DISS. ETH No. 24973

Fluorescent oligonucleotides as mechanistic probes

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH Zurich

(Dr. sc. ETH Zurich)

presented by

FLORENCE DANIELA BERGER

MSc ETH in Food Science, ETH Zurich, Switzerland

born on 04.08.1988

Citizen of

Fahrni (BE)

Switzerland

accepted on the recommendation of

Prof. Dr. Shana J. Sturla

Prof. Dr. Richard A. Manderville

Prof. Dr. Nathan Luedtke

2018

Für Mami & Papi

Table of Contents

Acknowledgements

First and foremost, my most profound gratitude goes to my advisor Prof. Dr. Shana Sturla for her guidance and for being an inspiring mentor. I am always thankful that you gave me the opportunity and support to pursue part of my research in Canada, and connected me to Prof. Dr. Richard Manderville. You trusted and believed in me from the beginning, and I am especially very thankful that you always promoted my personal strengths and interest. Your advices and support throughout these years shaped me into the person I am now. I will never forget the time in your lab and always be thankful for your great support.

My deep gratitude also goes to my advisor in Canada, Prof. Dr. Richard Manderville. The time in your lab has been a truly invaluable experience. You have been a great mentor and a very inspiring scientist, I could learn a lot from you. Your enthusiasm for science and new ideas have always motivated me, and it was a great pleasure to be part of your lab. Being back in Switzerland, I truly missed all the funny discussions we had about science and sports. I will never forget, sneaking out of the lab to go and watch a game of Roger Federer.

Thanks to Prof. Dr. Nathan Luedtke for accepting to be my co-examiner and for your valuable input regarding the research projects.

Thank you also to our collaborator Prof. Dr. Stacey Wetmore. I have known you throughout my PhD, but unfortunately never got to know you in person. Nevertheless, I appreciate our great collaboration, and the interesting skype talks about the base analogues. Big thanks also goes to Ryan Kung for the excellent collaboration.

I would also like to thank Dr. Chris Osuch for the great collaboration. I will never forget the endless hours at the fluorescence spectrometer, hoping to finally get a cool result and the discussions about skiing and Europe.

Thank you, Dr. Kaila Fadock, you have become a good friend during my stay as a master student, and an even more important person in my first year of my doctoral studies. I am very thankful for your patience in explaining me simple organic chemistry and introducing me to fluorescence spectroscopy. You are a great teacher and an even better friend.

A Bee, thank you for making my stay in Canada such an amazing experience. I remember the first few weeks in Canada, when we went together to class and barely talked, who would have believed that we became best friends. The drives to the soccer games and thirsty Thursday at Brass Taps are truly great memories, which I will never forget.

Big thanks goes to the rest of the Manderville laboratory. Thank you Dr. Prashant Deore, the synthesis god, I am thankful for always helping me, when I needed a second pair of hands. Darian and Thomas, thank you for your kind help and the funny moments in the office. Delaney and Jessica, for to fun nights at brass taps. Special thanks goes also to Dr. Aaron Witham and Dr. Mike Sproviero, you introduced me into the world of synthesis during my Master thesis and even though you graduated before I returned to the Manderville lab for my doctoral studies, your spirit remained in the lab.

Dr. Maureen McKeague, thank you for being truly Canadian. Your positive thinking and your enthusiasm for the most boring thing is just amazing. I was able to learn so much from you, and I am

very glad you always had an open door for all my questions. Thank you a lot for all the help and the very valuable advices.

Thank you, Dr. Arman Nilforoushan, for always helping me with my science question and teaching me some synthesis. Thank you for always motivate me to run more, and for putting so much effort into organizing the SOLA. All our discussion we had about life and your advices meant a lot to me.

Big thanks goes to the present and past members of the Sturla lab. You all contributed to the great spirit in this lab and thanks to you, I was every day more than happy to come to work. Dr. Stefano Malvezzi, thank you for welcoming me so warmly in your office and always bring happiness and fun to the lab. Michael Räz, thank you for being the best roommate and colleague, you are a great and impressive scientist. Thank you, Susanne Geisen, for all the Gin, especially the Susi Gin. Cécile Mingard, thank you for all the funny moments we shared together. Claudia Aloisi, thank you for you endless motivation to organize dinners, without you we would not have shared so many great moments in the lab. Bob and Junzhou, sorry for always mixing up your names, thank you for all the funny moments in your office and your dumplings. Nathalie Ziegler, I will never forget the nights out at the Ski-Week-End and in Barcelona, you were on fire. Dr. Hailey Gahlon, thank you that I could benefit a little bit from your enormous science knowledge. Daniela Kalbermatter, thank you for all your administrational help, and even more for all the nice discussion we had. Thank you to the past and present members of D13, Dr. Alessia Stornetta, Dr. Marianna Stamou, Dr. Simon Sieber, Fabrice Müller and the newest member Dr. Katie Hurley, whenever I needed a break I could come into your office and disturb you to have a little talk. Thank you, Dr. Ioannis Trantakis for bringing some Greek vibe into the lab. Thank you Aida Huber, Sabine Diederich and Mirjam Schneider for all the great talks we had in the kitchen, the quick stop to get tea ended a few times in a half hour break.

I would also like to thank all my friends. Especially Ursina, thank you for always having an open ear for my frustrations. I am a lucky person to have you as a friend. Thank you Béatrice, I could always count on you and the friendship we have is truly remarkable. Thank you, Lucie, I am glad I could always do sports with you and you always understood the frustrations of a doctoral student. Special thanks also goes to Matt Rocher. I am so glad I found the free room in your house. Living there turned out into a great friendship. Through these months in Canada, you became my best friend, and I am very thankful for all the talks we had, and that you always reminded me of just enjoying life.

A big thank you goes to Martin. You always knew what to say to calm me down when I was demotivated and frustrated throughout my doctoral studies. Whenever I spend time with you, I could always forget all my worries and just enjoy the moment.

Zum Schluss möchte ich noch meinen Eltern und meinem Bruder danken. Ohne eure Unterstützung und Liebe wäre ich nie wer ich jetzt bin. Ihr habt mich immer Motiviert, an mich geglaubt und es möglich gemacht, dass ich meine Träume und Wünsche folgen konnte. Nici, du bist der beste grosse Bruder den man sich wünschen kann. Die ewigen Konkurrenzkämpfe wer besser ist, haben mich immer Motiviert noch mehr zu geben, und an mich zu glauben.

Abstract

Exogenous and endogenous chemicals constantly attack DNA and result in DNA adducts, which when left unrepaired may lead to mutations and cancer. Mutations might arise due to a conformational changes to the DNA induced by chemical adducts. Studying the conformational changes induced by DNA adducts may help us to understand how mutations are generated or help to develop tools for early DNA mutation detection. For example, fluorescent modified nucleotides are commonly applied to study DNA conformational changes and to sense single nucleotide polymorphism. In this thesis, different synthetic nucleobase analogues were analyzed as tools to detect and understand DNA mutations and DNA alkylation products.

Chapter 1 provides an overview of the topics discussed in this thesis. The introduction discusses the basis of DNA, and gives a brief overview about DNA damage and repair. In particularly, the use of synthetic modified nucleotides or base analogues to detect DNA damage, mutations or DNA conformational changes is described. Finally, a series of artificial nucleotides are presented in the last part of the introduction.

In *Chapter 2*, a fluorescent modified C8-dG analogue is used to study the structural impact of C8-dG adducts. N-linked C8-dG adducts, derived from arylamines are known to induce frameshift mutations in the *Nar*I recognition sequence. Fluorescence nucleoside analogues are potential probes to study frameshift mutations. The fluorescent C-linked C8-4-fluorobiphenyl-dG (FBP-dG) is used to test the conformational impact induced by C8-dG adducts within the *Nar*I sequence. The stability, fluorescence response and conformational properties of a fully paired and a slipped mutagenic intermediate (SMI) duplex containing FPB-dG are measured. MD simulations and ^{19}F NMR show that the probe favors a *syn* conformation and is stacked within the DNA duplex. FBP-dG significantly stabilizes the slipped mutagenic intermediate (SMI) and exhibits fluorescence sensitivity to distinguish it from the fully paired duplex. Molecular dynamics simulations and optical spectroscopy predict increased rigidity of the biphenyl in the *syn* conformation. The turn-on emissive properties of FBP-dG in the SMI duplex may be a powerful tool for monitoring frameshift mutations upon synthesis with DNA polymerases.

In *Chapter 3*, the fluorescent FBP-dG is used to study the mechanism of frameshift mutations. Frameshift mutations arise through base slippage in a sequence dependent manner. We examined a series of DNA duplexes containing FBP-dG underlying the propensity of C8-dG adducts to form frameshift mutations. In the *Nar*I sequence, the fluorescence intensity increases upon inserting a base opposite FBP-dG (n+1 duplex). Thus, we suggest that FBP-dG bulges out after inserting a base opposite FBP-dG. In the CG3 sequence, the fluorescence does not increase after inserting one base opposite FBP-dG (n+1 duplex), therefore we suggest that in this sequence context FBP-dG can correctly pair with the incoming base. Next, we test extension of FBP-dG containing sequence using a model DNA polymerase Kf-. FBP-dG stalls DNA polymerase in the *Nar*I sequence more than the CG3 sequence. In this study, we provide a tool to study the frameshift mutations using a fluorescent probe, however the sensitivity of the probe is limited to applications without a complex background.

In *Chapter 4,* we explore the properties of two elongated nucleoside analogues (ExBIM and ExBenzi) as probes to detect O^6-MeG within a mutational hotspot of the human KRAS gene. Based on dangling end studies and thermodynamic analysis, we conclude that the elongated analogues increase the π-stacking interactions more than smaller base analogues. ExBenzi or ExBIM stabilize DNA duplexes when placed opposite O^6-MeG, whereas no significant stabilization is observed when the analogues are placed opposite a series of other modified DNA bases (O^6-BnG, N^2-MeG, 8-oxo-G, orN^6-MeA). Additionally ExBIM shows turn-on emissive properties opposite O^6-MeG, but not opposite the other

modified DNA bases. ExBenzi does not show fluorescence sensitivity to distinguish the different DNA lesion. Molecular dynamic simulations suggest that the analogues stabilize duplexes when placed opposite O^6-MeG through base stacking interactions with the flanking bases. The ExBIM:O^6-MeG pair is more shielded from the bulk solvent, and has less hydrogen bond interaction than the ExBenzi:O^6-MeG pair. Taken together this results provide the requirements for future probe development to detect specifically modified DNA bases.

In *Chapter 5*, the results of the doctoral thesis are summarized and critically evaluated. The possibilities and limitations of different artificial nucleosides analogues as tools to detect DNA mutations and DNA damage are provided.

In *Appendix A,* a study is presented which is the basis for the development of an exponential amplification method using artificial base pairs as markers opposite alkylated thymidine. dTPT3 and d5SICS are artificial nucleosides, which have a specific base partner. dTPT3 is efficiently incorporated opposite O^2- and O^4-methyltimidine. However, natural nucleosides are also incorporated opposite O^2- and O^4-methyltimidine, resulting in the loss of the DNA signal during exponential amplification. Therefore, research is ongoing to enrich the marked DNA, to keep the signal during amplification.

In *Appendix B*, a fluorescently modified thymidine analogue is used as a probe to determine the conformation of thymidine in the thrombin binding aptamer (TBA). 5-furyl-2'-deoxyuridine (^{Fur}dU) is a fluorescent probe with molecular rotor properties and exhibits dual probing characteristics, providing changes in emission wavelength and intensity with increased solvent rigidity. TBA is an aptamer, which specifically binds to the blood clotting enzyme thrombin, and binding of thrombin results in folding of the aptamer into a G-quadruplex (GQ). ^{Fur}dU serves as a diagnostic tool to sense interaction of T with the molecular target.

Zusammenfassung

Die Nukleobasen der Desoxyribonukleinsäure (DNA) werden konstant von exogenen und endogenen Stoffen angegriffen. Solche Stoffe finden sich in Abgasen, in der Nahrung oder im Rauch von Zigaretten. Metabolische Aktivierung führt zu elektrophilen chemischen Verbindungen, die die nukleophilen Basen angreifen, wodurch Schäden an der DNA entstehen. Modifikationen an Basen verhindern eine normale Basenpaarung, es können Mutationen entstehen, welche zu Krebs führen können. Ausserdem verändern modifizierte Nukleobasen die DNA Doppelhelix Struktur, welche zu Veränderung der Leseraster führen. Eine Veränderung des Leserasters resultiert in falscher Codierung von Proteinen, und schlussendlich kann Krebs entstehen. In dieser Doktorarbeit wird vor allem die Leserastermutationen erforscht.

In *Kapitel 1* gibt es eine Einleitung zum fachlichen Inhalt dieser Doktorarbeit. Zuerst werden die DNA und die Struktur der DNA im Allgemeinen erklärt. Es folgt dann die Beschreibung von verschiedenen sekundären Strukturen, sowie von verschiedenes alternativen Applikationen von DNA. Im nächsten Abschnitt werden verschiedene DNA Schäden und Mutationen diskutiert. Zum Schluss werden noch verschiedene synthetische Basen und Basenpaare vorgestellt.

Kapitel 2 beschreibt die Synthese eines fluoreszierenden diphenyl Desoxyguanosin Addukts (FBP-G), welches in eine DNA Sequenz inkorporiert wird. FBP-G ist in einem heterogenen Gleichgewicht zwischen einer normalen B-Typ und einer interagierenden Struktur. In der interagierenden Struktur hat FBP-G Wechselwirkungen mit den Nachbarsbasen. Diese Wechselwirkung stabilisieren auch eine Doppelhelix, welche eine 2-Basen Ausbuchtung hat. Es ist bekannt, dass die Stabilisation der 2-Basen Ausbuchtung zu einer Leserastermutation führen kann. Zusätzlich ist FBP-G stark fluoreszierend, wenn es in dieser Ausbuchtung platziert ist.

In Kapitel 2 konnten wir eine fluoreszierende DNA Basen herstellen, welche das Potential zur Detektion von Leserastermutationen hat. Leserastermutationen sind Sequenzabhängig, daher haben wir das FBP-Guanine Addukt in *Kapitel 3* in zwei Sequenzen inkorporiert, welche Unterschiedliche Basen Abfolgen haben. Die modifizierten Sequenzen wurden mit komplementären Sequenzen von unterschiedlichen Längen hybridisiert, welche die verschiedenen Schritte der Leserastermutation während der DNA Replikation simulieren sollen. Anhand der thermodynamischen Stabilität der Doppelhelices und der Veränderung der Fluoreszenz konnte einen Unterschied in der Entstehung der Leserastermutation zwischen den zwei Sequenzen bestimmt werden. Dieser Unterschied hat dann auch einen Einfluss auf die DNA Replikation, welche zu einer Sequenzabhängigen Blockierung der Polymerase führt. Zusammen mit den Resultaten aus *Kapitel 2* konnten wir nun zeigen, dass mit einfachen chemischen Methoden und synthetischen Basen, Mutationen besser verstanden werden können. Diese Studien dienen als Grundlagen für die Entwicklung von modifizierten Nukleobaen zur Nachweisung von DNA Schäden und Mutationen.

In *Kapitel 4* werden synthetische Nukleinsäuren (ExBIM und ExBenzi) zur spezifischen Basenpaarung von O^6-alkyl-Guanine Addukten als DNA Hybridisierungsproben untersucht. ExBIM und ExBenzi stabilisieren spezifisch Doppelhelices, wenn sie Gegenüber O^6-Methyl-Guanine (O^6-Me-Guanine) platziert sind. Alle anderen modifizierten DNA Basen (N^2-Me-Guanine, 8-oxo-Guanine, N^6-Me-Adenine) wurden durch diese Analoge nicht stabilisiert. Ausserdem waren Doppelhelices welche ExBIM gegenüber O^6-MeG enthielten stark fluoreszierend. ExBIM und ExBenzi haben starke Wechselwirkungen mit den umliegenden Basen, wenn sie gegenüber O^6-Me-Guanin platziert sind, was nicht für ein unmodifiziertes Guanin zutrifft. Ausserdem hat ExBIM weniger Hydrogen Verbindungen mit O^6-Me-Guanin als ExBenzi und ist besser vom herumliegenden Lösungsmittel geschützt. Die

Strukturellen Eigenschaften in dieser Studie dienen für zukünftige Entwicklungen von fluoreszierenden Analogen, welche zur Detektion DNA Schäden gebraucht werden.

In *Kapitel 5* wird die Arbeit zusammengefasst und die Forschungsergebnisse kritisch diskutiert. Synthetische Nukleinsäuren können für verschiedene Zwecke gebraucht werden. Die Ergebnisse dieser Doktorarbeit dienen als Grundlage für zukünftige Studien um biologische Prozesse mithilfe von synthetischen Nukleinsäuren zu verstehen oder Methoden zu entwickeln um Mutationen frühzeitig zu erkennen. Für diese Zwecke ist es wichtig, dass die synthetischen Nukleinsäuren intensiv fluoreszieren und in Zellen eingebaut werden können.

Appendix A beschreibt die Untersuchung von synthetischen DNA analogen als spezifische Nukleobasen für alkylierte Thymine (O^2- und O^4-Methyl-Thymine (O^2- und O^4-Me-Thymine). Durch eine effiziente Inkorporation von synthetischen Basen gegenüber O^2- und O^4-Me-Thymine, kann der Schaden markiert und amplifiziert werden. Dadurch werden DNA Schäden mit geringer Häufigkeit künstlich vermehrt und die Detektion ist erleichtert. Es wurde ein synthetisches Triphosphat gefunden, welches effizient gegenüber O^2- und O^4-Me-Thymine eingebaut wird. Leider ist die Inkorporation von natürlichen Nukleinsäuren gegenüber O^2- und O^4-Me-Thymine genauso gut. Durch die Inkorporation von natürlichen Basen, würde die Amplifikation der unmarkierten Sequenz viel schneller ansteigen, als die der markierten Sequenz und die Detektion der markierten Sequenz ist nicht möglich. Diese Studie dient als Basis um eine Methode zu entwickeln, um alkylierte Thymine zu detektieren.

Appendix B beschreibt die Anwendung von fluoreszierenden DNA Sequenzen als Werkzeug um ein Enzym der Blutgerinnung zu erkennen. Dabei wurde ein synthetisches Thymine (^{Fu}T) in ein Aptamer, welches das Enzym bindet, inkorporiert. ^{Fu}T reagiert sensitiv auf eine Veränderung der Mikroumgebung, das heisst wenn die Viskosität der Mikroumgebung zunimmt, steigt auch die Intensität der Fluoreszenz. Durch diese Eigenschaft konnten wir bestimmen, welche Thymine mit dem Enzym interagieren und konnten so die Bindungsaffinität des Aptamers erhöhen. Durch dieses Projekt konnten wir zeigen, dass modifizierte Aptamere geeignete Werkzeuge sind um Therapeutika zu entwickeln.

Abbreviations

2-AP	2-aminopurine
^{3n7z}A	8-aza-3,7-dideaza-2-deoxyadenosine
8-oxoG	8-Oxoguanine
A	Adenine
AAF	Acetylaminofluorene
ABP	4-aminobiphenyl
AF	Aminofluorene
AGT	O^6-alkylguanine alkyltransferase
alkylT	O^4- and O^2-alkylthymidine
AP	Abasic site
B[a]P	Benzo[a]pyrene
BDF	Base discriminating fluorescence
Benzi	2(1*H*)-Benzimidazolone
BER	Base excision repair
BIM	1*H*-Benzimidazole
C	Cytosin
CD	Circular dichroism
CNPh-dG	C8-(4-cyanophenyl(-2'-deoxyguanosine
DBS	Double strand breaks
DCM	Dichloromethane
ddNTPs	2',3'-Dideoxynucleoside-5'-*O*-triphosphate
^{DMA}C	N,N-dimethylaniline-20-deoxycytosine
^{DMA}T	N,N-dimethylaniline-20-deoxythymidine
DNA	Deoxyribonucleic acid
dNTP	2'-Deoxynucleoside-5'-*O*-triphosphate
Dpo4	DNA polymerase IV of *Sulfolobus solfataricus*
Ds	7-(2-thienyl)-imidazo[4,5-b]pyridine
EDTA	Ethylenediaminetetraacetic acid
ESF	Environmental sensitive fluorophore
EtOAc	Ethyl acetate
ExBenzi	1*H*-Naphtho[2,3-*d*]imidazole-2(3*H*)-one
ExBIM	1*H*-Naphtho[2,3-*d*]imidazole
FBP-dG	C8-F-biphenyl-2'deoxyguanosine
FRET	Fluorescence resonance energy transfer
Fu-dG	C8-furan-2'-deoxyguanosine
^{Fu}U	5-furyl-2'deoxyuridine
G	Guanine
GQ	G-quadruplexes
HCA	Heterocyclic aromatic amines

HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
IQ	2-amino-3-methylimidazo[4,5-*f*]quinolone
isoC	Isocytosine
isoG	Isoguanine
Kf	Klenow fragment of DNA polymerase I form *E. Coli*
KRAS	Kirsten rat sarcoma viral oncogene homologue
LC-MS	Liquid chromatropaghy-mass spectrometry
MD	Molecular Dynamics
^{MD}A	Methoxybenzodeazaadenine
^{MD}I	Methoxybenzoadeazainosine
MeOH	Methanol
Mg^{2+}	Magnesium ion
MMR	DNA mismatch repair
N^6-MeA	N^6-methyladenine
$NaHCO_3$	Sodium bicarbonate
*Nar*I	5′-CGGCGC-3′
NER	Nucleotide excision repapir
NHEJ	Nonhomologous endjoining
NMR	Nuclear magnetic resonance
NNAL	Methylnitrosamino-1-(3-pyridyl)-1-butanol
NNK	4-(methylnitrosamino)-1-(3-pyridil)-1-butanone
nt	Nucleotide
O^2-EtT	O^2-ethylthymidine
O^2-MeT	O^2-methylthymidine
O^2-PHBT	4-(3-pyridyl)-4-hydroxybut-1-ylthymidine
O^2-POBT	2^6-4-(3-Pyridyl)-4-oxobutylthymidine
O^4-EtT	O^4-ethylthymidine
O^4-MeT	O^4-methylthymidine
O^6-BnG	O^6-benzylguanine
O^6-CMG	O^6-carboxymethylguanine
O^6-MeG	O^6-methylguanine
OTA	Ochratoxin A
P	[2-amino-8- (1′-β-D-2-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)one
Pa	pyrrole-2-carbaldehyde
PAH	Polycyclic aromatic hydrocarbons
PCR	Polymerase chain reaction
Per	1*H*-Perimidin-2(3*H*)-one
Peri	1*H* Perimidine
PhIP	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine
Px	2-nitro-4-propynylpyrrole

Py-dG	C8-pyrene-2'-deoxyguanosine
qA	quadracyclic adenine analogues
RNA	Ribonucleic acid
SELEX	Systematic evolution of ligands by exponential enrichment
SMI	Slippage mutagenic intermediate
SMRT	Single molecule real-time sequencing
SNP	Single nucleotide polymorphism
ssDNA	Single strand DNA
SSO	Semi-synthetic organism
T	Thymine
Taq	DNA polymerase I from Thermus aquaticus
TBA	Thrombin binding aptamer
tC	tricyclic cytosine analogues
TEAA	Triethylammonium acetate
Th-dG	C8-thiophene-2'-deoxyguanosine
THF	Tetrahydrofuran
TLS	Translesion synthesis
T_m	Thermal melting temperature
U	Uracil
Z	6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone

Chapter 1

Introduction

1.1 DNA structure and function

The discovery of the DNA double helix by Watson and Crick in 1953 set a landmark for future research on genomic integrity.[1] DNA is composed of the four nucleobases, adenine (A), cytosine (C), guanine (G) and thymine (T). Two nucleobases (A with T, G with C, Figure 1) form a pair through Watson-Crick hydrogen bonds resulting in a right-handed DNA double helical structure, called B-DNA.[1] After the discovery of the structure, Crick described the central dogma of biology. The classical view states that DNA is transcribed into RNA (mRNA) and mRNA is synthesized into proteins.[2] This means that the hereditary information is stored in DNA, and is passed to the next generation by copying the DNA, or it is translated via mRNA into functional proteins. Therefore, it is fundamental that the DNA is replicated without error to ensure the integrity of the genome from generation to generation.

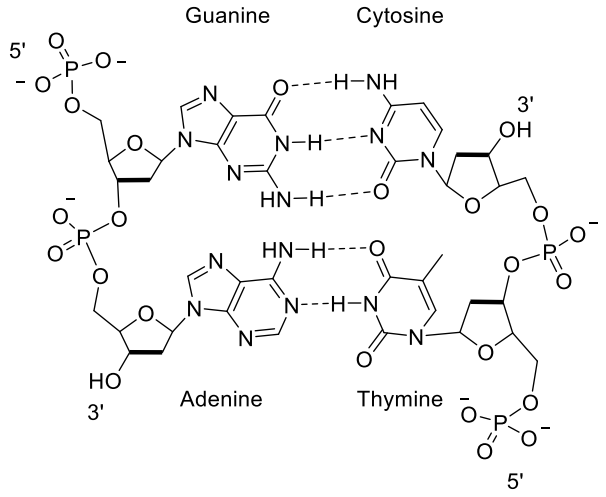


Figure 1: Watson-Crick hydrogen bonding between two anti-parallel DNA sequences to form a DNA double helix.

DNA is constantly replicated using an enzyme-catalyzed process that requires a single-stranded DNA template.[3] DNA polymerase maintains correct replication of DNA catalyzing the accurate and efficient incorporation of nucleotides (dNTPs) to a primer-template DNA in 5'→3' direction.[4] The cell has an elaborate mechanism for high fidelity replication (one error per 10^7-10^8 insertion[5,6]), which is a sequential multistep process involving several polymerases. The greatest contribution to the high fidelity is the nucleotide selectivity of the DNA polymerase.[6]

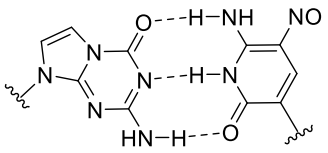
1.1.1 DNA damage and repair

The cell has several repair mechanisms and a DNA damage tolerance system to repair and tolerate any DNA damage,[7] which occurs through attack of endogenous and exogenous chemicals.[8–10] Repair processes include for example nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR). These mechanisms have a similar principle; splicing out the damaged region, inserting a new base into the gap and ligating of the strand.[11–13] DNA damage can escape the repair machinery, resulting in blocking of DNA replication. This stalling recruits specialized translesion synthesis (TLS) polymerases. These polymerases have lower fidelity and selectivity than replicative polymerases, therefore they may be error-free or error prone depending on which base is incorporated opposite the damaged base. Error-free bypass contributes to genomic stability, whereas error-prone bypass increases DNA mutations.[14,15]

1.1.2 Early DNA damage detection with polymerases

Discovered in 1958 by Arthur Kornberg,[16] polymerases catalyze the accurate and efficient incorporation of nucleotides (dNTPs) to a primer-template DNA in 5' →3' direction. Polymerases are important tools in research, for example allowing polymerase chain reaction (PCR) and sequencing. Their application has helped researchers study DNA mutations and early detection of DNA damage.[17,18] The Nobel Laureate (1993) Kary Mullis developed PCR in 1983, one of the largest breakthroughs in biotechnology of the 20th century. He simply put together already known techniques such as synthesis of oligonucleotides and use of DNA polymerases to synthesize new DNA copies.[19] In PCR, DNA is exponentially amplified with a DNA polymerase in multiple cycles of (1.) Denaturation, (2.) Primer annealing, and (3.) Synthesis of DNA.

PCR has recently become a standard tool for DNA damage detection using marker nucleotides, these marked sequences can then be sequenced. However, for specific and efficient incorporation of artificial nucleotides, new DNA polymerases need to be engineered. For example, large dye labelled nucleotides have been successfully incorporated into DNA sequences using a variant of *Pyrococcus furiosus* DNA polymerase (Pfu). In this manner Cy3-or Cy5-dCTP are incorporated in place of dCTP to generate brightly colored and highly fluorescent DNA oligonucleotides.[20] As another example, a variant of Taq polymerase was engineeredto accommodate the artificial DNA base pair (dZ:dP, Figure 2).[21] Once PCR has been successfully applied to amplify damaged DNA bases with an artificial marker nucleotide, the marker can bedetected using DNA sequencing technologies.[22]



dP dZ

Figure 2: dP-dZ base pair.

DNA sequencing is another important technique that relies on DNA polymerases. Sanger sequencing was first developed in 1977[23] and was used in 2001 to sequence the initial sequencing of the human genome.[24,25] While it took over decade to sequence the human genome with Sanger sequencing, next generation sequencing (NGS) takes less than a day.[26] Furthermore, the sequecing costs have dropped drastically, so that NGS is widely applied for genetic testing in families with histories of cancer genes (TP53 or KRAS).[27]

1.1.3 Nucleic acid structures and applications

In addition to the canonical right-handed double helix, alternative DNA structures can be found in biological systems. A wide variety of biologically-relevant structures exist,[28,29] which are formed when DNA gets unwound during DNA replication or transcription.[29] Examples for secondary structures are G-quadruplexes (GQ), triplex, cruciform and slipped structure are the most found secondary DNA structures.[30] Cruciform or hairpin are formed with intrastrand pairing with inverted repeat sequences.[29] They are targets for several proteins, which are involved in chromatin organization, transcription, replication and DNA repair.[31] The occurrence of hairpin structure, can induce DNA replication slippage. Triplexes are formed when a double helix that contains only purines binds to a third strand of polynucleotides.[29] They may be found *in vivo* and may play a role in molecular switches to modulate gene expression or other DNA metabolisms and regulations.[32] GQ form in G-rich regions through intra-

or inter-strand interactions and stabilized by four planar Gs, paired by Hoogsten bonding.[33,34] GQ are involved in biological processes as, for example, transcriptional regulation, DNA replication and genome stability.[35–37] GQ are mostly found in telomeres (tandemly arranged TTAGGG repeat), where they lead to telomere stabilization.[38,39]

Inspired by the diverse structure and function of nucleic acids, DNA molecules with new functions and properties have been engineered through direct synthesis and evolution. Aptamers are one prominent example. These are short oligonucleotides, which specifically bind to a molecular target. They are selected using systematic evolution of ligands by exponential enrichment (SELEX), which was developed in 1990 by Tuerk and Gold.[40] Upon target binding, aptamers fold into secondary structures (GQ, loop, stem, hairpin), which are essential for binding affinity and specificity.[41,42] Aptamers show potential for the development of early detection assays or targeted and personalized therapies. Using synthetic base analogues, the stability and binding affinity of aptamers can be improved. As for example, the dDs-dPx unnatural basepair (Figure 3) increased the binding affinity of the anti-VGEF aptamer, compared to the aptamer with natural bases.[43] Fluorescent modified bases can be used to detect target binding to the aptamer. A fluorescent modified C-linked aryl group has been incorporated into the thrombin binding aptamer (TBA) and shows turn-on emissive properties upon GQ folding induced by thrombin binding.[44]

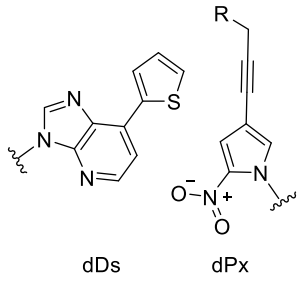


Figure 3: dDs-dPx base pair.

1.2 DNA damage and mutations

DNA bases are targets of damage by metabolites of genotoxic chemicals, which covalently bind to DNA bases resulting in DNA adducts. If these damaged DNA bases evade cellular repair, they might result in mutagenesis and ultimately carcinogenesis.[45,46] Endogenous sources are for example reactive oxygen or nitrogen species, and exogenous sources derive from a wide range of chemicals from the environment and food.[46] Polycyclic aromatic hydrocarbons (PAHs) are a large class of chemicals, derived from processing and burning of coal,[47,48], which generate bulky DNA adducts.[49] Alkylating N-nitroso compounds found in red and processed meat,[50–53] and tobacco smoke,[54–56] are known to form alkylated DNA adducts.[50,56,57] Exposure to aromatic amines and amides from industrial combustion, cigarette smoke and certain foods, can lead to the formation of C8-guanine adducts.[58,59]

1.2.1 Alkyl adducts

DNA alkylating agents are major contributors to DNA damage. Processed meat,[50,52,53] tobacco smoke,[54,60] and other exogenous or endogenous sources are known to contain DNA alkylating agents. DNA bases have a high electrostatic potential, therefore they are prone to react with alkylating agents.[61,62] 60 to 80 % of the total alkyl-adducts are N7-guanine adducts, they are the least cytotoxic or mutagenic.[63,64] N3-methyladenine is another abundant N-methylation product, which accounts for 10-20 %, however they are more mutagenic because they block the polymerase synthesis.[65,66] N1 of guanine and N3 of cytosine are minor alkylation sites in the DNA. O^6-alkylguanine is the most biologically-

relevant adduct, it results in G to A transitions in the cancer genes KRAS[67] and p53.[68] O^2- and O^4-alkylthymidine are less abundant, but are also important. They are resistant to repair and persist longer in mammalian tissues.[69,70]

1.2.2 C8-deoxyguanosine adducts

Arylamines and heterocyclic aromatic amines

2-aminofluorene (AF) and 2-acetylaminofluorene (AAF) belong to organic chemicals called arylamines and are representative of an important class of mutagens (Figure 4A). After metabolic activation of AF or AAF, major adducts are formed at the C8 position of guanine through covalent linkage of the amine or amide nitrogen.[71,72] AF and AAF were first developed as pesticides, but their mutagenic potential was discovered before they were used. However, they are perfect model carcinogens to understand the mutagenicity of genotoxic compounds and have been extensively studied. In 1941, Wilson et al. observed that metabolic activation of chemical compound was required for carcinogenesis.[73] But it was only until Watson and Crick solved the DNA structure,[1] that DNA adducts were discovered. The AAF-dG adduct was one of the first example of a synthesized carcinogen-DNA adduct and paved the way for studies of mutagenesis derived by genotoxic compounds.[74] Mutagenesis studies have shown that AAF-dG induces frameshift mutations, especially when placed in the *Nar*I sequence: 5'-$G_1G_2CG_3$CC-3' at position of the G_3[75–78] in *E.coli,*[75,79] whereas AF-dG in the same sequence result in point mutations. These model carcinogen links the conformational impact of the adduct on the DNA helix and the resulting mutational outcome.

AF-dG and AAF-dG are structurally similar, however AF-dG lacks the acetyl group at the nitrogen atom, which links the fluorene moiety with dG. Despite their structural resemblance, they have a different impact on the DNA duplex structure. Thus AF-dG and AAF-dG are interesting models for comparing the influence of the linkage on the DNA conformation and the therein resulting mutational outcome. AF-dG is in a conformational heterogeneity in dependence of the sequence context between the normal B-type and the base displaced stacked conformation.[80–83] Purine bases 3' to AF-dG favors the stacked conformation, whereas pyrimidine bases 5' to AF-dG favor the B-form.[84,83,85] X-ray studies of the DNA synthesis past the AF-dG showed, that the glycosidic angle switches during DNA synthesis form a *syn* conformation into an *anti* conformation. The *anti* conformation does not disturb the DNA helix and therefore it promotes the correct insertion of dCTP opposite the dG adduct.[86] On the other hand, the acetyl substituent on the N-linker of AAF-dG restricts the rotation of the glycosidic bond from the *syn* conformation into the natural-like *anti* conformation. Thus the polymerase is unable to incorporate any dNTP opposite the adduct in this manner AAF-dG stalls the polymerase.[87–89]

4-aminobiphenyl (ABP), another well studied arylamine, has been found in cigarette smoke and is related to bladder cancer.[90] ABP lacks the methylene bridge between the two phenyl rings, thus the biphenyl moiety is twisted and is more flexible. ABP-dG tends to adopt a B-form conformation and is nonplanar.[81,91,92] ABP-dG is expected to be less mutagenic than AF-dG and AAF-dG, due to the lower tendency to adopt the pro-mutagenic *syn* conformation.

2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ) is a heterocyclic aromatic amines (HCA) formed through pyrolysis of carbohydrates and amino acids (Figure 4B).[93–95] After metabolic activation, it forms adducts at the C8 and N7 positions of dG.[96–99] As AAF-dG, C8-IG-dQ adopts a base-displaced intercalated conformation in the *Nar*I recognition sequence.[100] Replication past C8-IQ-dG adduct in the *Nar*I sequence using different high fidelity and low fidelity polymerases results in frameshift mutations.[101]

2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) is another member of the HCAs and is formed during cooking of red meat (Figure 4B).[102] It also forms adducts at the C8 position of dG, C8-PhIP-dG adopts similar motifs in DNA as other arylamine adduct, however the adduct has greater flexibility and the rings can rapid rotate. The phenyl ring, has an out of plane geometry, and therefore contributes to a larger unwinding and twisting of the DNA helix,[103] the heterocyclic rings of PhIP can easier intercalate within the helix as the structural similar ABP-dG.[81] Replicative polymerase stall at the PhIP-dG adduct, whereas TLS polymerase κ has the potential to form a single base deletion through a slippage product.[104]

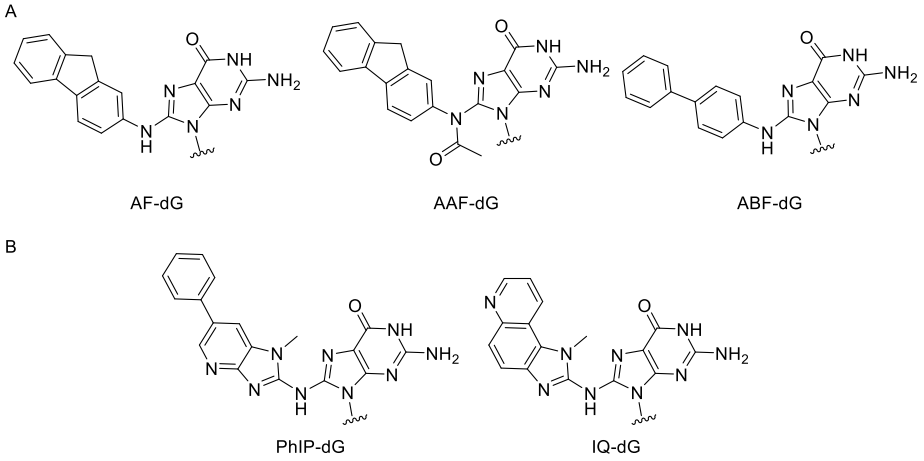


Figure 4: N-linked C8-dG adducts: A) Adducts derived from arylamines. B) Adducts derived from heterocyclic aromatic amines.

The studies on arylamines and heterocyclic amines show the importance of the adducts structure on the mutational outcome. Adducts which seem to be similar in shape or size, but small changes in the linkage, or flexibility can result in different conformation within the helix and thus result in a more or less severe mutagenic outcome. Several C8-dG adducts have been observed, arylamines mostly form an N-linkage between the carcinogen and the C8 position of dG. However, C8-dG adducts with a direct C8-C-linkage have also been detected.

C-linked C8-dG adducts

Carcinogens such as the food toxin ochratoxin A (OTA)[105] or benzo[a]pyrene (B[a]P), which is found in cigarette smoke[49] form a direct C8 C-linkage to dG (Figure 5). Despite playing an important role in carcinogenicity, relatively few studies exist that focus on the conformational preferences of C-linked adducts. It is unclear if the mutagenic and carcinogenic outcome of the N-linked adducts can be translated to C-linked adducts. In contrast to the N-linked C8-dG adducts, the C-linked adducts lack the flexible tether, and thus the impact on DNA conformation might be altered.[106,107] Indeed, an acetyl substituent on the N-linker of AAF-dG leads into different conformational heterogeneity, compared to AF-dG.

A series of C-linked C8-aryl-dG adducts with aryl rings of different sizes and shapes have been tested on their conformational preference within the helix and the resulting mutational outcome.[106,108] The impact on the DNA conformation and the resulting mutagenicity is dependent on the aryl ring size. Smaller adducts, like C8-furan-dG (Fu-dG) that prefer an *anti*-glycosyl conformation induce a realignment process, resulting in a C:C mismatch downstream of the adduct. Whereas bigger adducts, like C8-pyrene-dG (Py-dG), that prefer *syn*-glycosyl conformation cause a 2-base slippage (Figure

5).[106,109] Smaller single aryl carcinogens are not able to stable an intercalated conformation through π-π interactions, which larger aryl carcinogens can.

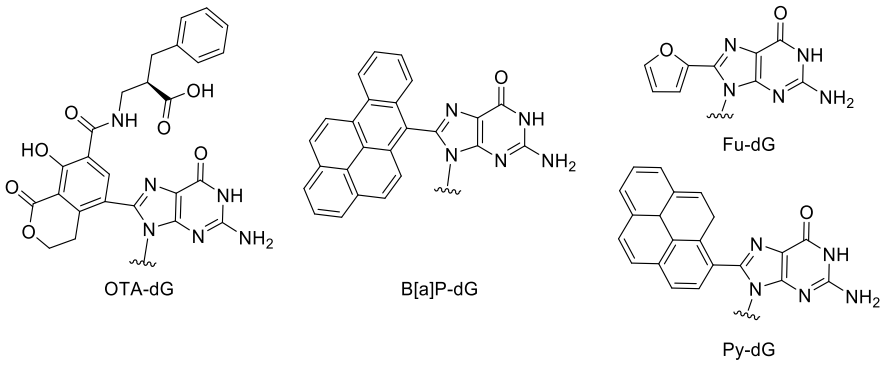


Figure 5: C-linked C8-dG adducts.

1.2.3 DNA mutations

Upon DNA replication, the polymerase may incorrectly copy DNA, resulting in DNA mutations. Point mutations include base substitutions, deletions or insertions. Base substitutions can be divided into two types: transition or transversion. In transition mutations, a purine base is replaced by a purine base, or a pyrimidine base by a pyrimidine base, e.g. G:C to A:T. Transversion mutations on the other hand, replaces a purine base by a pyrimidine base or a pyrimidine base by a purine base e.g. G:C to T:A.[110] Point mutations change the sequence, and three different outcomes are possible. (1) the new codons still codes for the same amino acid (silent substitution), (2) the new codon does not code for the same amino acid (missense mutation), (3) the new codon is a translation termination codon (nonsense mutations).[110,111] Frameshift mutations are a kind of point mutations and result in a change of the reading frame, which makes them more severe mutations. A central question in the research described in this thesis is how frameshift mutations are generated and to understand the mechanism of frameshift mutations.

One of the most dangerous forms of DNA damage is the double strand break (DSB). DSBs arise through ionizing radiation, chemicals, endogenously generated reactive oxygen species or from mechanical stress, but also endogenously during DNA replication.[112] When not recognized and repaired, DSBs result in cell death. Fortunately, DSBs can be repaired using several mechanisms, as for example nonhomologous end joining (NHEJ) and homologous recombination (HR).[112,113]

Genome sequencing has made it possible to find genome alterations in human cancer genes. The discovery of the first cancer-causing sequence, the codon 12 of the HRAS gene[114] launched the quest for more abnormal cancer genes. Mutations induced by exogenous carcinogens, like polycyclic aromatic hydrocarbons (PAH), alkylating agents or arylamine are common in the tumor suppressor gene TP53.[68] TP53 gene is mutated in 50 % of all cancers.[115,116] The proto-oncogene KRAS also is frequently mutated in lung[117] and colon cancer.[118] The mutational outcome is specific for each cancer type and is dependent on exogenous carcinogens. For example, exposure to PAHs result in a C to A mutation in lung cancer,[119] melanoma has a frequent C to T mutation due to misrepair of ultraviolet-induced pyrimidine dimers[120] and gastrointestinal tumors have more transition mutations at CpG dinucleotides, because of methylation of the C induced by alkylation agents.[118] Determination of cancer genomes and mutational signature is important to develop tools for early detection of cancer.

Frameshift mutations

Frameshift mutations, which are defined as the loss or gain of 1 or 2 bases, change the reading frame and result in proteins with diminished functions, or with complete loss of functions.[121] Frameshift mutations might occur spontaneously, especially in microsatellite instability (MSI), which are defined as monotonous runs of base repetitions (e.g. GGGGG). MSI in combination with mismatch repair deficiencies is associated with colorectal cancer.[122–124]

Chemicals, especially those with bulky, aromatic moieties induce frameshift mutations. These chemicals either react with the DNA to form a DNA adduct or they intercalate within the DNA helix without forming a DNA adduct. Acridines, for example, belong to model frameshift mutagens and are well-known examples for noncovalently intercalating agents.[125,126] The *Nar*I restriction enzyme recognition site 5'-$G_1G_2CG_3CC$-3' is a known hotspot for frameshift mutations.[75] For example, the frameshift mutation rate is higher, when the G_3 bears an AAF-dG adduct, however no mutation is observed when G_1 or G_2 are modified.[127] AAF-dG induces a rotation of the glycosidic angle from its normal *anti* into its pro-mutagenic *syn* conformation.[86] Due to the rotation of the glycosidic angle, AAF-dG forms a bulge out form with the flanking C into a -2 base misalignment through the Streisinger slippage mutagenic intermediate.[128] G_1 and G_2 lack the repetitive GC dinucleotide, and therefore frameshift mutations occurs at lower frequency. For frameshift mutations to occur it is important that the glycosidic angle of the adducted G is in a *syn* conformation, and is hindered to turn back into its natural *anti* conformation.[87–89,106,109] Frameshift mutations mainly arise through a change in DNA conformation resulting from these adducts, however we lack tools to detect and rapid study conformational changes.

1.3 Unnatural base analogues

1.3.1 Fluorescent base analogues

Fluorescence Spectroscopy is a dominant methodology in biological sciences, such as flow cytometry, medical diagnostics, DNA sequencing, cellular and molecular imaging. Fluorescence describes the emission of light from molecules that are excited. Electrons are normally in the lowest energy state S_0, and with energy the electron gets to an excited state S_1 or S_2. In this state the electron rapidly (10^{-12} seconds) loses its energy and falls back to the lowest level S_0.[129]

Nucleobases themselves are non-fluorescent under normal conditions, therefore a fluorophore has to be covalently or non-covalently introduced to the system.[130] The ideal fluorescent base analogue should have high fluorescence intensity with sensitivity to its local environment, should not destabilize base pairing interactions or be quenched in duplex DNA. Furthermore they should be applicable in phosphoramidite chemistry or be convertible into triphosphate.[131,132]

Fluorescent modified nucleobases are commonly applied to study nucleic acids, including single nucleotide polymorphism detection,[131,133–136] conformational and structural changes of DNA,[109,137–140] and enzyme activity testing.[141–143] External or non-covalently attached fluorophores are used, for example, in gel electrophoresis or cell microscopy. Ethidium bromide[144] or SYBR Gold[145] are dyes which are quenched in water, but fluoresce upon intercalating within single or double strand DNA. Fluorophores can also be attached to the backbone of the DNA strand either at the end or within the oligonucleotide sequence.[146,147] Attaching the fluorophore to the backbone of the DNA strand is generally less disruptive to the DNA conformation. Commercially available fluorophores such as

fluorescein or cyanine dyes have extremely high quantum yield and thus are applied in gel electrophoresis, fluorescent microscopy, and single molecule studies.[130]

Internal modifications involve the replacement of natural DNA bases with fluorescent modified base analogues. They either maintain the canonical scaffold or are planar aromatic analogues with different sizes or shapes. The design and choice of the nucleobase analogues is dependent on their application. Canonical fluorescent nucleobases are restricted to the purine or pyrimidine scaffold and require at least two Watson-Crick hydrogen-bonding groups.[148] The advantage of the canonical fluorescent nucleobases is that the pairing potential with natural bases may enable recognition by DNA polymerases and maintain DNA conformation. Non-canonical fluorescent analogues are not restricted to the Watson-Crick hydrogen-bonding, thus they have limited base pairing or recognition by DNA polymerases. However, they can be designed to have increased π-stacking interactions or improved optical properties.[148]

The earliest examples of fluorescent DNA bases were isomorphic heterocycles, for example the adenosine analogue 2-aminopurine (2-AP, Figure 6A).[149] They are very close to the natural nucleoside and minimally perturb the DNA helix. 2-AP is reported to be solvent sensitive. Fluorescence intensity is significantly quenched when paired opposite T or U, due to stacking interactions.[150] It is used as a dynamic probe to study dynamics and structure of DNA[151] and enzyme activity.[142,152] Simple aromatic hydrocarbon chromophores have been developed as nucleobase replacements and incorporated into the DNA π-stack. Hydrocarbon chromophore can replace a nucleoside without disturbing the natural deoxyribose scaffold. Usually these do not maintain Watson-Crick hydrogen bonding, however they increase stability by favoring stacking interactions.[132,153] Pyrene deoxyribose pairs with an abasic site, and in this manner slightly stabilizes the duplex (Figure 6A).[154] Pyrene exhibits both pairing selectivity and stability opposite the abasic site, and suggest that base stacking and geometric fit alone can stabilize duplex DNA.[154] As another example, the furan-conjugated dT (Fu-T, Figure 6A) derivative exhibits strong visible emission in aqueous environments and is quenched in apolar media.[155] The sensitivity for the DNA microenvironment by the Fu-T analogue was employed to detect abasic sites[156] and 8-oxoG.[157]

Extended thymidine and cytosine analogues (^{DMA}T and ^{DMA}C, Figure 6B) have been developed, which have similar stabilization as canonical base pairs and are not quenched in duplex DNA.[158–160] Such analogues have a wide range of applications, such as monitoring conformational changes, site specific metal binding and probe base pairing interactions.

Environmental sensitive fluorophores (ESF) sense changes in the microenvironment (pH or viscosity) and can be applied to study interactions of biomolecules with its environment.[135,161] Such fluorophore are able to detect single nucleotide polymorphisms. For example, 8-aza-3,7-dideaza-2-deoxyadenosine (^{3n7z}A, Figure 6C) is a highly sensitive base-discriminating fluorescent (BDF) analogue and discriminates a perfectly matched T against the unmatched bases.[162] Recently, a 3-deazaguanosine analog (3-(naphthalen-1-ylethynyl)-3-deaza-2′-deoxyguanosine, ^{3n7z}A, Figure 6C), which can discriminate a C in the complementary strand was reported.[163]

Besides modifying the canonical scaffold, nucleobases are converted into fluorophores by covalently attaching a chromophore to the nucleobase. Covalent attachment of an aryl moiety at the C8 position through a C-linkage extends the purines π-conjugated system, resulting in base analogue with fluorescent properties.[164] C-linked aryl-dG adducts are attractive fluorescent probes because the C8-position is not involved in canonical base-pairing interactions and thus the normal DNA conformation is not disturbed.[165] Unfortunately they destabilize the B-DNA duplex, as they favor a *syn* conformation around their glycosidic bond.[166,167] Nevertheless, C8-dG adducts can be exploited as tools to study the DNA conformational change, ie duplex -> G-Quadruplex (GQ) exchange or changes in secondary and

tertiary structure[165]. Examples for excellent C8-dG fluorescent probes are Fu-dG,[168] CNPh-dG[168] and Th-dG[169] (Figure 6D). These probes stabilize the GQ when placed into the thrombin binding aptamer (TBA) and show turn-on emissive properties, upon thrombin binding (exchange from duplex to GQ when thrombin binds).[168,169]

The laboratory of Wilhelmsson has developed numerous base analogue FRET-pairs to investigate the structure and dynamics of nucleic acids and their interactions with molecules. The tricyclic cytosine base analogues, tC° and tC_{nitro} (Figure 6E), are examples of analogues with high quantum yield in DNA duplexes and high stacking interaction within duplex DNA. Exchanging cytosine by tC° or tC_{nitro} does not disturb the DNA conformation and thus are perfect donor/acceptor FRET basepair.[139,170] Recently, they published two quadracyclic adenosine analogues, qAN1 and qA_{nitro}, which have been evaluated as FRET-donor and acceptor base pair (Figure 6E). They don't disturb the DNA conformation, form stable basepairs with T, and have excellent quantum yields. Furthermore, they can be used in combination with tC° as donor and tC_{nitro} as acceptor. An ultimate goal is to develop for each natural base a base analogue FRET-pair.[171]

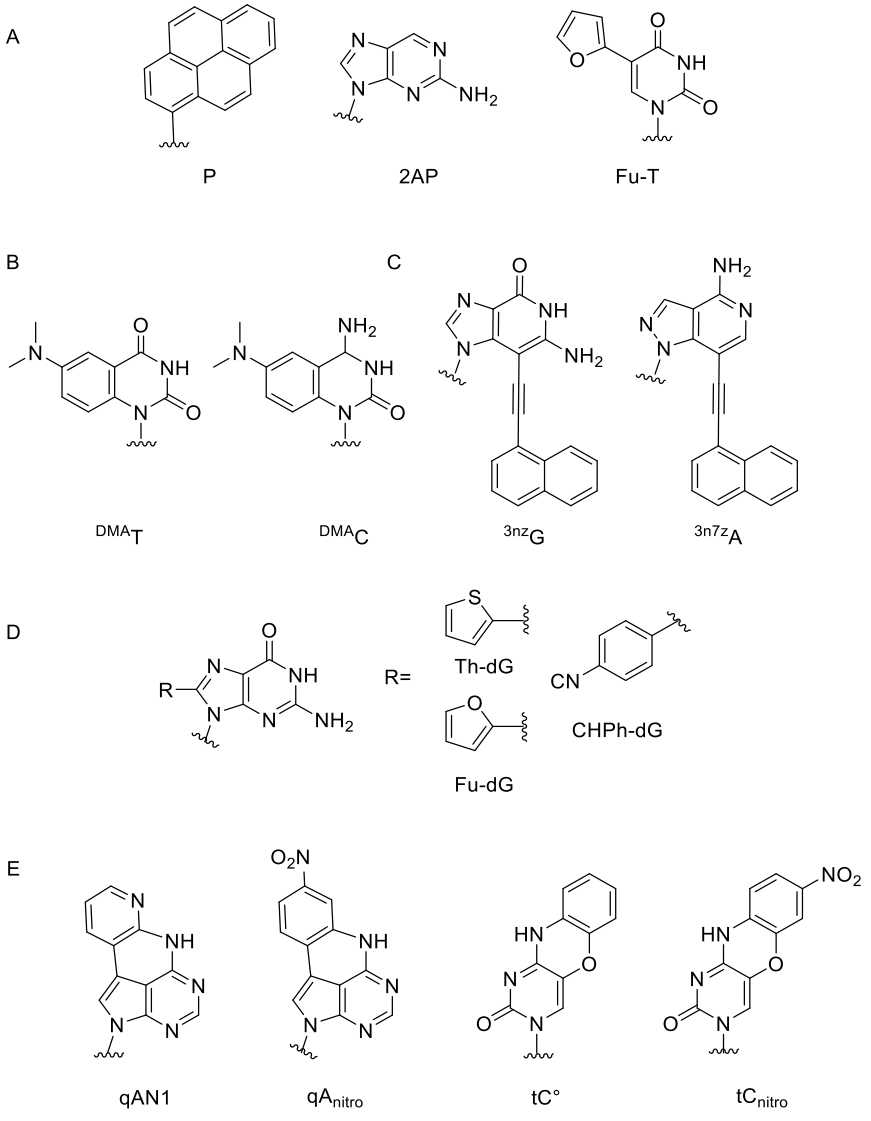


Figure 6: A) Fluorescent base analogues. B) Extended fluorescent base analogues. C) Environmental sensitive fluorescent base analogues. D) C-linked aryl-dG adducts. E) Base analog FRET pairs.

1.3.2 Artificial base pairs

Synthetic base pairs are developed based on a hydrogen bonding pattern that is orthogonal to the natural base pairs, or based on hydrophobic and packing forces. The design involves nucleobase shape mimics[172], hydrophobic pairs[173,174], and pairs with altered hydrogen-bonding motifs.[175,176] It has long been believed, that Watson-Crick hydrogen bonding is required for incorporation and efficient extension of artificial nucleobases by DNA polymerase.

The development of the first artificial base pair, therefore involved interaction via complementary hydrogen bonding pattern. isoG and isoC (Figure 7A), which pair with each other have been successfully replicated by Kf DNA polymerase.[177,178] isoG formed a mispairing with T and thus first PCR amplification had very low fidelity, which was improved by replacing dTTP by a 2-thioT to prevent mispairing. In this manner the PCR amplification fidelity was increased to 98 %.[179]

Artificial base pairs can be developed with the exploitation of further possible hydrogen bonding pattern, because the canonical base pairs do not exploit all possible hydrogen bonding patterns. The hydrogen bonded artificial base pair dZ and dP (Figure 2) was designed based on rearranging hydrogen bond donor and acceptor groups. The base pair was used in the generation of an aptamer against a cancer cell line and successfully amplified with PCR at a yield of 99.8 %.[22,180] Furthermore, the binding affinity of that aptamer was improved using the base pair dZ-dP.[181]

The ds-dy base pair (Figure 7A) was developed based on hydrogen bonding interactions and shape complementary between base pairs[182], addition of the thienyl group helps to stabilize the duplex with stacking interactions. ds-dy was successfully incorporated into RNA and a unnatural codon was recognized for site-specific incorporation of artificial amino acids into proteins.[182] Using the concept of shape-complementarity between base pairs the dDs-dPx base pair (Figure 3) was synthesized. This base pair was incorporated into am aptamer, resulting in an increased binding affinity.[183] Furthermore, the dPx scaffold can be modified with a variety of functional groups replacing R (Figure 3), for site-specific modification of DNA.[184]

Romesberg and coworkers have pursued the idea of using hydrophobic and packing forces to develop an artificial base pair.[185–187] A huge library of possible unnatural base pairs was created, evaluated, and the structure-activity relationship data was elucidated. After designing several generations of artificial base pairs, they determined the physicochemical properties required for efficient synthesis and extension. So far the effort is focused on the family of d5SICS-dMMO2, dNaM-d5SICS and dNaM-dTPT3 (Figure 7B), which have been incorporated into DNA and efficiently amplified by PCR with 99.9 % fidelity.[188–190] In 2014, the first semi synthetic organism (SSO), was created, which imported the unnatural triphosphate and replicated a single dNaM-d5SICS unnatural basepair in a second plasmid.[191] Very recently the development of two new codons containing the unnatural basepair dNaM-dTPT3, was decoded and the artificial amino acids could be incorporated into proteins.[186] Having a third basepair increases the potential to form new codons and therefore a cell could carry more functions than today. In addition, artificial DNA bases might serve as marker nucleotide for damaged DNA, the selective base partner allows for high-fidelity PCR amplification of the marked DNA.[192]

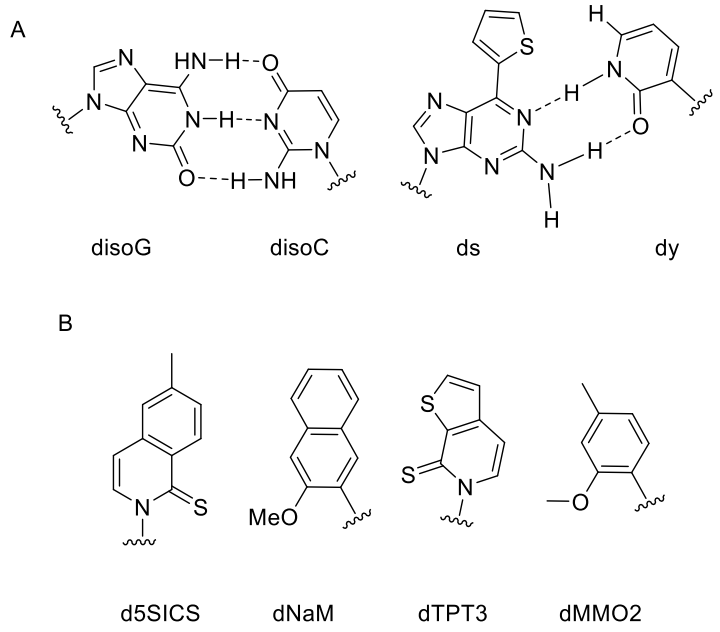


Figure 7: A) First artificial base pair based on hydrogen bonding interactions. B) Artificial base pairs based on hydrophobic and packing forces.

1.3.3 Base analogues as marker nucleotides

Fluorescent and non-fluorescent synthetic nucleotides are developed as marker nucleotides opposite DNA damage (Figure 8). Probes to pair opposite alkylated DNA are developed based on hydrogen-bonding patterns and hydrophobic stacking interactions. The first developed probe was Per, which forms more stable DNA duplexes when paired with O^6-BnG than with G[193], due to hydrophobic interactions via intercalation of Per into the DNA duplex.[194] Further development yielded Peri, Benzi and BIM, which stabilized DNA duplexes containing alkylated DNA.[195–197] Benzi was specifically incorporated opposite O^6-MeG and O^6-CMG versus G using a mutant Taq polymerase. Linear amplification was then applied to determine DNA adducts on a single base resolution.[198,199] Lastly two novel artificial nucleosides, ExBenzi and ExBIM were synthesized and gold nanoparticles-based probes were constructed to detect O^6-MeG in a sequence specific manner.[200]

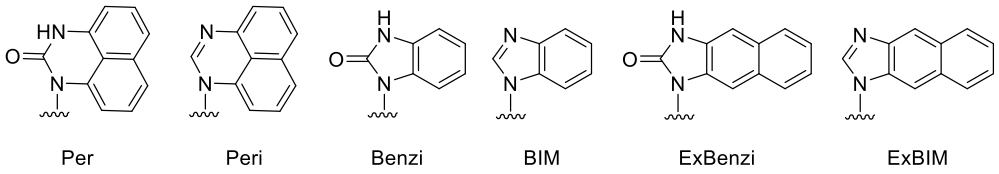


Figure 8: Artificial nucleosides as specific probes opposite O^6-alkylG.

1.4 Overview of thesis work

The goal of the work presented in this thesis was to use chemical tools to understand the impact of DNA adducts on DNA conformation and to understand frameshift mutations. Furthermore, we wanted to detect alkylated DNA using fluorescent modified nucleobases. Artificial nucleobases have a wide application to understand and detect DNA damage and mutations. We studied artificial nucleotides in

the context of conformational impact on the DNA helix and the resulting DNA mutations, and interactions of artificial nucleotides with damaged DNA for early detection of alkylated DNA.

In *Chapter 2,* the C-linked C8-F-biphenyl-dG (FBP-dG) adduct has been incorporated into the 12mer *Nar*I sequence. The fluorescent probe is designed based on the knowledge of donor-acceptor biphenyls that are aromatic chromophore, which exhibit strong fluorescence solvatochromism. To aim of this chapter is to determine the conformational preference of the FBP-dG within the *Nar*I duplex, and furthermore to determine the ability to distinguish the slipped mutagenic duplex from the fully paired duplex. Computational modeling and ^{19}F NMR indicate that FBP-dG shows preference for adopting a *syn* conformation in the fully paired *Nar*I duplex and to intercalate within the helix. Furthermore, the FBP-dG adduct stabilizes the SMI duplex formation and shows turn-on emissive properties, when forming the slipped mutagenic intermediate (SMI) duplex. Therefore the probe can distinguish between the fully paired and the SMI duplex. The results highlight the effect of the adduct linkage type on the expected mutational outcome. FBP-dG is an excellent probe to monitor frameshift mutations *in vitro.*

In *Chapter 3*, fluorescence properties of FBP-dG are exploited to study the mechanism of frameshift mutations within two different sequence contexts: the *Nar*I sequence (5'-GGCGCC-3') and the 3CG repeat (5'-CGCGCG-3'). It has been shown that a higher CG repeat has a higher frequency for frameshift mutations. The templates containing the fluorophore are annealed to primer of increasing lengths (n to n+6) to monitor stepwise the mechanism of frameshift mutations with fluorescence spectroscopy and UV-Vis thermal analysis. The bulge out form is created upon inserting one base opposite FBP-dG (n+1) in the *Nar*I sequence, whereas in the 3CG sequence after inserting two bases opposite FBP-dG (n+2). Interestingly, Kf- bypasses the damage in the 3CG sequence with increased capacity than in the *Nar*I sequence.

In *Chapter 4,* elongated nucleobases ExBenzi and ExBIM have been used to detect O^6-MeG within a mutational hotspot of the human KRAS gene. With dangling end studies and intensive duplex stability studies we show that elongated nucleosides have increased stacking interactions when placed opposite O^6-MeG than G. ExBenzi and ExBIM had higher affinity for O^6-MeG than a series of modified DNA bases (O^6-BnG, N^2-MeG, 8-oxo-G, N^6-MeA). We further measure the fluorescence intensity of the constructed duplexes, ExBIM shows turn-on emissive properties opposite O^6-MeG, but not opposite the modified DNA bases. With molecular dynamic simulations we provide structural insight into the interactions of the base analogues with O^6-MeG. This study provides basic requirements for further probe development to detect DNA lesions.

In *Appendix A,* synthetic base analogues have been tested as marker nucleotides opposite O^2- and O^4-MeT. The synthetic base analogues have a specific base partner, therefore exponential amplification can be performed, to increase the signal for the DNA damage for detection using masspectrometry. We test the incorporation of two artificial triphosphate dTPT3 and d5SICS opposite O^2- and O^4-MeT using standard DNA polymerases (Taq and Vent). Taq incorporates TPT3 opposite both lesions with high efficiency. Taq polymerases incorporates also A opposite O^2-MeT and G opposite O^4-MeT. Exponential amplification with dNaM as a specific base partner for dTPT3 would yield in very low copy number, due to incorporation of natural bases opposite the damage. This chapter is proof of concept for artificial base pairs to be incorporated opposite O^2- and O^4-MeT and sets a basis for future work to achieve exponential amplification using artificial nucleosides as markers opposite alkylate thymidine.

A study including a fluorescent labelled nucleoside analogue as a tool to determine the conformation of thymidine in the thrombin binding aptamer (TBA) is presented in *Appendix B.* In this study, a simple 5-furyl-2'-deoxyuridine (^{Fur}dU) nucleobase has been used to sense a conformational change of TBA with fluorescence spectroscopy. ^{Fur}dU has molecular rotor properties and exhibits dual probing

characteristics, providing changes in emission wavelength and intensity with increased solvent rigidity. In the TBA GQ ^{Fur}dU serves as a diagnostic tool to determine which T bases are solvent exposed and which T interact with the molecular target. Replacement of T-3 and T-12 with ^{Fur}dU increases thrombin binding affinity and provides a 2-fold increase in emission intensity. ^{Fur}dU is a powerful tool for further development of modified aptamers for diagnostic and therapeutic applications.

1.5 References

1. Watson, J. D. & Crick, F. H. C. A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 377–738 (1953).
2. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–563 (1970).
3. Alberts, B., Johnson, A. & Lewis, J. DNA Replication Mechanisms. in *Molecular Biology of the Cell* (Garland Science, 2002).
4. Steitz, T. A. A mechanism for all polymerases. *Nature* **391,** 231–2 (1998).
5. Zhong, X. *et al.* The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins. *Nucleic Acids Res.* **34,** 4731–4742 (2006).
6. McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* **18,** 148–161 (2008).
7. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
8. 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).
9. Gates, K. S. An overview of chemical processes that damage cellular DNA: Spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* **22,** 1747–1760 (2009).
10. Mazouzi, A., Velimezi, G. & Loizou, J. I. DNA replication stress: Causes, resolution and disease. *Exp. Cell Res.* **329,** 85–93 (2014).
11. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18,** 85–98 (2008).
12. Krokan, H. E. & Bjøras, M. Base Excision Repair. *Cold Spring Harb Perspect Biol* **5,** 1–22 (2013).
13. Schärer, O. D. Nucleotide excision repair in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5,** 1–19 (2013).
14. 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,** 141–152 (2012).
15. Lehmann, A. R. *et al.* Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst).* **6,** 891–899 (2007).
16. Inman, R. B., Schildkraut, C. L. & Kornberg, A. Enzymic synthesis of deoxyribonucleic acid. *J. Mol. Biol.* **21,** 197–198 (1965).
17. Grimaldi, K. A., McGurk, C. J., McHugh, P. J. & Hartley, J. A. PCR-based methods for detecting DNA damage and its repair at the sub-gene and single nucleotide levels in cells. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **20,** 181–196 (2002).
18. Chen, L., Liu, P., Evans, T. C. & Ettwiller, L. M. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* **355,** 752–756 (2017).
19. Saiki, R. K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **130,** 1350–1354 (1985).
20. Ramsay, N. *et al.* CyDNA: Synthesis and replication of highly Cy-Dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* **132,** 5096–5104 (2010).
21. Laos, R., Shaw, R., Leal, N. A., Gaucher, E. & Benner, S. Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry* **52,** 5288–5294 (2013).
22. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).
23. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74,** 5463–5467 (1977).
24. Venter, J. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
25. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
26. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch. Dis. Child. Educ. Pract.*

Ed. **98,** 236–238 (2013).
27. Meldrum, C., Doyle, M. a & Tothill, R. W. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. *Clin. Biochem. Rev.* **32,** 177–195 (2011).
28. Wang, G. & Vasquez, K. M. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst).* **19,** 143–151 (2014).
29. Kaushik, M. *et al.* A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Reports* **5,** 388–395 (2016).
30. Bacolla, A., Cooper, D. N. & Vasquez, K. M. DNA structure matters. *Genome Med.* **5,** 51–53 (2013).
31. Brázda, V., Laister, R. C., Jagelská, E. B. & Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **12,** 33–49 (2011).
32. Zain, R. & Sun, J.-S. Do natural DNA triple-helical structures occur and function in vivo? *C. Cell. Mol. life Sci.* **60,** 862–870 (2003).
33. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19,** 414–422 (2009).
34. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13,** 770–780 (2012).
35. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18,** 279–284 (2017).
36. Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA* **3,** 495–507 (2012).
37. Kwok, C. K. & Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **35,** 997–1013 (2017).
38. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12,** 847–854 (2005).
39. Fang, G. & Cech, T. R. Characterization of a G-Quartet Formation Reaction Promoted by the β-Subunit of the Oxytricha Telomere-Binding Protein. *Biochemistry* **32,** 11646–11657 (1993).
40. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Sci.* **249,** 505–510 (1990).
41. Gelinas, A. D., Davies, D. R. & Janjic, N. Embracing proteins: Structural themes in aptamer-protein complexes. *Curr. Opin. Struct. Biol.* **36,** 122–132 (2016).
42. Mayer, G. The chemical biology of aptamers. *Angew. Chemie - Int. Ed.* **48,** 2672–2689 (2009).
43. Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **31,** 453–457 (2013).
44. Sproviero, M. *et al.* Electronic tuning of fluorescent 8-aryl-guanine probes for monitoring DNA duplex–quadruplex exchange. *Chem. Sci.* **5,** 788–796 (2014).
45. Miller, J. A. Carcinogenesis by Chemicals: An Overview---G. H. A. Clowes Memorial Lecture. *Cancer Res.* **30,** 559–576 (1970).
46. Josephy, P. D. & Mennervik, B. *Molecular Toxicology*. (Oxford University Press, 2006).
47. Liu, G., Niu, Z., Niekerk, D. V., Xue, J. & Zheng, L. *Polycyclic Aromatic Hydrocarbons (PAHs) from coal combustion : emissions, analysis, and toxicology*. *Reviews of enviornmental contamination and toxicology* **192,** (2008).
48. Samanta, S. K., Singh, O. V. & Jain, R. K. Polycyclic aromatic hydrocarbons: Environmental pollution and bioremediation. *Trends Biotechnol.* **20,** 243–248 (2002).
49. Chen, L. *et al.* Expanded analysis of Benzo[a]pyrene-DNA adducts formed in vitro and in mouse skin: Their significance in tumor initiation. *Chem. Res. Toxicol.* **9,** 897–903 (1996).
50. Joosen, A. M. C. P. *et al.* Effect of processed and red meat on endogenous nitrosation and DNA damage. *Carcinogenesis* **30,** 1402–1407 (2009).
51. Lewin, M. H. *et al.* Red Meat Enhances the Colonic Formation of the DNA Adduct O^6-Carboxymethyl Guanine: Implications for Colorectal Cancer Risk. *Cancer Res.* **66,** 1859–1865 (2006).
52. Bouvard, V. *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol.* **16,** 1599–1600 (2015).

53. Cross, A. J. & Sinha, R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ. Mol. Mutagen.* **44,** 44–55 (2004).
54. Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **424,** 127–142 (1999).
55. Hecht, S. S. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* **3,** 733–744 (2003).
56. Hecht, S. S. Biochemistry, biology, and carcinogenicity of tobacco-specific N- nitrosamines. *Chem. Res. Toxicol.* **11,** 559–603 (1998).
57. Peterson, L. A. Formation, repair, and genotoxic properties of bulky DNA adducts formed from tobacco-specific nitrosamines. *J. Nucleic Acids* **2010,** 1–11 (2010).
58. Beland, F. A. *et al. Chemical Carcinogenesis and Mutagenesis I*. (1990).
59. Beland, F. a & Poirier, M. C. DNA Adducts and Their Consequences. *Methods to Assess DNA Damage Repair lnterspecies Comp.* 29–55 (1994).
60. Anna, L., Kovács, K., Gyorffy, E., Schoket, B. & Nair, J. Smoking-related O4-ethylthymidine formation in human lung tissue and comparisons with bulky DNA adducts. *Mutagenesis* **26,** 523–527 (2011).
61. Enoch, S. J., Ellison, C. M., Schultz, T. W. & Cronin, M. T. D. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit. Rev. Toxicol.* **41,** 783–802 (2011).
62. Ralhan, R. & Kaur, J. Alkylating agents and cancer therapy. *Expert Opin. Ther. Pat.* **17,** 1061–1075 (2007).
63. Boysen, G., Pachkowski, B. F., Nakamura, J. & Swenberg, J. A. The formation and biological significance of N7-guanine adducts. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **678,** 76–94 (2009).
64. Philippin, G., Cadet, J., Gasparutto, D., Mazon, G. & Fuchs, R. P. Ethylene oxide and propylene oxide derived N7-alkylguanine adducts are bypassed accurately in vivo. *DNA Repair (Amst).* **22,** 133–136 (2014).
65. Johnson, R. E., Yu, S.-L., Prakash, S. & Prakash, L. A Role for Yeast and Human Translesion Synthesis DNA Polymerases in Promoting Replication through 3-Methyl Adenine. *Mol. Cell. Biol.* **27,** 7198–7205 (2007).
66. Engelward, B. P. *et al.* A chemical and genetic approach together define the biological consequences of 3-methyladenine lesions in the mammalian genome. *J. Biol. Chem.* **273,** 5412–5418 (1998).
67. Arrington, A. K. *et al.* Prognostic and predictive roles of KRAS mutation in colorectal cancer. *Int. J. Mol. Sci.* **13,** 12153–12168 (2012).
68. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21,** 7435–7451 (2002).
69. Swenberg, J. A. *et al.* O4-ethyldeoxythymidine, but not O6-ethyldeoxyguanosine, accumulates in hepatocyte DNA of rats exposed continuously to diethylnitrosamine. *Proc Natl Acad Sci U S A* **81,** 1692–1695 (1984).
70. Engelse, L. Den, De Graaf, A., De Brij, R. J. & Menkveld, G. J. O- and O4-ethylthymine and the ethylphosphotriester dTp(Et)dT are highly persistent DNA modifications in slowly dividing tissues of the ethylnitrosourea-treated rat. *Carcinogenesis* **8,** 751–757 (1987).
71. Luch, A. Nature and nurture - lessons from chemical carcinogenesis. *Nat Rev Cancer* **5,** 113–125 (2005).
72. Cho, B. P. Dynamic Conformational Heterogeneities of Carcinogen-DNA Adducts and their Mutagenic Relevance. *J. Environ. Sci. Heal. Part C* **22,** 57–90 (2004).
73. Wilson, R. H., DeEds, F. & Cox, A. J. The Toxicity and Carcinogenic Activity of 2-Acetaminofluorene. *Cancer Res.* **1,** 595–608 (1941).
74. Kriek, E., Miller, J. A., Juhl, U. & Miller, E. C. 8-(N-2-Fluorenylacetamido) guanosine, an Arylamidation Reaction Product of Guanosine and the Carcinogen N-Acetoxy-N-2-fluorenylacetamide in Neutral Solution *. *Biochemistry* **6,** 177–182 (1967).
75. Fuchs, R. P. P., Schwartz, N. & Daune, M. P. Hot Spots of Frameshift Mutations Induced by the Ultimate Carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature* **294,** 657–659 (1981).
76. Fuchs, R. P. & Fujii, S. Translesion synthesis in Escherichia coli: Lessons from the NarI mutation hot spot. *DNA Repair (Amst).* **6,** 1032–1041 (2007).

77. Veaute, X. & Fuchs, R. P. P. Polymorphism in N -2-acetylaminofluorene induced DNA structure as revealed by DNase I footprinting. *Nucleic Acids Res.* **19,** 5603–5606 (1991).
78. Koffel-Schwartz, N. & Fuchs, R. P. Sequence determinants for -2 frameshift mutagenesis at NarI-derived hot spots. *J. Mol. Biol.* **252,** 507–513 (1995).
79. Koffel-Schwartz, N. *et al.* Carcinogen-induced mutation spectrum in wild-type, uvrA and umuC strains of Escherichia coli. Strain specificity and mutation-prone sequences. *J. Mol. Biol.* **177,** 33–51 (1984).
80. Zhou, L., Rajabzadeh, M., Traficante, D. D. & Cho, B. P. Conformational Heterogeneity of Arylamine-Modified DNA: 19F NMR Evidence. *J. Am. Chem. Soc.* **119,** 5384–5389 (1997).
81. Cho, B. P., Beland, F. A. & Marques, M. M. NMR Structural Studies of a 15-mer DNA Sequence from a ras Protooncogene, Modified at the First Base of Codon 61 with the carcinogen 4-Aminobiphenyl. *Biochemistry* **31,** 9587–9602 (1992).
82. Meneni, S., Shell, S. M., Zou, Y. & Cho, B. P. Conformation-specific recognition of carcinogen-DNA adduct in Escherichia coli nucleotide excision repair. *Chem. Res. Toxicol.* **20,** 6–10 (2007).
83. Meneni, S., Liang, F. & Cho, B. P. Examination of the Long-range Effects of Aminofluorene-induced Conformational Heterogeneity and Its Relevance to the Mechanism of Translesional DNA Synthesis. *J. Mol. Biol.* **366,** 1387–1400 (2007).
84. Vaidyanathan, V. G. & Cho, B. P. Sequence effects on translesion synthesis of an aminofluorene-DNA adduct: Conformational, thermodynamic, and primer extension kinetic studies. *Biochemistry* **51,** 1983–1995 (2012).
85. Liang, F. & Cho, B. P. Conformational and thermodynamic impact of bulky aminofluorene adduction on simulated translesion DNA synthesis. *Chem. Res. Toxicol.* **24,** 597–605 (2011).
86. Hsu, G. W. *et al.* Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. *J Biol Chem* **279,** 50280–50285 (2004).
87. Fuchs, R. & Daune, M. Physical studies on deoxyribonucleic acid after covalent binding of a carcinogen. *Biochemistry* **11,** 2659–2666 (1972).
88. O'Handley, S. F. *et al.* Structural Characterization of an N-Acetyl-2-aminofluorene (AAF) Modified DNA Oligomer by NMR, Energy Minimization, and Molecular Dynamics. *Biochemistry* **32,** 2481–2497 (1993).
89. Fuchs, R. & Daune, M. Physical basis of chemical carcniogenesis by N-2-fluorenylacetamide derivatives and analogs. *FEBS Lett.* **34,** 295–298 (1973).
90. Organization, W. H. *4-Aminobiphenyl*. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* **99,** (2010).
91. Cho, B. P., Beland, F. A. & Marques, M. M. NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. *Biochemistry* **33,** 1373–1384 (1994).
92. Jain, V., Vaidyanathan, V. G., Patnaik, S., Gopal, S. & Cho, B. P. Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19 F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **53,** 4059–4071 (2014).
93. Turesky, R. J. Heterocyclic aromatic amines: Potential human carcinogens. *Adv. Mol. Toxicol.* **4,** 37–83 (2010).
94. Turesky, R. J. Heterocyclic Aromatic Amine Metabolism, DNA Adduct Formation, and Carcinogenesis. *Drug Metab. Rev.* **34,** 625–650 (2002).
95. Turesky, R. J. & Vouros, P. Formation and analysis of heterocyclic aromatic amine-DNA adducts in vitro and in vivo. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **802,** 155–166 (2004).
96. Sugimura, T. Overview of carcinogenic heterocyclic amines. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **376,** 211–219 (1997).
97. Wakabayashi, K., Nagao, M., Esumi, H. & Sugimura, T. Food-derived mutagens and carcinogens. *Cancer Res.* **52,** 2092s–2098s (1992).
98. Wang, F. *et al.* Base-Displaced Intercalated Structure of the Food Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the Recognition Sequence of the NarI Restriction Enzyme, a Hotspot for −2 bp Deletions. *J. Am. Chem. Soc.* **128,** 10085–10095 (2006).
99. Wang, F., Elmquist, C. E., Stover, J. S., Rizzo, C. J. & Stone, M. P. DNA sequence modulates the conformation of the food mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in the recognition sequence of the NarI restriction enzyme. *Biochemistry* **46,** 8498–8516 (2007).

100. Elmquist, C. E., Wang, F., Stover, J. S., Stone, M. P. & Rizzo, C. J. Conformational differences of the C8-deoxyguanosine adduct of 2-amino-3-methylimidazo[4,5-f]quinoline (IQ) within the NarI recognition sequence. *Chem. Res. Toxicol.* **20,** 445–454 (2007).
101. Stover, J. S., Chowdhury, G., Zang, H., Guengerich, F. P. & Rizzo, C. J. Translesion Synthesis Past the C8- and N2-Deoxyguanosine Adducts of the Dietary Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the NarI Recognition Sequence by Prokaryotic DNA Polymerases. *Chem. Res. Toxicol.* **19,** 1506–1517 (2006).
102. Layton, D. W. *et al.* Cancer risk of hetercyclic amines in cooked foods: an analysis and implications for research. *Carcinogenesis* **16,** 39–52 (1995).
103. Brown, K. *et al.* Solution structure of the 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine adduct in duplex DNA. *PNAS* **98,** 8507–8512 (2001).
104. Fukuda, H. *et al.* Translesional DNA synthesis through a C8-guanyl adduct of 2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP) in vitro: REV1 inserts dC opposite the lesion, and DNA polymerase κ potentially catalyzes extension reaction from the 3′-dC terminus. *J. Biol. Chem.* **284,** 25585–25597 (2009).
105. Dai, J., Wright, M. W. & Manderville, R. A. Ochratoxin A Forms a Carbon-Bonded C8-Deoxyguanosine Nucleoside Adduct: Implications for C8 Reactivity by a Phenolic Radical. *J. Am. Chem. Soc.* **125,** 3716–3717 (2003).
106. Sproviero, M. *et al.* Structural and biochemical impact of C8-aryl-guanine adducts within the NarI recognition DNA sequence: influence of aryl ring size on targeted and semi-targeted mutagenicity. *Nucleic Acids Res.* **42,** 13405–13421 (2014).
107. Sharma, P., Manderville, R. A. & Wetmore, S. D. Structural and energetic characterization of the major DNA adduct formed from the food mutagen ochratoxin A in the NarI hotspot sequence: Influence of adduct ionization on the conformational preferences and implications for the NER propensity. *Nucleic Acids Res.* **42,** 11831–11845 (2014).
108. Witham, A. A. *et al.* Chlorine Functionalization of a Model Phenolic C8-Guanine Adduct Increases Conformational Rigidity and Blocks Extension by a Y-Family DNA Polymerase. *Chem. Res. Toxicol.* **28,** 1346–1356 (2015).
109. Sproviero, M. *et al.* Enhancing Bulge Stabilization through Linear Extension of C8-Aryl-Guanine Adducts to Promote Polymerase Blockage or Strand Realignment to Produce a C:C Mismatch. *Chem. Res. Toxicol.* **28,** 1647–1658 (2015).
110. Griffiths, A., Gelbart, W. & Miller, J. Modern Genetic Analysis. in *The Molecular Basis of Mutation* (ed. Freeman, W. H.) (1999).
111. Torgerson, T. & Ochs, H. *Genetics of Primary Immune Deficiencies*. *Stiehm's Immune Deficiencies* (2014).
112. Iliakis, G. The Role of DNA Double Strand Breaks in lonizing Radiation-Induced Killing of Eukaryotic Cells. *BioEssays* **13,** 641–648 (1991).
113. Rothkamm, K., Krüger, I., Thompson, L. H., Kru, I. & Lo, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol. Cell. Biol.* **23,** 5706–5715 (2003).
114. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300,** 149–152 (1982).
115. Hainaut, P. *et al.* Database of p53 gene somatic mutations in human tumors and cell lines: Updated compilation and future prospects. *Nucleic Acids Res.* **25,** 151–157 (1997).
116. Pfeifer, G. P. & Denissenko, M. F. Formation and repair of DNA lesions in the p53 gene: Relation to cancer mutations? *Environ. Mol. Mutagen.* **31,** 197–205 (1998).
117. Calvez, F. Le *et al.* TP53 and KRAS Mutation Load and Types in Lung Cancers in Relation to Tobacco Smoke : Distinct Patterns in Never , Former , and Current Smokers TP53 and KRAS Mutation Load and Types in Lung Cancers in Relation to Tobacco Smoke : Distinct Patterns in Never. *Cancer Res.* **65,** 5076–5083 (2005).
118. Cancer Genome Atlas Research Network, T. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
119. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
120. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer

genome. *Nature* **463,** 191–196 (2010).
121. Hoffmann, G. R. & Fuchs, R. P. P. Mechanisms of Frameshift Mutations: Insight from Aromatic Amines. *Chem. Res. Toxicol.* **10,** 347–359 (1997).
122. Boland, C. R. & Goel, A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* **138,** 2073–2087 (2010).
123. Wu, B. P., Zhang, Y. L., Zhou, D. Y., Gao, C. F. & Lai, Z. S. Microsatellite instability, MMR gene expression and proliferation kinetics in colorectal cancer with famillial predisposition. *World J. Gastroenterol.* **6,** 902–905 (2000).
124. Yurgelun, M. B. *et al.* Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps. *Cancer Prev. Res.* **5,** 574–582 (2012).
125. Hoffmann, G. R., Yin, C. C., Terry, C. E., Ferguson, L. R. & Denny, W. A. Frameshift mutations induced by four isomeric nitroacridines and their des-nitro counterpart in the lacZ reversion assay in Escherichia coli. *Environ. Mol. Mutagen.* **47,** 82–94 (2006).
126. Ferguson, L. R. & Denny, W. A. The genetic toxicology of acridine. *Mutat. Res.* **258,** 123–160 (1991).
127. Burnouf, D., Koehl, P. & Fuchs, R. P. Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci.* **86,** 4147–4151 (1989).
128. Garcia, A., Lambert, I. B. & Fuchst, R. P. P. DNA adduct-induced stabilization of slipped frameshift intermediates within repetitive sequences: Implications for mutagenesis. *Proc. Natl. Acad. Sci.* **90,** 5989–5993 (1993).
129. Lakowicz, J. R. *Principles of fluorescence spectroscopy, 3rd Edition, Joseph R. Lakowicz, editor*. *Principles of fluorescence spectroscopy, Springer, New York, USA, 3rd edn, 2006.* (2006).
130. Wilhelmsson, L. M. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.* **43,** 159–183 (2010).
131. Saito, Y. *et al.* An environmentally sensitive fluorescent purine nucleoside that changes emission wavelength upon hybridization. *Chem. Commun.* **49,** 5684–5686 (2013).
132. Teo, Y. N. & Kool, E. T. DNA-multichromophore systems. *Chem. Rev.* **112,** 4221–4245 (2012).
133. Cekan, P. & Sigurdsson, S. T. Single base interrogation by a fluorescent nucleotide: each of the four DNA bases identified by fluorescence spectroscopy. *Chem. Commun.* 3393–3395 (2008).
134. Mizuta, M. *et al.* A pyrimidopyrimidoindole nucleoside (dC PPI): photophysical properties and thermal stability of the modified DNA duplexes. *Nucleosides. Nucleotides Nucleic Acids* **26,** 1335–1338 (2007).
135. Okamoto, A., Tainaka, K. & Saito, I. Clear distinction of purine bases on the complementary strand by a fluorescence change of a novel fluorescent nucleoside. *J. Am. Chem. Soc.* **125,** 4972–4973 (2003).
136. Saito, Y., Miyauchi, Y., Okamoto, A. & Saito, I. Synthesis and properties of novel base-discriminating fluorescent (BDF) nucleosides: A highly polarity-sensitive fluorophore for SNP typing. *Tetrahedron Lett.* **45,** 7827–7831 (2004).
137. Choi, J. Y. *et al.* Biochemical basis of genotoxicity of heterocyclic arylamine food mutagens: Human DNA polymerase η selectively produces a two-base deletion in copying the N2-guanyl adduct of 2-amino-3-methylimidazo[4,5-f]quinoline but not the C8 adduct at the NarI G3 site. *J. Biol. Chem.* **281,** 25297–25306 (2006).
138. Hawkins, M. E. Fluorescent pteridine nucleoside analogs: a window on DNA interactions. *Cell Biochem. Biophys.* **34,** 257–281 (2001).
139. Börjesson, K. *et al.* Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems. *J. Am. Chem. Soc.* **131,** 4288–4293 (2009).
140. Godde, F., Toulmé, J. J. & Moreau, S. 4-amino-1H-benzo[g]quinazoline-2-one: a fluorescent analog of cytosine to probe protonation sites in triplex forming oligonucleotides. *Nucleic Acids Res.* **28,** 2977–2985 (2000).
141. Kirk, S. R., Luedtke, N. W. & Tor, Y. 2-Aminopurine as a real-time probe of enzymatic cleavage and inhibition of hammerhead ribozymes. *Bioorganic Med. Chem.* **9,** 2295–2301 (2001).
142. Raney, K. D., Sowers, L. C., Millar, D. P. & Benkovic, S. J. A Fluorescence-Based Assay for Monitoring Helicase Activity. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6644–6648 (1994).

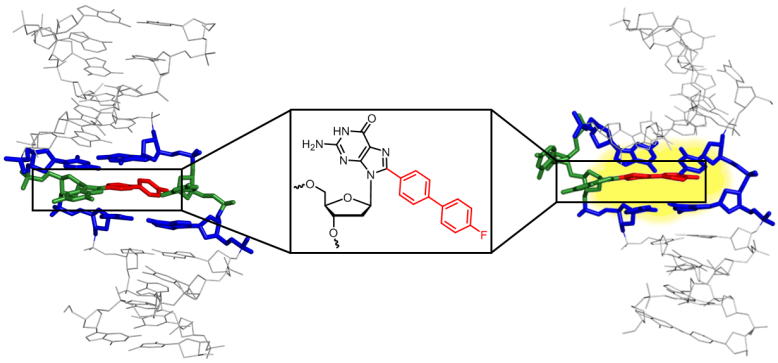
143. Beharry, A. A., Lacoste, S., O'Connor, T. R. & Kool, E. T. Fluorescence Monitoring of the Oxidative Repair of DNA Alkylation Damage by ALKBH3, a Prostate Cancer Marker. *J. Am. Chem. Soc.* **138,** 3647–3650 (2016).
144. LePecq, J. B. & Paoletti, C. A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization. *J. Mol. Biol.* **27,** 87–106 (1967).
145. Tuma, R. S. *et al.* Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal. Biochem.* **268,** 278–288 (1999).
146. Mayer, E. *et al.* 1-Ethynylpyrene as a tunable and versatile molecular beacon for DNA. *ChemBioChem* **5,** 865–868 (2004).
147. Varghese, R. & Wagenknecht, H. A. White-light-emitting DNA (WED). *Chem. - A Eur. J.* **15,** 9307–9310 (2009).
148. Xu, W., Chan, K. M. & Kool, E. T. Fluorescent nucleobases as tools for studying DNA and RNA. *Nat. Chem.* **9,** 1043–1055 (2017).
149. Ward, D. C. & Reich, E. Fluorescence Studies of Nucleotides and Polynucleotide. *J. Biochem. Chem.* **244,** 1228–1237 (1969).
150. Law, S. M., Eritja, R., Goodman, M. F. & Breslauer, K. J. Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine. *Biochemistry* **35,** 12329–12337 (1996).
151. Stivers, J. T. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: Metal-ion and base sequence effects. *Nucleic Acids Res.* **26,** 3837–3844 (1998).
152. Bloom, L. B. *et al.* Pre-Steady-State Kinetic Analysis of Sequence-Dependent Nucleotide Excision by the 3′-Exonuclease Activity of Bacteriophage T4 DNA Polymerase. *Biochemistry* **33,** 7576–7586 (1994).
153. Ren, R. X. F., Chaudhuri, N. C., Paris, P. L., Rumney IV, S. & Kool, E. T. Naphthalene, phenanthrene, and pyrene as DNA base analogues: Synthesis, structure, and fluorescence in DNA. *J. Am. Chem. Soc.* **118,** 7671–7678 (1996).
154. Matray, T. J. & Kool, E. T. Selective and stable DNA base pairing without hydrogen bonds. *J. Am. Chem. Soc.* **120,** 6191–6192 (1998).
155. Greco, N. J. & Tor, Y. Furan decorated nucleoside analogues as fluorescent probes: synthesis, photophysical evaluation, and site-specific incorporation. *Tetrahedron* **63,** 3515–3527 (2007).
156. Greco, N. J. & Tor, Y. Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. *J. Am. Chem. Soc.* **127,** 10784–10785 (2005).
157. Greco, N. J., Sinkeldam, R. W. & Tor, Y. An emissive C analog distinguishes between G, 8-oxoG, and T. *Org. Lett.* **11,** 1115–1118 (2009).
158. Mata, G., Schmidt, O. P. & Luedtke, N. W. A fluorescent surrogate of thymidine in duplex DNA. *Chem. Commun.* **52,** 4718–4721 (2016).
159. Schmidt, O. P., Mata, G. & Luedtke, N. W. Fluorescent Base Analogue Reveals T-HgII-T Base Pairs Have High Kinetic Stabilities That Perturb DNA Metabolism. *J. Am. Chem. Soc.* **138,** 14733–14739 (2016).
160. Mata, G. & Luedtke, N. W. Fluorescent probe for proton-coupled DNA folding revealing slow exchange of i -motif and duplex structures. *J. Am. Chem. Soc.* **137,** 699–707 (2015).
161. Okamoto, A., Tanaka, K., Fukuta, T. & Saito, I. Design of base-discriminating fluorescent nucleoside and its application to T/C SNP typing. *J. Am. Chem. Soc.* **125,** 9296–9297 (2003).
162. Suzuki, A., Saito, M., Katoh, R. & Saito, Y. Synthesis of 8-aza-3,7-dideaza-2′-deoxyadenosines possessing a new adenosine skeleton as an environmentally sensitive fluorescent nucleoside for monitoring the DNA minor groove. *Org. Biomol. Chem.* **13,** 7459–7468 (2015).
163. Suzuki, A., Yanagi, M., Takeda, T., Hudson, R. H. E. & Saito, Y. The fluorescently responsive 3-(naphthalen-1-ylethynyl)-3-deaza-2′-deoxyguanosine discriminates cytidine via the DNA minor groove. *Org. Biomol. Chem.* **15,** 7853–7859 (2017).
164. Nakatani, K. & Tor, Y. *Modified Nucleic Acids*. (2016).
165. Dumas, A. & Luedtke, N. W. Cation-mediated energy transfer in G-quadruplexes revealed by an internal fluorescent probe. *J. Am. Chem. Soc.* **132,** 18004–18007 (2010).
166. Manderville, R. A. *et al.* Fluorescent C-Linked C8-Aryl-guanine Probe for Distinguishing syn from anti Structures in Duplex DNA. *Chem. Res. Toxicol.* **25,** 1271–1282 (2012).
167. Fazakerley, G. V., Russel, J. C. & Wolfe, M. A. Determination of the syn- anti Equilibrium of Some Purine 3′:5′- Nucleotides by Nuclear- Magnetic- Relaxation Perturbation in the Presence

of a Lanthanide- Ion Probe. *Eur. J. Biochem.* **76,** 601–605 (1977).
168. Sproviero, M. & Manderville, R. A. Harnessing G-tetrad scaffolds within G-quadruplex forming aptamers for fluorescence detection strategies. *Chem. Commun.* **50,** 3097–3099 (2014).
169. Fadock, K. L., Manderville, R. A., Sharma, P. & Wetmore, S. D. Optimization of fluorescent 8-heteroaryl-guanine probes for monitoring protein-mediated duplex → G-quadruplex exchange. *Org. Biomol. Chem.* **14,** 4409–4419 (2016).
170. Engman, K. C. *et al.* DNA adopts normal B-form upon incorporation of highly flourescent DNA base analogue tC: NMR structure and UV-Vis spectroscopy characterization. *Nucleic Acids Res.* **32,** 5087–5095 (2004).
171. Füchtbauer, A. F. *et al.* Fluorescent RNA cytosine analogue-an internal probe for detailed structure and dynamics investigations. *Sci. Rep.* **7,** 1–8 (2017).
172. Kool, E. T., Morales, J. C. & Guckian, K. M. Mimicking the structure and function of DNA: Insights into DNA stability and replication. *Angew. Chemie - Int. Ed.* **39,** 990–1009 (2000).
173. Hirao, I. Unnatural base pair systems for DNA/RNA-based biotechnology. *Curr. Opin. Chem. Biol.* **10,** 622–627 (2006).
174. Henry, A. A. & Romesberg, F. E. Beyond A, C, G and T: Augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* **7,** 727–733 (2003).
175. Benner, S. A. Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res.* **37,** 784–797 (2004).
176. Geyer, C. R., Battersby, T. R. & Benner, S. A. Nucleobase Pairing in Expanded Watson-Crick-like Genetic Information Systems. *Structure* **11,** 1485–1498 (2003).
177. Switzer, C., Moronev, S. E. & Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and RNA. *J. Am. Chem. Soc.* **111,** 8322–8323 (1989).
178. Piccirilli, J. A., Krauch, T., Moronev, S. E. & Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and RNA extends the genetic alphabet. *Nature* **343,** 33–37 (1990).
179. Sismour, A. M. & Benner, S. A. The use of thymidine analogs to improve the replication of an extra DNA base pair: A synthetic biological system. *Nucleic Acids Res.* **33,** 5640–5646 (2005).
180. Yang, Z., Sismour, A. M., Sheng, P., Puskar, N. L. & Benner, S. A. Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.* **35,** 4238–4249 (2007).
181. Sefah, K. *et al.* In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci.* **111,** 1449–1454 (2014).
182. Hirao, I. *et al.* An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* **20,** 177–182 (2002).
183. Matsunaga, K. I. *et al.* Architecture of high-affinity unnatural-base DNA aptamers toward pharmaceutical applications. *Sci. Rep.* **5,** 1–7 (2015).
184. Okamoto, I., Miyatake, Y., Kimoto, M. & Hirao, I. High Fidelity, Efficiency and Functionalization of Ds-Px Unnatural Base Pairs in PCR Amplification for a Genetic Alphabet Expansion System. *ACS Synth. Biol.* **5,** 1220–1230 (2016).
185. Feldman, A. W. & Romesberg, F. E. In Vivo Structure–Activity Relationships and Optimization of an Unnatural Base Pair for Replication in a Semi-Synthetic Organism. *J. Am. Chem. Soc.* **139,** 11427–11433 (2017).
186. Zhang, Y. *et al.* A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551,** 644–647 (2017).
187. Zhang, Y. *et al.* A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc. Natl. Acad. Sci.* **114,** 1–6 (2017).
188. Malyshev, D. a. *et al.* Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci.* **109,** 12005–12010 (2012).
189. Leconte, A. M. *et al.* Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* **130,** 2336–2343 (2008).
190. Lavergne, T., Malyshev, D. A. & Romesberg, F. E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chem. - A Eur. J.* **18,** 1231–1239 (2012).
191. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509,** 385–8 (2014).
192. Riedl, J., Ding, Y., Fleming, A. M. & Burrows, C. J. Identification of DNA lesions using a third

base pair for amplification and nanopore sequencing. *Nat. Commun.* **6,** 1–11 (2015).
193. Gong, J. & Sturla, S. J. A synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA. *J. Am. Chem. Soc.* **129,** 4882–4883 (2007).
194. 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,** 7566–7576 (2013).
195. Gahlon, H. L. & Sturla, S. J. Hydrogen bonding or stacking interactions in differentiating duplex stability in oligonucleotides containing synthetic nucleoside probes for alkylated DNA. *Chem. - A Eur. J.* **19,** 11062–11067 (2013).
196. Angelov, T., Dahlmann, H. A. & Sturla, S. J. Oligonucleotide probes containing pyrimidine analogs reveal diminished hydrogen bonding capacity of the DNA adduct O6-methyl-G in DNA duplexes. *Bioorganic Med. Chem.* **21,** 6212–6216 (2013).
197. Gahlon, H. L., Schweizer, W. B. & Sturla, S. J. Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc.* **135,** 6384–6387 (2013).
198. 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,** 30–33 (2015).
199. 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 O 6 -alkylguanine DNA adducts. *Nucleic Acids Res.* **44,** 6564–6573 (2016).
200. Trantakis, I. A., Nilforoushan, A., Dahlmann, H. A., Stäuble, C. K. & Sturla, S. J. In-Gene Quantification of O6-Methylguanine with Elongated Nucleoside Analogues on Gold Nanoprobes. *J. Am. Chem. Soc.* **138,** 8497–8504 (2016).

Chapter 2

Conformational preference and fluorescence response of a C-linked C8-biphenyl-guanine lesion in the *Nar*I mutational hotspot: Evidence for enhanced *syn* adduct formation



Reprinted with permission from

Florence D. Berger, Shana J. Sturla, Ryan W. Kung, Tony Montina, Stacey D. Wetmore, and Richard A. Manderville, Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the NarI Mutational Hotspot: Evidence for Enhanced Syn Adduct Formation. Chem. Res. Toxicol., 2018, 31 (1), pp 37-47

F.D.B. designed the study, synthesized compounds and oligonucleotides, characterized them and performed the thermal melting, fluorescence and CD studies. R.W.K. and T.M. performed the molecular dynamic simulations. R.A.M., S.D.W. and S.J.S. conceived the research. F.D.B., R.A.M., S.D.W. and S.J.S. interpreted data and wrote the manuscript.

Chemical Research in Toxicology

Article

Cite This: Chem. Res. Toxicol. 2018, 31, 37−47

pubs.acs.org/crt

Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the *Nar*I Mutational Hotspot: Evidence for Enhanced *Syn* Adduct Formation

Florence D. Berger,[†] Shana J. Sturla,*[,†] Ryan W. Kung,[‡] Tony Montina,[‡] Stacey D. Wetmore,*[,‡] and Richard A. Manderville*[,§]

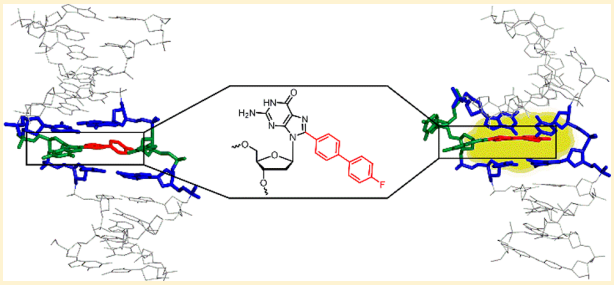
[†]Department of Health Sciences and Technology, Institute of Food, Nutrition, and Health, ETH Zürich, 8092 Zürich, Switzerland

[‡]Department of Chemistry and Biochemistry, and the Canadian Centre for Research in Advanced Fluorine Technologies, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

[§]Departments of Chemistry and Toxicology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Supporting Information

ABSTRACT: Aromatic chemical carcinogens can undergo enzymatic transformations to produce a range of electrophilic species that attach covalently to the C8-site of 2′-deoxyguanosine (dG) to afford C8-dG adducts. The most studied C8-dG adducts are formed from arylamines and contain a N-linkage separating the dG from the C8-aryl moiety. Other carcinogenic species result in direct aryl ring attachment to the dG moiety, resulting in C-linked adducts. The resulting C-linked adducts have reduced conformational flexibility compared to the corresponding N-linked C8-dG adducts, which can alter their orientation in the DNA duplex. Described herein are structural studies of a fluorescent C-linked 4-fluorobiphenyl-dG (FBP-dG) that has been incorporated into the reiterated G_3-postion of the 12-mer *Nar*I sequence and those containing other 5′-flanking nucleobases. FBP-dG displays a strong preference for adopting a *syn* conformation in the fully paired *Nar*I duplex to produce an intercalated structure that exhibits stacking interactions between the C-linked biphenyl and the flanking bases. FBP-dG is also shown to significantly stabilize the slippage mutagenic intermediate (SMI) duplex containing the lesion and 5′-flanking base within a 2-base bulge. FBP-dG exhibits fluorescence sensitivity to SMI duplex formation that can readily distinguish it from the fully paired duplex. Molecular dynamics simulations and optical spectroscopy for the *Nar*I oligonucleotides containing the C-linked FBP-dG predict increased rigidity of the biphenyl in the *syn* conformation. The greater propensity to generate the promutagenic *syn* conformation for the C-linked FBP-dG adduct compared to the N-linked 4-aminobiphenyl-dG adduct (ABP-dG) suggests greater mutagenicity for the C-linked analogue. These results highlight the effect of the adduct linkage type on the conformational properties of adducted DNA. The turn-on emission response of FBP-dG in the SMI duplex may be a powerful tool for monitoring SMI formation in the *Nar*I sequence upon synthesis with DNA polymerases.



INTRODUCTION

Endogenous and exogenous electrophiles can damage DNA to produce bulky covalent DNA adducts (addition products). If adducts evade DNA repair, they may initiate mutagenesis and ultimately carcinogenesis.[1,2] Bulky adducts at the C8-site of 2′-deoxyguanosine (C8-dG) are commonly produced by chemical carcinogens.[3] Examples include aromatic amines, such as 2-aminofluorene (AF), *N*-acetylaminofluorene (AAF), and 4-aminobiphenyl (ABP), which undergo metabolic activation to afford N-linked C8-dG adducts (Figure 1A).[4,5] Phenolic toxins, can produce O-linked C8-dG adducts[6−9] or C-linked C8-dG adducts[10−13] upon oxidative metabolism. C-linked varieties are also produced by the phenolic food toxin ochratoxin A (OTA),[11−15] the polycyclic aromatic hydrocarbon (PAH) benzo[*a*]pyrene,[16] or carcinogenic aryl hydrazines.[17,18]

Attachment of the bulky aryl moiety at the C8-site of dG can shift the glycosidic angle from the preferred *anti*-conformation for unmodified dG to the *syn*-conformation for C8-aryl-dG, which disrupts Watson−Crick hydrogen bonding with the opposite cytosine within the DNA duplex.[4,5] For duplexes containing a C8-aryl-dG, three unique conformations have been described:[19] the major-groove B-type conformer, in which the modified-dG adopts an *anti* conformation and the aryl moiety resides in the major groove;[20] the intercalated base-displaced stacked (S) conformer,[21,22] in which the glycosidic linkage of the modified-dG adopts the *syn* orientation and the aryl moiety is intrahelical; and the minor-groove wedge (W) conformer,[23−25] in which the adduct also adopts a *syn* glycosidic

Received: September 26, 2017
Published: November 29, 2017

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

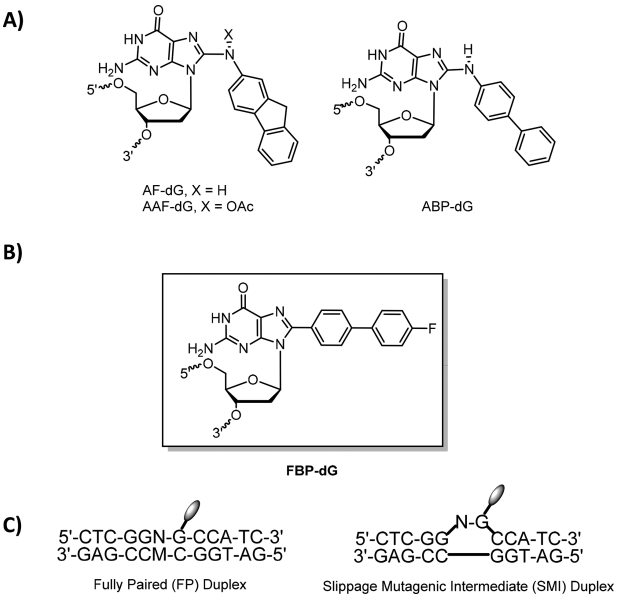


Figure 1. (A) N-linked AF-dG, AAF-dG, and ABP-dG adducts, (B) C-linked FBP-dG adduct, and (C) sequences of the fully paired and slippage mutagenic intermediate (SMI) duplexes, where N:M = C:G, G:C, A:T, or T:A in the fully paired duplex and N = C, G, A, or T in the SMI duplex. N:M = C:G represents the fully paired 12-mer *Nar*I sequence.

linkage, but the aryl moiety resides in the minor groove. The relative occupancy of each conformer depends on different helical twisting and π-stacking interactions, induced by the different linkage (C vs N vs O),[26,27] the size and planarity of the aryl moiety,[28−30] and the flanking sequence.[25,31,32] ABP-dG and AF-dG, examples for N-linked C8-dG adducts, exhibit S/B-heterogeneity, with AF-dG displaying a higher proportion of the population in the S-state due to increased planarity of the aryl moiety compared to the twisted biphenyl of ABP-dG.[33,34] AAF-dG, with an acetyl group attached to the C8-N atom, exhibits an even stronger *syn* preference than AF-dG leading to both S and W conformers.[34] In general, AF-dG and AAF-dG are more mutagenic than ABP-dG due to increased aryl planarity and stronger tendency to adopt the promutagenic *syn* conformation.[19,33−35]

The *Escherichia coli Nar*I endonuclease sequence (5′-$CG_1G_2CG_3C$-3′) has commonly been used to study the mutagenic potential of N-linked adducts. Adducts and the resulting conformational distortion of the helix influence the mutation frequency and characteristic in the *Nar*I sequence. Insertion of AAF-dG at the G_3-site is well-known to induce a higher frequency of GC deletion mutations in bacterial mutagenesis via a two-base slippage mechanism.[36,37] AAF-dG strongly stabilizes the so-called Streisinger slippage mutagenic intermediate (SMI, Figure 1C), which provides a rationale for its tendency to induce −2 deletion mutations.[38−40]

C-linked aryl rings lack the flexible tether, which results in different impacts on the DNA duplex conformation in the *Nar*I sequence context. Smaller aryl moieties, such as the phenyl-dG (Ph-dG), favor the B-type conformation and lack the ability to stabilize the SMI duplex.[29] However, adducts with linearly extended bulkier aryl moieties, such as the pyrene-dG (Py-dG), distort the duplex more and can stabilize the SMI duplex.[30] These studies were assisted by fluorescent properties of C-linked aryl-dG adducts, which permitted insight into the conformational change induced by the adducts. Py-dG in duplex DNA displayed dramatic changes in excitation wavelength compared to the single-strand form, as it was feasible for the pyrene moiety to engage in strong stacking interactions inside the duplex.[30] This suggested the potential design of a new probe to serve as a fluorescent reporter for predicting base flipping and SMI formation on the basis of changes in emission intensity and/or wavelength. To serve this purpose, aryl-dG adducts should be able to stabilize the SMI duplex and exhibit emission that is sensitive to solvent polarity. Unfortunately, Py-dG is not suitable for this purpose, as the adduct lacks strong fluorescence solvatochromism and failed to exhibit fluorescence that was sensitive to the adduct conformation.[30]

Donor−acceptor biphenyls are aromatic chromophores that exhibit strong fluorescence solvatochromism.[41,42] Attachment of electron-donating and electron-withdrawing groups at each end of the biphenyl system affords push−pull dyes with high sensitivity toward the polarity of the environment. This knowledge prompted the synthesis of a C-linked dG, containing a 4-fluoro-biphenyl moiety (FBP-dG, Figure 1B), in which dG could serve as the donor, while the fluoro-phenyl ring would act as the acceptor. Furthermore, the biphenyl system extends the phenyl ring in a linear fashion for predicted SMI duplex stabilization,[30] and the fluorine atom permits the use of ^{19}F NMR spectroscopy to define adduct conformation within the duplex.[31,34] This probe allows direct comparison with the properties of the corresponding N-linked ABP-dG adduct.[20,34] FBP-dG was incorporated into the *Nar*I sequence at the G_3-site containing various bases (N) 5′ to the lesion site of the adduct (Figure 1). Additionally, the fluorophenyl-dG (FPh-dG, Figure 1) was synthesized for ^{19}F NMR studies to complement the data on Ph-dG and permit full comparison to FBP-dG. By a combination of experimental and computational approaches, we determined the conformational preference of FBP-dG within the *Nar*I duplex and furthermore determined the ability of FBP-dG to stabilize the SMI duplex and distinguish the SMI duplex from the fully paired duplex using fluorescence spectroscopy. These data demonstrate the propensity of C-linked C8-F-biphenyl-dG to adopt an intercalated (I) structure and exhibit B/I-heterogeneity within the *Nar*I duplex and furthermore suggest FBP-dG as a chemical probe for monitoring SMI formation by fluorescence spectroscopy.

MATERIALS AND METHODS

Materials. Boronic acids (4-fluorobiphenylboronic acid and 4-fluorophenyl boronic acid), $Pd(OAc)_2$, 3,3′,3″-phosphinidynetris-(benzenesulfonic acid) trisodium salt (TPPTS), *N*,*N*-dimethylformamide diethyl acetal, 4,4′-dimethoxytrityl chloride, 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite, and other commercial products used for the synthesis of FBP-dG and FPh-dG phosphoramidites were used as received. All unmodified phosphoramidites (bz-dA-CE, ac-dC-CE, dmf-dG-CE, and dT-CE), activator (0.25 M 5-(ethylthio)-1H-tetrazole in CH_3CN), oxidizing agent (0.02 M I_2 in THF/pyridine/H_2O, 70/20/10, v/v/v), deblock (3% dichloroacetic acid in dichloromethane), cap A (THF/2,6-lutidine/acetic anhydride), cap B (methylimidazole in THF),and 1000 Å controlled pore glass (CPG) solid supports were purchased from Glen Research (Sterling, VA). Unmodified oligonucleotides were purchased from Sigma Genosys (Oakville, ON).

Methods. Suzuki cross-coupling reactions of boronic acids with 8-Br-dG to afford FBP-dG and FPh-dG were performed as described previously for other C-linked C8-dG adducts.[29,30] NMR spectra were recorded in $CDCl_3$ or DMSO-d_6 on 300 and 600 MHz Bruker

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

spectrometers, and chemical shifts were referenced to TMS (0 ppm) or the respective residual solvent peak. Full synthetic details and NMR spectra for modified phosphoramidites and NMR spectra and ESI-MS analysis of modified *Nar*I oligonucleotides are available in Supporting Information (SI). All adducted *Nar*I oligonucleotide substrates were prepared on a 1 μmol scale using a BioAutomation MerMade 12 automatic DNA synthesizer using standard or modified β-cyanoethyl-phosphoramidite chemistry.

The DNA concentration was determined by UV absorption, and the extinction coefficient was calculated using the IDT calculator at http://eu.idtdna.com/analyzer/applications/oligoanalyzer, assuming the same absorbance properties for dG and FBP-dG and FPh-dG (ε(5′-CTCGGCGCCATC) = 102,100 L/(mole·cm)). Thermal stability of *Nar*I oligonucleotides was measured by variable-temperature UV analysis using a Cary 300-Bio UV−vis spectrophotometer equipped with a 6 × 6 multicell Peltier block-heating unit and Hellma 114-QS 10 mm light path cells. T_m values were determined as previously outlined.[29,30] Oligonucleotide samples were prepared in 50 mM phosphate buffer, pH 7, with 100 mM NaCl, using equivalent amounts (6.0 μM) of the unmodified or FPh-dG and FBP-dG-modified *Nar*I oligonucleotide and its complementary strand. Circular dichroism (CD) spectra were recorded on a Jasco J-815 CD spectropolarimeter equipped with a 1 × 6 Multicell block thermal controller and a water circulator unit. Spectra were collected at 10 °C between 200 and 400 nm, with a bandwidth of 1 nm and scanning speed of 100 nm/min, as previously described.[29,30] All fluorescence spectra were recorded on a Cary Eclipse Fluorescence spectrophotometer equipped with a 1 × 4 Multicell block Peltier stirrer and temperature controller. Both excitation and emission spectra were recorded for the FPh-dG and FBP-dG-modified *Nar*I oligonucleotide hybridized to its complementary strand at 10 °C.

For dynamic ^{19}F NMR experiments, the single-stranded DNA (ssDNA) samples containing FPh-dG and FBP-dG were dissolved in a 100% D_2O solution and filtered. A stoichiometric amount of the complementary strand was added to form double-stranded DNA (dsDNA). ^{19}F NMR experiments were performed for both modified duplexes at temperatures ranging from 10 to 70 °C on a 16.4 T Bruker Avance III HD NMR spectrometer. Each experiment was carried out using proton decoupling, with 10k scans and a sweep width of 250 ppm. The spectra were analyzed, and figures were prepared using the MestReNova software.

Molecular dynamics (MD) simulations were conducted on the fully paired and SMI *Nar*I duplexes (N = C) bearing FPh-dG and FBP-dG at the G_3 position. The modified fully paired duplexes were examined in major-groove, wedge, intercalated, and base-displaced conformations, while FPh-dG and FBP-dG were considered in the *anti* and *syn* orientations in the SMI. Each system was placed in an octahedral TIP3P water box with charge equivalencing sodium ions. A total of 100 ns of production MD simulations were performed for each system using the PMEMD module of AMBER14.[43] The amber14SB force field was used,[43] which was supplemented by GAFF parameters[44] as required for the adducts. Key dihedral angles at the lesion sight, backbone RMSD, and representative structures were obtained using the cpptraj module of AMBERTOOLS.[45] Free energies were calculated using the MMPBSA method.[46] Full details of the simulations are provided in the SI.

RESULTS

Optical Properties of FBP-dG in *Nar*I. FBP-dG and FPh-dG containing *Nar*I 12-mer were synthesized using standard solid-phase synthesis (Scheme S1, SI). Solvatochromic properties of the FBP-dG nucleoside were determined as well as the conformational influence of FBP-dG in the 12-mer *Nar*I sequence, using UV−vis spectroscopy, fluorescence, and CD. *Nar*I bearing FPh-dG was used for ^{19}F NMR studies.

On the basis of UV derived thermal melting parameters of the various 12-mer *Nar*I duplexes, FBP-dG appeared to strongly destabilize all four fully paired duplexes compared to their unmodified duplex, whereas FBP-dG stabilized the SMI duplex (Table 1). When the 5′-flanking base was a G or C, the degree of the destabilization was slightly less (ΔT_m = −10.0 °C for G, −12.5 °C for C) than observed with an A or T 5′-flanking base (ΔT_m = −14.5 °C for A, −14.0 °C for T, Table 1). In the 12-mer *Nar*I duplex (N:M = C:G), FBP-dG was less destabilizing than Ph-dG (ΔT_m = −15.7 °C) in the same sequence context.[29] In the SMI duplexes containing FBP-dG, the modified duplexes were more stable when the 5′-flanking base was a pyrimidine (ΔT_m = +7.0 °C for T, +6.0 °C for C) vs a purine (ΔT_m = +4.5 °C for A, +1.5 °C for G).

Table 1. Thermal Melting Parameters of FBP-dG-Modified *Nar*I

fully paired duplex	T_m (°C)[a]	ΔT_m (°C)[b]	SMI duplex	T_m (°C)	ΔT_m (°C)
N:M = C:G, X = dG	63.5	–	N = C, X = dG	39.5	–
N:M = C:G, X = FBP-dG	51.0	−12.5	N = C, X = FBP-dG	45.5	+ 6.0
N:M = G:C, X = dG	62.0	–	N = G, X = dG	41.0	–
N:M = G:C, X = FBP-dG	52.0	−10.0	N = G, X = FBP-dG	42.5	+1.5
N:M = A:T, X = dG	61.0	–	N = A, X = dG	38.0	–
N:M = A:T, X = FBP-dG	46.5	−14.5	N = A, X = FBP-dG	42.5	+4.5
N:M = T:A, X = dG	61.0	–	N = T, X = dG	38.0	–
N:M = T:A, X = FBP-dG	47.0	−14.0	N = T, X = FBP-dG	45.0	+7.0

[a] T_m values of duplexes (6 μM) measured in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl, a heating rate of 1 °C/min and are reproducible within 3%. [b] ΔT_m = T_m (modified duplex) − T_m (unmodified duplex).

Inserting FBP-dG into the 12-mer *Nar*I sequence caused a 29 nm red-shift in excitation wavelength and little change in emission wavelength (Figure 2, Table S2), suggesting formation of a more planar biphenyl adduct structure. Emission from FBP-dG hybridized to the fully paired complementary strand was slightly quenched, when the 5′-flanking base was a pyrimidine (N = C or T), whereas virtually no change in emission intensity was observed when the 5′-flanking base was a purine (N = G or A) (Figure 2). The λ_{ex} in the fully paired duplexes, a red-shifted 4−12 nm relative to the ssDNA, was observed, suggesting increased π-stacking interactions (Table S2), while λ_{em} was blue-shifted, up to −4 nm, suggesting placement of FBP-dG in a less polar environment (Table S2). Comparing the emission intensity of the fully paired duplex to the SMI duplex, a 4.2-fold increase for the sequences containing a 5′-flanking pyrimidine and a 2.0-fold increase containing 5′-

The quantum yield and fluorescence properties of the FBP-dG nucleoside were initially determined in H_2O, CH_3CN, and $CHCl_3$ for comparison to the solvatochromic properties of other C-linked nucleosides (Table S1).[29] FBP-dG in the free nucleoside form is strongly emissive in H_2O (λ_{max} = 286 nm, λ_{em} = 424 nm, Φ_{fl} = 0.52) and is weakly sensitive to solvent polarity. These characteristics were also noted for the simple Ph-dG nucleoside (λ_{max} = 277 nm, λ_{em} = 395 nm, Φ_{fl} = 0.44) that lacks donor−acceptor (D−A) character.[29] Biphenyl systems that exhibit D−A character typically display quenched emission in water and light-up in nonpolar solvents. A red-shift in emission wavelength (λ_{em}), as observed for FBP-dG, is a typical behavior for biphenyl systems (Table S1).[41]

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

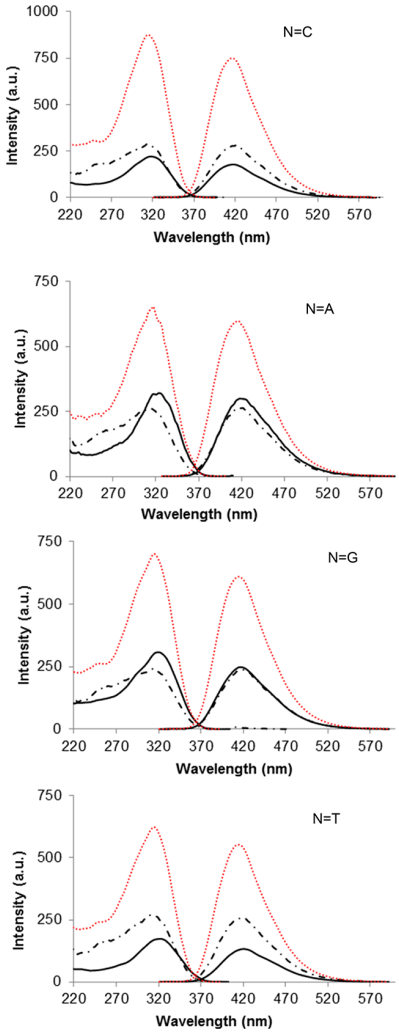


Figure 2. Excitation and emission spectra of FBP-dG-modified 12-mer *Nar*I oligonucleotides in the single-strand (black dashed-dotted traces), fully paired duplex (solid black traced), and SMI duplex (dotted red traces). All spectra of oligonucleotides (6 mM) were recorded in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl at 10 °C.

flanking purine were observed (Figure 2). This fluorescence response suggested a decrease in π-stacking interactions and less exposure of the biphenyl to the polar aqueous solvent.

CD spectral overlays (Figure S1, SI) illustrate comparisons between the adducted FP and SMI duplexes with the corresponding unmodified duplexes. In general, all duplexes gave rise to typical B-form CD patterns, with a major positive band at ~ 275 nm, a negative band at ~ 240 nm and crossover at ~ 260 nm.[47] For the four FP duplexes, the FBP-dG lesion caused a decrease in amplitude of the positive band centered at ~ 275 nm. In the SMI duplex, amplitudes of the major bands for the adducted and unmodified control duplexes were much more similar (Figure S1, SI).

Conformational Analysis of FBP-dG in *Nar*I. In order to determine the conformational preference of FBP-dG within the fully paired and SMI *Nar*I duplex (N = C), MD simulations were carried out on both duplex structures, and comparison was drawn to the corresponding structures determined for the smaller FPh-dG. Previous calculations indicate that the unsubstituted Ph-dG strongly favors the B-type conformation in the fully paired *Nar*I duplex structure.[29] In the SMI duplex, Ph-dG adopts the *anti* conformation (favored by 15.1 kJ/mol relative to the *syn* conformation). The fully paired *Nar*I duplexes containing FBP-dG or FPh-dG were also analyzed using variable-temperature ^{19}F NMR spectroscopy, to highlight the influence of the biphenyl on the duplex conformation.

MD simulations isolated four unique stable conformations for duplexes containing FPh-dG or FBP-dG (Figures 3 and 4,

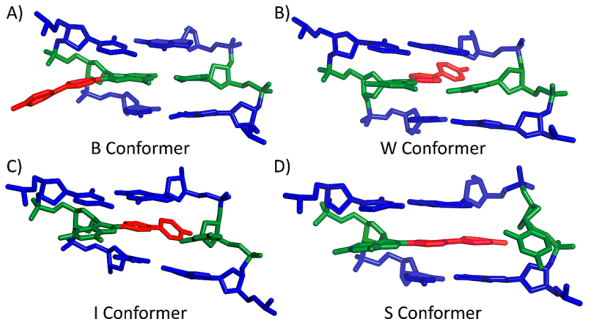


Figure 3. FBP-dG adducted *Nar*I (N = C) in the (A) major-groove conformation in which the biphenyl moiety is in the major groove, while the damaged G is in the *anti* conformation and maintains Watson–Crick hydrogen bonding with the opposing C, (B) wedge conformation in which the bulky moiety is in the minor groove, while the adducted G is *syn* and forms two Hoogsteen interactions with the pairing C, (C) intercalated conformation in which the damaged G is *syn*, and the biphenyl moiety and opposing C are intercalated within the helix, and (D) base-displaced intercalated conformation in which the adducted G is *syn*, and the bulky moiety displaces the opposing C into the extrahelical major groove location. Only the three base pairs about the lesion are shown for clarity, with the damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

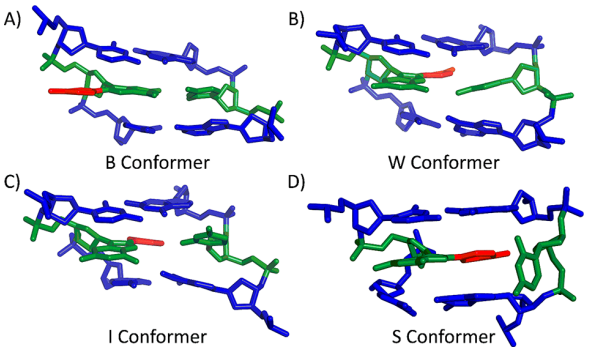


Figure 4. FPh-dG adducted *Nar*I (N = C) in the (A) major-groove conformation in which the phenyl moiety is in the major groove, while the damaged G is in the *anti* conformation and maintains Watson–Crick hydrogen bonding with the opposing C, (B) wedge conformation in which the bulky moiety is in the minor groove, while the adducted G is *syn* and forms two Hoogsteen interactions with the pairing C, (C) intercalated conformation in which the damaged G is *syn*, and the phenyl moiety and opposing C are intercalated within the helix, and (D) base-displaced intercalated conformation in which the adducted G is *syn* and the bulky moiety displaces the opposing C into the extrahelical major groove location. Only the three base pairs about the lesion are shown for clarity, with the damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37–47

Table 2 and Table S4). In all conformations, the Watson−Crick hydrogen bonding in the base pairs flanking the adduct remains

Table 2. MM-PBSA Relative Energies (kJ/mol) for Different Conformations of FPh-dG- and FBP-dG-Adducted Fully Paired *Nar*I Duplex

aryl group	major-groove	wedge	stacked	intercalated
FPh	0.0	24.8	65.0	60.8
FBP	0.0	43.4	36.4	2.3

intact, indicating that any structural deviations occur solely at the modified dG. In the major-groove or B-type conformation, FBP-dG adopts the *anti* glycosidic orientation. The bulky moiety is located in the major groove and adopts a planar arrangement about the guanine−aryl bond (θ, Figures 3A and 4A, Table S4), which minimizes helical distortion and maintains three stable Watson−Crick hydrogen bonds at the lesion site (>95% occupancy). The fluorine atom exclusively interacts with solvent. In all other adducted DNA conformations, the lesion adopts the *syn* conformation.

The wedge conformer places the biphenyl in the minor groove (Figure 3B), and the dG forms two hydrogen bonds with the opposing dC (≈ 35 and 75% occupancy). Similar to the major-groove conformation, the fluorine atom is exclusively exposed to solvent in this conformer. Since the biphenyl is positioned outside the helix in both the major-groove and wedge conformations, free rotation occurs about the ξ inter-ring dihedral angle within the bulky moiety, resulting in various relative orientations of the two rings throughout the MD simulations (i.e., standard deviation >100°; Table S4).

In the base-displaced stacked conformer, the biphenyl is located within the helix, and the opposing base is fully displaced into the major groove. Unlike FPh-dG-containing DNA, the fluorine atom has no strong interactions (e.g., hydrogen bonds or stacking contacts) with other parts of the DNA helix. MD simulations predict another closely related conformation (denoted the intercalated (I) conformer) in which the biphenyl stacks with the flanking bases and partially displaces the opposing dC such that the interactions with FBP-dG are disrupted (Figures 3C and 4C). Nevertheless, the opposing dC remains within the helix and stacks with the 5′-interstrand base. Another notable difference is the local environment of the fluorine atom. Specifically, since the aryl moiety intercalates within the helix, the fluorine atom of FPh-dG partially stacks with the 3′-dC and is more shielded from solvent than the fluorine atom in the major-groove and wedge conformations. In contrast, due to the larger bulky moiety, the fluorine atom of FBP-dG-containing DNA has no significant interactions with the DNA helix. The adducts in both intercalated conformations have reduced conformational flexibility within the aryl moiety with respect to the ξ dihedral angle compared to the B and W counterparts (i.e., standard deviation <30° for S and I, but >100° for B and W; Table S4 and Figure 5). Furthermore, the aryl moiety is on average less twisted in the stacked and intercalated conformations (average twist of 19° and 26°, respectively) compared to in the B and W conformers (33° and 31°, respectfully).

The major-groove orientation is the most stable conformation for both adducts (Table 2). Nevertheless, despite the structural similarities of the FBP- and FPh-containing duplexes, the relative energies for the next stable conformer vary significantly. Specifically, for the smaller FPh-dG lesion, the

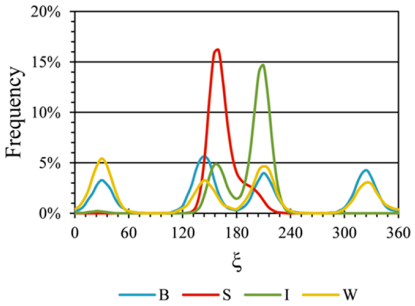


Figure 5. Frequency of the ξ dihedral angle (Figure S2) for FBP-dG-adducted fully paired *Nar*I duplex in the major-groove (B), stacked (S), intercalated (I), and wedge (W) conformations.

next most stable DNA conformation is the wedge conformer, which is 24.8 kJ/mol higher in energy. Furthermore, the small size of the phenyl moiety leads to energetically inaccessible intercalated conformations (>60 kJ/mol relative to the major-groove conformer). In contrast, the second most stable DNA conformation associated with FBP-dG is the intercalated conformer, which is only 2.3 kJ/mol higher in energy. Similarly, the base-displaced stacked conformation is slightly more stable than the wedge orientation, although both are >35 kJ/mol above the most stable major-groove orientation.

For the SMI duplexes, two stable conformations were isolated corresponding to the two glycosylic bond orientations of FBP-dG. The *anti* orientation about the gylcosidic bond places the aryl moiety within the major groove of the helix (Figure 6A) and is stabilized by Watson−Crick hydrogen

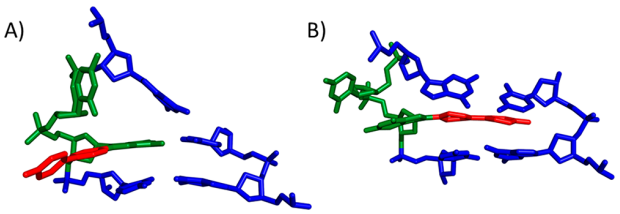


Figure 6. FBP-dG-adducted *Nar*I paired against a −2 deletion strand with the adduct in the (A) *anti* or (B) *syn* orientations. Only the lesion site including the two base pairs flanking the bulge are shown for clarity, with the damaged G and unpaired C in green, the bulky moiety in red, and flanking base pairs in blue.

bonding between the FBP-dG and the opposing 5′-dC. Due to the bulge formation, the 5′ base pair is weakened (<50% occupancy of Watson−Crick bonds), and the unpaired 5′-dC falls into the major groove. The position of the unpaired dC is highly flexible, alternating between a stacked orientation with the aryl moiety (≈40% of simulation time) and intrastrand hydrogen bonding via N4−H to a phosphate oxygen of the 5′ nucleotide (≈60% of simulation time). This predicted structure for FBP-dG is considerably similar to that previously reported for the *anti* Ph-dG in the same sequence context (Figure 7A). When FBP-dG adopts the *syn* glycosidic orientation, dG resides in the major groove and the biphenyl is intercalated such that it stacks with the base pairs flanking the bulge region (Figure 6B). As discussed for the fully paired structures, the biphenyl $\pi-\pi$ interactions in the *syn* orientation limit the rotational freedom about the ξ dihedral angle, resulting in a more planar biphenyl system in the *syn* orientation (average twist of 25°) compared to the *anti* orientation (average twist of 31°, Figure 8).

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

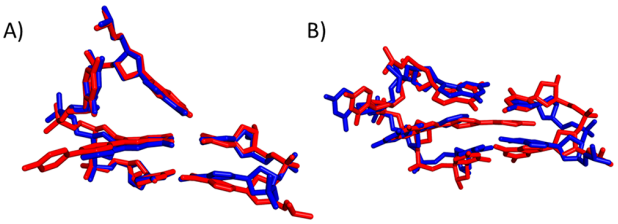


Figure 7. Overlay of the −2 deletion site for FPh-dG (blue) and FBP-dG (red) adducted helices with the adduct in (A) *anti* or (B) *syn* orientations. Only the lesion site including the two base pairs flanking the bulge are shown for clarity.

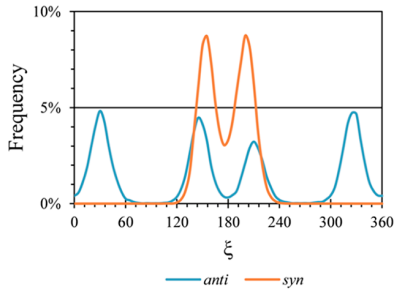


Figure 8. Frequency of the ξ dihedral angle (Figure S2) for FBP-dG-adducted SMI DNA in the *anti* and *syn* conformations.

Although the *syn* orientation of FBP-dG is similar to that previously reported for Ph-dG[29] (RMSD = 1.25 Å), the biphenyl is in a position that maximizes stacking interactions with all four nucleobases in the flanking base pairs (Figure 7B). As a result, the energy difference between the *anti* and *syn* conformations is small, with *anti* being only 2.8 kJ/mol more stable. These structural and energetic data correlate with the observed increased stability of the SMI duplexes containing FBP-dG compared to Ph-dG. Specifically, although both *anti* structures similarly lack helical interactions for the aryl group within the SMI duplex, FBP-dG *syn* structure exhibits increased stacking interactions between the aryl moiety and the flanking base pairs compared to Ph-dG, which increases the accessibility of this conformation and overall helical stability.

^{19}F NMR data were collected at variable temperatures for FPh-dG and FBP-dG-containing DNA duplexes (Figure 9). At 10 °C, the ^{19}F NMR spectrum of FPh-dG adducted DNA (Figure 9A) contains one major peak at 111.26 ppm (85% of the total adduct signal) and two small peaks at −110.96 ppm (10%) and −112.07 ppm (5%). The most downfield peak at −110.96 ppm corresponds to ssDNA since the variable-temperature data reveal that this peak gains intensity with increased temperature, while the other peaks decrease in intensity, which correlates with an increased concentration of ssDNA upon denaturing of dsDNA (Figure 9). Since the MD simulations suggest that the major-groove conformation is highly favored for the FPh-dG, the major NMR signal at −111.26 ppm (85%) was ascribed to the most stable B-form. The minor upfield peak at −112.07 ppm (5%) was attributed to the formation of the W-form. These assignments were further supported by the persistence of the peak at −111.26 ppm at higher temperatures, correlating with the greater stability of the B-form over the W-form dsDNA conformation (Table 2).

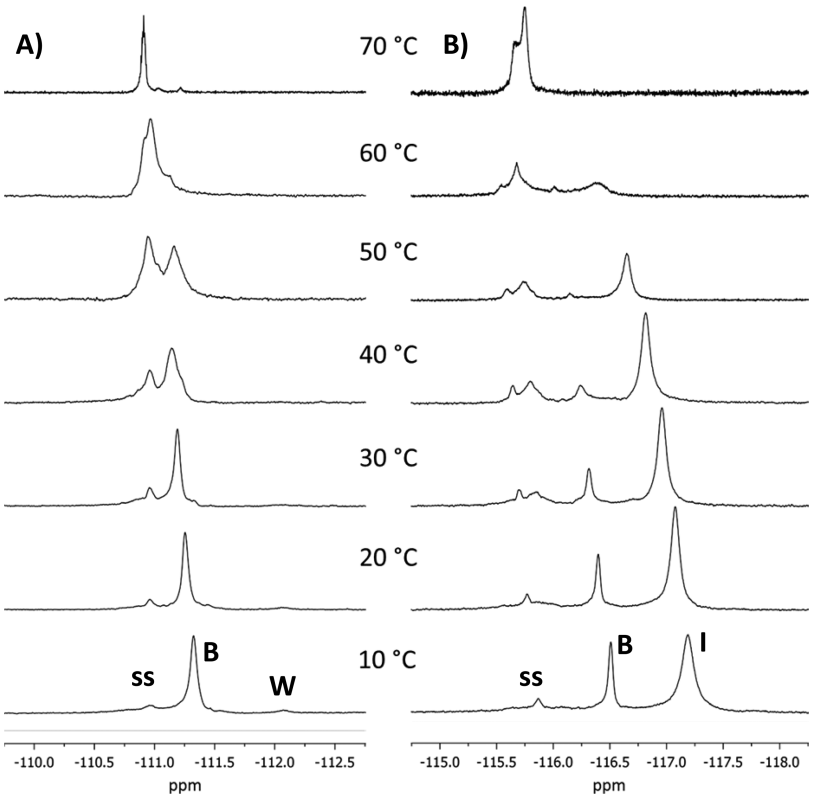


Figure 9. Variable-temperature proton decoupled ^{19}F NMR of (A) FPh-dG- and (B) FBP-dG-adducted fully paired *Nar*I duplex; single-strand (ss), major-groove (B), wedge (W), and intercalated (I).

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

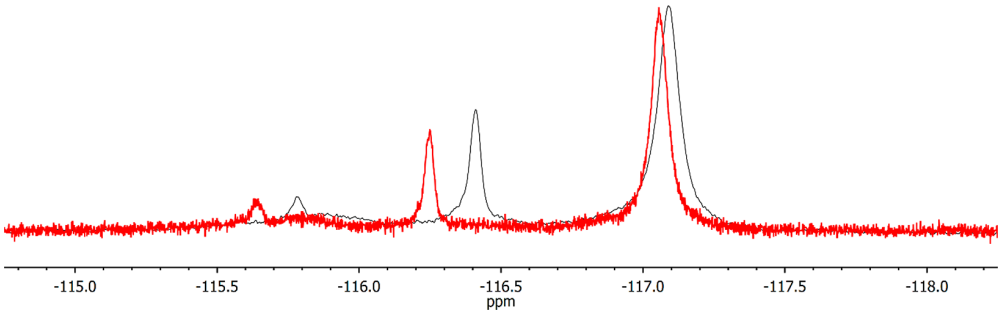


Figure 10. Isotopic effect for the proton decoupled ^{19}F NMR of the FBP-dG-adducted fully paired *Nar*I duplex in 100% D_2O (black) and 90% H_2O 10% D_2O (red).

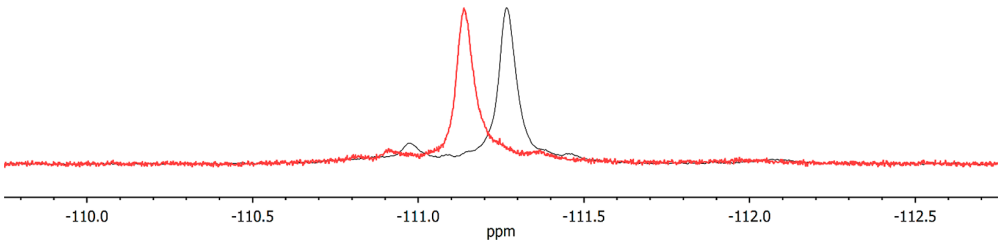


Figure 11. Isotopic effect for the proton decoupled ^{19}F NMR of the FPh-dG-adducted fully paired *Nar*I duplex in 100% D_2O (black) and 90% H_2O 10% D_2O (red).

These assignments suggest that the B-form makes up 94% of the total duplex for the FPh-dG adducted *Nar*I, with the W-form making up the remaining 6%, which is consistent with the relative energies of these duplexes determined by the MD simulations (Table 2).

The ^{19}F NMR spectra of FBP-dG-containing DNA at 10 °C contain three peaks corresponding to ssDNA and dsDNA (Figure 9B). The small downfield peak at −115.77 ppm (4%) corresponds to ssDNA, while the two major peaks at −116.40 ppm (19%) and −117.08 ppm (77%) correspond to dsDNA. The single-stranded peak was assigned based on the variable-temperature data, which indicates that the peak at −115.7 ppm increases with sample heating and corresponds to the formation of ssDNA due to annealing. Based on MD simulations, FBP-dG-containing DNA was predicted to be stable in two conformations, the major-groove and intercalated conformers (Table 2). The calculated energy difference between the major-groove and intercalated conformer of FBP-dG-containing DNA (2.3 kJ/mol) is considerably smaller than the difference between the two most stable FPh-dG adducted DNA conformations (24.8 kJ/mol), making unambiguous assignment of the peaks at −116.40 and −117.08 ppm difficult. For several N-linked C8-dG adducts, previously characterized by ^{19}F NMR, the fluorine atom in the S-type intercalated conformer resonates upfield relative to the B-type conformer.[31,34] Furthermore, at 50 °C the peak at −116.40 ppm disappears, while the peak at −117.08 ppm maintains half of the original area. Previous work on N-linked C8-dG adducts has shown that S-type intercalated structures with the aromatic moiety in the duplex are known to be more stable.[48] On the basis of these features, the downfield peak at −116.40 ppm (19%) that is closer to the ssDNA peak was ascribed to the B-type conformer, while the upfield peak at −117.08 ppm (77%) was assigned to the intercalated conformer. In this conformer, the fluorine atom is more shielded because the biphenyl is stacked within the DNA duplex. These assignments suggest that the intercalated conformer represents 80% of the FBP-dG adducted *Nar*I duplex, while the B-type conformer represents 20%.

To further clarify which peaks represent the intercalated (S and I) and groove (B and W) conformations, 1D ^{19}F NMR spectra were obtained for the adducted helices in 90% H_2O and 10% D_2O at 20 °C. By comparing the H−D isotopic effect, the degree of fluorine solvation can be determined. Specifically, the peaks with the greatest change in chemical shift will correspond to the conformation in which the fluorine label is most solvated. Previous work on fluorine-labeled adducts has demonstrated that a downfield chemical shift of ≈0.2 ppm when the deuterium content is decreased from 100% to 10% is indicative of a solvent exposed conformation.[49] Therefore, a chemical shift difference of ≈0.08 ppm corresponds to a conformation with the fluorine atom shielded from solvent. For the FBP-dG adducted DNA, the chemical shifts for the major (−117.08 ppm) and minor (−116.40 ppm) peaks move 0.04 ppm and 0.17 ppm downfield to −117.04 ppm and −116.23 ppm, respectively (Figure 10). This confirms that the major peak is solvent shielded and likely corresponds to the I conformation, while the minor peak is solvent exposed and corresponds to the B conformation. For FPh-dG-adducted DNA, the major peak shifts by 0.14 ppm (from −111.27 ppm to −111.13 ppm, Figure 11), while the minor peak shifts by 0.10 ppm (from −112.07 ppm to −111.97 ppm). Although these changes are smaller than previously reported for other solvent exposed conformations, the bulky group of FPh-dG is the smallest studied using this technique to date and does not extend the fluorine atom into the solvent sphere to the same extent as larger lesions. Therefore, the observed H−D isotopic effect supports our previous assignment of both peaks to DNA structures that are solvent exposed.

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

DISCUSSION

Conformation of FBP-dG in *Nar*I. *Comparison to Other C-Linked and N-Linked C8-dG Adducts.* On the basis of MD simulations, the *Nar*I 12-mer DNA duplex with FBP-dG at the G_3 position was predicted to display conformational heterogeneity between *anti* B-type and *syn* intercalated (I) structures. This model was consistent with ^{19}F NMR data, which indicated a 20:80 B:I ratio (Figure 9B). For the smaller FPh-dG, ^{19}F NMR analysis suggested a 94:6 B:W ratio (Figure 9C), consistent with the MD prediction of a strong *anti*-preference as well as previous MD predictions for Ph-dG adducted DNA.[29]

The conformational preferences of FBP-dG are comparable to other N-linked adducts.[19] On the basis of ^{19}F NMR data, FABP-dG and FAF-dG in a TG*A sequence context (G* represents the N-linked C8-dG adduct) exhibited B:S ratios of 100:0 for FABP-dG and 55:45 for FAF-dG.[20,49] In the 12-mer *Nar*I sequence context (i.e., CG*CC), FAF-dG exhibits a 25:75 B:S ratio.[31] There is no ^{19}F NMR study of the N-linked FABP-dG in the *Nar*I sequence; however, FAF-dG and FABP-dG have been studied in CG*CA and CG*CT sequence contexts, and FABP-dG was predicted to exhibit 95:5 and 67:33 B:S ratios, while the B:S ratios for FAF-dG were 36:64 and 20:80.[34] The B:S ratio for FAF-dG in the CG*CT sequence (20:80)[34] is almost identical to the corresponding ratio in the authentic *Nar*I sequence (25:75),[31] suggesting that the B:S ratio of 67:33 for FABP-dG in the CG*CT sequence[34] is representative of its expected ratio in the fully paired *Nar*I duplex. The B:I ratio of 20:80 of FBP-dG suggests a stronger *syn* preference, despite the calculated equilibrium between the *syn* and *anti* conformations. In fact, the B:I ratio determined for FBP-dG is much more similar to the B:S ratios determined for AF-dG. The conformational characteristics suggests greater mutagenicity for the C-linked FBP-dG compared to the N-linked ABP-dG due to its greater propensity to generate the promutagenic *syn* conformation.[19,33−35]

FBP-dG appears to stabilize the SMI duplex, especially containing a pyrimidine 5′ to the adduct (Table 1). The degree of SMI duplex stabilization by FBP-dG was essentially identical to that for the C-linked Py-dG (Table S3),[30] which was predicted to favor a *syn* conformation by 12.6 kJ/mol (Table S3). In the favored *syn* conformation, the pyrene can stack with its flanking bases.[30] The MD calculations suggested a conformational heterogeneity for FBP-dG between *anti* and *syn* in the SMI duplex. The predicted structure of the *anti* conformation is similar to that previously reported for the *anti* Ph-dG (Figure 7A),[29] which destabilizes the SMI duplex (ΔT_m vs unmodified = −2.9 °C, Table S3). On the other hand, in the *syn* conformation, increased stacking interactions between the biphenyl and its flanking bases are predicted compared to the smaller phenyl (Figure 7B). Thus, adoption of the *syn* conformation provides a rationale for the stability of the SMI duplex containing FBP-dG.

The conformational stability of the SMI is thought to be a critical determinant for the −2 frameshift mutation frequency in *Nar*I sequences and has been shown to correlate with the ability of N-linked C8-dG adducts to form stable S-conformers in the fully paired duplex.[31] Like FBP-dG, N-linked C8-dG adducts stabilize the *Nar*I SMI duplex. For example, AAF-dG is strongly stabilizing compared to the unmodified oligonucleotide (ΔT_m = +15 °C),[38,39] whereas AF-dG is only mildly stabilizing (ΔT_m values of +3.3 °C[31] and +6.0 °C[24]). AF-dG produces 10-fold fewer −2 deletion mutations compared to AAF-dG,[50] but AF-dG is expected to induce a higher frequency of deletion mutations than ABP-dG due to its greater tendency to adopt the S-type duplex structure.[34] Based on these comparisons, the deletion mutation frequency anticipated to arise from the C-linked FBP-dG adduct may be very similar to that from AF-dG, since the adducts exhibit similar *anti*/*syn* ratios in the fully paired duplex and have almost identical capacities to stabilize the SMI duplex, as evidenced by UV thermal melting experiments.

Fluorescence of FBP-dG in Fully Paired vs SMI *Nar*I Duplex. Based on fluorescence spectroscopy, the biphenyl nucleoside lacks D−A character.[29] However, when placed within the *Nar*I oligonucleotide, the λ_{ex} shifts to the red by 29 nm in the single-strand and up to 37 nm in the fully paired duplex. These changes suggest increased biphenyl planarity within the helical environment, which would be expected to increase the D−A character of FBP-dG. In the fully paired duplex with a pyrimidine 5′ to the FBP-dG, the emission was quenched compared to the single strand (Figure 2, Table 3).

Table 3. Properties of C-Linked C8-aryl-dG Adducts in Fully Paired and SMI *Nar*I 12-mer Duplexes

aryl group	duplex	E_{syn}/E_{anti} [a]	ΔT_m [b]	I_{rel} [c]
FBP	fully paired	2.3/0.0	−12.5	0.64
	SMI	2.8/0.0	+6.0	4.2

[a]Calculated relative energies (kJ/mol) of *syn* vs *anti* adduct conformation. [b]ΔT_m (°C) of adducted duplex vs unmodified control. [c]Relative emission intensity of ss vs fully paired or SMI vs fully paired duplexes.

The MD simulations coupled with the ^{19}F NMR data suggest that FBP-dG can adopt an intercalated *syn* conformation that favors π-stacking interactions between the biphenyl and the flanking bases. A wide variety of fluorescent nucleobase probes display quenched emission in duplex structures, and this is typically ascribed to π-stacking interactions, which can lead to radiationless decay.[51] In the SMI duplex, the emission intensity is turned on, because the biphenyl and the 5′-flanking base are unpartnered (Figure 2). The turn-on emission response of FBP-dG in the SMI duplex may be useful for monitoring SMI formation in the *Nar*I sequence upon synthesis with DNA polymerases.

To our knowledge, there is no literature reporting an O-linked biphenyl-dG adduct for comparison the corresponding C- and N-linked derivatives. However, the structural influence of the O-linked phenol-dG adduct has previously been published.[8] ^{19}F NMR studies and MD calculations indicate that O-linked phenyl-dG favors a B-form structure.[8] This is also suggested for the N-linked aniline-dG adduct[52] and the C-linked phenyl-dG.[29] Compared to the duplex containing the O-linked derivative, the B-form duplex containing the C-linked phenyl-dG adduct is less stable (ΔT_m = −15.7[29] vs −10.8 °C[8]) due to the stronger *syn* preference of the C-linked lesion that lacks the flexible tether separating the phenyl ring from the nucleobase. This increased *syn* preference for the C-linked lesion also provides a rationale for the greater propensity of the FBP-dG lesion to produce the I-type structure compared to the N-linked ABP-dG adduct that favors the B-type structure. In fact, the C-linked FBP-dG behaves more like N-linked AF-dG in the *Nar*I oligonucleotide than the corresponding ABP-dG. The methylene linkage between the two aromatic rings in AF-dG makes it planar for efficient π-stacking interactions, which

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

provides a rationale for its much greater propensity to produce the stacked S-type conformation and stabilize the SMI duplex.[31,34] For the C-linked FBP-dG adduct, its increased *syn* preference must stem from the difference in the length and flexibility of the linker between biphenyl and dG, and this may also influence how the biphenyl is situated in the helix.

SUMMARY

We have inserted a C-linked C8-biphenyl-dG bearing a fluorine atom (FBP-dG) into the G_3-site of the 12-mer *Nar*I sequence and characterized its emission response and impact on fully paired and slippage mutagenic intermediate (SMI) duplex stability. In the fully paired *Nar*I duplex, MD simulations combined with ^{19}F NMR demonstrate conformational heterogeneity between a major-groove B-type *anti*-adduct conformation (B) and an intercalated *syn*-adduct conformation (I), with a predicted B:I ratio of 20:80. This ratio contrasts the expected *anti*:*syn* ratio of 67:33 for the corresponding N-linked aminobiphenyl-dG (ABP-dG), but is similar to the corresponding ratio (25:75 in *Nar*I) for the N-linked 2-aminofluorene-dG (AF-dG), which adopts a planar adduct conformation due to the methylene linkage between the biphenyl rings. MD simulations for the C-linked FBP-dG adduct predict increased *syn* preference due to the lack of the flexible linker separating the dG component from the biphenyl moiety, which provides a rationale for its AF-dG-like properties within the fully paired and SMI *Nar*I duplexes. In the SMI duplex, FBP-dG and the 5′-flanking base are unpartnered within a 2-base bulge, and a turn-on fluorescence response is observed. This response suggests the potential of such modifications in DNA to enable emission monitoring the formation of SMI structures, which are mediated by polymerase processing of chemically damaged DNA, and thereby understanding lesion-induced mutagenicity.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemrestox.7b00266.

Figure S2 and Tables S1−S3 described in the text, including NMR spectra and experimental for synthetic samples; ESI-MS analysis and spectra of modified oligonucleotides; representative structures of each helix and computational details for the MD simulations (PDF)

AUTHOR INFORMATION

Corresponding Authors

*Tel:+41 44 632 91 75. E-mail: sturlas@ethz.ch.
*Tel: 403-329-2323. E-mail: stacey.wetmore@uleth.ca.
*Tel: 519-824-4120, x53963. E-mail: rmanderv@uoguelph.ca.

ORCID

Shana J. Sturla: 0000-0001-6808-5950
Stacey D. Wetmore: 0000-0002-5801-3942
Richard A. Manderville: 0000-0003-4035-8093

Funding

Support for this research was provided by the Swiss National Science Foundation (156280), the Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery grants to R.A.M. (311600-2013) and S.D.W. (2016-04568)), and the Canada Research Chair Program (950-228175 to S.D.W.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Computational resources from the New Upscale Cluster for Lethbridge to Enable Innovative Chemistry (NUCLEIC), and those provided by Westgrid and Compute/Calcul Canada, are greatly appreciated.

ABBREVIATIONS

dG, 2′-deoxyguanosine; AF, 2-aminofluorene; AAF, *N*-acetylaminofluorene; ABP, 4-aminobiphenyl; OTA, ochratoxin A; PAH, polycyclic aromatic hydrocarbon; SMI, slippage mutagenic intermediate; D−A, donor−acceptor; Ph-dG, C8-phenyl-dG; Py-dG, C8-pyrene-dG; FBP-dG, C8-fluoro-biphenyl-dG; FPh-dG, C8-fluoro-phenyl-dG; CD, circular dichroism; T_m, thermal melting temperature

REFERENCES

(1) Beland, F. A., and Kadlubar, F. F. (1990) Chemical carcinogenesis and mutagenesis, in *Handbook of Experimental Pharmacology* (Cooper, C. S., and Grover, P. L., Eds.) Vol. *94/I*, pp 267−325, Springer-Verlag, Heidelberg, Germany.

(2) Luch, A. (2005) Nature and nurture lessons from chemical carcinogenesis. *Nat. Rev. Cancer 5*, 113−125.

(3) Millen, A. L., Sharma, P., and Wetmore, S. D. (2012) C8-linked bulky guanosine DNA adducts: Experimental and computational insights into adduct conformational preferences and resulting mutagenicity. *Future Med. Chem. 4*, 1981−2007.

(4) Patel, D. J., Mao, B., Gu, Z., Hingerty, B. E., Gorin, A., Basu, A. K., and Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance. *Chem. Res. Toxicol. 11*, 391−407.

(5) Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health. 22*, 57−90.

(6) Dai, J., Wright, M. W., and Manderville, R. A. (2003) An oxygen-bonded C8-deoxy-guanosine nucleoside adduct of pentachlorophenol by peroxidase activation: Evidence for ambident C8 reactivity by phenoxyl radicals. *Chem. Res. Toxicol. 16*, 817−821.

(7) Dai, J., Sloat, A. L., Wright, M. W., and Manderville, R. A. (2005) Role of phenoxyl radicals in DNA adduction by chlorophenol xenobiotics following peroxidase activation. *Chem. Res. Toxicol. 18*, 771−779.

(8) Kuska, M. S., Witham, A. A., Sproviero, M., Manderville, R. A., Majdi Yazdi, M., Sharma, P., and Wetmore, S. D. (2013) Structural influence of C8-phenoxy-guanine in the *Nar*I recognition DNA sequence. *Chem. Res. Toxicol. 26*, 1397−1408.

(9) Witham, A. A., Verwey, A. M. R., Sproviero, M., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2015) Chlorine functionalization of a model phenolic C8-guanine adduct increases conformational rigidity and blocks extension by a Y-family DNA polymerase. *Chem. Res. Toxicol. 28*, 1346−1356.

(10) Manderville, R. A. (2005) Ambident reactivity of phenoxyl radicals in DNA adduction. *Can. J. Chem. 83*, 1261−1267.

(11) Dai, J., Wright, M. W., and Manderville, R. A. (2003) Ochratoxin A forms a carbon-bonded C8-deoxyguanosine nucleoside adduct: Implication for C8-reactivity by a phenolic radical. *J. Am. Chem. Soc. 125*, 3716−3717.

(12) Faucet, V., Pfohl-Leszkowicz, A., Dai, J., Castegnaro, M., and Manderville, R. A. (2004) Evidence for covalent DNA adduction by ochratoxin A following chronic exposure to rat and subacute exposure to pig. *Chem. Res. Toxicol. 17*, 1289−1296.

(13) Mantle, P. G., Faucet-Marquis, V., Manderville, R. A., Squillaci, B., and Pfohl-Leszkowicz, A. (2010) Structures of covalent adducts between DNA and ochratoxin A: A new factor in debate about

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

genotoxicity and human risk assessment. *Chem. Res. Toxicol. 23*, 89−98.
(14) Sharma, P., Manderville, R. A., and Wetmore, S. D. (2014) Structural and energetic characterization of the major DNA adduct formed from the food mutagen ochratoxin A in the NarI hotspot sequence: Influence of adduct ionization on the conformational preferences and implications for the NER propensity. *Nucleic Acids Res. 42*, 11831−11845.
(15) Manderville, R. A., and Wetmore, S. D. (2017) Mutagenicity of ochratoxin A: Role for a carbon-linked C8-deoxyguanosine adduct? *J. Agric. Food Chem. 65*, 7097−7105.
(16) Rogan, E. G., Cavalieri, E. L., Tibbels, S. R., Cremonesi, P., Warner, C. D., Nagel, D. L., Tomer, K. B., Cerney, R. L., and Gross, M. L. (1988) Synthesis and identification of benzo[*a*]pyrene-guanine nucleoside adducts formed by electrochemical oxidation and by horseradish peroxidase catalyzed reaction of benzo[*a*]pyrene with DNA. *J. Am. Chem. Soc. 110*, 4023−4029.
(17) Hiramoto, K., Kaku, M., Sueyoshi, A., Fujise, M., and Kikugawa, K. (1995) DNA base and deoxyribose modification by the carbon-centered radical generated from 4-(hydroxymethyl) benzenediazonium salt, a carcinogen in mushroom. *Chem. Res. Toxicol. 8*, 356−362.
(18) Gannett, P. M., Powell, J. H., Rao, R., Shi, X., Lawson, T., Kolar, C., and Toth, B. (1999) C8-Arylguanine and C8-aryladenine formation in calf thymus DNA from arenediazonium ions. *Chem. Res. Toxicol. 12*, 297−304.
(19) Cho, B. (2010) Structure-function characteristics of aromatic amine-DNA adducts, in *The Chemical Biology of DNA Damage* (Geacintov, N. E., and Broyde, S., Eds) pp 217−238, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
(20) Cho, B. P., Beland, F. A., and Marques, M. M. (1992) NMR structural studies of a 15-mer DNA sequence from *ras* protooncogene modified at the first base of codon 61 with the carcinogen 4-aminobiphenyl. *Biochemistry 31*, 9587−9602.
(21) Mao, B., Hingerty, B. E., Broyde, S., and Patel, D. J. (1998) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn*-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry 37*, 81−94.
(22) Wang, F., DeMuro, N. E., Elmquist, C. E., Stover, J. S., Rizzo, C. J., and Stone, M. P. (2006) Base-displaced intercalated structure of the food mutagen 2-amino-3-methylimidazo[4,5-*f*]quinoline in the recognition sequence of the *Nar*I restriction enzyme, a hotspot for − 2 bp deletions. *J. Am. Chem. Soc. 128*, 10085−10095.
(23) Shapiro, R., Hingerty, B. E., and Broyde, S. (1989) Minor-groove binding models for acetylaminofluorene modified DNA. *J. Biomol. Struct. Dyn. 7*, 493−513.
(24) Elmquist, C. E., Wang, F., Stover, J. S., Stone, M. P., and Rizzo, C. J. (2007) Conformational differences of the C8-deoxyguanosine adduct of 2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ) within the *Nar*I recognition sequence. *Chem. Res. Toxicol. 20*, 445−454.
(25) Wang, F., Elmquist, C. E., Stover, J. S., Rizzo, C. J., and Stone, M. P. (2007) DNA sequence modulates the conformation of the food mutagen 2-amino-3-methylimidazo[4,5-*f*]quinoline in the recognition sequence of the *Nar*I restriction enzyme. *Biochemistry 46*, 8498−8516.
(26) Sharma, P., Majdi Yazdi, M., Merriman, A., Manderville, R. A., and Wetmore, S. D. (2015) Influence of the linkage type and functional groups in the carcinogenic moiety on the conformational preferences of damaged DNA: Structural and energetic characterization of carbon- and oxygen-linked C8-phenolic-guanine adducts. *Chem. Res. Toxicol. 28*, 782−796.
(27) Manderville, R. A., and Wetmore, S. D. (2017) Understanding the mutagenicity of O-linked and C-linked guanine DNA Adducts: A combined experimental and computational approach. *Chem. Res. Toxicol. 30*, 177−188.
(28) Josephy, P. D., and Mannervik, B. (2006) *Molecular Toxicology*, 2nd ed., pp 517−518, Oxford University Press, Inc., New York.
(29) Sproviero, M., Verwey, A. M. R., Rankin, K. M., Witham, A. A., Soldatov, D. V., Manderville, R. A., Fekry, M. I., Sturla, S. J., Sharma, P., and Wetmore, S. D. (2014) Structural and biochemical impact of C8-aryl-guanine adducts within the *Nar*I recognition DNA sequence: Influence of aryl ring size on targeted and semi-targeted mutagenicity. *Nucleic Acids Res. 42*, 13405−13421.
(30) Sproviero, M., Verwey, A. M. R., Witham, A. A., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2015) Enhancing bulge stabilization through linear extension of C8-aryl-guanine adducts to promote polymerase blockage or strand realignment to produce a C: C mismatch. *Chem. Res. Toxicol. 28*, 1647−1658.
(31) Jain, N., Li, Y., Zhang, L., Meneni, S. R., and Cho, B. P. (2007) Probing the sequence effects on NarI-induced − 2 frameshift mutagenesis by dynamic ^{19}F NMR, UV, and CD spectroscopy. *Biochemistry 46*, 13310−13321.
(32) Jain, V., Hilton, B., Lin, B., Patnaik, S., Liang, F., Darian, E., Zou, Y., Mackerell, A. D., Jr., and Cho, B. P. (2013) Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Res. 41*, 869−880.
(33) Liang, F., and Cho, B. P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry 49*, 259−266.
(34) Jain, V., Vaidyanahan, V. G., Patnaik, S., Gopal, S., and Cho, B. P. (2014) Conformational insights into the lesion and sequence effects for arylamine-induced translesion DNA synthesis: ^{19}F NMR, surface plasmon resonance, and primer kinetic studies. *Biochemistry 53*, 4059−4071.
(35) Verghis, S. B. M., Essigmann, J. M., Kadlubar, F. F., Morningstar, M. L., and Lasko, D. D. (1997) Specificity of mutagenesis by 4-aminobiphenyl: Mutations at G residues in bacteriophage M13 DNA and G→C transversions at a unique dG8−ABP lesion in single-stranded DNA. *Carcinogenesis 18*, 2403−2414.
(36) Fuchs, R. P. P., Schwartz, N., and Daune, M. P. (1981) Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature 294*, 657−659.
(37) Burnouf, D., Koehl, P., and Fuchs, R. (1989) Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci. U. S. A. 86*, 4147−4151.
(38) Zhou, Y., and Romano, L. J. (1993) Solid-phase synthesis of oligonucleotides containing site-specific N-(2'-deoxyguanosin-8-yl)-2-(acetylamino)fluorene adducts using 9-fluorenylmethoxycarbonyl as the base-protecting group. *Biochemistry 32*, 14043−14052.
(39) Milhe, C., Fuchs, R. P. P., and Lefevre, J. F. (1996) NMR data show that the carcinogen N-2-acetylaminofluorene stabilises an intermediate of −2 frameshift mutagenesis in a region of high mutation frequency. *Eur. J. Biochem. 235*, 120−127.
(40) Xu, L., and Cho, B. P. (2016) Conformational insights into the mechanism of acetylaminofluorene-dG-induced frameshift mutations in the NarI mutational hotspot. *Chem. Res. Toxicol. 29*, 213−226.
(41) Maus, M., Rettig, W., Bonafoux, D., and Lapouyade, R. (1999) Photoinduced intramolecular charge transfer in a series of differently twisted donor−acceptor biphenyls as revealed by fluorescence. *J. Phys. Chem. A 103*, 3388−3401.
(42) Fukuda, R., and Ehara, M. (2013) Electronic excited states and electronic spectra of biphenyl: A study using many-body wavefunction methods and density functional theories. *Phys. Chem. Chem. Phys. 15*, 17426−17434.
(43) Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., and Kollman, P. A. (2014), *AMBER 14*; University of California, San Francisco.
(44) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general AMBER force field. *J. Comput. Chem. 25*, 1157−1174.
(45) Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

Woods, R. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668−1688.
(46) Miller, B. R., III, McGee, T. D., Jr, Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput. 8*, 3314−3321.
(47) Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlíčková, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res. 37*, 1713−1725.
(48) Sandineni, A., Lin, B., MacKerell, A. A., Jr., and Cho, B. P. (2013) Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis. *Chem. Res. Toxicol. 26*, 937−91.
(49) Zhou, L., Rajabzadeh, M., Traficante, D. D., and Cho, B. P. (1997) Conformational heterogeneity of aryl-amine-modified DNA: ^{19}F NMR evidence. *J. Am. Chem. Soc. 119*, 5384−5389.
(50) Shibutani, S. (2004) Requirements for frame-shift deletion during translesion synthesis. *Kankyo Hen'igen Kenkyu 26*, 135−141.
(51) Rankin, K. M., Sproviero, M., Rankin, K., Sharma, P., Wetmore, S. D., and Manderville, R. A. (2012) C8-Heteroaryl-2′-deoxyguanosine adducts as conformational fluorescent probes in the *Nar*I recognition sequence. *J. Org. Chem. 77*, 10498−10508.
(52) Shapiro, R., Ellis, S., Hingerty, B. E., and Broyde, S. (1998) Effect of ring size on conformations of aromatic amine−DNA adducts: The aniline−C8 Guanine adduct resides in the B-DNA major groove. *Chem. Res. Toxicol. 11*, 335−341.

DOI: 10.1021/acs.chemrestox.7b00266
Chem. Res. Toxicol. 2018, 31, 37−47

Supporting Information For:

Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the *Nar*I Mutational Hotspot: Evidence for Enhanced *syn*-Adduct Formation

Florence D. Berger, Shana J. Sturla,* Ryan W. Kung, Tony Montina, Stacey D. Wetmore,* and Richard A. Manderville*

**Corresponding Authors:*

Shana J. Sturla: Department of Health Sciences and Technology, Institute of Food, Nutrition and Health, ETH Zürich, 8092 Zürich, Switzerland
Tel:+41 44 632 91 75
E-mail: sturlas@ethz.ch

Stacey D. Wetmore: Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada, T1K 3M4
Tel: 403-329-2323
E-mail: stacey.wetmore@uleth.ca

Richard A. Manderville: Departments of Chemistry and Toxicology, University of Guelph, Guelph, ON, Canada, N1G 2W1
Tel: 519-824-4120, x53963
E-mail: rmanderv@uoguelph.ca

Table of Contents

List of Figures

List of Tables

List of Schemes

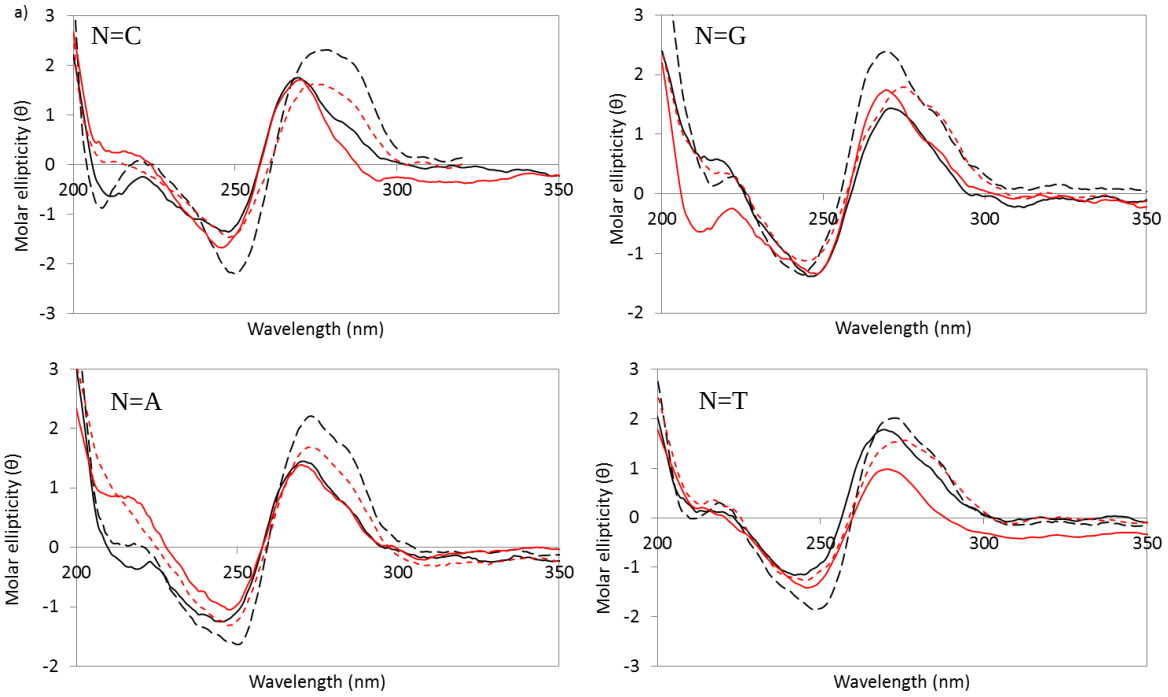


Figure S1. CD spectral overlays of *Nar*I duplexes, illustrating a general B-form pattern with a major positive band at 275 nm, a negative band at 240 nm and crossover at 260 nm. NarI(G):NarI(FP) (red lines), NarI(G):NarI(SMI) (dashed red lines), NarI(^{FBP}G):NarI(FP) (black lines), NarI(^{FBP}G):NarI(SMI) (dashed black lines) with N=A; C; G or T

Table S1. Solvatochromic Properties of FBP-dG.

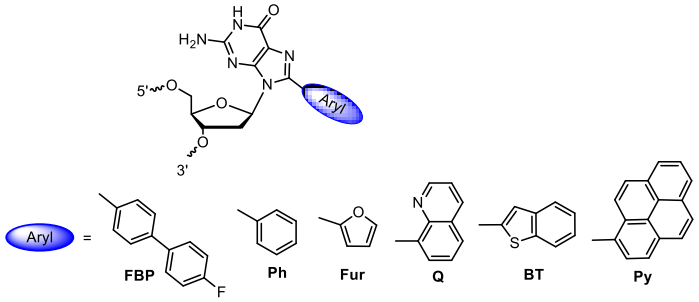
Solvent	**λ_{ex} (nm)**	**λ_{em} (nm)**	**λ_{fl}**
H_2O	**286**	**424**	**0.52**
$CHCl_3$	**303**	**410**	**0.48**
CH_3CN	**305**	**422**	**0.60**

Table S2. Photophysical parameters for the modified *Nar*I(12mer).

N	M	λ_{ex} (nm)[a]	$\Delta\lambda_{ex}$ (nm)[b]	λ_{em} (nm)	$\Delta\lambda_{em}$ (nm)	$\Delta\nu$ (cm^{-1})[c]
C	/[f]	314	/	421	/	8089
	G	318	4	417	-4	7466
	-2	314	0	417	-4	7866
G	/	313	/	420	/	8118
	C	319	6	416	-4	7305
	-2	315	2	414	-6	7576
A	/	315	/	422	/	8034
	T	323	8	420	-2	7141
	-2	315	0	413	-9	7533
T	/	311	/	420	/	8323
	A	323	12	420	0	7141
	-2	315	4	415	-5	7640

[a]All spectra for single strand and duplexes (6.0 μM) were recorded in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl.
[b] Change in excitation or emission maximum for duplex versus single strand.
[c] Stokes' shift ($\Delta\nu$) is calculated as $(1/\lambda_{ex} - 1/\lambda_{em})$
[d]$I_{rel} = I_{single\text{-}strand}/I_{FP}$
[f] $I_{rel} = I_{SMI}/ I_{FP}$
[f] / indicates optical properties of the modified base in the single strand.

Table S3. Properties of C-linked C8-aryl-dG adducts in fully paired and SMI *Nar*I 12-mer duplexes.



Aryl group	Duplex	fully paired E_{syn}/E_{anti} [a]	fully paired ΔT_m [b]	fully paired I_{rel} [c]
FBP	fully paired	2.3/0.0	−12.5	0.64
	SMI	2.8/0.0	+6.0	4.2
Ph[d]	fully paired	26.4/0.0	−15.7	0.38
	SMI	15.1/0.0	−2.9	0.80
Fur[d]	fully paired	14.2/0.0	-9.1	0.45
	SMI	14.2/0.0	-2.6	2.97
Q[d]	fully paired	25.1/0.0	−18.6	0.41
	SMI	0.0/23.8	0.0	4.92
BT[e]	fully paired	18.4/0.0	−15.8	1.45
	SMI	10.0/0.0	+3.8	1.05
Py[e]	fully paired	7.9/0.0	−20.1	0.85
	SMI	0.0/12.6	+6.2	0.61

[a] Calculated relative energies (kJ/mol) of *syn* versus *anti* adduct conformation. [b] ΔT_m (°C) of adducted duplex vs. unmodified control. [c] Relative emission intensity of ss vs fully paired or SMI vs fully paired duplexes. [d] Values taken from Ref. 29. [e] Values taken from Ref. 30.

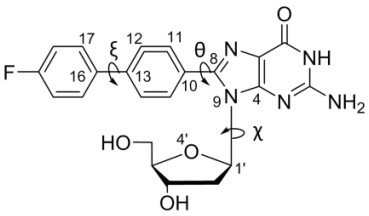
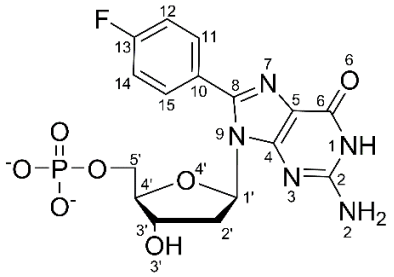


Figure S2. Chemical numbering and definition of key dihedral angles in the [FBP]dG adduct, with χ = ∠(O4′C1′N9C4), θ = ∠(N9C8C10C11), and ξ = ∠(C12C13C16C17).

Table S4: Average values of key dihedral angles (deg.) and backbone RMSD for clustered simulations on ^{FPh}dG or ^{FBP}dG-adducted DNA.

Adduct	Conformation	χ (°)	θ (°)	ξ (°)	RMSD (Å)
^{FPh}dG	Major-Groove	–133.4 (9.3)	–176.7 (6.8)	N/A	2.35 (0.55)
	Wedge	38.9 (10.0)	175.9 (8.4)	N/A	2.39 (0.43)
	Intercalated	38.9 (13.5)	–5.0 (8.5)	N/A	3.95 (0.67)
	Base-Displaced	35.8 (8.5)	-5.6 (7.3)	N/A	3.90 (0.57)
^{FBP}dG	Major-Groove	–132.5 (9.3)	–177.0 (6.9)	162.9 (100.8)	2.32 (0.52)
	Wedge	61.3 (22.7)	–176.7 (10.0)	42.7 (107.8)	2.79 (0.54)
	Intercalated	35.6 (10.7)	–4.1 (7.4)	–162.0 (28.9)	3.06 (0.53)
	Base-Displaced	40.3 (10.0)	–7.7 (7.5)	167.6 (16.8)	3.00 (0.51)
	–2 strand - *anti*	-131.6 (11.1)	4.6 (8.2)	18.9 (101.9)	3.22 (0.55)
	–2 strand - *syn*	40.2 (10.4)	176.0 (6.7)	-179.5 (25.3)	3.69 (0.65)

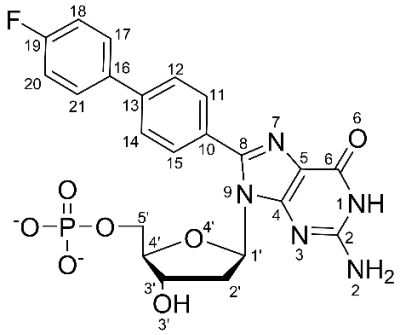
Table S5. MOL2 file of the F-Phenyl-dG adduct generated using the RED program (atom numbering provided in the figure below).



1	P	4.682	0.033	-1.207	P	1	LIG	1.2093
2	O5'	4.186	1.03	0	OS	1	LIG	-0.4938
3	OP1	4.581	0.757	-2.483	O2	1	LIG	-0.7864
4	OP2	5.894	-0.639	-0.716	O2	1	LIG	-0.7950
5	O3'	0	0	0	O	1	LIG	-0.5494
6	C5'	3.127	1.932	-0.26	CI	1	LIG	-0.0222
7	H5'	2.887	2.429	0.688	H1	1	LIG	0.0806
8	H5"	3.42	2.712	-0.98	H1	1	LIG	0.0806
9	C4'	1.869	1.246	-0.794	CT	1	LIG	0.1858
10	H4'	1.04	1.964	-0.816	H1	1	LIG	0.0992
11	O4'	2.118	0.795	-2.146	OS	1	LIG	-0.4181
12	C1'	1.887	-0.604	-2.254	CT	1	LIG	0.1229
13	H1'	0.883	-0.795	-2.639	H2	1	LIG	0.0912
14	C3'	1.429	0	0	CT	1	LIG	0.1455
15	H3'	1.825	0.015	1.023	H1	1	LIG	0.0749
16	C2'	2.004	-1.158	-0.825	CT	1	LIG	-0.0637
17	H2'	3.047	-1.33	-0.555	HC	1	LIG	0.0455
18	H2"	1.444	-2.09	-0.699	HC	1	LIG	0.0455
19	N9	2.815	-1.158	-3.232	N*	1	LIG	-0.0240
20	C8	2.515	-1.806	-4.443	CK	1	LIG	0.2328
21	C4	4.182	-1.246	-3.094	CB	1	LIG	0.1719
22	N7	3.587	-2.314	-5.013	NB	1	LIG	-0.5106
23	C5	4.634	-1.972	-4.19	CB	1	LIG	0.1374
24	N3	4.92	-0.696	-2.094	NC	1	LIG	-0.4924
25	C6	6.045	-2.234	-4.327	C	1	LIG	0.4724
26	C2	6.214	-0.915	-2.198	CA	1	LIG	0.5847
27	O6	6.669	-2.851	-5.169	O	1	LIG	-0.5308
28	N1	6.76	-1.631	-3.227	NA	1	LIG	-0.4411
29	H1	7.755	-1.825	-3.24	H	1	LIG	0.3482
30	N2	7.049	-0.449	-1.215	N2	1	LIG	-0.8464
31	H21	7.996	-0.206	-1.473	H	1	LIG	0.3871
32	H22	6.607	0.204	-0.577	H	1	LIG	0.3871
33	C10	1.172	-1.882	-5.041	CA	1	LIG	0.0197
34	C11	0.287	-0.789	-5.048	CA	1	LIG	-0.1202

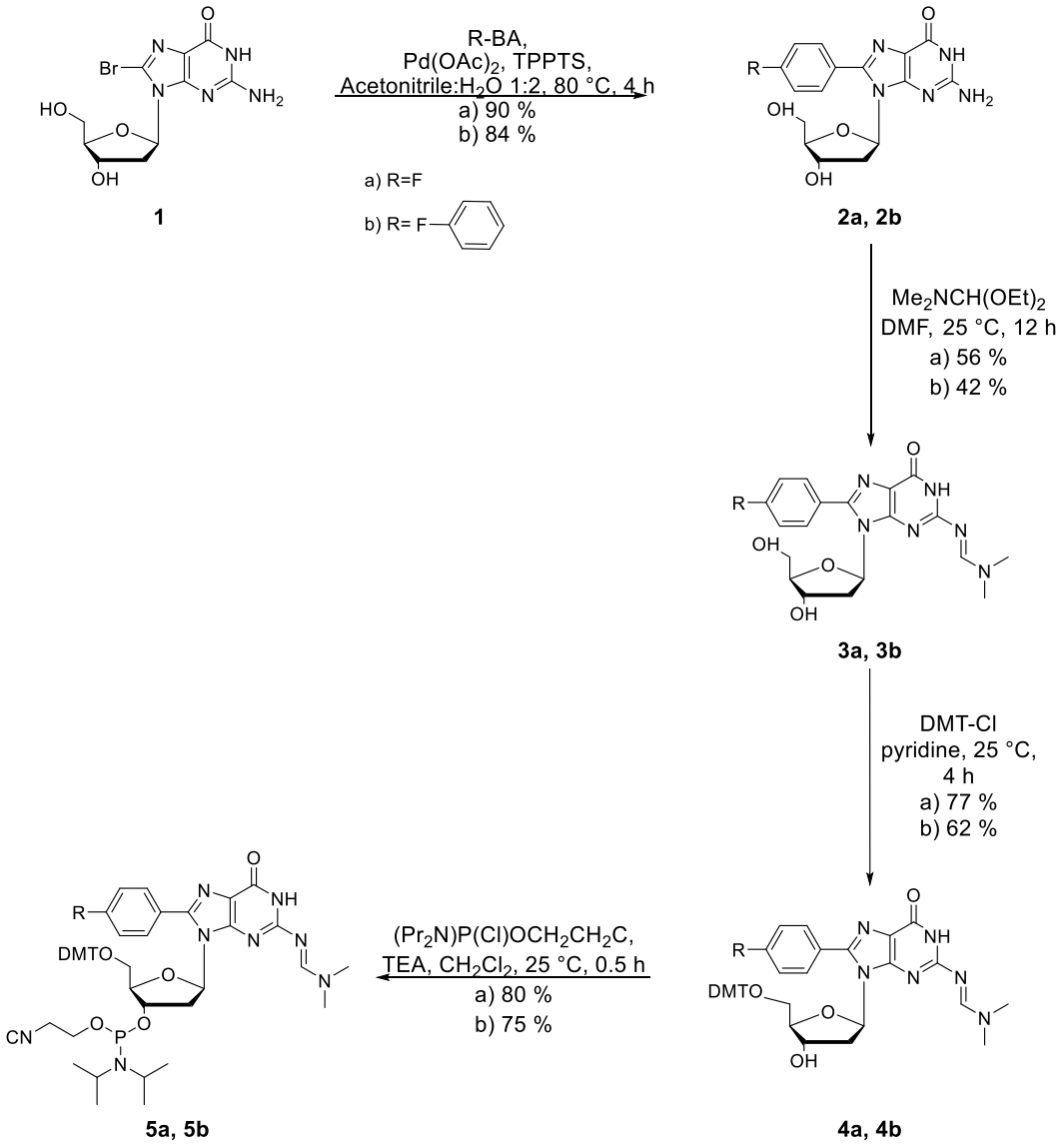
35	H11	0.581	0.153	-4.595	HA	1	LIG	0.1545
36	C15	0.796	-3.065	-5.7	CA	1	LIG	-0.1290
37	H15	1.492	-3.897	-5.719	HA	1	LIG	0.1692
38	C12	-0.957	-0.885	-5.669	CA	1	LIG	-0.2080
39	H12	-1.648	-0.049	-5.686	HA	1	LIG	0.1696
40	C14	-0.441	-3.169	-6.328	CA	1	LIG	-0.2164
41	H14	-0.746	-4.077	-6.837	HA	1	LIG	0.1699
42	C13	-1.303	-2.077	-6.295	CA	1	LIG	0.2515
43	F	-2.506	-2.175	-6.896	F	1	LIG	-0.1896

Table S6. MOL2 file of the FBP-dG adduct generated using the RED program (atom numbering provided in the figure below).



1	P	4.701	0.017	-1.112	P	1	LIG	1.2091
2	O5	4.162	1.098	0	OS	1	LIG	-0.4910
3	OP1	4.608	0.63	-2.446	O2	1	LIG	-0.7864
4	OP2	5.918	-0.584	-0.547	O2	1	LIG	-0.7952
5	O3'	0	0	0	OS	1	LIG	-0.5473
6	C5'	3.088	1.952	-0.351	CI	1	LIG	-0.0175
7	H5'	2.801	2.504	0.552	H1	1	LIG	0.0816
8	H5"	3.384	2.69	-1.114	H1	1	LIG	0.0816
9	C4'	1.872	1.189	-0.877	CT	1	LIG	0.1523
10	H4'	1.025	1.877	-0.991	H1	1	LIG	0.1052
11	O4'	2.196	0.641	-2.175	OS	1	LIG	-0.3937
12	C1'	1.937	-0.755	-2.206	CT	1	LIG	0.1104
13	H1'	0.946	-0.94	-2.629	H2	1	LIG	0.1002
14	C3'	1.429	0	0	CT	1	LIG	0.1513
15	H3'	1.824	0.085	1.02	H1	1	LIG	0.0733
16	C2'	2.005	-1.213	-0.741	CT	1	LIG	-0.0655
17	H2'	3.044	-1.38	-0.446	HC	1	LIG	0.0462
18	H2"	1.436	-2.131	-0.558	HC	1	LIG	0.0462
19	N9	2.898	-1.392	-3.104	N*	1	LIG	-0.0064

20	C8	2.64	-2.089	-4.3	CK	1	LIG	0.2410
21	C4	4.258	-1.167	-3.096	CB	1	LIG	0.1563
22	N7	3.735	-2.301	-4.998	NB	1	LIG	-0.5275
23	C5	4.75	-1.734	-4.267	CB	1	LIG	0.1547
24	N3	4.955	-0.501	-2.139	NC	1	LIG	-0.5181
25	C6	6.162	-1.658	-4.549	C	1	LIG	0.4914
26	C2	6.246	-0.405	-2.378	CA	1	LIG	0.6308
27	O6	6.817	-2.079	-5.484	O	1	LIG	-0.5370
28	N1	6.832	-0.942	-3.489	NA	1	LIG	-0.4811
29	H1	7.838	-0.893	-3.61	H	1	LIG	0.3568
30	N2	7.046	0.219	-1.455	N2	1	LIG	-0.8602
31	H21	7.89	0.667	-1.786	H	1	LIG	0.3885
32	H22	6.53	0.754	-0.762	H	1	LIG	0.3885
33	C10	1.316	-2.567	-4.726	CA	1	LIG	0.0296
34	C11	0.362	-3.091	-3.839	CA	1	LIG	-0.0793
35	H11	0.583	-3.163	-2.779	HA	1	LIG	0.1102
36	C15	1.015	-2.56	-6.098	CA	1	LIG	-0.1559
37	H15	1.76	-2.188	-6.793	HA	1	LIG	0.1604
38	C12	-0.863	-3.56	-4.307	CA	1	LIG	-0.1924
39	H12	-1.592	-3.937	-3.595	HA	1	LIG	0.1403
40	C14	-0.206	-3.035	-6.56	CA	1	LIG	-0.1317
41	H14	-0.404	-3.038	-7.628	HA	1	LIG	0.1382
42	C13	-1.177	-3.538	-5.675	cp	1	LIG	0.0223
43	C16	-2.485	-4.033	-6.171	cp	1	LIG	0.0372
44	C17	-3.103	-5.156	-5.593	CA	1	LIG	-0.1293
45	H17	-2.602	-5.688	-4.789	HA	1	LIG	0.1506
46	C21	-3.144	-3.392	-7.234	CA	1	LIG	-0.1522
47	H21'	-2.7	-2.508	-7.683	HA	1	LIG	0.1472
48	C18	-4.332	-5.625	-6.052	CA	1	LIG	-0.2504
49	H18	-4.807	-6.497	-5.616	HA	1	LIG	0.1744
50	C20	-4.372	-3.848	-7.707	CA	1	LIG	-0.2233
51	H20	-4.888	-3.35	-8.521	HA	1	LIG	0.1690
52	C19	-4.949	-4.96	-7.106	CA	1	LIG	0.3046
53	F	-6.139	-5.408	-7.557	F	1	LIG	-0.2078



Scheme S1. Synthesis of FPh-dG phosphoramidite and FBP-dG phosphoramidite

Synthesis of Modified *Nar*I. FPh-dG and FBP-dG were synthesized by Suzuki-Miyaura cross-coupling between 8-Br-dG and 4-fluorophenyl boronic acid or 4-fluorobiphenylboronic acid and converted into the phosphoramidite (Scheme S1) by adaption of the method described previously for other C-linked C8-aryl-dG adducts.[1-2] Solid phase DNA synthesis was then used to incorporate the corresponding phosphoramidite into the 12-mer *Nar*I sequence (5′-CTCGGNXCCATC, N=C, X=FBP-dG, FPh-dG) and additionally in oligonucleotides with varying 5′-flanking base (N = A, G or T). The various adducted 12-mer sequences were annealed to a complementary 12-mer sequence or a truncated 10mer sequence 5′-GATGGCCGAG modeling the SMI.

8-(aryl)-2'-deoxyguanosine (2a, 2b). The title compound was synthesized by adaption of a previously described method.[1] 8-Br-dG (1) (0.3 g, 0.9 mmol), $Pd(OAc)_2$ (0.005 g, 0.026 mmol),

TPPTS (0.03 mg, 0.05 mmol), Na_2CO_3 (0.3 g, 1.7 mmol) and **a) F-phenyl-boronic acid** (0.14 g, 1.2 mmol), **b) 4-(4-Fluorophenyl)phenylboronic acid** (0.23 g, 1.0 mmol) were dissolved in 2:1 H_20: acetonitrile allowed to react for 4 h under reflux at 80 °C. The reaction was diluted with ca. 20 mL of water and the pH adjusted to 6−7 with 1 M aq. HCl. The mixture was allowed to cool to 0 °C for 16 h hours and the precipitate was recovered by vacuum filtration as solid, **2a)** 0.25 g (80 %), ^{1}H NMR (300 MHz, DMSO-d_6) δ 1.97-2.04 (m, 1H), 3.07-3.17 (m, 1H), 3.48-3.54 (m, 1H), 3.60-3.66 (m, 1H), 3.76-3.80 (m, 1H), 4.30-4.32 (t, 1H), 5.15-5.21 (t, 1H), 5.99-6.04 (t, 1H), 6.59 (s 1H), 7.33-7.39 (t, 2H), 7.65-7.70 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 36.62, 62.12, 71.21, 84.67, 87.92, 115.53, 115.76, 117.12, 131.34, 131.42, 145.57, 152.05, 154.53, 161.36, 163.82. HRMS calcd for $C_{16}H_{17}FN_5O_4$ [M+H$^+$]: 362.1265, found: 362.1234. **b)** 0.34 g (90 %). ^{1}H NMR (400 MHz, DMSO-d_6) δ 2.04-2.08 (m, 1H), 3.23-3.27 (m, 1H), 3.59 (s, 1H), 3.70-3.72 (d, 1H), 3.86 (s, 1H), 4.39 (s, 1H), 5.20 (s, 1H), 5.87 (2, 1H), 6.14-6.17 (t, 1H), 6.74-6.75 (bt, 2H), 7.32-7.36 (t, 2H), 7.73-7.74 (d, 2H), 7.83-7.85 (d, 4H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 37.23, 62.68, 71.78, 85.44, 88.42, 115.93, 116.13, 117.86, 126.93, 129.011, 129.73, 135.95, 139.74, 145.55, 152.55, 156.94, 161.03, 162.54, 163.50. HRMS calcd for $C_{22}H_{21}FN_5O_4$ [M+H$^+$]: 438.1578, found: 438.1560

N^2-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (3a, 3b). The title compound was synthesized by adaption of a previously described method.[2] **2a)** (0.55 g, 1.322 mmol), **2b)** (0.34 g, 0.78 mmol) and dimethyl-formamide-diethyl-acetal (a) 1.1 mL, 6.08 mmol, b) 0.53 mL, 3.11 mmol) were dissolved in DMF in a round bottomed flask and allowed to react for 16 h at room temperature under argon atmosphere. The resulting solution was then evaporated to dryness, washed with MeOH and used without further purification to yield product**. a)** 0.32 g (51 %) ^{1}H NMR (400 MHz, DMSO-d_6) δ 2.065-2.11 (m, 1H), 3.04 (s, 3H), 3.15 (s, 3H), 3.18-3.25 (m, 1H), 3.38 (s, 1H), 3.54-3.57 (m, 1H), 3.65-3.68 (d, 2H), 3.81-3.82 (d, 2H), 4.92 (s, 1H), 5.25 (s, 1H), 6.06-6.09 (t, 1H), 7.38-7.42 (t, 2H), 7.70-7.73 (q, 2H), 8.51 (s, 1H). ^{13}C NMR (400 MHz, DMSO-d_6) δ 34.84, 37.17, 41.09, 48.84, 62.17, 71.21, 84.98, 87.95, 115.93, 116.15, 131.70, 131.78, 147.42, 150.92, 157.13, 157.78, 158.38, 161.80, 162.26, 164.26. HRMS calcd for $C_{19}H_{22}FN_6O_4$ [M+H$^+$]: 417.1687, found: 417.1664. **b)** 0.16 g (42 %). ^{1}H NMR (300 MHz, DMSO-d_6) δ 2.01-2.08 (m, 1H), 3.0 (s, 3H), 3.11 (s, 3H), 3.17-3.23 (m, 1H), 3.51-3.57 (m, 1H), 3.61-3.66 (m, 1H), 3.76-3.81 (m, 1H), 4.40-4.41 (d, 1H), 4.82-4.86 (t, 1H), 5.17-5.18 (d, 1H), 6.10-6.15 (t, 1H), 7.25-7.31 (m, 2H), 7.69-7.71 (d, 2H), 7.75-7.81 (m, 4H), 8.47 (s, 1H)**.** ^{13}C NMR (101 MHz, DMSO-d_6) δ 34.84, 37.17, 41.09, 48.84, 62.17, 71.21, 84.98, 87.95, 115.93, 116.15, 120.31, 126.89, 126.92, 131.69, 131.78, 147.42, 150.92, 157.13, 157.78, 158.38, 161.80, 164.26. HRMS calcd for $C_{25}H_{26}FN_6O_4$ [M+H$^+$]: 493.2000, found: 493.1969

5'-O-(4,4'-Dimethoxytrityl)-N^2-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (4a, 4b). The title compound was synthesized by adaption of a previously described method.[2] In a round bottomed flask N^2-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine **3a)** 0.23 g, 0.47 mmol, **3b)** 0.3 g, 0.6 mmol) was dissolved in anhydrous pyridine (10 mL) and the resulting solution was cooled to 0 °C. Subsequently, DMT-Cl (a) 0.190 g, 0.56 mmol, b) 0.25 g, 0.73 mmol) in pyridine (10 mL) was added dropwise. The mixture was allowed to react at room temperature under argon atmosphere. The reaction was monitored by TLC and after all starting material has been converted into the product (4 h), the resulting solution was diluted with CH_2Cl_2 and washed with water (2x10 mL). TEA (a few drops) was added, and the mixture was evaporated to dryness. The crude mixture was purified on with silica column chromatography and eluted with MeOH:CH_2Cl_2:TEA (5:19:5) to yield product. **a)** 0.240 g (70 %) ^{1}H NMR (300 MHz, $CDCl_3$) δ 2.22-2.31 (m, 1H), 3.06 (s, 3H), 3.11 (s, 3H), 3.20-3.25 (m, 1H), 3.31-3.33 (m, 1H), 3.45 (s, 1H),

3.77-3.78 (d, 6H), 4.00 (s, 1H), 4.70 (s, 1H), 5.38-5.40 (d, 1H), 6.19-6.23 dd, 1H), 6.79-6.85 (q, 4H), 7.24 (s, 2H), 7.25 (s, 5H), 7.37-7.48 (m, 4H), 7.86-7.91 (dd, 2H), 8.39 (s, 1H), 11.52 (bs, 1H)**.** ^{13}C NMR (101 MHz, $CDCl_3$) δ 35.08, 37.72, 41.29, 55.23, 64.24, 72.74, 83.74, 84.90, 86.33, 113.09, 115.48, 115.70, 120.45, 126.32, 126.35, 126.90, 127.85, 128.06, 129.96, 129.99, 131.71, 131.79, 135.78, 135.81, 144.72, 148.80, 151.03, 155.78, 157.79, 157.81, 158.50, 162.32, 164.80**.** HRMS calcd for $C_{40}H_{40}FN_6O_6$ [M+H$^+$]: 719.2993, found: 719.2936 **b)** 0.3 g (62 %) ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.19-2.26 (m, 1H), 2.81 (s, 3H), 2.93 (s, 3H), 3.19-3.25 (m, 1H), 3.34-3.38 (m, 1H), 3.49-3.53 (m, 1H), 3.69 (s, 6H), 3.99-4.04 (q, 1H), 4.80-4.85 (q, 1H), 6.21-6.24 (q, 1H), 6.68-6.72 (dd, 4H), 7.08-7.18 (m 5H), 7.23-7.26 (dd, 4H), 7.35-7.37 (d, 2H), 7.50-7.53 (m, 4H), 7.82-7.84 (d, 2H), 8.31 (s, 1H), 9.14-9.58 (bs, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 35.04, 37.84, 41.29, 55.20, 64.38, 72.60, 83.94, 85.11, 86.26, 113.05, 115.68, 115.89, 120.60, 128.12, 128.69, 128,77, 130.02, 130.11, 135,87, 135.93, 136.29, 136.32, 141.21, 144.79, 149.77, 151.16, 155.74, 157.74, 158.43, 161.47, 163.93. HRMS calcd for $C_{46}H_{44}FN_6O_6$ [M+H$^+$]: 795.3306, found: 795.3245

3'-O-[(2-Cyanoethoxy)(diisopropylamino)phosphine]-5'-O-(4,4'-Dimethoxytrityl)-N^2-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (5). The title compound was synthesized by adaption of a previously described method.[2] 5'-O-(4,4'-Dimethoxytrityl)-N^2-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine **4a)** 0.23 g, 0.3 mmol, **4b)** 0.25 g, 0.3 mmol) was dissolved in anhydrous CH_2Cl_2 (10 mL) and anhydrous TEA (a) 0.2 mL, 1.3 mmol b) 0.2 mL, 1.4 mmol), subsequently 2-cyanothyl-N,N-diisopropylchlorophosphoramidite (0.08 mL, 0.38 mmol) was added. The reaction was monitored by TLC and after all starting material has been converted into the product (30 min) the mixture was added dropwise to ice cold hexane and washed with hexane to yield product a) 0.13 g (45 %)[44] **a)** 0.97-1.02 (m, 6H), 1.06-1.10 (m, 6H), 2.25-2.41 (m, 2H), 2.69-2.71 (m, 3H), 2.89 (s, 3H), 3.02-3.03 (d, 3H), 3.43-3.49 (m, 5H), 3.68-3.69 (m, 7H), 4.11 (s, 2H), 4.9-5.14 (d, 1H), 6.1 (s, 1H), 6.65-6.70 (m, 4H), 7.07-7.15 (m, 9H), 7.22-7.35 (2H), 7.83 (2H), 8.3-8.4 (d, 1H), 10.1 (s, 1H) ^{13}C NMR (101 MHz, $CDCl_3$)δ 8.9,21.43, 24.5, 24.5, 35.20, 35.23, 37.14, 41.19, 41.26, 41.31, 43.23, 43.34, 43.35, 55.28, 55.24, 62.95, 63.55, 73.14, 73.36, 74.30, 74.49, 8.78, 84.86, 84.93, 112.98, 113.02, 115.49, 115.54, 115.70, 115.76, 117.09, 117.63, 117.73, 120.69, 120.70, 126.30, 126.33, 128.07, 128.13, 129.93, 130.00, 130.06, 131.70, 131.73, 131.79, 135.92, 144.76, 144.86, 148.82, 148.85, 155.92, 155.99, 157.83, 157.88, 158.25, 162.31, 162.33, 164.82. HRMS calcd for $C_{49}H_{57}FN_8O_7P$ [M+H$^+$]: 919.4072, found: 919.4013.

b) 0.18 g (75 %).^{1}H NMR (300 MHz, $CDCl_3$) δ 0.78-0.97 (m, 12H), 2.22-2.35 (m, 3H), 2.89 (s, 3H), 3.02 (s, 3H), 3.23-3.78 (m, 12H), 4.05-4.19 (m, 1H), 4.95-5.14 (m, 1H), 6.18-6.25 (m, 1H), 6.65-6.72 (m, 4H), 7.09-7.19 (m, 6H), 7.24-7.29 (m, 4H), 7.36-7.39 (m, 2H), 7.53-7.62 (d, 4H), 7.89-7.98 (m, 2H), 8.31-8.39 (d, 1H), 8.68-8.69 (bs, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 19.63, 20.26, 21.59, 22.7, 24.23, 24.34, 35.20, 36.79, 40.89, 41.94, 43.00, 43.13, 52.76, 55.10, 57.58, 58.29, 60.10, 63.09, 72.80, 74.16, 84.00, 85.84, 112.82, 115.56, 115.78, 116.49, 116.711, 117.52, 117.66. 120.49, 126.61, 126.85, 127.61, 127.89, 128.54, 128.62, 129.81, 135.63, 136.02, 141.11, 144.46, 144.71, 149.27, 150,70, 150.99, 155.71, 155.74, 155.90, 157.66, 157.79, 158.19, 161.29, 163.74. ^{31}P NMR (300 MHz, $CDCl_3$) δ 149.34, 148.99 (73.53 %), 14.29 (27.47 %). HRMS calcd for $C_{55}H_{61}FN_8O_7P$ [M+H$^+$]: 995.4385, found: 995.4304.

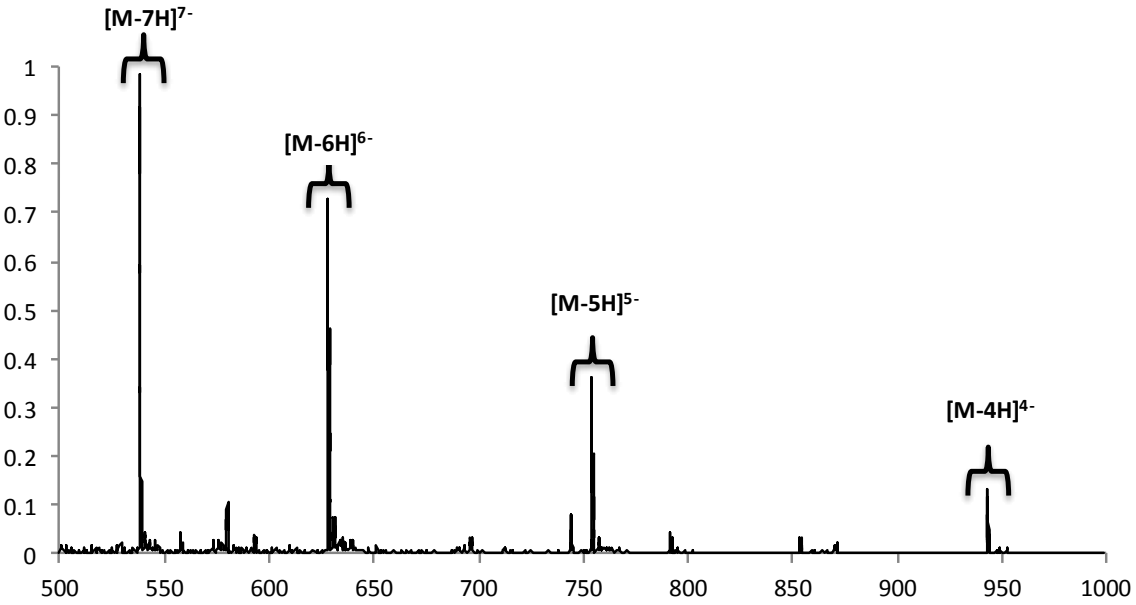


Figure S3: ESI$^-$ Spectrum of NarI(N=A, X=^{FBP}G).

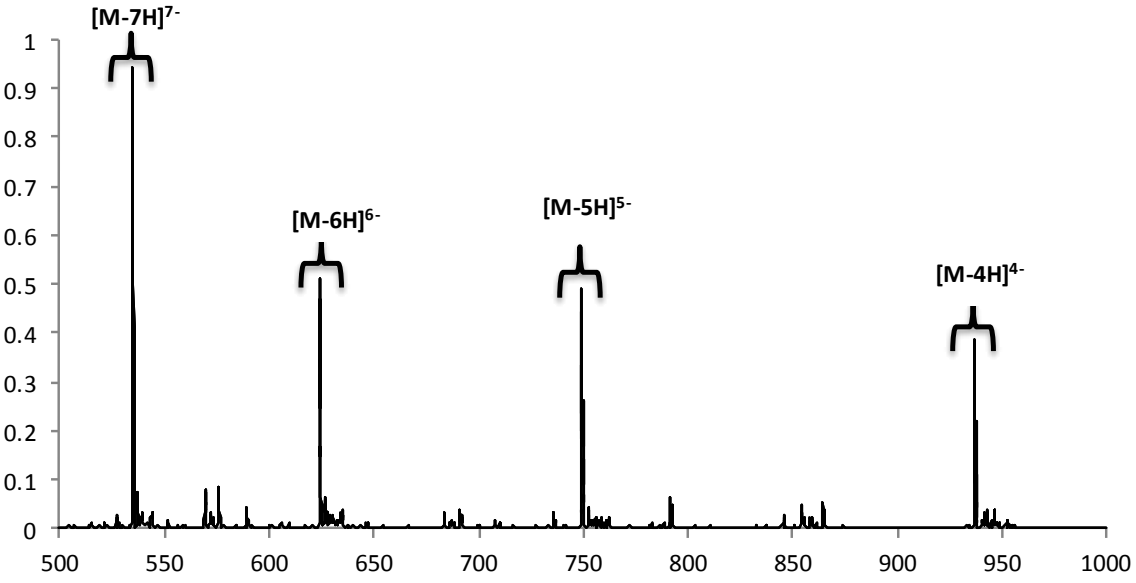


Figure S4: ESI$^-$ Spectrum of *Nar*I(N=C, X=^{FBP}G).

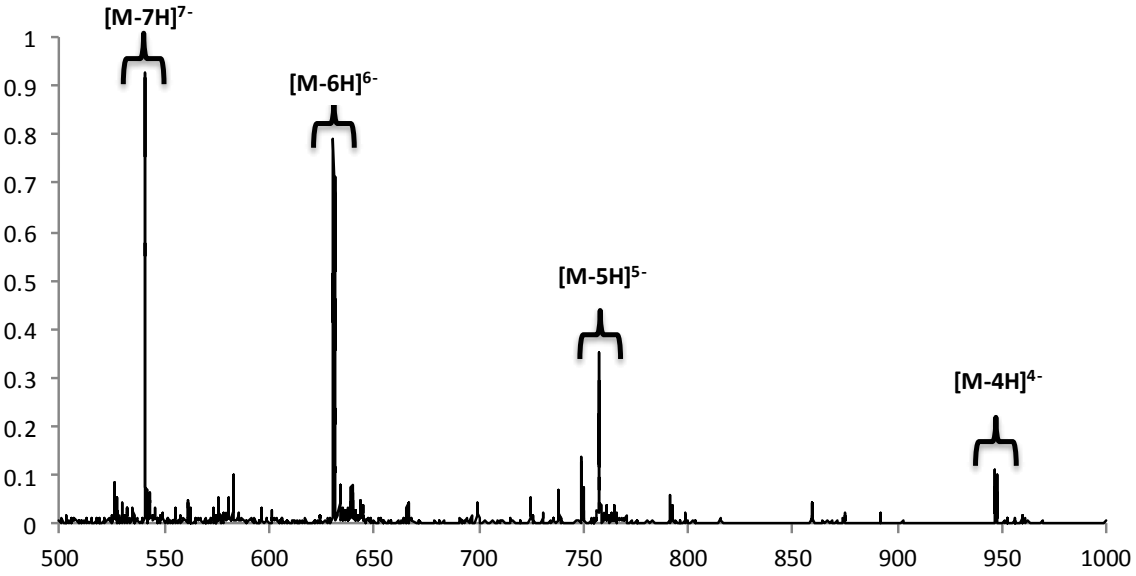


Figure S5: ESI$^-$ Spectrum of *Nar*I(N=G, X=^{FBP}G).

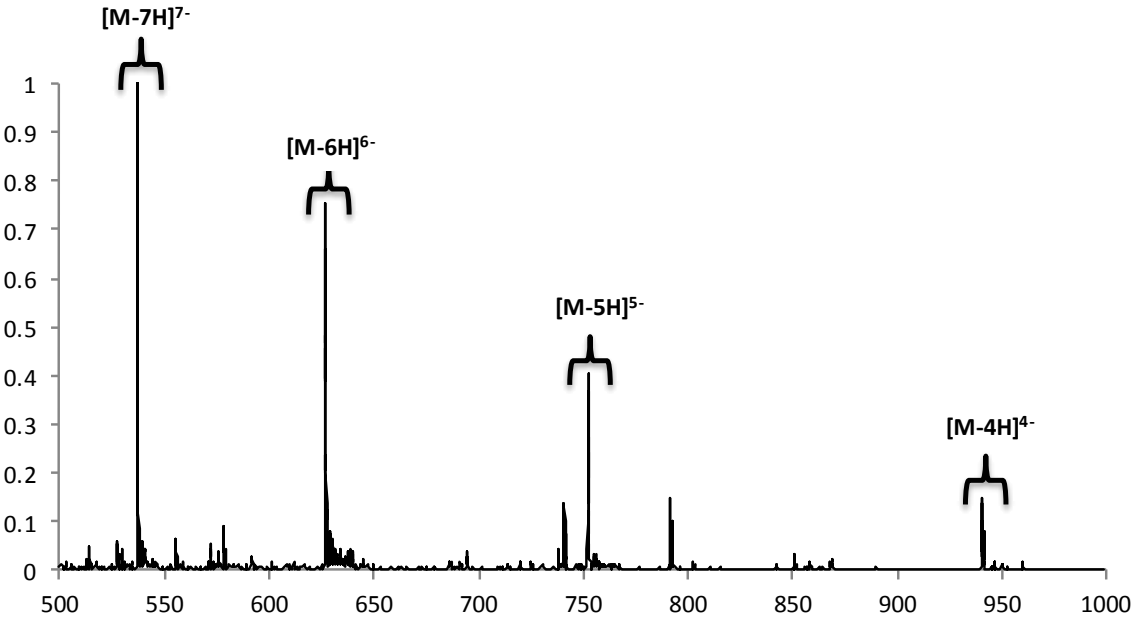


Figure S6: ESI$^-$ Spectrum of *Nar*I(N=T, X=^{FBP}G).

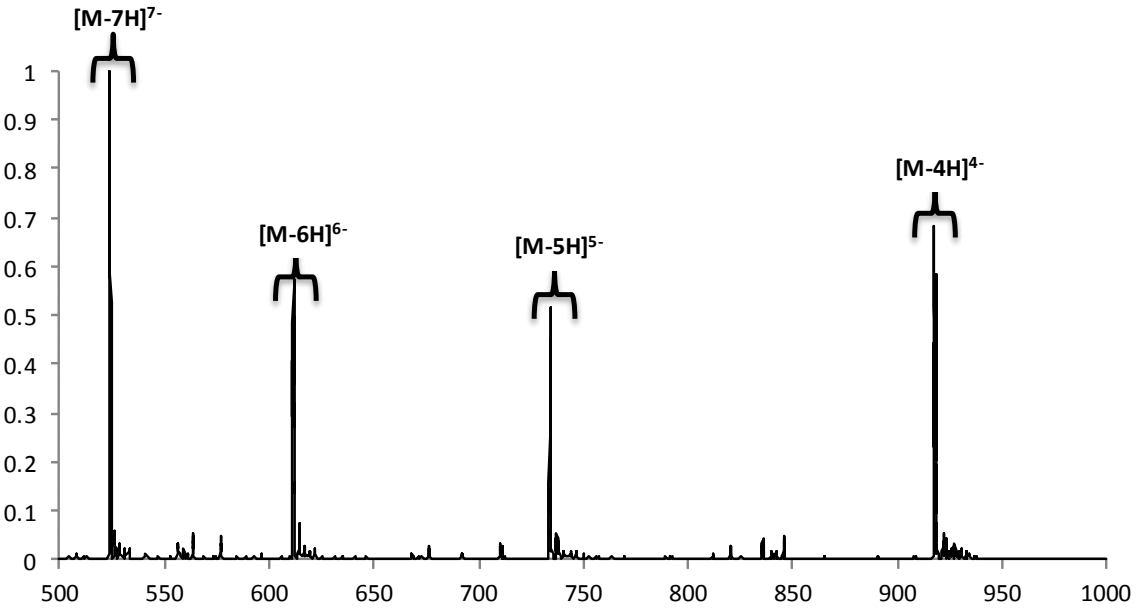


Figure S7: ESI$^-$ Spectrum of *Nar*I(N=C, X=^{FPh}G)

Table S7. ESI-MS Analysis of C8-Aryl-dG Modified *Nar*I (N,X) oligonucleotides.

Oligonucleotide	Formula	Calc Mass	Exptl m/z		Exptl Mass
*Nar*I(N=A, X=^{FBP}G)	$C_{127}H_{154}FN_{44}O_{70}P_{11}$	3774.64	$[M-7H]^{7-}$	538.28	3774.96
			$[M-6H]^{6-}$	628.16	3774.96
			$[M-5H]^{5-}$	753.97	3774.96
			$[M-4H]^{4-}$	942.69	3774.76
*Nar*I(N=C, X=^{FBP}G)	$C_{126}H_{154}FN_{42}O_{71}P_{11}$	3750.68	$[M-7H]^{7-}$	534.88	3751.16
			$[M-6H]^{6-}$	624.16	3750.96
			$[M-5H]^{5-}$	749.22	3751.11
			$[M-4H]^{4-}$	936.75	3751.00
*Nar*I(N=G, X=^{FBP}G)	$C_{127}H_{154}FN_{44}O_{71}P_{11}$	3790.96	$[M-7H]^{7-}$	540.56	3790.92
			$[M-6H]^{6-}$	630.89	3791.34
			$[M-5H]^{5-}$	757.19	3789.95
			$[M-4H]^{4-}$	946.47	3789.88
*Nar*I(N=T, X=^{FBP}G)	$C_{127}H_{155}FN_{41}O_{72}P_{11}$	3767.56	$[M-7H]^{7-}$	537.00	3766.00
			$[M-6H]^{6-}$	626.66	3765.96
			$[M-5H]^{5-}$	752.19	3765.95
			$[M-4H]^{4-}$	940.44	3765.76
*Nar*I(N=C, X=^{FPh}G)	$C_{120}H_{150}FN_{42}O_{71}P_{11}$	3674.65	$[M-7H]^{7-}$	524.00	3675.00
			$[M-6H]^{6-}$	611.5	3675.00
			$[M-5H]^{5-}$	734.00	3574.00
			$[M-4H]^{4-}$	917.24	3674.77

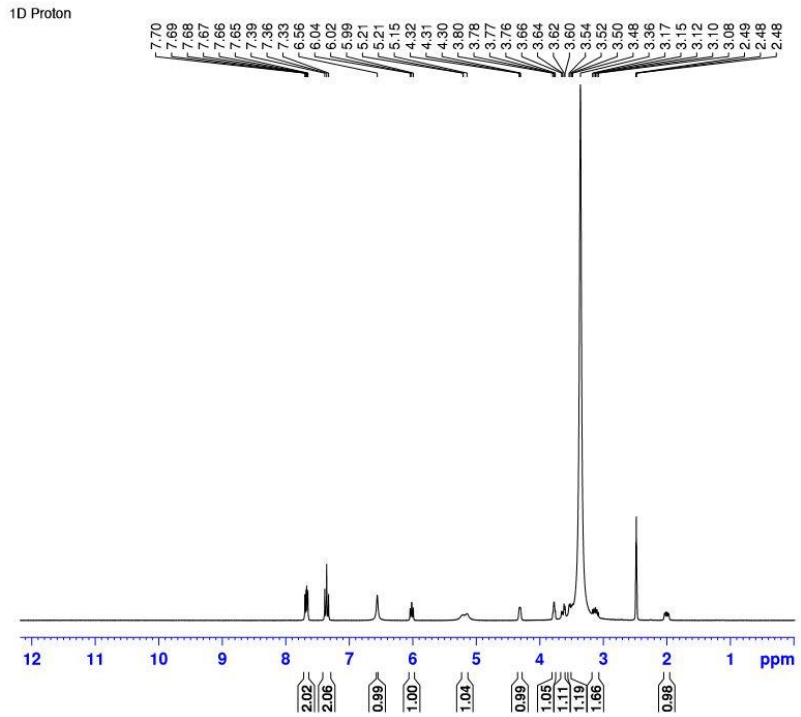


Figure S8: ^{1}H NMR of 2a in DMSO-d_6.

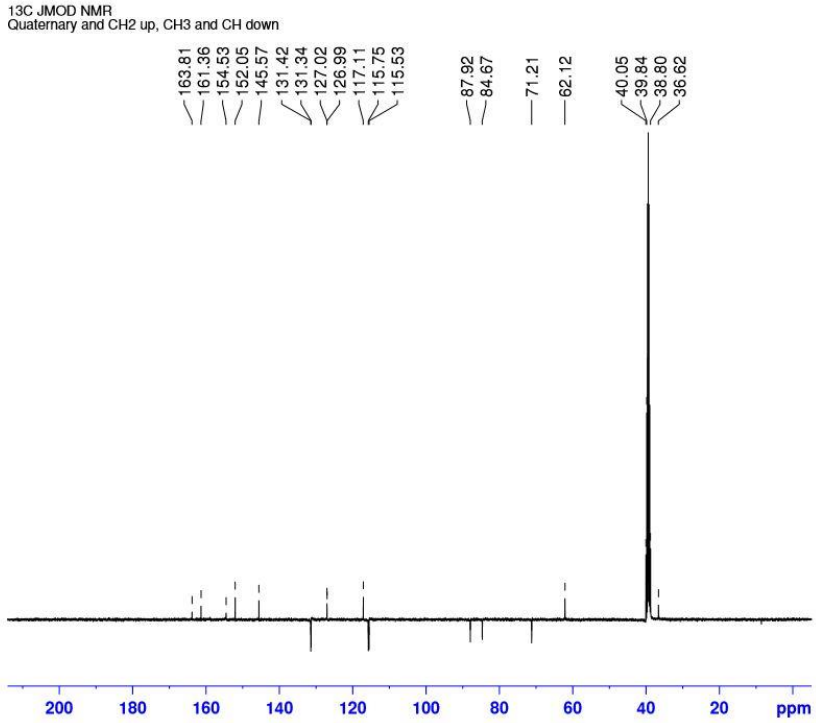


Figure S9: ^{13}C NMR of 2a in DMSO-d_6.

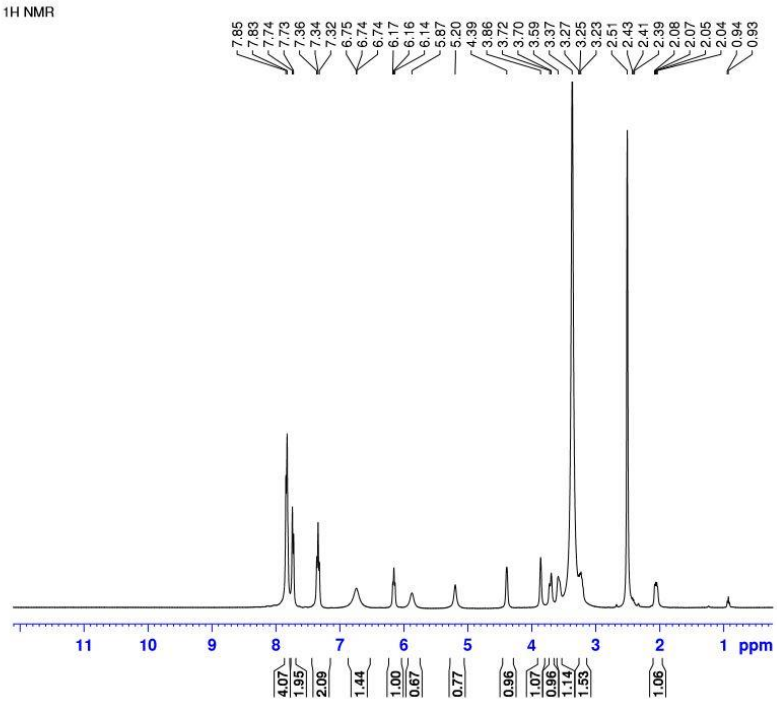


Figure S10: ^{1}H NMR of 2b in DMSO-d_6.

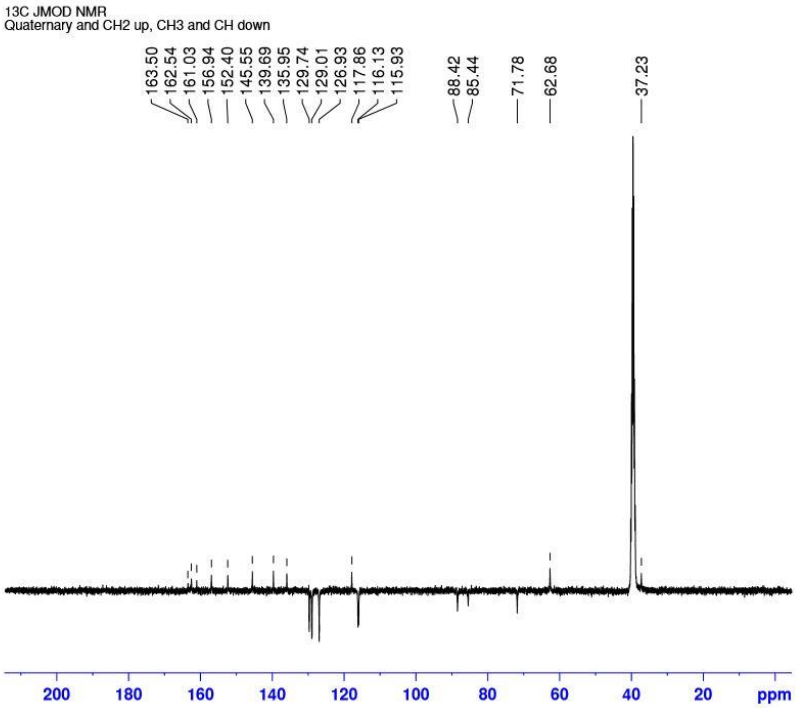


Figure S11: ^{13}C NMR of 2b in DMSO-d_6.

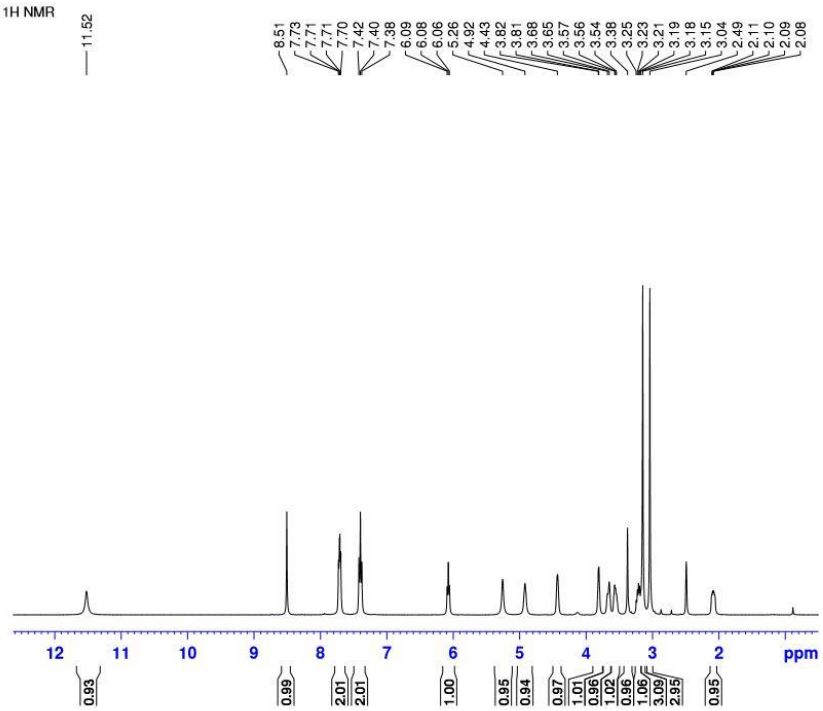


Figure S12: ^{1}H NMR of 3a in DMSO-d_6.

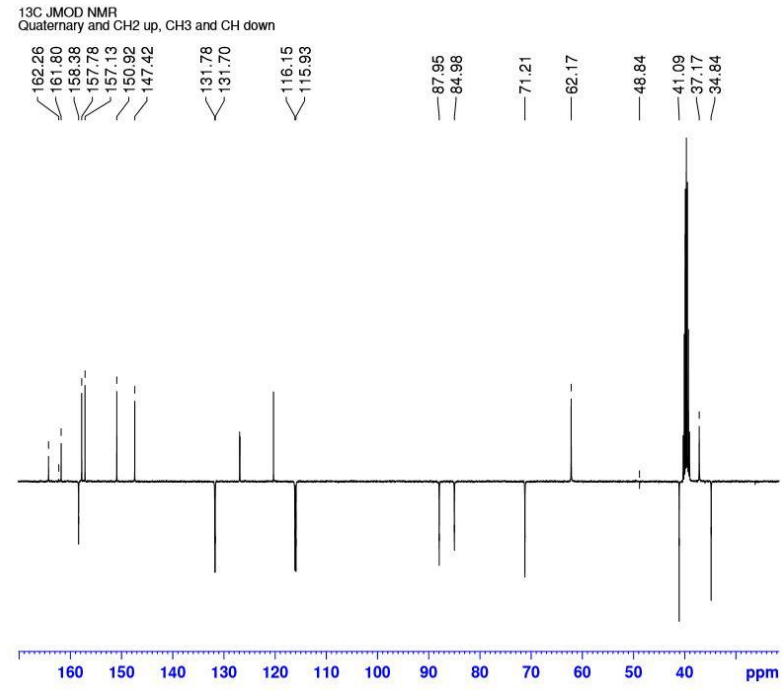


Figure S13: ^{13}C NMR of 3a in DMSO-d_6.

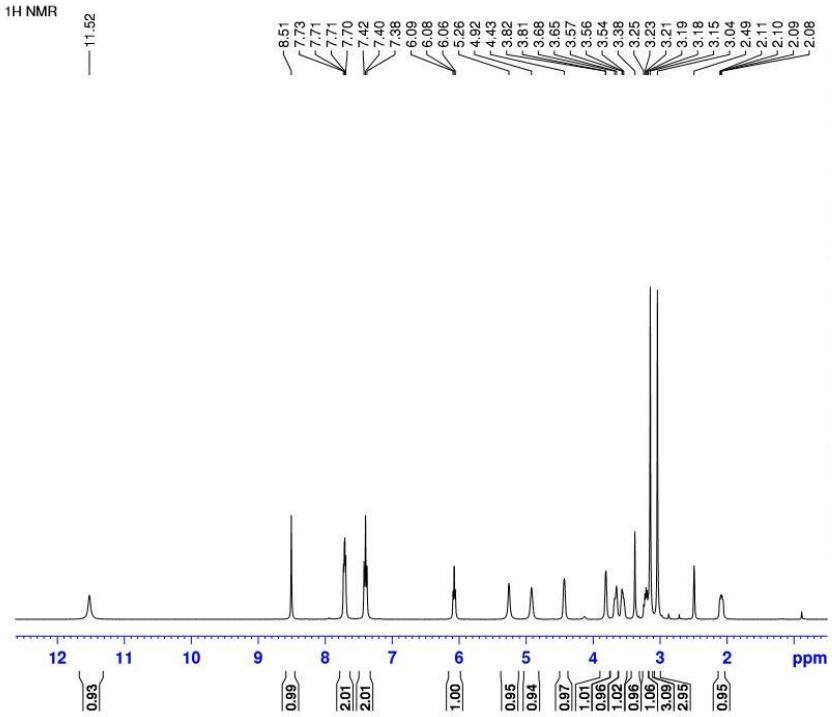


Figure S14: ^{1}H NMR of 3b in DMSO-d_6.

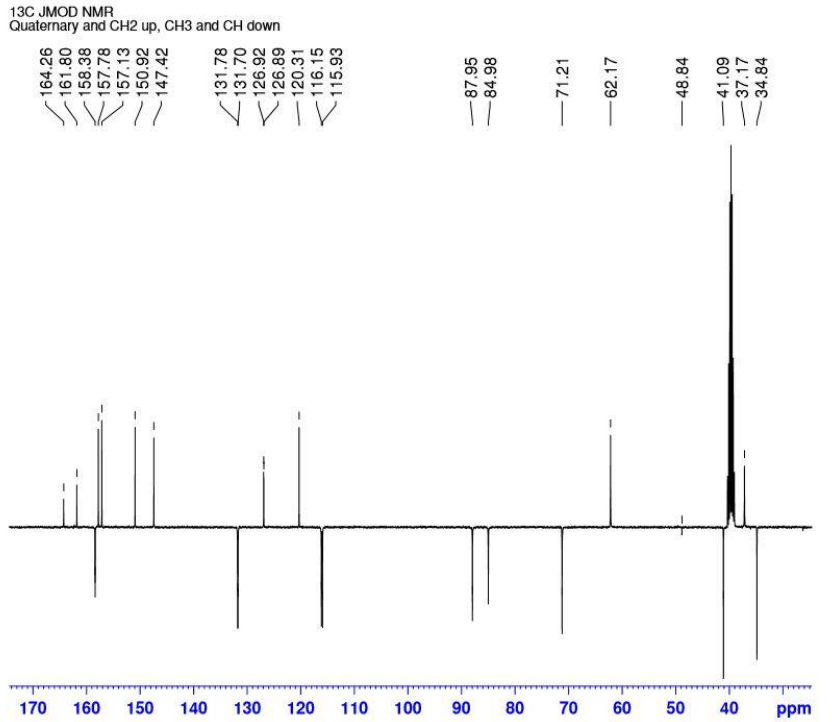


Figure S15: ^{13}C NMR of 3b in DMSO-d_6.

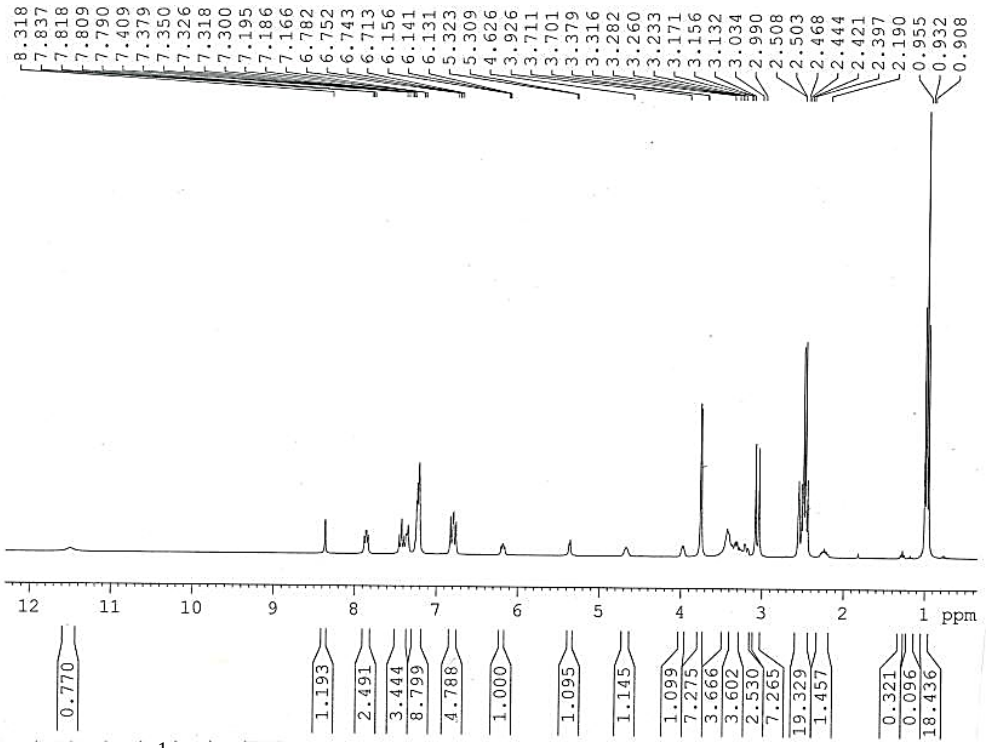


Figure S16: ^{1}H NMR of 4a in $CDCl_3$.

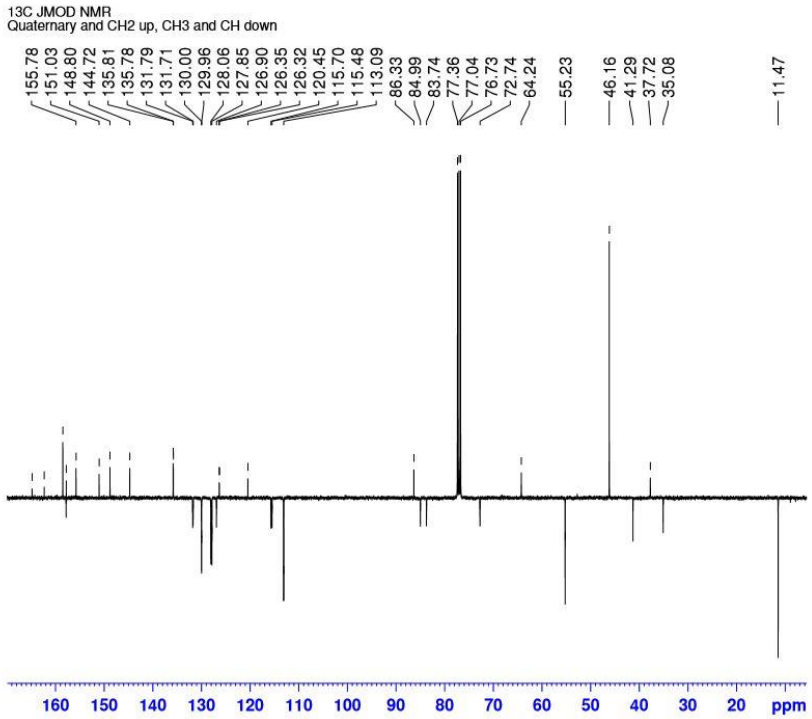


Figure S17: ^{13}C NMR of 4a in $CDCl_3$.

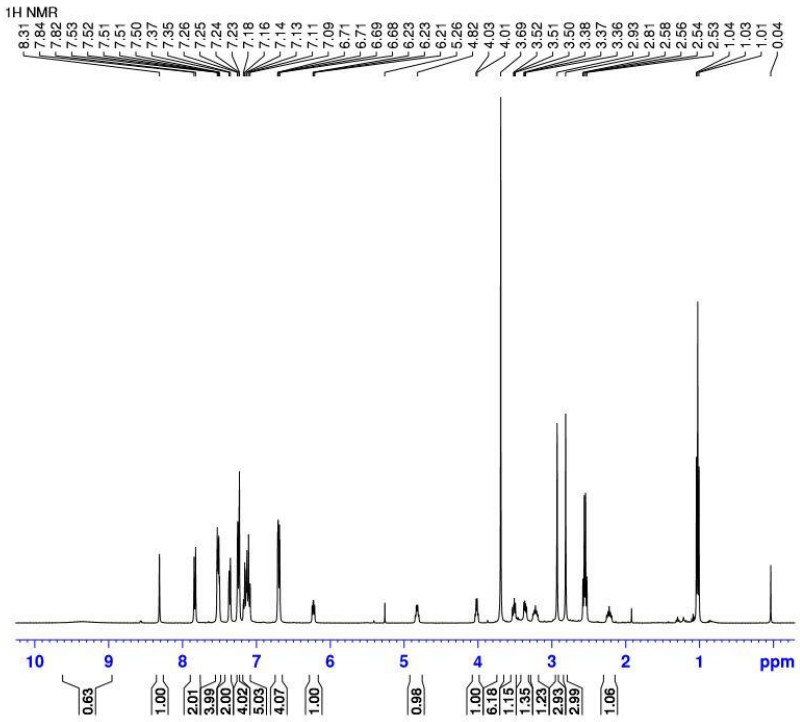


Figure S18: ^{1}H NMR of 4b in $CDCl_3$.

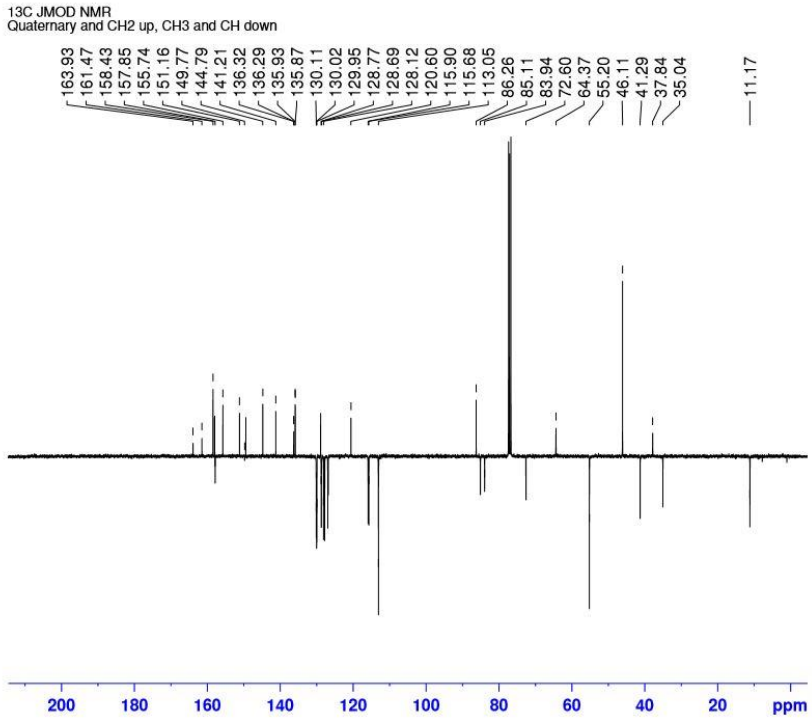


Figure S19: ^{13}C NMR of 4b in $CDCl_3$.

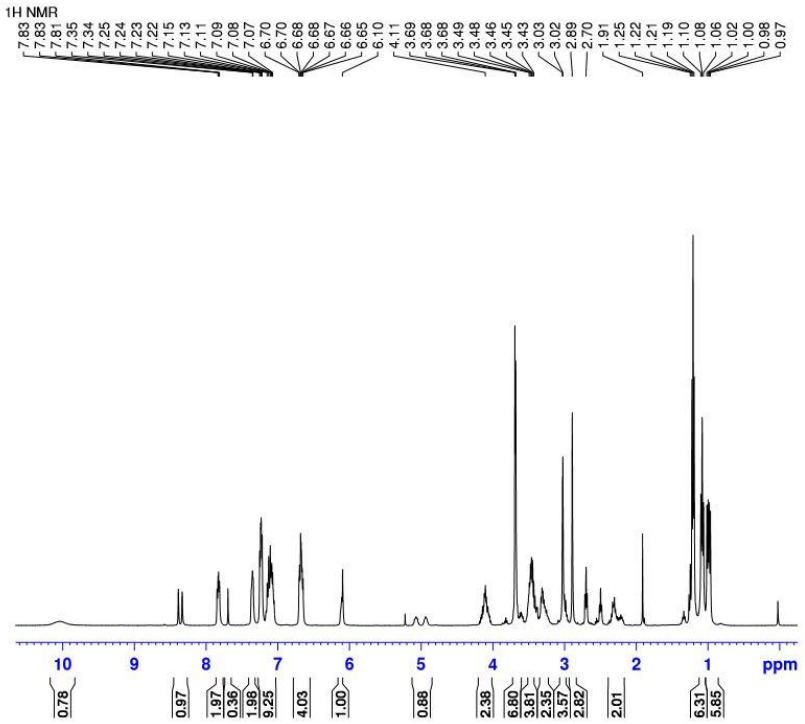


Figure S20: ^{1}H NMR of 5a in $CDCl_3$.

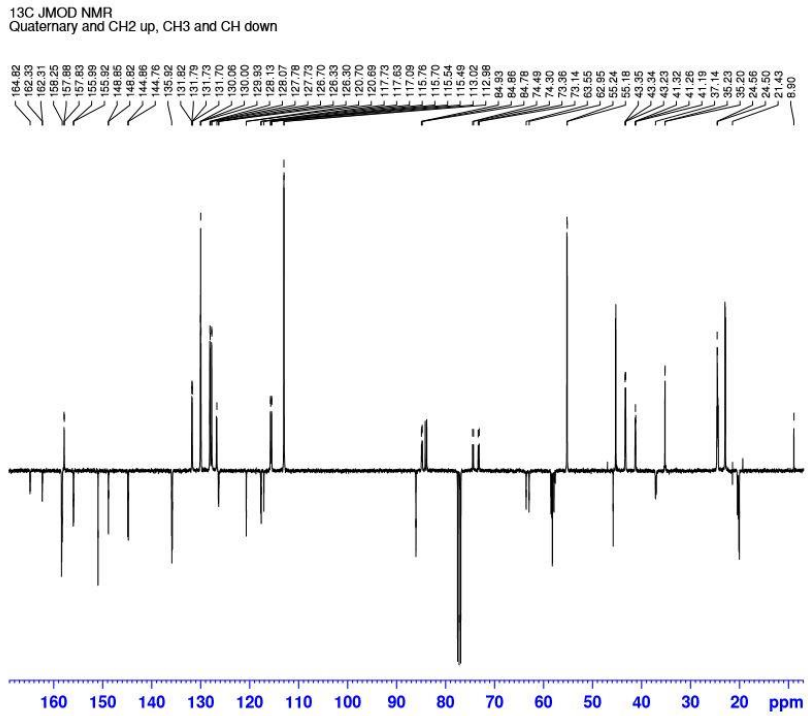


Figure S21: ^{13}C NMR of 5a in $CDCl_3$.

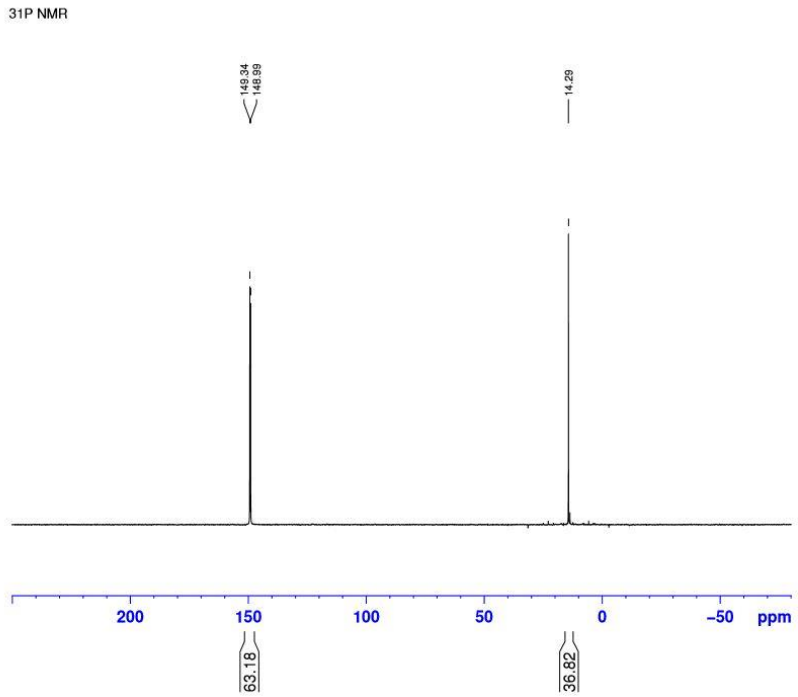


Figure S22: ^{31}P NMR of 5a in $CDCl_3$.

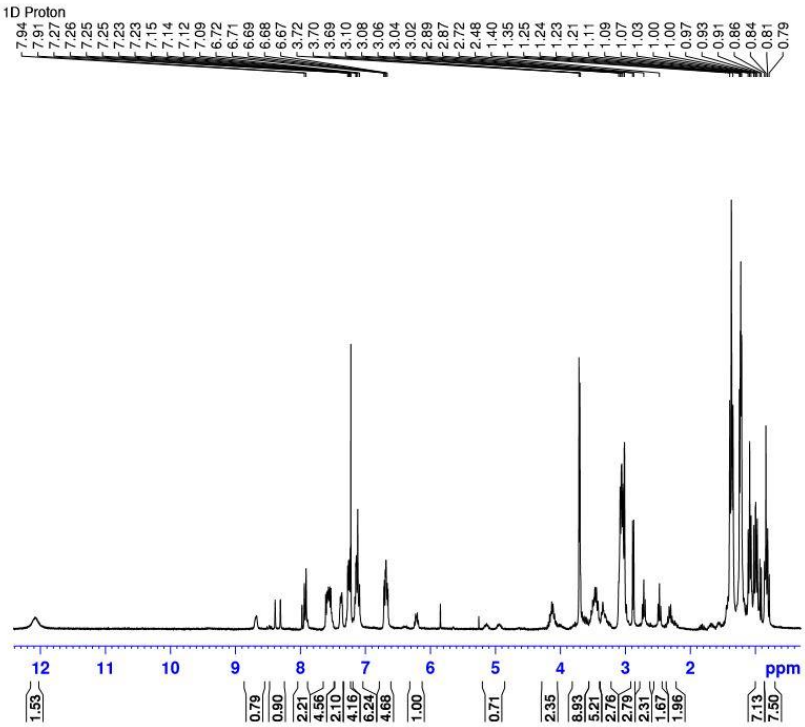


Figure S23: ^{1}H NMR of 5b in $CDCl_3$.

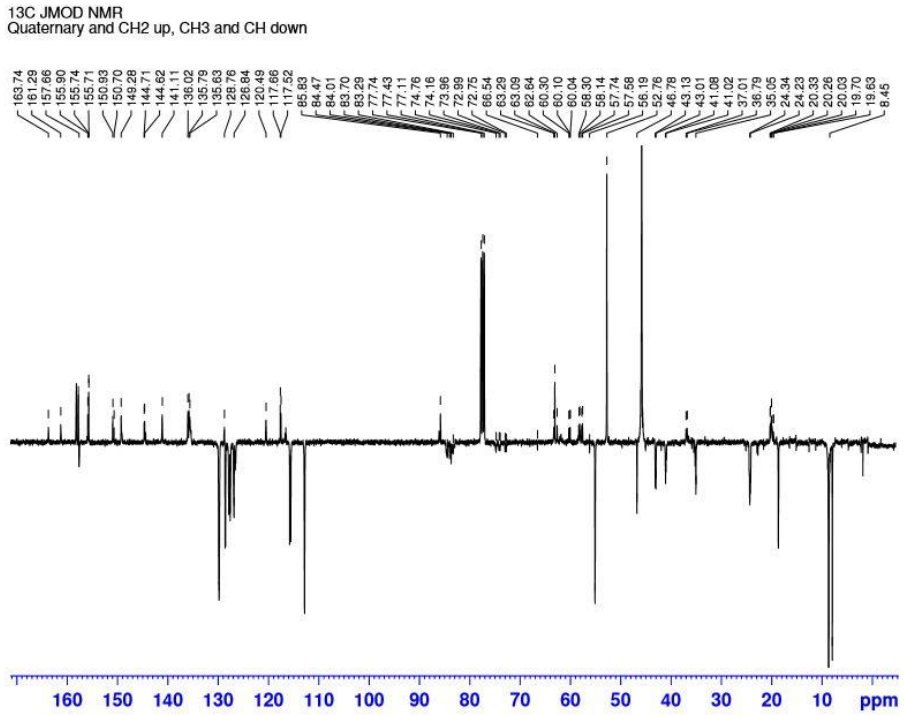


Figure S24: ^{13}C NMR of 5b in $CDCl_3$.

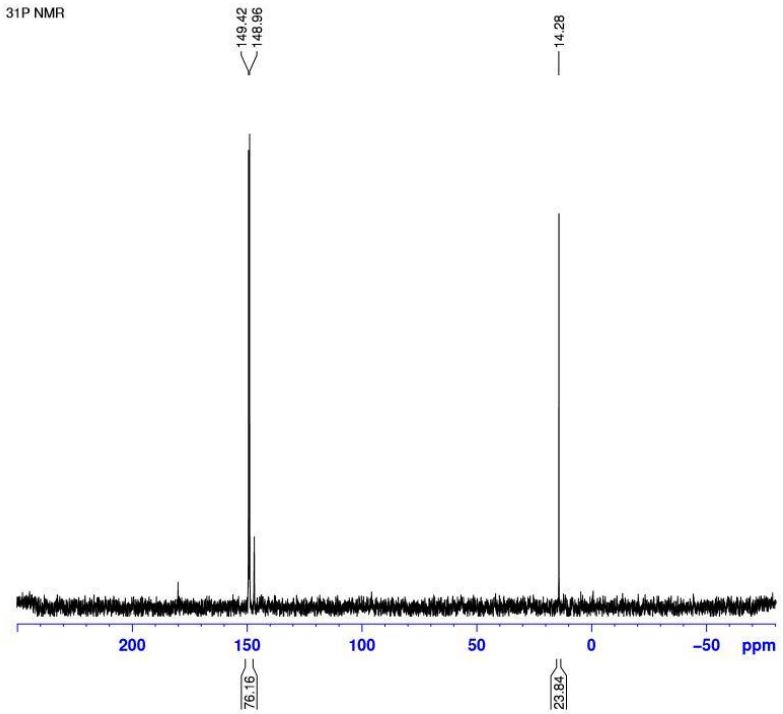


Figure S25: ^{31}P NMR of 5b in $CDCl_3$.

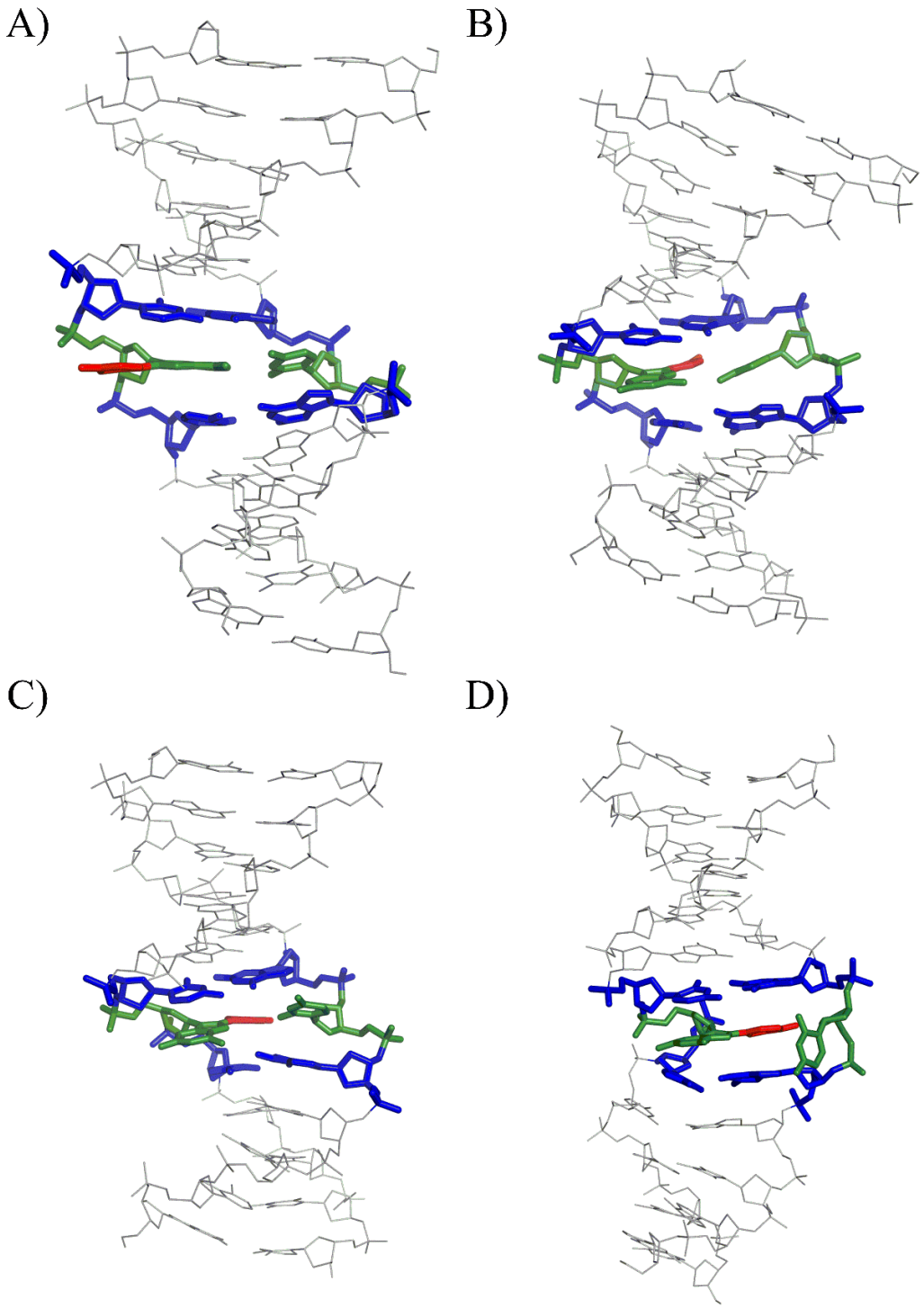


Figure S26: The A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FPh-dG adducted *Nar*I (N = C). The damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

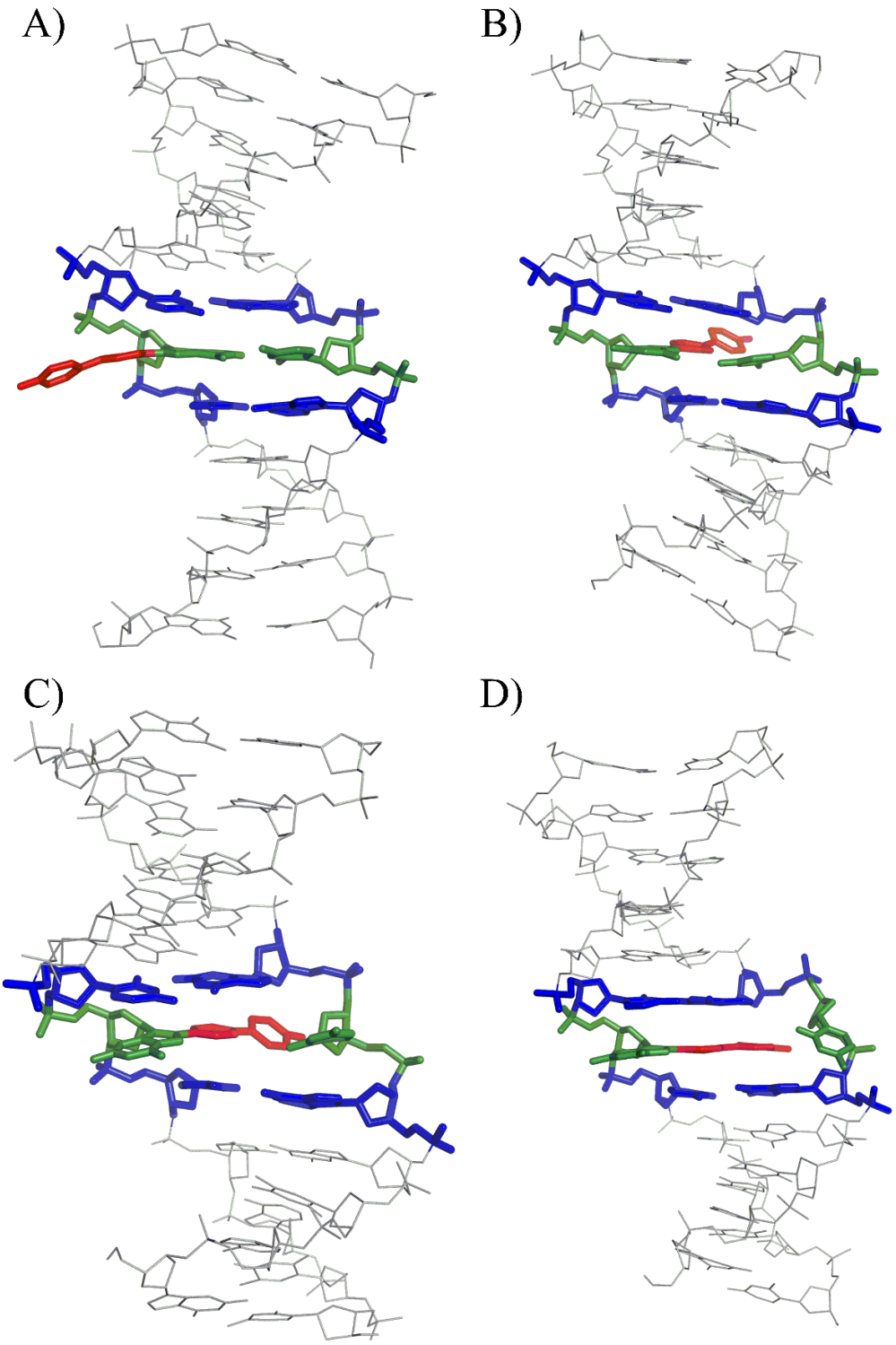


Figure S27: The A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FBP-dG adducted *Nar*I (N = C). The damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

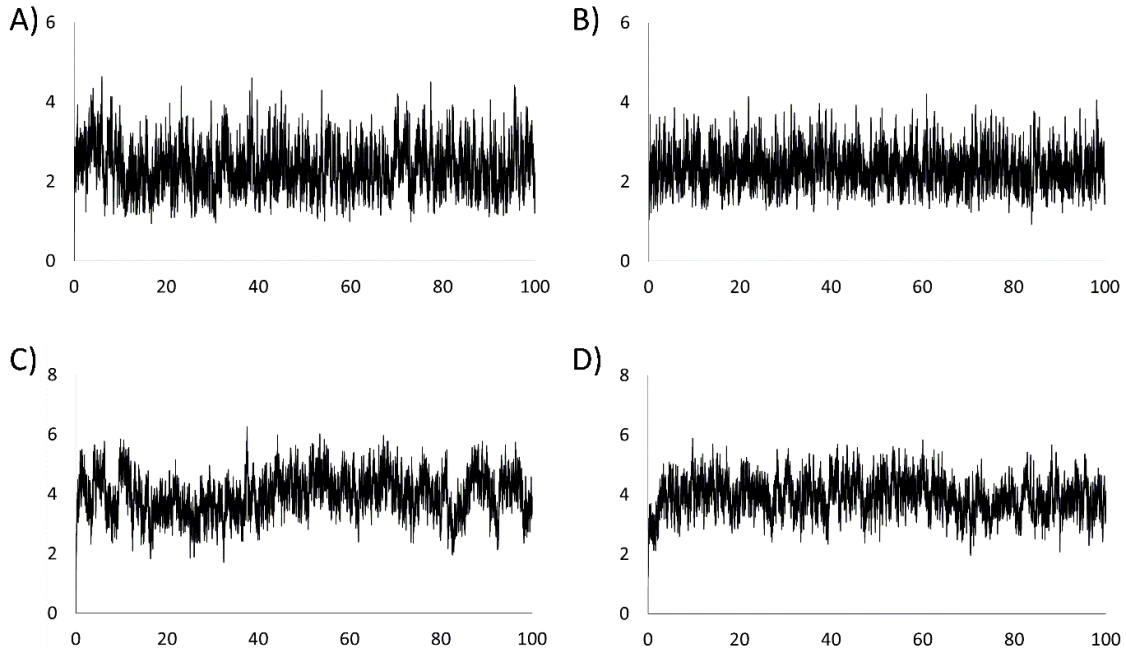


Figure S28: The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FPh-dG adducted fully paired *Nar*I (N = C).

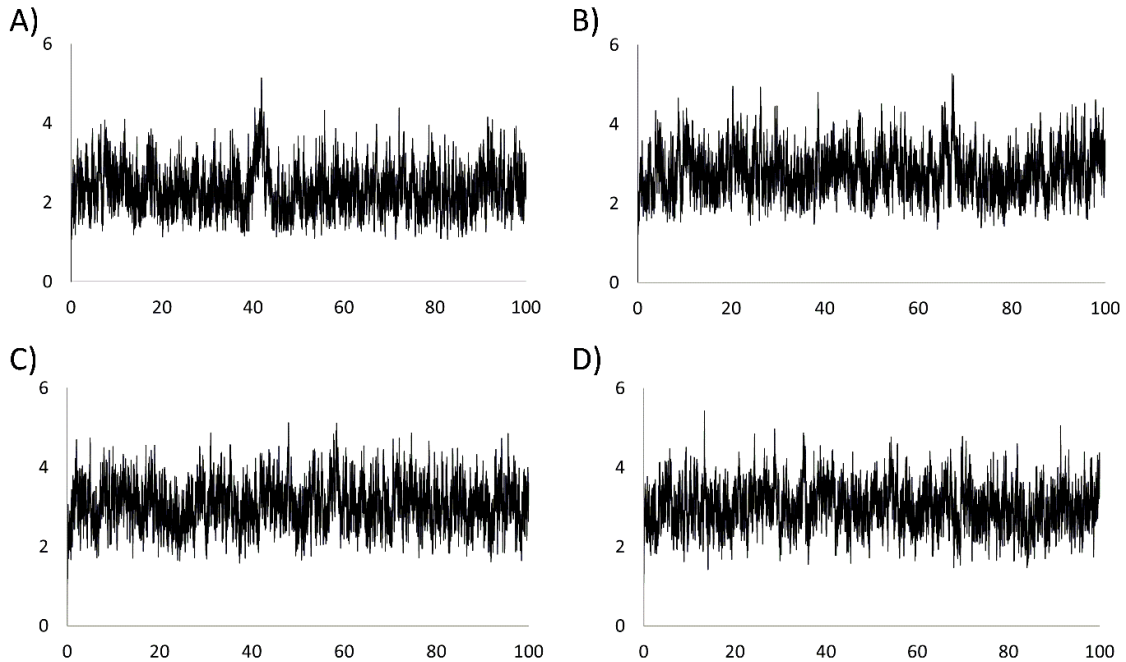


Figure S29: The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FBP-dG adducted fully paired *Nar*I (N = C).

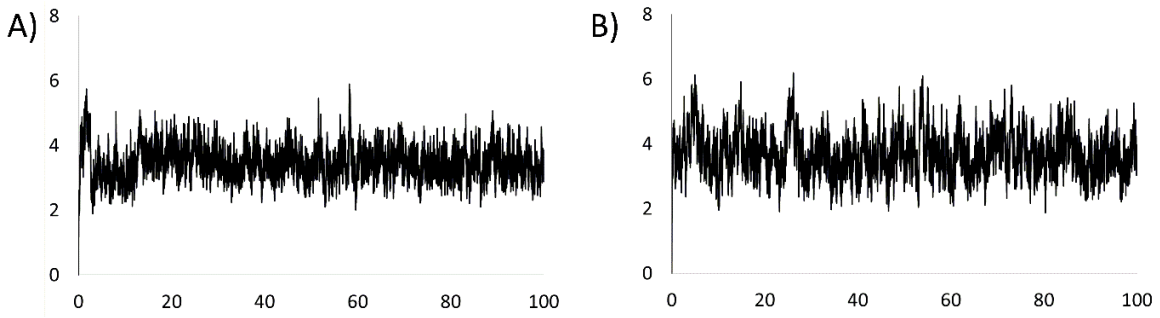


Figure S30: The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) *anti*, and B) *syn* conformations of FBP-dG adducted SMI *Nar*I (N = C).

Molecular Dynamics Simulations. Both the F-phenyl-dG (^{FPh}dG) and F-biphenyl-dG (^{FBP}dG) adducts were studied in the 5′–CTCGGCXCCATC–3′ 12-mer containing the *Nar*I sequence (underlined), with X being the fluorine-labeled adduct. The ^{FBP}dG adduct was also studied in a two-base bulge helix using the same primary 12-mer strand paired with the 5′–GAGCCGGTAG–3′ 10-mer strand as done previously for other DNA adducts.[1] The natural nucleotides were treated with the amber14SB force field, while GAFF parameters were used to supplement the amber14SB force field for the adducts. Antechamber of the AMBER toolset was used to assign atom types, while R.E.D. version III.4[3] was used to calculate the partial charges of the damaged residues based on B3LYP/6-31G(d) optimized structures of the *anti* and *syn* conformers. Specifically, the R.E.D. program uses the initial structures to calculate the molecular electrostatic potential (MEP) surfaces at the HF/6-31G* level of theory, and then fits the atomic charges to the MEP surfaces according to a restrained electrostatic potential. Importantly, the charges generated using the R.E.D. program are fully compatible with the Amber force field.[3] Although there are not suitable experimental quantities to test our force field against for these systems since the only experimental data available is presented in this work, the agreement between NMR and MD data provides confidence that our approach is reasonable. Molecular dynamics (MD) simulations were run on the adducted DNA duplexes using the PMEMD module of AMBER 14.

For the fully paired helices, initial structures were generated that correspond to the B-type (major-groove), W-type (wedge) and S-type (base-displaced intercalated) conformations discussed in the literature for N-linked C8-dG adducts. Specifically, the B-type conformer places the bulky moiety within the major groove of the helix, while the W-type conformer places the bulky moiety in the minor groove. In contrast, the S conformation intercalates the bulky moiety within the helix and displaces the opposing base into the major groove. These conformers were generated by modifying representative structures of previously studied C8-aryl-dG adducts[4] using GaussView 5.0. For the B and W conformations, the χ dihedral angle (χ = $\angle$(O4′C1′N9C4)) for each conformation initially considered was selected based on the value in the modified representative structure, while maintaining interactions between the damaged G and opposing C. For the S conformer, the χ dihedral was selected to keep the bulky group planar with respect to the flanking bases since the adduct has limited flexibility due to direct attachment of the aromatic group to C8 of dG. From these starting conformations, multiple MD simulations were performed for shorter timescales (20 ns) with bulky moiety orientations that differ in the

relative orientation with respect to the damaged nucleobase (θ = ∠(N9C8C10C11), Figure S2). Specifically, θ was initially set to 0°, 30° or 60° in order to enhance sampling of possible bulky moiety orientations, while the dihedral angle between the two rings of the biphenyl system (ξ = ∠(C12C3C16C17), Figure S2) was set equal to 0°, which was justified based on the sampling of many values over the production runs (Figure 5). It is worth noting that because of the lack of flexibility in the C–C linker, the bulky group cannot be directed in the 5′ or 3′ direction within a groove as reported in the literature for N-linked C8-dG adducts. In cases where these preliminary simulations led to new conformations, the corresponding representative structures were then used as initial coordinates for further simulations in an iterative process. Free energies estimated from the molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) approach were then used to determine the most favorable structures. Through this cycle, the intercalated conformation (I) was isolated where the bulky moiety and opposing C are both intercalated within the DNA duplex. The lowest energy structures obtained from the iterative phase sampling simulations were then used for the 100 ns production MD simulations described below. Similarly, preliminary conformations for the bulged structures were built based on the X-ray structure of the AF-dG adduct opposite a –2 deletion site (PDB ID: 1AX6), with GaussView 5.0 used to modify the lesion, generate *anti* lesion orientations and consider several input orientations with differing values of key dihedral angles.

Using the TLEAP module of AMBER, the DNA systems were solvated within a 8 Å octahedral TIP3P water box, such that there was a minimum of 8 Å between the DNA and the edge of the water box in any direction. Furthermore, the systems were neutralized with 22 or 20 sodium counter-ions for the full and deletion helices respectively. A cutoff distance of 8 Å was used for Leonard-Jones non-bonded interactions, with the particle-mesh Ewald method used to describe Coulombic interactions. The periodic boundary condition was used for all simulation steps. Minimization was performed using 500 steps of the steepest descent (SD) algorithm followed by 500 steps of the conjugate gradient (CG) minimization algorithm, while restraining the DNA with a 500 kcal mol^{-1} $Å^{-2}$ force constant. The DNA was subsequently optimized without constraints using 1000 steps of SD minimization and 1500 steps of CG minimization. Following minimization, each system was heated to 300 K over 20 ps under conditions of constant volume using the Langevin thermostat algorithm, while imposing a weak constraint on DNA (10 kcal mol^{-1} $Å^{-2}$). Finally, 100 ns production simulations were performed at constant temperature (300 K) and pressure (1 atm) using the Langevin and Berendsen algorithms, respectively, and the hydrogen atoms were constrained using SHAKE. A 2 fs time step was used for the production runs, with the trajectory sampled every 1250 steps. During the production phase, 25 kcal mol^{-1} $Å^{-2}$ distance restraints were used on the Watson-Crick hydrogen bonds in the terminal base pairs to prevent unravelling that is a common artifact of MD simulations on DNA.[5]

Structural analysis was carried out on each 100 ns trajectory using the PTRAJ and CPPTRAJ modules of AMBER 14. The stability of the systems was confirmed using the backbone root-mean-square deviation (RMSD) compared to the first frame of the production run (Table S4, Figures S28 and S29). Representative structures were obtained by clustering the MD simulations with respect to the relative orientation of the damaged nucleobase and 2′-deoxyribose (χ), and the bulky moiety and damaged nucleobase (θ). The reported structures occur over the entire simulation trajectory, with the exception of the ^{FPh}dG adducted DNA wedge and base-displaced intercalated conformations. The wedge structure interconverts between the reported conformer (46.6% of the simulation time) and a slightly higher energy wedge conformer (32.1 kJ/mol) that deviates in χ by approximately 40°, while the base-displaced intercalated

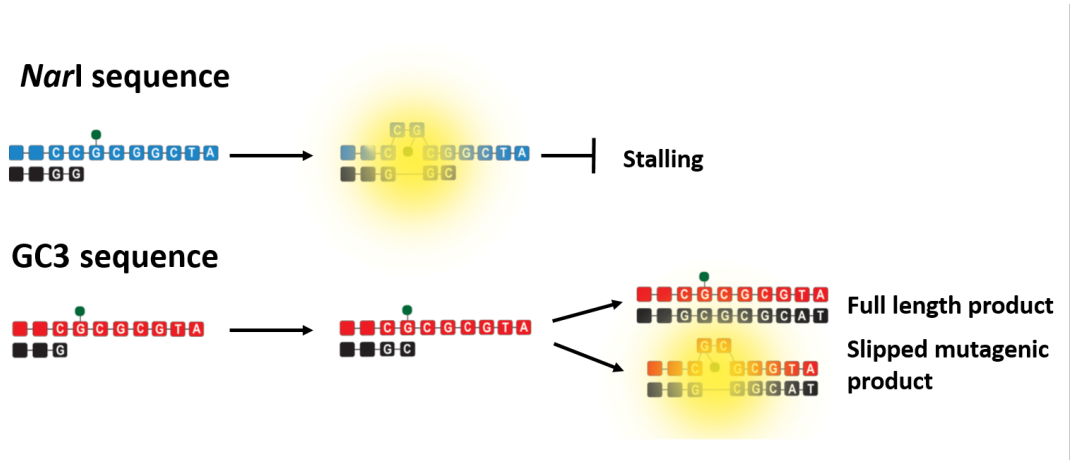
conformer dominates the trajectory (87.4% of the simulation time) and interconverts with a wedge conformation. A total of 400 frames of the water and sodium free trajectories corresponding to the clustered simulations were sampled to calculate the enthalpy according to the MMPBSA method, and the entropy was estimated using the normal mode analysis method in AMBER.

References

1. Sproviero, M.; Verwey, A. M.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D., Structural and Biochemical Impact of C8-Aryl-guanine Adducts within the NarI Recognition DNA Sequence: Influence of Aryl Ring Size on Targeted and Semi-Targeted Mutagenicity. *Nucleic Acids Res.* **2014,** *42* (21), 13405-13421.
2. Sproviero, M.; Verwey, A. M.; Witham, A. A.; Manderville, R. A.; Sharma, P.; Wetmore, S. D., Enhancing Bulge Stabilization through Linear Extension of C8-Aryl-Guanine Adducts to Promote Polymerase Blockage or Strand Realignment to Produce a C:C Mismatch. *Chem. Res. Toxicol.* **2015,** *28* (8), 1647-1658.
3. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010,** *12* (28), 7821-7839.
4. Sharma, P.; Majdi Yazdi, M.; Merriman, A.; Manderville, R. A.; Wetmore, S. D., Influence of the Linkage Type and Functional Groups in the Carcinogenic Moiety on the Conformational Preferences of Damaged DNA: Structural and Energetic Characterization of Carbon- and Oxygen-Linked C8-Phenolic-Guanine Adducts. *Chem. Res. Toxicol.* **2015,** *28* (4), 782-796.
5. Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P., Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theor. Comp.* **2014,** *10* (8), 3177-3189.

Chapter 3

Fluorescence profiles of synthetically modified DNA duplexes reveal how sequence-dependent base conformation heterogeneity gives rise to frameshift mutations



Florence D. Berger, Richard A. Manderville, and Shana J. Sturla

Manuscript in preparation

F.D.B. designed the study, synthesized compounds and oligonucleotides, characterized them and performed the thermal melting, fluorescence and primer extension studies. R.A.M. and S.J.S. conceived the research. F.D.B., R.A.M., and S.J.S. interpreted data and wrote the manuscript.

Fluorescence profiles of synthetically modified DNA duplexes reveal how sequence-dependent base conformation heterogeneity gives rise to frameshift mutations

Florence D. Berger,[a] Richard A. Manderville,[b] and Shana J. Sturla*,[a]

[a] Department of Health Science and Technology, Institute of Food, Nutrition, and Health, ETH Zurich, 8092 Zürich, Switzerland
[b] Department of Chemistry and Toxicology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Abstract: Bulky C8-dG adducts, which arise through aromatic chemical carcinogen might disturb the DNA duplex conformation and might result in DNA mutations. The DNA template slippes due to a C8-dG induced bulge out structure and leads into a -1 or -2 base deletion. The fluorescent C-linked 4-fluorobiphenyl-dG (FBP-dG) is an excellent probe to study the sequence dependent frameshift mutation. FBP-dG stabilizes a duplex construct that mimics a slipped mutagenic duplex in the 12mer *Nar*I sequence, furthermore it is emissive when it forms the slipped mutagenic duplex. Herein presented is a study of synthetically modified DNA duplex to characterize a sequence dependent base conformational heterogeneity of a bulky C8-dG adduct. We studied the mechanism of frameshift mutation in two different local sequence contexts. Two sequences containing FBP-dG are subjected full length extension using the model polymerase Klenow framgemnt exo$^-$ (Kf$^-$). A dramatic difference of polymerase bypass is observed for the two sequences. In the normal *Nar*I sequence persistence of the n+1 band is observed, indicating a strong stalling of Kf$^-$. Whereas in a sequence repeating 3 CGs (CG*CGCG), Kf$^-$ can bypass FBP-dG, resulting in a slipped mutagenic product and a full length product. To understand the sequence dependent mechanism of frameshift mutations, model DNA duplexes are constructed, annealing FBP-dG containing templates to primers of incrementally lengths. Upon creating the bulge-out structure, the fluorescence intensity increases of FBP-dG. FBP-dG induces the bulge out structure in the *Nar*I sequence upon inserting one base opposite the damage. On the other hand in the CG3 sequence the bulge out structure is generated upon inserting two bases opposite FBP-dG. This study demonstrates the use of a simple fluorescence-based chemical approach to understand sequence dependent frameshift mutation.

3.1 Introduction

Many known human carcinogens give rise to bulky DNA adducts *via* binding at the C8 position of guanine in DNA.[1–5] Persistence of these adducts during DNA replication can give rise to mutations that drive carcinogenesis. Bulky C8-dG adducts influence the DNA duplex conformation, which might result in a bulge out structure. This bulge out structure slippes the template DNA strand and a frameshift mutation might occur. Frameshift mutations ultimately result in a -1 or -2 base deletion, which cause a truncated protein with diminished or complete loss of function. However, we do not understand the structure-function relationship of bulky C8-adducts and their role in frameshift mutations.

Bulky C8-guanine alkylation adducts have been identified to arise from arylamines, arylhydrazines, heterocyclic amines, and polycyclic aromatic hydrocarbons. Depending on the aryl group they might form an N- or a C-linked C8-dG adduct. N-linked adducts, like 2-acetylaminofluorene-dG (AAF-dG), aminofluorene-dG (AF-dG) and aminobiphenyl-dG (ABP), are models from which significant structure-function relationships of DNA adducts with mutation have been elucidated. Despite their structural similarity, AF-dG is more flexible than the AAF-dG, because it lacks the acetyl, and adopts a regular B-form DNA Structure. AAF-dG, on the other hand, is rigid and more likely to adopt a stacked conformation.[6] ABP-dG finally, lacks a methylene bridge between the aryl rings, therefore is more flexible, and adopts a B-form conformation.[7–9] Analogous C8-dG adducts may also arise from phenolic chemicals, like the food mutagen Ochratoxin A (OTA),[10] or metabolites of benzo[a]pyrene (B[a]P),[11] in

which the resulting linkage between the chemical and the guanine can be either *via* carbon or oxygen, rather than nitrogen as in the case of AAF and AF. It is poorly understood how the O-or C-linked adduct influence the DNA conformation and how the conformation of the adduct relates to the mutational outcome.

AF-dG results in point mutations, AAF-dG is known to induce frameshift mutations[12] and ABP-dG, is expected to be less mutagenic as AAF-dG. The C-linked member, OTA result in deletion mutations[13,14] and double strand breaks (DBS),[5] and B[a]P leads into G to T and G to C transversion mutations.[4] The aforementioned examples illustrate that the mutagenic potential of C8-dG adducts depends on several factors: size of the adduct, linker type and finally sequence context.

In sequences containing short tandem repeats, i.e. runs of bases such as GGGGG or repeats of dinucleotides like CGCGCG, or microsatellites, spontaneous frameshift mutations tend to cluster,[15] and microsatellite instability in combination with mismatch repair deficiency is associated with cancer.[16–18] A high rate of frameshift mutations are observed to arise from DNA adducts in the *Nar*I recognition sequence (5'-$G^1G^2CG^3$CC-3') when G^3 is modified. AAF-dG induces frameshift mutations in *E.coli,*[19,20] especially when placed in the *Nar*I sequence.[19,21–23] N-linked adducts, which induce frameshift mutations within the *Nar*I sequence adopt generally a syn conformation, highlighting the importance of the conformation of the adduct to generate frameshift mutations[23–26] Moreover, AAF-dG in plasmids within 2GC and 3GC sequences induced frameshift mutations five-fold more frequently in the 3GC repeat compared to the 2GC repeat. [27]

The conformation of C8-dG adducts within the DNA helix is an indicator of their potential to induce frameshift mutations. AF-dG exhibits conformational heterogeneity between *anti* and *syn* conformations about the glycosidic bond.[28–31] If the *syn* conformation is unstable and the adduct rotates into an *anti* conformation, error-prone base incorporation opposite is possible.[32] The presence of the acetyl substituent in AAF-dG restricts rotation from *syn* to *anti* conformation, thus blocking polymerases and potentially inducing frameshift mutations.[12,33,34] Similar observations concerning the relationship with adducts conformation and mutagenicity was observed for C-linked C8-aryl-dG adducts.[35,36] Small adducts, for example C8-furan-dG (Fur-dG), preferentially adopt an *anti* orientation about the glycosidic bond, thus preserving the normal DNA conformation. On the other hand larger adducts, such as C8-pyrene-dG (Py-dG) prefer a *syn* orientation, significantly altering DNA conformation.

DNA Polymerase-mediated DNA primer extension studies carried out with DNA sequences containing the *Nar*I sequence 5'-$G^1G^2CG^3$CC-3' suggest two possible mutagenic outcomes that may be rationalized on the basis of conformational changes driven by the glycosidic bond orientation. In a first step, C is inserted opposite C8-aryl-dG adduct. As a result, a 2-base bulge is generated and another C can be inserted opposite the next G. In the case of smaller aryl-dG adducts that prefer the *anti* conformation, the glycosidic angle can rotate from the *syn* to *anti* conformation in the bulged out form. The switch from *syn* to *anti* results in a realignment of the template and primer, creating a C:C mismatch neighboring the damage position. Larger aryl-dG adduct that prefer syn conformation are a larger barrier to realign the strand. Therefore, the slipped mutagenic duplex is stabilized and synthesis continues and eventually result in a 2-base deletion mutation.[36] Orientation of the adduct within the helix (e.g. *syn* vs *anti*) exposes the adduct to either the non-polar environment outside the helix or the polar environment inside the helix.[37–40]

While knowledge of conformational preferences of C8-dG in DNA can help predict the mutagenic potential of chemicals on the basis of their DNA reactivity profiles, the structural details of their orientation can be difficult to evaluate and strongly depend on particular conditions. Thus, modified nucleic acid probes that are sensitive to the solvent environment are useful for reporting on

conformational changes within the helix. [41,42] C8-fluoro-biphenyl-dG (FBP-dG) was previously reported to stabilize a duplex construct that mimics a slipped mutagenic duplex in the 12mer *Nar*I sequence. Furthermore, it is emissive when it forms the slipped mutagenic duplex.[43] The adduct appears to prefer a *syn* conformation about the glycosidic angle and intercalate within the helix, as supported by ^{19}F NMR and computational modeling data.[43] The stacking interaction not only stabilized the truncated duplex, but also exposed the adduct to a less polar solvent environment.[43] Despite knowledge of how the modification impacts polymerase and DNA repair function, the environment-sensitive emissive properties of aryl-dG adducts have not been utilized as an indicator of how sequence features impact mutational likelihood.

In order to gain a deeper understanding of how C8-aryl adducts induce frameshift mutations, we characterized the mutational outcome of a C8-FBP-dG and the structure-relationship activity within the *Nar*I sequence. First, we evaluated the influence of the C8-aryl adduct FBP-dG on polymerase-catalyzed DNA synthesis using the high fidelity DNA polymerase Kf⁻ and how the capacity for full-length synthesis vs. frame-shift was modulated by the local sequence context of the adduct. To resolve the sequence dependent process of primer extension, we created DNA constructs containing FBP-dG in template strands that were annealed to primers with incrementally increasing lengths to model the stepwise process of DNA synthesis in an error free vs. frameshift inducing process. These constructs were compared on the basis of their thermal stability and fluorescence, as an indicator of the orientation of the adduct. Evaluating this process for two related sequences revealed a basis for why the propensity for frameshift mutation depends on sequence context utilized a simple fluorescence-based chemical approach.

3.2 Results

3.2.1 The FBP-dG adduct is extruded from a truncated DNA duplex

To evaluate the conformation of FBP-dG in a truncated duplex that results from frameshift mutagenesis, the modified base was incorporated into two 22mer sequences and its impact on thermal stability and emission was evaluated. The sequences were 22mer and contained either the NarI endonuclease sequence CG^{1}G^{2}CG^{3}C or a variant with the CG3 sequence G^{1}CG^{2}CG^{3}C. In both cases, G^3 was replaced by FBP-dG in both sequences using phosphoramidite chemistry and solid phase DNA synthesis as previously published.[43] The template sequences were annealed to two different complementary sequences, a full length complement and a truncated sequence two bases shorter than the full length sequence, thus modeling the outcome of frameshift mutagenesis. Replacement of dG with FBP-dG in both templates annealed to the full length complementary sequence destabilized the duplexes by 8.0 °C for the GC3 sequence and 9.0 °C for the NarI sequence (Table 1 and 2). Whereas replacing the dG with FBP-dG in both templates annealed to the truncated sequence stabilized the duplexes by 3.1 °C in the GC3 sequence and 2.0 °C in the NarI sequence (Table 1 and 2). These results were consistent with the data previously published for the 12mer *Nar*I sequence containing FBP-dG,[43] as well as for AAF-dG, which stabilized the truncated duplex in the NarI sequence, and has a tendency to induce frameshift mutations.[44–46] The fluorescence intensity of FBP-dG was higher for the truncated vs. fully duplex products, by 3-fold in in the CG3 sequence and by 2.5-fold in the NarI sequence (Figure 3). These data were consistent with previously reported properties of FBP-dG in a 12mer.[43]

3.2.2 Kf- can bypass FBP-dG in the CG3 sequence, but not in the NarI sequence

To test the propensity of polymerase-mediated bypass synthesis over FBP-dG to result in frameshift mutation, we carried out primer extension reactions using a model high fidelity polymerase Klenow fragment exo- (Kf⁻), first with single nucleotides (Figure 1b). In the *Nar*I sequence two bands were

observed for C, one for incorporation of one C (28 %) and one for incorporation of two Cs (47 %). Misincorporation of G (17 %) and A (36 %) was also observed. Incorporation of a second C past N-linked C8-dG adducts has previously been categorized as a two-base slippage mechanism and implied a frameshift mutation.[47] On the other hand, in the CG3 sequence only one C was incorporated and misincorporation of G (12 %) and A (22 %) was observed. Single nucleotide incorporation studies showed that C was incorporated the most efficient opposite FBP-dG in both sequence contexts, additionally we could conclude that in the *Nar*I sequence a base slippage might occurred, resulting in a frameshift mutation.

When all four dNTPs were added to the reaction, in the *Nar*I sequence, extension after the first nucleotide addition was blocked whereas in the CG3 sequence, full-length products were obtained (Figure 1c). We performed these full-length synthesis studies in the presence of all four nucleotides and with 10 nM Kf- concentration. Kf- was in in the *Nar*I sequence context strongly blocked. The polymerase was not able to bypass past FBP-dG. Bypass of FBP-dG was better in the CG3 sequence context, the n+1 band still persisted, however also a n+1 band was generated as well as an unanticipated n+4 band was observed. We increased the DNA polymerase concentration to 50 and 100 nM to test whether we could better promote bypass of FBP-dG. Indeed, high DNA concentration resulted in better bypass in both DNA sequence contexts. In the NarI sequence the n+1 band persisted and only a unusual n+4 band was generated. Whereas in the CG3 sequence the shorter products were all extended, generating full length and truncated products. As well as an unanticipated n+4 band is still produced. Concluding that Kf- could bypass FBP-dG better in the CG3 sequence than in the *Nar*I sequence, highlighting the impact of the sequence context on the mutational outcome. It remained unclear, how the conformation of FBP-dG impact the mutational outcome, and how frameshift mutation was generated. Therefore, in the next experiments we constructed DNA duplexes modeling the stepwise process of frameshift mutation and elucidated the mechanism with a simple fluorescence-based chemical approach.

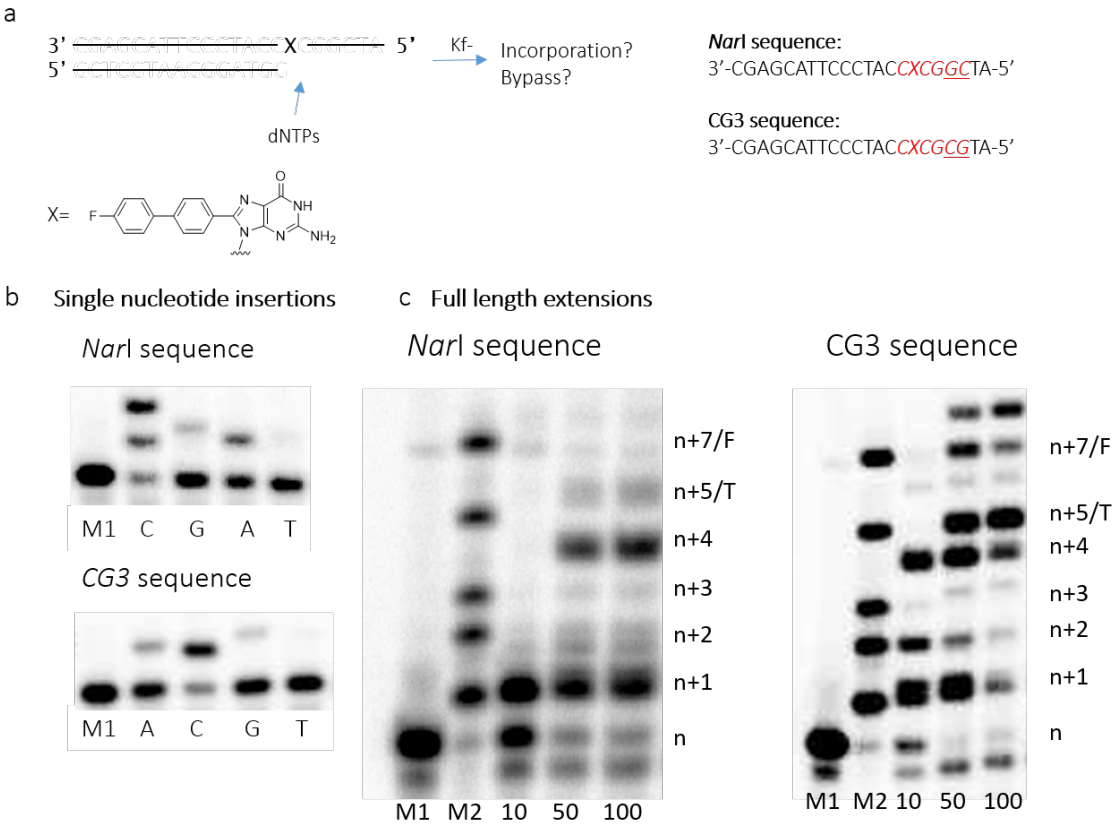


Figure 1: a) DNA polymerase-mediated primer extension experiments and sequences used in this study. b) Single Nucleotide extension with Kf⁻ (100 nM duplex, 100 µM dNTPs, 10 nM Kf-. Incubation: 15min 37 °C). c) Full length extension with Kf⁻ of the 22mer NarI sequence and 22mer CG3 sequence (100 nM duplex, 100 µM dNTPs, 10, 50, 100 nM Kf⁻. Incubation: 30 min, 37 °C). M1: 15mer primer. M2: mixture of n+1 primer, n+2 primer n+3 primer, truncated complementary sequence (n+5/T) and full length sequence (n+7/F).

3.2.3 Stepwise monitoring of frameshift mutation formation with fluorescence spectroscopy

The 22mer sequences containing FBP-dG at position 16 were annealed to primers of incrementally increasing lengths to model possible synthesis intermediates, in order to better understand the sequence dependent mechanism for frameshift mutation with a simple fluorescence-based approach. Addressing the *Nar*I sequence involved the construction of two model systems. The first was designed to test a high-fidelity pathway, where primers were added to create fully paired synthesis intermediates (Figure 2): n (22/15mer), n+1 (22/16-mer), n+2 (22/17-mer), n+3 (22/18mer) and n+7/F (22/22mer). For the slipped mutagenic pathway, primers were added to create a -2 CG deletion: n (22/15-mer), n+1 (22/16-mer), n+2 (22/17-mer), n+3 (22/18-mer) and n+5/T (22/20-mer). The CG3 sequence involved the construction of only one model system, as the synthesis intermediates between the fully paired and the slipped mutagenic pathways are the same (n to n+3): n (22/15mer), n+1 (22/16-mer), n+2 (22/17-mer), n+3 (22/18mer), and SMI/n+5 (22/20-mer) for the truncated duplex and FP/n+7 (22/22mer) for the full length duplex.

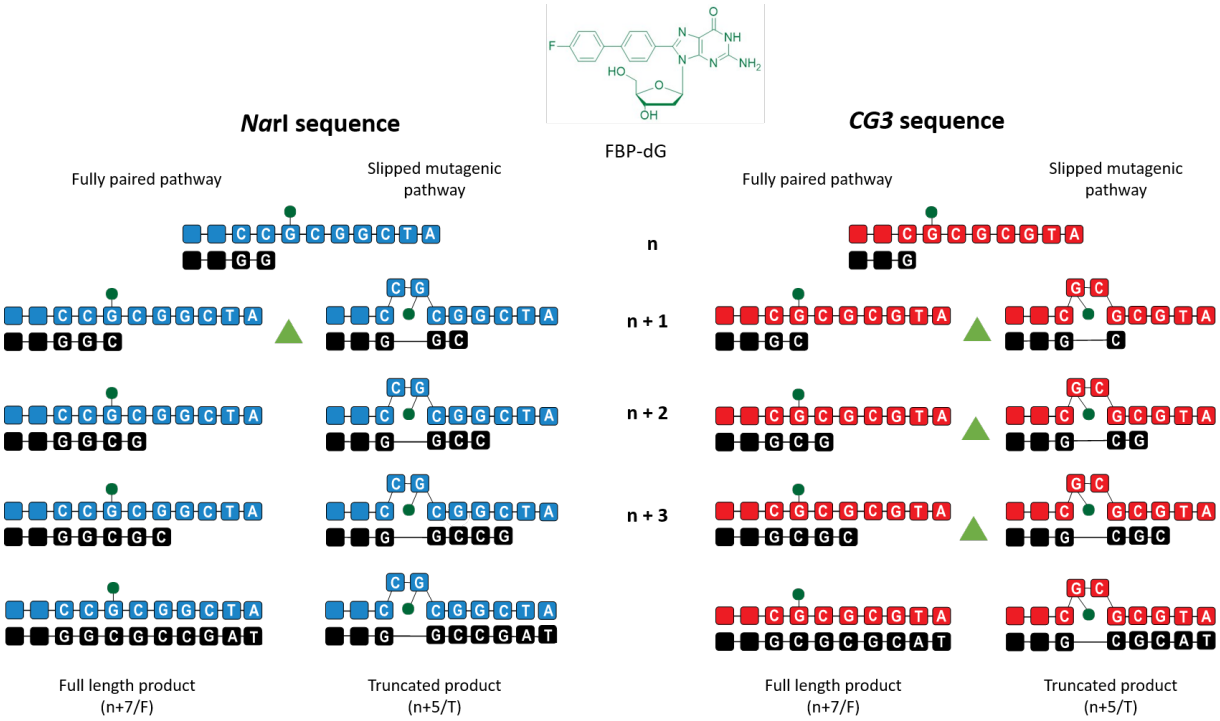


Figure 2: Constructed model duplexes: For duplexes n+1 for the *Nar*I and CG3 sequences, n+2 and n+3 in the CG3 sequence the primer sequences are the same (marked with green triangle).

The fluorescence intensity of FBP-dG in the n+1 duplex increased for the *Nar*I sequence (Figure 3), suggesting that FBP-dG induced a bulged out structure leading to a frameshift mutation. Indeed, following subsequent base additions for the slipped mutagenic pathway (n to n+5/T) the fluorescence intensity remained high, suggesting that the bulged structure was formed at the initial insertion step and persisted. In contrast, via the fully paired pathways, fluorescence intensity decreased as the primers were elongated, suggesting that the template realigned.

In the CG3 sequence, the fluorescence intensity of FBP-dG increased for the n+2 duplex, which was in contrast to the *Nar*I sequence. Upon addition of one base opposite the adduct (n+1), FBP-dG did not induce the bulge out form, suggesting the added base might correctly pair with FBP-dG, and upon adding of another base (n+2) FBP-dG induced the bulged out structure. Through further addition of bases the fluorescence intensity did not change until the formation of the truncated product (n+5/T), this supports that the bulge out structure was generated at position n+2 and remained until the truncated product was generated. In contrast, fluorescence intensity decreased for modeling the full length duplex (n+7/F), suggesting that correct pairing of FBP-dG decreased fluorescence intensity.

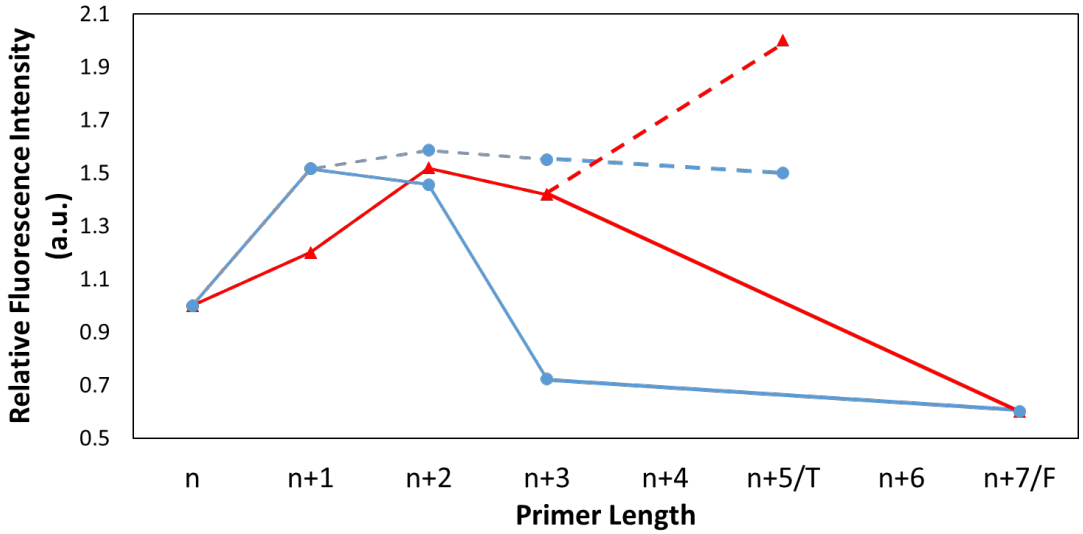


Figure 3: Relative fluorescence intensity (a.u.) as a function of primer elongation for the fully paired pathway (solid lines) and slipped mutagenic pathway (dashed lines) for the *Nar*I sequence (blue) and CG3 sequence (red).

Table 1: **Thermal stability data of CG3 sequence annealed to primers to model the synthesis intermediates**

	FBP-dG (°C)	G (°C)	ΔT (dG-FBP-dG)
n	59.4±0.5	59.6±0.5	-0.2
n+1	60.5±0.9	65.1±1.[a]	-4.6
n+2	62.8±0.6	66.6±0.3	-3.8
n+3	65.8±0.4	71.0±0.5	-5.2
FP/n+7	66.6±0.9	74.6±1.0	-8.0
SMI/n+5	68.1±1.1	65.0±0.5	+3.1

Table 2: Thermal stability of *Nar*I sequence annealed to primers to model the synthesis intermediates of fully paired and truncated duplexes

	Fully paired FBP-dG(°C)	Fully paired dG (°C)	ΔT (dG-FBP-dG)	Truncated FBP-dG (°C)	Truncated dG (°C)	ΔT (dG-FBP-dG)
n	59.3±1.3	60.7±1.7	-1.4	na		
n+1	60.4±2.2	65.7±1.5	-5.3	na		
n+2	62.2±1.3	68.3±1.3	-6.1	62.3±1.1	65.7±1.5	-3.4
n+3	62.6±1.1	72.0±1.3	-9.4	65.0±1.2	65.3±2.0	-0.3
FP/n+7	68.1±1.6	77.6±2.7	-9.0	na		
SMI/n+5	na			67.7±1.1	65.7±2.0	+2.0

The fluorescence data suggested that in the *Nar*I sequence the bulged out structure was formed after the insertion of one base (n+1 duplex), whereas in the CG3 sequence the bulged out structure was formed after the insertion of two bases (n+2 duplex, Figure 3), which was supportive for the enzymatic studies, where the *Nar*I sequence stalled DNA polymerase and in the CG3 sequence could better be bypassed. Replacing dG with FBP-dG in the CG3 sequence stabilized the n+1 duplex (Table 1), whereas replacing dG with FBP-dG in the *Nar*I sequence destabilized the n+1 duplex (Table 2), suggesting that in the CG3 sequence the incoming C could has favorable pairing interactions with FBP-G. In contrast, in the *Nar*I

sequence the incoming C could not be placed opposite FBP-dG, and therefore destabilized the duplex. With increasing primer length, melting temperature differences between the unmodified duplexes and modified duplexes (ΔTm) decreased for the *Nar*I sequence in the truncated intermediates. The truncated duplex n+5/T was more stable replacing dG with FBP-dG (Table 2). For the fully paired intermediates, the differences in melting temperatures of the FBP-dG duplexes increased with increasing primer length in the *Nar*I sequence. In the CG3 sequence, replacing dG with FBP-dG resulted in a destabilization of all constructed duplexes, only the truncated duplex was stabilized by replacing dG with FBP-dG. As previously determined, the glycosidic angle of the FBP-dG was preferentially in a syn conformation in the 12mer *Nar*I sequence and thus might intercalate into the helix and destabilized the duplex.[43] In contrast, the stability and fluorescence data for the CG3 sequence suggest that the glycosidic angle of FBP-dG to be in the *anti* conformation, positioning the modified base to correctly pair with the incoming base, and facilitating bypass by DNA polymerases.

3.3 Discussion

Using simple chemical tools and enzymatic studies, we gained new insight into how DNA sequence impacts the genesis of frameshift mutations associated with C8-dG adducts. The synthetic fluorescent deoxyguanosine adduct FPB-dG, bearing a biphenyl moiety at the C8 position of dG stabilized a DNA duplex, and was emissive when placed in a truncated duplex. Using the high fidelity polymerase Kf⁻, we observed a dramatic difference in bypass of FBP-dG depending on the sequence context. Despite the incorporation of two Cs opposite FBP-dG in the *Nar*I sequence, which suggested a possible frameshift mutation, FBP-dG strongly blocked Kf⁻. It appears that the formation of the bulged out structure in the *Nar*I sequence greatly alters DNA conformation, resulted the observed stalling. On the other hand, Kf⁻ could bypass FBP-dG in the CG3 sequence. In contrast to the *Nar*I sequence, the bulge out structure is created after incorporation of two bases past FBP-dG, resulting in a greater stalling of the DNA polymerase.

Based on the duplex stability in combination with the fluorescence data, we suggested that FBP-dG disturbs the DNA conformation, which results in stalling of the DNA polymerase. Previous studies using C-linked adducts have shown that a *syn* conformation about the glycosidic angle is more mutagenic than a glycosidic angle in an *anti* conformation.[35,36] AAF-dG destabilizes duplexes at the replication fork (n+1) and promotes the generation of frameshift mutations.[9] The glycosidic angle of modified bases is preferred to accommodate the *anti* conformation of the nucleoside for efficient polymerase extension.[9] In a previous study, it was observed that a fluorinated aminobiphenyl-dG (FABP-dG) had a larger stacked conformational population in the G*CA as in G*CT, while for fluorinated aminofluorene-dG (AF-dG) no difference was observed, furthermore the stacked conformational population did not change with increasing the primer length from 11 to 19 nucleotides.[9] Thus, C was more likely incorporated opposite FBP-dG than opposite FAF-dG, which is most likely due to the high stacked conformer in the FAF adduct.[9] Based on results of the FBP-dG within the 12mer *Nar*I sequence, the adduct preferred a *syn* conformation and the adduct intercalates within the helix.[43] Thus when FBP-dG intercalated within the helix, the incoming C cannot pair with the adduct, and created the frameshift mutation. Based on enzymatic studies and fluorescence spectroscopy a different conformation of FBP-dG was expected for the two sequence contexts, suggesting that the adduct did not adopt the *syn* conformation upon inserting one base opposite the adduct in the CG3 sequence. Thus the incoming base could correctly pair with the adduct and be bypassed more efficiently than in the *Nar*I sequence.

3.4 Conclusion

To study the impact of DNA sequence context on the propensity for C8-aryl-dG adducts, which are induce frameshift mutations, we characterized in different 22mer sequences the impact of a fluorescent modified guanine (FBP-dG) on polymerase-mediated bypass, fluorescence properties and thermal

stability. FBP-dG showed turn-on emissive properties when truncated intermediates or products associated with frameshift mutagenesis were generated, indicating that the creation of a bulged out adduct structure and its subsequent realignment process could explain DNA polymerase stalling during synthesis. However, in the *Nar*I sequence the bulged out structure was created upon inserting one base opposite the adduct, whereas in the CG3 sequence it was created upon inserting of two bases opposite the adduct, resulting from its orientation accommodating favorable interactions with the incoming nucleotide and yielding the product of full-length synthesis. This study highlight the importance of the sequence context for frameshift mutations and proposed a simple chemical tool to better understand the mechanism for frameshift mutations. T

3.5 Experimental details

3.5.1 DNA Synthesis

The DNA sequences (*Nar*I 22mer: 5'-CTCG^{1}G^{2}CG^{3}CCATCCCTTACGAGC-3', and the 22mer: 5'-CTG^{1}CG^{2}CG^{3}CCATCCCTTACGAGC-3', which contains 3GC repeats and where G^3 is FBP-dG[43] was prepared with a BioAutomation MerMade 12 automatic DNA synthesizer using standard or modified β-cyanoethylphosphoramidite chemistry.[36] Following synthesis, oligonucleotides were cleaved from the solid support and deprotected with 2 mL 30 % ammonium hydroxide solution at 55 °C for 12 h and purified by RP-HPLC. Oligonucleotides were analyzed by electrospray ionization mass spectrometry (ESI-MS).[48] Modified or unmodified oligonucleotide were annealed with primer strands at a final concentration of 3 μM in 50 mM Na_2HPO_4 buffer and 100 mM NaCl (pH 7). The strands were annealed by heating to 95 °C and then slowly cooling (1 °C/min) to 10 °C. The samples were used for melting temperature, fluorescence and circular dichroism measurements.

3.5.2 Melting Temperatures

Melting temperatures (T_m) of the duplexes were determined by variable temperature UV analysis using a Cary 300-Bio UV-vis spectrophotometer equipped with a 6x6 Multicell Peltier block-heating using Hellma 114-QS 10 mm light path cells. Modified or unmodified oligonucleotide were mixed with primer strands at a final concentration of 3 μM of each teamplate and primer in 1000 μL buffer (1X buffer: 50 mM Na_2HPO_4 buffer and 100 mM NaCl (pH 7)). The heating cycle (1 °C/min) then was repeated and UV absorption monitored at 280 nm. Melting temperatures (T_m) were determined by the first-derivative method with the application in the Cary Thermal software. All measurements were performed three times and the mean values (±standard deviation) are reported.

3.5.3 Fluorescence

Fluorescence spectra were recorded on a Cary Eclipse Fluorescence spectrophotometer equipped with a 1 x 4 multicell block Peltier stirrer and temperature controller, and excitation and emission slit widths of 2.5 nm. Modified oligonucleotides were annealed with primer strands at a final concentration of 3 μM of each teamplate and primer in 1000 μL buffer (1X buffer: 50 mM Na_2HPO_4 buffer and 100 mM NaCl (pH 7)).[36]

3.5.4 Primer elongation experiments

T4 polynucleotide kinase and [γ-^{32}P]ATP were used to label the 15mer primer strand at the 5'-end according to the manufacturers protocol.[49] Primers and templates were annealed by heating the sample to 95 °C for 5 min in a heating block and cooling over the course of 16 h in a block. Reaction mixtures contained 10-100 nM polymerase (Kf-), 100 nM 5' [^{32}P]primer:template DNA and 25 μM dNTPs in 10

μL of reaction buffer (1x Buffer: 50 mM NaCl, 10mM Tris-HCl (pH7.9), 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT))

Reaction mixtures were incubated at 37 °C for 60 min, then quenched by adding 20 μL PAGE gel loading buffer (80% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanole FF). The resulting mixtures (4 μL) were analyzed on 15% polyacrylamide/7M urea denaturing gels. Radioactive bands were visualized by autoradiography (Bio-Rad, Hercules, CA) and quantified using the Bio-Rad Quantity One software.[36]

3.6 References

(1) Luch, A. (2005) Nature and nurture - lessons from chemical carcinogenesis. *Nat Rev Cancer 5*, 113–125.

(2) Gannett, P. M., Powell, J. H., Rao, R., Shi, X., Lawson, T., Kolar, C., and Toth, B. (1999) C8-arylguanine and C8-aryladenine formation in calf thymus DNA from arenediazonium ions. *Chem. Res. Toxicol. 12*, 297–304.

(3) Hiramoto, K., Kaku, M., Sueyoshi, A., Fujise, M., and Kikugawa, K. (1995) DNA base and deoxyribose modification by the carbon-centered radical generated from 4-(hydroxymethyl)benzenediazonium salt, a carcinogen in mushroom. *Chem Res Toxicol 8*, 356–362.

(4) Sen, S., Bhojnagarwala, P., Francey, L., Lu, D., Pening, T. M., and Field, J. (2012) p53 Mutagenesis by Benzo[a]pyrene derived Radical Cations. *Chem. Res. Toxicol. 25*, 2117–2126.

(5) Kuroda, K., Hibi, D., Ishii, Y., Takasu, S., Kijima, A., Matsushita, K., Masumura, K. I., Watanabe, M., Sugita-Konishi, Y., Sakai, H., Yanai, T., Nohmi, T., Ogawa, K., and Umemura, T. (2013) Ochratoxin A induces DNA double-strand breaks and large deletion mutations in the carcinogenic target site of gpt delta rats. *Mutagenesis 29*, 27–36.

(6) Kriek, E. (1992) Fifty years of research on N-acetyl-2-aminofluorene, one of the most versatile compounds in experimental cancer research. *J. Cancer Res. Clin. Oncol. 118*, 481–489.

(7) Turesky, R. J. (2010) Heterocyclic aromatic amines: Potential human carcinogens. *Adv. Mol. Toxicol. 4*, 37–83.

(8) Cho, B. P., Beland, F. A., and Marques, M. M. (1994) NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. *Biochemistry 33*, 1373–1384.

(9) Jain, V., Vaidyanathan, V. G., Patnaik, S., Gopal, S., and Cho, B. P. (2014) Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19 F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry 53*, 4059–4071.

(10) Dai, J., Wright, M. W., and Manderville, R. A. (2003) Ochratoxin A Forms a Carbon-Bonded C8-Deoxyguanosine Nucleoside Adduct: Implications for C8 Reactivity by a Phenolic Radical. *J. Am. Chem. Soc. 125*, 3716–3717.

(11) Chen, L., Devanesan, P. D., Higginbotham, S., Ariese, F., Jankowiak, R., Small, G. J., Rogan, E. G., and Cavalieri, E. L. (1996) Expanded analysis of Benzo[a]pyrene-DNA adducts formed in vitro and in mouse skin: Their significance in tumor initiation. *Chem. Res. Toxicol. 9*, 897–903.

(12) O'Handley, S. F., Sanford, D. G., Xu, R., Lester, C. C., Hingerty, B. E., Broyde, S., and Krugh, T. R. (1993) Structural Characterization of an N-Acetyl-2-aminofluorene (AAF) Modified DNA Oligomer by NMR, Energy Minimization, and Molecular Dynamics. *Biochemistry 32*, 2481–2497.

(13) Faucet, V., Pfohl-Leszkowicz, A., Dai, J., Castegnaro, M., and Manderville, R. A. (2004) Evidence for Covalent DNA Adduction by Ochratoxin A following Chronic Exposure to Rat and Subacute Exposure to Pig. *Chem. Res. Toxicol. 17*, 1289–1296.

(14) Cavin, C., Delatour, T., Marin-Kuan, M., Holzhäuser, D., Higgins, L., Bezençon, C., Guignard, G., Junod, S., Richoz-Payot, J., Gremaud, E., Hayes, J. D., Nestler, S., Mantle, P., and Schilter, B. (2007) Reduction in antioxidant defenses may contribute to ochratoxin A toxicity and carcinogenicity. *Toxicol. Sci. 96*, 30–39.

(15) Fan, H., and Chu, J.-Y. (2007) A Brief Review of Short Tandem Repeat Mutation. *Genomics. Proteomics Bioinformatics 5*, 7–14.

(16) Boland, C. R., and Goel, A. (2010) Microsatellite Instability in Colorectal Cancer. *Gastroenterology 138*, 2073–2087.

(17) Wu, B. P., Zhang, Y. L., Zhou, D. Y., Gao, C. F., and Lai, Z. S. (2000) Microsatellite instability,

MMR gene expression and proliferation kinetics in colorectal cancer with famillial predisposition. *World J. Gastroenterol. 6*, 902–905.

(18) Yurgelun, M. B., Goel, A., Hornick, J. L., Sen, A., Turgeon, D. K., Ruffin, M. T., Marcon, N. E., Baron, J. A., Bresalier, R. S., Syngal, S., Brenner, D. E., Boland, C. R., and Stoffel, E. M. (2012) Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps. *Cancer Prev. Res. 5*, 574–582.

(19) Fuchs, R. P. P., Schwartz, N., and Daune, M. P. (1981) Hot Spots of Frameshift Mutations Induced by the Ultimate Carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature 294*, 657–659.

(20) Koffel-Schwartz, N., Verdier, J.-M. M., Bichara, M., Freund, A.-M. M., Daune, M. P., and Fuchs, R. P. P. P. (1984) Carcinogen-induced mutation spectrum in wild-type, uvrA and umuC strains of Escherichia coli. Strain specificity and mutation-prone sequences. *J. Mol. Biol. 177*, 33–51.

(21) Fuchs, R. P., and Fujii, S. (2007) Translesion synthesis in Escherichia coli: Lessons from the NarI mutation hot spot. *DNA Repair (Amst). 6*, 1032–1041.

(22) Veaute, X., and Fuchs, R. P. P. (1991) Polymorphism in N -2-acetylaminofluorene induced DNA structure as revealed by DNase I footprinting. *Nucleic Acids Res. 19*, 5603–5606.

(23) Koffel-Schwartz, N., and Fuchs, R. P. (1995) Sequence determinants for -2 frameshift mutagenesis at NarI-derived hot spots. *J. Mol. Biol. 252*, 507–513.

(24) Broschard, T. H., Koffel-Schwartz, N., and Fuchs, R. P. (1999) Sequence-dependent modulation of frameshift mutagenesis at NarI-derived mutation hot spots. *J. Mol. Biol. 288*, 191–199.

(25) Belguise-Valladier, P., and Fuchs, R. P. (1995) N-2-aminofluorene and N-2 acetylaminofluorene adducts: the local sequence context of an adduct and its chemical structure determine its replication properties. *J. Mol. Biol. 249*, 903–913.

(26) Belguise-Valladier, P., and Fuchs, R. P. P. (1991) Strong Sequence-Dependent Polymorphism in Adduct-Induced DNA Structure: Analysis of Single N-2-Acetylaminofluorene Residues Bound within the NarI Mutation Hot Spot. *Biochemistry 30*, 10091–10100.

(27) Wagner, J., Etienne, H., Janel-Bintz, R., and Fuchs, R. P. P. (2002) Genetics of mutagenesis in E. coli: Various combinations of translesion polymerases (Pol II, IV and V) deal with lesion/sequence context diversity. *DNA Repair (Amst). 1*, 159–167.

(28) Zhou, L., Rajabzadeh, M., Traficante, D. D., and Cho, B. P. (1997) Conformational Heterogeneity of Arylamine-Modified DNA: 19F NMR Evidence. *J. Am. Chem. Soc. 119*, 5384–5389.

(29) Cho, B. P., Beland, F. A., and Marques, M. M. (1992) NMR Structural Studies of a 15-mer DNA Sequence from a ras Protooncogene, Modified at the First Base of Codon 61 with the carcinogen 4-Aminobiphenyl. *Biochemistry 31*, 9587–9602.

(30) Meneni, S., Shell, S. M., Zou, Y., and Cho, B. P. (2007) Conformation-specific recognition of carcinogen-DNA adduct in Escherichia coli nucleotide excision repair. *Chem. Res. Toxicol. 20*, 6–10.

(31) Meneni, S., Liang, F., and Cho, B. P. (2007) Examination of the Long-range Effects of Aminofluorene-induced Conformational Heterogeneity and Its Relevance to the Mechanism of Translesional DNA Synthesis. *J. Mol. Biol. 366*, 1387–1400.

(32) Hsu, G. W., Kiefer, J. R., Burnouf, D., Becherel, O. J., Fuchs, R. P. P., and Beese, L. S. (2004) Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. *J Biol Chem 279*, 50280–50285.

(33) Fuchs, R., and Daune, M. (1972) Physical studies on deoxyribonucleic acid after covalent binding of a carcinogen. *Biochemistry 11*, 2659–2666.

(34) Fuchs, R., and Daune, M. (1973) Physical basis of chemical carcniogenesis by N-2-fluorenylacetamide derivatives and analogs. *FEBS Lett. 34*, 295–298.

(35) Sproviero, M., Verwey, A. M. R., Witham, A. A., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2015) Enhancing Bulge Stabilization through Linear Extension of C8-Aryl-Guanine Adducts to Promote Polymerase Blockage or Strand Realignment to Produce a C:C Mismatch. *Chem. Res. Toxicol. 28*, 1647–1658.

(36) Sproviero, M., Verwey, A. M. R., Rankin, K. M., Witham, A. A., Soldatov, D. V, Manderville, R. A., Fekry, M. I., Sturla, S. J., Sharma, P., and Wetmore, S. D. (2014) Structural and biochemical impact of C8-aryl-guanine adducts within the NarI recognition DNA sequence: influence of aryl ring size on targeted and semi-targeted mutagenicity. *Nucleic Acids Res. 42*, 13405–13421.

(37) Suzuki, A., Saito, M., Katoh, R., and Saito, Y. (2015) Synthesis of 8-aza-3,7-dideaza-2′-deoxyadenosines possessing a new adenosine skeleton as an environmentally sensitive fluorescent nucleoside for monitoring the DNA minor groove. *Org. Biomol. Chem. 13*, 7459–7468.

(38) Suzuki, A., Yanaba, T., Saito, I., and Saito, Y. (2014) Molecular design of an environmentally sensitive fluorescent nucleoside, 3-deaza-2'-deoxyadenosine derivative: Distinguishing thymine by probing the DNA minor groove. *ChemBioChem 15*, 1638–1644.

(39) Minakawa, N., Kawano, Y., Murata, S., Inoue, N., and Matsuda, A. (2008) Oligodeoxynucleotides containing 3-bromo-3-deazaadenine and 7-bromo-7-deazaadeinine 2'-deoxynucleosides as chemical probes to investigate DNA-protein interactions. *ChemBioChem 9*, 464–470.

(40) Sinkeldam, R. W., Greco, N. J., and Tor, Y. (2010) Fluorescent Analogs of Biomolecular Building Blocks Design Properties and Applications. *Chem. Rev. 110*, 2579–2619.

(41) Mata, G., and Luedtke, N. W. (2015) Fluorescent probe for proton-coupled DNA folding revealing slow exchange of i -motif and duplex structures. *J. Am. Chem. Soc. 137*, 699–707.

(42) Dumas, A., and Luedtke, N. W. (2010) Cation-mediated energy transfer in G-quadruplexes revealed by an internal fluorescent probe. *J. Am. Chem. Soc. 132*, 18004–18007.

(43) Berger, F. D., Sturla, S. J., Kung, R. W., Montina, T., Wetmore, S. D., and Manderville, R. A. (2017) Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the Nar I Mutational Hotspot: Evidence for Enhanced Syn Adduct Formation. *Chem. Res. Toxicol. 31*, 37–47.

(44) Zhou, Y., and Romano, L. J. (1993) Solid-Phase Synthesis of Oligonucleotides Containing Site-Specific N-(2'-Deoxyguanosin-8-yl)-2-(acetylamino)fluorene Adducts Using 9-Fluorenylmethoxycarbonyl as the Base-Protecting Group. *Biochemistry 32*, 14043–14052.

(45) Milhé, C., Fuchs, R. P., and Lefèvre, J. F. (1996) NMR data show that the carcinogen N-2-acetylaminofluorene stabilises an intermediate of -2 frameshift mutagenesis in a region of high mutation frequency. *Eur. J. Biochem. 235*, 120–127.

(46) Xu, L., and Cho, B. P. (2016) Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chem. Res. Toxicol. 29*, 213–226.

(47) Rechkoblit, O., Kolbanovskiy, A., Malinina, L., Geacintov, N. E., Broyde, S., and Patel, D. J. (2010) Mechanism of error-free and semitargeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4. *Nat Struct Mol Biol 17*, 379–388.

(48) Sproviero, M., Rankin, K. M., Witham, A. A., and Manderville, R. A. (2014) Utility of 5′-O-2,7-Dimethylpixyl for Solid-Phase Synthesis of Oligonucleotides Containing Acid-Sensitive 8-Aryl-Guanine Adducts. *J. Org. Chem. 79*, 692–699.

(49) Promega Corporation. (2012) DNA 5′ End-Labeling System Technical Bulletin. *https://ch.promega.com/resources/protocols/technical-bulletins/0/dna-5-end-labeling-system-protocol/*.

Supporting information

Table S1: **ESI⁻ MS Analysis of FBP-dG and Modified *Nar*I Oligonucleotides**

		expt M/Z (ESI-)	exptl mass
Chemical Formula: $C_{222}H_{276}FN_{78}O_{131}P_{21}$	$[M-12H]^{12-}$	565.69	6800.28
Exact Mass: 6799.18	$[M-11H]^{11-}$	617.19	6800.09
Molecular Weight: 6802.51	$[M-10H]^{10-}$	679	6800
	$[M-9H]^{9-}$	754.5	6799.5

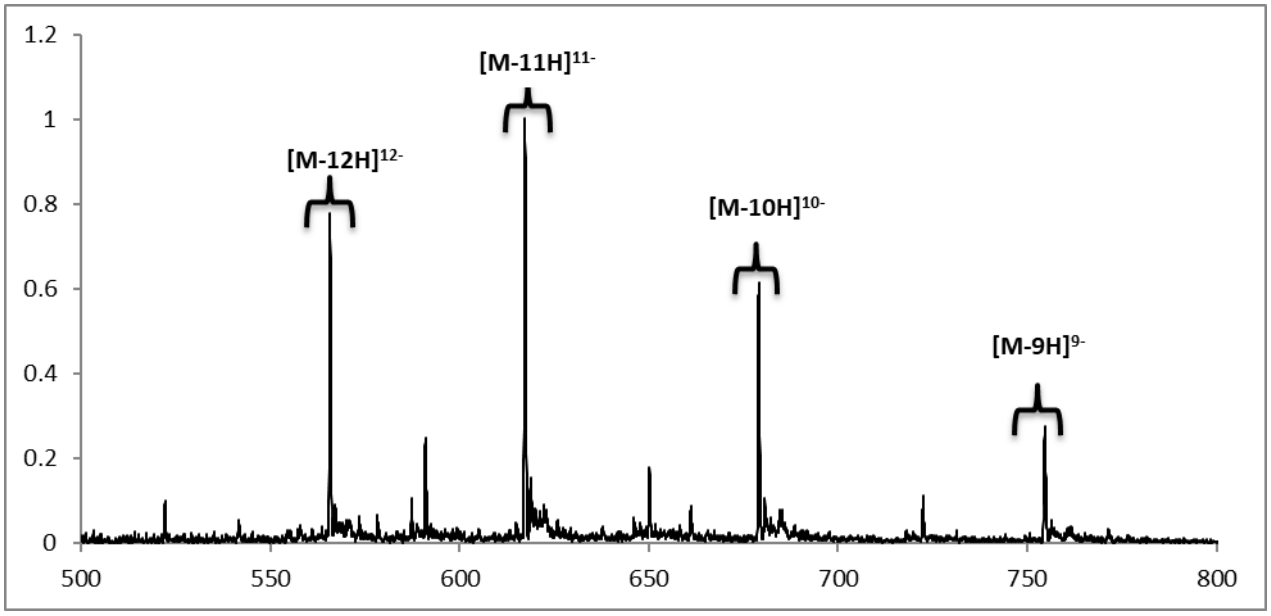


Figure S1: MS Spectrum for FBP-dG modified 22mer sequence

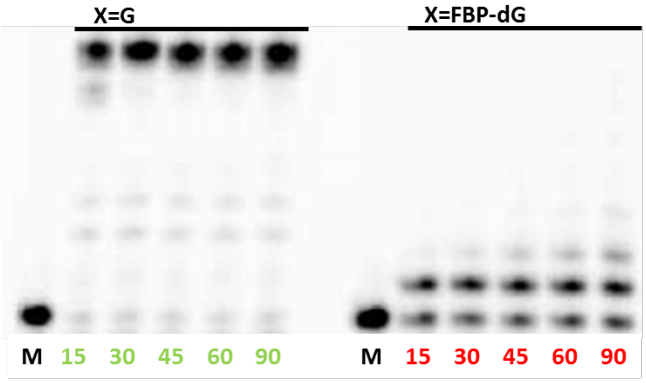
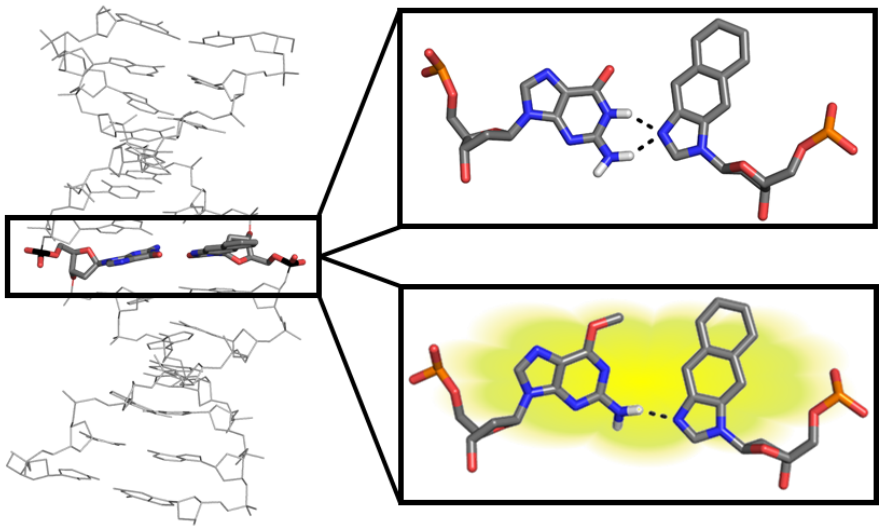


Figure S2: Full length extension of FBP-dG and dG containing 22mer *Nar*I sequence using TLS polymerase Dpo4. 10 µL reaction mixtures contained 10 nM polymerase (Dpo4), 100 nM 5' [^{32}P]primer:template DNA and 25 µM dNTPs in the reaction Buffer (1x Buffer for Dpo4 contained: 50nM NaCl, 50mM Tris (pH 8.0), 2.5 mM MgCl2, 5 mM DTT, 100 µg/mL bovine serum albumin, 5 % Glycerol). Reaction mixtures were incubated at 37 °C for 15 – 60 min, then quenched by adding 20 µL PAGE gel loading buffer (80% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanole FF). The resulting mixtures (4 µL) were analyzed on 15% polyacrylamide/7M urea denaturing gels.

Chapter 4

Fluorescent elongated hydrophobic nucleobase analogues stabilize DNA duplexes containing O^6-alkylguanine adducts



Heidi A. Dahlmann,[a] Florence D. Berger,[a] Ryan W. Kung, Laura A. Wyss, Irina Gubler, Maureen McKeague, Stacey D. Wetmore, and Shana J. Sturla

[a] Authors contributed equally to this work

Submitted and accepted for review at Helvetica Chimica Acta

H.A.D. and S.J.S. conceived of the project. H.A.D., F.D.B., R.W.K., L.A.W., I.G., and M.M. performed experiments. All authors contributed to the design of experiments, interpretation of data, and writing the paper.

Fluorescent elongated hydrophobic nucleobase analogues stabilize DNA duplexes containing O^6-alkylguanine adducts

Heidi A. Dahlmann,[‡a,b] Florence D. Berger,[‡a] Ryan W. Kung,[c] Laura A. Wyss,[a] Irina Gubler,[a] Maureen McKeague,[a] Stacey D. Wetmore,[c] and Shana J. Sturla[*,a]

[a] Department of Health Science and Technology, ETH Zürich, Schmelzbergstrasse 9, 8092 Zürich (Switzerland), sturlas@ethz.ch
[b] Current affiliation: Department of Chemistry, Hendrix College, 1600 Washington Ave, Conway, AR 72032 (USA)
[c] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada
[‡] Authors contributed equally to this work

Chemical alkylation of DNA produces potentially toxic and mutagenic damage such as O^6-alkylguanine (O^6-alkylG) adducts. Non-natural nucleoside analogues that pair with DNA adducts provide a potential basis for studying damaged DNA. Herein, we evaluated the base pairing properties of elongated nucleoside analogues containing napthalene-derived tricyclic nucleobases as DNA adduct-pairing nucleoside analogues in DNA hybridization probes. DNA duplex melting studies revealed that the elongated nucleoside analogs formed more stable base pairs opposite O^6-alkylG than G and were better able to distinguish between G, O^6-alkylG, and an abasic site than any previously described nucleoside analogue. DNA duplexes containing an elongated base analogue exhibited different fluorescence intensities when paired opposite O^6-alkylG versus G or abasic sites. Their selectivity for stabilizing alkylated DNA make the elongated hydrophobic base analogues improved candidates for incorporating into DNA hybridization probes targeting O^6-alkylG.

Keywords: DNA damage • nucleobase analogues • DNA recognition • fluorescent probes • stacking interactions

Introduction

Oligonucleotide hybridization comprises a fundamental basis for sequence-specific nucleic acid detection strategies ranging from DNA and RNA detection blots and gene arrays to fluorescence in situ hybridization and fluorescence resonance energy transfer probes.[1,2] Nucleic acid hybridization is also an essential feature of modern gene-targeting techniques such as antisense gene therapy[3] and CRISPR/Cas9.[4,5] The specificity and efficiency of hybridization between oligonucleotide strands is driven by hydrogen bonding and shape complementarity between nucleobases in opposite strands, which contribute to the overall nucleic acid duplex stability. Subsequently, duplex stability can be significantly impacted by non-natural nucleoside analogues such as fluorophores, which are frequently included in hybridization probes, or by chemically-modified natural nucleosides in target nucleic acid strands. In these non-canonical situations, additional forces such as pi-stacking interactions may contribute significantly to DNA duplex stability and the design of new nucleic acid probes.[6]

Modification by chemical alkylation gives rise to DNA adducts that can result in mutation and toxicity. Alkylation at the O^6-position of guanine in DNA to form adducts including O^6-methylguanine (O^6-MeG, Figure 1A) changes the hydrogen bonding and steric relationships in the G:C base pair.[7–9] O^6-alkylG adducts (Figure 1A) form upon exposure to chemical genotoxins[10,11] such as nitrosamines or DNA alkylating drugs[12,13] or endogenously.[14,15] Although O^6-alkylG adducts occur physiologically at extremely low levels and can be removed by direct repair, they are prone to inducing G to A transition mutations[16] and promoting carcinogenesis.[13]

Oligonucleotide hybridization probes containing nucleoside analogues provide a means for sequence-specific analysis of DNA containing alkylated bases and facilitate exploration of factors contributing to stability of DNA duplexes containing non-natural or alkylated nucleobases.[6] Synthetic nucleoside analogues containing environmentally sensitive fluorophores comprise an important class since they can be utilized in hybridization probes for specific pairing with modified DNA base sites and fluorescence spectroscopy is a sensitive, relatively inexpensive, and straightforward technique for measuring DNA interactions and analyzing alkylated DNA.[17–20] For example, a fluorescent hybridization probe containing the nucleoside analogue adenosine-1,3-diazaphenoxazine distinguishes 8-oxoG from G,[18,21] and a fluorescent thymidine analogue or 2-aminopurine can be used to detect abasic sites.[19,22,23] Furthermore, a variety of modified purines as well as aza- and deaza purine analogues have been used to probe DNA mismatches.[24–27] Finally, Sasaki and co-workers reported a hybridization probe containing a chemically reactive nucleoside analogue that transfers an alkynylated enone to O^6-MeG but not to G, which creates a synthetic handle for further fluorescence- or affinity-based labelling of O^6-MeG in DNA.[28]

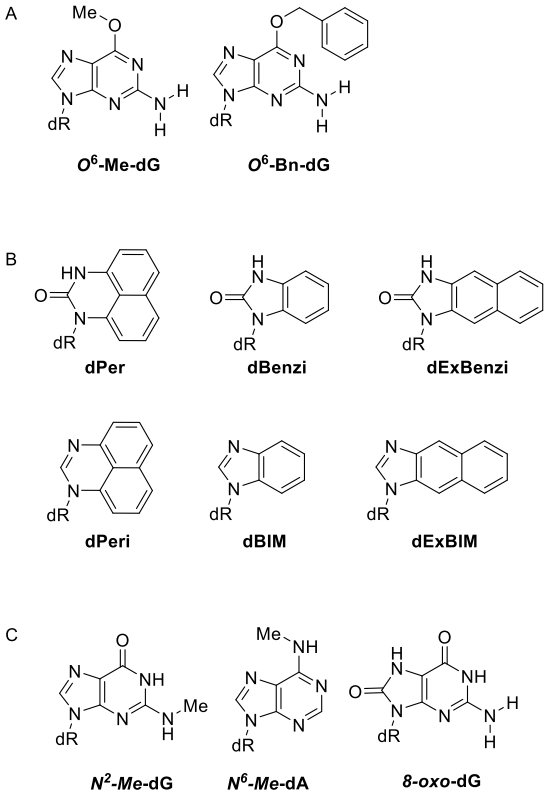


Figure 1 A) AlkylG. B) Nucleoside analogues. C) Modified DNA bases tested opposite ExBIM and ExBenzi. dR = deoxyribose

We developed a colorimetric DNA alkylation detection platform based on gold nanoparticles (AuNPs) functionalized with oligonucleotide hybridization probes containing hydrophobic nucleoside analogues that preferentially form base pairs with O^6-alkylG rather than G.[29,30] Earlier studies showed that a perimidinone-derived nucleoside analogue (Per, Figure 1B) could pair specifically with the model bulky DNA adduct O^6-BnG.[31] X-ray data indicated that the Per nucleoside assumed an anti-conformation with respect to the bulky nucleobase to intercalate into the duplex and form a π-stack with the benzyl group of O^6-BnG.[32] Further hydrophobic perimidinone- and benzimidazole-derived nucleoside analogues (Benzi, BIM, and Peri, Figure 1B) also formed more stable base pairs with O^6-alkylG adducts than with canonical G, A, T, and C in DNA duplexes.[33] Notably, the larger nucleosides Peri and Per were found to better stabilize O^6-BnG pairs than did the smaller analogues Benzi and BIM.[31,34]

We hypothesized that the larger synthetic bases were more effective for sensing alkylation sites because generally, larger nucleobases π-stack more strongly than smaller nucleobases due to having a larger area of overlap with neighbouring bases. This model is supported by data regarding nucleosides with benzene, naphthalene, and pyrene nucleobases[35] as well as expanded natural nucleobases (in so-called expanded "x"-and widened "y"-DNA).[36–45] Given that larger nucleobases facilitate π-stacking within the DNA duplex, we designed new expanded nucleoside analogues ExBenzi and ExBIM (Figure 1B), which both stabilize O^6-MeG in preference to G. These structures were critical for a first in-gene colorimetric quantification of O^6-alkylG, however several fundamental questions remained open concerning these structures such as the basis of their improved performance, selectivity and intrinsic fluorescence behaviour.[46]

In this study, we characterized the influence of elongated hydrophobic synthetic nucleosides ExBenzi and ExBIM on hybridization with chemically modified oligonucleotides. We determined the thermal stabilities of duplexes containing ExBenzi and ExBIM paired opposite G, O^6-MeG, O^6-BnG, N2-MeG, N6-MeA, 8-oxoG (Figure 1C) and a model abasic site (THF). To test the hypothesis that elongated nucleosides have increased π-stacking capacity, the stacking propensities of nucleoside analogues BIM, Benzi, Peri, Per, ExBIM, and ExBenzi were assessed. In particular, thermal melting studies were performed on short DNA duplexes containing unpaired nucleosides dangling from the 5'-ends of each strand. Additionally, we used molecular dynamic simulations to determine the relative contributions of hydrogen bonding and base stacking of each elongated hydrophobic synthetic nucleoside to stabilize DNA when paired opposite alkylguanine. Finally, we characterized the fluorescence properties of ExBIM and ExBenzi; both were found to have high quantum yields and showed turn-on emissive properties only opposite O^6-MeG.

Results and Discussion

Extended nucleosides stabilize duplexes containing O^6-alkyl-G

To address whether the extended nucleoside structures of ExBenzi and ExBIM increase the stability of duplexes with O6-alkylG:probe vs G:probe pairs, DNA duplexes with nucleoside analogues placed opposite O6-MeG, O6-BnG, or G were evaluated by performing thermal denaturation studies. Thermodynamic parameters of duplex melting were measured, namely free energy (ΔG°, Figure 2), which reflects duplex stability, as well as enthalpy (ΔH°) and entropy (ΔS°) (Table S1).[34] The hybridization probe sequence 5'-CCGPTATACCGACAA-3' (P = nucleoside analogue) was selected to allow direct comparison of our results with those previously reported for Per, Peri, Benzi, and BIM.[31,34] DNA duplexes containing nucleoside analogues paired opposite O6-alkylG adducts were generally more stable than those containing analogues paired opposite G (Figure 2, Table S1). This result is consistent with previous observations for Per[31,34] and may be rationalized on the basis of hydrophobic interactions favouring probe pairing with the more hydrophobic O6-alkylG adducts than with polar canonical nucleobases. This hydrophobic pairing phenomenon has been previously observed with other non-natural nucleosides.[41,43,44,47–53] The highest DNA duplex stabilities were observed for ExBenzi paired with either of the O6-alkylG adducts, while similar stabilization of O6-alkylG was observed for both ExBIM and Per (Figure 2, Table S1). These larger nucleoside analogues ultimately stabilize the adduct-containing DNA duplex more than smaller nucleosides BIM and Benzi.[34] An explanation for this observation may be that the larger nucleobases have more surface area and therefore higher propensity to π –stack within the duplex, consistent with evidence for π –stacking interactions in the crystal structure of a duplex with a Per: O6-BnG pair.[32]

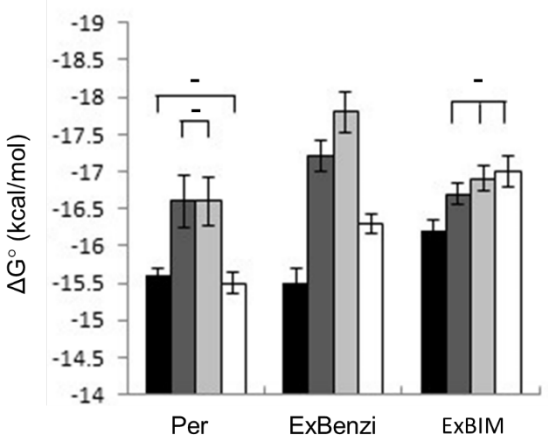


Figure 2: Comparison of standard free energies of duplex melting for DNA duplexes consisting of 5'-TTGTCGGTATAXCGG-3' complementary to 3'-AACAGCCATATPGCC-5', in which X = G, O^6-MeG, O^6-BnG, or THF and P = nucleoside analogue (Table S2). Error bars indicate 95% confidence intervals. Analogue:target pairs that do not differ significantly ($p > 0.05$) according to the Games-Howell test are denoted (-). black: G, dark grey: O^6-MeG, grey: O^6-BnG, white: THF

Elongated nucleosides have increased π-stacking interactions

To further assess the π-stacking propensities of the nucleoside analogues, their ability to stabilize the 5'-termini of short DNA duplexes was evaluated in a dangling-end experiment.[35–39,54–58] Thermal stabilities (ΔG37) of the self-complementary oligonucleotide 5'-PCGCGCG-3', in which the dangling nucleoside P is unpaired but can overlap with the adjacent terminal base pair, were determined (Table 1). In such studies, dangling residue nucleobase polarizability and surface area, which affect both π-stacking and shielding of the duplex from solvent, robustly correlate to dangling nucleoside-induced duplex stabilization.[35,55] The sequence tested has a G-C content that precludes minor groove binding of the dangling residue[35] and is known to form a B-DNA duplex structure at NaCl concentrations below 2.5 M.[35–39,54,55] Among the hydrophobic nucleoside analogues, ExBenzi, ExBIM, and Peri stabilized the duplex the most, while Per stabilized it more moderately. Notably, Benzi and BIM, which have similar size and shape as purine nucleosides, stabilized the duplex to a similar extent as A and G.[35,40]

Overall, our results are consistent with previous reports that larger nucleobases generally π-stack more strongly since they have more surface area of overlap with neighboring bases, which has been specifically demonstrated in studies evaluating the stacking capabilities of benzene, naphthalene, and pyrene nucleosides[35,40] or comparing A, C, and T to their elongated counterparts xA, xC, and xT (Table 1).[36,37] The lower dangling-end stabilization by Per vs. Peri may result from the carbonyl group on Per inducing a conformation less favourable for dangling-end π-stacking. Indeed, the shape and orientation of bases, in addition to size, were important factors in data from similar experiments involving pyrene nucleoside and a larger porphyrin nucleoside.[56]

Table 1: Dangling-end stabilization values.

Category	P[a]	ΔG [kcal/mol][b]
--	--	-8.9 ± 0.1 (-8.1 ± 0.1[35])
Hydrophobic	Benzi	-10.9 ± 0.2
	BIM	-10.2 ± 0.1
	Per	-10.5 ± 0.1
	Peri	-11.7 ± 0.1
Elongated hydrophobic	ExBenzi	-11.3 ± 0.2
	ExBIM	-11.8 ± 0.3
Canonical	C	-9.6 ± 0.1 (-8.9 ± 0.1[35])
	T	-9.9 ± 0.1 (-9.2 ± 0.1[35])
	A	-10.5 ± 0.1 (-10.0 ± 0.1[35])
	G	-10.5 ± 0.1 (-9.9 ± 0.1[35])
Other non-canonical	Benzene	-9.4 ± 0.7[35]
	Naphthalene	-10.9 ± 0.1[35]
	Pyrene	-11.3 ± 0.3[35]
	xC	-10.1 ± 0.3[37]
	xT	-11.1 ± 0.3[36]
	xA	-12.1 ± 0.2[36]

[a] Duplex = 5'-PCGCGCG-3' : 3'-GCGCGCP-5', in which P = designated nucleoside. [b] ΔG = free energy of melting at 37 °C. Errors indicate standard error from a minimum of 5 replicates.

Elongated nucleosides specifically stabilize O^6-MeG

To test the specificity of ExBenzi and ExBIM for preferentially forming base pairs opposite O^6-alkylG as opposed to other DNA lesions, we measured the melting temperatures of DNA duplexes containing ExBenzi or ExBIM paired opposite a broad variety of DNA lesions including alkylated, oxidized, or abasic DNA (O^6-MeG, N2-MeG, 8-oxoG, N6-MeA, or THF, Figure 3). The elongated nucleoside analogues were incorporated into hybridization probes complementary to Codon 13 of the human KRAS gene,[59] which is a hotspot for mutation in cancers (5'-CCTACGPCACCAG-3', P=ExBIM or Exbenzi). The melting data indicate that ExBIM and ExBenzi preferentially stabilized the duplexes when paired opposite O^6-MeG (Figure 3, Table S2) as opposed to other natural and modified nucleobases. The results complemented those obtained for the nucleoside analogues paired opposite G or O^6-MeG in a different sequence context (Figure 2) and were consistent with melting data determined previously for the same sequence.[46] For ExBenzi, the difference in melting temperature between the most stable duplex (O^6-MeG: ExBenzi) and second-most stable duplex (N2-MeG: ExBenzi) was 7.6 ± 0.8 °C (Figure 3A), while for ExBIM, the corresponding difference between the most stable duplex (O^6-MeG: ExBIM) and second-most stable duplex (THF:ExBIM) was 3.0 ± 1.4 °C (Figure 3B). These results indicate on a first level that ExBIM and ExBenzi discriminate between O^6-MeG and different modified DNA bases, and more specifically that the robustness of their selectivity stems from the unusually high stability of the O^6-MeG: ExBenzi/ExBIM duplexes (Tm 55.3 ± 0.6 and 53.0 ± 1.0 °C, respectively, Figure 3). Their high relative stabilities are remarkable considering that in this sequence context, O^6-Me is inherently more destabilizing than the other lesions, according to the melting temperatures of duplexes containing the lesions paired opposite their cognate canonical bases (O^6-Me:C = 50.7 ± 0.6, N2-MeG: C= 61.7 ± 0.6, 8-oxoG: C = 57.3 ± 0.6, and N6-MeA: T = 51.4 ± 0.6 °C) (Table S2).

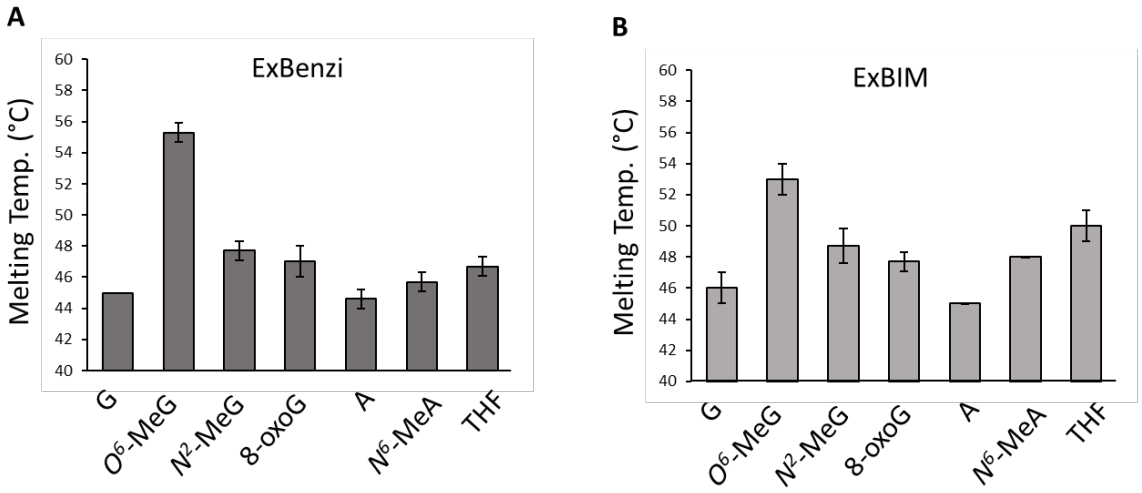


Figure 3: Comparison of melting temperatures of DNA duplexes consisting of 5'-CCTACGPCACCAG-3' paired opposite 3'-GGATGCXGTGGTC-5', in which X = the indicated natural or modified nucleobase and for A) P= ExBenzi and B) P= ExBIM. Error bars indicate standard deviation (n = 3).

MD simulations support increased stacking interactions of elongated base analogues

To further explore physical properties driving the duplex stability associated with ExBenzi and ExBIM, classical molecular dynamics (MD) simulations were carried out on model 13-mers containing the nucleoside analogues paired opposite O^6-MeG or G. When considering the relative contributions of hydrogen bonding and base stacking, the MD simulations implicated base-stacking interactions as the main influence on selectivity. In particular, O^6-methylguanine was calculated to more strongly stack with flanking 3' and 5' nucleobases when paired opposite either ExBenzi or ExBIM (lesion stacking energies of -17.4 and -20.0 kcal/mol, respectively) in a syn-conformation with the naphthalene-portion of the nucleobase projecting out of the duplex into the major groove, and allowing the heterocyclic portion of the nucleobase to hydrogen bond with the target nucleoside. When guanine was paired opposite either ExBenzi or ExBIM, the stacking energies were significantly reduced (-10.6 and -12.8 kcal/mol, respectively). Thus, lesion stacking interactions with neighboring bases may contribute to the increased stability of duplexes containing base pairs formed between the nucleoside analogues and O^6-methylG rather than G.

Base-stacking interactions, particularly between the nucleoside analogue and its flanking 3' and 5' nucleobases, appear to play a strong role in determining the relative stability of O^6-MeG:analogue pairs. The calculated base analogue stacking energies (Table 2) reflect the relative stabilities of O^6-methylG:analogue pairs, with stronger base analogue stacking energy calculated for the O^6-methylG:ExBenzi pair (-16.9 kcal/mol) than for the $O^6$6-methylG:ExBIM pair (-11.3 kcal/mol). The increased base stacking energies complement the effect of hydrogen bonding in strengthening O^6-methylG:Exbenzi pairs in comparison to O^6-methylG:ExBIM pairs. Additionally, the MD simulations indicate that ExBIM opposite O^6-MeG has the weakest hydrogen-bonding interaction, forming only one hydrogen bond throughout the MD trajectory. In contrast, ExBenzi opposite O^6-MeG forms two hydrogen bonds (Table 2, Sensitive fluorescent nucleosides for discriminating apurinic / apyrimidinic sites in DNA duplex Figure 4); however, hydrogen-bonding interactions do not appear to entirely account for duplex stabilization by elongated probes, as evidenced by experimental data showing that ExBIM:G is more stable than ExBenzi:G despite weaker hydrogen bonding (Table 2, Figure 4). In this case, guanine stacking energies may also influence on the base-pair stability, with values of -12.8 kcal/mol calculated for the ExBIM:G pair and -10.6 kcal/mol calculated for the ExBenzi:G pair. Thus, a combination of experimental and computational data suggests that the basis of selective formation of stable base pairs between nucleoside analogues and O^6-MeG, rather than G, derives from a combination of hydrogen bonding and base-stacking interactions, with the relative contribution of each of these factors being context-dependent.

Table 2: Interactions between ExBIM or ExBenzi and G or O^6-MeG. Reported values for each hydrogen bond include hydrogen-bonding occupancies, average donor–acceptor distance and average donor–hydrogen–acceptor angle calculated over the entire production run.

	Hydrogen-bonding energy (kcal/mol)	Base analogue stacking energy (kcal/mol)	Stacking energy (kcal/mol)
ExBIM:G	-10.6±2.0	-11.6±2.7	-12.8±4.3
ExBIM: O^6-MeG	-4.6±1.6	-11.3±2.8	-20.0±4.1
ExBenzi:G	-18.5±2.9	-15.5±2.4	-10.6±4.8
ExBenzi:O^6-MeG	-11.0±2.3	-16.9±2.4	-17.4±4.0

[a] Energies calculated as the sum of the van der Waals and electrostatic interactions between different nucleobases using the LIE command in CPPTRAJ; standard deviations indicate variance from the dynamics of the trajectory.

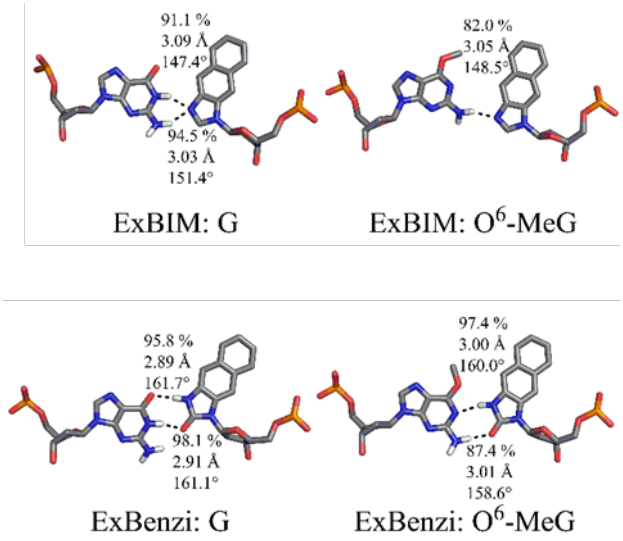


Figure 4: Interactions between ExBIM or ExBenzi and G or O^6-MeG. Reported values for each hydrogen bond include hydrogen-bonding occupancies, average donor–acceptor distance and average donor–hydrogen–acceptor angle calculated over the entire production run.

ExBIM fluorescence sensitivity opposite O^6-MeG

Fluorescent modified nucleobases are commonly applied to study nucleic acids. For example, fluorescent nucleoside analogues have been incorporated into hybridization probes for detection of 8-oxoG and abasic sites.[18,19,21–23] A major advantage of employing fluorescent modified nucleobases is that the sensitivity of fluorescence spectroscopy is much higher compared to UV spectroscopy. We compared the fluorescence properties of the nucleoside probes and found that BIM had the highest quantum yield, followed by ExBenzi, ExBIM, and Benzi. All nucleoside analogues tested herein had emission maxima in the UV range; those with larger surface area (ExBIM and ExBenzi) emitted at higher wavelengths compared to the smaller nucleosides BIM and Benzi. Per and Peri nucleosides were not measurably fluorescent. The quantum yield of Benzi is very similar as previously reported for its free nucleobase, benzimidazolone.[60]

The fluorescent properties of ExBenzi and ExBIM inspired us to explore their use as environment-sensitive DNA adduct probes. Therefore, we measured the fluorescence intensities of duplexes containing ExBenzi and ExBIM paired opposite different DNA modifications (Figure 5) as well as single-stranded DNA (ssDNA) containing ExBenzi and ExBIM. We found that the fluorescence intensities of ssDNA containing each analogue were 100-fold lower than those of the ExBenzi and ExBim free nucleosides. Indeed, there are many examples of fluorescent base analogues that are quenched when incorporated into DNA, for example 2AP.[23] It is also known that fluorescent nucleosides are quenched by G, both in ssDNA and in duplex DNA, when G is opposite or adjacent to the fluorescent nucleoside.[21,61–63] Upon hybridization of ssDNA containing ExBIM to complementary strands containing G, O^6-MeG, or any other modified base, fluorescence was quenched relative to the ssDNA (Figure S1, SI). For ExBenzi, fluorescence intensity did not significantly differ between the ssDNA and the corresponding target duplexes.

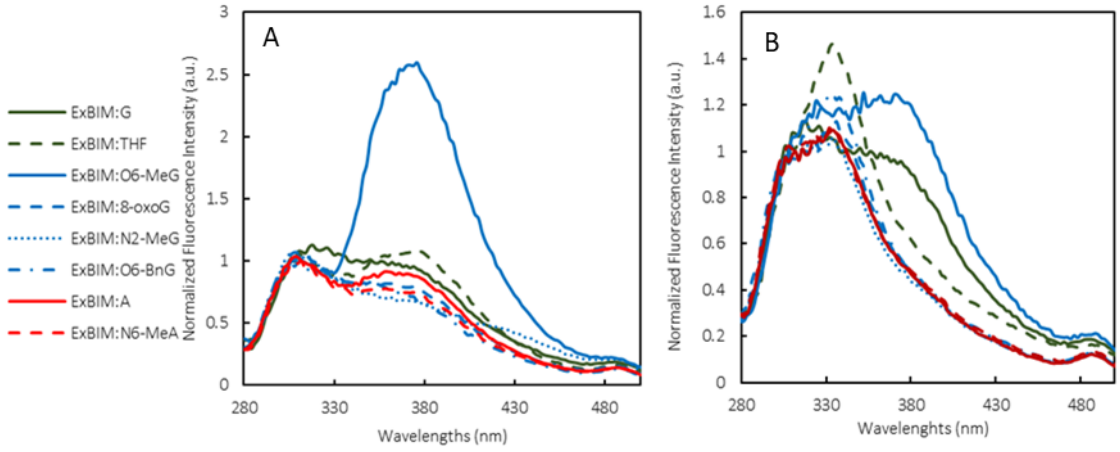


Figure 5: Fluorescence emission spectra of duplexes containing A) ExBIM and B) ExBenzi opposite modified Bases. Fluorescence spectra were normalized to the water Raman peak at 310 nm.

Duplexes containing ExBenzi or ExBIM paired opposite O^6-MeG showed unique fluorescence profiles compared to duplexes containing the analogues paired opposite other lesions or canonical nucleosides (Figure 5). Notably, the peak maxima shifted from 330 nm for the ssDNA to 370 nm when ExBenzi or ExBIM was paired opposite O6-MeG. Moreover, the ExBIM: O^6-MeG duplex had 2-4-fold higher fluorescence intensity at 370 nm than did the other duplexes tested; this effect was not seen for the ExBenzi duplexes. These results highlight the potential of ExBIM as a fluorescent probe to selectively discriminate between O^6-MeG and the other modified bases.

Table 3: Measured quantum yields Φ_f[a]

Compound (solvent)	**Φ_f** [b]	**λem, max**
Benzi (10% DMSO)	0.35	309 nm
BIM (10% DMSO)	0.82	297 nm
ExBenzi (10% DMSO)	0.69	338 nm
ExBIM (10% DMSO)	0.44	364 nm
Tryptophan (water)	0.10 [c]	346 nm
2-Aminopyridine (0.05 M H_2SO_4)	0.71 [d]	358 nm

[a] Obtained relative to standards tryptophan and 2-aminopyridine. [b] Excitation wavelength 250 nm. [c]Lit. value = 0.12[60]. dLit. value = 0.60[64]

The thermodynamic stability of duplexes afforded by the elongated nucleobases does not directly correlate to their selective fluorescence enhancement when paired opposite O^6-MeG. However, the quantum yield of fluorescent base analogues is highly sensitive to their surroundings including hydrogen bonding, single/double-stranded environment, neighbouring base, and opposite base.[65] The decrease in quenching observed by ExBIM when paired to O^6-MeG may be a combined result of the decreased hydrogen bonding in the analogue pair (Table 2, Figure 4), changes in base stacking of the base analogue/pairing base (Table 2), and lower average solvent accessible surface area of ExBIM (SASA, Figure S4) predicted according to the MD simulations described above. For example, decreased hydrogen bonding with the opposing strand and a lower solvent accessible surface area suggests that ExBIM may be better shielded from the discrete interactions with the pairing base in DNA and bulk solvent, resulting in an increase in the fluorescence intensity.

Conclusions

In this study, two large hydrophobic nucleobase analogues (ExBIM and ExBenzi) were incorporated into oligonucleotides to test the contribution of π-surface area on the stabilization of O^6MeG adducts in a duplex. The results indicated that ExBenzi was the nucleoside analogue most selective in distinguishing between G, alkylated purines, and other DNA adducts, and its performance exceeded those of any previously reported adduct-directed probe analogues.[31,34] These data supported the hypothesis that increasing the π-stacking capacity of hydrophobic base analogues increased the stability of duplexes containing O^6-alkylG rather than G. Although ExBIM was not as effective as ExBenzi at preferentially stabilizing O^6-alkylG versus G duplexes, it did prove more valuable as a fluorophore for discriminating between O^6-MeG and G, as well as other modified bases, in the context of duplex DNA. In addition to increasing our understanding of factors that impact

the stability of chemically modified DNA, data from the present study provides insight on discrete chemical interactions of hydrophobic base analogues with modified DNA bases as a basis of improved probe design and development for detecting DNA damage.

Experimental Section

UV absorption was measured at 260 nm using a Cary 100 UV-Vis spectrophotometer equipped with a Peltier thermal programmer. Absorbance spectra and fluorescence intensities were recorded on a Tecan infinite M200 Pro platereader. The Per, ExBenzi, and ExBIM nucleosides were prepared and incorporated into oligonucleotides as previously described[31,46]. For thermodynamic studies on duplexes containing nucleoside analogue:probe pairs, the sequence 5'-CCGPTATACCGACAA-3' (P = ExBenzi or ExBIM) was selected for the hybridization probe, while the complementary sequence 3'-GGCXATATGGCTGTT-5' (X = G, O^6-MeG, O^6-BnG, or THF) was used for the target strands. For dangling-end studies, the sequence 5'-PCGCGCG-3' (P = natural nucleoside or nucleoside analogue) was used. For the thermal melting and fluorescence studies on probe specificity the sequence for Codon 13 of the human KRAS 5'-CCTACGPCACCAG-3' (P=ExBIM or Exbenzi) and the complementary strand 5'-CTGGTGXCGTAGG-3' (X = G, O^6-MeG, or THF) was used.

Thermodynamic studies

The melting of DNA duplexes was observed using UV spectroscopy. For non-self-complementary duplexes, solutions of duplex DNA (in 0.25 M NaCl, 0.2 mM EDTA, 20 mM sodium phosphate, pH 7.0 buffer) were heated (in Teflon- stoppered 1 cm path-length quartz cuvettes) to 80 ºC, then cooled to 25 ºC and held at that temperature for 5 minutes to anneal the duplex; then, the heating cycle was repeated to obtain the melting curve for data analysis. Two different heating/cooling rates (0.75 ºC/min and 1 ºC/min) were tested; no significant differences in transition temperatures due to these different rates were observed, so data obtained using either heating/cooling rate were pooled. For each probe:target duplex, measurements of at least five different concentrations (ranging from 0.8 to 10 μM) were taken. Hystereses were observed between melting and cooling curves; only melting curves were analyzed.
For dangling-end experiments, DNA solutions (of single-strand concentrations ranging from 12 to 40 μM) were prepared in deionized water, then combined with equal volumes of buffer (2 M NaCl, 0.2 mM EDTA, 20 mM sodium phosphate, pH 7.0) to generate sample solutions (with duplex concentrations ranging from 3 to 10 μM). These samples were heated (in Teflon- stoppered 1 cm path-length quartz cuvettes) to 90 ºC, then cooled to 20 ºC and held at that temperature for 25-45 min to anneal the duplex; then, the heating cycle was repeated once more to obtain a melting curve for analysis. At least five samples per duplex were analyzed.

Calculating thermodynamic parameters by analyzing the shape of the melting curve (curve fit method).

Thermodynamic parameters were derived using the hyperchromicity method available in the Cary Thermal application software. Briefly, the software allows the user to select upper and linear baselines for the melting curve and subsequently graphs the fraction of annealed duplex (i.e., fraction-folded, or α) vs temperature. The linear region of the fraction-folded curve was designated between α = 0.4-0.6 for all curves. The linear region of the subsequently plotted ln Ka vs 1/T curve was selected between values of 1/T in which T corresponded to the upper and lower temperatures designated in the linear regions of the fraction-folded curve. Error values reported for thermodynamic parameters obtained with this method represent the standard deviation of the mean values obtained for each set of probe:target duplexes.

Statistical Analysis

The ΔG° values were analyzed using SPSS Statistics 17.0 software. The level of significance for each analysis was selected to be $p<0.05$. Data sets from each probe:target duplex passed the Shapiro-Wilk test for normality; however, the data collectively failed the Levene test for homogeneity of variances (Levene statistic = 3.354, df1 = 11, df 2 = 174, sig. = 0.000). Thus, it was not possible to use the commonly-used ANOVA method for comparing the statistical differences between data sets. Instead, the Welch test for equality of means was performed; results indicated that there was a significant difference between data sets (statistic = 65.422, df1 = 11, df2 = 57.234, sig. = 0.000). Finally, the post hoc Games-Howell test was performed to determine the significance of differences between data sets.

Duplex melting temperatures

Melting temperatures (Tm) of the duplexes were determined by variable temperature UV analysis using a Cary 300-Bio UV-vis spectrophotometer equipped with a 6x6 Multicell Peltier block-heating using Hellma 114-QS 10 mm light path cells. Modified or unmodified oligonucleotide were mixed with primer strands at a final concentration of 2.5 μM of each template and primer in 1000 μL PBS. The heating cycle (0.4 °C/min) then was repeated and UV absorption monitored at 280 nm. Melting temperatures (Tm) were determined by the first-derivative method with the application in the Cary Thermal software. All measurements were performed three times and the mean values (± standard deviation) are reported.

Fluorescence spectra of nucleosides and oligonucleotides

Fluorescence spectra were recorded on a Tecan infinite M200 Pro plate reader in 96-well UV-Star microplates (Greiner). Samples were prepared at 2.5 µM each template and primer sequence in 500 µL PBS and heated for 5 min at 95 °C in a heating block and slowly cooled down in the block. Spectra were recorded at an excitation wavelength of 244 nm (bandwidth: 5 nm, gain: 132 and Z-position: 15736) and measured from 280 to 600 nm.

To characterize the fluorescence of the hydrophobic nucleoside analogues exBIM and exBenzi, we determined their quantum yields (Φf) under neutral and acidic conditions.[66–68] Quantum yields (Φf) of nucleoside analogues in 10% DMSO/90% H2O were determined relative to standards tryptophan (Serva Feinbiochemica) in water (pH 7) and 2-aminopyridine (Sigma Aldrich) in 0.05 M H2SO4 (aq) using the comparative method of Williams et al.50 The quantum yield was calculated with the following equation:

$$\phi_x = \phi_{ST}\left(\frac{Grad_x}{Grad_{ST}}\right)\left(\frac{\eta_x^2}{\eta_{ST}^2}\right)$$

in which Φx = Φf of sample, ΦST = literature Φf value of standard, GradX = slope of sample integrated emission intensities vs absorbance, GradST = slope of standard integrated emission intensities vs absorbance, η2X = refractive index of sample solvent, and η2ST = refractive index of standard solvent. Values for ΦST were 0.12 for tryptophan51 and 0.60 for 2-aminopyridine.[68]

Absorbance of nucleoside analogues and standards was measured at 250 nm and plotted versus the respective concentrations of the samples (at least 5 data points). Theoretical absorbances of sample solutions at lower concentrations (i.e., those with absorbances less than 0.1) were extrapolated from the slope of the best linear fit of the absorbance versus concentration curves. Extrapolated values were needed for cases in which the emission intensity of higher-concentration samples exceeded the detection limit of the spectrometer. To obtain fluorescence emission spectra, samples were excited at 250 nm. To calculate the emission intensity, the area below the curve of relative fluorescence intensity was integrated (calculated using Microsoft Excel). Because the minimum emission wavelength of the platereader was 280 nm, the emission curves of nucleoside analogues BIM and Benzi were cut off at the lower wavelength. Therefore the integrals of fluorescence intensity for BIM and Benzi were calculated differently: the right half of the emission peak was integrated and multiplied by two. Integrated fluorescence intensities were then plotted against calculted absorbance values (at least 5 data points) and the slope of the corresponding linear trendlines were used to calculate the quantum yields with the equation above. The quantum yields of nucleoside analogues were calculated twice, using the experimentally determined slopes of each standard (tryptophan and 2-aminopyridine). Finally, the average of the two quantum yield values from both standards was used as the final value for each nucleoside analogue.

Computational methods

The initial structure of the undamaged 13mer DNA duplex (5′-CTGGTGPCGTAGG) was generated using the NAB module of AMBER 14.[69] The helix was modified at the P position to either the ExBIM or ExBenzi base analogue, resulting in a P:X base pair (P = ExBIM/ExBenzi and X = G or O6-MeG). The 13mer containing ExBenzi was generated such that the orientation of the modified base pair matched the NMR and crystal structures of similar modified bases,[32] while the ExBIM containing duplexes were generated such that the hydrogen-bonding interactions between the base analogue and the pairing base were optimized. Modifications to the natural nucleobases to form the base analogues and the O6-MeG lesion were performed using GaussView. The atom types for the base analogues and damaged nucleotides were generated with the ANTECHAMBER module of Amber 14 using the AMBER14SB[70] force field supplemented with GAFF.[71] Partial charges were calculated using RESP charge fitting at the HF/6-31G(d) level within the R.E.D. program (version III.4).[72] The remainder of the system was treated with the AMBER14SB force field.[70]

The DNA helices were solvated in a TIP3P water box, such that there was a minimum of 10 Å between the solute and the edge of the box, with a charge equilibrating amount of sodium ions added using the tleap module of Amber 14. Each system was minimized over four stages with each stage involving 2500 steps of steepest decent (SD) followed by 2500 steps conjugate gradient (CG). The first stage of minimization was performed with a constraint on all solute atoms, while the second stage was done with constraints on all heavy atoms. Subsequently, a minimization cycle was completed with the solute constrained. All constraints applied during the minimization were 50.0 kcal/(mol Å2). Finally, all constraints were removed and the entire system was converged to an energy minimum.

Heating was divided into six 20 ps phases in order to gradually bring the system to 310 K. Each phase was performed with constant volume and a 25.0 kcal/mol Å2 restraint on the solute. The first heating phase increased the temperature from 0 K to 60 K, while each additional phase raised the temperature by 50 K. Equilibration of the system was subsequently performed using five 20 ps stages, with the constraints on the solute being gradually decreased from 20 kcal/mol Å2 to 1.5 kcal/mol Å2. Specifically, the solute was restrained with 20, 15, 10, 5 and 1.5 kcal/mol Å2 for the five equilibration stages, respectively. Finally, a production run of 100 ns was obtained for each system using a 2 fs timestep.

All heating, equilibration and production steps were performed with a 25 kcal/mol Å2 restraint on the terminal Watson-Crick hydrogen bonds to prevent fraying. During the simulations, a nonbonded cutoff of 8 Å was applied to the system, with the long-range interactions treated using the particle mesh Ewald algorithm. The Langevin thermostat ($\gamma = 3.0$) was used during the heating, equilibration and production phases. The SHAKE algorithm was implemented to constrain the hydrogen atoms and the periodic boundary condition was applied.

The cpptraj module of Amber 14 was used to calculate hydrogen-bonding occupancies, interaction energies, backbone root-mean-square deviation (RMSD), average dihedral angles and the solvent accessible surface area (SASA) of the modified bases. Hydrogen bonding interactions were calculated using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. Linear interaction energies (LIE) were calculated as the stability gained between the base analogue and the pairing nucleobase (denoted base analogue hydrogen-bonding energy), the base analogue and the four flanking bases (denoted base analogue stacking energy), or the pairing base and the four flanking bases (denoted G/ O^6-MeG stacking energy).

Supplementary Material

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/MS-number.

Acknowledgements

This work was funded by the Swiss National Science Foundation (156280), the European Research Council (260341) and the Natural Sciences and Engineering Research Council of Canada (2016-04568). We thank Dr. Arman Nilforoushan for contributions to the synthesis of phosphoramidites.

Author Contribution Statement

H.A.D. and S.J.S. conceived of the project. H.A.D., F.D.B., R.W.K., L.A.W., I.G., M.M. performed experiments. All authors contributed to the design of experiments, interpretation of data, and writing the paper.

References

[1] E. Jensen, 'Technical Review: in situ hybridization', *Anat. Rec.*, **2014**, *297*, 1349–1353.

[2] V. V. Didenko, 'DNA probes using fluorescence resonance energy transfer (FRET): designs and applications', *Biotechniques*, **2001**, *31*, 1106–1121.

[3] E. Koller, W. A. Gaarde, and B. P. Monia, 'Elucidating cell signaling mechanisms using antisense technology', *Trends Pharmacol. Sci.*, **2000**, *21*, 142–148.

[4] J. A. Doudna and E. Charpentier, 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, **2014**, *346*, 1258096-1-1258096–1258099.

[5] M. Bartosiewicz, M. Trounstine, D. Barker, R. Johnston, and a Buckpitt, 'Development of a toxicological gene array and quantitative assessment of this technology.', *Arch. Biochem. Biophys.*, **2000**, *376*, 66–73.

[6] E. A. Meyer, R. K. Castellano, and F. Diederich, *Interactions with aromatic rings in chemical and biological recognition*, in *Angew. Chemie - Int. Ed.*, **2003**, *42*, 1210-1250

[7] D. J. Patel, L. Shapiro, S. A. Kozlowski, B. L. Gaffney, and R. A. Jones, 'Structural studies of the O6meG•C interaction in the d(C-G-C-G-A-A-T-T-C-O6meG-C-G) duplex', *Biochemistry*, **1986**, *25*, 1027–1036.

[8] D. J. Patel, L. Shapiro, S. A. Kozlowski, B. L. Gaffney, and R. A. Jones, 'Structural studies of the O6meG•C interaction in the d(C-G-C-G-A-A-T-T-C-O6meG-C-G) duplex', *Biochemistry*, **1986**, *25*, 1036–1042.

[9] B. L. Gaffney and R. A. Jones, 'Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O6-methylguanine with reference both to Watson-Crick pairs and to mismatched pairs', *Biochemistry*, **1989**, *28*, 5881–5889.

[10] S. S. Hecht, 'DNA adduct formation from tobacco-specific N-nitrosamines', *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, **1999**, *424*, 127–142.

[11] B. D. Reh, D. G. DeBord, M. A. Butler, T. M. Reid, C. Mueller, and J. M. Fajen, 'O(6)-methylguanine DNA adducts associated with occupational nitrosamine exposure', *Carcinogenesis*, **2000**, *21*, 29–33.

[12] E. S. Newlands, M. F. F. Stevens, S. R. Wedge, R. T. Wheelhouse, and C. Brock, 'Temozolomide: a review of its discovery, chemical properties, pre-clinical development and clinical trials', *Cancer Treat. Rev.*, **1997**, *23*, 35–61.

[13] G. P. Margison, M. F. Santibáñez Koref, and A. C. Povey, 'Mechanisms of carcinogenicity / chemotherapy by O6-methylguanine', *Mutagenesis*, **2002**, *17*, 483–487.

[14] B. Rydberg and T. Lindahl, 'Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction.', *EMBO J.*, **1982**, *1*, 211–216.

[15] B. Sedgwick, 'Nitrosated peptides and polyamines as endogenous mutagens in O6-alkylguanine-DNA alkyltransferase deficient cells', *Carcinogenesis*, **1997**, *18*, 1561–1567.

[16] A. M. Woodside and F. P. Guengerich, 'Effect of the O6 substituent on misincorporation kinetics catalyzed by DNA polymerases at O6-methylguanine and O6-benzylguanine', *Biochemistry*, **2002**, *41*, 1027–1038.

[17] L. M. Wilhelmsson, 'Fluorescent nucleic acid base analogues.', *Q. Rev. Biophys.*, **2010**, *43*, 159–183.

[18] N. J. Greco, R. W. Sinkeldam, and Y. Tor, 'An emissive C analog distinguishes between G, 8-oxoG, and T', *Org. Lett.*, **2009**, *11*, 1115–1118.

[19] N. J. Greco and Y. Tor, 'Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites', *J. Am. Chem. Soc.*, **2005**, *127*, 10784–10785.

[20] J. T. Stivers, '2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: Metal-ion and base sequence effects', *Nucleic Acids Res.*, **1998**, *26*, 3837–3844.

[21] Y. Taniguchi, Y. Koga, K. Fukabori, R. Kawaguchi, and S. Sasaki, 'OFF-to-ON type fluorescent probe for the detection of 8-oxo-dG in DNA by the Adap-masked ODN probe', *Bioorganic Med. Chem. Lett.*, **2012**, *22*, 543–546.

[22] J. M. Jean and K. B. Hall, '2-Aminopurine fluorescence quenching and lifetimes: role of base stacking', *Proc. Natl. Acad. Sci. U. S. A.*, **2001**, *98*, 37–41.

[23] J. K. Kelley, Shana. O., Barton, 'Electron Transfer Between Bases in Double Helical DNA', *Science*, **1999**, *283*, 375–381.

[24] A. Suzuki, N. Nemoto, I. Saito, and Y. Saito, 'Design of an environmentally sensitive fluorescent 8-aza-7-deaza-2'-deoxyadenosine derivative with dual fluorescence for the specific detection of thymine', *Org. Biomol. Chem.*, **2014**, *12*, 660–666.

[25] Y. Saito, A. Suzuki, Y. Okada, Y. Yamasaka, N. Nemoto, and I. Saito, 'An environmentally sensitive fluorescent purine nucleoside that changes emission wavelength upon hybridization', *Chem. Commun.*, **2013**, *49*, 5684–5686.

[26] Y. Saito, A. Suzuki, T. Yamauchi, and I. Saito, 'Design and synthesis of 7-naphthyl-8-aza-7-deaza-2′-deoxyadenosines as environmentally sensitive fluorescent nucleosides', *Tetrahedron Lett.*, **2015**, *56*, 3034–3038.

[27] T. Yamauchi, T. Takeda, M. Yanagi, N. Takahashi, A. Suzuki, and Y. Saito, 'Sensitive fluorescent nucleosides for discriminating apurinic / apyrimidinic sites in DNA duplex', *Tetrahedron Lett.*, **2017**, *58*, 117–120.

[28] K. Onizuka, T. Nishioka, Z. Li, D. Jitsuzaki, Y. Taniguchi, and S. Sasaki, 'An efficient and simple method for site-selective modification of O6-methyl-2′-deoxyguanosine in DNA', *Chem. Commun.*, **2012**, *48*, 3969–3971.

[29] I. A. Trantakis, S. Bolisetty, R. Mezzenga, and S. J. Sturla, 'Reversible aggregation of DNA-decorated gold nanoparticles controlled by molecular recognition', *Langmuir*, **2013**, *29*, 10824–10830.

[30] I. A. Trantakis and S. J. Sturla, 'Gold nanoprobes for detecting DNA adducts', *Chem. Commun.*, **2014**, *50*, 15517–15520.

[31] J. Gong and S. J. Sturla, 'A synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA', *J. Am. Chem. Soc.*, **2007**, *129*, 4882–4883.

[32] E. A. Kowal, R. R. Lad, P. S. Pallan, E. Dhummakupt, Z. Wawrzak, M. Egli, S. J. Sturla, and M. P. Stone, 'Recognition of O6-benzyl-2'-deoxyguanosine by a perimidinone-derived synthetic nucleoside: A DNA interstrand stacking interaction', *Nucleic Acids Res.*, **2013**, *41*, 7566–7576.

[33] H. L. Gahlon, M. L. Boby, and S. J. Sturla, 'O6-Alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding', *ACS Chem. Biol.*, **2014**, *9*, 2807–2814.

[34] H. L. Gahlon and S. J. Sturla, 'Hydrogen bonding or stacking interactions in differentiating duplex stability in oligonucleotides containing synthetic nucleoside probes for alkylated DNA', *Chem. - A Eur. J.*, **2013**, *19*, 11062–11067.

[35] K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, D. C. Tahmassebi, and E. T. Kool, 'Factors contributing to aromatic stacking in water: Evaluation in the context of DNA', *J. Am. Chem. Soc.*, **2000**, *122*, 2213–2222.

[36] J. Gao, H. Liu, and E. T. Kool, 'Expanded-size bases in naturally sized DNA: Evaluation of steric effects in Watson-Crick pairing', *J. Am. Chem. Soc.*, **2004**, *126*, 11826–11831.

[37] H. Liu, J. Gao, and E. T. Kool, 'Size-expanded analogues of dG and dC: Synthesis and pairing properties in DNA', *J. Org. Chem.*, **2005**, *70*, 639–647.

[38] H. Lu, K. He, and E. T. Kool, 'yDNA: A new geometry for size-expanded base pairs', *Angew. Chemie - Int. Ed.*, **2004**, *43*, 5834–5836.

[39] A. H. F. Lee and E. T. Kool, 'Novel benzopyrimidines as widened analogues of DNA bases', *J. Org. Chem.*, **2005**, *70*, 132–140.

[40] K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi, and E. T. Kool, 'Experimental measurement of aromatic stacking affinities in the context of duplex DNA', *J. Am. Chem. Soc.*, **1996**, *118*, 8182–8183.

[41] A. K. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz, and F. E. Romesberg, 'Efforts toward the expansion of the genetic alphabet: Information storage and replication with unnatural hydrophobic base pairs', *J. Am. Chem. Soc.*, **2000**, *122*, 3274–3287.

[42] Y. Wu, A. K. Ogawa, M. Berger, D. L. McMinn, P. G. Schultz, and F. E. Romesberg, 'Efforts toward expansion of the genetic

alphabet: Optimization of interbase hydrophobic interactions', *J. Am. Chem. Soc.*, **2000**, *122*, 7621–7632.

[43] C. Brotschi, A. Häberli, and C. J. Leumann, 'A stable DNA duplex containing a non-hydrogen-bonding and non-shape-complementary base couple: Interstrand stacking as the stability determining factor', *Angew. Chemie - Int. Ed.*, **2001**, *40*, 3012–3014.

[44] K. M. Guckian, J. C. Morales, and E. T. Kool, 'Structure and base pairing properties of a replicable nonpolar isostere for deoxyadenosine', *J. Org. Chem.*, **1998**, *63*, 9652–9656.

[45] T. Angelov, H. A. Dahlmann, and S. J. Sturla, 'Oligonucleotide probes containing pyrimidine analogs reveal diminished hydrogen bonding capacity of the DNA adduct O6-methyl-G in DNA duplexes', *Bioorganic Med. Chem.*, **2013**, *21*, 6212–6216.

[46] I. A. Trantakis, A. Nilforoushan, H. A. Dahlmann, C. K. Stäuble, and S. J. Sturla, 'In-Gene Quantification of O6-Methylguanine with Elongated Nucleoside Analogues on Gold Nanoprobes', *J. Am. Chem. Soc.*, **2016**, *138*, 8497–8504.

[47] M. Berger, A. K. Ogawa, D. L. McMinn, Y. Wu, P. G. Schultz, and F. E. Romesberg, 'Stable and selective hybridization of oligonucleotides with unnatural hydrophobic bases', *Angew. Chemie - Int. Ed.*, **2000**, *39*, 2940–2942.

[48] G. T. Hwang, Y. Hari, and F. E. Romesberg, 'The effects of unnatural base pairs and mispairs on DNA duplex stability and solvation', *Nucleic Acids Res.*, **2009**, *37*, 4757–4763.

[49] L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, and F. E. Romesberg, 'Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications', *J. Am. Chem. Soc.*, **2014**, *136*, 826–829.

[50] T. Lavergne, M. Degardin, D. A. Malyshev, H. T. Quach, K. Dhami, P. Ordoukhanian, and F. E. Romesberg, 'Expanding the scope of replicable unnatural DNA: Stepwise optimization of a predominantly hydrophobic base pair', *J. Am. Chem. Soc.*, **2013**, *135*, 5408–5419.

[51] K. Betz, M. Kimoto, K. Diederichs, I. Hirao, and A. Marx, 'Structural Basis for Expansion of the Genetic Alphabet with an Artificial Nucleobase Pair', *Angew. Chemie - Int. Ed.*, **2017**, *56*, 1–5.

[52] M. Kimoto and I. Hirao, 'Unique Thermal Stability of Unnatural Hydrophobic Ds Bases in Double-Stranded DNAs', *ACS Synth. Biol.*, **2017**, *6*, 1944–1951.

[53] A. K. Mishra, H. Weissman, E. Krieg, K. A. Votaw, M. McCullagh, B. Rybtchinski, and F. D. Lewis, 'Self-Assembly of Perylenediimide-Single-Strand-DNA Conjugates: Employing Hydrophobic Interactions and DNA Base-Pairing To Create a Diverse Structural Space', *Chem. - A Eur. J.*, **2017**, *23*, 1–11.

[54] J. S. Lai and E. T. Kool, 'Selective Pairing of Polyfluorinated DNA Bases', *J. Am. Chem. Soc.*, **2004**, *126*, 3040–3041.

[55] H. Rosemeyer, F. Seela, S. M. Freier, B. J. Burger, D. Alkema, T. Neilson, D. H. Turner, D. H. Turner, N. Sugimoto, R. Kierzek, S. D. Dreiker, P. J. Romaniuk, D. W. Hughes, R. J. Gregoire, T. Neilson, R. A. Bell, N. Sugimoto, R. Kierzek, D. H. Turner, S. M. Freier, D. Alkema, A. Sinclair, T. Neilson, T. H. Turner, M. Senior, R. A. Jones, K. J. Breslauer, E. T. Kool, S. Bommarito, N. Peyret, J. J. SantaLucia, T. Sugiyama, E. Schweinberger, Z. Kazimierczuk, N. Ramzaeva, H. Rosemeyer, F. Seela, M. J. Doktycz, T. M. Paner, M. Amaratunga, A. S. Benight, H. Grosjean, D. G. Söll, D. M. Crothers, S. Ball, M. A. Reeve, P. S. Robinson, F. Hill, D. M. Brown, D. Loakes, J. C. Williams, S. C. Case-Green, K. U. Mir, E. M. Southern, Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci, B. C. Froehler, K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi, E. T. Kool, L. E. Xodo, G. Manzini, F. Quadrifoglio, G. A. van der Marel, J. H. van Boom, T. J. Povsic, P. B. Dervan, Y. S. Sanghvi, G. D. Hoke, S. M. Freier, M. C. Zounes, C. Gonzales, L. Cummins, H. Sasmor, P. D. Cook, S. Wang, E. T. Kool, B. C. Froehler, S. Wadwani, T. J. Terhorst, S. R. Gerrard, J. Sagi, A. Szemzo, K. Ebinger, A. Szabolcs, G. Sagi, E. Ruff, L. Ötvös, L. C. Sowers, B. R. Shaw, W. D. Sedwick, K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi, E. T. Kool, F. Seela, M. Zulauf, N. Ramzaeva, F. Seela, F. Seela, G. Becher, F. Seela, H. Berg, H. Rosemeyer, N. Ramzaeva, C. Mittelbach, F. Seela, N. Ramzaeva, F. Seela, J. A. McDowell, D. H. Turner, F. Seela, R. Kröschel, Y. He, F. Seela, and S. Lampe, 'Modified purine nucleosides as dangling ends of DNA duplexes: the effect of the nucleobase polarizability on stacking interactions', *J. Chem. Soc. Perkin Trans. 2*, **2002**, *22*, 746–750.

[56] H. Morales-Rojas and E. T. Kool, 'A Porphyrin C -Nucleoside Incorporated into DNA', *Org. Lett.*, **2002**, *4*, 4377–4380.

[57] S.-I. Nakano, M. Fujii, and N. Sugimoto, 'Use of nucleic Acid analogs for the study of nucleic Acid interactions.', *J. Nucleic Acids*, **2011**, *2011*, 967098.

[58] J. Isaksson and J. Chattopadhyaya, 'A uniform mechanism correlating dangling-end stabilization and stacking geometry', *Biochemistry*, **2005**, *44*, 5390–5401.

[59] A. K. Arrington, E. L. Heinrich, W. Lee, M. Duldulao, S. Patel, J. Sanchez, J. Garcia-Aguilar, and J. Kim, 'Prognostic and predictive roles of KRAS mutation in colorectal cancer', *Int. J. Mol. Sci.*, **2012**, *13*, 12153–12168.

[60] Z. Lazar, B. Benali, K. Elblidi, M. Zenkouar, B. Lakhrissi, M. Massoui, B. Kabouchi, and C. Cazeau-Dubroca, 'Photophysical Study of Benzimidazolone and its Derivative Molecules in Solution', *J. Mol. Liq.*, **2003**, *106*, 89–95.

[61] C. Dohno and I. Saito, 'Discrimination of single-nucleotide alterations by G-specific fluorescence quenching', *ChemBioChem*, **2005**, *6*, 1075–1081.

[62] C. A. M. Seidel, A. Schulz, and M. H. M. Sauer, 'Nucleobase-Specific Quenching of Fluorescent Dyes. 1. Nucleobase One-Electron Redox Potentials and Their Correlation with Static and Dynamic Quenching Efficiencies', *J. Phys. Chem.*, **1996**, *100*, 5541–5553.

[63] Y. Shinohara, K. Matsumoto, K. Kugenuma, T. Morii, Y. Saito, and I. Saito, 'Design of environmentally sensitive fluorescent 2'-deoxyguanosine containing arylethynyl moieties: Distinction of thymine base by base-discriminating fluorescent (BDF) probe', *Bioorganic Med. Chem. Lett.*, **2010**, *20*, 2817–2820.

[64] H. Lee, E. Luna, M. Hinz, J. J. Stezowski, A. S. Kiselyov, and R. G. Harvey, 'Synthesis of oligonucleotide adducts of the bay region diol epoxide metabolites of carcinogenic polycyclic aromatic hydrocarbons', *J. Org. Chem.*, **1995**, *60*, 5604–5613.

[65] J. N. Wilson, Y. Cho, S. Tan, A. Cuppoletti, and E. T. Kool, 'Quenching of fluorescent nucleobases by neighboring DNA: The "insulator" concept', *ChemBioChem*, **2008**, *9*, 279–285.

[66] A. T. R. Williams, S. A. Winfield, and J. N. Miller, 'Relative fluorescence quantum yields using a computer-controlled luminescence spectrometer', *Analyst*, **1983**, *108*, 1067–1071.

[67] R. F. Chen, 'Fluorescence Quantum Yields of Tryptophan and Tyrosine', *Anal. Lett.*, **1967**, *1*, 35–42.

[68] R. R. and T. A.C., '2-Aminopyridine as a Standard for Low-Wavelength Spectrofluorimetry', *J. Phys. Chem.*, **1968**, *72*, 2680–2681.

[69] D. A. Case, V. Babin, J. T. Berryma, R. M. Betz, Q. Cai, D. S. Cerrutti, T. E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, and E. Al., 'Amber 14 Reference Manual', *Amber 14*, **2014**.

[70] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *J. Chem. Theory Comput.*, **2015**, *11*, 3696–3713.

[71] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. a Kollman, and D. a Case, 'Development and testing of a general amber force field', *J. Comput. Chem.*, **2004**, *25*, 1157–1174.

[72] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, 'The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building', *Phys. Chem. Chem. Phys.*, **2010**, *12*, 7821–7839.

Supporting information

Table S1: Enthalpies and entropies of probe target melting.[a]

P	X	*n* [b]	ΔH_{CF} (kcal/mol)	$T\Delta S_{CF}°$ (kcal/mol)	$\Delta G_{CF}°$ (kcal/mol) [c]	$\Delta\Delta G_{CF}°$ (kcal/mol)[d]
Per	**G**	12	-92.0 ± 0.5	-76.4 ± 0.2	-15.6 ± 0.1	
	***O*^{6}-MeG**	7	-98.6 ± 1.4	-82.0 ± 0.6	-16.6 ± 0.4	-1.0
	***O*^{6}-BnG**	9	-95.4 ± 1.7	-78.9 ± 0.6	-16.6 ± 0.4	-1.0
	THF	10	-94.6 ± 0.8	-79.0 ± 0.3	-15.5 ± 0.2	0.1
Exbenzi	**G**	18	-91.8 ± 1.1	-76.3 ± 1.0	-15.5 ± 0.4	
	***O*^{6}-MeG**	29	-100.0 ± 1.1	-82.9 ± 1.0	-17.2 ± 0.6	-1.7
	***O*^{6}-BnG**	12	-104.8 ± 1.2	-87.1 ± 1.0	-17.8 ± 0.4	-2.2
	THF	18	-97.8 ± 0.7	-81.5 ± 0.6	-16.3 ± 0.3	-0.8
ExBIM	**G**	17	-95.4 ± 0.8	-79.2 ± 0.7	-16.2 ± 0.3	
	***O*^{6}-MeG**	18	-97.8 ± 0.7	-81.1 ± 0.7	-16.7 ± 0.3	-0.5
	***O*^{6}-BnG**	18	-97.3 ± 0.8	-80.4 ± 0.7	-16.9 ± 0.3	-0.7
	THF	18	-99.6 ± 1.0	-82.7 ± 0.9	-17.0 ± 0.4	-0.8

[a] Duplex = 5'-TTGTCGGTATAXCGG-3' complementary to 3'-AACAGCCATATPGCC-5', in which X = nucleosides of G, *O*^{6}-MeG, or *O*^{6}-BnG and P = nucleoside analogue. Values derived from analyzing the shape of the melting curves ("curve-fit" method). Errors represent the standard deviation of the mean values obtained for each set of probe:target duplexes. [b] Number of separately prepared samples. [c] $\Delta G_{CF}° = \Delta H_{CF} - T\Delta S_{CF}°$ (kcal/mol). [b] $\Delta\Delta G_{CF}° = \Delta G_{CF}°$ (probe:adduct duplex) - $\Delta G_{CF}°$ (probe:G duplex).

Table S2: Melting temperatures

Duplex	Tm ºC	ΔTm[a]
G: ExBenzi	45.0 ± 0.006	na
G: ExBIM	46.0 ± 1.0	na
G: C	60.7 ± 0.6	na
O^6-MeG: ExBenzi	55.3 ± 0.6	+10.3
O^6-MeG: ExBIM	53.0 ± 1.0	+7.0
O^6-MeG: C	50.7 ± 0.6	-10.0
N^2-MeG: ExBenzi	47.7 ± 0.6	+2.7
N^2-MeG: ExBIM	48.7 ± 1.1	+2.7
N^2-MeG: C	61.7 ± 0.6	+1.0
8-oxoG: ExBenzi	47.0 ± 1.0	+2
8-oxoG: ExBIM	47.7 ± 0.6	+1.2
8-oxoG: C	57.3 ± 0.6	-3.4
A: ExBenzi	44.6 ± 0.6	na
A: ExBIM	45.0 ± 0.03	na
A: T	55.6 ± 0.6	na
N^6-MeA: ExBenzi	45.7 ± 0.6	+1.1
N^6-MeA: ExBIM	48.0 ± 0.03	+3.0
N^6-MeA: T	51.4 ± 0.6	-3.6
THF: ExBenzi	46.7 ± 0.6	na
THF: ExBIM	50.0 ± 1.0	na

[a] ΔTm=(X:P)-(G:P) when X = modified G and and =(X:P)-(A:P) when X = modified A; P = Exbenzi or ExBIM; na = not applicable.

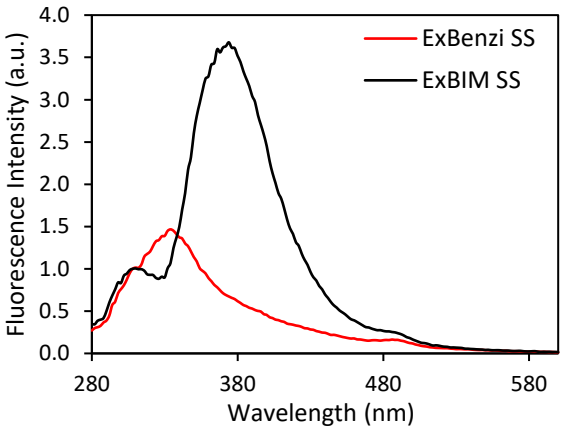


Figure S1: Fluorescence Intensity spectra of ExBIM (black) and ExBenzi (red) single strand

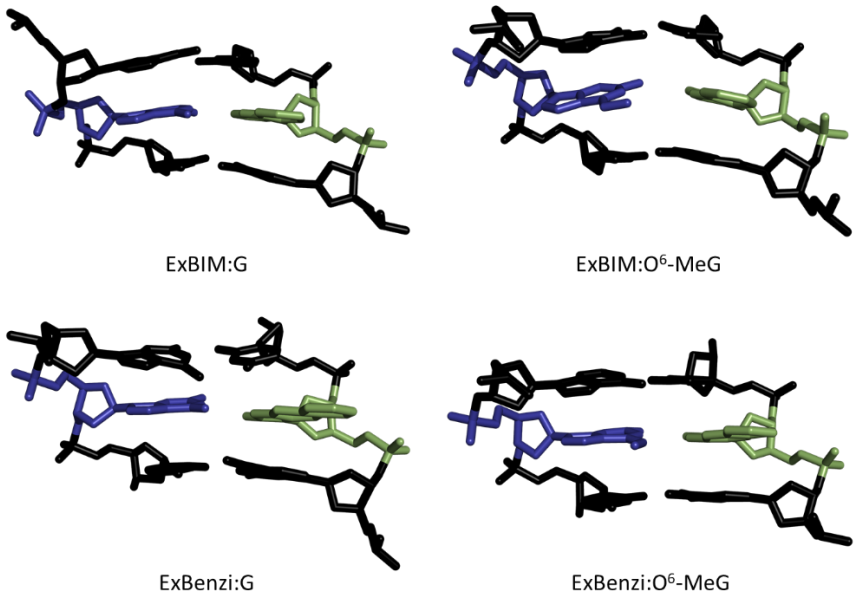


Figure S2: MD representative structures of ExBIM opposite G or O^6-MeG and ExBenzi opposite G or O^6-MeG in DNA duplexes. Only the trimer of base pairs is shown including the base analogue pair and the two flanking base pairs, with ExBIM/ExBenzi in green, G/O^6-MeG in blue and the flanking base pairs in black.

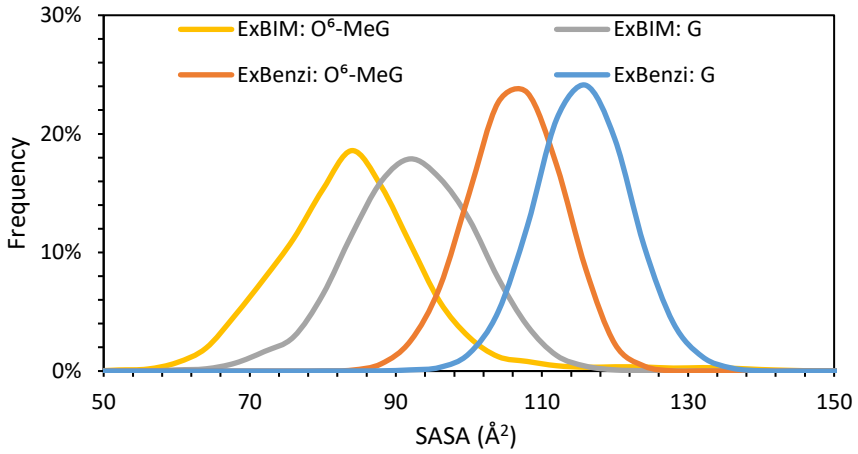


Figure S3: Frequency of the calculated solvent accessible surface area (SASA) of ExBIM and ExBenzi opposite G and O^6-MeG across MD trajectories for modified DNA duplexes.

Appendix A

Incorporation of synthetic triphosphates opposite O^2- and O^4-MeT for exponential amplification

A1 Introduction

Exposure to alkylating agents, found in cigarette smoke,[1–3] red and processed meat[4–7] are known to form alkylated DNA adducts (Figure 1).[4,3,8] In particular, exposure to smoke causes the formation of O^2- and O^4-methylT (MeT) and O^2- and O^4-ethylT (EtT).[2,9–11] In addition, the larger O^2-POBT and O^2-PHBT are formed by the tobacco-specific alkylating agent, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which is converted to methylnitrosamino-1-(3-pyridyl)-1-butanol (NNAL).[3] In lung and liver tissue from rats treated with NNK and NNAL, an accumulation of O^2-POBT and O^2-PHBT was found.[12,13] O^4- and O^2-alkylthymidine (alkylT) are 10 to 100 times less observed than O^6-alkylguanine (alkylG) adducts.[14]

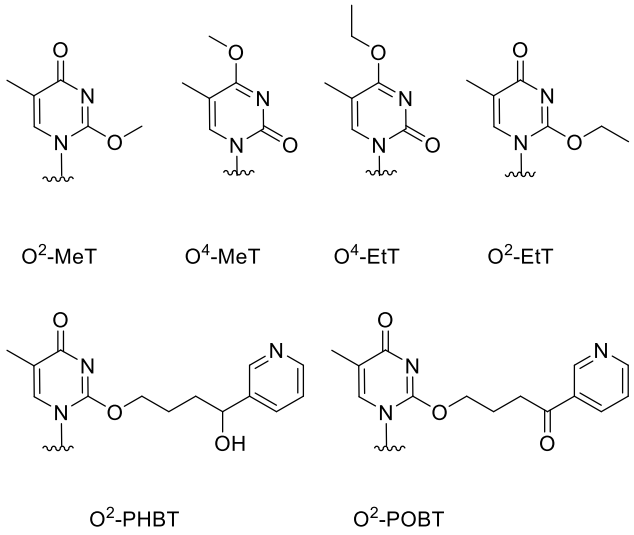


Figure 1: Examples of O^2- and O^4-alkylthymidine DNA adducts.

Despite the lower abundance, alkylT lesions induce mutations. In particular,O^4-MeT and O^4-EtT induce T to C transition mutations.[15–17] Furthermore, these lesions are highly persistence in mammalian tissues, which is related to the poor repair by mammalian enzymes.[18–20] Unlike the *E.coli* Ada repair enzyme[21,22] the mammalian O^6-alkylguanine alkyltransferase (AGT) cannot efficiently repair O^4-alkylT lesions.[18,21–23] Therefore, it is desirable to have methods for early detection of alkylT lesions.

Given that alkyl DNA adducts occur at very low abundancy, sensitive and high throughput methods are required for their detection. Typically, DNA is extracted and hydrolyzed to the nucleobases and analyzed through LC-MS/MS. Unfortunately, the DNA sequence context is lost using this method.[24] Alternatively, the sequence context can be kept if the damage template is extended with a primer sequence containing a dU. After polymerase extension, the product is cleaved at the uracil site to generate small enough DNA pieces that are detectable by mass spectrometry.[25] Another approach uses single molecular real time sequencing (SMRT)[26] with fluorescently phospholinked nucleotides to monitor base incorporation. In this method, the polymerase kinetics are altered in a damaged or artificial nucleobase containing template compared to the unmodified template.[27] Unfortunately, these single-molecule methods require large sample sizes for detection, therefore, amplification of DNA using the polymerase chain reaction (PCR) is required. Unfortunately, PCR cannot retain the damage identity and location. Therefore, in order to specifically detect alkylT lesions in reasonable sample sizes it is desirable to incorporate a marker nucleotide at the modification site, thus allowing PCR amplification.

The recognition of DNA damage by nucleotides requires a specific probe, which is incorporated with high selectivity and efficiency. Nucleoside analogs with different size, π surface area, and hydrogen bonding potential were synthesized to selectively pair with O^6-alyklG.[28] O^6-alkylG lesions have been located using an engineered DNA polymerase which can process the artificial base Benzi and specifically incorporate Benzi opposite the lesion.[29,30] Linear PCR amplification generated a pool of labeled amplicons. Unfortunately, Benzi has no specific synthetic base partner, which allows only linear amplification.[30] We are interested in using a similar approach for the alkylT lesions.

One possible source of new synthetic analogues for detecting alkylT lesions is the large library of synthetic base pairs developed by the Romesberg lab. Within their library, they identified two synthetic base pairs dNaM-d5SICS[31] and dNaM-dTPT3[32,33] (Figure 2). These artificial basepairs require a hydrophobic substituent at the primer and the template, and the subsequent extension requires a H-bond acceptor in primer nucleobase and a hydrophobic template nucleotide.[31] Efficient replication of these artificial basepair proceed in absence of hydrogen bond interaction, but with efficient packing interactions.[34] The synthetic base pairs show potential to be incorporated opposite damaged DNA through efficient packing interactions.

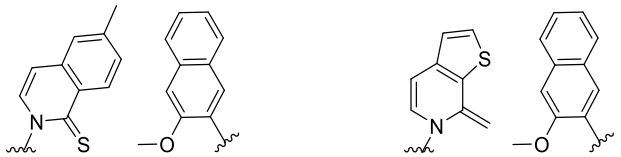


Figure 2: d5SICS-dNaM and dTPT3-dMaM synthetic basepairs.

We propose that d5SICS or dTPT3 are incorporated opposite alklT, due to the structural similarities of O^2-MeT to dNaM, which is the specific base partner of dTPT3 and d5SICS. The ultimate goal of this project is to amplify the damaged DNA and to directly detect it by MS/MS. Therefore the first step to achieve this goal is to test whether d5SICS or dTPT3 are efficient incorporated opposite O^4- and O^2-MeT with commercially available DNA polymerases.

A2 Results and discussion

We tested the incorporation of synthetic nucleosides TPT3 and 5SICS opposite O^2-MeT and O^4-MeT using two standard high fidelity polymerases (Taq and Vent). We performed single nucleotide incorporation experiments. Briefly, a 5'-end FAM labeled 23 nucleotide (nt) primer and a 45-nt template containing either O^2-MeT or O^4-MeT were incubated with DNA polymerase and natural or synthetic dNTPS. Products were resolved by denaturing polyacrylamide gel electrophoresis and visualized with a fluorescence imager (Figure 3). The synthetic nucleoside TPT3 was incorporated opposite O^2-MeT and O^4-MeT using Taq polymerase with very high efficiency (~ 100 %). Vent incorporated TPT3 only opposite O^2-MeT with high efficiency (~ 100 %), but only 50% opposite O^4-MeT. In general, 5SICS was incorporated with a lower efficiency opposite both O^2-MeT and O^4-MeT and using both DNA polymerases. The synthetic triphosphate TPT3 was incorporated with very high efficiency opposite both, O^2-MeT and O^4-MeT using Taq DNA polymerase (Figure 3B). Both polymerases were able to synthesize past O^2-MeT and O^4-MeT in the presence of only the 4 natural dNTPs.

As previously determined, Taq DNA polymerase correctly incorporate A opposite O^2-MeT and misincorporate G opposite O^4-MeT. Taq DNA polymerase correctly incorporated A opposite O^2-MeT and only minor misincorporation of the other natural bases was observed (Figure S3). On the other hand,

Taq misincorporated mostly G (>50 %) opposite O^4-MeT and correct incorporation of A was lower than 20 %. This data correlated with previously published data. Standard high fidelity polymerase Kf- incorporated the correct A opposite O^2-MeT and misincorporated G opposite O^4-MeT, the same result was observed with pol η.[35] In human cell extract G was misincorporated mainly opposite O^4-MeT.

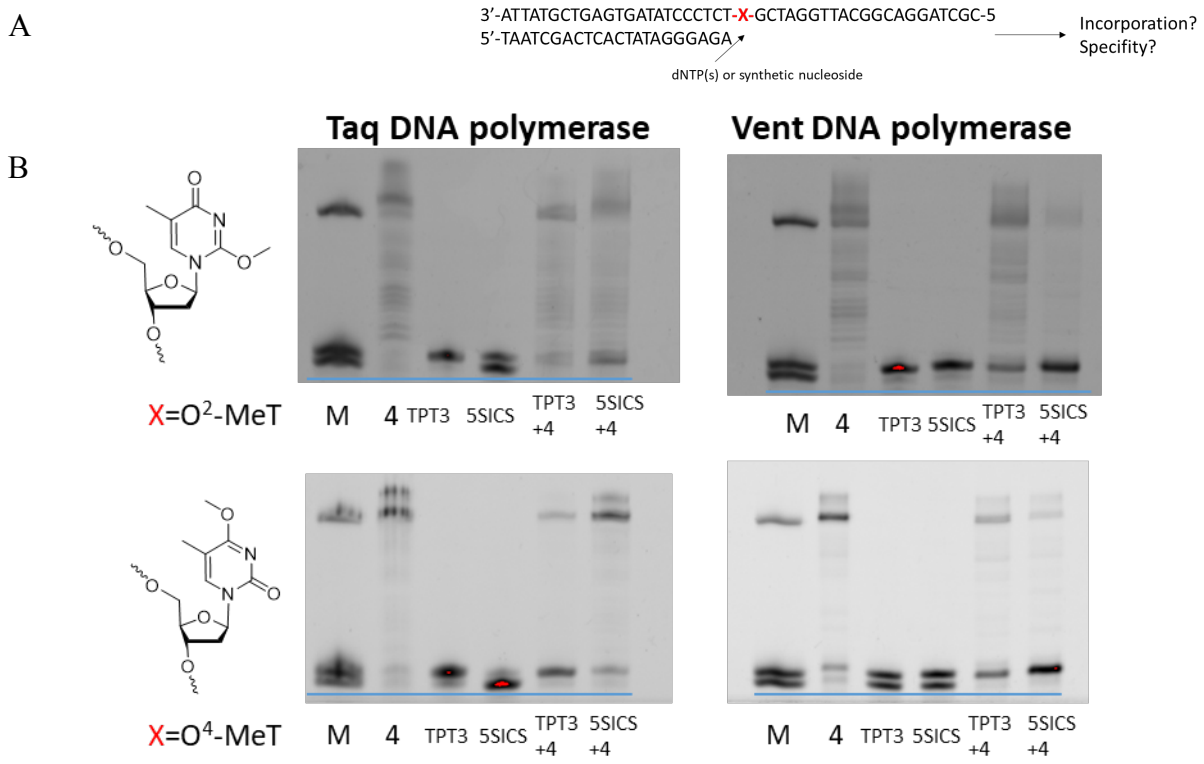


Figure 3: A) DNA polymerase-mediated primer extension experiments and sequences used in this study. B) Replication by Taq and Vent DNA polymerase past template with C = O^2-MeT and O^4-MeT. M, 23mer and 24mer primer, full length product; 4, all four dNTPs; TPT3, dTPTTP; 5SICS, d5SICSTP; TPT3+4, all four dNTPs plus dTPT3TP; 5SICS+4, all four dNTPs plus d5SICS. Final dNTP concentrations were 250 mM for each natural dNTPs, and 10mM for TPT3 and 5SICS, incubated at 55 °C for Taq and 37 °C for Vent for 15min for TPT3 or 5SICS and 30 min in presence of all 4 dNTPs.

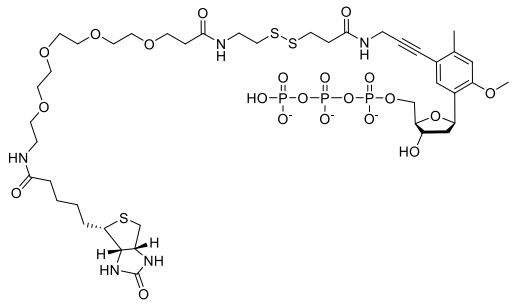
A3 Conclusion and outlook

Here we demonstrated that the synthetic nucleosides TPT3 and 5SICS could be incorporated opposite O^2- and O^4-MeT. However amplification is limited, as the Taq DNA polymerase was able to incorporate natural nucleosides opposite O^2-MeT and O^4-MeT. Specifically, Taq DNA polymerase incorporated the synthetic triphosphate TPT3 with almost 100 % efficiency opposite both O^2-MeT and O^4-MeT. The incorporation of TPT3 opposite O^2-MeT and O^4-MeT could not outcompete the natural dNTPs. The incorporation of natural bases opposite O^2-MeT and O^4-MeT gives rise to a much higher copy number of unmodified sequences. To overcome this limitation the marked oligonucleotide can be enriched and pulled out before exponential amplification in order to ensure high yield of the marked DNA by exponential amplification. The biotinylated dNaM analog, $dMMO2^{SSBIO}TP$ (Figure 4A) has been incorporated site specific to detect 8-oxoG in the cancer gene KRAS. The biotin marked DNA sequence was isolated with streptavidin beads and sequencing could locate the damage.[36]

The long term goal is to combine amplification with other detection strategies. In particular, once we achieve specific insertion of dTPT3TP opposite alkylT, it can be used for amplication with the biotinylated dNaMTP analog. Next, amplicons bearing $dMMO2^{SSBIO}$ (Figure 4A) can be isolated using a mobility gel shift assay. Furthermore, a fluorescent labelled dNaMTP analog can be synthesized, to detect the DNA damage direct on a PAGE gel using fluorescence imaging. The Biotin-PEG- moiety can

be replaced by a Cy5-PEG linker (Figure 4B). The fluorescent labelled analog has several advantages over the biotin labeled analog. In the biotin-labelled assay, the DNA is bound to streptavidin to detect. The detection includes several steps, which results in transfer or dilution of the DNA, resulting in DNA concentration loss and ultimately in DNA damage loss.[37]

A



B

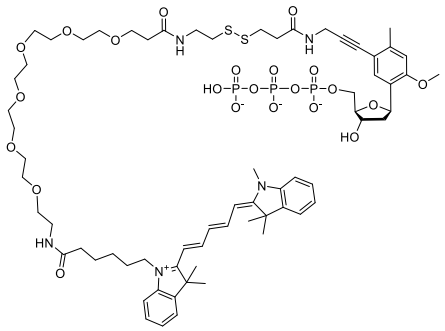


Figure 4: A) Biotin labeled dNaM analog: dMMO2^{SSBIO}TP and B) Cy3 labeled dNaM analog: dMMO2^{Cy3}TP.

A4 Methods

A4.1 Chemical reagents and materials

Reagents were purchased from Sigma-Aldrich and used without further purification. O^2-MeT was synthesized as described previously.[38,39] dNTPs, Vent and Taq polymerase were obtained by New England Biolabs. d5SICSTP was purchased from MyChem and dTPT3TP was kindly provided by Professor Floyd Romesberg. Silica gel 60 F254 plates with aluminum backing were used for thin layer chromatography. Flash column chromatography was performed on a *Biotage* system with pre-packed Flash+$KPSiO_2$ cartridges. ^1H, and ^{31}P NMR spectra were recorded on a Bruker Biospin 400 MHz NMR instrument.

A4.3 Oligonucleotides

FAM labelled oligonucleotides were purchased from Eurogentec. Modified oligonucleotides containing O^4-MeT were purchased from Gene Link^{TM} (Hawthorne, NY; USA) and containing O^2-MeT were synthesized using on a Mermade 4 DNA synthesizer (Bioautoamtion Corporation) using β-cyanophosphormidite chemistry. O^2-MeT containing 45mer oligonucleotide were purified by reverse phase HPLC on a Phenomenex Luna C-18 column (5 μm, 4.6 x 250 mm), with a linear gradient of 9 to 14 % (v/v) acetonitrile in 50 mM triethylammonium acetate over 25 min. The 45mer oligonucleotide containing O^2-MeT eluted with a retention time of 18.8 min. Corresponding oligonucleotide fractions were collected and combined, dried on a centrifugal vacuum concentrator. The purity of the 45mer oligonucleotide containing O^2-MeT was confirmed by HPLC analysis. The presence of the desired O2-MeT product was confirmed on an Agilent MSD ion trap mass spectrometer with electrospray ionization, operated in negative ion mode. The ssDNA concentration was determined on a UV Spectroscopy at 260 nm on a NanoDrop 1000 spectroscopy. Theoretical molar extinction coefficients

of the DNA sequences was calculated using Integrated DNA technologies online at http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/.

A4.2 Primer extension assays

Fam-labelled primer and templates were annealed by incubating at 95 °C for 3 min and slow colling in the block over 12 h. Final concentrations were 2.5 μM primer and 2 μM template. Full lengths primer extension reactions (10 μL) contained 1 x reactions buffer, 50 nM enzyme, 500 nM DNA (500 nM primer and 400 nM template), and 250 μM dNTPs. Single nucleotide primer extension reactions reactions (10 μL) contained 1 x reactions buffer, 25 nM enzyme, 250 nM DNA (250 nM primer and 200 nM template), and 250 μM each dNTPs or 10 μM d5SICSTP/dTPT3TP. Full lengths primer reactions were incubated for 30 min and single nucleotide reactions for 5 min. For Vent polymerase the standard ThermoPol® reaction buffer (20 mM Tris-HCl, 0.1 % Triton® X-100, 100 mM $(NH_4)_2SO_4$ 2 mM $MgSO_4$, 10 mM KCl, pH 8.8 @ 25 °C) was used and reactions were incubated at 37 °C and for Taq polymerase the standard Taq buffer (10 nM Tris-HCL, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3 @ 25 °C) was used and reactions were incubated at 55 °C. Reactions were quenched by adding 10 μL stop solution (95 % formamide and 50 mM EDTA) and the products were resolved on a 20 % polyacrylamide/7M urea denaturing gels and scanned by a BioRad Imager. Quantification and analysis was carried out with Image J software.

A5 References

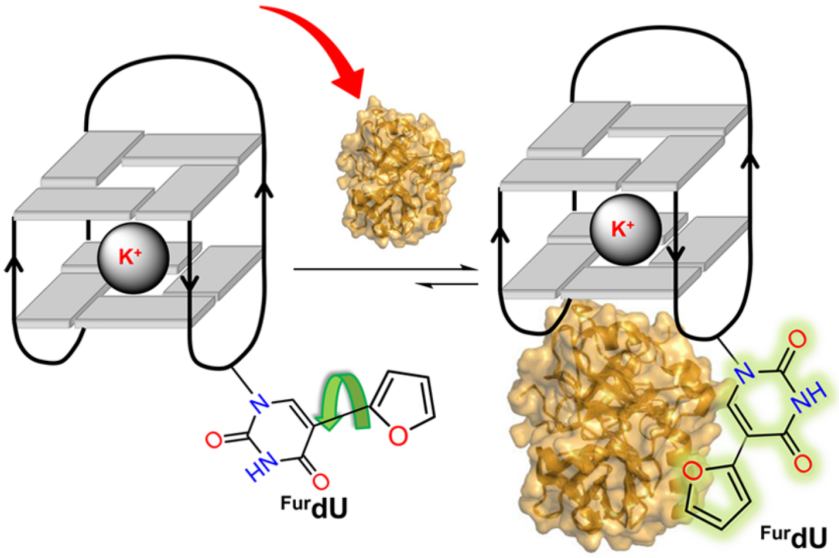
1. Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **424,** 127–142 (1999).
2. Hecht, S. S. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* **3,** 733–744 (2003).
3. Hecht, S. S. Biochemistry, biology, and carcinogenicity of tobacco-specific N- nitrosamines. *Chem. Res. Toxicol.* **11,** 559–603 (1998).
4. Joosen, A. M. C. P. *et al.* Effect of processed and red meat on endogenous nitrosation and DNA damage. *Carcinogenesis* **30,** 1402–1407 (2009).
5. Lewin, M. H. *et al.* Red Meat Enhances the Colonic Formation of the DNA Adduct O^6-Carboxymethyl Guanine: Implications for Colorectal Cancer Risk. *Cancer Res.* **66,** 1859–1865 (2006).
6. Bouvard, V. *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol.* **16,** 1599–1600 (2015).
7. Cross, A. J. & Sinha, R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ. Mol. Mutagen.* **44,** 44–55 (2004).
8. Peterson, L. A. Formation, repair, and genotoxic properties of bulky DNA adducts formed from tobacco-specific nitrosamines. *J. Nucleic Acids* **2010,** 1–11 (2010).
9. Hecht, S. S. Human urinary carcinogen metabolites: biomarkers for investigating tobacco and cancer. *Carcinogenesis* **23,** 907–922 (2002).
10. Godschalk, R. *et al.* Comparison of multiple DNA adduct types in tumor adjacent human lung tissue : effect of cigarette smoking * derived from tobacco and its pyrolysis products , including are thought to initiate lung carcinogenesis . The presence in tumor adjacent uninvolved. *Carcinogenesis* **23,** 2081–6 (2002).
11. Anna, L., Kovács, K., Gyorffy, E., Schoket, B. & Nair, J. Smoking-related O4-ethylthymidine formation in human lung tissue and comparisons with bulky DNA adducts. *Mutagenesis* **26,** 523–527 (2011).
12. Upadhyaya, P., Kalscheuer, S., Hochalter, J. B., Villalta, P. W. & Hecht, S. S. Quantitation of pyridylhydroxybutyl-DNA adducts in liver and lung of F-344 rats treated with 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.* **21,** 1468–1476 (2008).
13. Lao, Y., Yu, N., Kassie, F., Villalta, P. W. & Hecht, S. S. Formation and accumulation of pyridyloxobutyl DNA adducts in F344 rats chronically treated with 4-(methylnitrosamino)-1-(3-

pyridyl)-1-butanone and enantiomers of its metabolite, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.* **20,** 235–245 (2007).
14. Fu, D., Calvo, J. A. & Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **12,** 104–120 (2012).
15. Klein, J. C. *et al.* Use of shuttle vectors to study the molecular processing of defined carcinogen-induced DNA damage: mutagenicity of single O4-ethylthymidine adducts in HeLa cells. *Nucleic Acids Res.* **18,** 4131–4137 (1990).
16. Klein, J. C. *et al.* Role of nucleotide excision repair in processing of O4-alkylthymines in human cells. *J. Biol. Chem.* **269,** 25521–25528 (1994).
17. Altshuler, K. B., Hodes, C. S. & Essigmann, J. M. Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells: A comparison of the mutagenicities of O4-methylthymine and O6-methylguanine in cells with different DNA repair backgrounds. *Chem. Res. Toxicol.* **9,** 980–987 (1996).
18. Brent, T. P. *et al.* Repair of O-alkylpyrimidines in mammalian cells : A present consensus. *Biochem. J* **85,** 1759–1762 (1988).
19. Fang, Q., Kanugula, S., Tubbs, J. L., Tainer, J. A. & Pegg, A. E. Repair of O4-alkylthymine by O6-alkylguanine-DNA alkyltransferases. *J. Biol. Chem.* **285,** 8185–8195 (2010).
20. Wang, P., Amato, N. J., Zhai, Q. & Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in Escherichia colicells. *Nucleic Acids Res.* **43,** 10795–10803 (2015).
21. Wilkinson, M. C. *et al.* Purification of the E.coli ogt gene product to homogeneity and its rate of action on O6-methylguanine,, O6-ethylguanine and O4-methylguanine in dodecadeoxyribonucleotides. *Nucleic Acids Res.* **17,** 8475–8484 (1989).
22. Graves, R. J., Li, B. F. L. & Swann, P. F. Repair of C^-methylguanine, C^-ethylguanine, OMsopropylguanine and 0*-methylthymine in synthetic oligodeoxynudeotides by Escherichia coli ada gene 6^-alkylguanine-DNA-alkyltransferase. *Carcinogenesis* **10,** 661–666 (1989).
23. Bronstein, S. M., Skopek, T. R. & Swenberg, J. A. Efficient repair of O6-ethylguanine, but not O4 -ethylthymine or O2-ethylthymine, is dependent upon O6-alkylguanine-DNA alkyltransferase and nucleotide excision repair Activities in Human Cells. *Cancer Res.* **52,** 2008–2011 (2011).
24. Crain, P. F. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.* **193,** 782–790 (1990).
25. Chowdhury, G. & Guengerich, F. P. Liquid Chromatography-Mass Spectrometry Analysis of DNA Polymerase Reaction Products. in *Current Protocols in Nucleic Acid Chemistry* 1–11 (John Wiley & Sons, Inc., 2001).
26. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7,** 461–465 (2010).
27. Clark, T. A., Spittle, K. E., Turner, S. W. & Korlach, J. Direct Detection and Sequencing of Damaged DNA Bases. *Genome Integr.* **2,** 10–19 (2011).
28. Gahlon, H. L. & Sturla, S. J. Hydrogen bonding or stacking interactions in differentiating duplex stability in oligonucleotides containing synthetic nucleoside probes for alkylated DNA. *Chem. - A Eur. J.* **19,** 11062–11067 (2013).
29. 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,** 30–33 (2015).
30. 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 O 6 -alkylguanine DNA adducts. *Nucleic Acids Res.* **44,** 6564–6573 (2016).
31. Leconte, A. M. *et al.* Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* **130,** 2336–2343 (2008).
32. Dhami, K. *et al.* Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. *Nucleic Acids Res.* **42,** 10235–10244 (2014).
33. Lavergne, T., Malyshev, D. A. & Romesberg, F. E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chem. - A Eur. J.* **18,** 1231–1239 (2012).
34. Malyshev, D. a. *et al.* Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci.* **109,** 12005–

12010 (2012).
35. Andersen, N., Wang, J., Wang, P., Jiang, Y. & Wang, Y. In-Vitro Replication Studies on O2-Methylthymidine and O4-Methylthymidine. *Chem. Res. Toxicol.* **25,** 2523–2531 (2012).
36. Riedl, J., Ding, Y., Fleming, A. M. & Burrows, C. J. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* **6,** 1–11 (2015).
37. Condie, A. G., Yan, Y., Gerson, S. L. & Wang, Y. A fluorescent probe to measure DNA damage and repair. *PLoS One* **10,** 1–21 (2015).
38. Sturla, S. J., Scott, J., Lao, Y., Hecht, S. S. & Villalta, P. W. Mass spectrometric analysis of relative levels of pyridyloxobutylation adducts formed in the reaction of DNA with a chemically activated form of the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.* **18,** 1048–1055 (2005).
39. Xu, Y. Z. & Swann, P. F. Oligodeoxynucleotides containing O2-alkylthymine: Synthesis and characterization. *Tetrahedron Lett.* **35,** 303–306 (1994).

Appendix B

A simple molecular rotor for defining nucleoside environment within a DNA aptamer-protein complex



Thomas Z. Cservenyi, Abigail J. Van Riesen, Florence D. Berger, Ahmed Desoky, and Richard A. Manderville, A simple Molecular Rotor for Defining Nucleoside Environment within a DNA Aptamer-Protein Complex. ACS Chem. Biol., 2016, 11 (9), pp 2576-2582

T.Z.C. designed the study, performed the thermal melting, fluorescence and CD studies. A.J.vR. performed fluorescence studies, F.D.B. performed fluorescence studies, A.D. synthesized and purified oligonucleotides. R.A.M. conceived the research. R.A.M. and T.Z.C. interpreted data and wrote the manuscript.

ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

A Simple Molecular Rotor for Defining Nucleoside Environment within a DNA Aptamer–Protein Complex

Thomas Z. Cservenyi,[†] Abigail J. Van Riesen,[†] Florence D. Berger,[‡] Ahmed Desoky,[†,∥] and Richard A. Manderville*[,†]

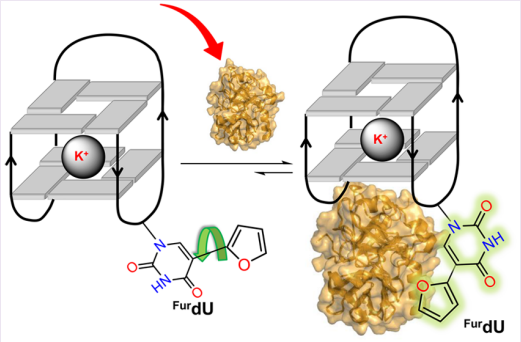
[†]Departments of Chemistry and Toxicology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

[‡]Department of Health Science and Technology, ETH Zurich, Schmelzbergstrasse 9, 8092 Zurich, Switzerland

[∥]Department of Chemistry, Faculty of Science, Suez Canal University, Ismailia, Egypt

Supporting Information

ABSTRACT: The simple 5-furyl-2′-deoxyuridine (^{Fur}dU) nucleobase exhibits dual probing characteristics displaying emissive sensitivity to changes in microenvironment polarity and to changes in solvent rigidity due to its molecular rotor character. Here, we demonstrate its ability to define the microenvironment of the various thymidine (T) loop residues within the thrombin binding aptamer (TBA) upon antiparallel G-quadruplex (GQ) folding and thrombin binding. The emissive sensitivity of the ^{Fur}dU probe to microenvironment polarity provides a diagnostic handle to distinguish T bases that are solvent-exposed within the GQ structure compared with probe location in the apolar duplex. Its molecular rotor properties then provide a turn-on fluorescent switch to identify which T residues within the GQ bind specifically to the protein target (thrombin). The fluorescence sensing characteristics of ^{Fur}dU make it an attractive tool for mapping aptamer–protein interactions at the nucleoside level for further development of modified aptamers for a wide range of diagnostic and therapeutic applications.



Nucleic acid research has expanded in recent years beyond investigation of their abilities related to storage of genetic information. This has led to the development of an *in vitro* methodology for the derivation of functional nucleic acids, capable of performing binding with high specificity and sensitivity. These novel classes of biomolecules, termed aptamers, have been employed as versatile tools for the detection of important biological targets for applications in imaging, diagnostics, and therapeutics.[1−3] Chemical modifications of aptamers increase chemical diversity for improved therapeutic use[4,5] and molecular target binding affinity.[6−10] Aptamer base modifications can also create a fluorophore with emission that is sensitive to the microenvironment, resulting in a versatile tool for target detection.[11] Knowledge of how aptamers bind to their targets can be useful for optimizing aptamer length and establishing preferred sites of aptamer modification for a wide range of bioanalytical and therapeutic applications.[12]

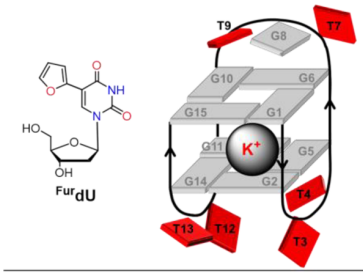
Our interest in aptamer-target interactions has focused on the utility of internal fluorescent nucleobase mimics for monitoring target binding on the basis of probe emission intensity and wavelength changes.[13] We have demonstrated that 8-heteroaryl-2′-deoxyguanosine residues, such as 8-furyl-dG (^{Fur}dG)[14−17] and 8-thienyl-dG (ThdG),[18] can be placed within G-tetrads of G-quadruplex (GQ) structures without perturbing GQ folding. The 8-heteroaryl-dG bases are fluorescent and exhibit quenched emission in the duplex that exhibits increased intensity in the GQ structure, possibly due to effective energy-transfer from the unmodified G's in the G-tetrad.[19,20] Using the 15-mer thrombin binding aptamer (TBA)[21] to establish proof-of-concept, the 8-heteroaryl-dG probes were employed in duplex → GQ exchange[22] to monitor thrombin binding through the increased emission intensity of the modified dG base triggered by thrombin-induced GQ formation.[15,18]

While the 8-heteroaryl-dG probes minimally perturb thrombin-binding affinity and exhibit useful fluorescence switching properties,[15,18] it was desirable to simplify the detection platform, as duplex → GQ exchange could provide false positives in complex matrices where certain metal cations and ligands that promote GQ formation would also induce a positive probe response. Thus, we sought an alternative nucleobase probe that could exhibit a change in emission response solely due to target (thrombin) binding. Given that thrombin binds to the loop regions of TBA[23−25] with preferential interaction with the two TT loop residues,[25] we reasoned that the emissive probe should rather be a T mimic than a G replacement. Furthermore, the fluorophore should possess molecular rotor characteristics[26] in which structural

Received: May 19, 2016
Accepted: July 22, 2016
Published: July 22, 2016

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

rigidification enhances emission intensity, making these viscosity-sensitive probes useful for monitoring biomolecular interactions.[27,28] The prototypical pyrimidine molecular rotor nucleobase probe is 5-furyl-2′-deoxyuridine (^{Fur}dU, Figure 1)



mTBA	Sequence
^{Fur}dU-3	5′-GG^{Fur}dUTGGTGTGGTTGG-3′
^{Fur}dU-4	5′-GGT^{Fur}dUGGTGTGGTTGG-3′
^{Fur}dU-7	5′-GGTTGG^{Fur}dUGTGGTTGG-3′
^{Fur}dU-9	5′-GGTTGGTG^{Fur}dUGGTTGG-3′
^{Fur}dU-12	5′-GGTTGGTGTGG^{Fur}dUTGG-3′
^{Fur}dU-13	5′-GGTTGGTGTGGT^{Fur}dUGG-3′

Figure 1. Structure of ^{Fur}dU, the antiparallel GQ produced by TBA in the presence of K^+ with T residues highlighted in red and the mTBA sequences containing the ^{Fur}dU modification.

developed by the Tor laboratory.[29−32] This emissive pyrimidine exhibits dual probing capabilities where environmental polarity impacts its Stokes shift ($\Delta\nu$), while structural rigidity impacts its brightness.[31] The sensitivity of ^{Fur}dU to polarity has been employed to measure duplex major groove polarity,[30] while its sensitivity to rigidity has been demonstrated in the duplex with the probe placed opposite an abasic site.[29,31] However, the probe has not yet been employed for monitoring protein−DNA interactions where its molecular rotor character could serve as a useful tool for determining DNA aptamer residues that bind to the protein target.

Here, we report the utility of ^{Fur}dU for mapping the site of thrombin binding to the antiparallel GQ produced by TBA. TBA was chosen for study in order to compare the emissive responses of ^{Fur}dU and ^{Fur}dG,[14−17] and because the TBA−thrombin complex has been characterized by high-resolution crystallographic studies,[24] permitting a direct comparison between probe response and structure. The ^{Fur}dU nucleobase was incorporated into the six different T sites of TBA (Figure 1), and its structural impact and emissive response within the duplex and GQ were determined for the six unique modified TBA (mTBA) strands. The emissive response of the ^{Fur}dU probe at the various T positions upon thrombin binding was in agreement with expectations derived from the X-ray structure of the TBA−thrombin complex.[24] Our studies establish the utility of emissive molecular rotor nucleobase probes for mapping protein-aptamer binding sites and provide a TBA-thrombin platform for optimizing molecular rotor probe performance.

RESULTS AND DISCUSSION

UV Thermal Melting Experiments and CD Measurements. The ^{Fur}dU nucleoside and its phosphoramidite, for site-specific incorporation into the TBA oligonucleotide using solid-phase DNA synthesis, were synthesized as previously described by Greco and Tor (see Supporting Information (SI) for ^{1}H NMR spectra of synthetic samples and ESI-MS data (Table S1) and spectra of mTBA oligonucleotides).[32] Thermal melting studies (Table 1) were initially carried out to determine the

Table 1. UV-Thermal Melting Parameters for Duplex (D) and GQ Formation by Native and ^{Fur}dU-Modified TBA[a]

mTBA	T_m D, Na^+	ΔT_m D, Na^+	T_m GQ, K^+	ΔT_m GQ, K^+
native	64.5		53.5	
^{Fur}dU-3	64.0	−0.5	51.5	−2.0
^{Fur}dU-4	63.5	−1.0	60.0	+6.5
^{Fur}dG-7	65.0	+0.5	52.0	−1.5
^{Fur}dG-9	64.5	0.0	56.8	+3.0
^{Fur}dU-12	64.5	0.0	52.0	−1.5
^{Fur}dU-13	64.0	−0.5	56.5	+3.0

[a] T_m values are in °C and reproducible within 3%.

positional impact of ^{Fur}dU on duplex (6 μM, T_m's in Na^+ solution) and GQ (T_m's in K^+ solution) stability. For TBA, we routinely carry out duplex studies in Na^+ solution,[14,15,17,18] as this eliminates the opportunity for competitive GQ formation (GQ $T_m \sim 24$ °C in Na^+ solution[24]). The ^{Fur}dU probe is an excellent T mimic and exhibited minimal impact on duplex stability regardless of probe location, as the probe can participate in WC base pairing with A to form stable duplexes.[29−31] However, in the GQ produced in K^+ solution, the ^{Fur}dU probe caused a stabilizing influence on GQ stability at positions 4 ($\Delta T_m = 6.5$ °C), 9, and 13 (ΔT_m's = 3.0 °C), and a slight destabilizing influence at positions 3, 7, and 12 compared to native TBA (T_m = 53.5 °C). The observed changes in thermal stability induced by ^{Fur}dU are in agreement with structural studies.[24,33−35] NMR suggests that T-4 and T-13 produce a base pair with strong stacking interactions with the G-tetrad, while T-9 interacts with G-8 of the TGT loop and the adjacent G-tetrad.[35] In contrast, T-3, T-7, and T-12 are not predicted to interact with adjacent nucleotides and are freely accessible to solvent.[24,33,34] Replacement of T with ^{Fur}dU at G-tetrad-stacking sites (i.e., 4, 9, and 13) would be expected to enhance stacking interactions and hence stabilize the GQ structure. However, the increased lipophilicity of ^{Fur}dU relative to T would slightly decrease GQ stability at solvent-exposed sites (i.e., 3, 7, and 12). Thus, the ^{Fur}dU probe proved to be an excellent indicator of T loop residue interactions or lack thereof within the GQ structure on the basis of thermal melting parameters. Other T residue replacements within TBA have lacked this ability. For example, replacement of T-4 or T-13 with C destabilizes the GQ structure due to diminished G-tetrad stacking interactions.[36] Replacement of T-4 or T-13 with A[23] or 2-aminopurine (2AP)[37] can slightly stabilize the GQ structure ($\Delta T_m = 2.0$ °C[37]). However, both A and 2AP exhibit a strong destabilizing influence at T-9.[23,37]

Circular dichroism (CD) was utilized to confirm the duplex and antiparallel GQ topology of the ^{Fur}dU mTBA samples. Typical duplex CD spectra were obtained with negative peaks at 240 nm and positive peaks at 260 nm, while GQ CD spectra confirmed an antiparallel structure with positive peaks at 245 and 290 nm and negative peaks at 260 nm (Figure S1, SI).[18,24]

Fluorescence Response: Duplex vs GQ. The emission and excitation spectra of the mTBA GQ (dashed traces) and duplex (solid traces) samples highlight the ability of ^{Fur}dU to distinguish the two topologies at the various T positions (Figure 2). The free ^{Fur}dU nucleoside displays two absorbances in aqueous buffer at 252 nm (ε = 13 800 cm^{-1} M^{-1}) and 316

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

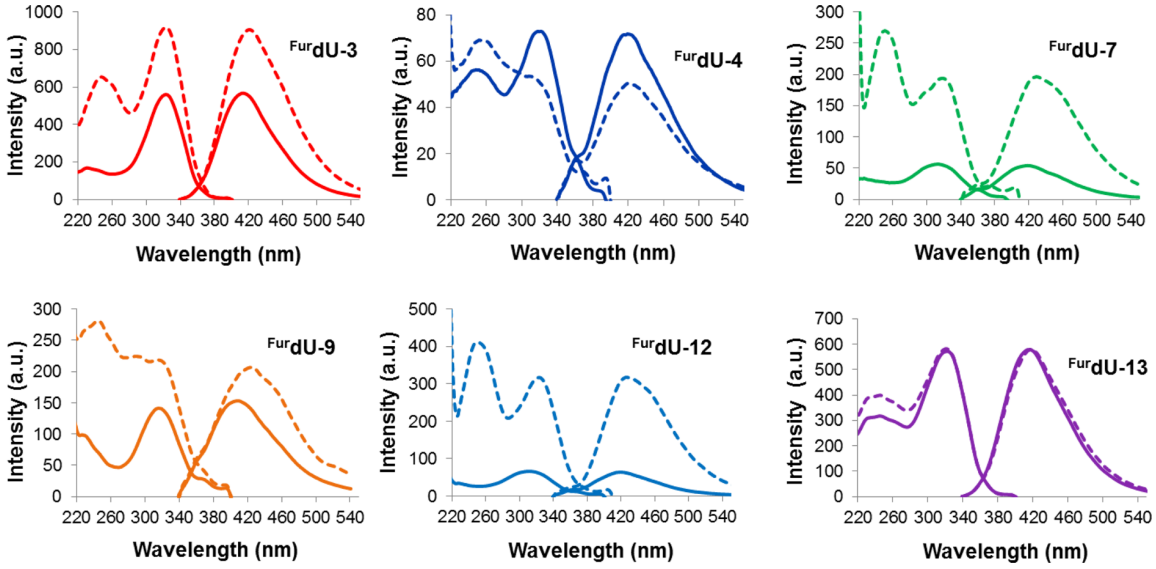


Figure 2. Fluorescence excitation and emission spectral overlays of mTBA oligonucleotides. Solid lines represent duplexes in Na^+ buffer; dashed lines represent GQs in K^+. Emission spectra were recorded with excitation at 316 nm.

nm (ε = 11 000 cm^{-1} M^{-1})[32] with λ_{em} at 431 nm (Φ_{fl} = 0.03).[29] The absorption maximum at 316 nm is insensitive to polarity changes, while the emission spectra are significantly impacted and display increased bathochromic (red-shift) and hyperchromic (increased intensity) effects with increasing solvent polarity.[29] Thus, emission spectra for the ^{Fur}dU-mTBA samples were recorded with λ_{ex} = 316 nm, and it was anticipated that probe emission would increase in the GQ compared with its emission in the apolar duplex environment[30] if the probe experienced greater solvent exposure.

The anticipated ^{Fur}dU probe light-up emission response was observed at ^{Fur}dU-7 and ^{Fur}dU-12 (Figure 2) where the probe exhibited ~4- and 5-fold higher fluorescence intensity compared with the corresponding probe emission in the duplex. In the GQ positions, 7 and 12 are fully solvent-exposed and represent sites where the ^{Fur}dU probe slightly decreased GQ stability (Table 1). The other probe site that is solvent-exposed is ^{Fur}dU-3. At this position, an intensity increase of only 1.6-fold was observed, partly due to the highly emissive nature of the probe within the duplex. In contrast, the ^{Fur}dU probe exhibited quenched GQ emission intensity (^{Fur}dU-4), no changes in emission intensity (^{Fur}dU-13) or only a minor enhancement (^{Fur}dU-9) compared with duplex emission at positions where the probe stabilized the GQ (Table 1) and is engaged in stacking interactions. A notable feature of the excitation spectra for the GQ samples (dashed traces) with the ^{Fur}dU probe at positions 3, 7, and 12 was the obvious presence of the excitation maxima at ~250 nm that were absent in the duplex excitation spectra (solid traces) but present in the absorbance spectrum of the free nucleoside ^{Fur}dU.[32] We have observed similar dual absorption maxima for push−pull 8-aryl-dG probes and have attributed the blue-shifted peak to a twisted nonplanar biaryl structure, while the red-shifted maxima would represent the fully conjugated planar structure.[38] The absence of the blue-shifted 250 nm maxima in the excitation spectra of the duplex samples is consistent with preferential formation of the planar probe structure due to π-stacking interactions. In the antiparallel GQ structures with the ^{Fur}dU probe exposed to solvent (sites 3, 7, and 12), the biaryl chromophore would be free to rotate between twisted and planar states.

The photophysical parameters of the various ^{Fur}dU-mTBA samples are summarized in Table 2. At positions ^{Fur}dU-3, -7, and -12 that provided the greatest increase in emission intensity (I_{rel} values), significant increases in Stokes shift ($\Delta(\Delta\nu)$ values (9−13 nm) were also observed. In contrast, positions ^{Fur}dU-4 and ^{Fur}dU-13 that failed to exhibit increases in emission intensity displayed small Stokes shift changes (1 and 3 nm). The only exception to the observed trend occurred at ^{Fur}dU-9, which produced a large Stokes shift change (13 nm) and a relatively small I_{rel} value (1.3). Overall, the observed emission changes of the ^{Fur}dU probe in TBA roughly correlated with emission changes noted for 2AP,[37] which also displayed emission intensity increases in the GQ relative to the duplex when placed at loop sites not involved in π-stacking interactions with the G-tetrads, i.e., positions 3, 7, and 12. For 2AP, the emission increase at specific sites within the TBA GQ relative to the duplex can be ascribed to diminished stacking interactions.[37]

Table 2. Photophysical Parameters for ^{Fur}dU-mTBA Duplex (D) and GQ Structures

^{Fur}dU	λ_{em}D (nm)[a]	$\Delta\nu$ (nm)[b]	λ_{em}GQ (nm)	$\Delta\nu$ (nm)	$\Delta(\Delta\nu)$ (nm)[c]	I_{rel}[d]
3	412	96	421	105	9	1.6
4	418	102	419	103	1	0.7
7	417	101	428	112	11	3.6
9	408	92	421	105	13	1.3
12	415	99	428	112	13	5.0
13	417	101	420	104	3	1.0

[a]Wavelengths of emission (λ_{em}) for D and GQ were recorded with excitation at 316 nm at 10 °C. [b]Stokes shift ($\Delta\nu$) (λ_{em} − 316 nm) for D and GQ. [c]Relative Stokes shift for GQ ($\Delta\nu$) − D ($\Delta\nu$). [d]Relative emission intensity for GQ/D.

During the course of our studies, Tanpure and Srivatsan reported that the 5-(benzofuran-2-yl)uracil (^{BFur}dU) nucleobase can be used to monitor GQ formation within human telomeric

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

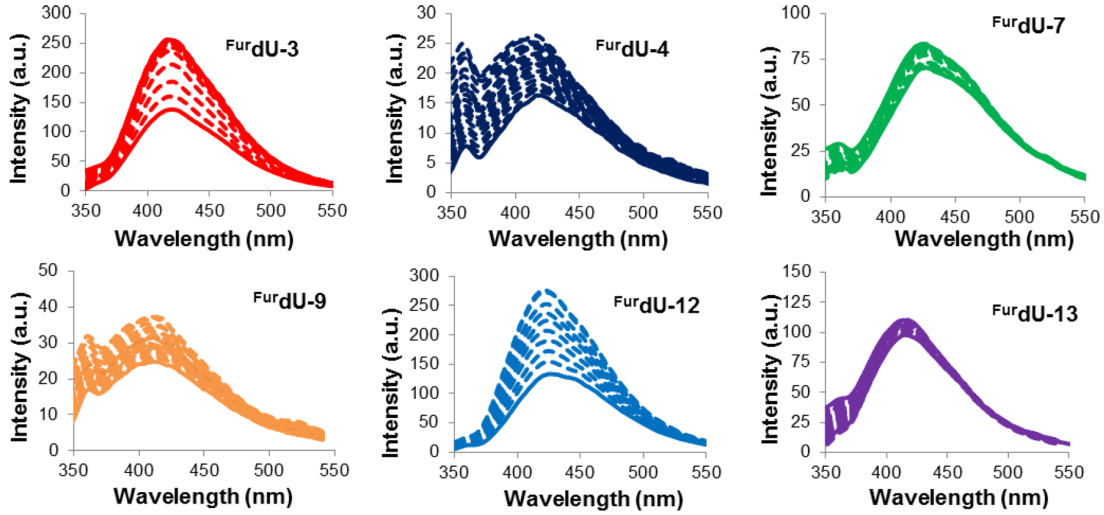


Figure 3. Fluorescence titrations (3 μM mTBA) carried out with thrombin at 25 °C. The initial trace of GQ is depicted by solid line, while dashed traces depict emission upon successive addition of thrombin. Emission spectra were recorded with excitation at 316 nm.

DNA and RNA GQs.[39] They noted that the probe exhibited ~4- to 9-fold fluorescence intensity enhancements in the GQ compared with the duplex depending on the specific GQ topology produced and position of the modification. At certain sites, the ^{BFur}dU probe exhibited quenched emission in the GQ relative to the duplex, as noted for ^{Fur}dU-4-mTBA. While they attributed the changes in probe emission response to a variety of factors—including desolvation−solvation effects, rigidification of the fluorophore, reduced electron transfer processes, and stacking interactions[39]—we propose that the main factors for the light-up ^{Fur}dU emissive response in the GQ compared with the duplex is the change in microenvironment polarity[40] coupled with the loss of π-stacking interactions. The molecular rotor properties of the ^{Fur}dU probe[31] cannot play a role in the enhanced emission in the GQ relative to the duplex because the greatest I_{rel} values (up to 5-fold) occurred at loop sites where the probe is fully solvent-exposed with decreased rigidification of the biaryl fluorophore.

Thrombin Binding Studies. Fluorescence thrombin-aptamer titrations were carried out in potassium phosphate buffer (0.1 M KCl, pH 7) at 25 °C in order to determine the thrombin binding affinity of the various mTBA oligonucleotides and to monitor the site-specific emissive response of the ^{Fur}dU probe to protein binding. Control titrations with bovine serum albumin (BSA) were also carried out. For these experiments, the mTBA samples (3 μM) were prefolded into the functional GQ structure required for thrombin recognition prior to protein addition. In this way, changes in probe emission would be in direct response to protein binding. The change in emission of the ^{Fur}dU probe at the various T sites within the mTBA GQ (solid trace) upon thrombin addition (dashed traces, Figure 3) demonstrated a 2-fold increase in emission intensity at ^{Fur}dU-3 and ^{Fur}dU-12. In contrast, only weak emissive responses were observed at the other locations, and less than 1% change in emission intensity was noted in titrations with BSA (Figure S2, SI). The solid-state structure of the unmodified thrombin−TBA complex reveals that T-3 and T-12 form hydrophobic contacts with exosite I of the thrombin protein.[23,24] In contrast, the TGT loop that includes T-7 and T-9 is far away from the protein binding site, while T-4 and T-13 contribute to GQ stability through G-tetrad stacking interactions and do not protrude into the hydrophobic protein binding pocket.[23,24]

Plots of the normalized fluorescence intensity versus [protein] (Figure S2, SI, includes plots for thrombin and BSA addition) indicated a 1:1 thrombin−mTBA interaction and provided the dissociation constants (K_d in μM) given in Table 3. Also provided in Table 3 are changes in emission wavelength ($\Delta(\lambda_{em})$) and emission intensity (I_{rel})) in the thrombin−mTBA complex compared to the corresponding photophysical parameters for the free mTBA GQ structure (Table 2). Sites that provided the greatest increase in emission intensity (^{Fur}dU-3 and ^{Fur}dU-12) also provided the largest wavelength changes, showing a blue-shift upon thrombin binding. This observation is consistent with their interaction with hydrophobic residues within the thrombin binding site and would be expected to decrease probe emission due to its sensitivity to polarity. However, the interaction of the probe with the protein binding pocket would also be expected to decrease free rotation of the biaryl fluorophore for increased brightness due to the molecular rotor properties of the nucleobase probe. Thus, at positions ^{Fur}dU-3 and ^{Fur}dU-12, it appears that the loss of probe rotation has a greater impact than the decrease in probe polarity to cause the overall 2-fold increase in probe emission.

Table 3. Photophysical Parameters and Dissociation Constants for Thrombin Binding by ^{Fur}dU-mTBA

^{Fur}dU	D(l_{em}) (nm)[a]	I_{rel}[b]	K_d (uM)
3	−3.0	2.0	0.8 ± 0.1
4	0.0	1.4	6.1 ± 0.4
7	−1.0	1.2	3.5 ± 0.4
9	0.0	1.4	5.9 ± 0.8
12	−7.0	2.2	1.5 ± 0.2
13	0.0	1.1	4.9 ± 0.5

[a]Change in emission wavelength of mTBA GQ + thrombin versus free GQ. [b]Relative emission intensity for GQ + thrombin/GQ.

Binding of native TBA to thrombin was determined using a 5′-fluorescein (FAM)-labeled TBA sample and fluorescence polarization (FP), which provided a K_d value of 4.9 μM.

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

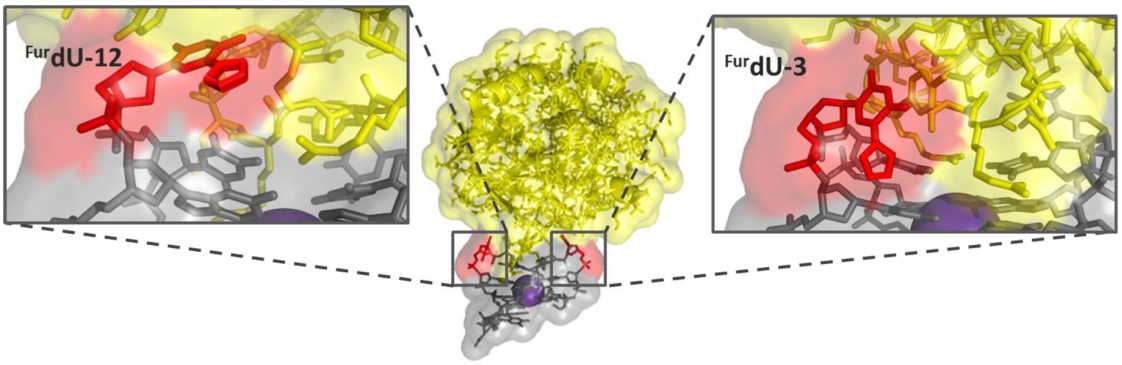


Figure 4. Structure of TBA in gray, highlighting positions ^{Fur}dU-3 and ^{Fur}dU-12 in red, bound to K^+ (purple ball) and thrombin protein in yellow (Protein Data Bank, 4DII[24]). Enlarged with the PyMOL molecular graphics system showing positions ^{Fur}dU-12 on the left and ^{Fur}dU-3 T3 on the right.

Previously, we demonstrated that placement of 8-aryl-dG probes within the *syn*-G-tetrad positions of TBA does not perturb thrombin binding compared to the 5′-FAM-labeled TBA sample.[18] Similar thrombin binding constants (4.4−6.8 μM) were obtained regardless of probe location within the G-tetrads. Inspection of the K_d values (Table 3) for the ^{Fur}dU-mTBA samples indicates that binding affinity is strongly location-dependent. With the ^{Fur}dU probe at positions 4, 7, 9, and 13, the K_d values ranged from 3.5 to 6.1 μM, which is similar to the affinity for thrombin displayed by the 5′-FAM-labeled TBA sample and the mTBA oligonucleotides containing 8-aryl-dG probes in the G-tetrad positions.[18] However, ^{Fur}dU-3- and ^{Fur}dU-12-mTBA exhibited enhanced binding affinity with K_d values of 0.8 and 1.5 μM, respectively. A model of the mTBA−thrombin structure utilizing the X-ray structure of the native TBA−thrombin complex[24] is presented in Figure 4. The model suggests that insertion of the ^{Fur}dU probe into the protein hydrophobic binding pocket would be expected to enhance stability compared with insertion of T due to the increased lipophilicity and enhanced stacking interactions of ^{Fur}dU with protein residues.

Utility of ^{Fur}dU for Aptasensor Development. We have demonstrated that the isosteric emissive nucleoside ^{Fur}dU[29−32] can be used to define the microenvironment of the loop T residues within the thrombin−TBA complex. The loop T residues of TBA are known to impact GQ stability and thrombin binding.[23−25] By comparing the emission of the ^{Fur}dU probe in the duplex to the GQ structure required for thrombin binding, it is possible to predict which T's are are solvent-exposed within the GQ, as this causes preferential light-up emission of the ^{Fur}dU nucleobase. Due to the molecular rotor properties of ^{Fur}dU, it also predicts which T's interact strongly with the protein thrombin (T-3 and T-12) by displaying enhanced fluorescence emission intensity and enhanced thrombin binding affinity (Table 3 and Figure 3). Compared with ^{Fur}dG for monitoring thrombin binding,[15,18] the ^{Fur}dU probe can avoid the requirement for duplex → GQ exchange to provide an emissive turn-on signal by responding directly to thrombin addition depending on its site of incorporation, which simplifies the detection platform and provides information on the site(s) of protein binding. This ability makes it a potentially useful tool for determining DNA aptamer residues that interact with protein targets in which structural information is lacking. Other approaches such as NMR and X-ray crystallography are critical for providing high-resolution structures with atomic detail, yet they are time-consuming and require significant quantities of the aptamer-protein complex. Newer developed methods such as FP for mapping aptamer-protein interactions[12] utilize complicated instrumentation (capillary electrophoresis (CE) with laser-induced FP detection) and produce detailed information that is not as straightforward to interpret as turn-on emission intensity of the ^{Fur}dU probe upon protein binding. Our work clearly establishes T-3 and T-12 residues of TBA as modification sites for molecular rotor probe development. The ^{Fur}dU probe is not optimized for detecting hydrophobic environments in proteins because it exhibits quenched emission in apolar environments, but enhanced emission with increased rigidity. Ideally, the molecular rotor probe should display enhanced emission with significant changes in emission wavelength with decreased solvent polarity for a dramatic increase in fluorescence signal upon insertion into the hydrophobic protein binding pocket. Thus, our current efforts are directed toward the use of the TBA−thrombin platform to optimize artificial molecular rotor DNA bases that possess turn-on visible excitation and emission wavelengths only upon thrombin binding. Such probes would then be employed in a wide range of aptamers for development of fluorescent aptasensors.

Conclusion. These results show the impact of the fluorescent ^{Fur}dU probe within the three unique loops of the 15-mer TBA, which folds into an antiparallel GQ upon target (thrombin) binding. The ^{Fur}dU probe exhibits dual probing characteristics, providing changes in emission wavelength and intensity upon changes in microenvironment polarity and enhanced fluorescence intensity with increased solvent rigidity due to its molecular rotor properties. Within the TBA GQ, its sensitivity to microenvironment polarity provides a diagnostic emissive handle to determine which T bases are solvent-exposed. Its molecular rotor properties then provide a fluorescent diagnostic handle to determine which T residues interact strongly with the molecular target, which restricts free rotation of the ^{Fur}dU probe for enhanced emission. For the TBA−thrombin interaction, replacement of T-3 and T-12 with ^{Fur}dU increases thrombin binding affinity and provides a 2-fold increase in emission intensity. This observation correlated with the solid-state structure of the TBA−thrombin complex, which demonstrates that T-3 and T-12 are inserted into the hydrophobic binding site of the protein (thrombin) target. These fluorescence-sensing characteristics make ^{Fur}dU a potentially useful tool for mapping aptamer-protein interactions

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

at the nucleoside level for further development of modified aptamers for a wide range of diagnostic and therapeutic applications.

METHODS

Chemicals. Native TBA (5′-GGTTGGTGTGGTTGG), its complementary strand (5′-CCAACCACACCAACC), and 5′-FAM-labeled-TBA were purchased from Sigma-Aldrich Ltd. (Oakville, ON). The oligonucleotides were purified by Sigma-Aldrich using HPLC. All unmodified phosphoramidites (bz-dA-CE, ac-dC-CE, dmf-dG-CE, and dT-CE), activator (0.25 M 5-(ethylthio)-1*H*-tetrazole in CH_3CN), oxidizing agent (0.02 M I_2 in THF/pyridine/H_2O, 70/20/10, v/v/v), deblock (3% dichloroacetic acid in dichloromethane), cap A (THF/2,6-lutidine/acetic anhydride), cap B (methylimidazole in THF), and 1000 Å controlled pore glass (CPG) solid supports were purchased from Glen Research (Sterling, VA). Bovine thrombin was purchased from BioPharm Laboratories LLC (Bluffdale, Utah), while BSA was from Sigma-Aldrich Ltd.

Oligonucleotide Synthesis and Purification. All ^{Fur}dU-mTBA oligonucleotides were prepared on a 1 μmol scale using a BioAutomation MerMade 12 automatic DNA synthesizer using standard or modified β-cyanoethylphosphoramidite chemistry. Full synthetic details of the ^{Fur}dU phosphoramidite and solid-phase DNA synthesis have been previously published.[32] Upon completion of DNA synthesis, the crude mTBA oligonucleotide solutions were deprotected and cleaved from their solid support in aqueous ammonium hydroxide, filtered using syringe filters (PVDF 0.20 μm), and concentrated under diminished pressure. Samples were then resuspended in Milli-Q water (18.2 MΩ) and purified using an Agilent HPLC instrument equipped with an autosampler, a diode array detector (monitored at 258 and 316 nm), fluorescence detector (monitored at λ_{ex} = 316 nm and λ_{em} = 430 nm), and autocollector. Separation was carried out at 50 °C using a 5 μm reversed-phase (RP) semipreparative C18 column (100 × 10 mm) with a flow rate of 3.5 mL/min, and various gradients of buffer B in buffer A (buffer A = 95:5 aqueous 50 mM TEAA, pH 7.2/acetonitrile; buffer B = 30:70 aqueous 50 mM TEAA, pH 7.2/acetonitrile).

ESI-MS Analysis. MS experiments for identification of the mTBA oligonucleotides were conducted on a Bruker amaZon quadrupole ion trap SL spectrometer (Bruker Daltonics Ltd., Milton, ON). Oligonucleotide samples were prepared in 90% Milli-Q filtered water/10% methanol containing 0.1 mM ammonium acetate. Masses were acquired in the negative ionization mode with an electrospray ionization source (see SI for representative ESI-MS spectrum).

UV Melting Experiments. All melting temperatures (T_m) of TBA oligonucleotides were measured using a Cary 300-Bio UV−vis spectrophotometer (Agilent Technologies Inc., Santa Clara, CA) equipped with a 6 × 6 multicell block-heating unit using quartz (114-QS) 10 mm light path cells. Oligonucleotides were quantified using extinction coefficients obtained from http://www.idtdna.com/analyzer/applications/oligoanalyzer with mTBA assumed to have the same extinction coefficient as native TBA. Oligonucleotide samples were prepared in 100 mM phosphate buffer, at a pH of 7, with 100 mM MCl (M = Na^+ or K^+), using equivalent amounts (6.0 μM) of the unmodified or mTBA oligonucleotide and its complementary strand. The UV absorption at 260 nm (for duplex formation) and 295 nm (for GQ formation) was monitored as a function of temperature and consisted of forward−reverse scans from 10 to 90 °C at a heating rate of 0.5 °C/min, which was repeated five times. The T_m values were determined using hyperchromicity calculations provided in the Thermal software.

Circular Dichroism. Circular dichroism (CD) spectra were recorded on a Jasco J-815 CD spectropolarimeter (Jasco Inc., Easton, MD) equipped with a 1 × 6 Multicell block thermal controller and a water circulator unit. Spectra were collected at 10 °C between 200 and 400 nm, with a bandwidth of 1 nm and scanning speed at 100 nm/min. Spectra of solutions containing 6.0 μM DNA duplexes or GQs, prepared as described above, were recorded in quartz glass cells (110-QS) and were the averages of five accumulations that were smoothed using the Jasco software.

Fluorescence Measurements and Protein Titrations. All fluorescence spectra were recorded on a Cary Eclipse Fluorescence spectrophotometer (Agilent Technologies Inc., Santa Clara, CA) equipped with a 1 × 4 multicell block stirrer and temperature controller. Samples of mTBA (6 μM) were prepared in 100 mM phosphate buffer, at a pH of 7, with 100 mM MCl (M = Na^+ or K^+). All measurements were made using quartz cells (108.002F-QS) with a light path of 10 × 2 mm, and excitation and emission slit-widths were either 2.5 or 5 nm. Titrations proceeded according to previously published protocols[15,18] with minor variations. The mTBA samples (3 μM) were prefolded into the GQ structure in 100 mM potassium phosphate buffer, at a pH of 7, with 100 mM KCl at 25 °C. Fluorescence scans were taken 10 min following thrombin additions (5 μL from a 100 mM protein solution in 100 mM sodium phosphate pH 7.0 with 0.1 M NaCl at RT) to the mTBA solutions, until a final concentration of two equivalents of protein had been added. Fluorescence titration data were transformed into binding isotherms by calculating the fraction bound using Fraction Bound = $(F_{obs} - F_i)/F_{max} - F_i)$, where F_{obs} = observed fluorescence intensity, F_i is the initial fluorescence intensity, and F_{max} is the fluorescence intensity of the oligonucleotide when fully bound by thrombin. Plots of the fraction of mTBA bound versus [thrombin] generated binding isotherms that were analyzed with SigmaPlot 13.0 using one-site saturation to obtain K_d values. A FP method provided a K_d value for thrombin binding by 5′-FAM-labeled-TBA, as previously described.[18]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.6b00437.

Figure S1 (CD spectra), Figure S2 (fluorescence titration plots), NMR spectra of ^{Fur}dU phosphoramidite, and ESI-MS spectra of ^{Fur}dU-mTBA (PDF)

AUTHOR INFORMATION

Corresponding Author

*Tel.: 1-519-824-4120, x53963. E-mail: rmanderv@uoguelph.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 311600-2013 to R.A.M.).

REFERENCES

(1) Jayasena, S. D. (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem. 45*, 1628−1650.

(2) Bunka, D. H., and Stockley, P. G. (2006) Aptamers come of age − at last. *Nat. Rev. Microbiol. 4*, 588−596.

(3) Ng, E. W. M., Shima, D. T., Calias, P., Cunningham, E. T., Jr., Guyer, D. R., and Adamis, A. P. (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discovery 5*, 123−132.

(4) Astakhova, I. K., and Wengel, J. (2014) Scaffolding along nucleic acid duplexes using 2′-amino-locked nucleic acids. *Acc. Chem. Res. 47*, 1768−1777.

(5) Diafa, S., and Hollenstein, M. (2015) Generation of aptamers with an expanded chemical repertoire. *Molecules 20*, 16643−16671.

(6) Vaught, J. D., Bock, C., Carter, J., Fitzwater, T., Otis, M., Schneider, D., Rolando, J., Waugh, S., Wilcox, S. K., and Eaton, B. E. (2010) Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc. 132*, 4141−4151.

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

(7) Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S., and Hirao, I. (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol. 31*, 453−457.
(8) Imaizumi, Y., Kasahara, Y., Fujita, H., Kitadume, S., Ozaki, H., Endoh, T., Kuwahara, M., and Sugimoto, N. (2013) Efficacy of base-modification on target binding of small molecule DNA aptamers. *J. Am. Chem. Soc. 135*, 9412−9419.
(9) Zhao, J., Katsube, S., Yamamoto, J., Yamasaki, K., Miyagishi, M., and Iwai, S. (2015) Analysis of ATP and AMP binding to a DNA aptamer and its imidazole-tethered derivatives by surface plasmon resonance. *Analyst 140*, 5881−5884.
(10) Tsvetkov, V. B., Varizhuk, A. M., Pozmogova, G. E., Smirnov, I. P., Kolganova, N. A., and Timofeev, E. N. (2015) A universal base in a specific role: tuning up a thrombin aptamer with 5-nitroindole. *Sci. Rep. 5*, 16337.
(11) Li, T., Fu, R., and Park, H. G. (2010) Pyrrolo-dC based fluorescent aptasensors for the molecular recognition of targets. *Chem. Commun. 46*, 3271−3273.
(12) Zhang, D., Lu, M., and Wang, H. (2011) Fluorescence anisotropy analysis for mapping aptamer−protein interaction at the single nucleotide level. *J. Am. Chem. Soc. 133*, 9188−9191.
(13) Manderville, R. A., and Wetmore, S. D. (2016) C-Linked 8-aryl guanine nucleobase adducts: biological outcomes and utility as fluorescent probes. *Chem. Sci. 7*, 3482−3493.
(14) Sproviero, M., Fadock, K. L., Witham, A. A., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2014) Electronic tuning of fluorescent 8-aryl-guanine probes for monitoring DNA duplex-quadruplex exchange. *Chem. Sci. 5*, 788−796.
(15) Sproviero, M., and Manderville, R. A. (2014) Harnessing G-tetrad scaffolds within G-quadruplex forming aptamers for fluorescence detection strategies. *Chem. Commun. 50*, 3097−3099.
(16) Sproviero, M., Fadock, K. L., Witham, A. A., and Manderville, R. A. (2015) Positional impact of fluorescently modified G-tetrads within polymorphic human telomeric G-quadruplex structures. *ACS Chem. Biol. 10*, 1311−1318.
(17) Blanchard, D. J., Cservenyi, T. Z., and Manderville, R. A. (2015) Dual fluorescent deoxyguanosine mimics for FRET detection of G-quadruplex folding. *Chem. Commun. 51*, 16829−16831.
(18) Fadock, K. L., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2016) Optimization of fluorescent 8-heteroaryl-guanine probes for monitoring protein-mediated duplex→G-quadruplex exchange. *Org. Biomol. Chem. 14*, 4409−4419.
(19) Dumas, A., and Luedtke, N. W. (2010) Cation-mediated energy transfer in G-quadruplexes revealed by an internal fluorescent probe. *J. Am. Chem. Soc. 132*, 18004−18007.
(20) Dumas, A., and Luedtke, N. W. (2011) Highly fluorescent guanosine mimics for folding and energy transfer studies. *Nucleic Acids Res. 39*, 6825−6834.
(21) Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., and Toole, J. J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature 355*, 564−566.
(22) Alberti, P., and Mergny, J. (2003) DNA duplex−guadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl. Acad. Sci. U. S. A. 100*, 1569−1573.
(23) Nagatoishi, S., Isono, N., Tsumoto, K., and Sugimoto, N. (2011) Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie 93*, 1231−1238.
(24) Krauss, I. R., Merlino, A., Randazzo, A., Novellino, E., Mazzarella, L., and Sica, F. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res. 40*, 8119−8128.
(25) Pica, A., Krauss, I. R., Merlino, A., Nagatoishi, S., Sugimoto, N., and Sica, F. (2013) Dissecting the contribution of thrombin exosite I in the recognition of thrombin binding aptamer. *FEBS J. 280*, 6581−6588.
(26) Haidekker, M. A., and Theodorakis, E. A. (2010) Environment-sensitive behavior of fluorescent molecular rotors. *J. Biol. Eng. 4*, 11.
(27) Goh, W. L., Lee, M. Y., Joseph, T. L., Quah, S. T., Brown, C. J., Verma, C., Brenner, S., Ghadessy, F. J., and Teo, Y. N. (2014) Molecular rotors as conditionally fluorescent labels for rapid detection of biomolecular interactions. *J. Am. Chem. Soc. 136*, 6159−6162.
(28) Dziuba, D., Pohl, R., and Hocek, M. (2015) Polymerase synthesis of DNA labelled with benzylidene cyanoacetamide-based fluorescent molecular rotors: fluorescent light-up probes for DNA-binding proteins. *Chem. Commun. 51*, 4880−4882.
(29) Greco, N. J., and Tor, Y. (2005) Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. *J. Am. Chem. Soc. 127*, 10784−10785.
(30) Sinkeldam, R. W., Greco, N. J., and Tor, Y. (2008) Polarity of major grooves explored by using an isosteric emissive nucleoside. *ChemBioChem 9*, 706−709.
(31) Sinkeldam, R. W., Wheat, A. J., Boyaci, H., and Tor, Y. (2011) Emissive nucleosides as molecular rotors. *ChemPhysChem 12*, 567−570.
(32) Greco, N. J., and Tor, Y. (2007) Synthesis and site-specific incorporation of a simple fluorescent pyrimidine. *Nat. Protoc. 2*, 305−316.
(33) Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U. S. A. 90*, 3745−3749.
(34) Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E., and Tulinsky, A. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem. 268*, 17651−17654.
(35) Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S., and Bolton, P. H. (1993) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry 32*, 1899−1904.
(36) Smirnov, I., and Shafer, R. H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry 39*, 1462−1468.
(37) Johnson, J., Okyere, R., Joseph, A., Musier-Forsyth, K., and Kankia, B. (2013) Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res. 41*, 220−228.
(38) Blanchard, D. J. M., Fadock, K. L., Sproviero, M., Deore, P., Cservenyi, T. Z., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2016) Photophysical properties of push−pull 8-aryl-deoxyguanosine probes within duplex and G-quadruplex structures. *J. Mater. Chem. C 4*, 2915−2924.
(39) Tanpure, A. A., and Srivatsan, S. G. (2015) Conformation-sensitive nucleoside analogues as topology-specific fluorescent turn-on probes for DNA and RNA G-quadruplexes. *Nucleic Acids Res. 43*, e149.
(40) De Tito, S., Morvan, F., Meyer, A., Vasseur, J.-J., Cummaro, A., Petraccone, L., Pagano, B., Novellino, E., Randazzo, A., Giancola, C., and Montesarchio, D. (2013) Fluorescence emhancement upon G-quadruplex folding: synthesis, structure, and biophysical characterization of a dansyl/cyclodextrin-tagged thrombin binding aptamer. *Bioconjugate Chem. 24*, 1917−1927.

DOI: 10.1021/acschembio.6b00437
ACS Chem. Biol. 2016, 11, 2576−2582

Supporting Information for:

A Simple Molecular Rotor for Defining Nucleoside Environment within a DNA Aptamer-Protein Complex

Thomas Z. Cservenyi, Abigail J. Van Riesen, Florence D. Berger, Ahmed Desoky and Richard A. Manderville*

Table of Contents:

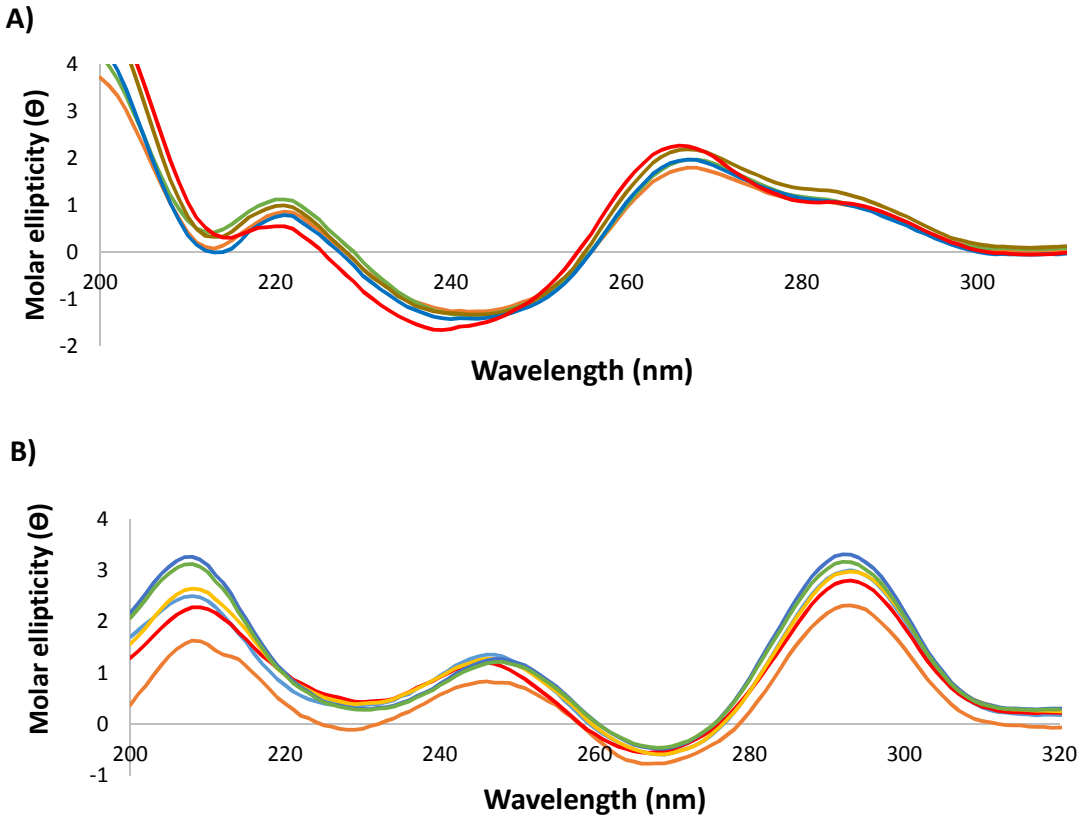
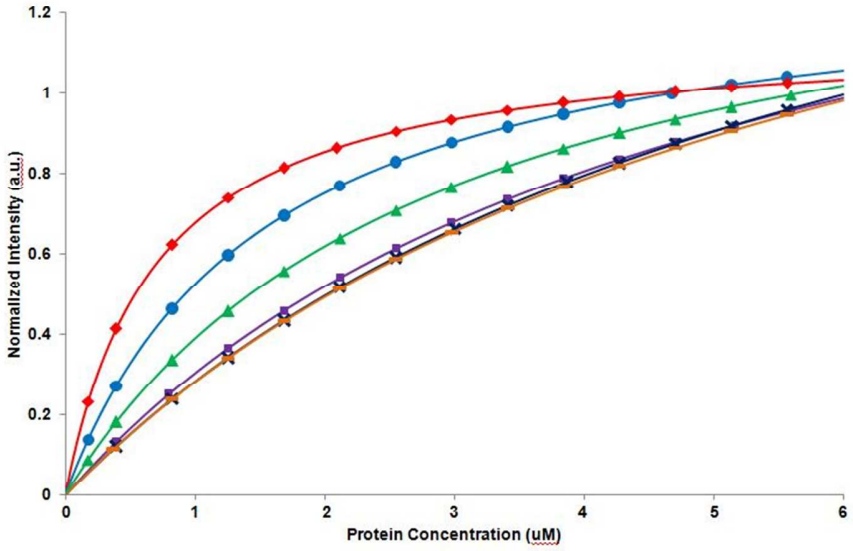


Figure. S1 Circular dichroism spectra of ^{Fur}dU-mTBA (6 μM) recorded at 10 °C. A) Duplexes with 1 equiv. complementary strand in 100 mM HNa_2PO_4 buffer pH 7.0, 100 mM NaCl. B) GQs in 100 mM HK_2PO_4 buffer pH 7.0, 100 mM KCl; ^{Fur}dU-3 (red), ^{Fur}dU-4 (green), ^{Fur}dU-7 (orange), ^{Fur}dU-9 (grey), ^{Fur}dU-12 (yellow), ^{Fur}dU-13 (blue).

A)



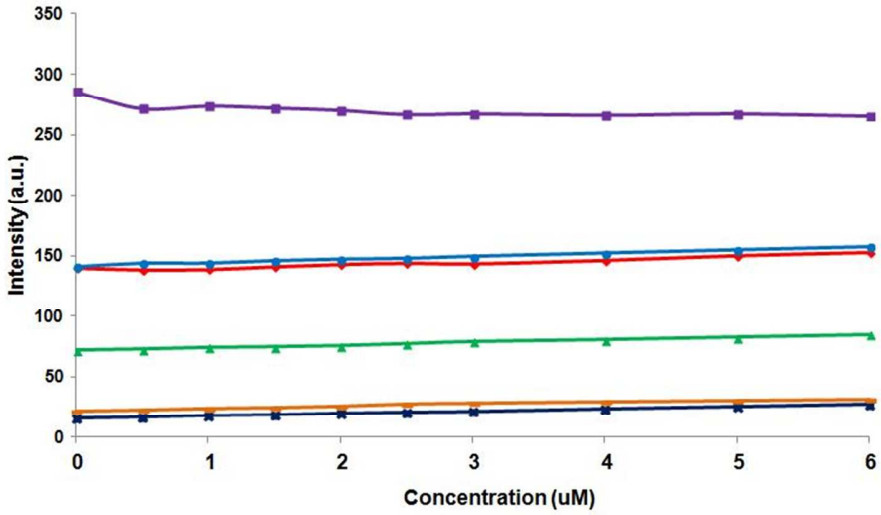
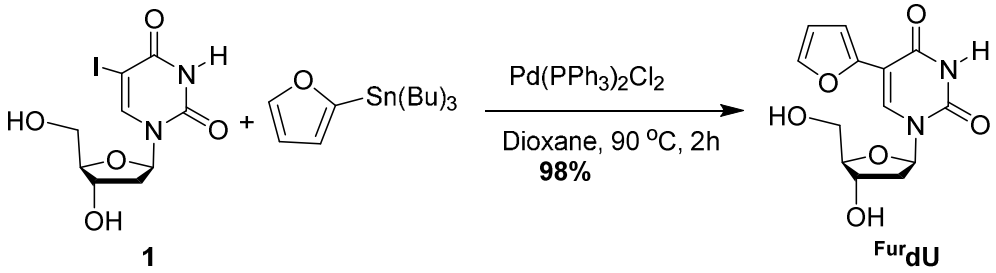


Figure S2. A) Normalized emission intensity of mTBA upon successive addition of thrombin, B) emission intensity changes upon successive addition of BSA; ^{Fur}dU-3 (red), ^{Fur}dU-4 (dark blue), ^{Fur}dU-7 (green), ^{Fur}dU-9 (orange), ^{Fur}dU-12 (light blue), ^{Fur}dU-13 (purple).

Synthesis of ^{Fur}dU-Phosphoramidite

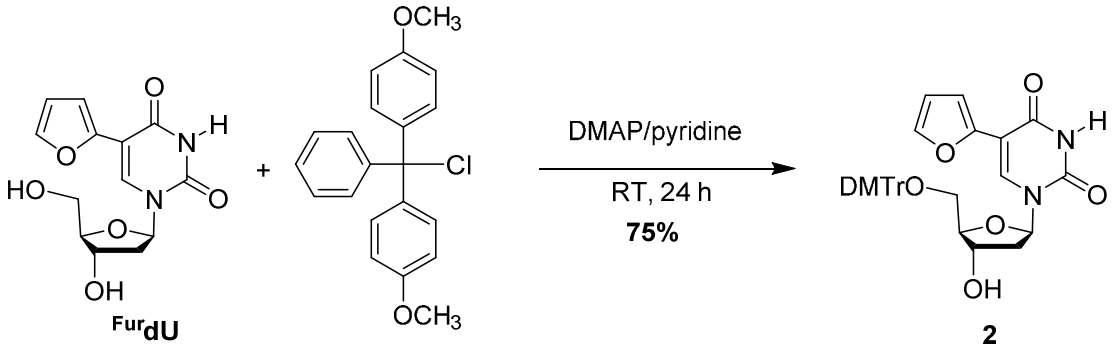
The ^{Fur}dU-phosphoramidte was prepared according to the literature procedure (Greco, N. J., and Tor, Y. (2007) Synthesis and site-specific incorporation of a simple fluorescent pyrimidine. *Nat. Protoc. 2*, 305-316).

(i) Synthesis of 5-(fur-2-yl)-2′-deoxyuridine (^{Fur}dU)



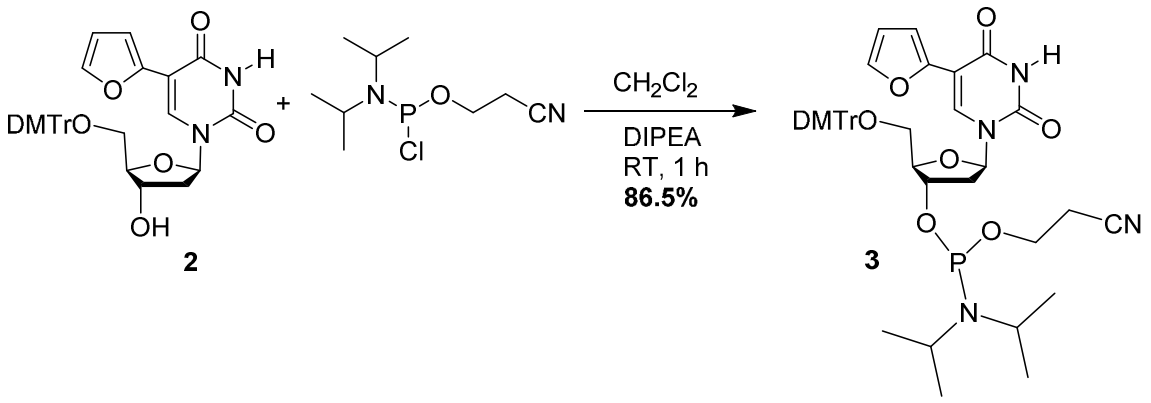
In a flame dried 2-neck 100-mL flask, a mixture of 5-iodo-2′-deoxyuridine (3g, 8.47 mmol) and dichlorobis(triphenylphosphine)Pd(II) (0.12 g, 0.17 mmol) under nitrogen atmosphere in anhydrous dioxane (60 ml) was added 2-(tributylstannyl)furan (11.5 ml, 36.43 mmol) via a syringe. The suspension was refluxed at 90 ºC under nitrogen atmosphere for 2 h, where TLC (10% MeOH in $CHCl_3$) indicates that the reaction is complete. The reaction mixture was cooled to room temperature and filtered through celite 545. The celite was washed with 150 mL methanol; the solvent was removed under reduced pressure where the resulting oil was triturated with hexanes (3x 100 mL) to produce a solid. The resulting solid was taken up in a minimum of hot solvent (1/1 methanol/chloroform) and precipitated from hexanes. The formed precipitate was filtered, washed with hexane then dried under vacuum to give the desired product as white solid (2.48 g, 98 % yield). ^{1}H NMR (300 MHz, DMSO-d_6): δ 11.62 (s, NH, 1H), 8.32 (s, 1H), 7.60 (s, 1H), 6.84 (d, *J* = 2.8 Hz, 1H), 6.51 (m, 1H), 6.22 (t, *J* = 6.42 Hz, 1H), 5.26 (d, *J* = 4.4 Hz, 1H), 5.08 (t, *J* = 4.4 Hz, 1H), 4.27 (m, 1H), 3.82 (m, 1H), 3.59 (m, 2H), 2.16 (t, *J* = 5.4 Hz, 2H).

(ii) Synthesis of 5′-dimethoxytrityl-5-(fur-2-yl)-2′-deoxyuridine (2)



A mixture of ^{Fur}dU (0.7 g, 2.37 mmol), DMTr-Cl (0.96 g, 2.85 mmol) and DMAP (0.08 g, 0.65 mol) in anhydrous pyridine (20 ml) was stirred at room temperature under nitrogen atmosphere for 16 h. The solvent was evaporated under reduced pressure. The product was purified by flash column chromatography ($CHCl_3$: MeOH; Et_3N; 97: 2: 1). The desired product was isolated as white foam (1.06 g, 75% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.16 (s, 1H), 7.43 (bd, 2H), 7.33- – 7.30 (m, 4H), 7.21- 7.12 (m, 4H), 6.95 (d, *J*= 3 Hz, 1H), 6.77 – 6.72 (m, 5H), 6.41 (t, *J* = 7.03 Hz, 1H), 6.26 (m, 1H), 4.4 (m, 1H), 3.71 (s, 6H), 3.5 (m, 1H), 3.32 (m, 1H), 2.51 (m, 1H), 2.33 – 2.24 (m, 1H).

(iii) 3′-(2-cyanoethyl)-diisopropylphosphoramidite-5′-dimethoxytrityl -5-(fur-2-yl)-2′-deoxyuridine (3)



In a flame dried 50-mL RBF, 5′-Dimethoxytrityl-5-(fur-2-yl)-2′-deoxyuridine **2** (0.78 g, 1.3 mmol) was dissolved in anhydrous CH_2Cl_2 (20 mL) in the presence of molecular sieves 4Å after removal of the air from the flask under nitrogen atmosphere. The reaction mixture was stirred at room temperature under nitrogen atmosphere for 20 minutes. *N,N*-Diisopropylethylamine (0.68 mL, 3.92 mmol) was added to the reaction mixture followed by 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (0.5 g, 2.09 mmol) via a syringe. The reaction mixture was stirred at room temperature for 1 h (following the disappearance of the starting material by TLC; ethyl acetate/hexane; 1/1 + 1% Et_3N). The reaction mixture was diluted with CH_2Cl_2, filtered to remove the molecular sieves and concentrated under reduced pressure. The product was purified by flash column chromatography (1/1 ethyl acetate/hexanes, 1% Et_3N). The product was obtained as a light brown foam (0.9 g, 86.5% yield). ^{1}H NMR (300 MHz, DMSO-*d*6): δ 11.60 (s, NH, 1H), 7.98 (s, 1H), 7.35 (d, 2H), 7.24 – 7.15 (m, 8H), 6.80 – 6.76 (m, 5H), 6.43- 6.41 (m, 1H), 6.14 (t, *J* = 6.4 Hz, 1H), 4.45 – 4.43 (m, 1H), 4.10 – 4.08 (m, 1H), 3.66 (s, 6H), 3.58- 3.44 (m, 4H), 3.20 – 3.15 (m, 2H), 2.63 – 2.59 (t, *J* = 6.6 Hz, 2H), 1.16– 1.05 (m, 12H); ^{31}PNMR (162 MHz, DMSO-*d*6, referenced to H_3PO_4): δ 150.06, 149.7.

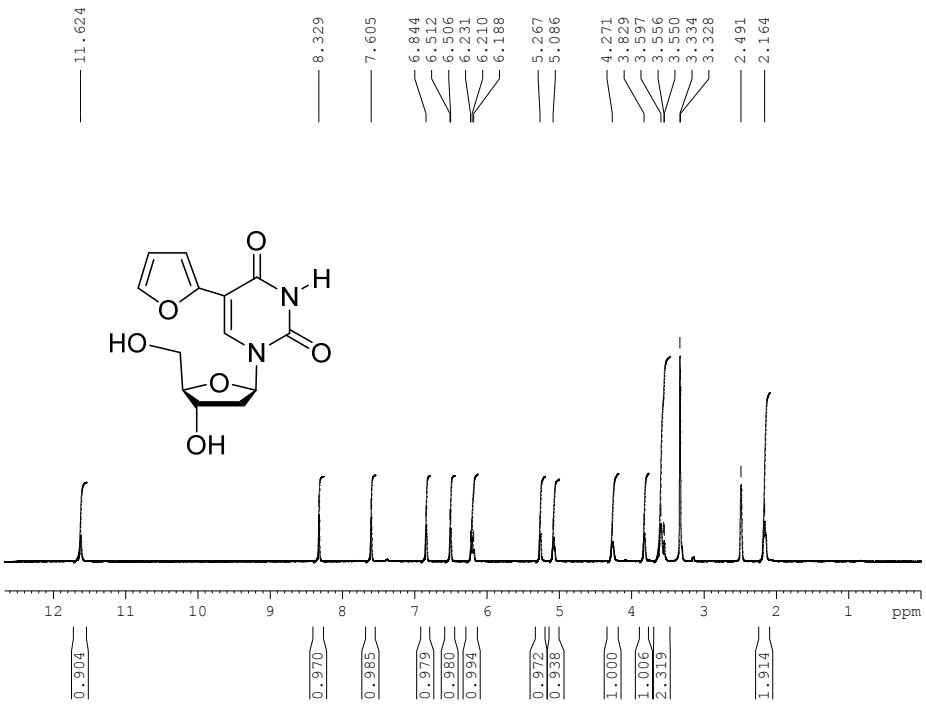


Figure S3. ^{1}H NMR spectrum of ^{Fur}dU in DMSO-d_6.

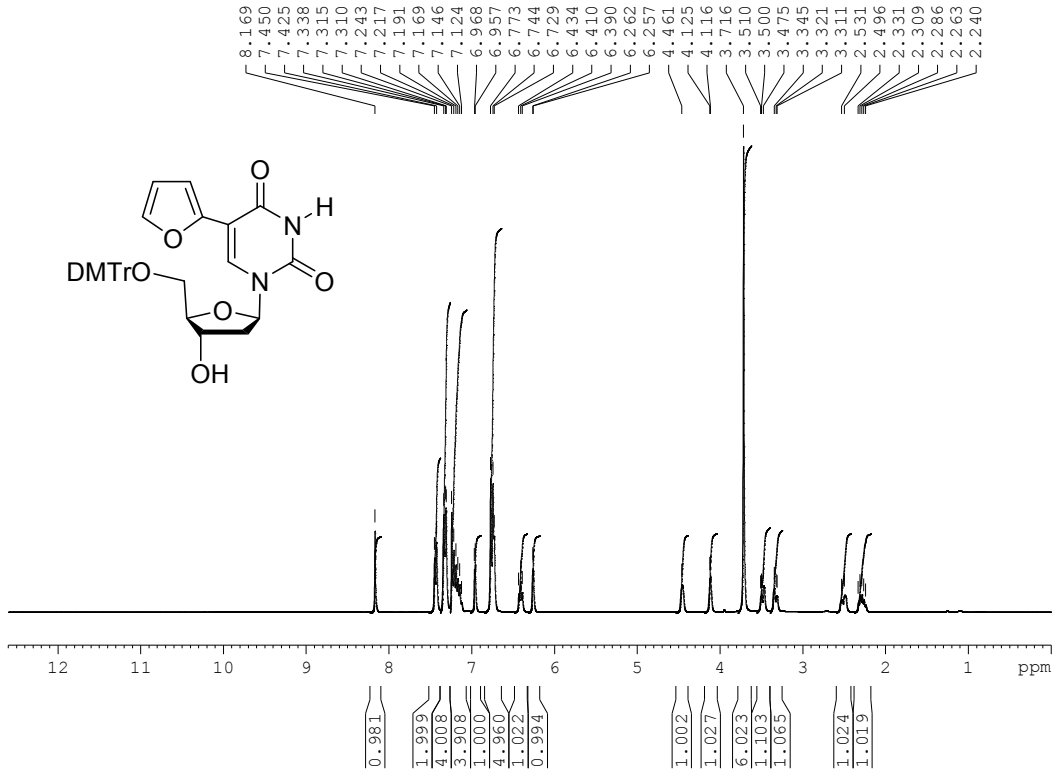


Figure S4. ^{1}H NMR spectrum of **2** in $CDCl_3$.

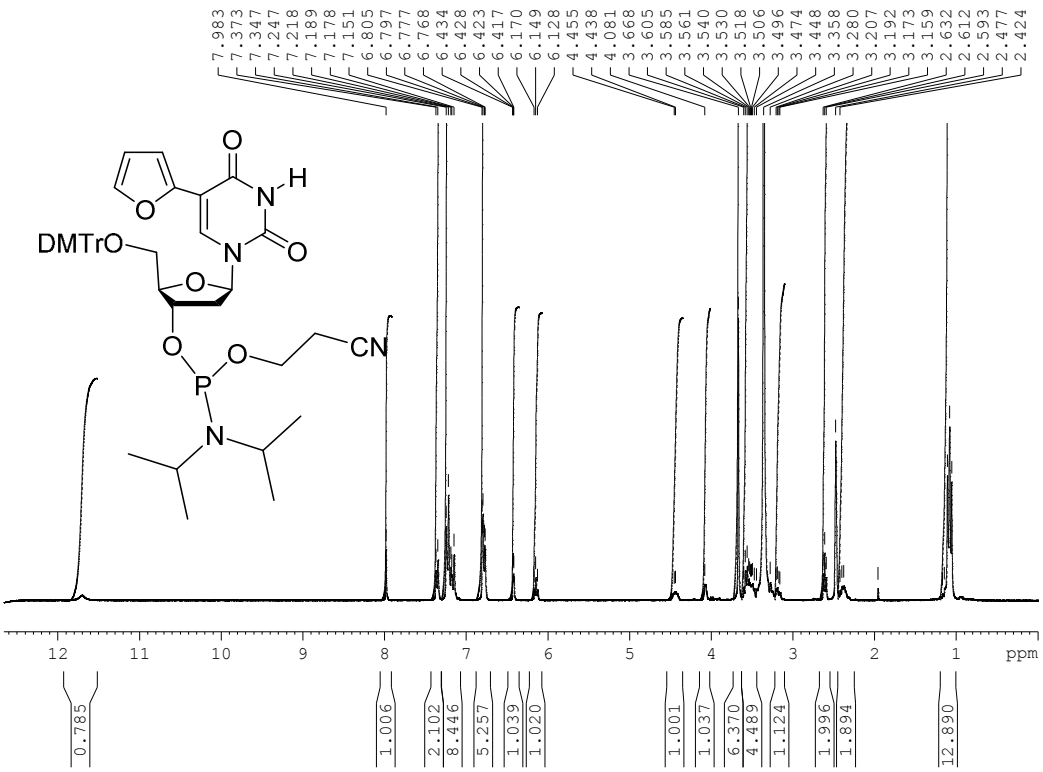


Figure S5. ^{1}H NMR spectrum of **3** in DMSO-d_6.

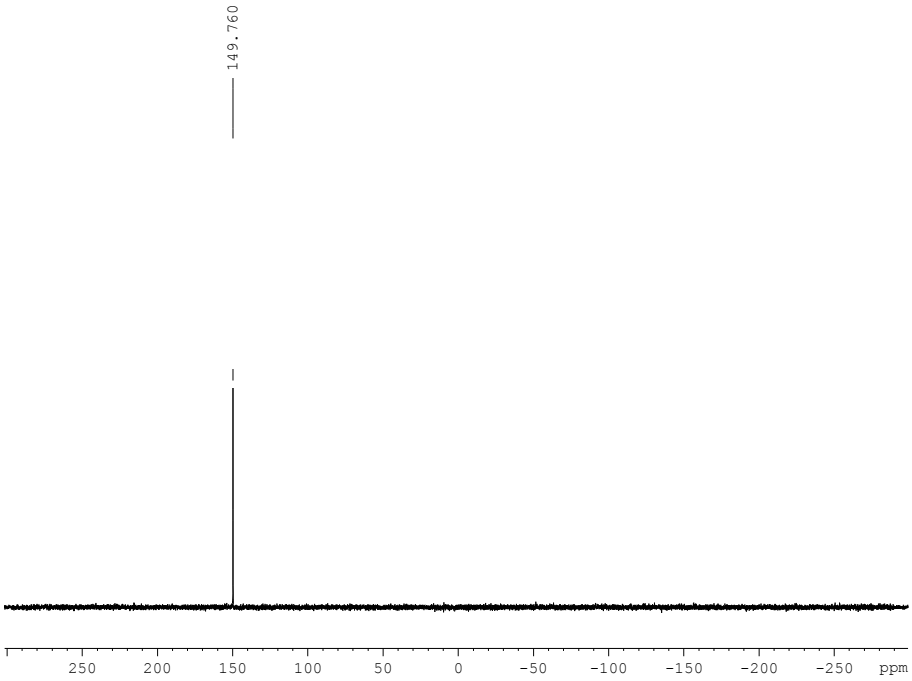


Figure S6. ^{31}P NMR spectrum of **3** in DMSO-d_6.

Table S1: Tabulated MS data of ^{Fur}dU mTBA strands.

mTBA	Formula	Calc. Mass	*m/z* (ESI$^-$)	Exp. Mass
^{Fur}dU-3	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 529.8	4777.2
			$[M-8H]^{8-}$ = 596.5	4780.0
			$[M-7H]^{7-}$ = 681.6	4778.2
			$[M-6H]^{6-}$ = 795.2	4777.2
			$[M-5H]^{5-}$ = 954.7	4778.5
^{Fur}dU-4	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 530.1	4779.9
			$[M-8H]^{8-}$ = 596.4	4779.2
			$[M-7H]^{7-}$ = 681.7	4778.9
			$[M-6H]^{6-}$ = 795.4	4778.4
			$[M-5H]^{5-}$ = 954.9	4779.5
^{Fur}dU-7	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 530.1	4779.9
			$[M-8H]^{8-}$ = 596.2	4777.6
			$[M-7H]^{7-}$ = 681.8	4779.6
			$[M-6H]^{6-}$ = 795.5	4779.0
			$[M-5H]^{5-}$ = 954.3	4776.5
^{Fur}dU-9	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 530.1	4779.9
			$[M-8H]^{8-}$ = 596.4	4779.2
			$[M-7H]^{7-}$ = 681.7	4778.9
			$[M-6H]^{6-}$ = 795.1	4776.6
			$[M-5H]^{5-}$ = 954.7	4778.5
^{Fur}dU-12	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 529.8	4777.2
			$[M-8H]^{8-}$ = 596.4	4779.2
			$[M-7H]^{7-}$ = 681.5	4777.5
			$[M-6H]^{6-}$ = 795.2	4777.2
			$[M-5H]^{5-}$ = 954.8	4779.0
^{Fur}dU-13	$C_{153}H_{187}N_{57}O_{95}P_{14}$	4778.1	$[M-9H]^{9-}$ = 530.2	4780.8
			$[M-8H]^{8-}$ = 596.4	4779.2
			$[M-7H]^{7-}$ = 681.8	4779.6
			$[M-6H]^{6-}$ = 795.1	4776.6
			$[M-5H]^{5-}$ = 954.5	4777.5

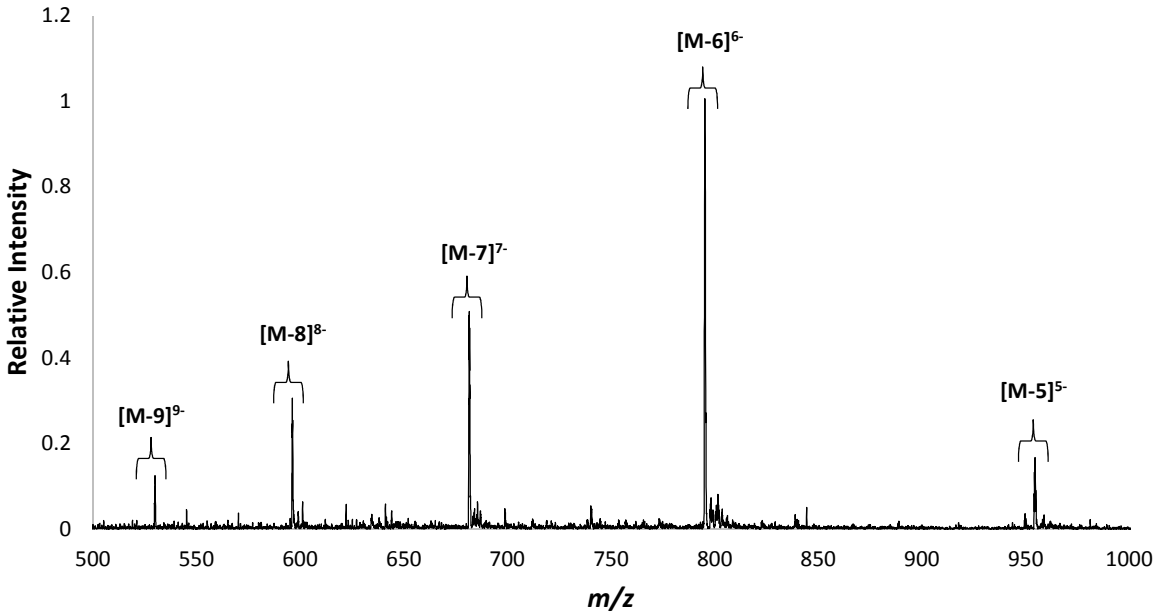


Figure S7. Representative ESI⁻ MS Spectrum of ^{Fur}dU-mTBA.

Curriculum vitae

Florence Daniela Berger

Date of birth: 04.08.2018
Nationality: Swiss
E-Mail: berger.florence@hotmail.com

Education

09/2014 – 02/2018	**Doctorate,** Laboratory of Toxicology, ETH Zürich, Switzerland Bioactive Compounds Doctoral Program Thesis title: *Fluorescent oligonucleotides as molecular mechanistic probes to understand mutagenicity and monitor target binding.* Supervisors: Prof. Shana J. Sturla, ETH Zürich, Switzerland and Prof. Richard A. Manderville, University of Guelph, Canada.
09/2014 – 12/2015	**Visiting Doctoral Student,** University of Guelph, ON, Canada
02/2012 – 04/2014	**Master of Science in Food Science,** ETH Zürich, Switzerland Major: Food Processing, Minor: Safety and Quality in Agri-Food-Chain Masterthesis title: *Synthesis and Properties of Duplex DNA Containing Oxygen-Linked Deoxyguanosine Adducts Derived from Phenolic Carcinogens* Masters research conducted at University of Guelph, Guelph, Canada. Supervisors: Prof. Shana J. Sturla, ETH Zürich, Switzerland and Prof. Richard A. Manderville, University of Guelph, Canada.
09/2008 – 07/2011	**Bachelor of Science in Food Science,** ETH Zürich, Switzerland
09/2004 – 06/2008	**Matura,** Kollegium Heilig Kreuz, Fribourg, Switzerland Major: Spanish

Professional Related Activities

11/2016 - 09/2017	**Ethical Approval:** Association of human colon health status with food-related cancer biomarkers measured in stool samples: single point feasibility study in participants scheduled for routine screening colonoscopy Successfully submission of ethical permission at the Swiss Ethics Committees on research involving humans
01/2015 – 12/2015	**Industry project** in collaboration with Aterica Health Inc., Waterloo, ON, Canada. Proof of concept for the development of a diagnostic tool to detect allergen, with a future patent application.

Publications

Berger, F.D., Manderville, R.A., Sturla, S.J. (2017) Fluorescence profiles of synthetically modified DNA duplexes reveal how sequence-dependent base conformation heterogeneity gives rise to frameshift mutations. In preparation

Dahlmann, H.D.,‡ Berger,F.D.‡ Kung, R.W., Wyss, L.A., Gubler, I., McKeague, M., Wetmore, S.D., and Sturla, S.J. (2017) Fluorescent elongated hydrophobic nucleobase analogues stabilize DNA duplexes containing *O*6-alkylguanine adducts. Submited and accepted for reviews at Helvetica Chimica Acta. ‡Designates equal contributions. In preparation.

Berger, F.D., Sturla, S.J., Kung, R. W., Montina, T., Wetmore, S. D., and Manderville, R.A. (2017) Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion

in the *Nar*I Mutational Hotspot: Evidence for Enhanced *syn*-Adduct Formation. Chem. Res. Toxicol. 10.1021/acs.chemrestox.7b00266

Cservenyi, T.Z., Van Riesen, A.J., Berger, F.D., Desoky, A. and Manderville, R.A. (2016) A Simple Molecular Rotor for Defining Nucleoside Environment within a DNA Aptamer-Protein Complex. *ACS Chem. Biol.*, **11**, 2576–2582.

Reif, C., Arrigoni, E., Berger, F., Baumgartner, D. and Nyström, L. (2013) Lutein and β-carotene content of green leafy Brassica species grown under different conditions. *LWT - Food Sci. Technol.*, **53**, 378–381.

Presentations

Annual Meeting 2017 of the Swiss Society of Toxicology: Berger, F.D., Manderville, R.A., and Sturla, S.J. *Mechanism of Frameshift Mutation induced by genotoxic compounds*. Basel, Switzerland, December 1, **2017.** (Selected for oral presentation)

VVIIth Symposium on Chemistry of Nucleic Acid Components 2017: Berger, F.D., Manderville, R.A., and Sturla, S.J. *Fluorescent Modified DNA to Detect Frameshift Mutations*, Cesky Krumlov, Czech Republic, June **2017.** (Poster)

Swiss Chemical Society Fall Meeting 2016: Berger, F.D., Manderville, R.A., and Sturla, S.J. *Fluorescent Modified DNA to Detect Frameshift Mutations*, Zurich, Switzerland, September **2016.** (Poster)

ACS 248th National Meeting 2014, Berger, F.D., Manderville, R.A., and Sturla, S.J. *Synthesis and Properties of Duplex DNA Containing Oxygen-Linked Deoxyguanosine Adducts Derived from Phenolic Carcinogens*, San Francisco, CA, USA, August **2014**. (Poster)

96th Canadian Chemistry Conference and Exhibition. Berger, F.D., Kuska M.S., Witham A.A., Sturla S.J., Manderville R.A. *Synthesis and Properties of Duplex DNA Containing Oxygen-Linked Deoxyguanosine Adducts Derived from Phenolic Carcinogens*. (BM6 00341) Quebec City, May **2013**. (Poster)

Languages

German	native language
English	business fluent
French	fluent
Spanish	intermediate

Continuing Education

Safety Pharmacology Course. Swiss Society of Toxicology part of DCPT Modular Training Programme in Toxicology. February 2017

Fremdstoffmetabolismus/Toxikokinetik Course. Swiss Society of Toxicology part of DCPT Modular Training Programme in Toxicology. August 2017