2013

Sequence Effects on the Repair and Replication of Arylamine-DNA Adducts

Vipin Jain
University of Rhode Island, vipinj_11@yahoo.co.in

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

---

Recommended Citation

Jain, Vipin, "Sequence Effects on the Repair and Replication of Arylamine-DNA Adducts" (2013). *Open Access Dissertations*. Paper 22.
https://digitalcommons.uri.edu/oa_diss/22

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

**SEQUENCE EFFECTS ON THE REPAIR AND REPLICATION OF
ARYLAMINE-DNA ADDUCTS**

**BY
VIPIN JAIN**

**A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOMEDICAL AND PHARMACEUTICAL SCIENCES**

UNIVERSITY OF RHODE ISLAND

2013

DOCTOR OF PHILOSOPHY DISSERTATION

OF

VIPIN JAIN

APPROVED:

Dissertation Committee:

Major Professor   Dr. Bongsup Cho

Dr. Roberta King

Dr. Brenton DeBoef

Dr. Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2013

ABSTRACT

Cancer is a collection of diseases defined by uncontrolled cell growth and dissemination of abnormal cells. It is the second most deadly disease in the United States. Exposure to chemicals, both endogenous and exogenous is one of the major causes of cancer initiation. The role of chemicals in carcinogenesis was observed back in the eighteenth century when surgeon Pott observed a large number of scrotal cancer cases in chimney sweeps and associated it with to the soot exposure. However, research in the field of chemical carcinogenesis was renewed when Yamagawa and Ichikawa reported the development of malignant tumor on the rabbit's ear due to the regular application of coal tar. Later, it was discovered that the presence of the polycyclic aromatic hydrocarbon benzo[a]pyrene in the coal tar was responsible for its tumorogenicity. Following that finding, researchers have made significant advancement in the field of chemical carcinogenesis.

Majority of the chemical carcinogens are genotoxic in nature which exert their effect primarily by reacting with the DNA. These chemicals are either reactive as such or undergo enzymatic activation into the electrophilic species that enable them to form covalent adducts with the electron-rich DNA. Therefore, DNA-adduct formation is considered as a hallmark of chemical carcinogenesis. DNA-adducts are formed regularly in the cells as the genomic DNA is under regular attack by various chemicals. Cells have developed various repair mechanisms such as nucleotide excision repair (NER), base excision repair (BER) to protect the DNA from the effect of these adducts. Despite the effectiveness of these defense mechanisms some of the adducts have the potential to

evade them. Unrepaired adducts can enter in the replication cycle and affect the process which creates a possibility of various kinds of mutation induction. Mutation on the genes that control the cell growth may trigger cancer initiation. For example, mutation of tumor suppressor p53 gene is regularly seen in the majority of sporadic human cancers.

Arylamines is an important group of chemical carcinogens that are implicated in the etiology of human cancers. 2-Naphthylamine, benzidine, 4-aminobiphenyl and 2-acetylaminofluorene are among the well-known arylamine carcinogens. 2-Acetylaminofluorene was originally developed as a pesticide, but was banned because of its liver carcinogenicity on animals. However, it is used continuously by researchers as a model carcinogen to study the chemical carcinogenesis. Arylamines are not reactive per se but undergo metabolic transformations in the body to produce highly reactive nitrenium ion, which bind with DNA specifically at C-8 position of guanine to yield two major C8-substituted dG adducts: *N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene (dG-C8-AF) and *N*-(2'-deoxyguanosin-8-yl)-2-acetylaminofluorene (dG-C8-AAF). Similarly, bladder carcinogen 4-aminobiphenyl forms *N*-(2'-deoxyguanosin-8-yl)-4-aminobiphenyl (dG-C8-ABP). The NMR studies on ABP, AF, and AAF adducted duplexes found that they exist in a mixture of two prototype conformers; *anti*-glycosidic B-conformer in which the fluorenyl moiety orients in the major groove of double helical DNA without affecting the Watson-Crick base pairing and *syn*-glycosidic stacked-conformer (S) in which the carcinogen is inserted into the duplex resulting into flipping out of the modified guanine. Along with the above two conformers, AAF additionally produces a *syn*-glycosidic wedge-conformer (W) with the fluorene moiety in the narrow minor

groove of the duplex. The population of these conformers was found to be dependent on the nature of the lesion and the bases surrounding it.

Arylamine-DNA adducts are good substrates for NER, a repair pathway responsible for removing bulky lesions. The repair efficiency of NER proteins are found to be affected by different chemical, structural and conformational factors. The nature of bases surrounding the lesion site is an important factor which is generally stated as sequence effect. The NER of arylamines exhibits dramatic sequence effects. One of the most significant effect was witnessed on the reparability of AAF adduct involving *E. coli* UvrABC and human exonuclease NER systems, where the lesion presents separately on each of the three guanines of the *NarI* sequence (5'-...CG₁G₂CG₃CC...-3'). It was discovered that in *E. coli* and human exonuclease the relative repair efficiencies of AAF at G₁, G₂ and G₃ were in a ratio of 100:18:66 and 38:100:68, respectively. But, the rationale behind this sequence dependent modulation of NER is not clear.

We hypothesize that adduct induced conformational heterogeneity is modulated by the base sequences around the lesion and the thermal/thermodynamic stability of the lesion containing DNA duplexes is an important factor for determining their repair efficiencies. In **Manuscript I** (published in *Nucleic Acids Research*), the objective was to test our hypothesis on the aforementioned sequence dependent NER of AAF in the *NarI* hot spot sequence. To that end, we prepared three 16-mer oligonucleotide duplexes that are site specifically modified by FAAF (fluorinated derivative of AAF) at G₁, G₂ or G₃ of the *NarI* sequence (5'-CTCTCG₁G₂CG₃CCATCAC-3'). We utilized ¹⁹F NMR and circular

dichroism (CD) to determine the conformational profiles of the FAAF in the three duplexes. We conducted melting experiments using UV-spectrophotometer and differential scanning calorimetry (DSC) for thermodynamic analyses. In addition, we carried out NER assay of these duplexes using the *E. coli* UvrABC system. Similarly, the repair work was also performed on FAF (fluorinated derivative of the N-deacetylated AF) modified duplexes as their conformational and thermodynamic properties were known.

Our ^{19}F NMR/CD data showed that the FAAF at G_1 and G_3 orients majorly in syn-glycosidic S- and W-conformers, whereas the anti glycosidic B-conformer is favored at G_2 . The thermodynamic data obtained from UV and DSC indicates that the S- and W-conformers induce greater distortion and thermodynamic destabilization of the modified duplexes. Interestingly, we found that the repair of FAAF occurred in a conformation-specific manner with highly S/W-conformeric AAF at G_3 and G_1 incised more efficiently than the B-type G_2 ($G_3 \sim G_1 > G_2$). The N-deacetylated AF adducts in the same *NarI* sequence were repaired 2-3 fold less than the bulky N-acetylated FAAF, but the order of reparability was different ($G_2 \sim G_1 > G_3$), a reverse order of the FAAF. We believe the presence of N-acetyl group raised the conformational barrier of FAAF vs. FAF. Overall, the results provided conformational and thermodynamic insight into the sequence dependent UvrABC incisions of the bulky aminofluorene DNA adducts.

In **Manuscript II** (to be submitted in *Chemical Research in Toxicology*), we extended our NER mono-adducts work (Manuscript I) to FAAF derived di-adducts formed from the *NarI* sequence. The objective was to understand how the cluster adducts are treated

by the NER proteins and to investigate the sequence effects in the repair efficiency of the arylamine di-adducts, if any. To do so, we prepared modified oligonucleotides that contain FAAF simultaneously on two (G_1G_2 , G_2G_3 and G_1G_3) of the three guanines (G_1 , G_2 , and G_3) of the *NarI* sequence (5'-CTCTCG₁G₂CG₃CCATCAC-3'). We conducted the NER assays on the three di-adducts using the *E. coli* UvrABC proteins. Moreover, we performed spectroscopic (^{19}F NMR and CD) and melting experiments for conformational and thermodynamic examination, respectively.

We found that the repair of the bulky FAAF di-adducts occurred with double the efficiency of the mono-adducts. Our structural data related the enhanced reparability with the greater thermal and thermodynamic destabilization produced in the di-adduct duplexes. In addition, the three di-adduct duplexes exhibited dramatic sequence effects in the repair efficiency. Although, the ^{19}F NMR data was not conclusive to provide their conformational profile, the respective mono-adducts data (Manuscript I) indicate that the repair seems to occur in a conformation-specific manner. The *NarI*- G_2G_3 and $-G_1G_3$ di-adducts are repaired more efficiently and they contain FAAF at G_3 position, in which FAAF is known to exist in highly S/W-conformation. In contrast, the *NarI*- G_1G_2 di-adduct contains the highly B-conformeric G_2 position. Apparently, duplex distortion by destacking and destabilization is important as the repair and destacking/destabilization trend followed in the same order: $G_2G_3 > G_1G_3 > G_1G_2$. The results of this study provide the first report on the structure-repair relationships of FAAF cluster adducts.

Arylamine adducts are also capable of inducing mutation. For example, AF produces point mutation and AAF results into a mixture of point and frameshift mutations. Aforementioned sequence effects in the *NarI* sequence are not limited to repair, but were also seen in the mutational outcome. Previously, it has been reported that AF and AAF adducts yield significantly higher frequency of -2 frameshift mutation only when adducted to third guanine (G₃) of the *NarI* sequence (5'--G₁G₂CG₃CN--3'). In addition, the mutational rate was found to be governed by the base at *N* position. The presence of cytosine (C) at position *N* resulted into relatively higher rate of mutation as compared to thymine (T). Structural studies from our laboratory have shown that the FAF adduct adopts stacked (S) conformer when *N* is = C, but displayed conformational heterogeneity when C is replaced by T. In **Manuscript III** (published in *Nucleic Acids Research*), our focus was to determine whether the above 3'-next flanking base (*N*) effect on the conformational property is limited to AF modified *NarI* sequence or can also exist in random G*CN duplexes modified by different arylamines (ABP, AF, AAF).

We prepared two 11mer duplexes (5'-CCATCGCNACC-3', N= T or A) having the same G*CN sequence context as in *NarI* sequence and modified them with FABP, FAF or FAAF. We utilized various techniques including ¹⁹F NMR, CD, DSC, gel electrophoresis (EMSA) to study the structural properties of the two duplexes. In addition we performed molecular dynamics/potential of mean force (MD/PMF) calculations. We also performed *E.coli* UvrABC based NER assays to determine the NER significance of 3'-next flanking base (*N*) effect. Our results showed that the A to T polarity swap in the G*CA/G*CT transition produced a dramatic increase in destabilized stacked

conformation, but resulted in unexpected 2~3-fold lower NER efficiencies. However, the results were consistent with lesion-induced DNA bending/distortion determined by CD, EMSA and MD/PMF calculations. As for the lesions, FAAF displayed 3~4 times better reparability than FABP and FAF lesions, which is consistent with the magnitude of bending and helix destabilization. Overall, the results suggest that lesion-induced DNA bending/thermodynamic destabilization is a more important NER factor than the usual S/B conformational heterogeneity, as has been observed previously for AF and AAF in certain sequence contexts.

In **Manuscript IV** (to be submitted in *Chemical Research in Toxicology*), we extended our previous work (Manuscript III) to examine the effect of 3'-next flanking T induced conformational heterogeneity on the DNA polymerase mediated translesion synthesis (TLS). Initially, we carried out chemically simulated TLS (without DNA polymerase) to deduce the conformational property (using ^{19}F NMR and CD) of FAF/FABP adducts at the replication fork and during primer elongation across modified G*CA and G*CT templates. Following conformational analysis, we performed DNA polymerase mediated running start and steady state kinetics experiments to probe the effect on the insertion efficiency and fidelity of DNA polymerases. FAF and FABP in the G*CA and G*CT sequences were found to adopt a similar conformational profiles at the replication fork which is in contrast to the usual duplex. Due to the similar conformational profiles, no substantial 3'-next flanking base effect was observed in the base insertion efficiency at the replication fork. However, we observed some lesion effect in polymerase stalling around the lesion site, i.e., the anti-conformer FABP showed more stalling at the

preinsertion site (n-1) whereas the syn-conformeric FAF at the lesion site (n). Overall, there was no significant 3'-next flanking base effect on the activity of polymerases. However, the results provide valuable insights on the role of lesion-induced conformational heterogeneity in modulating the translesion DNA synthesis.

In summary, the studies described in this Thesis provided the structural and thermodynamic rationales for the sequence- and lesion-dependent repair and replication of the bulky arylamine-DNA adducts. In addition, we demonstrated how the global duplex distortion overpowers the local conformational disturbances as an important damage recognition factor. To the best of our knowledge, this is the first report on the structure and repair characteristics of arylamine derived cluster adducts along with the dramatic sequence effects. Our results also provided a glimpse of replication across the arylamine adduct and the role of adduct conformation in modulating the process of TLS. Overall, we established a strong and sensible structural-function relationships between the structure/conformation of arylamine adducts and their mutational and repair outcomes.

ACKNOWLEDGEMENTS

The completion of this doctoral dissertation has been one of the most significant academic challenges I have ever had to face. It would not have been possible without the help and support of the kind people around me. There are many people to thank who have been cornerstones of my journey through graduate school and finish, to only some of whom it is possible to give particular mention here.

Above all, I am indebted to my major professor, **Dr. Bongsup Cho**, for giving me an opportunity to work with him as a graduate student and for leading me through the many challenging facets of my research career. This dissertation would not have been possible without his patience, constant words of encouragement and immense support, not to mention his guidance and unsurpassed knowledge of medicinal chemistry. He has been invaluable on both academic and personal fronts, for which I am grateful. I am extremely grateful to all my collaborators and coauthors. Without their inputs, ideas and support, this dissertation would not have been possible. I would also like to acknowledge the University of Rhode Island, College of Pharmacy for their assistance with various facilities.

I would like to extend sincere thanks to each of my doctoral committee members, Dr. Roberta King, Dr. Brenton DeBoef, Dr. Keykavous Parang and Dr. Mindy Levine.

A large part of my success and strength as a graduate student and researcher comes from my family. It is to them that I owe my deepest gratitude. A sincere thanks to my dad

Narendra Jain and mom **Madhuri Jain** for their profound belief in me and my endeavors. A special mention to my elder brother Nitin, sister-in-law Ritu, sister Shalu and brother-in-law Nikhil. It is their love, unselfish care and immense sacrifices that have been a catalyst in my journey till here. This effort would have been incomplete without the mention of the incomparable love, support and patience of my wife, **Anshu Jain**.

I would also like to express my gratitude to all my past and present lab members Dr. Nidhi Jain, Dr. Satyakam Patnaik, Dr. Vaidyanathan Ganesan, Varsha, Anusha, Lifang, Sathyaraj who have been supportive throughout my research. I would also like to thank all my friends especially Saumitra, Shalini, Abhishek, Shripad, Chinmay, Swapnil Malekar, Swapnil Dalavi, Rajesh, Rahul who have never lost faith in me and cheered me up in difficult times.

Without the inspiration of these people and others who have helped me throughout the past five years, I would not have been where I am.

PREFACE

This dissertation was prepared according to the University of Rhode Island “Guidelines for the Format of Theses and Dissertations” standards for Manuscript format. This dissertation comprises of four manuscripts that have been combined to satisfy the requirements of the department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island.

MANUSCRIPT-I: Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the *NarI* sequence

This manuscript has been published in ‘Nucleic Acids Research’ May 2012.

MANUSCRIPT-II: Nucleotide excision repair of cluster acetylaminofluorene adducts on the *NarI* sequence: sequence dependence and structural factors

This manuscript has been prepared for submission to ‘Chemical Research in Toxicology’ for Publication.

MANUSCRIPT-III: Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor

This manuscript has been published in ‘Nucleic Acids Research’ January 2013.

MANUSCRIPT-IV: Investigation of the 3'-next flanking base effect on the translesion synthesis across arylamine-DNA adducts: A structural perspective

This manuscript has been prepared for submission to ‘Chemical Research in Toxicology’ for Publication.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS.....	x
PREFACE.....	xii
TABLE OF CONTENTS.....	xiii
LIST OF FIGURES.....	xiv
LIST OF TABLES.....	xix
MANUSCRIPT I	1
MANUSCRIPT II.....	50
MANUSCRIPT III.....	96
MANUSCRIPT IV.....	142

LIST OF FIGURES

MANUSCRIPT-I

Figure 1: (a) Chemical structures of AAF, FAAF, AF and FAF adducts; (b) sequences of fully paired 16-mer and 12-mer <i>NarI</i> duplexes used in the present study; (c) the B/S and (d) B/S/W-conformer equilibrium of FAF and FAAF-modified duplexes.....	42
Figure 2: (a) Chromatogram of a reaction mixture between 16-mer <i>NarI</i> sequence (5'-CTCTCG ₁ G ₂ CG ₃ CCATCAC-3') and an activated FAAF (<i>N</i> -acetoxy- <i>N</i> -2-(acetylamino)-7-fluorofluorene); (b) on-line photodiode array UV/Vis spectra of mono-, di- and tri-FAAF adducts.....	43
Figure 3: (a) CD spectral overlays recorded at 15 °C and (b) differential scanning calorimetry (DSC) curves of fully paired 16-mer <i>NarI</i> duplexes.....	44
Figure 4: ¹⁹ F NMR spectra of (a) FAAF-modified <i>NarI</i> 16-mer, b) FAAF-modified non- <i>NarI</i> 12-mer, and (c) FAF-modified <i>NarI</i> 12-mer duplexes at 5 °C.....	45
Figure 5: Dynamic ¹⁹ F NMR spectra of fully paired 16-mer <i>NarI</i> duplexes. FAAF modification at (a) G ₁ , (b) G ₂ , and (c) G ₃	46
Figure 6: Absolute percent incision rates of a) FAAF and b) FAF- <i>NarI</i> duplexes; c) percent incision rates histogram relative to <i>NarI</i> -G ₃	47

MANUSCRIPT-II

Figure 1: Adduct and sequence used in this study (a and b); (c) major groove views of the central trimer segments of the B/S/W-conformer equilibrium of FAAF-modified mono-adduct duplexes.....	82
Figure 2: HPLC profile of FAAF-modified ODN. (a) Chromatogram of a reaction mixture (b) on-line photodiode array UV/Vis spectra of mono-, di- and tri-FAAF adducts.....	83
Figure 3: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 4.....	84
Figure 4: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 4 (complete time range).....	85
Figure 5: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 5.....	86
Figure 6: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 6.....	87
Figure 7: CD and UV-melting profiles. (a) CD spectral overlay recorded at 20 °C and (b) UV-melting curves fully paired 16-mer <i>NarI</i> di-adduct duplexes.....	88
Figure 8: Comparative thermodynamic parameters histogram of <i>NarI</i> di-adduct duplexes.....	89
Figure 9: ^{19}F NMR spectra of FAAF-modified di-adduct duplexes.....	90
Figure 10: Comparison of ^{19}F NMR spectra of FAAF-modified di-adduct duplexes with their respective mono-adducts.....	91
Figure 11: Imino proton spectra of FAAF-modified di-adduct duplexes.....	92

Figure 12: Polyacrylamide gel showing the incision products of the 5'-terminally labeled FAAF modified <i>NarI</i> di-adduct substrates.....	93
Figure 13: UvrABC nuclease incision kinetic results.....	94
 MANUSCRIPT-III	
Figure 1: Adducts and sequences used in this study (a and b); (c) Major groove views of the B, S, and W-conformers of ABP, AF, and AAF.....	133
Figure 2: ¹⁹ F NMR (a, b, c) and CD (d, e, f) spectra of FABP-, FAF- and FAAF-modified 11-mer duplexes, respectively, in the G*CA and G*CT sequence contexts at 20°C.....	134
Figure 3: Autoradiograph of 15% (w/v) native polyacrylamide gel showing the relative mobility of single strand (ss) and GCA/GCT 19-mer double strand (ds); both unmodified and modified.....	135
Figure 4: Comparative thermodynamic parameters histogram of FABP (Black), FAF (Blue) and FAAF (Pink) in (a) G*CA duplexes and (b) G*CT duplexes.....	136
Figure 5: Free energy profiles from PMF calculations over the sampling range of 0.5 ns to 3 ns for modified and unmodified sequences.....	137
Figure 6: Absolute incision rates (a, b) of modified 55-mer substrates in the G*CA and G*CT sequence contexts. (c) Relative percent incision rate histograms with respect to G*CA-FAAF duplex.....	138

Figure 7: Image of the a) G*CT-FAF adduct in the B state (350°)	
and b) G*CT-FAAF adduct in the B state (330°).....	139

MANUSCRIPT-IV

Figure 1: (a) Structures of ABP, AF and AAF (b) Major groove views of the B and S conformers of AF (c) 11-mer duplexes used in the present study (d) Primer-template duplexes used in simulated translesion synthesis.....	175

Figure 2: CD (a) and ¹⁹ F NMR (b) spectra of FABP-modified 11-mer duplexes.....	176

Figure 3: ¹⁹ F NMR (a) and imino proton (b) spectra of FABP-modified 11-mer duplexes at 20 °C.....	177

Figure 4: Dynamic ¹⁹ F NMR spectra of FABP-modified 19-mer (a) G*CT and (b) G*CA duplexes at different primer positions i.e. n-1, n, n+3 and full.....	178

Figure 5: Dynamic ¹⁹ F NMR spectra of FAF-modified 19-mer (a) G*CT and (b) G*CA duplexes at different primer positions i.e. n-1, n, n+3 and full.....	179

Figure 6: CD spectra of (a) FABP- and (b) FAF-modified 19-mer G*CT (right) and G*CA (left) duplexes at different primer positions i.e. n-1, n, n+3 and full.....	180

Figure 7: Running start experiments across (a) FABP- and (b) FAF-modified templates (G*CA and G*CT) in presence	

of all four dNTPs (250 μ M each) and Kf-exo ⁻ (2.5 nM) at different time intervals (M, marker).....	181
Figure 8: Running start experiments across (a) FABP- and (b) FAF-modified templates (G*CA and G*CT) in presence of all four dNTPs (250 μ M each) and Dpo4 (50 nM, FABP; 100 nM, FAF) at different time intervals (M, marker).....	182
Figure 9: Oligonucleotide sequences used for steady state kinetics.....	183
Figure 10: Plots showing the extension frequency vs. insertion site: (a) FABP, (b) FAF in match (left) and mismatch (right) series.....	184

LIST OF TABLES

MANUSCRIPT-I

Table 1: Thermal and thermodynamic parameters of FAAF modified <i>NarI</i> duplexes obtained from Differential Scanning Calorimetry.....	48
Table 2: Conformational heterogeneity (B/S/W), thermal destabilization and relative percent incision rates of FAAF- and FAF-modified <i>NarI</i> duplexes.....	49

MANUSCRIPT-II

Table 1: Thermal and thermodynamic parameters of FAAF modified <i>NarI</i> di-adduct duplexes obtained from UV-melting experiments.....	95

MANUSCRIPT-III

Table 1: Lesion-induced CD blue shifts.....	140
Table 2: Thermal and thermodynamic parameters derived from differential scanning calorimetry (DSC) curves.....	141

MANUSCRIPT-IV

Table 1: Summary of the conformational profile exhibited by FABP and FAF at different primer positions in GCA and GCT sequence contexts.....	185
Table 2: Kinetic parameters for dGTP insertion opposite the	

templating C one nucleotide before the FABP lesion site (n-1).....	186
Table 3: Kinetic parameters for nucleotide insertion opposite	
the FABP lesion (n).....	187
Table 4: Kinetic parameters for correct dGTP insertion opposite	
the templating C one nucleotide after the FABP lesion site (n+1).....	188
Table 5: Kinetic parameters for correct dTTP insertion opposite	
templating A at (n+3) position from the FABP lesion site.....	189
Table 6: Kinetic parameters for dGTP insertion opposite the	
templating C one nucleotide before the FAF lesion site (n-1).....	190
Table 7: Kinetic parameters for insertion opposite the FAF lesion (n).....	191
Table 8: Kinetic parameters for correct dGTP insertion opposite	
the templating C one nucleotide after the FAF lesion site (n+1).....	192
Table 9: Kinetics parameters for correct dTTP insertion opposite	
templating A at (n+3) position from the FAF lesion site.....	193
Table 10: Summary of the insertion efficiency against	
FABP- and FAF-modified template at different primer	
positions in GCA and GCT sequence contexts.....	194

Manuscript-I

Published in *Nucleic Acids Research*, 2012, 40, 3939-3951

Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the *NarI* sequence

Vipin Jain,¹ Benjamin Hilton,² Satyakam Patnaik,¹ Yue Zou,² M. Paul Chiarelli,³ and
Bongsup P. Cho*¹

¹Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island,
Kingston, RI, 02881, USA

² Dept. of Biochemistry and Molecular Biology, East Tennessee State University,
Johnson city, TN, 37614, USA

³Dept. of Chemistry, Loyola University, Chicago, IL 60626, USA

*To whom correspondence should be addressed: Telephone#: (+1) 401-874-5024; FAX#
(+1) 401-874-5766; E-mail: bcho@uri.edu

Present address: Department of Biomedical and Pharmaceutical Sciences, University of
Rhode Island, 7 Greenhouse Road, Kingston, RI, 02881, USA

Running Title: conformation-specific repair of aminofluorene adducts in the *NarI*
sequences

Abstract

Nucleotide excision repair (NER) is a major repair pathway that recognizes and corrects various lesions in cellular DNA. We hypothesize that damage recognition is an initial step in NER that senses conformational anomalies in the DNA caused by lesions. We prepared three DNA duplexes containing the carcinogen adduct *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene (FAAF) at G₁, G₂ or G₃ of *NarI* sequence (5'-CCG₁G₂CG₃CC-3'). Our ¹⁹F NMR/ICD results showed that FAAF at G₁ and G₃ prefer syn S- and W-conformers, whereas anti B-conformer was predominant for G₂. We found that the repair of FAAF occurs in a conformation-specific manner, *i.e.*, the highly S/W-conformeric G₃ and -G₁ duplexes incised more efficiently than the B-type G₂ duplex (G₃~G₁ > G₂). The melting and thermodynamic data indicate that the S- and W-conformers produce greater DNA distortion and thermodynamic destabilization. The *N*-deacetylated *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene (FAF) adducts in the same *NarI* sequence are repaired 2-3 fold less than FAAF: however, the incision efficiency was in order of G₂~G₁ > G₃, a reverse trend of the FAAF case. We have envisioned the so-called “*N*-acetyl factor” as it could raise conformational barriers of FAAF vs. FAF. The present results provide valuable conformational insight into the sequence dependent UvrABC incisions of the bulky aminofluorene DNA adducts.

Key Words: DNA adducts; conformational heterogeneity; sequence effect; ¹⁹F NMR; thermodynamics.

Abbreviations: AF, *N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene; AAF, *N*-(2'-deoxyguanosin-8-yl)-2-acetylamino-2-fluorene; FAAF, *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylamino-2-fluorene; FAF, *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene; DSC, differential scanning calorimetry; ¹⁹F NMR, fluorine nuclear magnetic resonance; ICD_{290-320 nm}, induced circular dichroism at 290-320 nm.

Introduction

Adduct formation is an important aspect of DNA damage: if unrepaired, various mutations in DNA could occur (1-3). The presence of mutations on specific oncogenes or tumor suppressor genes may trigger cancer initiation. Human cells are armed with various effective repair pathways to safeguard genomic DNA from continuous assault by exogenous and endogenous sources (4). Nucleotide excision repair (NER) is a major repair pathway that is known for the removal of stretches of bases containing various lesions, including single-base damages, bulky adducts, and cross-links, among others (5). Deficiencies in NER are closely associated with the development of several genetic diseases, such as xeroderma pigmentosum that increases the risk of skin cancer due to higher sensitivity to sunlight (6).

The NER pathway in *Escherichia coli* involves the UvrABC nuclease system and has been studied extensively for understanding DNA damage recognition and incision. *E. coli* NER is initiated following damage recognition by a dimeric UvrA protein. Next, UvrB protein reaches the damage site, forms a trimer, and verifies the damage. Departure of UvrA from the resulting complex recruits UvrC and UvrD proteins, which cleave and remove the lesion-bearing patch of DNA. Finally, DNA polymerase I synthesizes and ligase I seals a new patch to complete the repair process (5,7).

Arylamines are an important class of environmental pollutants that are implicated in the etiology of human cancers, especially of the bladder and liver (1). 2-Acetylaminofluorene was originally developed as an agricultural insecticide, but was later banned due to its

strong tumorigenic activity in rat livers (8). It has been used extensively as a model for studying chemical carcinogenesis. *In vivo*, metabolic activation of AAF produces a highly electrophilic nitrenium ion, which subsequently interacts with DNA to produce two major C8-substituted dG adducts: AAF and AF (Figure 1a)(8,9). *In vitro*, *N*-acetylated AAF blocks the activity of high-fidelity polymerases and require bypass polymerases for a translesion synthesis (TLS), whereas AF only slows down replication (10). In general, the bulky AAF exhibits greater susceptibility towards NER than AF(11), which is known to exist in a sequence-dependent equilibrium between *anti* B-conformer and *syn* S-conformer (Figure 1c)(10,12-14). We recently reported that AAF adducts also adopt a sequence-dependent S/B/W-conformational equilibrium (Figure 1d)(15).

Local sequence context plays an important role in the repair of arylamine-DNA adducts (16). Fuchs *et al.* constructed DNA sequences modified with AAF at each of three guanines of the most frequently studied mutational hotspot known as *NarI* sequence (5'-...CG₁G₂CG₃CC...-3') and tested their substrate reparability in the *E. coli* UvrABC and human exonuclease systems. In *E. coli*, the three AAFs were repaired in a sequence-dependent manner, with relative repair efficiencies of G₁:G₂:G₃ in a ratio of 100:18:66 (17,18). However, different repair efficiencies were observed for the same lesions by the human exonuclease, 38:100:68 for G₁, G₂, and G₃, respectively (18). AAF at G₃ of *NarI* sequence induces ~100-fold greater frequency of -2 frameshift (-2 deletion) mutations, even though the three guanines exhibit similar chemical reactivities (19). We have shown that the FAF-modified *NarI* -2 deletion duplex in the 5'--CG₁G₂CG₃*CC-3' context adopts a single looped-out bulge structure, whereas the 5'--CG₁G₂CG₃*CT--3' context

results in a local conformational heterogeneity (20). These results support the importance of the 3'-next flanking nucleotide to the lesion in modulation of mutation efficiency. The studies verified that the conformational stability of a slipped mutagenic intermediate is a critical determinant for the hotness (up to 30 to 50-fold) of G₃ in *NarI* sequence for -2 frameshift mutation (20-22). Mekovich *et al.* found a greater incision rate in *E. coli* systems when AAF was located at G₃ of the *NarI* sequence (5'---CG₁G₂CG₃*CC---3') than in a non-*NarI* sequence (5'---GATG*ATA---3')(23). Zou *et al.* have reported that the UvrABC incision efficiency is 70% more in the TG*T than in the CG*C sequence context when adducted with either AF- or AAF lesions (24).

The NER pathway is characterized by its unique ability to excise a wide array of structurally diverse DNA lesions. The structure of individual adducts *per se* is not as important as lesion-induced local distortions and destabilizations to trigger a NER response. Examples include disruption of Watson-Crick hydrogen bonding, DNA bending, thermodynamic destabilization, local conformational flexibility, and flipped-out bases in the unmodified complementary strand (25,26). However, mechanisms of sequence dependence that control NER efficiencies remained elusive. It could be that the observed local sequence effects described in the previous paragraph for AF and AAF are due to differences in the extent of distortions, which in turn depends on the conformation adopted in a particular sequence context.

In the present work, we investigated the role of conformational heterogeneity in the structure-repair relationships of AF and AAF. These two adducts are structurally similar,

but differ in the absence and presence, respectively, of an *N*-acetyl group on the central nitrogen. We prepared oligonucleotides that were site-specifically modified by the fluorine model FAF and FAAF at three different guanines (G₁, G₂, and G₃) of the *NarI* recognition sequences (Figure 1a,b). We conducted spectroscopic and melting experiments for conformational and thermodynamic analyses. Moreover, we performed NER studies of these adducts using the *E. coli* UvrABC system. The results present strong structural and thermodynamic evidences for the differential NER efficiencies exhibited by AF and AAF at different guanine residues of the *NarI* sequence.

Materials and Methods

Caution: *2-Aminofluorene derivatives are mutagens and suspected human carcinogens and therefore must be handled with caution.*

Crude oligodeoxynucleotides (ODN, 10 μ mol scale) in desalted form were purchased from Eurofins MWG operon (Huntsville, AL). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Preparation and Characterization of FAF- and FAAF-modified ODNs

We previously reported the preparation of 12-mer *NarI* ODNs (5'-CTCG₁G₂CG₃CCATC-3'), in which each of the three guanines were site-specifically modified by FAF (20). We have also demonstrated that the incorporation of fluorine atom at the longest axis position 7 does not affect the overall conformational and thermal/thermodynamic profiles of AF- or AAF-modified duplexes (14,20). The three FAF-modified *NarI* sequences were each

annealed with a complementary 12-mer sequence (5'-GATGGCGCCGAG-3') to form fully paired *NarI*-G₁-FAF, *NarI*-G₂-FAF, and *NarI*-G₃-FAF duplexes, respectively. These duplexes were thoroughly characterized by ¹⁹F NMR, CD, and UV melting experiments (20).

FAAF-modified 16-mer ODN were prepared using the general procedures described previously (15,27). Briefly, approximately 0.5-1 mg of *N*-acetoxy-*N*-2-(acetylamino)-7-fluorofluorene dissolved in absolute ethanol was added drop wise to a sodium citrate buffer (pH 6.0) containing 200-250 ODs of unmodified ODN (5'-CTCTCG₁G₂CG₃CCATCAC-3') and placed in a shaker for 5 min at 37 °C. Figure 2a shows a typical reversed-phase HPLC chromatogram derived from the resulting mixture. The FAAF modified oligomers appearing between 28 and 85 min were separated and purified up to >97% purity by repeated injections. The HPLC system consisted of a Hitachi EZChrom Elite HPLC unit with an L2450 diode array detector and a Phenomenex Luna C18 column (150 X 10 mm, 5.0 μm). We employed a gradient system involving 3-15% acetonitrile for 40 min followed by 15-20% and 20-35% acetonitrile for 20 min and 40 min, respectively, in pH 7.0 ammonium acetate buffer (100 mM) with a flow rate of 2.0 mL/min.

The three FAAF-modified 16-mer sequences were each annealed with the complementary sequence (5'-GTGATGGCGCCGAGAG-3') to form fully paired *NarI*-G₁-FAAF, *NarI*-G₂-FAAF, and *NarI*-G₃-FAAF duplexes for UV melting, DSC, CD and dynamic ¹⁹F NMR experiments.

LC/MS Characterization of FAAF-modified ODNs

Electrospray ionization and quadrupole time-of-flight mass spectrometry was utilized to verify the molecular weights and the position of FAAF attachment of the three oligomers. The 16-mer ODNs were sequenced using 3'-5' or 5'-3' exonucleases as described previously for the analysis of modified 12-mers (28). Normally, one microgram of a particular ODN was combined with 0.01 units of an exonuclease in a 1 mM solution of MgCl_2 and incubated for several hours. The digests were separated using a Phenomenex Aqua C18, 1.0×50 mm column (5 μm ; 120 Å). Solvent A was 5 mM in both ammonium acetate and dimethylbutyl amine. Acetic acid was added to solvent A to adjust the pH to 7.0. Solvent B was 0.1% formic acid in acetonitrile. The flow rate was 100 $\mu\text{L}/\text{min}$ and total run time was 20 min. All LC/MS spectra were acquired using a Waters SYNAPT quadrupole time-of-flight mass spectrometer (Milford, MA) operated in the negative ion and V-modes. The measured molecular masses of all three isomeric ODNs were within 0.1 daltons of their theoretical monoisotopic mass (5016.9 Daltons).

UV Melting

UV melting data were obtained using a Cary100 Bio UV/VIS spectrophotometer equipped with a 6×6 multi-cell block and 1.0 cm path length. Sample cell temperatures were controlled by an in-built Peltier temperature controller. Oligonucleotide duplexes with a concentration range of 0.4-6.4 μM were prepared in solutions containing 0.2 M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA at pH 7.0. Thermomelting curves were constructed by varying temperature of the sample cell (1 $^{\circ}\text{C}/\text{min}$) and monitoring absorbance at 260 nm. A typical melting experiment consisted of forward/reverse scans

and was repeated five times. Thermodynamic parameters were calculated using the program MELTWIN version 3.5 as described previously (12).

Circular Dichroism (CD)

CD measurements were conducted on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. Typically, 2 ODs of each strand were annealed with an equimolar amount of a complementary sequence. The samples were dissolved in 400 μ L of a neutral buffer (0.2 M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA) and placed in a 1.0 mm path length cell. The samples were heated at 85 °C for 5 min and then cooled to 15 °C, over a 10 min period to ensure complete duplex formation. Spectra were acquired every 0.2 nm with a 2 s response time from 200 to 400 nm at a rate of 50 nm/min, were the averages of 10 accumulations, and were smoothed using 17-point adaptive smoothing algorithms provided by Jasco.

Dynamic ^{19}F NMR:

Approximately 20 ODs of a pure FAAF-modified 16-mer ODN was annealed with an equimolar amount of a complementary sequence to produce a fully paired duplex (Figure 1b). The samples were then dissolved in 300 μ L of typical pH 7.0 NMR buffer containing 10% D₂O/90% H₂O, 100 mM NaCl, 10 mM sodium phosphate, and 100 μ M EDTA, and filtered into a Shigemi tube through a 0.2 μ m membrane filter. All ^1H and ^{19}F NMR results were recorded using a dedicated 5 mm $^{19}\text{F}/^1\text{H}$ dual probe on a Bruker DPX400 Avance spectrometer operating at 400.0 and 376.5 MHz, respectively, using acquisition parameters described previously (14,20,29). Imino proton spectra at 5°C were obtained

using a phase-sensitive jump-return sequence and referenced relative to that of DSS. ^{19}F NMR spectra were acquired in the ^1H -decoupled mode and referenced relative to that of CFCl_3 by assigning external C_6F_6 in C_6D_6 at -164.9 ppm. One and two-dimensional ^{19}F NMR spectra were measured between 5 and 60 $^\circ\text{C}$ with an increment of 5-10 $^\circ\text{C}$. Temperatures were maintained by a Bruker-VT unit with the aid of controlled boiling liquid N_2 in the probe. Line shape simulations were performed as described previously (30) using WINDNMR-Pro (version 7.1.6; *J. Chem. Educ. Software Series*; Reich, H. J., University of Wisconsin, Madison, WI).

DSC Experiments

Calorimetric measurements of the three FAAF-modified 16-mer duplexes were performed using a Nano-DSC from TA Instruments (Lindon, Utah). Prior to temperature scanning, samples were degassed for at least 10 min under house vacuum in a closed vessel. Solutions were loaded respectively into the sample and reference cells using a pipette by attaching a small piece of silicone tube at the end of the tip and were purged several times to get rid of air bubbles. After both cells were filled, they were capped and a slight external pressure (~3 atm) was applied to prevent evaporation of the sample solution. Raw data were collected as microwatts vs. temperature. Template-primer solutions were prepared by dissolving desalted samples in a pH 7.0 buffer solution consisting of 20 mM sodium phosphate, and 0.1 M NaCl. In a typical scan, a 0.1 mM template-primer solution was scanned against buffer from 15 to 90 $^\circ\text{C}$ at a rate of 0.75 $^\circ\text{C}/\text{min}$. At least 5 repetitions were obtained. A buffer vs. buffer scan was used as a control and subtracted from the sample scan and normalized for heating rate. This results

in base-corrected ΔC_p^{ex} versus temperature curves. Each transition shows negligible changes in the heat capacities between the initial and final states, thus $\Delta\Delta C_p^{\text{ex}}$ was assumed to be zero. The area of the resulting curve is proportional to the transition heat, which, when normalized for the number of moles of the sample, is equal to the transition enthalpy, ΔH . ΔH is an integration of ΔC_p^{ex} over temperature T . All sample solutions were 0.1 mM concentration. T_m was the temperature at half the peak area. ΔG and ΔS values have been determined according to the procedures described by Chakrabarti *et al* (31).

Substrate Construction and UvrABC Protein Purification

DNA substrates of 55 bp containing a FAAF adduct at each of three guanine residues were constructed as previously described (11,32). Briefly, an FAAF-modified 16-mer ODN (5'-CTCTCG₁G₂CG₃CCATCAC-3') was ligated with flanking 20-mer ODN (5'-GACTACGTACTGTTACGGCT-3') and 19-mer ODN (5'-GCAATCAGGCCAGATCTGC-3') ODN at the 5'- and 3'-end, respectively (Supplementary Figure S1). The 20-mer was 5'-terminally labeled with ³²P. The ligation product was purified by urea-PAGE under denaturing conditions. Following the purification, the substrate was annealed to the corresponding complementary strand, and then purified on an 8% native polyacrylamide gel. Similar procedures were employed to construct FAF-modified DNA substrates using 12-mer *NarI* sequence (5'-CTCG₁G₂CG₃CCATC-3'), which we prepared previously for ¹⁹F NMR/UV/ICD studies (20).

UvrA, UvrB, and UvrC proteins were over expressed in *E. coli* and then purified as previously described (33). The estimated purity of the three proteins was greater than 95%. A Bio-Rad Protein Assay was used to determine the protein concentration with BSA as the standard based on the manufacturer-recommended procedures.

Nucleotide Excision Assay and Quantification of Incision Products

The 5'-terminally labeled DNA substrates were incised by UvrABC as previously described (11,32). Briefly, the DNA substrates (2 nM) were incubated in the UvrABC reaction buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM MgCl₂, 5 mM DTT) at 37°C in the presence of UvrABC (UvrA, 10 nM, UvrB, 250 nM, and UvrC, 100 nM). The Uvr proteins were diluted and premixed in Uvr storage buffer before addition to the reaction. Aliquots were collected at 0, 5, 10, 15, and 20 minutes into the reaction. The reaction was terminated by heating at 95°C for five min. The products were denatured by addition of formamide loading buffer and heating to 95°C for five min, followed by quick chilling on ice. The incision products were then analyzed by electrophoresis on a 12% polyacrylamide sequencing gel under denaturing conditions with TBE buffer.

To quantify the incision products, radioactivity was measured using a Fuji FLA-5000 Image Scanner with MultiGauge V3.0 software. The DNA incised (in fmol) by UvrABC was calculated based on the total molar amount of DNA used in each reaction and the ratio of the radioactivity of incision products to total radioactivity of DNA. At least three independent experiments were performed for determination of the rates of incision.

Results

Model Sequences

We previously used FAF-modified 12-mer duplexes (5'-CTCG₁G₂CG₃*CNATC-3', *N* = C or T) to probe the impact of flanking and 3'-next flanking sequences on *NarI*-induced frameshift mutagenesis (20). Initially, we tried to use the same 12-mer *NarI* sequence for FAAF-modification for sake of comparison and consistency; however, the sequence was unsuitable for FAAF. Although FAAF adduction on the 12-mer *NarI* sequence was facile, a resulting reaction mixture was difficult to purify on the reverse phase HPLC system (see asterisked peaks in Supplementary Figure S2). Moreover, ligation efficiencies of the FAAF-modified 12-mers, particularly on the G₁- and G₂-positions, were very low. Accordingly, the length of DNA was increased to 16 (5'-CTCTCGGCGCCATCAC-3', Figure 1b) by adding two nucleotides (underlined) on either side of the 12-mer. The resulting FAAF-modified 16-mer ODNs were separated well on a reverse-phase HPLC and exhibited excellent ligation efficiencies (see below).

Figure 2a shows an HPLC profile of a work-up mixture after 20 min of reaction. Unreacted control ODN appeared at 21 min, followed by seven FAAF-modified ODNs in three retention time zones: Peaks 1-3 at 28-35 min, Peaks 4-6 at 42-60 min, and Peak 7 at 84 min. On-line UV (Figure 2b) of the modified ODNs displayed a small shoulder in the 290-320 nm range. The relative absorption intensities (290-320 nm) for the three peak groups were approximately 1:2:3. This finding is reminiscent of AF- or FAF-induced absorption shoulders observed in 290-350 nm, whose intensities correlate consistently with the number of adduct modifications (28). As a result, Peaks 1-3 and 4-6 were

assigned as mono- and di-adducts, respectively, and Peak 7 as a tri-adduct. These adducts were characterized by exonuclease digestion/ESI-TOF-MS-MS analyses, as described below. Depending on the location of FAAF, the mono-adducts were designated as *NarI*-G₁, *NarI*-G₂, or *NarI*-G₃, in which G₁, G₂, and G₃ signify the position of the FAAF-modified guanine. Details of the structural characterization and repair of the di- and tri-FAAF adducts will be published separately.

The HPLC elution profile of the FAAF-modified 16-mer ODNs in the present study is similar to that of the AAF-modified 15-mer *NarI* sequence (5'-TCCTCG₁G₂CG₃CCTCTC-3') reported by Tan *et al* (34). These results indicate that AAF and FAAF are chromatographically comparable, irrespective of sequence length, as long as the common *NarI* core (underlined) is included in the sequences. This is not surprising since conformational and thermodynamic compatibilities of fluorine containing AF and AAF models have well been documented (20,30,35). A similar elution pattern was observed for the FAF-modified 12-mer and FAAF-modified 16-mer *NarI* sequences; however, the order of elution of G₁ and G₃ was reversed (compare Figure 2a with Figure 2 in reference 20).

ESI-QTOF-MS Characterization

The molecular weights of all three FAAF-modified ODNs were measured by ESI-QTOF-MS prior to sequence verification by exonuclease digestion. Ionization of ODNs normally occurs by the loss of a proton from a phosphate group in the ODN backbone. As the number of nucleotides in an ODN increases, the average charge state observed in

the full scan mass spectra increases as well (28). As shown in Supplementary Figure S3, the ODNs containing sixteen nucleotides form $(M-4H)^{4-}$ ions predominantly unlike the 12-mers studied previously that form $(M-3H)^{3-}$ primarily upon electrospray (28). Exonucleases cleave terminal deoxynucleotides from the ODN chain until the FAAF-modified nucleotide is exposed at the end of the chain. At that point the digestion reaction slows down significantly. The position of modification is identified (in this case) when the fragment(s) formed by the loss of the unmodified guanine nucleotides is observed in the LC/MS spectra. This is shown in Figure S4 for the 3' digest of the -G₁(FAAF)G₂CG₃- ODN. The ions observed at m/z 659.5 and m/z 989.7 are the $(M-3H)^{3-}$ and $(M-2H)^{2-}$ ions formed from the 5'-CTCTCG₁(FAAF)-3' ODN digest fragment. The observation of these ions confirms that this ODN is modified on the G closest to the 5' end. The LC/MS analysis of the 5' digest of the second modified ODN to elute is shown in Supplementary Figure S5a and 5b. The exonuclease digestion of this particular reaction product was particularly slow and evidence for endonuclease activity is observed in the mass spectra. All the Y fragments observed in Supplementary Figure S5 are $(M-3H)^{3-}$ ions. The Y10 and Y9 fragments at m/z 1069.2 (5'-G₂CG₃(FAAF)CCATCAC-3') and m/z 959.5 (5'-CG₃(FAAF)CCATCAC-3') are formed by consecutive cleavages of unmodified guanines, confirming that this reaction product is modified on the guanine closest to the 3' end. LC/MS analysis of the exonuclease digests derived from the third singly-modified ODN to elute indicates that the FAAF group is attached to the central G in the sequence, -G₁G₂(FAAF)CG₃-. LC/MS analysis of the 3' digest (Supplementary Figure S6) show ions at m/z 1154.7 and m/z 769.5 corresponding to the $(M-2H)^{2-}$ and $(M-3H)^{3-}$ ions derived from the 5' -CTCTCG₁G₂(FAAF)-3' digest fragment. The mass

spectra acquired from the 5' -digest (Supplementary Figure S7) show $(M-2H)^{2-}$ and $(M-3H)^{3-}$ ions at m/z 1604.3 and m/z 1069.2 whose masses are consistent with 5'-G₂(FAAF)CG₃CCATCAC-3' fragment. No ions formed by the loss of two guanine deoxynucleotides were observed in any of the mass spectra. The observation of ODN fragments with two G's in both 3' and 5' digests confirms that the last singly-modified ODN to elute from the reaction mixture is modified on the middle G.

Circular Dichroism

Figure 3a shows an overlay of the CD spectra for the three FAAF-modified *NarI*-G₁, -G₂ and -G₃ duplexes relative to the unmodified control (red). Unmodified and FAAF-adducted duplexes both displayed a positive and negative ellipticity at around 270 and 250 nm, respectively, which is an S-curve characteristic of a B-form DNA double helix. The modified duplexes displayed significant blue shifts relative to the unmodified duplex, *NarI*-G₃ (6 nm) >> *NarI* G₁ ~ G₂ (3 nm), indicating adduct-induced DNA bending. A concomitant increase in the positive intensity around 270 nm was noted in the order of G₃ ~ G₁ > G₂, which could be due to the interaction of the intercalated S-conformeric FAAF with neighboring bases. We noted a similar blue shift and hyperchromic effect for the highly (75%) S-conformeric FAF-modified *NarI*-G₃ duplex (Figure 3b and Table 1 in Reference 20).

More importantly, FAAF-modified duplexes exhibited sequence-dependent induced CD in the 290-320 nm range (ICD_{290-320 nm}). This finding is reminiscent of ICD_{290-350 nm}, which has been used as a sensitive marker for the FAF-induced S/B/W-conformational

heterogeneity (positive for S- and W- and negative for B-conformer)(12,15,20,29,36). In the present case, however, the FAAF-modified *NarI* duplexes exhibited negative dips, with the *NarI*-G₂ duplex showing a greater dip than G₁ or G₃ duplexes (Figure 3a). This result could be due to a higher ratio (57%) of B-conformer for the *NarI*-G₂ duplex.

UV Melting Experiments

Supplementary Figure S8 shows the UV melting profiles of the three FAAF-*NarI* duplexes and an unmodified control duplex, all at 6.4 μ M. All duplexes showed typical monophasic, sigmoidal, helix-coil transitions with a strong linear correlation ($R^2 > 0.9$) between T_m^{-1} and $\ln C_t$. Thermal and thermodynamic parameters calculated from UV-melting are summarized in Supplementary Table S1. As expected, modified duplexes were destabilized thermally and thermodynamically relative to the control duplex. The magnitude of thermal (ΔT_m) and thermodynamic ($\Delta\Delta G$) destabilization was in the order of *NarI*-G₂ $\sim$ *NarI*-G₃ (-8.7 to -8.8 $^{\circ}$ C, 3.3 to 3.7 kcal/mol, respectively) $>$ *NarI*-G₁ (-4.6 $^{\circ}$ C and 2.0 kcal/mol, respectively).

Differential Scanning Calorimetry (DSC)

Figure 3b shows differential scanning calorimetry (DSC) plots of excess heat capacity C_p^{ex} vs. temperature for the FAAF-*NarI* duplexes relative to the unmodified control. Table 1 summarizes the thermal and thermodynamic parameters derived from these DSC curves. Consistent with the UV-melting data, the *NarI*-G₃ duplex was most destabilized ($\Delta\Delta H = 24.7$ kcal/mol, $\Delta\Delta G_{37^{\circ}\text{C}} = 4.7$ kcal/mol, $\Delta T_m = -8.3$ $^{\circ}$ C), followed by *NarI*-G₂

($\Delta\Delta G_{37^\circ\text{C}} = 4.1$ kcal/mol, $\Delta T_m = -7.9$ °C). *NarI*-G₁ was the least affected ($\Delta\Delta G_{37^\circ\text{C}} = 3.7$ kcal/mol, $\Delta T_m = -5.3$ °C).

Differences in the thermal and thermodynamic destabilizations must have arisen from the differences in the S/B/W-conformational characteristics. The most S-conformeric (61%) *NarI*-G₃ duplex causes disturbance of Watson-Crick base pairing, resulting in enthalpy reduction ($\Delta\Delta H = 24.7$ Kcal/mol)(Table 2). However, the large entropy ($\Delta\Delta S = 64.7$ eu) compensates for the enthalpy, thus resulting in the overall free energy loss of $\Delta\Delta G_{37^\circ\text{C}} = 4.7$ kcal/mol (37). In contrast, *NarI*-G₁ and -G₂ duplexes possess higher populations of B-conformer (46% and 57%, respectively), thus exhibiting lower differences in the enthalpy values ($\Delta\Delta H = 21.9$ kcal/mol and $\Delta\Delta H = 18.6$ kcal/mol, respectively)(Table 2). The relatively small enthalpy differences observed for the G₁ and G₂ duplexes could be attributed to the presence of S- and W-conformers in addition to B-conformer. As expected, entropy compensation was less in these two duplexes (*NarI*-G₁, $\Delta\Delta S = 58.9$ eu and *NarI*-G₂, $\Delta\Delta S = 47.0$ eu), yielding similar overall free energies (*NarI*-G₁, $\Delta\Delta G = 3.7$ kcal/mol and *NarI*-G₂, $\Delta\Delta G = 4.1$ kcal/mol).

S/B/W Conformational Heterogeneity

Figure 4a shows the ¹⁹F NMR spectra of FAAF-*NarI* 16-mer G₁-, G₂-, and G₃-duplexes measured at 5 °C, in which ¹⁹F signals are in slow chemical exchange. These *NarI*-FAAF duplexes exhibited three to five ¹⁹F signals, each representing a particular conformation. The percent population ratios shown were calculated on the basis of line simulations as shown in Supplementary Figure S9. Assignments of the different ¹⁹F signals of each

duplex were necessary to carry out meaningful structure-activity-relationship studies. The initial assignments (as shown in Fig 4a) were made on the basis of chemical exchange, ring current effect, and chemical shift pattern recognition as have been done for a number of FAAF- and FAF-adducts in various sequence contexts (30, 35). It has been demonstrated that AF and AAF adducts adopt the S/B- and S/B/W-conformational equilibrium, respectively (Figure 1c,d) and their ^{19}F chemical shifts are independent of overall sequence and its length, but strongly rely on the nature of the bases flanking the lesion (15,30). The major ^{19}F signals in Figure 4a correlate well with the S/B/W-profiles reported previously for FAAF adducts (15), *i.e.*, B-, S-, and W-conformers going from downfield to upfield, *i.e.*, -115.0 to -115.5 ppm, -115.5 to -117.0 ppm and -117.0 to -118.0 ppm, respectively. Additional signals were observed in the -114.0 ~ -115.0 ppm range for the *NarI*-G₂ and -G₃ duplexes (Figure 4a, see Supplementary Table S2 for exact chemical shifts). Consistent with this observation, their proton spectra displayed a mixture of broad imino signals arising not only from those involved in Watson-Crick hydrogen bonds (12~14 ppm), but also from the lesion site and its vicinity (11~12 ppm) (Supplementary Figure S10).

To complement the ^{19}F signal assignments, we additionally conducted a set of comparative spectral analyses using three FAAF-12-mer duplexes in the non-*NarI* sequences (Figure 4b) in otherwise identical flanking sequence contexts (CG*G, GG*C, and CG*C context for G₁, G₂, and G₃, respectively). The top trace in Figure 4b is the ^{19}F NMR spectrum of a FAAF-modified 12-mer duplex (5'-CTTCTCG*CCCTC-3'), whose S/B/W conformational profiles have been well characterized (15). It should be noted that

this non-*NarI* 12-mer duplex contains the identical CG*C flanking sequence context as the 16-mer *NarI*-G₃-FAAF duplex. Comparison of the two spectra (i.e., top traces of Figure 4a and 4b) revealed a parallel trend both in terms of chemical shifts and population ratios (Table 2 and Supplementary Table S2), supporting the conformational assignments. This is consistent with our previous findings that the electronic environment for the ¹⁹F signals of AF and AAF adducts are strongly modulated by the nature of flanking bases (15,30). Similarly, we prepared two additional FAAF-modified non-*NarI* 12-mer duplexes (5'-CTTCTCG*GCCTC-3' and 5'-CTTCTCGG*CCTC-3') with the same flanking base contexts (underlined) as the *NarI*-G₁ and -G₂ duplexes, respectively. Figure 4a, b compares the ¹⁹F NMR spectra of all three *NarI* 16-mer and non-*NarI* 12-mer duplexes side by side. The ¹⁹F signal profiles, as indicated by dotted lines (pink, B; red, S; green, W) for G₁ → G₂ → G₃ of each sequence context, match quite well overall despite of slight variations observed in chemical shifts and population ratios, particularly for the GG*C sequence context. Whereas the S- and W-conformer signals were prone to shift, the B-conformer signal appeared to be steady at -115.5 ppm. This trend is more apparent in Supplemental Figure S11, in which the two FAAF-modified sequence series (*NarI*-16-mer vs. non-*NarI*-12-mer) are compared in a pair for each -CG*G-, -GG*C-, and -CG*C- sequence contexts. It is plausible that the carcinogen moiety in the major groove of the B-conformer is not subjected to the ring current effect, as the S- and W-conformers would be (14). We were unable to identify the minor signals (asterisked) in the 16-mer *NarI*-G₂- (<19%) and -G₃-FAAF duplexes, although their downfield shifts relative to the B-conformer imply B-like conformers, in which the fluorine containing carcinogen moiety is exposed.

Figure 4c shows the ^{19}F NMR spectra of FAF-modified 12-mer duplexes with the same *NarI* sequence contexts, which have been thoroughly characterized (20). The B/S conformer population ratios were determined to be 42%:58%, 69%:31%, 35%:65% for FAF-modified *NarI*-G₁, G₂ and G₃, respectively, at 5 °C (Table 2)(20). Although the chemical shift difference (0.4 to 1.0 ppm) for the B and S conformer of FAAF (Figure 4a) is significantly smaller than that (~1.5 ppm) of the FAF counterparts, their overall S/B ratios appear to match (Figure 4a, c). The B/S/W population ratios for the FAAF-*NarI*-G₁, -G₂ and -G₃ 16-mer duplexes were 46:34:20, 57:15:9, and 13:61:26, respectively (Table 2). In both the FAAF- (Figure 4a) and FAF- (Figure 4c) *NarI* duplexes, the population of S-conformer decreased in the order of G₃ > G₁ > G₂ and that of the B-conformer decreased in the reverse order, G₂ > G₁ > G₃. This comparative analysis was based on the assumption that structurally similar FAAF would experience similar sequence effects on their conformational profiles as observed by FAF in different sequence contexts of the *NarI* sequence (20). As expected, the aminofluorene-induced B/S-heterogeneity is strongly dependent on the nature of the flanking sequences, regardless of whether the lesion has the bulky acetyl group on the central nitrogen linking the carcinogen and the modified guanine, thus validating our assumption.

Dynamic ^{19}F NMR

Figure 5 shows the ^{19}F NMR spectra of the three FAAF-*NarI* duplexes as a function of temperature (5-60 °C). Whereas the three ^{19}F signals in each duplex were in slow exchange at 5 °C, the two downfield B- and S-signals became exchange broadened, giving rise to coalescent signals at around 30, 40, and 25 °C for G₁, G₂, and G₃,

respectively. In all cases, the merged signals coalesced with the upfield W-signal at around 60 °C. All three *NarI* duplexes showed relatively strong off-diagonal contour peaks of the major signals in the exchange spectra (data not shown), confirming their chemical exchanges.

UvrABC Incisions of FAAF-adducts on *NarI* sequence

Figure 6 shows the kinetic assay results, in which 55-mer FAAF-modified DNA duplex substrates were incised by UvrABC nuclease. These substrates were radioactively labeled at the 5'-end of the adducted strand. The major incision products can be seen as 18-mer (*NarI*-G₁), 19-mer (*NarI*-G₂), or 21-mer (*NarI*-G₃) separated on a urea-PAGE gel under denaturing conditions (Supplementary Figure S12). The incision occurred at the 8th phosphate bond 5' to the modified nucleotide, which is consistent with the previously reported results of UvrABC incision (11,32).

Quantitative analysis of the incision indicated that the substrates were incised at different efficiencies, depending on where the damage site was located in the sequence (Figure 6). Specifically, the *N*-acetylated FAAF adducts at *NarI*-G₁ and *NarI*-G₃ displayed similar rates of incision, whereas *NarI*-G₂ had a much lower rate of incision, $G_3 (100\%) \geq G_1 (93\%) > G_2 (32\%)$ (Figure 6c, Table 2). For comparison, we also determined the UvrABC incision of FAF adducts in the same *NarI* sequence context. As shown in Figure 6b,c, the *N*-deacetylated FAF adducts in the same *NarI* sequence context were repaired 2 ~ 3 fold less than FAAF. Despite having similar B/S-conformer profiles (Figure 4 and Table 2) the incision efficiency of FAF adducts at the three different sites in the *NarI* sequence

followed the order of G_1 (44%) $\approx$ G_2 (43%) $>$ G_3 (25%), where the percentages were calculated relative to FAAF *NarI*- G_3 (which was the most efficiently incised).

Discussion

It is well known that DNA sequence is a major determining factor for repair outcomes of site-specifically modified bulky DNA lesions. In this study, we examined the conformational heterogeneity and thermodynamics of FAAF and FAF at three different guanine positions (G_1 , G_2 , and G_3) of the well-known *NarI* recognition sequence. Moreover, we obtained nucleotide excision repair (NER) data of these adducts using the *E. coli* UvrABC system. Table 2, which summarizes the S/B/W solution conformational heterogeneity, T_m , and NER efficiency results, presents strong evidence that the NER repair efficiencies of AAF and AF adducts in the *NarI* sequences are modulated by their conformational and thermodynamic properties.

FAAF-induced B/S/W-conformational heterogeneity

Our combined ^{19}F NMR/ICD results show that FAAF adduct in a well-known mutational hotspot *NarI* sequence exist in a mixture of B/S/W conformers with varying populations (Figure 4a, Table 2). A greater population of syn-glycosidic S-(61%) and W-(26%) conformers was observed in *NarI*- G_3 , in which the lesion is flanked with C on both 5' and 3' ends (-CG $_3$ *C-). This result is consistent with the preferred *syn*-conformation adopted by duplexes modified by AF (20), 2-amino-3-methylimidazo(4,5-f)quinolone (IQ)(38-40) and 2-amino-1-methyl-6-phenylimid-azo(4,5-b)pyridine (PHIP) (41), with the same -CG*C- contexts, either in *NarI* or non-*NarI* sequences (20,38,39,41). The

mostly *syn* *NarI*-G₃ duplex appeared to be distorted, bent, or possibly formed a B-Z junction, as evidenced by a significant blue shift and hyperchromic effect in CD (Figure 3a)(42). The latter was probably due to the π - π stacking interaction between the intercalated aminofluorene and flanking base pairs. On the other hand, the *NarI*-G₂ duplex (-G₁G₂*C-) exhibited largely the *anti* B-conformer (57%) along with S- (15%), W- (9%), and two unidentified minor conformers (~19%). In comparison to *NarI*-G₃, the *NarI*-G₂ duplex exhibited smaller blue shift and hyperchromic effect (Figure 3a), suggesting lesser disturbance of the double helical DNA structure.

These results indicate the heterogeneous nature of AAF in the *NarI* sequence and are consistent with a previous CD study that showed a major DNA distortion for AAF at G₃ adduct compared to G₁ and G₂(42). Similarly, Veaute *et al.* (43) conducted a DNase I footprint study on the *NarI* sequence and showed that AAF at the G₂ position inhibits DNase I digestion of DNA at up to 5 bases in the modified strand and 4 bases in the complementary strand. In contrast, inhibition at the G₁ and G₃ positions was extended to 8 and 6 bases, respectively in the modified strand.

The G₃ and G₂ duplexes are chemically isomeric, differing only in the direction of the G:C base pair at the 5'-position (e.g., C:G $\rightarrow$ G:C). Such a polarity swap is clearly responsible for the rather dramatic conformational shift from S- (61% to 15%) to B- (13% to 57%), and W-conformation (26% to 9%)(Table 2). A similar polarity switch at the 3'-end of the *NarI*-G₁ duplex resulted in varying degrees of conformational shift in S- (61% to 34%), B- (13% to 46%), and W-conformation (26% to 20%).

As expected, FAAF-modification at the three guanines of the 16-mer *NarI* sequence resulted in thermal ($\Delta T_m = -5.3 \sim -8.3$ °C) and thermodynamic ($\Delta\Delta G_{37^\circ C} = 3.7 \sim 4.7$ kcal/mol) destabilization relative to the unmodified control duplex (Table 1). The destabilizing effect of the FAAF modification was sequence-dependent and was related to the S/B/W-conformational profile. As summarized in Table 1 and 2, the highly S-conformeric *NarI*-G₃ duplex (61%) promoted lesion stacking and disrupted the lesion site Watson-Crick base pairs, resulting in thermal ($\Delta T_m = -8.3$ °C) and enthalpic destabilization ($\Delta\Delta H = 24.7$ kcal/mol). In contrast, the highly B-conformeric (57%) *NarI*-G₂ duplex exerted less enthalpy change ($\Delta\Delta H = 18.6$ kcal/mol). As expected, *NarI*-G₁ (46% B, 34% S) produced an intermediate change in enthalpy ($\Delta\Delta H = 21.9$ kcal/mol). In all cases, however, enthalpy-entropy compensation resulted in a small overall difference (~1 kcal/mol) in thermodynamic destabilization (Table 1). A similar case could occur if the W-conformer were considered as a thermodynamic destabilizer, G₃ (26%) > G₁ (20%) > G₂ (9%). We previously studied three fully paired FAF-modified 12-mer duplexes in the same *NarI* sequence contexts (20). The UV melting results showed that FAF modification destabilizes the duplexes ($\Delta T_m = -6.8$ to -9.4 °C, $\Delta\Delta G = 4.2$ to 4.6 kcal/mol) similarly. The highly S-conformeric (65%) *NarI*-G₃ duplex resulted in thermal destabilization ($\Delta T_m = -8.3$ °C), whereas the highly B-conformeric (69%) *NarI*-G₂ duplex exerted less destabilization ($\Delta T_m = -6.8$ °C)(Table 2)(20).

Conformation-specific Nucleotide Excision Repair

The *E. coli* UvrABC system displayed significant differences in repair of the FAAF adduct at each guanine position (G₁, G₂, and G₃) of the *NarI* sequence. The *NarI*-G₂

duplex showed considerably lower efficiency than *NarI*-G₁ or *NarI*-G₃, [G_3 (100%) $\geq$ G_1 (93%) $>$ G_2 (32%)] (Figure 6). It is clear from Table 2 that, these NER results are in good agreement with the order of the S-conformer population [G_3 (61%) $>$ G_1 (34%) $>$ G_2 (15%)], but are in exactly the reverse order of the population of B-conformer, [G_2 (57%) $>$ G_1 (46%) $>$ G_3 (13%)]. This data suggests that the S-conformation is recognized and incised by *E. coli* NER dominantly over the B-conformation. We reported previously a similar conformation-specific NER results on a series of FAF-modified duplexes (16,30).

The carcinogen in the highly S-conformer *NarI*-G₃ is base-displaced at the lesion site, thus resulting in a major disturbance in the DNA helical structure ($\Delta T_m = -8.3$ °C, $\Delta\Delta G_{37^\circ C} = 4.7$ kcal/mol)(Table 2). This finding is contrasted with the 57% B-conformer *NarI*-G₂ duplex, which maintains Watson-Crick base pairs at the lesion site ($\Delta T_m = -7.9$ °C, $\Delta\Delta G_{37^\circ C} = 4.1$ kcal/mol). Similar correlations could be made with either the W-conformer alone, G_3 (26%) $>$ G_1 (20%) $>$ G_2 (9%) or the *syn*-conformation (combined S and W), G_3 (87%) $>$ G_1 (54%) $>$ G_2 (24%).

For comparison, we also determined the UvrABC incisions of FAF-adducts in the same *NarI* sequence context. The two lesions revealed a similar B/S conformer heterogeneity in the *NarI* sequence context (Figure 4). Therefore, the expectation was that FAF would show a similar NER profile as FAAF, i.e., S-/W-conformer promotes NER over B-conformer. However, the NER results revealed that incision efficiency was in the order of $G_1 \approx G_2 > G_3$ (Figure 6, Table 2). At first, this result appears to be in line with the B-conformer population. It should be noted that FAF is repaired consistently 2- to 3-fold

less than FAAF (Figure 6c, Table 2). This result is a general trend reported in the literature, although much greater differences in incision efficiency between AF and AAF have been noted (11,44). As a result, the difference between G_1 and G_2 of FAF is not statistically significant ($P = 0.83$), but their difference with G_3 is significant ($P < 0.0001$).

The incision differences between FAAF and FAF seem to suggest that, in addition to the sequence-dependent adduct conformation, the acetyl group in FAAF may play a role in DNA damage recognition by UvrABC. The only structural difference between FAF and FAAF is the absence of a bulky acetyl group on the linking nitrogen of the former (Figure 1a). It has been documented that *N*-acetylated FAAF adducts in fully paired duplexes produce a mixture of complex S/B/W-conformers, whereas *N*-deacetylated FAF adopts a simple exchangeable S/B-equilibrium (15,30). Thus, it is clear that the *N*-acetyl group is responsible for generating up to 26% W-conformer in the *NarI* sequence (Figure 4). The bulkyness of the acetyl group with its possibility for *cis* and *trans* rotamer transitions about the amide bond (14,15,45) may facilitate the repositioning of the fluorenyl rings into the minor groove from the S conformation. This conformational rearrangement is relatively straightforward since it does not require a change in the glycosidic bond, which is *syn* in both cases. We observed a good correlation between the proportion of W-conformation and NER efficiency of FAAF.

Moreover, although FAF and FAAF have similar S/B-conformational profile (Figure 4), the *N*-acetyl group in the latter could act as a “conformational locker” to raise the energy barriers among conformers. Such a scenario, i.e., higher energy barriers of FAAF vs.

FAF, is plausible and might contribute to a greater disturbance in DNA, and thus greater repair. By contrast, the *N*-deacetylated FAF adopts a facile interchangeable B/S-equilibrium (<2 kcal/mol) that triggers weaker binding affinities with the damage-recognition protein UvrA. A recent crystal study indicated that the UvrA dimer does not contact the lesion site directly, but rather binds DNA regions on both sides of the modification and primarily recognizes adduct-induced unwinding, bending, and deformity in the overall DNA structure (25). Furthermore, DNA damage recognition in *E. coli* NER is achieved through a sequential 2-step mechanism (46). The initial step is to recognize the adduct-induced distorted DNA structure. After strand opening at the damage site, the DNA adduct structure is further recognized or verified in a second step, which may facilitate the flipping of the adducted nucleotide (47,48). Therefore, it is possible that, for FAF, the second step of recognition plays a more important role than the first step, whereas the first step is a dominate recognition for FAAF.

The order of NER efficiencies described here is roughly consistent with bacterial NER data on AAF adducts embedded in a similar *NarI* sequence (17): G₁ (100%), G₃ (66%), and G₂ (18%). Sequence dependence was also found in human NER of AAF adducted in the *NarI* sequence (18). In contrast to the *E. coli* NER data, however, the AAF adduct at G₂ (100%) was found to be more repairable, followed by G₃ (68%) and G₁ (38%). Despite differences in the nature of proteins involved in prokaryotic and eukaryotic NER, the two systems show similar involvement of β -hairpin intrusion as damage recognition factors (49). Liu *et al.*(50) found a general qualitative trend toward similar relative NER incision efficiencies for 65% of bulky benzo[a]pyrene and equine estrogen substrates.

Similar to bacterial UvrA, Rad4 (XPC) in yeast also recognizes helical distortion to sense DNA damage; unlike bacteria, yeast use a base-flipping mechanism for repair (26). Therefore, the efficiency of repair depends not only on the damage recognition step, but also on other factors, such as ease of base flipping.

In summary, our structural and thermodynamic data provide valuable conformational insights into the sequence-dependent UvrABC incisions of the bulky FAF and FAAF adducts in the *NarI* sequence context. Repair of the bulky *N*-acetylated FAAF adduct seems to occur in a conformation-specific manner, i.e., the highly S/W-conformeric G_3 and G_1 duplexes incised considerably more efficiently than the G_2 duplex ($G_3 \sim G_1 > G_2$)(Table 2). These results were supported by melting and thermodynamic data. Not surprisingly, FAF was repaired 2- to 3-fold less than FAAF; however, the order of incision efficiencies was the reverse of that in the FAAF case. We considered the so-called “*N*-acetyl factor” and lesion-specific recognition mechanism for the different orders of incision for FAF and FAAF. Finally, the temperature dependence of the S/B/W-conformational equilibria of the FAAF-adducts in the *NarI* sequence could provide valuable opportunities for conformation-specific NER utilizing thermophilic UvrABC proteins (51). Taken together, the results of this study demonstrate the complexity of NER mechanisms of bulky DNA lesions.

Funding

This work was supported by National Institutes of Health (Grant number R01CA098296). This research was also made possible in part by the use of the RI-INBRE Research

Core Facility supported by the National Center for Research Resources, National Institutes of Health (Grant number P20 RR016457).

Supplementary Data are available at www.etdadmin.com : Supplementary Table S1-S2, Supplementary Figures S1-S12.

References

1. Luch, A. (2005) Nature and nurture - lessons from chemical carcinogenesis. *Nat. Rev. Cancer*, **5**, 113-125.
2. Melchior, W.B., Jr., Marques, M.M. and Beland, F.A. (1994) Mutations induced by aromatic amine DNA adducts in pBR322. *Carcinogenesis*, **15**, 889-899.
3. Neumann, H.G. (2007) Aromatic amines in experimental cancer research: tissue-specific effects, an old problem and new solutions. *Crit. Rev. Toxicol.*, **37**, 211-236.
4. Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T. and Editors. (2006) *DNA Repair and Mutagenesis, Second Edition*. ASM Press, Washington.
5. Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2006) Prokaryotic nucleotide excision repair: the UvrABC system. *Chem. Rev. (Washington, DC, U. S.)*, **106**, 233-252.
6. Lehmann, A.R. (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie*, **85**, 1101-1111.

7. Van Houten, B. (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiol Rev*, **54**, 18-51.
8. Heflich, R.H. and Neft, R.E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutat. Res.*, **318**, 73-114.
9. Beland, F.A. and Kadlubar, F.F. (1990) *Handbook of Experimental Pharmacology*. Springer-Verlag, Heidelberg.
10. Cho, B.P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health. C. Environ. Carcinog. Ecotoxicol. Rev.*, **22**, 57-90.
11. Luo, C., Krishnasamy, R., Basu, A.K. and Zou, Y. (2000) Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, N-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease. *Nucleic Acids Res.*, **28**, 3719-3724.
12. Meneni, S.R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M.P. and Cho, B.P. (2006) Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res.*, **34**, 755-763.
13. 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.

14. 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.
15. Patnaik, S. and Cho, B.P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol.*, **23**, 1650-1652.
16. 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.
17. Seeberg, E. and Fuchs, R.P. (1990) Acetylaminofluorene bound to different guanines of the sequence -GGCGCC- is excised with different efficiencies by the UvrABC excision nuclease in a pattern not correlated to the potency of mutation induction. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 191-194.
18. Mu, D., Bertrand-Burggraf, E., Huang, J.C., Fuchs, R.P., Sancar, A. and Fuchs, B.P. (1994) Human and E.coli excinucleases are affected differently by the sequence context of acetylaminofluorene-guanine adduct. *Nucleic Acids Res.*, **22**, 4869-4871.
19. Burnouf, D., Koehl, P. and Fuchs, R.P. (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.
20. 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 19F NMR, UV, and CD spectroscopy. *Biochemistry*, **46**, 13310-13321.

21. 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.
22. 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.
23. Mekhovich, O., Tang, M. and Romano, L.J. (1998) Rate of incision of N-acetyl-2-aminofluorene and N-2-aminofluorene adducts by UvrABC nuclease is adduct- and sequence-specific: comparison of the rates of UvrABC nuclease incision and protein-DNA complex formation. *Biochemistry*, **37**, 571-579.
24. Zou, Y., Shell, S.M., Utzat, C.D., Luo, C., Yang, Z., Geacintov, N.E. and Basu, A.K. (2003) Effects of DNA adduct structure and sequence context on strand opening of repair intermediates and incision by UvrABC nuclease. *Biochemistry*, **42**, 12654-12661.
25. Jaciuk, M., Nowak, E., Skowronek, K., Tanska, A. and Nowotny, M. (2011) Structure of UvrA nucleotide excision repair protein in complex with modified DNA. *Nat. Struct. Mol. Biol.*, **18**, 191-197.
26. Min, J.H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570-575.
27. Cho, B.P. and Zhou, L. (1999) Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by ¹⁹F NMR spectroscopy. *Biochemistry*, **38**, 7572-7583.
28. Gao, L., Zhang, L., Cho, B.P. and Chiarelli, M.P. (2008) Sequence verification of oligonucleotides containing multiple arylamine modifications by enzymatic

- digestion and liquid chromatography mass spectrometry (LC/MS). *J. Am. Soc. Mass Spectrom.*, **19**, 1147-1155.
29. Jain, N., Meneni, S., Jain, V. and Cho, B.P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.*, **37**, 1628-1637.
  30. Meneni, S.R., Shell, S.M., Gao, L., Jurecka, P., Lee, W., Sponer, J., Zou, Y., Chiarelli, M.P. and Cho, B.P. (2007) Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. *Biochemistry*, **46**, 11263-11278.
  31. Chakrabarti, M.C. and Schwarz, F.P. (1999) Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry. *Nucleic Acids Res.*, **27**, 4801-4806.
  32. Zou, Y., Liu, T.M., Geacintov, N.E. and Van Houten, B. (1995) Interaction of the UvrABC nuclease system with a DNA duplex containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE. *Biochemistry*, **34**, 13582-13593.
  33. Zou, Y. and Van Houten, B. (1999) Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J.*, **18**, 4889-4901.
  34. Tan, X., Suzuki, N., Grollman, A.P. and Shibutani, S. (2002) Mutagenic events in *Escherichia coli* and mammalian cells generated in response to acetylaminofluorene-derived DNA adducts positioned in the Nar I restriction enzyme site. *Biochemistry*, **41**, 14255-14262.
  35. 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.
36. Liang, F., Meneni, S. and Cho, B.P. (2006) Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene-DNA adducts. *Chem. Res. Toxicol.*, **19**, 1040-1043.
  37. Liang, F. and Cho, B.P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry*, **49**, 259-266.
  38. 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 NarI recognition sequence. *Chem. Res. Toxicol.*, **20**, 445-454.
  39. 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 NarI restriction enzyme, a hotspot for -2 bp deletions. *J. Am. Chem. Soc.*, **128**, 10085-10095.
  40. 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 NarI restriction enzyme. *Biochemistry*, **46**, 8498-8516.
  41. Brown, K., Hingerty, B.E., Guenther, E.A., Krishnan, V.V., Broyde, S., Turteltaub, K.W. and Cosman, M. (2001) Solution structure of the 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine adduct in duplex DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8507-8512.

42. Koehl, P., Valladier, P., Lefevre, J.F. and Fuchs, R.P. (1989) Strong structural effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Nucleic Acids Res.*, **17**, 9531-9541.
43. Veaute, X. and Fuchs, R.P. (1991) Polymorphism in N-2-acetylaminofluorene induced DNA structure as revealed by DNase I footprinting. *Nucleic Acids Res.*, **19**, 5603-5606.
44. Gillet, L.C., Alzeer, J. and Scharer, O.D. (2005) Site-specific incorporation of N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (dG-AAF) into oligonucleotides using modified 'ultra-mild' DNA synthesis. *Nucleic Acids Res.*, **33**, 1961-1969.
45. Shapiro, R., Hingerty, B.E. and Broyde, S. (1989) Minor-groove binding models for acetylaminofluorene modified DNA. *J. Biomol. Struct. Dyn.*, **7**, 493-513.
46. Zou, Y., Luo, C. and Geacintov, N.E. (2001) Hierarchy of DNA damage recognition in Escherichia coli nucleotide excision repair. *Biochemistry*, **40**, 2923-2931.
47. Malta, E., Verhagen, C.P., Moolenaar, G.F., Filippov, D.V., van der Marel, G.A. and Goosen, N. (2008) Functions of base flipping in E. coli nucleotide excision repair. *DNA Repair (Amst)*, **7**, 1647-1658.
48. Malta, E., Moolenaar, G.F. and Goosen, N. (2006) Base flipping in nucleotide excision repair. *J Biol Chem*, **281**, 2184-2194.
49. Scharer, O.D. (2011) Multistep damage recognition, pathway coordination and connections to transcription, damage signaling, chromatin structure, cancer and aging: current perspectives on the nucleotide excision repair pathway. *DNA Repair (Amst)*, **10**, 667.

50. Liu, Y., Reeves, D., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, M., Kolbanovskiy, A., Bolton, J.L., Broyde, S., Van Houten, B., Geacintov, N. E. (2011). Probing for DNA damage with beta-hairpins: similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair (Amst)*, **10**, 684-696.
51. Ruan, Q., Liu, T., Kolbanovskiy, A., Liu, Y., Ren, J., Skorvaga, M., Zou, Y., Lader, J., Malkani, B., Amin, S. *et al.* (2007) Sequence context- and temperature-dependent nucleotide excision repair of a benzo[a]pyrene diol epoxide-guanine DNA adduct catalyzed by thermophilic UvrABC proteins. *Biochemistry*, **46**, 7006-7015.

Figure legends

Figure 1: (a) Chemical structures of AAF, FAAF, AF and FAF adducts; (b) sequences of fully paired 16-mer and 12-mer *NarI* duplexes used in the present study; major groove views of the central trimer segments of (c) the B/S and (d) B/S/W-conformer equilibrium of FAF and FAAF-modified duplexes. The modified dG and the complementary dC are shown in red and green sticks, respectively, and the aminofluorene moiety is highlighted with shiny grey CPK and the *N*-acetyl with pink CPK. In the B-type conformer, anti-[FAAF/FAF]dG maintains Watson-Crick hydrogen bonds, thereby placing the carcinogen moiety in the major groove. The carcinogens in the S- and W-conformers stack into the helix or wedged into the minor groove, respectively, with the modified dG in the syn conformation.

Figure 2: (a) Chromatogram of a reaction mixture between 16-mer *NarI* sequence (5'-CTCTCG₁G₂CG₃CCATCAC-3') and an activated FAAF (*N*-acetoxy-*N*-2-(acetylamino)-7-fluorofluorene). The mono- (G₁, G₃, G₂), di-, and tri-FAAF adducts eluted in the 28-35, 42-60, and 84 min were purified by reversed-phase HPLC (see Materials and Method for gradient condition); (b) on-line photodiode array UV/Vis spectra of mono-, di- and tri-FAAF adducts.

Figure 3: (a) CD spectral overlays recorded at 15 °C and (b) differential scanning calorimetry (DSC) curves recorded in 20 mM phosphate buffer containing 0.1 M NaCl at pH 7.0 of fully paired 16-mer *NarI* duplexes with FAAF modification at G₁ (green), G₂ (blue), and G₃ (pink).

Figure 4: ^{19}F NMR spectra of (a) FAAF-modified *NarI* 16-mer, b) FAAF-modified non-*NarI* 12-mer, and (c) FAF-modified *NarI* 12-mer duplexes at 5 °C. *unknown conformers; #impurity.

Figure 5: Dynamic ^{19}F NMR spectra of fully paired 16-mer *NarI* duplexes. FAAF modification at (a) G_1 , (b) G_2 , and (c) G_3 . *unknown conformers; #impurity.

Figure 6: Absolute percent incision rates of a) FAAF and b) FAF-*NarI* duplexes modified at G_1 , G_2 and G_3 ; c) percent incision rates histogram of FAF and FAAF at different positions relative to *NarI*- G_3 FAAF as 100%.

Table legends

Table 1: Thermal and thermodynamic parameters of FAAF modified *NarI* duplexes obtained from Differential Scanning Calorimetry

Table 2: Conformational heterogeneity (B/S/W), thermal destabilization and relative percent incision rates of FAAF- and FAF-modified *NarI* duplexes

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Table 1:

	5'-CTCTCG ₁ G ₂ CG ₃ CCATCAC-3' 3'-GAGAGC C GCGGTAGAG-5'							
	$-\Delta H$ kcal/mol	$-\Delta S$ eu	$-\Delta G_{37^\circ\text{C}}$ kcal/mol	T_m^b °C	$\Delta\Delta H^c$ kcal/mol	$\Delta\Delta S^d$ eu	$\Delta\Delta G_{37^\circ\text{C}}^e$ kcal/mol	ΔT_m^f °C
Control ^a	117.5	319.1	18.6	73.9	-	-	-	-
NarI-G ₁ -FAAF ^a	95.6	260.2	14.9	68.6	21.9	58.9	3.7	-5.3
NarI-G ₂ -FAAF ^a	98.9	272.1	14.5	66.0	18.6	47.0	4.1	-7.9
NarI-G ₃ -FAAF ^a	92.8	254.4	13.9	65.6	24.7	64.7	4.7	-8.3

a) The average standard deviations for $-\Delta G$, $-\Delta H$, and T_m are ± 0.4 , ± 3.0 , and ± 0.4 , respectively.

b) T_m values is the temperature at half the peak area.

c) $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex).

d) $\Delta\Delta S = \Delta S$ (modified duplex) - ΔS (control duplex).

e) $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).

f) $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Table 2:

<i>NarI</i> Duplexes	Population Ratios^a (%)			ΔT_m^b (°C)	Relative incision rate^c (%)
	<i>B</i>	<i>S</i>	<i>W</i>		
<i>NarI</i> -G ₁ -FAAF	46	34	20	-5.3	93
<i>NarI</i> -G ₂ -FAAF	57	15	9	-7.9	32
<i>NarI</i> -G ₃ -FAAF	13	61	26	-8.3	100
<i>NarI</i> -G ₁ -FAF	42	58	-	-9.4	44
<i>NarI</i> -G ₂ -FAF	69	31	-	-6.8	43
<i>NarI</i> -G ₃ -FAF	35	65	-	-8.3	25

^aThe percent population ratios were calculated at 5°C on the basis of line simulations

^b $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

^cPercent incision rate of modified duplexes with respect to *NarI*-G₃-FAAF (100 %).

Manuscript-II

To be submitted in *Chemical Research in Toxicology*

Nucleotide excision repair of cluster acetylaminofluorene adducts on the *NarI* sequence: sequence dependence and structural factors

Vipin Jain¹, Benjamin Hilton², Anshu Jain¹, Yue Zou², and Bongsup P. Cho¹

¹Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI, 02881, USA

² Department of Biomedical Sciences, East Tennessee State University, Johnson city, TN, 37614, USA

*To whom correspondence should be addressed:

Telephone#: (+1) 401-874-5024; FAX# (+1) 401-874-5766; E-mail: bcho@uri.edu

Present address:

Department of Biomedical and Pharmaceutical Sciences,
University of Rhode Island
7 Greenhouse Road,
Kingston, RI, 02881, USA

ABSTRACT

Cluster DNA damage refers to two or more lesions in a single turn of the DNA helix. Evidence indicates that the tandem oxidative DNA damages are less repairable than a single lesion, thus more genotoxic. Here, we describe the nucleotide excision repair (NER) of bulky cluster lesions of 2-acetylaminofluorene (AAF) as it is one of the most studied arylamine carcinogens and an excellent substrate for NER. The *E. coli* UvrABC proteins remove the dG-AAF in the *NarI* sequence (5'--CG₁G₂CG₃CC--3') in a sequence dependent manner ($G_3 \geq G_1 > G_2$) with highly syn-conformeric AAF at G₁ and G₃ incised more efficiently than highly anti-conformeric G₂. We prepared model duplexes in which two fluorine-labeled acetylaminofluorene adducts (dG-FAAF) are separated by none, one and two nucleotides in the *NarI* hot spot, i.e. 5'--CG₁*G₂*CG₃CC--3', 5'--CG₁G₂*CG₃*CC--3', and 5'--CG₁*G₂CG₃*CC--3' (G*=dG-FAAF), respectively. We performed NER (*E. coli* UvrABC) and structural studies on these cluster FAAF lesions. Our NER results showed greater reparability of the di-adducts when compared to the corresponding mono-adducts. Moreover, we observed dramatic sequence effects in their repair efficiency in the order of *NarI*-G₂G₃ > -G₁G₃ > -G₁G₂. Our NMR/CD and UV-melting results revealed that in contrast to the mono, di-adducts produced synergistic effect on duplex destabilization. In addition, FAAF at G₂G₃ and G₁G₃ destacks and destabilizes the duplexes more than at G₁G₂. Overall, the results indicate the importance of carcinogen induced destacking and related thermal/thermodynamic destabilization in the NER of bulky cluster adducts.

INTRODUCTION

The human genome is susceptible to various chemical assaults from external and internal sources including UV radiation and carcinogens. Adduct formation occurs throughout the genome: some sites showed higher adduct formation, yet little or no mutation induction and vice versa.¹⁻³ The so-called locally multiply damaged sites (LMDS) or clustered lesions are defined as two or more lesions occurring in a short stretch (<24 bp) of DNA.⁴ In general, cluster lesions are considered to be more genotoxic than a single lesion. It has also been hypothesized that cluster lesions are less repairable than individual ones and that mutation of one lesion is synergized by the presence of another lesion in the vicinity.^{5, 6} Studies have shown that the efficiency for the repair of clustered oxidative damage depends on the nature of chemical changes, inter-lesion distance, sequence contexts, and their relative orientation.⁷ Kalam et al⁵ have evaluated the mutagenic potential of 8-oxo-G adjacent to a uracil in simian kidney cells with the assumption that the uracil would be excised by uracil glycosylase to produce an apurinic site. A substantial fraction of tandem mutations was detected when the uracil was adjacent to 8-oxo-G. The presence of an AP site either 5' or 3' to 8-oxo-G increased the frequency of 8-oxo-G to T transversions with the effect much greater on the latter. It has also been suggested that cluster oxidative lesion could be converted into double strand breaks during repair process, highlighting the complexity of cluster-induced mutagenesis.⁸

Evidence suggests that a similar clustering occurs with bulky DNA lesions and may be responsible for their sequence dependent repair and mutational outcomes. Fuchs and coworkers have shown that the binding spectrum of 2-acetylaminofluorene (AAF) is

essentially random and all guanine residues exhibited equal reactivity towards N-AcO-AAF.^{9, 10} As in oxidative damages, GC rich regions and runs of contiguous guanines in the genome possess a higher probability of producing cluster adducts.¹¹ Like oxidative damages, the bulky adducts can also induce clustering either through an individual lesion or a mixture of multiple lesions. Since genomic DNA is under continuous assault by various chemicals, different combination of bulky cluster adducts are possible, for instance, combinations of arylamines and heterocyclic amines, arylamines, polycyclic aromatic hydrocarbons (PAH) and estrogen adducts etc. Moreover, the metabolic activation of these carcinogens produces reactive species which in-turn can form different DNA-adducts. Aside from binding, other factors such as repair and neighboring sequences play an important role in the formation of mutation hotspots.¹² 5-Methyl cytosine in CpG sites can control the dissemination of guanine adduct formation by BPDE, MNU, and AAF.¹¹ The cluster lesions in close vicinity may have an important ramification in NER damage recognition and mutational outcomes.⁷ Wei et al have shown that the rates of excision repair at locations modified by ($\pm$)-anti-BPDE varied extensively along exon 3 of the HPRT gene.¹³ Interestingly, mutations induced by bulky adducts are not always at the adduct-bearing nucleotides (non-targeted mutation, frameshift mutations, etc).

Arylamines are implicated in the etiology of human breast, liver and bladder cancers.¹⁴ The prototype AAF and its 2-amino and nitro derivatives produce two major DNA adducts upon *in-vivo* activation: *N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene (dG-C8-AF) and *N*-(2'-deoxyguanosin-8-yl)- 2-acetylaminofluorene (dG-C8-AAF)(Fig. 1a).¹⁵ Both AF and AAF adducts exist in an equilibrium of external binding anti-B-type and syn stacked S-

conformers. In the bulky N-acetylated AAF, however, the *syn*- adduct could also exist in the minor groove binding “wedge (W)-conformer” (Fig. 1c).^{16, 17} The population ratios of these different conformations were found to be strongly dependent on the nature of neighboring sequence context.¹⁶⁻²¹

A number of factors have been implicated to be important for the repair of bulky DNA lesions such as arylamines and PAHs.²²⁻²⁵ They include disruption of Watson-Crick hydrogen bonding, DNA bending, thermodynamic destabilization, local conformational flexibility etc.²³ Striking sequence effects were observed in the *E. coli* UvrABC mediated NER of fluorinated FAF and FAAF when present at different guanines of the *NarI* sequence (5'--G₁G₂CG₃CC--3'). We reported that FAAF exhibits conformation-specific repair, where highly S/W-conformeric G₃ and G₁ duplexes incised considerably more efficiently than the highly B-conformeric G₂ duplex (G₃ ~ G₁ > G₂). The NER efficiency of the N-deacetylated FAF was 2- to 3-fold less, but it too displayed strong sequence dependence.¹⁶ Mu et al have confirmed earlier human NER results of AF and AAF by Fuchs group and provided the molecular dynamics-based structural rationale for their sequence and lesion dependence in repair.²⁶ A similar dramatic sequence dependence of repair has also been observed in non *NarI* sequences. For example, we reported recently the NER efficiency on two duplexes with a G*CX sequence context, in which G* is a C8-substituted dG adducts derived from fluorinated analogs of 4-aminobiphenyl (FABP), 2-aminofluorene (FAF), and 2-acetylaminofluorene (FAAF), and X is either dA or dT. The results showed that FABP and FAF exist in a simple mixture of S and B conformers, while FAAF additionally adopts a W conformer. The bulky N-acetylated FAAF is repaired 3-4

times more efficiently than FABP and FAF. A simple A to T polarity swap in the G*CA/G*CT transition produced a dramatic increase in *syn*-conformation and resulted in 2~3-fold lower NER efficiencies in *E. coli*. These results indicate that lesion-induced DNA bending/thermodynamic destabilization is an important DNA damage recognition factor, more so than the local S/B-conformational heterogeneity that was observed previously for FAF and FAAF in certain sequence contexts.^{16, 21, 27} This work represents a novel 3'-next flanking sequence effect as a unique NER factor in *E. coli*.

Virtually nothing is known about the structure and repair consequence of cluster bulky lesions. As elaborated above, previous studies have mainly focused on single site-specifically modified sequences, primarily because of technical ease dealing with a single lesion than multiple adducts.²⁸ We hypothesize that cluster arylamine lesions could be formed wherever multiple guanines are present in the genome, resulting in major structural/ conformational consequences on repair and mutational outcomes.

To that end, we prepared the duplex NER substrates in which the bulky FAAF lesions are located on two (G₁G₂, G₂G₃ and G₂G₃) of the three guanines (G₁, G₂, and G₃) of the *NarI* recognition sequence (Fig. 1b). We conducted spectroscopic and melting experiments for conformational and thermodynamic analyses and NER measurements using the *E. coli* UvrABC system. The results present the first report on the NER of the bulky cluster FAAF lesions and its differential efficiency at different guanine pairs (G₁G₂, G₂G₃ and G₁G₃) of the *NarI* sequence.

MATERIALS AND METHODS

Caution: *2-Aminofluorene derivatives are mutagens and suspected human carcinogens and therefore must be handled with caution.*

Crude oligodeoxynucleotides (ODN, 10 μ mol scale) in desalted form were purchased from Eurofins MWG operon (Huntsville, AL). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Preparation and Characterization of FAAF-modified di-adduct ODNs

We reported previously the preparation of 16-mer *NarI* ODNs (5'-CTCTCG₁G₂CG₃CCATCAC-3'), in which each of the three guanine were adducted by FAAF.¹⁶ We have also documented that the incorporation of fluorine atom at the longest axis does not affect the overall conformational and thermodynamic profiles of aminofluorene-modified duplexes (e.g., FAF, FAAF).^{19, 29} FAAF-modified 16-mer di-adduct ODNs were prepared similarly with a longer reaction time.^{16, 17, 30} Briefly, approximately 0.5-1 mg of *N-acetoxy-N-2-(acetylamino)-7-fluorofluorene* dissolved in absolute ethanol was added drop wise to a sodium citrate buffer (pH 6.0) containing 200-250 ODs of unmodified ODN (5'-CTCTCG₁G₂CG₃CCATCAC-3') and placed in a shaker for 5 min at 37 °C. Figure 2a shows a typical reversed-phase HPLC chromatogram derived from the resulting reaction mixture. The FAAF modified oligomers appearing between 28-85 min range were separated and purified up to >97% purity by repeated injections. The HPLC system consisted of a Hitachi EZChrom Elite HPLC unit with an L2450 diode array detector and a Phenomenex Luna C18 column (150 X 10 mm, 5.0 μ m). We employed a gradient system involving 3-15% acetonitrile for 40 min followed

by 15-20% and 20-35% acetonitrile for 20 min and 40 min, respectively, in pH 7.0 ammonium acetate buffer (100 mM) with a flow rate of 2.0 mL/min.

The three di-FAAF-modified 16-mer sequences were each annealed with the corresponding complementary sequence (5'-GTGATGGCGCCGAGAG-3') to form fully paired duplexes, designated as *NarI*-G₁G₂, *NarI*-G₂G₃, and *NarI*-G₁G₃ depending upon the location of FAAF modification (Fig. 1b).

MALDI-TOF characterization of di-adduct ODNs

The modified 16-mer ODNs were characterized by following the general procedures on 3'-5' and 5'-3' exonuclease digestion/MALDI-TOF (time of flight) mass spectrometry.^{31, 32} In general, 1 µL of sample containing approximately 100-250 pmol of modified ODN was used for the digestion. In 5'-3' digestion, 0.01 unit of bovine spleen phosphodiesterase (BSP) was added to a solution containing 1 µL of ODN and 7 µL of water. Similarly, 0.1 unit of snake venom phosphodiesterase (SVP) was added to a solution of 1 µL ODN, 6 µL ammonium citrate (100mM, pH 9.4), and 6 µL of water for 3'-5' digestion. In each case, 1 µL of reaction mixture was withdrawn at regular time intervals until the digestion rate was significantly reduced, the reaction was quenched by mixing the aliquot with 1 µL of matrix (3-HPA and DHAC in 1:1 ratio). The samples were then immediately spotted on a MALDI plate and the spectra were collected on the Shimadzu Axima Performance mass spectrometer equipped with a 50 Hz nitrogen laser in the positive ion reflectron mode.

UV Melting

UV melting data were obtained using a Cary100 Bio UV/VIS spectrophotometer equipped with a 6×6 multi-cell block and 1.0 cm path length. Sample cell temperatures were controlled by a built-in Peltier temperature controller. Oligonucleotide duplexes with a concentration range of 1.2-6.0 μM were prepared in solutions containing 0.2 M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA at pH 7.0. Thermomelting curves were constructed by varying temperature of the sample cell (1 $^{\circ}\text{C}/\text{min}$) and monitoring absorbance at 260 nm. A typical melting experiment consisted of forward/reverse scans and was repeated five times. Thermodynamic parameters were calculated using the program MELTWIN version 3.5 as described previously.²⁰

Circular Dichroism (CD)

CD measurements were conducted on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. Typically, 1 ODs of each strand were annealed with an equimolar amount of a complementary sequence. The samples were dissolved in 200 μL of a neutral buffer (0.2 M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA) and placed in a 1.0 mm path length cell. The samples were heated at 85 $^{\circ}\text{C}$ for 5 min and then cooled to 15 $^{\circ}\text{C}$, over a 10 min period to ensure complete duplex formation. Spectra were acquired every 0.2 nm with a 2 s response time from 200 to 400 nm at a rate of 50 nm/min, were the averages of 10 accumulations, and were smoothed using the 17-point adaptive smoothing algorithms provided by Jasco.

Dynamic ^{19}F NMR

Approximately 10 ODs of a di-FAAF 16-mer ODN was annealed with an equimolar amount of a complementary sequence (5'-GTGATGGCGCCGAGAG-3') to produce a fully paired duplex (Fig. 1b). The duplex samples were then dissolved in 300 μL of a pH 7.0 NMR buffer containing 10% D_2O /90% H_2O , 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA, and filtered into a Shigemi tube through a 0.2 μm membrane filter. All ^1H and ^{19}F NMR results were recorded using a HFC probe on a Varian NMR spectrometer operating at 500.0 and 476.5 MHz, respectively, using the usual acquisition parameters described previously^{19, 29, 33}. Imino proton spectra were obtained at 5°C using a WET-1D sequence and referenced relative to that of DSS. ^{19}F NMR spectra were acquired in the ^1H -decoupled mode and referenced relative to that of CFCl_3 by assigning external C_6F_6 in C_6D_6 at -164.9 ppm. ^{19}F NMR spectra were measured between 5 and 65 °C with an increment of 5-10 °C. Temperatures were maintained by a Varian FTS control system.

Substrate Construction and UvrABC Protein Purification

DNA substrates of 55 bp containing two FAAF adducts simultaneously on two of three guanine residues were constructed as previously described.^{34, 35} Briefly, *NarI*-G₁G₂, *NarI*-G₂G₃ and *NarI*-G₁G₃ 16-mer ODNs was individually ligated with flanking 20-mer ODN (5'-GACTACGTACTGTTACGGCT-3') and 19-mer ODN (5'-GCAATCAGGCCAGATCTGC-3') ODN at the 5'- and 3'-end, respectively. The 20-mer was 5'-terminally labeled with ^{32}P . The ligation product was purified by urea-PAGE under denaturing conditions. Following the purification, the substrate was annealed to

the corresponding complementary strand, and then purified on an 8% native polyacrylamide gel.

UvrA, UvrB, and UvrC proteins were over expressed in *E. coli* and then purified as previously described.³⁶ The estimated purity of the three proteins was greater than 95%. A Bio-Rad Protein Assay was used to determine the protein concentration with BSA as the standard based on the manufacturer-recommended procedures.

Nucleotide Excision Assay and Quantification of Incision Products

The 5'-terminally labeled DNA substrates were incised by UvrABC as previously described.^{34, 35} Briefly, the DNA substrates (2 nM) were incubated in the UvrABC reaction buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM MgCl₂, 5 mM DTT) at 37°C in the presence of UvrABC (UvrA, 10 nM, UvrB, 250 nM, and UvrC, 100 nM). The Uvr proteins were diluted and premixed in Uvr storage buffer before addition to the reaction. Aliquots were collected at 0, 5, 10, 15, and 20 minutes into the reaction. The reaction was terminated by heating at 95°C for five min. The products were denatured by addition of formamide loading buffer and heating to 95°C for five min, followed by quick chilling on ice. The incision products were then analyzed by electrophoresis on a 12% polyacrylamide sequencing gel under denaturing conditions with TBE buffer (50 mM Tris-Borate, 1 mM EDTA, pH 8.0).

To quantify the incision products, radioactivity was measured using a Fuji FLA-5000 Image Scanner with MultiGauge V3.0 software. The DNA incised (in fmol) by UvrABC

was calculated based on the total molar amount of DNA used in each reaction and the ratio of the radioactivity of incision products to total radioactivity of DNA. At least three independent experiments were performed for determination of the rates of incision.

RESULTS

Model Sequences

We previously reported that a simply mixing of the *NarI* 16-mer ODN (5'-CTCTCG₁G₂CG₃CCATCAC-3') with the chemically reactive FAAF at room temperature produced approximately equal mixture of mono-, di- and tri-adducts.¹⁶ A longer reaction time increased the production of di- and tri-adducts which were clearly separable on a RP-HPLC as shown in Figure 2a. We grouped them previously into three: Peaks 1-3 (28-35 min), 4-6 (42-60 min), and 7 (84 min) as mono-, di-, tri-adducts on the basis of relative absorption intensity (1:2:3, respectively) of adduct induced shoulder in the 290-320 nm range.¹⁶ Moreover, we assigned Peak 1, 2 and 3 as G₁, G₃ and G₂, respectively, using the well-known exonuclease-LC/MS methods.¹⁶ In the present study, we have characterized Peak 4, 5 and 6 using MALDI-TOF as described below. The di-adducts were designated as *NarI*-G₁G₂, *NarI*-G₂G₃, or *NarI*-G₁G₃, in which the numbers signify the position of FAAF-modified guanines (Fig. 1b).

MALDI-TOF Characterization

The molecular weights of di-FAAF modified 16-mer were measured by MALDI-TOF prior to sequence verification by exonuclease digestion. Ionization of ODNs normally occurs due to the addition of a proton on the phosphate group in the ODN backbone.

Exonucleases cleave terminal deoxynucleotides from the ODN chain until the FAAF-modified dG is exposed at the end of the chain, at which point digestion reaction slows down significantly. The position of modification was identified (in this case) when fragment(s) are formed by the loss of an unmodified guanine nucleotide is observed in the MALDI-TOF spectra. Figure 3 (see Fig. 4 for complete time range) represents the 3'→5' (Fig. 3a) and 5'→3' (Fig. 3b) exonuclease digestion fragments for Peak 4 at different time intervals. In general, the 3'→5' digestion was much faster than 5'→3'. The ions observed at m/z 5257.2 at 30 sec (Fig. 3a) and m/z 5257.5 at 5 mins (Fig. 3b) represent $[M+H]^+$ ion of the doubly modified *NarI*-16mer ODN. Increase in incubation time leads to the digestion of subsequent unmodified bases. However, 3' and 5' digestions were significantly slowed down at m/z 3170.2 and 3781.0, respectively. The ion observed at m/z 3170.2 at 15 min of 3'-digestion represents $[M+H]^+$ ion formed from the doubly modified 5'-CTCTCG₁G₂CG₃-3' fragment (see Fig. 3 inset for theoretical MW values). These results indicate a FAAF modification at the G₃ position of *NarI* ODN. In 5'-digestion, the ion observed at m/z 3781.0 at 80 mins (Fig. 3b) represents $[M+H]^+$ ion formed from the doubly modified 5'-G₁G₂CG₃CCATCAC-3' fragment, suggesting another FAAF at the G₁ position. The ion at m/z 4070.2 of 5'-digestion (Fig. 3b) indicates that the digestion slowed down one base prior to the modification. Taken all together, peak 4 was assigned as *NarI*-G₁G₃ (5'-CTCTCG₁(FAAF)G₂CG₃(FAAF)CCATCAC-3').

Peak 5 and 6 were similarly characterized in the same manner and the results are shown in Figure 5 and 6, respectively. The 5'- and 3'-digest of peak 5 exhibit ions at m/z 3452.6 (Fig. 5a) and m/z 3169.5 (Fig. 5b), which correspond to the $[M+H]^+$ ions formed from the double modification of 5'-G₂CG₃CCATCAC-3' and 5'-CTCTCG₁G₂CG₃-3' fragments,

respectively. Therefore, Peak 5 was assigned as *NarI*-G₂G₃. The 5'- and 3'-digest of Peak 6 exhibited ions at *m/z* 3777.8 (Fig. 6a) and *m/z* 2551.6 (Fig. 6b) indicating [M+H]⁺ ions formed from the double modification of 5'-G₁G₂CG₃CCATCAC-3' and 5'-CTCTCG₁G₂-3' fragments, respectively. Hence, Peak 6 was assigned as *NarI*-G₁G₂ (5'-CTCTCG₁(FAAF)G₂(FAAF)CG₃CCATCAC-3'). The HPLC elution pattern and mass characterization of the FAAF di-adduct 16-mer ODNs in the present study is similar to that of the FAF di-adduct 12-mer *NarI* sequence (5'-CTCG₁G₂CG₃CCATCAC-3') reported by Gao *et al.*³¹

Circular Dichroism

Figure 7a shows an overlay of the CD spectra for the three FAAF di-adduct *NarI*-G₁G₃-, G₂G₃ and -G₁G₂ duplexes relative to the unmodified control (brown) all in ~35 μM concentration. Unlike FAF and FABP, the ICD_{290-320 nm} pattern for FAAF induced S/B/W-heterogeneity has not been established. Nonetheless, comparative pattern analysis provided some valuable conformational insight. Unmodified and FAAF di-adduct duplexes both displayed a positive and negative ellipticity at around 270 and 250 nm, respectively, which is characteristic for a B-form DNA double helix. The modified duplexes consistently exhibited a reduced intensity around 250 nm, signifying the alteration of the right handed helicity. All modified duplexes displayed slight blue shifts (2-3 nm) relative to the unmodified control duplex, indicating the lesion impact on the DNA backbone. The effect was most prominent in *NarI*-G₁G₃. Also, the modified duplexes exhibited a change in the positive ellipticity at around 270 nm, indicating the alterations of base stacking of a B-form DNA. While the *NarI*-G₁G₂ exhibited a slight

increase, the *NarI*-G₂G₃ and -G₁G₃ duplexes actually displayed a decrease in intensity with the greater impact in the former (Fig. 7a).

In addition, sequence-dependent induced CD in the 290-320 nm range (ICD_{290-320 nm}) was seen in all three di-adduct duplexes. Previously, we reported that the FAAF-modified *NarI* mono-adduct duplexes displayed negative dip, with the *NarI*-G₂ duplex exhibited a greater dip than G₁ or G₃ duplexes.¹⁶ Similar ICD_{290-320nm} was observed in the di-adduct duplexes. The *NarI*-G₁G₃ and -G₂G₃ duplexes showed a bigger dip than the *NarI*-G₁G₂ duplex.

UV Melting Experiments

Figure 7b shows the UV melting profiles of the three di-FAAF duplexes in comparison to the unmodified control at 4.8 μM. All duplexes exhibited typical monophasic, sigmoidal, helix to coil transitions with an excellent linear correlation ($R^2 > 0.9$) between T_m^{-1} and $\ln C_t$. It is clear that all modified duplexes were thermally and thermodynamically destabilized relative to the control. Table 1 summarizes the thermal and thermodynamic parameters calculated from the UV-melting experiments. The results were converted into the thermodynamic chart shown in Figure 8. The *NarI*-G₂G₃ duplex was most destabilized with $\Delta\Delta G_{37^\circ\text{C}} = 7.8$ kcal/mol, $\Delta\Delta H = 23.3$ kcal/mol and $\Delta T_m = -17.9$ °C. The ΔT_m value of *NarI*-G₁G₃ was -14.1 °C, approximately 4 °C lower than *NarI*-G₂G₃, but had a slightly higher $\Delta\Delta H = 24.1$ kcal/mol. However, the large entropy ($\Delta\Delta S = 56.0$ eu) compensated for the enthalpy (Fig. 8, Table 1), resulting in the overall free energy destabilization of $\Delta\Delta G_{37^\circ\text{C}} = 6.8$ kcal/mol. In comparison, the stability of the *NarI*-G₁G₂

was least affected ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol, $\Delta T_m = -10$ °C). The weak entropy compensation ($\Delta\Delta S = 6.0$ eu) for the substantial enthalpy reduction ($\Delta\Delta H = 6.2$ kcal/mol) resulted into the free energy destabilization ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol) (Fig. 8, Table 1). Overall, the magnitude of thermal (ΔT_m) and thermodynamic ($\Delta\Delta G$) destabilization was in the order of $NarI\text{-}G_2G_3 > NarI\text{-}G_1G_3 > NarI\text{-}G_1G_2$.

Dynamic ^{19}F NMR Spectroscopy

Figure 9 shows the ^{19}F NMR spectra of di-adduct duplexes as a function of temperature (5-65 °C). As expected, the di-adduct duplexes displayed an array of ^{19}F signals in the -114.0 to -118 ppm range. This is the same chemical shift range in which mono-FAAF adducts display their B- (-115.0 to -115.5 ppm), S- (-115.5 to -117.0 ppm), and W-conformers (-117.0 to -118.0 ppm) conformers.^{16, 17} The complexity observed in the ^{19}F NMR spectra of di- over mono-FAAF adduct was expected owing to the presence of two FAAF lesions, each of which is capable of exhibiting S/B/W heterogeneity. As a result, it was very difficult to assign signals as we had done for the mono-adducts.¹⁶

Nonetheless, all di adducts, exhibited typical dynamic NMR characteristics, i.e., exchange broadening of ^{19}F signals with increasing temperature, reaching an eventual coalescence signals at 65 °C and above, indicating thermal denaturation of a duplex (Fig. 9). In addition, overall signal patterns were found to be flanking sequence dependent. Figure 10a compares the ^{19}F NMR spectra of $NarI\text{-}G_2G_3$ with the mono-FAAF counterparts ($NarI\text{-}G_2$ and $-G_3$) at two different temperatures (5 and 20 °C). Similar comparisons were made for $NarI\text{-}G_1G_3$ and $NarI\text{-}G_1G_2$ (Fig. 10b and c). The expectation

was that the signal patterns of the di-adduct duplex would be the sum of their respective mono-adducts, but that turned out not to be the case. It is possible that the two nearby FAAF adducts would have some kind of complex conformational interactions with each other, influencing their conformational profiles.

It was noted that signals of *NarI*-G₁G₃ and G₂G₃ are broad even at 5 °C, indicating some conformational dynamic exchanges. This observation suggests a conformational flexibility at the lesion site, consistent with their high entropy values (Fig. 8, Table 1). In contrast, the ¹⁹F signals of *NarI*-G₁G₂ were comparatively sharper in line with their low entropy (Table 1). In all three duplexes, increasing temperatures resulted in coalescence of signals into a single signal (~115 ppm) which represents a free modified single strand. The *NarI*-G₁G₂ duplex showed a greater duplex melting signal at ~65 °C compared to 60 °C for the other two duplexes. These results are consistent with the UV-melting data above. The ¹⁹F conformational complexity was supported by their imino protons spectra (Fig. 11). The *NarI*-G₁G₃ and *NarI*-G₁G₂ duplexes exhibited multiple signals in the non-Watson-Crick base pairing 11-12 ppm region, which is known to occur from the imino protons of modified guanines. In *NarI*-G₂G₃, however, the signals were shifted downfield to 12.2-12.3 ppm.

UvrABC Incisions of *NarI* FAAF di-adducts

The NER kinetic assay results are shown in Figures 12 and 13, in which the 55-mer FAAF-diadduct duplex substrates were incised by UvrABC nuclease. These substrates were radioactively labeled at the 5'-end of the adducted strand. The major incision

products can be seen as ~18 to 20-mer fragments separated on a urea-PAGE gel under denaturing conditions (Fig. 12). The incision occurred at the 8th phosphate bond 5' to the modified nucleotide, which is consistent with the previously reported results of UvrABC incision.^{34, 35} Quantitative analysis of the incision products suggest that the di-adduct substrates were incised 1.5-2 times more efficiently than the previously reported mono-adducts (Fig. 13).¹⁶ Moreover, the substrates were incised at different efficiencies, depending upon where the two damage sites were located in the sequence (Fig. 13). Specifically, FAAF di-adducts in *NarI*-G₂G₃ (100%) displayed fastest rate of incision, followed by *NarI*-G₁G₃ (85%) whereas *NarI*-G₁G₂ (44%) had a much lower rate of incision; G₂G₃ (100%) > G₁G₃ (85%) > G₁G₂ (44%) (Fig. 13).

DISCUSSION

Impact of sequence context on duplex structure and thermodynamics

The CD results indicate that the overall DNA structure was greatly influenced by how the two bulky FAAF lesions are arranged in the *NarI* sequence (Fig. 7a). In particular, base stacking represented by changes in the positive ellipticity at 270 nm was sequence dependent. While a significant decrease was seen in *NarI*-G₂G₃ and *NarI*-G₁G₃, an increase in intensity was noted for *NarI*-G₁G₂. Similar sequence effect was observed in the thermal and thermodynamic instability (Fig. 8, Table 1): *NarI*-G₂G₃ > *NarI*-G₁G₃ > *NarI*-G₁G₂. This trend is not surprising since base stacking is the hallmark of FAAF modified S-conformation and linked to their thermal and thermodynamic instability, especially in the fully paired duplexes.^{19, 20} Conversely, the stability of *NarI*-G₁G₂ is due to the higher B-conformer population at G₁ (46%) and G₂ (57%).

Assuming that each lesion can exist in both B and S conformation, there are 4 major conformational possibilities: BB, BS, SB, and SS. There will be numerous other possibilities in between. Consequently, ^{19}F NMR spectra (Fig. 9 and 10) of di-FAAF adducts are complex and we were unable to make specific conformational assignments. However, we reported previously a dominance of S-conformation (61%) when a FAAF is present at G_3 position of the *NarI* sequence. Similar S-conformer preference at the same *NarI* G_3 position was observed for N-deacetylated AF¹⁹ and related C8-dG adducts derived from the heterocyclic amines 2-amino-3-methylimidazo(4,5-f)quinolone (IQ)³⁷ and 2-amino-1-methyl-6-phenylimid-azo(4,5-b)pyridine (PHIP)³⁸. Interestingly, we found a greater thermodynamic instability for the G_3 containing *NarI*- G_2G_3 and *NarI*- G_1G_3 duplexes in the present study. It could be argued that *NarI*- G_1G_3 produces a greater S-conformation collectively [95% with G_1 (34%) and G_3 (61%)] than G_2G_3 (76%). However, this argument is not consistent with a greater thermodynamic stability observed for the former (Fig. 8, Table 1). It is possible that the closer proximity of two FAAF in *NarI*- G_2G_3 (e.g. just one base apart) compared to *NarI*- G_1G_3 (two bases) could possibly induces a greater DNA distortion. It should be noted, however, that the most thermodynamically stable *NarI*- G_1G_2 has no base in between the lesions. Thus, it appears that assessing the overall thermodynamic stability of di adducts is complicated by various conformational and configurational factors.

Conformation-specific Nucleotide Excision Repair

As shown in Figure 13, the *E. coli* UvrABC system exhibited approximately 1.5-2.0 folds better efficiency in the reparability of the di-adducts (G_2G_3 , G_1G_3 and G_1G_2) as compared

to mono-adducts (G_1 , G_2 and G_3) in the same *NarI* sequence. This is not surprising considering the extent of thermal/thermodynamic destabilization of the di- vs. mono-FAAF duplexes. It is well known that the NER efficiency is highly modulated by the damage recognition (UvrA₂) and verification/recognition (UvrB) steps. Jaciuk et al have shown that the UvrA₂ does not interact with DNA lesion directly, but senses the structural/conformational perturbations induced by lack of Watson-Crick hydrogen bonding and thermodynamic destabilization.²³ Moreover, the thermodynamic destabilization of the duplex assists UvrB in the insertion of β -hairpin into the duplex which is required for the lesion verification/recognition.²⁴ We reported that the repair of FAAF mono-adducts in *NarI* sequence occurs in a conformation-specific manner, i.e. the highly S/W-conformeric G_3 and G_1 duplexes produced greater DNA distortion and thermodynamic destabilization and thus incised more efficiently than the B-type G_2 duplex ($G_3 \sim G_1 > G_2$).¹⁶ Therefore, the presence of two FAAF in the same *NarI* sequence is expected to synergize destabilization effect. Indeed, as shown in Table 1, the destabilization effect of di-adducts ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ to 7.8 kcal/mol, $\Delta T_m = -10.0$ to -17.9 °C) is 1.5-2.0 folds greater than that of the mono-adducts ($\Delta\Delta G_{37^\circ\text{C}} = 3.7$ to 4.7 kcal/mol, $\Delta T_m = -5.3$ to -8.3 °C).¹⁶ These results are in contrast to recent reports on oxidative DNA damages where the reparability is compromised due to clustering, leading to enhanced genotoxicity.⁷ The rationale behind this dissimilarity could well be the complicated repair process comprising of base excision repair, non-homologous end joining and homologous recombination proteins executing the processing of oxidative cluster damages.

In addition, we observed different repair efficiency depending on the location of the two FAAF adducts (G_2G_3 , G_1G_3 and G_1G_2). The *NarI*- G_1G_2 duplex showed considerably lower efficiency than *NarI*- G_2G_3 and *NarI*- G_1G_3 [G_2G_3 (100%) > G_1G_3 (85%) > G_1G_2 (44%)] (Fig. 13). These NER results are in good agreement with the extent of destacking and thermal and thermodynamic destabilization ($G_2G_3 > G_1G_3 > G_1G_2$). The *NarI*- G_2G_3 duplex displayed the highest repair rate and also underwent the most severe base destacking (Fig. 7a) and duplex destabilization (*NarI*- G_2G_3 : $\Delta\Delta G_{37^\circ\text{C}} = 7.8$ kcal/mol, $\Delta T_m = -17.9$ °C). By contrast, *NarI*- G_1G_2 duplex showed the weakest reparability, consistent with the minimum base stacking (Fig. 7a) and duplex destabilization (*NarI*- G_1G_2 : $\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol, $\Delta T_m = -10$ °C). *NarI*- G_1G_3 showed an intermediary reparability and destacking (Fig. 7a)/destabilization (*NarI*- G_1G_3 : $\Delta\Delta G_{37^\circ\text{C}} = 6.8$ kcal/mol, $\Delta T_m = -14.1$ °C). These results confirm the sequence dependence repair of the di-adduct duplexes on duplex destabilization.

In summary, to the best of our knowledge the present study represents the first report on the structure-NER relations of any bulky cluster lesions. The repair of di-FAAF adducts was found to occur 2-fold more efficiently than the corresponding mono-adducts in the *NarI* sequence. The greater reparability of the di-FAAF adducts was found to be primarily due to their greater thermal and thermodynamic destabilization. We also found that the incision efficiency of the di-FAAF adducts occurred in a conformation-specific manner i.e. the highly syn-conformeric *NarI*- G_2G_3 and *NarI*- G_1G_3 are repaired more efficiently than the highly anti-conformeric *NarI*- G_1G_2 . Taken together, these present

results indicate the importance of carcinogen induced destacking and related thermal/thermodynamic destabilization in the NER of di-FAAF adducts.

Funding:

This work was supported by the National Institutes of Health (Grant number# CA098296 to BC and CA86927 to YZ); the National Science Foundation/RI-EPSCoR (0554548) and the RI-INBRE core facility supported by National Institutes of Health (P20 RR016457).

References

- (1) Boldt, J., Mah, M. C., Wang, Y. C., Smith, B. A., Beland, F. A., Maher, V. M. and McCormick, J. J. (1991) Kinds of mutations found when a shuttle vector containing adducts of 1,6-dinitropyrene replicates in human cells. *Carcinogenesis* 12, 119-126.
- (2) Burnouf, D., Koehl, P. and Fuchs, R. P. (1989) Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci. USA* 86, 4147-4151.
- (3) Yang, J. L., Maher, V. M. and McCormick, J. J. (1987) Kinds of mutations formed when a shuttle vector containing adducts of (+/-)-7 beta, 8 alpha-dihydroxy-9 alpha, 10 alpha-epoxy-7,8,9, 10-tetrahydrobenzo[a]pyrene replicates in human cells. *Proc. Natl. Acad. Sci. USA* 84, 3787-3791.
- (4) Ward, J. F. (1988) DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Prog. Nucleic Acid Res. Mol. Biol.* 35, 95-125.

- (5) Kalam, M. A. and Basu, A. K. (2005) Mutagenesis of 8-oxoguanine adjacent to an abasic site in simian kidney cells: tandem mutations and enhancement of G-->T transversions. *Chem. Res. Toxicol.* 18, 1187-1192.
- (6) Shikazono, N., Pearson, C., O'Neill, P. and Thacker, J. (2006) The roles of specific glycosylases in determining the mutagenic consequences of clustered DNA base damage. *Nucleic Acids Res.* 34, 3722-3730.
- (7) Magnander, K. and Elmroth, K. (2012) Biological consequences of formation and repair of complex DNA damage. *Cancer Lett.* 327, 90-96.
- (8) Regulus, P., Duroux, B., Bayle, P. A., Favier, A., Cadet, J. and Ravanat, J. L. (2007) Oxidation of the sugar moiety of DNA by ionizing radiation or bleomycin could induce the formation of a cluster DNA lesion. *Proc. Natl. Acad. Sci. USA* 104, 14032-14037.
- (9) Fuchs, R. P. (1984) DNA binding spectrum of the carcinogen N-acetoxy-N-2-acetylaminofluorene significantly differs from the mutation spectrum. *J. Mol. Biol.* 177, 173-180.
- (10) Roy, A. and Fuchs, R. P. (1994) Mutational spectrum induced in *Saccharomyces cerevisiae* by the carcinogen N-2-acetylaminofluorene. *Mol. Gen. Genet.* 245, 69-77.
- (11) Ross, M. K., Said, B. and Shank, R. C. (2000) DNA-damaging effects of genotoxins in mixture: modulation of covalent binding to DNA. *Toxicol. Sci.* 53, 224-236.
- (12) Mah, M. C., Boldt, J., Culp, S. J., Maher, V. M. and McCormick, J. J. (1991) Replication of acetylaminofluorene-adducted plasmids in human cells: spectrum

- of base substitutions and evidence of excision repair. *Proc. Natl. Acad. Sci. USA* 88, 10193-10197.
- (13) Wei, D., Maher, V. M. and McCormick, J. J. (1995) Site-specific rates of excision repair of benzo[a]pyrene diol epoxide adducts in the hypoxanthine phosphoribosyltransferase gene of human fibroblasts: correlation with mutation spectra. *Proc. Natl. Acad. Sci. USA* 92, 2204-2208.
- (14) Luch, A. (2005) Nature and nurture - lessons from chemical carcinogenesis. *Nat. Rev. Cancer* 5, 113-125.
- (15) Heflich, R. H. and Neft, R. E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutation Research/Reviews in Genetic Toxicology* 318, 73-174.
- (16) Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M. P. and Cho, B. P. (2012) Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.* 40, 3939-3951.
- (17) Patnaik, S. and Cho, B. P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem Res Toxicol* 23, 1650-1652.
- (18) Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health C. Environ. Carcinog. Ecotoxicol. Rev.* 22, 57-90.

- (19) 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 19F NMR, UV, and CD spectroscopy. *Biochemistry* 46, 13310-13321.
- (20) Meneni, S. R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M. P. and Cho, B. P. (2006) Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res.* 34, 755-763.
- (21) Meneni, S. R., Shell, S. M., Gao, L., Jurecka, P., Lee, W., Sponer, J., Zou, Y., Chiarelli, M. P. and Cho, B. P. (2007) Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. *Biochemistry* 46, 11263-11278.
- (22) Geacintov, N., E. , Broyde, S., Buterin, T., Naegeli, H., Wu, M., Yan, S. and Patel., D. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers* 65, 202-210.
- (23) Jaciuk, M., Nowak, E., Skowronek, K., Tanska, A. and Nowotny, M. (2011) Structure of UvrA nucleotide excision repair protein in complex with modified DNA. *Nat. Struct. Mol. Biol.* 18, 191-197.
- (24) Liu, Y., Reeves, D., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, M., Kolbanovskiy, A., Bolton, J. L., Broyde, S., Van Houten, B. and Geacintov, N. E. (2011) Probing for DNA damage with beta-hairpins: similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair (Amst)* 10, 684-696.

- (25) Min, J. H. and Pavletich, N. P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449, 570-575.
- (26) Mu, H., Kropachev, K., Wang, L., Zhang, L., Kolbanovskiy, A., Kolbanovskiy, M., N, E. G. and Broyde, S. (2012) Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.* 40, 9675-9690.
- (27) 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.
- (28) Belguise-Valladier, P. and Fuchs, R. 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.
- (29) 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.
- (30) Cho, B. P. and Zhou, L. (1999) Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by ¹⁹F NMR spectroscopy. *Biochemistry* 38, 7572-7583.
- (31) Gao, L., Zhang, L., Cho, B. P. and Chiarelli, M. P. (2008) Sequence verification of oligonucleotides containing multiple arylamine modifications by enzymatic

- digestion and liquid chromatography mass spectrometry (LC/MS). *J. Am. Soc. Mass Spectrom.* 19, 1147-1155.
- (32) Zhang, L. K. and Gross, M. L. (2000) Matrix-assisted laser desorption/ionization mass spectrometry methods for oligodeoxynucleotides: improvements in matrix, detection limits, quantification, and sequencing. *J. Am. Soc. Mass Spectrom.* 11, 854-865.
  - (33) Jain, N., Meneni, S., Jain, V. and Cho, B. P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.* 37, 1628-1637.
  - (34) Luo, C., Krishnasamy, R., Basu, A. K. and Zou, Y. (2000) Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, N-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease. *Nucleic Acids Res.* 28, 3719-3724.
  - (35) Zou, Y., Liu, T. M., Geacintov, N. E. and Van Houten, B. (1995) Interaction of the UvrABC nuclease system with a DNA duplex containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE. *Biochemistry* 34, 13582-13593.
  - (36) Zou, Y. and Van Houten, B. (1999) Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J.* 18, 4889-4901.
  - (37) 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 NarI recognition sequence. *Chem. Res. Toxicol.* 20, 445-454.

- (38) Brown, K., Hingerty, B. E., Guenther, E. A., Krishnan, V. V., Broyde, S., Turteltaub, K. W. and Cosman, M. (2001) Solution structure of the 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine adduct in duplex DNA. *Proc. Natl. Acad. Sci. USA* 98, 8507-8512.

Figure legends

Figure 1: Adduct and sequence used in this study. (a) Chemical structures of AAF and FAAF adducts; (b) fully paired 16-mer *NarI* duplexes used in the present study; (c) major groove views of the central trimer segments of the B/S/W-conformer equilibrium of FAAF-modified mono-adduct duplexes. The modified dG and the complementary dC are shown in red and green sticks, respectively, and the aminofluorene moiety is highlighted with shiny grey CPK and the *N*-acetyl with pink CPK. In the B-type conformer, anti-glycosidic FAAF-dG maintains Watson-Crick hydrogen bonds, thereby placing the carcinogen moiety in the major groove. The carcinogens in the S- and W-conformers stack into the helix or wedged into the minor groove, respectively, with the modified dG in the syn conformation.

Figure 2: HPLC profile of FAAF-modified ODN. (a) Chromatogram of a reaction mixture between 16-mer *NarI* sequence (5'-CTCTCG₁G₂CG₃CCATCAC-3') and an activated FAAF (N-acetoxy-N-2-(acetilamino)-7-fluorofluorene). The mono- (G₁, G₃, G₂), di- (G₁G₃, G₂G₃ and G₁G₂), and tri-FAAF (G₁G₂G₃) adducts eluted in the 28-35, 42-60, and 84 min ranges were purified by reversed-phase HPLC (see materials and method for gradient condition). (b) on-line photodiode array UV/Vis spectra of mono-, di- and tri-FAAF adducts.

Figure 3: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 4 (see Fig. 2) (a) the 3'- exonuclease digestion spectrum at 30 seconds (bottom), 5 (middle) and 15 (top) minutes time points; (b) the 5'-exonuclease digestion spectrum at 5 (bottom), 30 (middle) and 80 (top) minutes time points. Insets

show the theoretical molecular weights of the fragments that should form after the 3'- and 5'-exonuclease digestion of peak 4 if ODN is modified at G₁ and G₃ positions.

Figure 4: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 4 (a) the 5'- exonuclease digestion spectrum and (b) the 3'- exonuclease digestion spectrum at different time points. Left side of the figure shows the theoretical molecular weights of the fragments that should form after the 3'- and 5'- exonuclease digestion of peak 4 if ODN is modified at G₁ and G₃ positions.

Figure 5: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 5 (a) the 5'- exonuclease digestion spectrum and (b) the 3'- exonuclease digestion spectrum at different time points. Left side of the figure shows the theoretical molecular weights of the fragments that should form after the 3'- and 5'- exonuclease digestion of peak 5 if ODN is modified at G₂ and G₃ positions.

Figure 6: MALDI-TOF mass spectrum of the hydrolysis fragments formed from exonuclease digestion of peak 6 (a) the 5'- exonuclease digestion spectrum and (b) the 3'- exonuclease digestion spectrum at different time points. Left side of the figure shows the theoretical molecular weights of the fragments that should form after the 3'- and 5'- exonuclease digestion of peak 6 if ODN is modified at G₁ and G₂ positions.

Figure 7: CD and UV-melting profiles. (a) CD spectral overlay recorded at 20 °C and (b) UV-melting curves in 20 mM phosphate buffer containing 0.1 M NaCl at pH 7 of fully paired 16-mer *NarI* di-adduct duplexes with FAAF modification at G₁G₂ (green), G₁G₃ (blue), and G₂G₃ (red).

Figure 8: Comparative thermodynamic parameters histogram of *NarI* di-adduct duplexes; G₁G₂ (green), G₁G₃ (blue), and G₂G₃ (red). The $\Delta\Delta$ values represent; $\Delta\Delta H =$

ΔH (modified duplex) - ΔH (control duplex), $T\Delta\Delta S = T\Delta S$ (modified duplex) - $T\Delta S$ (control duplex), $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).

Figure 9: ^{19}F NMR spectra of FAAF-modified di-adduct duplexes; *NarI*-G₁G₃ (left), *NarI*-G₂G₃ (middle), and G₁G₂ (right) at different time intervals.

Figure 10: Comparison of ^{19}F NMR spectra of FAAF-modified di-adduct duplexes with their respective mono-adducts; (a) *NarI*-G₂G₃ with *NarI*-G₂ and -G₃, (b) *NarI*-G₁G₃ with *NarI*-G₁ and -G₃ and (c) *NarI*-G₁G₂ with *NarI*-G₁ and -G₂ at 5°C (left) and 20 °C (right).

Figure 11: Imino proton spectra of FAAF-modified di-adduct duplexes; *NarI*-G₁G₃ (bottom), *NarI*-G₂G₃ (top), and G₁G₂ (middle) at 20°C.

Figure 12: The 5'-terminally labeled FAAF modified *NarI* di-adduct substrates (2 nM) were incubated with UvrABC (UvrA, 10 nM, UvrB, 250 nM, and UvrC, 100 nM) in UvrABC reaction buffer at 37°C for the different time periods. The incision products were then analyzed on a 12% polyacrylamide sequencing gel under denaturing condition. The 55-mer represents the intact DNA substrates, and the incision products represent the 5'-incised DNA fragments. The results are tabulated and relative incision efficiencies are plotted in Figure 13.

Figure 13: UvrABC nuclease incision kinetic results. a) comparative relative incision rate histogram of FAAF modified mono- and di-adducts located at different positions calculated by considering *NarI*-G₂G₃ FAAF incision rate as 100%. b) table showing the absolute incision rate of three di-adduct duplexes.

Table legend

Table 1: Thermal and thermodynamic parameters of FAAF modified *NarI* di-adduct duplexes obtained from UV-melting experiments.

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

Figure 8:

Figure 9:

Figure 10:

Figure 11:

Figure 12:

Figure 13:

Duplex	<i>NarI</i> -G ₁ G ₃	<i>NarI</i> -G ₂ G ₃	<i>NarI</i> -G ₁ G ₂
Incision rate	0.344±0.013	0.406±0.020	0.178±0.008

Incision rate (fmol/min)

Table 1:

5'-CTCTCG ₁ G ₂ CG ₃ CCATCAC-3' 3'-GAGAGC C GCGGTAGAG-5'								
	$-\Delta H$ kcal/mol	$-\Delta S$ eu	$-\Delta G_{37^\circ C}$ kcal/mol	T_m^b °C	$\Delta\Delta H^c$ kcal/mol	$\Delta\Delta S^d$ eu	$\Delta\Delta G_{37^\circ C}^e$ kcal/mol	ΔT_m^f °C
Control ^a	121.9	324.7	21.2	70.6	-	-	-	-
NarI-G ₁ G ₃ - FAAF ^a	97.8	268.7	14.4	56.5	24.1	56	6.8	-14.1
NarI-G ₂ G ₃ - FAAF ^a	98.6	274.6	13.4	52.7	23.3	50.1	7.8	-17.9
NarI-G ₁ G ₂ - FAAF ^a	115.7	318.7	16.9	60.6	6.2	6	4.3	-10

a) The average standard deviations for $-\Delta G$, $-\Delta H$, and T_m are ± 0.4 , ± 3.0 , and ± 0.4 , respectively.

b) T_m values is the temperature at half the peak area.

c) $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex).

d) $\Delta\Delta S = \Delta S$ (modified duplex) - ΔS (control duplex).

e) $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).

f) $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Manuscript-III

Published in *Nucleic Acids Research*, 2013, 41, 869-880

Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor

Vipin Jain¹, Benjamin Hilton², Bin Lin³, Satyakam Patnaik¹, Fengting Liang¹, Eva Darian³, Yue Zou², Alexander D. MacKerell Jr.³, and Bongsup P. Cho¹

¹Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI, 02881, USA

² Department of Biomedical Sciences, East Tennessee State University, Johnson city, TN, 37614, USA

³Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, 21201, USA

*To whom correspondence should be addressed:

Telephone#: (+1) 401-874-5024; FAX# (+1) 401-874-5766; E-mail: bcho@uri.edu

Present address:

Department of Biomedical and Pharmaceutical Sciences,
University of Rhode Island
7 Greenhouse Road,
Kingston, RI, 02881, USA

Abstract

The environmental arylamine mutagens are implicated in the etiology of various sporadic human cancers. Arylamine-modified dG lesions were studied in two fully paired 11-mer duplexes with a -G*CN- sequence context, in which G* is a C8-substituted dG adduct derived from fluorinated analogs of 4-aminobiphenyl (FABP), 2-aminofluorene (FAF), or 2-acetylaminofluorene (FAAF), and N is either dA or dT. The FABP and FAF lesions exist in a simple mixture of “stacked” (S) and “B-type” (B) conformers, whereas the N-acetylated FAAF also samples a “wedge” (W) conformer. FAAF is repaired 3-4 times more efficiently than FABP and FAF. A simple A- to -T polarity swap in the G*CA/G*CT transition produced a dramatic increase in *syn*-conformation and resulted in 2~3-fold lower NER efficiencies in *Escherichia coli*. These results indicate that lesion-induced DNA bending/thermodynamic destabilization is an important DNA damage recognition factor, more so than the local S/B-conformational heterogeneity that was observed previously for FAF and FAAF in certain sequence contexts. The present work represents a novel 3'-next flanking sequence effect as a unique nucleotide excision repair factor for bulky arylamine lesions in *E. coli*.

Key Words: DNA adducts, thermodynamics, conformational heterogeneity, flanking sequence, ^{19}F NMR

Abbreviations: FAAF, *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene; FAF, *N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene; FABP, *N*-(2'-deoxyguanosin-8-yl)-4-fluoro-4-aminobiphenyl; DSC, differential scanning calorimetry; ^{19}F NMR, fluorine

nuclear magnetic resonance; ICD, induced circular dichroism; EMSA, electrophoretic mobility shift assay; MD, molecular dynamics; PMF, potential of mean force.

Introduction

Structural and conformational damage in specific areas of the genome can trigger tumorigenesis. For example, disruption of a gene that encodes the tumor suppressor p53 protein has been found in the majority of sporadic human cancer (1). While human cells are equipped with repair pathways to safeguard the genome from various DNA damage, some lesions may go unrepaired, thereby serving as a faulty template to produce a complex array of mutations and genomic instability, ultimately leading to cancer initiation (2).

Aryl- and heterocyclic-amines are well known environmental mutagens/ carcinogens which have been implicated in the etiology of breast, liver, and bladder cancers in humans (3). Metabolic activation of these amines *in vivo* produces C8-substituted dG as the major DNA adducts (4). For example, the human bladder carcinogen 4-aminobiphenyl produces ABP (Fig. 1a). Similarly, AF and AAF are the major DNA adducts derived from 2-aminofluorene, 2-nitrofluorene, and 2-acetylaminofluorene (Fig. 1a). The ABP and AF adducts in fully-paired duplex DNA have been shown to adopt an equilibrium of two prototype conformers: “B-type”, in which the carcinogen resides in the major groove of a relatively unperturbed double helical DNA; and “stacked (S)”, in which the carcinogen is base-displaced and the glycosidic linkage to the modified guanine is *syn* (Fig. 1c) (5,6). The aromatic moieties of ABP are not coplanar as in AF, which results in a much lower S-state population than AF. AF-induced S/B-heterogeneity is dependent on the flanking sequence, which modulates mutational and repair outcomes (6,7). AAF is chemically identical to AF except for a single acetyl group

on the central nitrogen (Fig. 1a), leading to sampling of an additional W-conformation, in which the fluorene moiety is in the minor groove along with a *syn* glycosidic linkage (Fig. 1c) (7,8). The B and S conformers exhibited by AAF are similar to those obtained for ABP and AF.

Nucleotide excision repair (NER) is the major cellular pathway for removing bulky DNA lesions in cells. Accumulated evidence suggests that efficiency of NER is governed by various structural, cellular, and biological factors (9-11). Sequence context, in particular, plays an important role in NER of bulky DNA lesions (7,12). The most notable sequence effects were observed in the *NarI* sequence (5'-...CG₁G₂CG₃CC...-3') which is well-known for inducing higher frequency of -2 deletion mutations when adducted by AAF at G₃ position despite the similar chemical reactivities of three guanines (13,14). Fuchs and coworkers have shown that AAF in duplex is an excellent substrate for *E. coli* UvrABC and human exonuclease repair systems (15-18). They reported that in *E. coli* the relative repair efficiencies of AAF at G₁, G₂ and G₃ were in a ratio of 100:18:66, respectively whereas the human exonuclease exhibited 38:100:68 ratio (17,18). Mu et al (19) have recently carried out a human NER study of these lesions in HeLa cell extracts and found similar sequence dependent NER efficiencies. Their molecular dynamics (MD) simulation data indicated that the greater NER efficiencies are correlated with base sequence-dependent local untwisting and minor groove opening together with weaker stacking interactions (19). We recently conducted *E. coli* UvrABC NER studies on the *NarI* sequence duplexes (5'-G₁G₂CG₃CC--3'), in which guanines are modified by either AF or AAF (7). Results showed that the bulky AAF adducts repair in a conformation-

specific manner, with the highly S/W-conformeric G_3 and G_1 duplexes incised considerably more efficiently than the highly B-conformeric G_2 duplex ($G_3 \sim G_1 > G_2$). On the other hand, the repair rate of *N*-deacetylated AF was 2- to 3-fold lower than AAF, and the order of incision efficiencies was opposite of that observed for the AAF case. We have coined the term “*N*-acetyl factor” to describe the complexity of NER recognition of AF vs. AAF (7).

Here, we describe an unusual 3'-next flanking base effect on the conformational properties and *E. coli* NER efficiencies of three prototype arylamine adducts in the G^*CN sequence context (Fig. 1b: $G^* =$ ABP, AF, or AAF; $N =$ A or T). Results from spectroscopy (^{19}F NMR and ICD), thermodynamic quantification (DSC/UV), and gel electrophoresis, as well as molecular dynamics/potential of mean force (MD/PMF) calculations show that sequence-dependent lesion-induced DNA bending coupled with thermodynamic destabilization is responsible for the altered repair recognition of bulky arylamine-DNA adducts in *E. coli*. The present work represents a novel 3'- next flanking sequence effect as a unique nucleotide excision repair factor for bulky arylamine lesions in *E. coli*.

Materials and Methods

Adduct Synthesis:

Modified duplexes were prepared following the published procedures (7,8,20-23). The modified oligos were characterized by ESI-TOF-MS analysis as reported previously (24). An identical set of unmodified duplexes was also prepared as controls.

Differential Scanning Calorimetry (DSC):

DSC measurements were performed using a VP-DSC Micro-calorimeter from Microcal Inc. (Northampton, MA) according to the procedures published previously (22). All sample solutions were 0.12 mM concentration. T_m was the temperature at half the peak area. ΔG and ΔS values were determined by the procedures of Chakrabarti et al. (25). The uncertainties in the values of T_m , ΔH , ΔG and ΔS represent the random errors inherent in the DSC measurements.

UV-Melting (Cary100 Bio, Beckman) and **CD** (J-810, Jasco) experiments were performed using the previously reported procedures (7,20,23,26).

Dynamic ^{19}F NMR:

Duplex samples (about 20~30 ODS) were dissolved in 300 μL of pH 7.0 buffer (100 mM NaCl, 10 mM Na_3PO_4 , and 100 μM EDTA in 10% D_2O /90% H_2O) and filtered into through a Shigemi tube using a 0.2 μm membrane filter. All ^1H and ^{19}F NMR experiments were conducted using a dedicated 5-mm $^{19}\text{F}/^1\text{H}$ dual probe on a Bruker DPX400 Avance spectrometer operating at 400.0 and 376.5 MHz, respectively. Imino proton spectra were obtained using phase sensitive jump-return sequences at 5 $^\circ\text{C}$. ^{19}F NMR spectra were acquired in the ^1H -decoupled mode and referenced to CFCl_3 by assigning external C_6F_6 in C_6D_6 at -164.90 ppm. Temperature dependence spectra were processed as previously reported (20,27).

EMSA assay:

The FABP, FAF, and FAAF-modified 19-mer GCT and GCA sequences were each (100 nM) annealed with an equimolar complementary sequence, in which the 5'-end was γ - ^{32}P -labeled using T4 polynucleotide kinase and [γ - ^{32}P] ATP (Perkin-Elmer radiochemical, Boston, MA) in a buffer containing NaCl (25 mM) and Tris/HCl (25 mM). The mixture was heated at 95 °C for 5 minutes and then cooled to room temperature overnight. The duplexes were subjected to 15% non-denaturing polyacrylamide (acrylamide:bisacrylamide; 29:1, w/w) gel electrophoresis at 1,800 V and the temperature was maintained at 4-8 °C by regularly replacing the running buffer with the ice-cold TBE buffer. Gel was exposed to Kodak phosphor imaging screen overnight and scanned on Typhoon 9410.

Nucleotide Excision Assay:

DNA substrates of 58 bp containing a single FABP, FAF, or FAAF, each adducted at either G*CT or G*CA sequences were constructed as previously described (28,29). UvrA, UvrB, and UvrC proteins were over expressed in *E. coli* then purified as previously described (30). The 5'-terminally labeled DNA substrates were incubated and incised by UvrABC as previously described (28,29). Briefly, the DNA substrates (2 nM) were incubated in the UvrABC reaction buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM MgCl_2 , 5 mM DTT) at 37°C in the presence of UvrABC (UvrA, 10 nM, UvrB, 250 nM, and UvrC, 100 nM). The Uvr proteins were diluted and premixed in Uvr storage buffer before addition to the reaction. Aliquots were collected at 0, 5, 10, 15, and 20 minutes into the reaction. The reaction was terminated by heating at 95°C for five

min. The products were denatured by addition of formamide loading buffer and heating to 95°C for five min, followed by quick chilling on ice. The incision products were then analyzed by electrophoresis on a 12% polyacrylamide sequencing gel under denaturing conditions with TBE buffer.

To quantify the incision products, radioactivity was measured using a Fuji FLA-5000 Image Scanner with MultiGauge V3.0 software. The DNA incised (in fmol) by UvrABC was calculated based on the total molar amount of DNA used in each reaction and the ratio of the radioactivity of incision products to total radioactivity of DNA. At least three independent experiments were performed for determination of the rates of incision.

MD and PMF Calculations:

PMF calculations were performed on the GCA and GCT 11-mers initiated from the canonical B form of DNA for “anti” simulations where the glycosidic bond is in the *anti* form. “Syn” simulations were initiated from models based on an NMR structure (PDB: 1C0Y) in which the glycosidic bond is in the *syn* form. MD simulations were performed with the programs CHARMM and NAMD, using the CHARMM27 additive nucleic acid force field. Modified G* lesions were created based on the CHARMM General Force Field followed by additional optimization of the dihedral parameters linking the G base to the adduct. Determination of the PMFs followed the protocol of Banavali and MacKerell (31) with details of the simulations included in the supporting information.

Results

Model Systems

As model systems, 11-mer DNA duplexes [d(5'-CCATCG*CNACC-3').d(5'-GGTNGCGATGG-3')] were prepared, in which G* is FABP, FAF or FAAF and N is either dA or dT (designated as G*CA and G*CT duplexes, respectively) (Fig. 1b). The two sequences are chemically isomeric, differing only on the polarity of the 3'-next flanking A:T vs. T:A. The utility of fluorine-tagged lesions as effective structure probes has been documented (32). Both the G*CA and G*CT sequences have been used previously for studies of bulky adducts (22,33).

¹⁹F NMR Spectroscopy

Figure 2a-c compares the ¹⁹F NMR spectra (-114 ~ -121 ppm) of modified DNA duplexes at 20°C for the G*CT and G*CA sequence contexts (see Fig. S1 for full temperature ranges). ¹⁹F signal assignments were made based on the H/D solvent effect, exchange spectroscopy, adduct-induced CD (ICD_{290-350nm}), and chemical shifts as previously described (6,26,32,34).

FABP-Duplexes: A clear conformational difference exists between the two isomeric FABP-modified G*CA and G*CT duplexes (Fig. 2a). The single signal at -116.9 ppm for FABP-G*CA has been previously assigned to the B-conformer (22). In contrast, FABP-G*CT exhibited two signals at -116.9 (B) and -118.0 (S) ppm in a 40:60% ratio and adopted a two-site exchange (EXSY spectra at 5 and 17°C, inset, Fig. S1a). A large chemical shift gap (~1 ppm) between the two signals suggests differences of their

electronic environments. In addition, the -116.9 ppm signal revealed a large H/D effect (+0.24 ppm) compared to the -118.0 ppm signal (+0.14 ppm)(data not shown) upon increasing the D₂O content from 10 to 100%. The results indicate the exposed fluorine atom in the B-conformer, as observed in the MD/PMF simulations (Fig. S2).

FAF-Duplexes: Though not as dramatic, a similar sequence effect was observed for FAF. The FAF-G*CA duplex showed signals at -117.4 and -118.6 ppm in a 34:66% ratio (Fig. 2b), which have been assigned to B- and S-conformers, respectively (20,22,23,26,35). The S conformer population was increased to 90% in the FAF-G*CT duplex. Consistent with this assignment, the downfield -117.4 ppm signal revealed a larger H/D effect (+0.19 ppm) compared to the -118.8 ppm signal (+0.03 ppm) (data not shown), again consistent with the MD/PMF simulations (Fig. S2).

FAAF-Duplexes: At least three major signals are present (Fig. 2c) for FAAF in the G*CA and G*CT sequences. These signals are shifted considerably (~2 ppm) to the downfield compared to FAF, a phenomenon associated with the *N*-acetyl factor. We reported previously S/B/W-conformer assignment of FAAF-modified 12-mer and 16-mer duplexes in several sequence contexts (i.e., TG*A, CG*C, CG*G, and GG*C)(7,8). The results revealed that ¹⁹F signals of B-, S-, and W-conformers appear consistently going from downfield to upfield in order of -115.0 ~ -115.5 ppm, -115.5 ~ -117.0 ppm and -116.5 ~ -118.0 ppm, respectively. The major ¹⁹F signals in Fig. 2c fit that pattern. In particular, the signal patterns of the FAAF-modified G*CA/G*CT duplexes (5'-CCATCG*CNAC-3') match well with those observed for 16-mer (5'-

CTCTCG₁G₂CG₃*CCATCAC-3') (7) and 12-mer (5'-CTTCTCG*CCCTC-3') duplexes (8), both of which have the CG*C sequence context (underlined). In line with this observation, their proton spectra displayed a mixture of broad imino signals arising not only from those involved in Watson-Crick hydrogen bonds (12 ~ 14 ppm), but also from the lesion site and its vicinity (11-12 ppm)(Fig. S3). While the two sequences are similar in the total syn conformation (S+W)(i.e., 83 vs. 78%, respectively for G*CT and G*CA), the W-conformer population appears to be significantly greater for G*CT (30%) compared to G*CA (14%)(Fig. S4).

Induced Circular Dichroism (ICD)

Figure 2d-f show the CD of the modified G*CA and G*CT duplexes. We reported that B- and S-conformers are characterized by positive and negative ICD_{290-350 nm}, respectively (26). Accordingly, the B-conformeric FABP-G*CA displayed a strongly negative ICD_{290-350 nm}, whereas an S-shape curve was observed for the S/B-mixture G*CT duplex (Fig. 2d). These results are in good agreement with the ¹⁹F NMR results (Fig. 2a). Unlike FABP, FAF on both sequences exhibited a strong positive ICD_{290-350 nm} with the effect much greater for G*CT (Fig. 2e), consistent with the greater S-conformer population determined by ¹⁹F NMR (Fig. 2b). The ICD of FAAF (Fig. 2f), which is confined in the narrow 290-320 nm range, has not been defined as clearly as FAF (8).

In addition, the modified duplexes displayed significant blue shift relative to their respective control duplexes (Fig. S5a,b and Table 1). All except for FAF-G*CT exhibited significant blue shifts up to 8 nm. The bulky N-acetylated FAAF exhibited greater shifts

($\Delta_{G^*-G} = 4 \sim 8$ nm) than FAF and FABP ($\Delta_{G^*-G} = 0 \sim 4$ nm). GCA sequences, which are prone to the B-conformer, displayed greater blue shift ($\Delta_{G^*CA-G^*CT} = 2 \sim 4$ nm, Table 1) compared to GCT. It is well known that protein-induced DNA bending exhibits significant CD shift at 275 nm of regular B-type DNA (36-38). For instance, the HMG box protein SOX-5 bends DNA by $\sim 74^\circ$ upon binding, which resulted in a significant blue CD shift (37). These reports suggest that the blue shifts observed in the present study result from the distortion of the DNA backbone, particularly bending.

Gel Mobility Assay

Two 19-mer sequences (5'-CTTACCATCGCNACCATTTC-3', $N = T$ or A) were used to investigate the impact of the A/T polarity swap at N position on the gel mobility of the modified duplexes. Initially, the 11-mer sequences above were used but they denatured in the 15% native polyacrylamide gel at 1,800 V (data not shown). Figure 3 compares the electrophoretic mobility of the 19-mer GCA and GCT sequences with and without modifications. Differential mobility between the single strand (ss)- and double strand (ds)-oligonucleotides confirmed the integrity of the duplexes (Fig. 3). The modified duplexes exhibited retardation in the mobility in a lesion-dependent manner. In both sequences, major retardation effect was observed for FAAF followed by FABP, whereas no retardation was observed for FAF, results consistent with the CD blue shift data above (Table 1). It should be noted that the magnitude of retardation in mobility observed in the present study is significantly lower than what was previously observed in benzo[a]pyrene modified duplexes (39-41). We cannot rule out a possibility of C8-substituted dG (present study) vs. N²-substituted dG (benzo[a]pyrene) binding.

However, the rationale behind this small difference in mobility could well be due the significantly longer oligonucleotides (19-mer) used in the present study as opposed to the aforementioned benzo[a]pyrene cases (11-mer and 15-mer). The utilization of longer sequences will reduce the number of adducts per helical turn which might result into diminishing of the lesion-induced bending effect. Tsao et al reported similar effects of oligonucleotide length on the electrophoretic mobility of benzo[a]pyrene modified duplexes (41).

Thermodynamics

Thermodynamics results from UV-optical melting (Table S1) and differential scanning calorimetry (DSC), which is not dependent on melting patterns and stoichiometry (22) (Table 2), are comparable. Figure S6 shows the DSC thermograms of modified duplexes in the G*CA and G*CT sequences relative to the unmodified controls. These curves were transformed into the corresponding thermodynamic charts (Fig. 4a,b) and the results are tabulated in Table 2. According to the NMR results (Fig. 2), FABP and FAF display a S/B-equilibrium, whereas FAAF produces a complex S/B/W-equilibrium: thus, they will be compared separately.

FABP/FAF-DNA Duplexes: Both FABP and FAF resulted in destabilization (Fig. 4, Table 2). FABP reduced T_m for the G*CA and G*CT duplexes by -10.2 and -11.2 °C and $\Delta\Delta G_{37^\circ\text{C}}$ by 2.4 and 2.9 kcal/mol, respectively. The G*CA/G*CT transition produced major effect on $\Delta\Delta H$ (2.7 vs. 10.8 kcal/mol) and $\Delta\Delta S$ (0.8 vs. 25.6 eu), consistent with a significant increase in the S-conformer population from 0 to ~60% (Fig. 2a). The B-

conformer FABP-G*CA is expected to lead to small entropy compensation, and consequently the enthalpy reduction dominated the free energy destabilization (Fig. 4a). As expected, the structural disturbance caused by the S/B mixture FABP-G*CT duplex leads to a considerable reduction of melting enthalpy, however, most of it is compensated by entropy (Fig. 4b).

FAF modification resulted in a similar destabilization effect: ΔT_m by -9.1 and -9.6 °C and $\Delta\Delta G_{37^\circ\text{C}}$ by 2.7 and 2.9 kcal/mol, respectively, for G*CA and G*CT. However, compared to FABP, FAF in both sequences yielded significantly larger reduction in enthalpy ($\Delta\Delta H = 13.4$ and 16.4 kcal/mol) and entropy ($\Delta\Delta S = 34.2$ and 43.8 eu) compensation (Fig. 4 and Table 2). FAF exhibits more S-conformer than FABP in both sequences because of stronger stacking effect. As a result, sequence dependence on the thermodynamics was not as dramatic as in FABP. It is clear from Fig. 4 that FAF (over FABP) and G*CT (over G*CA) combinations produce consistently greater enthalpy/entropy compensation, *i.e.*, $\text{FAF/G*CT} > \text{FAF/G*CA} > \text{FABP/G*CT} > \text{FABP/G*CA}$ from the highest to the lowest. As expected the *N*-deacetylated FABP-G*CT and FAF-G*CA exhibited two site exchange (Table S2)

FAAF-Duplexes: FAAF modification resulted in the most significant reduction of ΔT_m by -21.8 and -23.3 °C and $\Delta\Delta H$ by 39.4 and 41.1 kcal/mol, respectively for G*CA and G*CT sequence contexts (Fig. 4a, 4b and Table 2). This is due to the bulky acetyl group in FAAF (Fig. S2). Like FABP and FAF, however, entropy compensation contributed a stabilization, *i.e.*, $\Delta\Delta G = 5.6$ and 5.4 kcal/mol, respectively, for the G*CA and G*CT.

Molecular Dynamics (MD)/Potential Mean Force (PMF) Calculations

To further understand the impact of lesion modification on the G*CA and G*CT duplexes, MD simulations were performed in combination with PMF calculations. PMF calculations yield the free energy as a function of the extent of flipping of the modified G* base (Fig. S7). Individual PMFs were determined with G* in either the *anti* or *syn* orientations, while only the *anti* orientation was studied for the unmodified duplexes.

Figure 5 shows the free energies from the PMFs. In the unmodified and *anti*-G* PMFs there is a deep minimum in 330 ~ 360° corresponding to the WC-base paired state (31), which in case of lesions corresponds to the B-state. Deep minima are also present in 330 ~ 360° in the *syn*-G* PMFs, which corresponds to the S-state. The conformer assignments were made based on published experimental NOE data (Fig. S8-S11, Tables S3-S6) (42-45). The relative energies of the flipped states are highest in the unmodified duplexes in all cases indicating that the lesions lower the relative energies of the flipped state and/or destabilize the low energy B- or S-states. Notably, the free energy surfaces show the relative energies of the flipped states to be lower in the *syn* PMFs, consistent with the conformational thermodynamic data discussed above. As such, the lowest energies of the flipped state occur with FAAF (Table 2) and moreover, that the *syn* FAAF may sample a wider range of conformations as compared to FABP and FAF, consistent with the ¹⁹F NMR data (Fig. 2). Further validation of the PMFs is the energies of the minima being about 15 kcal/mol, which is in good agreement with the experimental $\Delta G^\ddagger$ of 14.1 kcal/mol required for a B/S-conversion. Representative B/S/W-structures from the NOE-based PMFs are shown in Figures S12 to S14. For all three lesions the presence

of WC base pairing in the B-state and the stacking of the adduct into the duplex in the S-state is evident. In contrast, the structure of the W-state, which is only populated by FAAF, is highly distorted.

In the *anti*-G* PMFs, the average solvent accessible surface areas (SASA) of both the lesion and fluorine atom are high in the vicinity of the B state, consistent with the location adduct in the major groove (Fig. S2). For the *syn* PMFs, the SASA values are low in the regions corresponding to the S-state, consistent with adducts being stacked inside the helices. However, the SASA values are higher with FAAF as compared to FABP and FAF, suggesting an altered environment for FAAF. In addition, the *syn* PMFs of FAAF exhibits SASA minima in 60 ~ 120°, which encompasses the W state. These suggest a fundamental difference in conformational properties of FAF/FABP vs. FAAF, consistent with the significant difference in the ¹⁹F chemical shifts for the FAAF species as well as thermodynamic data.

Shown in Figure S15 are bending probability distributions for the B-, S- and W-states for the three lesions in both the G*CA and G*CT contexts. In the B- and S-states the extent of bending is significantly larger with FAAF (cyan) versus FABP (red) and FAF (blue). These results are consistent with the experimental data obtained from the greater blue shift in CD (Fig. S5, Table 1), although the changes in G*CA occur only in the S-state (Fig. S15). In addition, the simulations indicate the extent of bending for FABP to be similar to that of FAF. The significant increase in bending in FAAF is consistent with the greater destabilization of the duplexes caused by the FAAF lesion (Table 2).

Concerning the bending, calculation of local helicoidal parameters revealed significant differences in twist and tilt for base pair 8, where the A/T switch occurs (Table S7). For example, twist values are systematically larger and tilt values less negative in GCT versus GCA sequences. These differences suggest that the local structural alteration associated with the A/T switch is being propagated to the overall helix.

***E. coli* UvrABC incision**

DNA substrates containing lesions in the defined sequences were subjected to incisions by the *E. coli* UvrABC system in a kinetic assay. These substrates were radioactively labeled at the 5'-end of the adducted strand and the major incision products separated on a urea-PAGE gel running under denaturing conditions (Fig. S16). The incision occurred at the 8th phosphate bond 5' to G*, which is consistent with the currently accepted mechanisms of UvrABC-based NER (28,29). The substrates were incised at differing efficiencies depending on not only the type of DNA adduct but also the sequence context (Fig. 6). Specifically, the G*CA sequences were incised at higher rates by ~ 2 fold than G*CT, while the order of incision rate of adducts is: FAAF > FAF $\approx$ FABP for both sequences, with FAAF being incised with 2-3 folds greater efficiency than the other lesions.

It should be noted that the 5'-incision products appeared as doublet bands (Fig. S16). Similar incision products of this type of lesion have been observed previously (35,46,47). This is likely either due to the type of arylamine lesions or to the structural heterogeneity exhibited by this type of lesions as demonstrated in the present and previous studies,

suggesting that UvrABC may make the 5'-incision at the site different by one nucleotide for the different conformers of arylamine lesion.

DISCUSSION

Conformational and Thermodynamic Effects on the G*CA/G*CT Transition

The NMR/ICD results indicate that lesion stacking is affected considerably by a polarity swap at the 3'-next flanking base (GCA → GCT). The effect was most significant for FABP, which produced a dramatic increase in S-conformer (0% to 60%) (Fig. 2a). *This is an extraordinary DNA sequence effect.* A similar trend was observed for FAF, although the S-conformer was only 24% greater in G*CT than in G*CA (Fig. 2b). Unlike FABP and FAF, the impact of the A/T swap on FAAF was minimal; specifically, the *syn*-glycosidic S- and W-conformers remained relatively unchanged (~78% to ~83%)(Fig. 2c). Interestingly, the increase of W-conformer (14% to 30%) appeared to be compensated by a concomitant decrease of S-conformer (64% to 53%). This data indicates that the *N*-acetyl group in FAAF can push the low energy *syn*-S/W-equilibrium towards W (see “*N*-acetyl factor” below). Overall, these results indicate that the A/T swap has the largest impact on the most stable system, while the least stable FAAF lesion is the least impacted.

As expected, all modified duplexes were consistently destabilized as compared to the controls (Fig. 4, Table 2): FAAF > FAF ≈ FABP. The G*CA/G*CT transition led to further destabilization, which was associated with increases in lesion stacking (greater S/W) for all three lesions. Obviously, a higher population of the *syn*-S-/W-conformer

states is expected to disrupt the double helical structure which would significantly reduce the enthalpy, accompanied by a compensatory increase in entropy (22).

Lesion-Induced DNA Bending as a Major NER Recognition Factor

For each lesion, a greater proportion of B-conformer was observed in G*CA (FABP: 100%, FAF: 34%, FAAF: 22%) than G*CT (FABP: 40%, FAF: 10%, FAAF: 17%). Moreover, good correlation between the magnitudes of change in conformer populations and incision efficiencies was found among the lesions. In FABP, the A/T polarity swap caused a 60% change in S-conformer proportion and a 3-fold reduction in repair efficiency. The changes were significantly lower in FAF and FAAF (24% and 5-15%, respectively), as were the repair efficiencies (2.0- and 1.8-fold, respectively). At a glance, the results seem to suggest B-conformers have greater reparability than S-conformers. This feature is in clear contrast to the trend that has been observed previously for AF and AAF in certain sequence contexts, i.e., the S-conformer is more reparable than the B-conformer (7,35). This type of conformation-specific repair is not only restricted to arylamines, but also applied to other bulky lesions. For instance, Geacintov et al have reported that the base-displaced *cis*-N²-dG adducts of benzo[a]pyrene are incised more efficiently than the minor-groove orientated *trans*-N²-dG adducts (33).

The *E. coli* repair results in the present study, however, seem to match well with events of adduct-induced DNA bending/distortion, as evidenced by blue shifts in CD (Table 1) and retardation of mobility in EMSA (Fig. 3). The slowed mobility indicates flexibility at the lesion site as observed by Tsao et al for (+)-*trans*-anti-[BP]-N²-dG lesion in the TG*T

sequence context with concomitant thermal destabilization (41,48). Similarly, the bulky N-acetyl FAAF exhibited significantly slower electrophoretic mobility as compared to FAF and FABP within the same sequence context. In case of sequence, the G*CA duplex exhibited consistently greater bending than its G*CT counterpart, with the effect being significantly greater for FAAF than FAF and FABP. A similar CD pattern has been reported for AAF-modified *NarI* duplexes related to the formation of a B-Z junction (49). Clearly, the A/T swap alters the conformational equilibrium *anti* (B-) to *syn* (S- or W-). It should be noted that the G*CA (--TCG*CAA--) sequence contains a stretch of alternating pyrimidine:purine bases, which are predisposed to DNA bending (50-52). In contrast, such a stretch is interrupted in the highly S-conformeric G*CT (--TCG*CTA--) sequence. It is possible that the B-conformer may facilitate DNA bending, due to the exposure of the carcinogen moiety to the major groove's hydrophilic environment. In both sequences, a major effect was observed with FAAF, followed by FABP and FAF (Fig. 3). MD/PMF simulations indicate, however, that the major changes in G*CA occur in the S-state. Also, unlike the CD data, there were no significant differences in electrophoretic mobility between the two sequences (Fig. 3). The reason for the inconsistency in mobility, CD, and MD data demonstrating the sequence effect is not apparent, but the greater bending and flexibility of FAAF over FABP or FAF is in good agreement with the observed repair efficiencies (FAAF >> FAF $\approx$ FABP; Fig. 6).

The repair results in Figure 6 along with previously reported work on polycyclic aromatic hydrocarbons (41,53) and arylamines (7) indicate that lesion-induced destabilization of DNA is a major determining factor for repair. However, these lesions were consistently

repaired 2~3 times more efficiently in G*CA than in G*CT, which was not consistent with relative thermodynamic stabilities observed for each. The inconsistency is likely due to the second step of damage recognition (54) that becomes much more significant for FAAF vs. FABP and FAF within a given sequence. Unlike the initial step of damage recognition by UvrA₂ which depends on DNA conformation and sequence, the second step of recognition is well-known to be characterized by the direct interaction of UvrB with adduct itself upon DNA strand opening (47,54-56). In other words, the structure and chemistry of the lesions matter more with UvrB than UvrA₂. Recently, Liu et al reported the NER incision efficiencies of the bulky benzo[a]pyrene and equine estrogen substrates using human HeLa cell extracts and bacterial UvrABC proteins (53). They demonstrated that in spite of having differences in the prokaryotic and eukaryotic NER proteins, XPC-RAD23B and UvrB, respectively, exhibit common feature of β -hairpin intrusion for damage recognition. In addition, it was found that local thermodynamic destabilization near the lesion site assists the insertion of β -hairpin, thus recognition.

Clearly, the present study shows that the thermodynamic destabilization of the DNA duplex along with lesion flexibility promotes strand opening and thus the second step of damage recognition. The presence of the *N*-acetyl group (see below) may make FAAF more efficiently recognized than FAF and FABP at the second step due to its flexible nature and greater destabilization of the DNA double helix. As for the G*CA/G*CT transition, the initial recognition step conducted by UvrA₂ should be a major determinant factor as the same efficiency of recognition at the second step is expected for the same type of lesion. Thus, bending appears to be an important factor for the DNA damage

recognition. Indeed, a recent crystal study by Jaciuk et al (10) found that in the active site of UvrA, the fluorescein-modified duplexes were bent by $\sim 15^\circ$ and the structure was related to the kinked structure of psoralen and PAH adducts according to NMR (57). They concluded that the UvrA₂ protein does not have direct chemical contacts with a lesion per se, but indirectly senses the overall helical distortion (unwinding and bending)(10). Since energy is required for the bending, formation of the pre-bent DNA induced by bulky lesions would likely enhance the UvrA₂ binding and thus damage recognition.

Cai et al have reported a similar repair trend for the 5'-CACACCG*CACAC-3' sequence vs. 5'-CCATCCG*CTACC-3', in which G* is the major mutagenic lesion derived from the environmental carcinogen benzo[*a*]pyrene, 10*S* (+)-*trans-anti*-B[*a*]P-*N*²-dG (39). A greater repair (1.6-fold) of the CG*CA duplex over the CG*CT counterpart was attributed to its higher bending of the distant 5'-end sequences, as evidenced by gel experiments and MD simulations; these findings are consistent with the bending argument made in the present study. Although the sequence contexts (underlined above) near the lesion site, including the 3'-next flanking base, are identical to those used in the present study, they did not consider the structural and repair consequences of the GC*A/G*CT swap.

***N*-Acetyl Factor**

Though a relatively small modification, the *N*-acetyl group has an important structural consequence. As shown in Figure 7, the lack of the acetyl moiety in G*CT-FAF, allows the G* moiety (red licorice representation) to point away from the sugar and stay in the

plane of the GC base pair, where the N-H bond is directed towards the sugar. However, in G*CT-FAAF, the acetyl group will have a steric clash with the sugar moiety of G* (identified with a black arrow), thereby leading the fluorene moiety (cyan) to be perpendicular to the G* ring system. This persistent “perpendicular” lesion orientation is predicted to lead to more disruption of the DNA duplex. A similar observation regarding the differences in the orientation of AF and AAF was reported by Mu et al (19) who have conducted a NER study of these lesions in human HeLa cells. MD simulations in that work indicate that the greater repair susceptibility of AAF stems from steric hindrance effects of the acetyl group which significantly diminish the adduct-base stabilizing van der Waals stacking interactions relative to AF. The persistent “perpendicular” FAAF mentioned above could raise conformational barriers of FAAF, resulting in the overall lower free energy of the *syn*-G* PMFs for FAAF, as compared to FABP and FAF. In other words, the *N*-acetyl group in FAAF could act as a “conformational locker” (7) that orients the adduct in a position that will lead to greater destabilization of the DNA duplex (Fig. 4, Table 2), as well as the increased bending observed in CD (Table 1) and mobility assays (Fig. 3). As a result, FAAF lesions are repaired at significantly greater rate compared to the FABP and FAF lesions (7).

Conclusion

The A to T polarity swap in the arylamine-modified G*CA/G*CT transition produced a dramatic increase in destabilized stacked conformation, but resulted in unexpected 2~3-fold lower NER efficiencies. These results are consistent with lesion-induced DNA bending/distortion. As for lesions, FAAF was repaired 3~4 times more efficiently than

FABP and FAF lesions, which is consistent with the extent of bending and helix destabilization, as well as the steric constraint in the duplex (“*N*-acetyl factor”) (7). A number of different damage recognition parameters have been implicated in the molecular mechanisms of NER (9,55,58). Although, it is known that thermal/thermodynamic destabilization and DNA distortion/bending are important factors for damage recognition by repair proteins (9,39,53). The present results show that lesion-induced DNA bending/thermodynamic destabilization is a more important NER factor than the usual S/B conformational heterogeneity, as has been observed previously for AF and AAF in certain sequence contexts (7,35). The present work represents a novel 3'-next flanking sequence effect as a unique NER factor for bulky arylamine lesions in *E. coli*. Taken together, the present results demonstrate the complexity in DNA recognition factors for repair of bulky arylamine lesions in *E. coli*.

Supplementary Data: Supplementary Data are available on www.etdadmin.com online: the details of the PMF simulations, supplementary figures S1-S16 and supplementary tables (S1-S7).

Funding: This work was supported by the National Institutes of Health (Grant number# CA098296 to BC, GM051501 to ADM, and CA86927 to YZ); the National Science Foundation/RI-EPSCoR (0554548) and the RI-INBRE core facility supported by National Institutes of Health (P20 RR016457).

References:

1. Greenbaltt, M.S., Bennett, W.P., Hollstein, M.H. and Harris, C.C. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.*, **54**, 4855-4878.
2. Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T. and Editors. (2006) *DNA Repair and Mutagenesis, Second Edition*. ASM Press.
3. Luch, A. (2005) Nature and nurture - lessons from chemical carcinogenesis. *Nat. Rev. Cancer*, **5**, 113-125.
4. Beland, F.A. and Kadlubar, F.F. (1990) *Handbook of Experimental Pharmacology*. Springer-Verlag, Heidelberg, pp 267-325.
5. Cho, B. (2010) *Structure-function characteristics of aromatic amine-DNA adducts in The Chemical Biology of DNA Damage*, pp 217-233.
6. Meneni, S.R., Shell, S.M., Gao, L., Jurecka, P., Lee, W., Sponer, J., Zou, Y., Chiarelli, M.P. and Cho, B.P. (2007) Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. *Biochemistry*, **46**, 11263-11278.
7. Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M.P. and Cho, B.P. (2012) Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.* **40**, 3939-3951.

8. Patnaik, S. and Cho, B.P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol.* **23**, 1650-1652.
9. Geacintov, N., E. , Broyde, S., Buterin, T., Naegeli, H., Wu, M., Yan, S. and Patel., D. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers*, **65**, 202-210.
10. Jaciuk, M., Nowak, E., Skowronek, K., Tanska, A. and Nowotny, M. (2011) Structure of UvrA nucleotide excision repair protein in complex with modified DNA. *Nat. Struct. Mol. Biol.* **18**, 191-197.
11. Min, J.H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570-575.
12. Cai, Y., Patel, D.J., Broyde, S. and Geacintov, N.E.(2010) Base sequence context effects on nucleotide excision repair. *J. Nucleic Acids*, **2010**.
13. 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.
14. Burnouf, D., Koehl, P. and Fuchs, R.P. (1989) Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci. USA*, **86**, 4147-4151.
15. Delagoutte, E., Bertrand-Burggraf, E., Lambert, I.B. and Fuchs, R.P. (1996) Binding and incision activities of UvrABC excinuclease on slipped DNA intermediates that generate frameshift mutations. *J. Mol. Biol.* **257**, 970-976.

16. Fuchs, R.P. and Seeberg, E. (1984) pBR322 plasmid DNA modified with 2-acetylaminofluorene derivatives: transforming activity and in vitro strand cleavage by the Escherichia coli uvrABC endonuclease. *EMBO J.* **3**, 757-760.
17. Mu, D., Bertrand-Burggraf, E., Huang, J.C., Fuchs, R.P., Sancar, A. and Fuchs, B.P. (1994) Human and E.coli excinucleases are affected differently by the sequence context of acetylaminofluorene-guanine adduct. *Nucleic Acids Res.* **22**, 4869-4871.
18. Seeberg, E. and Fuchs, R.P. (1990) Acetylaminofluorene bound to different guanines of the sequence -GGCGCC- is excised with different efficiencies by the UvrABC excision nuclease in a pattern not correlated to the potency of mutation induction. *Proc. Natl. Acad. Sci. USA*, **87**, 191-194.
19. Mu, H., Kropachev, K., Wang, L., Zhang, L., Kolbanovskiy, A., Kolbanovskiy, M., N, E.G. and Broyde, S. (2012) Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.* (Epub ahead of print)
20. 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 19F NMR, UV, and CD spectroscopy. *Biochemistry*, **46**, 13310-13321.
21. Jain, N., Meneni, S., Jain, V. and Cho, B.P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.* **37**, 1628-1637.

22. Liang, F. and Cho, B.P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry*, **49**, 259-266.
23. Meneni, S.R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M.P. and Cho, B.P. (2006) Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res.* **34**, 755-763.
24. Gao, L., Zhang, L., Cho, B.P. and Chiarelli, M.P. (2008) Sequence verification of oligonucleotides containing multiple arylamine modifications by enzymatic digestion and liquid chromatography mass spectrometry (LC/MS). *J. Am. Soc. Mass Spectrom.* **19**, 1147-1155.
25. Chakrabarti, M.C. and Schwarz, F.P. (1999) Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry. *Nucleic Acids Res.* **27**, 4801-4806.
26. Liang, F., Meneni, S. and Cho, B.P. (2006) Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene-DNA adducts. *Chem. Res. Toxicol.*, **19**, 1040-1043.
27. Cho, B.P. and Zhou, L. (1999) Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by ¹⁹F NMR spectroscopy. *Biochemistry*, **38**, 7572-7583.
28. Luo, C., Krishnasamy, R., Basu, A.K. and Zou, Y. (2000) Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, N-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease. *Nucleic Acids Res.* **28**, 3719-3724.

29. Zou, Y., Liu, T.M., Geacintov, N.E. and Van Houten, B. (1995) Interaction of the UvrABC nuclease system with a DNA duplex containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE. *Biochemistry*, **34**, 13582-13593.
30. Zou, Y. and Van Houten, B. (1999) Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J.* **18**, 4889-4901.
31. Banavali, N.K. and MacKerell, A.D., Jr. (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **319**, 141-160.
32. 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.
33. Geacintov, N.E., Cosman, M., Hingerty, B.E., Amin, S., Broyde, S. and Patel, D.J. (1997) NMR solution structures of stereoisometric covalent polycyclic aromatic carcinogen-DNA adduct: principles, patterns, and diversity. *Chem. Res. Toxicol.* **10**, 111-146.
34. 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.
35. 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.

36. Blazy, B., Culard, F. and Maurizot, J.C. (1987) Interaction between the cyclic AMP receptor protein and DNA. Conformational studies. *J. Mol. Biol.* **195**, 175-183.
37. Connor, F., Cary, P.D., Read, C.M., Preston, N.S., Driscoll, P.C., Denny, P., Crane-Robinson, C. and Ashworth, A. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res.* **22**, 3339-3346.
38. Meierhans, D., Sieber, M. and Allemann, R.K. (1997) High affinity binding of MEF-2C correlates with DNA bending. *Nucleic Acids Res.* **25**, 4537-4544.
39. Cai, Y., Kropachev, K., Xu, R., Tang, Y., Kolbanovskii, M., Kolbanovskii, A., Amin, S., Patel, D.J., Broyde, S. and Geacintov, N.E. (2010) Distant neighbor base sequence context effects in human nucleotide excision repair of a benzo[a]pyrene-derived DNA lesion. *J. Mol. Biol.* **399**, 397-409.
40. Suh, M., Ariese, F., Small, G.J., Jankowiak, R., Liu, T.M. and Geacintov, N.E. (1995) Conformational studies of the (+)-trans, (-)-trans, (+)-cis, and (-)-cis adducts of anti-benzo[a]pyrene dilepoxide to N2-dG in duplex oligonucleotides using polyacrylamide gel electrophoresis and low-temperature fluorescence spectroscopy. *Biophys. Chem.* **56**, 281-296.
41. Tsao, H., Mao, B., Zhuang, P., Xu, R., Amin, S. and Geacintov, N.E. (1998) Sequence dependence and characteristics of bends induced by site-specific polynuclear aromatic carcinogen-deoxyguanosine lesions in oligonucleotides. *Biochemistry*, **37**, 4993-5000.

42. Abuaf, P., Hingerty, B.E., Broyde, S. and Grunberger, D. (1995) Solution conformation of the N-(deoxyguanosin-8-yl)aminofluorene adduct opposite deoxyinosine and deoxyguanosine in DNA by NMR and computational characterization. *Chem. Res. Toxicol.* **8**, 369-378.
43. 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.
44. 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.
45. Eckel, L.M. and Krugh, T.R. (1994) Structural characterization of two interchangeable conformations of a 2-aminofluorene-modified DNA oligomer by NMR and energy minimization. *Biochemistry*, **33**, 13611-13624.
46. Yang, Z., Colis, L.C., Basu, A.K. and Zou, Y. (2005) Recognition and incision of gamma-radiation-induced cross-linked guanine-thymine tandem lesion G[8,5-Me]T by UvrABC nuclease. *Chem. Res. Toxicol.* **18**, 1339-1346.
47. Zou, Y., Shell, S.M., Utzat, C.D., Luo, C., Yang, Z., Geacintov, N.E. and Basu, A.K. (2003) Effects of DNA adduct structure and sequence context on strand opening of repair intermediates and incision by UvrABC nuclease. *Biochemistry*, **42**, 12654-12661.
48. Ruan, Q., Liu, T., Kolbanovskiy, A., Liu, Y., Ren, J., Skorvaga, M., Zou, Y., Lader, J., Malkani, B., Amin, S. *et al.* (2007) Sequence context- and temperature-

- dependent nucleotide excision repair of a benzo[a]pyrene diol epoxide-guanine DNA adduct catalyzed by thermophilic UvrABC proteins. *Biochemistry*, **46**, 7006-7015.
49. Koehl, P., Valladier, P., Lefevre, J.F. and Fuchs, R.P. (1989) Strong structural effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Nucleic Acids Res.* **17**, 9531-9541.
  50. Lankas, F., Sponer, J., Langowski, J. and Cheatham, T.E., 3rd. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **85**, 2872-2883.
  51. McConnell, K.J. and Beveridge, D.L. (2001) Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* **314**, 23-40.
  52. Lyubchenko, Y.L., Shlyakhtenko, L.S., Appella, E. and Harrington, R.E. (1993) CA runs increase DNA flexibility in the complex of lambda Cro protein with the OR3 site. *Biochemistry*, **32**, 4121-4127.
  53. Liu, Y., Reeves, D., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, M., Kolbanovskiy, A., Bolton, J.L., Broyde, S., Van Houten, B. *et al.* (2011) Probing for DNA damage with beta-hairpins: similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair (Amst)*, **10**, 684-696.
  54. Zou, Y., Luo, C. and Geacintov, N.E. (2001) Hierarchy of DNA damage recognition in Escherichia coli nucleotide excision repair. *Biochemistry*, **40**, 2923-2931.

55. Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2006) Prokaryotic Nucleotide Excision Repair: The UvrABC System. *Chemical Reviews*, **106**, 233-252.
56. Van Houten, B., Croteau, D.L., DellaVecchia, M.J., Wang, H. and Kisker, C. (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutation Research*, **577**, 92-117.
57. Van Houten, B. and Snowden, A. (1993) Mechanism of action of the Escherichia coli UvrABC nuclease: clues to the damage recognition problem. *Bioessays*, **15**, 51-59.
58. Malta, E., Verhagen, C.P., Moolenaar, G.F., Filippov, D.V., van der Marel, G.A. and Goosen, N. (2008) Functions of base flipping in E. coli nucleotide excision repair. *DNA Repair (Amst)*, **7**, 1647-1658.

Figure captions:

Figure 1: (a) Structures of ABP [*N*-(2'-deoxyguanosin-8-yl)-4-aminobiphenyl], AF [*N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [*N*-(2'-deoxyguanosin-8-yl)-2-acetylaminofluorene] and their fluoro models, FABP [*N*-(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl], FAF [*N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene], and FAAF [*N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene]; (b) 11-mer GCA and GCT duplexes used in the present study; (c) Major groove views of the B, S, and W-conformers of ABP, AF, and AAF. Color code: modified-dG (red), dC (green) opposite the lesion site (orphan C), fluorene (grey CPK), acetyl (AAF only, magenta).

Figure 2: ¹⁹F NMR (a, b, c) and CD (d, e, f) spectra of FABP-, FAF- and FAAF-modified 11-mer duplexes, respectively, in the G*CA and G*CT sequence contexts at 20°C. The B, S, and W notation used in the ¹⁹F NMR (a, b, c) signal assignments represent major groove “B”, base displaced stacked “S” and minor groove “W” conformers, respectively (see Figure 1c legend). The CD spectra (d, e, f) of modified duplexes show the typical B-DNA characteristic (positive and negative ellipticity at 275 and 245 nm, respectively) and different ICD_{290-350nm} patterns represent the lesion conformation in the duplex (see CD in Results).

Figure 3: Autoradiograph of 15% (w/v) native polyacrylamide gel (acrylamide/bis-acrylamide 29:1, w/w) showing the relative mobility of single strand (ss) and GCA/GCT 19-mer double strand (ds); both unmodified and single site-specifically modified by FABP, FAF and FAAF.

Figure 4: Comparative thermodynamic parameters histogram of FABP (Black), FAF (Blue) and FAAF (Pink) in (a) G*CA duplexes and (b) G*CT duplexes. The $\Delta\Delta$ values

represent; $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex), $\Delta\Delta S = \Delta S$ (modified duplex) - ΔS (control duplex), $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).

Figure 5: Free energy profiles from PMF calculations over the sampling range of 0.5 ns to 3 ns for G*CT-FABP, G*CA-FABP, G*CT-FAF, G*CA-FAF, G*CT-FAAF, and G*CA-FAAF modified sequences (red and blue) as well as the unmodified GCT and GCA (black) as a function of the pseudo-dihedral angle PHI (Fig. S7).

Figure 6: Absolute incision rates (a, b) of FABP, FAF, and FAAF-modified 55-mer substrates in the G*CA and G*CT sequence contexts. (c) Relative percent incision rate histograms with respect to G*CA-FAAF modified duplex.

Figure 7: Image of the a) G*CT-FAF adduct in the B state (350°) and b) G*CT-FAAF adduct in the B state (330°). The flipping G6 base is red licorice, the orphan C6 is blue licorice, adducts are atom-colored thick CPK and the remainder of the DNA is atom-colored thin CPK. In b) the arrow indicates the sugar moiety with which the acetyl or aryl moiety on FAAF is suggested to form a steric clash.

Table captions:

Table 1: Lesion-induced CD blue shifts.

Table 2: Thermal and thermodynamic parameters derived from differential scanning calorimetry (DSC) curves.

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

Table 1:

<i>Lesion</i>	<i>GCA</i> (nm)	<i>GCT</i> (nm)	<i>Blue Shift</i> $\Delta^a_{(G^*G)}$ (nm)		<i>Blue shift</i> $\Delta^b_{(G^*CA-G^*CT)}$ (nm)
Control	271	271	GCA	GCT	
FABP	267	269	4	2	2
FAF	268	271	3	0	3
FAAF	263	267	8	4	4

a) Difference in the wavelength of positive band between the modified and control duplexes.

b) Difference in the wavelength of positive band between the G*CA and G*CT duplexes.

Table 2:

	CCATCG*CAACC GGTAGCGTTGG				CCATCG*CTACC GGTAGCGATGG			
	$-\Delta H$ kcal/mol	$-\Delta S$ eu	$-\Delta G_{37}$ kcal/mol	T_m °C	$-\Delta H$ kcal/mol	$-\Delta S$ eu	$-\Delta G_{37}$ kcal/mol	T_m °C ^b
Control^a	79.1	214.9	12.4	64.5	75.0	203.9	11.8	63.0
FABP^a	76.4	214.1	10.0	54.3	64.2	178.3	8.9	51.8
FAF^a	65.7	180.7	9.7	55.4	58.6	160.1	8.9	53.4
FAAF^a	39.7	106.1	6.8	42.7	33.9	88.8	6.4	39.7
	$\Delta\Delta H$ kcal/mol ^b	$\Delta\Delta S$ eu ^c	$\Delta\Delta G_{37}$ kcal/mol ^d	ΔT_m °C ^e	$\Delta\Delta H$ kcal/mol ^b	$\Delta\Delta S$ eu ^c	$\Delta\Delta G_{37}$ kcal/mol ^d	ΔT_m °C ^e
FABP^a	2.7	0.8	2.4	-10.2	10.8	25.6	2.9	-11.2
FAF^a	13.4	34.2	2.7	-9.1	16.4	43.8	2.9	-9.6
FAAF^a	39.4	108.8	5.6	-21.8	41.1	115.1	5.4	-23.3

a) The results were calculated from integration of the DSC curve directly. ΔG and ΔH represents the heat absorbed during duplex melting at 0.12 mM. The average standard deviations for ΔG , ΔH , and T_m are ± 0.2 , ± 2.0 , and ± 0.4 , respectively, b) $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex), c) $\Delta\Delta S = \Delta S$ (modified duplex) - ΔS (control duplex), d) $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex), e) $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Manuscript-IV

To be submitted in *Chemical Research in Toxicology*

Investigation of the 3'-next flanking base effect on the translesion synthesis across arylamine-DNA adducts: A structural perspective

Vipin Jain, Satyakam Patnaik[#], Vaidyanathan Ganesan[#], and Bongsup P. Cho*

Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island,
Kingston, RI, 02881, USA

[#]contributed equally

*To whom correspondence should be addressed:

Telephone#: (+1) 401-874-5024; FAX# (+1) 401-874-5766; E-mail: bcho@uri.edu

Present address:

Department of Biomedical and Pharmaceutical Sciences
University of Rhode Island
7 Greenhouse Road
Kingston, RI, 02881, USA

Abstract:

Sequence effect is a key structural factor that affects the mutational and repair consequences of the DNA-adduct. Previously, we reported the unusual sequence effect involving 3'-next flanking base (A vs T) on the conformational profile and *E.coli* UvrABC mediated nucleotide excision repair (NER) of arylamine lesions. In here, we extended our work to investigate the impact of the 3'-next flanking base (A vs T) on aminobiphenyl (ABP) and aminofluorene (AF)-induced major groove (B)/stacked (S) conformational heterogeneity during a simulated translesion synthesis. To do so, we carried out ^{19}F NMR/CD studies using two 19-mer template sequences, G*CA [d(5'-CTTACCATCG*CAACCATTC-3')] and G*CT [d(5'-CTTACCATCG*CTACCATTC-3')] annealed to the primers of different length. Moreover, we performed primer extension assay using both, Klenow fragment (exonuclease-deficient) of DNA polymerase I as well as Dpo4 to investigate the aforementioned effect on polymerase nucleotide insertion efficiency. Our ^{19}F NMR/ICD results showed that ABP and AF in the G*CT duplex exhibits a higher population of S-conformer than the G*CA duplex, but at the replication fork lesions exhibited similar conformational profiles irrespective of the sequence. In line with the structural data, no substantial effect was observed in the base insertion efficiency at the replication fork (n-1, n and n+1). However, the running start experiment clearly showed that the full length primer extension across modified templates (G*CA and G*CT; ABP and AF) was significantly faster in more B-conformeric G*CA sequence. In both lesions, significant stalling occurred at prelesion and lesion sites, but the anti-conformeric FABP showed more stalling at preinsertion site (n - 1) whereas the S conformer FAF at the lesion site (n). Kinetics data show that the relative nucleotide

insertion rate was significantly reduced at immediate upstream base pairs (n+1 and n+3). Taken together, although we were unable to observe any dramatic effect of 3'-next flanking base on the polymerase efficiency, still the results demonstrated the role of lesion-induced conformational heterogeneity in modulating the translesion DNA synthesis.

Introduction:

Constant interaction of the DNA with the variety of chemicals and radiation results into DNA adduct formation.¹ Body is well-equipped with the different repair pathways to remove most of these adducts.^{2,3} However, some of the lesions are capable of evading the repair process and successfully enters into the replication cycle. DNA replicative polymerase executes the replication across an undamaged DNA template with high accuracy and efficiency. But, the introduction of lesion interrupts the polymerase activity considerably. Therefore, replicative or specialized (bypass) polymerase mediated translesion synthesis (TLS) across the lesion are quiet mutagenic in nature, which could ultimately triggers the cancer initiation.^{2,3} The complex mutational spectra associated with a particular DNA lesion results due to the involvement of various factors such as, geometry of lesion in the polymerase active site, the nature of bases surrounding the lesion, the family of DNA polymerase involved in the replication process etc.⁴

Arylamines are classified as an important class of environmental carcinogens due to their genotoxic behavior. They have been implicated in the etiology of numerous human cancers such as breast, liver and bladder cancer.¹ Arylamines exert their effect upon metabolic activation by attacking the C-8 position of guanine to produce C8-substituted dG lesions. AF/AAF and ABP are the three major DNA adducts (Fig.1a) derived from the liver carcinogen 2-aminofluorene and bladder carcinogen 4-aminobiphenyl, respectively.^{5,6} In DNA duplex, ABP and AF adducts exist in a equilibrium of two major conformers: *anti*-glycosidic “B-type”, in which the adduct orients in the major groove of the duplex and *syn*-glycosidic “stacked (S)”, in which the carcinogen is base-displaced at

the adduct site (Fig. 1b, AF as an example).⁷⁻¹⁰ However, due to the non-coplanarity of biphenyl moiety, ABP does not stack that well as in case of AF, therefore predominantly adopts B-conformation.⁸ The B- and S-conformeric orientation of arylamine adducts is not only influenced by planarity of the adduct structure, but also from the nature of bases surrounding the lesion.^{7, 9-14} Strong literature evidences have bolster the hypothesis that the conformational profile of an adduct in a particular sequence context plays a significant role in determining its biological fate like mutation and repair.^{9, 11, 15-17} For instance, we have reported recently that arylamine lesions undergo nucleotide excision repair (NER) in a conformational specific manner where more distortive syn-glycosidic S-conformeric FAAF get repaired more efficiently than less-distortive B-conformation.¹¹

Involvement of structural and conformational factors is not limited to repair, but were also reported in the mutational outcome. For instance, AF and AAF are known to induce different mutational spectra where former induces solely point mutation and latter results into a mixture of point and frameshift mutations.¹⁸ These mutagenic dissimilarities between the two have been attributed to their structural and conformational differences. Furthermore, the mutagenicity of a particular adduct is also affected by its location within the DNA template i.e. neighboring base sequence context. One of the most striking reports on the sequence effect involves *NarI* sequence (5'--G₁G₂CG₃CN--3'), where AF and AAF adduct yield significantly higher frequency of -2 frameshift mutation, only when adducted to third guanine (G₃) of the sequence.¹⁹ Moreover, the mutational frequency was found to be governed by the base at *N* position. The presence of C at position *N* resulted into relatively higher rate of mutation as compared to T, suggesting

the involvement of next-flanking base.²⁰ Structural studies have shown that AF adduct adopts stacked (S) conformer when $N=C$, but displayed conformational heterogeneity when C was replaced by T.⁹

Recently, we discovered that the aforementioned 3'-next flanking T effect on the adduct conformation is not limited to AF or *NarI* sequence, but also persists in FABP or FAAF (fluorinated analogs of ABP and AAF) modified random G^*CT duplexes (5'-CCATCGCTACC-3'). Unexpectedly, FABP which is known to exist predominantly in B-conformation, exhibited conformational heterogeneity in G^*CT sequence context with 40:60 ratio of B- and S- conformers. But, when the polarity of 3'-next flanking base was changed from T:A to A:T in G^*CA duplex (5'-CCATCGCAACC-3'), the conformational heterogeneity of FABP was disappeared and resulted into exclusively B-conformation (100%). Similar effect was also seen in the FAF and FAAF modified G^*CN 11-mer duplexes, but with lesser intensity. In FAF, the presence of 3'-next flanking T (as opposed to A) increased the S-conformeric population from 66% to 90%, whereas FAAF exhibited ~5% increase in syn-glycosidic S- and W-conformers. The biological consequence of this effect was seen in the NER efficiencies of FABP, FAF and FAAF adducts. Wherein, all the three adducts were incised more efficiently when present in G^*CA duplex than G^*CT . Although, the magnitude of change in repair efficiencies was different in all three cases, it correlated well with the degree of change in their conformeric ratios.

Previously, we studied the chemically simulated TLS, although on different sequence contexts (CG*A and TG*A), to deduce the conformational behavior of FAF adduct at the replication fork.²¹ We found that the conformational heterogeneity exhibited by FAF in the duplex setting also exist at the replication fork and that the S-conformeric FAF in CG*A sequence context favors the insertion of A over C at the lesion site. In the present work, we carried out the similar type of simulated TLS on the aforementioned G*CT and G*CA duplex to investigate the impact of 3'-next flanking base on the conformation of FAF and FAPB during primer elongation. We also performed running start and steady state kinetic experiments to probe the effect on the insertion efficiency and fidelity of the DNA polymerases. Additionally, we studied the impact of flanking T on the conformational profile of FAPB in G*CT duplex by systematically replacing the flanking C with T.

Methods and Materials:

Caution: *4-Aminobiphenyl and 2-Aminofluorene derivatives are mutagens and suspected human carcinogens, therefore must be handled with caution.*

Crude oligodeoxynucleotides (ODN, 10 or 1 μ mol scale) in desalted form were purchased from Eurofins MWG operon (Huntsville, AL). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Preparation of FAF and FABP modified ODNs:

FABP modification of 11-mer ODNs (5'-CCATNGNTACC-3', Fig. 1c) and FAF/FABP modification of 19-mer ODNs (5'-CTTACCATCGCNACCATTTC-3', Fig. 1d) were performed by the general procedures described previously.^{9, 10, 13, 22} Briefly, approximately 5-10 mg of *N*-acetoxy-*N*-trifluoroacetyl-4'-fluoro-4-aminobiphenyl or *N*-acetoxy-*N*-trifluoroacetyl-7-fluorofluorene dissolved in absolute ethanol was added to a sodium citrate buffer (pH 6.0) containing 200-250 ODs of unmodified ODN and placed in a 37 °C shaker for overnight. The modified strands were purified by RP-HPLC system consisted of a Hitachi EZChrom Elite HPLC unit with an L2450 diode array detector and a Phenomenex Luna C18 column (150 X 10 mm, 5.0 μ m). We employed a gradient system involving 3-15% acetonitrile for 25 mins, in pH 7.0 ammonium acetate buffer (100 mM) with a flow rate of 2.0 mL/min.

Circular Dichroism (CD):

CD measurements were conducted on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller.²³ Typically, 2 ODs of each strand were annealed with an equimolar amount of a primer sequence. The samples were dissolved in 400 μ L of a neutral buffer (0.2 M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA) and placed in a 1.0 mm path length cell. The samples were heated at 85 °C for 5 min and then cooled to 15 °C, over a 10 min period to ensure complete duplex formation. Spectra were acquired every 0.2 nm with a 2 s response time from 200 to 400 nm at a rate of 50 nm/min, were the averages of 10 accumulations, and were smoothed using 17-point adaptive smoothing algorithms provided by Jasco.

Dynamic ^{19}F NMR:

Approximately 15 ODs of a modified template was annealed with an equimolar amount of a complementary primer sequence to produce different duplexes (Fig. 1c and 1d). The samples were then dissolved in 300 μL of typical pH 7.0 NMR buffer containing 10% D_2O /90% H_2O , 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA, and filtered into a Shigemi tube through a 0.2 μm membrane filter. All ^1H and ^{19}F NMR results were recorded using a dedicated 5 mm $^{19}\text{F}/^1\text{H}$ dual probe on a Bruker DPX400 Avance spectrometer operating at 400.0 and 376.5 MHz, respectively, using acquisition parameters described previously.^{9, 10, 24, 25} Imino proton spectra at 5°C were obtained using a phase-sensitive jump-return sequence and referenced relative to that of DSS. ^{19}F NMR spectra were acquired in the ^1H -decoupled mode and referenced relative to that of CFCl_3 by assigning external C_6F_6 in C_6D_6 at -164.9 ppm. ^{19}F NMR spectra were measured between 5 and 60 °C with an increment of 5-10 °C. Temperatures were maintained by a Bruker-VT unit with the aid of controlled boiling liquid N_2 in the probe.

Primer Extension Assay:

Standing start experiment: Steady state kinetic experiments were performed as described previously.^{21, 26} Briefly, the primers (29–33-mers) were 5'-radiolabeled using $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase (T4 PNK) by following the manufacturer's protocol. The 5' ^{32}P -labeled primer (100 pmol) was annealed to either an unmodified or adducted template oligonucleotide (120 pmol) by heating to 95 °C for 5 min and then slowly cooled to room temperature. The primer– template sequence (100 nM) was incubated with Kf-exo[−] (0.5 or 1.0 nM) for 5 min to form a binary complex. The reaction was initiated by adding a

dNTP/MgCl₂ (dNTP in variable concentration and 5 mM MgCl₂) solution to a binary mixture, the mixture was incubated at 20 °C for variable time intervals in Tris buffer [50 mM Tris-HCl (pH 7.4), 50 µg/mL BSA, and 5% (v/v) glycerol], and the reaction was quenched with 50 µL of 50 mM EDTA (pH 8.0)/95% formamide solution. The quenched sample was heated to 95 °C for 5 min and immediately cooled on ice. The products were resolved with a denaturing polyacrylamide gel [20% polyacryamide (w/v)/7 M urea] and electrophoresed at 2000 V for 4 h. The gel was exposed on a Kodak phosphor imaging screen overnight and scanned with a Typhoon 9410 variable mode imager. The band intensities were quantitated using ImageQuantTL from GE Healthcare.

Running start experiment: A 25-mer primer was used with the modified dG (FAF, FABP and FAAF) at position 31 of the 44-mer template. The extension was performed using both Kf-exo⁻ (2.5 nM), and Dpo4 (50 nM FABP, 100 nM FAF) polymerase in the presence of all four dNTPs. The aliquots were withdrawn at regular intervals of time and the reaction was quenched as mentioned above. The extension products were separated on a 20% denaturing polyacrylamide gel.

Determination of Kinetic Parameters of Nucleotide Incorporation.

Steady-state kinetic parameters for incorporation of the nucleotide opposite the unmodified and FAF- or FABP-modified templates were studied by following the procedures reported previously.^{21, 26} The reactions were performed at 20 °C. For the unmodified sequence, reactions were performed for ≤1 min for correct nucleotide incorporation and for up to 45 min for incorrect nucleotide incorporation. The percentage

of primer extended in kinetic assays was determined by taking the ratio of extended primer to the total amount of primer (unextended + extended primer). Initial velocities were calculated by plotting product concentration versus time. The kinetic parameters k_{cat} and K_m were determined by a nonlinear least-squares fit of the initial velocity versus dNTP concentration using SigmaPlot 11.0 (Systat Software Inc.). The efficiency of nucleotide insertion by Kf-exo⁻ was calculated as k_{cat}/K_m . The relative insertion (fins) or extension (fext) efficiencies were calculated as $(k_{cat}/K_m)_{mismatch}$ or $lesion/(k_{cat}/K_m)_{unmodified\ control}$.

Results:

Model Systems:

The original 11-mer G*CT sequence was revised (Fig. 1c) to study the effect of flanking base on the conformational heterogeneity induced as a result of 3'-next flanking T. As shown in Fig. 1c, the changes were made on the flanking cytosines by systematically replacing them with either G, T or A. The templates were modified with FAPB and annealed with their respective complementary sequences for ¹⁹F NMR and CD.

Figure 1d shows the G*CN (N= A or T) sequences designed for the simulated TLS system. In here, the original sequences were extended from 11-mer (Fig. 1c) to 19-mer (Fig. 1d) to impart stability to n-1 and n duplexes. Four model template–primer duplexes (Fig. 1d) were prepared by annealing the FAPB- or FAF-modified G*CN 19-mer strand with primers of variable length (n – 1, n, n + 3, and n + 9, where n is the lesion site).

Dynamic ^{19}F NMR Spectroscopy:

Flanking base effect:

The objective of this work was to investigate whether the presence of T around the lesion site (G^*) will alter the conformational heterogeneity (% S-conformer) induced by the 3'-next flanking T in G^*CT sequence context. We were also interested in deducing the role of 3'-flanking C in the G^*CT sequence, therefore changed it to A, T or G (Fig. 2). We used FABP adduct for this study since it exhibited the most dramatic differences in conformational ratio in the two sequences (Fig. 3); G^*CT (60% S-conformer) and G^*CA (100% B-conformer). Figure 2 shows the ^{19}F NMR of the different FABP modified sequences at different temperatures. In all the cases, one major signal was observed at ~ 117 ppm which falls in the chemical shift range of B-conformer as has been reported previously.²² The B-conformational nature of FABP was also confirmed from the CD where negative ellipticity at 290-350 nm ($\text{ICD}_{290-350 \text{ nm}}$)(Fig 2a) is a signature pattern of B-conformation.²² In addition, all the duplexes exhibited one minor signal in the upfield S-conformation range (-117.5 to -118.0 ppm) except G^*AT which had a minor downfield signal (~ 116.8 ppm). The results clearly indicate that the presence of T next to the lesion site does not produce a synergistic effect rather shifts the equilibrium back to the B-conformer (Fig. 3 shows comparison). The replacement of 3'-flanking C by A or G slightly maintained the conformational heterogeneity as evidenced from ^{19}F NMR and imino protons spectra (Fig 3b)(asterisk), but in general favored the B-conformation. Overall, the aforementioned observations indicate that the CG^*CT is a unique sequence context, where the next-flanking T produce maximum effect.

Simulated TLS

FABP: Figure 4a and 4b shows the dynamic (5–50 °C) ^{19}F NMR spectra of the G*CT and G*CA TLS system at various primer positions (n-1 to full), respectively. The assignments of ^{19}F signals in these spectra have been reported previously.^{21, 25} The full G*CT and G*CA duplexes (Fig. 1d) which represent an end point in the TLS process, displayed a clear conformational differences (Fig. 4a and 4b) between them, similar to the one that was previously observed in their respective 11-mer duplexes (Fig. 3). The FABP in G*CT duplex exhibited B- and S-conformational mixture whereas in G*CA it exclusively adopted the B-conformation (Fig. 4). Although, the conformational heterogeneity was intact in G*CT 19-mer duplex, but the percentage of S-conformation was dropped from 60% to 33% going from original 11-mer to 19-mer sequence. At n+3 primer position, similar conformational differences were retained between the two sequences as evidenced from their ^{19}F signals. In contrast to n+3 and full, the two sequences exhibited very similar ^{19}F NMR characteristics at (n) and before (n-1) the lesion site as both showed the two prominent ^{19}F signals. We assigned these signals as *anti*-type B* and *syn*-type S* conformers based on their similarity with the B- and S-chemical shift range. Apparently, the only noticeable difference between G*CT and G*CA (at n and n-1) was in the percentages of B*- and S*-conformers (~12%) with former having relatively higher S*-population (Table 1). This was surprising in G*CA case, as it exhibited predominantly B-conformeric signal at n+3 and full duplexes (Fig 4b). The overall summary of FABP induced conformers with their population ratios at different primer positions of the two templates are tabulated in Table 1.

FAF: Figure 5a and 5b shows the dynamic ^{19}F NMR spectra of the FAF adducted G*CT and G*CA TLS system (n-1 to full), respectively. The full G*CT and G*CA duplexes (Fig. 1d) displayed the conformational profile (Fig 5a and 5b) similar to their respective 11-mer duplexes that were reported previously. In both sequences, FAF exhibited B- and S-conformational heterogeneity with G*CT having ~16% more S population than G*CA. But, the extension of template from 11-mer to 19-mer affected the G*CT S-conformer population (reduced from 90% to 80%) as was observed in the FABP case above. At n+3 primer position, the conformational gap between the two sequences was increased to ~23%, specifically due to increase in B-conformer population in G*CA (Fig. 5b). Moving from n+3 to n and n-1 positions, FAF exhibited two broad ^{19}F signals at low temperatures (5 and 10 °C), which were well resolved at 15 °C and higher temperatures. Again, based on their chemical shifts the signals were assigned as B* and S* conformers. Between the two sequences, the gap was further increased to ~34% at the lesion site (n), reason being the significant increase in G*CA B* population to ~58% (Fig. 5b and Table 1). However, at n-1 position they displayed a very similar ^{19}F characteristic with approximately equal populations of B* and S*. The overall summary of FAF TLS system is tabulated in Table 1.

Circular Dichroism:

Figure 6 shows the FABP- (Fig. 6a) and FAF- (Fig. 6b) induced $\text{ICD}_{290-350\text{nm}}$ patterns at various primer extension steps for the G*CA and G*CT TLS system. Previously, $\text{ICD}_{290-350\text{nm}}$ has been used as a tool for probing FABP- and FAF-induced S/B heterogeneity where positive ellipticity represents syn-S conformer and negative

indicates anti-B conformer.²³ In all four cases (G*CT/G*CA and FABP/FAF), at $n-1$ and n positions the $ICD_{290-350\text{ nm}}$ was positive (Fig. 6) which could be due to the substantial population of S*-conformer in these duplexes. However, the S*-conformeric population at these positions did not match perfectly with the intensity of positive ellipticity, suggesting a unique environment of lesion in these open half formed duplexes. In FABP, the intensity of positive ellipticity decreased as the primer length increased and ended with a negative dip in G*CA and slight S-shape pattern in G*CT. The negative dip in G*CA- FABP indicate a transition from the B*/S* mixture at $n - 1$ and n to predominantly B conformation at $n + 3$ and full. Similarly, the S-shape pattern in G*CT- FABP indicate a change from the B*/S* mixture to well defined B/S conformational heterogeneity. In contrast, the FAF induced $ICD_{290-350\text{ nm}}$ at $n - 1$ and n ended with a positive hump owing to the significantly higher population of S-conformation in both G*CA and G*CT. The CD spectral data are consistent with the conformational heterogeneity observed in the ^{19}F NMR spectra described above (Fig. 4 and 5).

Primer extension assay:

Running start: In the present study, the running start experiments were performed across the 44-mer G*CN ($N= A$ or T) template modified by either FABP or FAF. A short ^{32}P labeled primer (25-mer) was annealed to the templates and primer extension were carried in the presence of all four dNTPs, DNA polymerase (KF exo^- or Dpo4) and appropriate reaction buffer as mentioned above in materials and methods. It is known that the presence of an adduct on the template strand retards the DNA polymerase mediated primer extension. Our focus was to investigate the effect of conformational differences

induces by the 3'-next flanking base on the primer extension across the modified templates (Fig. 7, 8).

Primer extension across FABP modified template:

a) *KF (exo⁻)*: Figure 7a shows the products of KF exo⁻ mediated primer extension across the FABP modified templates at different time intervals and room temperature. The KF exo⁻ was able to extend primer to full length product (44-mer) across GCT and GCA as early as 2-5 mins after start. However, two blockages in extension were seen, one before the lesion site (n-1, 30-mer) and the other at the lesion site (n, 31-mer). The blockage at the lesion site that was seen in the initial few minutes was relatively stronger. It appears that most of the full length product was formed from this blocked primer, because the primers at n-1 site were persistent even after 60 mins of reaction. Although, both sequence context show similar blockage pattern near the lesion site, but the duration of blockage were different. In the initial few minutes, GCT displayed a significant increase in blockage at the lesion site (green arrow) which starts to drop once the extension into full length product initiated. The GCA had a similar increase, but for relatively shorter period of time than GCT. Apparently, the effect of a prolonged blockage at the lesion site was propagated to the full length extension with GCA showing a faster rate of extension than GCT (see inset in Fig. 7a).

b) *Dpo4*: Similar experiments were performed with Dpo4 across the FABP modified templates (G*CT and G*CA) as shown in Figure 8a. Like KF exo⁻, Dpo4 extended the primer to full length products (44-mer) in both the sequences. Besides, blocks were seen

at two similar sites i.e. n-1 and n. But, in contrast to KF (exo⁻) the amount of full length extension by Dpo4 was much lesser, even at high enzyme concentration. Furthermore, the major blockage was one base prior to the lesion site (n-1) as opposed to lesion site (n) observed in KF exo⁻ case (32-mer). Despite these differences in extension efficiency and blockage sites, both exhibit similar sequence effects. Here too, GCT shows prolonged blockage at the lesion site (green arrow) resulting into slower full length extension (see inset Fig. 8a). These results indicate that KF (exo⁻) and Dpo4 experience similar conformational resistance during primer extension.

Primer extension across FAF modified template:

a) *KF (exo⁻)*: As depicted in Figure 7b, the study was extended to FAF lesion to see whether it experiences a similar sequence effect. In here, the primers were blocked at three sites with a major block at the lesion site and two minors were one base before (30-mer) and after the lesion site (32-mer). Despite having an extra blocking site, KF (exo⁻) displayed better efficiency of extension across FAF than FABP. In addition, the blunt end addition was observed here whereas no such addition was seen in the FABP case. Notably, the similar sequence effect was observed in FAF with GCA showing a faster rate of extension compared to GCT (see inset Fig. 7b).

b) *Dpo4*: The results of Dpo4 mediated primer extension are shown in Figure 8b. Here, there were two blocking sites with the major blockage at the lesion site (n). Unlike KF (exo⁻), no blockage after the lesion site (32-mer) was noticed. However, the differences in the rate of extension across the two sequences were noted with faster

extension across GCA than GCT. But the differences were not that significant. Overall, our running start experiment results show that the two DNA polymerases do experience sequence effects while extending primer across the above two lesions.

Steady state kinetics:

We performed steady-state kinetic experiments to investigate the influence of FABP- and FAF-induced conformational heterogeneity on nucleotide insertion kinetics around the lesion. To do so, the lesion was positioned one nucleotide downstream of the templating base ($n - 1$), at the templating position (n), one nucleotide upstream ($n + 1$), and three base pairs upstream of the primer terminus ($n + 3$) (Fig. 9). In addition, we also examined the role of sequence effects in insertion efficiency (kcat/Km). The results of the steady-state kinetic experiments are presented in Tables 2-10.

Insertion Efficiency: The presence of lesion (FABP or FAF) immediately downstream of the templating nucleotide ($n - 1$) barely affected the relative insertion efficiency (f_{ins}) of the correct nucleotide dGTP [FABP; f_{ins} = 0.22 and 0.23 (Table 2), FAF; f_{ins} = 0.40 and 0.31 (Table 6)] for the G*CA and G*CT sequences, respectively, compared to the controls. In FABP, the f_{ins} of dCTP opposite the lesion (at n) was further reduced to 0.11 and 0.09 for G*CA and G*CT, respectively, which represents ~10 fold reduction in rates, relative to the controls (Table 3). Interestingly, the insertion frequency of the wrong nucleotide dATP opposite the adducted guanine in the G*CA (3×10^{-4}) and G*CT (2×10^{-4}) context was further reduced, relative to the control. This suggest that in G*CA dCMP insertion was preferred over dAMP by ~367-fold, while 450-fold in G*CT (Table

3). In FAF, the f_{ins} of dCTP opposite the lesion (at n) was reduced to 0.046 and 0.12 for G*CA and G*CT, respectively, which represents ~22 and ~8 fold reduction in rates, respectively, relative to the controls. Here too, the insertion frequency of the wrong nucleotide dATP opposite the adducted guanine in the G*CA (1.5×10^{-3}) and G*CT (1.7×10^{-2}) context was further reduced, relative to the control. This suggest that in G*CA dCMP insertion was preferred over dAMP by ~31-fold, while just ~7-fold in G*CT (Table 7).

Extension Efficiency: In FABP, the extension efficiency of correct nucleotide (dGTP) at n+1 position in match series (dCTP opposite the lesion) was 1.0×10^{-5} and 9.1×10^{-6} in G*CA and G*CT, respectively. However, when dATP was present opposite the lesion site (mismatch series) the nucleotide insertion was blocked dramatically in both sequences and therefore the extension efficiency could not be determined (Table 4). In FAF, the extension efficiency of correct nucleotide (dGTP) at n+1 position in match series was almost similar in G*CA (5.1×10^{-5}) and G*CT (7.4×10^{-5}). Interestingly, the extension (insertion at n+1) of mismatch series (dATP opposite the lesion site) was not blocked and the extension efficiency was found to be 1.0×10^{-6} and 5.0×10^{-7} in G*CA and G*CT, respectively (Table 8). At n + 3, the effect of the lesion on the extension frequency was least as the extension frequency was FABP (Table 5); 0.49 (G*CA) and 0.08 (G*CT), FAF (Table 9); 1.5×10^{-1} (G*CA) and 7.2×10^{-2} (G*CT), in the dC-match series. In the mismatch series, the extension efficiency between the G*CA and G*CT sequences differed by only 10- and 2.5-fold in FABP and FAF, respectively.

Discussion:

The role of neighboring bases in modulating the mutational and repair outcomes of bulky DNA lesions is well reported.^{9, 11, 15-17} Previously, we determined the effect of 3'-next flanking base on the NER efficiency of FABP, FAF and FAAF adducts. Our NMR study on flanking base effect (Fig. 2 and 3) clearly demonstrates that the G*CA and G*CT are the unique sequence contexts that induce this effect. Cai et al²⁷ reported the similar outcome in the repair efficiencies of benzo[a]pyrene modified oligonucleotides having exactly same sequence contexts (5'..CGCA..3' and 5'..CG*CT..3'). Therefore, we extended our work to examine the effect of 3'-next flanking base (G*CA and G*CT) on the primer extension across FABP and FAF in light of structural data obtained from a simulated TLS.

Development of conformational differences between G*CA and G*CT sequences during simulated TLS:

FABP: The ¹⁹F NMR and ICD patterns of the simulated TLS (Fig. 4 and 6) across the FABP-modified G*CA and G*CT templates are very interesting. At preinsertion (n – 1) and insertion (n) sites of both the sequences, the FABP adopted a dynamic mixture of anti-B* and syn-S* conformers with a small conformational difference of ~12% (G*CT having more syn-S* conformer). This conformationally heterogeneous behavior of FABP was expected in G*CT sequence as it adopted a mixture of B (40%) and S (60%) conformers in 11-mer full duplex, as reported previously. However, the similar conformational heterogeneity observed in G*CA case was unanticipated, because it is known to adopt exclusively B-conformer in this sequence context. It appears that at these

positions the open structure of the undeveloped duplex provides flexibility to the lesion to move in (S*) and out (B*) of the duplex easily. Previously, we have observed similar scenario in FAF modified C*GA and TG*A sequences, where despite the dramatic differences in n+3 and full duplexes, FAF exhibited similar conformational profiles at n-1 and n sites.^{21, 28} Likewise, in here, the conformational difference between the two sequences started to build up at n+3 site with FABP adopting a mixture of B (65%) and S (35%) conformers in G*CT sequence whereas exclusively producing B-conformer in G*CA (Fig. 4, Table 1). These differences were then transmitted to the full duplexes. This data suggest that the insertion of nucleotide 3 bases upstream and thereafter of the damage site restricts the movement of the lesion, thus allowing it to adopt a more defined stable conformations. Notably, the extension of original 11-mer to 19-mer did not affect the conformational profile of FABP in G*CA sequence, but resulted into the reduction of S-conformer (60% to 35%) in G*CT sequence. Apparently, the reason behind the effect of template length on the percentage of S-conformer is not clear. Overall, the structural data suggest that the 3'-next flanking base effect observed previously in 11-mer duplexes do exist in simulated TLS involving 19-mer, but minimized significantly near the lesion site (n-1 and n).

FAF: The conformational behavior of the FAF was identical to FABP at n-1 and n positions as it too adopted a dynamic mixture of anti-B* and syn-S* conformers. In comparison to FABP, however, FAF exhibited much higher percentage of S-conformer which can be attributed to the planar nature of the fluorene moiety. Besides, the sequence effect (G*CA vs. G*CT) at n-1 position was negligible (~1% vs. 12% in FABP)(Table

1). Surprisingly, pronounced conformational differences were seen at the lesion site (n) with G*CT exhibiting ~34% more S-conformer population than G*CA (Table 1). This effect was much more dramatic than FABP (~12%, Table 1) and what was previously observed in FAF modified 11-mer full duplexes (~24%). The transition from n to n+3 position resulted into the increase in S-conformer population in both G*CA (42% to 58%, Table 1) and G*CT (76% to 81%, Table 1) sequences. But, due to the difference in the magnitude of increase, the overall conformational difference was reduced to 23% (from 34%). The completion of the duplex further increased the G*CA S-conformer population to 64%, whereas that of G*CT was largely unaffected (~81%). Again, the extension from 11-mer to 19-mer have negligible effect on the conformational profile of FAF in G*CA, but significantly reduced the S-conformer population (90% to 80%) in G*CT as observed in FABP case. Nevertheless, the structural data showed that the 3'-next flanking base effect noticed previously in 11-mer duplexes continued in the simulated TLS with maximum effect at the lesion site followed by n+3 and full duplexes, but largely no effect at the preinsertion site (n-1).

Structural insight into the primer extension across the lesion:

The TLS (running start, Fig. 7 and 8) across FAF/FABP by Kf-exo⁻ and Dpo4 experienced strong blockage near the damage site (n-1, n and n+1) in G*CA and G*CT sequences, respectively. In case of lesion, both polymerases showed greater stalling at the lesion site in FAF, whereas in FABP the extension was blocked more effectively at prelesion site n – 1 as well as lesion site. The basis of this difference in the blocking pattern of the lesions could be obtained from their structural data. Hsu et al²⁹ have

reported the crystal structures of AF modified dG while undergoing accurate replication across it by high fidelity DNA polymerase. They found that dG-AF exhibits *syn* conformation at the preinsertion site and undergoes a transition to an *anti* conformation at the insertion site which allows it to pair with an incoming dCTP. As shown in Table 1, the FAF at preinsertion site exhibited more *syn* type S*-conformer whereas FABP was oriented more in *anti* type B* conformer. This suggests that the *anti*-configuration of the FABP modified dG at the pre-insertion site does not allow the incorporation of the base at the lesion site, thus blocks the process at n-1 site. Similar effects was observed in FAF modified C*GA and TG*A sequences, where highly B-conformeric TG*A (full duplex) exhibited stronger blockage at n-1 site, whereas mostly S-conformeric CG*A (full duplex) blocks at the lesion site (n). In terms of sequence effect, both the sequences (in FAF and FABP) displayed similar blocking pattern as expected owing to their similar conformational profiles around the lesion site (Fig. 7 and 8, Table 1). However, the full length extension across G*CA sequence was relatively faster than G*CT in both the lesions and by both the DNA polymerases (inset in Fig 7 and 8). As mentioned above, the polymerase prefers modified dG in *anti*-configuration at the lesion site for the successful replication across the lesion. Therefore, relatively slower extension across G*CT might be due the higher percentages of *syn*-S* conformer of the lesion at the lesion site.

The sequence effect on the insertion efficiency (*f_{ins}*) of correct base (dGTP) at the preinsertion site is negligible in both the lesions (Fig. 10). This is in line with the structural data (summarized in Table 1) where the 3'-next flanking base barely affected the conformational patterns of FABP and FAF. The similar trend was seen at the lesion

site as the conformational characteristics of the lesions (FABP and FAF) and the insertion efficiency of the correct base (dCTP) were comparable between the two sequences. Although, both the lesions favor the insertion of correct base opposite them, but the preference for correct base was more prominent in FABP than FAF (see steady state kinetics in results section). This could be due to the relatively higher population of syn-S* (Table 1) conformer in FAF which makes it difficult for the polymerase to decide which base to be inserted opposite the lesion. In FABP, the insertion of dATP does not exhibit any sequence effect as expected from its structural data. In FAF, however, G*CT sequence favored dATP insertion by about 11 fold more than G*CA, which could be attributed to the relatively higher population of syn-S* (~34%) conformer in G*CT case. The extension of match series at n+1 site, immediately after the lesion (FAF and FABP) was significantly reduced due to the prolonged blockage at the lesion site as evident from the running start experiment. We were unable to determine the extension efficiency of FABP mismatch series due to the extremely low insertion rate of the base at n+1 site. In contrast, FAF mismatch series showed extension beyond the lesion site but with extremely low efficiency. These results indicate that the insertion of incorrect dATP was not tolerated by both FAF and FABP. In addition, no significant sequence effects were observed at this particular position in both match and mismatch series. At n+3 site, where ¹⁹F NMR data showed well defined conformations, the extension efficiency too exhibited some sequence effects. In FABP, the G*CA show exclusively B-conformation (against B/S mixture in G*CT) at this position so as 6- and 10-fold higher extension efficiency (than G*CT) of match and mismatch series, respectively. Whereas FAF, which have only

16% difference in the conformational profile in the two sequences, exhibited only 2-2.5 fold alteration in the extension efficiency of match and mismatch series.

In summary, our results revealed that (a) AF/ABP in the G*CA and G*CT sequences adopt a similar conformational profiles at the replication fork (n-1 and n site), (b) due to similar conformational profiles, no substantial 3'-next flanking base effect was observed in the base insertion efficiency at the replication fork (n, n and n+1), (c) significant stalling occurred at both the prelesion (n – 1) and lesion (n) sites but the anti-conformeric FABP showed more stalling at preinsertion site whereas the S conformer FAF at the lesion site (n), and (d) the extension efficiency was substantially reduced immediately following the lesion (n+1) and 3 bases downstream of the lesion site (n + 3). Overall, despite being unable to witness the effect of 3'-next flanking base on the activity of polymerases, our results demonstrated the role of lesion-induced conformational heterogeneity in modulating the translesion DNA synthesis.

Funding:

This work was supported by the National Institutes of Health (Grant number# CA098296) and the RI-INBRE core facility supported by National Institutes of Health (P20 RR016457).

References

- (1) Luch, A. (2005) Nature and nurture - lessons from chemical carcinogenesis. *Nat. Rev. Cancer* 5, 113-125.
- (2) Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., Ellenberger, T. and Editors (2006) *DNA Repair and Mutagenesis, Second Edition*, ASM Press.
- (3) Hubscher, U. and Maga, G. (2011) DNA replication and repair bypass machines. *Curr. Opin. Chem. Biol.* 15, 627-635.
- (4) Seo, K.-Y., Jelinsky, S. A. and Loechler, E. L. (2000) Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens. *Mutation Research/Reviews in Mutation Research* 463, 215-246.
- (5) Beland, F. A. and Kadlubar, F. F. (1990) *Handbook of Experimental Pharmacology*, Springer-Verlag, Heidelberg, pp 267-325.
- (6) Heflich, R. H. and Neft, R. E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutation Research/Reviews in Genetic Toxicology* 318, 73-174.
- (7) Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health C. Environ. Carcinog. Ecotoxicol. Rev.* 22, 57-90.
- (8) 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.

- (9) 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 19F NMR, UV, and CD spectroscopy. *Biochemistry* 46, 13310-13321.
- (10) Meneni, S. R., Shell, S. M., Gao, L., Jurecka, P., Lee, W., Sponer, J., Zou, Y., Chiarelli, M. P. and Cho, B. P. (2007) Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. *Biochemistry* 46, 11263-11278.
- (11) Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M. P. and Cho, B. P. (2012) Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.* 40, 3939-3951.
- (12) 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.
- (13) Meneni, S. R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M. P. and Cho, B. P. (2006) Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res.* 34, 755-763.
- (14) Patnaik, S. and Cho, B. P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol.* 23, 1650-1652.
- (15) Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S. and Patel, D. J. (1997) NMR solution structures of stereoisometric covalent polycyclic aromatic

- carcinogen-DNA adduct: principles, patterns, and diversity. *Chem Res Toxicol* 10, 111-146.
- (16) Liu, Y., Reeves, D., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, M., Kolbanovskiy, A., Bolton, J. L., Broyde, S., Van Houten, B. and Geacintov, N. E. (2011) Probing for DNA damage with beta-hairpins: similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair (Amst)* 10, 684-696.
  - (17) Mu, H., Kropachev, K., Wang, L., Zhang, L., Kolbanovskiy, A., Kolbanovskiy, M., N, E. G. and Broyde, S. (2012) Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.* (Epub ahead of publication).
  - (18) Shibutani, S., Suzuki, N. and Grollman, A. P. (1998) Mutagenic specificity of (acetylamino)fluorene-derived DNA adducts in mammalian cells. *Biochemistry* 37, 12034-12041.
  - (19) Burnouf, D., Koehl, P. and Fuchs, R. P. (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.
  - (20) 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.

- (21) Vaidyanathan, V. G. and Cho, B. P. (2012) Sequence effects on translesion synthesis of an aminofluorene-DNA adduct: conformational, thermodynamic, and primer extension kinetic studies. *Biochemistry* 51, 1983-1995.
- (22) Liang, F. and Cho, B. P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry* 49, 259-266.
- (23) Liang, F., Meneni, S. and Cho, B. P. (2006) Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene-DNA adducts. *Chem. Res. Toxicol.* 19, 1040-1043.
- (24) Jain, N., Meneni, S., Jain, V. and Cho, B. P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.* 37, 1628-1637.
- (25) 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.
- (26) Miller, H. and Grollman, A. P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry* 36, 15336-15342.
- (27) Cai, Y., Kropachev, K., Xu, R., Tang, Y., Kolbanovskii, M., Kolbanovskii, A., Amin, S., Patel, D. J., Broyde, S. and Geacintov, N. E. (2010) Distant neighbor base sequence context effects in human nucleotide excision repair of a benzo[a]pyrene-derived DNA lesion. *J. Mol. Biol.* 399, 397-409.

- (28) Liang, F. and Cho, B. P. (2011) Conformational and thermodynamic impact of bulky aminofluorene adduction on simulated translesion DNA synthesis. *Chem. Res. Toxicol.* 24, 597-605.
- (29) 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.

Figure Legends:

Figure 1: (a) Structures of ABP [*N*-(2'-deoxyguanosin-8-yl)-4-aminobiphenyl], AF [*N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [*N*-(2'-deoxyguanosin-8-yl)-2-acetylaminofluorene] and their fluoro models, FABP [*N*-(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl], FAF [*N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene], and FAAF [*N*-(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene]; (b) Major groove views of the B and S conformers of AF [color code: modified-dG (red), dC (green) opposite the lesion site, fluorene (grey CPK)]; (c) 11-mer duplexes used in the present study; (d) Primer-template duplexes used in simulated translesion synthesis.

Figure 2: ¹⁹F NMR (a) and CD (b) spectra of FABP-modified 11-mer duplexes (see Fig. 1c). The 5-mer on the top of each column represent core of 11-mer duplex and the underlined base indicates the change in the original 11-mer G*CT sequence (5'-CCATCGCTACC--3').

Figure 3: ¹⁹F NMR (a) and imino proton (b) spectra of FABP-modified 11-mer duplexes (see Fig. 1c) at 20 °C. The 5-mer in the middle represent core of 11-mer duplex with 3'-next flanking base in blue and the underlined base indicates the change in the original 11-mer G*CT sequence (5'-CCATCGCTACC--3'). * impurity

Figure 4: Dynamic ¹⁹F NMR spectra of FABP-modified 19-mer (a) G*CT and (b) G*CA duplexes at different primer positions i.e. n-1, n, n+3 and full (see Fig. 1d).

Figure 5: Dynamic ¹⁹F NMR spectra of FAF-modified 19-mer (a) G*CT and (b) G*CA duplexes at different primer positions i.e. n-1, n, n+3 and full (see Fig. 1d).

Figure 6: CD spectra of (a) FABP- and (b) FAF-modified 19-mer G*CT (right) and G*CA (left) duplexes at different primer positions i.e. n-1, n, n+3 and full (see Fig. 1d).

Figure 7: Running start experiments across (a) FABP- and (b) FAF-modified templates (G*CA and G*CT) in presence of all four dNTPs (250 μ M each) and Kf-exo⁻ (2.5 nM) at different time intervals (M, marker).

Figure 8: Running start experiments across (a) FABP- and (b) FAF-modified templates (G*CA and G*CT) in presence of all four dNTPs (250 μ M each) and Dpo4 (50 nM, FABP; 100 nM, FAF) at different time intervals (M, marker).

Figure 9: Oligonucleotide sequences used for steady state kinetics.

Figure 10: Plots showing the extension frequency vs. insertion site: (a) FABP, (b) FAF in match (left) and mismatch (right) series.

Table Legends:

Table 1: Summary of the conformational profile exhibited by FABP and FAF at different primer positions in GCA and GCT sequence contexts.

Table 2: Kinetic parameters for dGTP insertion opposite the templating C one nucleotide before the FABP lesion site (n-1).

Table 3: Kinetic parameters for nucleotide insertion opposite the FABP lesion (n).

Table 4: Kinetic parameters for correct dGTP insertion opposite the templating C one nucleotide after the FABP lesion site (n+1).

Table 5: Kinetic parameters for correct dTTP insertion opposite templating A at (n+3) position from the FABP lesion site.

Table 6: Kinetic parameters for dGTP insertion opposite the templating C one nucleotide before the FAF lesion site (n-1).

Table 7: Kinetic parameters for insertion opposite the FAF lesion (n).

Table 8: Kinetic parameters for correct dGTP insertion opposite the templating C one nucleotide after the FAF lesion site (n+1).

Table 9: Kinetics parameters for correct dTTP insertion opposite templating A at (n+3) position from the FAF lesion site.

Table 10: Summary of the insertion efficiency against FABP- and FAF-modified template at different primer positions in GCA and GCT sequence contexts.

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

5' - CCATCTTACCATC**G**CTACCATTCCCAATCACCATCATACCAATC - 3'
 ←-----*KF-exo*-----TAAGGGTTAGTGGTAGTATGGTTAG - 5'

G = FABP or FAF

Figure 8:

G = FABP or FAF

n-1 site: 5' - CCATCTTACCATC**GC****N**ACCATTCCCAATCACCATCATACCAATC -3'
NTGGTAAGGGTTAGTGGTAGTATGGTTAG -5'
↑
dGTP

n site: 5' - CCATCTTACCATC**GC****N**ACCATTCCCAATCACCATCATACCAATC -3'
GNTGGTAAGGGTTAGTGGTAGTATGGTTAG -5'
↑
dCTP or dATP

n+1 site: 5' - CCATCTTACCATC**GC****N**ACCATTCCCAATCACCATCATACCAATC -3'
XGNTGGTAAGGGTTAGTGGTAGTATGGTTAG -5'
↑
dGTP

n+3 site: 5' - CCATCTTACCATC**GC****N**ACCATTCCCAATCACCATCATACCAATC -3'
AG**X**GNTGGTAAGGGTTAGTGGTAGTATGGTTAG -5'
↑
dTTP

N: A or T; **X:** C or A; **G:** dG, dG-FABP or dG-FAF

Figure 10:

(a) FABP

(b) FAF

Table 1:

Duplex	Sequence Context	Conformational Profile			
		FABP		FAF	
n-1	G*CA	B*(82%)	S*(18%)	B*(32%)	S*(68%)
	G*CT	B*(70%)	S*(30%)	B*(33%)	S*(67%)
n	G*CA	B*(76%)	S*(24%)	B*(58%)	S*(42%)
	G*CT	B*(64%)	S*(36%)	B*(24%)	S*(76%)
n+3	G*CA	B(100%)		B(42%)	S(58%)
	G*CT	B(65%)	S(35%)	B(19%)	S(81%)
Full	G*CA	B(95%)	B*(5%)	B(36%)	S(64%)
	G*CT	B(67%)	S(33%)	B(20%)	S(80%)

Table 2:

Sequence context	k _{cat} (min ⁻¹)	K _m (nM)	V _{max} (nM/min)	<i>F_{ins}</i>
-GCA-	27.9±2.1	25.1±9.6	14.0±1.1	1
-G*CA-	10.9±1.2	44.7±20.3	5.4±0.6	0.22
-GCT-	24.5±1.7	30.0±10.1	12.2±0.8	1
-G*CT-	9.8±0.5	53.2±9.7	4.9±0.2	0.23

Table 3:

Sequence context	Incoming dNTP	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	F _{ins}
-GCA-	dCTP	11.41±0.5	1.47±0.4	5.7±0.3	1
	dATP	0.13±0.02	23.7±13.6	0.69±0.1	7×10 ⁻⁴
-G*CA-	dCTP	16.05±3.3	17.4±11.3	8.0±1.6	0.11
	dATP	0.04±0.004	16.9±8.6	0.2±0.02	3×10 ⁻⁴
-GCT-	dCTP	6.9±1.4	2.04±1.6	3.48±0.7	1
	dATP	0.20±0.01	8.3±2.8	1.01±0.05	7×10 ⁻³
-G*CT-	dCTP	0.85±0.13	13.5±6.8	0.42±0.07	0.09
	dATP	0.07±0.005	101±19.7	0.34±0.03	2×10 ⁻⁴

Table 4:

Sequence context	Nucleotide opposite G or G*	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	F _{ext}
-GCA-	C	17.08±1.42	0.019±0.002	8.54±0.71	1
	A				N.D
-G*CA-	C	7.09±5.85	821±985	3.55±2.93	1.003x10 ⁻⁵
	A				N.D
-GCT-	C	9.38±0.75	0.068±0.028	4.69±0.37	1
	A				N.D
-G*CT-	C	5.87±2.59	465±350	2.94±1.3	9.13x10 ⁻⁶
	A				N.D

Table 5:

Sequence context	Nucleotide opposite G or G*	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	<i>F_{ins}</i>
-GCA-	C	4.1±0.6	5.0±2.8	4.1±0.6	1.0
	A	3.02±0.58	4.25±3.6	6.04±1.2	0.87
-G*CA-	C	3.0±0.4	7.4±3.6	3.0±0.4	0.49
	A	0.24±0.02	14.3±3.6	4.9±0.4	0.02
-GCT-	C	5.5±0.4	0.8±0.3	5.5±0.4	1.0
	A	11.3±0.24	15.4±8.4	22.6±4.9	0.1
-G*CT-	C	2.2±0.2	3.9±1.7	2.2±0.2	0.08
	A	0.4±.001	29.0±4.6	8.06±0.5	.002

Table 6:

Sequence context	k _{cat} (min ⁻¹)	K _m (nM)	V _{max} (nM/min)	<i>F_{ins}</i>
-GCA-	20.3±1.7	31.0±12.9	10.2±0.9	1.0
-G*CA-	10.5±0.8	40.1±13.3	5.3±0.4	0.4
-GCT-	19.9±0.8	17.8±4	9.9±0.4	1.0
-G*CT-	12.8±1.3	37.5±16.4	6.4±0.7	0.31

Table 7:

Sequence context	Incoming dNTP	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	F _{ins}
-GCA-	dCTP	20.4±1.4	6.65±2.12	10.2±0.07	1
	dATP	0.13±0.004	22.8±2.3	0.69±0.02	1.0×10 ⁻³
-G*CA-	dCTP	1.9±0.15	13.3±4.5	0.94±0.07	0.046
	dATP	0.12±0.007	25±5.5	0.61±0.037	1.5×10 ⁻³
-GCT-	dCTP	14.7±2.4	14.5±10.8	7.35±1.4	1
	dATP	0.12±0.013	14.6±6.8	0.64±0.07	8.6×10 ⁻³
-G*CT-	dCTP	1.28±0.13	10.1±3.7	0.64±0.06	0.12
	dATP	0.05±0.007	3.18±2.4	0.28±0.03	1.7×10 ⁻²

Table 8:

Sequence context	Nucleotide opposite G or G*	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	F _{ext}
-GCA-	C	17.54±2.07	0.05±0.02	8.7±1.04	1
	A	0.018±0.017	63.33±15	0.37±0.03	8.7×10 ⁻⁷
-G*CA-	C	0.83±0.233	47.9±52.4	0.41±0.12	5.1×10 ⁻⁵
	A	0.02±0.002	64.5±17.5	0.44±0.04	1.008×10 ⁻⁶
-GCT-	C	9.7±0.9	0.01±0.006	4.86±0.48	1
	A	0.06±0.007	38.4±14.5	1.21±0.15	2.25×10 ⁻⁶
-G*CT-	C	0.42±0.19	8.18±21.0	0.21±0.09	7.4×10 ⁻⁵
	A	0.01±0.001	39.5±8.08	0.31±0.02	5.02×10 ⁻⁷

Table 9:

Sequence context	Nucleotide opposite G or G*	kcat (min ⁻¹)	K _m (μM)	V _{max} (nM/min)	<i>F_{ins}</i>
-GCA-	C	10.4±0.9	3.2±1.4	5.2±0.5	1.0
	A	13.8±1.0	8.93±2.4	13.8±1.0	4.8*10 ⁻¹
-G*CA-	C	11.5±1.9	24.1±11	5.7±1.0	1.5*10 ⁻¹
	A	1.2±0.1	29.4±8.0	11.9±1.3	1.3*10 ⁻²
-GCT-	C	9.1±0.6	1.00±0.4	4.5±0.3	1.0
	A	12.6±0.5	8.9±1.4	12.6±0.5	1.6*10 ⁻¹
-G*CT-	C	5.7±0.8	9.1±4.4	2.9±0.4	7.2*10 ⁻²
	A	1.4±2.0	30.2±1.2	13.7±0.2	5.2*10 ⁻³

Table 10:

Insertion site	Sequence Context		Insertion efficiency		Difference factor (FABP)		Difference factor (FAF)	
			FABP	FAF	G*CA/G*CT	G*CT/G*CA	G*CA/G*CT	G*CT/ G*CA
n-1	G*CA	dG*:dC	0.22	0.40	0.96		1.29	
	G*CT	dG*:dC	0.23	0.31		1.05		0.78
n	G*CA	dG*:dC	0.11	0.046	1.22		0.38	
		dG*:dA	3.0×10^{-4}	1.5×10^{-3}	1.5		0.088	
	G*CT	dG*:dC	0.09	0.12		0.82		2.61
		dG*:dA	2.0×10^{-4}	1.7×10^{-2}		0.67		11.3
n+1	G*CA	dG*:dC	1.0×10^{-5}	5.1×10^{-5}	1.1		0.67	
		dG*:dA	ND	1.0×10^{-6}	ND		2.0	
	G*CT	dG*:dC	9.1×10^{-6}	7.4×10^{-5}		0.91		1.45
		dG*:dA	ND	5.0×10^{-7}		ND		0.5
n+3	G*CA	dG*:dC	0.49	1.5×10^{-1}	6.1		2.1	
		dG*:dA	0.02	1.3×10^{-2}	10		2.5	
	G*CT	dG*:dC	0.08	7.2×10^{-2}		0.16		0.48
		dG*:dA	0.002	5.2×10^{-3}		0.1		0.4