Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Investigation of the Mutagenic Potential of Naturally Occurring Oxidized DNA Nucleobase Derivatives

Ulrike Lischke

aus

Dresden, Deutschland

2013

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 13. Mai 2013

Ulrike Lischke

Dissertation eingereicht am 28. Mai 2013

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 27. Juni 2013

für meinen Vater

“Paths are made by walking.”
Franz Kafka

Parts of this thesis were published or presented at conferences:

Publications

Gehrke, T.H.,[1] Lischke, U.,[1] Gasteiger, K. L.,[1] Schneider, S.,[1] Arnold, S., Müller, H. C., Stephenson, D. S., Zipse, H., Carell, T. *"Unexpected non-Hoogsteen-based mutagenicity mechanism of FaPy-DNA lesions" Nat. Chem. Bio.* **2013**, *9*, 455-461.

Stathis, D.,[1] Lischke, U., Koch, S.C. Deiml, C. A., Carell, T. *"Discovery and mutagenicity of a guanidinoformimine lesion as a new intermediate of the oxidative deoxyguanosine degradation pathway." J. Am. Chem. Soc.* **2012**, *134*, 4925-4930.

Münzel, M.,[1] Lischke, U., Stathis, D. Pfaffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K., Carell, T. *"Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine." Chemistry* **2011**, *17*, 13782-13788.

Further Publications

Heil, K.,[1] Kneuttinger, A.C., Schneider, S., Lischke, U., Carell, T. *„Crystal structures and repair studies reveal the identity and the base-pairing properties of the UV-induced spore photoproduct DNA lesion." Chemistry* **2011**, 17, 9651-9657.

Schmidt, A.,[1] Haferburg, G., Schmidt, A., Lischke, U., Merten, D., Ghergel, F., Büchel G., Kothe, E. *"Heavy metal resistance to the extreme: Streptomyces strains from a former uranium mining area" Geochemistry* **2009**, 69 Suppl. 2, 35-44.

Conference Presentations

Poster presentation: *"Replication of Stabilized Bioisosteric Formamidopyrimidine Analogues inside Various DNA Polymerases"*, interact, Munich, **2012**.

Poster presentation: *"Replication of Stabilized Bioisosteric Formamidopyrimidine Analogues inside Various DNA Polymerases"*, Structural Biology & DNA Repair, Amsterdam, Netherlands, **2012**.

[1] first author(s)

Acknowledgements

An dieser Stelle soll all den Menschen, die zum Gelingen der vorliegenden Arbeit beigetragen haben, gedankt werden.

Ein herzliches Dankeschön gebührt in erster Linie *Prof. Dr. Thomas Carell.* Vielen Dank für die Aufnahme in deine Gruppe und die Möglichkeit an interessanten und vielseitigen Themen arbeiten zu können. Dein unendlicher Optimismus, dein Engagement, deine Motivationsgabe und deine Unterstützung trugen zum Erfolg dieser Arbeit ebenso bei wie das mir entgegengebrachte Vertrauen.

Mein Dank gilt außerdem *Prof. Dr. Karl-Peter Hopfner*, welcher sich als Zweitgutachter dieser Arbeit zur Verfügung stellte.

Desweiteren danke ich den Mitgliedern der Prüfungskommission *Prof. Dr. Konstantin Karaghiosoff*, *PD Dr. Dietmar Martin*, *Prof. Dr. Klaus Förstemann* und *Prof. Dr. Angelika Vollmar.*

Slava Gärtner danke ich für ihre Unterstützung bei allen bürokratischen Belangen.

Dr. Sabine Schneider danke ich für die vielen Fahrten zum Synchrotron, das Messen unserer Kristalle und die anschließende Auswertung.

Mein großer Dank gebührt *Dr. Markus Müller* für sein unerschöpfliches Wissen und seine Hilfe in jeglichen Belangen. Dein Ideenreichtum und deine kritische, aber sehr wertvolle Meinung waren in vielerlei Hinsicht stets eine Bereicherung für mich.

Dr. Tim Gehrke, *Dr. Dimitrios Stathis, Dr. Martin Münzel, Sandra Koch* und *Karola Gasteiger* danke ich für die sehr gute Zusammenarbeit. Der stetige Wissensaustausch und die gegenseitige Unterstützung haben viel zum Gelingen unserer Arbeiten beigetragen.

Benjamin Hackner, *Caterina Brandmayr*, *Dr. Markus Müller* und *Dr. Tim Gehrke* danke ich für das Korrekturlesen meiner Doktorarbeit.

Der Lauftruppe und vor allem *Barbara Steigenberger* danke ich für den nötigen Ausgleich im Fürstenrieder Wald.

Vielen Dank auch an *Christian Deiml* und *Olga Kotljarova* für die tolle und sehr angenehme Laboratmosphäre sowie interessante, wissenschaftliche Diskussionen und Hilfe.

Johannes Harder und *Benjamin Hackner* danke ich für den schönen Radlstart in den Tag und viele kulinarisch sehr wertvolle Abende.

Olga Kotljarova und *Karola Gasteiger*, der morgendliche Kaffee mit Euch war ein Muss (und Genuss), welcher mir in Zukunft sehr fehlen wird.

Dr. Antje Hienzsch und *Caterina Brandmayr*, Euch danke ich für unsere schöne Zeit miteinander und die vielen Momente der Freude sowie die gegenseitige Motivation und Unterstützung in Momenten wo wir sie brauchten.

Dr. María Tomás Gamasa, muchas gracias por tu ayuda con mi español, nuestras mañanas con desayuno y por tu amistad.

Dr. Antje Hienzsch, *Dr. Danila Fazio*, *Dr. Dimitrios Stathis* und *Dr. David Pearson* danke ich für die wundervolle Zeit miteinander und vor allem für unsere lustigen Abende in griechischen oder italienischen Restaurants.

Ich danke der Gruppe für die tolle Atmosphäre, gute Teamarbeit, für schöne Feiern, den ganzen Kuchen, Grillabende, Mädelsabende und tolle Ausflüge. Es war eine wunderschöne Zeit mit Euch.

Meiner Familie und meinem Freund *Swen Wahl* danke ich für ihre stetige Unterstützung während meiner Promotion und in allen Lebenslagen. Ihr wart immer für mich da und habt mir stets viel Kraft gegeben. Vielen Dank dafür‼ Hervorzuheben ist die immerwährende fachliche Begeisterung meines Vaters *Henning Hacker* für meine Arbeit. Ich bin überzeugt, dass er den Grundstein für mein naturwissenschaftliches Interesse legte, indem er mir schon als Kind Zungenbrecher wie „Desoxyribonukleinsäure“ beibrachte ;-)

Index

Summary

The integrity of DNA in living organisms is permanently threatened by many endogenous and exogenous factors. Damages can lead to cell death and serious diseases and have to be repaired by the organism. The focus of this thesis is the investigation of the ability of high (*Bst* Pol I, Klenow fragment exo- (KF^{exo-})) and low fidelity polymerases (Pol κ, Pol η) to insert and bypass several oxidatively-derived purine-lesions. 8-Oxopurines (8-oxodA, 8-oxodG) and formamidopyrimidines (FaPydA, FaPydG) are the major oxidative DNA lesions, but also guanine-derived imidazolone (dIz) and its further degradation products oxazolone and guanidinoformimine are produced to a large extent. In order to examine the mutagenic potential of these lesions we first analyzed the ability of the polymerases to insert nucleotide triphosphates opposite the lesions by single nucleotide insertion studies. To further analyze and quantify the lesion bypass abilities of the polymerases we subsequently performed primer extension coupled pyrosequencing with fully elongated primers.

Whereas the adenine-derived lesions are mainly base-paired with the correct dTTP, the guanine-derived lesions exhibit substantial mutagenic properties. Next to correct base pairing with dCTP, misincorporation of dATP opposite carbocyclic FaPydG (*c*FaPydG) and 8-oxodG occurs to a significant extent. In addition, the Iz-derived lesions are base-paired with dATP, dCTP, and dGTP in a disperse manner and misincorporations range from 40-75% with the investigated polymerases. In contrast to the FaPy- and 8-oxo-lesions, which are replicated with good efficiencies, the Iz-derived lesions stall the polymerases to a large extent.

In order to support the biochemical data and to reveal the replication mechanism of a high fidelity polymerase with FaPy-lesions we performed co-crystallization studies with *Bst* Pol I. Due to the thermostability and remaining activity of *Bst* Pol I inside the crystal we could ob-

serve sequential primer extension of *c*FaPydA and *c*FaPydG. We obtained structures showing the error-free replication with dTTP and dCTP, respectively. No structure showing erroneous replication of *c*FaPydA was obtained underscoring the weak mutagenic potential of this lesion. In contrast, we obtained structures showing a *c*FaPydG:dA base pair after one and two rounds of replication. The crystal structures revealed that the mutagenic potential of 8-oxodG requires base pairing in the *syn* conformation with dATP. We expected a similar behavior for *c*FaPydG. Nevertheless, our structures of *c*FaPydG:dA clearly show base pairing in the *anti* conformation.

In comparison to undamaged base pairs all structures with FaPy-lesions in the post-insertion site of the polymerase revealed that the active center is distorted and that the template strands are shifted against each other. Despite this, several catalytic residues and the primer strand are retained in the correct position for further primer extension. Hence, despite the perturbations in the template strands, previously formed base pairs can be elongated.

After the discovery that oxidative modifications of the base methylcytosine (mC) occur naturally within the genome and play a key role during cellular development we sought to address the question of their mutagenic potential. Our studies reveal that none of the mC-derivatives is mutagenic, which supports their key role in epigenetics.

1 Introduction

1.1 DNA - The Molecule of Life

DNA is probably the most important macromolecule of living organisms. All the genetic information is saved in this molecule as chemical information; this is called the genetic code. DNA is the substance, which controls life and which is responsible for the correct translation of the code into proteins and thus for the correct "work cycle" in living organisms.

DNA is a polymer of four nucleobases (adenine, cytosine, guanine, and thymine), which are connected *via* desoxyribose and a phosphate backbone. Pyrimidines and purines are the two main building blocks for the nucleobases. Whereas adenine and guanine are purines, cytosine and thymine are pyrimidine bases.

In 1869, Friedrich Miescher isolated crude DNA for the first time, but it was only 84 years later, in 1953, that the structural composition of DNA was uncovered by James D. Watson and Francis H. C. Crick.[1] They revealed that DNA is constructed as a double helix, formed of two complementary strands. In general, the base pairs are formed between adenine and thymine, and between guanine and cytosine (**Fig. 1**). The stability of the duplex is assured by two main interactions; H-bonds between the opposing and extensive base stacking between the adjacent nucleobases.[2]

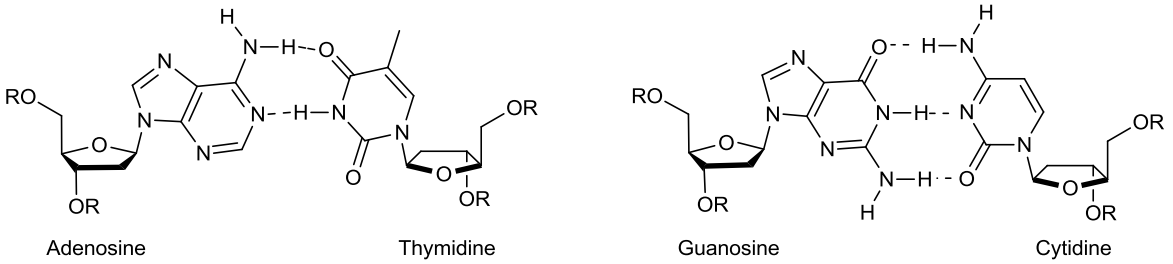


Figure 1: Base pairing of the nucleosides. The dashed lines represent the H-bonds between the nucleobases.

While in bacteria and viruses the genomic information is mostly stored in circular molecules, the eukaryotic system is more complex, which is also in relation to the different size of the DNA. Eukaryotic organisms, in comparison to prokaryotic organisms, have to store much more information. To circumvent the size problem, the DNA is compactly stored in chromosomes; specific constructions of proteins and DNA. Therein, the DNA is wrapped around a spool-like complex of 8 histones, the nucleosome, and these histone octamers built a long fiber with a diameter of 30 nm. Interactions between the nucleosomes result in folding of the chromatin fiber and thus further condensation (e.g. during mitosis) to the so-called chromosome.[3]

The number of chromosomes depends on the organism. In addition, organisms can either have one, two, or even more sets of chromosomes, consequently being called haploid, diploid, or polyploid. Humans, the example for a diploid organism, have two sets of 23 chromosomes each (22 pairs of autosomes and one pair of sex chromosomes), resulting in 46 chromosomes in each somatic cell nucleus.[4]

1.2 Epigenetics

All cells within an organism, in spite of their various functions, possess the same genotype. Within the DNA, specific base sequences stand for a huge variety of genes, but not all genes are incessantly transcribed into proteins. Depending on the cell types, the proteomes can differ enormously. Not all proteins are required at all times, therefore tight and specific control is needed to guarantee proper gene transcription.

Typical mechanisms to regulate gene expression are chemical modifications of bases and histones (e.g. methylation, phosphorylation, and acetylation),[3, 5] intervention of non-coding

RNAs,[6] and DNA-binding proteins (e.g. transcription factors).[7] These modifications can change the DNA-histone interaction strength, and therefore en- or disable transcription. This control is an extremely important process for versatile cellular differentiation and morphogenesis. In general, these modifications are reversible and belong to the huge field of the epigenetic marking system (Greek: *επί-*, *epi-;* over/above).

1.2.1 The Role of DNA-Methylation

DNA-methylation is a biological process, which affects transcription and whose main function is to maintain the genomic integrity. However, the key function of methylation is to silence certain genes, which is achieved by transfer of a methyl-group to the C(5) position of cytosines. Localization of the C(5) methyl group in the major groove of the DNA influences specific interactions with various proteins.[8] All in all, methylation is a crucial process involved in gene expression,[9] embryonic development,[10] differentiation,[11] imprinting,[12] X-chromosome inactivation,[8, 13] neuroplasticity,[14] and carcinogenesis.[15]

Given its importance, 5-methylcytosine (mC) is often referred to as the 5th base of the genome. In mammals, methylation occurs mostly at CpG dinucleotides, which are unevenly distributed within the genome, but to a lesser degree it has also been found at non-CpG sites.[16-17] In human DNA, albeit having a GC-content of 42%, only 1% of the cytosines are followed by a guanine.[18] Approximately 70-80% CpG dinucleotides throughout the whole mammalian genome are methylated.[19] The remaining unmethylated CpG dyads are often found in dense clusters (referred to as CpG islands) with close proximity to promoter regions and are involved in the control of gene expression.[12, 20]

In human peripheral blood mononuclear cells, almost 70% of annotated gene-promoters are linked with CpG islands, but less than 0.2% of non-CpG sites were methylated.[21] In contrast, studies with embryonic stem cells (ESC) show a completely different methylation pattern, where approximately 25% of non-CpG sites are methylated.[16]

Generally, methylation of CpG dyades undergoes widespread changes during cellular differentiation and the regulation of these patterns is of immense importance to ensure correct mammalian development. Whereas the CpG dinucleotides of the gametes are highly methylated,[22] the paternal and maternal methylation patterns are rapidly changed after fertilization.[23] During embryonic development a new wave of methylation takes place.[13] The promoters of several transcription factors (e. g. NANOG and OCT4), which are associated with pluripotency, are unmethylated in the early embryo to enable their gene expression, whereas most tissue-specific genes are methylated.[24] On the contrary, in the later phases of differentiation and in somatic cells, their promoters are mainly methylated and the tissue-specific genes expressed.[10] This is in accordance with the fact that mC-deficient ESC retain pluripotency but lose their differentiation ability.[25]

1.2.2 Methylation by DNMTs

In mammals, methylation is accomplished by three active DNA methyl-transferases (DNMT); DNMT1, DNMT3A, and DNMT3B. The formerly identified DNMT2 does not methylate DNA but RNA and has been renamed to tRNA aspartic acid methyltransferase 1 (TRDMT1), in accordance with its biological function.[26] While new methylation patterns are established by DNMT3A/B/L (*de novo* methyltransferases), DNMT1 preferentially works on hemimethylated strands and is associated with maintenance of existing DNA methylation patterns during rep-

lication. All DNMTs methylate the cytosine at carbon-5 and require *S*-adenosylmethionine as their cofactor, whose S-methyl group is transferred to the DNA base.[27] DNMT1 harbors a zinc-finger CXXC domain (C = cysteine, X = any amino acid) and a conserved region, which play a role in DNA recognition.[28-29] This region has also been found in other proteins, e.g. MBD1 (methyl binding domain protein),[30] TET1 (ten-eleven-translocation protein),[31] and TET3.[32] During S-phase, DNMT1 is localized to replication foci and functions to convert hemimethylated DNA strands into the fully methylated state.[33] In general, it methylates CpGs through an interaction with UHRF1 (ubiquitin-like with PHD and ring finger domains).[34-35]

In 2011 and 2012, Song *et al.* published two crystal structures of truncated mouse DNMT1 in complex with un- and hemimethylated DNA duplexes, thus allowing further insight into the binding mechanism.[36-37] DNMT1 recognizes and binds to un- and hemimethylated DNA with two distinct but supportive domains - the CXXC and the target recognition domain (TRD) - which together ensure accurate replication of methylation patterns. Binding to unmethylated CpGs is proposed to play an autoinhibitory role by preventing *de novo* methylation.[36] When bound to an unmethylated CpG, the entry of the unmethylated cytosine into the active site is prevented by an autoinhibitory loop.[36] In the crystal structure with hemimethylated DNA this loop is shifted and enables therefore the cytosine to enter the active site **(Fig. 2)**.[37]

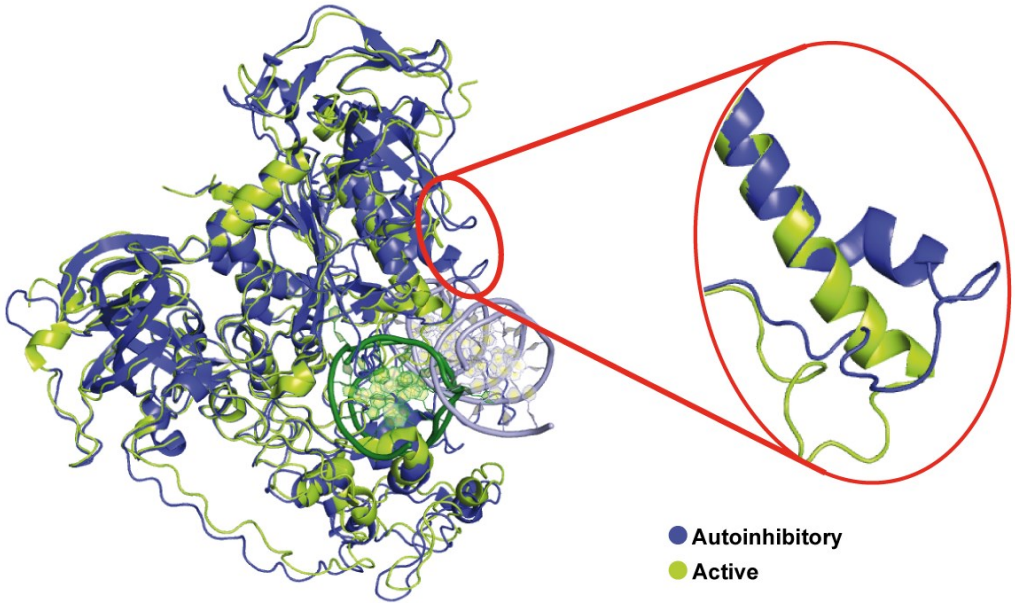


Figure 2: Overlay of the active with the autoinhibitive mDNMT1-DNA complex (PDB codes 4DA4 and 3PT6). The close-up shows the shifted catalytic loop.[36-37]

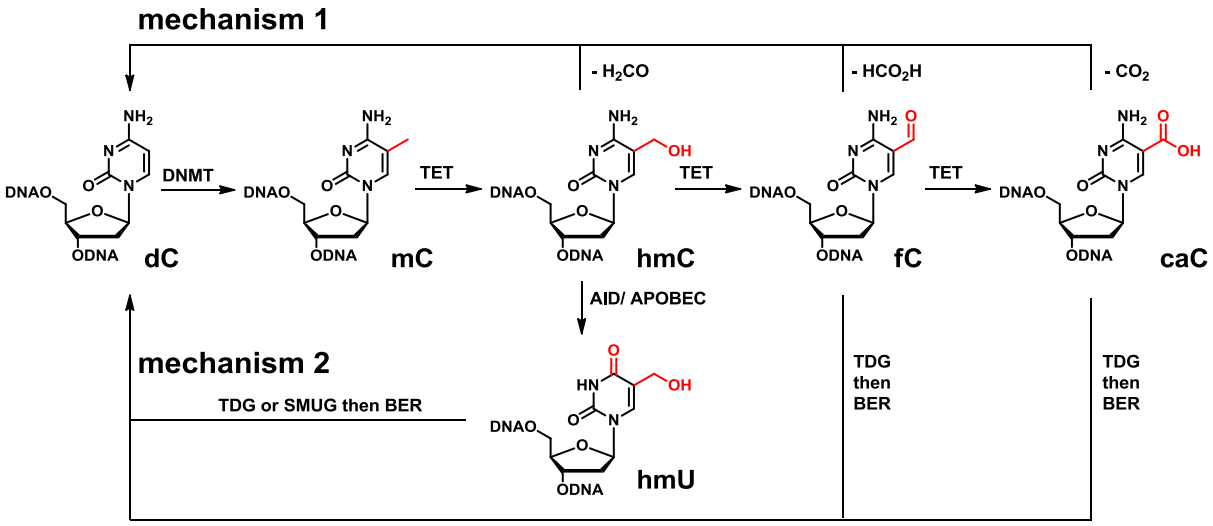
In order to enhance its accessibility for methylation, DNMT1 flips the to be methylated cytosine out of the helix,[37] thereby following a similar mechanism as has been shown previously for M.Hha1 from *Haemophilus haemolyticus*.[38-39]

DNMT3 (DNMT3A, 3B, and 3L) proteins are generally responsible for *de novo* methylation, particularly during gametogenesis and development, all exhibit distinct functions. DNMT3B seems to be essential during early development, whereas DNMT3A plays an important role during late development or post-natal stages.[40] DNMT3L (DNMT3-Like protein), despite its sequence similarities to DNMT3A and B, is catalytically inactive and is required during *de novo* methylation of DNA as a cofactor for DNMT3A and B.[41-44]

1.2.3 TET-Enzymes and Their Role in the Oxidation of mC

hmC, nowadays proposed to be the 6th base of the genome,[45] arises from the radical-based oxidation of the methylgroup of *S*-methylcytosines. The oxidation is carried out by the ten-

eleven-translocation proteins (TET1, TET2, and TET3), a family of 2-oxoglutarate- and Fe(II)-dependent dioxygenases. Their characteristic feature is the successive conversion of mC to hmC, formylcytosine (fC) and carboxylcytosine (caC) *in vivo* and *in vitro* **(Scheme 1)**.[46-48]



Scheme 1: Successive conversion of mC to hmC, fC, and caC by the TET enzymes and potential active demethylation pathways (**see chapter 1.2.4**). While mechanism 1 involves direct elimination of the functional groups, mechanism 2 involves base excision repair-based replacement of fC and caC by dC, or of hmC by dC after prior deamination of hmC to hmU.

The cofactor 2-oxoglutarate is produced by isocitrate dehydrogenases (IDH), which catalyze the oxidative decarboxylation of isocitrate.[31] In 1972, 5-hydroxymethylcytosine (hmC) had already been detected in DNA of various animals,[49] but it was only in 2009, that Rao and coworkers identified TET1 to carry out the conversion of mC to hmC in mouse embryonic stem cells (mESC).[50] At the same time Kriaucionis and Heintz discovered hmC in neurons.[51] Despite the short time since 2009, a huge amount of publications appeared describing the function of these bases. Genome wide mapping studies and even base resolution sequencing are reported.[31, 52-55] In a very simplified view, the hydroxylase activities of the TET enzymes catalyze the reversal of the silencing effect achieved by cytosine methylation.

All TET proteins have a conserved iron-binding subunit, and in the case of TET1 and TET3 an additional CXXC region.[31-32, 50] Nevertheless, *in vivo* studies of the CXXC domain of TET1 revealed no DNA binding activity and no function could be found.[56]

Figure 3: Schematic overview of the family of TET proteins.

The amount of mC and its oxidized products as well as the expression of the TET enzymes depends largely on the kind of tissue that is analyzed and the differentiation state of stem cells.[46, 57-58] While TET1 is mostly expressed in ESC and decreases during differentiation,[57, 59] TET2 and TET3 have been found more ubiquitously.[50, 57] TET1 and TET2 are associated with maintaining pluripotency and lineage specification,[60-61] TET3 seems to be important for differentiation processes and is also highly expressed in brain tissues.[23, 57, 62] The level of hmC decreases within the first days of differentiation but increases in further differentiation.[57]

It is known, that TET1 binds preferentially to CpG-rich promoters, in accordance with the elevated hmC-levels in these regulatory regions.[31, 63-64] This is further supported by TET1 knock-out studies in mESC, which showed a reduction of the hmC-levels by 35%.[62] TET1-depleted mice are viable and fertile but some have a slightly reduced body size. No further defect in maintaining pluripotency is detected.[59] In this scenario, the expected enhanced expression of TET2 or TET3 is not observed, thereby supporting the independent expression.[62]

TET2-deficient mice are also viable and exhibit a normal phenotype, but they seem to be more susceptible to hematopoietic malignancies.[65-67] In combination with *in vitro* data, this points towards a role of TET2 in self-renewal, proliferation, and differentiation of hematopoietic stem cells.[65]

Double knock-down of TET1 and TET2 further reduces the hmC level in comparison to single knock-down,[62] and leads to an elevated level of perinatal lethality.[68] However, these studies all support the hypothesis that TET1 and TET2 might be able to compensate for each other during early developmental stages.[68]

In contrast to TET1 and TET2, the levels of TET3 in ESC have been found to be very low.[57] This protein seems to play an important role in early eye and neural development in *Xenopus*.[32] It is the only TET enzyme, which has been found in oocytes and zygotes and is assumed to play a role in demethylation of the paternal pronucleus.[23, 62] During zygote development, the level of TET3 diminishes at the two-cell stage.[62]

1.2.4 Demethylation Mechanisms of DNA

(De)methylation is an important process, which influences transcriptional activity, therefore regulating gene expression, embryonic development and many more biological functions.[69] Genomic methylation patterns can be erased in a passive and in an active way. During cell division, the methylation patterns are maintained by methyltransferases (e.g. DNMT1), but their inhibition or absence leads to a gradual loss of methylated patterns.[70] This occurs for example in the paternal pronucleus in preimplantation embryos.[71] In addition, it was shown *in vitro* that DNMT1 is acting 50 times more efficiently on mC- than on hmC-hemimethylated DNA strands, supporting the role of hmC during passive demethylation.[72] In contrast, *in vivo*

studies revealed UHRF1-mediated DNMT1-binding to hmC-containing DNA strands and suggested that they play a role in dictating methylation sites on the daughter strand.[34] Further studies will be needed to clarify this issue.

In addition to passive loss of methylation, two pathways of active demethylation have been proposed. One involves BER, the other requires the direct removal of the functional groups **(Scheme 1)**. In both cases the result is an unmethylated cytosine, reversing the silencing effects caused by previous methylation.

All active demethylation pathways start with the oxidation of mC to hmC **(Scheme 1, mechanism 2)**. Subsequently, the TET enzymes can further oxidize hmC to fC and caC. Chemically, the oxidation of hmC to fC is about ten times faster than from fC to caC.[46-47, 73] During BER, fC and caC can be recognized by the thymine-DNA glycosylase (TDG). The resulting abasic site is replaced by an unmethylated cytosine.[48, 74-75] The study of He *et al*. supports this role of TDG by showing that TDG-deficient mESC accumulated caC.[48] Since hmC is not a good substrate for TDG,[48, 75-76] a similar pathway involving hmC-deamination was proposed with the actions of the deaminases AID (activation-induced deaminase) and APOBEC (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like). SMUG1 (single-strand-selective monofunctional uracil-DNA glycosylase 1) is also needed in this pathway. The AID/APOBEC enzymes are thought to deaminate hmC to 5hmU. This is excised by the 5hmU glycosylases, SMUG1, followed by repair *via* the BER.[77-78] In contrast to TDG, hmC is a good substrate for AID.[78]

Another possible active demethylation pathway of the oxidative mC-derivatives would include the direct removal of the functional group (dehydroxymethylation, deformylation, and decarboxylation), resulting in an unmethylated cytosine **(Scheme 1, mechanism 1)**. Chemi-

cally in comparison to dehydroxymethylation, deformylation and decarboxylation are the easier processes.[73] Based on *in vitro* data, a recent study suggests that decarboxylation of caC to C can take place *in vivo*, but so far, no enzyme carrying out these removals has been discovered.[79]

1.2.5 Base-Pairing Properties of mC and its Oxidized Forms

The mutagenic base pairing properties of the mC-derivatives are very low.[80-81] Kamiya *et al.* showed that fC can lead to various mutations in simian kidney cells (COS-7), nevertheless he also found that the mutation frequency is extremely low (0.03%-0.28%).[80] When the oxidized nucleobases fC and caC were discovered,[46-48] fC was believed to have a not negligible mutagenic potential. This was explained by a potential shift from its amino- to the imino-form, due to a putative strong intramolecular H-bond between the exocyclic NH_2(4) group and the carbonyl oxygen at C(5). Thus, base pairing with dATP would be enabled (**Fig. 4**).[80-82] Its potential role as major epigenetic control intermediates could in this case be problematic. A similar behavior could be possible for caC (**Fig. 4**). We assumed that the previously detected mutagenic behavior *in vivo* is not caused by the replicative polymerase but by other cellular processes. Therefore, the mutagenic potential would not lie within the base pairing properties but in a spontaneous deamination reactions caused e.g. by AID or APOBEC. Deamination of cytosine and its derivatives leads to the formation of U, 5mU, 5hmU, 5fU, and 5caU. Without repair, these deamination products could lead to spontaneous base pairing with dA and thus to C→T transversion mutations.[83-84]

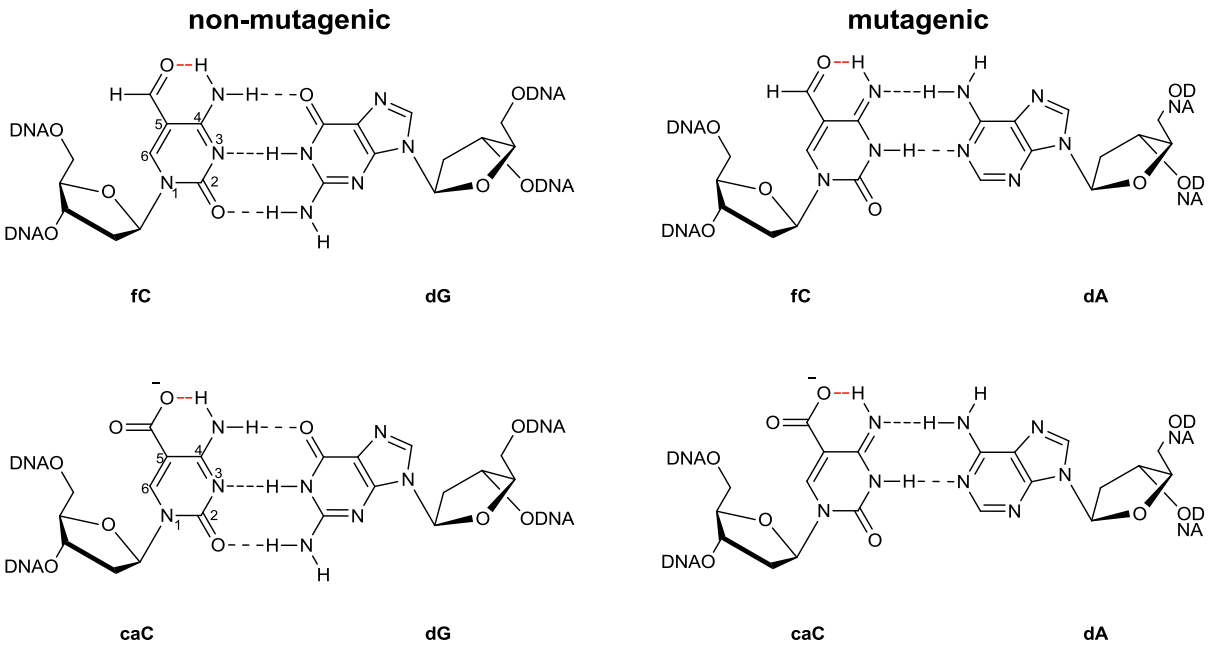


Figure 4: An intramolecular hydrogen bond (red) between the exocyclic NH_2(4) group and the carbonyl oxygen at C(5) of fC and caC was proposed to shift the amino/imino-equilibrium towards the imino tautomeric form. This could in consequence lead to base pairing with dA.

1.2.6 Cancer and Other Diseases

An aberrant methylation pattern in cells often leads to malignant tumors, which are known to show global hypomethylation, genomic instability, promoter hypermethylation, and further aberrant features associated with healthy tissue.[65] In this context, studies with DNMT3-deficient mice revealed various defects and the mice die at different developmental stages.[40] In humans, mutations in DNMT3B lead to hypomethylation of centromeric satellites and to heterochromatin instability, and thus to the ICF (Immunodeficiency, Centromere instability) syndrome. These findings suggest a major function of DNMT3B in maintaining chromosome stability.[40] Mutations within the DNMT3A gene are closely related to myeloid leukemia (AML) and myelodysplastic syndromes.[85] Decreased levels of hmC as an oxidative product of mC are often correlated with various types of cancer.[86] In the case of hepatocellular carcinoma (HCC), one of the most common type of cancer in humans, Chen *et al.* even

suggested hmC as a potential biomarker since its reduced appearance correlates largely with tumor stages.[87] Other studies reveal reduced hmC levels in squamous cell lung cancers, prostate, breast, colon, and brain tumors.[86] In healthy tissues, the highest level of hmC has been found in brain tissues, which are associated with higher cognitive functions.[45] In contrast, the hmC level in brain tumors is strongly reduced.[86, 88] Mutations within the TET enzymes, which are known to catalyze the oxidation of mC to hmC, also correlate with several types of cancers.[89] TET2 mutations are associated with hematopoietic malignancies, e.g. acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), and myeloproliferative neoplasms (MPN).[65] Expression of TET1 is significantly decreased in breast and prostate cancer tissues,[90] and TET1/2/3 are downregulated in melanomas.[91] Also mutations of the isocitrate dehydrogenases and subsequent inhibition of TET-mediated conversion of 5-mC to 5-hmC lead to several types of cancer.[91] In addition, upregulation of AID and APOBEC, which are expressed tissue-specifically, is also implicated in cancer development of the respective tissues, e.g. hepatocytes and B-cells.[92-94] In conclusion, the maintenance of methylation patterns is crucial for correct development in mammals and aberrant methylation due to mutations can lead to cancer.

1.3 DNA Damage

DNA is not inert, it can be damaged by many exogenous (e.g. UV light, X- and γ-rays, chemicals) and endogenous factors (e.g. products of cellular metabolism such as reactive oxygen or nitrogen species (ROS resp. RNS), oxidation, replication errors). These damages could slow down or block transcription, cause mutations, and they could also lead to cell death. If those

damages are not repaired they can cause serious diseases due to their cancerogenic or cytotoxic properties. In addition, they are implicated in aging.[95]

1.3.1 Biological Implications of Reactive Oxygen Species

In living organisms reactive oxygen species (ROS) are typical side products of cellular metabolism (respiration) in the mitochondria. In the case of overproduction or misregulation, this leads to oxidative stress. In general, cells rely on ROS for cellular signaling (growth, proliferation of stem cells, migration of cells, neurogenesis), however, the specific role of individual ROS is largely unknown. It is a complicated system with many stations of control, so if one layer of control is not working properly, unwanted damages to biomolecules can occur.[96]

The main biologically relevant ROS are superoxide $[O_2]^{\bullet -}$, hydrogen peroxide H_2O_2, hypochlorous acid HOCl, singlet oxygen $^1O_2$, lipid peroxides ROOH, ozone O_3 and hydroxyl radicals $[OH]^{\bullet}$.[96] They are generated in various organelles of the cell, particularly in mitochondria, endoplasmatic reticulum, cell membranes, phagosomes, and peroxisomes. They can mediate miscellaneous redox modifications on molecules, e.g. DNA, RNA, lipids and proteins, and regulate diverse physiological processes, e.g. cell migration, circadian rhythm, stem cell proliferation, and subsequent neurogenesis.[96] Many highly reactive oxidants have short half-lives within the cellular milieu. $[OH]^{\bullet}$ for example has a half-life of 10^{-9} s. For this reason they do not diffuse far from their production site. One exception is H_2O_2, which in the cellular milieu has a half-life of up to 1 ms and is therefore able to migrate further to reach its target.

The level of oxidized bases with approx. one per 10^8 normal nucleosides is rather low. While the hydroxyl radical $[OH]^{\bullet}$ reacts with all four nucleobases as well as with the desoxyribose, $^1O_2$ can only react with guanine.[97] $O_2^{\bullet -}$ is not reacting with the bases or sugars at all; its dismutation product H_2O_2 is mainly reacting with them in the presence of metal ions (Fenton

reaction).[97] Often, mainly as a results of $[OH]^{\bullet}$ or of one-electron oxidation, tandem-base lesion are produced. The chapter of oxidatively formed lesions is very complex and many studies have been performed in order to understand the various mechanisms.[96-102]

1.3.2 Oxidation of Purine Bases

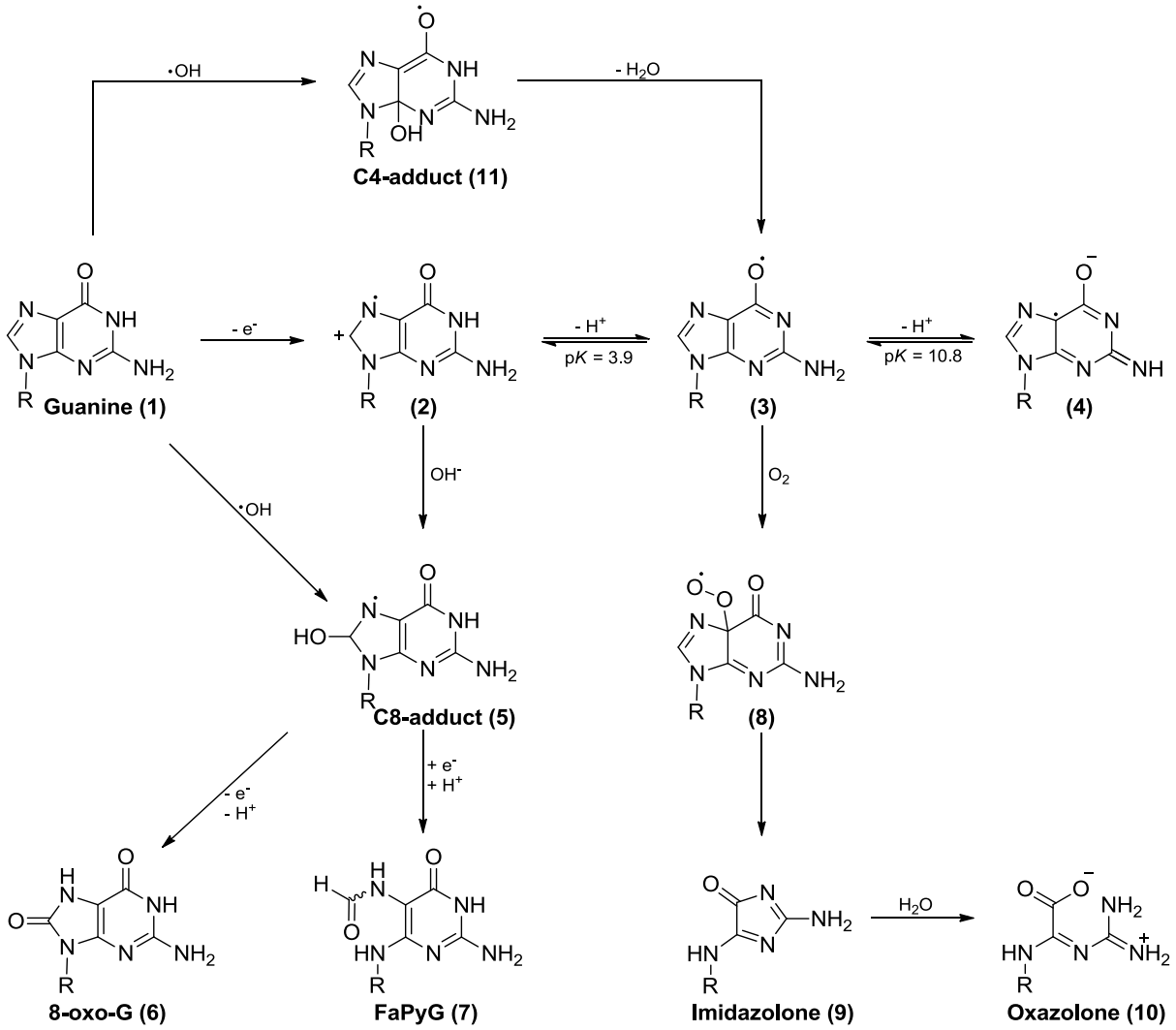
Out of the four canonical DNA bases, guanine has the lowest oxidation potential (guanine 1.29 V < adenine 1.42 V < cytosine 1.6 V < thymine 1.7 V vs. NHE) and can therefore be very easily oxidized.[100] This explains why most oxidized bases arise from guanine or adenine (**Scheme 2 and 3**).[103-104]

When DNA is irradiated in aqueous solution, ten times more FaPydG/8-oxodG is produced compared to the respective A-lesions.[105] Radiolysis experiments under O_2-deficient conditions in chromatin exhibit a higher formation of FaPy-lesions (dA and dG) than 8-oxo-lesions. When the experiment is performed in the presence of oxygen, the yield of the 8-oxo-lesions increases enormously, whereas the yield of the formamidopyrimidines stays roughly the same.[103, 106] In general, irradiation of DNA at 254 nm yields higher levels of FaPy- and 8-oxo-lesions in comparison to typical pyrimidine lesions.[107]

1.3.2.1 Oxidation of Guanine

Next to one- or two-electron oxidations, guanine can also be attacked by hydroxyl radicals or singlet oxygen **(Scheme 2)**. Oxidation through one-electron abstraction or oxidation through hydroxyl radicals leads to the same products. Single electron transfer can be achieved by ion-

izing radiation, type I photosensitizers, or by multiple oxidants, such as nitrosoperoxycarbonate.[108]



Scheme 2: Schematic overview of the oxidation of guanine.

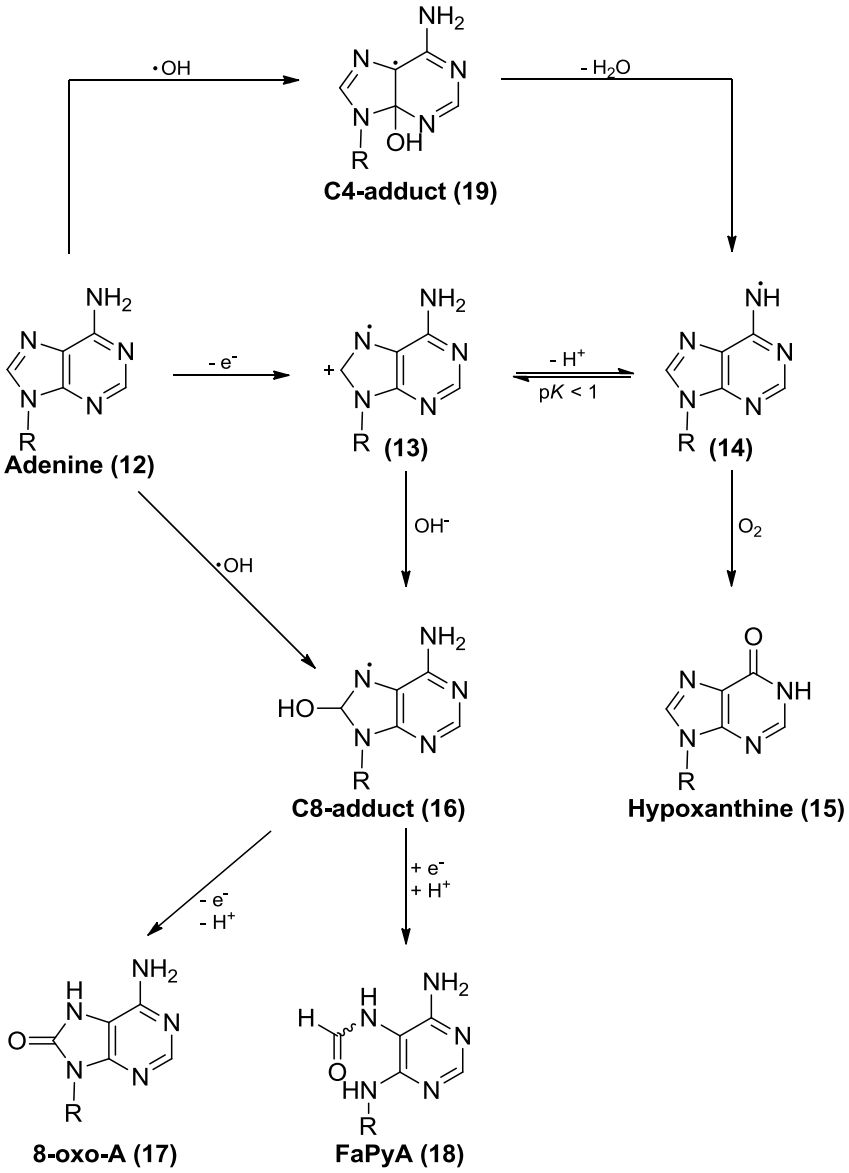
One-electron abstraction from guanine (**1**) generates a radical cation (**2**), which, under protic conditions, is in equilibrium with the guanine radical (**3**). Under basic conditions the latter can be deprotonated to the radical anion (**4**).[98, 109] The addition of a hydroxyl ion to the guanine radical cation (**2**) results in a "redox ambivalent" C8-hydroxy radical (**5**). The formation of the C8-adduct is also possible by a direct attack of a hydroxyl radical to the C8 of the undamaged guanine base (**1**).[109] Depending on the conditions, this adduct can be fur-

ther oxidized or reduced, leading either to 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodG, **6**) or 2,6-diamino-4-oxo-5-formamidopyrimidine (FaPydG, **7**). In the latter case, subsequent hydrolysis including the cleavage of the C8-N9 imidazole bond after the reduction is required. A reaction of oxygen with the guanine radical (**3**) results in the formation of a peroxyl radical (**8**). This can further react to the relatively unstable 2,5-diamino-4*H*-imidazol-4-one (imidazolon; dIz, **9**) and in aqueous solution will be hydrolyzed to 2,2,4-triamino-5-(2*H*)-oxazolone (oxazolone; dZ, **10**), which are both the predominant products of this pathway.[110-111] Also in this case, the guanine radical (**2**) can be generated by the attack of hydroxyl radicals to the C4 of the undamaged guanine (**1**).[109] The reaction of a hydroxyl radical with guanine leads to ≈17% to C8- (**5**) and ≈60-70% to C4-addition (**11**).[109]

With 0.74 V vs. NHE, 8-oxodG harbors an even lower oxidation potential than the undamaged guanine, explaining its high reactivity and fast further degradation to other more stable products.[112] These reactions generally include ring opening and subsequent rearrangements, leading to e.g. imidazolone (**9**), oxazolone (**10**), guanidinohydantoin, spiroiminodihydantoin, or even urea.[113-114]

1.3.2.2 Oxidation of Adenine

As has been shown previously for guanine, one-electron oxidation of adenine (**12**) leads to a radical cation (**13**), which exhibits very strong acidic properties (pKa ≤ 1, **Scheme 3**).[98, 115]



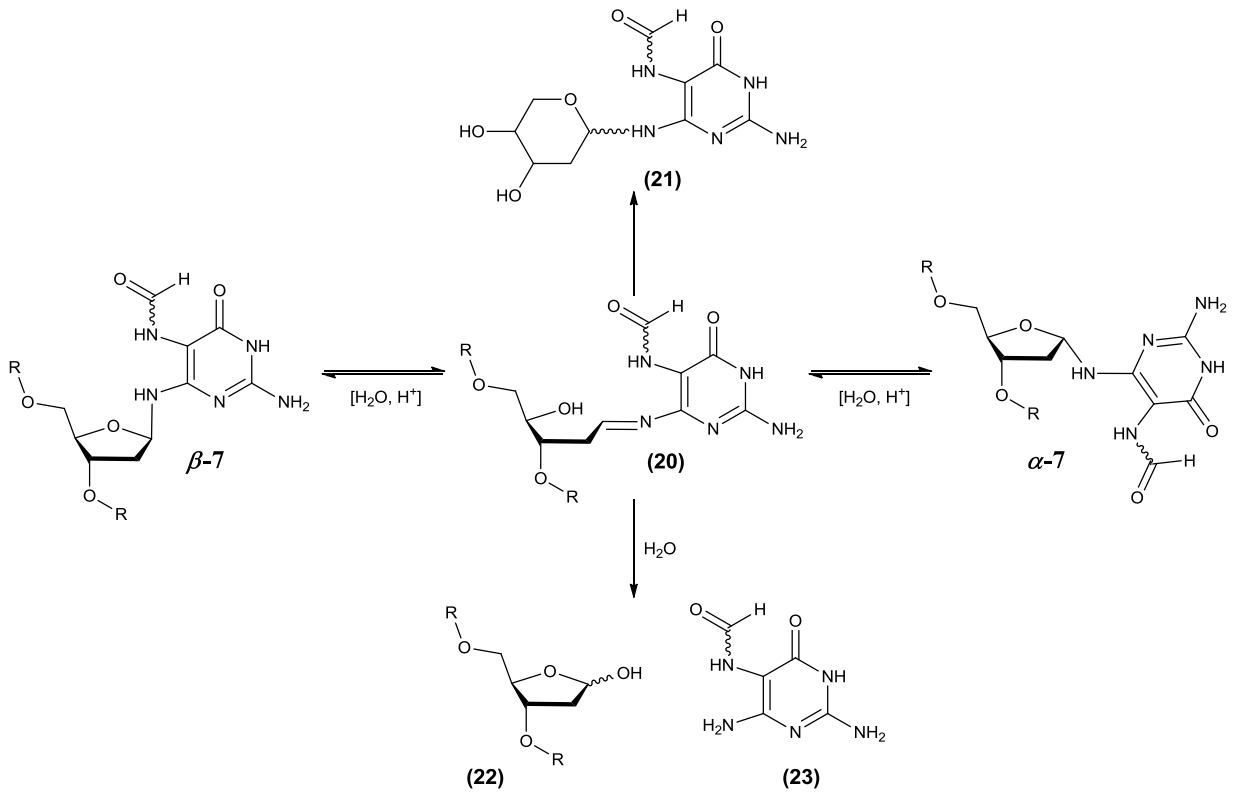
Scheme 3: Schematic overview of the oxidation of adenine.

Thus, the radical cation (**13**) can be directly protonated leading to the adenine radical (**14**). This can be further deaminated, leading to another major degradation product hypoxanthine (**15**). Nevertheless, the radical cation (**13**) can also be hydroxylated resulting in the formation of the C8-adduct (**16**). In the case of one-electron oxidation of this reducing radical 8-Oxo-7,8-dihydro-2'-deoxyadenosine (8-oxodA, **17**) is formed, and in the case of reduction and subsequent hydrolysis 4,6-diamino-5-formamidopyrimidine (FaPydA, **18**) is generated.[116] Again, both A-lesions arise from the same C8 intermediate. The reaction can also be initiated

by the attack of a hydroxyl radical, leading to the generation of the C4- (**19**) and C8-adduct (**16**). Whereas the first is quite unstable and remains at equilibrium with the educt, the formation of the C8-adduct is irreversible.[117] The reaction of a hydroxyl radical with adenine results in ≈37% to C8- and ≈50% to C4-addition.[98, 101] Analogous to 8-oxodG, 8-oxodA exhibits a low oxidation potential (0.92 V vs. NHE).[118] In contrast to 8-oxodG no further oxidation has been observed so far.

1.3.2.3 Epimerization and Deglycosylation of Formamidopyrimidines

A typical feature of formamidopyrimidines is their ability to readily epimerize and rearrange from the furanose (*α,β*-**7**) to the more stable pyranose form (**21**, **Scheme 4)**.[119] This is explained by the fact that the glycosidic bond does involve a nitrogen atom, which is not part of an aromatic, stabilized heterocycle. Furthermore, deglycosylation could occur, but this occurs at much slower rate **(22, 23)**.



Scheme 4: Epimerisation, anomerisation, and deglycosylation of FaPydG.

In deglycosylation studies with nucleosides FaPydG exhibits a half-life of more than 500 hours at 55°C, FaPydA in comparison has a half-life of 20.5 h. Even at 90°C FaPydG is still 16 times more stable than FaPydA with a half-life of around 91 h. Deglycosylation studies of FaPydA show at 37°C a half-life time of 103 h, while for FaPydG those studies were, due to time limitations, not done. Overall, these results reveal a higher stability of FaPydG in comparison to FaPydA.[120] The ratio of the α- and β-anomers changes within 6 hours at 37°C from α:β 1:1 to 1:1.33.[120] The α-anomer of FaPydG is known to block replication, whereas the β-anomer is bypassed during replication and can induce mutations.[121-122] The mutagenic β-anomer is known to be present in double-stranded DNA[123-124] whereas the α-anomer is the dominant species in single-stranded DNA (α:β → 2-3:1).[124] In the DNA-duplex, the lesion exists almost exclusively as the β-anomer. In general, it seems that in duplex DNA β-

anomers are more stable than α-anomers.[124-126] Based on their results of the work with the formamidopyrimidine-type DNA adduct of Aflatoxin B_1, Brown *et al.* conclude, that it is very unlikely that much α-anomer is present in genomic DNA.[124] Under their experimental conditions, the half-life of the β-anomer in double-stranded DNA is 16 h at pH 7.0. Therefore, anomerization during replication is quite unlikely.

The process of epimerization from the furanose to the more stable pyranose form makes synthesis of FaPy-containing oligonucleotides quite difficult. The epimerization cannot occur within the strand, where the 5'-substituent is a phosphate. Synthesis of FaPy-containing DNA oligos always results therefore in a mixture of α- and β-anomers, making it difficult to perform viable biological experiments and to obtain accurate results. Therefore the need for a synthesis that avoids rearrangement emerged, and in addition, stereodefined FaPy-analogs, so either α- or β-analogs, were synthesized. Insertion of the lesion-containing phosphoramidite as a dinucleotide circumvents epimerization during oligonucleotide synthesis.[127] To this end bioisosteric compounds were developed. Two well-known approaches for bioisosteric FaPy-analogs are the C-nucleosides from the Greenberg-group[128-129] and the carbocyclic FaPys from the Carell-group[130] (**see next chapter**).

1.3.2.4 Bioisosteric Analogs of the Formamidopyrimidines

In order to investigate the properties of formamidopyrimidines the need evolved for bioisosteric analogs. In comparison to the natural lesions these compounds should have similar structures and the same hydrogen bonding properties but should be anomerically stable during DNA synthesis and manipulation. This allows the preparation of FaPy-analog-containing oligonucleotides in sufficient quantity and purity. Given that DNA polymerases

choose nucleotides based on their sugar moiety,[131-134] it is important to use analogs with similar structural conformations compared to the desoxyribose. The compounds used in this thesis were found to be good mimics of the formamidopyrimidines and they enabled us to do co-crystallization studies with polymerases or repair proteins. Additionally, they were used as inhibitors of repair enzymes.[119]

1.3.2.4.1 C-nucleoside Analogs

One approach to investigate the biology of formamidopyrimidines in DNA was conducted by the Greenberg-group using the stereodefined C-nucleosides, in which the N(9) is replaced by a C (**Fig. 5**).[128-129] The geometry at the connection between the base and the sugar is therefore changed from trigonal planar (C_{sp^3}-N_{sp^3}-bond) to a tetraedric form (C_{sp^3}- C_{sp^3}-bond).

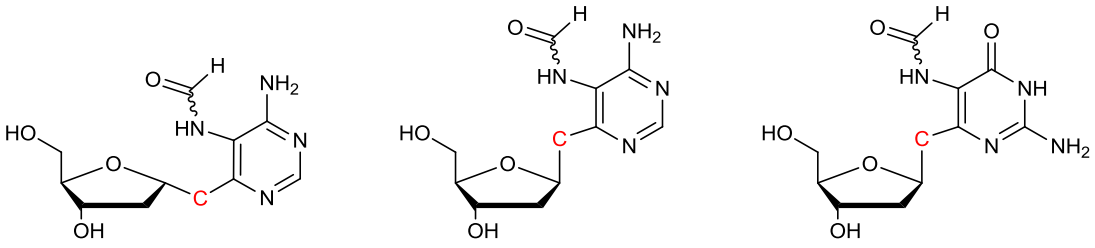
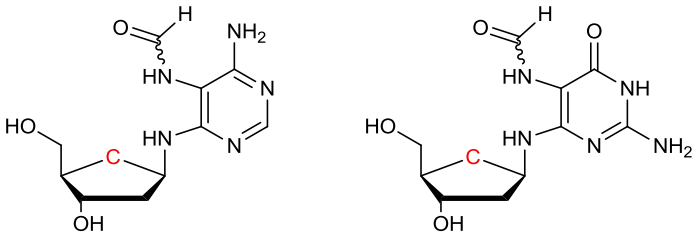


Figure 5: C-nucleoside FaPy-analogs of the Greenberg group.

When comparing thermodynamic properties of DNA duplexes containing either the canonical nucleoside, the natural formamidopyrimidines or the C-nucleoside analogs paired with the right base, the melting temperature of the duplex strand with the native base pair is the highest whereas the one with the C-nucleoside analog is the lowest.[129, 135] No difference between the α- and β-anomer is observed when measuring the melting temperature of DNA duplexes containing C-FaPydA.[135]

1.3.2.4.2 Carbocyclic Analogs

Another possibility to prevent anomerization of the FaPy-lesions that is applied by the Carell group is the use of bioisosteric lesion-analogs where the sugar oxygen is replaced by a methylene group (**Fig. 6**).[130]



β-*c*FaPydA *β*-*c*FaPydG

Figure 6: Carbocyclic FaPy-analogs developed by the Carell group.

The so-called carbocycle prevents anomerization and has no influence on the base pairing itself. To prove that the carbocycle does not change the thermodynamic properties, carbocyclic dG (*c*dG) and natural dG were compared. They were found to exhibit similar melting points when base paired with dA, dC, dG, or dT. In general, the base pairs with the carbocyclic nucleoside are slightly more stable.[136] Compared to the native bases dA or dG, a DNA duplex containing the carbocyclic FaPy-lesion is less stable.[136]

Previous investigations of how the carbocycle influences the *syn*/*anti*-conformation indicated only marginal differences between the 2'-deoxyribose and the carbocycle.[130] In agreement with NOESY experiments, the *anti* conformation is favored.[127,134] For standard nucleotides it was reported that the duplex stability and structure is only marginally affected by the mutation from the oxygen to the methylene group.[137-139]

To investigate of how the carbocycle influences the *syn/anti*-conformation during base pairing, theoretical studies were performed.[136] The torsion potentials of the C'-N bonds of natural and carbocyclic FaPydG were calculated using B3LYP/6-31G* density functional calculation. In both cases the *anti* conformation is about 6 kcal mol^{-1} more stable.[136] Additionally, the results are supported by 2D-torsional screens around the C1'-NH and around the NH-C4 bond of both monomers using the force field SYBYL mmff94s.[140]

1.4 Replication of Oxidative Lesions

In general, oxidative lesions reduce the replication efficiency of high-fidelity polymerases (**for further details see chapter 1.5**) and often serve as templating bases with mutagenic potential.[141-142] It should be noted that the lesions derived from adenine inhibit a much smaller mutagenic potential then the ones derived from guanine.[143] However, the level of mutagenicity depends much on the DNA sequence context,[144] the structure of the lesion itself, and on the structural deviations caused within the DNA strand by the lesion.[145] When a polymerase encounters a damaged DNA base, this lesion can either be blocking, coding or miscoding. The first two possibilities are rather harmless, but a miscoding lesion can lead to mutations and, in the worst case, to diseases.[146] 8-oxodG is one famous example of a miscoding lesion, since a stable base pair with dATP can even be elongated.[147] When a replicative polymerase cannot bypass a certain lesion or is too slow, a process called translesion synthesis (TLS) takes place, helping to bypass the lesion. In general, depending on the type of lesion, TLS requires two main polymerases.[148] One, which inserts one or two bases opposite the lesion, and a second one which elongates the strand past this lesion.[146, 149] When the lesion is bypassed, the original replicative polymerase can restart. Typical inserter polymerases

are Pol η, ι, δ, or γ, and typical extenders are Pol κ or ζ. PCNA (Proliferating Cell Nuclear Antigen) is the central scaffold promoting the access of TLS polymerases to the replication ensemble stalled at a lesion site. It is also called the replication sliding clamp and is largely regulated by ubiquitylation and sumoylation.[150] This mechanism is extremely important for lesions, which could otherwise not be handled by high-fidelity polymerases. Even though the active site of TLS polymerases is not as tightly structured as the one of high-fidelity polymerases, their relative "correct insertion fidelity" is sufficient to decrease the proportion of misinsertions.[146, 151]

1.5 Polymerases

In all living organisms there is the need to amplify the genetic information with high accuracy. To comply with this aim, nature has developed complex processes, relying on a variety of in concert working proteins including several classes of polymerases.[146] In general, all polymerases have a similar overall structure,[36,37] but their specific size and composition can vary enormously. Whereas Pol β (repair function) has a size of 39 kDa and works as a single subunit,[152] DNA Pol III from *E. coli* (replicative function) functions only as a complex, multichain enzyme (holoenzyme), which contains 10 different subunits.[153] Their main functions are replication, recombination, and repair events.[154-155] Based on sequence homologies, polymerases are grouped into seven different families (**Table 1**). The overall structure of a polymerase resembles a human right hand with fingers, thumb and palm. The palm is highly conserved in most polymerases and contains the catalytic center including residues responsible for metal-ion binding. They ensure the cleavage of the incoming dNTP when it is added to the 3'-OH group of the primer strand. This two-metal mechanism is found in every polymer-

ase and in most cases Mg^{2+} is used.[155] In contrast, the finger and thumb regions can differ in the various families.[155] In addition, some polymerases contain a conserved 3'-5' proofreading exonuclease domain.[156]

Table 1: Classification of the prokaryotic and eukaryotic polymerases. Table adapted from [146, 157]

Family	Prokaryotic	Function	Eukaryotic	Function
A	DNA Pol I	replication, repair	DNA Pol γ DNA Pol θ DNA Pol ν	replication in mitochondria and of plastid DNA replication of cross links, BER, hypermutation of immunglobulin genes repair
B	DNA Pol II	repair, replication, translesion synthesis	DNA Pol α DNA Pol δ DNA Pol ε DNA Pol ζ	replication (priming) replication (lagging strand), repair replication (leading strand), repair translesion synthesis (extension)
C	DNA Pol III DNA Pol E	Replication		
X	DNA Pol X	Repair	DNA Pol β DNA Pol λ DNA Pol μ DNA Pol σ TdT	BER BER, double strand break repair, immunglobulin recombinational repair (IRR), translesion synthesis immunglobulin recombinational repair sister chromatid cohesion antigen receptor diversity
Y	DNA Pol IV DNA Pol V	translesion synthesis	DNA Pol η DNA Pol ι DNA Pol κ Rev1L	translesion synthesis (incorporation)
RT			Telomerase	replication of chromosome ends

1.5.1 The Prokaryotic Polymerases of the A-Family

The A-family of polymerases is represented by high-fidelity polymerases and famous members are the DNA polymerases I from *E. coli,*[158] *B. subtilis,*[159] *G. stearothermophilus*[160] and *T. aquaticus*.[161] In general, they have the common architecture of a right hand with thumb, fingers, and palm domain, and in addition, they feature a 3'-5'-exonuclease domain responsible for proofreading. Therefore, they have two active sites, where one is responsible for polymerization and one for exonucleolytic proofreading.[156] However, some do only contain a vestigial, non-functional 3'-5'-exonuclease domain (e.g. *B. subtilis*,[159] *G. stearothermophilus*[160]). High fidelity polymerases play an important role in repair and replication and copy DNA templates in a very fast and accurate manner. Considerable sequence and struc-

tural deviations between the various members are observed within the finger and the thumb domains.[162] One characteristic aspect of a binary complex of Pol I and a dsDNA is that the acceptor template base for the incoming dNTP is not positioned in the insertion site next to the previously formed base pair; it has to enter a pre-insertion site first.[162] This mechanism might explain their ability to discriminate strongly against frameshifts, e.g. KF exhibits an error rate of $\leq 10^{-5}$.[162-163]

1.5.1.1 *E. coli* DNA Pol I and its Large Fragment Klenow

Escherichia coli is a gram-negative bacterium, which is naturally found in the intestine of birds and endothermic mammals. In 1956, Kornberg and coworkers discovered the first known polymerase, which was the DNA Pol I of *E. coli*.[158] Playing an important role in repair (e.g. BER[164] and NER[165]) and in replication, it binds efficiently to nicked and gapped DNA. It also removes the RNA primer of the Okazaki fragments and replaces the gaps of the lagging strand during DNA replication.[146] The optimal growth temperature of *E. coli* is 37°C. Its Pol I belongs to the A-family, and its polypeptide chain consists of 928 amino acids (103 kDa). As all A-family polymerases, it contains an 5'-3'-polymerase activity, 5'-3'-exonuclease activity as well as the 3'-5'-proofreading exonuclease activity. The large fragment of Pol I is called Klenow fragment (KF); it lacks the 5'-3'-exonuclease, but retains the polymerase and proofreading activity. KF is extensively used in molecular studies as a model system for determining polymerase structure and functions. The missing exonuclease makes it suitable for many applications. In 1985, Ollis *et al*. published the crystal structure of KF in complex with dT, which was the first high resolution structure of a high fidelity polymerase.[166] For certain ap-

plications where the proofreading exonuclease is unsuitable, another mutant, called KF^{exo-}, is used instead.

Opposite 8-oxodG, KF^{exo-} inserts efficiently 4-fold more dCTP than dATP.[167] Insertion studies with KF and the FaPy-lesions reveal a persistently lower mutagenic potential compared to the 8-oxo lesions.[141] Opposite FaPydA and FaPydG, the correct nucleotides dTTP and dCTP are inserted more slowly than opposite the native bases.[141-142] The misinsertion of dATP opposite FaPydG by KF^{exo-} is 1-5%, depending largely on the sequence context.[141] Additionally, KF^{exo-} is able to insert one nucleotide opposite imidazolone and oxazolone. After insertion, however, it generally stalls.

1.5.1.2 *G. stearothermophilus* DNA Pol I and its Large Fragment *Bst* Pol I

Geobacillus stearothermophilus is a thermophilic bacterium, which was isolated in a calcium carbonate mount of a hot spring in Yellowstone National Park in Idaho. The strain has an optimal growth temperature of 71°C and its DNA polymerase I is thus a thermostable member of the A-family, which is composed of 876 amino acids (97 kDa). It displays polymerase activity, strand displacement and 5'-3'-exonuclease activity, but it lacks the proofreading 3'-5'-exonuclease activity. We refer here to *Bst* Pol I as the large fragment of the polymerase I, which contains the 592 carboxy-terminal amino acids (67.7 kDa). In size and domain structure it is similar to the Klenow fragment from *E. coli* DNA polymerase I, but it lacks the 3'-5'-exonuclease activity. Nevertheless, it has a higher processivity than KF (*Bst* Pol I: 111 nucleotides/s; KF: 7.7 nucleotides/s).[160]

Sequence alignments of *Bst* Pol I with KF[160, 168] and Klentaq1[160, 169] (analogous large fragment of DNA polymerase I of the thermophilic bacterium *Thermus aquaticus*) revealed little

overall sequence similarities (37.7% within the polymerase domain and 9.8% within the N-terminal exonuclease domain), but the overall topology shows high structural resemblance between these three polymerases. In 1997, Kiefer *et al.* published the first high resolution structure of *Bst* Pol I,[160] therefore making it possible to analyze and compare the structures e.g. during DNA lesion bypass. Studies with various 8-oxodG-analogs with KF^{exo-} and *Bst* Pol I pointed out that KF^{exo-} is less sensitive to alterations in the imidazole ring of dG during incorporation of dCTP.[170]

The polymerase *Bst* Pol I consists of 4 main domains (**Fig. 7**): the vestigial 3'-5'-exonuclease domain, the palm domain, the thumb domain, and the finger domain. The N-terminal domain corresponds to the vestigial 3'-5'-exonuclease domain, which, in the case of *Bst* Pol I, is inactive due to lacking metal binding ability and structural and topological changes in the active-site cleft.[160, 171-172] The sequence identity between *Bst* Pol I and KF in the palm respectively polymerase domain is 71.0%.[160] In general, DNA polymerases have three highly conserved carboxylate residues within the active site, which are responsible for catalysis.[173-174] In *Bst* Pol I, the corresponding amino acids are Asp653, Asp830, and Glu831 (in KF: Asp705, Asp882, and Glu883). When the active site is formed, the thumb domain plays an important role in stabilizing the DNA. In contrast to KF[171] and Klentaq1,[175] the joint between two helices (I and I_1) is shorter in *Bst* Pol I. Conceivably this region is less flexible, therefore contributing to a more stable DNA-protein association, which could explain the higher processivity. The comparison of the finger domain reveals only marginal differences between the two polymerase fragments. In this region, the O-helices are located, which are essential for dNTP and template binding.[160, 174] Herein, two further highly conserved residues are located, Phe710 and Tyr714 (in KF: Phe762 and Tyr766).[160]

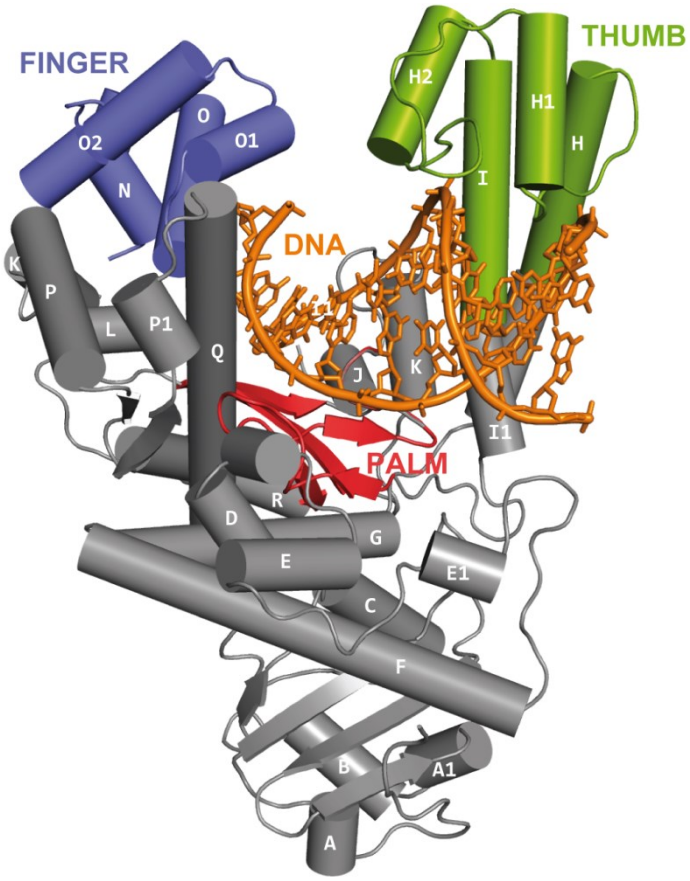
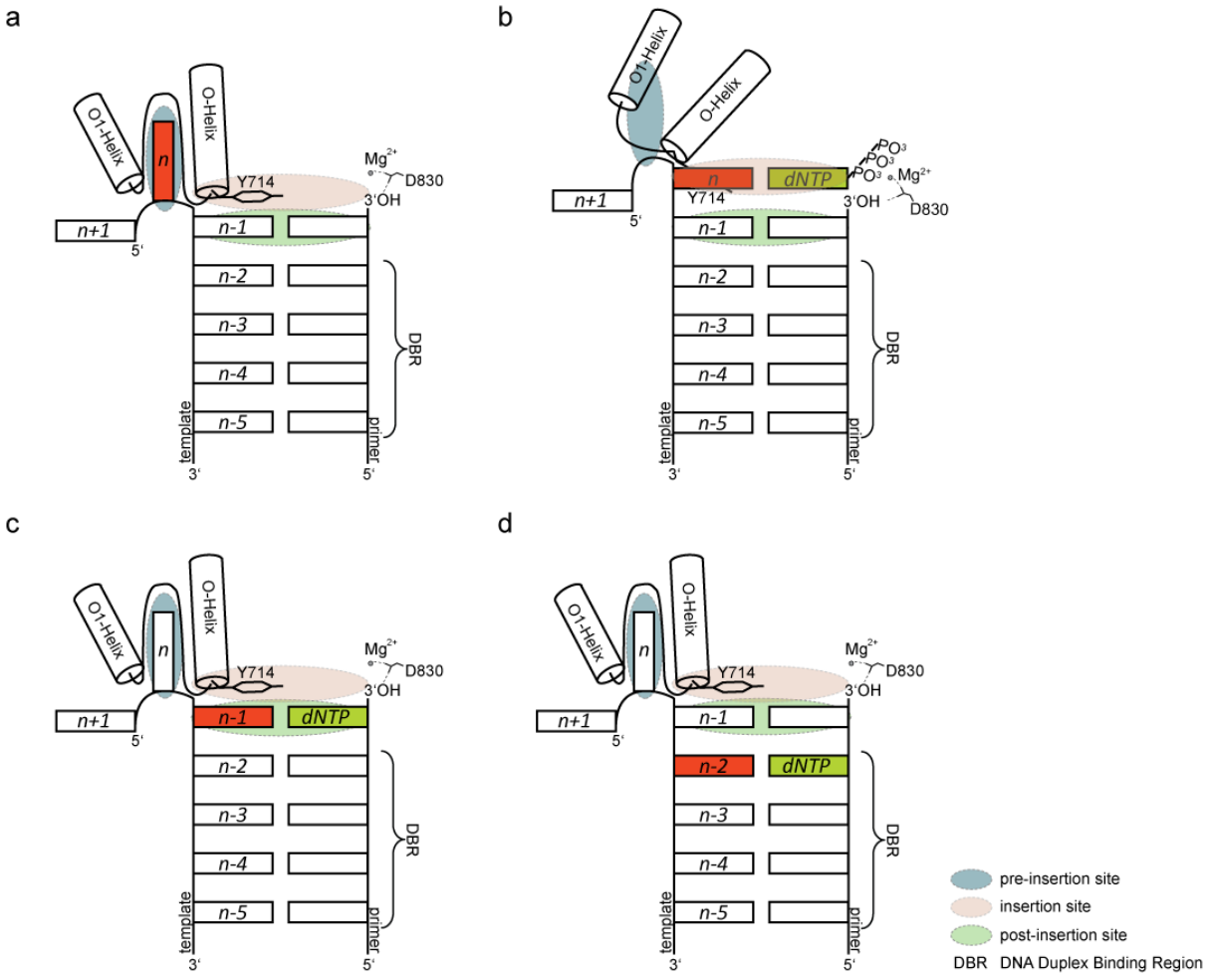


Figure 7: Crystal structure showing the topology of the large fragment of *Bst* Pol I in complex with DNA.

The close resemblance of *Bst* Pol I to KF and Klentaq1 makes it a good model system for other polymerases of the A-family. One advantage of *Bst* Pol I is its remaining activity in the crystal. Therefore, soaking crystals with additional nucleotides can result in further elongation of the DNA template strand.[162, 176] This makes it possible to crystallographically capture many stages of the replication cycle.[147, 162, 176-183] In general, the accuracy of the polymerase depends largely on the metal ions used in the reaction buffer, e.g. Mg^{2+} ensures a higher reliability then Mn^{2+}.[178] Co-crystallization studies with undamaged DNA duplex strands were reported before this study.[162, 176, 178] Additionally, lesion- (e.g. *c*FaPydG,[147] O6-methylguanine,[180] or bulky adducts[177, 179]) or mismatch-containing DNA duplexes were also co-crystallized with *Bst* Pol I.[178]

1.5.1.2.1 Replication of a DNA Lesion within *Bst* Pol I

When binding to a lesion containing DNA duplex strand, the polymerase is in an open conformation and the unpaired lesion is positioned between the helices O and O1, which corresponds to the pre-insertion site (pre-IS). In this state, the highly conserved amino acid Tyr714 (located in helix O) is stacking with the template base of the base pair at the post-insertion site (post-IS) n-1 and it is therefore blocking the insertion site (IS).[160, 176] Asp830 forms a hydrogen bond with the 3'-OH group of the primer strand (**Scheme 5a**).[160]



Scheme 5: Replication of a DNA-lesion (n; red) by *Bst* Pol I.

When a dNTP is bound, the polymerase is closing by moving its thumb towards the fingers, thereby defining the active center. During this movement the template base is shifted towards the newly formed insertion site and the triphosphate group of the nucleotide is coordinated by two magnesium ions (**Scheme 5b**).[162] Therefore, a nucleophilic attack is enabled and the base pair can be formed.[162] In this state, the Tyr714, formerly blocking the insertion site, is displaced, but now the pre-insertion site is blocked by the loop connecting the helices O and O1.[162] Subsequent release of pyrophosphate during covalent binding of the dNTP to (the 3'-OH terminus) of the primer strand destabilizes the closed conformation of the polymerase leading to translocation of the newly formed base pair to the post-insertion site (post-IS, n-1) (**Scheme 5c**).[131, 184] This translocation mechanism, whose driving force is the cleavage of the triphosphate, is an important step for frameshift prevention.[162] At the post-insertion site, Tyr714 stabilizes the templating base by π-stacking. Additionally, the highly conserved amino acids Arg615 and Gln797 form hydrogen bonds to the correct base pair and stabilize its position.[176]

After another elongation step the newly formed base pair is shifted to the n-2 position in the DNA duplex binding region (DBR) (**Scheme 5d**). This region binds 4 base pairs with the register positions n-2 to n-5.[162] An interesting observation made by Kiefer *et al.* was that the DNA, when being shifted to the duplex binding region, is underwinding significantly, resulting in a structural change from B- to A-conformation.[176] The DNA, which is not located inside the polymerase remains in the B-form conformation. The part of the DNA, which is bound in the polymerase changes back to the B-form when being shifted away from position n-5.[176]

Mispairs or lesions can lead to distortions within the DNA strand and subsequently in the polymerase active site,[185] possibly resulting in reduced processivity or dissociation of the polymerase.[178, 180, 183] There are many stages of proof within the replication cycle to ensure that the right nucleotide is paired to the templating base. One checkpoint is the correct formation of the insertion site. Only in the case of a suitable dNTP the insertion site can be fully constructed, which ensures proper base pairing of the nucleotide with the templating nucleobase, because only in this case the interactions between polymerase, nucleotide and template base are strong enough. To establish Watson-Crick base pairing, the conformation of the base pair in the insertion site is strictly defined.[186-187] Another checkpoint is the amino acid Asp830. When the polymerase is in the open conformation, Asp830 is essential for the activity of the enzyme and coordinates the 3'-OH group of the primer end. In most cases the geometry of a mismatch differs in comparison to a correctly formed base pair. If the primer end is not positioned correctly within the active site, no hydrogen bonds with Asp830 can be formed and no further nucleotide can be incorporated.[176, 178] A further amino acid to guarantee proper base pairing is Tyr714. Site-directed mutagenesis studies of this residue in the Klenow fragment revealed tremendous effects on the fidelity of replication.[188] One feature of oxidative lesions is their ability to enhance their replication *via* altered H-bond-formation.[147]

1.5.2 The Eukaryotic Y- Family Polymerases

The Y-family is represented by a group of repair polymerases, which play an important role in translesion synthesis (TLS) of DNA lesions. The members of this family (in eukaryotes: η, κ, ι, Rev1, **Table 1**) are known to be error prone or of low fidelity. This is mainly due to struc-

tural differences within the active site, which is more spacious than the one of high fidelity polymerases. Therefore, less contact between the polymerase, the template DNA and the nucleotide is established, allowing lesion bypass of e.g. more bulky lesions, which are generally excluded from the tighter active sites of other polymerases. Their thumb and fingers are smaller than in other polymerases and they lack the α-O helix, which in high fidelity polymerases is responsible for selectivity of nucleotides.[146, 149] They do not contain a proofreading 3'-5'-exonuclease domain, but compared to other polymerase families, they have a C-terminal polymerase-associated domain (PAD) or little finger domain (LF), which interacts with the DNA duplex.[149] In comparison to the replication efficiency of high fidelity polymerases (10^{-5}-10^{-6}), their error-rate is much higher (10^{-1}-10^{-4}) and their processivity much lower.[146, 148, 189] Each of the Y-family polymerases has a unique DNA damage bypass ability und fidelity.

1.5.2.1 Polymerase η

Polymerase η is the best analyzed polymerase of the Y-family and is well-known for its ability to replicate quite accurately past UV lesions. In humans, mutations or deletions in the gene for Pol η lead to Xeroderma pigmentosum (XP-V) resulting in extreme light sensitivity and a highly enhanced predisposition for skin cancer due to UV-induced lesions.[190] Opposite the major UV lesion cyclobutane pyrimidine/thymine dimers (CPD), which induce severe distortions within the DNA, Pol η correctly inserts two adenines.[191-192] It is also able to replicate past other types of lesions, albeit with less efficiency than with CPDs.[148] Crystal structures with cisplatin adducts (normally formed during chemotherapy with cisplatin) and yeast Pol η revealed correct bypass of these intra-strand crosslinks.[193-195] The enzyme is also able to

replicate past some bulky adducts, e.g. acetylated aromatic amine lesions of guanine (AAF-dG) or N^2-AAF-dG, where it mainly incorporates the correct dCTP, although with reduced efficiency.[196-198] Yeast Pol η can also bypass 8-oxodG lesions, correctly inserting dCTP and extending the newly formed base pair.[199-200] Crystal structures were obtained of a ternary complex with yeast Pol η, template-primer DNA with 8-oxodG and dCTP. They reveal an active site that can accommodate this lesion with marginal structural changes.[200] Next to the correct insertion of dC, human Pol η incorporates a low amount of dA as well.[199]

1.5.2.2 Polymerase κ

In contrast to the other Y-family polymerases, Pol κ is the only one having homologues in prokaryotes and archaea, including DinB (Pol IV) in *Escherichia coli* and Dbh and Dpo4 in *Sulfolobus solfataricus*.[201] In addition, Pol κ (as well as Rev1) contains an N-terminal subdomain (N-clasp), which is indispensable for catalytic activity and which allows the polymerase to completely encircle the DNA.[202-203] In contrast to Pol η, its active site is more constrained and smaller, and smaller is also the finger domain.[203] This might explain its inability to insert nucleotides opposite many types of lesions (e.g. CDP, (6-4)TT photoproduct),[204-205] but it may also explain its enhanced fidelity and higher processivity on undamaged DNA in comparison to the other Y-family polymerases.[206] Albeit it cannot insert nucleotides opposite certain lesions, Pol κ can extend lesion base pairs previously formed by another polymerase.[148, 201, 207] It is unable to insert nucleotides opposite the 3'-T of a cis-syn CPD, but it can extend very efficiently from a dG:CPD base pair.[204] It is not capable of extending a (6-4)TT photoproduct.[204-205] Opposite 8-oxodG, Pol κ inserts more efficiently dA than dC.[207] Carpio *et al.* showed, that Pol κ forms mostly a Hoogsteen base pair of 8-oxodG(*syn*):dA(*anti*). In the

structure it can be seen that the active site is well-adapted to accommodate 8-oxodG in the *syn* conformation and that the template strand is not affected.[208-209]

1.6 Base Excision Repair

Base excision repair (BER) is an essential repair process, which plays an important role in the removal of damaged DNA. In summary, repair initiation starts with recognition of (mainly) non-helix-distorting lesions by specialized excision proteins called glycosylases. They are responsible for cleavage of the glycosidic bond between the sugar and the lesion base, leaving an abasic site (also referred to as AP site (apurinic/apyrimidinic site)). Subsequently, this AP site is cleaved by an endonuclease and a polymerase refills the gap (**Fig. 8**).

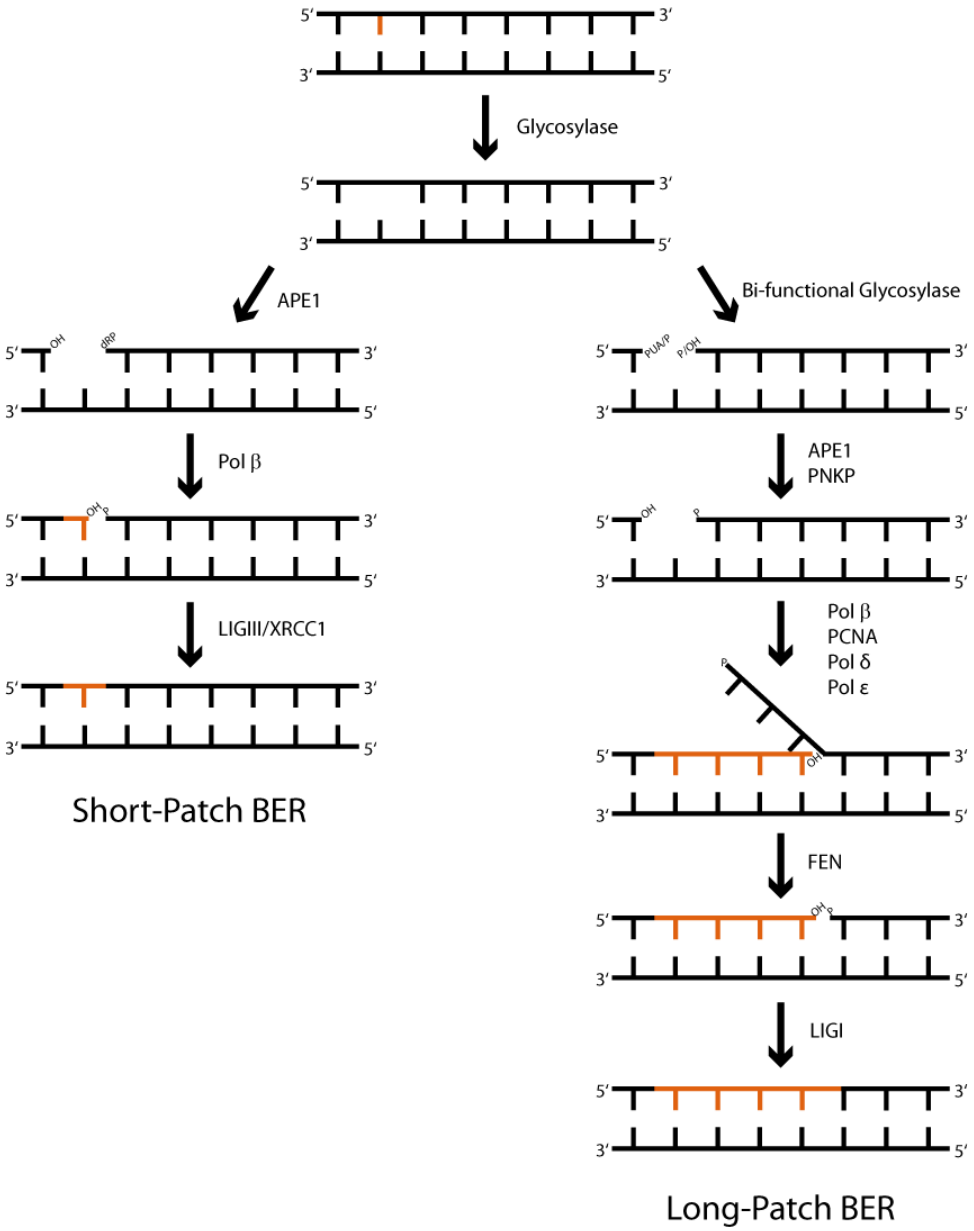


Figure 8: Mechanism of the human base excision repair.

Glycosylases are ubiquitous in all organisms. Each of them is specialized for the repair of certain DNA damages. Their main function is to recognize and subsequently excise the lesions. Some glycosylases possess only glycosylase activity (e.g. TDG, SMUG1, UNG), others additionally own a lyase function (e.g. FPG, OGG1, NTH1, NEI1-3), which is to cleave the preliminary formed AP site.[119] Hence, depending on their functions, they are either classified as mono- or bifunctional glycosylases.[210]

The lesion removal is initiated by flipping the base out of the DNA helix, consequently cleaving the glycosidic bond between the sugar and the base.[211] In the case of monofunctional glycosylases, subsequent incision by the apurinic endonuclease (APE1) generates a repair intermediate that contains a single-strand break with 3'-hydroxyl and 5'-deoxyribose-5'-phosphate (5'dRp) termini. The bifunctional glycosylases, after base removal, incise the DNA backbone 3' to the abasic site by β-elimination (e.g. OGG1, NTH1) or by β/δ-elimination (e.g. FPG, NEIL1/2),[212] thereby creating non-conventional 5'- or 3'-termini (e.g. 5'dRP, 5'-OH, 3'-PO_4, 3'-PG, or 3'-PUA). These blocking termini have to be modified before further polymerase actions are possible, which is achieved by APE1 or PNKP (polynucleotide kinase/ phosphatase).

Once a lesion is recognized and excised, two possible pathways of repair can follow. Depending on the size of the repair patch, either short patch (in the case of 1 nucleotide)[213] or long patch (2-12 nucleotides)[214-215] BER is initiated. In short-patch BER, Pol β, which possesses polymerase and 5'dRP lyase activity, inserts one nucleotide and excises the 5'dRP by creating the 5'-phosphate group, necessary for the final nick ligation by a complex of LIGIII (DNA ligase III) and XRCC1 (X-ray repair cross complementing 1). In long-patch BER, PCNA (proliferating cell nuclear antigen; sliding clamp for pols), RF-C (replication factor C, which loads PCNA on DNA) and DNA Pols β, δ and/or Pol ε are required for strand displacement synthesis, to excise a long DNA patch 5' of the lesion, which is finally removed by FEN1 (flap endonuclease 1). The last step of ligation is fulfilled by DNA ligase I (LIGI). In 2009, van Loon *et al.* analyzed the repair of 8-oxodG *via* immunofluorescence experiments in ROS-exposed cells and proposed a possible repair pathway of 8-oxodG.[216] Two major differences in the repair of this lesion have been observed, depending on the base opposite 8-oxodG and on the tissue. In

non-replicating tissues, an 8-oxodG:dC base pair is mainly recognized by OGG1 and removed by the short-patch BER. In replicating tissues, where 8-oxodG is often base paired with dA, MUTY firstly removes the opposing adenine via long-patch BER and subsequently short-patch BER will remove the 8-oxodG.[151]

The first glycosylase being found to recognize and efficiently excise formamidopyrimidine lesions was a glycosylase of *E. coli*, named after its function formamidopyrimidine glycosylase (FPG).[217-218] Being a bifunctional glycosylase it harbors an additional AP endonuclease activity, hence it can also act as a lyase.[219] In contrast to FaPydG and 8-oxodG, recognition and excision of FaPydA by FPG is independent from the nucleotide base paired with the lesions.[220] In general, FaPydG is excised more efficiently than 8-oxodG when base paired to dC, but both are less efficiently excised when opposite dA.[217, 221-222] A crystal structure of a binary complex of *c*FaPydG:dC-containing DNA duplex strand in complex with *Lactococcus lactis* FPG reveals a different recognition mechanism compared to 8-oxodG:dC. While FPG binds 8-oxodG in the *syn* conformation, FaPydG is bound in the *anti* conformation.[211, 223] Oxazolone is also excised by FPG, yet again with discriminating efficiency when base-paired with dA.[224] A similar behavior with respect to lesion recognition is shown by hOGG1 (human 8-oxodG-glycosylase 1), the functional human homolog of FPG. In contrast, when investigated with the A-lesions, hOGG1 discriminates strongly against 8-oxodA and FaPydA.[212, 225] 8-oxodA is only recognized and excised by OGG1 when base paired with dCTP.[226] While FPG and OGG1 excise the lesions from the DNA strand, another glycosylase, MUTY, is able to extrude the incorrect base opposite the lesion from the DNA duplex and to hydrolyze its glycosidic bond. While FPG is less efficient in excising e.g. FaPydG, 8-oxodG, or oxazolone when base paired with dA, MUTY is capable of excising dA, base-paired to the lesion.[221, 227] In hu-

mans, mainly six bifunctional glycosylases are associated with the repair of oxidative lesions: OGG1, MYH (MUTY homolog), NTH1 (endonuclease III homolog), and NEIL1-3 (NEI-like endonuclease VIII).[210] NTH1 and NEIL1-3 are known to excise pyrimidines and pyrimidine-derived lesions (FaPydG,[228-229] FaPydA,[225, 228-229] and oxazolone[224]) therefore preferentially acting on lesions base paired with purines.[228] Their ability to excise formamidopyrimidine but not 8-oxopurine lesions is assumed to depend on the recognition of the lesion structure, FaPys being recognized as modified pyrimidines.[228] Nevertheless, NEIL glycosylases excise FaPydA more efficiently than does NTH1.[228-229] All in all, a huge machinery of different enzymes acts in concert to remove DNA damage in order to sustain the genetic stability of organisms. Many biochemical and structural investigations reveal insights into the mechanisms of how the organism deals with oxidative lesion. However, further investigations have to be done.

1.7 Biological Implications of the FaPy- and 8-oxo-Lesions

In general, the oxidative lesions reduce the replication efficiency and can act as templing bases with mutagenic potential.[141-142] The lesions derived from adenine exhibit a much smaller mutagenic potential then those derived from guanine.[143] The mutagenicity depends also largely on the sequence context.[144]

The Klenow Fragment from *E. coli* (KF) lacking the 3'-5'-exconuclease is known to induce A→T transversion mutations by misinserting dATP opposite FaPydA,[142] but in general, this lesion has only a small effect on the Klenow fidelity.[125] *In vivo* studies in simian kidney cells (COS-7) reveal in less than 1% A→C transversions by 8-oxodA.[144, 230] In general, the mutagenic potential of 8-oxodA and its contribution to cellular damage is, in contrast to 8-oxodG

(6%-29%), assumed to be very low.[143, 231] Thereby, the vector survival rate of 8-oxodA (88-90%) is considerably higher compared to 8-oxodG (64-78%).[143]

The G-lesions mainly induce G→T transversion mutations and G→A transition mutations.[144, 232] Primer extension studies with stabilized analogs,[136, 233] natural or methylated versions of FaPydG[141, 234-236] show either correct base pairing with dC or misincorporation of dA.[141] G→T transversions are observed when FaPydG (<1.9%) or 8-oxodG (3.1% - 9.8%) are bypassed in wild-type *E. coli*[144] and those results are consistent with kinetic studies using KF and KF^{exo-}.[141, 167, 232, 237] Bypass of FaPydG in COS-7 cells exhibits slightly more G→T transversions than in 8-oxodG (8.2% vs. 6.1% in TXA sequence, and 29.6% vs. 23.5% in a TXT sequence).[144] In general, bypass of FaPydG is slower than bypass of 8-oxodG.[232] The bypass efficiency of FaPydG is comparable to those of oxazolone and imidazolone,[238-239] but in contrast, FaPydG and 8-oxodG have a lower mutation frequency.[232] It was further observed, that the 3'-5'-exonuclease activity has only little effect on FaPydG replication.[232] Kinetic primer extension studies with *S. cerevisiae* Pol η revealed that carbocyclic FaPydG is replicated in an error-free fashion.[136]

1.8 Biological Implications of Imidazolone and Oxazolone

Imidazolone can arise from oxidation of guanine and further from facile oxidation of 8-oxodG **(Scheme 2)**.[114, 240-241] Subsequently, the more stable oxazolone arises from hydrolyzation of imidazolone, which in aqueous solution at 20°C has a half life time of 18-24 h.[111] Both belong, next to 8-oxodG and FaPydG, to the major oxidation products of guanine.[110-111] In primer extension studies with imidazolone-containing DNA, KF, and KF^{exo-} only dGTP

was incorporated but bypass was inefficient.[240] Also *in vivo* studies showed mainly incorporation of dGTP opposite imidazolone, leading to G→C transversions mutations.[239]

Oxazolone is bypassed quite efficiently and exhibits a similar *in vivo* mutation frequency as imidazolone (oxazolone 86%,[238] imidazolone 91%[239]). However, dATP is the mainly inserted nucleotide opposite this lesion *in vivo* and *in vitro*, causing G⟶T transversion mutations. A small amount (≈5-10%) of dGTP is inserted as well.[224, 238] In primer extension studies with Pol β oxazolone represents a blocking lesion.[224] Altogether, the authors estimate its mutagenic potential to be higher than that of 8-oxodG (mutation frequency of 8-oxodG 7% and of oxazolone 95%), even though the replication efficiency of oxazolone is lower compared to 8-oxodG.[238]

1.9 Main Objectives of the Thesis

The main aim of the project was the investigation of how low and high fidelity polymerases deal with certain purine-derived DNA lesions, such as formamidopyrimidines, imidazolone, and its derivatives. The ability of polymerases to insert nucleotide triphosphates opposite the lesions at defined sites in the DNA strand and subsequent bypass of the lesion-containing base pair was thereby of major interest. In addition, quantitative analysis of the (mis)insertions should be performed. In order to support the biochemical data of the incorporation opposite the FaPy-lesions and to gain further insight into the insertion respectively bypass mechanisms co-crystallization experiments should be carried out with lesion-containing DNA and with the large fragment of the thermostable and high fidelity model polymerase I of *Geobacillus stearothermophilus* (*Bst* Pol I). A special feature of *Bst* Pol I is its remaining activity in the crystal, making it possible to directly monitor sequential elongation steps inside the crystal. In this context, investigation of the position of the formamide-group and its role during base pairing was of special interest.

Furthermore, given that an oxygen-containing group at the C(5) position of cytosine could in principle influence the base pairing properties, mC and its oxidative products, which are nowadays extensively found in genomes, should be investigated regarding their mutagenic potential. For this purpose, primer extension studies and subsequent pyrosequencing should be performed.

2 Abstracts of the Publications

2.1 Gehrke, T.H.,[1] Lischke, U.,[1] Gasteiger, K. L.,[1] Schneider, S.,[1] Arnold, S., Müller, H. C., Stephenson, D. S., Zipse, H., Carell, T. "Unexpected non-Hoogsteen-based mutagenicity mechanism of FaPy-DNA lesions." *Nat. Chem. Bio.*, 2013, 9, 455-461.

ABSTRACT: 8-Oxopurines (8-oxodG/8-oxodA) and formamidopyrimidines (FaPydG/FaPydA) are major oxidative DNA lesions involved in cancer development and aging. Their mutagenicity is believed to result from a conformational shift of the N9–C1'-glycosidic bonds from *anti* to *syn*, which allows the lesions to form noncanonical Hoogsteen-type base pairs with incoming triphosphates during DNA replication. Here we present biochemical data and the first crystal structures of carbocyclic FaPydA and FaPydG containing DNA in complex with a high-fidelity polymerase. Crystallographic snapshots show that the cFaPy lesions keep the *anti*-geometry of the glycosidic bond during error-free and error-prone replication. The observed dG•dC → dT•dA transversion mutations are the result of base shifting and tautomerization.

2.2 Stathis, D.,[1] Lischke, U., Koch, S.C. Deiml, C. A., Carell, T. "Discovery and mutagenicity of a guanidinoformimine lesion as a new intermediate of the oxidative deoxyguanosine degradation pathway." ***J. Am. Chem. Soc.*** **2012, 134, 4925-4930.**

ABSTRACT: Oxidative degradation of DNA is a major mutagenic process. Reactive oxygen species (ROS) produced in the course of oxidative phosphorylation or by exogenous factors are known to attack preferentially deoxyguanosine. The latter decomposes to give mutagenic lesions, which under physiological conditions are efficiently repaired by specialized maintenance systems in the cell. Although many intermediates of the degradation pathway are today well-known, we report in this study the discovery of a new intermediate with an interesting guanidinoformimine structure. The structure elucidation of the new lesion was possible by using HPLC − MS techniques and organic synthesis. Finally we report the mutagenic potential of the new lesion in comparison to the known lesions imidazolone and oxazolone using primer extension and pyrosequencing experiments.

2.3 Münzel, M.,[1] Lischke, U., Stathis, D. Pfaffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K., Carell, T. "Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine." *Chemistry* 2011, 17, 13782-13788.

ABSTRACT: 5-Formylcytosine (fC or $^{5\text{-CHO}}$dC) and 5-carboxylcytosine (caC or $^{5\text{-COOH}}$dC) have recently been identified as constituents of mammalian DNA. The nucleosides are formed from 5-methylcytosine (mC or $^{5\text{-Me}}$dC) via 5-hydroxymethylcytosine (hmC or $^{5\text{-HOMe}}$dC) and are possible intermediates of an active DNA demethylation process. Here we show efficient syntheses of phosphoramidites, which enable the synthesis of DNA strands containing these cytosine modifications based on Pd^0-catalyzed functionalization of 5-iododeoxycytidine. The first crystal structure of fC reveals the existence of an intramolecular H-bond between the exocyclic amine and the formyl group, which controls the conformation of the formyl substituent. Using a newly designed in vitro mutagenicity assay we show that fC and caC are only marginally mutagenic, which is a prerequisite for the bases to function as epigenetic control units.

References

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737-738. *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*.

[2] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **2006**, *34*, 564-574. *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix*.

[3] A. P. Wolffe, J. J. Hayes, *Nucleic Acids Res.* **1999**, *27*, 711-720. *Chromatin disruption and modification*.

[4] C. E. Ford, J. L. Hamerton, *Nature* **1956**, *178*, 1020-1023. *The chromosomes of man*.

[5] B. M. Turner, *Bioessays* **2000**, *22*, 836-845. *Histone acetylation and an epigenetic code*.

[6] C. A. Espinoza, T. A. Allen, A. R. Hieb, J. F. Kugel, J. A. Goodrich, *Nat. Struct. Mol. Biol.* **2004**, *11*, 822-829. *B2 RNA binds directly to RNA polymerase II to repress transcript synthesis*.

[7] T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074-1080. *Translating the histone code*.

[8] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068-1070. *The role of DNA methylation in mammalian epigenetics*.

[9] A. D. Riggs, Z. Xiong, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4-5. *Methylation and epigenetic fidelity*.

[10] A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079-1088. *Epigenetic modifications in pluri-potent and differentiated cells*.

[11] Y. Hirabayashi, Y. Gotoh, *Nat. Rev. Neurosci.* **2010**, *11*, 377-388. *Epigenetic control of neural precursor cell fate during development*.

[12] M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465-476. *DNA methylation landscapes: provocative insights from epigenomics*.

[13] S. Augui, E. P. Nora, E. Heard, *Nat. Rev. Genet.* **2011**, *12*, 429-442. *Regulation of X-chromosome inactivation by the X-inactivation centre*.

[14] J. J. Day, J. D. Sweatt, *Neuron* **2011**, *70*, 813-829. *Epigenetic mechanisms in cognition*.

[15] J. S. You, P. A. Jones, *Cancer Cell* **2012**, *22*, 9-20. *Cancer genetics and epigenetics: two sides of the same coin?*

[16] R. Lister, et al., *Nature* **2009**, *462*, 315-322. *Human DNA methylomes at base resolution show widespread epigenomic differences*.

[17] L. Laurent, et al., *Genome Res.* **2010**, *20*, 320-331. *Dynamic changes in the human methylome during differentiation*.

[18] M. Gardiner-Garden, M. Frommer, *J. Mol. Biol.* **1987**, *196*, 261-282. *CpG Islands in vertebrate genomes*.

[19] M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic Acids Res.* **1982**, *10*, 2709-2721. *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*.

[20] F. Larsen, G. Gundersen, R. Lopez, H. Prydz, *Genomics* **1992**, *13*, 1095-1107. *CpG islands as gene markers in the human genome*.

[21] Y. Li, et al., *PLoS Biol.* **2010**, *8*, e1000533. *The DNA methylome of human peripheral blood mononuclear cells*.

[22] J. P. Sanford, H. J. Clark, V. M. Chapman, J. Rossant, *Genes Dev.* **1987**, *1*, 1039-1046. *Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse*.

[23] K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabo, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3642-3647. *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine*.

[24] C. R. Farthing, G. Ficz, R. K. Ng, C. F. Chan, S. Andrews, W. Dean, M. Hemberger, W. Reik, *PLoS Genet.* **2008**, *4*, e1000116. *Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes*.

[25] M. Jackson, A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell, B. Ramsahoye, *Mol. Cell. Biol.* **2004**, *24*, 8862-8871. *Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells*.

[26] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Science* **2006**, *311*, 395-398. *Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2*.

[27] C. L. Ulrey, L. Liu, L. G. Andrews, T. O. Tollefsbol, *Hum. Mol. Genet.* **2005**, *14*, R139-147. *The impact of metabolism on DNA methylation*.

[28] T. H. Bestor, *EMBO J.* **1992**, *11*, 2611-2617. *Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain*.

[29] M. Pradhan, P. O. Esteve, H. G. Chin, M. Samaranayke, G. D. Kim, S. Pradhan, *Biochemistry (Mosc).* **2008**, *47*, 10000-10009. *CXXC domain of human DNMT1 is essential for enzymatic activity*.

[30] H. F. Jorgensen, I. Ben-Porath, A. P. Bird, *Mol. Cell. Biol.* **2004**, *24*, 3387-3395. *Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains*.

[31] Y. Xu, et al., *Mol. Cell* **2011**, *42*, 451-464. *Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells*.

[32] Y. Xu, et al., *Cell* **2012**, *151*, 1200-1213. *Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development*.

[33] J. K. Kim, M. Samaranayake, S. Pradhan, *Cell. Mol. Life Sci.* **2009**, *66*, 596-612. *Epigenetic mechanisms in mammals*.

[34] A. Kubosaki, et al., *Biochem. Biophys. Res. Commun.* **2012**, *426*, 141-147. *CpG site-specific alteration of hydroxymethylcytosine to methylcytosine beyond DNA replication*.

[35] T. H. Bestor, *Hum. Mol. Genet.* **2000**, *9*, 2395-2402. *The DNA methyltransferases of mammals*.

[36] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **2011**, *331*, 1036-1040. *Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation*.

[37] J. Song, M. Teplova, S. Ishibe-Murakami, D. J. Patel, *Science* **2012**, *335*, 709-712. *Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation*.

[38] R. Gerasimaite, E. Merkiene, S. Klimasauskas, *Nucleic Acids Res.* **2011**, *39*, 3771-3780. *Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase*.

[39] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, *76*, 357-369. *HhaI methyltransferase flips its target base out of the DNA helix*.

[40] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257. *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*.

[41] I. Suetake, F. Shinozaki, J. Miyagawa, H. Takeshima, S. Tajima, *J. Biol. Chem.* **2004**, *279*, 27816-27823. *DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction*.

[42] W. Reik, *Nature* **2007**, *447*, 425-432. *Stability and flexibility of epigenetic gene regulation in mammalian development*.

[43] S. K. Ooi, et al., *Nature* **2007**, *448*, 714-717. *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA*.

[44] D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, T. H. Bestor, *Science* **2001**, *294*, 2536-2539. *Dnmt3L and the establishment of maternal genomic imprints*.

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

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

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

[48] Y. F. He, et al., *Science* **2011**, *333*, 1303-1307. *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*.

[49] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781-790. *The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid*.

[50] M. Tahiliani, et al., *Science* **2009**, *324*, 930-935. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*.

[51] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930. *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain*.

[52] C. X. Song, et al., *Nat. Biotechnol.* **2011**, *29*, 68-72. *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine*.

[53] W. A. Pastor, et al., *Nature* **2011**, *473*, 394-397. *Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells*.

[54] M. Yu, et al., *Cell* **2012**, *149*, 1368-1380. *Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome*.

[55] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398-402. *Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation*.

[56] C. Frauer, et al., *PLoS One* **2011**, *6*, e16627. *Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1*.

[57] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181. *Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA*.

[58] M. Munzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468. *5-Hydroxymethylcytosine, the sixth base of the genome*.

[59] M. M. Dawlaty, et al., *Cell Stem Cell* **2011**, *9*, 166-175. *Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development*.

[60] K. P. Koh, et al., *Cell Stem Cell* **2011**, *8*, 200-213. *Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells*.

[61] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-1133. *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*.

[62] M. Wossidlo, et al., *Nat. Commun.* **2011**, *2*, 241. *5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming*.

[63] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343-348. *TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity*.

[64] H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Nature* **2011**, *473*, 389-393. *Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells*.

[65] M. Ko, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14566-14571. *Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice*.

[66] Z. Li, X. Cai, C. L. Cai, J. Wang, W. Zhang, B. E. Petersen, F. C. Yang, M. Xu, *Blood* **2011**, *118*, 4509-4518. *Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies*.

[67] K. Moran-Crusio, et al., *Cancer Cell* **2011**, *20*, 11-24. *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation*.

[68] M. M. Dawlaty, et al., *Dev. Cell* **2013**, *24*, 310-323. *Combined deficiency of tet1 and tet2 causes epigenetic abnormalities but is compatible with postnatal development*.

[69] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2013**. *DNA 5-Methylcytosine Demethylation Activities of the Mammalian DNA Methyltransferases*.

[70] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620. *Active DNA demethylation: many roads lead to Rome*.

[71] A. Inoue, Y. Zhang, *Science* **2011**, *334*, 194. *Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos*.

[72] V. Valinluck, L. C. Sowers, *Cancer Res.* **2007**, *67*, 946-950. *Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1*.

[73] S. Schiesser, T. Pfaffeneder, G. Kashiwasaki, T. Carell, *to be published* **2013**. *The Chemical Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxycytosine*.

[74] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, *8*, 328-330. *Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA*.

[75] C. G. Spruijt, et al., *Cell* **2013**. *Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives*.

[76] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338. *Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites*.

[77] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell Cycle* **2011**, *10*, 2662-2668. *Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond*.

[78] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, *145*, 423-434. *Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain*.

[79] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Muller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520. *Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing*.

[80] H. Kamiya, H. Tsuchiya, N. Karino, Y. Ueno, A. Matsuda, H. Harashima, *J. Biochem.* **2002**, *132*, 551-555. *Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells*.

[81] N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456-2463. *Synthesis and properties of oligonucleotides containing 5-formyl-2'-deoxycytidine: in vitro DNA polymerase reactions on DNA templates containing 5-formyl-2'-deoxycytidine*.

[82] C. J. La Francois, Y. H. Jang, T. Cagin, W. A. Goddard, 3rd, L. C. Sowers, *Chem. Res. Toxicol.* **2000**, *13*, 462-470. *Conformation and proton configuration of pyrimidine deoxynucleoside oxidation damage products in water*.

[83] Q. M. Zhang, H. Sugiyama, I. Miyabe, S. Matsuda, K. Kino, I. Saito, S. Yonei, *Int. J. Radiat. Biol.* **1999**, *75*, 59-65. *Replication in vitro and cleavage by restriction endonuclease of 5-formyluracil- and 5-hydroxymethyluracil-containing oligonucleotides*.

[84] K. Kemmerich, F. A. Dingler, C. Rada, M. S. Neuberger, *Nucleic Acids Res.* **2012**, *40*, 6016-6025. *Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice*.

[85] C. Holz-Schietinger, D. M. Matje, M. F. Harrison, N. O. Reich, *J. Biol. Chem.* **2011**, *286*, 41479-41488. *Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation*.

[86] S. G. Jin, Y. Jiang, R. Qiu, T. A. Rauch, Y. Wang, G. Schackert, D. Krex, Q. Lu, G. P. Pfeifer, *Cancer Res.* **2011**, *71*, 7360-7365. *5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations*.

[87] M. L. Chen, F. Shen, W. Huang, J. H. Qi, Y. Wang, Y. Q. Feng, S. M. Liu, B. F. Yuan, *Clin. Chem.* **2013**. *Quantification of 5-Methylcytosine and 5-Hydroxymethylcytosine in Genomic DNA from Hepatocellular Carcinoma Tissues by Capillary Hydrophilic-Interaction Liquid Chromatography/ Quadrupole Time-of-Flight Mass Spectrometry*.

[88] T. F. Kraus, et al., *Int. J. Cancer* **2012**, *131*, 1577-1590. *Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors*.

[89] H. Yang, et al., *Oncogene* **2013**, *32*, 663-669. *Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation*.

[90] C. H. Hsu, et al., *Cell Rep.* **2012**, *2*, 568-579. *TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases*.

[91] C. G. Lian, et al., *Cell* **2012**, *150*, 1135-1146. *Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma*.

[92] S. Yamanaka, M. E. Balestra, L. D. Ferrell, J. Fan, K. S. Arnold, S. Taylor, J. M. Taylor, T. L. Innerarity, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8483-8487. *Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals*.

[93] R. Pavri, M. C. Nussenzweig, *Adv. Immunol.* **2011**, *110*, 1-26. *AID targeting in antibody diversity*.

[94] M. B. Burns, et al., *Nature* **2013**, *494*, 366-370. *APOBEC3B is an enzymatic source of mutation in breast cancer*.

[95] J. H. J. Hoeijmakers, *New Engl. J. Med.* **2009**, *361*, 1475-1485. *DNA Damage, Aging, and Cancer*.

[96] B. C. Dickinson, C. J. Chang, *Nat. Chem. Biol.* **2011**, *7*, 504-511. *Chemistry and biology of reactive oxygen species in signaling or stress responses*.

[97] J. Cadet, T. Douki, J. L. Ravanat, *Free Radic. Biol. Med.* **2010**, *49*, 9-21. *Oxidatively generated base damage to cellular DNA*.

[98] S. Steenken, *Chem. Rev.* **1989**, *89*, 503-520. *Purine bases, nucleosides, and nucleotides: aqueous solution redox chemistry and transformation reactions of their radical cations and e- and OH adducts*.

[99] A. P. Breen, J. A. Murphy, *Free Radic. Biol. Med.* **1995**, *18*, 1033-1077. *Reactions of oxyl radicals with DNA*.

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

[101] A. J. S. C. Vieira, S. Steenken, *J. Am. Chem. Soc.* **1990**, *112*, 6986-6994. *Pattern of hydroxy radical reaction with adenine and its nucleosides and nucleotides. Characterization of two types of isomeric hydroxy adduct and their unimolecular transformation reactions*.

[102] J. Cadet, T. Douki, D. Gasparutto, J. L. Ravanat, *Mutat. Res.* **2003**, *531*, 5-23. *Oxidative damage to DNA: formation, measurement and biochemical features*.

[103] E. Gajewski, G. Rao, Z. Nackerdien, M. Dizdaroglu, *Biochemistry* **1990**, *29*, 7876-7882. *Modification of DNA bases in mammalian chromatin by radiation-generated free radicals*.

[104] G. V. Buxton, C. L. Greenstock, W. P. Helman, A. B. Ross, *J. Phys. Chem. Ref. Data* **1988**, *17*, 513-886. *Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ([center-dot]OH/[center-dot]O[sup -] in Aqueous Solution*.

[105] J. Cadet, et al., *J. Environ. Pathol. Toxicol. Oncol.* **2004**, *23*, 33-43. *Radiation-induced DNA damage: formation, measurement, and biochemical features*.

[106] T. Douki, R. Martini, J. L. Ravanat, R. J. Turesky, J. Cadet, *Carcinogenesis* **1997**, *18*, 2385-2391. *Measurement of 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 8-oxo-7,8-dihydroguanine in isolated DNA exposed to gamma radiation in aqueous solution*.

[107] P. W. Doetsch, T. H. Zasatawny, A. M. Martin, M. Dizdaroglu, *Biochemistry* **1995**, *34*, 737-742. *Monomeric base damage products from adenine, guanine, and thymine induced by exposure of DNA to ultraviolet radiation*.

[108] D. B. Medinas, G. Cerchiaro, D. F. Trindade, O. Augusto, *IUBMB Life* **2007**, *59*, 255-262. *The carbonate radical and related oxidants derived from bicarbonate buffer*.

[109] L. P. Candeias, S. Steenken, *Chemistry* **2000**, *6*, 475-484. *Reaction of HO* with guanine derivatives in aqueous solution: formation of two different redox-active OH-adduct radicals and their unimolecular transformation reactions. Properties of G(-H)**.

[110] S. Raoul, M. Berger, G. W. Buchko, P. C. Joshi, B. Morin, M. Weinfeld, J. Cadet, *J. Chem. Soc., Perkin Trans. 2* **1996**, 371-381. *1H, 13C and 15N nuclear magnetic resonance analysis and chemical features of the two main radical oxidation products of 2[prime or minute]-deoxyguanosine: oxazolone and imidazolone nucleosides*.

[111] J. Cadet, M. Berger, G. W. Buchko, P. C. Joshi, S. Raoul, J.-L. Ravanat, *J. Am. Chem. Soc.* **1994**, *116*, 7403-7404. *2,2-Diamino-4-[(3,5-di-O-acetyl-2-deoxy-.beta.-D-erythro-pentofuranosyl)amino]-5-(2H)-oxazolone: a Novel and Predominant Radical Oxidation Product of 3',5'-Di-O-acetyl-2'-deoxyguanosine*.

[112] S. Steenken, S. V. Jovanovic, M. Bietti, K. Bernhard, *J. Am. Chem. Soc.* **2000**, *122*, 2373-2374. *The Trap Depth (in DNA) of 8-Oxo-7,8-dihydro-2‘deoxyguanosine as Derived from Electron-Transfer Equilibria in Aqueous Solution*.

[113] J. L. Ravanat, C. Saint-Pierre, J. Cadet, *J. Am. Chem. Soc.* **2003**, *125*, 2030-2031. *One-electron oxidation of the guanine moiety of 2'-deoxyguanosine: influence of 8-oxo-7,8-dihydro-2'-deoxyguanosine*.

[114] N. R. Jena, P. C. Mishra, *Free Radic. Biol. Med.* **2012**, *53*, 81-94. *Formation of ring-opened and rearranged products of guanine: mechanisms and biological significance*.

[115] A. Adhikary, A. Kumar, D. Khanduri, M. D. Sevilla, *J. Am. Chem. Soc.* **2008**, *130*, 10282-10292. *Effect of base stacking on the acid-base properties of the adenine cation radical [A*+] in solution: ESR and DFT studies*.

[116] S. Raoul, M. Bardet, J. Cadet, *Chem. Res. Toxicol.* **1995**, *8*, 924-933. *Gamma irradiation of 2'-deoxyadenosine in oxygen-free aqueous solutions: identification and conformational features of formamidopyrimidine nucleoside derivatives*.

[117] M. M. Greenberg, *Radical and Radical Ion Reactivity in Nucleic Acid Chemistry*, Wiley, **2009**.

[118] H. Yanagawa, Y. Ogawa, M. Ueno, *J. Biol. Chem.* **1992**, *267*, 13320-13326. *Redox ribonucleosides. Isolation and characterization of 5-hydroxyuridine, 8-hydroxy-guanosine, and 8-hydroxyadenosine from Torula yeast RNA*.

[119] M. M. Greenberg, *Acc. Chem. Res.* **2012**, *45*, 588-597. *The formamidopyrimidines: purine lesions formed in competition with 8-oxopurines from oxidative stress*.

[120] M. M. Greenberg, Z. Hantosi, C. J. Wiederholt, C. D. Rithner, *Biochemistry* **2001**, *40*, 15856-15861. *Studies on N4-(2-deoxy-D-pentofuranosyl)-4,6-diamino-5-formamido-pyrimidine (Fapy.dA) and N6-(2-deoxy-D-pentofuranosyl)-6-diamino-5-formamido-4-hydroxypyrimidine (Fapy.dG)*.

[121] M. E. Smela, M. L. Hamm, P. T. Henderson, C. M. Harris, T. M. Harris, J. M. Essigmann, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6655-6660. *The aflatoxin B(1) formamidopyrimidine adduct plays a major role in causing the types of mutations observed in human hepatocellular carcinoma*.

[122] F. Busch, J. C. Pieck, M. Ober, J. Gierlich, G. W. Hsu, L. S. Beese, T. Carell, *Chemistry* **2008**, *14*, 2125-2132. *Dissecting the differences between the alpha and beta anomers of the oxidative DNA lesion FaPydG*.

[123] H. Mao, Z. Deng, F. Wang, T. M. Harris, M. P. Stone, *Biochemistry* **1998**, *37*, 4374-4387. *An intercalated and thermally stable FAPY adduct of aflatoxin B1 in a DNA duplex: structural refinement from 1H NMR*.

[124] K. L. Brown, J. Z. Deng, R. S. Iyer, L. G. Iyer, M. W. Voehler, M. P. Stone, C. M. Harris, T. M. Harris, *J. Am. Chem. Soc.* **2006**, *128*, 15188-15199. *Unraveling the aflatoxin-FAPY conundrum: structural basis for differential replicative processing of isomeric forms of the formamidopyrimidine-type DNA adduct of aflatoxin B1*.

[125] J. N. Patro, K. Haraguchi, M. O. Delaney, M. M. Greenberg, *Biochemistry* **2004**, *43*, 13397-13403. *Probing the configurations of formamidopyrimidine lesions Fapy.dA and Fapy.dG in DNA using endonuclease IV*.

[126] J. M. Aramini, S. H. Cleaver, R. T. Pon, R. P. Cunningham, M. W. Germann, *J. Mol. Biol.* **2004**, *338*, 77-91. *Solution structure of a DNA duplex containing an alpha-anomeric adenosine: insights into substrate recognition by endonuclease IV*.

[127] K. Haraguchi, M. M. Greenberg, *J. Am. Chem. Soc.* **2001**, *123*, 8636-8637. *Synthesis of oligonucleotides containing Fapy.dG (N6-(2-deoxy-alpha,beta-D-erythro-pentofurano-syl)-2,6-diamino-4-hydroxy-5-formami dopyrimidine)*.

[128] K. Haraguchi, M. O. Delaney, C. J. Wiederholt, A. Sambandam, Z. Hantosi, M. M. Greenberg, *Nucleic Acids Res. Suppl.* **2001**, 129-130. *Synthesis and characterization of oligonucleotides containing formamidopyrimidine lesions (Fapy.dA, Fapy.dG) at defined sites*.

[129] M. O. Delaney, M. M. Greenberg, *Chem. Res. Toxicol.* **2002**, *15*, 1460-1465. *Synthesis of oligonucleotides and thermal stability of duplexes containing the beta-C-nucleoside analogue of Fapy*dG*.

[130] M. Ober, U. Linne, J. Gierlich, T. Carell, *Angew. Chem. Int. Ed.* **2003**, *42*, 4947-4951. *The two main DNA lesions 8-Oxo-7,8-dihydroguanine and 2,6-diamino-5-formamido-4-hydroxypyrimidine exhibit strongly different pairing properties*.

[131] T. A. Steitz, Y. W. Yin, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359*, 17-23. *Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases*.

[132] T. W. Traut, *Mol. Cell Biochem.* **1994**, *140*, 1-22. *Physiological concentrations of purines and pyrimidines*.

[133] J. A. Brown, Z. Suo, *Biochemistry* **2011**, *50*, 1135-1142. *Unlocking the sugar "steric gate" of DNA polymerases*.

[134] W. Wang, E. Y. Wu, H. W. Hellinga, L. S. Beese, *J. Biol. Chem.* **2012**, *287*, 28215-28226. *Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides*.

[135] K. Haraguchi, M. O. Delaney, C. J. Wiederholt, A. Sambandam, Z. Hantosi, M. M. Greenberg, *J. Am. Chem. Soc.* **2002**, *124*, 3263-3269. *Synthesis and characterization of oligodeoxynucleotides containing formamidopyrimidine lesions and nonhydrolyzable analogues*.

[136] M. Ober, H. Muller, C. Pieck, J. Gierlich, T. Carell, *J. Am. Chem. Soc.* **2005**, *127*, 18143-18149. *Base pairing and replicative processing of the formamidopyrimidine-dG DNA lesion*.

[137] S. Smirnov, F. Johnson, R. Marumoto, C. de los Santos, *J. Biomol. Struct. Dyn.* **2000**, *17*, 981-991. *Structure of an 11-mer DNA duplex containing the carbocyclic nucleotide analog: 2'-deoxyaristeromycin*.

[138] S. D. Bruner, D. P. Norman, G. L. Verdine, *Nature* **2000**, *403*, 859-866. *Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA*.

[139] M. Ferrero, V. Gotor, *Chem. Rev.* **2000**, *100*, 4319-4348. *Biocatalytic selective modifications of conventional nucleosides, carbocyclic nucleosides, and C-nucleosides*.

[140] T. A. Halgren, *J. Comput. Chem.* **1999**, *20*, 720-729. *MMFF VI. MMFF94s option for energy minimization studies*.

[141] C. J. Wiederholt, M. M. Greenberg, *J. Am. Chem. Soc.* **2002**, *124*, 7278-7279. *Fapy.dG instructs Klenow exo(-) to misincorporate deoxyadenosine*.

[142] M. O. Delaney, C. J. Wiederholt, M. M. Greenberg, *Angew. Chem. Int. Ed.* **2002**, *41*, 771-773. *Fapy.dA induces nucleotide misincorporation translesionally by a DNA polymerase*.

[143] X. Tan, A. P. Grollman, S. Shibutani, *Carcinogenesis* **1999**, *20*, 2287-2292. *Comparison of the mutagenic properties of 8-oxo-7,8-dihydro-2'-deoxyadenosine and 8-oxo-7,8-dihydro-2'-deoxyguanosine DNA lesions in mammalian cells*.

[144] M. A. Kalam, K. Haraguchi, S. Chandani, E. L. Loechler, M. Moriya, M. M. Greenberg, A. K. Basu, *Nucleic Acids Res.* **2006**, *34*, 2305-2315. *Genetic effects of oxidative DNA damages: comparative mutagenesis of the imidazole ring-opened formamido-pyrimidines (Fapy lesions) and 8-oxo-purines in simian kidney cells*.

[145] J. E. Sale, *J. Cell. Sci.* **2012**, *125*, 1633-1643. *Competition, collaboration and coordination--determining how cells bypass DNA damage*.

[146] U. Hübscher, Spadari, S., Villani, G., Maga, G., *DNA POLYMERASES - Discovery, Characterizations and Functions in Cellular DNA Transactions*, World Scientific Publishing Co. PTE. Ltd., Singapure, **2010**.

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

[148] A. R. Lehmann, A. Niimi, T. Ogi, S. Brown, S. Sabbioneda, J. F. Wing, P. L. Kannouche, C. M. Green, *DNA Repair* **2007**, *6*, 891-899. *Translesion synthesis: Y-family polymerases and the polymerase switch*.

[149] S. Prakash, R. E. Johnson, L. Prakash, *Annu. Rev. Biochem.* **2005**, *74*, 317-353. *Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function*.

[150] M. Shaheen, I. Shanmugam, R. Hromas, *J. Nucleic Acids* **2010**, *2010*. *The Role of PCNA Posttranslational Modifications in Translesion Synthesis*.

[151] U. Hubscher, G. Maga, *Curr. Opin. Chem. Biol.* **2011**, *15*, 627-635. *DNA replication and repair bypass machines*.

[152] A. Schlabach, B. Fridlender, A. Bolden, A. Weissbach, *Biochem. Biophys. Res. Commun.* **1971**, *44*, 879-885. *DNA-dependent DNA polymerases from HeLa cell nuclei. II. Template and substrate utilization*.

[153] C. S. McHenry, *Annu. Rev. Biochem.* **1988**, *57*, 519-550. *DNA polymerase III holoenzyme of Escherichia coli*.

[154] T. A. Steitz, *J. Biol. Chem.* **1999**, *274*, 17395-17398. *DNA polymerases: structural diversity and common mechanisms*.

[155] C. M. Joyce, T. A. Steitz, *Annu. Rev. Biochem.* **1994**, *63*, 777-822. *Function and structure relationships in DNA polymerases*.

[156] I. V. Shevelev, U. Hubscher, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 364-376. *The 3' 5' exonucleases*.

[157] P. M. Burgers, et al., *J. Biol. Chem.* **2001**, *276*, 43487-43490. *Eukaryotic DNA polymerases: proposal for a revised nomenclature*.

[158] M. J. Bessman, A. Kornberg, I. R. Lehman, E. S. Simms, *Biochim. Biophys. Acta* **1956**, *21*, 197-198. *Enzymic synthesis of deoxyribonucleic acid*.

[159] T. Okazaki, A. Kornberg, *J. Biol. Chem.* **1964**, *239*, 259-268. *Enzymatic Synthesis of Deoxyribonucleic Acid. Xv. Purification and Properties of a Polymerase from Bacillus Subtilis*.

[160] J. R. Kiefer, C. Mao, C. J. Hansen, S. L. Basehore, H. H. Hogrefe, J. C. Braman, L. S. Beese, *Structure* **1997**, *5*, 95-108. *Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 A resolution*.

[161] A. J. Berdis, *Chem. Rev.* **2009**, *109*, 2862-2879. *Mechanisms of DNA polymerases*.

[162] S. J. Johnson, J. S. Taylor, L. S. Beese, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3895-3900. *Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations*.

[163] K. Bebenek, C. M. Joyce, M. P. Fitzgerald, T. A. Kunkel, *J. Biol. Chem.* **1990**, *265*, 13878-13887. *The fidelity of DNA synthesis catalyzed by derivatives of Escherichia coli DNA polymerase I*.

[164] G. Dianov, T. Lindahl, *Curr. Biol.* **1994**, *4*, 1069-1076. *Reconstitution of the DNA base excision-repair pathway*.

[165] P. K. Cooper, P. C. Hanawalt, *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 1156-1160. *Role of DNA polymerase I and the rec system in excision-repair in Escherichia coli*.

[166] D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, *Nature* **1985**, *313*, 762-766. *Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP*.

[167] L. G. Lowe, F. P. Guengerich, *Biochemistry* **1996**, *35*, 9840-9849. *Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by Escherichia coli polymerases I exo- and II exo*.

[168] C. M. Joyce, N. D. Grindley, *J. Bacteriol.* **1982**, *152*, 1211-1219. *Identification of two genes immediately downstream from the polA gene of Escherichia coli*.

[169] F. C. Lawyer, S. Stoffel, R. K. Saiki, K. Myambo, R. Drummond, D. H. Gelfand, *J. Biol. Chem.* **1989**, *264*, 6427-6437. *Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from Thermus aquaticus*.

[170] M. L. Hamm, et al., *Biochemistry* **2011**, *50*, 10713-10723. *Importance of the C2, N7, and C8 positions to the mutagenic potential of 8-Oxo-2'-deoxyguanosine with two A family polymerases*.

[171] L. S. Beese, T. A. Steitz, *EMBO J.* **1991**, *10*, 25-33. *Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism*.

[172] V. Derbyshire, N. D. Grindley, C. M. Joyce, *EMBO J.* **1991**, *10*, 17-24. *The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction*.

[173] A. H. Polesky, M. E. Dahlberg, S. J. Benkovic, N. D. Grindley, C. M. Joyce, *J. Biol. Chem.* **1992**, *267*, 8417-8428. *Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from Escherichia coli*.

[174] A. H. Polesky, T. A. Steitz, N. D. Grindley, C. M. Joyce, *J. Biol. Chem.* **1990**, *265*, 14579-14591. *Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from Escherichia coli*.

[175] S. Korolev, M. Nayal, W. M. Barnes, E. Di Cera, G. Waksman, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9264-9268. *Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-A resolution: structural basis for thermostability*.

[176] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304-307. *Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal*.

[177] G. W. Hsu, J. R. Kiefer, D. Burnouf, O. J. Becherel, R. P. P. Fuchs, L. S. Beese, *J. Biol. Chem.* **2004**, *279*, 50280-50285. *Observing Translesion Synthesis of an Aromatic Amine DNA Adduct by a High-fidelity DNA Polymerase*.

[178] S. J. Johnson, L. S. Beese, *Cell* **2004**, *116*, 803-816. *Structures of mismatch replication errors observed in a DNA polymerase*.

[179] G. W. Hsu, X. Huang, N. P. Luneva, N. E. Geacintov, L. S. Beese, *J. Biol. Chem.* **2005**, *280*, 3764-3770. *Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication*.

[180] J. J. Warren, L. J. Forsberg, L. S. Beese, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 19701-19706. *The structural basis for the mutagenicity of O(6)-methyl-guanine lesions*.

[181] M. Trostler, A. Delier, J. Beckman, M. Urban, J. N. Patro, T. E. Spratt, L. S. Beese, R. D. Kuchta, *Biochemistry* **2009**, *48*, 4633-4641. *Discrimination between Right and Wrong Purine dNTPs by DNA Polymerase I from Bacillus stearothermophilus*.

[182] W. Wang, H. W. Hellinga, L. S. Beese, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17644-17648. *Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis*.

[183] E. Y. Wu, L. S. Beese, *J. Biol. Chem.* **2011**, *286*, 19758-19767. *The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism*.

[184] A. A. Golosov, J. J. Warren, L. S. Beese, M. Karplus, *Structure* **2010**, *18*, 83-93. *The mechanism of the translocation step in DNA replication by DNA polymerase I: a computer simulation analysis*.

[185] H. Miller, A. P. Grollman, *Biochemistry* **1997**, *36*, 15336-15342. *Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis*.

[186] E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191-219. *Active site tightness and substrate fit in DNA replication*.

[187] M. F. Goodman, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10493-10495. *Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity*.

[188] S. S. Carroll, M. Cowart, S. J. Benkovic, *Biochemistry* **1991**, *30*, 804-813. *A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity*.

[189] K. Bebenek, T. A. Kunkel, *Adv. Protein Chem.* **2004**, *69*, 137-165. *Functions of DNA polymerases*.

[190] P. Kannouche, B. C. Broughton, M. Volker, F. Hanaoka, L. H. Mullenders, A. R. Lehmann, *Genes Dev.* **2001**, *15*, 158-172. *Domain structure, localization, and function of DNA polymerase eta, defective in xeroderma pigmentosum variant cells*.

[191] S. D. McCulloch, R. J. Kokoska, C. Masutani, S. Iwai, F. Hanaoka, T. A. Kunkel, *Nature* **2004**, *428*, 97-100. *Preferential cis-syn thymine dimer bypass by DNA polymerase eta occurs with biased fidelity*.

[192] R. E. Johnson, S. Prakash, L. Prakash, *Science* **1999**, *283*, 1001-1004. *Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta*.

[193] A. Alt, K. Lammens, C. Chiocchini, A. Lammens, J. C. Pieck, D. Kuch, K. P. Hopfner, T. Carell, *Science* **2007**, *318*, 967-970. *Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta*.

[194] T. D. Silverstein, R. E. Johnson, R. Jain, L. Prakash, S. Prakash, A. K. Aggarwal, *Nature* **2010**, *465*, 1039-1043. *Structural basis for the suppression of skin cancers by DNA polymerase eta*.

[195] C. Biertumpfel, et al., *Nature* **2010**, *465*, 1044-1048. *Structure and mechanism of human DNA polymerase eta*.

[196] M. Yasui, H. Dong, R. R. Bonala, N. Suzuki, H. Ohmori, F. Hanaoka, F. Johnson, A. P. Grollman, S. Shibutani, *Biochemistry* **2004**, *43*, 15005-15013. *Mutagenic properties of 3-(deoxyguanosin-N2-yl)-2-acetylaminofluorene, a persistent acetylaminofluorene-derived DNA adduct in mammalian cells*.

[197] C. Masutani, R. Kusumoto, S. Iwai, F. Hanaoka, *EMBO J.* **2000**, *19*, 3100-3109. *Mechanisms of accurate translesion synthesis by human DNA polymerase eta*.

[198] A. Bresson, R. P. Fuchs, *EMBO J.* **2002**, *21*, 3881-3887. *Lesion bypass in yeast cells: Pol eta participates in a multi-DNA polymerase process*.

[199] L. Haracska, S. L. Yu, R. E. Johnson, L. Prakash, S. Prakash, *Nat. Genet.* **2000**, *25*, 458-461. *Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta*.

[200] T. D. Silverstein, R. Jain, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, *Structure* **2010**, *18*, 1463-1470. *Structural basis for error-free replication of oxidatively damaged DNA by yeast DNA polymerase eta*.

[201] R. Vasquez-Del Carpio, T. D. Silverstein, S. Lone, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, *J. Mol. Biol.* **2011**, *408*, 252-261. *Role of human DNA polymerase kappa in extension opposite from a cis-syn thymine dimer*.

[202] S. Lone, S. A. Townson, S. N. Uljon, R. E. Johnson, A. Brahma, D. T. Nair, S. Prakash, L. Prakash, A. K. Aggarwal, *Mol. Cell* **2007**, *25*, 601-614. *Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass*.

[203] S. N. Uljon, R. E. Johnson, T. A. Edwards, S. Prakash, L. Prakash, A. K. Aggarwal, *Structure* **2004**, *12*, 1395-1404. *Crystal structure of the catalytic core of human DNA polymerase kappa*.

[204] M. T. Washington, R. E. Johnson, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1910-1914. *Human DINB1-encoded DNA polymerase kappa is a promiscuous extender of mispaired primer termini*.

[205] R. E. Johnson, S. Prakash, L. Prakash, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3838-3843. *The human DINB1 gene encodes the DNA polymerase Poltheta*.

[206] E. Ohashi, K. Bebenek, T. Matsuda, W. J. Feaver, V. L. Gerlach, E. C. Friedberg, H. Ohmori, T. A. Kunkel, *J. Biol. Chem.* **2000**, *275*, 39678-39684. *Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene*.

[207] L. Haracska, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16000-16005. *Role of human DNA polymerase kappa as an extender in translesion synthesis*.

[208] R. V.-D. Carpio, T. D. Silverstein, S. Lone, M. K. Swan, J. R. Choudhury, R. E. Johnson, S. Prakash, L. Prakash, A. K. Aggarwal, *PLoS One* **2009**, *4*, e5766. *Structure of Human DNA Polymerase κ Inserting dATP Opposite an 8-OxoG DNA Lesion*.

[209] Y. Zhang, F. Yuan, X. Wu, M. Wang, O. Rechkoblit, J. S. Taylor, N. E. Geacintov, Z. Wang, *Nucleic Acids Res.* **2000**, *28*, 4138-4146. *Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro*.

[210] Y. J. Kim, D. M. Wilson III, *Curr. Mol. Pharmacol.* **2012**, *5*, 3-13. *Overview of base excision repair biochemistry*.

[211] F. Coste, M. Ober, T. Carell, S. Boiteux, C. Zelwer, B. Castaing, *J. Biol. Chem.* **2004**, *279*, 44074-44083. *Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase*.

[212] P. A. van der Kemp, D. Thomas, R. Barbey, R. de Oliveira, S. Boiteux, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5197-5202. *Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine*.

[213] G. Dianov, A. Price, T. Lindahl, *Mol. Cell Biol.* **1992**, *12*, 1605-1612. *Generation of single-nucleotide repair patches following excision of uracil residues from DNA*.

[214] G. Frosina, P. Fortini, O. Rossi, F. Carrozzino, G. Raspaglio, L. S. Cox, D. P. Lane, A. Abbondandolo, E. Dogliotti, *J. Biol. Chem.* **1996**, *271*, 9573-9578. *Two pathways for base excision repair in mammalian cells*.

[215] A. Klungland, T. Lindahl, *EMBO J.* **1997**, *16*, 3341-3348. *Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1)*.

[216] B. van Loon, U. Hubscher, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18201-18206. *An 8-oxo-guanine repair pathway coordinated by MUTYH glycosylase and DNA polymerase lambda*.

[217] J. Tchou, V. Bodepudi, S. Shibutani, I. Antoshechkin, J. Miller, A. P. Grollman, F. Johnson, *J. Biol. Chem.* **1994**, *269*, 15318-15324. *Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA*.

[218] C. J. Chetsanga, M. Lozon, C. Makaroff, L. Savage, *Biochemistry* **1981**, *20*, 5201-5207. *Purification and characterization of Escherichia coli formamidopyrimidine-DNA glycosylase that excises damaged 7-methylguanine from deoxyribonucleic acid*.

[219] S. Boiteux, T. R. O'Connor, F. Lederer, A. Gouyette, J. Laval, *J. Biol. Chem.* **1990**, *265*, 3916-3922. *Homogeneous Escherichia coli FPG protein. A DNA glycosylase which excises imidazole ring-opened purines and nicks DNA at apurinic/apyrimidinic sites*.

[220] P. Jaruga, M. Birincioglu, T. A. Rosenquist, M. Dizdaroglu, *Biochemistry* **2004**, *43*, 15909-15914. *Mouse NEIL1 protein is specific for excision of 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 4,6-diamino-5-formamidopyrimidine from oxidatively damaged DNA*.

[221] C. J. Wiederholt, M. O. Delaney, M. A. Pope, S. S. David, M. M. Greenberg, *Biochemistry* **2003**, *42*, 9755-9760. *Repair of DNA containing Fapy.dG and its beta-C-nucleoside analogue by formamidopyrimidine DNA glycosylase and MutY*.

[222] N. Krishnamurthy, K. Haraguchi, M. M. Greenberg, S. S. David, *Biochemistry* **2008**, *47*, 1043-1050. *Efficient removal of formamidopyrimidines by 8-oxoguanine glycosylases*.

[223] J. C. Fromme, G. L. Verdine, *J. Biol. Chem.* **2003**, *278*, 51543-51548. *DNA lesion recognition by the bacterial repair enzyme MutM*.

[224] V. Duarte, D. Gasparutto, M. Jaquinod, J. Cadet, *Nucleic Acids Res.* **2000**, *28*, 1555-1563. *In vitro DNA synthesis opposite oxazolone and repair of this DNA damage using modified oligonucleotides*.

[225] J. Hu, N. C. de Souza-Pinto, K. Haraguchi, B. A. Hogue, P. Jaruga, M. M. Greenberg, M. Dizdaroglu, V. A. Bohr, *J. Biol. Chem.* **2005**, *280*, 40544-40551. *Repair of formamidopyrimidines in DNA involves different glycosylases: role of the OGG1, NTH1, and NEIL1 enzymes*.

[226] P. M. Girard, C. D'Ham, J. Cadet, S. Boiteux, *Carcinogenesis* **1998**, *19*, 1299-1305. *Opposite base-dependent excision of 7,8-dihydro-8-oxoadenine by the Ogg1 protein of Saccharomyces cerevisiae*.

[227] M. L. Michaels, C. Cruz, A. P. Grollman, J. H. Miller, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7022-7025. *Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA*.

[228] C. J. Wiederholt, J. N. Patro, Y. L. Jiang, K. Haraguchi, M. M. Greenberg, *Nucleic Acids Res.* **2005**, *33*, 3331-3338. *Excision of formamidopyrimidine lesions by endonucleases III and VIII is not a major DNA repair pathway in Escherichia coli*.

[229] T. K. Hazra, T. Izumi, I. Boldogh, B. Imhoff, Y. W. Kow, P. Jaruga, M. Dizdaroglu, S. Mitra, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 3523-3528. *Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA*.

[230] M. A. Kalam, A. K. Basu, *Chem. Res. Toxicol.* **2005**, *18*, 1187-1192. *Mutagenesis of 8-oxoguanine adjacent to an abasic site in simian kidney cells: tandem mutations and enhancement of G-->T transversions*.

[231] M. A. Graziewicz, T. H. Zastawny, R. Olinski, E. Speina, J. Siedlecki, B. Tudek, *Free Radic. Biol. Med.* **2000**, *28*, 75-83. *Fapyadenine is a moderately efficient chain terminator for prokaryotic DNA polymerases*.

[232] J. N. Patro, C. J. Wiederholt, Y. L. Jiang, J. C. Delaney, J. M. Essigmann, M. M. Greenberg, *Biochemistry* **2007**, *46*, 10202-10212. *Studies on the Replication of the Ring Opened Formamidopyrimidine, Fapy·dG in Escherichia coli*.

[233] M. Ober, M. Marsch, K. Harms, T. Carell, *Acta Crystallogr. E* **2004**, *60*, o1191-o1192. *A carbocyclic analogue of a protected β- d-2-deoxyribosylamine*.

[234] K. Asagoshi, H. Terato, Y. Ohyama, H. Ide, *J. Biol. Chem.* **2002**, *277*, 14589-14597. *Effects of a guanine-derived formamidopyrimidine lesion on DNA replication: translesion DNA synthesis, nucleotide insertion, and extension kinetics*.

[235] P. P. Christov, K. C. Angel, F. P. Guengerich, C. J. Rizzo, *Chem. Res. Toxicol.* **2009**, *22*, 1086-1095. *Replication past the N5-methyl-formamidopyrimidine lesion of deoxyguanosine by DNA polymerases and an improved procedure for sequence analysis of in vitro bypass products by mass spectrometry*.

[236] B. Tudek, M. Graziewicz, O. Kazanova, T. H. Zastawny, T. Obtulowicz, J. Laval, *Acta Biochim. Pol.* **1999**, *46*, 785-799. *Mutagenic specificity of imidazole ring-opened 7-methylpurines in M13mp18 phage DNA*.

[237] S. Shibutani, M. Takeshita, A. P. Grollman, *Nature* **1991**, *349*, 431-434. *Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG*.

[238] P. T. Henderson, J. C. Delaney, F. Gu, S. R. Tannenbaum, J. M. Essigmann, *Biochemistry* **2002**, *41*, 914-921. *Oxidation of 7,8-dihydro-8-oxoguanine affords lesions that are potent sources of replication errors in vivo*.

[239] W. L. Neeley, J. C. Delaney, P. T. Henderson, J. M. Essigmann, *J. Biol. Chem.* **2004**, *279*, 43568-43573. *In vivo bypass efficiencies and mutational signatures of the guanine oxidation products 2-aminoimidazolone and 5-guanidino-4-nitroimidazole*.

[240] K. Kino, H. Sugiyama, *Chem. Biol.* **2001**, *8*, 369-378. *Possible cause of G-C-->C-G transversion mutation by guanine oxidation product, imidazolone*.

[241] B. Matter, D. Malejka-Giganti, A. S. Csallany, N. Tretyakova, *Nucleic Acids Res.* **2006**, *34*, 5449-5460. *Quantitative analysis of the oxidative DNA lesion, 2,2-diamino-4-(2-deoxy-beta-D-erythro-pentofuranosyl)amino]-5(2H)-oxazolone (oxazolone), in vitro and in vivo by isotope dilution-capillary HPLC-ESI-MS/MS*.

Abbreviations

8-oxodA	8-Oxo-7,8-dihydro-2’-deoxyadenosine
8-oxodG	8-oxo-7,8-dihydro-2’-deoxyguanosine
Å	Angstrom
aa	amino acid
AID	activation induced cytidine deaminase
AP site	apurinic/apyrimidinic site= abasic site
APE	apurinic endonuclease
ATP	adenosine triphosphate
BER	base excision repair
c	carbocyclic
CPD	cyclobutane pyrimidine/thymine dimers
d	2'desoxy
dA	2'-desoxyadenosine
dATP	2'-desoxyadeninetriphosphat
dC	2'-desoxycytidine
dCTP	2'-desoxycytidinetriphosphat
dG	2'-desoxyguanosine
dGTP	2'-desoxyguaninetriphosphat
dIz	2,5-diamino-4*H*-imidazol-4-one, imidazolone
DNA	desoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	2'-desoxynucleotidetriphosphat
dZ	2,2,4-triamino-5-(2*H*)-oxazolone, oxazolone
DNA	deoxyribonucleic acid
dR	2'-desoxyribose

dRP	5'-deoxyribose-5'-phosphate
dT	2'-desoxythymidine
dTTP	2'-desoxythymidine triphosphat
et al.	lat. *et alii*, and others
FaPydA	4,6-diamino-5-formamidopyrimidine
FaPydG	2,6-diamino-4-oxo-5-formamidopyrimidine
FEN	Flap endonuclease
G	guanine
h	hour
His	histidine
HPLC	high performance liquid chromatography
Is	insertion site
Iz	2,5-diamino-4*H*-imidazol-4-one, imidazolone
KF^{exo-}	Klenow Fragment (without 5'-3' exonuclease)
L	liter
LIG	DNA ligase
LF	little finger domain
M	molecular weight (mol/L)
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation – time of flight
min	minute
ms	milli seconds
NER	nucleotide excision repair
NHE	normal hydrogen electrode
nm	nanometer
NMR	nuclear magnetic resonance
O_2	oxygen
ODN	oligodesoxynucleotide

PAD	polymerase-associated domain
PCNA	proliferating cell nuclear antigen
PG	phosphoglycolate
PNKP	Polynucleotide kinase / phosphatase
Pol	polymerase
post-IS	post-insertion site (n-1)
pre-IS	pre-insertion site
PUA	phosphor-α,β-unsaturated aldehyde
Py	pyridine
R	rest
RFC	replication factor C
RNA	ribonucleic acid
ROS	reactive oxygen species
RT	room temperature
s	seconds
T	temperature
TDG	thymidine-DNA glycosylase
TET	ten-eleven translocation protein
TLS	translesion synthesis
Tyr	tyrosine
UV	ultraviolett
V	Volt
vs.	Versus
XRCC	X-ray repair cross complementing
Z	2,2,4-triamino-5-(2*H*)-oxazolone, oxazolone
λ	wave length

Publications (Full Versions including Supplementary Informations)

Gehrke, T.H.,[1] Lischke, U.,[1] Gasteiger, K. L.,[1] Schneider, S.,[1] Arnold, S., Müller, H. C., Stephenson, D. S., Zipse, H., Carell, T. "Unexpected non-Hoogsteen-based mutagenicity mechanism of FaPy-DNA lesions." ***Nat. Chem. Bio.*****, 2013, 9, 455-461.**

nature chemical biology

ARTICLE

PUBLISHED ONLINE: 19 MAY 2013 | DOI: 10.1038/NCHEMBIO.1254

Unexpected non-Hoogsteen–based mutagenicity mechanism of FaPy-DNA lesions

Tim H Gehrke[1,3], Ulrike Lischke[1,3], Karola L Gasteiger[1,3], Sabine Schneider[1-3], Simone Arnold[1], Heiko C Müller[1], David S Stephenson[1], Hendrik Zipse[1] & Thomas Carell[1]*

8-Oxopurines (8-oxodG and 8-oxodA) and formamidopyrimidines (FaPydG and FaPydA) are major oxidative DNA lesions involved in cancer development and aging. Their mutagenicity is believed to result from a conformational shift of the N9-C1′ glycosidic bonds from *anti* to *syn*, which allows the lesions to form noncanonical Hoogsteen-type base pairs with incoming triphosphates during DNA replication. Here we present biochemical data and what are to our knowledge the first crystal structures of carbocyclic FaPydA and FaPydG containing DNA in complex with a high-fidelity polymerase. Crystallographic snapshots show that the cFaPy lesions keep the *anti* geometry of the glycosidic bond during error-free and error-prone replication. The observed dG·dC→dT·dA transversion mutations are the result of base shifting and tautomerization.

Oxidative DNA damage is a very frequent mutagenic event inside a cell responsible for spontaneous cancer development, inflammation and aging[1]. The bases for lesion formation are processes associated with aerobic respiration. These give rise to the formation of reactive oxygen species such as $O_2^{\cdot-}$, H_2O_2 and $OH^{\cdot}$, which attack the genetic material[2]. The purine nucleosides dG and dA are particularly vulnerable toward oxidative degradation owing to their comparably low oxidation potential[3–5]. Two major lesion types are formed (**Supplementary Results**, **Supplementary Fig. 1**). The two 8-oxo–type lesions 8-oxo-7,8-dihydro-2′-deoxyadenosine (8-oxodA) and 8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxodG) feature an oxidized C8 atom. In addition, the N7 atom carries a hydrogen atom, which changes this position from being hydrogen bond accepting to hydrogen bond donating. The second class of oxidative lesions are the ring-opened 4,6-diamino-5-formamidopyrimidine lesions FaPydA and FaPydG, which have a formamide group that has the potential to form a variety of different hydrogen bonds (**Fig. 1** and **Supplementary Fig. 1**).

Whereas the mechanisms behind the mutagenicity of the 8-oxo–type lesions are today quite well investigated biochemically[6,7] and structurally, with a variety of high-fidelity[6,8,9] and low-fidelity[10–14] polymerases, little is known about the mutation mechanisms associated with FaPy-type lesions[15]. Particularly, crystal structures that could provide insights into the mutation mechanisms are not available. Whereas the α-FaPy lesions are known to block high- and low-fidelity polymerases[16], the β-anomers do not stall replicative polymerases but reduce the replication efficiency and serve as templating bases with considerable mutagenic potential[17–19]. This seems to be responsible for their moderate toxicity[9,20]. Opposite FaPydA, the Klenow exo$^-$ fragment from *Escherichia coli* DNA polymerase I is known to misincorporate dA, leading to dA·dT→dT·dA transversion mutations[18]. In addition, the incorporation of dC opposite FaPydA was observed[17,21]. The oxidative dG-derived lesions are known to induce a plethora of mutations *in vivo*, of which dG·dC→dT·dA transversion mutations and dG·dC→dA·dT transition mutations are most frequently observed[22,23]. Primer extension studies with stabilized analogs of FaPydG[16,20] or methylated versions of FaPydG[21,24,25] showed either correct incorporation of dC or misincorporation of dA[19].

Previous crystal structures and biochemical data of the oxidized lesion 8-oxodG in complex with various polymerases clarified that this lesion can adopt either a *syn* or an *anti* conformation of the N9-C1′ (glycosidic) bond (**Fig. 1**)[6,8]. Whereas the *anti* conformation enables error-free bypass of the lesion by replicative polymerases, the *syn* conformation is responsible for the mutagenic effect[6,13]. This conformer is able to form a stable noncanonical Hoogsteen base pair with dA that induces only little distortion in the active site of the polymerase[6–8,10–14]. Although a similar mutation mechanism was proposed for the corresponding FaPy lesions[18,22], this has not been proven so far.

To enable biochemical and structural investigations of FaPy lesion–containing DNA, we used stabilized, bioisosteric FaPy lesion analogs in which the oxygen atom in the ribose is synthetically replaced by a methylene group (**Fig. 1** (X = CH_2) and **Supplementary Fig. 2**; details on synthesis of cFaPydA are in **Supplementary Note 1**)[16,20,26,27]. These carbocyclic analogs (cFaPydA and cFaPydG) feature the same hydrogen-bonding properties as the natural lesions but are anomerically stable during DNA synthesis and manipulation. This allows the preparation of carbocyclic β-FaPy lesion–containing oligonucleotides in sufficient quantity and excellent purity (>99%) as needed for crystallization studies. It is assumed that the naturally occurring FaPy lesions form initially as β-anomers but anomerize slowly to give a mixture of α- and β-anomers[28]. Although the equilibrium between α- and β-anomers may be affected upon binding of the lesion by the polymerases, which could increase the amount of the α-anomer, we report here crystallographic data of only the β-configured FaPy lesion in complex with the large fragment of the high-fidelity polymerase I from *Geobacillus stearothermophilus* (*Bst* Pol I). The crystal structures provide snapshot-type insights into the error-free and error-prone bypass reactions.

RESULTS

Chemical and biochemical studies

We first established that the carbocyclic analogs are suitable mimics of the natural lesions. To this end, we synthesized the carbocyclic version of the dG base (cdG; **Supplementary Note 1**) and compared it with dG with respect to the conformation of the glycosidic bond

[1]Center for Integrated Protein Science at the Department of Chemistry, Ludwig Maximilians University, Munich, Germany. [2]Present address: Department of Chemistry, Technical University, Munich, Germany. [3]These authors contributed equally to the work. *e-mail: thomas.carell@cup.uni-muenchen.de

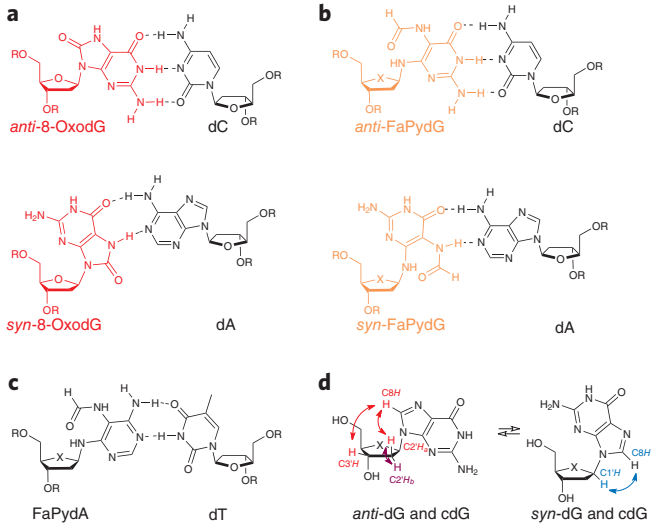


Figure 1 | Representation of the two main oxidation products of 2′-deoxyguanosine (8-oxodG, FaPydG) and FaPydA. (**a**) For 8-oxodG, the crystallographically observed 8-oxodG·dC and 8-oxodG·dA base pairs are shown. (**b**,**c**) Additionally, the putative base pairs formed by FaPydG and FaPydA are depicted. In the natural FaPy lesions, X = O, whereas in the stabilized carbocyclic analogs of FaPydG (**b**) and FaPydA (**c**), X = CH_2. (**d**) Depiction of the through-space interactions between C8*H* and the sugar protons in the *anti* conformer (red) and the *syn* conformer (blue). The reference NOESY signal between interactions was quantified, giving the same *syn*/*anti* distribution for both dG and cdG.

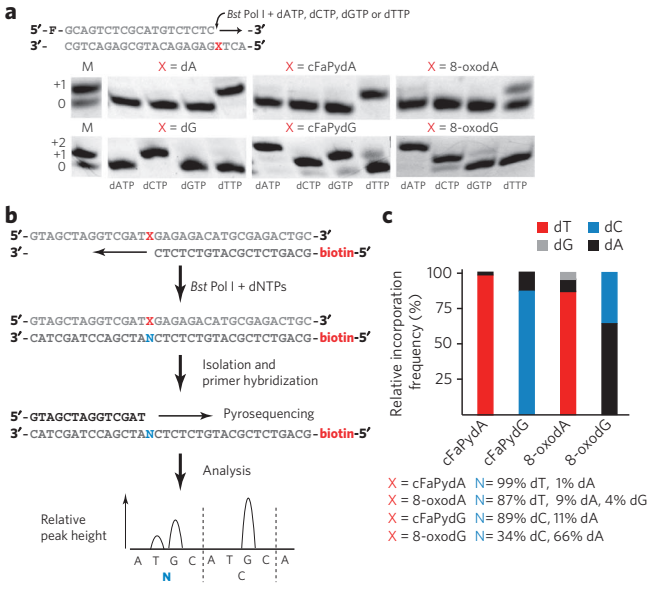


Figure 2 | Nucleotide insertion and bypass of oxidative lesions. (**a**) Single-nucleotide insertion reactions opposite cFaPydA and cFaPydG in comparison to those opposite of 8-oxodA and 8-oxodG. The primer was hybridized to a lesion-free control DNA or lesion (X)-containing template strand. F, fluorescein; M, marker. (**b**) Schematic presentation of the analysis of the misinsertion frequency and mutagenic signature. The biotinylated primer strands from the primer extension reaction with *Bst* Pol I and all four dNTPs were isolated and subjected to pyrosequencing analysis. (**c**) Relative frequency of correct and incorrect lesion bypass from the pyrosequencing experiment. The given percentage values are averages of three independent experiments.

using quantitative two-dimensional NOESY NMR experiments (**Fig. 1d** and **Supplementary Figs. 3** and **4**).

For both cdG and dG, we observed, in total, the same through-space interactions between C8*H* and the hydrogen atoms at C1′, C2′ and C3′. The C8*H*-C1*H*′ interaction is caused by the *syn* conformer, whereas the other interactions stem from the *anti* conformer. We quantified these interactions using the NOESY signal between C2′H_a and C2′H_b as internal standards and derived the same *syn*/*anti* ratio for dG and cdG, showing that the conformational equilibrium is unaffected by the exchange of O to CH_2 (**Supplementary Table 1**). This result was further secured by a theoretical study of cdG versus dG and also of cFaPydG versus FaPydG at the B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory. The obtained rotational profiles (**Supplementary Note 2**) around the N9-C1′ bonds were found to be similar, in agreement with the NMR data. We calculated that the positions of the conformational minima are the same and that the energetic positions of the barriers are influenced by the O to CH_2 replacement by only a small extent (≤8 kJ mol^{-1}).

With these results, we started biochemical studies. First, primer extension experiments were performed. The data shown in **Figure 2a** show that *Bst* Pol I[29], which we used in this study as a model for high-fidelity polymerases, is able to misincorporate nucleoside triphosphates into a primer strand opposite both cFaPydG and cFaPydA and that full extension of some primers finally occurs despite the conformational flexibility in DNA[30,31]. In single-nucleotide insertion studies, *Bst* Pol I was found to incorporate predominantly the correct dTTP opposite cFaPydA and 8-oxodA, enabling error-free bypass (**Fig. 2a** and **Supplementary Fig. 5**). Misincorporation is inefficient opposite both dA-derived lesions, showing that the misinstruction potential of cFaPydA and 8-oxodA is rather low (described below), which is in agreement with other reports and with kinetic data[18]. In contrast, for cFaPydG and 8-oxodG in addition to error-free bypass (insertion of dC), efficient misincorporation of dA is detected (**Fig. 2a**). With our primer sequence, this misincorporation results immediately in the +2 primer product, showing that extension of the cFaPydG·dA mispair is also possible. In addition, misincorporation of dG (the +1 product) is observed for cFaPydG. These data establish the cFaPydG lesion as a strongly misinstructing base, in agreement with a previous study[19].

To gain quantitative insight into the misincorporation events and, in particular, into the extension process that is often the limiting factor[24,25], we coupled a pyrosequencing step to the primer extension studies (**Fig. 2b**,**c**)[32]. In this assay, the fully extended primers are sequenced, giving quantitative data about the different bypass events. For the study, the cFaPy or 8-oxo lesions were inserted into a slightly longer DNA template strand (**Fig. 2b** and **Supplementary Table 2**). After hybridization to a biotinylated primer, primer extension was performed with the large fragment of *Bst* Pol I in the presence of all four canonical triphosphates. The elongated primers were extracted from the assay using streptavidin beads, annealed to a reverse primer and subsequently pyrosequenced[32,33]. The data shown in **Figure 2c** confirm that cFaPydA is indeed a relatively weak mutagenic lesion, as the lesion is correctly paired with dT with a frequency of 99%. In the case of 8-oxodA, replication across the lesion with *Bst* Pol I is more error-prone: in addition to correct dTTP insertion, 9% dATP and 4% dGTP are misincorporated as well. For both dG-derived lesions, we again detected dATP misincorporations to a large extent. For cFaPydG, dATP misincorporation was observed in 11% of all cases, whereas for 8-oxodG, misincorporation of dATP was the dominant process (66%). In the pyrosequencing data, no misincorporation of dG opposite cFaPydG was observed, showing that the cFaPydG·dG mispair cannot be extended by the polymerase.

To confirm the above results, we next measured steady-state kinetic data with *Bst* Pol I under standing start conditions. The data,

Table 1 | Kinetic parameters of dNTP insertion by *G. stearothermophilus* polymerase I opposite cFaPydA and cFaPydG

5′-ACTcFaPyGAGAGACATGCGAGACTGC-3′
3′ -CTCTCTGTACGCTCTGACG-5′

Incorporation	v_{max} (μM min^{-1})	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} μM^{-1})	f_{ins}[a]
dATP→cFaPydA	1.23×10^{-1}	359.63	2.76	7.66×10^{-3}	1.85×10^{-3}
dCTP→cFaPydA	6.49×10^{-2}	249.74	1.45	5.80×10^{-3}	1.40×10^{-3}
dGTP→cFaPydA	1.01×10^{-1}	815.96	2.25	2.76×10^{-3}	6.68×10^{-4}
dTTP→cFaPydA	9.80×10^{-2}	0.53	2.19	4.13	1.00
dATP→cFaPydG	5.43×10^{-2}	0.46	1.21	2.63	7.06×10^{-1}
dCTP→cFaPydG	6.25×10^{-2}	0.38	1.40	3.72	1.00
dGTP→cFaPydG	6.62×10^{-2}	39.87	1.48	3.71×10^{-2}	9.97×10^{-3}
dTTP→cFaPydG	4.07×10^{-2}	194.15	0.91	4.67×10^{-3}	1.26×10^{-3}

[a]$f_{ins} = (v_{max}/K_M)i / (v_{max}/K_M)r$, where r represents a correct base pair, and i represents an incorrect base pair. The kinetic studies were performed under standing start conditions, using 1 U *Bst* Pol I, 0.25 μM dsDNA and 0.1 nM–1 mM dNTPs. All of the reactions were carried out at 55 °C for 6 min.

shown in **Table 1** (**Supplementary Fig. 6**), support that *Bst* Pol I inserts a dTTP three to four orders of magnitude faster opposite cFaPydA than it inserts dATP, dCTP or dGTP. For cFaPydG, in contrast, dCTP was most efficiently incorporated, followed by incorporation of dATP (f_{ins} = 0.71). Incorporation of dGTP and dTTP was found to be two to three orders of magnitude less efficient. It was previously shown that the incorporation efficiency of dTTP opposite FaPydA is just 70%[18]. The incorporation of dCTP opposite FaPydG is approximately 50 times slower than formation of a dC:dG base pair[25]. These results show that the incorporation efficiencies of dTTP and dCTP opposite cFaPydA and cFaPydG are substantially reduced[9].

Error-free cFaPy bypass

In a previous crystal structure of the 8-oxodG lesion in complex with the large fragment of *Bst* Pol I (Protein Data Bank (PDB) code 1U49), Hoogsteen base pairing of the lesion with an incoming dATP in the post-insertion site (post-IS) was observed, explaining the mutagenic bypass reaction[6]. To investigate the mechanism behind error-free and error-prone replication through cFaPy lesions, we determined crystal structures of the large fragment of *Bst* Pol I in complex with DNA templates containing cFaPydA or cFaPydG. For the study, we either used prehybridized primer-template constructs or soaked the crystals in mother liquor containing different dNTPs. Because the polymerase is active in the crystal, as reported previously[34], this approach allows a direct, snapshot-like observation of the mutagenic and nonmutagenic primer extension reaction. Crystallographic data with 1.7- to 2.7-Å resolution for the error-free and error-prone translesion synthesis process through cFaPydA and cFaPydG were obtained. Data processing and structure refinement statistics are summarized in **Supplementary Tables 3** and **4**. Examples for the quality of the electron density for the structures are shown in **Supplementary Figures 7** and **8**. Overall, we found, on average, about 15% larger B-factors for the lesions in the complex structures, indicating higher flexibility of the formamidopyrimidines.

In the case of cFaPydA, we determined the structure of the lesion first in the (*n*+1) position inside the preinsertion site (pre-IS) of the polymerase (**Fig. 3a**,**b**). The lesion occupies the same position as an undamaged dA (PDB code 1L3S[34]). The polymerase is as expected in the open conformation. The O and O1 helices as well as Tyr714 occupy the same position as they do in a structure obtained with a lesion-free template strand. The lesion features the glycosidic bond in *anti* conformation. The residues Arg615, Gln797 and Tyr714 stabilize the last formed base pair in the template-primer duplex in the post-IS at position *n*–1, as reported previously[29].

Soaking the crystal in mother liquor containing dTTP allows the polymerase to perform the first primer extension reaction using the cFaPydA lesions as the templating base directly in the crystal. This structure is identical to a structure resulting from crystallization studies with a template-primer construct in which we synthetically preformed the cFaPydA·dT base pair (data not shown). The obtained structures show the cFaPydA lesion during error-free bypass in a Watson-Crick–type base-pairing situation with dT in the post-IS position (*n*–1; PDB code 4B9M) (**Fig. 3c**,**d**). A comparison with a structure featuring an intact dA·dT base pair (PDB code 1L3T[34]) at this position (**Supplementary Fig. 9a**) shows that cFaPydA·dT is less buckled (12°) but has a stronger propeller twist (–15°) than the cognate dA·dT base pair (buckle = 22°; propeller = –7.4°) in the polymerase *n*–1 position. The cFaPydA lesion is displaced by about 4.3 Å (C1′–C1′ distance; **Supplementary Fig. 9a**), which triggers repositioning of Tyr714 to optimize the π-stacking situation. The polymerase adopts an open-distorted conformation, which is accompanied by hinge movement of the O and O1 helices. This results in blocking of the pre-IS by the loop region between the O and O1 helix. The lesion-containing structure shows Tyr714 stacking with cFaPydA and dT. In contrast, in the reference crystal structure, which has no lesions, Tyr714 stabilizes only the template base dA. The O and O1 helices are shifted, preventing entry of a new templating base into the pre-IS. In the structure, cFaPydA forms a Watson-Crick–type base pair with dT with the formamide group oriented almost orthogonally to the heterocycle (**Supplementary Fig. 9a**). Despite the presence of the cFaPydA lesion, the primer strand and, in particular, the 3′ OH group of the primer end are in the correct position for primer extension. Asp830 forms the essential hydrogen bond with the 3′ OH primer end to position it for the coming reaction with the incoming triphosphate. This is most likely why the enzyme accomplishes error-free lesion bypass (insertion and extension) despite the structural distortion caused by the cFaPydA lesion.

To our surprise, we noted that the cFaPydA lesion also adopts a different conformation in the active site resulting from rotation around the C4-N9 bond. This is evident from the F_o-DF_c difference density obtained during structure refinement (**Supplementary Fig. 9b**). The occupancy adjustment of the conformations in the refinement process allowed us to estimate the abundance of the rotated state to about 5% of the molecules. This rotation establishes a relatively close distance between the N(6)H_2 of cFaPydA and N(3) of the T base (2.4 Å). The distance suggests the presence of an attractive hydrogen bond, which, however, requires one of the partners to switch into a different tautomeric state. Owing to the strong preference of dT for the keto form, despite the recent

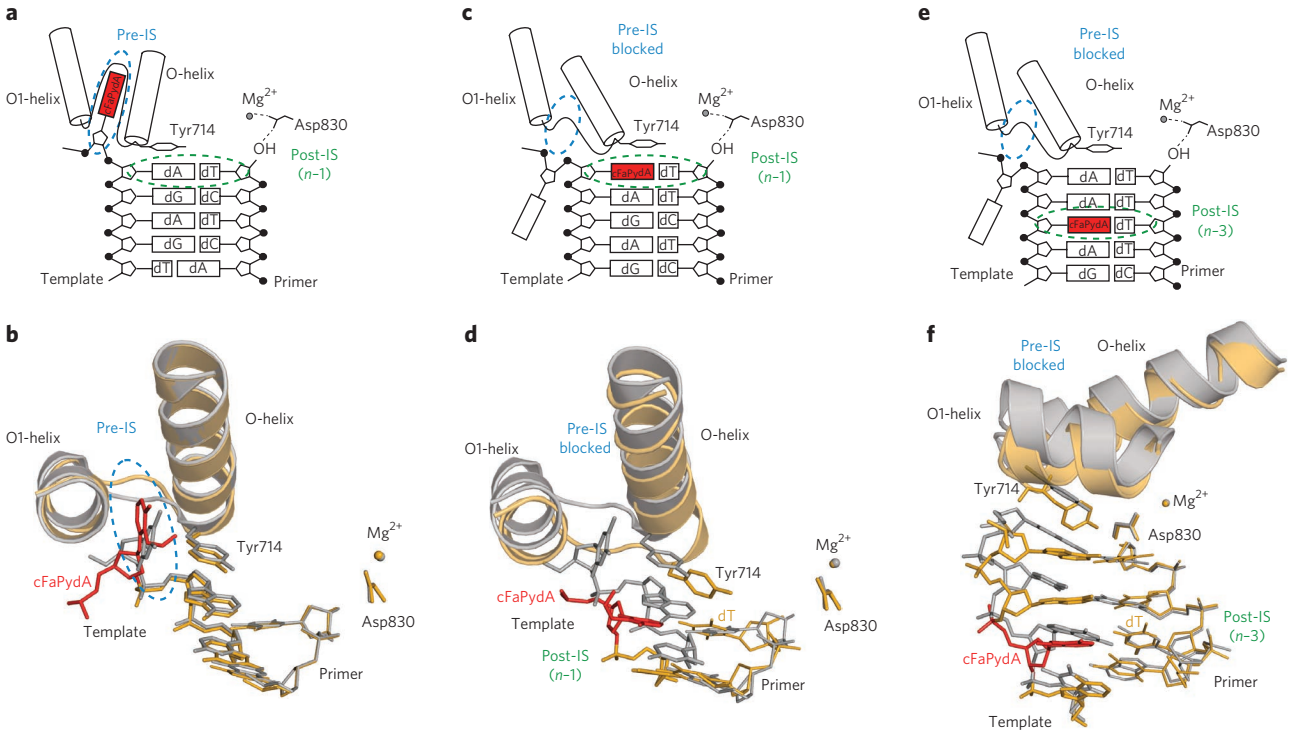


Figure 3 | Error-free reading of *Bst* Pol I through cFaPydA-containing template DNA. (**a**,**b**) Schematic view (**a**) and superposition of cFaPydA-DNA (red) in the pre-IS of the polymerase (gold; PDB code 4B9L) with the polymerase in complex with undamaged DNA (gray; PDB code 1L3S[34]) (**b**). (**c**–**f**) Structure after one (**c**,**d**; PDB code 4B9M) and three (**e**,**f**; PDB code 4B9N) rounds of correct template extension with the cFaPydA·dT base pair (red) at the post-IS (gold) *n*–1 and *n*–3, respectively. Overlaid in gray is the polymerase in complex with undamaged DNA (PDB codes 1L3T and 1L5U[34]).

proof of rare tautomers in polymerases[35,36], we believe that it is the lesion that is tautomerically more flexible.

Most importantly, the primer strand and the catalytic residues in the crystal structures remain in the correct position for the polymerase to extend the primer, despite the presence of cFaPydA. Indeed, when we used a template strand with a different sequence in which the cFaPydA lesion was followed by two further dA bases (oligodeoxynucleotide ODN4; **Supplementary Table 2**), soaking the crystals with dTTP caused elongation of the primer strand in the crystal by three dTTPs showing the cFaPydA·dT base pair (**Fig. 3e**,**f**) in the DNA duplex binding position *n*–3 with little residual distortion detectable in the DNA backbone (**Fig. 3e**,**f**). The formamide group is, most likely because of its flexibility, not defined in this structure.

We next analyzed the error-free bypass of the cFaPydG lesion. The crystal structure of *Bst* Pol I in complex with DNA containing the cFaPydG in the pre-IS is unexpected because it shows that the lesion resides outside the protein. Here, again, the pre-IS is blocked by the hinge movement of the O-helices (**Fig. 4a**,**b**), and Tyr714 is shifted to stack with the primer end without changing the position of the primer 3′ OH in the active site. The heterocycle of the cFaPydG lesion is only partially defined by the electron density, showing its flexible arrangement outside the pre-IS (**Supplementary Fig. 8**). To prove the presence of the lesion in the template strand in the absence of defining electron density, we reisolated the DNA from the crystal and confirmed the presence of the lesion by HPLC-MS (**Supplementary Table 2**) after total DNA digestion. In contrast, 8-oxodG does enter the pre-IS of the polymerase (**Fig. 4a**,**b**)[6].

Soaking the crystal with dCTP results in the formation of the cFaPydG·dC pair in the *n*–1 position with the glycosidic bond of the lesion, still in *anti* conformation, engaged in Watson-Crick base pairing with dC (**Fig. 4c**,**d** and **Supplementary Fig. 8b**). This shows that the whole structure with the lesion outside the pre-IS is sufficiently dynamic to enable loading of the cFaPydG lesion into the active site. In comparison, with the polymerase in complex with an intact dG·dC base pair (PDB code 1L5U[34]) at the same position, the lesion base pair shows a stronger propeller twist (–6.8°) than the dG·dC base pair (–2°) and no buckling (17.6° and 16°, respectively) of the base pair plane. Structural deviations associated with

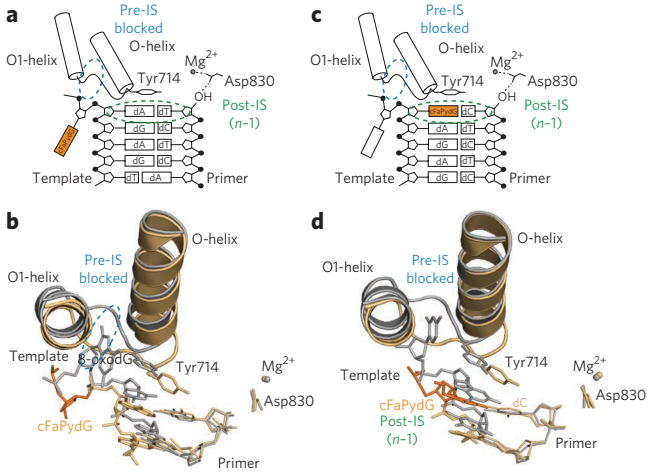


Figure 4 | Error-free reading of *Bst* Pol I through cFaPydG-containing template DNA. (**a**,**b**) Schematic view (**a**) and superposition of cFaPydG-DNA outside of the pre-IS (orange, yellow; PDB code 4B9S) with the polymerase in complex with 8-oxodG in the pre-IS (gray; PDB code 1U45 (ref. 6)) (**b**). (**c**,**d**) Schematic view (**c**) and structure after correct dCTP insertion with the cFaPydG·dC base pair (orange, yellow; PDB code 4B9T) in comparison with a dG·dC base pair (gray; PDB code 1L5U[34]) at the *n*–1 post-IS position (**d**).

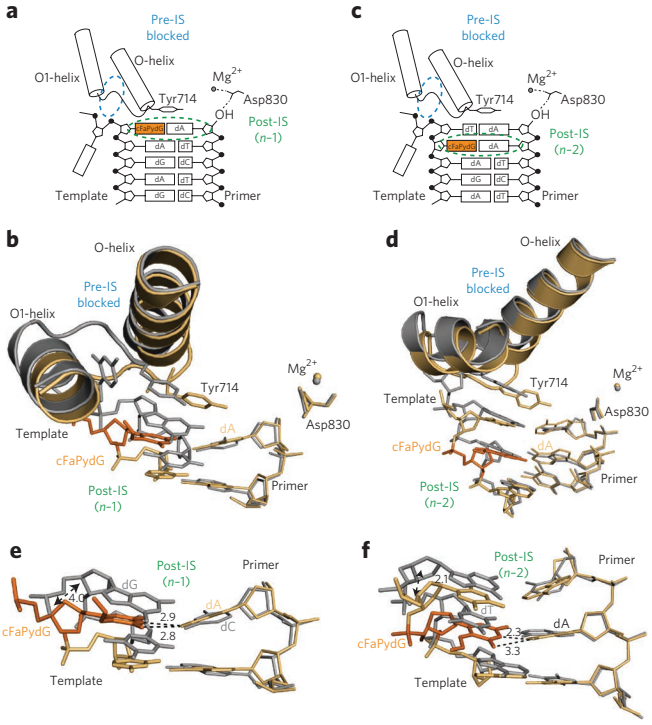


Figure 5 | Erroneous bypass and extension of cFaPydG. (**a**,**b**) Schematic view (**a**) and structure of *Bst* Pol I (yellow) in complex with cFaPydG (orange) after incorporation of a dATP in the crystal (PDB code 4B9U), superimposed with the polymerase in complex with the cognate dG·dC base pair (gray, PDB code 1L5U[34]) (**b**). (**c**,**d**) Schematic view (**c**) and structure after elongating the cFaPydG·dA base pair (orange) at *n*−2 (PDB code 4B9V) in comparison with structure involving undamaged DNA (gray; PDB code 1L5U) (**d**). (**e**,**f**) Side view of the base pairs. The cFaPydG·dA (orange, yellow) base pair is markedly buckled, the template strand is displaced, and the DNA is widened in comparison with the cognate base pair (gray) (**e**). After further strand elongation, the displacement is slightly relieved, but the cFaPydG·dA base pair remains out of planarity (**f**). Distances in **e** and **f** are given in Å.

the presence of the lesion are again limited to the template strand, which is shifted by about 3 Å. Again, Tyr714 is repositioned to optimize the stacking interactions with the newly formed cFaPydG·dC base pair. This in turn causes a shift of Gln797 (**Supplementary Fig. 10**). Only small differences are observed in the primer position, once again structurally explaining the biochemically observed efficient elongation of the primer after formation of the cFaPydG·dC base pair.

Error-prone bypass of cFaPydG

When we soaked crystals containing the cFaPydG lesion in the outer protein position in mother liquor containing dATP, we observed elongation of the primer and formation of the mutagenic cFaPydG·dA base pair in the crystal (**Fig. 5**). The cFaPydG lesion can obviously move into the active site from its outside position even when it involves formation of a 'purine'-purine mismatch. However, whereas the 8-oxodG lesion flips under these circumstances into the *syn* conformation to enable Hoogsteen base pairing with dA (PDB code 1U49[6]), the cFaPydG lesion forms the same base pair with the glycosidic bond remaining in *anti* conformation. The result is an interaction between cFaPydG and the dA counterbase that is stabilized by two hydrogen bonds (**Fig. 5a**,**b**). To form this *anti* cFaPydG·dA base pair, a 4.0-Å dislocation of the cFaPydG lesion is required (**Fig. 5e**), in contrast to a canonical dG·dC base pair at the same position. Again, this causes repositioning of Tyr714 and Gln797 to adjust the required stabilizing stacking interactions. The template strand is more strongly distorted in comparison to the cFaPydG·dC structure. The structural changes detected for the primer strand are surprisingly small in contrast, which again explains why the polymerase is able to extend a mutagenic cFaPydG·dA base pair. The strain transferred to the primer strand by the 'purine'-purine mispair cFaPydG·dA is relieved by turning the dA base by about 10° toward the major groove (**Fig. 5e** and **Supplementary Fig. 9e**). The whole base pair is substantially forced out of planarity and more strongly buckled (26°) than the respective undamaged dG·dC base pair (PDB code 1L5U[34]) inside the polymerase (16°). In addition, the propeller twist is more than twice as large for the damaged base (−5°) than for the dG·dC base pair (−2°). In comparison, a dG·dA mismatch, which cannot be extended by the polymerase, shows a much stronger distortion of the DNA strand, with a buckle of 38° and a propeller twist of −16° (ref. 37).

Soaking a crystal with dATP where the cFaPydG is followed by two dT in the template sequence results in a structure that shows extension of the cFaPydG·dA mispair to the position *n*−2 (**Fig. 5c**,**d**). All of the observed structural perturbations are now strongly reduced. The C1′-C1′ distance between the cFaPydG and a dG in the reference crystal structure is 2.1 Å (**Fig. 5f** and **Supplementary Fig. 9f**), demonstrating that once the lesion-containing base pair is formed and shifted away from the active site, further primer extensions are hard to avoid. In comparison with an analogous structure of a dG·dC base pair, the cFaPydG·dA (*n*−2) structure shows only small residual deviations.

DISCUSSION

The ability of cFaPydG to form an extendable base pair with an incoming dATP was expected from the biochemical data. However, it is unexpected and contradicts previous proposals that no Hoogsteen hydrogen bonding, enabled by rotation of the glycosidic bond from the canonical *anti* conformation into the *syn* geometry, is necessary. The *anti* cFaPydG·dA base pair is established without any involvement of the formamidopyrimidine substructure. Whereas for 8-oxodG, the C8-oxo group and the hydrogen-donating N7*H* functionality are essential to form the mutagenic 8-oxodG·dA base pair, the same atoms have no role in the case of the cFaPydG lesion. The observed *anti* cFaPydG·dA base pair consequently establishes a new mechanism for how the observed dG·dC→dT·dA transversion mutation, associated with oxidative damage, is formed. For 8-oxodG, rotation of the lesion into the *syn* conformation is essential to minimize steric problems associated with the close distance between the C8=O and the C5H_2-methylene groups of the deoxyribose in the canonical *anti* orientation[6]. The cFaPydG lesion solves this structural problem differently. The lesion turns the formamide group into an orthogonal position relative to the heterocycle and, in this way, minimizes the steric interaction between the C8=O and the C5H_2-methylene group. This reduces the need to shift the glycosidic bond from *anti* into the *syn* conformation. It is unexpected that the cFaPydG lesion can base pair with dA despite the *anti* conformation and that the cFaPydG·dA base pair is extended by the polymerase. This can be explained by the fact that the structural perturbations associated with the cFaPydG·dA base pair are limited to the template strand. Most important are the perturbations of the position of Tyr714 and Gln797, which give rise to a weakened hydrogen-bonding interaction between Gln797 and the cFaPydG·dA base pair; also, the stacking interaction of this mutagenic base pair with Tyr714 are compromised. The wider effects of these perturbations are, however, limited to the base pair in the *n*−2 position and not transmitted to the primer strand, which occupies the same position observed in structures with lesion-free template–primer complexes. The inability of the cFaPydG lesion to flip into the *syn* conformation clearly causes stronger structural

perturbations in comparison to 8-oxodG, which explains why mutagenic formation of the cFaPydG·dA base pair is reduced to 11% in comparison to 8-oxodG·dA, where it is 66%.

The lack of rotation influences, in particular, the planarity of the base pair. Such a deviation does not influence the strength of the hydrogen bonds but reduces the strength of stacking interactions with the other base pairs and with Tyr714. It seems that the ring-opened structure of the cFaPydG lesion, with its considerably smaller π-surface, decreases the energetic penalty for reduced stacking. A reduced aromatic character, in turn, may also allow the complex cFaPy lesion to shift more easily into other tautomeric states in comparison to the canonical bases. This is crystallographically observed for cFaPydA, but it is likely that similar tautomeric shifts are possible for cFaPydG as well.

The main difference between 8-oxo–type lesions and cFaPy-type lesions are therefore that the 8-oxo group can rotate as part of the formamide unit away from the $C5H_2$ group into an orthogonal position relative to the heterocycle. This gives the cFaPydG lesion more structural flexibility, which reduces the need for the lesion to rotate into the *syn* conformation. In addition, the lacking imidazole substructure reduces the energetic penalty associated with reduced stacking interactions, which makes it easier for cFaPy-containing base pairs to deviate from planarity. Although shown here only for cFaPydA, we think that cFaPy lesions can, in general, change more easily into different tautomeric states in comparison to natural purines, broadening their base-pairing properties without the need to use the formamide or the Hoogsteen base pairs. What is also interesting is the observation that cFaPydG·dG mispairs stall a replicative polymerase, whereas the cFaPydG·dA and 8-oxodG·dA mispairs, which are primarily formed, are extended to form full-length primers (**Fig. 2a**,**c**). This seems to be the reason why cFaPydG·dA and 8-oxodG·dA mismatches escape recognition by the repair system, which is triggered by a stalled replicative polymerase during replication. We would like to emphasize that our mechanistic studies show how a high-fidelity polymerase replicates cFaPydG[5] and that mutations arise. The *in vivo* mutagenicity of the bases depends of course also on other factors, such as repair or the action of other polymerases such as low-fidelity polymerases, which can explain different cell type–specific *in vivo* mutagenicities[9].

Received 1 October 2012; accepted 15 April 2013;
published online 19 May 2013

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. PDB: Atomic coordinates and structure factors of the protein in complex with DNA have been deposited with the following accession codes: cFaPydA, pre-IS (4B9L); cFaPydA·dT, *n*–1 (4B9M); cFaPydA·dT, *n*–3 (4B9N); cFaPydG, outside the pre-IS (4B9S); cFaPydG·dC, *n*–1 (4B9T); cFaPydG·dA, *n*–1 (4B9U); cFaPydG·dA, *n*–2 (4B9V). Cambridge Structural Database: Small-molecular crystallographic data have been submitted under the following accession codes: **9a** (CCDC-796512); **11** (CCDC-796513).

References

1. Finkel, T. & Holbrook, N.J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
2. Cadet, J., Douki, T. & Ravanat, J.L. Oxidatively generated base damage to cellular DNA. *Free Radic. Biol. Med.* **49**, 9–21 (2010).
3. Steenken, S. & Jovanovic, S.V. How easily oxidizible is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* **119**, 617–618 (1997).
4. Cadet, J., Douki, T. & Ravanat, J.L. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspects and formation in cells. *Acc. Chem. Res.* **41**, 1075–1083 (2008).
5. Jena, N.R. & Mishra, P.C. Formation of ring-opened and rearranged products of guanine: mechanisms and biological significance. *Free Radic. Biol. Med.* **53**, 81–94 (2012).
6. Hsu, G.W., Ober, M., Carell, T. & Beese, L.S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **431**, 217–221 (2004).
7. Duarte, V., Muller, J.G. & Burrows, C.J. Insertion of dGMP and dAMP during *in vitro* DNA synthesis opposite an oxidized form of 7,8-dihydro-8-oxoguanine. *Nucleic Acids Res.* **27**, 496–502 (1999).
8. Brieba, L.G. *et al.* Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J.* **23**, 3452–3461 (2004).
9. Greenberg, M.M. The formamidopyrimidines: purine lesions formed in competition with 8-oxopurines from oxidative stress. *Acc. Chem. Res.* **45**, 588–597 (2012).
10. Brieba, L.G. & Ellenberger, T. Hold tight (but not too tight) to get it right: accurate bypass of an 8-oxoguanine lesion by DNA polymerase β. *Structure* **11**, 1–2 (2003).
11. Krahn, J.M., Beard, W.A., Miller, H., Grollman, A.P. & Wilson, S.H. Structure of DNA polymerase β with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. *Structure* **11**, 121–127 (2003).
12. Eoff, R.L., Irimia, A., Angel, K.C., Egli, M. & Guengerich, F.P. Hydrogen bonding of 7,8-dihydro-8-oxodeoxyguanosine with a charged residue in the little finger domain determines miscoding events in *Sulfolobus solfataricus* DNA polymerase Dpo4. *J. Biol. Chem.* **282**, 19831–19843 (2007).
13. Vasquez-Del Carpio, R. *et al.* Structure of human DNA polymerase κ inserting dATP opposite an 8-OxoG DNA lesion. *PLoS ONE* **4**, e5766 (2009).
14. Batra, V.K. *et al.* Mutagenic conformation of 8-oxo-7,8-dihydro-2′-dGTP in the confines of a DNA polymerase active site. *Nat. Struct. Mol. Biol.* **17**, 889–890 (2010).
15. Dizdaroglu, M., Kirkali, G. & Jaruga, P. Formamidopyrimidines in DNA: mechanisms of formation, repair, and biological effects. *Free Radic. Biol. Med.* **45**, 1610–1621 (2008).
16. Büsch, F. *et al.* Dissecting the differences between the α and β anomers of the oxidative DNA lesion FaPydG. *Chemistry* **14**, 2125–2132 (2008).
17. Graziewicz, M.A. *et al.* Fapyadenine is a moderately efficient chain terminator for prokaryotic DNA polymerases. *Free Radic. Biol. Med.* **28**, 75–83 (2000).
18. Delaney, M.O., Wiederholt, C.J. & Greenberg, M.M. Fapy.dA induces nucleotide misincorporation translesionally by a DNA polymerase. *Angew. Chem. Int. Ed. Engl.* **41**, 771–773 (2002).
19. Wiederholt, C.J. & Greenberg, M.M. Fapy.dG instructs Klenow exo$^-$ to misincorporate deoxyadenosine. *J. Am. Chem. Soc.* **124**, 7278–7279 (2002).
20. Ober, M., Müller, H., Pieck, C., Gierlich, J. & Carell, T. Base pairing and replicative processing of the formamidopyrimidine-dG DNA lesion. *J. Am. Chem. Soc.* **127**, 18143–18149 (2005).
21. Tudek, B. *et al.* Mutagenic specificity of imidazole ring-opened 7-methylpurines in M13mp18 phage DNA. *Acta Biochim. Pol.* **46**, 785–799 (1999).
22. Kalam, M.A. *et al.* Genetic effects of oxidative DNA damages: comparative mutagenesis of the imidazole ring-opened formamidopyrimidines (Fapy lesions) and 8-oxo-purines in simian kidney cells. *Nucleic Acids Res.* **34**, 2305–2315 (2006).
23. Patro, J.N. *et al.* Studies on the replication of the ring opened formamidopyrimidine, Fapy.dG in *Escherichia coli*. *Biochemistry* **46**, 10202–10212 (2007).
24. Asagoshi, K., Terato, H., Ohyama, Y. & Ide, H. Effects of a guanine-derived formamidopyrimidine lesion on DNA replication: translesion DNA synthesis, nucleotide insertion, and extension kinetics. *J. Biol. Chem.* **277**, 14589–14597 (2002).
25. Christov, P.P., Angel, K.C., Guengerich, F.P. & Rizzo, C.J. Replication past the N5-methyl-formamidopyrimidine lesion of deoxyguanosine by DNA polymerases and an improved procedure for sequence analysis of *in vitro* bypass products by mass spectrometry. *Chem. Res. Toxicol.* **22**, 1086–1095 (2009).
26. Ober, M., Linne, U., Gierlich, J. & Carell, T. The two main DNA lesions 8-oxo-7,8-dihydroguanine and 2,6-diamino-5-formamido-4-hydroxypyrimidine exhibit strongly different pairing properties. *Angew. Chem. Int. Ed. Engl.* **42**, 4947–4951 (2003).
27. Ober, M., Marsch, M., Harms, K. & Carell, T. A carbocyclic analogue of a protected β-d-2-deoxyribosylamine. *Acta Crystallogr. Sect. E Struct. Rep. Online* **60**, o1191–o1192 (2004).
28. Patro, J.N., Haraguchi, K., Delaney, M.O. & Greenberg, M.M. Probing the configurations of formamidopyrimidine lesions Fapy·dA and Fapy·dG in DNA using endonuclease IV. *Biochemistry* **43**, 13397–13403 (2004).
29. Kiefer, J.R. *et al.* Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95–108 (1997).

30. Raoul, S., Bardet, M. & Cadet, J. Gamma irradiation of 2′-deoxyadenosine in oxygen-free aqueous solutions: identification and conformational features of formamidopyrimidine nucleoside derivatives. *Chem. Res. Toxicol.* **8**, 924–933 (1995).
31. Lukin, M. *et al.* Novel post-synthetic generation, isomeric resolution, and characterization of Fapy-dG within oligodeoxynucleotides: differential anomeric impacts on DNA duplex properties. *Nucleic Acids Res.* **39**, 5776–5789 (2011).
32. Münzel, M. *et al.* Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* **17**, 13782–13788 (2011).
33. Stathis, D., Lischke, U., Koch, S.C., Deiml, C.A. & Carell, T. Discovery and mutagenicity of a guanidinoformimine lesion as a new intermediate of the oxidative deoxyguanosine degradation pathway. *J. Am. Chem. Soc.* **134**, 4925–4930 (2012).
34. Johnson, S.J., Taylor, J.S. & Beese, L.S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl. Acad. Sci. USA* **100**, 3895–3900 (2003).
35. Topal, M.D. & Fresco, J.R. Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285–289 (1976).
36. Wang, W., Hellinga, H.W. & Beese, L.S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl. Acad. Sci. USA* **108**, 17644–17648 (2011).
37. Johnson, S.J. & Beese, L.S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116**, 803–816 (2004).

Acknowledgments

We are grateful to the beamline scientists at the Swiss Light Source and European Synchrotron Radiation Facility for setting up the beamlines. This research project was supported by the Deutsche Forschungsgemeinschaft through SFB 646 and SFB 749. Further support was obtained by the Volkswagen Foundation and in particular by the Excellence Cluster CiPS[M]. We thank K. Karaghiosoff and K. Lux for solving the crystal structures of the small molecules. We thank M. Müller for critical reading of the manuscript and many helpful discussions.

Author contributions

T.C. conceived and directed the study. He wrote the manuscript and designed experiments. T.H.G. and U.L. designed experiments. T.H.G. performed the synthesis of the lesions and of the DNA strands. U.L. and T.H.G. performed the biochemical experiments. U.L. purified the protein. K.L.G. performed the synthesis of cdG. S.A. developed the synthesis of cFaPydA. H.C.M. developed the synthesis of cdG. S.S. conducted crystallographic data collection and solved the crystal structures. H.Z. performed the theoretical studies. D.S.S. performed the NMR studies.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to T.C.

ONLINE METHODS

General experimental procedures. All reagents and solvents were purchased from Acros, Fischer Scientific (KMF Laborchemie Handels GmbH), Fluka (Sigma-Aldrich-Chemie GmbH) and were used as delivered. All of the experiments involving water-sensitive compounds were performed in oven-dried glassware under a nitrogen atmosphere. Reactions were monitored by analytical thin-layer chromatography (TLC) on VWR precoated aluminum plates 60 F_{254}, visualized by UV light, anisaldehyde or ninhydrin staining. Melting points were obtained in open-glass capillaries with a Büchi Smp 20 melting point apparatus and are not corrected. IR spectra of the compounds were recorded on a Perkin-Elmer FT-IR Spektrum 100 on an ATR unit. ^{1}H NMR spectra were obtained on the following spectrometers: Bruker AMX-200, ARX-300, DRX-500 and ARX-600 or Varian Mercury 200VX and VXR400S. The chemical shifts were referenced to $CHCl_3$ (δ 7.26) in $CDCl_3$, DMSO (δ 2.50) in DMSO–d_6, MeOH (δ 3.31) in CD_3OD–d_4, H_2O (δ 4.80) in D_2O and CH_3CN (δ 1.94) in CD_3CN–d_3. If necessary, peak assignment was carried out with the help of COSY, HMBC, HMQC or NOESY experiments. ^{13}C spectra were obtained on a Bruker DRX-200 (50 MHz), ARX-300 (75 MHz), DRX-500 (125 MHz) or Varian VXR400S (100 MHz) spectrometer and referenced to the solvent signal. ^{31}P spectra were obtained on the spectrometer Varian Mercury 200VX (81 MHz). ^{31}P chemical shifts are quoted in p.p.m. using 85% H_3PO_4 as an external standard. EI mass spectra were recorded on a Varian MAT CH7A or MAT90 (Varian) mass spectrometer. FAB spectra in low and high resolution were recorded on a Finnigan MAT95 mass spectrometer. ESI spectra and high-resolution ESI spectra were obtained on the mass spectrometers Finnigan MAT95S, JEOL JMS-700 or PE ScieX API QStar Pulsar i. FT-ICR-ESI mass spectra were measured on a Thermo Finnigan LTQ Orbitrap XL FTMS. Final concentrations of the purified oligonucleotides and enzyme concentrations were determined with a NanoDrop ND-1000 spectrophotometer (Peqlab). MALDI-TOF mass spectra of the oligonucleotides were confirmed using a Bruker Autoflex II spectrometer. DNA-melting experiments to determine curves were performed using a Carey 100 Bio-spectrophotometer equipped with Carey temperature controller, sample transport accessory and multi cell block (Varian). Commercial oligonucleotides were purchased from Metabion or Ella Biotech (Martinsried, Germany).

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an Expedite 8909 Nucleic Acid Synthesis System (PerSeptive Biosystems) using standard DNA synthesis conditions. Phosphoramidites for dA, dC, dG, dT and CPG carriers were obtained from Amersham, Glen Research or PE Biosystems. Oligonucleotides containing the cFaPydA building block were not compatible with the regular capping procedure. For these strands, the synthesis protocol was modified. The coupling time for **1** was extended to 2 × 7 min. The standard capping solution was replaced by a mixture of pivaloyl anhydride or butyric anhydride (0.54 M) and 2,6-lutidine (10.5%) in acetonitrile. The mild capping period was 25 s after incorporation of cFaPydA and each of the following bases. In addition, deblocking cycles were shortened using dichloroacetic acid to avoid detritylation of the N6 position of the lesion. Also, the coupling time for the first following phosphoramidite after the lesion was extended to 144 s. The terminal DMT-protecting group was removed from the oligonucleotides (see deprotection and purification).

Deprotection and purification. Deprotection and cleavage of the oligodeoxynucleotides from the CPG carrier containing cFaPydA were carried out in a mixture of saturated ammonia solution in water (7 M) and ethanol (3:1) at 17 °C overnight. DNA purification was conducted on analytical and preparative HPLC (Waters) using Nucleodur or Nucleosil columns (250 × 4 mm, C18ec, particle size 3 µm or 250 × 10 mm, C18ec, particle size 5 µm) from Machery-Nagel. The applied buffer was 0.1 M triethylammonium acetate in water and 0.1 M triethylammonium acetate in an 80% aqueous MeCN buffer system. The fractions were checked for purity by analytical HPLC. The purified oligonucleotides were concentrated *in vacuo* using a Savant Speed Vac and were desalted with Sep-Pak cartridges (Waters) before use. The oligodeoxynucleotides still containing the trityl group were deprotected by addition of 100 µL of an 80% acetic acid solution. After incubation for 20 min at room temperature, 100 µL of water together with 60 µL of a 3-M solution of sodium acetate were added. To this solution, 1.6 mL of ethanol were added, and the DNA was precipitated at least for 60 min at −20 °C. The DNA was isolated, decanting off the solution, and lyophilized, and samples were purified via HPLC as mentioned above.

Alternatively, the DNA was detritylated via SepPac cartridges. The cartridges were activated with acetonitrile (10 mL) and equilibrated with distilled water (10 mL). The lyophilized oligonucleotides were loaded onto the cartridge (0.2 mL/min). The column was washed with distilled water (6 mL), and then 5 mL trifluoroacetic acid (0.4%) in water was passed through it (during 90 s), followed immediately by 6 mL of 0.1 M triethylammonium acetate in water (pH 7) and then with 6 mL of water. Finally, the DNA was eluted two times each with 4 mL acetonitrile (80%) in water.

Enzymatic digestion. The DNA (2 nmol in 100 µL H_2O) was incubated with 10 µL buffer A (300 mM ammonium acetate, 100 mM $CaCl_2$, 1 mM $ZnSO_4$, pH 5.7) and 20 U nuclease S1 (*Aspergillus oryzae* from Roche) for 3 h at 37 °C. After adding 12 µL buffer B (500 mM Tris-HCl, 1 mM EDTA, pH 8.0), 5 U antarctic phosphatase (New England Biolabs) and 0.1 unit snake venom phosphodiesterase I (*Crotalus adamanteus* venom from USB corporation), the solution was incubated for a further 3 h at 37 °C until the digestion was completed. The sample was centrifuged (6,000 r.p.m., 10 min) and analyzed by HPLC and FT-ICR-HPLC/MS. For the analysis of the enzymatic digestion, a 3 HDO column (150 × 2.0 mm) from Interchim Interchrom Uptisphere was used. Eluting 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)). Gradients were as follows: 0–12 min; 0%–3% B; 12–60 min; 3%–60% B; 60–62 min; 60%–100% B; 62–90 min; 100% B; 90–95 min; 100% B–0% B; 95–130 min; 0% B, with a flow of 0.15 mL/min. The elution was monitored at 260 nm and 325 nm.

Single-nucleotide insertion. For the single-nucleotide insertion a 16-mer 5′-fluorescein–labeled primer was hybridized to either the lesion-containing template strand or the lesion-free control. The experiment was conducted by using 0.5 µM primer (ODN2-Primer) and 1 µM template (ODN2) to ensure all of the primer was fully hybridized. Together with *Bst* Pol I (1 U, NEB), 5 mM Mg^{2+} and 50 µM each of the individual dNTPs, each duplex was incubated for 6 min at 55 °C. The reactions were terminated by adding 20 µL TBE-urea sample buffer and analyzed by denaturating PAGE.

Steady-state kinetics. The primer extension was performed following the same protocol as for single-nucleotide insertion studies, except that the dNTP concentration was varied from 0.5 nM to 1 mM. Kinetic constants were derived as described in ref. 38.

Primer extension–based pyrosequencing. In a ratio of 2:1, a 34-mer template DNA strand (ODN1) was annealed to a 5′-biotinylated 19-mer primer strand (ODN1–Primer1, Metabion), ending one base 3′ to the lesion. Hence, the primer extension experiments were performed under standing start conditions with *Bst* Pol I. *Bst* Pol I (1 U; NEB), 5 mM $MgSO_4$, 50 µM dNTPs and 0.5 µM dsDNA were incubated in a total volume of 20 µl 1× ThermoPol Reaction Buffer (NEB) for 30 min at 55 °C. Subsequently, 2 µl of streptavidin Sepharose beads (GE Healthcare, Uppsala, Sweden), 40 µl Binding Buffer (PyroMark, Qiagen) and 18 µl ddH2O were added to each sample. After agitation for 15 min at 1,400 r.p.m., the beads were captured with a Vacuum Prep Tool (Qiagen) and washed with 70% ethanol, 0.2 M NaOH and washing buffer (Qiagen). The elongated biotinylated primers (immobilized on the bead surfaces) were annealed to 10 pmol of the sequencing primer (ODN1–Primer2, Metabion) in 30 µL annealing buffer (Qiagen). Pyrosequencing was performed on a PyroMark Q24 Pyrosequencer using standard conditions (Qiagen). For each dNTP, blanks were measured. The data were analyzed using the provided software, PyroMark Q24. The average values of the peak heights were calculated using Microsoft Excel, and the average blank values were subtracted. Hence, relative incorporation at every variable position was calculated and averaged from three independent experiments.

Protein crystallization. *Bst* Pol I was expressed and purified as previously described[39]. For co-crystallization, the lesion-containing template (ODN3–ODN6) was annealed to the primer ending one base 3′ to the lesion in 10 mM Na-cacodylate, pH 6.5, 50 mM NaCl, 0.5 mM EDTA and 10 mM $MgSO_4$. Prior to crystallization, protein and DNA were mixed in a 1:3 molar ratio, resulting in final concentrations of 5 mg/ml and 0.5 mM for *Bst* Pol I and dsDNA, respectively. Crystals were grown by mixing an equal volume of protein–DNA complex with 47–51% ammonium sulfate, 3.0–4.1% MPD and 100 mM MES, pH 5.8, using the hanging-drop vapor diffusion method. The crystallization plates were incubated at 18 °C, and crystals appeared after 2 to 6 d. For primer extensions,

doi:10.1038/nchembio.1254

the crystals were soaked overnight (single extension) or for 2–3 d (double and triple extension) in an artificial mother liquor supplement with 30 mM of the respective triphosphate. Crystals were frozen in 24% sucrose, 55% ammonium sulfate, 3.5% MPD, 100 mM MES, pH 5.8, and stored in liquid nitrogen until data collection. Data were collected at the beamlines ID14-1 (European Synchrotron Radiation Facility (ESRF), Grenoble, France) and PX I and PXIII (Swiss Light Source (SLS), Villigen, Switzerland) and processed with the programs XDS[40] and SCALA[41,42]. Structure solution was carried out by molecular replacement with PHASER[43] using the coordinates of PDB code 1U45 (ref. 6). To reduce model bias, before model building in COOT[44], the temperature factors were reset to the Wilson B factor, and simulated annealing omit maps, removing the area around the lesion, were calculated with PHENIX[45]. TLS and restrained refinement were carried out in REFMAC5[46,47] and PHENIX. All of the structural superpositions were done with LSQKAB[41], and structural figures were prepared with Pymol (Delano Scientific). Data processing and refinement statistics are summarized in **Supplementary Tables 3** and **4**. The DNA geometry was analyzed with CURVES+[48,49].

38. Creighton, S., Bloom, L.B. & Goodman, M.F. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.* **262**, 232–256 (1995).
39. Münzel, M., Lercher, L., Müller, M. & Carell, T. Chemical discrimination between dC and 5MedC via their hydroxylamine adducts. *Nucleic Acids Res.* **38**, e192 (2010).
40. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
41. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
42. Evans, P. Joint CCP4 and ESF-EACMB. *Newsletter Prot. Crystallogr.* **33**, 22–24 (1997).
43. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
44. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
45. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
46. Winn, M.D., Murshudov, G.N. & Papiz, M.Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
47. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
48. Olson, W.K. *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237 (2001).
49. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929 (2009).

Supplementary Information

Unexpected non-Hoogsteen–based mutagenicity mechanism of FaPy-DNA lesions

Tim H. Gehrke,[1, #] Ulrike Lischke,[1, #] Karola L. Gasteiger,[1, #] Sabine Schneider,[1,2,#] Simone Arnold,[1] Heiko C. Müller,[1] David S. Stephenson,[1] Hendrik Zipse,[1] and Thomas Carell[1,*]

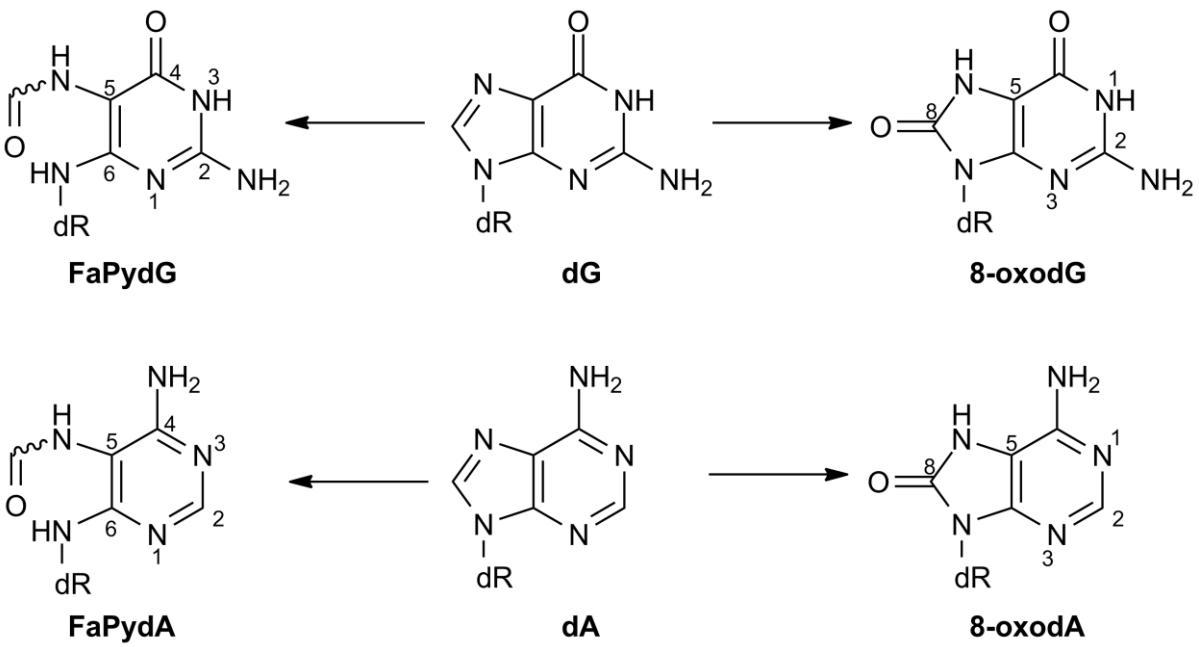
*Corresponding author: Thomas Carell, Center for Integrated Protein Science CIPS[M] at the Department of Chemistry, Ludwig-Maximilians-University Munich, Butenandtstr. 5-13 (F), D-81377 Munich, Germany Tel.: +49 (0)89 2180 77750; Fax: +49 (0)89 2180 77756, Email: thomas.carell@cup.uni-muenchen.de

# These authors contributed equally to the work.

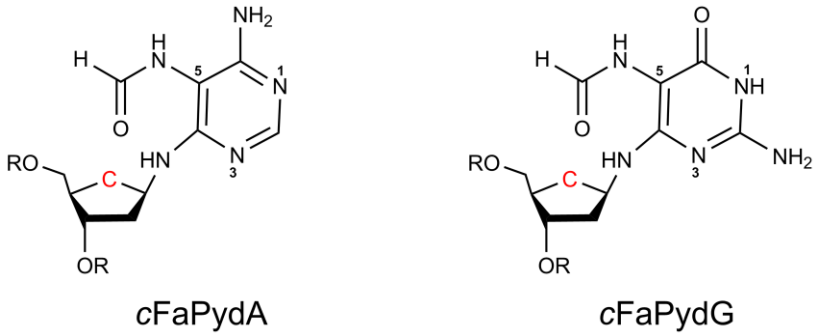
[1]Department of Chemistry, Ludwig-Maximilians-University, Butenandtstr. 5-13 (F), D-81377 Munich (Germany)

[2]Present address: Department of Chemistry, Technical-University, Lichtenbergstr. 4, D-85747 Garching, Munich (Germany)

Supplementary Results

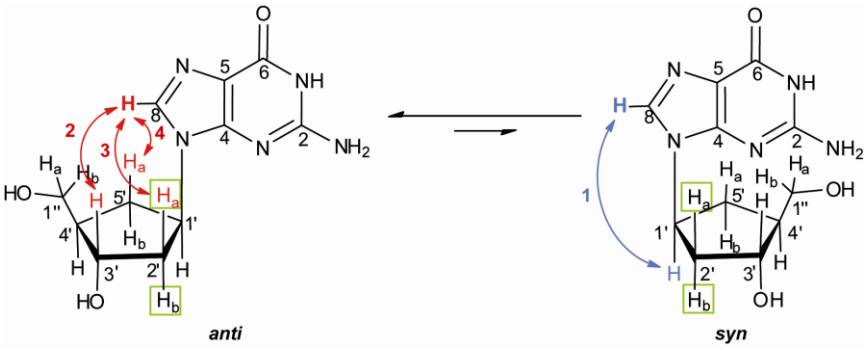


Supplementary Figure 1: The two main oxidation products of 2´-deoxyguanosine (dG), 8-oxodG and FaPydG and their counterparts from 2´-deoxyadenosine (dA), 8-oxodA and FaPydA, respectively (dR = 2´deoxribose).

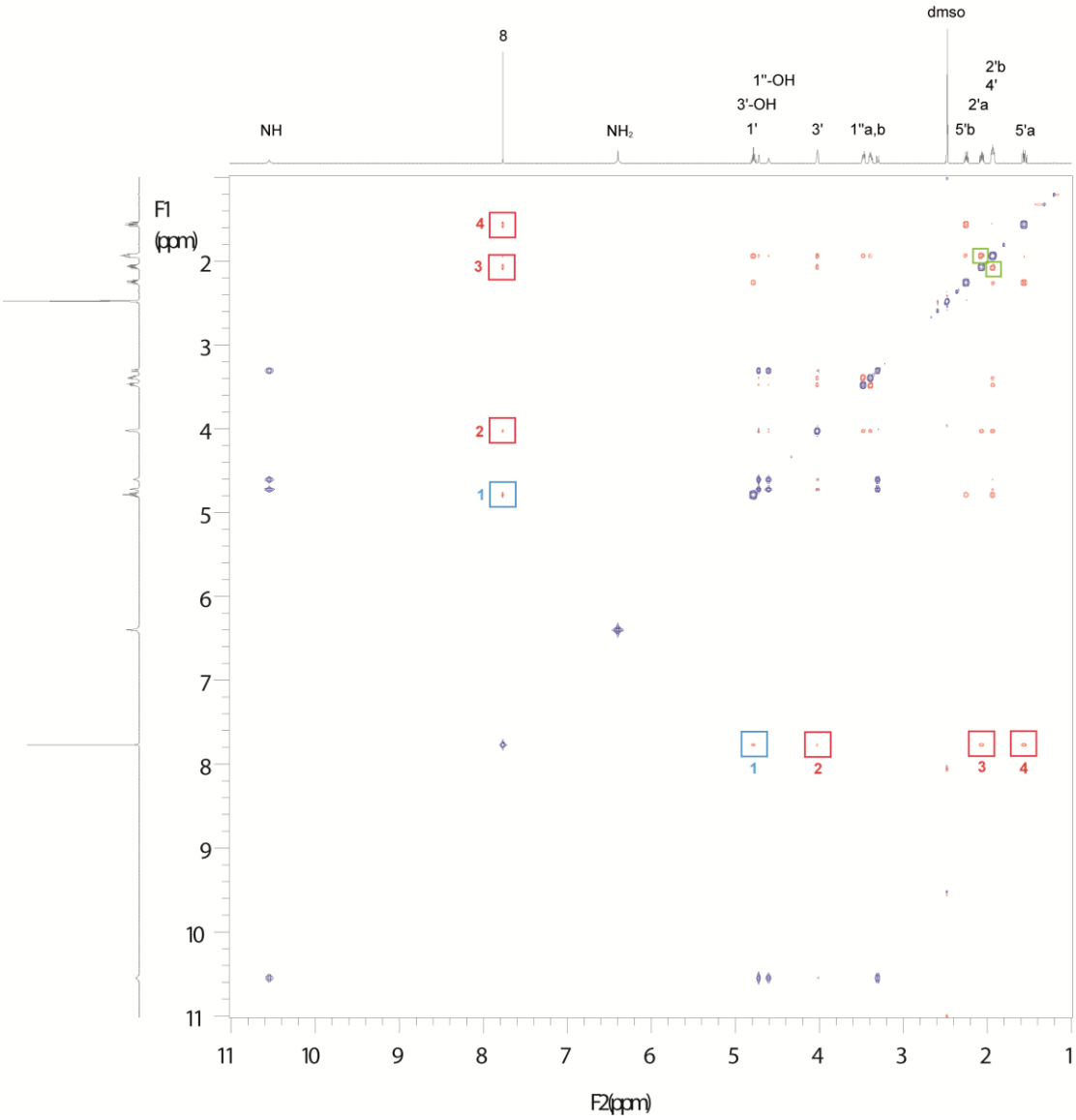


Supplementary Figure 2: Depiction of the carbocyclic analogs of FaPydA and FaPydG used for this study.

a

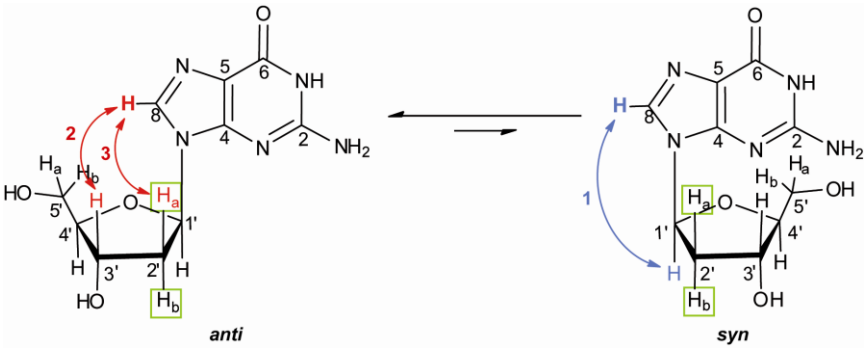


b

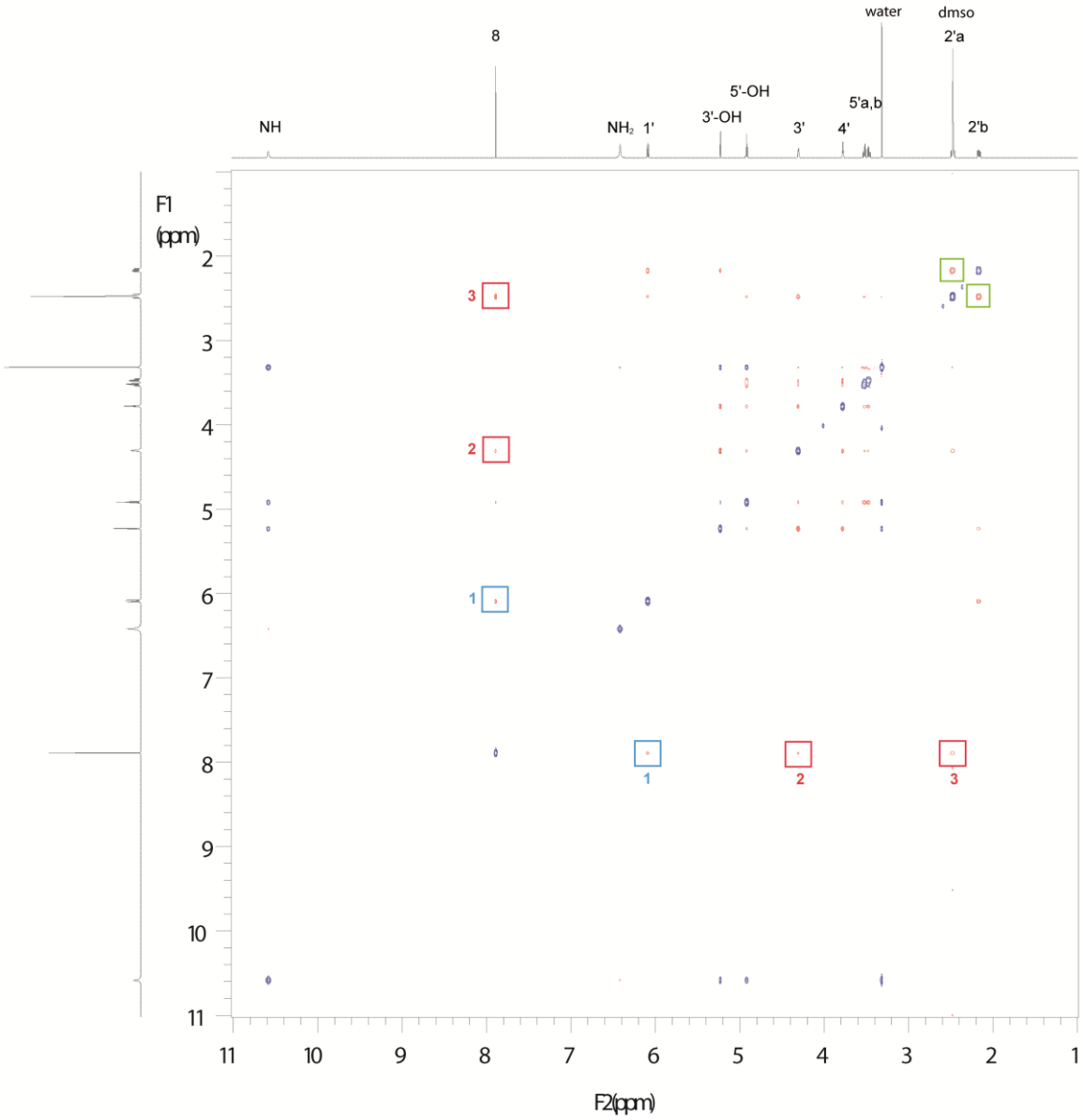


Supplementary Figure 3: Noesy (600 MHz, DMSO–d_6) of carbocyclic dG **16**. The interactions between the C8*H* and the C3´*H*, C2´H_a and C5´H_a of the *anti*-*c*dG are shown in red and the interactions between the C8*H* and the C1´H of the *syn*-*c*dG are shown in blue. The reference interactions of C2´H_a and C2´H_b are shown in green.

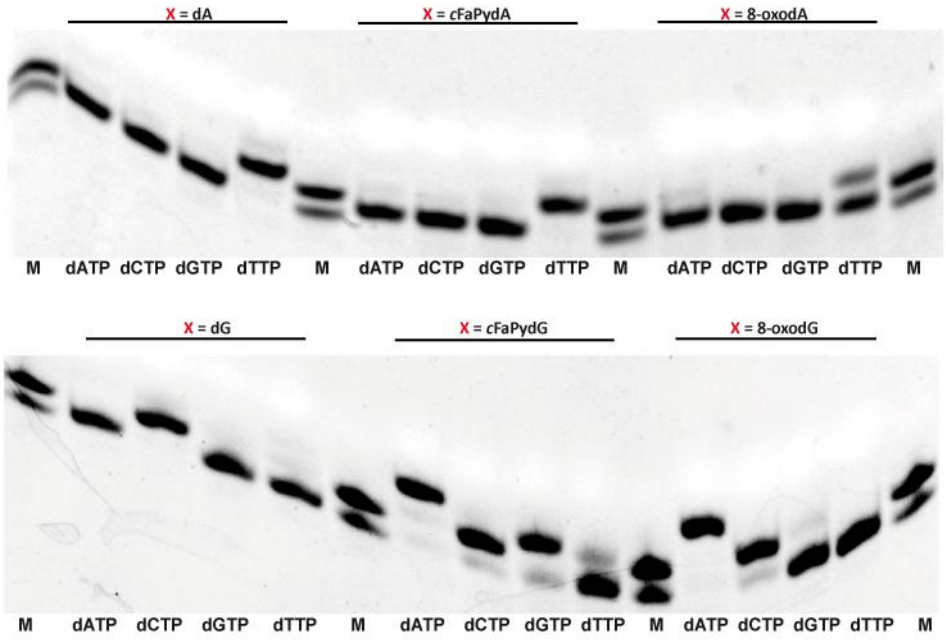
a



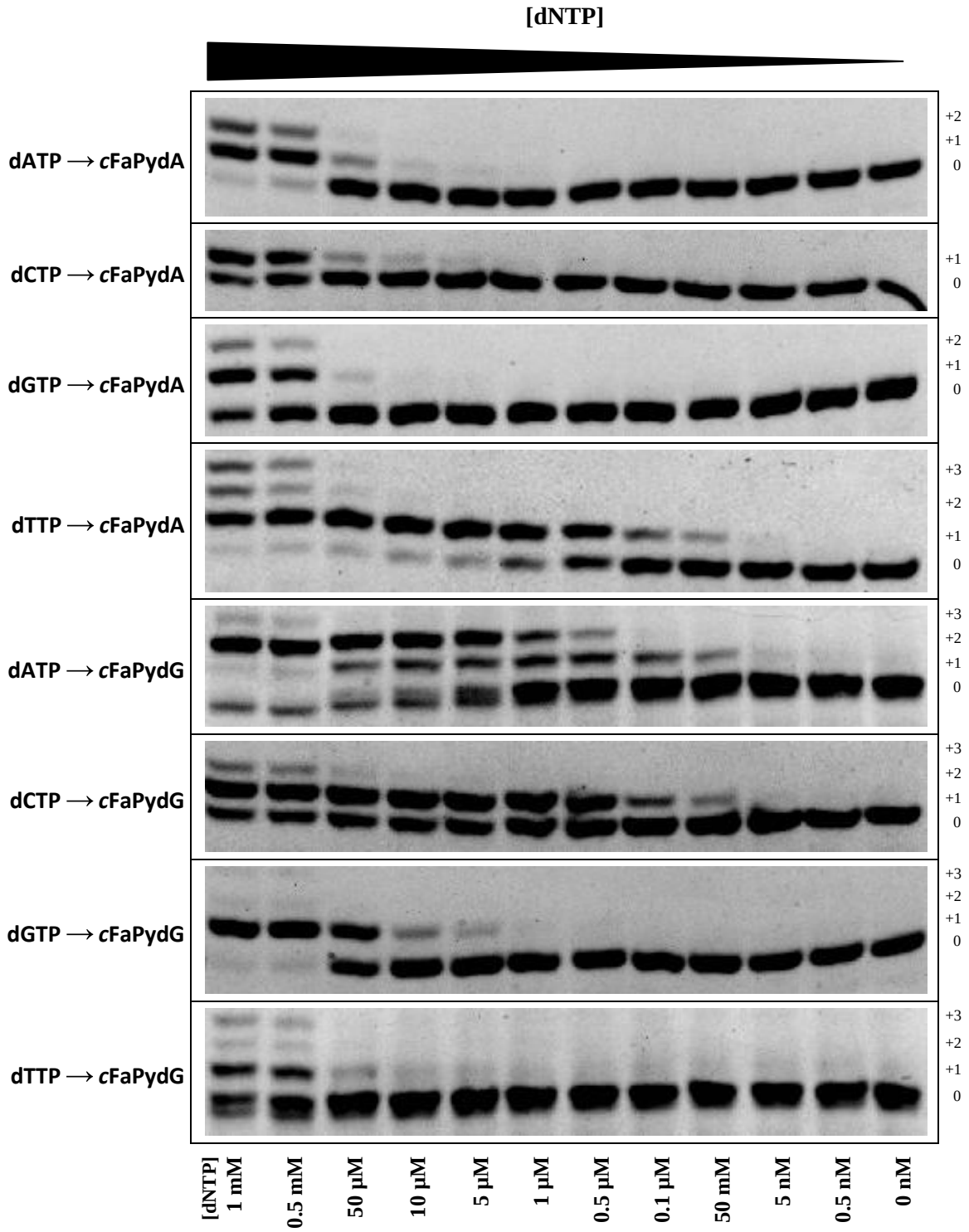
b



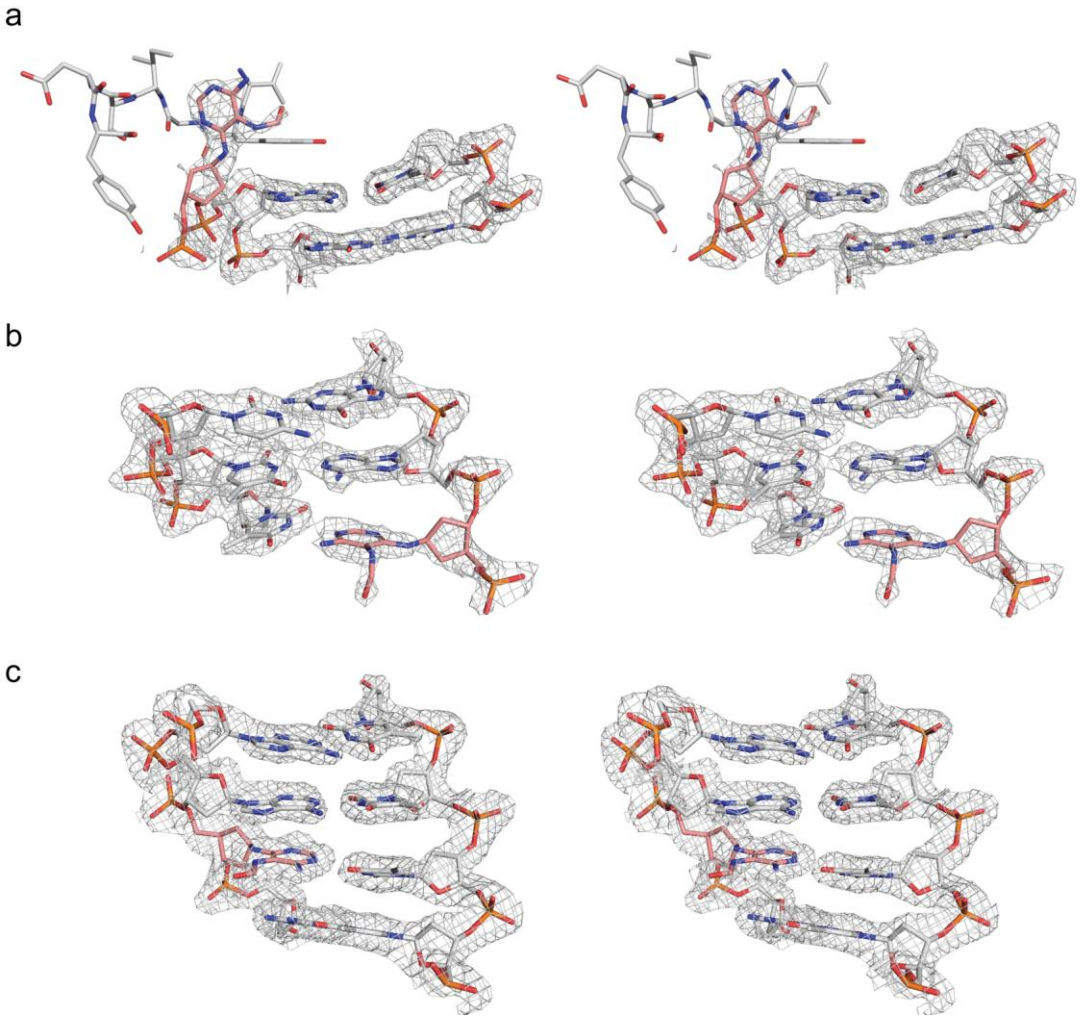
Supplementary Figure 4: Noesy (600 MHz, DMSO–d_6) of dG. The interactions between the C8*H* and the C3´*H* and C2´H_a of the *anti*-dG are shown in red and the interactions between the C8*H* and the C1´*H* of the *syn*-dG are shown in blue. The reference interactions of C2´H_a and C2´H_b are shown in green.



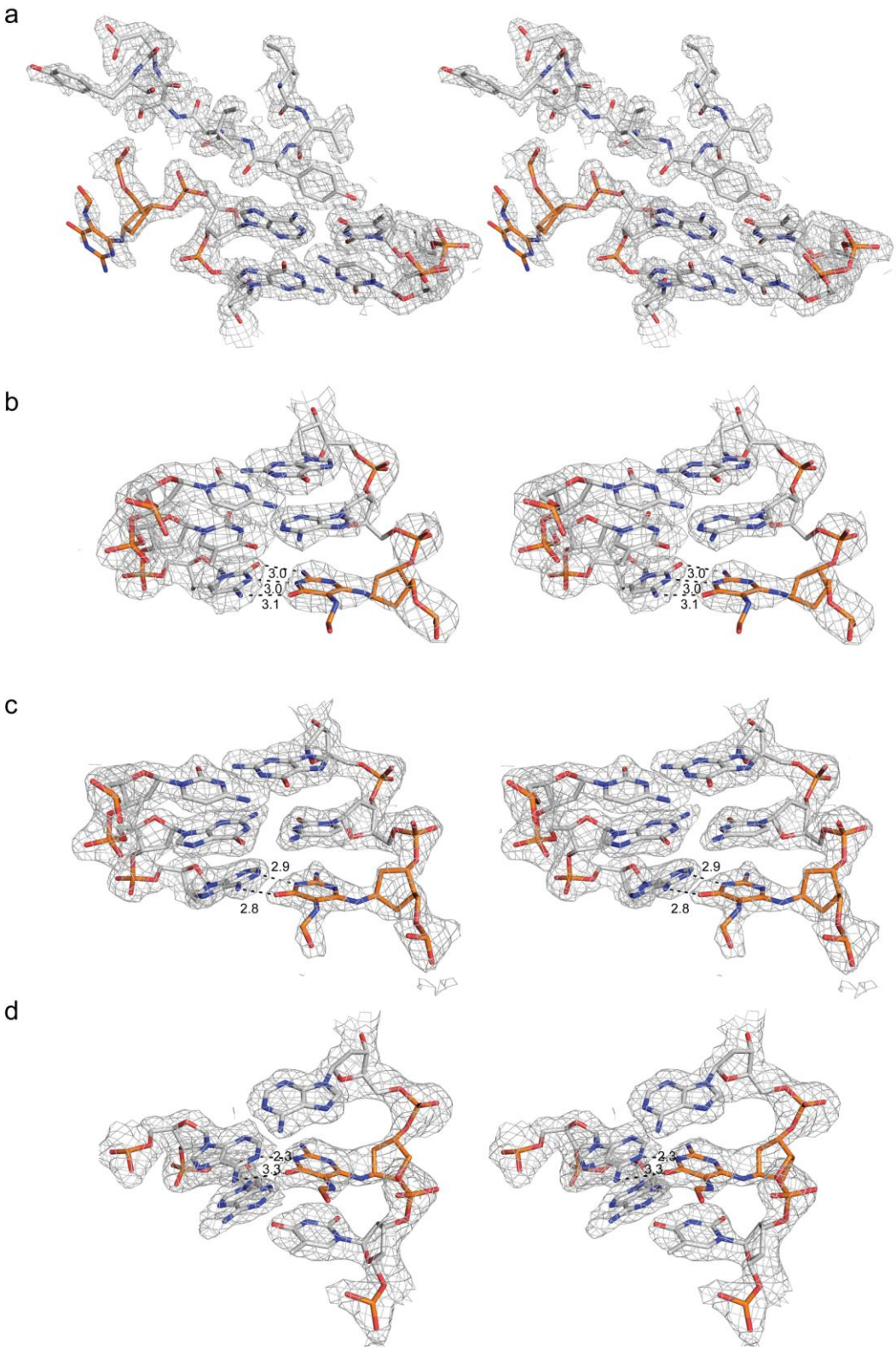
Supplementary Figure 5: Nucleotide insertion and bypass of oxidative lesions. Single nucleotide insertion reactions opposite *c*FaPydA and *c*FaPydG in comparison to 8-oxodA and 8-oxodG. f = fluorescein, M=Marker. The primer was hybridized to a lesion-free control DNA or lesion (= X) containing template strand.



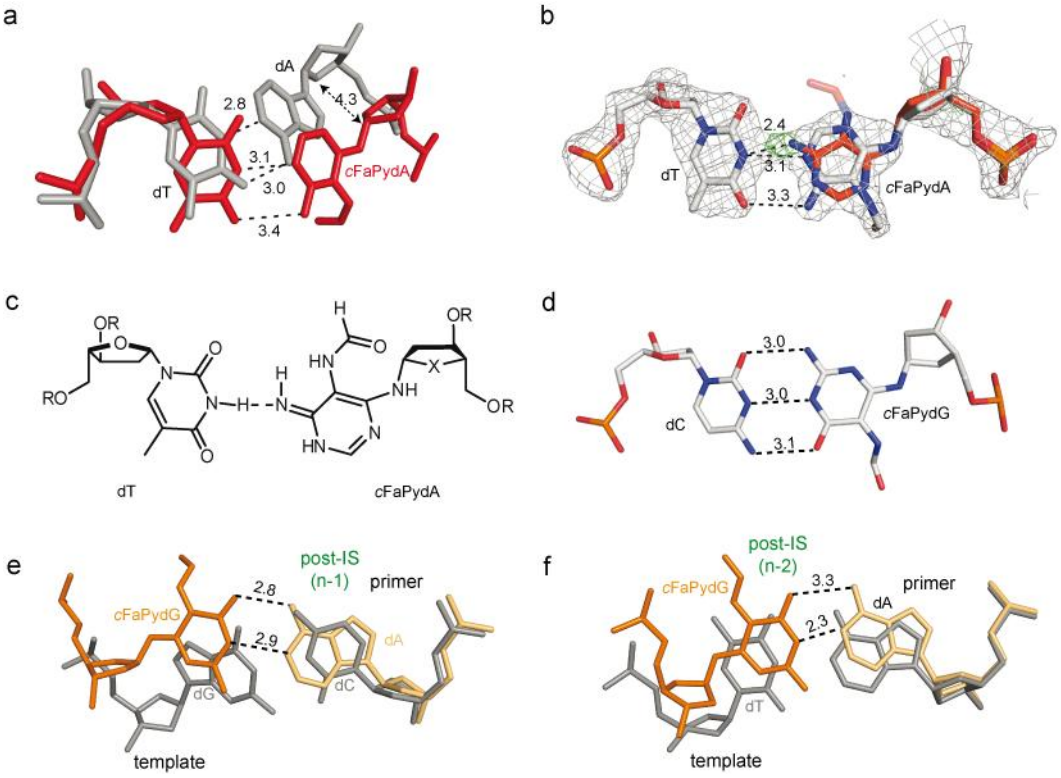
Supplementary Figure 6: Incorporation of dNTP by *G. stearothermophilus* Polymerase I opposite *c*FaPydA and *c*FaPydG. The concentration of *Bst* Pol I and dsDNA was constant throughout the experiment. Experiments were carried out at 55°C.



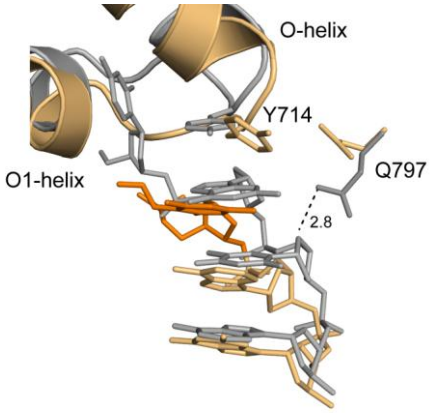
Supplementary Figure 7: Stereo view of the electron density for the *c*FaPydA lesion complexes. (a) *c*FaPydA in the pre-IS of the polymerase. (b) *c*FaPydA•dT at the post insertion site (n-1) after a single insertion event. (c) *c*FaPydA•dT base pair after 3 extension reactions in the crystal at n-3. Composite-omit 2Fo-*D*Fc electron density map is contoured at 1σ level. The lesion is highlighted in red.



Supplementary Figure 8: Stereo view of the electron density for the *c*FaPydG lesion complexes. a) *c*FaPydG at the pre-IS. Due to the flexibility only the sugar is defined in the electron density. b) *c*FaPydG•dC at the post-IS (n-1). Despite the dynamic character of the lesion, which is emphasized by the higher B-factors and weaker electron density, the lesion can be clearly placed. c) *c*FaPydG•dA mismatch at n-1 and d) n-2. Composite-omit 2Fo-*D*Fc electron density map is contoured at 1σ level. The lesion is highlighted in orange.



Supplementary Figure 9: *Bst* Pol I reading through *c*FaPydA- and *c*FaPydG-containing template DNA. (a) Comparison of *c*FaPydA•dT (red) with the correct dA•dT base pair (grey; PDB code 1L3T[1]) at the post-IS (n-1). Superposition was carried out with the coordinates of the polymerase used as reference frame. (b) Alternative conformation (red) of *c*FaPydA at the post-IS as indicated by the positive 2Fo-*D*Fc difference electron density (green). The 2Fo-DFc (blue) and Fo-DFc (green) electron density are contoured at 1 σ and 2.5 σ., respectively. (c) Depiction of the alternative *c*FaPydA•dT base pair. (d) Structure of the *c*FaPydG•dC base pair. (e) and (f) Structure of *Bst* Pol I in complex with *c*FaPydG after incorporation of a dATP at n-1 (PDB Code 4B9U) and n-2 (PDB code 4B9V). (e) Side view of superimposition with the cognate dG•dC base pair (grey, PDB code 1L5U[1]) (f) Side view of superimposition with undamaged DNA (grey; PDB Code 1L5U).



Supplementary Figure 10: Rotamer change of Q797. Superposition of the *Bst* Pol I-DNA complexes of *c*FaPydG•dC (n-1) and dG•dC (PDB code 1L5U). In the complex with the lesion-free DNA Q797 forms a hydrogen bond with the sugar O5 of the base at the post-IS n-2. For clarity the primer strand was removed.

Supplementary Table 1: Quantitative integrals of the interactions between the C8*H* and the C3´*H* and C2´H_a of the *anti*-conformers and the C8*H* and the C1´*H* of the *syn*-dG. As reference the integral of the C2´H_a and C2´H_b interactions were chosen. In brackets are the integrals divided by the reference integrals.

Interactions	**Peak integral (relative integral)**	
	Carbocyclic dG	**dG**
C2H_a-H_b (reference)	464	559
C8*H*-C1´*H*	62 (0.13)	74 (0.13)
C8*H*-C2´H_a	23 (0.06)	33 (0.06)
C8*H*-C3´*H*	107 (0.23)	156 (0.28)

Supplementary Table 2: Synthesized oligonucleotides containing the carbocyclic analogues of FaPydA and FaPydG and the 8-oxo-type lesions.

Name	Sequence	$M_{(calc)}$ (amu)	$M_{(found)}$ [a] (amu)
Strands for Primer Extension coupled Pyrosequencing			
ODN1-*c*FaPydA	5'-GTAGCTAGGTCGAT*c*FaPydAGAGAGACATGCGAGACTGC-3'	10596.8	10598.7
ODN1-*c*FaPydG	5'-GTAGCTAGGTCGAT*c*FaPydGGAGAGACATGCGAGACTGC-3'	10612.9	10612.7
ODN1-8-oxodA	5'-GTAGCTAGGTCGAT8-oxodAGAGAGACATGCGAGACTGC-3'	10597.9	10597.2
ODN1-8-oxodG	5'-GTAGCTAGGTCGAT8-oxodGGAGAGACATGCGAGACTGC-3'	10612.8	10612.4
ODN1-Primer 1	5'-Biotin-GCAGTCTCGCATGTCTCTC-3'		
ODN1-Primer 2	5'-GTAGCTAGGTCGA-3'		
Strands for Single Nucleotide Insertions and Primer Extensions			
ODN2-*c*FaPydA	5'-ACT*c*FaPydAGAGAGACATGCGAGACTGC-3'	7122.6	7124.4
ODN2-*c*FaPydG	5'-ACT*c*FaPydGGAGAGACATGCGAGACTGC-3'	7138.9	7139.7
ODN2-8-oxodA	5'-ACT8-oxodAGAGAGACATGCGAGACTGC-3'	7122.6	7122.0
ODN2-8-oxodG	5'-ACT8-oxodGGAGAGACATGCGAGACTGC-3'	7138.6	7137.2
ODN2-Primer	3'-CTCTCTGTACGCTCTGACG-fluo-5'		
Strands for Crystallization			
ODN3-*c*FaPydA	5'-CAG*c*FaPydAAGAGTCAGGCT-3'	4643.0	4640.4
ODN4-*c*FaPydA	5'-CAA*c*FaPydAAGAGTCAGGCT-3'	4626.2	4625.3
ODN5-*c*FaPydG	5'-CAT*c*FaPydGAGAGTCAGGCT-3'	4634.0	4632.0
ODN6-*c*FaPydG	5'-CAA*c*FaPydGCGAGTCAGGCT-3'	4618.0	4618.8
ODN7-*c*FaPydA	5'-CAT*c*FaPydACGAGTCAGGCT-3'	4593.9	4593.6

[a]MALDI-TOF; matrix: hydroxypicolinic acid

Supplementary Table 3: Data collection, processing and structure refinement statistics of the carbocyclic FaPydA-structures.

PDB code	***c*FaPydA (pre-IS) 4B9L**	**dT-*c*FaPydA (n-1) 4B9M**	**dT-*c*FaPydA (n-3) 4B9N**
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	88.4, 93.7, 105.1	88.4, 93.7, 105.1	88.4, 93.7, 105.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	45.4-2.05 (2.16-2.05)	50.0-2.05 (2.2-2.05)	45.9- 2.2 (2.32-2.2)
R_{merge}	0.152 (0.604)	0.061 (0.458)	0.091 (0.586)
$I / \sigma I$	6.6 (2.4)	17.5 (3.9)	16.0 (3.2)
Completeness (%)	99.8 (100)	99.7 (100.0)	99.8 /99.1
Redundancy	4.2 (4.2)	4.1 (4.1)	4.9 (4.9)
Refinement			
Resolution (Å)	45.4-2.05	50.0-2.05	45.9-2.2
No. reflections	54,997	52,362	44,989
R_{work} / R_{free}	0.187 / 0.226	0.18 /0.223	0.183 / 0.2173
No. atoms	5,385	5,384	5,395
Protein	4,683	4,680	4,673
DNA	432	472	433
Water	270	232	289
B-factors			
Protein	45.2	32.0	57.1
DNA	76.9	55.2	110.4
Water	35.4	46.0	42.0
R.m.s. deviations			
Bond lengths (Å)	0.009	0.014	0.006
Bond angles (°)	1.21	1.39	1.0

*Values in parentheses are for the highest-resolution shell.

Supplementary Table 4: Data collection, processing and structure refinement statistics of the carbocyclic FaPydG-structures.

PDB code	***c*FaPydG (pre-IS) 4B9S**	**dC-*c*FaPydG (n-1) 4B9T**	**dA-*c*FaPydG (n-1) 4B9U**	**dA-*c*FaPydG (n-2) 4B9V**
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	88.1, 93.7, 105.6	88.3, 93.8, 105.6	88.7, 93.9, 105.6	87.9, 94.0, 105.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	46.8-1.75 (1.84-1.75)	45.0-2.65 (2.8-2.65)	46.3–2.1 (2.21-2.1)	47.0-2.0 (2.11-2.0)
R_{merge}	0.043 (0.507)	0.102 (0.474)	0.103 (0.666)	0.060 (0.441)
$I / \sigma I$	24.3 (3.9)	17.6 (4.0)	11.6 (3.3)	18.9 (4.6)
Completeness (%)	99.7 (99.4)	95.3 (97.1)	99.3 (99.0)	99.8 (99.8)
Redundancy	5.9 (6.1)	5.9 (5.8)	4.5 (4.6)	5.9 (6.1)
Refinement				
Resolution (Å)	46.8-1.75	45.0-2.65	46.3-2.1	47.0-2.0
No. reflections	86,204	34,272	49,156	56,504
R_{work} / R_{free}	0.181 / 0.206	0.221 / 0.273	19.1 / 22.4	0.187 / 0.215
No. atoms	5,484	5,201	5,368	5,362
Protein	4,677	4,658	4,658	4,675
DNA	439	473	495	473
Water	358	70	217	206
B-factors				
Protein	26.0	60.7	30.3	30.0
Ligand/ion	39.4	121.0	52.4	55.5
Water	32.0	92.9	34.4	31.4
R.m.s. deviations				
Bond lengths (Å)	0.017	0.011	0.014	0.0139
Bond angles (°)	1.8	1.56	1.6	1.6

*Values in parentheses are for the highest-resolution shell.

References

1. Johnson, S.J., Taylor, J.S. & Beese, L.S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl. Acad. Sci. USA* **100**, 3895-3900 (2003).

Supplementary Note 1

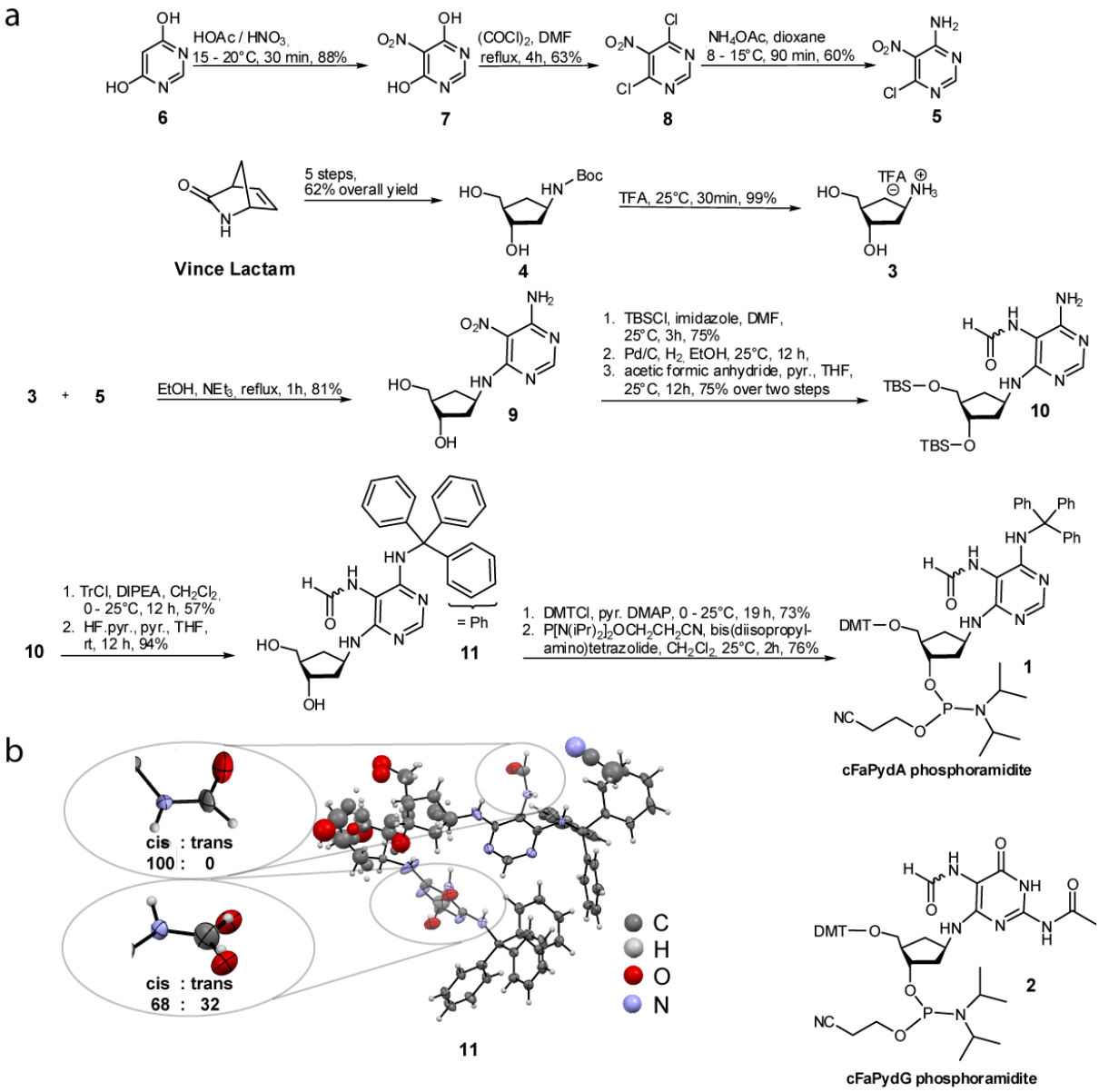
Overview of the synthesis of carbocyclic FaPydA

Here we report the synthesis of a new phosphoramidite building block for the preparation of carbocyclic FaPydA containing DNA (*c*FaPydA) **1** (**Supplementary Fig. 1**). Together with the corresponding building block for FaPydG (*c*FaPydG) **2** lesion containing DNA duplexes were prepared and crystallized in complex with the large fragment of the high fidelity polymerase I from *Geobacillus stearothermophilus*.

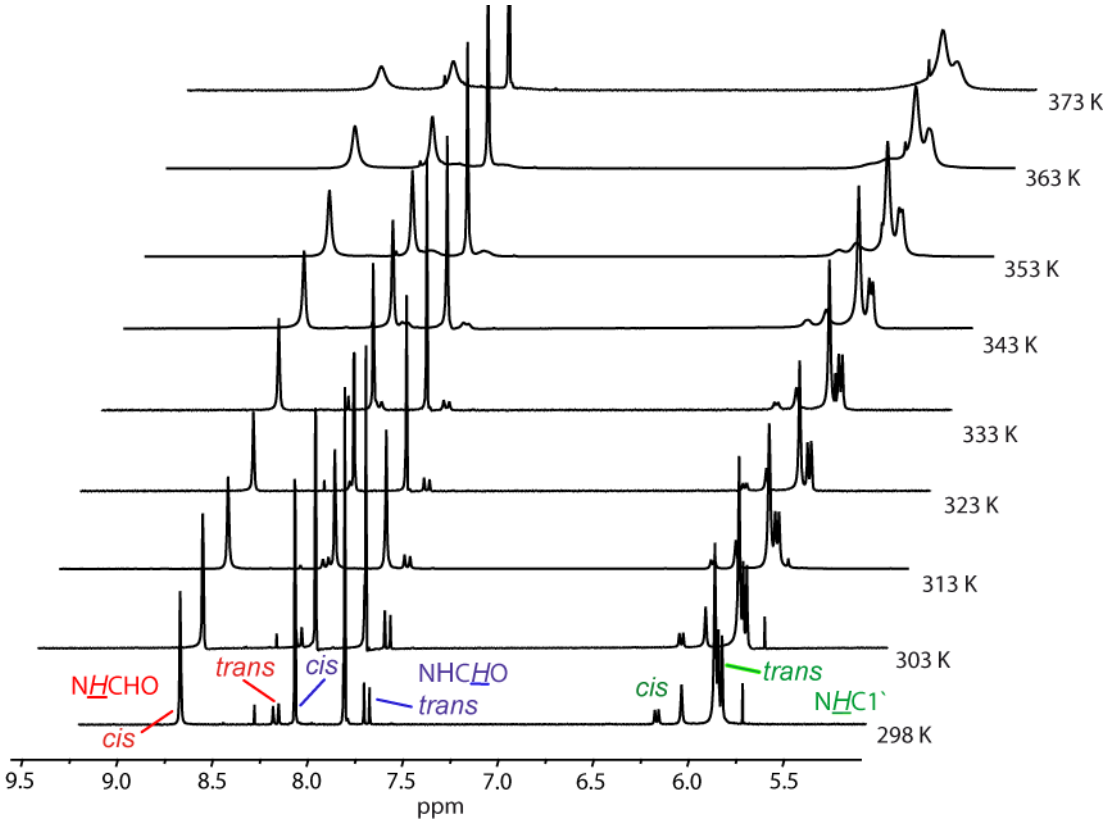
For the synthesis of *c*FaPydA we prepared the enantiomerically pure amino cyclopentane unit (**3**) starting from Vince lactam *via* the Boc-protected aminocyclopentane (**4**) (**Supplementary Fig. 1**). The pyrimidine moiety (**5**), which serves as a precursor for the pyrimidine heterocycle, was obtained from 4,6-dihydroxy-pyrimidine (**6**) *via* compound (**7**) and (**8**) in three steps. Coupling of (**3**) and (**5**) was gave compound (**9**) in yields of above 80%. Subsequent protection of the two OH-groups with TBSCl, followed by reduction of the nitro group with Pd/C under hydrogen atmosphere, provided the unstable amine, which was not isolated but *in situ* converted into the formamide (**10**) with formyl acetic anhydride. Due to the directing effect of the two pyrimidine ring nitrogens the formylation occurred exclusively at position N5. Compound (**11**) was obtained after protection of the N4 nitrogen with triphenylmethyl-chloride (TrCl) in pyridine, which is an unusual protecting group for this position. Deprotection of the 3´ and 5´ OH groups under mild conditions was conducted using HF·pyridine in pyridine. The pseudo-*β*-furanosidic configuration of the precursor (**11**) was proven by X-ray crystallography. The structural data show the *cis* rotamer of the formamide group to be the main isomer in the crystal (-N5H-CHO) (**Supplementary Figure 1b)**. The formamide group adopts a position orthogonal to the pyrimidine ring in order to avoid unfavorable steric interactions. NMR spectroscopic analysis of the compound (**10**) in different solvents confirms a strong solvent-dependency of the *cis*/*trans* equilibrium with the compound preferring the *cis* conformation as observed in the crystal structure. The rotation barrier calculated from temperature dependent NMR experiments to 71.7 ± 2.9 $kJmol^{-1}$ (DMSO-d_6) is in good agreement with literature values for unsubstituted formamide (75.5 ± 1.1 $kJmol^{-1}$, DMSO-d_6) (**Supplementary Fig. 2**).

The final functionalization of the 5´ hydroxyl group with 4,4'-dimethoxytrityl chloride (DMTCl) for DNA synthesis was carried out under standard conditions followed by conver-

sion of the 3´ OH group into the phosphoramidite (**1)** using 3-(bis(diisopropylamino)phosphanooxy)-propanenitrile (for detailed procedures see below).



Supplementary Figure 1: (a) Synthesis of the protected carbocyclic lesion analog *c*FaPydA (**1**). The carbocyclic FaPydG (**2**) was synthesized as previously reported. (b) Crystal structure of compound **11** with resolved *cis* and *trans* rotamers of the formamide units.

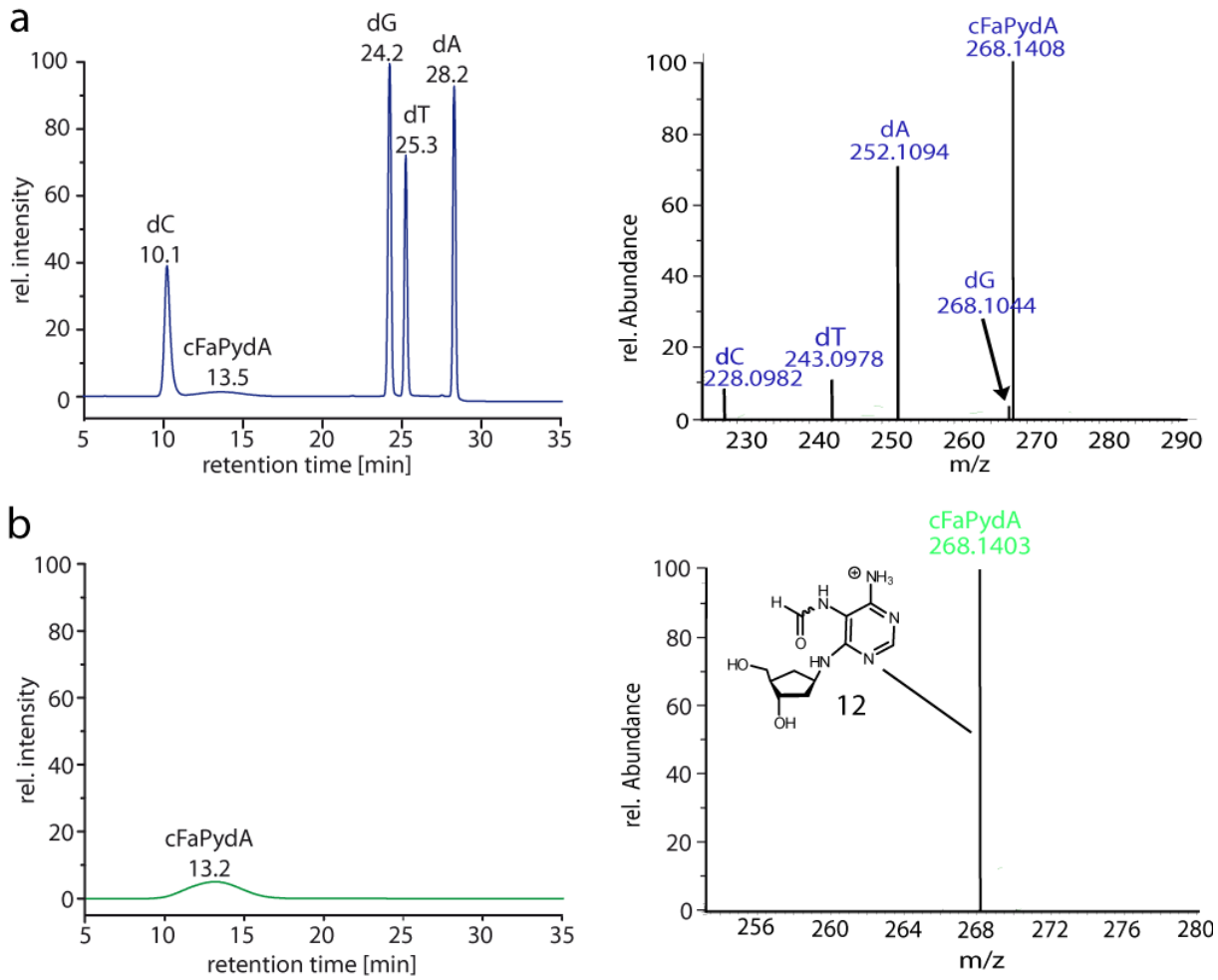


Supplementary Figure 2: ^{1}H-NMR (400 MHz, DMSO–d_6) of compound **1** at different temperatures. The formamide protons are marked in red and blue. The *cis* protons are shifted upfield due to the anisotropic effect. Protons of the C1`N*H* are highlighted in green.

Both phosphoramidites (**1** and **2**) were subsequently incorporated into different oligonucleotides (**Supplementary Information - Table 2**).

To this end the standard DNA synthesis protocol was adjusted. The coupling time of the carbocyclic FaPy-phosphoramidites was extended from 96 sec to 10 min for the *c*FaPydG compound (**2**) and even to 14 min for the *c*FaPydA (**1**) lesion due to the steric problems associated with the bulky triphenylmethyl (trityl) group of **1**. The acidic deprotection steps during DNA synthesis were shortened in order to avoid cleavage of the acid sensitive trityl group. The DNA was subsequently cleaved from the solid support and deprotected using saturated ammonia solution in ethanol. Final deprotection of the trityl group of **1** was achieved with 80% acetic acid in aqueous solution for 20 min at rt. After precipitation of the DNA purification was performed by HPL-chromatography. Correct incorporation of the *c*FaPydA building block was proven by enzymatic digestion, followed by high resolution LC-MS characterization of the digest (**Supplementary Fig. 3**). In these experiments we detected next to the four canonical nucleotides dA, dC, dG and dT, a fifth broad peak with a retention time of 13.5 min

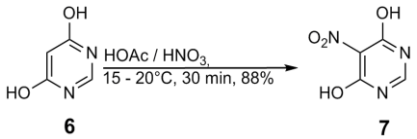
and a moleculer weight of *m/z* = 268.1408, is excellent agreement with the expected molecular weight (*m/z* = 268.1404) for the fully deprotected *c*FaPydA compound (**12**) (**Supplementary Fig. 3b**). The broad appearance of the peak is attributed to the *cis*/*trans* interconversion of the formamide group on the column.



Supplementary Figure 3: (a) HPLC/MS (ESI$^+$) of the digested oligonucleotide **ODN7** with *c*FaPydA (top, blue), (b) monomeric carbocyclic nucleoside *c*FaPydA **12** (bottom, green). The numbers in the HPLchromatograms are retention times in min. The numbers in the mass spectra are the found molecular weights.

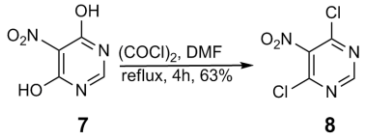
Detailed synthesis of carbocyclic FaPydA nucleoside phosphoramidite 1

4,6-dihydroxy-5-nitropyrimidine 7



To a cooled solution consisting of conc. acetic acid (60 mL) and nitric acid (20 mL) 4,6-dihydroxypyrimidine **6** (20.0 g, 178 mmol) was added slowly over 30 min**.** The solution was kept between 10 and 20 °C. After stirring for another 30 min between 10 and 20 °C, the solution was additionally stirred for 10 min at rt. The reaction was then poured onto ice (150 mL) and the resulting precipitate was filtered off and dried under high vacuum. 4,6-dihydroxy-5-nitropyrimidine **7** (25 g, 160 mmol) was isolated as a yellow solid. mp: 320 °C(lit °C); ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.72 (s, 1H, H_{Ar}), 13.21 (br s, 2H, 2 x OH); ^{13}C NMR (100 MHz, DMSO-d_6): δ 119.9 (C-NO_2), 150.9 (CH), 155.5 (2 x C-OH); IR (ATR): 3027, 2937, 1726, 1664 1555, 1468, 1350, 1310, 1280, 1196, 884, 788, 631 cm^{-1}; EI-HRMS (m/z): $[M]^{\cdot+}$ calcd. for $C_4H_3N_3O_4$, 157.0118; found, 157.0118.

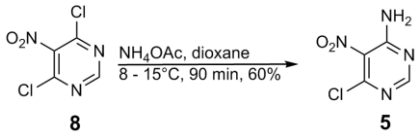
4,6-dichloro-5-nitropyrimidine 8



To a suspension consisting of 4,6-dihydroxy-5-nitropyrimidine **7** (500 mg, 3.18 mmol) and oxalyl chloride (2.0 mL, 22.4 mmol, 7 eq), 1 mL of DMF (12.8 mmol, 4 eq) was added carefully. The reaction mixture was heated under reflux for 4 h while the reaction mixture changed from red to yellow. The mixture was dried under high vacuum and to the resulting solid was added ice (5 mL). The solution was extracted with toluene (3 x 20 mL) and the organic phase was washed another time with water. The organic layer was dried over $MgSO_4$ and the organic solvent was removed under reduced pressure. 4,6-dichloro-5-nitropyrimidine **8** was recrystallized from *n*-hexane to give colorless plates.

mp: 101 °C; TLC (toluene:EtOAc, 10:1 v/v): R_f = 0.60; ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.75 (s, 1H, CH_{Ar}); ^{13}C NMR (100 MHz, DMSO-d_6): δ 119.9 (C-NO_2), 150.9 (C-H), 155.5 (C-Cl); IR (ATR) 3055, 1682, 1588 , 1534, 1428, 1357, 1313, 1127, 1050, 976, 855, 834, 785, 718 cm^{-1}; EI-HRMS (m/z): $[M]^{\cdot+}$ calcd. for $C_4H^{35}Cl_2N_3$, 192.9440; found, 192.9433.

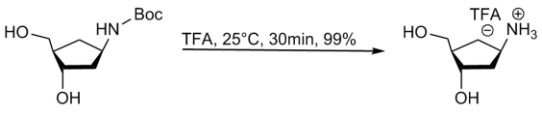
4-amino-6-chloro-5-nitropyrimidine 5



8 5

To a suspension of 4,6-dichloro-5-nitropyrimidine **8** (800 mg, 4.12 mmol) in dioxane, kept at 8 °C in an ice bath, a freshly prepared solution of conc. acetic acid (1.6 mL) in ammonium hydroxide (28 %, 3.2 mL) was slowly added *via* syringe pump to the suspension. The solution did not warm up over 15 °C and was stirred for another 90 min under ice cooling. The product was then precipitated by adding ice cold distilled water (18 mL). The resulting yellow solid was filtrated and recrystallized from benzene to separate it from the diaminated product. The recrystallized product **5** could be obtained as small yellow needles (425 mg, 2.43 mmol, 60 %). mp: 155 °C; TLC ($CHCl_3$:MeOH, 10:1 v/v): R_f = 0.36; FTIR (ATR): 3307, 3144, 1647, 1575, 1519, 1396, 1340, 1233, 1042, 964, 861, 777 cm^{-1}; ^1H NMR (300 MHz, DMSO–d_6): δ 7.95 (s, 1H, H_{Ar}), 8.67 – 8.80 (br s, 2H, NH_2); ^{13}C NMR (100 MHz, DMSO–d_6): δ 115.3 (C_{Ar}), 152.0 (C_{Ar}), 155.4 (C_{Ar}), 160.31 (C_{Ar}); EI-HRMS (m/z): $[M]^{\cdot+}$ calcd. for $C_4H_3{}^{35}ClN_4O_2$, 173.9945; found, 173.9948.

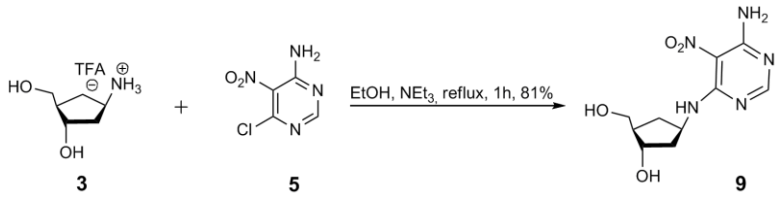
(1*S*,2*R*,4*R*)-4-ammonium-2-(hydroxymethyl)cyclopentanol-trifluoroacetate 3



4 3

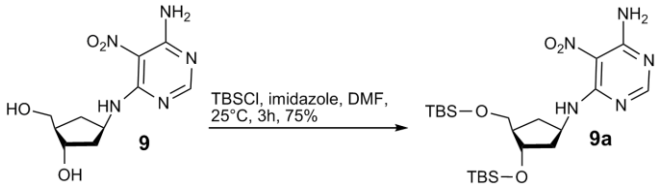
The carbamate *tert*-Butyl *N*-[(1*R*,3*S*,4*R*)-3-hydroxy-4-(hydroxymethyl)cyclopentyl]carbamate **4** was synthesized according to literature from Vince lactam and added to 10 ml of trifluoroacetic acid in water (95:5). The reaction was stirred for 30 min at rt. This was followed by removal of acid and water under reduced pressure. The resulting colorless oil of **3** (420 mg, 1.71 mmol, 99%) was dried under high vacuum overnight und used without any further purification. mp: 3 – 4 °C; FTIR (ATR): 3295, 3045, 2938, 1785, 1668, 1532, 1432, 1392, 1369, 1350, 1178, 1130, 1077, 1017, 920, 880, 839, 798, 777, 722 cm^{-1}; ^1H NMR (400 MHz, D_2O–d_2): δ 1.07 (dt, $^2J_{H\text{-}H}$ (C5´H_b,C5´H_a) = 13.3 Hz, $^3J_{H\text{-}H}$ (C5´H_b,C1´H,C4´H) = 9.1 Hz, 1 H, C5´H_b), 1.63–1.70 (m, 1 H, C2´H_b), 1.72–1.81 (m, 2 H, C2´H_a, C4´H), 2.03–2.11 (m, C5´H_a), 3.27 (dd, $^2J_{H\text{-}H}$ (C1´´H_b,C1´´H_a) = 11.1 Hz, $^3J_{H\text{-}H}$ (C1´´H_b,C4´H) = 6.27 Hz, 1 H, C1´´H_b), 3.33 (dd, $^2J_{H\text{-}H}$ (C1´´H_a,C1´´H_b) = 11.2 Hz, $^3J_{H\text{-}H}$ (C1´´H_a,C4´H) = 5.8 Hz, 1 H, C1´´H_a), 3.44–3.52 (m, 1 H, C1´H), 3.83 (m, 1H, C3´H); ^{13}C NMR (100 MHz, D_2O–d_2): δ 30.24 (C´5), 38.95 (C´2), 48.90 (C´4), 49.65 (C´1), 63.18 (C´´1), 73.08 (C´3), 116.60 (q, TFA), 164.20 (q, TFA); ^{19}F NMR (200 MHz, D_2O–d_2): δ –76.30 (s, 3 F, C$\underline{F}_3CO_2^-$); EI-HRMS (m/z): [M + H^+]: calcd. for $[C_6H_{13}NO_2+H]^+$, 132.1019; found, 132.1020.

***N*-4-{[(1´*R*,3´*S*,4´*R*)-3´-Hydroxy-4´-(hydroxymethyl)-cyclopentyl]amino}-5-nitro-6-amino-pyrimidine 9.**



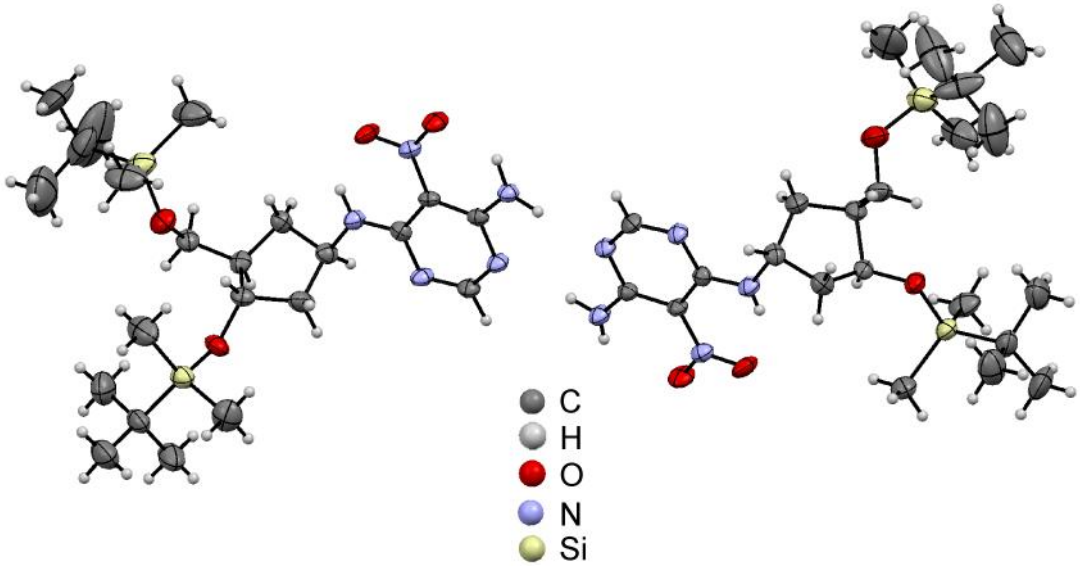
The crude oil of **3** (420 mg, 1.71 mmol), was dissolved in 2.1 mL of dry ethanol, and 4-amino-6-chloro-5-nitropyrimidine **5** (302 mg, 1.73 mmol, 1 eq) together with triethylamine (1.68 mL, 12.1 mmol, 7 eq) was added. The mixture was refluxed for 1 h. After cooling to room temperature, the resulting yellow precipitate **9** (377 mg, 1.40 mmol, 81 %) was filtered, washed with ethanol, dried under high vacuum and used in the following reaction without any further purification. mp: decomp. > 169 °C; TLC ($CHCl_3$:MeOH, 4:1 v/v): R_f = 0.40; FTIR (ATR): 3447, 2928, 1651, 1605, 1515, 1398, 1369, 1257, 1137, 1043, 1004, 794, 586, 557, 425 cm^{-1}; ^{1}H NMR (400 MHz, DMSO–d_6): δ 1.24–1.34 (m, 1 H, C5´H_a), 1.76 (ddd, $^2J_{H\text{-}H}$ (C2´H_b,C2´H_a) = 13.2 Hz, $^3J_{H\text{-}H}$ (C2´H_b, C1´H) = 7.72 Hz, $^3J_{H\text{-}H}$ (C2´H_b, C3´H) = 6.53 Hz, 1 H, C2´H_b), 1.85–1.93 (m, 2 H, C2´H_a, C4´H), 2.25 (dt, $^2J_{H\text{-}H}$ (C2´H_a,C2´H_b) = 13.0 Hz, $^3J_{H\text{-}H}$ (C2´H_a,C1´H,C3´H) = 8.1 Hz), 1.31 (dt, $^2J_{H\text{-}H}$ (C5´H_a,C5´H_b) = 13.1 Hz, $^3J_{H\text{-}H}$ (C5´H_a,C1´H,C4´H) = 7.88 Hz, 1 H, C5´H_a), 3.38 (dd, $^2J_{H\text{-}H}$ (C1´´H_b,C1´´H_a) = 10.5 Hz, $^3J_{H\text{-}H}$ (C1´´H_b,C4´H) = 5.72 Hz, 1 H, C1´´H_b), 3.44 (dd, $^2J_{H\text{-}H}$ (C1´´H_a,C1´´H_b) = 10.5 Hz, $^3J_{H\text{-}H}$ (C1´´H_a,C4´H) = 5.50 Hz, 1 H, C1´´H_a), 3.96 (dt, $^3J_{H\text{-}H}$ (C3´H,C2´H_b) = 6.25 Hz, $^3J_{H\text{-}H}$ (C3´H,C2´H_a,C4´H) = 4.41 Hz, 1 H, C3´H), 4.60 (br, 2 H, 2 x OH), 4.71 (m, 1 H, C1´H), 7.97 (s, 1 H, C1H), 8.51 (br, 1 H, NH), 8.58 (br, 1 H, NH), 9.15 (d, $^3J_{H\text{-}H}$ (C1´NH, C1´H) = 7.52 Hz, 1 H, C1´NH);^{13}C NMR (100 MHz, DMSO–d_6): δ 34.11 (C´5), 41.50 (C´2), 48.90 (C´4), 50.12 (C´1), 62.52 (C´´1), 72.08 (C´3), 111.68 (C5), 155.99 (C6), 158.77 (C4), 159.42 (C2); ESI-HRMS (m/z): [M + H^+] calcd. for $[C_{10}H_{15}N_5O_4+H]^+$, 270.1197; found, 270.1199.

***N*-4-{[(1´*R*,3´*S*,4´*R*)-3´-[(*tert*-Butyldimethylsilyl)oxy]-4´-{[(*tert*-butyldimethylsilyl)oxy]-methyl}-cyclopentyl]amino}-5-nitro-6-amino-pyrimidine 9a.**



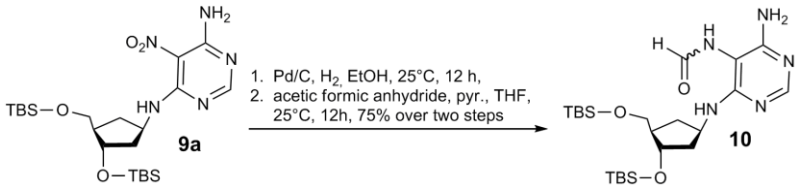
Compound **9** (450 mg, 0.90 mmol) was dissolved in dry DMF (15 mL). To the solution was given imidazole (184 mg, 2.71 mmol, 3 eq) and *tert*-butyldimethylsilylchloride (408 mg, 2.71 mmol, 3 eq) and the reaction was stirred at rt for 3 h. After complete conversion of the starting material, the reaction mixture was diluted with ethylacetate (125 mL) and washed with saturated $NaHCO_3$ (125 mL) and brine (4 x 125 mL.). The combined aqueous phases were extracted once again with 100 mL of ethylacetate. The collected organic phases were dried over anhydrous Na_2SO_4, filtrated and the solvent was removed under reduced pressure. The crude product was purified *via* column chromatography ($CHCl_3$: MeOH 20:1 → 10:1). The product was obtained as a yellow solid (336 mg, 0.68 mmol, 75 %). Crystals for X-ray analy-

sis were received by dissolving **9a** in *n*-heptane at 40 °C. After slowly cooling the solution to rt, **9a** crystallized in colorless plates over a period of 14 days at rt (**Supplementary Figure 4)**. mp: 50 °C; TLC ($CHCl_3$:MeOH, 10:1 v/v): R_f = 0.59; FTIR (ATR): 3397, 3344, 3274, 3122, 2953, 2927, 2855, 1583, 1546, 1522, 1471, 1390, 1358, 1247, 1116, 1094, 1038, 1006, 960, 936, 870, 832, 796, 775 cm^{-1}; ^1H NMR (600 MHz, $CDCl_3$): δ 0.03 (s, 3 H, Si-CH_3), 0.04 (s, 3 H, Si-CH_3), 0.04 (s, 3 H, Si-CH_3), 0.04 (s, 3 H, Si-CH_3), 0.87 (s, 9 H, Si-C(CH_3)$_3$), 0.88 (s, 9 H, Si-C(CH_3)$_3$), 1.31 (dt, $^2J_{H\text{-}H}$ (C5´H_b,C5´H_a) = 13.1 Hz, $^3J_{H\text{-}H}$ (C5´H_b,C1´H,C4´H) = 8.8 Hz, 1 H, C5´H_b), 1.75 (ddd, $^2J_{H\text{-}H}$ (C2´H_b,C2´H_a) = 13.1 Hz, $^3J_{H\text{-}H}$ (C2´H_b,C1´H) = 7.7 Hz, $^3J_{H\text{-}H}$ (C2´H_b,C3´H) = 6.7 Hz, 1 H, C2´H_b), 2.04 (m, 1 H, C4´H), 2.09 (ddd, $^2J_{H\text{-}H}$ (C2´H_a,C2´H_b) = 12.6 Hz, $^3J_{H\text{-}H}$ (C2´H_a,C1´H) = 7.8 Hz, $^3J_{H\text{-}H}$ (C2´H_a,C3´H) = 4.4 Hz, 1 H, C2´H_a), 1.31 (dt, $^2J_{H\text{-}H}$ (C5´H_a,C5´H_b) = 13.1 Hz, $^3J_{H\text{-}H}$ (C5´H_a,C1´H,C4´H) = 7.88 Hz, 1 H, C5´H_a), 3.56 (dd, $^2J_{H\text{-}H}$ (C1´´H_b,C1´´H_a) = 10.1 Hz, $^3J_{H\text{-}H}$ (C1´´H_b,C4´H) = 5.0 Hz, 1 H, C1´´H_b), 3.61 (dd, $^2J_{H\text{-}H}$ (C1´´H_a,C1´´H_b) = 10.1 Hz, $^3J_{H\text{-}H}$ (C1´´H_a,C4´H) = 5.0 Hz, 1 H, C1´´H_a), 4.17 (m, 1 H, C3´H), 4.77 (m, 1 H, C1´H), 7.20 (br, 1 H, NH), 8.01 (s, 1 H, C1H), 8.57 (br, 1 H, NH), 9.02 (d, $^3J_{H\text{-}H}$ (C1´NH, C1´H) = 7.2 Hz, 1 H, C1´NH); ^{13}C NMR (150 MHz, $CDCl_3$): δ –5.49 (Si-CH_3), –5.41 (Si-CH_3), –4.78 (Si-CH_3), –4.61 (Si-CH_3), 17.98 (Si-$\underline{C}$(CH_3), 18.31 (Si-$\underline{C}$(CH_3), 25.80 (Si-C($\underline{C}H_3$)$_3$, 25.91 (Si-C($\underline{C}H_3$)$_3$, 33.97 (C´5), 42.20 (C´2), 49.75 (C´4), 50.27 (C´1), 63.27 (C´´1), 73.10 (C´3), 112.69 (C5), 156.57 (C6), 159.33 (C4), 159.48 (C2); ESI-HRMS (m/z): [M + H^+] calcd. for [$C_{22}H_{43}N_5O_4Si_2$+H]$^+$, 498.2926; found, 498.2926.



Supplementary Figure 4: Crystal structure of compound 9a recrystallized from *n*-heptane.

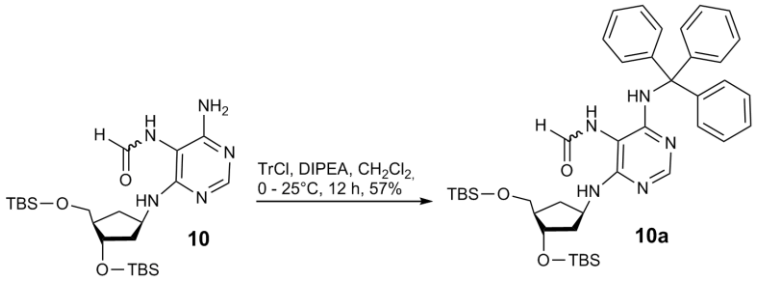
***N*-4-{[(1´*R*,3´*S*,4´*R*)-3´-[(*tert*-Butyldimethylsilyl)oxy]-4´-{[(*tert*-butyldimethylsilyl)oxy]-methyl}-cyclopentyl]amino}-5-formylamino-6-amino-pyrimidine 10.**



The silylether **9a** (200 mg, 0.40 mmol) was dissolved in dry ethanol (5 mL) and the reaction mixture was degassed carefully. The solution was given to a degassed suspension of Pd/C (100 mg) in dry ethanol (5 mL). The mixture was flushed with hydrogen and stirred for 3 h under the hydrogen atmosphere (1.2 bar) at rt. Under exclusion of oxygen, the charcoal and palladium were filtered off, using a 0.2 µm nylon syringe filter (Millipore^{TM}). The solvent was evaporated under reduced pressure at rt and the air sensitive amine was carefully co-evaporated three times using dry pyridine (3 x 1 mL) under high vacuum. The amine was then dissolved in 10 mL of degassed THF (abs.), 815 µL of dry pyridine were added and the solution was cooled to 0 °C. Meanwhile a solution consisting of formic acid (501 µL) and acetic acid anhydride (1.28 mL) was heated to 60 °C for 15 min and the mixed anhydride was afterwards cooled to 0 °C and carefully degassed. From this solution 617 µL of the mixed anhydride were promptly added to the amine and the reaction was stirred at 0 °C overnight reaching rt. After evaporation of the solvents the resulting crude reaction mixture was purified by column chromatography ($CHCl_3$:MeOH 20:1 → 10:1). The resulting formamide **10** could be obtained as white foam (147 mg, 0.30 mmol, 75 % over two steps). The *cis*- and *trans*-formamide ratio in DMSO was 3:1 and in $CDCl_3$ 1:1. mp: 141 – 143 °C; TLC ($CHCl_3$:MeOH 10:1 v/v): R_f= 0.19; FTIR (ATR): 3338, 2929, 2856, 1685, 1589, 1501, 1470, 1360, 1327, 1250, 1077, 1005, 938, 832, 772, 667 cm^{-1}; ^{1}H NMR (400 MHz, DMSO–d_6): δ 0.03 (s, 12 H, 4 x Si-CH_3), 0.87 (s, 9 H, Si-C$(CH_3)_3$, 0.87 (s, 9 H, Si-C$(CH_3)_3$, 1.10 (ddd, $^2J_{H\text{-}H}$ (C5´H_b,C5´H_a) = 12.6 Hz, $^3J_{H\text{-}H}$ (C5´H_b,C1´H,C4´H) = 9.3 Hz, 1 H, C5´H_b), 1.68–1.82 (m, 2 H,C2´H_b,C2´H_a), 1.84–1.92 (m, 1 H,C4´H), 2.11 (ddd, $^2J_{H\text{-}H}$ (C5´H_a,C5´H_b) = 12.8 Hz, $^3J_{H\text{-}H}$ (C5´H_a,C1´H,C4´H) = 9.3 Hz, 1 H, C5´H_a), 3.51 (dd, $^2J_{H\text{-}H}$ (C1´´H_b,C1´´H_a) = 9.5 Hz, $^3J_{H\text{-}H}$ (C1´´H_b,C4´H) = 5.0 Hz, 1 H, C1´´H_b), 3.55 (dd, $^2J_{H\text{-}H}$ (C1´´H_a,C1´´H_b) = 9.5 Hz, $^3J_{H\text{-}H}$ (C1´´H_a,C4´H) = 5.3 Hz, 1 H, C1´´H_a), 4.08 (ddd, $^3J_{H\text{-}H}$ (C3´H,C2´H_a) = 7.2 Hz, $^3J_{H\text{-}H}$ (C3´H,C2´H_b,C4´H) = 4.30 Hz, 1 H, C3´H), 4.50 (m, 1 H, C1´H), 5.86 (d, $^3J_{H\text{-}H}$ (C1´NH, C1´H) = 7.80 Hz, 0.75 H, [*cis*-formamide]C1´NH), 5.89 (br, 1 H, NH), 6.07 (br, 1 H, NH), 6.20 (d, $^3J_{H\text{-}H}$ (C1´NH,C1´H) = 7.80 Hz, 0.25 H, [*trans*-formamide]C1´NH), 7.72 (d, $^3J_{H\text{-}H}$ (N*H*CHO,NHC*H*O) = 11.6 Hz, 0.25 H [*trans*-Formamide]N*H*CHO, 7.84 (s, 0.75 H, [*cis*-formamide]C2H), 7.84 (s, 0.25 H, [trans-formamide]C2H), 8.10 (d, $^3J_{H\text{-}H}$ (NHC*H*O,N*H*CHO) = 1.40 Hz, 0.75 H [*cis*-formamide]NHC*H*O, 8.20 (d, $^3J_{H\text{-}H}$ (NHC*H*O,N*H*CHO) = 11.8 Hz, 0.25 H [*trans*-formamide]NHC*H*O, 8.70 (d, $^3J_{H\text{-}H}$ (N*H*CHO,NHC*H*O) = 1.0 Hz , 0.75 H [*cis*-formamide]N*H*CHO); ^{13}C NMR (100 MHz, DMSO–d_6): δ –5.44 (Si-CH_3), –5.42 (Si-CH_3), –4.82 (Si-CH_3), –4.62 (Si-CH_3), 17.71 (Si-*C*(CH_3), 18.00 (Si-*C*(CH_3), 25.72 (Si-C(*C*$H_3)_3$, 25.80 (Si-C(*C*$H_3)_3$, 33.67([*trans*-formamide]C´5), 33.82 ([*cis*-formamide]C´5), 40.73 ([*trans*-formamide]C´2), 41.40 ([*cis*-formamide]C´2), 48.69 (C´1), 49.49 ([*trans*-formamide]C´4), 49.53 ([*cis*-formamide]C´4), 63.74 ([*cis*-formamide]C´´1), 63.85 ([*trans*-formamide]C´´1), 72.84 ([*cis*-formamide]C´3), 72.94 ([*trans*-formamide]C´3), 93.99 ([*cis*-formamide]C5),

94.18 ([*trans*-formamide]C5), 155.46 ([*cis*-formamide]C2), 155.89 ([*trans*-formamide]C2), 157.72 ([*cis*-formamide]C4), 158.69 ([*cis*-formamide]C6), 159.08 ([*trans*-formamide]C4), 160.25 ([*trans*-formamide]C6), 160.92 ([*cis*-formamide]CHO), 165.58 ([*trans*-formamide]CHO); ESI-HRMS (m/z): [M + H^+] calcd. for $[C_{23}H_{45}N_5O_3Si_2+H]^+$; 496.3134; found, 496.3131.

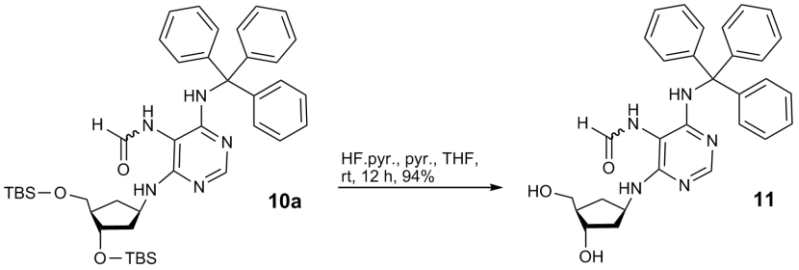
***N*-4-{[(1´*R*,3´*S*,4´*R*)-3´-[(*tert*-Butyldimethylsilyl)oxy]-4´-{[(*tert*-butyldimethylsilyl)oxy]-methyl}-cyclopentyl]amino}-5-formylamino-6-tritylamino-pyrimidine 10a**



The protected silylether **10** (555 mg, 1.12 mmol) was dissolved in dry CH_2Cl_2 and stirred in a Schlenk-flask over molecular sieve (4 Å) for 30 min. The solution was cooled in an ice bath and triphenylmethyl chloride (376 mg, 1.35 mmol, 1.2 eq) was added to the solution. The mixture was stirred overnight reaching rt. The solvent was removed under vacuum and the crude product was purified *via* column chromatography ($CHCl_3$:MeOH: 200:1→ 20:1). The resulting product **10a** could be isolated as white foam (462 mg, 0.63 mmol, 57 %). The *cis*- and *trans*- formamide ratio in AcCN was 2:1 in $CDCl_3$ both conformers were observed in a ratio of 1:1. mp: 133 – 134 °C; TLC ($CHCl_3$:MeOH, 20:1 v/v): R_f = 0.48; FTIR (ATR): 3335, 2914, 2360, 2336, 2060, 2009, 1678, 1588, 1478, 1442, 1326, 1266, 1184, 1155, 1032, 899, 748, 728, 698, 655, 636, 626, 614 cm^{-1}; ^{1}H NMR (400 MHz, MeCN–d_3): δ 0.05 (s, 4 H, [*cis*-Formamide]Si-CH_3), 0.06 (s, 8 H, [*trans*-Formamide]Si-CH_3), 0.88 (s, 9 H, Si-C(CH_3)$_3$, 0.90 (s, 9 H, Si-C(CH_3)$_3$), 1.09–1.17 (m, 1 H, C5´H_b), 1.65 (ddd, $^2J_{H\text{-}H}$(C2´H_b,C2´H_a) = 13.0 Hz, $^3J_{H\text{-}H}$(C2´H_b,C1´H) = 8.6 Hz, $^3J_{H\text{-}H}$ (C2´H_b,C3´H) = 6.6 Hz, 1 H, C2´H_b), 1.84–1.96 (m, 2 H, C2´H_a,C4´H), 2.14–2.22 (m, 1 H, C5´H_a), 3.50 (d, $^2J_{H\text{-}H}$(C1´´H_b,C1´´H_a) = 5.6 Hz, 0.5 H, [*trans*-formamide]C1´´H_b,C1´´H_a), 3.58 (d, $^2J_{H\text{-}H}$(C1´´H_b,C1´´H_a) = 5.7 Hz, 1.5 H, [*cis*-formamide]C1´´H_b,C1´´H_a), 4.12–4.16 (dt, $^3J_{H\text{-}H}$ (C3´H,C2´H_b) = 6.6 Hz, $^3J_{H\text{-}H}$ (C3´H,C2´H_a,C4´H) = 4.3 Hz, 1 H, C3´H), 4.43–4.54 (m, 1 H, C1´H), 5.14 (d, $^3J_{H\text{-}H}$ (C1´NH, C1´H) = 7.6 Hz, 0.75 H, [*cis*-formamide]C1´NH), 5.40 (d, $^3J_{H\text{-}H}$ (C1´NH, C1´H) = 7.7 Hz, 0.25 H, [*trans*-formamide]C1´NH), 6.93 (d, $^3J_{H\text{-}H}$(N*H*CHO,NHC*H*O) = 11.7 Hz, 0.25 H, [*trans*-formamide]N*H*CHO), 7.18–7.28 (m, 10 H, -Trityl), 7.31–7.34 (m, 5 H, -Trityl), 7.42 (s, 0.75 H, [*cis*-formamide]NHC*H*O), 7.51 (s, 0.25 H, [*trans*-formamide]C2H), 7.59 (s, 0.75 H, [*cis*-formamide]C2H), 7.92 (d, $^3J_{H\text{-}H}$(NHC*H*O,N*H*CHO) = 11.7 Hz, 0.25 H, [*trans*-formamide]NHC*H*O), 8.32 (d, $^3J_{H\text{-}H}$(NHC*H*O,N*H*CHO) = 1.2 Hz, 0.75 H, [*cis*-formamide]N*H*CHO), 9.02 (br, 0.25 H, [*trans*-formamide]NH), 9.04 (br, 0.75 H, [*cis*-formamide]NH); ^{13}C NMR (100 MHz, MeCN–d_3): δ –4.73 (Si-CH_3), –4.71 (Si-CH_3), –4.02 (Si-CH_3), –3.83 (Si-CH_3), 19.08 (Si-*C*(CH_3), 19.41 (Si-*C*(CH_3), 26.69 (Si-C(*C*H_3)$_3$, 26.79 (2 x Si-C(*C*H_3)$_3$, 26.85 (Si-C(*C*H_3)$_3$, 35.50 (C5´), 43.10 (C2´), 50.83 (C1´), 51.34 (C4´), 65.29 (C1´´), 71.77 (Cq from -trityl), 74.75 (C3´), 97.66 (C5), 127.91 (CH-Trityl), 128.98 (CH-Trityl), 130.2 (CH-Trityl), 147,35 (C- from Trityl), 155,93 (C2), 158,28 (C4), 158.99 (C6),

162.58 (<u>C</u>HO); ESI-HRMS (m/z): [M + H^+] calcd. for [$C_{42}H_{59}N_5O_3Si_2$+H]$^+$, 738.4229; found, 738.4238.

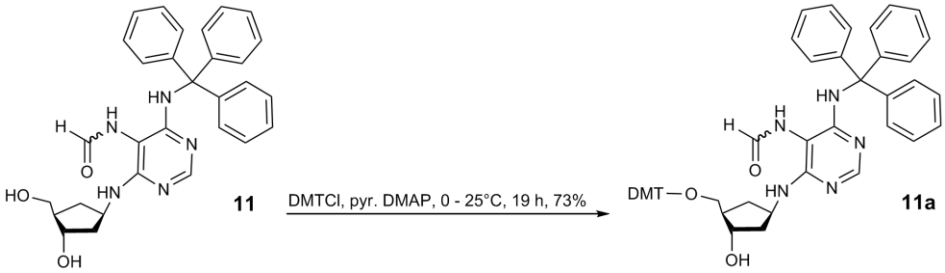
_N_-4-{[(1´_R_,3´_S_,4´_R_)-3´-hydroxy-4´-(hydroxymethyl)cyclopentyl]amino}-5-formylamino-6-tritylamino-pyrimidine 11



The silylether **10a** (3 x 125 mg, 3 x 169 µmol) and dry tetrahydrofurane (3 x 3.5 mL) were given into three sealable polypropylene tubes. Dry pyridine (3 x 131 µL) and HF·pyridine complex (~ 65% HF content, 3 x 131 µL) were added, and the sealed tubes were stirred overnight at room temperature. To the reaction was given methoxytrimethylsilane (3 x 405 µL) and the reaction mixtures were stirred again for 2 h. Afterwards the solvent was evaporated under reduced pressure and the resulting white solid was dried *in vacuo*. The diol was purified *via* column chromatography (chloroform: methanol: 20:1→ 10:1) (243 mg, 477 µmol, 94%). Crystals suitable for X-ray analysis were obtained by dissolving **11** in acetonitrile and appeared after 7 days at rt as colorless crystals (**Supplementary Fig. 1b)**.

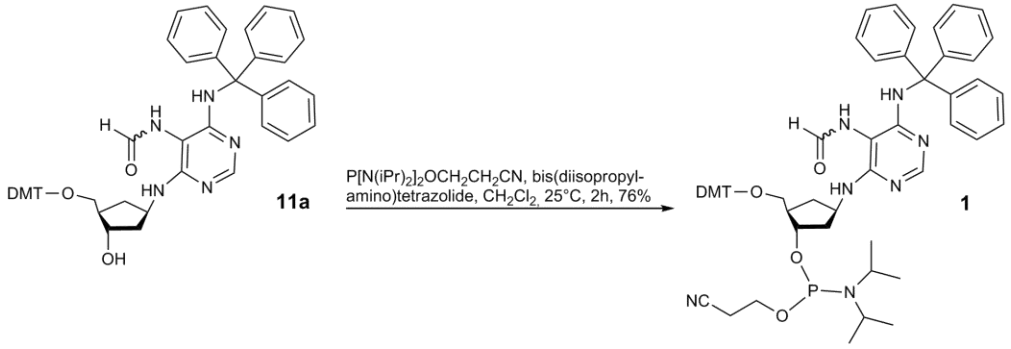
TLC ($CHCl_3$:MeOH, 10:1 v/v): R_f = 0.12; FTIR (ATR): 3423, 2956, 2928, 3856, 2360, 1725, 1700, 1580, 1492, 1470, 1443, 1255, 1117, 1073, 1038, 1004, 938, 874, 834, 773, 745 cm^{-1}; ^{1}H NMR (400 MHz, MeCN–d_3): δ 1.17 (dt, $^2J_{H\text{-}H}$(C5´H_b,C5´H_a) = 13.0 Hz, $^3J_{H\text{-}H}$(C5´H_b,C1´H,C4´H) = 8.3, 1 H, C5´H_b, 1.69 (m, 1 H, C2´H_b), 1.82–1.96 (m, 2 H, C2´H_a,C4´H), 2.20 (dt, $^2J_{H\text{-}H}$(C5´H_a,C5´H_b) = 12.9 Hz, $^3J_{H\text{-}H}$(C5´H_b,C1´H,C4´H) = 8.3, 1 H, C5´H_a, 3.47–3.55 (m, 2 H, C1´´H_b,C1´´H_a), 4.00–4.04 (m, 1 H, C3´H), 4.41–4.50 (m, 1 H, C1´H), 5.35 (d, $^3J_{H\text{-}H}$ (C1´NH,C1´H) = 7.67 Hz, 0.75 H, [*cis*-formamide]C1´NH), 5.95 (d, $^3J_{H\text{-}H}$ (C1´NH,C1´H) = 7.7 Hz, 0.25 H, [*trans*-formamide]C1´NH), 6.99 (d, $^3J_{H\text{-}H}$ (N<u>*H*</u>CHO,NHC<u>*H*</u>O) = 11.4 Hz, 0.25 H [*trans*-formamide]N<u>*H*</u>CHO), 7.18–7.22 (m, 3 H, 3 x *para*-H of trityl), 7.24–7.29 (m, 6 H, 6 x CH-Trityl), 7.31–7.35 (m, 6 H, 6 x CH-Trityl), 7.50 (s, (N<u>*H*</u>CHO,NHC<u>*H*</u>O), 0.75 H [*cis*-formamide]NHC<u>*H*</u>O), 7.51 (s, 0,75 H, [*cis*-formamide]C2H), 7.59 (s, 0.25 H, [*trans*-formamide]C2H), 7.92 (d, $^3J_{H\text{-}H}$(NHC<u>*H*</u>O,N<u>*H*</u>CHO) = 11.75 Hz, 0.25 H, [*trans*-formamide]NHC<u>*H*</u>O), 8.31 (d, $^3J_{H\text{-}H}$(NHC<u>*H*</u>O,N<u>*H*</u>CHO) = 1.16 Hz, 0.75 H, [*cis*-formamide]N<u>*H*</u>CHO), 8.57 (br, 1 H, NH), 2 x OH not observed; ^{13}C NMR (100 MHz, MeCN–d_3): δ 35.76 (C5´), 43.09 (C2´), 50.44 (C4´), 50.94 (C1´), 65.04 (C1´´), 71.73 (C- from Trityl), 74.77 (C3´), 97.61 (C5), 127.89 (CH-Trityl), 128.96 (CH-Trityl), 130.18 (CH-Trityl), 147.35 (CH- from Trityl), 155.11 (C2), 155,90 (C4), 158.27 (C6), 162.65 (<u>C</u>HO); ESI-HRMS (m/z): [M + H^+] calcd. for [$C_{30}H_{31}N_5O_3$+H]$^+$, 510.2500; found, 510.2503.

***N*-4-{[(1´*R*,3´*S*,4´*R*)-4´-{[(Dimethoxytrityl)oxy]methyl}-3´-hydroxycyclopentyl]amino}-5-(formylamino)-6-tritylamino-pyrimidine 11a**



Compound **11** (214 mg, 0.42 mmol) was dissolved two times in dry pyridine (2 x 1 mL) and concentrated *in vacuo* two times. The residue was dissolved again in dry pyridine (8.5 mL) and stirred for 12 h over 4 Å molecular sieves. The solution was then cooled in an ice bath and dimethoxytritylchloride (DMTCl) (86 mg, 0.25 mmol, 0.6 eq) was given to the solution. After another hour additional DMTCl (86 mg, 0.25 mmol, 0.6 eq) was added. The mixture was stirred for another hour at 0 °C and finally for 18 h at rt. The reaction was terminated adding 5 mL of MeOH. The molecular sieves were filtered off and the solvent was removed under reduced pressure. The resulting crude yellow product was purified *via* column chromatography using a $CHCl_3$:MeOH gradient: 200:1 → 100:1 →50:1 → 10:1 with 0.5% pyridine. Compound **11a** was isolated as a white foam (249 mg, 0.31 mmol, 73 %)); mp: 60–61 °C; ($CHCl_3$:MeOH, 20:1 v/v): R_f= 0.29; FTIR (ATR): 3444, 2968, 2915, 2836, 2360, 2338, 1724, 1607, 1508, 1456, 1446, 1285, 1244, 1175, 1152, 1136, 1068, 1030, 982, 903, 861, 822, 791, 726, 704, 651 cm^{-1}; ^{1}H NMR (400 MHz, MeCN–d_3): δ 1.17 (dt, $^2J_{H\text{-}H}$(C5`$H_b$,C5`H_a) = 12.6 Hz, $^3J_{H\text{-}H}$(C5`$H_b$,C1´H,C4´H) = 9.6 Hz, 1 H, C5´$H_b$), 1.69 (ddd, $^2J_{H\text{-}H}$(C2`H_b,C2`$H_a$) = 13.4 Hz, $^3J_{H\text{-}H}$(C2`H_b,C1`H,C5´$H_b$) = 8.1 Hz, $^3J_{H\text{-}H}$(C2`H_b,C1`H,C5´$H_a$) = 7.2 Hz, 1 H, C2´$H_b$), 1.86–1.92 (m, 1 H, C2´$H_a$), 1.99–2.10 (m, 1 H, C4´H), 2.23–2.33 (m, 1 H, C5`H_a), 3.01–3.05 (m, 1 H, C1``$H_b$), 3.07–3.10 (m, 1 H, C1``H_a), 3.77 (s, 6 H, 2 x OCH_3), 3.94–4.03 (m, 1 H, C3´H), 4.43–4.53 (m, 1 H, C1`H), 5.20 (d, $^3J_{H\text{-}H}$ (C1´NH,C1´H) = 7.6 Hz, 0.75 H, [*cis*-Formamide]C1´NH), 5.43 (d, $^3J_{H\text{-}H}$ (C1´NH,C1´H) = 7.7 Hz, 0.25 H, [*trans*-Formamide]C1´NH), 6.09 (s, 1 H, OH), 6.85–6.89 (m, 4 H, 4 x *meta*-H of *para*-CH_3O*Ph*), 6.99 (d, $^3J_{H\text{-}H}$ (N*H*CHO,NHC*H*O) = 12.0 Hz, 0.25 H [*trans*-formamide]N*H*CHO), 7.18–7.23 (m, 5 H, C*H*- of DMT and Trityl), 7.24–7.35 (m, 20 H, C*H*- of DMT and Trityl), 7.49 (s, 0.75 H, [*cis*-formamide]NHC*H*O), 7.53 (s, 1 H, C2H), 7.92 (d, $^3J_{H\text{-}H}$(NHC*H*O,N*H*CHO) = 11.7 Hz, 0.25 H, [*trans*-formamide]NHC*H*O), 8.34 (d, $^3J_{H\text{-}H}$(NHC*H*O,N*H*CHO) = 1.1 Hz, 0.75 H, [*cis*-Formamide]N*H*CHO), 8.58 (br, 1 H, NH); ^{13}C NMR (100 MHz, MeCN-d_3): δ 36.38 (C5´), 42.52 (C2´), 48.55 (C4´), 50.41 (C1´), 55.97 (2 x OCH_3) 66.28 (C1´´), 71.42 (C- from Trityl), 74.31 (C3´), 68.69 (C- from DMT), 97.32 (C5), 114.06 (4 x *meta*-C of *p*-CH_3OPh), 127.55 (*C*H- from Trityl), 128.63 (*C*H- fromTrityl), 128.84 (*C*H- from DMT), 129.13 (*C* from DMT), 129.85 (*C*H- from Trityl), 131.05 (*C*H- from DMT), 131.06 (*C*H- from DMT), 132.29 (*C*H- from DMT), 137.43 (*C*H- from DMT), 137.48 (*C*H- from DMT), 146.54 (*C*H- from DMT), 147.00 (*C*- from Trityl), 155.58 (C2), 157.91 (C4), 158.59 (C6), 159.64 (C- from DMT), 162.29 (*C*HO); ESI-HRMS (m/z): [M + H^+] calcd. for [$C_{51}H_{49}N_5O_5$+H]$^+$, 812.3806; found, 812.3813.

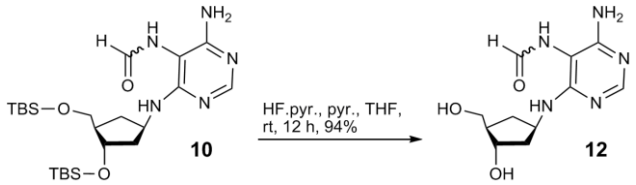
(1´S,2´R,4´R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((5-formamido-6-(tritylamino)-pyrimidin-4-yl)amino)cyclopentyl-(2-cyanoethyl)-diisopropyl-phosphoramidite 1



Under Argon, compound **11a** (228 mg, 0.28 mmol) was dissolved in degassed dry dichloromethane (3 mL) together with diisopropylaminotetrazolate (24.3 mg, 0.14 mmol, 0.5 eq). The reaction mixture was degassed again before cooling in an ice bath and 3-(bis(diisopropylamino)phophanooxy)propanenitrile (100 µL, 0.32 mmol, 1.1 eq) was added and stirred overnight at room temperature. The resulting mixture was dried under reduced pressure. Pure phosphoramidite could be obtained by column chromatography ($CHCl_3$: MeOH 20:1; 0.5 % pyridine) on desactivated silicagel (2 % pyridine). The dried white product was dissolved in a small amount of dry degassed dichloromethane and slowly added degassed *n*-pentane (100 volume equivalents). The resulting white precipitate was isolated decanting off the solvent under argon and dried under high vacuum. The phosphoramidite **1** could be obtained as a white crystalline solid (215 mg, 0.21 mmol, 76%). The dried product could be stored under an argon atmosphere at -20 °C for at least one year without any oxidation. mp: decomp. > 113 °C (decomp.); TLC ($CHCl_3$:MeOH. 10:1 v/v): R_f = 0.68 mixture of stereoisomers); FTIR (ATR): 3130, 2943, 2360, 1711, 1630, 1542, 1483, 1455, 1417, 1368, 1353, 1314, 1266, 1221, 1368, 1174, 1153, 1098, 1050, 1018, 979, 913, 893, 862, 862, 817, 794, 758, 716, 676, 643 cm^{-1}; ^1H NMR (600 MHz, $CDCl_3$): δ 1.00–1.02 (m, 3 H , [isomer 1]PN[CH(CH_3)$_2$]$_2$), 1.08–1.13 (m, 3 H , [isomer 2]PN[CH(CH_3)$_2$]$_2$), 1.17–1.22 (m, 6 H , [isomer 1 + 2]PN[CH(CH_3)$_2$]$_2$), 1.24–1.28 (m, 1 H, C5`$H_b$), 1.54–1.67 (m, 1 H, C2`H_b), 2.13–2.17 (m, 1 H, C2`$H_a$), 2.23–2.30 (m, 1 H, C4`H), 2.38–2.44 (m, 1 H, C5`H_a), 2.53–2.58 (m, 2 H, CH_2CN), 2.97–3.08 (m, 1 H, C1``$H_a$), 3.16–3.21 (m, 1 H, C1``H_b), 3.47–3.61 (m, 2 H, [$(CH_3)_2$C*H*N]$_2$PR_2), 3.67–3.76 (m, 2 H, C$H_2$$CH_2$CN), 3.77–3.79 (m, 6 H, CH_3OPh), 4.08–4.13 (m, 1 H, C3`H), 4.44–4.50 (m, 1 H, C1`H), 6.62 (br s, 1 H, N*H*CHO), 6.85–6.89 (m, 4 H, 4 x *meta*-H of *para*-CH_3O*Ph*), 7.17–7.21 (m, 5 H, C*H*- of DMT and Trityl), 7.23–7.30 (m, 15 H, C*H* of DMT and Trityl), 7.37–7.42 (m, 4 H, C*H*- from of DMT and Trityl), 7.76 (s, 1 H, C2H), 8.09 (br s, 1 H, NHC*H*O), 8.68 (br, 1 H, NH); ^{13}C NMR (150 MHz, $CDCl_3$): δ 20.16, 20.39 (*C*H_2CN), 24.34, 24.44, 24.53, 24.56 (4 x PN[CH(*C*H_3)$_2$]$_2$), 36.09, 36.33 (C5´), 41.12, 41.31 (C2´), 43.04, 43.09 (2 x PN[*C*H(CH_3)$_2$]$_2$), 46.96, 47.01 (C4´), 49.87 (C1´), 55.17, 55.20 (2 x OCH_3), 58.12, 58.25 (O*C*$H_2$$CH_2$CN), 64.14, 64.59 (C1´´), 74.81, 74.92 (C- from Trityl), 75.51, 75.62 (C3´), 85.71, 85.74 (O*C*(*p*-CH_3OPh)$_2$Ph), 85.77, 85.80 (C5), 112.95, 112.97 (4 x *meta*-C from *p*-CH_3OPh), 117.65, 117.66 (*C*N), 123.68 (*C*- from DMT), 123.68–130.11 127.55 (*C*H- from Trityl), 128.63 (*C*H- from Trityl), 128.84 (*C*H- from DMT), 129.13 (*C*H- from DMT), 129.85 (*C*H- from Trityl), 131.05 (*C*H- from DMT), 131.06 (*C*H- from DMT), 132.29 (*C*H- from DMT), 137.43 (*C*H- from DMT), 137.48 (*C*H- from DMT), 146.54 (*C*H-

from DMT), 147.00 (C- from Trityl), 155.58 (C2), 156.59 (C4), 158.34, 158.35 (C6), 160.25 (*C*HO). ^{31}P NMR (200 MHz, $CDCl_3$): δ 148.65, 149.04 (ROP(N(*i*Pr)$_2$)CH_2CH_2CN); ESI-HRMS (m/z): $[MH]^-$ calcd. for $[C_{51}H_{49}N_5O_5H]^-$, 1010.4739; found, 1010.4767.

_N_-(5-(Formylamino)-4-{[(1'*R*,3'*S*,4'*R*)-3'-hydroxy-4'-(hydroxymethyl)cyclopentyl]-amino}-6-aminopyrimidine 12



Compound **10** (78 mg, 0.16 mmol) was dissolved in dry THF (3 mL) with dry pyridine (122 µL) and afterwards pyridine·HF complex (122 µL, 70 %) was added and the mixture was stirred overnight at rt in a sealable polypropylene tube. The diol precipitated as a white fine solid, which was centrifuged afterwards. The isolated product was resuspended in dry ethyl acetate (1 mL) and afterwards stirred for another two hours together with 370 µL MeOTMS. The product was again centrifuged, and to the decanted supernatant were given another was 370 µL of MeOTMS and the solution was stirred for another hour. Both fractions were combined and the solvent was removed under reduced pressure. The resulting white solid was dried under high vaccum yielding 40.0 mg (0.15 mmol)of **12** (94 %). mp: decomp.> 147 °C; TLC (Reversed Phase; MeOH:H_2O, 2:1 v/v): R_f = 0.88; FTIR (ATR): 3157, 2957, 2879, 2680, 2499, 1669, 1646, 1635, 1595, 1503, 1467, 1449, 1441, 1416, 1374, 1331, 1285, 1270, 1249, 1197, 1159, 1151, 1136, 1077, 1035, 1004, 859, 825, 802, 765, 752, 731, 707, 682 cm^{-1}; ^{1}H-NMR (400 MHz, MeOH-d_4): δ 1.30–1.37 (m, 1 H C5´H_b), 1.82–2.02 (m, 3 H, C4´H,C2´H_b,C2´H_a), 2.11 (dt, $^2J_{H\text{-}H}$ (C5´H_a,C5´H_b) = 13.3 Hz, $^3J_{H\text{-}H}$ (C5´H_a,C1´H,C4´H) = 7.7 Hz, 1 H, C5´H_a), 3.53–3.58 (m, 2 H, C1´´H_b,C1´´H_a), 4.05–4.09 (m, 1 H, C3´H), 4.56 – 4.63 (m, 1 H, C1´H), 8.02 (br s, 1 H, N*H*CHO), 8.24 (br s, 1 H, C*H*O), 8.50 (s, 1 H, C2H), 2 x OH, C1´N*H*, 2 x NH not detected; ^{13}C NMR (100 MHz, MeOH-d_4): δ 35.48(C5´), 42.61 (C2´), 50.21 (C4´), 51.67 (C1´), 64.59 (C1´´), 74.34 (C3´), 93.88 (C5), 149.62 (C2), 155.96 (C4), 159.32 (C6), 164.56 (*C*HO); ESI-HRMS (m/z): $[MH]^+$ calcd. for $[C_{11}H_{17}N_5O_3H]^+$, 268.1404; found, 268.1404.

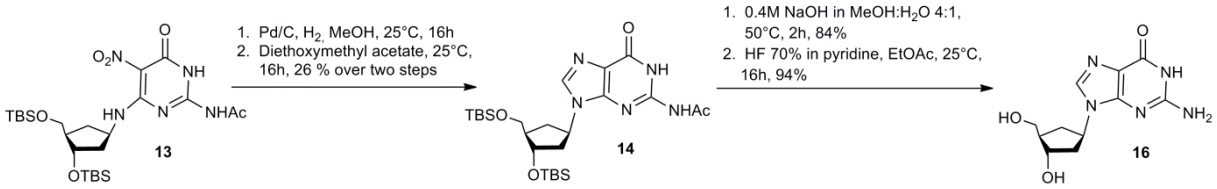
X-Ray crystallographic structures of compound 9a and 11

Single crystals of **9a** and **11** suitable for X-ray diffraction were carefully mounted on the top of a thin glass wire. Data collection was performed with an Oxford Xcalibur3 diffractometer equipped with a Spellman generator (50 kV, 40 mA) and a Kappa CCD detector, operating with Mo Kα radiation (λ = 0.71071 Å). Data collection was performed with the Crysalis CCD software, for the data reduction the Crysalis RED software[1] was used. Absorption correction using the SCALE3 ABSPACK multiscan method[2] was applied. Structures were solved an refined with SHELX-97[3] and finally checked using PLATON[4]. Details for data collection and structure refinement are summarized in the corresponding cif files. CCDC-796512 (**9a**) and

CCDC-796513 (**11** 2 MeCN) contain the supplementary crystallographic data. The data can be obtained free of charge from the Cambridge Crystallographic Data Centre *via* www.ccdc.cam.ac.uk/data_request/cif.

Overview over the synthesis of carbocyclic 2´-deoxyguanosine

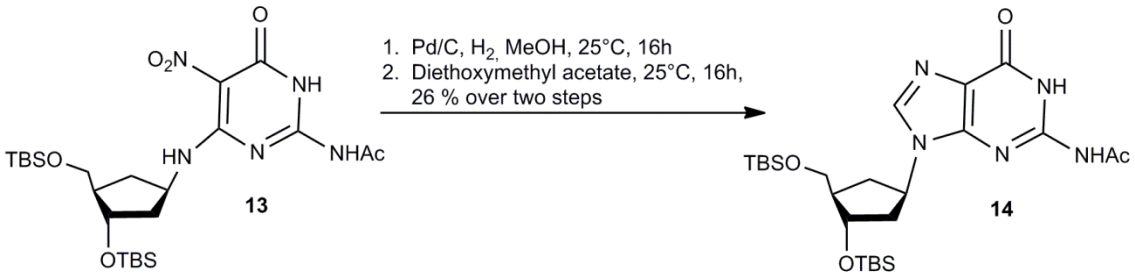
The synthesis of the carbocyclic 2´-deoxyguanosine (**Supplementary Fig. 5**) starts from **13** with reduction of the nitro group with Pd/C under hydrogen and reaction of the resulting amine with diethoxymethyl acetate to give the protected purine **14** in a 26 % yield. For the acetyl deprotection **14** was stirred in 0.4 M NaOH to give **15** in a 84 % yield, followed by TBS deprotection in HF/pyridine with 94 % yield of **16**.



Supplementary Figure 5: Synthesis of carbocyclic dG (**16**).

Detailed synthesis of carbocyclic 2´-deoxyguanosine

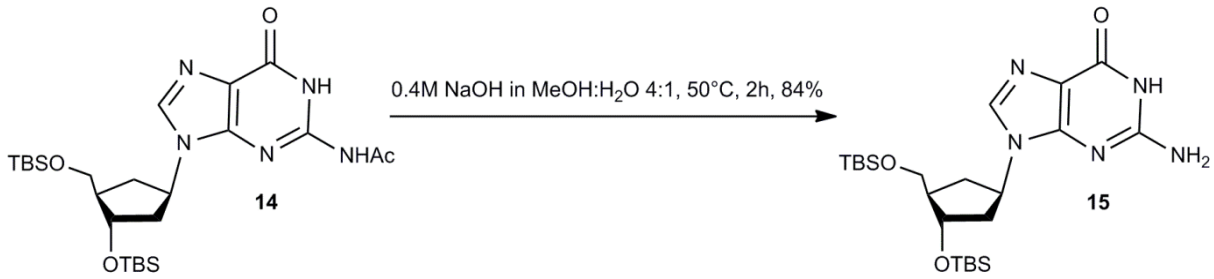
***N*-[9-[(1´*R*,3´*S*,4´*R*)-3´-[(*tert*-Butyldimethylsilyl)oxy]-4´-[[(*tert*-butyldimethylsilyl)oxy]-methyl]cyclopentyl]-6-oxo-6,9-dihydro-1*H*-purin-2-yl]-acetamide 14**



13 (1.60 g, 2.88 mmol) was dissolved in dry methanol (30 mL). 700 mg 10 % Pd/C were added and the reaction mixture was degassed carefully. The argon atmosphere was changed into a hydrogen atmosphere and the mixture was stirred over night under hydrogen (1.2 bar) at rt. The charcoal and palladium were filtered off under exclusion of oxygen, using a a 0.2 µm nylon syringe filter (Millipore^{TM}). The solvent was evaporated under reduced pressure at rt and the air sensitive amine was diluted in diethoxymethyl acetate (16 mL). The mixture was stirred at rt for 18 h and then heated up to 100 °C for 1 h. The cooled solution was diluted with ethyl acetate (200 mL) and washed with brine (3 x 250 mL). The organic layer was dried with $MgSO_4$ and evaporated under reduced pressure. The crude reaction mixture was purified by column chromatography ($CHCl_3$:MeOH 80:1 → 40:1). The resulting purine **14** could be

obtained as an orange foam (400 mg, 0.75 mmol, 26 % over two steps). TLC ($CHCl_3$:MeOH, 10:1 v/v): R_f = 0.30; FTIR (ATR): 3155, 2952, 2927, 2855, 2739, 17112, 1680, 1611, 1553, 1471, 1463, 1397, 1375, 1361, 1321, 1251, 1116, 10889, 1051, 1004, 938, 834, 813, 775, 742, 724, 668 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$): δ 0.04 (s, 3H, Si-CH_3), 0.05 (s, 3H, Si-CH_3), 0.06 (s, 6H, Si-CH_3), 0.87 (s, 9H, Si-C(CH_3)$_3$), 0.89 (s, 9H, Si-C(CH_3)$_3$), 1.71 – 1.81 (dt, $^2J_{H\text{-}H}$ (C5´H_a, C5´H_b) = 13.0 Hz, $^3J_{H\text{-}H}$ (C5´H_a, C1´H, C4´H) = 9.3 Hz, 1H, C5´H_a), 2.07 – 2.16 (m, 3H, C2´H_a, C2´H_b, C4´H), 2.30 (s, 3H, acetyl-CH_3), 2.36 – 2.45 (dt, $^2J_{H\text{-}H}$ (C5´H_b, C5´H_a) = 13.0 Hz, $^3J_{H\text{-}H}$ (C5´H_b, C1´H, C4´H) = 8.1 Hz, 1H, C6´H_b), 3.57 – 3.62 (dd, $^2J_{H\text{-}H}$ (C1´´H_a, C1´´H_b) = 10.1 Hz, $^3J_{H\text{-}H}$ (C1´´H_a, C1´´OH) = 4.7 Hz, 1H, C1´´H_a), 3.66 – 3.72 (dd, $^2J_{H\text{-}H}$ (C1´´H_b, C1´´H_a) = 10.1 Hz, $^3J_{H\text{-}H}$ (C1´´H_b, C1´´OH) = 4.5 Hz, 1H, C1´´H_b), 4.28 – 4.33 (m, 1H, C3´H), 4.85 – 4.94 (p, $^2J_{H\text{-}H}$ (C1´H, C2´H_a, C2´H_b, C5´H_a, C5´H_b) = 8.6 Hz, 1H, C1´H), 7.75 (s, 1H, C8H), 8.96 (s, 1H, NH), 11.95 (s, 1H, NH). ^{13}C NMR (101 MHz, $CDCl_3$): δ -5.46 (Si-CH_3), -5.42 (Si-CH_3), -4.75 (Si-CH_3), -4.62 (Si-CH_3), 17.98 (Si-$\underline{C}$(CH_3)$_3$), 18.34 (Si-$\underline{C}$(CH_3)$_3$), 24.43 (acetyl-$\underline{C}$$H_3$), 25.79 (Si-C($\underline{C}$$H_3$)$_3$), 25.93 (Si-C($\underline{C}$$H_3$)$_3$), 33.36 (C5´), 41.58 (C2´), 49.73 (C4´), 53.18 (C1´), 62.93 (C1´´), 72.84 (C3´), 121.55 (C5), 137.16 (C8), 146.74 (C2), 148.16 (C4), 155.68 (C6), 171.52 (acetyl-C). ESI-HRMS (m/z): $[MH]^-$, calcd. for $[C_{25}H_{46}N_5O_4Si_2H]^-$: 536.3083; found, 536.3075.

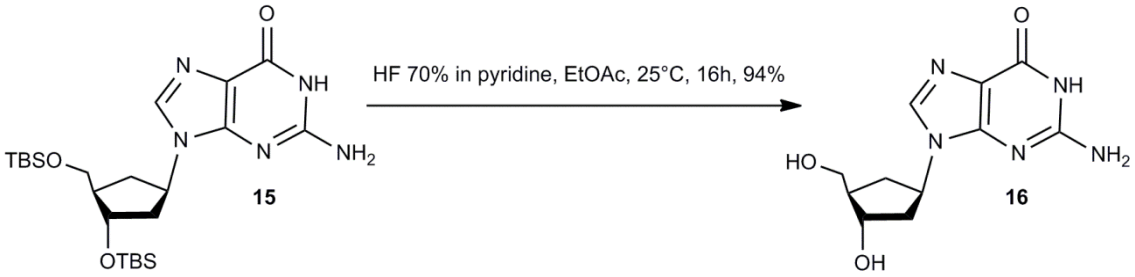
2-Amino-9-[(1´*R*,3´*S*,4´*R*)-3´-[(*tert*-butyldimethylsilyl)oxy]-4´-[[(*tert*-butyldimethylsilyl) oxy]-methyl]cyclopentyl]-1,9-dihydro-purine-6-one 15



14 (250 mg, 0.47 mmol) was dissolved in 0.4 M NaOH in MeOH:H_2O 4:1 and stirred for 2 h at 50 °C until the reaction was complete. The mixture was neutralized with acetic acid and evaporated under reduced pressure. The crude residue was purified by column chromatography (CH_2Cl_2:MeOH 10:1). The resulting purine **15** could be obtained as a yellow foam (195 mg, 0.39 mmol, 84 %). TLC (CH_2Cl_2:MeOH 10:1 v/v): R_f = 0.31; FTIR (ATR): 3317, 3156, 2952, 2928, 2855, 1695, 1627, 1597, 1569, 1536, 1471, 1361, 1253, 1164, 1111, 1005, 938, 834, 813, 775, 669 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$): δ 0.06 (s, 3H, Si-CH_3), 0.06 (s, 3H, Si-CH_3), 0.06 (s, 3H, Si-CH_3), 0.07 (s, 3H, Si-CH_3), 0.89 (s, 9H, Si-C(CH_3)$_3$), 0.90 (s, 9H, Si-C(CH_3)$_3$), 1.69 – 1.79 (dt, $^2J_{H\text{-}H}$ (C5´H_a, C5´H_b) = 13.0 Hz, $^3J_{H\text{-}H}$ (C5´H_a, C1´H, C4´H) = 9.4 Hz, 1H, C5´H_a), 1.94 – 2.29 (m, 3H, C2´H_a, C2´H_b, C4´H), 2.37 – 2.47 (dt, $^2J_{H\text{-}H}$ (C5´H_b, C5´H_a) = 12.9 Hz, $^3J_{H\text{-}H}$ (C5´H_b, C1´H, C4´H) = 7.9 Hz, 1H, C5´H_b), 3.60 – 3.65 (m, 1H, C1´´H_a), 3.66 – 3.72 (m, 1H, C1´´H_b), 4.29 – 4.34 (dt, $^3J_{H\text{-}H}$ (C3´H, C2´H) = 6.1 Hz, $^3J_{H\text{-}H}$ (C3´H, C4´H) = 4.5 Hz, 1H, C3´H), 4.83 – 4.96 (p, $^2J_{H\text{-}H}$ (C1´H, C2´H_a, C2´H_b, C5´H_a, C5´H_b) = 8.5 Hz, 1H, C1´H), 6.23 (s, 2H, NH_2), 7.64 (s, 1H, C8H), 11.92 (s, 1H, NH). ^{13}C NMR (101 MHz, $CDCl_3$): δ -5.43 (Si-CH_3), -5.40 (Si-CH_3), -4.74 (Si-CH_3), -4.59 (Si-CH_3), 18.00 (Si-$\underline{C}$(CH_3)$_3$), 18.35 (Si-$\underline{C}$(CH_3)$_3$), 20.95, 25.82 (Si-C($\underline{C}$$H_3$)$_3$), 25.95 (Si-C($\underline{C}$$H_3$)$_3$), 33.30 (C5´), 41.29 (C2´), 49.80 (C4´), 52.92 (C1´), 63.13 (C1´´), 72.85 (C3´), 76.68, 117.17

(C5), 136.13 (C8), 151.72 (C2), 153.21 (C4), 159.11 (C6). ESI-HRMS (m/z): $[MH]^-$ calcd. for $[C_{23}H_{44}N_5O_3Si_2H]^-$, 494.2977; found, 494.2974.

2-Amino-9-((1´R,3´S,4´R)-3-hydroxy-4-(hydroxymethyl)cyclopentyl)-1,9-dihydro-6H-purin-6-one 16



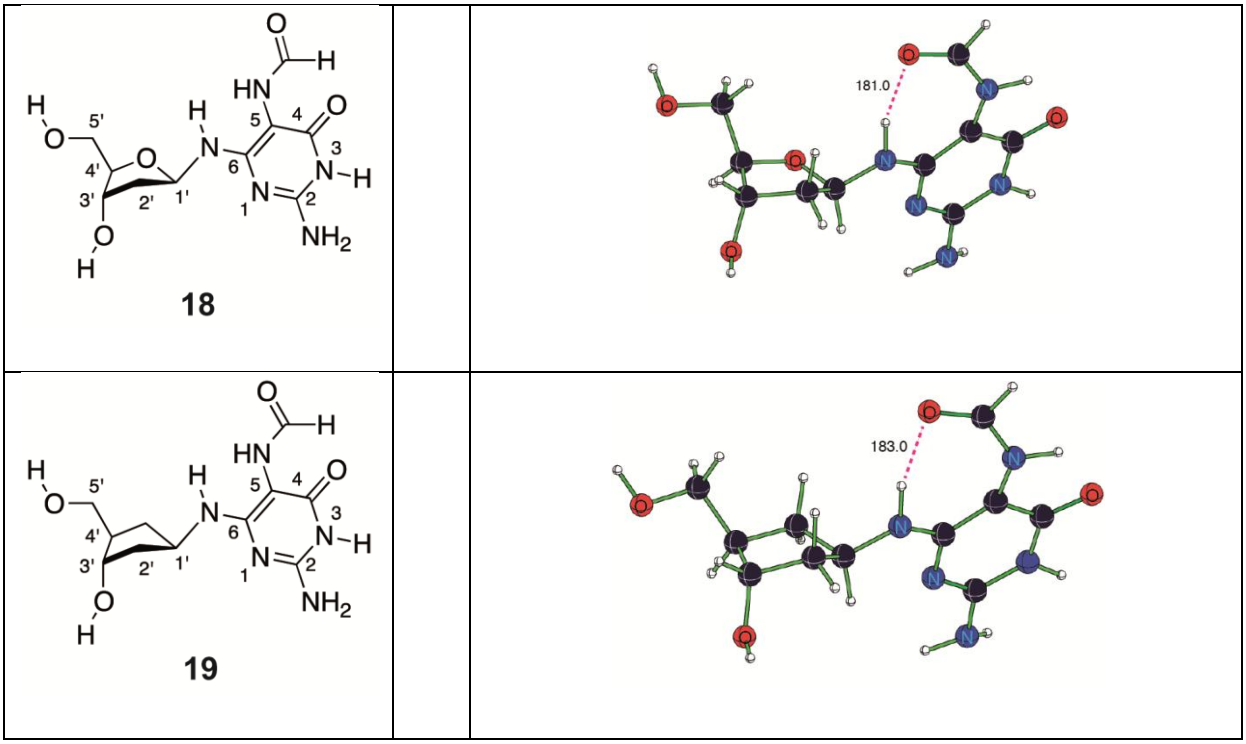
15 (180 mg, 0.36 mmol) was dissolved in ethyl acetate (9.6 mL) and HF (70 % in pyridine; 0.36 mL) was added and stirred over night. The reaction was quenched with methoxytrimethylsilane (2 mL). After evaporation under reduced pressure, the crude reaction mixture was purified by column chromatography (CH_2Cl_2:MeOH 20:1 → 5:1). The carbocyclic deoxyguanosine **16** could be obtained as a white powder (90 mg, 0.34 mmol, 94 %). TLC (CH_2Cl_2:MeOH, 2:1 v/v): R_f = 0.29; FTIR (ATR): 3497, 3320, 3214, 2730, 2362, 2335, 1727, 1629, 1569, 1538, 1488, 1411, 1380, 1380, 1322, 1252, 1177, 1158, 1068, 1047, 1033, 1006, 916, 849, 776, 673 cm^{-1}; ^1H NMR (599 MHz, DMSO-d_6): δ 1.55 – 1.61 (dt, $^2J_{H\text{-}H}$ (C5´H_b, C5´H_a)= 12.8 Hz, $^3J_{H\text{-}H}$ (C5´H_b, C1´H, C4´H) = 9.4 Hz, 1H, C5´H_b), 1.92 – 1.99 (m, 2H, C2´H_b, C4´H), 2.05 – 2.12 (ddd, $^2J_{H\text{-}H}$ (C2´H_a, C2´H_b)= 12.8 Hz, $^3J_{H\text{-}H}$ (C2´H_a, C3´H) = 10.0 Hz, (C2´H_a, C1´H) = 6.3 Hz, 1H, C2´H_a), 2.25 – 2.30 (m, 1H C5´H_a), 3.38 – 3.45 (m, 1H, C1´´H_a), 3.47 – 3.54 (m, 1H, C1´´H_b), 4.02 – 4.08 (m, 1H, C3´H), 4.60 – 4.65 (t, $^2J_{H\text{-}H}$ (C1´´OH, C1´´H_a, C1´´H_b) = 5.2 Hz, 1H, C1´´OH), 4.73 – 4.77 (d, $^2J_{H\text{-}H}$ (C3´OH, C3´H) = 4.0 Hz, 1H, C3´OH), 4.76 – 4.85 (tt, $^3J_{H\text{-}H}$ (C1´H, C2´H_b, C5´H_b) = 9.9 Hz, $^3J_{H\text{-}H}$ (C1´H, C5´H_a) = 7.6 Hz, 1H, C1´H), 6.42 (s, 2H, NH_2), 7.79 (s, 1H, C8H), 10.57 (s, 1H, NH). ^{13}C NMR (101 MHz, DMSO- d_6): δ 34.06 (C5´), 40.66 (C2´), 49.23 (C4´), 52.06 (C1´), 62.77 (C1´´), 71.50 (C3´), 116.74 (C5), 135.35 (C8), 151.07 (C2), 153.21 (C4), 156.80 (C6). ESI-HRMS (m/z): $[MH]^-$ calcd. for $[C_{11}H_{16}N_5O_3H]^-$, 266.1248; found, 266.1247.

1. **CrysAlis RED, O.D.L., Version 1.171.27p5 beta (release 01-04-2005 CrysAlis171.NET) (compiled Apr 1 2005, 17:53:34).**
2. **ABSPACK, S. An Oxford Diffraction program (1.0.4, gui:1.0.3) (C) Oxford Diffraction Ltd. (2005).**
3. **Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr A* 64, 112-22 (2008).**
4. **Spek, A.L. PLATON, A Multipurpose Crystallographic Tool, Utrecht University, Utrecht, The Netherlands,. (1999).**

Supplementary Note 2

Theoretical studies

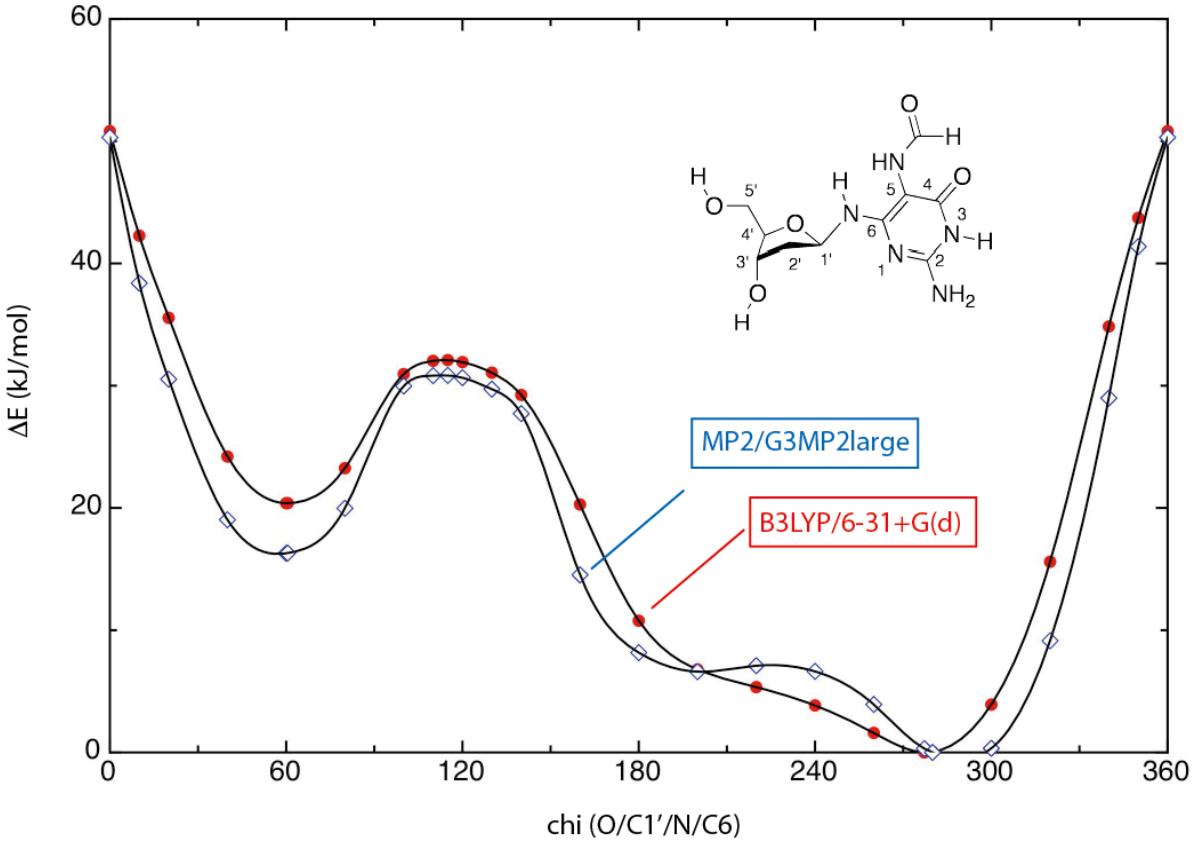
In order to identify possible differences in conformational preferences between 2,6-diamino-4-hydroxy-5-formamidopyrimidine of 2'-desoxyguanosine (**18**) and its cyclopentane analogue **19**, a conformational search for both systems has been conducted at the B3LYP/6-31+G(d) level of theory. This method has recently been used by Schlegel *et al.* in a series of theoretical studies on the formation of damaged nucleobases.[1] Thermal corrections to enthalpies at 298.15 K have been calculated at the same level using the rigid rotor/harmonic oscillator model. Single point energies have subsequently been calculated at the MP2(FC)/G3MP2large level of theory and combined with thermal corrections obtained at B3LYP/6-31+G(d) level in order to calculate enthalpies at 298.15 K. The MP2(FC)/G3MP2large calculations form an integral part of the G3(MP2)B3 scheme developed for the calculation of accurate thermo-chemical data and the G3MP2large basis set corresponds effectively to a 6-311++G(2df,2p) basis for the systems studied here.[2] All calculations have been performed with *Gaussian 09*.[3] In order to avoid hydrogen bonding interactions between the C3' and C5' hydroxy groups with the nucleobase, the orientation of these former groups was chosen such that the O-H bonds point away from the central furan/pentane ring system. The most favorable conformations found in the gas phase are shown in **Supplementary Figure 8** for both systems.



Supplementary Figure 1. 2D and 3D structures of formamidopyrimidine desoxynucleoside **18** and its carbocyclic analogue **19**.

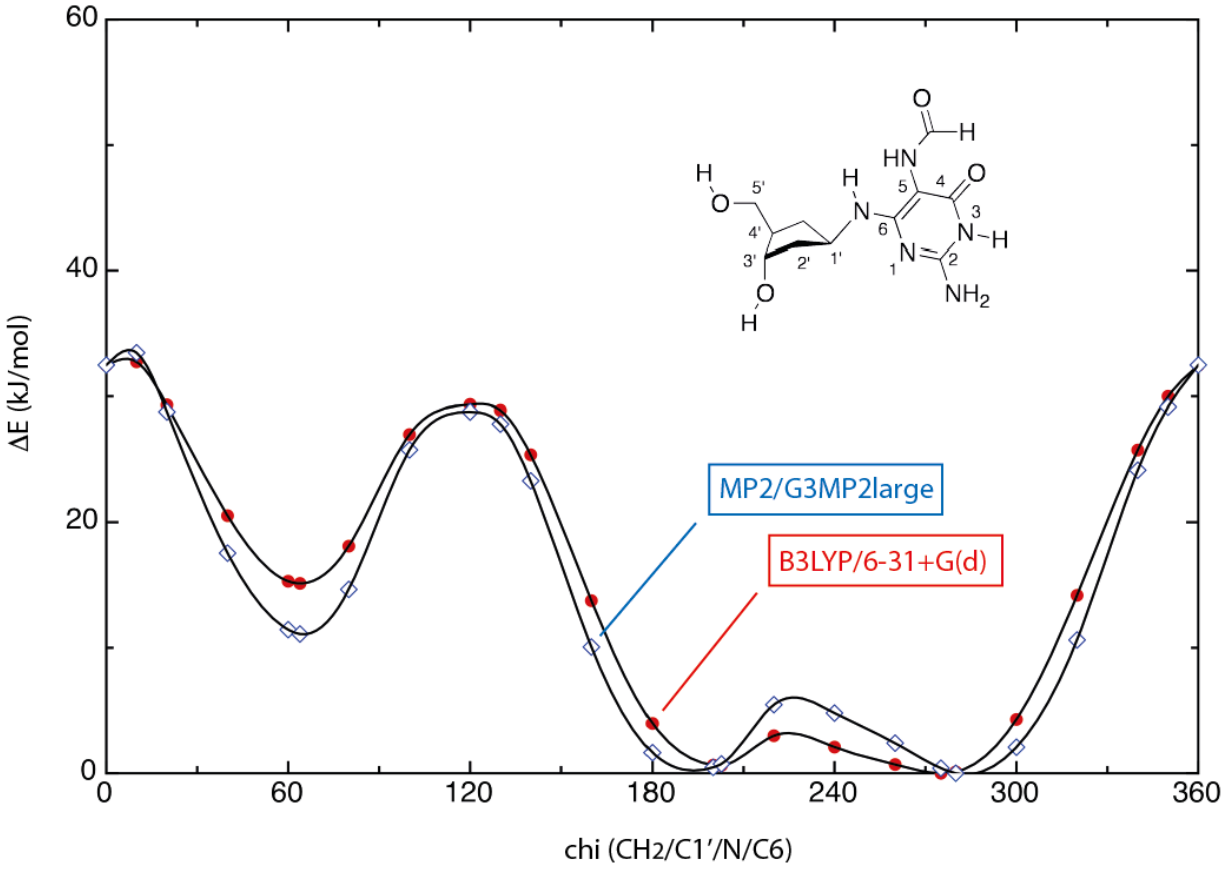
It is noteworthy that, in the absence of the interaction partners present in a polar solvent or a protein surrounding, the formamido groups form intramoleculare hydrogen bonds to the anomeric amino N-H bond. The rotation angle *chi*[4,5] around the C-N bond defined by the centers O/C1'/N/C6 in **18** amounts to *chi*(**18**) = -82.7°. In carbocyclic analog **19** an almost identical rotational angle of *chi*(**19**) = -84.9° is found.

The rotational profile in **18** was explored starting from the most favorable conformation shown in **Supplementary Figure 1** driving the *chi* angle with a step size of 20° in positive and negative directions. In selected regions additional points were added with a step size of 10°. The resulting rotational profile around the C1'-N bond is surprisingly similar at B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory (**Supplementary Fig. 2**). Minor differences concern the flat region between 180 and 210°, where MP2 theory hints at formation of a shallow minimum, while B3LYP theory does not.



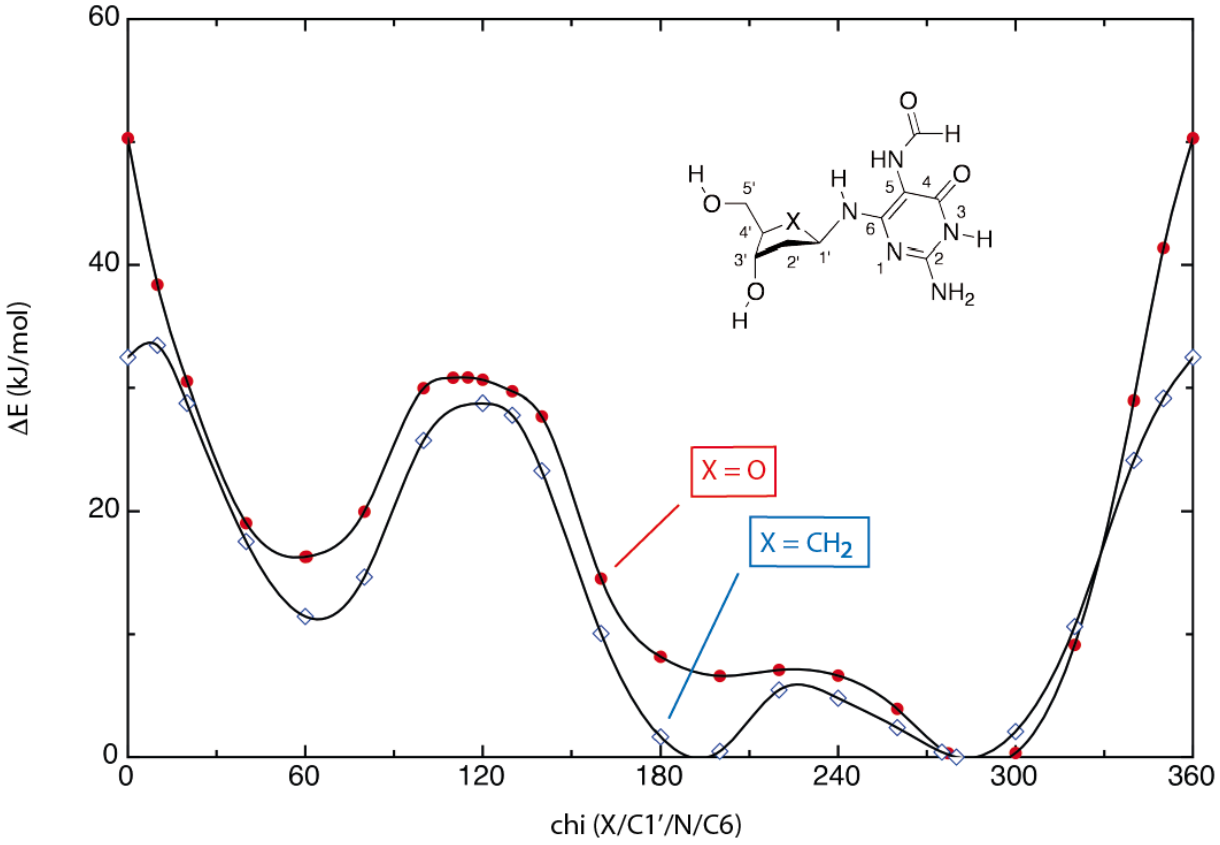
Supplementary Figure 2: Rotational profile for dihedral angle *chi* (O/C1'/N/C6) in formamidopyrimidine **18** at the B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory.

In analogy to the procedure chosen for **18** the rotational profile in carbocyclic analogue **19** was explored starting from the most favorable conformation shown in **Supplementary Figure 1** and a step size of 20°. Additional data points were again added in seletced regions. It is again found that the rotational potential is largely similar at B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory **(Supplementary Fig. 3)**. Minor differences again concern the region with *chi* angles from 180 - 210°, where MP2 theory predicts a more shallow potential energy surface. In contrast to heterocyclic reference system **18**, however, both methods agree in that a true minimum exists at a rotational angle of 203°, whose energy is almost identical to the global rotational minimum at chi = +275°.



Supplementary Figure 3: Rotational profile around angle *chi* (CH_2/C1'/N/C6) in carbocyclic formamidopyrimidine **19** at the B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory.

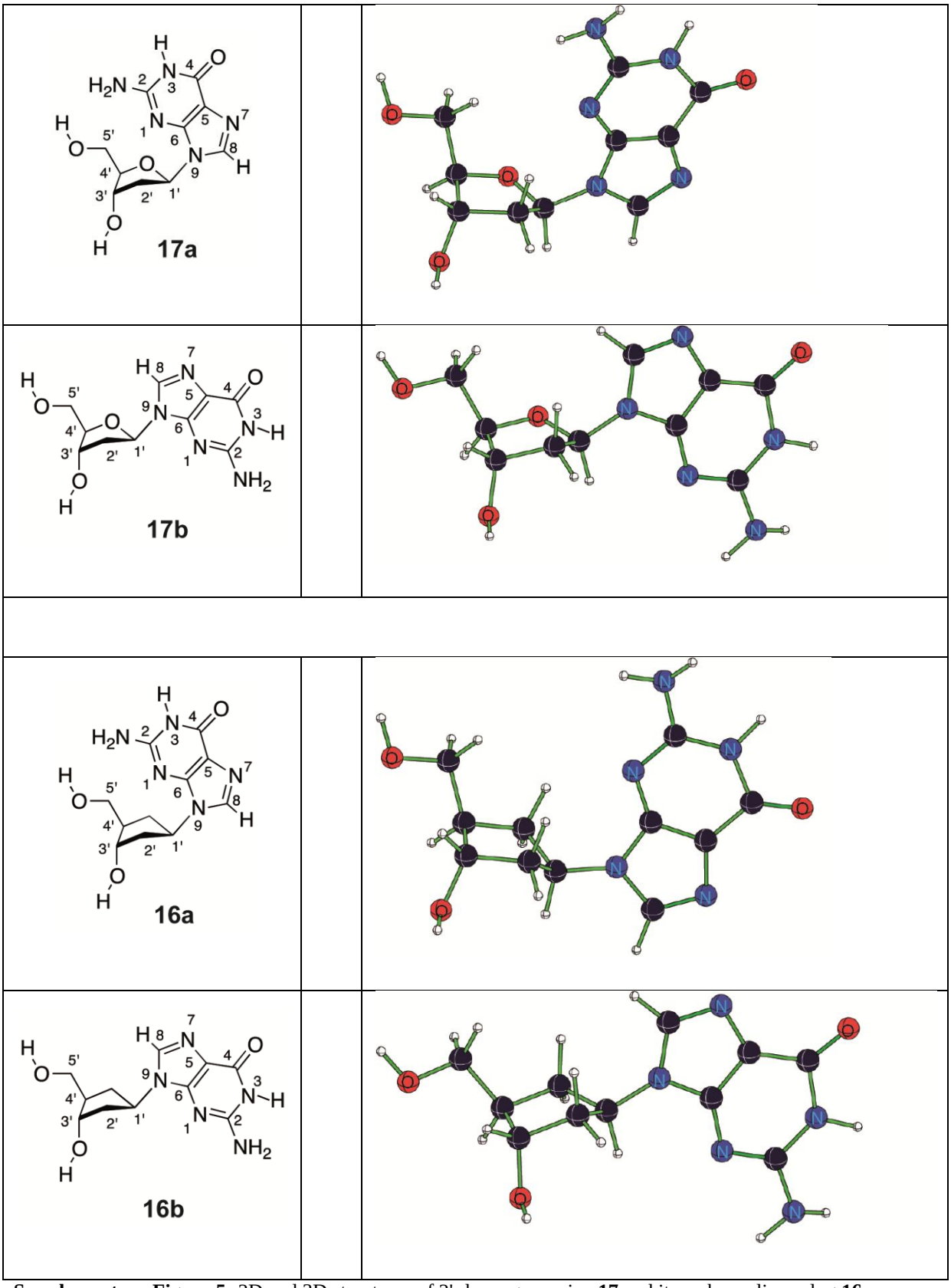
A direct comparison of the rotational potentials for heterocyclic formamidopyrimidine **18** and its carbocyclic analogue **19** as calculated at the MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level is shown in **Supplementary Figure 4**. Both rotational potentials are largely similar in that a global minimum exists at *chi* angles around 275^o and a second, less stable minimum at *chi* values around 60^o. The rotational potentials differ in the magnitude of the rotational barrier through the fully eclipsed conformation at chi = 0^o, and also in the flat region with chi = 180 - 210^o, where a true minimum exists for **19**, but (apparently) not for **18**.



Supplementary Figure 4: Rotational profile around angle *chi* (X/C1'/N/C6) in formamidopyrimidine **18** (red dots) and its carbocyclic analogue **19** (blue diamonds) at the MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level of theory.

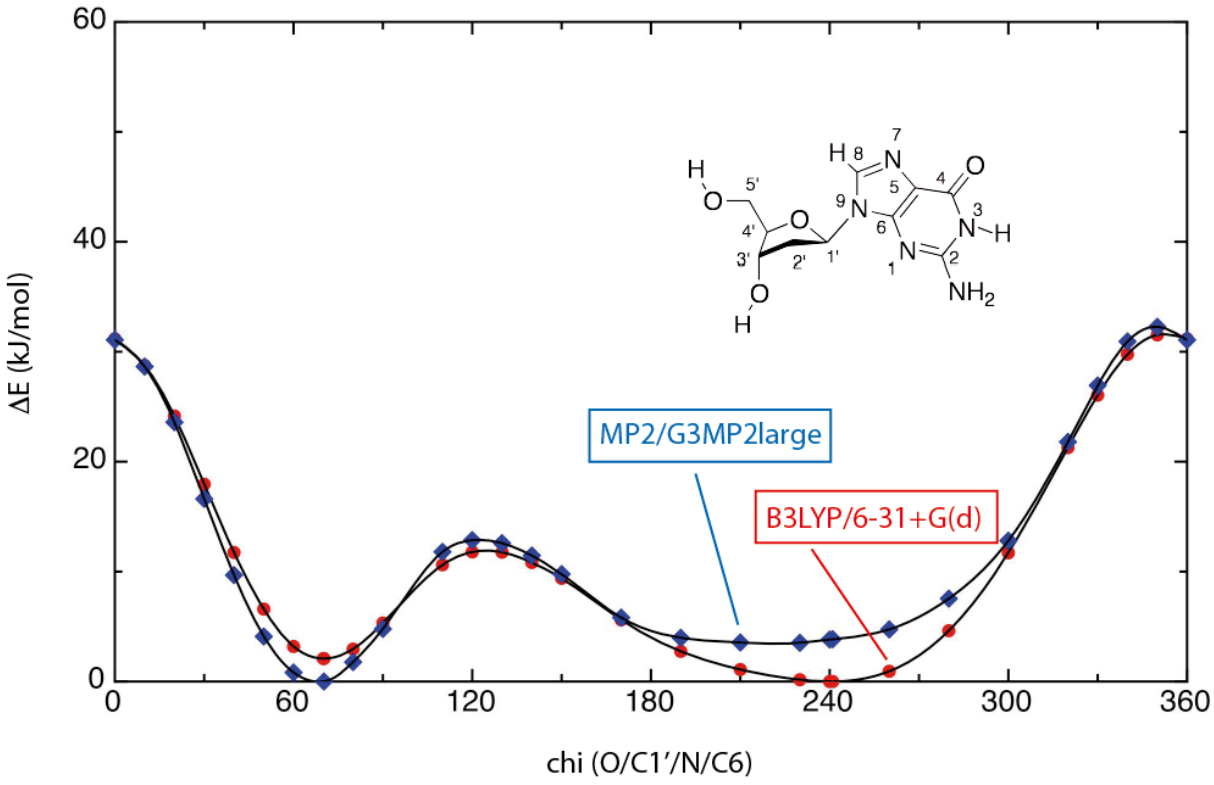
The surprisingly similar rotational potentials for **18** and **19** identified here are closely similar to those obtained at the B3LYP/6-31G(d) level for model systems carrying methyl protecting groups on the C3' and C5' hydroxy groups.[6] This implies that the conformational similarity between the hetero- and carbocyclic systems **18** and **19** is neither dependend on the exact details of the model systems chosen for study nor on the particular level of theory.

How much of the similarity between hetero- and carbocyclic formamidopyrimidine derivatives is retained in the undamaged nucleosides was subsequently explored for 2'-deoxyguanosine **17** and its carbocyclic analogue **16** using the same theoretical approach as before. In order to avoid hydrogen bonding interactions between the C3' and C5' hydroxy groups with the nucleobase, the orientation of these former groups was again chosen such that the O-H bonds point away from the central furan/pentane ring system. The most favorable conformations found for these systems in the gas phase at MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level are shown in **Supplementary Figure 5**.



Supplementary Figure 5: 2D and 3D structures of 2'-deoxyguanosine **17** and its carbocyclic analog **16**.

In contrast to the formamidopyrimidine derivatives studied before, two conformations **17a** and **17b** of similar stability can be identified for 2'-deoxyguanosine. The more stable conformation **17a** with *chi* = +70.2^{o} orients the guanine base such that the N1 nitrogen atom points towards the ribose ring, while the less stable conformation **17b** with *chi* = -119.0^{o} orients the guanine base in the opposite direction. This latter conformer is less stable than **17a** by 3.70 kJ/mol (MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level). The relative stability of these conformers was reconfirmed at full G3(MP2)B3 level, at which the energy difference amounts to 2.9 kJ/mol. A rather similar situation is found for the carbocyclic analogue **16** with the more favourable conformer **16a** (*chi* = +61.5^{o}) and the slightly less stable conformer **16b** (*chi* = -118.1^{o}). The energy difference is again rather small with +3.9 kJ/mol at MP2(FC)/G3MP2large//B3LYP/6-31+G(d) and +3.0 kJ/mol at full G3(MP2)B3 level.

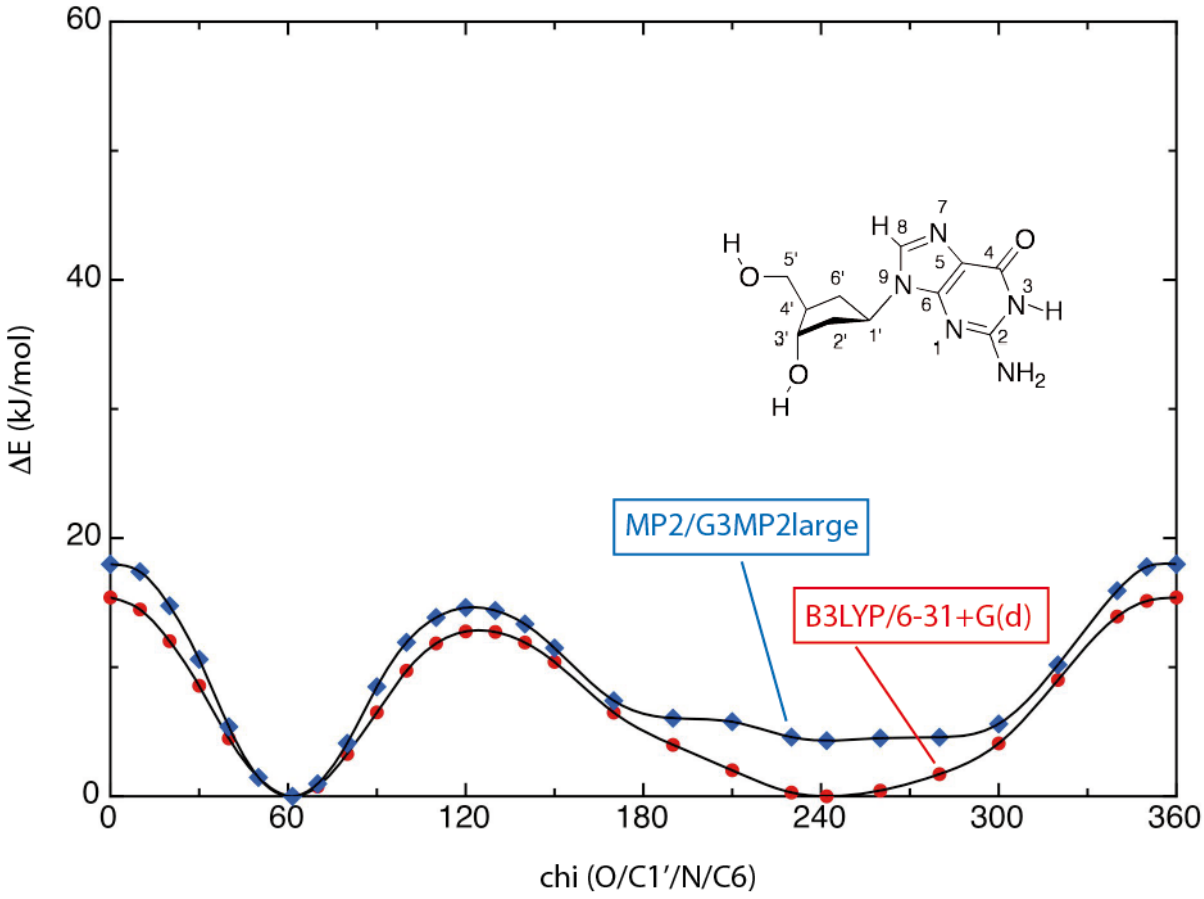


Supplementary Figure 6: Rotational profile for dihedral angle *chi* (O/C1'/N9/C6) in 2'-deoxyguanosine **17** at the B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory.

The rotational profile in **17** was explored starting from the local minimum shown in **Supplementary Figure 5** driving the *chi* angle with a step size of 20^{o} in positive and negative directions. In selected regions additional points were added with a step size of 10^{o}. The resulting

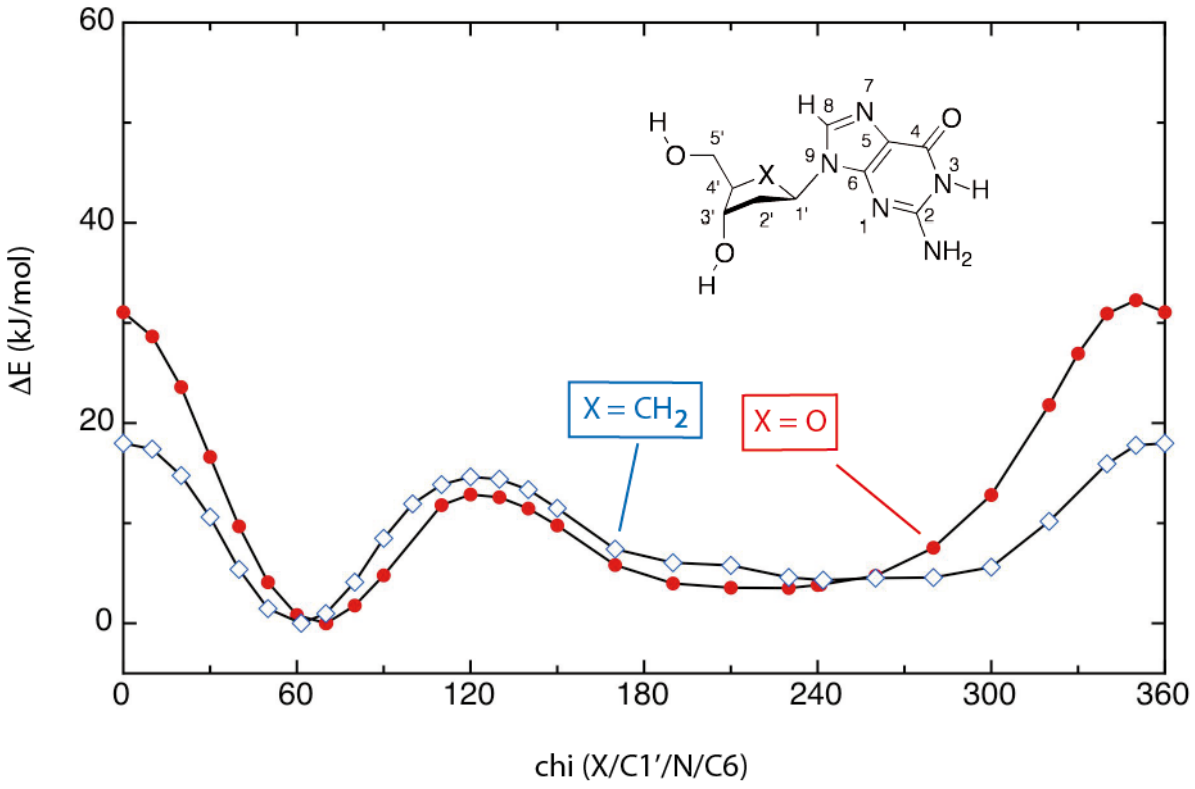
rotational profile around the C1'-N9 bond is again quite similar at B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory (**Supplementary Figure 6**). Minor differences again concern the flat region between 180 and 260^{o}, where MP2 theory places the minimum at somewhat smaller angles (at around 210^{o}) as compared to the B3LYP minim at 240^{o}. The rotational barriers are also quite similar at both levels of theory, the larger barrier of just over 30 kJ/mol leading through the structure with fully eclipsed O/C/N/C dihedral angle. This is significantly lower as compared to the FaPydG derivative, where the corresponding rotational barrier is more than 20 kJ/mol higher (**Supplementary Figure 2**). A second major difference between FaPydG derivative **18** and 2'-deoxyguanosine **17** concerns the relative energies of the *gauche* conformations at rotational angles around -80 and +60 degrees. While the *gauche* conformation at -80^{o} is the only low-.energy conformation in FaPydG derivative **18**, both structures are practically isoenergetic in 2'-deoxyguanosine **17** (**Supplementary Figure 2** vs. **Figure 6**).

Repeating the above approach for carbocyclic analogue **16** and starting again from the structure shown in **Supplementary Figure 4** yields the rotational profile shown in **Supplementary Figure 7**. The latter is quite similar to that of the parent heterocyclic system **17** and displays a larger shallow region between dihedral angles of 190 to 270^{o} and a more clearly defined *gauche* conformation with dihedral angle of *chi*(**16**) = +61.5^{o}. Relative energies for these two minima are rather small at the B3LYP/6-31+G(d) level used for geometry optimizations, but differ slightly at MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level at which the *gauche* conformation at around 60^{o} is energetically preferred by 4.3 kJ/mol.



Supplementary Figure 7: Rotational profile around angle *chi* (CH_2/C1'/N/C6) in carbocyclic 2'-deoxyguanosine **16** at the B3LYP/6-31+G(d) and MP2(FC)/G3MP2large//B3LYP/6-31+G(d) levels of theory.

A direct comparison of the rotational potentials for 2'-deoxyguanosine **17** and its carbocyclic analogue **16** as calculated at the MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level is shown in **Supplementary Figure 8**. Both rotational potentials are largely similar in that a global minimum exists at *chi* angles around 60° and a second, less stable minimum at *chi* values around 180 - 260°. The rotational potentials differ mainly in the magnitude of the rotational barrier through the fully eclipsed conformation at chi = 0°, which is higher in the heterocyclic parent system **17** by approx. 14.3 kJ/mol as compared to carbocyclic analogue **16**. This barrier difference is most likely due to unfavorable electrostatic interactions between the N1 nitrogen atom of the guanine ring system and the ring oxygen atom of the ribose ring, both of which carry a partial negative charge. Replacement of the oxygen by CH_2 in analogue **16** replaces this repulsive interaction by a weakly attractive component. In the more favorable low-energy region of the rotational potential between *chi* = +50 - +280° there is, however, hardly any difference between parent system **17** and its carbocyclic analogue **16**.



Supplementary Figure 8: Rotational profile around angle *chi* (X/C1'/N/C6) in 2'-deoxyguanosine **17** (red dots) and its carbocyclic analogue **16** (blue diamonds) at the MP2(FC)/G3MP2large//B3LYP/6-31+G(d) level of theory.

1. Ye, Y. *et al.* Mechanistic aspects of the formation of guanidinohydantoin from spiroiminodihydantoin under acidic conditions. *Chem. Res. Toxicol.* **22**, 526-35 (2009).
2. Curtiss, L.A., Redfern, P.C., Raghavachari, K., Rassolov, V. & Pople, J.A. Gaussian-3 theory using reduced Moller-Plesset order. *J. Chem. Phys.* **110**, 4703-4709 (1999).
3. M. J. Frisch *et al*. Gaussian 09, Revision B.01 edn (Gaussian, Inc., Wallingford CT, 2010).
4. www.chem.qmul.ac.uk/iupac/misc/pnuc1.html.
5. www.chem.qmul.ac.uk/iupac/misc/pnuc2.html.
6. Ober, M., Müller, H., Pieck, C., Gierlich, J. & Carell, T. Base pairing and replicative processing of the formamidopyrimidine-dG DNA lesion. *J. Am. Chem. Soc.* **127**, 18143-9 (2005).

Stathis, D.,[1] Lischke, U., Koch, S.C. Deiml, C. A., Carell, T. "Discovery and mutagenicity of a guanidinoformimine lesion as a new intermediate of the oxidative deoxyguanosine degradation pathway." ***J. Am. Chem. Soc.*** **2012, 134, 4925-4930.**

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

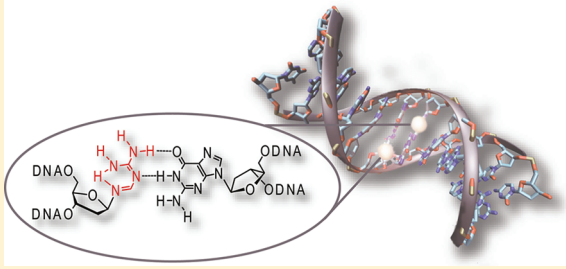
Discovery and Mutagenicity of a Guanidinoformimine Lesion as a new Intermediate of the Oxidative Deoxyguanosine Degradation Pathway

Dimitrios Stathis, Ulrike Lischke, Sandra C. Koch, Christian A. Deiml, and Thomas Carell*

Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität, Munich, Butenandtstraße 5-13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: Oxidative degradation of DNA is a major mutagenic process. Reactive oxygen species (ROS) produced in the course of oxidative phosphorylation or by exogenous factors are known to attack preferentially deoxyguanosine. The latter decomposes to give mutagenic lesions, which under physiological conditions are efficiently repaired by specialized maintenance systems in the cell. Although many intermediates of the degradation pathway are today well-known, we report in this study the discovery of a new intermediate with an interesting guanidinoformimine structure. The structure elucidation of the new lesion was possible by using HPLC−MS techniques and organic synthesis. Finally we report the mutagenic potential of the new lesion in comparison to the known lesions imidazolone and oxazolone using primer extension and pyrosequencing experiments.



INTRODUCTION

Guanine is the nucleobase with the lowest oxidation potential of the four canonical nucleobases.[1] Therefore, reactive oxygen species (ROS) degrade preferentially dG bases to give dG-derived lesions.[2−6] These lesions are mutagenic and typically repaired by dedicated DNA repair enzymes.[7−12] Scheme 1 summarizes the main degradation pathway of 2′-deoxyguanosine. Most of the oxidation products are today known.[13] The structures of certain putative intermediates along the dG degradation pathway and their mutagenic properties remain, however, to be uncovered. Particularly the reaction of the oxazolone lesion (dZ) to the aminoriboside, which hydrolyzes to give an AP site as the final product of the oxidative dG degradation, is an enigmatic transformation. One-electron oxidation of dG followed by reaction of the dG-radical cation ($dG^{\bullet+}$) with water and finally either oxidation or reduction of the 8-OH-dG radical give rise to 8-oxo-guanosine (8-oxo-dG)[14] or formamidopyrimidine (FaPy-dG) lesions,[15−17] respectively (Scheme 1). Both are well studied oxidative lesions, which strongly contribute to the mutagenic effect of ROS. A second more complex degradation pathway involves deprotonation of $dG^{\bullet+}$ and trapping of the radical by $O_2^{\bullet-}$. The corresponding 5-OOH-dG intermediate predominantly leads to the imidazolone (dIz) lesion,[5,18] known to rearrange to oxazolone (dZ).[19,20] The formation of spiroiminohydantoin (dSp), guanidinohydantoin (dGh), and iminoalantoin (dIa), reported before,[21] is assumed to take place via an intramolecular redox reaction in which the C5-peroxyl group participates in the oxidation of C8 to generate 8-hydroxy-dG as the precursor of these compounds. Oxazolone (dZ) is the last characterized intermediate before the occurrence of 1′-aminoriboside, which hydrolyzes to give an abasic site. The mutagenic effect of ROS is caused by all of these degradation intermediates, which exist in damaged DNA in unknown relative amounts. They all have been implicated in human pathogenesis and disease.[22−24] For example, 8-oxo-dG[25] and FaPy-dG lesions[26,27] are known to produce G to T transversion mutations, while imidazolone lesions are assumed to induce G to C transversion mutations.[28]

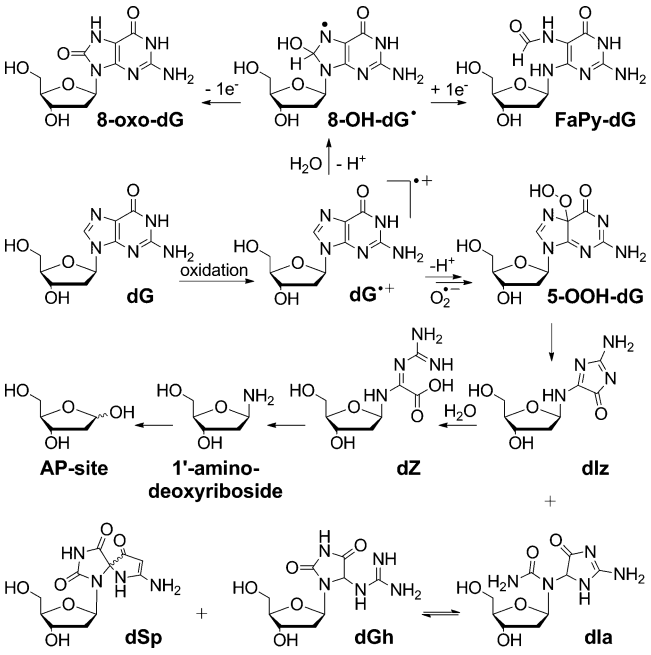
RESULTS AND DISCUSSION

In order to get more insight into the imidazolone/oxazolone based degradation pathway we performed HPLC−MS experiments with purified dIz lesion-containing DNA strands. For the study we prepared the oligonucleotide ODN1 (Figure 1A) and irradiated the DNA strand, containing a single dG base in the middle of the sequence, in the presence of riboflavin. The latter is a well-known dG photooxidant, which leads to the formation of imidazolone lesions.[20,29] Indeed, irradiation of ODN1 gave a mixture of products with one major component. This oligonucleotide (dIz-ODN1) was isolated and the presence of the imidazolone lesion was confirmed by high resolution ESI mass spectrometry. Figure 1A shows the HPL-chromatogram of the purified dIz-containing ODN1 strand, together with the high resolution mass spectrum. The molecular peak for dIz-ODN1 appears at m/z = 1367 ($[M - 3H]^{3-}$), while the second major peak at m/z = 1374 is caused by the corresponding monosodium adduct. To further confirm the presence of the dIz lesion in the purified oligonucleotide, we digested the DNA

Received: December 19, 2011
Published: February 13, 2012

   dx.doi.org/10.1021/ja211435d | J. Am. Chem. Soc. 2012, 134, 4925−4930

Scheme 1. Oxidative Degradation of 2′-Deoxyguanosine by Direct Electron Abstraction and the Corresponding Base Modifications



with P1 nuclease and subsequently incubated the reaction mixture with alkaline phosphatase and snake venom phosphodiesterase. The obtained nucleoside mixture was again analyzed by HPLC–MS. Figure 1B shows the HPLC signals of the natural nucleosides present in the digest, as well as a new signal which appears with a retention time of 9.6 min. This signal gives an m/z value of 229.0930, which is in perfect agreement with the calculated molecular weight of the dIz nucleoside. These results confirm that the first, major reaction product, formed during riboflavin induced dG oxidation, is indeed the imidazolone lesion.

While handling the dIz-containing ODN1 strand, we noticed that the imidazolone lesion is rather unstable. The lesion degraded quickly to give further reaction products. Figure 2 shows the HPL-chromatogram of the formerly pure dIz-containing DNA strand after standing of the solution for 24 h at room temperature. It is clearly visible that at least two additional compounds were formed within the DNA. To investigate the stability in more detail, we dissolved the dIz-containing DNA in different buffers at various pH values and monitored the decomposition reaction by LC–MS.

Indeed, the imidazolone lesion is unstable and rapidly degrades ($t_{1/2}$ = 18 h at pH = 7.0) to give two further products with a rate that depends on the salt concentration, the temperature, and more importantly, the pH of the buffer. It is most stable when the DNA solution is kept cold (<10 °C) at pH = 7.0. Changing the pH value (Supplementary Figure S2) or increasing the temperature of the solution, e.g., to 37 °C, caused rapid decomposition.

As previously reported,[30] analysis of the dIz decomposition reaction by HPLC–MS showed that the dIz-ODN1 initially forms the oxazolone product dZ, which is detectable at m/z = 686.27 (z = 6). If a solution of the dIz-ODN1, for example, was allowed to stand for 24 h at room temperature at pH 7.0, the solution contained finally only residual amounts of dIz-ODN1. Instead, the dZ-ODN1 dominated (>90%) and another, so far

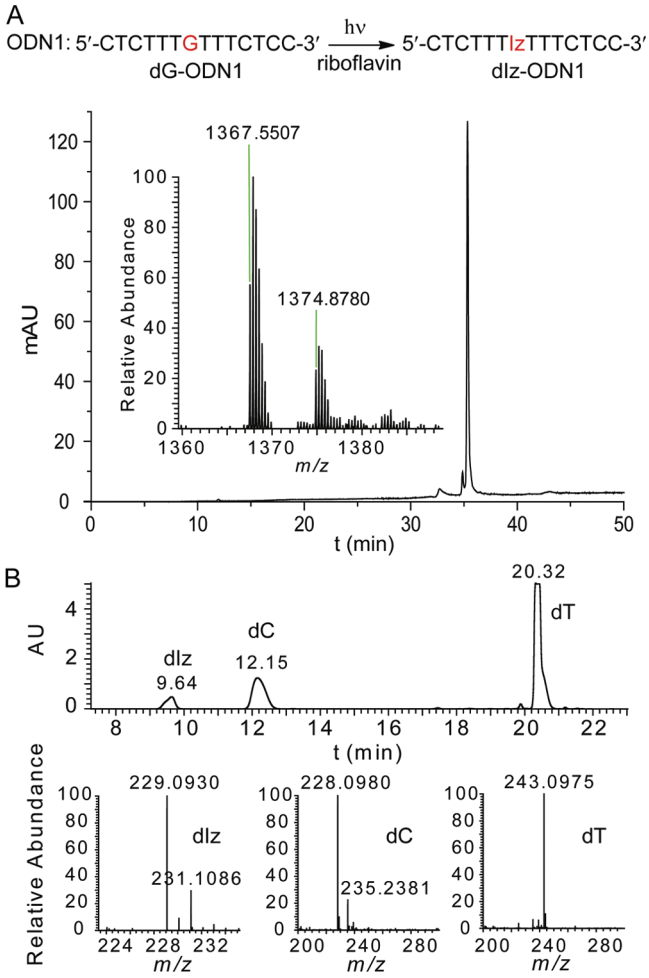


Figure 1. (A) HPLC of the purified dIz-ODN1 and the corresponding high resolution mass spectrum ($[M - 3H]^{3-}_{calcd}$ = 1367.5547, $[M - 4H + Na]^{3-}_{calcd}$ = 1374.8820). (B) HPLC–MS spectra obtained from the total enzymatic digestion of the dIz-ODN1. dIz $[M + H]^{+}_{calcd}$ = 229.0931, dC $[M + H]^{+}_{calcd}$ = 228.0979, and dT $[M + H]^{+}_{calcd}$ = 243.0975.

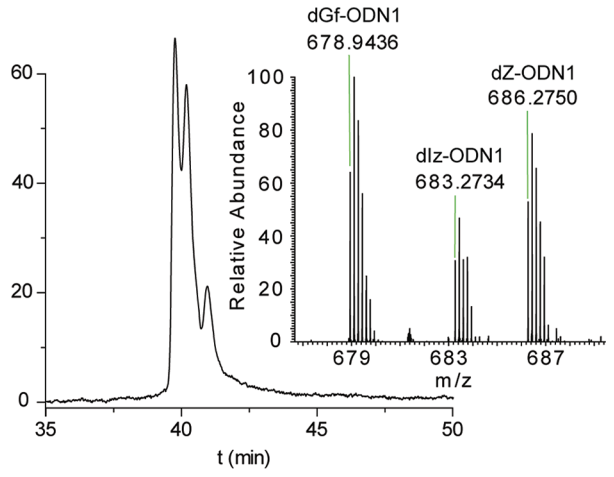


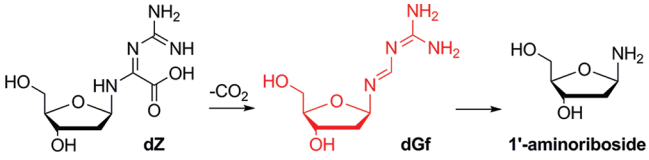
Figure 2. HPLC of dIz-ODN1 after standing for 24 h at room temperature and the corresponding high resolution mass spectrum. dIz-ODN $[M - 6H]^{6-}_{calcd}$ = 683.2737, dZ-ODN $[M - 6H]^{6-}_{calcd}$ = 686.2755, and dGf-ODN $[M - 6H]^{6-}_{calcd}$ = 678.9438.

dx.doi.org/10.1021/ja211435d | J. Am. Chem. Soc. 2012, 134, 4925–4930

unknown reaction product, appeared. Analysis of the DNA by mass spectrometry after 48 h led to the surprising discovery of a new compound that accumulates with time to significant amounts (>25%). Certainly the new compound is more stable than the original imidazolone lesion.

For this so far unknown compound (dGf) formed in ODN1, we determined a molecular weight only 26 mass units lighter than the original dIz-ODN1 (Supplementary Figure S3). Because the dGf-ODN1 accumulated only after formation of the oxazolone lesion dZ, we reasoned that the new compound, dGf, must be a further dZ-derived degradation product that builds up in solution and that could be the unknown link between the dZ lesion and the amino-riboside. Based on the molecular weight and the fact that dGf is formed likely from dZ, we speculated that the compound forms by decarboxylation and therefore proposed the guanidinoformimine structure shown in Scheme 2.

Scheme 2. Decomposition of the Oxazolone Containing Oligonucleotide via the New Lesion As an Intermediate to 1′-Aminoriboside



The high resolution mass spectrum shown in Figure 2 provides in total three sets of signals caused by dIz-, dZ-, and dGf-ODN1. The calculated value of the dGf-containing DNA strand (m/z = 678.9438, z = 6) is in excellent agreement with the experimentally determined value of m/z = 678.9436 (z = 6), strongly supporting the decarboxylation idea.

In order to gain further support for the structure, we next performed a total enzymatic digest of dIz-ODN1, which was allowed to stand in solution for 2 days, and analyzed the digest by HPLC–MS (Figure 3). The experiment proved the simultaneous presence of DNA containing dIz, dZ, and dGf lesions. Again, the molecular weight of dGf, now as the nucleoside, with m/z = 203.1140, was in excellent agreement with the proposed structure. Furthermore, we were able to observe a signal corresponding to the putative dGf-aglycon (Gf) in an MS/MS experiment. The fragment has the molecular weight of m/z = 87.0668.

In order to investigate if the lesion composition changes in double-stranded DNA, we repeated the study with a dIz-containing DNA double strand. The same decomposition products dIz, dZ, and dGf were observed at a similar ratio as in single-stranded DNA. However, the decomposition process was slightly slower in the duplex, probably due to partial shielding of the lesions from the aqueous environment.

To finally prove that the dGf product possesses the proposed guanidinoformimine structure, we next wanted to show that the key fragmentation product with m/z = 87.0668 is indeed the deribosylated fragment. To this end we synthesized the 3′,5′-bisacetyl-dG derivative (ac_2dG).[31,32] After 8 min of irradiation of the compound in the presence of riboflavin, we isolated the bisacetyl-protected ac_2dIz nucleoside in 90% yield. An aqueous solution of the ac_2dIz derivative was allowed to stand at room temperature for several days. Analysis of the solution during this time showed that within 12 h the ac_2dIz nucleoside fully converted to the corresponding ac_2dZ nucleoside. After about 3

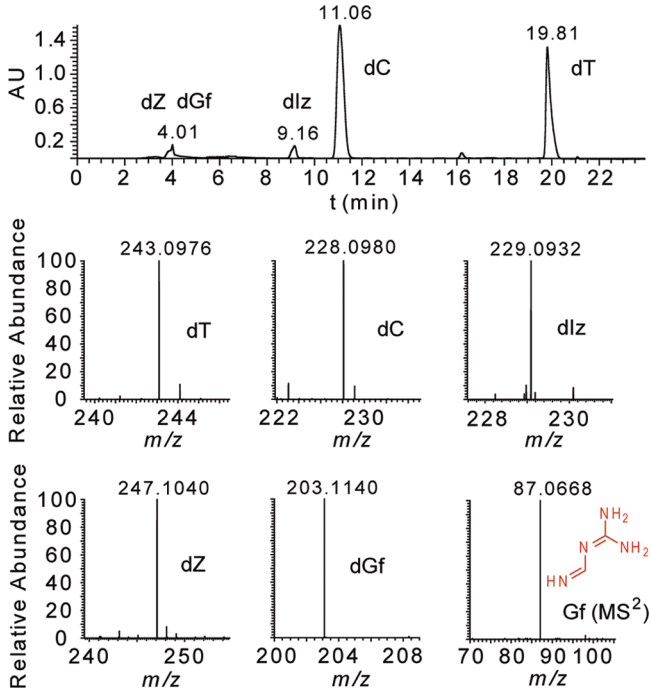


Figure 3. Enzymatic digest and HPLC–MS/MS of the dIz-ODN1 after 2 days. The experimental exact masses are in agreement with the calculated values. dT $[M + H]^+_{calcd}$ = 243.0975, dC $[M + H]^+_{calcd}$ = 228.0979, dIz $[M + H]^+_{calcd}$ = 229.0931, dZ $[M + H]^+_{calcd}$ = 247.1037, dGf $[M + H]^+_{calcd}$ = 203.1139, and Gf $[M + H]^+_{calcd}$ = 87.0665.

days, a new compound with a molecular weight (m/z = 287.1349) calculated for ac_2dGf was formed in about 20% yield. In the bisacetylated form, the reaction to the dGf compound is hence significantly slower than in DNA. The ac_2dGf compound was next used for MS/MS experiments, which allowed us to detect again the fragment at m/z = 87.0668. This proves that the new intermediate is the deribosylated species with the proposed guanidinoformimine structure (Supplementary Figure S6). In order to collect additional data which support the structure, we also performed NMR studies, although this was possible only for the obtained compound mixture. In support of the structure, we detected a signal at 7.78 ppm that corresponds to the amidinium-C*H* (Supplementary Figure S7). The signal grew with time in line with increasing amounts of dGf in solution.

From our study we conclude that the first dG degradation product is the dIz lesion, which rearranged within a few hours to give the dZ lesion as the dominating species. Subsequently decarboxylation occurs giving the new dGf intermediate. This compound is already detectable after 1 day and accumulates while the amount of the dZ lesion decreases (20 mM in water, rt). Finally 1′-amino-ribose is detected as the hydrolysis product of the dGf compound. The aminoriboside hydrolyses thereafter to form the abasic site. In this scenario, the dIz lesion is only able to contribute to the mutagenic effect of ROS in the first hours after lesion formation. Thereafter the dZ lesion will contribute dominantly followed by the newly discovered dGf lesion.

In order to determine the mutagenic effect of the lesions (dIz, dZ, and dGf), we performed primer extension experiments using one high (Klenow exo$^-$) and two low fidelity polymerases (hPolκ and *sc*Polη, respectively) and lesion-containing oligonucleotides X-ODN2 (X = dG, dIz, dZ, or

dx.doi.org/10.1021/ja211435d | *J. Am. Chem. Soc.* 2012, 134, 4925–4930

dGf). The polymerases were allowed to extend a complementary primer strand along the lesion containing template in the presence of all four canonical triphosphates. In order to estimate the bypass efficiencies exhibited by the polymerases mentioned above we performed primer extension studies and followed the reaction using polyacrylamide gel electrophoresis. The experiment revealed that all lesions can be bypassed although the bypass efficiency was found to be limited (Figure 4A). We discovered that the efficiencies vary significantly depending on the polymerase and the lesion. Polκ is almost completely stalled and is hardly able to incorporate even a single nucleotide opposite the lesions. The Klenow fragment in contrast incorporates one nucleotide relatively efficiently, after which it stalls to a large extent. Polη has the highest bypassing capabilities. For all three polymerases the dGf lesion constitutes the most severely blocking lesion. Despite the low bypass efficiency, still significant amounts of the primers were fully extended. To characterize and quantify the mutations associated with the bypass, we sequenced the fully extended strands by pyrosequencing.[33,34] The results of the primer extensions and of the pyrosequencing experiments are summarized in Figure 4.

As a control, all studies were also performed with a dG-ODN2 template strand, always resulting in quantitative incorporation of dC opposite the dG residue, independently of the polymerase used. The Klenow fragment, as a model for a high fidelity polymerase, inserts only dC opposite dG into the primer strand. The dIz lesion is preferentially base paired with dC (46%), but also dA (21%) and dG (33%) are inserted. The dZ lesion, in contrast, instructs the polymerase to insert a dA (59%) with remarkable efficiency. The dGf compound induces the incorporation of dC, dA, and dG with very similar efficiencies of about 30%. The low fidelity polymerase Polκ pairs the lesions dIz and dZ preferentially with dC (58%, 41%). However, incorporation of dG is the second best option opposite dIz and the counterbase of preference for the dGf lesion. Polη, in contrast, base pairs all lesions preferentially with dG ($\approx$ 60%) followed by dC ($\approx$ 40%). Finally, we observed that none of the polymerases performs any base pairing with dT, which is very surprising.

The results show that all investigated polymerases are able to bypass the three lesions to a certain extent. Polη is the polymerase that will generate dG to dC transversion mutations with the highest rate *in vitro*. *In vivo* studies have, however, shown that in cells small oxidative lesions are likely bypassed with a high fidelity polymerase[35] and by the low fidelity polymerase Polκ.[36] Bypass by Polη *in vivo* is consequently unlikely.[37] We observe that in the cases of dZ and dGf the high fidelity Klenow fragment incorporates the correct base dC in roughly only 30% of all bypass reactions. Polκ shows a comparable selectivity. Overall, Polκ reduces the error-rate during bypass only marginally. The unusually high incorporation of dA, by both Klenow fragment and Polκ, can be partially explained by the A-rule of polymerases.[38−41] We explain the small differences between our data and data in the literature also with the different method that we utilized here to investigate the error-rate during bypass.[42−44] In our case only fully elongated primers are sequenced to provide data. Primers, in which the polymerase inserted a base opposite the lesion which are not fully extended, are consequently not analyzed.

Quite interesting is the observation that the new dGf intermediate is bypassed by all polymerases despite its open structure. dG incorporation opposite this lesion is slightly

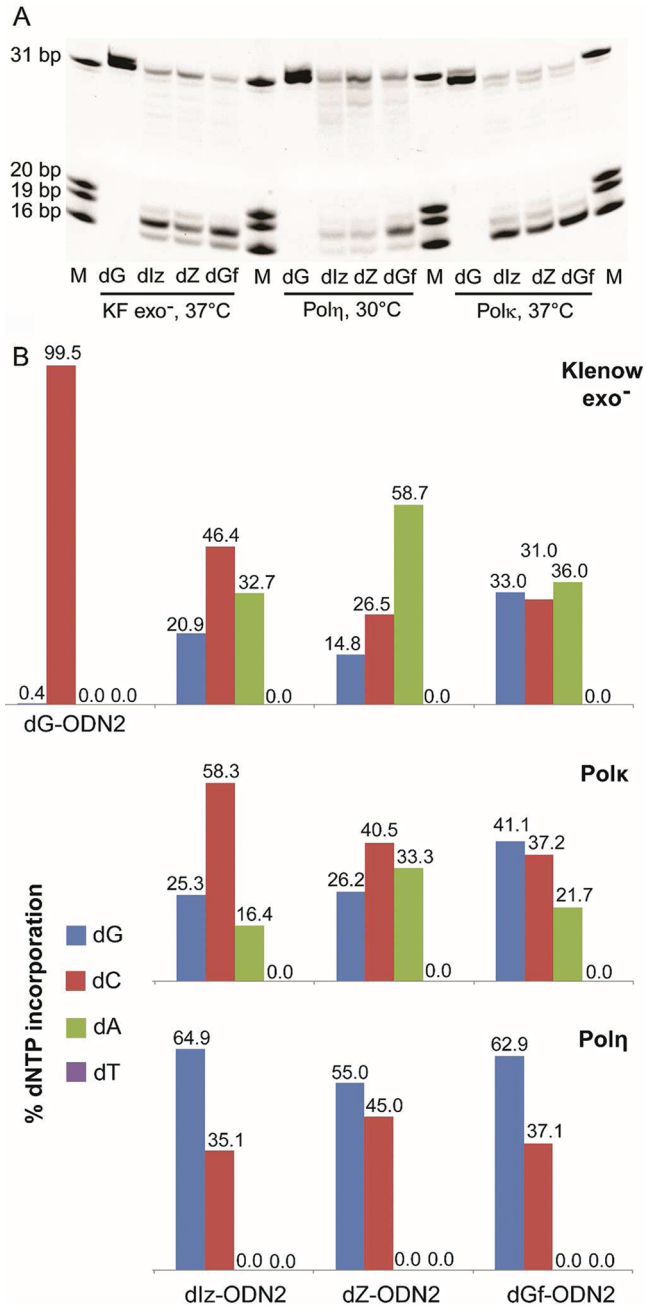


Figure 4. (A) Polyacrylamide gel electrophoresis of the primer extension experiments and (B) graphical representation of the mutagenesis assay results. X-ODN2 (where X = dG, dIz, dZ, or dGf) is annealed to a biotinylated primer, which is extended *in vitro* by the polymerase of interest. The corresponding biotinylated strands, containing the lesion's counter-base Y, are annealed to a reverse primer and analyzed by pyrosequencing. The mutagenic "fingerprint" of each dG degradation product is directly visualized. A dG-containing control strand was used for each polymerase. Here incorporation of dC was always 100%. Incorporation of dT was not observed during the analysis.

favored in the low fidelity systems, while the high fidelity Klenow fragment incorporates dG, dC, and dA with similar probability. We believe that the non-instructive character of the new lesion is explainable by its small size, its flexibility, and possibly the existence of different tautomers. The molecules themselves should exist in cyclic fully conjugated quasi "homoaromatic" structure[45−47] due to the presence of an

dx.doi.org/10.1021/ja211435d | J. Am. Chem. Soc. 2012, 134, 4925−4930

internal H-bond, known to be quite strong in similar systems. It seems that this H-bond allows the dGf lesion to adopt a quasi-heterocyclic structure, with the result that the dGf structure can function as a standard nucleobase (Figure 5), which explains

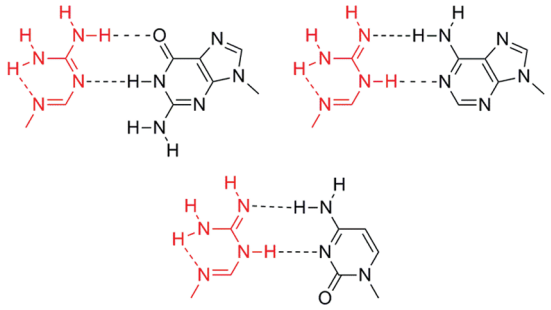


Figure 5. Possible base pairs formed between dGf and dG, dA, and dC.

the efficient bypass. In this cyclic structure, the dGf compound can, however, exist in different tautomeric forms, which could be responsible for the flexible coding properties as depicted in Figure 5.

CONCLUSION

In summary we report the discovery of a new intermediate along the oxidative dG degradation pathway. The new dGf-lesion constitutes a significant barrier to replication *in vitro*, but depending on the polymerase, substantial bypass is observed as well. The flexible base pairing properties of the compound direct polymerases to pair it in a severely error-prone manner, thus rendering it a highly mutagenic lesion. This property is explained by us with a rigid, cyclic structure of the lesion held together by an intramolecular hydrogen bond and by the existence of different tautomeric states.

ASSOCIATED CONTENT

Supporting Information

Experimental procedures, supplementary figures, and detailed mass spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Thomas.Carell@cup.uni-muenchen.de

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB749 and CA275/8-4) and the Volkswagen Foundation. We thank Toni Pfaffeneder for his help with the HPLC-MS/MS experiments and Dr. Markus Müller for critical reading of the manuscript.

REFERENCES

(1) Steenken, S.; Jovanovic, S. V. *J. Am. Chem. Soc.* **1997**, *119*, 617.
(2) Cadet, J.; Douki, T.; Gasparutto, D.; Ravanat, J.-L. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2003**, *531*, 5.
(3) Cadet, J.; Courdavault, S.; Ravanat, J.-L.; Douki, T. *Pure Appl. Chem.* **2005**, *77*, 947.
(4) Gimisis, T.; Cismas, C. *Eur. J. Org. Chem.* **2006**, 1351.
(5) Pratviel, G.; Meunier, B. *Chem.—Eur. J.* **2006**, *12*, 6018.
(6) Cadet, J.; Douki, T.; Ravanat, J.-L. *Free Radical Biol. Med.* **2010**, *49*, 9.
(7) Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L. *Nature* **2009**, *462*, 762.
(8) Coste, F.; Ober, M.; Carell, T.; Boiteux, S.; Zelwer, C.; Castaing, B. *J. Biol. Chem.* **2004**, *279*, 44074.
(9) Leipold, M. D.; Workman, H.; Muller, J. G.; Burrows, C. J.; David, S. S. *Biochemistry* **2003**, *42*, 11373.
(10) Scharer, O. D. *Angew. Chem., Int. Ed.* **2003**, *42*, 2946.
(11) Serre, L.; De, J. K. P.; Boiteux, S.; Zelwer, C.; Castaing, B. *EMBO J.* **2002**, *21*, 2854.
(12) Fromme, J. C.; Verdine, G. L. *Nat. Struct. Biol.* **2002**, *9*, 544.
(13) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109.
(14) Cadet, J.; Douki, T.; Ravanat, J.-L. *Acc. Chem. Res.* **2008**, *41*, 1075.
(15) Dizdaroglu, M.; Kirkali, G.; Jaruga, P. *Free Radical Biol. Med.* **2008**, *45*, 1610.
(16) Ober, M.; Linne, U.; Gierlich, J.; Carell, T. *Angew. Chem., Int. Ed.* **2003**, *42*, 4947.
(17) Tudek, B. *J. Biochem. Mol. Biol.* **2003**, *36*, 12.
(18) Neeley, W. L.; Essigmann, J. M. *Chem. Res. Toxicol.* **2006**, *19*, 491.
(19) Raoul, S.; Berger, M.; Buchko, G. W.; Joshi, P. C.; Morin, B.; Weinfeld, M.; Cadet, J. *J. Chem. Soc., Perkin Trans. 2* **1996**, 371.
(20) Cadet, J.; Berger, M.; Buchko, G. W.; Joshi, P. C.; Raoul, S.; Ravanat, J.-L. *J. Am. Chem. Soc.* **1994**, *116*, 7403.
(21) Luo, W.; Muller, J. G.; Burrows, C. *J. Org. Letters* **2001**, *3*, 2801.
(22) Tudek, B.; Winczura, A.; Janik, J.; Siomek, A.; Foksinski, M.; Olinski, R. *Am. J. Transl. Res.* **2010**, *2*, 254.
(23) Sedelnikova, O. A.; Redon, C. E.; Dickey, J. S.; Nakamura, A. J.; Georgakilas, A. G.; Bonner, W. M. *Mutat. Res., Rev. Mutat. Res.* **2010**, *704*, 152.
(24) Radak, Z.; Boldogh, I. *Free Radical Biol. Med.* **2010**, *49*, 587.
(25) Hsu, G. W.; Ober, M.; Carell, T.; Beese, L. S. *Nature* **2004**, *431*, 217.
(26) Delaney, M. O.; Wiederholt, C. J.; Greenberg, M. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 771.
(27) Ober, M.; Mueller, H.; Pieck, C.; Gierlich, J.; Carell, T. *J. Am. Chem. Soc.* **2005**, *127*, 18143.
(28) Kino, K.; Sugiyama, H. *Chem. Biol.* **2001**, *8*, 369.
(29) Kino, K.; Saito, I.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 7373.
(30) Gasparutto, D.; Ravanat, J. L.; Gerot, O.; Cadet, J. *J. Am. Chem. Soc.* **1998**, *120*, 10283.
(31) Matsuda, A.; Shinozaki, M.; Suzuki, M.; Watanabe, K.; Miyasaka, T. *Synthesis* **1986**, *1986*, 385.
(32) Wei, G.; Loktionova, N. A.; Pegg, A. E.; Moschel, R. C. *J. Med. Chem.* **2005**, *48*, 256.
(33) Ronaghi, M.; Uhlén, M.; Nyrén, P. *Science* **1998**, *281*, 363.
(34) Munzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chemistry* **2011**, *17*, 13782−13788.
(35) Neeley, W. L.; Delaney, J. C.; Henderson, P. T.; Essigmann, J. M. *J. Biol. Chem.* **2004**, *279*, 43568.
(36) Prakash, S.; Johnson, R. E.; Prakash, L. *Annu. Rev. Biochem.* **2005**, *74*, 317.
(37) Hubscher, U.; Maga, G.; Spadari, S. *Annu. Rev. Biochem.* **2002**, *71*, 133.
(38) Bernard S, S. *DNA Repair* **2002**, *1*, 125.
(39) Sagher, D.; Strauss, B. *Biochemistry* **1983**, *22*, 4518.
(40) Kunkel, T. A.; Schaaper, R. M.; Loeb, L. A. *Biochemistry* **1983**, *22*, 2378.
(41) Boiteux, S.; Laval, J. *Biochemistry* **1982**, *21*, 6746.
(42) Duarte, V.; Gasparutto, D.; Jaquinod, M.; Cadet, J. *Nucleic Acids Res.* **2000**, *28*, 1555.
(43) Kino, K.; Sugiyama, H. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2005**, *571*, 33.
(44) Kino, K.; Sugasawa, K.; Mizuno, T.; Bando, T.; Sugiyama, H.; Akita, M.; Miyazawa, H.; Hanaoka, F. *ChemBioChem* **2009**, *10*, 2613.

(45) Winstein, S. *J. Am. Chem. Soc.* **1959**, *81*, 6524.
(46) Williams, R. V. *Chem. Rev.* **2001**, *101*, 1185.
(47) Freeman, P. K. *J. Org. Chem.* **2005**, *70*, 1998.

Supporting Information

Discovery and Mutagenicity of a Guanidinoformimine Lesion as a new Intermediate of the Oxidative Deoxyguanosine Degradation Pathway

Dimitrios Stathis, Ulrike Lischke, Sandra C. Koch, Christian A. Deiml, and Thomas Carell*

Department of Chemistry, Ludwig-Maximilians-Universität, Munich, Butenandtstraße 5-13, 81377, Munich, Germany

*Thomas.Carell@cup.uni-muenchen.de

Materials and Methods: All solvents and reagents were obtained in commercially available qualities puriss., p. a., or purum. Dry solvents were purchased from *Fluka* or *Acros* and were used as received. Acetonitrile for HPLC and MS was purchased from *VWR* in HPLC grade. Unmodified oligonucleotides were purchased from *Metabion international AG*. The exonuclease deficient DNA polymerase I from *E. coli* (Klenow exo$^-$ fragment) was purchased by *New England BioLabs GmbH*. ^{1}H NMR spectra were obtained on *Varian* Mercury-200 MHz, *Varian* INOVA-400 MHz or *Varian* 600 MHz spectrometers. ^{13}C NMR spectra were obtained on a *Varian* INOVA-400 MHz or *Varian* 600 MHz spectrometer. ESI spectra were obtained on a *Thermo* Finnigan LTQ-FT ICR spectrometer. Analytical HPLC analysis was performed on a *Waters* 2695 apparatus equipped with thermostatic sample chamber, *Waters* 2996 Photo Diode Array detector and a *Macherey - Nagel* cc250/4 Nucleosil 120-3 C18 column. Preparative HPLC separations were achieved on a *Waters* 1525 system equipped with a *Waters* 2487 UV detector and a vp 250/10 Nucleodur 100-5 C18ec column from *Macherey - Nagel*. The elution system used both for analytical and preparative HPLC consisted of buffer A (100 mM $NHEt_3OAc$ in water, pH=7.0) and buffer B (100 mM $NHEt_3OAc$ in 80% acetonitrile, pH=7.0) using gradients and flow rates depending on the experiment. The

digestion experiments were analyzed by LC-ESI-MS on a *Thermo Finnigan* LTQ Orbitrap XL coupled with a *Dionex* Ultimate 3000 HPLC system, having a flow of 0.15 mL/min over an Uptisphere120-3HDO column from *Interchim*. Elution buffers were buffer C (2 mM NH_4CHOO in H_2O (pH 5.5)) and buffer D (2 mM NH_4CHOO in 80% MeCN (pH 5.5)). The gradient applied was 0 → 12 min: 0 % → 3 % buffer D; 12 → 60 min: 3 % → 60 % buffer D; 60 → 62 min: 60 % → 100 % buffer D; 62 → 70 min: 100 % buffer D; 70 → 85 min: 100 → 0 % buffer D; 85 → 95 min: 0 % buffer D. The elution was monitored at 260 nm (*Dionex* Ultimate 3000 Diode Array Detector).

Photooxidation of dG-ODN1: 200 nmol lyophilised, HPLC purified dG-ODN1 (5′-CTCTTTGTTTCTCC-3′) was dissolved in sodium cacodylate buffer (0.2 mL, 1 M) and a saturated solution of riboflavin in water (1.8 mL). The mixture was placed in a 10 mm quartz cuvette, subjected to oxygen bubbling for 15 min and irradiated under bubbling for 5 min at 5°C, at a distance of 7 cm from the source. For the irradiation, a 500 W high pressure Mercury UV lamp equipped with a 350 nm cut-off filter was used. The crude photolysis mixture was subjected to centrifugal filtration at 4°C using a *Millipore* Amicon 3000 MWCO centrifugal filter, in order to remove excess of riboflavin. HPLC purification was accomplished using an elution system of 5-20% buffer B in 60 min. The pure dIz containing oligonucleotide (5′-CTCTTT(dIz)TTTCTCC-3′, dIz-ODN1) was kept cold in order to avoid further decomposition of the lesion.

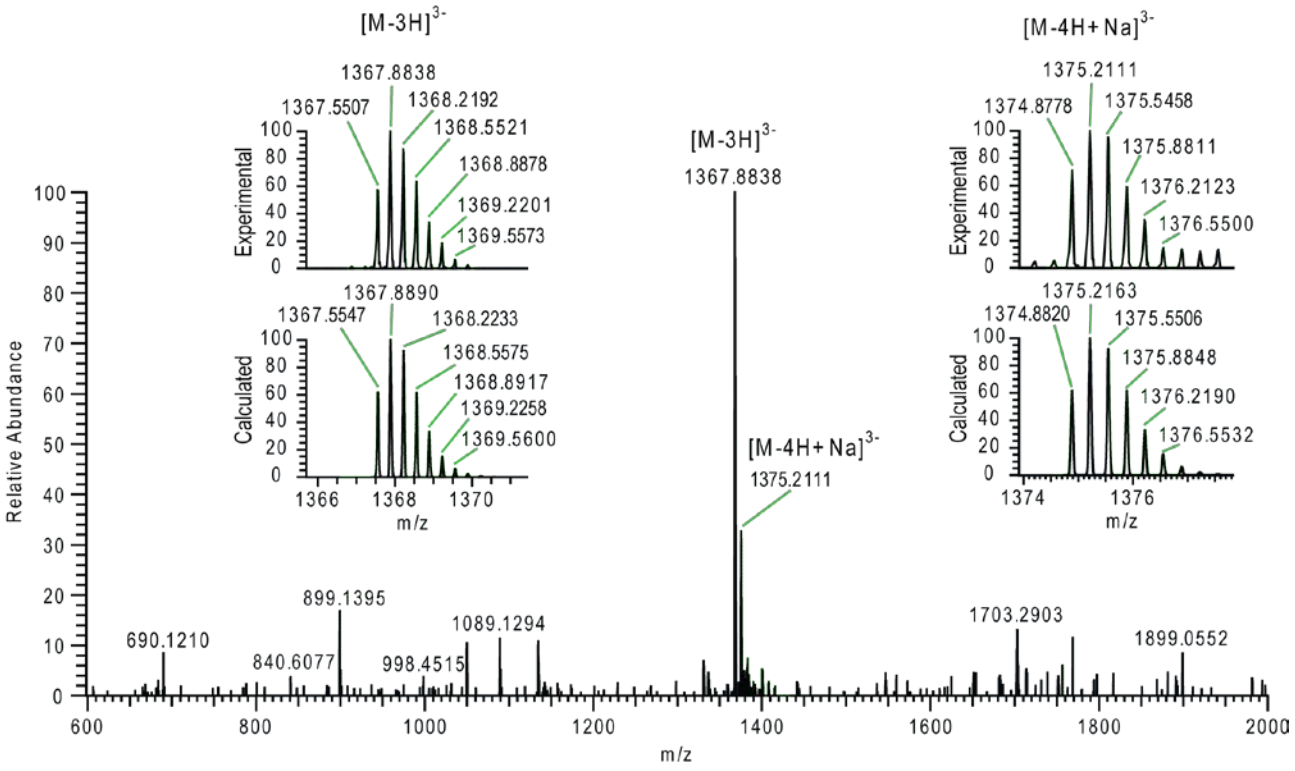


Figure S1: High resolution mass spectrum of dIz-ODN1 (5′-CTCTTT(dIz)TTTCTCC-3′).

pH dependency of dIz-ODN1 decomposition: 5 nmol of lyophilized, HPLC purified dIz-ODN1 were dissolved in 200 µL of triethylammonium acetate buffer of pH=5.4 (Fig. S2 A), pH=7.0 (Fig. S2 B) or pH=8.4 (Fig. S2 C). The decomposition of the dIz lesion (T=25°C) was followed by analytical HPLC using a gradient of 0-25 % buffer B in 45 min. The half life times of dIz at the oligonucleotide level, were found to be 3, 18 and 15 h at pH values of 5.4, 7.0 and 8.5 respectively.

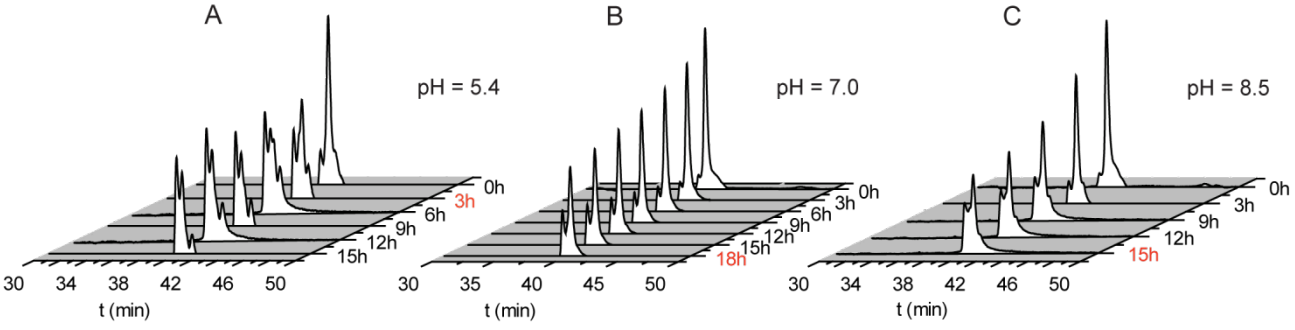


Figure S2: Stability of dIz-ODN1 (5′-CTCTTT(dIz)TTTCTCC-3′) at different pH conditions.

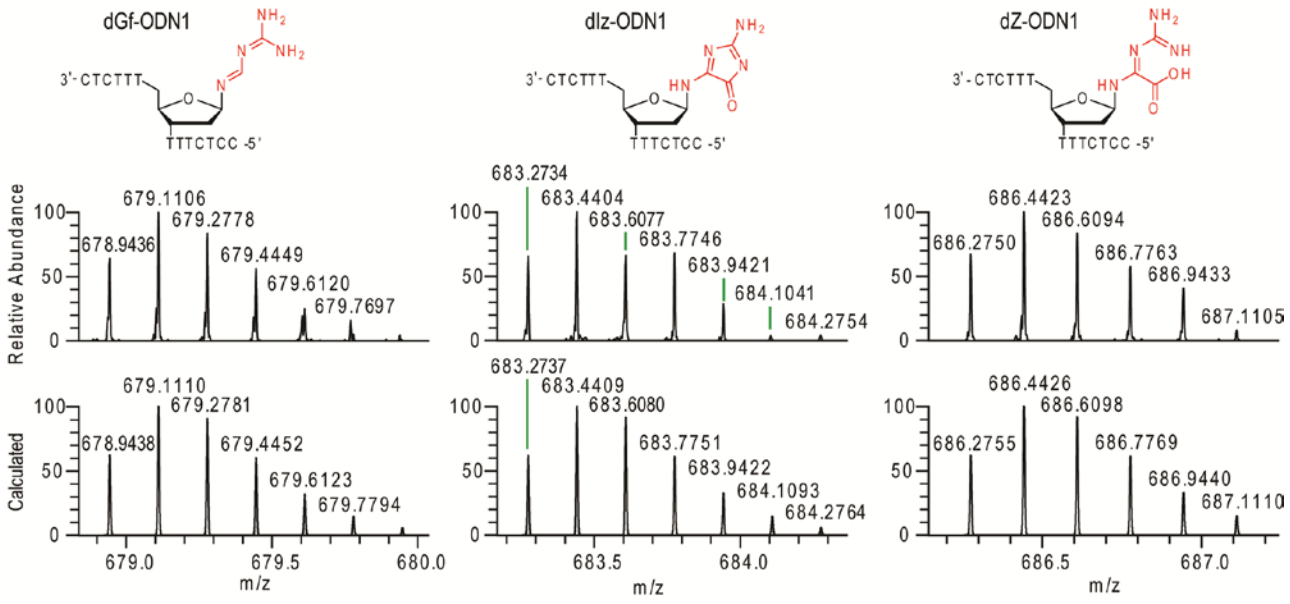


Figure S3: High resolution mass spectrum of the dIz-ODN1 and its observed decomposition products dZ-ODN1 and dGf-ODN1.

Synthesis of the 3′,5′-bisacetyl-dG (ac_2dG)[1,2]:

Deoxyguanosine (3.00 g, 11.22 mmol, 1 eq) and 4-dimethylaminopyridine (102.8 mg, 0.84 mmol, 0.075 eq) were dissolved in a mixture of acetonitrile (130 mL) and triethylamine (3.75 mL, 26.9 mmol, 2.4 eq). Acetic anhydride was added (2.54 mL, 26.9 mmol, 2.4 eq - 1.2 molar equivalents for each hydroxyl group) and the solution was stirred for 3 h at room temperature. 5 mL methanol were added and stirring continued for 5 min. The suspended material was directly collected by filtration and washed with cold ethanol and ether. After drying under high vacuum, ac_2dG was obtained pure (Figure S4) as a white powder (3.8 g, 96%). R_f = 0.25 (CH_2Cl_2/EtOH

90:10); m.p.: 224°C; ESI-MS: $[M-H]^+_{calc.}$= 350.1106, $[M-H]^+_{exp.}$ = 350.1106; ^{1}H-NMR and ^{13}C-NMR were in absolute accordance with those previously reported.

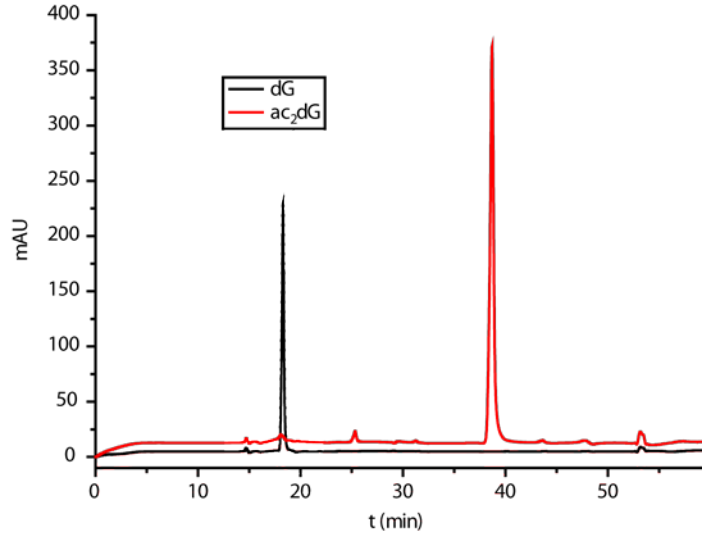


Figure S4: HPLCs of the starting material (dG), eluting at 18.5 min, and the pure product (ac_2dG), eluting at 38 min. Elution system: 0 to 30 % buffer B in 45 min.

Photooxidation of the 3′,5′-bisacetyl-dG (ac_2dG): 2 mL saturated solution of riboflavin in water was saturated with 3′,5′-bisacetyl-dG (ac_2dG) and subjected to O_2 bubbling for 15 min. The mixture was irradiated as described above for the dG-ODN1. The crude photolysis mixture was purified by HPLC using a gradient from 0 to 30 % buffer B in 45 min.

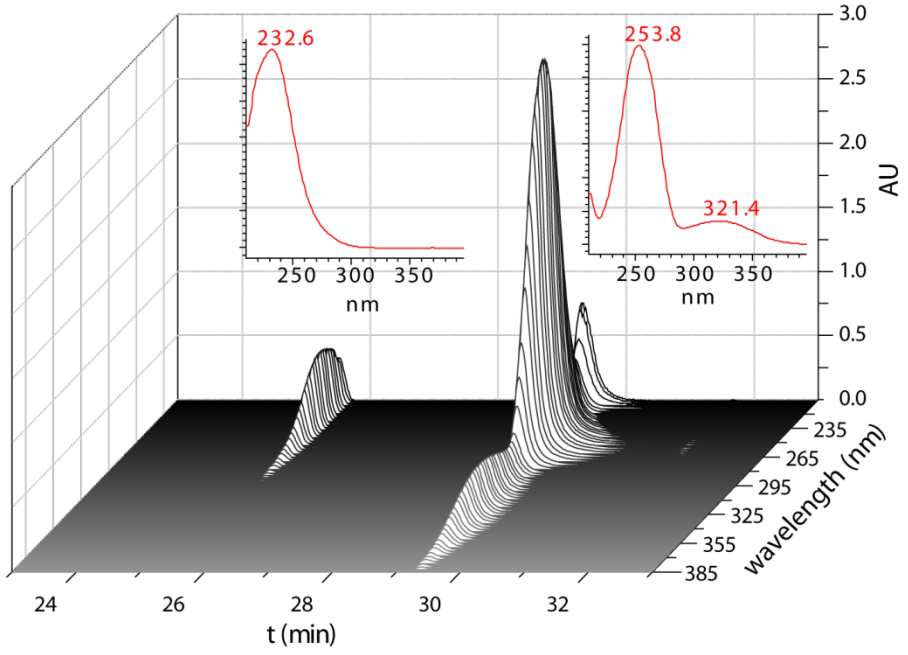


Figure S5: 3D-HPLC and corresponding UV spectra of ac_2dIz (R_t = 29.4 min), starting to decompose to ac_2dZ (R_t = 26.0 min). Elution system 0 to 30 % buffer B in 45 min.

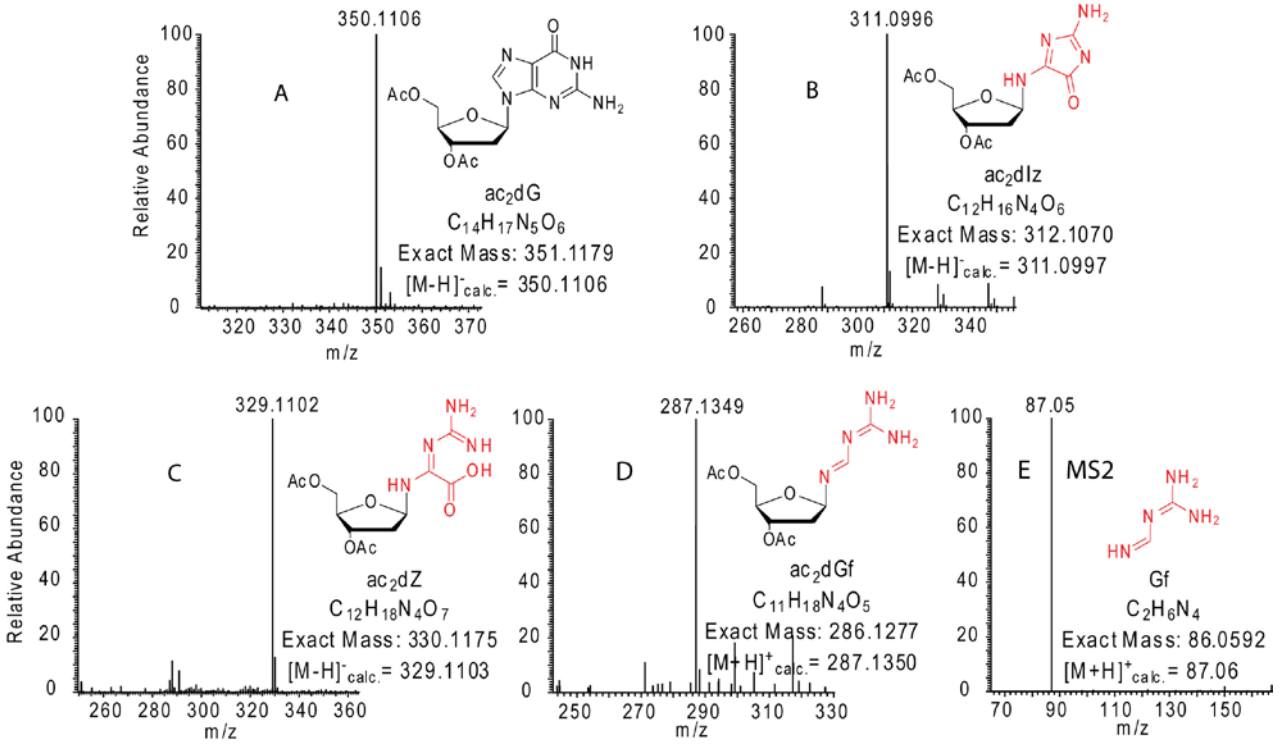


Figure S6: High resolution mass spectra of ac_2dG (A), ac_2dIz (B), ac_2dZ (C), ac_2dGf (D) and the low resolution MS/MS spectrum of the Gf-aglycone fragment (E), as measured by an ion trap experiment.

^{1}H-NMR data:

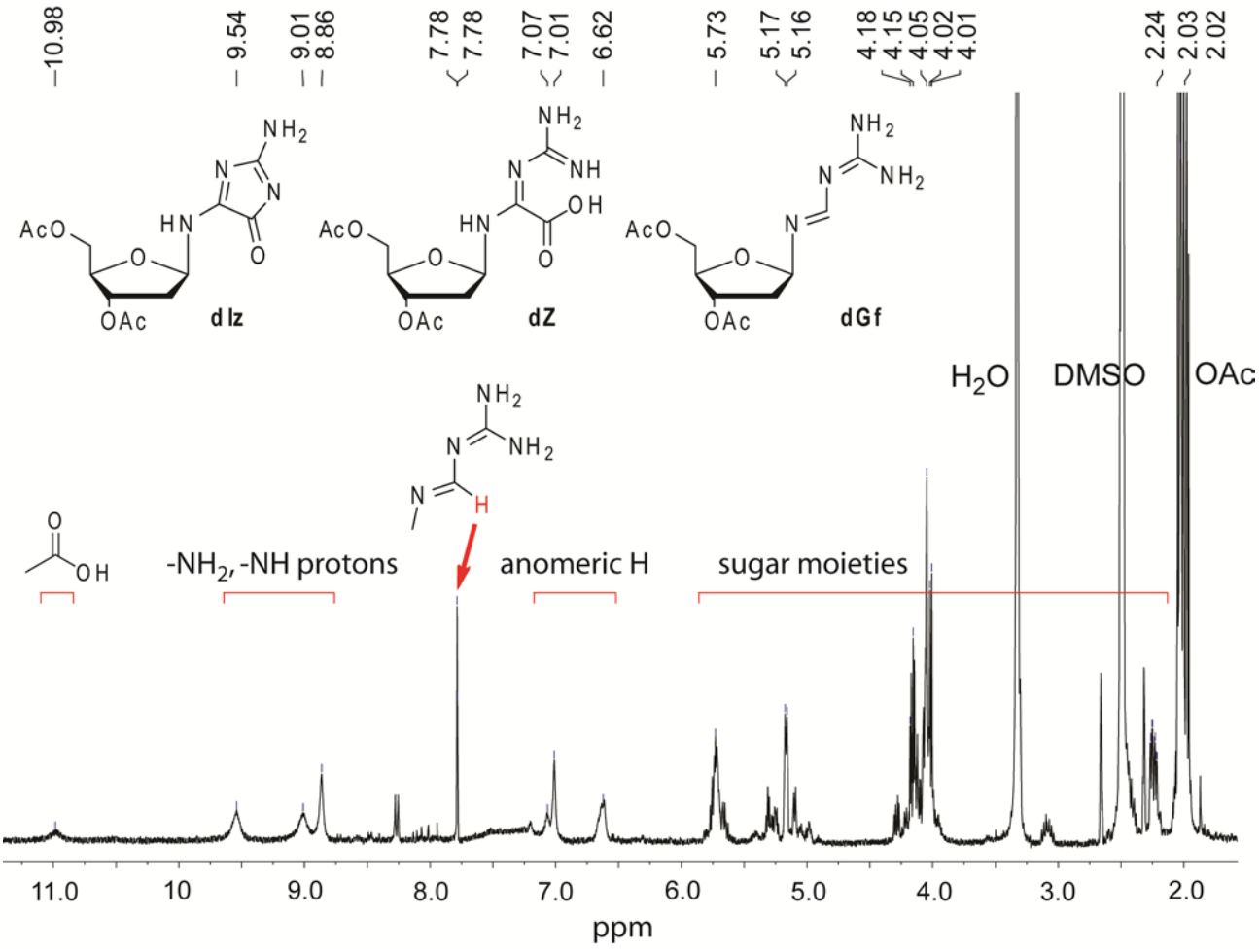


Figure S7: ^{1}H-NMR of a mixture of ac_2dIz, ac_2dZ and ac_2dGf compounds as obtained after incubation of an HPLC purified ac_2dIz water solution at 37°C for 24 h. The characteristic signal for the formimine proton appears at 7.78 ppm.

Enzymatic digestion of dIz-ODN1 (5′-CTCTTT(dIz)TTTCTCC-3′): 11 units P1 nuclease were added to a solution containing 1.5 nmol dIz-ODN1 (60 µL, 24.4 µM solution) and the mixture was incubated at 0°C for 6 h. 10 units alkaline phosphatase and 0.2 units snake venom

phosphodiesterase were added and incubation at 0°C was continued for another 12 h. The mixture was centrifuged (13400 rpm) for 15 min and the supernatant injected to the HPLC-MS.

Enzymatic digestion of dZ-ODN1 (5′-CTCTTT(dZ)TTTCTCC-3′): 1.5 nmol dIz-ODN1 (60 µL, 24.4 µM solution) were left to stand at room temperature over night in order to form the corresponding hydrolysis product 5′-CTCTTT(dZ)TTTCTCC-3′ (dZ-ODN1). 11 units P1 nuclease were added and the mixture was incubated at 37°C for 3 h. 10 units alkaline phosphatase and 0.2 units snake venom phosphodiesterase were added and incubation at 37°C was continued for another 3 h. The mixture was centrifuged (13400 rpm) for 15 min and the supernatant injected to the HPLC-MS.

Purification of human polymerase κ (hPolκ$_{19-526}$): Human polymerase kappa was cloned from human cDNA. The Stargate® system was used as a cloning strategy and the first codons were optimized for *E.coli* usage by using the primers below:

5′ P-AAT GGG CCT GAA CGA TAA TAA AGC AGG AAT GGA AGG AT-3′

5′ P-TCC CTT GTT GGT GTT TCC TGT CCT CTT C-3′

The PCR-product was applied in a Stargate® reaction to form an entry clone (pPSG-IBA10) containing a short version of human polymerase kappa (hPolκ$_{19-526}$). The pPSG-IBA33 was used as expression plasmid, which was transformed in *E. coli* Rosetta 2 (DE3) cells. hPolκ$_{19-526}$ was expressed using an autoinduction medium (ZYP-5052) adopted from Studier.[3] The cells were grown at 37°C for 4 h and then for 16 h at 25°C and harvested by centrifugation. Subsequently, the pellet was lysed in His-Buffer A (50 mM Tris pH=7.4, 500 mM NaCl, 20 mM Imidazol, 0.2 % Tween 20) using a french press and applied to a Ni-NTA superflow column. Polκ$_s$ was eluted in buffer B (50 mM Tris pH=7.4, 500 mM NaCl, 250 mM Imidazol). The protein was then transferred

to Heparin buffer A (50 mM Tris 7.4, 100 mM NaCl 5 mM DTT, 5 % glycerol) and injected into a heparin column. A high salt buffer (Heparin buffer B: 50 mM Tris 7.4, 800 mM NaCl, 5mM DTT, 5 % glycerol) was used for elution. The central peak fractions were collected and concentrated using a *Millipore* Amicon 30K 15 ml filter. As a polishing step the protein was applied to Superdex G75 with gel filtration buffer (25 mM HEPES pH=7.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP) as mobile phase. The protein was again concentrated and stored at -20°C by supplementing the buffer with 50% glycerol.

Purification of yeast polymerase η (*sc*Polη): *sc*Polη was purified as described before.[4-6]

Primer extension and denaturing polyacrylamide gel electrophoresis: 5 pmol of X-ODN2 (5′-CTA CCT ACC TCC ATX ACA CAT CCC ACA CTC C-3′) containing either dG, dIz, dZ or dGf at position X, were hybridized (2 min 50°C, then cooling -1°C/30sec) with 5 pmol of a fluorescein labeled primer (5′-Fluo-GGA GTG TGG GAT GTG T-3′). Subsequently, primer extension experiments were performed with the exonuclease deficient DNA polymerase I from *E. coli* (Klenow exo$^-$), hPolκ$_{19\text{-}526}$ and *sc*Polη. The polymerases (1 U Klenow exo$^-$, 1 µM *sc*Polη, or 0.5 µM hPolκ$_{19\text{-}526}$), 50 µM dNTPs and 0.5 µM dsDNA were incubated in a total volume of 10 µL 1×NEBuffer 2 (*New England Biolabs*; 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9) for 30 min, at optimized temperatures (30°C for *sc*Polη, 37°C for Klenow exo$^-$ and hPolκ$_{19\text{-}526}$). The reactions were terminated by adding 10 µL TBE-Urea sample buffer and analyzed by denaturating polyacrylamide gel electrophoresis.

Pyrosequencing analysis:[7] 10 pmol of X-ODN2 (5′-CTA CCT ACC TCC ATX ACA CAT CCC ACA CTC C-3′) containing either dG, dIz, dZ or dGf at position X, were hybridized with a biotin labeled primer (5′-Biotin-GGAGTGTGGGATGTGT-3′). Subsequently, primer extension

experiments were performed with the Klenow exo$^-$, hPolκ$_{19\text{-}526}$ and *sc*Polη. The polymerases (1 U Klenow exo$^-$, 1 µM *sc*Polη, or 0.5 µM hPolκ$_{19\text{-}526}$), 50 µM dNTPs and 0.5 µM dsDNA were incubated in a total volume of 20 µL 1×NEBuffer 2 (*New England Biolabs*; 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9) for 30 min, at optimized temperatures (30°C for *sc*Polη, 37°C for Klenow exo$^-$ and hPolκ$_{19\text{-}526}$). 2 µL Streptavidin Sepharose beads (*GE Healthcare*, Uppsala, Sweden), 40 µL Binding Buffer (*Qiagen*, Hilden, Germany) and 18 µL ddH_2O were added. After agitation at 1400 rpm for 15 min the beads were captured with a Vacuum Prep Tool (*Qiagen*, Hilden, Germany), washed with 70% EtOH, 0.1 M NaOH and Washing Buffer (*Qiagen*, Hilden, Germany). The beads were dissolved in 25 µL Annealing Buffer (*Qiagen*, Hilden, Germany) containing 5 pmol sequencing primer (5′-CTACCTACCTCCAT-3′), *Metabion*, Martinsried, Germany). Pyrosequencing was performed on a PyroMark Q24 Pyrosequencer using standard conditions (*Qiagen*, Hilden, Germany). The data was analyzed by the software provided by the manufacturer. Peak heights were exported to *Microsoft* Excel and the average value for each signal was calculated. The average blank values obtained for each of the four dNTPs were subtracted from the values obtained for the incorporation of the individual triphosphates. The corresponding values were used for the calculation of the relative incorporation at every variable position. The obtained data set was the average of three measurements.

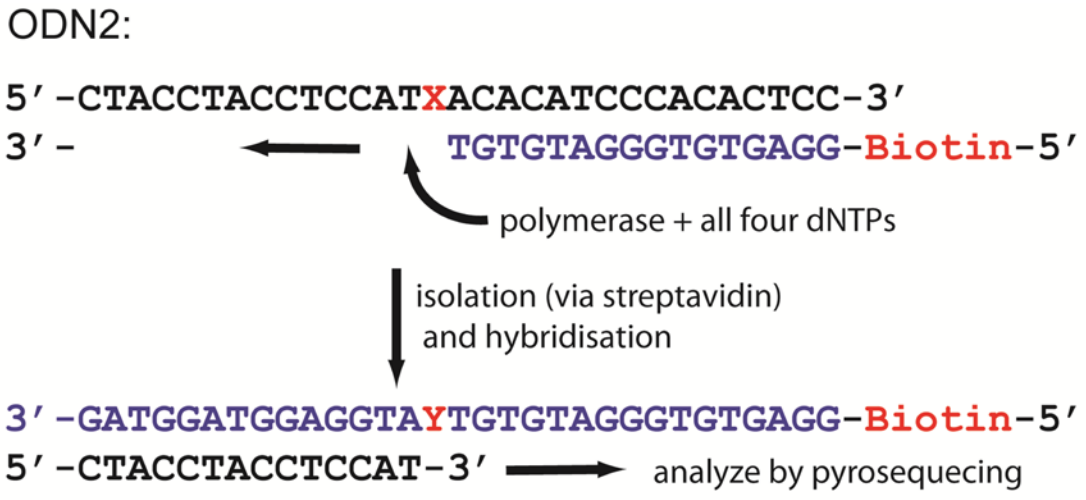


Figure S8: Graphical representation of the pyrosequencing assay.

(1) Wei, G.; Loktionova, N. A.; Pegg, A. E.; Moschel, R. C. *J. Med. Chem.* **2005**, *48*, 256.

(2) Matsuda, A.; Shinozaki, M.; Suzuki, M.; Watanabe, K.; Miyasaka, T. *Synthesis* **1986**, *1986*, 385.

(3) F. William, S. *Protein Expression Purif.* **2005**, *41*, 207.

(4) Schorr, S.; Schneider, S.; Lammens, K.; Hopfner, K.-P.; Carell, T. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20720.

(5) Ober, M.; Mueller, H.; Pieck, C.; Gierlich, J.; Carell, T. *J. Am. Chem. Soc.* **2005**, *127*, 18143.

(6) Alt, A.; Lammens, K.; Chiocchini, C.; Lammens, A.; Pieck, J. C.; Kuch, D.; Hopfner, K.-P.; Carell, T. *Science* **2007**, *318*, 967.

(7) Munzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chemistry* **2011**, *17*, 13782.

Münzel, M.,[1] Lischke, U., Stathis, D. Pfaffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K., Carell, T. "Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine." *Chemistry* 2011, 17, 13782-13788.

CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.201102782

Improved Synthesis and Mutagenicity of Oligonucleotides Containing 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine

Martin Münzel, Ulrike Lischke, Dimitrios Stathis, Toni Pfaffeneder, Felix A. Gnerlich, Christian A. Deiml, Sandra C. Koch, Konstantin Karaghiosoff, and Thomas Carell*[a]

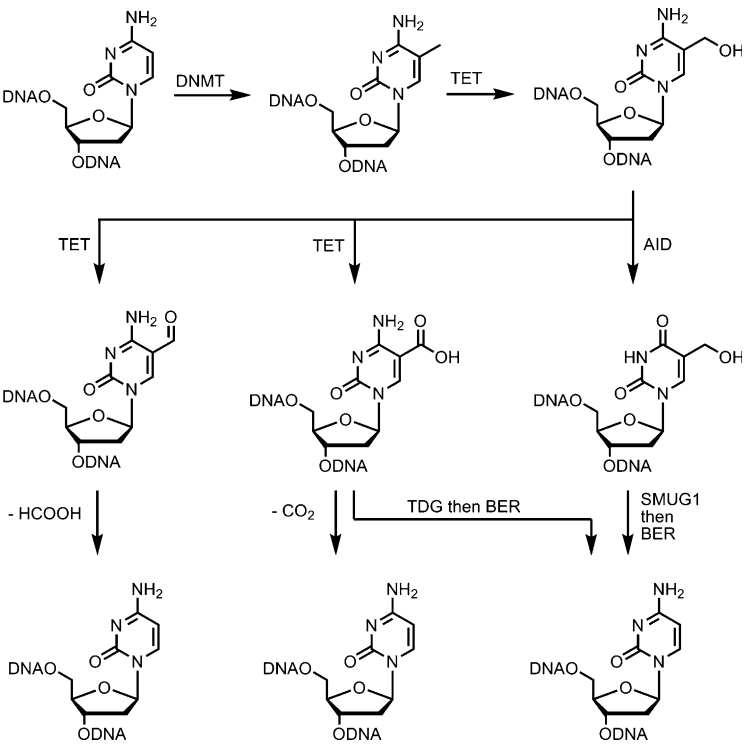
Abstract: 5-Formylcytosine (fC or $^{5\text{-CHO}}$dC) and 5-carboxylcytosine (caC or $^{5\text{-COOH}}$dC) have recently been identified as constituents of mammalian DNA. The nucleosides are formed from 5-methylcytosine (mC or $^{5\text{-Me}}$dC) via 5-hydroxymethylcytosine (hmC or $^{5\text{-HOMe}}$dC) and are possible intermediates of an active DNA demethylation process. Here we show efficient syntheses of phosphoramidites which enable the synthesis of DNA strands containing these cytosine modifications based on Pd^0-catalyzed functionalization of 5-iododeoxycytidine. The first crystal structure of fC reveals the existence of an intramolecular H-bond between the exocyclic amine and the formyl group, which controls the conformation of the formyl substituent. Using a newly designed in vitro mutagenicity assay we show that fC and caC are only marginally mutagenic, which is a prerequisite for the bases to function as epigenetic control units.

Keywords: cytosine · DNA · DNA methylation · nucleosides · phosphoramidite

Introduction

Hydroxymethylcytosine (hmC) is a new DNA base that was detected in various tissues and in embryonic stem cells.[1] It is currently believed that this so-called sixth base of the genome is involved in epigenetic programming processes.[2] Hydroxymethylcytosine is generated from the epigenetic marker base 5-methylcytosine (mC) by hydroxylation of the C5-methyl group performed by ketoglutarate-dependent TET enzymes (TET1-3).[3] The presence of mC in promoter regions is considered to be a main signal that finally leads to the silencing of the corresponding gene.[4] Others and us speculated that the new base hmC is an intermediate in active demethylation pathways (Scheme 1) that allow cells to actively remove the C5-methyl groups of mC in the absence of replication.[1e,3,5] This removal would in principle enable cells to unblock silenced genes and therewith to dynamically modulate their genetic program.

Unblocking of genes via hmC can in principle occur in multiple ways (Scheme 1). First, hmC could be an early intermediate of an oxidative demethylation pathway in which hmC is further oxidized to 5-formylcytosine (fC) and 5-carboxylcytosine (caC). Demethylation is thereafter possible via either elimination of formic acid from fC or decarboxylation of caC. Formic acid elimination is a known process that occurs during steroid biosynthesis,[6a] while decarboxylation at pyrimidine bases occurs during uridine biosynthesis



Scheme 1. Postreplicative modification of cytosine to hmC and putative active demethylation pathways. DNMT: DNA methyltransferase, TET: ten-eleven-translocation (methylcytosine dioxygenase), AID: activity induced cytosine deaminase, TDG: thymine DNA glycosylase, SMUG1: single-strand selective monofunctional uracil DNA glycosylase, BER: base excision repair.

[a] Dipl.-Chem. M. Münzel, Dipl.-Biol. U. Lischke, M. Sc. D. Stathis, M. Sc. T. Pfaffeneder, Dipl.-Chem. F. A. Gnerlich, Dipl.-Biochem. C. A. Deiml, M. Sc. S. C. Koch, Prof. Dr. K. Karaghiosoff, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPS^M)
at the Department of Chemistry, Ludwig-Maximilians-University
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.uni-muenchen.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201102782.

and in the pyrimidine salvage pathways of certain eukaryotes.[6a–d] Alternatively, it is possible that hmC or the further oxidized bases fC and caC are substrates of special DNA-repair glycosylases. These would generate abasic sites that are repaired by gap-filling with 2′-deoxycytidine (dC) using well established DNA repair pathways.[8a,9b]

Recently, others and us indeed discovered the oxidized nucleobases fC and caC in embryonic stem cells and in selected other tissues.[7–8] We could show that the fC levels decrease during stem cell development suggesting that fC is indeed involved in stem cell differentiation.[7] For fC, caC and hmC it was shown that these bases are substrates for repair enzymes. fC and caC are processed by the thymine-DNA glycosylase (TDG),[8a,9] while hmC was found to be converted to hydroxymethyluridine (hmU) by the enzyme cytosine deaminase (AID). hmU is removed from DNA by the glycosylase SMUG1 (single-strand selective monofunctional uracil DNA glycosylase).[10]

Currently the function of the new bases fC and caC is unknown and their role in a putative epigenetic active demethylation process is still unclear. Regarding their potential role as key epigenetic control intermediates during cell differentiation, the previously observed mutagenic potential of fC is problematic (Figure 1).[11] It was speculated that a putative strong intramolecular H-bond between the exocyclic N4 amino group and the carbonyl oxygen at C5 shifts the amino–imino equilibrium towards the mutagenic imino form, which would enable fC to form a wrong base pair with dA (Figure 1).[11–12] For caC mutagenicity data are so far not reported. The existence of a strong H-bond in caC is supported by the unusual low pK_a value of the C5-carboxylic acid of $pK_a < 1.5$.[13] In line with the electron-deficient character of the new bases fC and caC, the acidity of the corresponding protonated bases $[HfC]^+$ and $[HcaC]^+$ (protonation occurs at N3) is unusually high with $pK_{a[HfC]^+} = 2.4$ and $pK_{a[HcaC]^+} = 4.0$ (for comparison: $pK_{a[HC]^+} = 4.4$, $pK_{a[HhmC]^+} = 4.4$, and $pK_{a[HmC]^+} = 4.5$).[11b,13]

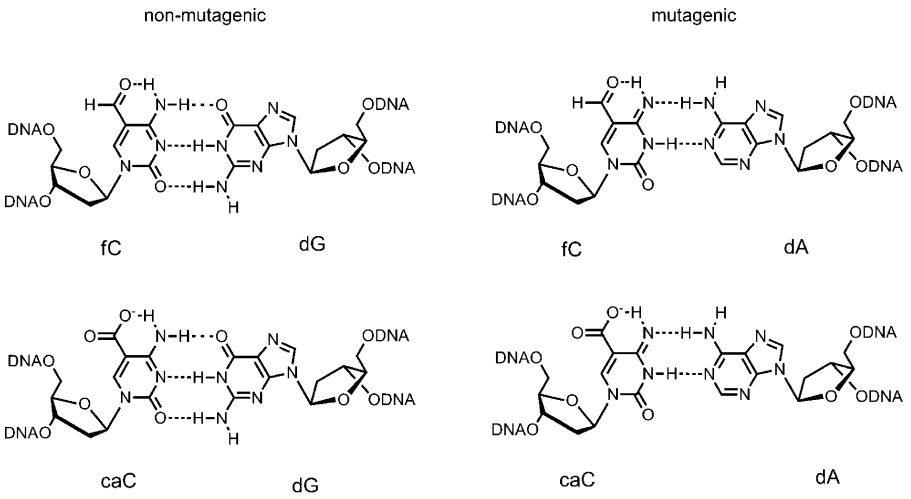


Figure 1. fC and caC exhibit an intramolecular hydrogen bond that was proposed to shift the amino/imino equilibrium towards the imino tautomeric form which would then form a mutagenic base pair with dA.

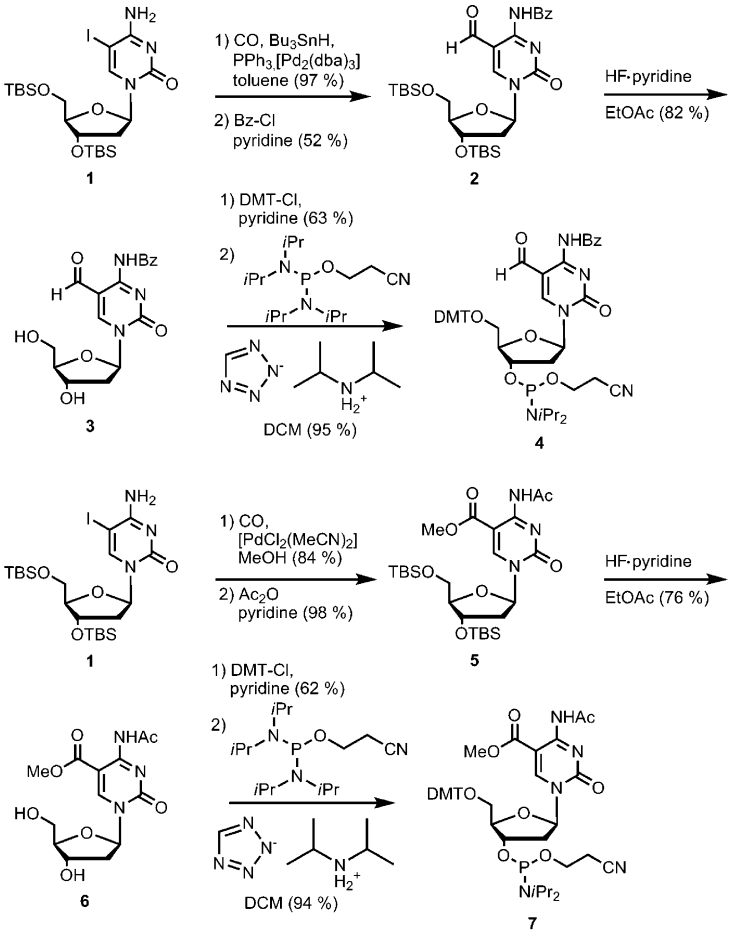
Results and Discussion

In order to investigate the mutagenic properties of caC and to re-examine the mutagenic potential of hmC and fC we prepared hmC, fC and caC oligonucleotides (Scheme 1).[14] Strands containing hmC were prepared as described by us before.[15] To access oligonucleotides containing fC in the required quantity and quality as needed for the current study we developed an improved fC phosphoramidite reagent using a carbonylative coupling strategy[1e,15a] with 5-iodo-2′-deoxycytidine (**1**) that was recently also employed by the He group to synthesize similar fC and caC phosphoramidites.[16] The improved fC phosphoramidite features a benzoyl protecting group at the N4-position (Scheme 2), which we found to be more stable than the acetyl protecting group. The protected fC building blocks give in our hands higher overall yields than the corresponding fC phosphoramidite without any protecting group at N4.[16] The direct incorporation of fC also avoids the oxidative diol cleavage that was developed by Karino et al. to generate fC in oligonucleotides from a diol-precursor phosphoramidite.[11b] The direct incorporation of an fC building block via phosphoramidite chemistry is also possible for RNA. Here, however, a DMF protecting group was chosen for N4.[17]

The caC phosphoramidite reagent was developed based on results of Sekine and co-workers.[13] We chose to protect the N4 amine groups with an acetyl group and we used a methyl ester to protect the C5 carboxylic acid (Scheme 2). Both protecting groups are cleaved under the standard basic conditions typically needed to cleave the oligonucleotide from the solid support.

Starting point for the synthesis of the fC phosphoramidite building block is TBS-protected 5-iodo-2′-deoxycytidine (**1**) which was synthesized as described previously.[15a] The formyl group was introduced by a carbonylative coupling, which we already employed for the synthesis of the free fC nucleoside and an hmC phosphoramidite.[1e,15a] The resulting TBS-protected 5-formyl-2′-deoxycytosine was benzoylated to give **2**. Cleavage of the TBS groups was achieved with HF in pyridine.[18] When the latter reaction is carried out in EtOAc, the free nucleoside **3** precipitates and no further purification is needed. Finally, the 5′-hydroxyl group of compound **3** was protected with a DMT group and the 3′-hydroxyl group phosphitylated to give the final compound **4**.

In the first step towards the caC phosphoramidite building block, **1** was converted to the methyl ester in a Pd^0-catalyzed reaction.[19] During the reaction the CO pressure has to be ad-



Scheme 2. Synthesis of the fC phosphoramidite **4** and of the caC phosphoramidite **7**. X = fC or caC.

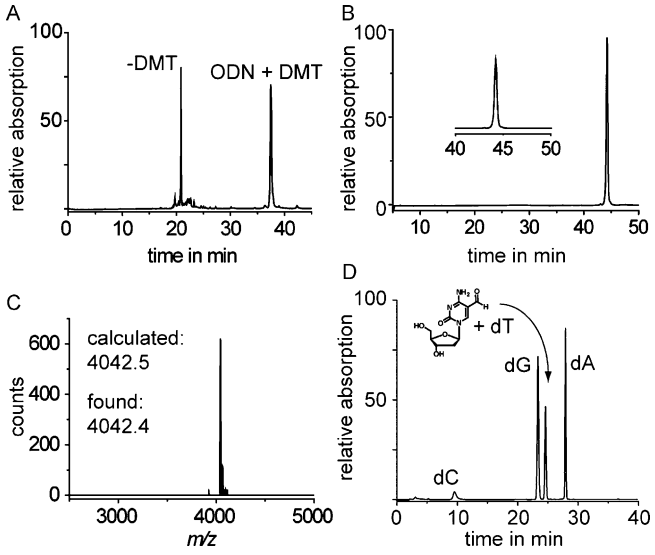


Figure 2. Analysis of ODN1 containing fC (Sequence of ODN1: 5′-GTAATGXGCTAG-3′; X = fC): A) Reversed-phase HPLC chromatogram of ODN1 (X = fC) directly after cleavage from the resin (0–50% buffer B in 45 min). B) Reversed-phase HPL chromatogram of ODN1 (X = fC) after cleavage of the DMT group and purification (0–20% buffer B in 45 min). C) MALDI-TOF spectrum of the purified strand ODN1 (X = fC). D) Digestion of the purified DNA strand ODN1 (X = fC) yields the canonical nucleosides plus fC, which co-elutes with dT.

justed carefully because otherwise a carbonylative Buchwald coupling occurs as a side reaction. Compound **5** was subsequently obtained after acetylation. In line with the reduced electron deficiency of the caC compound, the N4 acetyl protecting group is considerably more stable in compound **5** than in the fC derivative. For the cleavage of the silyl esters, we treated **5** with HF in pyridine/ethyl acetate which caused precipitation of pure **6**. The free nucleoside was finally DMT-protected and phosphitylated to give the caC phosphoramidite building block **7**.

For the incorporation of fC into oligonucleotides a standard solid-phase DNA synthesis was performed with phosphoramidite **4**. Coupling times for **4** were doubled to ensure good yields. As shown in Figure 2, the incorporation of fC proceeds without difficulty. After cleavage of the oligonucleotide from the resin and deprotection, the desired ODN1 with X = fC was obtained as the main product (Figure 2A, B). The same results were obtained for the 37mer ODN2 (Scheme 2). Deprotection was in both cases possible with either conc. NH_4OH or with 0.4 M NaOH in MeOH. MALDI-TOF analysis of the oligonucleotides proved the correct molecular weight of the synthesized strands (Figure 2C). After total digestion of the prepared oligonucleotides the four canonical bases were obtained with the fC base co-eluting with dT. Detection of the fC component was ensured by mass spectrometry (see Supporting Information).[7]

Incorporation of caC was possible with the phosphoramidite **7** using again rather standard solid-phase DNA synthesis conditions. The coupling times were extended to ensure good yields. In this case, however, standard deprotection conditions could not be employed, because treatment with ammonia or with K_2CO_3 in methanol would not yield the desired acid but either the amide or the methyl ester. Deprotection was therefore performed with NaOH in water/methanol.[15a] As shown in Figure 3A, this procedure allows to obtain the caC containing DNA strands as the only product. The caC containing DNA strands were subsequently purified by reversed phase HPLC (Figure 3B) and the correct composition was again proven by MALDI-TOF mass spectrometry (Figure 3C) and HPLC-MS analysis of the nucleotides obtained after total enzymatic digest. Under our conditions, the caC nucleoside eluted with a retention time of 5.3 min. Thus, fC- and caC-containing oligonucleotides can be efficiently prepared in high yields. With the improved phosphoramidite reagents the DNA strands are obtained under the described coupling and deprotection conditions in excellent purity.

In order to characterize and compare the base-pairing properties of hmC, fC and caC we measured melting points of sequence identical oligonucleotides under exactly the

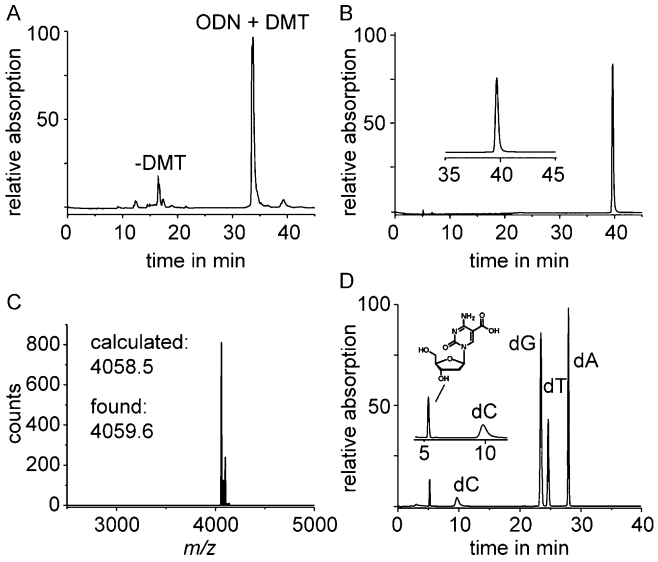


Figure 3. Analysis of ODN1 containing caC (Sequence of ODN1: 5′-GTAATGXGCTAG-3′; X = caC): A) Reversed-phase HPLC chromatogram of ODN1 (X = caC) directly after cleavage from the resin (0–50% buffer B in 45 min). B) Reversed-phase HPLC chromatogram of ODN1 (X = caC) after cleavage of the DMT group and purification (0–20% buffer B in 45 min). C) MALDI spectrum of the purified strand ODN1 (X = caC). D) HPLC chromatogram of purified ODN1 (X = caC) after enzymatic digestion.

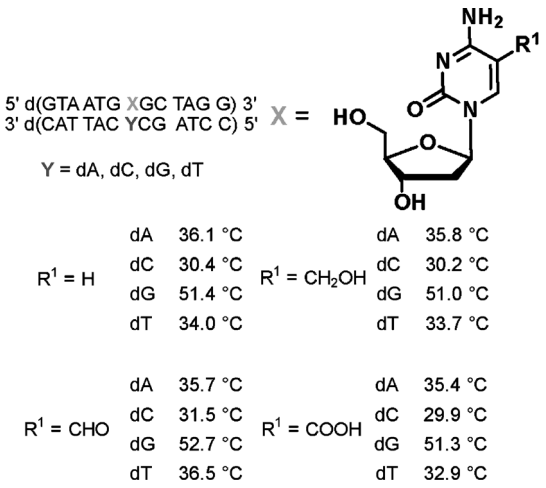
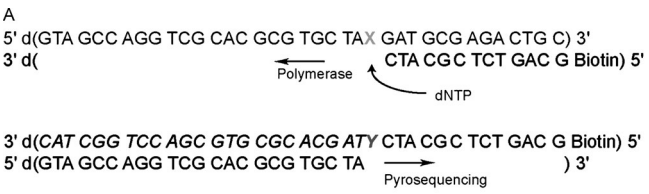


Figure 4. Melting point analysis of ODNs, which contain different 5-substituted cytosines. 13 mer duplexes (ODN1: 5′-GTAATGXGCTAG-3′) containing dC, hmC, fC, and caC as X was hybridized with different ODN3s (5′-CCTAGCYCATTAC-3′) containing either dA, dC, dG, or dT in the Y-position opposite the cytosine bases. The duplexes (ODN1:ODN3) were repeatedly heated and cooled between 20 and 80 °C. Melting points of all individual base–counterbase combinations are shown.

same conditions (Figure 4). To this end, ODN1, with the respective modified dC-bases present at position X, was mixed with the ODN3 counter strand, which contained individual canonical nucleosides at position Y. The solutions were repeatedly heated and cooled between 80 and 20 °C. In all cases the cytosine derivatives paired best with dG (R^1 = H: 51.4 °C; R^1 = CH_2OH: 51.0 °C; R^1 = CHO: 52.7 °C; R^1 = COOH: 51.3 °C). Second best base pairing was in most cases observed with dA. For fC we noted a slightly increased stability of the fC:dT base pair. As already observed for hmC[20] the base pairing properties of fC and caC differ only marginally from dC. In addition, the stability differences between the nC:dG (n = hm, f, ca) base pairs and all mismatches are so high that a severe mutagenic effect of the modified dC bases can be excluded based on the melting point data.

To investigate the mutagenic potential further we developed a novel primer extension assay which we coupled directly to a pyrosequencing procedure. For the study we hybridized the nC containing oligonucleotides (ODN2) with a biotinylated primer strand (ODN4) and used various high and low fidelity polymerases to extend the primer past the modified cytosines to the full length product as shown in Figure 5. The resulting transcripts were isolated using streptavidine coated sepharose beads and subsequently used as new templates for pyrosequencing. As such they were hybridized to the reverse primer (ODN5) and subjected to standard pyrosequencing conditions using the Klenow polymerase.[21] The experiment has the advantage that it provides incorporation data in a situation in which all triphosphates are present simultaneously and compete for incorporation.



B

	Klenow exo-	Pol η	Pol κ
dC	100 % dG	100 % dG	100 % dG
mC	100 % dG	100 % dG	100 % dG
hmC	100 % dG	100 % dG	100 % dG
fC	99 % dG 1 % dA	99 % dG 1 % dA	98 % dG 2 % dA
caC	100 % dG	100 % dG	100 % dG
$^{4\text{-NHOH}}$dC	52 % dG 48 % dA	56 % dG 44 % dA	65 % dG 35 % dA
dT	100 % dA	100 % dA	100 % dA

Figure 5. Sequencing based mutagenicity assay with different cytosine derivatives. All values are the average of three independent measurements. A) ODN2, containing the modification, was elongated with different polymerases, the transcript was isolated, hybridized to a reverse primer and sequenced in a pyrosequencer. B) Nucleosides that are incorporated opposite the different cytosine derivatives. In contrast to the known mutagenic $^{4\text{-NHOH}}$dC none of the other C5-modified cytosines are strongly mutagenic.

We examined the mutagenic potential of hmC, fC and caC first with the high fidelity polymerase Klenow *exo*$^-$.

However, due to the fact that modified bases are in vivo potentially replicated with a low fidelity polymerase, we also studied the common eukaryotic low fidelity polymerase eta (*Saccharomyces cerevisiae*) and polymerase kappa (*Homo sapiens*). For the study we used as control nucleosides dT as a non-mutagenic base and the mutagenic base $^{4\text{-NHOH}}$dC, in which the exocyclic amino group is replaced by a hydroxylamine group.[22] This base is known to exhibit its mutagenicity due to the increased presence of the imino tautomeric form, which is also believed to be the basis for the mutagenicity of fC.[23] All data obtained are compiled in Figure 5. As expected, dC directs exclusively the incorporation of dG into the primer, while dT instructs the incorporation of dA. $^{4\text{-NHOH}}$dC directs the incorporation of both dA and dG, which is in accordance with the proposal that it derives its mutagenic effect from the fact that it exists partially in the imino tautomeric form.[21b,23] To our surprise, we noted that fC in our coupled assay is, in contrast to previous reports, but in agreement with our melting point studies, only marginally mutagenic. All examined polymerases efficiently base paired fC with dG. In addition we indeed observed a small additional incorporation of dA. The incorporation is detected slightly above the error limit of the experiment, indicating that the formyl group promotes formation of the imino tautomeric form only to a very small extent. For caC and hmC we could not detect any other incorporation but dG. Our data clearly show that none of the new C5-modified nucleobases hmC, fC and caC are mutagenic. They all retain the base pairing behaviour of dC.

To investigate if the previously detected mutagenicity of fC may be caused by degradation products such as abasic sites, which may be generated by an accelerated proton catalyzed depyrimidination of fC, we performed stability studies. To this end fC and caC containing oligonucleotides were exposed to different pH conditions for 18 h and the decomposition was studied by HPLC. We noticed that neither the fC nor the caC containing ODN1 decomposed during this time in solutions buffered between pH 7 and 5 (37 °C). Under physiological conditions fC and caC are hence stable DNA modifications that show no increased lability. Only when we treated ODN1 (with X = fC and X = caC) under nonphysiological, mildly acidic conditions (0.01 % formic acid), we observed rapid decomposition of the strands. Under these conditions we observed an increased lability of fC and caC compared to dC because the dC containing control strands decomposed significantly slower.

We next recrystallized the fC base in order to obtain deeper insight into the intramolecular H-bond between the exocyclic amino group $NH_2(4)$ and the carbonyl–oxygen at C5, which was suggested to be the molecular reason for the increased mutagenicity. The structure of fC is depicted in Figure 6. The H-bond is indeed present in line with NMR studies carried out before.[12b,24] The low mutagenic potential of fC, however, suggests that this H-bond does not shift the amino/imino equilibrium substantially towards the imino-tautomeric form.

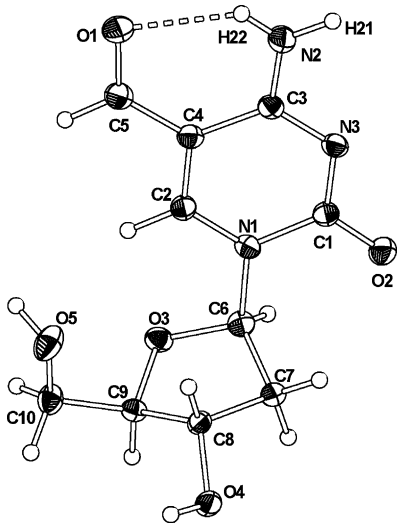


Figure 6. Molecular structure of fC in the crystal showing the intramolecular hydrogen bond between the aldehyde oxygen atom and the amino group. ORTEP representation, thermal ellipsoids are drawn at 50 % probability level.

Summary and Conclusion

In the last two years, it was discovered that the epigenetic base mC is further oxidized to hmC, fC and caC. Although the function of these new bases is still unclear at this point, they are present in substantial amounts in stem cells and neuronal cells. In principle oxidation of mC could occur in order to stimulate cellular mutagenic events. This is for example observed in the case of adenine, which is deaminated by special enzymes in immune cells to generate inosine, which base pairs with dC and not dT.[25] This mutagenic process is vital for a correct immune response. Alternatively, hmC, fC and caC could be new bases that have epigenetic functions. The previously reported mutagenicity of fC, which was found to be only slightly less mutagenic than $^{8\text{-oxo}}$dG,[11a] argued against a pure epigenetic function. The strong H-bond between the $NH_2(4)$ group and the carbonyl oxygen at C5 was suggested to promote the imino tautomeric state, which was proposed to stimulate pairing possibilities with dA.[11a,12a] In this study we re-examined the mutagenicity of fC and we measured for the first time the mutagenic effect of caC. Both compounds were measured against $^{4\text{-NHOH}}$dC, which is a known mutagenic dC modification due to an increased imine character. We also report a crystal structure of fC, which indeed shows the expected H-bond. Our data show that $^{4\text{-NHOH}}$dC is mutagenic and that it pairs both with dG and dA in accordance with increased existence of the imino-tautomeric form. The new bases fC and caC, however, only pair with dG and do not show such dual base pairing characteristics. They were also found to be stable under physiological conditions, which eliminates the possibility that they exhibit a mutagenic potential due to rapid depyrimidination. We believe that the previously observed mutagenicity of fC in vivo is therefore not caused by

a polymerase acting on the fC base but by molecular structures that are formed due to the in vivo action of either base- or nucleotide excision repair acting on the modifications. Indeed, it was recently shown that fC and caC are selectively excised by the enzyme TDG, which would generate an abasic site.[8a,9b] In summary, our data show that fC and caC are non-mutagenic and therefore can be involved in epigenetic programming of cells without disturbing the genetic sequence information.

Experimental Section

General remarks: Chemicals and solvents were purchased from ABCR, Alfa Aesar, Acros, Fluka, Sigma–Aldrich or TCI in the qualities puriss., p.a. or purum. Dry solvents (<50 ppm H_2O) were obtained from Fluka and Acros. All reactions employing dry solvents were performed under inert atmosphere (N_2). Technical grade solvents were distilled prior to use for column chromatography and liquid–liquid extractions on a rotary evaporator (Heidolph Laborota 4000). Reaction products were dried at high vacuum (0.1 mbar). Aqueous solutions were concentrated on a SpeedVac plus CS110 A or SPD 111V from Savant or lyophilized (Christ ALPHA 2-4). Thin layer chromatography (TLC) was performed with aluminium plates (silica gel 60 F254, 10×5 cm). Substances were visualized by illumination with UV light ($\lambda = 254$ nm). ESI-MS was performed on a Finnigan LTQ FTICR. MALDI-TOF was performed on a Bruker Autoflex II. NMR spectra were recorded on the following spectrometers: Varian Oxford 200, Bruker AC 300, Varian XL 400 and Bruker AMX 600. The chemical shifts (δ) are given in ppm, the coupling constants (J) in Hz. Exonuclease-deficient DNA polymerase I from *E. coli* (Klenow *exo*$^-$) was obtained from NEB, Pol η was purified as described before[26] and Pol κ was cloned from human cDNA and expressed in Rosetta 2 (DE3) cells.

Oligonucleotide synthesis and deprotection: DNA synthesis was performed on an Expedite 8909 Nucleic Acid Synthesis System (PerSeptive Biosystems) or an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions. Phosphoramidites for dA, dC, dG, dT and CPG carriers were obtained from Amersham, Glen Research or PE Biosystems. The hmC and fC phosphoramidites were dissolved in dry MeCN, the caC phosphoramidite in dry toluene. The oligonucleotides were removed from the resin under concomitant cleavage of the standard nucleobase protecting groups by treatment with 750 μL conc. NH_3 and 250 μL ethanol at room temperature for 18 h (fC) or treatment with 0.4 M NaOH in water/methanol 1:4. Subsequently, the solution was decanted from the resin. When the deprotection was carried out with NaOH, a 1 M TEA/AcOH solution (600 μL) was added and the solution concentrated to a final volume of 750 μL. After deprotection with NH_4OH the solution was concentrated to dryness and subsequently taken up in water. Please note that deprotection with NaOH is not compatible with DMF protecting groups.

HPLC and cleavage of DMT groups: Purification and analysis of ODNs was performed on a Waters system (Alliance 2695 with PDA 2996; preparative HPLC: 1525EF with 2484 UV detector) with VP 250/10 Nucleosil 100-7 C18 and VP 250/4 Nucleosil 20-3 C18 columns from Macherey–Nagel using a gradient of 0.1 M triethylamine/acetic acid in water and 80% acetonitrile. The oligonucleotides still containing the trityl group were deprotected by addition of 100 μL of an 80% acetic acid solution. After incubation at RT for 20 min, 100 μL of water together with 60 μL of a 3 M solution of sodium acetate were added. Finally, the ODN was precipitated by the addition of 1600 μL ethanol at −20 °C (30 min). The strands were stored in *dd*H_2O.

Enzymatic digestion: For the enzymatic digestion DNA mixtures (4 to 10 μg in a final volume of 100 μL H_2O) were heated to 100 °C for 5 min to denature the DNA and rapidly cooled on ice. Buffer E (10 μL, 300 mM ammonium acetate, 100 mM $CaCl_2$, 1 mM $ZnSO_4$, pH 5.7) and nuclease S1 (80 units, *Aspergillus oryzae*) were added to the mixture and incubated for 3 h at 37 °C. Addition of buffer F (12 μL, 500 mM Tris-HCl, 1 mM EDTA), antarctic phosphatase (10 units), snake venom phosphodiesterase I (0.2 units, *Crotalus adamanteus* venom) and incubation for further 3 h at 37 °C completed the digestion. The supernatant was removed, the volume reduced to 100 μL and measured with HPLC-ESI-MS.

HPLC-ESI-MS: The samples (100 μL injection volume) were analyzed by HPLC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system using gradient of 2 mM ammonium format in water and 80% acetonitrile over an Uptisphere120–3HDO column from Interchim.

Melting points: Melting profiles were measured on a Jasco V-650 spectrophotometer using quartz glass cuvettes with 1 cm path length. In these were in a total volume of 1 mL: 1 μmol ODN, 1 μM counterstrand, 150 mM NaCl and 10 mM Tris pH: 7.4. First, the oligonucleotides were hybridized by slowly cooling the samples down from 80 °C to 20 °C. The melting profiles started with a denaturing run (20 to 80 °C) with a slope of 0.5 °C min^{-1}. At least two denaturing and two renaturing ramps were performed and averaged for evaluation of the melting point. (T_M = zero-crossing of second derivative of the 400 nm-background corrected change in hyperchromicity at 260 nm.) For analysis of the data, the program Origin (Microcal) was used.

Pyrosequencing analysis: A 37 mer template DNA strand (ODN 2: 5′-d(GTA GCC AGG TCG CAC GCG TGC TAX GAT GCG AGA CTG C)-3′ was prepared containing either dC, mC, hmC, fC, caC, or $^{4\text{-NHOH}}$dC, or dT at position X. 10 pmol of the template were hybridized with a biotinylated primer (ODN 4: 5′-d(biotin-GCAGTCTCGCATC)-3′, Metabion). Subsequently, the primer extension experiments were performed with the exonuclease deficient DNA polymerase I from *E. coli* (Klenow *exo*$^-$), Pol η and Pol κ. The polymerases (1 U Klenow *exo*$^-$, 1 μM Pol η, or 0.5 μM Pol κ), 50 μM dNTPs and 0.5 μM dsDNA were incubated in a total volume of 20 μL 1× NEBuffer 2 for 30 min at different temperatures (30 °C for Pol η, 37 °C for Klenow *exo*$^-$ and Pol κ). To this solution 2 μL streptavidin sepharose beads (GE Healthcare, Uppsala, Sweden), 40 μL binding buffer (Qiagen, Hilden, Germany) and 18 μL ddH_2O were added. After agitation at 1400 rpm for 15 min the beads were captured with a Vacuum Prep Tool (Qiagen, Hilden, Germany), washed with 70% EtOH, 0.1 M NaOH and Washing Buffer (Qiagen, Hilden, Germany). The beads were dissolved in 25 μL Annealing Buffer (Qiagen, Hilden, Germany) containing 5 pmol sequencing primer ODN5 (5′-d(GTAGCCAGGTCGCACGCGTGCTA)-3′, Metabion, Martinsried, Germany). Pyrosequencing was performed on a PyroMark Q24 Pyrosequencer using standard conditions (Qiagen, Hilden, Germany). The data was analyzed by the software provided by the manufacturer. Peak heights were exported to Microsoft Excel and the average of all blank sites for the individual nucleobases calculated. This resulted in one blank for dT, one blank for dC, one blank for dG and one blank for dA. This value was subtracted from the values for the incorporation of the individual triphosphates. With these data the relative incorporation at every variable position was calculated. The data are the average of three measurements.

Recrystallization: Single crystals, suitable for X-ray diffraction, were obtained by very slow evaporation of the solvent of a fC solution in methanol. The crystals were introduced into perfluorinated oil and a suitable single crystal was carefully mounted on the top of a thin glass wire. Data collection was performed with an Oxford Xcalibur 3 diffractometer equipped with a Spellman generator (50 kV, 40 mA) and a Kappa CCD detector, operating with $Mo_{K\alpha}$ radiation ($\lambda = 0.71071$ Å). Data collection was performed with the CrysAlis CCD software;[27] CrysAlis RED software[27] was used for data reduction. Absorption correction using the SCALE3 ABSPACK multiscan method[28] was applied. The structures were solved with SHELXS-97,[29] refined with SHELXL-97[30] and finally checked using PLATON.[31] All hydrogen atoms involved in hydrogen bonding were found in the differential Fourier map and refined. Details for data collection and structure refinement are summarized in Table S4.

CCDC 843055 contains the supplementary crystallographic data for this paper. The data can be obtained free of charge from The Cambridge Crystallographic Date Centre via www.ccdc.cam.ac.uk/data_request/cif.

Acknowledgements

We thank Benjamin Clanner-Engelshofen (LMU Munich) for preparative work and we thank Markus Müller for critical reading of the manuscript. We thank the DFG Normalverfahren CA275/8-4, SFB 749 and SFB 646 for financial support. M.M. and T.P. thank the Fonds der Chemischen Industrie for pre-doctoral fellowships.

[1] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) 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; c) M. Münzel, D. Globisch, T. Carell, *Angew. Chem.* **2011**, *123*, 6588–6596; d) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503–5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; e) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367; f) A. Szwagierczak, A. Brachmann, C. S. Schmidt, S. Bultmann, H. Leonhardt, F. Spada, *Nucleic Acids Res.* **2011**, *39*, e181.

[2] a) Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451–464; b) H. Wu, A. C. D′Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes Dev.* **2011**, *25*, 679–684; c) H. Wu, A. C. D/′Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, Y. Zhang, *Nature* **2011**, *473*, 389–393; d) M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*, 241; e) K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. C. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343–348; f) J. Walter, *Cell Stem Cell* **2011**, *8*, 121–122; g) W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394–397; h) K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, *Cell Stem Cell* **2011**, *8*, 200–213; i) K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642–3647; j) G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398–402; k) C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–72; l) M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839–843; m) S. Ito, A. C. D′Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133.

[3] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583.

[4] R. Bonasio, S. J. Tu, D. Reinberg, *Science* **2010**, *330*, 612–616.

[5] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620.

[6] a) M. Akhtar, M. R. Calder, D. L. Corina, J. N. Wright, *Biochem. J.* **1982**, *201*, 569–580; b) P. M. Shaffer, C. A. Hsu, M. T. Abbott, *J. Bacteriol.* **1975**, *121*, 648–655; c) J. M. Simmons, T. A. Muller, R. P. Hausinger, *Dalton Trans.* **2008**, 5132–5142; d) J. A. Smiley, J. M. Angelot, R. C. Cannon, E. M. Marshall, D. K. Asch, *Anal. Biochem.* **1999**, *266*, 85–92.

[7] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.

[8] a) 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; b) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.

[9] a) 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; b) A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.

[10] J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, *Cell* **2011**, *145*, 423–434.

[11] a) H. Kamiya, H. Tsuchiya, N. Karino, Y. Ueno, A. Matsuda, H. Harashima, *J. Biochem.* **2002**, *132*, 551–555; b) N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456–2463.

[12] a) C. J. La Francois, Y. H. Jang, T. Cagin, W. A. Goddard, L. C. Sowers, *Chem. Res. Toxicol.* **2000**, *13*, 462–470; b) A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **2002**, *30*, 4068–4074; c) G. Kawai, T. Yokogawa, K. Nishikawa, T. Ueda, T. Hashizume, J. A. McCloskey, S. Yokoyama, K. Watanabe, *Nucleosides Nucleotides* **1994**, *13*, 1189–1199; d) Y. El Safadi, J. C. Paillart, G. Laumond, A. M. Aubertin, A. Burger, R. Marquet, V. Vivet-Boudou, *J. Med. Chem.* **2010**, *53*, 1534–1545.

[13] M. Sumino, A. Ohkubo, H. Taguchi, K. Seio, M. Sekine, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 274–277.

[14] T. Carell, M. Münzel, "Building Blocks and Methods for the Synthesis of 5-Hydroxymethylcytosine Containing Nucleic Acids", European Patent Application EP10191078.4, filed Nov. 12, 2010.

[15] a) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671–5673; b) S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **1997**, *25*, 553–558; c) A. S. Hansen, A. Thalhammer, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1181–1184; d) M. de Kort, P. C. de Visser, J. Kurzeck, N. J. Meeuwenoord, G. A. van der Marel, W. Ruger, J. H. van Boom, *Eur. J. Org. Chem.* **2001**, 2075–2082; e) Q. Dai, C.-X. Song, T. Pan, C. He, *J. Org. Chem.* **2011**, *76*, 4182–4188.

[16] Q. Dai, C. He, *Org. Lett.* **2011**, *13*, 3446–3449.

[17] a) H. Lusic, E. M. Gustilo, F. A. P. Vendeix, R. Kaiser, M. O. Delaney, W. D. Graham, V. A. Moye, W. A. Cantara, P. F. Agris, A. Deiters, *Nucleic Acids Res.* **2008**, *36*, 6548–6557; b) G. H. Clever, K. Polborn, T. Carell, *Angew. Chem.* **2005**, *117*, 7370–7374; *Angew. Chem. Int. Ed.* **2005**, *44*, 7204–7208.

[18] I. E. Szabo, T. C. Bruice, *Bioorg. Med. Chem.* **2004**, *12*, 4233–4244.

[19] Y. Nomura, N. Haginoya, Y. Ueno, A. Matsuda, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2811–2816.

[20] A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Chem. Commun.* **2011**, *47*, 5325–5327.

[21] a) M. Ronaghi, M. Uhlen, P. Nyren, *Science* **1998**, *281*, 363–365; b) M. Münzel, L. Lercher, M. Müller, T. Carell, *Nucleic Acids Res.* **2010**, *38*, e192.

[22] M. Münzel, C. Szeibert, A. F. Glas, D. Globisch, T. Carell, *J. Am. Chem. Soc.* **2011**, *133*, 5186–5189.

[23] D. M. Brown, M. J. E. Hewlins, P. Schell, *J. Chem. Soc. C* **1968**, 1925–1929.

[24] G. Kawai, T. Yokogawa, K. Nishikawa, T. Ueda, T. Hashizume, J. A. McCloskey, S. Yokoyama, K. Watanabe, *Nucleosides Nucleotides Nucleic Acids* **1994**, *13*, 1189–1199.

[25] E. Giblett, J. Anderson, F. Cohen, B. Pollara, H. Meuwissen, *Lancet* **1972**, *300*, 1067–1069.

[26] S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20720–20725.

[27] CrysAlis, 1.171.27p5 beta (release 01-04-2005 CrysAlis171.NET) (compiled Apr 1, 2005, 17:53:34) ed., Oxford Diffraction Ltd., **2005**.

[28] SCALE3, 1.0.4, gui:1.0.3 ed., Oxford Diffraction Ltd., **2005**.

[29] G. M. Sheldrick, SHELXS-97: Program for Crystal Structure Solution, University of Göttingen, Göttingen (Germany), **1997**.

[30] G. M. Sheldrick, SHELXL-97: Program for the Refinement of Crystal Structures, University of Göttingen, Göttingen (Germany), **1997**.

[31] A. L. Spek, PLATON: A Multipurpose Crystallographic Tool, Utrecht University, Utrecht (The Netherlands), **1999**.

Received: September 6, 2011
Published online: November 8, 2011

CHEMISTRY
A EUROPEAN JOURNAL

Supporting Information

Improved Synthesis and Mutagenicity of Oligonucleotides Containing 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine

Martin Münzel, Ulrike Lischke, Dimitrios Stathis, Toni Pfaffeneder, Felix A. Gnerlich, Christian A. Deiml, Sandra C. Koch, Konstantin Karaghiosoff, and Thomas Carell*[a]

chem_201102782_sm_miscellaneous_information.pdf

Buffer Systems for HPLC

buffer system I: analytical and preparative reversed phase HPLC

Buffer A: 0.1 M $NHEt_3OAc$

Buffer B: 0.1 M $NHEt_3OAc$ in 80% MeCN

buffer system II: analytical HPLC of enzymatic digestion products; LC-MS

Buffer C: 2 mM NH_4CHOO

Buffer D: 2 mM NH_4CHOO in 80% MeCN

always used with standard gradient: 0 ? 12 min; 0 % ? 3 % buffer D; 12 ? 60 min; 3 % ? 60 % buffer D; 60 ? 62 min; 60 % ? 100 % buffer D; 62 ? 70 min; 100 % buffer D; 70 ? 85 min; 100 ? 0 % buffer D; 85 ? 95 min; 0 % buffer D.

MALDI-MS

MALDI spectra were recorded on a *Bruker* autoflex II unit with an *MTP* AnchorChip var/384 target. Prior to the measurements the samples were desalted using *MF-Millipore* membrane filters (0.025 µM). Desalting times depended on the molecular weight of the analyte. (4000 Da: 30 min, 12000 Da: over night)

MALDI matrix:

HPA Crown: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 µl), ammonium hydrogencitrate (10 mg) in 500 µL ddH_2O and 500 µL MeCN

Sequences of DNA Strands

ODN1: 5' d(GTA ATG XGC TAG G)3' X = dC, mC, hmC, fC, caC

ODN2: 5' d(GTA GCC AGG TCG CAC GCG TGC TAX GAT GCG AGA CTG C) 3' X = dC, mC, hmC, fC, caC

ODN3: 5' d(CCT AGC YCA TTA C)3' Y = dA, dG, dC, dT

ODN4: 5' d(Bio GCA GTC TCG CAT C)3'

ODN5: 5' d(GTA GCC AGG TCG CAC GCG TGC TA)3'

Stability assay of fC and caC at the oligonucleotide level

The oligonucleotides used for this assay where HPLC purified and tested for analytical purity prior to the stability analysis. A solution of ODN1 (18 pmol/µL, X = dC, fC, caC) was spiked with 0.01 % HCOOH and directly incubated at 37°C. Additionally, ODN1 was incubated in buffered solution at pH 7.0 and 5.0 at 37°C. Sequential injections (10 µL, 180 pmol DNA)

every six hours and over a time frame of 18 h reveals the decomposition of the fC and caC strands in direct comparison with the control dC containing strand. The analytical HPLC runs where performed by application of buffer system I, with pH value strictly regulated to 7.00 (flow rate: 0.5 mL/min, gradient: 0? 20 % buffer B in 45 min).

LC MS analysis of digested DNA strands

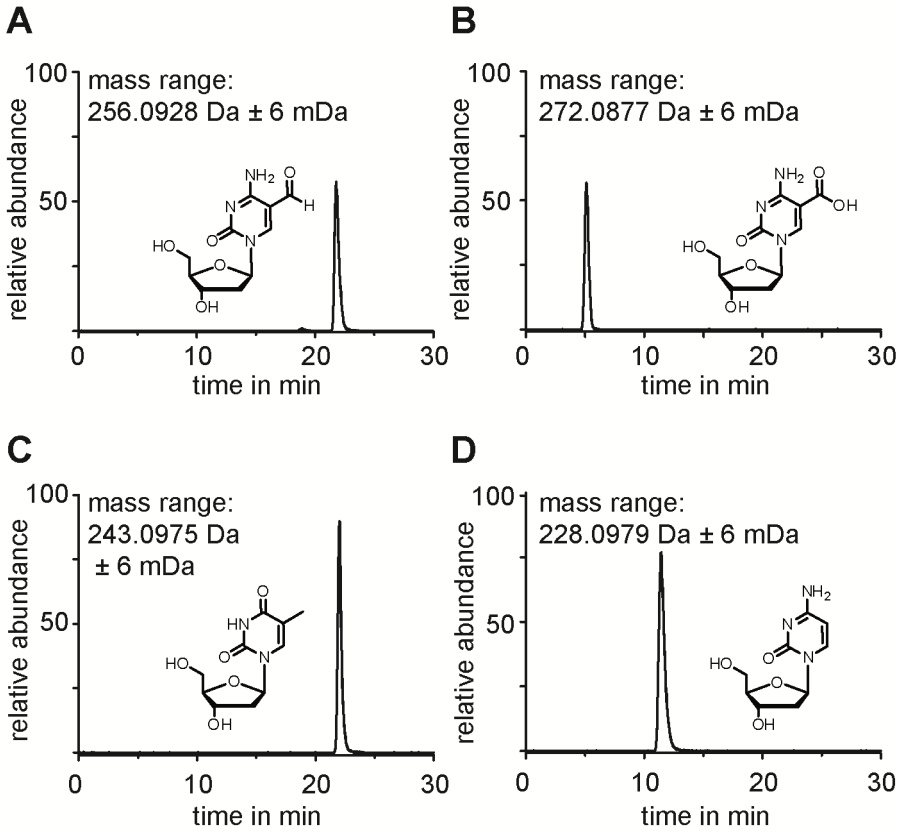


Figure S1: LC/MS analysis of the digested nucleoside mixtures. A) Mass filter for $^{5\text{-CHO}}$dC. B) Mass filter for $^{5\text{-COOH}}$dC. C) Mass filter for dT. d) Mass filter for dC.

Purification of human polymerase ?

Human polymerase kappa was cloned from human cDNA. The Stargate® system was used as a cloning strategy and the first codons were optimized for *E.coli* usage by using the primers below.

Primer:

5’P – AATG GGC CTG AAC GAT AAT AAA GCA GGA ATG GAA GGA T

5’P – TCCC TTG TTG GTG TTT CCT GTC CTC TTC

Furthermore the PCR-product was applied in a Stargate® reaction to form an entry clone (pPSG-IBA10) containing a short version of human polymerase kappa (pol?s, 19-526). The pPSG-IBA33 was used as expression plasmid, which was transformed in Rosetta 2 (DE3) cells. Pol?s was expressed using an autoinduction medium (ZYP-5052) adopted from Studier (Protein Expression and Purification 41 (2005) 207–234). The cells were grown at 37°C for 4 h and then for 16 h at 25°C and harvested by centrifugation. Subsequently the pellet was lysed in His-Buffer A (50 mM Tris pH = 7.4, 500 mM NaCl, 20 mM imidazol, 0.2 % Tween 20) using a french press and applied to a Ni-NTA superflow column. Pol?s was eluted in buffer B (50 mM Tris pH=7.4, 500 mM NaCl, 250 mM imidazol). Furthermore the protein was transferred to Heparin buffer A (50 mM Tris 7.4, 100 mM NaCl, 5 mM DTT, 5 % glycerol) and injected into a heparin column. A high salt buffer (Hep B, 50 mM Tris pH = 7.4, 800 mM NaCl, 5mM DTT, 5 % glycerol) was used for elution. Afterwards the central peak fractions were pooled and concentrated using an Amicon 30K 15 ml. As a polishing step the protein was applied to Superdex G75 with gel filtration buffer (25 mM HEPES pH = 7.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP) as mobile phase. The protein was again concentrated and stored at -20°C by supplementing the buffer with 50 % glycerol.

Purification of yeast polymerase ?

Pol ? was purified as described before.[1]

Crystal Structure of fC

A characteristic feature of the structure of fC is a strong intramolecular hydrogen bond (Table 1) between the oxygen atom of the CHO group and the adjacent NH_2 group involving the hydrogen atom H22. Due to this hydrogen bonding the orientation of the CHO-group is fixed and there is no disorder of this group with respect to a rotation around the C_{ring}-CHO bond observed. The most fascinating feature is the crystal packing of fC. It is determined by an extensive network of intermolecular hydrogen bonds. To this network participates the NH_2 group, both OH groups and the oxygen atoms of both carbonyl groups of each molecule. Interestingly, while the primary OH-group acts only as donor, the secondary OH-group acts both as donor and as acceptor (Fig. S2). In the crystal every molecule undergoes O–H···O hydrogen bonds with four neighboring molecules acting two times as acceptor via the carbonyl oxygen atoms and two times as donor via both OH-groups. With two further molecules N–H···O hydrogen bonds are formed. This results in the formation of layers (Fig. S3), which are strongly puckered due to the fixed angle between the five-membered and the six-membered ring of fC. These layers are packed to form the crystal.

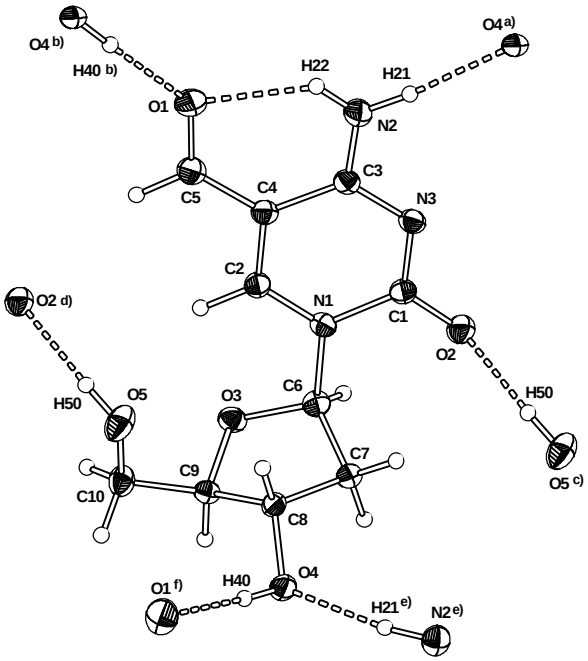


Figure S2: Crystal structure of fC showing the hydrogen bonds in which one molecule of fC is involved in the crystal. ORTEP representation, thermal ellipsoids are drawn at 50% probability level. Symmetry codes: a) 1-x, -0.5+y, 1-z; b) 2-x, -0.5+y, 2-z; c) -1+x, y, -1+z; d) 1+x, y, 1+z; e) 1-x, 0.5+y, 1-z; f) 2-x, 0.5+y, 2-z.

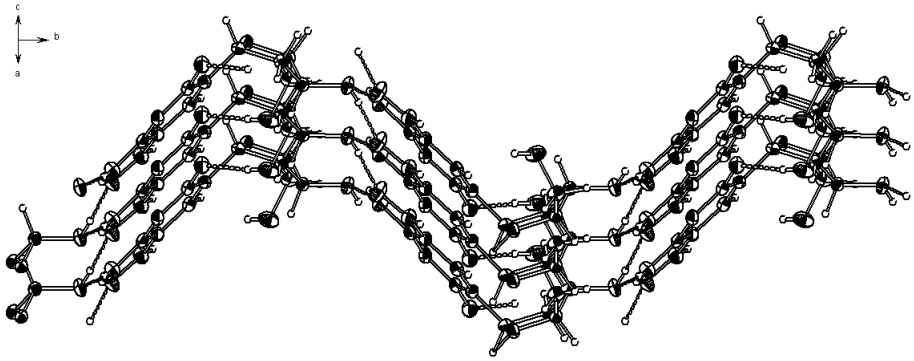


Figure S3: Crystal structure of fC, view of a puckered layer of hydrogen bonded molecules along the b-axis; thermal ellipsoids are drawn at 50% probability level.

Table S1: Hydrogen bonding parameters for fC in the crystal.

	D – H	H ··· A	D ··· A	D – H ··· A
N2 – H21 ··· O4[a)]	0.93(3)	1.92(3)	2.848(3)	173(3)
N2 – H22 ··· O1	0.83(3)	2.09(3)	2.745(3)	136(3)
O4 – H40 ··· O1[b)]	0.84(3)	1.86(3)	2.706(3)	176(3)
O5 – H50 ··· O2[c)]	0.86(3)	1.86(3)	2.719(3)	176(3)

Symmetry codes: [a)] 1-x, -0.5+y, 1-z; [b)] 2-x, -0.5+y, 2-z; [c)] -1+x, y, -1+z.

Table S2: Selected atom distances (in ?) in the molecular structure of fC in the crystal.

O1 – C5	1.242(3)	C2 – C4	1.371(4)
O2 – C1	1.244(3)	C3 – C4	1.446(4)
O3 – C6	1.420(3)	C4 – C5	1.435(4)
O3 – C9	1.450(3)	C6 – C7	1.525(4)
O4 – C8	1.424(3)	C7 – C8	1.517(4)
O6 – C10	1.421(3)	C8 – C9	1.531(4)
C9 – C10	1.513(4)	N1 – C2	1.339(3)
N1 – C6	1.509(3)	N1 – C1	1.429(3)
N2 – C3	1.331(4)	N3 – C3	1.335(3)
N3 – C1	1.344(3)		

Table S3: Selected bond angles (in °) in the molecular structure of fC in the crystal.

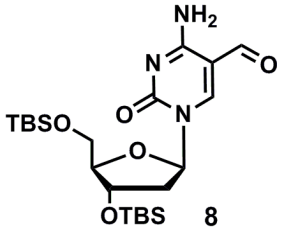
C6 – O3 – C9	110.3(2)	C6 – C9 – C10	115.2(2)
O3 – C9 – C8	104.4(2)	O3 – C9 – C10	110.3(2)
C1 – N1 – C2	120.5(2)	O5 – C10 – C9	112.0(2)
C1 – N1 – C6	117.3(2)	C2 – N1 – C6	122.2(2)
C1 – N3 – C3	120.0(2)	O2 – C1 – N3	123.4(2)
N1 – C1 – N3	119.6(2)	O2 – C1 – N1	117.0(2)
N1 – C2 – C4	121.3(2)	N2 – C3 – C4	121.3(2)
N3 – C3 – C4	122.1(2)	N2 – C3 – N3	118.5(2)
C2 – C4 – C3	116.6(2)	C2 – C4 – C5	118.4(2)
C3 – C4 – C5	125.0(2)	O1 – C5 – C4	124.0(3)
N1 – C6 – C7	112.0(2)	O3 – C6 – C7	106.5(2)
O3 – C6 – N1	108.3(2)	C6 – C7 – C8	102.1(2)
O4 – C8 – C9	113.5(2)	C7 – C8 – C9	101.7(2)
O4 – C8 – C7	110.4(2)		

Table S4: Details for X-ray data collection and structure refinement for compound fC

Empirical formula	$C_{10}H_{13}N_3O_5$
Formula mass	255.23
T[K]	173(2)
Crystal size [mm]	0.2×0.15×0.05
Crystal description	pale yellow block
Crystal system	monoclinic
Space group	*P*21
a [?]	5.0787(4)
b [?]	16.7550(12)
c [?]	6.5815(5)
ß [°]	103.791(9)
V [? ^{3}]	543.90(7)
Z	2
?$_{calcd.}$ [g cm^{-3}]	1.559
µ [mm^{-1}]	0.127
F(000)	268
T range [°]	4.30 – 33.51
Index ranges	-4=*h*=7
	-24=*k*=21
	-10=*l*=9
Reflns. collected	3819
Reflns. obsd.	2105
Reflns. unique	2933
	(R_{int} = 0.0322)
R_1, wR_2 (2s data)	0.0502, 0.0799
R_1, wR_2 (all data)	0.0806, 0.0895
GOOF on F^2	0.956
Peak/hole [e ? $^{-3}$]	0.28/-0.24

Synthesis of compounds 1-11

5-formyl-3', 5'-(*O-tert*-butyl-dimethylsilyl)-2'deoxycytidine (8)



8

The reaction was carried out as described previously. The scale could be increased to 16 g. [2]

^{1}H-NMR (300 MHz, $CDCl_3$): d (ppm) = 9.51 (s, 1H, C-C**H**=O), 8.57 (s, 1H, N-C**H**=C-CHO), 8.37 (s, 1H, 1x N**H**$_2$), 7.46 (s, 1H, 1x N**H**$_2$), 6.19 (t, ^{3}J = 6.1, 1H, O-C**H**-N), 4.40 – 4.32 (m, 1H, O-C**H**-CH_2-CH-N), 4.08 – 4.02 (m, 1H, O-CH-C**H**-CH_2-O), 3.95 (dd, ^{3}J = 2.7, ^{2}J = 11.7, 1H, 1x O-C**H**$_2$-CH-O), 3.78 (dd, ^{3}J = 2.6, ^{2}J = 11.6, 1H, 1x O-C**H**$_2$-CH-O), 2.59 (ddd, ^{3}J = 3.6, ^{3}J = 5.8, ^{2}J = 10.3, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.20 – 2.08 (m, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.89 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.88 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.10 (s, 3H, O-Si-C**H**$_3$), 0.08 (s, 6H, O-Si-C**H**$_3$), 0.07 (s, 3H, O-Si-C**H**$_3$).

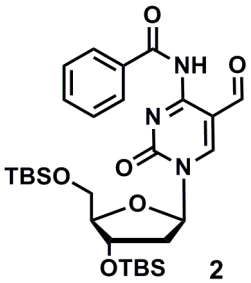
^{13}C-NMR (75 MHz, $CDCl_3$): d (ppm) = 187.1, 162.1, 153.1, 152.6, 104.9, 88.8, 87.9, 71.5, 62.6, 42.8, 25.9 (3C), 25.7 (3C), 18.4, 17.9, -4.5, -4.9, -5.2, -5.4.

HRMS (ESI+): calculated for $C_{22}H_{42}N_3O_5Si_2^+$ $[M+H^+]^+$: 484.2658, found: 484.2654

melting range: 150 - 152 °C

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3365 (w), 2952 (w), 2929 (w), 2857 (w), 1651 (s), 1245 (m), 1083 (s), 829 (s), 776 (s).

3',5'-(*tert*-butyl-dimethylsilyl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (2)



2

A round bottom flask was charged with 680 mg (1.41 mmol, 1.0 eq) **8** and 20 mL pyridine and was subsequently cooled to 0 °C. 875 µL (7.03 mmol, 5.0 eq) Bz-Cl were added slowly, the reaction mixture stirred at 0 °C for 30 min and for 2 h at room temperature. The reaction was stopped by addition of ice and diluted with 300 mL EtOAc. The solution was washed with sat. $NaHCO_3$ (1x 300 mL) and sat. NH_4Cl (1x 300 mL). The organic layer was separated, dried over $MgSO_4$ and subsequently evaporated to dryness. The crude product was purified by

column chromatography (*i*Hex/EtOAc 99:1? 29:1? 9:1) to give 435 mg (52 %) of **2** as a colorless oil along with 197 mg (20 %) of the double benzoylated derivative.

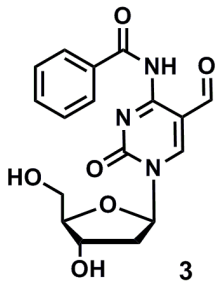
^{1}H-NMR (400 MHz, $CDCl_3$): d (ppm) = 9.60 (br s, 1H, N**H**), 8.75 (s, 1H, C**H**O), 8.11 (ddd, ^{4}J = 2.8, ^{3}J = 3.9, ^{3}J = 8.5, 3H, Ph-**H**, N-C**H**=C-CHO), 7.63 – 7.44 (m, 3H, Ph-**H**), 6.22 (t, ^{3}J = 6.2, 1H, O-C**H**-N), 4.40 – 4.34 (m, 1H, O-C**H**-CH_2-CH-N), 4.10 (dd, ^{3}J = 2.6, ^{3}J = 5.7, 1H, O-CH-C**H**-CH_2-O), 3.97 (dd, ^{3}J = 2.6, ^{2}J = 11.6, 1H, 1x O-C**H**$_2$-CH-O), 3.80 (dd, ^{3}J = 2.5, ^{2}J = 11.6, 1H, 1x O-C**H**$_2$-CH-O), 2.71 – 2.61 (m, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.13 (dt, ^{3}J = 6.1, ^{2}J = 13.5, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.90 (2x s, 18H, O-Si-C(C**H**$_3$)$_3$), 0.11 (s, 3H, O-Si-C**H**$_3$), 0.10 (s, 3H, O-Si-C**H**$_3$), 0.09 (s, 3H, O-Si-C**H**$_3$), 0.07 (s, 3H, O-Si-C**H**$_3$).

^{13}C-NMR (101 MHz, $CDCl_3$): d (ppm) = 187.5, 174.7, 170.5, 159.4, 142.9, 133.6, 133.0, 130.1 (2C), 128.4 (2C), 91.6, 89.1, 88.4, 71.0, 62.6, 42.7, 25.9 (3C), 25.7 (3C), 18.4, 18.0, -4.5, -4.9, -5.3, -5.4.

HRMS (ESI+): calculated for $C_{29}H_{46}N_3O_6Si_2^+$ $[M+H^+]^+$: 588.2920, found: 588.2920.

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2954 (w), 2929 (w), 2885 (w), 2856 (w), 1718 (m), 1692 (m), 1658 (s), 1567 (s), 1471 (s), 1250 (s), 1080 (s), 829 (s), 776 (s).

4-*N*-benzoyl-5-formyl-2'deoxycytidine (3)



3

In a polypropylene tube 395 mg (0.67 mmol, 1.0 eq) of the TBS-protected nucleoside **2** were dissolved in dry EtOAc (final concentration 60 mM). Subsequently 272 µL (3.36 mmol, 5.0 eq) pyridine and 262 µL (10.1 mmol, 15 eq) HF (70 % in pyridine) were added and the reaction mixture stirred 14 h at rt. During this time a white solid precipitated. 1 mL/mmol TMSOMe were added and the reaction mixture stirred another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min). The crude product (purity >90 %) was not further purified and directly taken on to the next reaction. Yield: 198 mg (82 %).

^{1}H-NMR (400 MHz, pyridine-D_5/DMSO-D_6): d (ppm) = 9.82 (br s, 1H, N**H**), 9.64 (s, 1H, C**H**O), 8.22 (d, ^{3}J = 7.1, 2H, Ph-**H**), 7.62 (s, 1H, N-C**H**=C-CHO), 7.55 (d, ^{3}J = 7.4, 1H, Ph-**H**), 7.48 (t, ^{3}J = 7.4, 2H, Ph-**H**), 6.60 (dd, ^{3}J = 4.9, ^{3}J = 6.3, 1H, O-C**H**-N), 4.90 (dd, ^{3}J = 5.8, ^{3}J = 10.5, 1H, O-C**H**-CH_2-CH-N), 4.46 (dt, ^{3}J = 2.9, ^{3}J = 4.7, 1H, O-CH-C**H**-CH_2-O), 4.22 (dd, ^{3}J = 3.0, ^{2}J = 12.0, 1H, 1x O-C**H**$_2$-CH-O), 4.12 (dd, ^{3}J = 2.9, ^{2}J = 12.0, 1H, 1x O-C**H**$_2$-CH-O), 2.94 (ddd, ^{3}J = 5.8, ^{3}J = 6.3, ^{2}J = 13.5, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.69 (ddd, ^{3}J = 4.9, ^{3}J = 6.1, ^{2}J = 13.5, 1H, 1x O-CH-C**H**$_2$-CH-N).

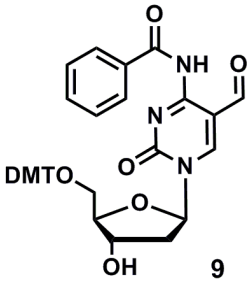
^{13}C-NMR (101 MHz, $CDCl_3$): d (ppm) = 188.4, 159.1, 153.8, 152.2, 149.6, 134.1, 132.6, 128.6 (2C), 127.9 (2C), 106.5, 88.9, 87.9, 69.2, 60.5, 41.8.

HRMS (ESI+): calculated for $C_{17}H_{18}N_3O_6^+$ $[M+H^+]^+$: 360.1190, found: 360.1191.

Melting Range: >200 °C decomposition.

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3477 (br, m), 3306 (br, m), 2988 (w), 2931 (w), 2906 (w), 2851 (w), 1721 (s), 1650 (s), 1570 (s), 1473 (s), 1251 (s), 1238 (s), 1092 (s), 795 (m).

5'-(dimethoxytrityl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (9)



9

A round bottom flask was charged with rigorously dried 3 Å molecular sieves. 146 mg (0.41 mmol, 1.0 eq) of the nucleoside **3** (final concentration: 100 mM) and 151 mg (0.45 mmol, 1.1 eq) DMT-Cl were dissolved in pyridine. The reaction mixture was stirred for 14 h and the molecular sieves subsequently filtered off. The filtrate was evaporated to dryness and the crude product purified by column chromatography. Eluent for column chromatography: DCM/MeOH (49:1; 0.1 % NEt_3). Yield: 170 mg (63 %) of **9** as a pale yellow solid.

^{1}H-NMR (400 MHz, DMSO-D_6): d (ppm) = 8.97 (s, 1H, C-C**H**=O), 8.81 (s, 1H, N-C**H**=C-CHO), 8.00 – 7.89 (m, 2H, Ph-**H**), 7.70 (t, ^3J = 7.3, 1H, Ph-**H**), 7.61 (t, ^3J = 7.3, 2H, Ph-**H**), 7.33 (m, 4H, DMT-**H**), 7.27 – 7.18 (m, 5H, DMT-**H**), 6.88 (dd, ^4J = 0.7, ^3J = 9.1, 4H, DMT-**H**), 6.09 (dd, ^3J = 5.1, ^3J = 6.4, 1H, O-C**H**-N), 5.42 (d, ^3J = 4.8, 1H, O**H**), 4.34 – 4.26 (m, 1H, O-C**H**-CH_2-CH-N), 4.07 (dd, ^3J = 4.3, ^3J = 7.9, 1H, O-CH_2-C**H**-CH-O), 3.71 (2x s, 6H, O-C**H**$_3$), 3.34 – 3.26 (m, 2H, O-C**H**$_2$-CH-O), 2.49 – 2.43 (m, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.34 (ddd, ^3J = 4.5, ^3J = 6.3, ^2J = 11.3, 1H, 1x O-CH-C**H**$_2$-CH-N).

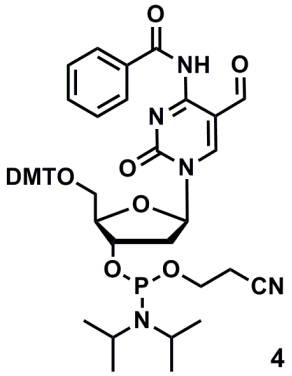
^{13}C-NMR (101 MHz, DMSO-D_6): d (ppm) = 189.0, 163.8, 159.3, 158.2 (2C), 154.3, 152.7, 144.5, 135.4, 135.2, 133.3, 133.2, 129.77 (2C), 129.76 (2C), 129.1 (4C), 128.0 (2C), 127.7 (2C), 126.9, 114.1 (2C), 113.3 (2C), 105.8, 87.7, 86.4, 86.0, 69.4, 63.0, 55.0 (2C), 40.7.

HRMS (ESI+): calculated for $C_{38}H_{36}N_3O_8^+$ $[M+H^+]^+$: 662.2497, found: 662.2500.

Melting Range: 115 – 123 °C

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2934 (w), 1716 (m), 1650 (s), 1601 (m), 1565 (s), 1508 (m), 1470 (s), 1247 (s), 1174 (m), 1089 (m), 1068 (m), 1031 (m), 827 (m), 790 (m), 701 (s).

3'-(di*iso*propylamino-cyanoethoxyphosphino)-5'-(dimethoxytrityl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (4)



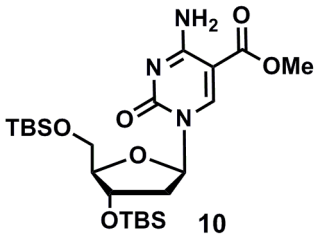
4

In a Schlenck tube 100 mg (0.15 mmol, 1.0 eq) of DMT-protected nucleoside **9**, 14 mg (0.13 mmol, 0.5 eq) di*iso*propyltetrazolide, and 55 µL (0.18 mmol, 1.2 eq) bis(di*iso*propylamino)(2-cyanoethoxy)phosphine, were dissolved in 10 mL rigorously degassed DCM and the solution degassed three more times (freeze, pump, thaw). The reaction mixture was stirred at rt for 15 h and subsequently concentrated in an argon atmosphere. The crude product was purified by column chromatography. The test tubes that were used to collect the fractions were flushed with argon prior to use. Eluent for column chromatography: DCM/MeOH (49:1, 0.1 % NEt_3). Yield: 113 mg (95 %, colorless foam) of **4** as a mixture of two diastereomers on P.

^{31}P-NMR (81 MHz, Acetone): d (ppm) = 149.64, 149.45.

HRMS (ESI+): calculated for $C_{47}H_{53}N_5O_9P^+$ $[M+H^+]^+$: 862.3575, found: 862.3580.

3',5'-(*tert*-butyl-dimethylsilyl)-5-carboxymethyl-2'deoxycytidine (10)



10

In a glass autoclave 1.00 g (1.72 mmol, 1.0 eq) **1** and 49 mg (0.19 mmol, 0.1 eq) $Pd(MeCN)_2Cl_2$ were dissolved in 20 mL methanol and subsequently 600 µL (3.44 mmol, 2.0 eq) DIPEA added. The autoclave was flushed once with CO and the reaction mixture then stirred for 18 h at 60 °C with a CO pressure of 3 bar. After completion of the reaction the gas was discharged and the solvent evaporated *in vacuo*. The crude product was purified by column chromatography (DCM/MeOH 99:1, dry loaded) to give 696 mg (84 %) of **10** as a yellowish solid.

^{1}H-NMR (300 MHz $CDCl_3$): d (ppm) = 8.59 (s, 1H, N-C**H**=C-COOMe), 7.93 (s, 1H, 1x N**H**$_2$), 7.31 (s, 1H, 1x N**H**$_2$), 6.15 (dd, *J*=5.9, 7.3, 1H, O-C**H**-N), 4.36 – 4.29 (m, 1H, O-C**H**-

CH_2-CH-N), 4.03 (dd, ^3J = 2.9, ^3J = 5.6, 1H, O-CH_2-C**H**-CH-O), 3.83 (dd, ^3J = 3.2, ^2J = 11.3, 1H, 1x O-C**H**$_2$-CH-O), 3.79 (s, 3H, C-CO-O-C**H**$_3$), 3.74 (dd, ^3J = 3.1, ^2J = 11.3, 1H, 1x O-C**H**$_2$-CH-O), 2.58 (ddd, ^3J = 2.4, ^3J = 5.8, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 1.94 (ddd, ^3J = 6.1, ^3J = 7.3, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.86 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.85 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.05 (s, 6H, Si-C**H**$_3$), 0.04 (s, 6H, Si-C**H**$_3$).

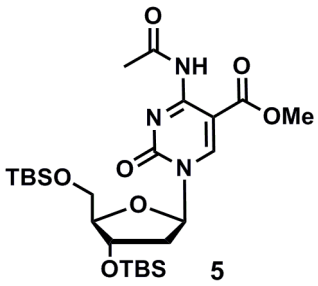
^{13}C-NMR (75 MHz, $CDCl_3$): d (ppm) = 165.1, 163.7, 154.0, 147.4, 95.1, 88.7, 87.7, 72.7, 63.0, 51.8, 42.8, 25.8 (3C), 25.7 (3C), 18.2, 17.9, -4.7, -5.0, -5.6, -5.7.

HRMS (ESI+): calculated for $C_{23}H_{44}N_3O_6Si_2^+$ $[M+H^+]^+$: 514.2763, found: 514.2762

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2930 (m), 2857 (m), 1720 (s), 1674 (s), 1501 (m), 1101 (s), 1075 (s), 822 (s), 776 (s).

melting range: 129 - 131 °C

3',5'-(*tert*-butyl-dimethylsilyl)-4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (5)



5

In a round bottom flask 550 mg (1.07 mmol, 1.0 eq) **57** were dissolved in 50 mL THF. Subsequently 7 mL (53 mmol, 50 eq) Ac_2O, 87 µL (1.07 mmol, 1.0 eq) pyridine and 7 mg (0.05 mmol, 0.1 eq) DMAP were added. The mixture was stirred for 15 h at rt. After completion of the reaction 5 mL MeOH were added and the solution was concentrated *in vacuo* to give 585 mg (98 %) of **5** as a yellowish solid.

^{1}H-NMR (599 MHz, $CDCl_3$): d (ppm) = 10.71 (s, 1H, N**H**), 8.80 (s, 1H, N-C**H**=C-COOMe), 6.13 (dd, ^3J = 6.1, ^3J = 7.1, 1H, O-C**H**-N), 4.35 (dt, ^3J = 2.2, ^3J = 5.5, 1H, O-C**H**-CH_2-CH-N), 4.11 (q, ^3J = 2.7, 1H, O-CH_2-C**H**-CH-O), 3.88 – 3.84 (m, 4H, CO-O-C**H**$_3$, 1x O-C**H**$_2$-CH-O), 3.77 (dd, ^3J = 2.9, ^2J = 11.4, 1H, 1x O-C**H**$_2$-CH-O), 2.69 (ddd, ^3J = 2.3, ^3J = 5.9, ^2J = 13.5, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.64 (s, 3H, N-CO-C**H**$_3$), 1.99 (ddd, ^3J = 6.0, ^3J = 7.3, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.89 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.84 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.08 (s, 3H, Si-C**H**$_3$), 0.07 (s, 3H, Si-C**H**$_3$), 0.06 (s, 3H, Si-C**H**$_3$), 0.04 (s, 3H, Si-C**H**$_3$).

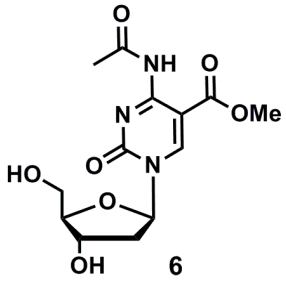
^{13}C-NMR (151 MHz, $CDCl_3$): d (ppm) = 172.0, 164.9, 159.6, 153.1, 148.7, 95.7, 89.4, 88.7, 72.9, 63.1, 52.4, 42.9, 27.2, 25.8 (3C), 25.7 (3C), 18.2, 18.0, -4.6, -4.9, -5.6, -5.7.

HRMS (ESI+): calculated for $C_{25}H_{46}N_3O_7Si_2^+$ $[M+H^+]^+$: 556.2869, found: 556.2866

Melting Range: 132 - 134 °C

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2927 (w), 2857 (w), 1644 (s), 1537 (m), 1284 (s), 1255 (s), 1119 (s), 1006 (m), 817 (s), 761 (s).

4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (6)



6

In a polypropylene tube 473 mg (0.85 mmol, 1.0 eq) of the TBS-protected nucleoside **5** were dissolved in dry EtOAc (final concentration 60 mM). Subsequently 332 µL (12.8 mmol, 5.0 eq) pyridine and 343 µL (4.26 mmol, 15 eq) HF (70 % in pyridine) were added and the reaction mixture stirred 14 h at rt. During this time a white solid precipitated. 1 mL/mmol TMSOMe were added and the reaction mixture stirred another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min). The crude product was further purified by column chromatography (DCM/MeOH 19:1) to give 212 mg (76 %) of **6** as a colorless solid.

^{1}H-NMR (400 MHz, CD_3OD): d (ppm) = 9.12 (s, 1H, N-C**H**=C-COOMe), 6.18 (t, ^{3}J = 6.0, 1H, O-C**H**-N), 4.38 (dt, ^{3}J = 4.6, ^{3}J = 6.2, 1H, O-C**H**-CH_2-CH-N), 3.99 (dd, ^{3}J = 3.4, ^{3}J = 7.5, 1H, O-CH_2-C**H**-CH-O), 3.89 – 3.82 (m, 4H, CO-O-C**H**$_3$, 1x O-C**H**$_2$-CH-O), 3.76 (dd, ^{3}J = 3.4, ^{2}J = 12.0, 1H, 1x O-C**H**$_2$-CH-O), 2.47 (ddd, ^{3}J = 4.8, ^{3}J = 6.4, ^{3}J = 13.7, 1H, O-CH-C**H**$_2$-CH-N), 2.26 – 2.16 (m, 1H, O-CH-C**H**$_2$-CH-N), 2.02 (s, 3H, N-CO-C**H**$_3$).

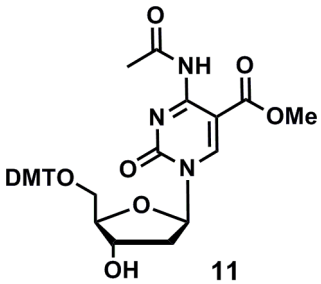
^{13}C-NMR (101 MHz, CD_3OD): d (ppm) = 173.6, 166.8, 165.4, 156.7, 150.2, 97.2, 89.4, 88.7, 71.4, 62.3, 52.6, 42.8, 20.6.

HRMS (ESI+): calculated for $C_{13}H_{18}N_3O_7^+$ $[M+H^+]^+$: 328.1139, found: 328.1138

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3305 (br, m), 2928 (w), 1634 (s), 1584 (m), 1484 (m), 1437 (s), 1326 (m), 1296 (s), 1255 (s), 1101 (s), 792 (s).

melting range: > 200 °C decomposition

5'-(dimethoxytrityl)-4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (11)



11

A round bottom flask was charged with rigorously dried 3 Å molecular sieves. 173 mg (0.53 mmol, 1.0 eq) of the nucleoside **6** (final concentration: 66 mM) and 185 mg (0.53 mmol, 1.0 eq) DMT-Cl were dissolved in pyridine. The reaction mixture was stirred for 14 h and the molecular sieves subsequently filtered off. The filtrate was evaporated to dryness and the crude product purified by column chromatography. Eluent for column chromatography: DCM/MeOH (49:1). Yield: 158 mg (48 %) of **11** as a colorless solid.

^{1}H-NMR (400 MHz, CD_3CN): d (ppm) = 10.69 (s, 1H, N**H**), 8.89 (s, 1H, N-C**H**=C-COOMe), 7.46 – 7.41 (m, 2H, DMT), 7.35 – 7.21 (m, 7H, DMT), 6.88 – 6.82 (m, 4H, DMT), 6.05 (t, ^{3}J = 5.9, 1H, O-C**H**-N), 4.35 (m, 1H, O-C**H**-CH_2-CH-N), 4.05 (dd, ^{3}J = 3.9, ^{3}J = 7.5, 1H, O-CH_2-C**H**-CH-O), 3.76 (s, 6H, CH=C-O-C**H**$_3$), 3.44 – 3.38 (m, 1H, 1x O-C**H**$_2$-CH-O), 3.23 (dd, ^{3}J = 3.9, ^{2}J = 10.9, 1H, 1x O-C**H**$_2$-CH-O), 3.18 (s, 3H, CO-O-C**H**$_3$), 2.56 – 2.47 (m, 4H, 1x O-CH-C**H**$_2$-CH-N, N-CO-C**H**$_3$), 2.34 – 2.26 (m, 1H, 1x O-CH-C**H**$_2$-CH-N).

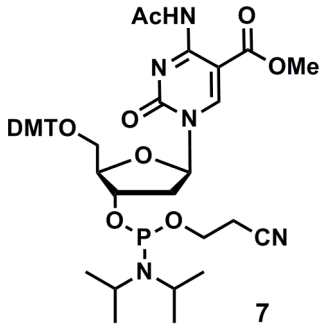
^{13}C-NMR (101 MHz, CD_3CN): d (ppm) = 172.0, 166.5, 160.7, 159.82, 159.81, 154.2, 150.1, 145.9, 136.94, 136.87, 131.08 (2C), 131.04 (2C), 129.1 (2C), 128.9 (2C), 128.0, 114.1 (4C), 97.1, 88.9, 87.8, 87.4, 71.1, 63.7, 56.0 (2C), 52.8, 42.3, 27.1.

HRMS (ESI+): calculated for $C_{34}H_{36}N_3O_9^+$ $[M+H^+]^+$: 630.2446, found: 630.2443

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3277 (m), 3223 (m), 2952 (m), 2935 (w), 2837, 1659 (s), 1547 (m), 1506 (s), 1439 (m), 1329 (m), 1296 (s), 1255 (s), 1174 (s), 1098 (s), 837 (m), 792 (s), 701 (s).

melting range: 97 - 98 °C

3'-(di*iso*propylamino-cyanoethoxyphosphino)-5'-(dimethoxytrityl)-5-carboxymethyl-2'deoxycytidine (7)



7

In a Schlenck tube 158 mg (0.25 mmol, 1.0 eq) of DMT-protected nucleoside **11** (final concentration: 50 mM), 22 mg (0.13 mmol, 0.5 eq) di*iso*propyltetrazolide, and 98 µL (0.30 mmol, 1.2 eq) bis(di*iso*propylamino)(2-cyanoethoxy)phosphine, were dissolved in 10 mL rigorously degassed DCM and the solution degassed three more times (freeze, pump, thaw). The reaction mixture was stirred at rt for 15 h and subsequently concentrated in an argon atmosphere. The crude product was purified by column chromatography. The test tubes that were used to collect the fractions were flushed with argon prior to use. Eluent for column chromatography: DCM/MeOH (49:1, 0.1 % NEt_3). Yield: 196 mg (94 %, colorless foam) of **7** as a mixture of two diastereomers on P.

The compound was directly used for solid phase DNA synthesis. Its identity was proven by successful incorporation into DNA.

^{31}P-NMR (81 MHz, Acetone): d (ppm) = 149.46, 149.32.

HRMS (ESI+): calculated for $C_{43}H_{53}N_5O_{10}P^+$ $[M+H^+]^+$: 830.3525, found: 830.3530.

[1] S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell, *Proceedings of the National Academy of Sciences* **2010**, *107*, 20720-20725.

[2] a) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367; b) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Organic Letters* **2010**, *12*, 5671-5673.