this project was besides the isolation and digestion of the DNA of fed stem cells also the development of a novel UHPLC-MS/MS method for the detection and quantification of synthetically modified DNA nucleosides. Furthermore, I did the evaluation and interpretation of the resulting data. Together with *Dr. Alexander Schön*, who did the synthesis work for this publication, I developed a protocol for the sensitive and reliable quantification of the nucleoside aza-5-

formyldeoxycytidine (a-fdC). Here, derivatization was needed to improve the detection of a-fdC in UHPLC-MS/MS measurements which showed a difficult to evaluate mass spectrometric signal. The needed calibration curves of the derivatized a-fdC were done by me. An important step for the analysis of the DNA samples, was the splitting of the samples in equal amounts before derivatisation. This was necessary since the derivatisation for a-fdC was quantitative, whereas fdC and 2'-Fluoro-formylcytidine (F-fdC) did not result in a quantitative conversion. This ensured an accurate quantification of all three nucleosides fdC, a-fdC and F-fdC. Furthermore, it was my idea to change the UHPLC column material from C8 to C18. This was necessary because the signals of derivatised fdC and a-fdC had the same retention time on a C8 column, and their mass difference of m/z 1 is not distinguishable on a triple quadrupole mass spectrometer. The final quantification of these nucleosides, as well as the graphical presentation of the results was done by me, too. *Ewelina Kaminska* did the cell feeding studies. *Eveliina Ponkkonen* performed the bisulfite studies to investigate the reactivities of fdC, F-fdC and a-fdC. *Eva Korytiaková* synthesized F-fdC. *Dr. Sarah Schiffers* did a first method development for the analysis of the 6-aza-nucleosides.

Article:

https://onlinelibrary.wiley.com/doi/full/10.1002/anie.202000414

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.

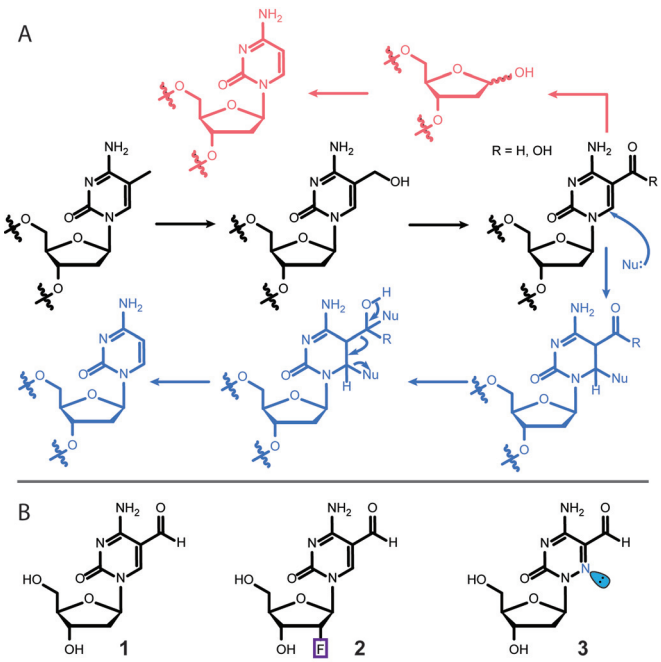


Figure 1. A) The mdC removal pathways that involve oxidation to hmdC, fdC, and cadC followed by either base-excision repair (magenta) or C–C bond cleavage (blue). B) Structures of fdC (**1**) and the two probe molecules **2** and **3** used for this study.

[*] 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

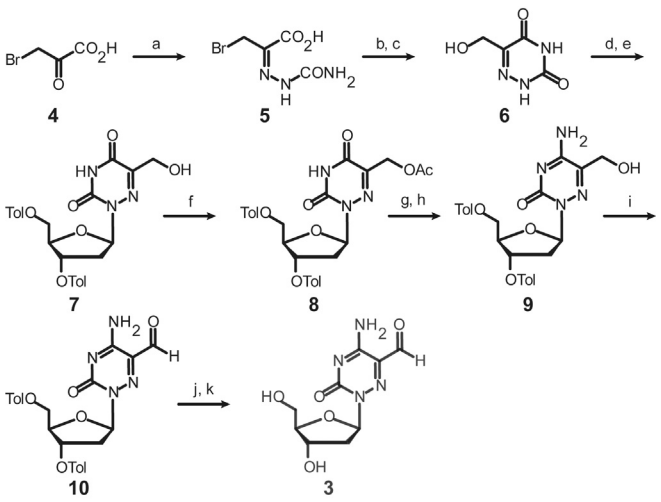
[+] These authors contributed equally to this work.

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).

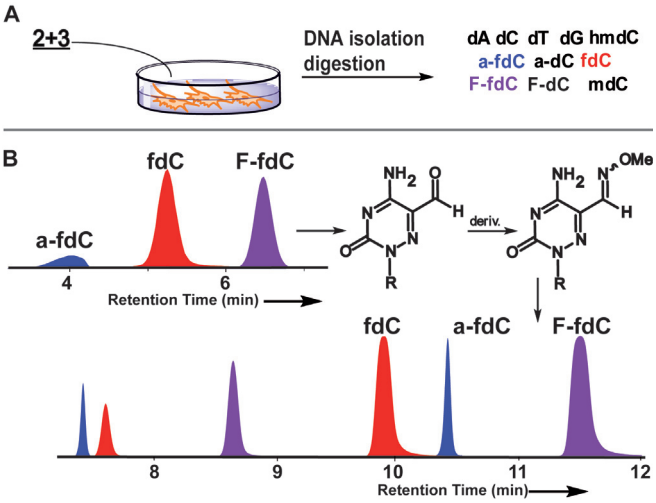


Figure 2. A) Overview of the experimental steps with the feeding and analysis. B) Analysis scheme and the reaction of a-fdC with methoxyamine to block hydrate formation and of a typical UHPL-chromatogram before (C-8 column) and after derivatization (C-18 column) for exact quantification. Peak splitting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

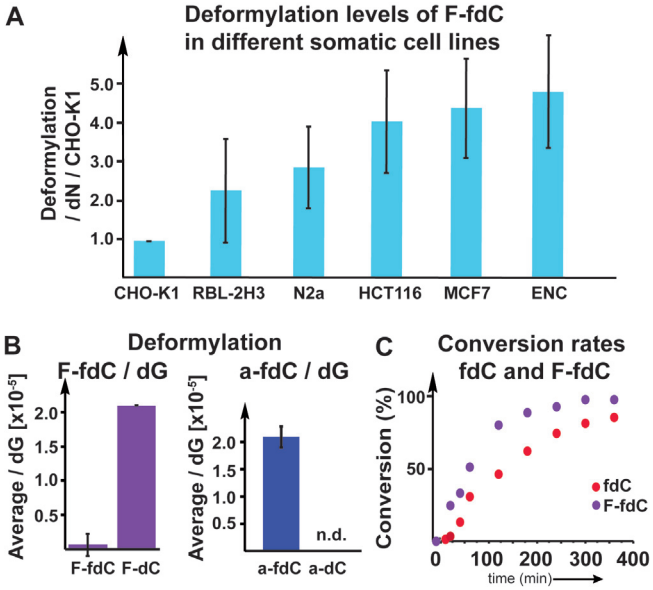


Figure 3. A) Deformylation data for F-fdC in different cell types, showing that F-fdC is deformylated in very different cells. Deformylation rate was calculated by the F-dC + F-mdC/dN per F-fdC/dN, then the values were normalized to the cell line with the lowest deformylation level (CHO-K1 = 1). B) The 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

4. Unpublished work

4.1 Analysis of lung tissue DNA

In the following, we were interested in the DNA modification levels of cancer tissues compared to healthy tissues. In particular, we wanted to determine the background levels of these modifications in tissues resulting from non-enzymatic oxidative processes. To this end, the DNA modification levels of the mushroom *Amanita muscaria,* which has no TET-enzymes, were taken as reference values.

4.1.1 Results and discussion of the mass spectrometric evaluation of the lung tissues

We started our investigation by comparing the DNA modification levels of mdC, hmdC, fdC, cadC and 8-oxo-dG. For this reason, the DNA of the samples was isolated using standard methods, enzymatically digested to the nucleoside level and analysed by UHPLC-MS/MS. In order to classify the order of magnitude of the epigenetic activity in healthy and tumor tissue, the values of the mushroom *Amanita muscaria* (fly agaric), from an earlier publication[170], were included in the following graphical representations and shown together with results of these measurements in the following figures.

Figure **4-1** shows the levels of the highly abundant oxidative DNA lesion 8-oxo-dG. The graph shows that the values of oxidative damage are in a similar range in all sample sources tested. This could mean that the cells of all tissues investigated here have undergone similar levels of oxidative stress. Since this is rather unlikely, we considered that the oxidative damage occurring to DNA during the extraction process is very high in comparison to the natural background, despite all efforts to prevent it *via* the proper use of antioxidants. This would level out the natural differences in the 8-oxo-dG-levels. The oxidation products of N^1-methyl-2'deoxyadenosine, 6-hydroxymethyl-2'-deoxyadenosine (N^6-hmdA) and 8-oxo-2′-deoxyadenosine (8-oxo-dA), could be other suitable markers for oxidative stress to cells. Their levels should be analysed in future experiments. Another possibility would be spiking of heavy labelled dG and quantifying the oxidation product of the heavy labelled 8-oxo-dG.

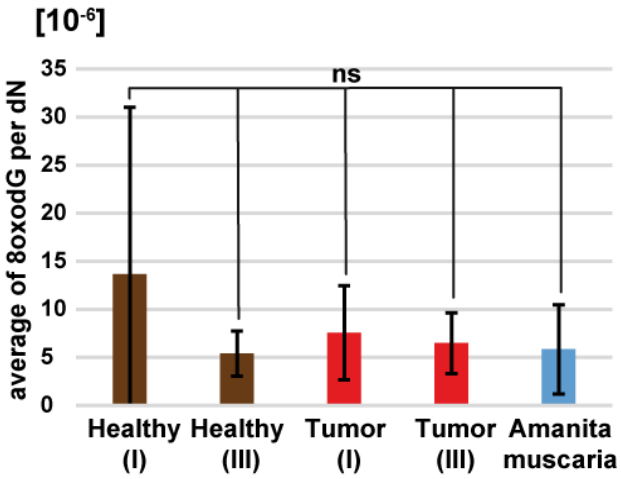


Figure 4-1: Comparison of the 8-oxo-dG levels in healthy tissue (brown), cancer stadium (tumor) I and III (red) and *Amanita muscaria* (light blue). Graphs show data for healthy and cancer tissues from 12 different patients and for *Amanita muscaria* from six biological samples. Bars show mean, error bars show standard deviation (s.d.). (Ordinary one-way ANOVA with Tukey's multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

Figure **4-2** presents the levels of mdC. It shows comparable levels of mdC in all different biological samples. This makes the mushroom *Amanita muscaria,* a phylogenetically distant eukaryote, a useful system for an epigenetic comparison. It further confirms the finding of the recent publication by *Schelter et al.*[170] that *Amanita muscaria* features similarly high mdC levels as human tissues.

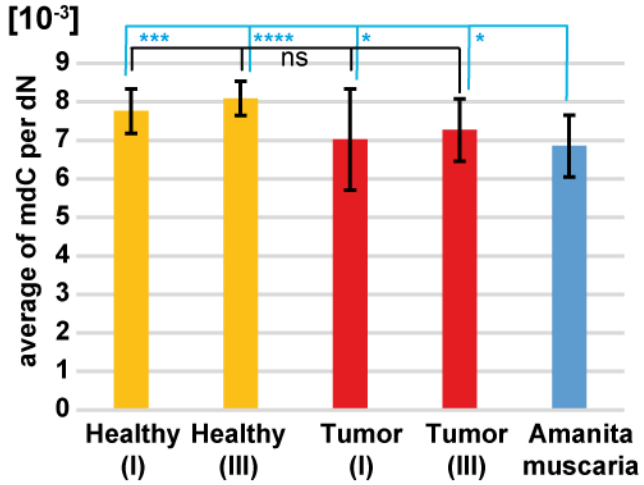


Figure 4-2: Comparison of the mdC levels in healthy tissue (yellow), cancer stadium (tumor) I and III (red) and *Amanita muscaria* (blue). Graphs show data for healthy and cancer tissues from 12 different patients and for *Amanita muscaria* from six biological samples. Bars show mean, error bars show standard deviation (s.d.). (Ordinary one-way ANOVA with Tukey's multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

An analysis of the hmdC levels (Figure **4-3)** shows large deviations between the biological samples. A closer look onto Figure **4-3,** shows a reduction by half of the hmdC levels of healthy tissues (2.9 x 10^{-4} per dN) vs. tumor tissues (1.2 x 10^{-4} per dN). This is a well-known phenomenon in cancerous tissue[171], since they are hypoxic and therefore show reduced activity of the O_2-dependent Tet enzyme in most of the cases[122, 172]. Remarkable is the hmdC value of *Amanita muscaria* (5.24 x 10^{-6} per dN) compared to human healthy lung tissue (nearly 55-times less) and tumor tissue (~ 25-times less). These values show that despite the hypoxic conditions, Tet activity in tumor tissues is still present, but is greatly reduced compared to the healthy tissues. In further studies, an analysis of the Tet expression levels of healthy and tumor tissue should be performed to directly correlate them with the obtained data.

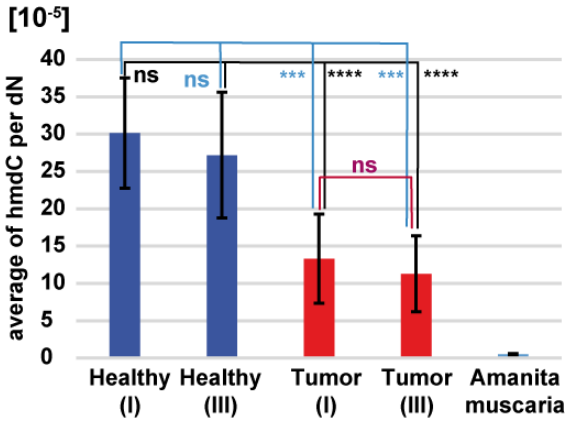


Figure 4-3: Comparison of the hmdC levels in healthy tissue (dark-blue), cancer stadium (tumor) I and III (red) and *Amanita muscaria* (light blue). Graphs show data for healthy and cancer tissues from 12 different patients and for *Amanita muscaria* from six biological samples. Bars show mean, error bars show standard deviation (s.d.). (Brown-Forsythe and Welch ANOVA with Dunnett's T3 multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

A quantification of the fdC level yielded the results shown in Figure **4-4**. The data show that the fdC levels are similar for all investigated biological samples, including *Amanita muscaria*. Since *Amanita muscaria* does not possess TET enzymes, the fdC levels represented by the blue bar is generated purely by ROS. The here presented data show, that, the fdC levels of healthy and tumor human tissues are comparable to those found in *Amanita muscaria*. This suggest that the fdC-data represent ROS-caused basal levels of this modified base in human tissues. The data

suggest that in healthy and tumor tissues the Tet-activity is limited to the production of hmdC.

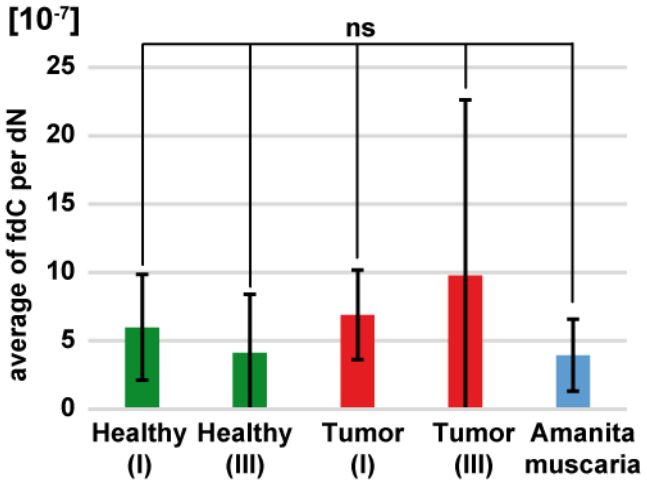


Figure 4-4: Comparison of the fdC levels in healthy tissue (dark green), cancer stadium (tumor) I and III (red) and *Amanita muscaria* (light blue). Graphs show data for healthy and cancer tissues from 12 different patients and for *Amanita muscaria* from six biological samples. Bars show mean, error bars show standard deviation (s.d.). (Ordinary one-way ANOVA with Tukey's multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

The same applies to the cadC levels shown in Figure **4-5**. However, the cadC values of the lung samples are three times higher compared to *Amanita muscaria*. This could mean, that the measured cadC values are mostly of enzymatic origin, whereas the fdC values indicated ROS induced basal levels. *Coprinopsis cinerea* showed faster conversions of hmdC to fdC than fdC to cadC which would result in comparable fdC values (Fig. **4-4**) but reduced cadC levels (Fig. **4-5**)[173].

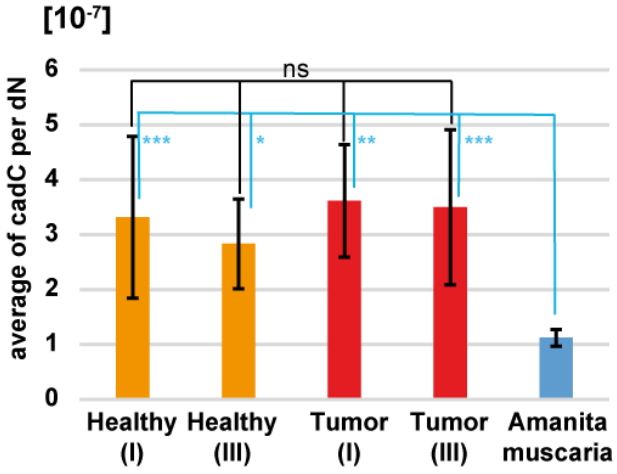


Figure 4-5: Comparison of the cadC levels in healthy tissue (orange), cancer stadium (tumor) I and III (red) and *Amanita muscaria* (light blue). Graphs show data for healthy and cancer tissues from 12 different patients and for *Amanita muscaria* from six biological samples. Bars show mean, error bars show standard deviation (s.d.). (Ordinary one-way ANOVA with Tukey's multiple comparisons test: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

Our data shows the opposite, where the cadC levels are still increased compared to *Amanita muscaria*. Finally, one can see that the values for tumor tissues are slightly elevated compared to the healthy ones.

All together the data show the levels of the different oxidized mdC derivatives. The nucleosides fdC and cadC, hardly show any difference regarding the amounts between tumor stage I and III. But both stages show the same change compared to the healthy tissue. Our study shows, in line with previously published studies[174], that especially hmdC, and in some cases also mdC, can be used as markers of tumor stage, whereas this is not the case for fdC and cadC and 8-oxo-dG. Until now, the presented data of the quantification studies do not allow to distinguish tumor stages.

4.2 Analysis of lung tissue RNA

Next to the levels of modified DNA bases, the levels of non-canonical RNA nucleosides could potentially also function as diagnostic tools. RNA has a similar structure to DNA but differs in some points. RNA is mostly a single-stranded molecule, with a ribose “backbone” and an exchange of thymidine to uracil. The most important RNAs are the messenger RNA (mRNA) which transports the information of the DNA to the ribosome. Beside the mRNA, transfer RNA (tRNA) and ribosomal RNA (rRNA) play an important role in protein biosynthesis due to the translation of mRNA to the protein level[175]. In the following, we were interested in potential differences of RNA modification levels in human tumor tissues compared to healthy tissues.

4.2.1 Results and discussion of the mass spectrometric evaluation of the lung tissues

Figure **4-6** shows the chemical structures of the very abundant rRNA and tRNA modifications, 2'O-methyl-adenosine (Am), 2'O-methyl-guanosine (Gm), 2'O-methyl-cytidine (Cm) and pseudouridine (Ψ). The analysis of the levels of these modified RNA nucleosides yielded the results shown in Figure **4-7**.

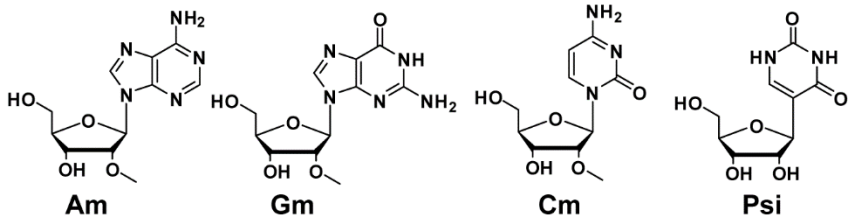


Figure 4-6: Most abundant RNA modifications.

The obtained values of the nucleosides, depicted in Figure **4-6**, show no differences between healthy and tumor tissue.

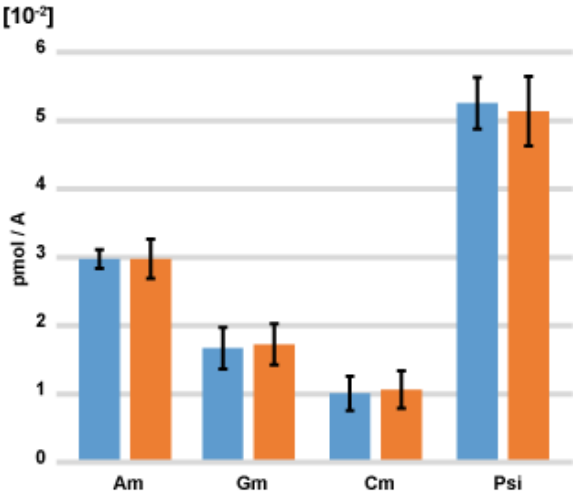


Figure 4-7: Depiction of the RNA modified nucleosides Am, Gm, Cm and Ψ in healthy (light blue) and tumor tissues (orange). Graphs show data for healthy and cancer tissues from 12 different patients measured in technical triplicates.

Figure **4-8** shows the less abundant modifications measured in a second quantitative MS-experiment.

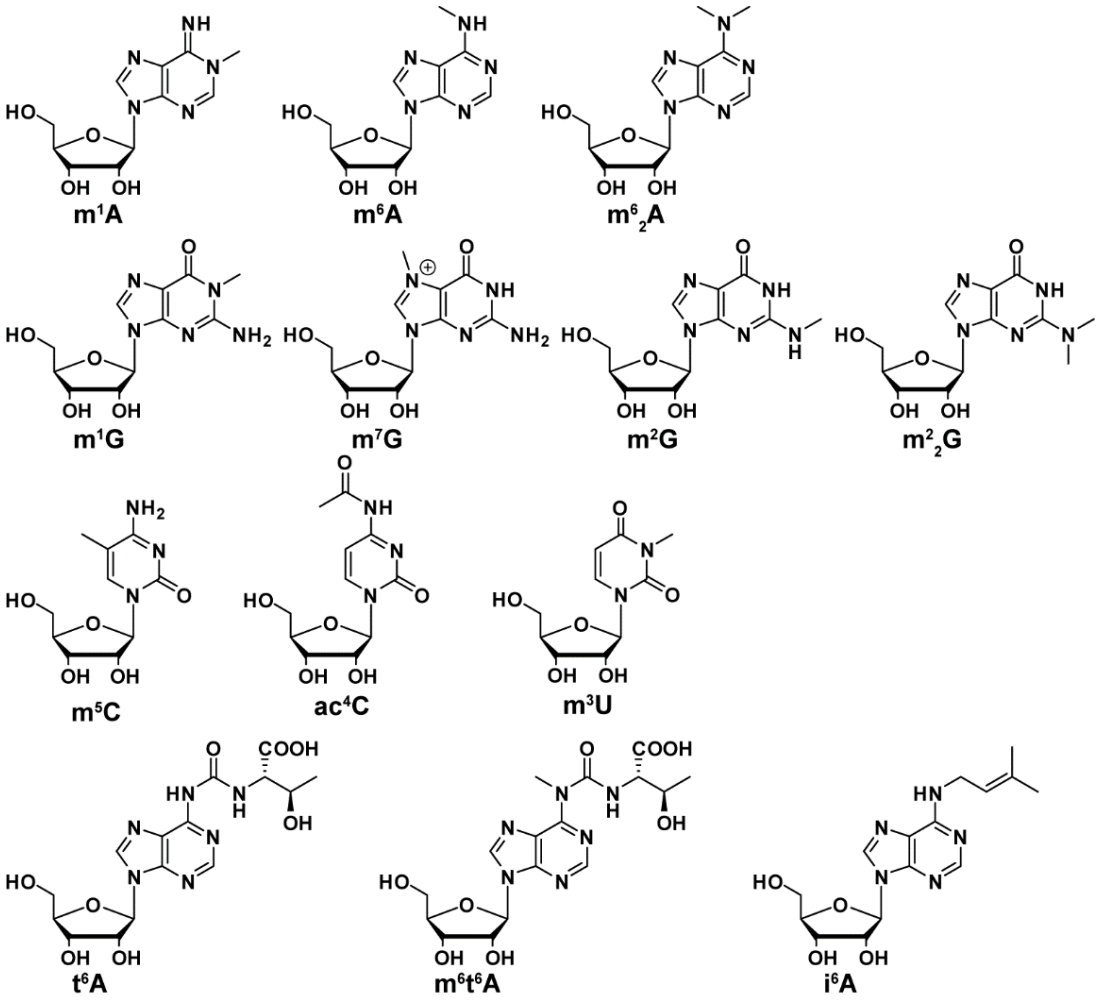


Figure 4-8: Less abundant RNA modifications.

Figure **4-9** summarizes the data about, the levels of all the modifications shown in Figure **4-8**: N^1-methyladenosine (m^1A, rRNA/tRNA), N^6-methyladenosine (m^6A, rRNA/tRNA/mRNA), N^6,N^6-dimethyladenosine ($m^6{}_2A$, rRNA), threonylcarbamoyladenosine (t^6A, tRNA), N^6-isopentenyladenosine (i^6A, tRNA), N^6-methyl-N^6-threonylcarbamoyladenosine (*m^6t^6A,* tRNA), 1-methylguanosine (m^1G, rRNA/tRNA), 2-methylguanosine (m^2G, rRNA/tRNA), 7-methylguanosine (m^7G, rRNA/tRNA), 2,2-dimethylguanosine ($m^2{}_2G$, rRNA/tRNA), 5-methylcytidine (m^5C, mRNA), N^4-acetylcytidine (ac^4C, rRNA/tRNA), and 3-methyluridine (m^3U, rRNA). To our surprise, also for these modified RNA nucleosides no significant differences were obtained between the levels found in human tumour and healthy tissues. However, we noticed consistently slightly lower levels in tumor tissues.

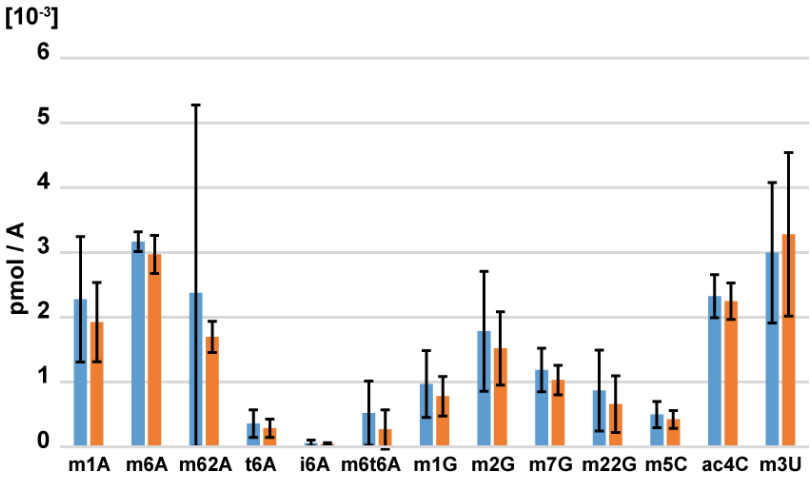


Figure 4-9: Depiction of the detected RNA modified nucleosides in healthy (light blue) and tumor tissues (orange). Graphs show data for healthy and cancer tissues from 12 different patients measured in technical triplicates.

The dysregulation of epigenetic RNA pathways, involving writers, erasers and readers of modifications, might be involved in the pathogenesis of human diseases like cancer[175]. These writers, erasers and readers, can have promoting or inhibitory effects on the hallmarks (shown by *Hanahan et al.*[176]) of cancer development[175]. These hallmarks consist of factors like evading growth suppressors, resisting cell death, inducing angiogenesis and an activation of invasion and metastasis[176]. Lung cancer can be caused by an upregulation of specific oncogenes[175]. This upregulation can be affected by the m^6A writer, a complex of methyltransferase-like 3 (METTL3) and METTL14, or m^6A eraser like fat mass and obesity-associated protein (FTO)[175]. These m^6A players, play a certain role in cell proliferation, growth suppression and metastatic potential[175] and could play a crucial role in angiogenesis, the interaction of cancer cells and the host immune

system and a possibility in using those enzymes as targets for immunotherapy[175]. It is interesting that all this has no effect on the levels of critical RNA modifications. It seems that the modification levels are critical so that cells with mis-modifications even if they are cancer cells cannot survive.

All together the here reported data show no changes on the RNA level. This is a quite interesting result, since RNA modifications are an intermediate in protein synthesis (mRNA) or an effector molecule (rRNA or tRNA) or can have direct functional effects on gene expression[175]. Furthermore, modifications like m^1A, m^5C, Ψ and m^6A form the epitranscriptome and code for a new layer of information, controlling the protein synthesis[177]. Thus, one could have expected changes from healthy to cancerous tissues, comparable to DNA level. Maybe there is no change on the quantities of these modifications but on their location within the different strands. Therefore, sequencing studies could give further information, whether the modifications are located on a different position in healthy and cancerous tissues.

4.3 Effect of smoking conditions on DNA modification levels in cell studies

In a collaboration with the Stathopoulos group of the Comprehensive Pneumology Center, who did the cell culture studies, we wanted to investigate the effect of cigarette smoking on the DNA modification levels by treating cell lines with cigarette smoke extract (CSE). We were especially interested in the ROS caused oxidative lesions of this treatment, to gain first information about the direct effect of smoking.

4.3.1 Results and discussion

As mentioned above, smoking is one of the main key players of externally generated oxidative stress to the cells. Figure **4-10** shows the 8-oxo-dG levels for two different non-small cell lung carcinoma cell lines (A549 and EKVX) under smoked and normal conditions. For the analysis, two cell lines A549 and EVKX, were treated with CSE over a period of two weeks before they were directly measured by mass spectrometry. This CSE was generated by bubbling cigarette smoke through normal cell medium.

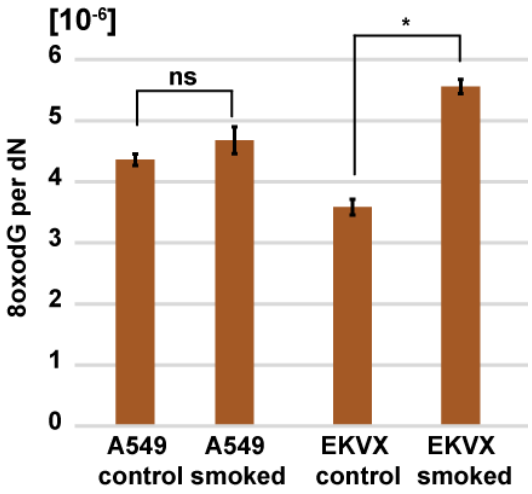


Figure 4-10: Comparison of the averaged 8-oxo-dG levels in treated and control A549 and EKVX cells lines. Graphs show data for smoked and control from three biological replicates, measured as technical triplicates. Bars show mean, error bars show standard deviation (s.d.). (Students t-test, Two-sample t-test assuming equal variances: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$)

The data show for A549 cell line induced increased 8-oxo-dG levels, which are very strong for EKVX cells and shows an increase of ~ 2 x 10^{-6} per dN. Thus, there is a greater impact on the EKVX cell lines by smoking conditions compared to A549. An earlier publication by *Liu et al.*[178] showed for a liquid cigarette smoke extract (CSE) treatment of A549 cells a modified cell structure

with an elongated shape and reduced cell-cell contact and a more fibroblast-like morphology[178]. This might prevent a greater exchange of medium through a thicker cell membrane, meaning less oxidative damage of the DNA caused by CSE medium. But the measured 8-oxo-dG levels show comparable amounts to *Amanita muscaria* with ~ 5 x 10^{-6} per dN. This suggests that although there is an impact of CSE on the cells, there is no dramatic increase of 8-oxo-dG levels compared to *Amanita muscaria.* The EVKX cell line showed increased 8-oxo-dG level after treatment with CSE, since CSE contains free radical oxygen species, which can influence dG.

Figure **4-11** shows the results of mdC quantification. Firstly, one can see again less influence of the CSE-medium on the mdC levels of A549 whereas the mdC levels for EKVX are dramatically reduced. As mentioned in the introduction, oxidative stress effects the epigenetics and thereby the methylation of dC. A comparison with the lung cancer samples (chapter **4.1**) shows comparable results for the tumor stage and A549/EKVX cells we picked.

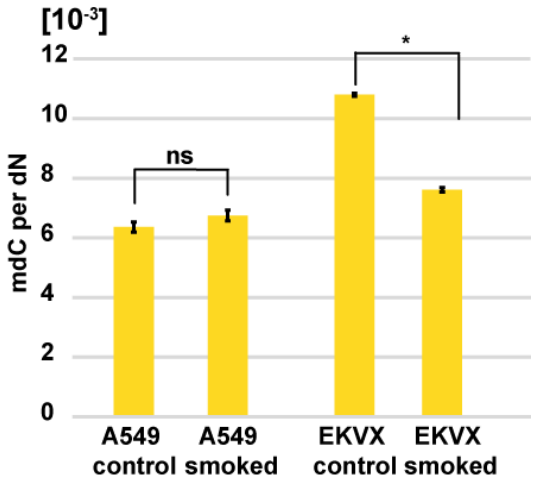


Figure 4-11: Comparison of the averaged mdC levels in treated and control A549 and EKVX cells lines. Graphs show data for smoked and control from three biological replicates, measured as technical triplicates. Bars show mean, error bars show standard deviation (s.d.). (Students t-test, Two-sample t-test assuming equal variances: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$)

When comparing the hmdC (Figure **4-12**) levels one can see higher starting levels of A549 of about 4 x 10^{-6} per dN compared to EKVX. A549 again shows no difference, whereas the hmdC levels of EKVX are slightly decreased for the CSE conditions. These data are ~5 x 10^{-5} per dN lower than the data for the tumor tissues. This might be affected by controlled cell culture conditions.

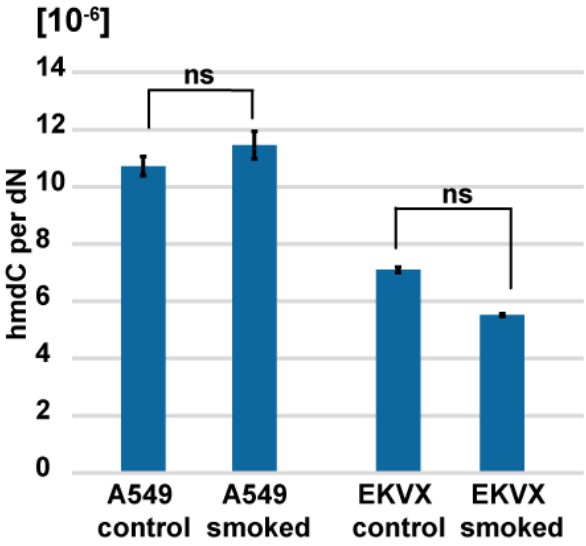


Figure 4-12: Comparison of the averaged hmdC levels in treated and control A549 and EKVX cells lines. Graphs show data for smoked and control from three biological replicates, measured as technical triplicates. Bars show mean, error bars show standard deviation (s.d.). (Students t-test, Two-sample t-test assuming equal variances: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$)

The fdC-levels, depicted in Figure **4-13**, show higher values for EKVX cells compared to A549. The fdC data for EKVX cells have slightly decreased, whereas the values for A549 cells do not show a change. Both cell lines show 5-times less fdC values compared to *Amanita muscaria* and the tumor tissues (see Chapter **4.1**). An analysis of the cadC was not performed, since the data were below the quantification limit.

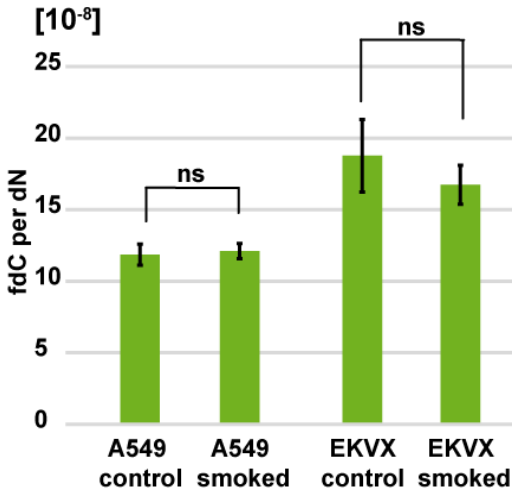


Figure 4-13: Comparison of the averaged fdC levels in treated and control A549 and EKVX cells lines. Graphs show data for smoked and control from three biological replicates, measured as technical triplicates. Bars show mean, error bars show standard deviation (s.d.). (Students t-test, Two-sample t-test assuming equal variances: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$)

In summary, the A549 cell line is probably an insufficient test cell system for investigating the influence of cigarette smoked extract treatment, because they do not show any changes due to the change of their cell surface. The EKVX cell line, on the other hand, shows significant changes for the 8-oxo-dG, mdC and hmdC levels. Firstly, the 8-oxo-dG values show the influence of CSE-medium on the cells and indicates that this cell line functions well. Secondly, the oxidative damage caused by the CSE medium represented by the mdC and hmdC values, which are lower compared to the control. It was already shown before, that oxidative stress exposure is associated with a decrease in DNA methylation levels[179]. Dnmt1 could play a role here, as it was recently proposed to have a mechanistic link to oxidative stress[179]. Therefore, the effect of remethylation caused by cell division as well as the linkage between oxidative stress on Dnmt1 should be analysed in further experiments.

4.4 Effect of hyperoxia treatment on modification levels in mice

With this first study we were interested to investigate the effect of hyperoxia treatment of new-born mice, old mice and the long-term effect of short time ventilation. The effect of short time hyperoxia treatment on lung tissue, including its long-term damage and hence a potential impact on the DNA modification levels, is still a poorly understood area. Short time hyperoxia treatment plays a major role in ill-adapted patients during the new-born period[180] and in times like the Corona pandemic in 2020/21, when many patients had to be artificially ventilated due to the disease. *Chu et al.*[181] and *Girardis et al.*[182] showed already a correlation of oxygen treatment and mortality. We hypothesized that the epigenetic modification 8-oxo-dG could function as a potential marker for fast measurements to minimize the mortality rates of ventilated patients. One could suspect that oxidative damage may have a significant contribution on the DNA level next to the initial disease[183]. First studies of hyperoxia treated mice in 8-day experiments, showed changes in airway function and airway hyperreactivity[184]. Although artificial respiration of higher oxygen concentrations is known to have negative effects on the lung structure and function [185], hardly anything is known about the effect on the DNA levels of these patients besides a correlation of increased oxygen therapy and increased 8-oxo-dG levels[186].

4.4.1 Results and Discussion

The mice, needed for these experiments, were exposed for 8 hours to either a hyperoxia (40% O_2) condition (treated) or ambient air (control). Three experimental groups were generated. The first group of mice were treated with hyperoxia as new-borns (day 4-7) and sacrificed immediately after treatment. In a second group, older mice (82 weeks) were treated with hyperoxia and sacrificed at the end of the experiment and in a third experiment, new-born mice (~ 5days) were treated with hyperoxia and sacrificed later (43 weeks), to visualise possible long-term effects of short hyperoxia treatments. The data obtained in the process are presented in the Figures **4-14** and **4-15**.

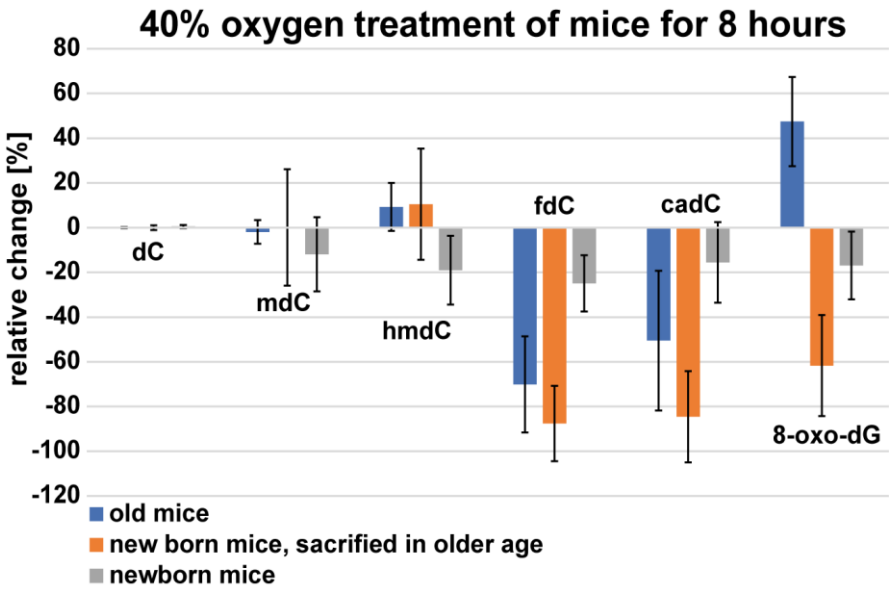


Figure 4-14: Hyperoxia treatment of old (blue) or new-born mice (grey), as well as long-term effects of hyperoxia treatment on the new-born mice (orange). Graphs show data of three different conditions from different biological replicates (adult: 3; new-born: 4; new-born, sacrificed later: 6) measured as technical duplicates.

At first, we discuss the established marker for oxidative damage, 8-oxo-dG. It can be noticed that the data show large deviations (see Figures **4-14** and **4-15**). It seems that hyperoxia treatment has indeed a large effect on the lungs of adult mice. The 8-oxo-dG level is increased by ~40%. The level in new-born mice shows no change, which could be caused due to not fully developed lung structure. More dramatically are the results of new-born mice sacrificed as adults where the 8-oxo-dG levels are decreased by nearly 60%. One could suggest, that the short-term hyperoxia treatment may lead to an upregulation of DNA repair mechanism or an activation of other genes. This needs to be further investigated. It becomes clear by comparing the new-born data of the long-term study with the short-term study, where the 8-oxo-dG levels are neither up nor down regulated. In the future, the expression of genes like *Ogg*1 or Tet enzymes need to be analysed. A transcriptome wide analysis showed already for 8 h of mechanical ventilation 19 upregulated and 5 downregulated genes[187]. In cell studies, within this project, Dnmt1 has been identified as a downregulated gene, which can be confirmed by the depicted data in Figure **4-14**.

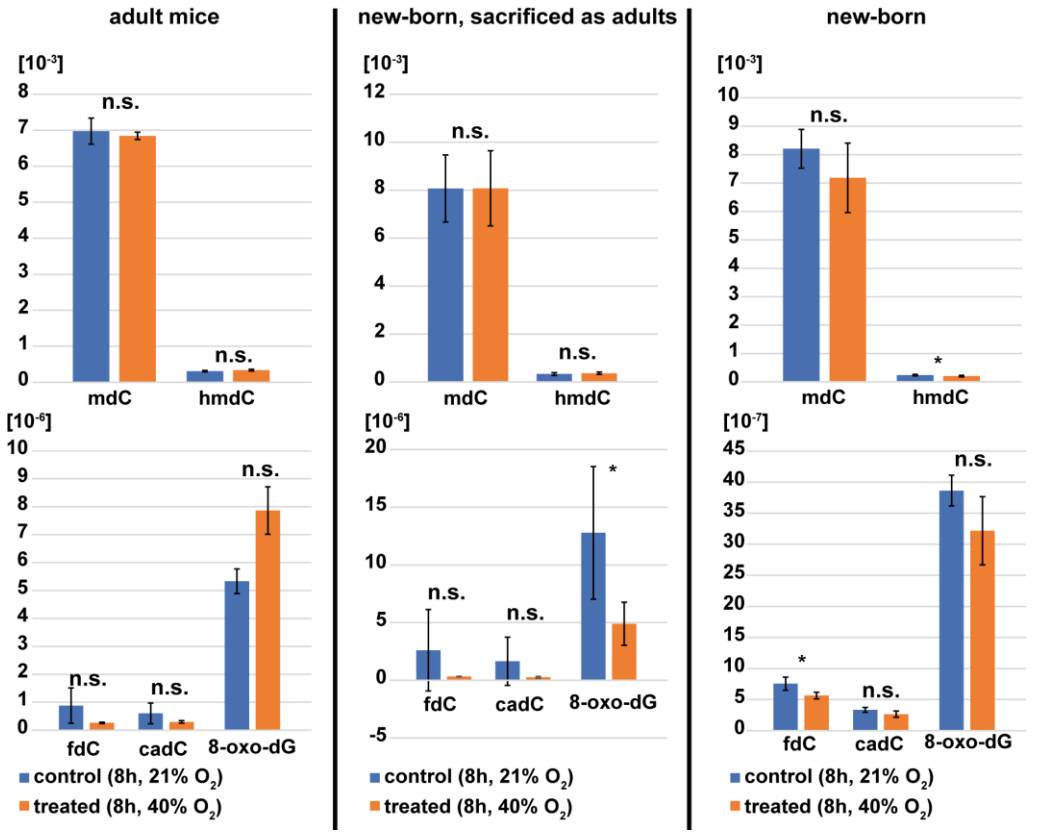


Figure 4-15: Hyperoxia treatment of old (blue) or new-born mice (grey), as well as long-term effects of hyperoxia treatment on the new-born mice (orange). Graphs show data of three different conditions from different biological replicates measured as technical duplicates (adult: 3; new-born: 4; new-born, sacrificed later: 6). Bars show mean, error bars show standard deviation (s.d.). (Students t-test, Two-sample t-test assuming equal variances: ns: $p \geq 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$)

Amarelle et al.[183] showed, that mechanical ventilation has an impact on the lung structure and hyperoxia is known to alter growth factor signalling, cell proliferation, apoptosis and vascular development. *Krokidis et al.*[188] revealed a combined scenario of membrane lipid remodelling and oxidatively-induced 8-oxo-purine adducts under variable radical reactivity and stress caused by high oxygen treatment.

A closer inspection of the mdC levels, shown in Figure **4-14**, shows no change for old mice and the long-term study whereas new-born mice show a decrease. For hmdC (~10%) long-term studied mice and adult mice show a similar trend with increased hmdC levels, whereas new-borns show the opposite. Old mice with a fully developed pulmonary lung system might have a different cellular process compared to new-born mice with an undeveloped pulmonary system. A comparison of the increasing mdC and hmdC values, of old mice, with the decreasing values of fdC (~80%) and cadC (~50%) suggests that there is an epigenetic link between the short-term

hyperoxia treatment and the resulting effects. This effect is corroborated by the long-term study of new-born treated mice, which also shows a decrease in fdC and cadC levels after they were sacrificed as adults. In future investigations, one should sequence the methylation pattern of these mice and could compare the Tet and *Ogg*1 activity.

These initial data need to be followed up to reveal potential late effects of oxidative stress on the DNA modification level. One could analyse the lungs regarding changes within the structure, since hyperoxia treatment might influence a process to prevent high levels of oxidative lesions. *Dylag et al.*[184] already described in a study, that long term oxygen treatment resulted in a change in the airway functions as well as alveolar simplification, airway tethering and elastin redistribution[184].

4.5 Methods

4.5.1 DNA analysis of lung tissues

4.5.1.1 Sample origin and use

Within the framework of a collaboration with PD Dr. med. Christian Schneider of Thoracic Surgery and Human Tissue and Cell Research (HTCR), patient material was obtained for the following analyses. All healthy and tumor-like sample material from each patient was immediately snap-frozen after removal. In a first approach, the patient material from patients with UICC classification I and III was used, a classification according to the system of the International Association Against Cancer. A regional limitation without lymphatic invasion was used for stage I and stage II with and without lymphatic invasion.

4.5.1.2 DNA isolation and digestion

The lung tissues (0.1 g) were frozen in liquid nitrogen and then crushed for 2 min at 30 Hz by a ball mill (Tissue Lyzer Qiagen). The powder, was then directly used for DNA isolation, performed according to work published earlier[189]. To inhibit and reduce background oxidation during the isolation process, all buffers were supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 µM) and deferoxamine mesylate salt (Desferal, 200 µM)[189-190].

The chemicals used for the digestion of the obtained DNA are shown exemplified in Table **1**. The amount of DNA was calculated after photometrically measuring the DNA concentration via the Nanodrop (Implen NanoPhotometer, N60, Version NPOS 4.2e build 14900), before been digested to nucleosides by Nucleoside Digestion Mix (M0649S) kit (New England BioLabs Inc.) according to the manufacturer's instructions. The spiked nucleoside mix was prepared from heavy labelled with the final concentrations of mdC (10.2 µM), hmdC (0.2 µM), fdC (0.005 µM), cadC (0.004 µM) and 8-oxo-dG (0.022 µM)[170]. The final volume of 50 µL was then incubated at 37 °C for 1.5 hours. Before submitting the samples to mass spectrometry, all samples where filtered by using an Acro-Prep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). Information about the mass spectrometry measurements are show in chapter **4.5.5.1.**

Table 1: General digestion sheet for the isolated DNA of the obtained tissues.

	Sample	m (DNA) [ng] dissolved in water	V (Reaction buffer 10X) [µL]	V (Enzyme mix) [µL]	V (Nucleosides mix) [µL]	Total Volume [µL]	Incubate 1 h 30 min 37 °C	Injection volume [µL]
Blank 1	**1**	0	5.0	2.0	3.0	50.0		39
Blank 2	**2**	0	5.0	2.0	3.0	50.0		39
Lung tissue (healthy)	**3**	1000 / 1500	5.0	2.0	3.0	50.0		39
Lung tissue (tumor)	**4**	1000	5.0	2.0	3.0	50.0		39

4.5.2 RNA analysis of lung tissues

The flow-throughs obtained during the DNA isolation described above[189], were further extracted to gain pure RNA samples. For this purpose, 2 vol. ethanol (100%) was added to the collected flow-throughs from the DNA isolation. This mixture was pipetted onto Zymo-Spin IIC columns and spun down at 10.000 *x* g for 2 min. In a next step, the RNA samples contained in the column were washed with 400 µL of an RNA wash solution (1 min, 10.000 *x* g), before potential contaminating DNA was digested using a protocol of peqGOLD (peqlab, DNas I Digest kit). The columns were then rinsed with 400 µL RNA Prep Buffer, 600 µL RNA Wash Solution. Purified RNA was eluted in 25 µL RNAse free water. The chemicals used for the digestion of the obtained RNA are shown exemplified in Table **2.** The amount of RNA was calculated after photometrically measuring the RNA concentration via the Nanodrop (Implen NanoPhotometer, N60, Version NPOS 4.2e build 14900), before been digested to nucleosides by Nucleoside Digestion Mix (M0649S) kit (New England BioLabs Inc.) according to the manufacturer's instructions. The final volume of 50 µL was then incubated at 37 °C for 1.5 hours. The spiked in nucleoside mix was prepared from heavy labelled RNA modifications produced by the Carell group. The amounts of spiked nucleosides were estimated in preliminary trials with a mixture of randomly chosen healthy and tumor samples.

Before submitting the samples to mass spectrometry, all samples were filtered by using an Acro-Prep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). Information about the mass spectrometry measurements are show in chapter **4.5.5.2.**

Table 2: General digestion sheet for the isolated RNA of the obtained tissues.

	Sample	m (RNA) [ng] dissolved in water	V (Reaction buffer 10X) [µL]	V (Enzyme mix) [µL]	Total Volume [µL]	Incubate 1 h 30 min 37 °C	V (H_2O) [µL]	(V) Nucleoside Mix	Total Volume [µL]	Injection volume [µL]
Blank 1	**1**	0	5.0	2.0	50.0		40 - V_{nuc}		90	75
Blank 2	**2**	0	5.0	2.0	50.0		40 - V_{nuc}		90	75
Lung tissue (healthy)	**3**	1000 / 1500	5.0	2.0	50.0		40 - V_{nuc}		90	75
Lung tissue (tumor)	**4**	1000	5.0	2.0	50.0		40 - V_{nuc}		90	75

4.5.3 Effect of smoking conditions on DNA modification levels in cell studies

4.5.3.1 Sample origin and use

In a collaboration with Dr. Lamort, in the group of Prof. Dr. Stathopoulos, we analysed the effect of two weeks smoking conditions on two A549 (KRAS WT) and EKVX (KRAS MUT) cell lines. For this purpose, cigarette smoke is bubbled through the cell medium and then used in cell experiments.

4.5.3.2 DNA isolation and digestion

The cells were crushed for 2 min at 30 Hz by a ball mill (Tissue Lyzer Qiagen). The powder, was then directly used for DNA isolation, performed according to a work published earlier[189]. To inhibit and reduce background oxidation during the isolation process, all buffers were supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 µM) and deferoxamine mesylate salt (Desferal, 200 µM)[189-190].

The used chemicals for the digestion of the obtained DNA are shown exemplified in Table **3.** The amount of DNA was calculated after photometrically measuring the DNA concentration via the Nanodrop (Implen NanoPhotometer, N60, Version NPOS 4.2e build 14900), before been digested to nucleosides by Nucleoside Digestion Mix (M0649S) kit (New England BioLabs Inc.) according to the manufacturer's instructions. The spiked nucleoside mix was prepared from heavy labelled with the final concentrations of mdC (10.2 µM), hmdC (0.2 µM), fdC (0.005 µM), cadC (0.004 µM) and 8-oxo-dG (0.022 µM)[170]. The final volume of 50 µL was then incubated at 37 °C for 1.5 hours. Before submitting the samples to mass spectrometry, all samples where filtered by using an Acro-Prep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). Information about the mass spectrometry measurements are show in chapter **4.5.5.1.**

Table 3: General digestion sheet for the isolated DNA of the obtained tissues.

	Sample	m (DNA) [ng] dissolved in water	V (Reaction buffer 10X) [µL]	V (Enzyme mix) [µL]	V (Nucleosides mix) [µL]	Total Volume [µL]	Incubate 1 h 30 min 37 °C	Injection volume [µL]
Blank 1	**1**	0	5.0	2.0	3.0	50.0		39
Blank 2	**2**	0	5.0	2.0	3.0	50.0		39
DNA (control)	**3**	1000 / 1500	5.0	2.0	3.0	50.0		39
DNA (smoked)	**4**	1000	5.0	2.0	3.0	50.0		39

4.5.4 Effect of hyperoxia on treatment on modification levels in mice

4.5.4.1 Sample origin and use

In a first experiment, mice were exposed to an increased oxygen concentration for a short time period of 8 hours. The data required for these measurements were obtained in collaboration with the working group of PD Dr. Anne Hilgendorff.

4.5.4.2 DNA isolation and digestion

The lung tissues (0.1 g) were frozen in liquid nitrogen and then crushed for 2 min at 30 Hz by a ball mill (Tissue Lyzer Qiagen). The powder, was then directly used for DNA isolation, performed according to a work published earlier[189]. To inhibit and reduce background oxidation during the isolation process, all buffers were supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 µM) and deferoxamine mesylate salt (Desferal, 200 µM)[189-190].

The used chemicals for the digestion of the obtained DNA is shown exemplified in Table **4.** The amount of DNA was calculated after photometrically measuring the DNA concentration via the Nanodrop (Implen NanoPhotometer, N60, Version NPOS 4.2e build 14900), before been digested to nucleosides by Nucleoside Digestion Mix (M0649S) kit (New England BioLabs Inc.) according to the manufacturer's instructions. The spiked nucleoside mix was prepared from heavy labelled with the final concentrations of mdC (10.2 µM), hmdC (0.2 µM), fdC (0.005 µM), cadC (0.004 µM) and 8-oxo-dG (0.022 µM)[170]. The final volume of 50 µL was then incubated at 37 °C for 1.5 hours. Before submitting the samples to mass spectrometry, all samples where filtered by using an Acro-Prep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). Information about the mass spectrometry measurements are show in chapter **4.5.5.1.**

Table 4: General digestion sheet for the isolated DNA of the obtained tissues.

	Sample	m (DNA) [ng] dissolved in water	V (Reaction buffer 10X) [µL]	V (Enzyme mix) [µL]	V (Nucleosides mix) [µL]	Total Volume [µL]	Incubate 1 h 30 min 37 °C	Injection volume [µL]
Blank 1	**1**	0	5.0	2.0	3.0	50.0		39
Blank 2	**2**	0	5.0	2.0	3.0	50.0		39
Lung tissue (control)	**3**	1000 / 1500	5.0	2.0	3.0	50.0		39
Lung tissue (ventilated)	**4**	1000	5.0	2.0	3.0	50.0		39

4.5.5 LC-MS/MS measurements

4.5.5.1 DNA Measurements with MS/MS

All experiments were performed in biological and technical triplicates to ensure reproducibility. Solvents for LC-MS/MS analysis were purchased from Honeywell, Roth and used without further purification.

The analysis was performed using an UHPLC-QQQ-MS/MS system consisting of a Triple Quadrupole 6490 mass spectrometer (Agilent) with an ESI source and an Agilent Infinity 1290 UHPLC. The elution was monitored at 260 nm (Agilent InfinityLab Deuterium Lamp G1314). Data Acquisition and processing were performed using MassHunter Workstation Software Version B.07.01 (Agilent).

mdC, hmdC, fdC, cadC and 8-oxo-dG were separated on an InfinityLab Poroshell 120 SB-C8 column (2.1 mm x 150 mm, 2.7 µm, Agilent Technologies, USA) at 35 °C. Water containing 0.0085% FA (v/v, solvent A) and MeCN containing 0.0085% FA (v/v, solvent B) was used as the mobile phase. A gradient of 0 - 3.5% B for 4 min, 3.5 - 5% B for 2.9 min, 5 - 80% B for 0.3 min, 80% B for 3.3 min was used. The flow rate of the mobile phase was set to 0.35 mL min^{-1}[189].

The source-dependent parameters were as follows, gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N_2), capillary voltage 2,500 V in the positive ion mode, and nozzle voltage 500 V. Delta EMV was set to 500 (positive mode). The nucleosides and labelled products were monitored using the multiple reaction monitoring (MRM) mode. The MRM parameters were optimized to achieve maximal detection sensitivity (Tables **5**)[170].

Table 5: MRM parameters for the quantification and detection of the modified dC and dG molecules.

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity	Collision Energy	Fragmentor
UV	1	0,0 -> 0,0	MRM	Target	0	0	6	Positive		
mC-dN-d3	3	245.1 -> 129.1	MRM	ISTD	245,1	129,1	3,472	Positive	1	380
mC-dN	3	242.1 -> 126.1	MRM	Target	242,1	126,1	3,533	Positive	1	380
hmC-dN-d2-15n2	2	262.1 -> 146.1	MRM	ISTD	262,1	146,1	2,082	Positive	2	380
hmC-dN	2	258.1 -> 142.1	MRM	Target	258,1	142,1	2,121	Positive	2	380
fC-dN-15n2	5	258.1 -> 142.0	MRM	ISTD	258,1	142	6	Positive	5	380
fC-dN	5	256.1 -> 140.1	MRM	Target	256,1	140,1	6	Positive	5	380
fC	5	140.0 -> 97.0	MRM	Target	140	97	6	Positive	13	380
dG-UV	3	0,0 -> 0,0	MRM	Target	0	0	6,3	Positive		
dC-UV	3	0.0 -> 0.0	MRM	Target	0	0	1,838	Positive		
caC-dN-15n2	3	274.1 -> 158.0	MRM	ISTD	274,1	158	3,195	Positive	5	380
caC-dN	3	272.1 -> 156.0	MRM	Target	272,1	156	3,214	Positive	5	380
caC	3	156.0 -> 138.0	MRM	Target	156	138	3,21	Positive	13	380
8oxoG-dN-15n5	5	289.1 -> 173.0	MRM	ISTD	289,1	173	6,5	Positive	9	380
8oxoG-dN	5	284.1 -> 168.1	MRM	Target	284,1	168,1	6,5	Positive	9	380

4.5.5.2 RNA measurements with Orbitrap

A Dionex Ultimate 3000 HPLC system coupled to a Thermo Fisher LTQ Orbitrap XL mass spectrometer was used for the quantitative HPLC-HESI-MS analysis of the enzymatically digested RNA samples with an injection volume of 75 µL. Nucleosides were separated on an Interchim Uptisphere120-3HDO-C18 column at 30 °C. Elution buffers were buffer A (2 mM NH_4HCOO in H_2O; pH 5.5) and buffer B (2 mM NH_4HCOO in H_2O/MeCN 20/80 v/v; pH 5.5) with a flow rate of 0.15 mL/min with the followed gradient: 0→10 min, 0% B; 10→50 min, 0→5 % B; 50→75 min, 5→60 % B; 75→80 min, 60→100 % B. The chromatogram was recorded at 260 nm with a Dionex Ultimate 3000 Diode Array Detector, and the chromatographic eluent was directly injected into the ion source of the mass spectrometer without prior splitting. Ions were scanned in the positive polarity mode over a full-scan range of m/z = 225-800 with a resolution of 60,000. Further parameters are: Capillary temperature 275 °C; source voltage 4.80 kV; vaporizer temperature 100 °C, capillary voltage 0 V; tube lens voltage 45 V. The ion chromatograms of the compounds of interest were extracted from the total ion current (TIC) chromatogram, and the areas under the curves were integrated for further calculating of final amounts.

References

[1] J. Rissler, A. Gudmundsson, H. Nicklasson, E. Swietlicki, P. Wollmer, J. Löndahl, *Deposition efficiency of inhaled particles (15-5000 nm) related to breathing pattern and lung function: an experimental study in healthy children and adults, Part. Fibre Toxicol.* **2017**, *14*, 10.

[2] D. Bösch, *Lunge und Atemwege*, 1 ed., Springer, Berlin, Heidelberg, **2014**.

[3] S. S. Salvi, P. J. Barnes, *Chronic obstructive pulmonary disease in non-smokers, Lancet* **2009**, *374*, 733-743.

[4] C. F. Vogelmeier, G. J. Criner, F. J. Martinez, A. Anzueto, P. J. Barnes, J. Bourbeau, B. R. Celli, R. Chen, M. Decramer, L. M. Fabbri, P. Frith, David M. G. Halpin, M. V. López Varela, M. Nishimura, N. Roche, R. Rodriguez-Roisin, D. D. Sin, D. Singh, R. Stockley, J. Vestbo, J. A. Wedzicha, A. Agusti, *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary, Eur. Respir. J.* **2017**, *49*, 1700214.

[5] D. Plass, T. Vos, C. Hornberg, C. Scheidt-Nave, H. Zeeb, A. Krämer, *Entwicklung der Krankheitslast in Deutschland, Dtsch. Arztebl. Int.* **2014**, *111*, 629-638.

[6] S. Bundesamt, in *Mortalität und Todesursachen (2021) Sterbefälle, Sterbeziffern (je 100.000 Einwohner, altersstandardisiert) (ab 1998).* Statistisches Bundesamt, Gesundheitsberichterstattung des Bundes, **2021**.

[7] V. Navaratnam, R. B. Hubbard, *The Mortality Burden of Idiopathic Pulmonary Fibrosis in the United Kingdom, Am. J. Respir. Crit. Care Med.* **2019**, *200*, 256-258.

[8] R. Li, Z. Jia, M. A. Trush, *Defining ROS in Biology and Medicine, React. Oxyg. Species (Apex)* **2016**, *1*, 9-21.

[9] D. P. Rosanna, C. Salvatore, *Reactive oxygen species, inflammation, and lung diseases, Curr. Pharm. Des.* **2012**, *18*, 3889-3900.

[10] W. Domej, K. Oettl, W. Renner, *Oxidative stress and free radicals in COPD--implications and relevance for treatment, Int. J. Chron. Obstruct. Pulmon. Dis.* **2014**, *9*, 1207-1224.

[11] U. S. Srinivas, B. W. Q. Tan, B. A. Vellayappan, A. D. Jeyasekharan, *ROS and the DNA damage response in cancer, Redox. Biol.* **2019**, *25*, 101084.

[12] A. L. Durham, I. M. Adcock, *The relationship between COPD and lung cancer, Lung Cancer* **2015**, *90*, 121-127.

[13] W. A. Pryor, K. Stone, *Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite, Ann. N. Y. Acad. Sci.* **1993**, *686*, 12-27; discussion 27-18.

[14] P. J. Barnes, *Oxidative stress-based therapeutics in COPD, Redox Biol* **2020**, *33*, 101544.

[15] S. S. Hecht, *Tobacco Smoke Carcinogens and Lung Cancer, J. Natl. Cancer Inst.* **1999**, *91*, 1194-1210.

[16] S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell, *Mechanism of replication blocking and bypass of Y-family polymerase {eta} by bulky acetylaminofluorene DNA adducts, Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20720-20725.

[17] G. Caramori, P. Kirkham, A. Barczyk, A. Di Stefano, I. Adcock, *Molecular pathogenesis of cigarette smoking-induced stable COPD, Ann N Y Acad Sci* **2015**, *1340*, 55-64.

[18] T. Schaberg, U. Klein, M. Rau, J. Eller, H. Lode, *Subpopulations of alveolar macrophages in smokers and nonsmokers: relation to the expression of CD11/CD18 molecules and superoxide anion production, Am J Respir Crit Care Med* **1995**, *151*, 1551-1558.

[19] A. Noguera, S. Batle, C. Miralles, J. Iglesias, X. Busquets, W. MacNee, A. G. Agustí, *Enhanced neutrophil response in chronic obstructive pulmonary disease, Thorax* **2001**, *56*, 432-437.

[20] H.-Y. Tan, N. Wang, S. Li, M. Hong, X. Wang, Y. Feng, *The Reactive Oxygen Species in Macrophage Polarization: Reflecting Its Dual Role in Progression and Treatment of Human Diseases, Oxid. Med. Cell. Longev.* **2016**, *2016*, 2795090.

[21] A. Soltani, D. W. Reid, S. S. Sohal, R. Wood-Baker, S. Weston, H. K. Muller, E. H. Walters, *Basement membrane and vascular remodelling in smokers and chronic obstructive pulmonary disease: a cross-sectional study, Respir Res* **2010**, *11*, 105.

[22] A. Corlăţeanu, I. Odajiu, V. Botnaru, S. Cemirtan, *From smoking to COPD--current approaches, Pneumologia* **2016**, *65*, 20-23.

[23] A. Papi, G. Casoni, G. Caramori, I. Guzzinati, P. Boschetto, F. Ravenna, N. Calia, S. Petruzzelli, L. Corbetta, G. Cavallesco, E. Forini, M. Saetta, A. Ciaccia, L. M. Fabbri, *COPD increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma, Thorax* **2004**, *59*, 679-681.

[24] M. Reck, D. F. Heigener, T. Mok, J. C. Soria, K. F. Rabe, *Management of non-small-cell lung cancer: recent developments, Lancet* **2013**, *382*, 709-719.

[25] C. Zappa, S. A. Mousa, *Non-small cell lung cancer: current treatment and future advances, Transl. Lung Cancer Res.* **2016**, *5*, 288-300.

[26] W. D. Travis, E. Brambilla, M. Noguchi, A. G. Nicholson, K. R. Geisinger, Y. Yatabe, D. G. Beer, C. A. Powell, G. J. Riely, P. E. Van Schil, K. Garg, J. H. Austin, H. Asamura, V. W. Rusch, F. R. Hirsch, G. Scagliotti, T. Mitsudomi, R. M. Huber, Y. Ishikawa, J. Jett, M. Sanchez-Cespedes, J. P. Sculier, T. Takahashi, M. Tsuboi, J. Vansteenkiste, I. Wistuba, P. C. Yang, D. Aberle, C. Brambilla, D. Flieder, W. Franklin, A. Gazdar, M. Gould, P. Hasleton, D. Henderson, B. Johnson, D. Johnson, K. Kerr, K. Kuriyama, J. S. Lee, V. A. Miller, I. Petersen, V. Roggli, R. Rosell, N. Saijo, E. Thunnissen, M. Tsao, D. Yankelewitz, *International association for the study of lung cancer/american thoracic society/european*

respiratory society international multidisciplinary classification of lung adenocarcinoma, J. Thorac. Oncol. **2011**, *6*, 244-285.

[27] N. Taghizadeh, J. M. Vonk, H. M. Boezen, *Lifetime Smoking History and Cause-Specific Mortality in a Cohort Study with 43 Years of Follow-Up, PLoS One* **2016**, *11*, e0153310.

[28] A. Agudo, C. Bonet, N. Travier, C. A. González, P. Vineis, H. B. Bueno-de-Mesquita, D. Trichopoulos, P. Boffetta, F. Clavel-Chapelon, M.-C. Boutron-Ruault, R. Kaaks, A. Lukanova, M. Schütze, H. Boeing, A. Tjonneland, J. Halkjaer, K. Overvad, C. C. Dahm, J. R. Quirós, M.-J. Sánchez, N. Larrañaga, C. Navarro, E. Ardanaz, K.-T. Khaw, N. J. Wareham, T. J. Key, N. E. Allen, A. Trichopoulou, P. Lagiou, D. Palli, S. Sieri, R. Tumino, S. Panico, H. Boshuizen, F. L. Büchner, P. H. M. Peeters, S. Borgquist, M. Almquist, G. Hallmans, I. Johansson, I. T. Gram, E. Lund, E. Weiderpass, I. Romieu, E. Riboli, *Impact of Cigarette Smoking on Cancer Risk in the European Prospective Investigation into Cancer and Nutrition Study, J. Clin. Oncol.* **2012**, *30*, 4550-4557.

[29] A. W. Caliri, S. Tommasi, A. Besaratinia, *Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer, Mutat. Res. Rev. Mutat. Res.* **2021**, *787*, 108365.

[30] F. J. Rang, J. Boonstra, *Causes and Consequences of Age-Related Changes in DNA Methylation: A Role for ROS?, Biology* **2014**, *3*, 403-425.

[31] M. P. Murphy, *How mitochondria produce reactive oxygen species, Biochem. J.* **2009**, *417*, 1-13.

[32] C. C. Winterbourn, *Reconciling the chemistry and biology of reactive oxygen species, Nat. Chem. Biol.* **2008**, *4*, 278-286.

[33] M. Mittal, M. R. Siddiqui, K. Tran, S. P. Reddy, A. B. Malik, *Reactive oxygen species in inflammation and tissue injury, Antioxid. Redox Signal.* **2014**, *20*, 1126-1167.

[34] T. Fukai, M. Ushio-Fukai, *Superoxide dismutases: role in redox signaling, vascular function, and diseases, Antioxid. Redox. Signal.* **2011**, *15*, 1583-1606.

[35] M. Giorgio, M. Trinei, E. Migliaccio, P. G. Pelicci, *Hydrogen peroxide: a metabolic by-product or a common mediator of ageing signals?, Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 722-728.

[36] A. M. Fleming, C. J. Burrows, *Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters, J. Am. Chem. Soc.* **2020**, *142*, 1115-1136.

[37] M. Fransen, M. Nordgren, B. Wang, O. Apanasets, *Role of peroxisomes in ROS/RNS-metabolism: Implications for human disease, Biochim. Biophys. Acta. Mol. Basis. Dis.* **2012**, *1822*, 1363-1373.

[38] S. S. Cao, R. J. Kaufman, *Endoplasmic reticulum stress and oxidative stress in cell fate decision and human disease, Antioxid. Redox Signal.* **2014**, *21*, 396-413.

[39] R. Fabrini, A. Bocedi, V. Pallottini, L. Canuti, M. De Canio, A. Urbani, V. Marzano, T. Cornetta, P. Stano, A. Giovanetti, L. Stella, A. Canini, G.

Federici, G. Ricci, *Nuclear shield: a multi-enzyme task-force for nucleus protection, PLoS One* **2010**, *5*, e14125.

[40] T. B. Kryston, A. B. Georgiev, P. Pissis, A. G. Georgakilas, *Role of oxidative stress and DNA damage in human carcinogenesis, Mutat. Res.* **2011**, *711*, 193-201.

[41] T. Kalogeris, Y. Bao, R. J. Korthuis, *Mitochondrial reactive oxygen species: A double edged sword in ischemia/reperfusion vs preconditioning, Redox. Biol.* **2014**, *2*, 702-714.

[42] M. Malinouski, Y. Zhou, V. V. Belousov, D. L. Hatfield, V. N. Gladyshev, *Hydrogen peroxide probes directed to different cellular compartments, PLoS One* **2011**, *6*, e14564.

[43] E. Hatem, V. Berthonaud, M. Dardalhon, G. Lagniel, P. Baudouin-Cornu, M. E. Huang, J. Labarre, S. Chdin, *Glutathione is essential to preserve nuclear function and cell survival under oxidative stress, Free Radic. Biol. Med.* **2014**, *75 Suppl 1*, S25-26.

[44] W. Luo, J. G. Muller, E. M. Rachlin, C. J. Burrows, *Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model, Chem. Res. Toxicol.* **2001**, *14*, 927-938.

[45] J. Cadet, T. Douki, J.-L. Ravanat, *Oxidatively Generated Damage to the Guanine Moiety of DNA: Mechanistic Aspects and Formation in Cells, Acc. Chem. Res.* **2008**, *41*, 1075-1083.

[46] H. Hong, H. Cao, Y. Wang, Y. Wang, *Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate, Chem. Res. Toxicol.* **2006**, *19*, 614-621.

[47] S. Delaney, J. C. Delaney, J. M. Essigmann, *Chemical–Biological Fingerprinting: Probing the Properties of DNA Lesions Formed by Peroxynitrite, Chem. Res. Toxicol.* **2007**, *20*, 1718-1729.

[48] R. Abbotts, S. Madhusudan, *Human AP endonuclease 1 (APE1): from mechanistic insights to druggable target in cancer, Cancer Treat. Rev.* **2010**, *36*, 425-435.

[49] Z. D. Parsons, K. S. Gates, *Thiol-dependent recovery of catalytic activity from oxidized protein tyrosine phosphatases, Biochemistry* **2013**, *52*, 6412-6423.

[50] V. A. Kobliakov, *Mechanisms of tumor promotion by reactive oxygen species, Biochemistry (Mosc)* **2010**, *75*, 675-685.

[51] A. Ostman, J. Frijhoff, A. Sandin, F. D. Böhmer, *Regulation of protein tyrosine phosphatases by reversible oxidation, J. Biochem.* **2011**, *150*, 345-356.

[52] P. C. Dedon, S. R. Tannenbaum, *Reactive nitrogen species in the chemical biology of inflammation, Arch. Biochem. Biophys.* **2004**, *423*, 12-22.

[53] L. Aguilera-Aguirre, K. Hosoki, A. Bacsi, Z. Radák, S. Sur, M. L. Hegde, B. Tian, A. Saavedra-Molina, A. R. Brasier, X. Ba, I. Boldogh, *Whole*

transcriptome analysis reveals a role for OGG1-initiated DNA repair signaling in airway remodeling, Free Radic. Biol. Med. **2015**, *89*, 20-33.

[54] K. L. Rock, H. Kono, *The inflammatory response to cell death, Annu. Rev. Pathol.* **2008**, *3*, 99-126.

[55] A. Porzionato, M. M. Sfriso, A. Mazzatenta, V. Macchi, R. De Caro, C. Di Giulio, *Effects of hyperoxic exposure on signal transduction pathways in the lung, Respir. Physiol. Neurobiol.* **2015**, *209*, 106-114.

[56] S. G. Rhee, Y. S. Bae, S.-R. Lee, J. Kwon, *Hydrogen Peroxide: A Key Messenger That Modulates Protein Phosphorylation Through Cysteine Oxidation, Sci. STKE* **2000**, *2000*, pe1-pe1.

[57] Laura A. Sena, S. Li, A. Jairaman, M. Prakriya, T. Ezponda, David A. Hildeman, C.-R. Wang, Paul T. Schumacker, Jonathan D. Licht, H. Perlman, Paul J. Bryce, Navdeep S. Chandel, *Mitochondria Are Required for Antigen-Specific T Cell Activation through Reactive Oxygen Species Signaling, Immunity* **2013**, *38*, 225-236.

[58] M. Sundaresan, Z.-X. Yu, V. J. Ferrans, K. Irani, T. Finkel, *Requirement for Generation of H₂O₂ for Platelet-Derived Growth Factor Signal Transduction, Science* **1995**, *270*, 296-299.

[59] G. Antoniali, M. C. Malfatti, G. Tell, *Unveiling the non-repair face of the Base Excision Repair pathway in RNA processing: A missing link between DNA repair and gene expression?, DNA Repair (Amst)* **2017**, *56*, 65-74.

[60] E. A. Veal, A. M. Day, B. A. Morgan, *Hydrogen peroxide sensing and signaling, Mol. Cell* **2007**, *26*, 1-14.

[61] N. Sambiagio, J.-J. Sauvain, A. Berthet, R. Auer, A. Schoeni, N. B. Hopf, *Rapid Liquid Chromatography—Tandem Mass Spectrometry Analysis of Two Urinary Oxidative Stress Biomarkers: 8-oxodG and 8-isoprostane, Antioxidants* **2021**, *10*, 38.

[62] N. T. González, E. Otali, Z. Machanda, M. N. Muller, R. Wrangham, M. E. Thompson, *Urinary markers of oxidative stress respond to infection and late-life in wild chimpanzees, PLoS One* **2020**, *15*, e0238066.

[63] S. S. Davies, L. J. Roberts, *F2-isoprostanes as an indicator and risk factor for coronary heart disease, Free Radic. Biol. Med.* **2011**, *50*, 559-566.

[64] S. Loft, A. Olsen, P. Møller, H. E. Poulsen, A. Tjønneland, *Association between 8-oxo-7,8-dihydro-2'-deoxyguanosine Excretion and Risk of Postmenopausal Breast Cancer: Nested Case–Control Study, Cancer Epidemiol. Biomarkers Prev.* **2013**, *22*, 1289-1296.

[65] S. Loft, P. Svoboda, K. Kawai, H. Kasai, M. Sørensen, A. Tjønneland, U. Vogel, P. Møller, K. Overvad, O. Raaschou-Nielsen, *Association between 8-oxo-7,8-dihydroguanine excretion and risk of lung cancer in a prospective study, Free Radic. Biol. Med.* **2012**, *52*, 167-172.

[66] U. Deichmann, *Epigenetics: The origins and evolution of a fashionable topic, Dev. Biol.* **2016**, *416*, 249-254.

[67] C. H. Waddington, *The epigenotype. 1942, Int. J. Epidemiol.* **2012**, *41*, 10-13.

[68] R. A. Martienssen, A. D. Riggs, V. E. A. Russo, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor (N.Y.) : Cold Spring Harbor laboratory, **1996**.

[69] G. Felsenfeld, *A brief history of epigenetics, Cold Spring Harb. Perspect. Biol.* **2014**, *6*.

[70] K. J. O'Donnell, M. J. Meaney, *Epigenetics, Development, and Psychopathology, Annu. Rev. Clin. Psychol.* **2020**, *16*, 327-350.

[71] B. Jin, Y. Li, K. D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?, Genes Cancer* **2011**, *2*, 607-617.

[72] Z. Li, H. Tan, H. Yu, Z. Deng, X. Zhou, M. Wang, *DNA methylation and gene expression profiles characterize epigenetic regulation of lncRNAs in colon adenocarcinoma, J. Cell. Biochem.* **2020**, *121*, 2406-2415.

[73] J. Loaeza-Loaeza, A. S. Beltran, D. Hernández-Sotelo, *DNMTs and Impact of CpG Content, Transcription Factors, Consensus Motifs, lncRNAs, and Histone Marks on DNA Methylation, Genes* **2020**, *11*, 1336.

[74] H. J. Lee, T. A. Hore, W. Reik, *Reprogramming the methylome: erasing memory and creating diversity, Cell Stem Cell* **2014**, *14*, 710-719.

[75] M. F. Fraga, E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suñer, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y. Z. Wu, C. Plass, M. Esteller, *Epigenetic differences arise during the lifetime of monozygotic twins, Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10604-10609.

[76] Casey A. Gifford, Michael J. Ziller, H. Gu, C. Trapnell, J. Donaghey, A. Tsankov, Alex K. Shalek, David R. Kelley, Alexander A. Shishkin, R. Issner, X. Zhang, M. Coyne, Jennifer L. Fostel, L. Holmes, J. Meldrim, M. Guttman, C. Epstein, H. Park, O. Kohlbacher, J. Rinn, A. Gnirke, Eric S. Lander, Bradley E. Bernstein, A. Meissner, *Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells, Cell* **2013**, *153*, 1149-1163.

[77] S. R. Kinney, S. Pradhan, *Regulation of expression and activity of DNA (cytosine-5) methyltransferases in mammalian cells, Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 311-333.

[78] M. Okano, D. W. Bell, D. A. Haber, E. Li, *DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development, Cell* **1999**, *99*, 247-257.

[79] D. Cheishvili, L. Boureau, M. Szyf, *DNA demethylation and invasive cancer: implications for therapeutics, Br. J. Pharmacol.* **2015**, *172*, 2705-2715.

[80] M. Kulis, M. Esteller, in *Advances in Genetics, Vol. 70* (Eds.: Z. Herceg, T. Ushijima), Academic Press, **2010**, pp. 27-56.

[81] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Human DNA methylomes at base resolution show widespread epigenomic differences, Nature* **2009**, *462*, 315-322.

[82] M. V. C. Greenberg, D. Bourc'his, *The diverse roles of DNA methylation in mammalian development and disease, Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590-607.

[83] A. Breiling, F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond, Epigenetics & Chromatin* **2015**, *8*, 24.

[84] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *The discovery of 5-formylcytosine in embryonic stem cell DNA, Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7008-7012.

[85] C. B. Mulholland, A. Nishiyama, J. Ryan, R. Nakamura, M. Yiğit, I. M. Glück, C. Trummer, W. Qin, M. D. Bartoschek, F. R. Traube, E. Parsa, E. Ugur, M. Modic, A. Acharya, P. Stolz, C. Ziegenhain, M. Wierer, W. Enard, T. Carell, D. C. Lamb, H. Takeda, M. Nakanishi, S. Bultmann, H. Leonhardt, *Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals, Nat. Commun.* **2020**, *11*, 5972.

[86] R. Matuleviciute, P. P. Cunha, R. S. Johnson, I. P. Foskolou, *Oxygen regulation of TET enzymes, The FEBS Journal* **2021**, *n/a*.

[87] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids, Cell Cycle* **2009**, *8*, 1698-1710.

[88] L. M. Iyer, D. Zhang, L. Aravind, *Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification, Bioessays* **2016**, *38*, 27-40.

[89] C.-W. J. Lio, X. Yue, I. F. López-Moyado, M. Tahiliani, L. Aravind, A. Rao, *TET methylcytosine oxidases: new insights from a decade of research, J. Biosci.* **2020**, *45*, 21.

[90] 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, *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, Science* **2011**, *333*, 1303-1307.

[91] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science* **2011**, *333*, 1300-1303.

[92] Y. F. Xu, F. Z. Wu, L. Tan, L. C. Kong, L. J. Xiong, J. Deng, A. J. Barbera, L. J. Zheng, H. K. Zhang, S. Huang, J. R. Min, T. Nicholson, T. P. Chen, G. L. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells, Mol. Cell* **2011**, *42*, 451-464.

[93] H. K. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, *TET1 is a DNA-binding protein that modulates DNA methylation and*

gene transcription via hydroxylation of 5-methylcytosine, Cell Res. **2010**, *20*, 1390-1393.

[94] F. Spada, S. Schiffers, A. Kirchner, Y. Zhang, G. Arista, O. Kosmatchev, E. Korytiakova, R. Rahimoff, C. Ebert, T. Carell, *Active turnover of genomic methylcytosine in pluripotent cells, Nat. Chem. Biol.* **2020**, *16*, 1411-1419.

[95] M. Münzel, D. Globisch, T. Carell, *5-Hydroxymethylcytosine, the Sixth Base of the Genome, Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468.

[96] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates, PLoS One* **2010**, *5*, e15367.

[97] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Quantification of the sixth DNA base hydroxymethylcytosine in the brain, Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377.

[98] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA, Angew. Chem. Int. Ed.* **2018**, *57*, 4296-4312.

[99] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *5-Formylcytosine can be a stable DNA modification in mammals, Nat. Chem. Biol.* **2015**, *11*, 555-557.

[100] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Analysis of an Active Deformylation Mechanism of 5-Formyl-deoxycytidine (fdC) in Stem Cells, Angew. Chem. Int. Ed.* **2020**, *59*, 5591-5594.

[101] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair, Cell* **2011**, *146*, 67-79.

[102] 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, *Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine, J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.

[103] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells, J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.

[104] P. Fortini, E. Dogliotti, *Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways, DNA Repair (Amst).* **2007**, *6*, 398-409.

[105] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T.-P. Gu, B. Hu, Colum P. Walsh, J. Li, F. Tang, G.-L. Xu, *Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote, Cell Stem Cell* **2014**, *15*, 447-459.

[106] K. D. Rasmussen, K. Helin, *Role of TET enzymes in DNA methylation, development, and cancer, Genes Dev.* **2016**, *30*, 733-750.

[107] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S. H. Kim, S. Ito, C. Yang, P. Wang, M. T. Xiao, L. X. Liu, W. Q. Jiang, J. Liu, J. Y. Zhang, B. Wang, S. Frye, Y. Zhang, Y. H. Xu, Q. Y. Lei, K. L. Guan, S. M. Zhao, Y. Xiong, *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases, Cancer Cell* **2011**, *19*, 17-30.

[108] C. Liu, L. Liu, X. Chen, J. Shen, J. Shan, Y. Xu, Z. Yang, L. Wu, F. Xia, P. Bie, Y. Cui, X.-w. Bian, C. Qian, *Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1, PLoS One* **2013**, *8*, e62828.

[109] Christine G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C.-w. Lee, D. Hu, Bill Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, Abraham J. Khorasani, R. Fang, C. Lezcano, Lyn M. Duncan, Richard A. Scolyer, John F. Thompson, H. Kakavand, Y. Houvras, Leonard I. Zon, Martin C. Mihm, Ursula B. Kaiser, T. Schatton, Bruce A. Woda, George F. Murphy, Yujiang G. Shi, *Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma, Cell* **2012**, *150*, 1135-1146.

[110] Y. Kudo, K. Tateishi, K. Yamamoto, S. Yamamoto, Y. Asaoka, H. Ijichi, G. Nagae, H. Yoshida, H. Aburatani, K. Koike, *Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation, Cancer Sci.* **2012**, *103*, 670-676.

[111] I. Afanas'ev, *New nucleophilic mechanisms of ros-dependent epigenetic modifications: comparison of aging and cancer, Aging Dis.* **2014**, *5*, 52-62.

[112] Q. Wu, X. Ni, *ROS-Mediated DNA Methylation Pattern Alterations in Carcinogenesis, Curr. Drug Targets* **2015**, *16*, 13-19.

[113] Heather M. O'Hagan, W. Wang, S. Sen, C. DeStefano Shields, Stella S. Lee, Yang W. Zhang, Eriko G. Clements, Y. Cai, L. Van Neste, H. Easwaran, Robert A. Casero, Cynthia L. Sears, Stephen B. Baylin, *Oxidative Damage Targets Complexes Containing DNA Methyltransferases, SIRT1, and Polycomb Members to Promoter CpG Islands, Cancer Cell* **2011**, *20*, 606-619.

[114] N. Bhutani, David M. Burns, Helen M. Blau, *DNA Demethylation Dynamics, Cell* **2011**, *146*, 866-872.

[115] S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine, Nucleic Acids Res.* **1997**, *25*, 553-559.

[116] O. Tunc, K. Tremellen, *Oxidative DNA damage impairs global sperm DNA methylation in infertile men, J. Assist. Reprod. Genet.* **2009**, *26*, 537-544.

[117] P. W. Turk, A. Laayoun, S. S. Smith, S. A. Weitzman, *DNA adduct 8-hydroxyl-2'-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase, Carcinogenesis* **1995**, *16*, 1253-1255.

[118] S. A. Weitzman, P. W. Turk, D. H. Milkowski, K. Kozlowski, *Free radical adducts induce alterations in DNA cytosine methylation, Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1261-1264.

[119] S. Steenken, S. V. Jovanovic, *How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution, J. Am. Chem. Soc.* **1997**, *119*, 617-618.

[120] M.-H. Baik, J. S. Silverman, I. V. Yang, P. A. Ropp, V. A. Szalai, W. Yang, H. H. Thorp, *Using Density Functional Theory To Design DNA Base Analogues with Low Oxidation Potentials, J. Phys. Chem. B* **2001**, *105*, 6437-6444.

[121] T. Arai, V. P. Kelly, O. Minowa, T. Noda, S. Nishimura, *High accumulation of oxidative DNA damage, 8-hydroxyguanine, in Mmh/Ogg1 deficient mice by chronic oxidative stress, Carcinogenesis* **2002**, *23*, 2005-2010.

[122] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA, Nat. Chem. Biol.* **2014**, *10*, 574-581.

[123] A. Mangerich, C. G. Knutson, N. M. Parry, S. Muthupalani, W. Ye, E. Prestwich, L. Cui, J. L. McFaline, M. Mobley, Z. Ge, K. Taghizadeh, J. S. Wishnok, G. N. Wogan, J. G. Fox, S. R. Tannenbaum, P. C. Dedon, *Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer, Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1820-1829.

[124] R. Yamaguchi, T. Hirano, S. Asami, M. H. Chung, A. Sugita, H. Kasai, *Increased 8-hydroxyguanine levels in DNA and its repair activity in rat kidney after administration of a renal carcinogen, ferric nitrilotriacetate, Carcinogenesis* **1996**, *17*, 2419-2422.

[125] C. M. Gedik, A. Collins, *Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study, FASEB J.* **2005**, *19*, 82-84.

[126] G. Boysen, L. B. Collins, S. Liao, A. M. Luke, B. F. Pachkowski, J. L. Watters, J. A. Swenberg, *Analysis of 8-oxo-7,8-dihydro-2'-deoxyguanosine by ultra high pressure liquid chromatography-heat assisted electrospray ionization-tandem mass spectrometry, J Chromatogr B Analyt Technol Biomed Life Sci* **2010**, *878*, 375-380.

[127] S. J. Culp, B. P. Cho, F. F. Kadlubar, F. E. Evans, *Structural and conformational analyses of 8-hydroxy-2'-deoxyguanosine, Chem. Res. Toxicol.* **1989**, *2*, 416-422.

[128] J. Wu, M. McKeague, S. J. Sturla, *Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq, J. Am. Chem. Soc.* **2018**, *140*, 9783-9787.

[129] Y. Ding, A. M. Fleming, C. J. Burrows, *Sequencing the Mouse Genome for the Oxidatively Modified Base 8-Oxo-7,8-dihydroguanine by OG-Seq, J. Am. Chem. Soc.* **2017**, *139*, 2569-2572.

[130] N. O. Hudson, B. A. Buck-Koehntop, *Zinc Finger Readers of Methylated DNA, Molecules* **2018**, *23*.

[131] F. R. Traube, T. Carell, *The chemistries and consequences of DNA and RNA methylation and demethylation, RNA Biol* **2017**, *14*, 1099-1107.

[132] S. S. David, V. L. O'Shea, S. Kundu, *Base-excision repair of oxidative DNA damage, Nature* **2007**, *447*, 941-950.

[133] J. W. Drake, *The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes, Ann. N. Y. Acad. Sci.* **1999**, *870*, 100-107.

[134] H. Oda, A. Taketomi, R. Maruyama, R. Itoh, K. Nishioka, H. Yakushiji, T. Suzuki, M. Sekiguchi, Y. Nakabeppu, *Multi-forms of human MTH1 polypeptides produced by alternative translation initiation and single nucleotide polymorphism, Nucleic Acids Res.* **1999**, *27*, 4335-4343.

[135] G. W. Hsu, M. Ober, T. Carell, L. S. Beese, *Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase, Nature* **2004**, *431*, 217-221.

[136] M. Giorgio, G. I. Dellino, V. Gambino, N. Roda, P. G. Pelicci, *On the epigenetic role of guanosine oxidation, Redox. Biol.* **2020**, *29*, 101398.

[137] S. Boiteux, F. Coste, B. Castaing, *Repair of 8-oxo-7,8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the Fpg and OGG1 DNA N-glycosylases, Free Radic. Biol. Med.* **2017**, *107*, 179-201.

[138] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, W. Dean, *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair, Epigenetics & Chromatin* **2013**, *6*, 39-39.

[139] S. Cogoi, A. Ferino, G. Miglietta, E. B. Pedersen, L. E. Xodo, *The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription, Nucleic Acids Res.* **2018**, *46*, 661-676.

[140] R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, S. Balasubramanian, *G-quadruplex structures mark human regulatory chromatin, Nat. Genet.* **2016**, *48*, 1267-1272.

[141] Z.-Y. Sun, X.-N. Wang, S.-Q. Cheng, X.-X. Su, T.-M. Ou, *Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to*

Interfering with Nucleic Acid–Protein Interaction, Molecules **2019**, *24*, 396.

[142] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *High-throughput sequencing of DNA G-quadruplex structures in the human genome, Nat. Biotechnol.* **2015**, *33*, 877-881.

[143] D. J. Patel, A. T. Phan, V. Kuryavyi, *Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, Nucleic Acids Res.* **2007**, *35*, 7429-7455.

[144] A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Stability and kinetics of G-quadruplex structures, Nucleic Acids Res.* **2008**, *36*, 5482-5515.

[145] S. Q. Mao, A. T. Ghanbarian, J. Spiegel, S. Martínez Cuesta, D. Beraldi, M. Di Antonio, G. Marsico, R. Hänsel-Hertsch, D. Tannahill, S. Balasubramanian, *DNA G-quadruplex structures mold the DNA methylome, Nat. Struct. Mol. Biol.* **2018**, *25*, 951-957.

[146] A. M. Fleming, Y. Ding, C. J. Burrows, *Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair, Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2604-2609.

[147] S. Bielskutė, J. Plavec, P. Podbevšek, *Oxidative lesions modulate G-quadruplex stability and structure in the human BCL2 promoter, Nucleic Acids Res.* **2021**, *49*, 2346-2356.

[148] O. Ramon, S. Sauvaigo, D. Gasparutto, P. Faure, A. Favier, J. Cadet, *Effects of 8-oxo-7,8-dihydro-2'-deoxyguanosine on the binding of the transcription factor Sp1 to its cognate target DNA sequence (GC box), Free Radic. Res.* **1999**, *31*, 217-229.

[149] C. Vizcaíno, S. Mansilla, J. Portugal, *Sp1 transcription factor: A long-standing target in cancer chemotherapy, Pharmacol. Ther.* **2015**, *152*, 111-124.

[150] L. Pan, B. Zhu, W. Hao, X. Zeng, S. A. Vlahopoulos, T. K. Hazra, M. L. Hegde, Z. Radak, A. Bacsi, A. R. Brasier, X. Ba, I. Boldogh, *Oxidized Guanine Base Lesions Function in 8-Oxoguanine DNA Glycosylase-1-mediated Epigenetic Regulation of Nuclear Factor κB-driven Gene Expression, J. Biol. Chem.* **2016**, *291*, 25553-25566.

[151] M. K. Hailer-Morrison, J. M. Kotler, B. D. Martin, K. D. Sugden, *Oxidized guanine lesions as modulators of gene transcription. Altered p50 binding affinity and repair shielding by 7,8-dihydro-8-oxo-2'-deoxyguanosine lesions in the NF-kappaB promoter element, Biochemistry* **2003**, *42*, 9761-9770.

[152] Q. Du, P.-L. Luu, C. Stirzaker, S. J. Clark, *Methyl-CpG-binding domain proteins: readers of the epigenome, Epigenomics* **2015**, *7*, 1051-1073.

[153] E. Touati, V. Michel, J. M. Thiberge, P. Avé, M. Huerre, F. Bourgade, A. Klungland, A. Labigne, *Deficiency in OGG1 protects against inflammation and mutagenic effects associated with H. pylori infection in mouse, Helicobacter* **2006**, *11*, 494-505.

[154] H. Sampath, V. Vartanian, M. R. Rollins, K. Sakumi, Y. Nakabeppu, R. S. Lloyd, *8-Oxoguanine DNA glycosylase (OGG1) deficiency increases*

susceptibility to obesity and metabolic dysfunction, PLoS One **2012**, *7*, e51697.

[155] G. V. Mechetin, A. V. Endutkin, E. A. Diatlova, D. O. Zharkov, *Inhibitors of DNA Glycosylases as Prospective Drugs, Int. J. Mol. Sci.* **2020**, *21*, 3118.

[156] C. A. Hammett-Stabler, U. Garg, *The evolution of mass spectrometry in the clinical laboratory, Methods Mol Biol* **2010**, *603*, 1-7.

[157] D. S. Millington, N. Kodo, D. L. Norwood, C. R. Roe, *Tandem mass spectrometry: A new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism, Journal of Inherited Metabolic Disease* **1990**, *13*, 321-324.

[158] G. Chowdhury, F. P. Guengerich, in *Encyclopedia of Analytical Chemistry*, **2015**, pp. 1-39.

[159] E. Dudley, L. Bond, *Mass spectrometry analysis of nucleosides and nucleotides, Mass Spectrom Rev* **2014**, *33*, 302-331.

[160] J. Dong, **2009**.

[161] C. M. Alymatiri, M. G. Kouskoura, C. K. Markopoulou, *Decoding the signal response of steroids in electrospray ionization mode (ESI-MS), Analytical Methods* **2015**, *7*, 10433-10444.

[162] P. Lössl, *A highly charged topic, Nature Physics* **2020**, *16*, 368-368.

[163] J. Faktor, M. Dvorakova, J. Maryas, I. Procházková, P. Bouchal, *Identification and characterisation of pro-metastatic targets, pathways and molecular complexes using a toolbox of proteomic technologies, Klinická onkologie : casopis Ceské a Slovenské onkologické spolecnosti* **2012**, *25 Suppl 2*, 2S70-77.

[164] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo, Nat. Chem. Biol.* **2018**, *14*, 72-78.

[165] A. Kohler, A. Kuo, L. G. Nagy, E. Morin, K. W. Barry, F. Buscot, B. Canbäck, C. Choi, N. Cichocki, A. Clum, J. Colpaert, A. Copeland, M. D. Costa, J. Doré, D. Floudas, G. Gay, M. Girlanda, B. Henrissat, S. Herrmann, J. Hess, N. Högberg, T. Johansson, H. R. Khouja, K. LaButti, U. Lahrmann, A. Levasseur, E. A. Lindquist, A. Lipzen, R. Marmeisse, E. Martino, C. Murat, C. Y. Ngan, U. Nehls, J. M. Plett, A. Pringle, R. A. Ohm, S. Perotto, M. Peter, R. Riley, F. Rineau, J. Ruytinx, A. Salamov, F. Shah, H. Sun, M. Tarkka, A. Tritt, C. Veneault-Fourrey, A. Zuccaro, A. Tunlid, I. V. Grigoriev, D. S. Hibbett, F. Martin, *Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists, Nat. Genet.* **2015**, *47*, 410-415.

[166] M. Y. Liu, J. E. DeNizio, R. M. Kohli, in *Methods in Enzymology, Vol. 573* (Ed.: R. Marmorstein), Academic Press, **2016**, pp. 365-385.

[167] K. Hochedlinger, R. Jaenisch, *Induced Pluripotency and Epigenetic Reprogramming, Cold Spring Harb. Perspect. Biol.* **2015**, *7*.

[168] Y. Gao, J. Chen, K. Li, T. Wu, B. Huang, W. Liu, X. Kou, Y. Zhang, H. Huang, Y. Jiang, C. Yao, X. Liu, Z. Lu, Z. Xu, L. Kang, J. Chen, H. Wang, T. Cai, S. Gao, *Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming, Cell Stem Cell* **2013**, *12*, 453-469.

[169] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing, Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.

[170] F. Schelter, A. Kirchner, F. R. Traube, M. Müller, W. Steglich, T. Carell, *5-Hydroxymethyl-, 5-Formyl- and 5-Carboxydeoxycytidines as Oxidative Lesions and Epigenetic Marks, Chem. Eur. J.* **2021**, *27*, 8100-8104.

[171] J.-B. Alberge, F. Magrangeas, M. Wagner, S. Denié, C. Guérin-Charbonnel, L. Campion, M. Attal, H. Avet-Loiseau, T. Carell, P. Moreau, S. Minvielle, A. A. Sérandour, *DNA hydroxymethylation is associated with disease severity and persists at enhancers of oncogenic regions in multiple myeloma, Clin. Epigenetics* **2020**, *12*, 163.

[172] K. Chen, J. Zhang, Z. Guo, Q. Ma, Z. Xu, Y. Zhou, Z. Xu, Z. Li, Y. Liu, X. Ye, X. Li, B. Yuan, Y. Ke, C. He, L. Zhou, J. Liu, W. Ci, *Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer, Cell Res.* **2016**, *26*, 103-118.

[173] L. Zhang, W. Chen, L. M. Iyer, J. Hu, G. Wang, Y. Fu, M. Yu, Q. Dai, L. Aravind, C. He, *A TET Homologue Protein from Coprinopsis cinerea (CcTET) That Biochemically Converts 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine, J. Am. Chem. Soc.* **2014**, *136*, 4801-4804.

[174] M. Guo, L. Zhang, Y. Du, W. Du, D. Liu, C. Guo, Y. Pan, D. Tang, *Enrichment and Quantitative Determination of 5-(Hydroxymethyl)-2'-deoxycytidine, 5-(Formyl)-2'-deoxycytidine, and 5-(Carboxyl)-2'-deoxycytidine in Human Urine of Breast Cancer Patients by Magnetic Hyper-Cross-Linked Microporous Polymers Based on Polyionic Liquid, Anal. Chem.* **2018**, *90*, 3906-3913.

[175] I. Barbieri, T. Kouzarides, *Role of RNA modifications in cancer, Nat. Rev. Cancer* **2020**, *20*, 303-322.

[176] D. Hanahan, R. A. Weinberg, *Hallmarks of cancer: the next generation, Cell* **2011**, *144*, 646-674.

[177] B. S. Zhao, I. A. Roundtree, C. He, *Post-transcriptional gene regulation by mRNA modifications, Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 31-42.

[178] Y. Liu, W. Gao, D. Zhang, *Effects of cigarette smoke extract on A549 cells and human lung fibroblasts treated with transforming growth factor-beta1 in a coculture system, Clin. Exp. Med.* **2010**, *10*, 159-167.

[179] A. R. Seddon, Y. Liau, P. E. Pace, A. L. Miller, A. B. Das, M. A. Kennedy, M. B. Hampton, A. J. Stevens, *Genome-wide impact of hydrogen peroxide on maintenance DNA methylation in replicating cells, Epigenetics & Chromatin* **2021**, *14*, 17.

[180] I. Millán, J. D. Piñero-Ramos, I. Lara, A. Parra-Llorca, I. Torres-Cuevas, M. Vento, *Oxidative Stress in the Newborn Period: Useful Biomarkers in the Clinical Setting, Antioxidants* **2018**, *7*, 193.

[181] D. K. Chu, L. H. Kim, P. J. Young, N. Zamiri, S. A. Almenawer, R. Jaeschke, W. Szczeklik, H. J. Schünemann, J. D. Neary, W. Alhazzani, *Mortality and morbidity in acutely ill adults treated with liberal versus conservative oxygen therapy (IOTA): a systematic review and meta-analysis, Lancet.* **2018**, *391*, 1693-1705.

[182] M. Girardis, S. Busani, E. Damiani, A. Donati, L. Rinaldi, A. Marudi, A. Morelli, M. Antonelli, M. Singer, *Effect of Conservative vs Conventional Oxygen Therapy on Mortality Among Patients in an Intensive Care Unit: The Oxygen-ICU Randomized Clinical Trial, Jama.* **2016**, *316*, 1583-1589.

[183] L. Amarelle, L. Quintela, J. Hurtado, L. Malacrida, *Hyperoxia and Lungs: What We Have Learned From Animal Models, Front. Med.* **2021**, *8*.

[184] A. M. Dylag, J. Haak, M. Yee, M. A. O'Reilly, *Pulmonary mechanics and structural lung development after neonatal hyperoxia in mice, Pediatr. Res.* **2020**, *87*, 1201-1210.

[185] E. R. Weibel, *Oxygen Effect on Lung Cells, Arch. Intern. Med.* **1971**, *128*, 54-56.

[186] R. Solberg, J. H. Andresen, R. Escrig, M. Vento, O. D. Saugstad, *Resuscitation of Hypoxic Newborn Piglets With Oxygen Induces a Dose-Dependent Increase in Markers of Oxidation, Pediatr. Res.* **2007**, *62*, 559-563.

[187] X. Zhang, E. Gonzalez Rodriguez, M. Markmann, P. Oak, A. Hilgendorff, *Expression profiling of genes differentially regulated under hyperoxia andmechanical ventilation in mouse model of neonatal chronic lung disease, Eur. Respir. J.* **2019**, *54*, PA4113.

[188] M. G. Krokidis, E. Parlanti, M. D'Errico, B. Pascucci, A. Pino, A. Alimonti, D. Pietraforte, A. Masi, C. Ferreri, C. Chatgilialoglu, *Purine DNA Lesions at Different Oxygen Concentration in DNA Repair-Impaired Human Cells (EUE-siXPA), Cells* **2019**, *8*, 1377.

[189] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides, Nat. Protoc.* **2019**, *14*, 283-312.

[190] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry, Nat. Protoc.* **2008**, *3*, 1287-1298.

Appendix A: 5-hydroxymethyl-, 5-formyl- and 5-carboxydeoxycytidines as oxidative lesions and epigenetic marks

Chemistry–A European Journal

Supporting Information

5-Hydroxymethyl-, 5-Formyl- and 5-Carboxydeoxycytidines as Oxidative Lesions and Epigenetic Marks

Florian Schelter[+], Angie Kirchner[+], Franziska R. Traube, Markus Müller, Wolfgang Steglich, and Thomas Carell*

Table of contents

Supporting Figure

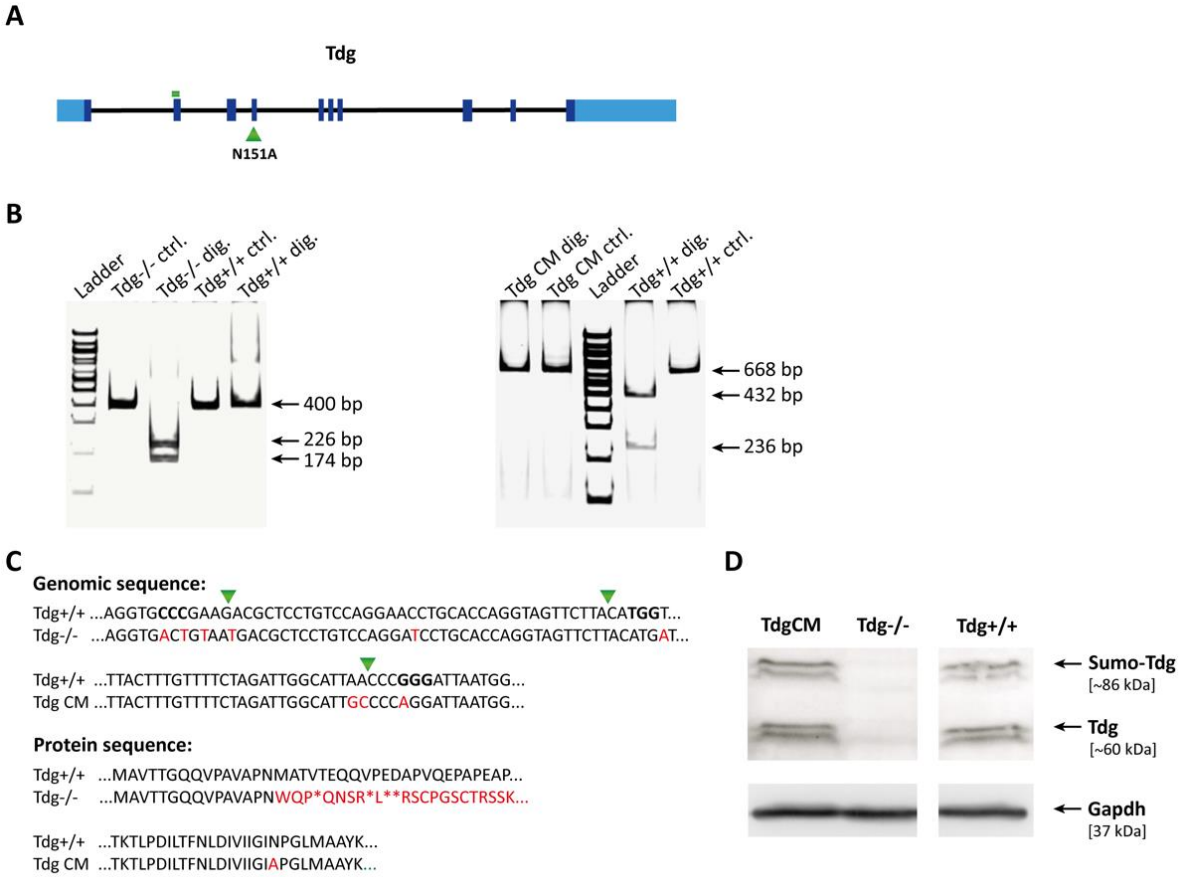


Figure S1: Generation and validation of Tdg$^{-/-}$ and Tdg^{CM} mESCs. A) Schematic representation of the mouse Tdg gene body. Exons are represented by blue boxes and UTR in lighter colour. The positions of gRNAs are indicated by a green rectangle and triangle, respectively. **B**) PAGE-analysis after amplification of the target locus and restriction digest. Insertion of the repair template leads to gain of a *BamHI* (NEB) restriction site and frameshift mutation. Insertion of the catalytic mutant repair template leads to a loss of the *SmaI* (NEB) site. Undigested sample (Sample ctrl.), digested sample (Sample dig.), Ladder (100 bp DNA ladder, NEB); **C**) Genomic sequences of Tdg$^{-/-}$ and Tdg^{CM} alleles, as well as respective protein sequences. Antisense strands are shown so as to highlight the PAM sequences (bold in wild type sequence). The expected Cas9 cut sites are indicated by green triangles. Alteration to the genomic DNA and protein sequence are highlighted in red, whereby the asterix represents a stop codon. **D**) Western blot with an anti-Tdg antibody on whole cell extracts from Tdg$^{+/+}$ and CRISPR-edited mESCs. An anti-Gapdh antibody was used as loading control. The observed double band may result from an isoform or a protein degradation product.[1]

Cell culture

All cells were routinely cultured at 37 °C in water saturated, CO_2-enriched (5%) atmosphere and regularly tested for Mycoplasma contamination using the Mycoplasma Detection Kit (JenaBioscience PP-401L). HEK 293T cells (ATCC, cat. no. CRL-3216) were maintained in DMEM medium (Merck) containing 10% (v/v) fetal bovine serum (FBS, Invitrogen), 1% (v/v) *L*-alanyl-*L*-glutamine (Merck) and 1% (v/v) penicillin-streptomycin (Merck). Human induced pluripotent stem cells (iNGNs; hiPSCreg no. HVRDi004-B-1)[2] were cultured on Geltrex (Thermo Fisher) coated culture vessels as described.[3] Cells were regularly maintained in E9 medium containing DMEM medium, Nutrient Mix Hams F-12 (Merck), 2 mM *L*-alanyl-*L*-glutamine, 64 mg/L *L*-ascorbic acid 2-phosphate (Merck), 1x Penicillin-Streptomycin (Merck), 0.1 mM nicotinamide, 77.6 nM sodium selenite, 11.18 g/L sodium chloride and 10 μg/mL holo-transferrine (Merck) and 20 μg/mL insulin (Merck). After sterile filtration (0.22 μM, Merck), E9 was completed by addition of human FGF-2 100 ng/mL (PeproTech) and human TGF-ß1 2 ng/mL (PeproTech). Mouse embryonic stem cells (Kindlin3$^{+/+}$, M. Moser, Max Planck Institute for Biochemistry)[4] were cultured in basal medium for mESCs on gelatin (0.2% (w/v) solution, Merck) coated culture vessels as described previously.[5] mESC lines were routinely maintained in basal medium supplemented with 1000 U/mL LIF (ORF Genetics), 3 µM CHIR99021 (Selleckchem) and 7.5 µM CGP77675 (Axon Medchem). For LC-MS analysis, mESC basal medium was supplemented exclusively with 1000 U/mL LIF and cells were cultured for four days with routine passaging on the second day.

Generation and validation of CRISPR-edited mESCs

Guide RNAs (gRNAs) were designed over exon 2 and 4 of Tdg, respectively (Figure S1), using the CRISPR design tool (http://crispr.mit.edu/). Oligonucleotides used for construction of dual Cas9 and gRNA expression vectors are listed in Table S1. The oligonucleotides were annealed and cloned in the *BbsI* site of pSpCas9-2A-Puro (PX459; Addgene Plasmid 48139).[6]

All synthetic oligonucleotides were purchased from Merck or IDT.

Table S1: Oligonucleotides used in this study.

Name	**Sequence**	**Application**
mTdg_-/- gRNA FF	5´-CACCGTGCACCAGGTAGTTCTTACA-3´	*guide* RNA
mTdg_-/- gRNA FR	5´-AAACTGTAAGAACTACCTGGTGCAC-3´	*guide* RNA
mTdg_-/- gRNA RF	5´-CACCGTCCTGGACAGGAGCGTCTTC-3´	*guide* RNA
mTdg_-/- gRNA RR	5´-AAACGAAGACGCTCCTGTCCAGGAC-3´	*guide* RNA
mTdg_CM gRNA F	5´-CACCGTTCTAGATTGGCATTAACCC-3´	*guide* RNA
mTdg_CM gRNA R	5´-AAACGGGTTAATGCCAATCTAGAAC-3´	*guide* RNA
mTdg_-/- FWD	5´-TTGTTCTTTCAGCTATTCTCTGGA-3´	PCR
mTdg_-/- REV	5´-CCCACCCCCTGTTGTTCAC-3´	PCR
mTdg CM FWD	5´-ATTAGCCTTACTTGGCAGGGG-3´	PCR
mTdg CM REV	5´-ACTCCGCTATTACCAACGCT-3´	PCR
mTdg-/-	5´-CCAGCAGTAGCTCCTAACATGGCAACCGT GACAGAACAGCAGGTGACTGTAATGACG CTCCTGTCCAGGATCCTGCACCAGGTAGTT CTTACATGATCTCCCTTGAATGTCTCCGCCT CCAGCAGAGCAATCAGTGAACAC-3´	Repair template
mTdg CM	5´-GTCTGAGTTTTTAAAAACAAAAGTGTTCA ATTACTTTGTTTTCTAGATTGGCATTGCCC CAGGATTAATGGCTGCTTACAAAGGACAT CACTACCCTGGGCCT-3´	Repair template

Transfection

mESCs were routinely cultured as described. Cas9-sgRNA expression vectors and respective repair templates targeting Tdg were cotransfected in kindlin3$^{+/+}$ mESCs using Lipofectamine 2000 (Life Technologies) according to the manufacturer's recommendations. To minimize non-homologous end joining events, the inhibitor Scr7 (1 μM, Merck) was added to the culture medium until subcloning.[7] 24 h post transfection, cultures were selected with 0.5 μg/mL puromycin for 48 h and subcloned by limiting dilution.

Screen and sequencing of clonal populations

Clonal polpulations were lysed in QuickExtraction buffer (Epicentre) according to the manufacturer. Amplification of the target locus was achieved by a standard procedure using HotStarTaq (Qiagen). Successfully targeted populations were identified by digestion of the amplification product with either *BamHI* (Tdg$^{-/-}$) or SmaI (Tdg^{CM}; both purchased from NEB), as recommended. In addition, positive clones were analysed through sanger sequencing (Eurofins Genomics).

Western blot

For further analysis of CRISPR-edited mESCs, whole-cell extracts (RIPA buffer: 10 mM TRIS pH 7.4; 0.5 mM EDTA; 1% Triton-X; 1% deoxycholate; 150 mM sodium chloride; 2.5 mM magnesium chloride; 1 mM DTT; 1x protease inhibitor tablet, Roche) were probed with a mouse anti-TDG (1:200, D-11, sc-376652) and a mouse anti-GAPDH (1:10000, 6C5, sc-32233, both purchased from Santa Cruz Biotechnology) antibody after standard western blot procedures. As secondary antibody an HRP-conjugated goat antibody (1:10000, A3682, Merck) was used. For visualization, the SuperSignal West Pico Chemiluminescent Substrate Kit (Thermo Scientific; 34077) was used according to the manufacturer's recommendations.

Harvest and genomic DNA isolation from Amanita muscaria

The mushrooms were harvested during fall 2020 in the woods around munich (one mushroom per colony) (see Figure S2). After harvest, they were washed once with soap and twice with destilled water to remove bacteria and small animals. Afterwards they were wrapped in paper towels and dried in warm air (30°C for 10 min).

Figure S2: Amanita muscaria from three different colonies.

For genomic DNA isolation, stem and cap of three different mushrooms were separated and peeled to remove residual contaminations before they were freeze-dried. Then the mushroom (5 g) was crushed for 5 min at 30 Hz by a ball mill (Tissue Lyzer Qiagen). 200 mg of powder was used for the DNA-isolation. Extraction of the DNA was performed according to work published earlier.[3] All buffers were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 µM) and deferoxamine mesylate salt (Desferal, 200 µM) to reduce background oxidation and deamination.[3, 8]

Derivatization and digestion of genomic DNA

Derivatization of genomic DNA of the mushroom was performed according to an earlier published work.[9]

As exemplified in Table S2, we used the following chemicals for the analysis of our samples. Reaction buffer 10X and Enzyme mix was bought as a Nucleoside Digestion Mix (M0649S) kit (New England BioLabs Inc.). The nucleoside mix was prepared from heavy labelled mdC (10.2 µM), hmdC (0.2 µM), fdC (0.005 µM), cadC (0.004 µM) and 8-oxo-dG (0.022 µM). Furthermore, we spiked certain amounts of heavy labelled derivatized-AP-sites product and derivatized-β-Elimination product for later quantification of the AP-sites.[9] The concentration of the DNA for the digestion was received from the Nanodrop. The total volume of 50 µL was then incubated at 37 °C for 1.5 hours. Afterwards the samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences) before applying to LC-MS/MS.

Table S1: Digestion sheet of the gDNA samples.

	Sample	m (DNA) [ng] dissolved in water	V (Reaction buffer 10X) [µL]	V (Enzyme mix) [µL]	V (Nucleosides mix) [µL]	derivatized-AP (d9) [µL]	derivatized -β-Elimination (d9) [µL]	Total Volume [µL]	1 h 30 min 37 °C	Injection volume [µL]
Blank 1	1	0	5.0	2.0	3.0	0.31	0.394	50.0		39
Blank 2	2	0	5.0	2.0	3.0	0.31	0.394	50.0		39
Control mushroom stem	3	3000 / 3500	5.0	2.0	3.0	0.31	0.394	50.0		39
Control mushroom cap	4	3000 / 3500	5.0	2.0	3.0	0.31	0.394	50.0		39
Mushroom stem	5	3000	5.0	2.0	3.0	0.31	0.394	50.0		39
Mushroom cap	6	3000	5.0	2.0	3.0	0.31	0.394	50.0		39

LC-MS/MS measurements

All experiments were performed in biological and technical triplicates to ensure reproducibility. Solvents for LC-MS/MS analysis were purchased from Honeywell, Roth and used without further purification.

The analysis was performed using an UHPLC-QQQ-MS/MS system consisting of a Triple Quadrupole 6490 mass spectrometer (Agilent) with an ESI source and an Agilent Infinity 1290 UHPLC. The elution was monitored at 260 nm (Agilent InfinityLab Deuterium Lamp G1314). Data Acquisition and processing were performed using MassHunter Workstation Software Version B.07.01 (Agilent).

mdC, hmdC, fdC, cadC and 8oxodG were seperated on an InfinityLab Poroshell 120 SB-C8 column (2.1 mm x 150 mm, 2.7 µm, Agilent Technologies, USA) at 35 °C. Water containing 0.0085% FA (v/v, solvent A) and MeCN containing 0.0085% FA (v/v, solvent B) was used as the mobile phase. A gradient of 0 - 3.5% B for 4 min, 3.5 - 5% B for 2.9 min, 5 - 80% B for 0.3 min, 80% B for 3.3 min was used. The flow rate of the mobile phase was set to 0.35 mL min^{-1}.[3]

The derivatized abasic sites were separated on an InfinityLab Poroshell 120 SB-C8 column (2.1 mm x 150 mm, 2.7 µm, Agilent Technologies, USA) at 35 °C. Water containing 0.0085% formic acid (FA) (v/v, solvent A) and MeCN containing 0.0085% FA (v/v, solvent B) was used as the mobile phase. A gradient of 0 - 3.5 % B for 5 min, 3.5 - 5% B for 1.9 min, 5 - 80% B for 6.2 min, 80% B for 1.2 min, 80 – 5% B for 2.7 min and 0% B for 1.7 min was used. The flow rate of the mobile phase was set to 0.35 mL min^{-1}.[9]

The source-dependent parameters were as follows, gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N_2), capillary voltage 2,500 V in the positive ion mode, and nozzle voltage 500 V. Delta EMV was set to 500 (positive mode). The nucleosides and labelled products were monitored using the multiple reaction monitoring (MRM) mode. The MRM parameters were optimized to achieve maximal detection sensitivity (Tables S3 and S4).

Table S2: MRM parameters for the detection of mdC, hmdC, fdC, cadC and 8oxodG.

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity	Collision Energy	Fragmentor
UV	1	0,0 -> 0,0	MRM	Target	0	0	6	Positive		
mC-dN-d3	3	245.1 -> 129.1	MRM	ISTD	245,1	129,1	3,472	Positive	1	380
mC-dN	3	242.1 -> 126.1	MRM	Target	242,1	126,1	3,533	Positive	1	380
hmC-dN-d2-15n2	2	262.1 -> 146.1	MRM	ISTD	262,1	146,1	2,082	Positive	2	380
hmC-dN	2	258.1 -> 142.1	MRM	Target	258,1	142,1	2,121	Positive	2	380
fC-dN-15n2	5	258.1 -> 142.0	MRM	ISTD	258,1	142	6	Positive	5	380
fC-dN	5	256.1 -> 140.1	MRM	Target	256,1	140,1	6	Positive	5	380
fC	5	140.0 -> 97.0	MRM	Target	140	97	6	Positive	13	380
dG-UV	3	0,0 -> 0,0	MRM	Target	0	0	6,3	Positive		
dC-UV	3	0.0 -> 0.0	MRM	Target	0	0	1,838	Positive		
caC-dN-15n2	3	274.1 -> 158.0	MRM	ISTD	274,1	158	3,195	Positive	5	380
caC-dN	3	272.1 -> 156.0	MRM	Target	272,1	156	3,214	Positive	5	380
caC	3	156.0 -> 138.0	MRM	Target	156	138	3,21	Positive	13	380
8oxoG-dN-15n5	5	289.1 -> 173.0	MRM	ISTD	289,1	173	6,5	Positive	9	380
8oxoG-dN	5	284.1 -> 168.1	MRM	Target	284,1	168,1	6,5	Positive	9	380

Table S3: MRM parameters for the derivatized AP-sites.

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity	Collision Energy	Fragmentor
Superfly	5	362,2 -> 334,2	MRM	Target	362,2	334,2	8,826	Positive	19	380
Superfly-AP_1_d9	5	487,3 -> 459,3	MRM	ISTD	487,3	459,3	9,583	Positive	19	380
Superfly-AP_1	5	478,2 -> 450,2	MRM	Target	478,2	450,2	10,856	Positive	19	380
Superfly-AP_2_d9	5	487,3 -> 201,2	MRM	ISTD	487,3	201,2	9,583	Positive	40	380
Superfly-AP_2	5	478,2 -> 192,1	MRM	Target	478,2	192,1	11,688	Positive	40	380
Superfly-beta-Eli_1_d9	5	469,3 -> 441,3	MRM	ISTD	469,3	441,3	9,84	Positive	19	380
Superfly-beta-Eli_1	5	460,2 -> 432,2	MRM	Target	460,2	432,2	9,917	Positive	19	380
Superfly-beta-Eli_2_d9	5	469,3 -> 201,2	MRM	ISTD	469,3	201,2	9,84	Positive	33	380
Superfly-beta-Eli_2	5	460,2 -> 192,1	MRM	Target	460,2	192,1	9,926	Positive	34	380
Superfly-Naphtyl	5	500,2 -> 472,2	MRM	Target	500,2	472,2	10,895	Positive	19	380

Literature

[1] S. Um, M. Harbers, A. Benecke, B. t. Pierrat, R. Losson, P. Chambon, *Journal of Biological Chemistry* **1998**, *273*, 20728-20736.

[2] V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, *Mol Syst Biol* **2014**, *10*, 760.

[3] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nature Protocols* **2019**, *14*, 283-312.

[4] M. Moser, B. Nieswandt, S. Ussar, M. Pozgajova, R. Fässler, *Nat Med* **2008**, *14*, 325-330.

[5] F. Spada, S. Schiffers, A. Kirchner, Y. Zhang, G. Arista, O. Kosmatchev, E. Korytiakova, R. Rahimoff, C. Ebert, T. Carell, *Nature Chemical Biology* **2020**, *16*, 1411-1419.

[6] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, F. Zhang, *Nature protocols* **2013**, *8*, 2281-2308.

[7] G. Li, X. Zhang, C. Zhong, J. Mo, R. Quan, J. Yang, D. Liu, Z. Li, H. Yang, Z. Wu, *Scientific Reports* **2017**, *7*, 8943.

[8] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nature protocols* **2008**, *3*, 1287-1298.

[9] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *Journal of the American Chemical Society* **2017**, *139*, 10359-10364.

Appendix B: Analysis of an active deformylation mechanism of 5-formyl-2'-deoxycytidine (fdC) in stem cells

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Analysis of an Active Deformylation Mechanism of 5-Formyldeoxycytidine (fdC) in Stem Cells

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

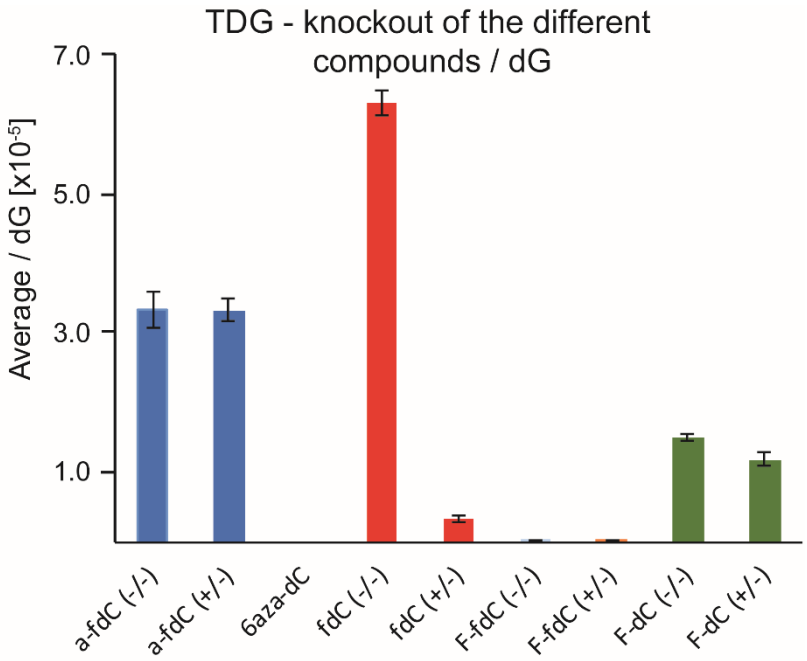
anie_202000414_sm_miscellaneous_information.pdf

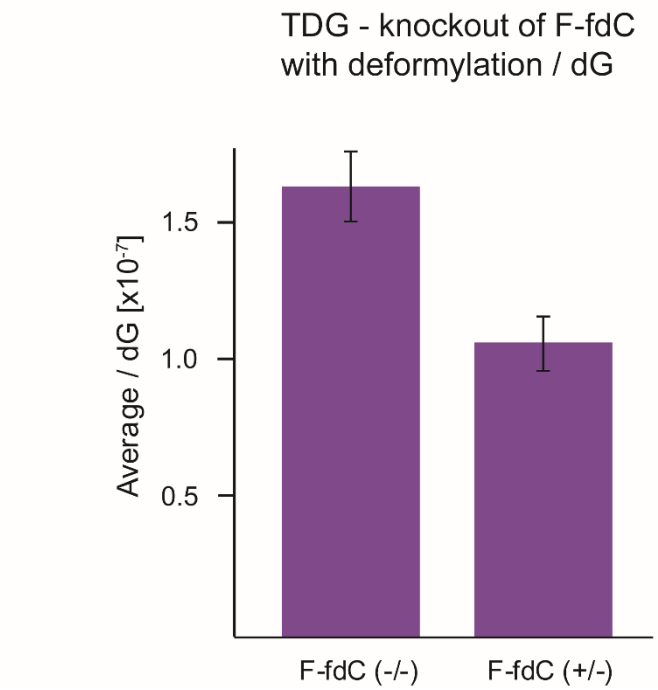
Supporting information

Table of contents

Supplementary figures

Table 1: TDG knockout experiments of the fed nucleosides





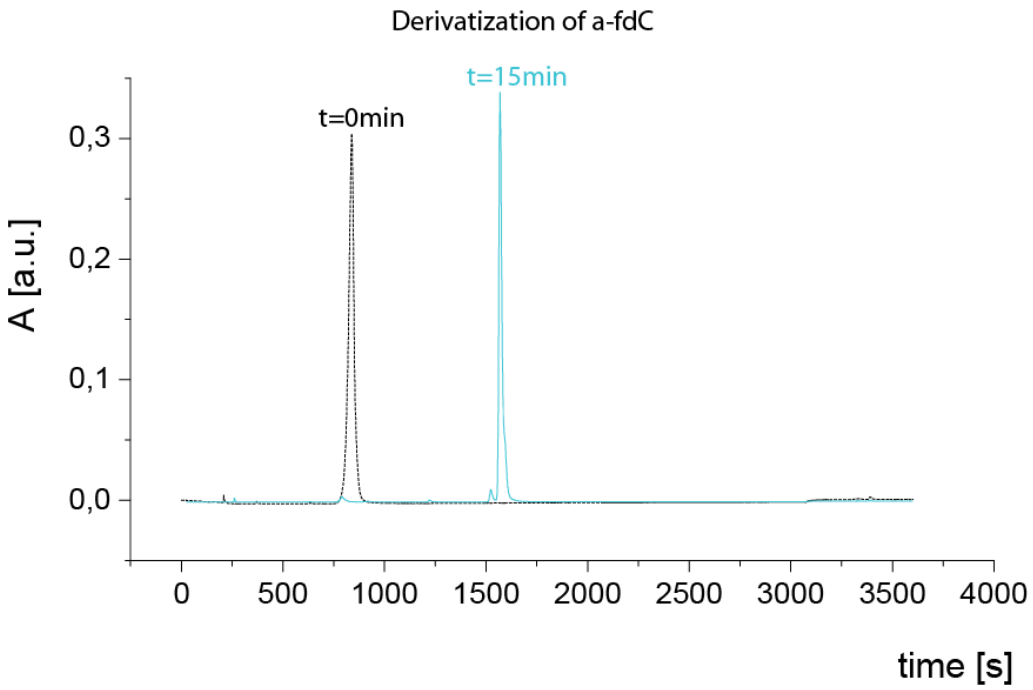


Figure 1: Quantitative derivatization of a-fdC with $MeONH_2$ after 15 min.

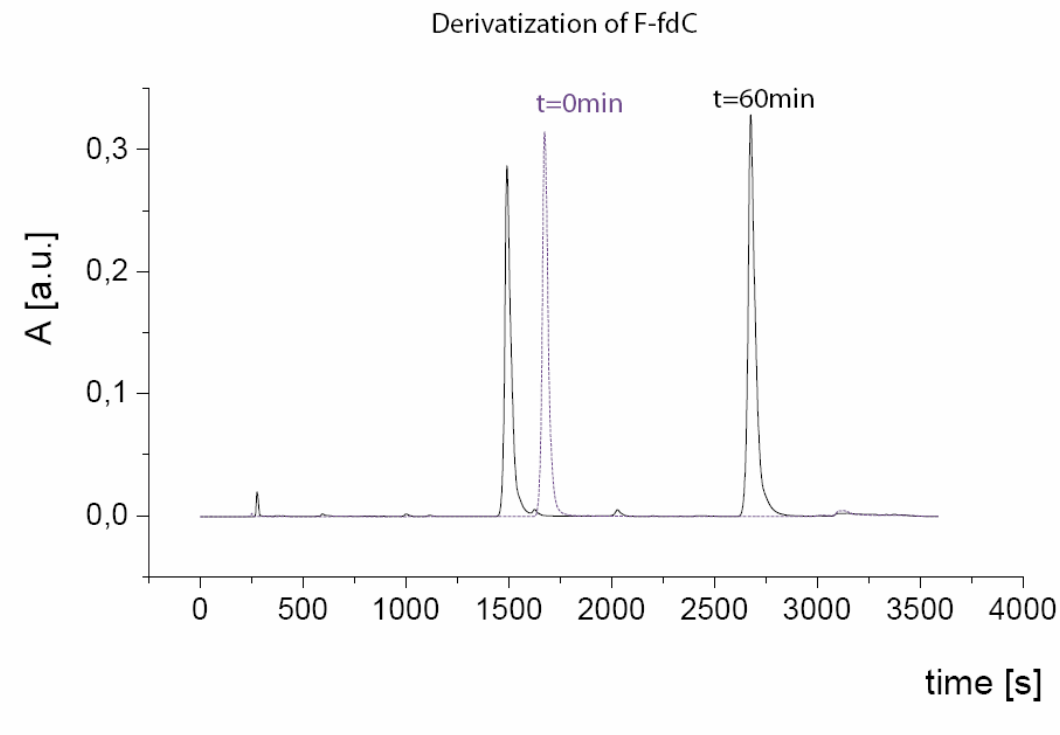


Figure 2: Derivatisation of F-fdC after 60 min, mixture of product and starting material.

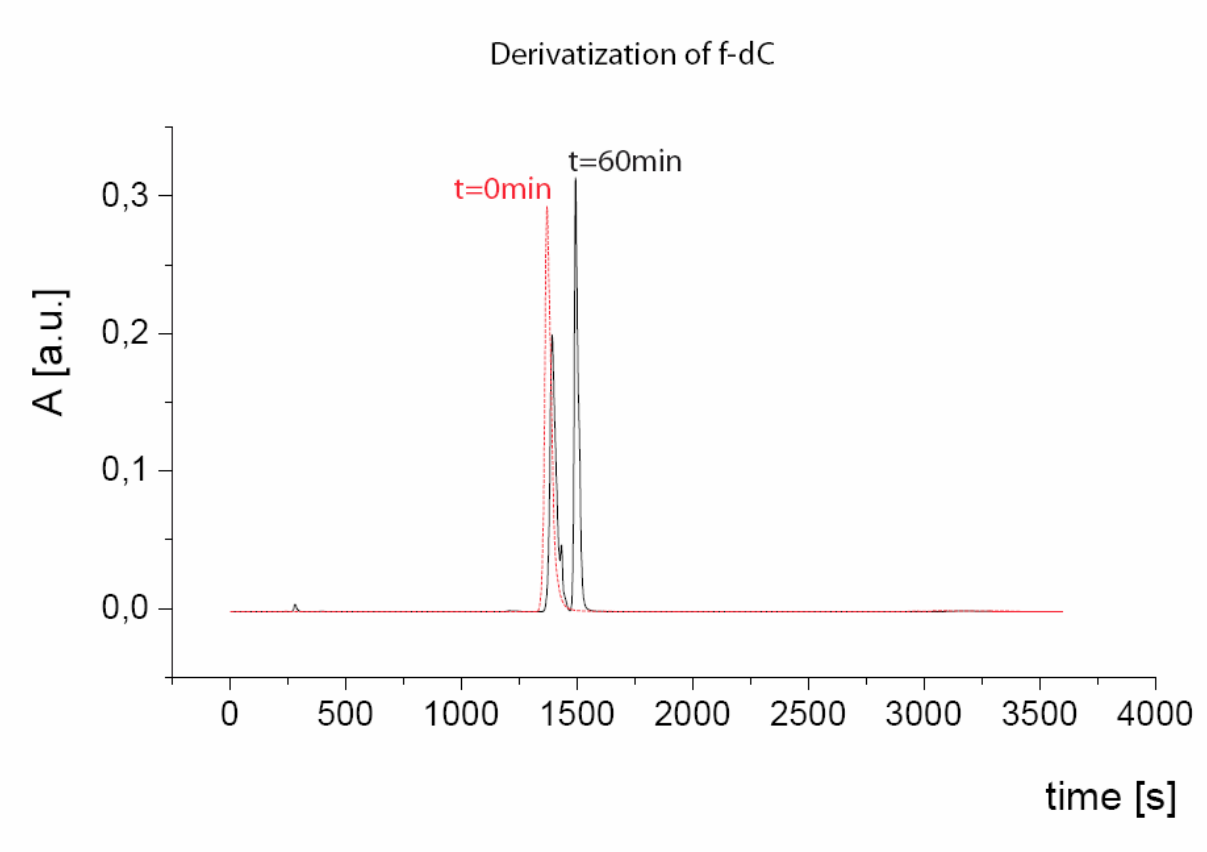


Figure 3: Derivatisation of fdC after 60 min, mixture of product and starting material.

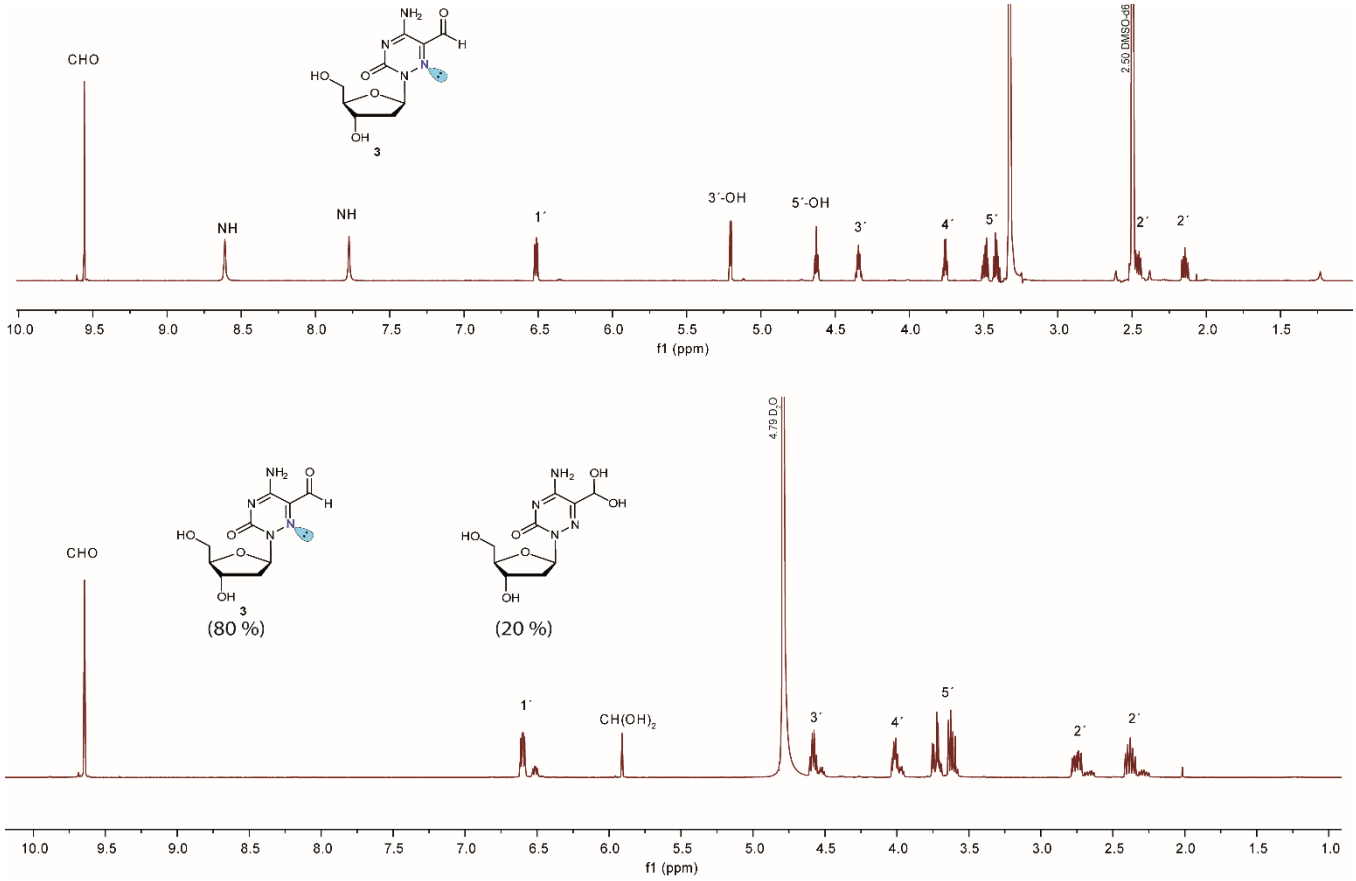


Figure 4: NMR spectra of a-fdC showing the hydrate formation of a-fdC in D_2O. Upper panel: NMR in DMSO-d6. Lower panel: NMR in D_2O.

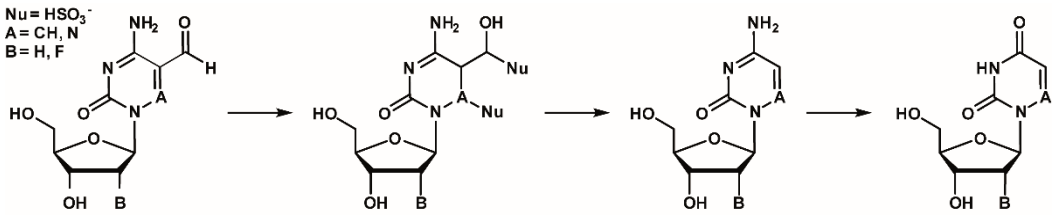


Figure 5: Proposed reaction mechanism with bisulfite

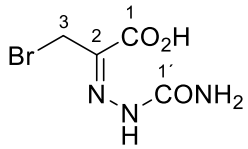
General Methods

Chemicals were purchased from *Sigma-Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros Organics* and used without further purification. Solvents of reagent grade were purified by distillation. Reactions and column chromatography fractions were monitored by thin-layer chromatography (TLC) on silica gel F254TLC plates from *Merck KGaA*. Flash column chromatography was performed on *Geduran*® Si60 (40-63 µm) silica gel from *Merck KGaA* applying slight nitrogen pressure. Chemical transformations were conducted under nitrogen or argon atmosphere in oven-dried glassware unless otherwise specified. NMR spectra were recorded on *Bruker AVIIIHD 400* (400 MHz) or *Varian NMR-System600* (600 MHz) spectrometers. ^{1}H-NMR shifts were calibrated to the residual solvent resonances: $CDCl_3$ (7.26 ppm), DMSO-d6 (2.50 ppm) and D_2O (4.79 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: $CDCl_3$ (77.16 ppm), DMSO-d6 (39.52 ppm). All NMR spectra were analysed using the program *MestReNova* 10.0.1 from *Mestrelab Research S. L*. Low resolution mass spectra were measured on a LT Q FT-ICR by *Thermo Finnigan GmbH*. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwigs-Maximilians-Universität München on a MAT 90 (ESI) from *Thermo Finnigan GmbH*. IR spectra were recorded on a *PerkinElmer Spectrum BX II FT-IR* system. Substances were applied as a film or directly as solids on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC *Waters Alliance* (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column *Nucleosil* 120-2 C18 from *Macherey Nagel* applying eluent flow of 0.5 mL/min. Preparative RP-HPLC was performed on a HPLC *Waters Breeze* (2487 Dual λArray Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from *Macherey Nagel*. A flowrate of 5 mL/min was applied.

Synthetic Procedures

Synthesis of 6-Aza-5-formyl-2´-deoxycytidine 3

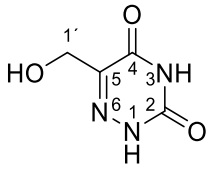
3-Bromopyruvic acid semicarbazone **5**



A solution of semicarbazide hydrochloride (3.34 g, 30.0 mmol, 1.0 eq.) and NaOAc (3.20 g, 38.6 mmol, 1.3 eq.) in 25.0 mL ddH_2O was added to a solution of 5.00 g 3-bromopyruvic acid (30.0 mmol, 1.0 eq.) in 15.0 mL glacial HOAc and 5.0 mL ddH_2O at 0 °C. The mixture was allowed to stand at room temperature for 2.5 h and at 0 °C for 19 h. The resulting white precipitate was filtered, washed with ice cold ddH_2O and Et_2O and dried subsequently under high vacuum to yield 3.26 g of the semicarbazone **5** as a mixture of (E)- and (Z)-isomers (14.6 mmol, 49 %) as a colorless solid.

^{1}H-NMR (400 MHz, DMSO-*d6*): δ/ppm = 10.58 (s, 0.6H, NN**H**) , 10.52 (s, 0.4H, NN**H**), 7.39 (br s, 0.8H, CON**H_2**), 6.92 (br s, 1.3H, CON**H_2**), 4.57 (s, 1.3H, 3-H), 4.46 (s, 0.8H, 3-H). **^{13}C-NMR** (101 MHz, DMSO-*d6*): δ/ppm = 163.51 (1-C), 155.63 (1´-C), 132.48 (2-C), 32.48 (3-C). **IR (ATR):** ν (cm^{-1}) = 3462 (w), 3245 (w), 2394 (br w), 1894 (br w), 1693 (m), 1445 (m), 1421 (s), 1364 (m), 1297 (m), 1234 (w), 1201 (s), 1164 (m), 1143 (m), 992 (m), 821 (s), 695 (m), 658 (s).

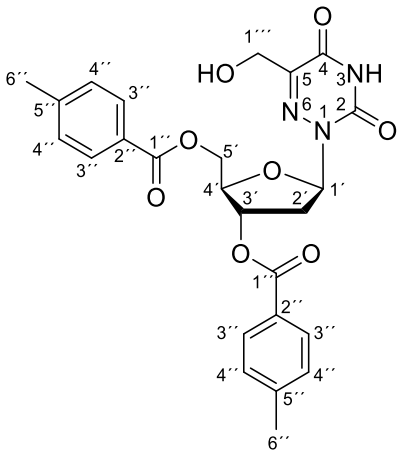
5-Hydroxymethyl-6-azauracil **6**



Compound **6** was synthesized according to a modified procedure of *Alekseeva et al.*[1] Under Ar atmosphere 1.40 g semicarbazone **5** (6.2 mmol, 1.0 eq.) were suspended in 50.0 mL $SOCl_2$ and two drops of pyridine were added. The reaction mixture was refluxed at 80 °C for 75 min, allowed to cool to room temperature, filtered through Celite and concentrated *in vacuo* to approximately 15 mL. Crystallization at 0 °C for 4 d resulted in a yellow precipitate which was filtered, dissolved in 30 ml dry CH_2Cl_2 and concentrated to dryness. After drying under high vacuum the obtained yellow solid was suspended in 20.0 mL ddH_2O and refluxed at 110 °C for 17 h. Subsequently, the solution was concentrated *in vacuo*, the residue was dissolved in 15.0 mL ddH_2O and lyophillized to yield 657 mg 5-hydroxy-6-azauracil **6** (4.60 mmol, 74 %) as a beige solid.

^{1}H-NMR (400 MHz, DMSO-*d6*): δ/ppm = 12.16 (s, 1H), 11.93 (s, 1H), 4.25 (s, 2H, 1´-H). **^{13}C-NMR** (101 MHz, DMSO-*d6*): δ/ppm = 156.75 (4-C), 149.55 (2-C), 144.33 (5-C), 58.02 (1´-C). **HRMS (ESI$^-$)**: calculated for $C_4H_4N_3O_3^-$ [M-H]$^-$ 142.0258; found: 142.0258. **IR (ATR):** ν (cm^{-1}) = 3429 (w), 3197 (w), 3034 (w), 2846 (w), 1682 (s), 1476 (m), 1446 (m), 1415 (m), 1244 (m), 1207 (m), 1061 (m), 1030 (m), 869 (m), 811 (m), 742 (s).

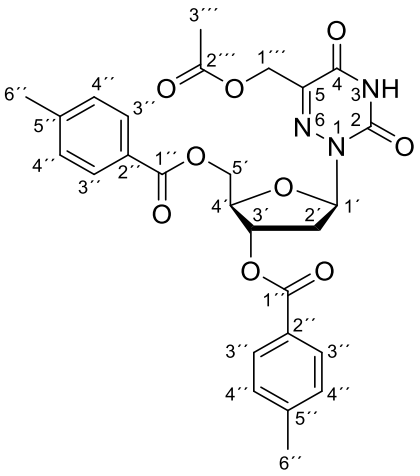
5-Hydroxymethyl-(3´,5´-di-*O*-p-toluoyl)-6-aza-2´-deoxyuridine **7**



In a heat-dried round bottom flask and under argon atmosphere 290 mg 5-hydroxymethyl-6-azauracil **6** (2.03 mmol, 1.0 eq.) were suspended in 10.64 mL hexamethyldisilazane and 351 µL TMSCl were added. The mixture was refluxed at 135 °C for 75 min, the resulting brown solution was cooled to room temperature and concentrated *in vacuo* at 30 °C. The oily residue was dried under high vacuum for 1 h and subsequently dissolved in 6.38 mL dry $CHCl_3$. 867 mg Hoffer´s chlorosugar (2.23 mmol, 1.1 eq.) were added and the reaction mixture was stirred at room temperature for 17 h. The reaction mixture was poured into 250 mL 0.2 M aq. HCl and extracted with EtOAc (3 x 200 mL). Combined organic layers were dried over Na_2SO_4 and concentrated to dryness. Purification via silica gel column chromatography (iHex:EtOAc 3:1 → iHex:EtOAc 1:1 → EtOAc) yielded a 9:1 mixture of diastereomeric β- and α- nucleosidation products (760 mg, 1.53 mmol, 75 %) as a colorless solid. Recrystallization from absolute EtOH yielded 558 mg of the pure β-nucleoside **7** (1.13 mmol, 56%) as a colorless solid.

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 9.52 (s, 1H, 3-H), 7.96 – 7.88 (m, 4H, 3´´-H), 7.26 – 7.19 (m, 4H, 4´´-H), 6.73 (dd, *J* = 6.7, 6.6 Hz, 1H, 1´-H), 5.65 (ddd, *J* = 6.3, 3.1, 3.0 Hz, 1H, 3´-H), 4.73 (dd, *J* = 11.5, 7.0 Hz, 1H, 5´-H), 4.61 – 4.46 (m, 4H, 4´-H, 5´-H, 1´´´-H), 2.99 (ddd, *J* = 14.3, 6.7, 6.5 Hz, 1H, 2´-H), 2.50 (ddd, *J* = 14.2, 6.8, 3.3 Hz, 1H, 2´-H), 2.42 (s, 3H, 6´´-H), 2.38 (s, 3H, 6´´-H). **^{13}C-NMR** (151 MHz, $CDCl_3$): δ/ppm = 166.70 (1´´-C), 166.19 (1´´-C), 155.59 (4-C), 148.47 (2-C), 145.58 (5-C), 144.54 (5´´-C), 144.29 (5´´-C), 129.94 (3´´-C), 129.37 (4´´-C), 129.30 (4´´-C), 126.85 (2´´-C), 126.59 (2´´-C), 86.27 (1´-C), 82.74 (4´-C), 75.20 (3´-C), 64.36 (5´-C), 60.29 (1´´´-C), 35.04 (2´-C), 21.87 (6´´-C), 21.82 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{25}H_{26}N_3O_8^+$ $[M+H]^+$ 496.1714, found: 496.1715; calculated for $C_{25}H_{29}N_4O_8^+$ $[M+NH_4]^+$ 513.1980, found: 513.1975; calculated for $C_{25}H_{25}N_3O_8Na^+$ $[M+Na]^+$ 518.1534, found: 518.1529. **HRMS (ESI$^-$)**: calculated for $C_{25}H_{24}N_3O_8^-$ $[M-H]^-$ 494.1566; found: 494.1572. **IR (ATR):** ν (cm^{-1}) = 1714 (s), 1611 (m), 1450 (w), 1273 (s), 1178 (m), 1105 (s), 1020 (w), 753 (m). **R_f** (iHex: EtOAc 1:1): 0.27.

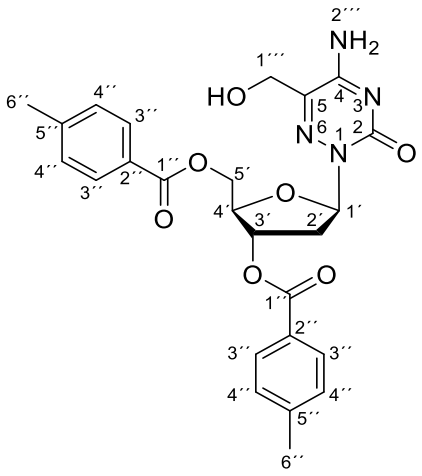
5-Methylacetate-(3´,5´-di-*O*-p-toluoyl)-6-aza-2´-deoxyuridine **8**



273 mg hydroxymethyl nucleoside **7** (0.55 mmol, 1.0 eq.) were dissolved in 3.30 mL dry pyridine and 220 µL acetic anhydride (2.33 mmol, 4.2 eq.) were added under argon atmosphere. The reaction mixture was stirred at room temperature for 22 h and stopped by addition of 5.0 mL absolute EtOH. The solution was concentrated *in vacuo*, coevaporated twice from dry toluene (2 x 15 mL) and the residue was purified by silica gel column chromatography (CH_2Cl_2 → CH_2Cl_2:MeOH 20:1) to obtain 283 mg of acetyl protected nucleoside **8** (0.53 mmol, 96 %) as a colorless foam.

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 8.54 (s, 1H, 3-H), 7.94 – 7.88 (m, 4H, 3´´-H), 7.26 – 7.18 (m, 4H, 4´´-H), 6.67 (dd, *J* = 6.9, 5.4 Hz, 1H, 1´-H), 5.69 (ddd, *J* = 7.4, 4.5, 3.2 Hz, 1H, 3´-H), 5.04 (d, *J* = 14.0 Hz, 1H, 1´´´-H), 4.91 (d, *J* = 14.0 Hz, 1H, 1´´´-H), 4.58 – 4.49 (m, 3H, 4´-H, 5´-H), 2.96 (ddd, *J* = 14.1, 6.8, 5.4 Hz, 1H, 2´-H), 2.51 (ddd, *J* = 14.1, 7.0, 4.5 Hz, 1H, 2´-H), 2.43 (s, 3H, 6´´-H), 2.39 (s, 3H, 6´´-H), 2.15 (s, 3H, 3´´´-H). **^{13}C-NMR** (151 MHz, $CDCl_3$): δ/ppm = 170.33 (2´´´-C), 166.33 (1´´-C), 166.07 (1´´-C), 154.46 (4-C), 147.79 (2-C), 144.52 (5´´-C), 144.13 (5´´-C), 141.45 (5-C), 129.94 (3´´-C), 129.89 (3´´-C), 129.38 (4´´-C), 129.27 (4´´-C), 127.05 (2´´-C), 126.65 (2´´-C), 86.29 (1´-C), 82.54 (4´-C), 74.86 (3´-C), 64.33 (5´-C), 59.91 (1´´´-C), 35.38 (2´-C), 21.88 (6´´-C), 21.82 (6´´-C), 20.73 (3´´´-C). **HRMS (ESI$^+$):** calculated for $C_{27}H_{31}N_4O_9^+$ $[M+NH_4]^+$ 555.2086, found: 555.2019. **HRMS (ESI$^-$)**: calculated for $C_{27}H_{26}N_3O_9^-$ $[M-H]^-$ 536.1675; found: 536.1677. **IR (ATR)**: ν (cm^{-1}) = 1952 (w), 1711 (s), 1610 (m), 1442 (m), 1380 (m), 1269 (s), 1177 (m), 1100 (s), 1019 (m), 839 (w), 751 (s). **R_f** (CH_2Cl_2:MeOH 20:1): 0.52.

5-Hydroxymethyl-(3´,5´-di-*O*-p-toluoyl)-6-aza-2´-deoxycytidine **9**

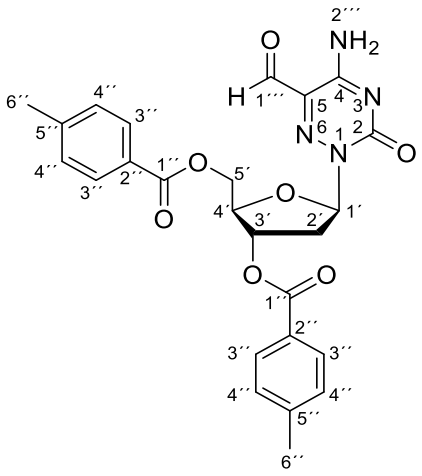


In a heat dried Schlenk-flask 487 mg 1,2,4-triazole (7.05 mmol, 9.0 eq.) were dissolved in 19.50 mL dry MeCN and cooled to 0 °C under Ar atmosphere. 146 µL $POCl_3$ (1.57 mmol, 2.0 eq.) were added dropwise and the mixture was stirred at 0 °C for 10 min. Subsequently, 966 µL NEt_3 (6.97 mmol, 8.9 eq.) were added and the mixture was stirred for another 20 min at 0 °C before 421 mg 2´-deoxyuridine derivative **8** (0.78 mmol, 1.0 eq.) were added. The reaction mixture was allowed to warm to room temperature and stirred for 18 h. After complete conversion the mixture was poured into 100 mL saturated aq. $NaHCO_3$ solution and extracted with CH_2Cl_2 (3 x 150 mL). Combined organic layers were dried over Na_2SO_4 and concentrated to dryness *in vacuo*.

The residue was dissolved in 15.60 mL 1,4-dioxane, 4.91 mL conc. NH_4OH were added and the mixture was stirred at 40 °C for 5 h. After cooling to room temperature the solution was poured into saturated aq. NH_4Cl solution (100 mL) and extracted with CH_2Cl_2 (3 x 100 mL). Combined organic layers were dried over Na_2SO_4, concentrated to dryness *in vacuo* and purified by silica gel column chromatography (CH_2Cl_2:MeOH 80:1 → CH_2Cl_2:MeOH 60:1 → CH_2Cl_2:MeOH 50:1 → CH_2Cl_2:MeOH 30:1 → CH_2Cl_2:MeOH 20:1 → CH_2Cl_2:MeOH 15:1) to obtain 324 mg of the 2´-deoxycytidine derivative **9** (0.66 mmol, 84 %) as an off-white solid.

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 9.96 (s, 1H, 2´´´-H), 9.21 (s, 1H, 2´´´-H), 7.91 – 7.84 (m, 4H, 3´´-H), 7.24 – 7.16 (m, 4H, 4´´-H), 6.67 (d, *J* = 9.6 Hz, 1H, 1´-H), 5.64 (dd, *J* = 6.5, 3.4 Hz, 1H, 3´-H), 4.66 – 4.48 (m, 6H, 4´-H, 5´-H, 1´´´-H), 3.02 – 2.94 (m, 1H, 2´-H), 2.56 – 2.48 (m, 1H, 2´-H), 2.39 (s, 3H, 6´´-H), 2.36 (s, 3H, 6´´-H). **^{13}C-NMR** (151 MHz, $CDCl_3$): δ/ppm = 166.45 (1´´-C), 166.15 (1´´-C), 154.50 (4-C), 144.47 (5´´-C), 144.30 (5´´-C), 137.91 (5-C), 129.94 (3´´-C), 129.87 (3´´-C), 129.42 (4´´-C), 129.36 (4´´-C), 126.91 (2´´-C), 126.63 (2´´-C), 87.22 (1´-C), 83.01 (4´-C), 75.07 (3´-C), 64.27 (5´-C), 60.98 (1´´´-C), 35.38 (6´´-C), 21.85 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{25}H_{27}N_4O_7^+$ $[M+H]^+$ 495.1874, found: 495.1872; calculated for $C_{25}H_{26}N_4O_7Na^+$ $[M+Na]^+$ 517.1694, found: 517.1691. **HRMS (ESI$^-$)**: calculated for $C_{25}H_{25}N_4O_7^-$ $[M-H]^-$ 493.1729; found: 493.1734. **IR (ATR)**: ν (cm^{-1}) = 3229 (br w), 1716 (s), 1611 (m), 1450 (w), 1272 (s), 1178 (m), 1104 (s), 1020 (m), 752 (s). **R_f** (CH_2Cl_2:MeOH 20:1): 0.26.

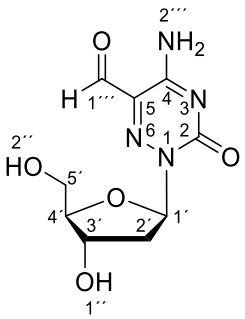
5-Formyl-(3´,5´-di-*O*-p-toluoyl)-6-aza-2´-deoxycytidine **10**



In a heat dried Schlenk flask 306 mg of 5-hydroxymethyl-2´-deoxycytdidine derivative **9** (0.62 mmol, 1.0 eq.) were dissolved in 18.6 mL dry CH_2Cl_2 and cooled to -15 °C in a NaCl/ice cooling bath. Subsequently, 289 mg Dess-Martin periodinane (0.68 mmol, 1.1 eq.) were added, the mixture was allowed to warm to room temperature and stirred for 1 h at room temperature. After complete conversion the reaction was stopped by addition of a solution of 294 mg $Na_2S_2O_3$ (1.86 mmol, 3.0 eq.) in 100 mL saturated aq. $NaHCO_3$ and extracted with with EtOAc (3 x 150 mL). Combined organic layers were dried over Na_2SO_4, concentrated to dryness *in vacuo* and purified by silica gel column chromatography (CH_2Cl_2 → CH_2Cl_2:MeOH 40:1) to yield 272 mg of aldehyde **10** (0.55 mmol, 89 %) as a yellow foam.

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 9.40 (s, 1H, 1´´´-H), 8.31 (s, 1H, 2´´´-H), 7.94 – 7.86 (m, 4H, 3´´-H), 7.28 – 7.20 (m, 4H, 4´´-H), 6.87 (dd, *J* = 5.9 Hz, 1H, 1´-H), 5.74 (dd, *J* = 5.9, 2.9 Hz, 1H, 3´-H), 4.72 – 4.52 (m, 3H, 4´-H, 5´-H), 3.01 – 2.94 (m, 1H, 2´-H), 2.66 – 2.59 (m, 1H, 2´-H), 2.43 (s, 3H, 6´´-H), 2.41 (s, 3H, 6´´-H). **^{13}C-NMR** (151 MHz, $CDCl_3$): δ/ppm = 188.81 (1´´´-C), 166.26 (1´´-C), 166.17 (1´´-C), 155,17 (4-C), 144.69 (5´´-C), 144.43 (5´´-C), 129.97 (3´´-C), 129.83 (3´´-C), 129.44 (4´´-C), 129.37 (5-C), 126.93 (2´´-C), 126.48 (2´´-C), 88.08 (1´-C), 83.34 (4´-C), 74.73 (3´-C), 63.87 (5´-C), 36.14 (2´-C), 21.89 (6´´-C), 21.84 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{25}H_{25}N_4O_7^+$ $[M+H]^+$ 493.1718, found: 493.1717; calculated for $C_{25}H_{24}N_4O_7Na^+$ $[M+Na]^+$ 515.1537, found: 515.1534. **HRMS (ESI$^-$)**: calculated for $C_{25}H_{23}N_4O_7^-$ $[M-H]^-$ 491.1572; found: 491.1577. **IR (ATR)**: ν (cm^{-1}) = 3237 (br w), 1720 (s), 1612 (m), 1272 (s), 1178 (w), 1104 (m), 753 (m). **R_f** (CH_2Cl_2:MeOH 20:1): 0.34.

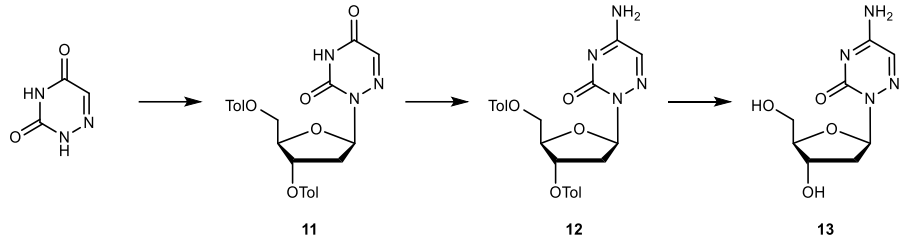
6-Aza-5-formyl-2´-deoxycytidine **3**



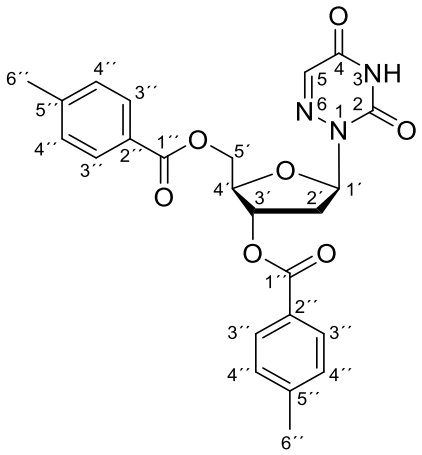
The deprotcetion was performed according to a modified procedure of *Mitchell et al.*[2] 36 mg of toluoyl protected 6-Aza-nucleoside **10** (73 µmol, 1.0 eq.) were dissolved in 0.73 mL dry MeOH and 0.73 dry benzene. To the solution 540 µL 0.5 M NaOMe in dry MeOH (270 µmol, 3.7 eq.) were added and the reaction mixture was stirred for 1.5 h at room temperature. After complete conversion the reaction mixture was neutralized with 2 M aq. HCl and evaporated to dryness. The resulting residue was dissolved in 10 mL ddH_2O and extracted with CH_2Cl_2 (5 x 10 mL). The aqueous layer was lyophyllized and the residue was purified by *reversed phase* HPLC (0 % → 3 % MeCN in H_2O in 45 min) to yield 10.1 mg of the desired product **3** (39 µmol, 54 %) as a colorless solid.

^{1}H-NMR (600 MHz, DMSO-*d6*): δ/ppm = 9.56 (s, 1H, 1´´´-H), 8.61 (s, 1H, 2´´´-H), 7.78 (s, 1H, 2´´´-H), 6.52 (dd, *J* = 7.0, 5.0 Hz, 1H, 1´-H), 5.21 (d, *J* = 4.7 Hz, 1H, 1´´-H), 4.63 (t, *J* = 5.9 Hz, 1H, 2´´-H), 4.35 (ddd, *J* = 10.2, 5.3, 5.1 Hz, 1H, 3´-H), 3.76 (dd, *J* = 5.2, 5.1 Hz, 1H, 4´-H), 3.49 (dt, *J* = 11.2, 5.5 Hz, 1H, 5´-H), 3.41 (dt, *J* = 11.8, 6.0 Hz, 1H, 5´-H), 2.47 (ddd, *J* = 13.3, 6.3, 4.9 Hz, 1H, 2´-H), 2.15 (ddd, *J* = 13.1, 7.0, 5.3 Hz, 1H, 2´-H). **^{13}C-NMR** (151 MHz, DMSO-*d6*): δ/ppm = 189.77 (1´´´-C), 156.56 (4-C), 151.66 (2-C), 129.61 (5-C), 87.87 (4´-C), 86.48 (1´-C), 70.50 (3´-C), 62.07 (5´-C), 37.74 (2´-C). **HRMS (ESI$^+$)**: calculated for $C_9H_{13}N_4O_5^+$ $[M+H]^+$ 257.0881, found: 257.0880; calculated for $C_9H_{12}N_4O_5Na^+$ $[M+Na]^+$ 279.0700, found: 279.0698. **HRMS (ESI$^-$)**: calculated for $C_9H_{11}N_4O_5^-$ $[M-H]^-$ 255.0735; found: 255.0735.

Synthesis of 6-Aza-2´-deoxycytidine



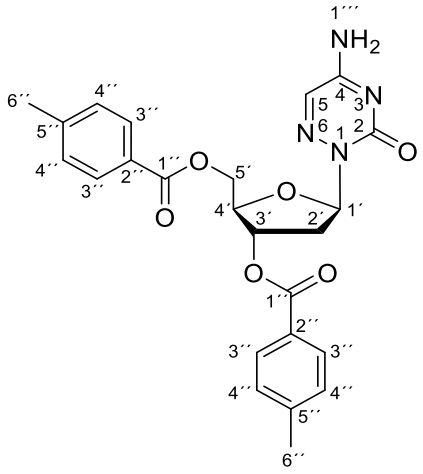
(3´,5´-Di-*O*-p-toluoyl)-6-aza-2´-deoxyuridine **11**



In a heat-dried round bottom flask and under argon atmosphere 300 mg 6-azauracil (2.65 mmol, 1.0 eq.) were suspended in 2.88 mL hexamethyldisilazane and 100 µL TMSCl were added. The mixture was refluxed at 145 °C for 1 h and concentrated *in vacuo* at 30 °C after cooling to room temperature. The obtained dark oil was dried under high vacuum for 1 h and subsequently dissolved in 8.33 mL dry $CHCl_3$. 1.13 g Hoffer´s chlorosugar (2.92 mmol, 1.1 eq.) were added and the reaction mixture was stirred at room temperature for 16 h. The reaction mixture was poured into 250 mL saturated aq. $NaHCO_3$ solution and extracted with EtOAc (3 x 200 mL). Combined organic layers were dried over Na_2SO_4 and concentrated to dryness. Purification via silica gel column chromatography (iHex:EtOAc 3:1 → iHex:EtOAc 1:1 → iHex:EtOAc 1:2) yielded a mixture of diastereomeric β- and α- nucleosidation products (868 mg, 1.86 mmol, 70 %) as a colorless solid. Recrystallization from absolute EtOH yielded 742 mg of the pure β-nucleoside **11** (1.59 mmol, 60%) as a colorless solid.

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 8.82 (s, 1H,3-H), 7.97 – 7.91 (m, 4H, 3´´-H), 7.29 – 7.20 (m, 5H, 5-H, 4´´-H), 6.68 (dd, *J* = 6.9, 6.7 Hz, 1H, 1´-H), 5.72 (ddd, *J* = 6.8, 3.9, 3.5 Hz, 1H, 3´-H), 4.66 (dd, *J* = 11.6, 4.3 Hz, 1H, 5´-H), 4.53 (dd, *J* = 9.7, 4.4 Hz, 1H, 4´-H), 4.46 (dd, *J* = 11.6, 5.0 Hz, 1H, 5´-H), 2.98 (dt, *J* = 13.4, 6.5 Hz, 1H, 2´-H), 2.48 (ddd, *J* = 14.1, 6.8, 3.9 Hz, 1H, 2´-H), 2.43 (s, 3H, 6´´-H), 2.41 (s, 3H, 6´´-H). **^{13}C-NMR** (101 MHz, $CDCl_3$): δ/ppm = 166.36 (1´´-C), 166.16 (1´´-C), 155.30 (4-C), 147.77 (2-C), 144.56 (5´´-C), 144.11 (5´´-C), 136.10 (5-C), 129.94 (3´´-C), 129.90 (3´´-C), 129.39 (4´´-C), 129.27 (4´´-C), 127.11 (2´´-C), 126.61 (2´´-C), 86.05 (1´-C), 82.73 (4´-C), 75.03 (3´-C), 64.06 (5´-C), 35.17 (2´-C), 21.90 (6´´-C), 21.86 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{24}H_{24}N_3O_7^+$ $[M+H]^+$ 466.1609, found: 466.1610; calculated for $C_{24}H_{27}N_4O_7^+$ $[M+NH_4]^+$ 483.1874, found: 483.1874. **HRMS (ESI$^-$)**: calculated for $C_{24}H_{22}N_3O_7^-$ $[M-H]^-$ 464.1463; found: 464.1469. **IR (ATR)**: ν (cm^{-1}) = 3200 (w), 1727 (s), 1699 (s), 1610 (w), 1437 (w), 1392 w), 1374 (w), 1329 (m), 1265 (s), 1173 (w), 1100 (s), 1080 (s), 954 (m), 808 (m), 752 (s). **R_f** (iHex: EtOAc 1:1): 0.44.

(3´,5´-Di-*O*-p-toluoyl)-6-aza-2´-deoxycytidine **12**

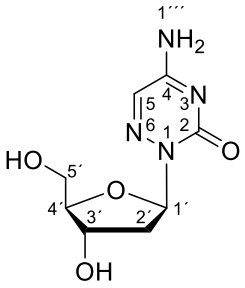


In a heat dried Schlenk-flask 668 mg 1,2,4-triazole (9.67 mmol, 9.0 eq.) were dissolved in 26.80 mL dry MeCN and cooled to 0 °C under Ar atmosphere. 201 µL $POCl_3$ (2.15 mmol, 2.0 eq.) were added dropwise and the mixture was stirred at 0 °C for 10 min. Subsequently, 1.33 mL NEt_3 (9.56 mmol, 8.9 eq.) were added and the mixture was stirred for another 20 min at 0 °C before 500 mg of protected 6-aza-2´-deoxyuridine **11** (1.07 mmol, 1.0 eq.) were added. The reaction mixture was allowed to warm to ambient temperature and stirred for 17 h at room temperature. After complete conversion the mixture was poured into 250 mL saturated aq. $NaHCO_3$ solution and extracted with CH_2Cl_2 (3 x 200 mL). Combined organic layers were washed with brine (150 mL), dried over Na_2SO_4 and concentrated to dryness *in vacuo*.

The residue was dissolved in 21.40 mL 1,4-dioxane and 6.74 mL conc. NH_4OH were added. After stirring the mixture at 45 °C for 3 h the solution was poured into saturated aq. NH_4Cl solution (150 mL) and extracted with EtOAc (3 x 150 mL). Combined organic layers were washed with brine (150 mL), dried over Na_2SO_4 and concentrated to dryness *in vacuo*. The crude product was purified by silica gel column chromatography (CH_2Cl_2:MeOH 40:1 → CH_2Cl_2:MeOH 20:1 → CH_2Cl_2:MeOH 10:1) to obtain 460 mg of the amination product **12** (0.99 mmol, 93 %) as a slightly yellowish solid.

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 8.57 (s, 1H, 1´´´-H)), 7.92 (m, 4H, 3´´-H), 7.71 (s, 1H, 5-H), 7.32 (s, 1H, 1´´´-H), 7.25 – 7.16 (m, 4H, 4´´-H), 6.72 (dd, *J* = 6.6, 6.3 Hz, 1H, 1´-H), 5.73 (ddd, *J* = 7.2, 3.9, 3.1 Hz, 1H, 3´-H), 4.58 (ddd, *J* = 11.1, 5.4, 1.4 Hz, 1H, 5´-H), 4.55 – 4.52 (m, 1H, 4´-H), 4.47 (ddd, *J* = 11.2, 5.2, 1.2 Hz, 1H, 5´-H), 3.01 (ddd, *J* = 13.6, 6.8, 6.4 Hz, 1H, 2´-H), 2.45 (ddd, *J* = 13.9, 6.9, 4.2 Hz, 1H, 2´-H), 2.41 (s, 3H, 6´´-H), 2.36 (s, 3H, 6´´-H). **^{13}C-NMR** (151 MHz, $CDCl_3$): δ/ppm = 166.60 (1´-C), 166.20 (1´´-C), 159.07 (4-C), 154.24 (2-C), 144.34 (5´´-C), 143.96 (5´´-C), 129.94 (3´´-C), 129.94 (3´´-C), 129.32 (4´´-C), 129.21 (4´´-C), 128.61 (5-C), 127.10 (2´´-C), 126.84 (2´´-C), 87.76 (1´-C), 82.23 (4´-C), 75.41 (3´-C), 64.55 (5´-C), 35.17 (2´-C), 21.85 (6´´-C), 21.79 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{24}H_{25}N_4O_6^+$ $[M+H]^+$ 465.1769, found: 465.1770. **HRMS (ESI$^-$)**: calculated for $C_{24}H_{23}N_4O_6^-$ $[M-H]^-$ 463.1623; found: 463.1630**. IR (ATR)**: ν (cm^{-1}) = 3305 (br w), 3059 (br w), 1715 (m), 1643 (m), 1610 (m), 1536 (w), 1466 (w), 1334 (w), 1265 (s), 1178 (m), 1102 (s) 1020 (m), 961 (w), 839 (w), 730 (s). **R_f** (CH_2Cl_2:MeOH 20:1): 0.30.

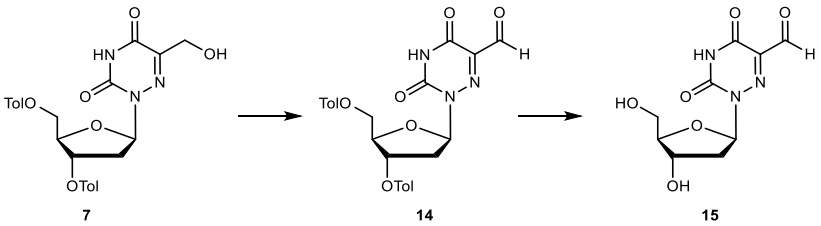
6-Aza-2´-deoxycytidine **13**



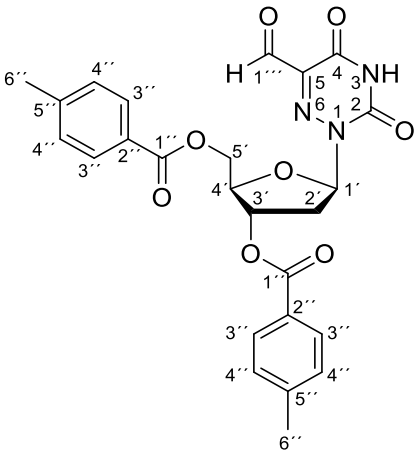
In a round bottom flask 10.0 mL of a solution of $MeNH_2$ in EtOH (33%) were added to 200 mg of toluoyl protected compound **12** (0.43 mmol, 1.0 eq.) and the mixture was stirred at room temperature for 14 h. As reaction control still showed protected starting material another 5.0 mL 33 % $MeNH_2$ in EtOH were added and the mixture was heated to 45 °C for 1 h. The solution was evaporated to dryness, dissolved in 10 mL ddH_2O and 10 mL CH_2Cl_2. Phases were separated, the aqueous layer was extracted with CH_2Cl_2 (4 x 10 mL) and subsequently concentrated *in vacuo*. The crude product was dissolved in ddH_2O (24 mL) and purified by reversed phase HPLC (0 % → 5 % MeCN in H_2O in 45 min) to yield 57.6 mg of the deprotected 6-Aza-nucleoside **13** as a white solid (0.25 mmol, 58 %).

^{1}H-NMR (600 MHz, D_2O): δ/ppm = 7.75 (s, 1H, 5-H), 6.57 (dd, *J* = 7.2, 4.8 Hz, 1H, 1´-H), 4.57 (ddd, *J* = 6.9, 4.4, 1.7 Hz, 1H, 3´-H), 4.05 – 4.01 (m, 1H, 4´-H), 3.85 – 3.69 (m, 1H, 5´-H), 3.71 – 3.61 (m, 1H, 5´-H), 2.72 (ddd, *J* = 13.9, 7.1, 4.8, 2.0 Hz, 1H, 2´-H), 2.35 (ddd, *J* = 14.7, 7.4, 4.0, 1.9 Hz, 1H, 2´-H). **^{13}C-NMR** (151 MHz, D_2O): δ/ppm = 158.97 (4-C), 155.80 (2-C), 129.37 (5-C), 86.52 (1´-C), 86.44 (4´-C), 70.93 (3´-C), 61.80 (5´-C), 36.67 (2´-C). **HRMS (ESI$^-$)**: calculated for $C_8H_{11}N_4O_4^-$ $[M-H]^-$ 227.0786; found: 227.0787.

Synthesis of 6-Aza-5-formyl-2´-deoxyuridine



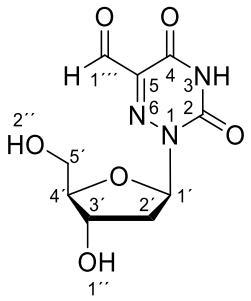
5-Formyl-(3´,5´-di-*O*-p-toluoyl)-6-aza-2´-deoxyuridine **14**



In a Schlenk flask 158 mg of 5-hydroxymethyl-2´-deoxyuridine derivative **7** (0.32 mmol, 1.0 eq.) were dissolved in 9.6 mL dry CH_2Cl_2 and cooled to -15 °C. At this temperature 162 mg Dess-Martin periodinane (0.38 mmol, 1.2 eq.) were added, the mixture was allowed to warm to room temperature slowly and was furthe stirred for 2.5 h. The reaction was stopped by addition of a solution of 182 mg $Na_2S_2O_3$ (1.15 mmol, 3.6 eq.) in 50 mL saturated aq. $NaHCO_3$ and the mixture was extracted with with EtOAc (3 x 100 mL). Combined organic layers were dried over Na_2SO_4, concentrated to dryness *in vacuo* and purified by silica gel column chromatography (iHex:EtOAc 1:1 → EtOAc) to yield 152 mg of aldehyde **14** (0.31 mmol, 96 %) as an off-white foam.

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 9.58 (s, 1H, 1´´´-H), 9.02 (s, 1H, 3H), 8.01 – 7.79 (m, 4H, 3´´-H), 7.29 – 7.18 (m, 4H, 4´´-H), 6.70 (t, *J* = 6.3 Hz, 1H, 1´-H), 5.72 (dd, *J* = 6.7, 3.4 Hz, 1H, 3´-H), 4.70 – 4.49 (m, 3H, 4´-H, 5´-H), 3.07 (dt, *J* = 14.2, 6.4 Hz, 1H, 2´-H), 2.58 (ddd, *J* = 14.2, 6.7, 4.0 Hz, 1H, 2´-H), 2.43 (s, 3H, 6´´-H), 2.40 (s, 3H, 6´´-H). **^{13}C-NMR** (101 MHz, $CDCl_3$): δ/ppm = 184.16 (1´´´-C), 166.33 (1´´-C), 166.14 (1´´-C), 152.79 (4-C), 147.15 (2-C), 144.68 (5´´-C) , 144.40 (5´´-C), 137.22 (5-C), 129.94 (3´´-C), 129.81 (3´´-C), 129.42 (4´´-C), 129.40 (4´´-C), 126.81 (2´´-C), 126.41 (2´´-C), 87.16 (1´-C), 83.20 (4´-C), 74.59 (3´-C), 63.74 (5´-C), 35.40 (2´-C), 21.91 (6´´-C), 21.85 (6´´-C). **HRMS (ESI$^+$)**: calculated for $C_{25}H_{27}N_4O_8^+$ $[M+NH_4]^+$ 511.1823, found:511.1823. **HRMS (ESI$^-$)**: calculated for $C_{25}H_{22}N_3O_8^-$ $[M-H]^-$ 492.1412; found: 492.1413**. IR (ATR)**: ν (cm^{-1}) = 2963 (w), 1712 (s), 1611 (m), 1439 (w), 1398 (w), 1309 (w), 1260 (s), 1178 (m), 10963 (s), 1020 (s), 909 (m), 800 (s), 752 (s), 732 (s). **R_f** (iHex: EtOAc 1:1): 0.17.

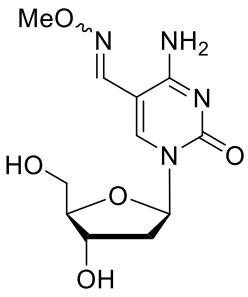
5-Formyl-6-aza-2´-deoxyuridine **15**



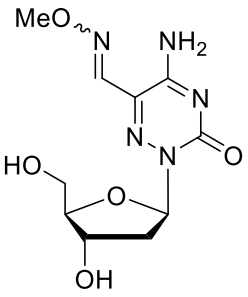
97 mg of toluoyl protected 6-Aza-nucleoside **14** (0.20 mmol, 1.0 eq.) were dissolved in 2.0 mL dry MeOH and 2.0 dry benzene. To the solution freshly prepared 0.5 M NaOMe in dry MeOH (1.46 mL, 0.73 mmol, 3.7 eq.) was added and the reaction mixture was stirred for 2 h at room temperature. Subsequently, the reaction mixture was neutralized with 2 M aq. HCl, evaporated to dryness and redissolved in 10 mL ddH_2O. The mixture was extracted with CH_2Cl_2 (4 x 10 mL). The aqueous layer was lyophyllized and the residue was purified by *reversed phase* HPLC (0 % → 3 % MeCN in H_2O in 45 min) to yield 16.3 mg of the product **15** (63 µmol, 32 %) as a colorless solid.

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 9.62 (s, 0.2H, 1´´´-H), 6.55 (dd, *J* = 7.2, 4.5 Hz, 0.2H, 1´-H), 6.48 (dd, *J* = 7.2, 5.2 Hz, 0.8H, 1´´-H), 5.87 (s, 0.7H, 1´´´-H), 4.62 – 4.48 (m, 1H, 3´-H), 3.96 (m, 1H, 4´-H), 3.75 – 3.55 (m, 2H, 5´-H), 2.71 (m, 1H, 2´-H), 2.28 (m, 1H, 2´-H). **^{13}C-NMR** (101 MHz, D_2O): δ/ppm = 190.31 (1´´´-C), 165.12 (4-C), 162.87 (4-C)), 156.31 (2-C), 155.12 (2-C), 142.78, 137.09 (5-C), 87.81 (4´-C), 86.93 (4´-C), 86.48 (1´-C), 86.42 (1´-C), 86.14 (1´´´-C), 71.05 (3´-C), 70.83 (3´-C), 61.81 (5´-C), 61.70 (5´-C), 36.89 (2´-C), 36.57 (2´-C). **HRMS (ESI$^-$)**: calculated for $C_9H_{10}N_3O_6^-$ $[M-H]^-$ 256.0575; found: 256.0576.

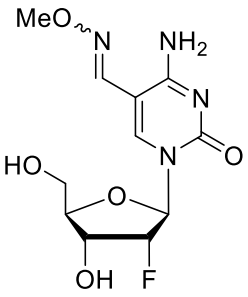
Reactions of **1**, **2**, **3** and **15** with Methoxyamine



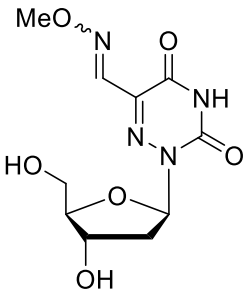
16



17



18



19

To a solution of 100 nmol nucleoside (1.0 eq.) in 35.0 µL aq. NaOH (pH = 10) was added 33.33 µL aq. $MeONH_2$ (150 mM, 50.0 eq.) and the mixture was incubated at 25 °C for 1 h. The reaction was stopped by the addition of 35.0 µL of aq. HCOOH (pH = 3) and analyzed by reversed phase HPLC. The resulting reaction products were collected and analyzed via HRMS.

	16	**17**	**18**	**19**
Conversion of SM	50 %	quant.	57 %	quant.
HPLC gradient [MeCN in H_2O in 45 min]	0 % → 15 %	0 % → 30 %	0 % → 13 %	0 % → 30 %
HRMS (calcd.) $ESI^{(+)}$	285.1194 $C_{11}H_{17}O_5N_4^+$ $[M+H]^+$	286.1146 $C_{10}H_{16}O_5N_5^+$ $[M+H]^+$	303.1099 $C_{11}H_{16}O_5N_4F^+$ $[M+H]^+$	
$ESI^{(-)}$	283.1048 $C_{11}H_{15}O_5N_4^-$ $[M-H]^-$	284.1000 $C_{10}H_{14}O_5N_5^-$ $[M-H]^-$		285.0841 $C_{10}H_{13}O_6N_4^-$ $[M-H]^-$
HRMS (found) $ESI^{(+)}$	285.1193 $[M+H]^+$	286.1151 $[M+H]^+$	303.1098 $[M+H]^+$	
$ESI^{(-)}$	283.1048 $[M-H]^-$	284.1000 $[M-H]^-$		285.0840 $[M-H]^-$

Reaction of **1**, **2** and **3** with $NaHSO_3$

20 mM nucleoside solution in ddH_2O was shaken with 4.36 M aq. $NaHSO_3$ (435 eq, pH = 5) at 55 °C. Samples were taken at time points of 0, 5 min, 20 min, 30 min, 40 min, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h and 15 h. Bisulfite was cleaved by preparing 1 mM solution of reaction mixture and 2 M NaOH (1700 eq), shaken for 10 min at rt and followed by immediate injection to HPLC (Gradients: **1** 0% to 13% MeCN in H_2O in 45 min, **2** 0% to 15% MeCN in H_2O in 45 min, **3** 0% to 3% MeCN in H_2O in 45 min).

Chromatograms of each time point were normalized to the integral of the reagent and corrected accordingly. Conversion was determined by the decrease of the corrected integrals of the starting materials.

Cell culture

DMEM high glucose (Sigma Aldrich) containing 10% heat-inactivated FBS (Gibco), 100 U/mL penicillin, 100 µg/mL streptomycin, 2 mM L-glutamine were used to culture Neuro-2a and RBL-2H3 cell lines. DMEM high glucose and Ham's Nutrient Mixture F12 (Sigma Aldrich) containing 10% heat-inactivated FBS (Gibco), 100 U/mL penicillin, 100 µg/mL streptomycin, 2 mM L-glutamine were used to culture CHO-K1 cells. Metabolic-labelling experiments were conducted by plating cells in their dedicated medium containing 350 µM of F-fdC or a-fdC (in case of the co-feeding study: 350 µM of each) for three days.

DMEM high glucose (Sigma Aldrich) containing 10% FBS (PAN Biotech), 2 mM L-glutamine, 1x MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol (Sigma Aldrich) were used as basal medium for E14 TDG +/- and -/- (obtained from Cortázar *et al.)* mESC cultures. The mESC lines were maintained in naïve state on gelatin coated plates by supplementing basal medium with 1000 U/mL LIF (ORF Genetics), 3.0 µM GSK3 inhibitor CHIR99021 and 1.0 µM MNK inhibitor CGP57380 (a2i medium). Metabolic labelling experiments with isotope-labeled nucleosides were performed by plating mESCs in priming conditions consisted of basal mESC medium supplemented with 3. µM GSK3 inhibitor CHIR99021 and 2.5 µM tankyrase inhibitor IWR1-endo. Labelled nucleosides were added at the concentration of 350 µM each to the primming medium over three days.

All samples were washed with PBS (Sigma Aldrich) once befor harvesting and lysed directly in the plates by adding RLT buffer (Qiagen) supplemented with 400 µM 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM). Next, DNA was sheered in MM400 bead mill (Retsch) at 30 Hz for 1 min in 2 mL microcentrifuge tubes (Eppendorf) with 5 mm diamiter stainless steel beads (one per tube) and centrifuged at 21000rcf for 5 minutes. Genomic DNA (gDNA) was extrated using Zymo Quick gDNA mini-prep® kit according to the protocol with an addition of RNAse A treatment (Qiagen) at 0.2 mg/mL in Genomic Lysis Buffer – 400 µL of the solution transferred directly on the column and incubated for 15 minutes. All samples were eluted in DNAse-free ddH_2O with 20 µM BHT, the concentration of gDNA was measured on Nanodrop.

While investigating the deformylation levels using F-fdC as a metabollic label, the following conditions were used for the gDNA digestion procedure: Degradase (1.5 µL) per up to 10 µg of F-fdC labelled gDNA in 35 µL in H_2O. The digestion misture was incubated at 37 ^oC for 4 hours. Then, the sampes were filtered using AcroPrep Advance 0.2 µm 96-well filter plate (Pall Life Sciences) prior to LC-MS/MS analysis (39 µL injection volume at 4 °C). This method was compared to NEB enzyme mix digestion (described below) and showed no

difference in the QQQ measurement outcome, thus we concluded that both can be used at equal efficiency.

Chemical labelling and spiking

Chemical labelling of a-fdC was needed to receive a sharper signal during the MS analysis. Therefore, methoxyamine was used as a derivatization reagent. We optimized the reaction conditions including time and temperature.

The quantification of the the Fluoro compounds as well as the other known nucleosides was done by spiking of heavy labelled nucleosides which were synthesized by ourselves.

For the determination of a-fdC in the gDNA samples we did an external calibration.

Analysis of labelled and unlabelled products of F-fdC and a-fdC

The analysis was performed using an UHPLC-QQQ-MS/MS system consisting of a Triple Quad™ 6490 mass spectrometer (Agilent) with an ESI source and an Agilent Infinity 1290 UHPLC. The elution was monitored at 260 nm (Agilent InfinityLab Deuterium Lamp G1314). Data Acquisition and processing were performed using MassHunter Workstation Software Version B.07.01 (Agilent).
The UHPLC separation was performed for the Fluoro-fdC as well as mdC, hmdC, fdC on an InfinityLab Poroshell 120 SB-C8 column (2.1 mm x 150 mm, 2.7 µm, Agilent Technologies, USA) at 35 °C. Water containing 0.0085% FA (v/v, solvent A) and MeCN containing 0.0085% FA (v/v, solvent B) was used as the mobile phase. A gradient of 0 - 3.5% B for 4 min, 3.5 - 5% B for 2.9 min, 5 - 80% B for 0.3 min, 80% B for 3.3 min was used. The flow rate of the mobile phase was set to 0.35 mL min^{-1}.

The derivatized a-fdC, F-fdC and fdC were separated on an InfinityLab Poroshell 120 SB-C18 column (2.1 mm x 150 mm, 2.7 µm, Agilent Technologies, USA) at 35 °C. Water containing 0.0085% FA (v/v, solvent A) and MeCN containing 0.0085% FA (v/v, solvent B) was used as the mobile phase. A gradient of 0% B for 3 min, 0 - 15% B for 27 min, 15 - 100% B for 5 min was used. The flow rate of the mobile phase was set to 0.35 mL min^{-1}.

The mass spectrometry detection was performed under positive ESI mode. The nucleosides and labelled products were monitored using the multiple reaction monitoring (MRM) mode. The MRM parameters were optimized to achieve maximal detection sensitivity (Tables 2 and 3).

Table 2 MRM parameters for the detection of F-fdC and its derivatives

Name	Transition	Scan	Type	Precursor Ion	Product Ion	Ion Polarity
UV-dG	0,0 -> 0,0	MRM	Target	0,0	0,0	Positive
UV-dC	0,0 -> 0,0	MRM	Target	0,0	0,0	Positive
UV total	0,0 -> 0,0	MRM	Target	0,0	0,0	Positive
Fluoro-fdC-dN	274,1 -> 140,1	MRM	Target	274,1	140,1	Positive
Fluoro-fdC-dN-$^{15}N_2$	276,1 -> 142,0	MRM	ISTD	276,1	142,0	Positive
Fluoro-dU	245,1 -> 225,1	MRM	Target	245,1	225,1	Negative
Fluoro-dC	246,1 -> 112,1	MRM	Target	246,1	112,1	Positive
F-fdU	273,1 -> 253,1	MRM	Target	273,1	253,1	Negative
fdC-dN-$^{15}N_2$	258,1 -> 142,0	MRM	ISTD	258,1	142,0	Positive
fdC-dN	256,1 -> 140,1	MRM	Target	256,1	140,1	Positive
$^{15}N_2$-FdC	248,1 -> 114,0	MRM	ISTD	248,1	114,0	Positive

Table 3 MRM parameters for the derivatized a-fdC samples

Name	Transition	Scan	Type	Precursor Ion	Product Ion	Ion Polarity
MeON-a-fdC	286,1 -> 170,1	MRM	Target	286,1	170,1	Positive
a-fdC	257,1 -> 141,1	MRM	Target	257,1	141,1	Positive
a-dC	229,1 -> 188,1	MRM	Target	229,1	188,1	Positive
a-dC	229,1 -> 112,1	MRM	Target	229,1	112,1	Positive

Digestion of the DNA of N2a cell line

As shown in Table 4 we used the following chemicals for the analysis of our samples. Reaction buffer 10X and Enzyme mix was bought as a Nucleoside Digestion Mix (M0649S) kit (*New England BioLabs Inc.*). The nucleosides mix was prepared by ourselve and contains heavy labelled mdC, hmdC, fdC, cadC, 8oxodG and hmdU. Furthermore we spiked heavy labelled Fluoro-dC and Fluoro-fdC for later quantification of the Fluoro-compounds. The concentration of the DNA for the digestion was received from the Nanodrop. The total volume was then incubated at 37 °C for 1.5 hours.

Table 4: Digestion sheet of the gDNA samples

	Sample	c(ng/µL)	m (DNA) [ng]	V (H2O) [µL]	V (DNA in H_2O) [µL]	V (Reaction buffer 10X)	V (Enzyme mix)	V (Nucleosides mix)	Spiking heavy Fluoro-dC	Spiking heavy Fluoro-fdC	Total Volume [µL]	1 h 30 min 37 °C	Injection volume [µL]
Blank 1	1	0	0	38,4	0,0	5,0	3,0	3,0	0,35	0,25	50,0		39
Blank 2	2	0	0	38,4	0,0	5,0	3,0	3,0	0,35	0,25	50,0		39
Control N2a Cofeeding 10 µg	3	528	10000	19,5	18,94	5,0	3,0	3,0	0,35	0,25	50,0		39
Control N2a Cofeeding 10 µg	4	528	10000	19,5	18,94	5,0	3,0	3,0	0,35	0,25	50,0		39
Control N2a Cofeeding 5 µg	5	528	5000	28,9	9,47	5,0	3,0	3,0	0,35	0,25	50,0		39
Control N2a Cofeeding 5 µg	6	528	5000	28,9	9,47	5,0	3,0	3,0	0,35	0,25	50,0		39
Sample N2a Cofeeding 10 µg	7	412	10000	14,1	24,27	5,0	3,0	3,0	0,35	0,25	50,0		39
Sample N2a Cofeeding 10 µg	8	412	10000	14,1	24,27	5,0	3,0	3,0	0,35	0,25	50,0		39
Sample N2a Cofeeding 5 µg	9	412	5000	26,3	12,14	5,0	3,0	3,0	0,35	0,25	50,0		39
Sample N2a Cofeeding 5 µg	10	412	5000	26,3	12,14	5,0	3,0	3,0	0,35	0,25	50,0		39

Derivatization of a-fdC in digested gDNA of N2a cell line

Derivatization of a-fdC from the digested DNA (5 µg) of biological samples by methoxyamine was performed under the following conditions. Briefly, the nucleoside mixture was derivatized in 50 µL H_2O with 40 µL aq. NaOH (pH 10) and 8 µL methoxyamine (163 mM in H_2O) for 45 min at 25 °C. Afterwards the solution was neutralized with 40 µL aq. FA (pH 3). The derivatized nucleosides were then lyophylized and resuspended in 50 µL H_2O.

After resuspension the samples were filtered utilizing a 0.2 µm Supor filtration plate (*Pall Corporation*) and subjected to UHPLC-QQQ-MS/MS.

The determination of F-fdC and some other nucleosides in gDNA (10 µg) was performed by using heavy labelled compounds. After the digestion the samples were filtered utilizing a 0.2 µm Supor filtration plate (*Pall Corporation*) and subjected to UHPLC-QQQ-MS/MS.

Calibration curve for external quantification of a-fdC in gDNA of N2a cell lines

The quantification of a-fdC was done by an external calibration.[3] The external calibration curve was done by serially diluting pure a-fdC (see Table 5) and measured in technical triplicates prior to each measurment. Linear regression was done by Microsoft Excel (Figure 6). The injection volume was 29 µL.

Table 5 Measured values for the external calibration curve of a-fdC

n [fmol]	Average area	standard deviation
0,11	1006	110
0,22	1822	98
0,44	3160	414
0,88	4754	428
1,77	8439	333
3,53	14692	346
7,07	28982	522
14,14	56445	1107
28,28	111343	2041
56,55	220616	2561
113,10	439962	2460

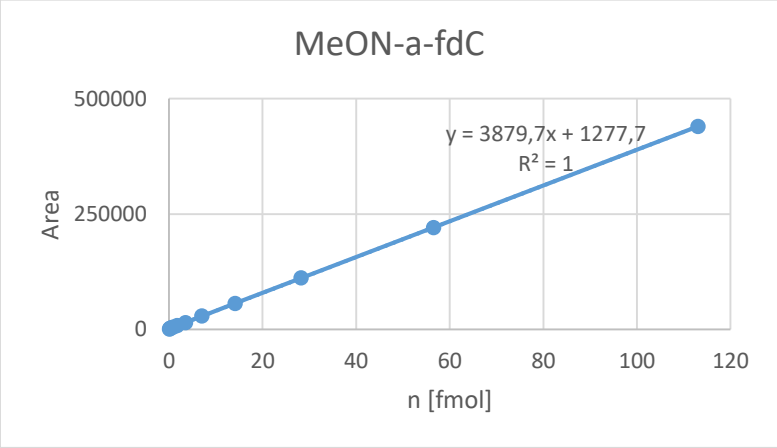


Figure 6: Calibration curve for the external quantification of a-fdC

Quantifiation of concentration of the fed compounds

The quantification of the fed compounds was performed in two technical replicates of a biological triplicate.

For the quantification of a-fdC one has to say, that there was a small background in the a-fdC peak, which was continously substracted from the a-fdC value. This value was determined by the control sample.

The quantification of the Fluoro-fdC and the deformylated Fluoro-dC was done over the ratio of unlabelled to labelled compound. Furthermore it was compared to the amount of deoxy-cytidine and deoxy-Guanosine and dN. The final results can be seen in Figures 7.

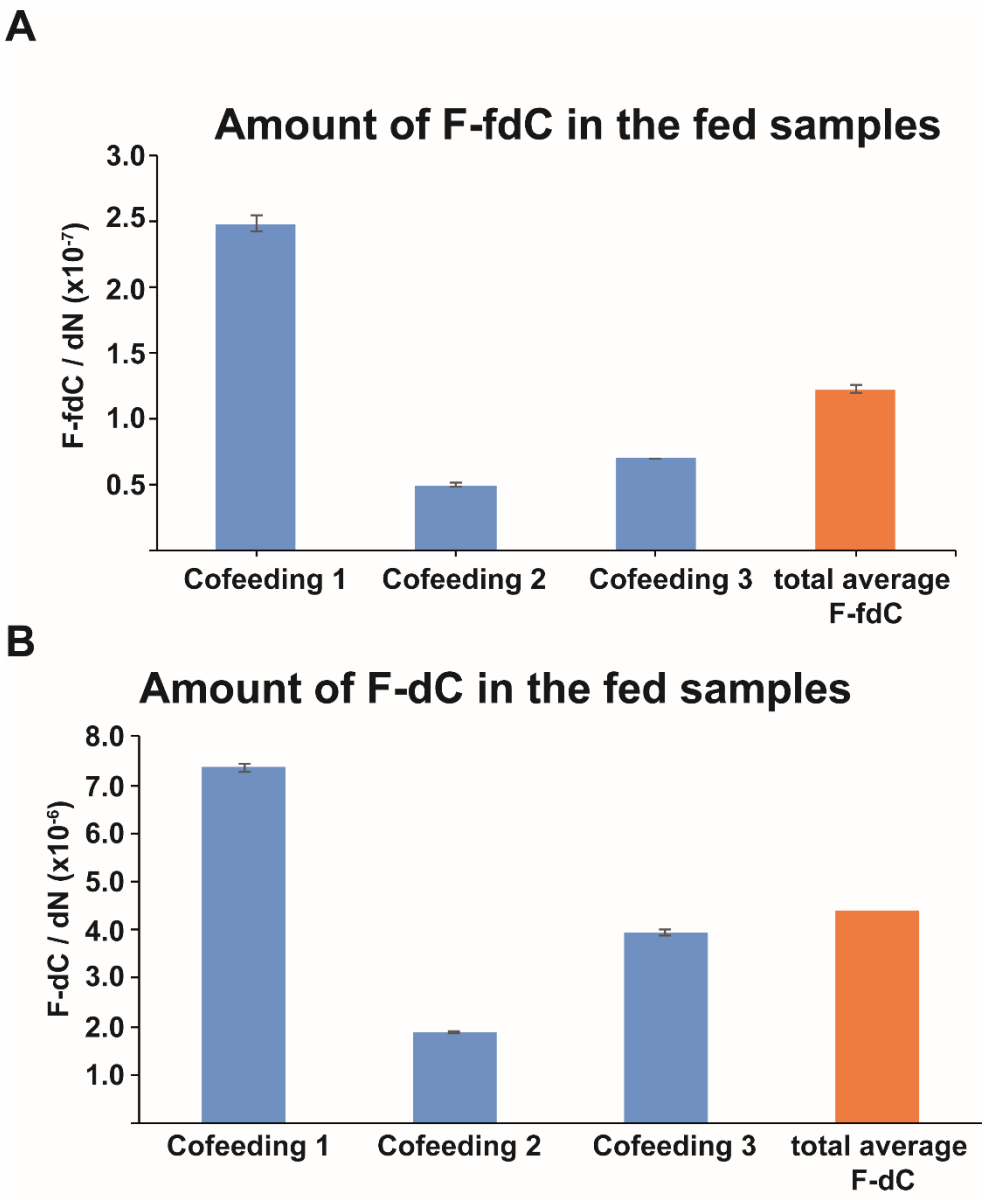


Figure 7 Measured amount of F-fdC (A) and F-dC (B) in the fed samples of N2a cells

Furthermore we were able to calculate the amount of deformylation of Fluoro-fdC by the amounts of Fluoro-fdC and Fluoro-dC. This was able to be done by the unnatural Fluoro compound. In Figure 8 the deformylation rates are shown.

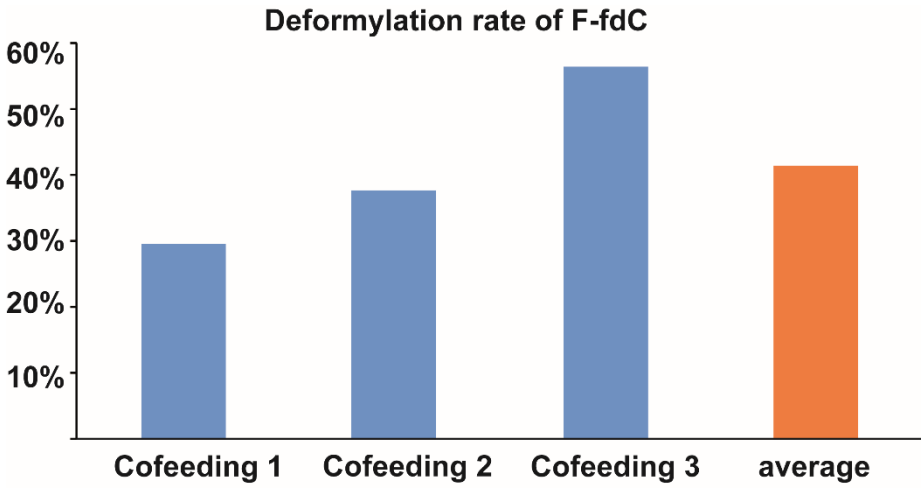
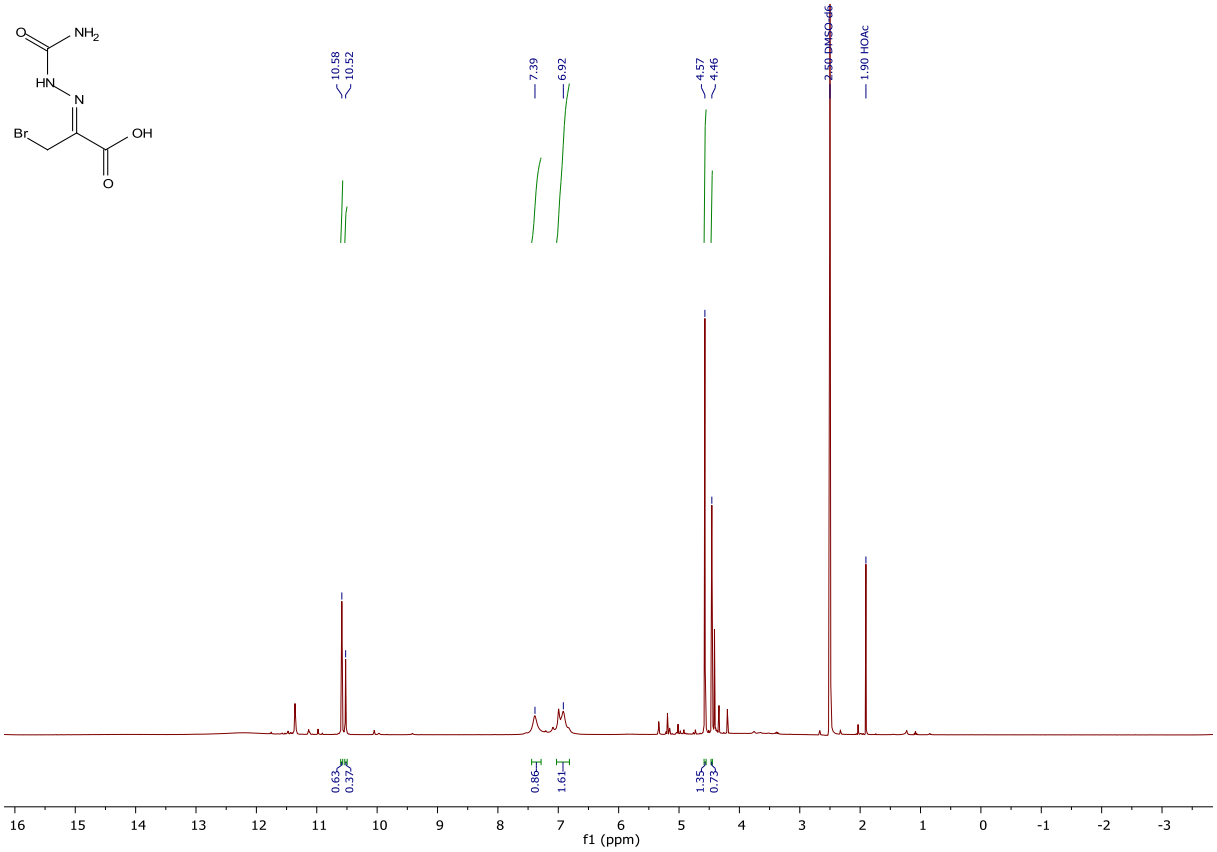
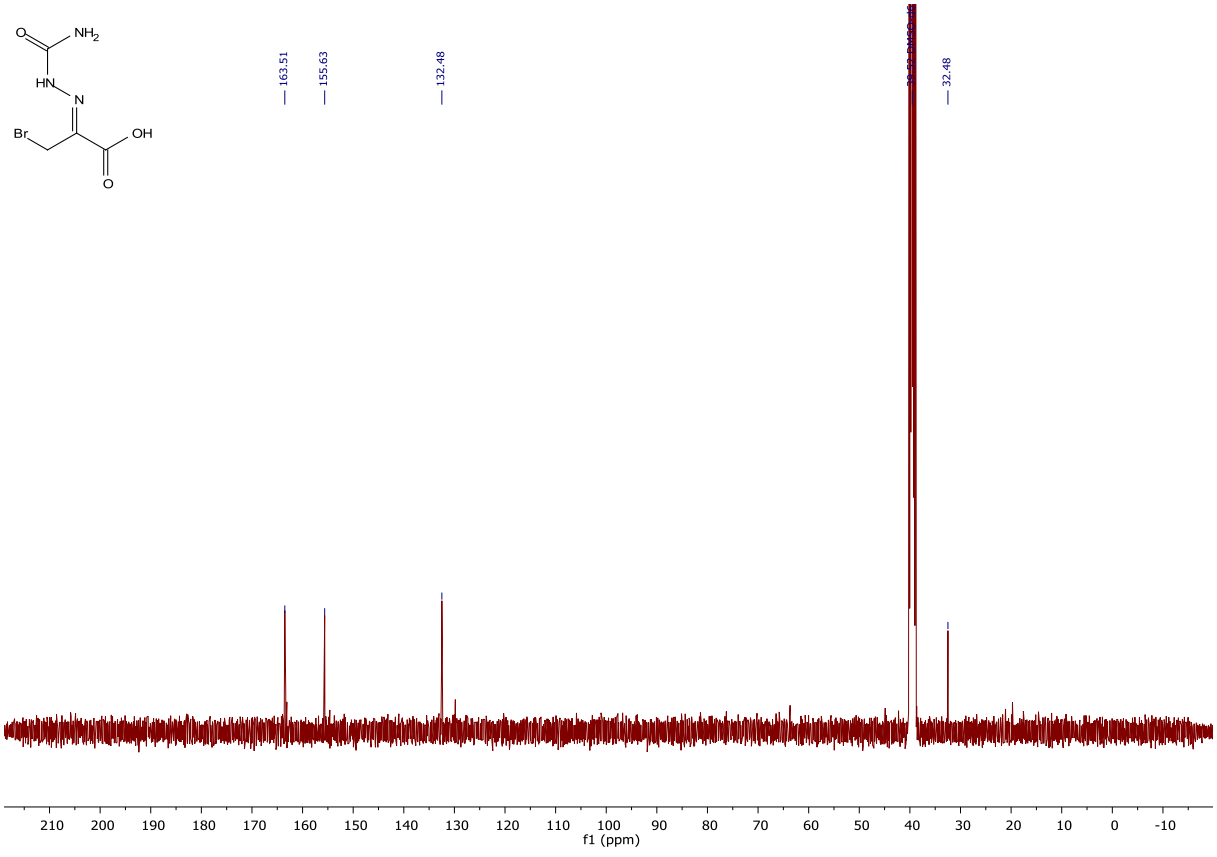
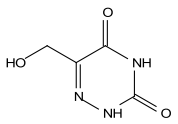


Figure 8 Calculated deformylation rate of F-fdC in N2a cells

NMR spectra of synthesized compounds

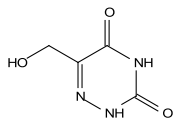





O
HO
NH
N
NH
O

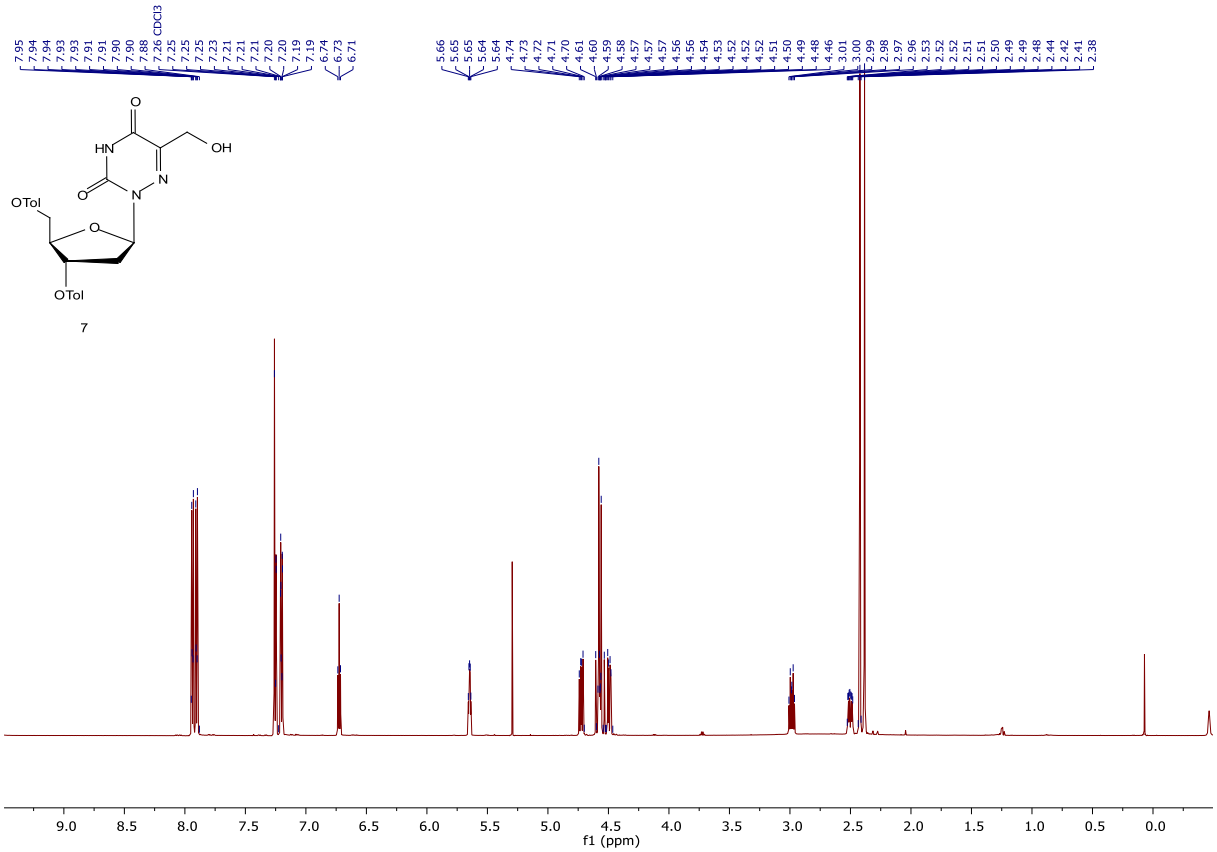
6

12.16
11.93
4.25
2.50 DMSO-d6
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
f1 (ppm)

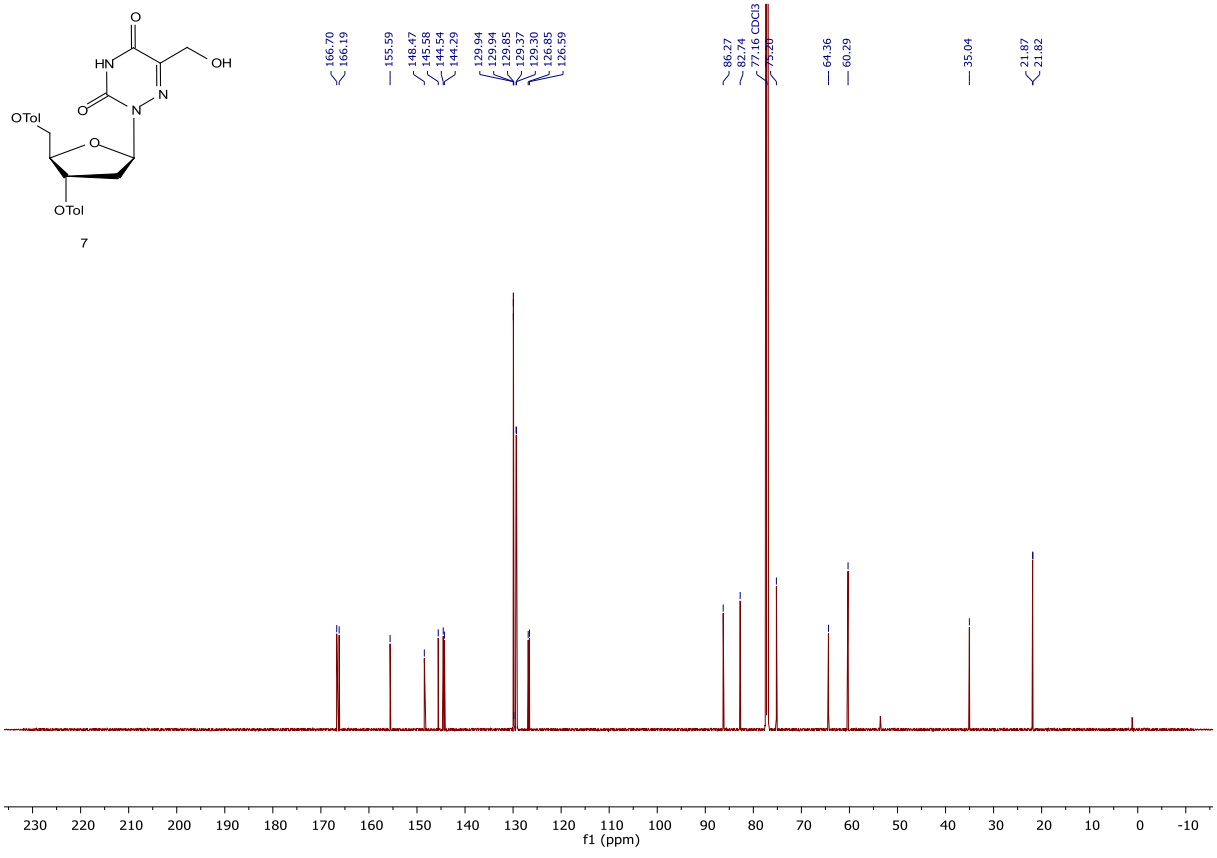

O
HO
NH
N
NH
O

6

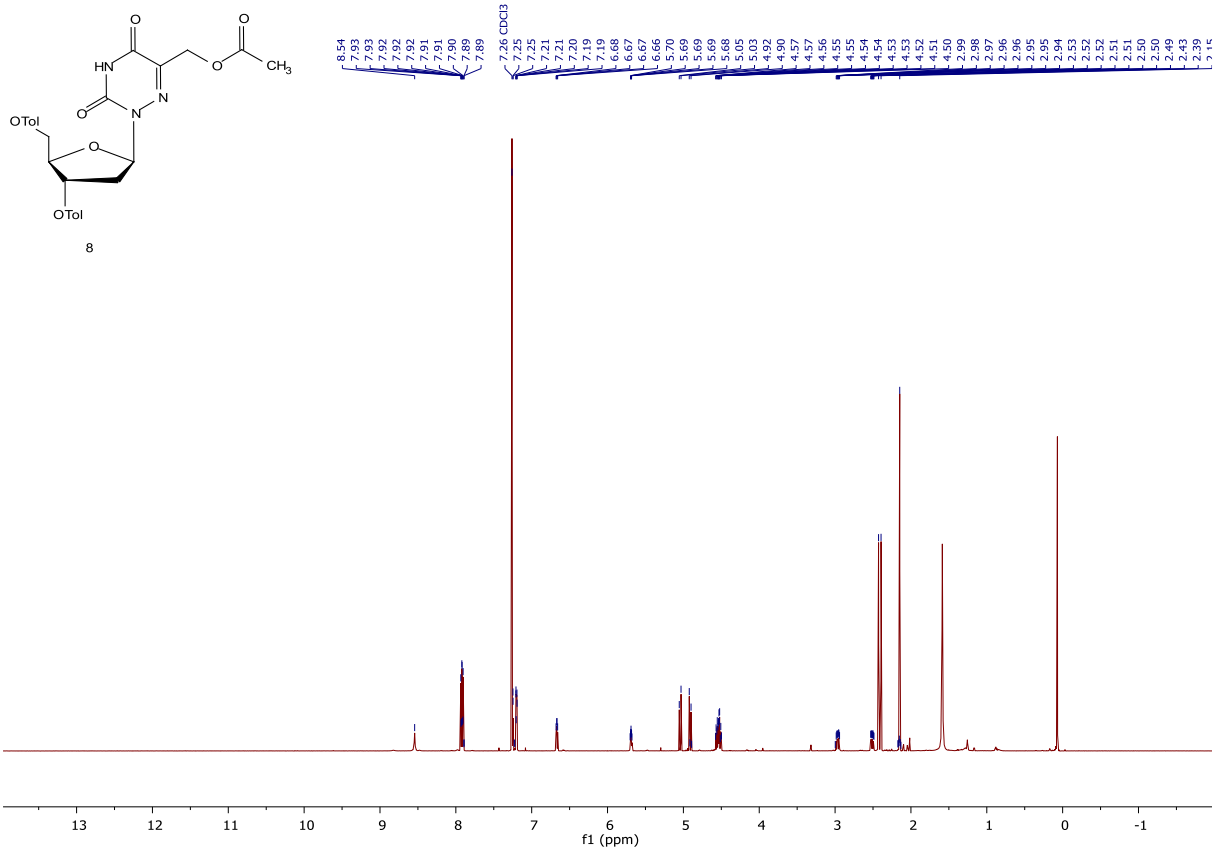
156.75
149.55
144.33
58.02
39.52 DMSO-d6
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-1
f1 (ppm)

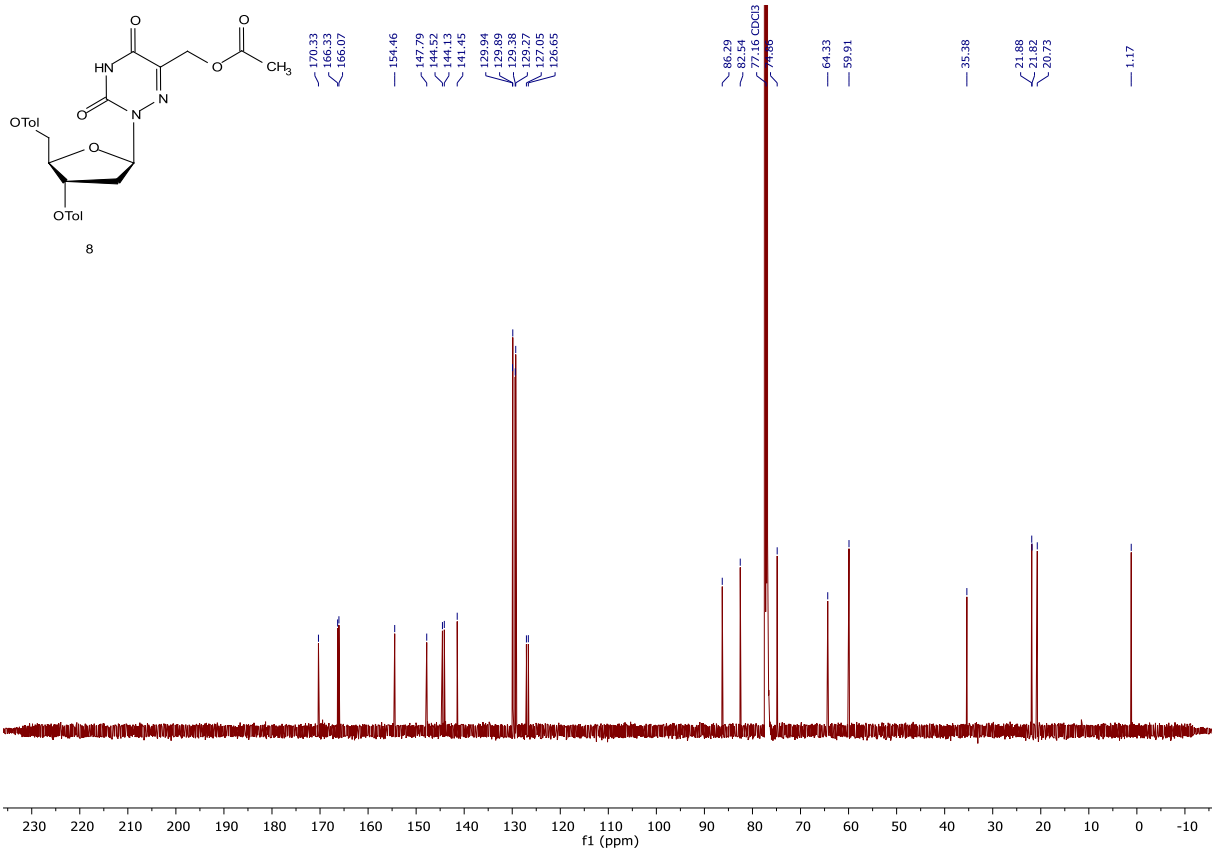

7.95 7.94 7.94 7.93 7.93 7.91 7.91 7.90 7.90 7.88 7.26 CDCl3 7.25 7.25 7.25 7.23 7.21 7.21 7.21 7.20 7.20 7.19 7.19 6.74 6.73 6.71
5.66 5.65 5.65 5.64 5.64 4.74 4.73 4.72 4.71 4.70 4.61 4.60 4.59 4.58 4.57 4.57 4.57 4.56 4.56 4.54 4.53 4.52 4.52 4.52 4.51 4.50 4.49 4.48 4.46 3.01 3.00 2.99 2.98 2.97 2.96 2.53 2.52 2.52 2.51 2.51 2.50 2.49 2.49 2.48 2.44 2.42 2.41 2.38
HN
O
N
OH
OTol
OTol
9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

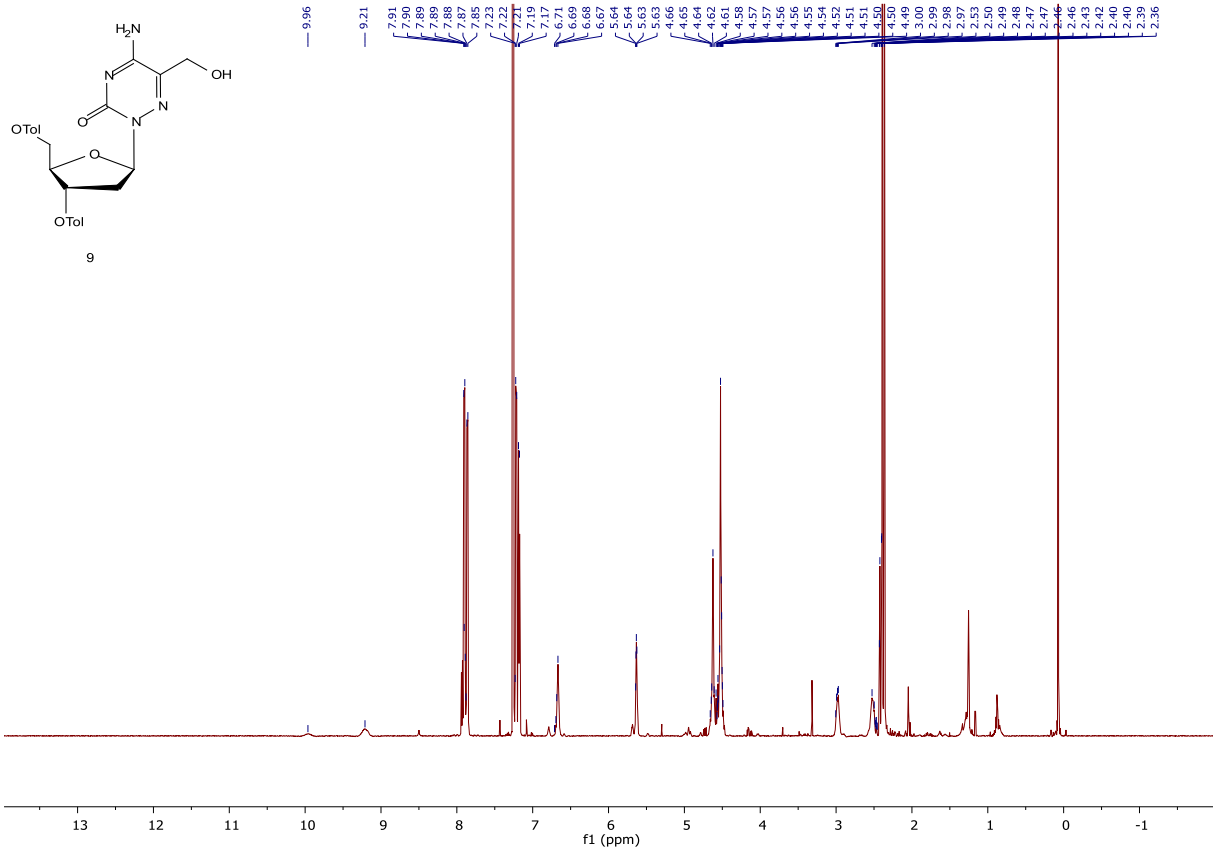
7

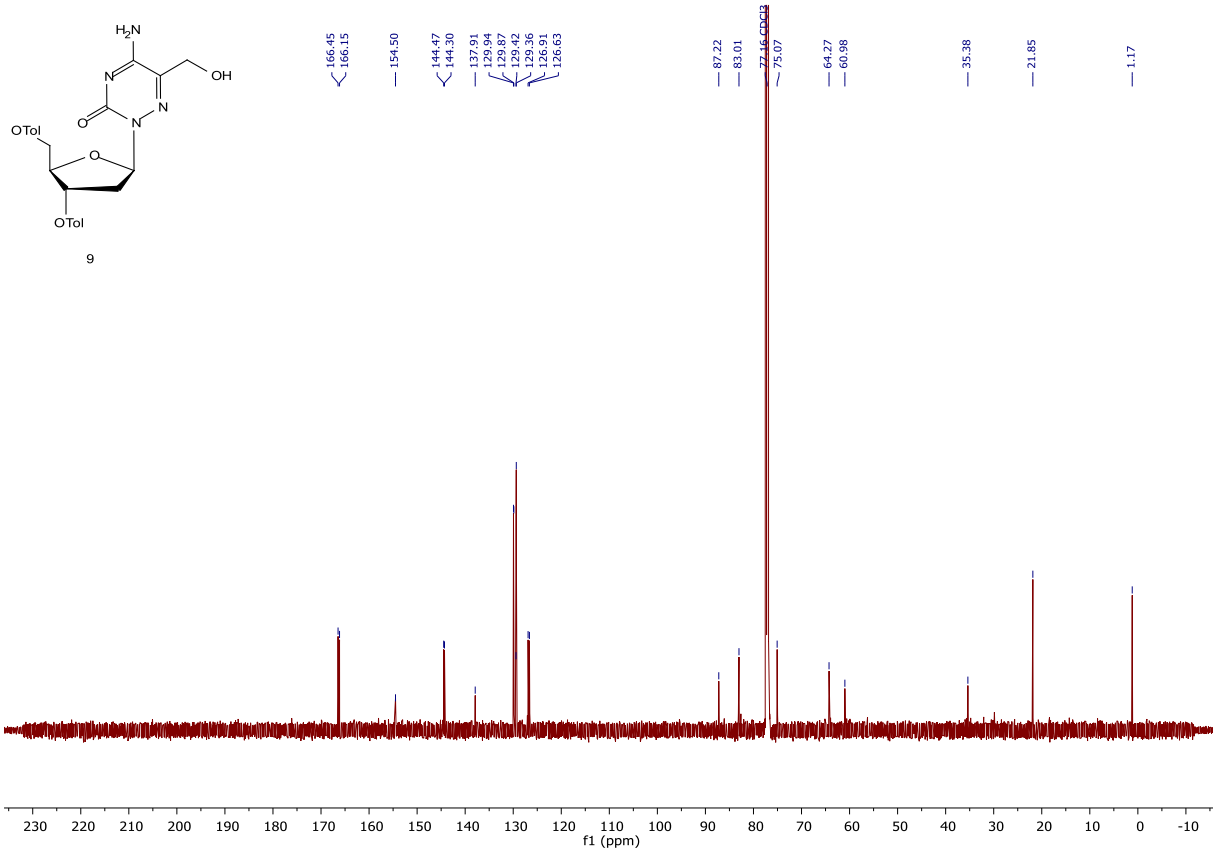

166.70 166.19 155.59 148.47 145.58 144.54 144.29 129.94 129.94 129.85 129.37 129.30 126.85 126.59 86.27 82.74 77.16 CDCl3 75.20 64.36 60.29 35.04 21.87 21.82
HN
O
N
OH
OTol
OTol
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

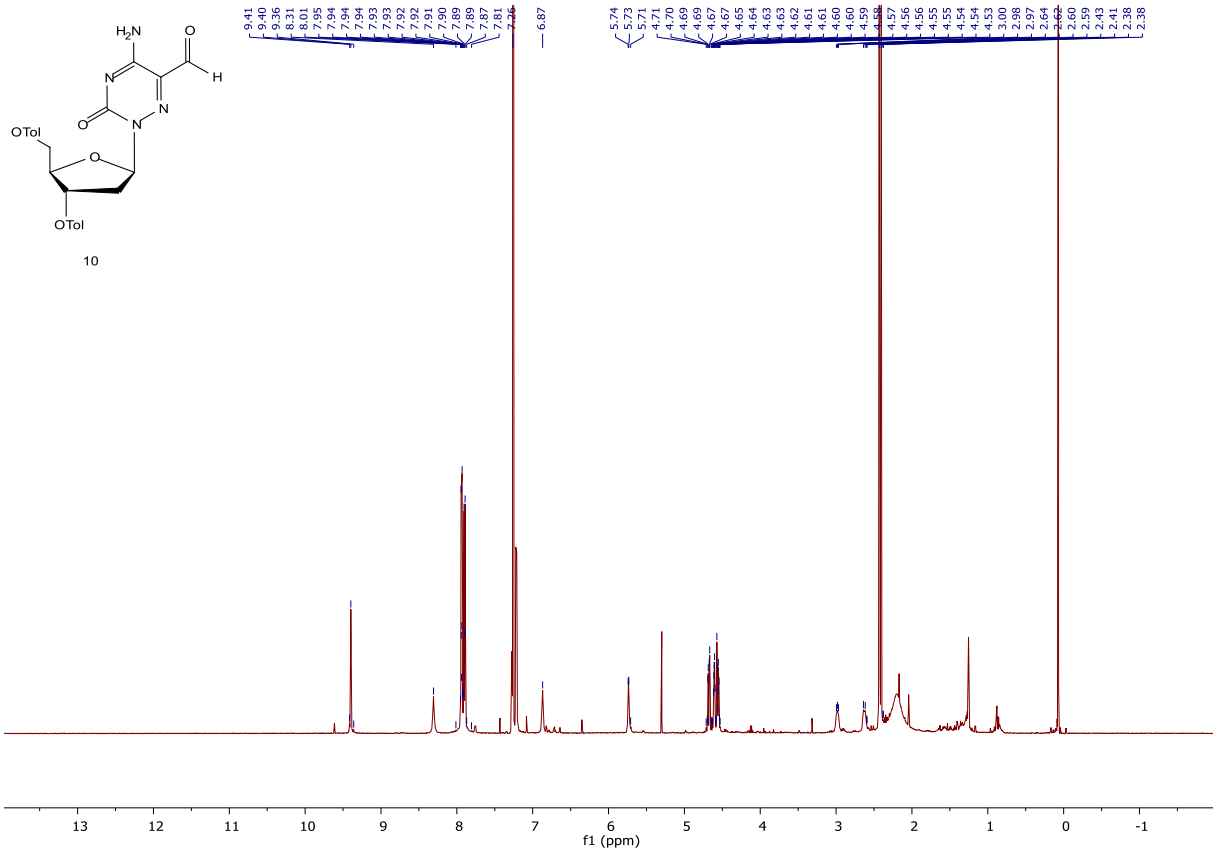
7

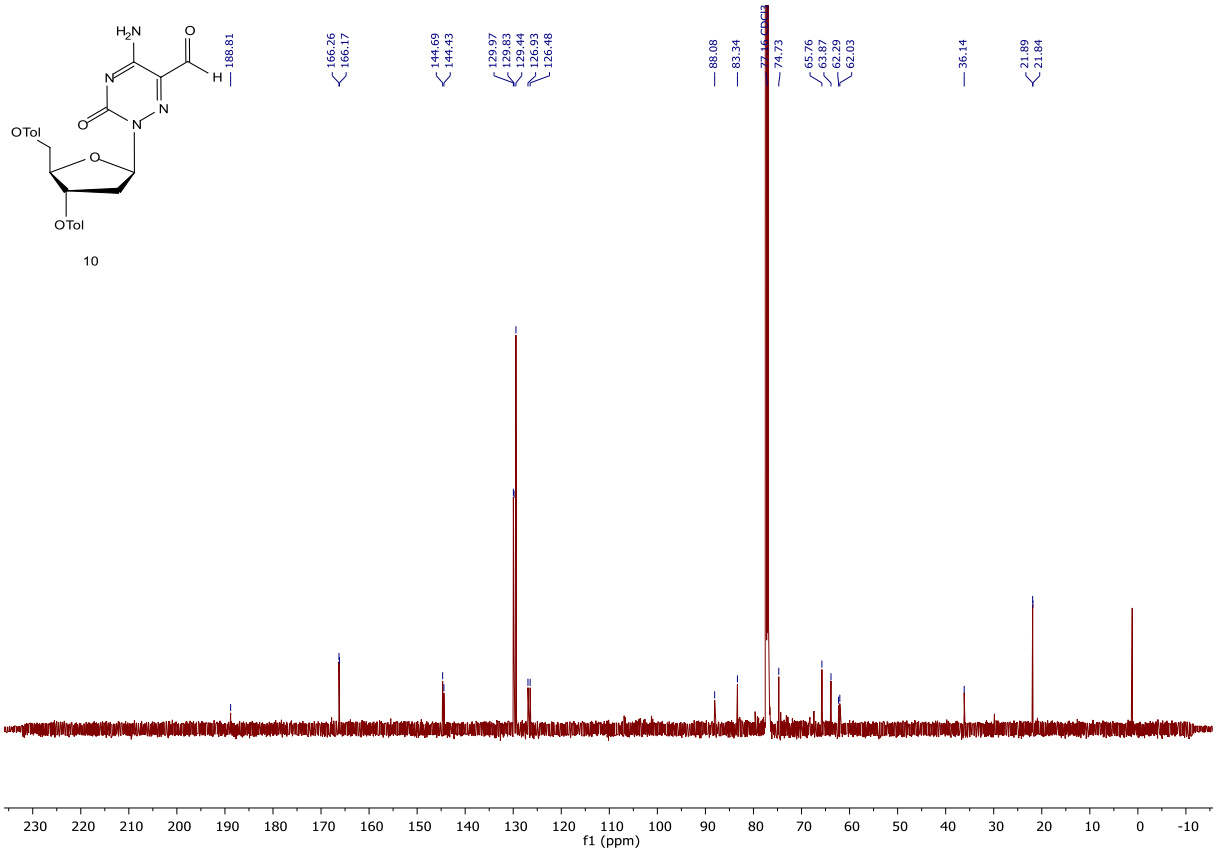

OTol
OTol
8
8.54
7.93
7.93
7.92
7.92
7.92
7.91
7.91
7.90
7.89
7.89
7.26 CDCl3
7.25
7.25
7.21
7.21
7.20
7.19
7.19
6.68
6.67
6.67
6.66
5.70
5.69
5.69
5.69
5.68
5.05
5.03
4.92
4.90
4.57
4.57
4.56
4.55
4.55
4.54
4.54
4.53
4.53
4.52
4.51
4.50
2.99
2.98
2.97
2.96
2.96
2.95
2.95
2.94
2.53
2.52
2.52
2.51
2.51
2.50
2.50
2.49
2.43
2.39
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f1 (ppm)

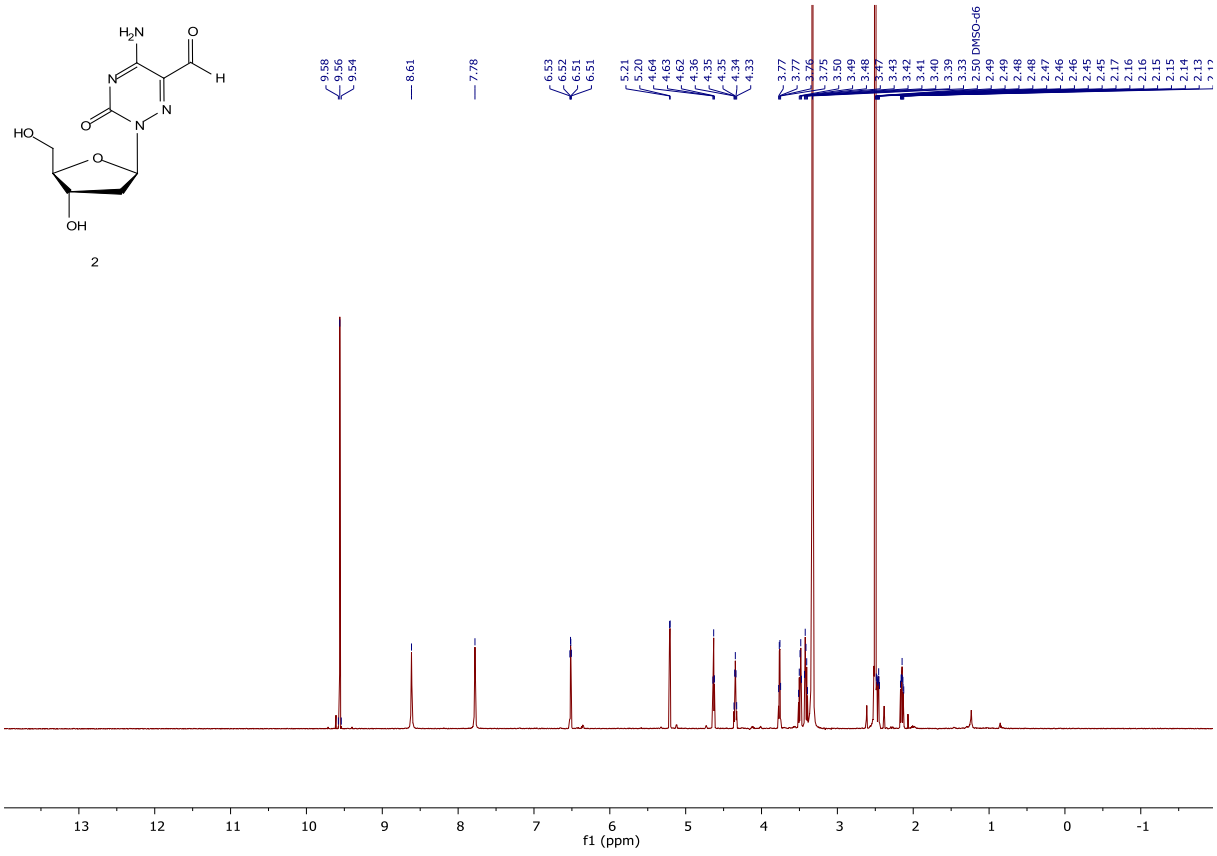

OTol
OTol
8
170.33
166.33
166.07
154.46
147.79
144.52
144.13
141.45
129.94
129.89
129.38
129.27
127.05
126.65
86.29
82.54
77.16 CDCl3
74.88
64.33
59.91
35.38
21.88
21.82
20.73
1.17
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


9.96
9.21
7.91
7.90
7.89
7.89
7.88
7.87
7.85
7.23
7.22
7.21
7.19
7.17
6.71
6.69
6.68
6.67
5.64
5.64
5.63
5.63
4.66
4.65
4.64
4.62
4.61
4.58
4.57
4.57
4.56
4.56
4.55
4.54
4.52
4.51
4.51
4.50
4.50
4.49
3.00
2.99
2.98
2.97
2.53
2.50
2.49
2.48
2.47
2.47
2.46
2.46
2.43
2.42
2.40
2.40
2.39
2.36
H2N
OH
N
N
O
N
OTol
O
OTol
9
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f1 (ppm)

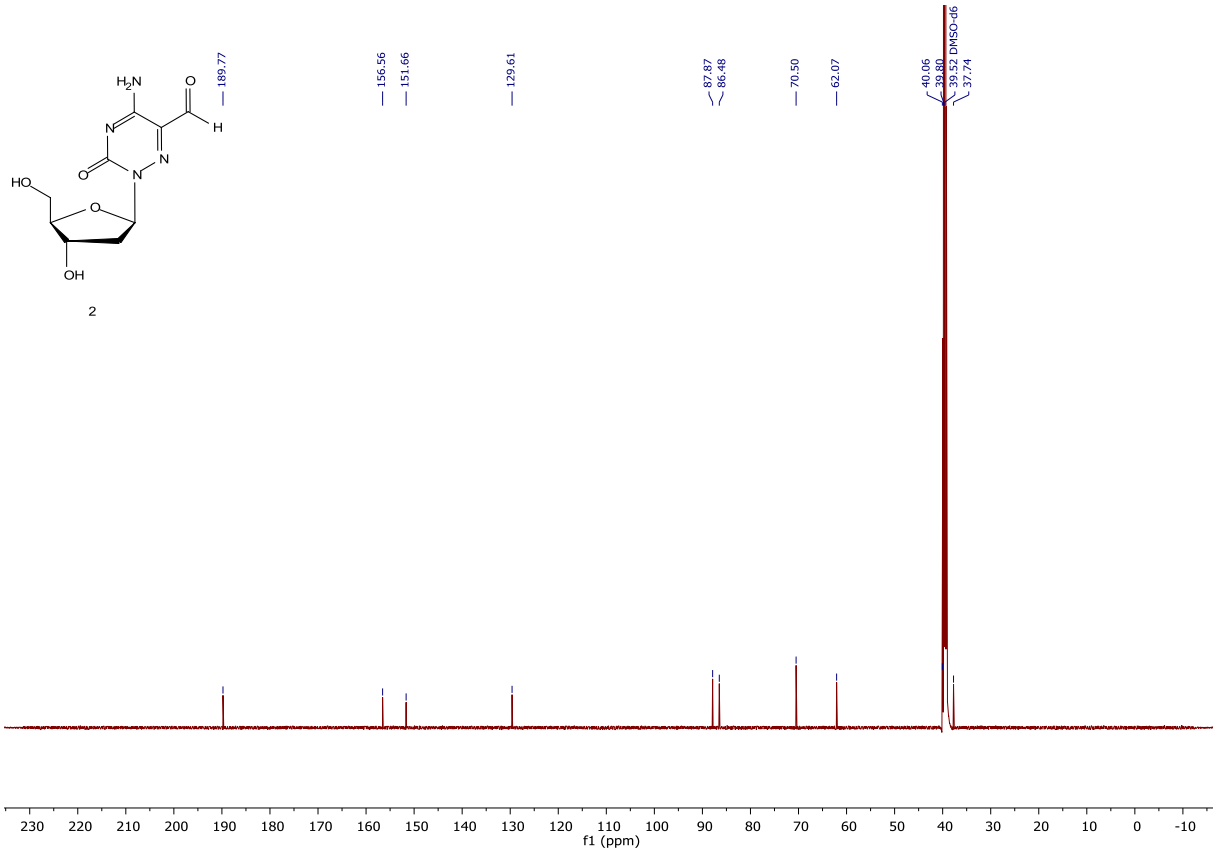

166.45
166.15
154.50
144.47
144.30
137.91
129.94
129.87
129.42
129.36
126.91
126.63
87.22
83.01
77.16 CDCl3
75.07
64.27
60.98
35.38
21.85
1.17
H2N
OH
N
N
O
N
OTol
O
OTol
9
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


9.41
9.40
9.36
8.31
8.01
7.95
7.94
7.94
7.93
7.93
7.92
7.92
7.91
7.90
7.89
7.89
7.87
7.81
7.26
6.87
5.74
5.73
5.71
4.71
4.70
4.69
4.69
4.67
4.67
4.65
4.64
4.63
4.63
4.62
4.61
4.61
4.60
4.60
4.59
4.58
4.57
4.56
4.56
4.55
4.55
4.54
4.54
4.53
3.00
2.98
2.97
2.64
2.62
2.60
2.59
2.43
2.41
2.38
2.38
OTol
OTol
10
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f1 (ppm)

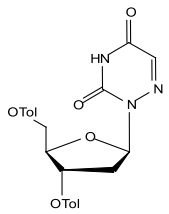

188.81
166.26
166.17
144.69
144.43
129.97
129.83
129.44
126.93
126.48
88.08
83.34
77.16 CDCl3
74.73
65.76
63.87
62.29
62.03
36.14
21.89
21.84
OTol
OTol
10
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


9.58
9.56
9.54
8.61
7.78
6.53
6.52
6.51
6.51
5.21
5.20
4.64
4.63
4.62
4.36
4.35
4.35
4.34
4.33
3.77
3.77
3.76
3.75
3.50
3.49
3.48
3.47
3.43
3.42
3.41
3.40
3.39
3.33
2.50 DMSO-d6
2.49
2.49
2.48
2.48
2.47
2.46
2.46
2.45
2.45
2.17
2.16
2.16
2.15
2.15
2.14
2.13
HO
O
OH
H
O
N
N
N
O
H_2N
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f1 (ppm)

2

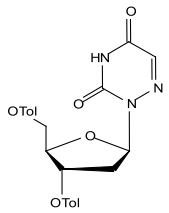

189.77
156.56
151.66
129.61
87.87
86.48
70.50
62.07
40.06
39.80
39.52 DMSO-d6
37.74
HO
O
OH
H
O
N
N
N
O
H_2N
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

2


O
HN
O
N
N
OTol
O
OTol

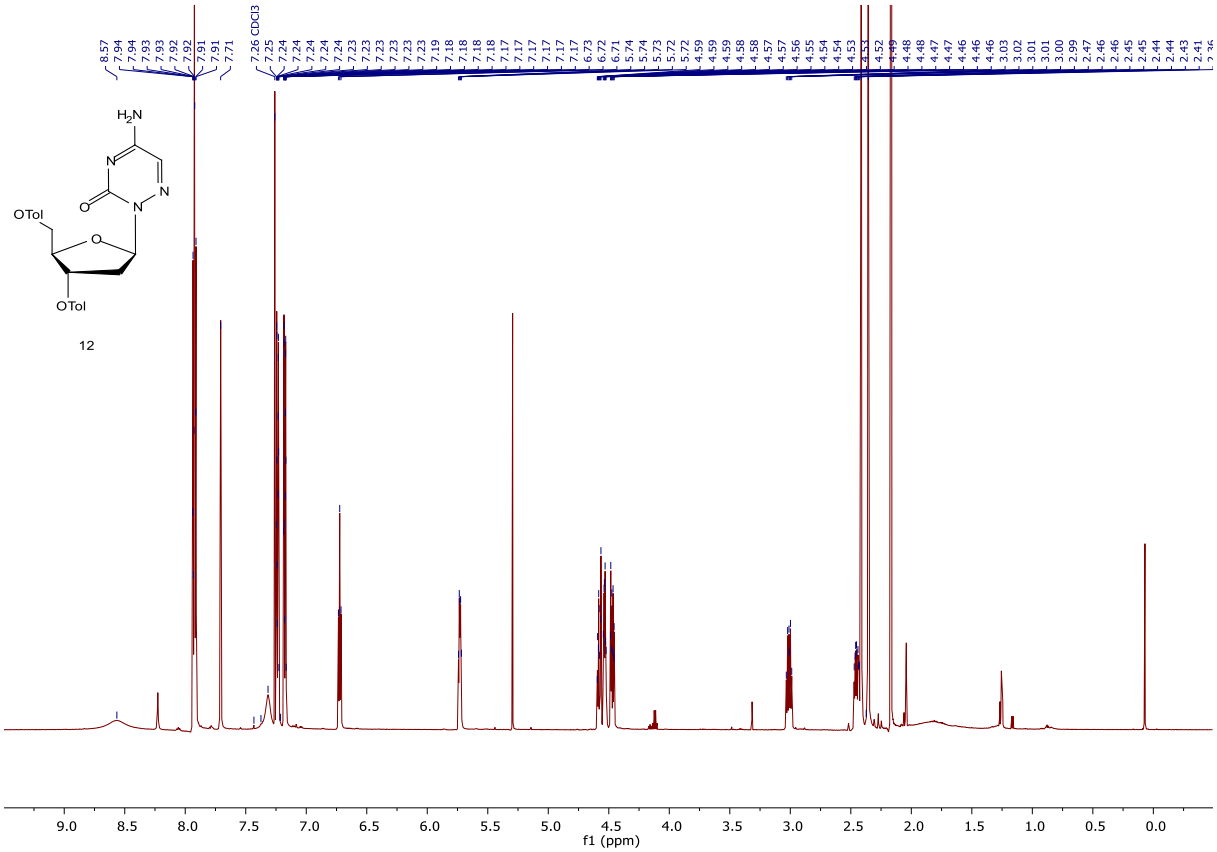
11

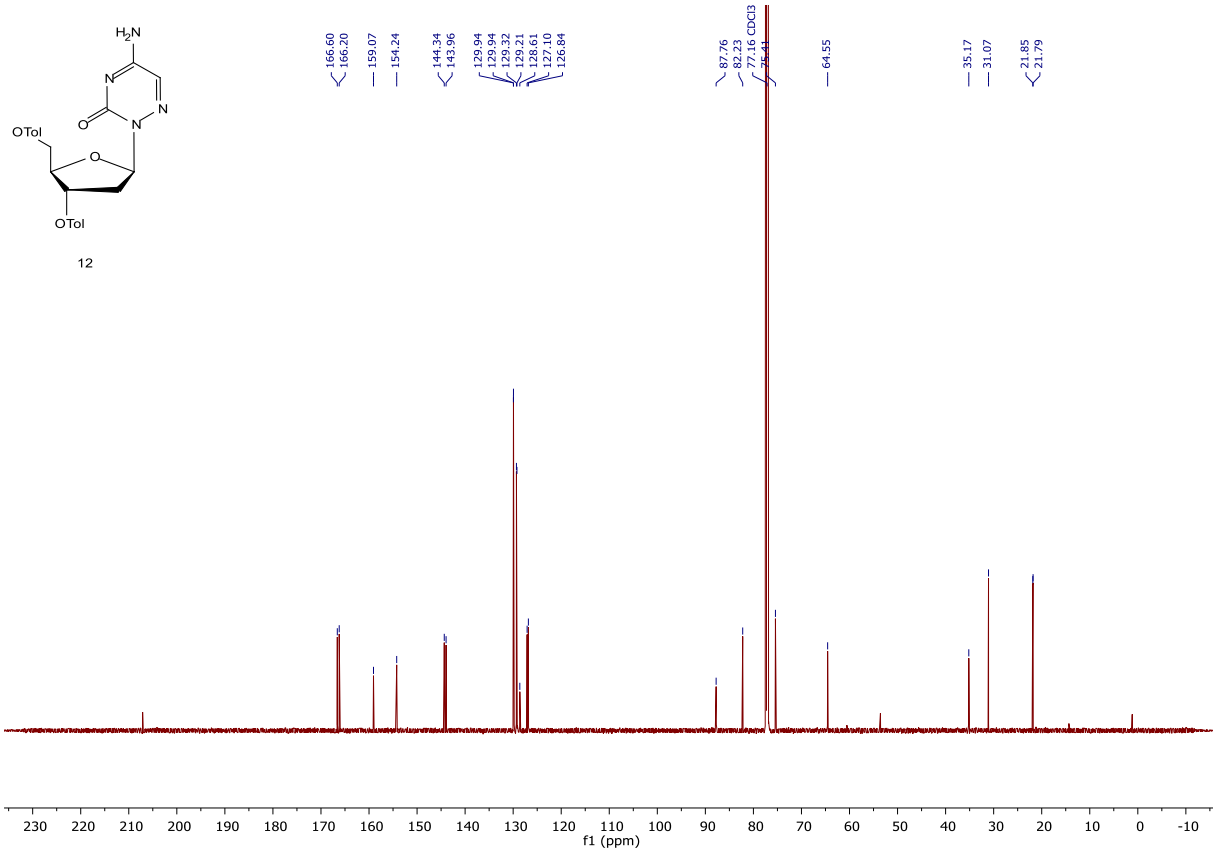
8.86
8.82
8.82
7.96
7.95
7.94
7.93
7.92
7.91
7.90
7.26 CDCl3
6.70
6.68
6.67
5.73
5.73
5.72
5.71
5.70
4.68
4.66
4.65
4.64
4.55
4.53
4.52
4.51
4.48
4.47
4.45
4.44
3.02
3.00
2.98
2.97
2.95
2.51
2.50
2.50
2.49
2.48
2.47
2.46
2.45
2.43
2.41
16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3
f1 (ppm)

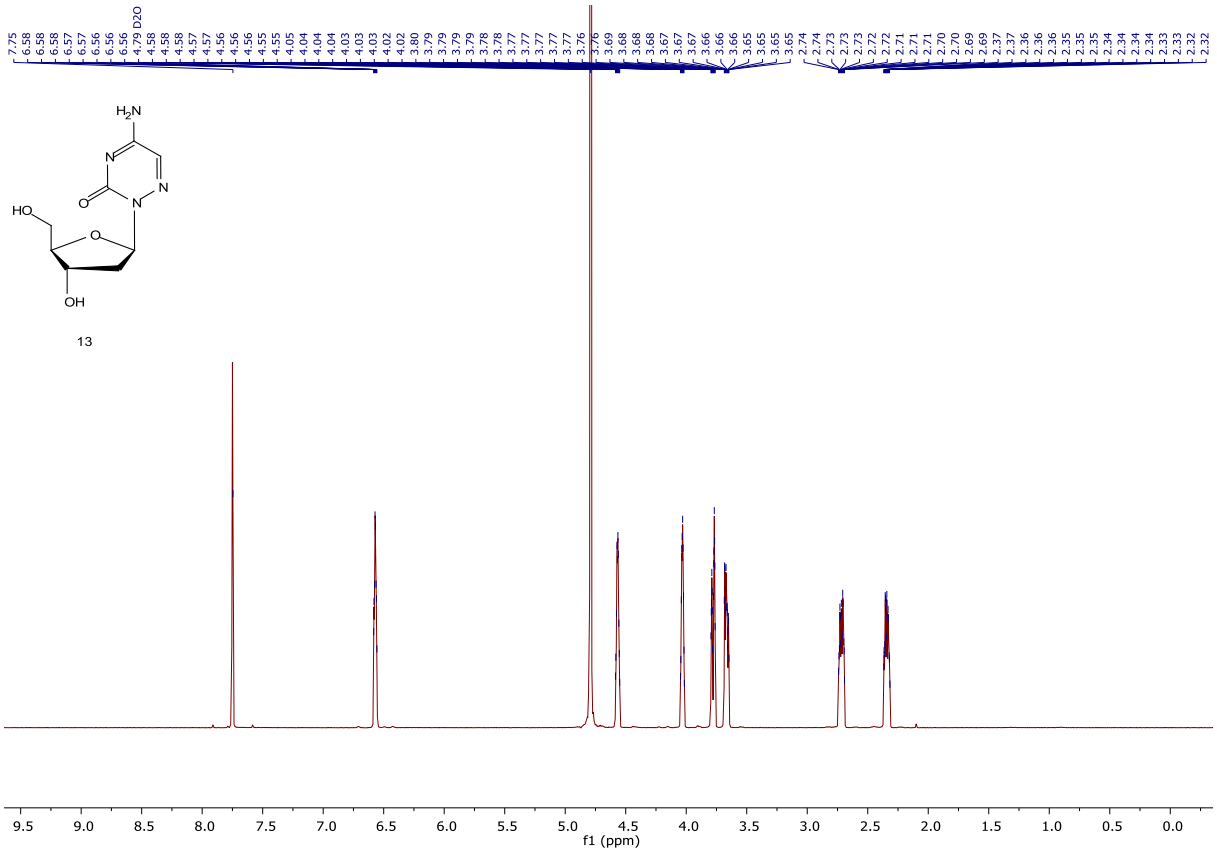

O
HN
O
N
N
OTol
O
OTol

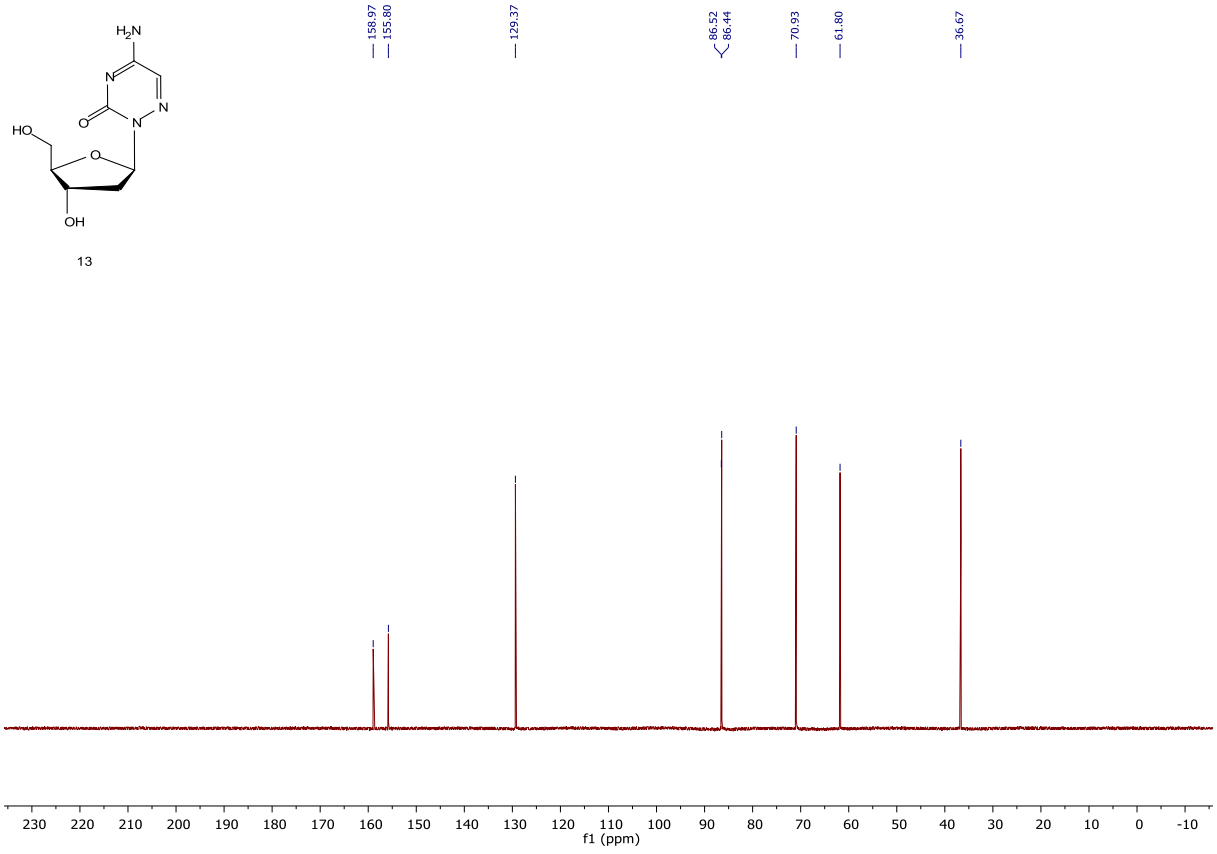
11

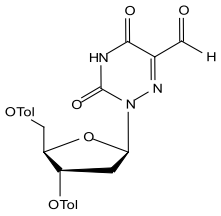
166.36
166.16
155.30
147.77
144.56
144.11
136.10
129.94
129.90
129.39
129.27
127.11
126.61
86.05
82.73
77.16 CDCl3
75.03
64.06
35.17
21.90
21.86
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


8.57
7.94
7.94
7.93
7.93
7.92
7.92
7.91
7.91
7.71
7.26 CDCl3
7.25
7.24
7.24
7.24
7.24
7.24
7.23
7.23
7.23
7.23
7.23
7.23
7.19
7.18
7.18
7.18
7.18
7.17
7.17
7.17
7.17
7.17
7.17
6.73
6.72
6.71
5.74
5.74
5.74
5.73
5.72
5.72
4.59
4.59
4.59
4.58
4.58
4.57
4.57
4.56
4.55
4.54
4.54
4.53
4.52
4.48
4.48
4.47
4.47
4.46
4.46
4.46
3.03
3.02
3.01
3.01
3.00
2.99
2.47
2.46
2.46
2.45
2.45
2.44
2.44
2.43
2.41
H2N
N
N
O
N
OTol
O
OTol
12
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)


H2N
N
N
O
N
OTol
O
OTol
12
166.60
166.20
159.07
154.24
144.34
143.96
129.94
129.94
129.32
129.21
128.61
127.10
126.84
87.76
82.23
77.16 CDCl3
76.41
64.55
35.17
31.07
21.85
21.79
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

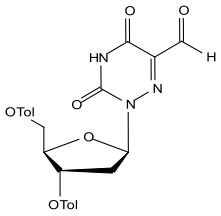

7.75
6.58
6.58
6.58
6.57
6.57
6.56
6.56
6.56
4.79 D2O
4.58
4.58
4.58
4.57
4.57
4.56
4.56
4.56
4.55
4.55
4.05
4.04
4.04
4.04
4.03
4.03
4.03
4.02
4.02
3.80
3.79
3.79
3.79
3.79
3.78
3.78
3.77
3.77
3.77
3.77
3.77
3.76
3.76
3.69
3.68
3.68
3.68
3.67
3.67
3.67
3.66
3.66
3.66
3.65
3.65
3.65
3.65
2.74
2.74
2.73
2.73
2.73
2.72
2.72
2.71
2.71
2.71
2.70
2.70
2.70
2.69
2.69
2.37
2.37
2.36
2.36
2.36
2.35
2.35
2.35
2.34
2.34
2.34
2.34
2.33
2.33
2.32
2.32
H2N
N
N
N
O
HO
O
OH
13
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)


158.97
155.80
129.37
86.52
86.44
70.93
61.80
36.67
H2N
N
N
N
O
HO
O
OH
13
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


O
O
HN
H
N
O
N
OTol
O
OTol

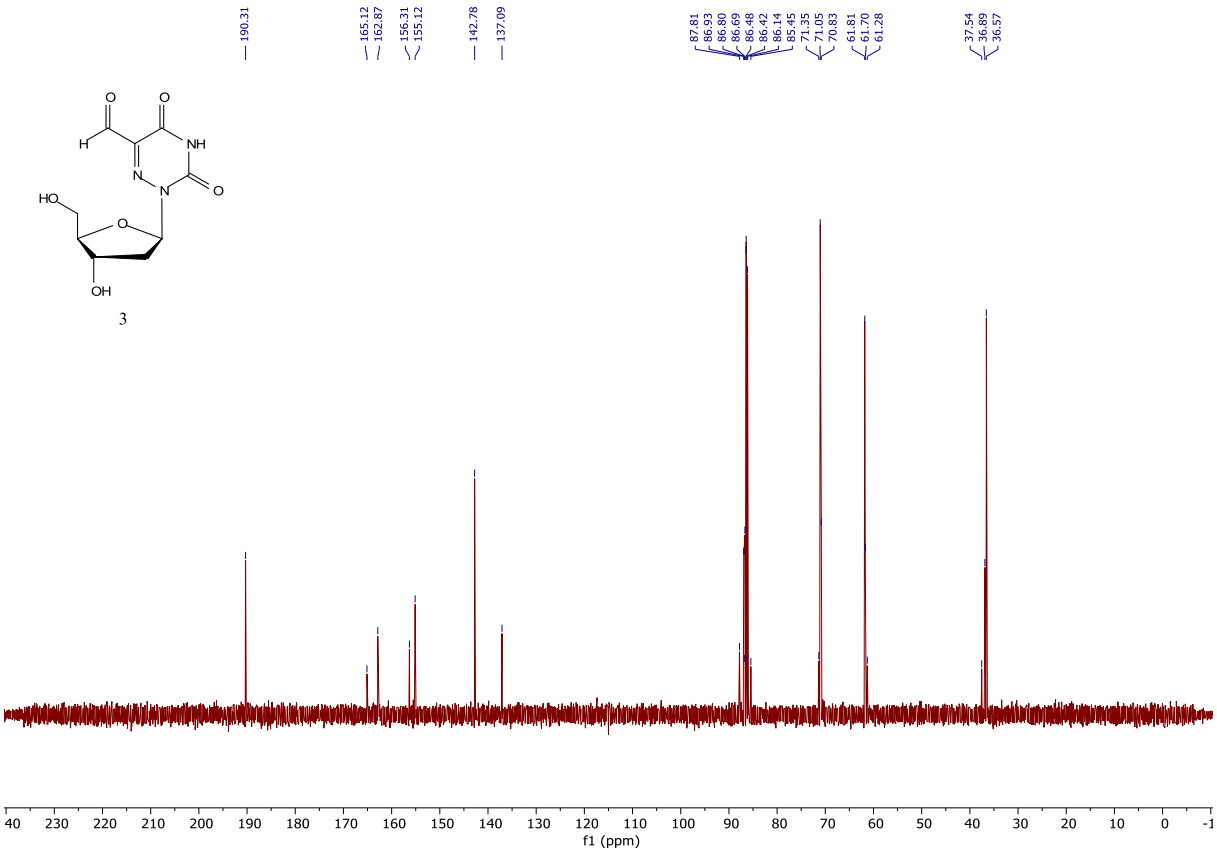
14

9.60
9.58
9.55
9.02
7.95
7.94
7.93
7.93
7.92
7.92
7.91
7.90
7.90
7.89
7.88
7.87
7.85
7.26 CDCl3
6.73
6.71
6.70
6.68
6.67
6.66
5.73
5.72
5.72
5.71
4.67
4.66
4.64
4.64
4.62
4.62
4.60
4.59
4.58
4.57
4.57
4.56
4.54
4.53
3.10
3.08
3.07
3.06
3.05
3.03
2.61
2.60
2.59
2.58
2.57
2.56
2.56
2.55
2.45
2.43
2.41
2.41
2.40
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
f1 (ppm)

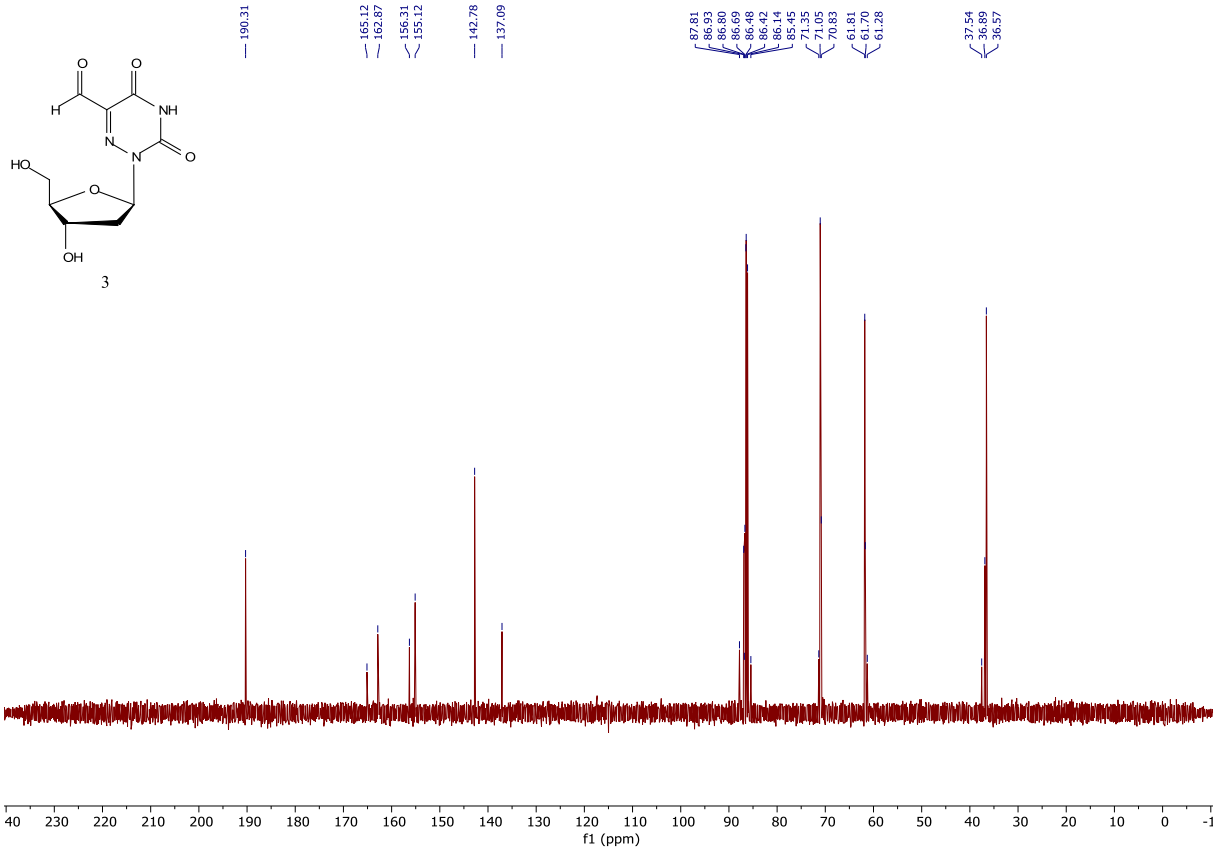

O
O
HN
H
N
O
N
OTol
O
OTol

14

184.16
166.33
166.14
152.79
147.15
144.68
144.40
137.22
129.94
129.84
129.81
129.70
129.54
129.42
129.40
126.81
126.41
87.16
84.08
83.20
77.48 CDCl3
77.36
77.16
76.84 CDCl3
74.59
63.74
35.40
29.85
21.91
21.90
21.85
1.17
40
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-1
f1 (ppm)


190.31
165.12
162.87
156.31
155.12
142.78
137.09
87.81
86.93
86.80
86.69
86.48
86.42
86.14
85.45
71.35
71.05
70.83
61.81
61.70
61.28
37.54
36.89
36.57
O
O
H
NH
N
N
O
HO
O
OH
40 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -1
f1 (ppm)

3


190.31
165.12
162.87
156.31
155.12
142.78
137.09
87.81
86.93
86.80
86.69
86.48
86.42
86.14
85.45
71.35
71.05
70.83
61.81
61.70
61.28
37.54
36.89
36.57
O
O
H
NH
N
N
O
HO
O
OH
40 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -1
f1 (ppm)

3

Bibliography

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

[2] W. L. Mitchell, P. Ravenscroft, M. L. Hill, L. J. S. Knutsen, B. D. Judkins, R. F. Newton, D. I. C. Scopes, *J. Med. Chem.* **1986**, *29*, 809-816.

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

Acknowledgements

Firstly, I want to thank my PhD advisor Prof. Dr. Thomas Carell for the opportunity, support and guidance he has provided throughout the last 3 years I spent in his group. It was an honour for me to work in your group with all the possibilities you offered me and the whole group during this time. It was a great feeling that we were able to focus just on science.

I also want to thank Dr. Pavel Kielkowski for taking over the position of the second reviewer and all the help he offered, when I struggled with mass spec issues.

I would like to thank Prof. Dr. med. Thomas Gudermann and PD. Dr med. Claudia Staab-Weijnitz, for their work as my thesis advisors and their work for the GRK2338.

I would like to thank Dr. Mirko Wagner and Dr. Katharina Iwan who were great teacher and co-worker in mass spectrometry. I am thankful for the discussions and collaborations we had.

Further I would like to thank Dr. Markus Müller for the interesting discussions we had regarding chemistry, biochemistry and other topics during my PhD. I always felt delighted for the effort you showed everyday organizing stuff for us.

I thank Mrs. Slava Gärtner for the management of all bureaucratic issues.

Further I want to thank Kerstin Kurz for her immeasurable help with Elfriede and Paul as well as her help with several digestions and DNA and RNA isolations during my PhD.

I would like to thank Dr. Markus Müller, Dr. Mirko Wagner, Dr. Alexander Schön, Jonas Feldmann and Stefan Wiedemann for finding time to comment and help with corrections of my PhD thesis.

In addition, I would like to thank all the people that I met in Carell group for this awesome time. I really enjoyed the time here. Especially the “Perfect-Dinner”-Cooking-Group with the chefs Stefan (ich lose die Bewertungen, ah verdammt nur 4 Punkte), Simon (das ist mein absolutes Lieblingsessen), Johann (gebratener Reis rules the world), Jonas (Mettwoch) and Felix (Fertigessen olé olé). Also the “Kartler” Markus (Grünsolo was anderes mag ich ned), Felix (was mach ich jetzt, wenn ich das nicht habe), Stefan (wenn keiner mag dann spiele ich wohl) and Alexander (ach komm scho, wieder kein Geld mehr).

Thanks to my sport, working and partying friends Olli (wie spontan bist Du?-Treffen), Julia (für die wunderschönen letzten 7 Jahre mit Dir und wenn selbst die Eltern wissen, dass eine Wohnungsparty stattfindet Du aber nicht :-D), Hippo (zahlreiche „intellektuelle“ Gespräche und aufbauende

Worte), Muskel-Martin (Reisen Sport und vieles mehr, #heuteweglöten), Rene (die vielen interessanten Gespräche beim Essen), Burny (weiterhin on fire), Chris (ahoi moje auto se zastavilo), Jochen (die vielen Radtouren und die Möglichkeit immer Deine Wohnung in Regensburg nutzen zu dürfen), Alex und Stefan (für die immerwährende Unterstützung), Michi (die Ausflüge, Urläube, Messen und Trainings) and Juliana the Smurf (for all the conversations we had and don't forget, boarisch musst Du aba no lerna) and many more who went this way with me.

Special thanks go to my parents, my grandparents, my brother and sister who always supported me, believed in me and my work. They always offered me help of all kind during the time beginning with school until the thesis submission today. I really appreciated it!