DISS. ETH No.29125

Quantification and mapping of *O*^{6}-alkylguanine repair in the nucleosome

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

EMMA SOFIA SANDELL

M. Sc. in Food Science, ETH Zürich

Born on 05.01.1993

accepted on the recommendation of

Prof. Dr. Shana J. Sturla

Prof. Dr. Dennis Gillingham

Prof. Dr. Ferdinand von Meyenn

2023

“Okay, let’s go!”

Table of Contents

Acknowledgements

I joined the laboratory of toxicology, because I liked the research and was up to a challenging experience. The Ph.D. journey was indeed like a ride in the Superpolyp, with ups and downs and often also just going in circles. But I was always happy to go to work, even in the periods when nothing seemed to work, because I had so many amazing and fun people to share the ride with.

First and foremost, I would like to thank Prof. Dr. Shana Sturla for giving me the opportunity to join this great research group. Your support and trust to try out new things, and the freedom you gave me to follow own ideas, pushed me to do and learn so much more than I could have ever imagined. In these years I've grown to become a more independent researcher and a more confident person. Also thank you for providing me with the opportunities to attend conferences and grant retreats.

Also I want to thank Prof. Dr. Dennis Gillingham and Prof. Dr. Ferdinand von Meyenn for being part of my doctoral committee. I am grateful for your scientific advice and experimental suggestions on my research projects throughout the years.

Thank you, Dr. Yael David for all the brainstorming meetings and for the great collaboration and contribution to Chapter 3 of this thesis.

And of course, I would like to thank Dr. Claudia Aloisi; how lucky I was to have the best second supervisor I could have ever wished for! You have been there from the very beginning of my Ph.D., helping me draft my research plan, discussing results and procedures of my projects. You have taught me so much, supported me and boosted my confidence whenever needed. Thank you!

And now to all dear lab members. Laura, thanks for all the laughter ("I like your laugh"), talks and discussions, the office and youtube quotes, the chemdraw help and most importantly for introducing me to Martijn. I'm sorry that I still cannot do the lyrics right. YangJiang, thank you for playing beautiful guitar music in the office, for the bioinformatics support and for being the most thankful prank-receiver. Jeanne, thanks for always having an open ear, helping me out of the forest and all the mr. bean jokes and horse filters. Ally-gator, thanks for ice-skating, coffee drinking and for the most amazing salt collection. Cécile, my twin sister, I still miss to hear your jokes and laughter across the hallway. You had simply the best jokes! And Fabrice Pastice, you set the corner stone to the snack museum, which had an impact on everyone's life. Also thank you for your input of the scientific steepness curve. I hope you can publish that soon! Maja, thank you for your endless supply with coffee and popcorn and Raul, for funny faces on mandarins and the crab-raves. I think it's time you stopped drinking lemon water. Thanks to the current group members Niko, Jasmina, Lina, Navnit, Hélène, and Emma D. for having been part of my PhD journey. Vakil, thank you for your Nordic vibes and your juicy stories. The best ones always started with "in my village..". Now it is only between you and Laura, who is the true Ryan. Thank you, Frida, for helping me with enzyme kinetics, and revising parts of this thesis. Also Jake, thank you so much for revising part of my thesis. Sabrina, it was fun doing and discussing science with you and Katie, thank you for (still) joining me to superkondi and all your advice on PhD and life. Gabriele and Georg, thank you for supporting me with your knowledge and also Nathalie and Sabine for helping out with all my cell studies and MS problems. I was very happy you were there, and I could always rely on you! Also I would like to thank former members, Susanne Geisen, Jingjing Sun, Junzhou Wu, and my master thesis advisor Michael Räz, who introduced me to the lab and

the science. Prisca Iseli and Daniela Kalbermatter, thank you for the administrative support. Many thanks also go to all the students I had the chance to work with: Kento Schreier, Jana Bangerter, Katrin Halter, Vera Stäheli, Alona Slastennikova and Mara Bless.

I also want to thank my family for everything you have done for me. How lucky I am to have such a caring, warm-hearted, and supportive family!

Finally, thank you Vali for being so encouraging and supporting throughout the years and for all the laughter and fun. I'm very much looking forward to finally set the sails and discover unknown waters with you!

Abstract

Alkylation is a major form of DNA damage induced by *N*-nitroso compounds. These genotoxic chemicals are present in tobacco smoke, fuel combustion products, and are produced endogenously in the stomach and colon. Alkylation at the O^6-position of guanine results in the formation of O^6-methylguanine (O^6-MeG) or O^6-carboxymethlyguanine (O^6-CMG). These damaged forms of guanine alter the capacity of the nucleobase for hydrogen bonding, thus leading to genomic instability. O^6-CMG is important in human health as it has been linked to colon carcinogenesis, and high dietary intake of processed and red meat. To address whether O^6-CMG formation increases cancer risk, we need to better understand the formation, occurrence, and endogenous DNA repair process of O^6-CMG. The goals of this thesis were to develop new tools to gain mechanistic insights on DNA damage response and to perform experiments to assess the biological significance of alkylation adducts. The main achievements of this thesis were the development of a quantification strategy based on nanopore sequencing, which was used to investigate the efficiency of alkyl damage repair in the nucleosome, and a quantification tool based on fluorescent chemical labeling of O^6-CMG.

Chapter 1 reviews the current knowledge on O^6-CMG. First, the relationship between O^6-CMG, human health, and its link to red meat consumption is discussed. Next, the cellular processing of DNA adducts by DNA repair and translesion synthesis are presented in the context of their contribution to mutagenesis. Finally, we review synthesis, detection and mapping strategies that have been central to O^6-CMG research in the recent years.

Chapter 2 presents the development and validation of a quantification method for O^6-methylguanine (O^6-MeG) and O^6-carboxymethylguanine (O^6-CMG) using nanopore sequencing. This novel method was used to investigate its repair by the human enzyme alkylguanine transferase (hAGT). Nanopore sequencing measures changes in electrical current as DNA passes through a pore to identify nucleotides. This current is directly dependent on the structural characteristics of the nucleobases passing through the nanopore. Therefore, chemical structure variations of damaged nucleotides can be detected and distinguished using this technology. Using raw current data and a predictive statistical model, we show that DNA damage can be accurately quantified in several sequence-specific contexts. This method provides insight into enzyme directionality, sequence context dependency, and the importance of damage positioning in the nucleosome for DNA. This newly developed method enables quantification of biologically relevant, mutagenic DNA damage while preserving DNA sequence information, thus leading to a better understanding of how adduct repair mechanisms affect mutation distribution.

Chapter 3 presents the development of a simple and selective fluorescence-based method for quantifying O^6-CMG. The method takes advantage of the carboxylic acid functional group of the adduct, which allows amine ligation in combination with click-chemistry to attach a fluorophore to O^6-CMG. This approach was developed and optimized for O^6-CMG in oligonucleotides and further tested in samples of double stranded DNA containing O^6-CMG as well as in chemically treated DNA and cells. Our method is sensitive, with a detection limit of 16 O^6-CMG /10^6 nucleotides. Detection is also specific for O^6-CMG, as we have found that the fluorescent probe does not label the epigenetic and therefore abundant 5caC modification. This

method allows, in principle, any click-reactive chemical to be coupled to a carboxymethyl adduct and provides the basis for the development of further biochemical tools to study the effects of O^6-CMG mutagenicity.

Chapter 4 summarizes the findings of this thesis. Future opportunities and challenges are critically discussed and recommendations for future directions are presented.

Appendix A reports a structure-activity relationship study to identify structural properties of O^6-CMG that affect the fidelity of DNA polymerases. An amine ligation approach, similar to that described in Chapter 2, was used to generate O^6-CMG analogs of varying size and charge. Primer extension assays were used to investigate the influence of structural variations of the O^6-CMG analogs on polymerase fidelity. While Pol κ was more mutagenic in the bypass of O^6-CMG analogs with larger size or a positive charge, Pol ι and Pol η were only slightly affected by steric effects or different charges.

Appendix B reports a study on the chemical mechanism of O^6-CMG formation in DNA by L-azaserine. The results indicated that hydrolysis of L-azaserine to L-serine is promoted by acid but is stable under neutral conditions. Evidence for the formation of an O^6-Ser-CMdG intermediate in azaserine-induced O^6-CMdG formation was provided for the first time. In addition, a nano-liquid chromatography high-resolution tandem mass-spectrometry method was developed to assess the dose- and time-dependent formation of adducts to further describe L-azaserine-induced O^6-CMdG and O^6-MedG formation. The results are important to consider for the design of future studies using L-azaserine to study NOC exposure.

Zusammenfassung

Alkylierung ist eine der häufigsten Formen der DNA-Schädigung durch *N*-Nitroso-Verbindungen. Diese genotoxischen Chemikalien sind in Tabakrauch und Verbrennungsprodukten enthalten und werden endogen im Magen und Dickdarm gebildet. Die Alkylierung an der O^6-Position von Guanin führt zur Bildung von O^6-Methylguanin (O^6-MeG) oder O^6-Carboxymethlyguanin (O^6-CMG). Diese geschädigten Formen von Guanin verändern die Fähigkeit der Nukleobase Wasserstoffbrücken zu bilden und führen so zu genomischer Instabilität. O^6-CMG ist für die menschliche Gesundheit von Bedeutung, da es mit der Entstehung von Dickdarmkrebs und einem hohen Konsum von verarbeitetem und rotem Fleisch in Verbindung wird. Um herauszufinden, ob die Bildung von O^6-CMG das Krebsrisiko erhöht, müssen wir die Bildung, das Vorkommen und den endogenen DNA-Reparaturprozess von O^6-CMG besser verstehen. Ziel dieser Arbeit war es, neue Werkzeuge zu entwickeln, um mechanistische Einblicke in die DNA-Schadensreaktion zu gewinnen und Experimente durchzuführen, um die biologischen Auswirkungen von Alkylierungsaddukten zu bewerten. Die wichtigsten Ergebnisse dieser Arbeit waren die Entwicklung einer Quantifizierungsstrategie auf der Grundlage der Nanoporen-Sequenzierung, die zur Untersuchung der Effizienz der Reparatur von Alkylschäden im Nukleosom verwendet wurde, und ein Quantifizierungsinstrument auf der Grundlage der fluoreszierenden chemischen Markierung von O^6-CMG.

Kapitel 1 gibt einen Überblick über den aktuellen Wissensstand zu O^6-CMG. Zunächst wird die Beziehung zwischen O^6-CMG und der menschlichen Gesundheit sowie der Zusammenhang mit dem Verzehr von rotem Fleisch diskutiert. Anschliessend wird die zelluläre Verarbeitung von DNA-Addukten durch DNA-Reparatur und Transläsionssynthese im Zusammenhang mit ihrem Beitrag zur Mutagenese dargestellt. Schliesslich werden Synthese-, Detektions- und Kartierungsstrategien vorgestellt, die in den letzten Jahren für die O^6-CMG-Forschung von zentraler Bedeutung waren.

In **Kapitel 2** wird die Entwicklung und Validierung einer Quantifizierungsmethode für O^6-Methylguanin (O^6-MeG) und O^6-Carboxymethylguanin (O^6-CMG) mittels Nanoporen-Sequenzierung vorgestellt. Diese neuartige Methode wurde zur Untersuchung ihrer Reparatur durch das menschliche Enzym Alkylguanin-Transferase (hAGT) eingesetzt. Bei der Nanoporen-Sequenzierung passieren DNA-Stränge einzelne Poren und bewirken dadurch eine messbare Änderung des angelegten elektrischen Stroms, wodurch Nukleotide identifiziert werden können. Dieser Strom hängt direkt von den strukturellen Eigenschaften der Nukleobasen ab, welche die Nanopore passieren. Daher können mit dieser Technologie Variationen in der chemischen Struktur geschädigter Nukleotide erkannt und unterschieden werden. Anhand von Rohstromdaten und einem prädiktiven statistischen Modell zeigen wir, dass DNA-Schäden in verschiedenen sequenzspezifischen Kontexten genau quantifiziert werden können. Diese Methode ermöglicht Einblicke in die Richtungsabhängigkeit der Reparaturenzyme, die Abhängigkeit vom Sequenzkontext und die Bedeutung der Positionierung der Schäden im Nukleosom für die DNA-Reparatur. Die neu entwickelte Methode ermöglicht die Quantifizierung biologisch relevanter, mutagener DNA-Schäden unter Beibehaltung der DNA-Sequenzinformation und führt so zu einem besseren Verständnis der Auswirkungen von Addukt-Reparaturmechanismen auf die Mutationsverteilung.

In **Kapitel 3** wird die Entwicklung einer einfachen und selektiven fluoreszenzbasierten Methode zur Quantifizierung von O^6-CMG vorgestellt. Die Methode nutzt die funktionelle Carbonsäuregruppe des Addukts, die eine Amin-Ligation in Kombination mit Click-Chemie ermöglicht, um ein Fluorophor an O^6-CMG zu binden. Dieser Ansatz wurde für O^6-CMG in Oligonukleotiden entwickelt und optimiert und in doppelsträngigen O^6-CMG enthaltenden DNA-Proben, sowie in chemisch behandelter DNA und Zellen getestet. Unsere Methode ist sensitiv, mit einer Nachweisgrenze von 16 O^6-CMG /10^6 Nukleotiden. Die Detektion ist auch spezifisch für O^6-CMG. Wir zeigen, dass die entwickelte Methode die epigenetische und daher häufig vorkommende 5caC-Modifikation nicht markiert. Die Methode erlaubt prinzipiell die Kopplung jeder klickreaktiven Chemikalie an ein Carboxymethyl-Addukt und bildet die Grundlage für die Entwicklung weiterer biochemischer Werkzeuge zur Untersuchung der Auswirkungen der O^6-CMG-Mutagenität.

Kapitel 4 fasst die Ergebnisse dieser Arbeit zusammen. Zukünftige Möglichkeiten und Herausforderungen werden kritisch diskutiert und Empfehlungen für künftige Richtungen werden vorgestellt.

Anhang A berichtet über eine Struktur-Aktivitäts-Beziehungsstudie, die durchgeführt wurde, um strukturelle Eigenschaften von O^6-CMG zu identifizieren, die die Genauigkeit von DNA-Polymerasen beeinflussen. Ein Amin-Ligationsansatz, ähnlich jenem in Kapitel 2 beschriebenen, wurde verwendet, um O^6-CMG-Analoga mit unterschiedlicher Grösse und Ladung zu erzeugen. Anhand von Primer-Extension-Experiment wurde der Einfluss struktureller Variationen der O^6-CMG-Analoga auf die Polymerase-Genauigkeit untersucht. Während Pol κ mutagener gegenüber O^6-CMG-Analoga mit grösserer Masse oder positiver Ladung war, wurden Pol ι und Pol η nur geringfügig durch sterische Effekte oder Ladungensunterschiede beeinflusst.

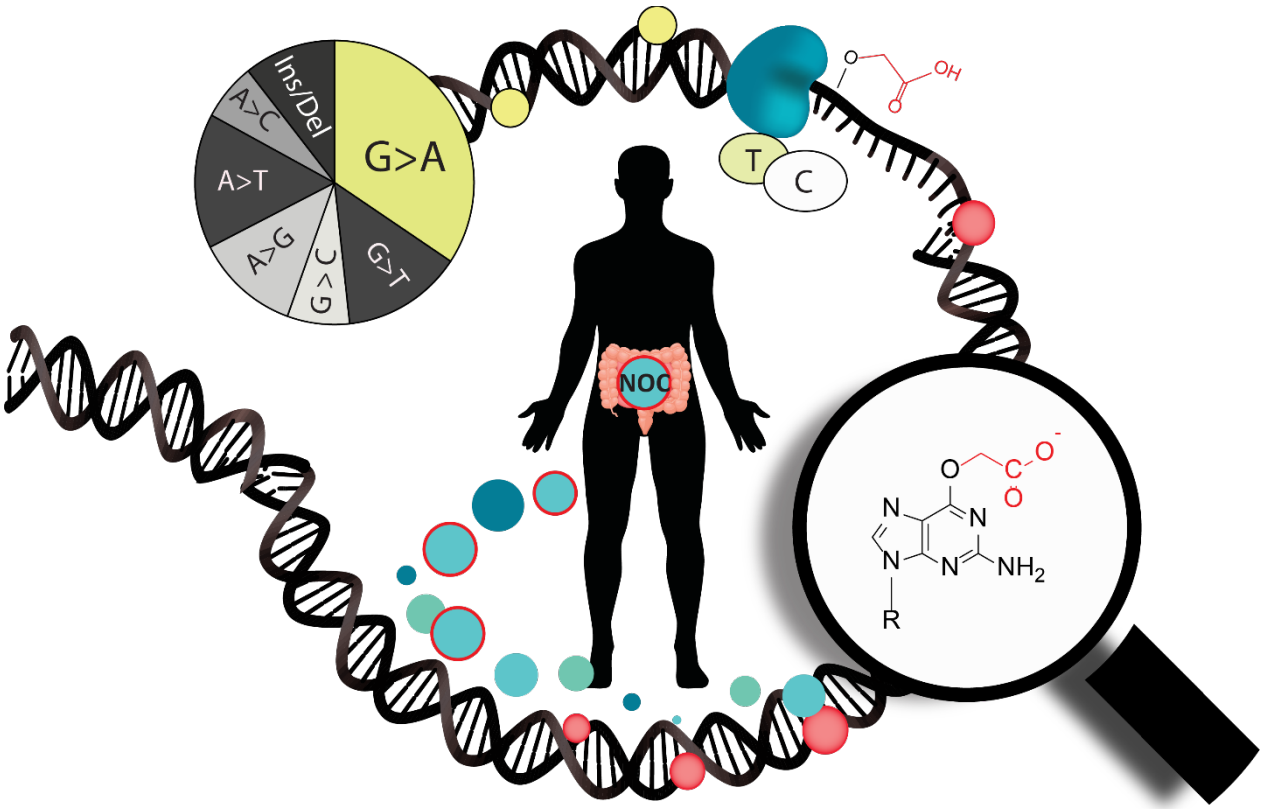
Anhang B enthält eine Studie über den chemischen Mechanismus der Bildung von O^6-CMG in der DNA durch L-Azaserin. Die Ergebnisse zeigen, dass die Hydrolyse von L-Azaserin zu L-Serin durch Säure gefördert wird, aber unter neutralen Bedingungen stabil ist. Erstmals konnte die Bildung eines O^6-Ser-CMdG-Zwischenproduktes bei der Azaserin-induzierten O^6-CMdG-Bildung nachgewiesen werden. Darüber hinaus wurde eine hochauflösende Nano-Flüssigchromatographie-Tandem-Massenspektrometrie-Methode entwickelt, um die dosis- und zeitabhängige Bildung von Addukten zu bewerten und die L-Azaserin-induzierte O^6-CMdG- und O^6-MedG-Bildung weiter zu charakterisieren. Die Ergebnisse sind wichtig für die Planung zukünftiger Studien, in denen L-Azaserin zur Untersuchung der NOC-Exposition eingesetzt wird.

Abbreviations

O^6-CMG	O^6-Carboxymethylguanine
O^6-MeG	O^6-Methylguanine
O^6-alkylG	O^6-Alkylguanine
5meC	5-Methylcytosine
5caC	5-Carboxylcytosine
8-oxoG	8-Oxoguanine
CRC	Colorectal cancer
MGMT	Methylguanine methyltransferase
NER	Nucleotide Excision Repair
ATL	Alkyltransferase-like enzymes
AGT	alkylguanine alkyltransferase
hAGT	Human AGT
MGMT	O^6-methylguanine-DNA methyltransferase
TLS	Translesion synthesis
Pol	Polymerase
NCP	Nucleosome core particle
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ssDNA	Single strand DNA
dsDNA	Double strand DNA
ctDNA	Calf-thymus DNA
gDNA	Genomic DNA
NTP	Nucleoside triphosphate
G	Guanine
A	Adenine
C	Cytosine
T	Thymine
NOCs	*N*-nitroso compounds
KDA	Potassium Diazoacetate
MNU	*N*-methyl-*N*-nitrosourea
AF-594	Alexafluor 594
ESI	Electrospray ionization
Em	Emission
Ex	Excitation
HRMS	High-resolution Mass-Spectrometry
HPLC	High-performance liquid-chromatography
RP-HPLC	Reverse-phase high-performance liquid chromatography
MS	Mass-Spectrometry

NMR	Nuclear magnetic resonance
LC-MS	Liquid-chromatography mass-spectrometry
nt	Nucleotide
bp	basepair
Pol	Polymerase
PCR	Polymerase chain reaction
ONT	Oxford nanopore sequencing
NNM	Neural network model
LR	Logistic regression
FP	False-positive
FN	False-negative
TP	True-positive
TN	True-negative
DMSO	Dimethylsulfoxide
DMF	Dimethylformamide
DMTMM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride
RT	Room temperature
Equiv	Equivalent
LOD	Limit of detection

Chapter 1 – Introduction: A Chemical Link between Meat Consumption and Colorectal Cancer Development?



E. S. Sandell wrote sections on meat and O^6-CMG occurrence and synthetic strategies, edited the document and addressed reviewers' comments. C. M. N. Aloisi wrote sections on cellular response to O^6-CMG, detection and mapping strategies of O^6-CMG, edited the document, and addressed reviewers' comments. S. J. Sturla wrote the introduction, edited the document, and addressed to reviewers' comments.

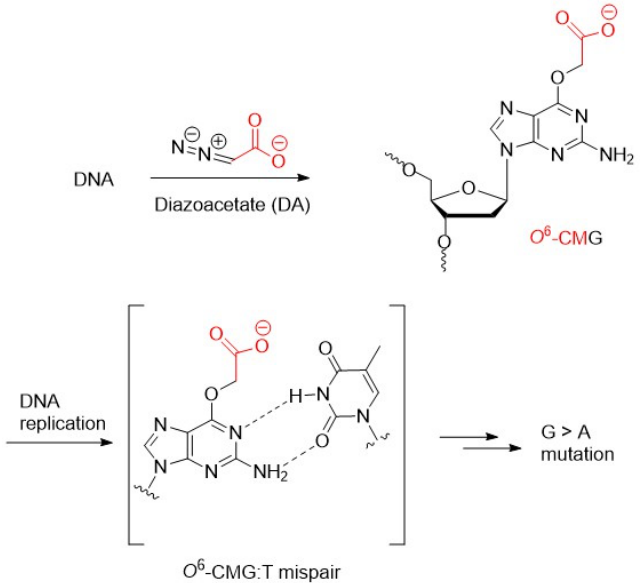
1.1 Abstract

O^6-carboxymethylguanine (O^6-CMG) is a mutagenic DNA adduct that forms at increased levels when people eat meat. It has been studied as a potential initiating event in colorectal carcinogenesis. It can arise from alkylation of guanine in DNA by electrophilic degradation products of *N*-nitroso compounds. There is significant data regarding biochemical and cellular process, including DNA repair and translesion DNA synthesis that control O^6-CMG accumulation, persistence, and mutagenicity. Mutation spectra arising from the adduct closely resemble common mutations in colorectal cancer; however, gaps remain in understanding the biochemical processes that regulate how and where the damage persists in the genome. Addressing such questions relies on advances in chemistry such as synthesis approaches and bioanalytical methods. Results of research in this area help advance our understanding of the toxicological relevance of O^6-CMG-modified DNA. Further attention should focus on understanding how a combination of genetic and environmental factors control its biological persistence and how this information can be used as a basis of biomonitoring and prevention efforts to help mitigate colon cancer risk.

1.2 Introduction

Colorectal cancer has the third highest incidence number worldwide.[1] It is a heterogeneous disease with a complex etiology, and while colorectal cancers can arise from inherited genetic defects, such as Lynch syndrome and polyposis syndromes, these account for <5% of cases.[2] Thus, for a large majority of cases, environmental and lifestyle factors drive colorectal cancer (CRC) risk.[1] Among these, the consumption of processed meat causes CRC, and consuming red meat is probably carcinogenic.[3] For individuals who consume processed or red meat, the excess risk arises from exposure to carcinogenic chemicals in the food.

In the search to identify chemicals in red and processed meat that initiate colon carcinogenesis, colon tissues were exposed to meat digests, and a DNA alkylation adduct with a negatively charged carboxymethyl group on guanine (G) in DNA (O^6-CMG) was discovered.[4] Furthermore, the formation of O^6-CMG and other O^6-alkylguanine (O^6-alkylG) has been established to form from reactions with *N*-nitroso compounds (NOCs). NOC carcinogenesis has been investigated for more than 50 years.[4-13] NOCs are commonly found in cigarette smoke and can also form from amines upon cooking or digestion of nitrite-containing meat.[14] Alkylation of the O^6-position of guanine to form DNA adducts results from nucleophilic attack of alkyl-diazonium electrophiles derived from metabolic activation or decomposition of NOCs.



Scheme 1. O^6-CMG-DNA Formation from Diazoacetate-Forming Chemicals

Damage to genomic DNA, such as the formation of DNA adducts, can be toxic to cells, and its detrimental phenotype has been investigated since the late 1940s.[15,16]DNA damage was

originally detected as a disruption of chromosome integrity.[17,18] A decade later, chemical modifications to the nucleobases were suggested as a basis for genomic instability.[19-22] Despite being postulated and associated with processes of toxicity, mutagenicity, and carcinogenicity,[23-32] DNA adducts were structurally characterized several years later, thanks to the advancement of bioanalytical chemistry and detection techniques.[33] Many environmental sources of DNA adducts have been identified, and corresponding DNA adducts characterized;[10,34-37] however, understanding cancer-driving mechanisms at a biochemical and molecular level remains critical to validate and mitigate the contributions of chemical exposures to human cancer incidence, such as in CRC. A substantive body of evidence bridging *in vitro*, *in vivo*, and population-based studies of incidence, mechanisms, and outcomes exists for a few DNA damage-forming carcinogens, such as aristolochic acid and upper urinary tract cancer[37,38] and aflatoxin and liver cancer.[36,39] However, the remaining data gaps for specific exposure-adduct structures as well as emerging untargeted profiles of DNA base modification signatures in exposure scenarios or disease states are challenging and active areas of research.[40-43] The focus of this article is on the chemistry and biology of O^6-carboxymethylguanine (O^6-CMG) in relation to CRC.

The misincorporation of thymine or adenine opposite O^6-CMG during DNA replication, due to the diminished hydrogen-bonding capacity of O^6-CMG compared to G, induces point mutations, a key event in carcinogenesis (Scheme 1).[44,45] Biochemical, molecular, and structural studies have helped characterize the O^6-CMG mutagenic phenotype.[46-49] However, reports concerning the repair of O^6-CMG are scarce or draw contradictory conclusions; therefore, a better understanding of O^6-CMG clearance is still needed to help understand adverse outcomes. There is also information lacking concerning O^6-CMG distribution in the genome, such as whether O^6-CMG accumulates in cancer-relevant genes. Emerging research concerning detection strategies that retain information concerning O^6-CMG damage location and persistence is expected to further clarify relationships between O^6-CMG and cancer.[50,51]

Herein, we review evidence linking meat consumption with O^6-CMG occurrence, present mechanistic data elucidating the role of DNA synthesis and repair in O^6-CMG-induced mutagenicity, and present emerging strategies to better address contributions to carcinogenesis. Focusing on chemical advances in this area, we describe synthetic and analytical strategies that have been instrumental to the elucidation of O^6-CMG-induced phenotypes. Finally, we anticipate how new detection strategies that provide information on the genomic location of O^6-CMG can help delineate O^6-CMG formation, repair, persistence, and evolution of cancer-relevant mutations in human cells.

1.3 Red meat consumption and O^6-CMG occurrence

It was suggested in 1975 that a diet high in meat consumption may increase CRC risk.[52] This seminal study related dietary habits across countries with national cancer incidence and catalyzed efforts to uncover the potential basis of this observation. Thus, several controlled intervention and cohort studies followed (Figure 1A). [3,13,53-57] For example, higher NOC levels were measured in stool samples from people with a high-meat diet compared to individuals with a low-meat or vegetarian diet,[53,54,58] particularly if red or processed meat was consumed.[56,59] Heme iron was found to stimulate endogenous NOC production.[55,60] Additionally, NOC formation upon consumption of meat was ascribed largely to the use of nitrate and nitrite as food preservatives and the reaction of nitrites with amines in the meat or in the human body (Figure 1B). These processes can be promoted by low pH[14,61-65] or catalyzed by gut microbial enzymes at

neutral pH.[66] Nitrite-dependent nitrosation also occurs in mammalian cells owing to the endogenous production of nitric oxide by nitric oxide synthase.[44,67,68] Thus, the mutagenic effects of NOC exposure, initiated by the formation of NOC-induced DNA adducts, are viewed as a likely molecular initiating event in colorectal carcinogenesis.

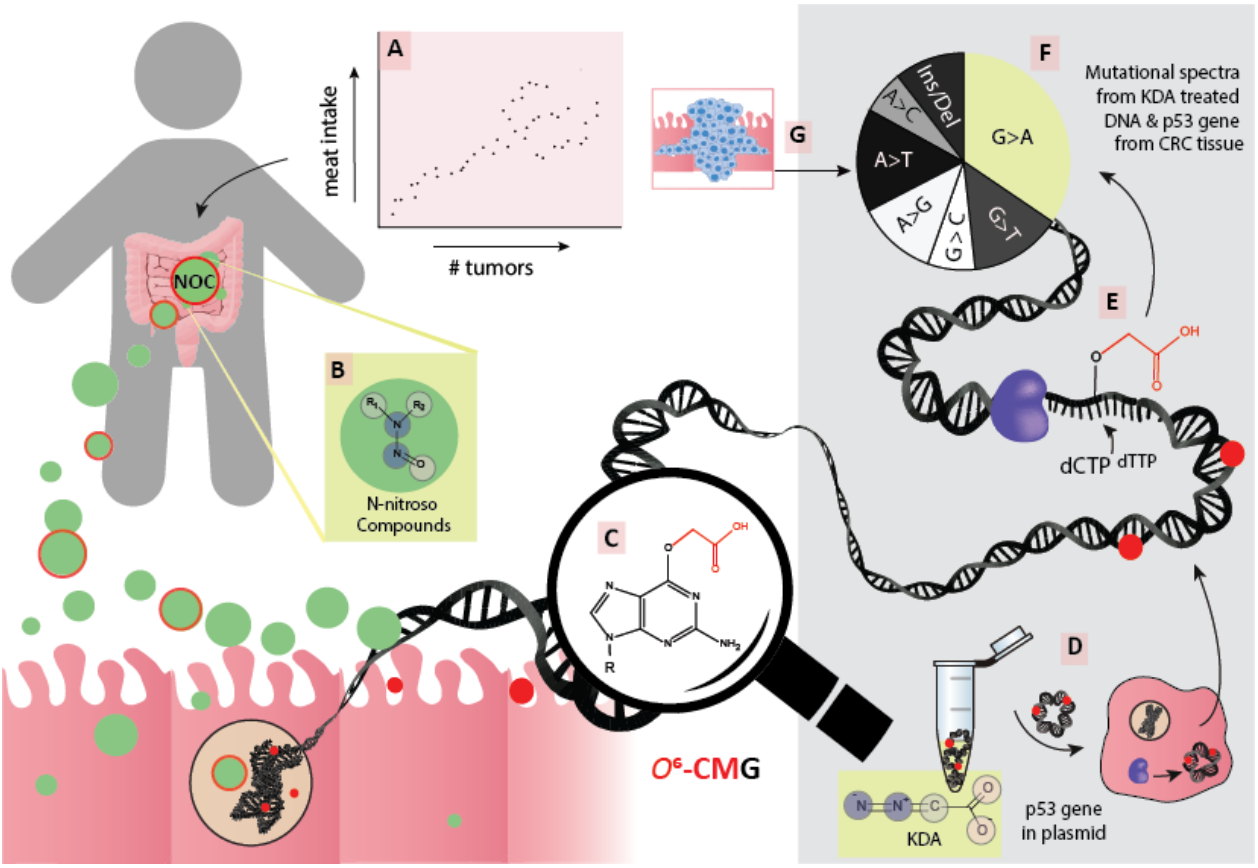


Figure 1. Graphical summary of this review. (A) Correlation between meat intake and the development of tumors in humans. (B) NOCs are formed by reaction of amines with NO-forming molecules, including nitrogenous dietary compounds found in meat. (C) NOCs can form O^6-CMG in DNA. (D, F) O^6-CMG induces a mutational spectrum in p53 that resembles (G) those found in the same gene in cancer tissues. (E) The mutations induced upon replication of O^6-CMG-modified DNA are probably due to the misincorporation of nuclear bases opposite O^6-CMG.

NOCs such as *N*-(*N*-acetyl-L-prolyl)-*N*-nitrosoglycine, *N*-nitrosoglycocholic acid (NOGC), and L-azaserine, and the reactive intermediates potassium diazoacetate (KDA) all have an *N*-carboxymethyl-*N*-nitroso moiety, and thus the potential to react with nucleophilic positions on DNA bases, resulting in the addition of a carboxymethyl group. Azaserine, originally developed as an antibiotic drug, recently has been used as a convenient model compound to induce carboxymethylation of DNA.[8,69-72] Various nucleophilic positions in DNA react with NOC metabolites or decomposition products. For example, reaction of calf thymus DNA with KDA gave rise to several carboxymethylated DNA adducts including N^4-carboxymethyl-cytosine (N^4-CMC), N^3-carboxymethylthymine (N^3-CMT), O^4-carboxymethylthymine (O^4-CMT), and N^6-carboxymethyl-2-deoxyadenine (N^6-CMA).[49,73,74] Only N^6-CMA, O^6-CMG, and N^2-CMG have been detected in cells exposure to azaserine or glyoxal,[49,73] and only O^6-CMG was associated with meat consumption in humans.[74]

To determine if O^6-CMG is induced by dietary NOCs, a study was conducted involving healthy human volunteers who ate a high red meat or vegetarian diet for 15 days.[10] Consistent with previous findings,[53,75,76] there was a significant increase in NOC levels in feces of people eating a high red meat diet. Moreover, the level of O^6-CMG in colonic exfoliated cells was significantly increased. This compelling observation suggested that O^6-CMG arises from meat-derived NOCs (Figure 1C).[10]

Increased O^6-CMG formation in relation to meat consumption and digestion has been confirmed with a number of studies.[9,12,13] For example, digestion of white, red, and well-cooked meat was simulated in an *in vitro* gastrointestinal digestion model[77] for the stomach and small and large bowels, using fecal-inocula from individual human colonic digests.[9,12,13] Exposure of resulting digestive fluids to cellular genomic DNA and analysis with a mass spectrometry (MS)-based DNA adductomics workflow revealed that O^6-CMG was consistently increased upon exposure to red meat digest and did so in a dose-dependent manner. It was interesting also that O^6-CMG formation appeared to vary considerably among samples with different microbiome composition, supporting the role of microbial enzymes in endogenous nitrosation processes leading to its formation.[37,38] Several other adducts were detected in the *in vitro* digestion studies,[9,12,13] including dimethylthymine, hydroxybutylcytosine, and hydroxyethylthymine, but only for O^6-MeG and O^6-CMG, there was also a corroborating association between adduct levels and actual red meat consumption.[10,13]

1.4 O^6-CMG mutagenicity and cancer mutational spectra

The combination of (1) epidemiological data, suggesting that diets high in meat consumption contribute to increased CRC risk, and (2) exposure data, revealing that increased levels of O^6-CMG are formed in intestinal DNA after consumption or digestion of meat, raised the question of whether and how O^6-CMG causes mutations and cancer. In carcinogenesis bioassays[78-80] and bacterial mutagenicity tests, increased tumor numbers and mutation rates were associated with O^6-CMG formation.[70,81] These studies were performed with azaserine[69,82] as a chemical probe for exposures that give rise to O^6-CMG formation.[70-72] Azaserine induces the formation of both O^6-CMG and O^6-MeG, but O^6-CMG levels are at least 10 times higher than O^6-MeG.[72] Mutation spectra from yeast, in this case exposed to KDA, were enriched in G > A mutations, accounting for about 40% of the observed mutations in p53 plasmids (Figure 1D), whereas those arising from *N*-methyl-*N*-nitrosourea (MNU), a methylating agent that forms O^6-MeG,[46] produced a significantly different mutation spectrum with almost exclusively G > A mutations. While it is difficult to dissect the contribution of O^6-CMG vs O^6-MeG to the spectrum on the basis of these data, minor aspects of the KDA or azaserine spectra seem further consistent with replication past O^6-CMG, such as significant G > T mutations that are almost completely absent in the MNU spectrum. Furthermore, these minor mutations are also consistent with observations that dAMP is occasionally inserted opposite O^6-CMG.[47]

Remarkably, the type and frequency of mutations induced by KDA in p53 plasmids were extremely similar to those found in the p53 gene of CRC and stomach cancer tissues from patients (Figure 1G).[46] To identify important cellular factors that contribute to the O^6-CMG mutagenic phenotype, bacterial cells with defined gene deficiencies were exposed to azaserine, and the mutation frequency was evaluated. The mutagenicity of azaserine was enhanced by cellular proficiency in TLS[70] and reduced by DNA repair activity, particularly factors for the nucleotide excision repair (NER) pathway[70] and alkyltransferase repair enzymes.[81] The reported high frequency of G > A point mutations is consistent with the misincorporation

of T opposite O^6-CMG as a basis for azaserine and KDA mutagenicity. These observations raise many questions concerning the structural and enzymological basis of error-prone O^6-CMG translesion DNA synthesis.

1.5 Cellular Response to O^6-CMG: DNA translesion synthesis

To characterize how O^6-CMG's mispairing potential may lead to mutations during DNA replication, DNA synthesis over O^6-CMG was evaluated using site-specifically modified oligonucleotides and purified DNA Pols (Figure 2).[47-49,83,84] In a first biochemical study of replication of DNA containing O^6-CMG, four human TLS Pols η, ι, κ, and ζ, which are error prone TLS Pols that replicate damaged DNA templates, along with replicative Pol δ were tested for their capacity to bypass O^6-CMG in a mutational DNA hotspot region of the colon cancer-relevant KRAS gene.[47] Replicative Pol δ and TLS Pol η incorporated dCMP or dAMP opposite O^6-CMG, whereas Pol κ incorporated dCMP only, and Pol ι incorporated dTMP only. Earlier structural evaluation of O^6-CMG paired with different bases in DNA suggests that the insertion of T may arise from a Watson–Crick-type pair, similar to the canonical A:T pair.[85] In the case of binding with C in a duplex, a reverse wobble pair formed. All polymerases except Pol ι efficiently extended past O^6-CMG. Despite the observation that error-free bypass of O^6-CMG is a kinetically prevalent scenario, the researchers identified TLS Pols, namely Pol δ, η, and ι, whose function is consistent with O^6-CMG inducing a combination of G > A and G > T mutations, which are frequently associated with signatures in CRC.[47]

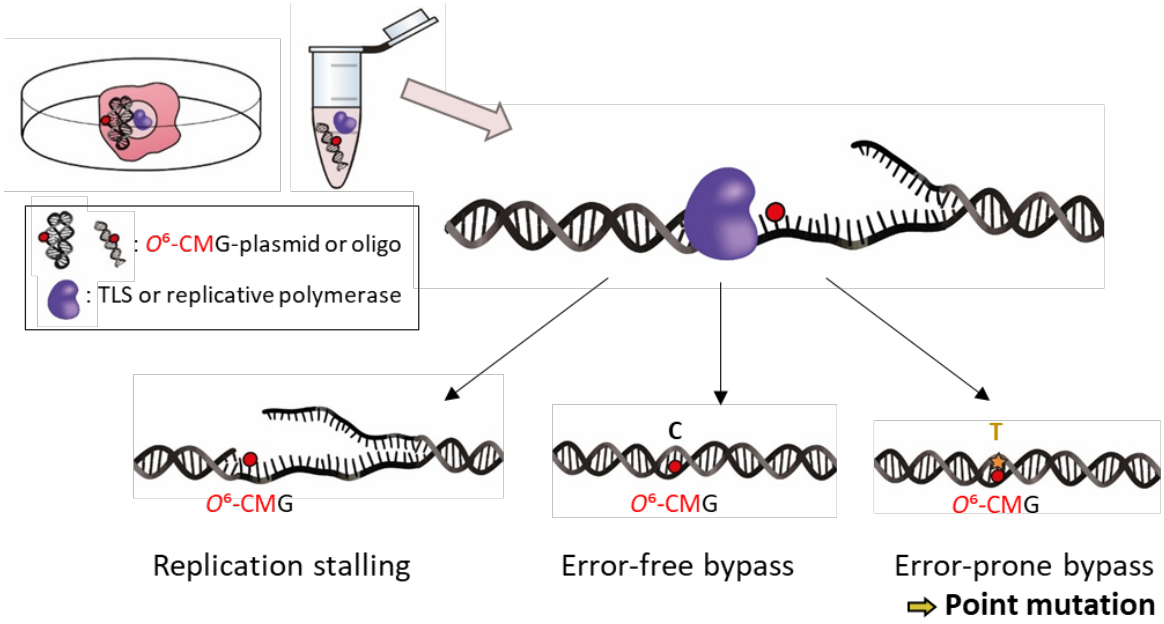


Figure 2. Outcomes of O^6-CMG DNA replication. The replication of O^6-CMG has been investigated both in cells and on artificial DNA with isolated enzymes. Outcomes vary according to the replicating machinery and are: stalling of the replication (bypass efficiency was of 10% in bacteria and of 38% in human cells), error-free bypass (incorporation of dCTP), and error-prone bypass (incorporation of dTTP only in cell studies, of dTTP and dATP by isolated enzymes).

O^6-CMG is a unique DNA adduct in placing a negatively charged carboxylate in the polymerase active site, raising the question of whether steric or electronic interactions drive its miscoding potential. In order to investigate this question, replication over O^6-CMG and structural analogs by Pol η, κ, and ι was tested.[48]

The structural analogs of O^6-CMG were all amides or esters with side chains of varying length and size (H, methyl, allyl, isopentyl), and one that was positive, rather than negatively, charged (ethylenamine) was also tested. The structure of the DNA adduct had a strong effect on dNTP selection by Pol κ, but not by Pol η and ι; however, Pol η lost fidelity with increased size of the adduct. Structural modeling in the active site of Pol κ suggested that hydrogen bonding between the templating base and the incoming nucleotide is an important driving force for the selection of the preferred NTP, and in the case of O^6-CMG bypass by Pol κ, stabilization by ionic interactions with the carboxylate might promote error-free bypass.[48]

When tested in a more complex cellular environment, O^6-CMG replication was moderately blocked in human cells (38% bypass efficiency)[49] and strongly inhibited in Escherichia coli (10% bypass efficiency).[83] Pol κ and Pol ζ seemed the most efficient at overcoming O^6-CMG replication blockages, as their genetic depletion led to a decreased bypass efficiency (30% and 26%, respectively, compared to 38% bypass measured in the presence of a fully proficient replication machinery). Meanwhile, Pol η seemed scarcely involved since the bypass efficiency in its absence was similar to wildtype.[49] Depletion of Pol ι resulted in a decreased frequency of G > A mutations but not in decreased bypass efficiency, suggesting that Pol ι outcompetes other higher fidelity enzymes and functions as a prominent mis-inserter of T opposite O^6-CMG. While the relative activity of individual TLS Pols was comparable when isolated enzymes were tested[47] and when replication was measured in cells,[49,83] G > A mutations were exclusively induced in cells.[49,83] A similar mutational pattern was found when other aliphatic O^6-alkyl Gs were assayed.[86] These data provide a working model for how mutations are induced inhuman cells when O^6-CMG is formed and persists, but does not account for repair processes that may effectively mitigate O^6-CMG mutagenicity.

1.6 Cellular Response to O^6-CMG: DNA repair

In spite of early observations that bacterial cells defective in DNA repair were more susceptible to azaserine-induced mutagenicity[70,81], current knowledge of O^6-CMG repair remains incomplete, especially concerning the role of specific human repair pathways. Repair of O^6-CMG was first investigated by assessing azaserine mutagenicity in bacteria and yeast with defined deficiencies in repair factors.[70,81] Alkyltransferase repair enzymes[81] and NER assisted by alkyltransferase-like (ATL) enzymes[70,87,88] protected bacterial and yeast cells from mutations resulting from exposure to azaserine. ATL proteins are homologous to O^6-methylguanine-DNA methyltransferase (MGMT), but do not have the transferase domain; therefore, they bind DNA at modified bases but do not catalyze their removal.[88-91] Binding and conformational studies supported that ATL proteins serve as signaling factors of O^6-alkylG, including O^6-CMG, to the NER pathway.[70,87,88] Although ATLs have not been found in higher eukaryotes yet, these results raise the question of whether alkyltransferases such as MGMT might have a similar role in facilitating and signaling for the repair of O^6-CMG in higher organisms. However, this speculation has never been confirmed, and the role of MGMT in relation to O^6-CMG persistence in human cells remains equivocal.

To our knowledge, in three studies researchers interrogated which human DNA repair proteins recognize and remove O^6-CMG (Figure 3).[71,92,93] In the first study reported in 1999, human cell lines with defined defects in DNA repair pathways were exposed to azaserine, and cell survival was measured as an indication of sensitivity to azaserine and O^6-CMG. Cells deficient in NER factors (e.g., NER damage recognition factor XPC and DNA helicases XPB and XPD) were the most sensitive to azaserine, whereas deficiencies in mismatch repair and MGMT did not influence cell survival.[71] The lack of MGMT activity toward O^6-CMG

was in agreement with previous work involving mass spectrometric measurements of alkyl transfer in which the carboxymethyl group was not transferred from O^6-CMG-modified oligonucleotides to MGMT when incubated with purified bacterial alkyltransferases or with human cellular extracts.[93] Finally, when a similar study was subsequently performed with purified MGMT, the carboxymethyl group was transferred from O^6-CMG to the purified human MGMT[92], supporting for the first time a direct role of MGMT in removing O^6-CMG. In recent work, MGMT and NER have been suggested to repair also other O^6-alkylGs, with NER repair machinery targeting bulkier alkylGs and MGMT transferring smaller alkyl groups, for example, methyl and ethyl; however, O^6-CMG was not addressed.[86]

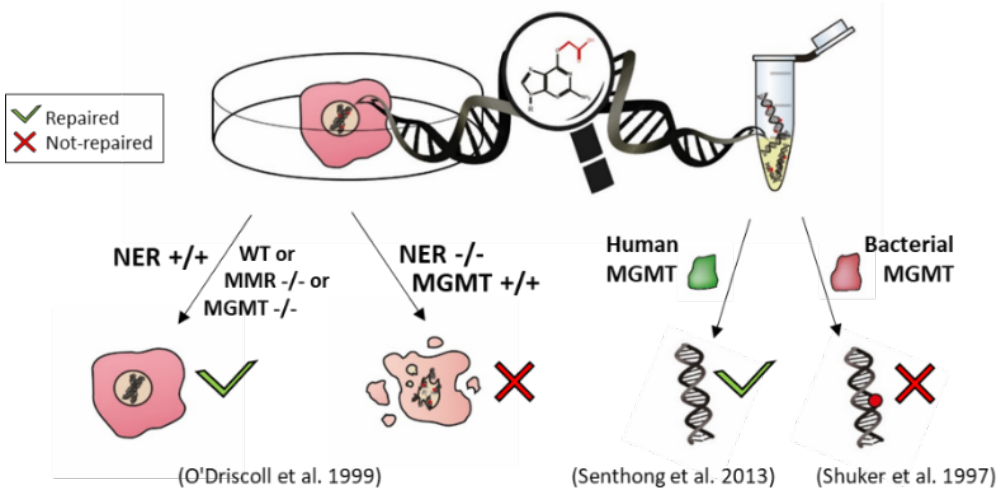
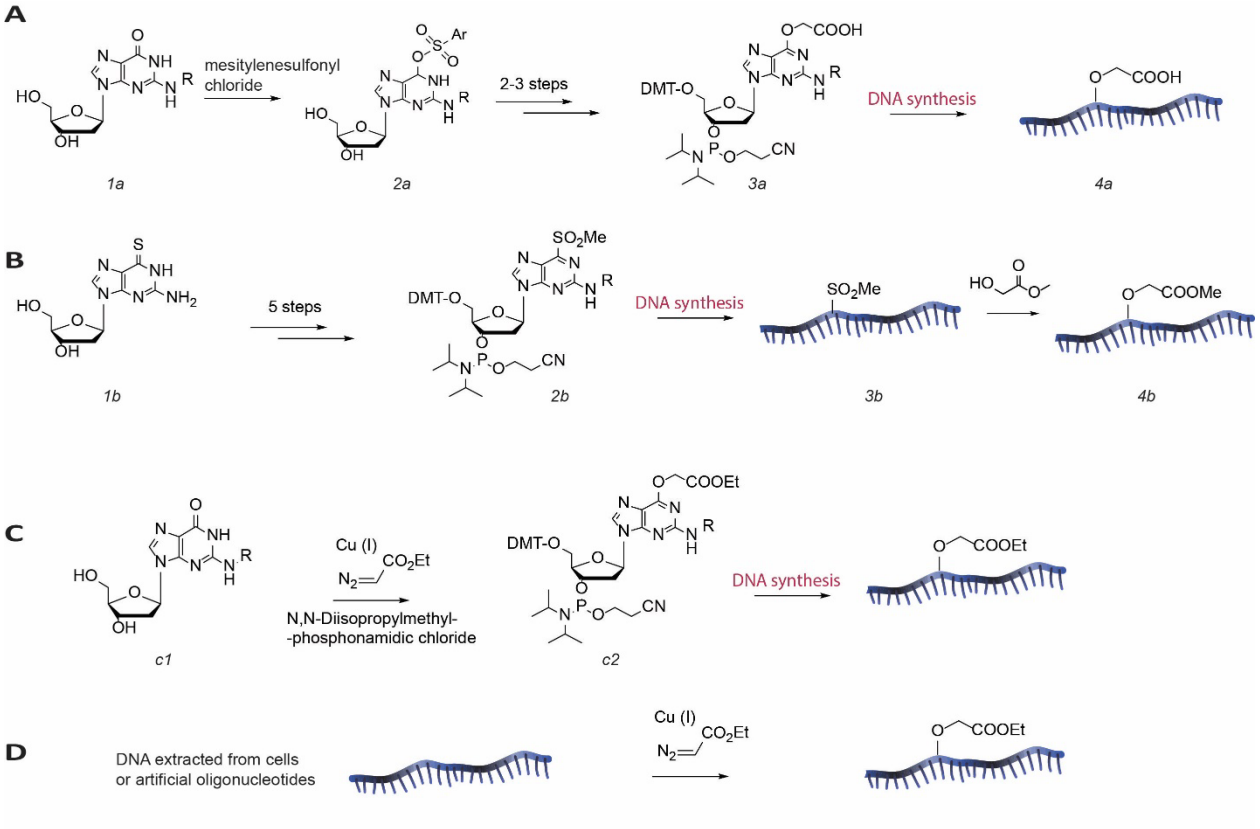


Figure 3. Experimental evidence for how O^6-CMG DNA is repaired. The repair of O^6-CMG has been investigated both with cells and synthetic DNA with isolated enzymes. Cells deficient in NER but not MGMT are sensitive to O^6 -CMG. Purified human but not bacterial MGMT repaired O^6-CMG (WT, wild-type, that is, repair proficient; + /+ proficient; − /− deficient; NER, nucleotide excision repair; MMR, mismatch repair; MGMT, methylguanine methyltransferase).

1.7 Synthetic strategies for O^6-CMG-modified DNA to enable molecular studies

Synthesizing DNA strands modified with O^6-CMG have enabled *in vitro* biochemical studies and have contributed to the identification of key factors that promote or mitigate O^6-CMG mutagenicity. DNA strands containing O^6-CMG can be either generated site-specifically by solid-phase synthesis of oligonucleotides with a modified phosphoramidite (Scheme2A−C)[94-99] or in an indiscriminate fashion by modification of any G in DNA (Scheme 2D).[99] An early strategy for the synthesis of O^6-CMG phosphoramidite was based on the formation of an O^6-arylsulfonate intermediate (**2a**) from guanosine (**1a**). The O^6-arylsulfonate group was replaced by a trialkylamine to form a reactive trialkylammonium, which was readily displaced by alcohols in the presence of the base catalyst 1,8-diazabicyclo[5.4.0]undec-7-ene. By reacting the trialkylammonium intermediate with methyl glycolate, the corresponding ester derivative of O^6-CMG was obtained with a final yield ranging from 40% to 50%.[94-96] The ester derivative of O^6-CMG was finally converted to the O^6-CMG phosphoramidite **3a** and used for solid-phase synthesis of DNA to obtain O^6-CMG-modified DNA (**4a**,Scheme 2A).



Scheme 2. Synthetic Strategies for ***O*^{6}-CMG-Containing DNA. (A) Via synthesis of a modified methylsulfoneguanine phosphoramidite, (B) via synthesis of methylsulfone intermediate, and post-synthetic modification (C) via a copper-directed alkylation of isolated guanosine or (D) of guanines in DNA.

An alternative strategy for synthesis of DNA with *O*^{6}-CMG entails the synthesis of a modified methylsulfoneguanine phosphoramidite (**2b**), solid-phase DNA synthesis and a postsynthetic DNA modification step, to yield site-specifically modified *O*^{6}-CMG DNA oligonucleotides (Scheme 2B). The 2'-deoxy-6-thioguanosine analog (**1b**) was converted to amethylsulfone derivate. The corresponding phosphoramidite (**2b**) was obtained in around 30% yield and used for solid-phase DNA synthesis. The resulting DNA (**3b**) contained a 2-amino-6-methylsulfonyl group, which then was converted to an *O*^{6}-methylcarboxymethylester via nucleophilic substitution at the C-6 position by methyl glycolate. After hydrolysis of the methylester (**4b**), the *O*^{6}-CMG-modified nucleotide was obtained (Scheme 2B).[98] While these strategies were critical for advancement of this area, developments in metal-catalyzed reactions of nucleic acids have greatly simplified access to substrates for further research.

Recently, a selective metal-catalyzed nucleotide functionalization strategy that provides a reliable and fast means for the synthesis of *O*^{6}-CMG has been reported. It entails a copper-catalyzed alkylation reaction to selectively modify dG analogues,[99] followed by standard phosphoramidite conversion.[98] An *O*^{6}-CMG precursor, COOEt-phosphoramidite (**2c**), was synthesized from 2'-deoxy-*N*^{2}-isobutyrylguanosine (**1c**) via a copper-catalyzed alkylation with ethyl diazoacetate. The reaction entails chemoselective introduction of a methylcarboxyethyl ester group at the *O*^{6}-position of dG. The selectivity is thought to arise from pre-association of copper with the 7-position N of guanine.[99] The corresponding *O*^{6}-methylcarboxyethylester-phosphoramidite (**2c**) was prepared and incorporated into a DNA strand by standard solid-phase DNA

synthesis. The desired O^6-CMG was obtained following hydrolysis of the ethyl ester.[98] A benefit of this copper-catalyzed approach is that its high selectivity allows it to be performed directly on DNA, including oligonucleotides, plasmids, ctDNA, and genomic DNA (Scheme 2C), therefore facilitating molecular biology and biochemistry studies of O^6-CMG.

1.8 Detection of O^6-CMG

The detection of O^6-CMG in biological samples from *in vivo*[10,100,101] and *in vitro*[4] studies has been pivotal to link O^6-CMG to the consumption of meat and to address the hypothesis that it is a molecular initiating event in CRC. Established O^6-CMG detection strategies (Figure 4) include immunoblotting,[100] affinity chromatography,[8,94]. ^{32}P postlabeling,[101] and MS.[4,72,102,103] The first two methods are based on the selective recognition of the adduct by an antibody obtained via immunization of animals.[94] The antibody was either used in combination with a fluorescent secondary antibody for the blotting of purified DNA samples (Figure 4A)[100] or placed in a matrix for chromatographic separation of DNA (Figure 4B) followed by reversed-phase high-performance liquid chromatography with fluorescence-based detection of O^6-CMG.[8,94] Antibody strategies enabled the detection of O^6-CMG in ctDNA exposed to azaserine and KDA with a limit of quantification of 0.1 pmol and in human samples with a limit of detection of 0.1 fmol,[100] allowing the detection of 15 adducts/10^8 nucleotides, measured in 1µg of DNA extracted from human blood of volunteers who consumed a high-meat diets for 43 days.

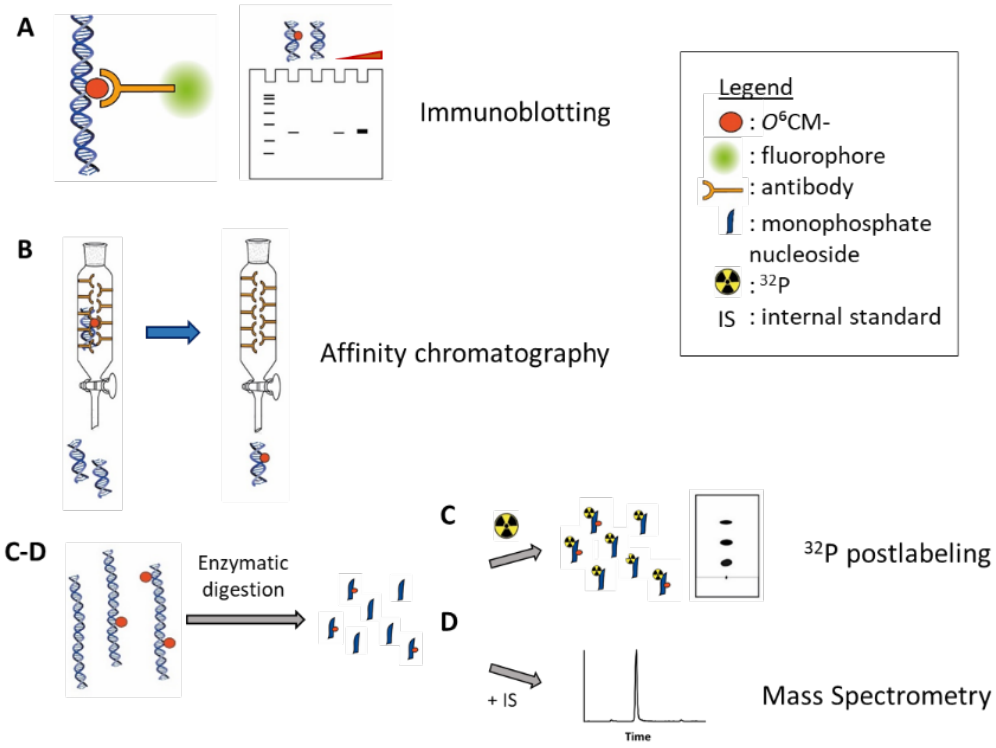


Figure 4. Strategies for O^6-CMG detection. (A) Immunoblotting, (B) affinity chromatography, (C) ^{32}P-postlabeling, and (D) mass spectrometry.

For MS and ^{32}P postlabeling-based detection of O^6-CMG, DNA samples derived from cells exposed to azaserine or KDA,[4,72,102] from healthy volunteers on controlled diets,[103] or from unexposed rat tissue,[101] were enzymatically digested to nucleoside 3'-monophosphate (dNp).[4,72,101-103] For ^{32}P post-labeling, dNps including O^6-CMdGp were phosphorylated with radioactive ^{32}P-ATP, separated by two-dimensional chromatography, and quantified via autoradiography with a limit of quantification of 1 adduct/10^9 nucleotides, allowing the detection of endogenous O^6-CMG adducts in rat stomach tissue at levels ranging from 3 to 56 adducts/10^8 nucleotides (Figure 4C). Once the signal was quantified, dNps were separated by chromatography and dephosphorylated, and their structure was confirmed by UV and LC/MS.[101] MS-based strategies rely on the use of a stable isotope of the adduct of interest as an internal standard for O^6-CMG quantification (Figure 4D).[4,72,102,103] The most sensitive reported measurement of O^6-CMG had a limit of quantification of 73.4 amol, achieved by MS using nanoflow-LC/nanoESI-MS^3.[72] By this approach, the researchers could detect endogenous levels of O^6-CMG as low as 5.4 adducts/10^5nucleotide, which is around 30 adducts per cell. Adduct levels increased by almost a factor of 100 when exposed to up to 450µm azaserine. Despite being more sensitive than antibody-based methods, to our knowledge, MS strategies have not been applied so far (or not successfully)[103] to the quantification of O^6-CMG in samples derived from humans. Nonetheless, the results of these studies provide critical confirmation concerning the validity of using model carboxymethylating drugs to investigate O^6-CMG cellular phenotypes.[4,72,102]

1.9 Strategies to map O^6-CMG In DNA

Typical O^6-CMG detection methods, such as immunoaffinity, postlabeling, or MS approaches as described in Section 7, are extremely valuable for determining the total levels of O^6-CMG in a genome, but information regarding genomic location is completely lost. In the context of a modern epigenetic understanding of genome function, seminal reports have recently emerged concerning strategies for genomic mapping of DNA damage.[41-43] In the case of O^6-CMG, two significant advances of the last two years, namely, a nanopore-based O^6-CMG sequencing approach[50] and an artificial nucleoside/engineered

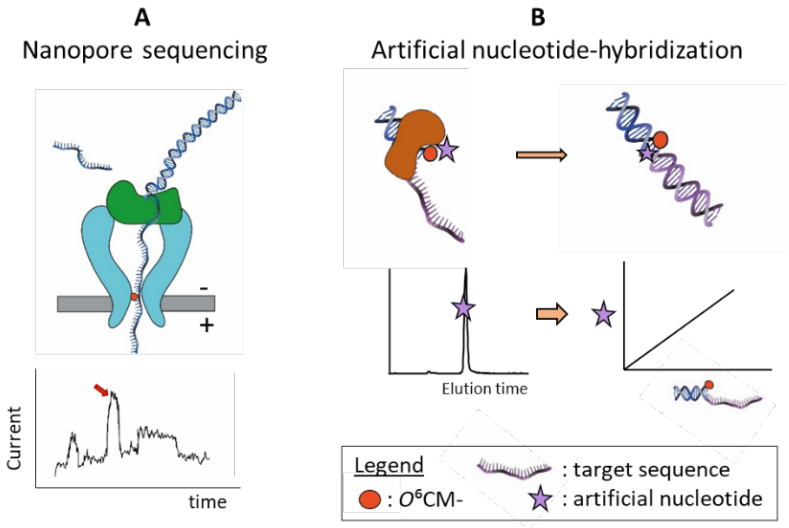


Figure 5. Sequence-targeted detection strategies for O^6-CMG. (A) Nanopore sequencing and (B) artificial nucleotide-based replication and MS detection. These strategies integrate detection of O^6-CMG with information regarding the surrounding DNA sequence.

polymerase-based analysis strategy,[51] are anticipated to enable future research involving the identification of the presence and effects of O^6-CMG in selected DNA regions, such as cancer genes.

The nanopore-based sequencing data reported for samples of DNA containing O^6-CMG entailed the preparation of a library of O^6-CMG-modified oligonucleotides, tagged with a deca(deoxythymidine) sequence.[50] The library was electrophoretically driven into a customized pore, and the passage of ssDNA through the pore caused a variation in the applied current that enabled the identification of the different bases. A high-current signal with flickering characteristics was recorded with high accuracy (95%) at the passage of O^6-CMG (Figure5A).[50] In the same study, the replication of O^6-CMG by Phi29DNA Pol, which serves as a motor protein and controls the passage of the DNA through the nanopore, was characterized. O^6-CMG was a replication blockage for Phi29 DNA Pol. Nevertheless, because the passage of O^6-CMG-DNA through the nanopore occurred and resulted in a high and accurate signal, nanopore sequencing is suggested as a promising strategy for damage sequencing, particularly for damage that is a replication obstacle for Pols and thus hinders the applicability of most next-generation sequencing strategies.

Another emerging strategy for locating DNA adducts in defined sequences, and that has been recently applied to O^6-CMG, involves an artificial nucleotide triphosphate designed to be inserted specifically at O^6-alkyl-guanine damage sites and coupling of this enzymatic amplification process with a mass spectrometric method for quantification of nucleotide insertion events (Figure 5B).[51] The key artificial nucleotide used for this study was a heterocyclic imidic nucleotide that is inserted opposite O^6-CMG but not G, therefore, marking the presence of O^6-CMG. This approach builds on previous research concerning the interaction of artificial nucleotides with DNA adducts and polymerase-mediated amplification of damaged DNA.[104-111] The artificial nucleotide was selectively inserted opposite O^6-CMG by a bypass-proficient KlenTaq mutant DNA Pol and required to efficiently extend a primer annealed to a DNA template containing O^6-CMG, promoting DNA synthesis past O^6-CMG. While the strategy has the potential to target any DNA sequence by employing the corresponding complementary sequence as a DNA primer, increased sensitivity is required to enable its utilization to interrogate defined genomic sequences for O^6-CMG formation and persistence in biological samples to relate with CRC mutation landscapes.

1.10 Conclusions

The O^6-CMG DNA adduct is formed in elevated amounts when people eat a diet high in meat, a known human carcinogen. To address whether the formation of this damage may contribute to the increased risk of cancer associated with diets high in meat, numerous mechanistic studies have addressed the biological impacts of O^6-CMG formation. It induces a spectrum of point mutations that resemble those found in cancer tissues. Via a combination of structural, biochemical, and cellular studies, endogenous processes of DNA repair and replication over O^6-CMG have been investigated. DNA translesion synthesis was identified as a basis for O^6-CMG mutagenicity, while excision or reversal mitigates toxic and mutagenic properties of O^6-CMG.Chemical research in the synthesis and analysis of O^6-CMG has enabled studies leading to these findings, and there are several recent advances anticipated to address the remaining unknown information about O^6-CMG, such as how exactly is it recognized for repair in mammalian cells and how is it distributed in the genome. For example, there is a highly selective metal-catalyzed reaction by which O^6-CMG can be introduced into various forms of DNA in a simple and reliable synthesis protocol. Additionally, two emerging strategies, involving nanopore sequencing and artificial nucleotide-mediated

amplification and detection, are being developed to locate O^6-CMG in DNA. It is anticipated that these emerging technologies will spur further answers concerning how O^6-CMG formation and persistence drive carcinogenesis. The evidence reported herein and the gaps that can be addressed with further research are pivotal to harness the knowledge of O^6-CMG formation and its biological consequences in order to implement monitoring of O^6-CMG in human tissues for reducing cancer risk associated with diet.

Conflicts of interest

The authors declare no competing financial interest.

Funding

Swiss National Science Foundation (156280, 185020), European Research Council (260341), and Krebsliga (KFS-4443-02-2018)

1.11 Acknowledgements

We thank Dr. Susanne Geisen, Dr. Katherine Hurley, Nikolai Püllen, and Vera Hürlimann for valuable feedback on the manuscript. We acknowledge all co-workers and collaborators who contributed to some of the work described in this article, especially Dr. Hailey Gahlon, Dr. Susanne Geisen, Dr. Arman Nilforoushan, Dr. Michael Räz, Dr. Laura Wyss, and Nathalie Ziegler. We also acknowledge Prof. Sarah Shuck, Editorial Advisory Board Member of Chemical Research in Toxicology, for the invitation to submit this article and for her vision and guidance in creating the Building Chemical Bonds series.

1.12 References

1. International Agency for Research on Cancer. Vol. 2020 (https://gco.iarc.fr/today/home).
2. Sehgal, R. *et al.* Lynch syndrome: an updated review. *Genes (Basel)* **5**, 497-507 (2014).
3. Cross, A.J. *et al.* A large prospective study of meat consumption and colorectal cancer risk: an investigation of potential mechanisms underlying this association. *Cancer Res* **70**, 2406-14 (2010).
4. Hemeryck, L.Y. *et al.* In vitro DNA adduct profiling to mechanistically link red meat consumption to colon cancer promotion. *Toxicol Res (Camb)* **5**, 1346-1358 (2016).
5. Friedman, O.M., Mahapatra, G.N. & Stevenson, R. The methylation of deoxyribonucleosides by diazomethane. *Biochim Biophys Acta* **68**, 144-6 (1963).
6. Friedman, O.M., Mahapatra, G.N., Dash, B. & Stevenson, R. Studies on the action of diazomethane on deoxyribonucleic acid. The action of diazomethane on deoxyribonucleosides. *Biochim Biophys Acta* **103**, 286-97 (1965).
7. Montesano, R. Alkylation of DNA and tissue specificity in nitrosamine carcinogenesis. *J Supramol Struct Cell Biochem* **17**, 259-73 (1981).
8. Harrison, K.L., Jukes, R., Cooper, D.P. & Shuker, D.E. Detection of concomitant formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine in DNA exposed to nitrosated glycine derivatives using a combined immunoaffinity/HPLC method. *Chem Res Toxicol* **12**, 106-11 (1999).
9. Vanden Bussche, J. *et al.* O(6)-carboxymethylguanine DNA adduct formation and lipid peroxidation upon in vitro gastrointestinal digestion of haem-rich meat. *Mol Nutr Food Res* **58**, 1883-96 (2014).
10. Lewin, M.H. *et al.* Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res* **66**, 1859-65 (2006).

11. Farmer, P.B., Foster, A.B., Jarman, M. & Tisdale, M.J. The alkylation of 2'-deoxyguanosine and of thymidine with diazoalkanes. Some observations on o-alkylation. *Biochem J* **135**, 203-13 (1973).
12. Van Hecke, T. *et al.* Increased oxidative and nitrosative reactions during digestion could contribute to the association between well-done red meat consumption and colorectal cancer. *Food Chem* **187**, 29-36 (2015).
13. Hemeryck, L.Y., Rombouts, C., De Paepe, E. & Vanhaecke, L. DNA adduct profiling of in vitro colonic meat digests to map red vs. white meat genotoxicity. *Food Chem Toxicol* **115**, 73-87 (2018).
14. Mirvish, S.S. Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC. *Cancer Lett* **93**, 17-48 (1995).
15. Herskowitz, I.H. The relationship of x-ray induced recessive lethals to chromosomal breakage. *Am Nat* **80**, 588-92 (1946).
16. Allsopp, C.B. & Catcheside, D.G. Chemical breakage of chromosomes. *Nature* **161**, 1011 (1948).
17. Lane, G.R. X-ray fractionation and chromosome breakage. *Heredity (Edinb)* **5**, 1-35 (1951).
18. Nasrat, G.E., Kaplan, W.D. & Auerbach, C. A quantitative study of mustard gas induced chromosome breaks and re-arrangements in Drosophila Melanogaster. *Z Indukt Abstamm Vererbungsl* **86**, 249-62 (1954).
19. Freifelder, D., Davison, P.F. & Geiduschek, E.P. Damage by visible light to the acridine orange--DNA complex. *Biophys J* **1**, 389-400 (1961).
20. Rupert, C.S. Repair of ultraviolet damage in cellular DNA. *J Cell Comp Physiol* **58(3)Pt 2**, 57-68 (1961).
21. Setlow, R.B. & Setlow, J.K. Evidence that ultraviolet-induced thymine dimers in DNA cause biological damage. *Proc Natl Acad Sci U S A* **48**, 1250-7 (1962).
22. Hotz, G. Photoreactivation of Uv-Damage in Phage Containing 5-Bromouracil-DNA. *Z Vererbungsl* **95**, 211-4 (1964).
23. Strauss, B., Scudiero, D. & Henderson, E. The nature of the alkylation lesion in mammalian cells. *Basic Life Sci* **5A**, 13-24 (1975).
24. Bartsch, H. Metabolic activation of aromatic amines and azo dyes. *IARC Sci Publ*, 13-30 (1981).
25. Setlow, R.B., Cao, E.H. & Delihas, N.C. Enzymology of repair of DNA adducts produced by N-nitroso compounds. *IARC Sci Publ*, 561-70 (1984).
26. Leng, M. Immunological detection of lesions in DNA. *Biochimie* **67**, 309-15 (1985).
27. Jeffrey, A.M. DNA modification by chemical carcinogens. *Pharmacol Ther* **28**, 237-72 (1985).
28. Strauss, B.S. Molecular biology of the response of cells to radiation and to radiomimetic chemicals. *Cancer* **40**, 471-80 (1977).
29. Bresnick, E. & Eastman, A. Alkylation of mammalian cell DNA, persistence of adducts, and relationship to carcinogenesis. *Drug Metab Rev* **13**, 189-205 (1982).
30. Montesano, R. *et al.* Repair of DNA alkylation adducts in mammalian cells. *Biochimie* **67**, 919-28 (1985).
31. Drake, J.W. & Baltz, R.H. The biochemistry of mutagenesis. *Annu Rev Biochem* **45**, 11-37 (1976).
32. Chambers, R.W. Chemical carcinogenesis: a biochemical overview. *Clin Biochem* **18**, 158-68 (1985).
33. Klaene, J.J., Sharma, V.K., Glick, J. & Vouros, P. The analysis of DNA adducts: the transition from (32)P-postlabeling to mass spectrometry. *Cancer Lett* **334**, 10-9 (2013).
34. Ma, B., Stepanov, I. & Hecht, S.S. Recent Studies on DNA Adducts Resulting from Human Exposure to Tobacco Smoke. *Toxics* **7**(2019).
35. Savoye, I. *et al.* Patterns of Ultraviolet Radiation Exposure and Skin Cancer Risk: the E3N-SunExp Study. *J Epidemiol* **28**, 27-33 (2018).
36. Viegas, S. *et al.* Occupational exposure to aflatoxin (AFB(1)) in poultry production. *J Toxicol Environ Health A* **75**, 1330-40 (2012).
37. Shaw, D. Toxicological risks of Chinese herbs. *Planta Med* **76**, 2012-8 (2010).

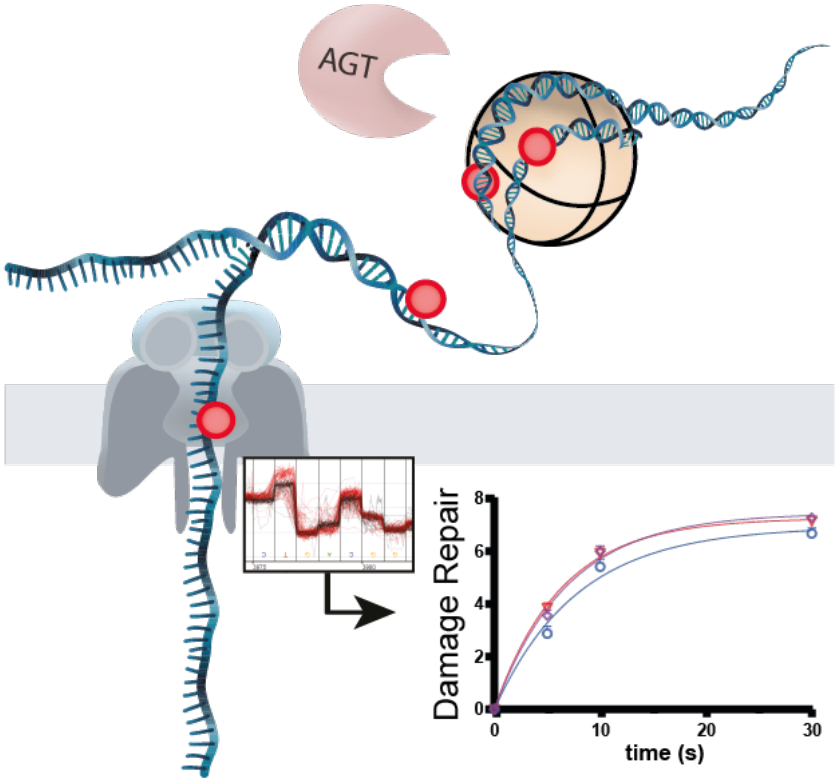
38. Arlt, V.M. *et al.* Aristolochic acid mutagenesis: molecular clues to the aetiology of Balkan endemic nephropathy-associated urothelial cancer. *Carcinogenesis* **28**, 2253-61 (2007).
39. Marchese, S. *et al.* Aflatoxin B1 and M1: Biological Properties and Their Involvement in Cancer Development. *Toxins (Basel)* **10**(2018).
40. Balbo, S., Turesky, R.J. & Villalta, P.W. DNA adductomics. *Chem Res Toxicol* **27**, 356-66 (2014).
41. Yu, Y., Wang, P., Cui, Y. & Wang, Y. Chemical Analysis of DNA Damage. *Anal Chem* **90**, 556-576 (2018).
42. Li, W. & Sancar, A. Methodologies for detecting environmentally induced DNA damage and repair. *Environ Mol Mutagen* **61**, 664-679 (2020).
43. Mingard, C., Wu, J., McKeague, M. & Sturla, S.J. Next-generation DNA damage sequencing. *Chem Soc Rev* **49**, 7354-7377 (2020).
44. Leaf, C.D., Wishnok, J.S. & Tannenbaum, S.R. Mechanisms of endogenous nitrosation. *Cancer Surv* **8**, 323-34 (1989).
45. Loveless, A. Possible relevance of O-6 alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. *Nature* **223**, 206-7 (1969).
46. Gottschalg, E., Scott, G.B., Burns, P.A. & Shuker, D.E. Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis* **28**, 356-62 (2007).
47. Raz, M.H. *et al.* Bypass of Mutagenic O(6)-Carboxymethylguanine DNA Adducts by Human Y- and B-Family Polymerases. *Chem Res Toxicol* **29**, 1493-503 (2016).
48. Raz, M.H., Sandell, E.S., Patil, K.M., Gillingham, D.G. & Sturla, S.J. High Sensitivity of Human Translesion DNA Synthesis Polymerase kappa to Variation in O(6)-Carboxymethylguanine Structures. *ACS Chem Biol* **14**, 214-222 (2019).
49. Wu, J. *et al.* Replication studies of carboxymethylated DNA lesions in human cells. *Nucleic Acids Res* **45**, 7276-7284 (2017).
50. Wang, Y. *et al.* Nanopore Sequencing Accurately Identifies the Mutagenic DNA Lesion O(6) - Carboxymethyl Guanine and Reveals Its Behavior in Replication. *Angew Chem Int Ed Engl* **58**, 8432-8436 (2019).
51. Aloisi, C.M.N., Nilforoushan, A., Ziegler, N. & Sturla, S.J. Sequence-Specific Quantitation of Mutagenic DNA Damage via Polymerase Amplification with an Artificial Nucleotide. *J Am Chem Soc* **142**, 6962-6969 (2020).
52. Armstrong, B. & Doll, R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int J Cancer* **15**, 617-31 (1975).
53. Hughes, R., Cross, A.J., Pollock, J.R. & Bingham, S. Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis* **22**, 199-202 (2001).
54. Bingham, S.A., Hughes, R. & Cross, A.J. Effect of white versus red meat on endogenous N-nitrosation in the human colon and further evidence of a dose response. *J Nutr* **132**, 3522S-3525S (2002).
55. Cross, A.J., Pollock, J.R. & Bingham, S.A. Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. *Cancer Res* **63**, 2358-60 (2003).
56. Norat, T. *et al.* Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst* **97**, 906-16 (2005).
57. Cross, A.J., Greetham, H.L., Pollock, J.R., Rowland, I.R. & Bingham, S.A. Variability in fecal water genotoxicity, determined using the Comet assay, is independent of endogenous N-nitroso compound formation attributed to red meat consumption. *Environ Mol Mutagen* **47**, 179-84 (2006).
58. Bingham, S.A. High-meat diets and cancer risk. *Proc Nutr Soc* **58**, 243-8 (1999).

59. Goldbohm, R.A. *et al.* A prospective cohort study on the relation between meat consumption and the risk of colon cancer. *Cancer Res* **54**, 718-23 (1994).
60. Cross, A.J., Pollock, J.R. & Bingham, S.A. Red meat and colorectal cancer risk: the effect of dietary iron and haem on endogenous N-nitrosation. *IARC Sci Publ* **156**, 205-6 (2002).
61. Loeppky, R.N., Shi, J., Barnes, C.L. & Geddam, S. A diazonium ion cascade from the nitrosation of tolazoline, an imidazoline-containing drug. *Chem Res Toxicol* **21**, 295-307 (2008).
62. Loeppky, R.N. & Shi, J. Nucleoside and DNA adducts from N-nitrosotolazoline. *Chem Res Toxicol* **21**, 319-29 (2008).
63. Loeppky, R.N. & Shi, J. N-nitrosotolazoline: decomposition studies of a typical N-nitrosoimidazoline. *Chem Res Toxicol* **21**, 308-18 (2008).
64. Hecht, S.S. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem Res Toxicol* **11**, 559-603 (1998).
65. Radwan, G., Hecht, S.S., Carmella, S.G. & Loffredo, C.A. Tobacco-specific nitrosamine exposures in smokers and nonsmokers exposed to cigarette or waterpipe tobacco smoke. *Nicotine Tob Res* **15**, 130-8 (2013).
66. Leach, S.A., Thompson, M. & Hill, M. Bacterially catalysed N-nitrosation reactions and their relative importance in the human stomach. *Carcinogenesis* **8**, 1907-12 (1987).
67. Forstermann, U. *et al.* Isoforms of nitric oxide synthase. Characterization and purification from different cell types. *Biochem Pharmacol* **42**, 1849-57 (1991).
68. Knekt, P., Jarvinen, R., Dich, J. & Hakulinen, T. Risk of colorectal and other gastro-intestinal cancers after exposure to nitrate, nitrite and N-nitroso compounds: a follow-up study. *Int J Cancer* **80**, 852-6 (1999).
69. Sartorelli, A.C. & Booth, B.A. Inhibition of the synthesis of thymine nucleotides by azaserine. *Mol Pharmacol* **3**, 71-80 (1967).
70. Kubitschek, H.E. & Sepanski, R.J. Azaserine: survival and mutation in Escherichia coli. *Mutat Res* **94**, 31-8 (1982).
71. O'Driscoll, M., Macpherson, P., Xu, Y.Z. & Karran, P. The cytotoxicity of DNA carboxymethylation and methylation by the model carboxymethylating agent azaserine in human cells. *Carcinogenesis* **20**, 1855-62 (1999).
72. Yu, Y., Wang, J., Wang, P. & Wang, Y. Quantification of Azaserine-Induced Carboxymethylated and Methylated DNA Lesions in Cells by Nanoflow Liquid Chromatography-Nanoelectrospray Ionization Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal Chem* **88**, 8036-42 (2016).
73. Wang, H., Cao, H. & Wang, Y. Quantification of N2-carboxymethyl-2'-deoxyguanosine in calf thymus DNA and cultured human kidney epithelial cells by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope dilution method. *Chem Res Toxicol* **23**, 74-81 (2010).
74. Jianshuang Wang, Y.W. Chapter Six - Carboxymethylation of DNA Induced by N-Nitroso Compounds and Its Biological Implications,. in *Advance in Molecular Toxicology*, Vol. 5 (ed. Fishbein, J.C.) 219-243 (Elsevier, Netherlands, 2011).
75. Bingham, S.A. *et al.* Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? *Carcinogenesis* **17**, 515-23 (1996).
76. Habermeyer, M. *et al.* Nitrate and nitrite in the diet: how to assess their benefit and risk for human health. *Mol Nutr Food Res* **59**, 106-28 (2015).
77. Molly, K., Vande Woestyne, M. & Verstraete, W. Development of a 5-step multi-chamber reactor as a simulation of the human intestinal microbial ecosystem. *Appl Microbiol Biotechnol* **39**, 254-8 (1993).

78. Konishi, Y. *et al.* Pancreatic tumors induced by a single intraperitoneal injection of azaserine in partial pancreatectomized rats. *Cancer Lett* **9**, 43-6 (1980).
79. Longnecker, D.S. & Curphey, T.J. Adenocarcinoma of the pancreas in azaserine-treated rats. *Cancer Res* **35**, 2249-58 (1975).
80. Lilja, H.S., Hyde, E., Longnecker, D.S. & Yager, J.D., Jr. DNA damage and repair in rat tissues following administration of azaserine. *Cancer Res* **37**, 3925-31 (1977).
81. Sedgwick, B. Nitrosated peptides and polyamines as endogenous mutagens in O6-alkylguanine-DNA alkyltransferase deficient cells. *Carcinogenesis* **18**, 1561-7 (1997).
82. Ehrlich, J. *et al.* Antibiotic studies of azaserine. *Nature* **173**, 72 (1954).
83. Wang, P., Leng, J. & Wang, Y. DNA replication studies of N-nitroso compound-induced O (6)-alkyl-2'-deoxyguanosine lesions in Escherichia coli. *J Biol Chem* **294**, 3899-3908 (2019).
84. Prakash, S. & Prakash, L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev* **16**, 1872-83 (2002).
85. Zhang, F. *et al.* Structures of DNA duplexes containing O6-carboxymethylguanine, a lesion associated with gastrointestinal cancer, reveal a mechanism for inducing pyrimidine transition mutations. *Nucleic Acids Res* **41**, 5524-32 (2013).
86. Du, H., Wang, P., Li, L. & Wang, Y. Repair and translesion synthesis of O (6)-alkylguanine DNA lesions in human cells. *J Biol Chem* **294**, 11144-11153 (2019).
87. Aramini, J.M. *et al.* Structural basis of O6-alkylguanine recognition by a bacterial alkyltransferase-like DNA repair protein. *J Biol Chem* **285**, 13736-41 (2010).
88. Wilkinson, O.J. *et al.* Alkyltransferase-like protein (Atl1) distinguishes alkylated guanines for DNA repair using cation-pi interactions. *Proc Natl Acad Sci U S A* **109**, 18755-60 (2012).
89. Tubbs, J.L. *et al.* Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808-13 (2009).
90. Latypov, V.F. *et al.* Atl1 regulates choice between global genome and transcription-coupled repair of O(6)-alkylguanines. *Mol Cell* **47**, 50-60 (2012).
91. Tomaszowski, K.H. *et al.* The bacterial alkyltransferase-like (eATL) protein protects mammalian cells against methylating agent-induced toxicity. *DNA Repair (Amst)* **28**, 14-20 (2015).
92. Senthong, P. *et al.* The nitrosated bile acid DNA lesion O6-carboxymethylguanine is a substrate for the human DNA repair protein O6-methylguanine-DNA methyltransferase. *Nucleic Acids Res* **41**, 3047-55 (2013).
93. Shuker, D.E. & Margison, G.P. Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Res* **57**, 366-9 (1997).
94. Harrison, K.L., Fairhurst, N., Challis, B.C. & Shuker, D.E. Synthesis, characterization, and immunochemical detection of O6-(carboxymethyl)-2'-deoxyguanosine: a DNA adduct formed by nitrosated glycine derivatives. *Chem Res Toxicol* **10**, 652-9 (1997).
95. B.L. Gaffney, R.A.J. Synthesis of O-6-alkylated deoxyguanosine nucleosides. *Tetrahedron Letters* **23** 2253-2256 (1982).
96. Xu, Y.-Z. Synthesis and Characterization of DNA Containing O6-Carboxymethylguanine. *Tetrahedron* **56**, 6075-6081 (2000).
97. Shibata, T. *et al.* Novel synthesis of O6-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, O6-methylguanine DNA methyltransferase (MGMT). *Nucleic Acids Res* **34**, 1884-91 (2006).
98. Millington, C.L. *et al.* Convenient and efficient syntheses of oligodeoxyribonucleotides containing O(6)-(carboxymethyl)guanine and O(6)-(4-oxo-4-(3-pyridyl)butyl)guanine. *Nucleosides Nucleotides Nucleic Acids* **31**, 328-38 (2012).
99. Geigle, S.N., Wyss, L.A., Sturla, S.J. & Gillingham, D.G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem Sci* **8**, 499-506 (2017).

100. Cupid, B.C., Zeng, Z., Singh, R. & Shuker, D.E. Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chem Res Toxicol* **17**, 294-300 (2004).
101. Terasaki, M. *et al.* Detection of endogenous DNA adducts, O-carboxymethyl-2'-deoxyguanosine and 3-ethanesulfonic acid-2'-deoxycytidine, in the rat stomach after duodenal reflux. *Cancer Sci* **99**, 1741-6 (2008).
102. Vanden Bussche, J., Moore, S.A., Pasmans, F., Kuhnle, G.G. & Vanhaecke, L. An approach based on ultra-high pressure liquid chromatography-tandem mass spectrometry to quantify O6-methyl and O6-carboxymethylguanine DNA adducts in intestinal cell lines. *J Chromatogr A* **1257**, 25-33 (2012).
103. Da Pieve, C., Sahgal, N., Moore, S.A. & Velasco-Garcia, M.N. Development of a liquid chromatography/tandem mass spectrometry method to investigate the presence of biomarkers of DNA damage in urine related to red meat consumption and risk of colorectal cancer. *Rapid Commun Mass Spectrom* **27**, 2493-503 (2013).
104. Gong, J. & Sturla, S.J. A synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA. *J Am Chem Soc* **129**, 4882-3 (2007).
105. Gahlon, H.L. & Sturla, S.J. Hydrogen bonding or stacking interactions in differentiating duplex stability in oligonucleotides containing synthetic nucleoside probes for alkylated DNA. *Chemistry* **19**, 11062-7 (2013).
106. Kowal, E.A. *et al.* Recognition of O6-benzyl-2'-deoxyguanosine by a perimidinone-derived synthetic nucleoside: a DNA interstrand stacking interaction. *Nucleic Acids Res* **41**, 7566-76 (2013).
107. Stornetta, A., Angelov, T., Guengerich, F.P. & Sturla, S.J. Incorporation of nucleoside probes opposite O(6)-methylguanine by Sulfolobus solfataricus DNA polymerase Dpo4: importance of hydrogen bonding. *Chembiochem* **14**, 1634-9 (2013).
108. Wyss, L.A., Nilforoushan, A., Williams, D.M., Marx, A. & Sturla, S.J. The use of an artificial nucleotide for polymerase-based recognition of carcinogenic O6-alkylguanine DNA adducts. *Nucleic Acids Res* **44**, 6564-73 (2016).
109. Wyss, L.A. *et al.* Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *J Am Chem Soc* **137**, 30-3 (2015).
110. Aloisi, C.M.N., Sturla, S.J. & Gahlon, H.L. A gene-targeted polymerase-mediated strategy to identify O(6)-methylguanine damage. *Chem Commun (Camb)* **55**, 3895-3898 (2019).
111. Raz, M.H., Aloisi, C.M.N., Gahlon, H.L. & Sturla, S.J. DNA Adduct-Directed Synthetic Nucleosides. *Acc Chem Res* **52**, 1391-1399 (2019).

Chapter 2 – Through the nanopore: Quantification and mapping of O^6-alkylguanine repair in nucleosomes



Emma S. Sandell, Qingzeng Gao, Alona Slastennikova, Sabrina M. Huber, Claudia M. N. Aloisi, Yael David, Shana J. Sturla*, Through the nanopore: Quantification and mapping of O^6-alkylguanine repair in the nucleosome, *Manuscript in preparation*

2.1 Abstract

O^6-alkylguanine modifications in DNA can arise from chemical exposures—drugs, the environment and diet—and are repaired by alkylguanine transferase (AGT). This direct reversion enzyme transfers the alkyl group from DNA to cysteine residue 145 of the AGT protein. The efficiency of DNA repair may be influenced by its packaging of chromosomal DNA in nucleosomes raising questions concerning the impact of DNA damage accessibility and repair efficiency in nucleosomes. However, there is a lack of methods to quantify O^6-alkylguanine levels at structurally diverse positions in the nucleosome as a basis for tracking relative repair proficiency. Nanopores sequence DNA based on structural information and have mainly been used for mapping epigenetic modifications in DNA. Here, we developed a nanopore-based method to sequence the biologically relevant O^6-alkylguanine modifications O^6-methylguanine and O^6-carboxymethylguanine. Furthermore, we investigated the influence of guanine alkylation on nucleosome stability and repair of site-specific base damage by AGT. By using nanopore raw current data, we quantified DNA damage with high specificity and accuracy. Using this approach, we confirmed that AGT repair occurs in a bi-directional matter. In addition, we found that the positioning of methyl groups at the O^6-position of guanine inward rather than outward in the nucleosome impedes repair. These results enable simultaneous relative quantification of biologically relevant, mutagenic DNA modifications, while preserving DNA sequence information. Moreover, they provide unprecedented insight on how positioning of spurious methyl groups in DNA protects them from enzymatic repair.

2.2 Introduction

Alkylating agents, such as anticancer drugs, antibiotics[1] as well as chemicals in the environment[2] and food,[3,4] can react with nucleophilic positions of DNA nucleobases to form cytotoxic or mutagenic DNA damage. These chemical modifications can distort the DNA helix and alter nucleosome stability by impairing bending and wrapping around histones.[5,6] Wrapping also influences enzyme access to DNA and might impair the function of DNA-interacting proteins such as those critical in repairing or erasing DNA modifications.[7-10] How alkylation affects nucleosome stability and if nucleosome affects DNA repair is crucial to understand how chemically induced DNA alkylation drives genomic instability.

DNA alkylation occurs infrequently at the O^6-position of guanine, but the resulting damage have significant biological implications for cell viability and mutagenicity. For example, O^6-carboxymethylguanine (O^6-CMG) is formed by *N*-nitroso compounds found in high red or processed meat[4] and O^6-Methylguanine (O^6-MeG) arises from tobacco smoke or anticancer drugs[2]. O^6-MeG is efficiently repaired by a highly conserved protein O^6-alkylguanine DNA alkyltransferase (AGT)[11,12], but the true biological extent of O^6-CMG as an AGT substrate is yet unclear. O^6-CMG repair studies carried out with isolated enzymes have shown contradictory results[13,14], while studies using knock-out cells have supported that AGT does contribute to O^6-CMG repair.[15,16]

AGT directly transfers an alkyl group from DNA to its cysteine residue (Cys145) and as such, the enzyme is irreversibly inactivated.[11] Monomeric AGT slides bidirectionally along the DNA, and forms stable cooperative binding complexes at target sites, consisting of 7 AGT molecules covering around 30 bp to support repair.[17,18] To form cooperative assemblies, DNA sites need to be unobstructed and formation depends strongly on the helical twist and the torsional flexibility of the AGT-DNA complex.[19] DNA repair

enzymes are influenced by the compact and dynamic structure of chromatin and understanding variations in the positioning of DNA alkylation damage and their repair in nucleosome, together with their effects on nucleosome stability, is crucial to understand genomic instability and damage distribution. Although it has been suggested that binding complexes may fit in between nucleosomes, or in open regions near DNA replication complexes[19] it is not known how well AGT can access damage contained within the nucleosome and whether efficiency of damage removal is impaired.

Nucleosome core particles (NCP) are the core unit of chromatin and consist of 147 base pairs of DNA wrapped around an octameric protein core with a dyad axis running through the center of the NCP.[7] Early studies showed increased accessibility and activity to sites closer to ends of nucleosome DNA compared to the dyad axis due to the spontaneous unwrapping of DNA and temporary exposure of sites that are normally occluded.[9,20] Furthermore, it has been proposed that damage with outward rotational position from the histone are more easily repaired than inward facing damage.[20,21] However, the outward position was not consistently favored over inward position when five different glycosylases were tested, suggesting that activity on NCPs is highly variable and might depend on an array of factors such as solvent accessibility, size and structure of the damaged base and mechanism of the glycosylase.[7] These principles of damage repair in nucleosome have been tested for glycosylases but have not been established for AGT repair of alkyl damage. In particular, the AGT reliance on cooperative binding and on unobstructed DNA raise important questions on the impact of DNA damage accessibility for AGT damage repair. Being able to quantify O^6-MeG and O^6-CMG at different positions in the nucleosome is essential to investigate on how nucleosome stability is affected by damage and how well damage is repaired. However, a limitation in studying repair of O^6-alkyl guanine damage in nucleosomes relates to a lack of methods for accurately quantifying O^6-alkylguanine repair at multiple positions simultaneously, such that the sequence context remains intact.

Existing insights on O^6-alkylguanine repair has been gained by studies in which restriction enzymes coupled with human AGT was used, and products were separated and visualized by gel electrophoresis.[13] More accurate quantification of alkyl guanine damage was achieved with LC-MS approaches[22] or by hybridizing probes attached to gold nanoparticles for colorimetric quantification.[23] However, both methods do not give direct sequence information. Artificial nucleosides have also been designed to be inserted specifically at O^6-CMG damage sites and coupled with mass spectrometry to locate and quantify O^6-CMG.[24] However, sensitivity remains a limitation. Nanopores seems to be particularly well suited to quantify and detect different kinds of damage as it sequences DNA based on structural information, and it has been used to locate O^6-MeG and O^6-CMG.[25] By this approach, structural variations can be detected and distinguished by measuring current changes as DNA passes through a nanopore. To drive ssDNA molecules through a nanopore, which is embedded into a lipid membrane, electrostatic potential is applied. With current as the sole driving force, DNA would pass the pore too quickly to achieve an adequate signal, therefore, pores are functionalized with helicases or polymerases, to allow for an increased level of control.[26]

Aside from O^6-MeG and O^6-CMG, nanopores have been used to locate modifications with diverse structures in DNA strands, including 8oxoG[27], cisplatin damage[28] and benzo(a)pyrene[29]. Initial research involved the use of wild-type alpha-hemolysin nanopores and immobilization of DNA within ion channel with biotin-streptavidin complex to reduce translocation speed for accurate single-base identification. In this way, 8-oxoG and abasic site derivatized with a crown-ether in single stranded DNA, induced enough

current blockade to detect modification.[27,30] Oxford nanopore (ONT), which is commercially available, relies on bacterial amyloid secretion pore CsgG as sequencing nanopore, and a Hel308 helicase as a motor protein to reduce signal-to-noise ratio and increase sequencing accuracy.[31] It has successfully been used to detect at least 10 different modification structures inserted in plasmid DNA or via PCR.[32,33] Especially bulky modifications induced a larger signal change than smaller modifications, thus were easier to detect, however O^6-CMG was not analyzed in this study.[32,33] Using mycobacterium smegmatis porin A (MspA) nanopores functionalized with Phi29 DNA polymerase as motor protein, O^6-CMG was effectively analyzed based on the induced shifts in current or using polymerase stalling kinetics data.[25,34] The later analysis strategy was more accurate since the polymerase stalling is not significantly influenced by sequence contexts.[34] The use of this approach for biochemical studies of O^6-CMG is limited, however, by the low throughput and more limited availability of the MspA nanopore.

Oxford nanopore technology (ONT) was used to quantify O^6-MeG and O^6-CMG in specific sequence contexts in DNA to investigate efficiency of repair in the nucleosome. By recording current data of O^6-MeG- and O^6-CMG-modified DNA passing through the nanopore, we trained and validated a model to identify each sequenced strand to be modified or unmodified in order to use it as a tool for damage quantification. The linear relationship between nanopore current signal and DNA damage was tested with different mixtures of modified plasmid and cross-validated with LC-MS measurements. The model was used to test hAGT repair of O^6-MeG and O^6-CMG in different sequence contexts, directional movement of the enzyme during repair and efficiency of repair in the nucleosome compared to free DNA. Furthermore, we tested how DNA damage affects nucleosome stability.

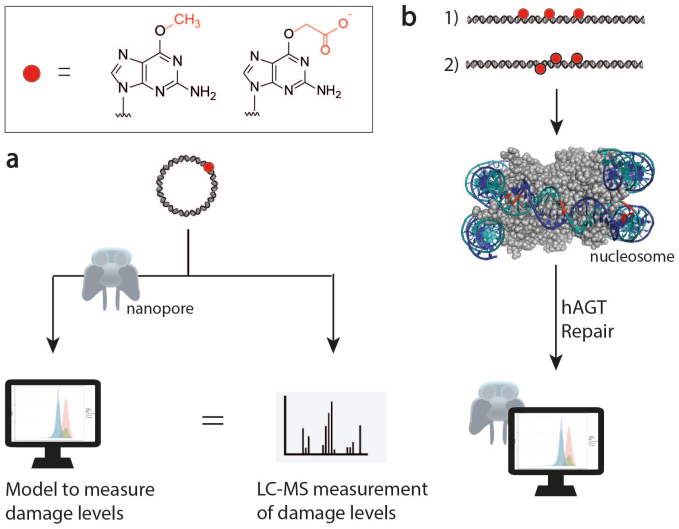


Figure 1. Overview of experimental design. a) Development of a nanopore-based predictive model for b) measuring repair of O^6-alkylguanine damage by hAGT.

2.3 Results

***O*^{6}-alkylguanines can be located and distinguished based on raw current data.** To detect O^6-CMG or O^6-MeG with a nanopore sequencer, we prepared DNA templates containing a single modification in a defined location of a plasmid. The O^6-CMG was prepared by copper-catalyzed alkylation with ethyldiazoacetate[35], allowing to introduce a carboxyethylester moiety at the O^6 position of guanosine. The modification was incorporated into oligonucleotides via phosphoramidite chemistry and hydrolyzed as described previously.[36] The identity of the modified strand was confirmed by mass spectrometry (Figure S1). Using endonuclease nb.BbvCI we nicked the pEGFP-W plasmid at two sites on its minus strand to excise a 20 bp region of ssDNA. This portion of DNA was replaced with a synthetic oligonucleotide (Table S1) containing

either O^6-MeG, O^6-CMG or G (Figure 1a). Following the replacement reaction, the vector was linearized with the restriction enzyme HF NotI to generate a 4728 bp double-stranded DNA duplex with one modified guanine on the minus strand at position 752. The length of the modified plasmids was confirmed by automated gel electrophoresis (Figure S5). The duplexes with each of the two modifications, or the control with no modification, were used for sequencing studies.

Using an ONT MinION sequencer, we tested whether the O^6-alkylguanine modifications could be distinguished on the basis of raw current data. For nanopore sequencing, double-stranded DNA is first unzipped by a DNA helicase and is then translocated through the nanopore by an ionic current. Depending on the sequence and structure of the DNA transitioning through the pore, changes in ionic current are measured, and the output data are called squiggles.[31] If an alkyl group is present on a nucleobase, the signal will deviate at and around the alkylated position. In a first step, squiggles were re-annotated to a reference sequence based on expected current values, using the nanopore sequencing analysis tool called tombo.[37] Then, the re-squiggled signals derived from either control plasmid and a damage-containing plasmid were overlaid for comparison (Figure 2a). Damage-containing plasmids (blue or green) differed from the control plasmid (black lines) at the modified and two neighboring bases. Furthermore, current measured for O^6-MeG and O^6-CMG were significantly different from each other, especially at the base downstream of the modification (Figure 2b), suggesting nanopore sequencing could be used to detect O^6-alkylguanine damage and distinguish alkyl groups.

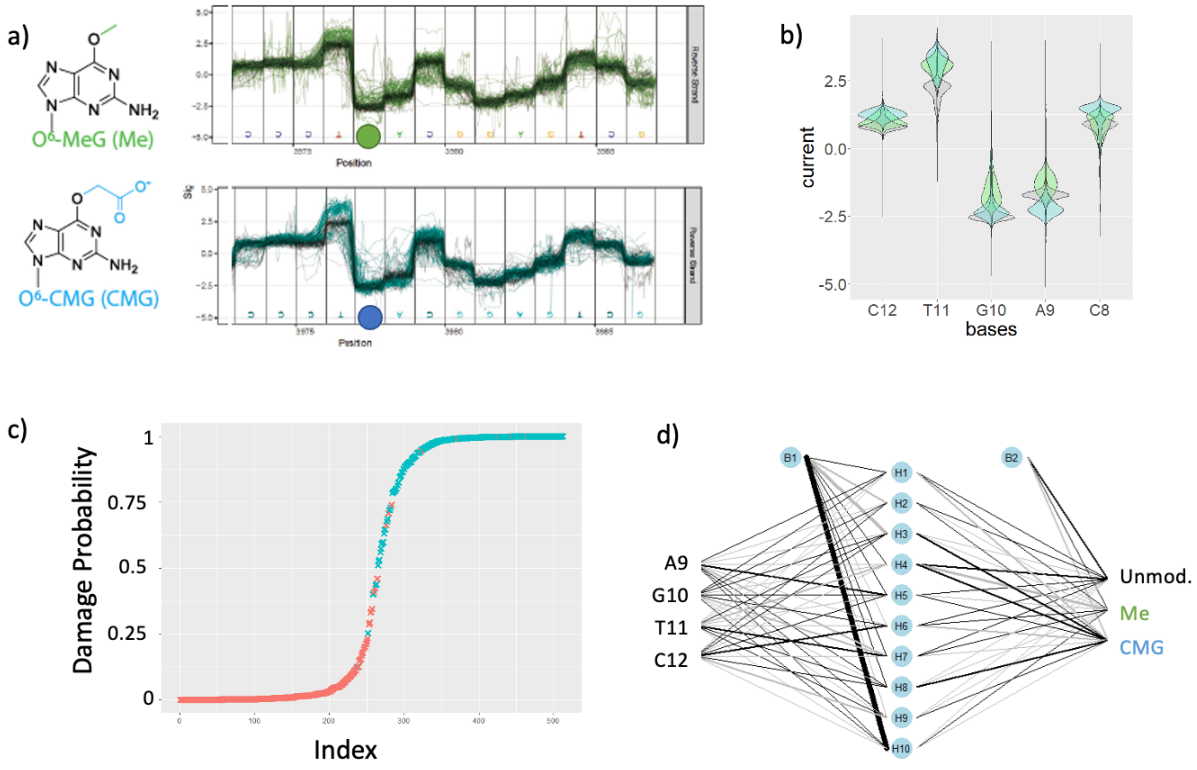


Figure 2 a) current (squiggles) measured at damage location of unmodified (black), O^6-MeG (green) and O^6-CMG (blue) of plasmid. b) extracted current data at and around modified base based on which c) binary logistic regression model (red-unmodified / blue-modified) and d) neural network model was built.

To check if the damage could also be identified by the time of translocation through the pore, and not only by the current change, we investigated the translocation kinetics of the DNA through the pore. In previous studies involving the use of ONT nanopores to sequence modified RNA, it was found that DNA translocation through the pore slows down when the modified base passes the helicase active site and a second time when it passes the pore, which leads to slower translocation 10-15 bp downstream of the modified base and at the modification site itself. Although in previous studies, it was claimed that translocation rate data were more accurate than current data for damage detection, we did not find a significant decrease of translocation rates for the modified strands, neither at the location of the modification, nor downstream of the modification (Figure S6). Therefore, for further studies we only used current data, and not translocation rate data as input for our models.

Relative quantification of O^6-alkylguanines using statistical models based on raw current data. To test the extent of O^6-alkylguanine modification at specific positions can be relatively quantified using statistical models, we evaluated the use of LR to estimate the probability of a read to result from DNA being modified at a specific position. Re-squiggled signals from the unmodified and modified plasmids were extracted from tombo fast5 files and analyzed using R by plotting current data of the modified and neighbor bases. We observed that the current signal measured at the modified position partly overlapped with the one for unmodified DNA and was otherwise comprised of a unique current signal we attributed to the presence of the alkyl group (Figure S8). Observing the signal for the unmodified strand was consistent with the presence of 20-25% unmodified strands, resulting from incomplete replacement of segments within a plasmid with a modified oligonucleotide, consistent with previous examples.[38] Thus, to extract the signals for the alkylated strand, we used k-means (k=2) clustering (stats package in R) on the raw data measured for the modified and neighbor bases. The cluster-groups for which the signals did not overlap with the unmodified signals were used to construct a predictive statistical model to call the modification positions. (Figure S8)

To call unmodified from O^6-Me or O^6-CMG positions, a binary model was built (Figure 2c) and to distinguish between all three modifications, a neural network model (figure 2d) was used. The binary model was built based on logistic regression (LR) using the generalized linear function model (glm.fit) in R, with family set as binomial. Three independent variables were used as input: current data from modified base and from the neighboring bases up- and downstream. The three-way classifier was built using a neural network model (NNM) with NeuralNetTools package in R. To train and test the models, unmodified and extracted modified data was randomly split, and 70% were used as training (n=2940) and 30% as test (n=1260) data. For the binary model, the glm.pred function in R was used to classify each strand as modified or unmodified. Model performance was tested with the test set and evaluated based on accuracy, sensitivity, and specificity of the predictions. The binary model performed almost equally well for both chemical modifications: O^6-MeG was identified with 96.5% sensitivity, 97.7% specificity, and 97.1% accuracy and O^6-CMG was identified with 96.6% sensitivity, 96.4% specificity, and 96.5% accuracy. The accuracy of three-way classifier (calculated with the 'ypred' function) varied slightly depending on the target: 98.2% for unmodified, 90.8% for O^6-MeG, and 93% for O^6-CMG. The capabilities and high performance of the binary model suggested its utility as a relative quantification tool for alkylated DNA as a minority component of DNA.

Logistic regression outperformed tombo in relative damage quantification. To determine the proportion of modified bases in a bulk sample of DNA plasmid, we mixed modified and unmodified plasmids in varying ratios (5%-100%), sequenced these mixtures with ONT, and analyzed the output with our LR binary prediction model. The modified proportions were also determined with LC-MS. To account for the presence of unmodified plasmid in the modified sequence, and inherent sequencing error, the levels of modification extracted from the nanopore raw current data were determined relative to the highest value, which was fixed to 100%, for both ONT and, LC-MS measurements. For both O^6-MeG and O^6-CMG, we could use the LR model to accurately determine the proportion of damage in the plasmid, except for the plasmid with 0% damage, where the model overestimated the levels (Figure 3). When we compared the performance of the LR model with tombo's damage detection tool, we found that the damage-proportion could be detected at specific locations after re-squiggling, using the "detect_modification -model_ sample_compare" option and "text_output –fractions" command in tombo, and the LR tool performed considerably better in the lower range of modification levels, i.e. below 40%. At these low levels, the use of tombo resulted in large overestimates of the proportion of modification. Thus, LR appears to be more accurate when low levels of DNA modifications need to be quantified with the nanopore.

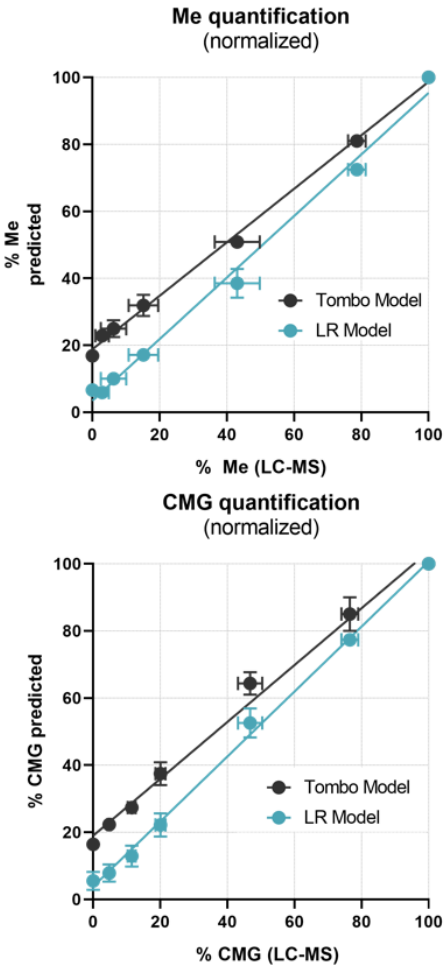


Figure 3. LR model were developed to quantify proportions of modified plasmid better than the tombo model (tool for DNA modification detection) at low concentrations of damage. LC-MS measurement was used as a reference.

Alkylation induced significant current deviations in the nanopore signal of a nucleosome positioning sequence. Having established a method to relatively quantify DNA base modification on the basis of raw current changes acquired with an ONT nanopore, we aimed to use this approach to characterize how the positioning of O^6-alkylguanine in the nucleosome influences the efficiency with which it can be repaired by the direct reversion repair enzyme hAGT. The nucleosome positioning Widom 601 sequence, which binds strongly around histone octamers and provides a homogeneous and predictable particle geometry, was used for this purpose. We placed DNA damage in one of three positions near the center of the duplex: outward, towards the solution (out); in a mid-position (mid); or inward, towards the histone core (in) as defined by the 601 NCP crystal structure (PDB ID: 3LZ0, Table S1). Oligonucleotides containing O^6-CMG were synthesized as previously described,[35,36] and their identity of the modified strand was confirmed by mass spectrometry (Figure S2-4). Seven oligonucleotides, of which three were modified with O^6-CMG, O^6-MeG or unmodified, were annealed and ligated to form the 601 sequence and the length was confirmed by automated gel electrophoresis (Figure S5).

The reads obtained from nanopore analysis of the 601-sequence containing O^6-MeG (green signal, Figure 4) or O^6-CMG (blue) had significantly different current levels at the positions corresponding to the modified

base and the neighboring bases, as compared to reads from corresponding control 601 sequence (black) at all three positions. Additionally, the currents measured for O^6-MeG and O^6-CMG were significantly different from each other, which would potentially allow to also distinguish between different types of damage.

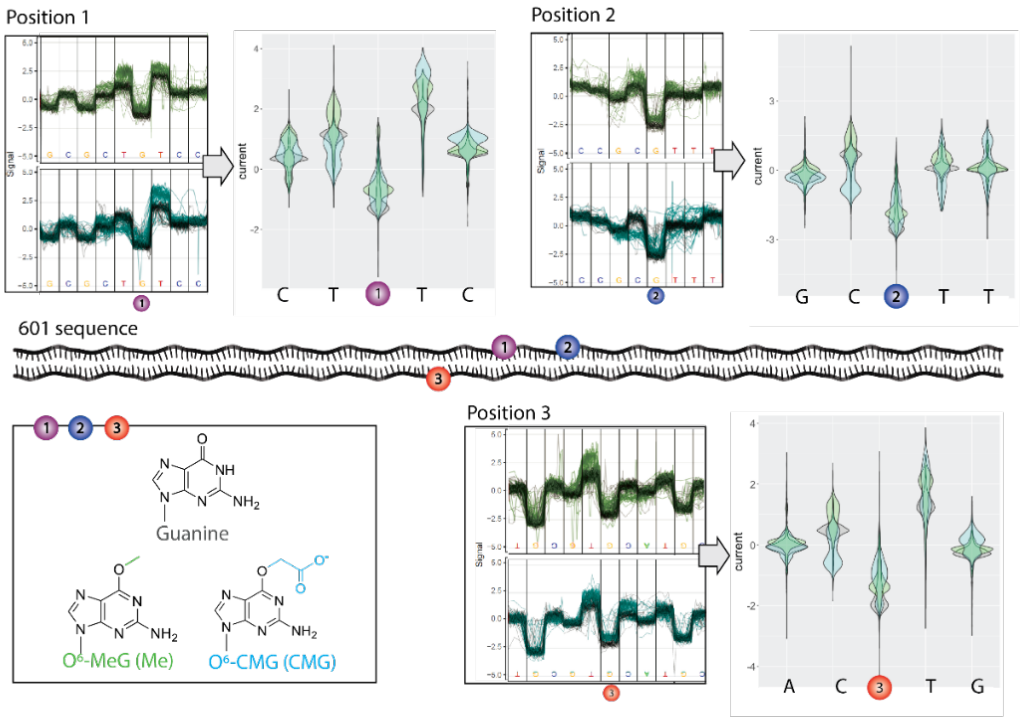


Figure 4. Current signals/squiggle plots and extracted data at and around modified base is shown for damage position 1-3 in the 601 sequence. At all positions measured current deviates for both O^6-MeG (green) and O^6-CMG (blue) from unmodified control (black).

Re-squiggled signals from the unmodified and modified 601 were extracted from tombo fast5 files to build a LR regression model for each position as described previously in the chapter "Relative quantification of O^6-alkylguanines using statistical models based on raw current data". The model performed well in terms of accuracy, specificity, and sensitivity for all tested positions for both modifications. (Table S2) Since the damage levels were known for the input 601 sequence, we were able to obtain absolute values for the damage.

Nucleosomes are formed with O^6-alkylguanines but are less stable. To test whether stable nucleosomes are formed with damaged strands, the modified 601 sequences modified with O^6-CMG or O^6-MeG at single positions, or all three positions were reconstituted with histones to form nucleosomes. The nucleosomes formed were analyzed by non-denaturing gels (Figure 5). It was confirmed that NCPs were formed for all duplexes containing

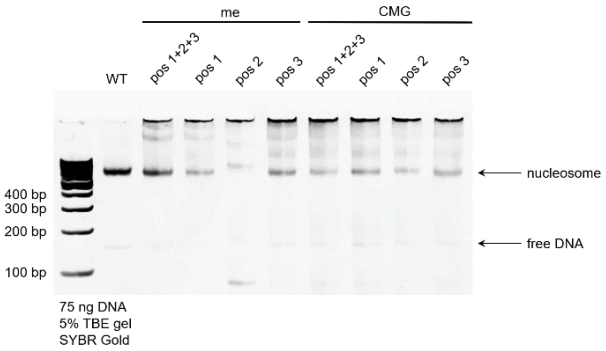


Figure 5. Non-denaturing gel were used to confirm nucleosome formation with damaged DNA

modifications. However, for all nucleosomes containing alkylated DNA it looked blurry near the top end of the gel, indicating that the nucleosomes were less stable than the non-modified 601 sequence. Thermal shift assays will be performed with the same damaged nucleosome complexes to compare the thermal stability of non-modified and modified nucleosome complex, however, this experiment is still in progress.

Orientation of the damage towards the nucleosome affects repair efficiency. Before conducting repair studies in nucleosomes comprised of DNA wrapped around histones, we tested for directional bias (ie proximity to 3' vs 5' end of the DNA) of O^6-MeG repair, and found no evidence for it (Figure 6A, Table 1), which is in line with previous findings.[17] Thus, we placed three methyl or carboxymethyl groups at three different positions in the 601 nucleosome sequence and measured initial reaction rates by allowing the construct to react with hAGT in either packed as a nucleosome or as a free DNA. After 30s, a maximum of 70% repair was observed to occur for all three positions in free duplex substrates, and repair rates were the same for all three positions (Figure 6B). the relatively low maximal repair was due to the low amount of enzyme we used compared to substrate (8:1) but nearly full repair can be achieved by increasing the enzyme amount.[12,39] These results suggest that there are no preferences for hAGT repair amongst the tested sequence contexts (Table 1, Free DNA). When incubating O^6-CMG with the hAGT, no repair was detected after 2minutes incubation, even when testing a much larger excess than what was used to repair O^6-MeG (3-fold), suggesting that O^6-CMG is not a substrate for isolated hAGT.

When we performed the same experiment but wrapped the 601 sequence with O^6-MeG damage around histone octamers, repair was affected by the position and orientation of the damage towards nucleosome. The more accessible positions 1 and 2 (placed in an outward or

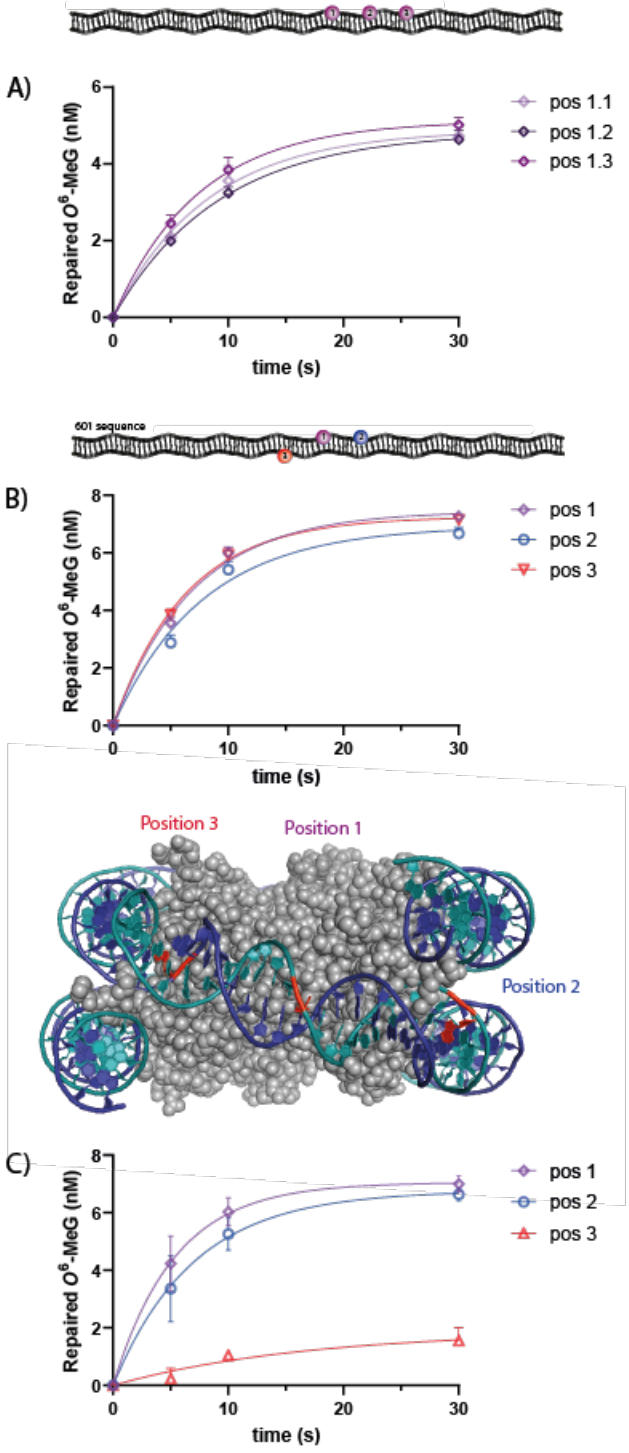


Figure 6. hAGT repair of O^6-MeG was tested in different contexts. A) a template with three modifications in the same sequence and strand, each 20 bases apart, was incubated with hAGT for test for directionality of repair. B) a 601 template containing three modifications with different rotational positions (when in nucleosome) were tested for repair in free and C) nucleosome DNA.

mid/position respectively, Figure 6C) were repaired at 70%, comparably to results for free DNA. On the other hand, Position 3, which was equally well repaired as position 1 and 2 in free DNA, was only poorly repaired in the nucleosome. Namely, only 15% of O^6-MeG at position 3 was repaired after 30 seconds. These results indicate that efficiency of damage repair is not influenced by lateral positioning (e.g. damage towards 5' or 3' end) or sequence context but by the positioning in the nucleosome.

Table 1. The time-course of repaired O^6-MeG was plotted, and curves were fit to the first-order equation.[a]

	Position	K_{obs}/s	Fold change
Directionality	1.1	0.122±0.029	2
(Fig 4A)	1.2	0.109±0.013	1.8
	1.3	0.134±0.021	2.2
Free DNA	1	0.145±0.019	2.4
(Fig 4B)	2	0.128±0.027	2.1
	3	0.158±0.016	2.6
Nucleosome	1	0.187±0.054	3.1
(Fig 4C)	2	0.144±0.051	2.4
	3	0.06 ±0.075	1

[a] $P=A(1-e^{-kt})$ where P is the concentration of product (repaired O^6-MeG) at time t, A is the total concentration of O^6-MeG duplex substrate (10nM) and k is the first-order rate constant (k_{obs}).

2.4 Discussion

hAGT has been extensively studied *in vitro* for the repair of O^6-alkylguanine, but to measure repair in structurally complex and biologically relevant environments, such as the nucleosome, methods are needed that can directly measure alkylation levels at multiple positions. We used, therefore, raw current data acquired by oxford nanopore technology to assess hAGT repair of O^6-MeG and O^6-CMG. A unique current signal at and around the modified positions in the plasmid or 601 strand was evident, and current signals of O^6-MeG and O^6-CMG were significantly different from each other, suggesting that current data can be used to distinguish between alkyl groups. Using a predictive statistical model based on LR, we were able to accurately measure levels of O^6-CMG and O^6MeG. Stable nucleosomes were formed with damaged DNA, and we found that that the efficiency of damage repair in the nucleosome is influenced by the orientation of the damaged base, as O^6-MeG was repaired only when it was facing outwards towards solution, but not inwards towards the histone molecule. Furthermore, we found no sequence context dependence or a directional bias for hAGT's function. Interestingly, O^6-CMG was not repaired by hAGT, even at large excess of the enzyme. Thus, the nanopore-based DNA alkylation sequencing, based on nanopore raw data and a predictive statistical model, proved to be a useful tool for investigating several repair-related questions.

Nanopore raw data. Examining current plots (Figure 2, 4) clearly illustrated that DNA modifications significantly alter current signals up to two bases around the modified base. This is expected since a pore can accommodate approximately five nucleotides during measurement.[40] Furthermore, we found that current signals differed between alkyl groups. Compared to the nonmodified strand, the deviation was slightly larger for O^6-CMG compared to O^6-MeG. Moreover, current was a more effective way to distinguish alkylation groups than their pore passage rates. The differential alterations in current for the carboxymethyl vs methyl groups might be due either to the negative charge of the carboxymethyl or its larger size. Bulky modifications generally lead to a more significant or wider spread in signal change, which has also been seen in previous studies.[32,33] So far, there is no evidence that charge hugely influences current data, however, data is scarce. In one study, C5-cytosine variants were sequenced, including the charged modification 5-carboxylcytosine, using a phi29 DNA polymerase-controlled pore, and the current signal difference to the control was more significant than for the other uncharged modifications, however, it was also the bulkiest modification of all.[41] While we could distinguish the presence or absence, as well as the identity of alkyl groups, we did not observe large differences in the rates with which unmodified or modified strands passed through the nanopore. Most experiments where kinetic data obtained by ONT was successfully used for damage detection, mainly concerned RNA, not DNA.[40] ONT processes DNA at a speed of 450 bp/s for DNA vs only 70 bp/s for RNA[42], which might account for the higher resolution in kinetics signals for RNA analysis. Also, in studies using self-assembled pores functionalized with a polymerase, kinetic data was used to detect modifications, which can significantly lower the speed of the strand going through the pore, increasing the resolution.[34] Furthermore, structure activity relationship studies with analogs of O^6-CMG have shown that charged DNA modifications may have influence on the processivity of polymerases, which may lead to slower transition through pore.[43] Current data was found to be more important than kinetic data in distinguishing damaged from undamaged base, and was therefore used to build predictive models for processing the data.

Statistical Models. Using nanopore current data and a LR model, we accurately quantified O^6-CMG and O^6-MeG in specific sequence contexts, even at low proportions of damage (<40%). At these levels, we found that the tombo prediction model overestimated the damage. This result is consistent with findings from a previous study where damage detection tools tombo, nanopolish and deepsignal were used to predict DNA 5-metylcytosine modifications at specific locations in yeast DNA. Damage fraction and damage probability were predicted and compared with LC-MS and PacBio data. All three tested damage detection tools overestimated damage at low levels.[44] ELIGOS, another damage detection tool, was used to analyze a library of 15 plasmids containing site-specifically inserted damage of varying sizes. Sequencing errors of base-calling data were used to identify *in silico* mixtures with different proportions of DNA containing vs. lacking modification inserted at a specific spot. In that study, most adducts were detectable as low as 5%, similar to what we found. However, base-calling data could not be used to distinguish between different damaged bases.[33]

When analyzing our data acquired from the repair experiments with the tombo tool (Figure S11), we found a similar enzymatic repair rate (k_{obs}) as with the LR model (Table S3), since for enzyme kinetics, initial repair rates are measured when damage levels were relatively high. Also, hAGT-mediated repair of O^6-MeG did not proceed to completion, and plateaued already at 60-70% repair, which corresponds to 30-40% O^6-MeG remaining in the strand. At these damage levels the tombo fraction tool performs still relatively well.

However, if the repair reaction were to go to completion, damage would be expected at low levels, where LR model performs better than tombo.

Nucleosome Repair. *O*^{6}-MeG was efficiently repaired in the nucleosome when the damage was facing outwards towards solution, but not when it was oriented toward the nucleosome, where repair was almost completely inhibited. Previous studies concerning enzymatic repair of DNA in nucleosomes, focusing on processes such as glycosylase-mediated removal of oxidative damage,[45] nucleotide excision repair and photolyase repair of cyclobutene pyrimidine dimers [8] have also suggested that outward-facing modifications are more easily repaired than inward facing ones, but that repair was generally inhibited on damage positioned at the dyad axis. We also found repair was inhibited when the damage was facing inwards toward the histones. However, outward facing damage was repaired at the same rate or even slightly faster (position 1) than in free DNA (Table 1). This may be explained by the presence of histones, as in a previous study, increased repair activity of hAGT was enhanced in the presence of histones.[46] Furthermore, in a study testing repair of oxidative damage in nucleosomes by glycosylase enzymes, it was found that uracil-DNA glycosylase (UDG) and 3-alkyladenine DNA glycosylase (AAG) were equally active in excising the damaged base in nucleosomal and free DNA when the damage was facing outwards. However, excision was inhibited at inward facing positions. DNA-formamidopyrimidine glycosylase (Fpg) and human 8-oxoguanine DNA *N*-glycosylase 1 (hOGG1) were significantly inhibited independent of the positioning of adducts in the nucleosome. UDG and AAG are both small enzymes around 25 kDa, whereas Fpg and hOGG1 are 30 kDa and 38 kDa, suggesting that nucleosome structures may hinder accessibility particularly for larger enzymes.[45]

It was suggested that bending of DNA during repair reduces repair efficiency. For example, crystal structures suggest that hAAG and UDG, bend their DNA substrates at angles of 22° and 45°, while Fpg, and hOGG1, which show little or no reactivity on NCPs, bend their substrates more severely, at angles of 66°, and 70°, respectively.[45] The greater the DNA bending angle the more contacts between the DNA backbone and the histone core must be disrupted, which may make it energetically costly for these enzymes to act on NCPs, thus, may require chromatin remodeling factors to expose packed DNA for repair. With 21 kDa, hAGT is smaller than all the glycosylases tested, and for hAGT repair, only a ∼30° bending is required.[17] These DNA structural changes seem suitable to help flip the damaged nucleotide out from the major groove and to allow hAGT repair.[17]

These data further suggest that the formation of a cooperative repair complex involving multiple hAGT enzymes is unlikely when acting on DNA damage at the nucleosome, since the enzymes can only access the DNA from one side due to the nucleosome barrier. Therefore, we suggest that an hAGT monomer repairs *O*^{6}-MeG within a DNA nucleosome. Possibly, hAGT can disrupt enough histone-DNA contacts to adopt an active configuration when binding to the outer surface of nucleosomal DNA, but not when presented with inward-facing damage where DNA is constrained by multiple contacts. Nonetheless, chromosomal DNA damage repair in the cell differs from the model system investigated in this study. The DNA sequence used here binds more strongly to the histone core than other sequences, and additional factors in the nucleus, such as chromatin remodeling factors or decondensation occurring upon DNA damage, which are necessary to expose packaged DNA to repair, were not considered.

hAGT repair of O^6-CMG. O^6-CMG was not repaired by hAGT, even at large excess of enzyme (25 fold). This is in line with early studies with isolated AGT enzymes from E.Coli and O^6-CMG where no repair was detected, and where the lack of repair was hypothesized to be the due to the bacterial source of AGT.[14] In a later study, however, O^6-CMG was found to be repaired by human AGT at almost the same rate as O^6-MeG, compared to bacterial AGT.[13] Although we tried to mimic conditions used in that study, we did not detect any repair for O^6-CMG, while O^6-MeG was very efficiently repaired by hAGT. While *in vitro* studies with isolated enzymes found contradictory results, cell studies indicate that AGT is involved in repair of O^6-CMG, as inhibition by O^6-benzylguanine or cells low in AGT lead to increased O^6-CMG levels.[16,22] Recently, we found that the combination of hAGT with nucleotide excision repair (NER) have a synergistic effect on O^6-CMG removal.[15] That AGT may also be involved in other pathways than just repair, is also supported by the fact that multiple protein interaction partners have been identified for hAGT, suggesting a possible signaling role for hAGT to lead recruitment of protein factors to the DNA.[46]

We demonstrate here for the first time that hAGT repairs O^6-MeG contained within the nucleosome. However, the rotational position affects repair and is repaired poorly when damage is facing inwards. Due to the tight binding of the DNA around the histone octamers, we speculate that hAGT may function in this context as a monomer vs. cooperative binding. In summary, using a nanopore, we were able to confirm and expand knowledge on mechanistic aspects of O^6-MeG repair by hAGT. Nanopore sequencing of DNA modification locations could be further used to test repair enzymatic processing of other types of DNA modification. Expanding this method to measure repair in any sequence context should be possible if enough coverage of the tested DNA strands can be reached and if training data can easily be obtained e.g. by insertion of modified bases via PCR.

2.5 Materials & Methods

Chemical Reagents and Materials: 2'-Deoxy-N2-isobutyrylguanosine was purchased from Carbosynth (Compton, UK). 4,4'-Dimethoxytrityl chloride from LinkTechnologies Ltd. (Lanarkshire, Scotland) for O^6-CMG phosphoramidite was prepared following a previously reported method.[35,36] Reagents for solid-phase DNA synthesis were obtained from Glen Research Co. (Sterling, VA) and Link Technologies Ltd. (Lanarkshire, Scotland). PCR & DNA Cleanup Kit (5 µg), Nb. BbvCI, T4 DNA ligase, NotI-HF, T4-polynucleotide kinase was purchased from New England Biolabs (Ipswich, MA, USA). Unmodified and methylated oligonucleotides, (list in supplementary information) were ordered high-performance liquid chromatography (HPLC)-pure from eurogentec (Seraing, Belgium). 0.45-micron syringe filter and sybr gold nucleic acid stain was obtained from Thermo Fisher (Waltham, US). Recombinant Human MGMT protein (ab79251) was purchased from Abcam (Cambridge, UK). Tris-HCl pH 8, DTT (TRIS-HCl (pH 7.0 at 25°C), protein thermal shift buffer, dithiothreitol (DTT) were all purchased from Invitrogen (Switzerland) and ProNex® Size-Selective Purification System, QuantiFluor® ONE dsDNA System from Promega (Madison, USA). DNA degradase plus was acquired from Lucerna Chem (Luzern, Switzerland) and Centrifugal 10kD filters and kanamycin sulphate from VWR (Switzerland). Flongle cells, native barcoding expansion 1-12 kit (EXP-NBD104), ligation sequencing kit (SQK-LSK109) and flow cell priming kit (EXP-FLP002) was obtained from Oxford Nanopore (Oxford, UK).

Oligonucleotide Synthesis. The O^6-CMG was prepared by copper-catalyzed alkylation with ethyldiazoacetate[35], allowing to introduce a carboxyethylester moiety at the O^6 position of guanosine. The

modification was incorporated into oligonucleotides via solid-phase synthesis on a Mermade 4 DNA synthesizer from Bio Automation Corporation (Plano, TX) using base-labile phosphoramidites (dmf-dG-CE, Ac-dC-CE, dT-CE and dA-CE). After DNA synthesis the O^6-carboxymethylester containing sequences were deprotected by NaOH and ammonia, as described previously by Millington et. al.[36] to yield the O^6-CMG modification. Subsequently, the template was purified by reverse phase high- performance liquid chromatography (RP-HPLC) on a Phenomenex Luna® (5 µm, C18, 100 Å) with a linear gradient from 10-13 % (v/v) acetonitrile in 50 mM triethylammonium acetate over 25 min. The mass of the collected compound was determined by mass spectrometry (MS; Thermo Scientific LGT Velos), operated in negative ion mode (Figure S1-4). The single strand DNA concentration was determined by monitoring UV absorption at 260 nm on a NanoDrop 1000 spectrophotometer (Thermofisher; Massachusetts, USA).

Plasmid Expression. DH5a *E. Coli* transformed with pEGFP-W plasmid was incubated for 16 hours at 37° C at 400 rpm in 40 mL sterile LB media (10 g tryptone, 10 g NaCl, 5 g Yeast extract per L). Addition of kanamycin allowed for selective cultivation of the desired *E.coli* strains containing pEGFP-W plasmids. After incubation, *E. Coli* was transferred to 50 mL falcon tube and the plasmid was extracted using GenElute Plasmid DNA midiprep Kit following the manufacturer's protocol. Purity was checked by nanodrop. The extracted plasmid was visualized with agarose gel electrophoresis (0.8%, 80V, 3h, GelRed staining) either supercoiled, or linearized with NOTI or HindIII to confirm the correct length. Concentration was measured with Quantus.

Insertion of modified oligonucleotide into a plasmid. The DNA damage of interest was inserted at position G752 of the pEGFP-W shuttle vector. The modification spot is flanked by two sequences that can be recognized and nicked by Nb. BbvCI (NEB) allowing the excision of the 20mer sequence GCCCGCCCGGGACTGCCTCA in one step. In short, 20 ug of pEGFP-W was digested with Nb. BbvCI at 37°C for 1 hour. Then, 500 pmol of 20mer complementary strand were added and heated to 80°C and gradually cooled down to rt. The plasmid was purified using Monarch kit according to manufacturer's instructions. Following, 500 pmol of 20mer insert strand with DNA modification was added and heated to 80°C and gradually cooled down to rt. The two nicks were ligated by T4 DNA ligase at rt overnight. T4 ligase was deactivated at 65°C for 10 min and the plasmid was linearized using the restriction enzyme NOTI for 1 hour at 37°C. After inactivation at 65°C for 20 min the plasmid was purified with monarch kit and purity and concentration was checked via nanodrop. Automated gel electrophoresis was used to check the length of the plasmid.

Synthesis of modified nucleosome positioning sequence. Complementary, overlapping modified or unmodified 5' phosphorylated sequences (7 sequences) were pooled and annealed at equimolar amount to get the full 601 sequence (147bp) with modifications at specific spots. (list with oligonucleotides in Table S1). The pooled oligonucleotides were heated at 95°C for 5min and then cooled down to rt for annealing and ligated by T4-ligase for four hours. The strands were purified using pronex beads (3X), to enrich for sequences above 100bp. The concentration was measured on quantus. Length of the sequence was checked by agarose gel (2%, 80V, 100min, GelRed staining) or automated gel electrophoresis (Figure S5) and ligation was checked with denaturing 20% PAGE gel (270V, 3h, SybrGold staining). The same method was used for producing a DNA template to measure directionality of repair by ligating seven oligonucleotides, of which three oligonucleotides were modified with position 1 (Table S1).

Recombinant histone expression and purification. Recombinant human histones were expressed and purified as previously described.[1] Briefly, H2A, H2B, H3.1, and H4 were expressed in *E. coli* BL21(DE3) cells at 37 °C until reaching an OD_{600} of 0.6-0.8, before induction with 500 μM IPTG for 4 hours at 37 °C. Bacteria were then harvested by centrifugation at 6000 x g for 25 minutes. Cell pellets were resuspended in lysis buffer (200 mM NaCl, 20 mM Tris-HCl pH 7.6, 1 mM EDTA, 1 mM β-mercaptoethanol) and then lysed via sonication. Lysate was cleared by centrifugation at 30,000 x g for 20 minutes and the remaining insoluble pellet was resolubilized in extraction buffer (6 M guanidine HCl, 1 mM DTT, 1x PBS, pH 7) and nutated at 4 °C overnight. The extraction was cleared by centrifugation at 30,000 x g for 40 minutes and filtered using a 0.45-micron syringe filter. Filtered extraction was diluted 1:1 with HPLC buffer A (0.1% trifluoroacetic acid (TFA) in water) before purification with reverse-phase HPLC on a 30-70% buffer B (0.1% TFA in 90% acetonitrile and 10% water). Absorbance at 280 nm was used to observe desired peaks and fractions were collected using an automated fraction collector. Purity of fractions was determined using a quadrupole LC-MS. Fractions deemed pure were combined, lyophilized and stored at -80 °C.

Octamer formation. Lyophilized histones were dissolved in unfolding buffer (20 mM Tris-HCl pH 7.6, 6 M guanidine HCl, 0.5 mM EDTA, 1 mM DTT) and combined in a 1:1:0.95:0.95 molar ratio (H2A:H2B:H3.1:H4). Combined histones were diluted to a final concentration of 1 mg/mL with unfolding 3 x 1 L refolding buffer (20 mM Tris (pH 7.6), 2 M NaCl, 0.5 mM EDTA, and 1 mM DTT) at 4 °C for at least 4 hours each step. The assembled octamers were purified by size-exclusion chromatography using a GE Superdex 200 increase 10/300 GL column. Fractions were pooled, concentrated, diluted in 50% glycerol, and stored at -20 °C.

Nucleosome formation. Nucleosomes were assembled by salt gradient dialysis. Purified octamers were mixed with unmodified or modified Widom-601 DNA (synthesis described above) at molar ratios of 1.2, 1.24, and 1.28 in buffer (2 mM NaCl, 5 mM Tris-HCl pH 7.6, 0.5 mM EDTA, 0.5 mM DTT) in 20 μL volumes. Mixtures were pipetted into Slide-A-Lyzer MINI dialysis 3.5 kDa molecular weight cutoff devices (ThermoFisher Scientific) and dialyzed against 200 mL initializing buffer (10 mM Tris-HCl pH 7.6, 1.4 M NaCl, 0.1 mM EDTA, and 1 mM DTT) for 1 hour at 4 °C. Afterwards, mixtures continued to dialyze at 4 °C while 350 mL of dilution buffer (10 mM Tris (pH 7.6), 10 mM NaCl, 0.1 mM EDTA, 1 mM DTT) was added at a rate of 1 mL/min using a peristaltic pump. Assembly reactions were next transferred for another dialysis step to an additional 350 mL of dilution buffer for 6 hrs. Assembly reactions were moved to a final dialysis step in 300 mL of dilution buffer for 1-2 hrs at 4 ºC. Assembled nucleosomes were concentrated using a 30 kDa molecular weight cutoff centrifugal filter device and formation was analyzed by native gel electrophoresis (5% acrylamide gel, 0.5 X TBE, 120 V, 45 minutes) using SYBR Gold nucleic acid stain.

Nucleosome differential scanning fluorimetry (DSF) assays. Nucleosome stability was measured using a Protein Thermal Shift kit (Applied Biosystems) with modifications. The assay was conducted at 10 μL volumes with 5X SYPRO Orange dye (Invitrogen), 1 μL Protein Thermal Shift Buffer (Applied Biosystems) and nucleosome dilution buffer (10 mM Tris-HCl pH 7.6, 10 mM NaCl, 0.1 mM EDTA, and 1 mM DTT) in 384-well plates. Final nucleosome concentrations used were approximately 20 ng/μL of DNA. Fluorescence melt curve data was acquired using a QuantStudio 5 Real-Time PCR System (Applied Biosystems) with the following method: initial ramp rate of 1.6 °C/s to 25 °C with a 5-minute hold time at 25 °C, followed by a second ramp rate of 0.05 °C/s to 99.9 °C with a 2-minute hold time at 99.9 °C. Melting temperatures were calculated using the Protein Thermal Shift software (Applied Biosystems).

AGT Repair Experiments. The modified oligonucleotides and their complement strand were annealed in water at final concentrations of 10 nM (2 pmol), by heating for 3 min at 90°C followed by cooling to rt. Recombinant hMGMT (250 nM) was incubated with double stranded DNA substrates at 37 °C with reaction buffer (50mM Tris-HCl pH 8.3, 0.1mM EDTA, 0.5mM DTT, 50mM NaCl, 0.5mg/mL BSA). A mastermix was prepared with sufficient material for four time points. Once MGMT was added to the mixture (t = 0 s), 200 µL aliquots were removed (containing 50 pmol hAGT and 2 pmol of fully formed duplex) and added to 200 uL 0.2% TFA to stop the reaction. The mixture was neutralized by adding 95 uL NaOH (50 mM) and purified with pronex beads (3X). Repair reactions were performed in triplicate.

Analysis of DNA alkylation levels by Mass Spectrometry. Purified samples were digested with DNA degradase plus for 2 h at 37°C and 300 rpm. Then, samples were filtered through 10K MWCO spin filters (VWR, 516-0229, 10 min, 16.000xg), filter membranes were washed once (300 µL Milli Q water), and filtrates were combined by centrifugation (10 min, 16.000 xg). The eluent samples were vacuum-centrifuged to dryness in HPLC glass inserts and reconstituted with 10 µL water and sonicated for LC-MS/MS analysis. Quantification of O^6-CMdG and O^6-MedG was achieved by LC–MS/MS using a nanoAcquity UPLC system (Waters) HPLC and tandem quadrupole mass spectrometer (Thermo TSQ Vantage) with an electrospray ionization source (ESI). Liquid chromatography was performed with a Luna Omega PS C18 column (100 Å, 3 µm, 150 µm x 0.5 mm, Phenomenex) at a column temperature of 40 °C. The flow rate was 10 µL/min with the mobile phases A and B consisted of MS grade H2O with 0.1% FA and ACN with 0.1% FA, respectively (Biosolve Chimie, Dieuze, France). 2 µL samples were injected and the chromatographic separation was achieved by a 1 min hold at 0% B, followed by a linear gradient from 0% to 10% B over 9 min, following a linear gradient from 10% to 25% B over 5 min, and finally a linear gradient from 25% to 90% B over 5 min at a flow rate of 10 µL/min. The mass spectra were recorded in positive ionisation SRM mode with following parameters: capillary temperature, 250 °C; spray voltage, 3000 V; sheath gas pressure, 30; ion sweep gas pressure, 0; aux gas pressure, 5; collision gas, argon; scan width, m/z 0.1; scan time, 0.5 s. The transitions used were m/z 326.1096 à 210.0622 for O^6-CMdG nucleoside and m/z 282.1995 à 166.0722 for O^6-MedG nucleoside, both with Collision Energy 16V. Xcalibur software (Thermo) was used for data acquisition and processing.

DNA Sequencing and Data Acquisition. 200-250 ng of damage-containing plasmids or 20-30ng of damage-containing 601 sequence were used as the input for preparing the DNA sequencing library with the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies) and native barcoding expansion 1-12 kit (EXP-NBD104) following the native barcoding genomic DNA Protocol (NBE_9065_V109_revAB_ 14Aug2019). The constructed libraries, 20-40ng of plasmid or 2-4ng of pooled 601 sequence, were loaded onto the R9.4.1/FLO-FLG001 flow cell equipped on the MinION Mk1B (ONT). The sequencing and data acquisition were performed under MinKNOW software v18.12.6 (ONT) to generate a fast5 file containing raw ionic signal for individual molecules.

Data Analysis for Damage Localization. The raw data (fast5) was resquiggled using the analysis tool Tombo. To quantify the amount of modification at a specific position the detect_modification was applied on non-modified and modified sample and modification percentage was given with the text_output –fractions option, which was imported into integrative genomics viewer (IGV). To build a more accurate model, tailored to our needs, raw current data was extracted from re-squiggled signals and imported in R. We used k-means clustering over the modified and neighbor bases, using the “kmeans” function (k=2) in

the cluster package in R version 4.2.1 to remove noise and signals originating from unmodified strands. Only the cluster-group, where the signals did not overlap with the unmodified signals, were used further for model building. A two-way classifier was built using LR model with generalized linear function model (glm.fit), with family set as binomial and a three-way classifier was built using a neural network model (NNM) with nnet and NeuralNetTools package in R (size = 10). Three independent variables were used as input; current data from modified base and from the neighboring bases up- and downstream. To train and test the model, unmodified and extracted modified data was randomly split in 70% training and 30% test data. In R, the glm.pred function was used to predict the probability that each subject would have the outcome and ypred was used to calculate accuracy for NNM. If the probability was above a threshold 0.5, the subject was classified as "modified". Model performance was tested with the test set (n=1260) and evaluated based on the accuracy, sensitivity, and specificity of the predictions, described in terms of true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP). Sensitivity was defined as the proportion of actual positives and identified as such (TP/(TP + FN)). Specificity was defined as the proportion of actual negatives and identified as such (TN/(TN + FP)). Accuracy was defined as the proportion of TP and TN in all assessments and identified as (TP + TN)/(TP + TN + FP + FN). An overview of the analysis can be found in Figure S7.

Limit of detection with computationally simulated mixtures. The limit of detection was assessed by comparing in silico mixtures (i.e., by mixing reads for damage-containing plasmids with those of the control) at different ratios of DNA damage in mixture with control reads. The analysis was performed based on 5,000 reads per mixture. We were able to detect damage as low as 5% for both damage and models. (Figure S10)

2.6 References

1. Fu, D., Calvo, J.A. & Samson, L.D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **12**, 104-20 (2012).
2. Drablos, F. *et al.* Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)* **3**, 1389-407 (2004).
3. Tricker, A.R. & Preussmann, R. Carcinogenic N-nitrosamines in the diet: occurrence, formation, mechanisms and carcinogenic potential. *Mutat Res* **259**, 277-89 (1991).
4. Lewin, M.H. *et al.* Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res* **66**, 1859-65 (2006).
5. Suquet, C. & Smerdon, M.J. UV damage to DNA strongly influences its rotational setting on the histone surface of reconstituted nucleosomes. *J Biol Chem* **268**, 23755-7 (1993).
6. Norabuena, E.M. *et al.* Effect of the Spiroiminodihydantoin Lesion on Nucleosome Stability and Positioning. *Biochemistry* **55**, 2411-21 (2016).
7. Olmon, E.D. & Delaney, S. Differential Ability of Five DNA Glycosylases to Recognize and Repair Damage on Nucleosomal DNA. *ACS Chem Biol* **12**, 692-701 (2017).
8. Guintini, L., Charton, R., Peyresaubes, F., Thoma, F. & Conconi, A. Nucleosome positioning, nucleotide excision repair and photoreactivation in Saccharomyces cerevisiae. *DNA Repair (Amst)* **36**, 98-104 (2015).
9. Bilotti, K., Kennedy, E.E., Li, C. & Delaney, S. Human OGG1 activity in nucleosomes is facilitated by transient unwrapping of DNA and is influenced by the local histone environment. *DNA Repair (Amst)* **59**, 1-8 (2017).

10. Kosmoski, J.V., Ackerman, E.J. & Smerdon, M.J. DNA repair of a single UV photoproduct in a designed nucleosome. *Proc Natl Acad Sci U S A* **98**, 10113-8 (2001).
11. Pegg, A.E. Repair of O(6)-alkylguanine by alkyltransferases. *Mutat Res* **462**, 83-100 (2000).
12. Zang, H., Fang, Q., Pegg, A.E. & Guengerich, F.P. Kinetic analysis of steps in the repair of damaged DNA by human O6-alkylguanine-DNA alkyltransferase. *J Biol Chem* **280**, 30873-81 (2005).
13. Senthong, P. *et al.* The nitrosated bile acid DNA lesion O6-carboxymethylguanine is a substrate for the human DNA repair protein O6-methylguanine-DNA methyltransferase. *Nucleic Acids Res* **41**, 3047-55 (2013).
14. Shuker, D.E. & Margison, G.P. Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Res* **57**, 366-9 (1997).
15. Aloisi, C.M.N. *et al.* A combination of direct reversion and nucleotide excision repair counters the mutagenic effects of DNA carboxymethylation. *DNA Repair (Amst)* **110**, 103262 (2022).
16. Kostka, T. *et al.* Repair of O6-carboxymethylguanine adducts by O6-methylguanine-DNA methyltransferase in human colon epithelial cells. *Carcinogenesis* **42**, 1110-1118 (2021).
17. Kono, S. *et al.* Resolving the subtle details of human DNA alkyltransferase lesion search and repair mechanism by single-molecule studies. *Proc Natl Acad Sci U S A* **119**, e2116218119 (2022).
18. Daniels, D.S. *et al.* DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat Struct Mol Biol* **11**, 714-20 (2004).
19. Melikishvili, M. & Fried, M.G. Quaternary interactions and supercoiling modulate the cooperative DNA binding of AGT. *Nucleic Acids Res* **45**, 7226-7236 (2017).
20. Odell, I.D., Wallace, S.S. & Pederson, D.S. Rules of engagement for base excision repair in chromatin. *J Cell Physiol* **228**, 258-66 (2013).
21. Bilotti, K., Tarantino, M.E. & Delaney, S. Human Oxoguanine Glycosylase 1 Removes Solution Accessible 8-Oxo-7,8-dihydroguanine Lesions from Globally Substituted Nucleosomes Except in the Dyad Region. *Biochemistry* **57**, 1436-1439 (2018).
22. Geisen, S.M. *et al.* Direct Alkylation of Deoxyguanosine by Azaserine Leads to O(6)-Carboxymethyldeoxyguanosine. *Chem Res Toxicol* **34**, 1518-1529 (2021).
23. Trantakis, I.A., Nilforoushan, A., Dahlmann, H.A., Stauble, C.K. & Sturla, S.J. In-Gene Quantification of O(6)-Methylguanine with Elongated Nucleoside Analogues on Gold Nanoprobes. *J Am Chem Soc* **138**, 8497-504 (2016).
24. Aloisi, C.M.N., Nilforoushan, A., Ziegler, N. & Sturla, S.J. Sequence-Specific Quantitation of Mutagenic DNA Damage via Polymerase Amplification with an Artificial Nucleotide. *J Am Chem Soc* **142**, 6962-6969 (2020).
25. Wang, Y. *et al.* Nanopore Sequencing Accurately Identifies the Mutagenic DNA Lesion O(6) - Carboxymethyl Guanine and Reveals Its Behavior in Replication. *Angew Chem Int Ed Engl* **58**, 8432-8436 (2019).
26. Nawy, T. Nanopores and the helicase two-step. *Nat Methods* **12**, 1119 (2015).
27. Schibel, A.E. *et al.* Nanopore detection of 8-oxo-7,8-dihydro-2'-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J Am Chem Soc* **132**, 17992-5 (2010).
28. Ma, F. *et al.* Nanopore Sequencing Accurately Identifies the Cisplatin Adduct on DNA. *ACS Sens* **6**, 3082-3092 (2021).
29. Perera, R.T., Fleming, A.M., Johnson, R.P., Burrows, C.J. & White, H.S. Detection of benzo[a]pyrene-guanine adducts in single-stranded DNA using the alpha-hemolysin nanopore. *Nanotechnology* **26**, 074002 (2015).
30. An, N., Fleming, A.M., White, H.S. & Burrows, C.J. Crown ether-electrolyte interactions permit nanopore detection of individual DNA abasic sites in single molecules. *Proc Natl Acad Sci U S A* **109**, 11504-9 (2012).

31. Rang, F.J., Kloosterman, W.P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* **19**, 90 (2018).
32. Georgieva, D., Liu, Q., Wang, K. & Egli, D. Detection of base analogs incorporated during DNA replication by nanopore sequencing. *Nucleic Acids Res* **48**, e88 (2020).
33. Nookaew, I. *et al.* Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing. *Chem Res Toxicol* **33**, 2944-2952 (2020).
34. Zhang, J. *et al.* Discrimination between Different DNA Lesions by Monitoring Single-Molecule Polymerase Stalling Kinetics during Nanopore Sequencing. *Nano Lett* **22**, 5561-5569 (2022).
35. Geigle, S.N., Wyss, L.A., Sturla, S.J. & Gillingham, D.G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem Sci* **8**, 499-506 (2017).
36. Millington, C.L. *et al.* Convenient and efficient syntheses of oligodeoxyribonucleotides containing O(6)-(carboxymethyl)guanine and O(6)-(4-oxo-4-(3-pyridyl)butyl)guanine. *Nucleosides Nucleotides Nucleic Acids* **31**, 328-38 (2012).
37. Technologies, O.N. Tombo tool. Vol. 2023 (https://nanoporetech.github.io/tombo/, 2023).
38. Luzzietti, N., Knappe, S., Richter, I. & Seidel, R. Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat Protoc* **7**, 643-53 (2012).
39. McKeague, M., Otto, C., Raz, M.H., Angelov, T. & Sturla, S.J. The Base Pairing Partner Modulates Alkylguanine Alkyltransferase. *ACS Chem Biol* **13**, 2534-2541 (2018).
40. Fleming, A.M., Mathewson, N.J., Howpay Manage, S.A. & Burrows, C.J. Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. *ACS Cent Sci* **7**, 1707-1717 (2021).
41. Wescoe, Z.L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* **136**, 16582-7 (2014).
42. Neumann, D., Reddy, A.S.N. & Ben-Hur, A. RODAN: a fully convolutional architecture for basecalling nanopore RNA sequencing data. *BMC Bioinformatics* **23**, 142 (2022).
43. Raz, M.H., Sandell, E.S., Patil, K.M., Gillingham, D.G. & Sturla, S.J. High Sensitivity of Human Translesion DNA Synthesis Polymerase kappa to Variation in O(6)-Carboxymethylguanine Structures. *ACS Chem Biol* **14**, 214-222 (2019).
44. Pai, S.S., Ranjan, S., Mathew, A.R., Anindya, R. & Meur, G. Analysis of the long-read sequencing data using computational tools confirms the presence of 5-methylcytosine in the Saccharomyces cerevisiae genome. *Access Microbiol* **4**, acmi000363 (2022).
45. Olmon, E.D. & Delaney, S. Differential Ability of Five DNA Glycosylases to Recognize and Repair Damage on Nucleosomal DNA. *Acs Chemical Biology* **12**, 692-701 (2017).
46. Niture, S.K., Doneanu, C.E., Velu, C.S., Bailey, N.I. & Srivenugopal, K.S. Proteomic analysis of human O6-methylguanine-DNA methyltransferase by affinity chromatography and tandem mass spectrometry. *Biochem Biophys Res Commun* **337**, 1176-84 (2005).

Supporting Information – Chapter 2

Table 1

Oligonucleotides used in the study and 601 sequence with modifications.

Plasmid insert	
20mer fwd	P-GCCCGCCCGGGACTGCCTCA
20mer fwd Me /CMG	P-GCCCGCCCGGGACTGCCTCA
20mer rev	P-GCCCGCCCGGGACTGCCTCA
601 sequence	
601 fwd 1	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAA
601 fwd 2	P-CGCACGTACGCGCTGTCCC
601 fwd 3	P-CCGCGTTTTAACCGCC
601 fwd 4	P-AAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT
601 rev 1	ACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCG
601 rev 2	P-CGTACGTGCGTTTAAGCG
601 rev 3	P-GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAG
601 fwd 2 Me / CMG	P-CGCACGTACGCGCTGTCCC
601 fwd 3 Me / CMG	P-CCGCGTTTTAACCGCC
601 rev 2 Me / CMG	P-CGTACGTGCGTTTAAGCG

601 sequence ligated with modifications Pos 1, Pos 2, Pos 3

NCP position 1 (+2) and position 2 (+11) from the central base of the DNA forward strand and position 3 (+8) from central base in reverse DNA strand.

5′ GGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCC
3′ CCACGGCTCCGGCGAGTTAACCAGCATCTGTCGAGATCGTGGCGAATTTGCGTGCATGCGCGACAGGGGGCGCAAAATTGGCGGTTCCCCTAATGAGGGATCAGAGG

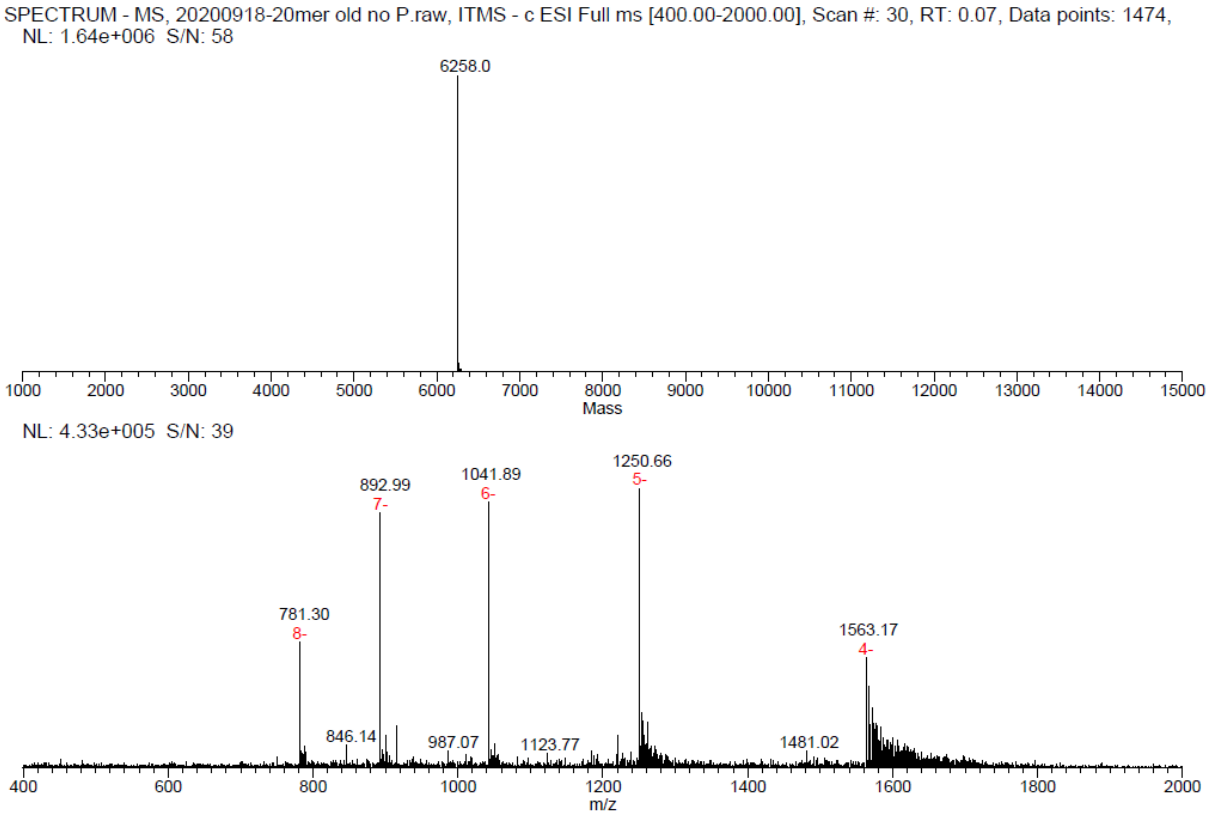


Figure S1, Mass spectrometry characterization and Magtran deconvolution of the *O*^{6}-CMG containing 20mer (parent mass 6259.3).

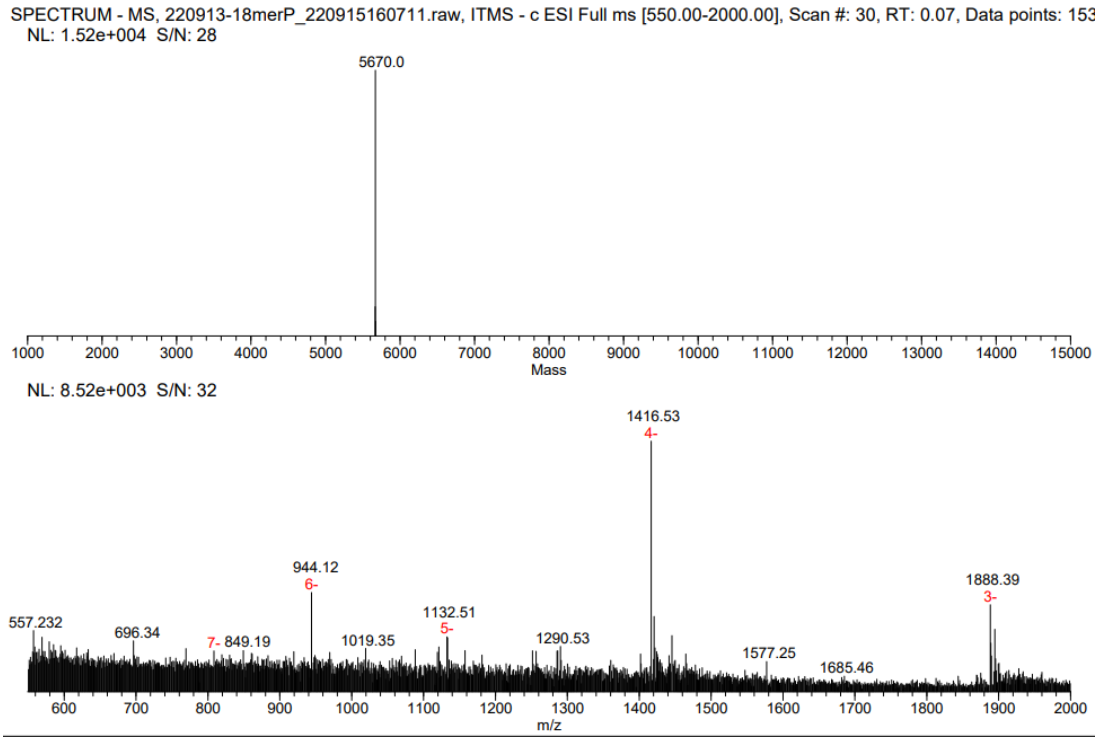


Figure S2, Mass spectrometry characterization and Magtran deconvolution of the phosphorylated *O*^{6}-CMG containing 18mer (parent mass 5669.5).

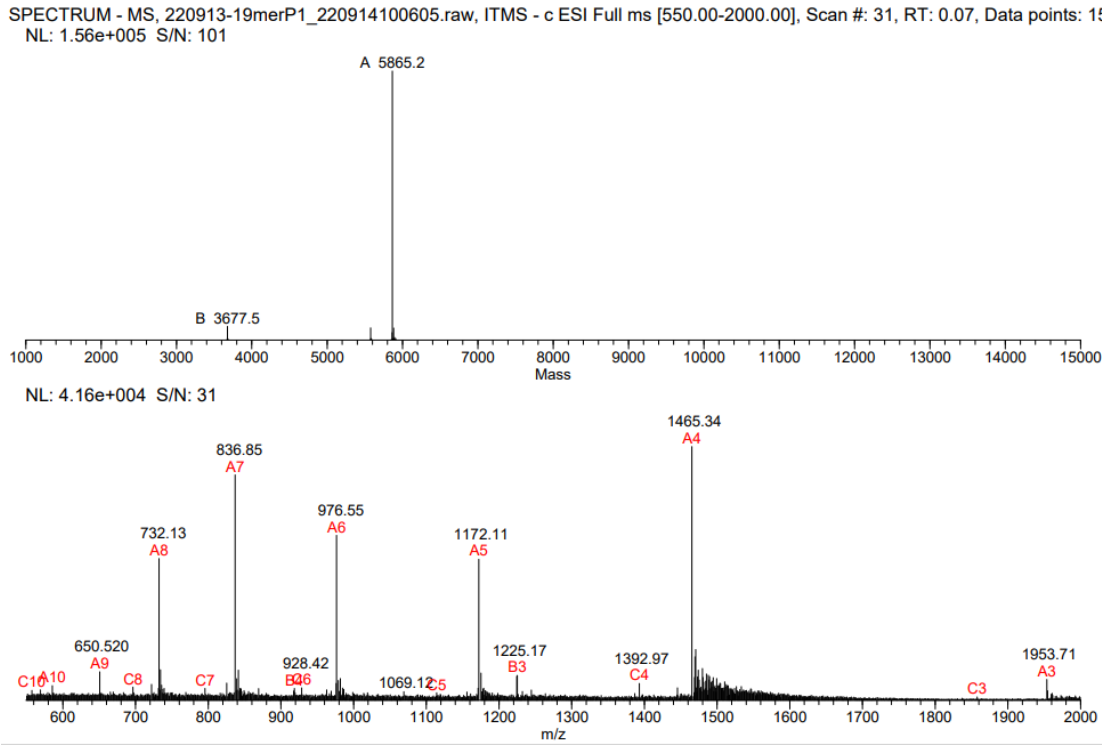


Figure S3, Mass spectrometry characterization and Magtran deconvolution of the phosphorylated O^6-CMG containing 19mer (parent mass 5864.6).

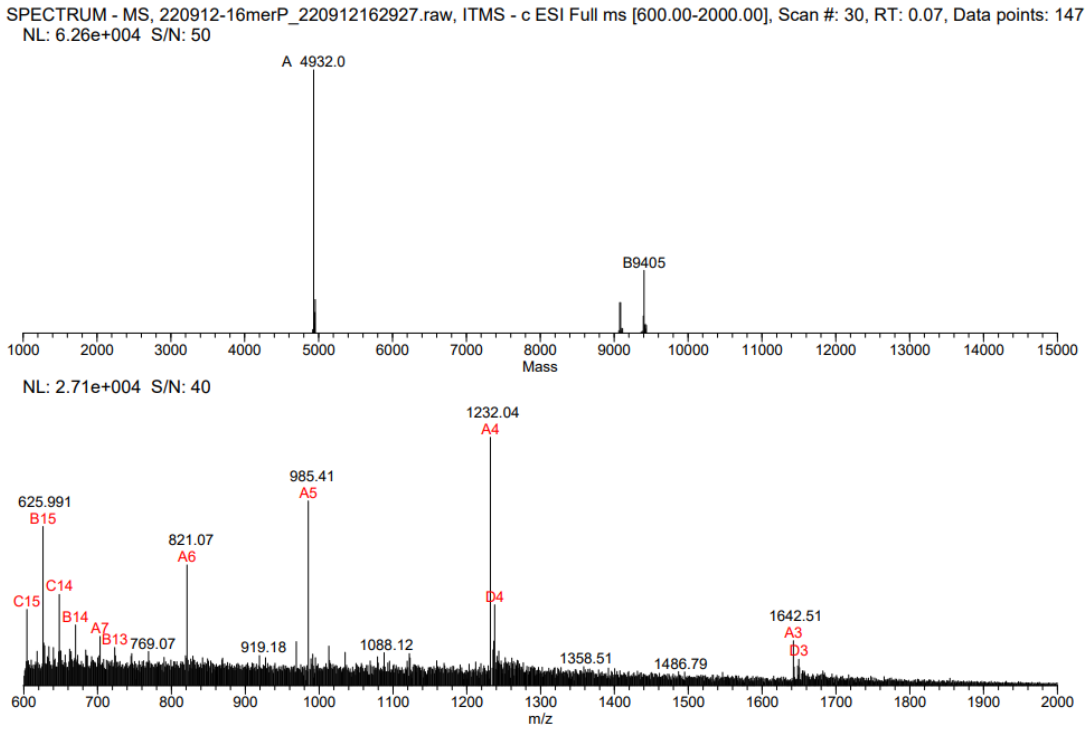


Figure S4, Mass spectrometry characterization and Magtran deconvolution of the phosphorylated O^6-CMG containing 18mer (parent mass 5669.5).

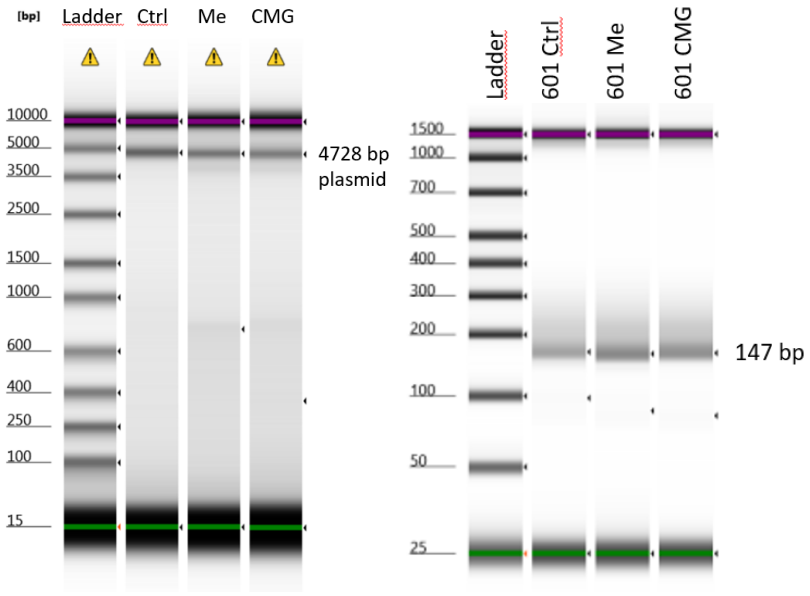


Figure S5. Tape station results of modified plasmid samples confirm length and purity of the plasmid (left). On the right are tape station results of 601 sequence (147bp) containing triple methylated, carboxymethylated or no damage (601 Ctrl).

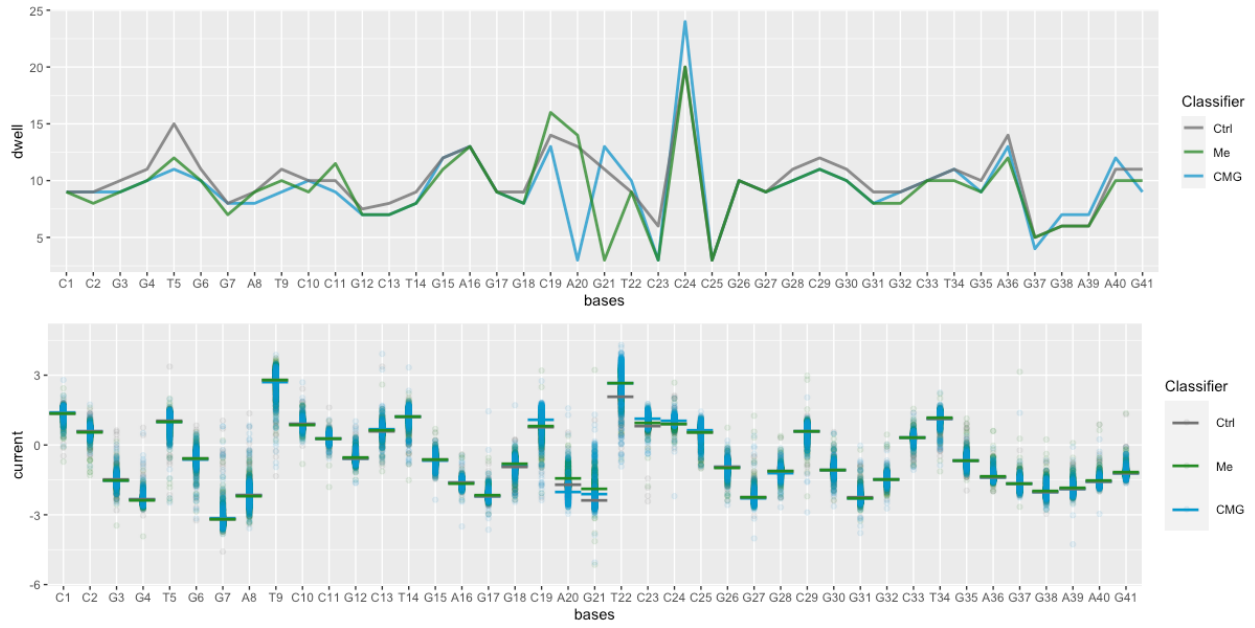


Figure S6. Kinetic (dwell) and current data obtained by ONT for the plasmid DNA. The modified base is at position G21. There was no significant deceleration at or after the modified base.

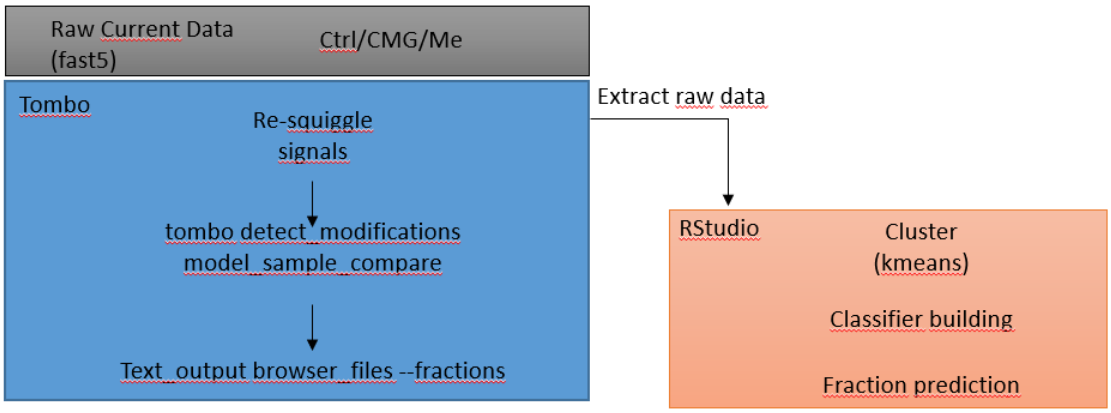


Figure S7. Workflow of data analysis and model building. Raw current data was re-squiggled using tombo tool. Re-squiggled raw data was either analyzed further with the inbuilt tombo fraction detection tool or extracted and clustered in R to prepare training and test data for model building.

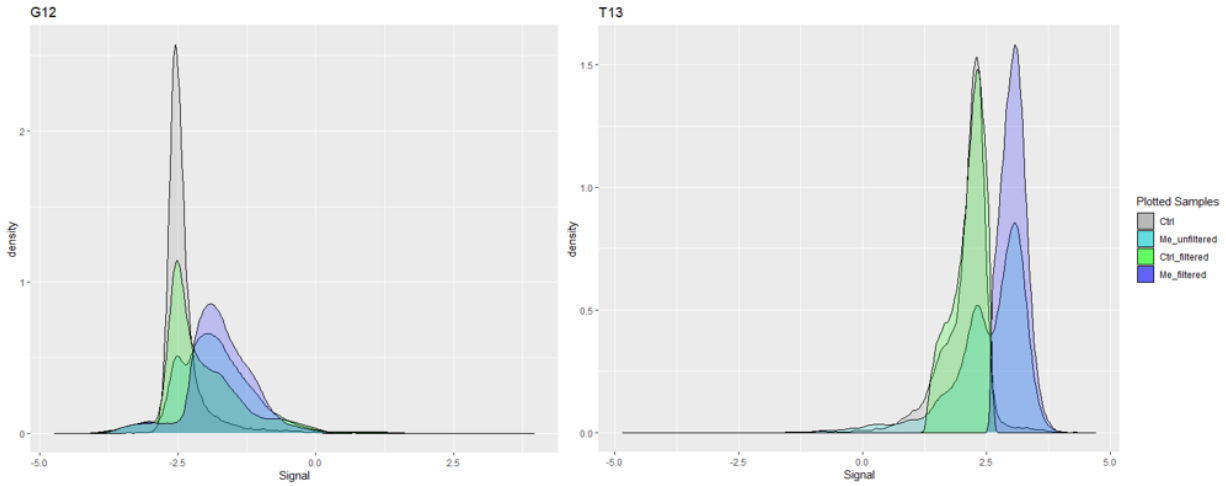


Figure S8. Current of O^6-MeG modified plasmid measured at modified (G12) and neighbor position (T13). The data obtained by nanopore "Me_unfiltered" (lighblue) was clustered into two groups – "Ctrl_filtered" (green) and "Me_filtered (darkblue"). "Ctrl_filtered" was overlapping with the current signals measured for unmodified sample (grey), while signals for "Me_filtered" presumably were induced by the O^6-MeG modification. "Ctrl" and "Me_filtered" were used as training and test data for the model.

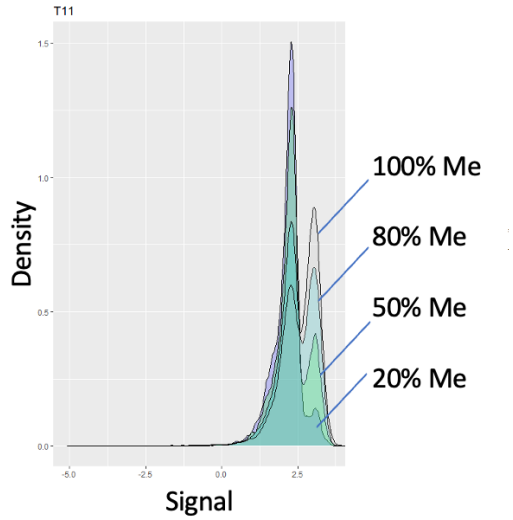


Figure S9. Various proportions of methylated and unmethylated plasmid was mixed and current was measured by ONT. The current signal measured at base next to modified *O*^{6}-MeG in the plasmid shifted toward the expected current of unmodified base with decreasing methylation ratio.

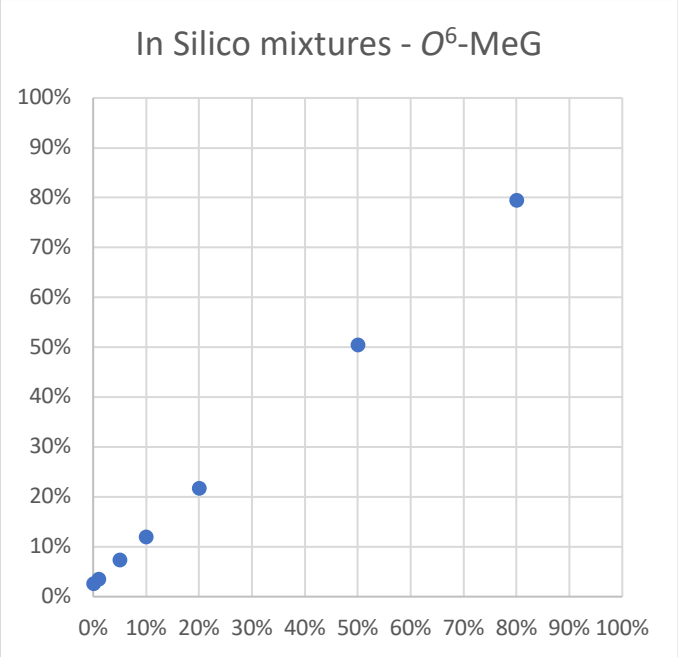


Figure S10 Accurate detection level of a DNA adduct was assessed by comparing in silico mixtures (i.e., by mixing reads for adduct-containing plasmids with those of the control) at different ratios of DNA adduct in mixture with control reads. Accuracy drops below 5%, as the model is overestimating the amount of modifications.

Table 2. LR Model performance for modified positions in 601 sequence.

	Pos 1 (Me)	Pos 2 (Me)	Pos 3 (Me)	Pos 1 (CMG)	Pos 2 (CMG)	Pos 3 (CMG)
Sensitivity	95.90%	89.80%	91.80%	97.1%	95.6%	95.7%
Specificity	97.70%	95.50%	94%	97.24%	98.0%	97.7%
Accuracy	96.90%	92.60%	92.90%	97.8%	96.8%	96.6%

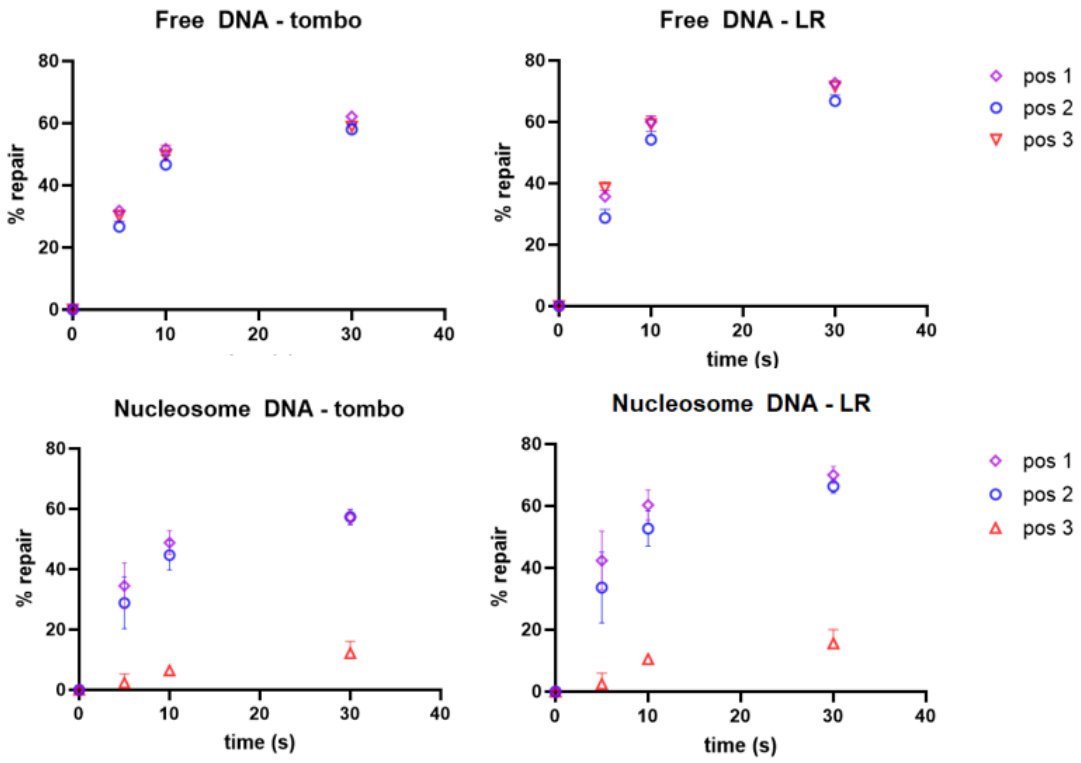


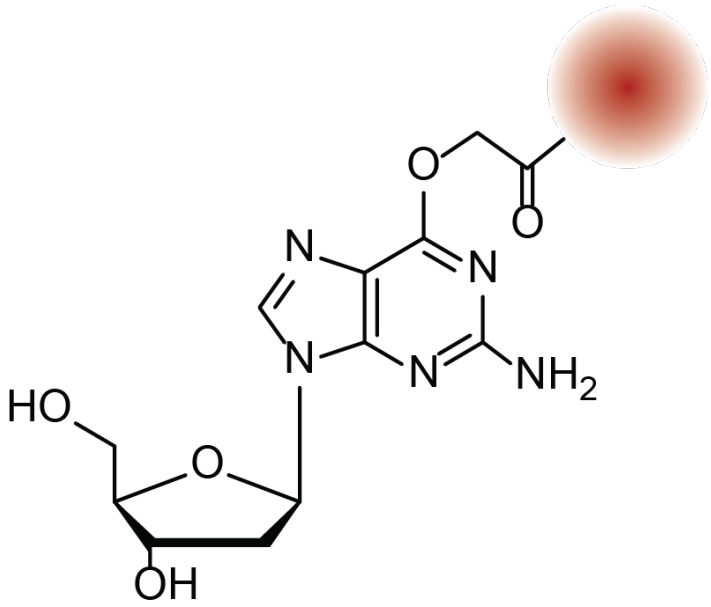
Figure S11. Damage quantification by tombo and LR for the repair experiments.

Table S3. Comparison for K_{obs} obtained by tombo or LR analysis.

		K_{obs}/s (LR)	K_{obs}/s (tombo)
Free DNA	Pos 1	0.145	0.152
	Pos 2	0.128	0.133
	Pos 3	0.158	0.153
Nucleosome	Pos 1	0.187	0.185
	Pos 2	0.144	0.141
	Pos 3	0.06	n.d.

.

Chapter 3 – Amine-fluoroclick labeling of carboxymethylated DNA adducts



Emma S. Sandell, M. Bless, V. Stäheli, R. Fernández Cereijo, S. Diedrich, Shana J. Sturla*, Amine-Fluoroclick labeling of carboxymethylated DNA adducts, *Manuscript in preparation*

3.1 Abstract

Exposure to genotoxins and endogenous processes lead to the formation of DNA adducts, which can lead to mutation and cancer formation. The DNA adduct O^6-CMG can be formed by *N*-nitroso compounds arising from red or processed meat consumption. O^6-CMG is mutagenic and suggested to be a molecular initiating event in colorectal carcinogenesis, thus, it has the potential to serve as a predictive biomarker for diagnosis or risk assessment of colorectal cancer. To better understand the relationship between elevated O^6-CMG and colorectal cancer, there is a need for reliable and efficient detection methods for O^6-CMG. In the present work, we developed a method that combines amine ligation and click-mediated fluorophore ligation to quantify O^6-CMG. This process entails reacting oligonucleotides containing O^6-CMG with amino-alkynes to yield alkyne modified DNA. Next, an azide-fluorophore is reacted via a copper-mediated click reaction resulting in a fluorophore attached to the original O^6-CMG residue. Having established a working method, we tested sensitivity and selectivity, by applying our labeling method on mixtures of synthesized O^6-CMG and the epigenetic 5caC containing oligonucleotides. We demonstrated that fluorescence strongly correlates with O^6-CMG levels and does not label 5caC. Finally, different ratios of carboxymethylated ctDNA were fluorescently labelled and a limit of detection of 16 O^6-CMG /10^6 nucleotides was determined by LC-MS. The labeling strategy presented here is a simple and selective approach to quantify O^6-CMG in a rapid and cost-efficient manner.

3.2 Introduction

DNA modifications occur physiologically, such as the epigenetic modifications 5meC or 5caC, or can be induced by genotoxic agents to form DNA damage. While epigenetic modifications are abundant and critical regulatory factors for gene expression[1], DNA damage can be detrimental to health by inducing mutations and promoting cancer development.[2] Because damaged DNA residues are often chemically stable and specific to genotoxin characteristics, they have the potential to be used as predictive biomarkers of disease and toxin exposure.[3,4] For example, elevated levels of adducts excreted in the urine of individuals exposed to aflatoxin were significantly correlated with an increased risk of liver cancer,[5] and acetaldehyde-induced DNA adducts measured in the blood have been linked to risk of esophageal cancer[6]. DNA adducts occur at very low levels, so detection methods typically rely heavily on mass spectrometry analysis. Mass spectrometry can provide high sensitivity and structural information about the adduct, but these methods require tedious sample preparation, expensive equipment, and are difficult to scale up. For this reason, we investigate an alternative approach to measure O^6-CMG.

The DNA adduct O^6-CMG has been identified in exfoliated colon cells as a consequence of increased fecal NOC arising from red meat in the colon,[7] and in DNA from the blood of people on a high meat diet.[8] There are endogenous repair systems to address DNA damage, and how O^6-CMG is repaired is still a matter of debate[9,10]. Cell studies suggest that alkylguanine transferase (AGT) is involved in the repair of O^6-CMG, since inhibition by O^6-benzylguanine or cells low in AGT lead to increased O^6-CMG levels.[11,12] Moreover, we have recently found that a combination of AGT and nucleotide excision repair has a synergistic effect on O^6-CMG removal.[13] Repair of this damaged residue is important because O^6-CMG has a strong miscoding potential during DNA synthesis, as DNA polymerases tend to misincorporate a thymine instead of the correct cytocine, resulting in GC → AT mutations.[14] The mutation spectra resulting from DNA synthesis templated by a plasmid exposed to the carboxymethylating agent KDA resembled the mutation

spectra found in colon and gastric cancers.[15] These data suggest the formation of O^6-CMG as a molecular initiating event in colorectal carcinogenesis, but it has not yet been characterized as a predictive biomarker for colorectal cancer due to the lack of suitable quantification methods.

Previous studies of the formation of O^6-CMG in DNA derived from human cells, such as from blood,[8] tissue biopsy,[16] and exfoliated colon cells[7] have relied on the use of immunoblotting.[8] For the immunoblotting, denatured DNA is trapped on nitrocellulose filter where it is incubated with an O^6-CMG primary antibody, labelled with a secondary antibody, and quantified by chemiluminescence. The limit of detection for this approach is 15 CMG per 10^8 bases or 0.4 fmol O^6-CMG. The sensitivity of this method is not always sufficient to measure O^6-CMG. While it was possible to detect O^6-CMG levels in DNA from the blood of three volunteers,[8] quantitation of O^6-CMG adducts in DNA extracted from exfoliated colon cells were below the level of detection.[7] Using cell counting techniques, a significant relationship was observed between the percentage of cells staining positive for O^6-CMG and the fecal level of NOC arising from red meat in the colon, however the method is not quantitative.[7] Widespread use of the immunoblotting method is limited by the fact that the antibody used in this technique is not commercially available, the detection limit is high, and the assay can only be used on intact DNA.

Although mass spectrometry (MS) is one of the most sensitive methods, it has not been successfully used to quantify O^6-CMG in samples derived from humans (except in one study of colon biopsy samples).[17] Most targeted mass spectrometry methods have quantified both O^6-MedG and O^6-CMdG using UHPLC-MS/MS[18], nLC-nESI-MS3[19] or nanoLC-ESI-HRMS[12]. These methods use internal standards for absolute quantification with lower detection limits down to the fmole-amole levels. These methods were successfully applied on DNA extracted from azaserine treated cells or KDA treated ctDNA. Furthermore, an LC-MS/MS method was developed to quantify O^6-CMG in human urine samples, but the levels in the urine were too low to be detected.[17] O^6-CMG was also detected using an HRMS methodology in an untargeted analysis of DNA adducts in colon biopsy samples[16] and in gastrointestinal digestion samples[20]. However, LC-MS analyses in general have tedious sample preparation procedures, require expensive instrumentation, and are difficult to scale up. Furthermore, these methods require internal standards for absolute quantification, which adds an additional barrier to LC-MS quantification of O^6-CMG.

DNA adducts have been detected and quantified at low levels using fluorescent labeling methods.[4,7,21] For instance, there are several biochemical labeling methods that rely on base excision repair enzymes to excise damaged bases and create AP sites. These AP sites can be labelled with specifically designed aminooxy-substituted rotor dyes [22] or spectrally distinct fluorescent nucleotides to allow simultaneous quantification of different lesions[23]. A similar method has been developed in our group for the quantification of 8-oxoG using a glycosylase to excise the damage which is replaced by an alkyl-modified base that can be fluorescently labelled via click-chemistry.[24,25] Chemical ligation reactions involving carboxyl-DNA modifications, such as O^6-CMG and epigenetic 5caC, with amine ligation have been reported previously[26,27] but to our knowledge, there are no examples of chemical ligation of low-abundance carboxymethyl-containing DNA adducts with a fluorophore, that avoids labelling the more abundant epigenetic marker 5caC.

In this work, we have developed a fluorescence-based ligation method for the selective detection of carboxymethylated adducts regardless of the presence of the epigenetic modification 5caC. The labeling

method is based on a combination of amine ligation and click reaction and was developed using O^6-CMG containing oligonucleotides. Optimization reactions of the amine ligation were performed with different temperatures and co-solvents on single- and double-stranded (ds) oligonucleotides. Selectivity was tested using different ratios of ds oligonucleotides containing O^6-CMG and 5caC, and sensitivity of the labeling method was tested on mixtures of ctDNA treated with potassium diazoacetate (KDA) and quantified by fluorescence and LC-MS. Finally, we aimed to validate the method using DNA extracted from azaserine-treated cells.

3.3 Results and Discussion

Fluorescent labeling of *O*^{6}-CMG using amine ligation and click chemistry. The labeling method was developed on a 9mer oligonucleotides (dTTTTG^{cm}TTTT) containing an *O*^{6}-CMG. *O*^{6}-CMG was induced, as previously reported, by a copper-catalyzed alkylation with ethyl diazoacetate, which selectively introduced a carboxyethyl ester at the *O*^{6}-position of guanine which was further hydrolyzed to obtain an *O*^{6}-CMG modified DNA strand.[28] Using this template, we explored the susceptibility of ligating a fluorophore to *O*^{6}-CMG with a bound primary amine. First, a direct labeling strategy (Figure 1, Strategy 1) of an amine-fluorophore to *O*^{6}-CMG was tested by activating the carboxyl group with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) but no conversion was observed in 16 h reaction time. Therefore, a two-step ligation approach (Figure 1, Strategy 2) was tested. This approach entailed primary installation of an alkyne group via amine ligation, which allows further attachment of any clickable moiety, in this case a fluorophore, via click chemistry. Our hypothesized rational behind this was that in the first strategy the fluorophore was too large and/or too close to the carboxyl group, thereby sterically hindering the possibility of amide formation, opposed to the incorporation of an alkyne amine that prevented this steric hinderance. To test our hypothesis, we chose propargylamine and 3-butynylamine and conversion of 42% and 30% were observed after 16 h at room temperature. (Table 1). The alkyne modified oligonucleotides (Figure 1) were purified by HPLC and the mass of the collected compound was checked by mass spectrometry (Figure S1, S2). Fluorescent azide AF-594 was then reacted with the alkyne-containing oligonucleotides by copper(I)-catalyzed cycloaddition. We observed 95%-100% conversion for *O*^{6}-CMG-butynyl and 0% conversion for *O*^{6}-CMG-propyne after 10 min reaction time at 37 °C. The mass of the AF-594 labelled oligonucleotide was confirmed by mass spectrometry (Figure S3.)

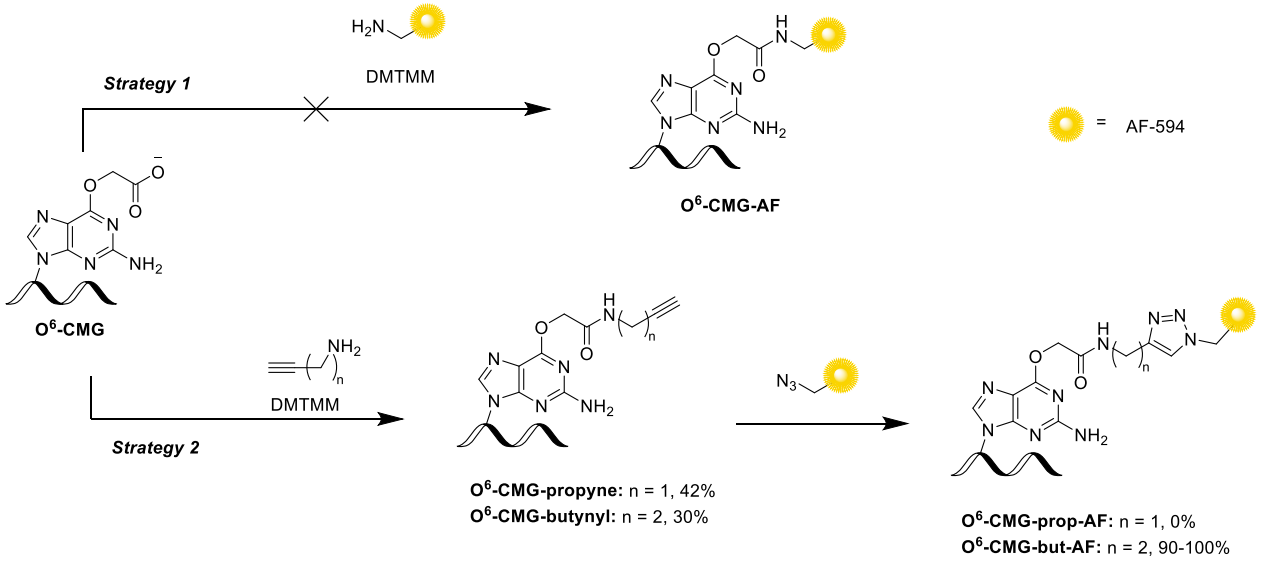


Figure 1. General Labeling Strategies for attaching a fluorophore to *O*^{6}-CMG adduct. **Strategy 1**: direct ligation of *O*^{6}-CMG by fluorophore with primary amine, resulting in no product formation. **Strategy 2**: two step approach by first incorporating small clickable handle followed by subsequent copper-mediated click-reaction to install fluorophore on the *O*^{6}-CMG adduct

Co-solvent and temperature modulate reactivity. We next aimed to optimize the amine ligation of 3-butynylamine towards the *O*[6]-CMG containing oligonucleotide due to the relatively low yield of 30%. To optimize the conversion, we tested different reaction times, temperatures, and co-solvents. The preliminary 30% conversion was obtained by using the 9-mer oligonucleotide (25 µmol), 3-butynylamine (200 equivalents) and DMTMM (1000 equivalents) in buffer (MOPS, pH 8) for 16 h at room temperature. To test whether higher conversions can be achieved over

time, a time-course study over six days increased the conversion to 90% (Figure 2A). However, reaction time longer than 16 h are considered impractical for the labeling method to be applied as quantification tool. Therefore, a higher temperature was tested to drive the reaction to completion in a shorter time span, together with co-solvents to increase solubility of the reagents. Co-solvents were tested by replacing 20% of the aqueous buffer with DMSO, DMF or Dioxane. After reacting for 16 h at room temperature, we found that both DMSO and DMF increased conversion by more than 10%, while dioxane prevented the reaction, as no product was found. Incubating the reaction for 16 h at 60°C with no co-solvent increased the conversion from 30% to 57%, but additional DMSO decreased this yield to 51%. Of the three co-solvents tested, dioxane is the most polar, and the reduced reactivity may be due to lower solubility of the primary amine or instability of DMTMM. DMTMM is known to be instable in inorganic solutions, as it undergoes self-immolative degradation. The half-life of DMTMM in pure DMSO was found to be 120 min and in DMF 15 min, however, it was never tested in dioxane.[29] Although decreased instability was observed for DMTMM in DMSO and DMF, we found increased conversions with both co-solvents at room temperature. As the DMTMM activation step is only 30 minutes and the reaction contains only 20% of the co-solvent, it might not be enough to destabilize DMTMM entirely, while the primary amine might be better soluble.

Table 1. Optimizations of amine ligation on *O*[6]-CMG 9mer (dTTTTG[cm]TTTT) with a reaction time of 16 h at RT and conversion was obtained by HPLC.

Primary Amine	Co-Solvent	Temp. (°C)	Conversion (%)
CF-594 Amine	-	RT	0%
Propargylamine (n=1)	-	RT	42%
3-Butynylamine (n=2)	-	RT	30%
	DMSO	RT	42%
	DMF	RT	45%
	Dioxane	RT	0%
	-	60°C	57%
	DMSO	60°C	51%

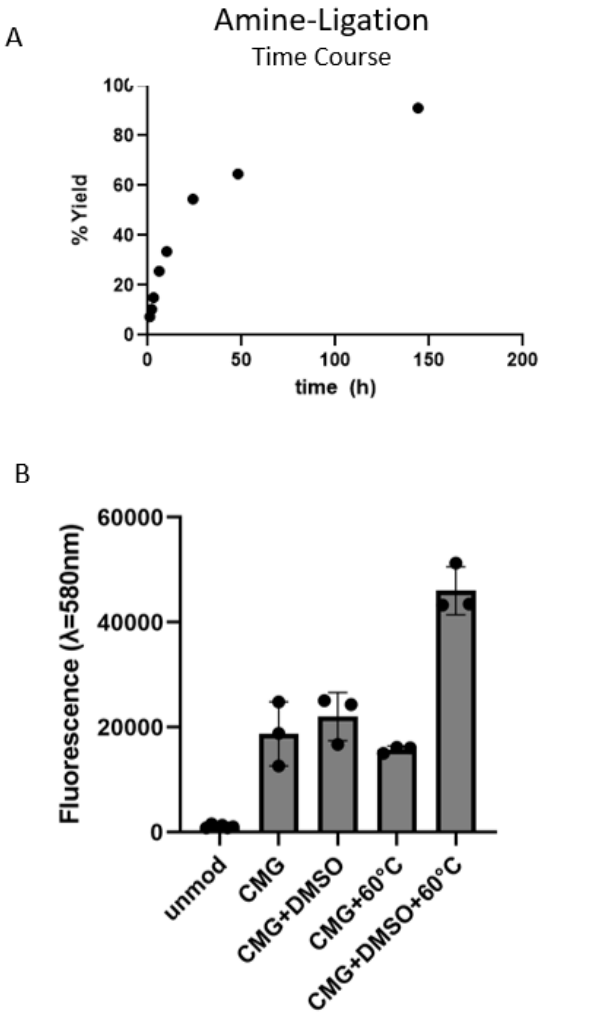


Figure 2. A) Time-course measurement of the amine ligation with 3-butynylamine, without co-solvents and at RT. Over 90% conversion rate was found after 6 days of reaction time. (n=1). B) Optimizations of the amine ligation were done on ds 20mer *O*[6]-CMG oligonucleotide using fluorescence as read-out (λ=580 nm).

To test the efficiency of the amine-fluoroclick ligation in dsDNA oligonucleotide, a longer oligonucleotide was required with a higher melting temperature. Therefore, we synthesized a 20mer sequence (TGAGGCAG^{cm}TCCCGGGCGGGC) containing O^6-CMG. Carboxyethylester phosphoramidite was first synthesized via previously reported method[28] and subsequently incorporated into the oligonucleotide by chemical DNA synthesis followed by hydrolysis[30] to obtain the 20mer sequence containing O^6-CMG. Next, the 20mer O^6-CMG containing strand was annealed to a commercially obtained reverse strand. We tested the amine ligation with DMSO as co-solvent at room temperature and 60 °C. Fluorescence was measured with a plate reader at 16 hours with 2 ng of DNA. We found a clear increase in fluorescence when O^6-CMG is present in the sample. The addition of the co-solvent and elevated temperature alone did not have a great impact on the fluorescent signal, but a dramatic increase was observed when DMSO and increased temperature were used in combination during the amine ligation reaction. We did not see such an effect in single stranded DNA, where we rather saw reduced coupling rates when this combination was used (Table 1). We propose that double-stranded DNA causes a sterical hindrance for the amine ligation reaction. Indeed, it was previously found that single-stranded DNA reacted in significantly higher yield (92%) compared to double-stranded DNA (78%) during a copper-catalyzed azide–alkyne cycloaddition reaction, which was attributed to steric hindrance at the site of the alkyne moiety.[31] Furthermore, it is known that both DMSO[32] and high temperatures lead to DNA melting, making it more accessible to the amine ligation reaction.

5caC is not labeled by the amine-fluoroclick ligation. Multiple carboxylated nucleobases may be present in the genome. Among these is 5caC which originates from oxidation of the epigenetic marker 5meC by TET enzymes and is expected to occur more frequently than carboxymethylation damage. It is therefore crucial for our labeling strategy to detect O^6-CMG and not 5caC. To test selectivity of our method, we used the two-step ligation approach (Figure 3) on commercially available 20mer containing 5caC sequence (Table S1) and measured product formation of the amine ligation by integration of chromatographic peaks. We found 60% conversion for **5caC-propyne** after 16 h and 92% for **5caC-butynyl** after 6 days at RT. Molecular mass of HPLC collected products were confirmed by mass spectrometry (Appendix S4). Subsequently, the final step of the fluorescence labeling process was performed to click the AF-594 fluorophore to **5caC-propyne** and 5**caC-butynyl**. To our surprise, there was no reactivity of either **5caC-propyne** or **5caC-butynyl**. One possible explanation that the click reaction did not work on alkyl modified 5caC is that the carboxy group in O^6-CMG has an additional CH_2 between the carboxy group and the nucleoside aromatic ring preventing the fluorophore from reacting to the butynyl group. For O^6-CMG, we also observed that the distance between the nucleotide played an important role by preventing the strategy 1 from succeeding (Figure 1).

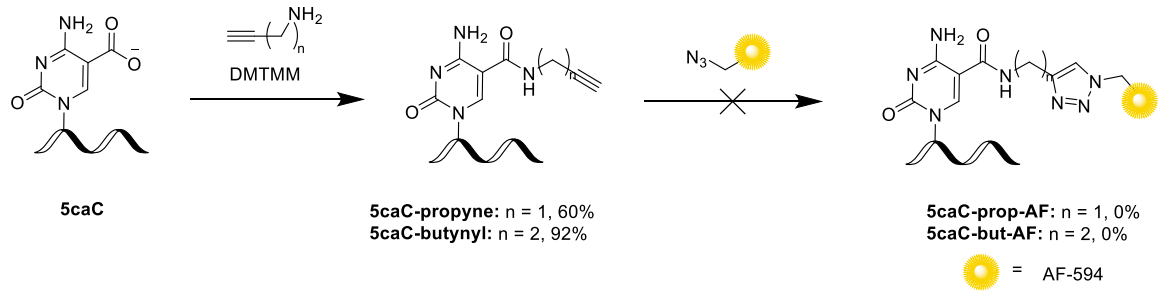


Figure 3. Reaction scheme overview for amine ligation followed by fluorophore attachment by click chemistry for 5caC.

Dose-response increase in fluorescent signal for *O*^{6}-CMG but not for 5caC. To confirm the observed specificity of the fluorescent-labeling method towards *O*^{6}-CMG over 5caC, we prepared mixtures containing 0.01-1 µM of either *O*^{6}-CMG or 5caC modified strands and unmodified 20mer oligonucleotide. We performed the fluorescence labeling approach explained above to analyze these sample mixtures. We found that with increasing amounts of *O*^{6}-CMG, the fluorescent signal increased, while there was no significant increase in fluorescence with higher 5caC amounts (Figure 4). To test whether this observation was sequence related, we obtained a second 20mer with 5caC incorporated in a different sequence context. Also here, we observed no noticeable increase in fluorescence with increase of 5caC. Measured fluorescence and *O*^{6}-CMG concentration showed a strong linear relationship (R^2 = 0.9501). Limit of detection (LOD) was calculated with the formula LOD = 3,3 x σ / S, where σ is the standard deviation and S is the slope. LOD was found to be 80.79 fmol of *O*^{6}-CMG or 12.5 *O*^{6}-CMG per 10^3 nucleotides. This LOD was higher than expected, probably due to the outlier measurement at the *O*^{6}-CMG concentration of 0.5 µM. There may have been sample contamination or an error in the dilution series preparation. This experiment should be performed again with a new dilution series and the concentrations should be cross-verified by LC-MS.

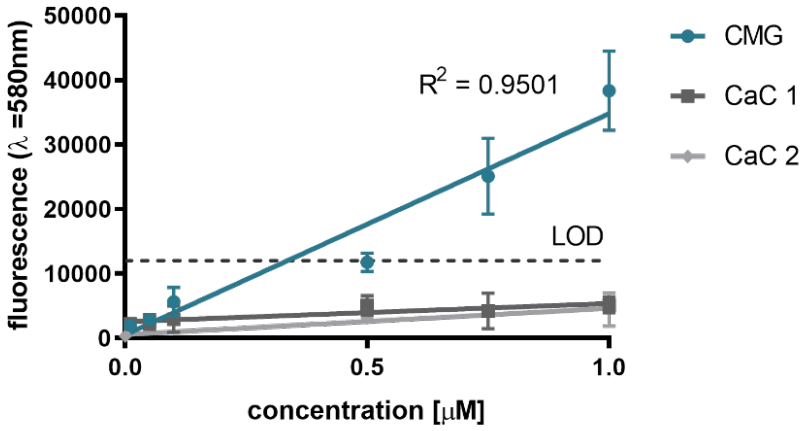


Figure 4. Dilution series of ds 20mer *O*^{6}-CMG and ds 20mer 5caC (5caC in two different sequence context) were reacted with the amine-fluoroclick ligation (20% DMSO, 60 °C, 16 h). There was a fluorescence increase with increasing amount of *O*^{6}-CMG, which was not observed for 5caC.

Potassium diazoacetate (KDA) treated ctDNA is labelled by amine-fluoroclick ligation approach. Having established a selective and robust labeling method for *O*^{6}-CMG adducts, we wanted to test the labeling method on more complex genomic DNA and compare it to a previously defined method of *O*^{6}-CMG detection. Therefore, we synthesized the carboxymethylating agent potassium diazoacetate (KDA) to treat calf-thymus DNA (ctDNA). A dilution series was prepared with proportions of 100%, 50%, 20%, 10%, 5% and 0% of KDA-treated and untreated ctDNA. These samples were measured both by LC-MS, following a previously reported method[12], and the amine-fluoroclick labeling method. We found that higher ratios of KDA treated ctDNA in a sample led to higher levels of *O*^{6}-CMG and to brighter fluorescence. The R^2 of 0.998 (Figure 5) indicates a strong correlation between the fluorescence and *O*^{6}-CMG amount. We calculated the LOD for the fluorescence measurement, as previously described, and found that it lies at the level of samples containing 5% of KDA-treated ctDNA, which translates to 16 *O*^{6}-CMG /10^6 nt. This result

demonstrates that amine-fluoroclick ligation is efficient even on genomic DNA and has the potential to be used as simple quantification tool that does not rely on LC-MS.

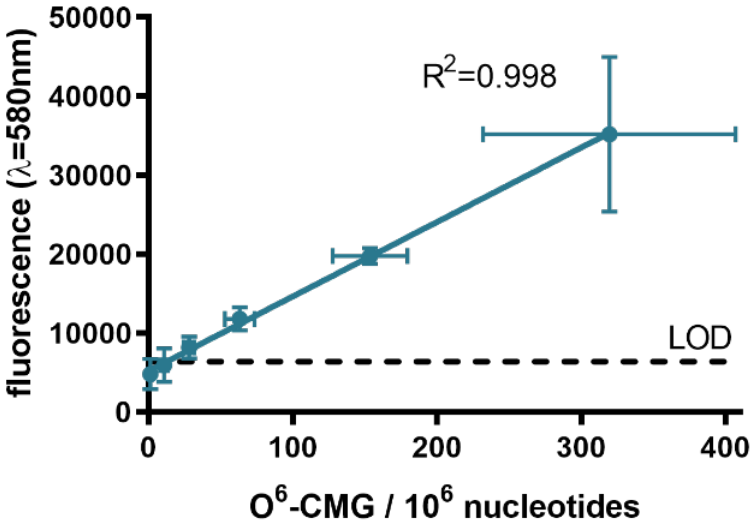


Figure 5. Dilution series of KDA treated ctDNA mixed with untreated ctDNA were prepared and measured with fluorescent amine-fluoroclick ligation and LC-MS. There was a strong correlation between O^6-CMG amount and fluorescence, and the LOD was calculated at 16 O^6-CMG /10^6 nt.

The LOD obtained from the experiment with mixtures of oligonucleotides (Figure 4) was 80 fmol which corresponded to 12.5 CMG/10^3 nt. However, when using ctDNA treated with KDA we found an LOD of 16 CMG/10^6 nt. These results were unexpected, since we anticipated the reaction to be more efficient in oligonucleotides. Oligonucleotides usually have a lower melting temperature, which makes it easier to denature dsDNA, which reduces sterical hindrance for the amine ligation to work. One reason to explain why the LOD was lower for ctDNA treated with KDA could be that KDA also induces other DNA carboxymethylations other than O^6-CMG,[33,34] which could potentially also be labeled. Similar samples were used when the LOD was measured in the immunoslot blot method. There, ctDNA was treated with 10 mM KDA and known amounts of the adducted DNA were mixed with ctDNA without adducts to give a series of DNA standards with 0−10 fmol O^6-CMG/µg DNA.[8] Using these samples, a standard curve was built, and the LOD was determined to be at 15 O^6-CMG/10^8 nt, which is 100x lower than what we observed for our method (16 CMG/10^6 nt). Thus, further optimizations on the amine-fluoroclick method are required for application on human samples.

Amine-fluoroclick labeling of cellular DNA. To test whether the labeling method is suitable for damage detection in cellular DNA, we exposed human skin fibroblast cells (BJ5-ta) with 900 µM of azaserine. Isolated genomic DNA was fluorescently labeled using both the amine-fluoroclick method and quantified using LC-MS. Fluorescence measurement indicated a slight increase in fluorescence when comparing the exposed to the unexposed cell DNA samples (Figure 6B), but only when the gain, which determines the amplification of a detected signal, was set to maximum (250) while we usually measured with a gain of 170. With quantification by LC-MS, only low damage levels were found for both the azaserine exposed and unexposed DNA (Figure 6A). The measured value for azaserine exposed cell DNA was 3 O^6-CMG/10^6 nt

which was below the LOD previously calculated at 16 O^6-CMG/10^6 nt (Figure 6C). Therefore, a reliable quantification by fluorescence was not possible.

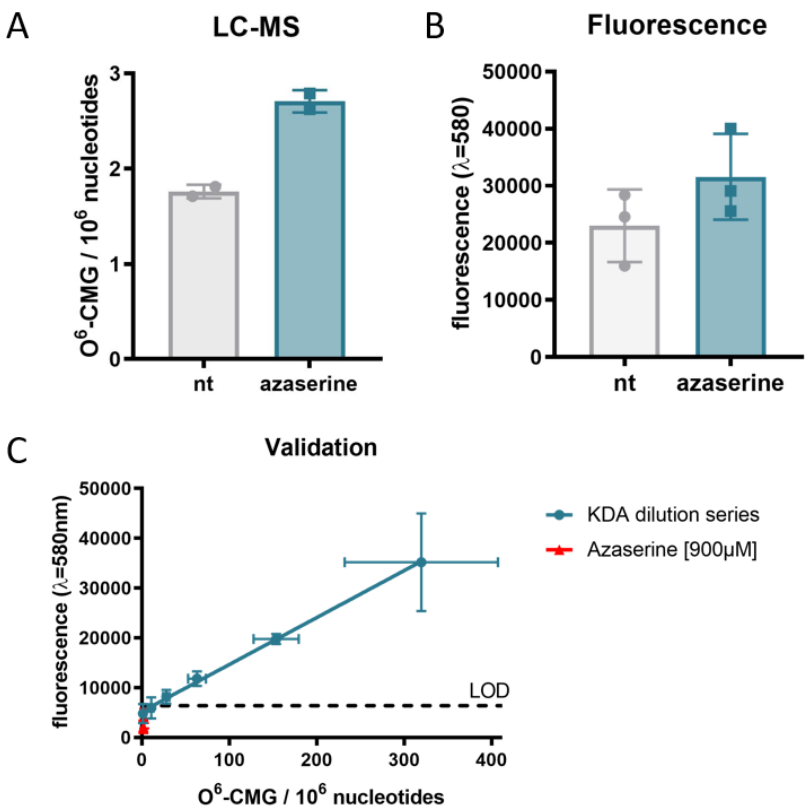


Figure 6. Human skin fibroblast cells (BJ5-ta) were exposed to 900 µM azaserine for 4 h to test the labeling strategy with cellular DNA. A) LC-MS was used to quantify damage in azaserine exposed and unexposed (nt) genomic DNA but only a slight increase in O^6-CMG damage was found for the DNA from the treated cells B) Applying the amine-fluoroclick labeling method on the same samples, only a slight fluorescent increase was seen for the exposed samples compared to unexposed, and only with an increased gain (250). C) Measuring the same samples with a gain of 170, corresponding to what was used for the KDA treated ctDNA samples, the detected fluorescence (in red) was below limit of detection.

Exposure of cells with azaserine usually leads to high levels of DNA damage. In a previously published study, where XPA-deficient fibroblast cells (GM04429) were treated with 450 µM azaserine for 4 h, levels of 49.5 O^6-CMG/10^6nt were found and in azaserine exposed human colorectal carcinoma cells levels of 91 O^6-CMG /10^6 were detected.[19] The wild-type fibroblast cells in our experiment (BJ5-ta) had an intact repair system, therefore, most likely many O^6-CMG adducts were repaired. Using longer treatment times, inhibition of the O^6-alkylguanine repair enzyme AGT, or using XPA-deficient fibroblast cells, would be expected to increase damage to levels needed to validate the current method.

For the previously reported LOD for the O^6-CMG detection in cells by immunoblotting, an LOD of 15 O^6-CMG per 10^8 bases was calculated which was enough to detect O^6-CMG levels in DNA from blood from three volunteers[8] but not in the DNA extracted from exfoliated colon cells which were below the LOD[7]. The LOD of our method was at 16 O^6-CMG/10^6 nt, suggesting that optimizations are needed to improve the sensitivity. Typical detection limits of immune assays using chemiluminescence and microtiter plates range from several adducts/10^7 -10^8 nts but even detection limits of several adducts/10^9 nts for

benzo[*a*]pyrene DNA adducts have been achieved.[4,21] To improve sensitivity in our case, the simplest approach would be to increase the gain of the plate reader (170 -> 250), as it was not set to its maximum during the measurement. The gain determines how much a detected signal is amplified. Therefore, through higher amplification, lower fluorescence can be detected. Furthermore, the amine ligation can be optimized to label more O^6-CMG damage by testing other co-solvents, increasing the temperature above 60 °C, increasing reaction times or a more fluorescent dye could be used.

3.4 Conclusion

We report for the first time a selective carboxymethyl labeling method based on amine ligation and click-chemistry to attach a fluorophore to O^6-CMG using readily available materials. Here, we used a two-step strategy of first ligating a small alkyne chain to the carboxy group, followed by selective incorporation of a fluorophore by click chemistry. DNA labelled with this amine-fluoroclick ligation method were analyzed by fluorescence, which correlated strongly with O^6-CMG levels in both oligonucleotides and genomic DNA with an LOD of 16 O^6-CMG /10^6 nt. Biological validation was attempted using DNA from azaserine-exposed cells, but repair of the O^6-CMG by the intact repair system in the cells resulted in damage levels below the detection limit. By improving the sensitivity, the method can potentially be applied to human DNA, where low levels of O^6-CMG damage are expected, to characterize O^6-CMG as a biomarker and to correlate damage levels with risk factors such as diet, ultimately providing a means to determine an individual's risk of developing colorectal cancer.

3.5 Methods

Chemical Reagents and Materials. 2'-Deoxy-N2-isobutyrylguanosine was purchased from Carbosynth (Compton, UK) and 4,4'-Dimethoxytrityl chloride and TEAA (Aqueous Triethylammonium Acetate, 2 M, pH 7, HPLC grade) from LinkTechnologies Ltd. (Lanarkshire, Scotland). Protected phosphoramidites and chemicals for solid-phase DNA synthesis were obtained from Glen Research Co. (Sterling, VA) and Link Technologies Ltd. (Lanarkshire, Scotland). Two oligonucleotides containing 5caC in different sequence contexts and unmodified oligonucleotides were ordered HPLC-pure from Eurofins Genomics (Germany). PCR & DNA Cleanup Kit (5 µg), was purchased from New England Biolabs (Ipswich, MA, USA). But-3-yn-1-amine (3-butynylamine) was obtained from Fluorochem (Glossop, UK) and Alexa Fluor 594 ((AF594-Picoyl-Azide) Jena Bioscience (Jena, Germany). The click reaction ligand THPTA (Tris(3-hydroxypropyltriazolylmethyl)amine) was manufactured by Lumiprobe (Hannover, Deutschland) and QuantiFluor® ONE dsDNA System by Promega (Madison, USA). All media and buffers for cell culture were purchased from Invitrogen (Carlsbad, California) and centrifugal 10 kD filters from VWR (Switzerland). Labeled and unlabeled CMG standards for LC-MS analysis, were synthesized as described earlier.[12] All other reagents were purchased from Sigma Aldrich.

HPLC methods. Reverse phase high-performance liquid chromatography (Agilent 1100 series/ 1260 infinity) and a Phenomenex Luna (5 µm, C18, 100 Å) column were used. For the amine-ligations on both 20mer and 9mer **Method A** was used to purify product from starting material: linear gradient from 8 to 15% (v/v) acetonitrile in 50 mM triethylammonium acetate over 30 min. To purify the click-product from the alkyne-modified DNA **Method B** was used: linear gradient from 0-16% (v/v) in 0-18 min, followed by a

linear increased up 20% until min 23, and then up to 50% by min 33, using acetonitrile in 50 mM triethylammonium acetate.

Oligonucleotides. The 9mer (TTTTG^{cm}TTTT) containing CMG was produced via copper carbene mediated reaction, as described earlier.[28] A 20mer sequence (TGAGGCAG^{cm}TCCCGGGCGGGC) containing O^6-CMG was synthesized on a Mermade 4 DNA synthesizer from Bio Automation Corporation (Plano, TX) using base-labile phosphoramidites (dmf-dG-CE, Ac-dC-CE, and dA-CE) and CMG-phosphoramidite, which was synthesized as previously reported.[27] Following solid-phase DNA synthesis, the oligonucleotide containing O^6-carboxymethylester was deprotected as described previously to yield the O^6-CMG modification.[30] **Method A** (HPLC methods) was used to purify oligonucleotides by HPLC. The mass of the collected compound was determined by mass spectrometry (MS; Thermo Scientific LGT Velos), operated in negative ion mode. The single strand DNA concentration was determined by monitoring UV absorption at 260 nm on a NanoDrop 1000 spectrophotometer (Thermofisher; Massachusetts, USA).

Labeling strategy in single stranded oligonucleotides. The single stranded 9mer containing O^6-CMG was used to develop and optimize the labeling strategy. A primary amine was coupled via an amine coupling reaction with the carboxylate group in the O^6-position of CMG. Coupling reactions were carried out with 25 µM of the O^6-CMG-containing 9mer template in MOPS buffer (50 mM, 0.5 M NaCl, pH 8). The carboxylate group was activated with DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride) (1000 equiv.) for 30 min at RT (1000 rpm). Then, primary amine (200 equiv) was added, and the reaction was kept on the Thermo-Shaker overnight at RT, at 500 rpm. Different reaction conditions were tested by adding co-solvents, replacing 20% of the water with either DMSO, DMF or Dioxane or incubating the reaction in a PCR machine at 60°C. After the incubation, the DNA was purified by RP-HPLC (**Method A**) and the mass was determined by MS operated in negative mode. Purified O^6-alkyl-modified oligonucleotides were used further for fluorophore click-reactions. The click-reaction was optimized based on a previous published study [24]. Briefly, the alkyne-modified oligonucleotide was mixed in the following order with 100 µM azide-fluorphore, 20% DMSO, Cu/THPTA complex (1 mM/10 mM), and 0.1 M phosphate buffer in a final volume of 20 µL . Freshly prepared sodium ascorbate was added to start the reaction which was incubated at 37 °C on a shaker at 500 rpm. The conversion of each reaction was calculated by integration of chromatographic peaks (HPLC **Method B**).

Reaction optimizations in double stranded oligonucleotides. The 20mer oligonucleotides containing CMG was pooled with equimolar amount of unmodified complementary oligonucleotide and annealed by heating at 95 °C for 5 min and then cooled down to rt. Coupling reactions were carried out with 1 µM O^6-CMG 20mer in MOPS buffer (50 mM, 0.5 M NaCl, pH 8). Co-solvents were tested by replacing 20% of the water with DMSO. The carboxylate group was activated with DMTMM (1000 equiv.) for 30 min at RT (1000 rpm). Then, 1But-3-yn-1-amine (200 eqiv.) was added and reacted at RT or 60°C for 16 h. Final volume was 40 µL. Oligonucleotide was purified from unreacted but-3-yn-1-amine by Monarch® PCR & DNA Cleanup Kit (5 µg) following manufacturer's instructions. Purified DNA was used for copper mediated click-reaction, as described previously, and purified by Monarch® PCR & DNA Cleanup Kit (5 µg). Using the QuantiFluor® ONE dsDNA System the concentration of each sample was determined, and equal amounts of DNA (2 ng) were loaded in triplicates in a black non-clear bottom 96 well plate and filled up to 150 µL with PBS and a pure PBS sample as blank. The fluorescence was read in a Tecan reader at 580 nm excitation wavelength, 623 nm emission wavelength, 25 °C, 25 flashes per well and 170 gain.

Reactivity towards 5caC. A primary amine was coupled via an amine coupling reaction with the 5caC in a 20mer oligonucleotide. Coupling reactions were carried out with 25 µM of the 5caC oligonucleotide in MOPS buffer (50 mM, 0.5 M NaCl, pH 8). The carboxylate group was activated with DMTMM (1000 equiv.) for 30 min at RT (1000 rpm). Then, primary amine (200 equiv) was added, and the reaction was kept on the Thermo-Shaker overnight at RT, at 500 rpm. Subsequently, the DNA was purified by RP-HPLC (**Method A**) and the mass was determined by MS, as previously described. Purified O^6-alkyl-modified oligonucleotide was used further for fluorophore click-reactions. Briefly, the alkyne-modified oligonucleotide was mixed in following order with 100 µM azide-fluorphore, 20% DMSO, Cu/THPTA complex (1 mM/10 mM), and 0.1 M phosphate buffer. Freshly prepared sodium ascorbate was added to start the reaction which was incubated at 37 °C on a shaker at 500 rpm. The conversion of each reaction was calculated by integration of chromatographic peaks (HPLC **Method B**).

Sensitivity and selectivity in oligonucleotides. Oligonucleotides containing O^6-CMG and 5caC at two different sequence contexts were annealed to their complementary strand and dilution series were prepared. Dilutions containing 0.1 uM-1 uM carboxylated DNA were prepared by mixing non-modified oligonucleotides with modified oligos. The optimized amine-fluoroclick labeling was applied on the samples and read-out on the plate reader.

Sensitivity in KDA treated ctDNA. A 800 mM KDA solution was prepared by mixing 0.26 mL of ethyl diazoacetate with 2.85 mL potassium hydroxide (1.8 M) for 4 h in a round-bottom flask on a magnetic stirrer in the dark. This solution was diluted with PBS to generate a 20 mM stock solution of KDA. 300 µg ctDNA split into three 1.5 mL DNA Lo-Bind Eppendorf tubes were exposed to a total of 120 µL KDA stock solution and filled up with PBS to a total reaction volume of 240 µL. The samples were left to react for 22 h at room temperature. The KDA-treated ctDNA was purified by ethanol precipitation and dissolved in 150 µL of MilliQ. The concentration was determined using a Quantus Fluorometer. A dilution series with 100%, 50%, 20%, 10%, 5% and 0% proportion of KDA-treated and untreated ctDNA was prepared.

Azaserine treatment of cells. Human skin fibroblast cells (BJ5-ta) were cultured at 37 °C in 5% CO_2 atmosphere in ATCC-recommended medium containing 10% fetal bovine serum and 100 IU/mL of penicillin. Cells were seeded in 15 cm dishes in complete medium. The cells were treated at 80% confluency with 900 µM of azaserine. After 4 h of treatment, the cells were detached by using trypsin-EDTA and harvested by centrifugation. The cell pellets were subsequently washed twice with phosphate-buffered saline (PBS) to remove residual medium and azaserine. The DNA was isolated using a Monarch kit.

Acknowledgements: The authors wish to thank Laura Slappendel and Jakob Folz for scientific advice and revising the manuscript.

3.6 References

1. Liyanage, V.R. *et al.* DNA modifications: function and applications in normal and disease States. *Biology (Basel)* **3**, 670-723 (2014).
2. Phillips, D.H. & Arlt, V.M. Genotoxicity: damage to DNA and its consequences. *EXS* **99**, 87-110 (2009).

3. Hernandez-Castillo, C., Termini, J. & Shuck, S. DNA Adducts as Biomarkers To Predict, Prevent, and Diagnose Disease-Application of Analytical Chemistry to Clinical Investigations. *Chem Res Toxicol* **33**, 286-307 (2020).
4. Turesky, J.G.B.H.Y.R.J. Chapter 1: Biomonitoring of DNA Damage in Humans. in *DNA damage, DNA repair and Disease* 400 (Royal Society of Chemistry, 2020).
5. Poirier, M.C. DNA adducts as exposure biomarkers and indicators of cancer risk. *Environ Health Perspect* **105 Suppl 4**, 907-12 (1997).
6. Heymann, H.M., Gardner, A.M. & Gross, E.R. Aldehyde-Induced DNA and Protein Adducts as Biomarker Tools for Alcohol Use Disorder. *Trends Mol Med* **24**, 144-155 (2018).
7. Lewin, M.H. *et al.* Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res* **66**, 1859-65 (2006).
8. Cupid, B.C., Zeng, Z., Singh, R. & Shuker, D.E. Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chem Res Toxicol* **17**, 294-300 (2004).
9. Senthong, P. *et al.* The nitrosated bile acid DNA lesion O6-carboxymethylguanine is a substrate for the human DNA repair protein O6-methylguanine-DNA methyltransferase. *Nucleic Acids Res* **41**, 3047-55 (2013).
10. Shuker, D.E. & Margison, G.P. Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Res* **57**, 366-9 (1997).
11. Kostka, T. *et al.* Repair of O6-carboxymethylguanine adducts by O6-methylguanine-DNA methyltransferase in human colon epithelial cells. *Carcinogenesis* **42**, 1110-1118 (2021).
12. Geisen, S.M. *et al.* Direct Alkylation of Deoxyguanosine by Azaserine Leads to O(6)-Carboxymethyldeoxyguanosine. *Chem Res Toxicol* **34**, 1518-1529 (2021).
13. Aloisi, C.M.N. *et al.* A combination of direct reversion and nucleotide excision repair counters the mutagenic effects of DNA carboxymethylation. *DNA Repair (Amst)* **110**, 103262 (2022).
14. Raz, M.H. *et al.* Bypass of Mutagenic O(6)-Carboxymethylguanine DNA Adducts by Human Y- and B-Family Polymerases. *Chem Res Toxicol* **29**, 1493-503 (2016).
15. Gottschalg, E., Scott, G.B., Burns, P.A. & Shuker, D.E. Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis* **28**, 356-62 (2007).
16. Hemeryck, L.Y., Decloedt, A.I., Vanden Bussche, J., Geboes, K.P. & Vanhaecke, L. High resolution mass spectrometry based profiling of diet-related deoxyribonucleic acid adducts. *Anal Chim Acta* **892**, 123-31 (2015).
17. Da Pieve, C., Sahgal, N., Moore, S.A. & Velasco-Garcia, M.N. Development of a liquid chromatography/tandem mass spectrometry method to investigate the presence of biomarkers of DNA damage in urine related to red meat consumption and risk of colorectal cancer. *Rapid Commun Mass Spectrom* **27**, 2493-503 (2013).
18. Vanden Bussche, J., Moore, S.A., Pasmans, F., Kuhnle, G.G. & Vanhaecke, L. An approach based on ultra-high pressure liquid chromatography-tandem mass spectrometry to quantify O6-methyl and O6-carboxymethylguanine DNA adducts in intestinal cell lines. *J Chromatogr A* **1257**, 25-33 (2012).
19. Yu, Y., Wang, J., Wang, P. & Wang, Y. Quantification of Azaserine-Induced Carboxymethylated and Methylated DNA Lesions in Cells by Nanoflow Liquid Chromatography-Nanoelectrospray Ionization Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal Chem* **88**, 8036-42 (2016).
20. Hemeryck, L.Y., Rombouts, C., De Paepe, E. & Vanhaecke, L. DNA adduct profiling of in vitro colonic meat digests to map red vs. white meat genotoxicity. *Food Chem Toxicol* **115**, 73-87 (2018).
21. Hernandez-Ramon, E.E. *et al.* Tamoxifen-DNA adduct formation in monkey and human reproductive organs. *Carcinogenesis* **35**, 1172-6 (2014).

22. Jun, Y.W., Harcourt, E.M., Xiao, L., Wilson, D.L. & Kool, E.T. Efficient DNA fluorescence labeling via base excision trapping. *Nat Commun* **13**, 5043 (2022).
23. Torchinsky, D., Michaeli, Y., Gassman, N.R. & Ebenstein, Y. Simultaneous detection of multiple DNA damage types by multi-colour fluorescent labelling. *Chem Commun (Camb)* **55**, 11414-11417 (2019).
24. Wu, J., McKeague, M. & Sturla, S.J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *J Am Chem Soc* **140**, 9783-9787 (2018).
25. Nikolai Püllen, a., Vakil Takhaveev, Navnit Singh, Emma Dillier, Sabine Diedrich, Shana J Sturla. Click chemistry in DNA enables quantification and sequencing of DNA modifications in the human genome. *manuscript in preparation*.
26. Lu, X. *et al.* Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J Am Chem Soc* **135**, 9315-7 (2013).
27. Raz, M.H., Sandell, E.S., Patil, K.M., Gillingham, D.G. & Sturla, S.J. High Sensitivity of Human Translesion DNA Synthesis Polymerase kappa to Variation in O(6)-Carboxymethylguanine Structures. *ACS Chem Biol* **14**, 214-222 (2019).
28. Geigle, S.N., Wyss, L.A., Sturla, S.J. & Gillingham, D.G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem Sci* **8**, 499-506 (2017).
29. Raw, S.A. An improved process for the synthesis of DMTMM-based coupling reagents. *Tetrahedron Letters* **50**, 946-948 (2009).
30. Millington, C.L. *et al.* Convenient and efficient syntheses of oligodeoxyribonucleotides containing O(6)-(carboxymethyl)guanine and O(6)-(4-oxo-4-(3-pyridyl)butyl)guanine. *Nucleosides Nucleotides Nucleic Acids* **31**, 328-38 (2012).
31. El-Sagheer, A.H. & Brown, T. Click chemistry with DNA. *Chem Soc Rev* **39**, 1388-405 (2010).
32. Tuncer, S. *et al.* Low dose dimethyl sulfoxide driven gross molecular changes have the potential to interfere with various cellular processes. *Sci Rep* **8**, 14828 (2018).
33. Jianshuang Wang, Y.W. Chapter Six - Carboxymethylation of DNA Induced by N-Nitroso Compounds and Its Biological Implications,. in *Advance in Molecular Toxicology*, Vol. 5 (ed. Fishbein, J.C.) 219-243 (Elsevier, Netherlands, 2011).
34. Hemeryck, L.Y. *et al.* In vitro DNA adduct profiling to mechanistically link red meat consumption to colon cancer promotion. *Toxicol Res (Camb)* **5**, 1346-1358 (2016).

Supporting Information - Chapter 3

Oligonucleotides used in the study:

9mer : 5’- TTTTGTTTT -3’

20mer CMG: 5’-dTGAGGC*AGTCCCGGGCGGGC-3’

20mer 5CaC (strand 1): 5’-dTGAGGC*AGTCCCGGGCGGGC-3’

20mer 5CaC (strand 2): 5’ -dGCCCGCC*GGGACTGCCTCA-3’

20mer unmodified: 5’-dTGAGGC*AGTCCCGGGCGGGC-3’

20mer unmodified: 5’-dGCCCGCCCGGGACTGCCTCA-3’

O^6-CMG-propyne – 9mer

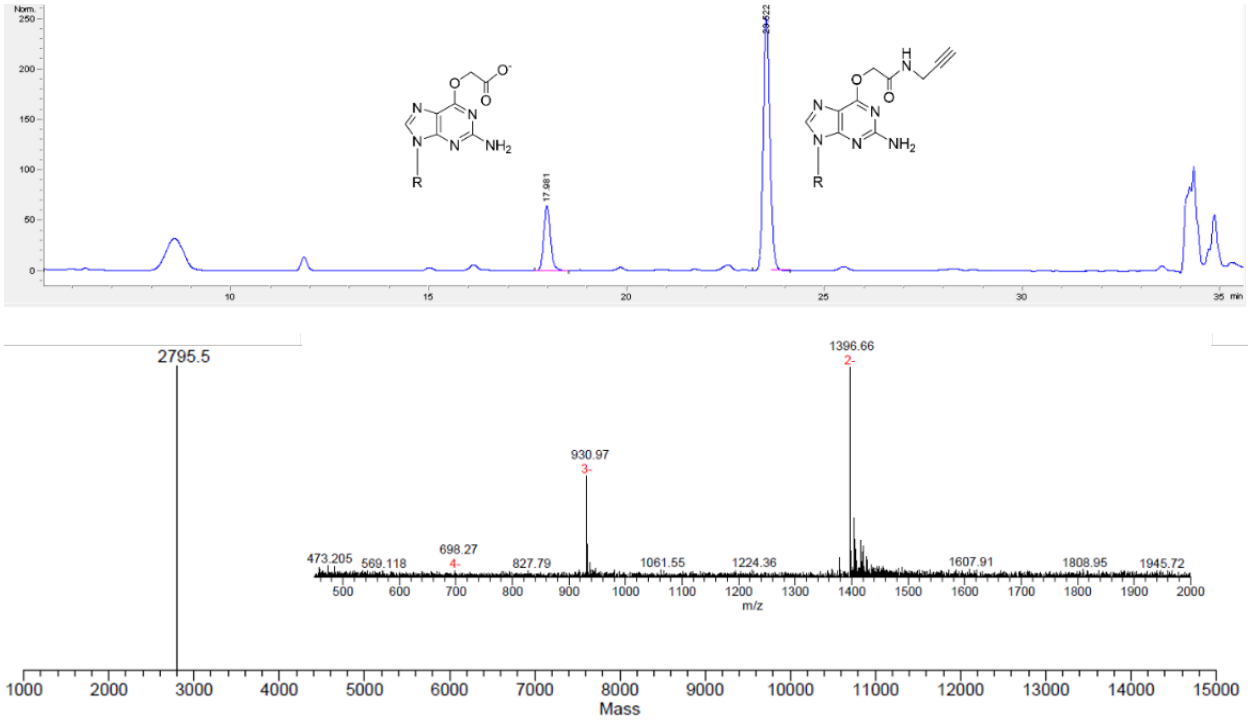


Figure S1: Reversed phase HPLC chromatogram of the amine ligation reaction with propargylamine and 9mer oligonucleotide containing the O^6-CMG adduct with a linear gradient from 8-15% (v/v) acetonitrile in 50 mM triethylammonium acetate over 30 min. Mass spectrometry characterization and Magtran deconvolution of the alkyne containing oligonucleotide (parent mass 2795.69).

O^6-CMG-butynyl – 9mer

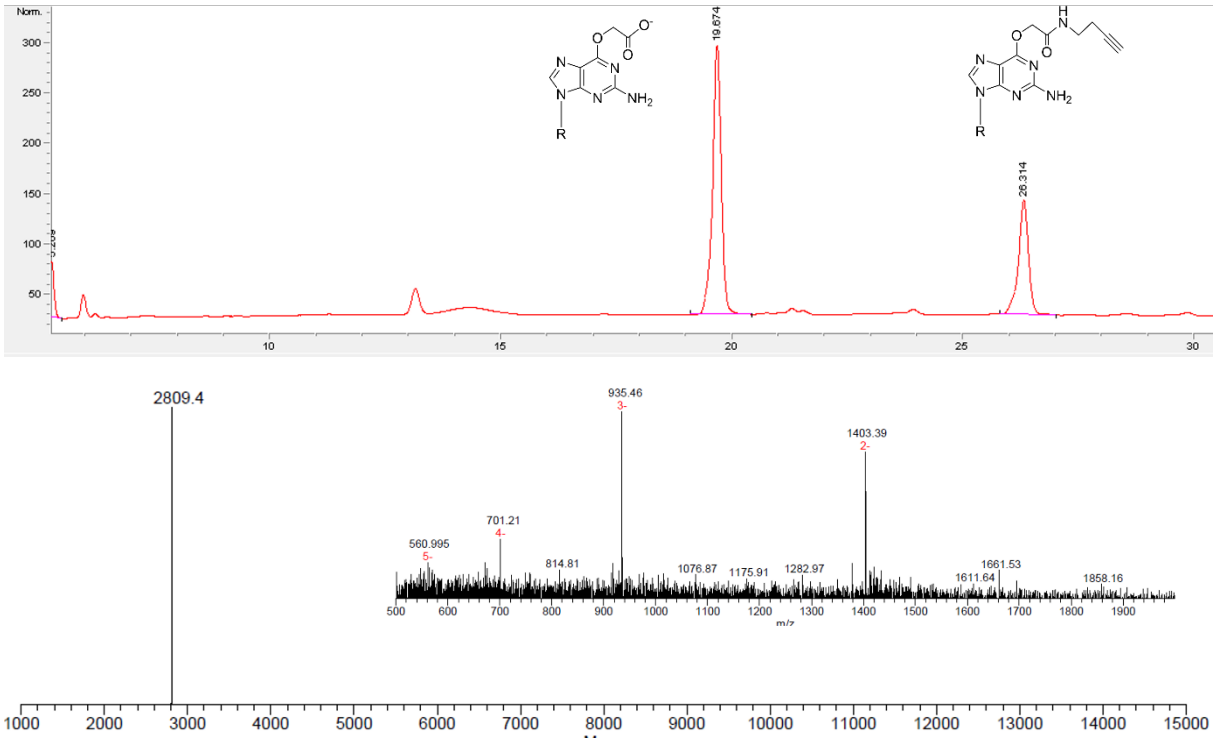


Figure S2: Reversed phase HPLC chromatogram of the amine ligation reaction with butynylamine and 9mer oligonucleotide containing the *O*^{6}-CMG adduct with a linear gradient from 8-15% (v/v) acetonitrile in 50 mM triethylammonium acetate over 30min. Mass spectrometry characterization and Magtran deconvolution of the alkyne containing oligonucleotide (parent mass 2809.692).

O^6-CMG-but-AF – 9mer

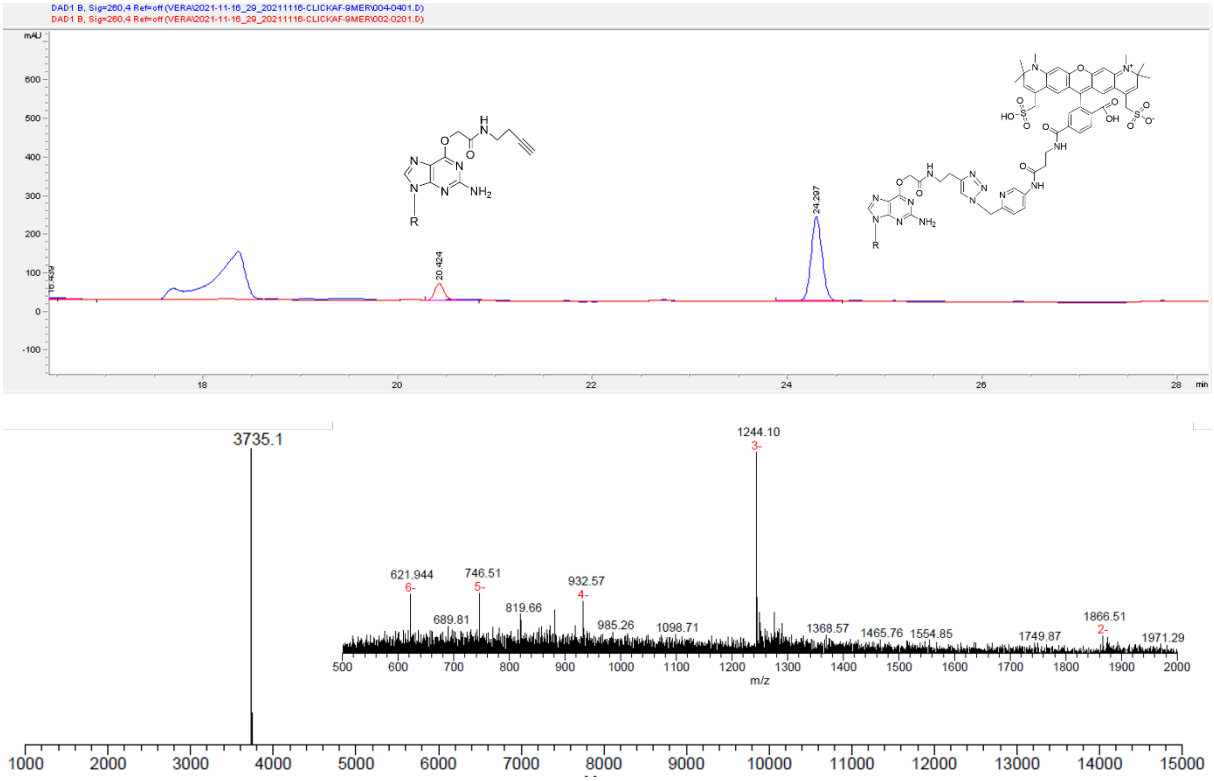


Figure S3: Reversed phase HPLC chromatogram of the amine ligation reaction with butynylamine and 9mer oligonucleotide containing the O^6-CMG adduct with a linear gradient from 0-16% (v/v) in 0-18- min, followed by a linear increased up 20% until min 23, and then up to 50% by min 33, using acetonitrile in 50 mM triethylammonium acetate. Mass spectrometry characterization and Magtran deconvolution of the alkyne containing oligonucleotide (parent mass 3735.923).

5caC-butynyl – 20mer

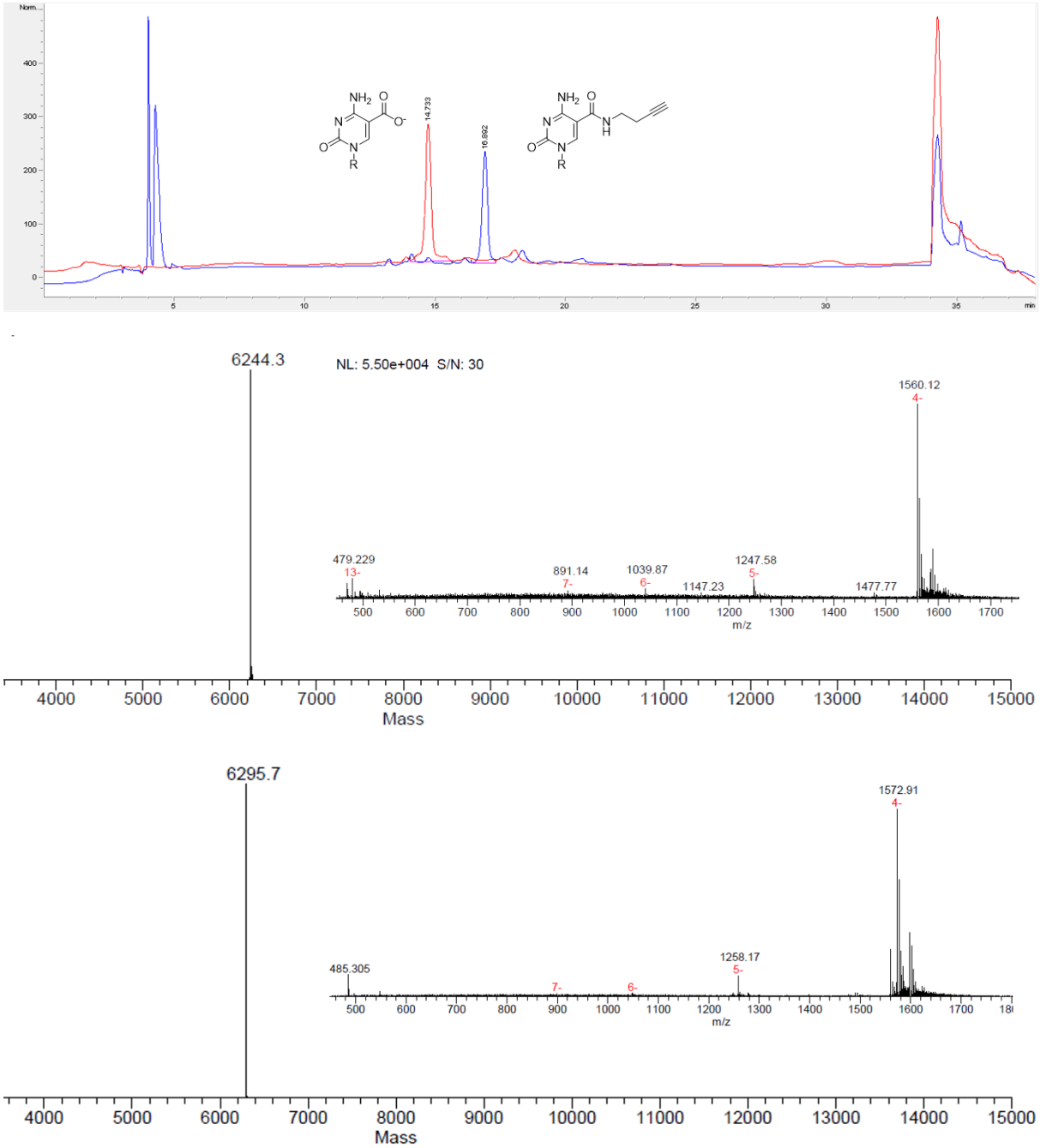


Figure S4: Reversed phase HPLC chromatogram of the amine ligation reaction with 3-butynylamine and 20mer oligonucleotide (strand 1) containing the 5caC modification with a linear gradient from 8-15% (v/v) acetonitrile in 50 mM triethylammonium acetate over 30 min. Mass spectrometry characterization and Magtran deconvolution of the alkyne containing oligonucleotide (parent mass 5CaC oligonucleotide: 6243.95, parent mass alkyl-modified 20mer (strand 1): 6294.95).

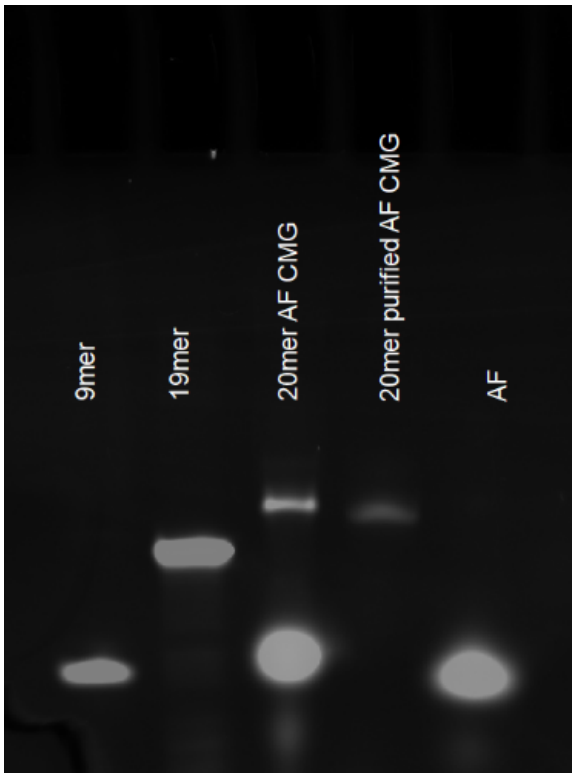


Figure S5: Labeling of 20mer oligonucleotide containing O^6-CMG.

Chapter 4 – Summary and Outlook

The goal of this thesis was to develop new tools to study DNA damage repair and to gain insight into the abundance of O^6-CMG and its impact on human health. In Chapter 1, I give an overview of chemical tools that have been developed to detect O^6-alkylguanines, to investigate their chemical properties and to study their impact on DNA replication and repair. In Chapter 2, an alternative detection and quantification technology based on nanopore sequencing is described and employed to study AGT repair of O^6-alkylguanine (O^6-alkylG) adducts in the nucleosome. In Chapter 3, a chemical labeling method to quantify O^6-CMG in a simple and efficient manner without the use of LC-MS is described. In this chapter, I discuss the limitations and possible solutions, and suggest future studies to advance detection and quantification technologies, towards expanding our knowledge of O^6-CMG and its implications for human health.

Future Directions for Nanopore Damage Sequencing (1) and O^6-alkylguanine Repair (2). Currently, the nanopore-based O^6-alkylG detection method presented in Chapter 2 is limited for targeted detection of DNA adducts in specific sequence contexts. To expand the knowledge and application of nanopore sequencing for DNA modification detection in a broader context, the following factors must be considered.

1.1) An important condition for extending the detection method is allowing to detect DNA damage in any sequence context. For this, training data containing the modified base in any sequence context is required. This has been attempted so far by inducing modifications enzymatically (e.g. 5MeC) or by introducing modified bases via PCR. To introduce O^6-alkylGs, we performed primer extension assay and PCR, and tested an array of A- and B-family polymerases to insert O^6-MedGTP and O^6-CMdGTP. We found

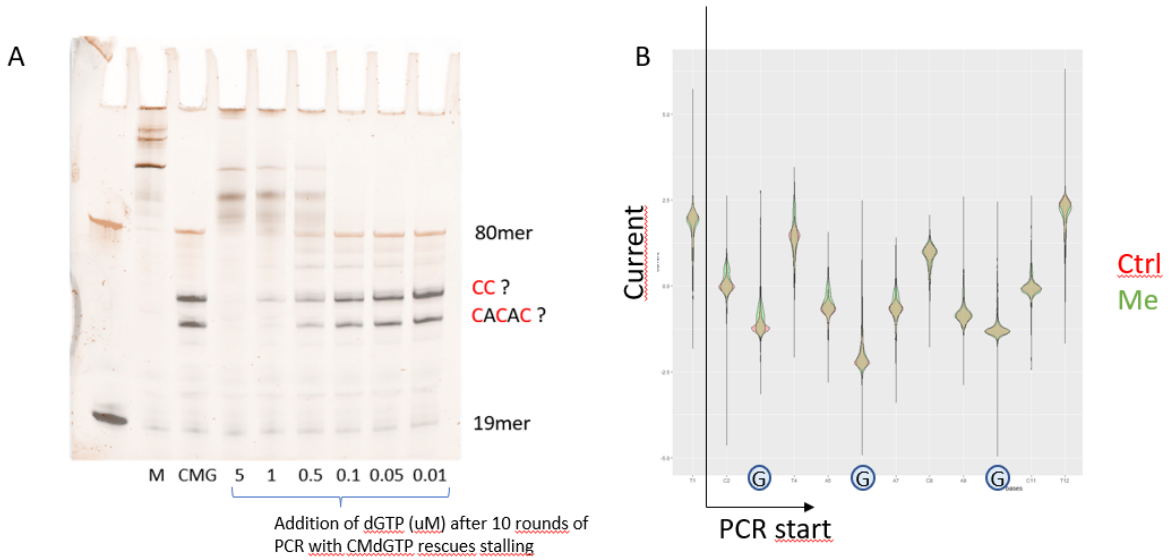


Figure 1. A) PCR was performed with the best performing enzyme VentExo. M = canonical dNTP, CMG=canonical dNTP but dGTP replaced by CMdGTP. After 10 rounds of PCR, different concentrations of dGTP were added to rescue stalling and allow full-length extension. Overextension was found with canonical dNTP, while there was clear stalling where dGTP was replaced by CMdGTP. **B)** 601 sequence was amplified by PCR using VentExo and CMdGTP. Canonical dGTP was added after 10 PCR rounds to allow for full length extension. Current data obtained by nanopore sequencing, indicated only a difference in current between unmodified ctrl (red) and methylated (green) DNA at the start of the strand, implying that modified base was only inserted there.

that all polymerases were stalling after inserting some modified bases, even the most promising one, VentExo. Adding canonical dGTP after ten cycles of PCR rescued the stalling, yielding full extension products (Figure 1A). However, when sequencing by nanopore, we found that the modified base was only inserted at the beginning of the sequence (Figure 1B). This implied that Pols failed at extending DNA for longer than 5-10 bases in the exclusive presence of modified dGTP, and that when canonical dGTP was introduced to rescue the full extension synthesis, dGTP was exclusively introduced opposite C. Thus, for the tested polymerases, O^6-MeG and O^6-CMG were poor substrates, possibly due to the modification at the O^6 position of the guanine, which is involved in hydrogen bonding. To introduce the modifications in a randomized manner via PCR, a bigger number of polymerases, possibly including some novel/engineered ones, need to be screened so that we could have an enzyme accepting O^6-alkylG as substrates hybridized opposite C and retaining high fidelity during DNA replication.

1.2) Only limited data exist on the effect of charged adducts on the current or kinetic data obtained by Oxford nanopore sequencing. We found that O^6-CMG induced a greater current signal change than O^6-MeG, but we could not conclude whether this was due to size or charge. Similar results were found in a study with epigenetic modifications, where the largest current change was seen for the charged modification.[1] To test how the charge influences the current signals and the speed of passage through the nanopore, O^6-CMG analogues of similar size but different charge could be prepared. For instance, similar structures could be synthesized as for the study presented in Appendix B, where we examined the effect of O^6-CMG analogues of different charge and size on TLS repair polymerases.[2] (Appendix B)

1.3) An important limitation of nanopore sequencing for the detection of DNA modifications in biological samples is sensitivity. So far, nanopore has been successfully used to detect epigenetic modifications that occur at high levels and in specific sequence contexts (e.g. CpG), and training data can be easily obtained by enzymatic production. DNA damage, on the other hand, occurs at low levels and training data are difficult to produce. In particular, the low frequency of damage combined with the high error rate of nanopore sequencing (5%) makes it difficult to detect low-level damage in genomic DNA. One option, which has been widely used for other damage detection strategies, are damage enrichment methods to increase the coverage of damaged regions. This has been achieved mainly with specific antibodies coupled to biotin, to extract damaged DNA by streptavidin binding. Using the amine-click ligation method described in Chapter 3, but coupling a biotin instead of a fluorophore, could serve as an excellent basis for a damage enrichment tool.

Finally, further experiments are required to validate the observations reported on the repair by nanopore sequencing.

2) Using nanopore sequencing as quantification tool, several questions related to hAGT repair of O^6-alkylguanine have been investigated. One finding was that O^6-MeG is well repaired, even in the nucleosome. For AGT repair, it is known that clusters of several AGT enzymes are formed around DNA at damaged sites, but due to the histone barrier, accessibility and formation of such enzyme clusters are questionable. Therefore, we proposed monomers to repair the damage in the nucleosome. However, to support this hypothesis, modelling approaches or crystal structures should be obtained with hAGT and damaged nucleosomes. Furthermore, the repair of O^6-CMG should be further investigated as it was not repaired by hAGT in our study, although several studies suggest that AGT is involved in its repair. In

particular, the fact that AGT has many interacting sites and can bind to several other cellular enzymes,[3] as well as the potential synergistic effect of AGT and NER seen in cell studies,[4] suggests that AGT may have a signaling role in O^6-CMG repair. This could be tested in an in vitro assay by combining different isolated cellular enzymes, such as AGT and NER enzymes.

Future directions of fluorescent amine ligation for the quantification of DNA damage. To expand the application of strategies presented in Chapter 3 of this thesis and allow the quantification of DNA adducts in biological samples, three key topics should be addressed: (1) increased sensitivity for detection of DNA adducts present at low cellular level (2) selectivity with regards to DNA adduct structures, and (3) characterization of O^6-CMG as biomarker.

1) In Chapter 3, we developed a chemical fluorescent labeling approach to quantify the adduct O^6-CMG in DNA. Although the method was efficient and robust in measuring O^6-CMG, it lacked sensitivity. Sensitivity is an intrinsic challenge for the detection of DNA damage and a common limitation of detection methods since DNA damage is a rare event compared to the number of bases in the entire genome. With our method, we obtained an LOD of 16 O^6-CMG/10^6 nt, indicating that the sensitivity needs to be improved to apply it to DNA with biological adduct levels. How sensitivity should be improved was discussed in detail in Chapter 4, but briefly 1) amine ligation efficiency should be improved to ligate more O^6-CMG adducts 2) brighter fluorescent dyes can be tested to detect fluorescence at a lower damage level 3) plate reader settings should be optimized, and gain increased to the maximum.

2) To detect DNA adducts in biological samples, selectivity for the target DNA adduct should be tested. Generally, for fluorescence methods relying on antibodies to bind the damage, it is crucial to test antibodies on cross-reactivity and unspecific binding. Our method, however, is based on a chemical reaction that targets only carboxyl groups, which should prevent many non-specific binding interactions towards other DNA adducts not owing any carboxylic acid group. The experiments reported in Chapter 3 have focused only on carboxymethyl adduct O^6-CMG and carboxyl adduct 5caC. However, previous studies have found that carboxymethylating agents such as azaserine can induce various carboxymethyl adducts.[5] For this, further studies should investigate which other carboxyl-adducts occur in DNA after azaserine treatment and the most abundant ones should be synthesized. On these adducts, the amine-fluoroclick labeling strategy should be tested to validate the selectivity of the method.

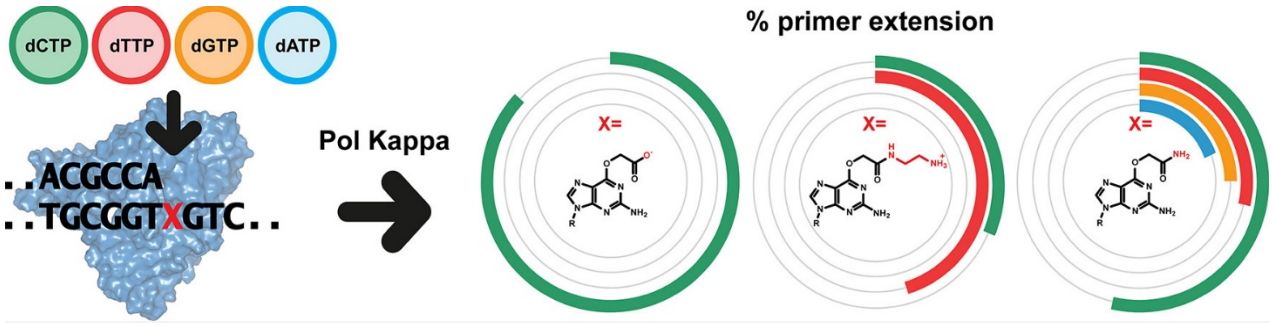
3) After optimizing the labeling strategy for sensitivity and selectivity, the potential of O^6-CMG as a biomarker for early disease detection should be characterized. First, information on which tissues are enriched for O^6-CMG is lacking and needs to be determined. For human studies, it would be optimal to use DNA from blood or urine since DNA can be easily extracted in sufficient amounts. Urine samples would be preferred, however, since they can be obtained non-invasively. To date, O^6-CMG has only been detected in blood,[6] not in urine.[7] Likely, O^6-CMG has not been found in urine, since repair by AGT removes the damaging alkyl group by direct repair, while damaged base repaired by NER or BER are excreted in the urine.[7,8] However, whether O^6-CMG is repaired by AGT or NER or a combination of both remains unclear. Gaining more insight into this will be of potentially great relevance for early-stage CRC detection, where AGT is often inactivated.

With this work, we have developed techniques to advance knowledge aimed at understanding mechanisms of mutagenicity induced by O^6-alkylguanine damage that are of high biological and potentially

clinical relevance. The knowledge gained from this work and from future studies will have far-reaching implications for understanding cancer development and improving cancer prevention strategies.

1. Wescoe, Z.L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* **136**, 16582-7 (2014).
2. Raz, M.H., Sandell, E.S., Patil, K.M., Gillingham, D.G. & Sturla, S.J. High Sensitivity of Human Translesion DNA Synthesis Polymerase kappa to Variation in O(6)-Carboxymethylguanine Structures. *ACS Chem Biol* **14**, 214-222 (2019).
3. Niture, S.K., Doneanu, C.E., Velu, C.S., Bailey, N.I. & Srivenugopal, K.S. Proteomic analysis of human O6-methylguanine-DNA methyltransferase by affinity chromatography and tandem mass spectrometry. *Biochem Biophys Res Commun* **337**, 1176-84 (2005).
4. Aloisi, C.M.N. *et al.* A combination of direct reversion and nucleotide excision repair counters the mutagenic effects of DNA carboxymethylation. *DNA Repair (Amst)* **110**, 103262 (2022).
5. Jianshuang Wang, Y.W. Chapter Six - Carboxymethylation of DNA Induced by N-Nitroso Compounds and Its Biological Implications,. in *Advance in Molecular Toxicology*, Vol. 5 (ed. Fishbein, J.C.) 219-243 (Elsevier, Netherlands, 2011).
6. Cupid, B.C., Zeng, Z., Singh, R. & Shuker, D.E. Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chem Res Toxicol* **17**, 294-300 (2004).
7. Da Pieve, C., Sahgal, N., Moore, S.A. & Velasco-Garcia, M.N. Development of a liquid chromatography/tandem mass spectrometry method to investigate the presence of biomarkers of DNA damage in urine related to red meat consumption and risk of colorectal cancer. *Rapid Commun Mass Spectrom* **27**, 2493-503 (2013).
8. Hernandez-Castillo, C., Termini, J. & Shuck, S. DNA Adducts as Biomarkers To Predict, Prevent, and Diagnose Disease-Application of Analytical Chemistry to Clinical Investigations. *Chem Res Toxicol* **33**, 286-307 (2020).

Appendix A: High Sensitivity of Human Translesion DNA Synthesis Polymerase κ to Variation in *O*^{6}-Carboxymethyl-guanine Structures



Reproduced with permission from

Michael H. Räz, Emma S. Sandell, Kiran M. Patil, Dennis G. Gillingham, Shana J. Sturla*, High sensitivity of human translesion DNA synthesis polymerase κ to variation in *O*^{6}-CMG structures, *ACS Chem Biol*., **2019**, 15;14(2):214-222

E.S. Sandell synthesized oligonucleotides, performed amine ligations and performed primer extension assays and responded and addressed reviewers' comments. M. H. Räz did the modeling, interpreted data and wrote the manuscript, K. M. Patil synthesized CMG-phosphoramidite, D. G. Gillingham revised the manuscript, S. J. Sturla interpreted data and wrote the manuscript and responded and addressed reviewers' comments.

5.1 Abstract

Carboxymethylation of DNA, including the formation of the DNA adduct O^6-carboxymethylguanine (O^6-CMG), is associated with lifestyle factors, such as diet. It can impede replicative polymerases (Pols) and lead to replication fork stalling, or an alternative means for replication to proceed by translesion DNA synthesis (TLS). TLS requires specialized DNA Pols characterized by open and preformed active sites capable of preferential bypass of alkylated DNA adducts but that have high error rates, leading to mutations. Human TLS Pols can bypass O^6-CMG with varying degrees of accuracy, but it is not known how the chemical structure of the O^6-CMG adduct influences polymerase proficiency or fidelity. To better understand how adduct structure determines dNTP selection at lesion sites, we prepared DNA templates with a series of O^6-CMG structural analogs and compared the primer extension patterns of Y- and X-family Pols in response to these modifications. The results indicate that the structure of the DNA adduct had a striking effect on dNTP selection by Pol κ and that an increased steric size influences the fidelity of Pol η, whereas Pol ι and β function were only marginally affected. To test the hypothesis that specific hydrogen bonding interactions between the templating base and the incoming dNTP are a basis of this selection, we modeled the structural analogs with incoming dNTP in the Pol κ active site. These data indicate that the base pairing geometry and stabilization by a dense hydrogen bonding network are important molecular features for dNTP incorporation, providing a basis for understanding error-free bypass of O^6-CMG by Pol κ.

5.2 Introduction

Strong electrophiles are potent genotoxins, damaging the structure of DNA and forming covalent DNA adducts. DNA damage that is left unrepaired can contribute to genomic instability, disrupt cellular functions, and ultimately lead to diseases, including cancer.[1] Carcinogenesis is characterized by an accumulation of somatic mutations, often caused by genotoxic chemicals such as N-nitroso compounds (NOCs).[2,3] NOCs are present in meat,[4] tobacco smoke, and fuel combustion products [5,6] and also are produced endogenously in the stomach and large intestine.[7,8] Their metabolic activation gives rise to reactive alkylating agents, which can modify exocyclic oxygen and ring nitrogens in DNA.[9] However, the formation of O^6-alkylguanine adducts is of particular concern due to their strong miscoding potential during DNA synthesis, where DNA polymerases (Pols) tend to misincorporate T, resulting in GC → AT mutations.[10,11]

Colorectal cancer (CRC) is characterized by the sequential accumulation of mutations in tumor suppressor and proto-oncogenes,[12,13] with point mutation hot-spots located in the kirsten rat sarcoma (KRAS) gene and comprised of GC → AT single base substitutions in up to 50% of colon cancers.[14,15] Among pro-mutagenic O^6-alkylguanine adducts, O^6-carboxymethylguanine (O^6-CMG) may contribute to GC → AT and GC → TA mutation patterns observed in human CRC tumors.[16,17] Furthermore, high O^6-CMG levels were measured in DNA from human colon cells [18,19] and in the sigmoid colon and rectum.[20] In addition, increased O^6-CMG adduct levels in colon cells have also been linked to diets high in red meat.[19-22] Consuming high levels of red meat is hypothesized to facilitate the gastrointestinal bacterial nitrosation of glycine, thereby forming diazoacetate intermediates with the potential to transfer carboxymethyl groups to nucleobases.[7,8] Despite the correlation between red meat, O^6-CMG, and CRC, no direct biological evidence between the occurrence of O^6-CMG and mutations driving colon carcinogenesis have been

established. Detailed knowledge about the enzymatic processing of O^6-CMG will be crucial in determining its relevance in the initiation and progression of colon cancer and in establishing biomarkers of CRC risk. Even more, O^6-CMG stands out among DNA alkylation adducts as one of the few reported negatively charged (COO^-) DNA adducts [23,24] under physiological conditions.

Biological and biochemical effects of O^6-CMG during DNA replication have been characterized in cell lines lacking translesion DNA synthesis (TLS) Pols and using isolated TLS Pols. In HEK293 cells transfected with plasmids containing a site-specific O^6-CMG, the adduct was a modest block for DNA synthesis and was moderately mutagenic.[25] Adduct bypass decreased in cells deficient in Pol κ and ζ, and O^6-CMG induced exclusively GC → AT mutations at a frequency of 6.4% when Pol ι was genetically depleted.[25] Therefore, the data suggested an involvement of TLS Pols κ and ζ in error-free bypass of O^6-CMG, whereas Pol ι seems to be responsible for the mutagenic bypass of the adduct.[25] In line with these findings, primer extension experiments demonstrated that the replicative Pol δ was only moderately blocked. Additionally, Pol κ, which only inserted dCTP, was the highest fidelity Pol among the TLS Pols tested, [26] and Pols η and ι frequently misincorporated dATP and dTTP opposite O^6-CMG.[26] Interestingly, the high rate of dATP insertion opposite O^6-CMG by Pol η contradicted results in transfected cells, where Pol η appeared to have no function in O^6-CMG bypass. Overall, the data from both isolated enzymes and cells suggest the miscoding potential of O^6-CMG depends on which Pol is functional in the TLS bypass of O^6-CMG and its mechanism of processing the adduct.[25,26]

The chemical determinants defining the fidelity of DNA Pols vary, [27-31] and there exists a complex interplay of interactions between the Pol, the DNA, and the damaged base.[32] Pols η, ι, and κ are characterized by a preformed active site, tailored to accommodate DNA adducts and make functional contacts with the DNA template and the incoming dNTP.[33,34] These Pols usually poorly overcome or extend from non-hydrogen bonding nucleobases and largely depend on Watson–Crick hydrogen bonding for faithful and efficient DNA synthesis.[29-31,35] The previous O^6-CMG bypass studies with isolated Pols revealed specific dNTP incorporation patterns opposite O^6-CMG. However, it is unclear how the adduct structure determines the fidelity of Pol η, ι, and κ and in particular what interactions confer the error-free bypass observed by Pol κ.[26]

To address how changes in size and charge of the O^6-CMG adduct impact Pol bypass and contribute to mutations, we performed structure–activity profiling studies. We prepared six different DNA templates with minimally altered O^6-CMG structures (Figure 1A–F) by coupling primary amines to the carboxylic acid group of a site-specific O^6-CMG at codon 12 in the KRAS sequence. Standing-start primer extension assays were carried out with the Y-family Pols η, ι, and κ and the X-family Pol β. Further analysis into relevant hydrogen-bonding interactions during the bypass of modified structures, potentially contributing to error-prone or error-free bypass of TLS Pols, was obtained by computational modeling. Understanding mechanisms that define the fidelity of TLS Pols as well as how adduct structure dictates dNTP selection will advance our capacity to predict mutagenicity on the basis of chemical structures.

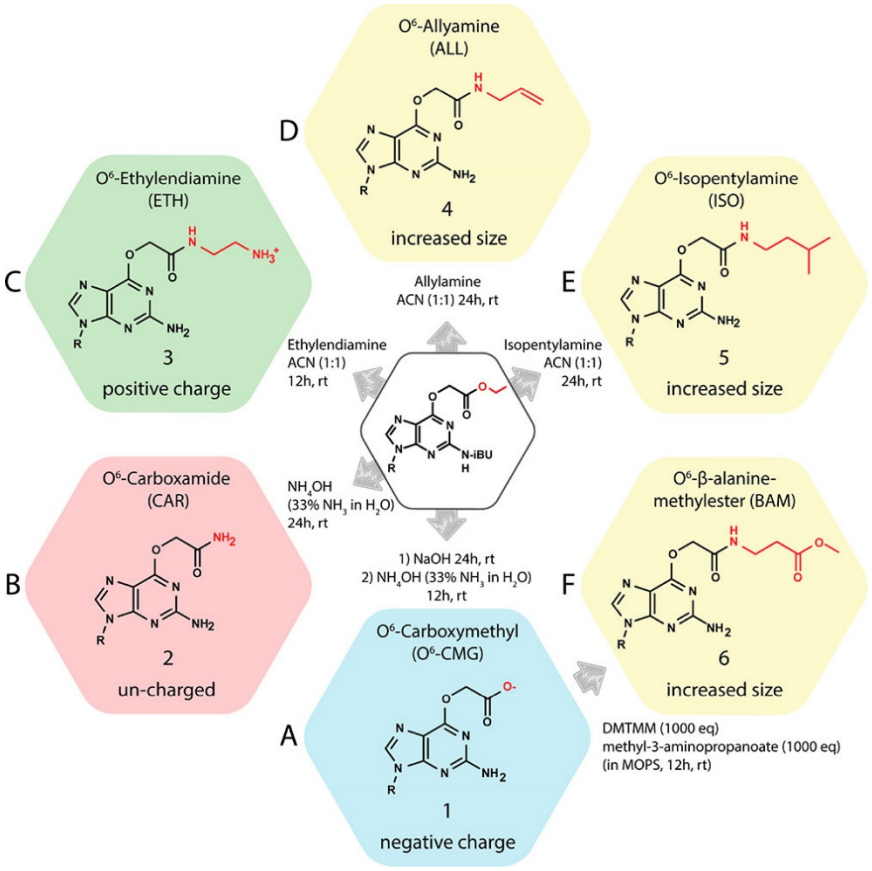


Figure 1. ***O*⁶-**CMG modifications used for the SAR study. *O*⁶-CMG (A) and the analogs CAR, ETH, ALL, and ISO (B–E) were obtained via post synthetic modification of the carboxyethyl ester; different adducts could be obtained depending on whether NaOH, NH_4OH, or primary amines were used. BAM was obtained from *O*⁶-CMG via activation of the carboxylic group with DMTMM followed by coupling with methyl-3-aminopropanoate. R = DNA.

5.3 Results and Discussion

Preparation of Carboxymethyl Adduct Analogs. *O*⁶-CMG was prepared using a recently reported copper-catalyzed alkylation reaction with diazo compounds, allowing us to chemoselectively introduce a carboxyethylester moiety at the O6 position of guanosine.[36] The modified nucleoside was incorporated into oligonucleotides via standard phosphoramidite chemistry and subsequently deprotected as previously described [37] to obtain an *O*⁶-CMG modified DNA template (Figure 1A). To probe the structural determinants of the *O*⁶-CMG bypass, we synthesized carboxymethyl adducts with variations in size and charge postsynthetically using two different strategies via an alkoxide replacement or peptide coupling reaction.

The alkoxide replacement reaction was carried out on oligonucleotides with dmf-dG, Ac-dC, Bz-dA, and iBu-dCMG protected nucleobases, which were treated with NH_4OH or the primary amines isopentylamine, ethylenediamine, or allylamine. The basicity/nucleophilicity of NH_4OH or the corresponding primary amines led to the removal of all protecting groups from the nucleobases or the phosphate backbone and replaced the alkoxide of the *O*⁶-carboxyethylester with NH_3 or the corresponding primary amine (Figure 1B–E). Oligonucleotides with modifications 2–5 (Figure 1) were obtained in yields between 50 and

80%. But we observed the formation of a side product that was 13 Da higher in mass than the theoretically calculated mass for either the ISO, ETH, or ALL modified oligonucleotide. We were unable to verify the adduct's chemical identity but attributed the occurrence of the second species to the N2-isobutyryl (iBu) protecting group because in a test reaction with an N2-dimethylformamide (dmf) protected O^6-carboxyethylesterguanine (in the DNA template) we did not detect the formation of the side-product.

The second strategy used to prepare the β-alanine-methylester modified adduct involved a 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) mediated activation of the carboxylic acid of the deprotected O^6-CMG template followed by an amide coupling reaction (Figure 1, F). Coupling of the β-alanine-methylester to the carboxylic acid resulted in the desired O^6-β-alanine-methylester modified oligonucleotide (**6**) in 40–50% yield.

We also tested whether DMTMM could mediate the coupling with ethylenediamine, isopentylamine, and allylamine. This reaction led to comparable yields as observed for the β-alanine-methylester reaction.

In total, we prepared six different DNA templates containing O^6-CMG or an analog with systematically altered size and charge characteristics (Figure 1). Thus, the O^6-carboxamide analog (CAR, Figure 1B) is similar in size to O^6-CMG (Figure 1A, based on surface volume calculations, Table S1) but is neutral, not negatively charged under physiological conditions. On the other hand, a positive charge under physiological conditions was introduced by positioning an alkyl amine beta to the carbonyl (ETH, Figure 1C). To probe steric interactions in the Pol active site, O^6-isopentylamine (ISO, Figure 1E) and O^6-β-alanine-methylester (BAM, Figure 1F), which contain long uncharged alkyl chains, were synthesized. Finally, to probe combined influences of charge and size, O^6-allylamine (ALL, Figure 1D) was designed to be similar in size to ETH (Table S1) but lacks a charged group.

Pol κ Bypasses O^6-CMG Specifically but Adduct Alterations Significantly Reduce Efficiency and Fidelity. The ability of Pol κ to bypass O^6-CMG and the five carboxymethyl analogs was evaluated by primer extension analysis (Figure 2). Opposite O^6-CMG, Pol κ inserted dCMP almost exclusively as evidenced by a prominent n + 2 band (82%, Figure 2, lane 4). Interestingly, this stringent insertion pattern was unique to O^6-CMG, and Pol κ was more promiscuous for the other five carboxymethyl adducts. We observed that minor alterations of the O^6-CMG structure, e.g., changing the carboxylate group of O^6-CMG to a carboxamide group (CAR), led to insertion of all four dNMPs (Figure 2, lanes 10–13). The volumes of the modifications at the O^6-position for O^6-CMG and CAR are almost identical (e.g., 51 Å3 and 58 Å3, respectively, Table S1); hence the change in the bypass fidelity of CAR might be attributed to the change of the functional group (COO^- to $CONH_2$) rather than to the minimally increased size of the adduct. The apolar and sterically larger adducts ALL, BAM, and ISO also reduced the fidelity of Pol κ, but there was little difference in the observed primer extension pattern between ISO, ALL, and BAM (Figure 2, lanes 15–26). Indeed, the similar mutagenic bypass observed for the small CAR as well as for the larger ALL, BAM, and ISO adducts suggested that steric effects do not account for the observed differences in Pol fidelity. Due to the charge of O^6-CMG, we were therefore particularly interested if electrostatic changes may affect the fidelity of Pol κ.

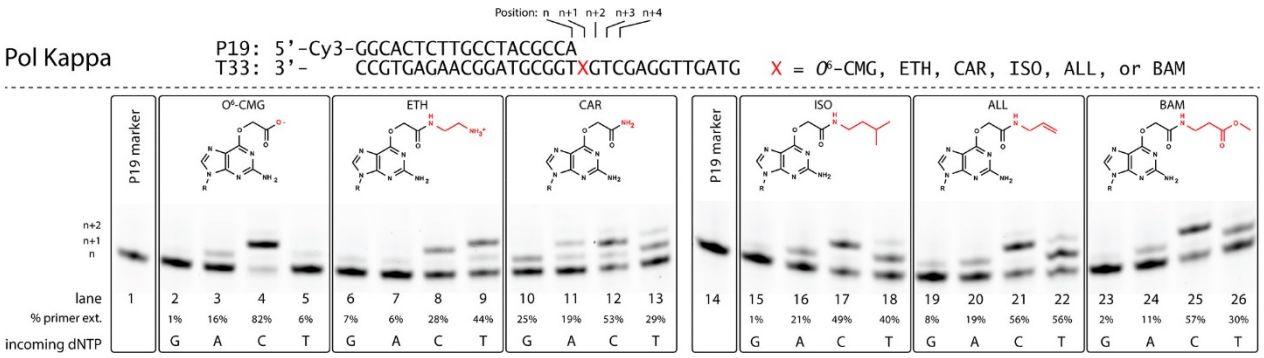


Figure 2. Polymerase extension products observed from Pol κ in the bypass of *O*[6]-CMG and the synthetic carboxymethyl adducts ETH, CAR, ISO, ALL, and BAM.

To investigate the hypothesis that the *O*[6]-CMG charge highly influenced nucleotide selection by Pol κ, we investigated the ETH analog, which is positively charged under physiological conditions due to the ammonium moiety. The volume of ETH (108 Å3) is about twice that of CAR (58 Å3) and almost identical to that of ALL (100 Å3). The primer extension data revealed that Pol κ incorporated dTMP and dCMP opposite ETH. It was therefore bypassed with higher fidelity than CAR or ALL but with lower fidelity than *O*[6]-CMG. Interestingly, the bypass of ETH was characterized by preferential incorporation of dTMP (44%) over dCMP (28%; Figure 2, lane 9 and 8). This result was unique, since the bypass of ETH was the only instance among all tested adducts where incorporation of dTMP seemed to be preferred over dCMP (Figure S18A).

To quantitatively assess the mis-insertion frequency of Pol κ, we determined steady-state kinetic rates for the negatively charged *O*[6]-CMG adduct, the neutral CAR analog (same size as *O*[6]-CMG), as well as for the positively charged ETH analog and the uncharged ALL analog (same size as ETH). Kinetic parameters K_M and k_{cat} were determined under enzyme-limiting conditions with at least a 10-fold excess of DNA and varying dCTP and dTTP concentrations. Among the four DNA substrates tested, Pol κ most efficiently incorporated dCMP opposite CAR with a k_{cat}/K_M of 0.65 μM^{-1} min^{-1} (Table 1, entry 2), which was 11-fold more efficient than incorporation of dCMP opposite *O*[6]-CMG. A large contribution of the efficient Pol κ catalysis opposite CAR seems to derive from the favorable K_M (1.3 μM), which was the lowest K_M observed. However, incorporation of dCMP opposite *O*[6]-CMG (k_{cat}/K_M of 0.06 μM^{-1} min^{-1}, Table 1, entry 1) was 17- and 20-fold more efficient than opposite ALL and ETH, respectively (Table 1, entries 3 and 4).

Interestingly, Pol κ incorporated dTMP opposite ETH with an almost identical k_{cat}/K_M (0.07 μM^{-1} min^{-1}, Table 1, entry 8) as observed for the incorporation of dCMP opposite *O*[6]-CMG, and dTMP incorporation was 23-fold more efficient than incorporation of dCMP opposite ETH. No such drastic difference was observed between the incorporation of dCMP and dTMP opposite ALL (k_{cat}/K_M of 0.004 μM^{-1} min^{-1} and 0.003 μM^{-1} min^{-1}, respectively, Table 1, entries 3 and 7), which is the same size as ETH.

Thus, the change in nucleotide insertion fidelity between ETH and ALL might be a direct effect of the positive charge of ETH. Of note, the K_M as well as the turnover number (k_{cat}) were almost identical, and the highest observed for dCMP insertion opposite *O*[6]-CMG (k_{cat} 2.9 min–1) and dTMP insertion opposite ETH (k_{cat} 3.4 min^{-1}).

Table 3. Steady-State Kinetic Analysis for the Insertion of dCTP or dTTP on Templates Containing O^6-CMG, CAR, ETH, and ALL.

Pol	incoming dNTP	entry	template	k_{cat} [min^{-1}]	K_M[μM]	k_{cat}/K_M [μM^{-1} min^{-1}]	fold change
Kappa (κ)	dCTP	1	O^6-CMG	2.9 ± 0.25	47 ± 10	0.06 ± 0.014	1
		2	CAR	0.82 ± 0.021	1.3 ± 0.23	0.65 ± 0.120	+11
		3	ALL	0.66 ± 0.077	181 ± 40	0.004 ± 0.0010	−17
		4	ETH	0.40 ± 0.033	131 ± 27	0.003 ± 0.0010	−20
	dTTP	5	O^6-CMG	n.d.[a]	n.d.[a]	n.d.[a]	n.d.[a]
		6	CAR	0.16 ± 0.010	15 ± 3.2	0.010 ± 0.0020	−6
		7	ALL	0.42 ± 0.033	124 ± 22	0.003 ± 0.0010	−18
		8	ETH	3.4 ± 0.18	49 ± 7.6	0.070 ± 0.010	+1.1
Eta (η)	dCTP	9	O^6-CMG	0.13 ± 0.004	20 ± 3.0	0.0065 ± 0.0010	1
		10	CAR	0.10 ± 0.003	8.3 ± 1.4	0.012 ± 0.0020	+1.8
		11	ALL	0.15 ± 0.006	61 ± 10	0.0024 ± 0.0010	−2.7
		12	ETH	0.060 ± 0.002	18 ± 2.4	0.0035 ± 0.0010	−1.9
	dTTP	13	O^6-CMG	0.090 ± 0.004	22 ± 4.2	0.004 ± 0.0010	−1.6
		14	CAR	0.050 ± 0.002	6.8 ± 1.2	0.008 ± 0.0010	+1.2
		15	ALL	0.10 ± 0.004	8.1 ± 1.4	0.012 ± 0.0020	+1.9
		16	ETH	0.19 ± 0.004	2.4 ± 0.40	0.080 ± 0.014	+12

[a]Not detected

These results are partially in line with previous findings where Pol κ preferentially incorporated dCMP over dTMP when bypassing other O^6-alkylguanine adducts such as O^6-MeG, O^6-benzylguanine (O^6-BnG), and O^6-[4-(3-pyridyl)-4-oxobutyl]guanine (O^6-PobG). The increasing steric size of the adducts (O^6-MeG < O^6-BnG < O^6-PobG) reduced the incorporation of dCMP and dTMP but promoted incorporation of dGMP opposite O^6-PobG. (38) In our study, dCMP insertion by Pol κ was preferred over dTMP incorporation for O^6-CMG, CAR, and ALL but not for ETH (Figure S18A). The increasing size of the adducts had an impact on the efficiency of Pol κ but did not drastically affect the fidelity (dCTP > dTTP), with the exception of ETH, which was more efficiently bypassed by insertion of dTMP by Pol κ.

Adduct Size Enhances Mutagenicity of Pol η. To investigate the tolerance of the Pols to the O^6-CMG adducts in more detail and compare the substrate specificity of Pol κ with other Pols involved in DNA damage bypass, we performed primer extension assays with Pol ι and η as well as with the X-family Pol β. The substrate specificity of all three Pols was less distinct as compared to Pol κ. Among all Pols tested, Pol η was the most mutagenic in the bypass of all carboxymethyl adducts. The observed primer extension patterns did not show much variation for all adducts tested, and high primer extension percentages for incorporation of dAMP and dTMP opposite all carboxymethyl adducts were observed (Figure S13).

To quantify differences between Pol κ and Pol η bypass function, we also determined steady-state kinetic rates for the bypass of O^6-CMG, CAR, ALL, and ETH by Pol η. Interestingly, Pol η most efficiently incorporated dTMP opposite ETH (k_{cat}/K_M of 0.080 μM^{-1} min^{-1}, Table 1, entry 16) and did so 23-fold more efficiently as compared to incorporation of dCMP opposite ETH (k_{cat}/K_M of 0.0035 μM^{-1} min^{-1}, Table 1, entry 12). Contrary to the high k_{cat} observed for insertion of dTMP opposite ETH by Pol κ, the increased efficiency can be attributed to a favorable K_M of 2.4 μM, which was the most favorable observed for all Pol η kinetic parameters. In addition, Pol η preferred incorporation of dCMP over dTMP only for the small O^6-CMG and CAR analogs (k_{cat}/K_M of 0.0065 μM^{-1} min^{-1} and 0.012 μM^{-1} min^{-1}, respectively, Table 1, entries 9 and 10) and a decreased dCMP incorporation-efficiency was observed for the bulkier ALL and ETH analogs (k_{cat}/K_M of 0.0024 μM^{-1} min^{-1} and 0.035 μM^{-1} min^{-1}, respectively, Table 1, entries 11 and 12). The mutagenic bypass of O^6-CMG was previously kinetically characterized by predominant insertion of dAMP, [26] which was unique, compared with other O^6-alkylG adducts, i.e., O^6-MeG, O^6-BnG, and O^6-PobG.[38]

The differences in the structures of O^6-MeG, O^6-BnG, O^6-PobG, and O^6-CMG, together with the similarities in the Pol η primer extension patterns observed for the carboxymethyl adducts, suggest that the preference for dAMP insertion of Pol η opposite O^6-CMG is likely a consequence of interactions with the CO itself rather than with atoms bonded to it and potential steric or charge–charge interactions with them.

Previous kinetic studies demonstrated that the bypass of Pol η past O^6-alkylG adducts was dominated by incorporation of dCMP and dTMP opposite the smaller O^6-MeG and O^6-BnG adducts,[38] similar to our observation for the bypass of O^6-CMG and CAR (Figure S18B). However, for the larger O^6-PobG adduct, dTMP incorporation efficiency decreased, and more dGMP insertion was observed.[38] Here, our kinetic data revealed that opposite the larger ETH and ALL analogs, dTMP was more efficiently incorporated and that the dCMP incorporation efficiency decreased (Figure S18B). Thus, the increased steric size of ALL and ETH might contribute to a more efficient mutagenic bypass of Pol η and lead to an increased incorporation of dTMP, whereas the smaller O^6-CMG and CAR analogs favored incorporation of dCMP (Figure S18B).

In the cases of Pol ι and β, processivity was similar to Pol κ, but it lacked the sensitivity toward a specific adduct. For all tested adducts the insertion of dTMP was preferred over dCMP, whereas only little to no incorporation of dAMP and no incorporation of dGMP was observed (Figures S14 and S15). In the case of Pol ι, the preference for dTMP insertion is probably due to the architecture of the Pol active site and the propensity for template purines to adopt a syn conformation, which forms a Hoogsteen base pair with an incoming dTTP.[39] In summary, therefore, Pol κ, ι, η, and β were effective in O^6-CMG bypass, and the structure variations of the carboxymethyl moiety seem to uniquely influence the fidelity of Pol κ. Moreover, the increased steric size of ALL and ETH decreased the fidelity of Pol η, whereas the dNTP preference of the other Pols was not affected. Nonetheless, it remained unclear what particular interactions account for these observations.

Computational Modeling Suggests a Role for Interbase Hydrogen Bonding for Error-Free Bypass of O^6-CMG by Pol κ. The primer extension assays revealed that Pol κ is specific in the bypass of O^6-CMG; however, every variation of the O^6-CMG structure (tested in this study) resulted in a more mutagenic bypass of the adducts by the Pol. Replicative Pols often use a combination of geometric fit and shape complementarity to discriminate matched from mismatched base pairs. In contrast, Y-family Pols cannot use a conformation-induced selection mechanism due to their large and solvent exposed active site. Instead, there is evidence that the TLS Pols κ, η, and Dpo4 employ hydrogen bonding to select the correct nucleotide.[29-31,40] Therefore, the specific bypass of O^6-CMG by Pol κ could be directly related to an energetically more favorable hydrogen bonding pattern between O^6-CMG and dCTP as compared to dTTP, dGTP, and dATP. Therefore, we used computational modeling to identify differences in the interbase hydrogen bonding pattern between the templating base and the incoming nucleotide.

To investigate the molecular basis for the specific and error-free bypass of O^6-CMG by Pol κ, we compared computational models of O^6-CMG:dCTP, O^6-CMG:dTTP, O^6-CMG:dATP, and O^6-CMG:dGTP (Figure 3B–E) in the Pol κ active site. These models were built starting from a ternary crystal structure of Pol κ (PDB: 5W2A) [41] in complex with a primer-template strand and an incoming dCTP with Molecular Operating Environment (MOE) software. All modified structures were prepared with the MOE QuickPrep functionality and energy minimized with the Amber 99 force field. The O^6-CMG:dCTP base pair (Figure 3B) adopts a coplanar wobble base pair with three hydrogen bonds between the N1, the N2, and the

carboxylate residue of O^6-CMG with the pairing dCTP. The O^6-CMG:dTTP base pair forms two hydrogen bonds in a Watson–Crick geometry, but the potential repulsion between the O^6 of CMG and the O^4 of dTTP prevents optimal alignment and moves the incoming dTTP out of plane relative to the templating base (Figure 3C). O^6-CMG:dATP as well as O^6-CMG:dGTP can both form hydrogen bonds with the carboxylate residue of O^6-CMG; however, both dNTPs are displaced in the active site and out of plane relative to O^6-CMG (Figure 3D,E). We found that the O^6-CMG:dCTP pair formed the most dense hydrogen bonding pattern, while being in a coplanar geometry, whereas dTTP and in particular dATP and dGTP were misaligned in the Pol active site. Hydrogen bonding, base pair geometry, and also the base stacking are important factors that directly contribute to the polymerase fidelity [42,43] and might be responsible for the specific bypass of O^6-CMG by Pol κ.

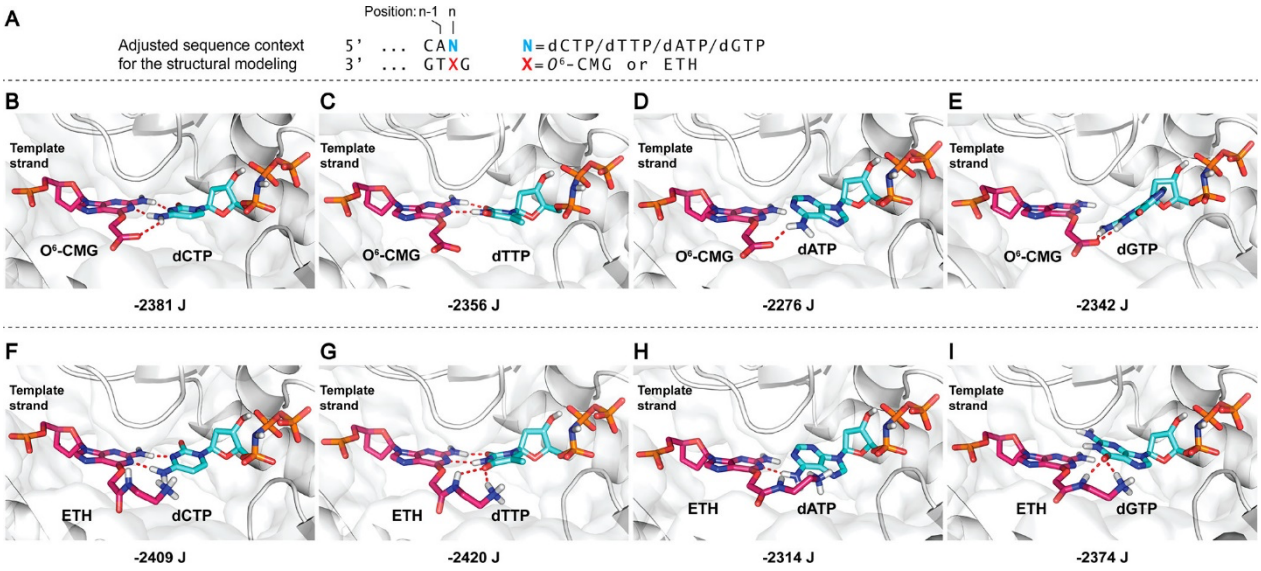


Figure 3. Computational modeling of O^6-CMG or ETH in the Pol κ active site. (A) Adjusted sequence context with X = O^6-CMG or ETH and N = the incoming dNTP. Structural models of (B) O^6-CMG:dCTP, (C) O^6-CMG:dTTP, (D) O^6-CMG:dATP, and (E) O^6-CMG:dGTP. (G) Structural model of (F) ETH:dCTP, (G) ETH:dTTP, (H) ETH:dATP, and (I) ETH:dGTP. The numbers below each modeled structure represent the minimized potential energy (J) of the corresponding base pair. The incoming dTTP is better stabilized upon pairing with ETH due to two additional hydrogen bonds, whereas dCTP forms only two hydrogen bonds with ETH (Figure 3F and G). The better stabilization of ETH:dTTP in comparison to ETH:dCTP may serve as an explanation for the preferred incorporation of dTMP and is supported by the low potential energy (e.g., −2420 J, Figure 3G) predicted by modeling.

Support for the importance of the base pair geometry and hydrogen bonding for the fidelity of Pol κ can be derived from previous research with Sulfolobus solfataricus DNA polymerase Dpo4 (an archaeal homologue of Pol κ). Dpo4-catalzyed primer extension products demonstrated that O^6-MeG was bypassed mainly by insertion of dCTP and only to a minor extent by dTTP. [44] A crystal structure for dCTP paired opposite O^6-MeG in the active site of Dpo4 revealed that the bypass proceeds by pairing dCMP in a wobble geometry with the O^6-MeG, forming two hydrogen bonds.[45] In contrast, structural data for dTTP paired opposite O^6-MeG in the same enzyme displayed disordered base pairing, potentially in a pseudo-Watson–Crick geometry.

In total, three possible O^6-MeG:dTTP structures were predicted, thus indicating that there is not one which is strongly energetically favored.[44] Of note, the same base pairing modes were detected crystallographically for the bulky adduct O^6-benzylguanine (O^6-BnG) in the active site of Dpo4,[45] in

agreement with the base pairing patterns observed in the models herein (Figure 3B,C). Similar to Dpo4, the O^6-CMG:dCTP wobble base pair, modeled in the Pol κ active site, might be energetically more favored due to the three hydrogen bonds as compared to the pseudo-Watson–Crick O^6-CMG:dTTP base pair.

Polymerase bypass of the ETH adduct by Pol κ was less specific as compared to O^6-CMG but not as mutagenic as for the other four O^6-CMG adducts. In order to identify differences in hydrogen bonding, potentially responsible for the altered dNTP selectivity, we thought to model ETH paired with an incoming dNTP in the Pol κ active site and compare these models to O^6-CMG structures. The model of the ETH:dCTP base pair suggested a coplanar wobble geometry and forms two hydrogen bonds between the N^1 and N^2 of ETH and dCTP (Figure 3F). ETH:dTTP adopts a coplanar Watson–Crick base-pair geometry with two hydrogen bonds emanating from N^1 and N^2 of ETH and the other two from the amine and ammonium residue (Figure 3G). The computational model of ETH:dATP displayed one hydrogen bond between the N^1 of ETH and the N^6 of dATP (Figure 3H), whereas the model for ETH:dGTP predicted two hydrogen bonds between the amine and ammonium residue of ETH and the O^6 of dGTP. In addition, dATP and also dGTP were moved out of plane relative to the templating ETH, thereby adopting an unfavorable geometry within the active site. The observed base pairing geometries for ETH were very similar to the modeled O^6-CMG base pairs, however, with particular differences in the hydrogen bonding patterns.

To further test the influence of hydrogen-bonding suggested by the above model, we performed primer extension experiments with modified nucleosides zebularine (dZEB), closely resembling C but lacking the 4′-amino group, and dihydropyrimido[4,5-c][1,2]oxazin-7-one (dPTP), which has two tautomeric forms and can hydrogen bond similar to C or T but is larger (Figure S19C). Modeling the incoming dZEB in the Pol κ active site revealed that dZEB forms only one hydrogen bond with O^6-CMG as well as with ETH (Figure S20B and D), and the primer extension assays revealed that dZEB was incorporated neither opposite O^6-CMG nor opposite ETH (Figure S19A and B). This observation further supports that hydrogen bonding interactions with the 4′-amino group of dCTP are important for dNTP selectivity by Pol κ and stabilize the positioning of the carboxylate group for high dCTP selectivity by Pol κ (compare Figure 3B and F and Figure S20 B and D). In contrast to dZEB, dPTP paired with O^6-CMG or ETH suggested geometries and hydrogen-bonding patterns similar to dTTP. Despite dPTP's increased size, it was incorporated opposite ETH and to some extent opposite O^6-CMG (Figure S19A and B). These data reinforce hydrogen bonding stabilization as a basis for nucleotide selection opposite O^6-CMG.

Since the bypass of ETH by Pol κ was also characterized by a strong n + 2 primer extension band (Figure 2, lanes 8 and 9), we hypothesized that the ETH adduct might have a stabilizing effect on the subsequent primer extension step. To investigate this possibility, we modeled ETH:T and ETH:C in the post-insertion step (Figure S16). The resulting model suggested that ETH may stabilize the incoming dTTP as well as the templating G through hydrogen bonds emanating from the ETH ammonium residue. At the same time, ETH maintains three hydrogen bonds with the paired T in a Watson–Crick geometry (Figure S16B). In contrast, when C is paired with ETH, the ammonium side-chain at the O^6 position of ETH is positioned away from the Watson–Crick face. Thus, it does not interfere with the incoming dCTP but also does not confer any favorable interactions (Figure S16C).

Since O^6-CMG as well as ETH contain pH-sensitive functional groups, we were also curious if changes of the charge state may influence the fidelity of Pol κ. Previous work showed that E. coli. Pol II and IV benefit

from electrostatic effects and are insensitive to steric changes of base pairs.[46] Thus, we adjusted the structure of O^6-CMG or ETH to the corresponding protonated (COOH for O^6-CMG) or deprotonated state (NH_2 for ETH) in the MOE software and assessed the changes in hydrogen bonding with the incoming dCTP or dTTP by energy-minimization. O^6-$CMG_{(COOH)}$ paired with dCTP lost the hydrogen bond between the carboxylate group and the N4 of dCTP (Figure S17B), and ETH(NH_2) paired with dCTP had an additional hydrogen bond (Figure S17C). In contrast, no changes in the hydrogen bonding pattern for ETH(NH_2):dTTP were observed (Figure S17D). These models suggested that charges within carboxymethyl adducts can determine nucleotide selection via interaction with the incoming dNTP.

5.4 Conclusion

The results obtained from this study demonstrate that Pol κ is specific in the bypass of O^6-CMG, but minor changes of the carboxymethyl structure strongly reduce this fidelity. In addition, hydrogen bonding interactions between the templating base and the incoming nucleotide appears to drive the selection of the preferred dNTP. Despite the large degree of structural overlap shared among the Y-family TLS Pols, Pol η and ι did not display this strong bypass selectivity toward a particular carboxymethyl adduct. A unique structural feature of Pol κ is the N-clasp motif on the major groove side, which stabilizes the catalytic complex, and is absent in other Y-family Pols.[47,48] The modifications addressed in this study at the O^6-position are in the major groove and possibly interact with the N-clasp motif, which may explain the higher sensitivity of Pol κ toward the structural variations. Therefore, further research to delineate molecular interactions in Pol κ vs other Pols as a basis of understanding mutagenicity would be of great interest.

5.5 Methods

Chemical Reagents and Materials. dNTPs were purchased from Bioconcept (Switzerland). dZEB and dPTP were purchased from TriLink Biotechnologies (USA). TRIS-HCl (pH 7.0 at 25 °C), TRIS-HCl (pH 8.0 at 25 °C), NaCl, MgCl2, dithiothreitol (DTT), and glycerol were all purchased from Invitrogen (Carlsbad, California). Bovine serum albumin (BSA) was obtained from New England Biolabs (Ipswich, MA, USA). Sep-Pak C18 classic cartridges were purchased from Waters (Milford, MA, USA). All other reagents were purchased from Sigma-Aldrich (Switzerland). Pols η, κ, and ι were purchased from Enzymax (Lexington, KY, USA). Pol β was a generous gift from Prof. Samuel Wilson, National Institutes of Environmental Health Sciences (Durham, USA). O^6-CMG phosphoramidite was prepared by adaptation of a previous method.[36] Synthetic details and characterization of nucleoside intermediates are described in the Supporting Information.

Oligonucleotides. Template sequences were (33mer) 5′-GTAGTTGGAGCTGXTGGCGTAGGCAAGAGTGCC-3′ where X = G, O^6-CMG, ETH, CAR, ALL, ISO, or BAM. The G-containing strand was ordered HPLC-pure from Eurofins Genomics (Germany). The O^6-CMG (CMG) containing template was synthesized on a Mermade 4 DNA synthesizer from Bio Automation Corporation (Plano, TX) using standard conditions and base-labile phosphoramidites (dmf-dG-CE, Ac-dC-CE, and dA-CE). After DNA synthesis, the O^6-carboxymethylester containing 33mer sequence was deprotected as described previously by Millington et al. [37] to yield the CMG modification. The synthesis was performed in trityl-on mode, and the oligonucleotides were postsynthetically purified on a Sep-Pak C18 cartridge (www.waters.com) to remove truncated sequences. Elution of the desired oligonucleotide consisted of cleaving the DMT group and collection of the desired

DNA. Subsequently, the template was purified by reverse phase high-performance liquid chromatography (RP-HPLC) with a linear gradient from 10 to 13% (v/v) acetonitrile in 50 mM triethylammonium acetate over 25 min.

Modification of O^6-CMG by Ester Deprotection. The CAR-containing 33mer oligonucleotide was obtained by deprotection of the O^6-carboxyethylester-containing oligonucleotide with NH4OH (33% NH3 in H2O) at RT (RT) for 24 h. The ISO-, ALL-, and ETH-containing sequences were obtained by deprotecting the O^6-carboxyethylester-containing oligonucleotide with isopentylamine, allylamine, or ethendiamine, respectively, in acetonitrile (1:1) for 24 h at RT. The solution of deprotection reactions of CAR, ISO, ALL, and ETH were evaporated to dryness, resuspended in 100 mM TEAA buffer, and prepurified with a Sep-Pak C18 cartridge as described in the oligonucleotide section. All modified templates were subsequently purified by RP-HPLC on a Phenomenex Luna (5 µm, C18, 100 Å) column with linear gradients from 11 to 16% (ISO and ALL), 10–14% (ETH), and 10–15% (CAR; v/v) acetonitrile in 50 mM triethylammonium acetate over 25 min.

Peptide Coupling Reaction. The BAM modification was introduced via an amine coupling reaction with the carboxylate group in the O^6-position of O^6-CMG. Coupling reactions were carried out with 1–20 nM of the O^6-CMG-containing 33mer template in MOPS buffer (50 mM, 0.5 M NaCl, pH 8). The carboxylate group was activated with DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride) (1000 equiv) at RT for 45 min on a Thermo-Shaker (1000 rpm). Then, methyl-3-aminopropanoate (1000 equiv) was added, and the reaction was kept on the Thermo-Shaker overnight (rt, 500 rpm). Subsequently, the DNA was isolated via ethanol precipitation and purified by RP-HPLC on a Phenomenex Luna (5 µm, C18, 100 Å) column with a linear gradient from 10 to 14% (v/v) acetonitrile in 50 mM triethylammonium acetate over 25 min.

DNA Primer Strand. The sequence for the Cy3-labeled oligonucleotide primer strands was as follows: (19mer) 5′-Cy3-GGCACTCTTGCCTACGCCA-3′. The primer was purchased from Eurofins Genomics (Germany) and purified using polyacrylamide gel electrophoresis (PAGE). The presence of all desired DNA products was confirmed by mass spectrometry (Supporting Information) performed on an Agilent MSD ion trap mass spectrometer with electrospray ionization, operated in negative ion mode. The concentration was determined using UV spectroscopy at 260 nm on a NanoDrop 1000.

Primer Extension Assays. A Cy3-labeled primer was annealed to a complementary 33mer template (5′-GTAGTTGGAGCTGXTGGCGTAGGCAAGAGTGCC-3′ with X = O^6-CMG, CAR, ISO, ALL, ETH, or BAM). The annealing reactions were carried out by heating at 95 °C for 5 min and allowing to cool to RT for 3 h. The primer extension reactions were carried out in a reaction buffer containing 40 mM Tris-HCl (pH 8.0, 25 °C), 1 mM MgCl2, 10 mM DTT, 250 µg/mL BSA, and 3% glycerol. Further, 200 nM 5′-Cy3-primer-template DNA and 40 nM of the polymerase were used. The reaction was started with the addition of 100 µM dATP, dGTP, dCTP, dTTP, and 10 µM, 50 µM, or 100 µM dZEB or dPTP, to the sample containing the buffer, 5′-Cy3-primer-template DNA, and the Pol. The reaction mixture was incubated for 25 min at 37 °C, and the reaction was terminated by adding 10 µL of quenching solution (80% formamide with 10 mM NaOH, 0.1 M EDTA, and a trace of bromophenol blue and xylene cyanol). The resulting solution was loaded on a 15% polyacrylamide 7 M urea gel, run for 2.5 h at 500 V, and visualized with a ChemiDoc Imager (BioRad, Hercules, CA).

Steady-State Kinetics. Steady-state kinetics were performed with Pol κ and η under a range of dCTP and dTTP concentrations. A Cy3-labeled primer was annealed to the complementary template for the primer extension assays, which contained O^6-CMG, CAR, ETH, or ALL. For the enzymatic reactions, 5 nM of Pol κ or 10 nM or Pol η was allowed to react with 100 nM 5′-Cy3-primer-template DNA at 37 °C in reaction buffer containing 40 mM Tris-HCl (pH 8.0 at 25 °C), 1 mM MgCl2, 5 mM DTT, 250 μg/mL BSA, and 3% glycerol. All reactions were performed three times. Reactions were quenched at various time points with stop dye (80% formamide with 10 mM NaOH, 0.1 M EDTA, and a trace of bromophenol blue and xylene cyanol). The aliquots were loaded on a 15% polyacrylamide 7 M urea gel and separated by electrophoresis (500 V, 2.5 h). The product bands were analyzed with a ChemiDoc Imager and quantified with ImageLab 5 software (BioRad Hercules, CA). The nucleotide incorporation rate (v_{obs}) was plotted as a function of the dCTP or dTTP concentration and fit by a rectangular hyperbola using the Michaelis–Menten equation to obtain K_M and V_{max}. The k_{cat} was obtained by dividing V_{max} by the total enzyme concentration.

***In Silico* Modeling Experiments.** Model structures were computed with Molecular Operating Environment (MOE) software suite (Chemical Computing Group Inc., Montreal, QC, Canada). A tertiary crystal structure of Pol κ (PDB: 5W2A) was used for the modeling experiments with incoming ddCTP, a 13-mer template (5′-ATGXCTGATCCGC-3′ with X = lucidin-N_2-G) and a 9-mer primer (5′-GCGGATCAG-3′) were used for all modeling experiments. The crystal structures were modified with the MOE builder tool, to match the sequence context in our in vitro assays. The template was changed to 5′-ATGXTGGATCCGC-3′ with X = O^6-CMG or ETH and the primer to 5′-GCGGATCCA-3′. The polymerase structure was prepared with the MOE QuickPrep function with default settings, including corrections of structural errors, addition of hydrogens, 3D optimization of H-bonding networks, and deletion of water molecules further than 4.5 Å from the protein, and minimization restrained within 8 Å from the modified base pairs. For energy minimization, the Amber99 force field was used. Graphical visualization of results was performed in PyMol software (Schrödinger, New York).

Acknowledgements. We thank D. Rasale and S. Coomar for help with the preparation of the O^6-CMG nucleoside and D. Neri, G. Bassi, and C. Stress for helpful input regarding the peptide coupling reaction. We also thank D. M. Williams for helpful discussion concerning the alkoxide replacement reaction. This work was supported by the Swiss National Science Foundation (156280).

5.6 References

1. Friedberg, E.C., McDaniel, L.D. & Schultz, R.A. The role of endogenous and exogenous DNA damage and mutagenesis. *Curr Opin Genet Dev* **14**, 5-10 (2004).
2. Mirvish, S.S. Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC. *Cancer Lett* **93**, 17-48 (1995).
3. Fahrer, J. & Kaina, B. O6-methylguanine-DNA methyltransferase in the defense against N-nitroso compounds and colorectal cancer. *Carcinogenesis* **34**, 2435-42 (2013).
4. Tricker, A.R. & Preussmann, R. Carcinogenic N-nitrosamines in the diet: occurrence, formation, mechanisms and carcinogenic potential. *Mutat Res* **259**, 277-89 (1991).

5. Drablos, F. *et al.* Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)* **3**, 1389-407 (2004).
6. Fu, D., Calvo, J.A. & Samson, L.D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **12**, 104-20 (2012).
7. Shuker, D.E. & Margison, G.P. Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Res* **57**, 366-9 (1997).
8. Lundberg, J.O. & Weitzberg, E. Biology of nitrogen oxides in the gastrointestinal tract. *Gut* **62**, 616-29 (2013).
9. Gates, K.S. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol* **22**, 1747-60 (2009).
10. Mitra, G. *et al.* Molecular analysis of O6-substituted guanine-induced mutagenesis of ras oncogenes. *Proc Natl Acad Sci U S A* **86**, 8650-4 (1989).
11. Pauly, G.T. & Moschel, R.C. Mutagenesis by O(6)-methyl-, O(6)-ethyl-, and O(6)-benzylguanine and O(4)-methylthymine in human cells: effects of O(6)-alkylguanine-DNA alkyltransferase and mismatch repair. *Chem Res Toxicol* **14**, 894-900 (2001).
12. Fearon, E.R. Molecular genetics of colorectal cancer. *Annu Rev Pathol* **6**, 479-507 (2011).
13. Armaghany, T., Wilson, J.D., Chu, Q. & Mills, G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res* **5**, 19-27 (2012).
14. Seo, K.Y., Jelinsky, S.A. & Loechler, E.L. Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens. *Mutat Res* **463**, 215-46 (2000).
15. Prior, I.A., Lewis, P.D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. *Cancer Res* **72**, 2457-67 (2012).
16. Burns, P.A., Gordon, A.J. & Glickman, B.W. Mutational specificity of N-methyl-N-nitrosourea in the lacI gene of Escherichia coli. *Carcinogenesis* **9**, 1607-10 (1988).
17. Gottschalg, E., Scott, G.B., Burns, P.A. & Shuker, D.E. Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis* **28**, 356-62 (2007).
18. Cupid, B.C., Zeng, Z., Singh, R. & Shuker, D.E. Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chem Res Toxicol* **17**, 294-300 (2004).
19. Lewin, M.H. *et al.* Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res* **66**, 1859-65 (2006).
20. Povey, A.C., Hall, C.N., Badawi, A.F., Cooper, D.P. & O'Connor, P.J. Elevated levels of the pro-carcinogenic adduct, O(6)-methylguanine, in normal DNA from the cancer prone regions of the large bowel. *Gut* **47**, 362-5 (2000).
21. Hall, C.N., Badawi, A.F., O'Connor, P.J. & Saffhill, R. The detection of alkylation damage in the DNA of human gastrointestinal tissues. *Br J Cancer* **64**, 59-63 (1991).
22. Bingham, S.A. *et al.* Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? *Carcinogenesis* **17**, 515-23 (1996).
23. Ochs, S. Reaction of 2′ -Deoxyguanosine with Glyceraldehyde. *Liebigs Annalen der Chemie* (1994).
24. Schneider, M. *et al.* Determination of glycated nucleobases in human urine by a new monoclonal antibody specific for N2-carboxyethyl-2'-deoxyguanosine. *Chem Res Toxicol* **17**, 1385-90 (2004).
25. Wu, J. *et al.* Replication studies of carboxymethylated DNA lesions in human cells. *Nucleic Acids Res* **45**, 7276-7284 (2017).
26. Raz, M.H. *et al.* Bypass of Mutagenic O(6)-Carboxymethylguanine DNA Adducts by Human Y- and B-Family Polymerases. *Chem Res Toxicol* **29**, 1493-503 (2016).

27. Morales, J.C. & Kool, E.T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat Struct Biol* **5**, 950-4 (1998).
28. Morales, J.C. & Kool, E.T. Varied Molecular Interactions at the Active Sites of Several DNA Polymerases: Nonpolar Nucleoside Isosteres as Probes. *J Am Chem Soc* **122**, 1001-1007 (2000).
29. Washington, M.T., Helquist, S.A., Kool, E.T., Prakash, L. & Prakash, S. Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase eta. *Mol Cell Biol* **23**, 5107-12 (2003).
30. Wolfle, W.T. *et al.* Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase kappa. *Mol Cell Biol* **25**, 7137-43 (2005).
31. Gahlon, H.L., Boby, M.L. & Sturla, S.J. O6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem Biol* **9**, 2807-14 (2014).
32. Yang, W. & Gao, Y. Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu Rev Biochem* **87**, 239-261 (2018).
33. Sale, J.E., Lehmann, A.R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* **13**, 141-52 (2012).
34. Yang, W. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase eta. *Biochemistry* **53**, 2793-803 (2014).
35. Krueger, A.T. & Kool, E.T. Model systems for understanding DNA base pairing. *Curr Opin Chem Biol* **11**, 588-94 (2007).
36. Geigle, S.N., Wyss, L.A., Sturla, S.J. & Gillingham, D.G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem Sci* **8**, 499-506 (2017).
37. Millington, C.L. *et al.* Convenient and efficient syntheses of oligodeoxyribonucleotides containing O(6)-(carboxymethyl)guanine and O(6)-(4-oxo-4-(3-pyridyl)butyl)guanine. *Nucleosides Nucleotides Nucleic Acids* **31**, 328-38 (2012).
38. Choi, J.Y. *et al.* Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J Biol Chem* **281**, 38244-56 (2006).
39. Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L. & Aggarwal, A.K. Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature* **430**, 377-80 (2004).
40. Mizukami, S., Kim, T.W., Helquist, S.A. & Kool, E.T. Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* **45**, 2772-8 (2006).
41. Yockey, O.P. *et al.* Mechanism of Error-Free DNA Replication Past Lucidin-Derived DNA Damage by Human DNA Polymerase kappa. *Chem Res Toxicol* **30**, 2023-2032 (2017).
42. Kunkel, T.A. DNA replication fidelity. *J Biol Chem* **279**, 16895-8 (2004).
43. Lee, H.R., Helquist, S.A., Kool, E.T. & Johnson, K.A. Importance of hydrogen bonding for efficiency and specificity of the human mitochondrial DNA polymerase. *J Biol Chem* **283**, 14402-10 (2008).
44. Eoff, R.L., Irimia, A., Egli, M. & Guengerich, F.P. Sulfolobus solfataricus DNA polymerase Dpo4 is partially inhibited by "wobble" pairing between O6-methylguanine and cytosine, but accurate bypass is preferred. *J Biol Chem* **282**, 1456-67 (2007).
45. Eoff, R.L., Angel, K.C., Egli, M. & Guengerich, F.P. Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: steady-state and pre-steady-state kinetics and x-ray crystallography of correct and incorrect pairing. *J Biol Chem* **282**, 13573-84 (2007).
46. Silverman, A.P., Jiang, Q., Goodman, M.F. & Kool, E.T. Steric and electrostatic effects in DNA synthesis by the SOS-induced DNA polymerases II and IV of Escherichia coli. *Biochemistry* **46**, 13874-81 (2007).
47. Lone, S. *et al.* Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol Cell* **25**, 601-14 (2007).

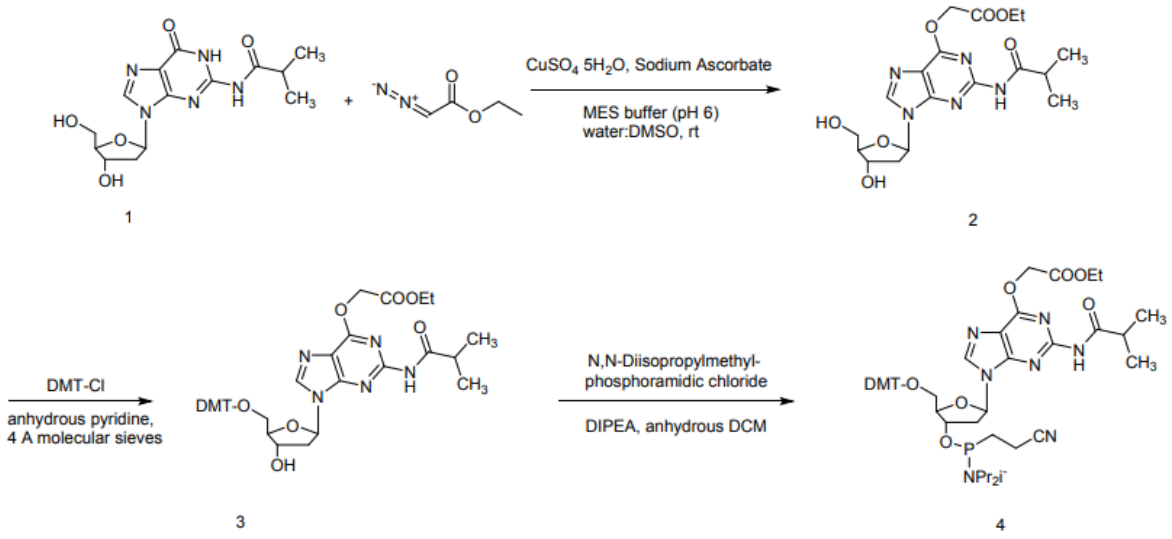
48. Pillaire, M.J., Betous, R. & Hoffmann, J.S. Role of DNA polymerase kappa in the maintenance of genomic stability. *Mol Cell Oncol* **1**, e29902 (2014).

5.7 Supporting Information

Chemical reagents and materials. Reagents were purchased from Sigma-Aldrich and were used without further purification unless otherwise specified. Solvents were purchased from Sigma-Aldrich and Fisher Scientific, and dried by standard techniques. All reactions were monitored with analytical TLC (Merck Kieselgel 60 F254). Flash column chromatography was performed on a Biotage system with pre-packed Flash+ KP-SiO2 cartridges. 1H, 13C, and 31P NMR spectra were recorded on a Bruker Biospin 400 MHz NMR instrument, and chemical shifts are reported in parts per million (ppm, δ) relative to the chemical shift of the respective NMR solvent.

O^6-carboxymethylguaninephosphoramidite synthesis overview.

Scheme 3 Synthesis of ***O*^{6}**-carboymethylguaninephosphoramidite



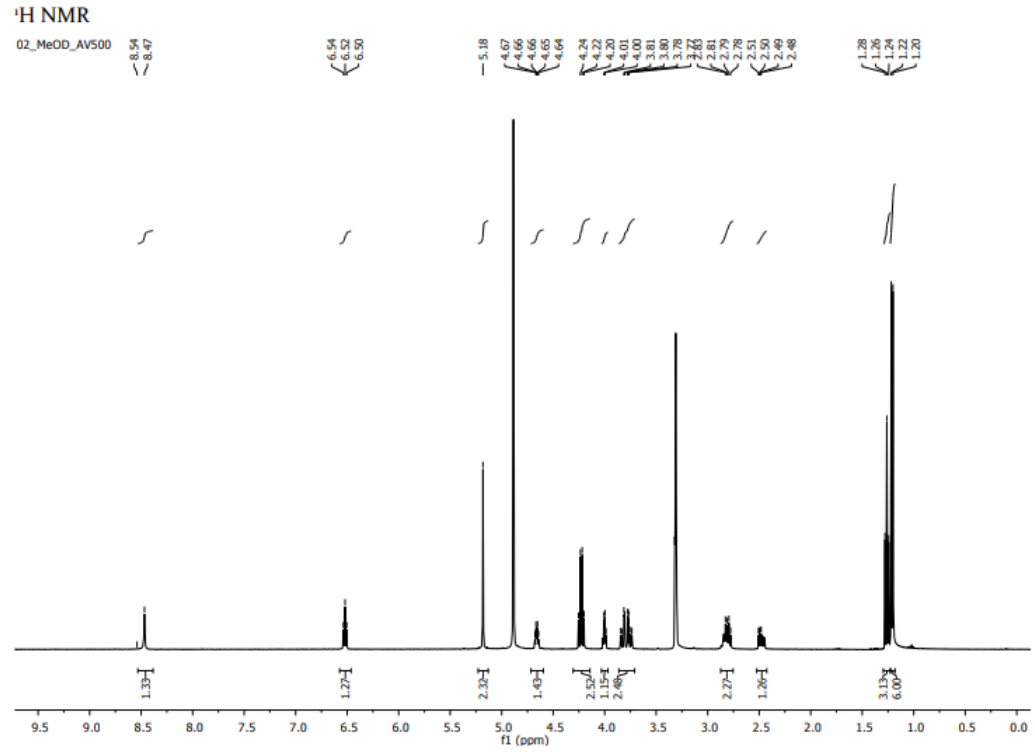
N-(9-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-oxo-6,9-dihydro-1H-purin-2-yl)isobutyramide (2)

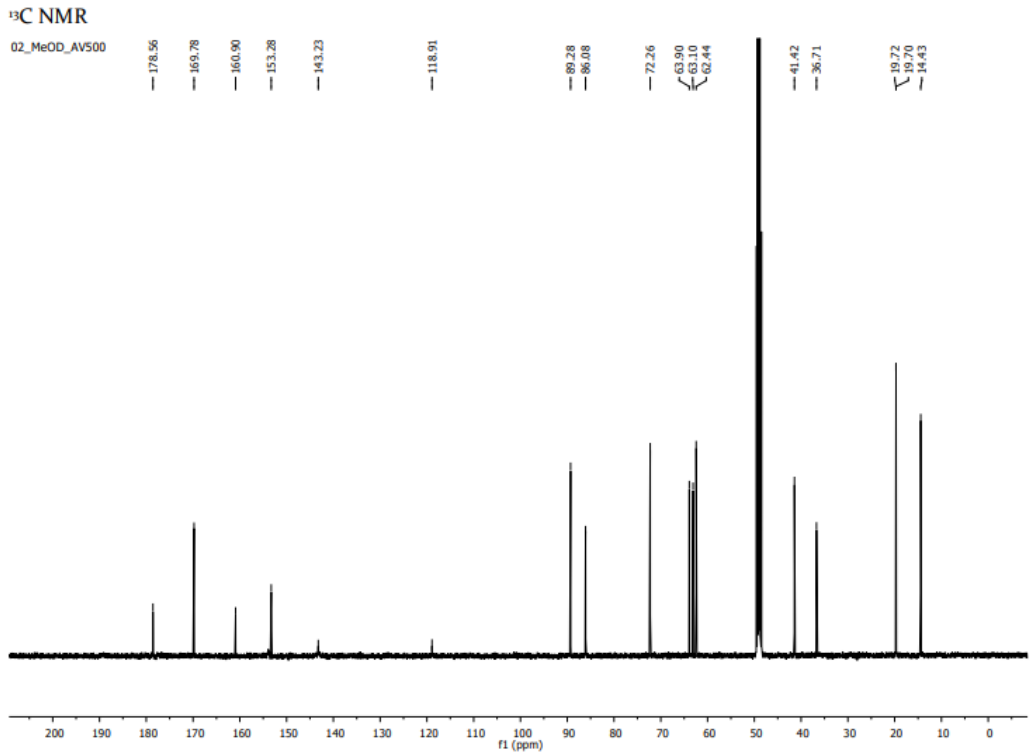
A 500 ml round-bottom flask was charged with H2O (235 ml) and MES buffer (0.5 M, pH adjusted to 6.0 with conc. NaOH, 56.1 ml). The solution was first degassed with N2 and then CuSO4 · 5H2O (0.07 g, 0.28 mmol) was added. A solution of 2'-Deoxy-N2-isobutyrylguanosine (1, 0.4 g, 1.40 mmol) in DMSO (1.0 ml) (dissolved via ultra-sonication) followed by ethyldiazoacetate (1.84 g, 14.00 mmol) and sodium ascorbate (0.28 g, 1.40 mmol) were added. The reaction mixture was stirred overnight at room temperature. After 24 hours TLC showed no presence of the starting material compound 1. The reaction mixture was lyophilized overnight to yield brown solid as a crude reaction product. This crude was re-dissolved in H2O (20 ml) and purified with reversed phase column chromatography using water and acetonitrile solvent gradient. Appropriate fractions were collected and lyophilised to yield 0.4 g (79%) of 2 as a white solid.

^{1}H NMR (400 MHz, MeOD) δ 8.47 (s, 1H), 6.56 – 6.49 (t, J = 6.5 Hz, 1H), 5.18 (s, 1H), 4.69 – 4.62 (m, 1H), 4.27 – 4.18 (q, J = 7.1 Hz, 2H), 4.02 – 3.97 (m, 1H), 3.86 – 3.73 (m, 2H), 2.87 – 2.76 (m, 2H), 2.51 – 2.44 (m, 1H), 1.28 – 1.25 (t, J = 7.1 Hz, 3H), 1.23 – 1.19 (d, J = 6.8 Hz, 6H).

^{13}C NMR (126 MHz, MeOD) δ 178.56, 169.78, 160.90, 153.28, 143.23, 118.91, 89.28, 86.08, 72.26, 63.90, 63.10, 62.44, 41.42, 36.71, 19.72, 19.70, 14.43.

HRMS (ESI+) m/z calculated for C18H26N5O7 [M+H]+ : 424.1830; found 424.1827





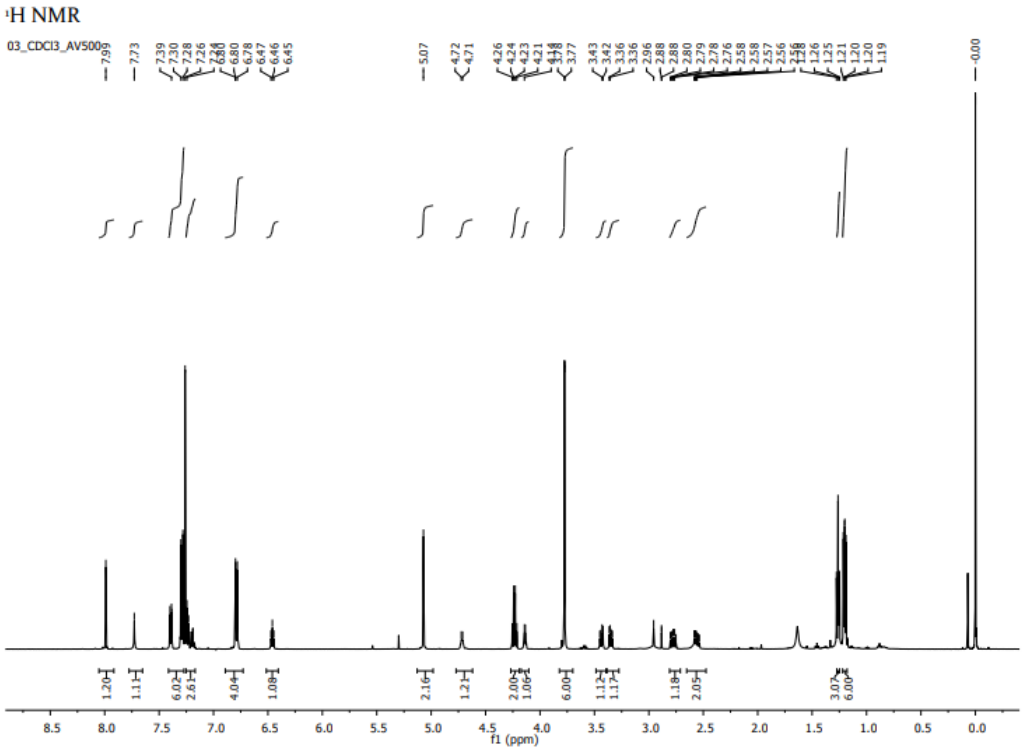
Ethyl-2-((9-((2R,4S,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-2- isobutyramido-9H-purin-6-yl)oxy)acetate (3)

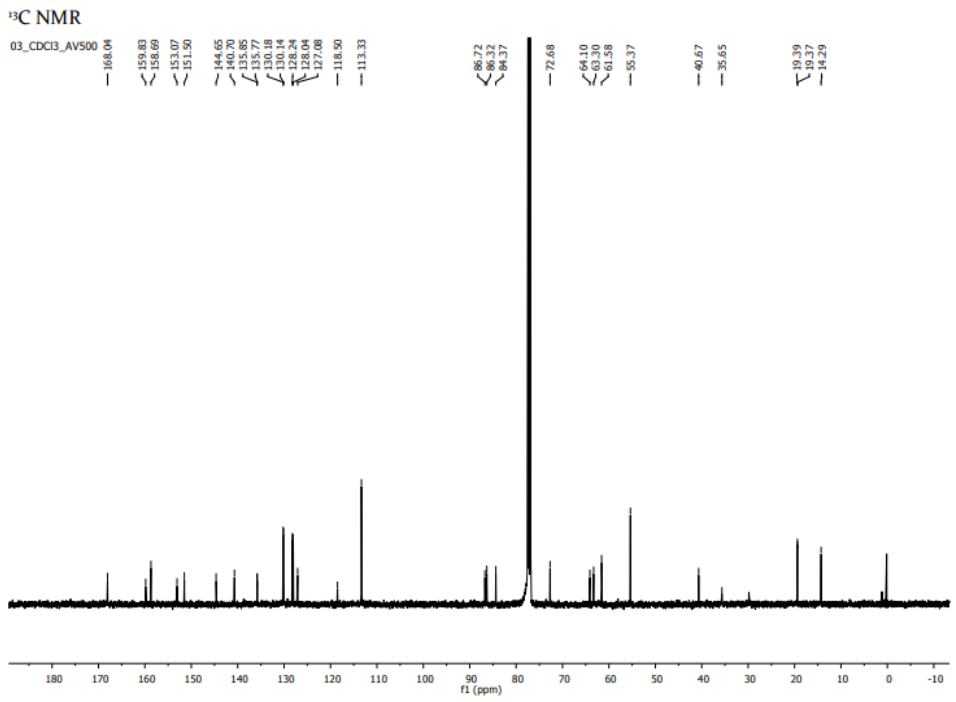
To oven dried 50 mL round-bottom flask, compound 2 (350 mg, 0.826 mmol) and 4 Å molecular sieves (500 mg) were added and to that anhydrous pyridine (5.0 mL) was added under inert gas atmosphere. Stirred the reaction mixture for next 20 min at room temperature and cooled on ice bath. In another flask, the 4,4-Dimethoxytrityl chloride (700 mg, 2.06 mmol) was dissolved into anhydrous pyridine (2.0 mL), which was then transferred to an above mentioned 50 mL roundbottom flask in a dropwise fashion under ice cold condition. Allowed the reaction mixture to warm up to room temperature and continued the stirring for next 20 h. Reaction progress was monitored by TLC. After 20 h, methanol (1.0 mL) was added to quench the excess of reagent added. All the solvents were evaporated and crude compound was purified by normal phase column chromatography (triethylamine neutralized silica gel) using cyclohexane and ethyl acetate solvents containing 0.5% (v/v) trimethylamine. Appropriate fractions were collected and evaporated on rotatory evaporator to yield 340 mg (56%) white foamy compound 3.

^{1}H NMR (500 MHz, CDCl3) δ 7.99 (s, 1H), 7.73 (s, 1H), 7.43 – 7.16 (m, 8H), 6.80 – 6.77 (d, J = 8.5 Hz, 4H), 6.48 – 6.44 (t, J = 6.5 Hz, 1H), 5.07(s, 2H), 4.73-4.70 (m, 2H), (s, 2H), 4.26 – 4.21 (q, J = 7.1 Hz, 2H), 4.16 – 4.12 (m, 1H), 3.79 – 3.76 (d, J = 1.2 Hz, 6H), 3.46 – 3.42 (m, 1H), 3.36 – 3.34 (m, 1H), 2.81 – 2.75 (m, 1H), 2.59 – 2.52 (m, 2H), 1.28 – 1.25 (t, J = 7.1 Hz, 3H), 1.21 – 1.19 (m, 6H).

^{13}C NMR (126 MHz, CDCl3) δ 168.04, 159.83, 158.69, 153.07, 151.50, 144.65, 140.70, 135.85, 135.77, 130.18, 130.14, 128.24, 128.04, 127.08, 118.50, 113.33, 86.72, 86.32, 84.37, 72.68, 64.10, 63.30, 61.58, 55.37, 40.67, 35.65, 19.39, 19.37, 14.29.

HRMS (ESI+) m/z calculated for C39H44N5O9 [M+H]+ : 726.3125; found 726.3134





Ethyl-2-((9-((2R,4S,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-2-isobutyramido-9H-purin-6-yl)oxy)acetate (4)

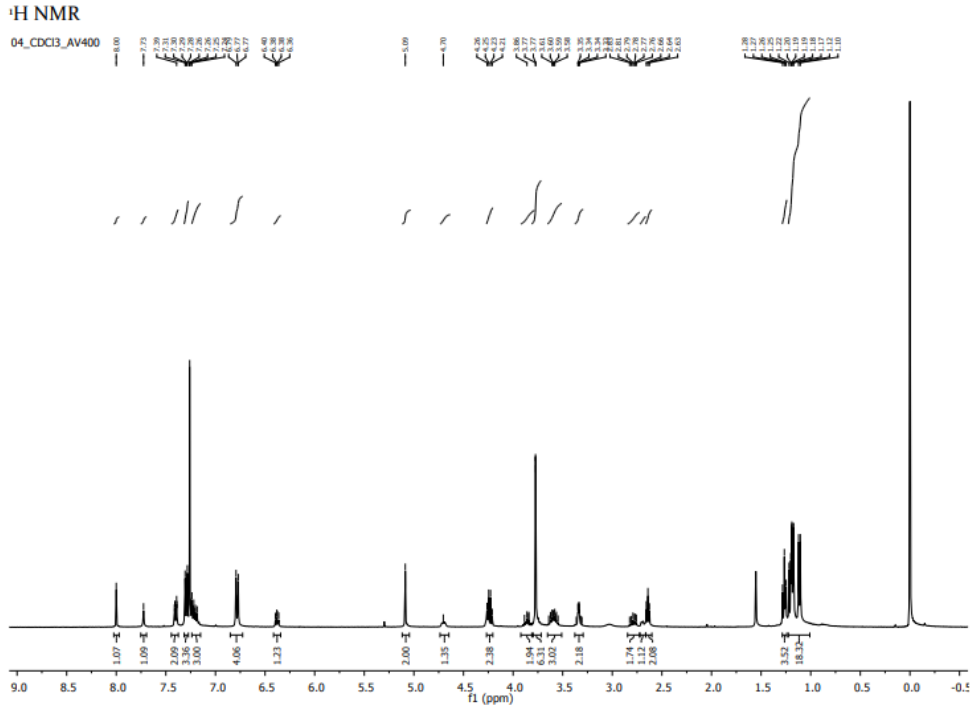
To oven dried 50 mL round-bottom flask, compound 3 (150 mg, 0.35 mmol) and 4 Å molecular sieves (500 mg) were added, to that, anhydrous dichloromethane (3.0 mL) and DIPEA (0.87 mmol) were added under inert gas atmosphere and cooled the flask on ice bath. After 10 min, 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite (0.52 mmol) was added dropwise to above reaction mixture. An ice-bath was removed after next 5 min and allowed the reaction mixture to warm up to room temperature and continued the stirring for next 2 h. Reaction progress was monitored by TLC. The reaction was quenched by addition of methanol (0.5 ml) and stirred for next 5 min. An excess of DCM was then added to reaction mixture and organic phase washed with sat. NaHCO3 and brine solutions, and dried over Na2SO4. Crude product was purified by normal phase column chromatography (triethylamine neutralized silica gel) using cyclohexane and ethyl acetate solvents containing 0.5% (v/v) trimethylamine. Appropriate fractions were collected and evaporated on rotatory evaporator to yield 130 mg (59%) of white foamy compound 4. Two diastereoisomers were separated by column chromatography and NMR analysis was performed on one of the diastereoisomers.

^{1}H NMR (400 MHz, CDCl3) δ 8.00 (s, 1H), 7.73 (s, 1H), 7.41 – 7.39 (m, 2H), 7.18-7.31 (m, 6H), 6.81 – 6.74 (m, 4H), 6.41 – 6.35 (m, 1H), 5.11 – 5.06 (s, 2H), 4.74 – 4.66 (m, 1H), 4.29 – 4.17 (q, J = 7.0 Hz, 2H), 3.91 – 3.80 (m, 2H), 3.79 – 3.75 (d, J = 1.0 Hz, 6H), 3.65 – 3.54 (m, 3H), 3.36 – 3.30 (dd, J = 4.2, 1.8 Hz, 2H), 2.84 – 2.74 (m, 2H), 2.72 – 2.68 (m, 1H), 2.66 – 2.61 (m, 2H), 1.29 – 1.23 (t, J = 7.1 Hz, 3H), 1.23 – 1.05 (m, 18H).

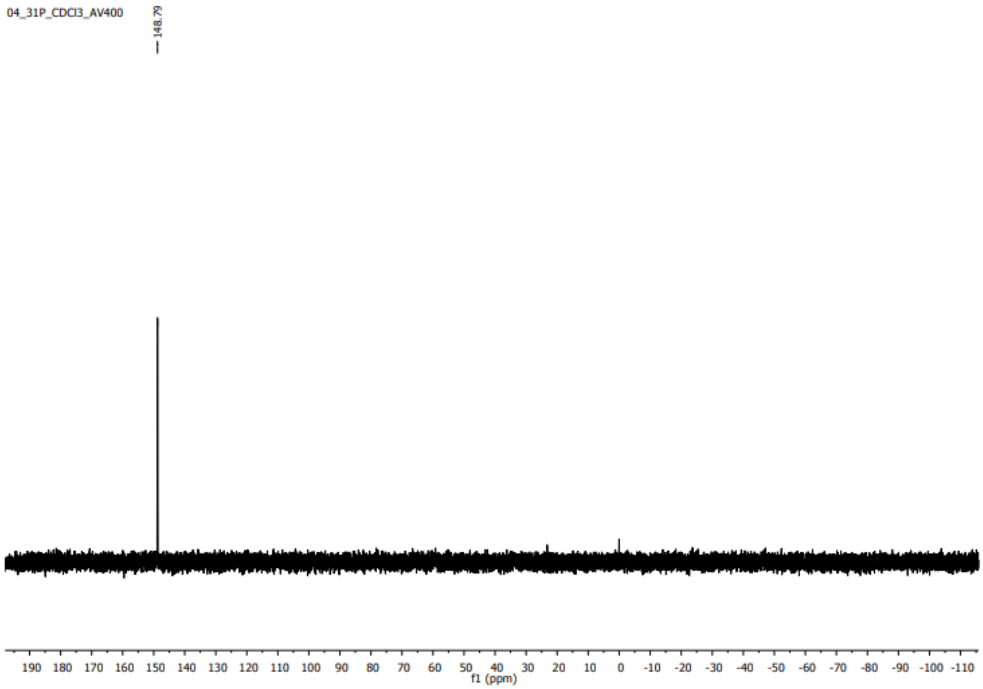
^{31}P NMR (162 MHz, CDCl3) δ 148.79.

HRMS (ESI+) m/z calculated for C48H61N7O10P [M+H]+ : 926.4212; found 926.4212

^{1}H NMR



^{31}P NMR



Carboxymethylguanine template (T33-*O*^{6}-CMG)

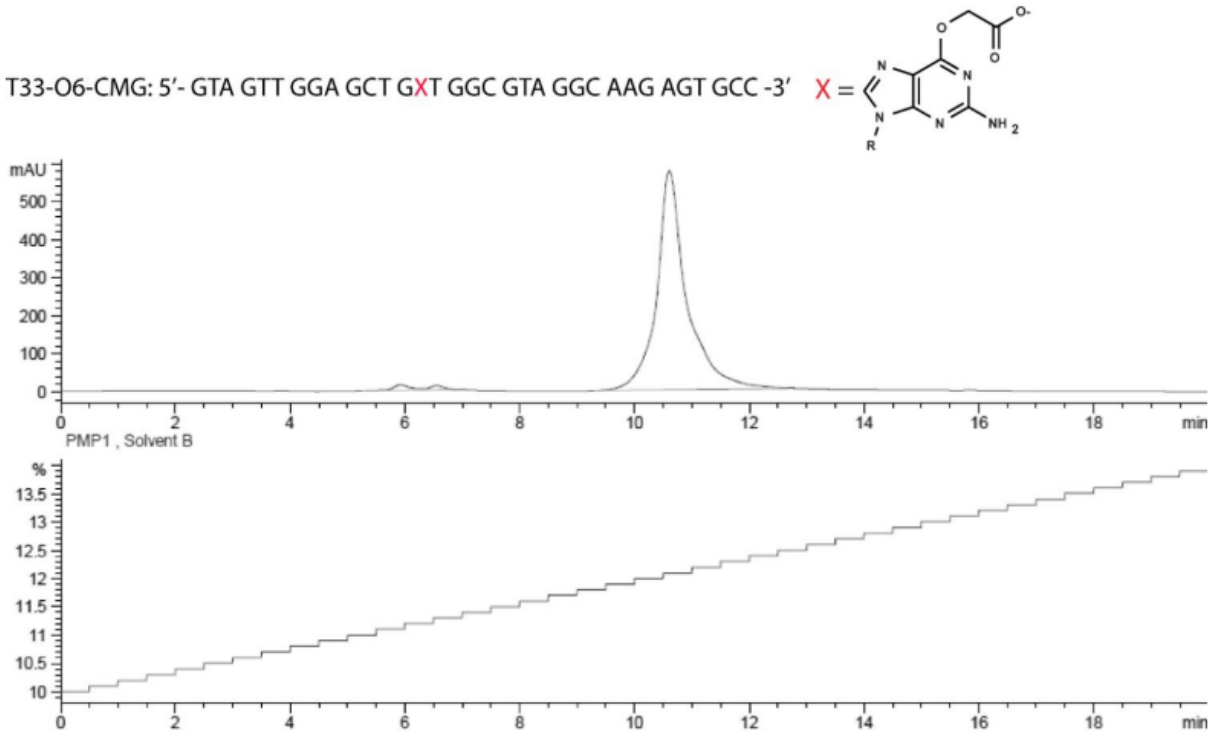


Figure S1: Reversed phase HPLC chromatogram of the 33mer template containing the carboxymethylguanine adduct (T33-O^6 -CMG) with a linear gradient from 10-14% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

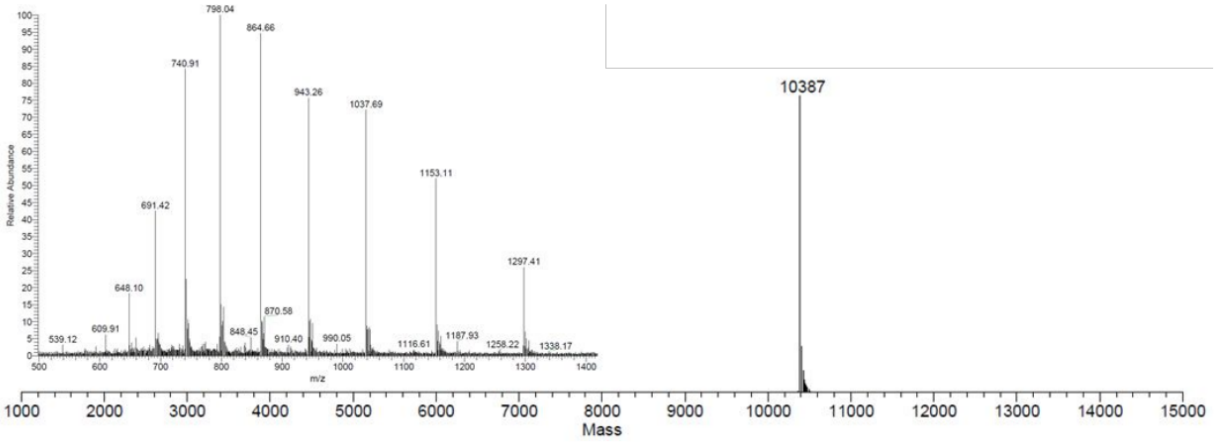


Figure S2: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the carboxymethylguanine adduct (parent mass 10388.62)

Isopentylamine template (T33-*O*^{6}-ISO)

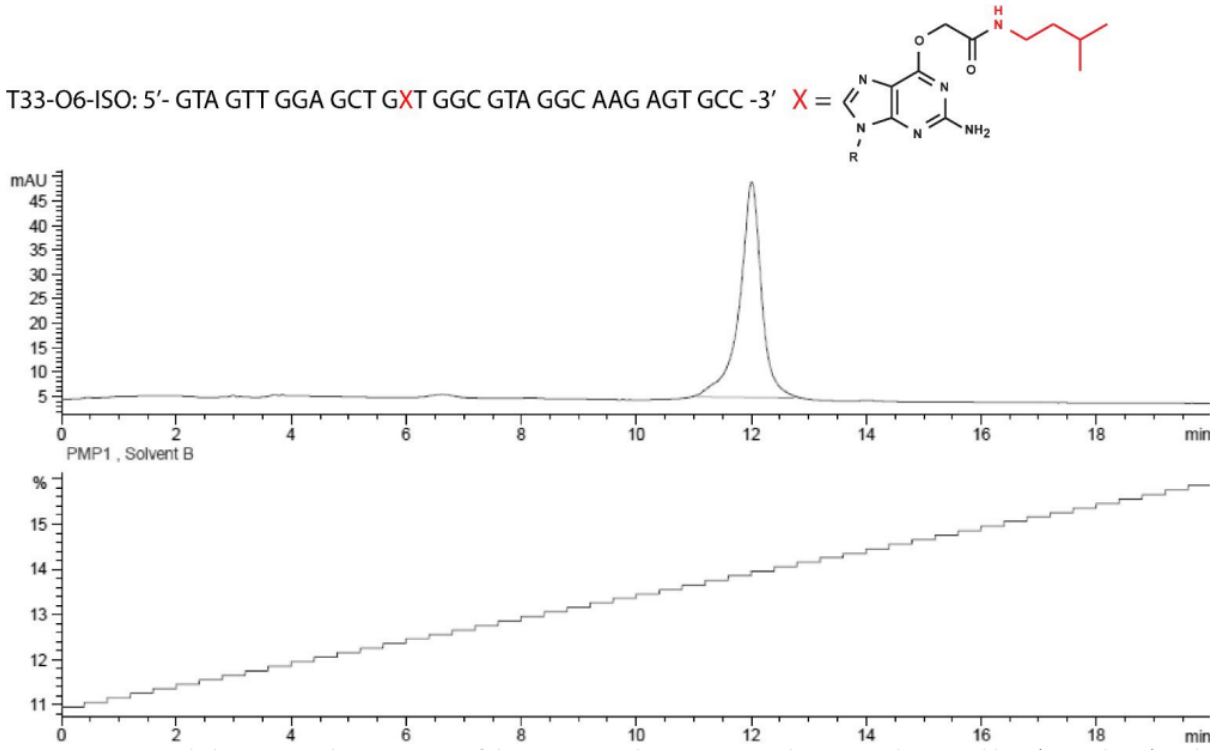


Figure S3: Reversed phase HPLC chromatogram of the 33mer template containing the isopentylamine adduct (T33-O^6 -ISO) with a linear gradient from 11-16% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

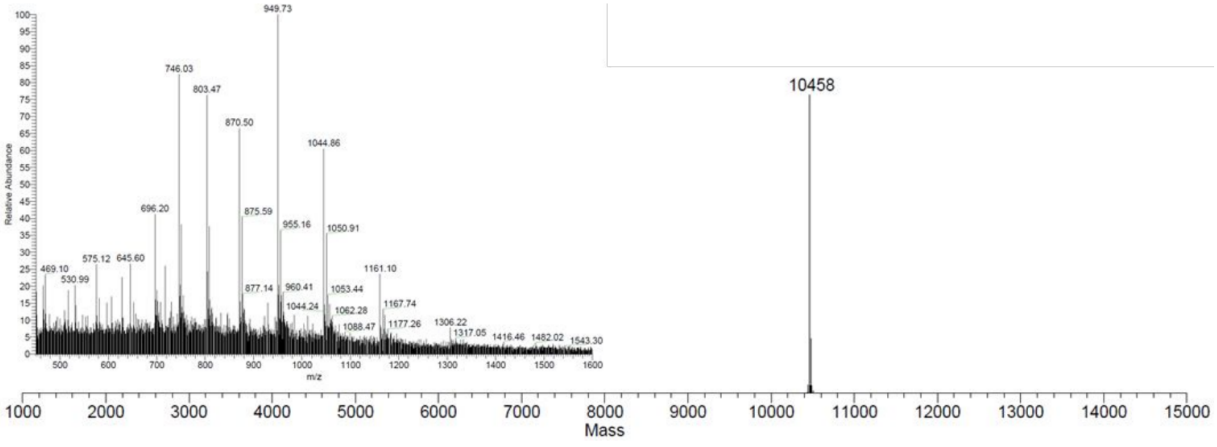


Figure S4: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the isopentylamine adduct (parent mass 10457.62)

Allylamine template (T33-*O*^{6}-All)

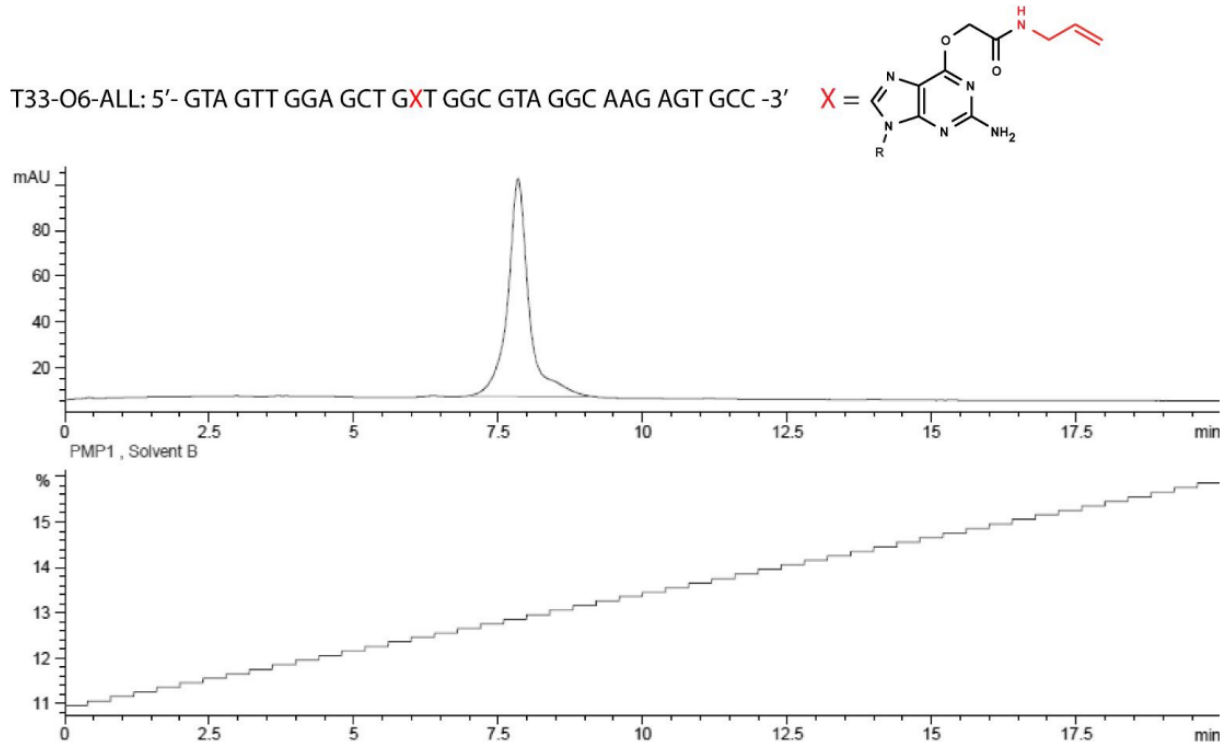


Figure S5: Reversed phase HPLC chromatogram of the 33mer template containing the allylamine adduct (T33-O^6 -ALL) with a linear gradient from 11-16% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

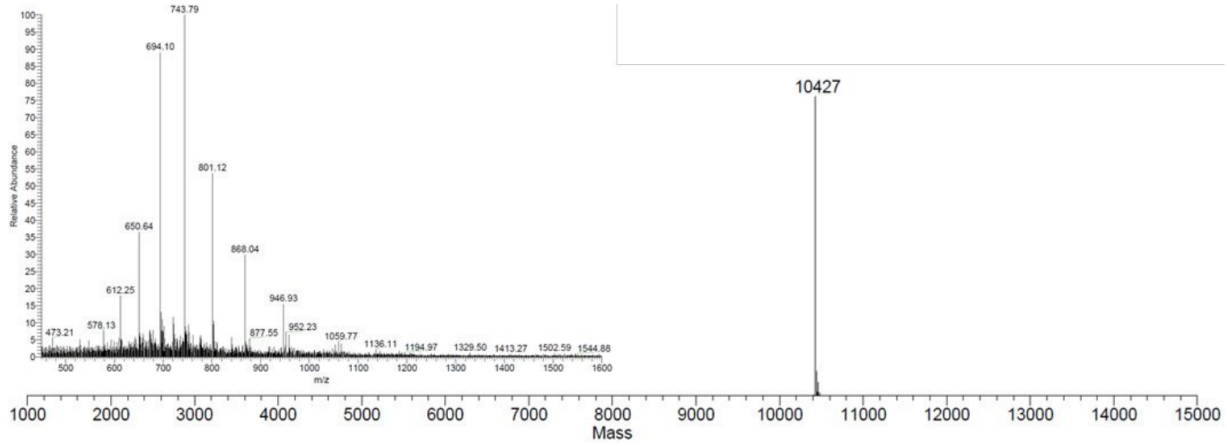


Figure S6: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the allylamine adduct (parent mass 10427.62)

Ethylenediamine template (T33-*O*^{6}-ETH)

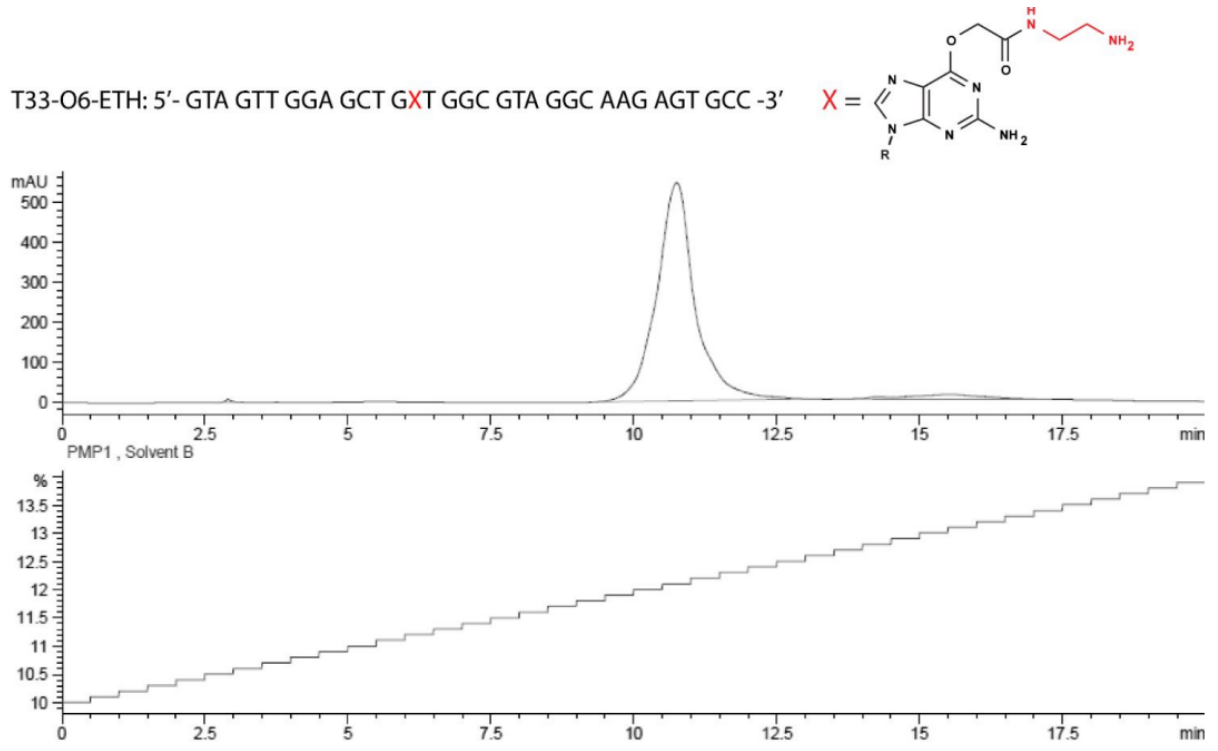


Figure S7: Reversed phase HPLC chromatogram of the 33mer template containing the ethylendiamine adduct (T33-O^6 -ETH) with a linear gradient from 10-14% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

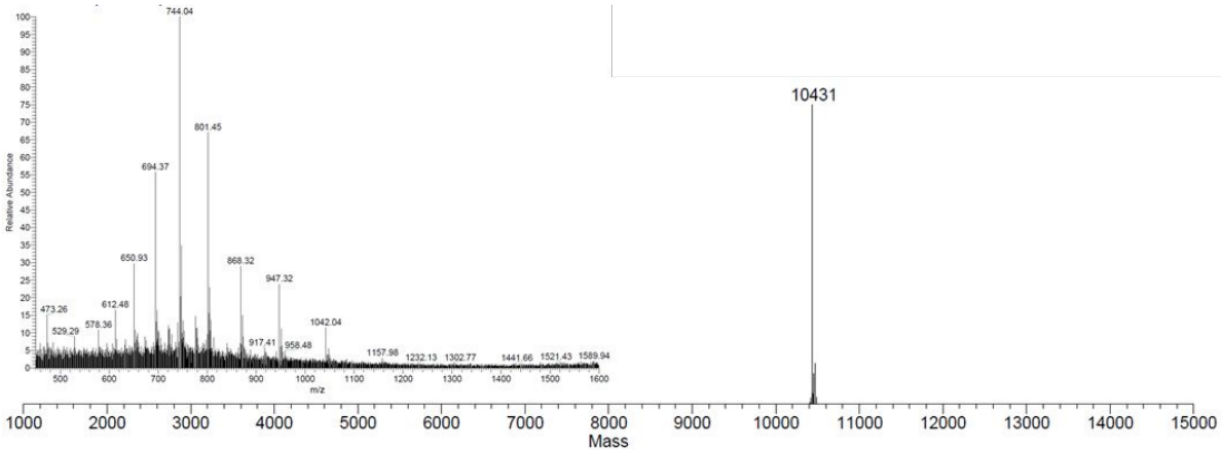


Figure S8: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the ethylendiamine adduct (parent mass 10430.62)

Carboxamide template (T33-*O*^{6}-CAR)

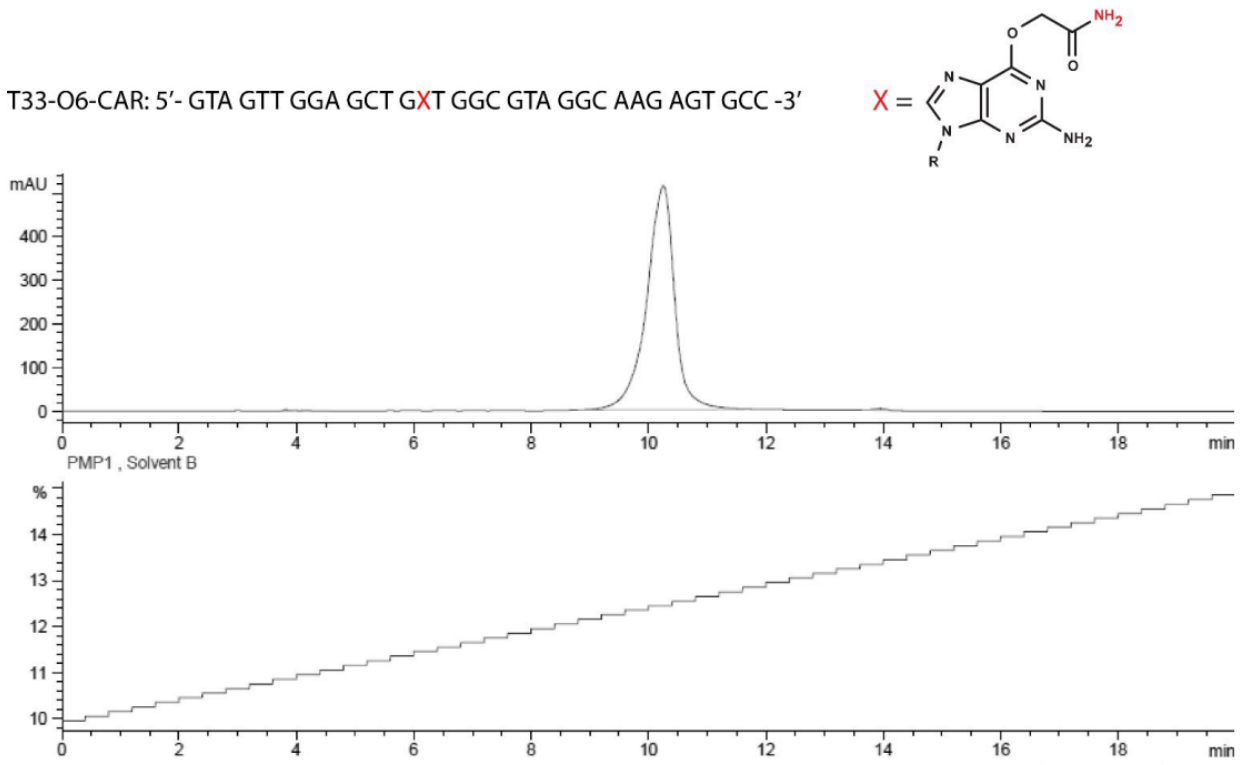


Figure S9: Reversed phase HPLC chromatogram of the 33mer template containing the carboxamide adduct (T33-*O*^{6} -CAR) with a linear gradient from 10-15% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

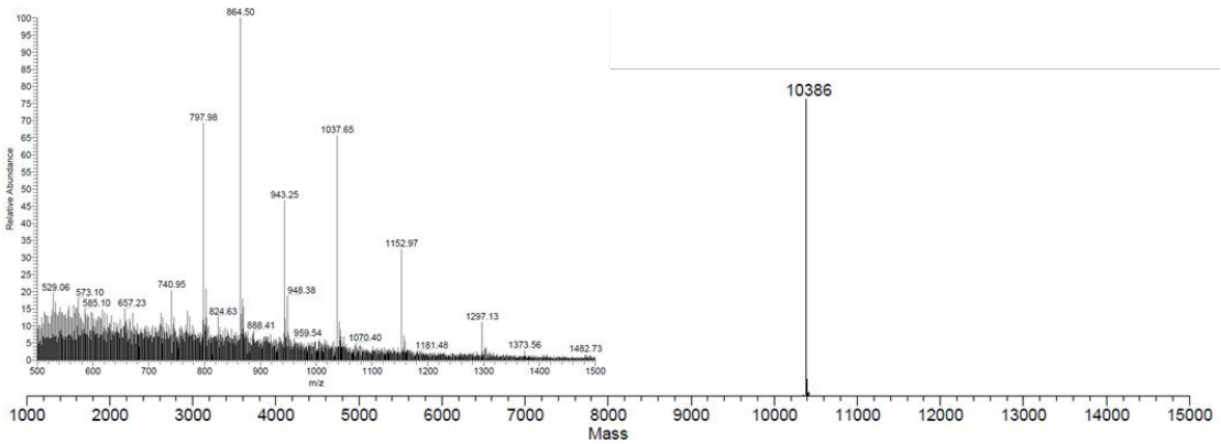


Figure S10: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the carboxamide adduct (parent mass 10387.03)

β-Alaninemethylester template (T33-O^6 -BAM)

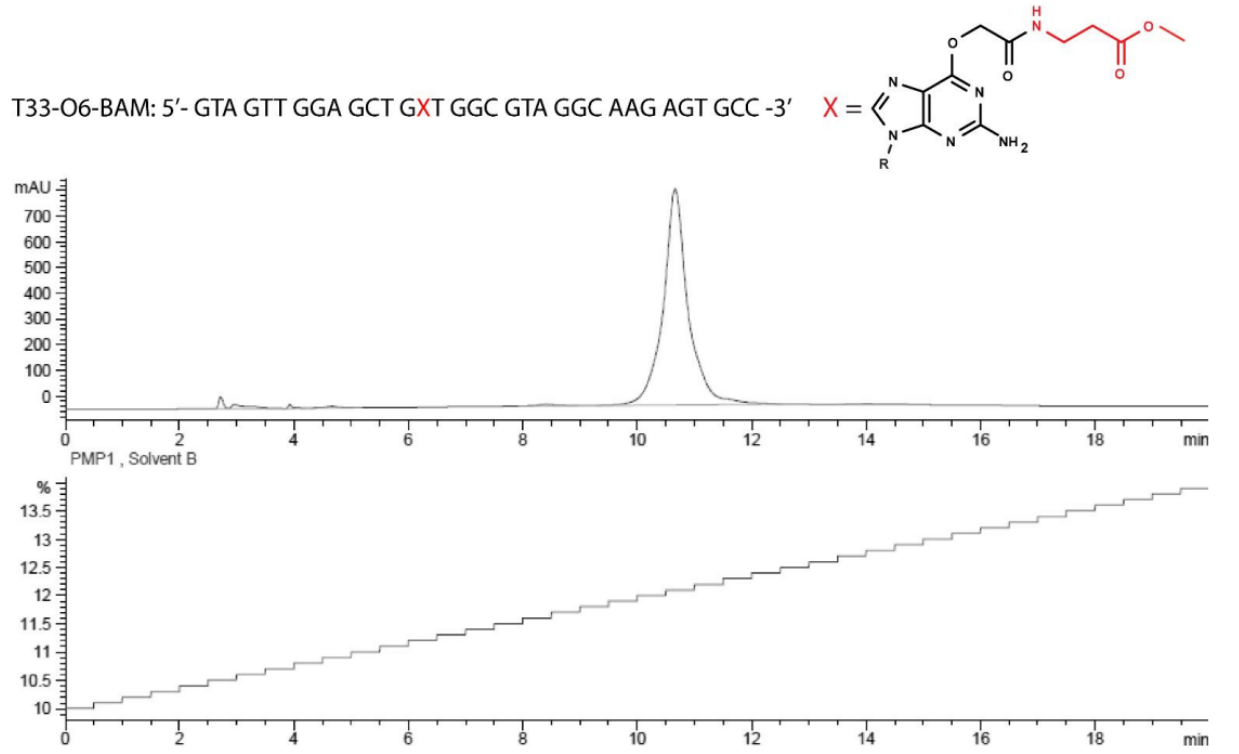


Figure S11: Reversed phase HPLC chromatogram of the 33mer template containing the β-alaninemethylester adduct (T33-O^6 -BAM) with a linear gradient from 10-14% (v/v) acetonitrile in 50 mM triethylammonium acetate over 20 min

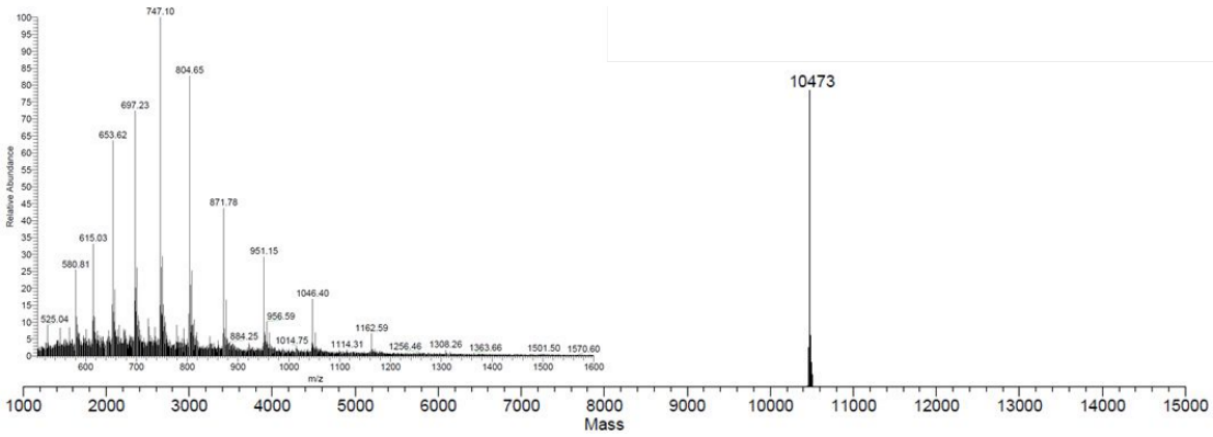


Figure S12: Mass spectrometry characterization and Magtran deconvolution of the 33mer template containing the βalaninemethylester adduct (parent mass 10473.62)

Surface volume

The surface volume was calculated with Chem3D software (Cambridge Software Ltd.) base on the Connolly surface algorithm (solvent excluded area).

Table S1. Surface volume of the modification at the O^6-position of guanine. The modifications are ranked in ascending order of their surface volume.

Size rank	Modification	Volume of modification at O^6-position of G (Å^3)
1	O^6-CMG	51
2	CAR	58
3	ALL	100
4	ETH	108
5	BAM	131
6	ISO	143

Polymerase primer extension experiments

Figure S13: Polymerase extension products observed from Pol η in the bypass of O^6 -CMG and the synthetic carboxymethyl adducts ETH, CAR, ISO, ALL, and BAM

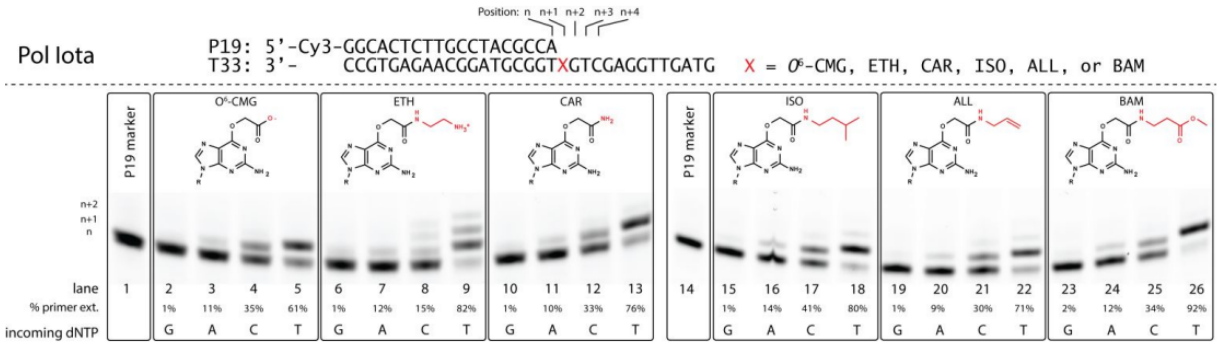


Figure S14: Polymerase extension products observed from Pol ι in the bypass of O^6 -CMG and the synthetic carboxymethyl adducts ETH, CAR, ISO, ALL, and BAM

Figure S15: Polymerase extension products observed from Pol β in the bypass of O^6-CMG and the synthetic carboxymethyl adducts ETH, CAR, ISO, ALL, and BAM

Computed structures of ETH in the post-insertion step

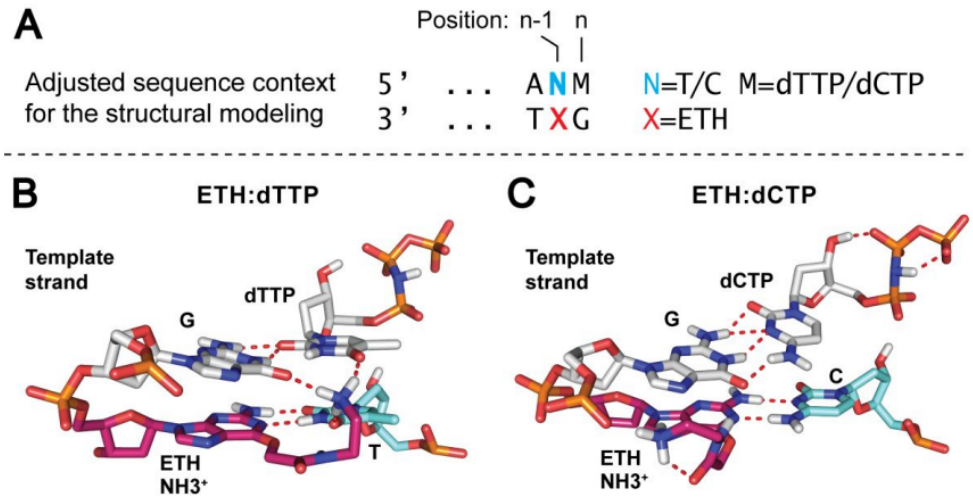


Figure S16: Computational modeling of ETH in the post-insertion step of Pol κ. (A) Adjusted sequence context with X = ETH and N= T or C and M = the incoming dCTP or dTTP. Structural models of (B) ETH:dTTP and (C) ETH:dCTP.

Computed structure of O^6-CMG and ETH in an uncharged state

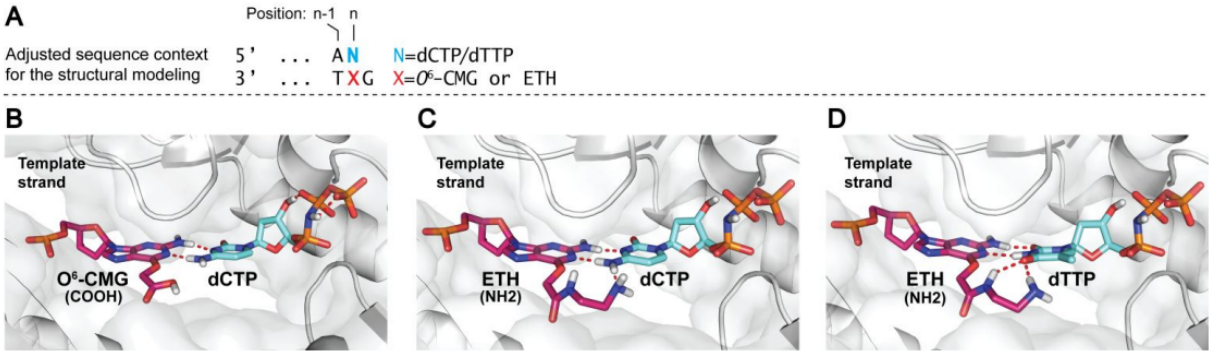


Figure S17: Computational modeling of O^6-CMG or ETH in an uncharged state in the Pol κ active site. (A) Adjusted sequence context with X = O^6-CMG or ETH and N = the incoming dCTP or dTTP. Structural models of (B) O^6-$CMG_{(COOH)}$:dCTP, (B) ETH(NH2):dCTP and (D) ETH(NH2):dTTP.

Catalytic efficiencies plotted as a function of the calculated volume of the modification at the O^6 -position.

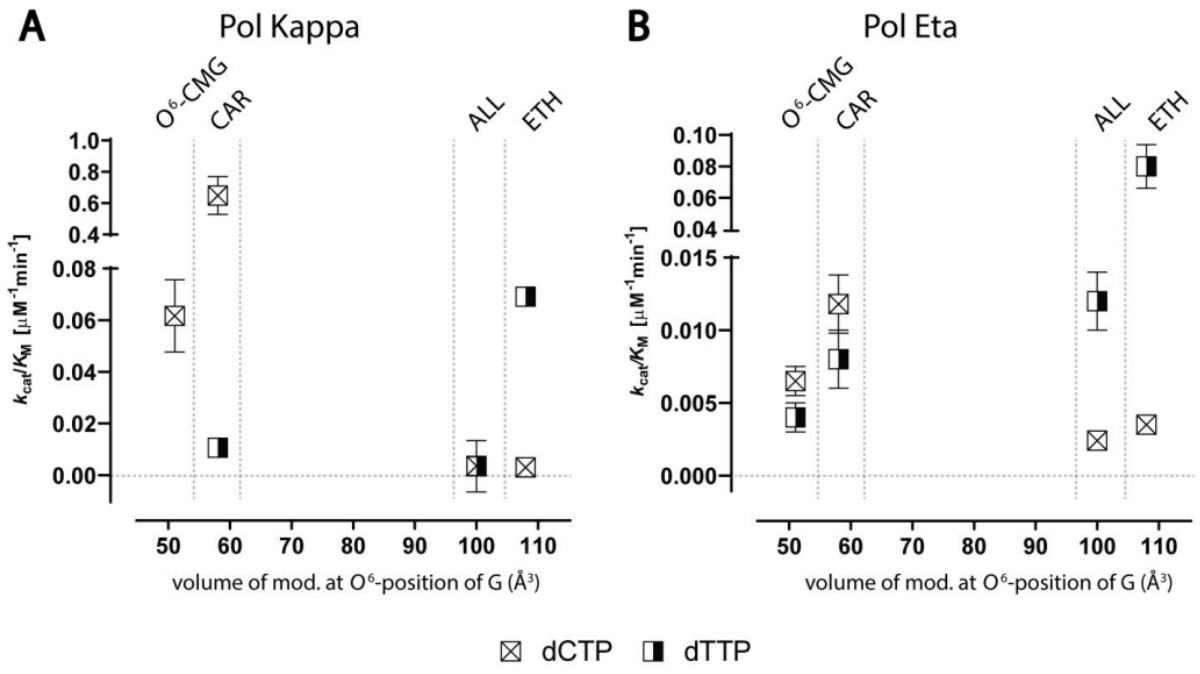


Figure S18. Primer extension efficiencies (k_{cat}/K_M) plotted as function of the increasing volumes of the modification at the O^6-position of G. The volumes were calculated based on the Connolly surface algorithm (solvent excluded area).

Polymerase primer extension experiments

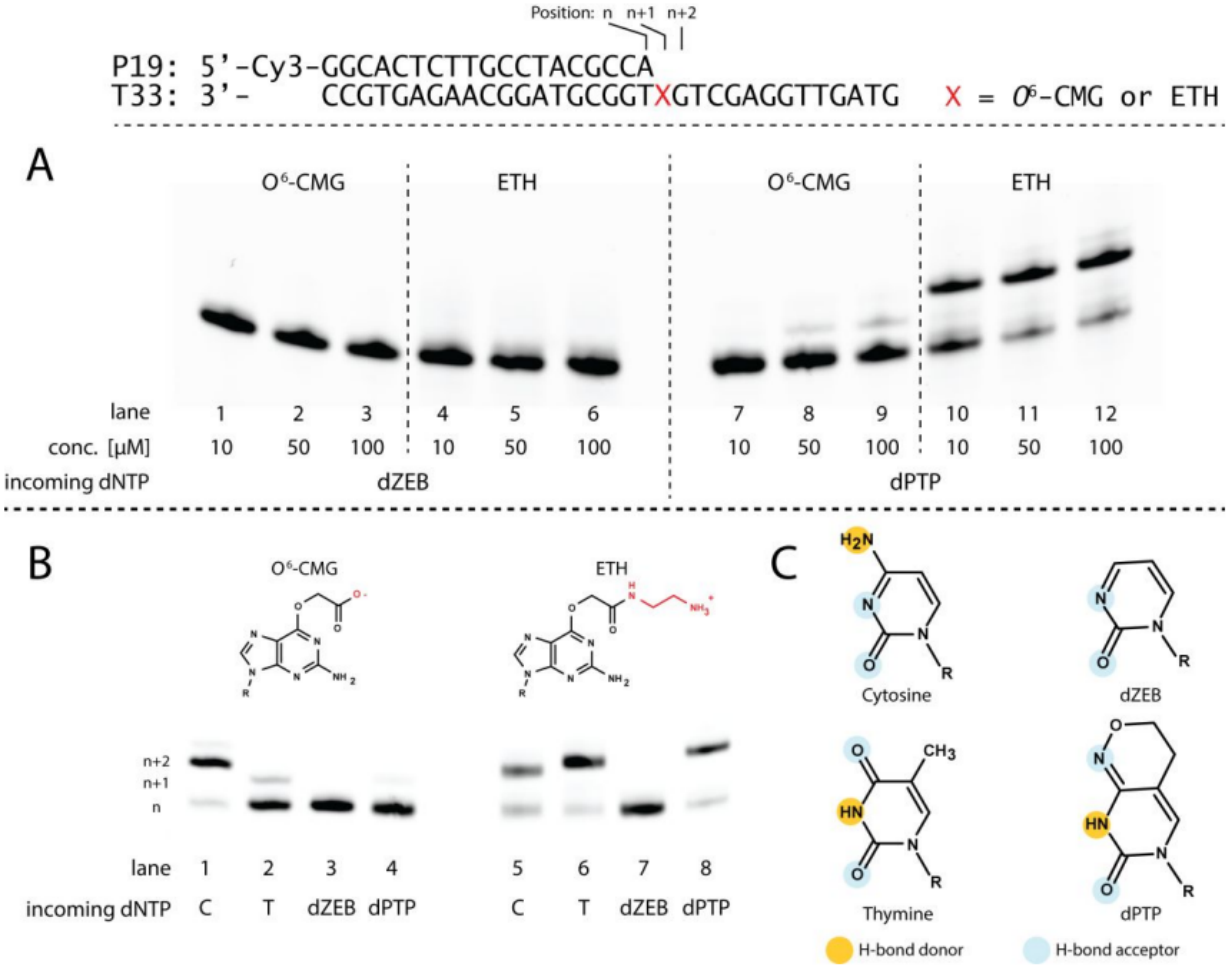


Figure S19. Polymerase extension products observed from Pol κ in the bypass of O^6-CMG and ETH with the modified nucleosides dZEB and dPTP. (A) Primer extension reactions in the presence of increasing concentrations of dZEB and dPTP (10, 50, or 100 μM). (B) Direct comparison of incorporation of dCTP, dTTP, dZEB or dPTP opposite O^6-CMG or ETH. (C) Structural renderings of cytosine, thymine, dZEB and dPTP.

Computed structures of O^6-CMG and ETH paired with dZEB or dPTP in the Pol κ active site

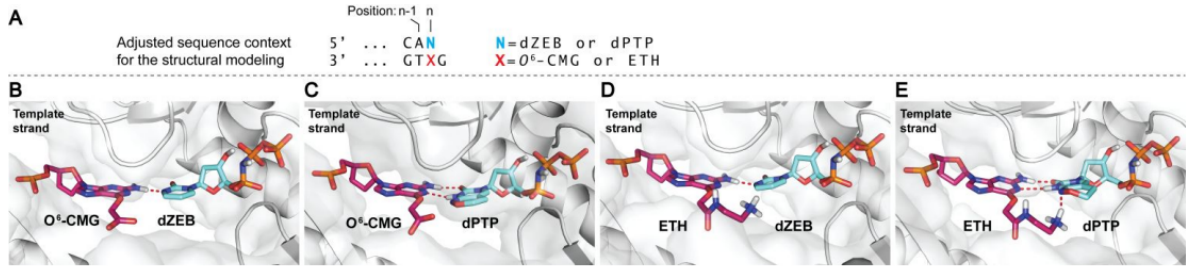
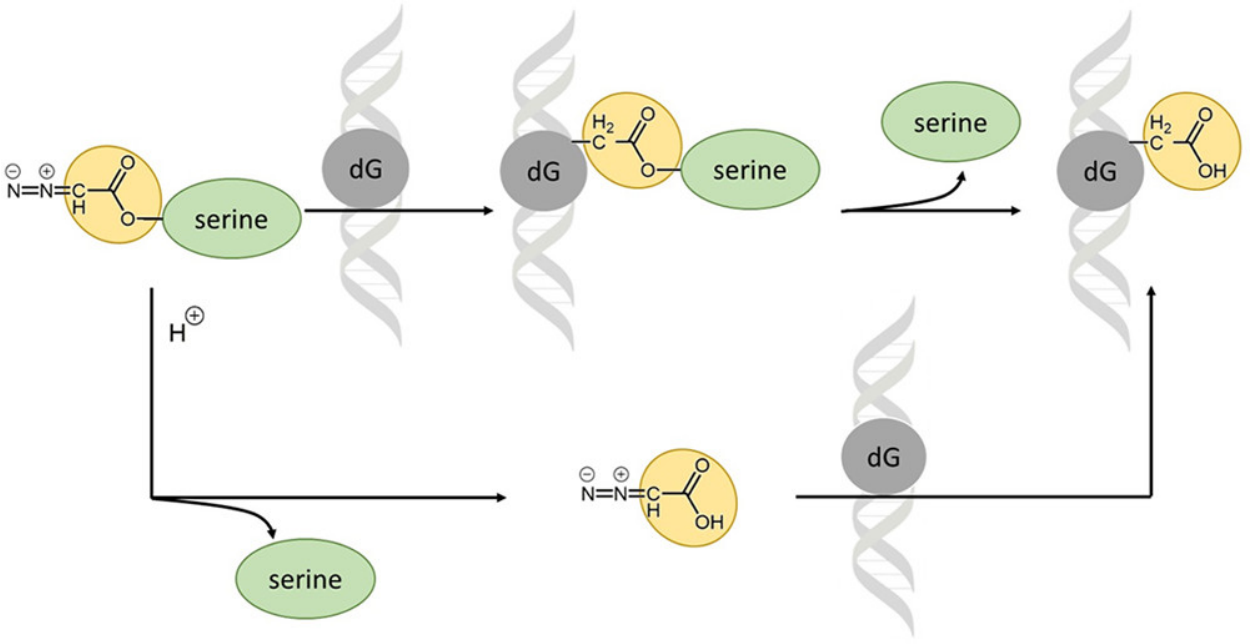


Figure S20: Computational modeling of O^6-CMG or ETH in the Pol κ active site. (A) Adjusted sequence context with X = O^6- CMG or ETH and N = the incoming dZEB or dPTP. Structural models of (B) O^6-CMG:dZEB, (C) O^6-CMG:dPTP, (D) ETH:dZEB, and (E) ETH:dPTP.

Appendix B: Direct alkylation of deoxyguanosine by azaserine leads to O^6-Carboxymethylguanosine



Reproduced with permission from

Susanne M. Geisen, Claudia M. N. Aloisi, Sabrina M. Huber, Emma S. Sandell, Nora A. Escher, and Shana J. Sturla*, Direct alkylation of deoxyguanosine by azaserine leads to O^6-Carboxymethylguanosine, *Chem. Res. Toxicol.*, **2021**, 34, 6, 1518-1529

E.S. Sandell performed experiments on oligonucleotides and wrote the relevant experimental and results section. S.M. Geisen performed stability experiments of azaserine in vitro and performed NMR and LC-MS measurements, interpreted data and wrote the manuscript. C. M. N. Aloisi gave scientific advice, interpreted data and wrote the manuscript. S. M. Huber performed stability measurements of azaserine in biological matrices and wrote the relevant experimental and results section. N.A. Escher performed cell studies. S. J. Sturla desvised the project and wrote the manuscript.

6.1 Abstract

The O^6-alkylguanosine adduct O^6-carboxymethyldeoxyguanosine (O^6-CMdG) has been detected at elevated levels in blood and tissue samples from colorectal cancer patients and from healthy volunteers after consuming red meat. The diazo compound L-azaserine leads to the formation of O^6-CMdG as well as the corresponding methyl adduct O^6-methyldeoxyguanosine (O^6-MedG) in cells and is therefore in wide use as a chemical probe in cellular studies concerning DNA damage and mutation. However, there remain knowledge gaps concerning the chemical basis of DNA adduct formation by L-azaserine. To characterize O^6-CMdG formation by L-azaserine, we carried out a combination of chemical and enzymatic stability and reactivity studies supported by liquid chromatography tandem mass spectrometry for the simultaneous quantification of O^6-CMdG and O^6-MedG. We found that L-azaserine is stable under physiological and alkaline conditions as well as in active biological matrices but undergoes acid-catalyzed hydrolysis. We show, for the first time, that L-azaserine reacts directly with guanosine (dG) and oligonucleotides to form an O^6-serine-CMdG (O^6-Ser-CMdG) adduct. Moreover, by characterizing the reaction of dG with L-azaserine, we demonstrate that O^6-Ser-CMdG forms as an intermediate that spontaneously decomposes to form O^6-CMdG. Finally, we quantified levels of O^6-CMdG and O^6-MedG in a human cell line exposed to L-azaserine and found maximal adduct levels after 48 h. The findings of this work elucidate the chemical basis of how L-azaserine reacts with deoxyguanosine and support its use as a chemical probe for N-nitroso compound exposure in carcinogenesis research, particularly concerning the identification of pathways and factors that promote adduct formation.

6.2 Introduction

The structural integrity of nucleobases in DNA is essential for the correct functioning of cellular processes. The reaction of DNA with electrophilic exogenous and endogenous chemicals can result in DNA adducts that are mutagenic and initiate carcinogenesis. The DNA adducts O^6-carboxymethyldeoxyguanosine (O^6-CMdG) and O^6-methyldeoxyguanosine (O^6-MedG) are pro-mutagenic lesions of current interest in the context of colon carcinogenesis associated with meat consumption.[1-5] Both adducts were elevated in blood and tissue samples from colorectal cancer patients.[6-9] Further, O^6-CMdG was detected in colonic exfoliated cells from human feces[10] and in human blood samples from individuals on high-red-meat diets.[11] The association with meat, particularly red, was confirmed in other studies where O^6-CMdG was detected at higher levels in tissues from volunteers with a high-red-meat diet compared to a vegetarian one[12] and in studies where digestion of red and white meat was modeled in vitro. [13-15] The formation of O^6-CMdG and O^6-MedG was also observed after exposure of DNA to *N*-nitroso compounds (NOCs),[11,16,17] which are present in red and cured meat [18] or can be derived from nitrogenous dietary precursors via nitrite-mediated nitrosation of secondary amines that occurs largely endogenously.[19] If unrepaired, the high miscoding potential of O^6-MedG to pair with thymine results in GC → AT transition mutations upon replication.[4,5] For O^6-CMdG, the misincorporation of thymine, and to a smaller extent of adenine, could be demonstrated, [20,21] therefore providing a mechanistic basis for the GC→ AT transition and GC → TA transversion mutations found in the p53 gene after treatment with a carboxymethylating agent and replication in yeast. [22] Interestingly, these are also the two most common mutations found in colorectal cancer. [22,23]

Carboxymethlyating agents, such as compounds derived from *N*-carboxmethyl-NOCs and other chemical probes, [24] promote the formation of O^6-CMdG (Figure 1) and to a lesser extent O^6-MedG. [11,17,25] For example, L-azaserine and potassium diazoacetate are commonly used to generate O^6-CMdG and O^6-MedG. [8,10,17,26,27] In previous studies, potassium diazoacetate induced higher absolute levels of O^6-CMdG and O^6-MedG than L-azaserine, but the carboxymethyl:methyl ratio was lower for potassium diazoacetate. [16,17] The average level of O^6-CMdG and O^6-MedG adducts after exposure to potassium diazoacetate was similar in cells and naked DNA, suggesting that no enzymatic activation is required for potassium diazoacetate. However, the high activity of potassium diazoacetate limits its use as a carboxymethylating agent, and no reproducible results were obtained in cells.[10] Moreover, it is difficult to determine potassium diazoacetate concentration, due to its high instability and fast conversion during characterization. Hence, a nominal concentration is normally assumed based on quantitative production from alkaline hydrolysis of ethyldiazoacetate, [10,28] but a true yield to our knowledge has never been determined. Based on technical limitations in the use of potassium diazoacetate, L-azaserine seems to be the preferred chemical for inducing O^6-CMdG and O^6-MedG, and exposure of cells to L-azaserine resulted in reproducible adduct levels. [27]

L-azaserine was discovered in 1954 as the active component of a crude filtrate from a culture of Streptomyces bacteria with antibiotic activity and potent antitumor activity in rodents. [29] It was positive in the Ames test and induced adenocarcinoma in rats[30] via the formation of DNA damage.[31] The formation of O^6-CMdG by L-azaserine was demonstrated by reaction with calf thymus DNA (ctDNA) and exposure in cells.[16,17,26,27] Despite the long-standing knowledge that L-azaserine gives rise to O^6-CMdG, details concerning the mechanism of this reaction remain limited.

It has been hypothesized that L-azaserine gives rise to carboxymethyl adducts by conversion to diazoacetate, which can carboxymethylate DNA (Figure 1). Two potential mechanisms for the conversion of L-azaserine in diazoacetate have been suggested: enzymatic β-elimination of L-azaserine to diazoacetate and pyruvate (Figure 1C) or spontaneous hydrolysis to yield diazoacetate and L-serine (Figure 1B). Evidence for enzymatic β-elimination of L-azaserine was established using enzymatic extracts from mouse liver, and a stoichiometric relationship could be demonstrated for L-azaserine disappearance and glycolic acid, ammonia, and pyruvate formation, with the last two being the hydrolysis products of the β-elimination intermediate 2-aminoacrylate.[32] The responsible enzyme was presumed to be a dehydrogenase.[33] Decomposition of L-azaserine to pyruvate and ammonia was also demonstrated in bicarbonate buffer, but only in the presence of pyridoxal ion as a cofactor. A pyridoxal ion-catalyzed β-elimination of L-azaserine via a Schiff base was suggested, resulting in diazoacetate and aminoacrylate that subsequent hydrolyses to pyruvate and ammonia.[34] However, the observation that O^6-CMdG and O^6-MedG form from L-azaserine and ctDNA in the absence of enzymes suggested an additional, nonenzymatic mechanism of L-azaserine breakdown that could yield DNA adducts. This mechanism has been proposed to involve spontaneous hydrolysis to yield serine and diazoacetate (Figure 1B). [17] A further possibility is that L-azaserine may react directly with DNA and that subsequent hydrolysis yields O^6-CMdG (Figure 1A).

The objective of this study was to elucidate the chemical mechanism of O^6-CMdG formation by L-azaserine and establish a chemical basis for its use to address the corresponding DNA damage in cells. Thus, we evaluated the stability of L-azaserine under varying pH conditions and in the presence of active biological matrices. We characterized in detail the basis of O^6-CMdG formation from the direct reaction of L-

azaserine with guanosine and oligonucleotides. Finally, we quantified the formation of O^6-CMdG and O^6-MedG by L-azaserine in human colon epithelial cells (HCEC) on the basis of dose and time by establishing a mass spectrometry approach for the simultaneous and sensitive quantification of O^6-CMdG and O^6-MedG. The findings suggest a new direct reaction mechanism between L-azaserine and dG that gives rise to an O^6-Ser-CMdG intermediate, which undergoes spontaneous hydrolysis to form O^6-CMdG. Further, the characterization of adduct formation in cells as a function of L-azaserine dose and time of exposure supports the use of L-azaserine as a chemical probe in future studies addressing biological consequences of O^6-CMdG.

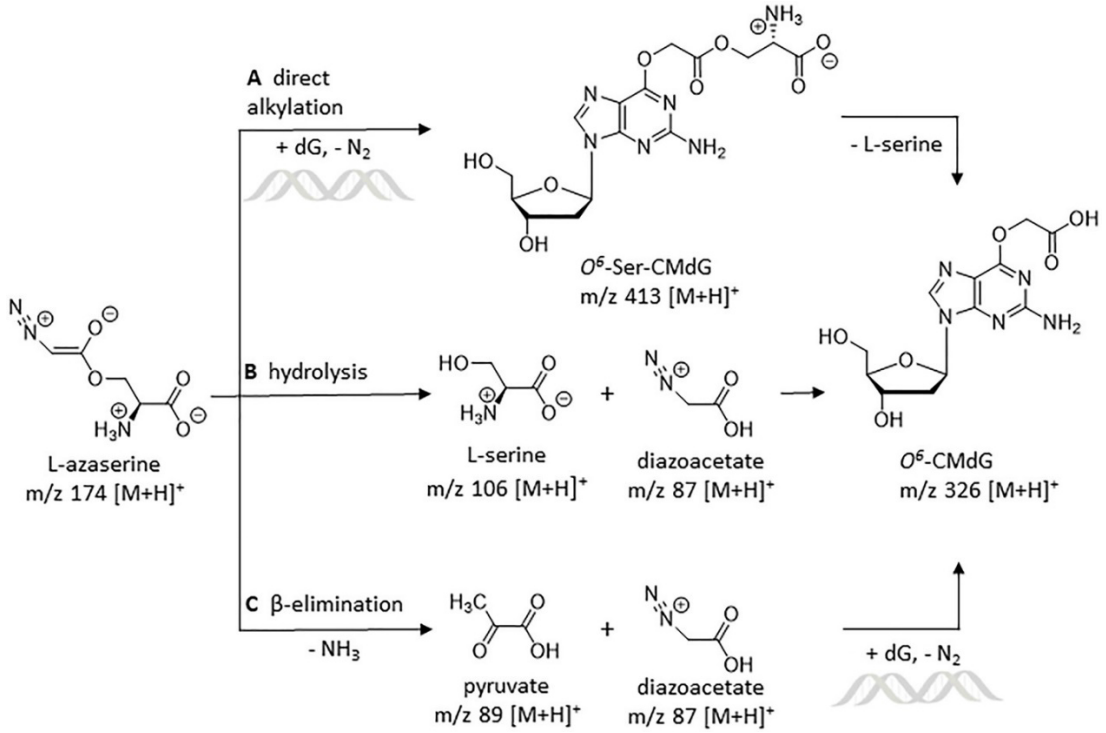


Figure 1 Proposed mechanisms for L-azaserine-induced O^6-CMdG formation. (A) Nucleophilic attack of dG on L-azaserine results in an O^6-Ser-CMdG intermediate followed by hydrolysis of L-serine to yield O^6-CMdG. (B) L-azaserine decomposes spontaneously to L-serine and diazoacetate, which subsequently carboxymethylates dG. (C) Enzymatic β-elimination of L-azaserine results in pyruvate and diazoacetate.

6.3 Experimental Procedures

Methods. Potassium diazoacetate was synthesized via alkaline hydrolysis of ethyldiazoacetate as previously described, and neutral solutions for further usage were obtained by neutralizing with 0.1 M HCl.[10,17,22] L-azaserine was purchased from Abcam (Cambridge, UK) and purity confirmed by ^{1}H NMR. Ethyldiazoacetate was obtained from Pfaltz & Bauer Inc. (Waterbury, Connecticut), deoxygunaosine (dG) from Fluorochem (Hadfield, UK), and labeled ^{15}N5-dG from Cambridge Isotope Laboratories Inc., (Tewskbury, Massachusetts) and were used without further purification. N^2-MedG standard was purchased from Carbosynth (Compton, UK). Oligonucleotide was ordered high-performance liquid chromatography (HPLC)-pure from Eurogentec. If not stated otherwise, all chemicals were purchased from Sigma-Aldrich (Buchs, Switzerland) and used without further purification. Milli-Q water was from a

Millipore Synergy UV device. Mobile phases for liquid chromatography (LC) were obtained from Sigma-Aldrich (Buchs, Switzerland), and mobile phases for LC coupled with mass spectrometry (LC-MS) were purchased from Biosolve Chimie (Dieuze, France). All media and buffers for cell culture were purchased from Invitrogen (Carlsbad, California). HCEC-1CT cell lines were kindly provided by Prof J.W. Shay (University of Texas, USA).[35] NMR analysis was performed on a Bruker Biospin 400 MHz NMR instrument (Bruker, Billerica, Massachusetts), and chemical shifts are reported in parts per million (ppm, δ) referenced to the chemical shift of the respective solvent. NMR solvents were purchased from Sigma-Aldrich (Buchs, Switzerland). NMR files were evaluated by Mnova version 11 (Mestrelab Research, Santiago de Compostela, Spain). Reversed-phase HPLC analysis was performed on either an Agilent 1100 or 1200 Series HPLC (Agilent Technologies, Santa Clara, California).

Data were analyzed and plotted using GraphPad Prism, Version 8 (GraphPad Software Inc., San Diego, California). Applied statistical tests are specified for each data set in the corresponding figure caption.

pH Stability of L-azaserine. L-azaserine (25 mM) was dissolved in 500–980 µL of 0.1 M buffer solutions at a pH ranging from pH 2 to 11: pH 2, 2.4, and 3.4 glycine-HCl buffer, pH 5.8 sodium phosphate, pH 7.2 sodium phosphate buffer, pH 8.5 carbonate buffer, and pH 10.5 tris-buffer. Solutions were analyzed by reversed-phase HPLC with a diode array detector set to monitor absorbance at λ = 254 nm (L-azaserine) and 210 nm (serine, pyruvate) on a Luna C18 column, 4.6 × 250 mm, particle size 5 µm, and pore size 100 Å (Phenomenex, Torrance, California) using H2O as mobile phase A and ACN as mobile phase B. Chromatographic separation was achieved by an initial isocratic hold at 100% A for 5 min, followed by a gradient to 40% B in 10 min at a flow rate of 1 mL/min. Injection volume was 10 µL. The starting concentration of L-azaserine was chosen to allow for detection of 0.5% and quantification of 2% hydrolysis to l-serine (LOQ 0.5 mM l-serine). Each sample (pH 2–11) was analyzed every 3 h for a total of 120 h. The analytical method was shortened for samples at pH 2–5.8 to an isocratic hold at 100% A for 6.0 min. l-Serine eluted at retention time 2.6 min, pyruvate at 2.9 min, and L-azaserine at 3.8 min. A calibration curve for L-azaserine and l-serine from 0.5, 1, 6, 12, 25, and 50 mM was analyzed.

Activity of Whole Cell Lysate toward L-azaserine**.** L-azaserine (5 µL of 2 mM stock solution) was added to freshly prepared whole cell lysate (100 µL) derived from human lung epithelial cells (BEAS-2B cells, ATCC CRl-9609) cultured in bronchial epithelial basal medium (BEBM, Lonza) at 37 °C in 5% CO2 atmosphere. To obtain the lysate, growth medium was removed when the cells reached a confluency of ∼80%, and 1 × DPBS (10 mL) was added to wash the cells. Per 100 mm cell culture dish, 1 mL of CelLytic M (Sigma, C2978) was added followed by incubation at room temperature for 30 min on a shaker. Cell lysate was collected using a cell scraper and transferred into a 1.5 mL Eppendorf tube for immediate use. The mixture was incubated at 37 °C between 30 min and 24 h. After incubation, the mixture was loaded onto a prewashed 10 kDa MWCO filter unit (VWR, prewashed with H2O (300 µL) by centrifugation at 12,000g for 5 min) and centrifuged at 12,000g for 10 min. The flow through was loaded onto a Phenomenex Strata-X 33 µm Polymeric Reversed Phase 30 mg/1 mL SPE column that was prewashed with 2 × 1 mL MeOH followed by 2 × 1 mL H2O. The L-azaserine-containing fraction was eluted using 400 µL of H_2O, concentrated to dryness by vacuum centrifugation and redissolved in 50 µL H2O for HPLC analysis using the identical method as described in the pH Stability of L-azaserine section. Results are shown in Figure S1A. As a positive control, fluorescein diacetate (FDA) was also tested as a substrate. Stock solutions (20 mg/mL in acetone) were stored at −20 °C, and immediately before use, an aliquot of FDA stock solution was diluted in acetone to a

final concentration of 2 mg/mL. FDA (2 µL of 2 mg/mL acetone solution) was added to fresh cell lysate (400 µL) or to CelLytic M (400 µL) to obtain a final FDA concentration of 24 µM. The resulting mixtures were allowed to react at room temperature for 2 h and applied to prewashed 10 kDaA MWCO filter units and centrifuged (12,000g for 10 min). The 50 µL aliquots of the flow throughs were loaded on a black 96 half-well plate for analysis. The experiment was performed three times.

To quantify the amount of FDA converted to fluorescein, a stock solution was prepared fresh at a concentration of 2 mg/mL in acetone. This stock solution was used to generate a fluorescein calibration line in CelLytic M with fluorescein concentrations of 75 µM, 50 µM, 25 µM, 10 µM, 5 µM, 1 µM, and 0.5 µM. The 50 µL aliquots (triplicates) of each concentration were loaded on a black 96 half-well plate for analysis. Fluorescence measurements were acquired on an Infinite M200 Pro plate reader from Tecan at 18 °C using excitation wavelength of 490 nm and an emission wavelength of 525 nm. No fluorescence readout was obtained for the FDA in CelLytic control, confirming that there is no spontaneous hydrolysis of FDA occurring.

Incubation of L-azaserine with Porcine Liver Esterase. Lyophilized powder of porcine liver esterase (Sigma, E3019) was reconstituted in 10 mM borate buffer pH 8.0 at a concentration of 40 mU/µL. L-azaserine (100 µM) in 10 mM borate buffer pH 8.0 was incubated at 37 °C with varying amounts of porcine liver esterase (0.04–4 U) between 30 min and 24 h. After incubation, the esterase was removed by filtration over a 10 kDa MWCO filter unit at 12,000g for 5 min. The flow through was directly used for HPLC quantification of L-azaserine using the identical method as described in the pH Stability of L-azaserine section. Results are shown in Figure S1B. FDA was tested as positive control as described in Activity of Whole Cell Lysate toward L-azaserine section.

Reaction of an Oligonucleotide with L-azaserine. A 9-mer oligonucleotide (TTTTGTTTT, 5 mM) was allowed to react with L-azaserine (50 mM) in 1 × sodium phosphate buffer (20 µL) at 37 °C for 24 h. The resulting mixture (20 µL) was directly analyzed by reversed-phase HPLC with a diode array detector set to monitor absorbance at λ = 260 nm on a Clarity oligo-RP column, 250 × 4.6 mm, particle size 5 µm, pore size 110 Å (Phenomenex, Torrance, California) using TEAA 50 mM as mobile phase A and acetonitrile (ACN) as mobile phase B. Chromatographic separation was achieved by a gradient from 92% A to 85% A in 30 min at a flow rate of 1 mL/min. The unreacted oligonucleotide eluted with a retention time of 21 min, followed by three small peaks with retention times 21.5, 22.0, and 22.7 min (Figure S2A). Fractions of these peaks were collected and further analyzed by electrospray ionization mass spectrometry on a Velos Linear Ion Trap (Thermo Scientific, Waltham, Massachusetts). The electrospray mass spectrum for the material that eluted at 21.96 min in the UV-chromatogram corresponded to the anticipated m/z value of the electrospray ionization mass for the O^6-Ser-CMG-oligonucleotide (2845.46, Figure S2C). To verify the structure and sequence, this material was further fragmented to obtain its MS2 mass spectrum. The collision energy was set at 30 with parent m/z 948, which predominantly generated a-B and w type fragment ions (Figure S2B). The CID mass fragments corresponded to the calculated mass fragments of the O^6-Ser-CMG-oligonucleotide, indicating the anticipated sequence context and the serine modification to be on guanine. HRMS was calculated for TTTT-G-TTTT: $[M - H]^-$m/z 2699.813, found 674.12, 898.994, 1348.94, Magtran deconvolution 2700.26, TTTT- O^6-CMG-TTTT: $[M - H]^-$m/z 2759.7, found 688.32, 918.46, 1377.94, Magtran deconvolution 2758.48, TTTT- O^6-Ser-CMG-TTTT: $[M - H]^-$m/z 2846.69, found 568.044, 710.483, 947.45, 1421.45, Magtran deconvolution 2845.46

Preparation of Standards for Mass Spectrometry. O^6-CMdG was prepared by the copper carbene-based method described by Geigle et al. [36] In short, 2.5 mg (9.35 µmol) of dG was allowed to react with 10.7 mg (93.5 µmol) of ethyldiazoacetate in the presence of 300 µg (1.87 µmol) CuSO4. O^6-CMdG was obtained by alkaline hydrolysis followed by reversed-phase HPLC purification with a diode array detector set to monitor absorbance at λ = 280 and 250 nm on a Luna C18 column, 4.6 × 250 mm, particle size 5 µm, pore size 100 Å (Phenomenex, Torrance, California) using 0.1% HOAc as mobile phase A and ACN as mobile phase B. Chromatographic separation was achieved by an initial isocratic hold at 100% A for 10 min, followed by a gradient to 16% B in 45 min at a flow rate of 1 mL/min. Injection volume was 20 µL. The product (RT 41.3 min, λ = 280, 250 nm) was obtained in 40% yield as a white powder. HRMS (ESI) calculated for $C_{12}H_{15}N_5O_6$: $[M + H]^+$m/z 326.1101, found 326.1095, MS2 calculated for $C_7H_7N_5O_3$: [M + H – Gua]+m/z 210.0628, found 210.0617. 1H NMR (Figure S3) matched published data. $^{15}N_5$-O^6-CMdG was prepared by the same protocol using 2.5 mg (9.35 µmol) of dG $^{15}N_5$-dG as starting material, yielding 1 mg (40% yield) of a white powder of $^{15}N5$-O^6-CMdG. HRMS (ESI) calculated for $C_{12}H_{15}{}^{15}N_5O_6$: $[M + H]^+$m/z 331.0952, found 331.0944, MS2 calculated for $C_7H_7{}^{15}N_5O_3$: $[M + H – Gua]^+$m/z 215.0479, found 215.0469 $[M + H – Gua]^+$.

O^6-MedG was synthesized by the method reported by Reza et al.[37] In short, 0.57 g (2 mmol) dG was allowed to react in 10 mL of pyridine with 4.3 g (79.6 mmol) of sodium methoxide in 300 mL of MeOH. The product was purified by reversed-phase HPLC with a diode array detector set to monitor absorbance at λ = 260 nm on a Luna C18 column 10 × 250 mm, particle size 5 µm, pore size 100 Å (Phenomenenx, Torrance, California) using 5% ACN in H2O as mobile phase A and ACN as mobile phase B. Chromatographic separation was achieved by an initial isocratic hold at 100% A for 15 min, followed by gradient to 35% B in 10 min at a flow rate of 2 mL/min. Injection volume was 200 µL. Product (RT 20.7 min, λ = 260 nm) was obtained in a 60% yield as a light yellow powder. HRMS (ESI) calculated for C11H15N5O4: [M + H]+m/z 282.1202, found 282.1197, MS2 calculated for C6H7N5O: [M + H – Gua]+m/z 166.0729, found 166.0715. 1H NMR (Figure S4) matched the published data.[37] D3-O^6-MedG was synthesized accordingly with 0.19 g (0.67 mmol) of dG and 1.51 g (26.5 mmol) of D3-sodium methoxide in 100 mL MeOD (both Armar chemicals, Switzerland). Product was purified by reversed-phase LC as described for unlabeled O^6-MedG and obtained as a light-yellow powder in 60% yield. HRMS (ESI) calculated for $C_{11}H_{12}D_3N_5O_4$: $[M + H]^+$m/z 285.1391, found 285.1383, MS2 calculated for C6H5D3N5O: $[M + H – Gua]^+$m/z 169.0918, found 169.0904. We did not observe any D/H exchange by analyzing a 100 nM D3-O^6-MedG standard solution by nanoLC-ESI-HRMS2 and following the intensity in the extracted ion chromatograms (XICs) for D3-O^6-MedG (m/z 285.1383 $[M + H]^+$) and O^6-MedG (m/z 282.1197 $[M + H]^+$, not detectable (nd)) over time of the analysis.

Reaction of dG with L-azaserine. dG (5 µM in 500 µL Milli-Q or tris-buffer (10 mM) containing 2 mM MgCl2, at pH 7, 8, 9, and 10.6) was combined with L-azaserine or potassium diazoacetate (5 mM), and the reaction mixture was stirred for 0.5, 1, 2, 3, and 24 h at 37 °C. After the indicated time, samples were processed for quantification of O^6-CMdG and O^6-MedG and analyzed by nanoLC-ESI- $HRMS^2$, as described in the detailed method validated for analysis of cell samples elaborated below, with the following changes: Trapping time was set to 0.1 min. Resolution was set to 60 K for the orbitrap mass analyzer in MS^1, 68–444 m/z, and 1 microscan. Target inclusion list was expanded to include O^6-Ser-CMdG m/z 413.1421 $[M + H]^+$, O^6-Ser-CMG m/z 297.0947 $[M + H]^+$, L-azaserine m/z 174.0515 $[M + H]^+$, serine m/z 106.0504 $[M + H]^+$, and dG m/z 268.1046 $[M + H]^+$. The mass tolerance was set to ±50 ppm. Fragmentation of a O^6-Ser-CMdG was

targeted in MS^2 XICs with neutral loss of dR, O^6-Ser-CMG m/z 297.0947 $[M + H - dR]^+$, and serine fragment,[38-40] with m/z 251.0893 $[M + H - (dR + H2O + CO)]^+$. Mass tolerance for all XICs was ±10 ppm.

Synthesis of O^6-Ser-CMdG. Four mM dG was dissolved in 1200 µL of 0.1 M sodium phosphate buffer at pH 7.2. Twenty-five mM L-azaserine was added, and the mixture was allowed to react at 22 °C for 120 h while being continuously analyzed by reversed-phase HPLC with a diode array detector s set to monitor absorbance at λ = 254 nm (L-azaserine, dG, G, O^6-CMdG, O^6-Ser-CMdG), 210 nm (L-serine), and 280 nm (O^6-CMdG, O^6-Ser-CMdG) on a Luna C18 column, 4.6 × 250 mm, particle size 5 µm, pore size 100 Å (Phenomenex, Torrance, California) using H2O as mobile phase A and ACN as mobile phase B. Chromatographic separation was achieved by an initial isocratic hold at 100% A for 5 min, followed by a gradient to 45% B in 45 min at a flow rate of 1 mL/min. Injection volume was 10 µL. All eluting peaks were identified by m/z on a Velos Linear Ion Trap (Thermo Scientific, Waltham, Massachusetts) and in case of L-azaserine, dG, guanine, and l-serine with coelution of respective standards and corresponding DAD spectrum. L-azaserine eluted at retention time 3.9 min (m/z 174), dG at 16.8 min (m/z 268), l-serine at 2.8 min (m/z 106), and guanine at 11.4 min (m/z 152) A peak at retention time 7.9 min had m/z 297 matching the calculated mass for Ser-G, while the peak of interest eluted at retention time of 18.7 min with m/z 413 as calculated for O^6-Ser-CMdG (Figure S5A). This peak was collected over 200 runs and characterized. HRMS was calculated for $C_{15}H_{20}N_6O_8$: $[M + H]^+$m/z 413.1421, found 413.1411 (Figure S5B), and MS^2 was calculated for $C_{10}H_{12}N_6O_5$: $[M + H - Gua]^+$m/z 297.0947, found 297.0938 ^{1}H NMR in deuterium oxide confirmed presence of O^6-Ser-CMdG (Figure S5C). ^{1}H NMR (400 MHz, deuterium oxide) δ 8.14 (s, 1H, Ar–H), 6.38 (t, J = 8.2 Hz, 1H, 1′-H), 5.07 (s, 1H, NH2–CH), 4.83–4.81 (m, 2H, O–CH2–COO), 4.67–4.62 (m, 1H, 4′-H), 4.18–4.13 (m, 1H, 3′-H), 3.93–3.84 (m, 2H, O–CH2–CNH2), 3.83–3.74 (m, 2H, 5′-H), 2.84 (dt, J = 14.7, 7.4 Hz, 1H, 2′-H), 2.57–2.46 (m, 1H, 2′-H).

Hydrolysis of O^6-Ser-CMdG. O^6-Ser-CMdG was dissolved in 0.1 M phosphate buffer at pH 7.2, and the solution was analyzed by reversed-phase HPLC every hour for a total of 30 h and peaks identified as described above but using HOAc as mobile phase A and a gradient consisting of an initial isocratic hold for 10 min at 100% A, followed by a gradient to 16% B in 45 min. A product eluting at 42.5 min (Figure S6A) was isolated and confirmed to be O^6-CMdG on the basis of ^{1}H NMR (Figure S6C) and HRMS (ESI) calculated for $C_{12}H_{15}N_5O_6$: $[M + H]^+$ 326.1101, found 326.1094, matching the authentic standard.

Cell Culture of HCEC Cells and L-azaserine Exposure. Cells were maintained as monolayers in 10 cm dishes in a humidified, 5% CO2 atmosphere at 37 °C. Media consisted of 80% DMEM and 20% M199 Earle's salt medium, supplemented with 2% Hyclone fetal bovine serum (Hyclone Laboratories Inc., San Angelo, Texas), 25 ng/L epidermal growth factor, 1 µg/L hydrocortisone, 10 µg/L insulin, 2 µg/L transferrin, 50 µg/L gentamycin, and 0.9 ng/L sodium selenite. Cells were regularly confirmed to be mycoplasma free using the MycoAlert Kit (Lonza, Basel, Switzerland).[41] L-azaserine stock solutions were prepared in Milli-Q water. O^6-Benzylguanine (O^6-BG) stock solutions were prepared in dimethyl sulfoxide (DMSO), and the final concentration of DMSO was 0.1%. To assess cell viability, cells were seeded in 96-well plates at a density of 1 × 104 and exposed to increasing L-azaserine concentrations (0, 1, 10, 50, 250, 750, 1000, 2500 µM) for 120 h. Cells were exposed to TritonX (Sigma-Aldrich, Buchs, Switzerland) as a positive control for cytotoxicity. Cell survival was measured using the CellTiterGlo assay (Promega, Madison, Wisconsin) following the manufacturer's instructions.[42]

For DNA adduct analysis, cells were seeded in 10 cm dishes at a density of 2 × 106. Cells were exposed to 0, 125, 250, 500, and 1000 μM L-azaserine for 4 h to evaluate thhe dose–response relationship and with 500 μM L-azaserine for 0.5, 4, 10, 24, 48, 72, 96, and 120 h to determine the time-dependent O^6-CMdG and O^6-MedG formation. Cell viability, dose–response, and time course studies were each performed three independent times.

DNA Isolation and Sample Preparation for Adduct Analysis by Mass Spectrometry. DNA was isolated from cell pellets using the QIAamp DNA Mini Kit (©QIAGEN, Switzerland). Extracted and dried DNA (20 μg) was dissolved in buffer (10 μL, 10 mM Tris pH 7, 2 mM MgCl2). To this sample was added 50 μL of a master mix solution comprised of 2.5 U/μg DNA benzonase, 3 mU/μg DNA phosphodiesterase I, 2 U/μg DNA alkaline phosphatase, 2 nM D3-O^6-MedG and 4 nM 15N5-O^6-CMdG in the aforementioned buffer. A solution of ctDNA (20 μg in 10 μL buffer) and a blank sample (10 μL buffer) were combined with 50 μL of the master mix as controls for DNA digestion and background signal, respectively. Following the digestion, MS grade H_2O (350 μL) was added to the sample (total volume 410 μL), and enzymes were removed by molecular weight filtration with a cutoff of 10 kDa (VWR, Radnor, Pennsylvania). An aliquot (60 μL) of the resulting solution was removed for dG quantification by HPLC (described below). Samples were concentrated to dryness in a Thermo Scientific Savant Universal SpeedVac Vacuum system, resuspended in 200 μL MS grade H2O + 0.1% formic acid (FA), and further processed by solid-phase extraction on a C18 cartridge, Sep-Pak Vac 1 cm3 (50 mg) (Waters, Milford, Massachusetts) using MS grade CH_3OH, CH_3OH + FA, and H_2O. Cartridges were washed with 2 × 1 mL CH_3OH and 1 mL CH_3OH + 0.1% FA followed by 2 × 1 mL H_2O + 0.1% FA. The samples were loaded on the cartridge, washed with 2 × 1 mL H_2O + 0.1% FA and 1 mL 3% CH_3OH, and eluted with 2 × 500 μL of 80% CH_3OH in H_2O. The eluted fractions were dried under vacuum and dissolved in 10 μL H_2O for subsequent MS analysis.

High-Resolution Mass Spectrometry Method for Simultaneous Quantification of O^6-CMdG and O^6-MedG. Simultaneous quantification of O^6-CMdG and O^6-MedG was achieved on an Orbitrap Fusion Lumos (Thermo Scientific, Waltham, Massachusetts) equipped with a nano electrospray ionization source and a nanoAcquity UPLC M-class system (Waters, Milford, Massachusetts). Liquid chromatography was performed with an Acquity UPLC M-Class Symmetry C18 Trap column (100 Å, 5 μm, 180 μm x 20 mm) and an Acquity UPLC M-Class HSS T3 column (100 Å, 1.8 μm, 75 μm × 250 mm) at a column temperature of 40 °C. The sample loop volume was 5 μL, with an injection volume of 2 μL. Mobile phases A and B consisted of mass spectrometry grade H2O with 0.1% FA and ACN with 0.1% FA, respectively (Biosolve Chimie, Dieuze, France). The chromatographic method included trapping for 0.5 min at 99.5% A at a flow rate of 15 μL/min. The chromatographic separation was achieved by a 1 min hold at 3% B, followed by a linear gradient from 3% to 45% B over 19 min, at a flow rate of 0.3 μL/min. Nano electrospray ionization was used in positive mode, 2.2 kV at 270 °C. Mass spectrometry-based detection was performed in parallel reaction monitoring (PRM) with higher-energy collisional dissociation in orbitrap technology (HCD-OT) detection mode with the following settings: The RF lens was set to 35%. Resolution of the orbitrap was set to 120 K in MS1, 160–400 m/z, and 1 microscan. AGC target was set to 4 × 105 with a maximum injection time of 100 ms. Targeted inclusion list was comprised of O^6-CMdG with m/z 326.1096 $[M + H]^+$, heavy $^{15}N_5$-O^6-CMdG with m/z 331.0946 $[M + H]^+$, O^6-MedG with m/z 282.1995 $[M + H]^+$, and heavy D3-O^6-MedG with m/z 285.1384 $[M + H]^+$. The mass tolerance was set to ±25 ppm. Isolation for MS^2 was performed in the quadrupole with an isolation window of 1.6 m/z and the first mass set to 140 m/z. HCD was used for

fragmentation with a collision energy of 25%. Fragments were detected in the orbitrap with a resolution of 60 K. The AGC target was set to 5×10^4 with a maximum injection time of 118 ms. For quantification, neutral loss of deoxyribose was targeted in XICs using O^6-CMG m/z 210.0622 $[M + H - Gua]^+$, heavy $^{15}N_5$-O^6-CMG m/z 215.0472 $[M + H - Gua]^+$, O^6-MeG m/z 166.0722 $[M + H - Gua]^+$, and heavy D3-O^6-MeG m/z 169.0911 $[M + H - Gua]^+$ with a mass tolerance of ±10 ppm. Calibration standards containing 0.5, 1, 5, 10, 25, 50, and 100 nM of O^6-CMdG and O^6-MedG standards undergoing sample preparation served for quantification (Figure S7).

Specificity was guaranteed by processing a blank sample for each analysis and selectivity by coelution with internal standards and accurate parent and fragment mass. Pure standards were used to determine the limit of detection (LOD) and limit of quantification (LOQ) defined as the amount of analyte generating a signal-to-noise ratio (S/N) of 3 or 10 in the XICs. LOD (S/N > 3) was 5 pM (10 amol) and 1 pM (2 amol) for O^6-CMdG and O^6-MedG, respectively, and LOQ (S/N > 10) was 10 pM (20 amol) for both adducts. We further assessed LOQ in matrix, intra- and interday accuracy (expressed as % recovery), and precision (expressed as the relative standard deviation, RSD %) using quality control samples consisting of calf thymus DNA (ctDNA, 15 µg) or HCEC genomic DNA (15 µg) spiked with increasing amounts of O^6-CMdG and O^6-MedG (0.01, 0.05, 0.1, 1.0, 10, 50 100 and 200 nM). LOQ in matrix was determined to be 50 pM (100 amol) for O^6-CMdG and O^6-MedG. The precision and accuracy for the measurement of O^6-CMdG and O^6-MedG in ctDNA and HCEC genomic DNA as matrix (Table S1) had a relative standard deviation (RSD) ranging from 1.7 to 25.6% (O^6-CMdG) or 0.8 to 23.5% (O^6-MedG) and recovery 83.7 to 113.4% (O^6-CMdG) or 92.1 to 122.6% (O^6-MedG) of the actual adduct level.

dG Quantification by Liquid Chromatography. Quantitation of dG was carried out by reversed-phase HPLC with a diode array detector set to monitor absorbance at λ = 254 nm on a C18 Kinetex column, 2.1 × 150 mm, particle size 2.6 µm, pore size 100 Å column (Phenomenex, Torrance, California) using 3% ACN in H2O as mobile phase A and ACN as mobile phase B. Chromatographic separation was achieved by a gradient from 100% A to 88% A in 6 min at a flow rate of 0.2 mL/min. The injection volume was 20 µL. The retention times for nucleosides were the following: dC 3.5 min, dG 8.4 min, dT 9.2 min, and dA 10.6 min. A calibration curve for dG (0.1, 1, 10, 25, 50, and 100 µM) was analzyed. The amount of dG was then used to calculate the total number of nucleotides based on the assumption that the GC content is 40.9% in human DNA.[43]

6.4 Results

Stability of L-azaserine at Varying pH and Biological Matrices. The hydrolysis of L-azaserine to yield l-serine and diazoacetate has been previously suggested as a basis for the formation of O^6-CMdG (Figure 1B).[17] However, this mechanism raises concern about the wide use of L-azaserine in cell-based studies and whether the compound is stable under the experimental conditions of in vitro toxicological assays. Therefore, we sought to characterize the stability of L-azaserine as a function of pH and in the presence of biological matrices. First, levels of L-azaserine and its hydrolysis product L-serine were measured in solutions ranging in pH from pH 2 to 11 at 25 °C over the course of 5 days. No decomposition was observed between pH 5.8 and 10.5, while at pH 2, L-azaserine was completely hydrolyzed to l-serine after 2 h (Figure S8). The -catalyzed hydrolysis was modeled as a pseudo-first-order reaction (Figure 2A), and the half-life (t1/2) was calculated to be 0.4 h at pH 2, 0.7 h at pH 2.4, and 4.7 h at pH 3.4. Next, we assessed the stability of L-azaserine in the presence of cell culture medium, active whole cell lysates, and excess commercial porcine liver esterase. None of these biological matrices were active in catalyzing any conversion of L-azaserine at 37 °C from 10 min to 24 h (Figure S1). On the other hand, fluorescein diacetate, tested as a positive control at the same concentration, was completely hydrolyzed within a few minutes. Together, these data emphasize that L-azaserine resists hydrolysis under neutral physiological conditions in the presence of biological matrices and suggest against it being the initial step in the reaction of L-azaserine with DNA to give rise to O^6-CMdG.

A

$$rate = \frac{k_{hydrolysis}}{K_a}[H_2O][H^+][\text{L-azaserine}] \quad (1)$$

$$k_{obs} = \frac{k_{hydrolysis}}{K_a}[H_2O][H^+] \quad (2)$$

$$k_{obs} = \frac{ln\,2}{t_{1/2}} \quad (3)$$

$$\frac{k_{hydrolysis}}{K_a}[H_2O] = 0.05\ M^{-1}s^{-1} \quad (4)$$

$$\frac{k_{hydrolysis}}{K_a} = \frac{0.05\ M^{-1}s^{-1}}{55.5\ M} = 0.90 \times 10^{-4} M^{-2}s^{-1} \quad (5)$$

B

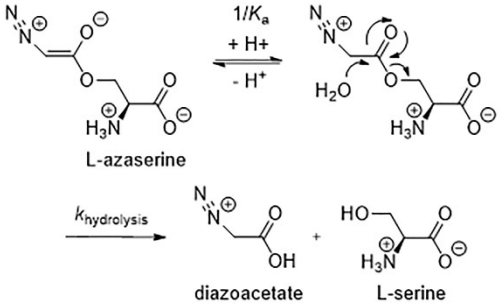


Figure 2. Acid-promoted hydrolysis of L-azaserine. (A) Rate law for hydrolysis of L-azaserine and relationship with observed values. (B) Proposed mechanism of acid-catalyzed hydrolysis of L-azaserine to L-serine

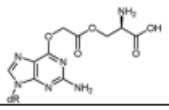
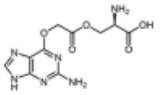
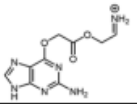
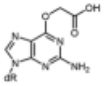
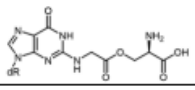
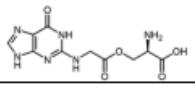
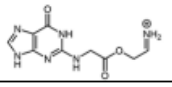
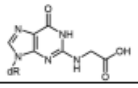
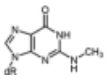
Evidence for O^6-Ser-CMdG as the Intermediate in the Formation of O^6-CMdG by L-azaserine. Having established that L-azaserine does not undergo hydrolysis under physiologically relevant conditions, we considered whether under those conditions L-azaserine may directly react with dG to give rise to the direct adduct O^6-Ser-CMdG, which then yields O^6-CMdG upon spontaneous hydrolysis (Figure 1A). To gain support for this mechanism, we first explored whether O^6-Ser-CMdG could be formed by reacting a 9-mer nucleotide with an excess L-azaserine in sodium phosphate buffer at 37 °C for 24 h. The crude reaction mixture was then analyzed by HPLC for the presence of newly formed oligonucleotide peaks (Figure S2), which were collected and further analyzed by electrospray ionization mass spectrometry. The electrospray mass spectrum and the fragmentation for the oligonucleotide that eluted at 22.0 min in the UV-chromatogram corresponded to the anticipated m/z values of the electrospray ionization mass for the O^6-Ser-CMG-oligonucleotide (Figure S2), providing the first support for the direct reaction of L-azaserine with DNA. To allow robust and unambiguous characterization of the direct adduct, we decided to reduce the

complexity of the reaction and allow dG to react with an excess of L-azaserine in tris-buffer at pH in the range of 7–11 for 0.5–24 h. The resulting reaction mixtures were analyzed for the presence of O^6-Ser-CMdG, O^6-CMdG, and O^6-MedG by nanoLC-ESI-HRMS^2 in PRM mode. We monitored mass transitions accounting for neutral loss of deoxyribose followed by possible fragmentation of serine (Table 1).[38-40] XICs of all reaction mixtures revealed the appearance of two peaks with the mass corresponding to O^6-Ser-CMdG at a retention time of 17.9 and 18.1 min (Figure 3A). Their mass-fragments were the same, consistent with being two positional isomers of Ser-CMdG. Further, two peaks were detected with m/z corresponding to carboxymethylated dG with retention times close to the potential Ser-CMdG isomers, and one of them coeluted with the standard $^{15}N_5$-O^6-CMdG (18.35 min), thereby assigned as O^6-CMdG. Similarly, two peaks were detected with m/z corresponding to isomers of MedG; one peak coeluted with an N^2-MedG standard (18.01 min), while the second peak coeluted with the standard for O^6-MedG (18.75 min). As a negative control, we performed the same reaction with potassium diazoacetate, and no peaks were detected corresponding to Ser-CMdG or fragments thereof, but the same two peaks for each carboxymethylation and methylation of dG as for the L-azaserine treated samples were observed (Figure 3B). None of the targeted m/z values were detected in control samples with any of the single reagents. These results suggest that L-azaserine reacts directly with dG, forming adducts that retain the serine moiety at the O^6- and N^2-positions.

***O*^{6}-Ser-CMdG Decomposes to *O*^{6}-CMdG.** Having gained evidence for the formation of O^6-Ser-CMdG intermediate in the reaction of dG with L-azaserine by mass spectrometry, we further characterized this novel structure and confirmed that it can be converted to O^6-CMdG. Thus, by HPLC analysis, Ser-CMdG was observed in the reaction of dG with L-azaserine already after 30 min and remained detectable up to 120 h (18.7 min, m/z of 413, Figure S5A). The compound was purified by HPLC, and its ^{1}H NMR spectrum was consistent with Ser-CMdG due to the presence of three characteristic peaks at 5.07 (s, 1H, NH_2–CH), 4.83–4.81 (m, 2H, O–CH_2–COO), and 3.93–3.84 ppm (m, 2H, O–CH_2), accounting for five additional Hs in addition to those present in dG (Figure S5C). The HRMS of the isolated peak revealed m/z 413.1411 (RT 17.56 min), corresponding to Ser-CMdG (Figure 3C). After isolation of Ser-CMdG by HPLC, analysis of the sample by mass spectrometry always revealed the presence of O^6-CMdG, but not N^2-CMdG, (Figure 3C), suggesting that the Ser-CMdG adduct eluting at retention time 17.56 is O^6-Ser-CMdG.

Having established that O^6-Ser-CMdG forms as a direct product of the reaction between dG and L-azaserine, we characterized further its aforementioned apparent hydrolysis to yield O^6-CMdG. At pH 7.2, the disappearance of O^6-Ser-CMdG (HPLC retention time 23 min) coincided with the appearance of O^6-CMdG (HPLC retention time 43 min; Figure S6, identity confirmed by coelution with the authentic standard and consistent HRMS data (Figure 4)). We concluded that O^6-Ser-CMdG is an intermediate in L-azaserine-induced O^6-CMdG formation, thereby suggesting a possible mechanism for the formation of O^6-CMdG in DNA under physiological conditions in which L-azaserine is stable.

Table 4. Targeted m/z with proposed structures for ***O*⁶**-Ser-CMdG and transitions

Compound[a]	*m/z* MS^n	Structure[b]	RT (min)	diazoacetate	L-azaserine	control	LC RT 18.7
***O*⁶-Ser-CMdG**	413.1421 MS^1		18.14	nd	413.1415	nd	413.1419
***O*⁶-Ser-CMG**	297.0947 MS^2		18.14	nd	297.0941	nd	297.0937
***O*⁶-Ser-CMG – (H_2O+CO)**	251.0893 MS^2		18.15	nd	251.0883	nd	251.0884
***O*⁶-CMdG**	326.1096 MS^1		18.37	326.1096	326.1094	nd	326.1099
$^{15}N_5$-*O*⁶-CMdG	331.0952 MS^1		18.37	331.0944	331.0944	331.0944	331.0944
***O*⁶-MedG**	282.1195 MS^1		18.84	282.1196	282.1187	nd	282.1195
D_3-*O*⁶-MedG	285.1391 MS^1		18.84	285.1383	285.1383	285.1383	285.1383
***N*²-Ser-CMdG**	413.1421 MS^1		17.91	nd	413.1418	nd	nd
***N*²-Ser-CMG**	297.0947 MS^2		17.91	nd	297.0948	nd	nd
***N*²-Ser-CMG – (H_2O+CO)**	251.0893 MS^2		17.89	nd	251.0882	nd	nd
***N*²-CMdG**	326.1096 MS^1		18.03	326.1095	326.1095	nd	nd
***N*²-MedG**	282.1195 MS^1		18.00	282.1195	282.1197	nd	nd

[a] *O*⁶ substitution was confirmed by isotopically labeled *O*⁶-CMdG and *O*⁶-MedG standards. *N*² -coordination is confirmed by *N*²-MedG standard.

[b] Structures and m/z of *O*⁶-CMdG and *O*⁶-MedG as hydrolysis products are included. Detected structures are indicated for diazoacetate, L-azaserine, and control samples from reaction mixture with dG and for the isolated compound by reversed-phase liquid chromatography (LC RT 18.7 min). nd = not detectable

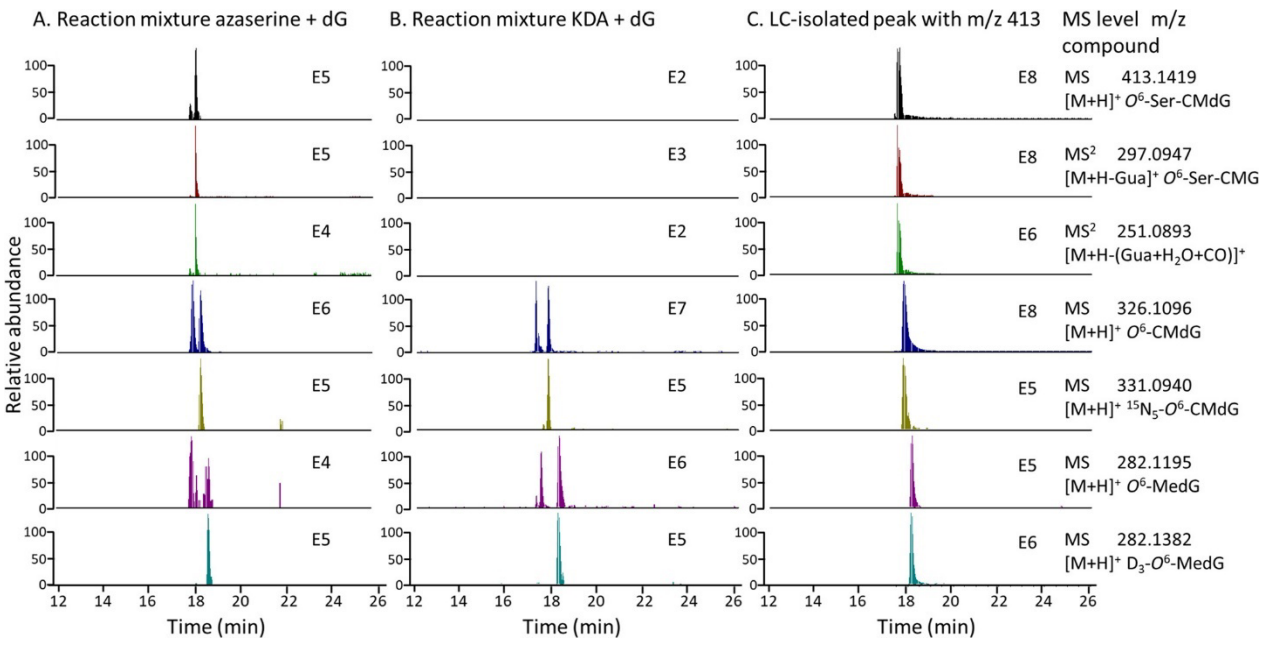


Figure 3. Representative chromatograms for the reaction of dG with L-azaserine or potassium diazoacetate analyzed by nanoLC-ESI-HRMS^2. (A, B) 5 µM dG was reacted with 5 mM L-azaserine (A) or potassium diazoacetate (B) as control in 10 mM tris-buffer at pH 7, followed by SPE and nanoLC-ESI-HRMS^2. (C) 4 mM dG was reacted with 25 mM L-azaserine in 0.1 M sodium phosphate buffer and peak with m/z 413 isolated by LC, concentrated, and analyzed by nanoLC-ESI-HRMS^2. XICs represent parent mass for ***O*^{6}-**Ser-CMdG and two fragmentations, followed by *O*^{6}-CMdG, $^{15}N_5$-*O*^{6} -CMdG as internal standard, *O*^{6}-MedG and D_3-*O*^{6}-MedG as internal standard. The sample represented in (C) was analyzed on a new column, explaining the shift in absolute retention time, while the Δ between *O*^{6}-Ser-CMdG and *O*^{6}-CMdG was equal.

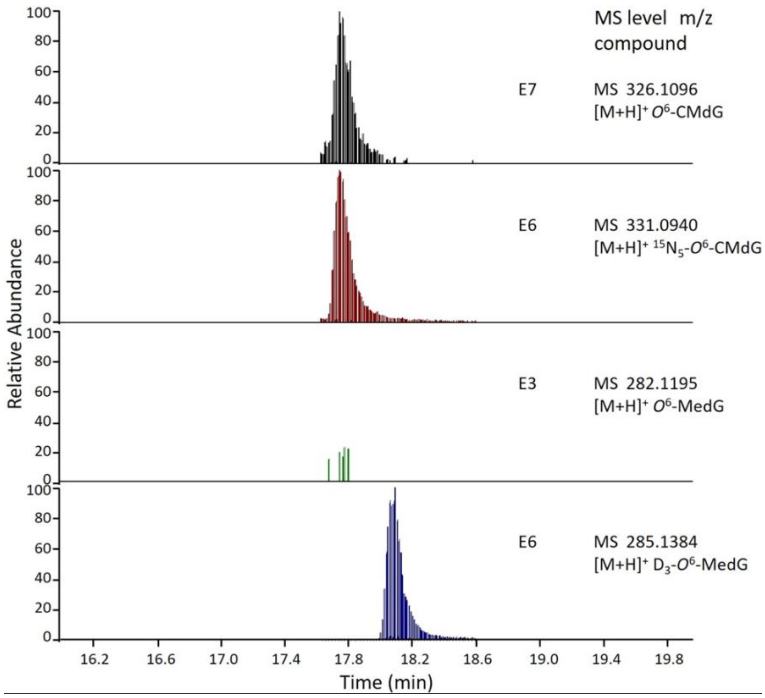


Figure 4. NanoLC-ESI-HRMS^2 chromatogram for the decomposition product of *O*^{6}-Ser-CMdG. Decomposition product was isolated by reversed-phase LC and identified as *O*^{6} -CMdG by nanoLC-ESI HRMS^2. No methylated dG could be detected. The m/z values are determined experimentally.

Exposure of Cells to L-azaserine and Quantification of Resulting Adducts. Having established that under neutral conditions direct alkylation of dG by L-azaserine occurs to form O^6-Ser-CMdG that spontaneously hydrolyzes to yield O^6-CMdG, we were interested in characterizing this process in cells and understanding the kinetic basis of adduct formation upon exposure to L-azaserine. Immortalized but normal diploid human colonic epithelial HCEC cells were employed as a model for adduct formation in healthy colon tissue.[35] HCEC cells express epithelial and stem cell markers and do not show mutations in genes known to be involved in colon cancer progression (APC, KRAS, TP53).[44] HCEC cells were exposed to L-azaserine for 0.5 to 120 h and at concentrations ranging from 125 to 1000 μM. Following DNA isolation and hydrolysis, products O^6-CMdG and O^6-MedG were detected, but no O^6-Ser-CMdG intermediate in cells, either because levels were too low or because it is not formed in cells. While inconclusive regarding whether the chemical mechanism established herein is the same as what occurs in cells, the fact that we could not detect it was not surprising, given our observation of the high instability of the serine intermediate on the nucleotide level, and the procedure required to digest genomic DNA for analysis of modified nucleosides from cells. Despite not observing the transient intermediate in samples obtained from cells, in order to characterize overall rates of DNA damage induced by L-azaserine in cells, we validated the nanoLC-ESI-HRMS^2 method for the concomitant quantification of O^6-CMdG and O^6-MedG in cells with a stable isotope dilution method that accounts for losses during sample preparation and ionization and allows for accurate quantification of the analyte based on the ratio of its signal area to that of labeled internal standard. PRM was applied for the mass transitions corresponding to a neutral loss of deoxyribose (dR) from the protonated analytes: O^6-CMdG and O^6-MedG, and their isotopically labeled internal standards ^{15}N5-O^6-CMdG and D3-O^6-MedG.

***O*^{6}-CMdG and *O*^{6}-MedG Levels in HCEC Cells.** The robust nanoLC-ESI-HRMS^2 approach for simultaneously quantifying O^6-CMdG and O^6-MedG enabled the measurement of L-azaserine-induced adducts in HCEC cells in a dose- and time-dependent manner. The level of both adducts increased linearly with L-azaserine concentration (0–1000 μM, 4 h exposure), with maximum adduct levels of 32.8 ± 7.8 O^6-CMdG and 4.3 ± 0.9 O^6-MedG lesions/10^7 nucleotides (Figure 5, for representative MS2 spectra, see Figures S9 and S10). Levels of O^6-CMdG were low compared to O^6-MedG, as reported previously.[17,27] We also detected O^6-CMdG in unexposed control samples but at levels below LOQ, while 0.7 ± 1.3 O^6-MedG adducts/10^7 nucleotides could be quantified in unexposed control samples. Endogenous O^6-CMdG was previously reported in human colorectal cells (HCT-166) and in XPA-deficient human fibroblast cells, and O^6-MedG is very well-known to be formed endogenously. [27,45]

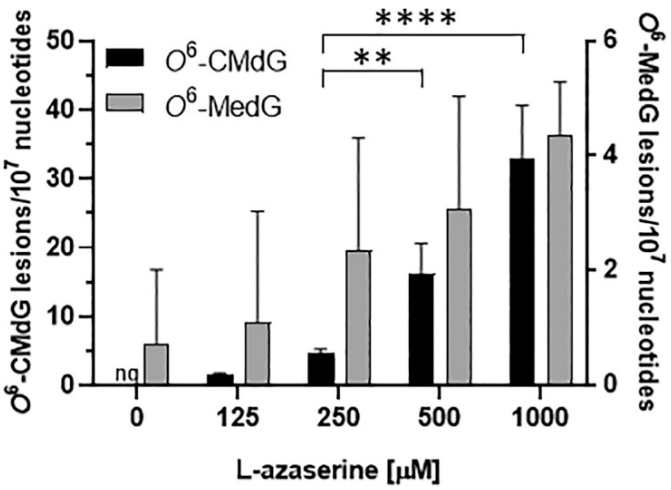


Figure 5. Dose-dependent formation of O^6-CMdG and O^6-MedG in HCEC cells. Cells were exposed for 4 h to 0, 125, 250, 500, or 1000 μM L-azaserine, and extracted DNA was analyzed for O^6-CMdG and O^6-MedG by nanoLC-ESI-HRMS^2. Adduct levels are expressed as lesions/107 nucleotides (mean ± SD, n = 3). P-values were calculated with one-way ANOVA and Tukey's multiple comparison's test, $^{**}p < 0.005$, $^{****}p < 0.0001$.

Having established the dose-dependent formation of O^6-CMdG and O^6-MedG after 4 h exposure to L-azaserine, we were interested in the temporal profiles of the DNA adducts. Therefore, we exposed HCEC cells to 500 µM L-azaserine (EC40 after 120 h, Figure S11) for various durations (0.5, 4, 10, 24, 48, 72, 96, 120 h). After 0.5 h, adduct levels were below the LOQ for O^6-CMdG and O^6-MedG. O^6-CMdG levels increased during the first 48 h to a maximum adduct level of 116 ± 21 O^6-CMdG lesions/10^7 nucleotides (Figure 6) and plateaued after, displaying no changes in adduct level for the remaining 3 days. Also, O^6-MedG levels reached a maximum value after 48 h; however, we could not establish a time-dependent relationship due to the overall low level of O^6-MedG adducts formed in HCEC exposed to 500 µM L-azaserine.

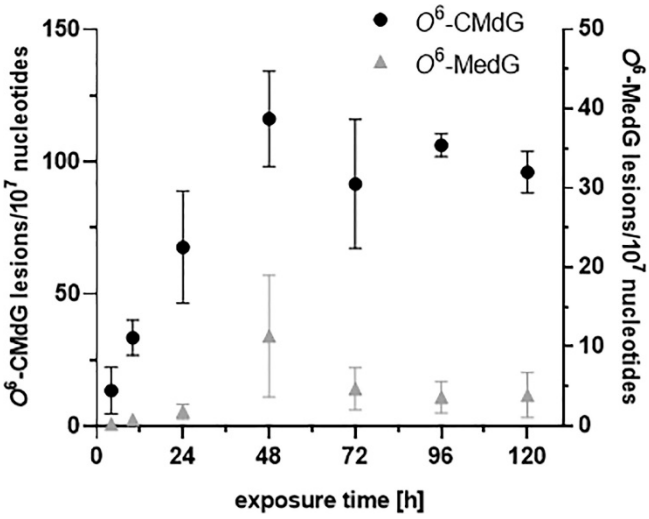


Figure 6. Time-dependent formation of ***O*⁶-**CMdG and ***O*⁶-**MedG in HCEC cells. Cells were exposed to 500 µM L-azaserine, and DNA was extracted after 0.5, 4, 10, 24, 48, 72, 96, and 120 h and further analyzed for ***O*⁶-**CMdG and ***O*⁶-**MedG levels by nanoLC-ESI-HRMS^2. Adduct levels are expressed as lesions/10^7 nucleotides (mean ± SD, n = 3).

6.5 Discussion

NOCs are well-known carcinogens linked to diet and lifestyle. The formation of the pro-mutagenic DNA adducts O^6-CMdG and O^6-MedG by NOCs is well established; [8,10,16,17,27] however, understanding of the chemical mechanism of NOC-induced O^6-CMdG formation and factors that influence their biological consequences is limited. It has been hypothesized that the release of diazoacetate from NOCs and analogs is responsible for the carboxymethylation and methylation of DNA. However, if the reactive intermediate is the same for various NOCs, such as *N*-(*N*-acetyl-l-prolyl)-*N*-nitrosoglycine (APNG) and *N*-nitrosoglycocholic acid (NOGC), and for the reactive intermediates L-azaserine and diazoacetate, it remains unclear why they all result in different ratios of O^6-CMdG to O^6-MedG. [17,25] Further, the carboxymethylating decomposition products of *N*-nitroso bile acid conjugates have not been recovered quantitatively. [46] The lack of characterization of the metabolism of NOCs and of reactive intermediates raises the question of what the mechanisms of carboxymethylation and methylation of DNA by NOCs are, including the possibility of several pathways and intermediates.

Hydrolysis of L-azaserine to L-serine and diazoacetate has been suggested earlier by Harrison et al.,[17] and decomposition to diazoacetate as the intermediate in O^6-CMdG formation has been demonstrated for *N*-nitroso-glycine under physiological conditions.[11] However, in this study, we demonstrated that the carboxymethylating agent L-azaserine undergoes hydrolysis to L-serine only under acidic conditions and that under physiologically relevant conditions in the presence of biological matrices, L-azaserine is stable. The half-life of L-azaserine was 0.4 h at pH 2, and its stability increased with increasing pH, consistent with an acid-catalyzed hydrolysis mechanism. Protonation of the diazocarbonyl moiety, possibly at the carbon bound to both the diazo and carbonyl groups, is expected to promote hydrolysis of the ester and release L-

serine (Figure 2B). On the basis of the pH dependence of the protonation/hydrolysis rates and having derived k_{obs} for each pH value (Table S2), we estimated the ratio $k_{hydrolysis}/K_a$ to be 0.90 × 10–4 M^{-1} s^{-1} (Figure 2A, eq 5). The pKa of the protonated L-azaserine (Figure 2A) could not be extracted, while similar diazo compounds such as diazomethane have a pKa of 10, such a high value would be inconsistent with the pH-stability profile of L-azaserine.[47] The observation of acid-promoted decomposition of L-azaserine is consistent with a previous report of loss of antibiotic activity and N^2 formation at pH 2 for L-azaserine.[48] Indeed, α-diazo compounds are acid labile, resulting in formation of reactive alkenium ions and loss of N^2.[49] Acid-promoted decomposition was also demonstrated for diazopeptides and *N*-nitrosopeptides, while under physiological conditions they are stable.[50,51] Finally, under neutral conditions, diazoacetate was demonstrated to be an intermediate of *N*-nitroso-glycine-induced O^6-CMdG and O^6-MedG formation, potentially via formation of carboxymethyldiazonium ion and diazomethane, whereas, to our knowledge, the decomposition of NOCs with more stabilized α-diazocarboxy-groups,[52] for example, diazopeptides, has not been characterized under physiological conditions.[11]

Based on the lack of evidence for the decomposition of L-azaserine at neutral pH in the presence of active whole cell lysates, we hypothesized that L-azaserine reacts directly with DNA under physiological conditions. The formation of a direct reaction product Ser-CMdG from L-azaserine has been previously proposed, but to our knowledge was never investigated. [17,53] Data from reactions with oligonucleotides and with dG revealed the existence of the intermediate O^6-Ser-CMdG (Figure 1). We did not observe any correlation of levels of the direct adduct with time or pH; however, the amount that could be obtained from oligonucleotides was extremely small and transient on the nucleotide level, making any clear kinetic conclusions elusive. The other regioisomer observed was assigned as N^2-SerCMdG. On the basis of peak area, and approximating the same recovery and ionization behavior, it appears to be on average about 3 times less abundant than O^6-SerCMdG. However, for CMdG and MedG adducts, the N^2:O^6 ratio was 1.3 and 1.1, respectively. Interestingly, for the reaction with diazoacetate, the N^2:O^6 ratio was 2 for O^6-CMdG and 0.7 for O^6-MedG, suggesting the adducts were formed via a different chemical mechanism. Diazo compounds with stabilizing electron withdrawing groups are a useful tool in the modification and detection of biomolecules,[52] as demonstrated by the use of ethyldiazoacetate in carboxylation of dG.[36] Next to the O^6- position, N^7, N^3, N^2, and N^1 represent nucleophilic sites in dG that are susceptible to modification by electrophilic reagents.[54] N^7-CMG and N^7-MeG have been detected after in vitro NOCs exposure, however, these readily depurinate.[25,53] Further, two methylated-dG adducts with modification other than at the O^6-position were detected previously after exposure of ctDNA to potassium diazoacetate, and one carboxymethylated-dG adduct other than O^6 was present in colon tumor biopsy samples analyzed by a $HRMS^2$ approach for establishing a diet-related DNA adduct database.[8]

While evidence for the direct serine adduct could not be gained from samples of genomic DNA from L-azaserine-exposed cells, we characterized the formation and persistence of O^6-CMdG and O^6-MedG as a function of L-azaserine dose and time of exposure with a robust and sensitive nanoLC-ESI-$HRMS^2$ PRM method validated for analysis of cell samples. To our knowledge, this is the first HRMS approach for the absolute quantification of O^6-MedG and O^6-CMdG in biological samples using stable isotope-labeled internal standards for accurate quantification of both adducts. A few methods involving triple quadrupole or ion traps as mass analyzer exist for the detection and quantification of O^6-CM(d)G in biological samples, [55] also in combination with O^6-Me(d)G.[10,27] A HRMS methodology targeting O^6-MeG and O^6-CMG

and other targeted and untargeted DNA adducts was successfully applied to colon biopsy samples to establish a diet-related DNA adducts database and to compare the DNA adduct profile from white vs red meat in gastrointestinal digestion samples. [8,15] O^6-MedG was further one of the targeted adducts in a reported adductomic approach for nitrosamine-induced DNA adducts.[56] Validation of our nanoLC-ESI-HRMS^2 approach revealed LOQs in matrix were in the low amol range and comparable to LOQs previously reported for pure standards on nanoLC-ESI-MS3,[27] and improved compared to a HRMS method targeting four adducts in biological matrix.[8] Applying our methodology to HCEC cell exposed to L-azaserine, we found a linear response between L-azaserine dose and O^6-CMdG and O^6-MedG adduct levels in HCEC cells (Figure 5). Both adducts were present in unexposed control samples, though not quantifiable for O^6-CMdG and below one O^6-MedG adduct/10^7 nucleotides. We measured maximal O^6-CMdG adduct levels in cells after 48 h of exposure, followed by a steady level up to 120 h (Figure 6). The O^6-MedG profile showed a similar trend, but the adduct levels were too low for statistical analysis. The low O^6-MedG levels observed are consistent with the more potent O^6-CMdG formation by L-azaserine.[16,17] Additionally, low levels of O^6-MedG measured in cells might be due to efficient repair of O^6-MedG by O^6-methylguanine DNA methyl transferase.[57] In a previous study, a maximum N^7-MeG adduct level was achieved also after 48 h exposure to NDMA, however, in a process requiring metabolic activation by CYP enzymes.[58] Adduct levels were on the same order of magnitude as previously reported for colorectal cancer cells exposed to L-azaserine, [27] which was more potent in inducing O^6-CMdG.[27] Endogenous O^6-CMdG and O^6-MedG formation has been reported previously and is likely due to the endogenous formation of s-adenosylmethionine and NOCs, [27,59-61] which account for 45–75% of the total human NOC exposure.[19,62]

6.6 Conclusion

In this study, we characterized the chemical basis of L-azaserine-induced O^6-CMdG formation. We identified two mechanisms, occurring under acidic and physiological conditions, respectively, by which L-azaserine converts into reactive intermediates and induces O^6-CMdG. Our findings have implications for endogenously formed NOCs that do not breakdown to diazoacetate under physiological conditions but can form DNA adducts via nucleophilic attack. We established a highly sensitive nanoLC-ESI-HRMS^2 methodology and measured a linear dose-response relationship for O^6-CMdG lesions in cells after exposure to L-azaserine. The knowledge of L-azaserine-induced DNA adduct formation has important implications to understand O^6-CMdG formation and accumulation in cells and to study its biological consequences. Overall, this work provides fundamental chemical insight concerning the basis of how L-azaserine may give rise to DNA damage and supports its use as a chemical probe in NOC-induced carcinogenesis research, particularly concerning the identification of factors that mitigate adverse effects of O^6-CMdG.

Funding

This work was supported by the Swiss National Science Foundation (185020, 186332) and the Swiss Cancer Research Foundation (KFS-4443-02-2018-R).

Notes

The authors declare no competing financial interest.

Acknowledgements

The authors wish to thank the staff of the Functional Genomics Center Zurich (FGCZ) for providing nanoLC-ESI-HRMS^2 instrumentation and technical support, namely Dr. Bernd Roschitzki, Dr. Peter Gehrig, Dr. Serena di Palma, Dr. Endre Laczko, and Laura Kunz. We also thank Dr. Hailey Gahlon for scientific advice and Jonne van Dijk for her technical support in sample preparation.

6.7 References

1. Bouvard, V. *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* **16**, 1599-600 (2015).
2. Aloisi, C.M.N., Sandell, E.S. & Sturla, S.J. A Chemical Link between Meat Consumption and Colorectal Cancer Development? *Chem Res Toxicol* **34**, 12-23 (2021).
3. Turesky, R.J. Mechanistic Evidence for Red Meat and Processed Meat Intake and Cancer Risk: A Follow-up on the International Agency for Research on Cancer Evaluation of 2015. *Chimia (Aarau)* **72**, 718-724 (2018).
4. Loechler, E.L., Green, C.L. & Essigmann, J.M. In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc Natl Acad Sci U S A* **81**, 6271-5 (1984).
5. Pauly, G.T. & Moschel, R.C. Mutagenesis by O(6)-methyl-, O(6)-ethyl-, and O(6)-benzylguanine and O(4)-methylthymine in human cells: effects of O(6)-alkylguanine-DNA alkyltransferase and mismatch repair. *Chem Res Toxicol* **14**, 894-900 (2001).
6. Hall, C.N., Badawi, A.F., O'Connor, P.J. & Saffhill, R. The detection of alkylation damage in the DNA of human gastrointestinal tissues. *Br J Cancer* **64**, 59-63 (1991).
7. O6-methylguanine in blood leucocyte DNA: an association with the geographic prevalence of gastric cancer and with low levels of serum pepsinogen A, a marker of severe chronic atrophic gastritis. The EUROGAST Study Group. *Carcinogenesis* **15**, 1815-20 (1994).
8. Hemeryck, L.Y., Decloedt, A.I., Vanden Bussche, J., Geboes, K.P. & Vanhaecke, L. High resolution mass spectrometry based profiling of diet-related deoxyribonucleic acid adducts. *Anal Chim Acta* **892**, 123-31 (2015).
9. Povey, A.C., Hall, C.N., Badawi, A.F., Cooper, D.P. & O'Connor, P.J. Elevated levels of the pro-carcinogenic adduct, O(6)-methylguanine, in normal DNA from the cancer prone regions of the large bowel. *Gut* **47**, 362-5 (2000).
10. Vanden Bussche, J., Moore, S.A., Pasmans, F., Kuhnle, G.G. & Vanhaecke, L. An approach based on ultra-high pressure liquid chromatography-tandem mass spectrometry to quantify O6-methyl and O6-carboxymethylguanine DNA adducts in intestinal cell lines. *J Chromatogr A* **1257**, 25-33 (2012).
11. Cupid, B.C., Zeng, Z., Singh, R. & Shuker, D.E. Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chem Res Toxicol* **17**, 294-300 (2004).
12. Lewin, M.H. *et al.* Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res* **66**, 1859-65 (2006).
13. Vanden Bussche, J. *et al.* O(6)-carboxymethylguanine DNA adduct formation and lipid peroxidation upon in vitro gastrointestinal digestion of haem-rich meat. *Mol Nutr Food Res* **58**, 1883-96 (2014).
14. Hemeryck, L.Y. *et al.* In vitro DNA adduct profiling to mechanistically link red meat consumption to colon cancer promotion. *Toxicol Res (Camb)* **5**, 1346-1358 (2016).
15. Hemeryck, L.Y., Rombouts, C., De Paepe, E. & Vanhaecke, L. DNA adduct profiling of in vitro colonic meat digests to map red vs. white meat genotoxicity. *Food Chem Toxicol* **115**, 73-87 (2018).

16. Harrison, K.L., Fairhurst, N., Challis, B.C. & Shuker, D.E. Synthesis, characterization, and immunochemical detection of O6-(carboxymethyl)-2'-deoxyguanosine: a DNA adduct formed by nitrosated glycine derivatives. *Chem Res Toxicol* **10**, 652-9 (1997).
17. Harrison, K.L., Jukes, R., Cooper, D.P. & Shuker, D.E. Detection of concomitant formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine in DNA exposed to nitrosated glycine derivatives using a combined immunoaffinity/HPLC method. *Chem Res Toxicol* **12**, 106-11 (1999).
18. Stuff, J.E., Goh, E.T., Barrera, S.L., Bondy, M.L. & Forman, M.R. Construction of an N-nitroso database for assessing dietary intake. *J Food Compost Anal* **22**, S42-S47 (2009).
19. Mirvish, S.S. Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC. *Cancer Lett* **93**, 17-48 (1995).
20. Wu, J. *et al.* Replication studies of carboxymethylated DNA lesions in human cells. *Nucleic Acids Res* **45**, 7276-7284 (2017).
21. Raz, M.H. *et al.* Bypass of Mutagenic O(6)-Carboxymethylguanine DNA Adducts by Human Y- and B-Family Polymerases. *Chem Res Toxicol* **29**, 1493-503 (2016).
22. Gottschalg, E., Scott, G.B., Burns, P.A. & Shuker, D.E. Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of O6-carboxymethyl- and O6-methyl-2'-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis* **28**, 356-62 (2007).
23. Andreyev, H.J. *et al.* Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer* **85**, 692-6 (2001).
24. Jianshuang Wang, Y.W. Chapter Six - Carboxymethylation of DNA Induced by N-Nitroso Compounds and Its Biological Implications,. in *Advance in Molecular Toxicology*, Vol. 5 (ed. Fishbein, J.C.) 219-243 (Elsevier, Netherlands, 2011).
25. Shuker, D.E. & Margison, G.P. Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Res* **57**, 366-9 (1997).
26. O'Driscoll, M., Macpherson, P., Xu, Y.Z. & Karran, P. The cytotoxicity of DNA carboxymethylation and methylation by the model carboxymethylating agent azaserine in human cells. *Carcinogenesis* **20**, 1855-62 (1999).
27. Yu, Y., Wang, J., Wang, P. & Wang, Y. Quantification of Azaserine-Induced Carboxymethylated and Methylated DNA Lesions in Cells by Nanoflow Liquid Chromatography-Nanoelectrospray Ionization Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal Chem* **88**, 8036-42 (2016).
28. Wang, J. & Wang, Y. Chemical synthesis of oligodeoxyribonucleotides containing N3- and O4-carboxymethylthymidine and their formation in DNA. *Nucleic Acids Res* **37**, 336-45 (2009).
29. Stock, C.C., Reilly, H.C., Buckley, S.M., Clarke, D.A. & Rhoads, C.P. Azaserine, a new tumour-inhibitory substance; studies with Crocker mouse sarcoma 180. *Nature* **173**, 71-2 (1954).
30. Longnecker, D.S. & Curphey, T.J. Adenocarcinoma of the pancreas in azaserine-treated rats. *Cancer Res* **35**, 2249-58 (1975).
31. Lilja, H.S., Hyde, E., Longnecker, D.S. & Yager, J.D., Jr. DNA damage and repair in rat tissues following administration of azaserine. *Cancer Res* **37**, 3925-31 (1977).
32. Rosenkrantz, H. *et al.* The stoichiometric fate of azaserine metabolized in vitro by tissues from azaserine-treated dogs and mice. *Toxicol Appl Pharmacol* **22**, 607-20 (1972).
33. Jacquez, J.A. & Sherman, J.H. Enzymatic degradation of azaserine. *Cancer Res* **22**, 56-61 (1962).
34. Longenecker, J.B. & Snell, E.E. Pyridoxal and metal ion catalysis of alpha, beta elimination reactions of serine-3-phosphate and related compounds. *J Biol Chem* **225**, 409-18 (1957).
35. Zhang, L. *et al.* Exome Sequencing of Normal and Isogenic Transformed Human Colonic Epithelial Cells (HCECs) Reveals Novel Genes Potentially Involved in the Early Stages of Colorectal Tumorigenesis. *BMC Genomics* **16 Suppl 1**, S8 (2015).

36. Geigle, S.N., Wyss, L.A., Sturla, S.J. & Gillingham, D.G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem Sci* **8**, 499-506 (2017).
37. Reza Fathi, B.G., Pei-Pei Kung, Gaffney Barbara L., & Jones, Roger A. . Synthesis of 6-substituted 2′ -deoxyguanosine derivatives using trifluoroacetic anhydride in pyridine. *Tetrahedron Letters* **31**, 319 (1990).
38. Zhang, P. *et al.* Revisiting Fragmentation Reactions of Protonated alpha-Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation. *Sci Rep* **9**, 6453 (2019).
39. Françoise Rogalewicz, Y.H., Gilles Ohanessiana. Fragmentation mechanisms of α-amino acids protonated under electrospray ionization: a collisional activation and ab initio theoretical study. *International Journal of Mass Spectrometry* **195-196**, 565-590 (2000).
40. Piraud, M. *et al.* ESI-MS/MS analysis of underivatised amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. *Rapid Commun Mass Spectrom* **17**, 1297-1311 (2003).
41. MycoAlert Assay Control Set. Vol. 2021 (Lonza online protocol, https://bioscience.lonza.com/lonza_bs/CH/en/download/product/asset/27672).
42. CellTiter-Glo Luminescent Cell Viability Assay. Vol. 2021 (Promega, https://ch.promega.com/-/media/files/resources/protocols/technical-bulletins/0/celltiter-glo-luminescent-cell-viability-assay-protocol.pdf).
43. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Res Notes* **12**, 106 (2019).
44. Roig, A.I. *et al.* Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. *Gastroenterology* **138**, 1012-21 e1-5 (2010).
45. Nakamura, J. *et al.* The endogenous exposome. *DNA Repair (Amst)* **19**, 3-13 (2014).
46. David E. G. Shuker, S.R.T., and John S. Wishnok. N-Nitroso bile acid conjugates. 1. Synthesis, chemical reactivity and mutagenic activity. *Journal of Organic Chemistry* **46**, 2092-2096 (1981).
47. Smyth, J.F.M.a.T. Hydrolysis of diazomethane-kinetics and mechanism. *Journal American Chemical Society* **102**, 7303-7308 (1980).
48. Bartz, Q.R. *et al.* Isolation and characterization of azaserine. *Nature* **173**, 72-3 (1954).
49. M, R. Diazo Compounds: Properties and Synthesis. *Academic Press* (2012).
50. Challis, B.C. Chemistry and biology of nitrosated peptides. *Cancer Surv* **8**, 363-84 (1989).
51. Challis, B.C., Hopkins, A.R., Milligan, J.R., Mitchell, R.C. & Massey, R.C. Nitrosation of peptides. *IARC Sci Publ*, 61-70 (1984).
52. Mix, K.A., Aronoff, M.R. & Raines, R.T. Diazo Compounds: Versatile Tools for Chemical Biology. *ACS Chem Biol* **11**, 3233-3244 (2016).
53. Zurlo, J., Curphey, T.J., Hiley, R. & Longnecker, D.S. Identification of 7-carboxymethylguanine in DNA from pancreatic acinar cells exposed to azaserine. *Cancer Res* **42**, 1286-8 (1982).
54. Liu, S. & Wang, Y. Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem Soc Rev* **44**, 7829-54 (2015).
55. Da Pieve, C., Sahgal, N., Moore, S.A. & Velasco-Garcia, M.N. Development of a liquid chromatography/tandem mass spectrometry method to investigate the presence of biomarkers of DNA damage in urine related to red meat consumption and risk of colorectal cancer. *Rapid Commun Mass Spectrom* **27**, 2493-503 (2013).
56. Balbo, S., Hecht, S.S., Upadhyaya, P. & Villalta, P.W. Application of a high-resolution mass-spectrometry-based DNA adductomics approach for identification of DNA adducts in complex mixtures. *Anal Chem* **86**, 1744-52 (2014).
57. Pegg, A.E. Repair of O(6)-alkylguanine by alkyltransferases. *Mutat Res* **462**, 83-100 (2000).

58. Lin, H. & Hollenberg, P.F. N-nitrosodimethylamine-mediated formation of oxidized and methylated dna bases in a cytochrome P450 2E1 expressing cell line. *Chem Res Toxicol* **14**, 562-6 (2001).
59. Sharma, V. *et al.* Molecular dosimetry of endogenous and exogenous O(6)-methyl-dG and N7-methyl-G adducts following low dose [D3]-methylnitrosourea exposures in cultured human cells. *Chem Res Toxicol* **27**, 480-2 (2014).
60. Konishi, Y. *et al.* Pancreatic tumors induced by a single intraperitoneal injection of azaserine in partial pancreatectomized rats. *Cancer Lett* **9**, 43-6 (1980).
61. De Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-85 (2004).
62. Tricker, A.R. N-nitroso compounds and man: sources of exposure, endogenous formation and occurrence in body fluids. *Eur J Cancer Prev* **6**, 226-68 (1997).

6.8 Supporting Information

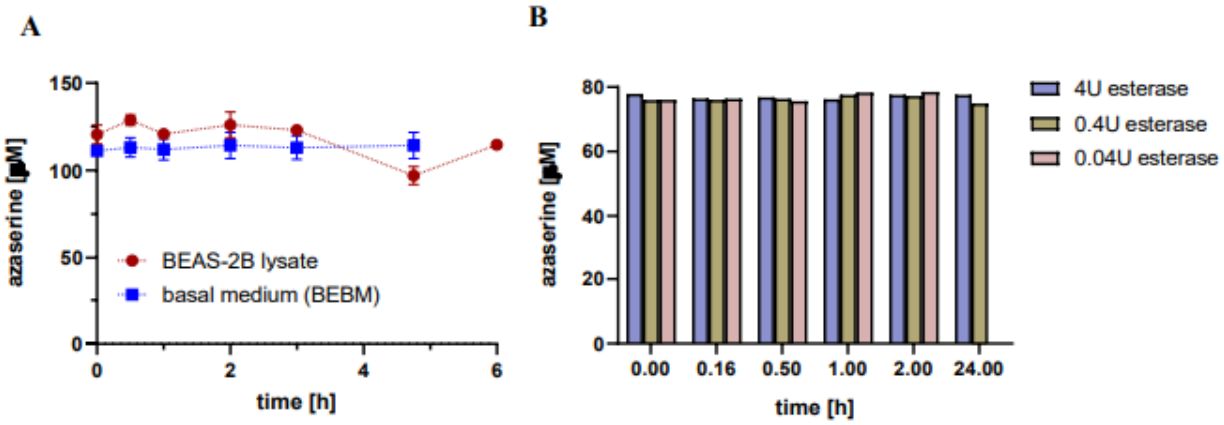


Figure S1. L-azaserine is stable in the presence of biological matrices. L-azaserine was incubated with active whole cell lysates and medium from BEAS-2B (A) as well as excess commercial porcine liver esterase (B). No decomposition of L-azaserine could be observed via HPLC.

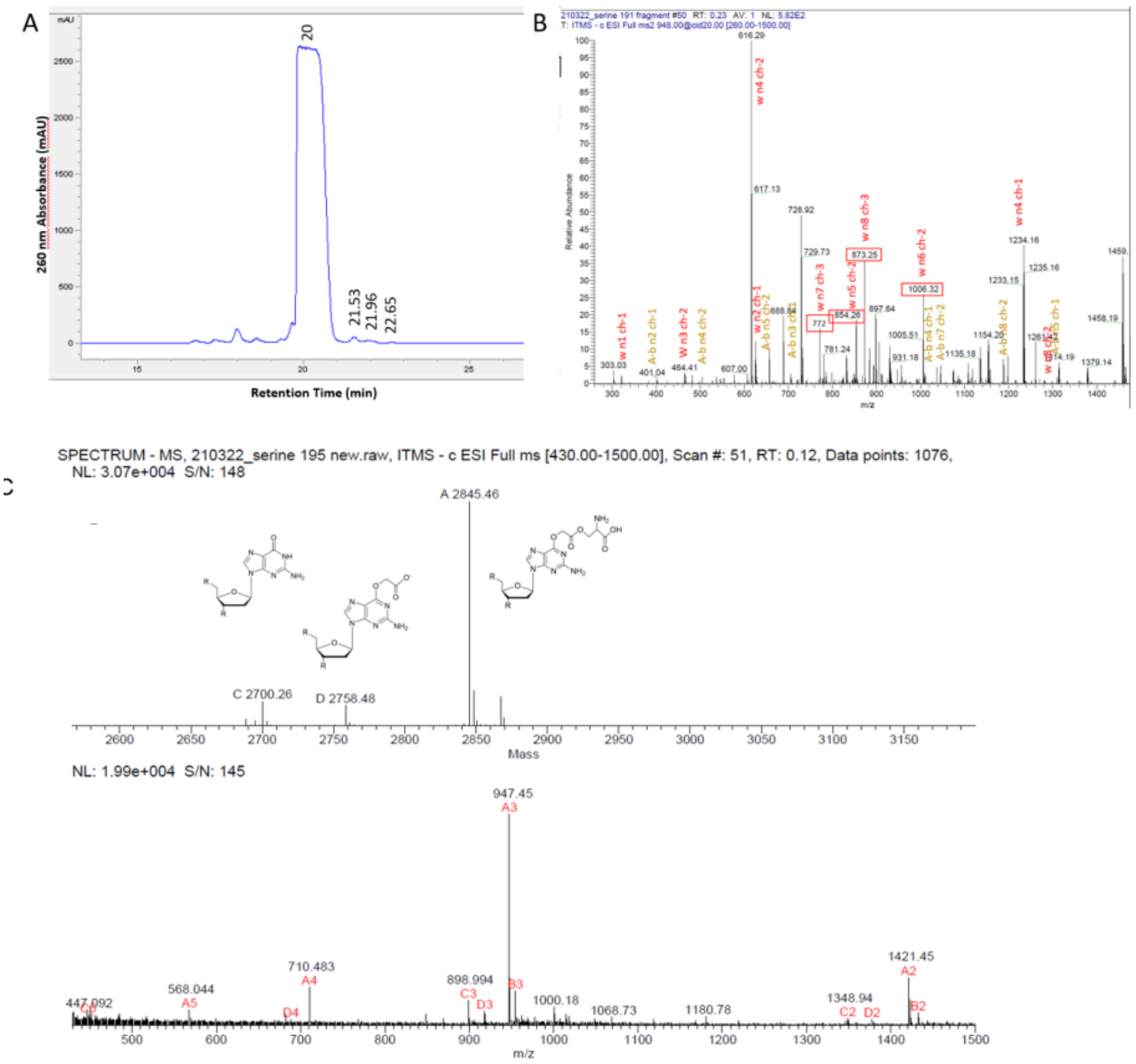


Figure S2. Formation of *O*^{6}-Ser-CMG in an oligonucleotide 9-mer. UV chromatogram of the products of the reaction of an oligonucleotide with L-azaserine. (A) The starting oligonucleotide eluted at 20 min, and the *O*^{6} -Ser-CMG at 21.96 min. (B) MS^2 fragments from HRMS analysis of the peak eluting at 21.96 min in the chromatogram shown in (A). B (yellow) and w (red) type ions were mainly observed. The values were consistent with modification on guanine in the oligonucleotide 9-mer sequence. (C) Fraction that eluted at 21.96 min (A) was analyzed by mass spectrometry. m/z value of the electrospray ionization mass of the non-modified oligonucleotide, *O*^{6}-CMG-oligonucleotide and Ser-*O*^{6} -CMG-oligonucleotide were found.

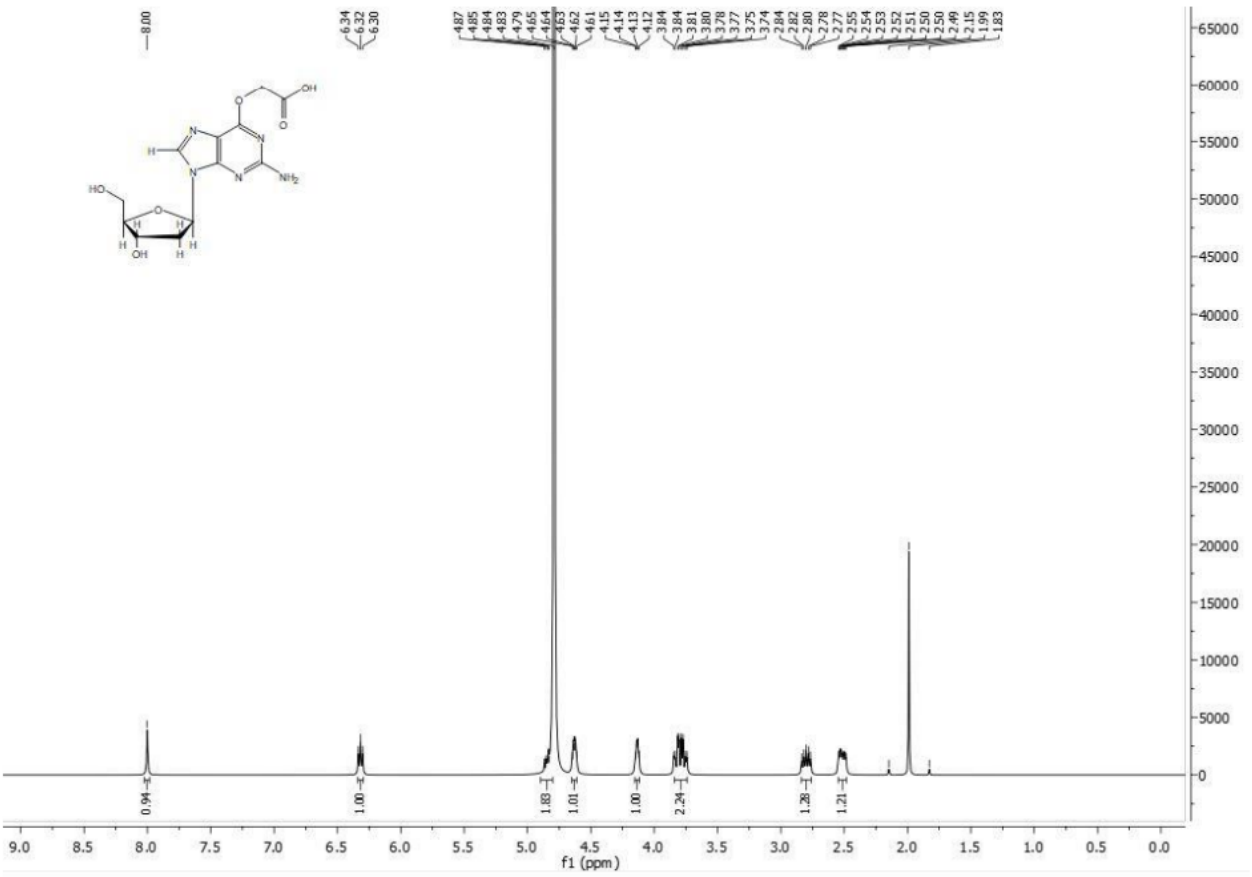


Figure S3. ^{1}H NMR (400 MHz, Deuterium Oxide) of *O*^{6}-CMdG as standard for quantification by nanoLC-ESI-HRMS^{2}. *O*^{6} -CMdG: δ 8.00 (s, 1H, Ar-H), 6.32 (t, J = 6.9 Hz, 1H, 1'-H), 4.87 – 4.83 (m, 2H, O-CH_2), 4.63 (dt, J = 6.5, 3.4 Hz, 1H, 4'-H), 4.14 (q, J = 3.8 Hz, 1H, 3'-H), 3.79 (qd, J = 12.5, 4.2 Hz, 2H, 5'-H), 2.80 (dt, J = 13.9, 6.8 Hz, 1H, 2'-H), 2.52 (ddd, J = 14.1, 6.4, 3.7 Hz, 1H, 2'-H).

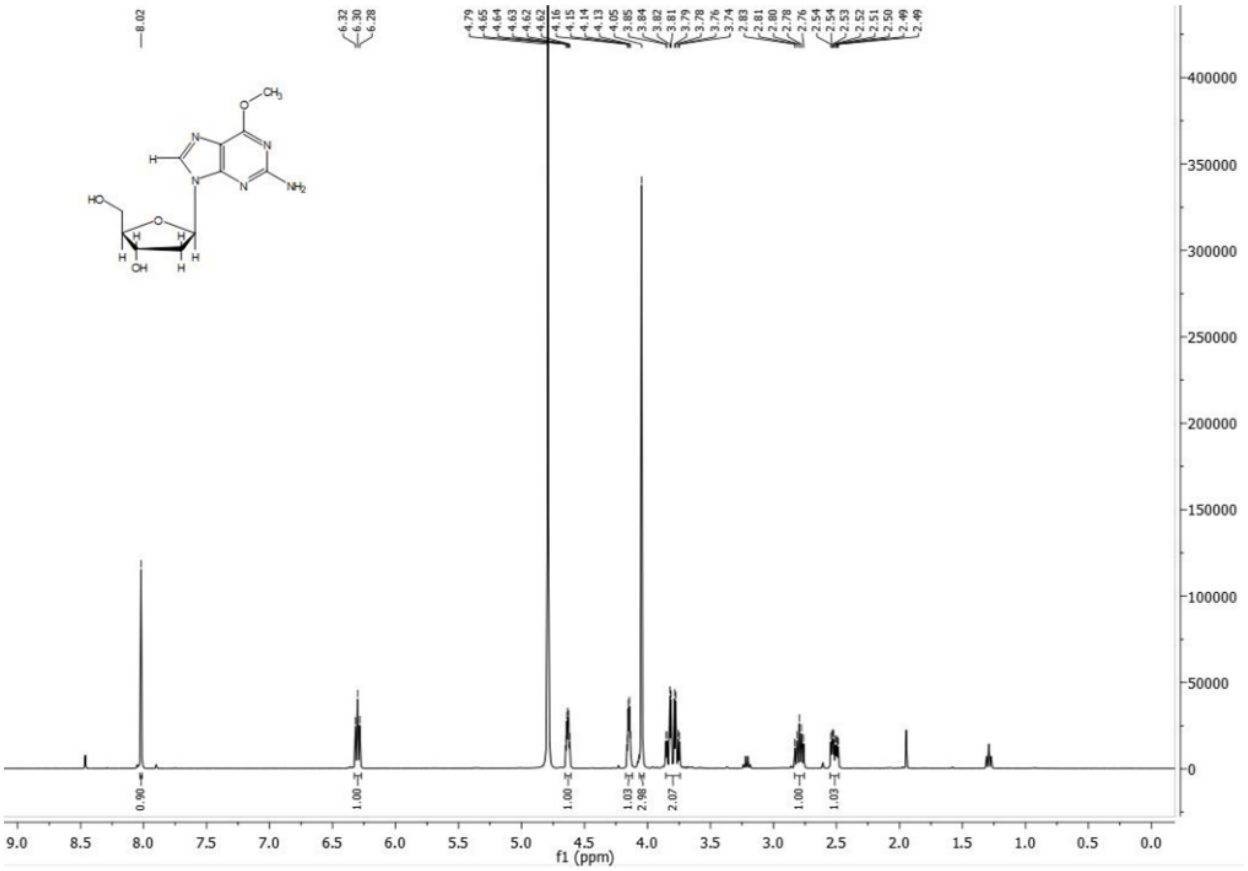


Figure S4. ^{1}H NMR (400 MHz, Deuterium Oxide) of *O*^{6}-MedG as standard for quantification by nanoLC-ESI-HRMS^{2}. *O*^{6} -MedG: δ 8.02 (s, 1H, Ar-H), 6.30 (t, J = 6.9 Hz, 1H, 1'-H), 4.63 (dt, J = 6.1, 3.1 Hz, 1H, 4'-H), 4.15 (q, J = 3.5 Hz, 1H, 3'-H), 4.05 (s, 3H, O-CH_3), 3.80 (qd, J = 12.6, 3.9 Hz, 2H, 5'-H), 2.80 (dt, J = 13.9, 6.8 Hz, 1H, 2'-H), 2.52 (ddd, J = 14.0, 6.2, 3.3 Hz, 1H, 2'-H).

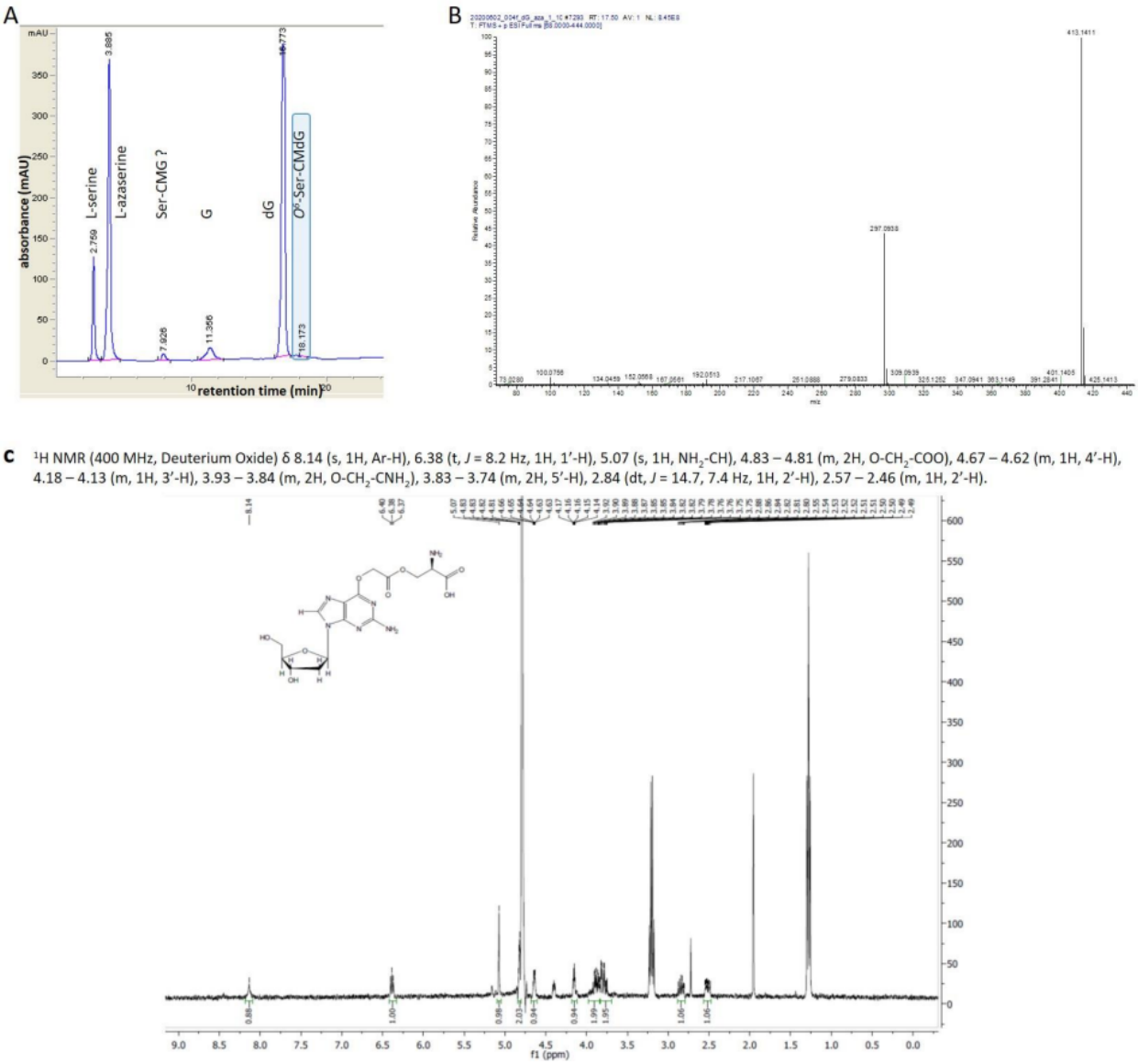


Figure S5. Characterization of O^6 -Ser-CMdG. (A). A representative liquid chromatography chromatogram at λ=210 nm of a dG azaserine reaction mixture in water after 102 h. The peak at retention time of 18.7 min was collected for further characterization. (B). nanoLC-ESI-HRMS on an Orbitrap FusionTM LumosTM (Thermo Scientific, Waltham, Massachusetts) characterization of the peak isolated by liquid chromatography. Most intense peak is m/z 413.1411 $[M+H]^+$ being identical to the calculated m/z for an O^6-Ser-CMdG intermediate (413.1421), second intense peak is 297.0938 $[M+H]^+$, being identical to the calculated m/z of a O^6-Ser-CMG intermediate (297.0947) most likely from loss of deoxyribose during the ionization process. (C). ^{1}H NMR of the isolated compound confirming the O^6-Ser-CMdG structure.

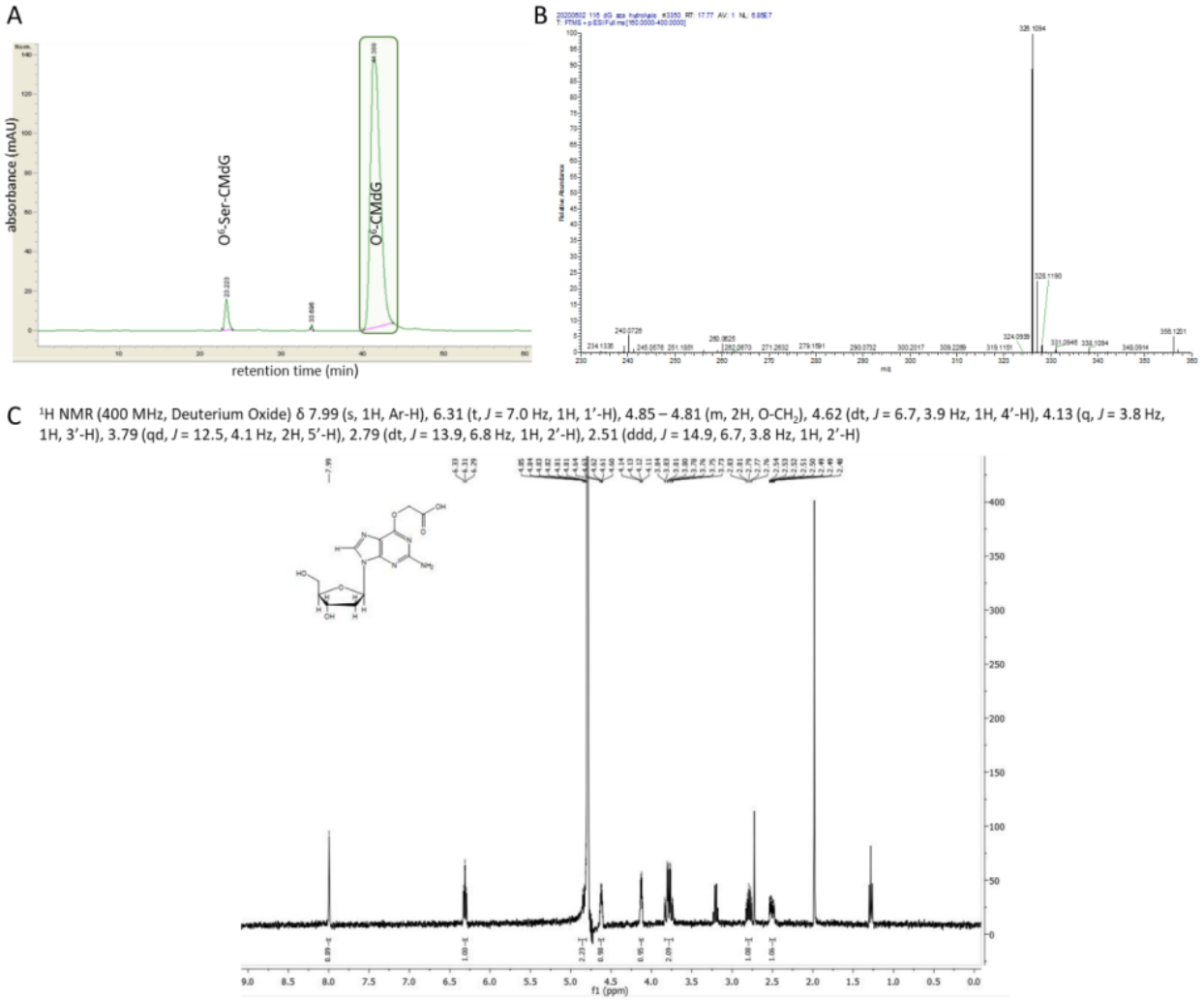


Figure S6. Characterization of the decomposition product of *O*^{6}-Ser-CMdG. (A) A representative liquid chromatography chromatogram at λ=280 nm of *O*^{6}-Ser-CMdG after 30 h in D_2O. *O*^{6}-Ser-CMdG (22.2 min) is not completely hydrolyzed. The peak at retention 41.3 min was collected for further characterization. (B) nanoLC-ESI-HRMS characterization of the peak isolated by liquid chromatography. Most intense peak is m/z 326.1094 $[M+H]^+$ being identical to the calculated m/z for *O*^{6}-CMdG (326.1101). (C) ^{1}H NMR of the isolated compound confirming the *O*^{6}-CMdG structure

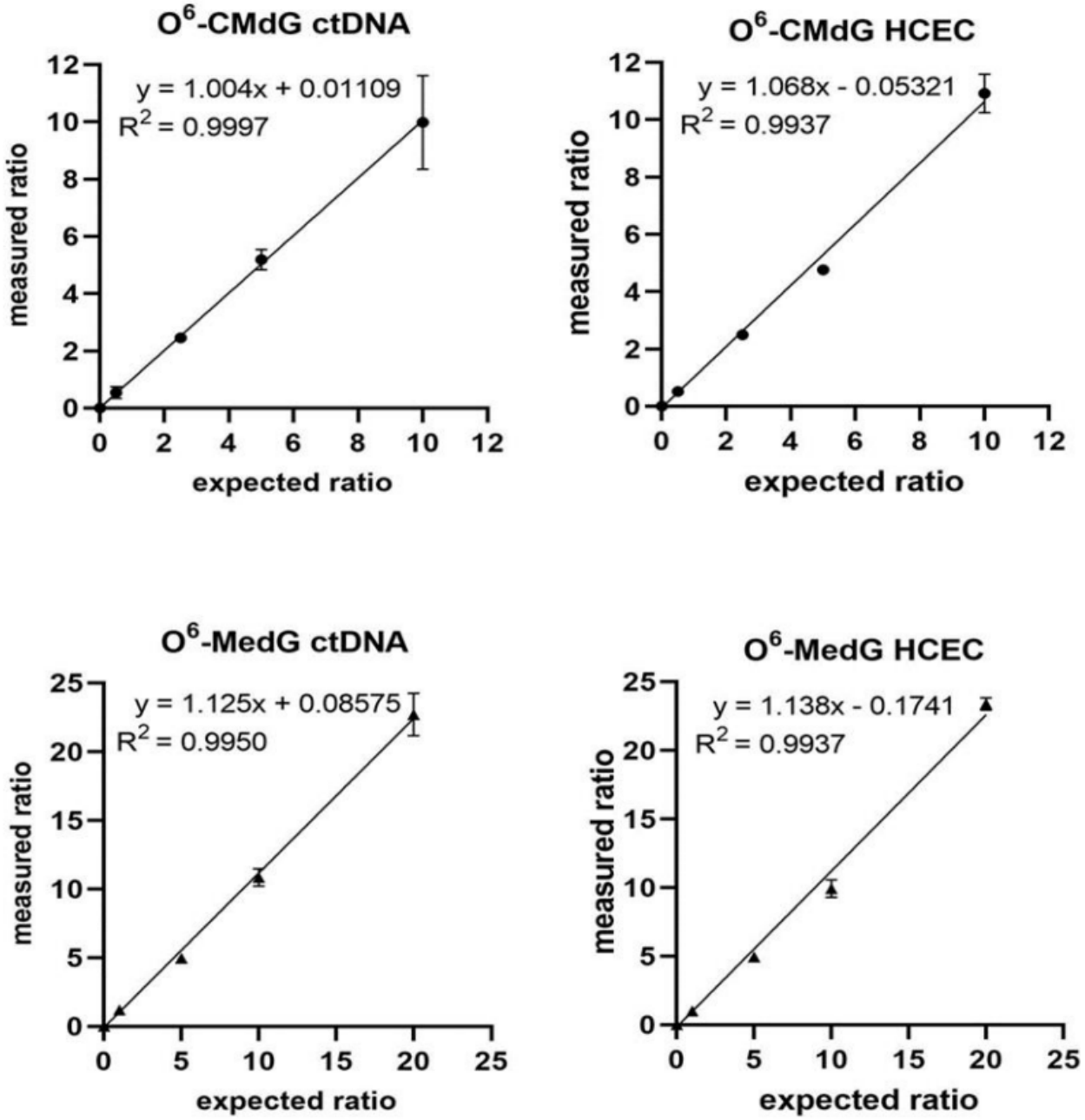


Figure S7. Validation results for the quantification of *O*[6]-CMdG and *O*[6]-MedG by nanoLCESI-HRMS[2]. 0.01-200 nM *O*[6] -CMdG and *O*[6] -MedG were spiked in ctDNA (15 µg) or isolated DNA (15 µg) from unexposed HCEC cells followed by sample preparation and analysis as described. Results were plotted as measured ratio of analyte /internal standards over expected ratio of analyte/internal standard. Reasonable linearity was achieved for the analysis of *O*[6]-CMdG and *O*[6]-MedG in matrix.

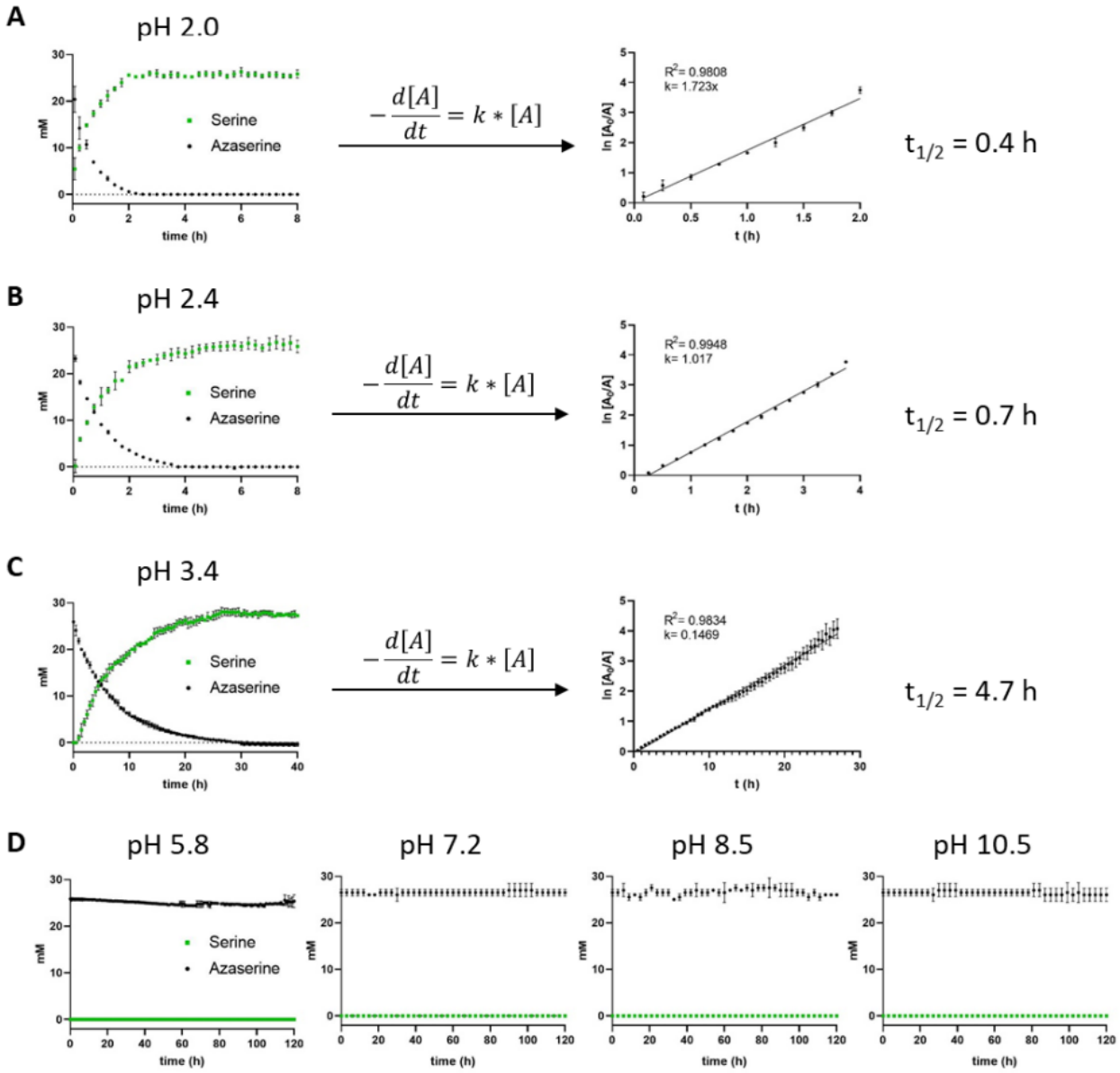


Figure S8. Acid-promoted hydrolysis of L-azaserine. Hydrolysis of L-azaserine to L-serine at (A) pH 2, $t_{1/2}$=0.4 h (B) pH 2.4, $t_{1/2}$=0.7 h. and (C) pH 3.4, $t_{1/2}$=4.7 h. Hydrolysis of L-azaserine was modeled as a reaction of pseudo first order and half-life ($t_{1/2}$) calculated with the following equations: ln[$A0$] ln[A] = kt, $t_{1/2}$ = $ln2\ k$ where [A0]=starting concentration of L-azaserine (mM), [A]=concentration of L-azaserine at t (mM), k=slope (1/h), (n=3). (D) No hydrolysis of L-azaserine. L-serine was observed when pH was 5.8, 7.2, 8.5 and 10.5 over 120 h (n =2).

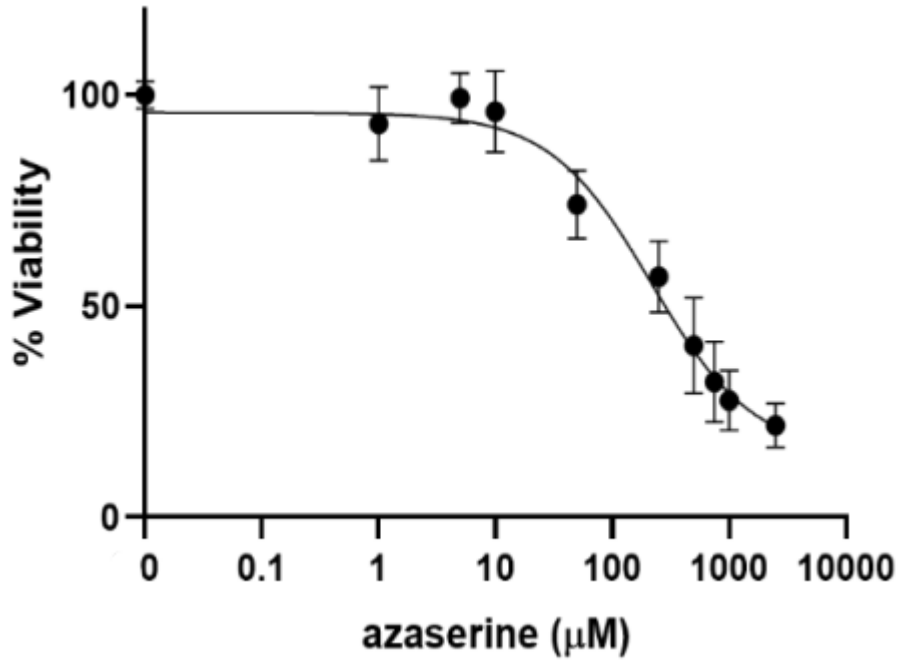


Figure S9. Viability of HCEC cells exposed to azaserine. Cell were exposed to 0, 1, 10, 50, 250, 750, 1000 and 2500 µM azaserine for 120 h and cell viability was determined by CellTiter-Glo. Cell viability is expressed in % to unexposed control cells (100 %viability) (mean ± SD, n=3). When exposed to 500 µM azaserine, concentration chosen for time-dependent adduct level measurement (Figure 7), 40 % of cells were viable after 120 h. Data were fitted in a nonlinear curve using the function [Inhibitor] vs response of GraphPad (R 2 was 0.92). EC50 was calculated to be 220 µM.

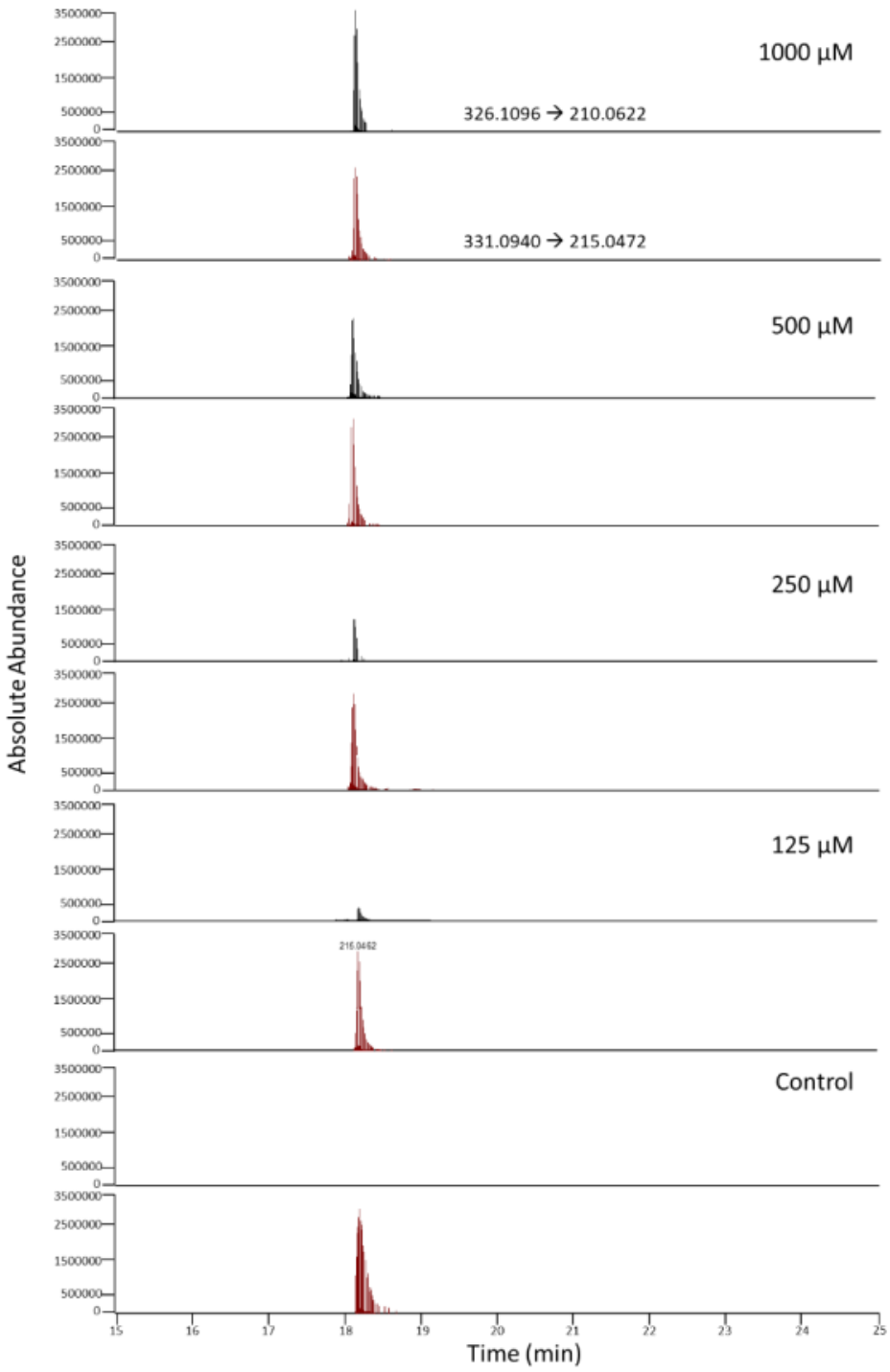


Figure S10. Representative MS^2 Spectra for L-azaserine induced, dose-dependent quantification of O^6-CMdG in HCEC cells. Cells were exposed for 4 h with 0, 125, 250, 500, or 1000 µM L-azaserine, and extracted DNA was analyzed for O^6-CMdG by nanoLC-ESI-$HRMS^2$

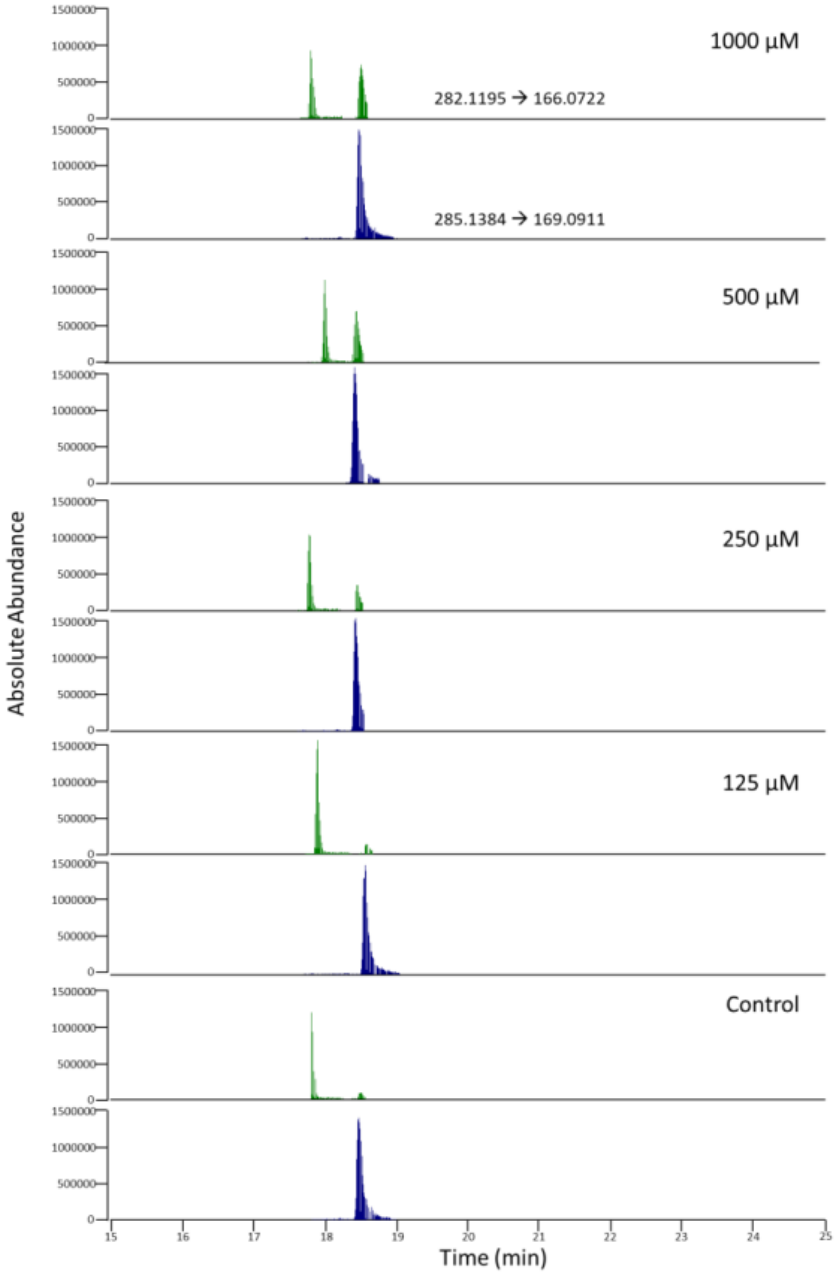


Figure S11. Representative MS^2 Spectra for L-azaserine induced, dose-dependent quantification of O^6 -MedG in HCEC cells. Cells were exposed for 4 h with 0, 125, 250, 500, or 1000 µM L-azaserine, and extracted DNA was analyzed for O^6-MedG by nanoLC-ESI-$HRMS^2$.

Table 5. Half-life ($t_{1/2}$) and k_{obs} values for L-azaserine at varying pHs.

pH	$[H^+]$	$t_{1/2}$[a] (s)	K_{obs}[b] (s^{-1})
2	0.01	1440	0.000481
2.4	0.003981	2520	0.000275
3.4	0.000398	16920	$4.1 \cdot 10^{-5}$
5.8	$1.58 \cdot 10^{-6}$	nd	nd
7.2	$6.31 \cdot 10^{-8}$	nd	nd
8.5	$3.16 \cdot 10^{-9}$	nd	nd
10.5	$1.58 \cdot 10^{-11}$	nd	nd

* $t_{1/2}$ values are derived via the eq. 3 (Fig. 3B); [b] k_{obs} values were derived via HPLC measurements (Fig. S1); nd = not determined

Table S2. Interday and intraday precision and accuracy for the nanoLC-ESI-HRMS^2 method (S/N > 10).

nM	Inter/intraday	***O*^{6}-CMdG**				***O*^{6}-MedG**			
		ctDNA		HCEC		ctDNA		HCEC	
		RSD%[a]	Rec%[b]	RSD%	Rec%	RSD%	Rec%	RSD%	Rec%
0.05	interday	8.9	83.7	6.7	97.8	4.9	118.1	23.5	110.4
0.1	interday	2.6	106.8	12.0	113.4	12.9	104.6	12.6	120.9
1.0	interday	25.6	101.5	1.8	103.5	9.8	122.6	0.8	92.1
50	interday	1.7	99.3	4.9	99.9	3.3	99.4	1.0	98.7
	intraday	2.9	96.4	7.7	98.1	0.9	100.9	2.0	101.1

0.05, 0.1, 1.0 and 50 nM *O*^{6} -CMdG and *O*^{6} -MedG were spiked in ctDNA (15 µg) and isolated DNA (15 µg) from unexposed HCEC cells followed by sample preparation and analysis as described. [a]RSD=relative standard deviation; [b]Rec=recovery (accuracy)