2014

STRUCTURAL AND CONFORMATIONAL INSIGHTS INTO BULKY ARYLAMINE-INDUCED MUTAGENESIS

Lifang Xu
University of Rhode Island, fountainxu@gmail.com

STRUCTURAL AND CONFORMATIONAL INSIGHTS INTO BULKY ARYLAMINE-INDUCED MUTAGENESIS

BY

LIFANG XU

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

2014

DOCTOR OF PHILOSOPHY DISSERTATION

OF

LIFANG XU

APPROVED:

Dissertation Committee:

Major Professor Bongsup Cho

Roberta King

Brett Lucht

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2014

ABSTRACT

Cancer is the second deadliest disease in the United States. Over 100 different types of cancers exist, among which lung, breast and prostate cancers are those most frequently diagnosed. Genetic factors are important. However, exposures to tobacco smoke and environmental pollutants are considered to be responsible for 75%–80% of cancer. About 6% of cancer deaths every year in the US are reportedly to be directly linked to known carcinogen exposures. Therefore, it is important to study the mechanisms of how the environmental carcinogens trigger cancer initiation. Most chemical carcinogens are metabolized into reactive species in vivo to interact with DNA, consequently producing covalent DNA adducts. These harmful lesions can be removed by various repair systems including base excision and nucleotide excision repair machinery in the cell. However, unrepaired lesions can enter into cell's DNA replication cycle and generate various point and frameshift mutations. In particular, the latter represents a gain or loss of base pairs, which alters the genome information. As an example, mutations on the specific genes such as the tumor suppressor p53 may trigger cancer initiation.

Arylamine is known as an important group of environmental chemical carcinogens. Some members of this group, such as 4-aminobiphenyl (ABP), benzidine and 2-naphthylamine, are classified as human bladder carcinogens. These chemicals are found commonly in cigarette smoke, incomplete diesel exhausts, and hair dye products. 2-Aminofluorene is a prototype animal carcinogen that undergoes metabolic activation by liver enzymes to form electrophilic nitrenium ion to form two major C8 substituted DNA-adducts: *N*-(2′-deoxyguanosin-8-yl)-2-aminofluorene (dG-C8-AF) and *N*-(2′-

deoxyguanosin-8-yl)-2-acetylaminofluorene (dG-C8-AAF). Similarly, the human carcinogen ABP produces *N*-(2′-deoxyguanosin-8-yl)-4-aminobiphenyl (dG-C8-ABP). Encountering these lesions in a replicative or a bypass polymerase will result in different types of biological outcomes, such as error-free, error-prone, or frameshifts.

Manuscript I (published in *Chemical Research in Toxicology,* 2012) is a rapid report. In this communication, we used a real-time, label-free chip-based technique named surface plasmon resonance (SPR) to determine the binding interaction between the DNA replicative polymerase exonuclease-free Klenow fragment and three arylamine DNA lesions (FAF/FAAF/FABP). We designed biotin labeled DNA hairpin construct with modified lesions and immobilized the DNA on the streptavidin coated chip. The analyte Kf-exo$^-$ was added over the DNA surface in the presence or absence of dNTP. The results showed a tight binding between the enzyme and unmodified DNA with great dNTP selectivity. In contrast, the dNTP selectivity was minimal in adduct modified DNA. Moreover, lesion included DNA tended to have better and stronger binding than unmodified DNA.

Manuscript II (published in *Chemical Research in Toxicology,* 2014) contains the full details of **Manuscript I**. The full paper involves two 5'-flanking sequence (CG*A and TG*A), two adducts (FAAF and FABP), and two different polymerases (*E. coli* replicative polymerase Kf-exo$^-$ and human repair polymerase β). We employed the same SPR methodology to study the binding interaction and complementary ^{19}F NMR and primer steady-state kinetics. Results showed significant substrate specificity for Kf-exo$^-$ and polymerase β, which are double-stranded/single-stranded junction and a double-stranded DNA with a nucleotide gap structure, respectively. Tight binding with native

DNA was observed, as well as the high nucleotide selectivity. However, Kf-exo$^-$ binds tightly to lesion DNA, but not for polymerase β. A minimal nucleotide selectivity for modified was observed with both enzymes. Moreover, the dynamic ^{19}F NMR and primer steady-state kinetics results indicated the importance of lesion-induced conformational heterogeneity in polymerase binding.

In **Manuscript III** (to be submitted to *Journal of Molecular Biology*), we conducted a series of systematic studies to probe the conformational mechanisms of arylamine-induced -2 base deletion mutations frequently observed in the *Nar*I mutational hot sequence (5'---TCGGCG*C*N*---3'; *N*= dC and dT) of *E. coli* during translesion synthesis (TLS). We employed two well-characterized fluorinated bulky DNA lesions FAAF and FABP that were derived from the environmental carcinogens 2-aminofluorene and 4-aminbiphenyl. Our work focused primarily on elucidating the effects of lesion size, bulkiness, and overall topology and the 3'-next flanking base N in producing the bulge structure responsible for -2 frameshift mutations. Two chemical simulated TLS models were examined, in which the FAAF/FABP lesion is positioned at G_3 position of two 16-mer *Nar*I sequences, which were annealed systematically with increasing primer lengths in the full length and -2 deletion pathways. Their thermodynamic, conformational, and binding profiles at each elongation step were measured by various biophysical techniques including spectroscopic (dynamic ^{19}F NMR/CD), thermodynamic (UV-melting/DSC), and affinity binding (SPR). Results showed two different -2 bulge formations, which are triggered by the conformational stability of the G_3*: C base pair at the replication fork, as well as the nature of base sequences surrounding the lesion site. Each bulge structure exists in a mixture of "external solvent exposed" B-type (B-SMI) and "inserted solvent

protected “stacked” S-type (S-SMI), and their conformational rigidity increases as a function of primer lengths. The results indicate the importance of conformational stability, heterogeneity, and flexibility in the mechanisms of bulky arylamine-induced frameshift mutagenesis.

ACKNOWLEDGEMENTS

A completion of my doctoral program is the most significant milestone in my life. It would not be possible without the help and support from many people around me. First of all, I thank my major professor, Dr. Bongsup Cho, for giving me the opportunity to be his graduate student and for his guidance and encouragement during my program. Without his patience and immense support and input, this dissertation would not have been possible. He has been my mentor not just academically but also for my personal aspects. I am extremely grateful to learn so many things from him to be a better and mature person.

I would also like to thank my committee members, Dr. Roberta King, Dr. Brett Lucht, Dr. Mindy Levine and Dr. Navindra Seeram for their time and support. I also thank University of Rhode Island, College of Pharmacy and the National Institutes of Health (NIH) for supporting me with teaching and research assistantships.

I appreciate the past and present lab members, Drs. Vaidyanathan Ganesan, Sathyakam Patnaik, Sathyaraj Gopal, Vipin Jain for their constant support. In particular, I would like to convey my special thanks to Vaidya who was always available to help and encourage me. Thanks to all my friends especially Mengyun, Yajuan, Yixin and Wenjing who cheered me up during my difficult and challenging periods.

Finally, I have to thank my parents for supporting all my choices regardless right or wrong and thank their understanding of not being with them.

PREFACE

This dissertation was prepared following the standards of Manuscript format of "Guidelines for the Format of Theses and Dissertations" (University of Rhode Island). This dissertation consists of three manuscripts to meet the requirement of the department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island.

MANUSCRIPT-I: Binary and Ternary Binding Affinities between exonuclease-deficient Klenow fragment (Kf-exo$^-$) and Various Arylamine DNA Lesions Characterized by Surface Plasmon Resonance.

This manuscript has been published in '*Chemical Research in Toxicology*' August 2012.

MANUSCRIPT-II: Real-time Surface Plasmon Resonance Study of Biomolecular Interactions between Polymerase and Bulky Mutagenic DNA Lesions.

This manuscript has been published in '*Chemical Research in Toxicology'* September 2014.

MANUSCRIPT-III: A Systematic Spectroscopic and Thermodynamic Investigation of Slippage Mediated Frameshift Mutagenesis.

This manuscript has been prepared for submission to '*Journal of Molecular Biology*' for Publication.

APPENDIX: Binding Kinetics of DNA-protein Interaction using Surface Plasmon Resonance.

This appendix has been published in '*Nature Protocol Exchange*' May 2013.

TABLE OF CONTENTS

LIST OF FIGURES

MANUSCRIPT-I

MANUSCRIPT-II

MANUSCRIPT-III

APPENDIX

LIST OF TABLES

MANUSCRIPT-I

MANUSCRIPT-II

MANUSCRIPT-III

Manuscript I

Published in ***Chemical Research in Toxicology***, 2012, 25, 1568-1570

Binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo$^-$) and various arylamine DNA lesions characterized by surface plasmon resonance

V.G. Vaidyanathan, Lifang Xu and Bongsup P. Cho*

Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy,

University of Rhode Island, Kingston, RI 02881

***Correspondence to Bongsup P. Cho:**

Phone: +1 401 874 5024

Fax: +1 401 874 5766

E-mail: bcho@uri.edu

ABSTRACT

We used surface plasmon resonance (SPR) to characterize the binding interactions between exonuclease-free Klenow fragment (Kf-exo$^-$) and unmodified dG and dG adducts derived from arylamine carcinogens: fluorinated 2-aminofluorene (FAF), 2-acetylaminofluorene (FAAF), and 4-aminobiphenyl (FABP). Tight polymerase binding was detected with unmodified dG and the correct dCTP. The discrimination of correct versus incorrect nucleotides was pronounced with K_D values in order of dCTP << dTTP < dATP < dGTP. In contrast, minimal selectivity was observed for the modified templates with Kf-exo$^-$ binding tighter to the FAAF-dG (k_{off}: 0.02s^{-1}) and FABP-dG (k_{off}: 0.01s^{-1}) lesions than to FAF-dG (k_{off}: 0.04s^{-1}).

DNA is under constant assault by various endogenous and exogenous pathways, which result in different types of DNA damage. When a polymerase encounters a lesion, it can bypass by replicative polymerase, either inserting the correct base (error-free) or incorrect base (error-prone)[1]. The environmental arylamine carcinogens are known to form C8-substituted dG adducts *in vivo*. We have shown that these lesions exist in a mixture of the base-displaced stacked, major groove B-type, and wedge conformers, with each leading to potentially unique sequence-dependent mutation and nucleotide excision repair outcomes[2].

It is important to understand the nature of interactions between polymerase and DNA lesions. Crystal structure and kinetic analyses have been used to elucidate details of polymerase action at an atomic resolution[1]. However, similar structural details of bulky DNA lesions have been challenging due to difficulties with obtaining crystals[3,4].

Consequently, various alternative techniques such as fluorescence, circular dichroism spectroscopy, gel mobility shift assays, and footprinting assays have been used[5,6] However, these techniques are either qualitative or semi-quantitative, non-compatible with fast dissociation rates, and require labeling of at least one of the components of interest. Although gel-based assay is relatively simple and robust, samples of interest will not be in chemical equilibrium and the system's components are not amenable to testing across temperatures or salt concentrations[6]. Surface plasmon resonance (SPR) is a chip-based, label free solution technique that allows real-time monitoring of binding interactions between DNA and proteins[5-8].

In this report, a SPR study was conducted to examine polymerase interactions of DNA lesions derived from three fluorinated prototype arylamine carcinogens: 2-aminofluorene (FAF), 2-acetylaminofluorene (FAAF), and 4-aminobiphenyl (FABP) (Fig. 1c). We employed exonuclease-free *E. coli* DNA polymerase I Klenow fragment (Kf-exo$^-$) as it avoids complication of proofreading activity. The features of fluorinated arylamines as effective conformational probes are well documented[2,9]. The present study takes advantage of the sensitivity of Biacore T200 to conduct SPR analysis of the binary and ternary polymerase complexes of bulky carcinogen-DNA adducts.

Figure 1a and S1a show the construction scheme for a biotinylated hairpin-based template-primer strand on a gold sensor chip. The hairpin-DNA was used to improve stability of oligonucleotides during performance of kinetics experiments. Arylamine-modified 31-mer oligonucleotides were purified by HPLC and characterized by mass spectrometry (Fig. S2). The biotin-hairpin-template/primer strands were annealed, ligated, and purified by denaturing polyacrylamide gel (Fig. S1b). The incorporation of

dideoxythymidine (ddT) was carried out using Kf-exo$^-$ and the 3′ terminal ddT allowed capture of the ternary polymerase/template-primer/ dNTP complex without primer extension.

The kinetic assays were optimized with respect to regeneration buffer, surface density, and surface testing, as described elsewhere[10] (Fig. S3). The binding kinetics analysis was performed by injecting varying amounts of Kf-exo$^-$ to cover the hairpin template-primer DNA (Fig. 1b) coated on streptavidin surface in the absence (binary) and presence (ternary) of dNTPs (100 μM). The injections were repeated three times for each concentration in random, and the resulting data were fitted to the Langmuir model (1:1) (Fig. 2). From the fitting, binding constants (k_{on}, k_{off} and K_D) were calculated (Table 1 and S1) using Biacore's BIAsimulation software. The Chi-squared values for the 1:1 fitting were less than 1% of R_{max} (0.002–0.003 for all experiments with R_{max} in the range of 0.7–3.5RU) (Figs. S4 and S5). The K_D values for ternary systems were determined using affinity analysis as the association rate (k_{on}) reaches the near-diffusion limit. This procedure allowed the monitoring of interactions between unmodified or adducted DNA with different polymerases on a single chip. Furthermore, DNA over the chip surface was found to be stable for at least 7–10 days, without loss in binding activity under buffered reaction conditions.

The results from the binding assay (Fig. S6) are summarized in Table 1. The Kf-exo$^-$ bound tightly to unmodified DNA in the presence of a correct incoming dCTP opposite the templating dG. However, relative to dCTP binding, binding tightness was reduced by 30-, 60-, 34-, and 264-fold in binary, dATP, dTTP, and dGTP, respectively (Fig. 2b and Table 1). The discrimination ability of correct versus incorrect nucleotides was

significant, as the Watson-Crick base pair dCTP bound tightly and dGTP does not bind significantly. In contrast, the discrimination effect on Kf-exo$^-$ binding was weaker for binding to FAF than for binding to unmodified DNA. The specificity of binding between the correct dCTP and incorrect nucleotides, as well as for the binary system, differed by only 2- to 16-fold. The tightness of Kf-exo$^-$ binding in the presence of dCTP was reduced by 4-fold, as compared to that of the unmodified control.

Moreover, the difference in binding affinity between dCTP and dATP was less for FAF-dG (10-fold), as compared to that of unmodified DNA (60-fold) (Fig. 2b). The Kf-exo$^-$ bound more tightly to FAAF ($k_{off} = 0.02s^{-1}$) and FABP-dG ($k_{off} = 0.01s^{-1}$) lesion sites than to the unmodified control ($k_{off} = 0.13s^{-1}$) while k_{on} values are similar. However, discrimination between correct and incorrect nucleotides was not maintained with FAAF and FABP-dG, for which binding affinities differed by only 1- to 3-fold (Fig. 2b).

Highly specific binding of Kf-exo$^-$ to unmodified DNA in the presence of dCTP opposite a dG templating base is in line with the polymerase undergoing conformational change from an open to a closed system to form Watson-Crick base pairs[11]. However, Kf-exo$^-$ does bind weakly with incorrect nucleotides, probably retaining the open polymerase conformation. In particular, the binding of dGTP is very poor compared to other nucleotides.

To further confirm that the binding of polymerase to DNA is 1:1, theoretical R_{max} values were calculated and compared with experimental values. The data presented here are consistent with data from sedimentation studies in which polymerase was shown to bind template-primer junction in a 1:1 ratio[12]. Interestingly, the K_D value for Kf-exo$^-$ binding to FAF adducts was higher in the presence of dCTP than with unmodified DNA (Table 1),

indicating that the lesion prevents the nucleotide-induced, catalytically-favored closed conformation. Previous studies have shown that the carcinogenic aminofluorene orients into the energetically favorable solvent-exposed major groove, which causes less disruption at the replication fork, but may perturb the groove structures and the geometry in the active site of the polymerase[3].

The aforementioned crystal structure of AF on T7 DNA polymerase showed fuzzy electron densities around the carcinogenic aminofluorene moiety in line with sequence-dependent conformational heterogeneity in solution[4]. The present kinetics data also fit with previously published findings from a single nucleotide insertion assay study in which dATP was the next preferred nucleotide after dCTP[13].

The higher binding affinity of Kf-exo$^-$ to the bulky *N*-acetylated FAAF lesion, compared to unmodified DNA, could be due to the adduct perturbing the template-primer junction while maintaining some specific interactions with amino acids on the active site of the polymerase. It has been shown that the AAF lesion has two hydrogen bond interactions between the N^2-amino group of the modified guanine and Asp-534, as well as between the N_7-guanine and Arg-566[4]. In addition, the lesion adopts a syn-glycosidic conformation wherein the fluorene moiety is inserted between the hydrophobic pocket of the O-helix finger subdomain. These changes also keep the polymerase in the open and maintain a distorted conformation of the subdomain fingers, causing the Tyr-530 residue to occupy the binding region of the nucleotide and preventing interaction between the incoming nucleotide and polymerase[4]. The present data are also in agreement with previous results from tryptic digestion studies, in which the polymerase was shown to bind very tightly to unmodified DNA in the presence of the correct nucleotide and to be

insensitive to digestion; FAAF did not exhibit any additional stability in relation to the incoming nucleotide[14]. FAF adducts are known to exist in a sequence-dependent equilibrium of B and S conformers[2,9]. FABP is similarly N-deacetylated; however, its biphenyl moiety is not as coplanar as fluorene, thereby resulting in a lesser base-displaced stacked conformer population[15]. Consequently, FABP may behave similar to FAAF at the replication fork in the active site of a polymerase.

In summary, tight binding of Kf-exo$^-$ was observed with unmodified dG in the presence of a correct dCTP in this study. Nucleotide selectivity was pronounced with K_D values in the order of dCTP << dTTP < dATP < dGTP. In contrast, minimal selectivity was observed for the modified templates: Kf-exo$^-$ bound tightly to FAAF-dG and FABP-dG lesions as compared to FAF-dG. The SPR results for FAF and FAAF agreed with those obtained from gel-based assays,[16] demonstrating SPR as a powerful and superior tool for studying protein/DNA interactions with bulky DNA lesions as it provides k_{on} and k_{off} rates.

ASSOCIATED CONTENT

Supporting Information. The synthesis and mass spectrum of adducts; binding profiles; simulated data are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding author

***** Phone**:** +1 401 874 5024. Fax: +1 401 874 5766.

Email: bcho@uri.edu

Funding

This research is supported by NCI/NIH (CA098296) and NCRR/NIH (P20 RR016457).

Acknowledgements

We thank Dr. Paul Belcher (GE Health Sciences) for his valuable suggestions in developing methodologies.

Abbreviations

SPR, surface plasmon resonance; Kf-exo$^-$, Klenow fragment exonuclease deficient; FAF-dG, *N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene; FAAF-dG, *N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-acetyl-aminofluorene; FABP-dG, *N*-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl.

REFERENCES

(1) Guengerich, F. P. *Chem. Rev.* **2006**, *106*, 420-52.

(2) Meneni, S. R.; Shell, S. M.; Gao, L.; Jurecka, P.; Lee, W.; Sponer, J.; Zou, Y.; Chiarelli, M. P.; Cho, B. P. *Biochemistry* **2007**, *46*, 11263-78.

(3) Hsu, G. W.; Kiefer, J. R.; Burnouf, D.; Becherel, O. J.; Fuchs, R. P.; Beese, L. S. *J. Biol. Chem.* **2004**, *279*, 50280-5.

(4) Dutta, S.; Li, Y.; Johnson, D.; Dzantiev, L.; Richardson, C. C.; Romano, L. J.; Ellenberger, T. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16186-91.

(5) Ritzefeld, M.; Sewald, N. *J. Amino Acids* **2012**, *2012*, 816032.

(6) Dey, B.; Thukral, S.; Krishnan, S.; Chakrobarty, M.; Gupta, S.; Manghani, C.; Rani, V. *Mol. Cell. Biochem.* **2012**, *365*, 279-99.

(7) Stengel, G.; Knoll, W. *Nucleic Acids Res.* **2005**, *33*, e69.

(8) Tsoi, P. Y.; Zhang, X.; Sui, S. F.; Yang, M. *Analyst* **2003**, *128*, 1169-74.

(9) Meneni, S.; Liang, F.; Cho, B. P. *J Mol Biol* **2007**, *366*, 1387-400.

(10) Myszka, D. G. *J. Mol. Recognit.* **1999**, *12*, 279-84.

(11) Joyce, C. M.; Potapova, O.; Delucia, A. M.; Huang, X.; Basu, V. P.; Grindley, N. D. *Biochemistry* **2008**, *47*, 6103-16.

(12) Delagoutte, E.; Von Hippel, P. H. *J. Biol. Chem.* **2003**, *278*, 25435-47.

(13) Vaidyanathan, V. G.; Cho, B. P. *Biochemistry* **2012**, *51*, 1983-95.

(14) Dzantiev, L.; Romano, L. J. *Biochemistry* **2000**, *39*, 5139-45.

(15) Zhou, L., Rajabzadeh, M., Traficante, D. D., Cho, B. P. *J Am Chem Soc* **1997**, *119*, 5384-5389.

(16) Dzantiev, L.; Romano, L. J. *J. Biol. Chem.* **1999**, *274*, 3279-84.

Figure legends

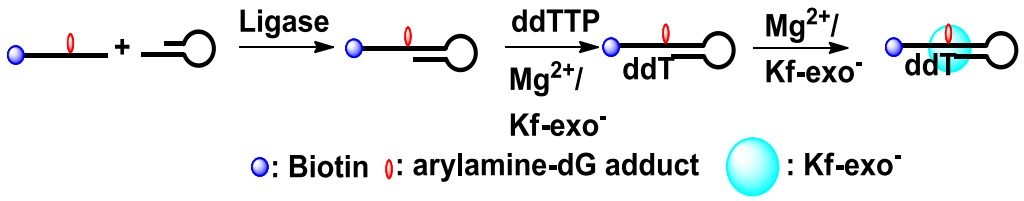
Figure 1: (a) Schematic representation of template-primer DNA constructs (b) oligonucleotide sequence used in the SPR assay and (c) the structures of arylamine-dG adducts

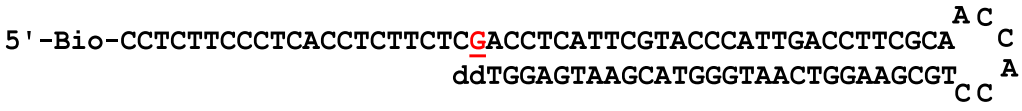
Figure 2: (a) Sensorgrams of Kf-exo- binding with unmodified and arylamine-modified DNA adducts (fitted curves were overlaid as red lines) (b) Plot of specificity ratio of binary or wrong nucleotide to correct dCTP vs. unmodified DNA and modified adducts

Table legends

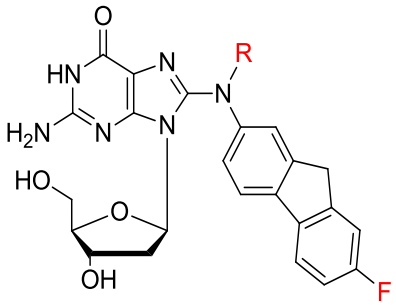
Table 1: Dissociation constants (K_D) for the unmodified dG and dG-arylamine adducts with Kf-exo$^-$ using steady-state affinity analysis

Figure 1:

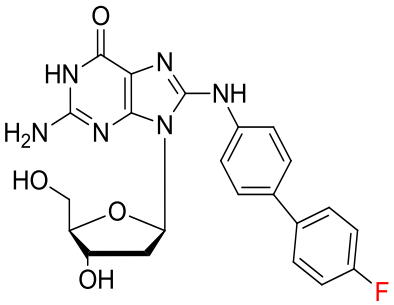




G: dG/dG-FAF/dG-FAAF/dG-FABP



FAF, R= H;
FAAF, R= $COCH_3$;



FABP

Figure 2:

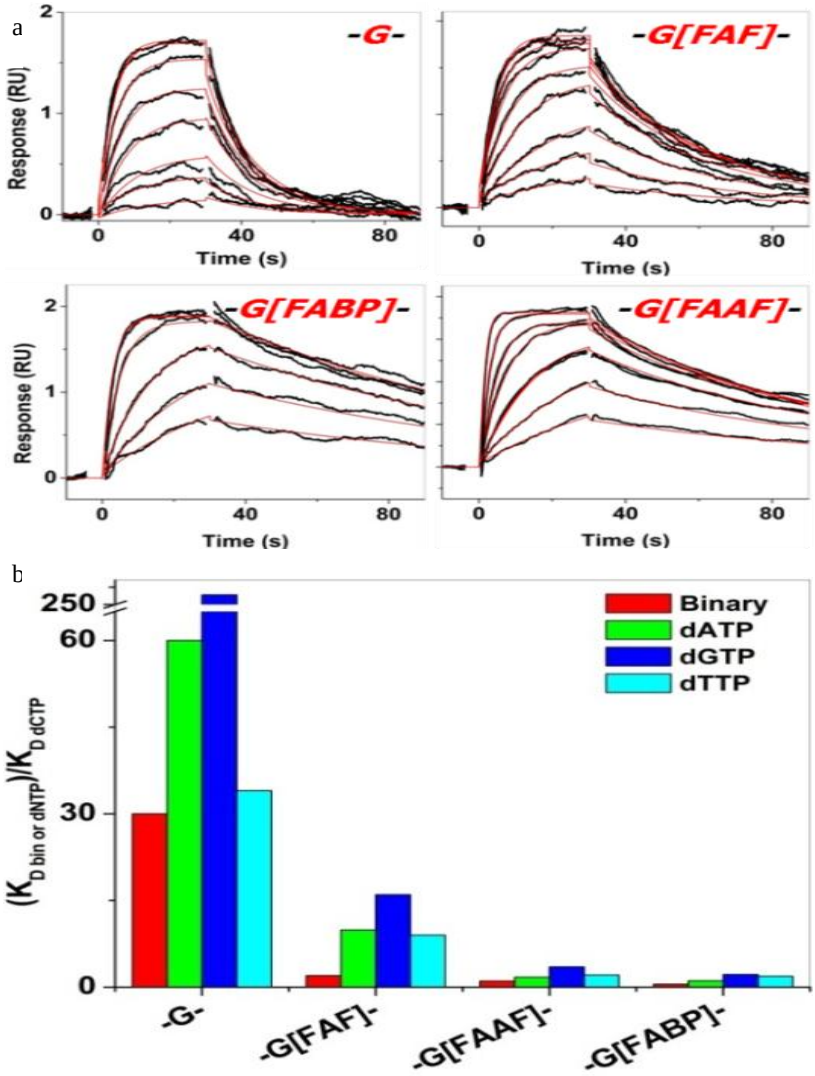

a
-G-
-G[FAF]-
-G[FABP]-
-G[FAAF]-
Response (RU)
Time (s)
b
Binary
dATP
dGTP
dTTP
250
60
30
0
-G-
-G[FAF]-
-G[FAAF]-
-G[FABP]-

Table 1:

Template	Binary	dCTP	dATP	dGTP	dTTP
-G-	1.5±0.5	0.05±0.1	3±2	13.2±12	1.7±0.5
-G[FAF]-	0.4±0.3	0.2±0.1	1.98±1.7	3.2±1.4	1.8±1.8
-G[FAAF]-	0.2±0.05	0.19±0.1	0.33±0.05	0.67±0.07	0.4±0.08
-G[FABP]-	0.14±0.09	0.29±0.1	0.31±0.1	0.63±0.11	0.54±0.17

‡K_D values are in nM

Manuscript II

Published in ***Chemical Research in Toxicology***, 2014, 27, 1796-1807

Real-time Surface Plasmon Resonance Study of Biomolecular Interactions between Polymerase and Bulky Mutagenic DNA Lesions

Lifang Xu[⊥], V.G. Vaidyanathan[⊥,¶] and Bongsup P. Cho*

Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States

[¶] Present Address: Chemical Laboratory, CSIR-CLRI, Adyar, Chennai 600020, India

[⊥] These authors contributed equally to this work

***Correspondence to Bongsup P. Cho:**

Phone: +1 401 874 5024

Fax: +1 401 874 5766

E-mail: bcho@uri.edu

ABSTRACT

Surface plasmon resonance (SPR) was used to measure polymerase binding interactions of the bulky mutagenic DNA lesions *N*-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl (FABP) or *N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene (FAAF) in the context of two unique 5'-flanking bases (CG*A and TG*A). The enzymes used were exo-nuclease-deficient Klenow fragment (Kf-exo$^-$) or polymerase β (pol β). Specific binary and ternary DNA binding affinities of the enzymes were characterized at sub-nanomolar concentrations. The SPR results showed that Kf-exo$^-$ binds strongly to a double strand /single strand template/primer junction, whereas pol β binds preferentially to double-stranded DNA having a one-nucleotide gap. Both enzymes exhibited tight binding to native DNA, with high nucleotide selectivity, where the K_D values for each base pair increased in the order dCTP << dTTP ~ dATP << dGTP. In contrast to pol β, Kf-exo$^-$ binds tightly to lesion-modified templates; however, both polymerases exhibited minimal nucleotide selectivity towards adducted DNA. Primer steady-state kinetics and ^{19}F NMR results support the SPR data. The relative insertion efficiency f_{ins} of dCTP opposite FABP was significantly higher in the TG*A sequence compared to CG*A. Although the Kf-exo$^-$ was not sensitive to the presence of a DNA lesion, FAAF-induced conformational heterogeneity perturbed the active site of pol β, weakening the enzyme's ability to bind to FAAF adducts compared to FABP adducts. The present study demonstrates the effectiveness of SPR for elucidating how lesion-induced conformational heterogeneity affects the binding capability of polymerases, and ultimately the nucleotide insertion efficiency.

.INTRODUCTION

Polymerases are critical to the replication and repair of DNA.[1] While replication of DNA is an essential first step for cell division, repair of DNA is needed when insults such as UV rays, environmental toxins, and some drugs chemically modify DNA.[2] These modifications can yield a diverse array of mutations.[3] To understand the mechanisms of DNA replication and repair, it is crucial to understand how a polymerase processes DNA lesions.[4, 5]

As part of ongoing carcinogenesis research, and to understand the mechanisms of DNA mutation and repair, we have been studying how the bulky and mutagenic arylamine-DNA lesions (Figure 1a) interact with a polymerase or a repair protein.[6-11] Using ^{19}F NMR, microcalorimetric and other biophyisical methods, we have shown that the arylamine lesions adopt three unique conformations: base-displaced stacked (*S*), major groove *B*-type (*B*) and minor-groove wedge (*W*) depending on the location of the lesion (Figure 1b).[10-13] The relative populations of *S*-, *B*-, and *W*- conformers depend on the nature of attachment on the central nitrogen (*N*-acetyl *vs.* *N*-deacetylated) and the hydrophobic carcinogen ring moiety (planar *vs.* twisted) as well as the base sequences (flanking *vs*. near long-range) surrounding the lesion.[13-15]

It has been shown that most replicative polymerases easily bypass the planar and *N*-deacetylated aminofluorene (AF) adducts after a brief stall at the lesion site. On the other hand, the bulkier *N*-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF) analogs cannot be readily bypassed, and thus stall DNA synthesis.[16] In vitro studies with X-family polymerase *β*, AAF adducts lead to -2 base deletion mutations, while AF extends full length primers.[17] A recent study via single-molecule fluorescence spectroscopy

showed that high-fidelity polymerases cannot extend a primer whose terminus occurs across from AAF.[18] In *E. coli*, AAF adducts results mostly in frameshift mutations, while both AF and AAF adducts cause point mutations .[19] In mammals, both adducts afford point mutations. [17] This difference in mutagenic profiles has been attributed to the presence of a bulky acetyl group on the central nitrogen, which causes the AAF adduct to adopt a *syn* conformation. [20] In contrast, the AF adduct adopts an *anti*-/*syn*- conformation, while the *N*-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl (FABP) adduct adopts exclusively an *anti*-conformation. [9] Other factors influencing adduct-induced mutations include topology, insertion of the nucleotide opposite the lesion site, and the characteristics of the polymerase.[21, 22]

Numerous crystal structure and kinetic analysis studies are available and provide information on actions of native[23-25] and damaged[2, 5, 26-34] DNA with various polymerases. However, only few examples of replicative polymerases complexed with bulky arylamine modified-DNA are available with atomic resolution details, [26, 27] presumably due to difficulties with obtaining crystals. High-resolution solution NMR can offer dynamic information alternative to the static crystallography.[35,36] However, some bulky DNA lesions cause conformational variation in the DNA and upon binding with a polymerase, which introduces additional challenges to the use of this method.[14,15,37-39] As a result, most NMR studies thus far are limited to adducted DNA without full presence of polymerases and repair proteins.[40,41] Theoretical/molecular dynamic simulations in conjunction with limited NMR and crystal data have been useful. [4,40, 42, 43]

Other available techniques for biomolecular interactions such as electrophoretic mobility (gel shift or gel retardation assay), and filter-binding assays provide valuable

information on binding affinity. However, these approaches either contribute little or no insight on the kinetic parameters underlying complex formation. Moreover, these techniques require strenuous work to determine binding parameters.[44] In addition, gel assays do not allow the samples of interest to be in chemical equilibrium due to fast dissociation rate during electrophoresis, and thus it is difficult to measure proper binding kinetics and thermodynamics.[44] Finally, microcalorimetry such as isothermal titration calorimetry (ITC) is a fast and robust method that certainly could be used to characterize binding interactions and the thermodynamics of polymerase DNA interactions in free solution, but low affinity interactions would require higher protein concentrations.[44]

Surface plasmon resonance (SPR) is a powerful, chip-based, and label-free solution technology that can provide real-time information on kinetics and thermodynamics.[44-48] SPR relies on changes in the refractive index that are due to changes in mass, and can thus measure a small difference in binding (K_D) at sub-nanomolar level. SPR is thus ideal for probing interactions of binary and ternary polymerase-DNA interaction. We have recently communicated our initial SPR work on the binding affinities of Kf-exo$^-$ to arylamine DNA lesions.[6,49] Subsequently, a similar study was conducted to elucidate how FAF lesions affect the active site conformation of the human repair enzyme pol β, and how the structure and sequence of the DNA affects its ability to be repaired.[7]

In the present study, we are providing a complete set of SPR data on the binding of Kf-exo$^-$ or pol β to FAAF and FABP lesions in two different sequences (CG*A and TG*A). To complement the SPR binding results, we also conducted dynamic ^{19}F NMR as well as steady-state nucleotide insertion kinetics. The results are discussed in terms of adduct-induced conformational heterogeneity, the effect of the 5′-flanking base sequence,

substrate specificity, and the nature of a polymerase. The purpose of the present paper is two-fold: 1) to give the full details of our previous SPR work ("Rapid Report")[6] and 2) to introduce SPR to the chemical toxicology community as a powerful alternative to existing techniques for investigating protein-DNA interactions. As a result, the choice of polymerases used in the present study was based largely on the experimental systems in our previous work.[7,8,50] Obviously, future SPR studies should be expanded to a range of Y-family bypass polymerases, which is more likely to be involved in replication of bulky DNA lesions.

MATERIALS AND METHODS

DNA sequences containing 5′-biotin labeled 31-mer oligonucleotides, phosphorylated 52-mer hairpin and 21-mer complementary sequences (Figure 2b, c) were purchased from Operon (Eurofin, Huntsville, AL) in desalted form and purified by reverse phase high-performance liquid chromatography (RP-HPLC). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA) and used as received. The HPLC system was consisted of a Hitachi EZChrom Elite HPLC system with an L2450 diode array detector and a Clarity column (10 mm × 150 mm, 3 μm) (Phenomenex, Torrance, CA). The mobile phase system involved a 20 min linear gradient profile from 3 to 16% (v/v) acetonitrile with 100 mM ammonium acetate buffer (pH 6.5) at a flow rate of 2.0 mL/min. Kf-exo$^-$ (D424A) and pol β were received as gifts from Dr. Catherine Joyce (Yale University, New Haven, CT) and Dr. William Beard (NIEHS, Research Triangle Park, NC).

^{19}F NMR

Approximately 70 μM of a FAAF- or FABP-dG modified 16-mer template was annealed with a 9-mer primer in a 1:1 molar ratio to produce ds/ss junction containing duplexes (Figure 3). The samples were lyophilized and dissolved in 300 μL of typical pH 7.0 NMR buffer containing 10% D_2O/90% H_2O with 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA. All ^{19}F NMR spectra were recorded using a dedicated 5 mm $^{19}F/^{1}H$ dual probe on a Varian 500 MHz spectrometer operating at 476.5 MHz, using acquisition parameters described previously.[11,51,52] The spectra were acquired in the ^{1}H-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 at two different temperatures, 5 and 25 °C.

Primer extension assay

Standing start experiments

Single nucleotide/full length extension experiments for both FABP- and FAAF-dG adducts in Kf-exo$^-$ were performed as described previously.[8] Briefly, the 9-mer primer was 5′-radiolabeled using [γ-^{32}P] ATP and T4 polynucleotide kinase (T4 PNK) following the manufacturer's protocol. The ^{32}P-labeled primer (50 pmol) was annealed to either an unmodified or adducted template oligonucleotide (60 pmol) by heating to 95 °C for 5 min and then slowly cooling to room temperature in 3 h. For pol β assays, 1 nt-gap was generated by adding downstream 9-mer primer with 5′-phosphate group while annealing with radiolabeled primer (9-mer) and template (19-mer).[7] The ds/ss primer-template sequence (20 nM) was incubated with Kf-exo$^-$ (0.5 or 1.0 nM) for 5 min to form a binary complex in Tris buffer (Tris, 50 mM pH 7.4; BSA, 50 μg/mL; 5% (v/v) glycerol). The

reaction was initiated by adding a dNTP (100 μM)/$MgCl_2$ (5 mM) solution to a binary mixture and incubated at 22°C for 10 min. The reaction was arrested with gel loading buffer (containing 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) electrophoresed at 2500 V for 4 h. The gel was exposed on a Kodak phosphor imaging screen overnight and scanned with a Typhoon 9410 variable mode imager.

Steady-state kinetics analysis

To determine the efficiency of dCTP insertion opposite the adducted site, steady-state kinetic parameters for incorporation of the nucleotide opposite the unmodified and FABP-modified templates were determined by using the reported literature procedures.[7,8] The reactions were performed with pol β (0.5 nM) and oligonucleotide (20 nM) at 22°C. For the unmodified sequence, reactions were performed in shorter time period of 0.5-10 min for nucleotide incorporation and up to 30 min in the case of modified templates. The band intensities were quantitated using ImageQuantTL from GE Healthcare. 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). The kinetic parameters k_{cat} and K_m were determined as described earlier.[7, 8]

SPR Measurements

Arylamine-modified hairpin template/primer constructs

The modification of 5′-biotin CGA/TGA sequences (31-mer) was carried out using the previously reported procedures (Figure 2)[6, 7] and the modified products were purified by by RP-HPLC and characterized by MALDI-TOF mass spectrometer. Biotinylated

unmodified (20 μM) or modified 31-mer (20 μM) was annealed with 20 μM of 52-mer hairpin by heating to 95°C for 5 min and cooling down to room temperature (Figure 2). The annealed mixture was ligated by using 4000 U T4 DNA ligase in 1× ligase buffer for 16 h at room temperature. The ligated 83-mer oligonucleotide was purified by 10% denaturing polyacrylamide gel (Figure S1) and extracted using crush and soak method. The extracted oligonucleotide was desalted using Illustra G-25 spin column. The desalted oligonucleotide was incubated with 2', 3'-dideoxy-thymidine-5'-triphosphate (ddTTP) (1 mM) in the presence of Kf-exo$^-$ (1 μM) and 5 mM $MgCl_2$ for 12 h. The dideoxy-terminus DNA was purified by RP-HPLC (Figure S2) after precipitation of protein using phenol-chloroform-isoamyl alcohol (25:24:1) followed by ethanol extraction.

Characterization of oligonucleotides by MALDI-TOF

Either biotinylated 31-mer, 83-mer or 84-mer DNA sequences (100 pmol) was mixed with 2 μL matrix containing 1 μL of 3-hydroxy picolinic acid (3-HPA) (50 mg/mL dissolved in acetonitrile/water 50% v/v) and 1 μL of diammonium hydrogen citrate (DAHC) (50 mg/mL dissolved in acetonitrile/water 50% v/v). MALDI-TOF experiments were performed using Axima Performance from Shimadzu Biotech. The mass spectrometric measurement of 31-mer oligonucleotides was carried out in a reflectron positive mode. The calibration of the instrument in reflectron positive mode was performed using low molecular weight oligonucleotide or peptide standard calibration kit. For high molecular weight oligonucleotides (>10,000 Da), calibration was done in a linear negative mode using 52-, 80-, 90-, 100-mer standards with laser power 120 in order to enhance the signal intensity. The spectral data was processed by using Shimadzu

Biotech MALDI-MS software with processing parameters as follows: smoothing filter width as 20 channels; baseline filter width as 80 channels and double threshold.

DNA coating on biosensor chip

SPR measurements were conducted with Biacore™ T200 (GE Healthcare). A carboxymethylated dextran coated CM5 chip supplied by GE Healthcare was used to immobilize streptavidin (SA) via the amine coupling kit on flow cells by following the previously reported literature.[6,7,49] The EDC/NHS mixture was injected over the surface for 7 minutes followed by SA (50 μg/mL dissolved in sodium acetate buffer, pH 4.5). The unreacted reactive esters were blocked with 1 M ethanolamine for 7 min. The running buffer used for immobilization was 1× HBS-EP$^+$ buffer containing 10 mM Hepes (pH 7.4), 150 mM NaCl, 3 mM EDTA and 0.05% non-ionic surfactant P20. The flow cells were immobilized with SA around 2,500 RU. After SA immobilization, the surface was washed with 50 mM NaOH for 60 s pulse and repeated for 5 times to remove the free SA until the change in response unit reaches below 20 RU. The surface was stabilized by injecting 3-4 times running buffer followed by equilibration with running buffer for 1 h. The 84-mer biotinylated DNA-hairpin sequences of either unmodified or adducted DNA (0.25-0.3 nM) were injected over the flow cells 2 or 4 for 60-120 s individually to achieve 0.7-3.5 RU. The flow cells were washed with running buffer to remove the unbound DNA and to stabilize the surface. Before conducting kinetics experiments, 1 mM dideoxythymidine triphosphate (ddTTP) in the presence of Kf-exo$^-$ (1 μM) and 5 mM $MgCl_2$ was injected over the surface for 5 min followed by 0.05% SDS to remove the polymerase. For pol β experiments, 1 nt-gap was created by using the same DNA

coating approach and in addition corresponding downstream complementary sequence (21-mer) containing 5′-phosphate group (2 nM) was injected over the surface for 5 min.

Real-time kinetic analysis

Kf-exo$^-$ was injected with or without dNTPs (100 μM) over the DNA surface in random order (neither ascending nor descending concentrations). Each concentration was repeated twice. For binary system, varying concentrations of Kf-exo$^-$ (0-10 nM) prepared in running buffer containing 1× HBS-P$^+$ along with 100 μg/mL bovine serum albumin (BSA) and 5 mM $MgCl_2$ was used. The polymerase was injected for 30 s with flow rate of 100 μL/min followed by dissociation of polymerase. The surface was regenerated using 0.05% SDS with flow rate of 100 μL/min and injection time was 30 s followed by extra wash with running buffer. After regeneration of the surface, the surface was stabilized with running buffer for 15 min. Initially three startup steps with running buffer and four times of zero concentration injection were performed to condition the surface. For ternary system, individual dNTP (100 μM) was mixed with varying concentrations of Kf-exo$^-$ and injected over the surface. The sensorgrams were double referenced and fitted using a 1:1 Langmuir model. The binding affinity constants (K_D) for binary and ternary systems were calculated using steady-state affinity analysis in BIAevaluation software v1.0 as the association rate for the ternary system particularly with dCTP and unmodified dG reaches near diffusion limit. The sensorgrams for binary systems were globally fitted with BIAsimulation Basic kinetics module software by using experimental k_a and k_d values (Figure S7).

Similar experiments were carried out for pol β interaction studies with adduct present

both at the non-gapped duplex DNA and 1 nt-gap DNA. Single nucleotide (1 nt)-gapped DNA was generated by annealing corresponding downstream primer. For non-gapped DNA, the concentration of pol β was varied up to 1000 nM while for 1 nt-gap was 0-100 nM depending on dG adduct embedded. The binding constants were obtained using 1:1 Langmuir model.

RESULTS

Model hairpin template/primer constructs

An overall scheme for the construction of the biotinylated hairpin-based template-primer strands is depicted in Figure 2a. FABP- or FAAF-modified biotin-31-mer oligonucleotides were prepared according to published procedures.[7, 10, 12] The 52-mer hairpin-DNA was annealed and ligated to the biotinylated 31-mer (Figure 2b, c). ddTTP was incorporated at the 3′ primer terminus using Kf-exo$^-$.[25, 53] The hairpin structure was created to improve the thermal stability of the oligonucleotide constructs on a gold chip during kinetics experiments. As a result, the same oligonucleotide constructs could be used multiple times with different polymerases and buffer conditions. Finally, the lesion was positioned at the 22nd base, with 21 bases on the 5'-side and 28 bases on the 3'-side, in order to avoid close contact between the polymerase and the chip surface. The resulting template/primer strands, containing the biotinylated 84-mer hairpin, were purified by denaturing polyacrylamide gel (Figure S1) and used for further study.

MALDI-TOF spectrum, obtained in reflectron mode, of the FAAF-modified biotin-TG*A- 31-mer sequence is shown in Figure S3. A distinctive peak at 9841.30 Da is in close agreement with theory (9839.90 Da, $\Delta m/z$: +1.40) and the inset is a linear negative

mode spectrum. The inset of b at 25925.76 Da corresponds to the 83-mer strand consisting of the biotin-31-mer -TG[FAAF]A- and the 52-mer hairpin in the absence of ddT at the primer terminus (theoretical 25923.00: Δ*m/z*: +2.76). The inset of c at 26206.70 Da corresponds to the 84-mer strand formed by adding ddT to the primer terminus of the 83-mer strand (26211.00 Da: Δ*m/z*: -4.30). The corresponding TG[FABP]A, CG[FAAF]A and CG[FABP]A sequences were similarly characterized (Figure S4-S6). All of the calculated and experimental *m/z* values are shown in Table S1.

^{19}F NMR

To examine lesion-induced conformational heterogeneity, we measured ^{19}F NMR spectra of modified 16/9-mer template/primer duplexes. As shown in Figure 3, the ^{19}F NMR spectra of FABP- and FAAF-modified duplexes in the CG*A and TG*A sequences are compared at 25 °C. FABP-duplexes exhibited a single peak at -116.4 ppm in both sequences, which is consistent with the chemical shift range observed previously for the *anti-B*-type FABP conformer.[9] The bulky FAAF displayed three ^{19}F signals with two prominent peaks of similar intensity at around -114 to -116 ppm, for both sequences. We have previously reported the chemical shift ranges that correspond to the *B*-, *S*-, and *W*-conformers of FAAF-modified duplexes, i.e., -115.0 to -115.5 ppm for the *B*-conformer, -115.5 to -117.0 ppm for the *S*-conformer, and -117.0 to -118.0 ppm for the *W*-conformer.[11,13] Hence, the present FAAF-induced heterogeneity could be a variation of the *B*/*S*/*W* heterogeneity. In contrast to the aforementioned study, however, the ^{19}F signals in the present study are derived from the lesions at the ds/ss junction, not fully paired double helical duplexes.[8,12] The relative shielding of ^{19}F signals and the narrow the narrow chemical shift range (~ 2 ppm) in the present work are probably due to the

flexible lesions at the ds/ss junction. As a result, we could not unequivocally assign the signals to the *B*-, *S*-, or *W*-conformer.

Primer extension assay

Single nucleotide incorporation was carried out using the *E. coli* exonuclease-deficient Klenow fragment (Kf-exo$^-$) and the human base excision repair polymerase β (pol β) (Figure 4). Like any other high-fidelity replicative polymerase, Kf-exo$^-$ prefers the ds/ss replication fork as a template/primer DNA substrate. In the unmodified DNA control, the primer was immediately elongated to full length in the presence of all four nucleotides and Kf-exo$^-$ (data not shown). With the FABP-modified template, however, primer elongation was largely stalled at the lesion site, with some insertion of the correct dCTP opposite the lesion (Figure 4a).

Unlike Kf-exo$^-$, pol β prefers a single nucleotide gap as a substrate.[54, 55] With pol β, there was no full extension of either the unmodified (not shown) or FABP-modified template (Figure 4b). We observed preferential dCTP incorporation opposite the lesion. As for FAAF, no nucleotide insertion was observed with either Kf-exo$^-$ or pol β, even at high enzyme concentrations or longer incubation period (data not shown) because the lesion had completely blocked elongation.

Steady-state kinetics

We conducted steady-state experiments to investigate the impact of conformational heterogeneity on nucleotide insertion kinetics. The results for Kf-exo$^-$ and pol β are summarized in Tables 1 and 2, respectively. To examine the influence of lesions, we used the relative insertion efficiency f_{ins}, which was defined as $(k_{cat}/K_m)_{\text{modified or mismatched}}$

/$(k_{cat}/K_m)_{unmodified}$. With Kf-exo$^-$, the f_{ins} of dCTP opposite -CG[FABP]A- was 500-fold lower than that of the unmodified control (Table 1). This is contrasted with -TG[FABP]A- which was reduced only 33-fold. In the pol β assay (Table 2), the f_{ins} of dCTP opposite FABP in the CGA sequence was 142-fold lower than that of the control, while in the TGA sequence the f_{ins} was 59-fold lower than that of the control. These results indicate that the nucleotide insertion efficiency is consistently greater in the TGA sequence compared to the CGA sequence, regardless of the polymerase structure. We were unable to perform similar steady-state kinetics experiments for FAAF because this lesion caused a major blockage at the replication fork.

SPR binding experiments

DNA coating and mass transport limitation studies After activation with streptavidin (SA), flow cells 1 and 3 were retained as blank references, and DNA was coated on the SA surface of flow cells 2 and 4. Surface testing, regeneration buffer scouting, and the mass transport limitation test were performed before the kinetics experiments as described previously.[6] DNA coating at 0.7 resonance units (RU) did not show any influence of mass transport; an increase in flow rate of the analyte did not alter the association rate. However, at 10 RU, mass transport became a limiting factor, as the association rate deviated with the flow rate of the analyte (data not shown). Based on this study of mass transport limitation, all the experiments were carried out in the DNA coating range between 0.7 and 3.5 RU.

Kf-exo$^-$ The sensorgrams for the binary binding between Kf-exo$^-$ and the unmodified TGA controls or the modified TG*A oligonucleotide constructs are shown in

Figure 5a. We performed steady-state affinity analysis of the binary and ternary complexes in the presence of four dNTPs (Figure 6). A similar set of results for the CGA sequence have been reported previously[6] and the results on the binding affinity of Kf-exo$^-$ to both TGA and CGA sequences are summarized in Table 3.

As for the unmodified controls, Kf-exo$^-$ binds tightly in both sequences in the presence of the correct dCTP. The affinity of binding for the CGA sequence was reduced by 30-, 62-, 264-, and 34-fold in binary, dATP, dGTP and dTTP, respectively, compared to the correct dCTP binding (Table 3). Similar results were obtained for TGA, where the binding affinity was reduced by 15-, 39-, 180-, and 40-fold in binary, dATP, dGTP and dTTP, respectively (Table 3). These results are consistent with those of the nucleotide insertion assay, which showed preferential insertion of the correct dCTP.

Kf-exo$^-$ bound strongly to the modified TG*A templates. In the TG*A sequence, the K_D value for FABP was 4.9- fold greater than the control, and the K_D value for FAAF was 8.8-fold greater than the control. Similar changes were observed in the CG*A sequence, where the K_D for FABP was 10.8-fold larger than for the control, and the K_D for FAAF was 7.2-fold larger than for the control. These differences are primarily due to the much slower dissociation rates observed for the modified template/primer for both the CG*A sequence (FAAF, k_d: 0.02 s^{-1}; FABP k_d: 0.01 s^{-1}) and the TG*A sequence (FAAF, k_d: 0.01 s^{-1}, FABP, k_d: 0.01 s^{-1}). The net stabilization energies were positive and ranged from 1.10 to1.47 kcal/mol (Table S2).

Nucleotide selectivity was low in the modified ternary complexes. K_D for the correct nucleotide was 0.19 – 0.25 nM with FAAF and 0.29 – 0.30 nM with FABP, while for the incorrect nucleotide, K_D was 0.28 – 0.67 nM with FAAF and 0.31 – 0.66 nM with FABP.

Pol β For pol *β*, binding assays were performed on two distinct substrates: non-gapped ds/ds and 1 nt-gap. The results for the binary and ternary systems on both CGA and TGA sequences are summarized in Table 4. Weak binding was observed for the non-gapped DNA, with K_D values of ~0.8 μM (data not shown). In contrast, the binding affinity of pol *β* increased 1,000 fold with the 1 nt-gap.

As for the unmodified controls, pol *β* binds to the correct dCTP more tightly. The binding affinity for the dCTP is 2.7-fold higher in the TGA sequence, and 4.5-fold higher in the CGA sequence (Table 4). In contrast to Kf-exo$^-$, the binding in the binary complex between the modified template and pol *β* is less tight than that in the complex containing the unmodified template, where the differences in binding are approximately 3-fold for FABP and 5- to 6-fold for FAAF, respectively. Similar to Kf-exo$^-$, the binary complex with FAAF showed slower off rates (k_d: 0.01 s^{-1}) with pol *β* in both sequences. The curve fits for dG-FAAF (Figure 5b) are relatively poor: however, the residual plots for the dG-FAAF/plo *β* binary complex (Figure S10) indicate a good curve fit within 1% chi^2 values of R_{max}. The complexes with *B*-conformeric FABP exhibited unusually faster dissociation rates for both the CG*A and TG*A sequences, where the k_d values were 0.76 s^{-1} and 0.40 s^{-1}, respectively, and the negative net stabilization energy was -1.04 and -0.27 kcal/mol, respectively (Table S2).

Figure S9 show the sensorgrams for the ternary complexes between pol *β* and the FAAF- and FABP-modified CG*A constructs. We have recently reported a similar set of binding results for the *N*-deacetylated FAF.[7] With the correct nucleotide dCTP, the pol *β* binds 2.7- fold more tightly in the ternary complex than in the binary complex and ~3,000-fold more tightly than to the non-gapped DNA. The binding affinity to the

incorrect nucleotide was 4 to 5-fold lower than to the correct dCTP. The lesion in the 1 nt-gap reduced the binding affinity of pol β by 6-fold for FAAF and 3-fold for FABP, virtually eliminating the nucleotide selectivity of pol β at the lesion site. The affinity for pol β binding decreased in the order dG > FABP > FAAF.

DISCUSSION

In the present study, we have employed SPR to investigate the binary and ternary binding interactions of Kf-exo$^-$ and pol β to two prototype arylamine-DNA lesions (FABP and FAAF) in the context of two different sequences (CG*A and TG*A). Kf-exo$^-$ is a 68-kDa high fidelity replicative A-family bacterial DNA polymerase,[56] which carries a polymerase and 3'-5'-exonuclease activities and has been used extensively as a model enzyme for studying adduct-induced DNA synthesis. Pol β is the smallest (39 kDa) eukaryotic polymerase, belonging to the X-family of base-excision repair DNA polymerases, and has been characterized extensively.[57] With pol β, primer extension past AAF adduct was blocked, but full length products were shown to contain exclusively -2 deletion mutations.[17] Although its role is limited in base excision repair, pol β has been additionally implicated in the replication of various DNA damage. For example, deregulation of pol β may enhance the genetic instability induced by bulky lesions such as *cis*-platin[32] and UV radiation.[33] Pol β can also bypass abasic site[58] and bulky polyaromatic hydrocarbons adducts.[34] FABP and FAAF are C8-substituted dG adducts adducts which contain structurally unique arylamine structures, i.e., *N*-acetylated/coplanar-fluorene and *N*-deacetylated/twisted-biphenyl, respectively (Figure 1a). Finally, the two sequences (CG*A *vs.* TG*A) were selected because of their marked difference in the *S*/*B* population ratios observed with the *N*-deacetylated FAF.[11] The

SPR results, along with data from ^{19}F NMR and steady-states primer kinetics, elucidate how lesion-induced conformational heterogeneity alters the binding capacity of a polymerase and thus its nucleotide insertion efficiency.

Model hairpin oligonucleotide constructs for SPR binding assays

We constructed the 84-mer hairpin-based oligonucleotides for SPR (Figure 2) based on the following considerations. First, the incorporation of ddT at the 3′-end of the primer prevents the usual nucleophilic attack of the 3'-hydroxyl to the incoming dNTP, and thus blocks the formation of a phosphodiester bond.[25,59] This ensures the stability of the ternary complex polymerase/template-primer/dNTP for SPR measurements. Previous assays using gel electrophoresis, single-molecule FRET, or crystallography have consistently shown that the absence of 3′-OH at the primer terminus does not affect the affinity with which polymerases bind to binary and ternary complexes of DNA.[25,59] Second, while Kf-exo$^-$ requires a minimum of 11 bases, because it covers approximately 5 bases downstream from the primer/template junction and 6-7 bases upstream to the 3′-primer terminus,[60] pol β can operate on any length of DNA containing a 1 nt-gap.

Binary and ternary binding affinities with unmodified control DNA

We observed very tight binding of Kf-exo$^-$ with native unmodified dG, in the presence of the correct incoming nucleotide dCTP. This system exhibited high nucleotide selectivity, with K_D values increasing in the order dCTP << dTTP ~ dATP << dGTP (Table 3). The SPR results are in agreement with nucleotide insertion assays, which showed exclusive insertion of the correct dCTP over other dNTPs. Crystal structures usually indicate 1:1 DNA polymerase-DNA complexes. 2:1 and higher order complexes

have also been observed in solution by various biochemical and biophysical methods.[60] The stoichiometry, however, is highly concentration dependent. As shown in Figure S11, comparison between theoretical and experimental *Rmax* for pol β and Kf-exo$^-$ are in good match, indicating a 1:1 complex.

Initially, we carried out a SPR binding assay of pol β using the non-gapped ds/ss junction replication fork. The binding was very weak, with K_D values in the range of micromolar concentration.[7] However, upon introduction of the 1 nt-gap (Figure 2c), the DNA binding affinity of pol β increased 200- to 1,000-fold. These results indicate that the presence of 5′-PO_4 enhances the binding affinity of 8-kDa lyase domain as well as the 31-kDa catalytic domain. The observed differences in binding affinity are consistent with previous reports in which the lyase domain in the duplex (non-gapped) DNA was flexible. Introduction of the 1 nt-gap enhances the binding affinity of the polymerase to DNA.[61] The results are also in agreement with gel assays, which had previously shown that addition of the correct dCTP opposite unmodified DNA enhances the binding affinity of polymerase compared to other nucleotides, by an induced-fit model adopted by pol β.[61]

Lesion and sequence effects on binary binding affinities with modified DNA

An unusually greater binding of Kf-exo$^-$ was observed for modified dG, where the K_D of this interaction was 5 – 11-fold higher than the K_D for interaction with the unmodified native DNA substrate. The binary binding affinity decreased in the order FABP > FAAF > dG for the CG*A sequence, and FAAF> FABP > dG for the TG*A sequence (Table 3). Previous studies have also shown tighter binary binding of Kf-exo$^-$ with the AAF adduct.[62] Using gel-retardation assays, Dzantiev and Romano[62] showed that the bulky

and hydrophobic AAF interacts with nearby hydrophobic amino acid residues, strengthening its binding to the active site of Kf-exo$^-$. The authors suggested that such lesion-induced conformational adjustment may block the conformational change required to properly accommodate an incoming nucleotide.[27]

It is well established that the *N*-deacetylated fluorinated analog FAF adducts (Figure 1) adopt sequence-dependent equilibrium between *B*- and *S*-conformers. FABP is similarly *N*-deacetylated, but lacks a methylene bridge, resulting in a bulky twisted biphenyl moiety.[21] In other words, FABP may behave like FAAF at the replication fork of the template in the active site of a polymerase. In contrast to the unmodified control, modified adducts displayed a significant decrease (7- to 13-fold) in dissociation rate, with positive net stabilization energy (Table S2). The markedly slower off-rates are consistent with single-molecule FRET studies as well as gel shift assay in which the presence of the bulky DNA adduct stabilizes the binary complex and does not induce dissociation before the nucleotide incorporation.[16,61]

In contrast to Kf-exo$^-$, pol β exhibited significantly lesser binary binding affinity to the modified templates. Furthermore, the modified sequences exhibited significantly faster dissociation rates and more negative net stabilization energies. As in the ds/ss situation discussed above, it is likely that FAAF promote conformational heterogeneity in a sequence containing a 1-nt gap. Such heterogeneity may hinder the interaction of that sequence with key amino acids in the polymerase, thus preventing the polymerase from undergoing conformational change that is necessary for strong binding.

Lesion and sequence effects on ternary binding affinities with modified DNA

Nucleotide selectivity was low in the ternary complexes with Kf-exo$^-$, where the K_D

values indicate poor discrimination between the correct (K_D 0.19 – 0.30 nM) and incorrect (K_D 0.28 – 0.67 nM) nucleotides. Variance in these values ranged from 1.5- to 3.5- fold (Table 3). This poor selectivity does not depend on the nature of the lesion (FABP *vs*. FAAF) or the 5′-flanking base (*C*G*A vs. *T*G*A). The lack of nucleotide selectivity appears to be in agreement with the results of tryptic digestion studies, in which the AAF-polymerase complex maintains an unstable non-catalytic open conformation in the presence of any dNTP.[63] In other words, AAF-modification did not stabilize the complexes in relation to the incoming nucleotide. This is contrasted with native DNA, to which the polymerase binds very tightly in the presence of the correct nucleotide dCTP, and is insensitive to digestion. Our ^{19}F NMR results (Figure 3) indicate a complex conformational heterogeneity of the bulky FAAF at the ds/ss templating position, which may prevent the polymerase from properly accommodating an incoming dNTP. This reasoning is in accord with the weak electron densities observed for the arylamine base in the active site of T7 DNA polymerase,[27] where the authors of the previous study also concluded that conformational heterogeneity may hinder the insertion of an incoming nucleotide.

The low selectivity for incoming nucleotides could also arise from the high stability of binary complex, which may hinder the polymerase's ability to recognize the incoming nucleotide. No crystal structures or high-resolution NMR structures are currently available for complexes between any DNA polymerase and ABP or the fluorinated FABP. In the present study, FABP in both sequences exhibited a single ^{19}F signal possibly for a *B*- or a *B*/*S*-conformational mix owing to the presumed conformational flexibility at the ds/ss junction. These NMR data, albeit in the absence of a polymerase, are in agreement

with the gel based kinetics data, which reveal a preference towards inserting the correct nucleotide over other nucleotides (Table 1).

In the case of Kf-exo$^-$, TG*A sequence favored the insertion of dCTP more efficiently than the CG*A sequence. The relative insertion efficiency f_{ins} of dCTP opposite FABP was significantly lower in the CG*A (500-fold) and TG*A (33-fold) sequences compared to the unmodified controls (Table 1). This 15-fold difference in f_{ins} is puzzling because FABP at ds/ss junction exhibited a single ^{19}F signal in both sequences (Figure 3). However, we have shown previously that FAF in the duplex setting displayed a greater *S*-conformer in the CG*A duplex (50%) relative to the TG*A (38%). As mentioned above, it is likely that the absence of co-planarity in FABP would embrace intermediate structures between FAAF and FAF, as observed from ^{19}F NMR, gel and SPR assays.

The SPR results with pol β (Table 4) indicated that a modified templating base weakens the polymerase binding affinity and the nucleotide selectivity (Figure S9, Table 4). The reduced binding affinity of pol β to the modified template DNA could be related to the lesion-induced conformational heterogeneity in the active site of the polymerase. In the closed conformation, key amino acids such as Lys 234 and Tyr 271 interact with the minor groove of the primer strand, while Arg 283 interacts with the template strand of DNA. As mentioned above, it is possible that the FAAF at the 1 nt-gap may hinder the active site geometry, and thus prevent the conformational change necessary to form the catalytic ternary complex. We previously observed similar conformational heterogeneity caused by FAF bound to 1 nt-gap DNA in both the absence and presence of pol β.[7] The results are also consistent with translesion synthesis studies in which the minor groove conformation benzo[*a*]pyrene diol epoxide- N^2-dG adducts creates steric clash with the

active site of pol β, thereby reducing the insertion rate.[64] These results are in agreement with the steady-state kinetics data that show significant reductions in the f_{ins} of dCTP opposite FABP in the CG*A and TG*A sequences (142- and 59-fold, respectively), relative to the corresponding unmodified controls.

The question is how to reconcile the apparent lack of discrimination between dNTP at the binding step (Figure 7; Tables 3 and 4) with the clear preference for accurate insertion of dCTP (Figure 4; Tables 1 and 2). We have recently shown that the AF adduct can change its binding characteristics at the replication fork or in a single nucleotide gap in the active sites of DNA polymerases.[7] Similarly, it is plausible that the dynamics of FABP and FAAF-induced conformational heterogeniety could be altered to accommodate an incoming dNTP within the active site of polymerases in a way that favors the incorporation of the correct base dCTP.

.

SPR as a powerful tool for probing polymerase action

In the present study we have taken advantage of the sensitivity of SPR, which allowed us to probe the delicate interaction between polymerases and DNA strands containing arylamine-DNA lesions at the binary and ternary complex levels. We were able to measure a sub-nanomolar difference in binding affinity among dNTPs. We found that 0.7 – 3.5 RU of DNA coating was sufficient, with no significant interference from mass transport limitation.

The binding specificity ratios (K_D of the control binary complex over the K_D of a ternary complex) in the presence of dNTPs, for the unmodified (dG) and FAAF- and FABP-modified lesions are plotted as in Figure 7. The dNTPs are color-coded in the plot.

We observed highly specific binding between Kf-exo$^-$ and the native DNA substrates in the presence of the correct dCTP (green) opposite a dG templating base (Figure 7a,b). This is consistent with the polymerase undergoing conformational change, from open to closed, to form Watson-Crick base pairs. Kf-exo$^-$ binds weakly with the incorrect dNTPs, probably retaining the catalytically incompetent open conformation. The binding of dGTP (pink) with Kf-exo$^-$ was particularly poor. Similar binding results were obtained with pol β (Figure 7c,d) although the affinities for modified ternary complexes were generally weaker than those with Kf-exo$^-$. In both enzymes, however, we observed no discernible nucleotide specificity (dNTPs) and sequence effects (CG*A *vs* T*GA).

K_D values for the ternary complexes for unmodified DNA were determined using affinity analysis because the association rate (k_a) reaches the near-diffusion limit in native DNA. This procedure allowed for the monitoring of interactions between unmodified or adducted DNA, with different polymerases on a single chip. The present work also demonstrates the utility of SPR in distinguishing the substrate preference of different polymerases (e.g., ds/ss *vs*.1-nt gap for pol β). To our knowledge, this is the first comprehensive use of SPR to probe nucleotide insertion kinetics during the action of a polymerase. Furthermore, the present SPR work advances the limits of SPR technology,[48,65] demonstrating that SPR can measure sub-nanomolar affinity differences between incoming nucleotides and the active site of a polymerase.

In conclusion, we have characterized the SPR binding affinity of the mutagenic FABP and FAAF lesions bound to Kf-exo$^-$ and pol β. Kf-exo$^-$ binds strongly to ds/ss template/primer DNA, whereas pol β prefers gapped DNA. Tighter binding was observed between unmodified dG and Kf-exo$^-$ or pol β . The systems exhibited

nucleotide selectivity, with K_D values increasing in the order of dCTP << dTTP ~ dATP << dGTP. Unlike pol β, Kf-exo$^-$ binds tightly to both FAAF and FABP lesions in the binary systems. With lesion-modified templates, both polymerases exhibited minimal nucleotide selectivity. The relative insertion efficiency f_{ins} of dCTP opposite FABP was significantly higher in the TG*A compared to the CG*A sequence and the unmodified controls. While the lesion effect was not significant in Kf-exo$^-$, the active site of pol β is sensitive to the FAAF-induced conformational heterogeneity. Our SPR data are complemented by primer steady-state kinetics and ^{19}F NMR data, and provide valuable insights into how lesion-induced conformational heterogeneity in DNA alters the action of polymerases, and thus affects the nucleotide insertion efficiency and coding potential.

AUTHOR INFORMATION

Corresponding Author

Bongsup P. Cho: Phone: +1 401 874 5024. Fax: +1 401 874 5766. E-mail: bcho@uri.edu

Funding

This research is supported by NCI/NIH (CA098296) and NCRR/NIGMS (P20 GM103430-12).

ACKNOWLEDGMENTS

We thank Drs. Catherine Joyce / Olga Potapova of Yale University and Samuel H. Wilson and William A. Beard of NIEHS for providing Kf-exo$^-$ (D424A) and pol β. V.G.V acknowledges DST, India for Ramanujan Fellowship. We also thank Dr. Matthew Blome of GE HealthCare for helpful comments on SPR experiments.

ASSOCIATED CONTENT

Supporting Information. The details of sample preparation, MALDI characterization, sensorgram simulation and binding kinetics; denature gel separation of ligated and non-ligated oligonucleotides (Figure S1); HPLC chromatography of 83 and 84 mer TG[FAAF]A modified sequences (Figure S2); MALDI-TOF characterization of TG[FAAF]A 31, 83 and 84 mer (Figure S3), TG[FABP]A (Figure S4), CG[FAAF]A (Figure S5); CG[FABP]A (Figure S6); fitted and simulated curves of TG[FABP]A with Kf-exo- binding (Figure S7); sensorgrams of ternary Kf-exo$^-$ complexed with TGA (Figure S8); sensorgrams of ternary pol β complexed with CGA (Figure S9); sensorgram and fitted residuals of pol β binding with FAAF-dG (Binary) (Figure S10); theoretical calculations of binding ratio (Figure S11); tabulated values for spectral data of arylamine modified 31, 84 mer (Table S1) and kinetics details of sequence binding with Kf-exo$^-$ and pol β in binary system (1:1 binding)(Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

ABBREVIATIONS

FABP, *N*-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl; FAAF, *N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene; Kf-exo$^-$, Klenow fragment exonuclease deficient; pol β, human DNA polymerase β; SPR, surface plasmon resonance.

REFERENCES

(1) Delagoutte, E. (2012) DNA polymerases: mechanistic insight from biochemical and biophysical studies. *Front Biosci. (Landmark Ed) 17*, 509-544.

(2) Federley, R. G., and Romano, L. J. (2010) DNA polymerase: structural homology, conformational dynamics, and the effects of carcinogenic DNA adducts. *J. Nucleic Acids 2010*.

(3) Kozack, R., Seo, K. Y., Jelinsky, S. A., and Loechler, E. L. (2000) Toward an understanding of the role of DNA adduct conformation in defining mutagenic mechanism based on studies of the major adduct (formed at N(2)-dG) of the potent environmental carcinogen, benzo[*a*]pyrene. *Mutat. Res*. *450*, 41-59.

(4) Broyde, S., Wang, L., Rechkoblit, O., Geacintov, N. E., and Patel, D. J. (2008) Lesion processing: high-fidelity versus lesion-bypass DNA polymerases. *Trends Biochem. Sci*. *33*, 209-219.

(5) Freisinger, E., Grollman, A. P., Miller, H., and Kisker, C. (2004) Lesion (in)tolerance reveals insights into DNA replication fidelity. *EMBO J. 23*, 1494-1505.

(6) Vaidyanathan, V. G., Xu, L., and Cho, B. P. (2012) Binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo(-)) and various arylamine DNA lesions characterized by surface plasmon resonance. *Chem. Res. Toxicol*. *25*, 1568-1570.

(7) Vaidyanathan, V. G., Liang, F., Beard, W. A., Shock, D. D., Wilson, S. H., and Cho, B. P. (2013) Insights into the conformation of aminofluorene-deoxyguanine adduct in a DNA polymerase active site. *J. Biol. Chem. 288*, 23573-23585.

(8) 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.

(9) Jain, V., Hilton, B., Lin, B., Patnaik, S., Liang, F., Darian, E., Zou, Y., Mackerell, A. D., Jr., and Cho, B. P. (2013) Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Res*. *41*, 869-880.

(10) 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.

(11) 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.

(12) 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.

(13) Patnaik, S., and Cho, B. P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol*. *23*, 1650-1652.

(14) Cho, B. P. (2010) Structure-Function Charateristics of Aromatic Amine-DNA Adducts, in *The Chemical biology of DNA damage*. Wiley-VCH, Weinheim. 217-233.

(15) Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health,Part C: Environ. Carcinog.*

Ecotoxicol. Rev. *22*, 57-90.

(16) Michaels, M. L., Johnson, D. L., Reid, T. M., King, C. M., and Romano, L. J. (1987) Evidence for in vitro translesion DNA synthesis past a site-specific aminofluorene adduct. *J. Biol. Chem. 262*, 14648-14654.

(17) Shibutani, S., Suzuki, N., and Grollman, A. P. (1998) Mutagenic specificity of (acetylamino)fluorene-derived DNA adducts in mammalian cells. *Biochemistry 37*, 12034-12041.

(18) Vrtis, K. B., Markiewicz, R. P., Romano, L. J., and Rueda, D. (2013) Carcinogenic adducts induce distinct DNA polymerase binding orientations. *Nucleic Acids Res*. *41*, 7843-7853.

(19) Shibutani, S., Suzuki, N., Tan, X., Johnson, F., and Grollman, A. P. (2001) Influence of flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells. *Biochemistry 40*, 3717-3722.

(20) 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.

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

(22) Seo, K.-Y., Jelinsky, S. A., and Loechler, E. L. (2000) Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens. *Mutat. Res-Rev. Mutat*.

463, 215-246.

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

(24) Doublie, S., and Ellenberger, T. (1998) The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol. 8*, 704-712.

(25) Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science 264*, 1891-1903.

(26) 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.

(27) Dutta, S., Li, Y., Johnson, D., Dzantiev, L., Richardson, C. C., Romano, L. J., and Ellenberger, T. (2004) Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis. *Proc. Natl. Acad. Sci. U.S.A. 101*, 16186-16191.

(28) McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N., and Brown, T. (1994) Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs. *Biochemistry 33*, 10266-10270.

(29) Efrati, E., Tocco, G., Eritja, R., Wilson, S. H., and Goodman, M. F. (1999) "Action-at-a-distance" mutagenesis. 8-oxo-7, 8-dihydro-2'-deoxyguanosine causes base substitution errors at neighboring template sites when copied by DNA polymerase beta. *J.*

Biol. Chem. 274, 15920-15926.

(30) Hoffmann, J. S., Pillaire, M. J., Garcia-Estefania, D., Lapalu, S., and Villani, G. (1996) In vitro bypass replication of the cisplatin-d(GpG) lesion by calf thymus DNA polymerase beta and human immunodeficiency virus type I reverse transcriptase is highly mutagenic. *J. Biol. Chem. 271*, 15386-15392.

(31) Hoffmann, J. S., Pillaire, M. J., Maga, G., Podust, V., Hubscher, U., and Villani, G. (1995) DNA polymerase beta bypasses in vitro a single d(GpG)-cisplatin adduct placed on codon 13 of the HRAS gene. *Proc. Natl. Acad. Sci. U.S.A. 92*, 5356-5360.

(32) Canitrot, Y., Cazaux, C., Frechet, M., Bouayadi, K., Lesca, C., Salles, B., and Hoffmann, J. S. (1998) Overexpression of DNA polymerase beta in cell results in a mutator phenotype and a decreased sensitivity to anticancer drugs. *Proc. Natl. Acad. Sci. U.S.A. 95*, 12586-12590.

(33) Servant, L., Cazaux, C., Bieth, A., Iwai, S., Hanaoka, F., and Hoffmann, J. S. (2002) A role for DNA polymerase beta in mutagenic UV lesion bypass. *J. Biol. Chem. 277*, 50046-50053.

(34) Batra, V. K., Shock, D. D., Prasad, R., Beard, W. A., Hou, E. W., Pedersen, L. C., Sayer, J. M., Yagi, H., Kumar, S., Jerina, D. M., and Wilson, S. H. (2006) Structure of DNA polymerase beta with a benzo[c]phenanthrene diol epoxide-adducted template exhibits mutagenic features. *Proc. Natl. Acad. Sci. U.S.A. 103*, 17231-17236.

(35) Pustovalova, Y., Maciejewski, M. W., and Korzhnev, D. M. (2013) NMR mapping of PCNA interaction with translesion synthesis DNA polymerase Rev1 mediated by Rev1-BRCT domain. *J. Mol. Biol. 425*, 3091-3105.

(36) Berlow, R. B., Swain, M., Dalal, S., Sweasy, J. B., and Loria, J. P. (2012)

Substrate-dependent millisecond domain motions in DNA polymerase beta. *J. Mol. Biol. 419*, 171-182.

(37) 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.

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

(39) 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.

(40) Broyde, S., Wang, L., Zhang, L., Rechkoblit, O., Geacintov, N. E., and Patel, D. J. (2008) DNA adduct structure-function relationships: comparing solution with polymerase structures. *Chem. Res. Toxicol. 21*, 45-52.

(41) Lukin, M., and de Los Santos, C. (2006) NMR structures of damaged DNA. *Chem. Rev. 106*, 607-686.

(42) Tang, Y., Liu, Z., Ding, S., Lin, C. H., Cai, Y., Rodriguez, F. A., Sayer, J. M., Jerina, D. M., Amin, S., Broyde, S., and Geacintov, N. E. (2012) Nuclear magnetic resonance solution structure of an N(2)-guanine DNA adduct derived from the potent tumorigen dibenzo[a,l]pyrene: intercalation from the minor groove with ruptured Watson-Crick base pairing. *Biochemistry 51*, 9751-9762.

(43) Mu, H., Kropachev, K., Wang, L., Zhang, L., Kolbanovskiy, A., Kolbanovskiy,

M., Geacintov, N. E., 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.

(44) Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., and Rani, V. (2012) DNA-protein interactions: methods for detection and analysis. *Mol. Cell. Biochem. 365*, 279-299.

(45) Stengel, G., and Knoll, W. (2005) Surface plasmon field-enhanced fluorescence spectroscopy studies of primer extension reactions. *Nucleic Acids Res. 33*, e69.

(46) Sedletska, Y., Culard, F., Midoux, P., and Malinge, J. M. (2013) Interaction studies of muts and mutl with DNA containing the major cisplatin lesion and its mismatched counterpart under equilibrium and nonequilibrium conditions. *Biopolymers 99*, 636-647.

(47) Lebbink, J. H., Fish, A., Reumer, A., Natrajan, G., Winterwerp, H. H., and Sixma, T. K. (2010) Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS. *J. Biol. Chem. 285*, 13131-13141.

(48) Ritzefeld, M., and Sewald, N. (2012) Real-Time Analysis of Specific Protein-DNA Interactions with Surface Plasmon Resonance. *J. Amino Acids. 2012*, 816032.

(49) Vaidyanathan, V. G., Xu, L., and Cho, B. P. (2013) Binding kinetics of DNA-protein interaction using surface plasmon resonance. *Protocol Exchange* doi: 10.1038/protex.2013.1054.

(50) Jain, V., Vaidyanathan, V. G., Patnaik, S., Gopal, S., and Cho, B. P. (2014) Conformational insights into the lesion and sequence effects for arylamine-induced

translesion DNA synthesis: ^{19}F NMR, surface plasmon resonance, and primer kinetic studies. *Biochemistry 53*, 4059-4071.

(51) 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.

(52) 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.

(53) Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999) An open and closed case for all polymerases. *Structure 7*, R31-35.

(54) Singhal, R. K., and Wilson, S. H. (1993) Short gap-filling synthesis by DNA polymerase beta is processive. *J. Biol. Chem. 268*, 15906-15911.

(55) Beard, W. A., and Wilson, S. H. (2000) Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta. *Mutat. Res. 460*, 231-244.

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

(57) Beard, W. A., and Wilson, S. H. (2006) Structure and mechanism of DNA polymerase *β*. *Chem. Rev. 106*, 361-382.

(58) Efrati, E., Tocco, G., Eritja, R., Wilson, S. H., and Goodman, M. F. (1997) Abasic translesion synthesis by DNA polymerase beta violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase beta at an abasic lesion in different sequence contexts. *J. Biol. Chem. 272*, 2559-2569.

(59) Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature 391*, 251-258.

(60) Tang, K. H., and Tsai, M. D. (2008) Structure and function of 2:1 DNA polymerase. DNA complexes. *J. Cell. Physiol. 216*, 315-320.

(61) Ahn, J., Kraynov, V. S., Zhong, X., Werneburg, B. G., and Tsai, M. D. (1998) DNA polymerase beta: effects of gapped DNA substrates on dNTP specificity, fidelity, processivity and conformational changes. *Biochem. J. 331 (Pt 1)*, 79-87.

(62) Dzantiev, L., and Romano, L. J. (1999) Interaction of Escherichia coli DNA polymerase I (Klenow fragment) with primer-templates containing N-acetyl-2-aminofluorene or N-2-aminofluorene adducts in the active site. *J. Biol. Chem. 274*, 3279-3284.

(63) Dzantiev, L., and Romano, L. J. (2000) Differential effects of N-acetyl-2-aminofluorene and N-2-aminofluorene adducts on the conformational change in the structure of DNA polymerase I (Klenow fragment). *Biochemistry 39*, 5139-5145.

(64) Chary, P., Beard, W. A., Wilson, S. H., and Lloyd, R. S. (2012) DNA polymerase β gap-filling translesion DNA synthesis. *Chem. Res. Toxicol. 17*, 2744-2754.

(65) Schlachter, C., Lisdat, F., Frohme, M., Erdmann, V. A., Konthur, Z., Lehrach, H., and Glokler, J. (2012) Pushing the detection limits: the evanescent field in surface plasmon resonance and analyte-induced folding observation of long human telomeric repeats. *Biosens. Bioelectron.31*, 571-574.

FIGURE LEGENDS

Figure 1. (a) Chemical structures of FAAF- and FABP-dG adducts (b) Major (upper image) and minor (lower image) groove views of the prototype *B*-, *S*-, and *W*- conformers of arylamine dG-lesions in CPK model with the DNA duplex in grey surface (color code: arylamine lesion, red; modified-dG, cyan; dC opposite the lesion site, green). Note that the arylamine lesion (red) in *W*-conformation is wedged in the narrow minor groove.

Figure 2. (a) Schematic representation of template–primer DNA constructs for SPR assays; Hairpin template-primer oligonucleotide constructs for (b) Kf-exo$^-$ and (c) pol β.

Figure 3. ^{19}F NMR spectra of FABP and FAAF adducts in the CGA and TGA duplexes at ds/ss junction at 25 °C.

Figure 4. Assays of full-length and single-nucleotide incorporation into FABP-adducted CG*A and TG*A sequences with (a) Kf-exo$^-$ and (b) pol β.

Figure 5. Sensorgrams of binary complexes of (a) Kf-exo$^-$ and (b) pol β with unmodified and modified TGA sequences (1:1 binding fitted curves are overlaid as red lines).

Figure 6. Steady-state affinity analysis of interaction of Kf-exo$^-$ with (a) -TG[FAAF]A- and (b) -TG[FABP]A- sequences.

Figure 7. Plots of nucleotide specificity ratio (K_D-binary/K_D) with (a, b) Kf-exo$^-$ and (c, d) pol β for unmodified and modified TG*A and CG*A DNA templates. The dNTPs are color-coded in the plots. K_D-binary-dG represents K_D of unmodified DNA-polymerase binary complex and denominator K_D represents the ternary complex of unmodified DNA (or) binary and ternary complexes of adducted DNA.

TABLE LEGENDS

Table 1. Steady-state kinetics parameters for insertion of dCTP opposite unmodified and FABP-dG adduct with Kf-exo-

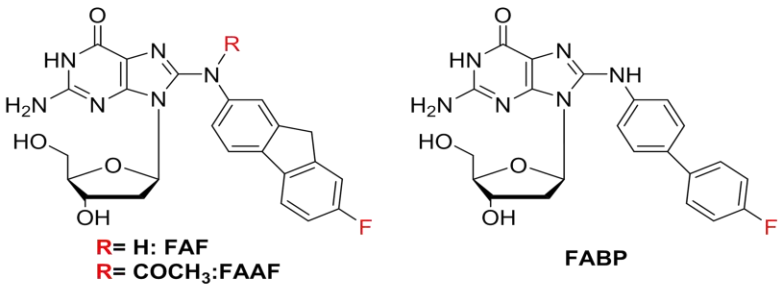
Table 2. Steady-state kinetics parameters for insertion of dCTP opposite unmodified and FABP-dG adduct 1 nt gap with pol β

Table 3. SPR binding affinities (K_D)* of unmodified TGA/CGA and arylamine dG-adducts with Kf-exo$^-$ (steady-state affinity analysis) in the binary and ternary systems

Table 4. SPR binding affinities (K_D)* of unmodified TGA/CGA and arylamine dG-adducts with pol β (1:1 binding) in the binary and ternary systems

Figure 1

a)



b)

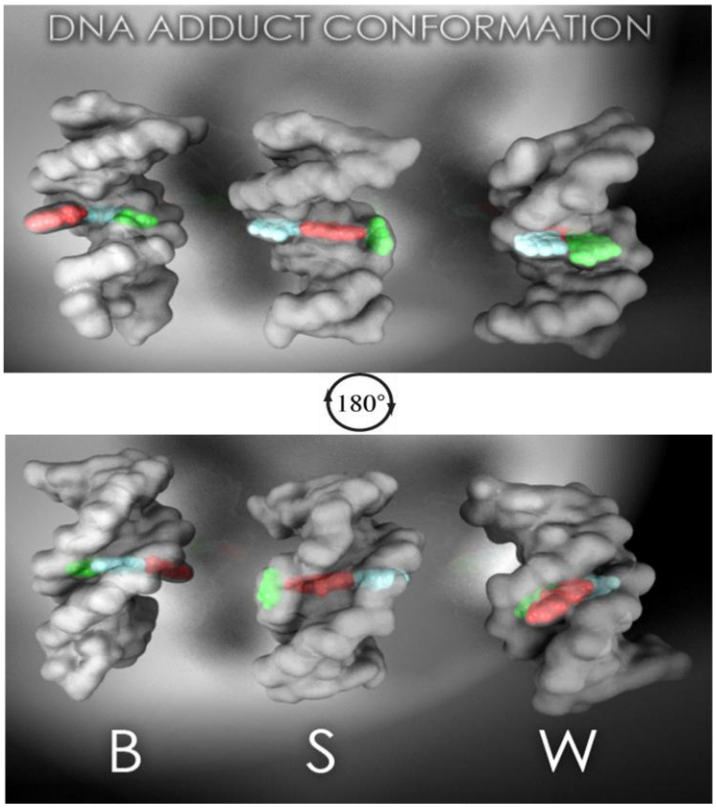


Figure 2

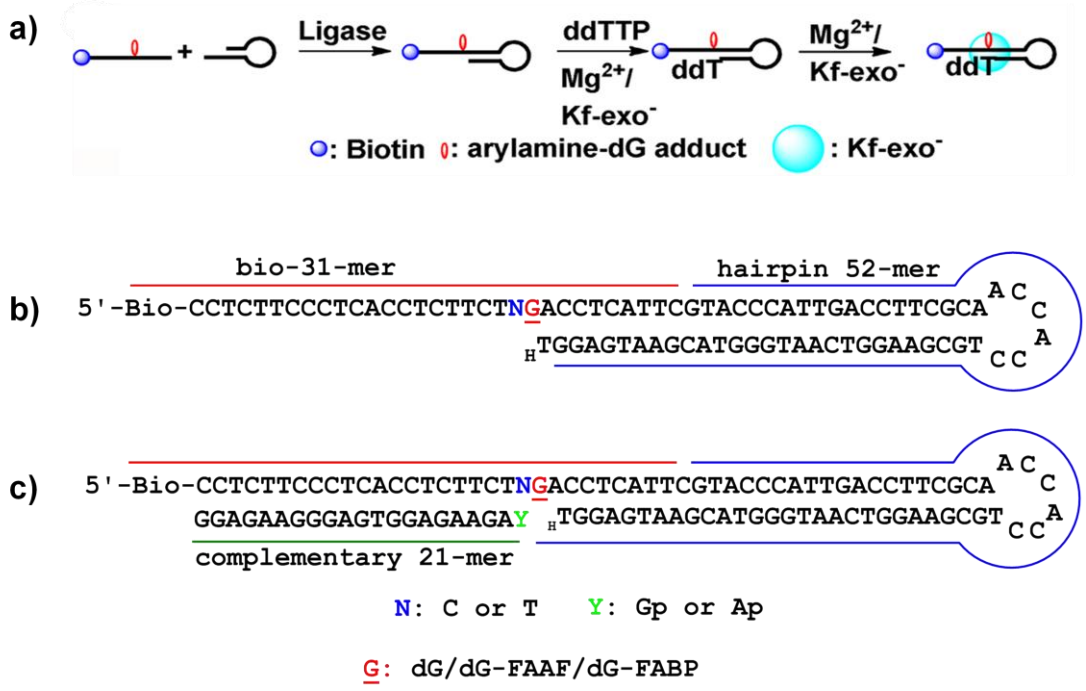

a)
Ligase
ddTTP
Mg2+/
Kf-exo-
ddT
Mg2+/
Kf-exo-
ddT
: Biotin
: arylamine-dG adduct
: Kf-exo-
b)
bio-31-mer
hairpin 52-mer
5'-Bio-CCTCTTCCCTCACCTCTTCTNGACCTCATTCGTACCCATTGACCTTCGCA
TGGAGTAAGCATGGGTAACTGGAAGCGT
c)
5'-Bio-CCTCTTCCCTCACCTCTTCTNGACCTCATTCGTACCCATTGACCTTCGCA
GGAGAAGGGAGTGGAGAAGAY
TGGAGTAAGCATGGGTAACTGGAAGCGT
complementary 21-mer
N: C or T
Y: Gp or Ap
G: dG/dG-FAAF/dG-FABP

Figure 3

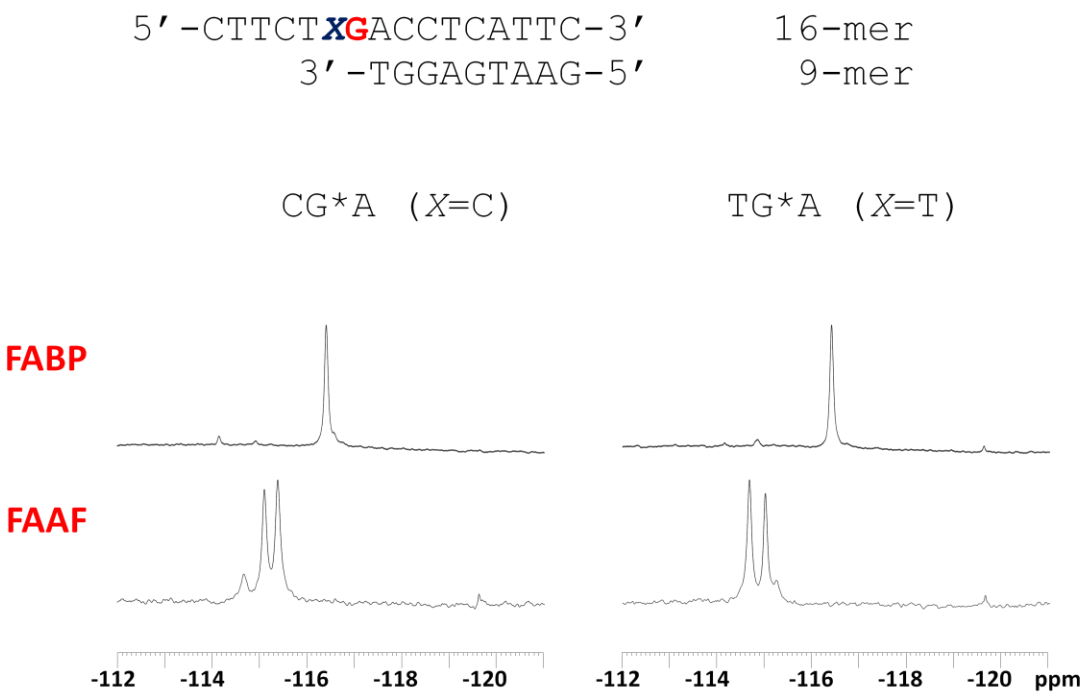

5′-CTTCTXGACCTCATTC-3′ 16-mer
3′-TGGAGTAAG-5′ 9-mer
CG*A (X=C)
TG*A (X=T)
FABP
FAAF
-112
-114
-116
-118
-120
-112
-114
-116
-118
-120
ppm

Figure 4

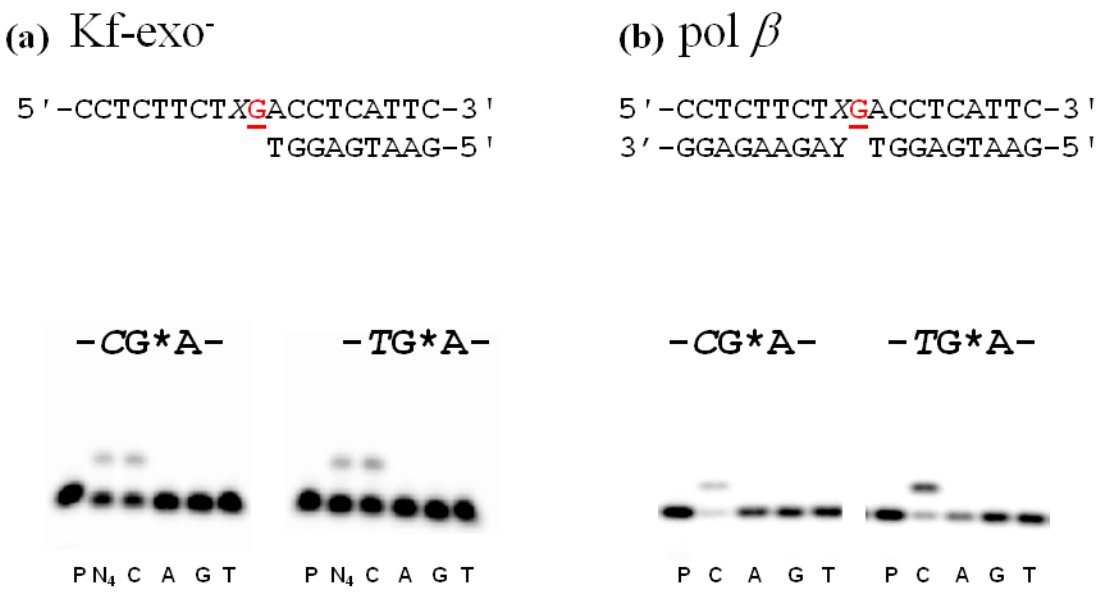

(a) Kf-exo-
5'-CCTCTTCTXGACCTCATTC-3'
TGGAGTAAG-5'
(b) pol β
5'-CCTCTTCTXGACCTCATTC-3'
3'-GGAGAAGAY TGGAGTAAG-5'
-CG*A-
-TG*A-
-CG*A-
-TG*A-
P N4 C A G T
P N4 C A G T
P C A G T
P C A G T

Figure 5

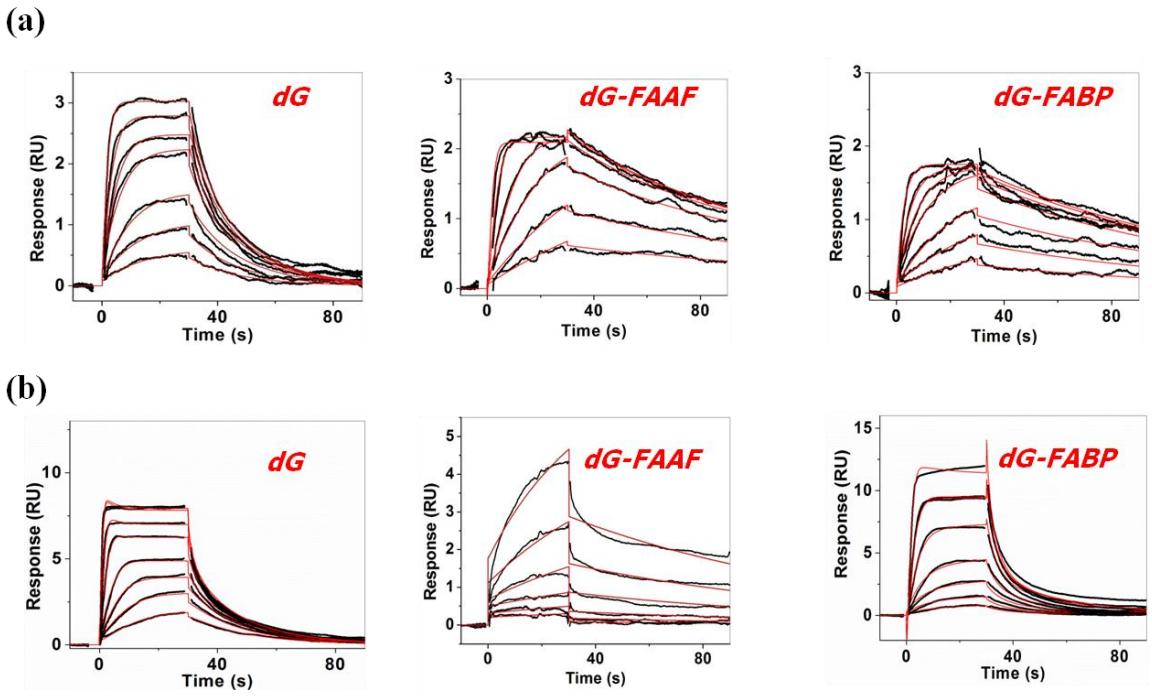


Figure 6

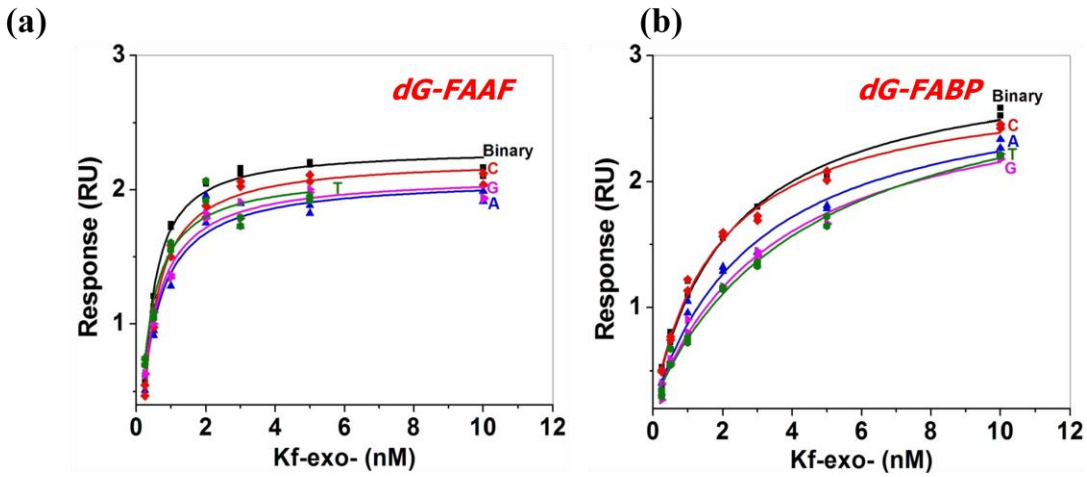

(a)
dG-FAAF
Binary
C
T
G
A
Response (RU)
Kf-exo- (nM)
(b)
dG-FABP
Binary
C
A
T
G
Response (RU)
Kf-exo- (nM)

Figure 7

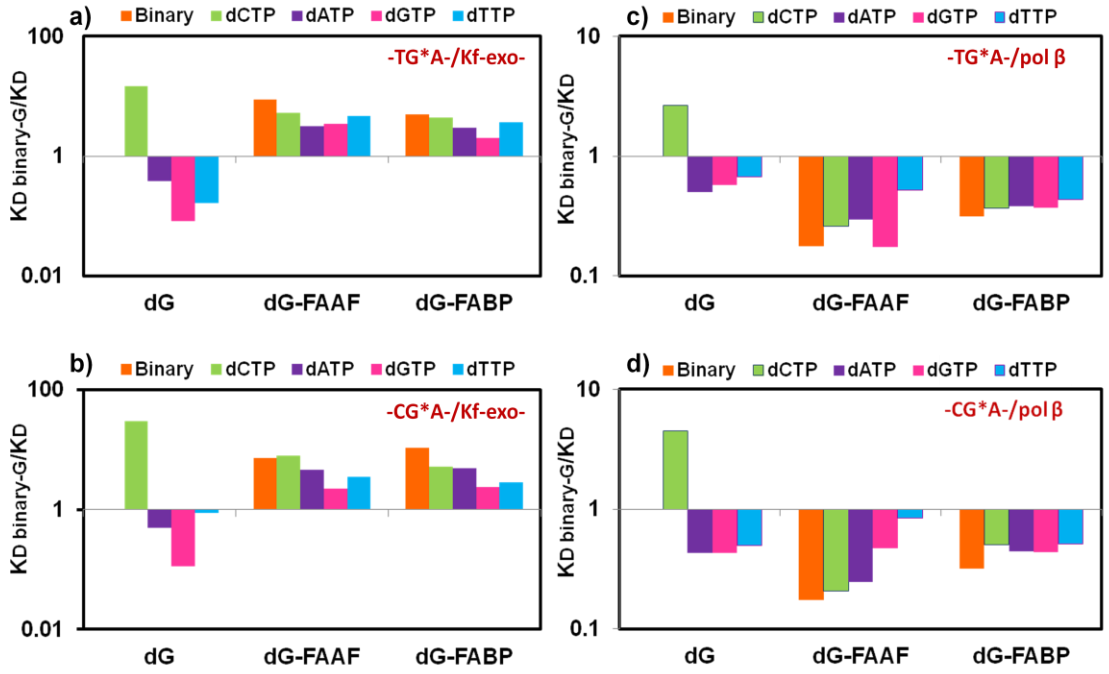

a)
Binary
dCTP
dATP
dGTP
dTTP
-TG*A-/Kf-exo-
KD binary-G/KD
100
1
0.01
dG
dG-FAAF
dG-FABP
c)
-TG*A-/pol β
10
0.1
b)
-CG*A-/Kf-exo-
d)
-CG*A-/pol β

Table 1

Sequence Context	Incoming dNTP	k_{cat} (min^{-1})	$K_{m,\ dCTP}$ (μM)	k_{cat}/K_m ($\mu M^{-1}\ min^{-1}$)	f_{ins}
-C**G**A-	dCTP	21.9(1.4)	0.80 (0.24)	27.3(8.4)	1.00
-C**G[FABP]**A-	dCTP	0.44(0.05)	6.62(3.37)	0.06(0.03)	0.002
-T**G**A-	dCTP	3.10 (0.31)	0.23 (0.14)	13.8 (8.3)	1.00
-T**G[FABP]**A-	dCTP	0.32(0.02)	0.66 (0.30)	0.48(0.22)	0.03

$f_{ins} = (k_{cat}/K_m)_{\text{modified}}/(k_{cat}/K_m)_{\text{unmodified dG control}}$

Table 2

Sequence Context	Incoming dNTP	k_{cat} (min^{-1})	$K_{m,\ dCTP}$ (μM)	k_{cat}/K_m ($\mu M^{-1}\ min^{-1}$)	f_{ins}
-C**G**A-	dCTP	1.14(0.08)	1.98(0.73)	0.58(0.21)	1.00
-C**G[FABP]**A-	dCTP	0.60(0.09)	135(41)	0.004(0.001)	0.007
-T**G**A-	dCTP	0.83(0.06)	4.75(1.48)	0.17(0.05)	1.00
-T**G[FABP]**A-	dCTP	1.02(0.14)	298(69)	0.003(0.001)	0.017

$f_{ins} = (k_{cat}/K_m)_{\text{modified}}/(k_{cat}/K_m)_{\text{unmodified dG control}}$

Table 3

Sequence	Binary	dCTP	dATP	dGTP	dTTP
-T**G**A-	1.3(0.3)	0.09(0.08)	3.5(1.1)	16(8)	3.60(0.95)
-C**G**A-[a]	1.5 (0.5)	0.05(0.02)	3.1(2.2)	13(12)	1.70(0.53)
-T**G[FAAF]**A-	0.15(0.05)	0.25(0.06)	0.42(0.18)	0.38(0.10)	0.28(0.19)
-C**G[FAAF]**A-[a]	0.21(0.05)	0.19(0.11)	0.33(0.05)	0.67(0.07)	0.43(0.08)
-T**G[FABP]**A-	0.27(0.02)	0.30(0.03)	0.44(0.01)	0.66(0.08)	0.36(0.07)
-C**G[FABP]**A-[a]	0.14(0.10)	0.29(0.12)	0.31(0.13)	0.63(0.11)	0.54(0.17)

[a] K_D values were taken from ref. 6
*K_D values are in nanomolar (nM).

Table 4

Sequence	Binary	dCTP	dATP	dGTP	dTTP
-T**G**A-	0.80(0.17)	0.30(0.09)	1.6(0.3)	1.40(0.15)	1.2(0.2)
-C**G**A-	0.90(0.10)	0.20(0.12)	2.10(0.09)	2.10(0.09)	1.8(0.2)
-T**G[FAAF]**A-	4.50(0.15)	3.10(0.16)	2.70(0.09)	4.60 (0.12)	1.50(0.04)
-C**G[FAAF]**A-	5.20(0.12)	4.40(0.08)	3.7(0.1)	1.90 (0.05)	1.10(0.21)
-T**G[FABP]**A-	2.60(0.25)	2.20(0.23)	2.10(0.17)	2.20(0.16)	1.80(0.12)
-C**G[FABP]**A-	2.80(0.21)	1.80(0.08)	2.00 (0.08)	2.10(0.07)	1.80(0.06)

For non-gapped duplex DNA (ds-DNA) with pol β, K_D values exceeds 0.8 μM.

*K_D values are in nanomolar (nM).

Manuscript III

Prepared for submission to ***Journal of Molecular Biology***

A Systematic Spectroscopic and Thermodynamic Investigation of Slippage Mediated Frameshift Mutagenesis

Lifang Xu and Bongsup P. Cho*

Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy,
University of Rhode Island,
Kingston, Rhode Island 02881, United States

***Correspondence to Bongsup P. Cho:**

Phone: +1 401 874 5024

E-mail: bcho@uri.edu

Abstract

We have conducted a series of systematic studies to probe the conformational mechanisms of arylamine-induced -2 base deletion mutations frequently observed in the *E. coli Nar*I mutational hot sequence (5'---CGGCG*C*N*---3'; *N*= dC and dT) during translesion synthesis (TLS). We employed two well-characterized fluorinated bulky DNA lesions [*N*-(2'-deoxyguanosin-8-yl)-2-fluoro-2-aminofluorene] (FAAF) and [*N*-(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl] (FABP) derived from the environmental carcinogens 2-aminofluorene and 4-aminbiphenyl. Our work focused primarily on elucidating the effects of lesion size, bulkiness and overall topology and the 3'-next flanking base N in producing an -2 slipped mutagenic intermediate (SMI), the bulge structure responsible for arylamine-induced -2 frameshift mutagenesis. To that end, we examined two chemical simulated TLS models, in which the FAAF/FABP lesion was positioned at G_3 position of two 16-mer *Nar*I sequences (5'-CTCTCG$_1$G$_2$CG$_3$*C*N*ATCAC-3', *N*=C: *Nar*I-dC Series; *N*=T: *Nar*I-dT Series). These templates were each annealed systematically with increasing primer lengths in the full length and -2 deletion pathways and their thermodynamic, conformational, and binding profiles at each elongation step were measured by various biophysical techniques including spectroscopic (dynamic ^{19}F NMR/CD), thermodynamic (UV-melting/DSC) and affinity binding (SPR). The results showed that the Streisinger-based -2 bulge formation is initially triggered by the conformational stability of the G_3*: C base pair at the ds/ss replication fork as well as the nature of base sequences surrounding the lesion site. The extent of conformational instability of the G_3*: C pair determines the nature of a slippage

('CG' vs. 'C') and subsequent primer elongation yields the respective -2 G_3*C or CG_3* bulge structures for FAAF and FABP, respectively. Each bulge structure exists in a mixture of B-SMI and S-SMI, in which the bulky lesion is located outside the bulge ('solvent exposed') and inserted into the bulge ('solvent protected') respectively, and their conformational rigidity increases as a function of primer lengths. We found that the B-/S-SMI population ratios are dependent on various structural characteristics primarily the bulkiness ('*N*-acetyl), coplanarity, and overall topology as well as the 3'-base sequence (N) next to the bulge formation. The results indicate the importance of conformational stability, heterogeneity and flexibility in the mechanisms of bulky arylamine-induced frameshift mutagenesis.

Introduction

Arylamine is an important group of 'bulky' environmental pollutants that has been implicated in various sporadic human cancers such as the bladder, breast, and liver cancer.[1] 2-Aminofluorene and its derivatives have been most extensively studied as model bulky carcinogens. In vivo, these chemicals are reduced to *N*-hydroxylamine and subsequently activated to the acetyl or sulfate derivatives by the action of ubiquitous *N*-acetyltransferase or sulfotransferase enzymes.[2] Consequently, these pro-carcinogenic esters produce highly reactive electrophilic nitrenium ions, which are known to interact directly with cellular DNA to form DNA adducts.[2] In vivo, 2-aminoflurene produces two major C8-subsituted dG adducts, *N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene (dG-C8-AF, simply designated as AF here on), and *N*-acetyl-(2'-deoxyguanosin-8-yl)-2-

aminofluorene (dG-C8-AAF, simply designated as AAF here on).[3] The related arylamine 4-aminobiphenyl is a known human bladder carcinogen that also binds to dG at C8 to form N-(2'-deoxyguanosin-8-yl)-4-aminobiphenyl (dG-C8-ABP, designated as ABP here on) as a major adduct. The structures of AF and AAF differ only in that the latter contains a bulky acetyl group on the central linking nitrogen (Figure 1). Despite the structural similarity, they produce different mutational and repair outcomes. In E. coli, AF produces both point and frameshift mutations, whereas AAF results in mostly frameshift mutations.[4-7] However, both lesions produce primarily G to T point mutations in the COS-7 mammalian cells replication.[8, 9] The *N*-deacetylated AF adduct in fully paired duplexes adopts an equilibrium between syn-glycosidic stacked (S) and anti-glycosidic major groove (B)-type conformations.[4-6, 10] In contrast, the N-acetylated AAF adduct adopts a mixture of the base-displaced stacked (S)-, the major groove binding B-type (B), and the minor groove binding wedge (W) conformations.[4-7] Both AF and AAF induce S/B/W-conformational heterogeneity and their population ratios depend on the nature of the base sequence neighboring the lesion-modified dG, thus having a direct impact on their mutational and repair outcomes. As an example, S-conformation destabilizes DNA duplexes and causes DNA bending, allowing a greater nucleotide excision repair.[11] The S/B conformeric AF is processed by high fidelity polymerases albeit with low frequency, after a short stall at the lesion site. On the contrary, the S/W-conformeric and distorting AAF lesion is a strong blocker of replicative polymerase, consequently requiring the recruitment of special bypass polymerase for translesion synthesis (TLS).[12] The S/B-conformational heterogeneity has also been observed in recent crystal studies of several mutagenic arylamines

complexed with replicative and bypass polymerases.[10, 13-16] Very little is known about the structure and conformation of the ABP adduct. However, accumulated evidence indicates that ABP exists mostly in B-type conformation in fully-paired complementary DNA duplexes.[17] This is rationalized by the fact that ABP is not as coplanar as AF because of the missing methylene carbon at C9.

Arylamine carcinogens produce two general types of mutation, the base and frameshift mutations. The latter usually involves a shift of one to two or multiple bases, causing the loss of genetic information. The molecular mechanisms of frameshift mutation have been studied in certain mutagenic sequences. The G: C rich *Nar*I sequence (5′-$G_1G_2CG_3CN$A-3′) in *E. coli* is one such example and has been considered as a unique mutational hotspot especially for AF- and AAF-induced frameshifts mutations. The bulky lesion AAF at G_3 position induces frameshifts at greater frequency and their propensity is modulated by the nature of the nucleotide in the *N* position (C ~ A > G >> T).[18-20] As an example, -2 deletion mutations arise primarily due to the extrusion of AAF-G_3 with neighboring cytosine bases into two bases misalignment through the formation of a -2 base slippage mutagenic intermediate (SMI), resulting into dinucleotide GC deletion. In contrast, adduction at G_1 and G_2 does not induce -2 deletion mutations because of the lacking repetitive GC dinucleotide. Similarly in the monotonous runs of G the extrusion of AAF-G_3 into '-1 base SMI' during replication produces -1 deletion mutations. This process for frameshift mutagenesis is known as "Streisinger Slippage Model", which is proposed by Streisinger and colleagues decades ago.[21, 22] Compared to AF, the bulkier *N*-acetylated AAF has shown much greater propensity to induce frameshift mutation in the *Nar*I sequence. AAF is a strong blocker in highly replicative

polymerase, and is bypassed by the low fidelity polymerase, ultimately producing various deletion mutations.[6, 7, 23, 24]

Clearly, the nature of polymerases can also contribute to the efficiency of deletion mutations. Gill and Romano have shown that the AAF in the *Nar*I sequence specifically interferes with the active site of *E. coli* DNA polymerase I Klenow fragment to induce deletion of the GC dinucleotides past AAF. In addition, they showed a very different SMI structure on non-*Nar*I sequence in the active site of the polymerase.[25] Fuchs group showed that DNA synthesis past G_3-AAF lesion in the *Nar*I sequence in the presence of pol II also leads a GC dinucleotide deletion, whereas pol V is responsible for its error-free bypass in *E. coli*.[19] The crystal structures of AAF-modified template-primer replication fork bound to the tight active site of T7 DNA polymerase showed that the hydrophobic AAF lies behind the O-helix and stuck in a hydrophobic pocket of finger subdomain, thus allowing the polymerase to adopt an open conformation. Such conformational anomaly results in strong blockage, triggering a slippage leading to various frameshift mutations during TLS.[7] Bulky lesions have also produced multiple conformations in the spacious active sites of various bypass polymerases, which may account for the different replication efficiencies including frameshift mutations.[24]

Using primer extension assays coupled with MALDI-TOF mass spectrometry, Schorr and Carell have shown that frameshift mutation is triggered by the unstable molecular association of the AAF-dG lesion with the correct incoming nucleotide dC.[26] Such configurations have been observed in both replicative and bypass polymerases and are likely to promote the lesion-containing dG and flanking bases to slip to form bulge structures. Hence, the stability of bulged-out structures and subsequent elongation will

determine the propensity for frameshift mutagenesis. To that end, we recently performed systematic structure and conformational studies of FAAF-modified *Nar*I-sequence based −1, −2, and −3 deletion duplexes.[27] FAAF is the ^{19}F analog of AAF. These SMIs existed in a mixture of the so-called external "solvent exposed" B-type (B-SMI) and inserted "solvent protected" "stacked" S (S-SMI) conformers, with the population of the S conformer and thermodynamic stability in the order of −1 > −2 > −3 deletion duplexes. The results showed greater thermal and thermodynamic stabilities of S-SMI over the flexible B-SMI, which supports the aforementioned Carell's hypothesis. We also studied *Nar*I-based -2 deletion [(5'-CTCGGCG*C***N**ATC-3') (5'-GATNGCCGAG-3'), ***N*** = dC or dT] duplexes, in which G* was FAF, the ^{19}F analog of AF. These sequences mimic a SMI for -2 deletion mutations. The results indicated that the *Nar*I-dC/-2 deletion duplex adopts mostly a S-SMI conformer, whereas the *Nar*I-dT/-2 deletion duplex exists as a mixture of S-SMI and various 'exposed" B-SMI (Figure 1).[28]

In the present study, we hypothesize that the *Nar*I-induced frameshift mutagenesis is stimulated by the conformational stability of SMI formed during TLS. The conformational, thermodynamics, and binding affinity details of the two progressive TLS models were examined, in which the FAAF/FABP lesion is positioned at G_3 position of 16-mer *Nar*I sequence (5'-CTCTCG$_1$G$_2$CG$_3^*$C*N*ATCAC-3', *N*=C: *Nar*I-dC Series; *N*=T: *Nar*I-dT Series). These templates were both annealed systematically with increasing primer lengths (full length extended or -2 deletion), and their thermodynamic, conformational, and binding profiles at each elongation step were investigated and analyzed. We have utilized a powerful array of biophysical techniques such as differential scanning calorimetry (DSC), surface plasmon resonance (SPR), as well as

circular dichroism (CD) and dynamic ^{19}F and imino NMR spectroscopy. The results are discussed the critical role of conformational stability, heterogeneity and flexibility in the mechanisms of bulky arylamine-induced frameshift mutagenesis.

Materials and Methods

Caution: Aminofluorene and aminobipheyl are animal, human carcinogens respectively therefore caution is required when handling.

All crude oligodeoxynucleotides (oligo, 2-10 μmol scale) in desalted form were obtained from Eurofins MWG operon (Huntsville, Al, USA) and purified by using reverse phase high performance liquid chromatography (RP-HPLC). The HPLC system consisted of a Hitachi EZChrom Elite HPLC unit with an L2450 diode array detector and a Phenomenex Clarity C18 column (150*10mm, 5.0 um). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA, USA).

Preparation of FAAF modified DNA. The 16-mer *Nar*I dC/dT series were used and their respective primers are listed in Figure 2. The G_3 in the sequence 5'-CTCTCG$_1$G$_2$CG*$_3$C*N*ATCAC-3' (G*=FAAF; *N* = C or T) was site-specifically modified by FAAF adduct according to published procedures.[4, 29] Briefly, 1 mg of *N*-acetoxy-*N*-2-acetylamino-7-fluorofluorene was first dissolved in absolute ethanol and added to 3 mL sodium citrate buffer (10mM, pH 6.0) containing approximately 200 μM of unmodified template (-CTCTCG$_1$G$_2$CG$_3$C*N*ATCAC-3', *N* = C or T). The mixed solution was placed in a 37 °C water bath shaker for 5 min. 3 mL Ether was added into reaction mixture to extract the extra FAAF diester and water layer of the mixture was collected and filtered with 0.2 μm filter paper. The worked out reaction mixture was injected to

reverse phase preparative HPLC (Figure 5a) and appropriate peaks were purified up to 99% purity. In theory there should be a total of seven FAAF adducts due to the presence of three guanines in the model sequence: three mono-, three di- and one tri-adduct. As such, a stringent HPLC condition is required. Our HPLC mobile systems entails a gradient of 3-9% acetonitrile for 5min followed by 9-30% acetonitrile for 20-min, in pH 7.0 100 mM of ammonium acetate buffer with a flow rate of 2.0 mL/min.

Preparation of FABP modified DNA. Similarly as FAAF modification, FABP was synthesized and was used to modify 16-mer *Nar*I dC/dT sequence.[5, 17] Generally, 3 mg of *N*-acetoxy-*N*-trifluoroacetyl-7-fluoro-4-aminobiphenyl was dissolved in absolute ethanol and mixed with approximately 200 μM of unmodified sequence in sodium citrate buffer at 37 °C water bath for 30 min. The reacted mixture was ether extracted and filtered then injected in RP-HPLC system and chromatogram was shown in Figure 7a. The modified DNA was collected and purified up to 99% purity by repeating the mixture injections following the method involving a gradient system of 7.5-12.2% acetonitrile in 100 mM of ammonium acetate buffer with a flow rate of 2.0 mL/min for 30min, followed by 12.2-40% for 5min and then 40%-7.5% for 5min.

All seven modified adducts were isolated and three mono-adducts were characterized by MALDI-TOF using 3'→5' or 5'→3' exonuclease enzyme digestion method. The isolated G_3-FAAF/FABP modified 16-mer sequences were each annealed with appropriate primers with different length to form the various ds/ss duplexes starting from n-1, n, n+1, n+2, n+3, n+6 to full duplex (Figure 2) for structural studies. A similar set of unmodified templates with appropriate primers was also prepared as controls.

G_3 adduct characterization. The FAAF/FABP modified mono-adducts were

characterized by enzyme digestion using matrix assisted laser desorption ionization-time of flight (MALDI-TOF). The MALDI matrix solution was prepared by mixing 1:1 of 3-hydroxypicolinic acid (3-HPA, 50 mg/mL) and ammonium citrate dibasic (50 mg/mL). The DNA samples (200 pmol) were mixed with 1 μl snake venom phosphodiesterase (SVP, 0.1 unit/μl) and 1 μl bovine spleen phosphodiesterase (BSP, 0.01 unit/μl) respectively for 3'- or 5'- enzyme digestion. Spot 1 μl digest solution with 1 μl matrix mixture on the plate every 30 seconds and air-dried. The MALDI-MS spectra were obtained in reflectron mode and analyzed using Shimadzu Axima performance.

UV thermal melting experiments. UV thermal melting experiments were carried out using a Cary100 Bio UV/Vis spectrophotometer equipped with a 6*6 multi-cell chamber and 1.0 cm path length. The cells temperatures were controlled by a built-in Peltier temperature controller. Various duplex solutions were prepared in solutions containing 0.2 M NaCl, 10 mM sodium phosphate and 0.2 mM EDTA (pH 7.0) with a concentration range of 1.2-6.4 μM. Thermal melting curves were monitored and conducted at 260 nm absorbance by varying the temperatures of the cell (1^{o}C/min). Each melting experiment contained forward/reverse scans and was repeated five times. Thermodynamic parameters of bimolecular reactions were obtained and calculated using the program MELTWIN version 3.5.[30]

Differential Scanning Calorimetry (DSC). All calorimetric samples were measured using Nano-DSC from TA Instrument (Lindon, UT, USA). 100 μM solutions containing unmodified or G_3-FAAF modified template with various primers were prepared by dissolving in a pH 7.0 buffer containing 20 mM sodium phosphate and 0.1 M NaCl and degassed at least 10 min under vacuum. The TLS samples were scanned

against the blank buffer from 15 to 85 °C at a rate of 0.75 °C/min; at least five repetitions including forward/reverse were measured. Raw data were collected in the form of microwatts *vs.* temperature. A buffer *vs.* buffer scan was provided as a blank to be subtracted from the sample scan and normalized for heating rate. The area under the resulting curves was proportional to the transition enthalpy, ΔH. ΔG and ΔS can be calculated according to previous described procedures.[31] Due to the separation difficulty of G_3-FABP mono-adduct, DSC experiment for FABP series was not conducted.

Circular Dichroism (CD) spectra. CD experiments were obtained on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. G_3-FAAF/FABP modified template (10 μM) as well as the unmodified control template were annealed with an equimolar amount of primer in 400 μl of a buffer which contains 0.2 M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA (pH 7.0) and placed in a 1.0 mm path-length cell. All the CD samples were incubated at 85 ^{o}C for 5 min and cooled to room temperature to ensure duplex formation. CD spectra were acquired from 200 nm to 400 nm at a scanning rate of 50 nm/min, along with every 0.2 nm with 2 s response time. The final data were the average of 10 accumulations scan with 25-point adaptive smoothing algorithms.

Dynamic ^{19}F-NMR experiments. Approximately 100 μM of G_3-FAAF/FABP modified 16-mer template was annealed with an equimolar amount of various primers to produce appropriate template-primer samples and lyophilized. The samples were dissolved in 250 μl of NMR buffer (10% D_2O/90% H_2O, pH 7.0 containing 100 mM NaCl, 10 mM sodium phosphate and 100 μM EDTA) and filtered through a 0.2 μm membrane filter into a Shigemi NMR tube. All ^{1}H and ^{19}F NMR results were obtained

using Varian NMR spectrometer with a HFC probe operating at 500.0 and 476.5 MHz, respectively, following the previous reported acquisition parameters.[28][32] Each spectra was recorded with 1.0 s recycle delay. Imino proton spectra was recorded at 5 °C to 60 °C using a phase sensitive jump return sequence and referenced to DSS. ^{19}F NMR spectra was acquired in the 5-70 °C temperature range with increment of 5 or 10 °C in the ^{1}H-decoupled mode and referenced to C_6F_6 in C_6D_6 at -164.9 ppm. A total of 1200 scans were acquired for each dynamic ^{19}F NMR. The line shape simulations were analyzed using WINDNMR-Pro version 7.1.6 according to the reported procedures.[10]

Surface Plasmon Resonance (SPR). Biacore T200 instrument from GE Healthcare was employed to measure the strength of duplex binding affinity during model TLS experiments in polymerase free solutions. In order to coat DNA on the streptavidin immobilized chip surface, 5' biotin labeled 16-mer *Nar*I sequence (5'-biotin-CTCTCGGCGCNATCAC-3', N = C or T) was designed and purchased from Eurofins MWG operon in desalted form. The *Nar*I sequence was used exactly the same as thermodynamic/conformational studies except the 5' biotin attachment. Around 10 μM of 5'-biotylated *Nar*I 16-mer dC/dT was modified with FAAF and FABP following the routine reaction procedure and purified by RP-HPLC and later characterized by MALDI-TOF using 3' enzyme digestion. The HPLC elution method for biotin labeled DNA modification was different from non-biotin labeled reaction because of the increasement of hydrophobicity biotin attachment. The FAAF-dC G_3 modified template was purified using RP-HPLC with a gradient method system of 5-15% acetonitrile with 100 mM of ammonium acetate buffer at a flow rate of 2.0 ml/min for 10 min followed by 15-38% acetonitrile for 5 min and 38%-5% for 5min, while FAAF-dT G_3 mono-adduct was

purified starting from 8%-15% acetonitrile for 20 min then followed by 15-38% acetonitrile for 10min and 38-8% for 5min in ammonium acetate buffer with 2ml/min flow rate. The elution method for FABP modification was also different where starting from 7.5% to 13% for 40 min, followed by 13-38% for 25 min and 38-7.5% for 5 min at 2 ml/min flow rate. The mono-adducts were characterized by MALDI-TOF using 3' enzyme digestion, since 5' of DNA was labeled by biotin, the 5' exonuclease enzyme did not apply in this case.

Carboxymethylated CM 5 chip was activated by amine coupling kit according to published procedures.[33] Generally, EDC/NHS mixture was injected over the chip for 7 minutes and followed by coating streptavidin (SA, 50 ug/ml dissolved in sodium acetate buffer, pH 4.5) on flow cell 2 and 4, while flow cell 1 and 3 were left as blank. 1M ethanolamine was injected over the surface to block the unreacted esters. Five pluses of 50 mM NaOH was injected over the chip to remove the unbound SA and running buffer stabilized the surface for 20-30 min before DNA coating. Flow cell 2 and 4 were coated around 200 RU unmodified DNA and FAAF/FABP modified G_3 dC/dT, respectively under the manual control mode. Different lengths of complementary sequences were prepared in HBS-P^+ buffer (10 mM Hepes, 150 mM NaCl, 0.05% surfactant P20 at pH 7.4) and injected over the chip surface at 25 °C with 100 s contact time and 360 s dissociate time at 15 µl/min flow rate. 50 mM NaOH was used as regeneration solution and injected over the chip for 30 s to remove the complementary sequences. Dissociation rate constants (kd) were determined and analyzed using Scrubber software, version 2.0 (Myszka and collaborators, BioLogic Software) in Kd-alone fitting mode.

Results

Translesion synthesis (TLS) model Systems:

Two TLS models were designed, in which FAAF or FABP lesion is at G_3 position of a 16-mer *Nar*I sequence (5'-CTCTCG$_1$G$_2$CG$_3$*CNATCAC-3', *N* = C: *Nar*I-dC Series; *N* = T: *Nar*I-dT Series). FAAF and FABP are fluorine-tagged AAF and ABP lesions, which are intended for obtaining dynamic ^{19}F NMR spectra. The underline 12-mer portion of this 16-mer *Nar*I sequence is identical to that used in our previous study, in which the sequence effect of the FAF was investigated in the context of -2 deletion mutation.[28] In that study, the *Nar*I-dC/-2 deletion duplex was found to adopt the S-SMI conformer exclusively, whereas the *Nar*I-dT/-2 deletion duplex showed multiple conformers, presumably consisting of S- and B-SMI conformers among others. Initially, we tried to use the same TLS sequences; however, the initial 12/5-mer template/primer (e.g., n-1) was too short to form proper duplexes to give meaningful thermo-melting and thermodynamic parameters. As such, two more bases were included on both sides (CT on the 5' and AC on the 3') to make a 16-mer, whereas the inner core was kept exactly the same.

A total of four -2 deletion SMI TLS models were produced for each FAAF and FABP lesion (Figure 2) in the dC and dT series and the corresponding unmodified controls. Four unmodified control TLS models were also prepared, such as fully paired complementary and -2 deletion duplexes. Figure 2A shows the fully extended duplex control models formed by annealing the FAAF/FABP-16-mer templates with appropriate primers, i.e., specifically n-1(16/7-mer), n (16/8-mer), n+1(16/9-mer), n+2(16/10-mer), n+3(16/11-mer), and n+8(16/16-mer). As for the SMI models, the 16-mer templates were similarly annealed primers to produce appropriate -2 SMI, specifically, n-1(16/7-

mer), n (16/8-mer), n+1 (16/9-mer), n+2 (16/10-mer), n+3 (16/11-mer), and n+6 (16/14-mer).

Sequence Issues:

Figure 3b shows two different -2 SMI models assumed for each lesion, i.e., G_3*C and CG_3* bulges for FAAF and FABP respectively, based on the previous high resolution ^{1}H NMR and fluorescence results that are described below. Using the simple Streisinger model depicted in Figure 3a, insertion of the correct cytosine opposite the lesion at G_3* is the first step.[34] The potentially unstable G_3*: C pair causes a polymerase to pause at the replication fork, triggering a slippage of the nascent strand and leaves two bases bulge out in the template. However, there are two slippage possibilities, either a G_3*C or a CG_3* bulge out. As detailed in Figure 4, the G_3*C bulge out involves a slippage of two terminal bases ("CG" slip) in the primer hydrogen bonded with the downstream complementary 5'-G_2C-3' dinucleotide. Alternatively, CG_3* bulge out can be formed by a single base "C" slippage. Regardless, continued replication of either scenario will lead to a chemically identical daughter strand that is two bases shorter than the parent strand. Figure 4 shows that each of the two pathways (two bases "CG" or one base "C" slippage) is expected to produce a conformational mixture of S-SMI and B-SMI. Both the G_3*C or CG_3* bulge out scenarios will lead to the same -2 deletion mutation. The biological outcome of the two models is identical, however, it is important to understand the structural and sequence aspects of the SMI involved in the different lesions. Evidence indicates the importance of the thermodynamic stabilities of the initial base pairing of G_3*: C at the replication fork. The delicate conformational structures of bulged-out SMI may determine the propensity for frameshift mutagenesis.

There are conflicting reports as to which SMI structure is responsible for the *Nar*I-based -2 deletion. Mao *et al* [35] conducted NMR/molecular modeling studies on a 12/10-mer -2 deletion duplex [(5'-CTCG$_1$G$_2$CG$_3$*CCATC-3') (5'-GATGGCCGAG-3')], in which G$_3$ is modified with AF. Their NMR results showed the exclusive presence of the CG$_3$* bulge out S-SMI (underlined above), in which the AF-modified guanine in the *syn* conformation and 5'-C reside in the major groove and the aminofluorene moiety is fully inserted into the bulge. This result is consistent with the results from our ^{19}F NMR and thermodynamic investigation in the same sequence context, which showed a conformational rigid S-SMI structure. On the contrary, NMR studies by Milhe *et al* on a AAF-modified on a similar 12/10-mer *Nar*I duplex (5'-ACCG$_1$G$_2$CG$_3$*CCACA-3') (5'-TGTGGCCGGT-3')] revealed about ~80% of the G$_3$*C bulge out SMI structure (underlined above), in which the AAF moiety is inserted into the duplex.[36] Unlike the AF case above, however, this S-SMI structure was not defined into a three-dimensional model because of conformational heterogeneity. Furthermore, the conformational nature of the remaining 20% sample was not clearly defined. Nevertheless, these results taken together indicate two very different lesion dependent slippage pathways, i.e., CG$_3$* and G$_3$*C bulge structures for AF and AAF, respectively.

The structures of AF and AAF are essentially identical except that AAF possesses a bulky *N*-acetyl group on the central nitrogen of adduct, thereby exhibiting unique conformational features and different mutational and repair outcomes. The term "*N*-acetyl factor" was previously coined to describe their repair differences. Schorr and Carell [26] showed that AAF-induced -2 frameshift mutation on *Nar*I sequence by the bypass polymerase pol η indeed follow the Milhe's [36] G$_3$*C bulge out model (Figure

3b). We have utilized fluorescence spectroscopy (unpublished) to investigate the two SMI pathways by using sequences, which include the fluorescent tag pyrrolo-deoxycytidine (^{P}C) in either 5'- or 3'-side of the lesion. The fluorescence results indicated that AAF and AF induce G_3*C and CG_3* slipped mutagenic structures, respectively, supporting the NMR results discussed above. Therefore, the conformational stability and flexibility of the G_3*: C base pairing at the replication fork dictates the types of a slippage, i.e., the conformationally flexible *N*-deacetylated AF promotes one base (C) slip, whereas the bulky and rigid *N*-acetylated AAF induces two base (CG) slippage. Evidently, the nature of the adduct structure (*N*-acetyl, bulkiness, coplanarity, overall topology) and base sequence contexts surrounding the lesion are important factors for determining the types of -2 frameshifts. FABP is considered as an analog of FAF because both are *N*-deacetylated, thus susceptible for conformational heterogeneity; however, FABP lacks a bridging methylene group, therefore less coplanar than the FAF. Hence, the G_3*C bulge model was selected for FAAF and the CG_3* bulge model for FABP.

Preparation and characterization of modified template sequences

The 16-mer *Nar*I template sequence (5'-CTCTC$G_1G_2CG_3$C*N*ATCAC-3', *N* = C or T for dC and dT series, respectively) was treated with either an activated FAAF or FABP, according to the biomimetic procedures published previously.[4, 28, 29] In principle, there should be at least seven adducts because of the three guanines in the *Nar*I sequence; such as three mono-, three di- and one tri-adduct. The guanines in the sequence maintain similar chemical reactivity, and consequently it is possible to regulate the relative ratios of mono-, di- and tri- adducts by adjusting reaction time. The complexity of the adduct profiles called for development of an efficient HPLC separation method.

Previously, the separation of this complex mixture took 90 min to collect all seven modified peaks (Figure 6b).[11] In the present study, an efficient HPLC method was developed to purify the same reaction mixture in a much shorter time frame (Figure 6a) (see Material and Methods for details). Figure 6 compares the two HPLC chromatograms. Figure 6a shows the separation of all seven modified sequences in near base-line resolution in less than 20 minutes. The un-reacted 16-mer oligo appeared at 5.5 min, while the FAAF-modified sequences are in the range of 11–14 min (mono-FAAF 1, 2, 3), 16–19 min (di-FAAF 4, 5, 6), and 20 min (tri-FAAF), respectively. This result is in clear contrast to the old HPLC profile wherein the mono-adducts (peaks 1–3) appeared in 28–35 min, di adducts (peaks 4–6) in 45–56 min, and the tri adduct peak 7 at around 85 min (Figure 6b). The initial adduct mono-, di- and tri-FAAF assignments were based on the UV intensity of the absorption shoulders in the range of 300–320 nm, the intensity of which is known to be proportional to the aminofluorene chromophores, 1:2:3 for mono-, di- and tri-adducts, respectively (Figure 5b). The structural identities of the FAAF-adducts were characterized by exonuclease enzyme digestions-MALDI-TOF mass spectrometry as described below. The results showed that peak 1was G_1, peak 2 was G_3, and peak 3 was G_2 in both the *Nar*I dC and dT sequences (Figures 8–13).

Similarly, the treatment of the same 16-mer *Nar*I dC/dT sequence with FABP (5'-CTCTCG$_1$G$_2$CG$_3$C*N*ATCAC-3', *N* = C or T for dC and dT series) gave a reaction mixture that showed all three group of adducts in less than 45 min, such as the three mono-adducts at 19–24 min, three di-adducts at 35–38 min and the tri-adduct at 42 min (Figure 7a). The UV shoulder absorbance in the range of 300-320 nm indicated the number of FABP adducts (mono di, and tri). The three mono-FABP adducts were

collected with repeated HPLC injections and were characterized by MALDI-TOF as detailed below. The results indicated that peak 1was G_1, peak 2 was G_3, and peak 3 was G_2 in both the *Nar*I dC and dT sequences (Figures 14–19). The order of elution was same as the FAAF case above.

***Nar*I-FAAF-16-mer dC sequence**:

The FAAF modified *Nar*I 16-mer dC sequence (5'-CTCTCG$_1$G$_2$CG$_3$*CCATCAC-3') was characterized previously using ESI-QTOF-MS.[11] The overall HPLC elution patterns were similar. All three mono-FAAF adducts have been characterized by the analysis of MALDI-TOF spectra (Figures 8-10). Here details of the characterization of peak 2 as G_3 modification is presented, which is relevant to the present study.

Figure 9 shows the MALDI-TOF spectra of 3'-5' SVP (a) and 5'-3' BSP (b) exonuclease digestions of peak 2 at different time points (0–120 s for SVP and 0–30 min for BSP). These two enzymes are known to remove one base at a time from the 3' and 5' side, respectively. The peak at 5017 m/z at 0 s was the mass of the modified template as a control (i.e., before digestion). However, at 30 s of digestion, the control 16-mer 5017 m/z was replaced, with the appearance of three lower molecular weights 3508, 3218, and 2929 m/z. These fragments correspond to the 11-, 10, and 9-mer fragments with two, one, and no extra base on the 3'-flanking side of the FAAF-lesion site (see inset), respectively. However, these signals disappeared quickly, leaving the 9-mer 2929 m/z (theoretical 2928.67 m/z; 5'-CTCTCG$_1$G$_2$CG$_3$[FAAF]-3') as the only one remaining peak at 120 s. The results indicated peak 2 as FAAF-G_3 modified. The 5'-3' exonuclease digestion (Figure 9b) was carried out similarly, which confirmed the 3'-5' exonuclease digestion

above. The parent 16-mer control signal m/z 5017 at 0 s was replaced with the exclusive signal at m/z 2883, which corresponds to the fragment, cleaved one base before the lesion (see inset). An additional peak was observed at m/z 2594 after 30 min of digestion which corresponds to the 8-mer containing FAAF at G_3. The 5'-3' exonuclease action was much slower than the 3'-5'-exonuclease counterpart. Therefore, these results confirmed peak 2 as the G_3-FAAF-modified 16-mer dC series, which are consistent with the ESI-MS results published previously.[11] The HPLC peaks 1 and 3 were similarly characterized and their spectra are included in Figure 8 and 10. The results identified the HPLC peaks 1, 2, and 3 as the FAAF at G_1, G_3, and G_2, respectively.

***Nar*I-FAAF-16-mer dT sequence**:

The FAAF modified dT sequence has not been characterized previously. Three mono-adduct peaks were characterized by MALDI using both 3' and 5' enzyme digestions. Figure 11 shows the 3' (a) and 5' (b) enzyme digestions of peak 1, where 5031 m/z peak at 0 s corresponds to the control peak of 16-mer FAAF modified dT sequence, whereas at 90 s and 120 s, peaks at 4742 and 4428 m/z showed the gradual digestion fragments, respectively. Peaks of 2310 and 1980 m/z at 150 s correspond to the 7- and 6-mer fragments with one and no extra base on the 3' side of FAAF lesion (see inset), respectively. These data indicate the FAAF-modification at G_1 (5'-CTCTCG$_1$[FAAF]-3'; see inset). The 5'-3' digestion shown in Figure 11b indicated 3844 m/z at 60 min which corresponds to one extra base at 5' side of the lesion. The results confirmed peak 1 as FAAF-modification at G_1.

Figure 12 shows the MALDI spectra of peak 2 with 3'-5' (a) and 5'-3' (b) digestions. The 5031 m/z peak at 0 s corresponds to the control peak before digestion,

whereas the 3218 and 2928 m/z peaks at 120 s −180 s correspond to 10- and 9-mer fragments, with one and no extra base at the 3'-side of the lesion site, respectively. The results indicate a FAAF-modification at G_3 position (CTCT$G_1G_2CG_3$[FAAF]-3'). The 5'-3 digestion at 60 m showed 2896 m/z, which corresponds a signal to one base extra at the 5'-side of the lesion G_3(CG_3[FAAF]CTATCAC-3'). These digestion results confirmed that peak 2 was G_3.

As for peak 3 (Figure 13), the 5031 m/z in both digestions corresponds to the molecular ion. In 3'-5' digestion, the 2599 and 2310 m/z peaks at 150 s correspond to 9- and 8-mer fragments containing modified G_2. On the contrary, the 5'-3' digestion profiles at 30-60 m revealed signals at 3844 and 3555 m/z, corresponding to 11- and 10-mer with one or no base extra to the 5' side of the lesion, respectively. These data confirmed that peak 3 was G_2. Hence, the results identified the HPLC peaks 1, 2, and 3 as the FAAF at G_1, G_3, and G_2, respectively.

***Nar*I-FABP-16-mer dC sequence:**

FABP modification produced three mono adduct peaks (peak 1, 2, and 3) as expected. Figure 14, 15, and 16 show the MALDI-TOF spectra of both 5'-3' and 3'-5'-exonuclease digestion mixtures derived from the peak 1, 2, and 3, respectively. Figure 14 shows the parent ion 4963 m/z at 0 s before digestion. Upon digestion, fragments at 1929 and 3487 m/z were persisted after 3'-5' and 5'-3' exonuclease digestion, respectively. These fragments correspond to bond cleavages right at the lesion site of G_1. As for peak 2 (Figure 15), persistent fragments were observed at 2876 and 2540 m/z after 3'-5' and 5'-3' digestions, respectively. The results indicate FABP modification at G_3. Peak 3 (Figure 16) produced persistent fragments at 2258 and 3159 m/z after 3'-5' and 5'-3' digestions,

respectively, which are consistent with the G_2 modification. Hence, these results identified peaks 1, 2, and 3 as FABP modification at G_1, G_3, and G_2.

***Nar*I-FABP-16-mer dT sequence:**

Figures 17–19 show the digestion results of *Nar*I-FABP-16-mer dT sequences. In all cases, the parent ions were observed at 4980 m/z before digestion. Upon 3'-5 enzyme digestion, peak 1, 2, and 3 produced fragments at 1930, 2878, and 2259 m/z, respectively, which indicated FABP modification at G_1, G_3, and G_2, respectively. Upon 5' enzyme digestion, peak 1, 2, and 3 produced fragments persisting at 3504, 2844, and 3175 m/z, confirming the same assignments.

The HPLC order of elution should be noted to be identical with FAAF/FABP modified *Nar*I 16-mer sequences, i.e., peak 1, 2, and 3 were G_1, G_3, and G_2, respectively, regardless of lesion and next flanking base sequences.

UV melting: All TLS model duplexes showed mostly monophasic sigmoidal curves on UV melting (Figures 20 and 21). A correlation ($R^2 > 0.9$) between $\ln C_t$ and Tm^{-1} was observed, confirming typical helix-coil melting transitions. Tables 1-4 summarize the thermal and thermodynamic parameters calculated from UV melting curves.

UV melting Curves:

FAAF series: Figure 20 shows UV-melting curves of FAAF-modified full (*Nar*I-FAAF-Full-dC and *Nar*I-FAAF-Full-dT) and -2 deletion duplexes (*Nar*I-FAAF-SMI-2-dC and *Nar*I-FAAF-SMI-2-dT) in the dC and dT series along with corresponding unmodified control models. The unmodified n-1 duplex (16/7-mer) in the dC series (Figure 20a, dotted black), in which the primer is elongated to the one base before the

lesion, was not clearly defined presumably because of a short primer. However, the duplex melting gradually improved to produce well-behaved sigmoidal curves, i.e., increase of Tm as function of temperature (n to n+8). In contrast with the FAAF modified model, the Tm from n to n+3 barely increased. The results indicated a lesion-induced destabilization. By contrast, the corresponding -2 SMI (FAAF-SMI-2-dC) duplex exhibited well-behaved melting curves of all duplexes including the n-1, with generally higher melting (for n to n+2). However, for the unmodified -2 SMI model, the Tm did not change between n and n+3. A higher melting of -2 SMI over the full duplex at n and n+1 indicated lesion-induced duplex stabilization. The opposite result was observed in the dT series, in which the Tm -2 SMI at n and n+1 was lower than that of the full duplex. This finding indicated the direct effect of the next flanking base N (e.g., T over C) on the bulge stability of FAAF at G_3 (Figure 20).

FABP series: Figure 21 shows the UV-melting curves of FABP-modified full and -2 SMI duplexes in both dC and dT series along with the unmodified controls. As in FAAF, FABP stabilized the duplex at n-1 in both the dC and dT series. FABP modified -2 SMI models showed a gradual increase of Tm, suggesting FABP-induced stabilization in the -2 bulge structure.

UV melting thermodynamics:

FAAF-dC series: Figures 22a and 22b show plots of UV-based thermal-melting (Tm) and thermodynamics (ΔG) for the FAAF modified full (*Nar*I-FAAF-Full-dC; left) and -2 SMI (*Nar*I-FAAF-SMI-2-dC; right) duplexes of dC Series with increasing length of primers (n+8 and n+6 for full duplex and -2 SMI, respectively). FAAF-modified duplexes (red, empty circles) are compared with unmodified ones (blue, filled circles). In

the unmodified full duplex model, values increased consistently as expected from standard primer elongation (blue lines). However, for the unmodified -2 SMI models (blue lines), the thermal and thermodynamic values slightly changed from n to n+3. In both full and -2 SMI cases, the lesion effects were minimal at and prior to the lesion site (n-1 to n+1), but became significant between n+2 and full (n+6 and n+8 for -2 SMI and Full, respectively). In the -2 SMI models, the modified (*Nar*I-FAAF-SMI-2-dC) showed greater stability (higher T_m and lower ΔG) than the unmodified controls (ΔTm, 2.10 °C to 13.06 °C, ΔΔG, -0.27 to -4.10 kcal/mol) for the n+2 to n+6 positions. By contrast, the FAAF modified full-length duplexes (*Nar*I-FAAF-Full-dC) showed lower thermal and thermodynamic stabilities (lower ΔTm -6.20 °C to -15.61°C) and higher ΔΔG (1.77 to 4.98 kcal/mol) values compared with those of the unmodified controls (Table 1).

These results indicate that the lesion effect at n-1 to n+1 is minimal in both -2 SMI and full TLS models. Remarkably, no SMI is expected to form up to this point although some discernible differences appear at n+1. The lesion effect was quite consistent between n+2 and n+6/n+8. The thermal and thermodynamic stability of the -2 SMI model over the control SMI was clearly due to the formation of a stable -2 SMI structure in which FAAF is stacked in the solvent protected bulge environment. By contrast, the negative thermodynamic effect on the full-length duplex models is contributed to the FAAF-induced S/B/W-conformational heterogeneity at both replication fork and duplex settings.

FAAF-dT series: Similar trend was observed in the dT series. Figure 23 shows the Tm and ΔG comparison between Full (*Nar*I-FAAF-Full-dT; left) and -2 SMI (*Nar*I-FAAF-SMI-2-dT; right) models for the FAAF dT series as a function of increasing length

of primers. In the fully extended model, FAAF-modified duplex is thermally and thermodynamically less stable than the unmodified control (lower ΔTm -9.61 °C to -19.40 °C and higher ΔΔG 2.42 kcal/mol to 5.02 kcal/mol). However, in the -2 SMI model, thermal and thermodynamic stabilities are significantly increased from n+2 to n+6 (higher ΔTm 10.48 °C to 11.13°C and lower ΔΔG -2.18 kcal/mol to -2.88 kcal/mol) (Table 2). A higher thermal stability encountered for the dC (ΔT_m 2.10 °C to 13.06 °C) over dT (ΔT_m 0.85 °C to 11.13 °C) series -2 SMI indicates greater FAAF's ability to form a stable bulge structure in dC than in dT.

FABP-dC series: Figure 24 shows the Tm and ΔG comparison between FABP-modified full and -2 SMI models with increasing length of primers for the dC series. For FAAF, the fully extended FABP duplexes are thermally and thermodynamically less stable than the unmodified controls in the dC (*Nar*I-FABP-Full-dC) (ΔTm -7.17 °C to -14.47 °C) series. However, in the -2 SMI models, thermal and thermodynamic stability significantly increased from n+2 to n+6 (ΔTm 0.48 °C to 8.17 °C and ΔΔG -0.1 kcal/mol to -2.18 kcal/mol) (Table 3).

FABP-dT series: Figure 25 shows the Tm and ΔG comparison between FABP-modified fully extended and -2 SMI models with increasing length of primers for the dT series. In the fully extended model, the FABP modified duplex destabilized the structure by ΔTm -7.32 °C to -15.25 °C and higher ΔΔG 2.16 kcal/mol to 4.27 kcal/mol (Table 4). The FABP modified duplex stabilized the -2 SMI bulge structure with higher ΔTm 2.48 °C to 10.57°C and lower ΔΔG -0.20 kcal/mol to -2.73 kcal/mol.

DSC

We also conducted DSC experiments on FAAF-modified and unmodified control

-2 SMI TLS models in the dC and dT series. Figure 26 shows the overlays of plots of heat capacity change with increasing temperatures. The maximum point of the Gaussian bell curves in the DSC thermograms represents duplex melting (Tm), and the areas under the curve denote transition enthalpy values (ΔH). The DSC results are independent of concentration and thus provide reliable thermal and thermodynamic parameters compared with those of UV melting.

***Nar*I-SMI-2-dC**: Figure 26a shows an overlay of the unmodified dC series (*Nar*I-SMI-2-dC) from n-1 to n+6 as controls. The n-1 curve (cyan), which represents a 16/7-ds/ss duplex, shows a broad curve with Tm of 35.1 °C and ΔH of -45.0 kcal/mol (Table 5). The curve shapes up nicely with one additional base (n) with Tm of 48.5 °C and ΔH of -48.5 kcal/mol. Both Tm and ΔH have mostly stalled between n+1 and n+3. However, a significant increase existed at n+6 in Tm (57.3 °C) and ΔH (-120.9 kcal/mol). This DSC profile is inconsistent with the regular full-paired TLS cases, which generally show an incremental Tm/ΔH increases with increasing primer elongation.[37] Therefore, these results reflect the presence of a -2 bulge duplex formation.

***Nar*I-FAAF-SMI-2-dC:** Figure 26b shows FAAF-modified -2 SMI bulge structure with increasing length of primers (n-1 to n+6). The major difference compared with the unmodified control (Figure 26a) indicated that Tm and ΔH increased progressively with increasing primer elongation from n-1 to n+3 TLS. In contrast with the unmodified control of -2 SMI, the curves for n to n+3 were all clustered together around the Tm of 48 °C (Figure 26a). These DSC patterns resemble those obtained from melting of a regular full-length unmodified DNA duplex. These results support a unique stabilizing effect of the bulky FAAF though insertion and hydrophobic stacking.

***Nar*I-SMI-2-dT**: Figure 26c shows the DSC profiles for the unmodified -2 SMI TLS models in the dT series. The DSC profile trend was similar to that of the corresponding dC series (Figure 26a) with slightly better Tm dispersion for n to n+3. A major exception demonstrated that the Tm and ΔH values are generally smaller in the dT series.

***Nar*I-FAAF-SMI-2-dT**: Figure 26d shows the DSC curves for the FAAF-modified -2 SMI TLS models. The profile trend was very similar to that of the corresponding FAAF-modified -2 SMI dC series with consistently smaller Tm and ΔH values. The increase in ΔH was not as incremental as Tm in the dC series above. The melting Tm of the n-1 duplex was relatively lower (32.3 °C) than that (41.9°C) (Table 5) of the dC series. This finding is ascribed to the presence of a weak T: A base pair instead of a more stable C: G base pair at the 3'-next flanking base, i.e., dT versus dC at N position (5'-CGGCG*CN-3').

Table 5 summarizes the thermal and thermodynamic parameters from the DSC results. Consistent with the UV melting results above, FAAF modified -2 SMI TLS models are more stable thermally and thermodynamically than the unmodified SMI controls. The same trends are also verified with dC (ΔT_m= 2.60 °C to 13.80 °C, ΔΔG= -0.55 kcal/mol to -5.63 kcal/mol) over dT (ΔT_m= 0.10 °C to 11.80 °C, ΔΔG= -0.17 kcal/mol to -3.40 kcal/mol) series.

Circular dichroism

Circular dichroism (CD) is a sensitive technique for distinguishing different types of DNA duplexes. For example, a typical B-form DNA helix displays a +/− "S-shape" ellipticity at 270 and 250 nm, respectively. The + intensity at 270 nm particularly

indicates the base stacking strength of a duplex DNA. We showed previously that AF- and AAF-modified duplexes exhibit lesion-induced ellipticity changes in the 300 nm to 320 nm ranges depending on their S/B/W-conformational heterogeneity.[11, 28]

Figure 27 compares the overlay of CD spectra of *Nar*I-FAAF/FABP-SMI-2 duplexes in the dC and dT series during the early stage of TLS, i.e., at n-1 (green), n (blue), and n+1 (red). A gradual increase of CD intensity in the FAAF modified dC series was observed at 270 nm in the n-1 → n → n+1 sequence, which indicates a progressive strengthening of base stacking. This finding is contrasted to the lack of such change at n and n+1 in the dT series. The CD results may indicate a greater stacking for the dC series than for the dT. A similar dC versus dT comparison was conducted for the FABP adduct. Interestingly, FABP-modified SMI showed a gradual decrease of intensity at 270 nm in the n-1 → n → n+1 sequence, whereas the dT series exhibited slight changes in intensity. These CD results indicate that FABP and FAAF are involved in uniquely different mechanisms in the formation of -2 bulge adduct structures.

Figure 28 shows the overlays of the CD spectra of FAAF-modified duplexes (red) with those of the unmodified controls (blue) for all TLS steps from n-1 to n+6 in both dC and dT series. In every case, FAAF-modified duplexes exhibited significant blue shifts compared with the unmodified controls. The effects were also greater for dC over dT series (dC series: 6 nm at n-1, 5 nm at n to n+3, 7 nm at n+6; dT series: 4 nm at n-1, 2 nm at n to n+3, 1 nm at n+6) (Table 6). These data suggest an adduct-induced DNA backbone bending. No significant changes existed in the CD intensity at 270 nm in both series throughout TLS except for the n+3/n+6 in the dT series, indicating a different pathway for the formation of -2 bulge adduct structures.

Figure 29 shows similar CD overlays for the FABP-modified duplexes (red) with those of the unmodified controls (blue) in both dC and dT series. We observed FABP-induced blue shifts. However, they were generally smaller (dC series: 4 nm at n-1, 5 nm at n to n+3, 3 nm at n+6; dT series: 1 nm at n-1, 2 nm at n to n+3, 1 nm at n+6) (Table 6) than the FAAF series. These data indicated a relatively smaller DNA backbone bending in the FAAF case. In the dC series, the intensity at 270 nm was greater than that in the controls during the early stage of bulge formation (n-1 to n+1). However, subsequent TLS decreased from n+2 to n+6. By contrast, FABP in the dT series showed consistently low intensity at 270 nm relative to the unmodified controls with minimal blue shifts (1 nm).

Dynamic ^{19}F-NMR

***Nar*I-FAAF-SMI-2-dC:** Figure 30 shows dynamic ^{19}F NMR spectra of FAAF-modified -2 SMI TLS models (n-1, n, n+1, n+3 and n+6) for the dC series (see Figure 2 for all sequences). These -2 SMI duplexes exhibited a mixture of ^{19}F signals, each representing a unique conformation with different electronic environments. The n+6 represents a full -2 deletion duplex. As discussed earlier, the G_3*C bulge -2 structure was selected for the FAAF-modified 16/14-mer -2 SMI model based on ^{1}H NMR and fluorescence results.

The ^{19}F NMR measurements were performed at 5 °C to 70 °C temperature range. All ^{19}F signals coalesce into a sharp single peak above 60 °C at around −115 ppm, which represents a fast averaging FAAF-modified single-stranded 16-mer template. The data indicated that conformational heterogeneity exists at the n-1 stage, where the 3'-end of primer was located at one base before the lesion site and the heterogeneity was

maintained even at 40 °C. The heterogeneity became more complex as bulge formation was about to occur at n and n+1. The bulge structure began maturing at n+3, and was completed at n+6. We have previously shown the ^{19}F signals owing to the B-, S-, and W-conformation of a fully paired FAAF-modified duplex to appear at −115.0 to −115.5 ppm, −115.5 to −117.0 ppm and −116.5 to −118.0 ppm ranges, respectively.[11] As mentioned above, the −115 ppm signals at the coalescence temperatures are attributed to the denatured single strand in which the ^{19}F tag is fully exposed to the solvent. This signal is usually in sync with B-type conformer, in which the ^{19}F tag is exposed and thus shifted to downfield. The shielded signal at −116.3 ppm can arise from the Van der Waals interactions between the ^{19}F tag and neighboring base pair as in the S- or W-conformer. However, the current model is a -2 bulge structure without discernible major or minor groove configurations. As a result, two major ^{19}F signals at −115.5 and −116.4 ppm at 20 °C in the n+6 duplex (e.g., completed -2 bulge structure) could be assigned to either "lesion-exposed" (B-SMI) or "lesion-stacked" (S-SMI) conformers (Figure 1). A small signal at −114.8 ppm was observed at lower temperatures (5 °C to 10 °C) and coalesced with the B-SMI signal at 20 °C. The identity of this minor thermally unstable conformer could not be characterized. The B- and S-SMI designation can only be made at the n+1, n+3, and n+6 duplexes, where two well-defined signals were obtained. The conformers observed for the n-1 to n+1 duplexes comprise a mixture of narrow and broad signals, which could be assigned to various conformationally flexible species, including the B- and S-SMI originated from the immaturity of the corresponding -2 bulge structures.

***Nar*I-FAAF-SMI-2-dT**: Similar dynamic 19F NMR experiments were performed for the dT series (Figure 31). We observed a much greater heterogeneity at the n-1 and n

duplexes than at the dC series; at least four different conformations were found in the −113 ppm to −117 ppm range. This ^{19}F signal complexity could be ascribed to numerous intermediate conformers possibly near the lesion at the beginning of bulge formation. The spectral pattern was simplified at n+1 presumably because of an increased conformational stability, indicating a near completion of the G_3*C bulge structure. A similar pattern persisted at n+3 with one major and one minor signal at −115.3 and −116.2 ppm, respectively. These results indicate that a primer elongation of three bases after the lesion site is enough to produce a stable -2 SMI and the pattern continue into a full bulge duplex at n+6 with slight changes.

In the dC series, the major downfield and minor upfield signals in the n+6 duplex were assigned to the solvent exposed B- and inserted stacked S-SMI conformers, respectively. As expected, the minor upfield S-SMI signal gradually coalesces into the B-SMI signal at 55 °C. The merged signal broadened at 60 °C and then sharpened at 70 °C owing to the denaturation to a single-stranded template. An exclusive presence of the B-SMI-2 in the dT series is contrasted to a 59:41 mixture of B- and S-SMI observed for the dC series. The results are consistent with the thermal and thermodynamic instabilities observed from UV melting and DSC. The 20 °C spectra of the n+6 SMI were simulated by line fittings (Figure 34). The simulation results showed 59% B-SMI (−115.5 ppm) and 41% S-SMI (−116.3 ppm) conformer in the dC series and 86% B-SMI (−115 ppm) and 14% S-SMI (−116 ppm) conformers in the dT duplex.

***Nar*I-FABP-SMI-2**: Figures 35 and 36 show the dynamic ^{19}F NMR spectra of the FABP- TLS for the formation of a -2 SMI (n-1, n, n+1, n+3, n+6) in the dC and dT series. Unlike the FAAF duplex cases, FABP exhibited a relatively simple

conformational heterogeneity throughout TLS. The simplicity is more pronounced in the dC series relative to the dT. The n+6 duplex exhibited 86: 14 ratio of the B- and S-SMI conformers. One major signal dominated in the n+1to n+6 sequence at 5 °C. However, a small peak at around −117 ppm increased along with the temperature increase in n+3 and n+6 duplexes. In particularly, two major peaks existed in n+6 at 50 °C, then exchanged, and eventually merged at 65 °C. Unlike the dC series, the dT series simplified as one major peak from n-1 to n+6. Except at 40 °C, a second minor peak showed up at n+1, n+3, and n+6, although two peaks merged at 60 °C.

Imino proton NMR

Figures 32 and 33 show the dynamic imino proton spectra of the *Nar*I-FAAF-SMI-2-dC and *Nar*I-FAAF-SMI-2-dT series under the same primer elongation and temperature conditions. Generally, A:T and G:C imino proton signals owing to Watson−Crick base pairs appear in the 13 ppm to 14 ppm and 12 ppm to 13 ppm ranges, respectively. The imino proton signal intensity decreases with increasing temperatures caused by fast proton exchanges. As expected, the G: C imino signals are more resistant to temperature than the A: T ones. Notably, the imino protons of the lesion-modified dG in the highly shielded 11 to 12 ppm range were relatively resistant to temperature and solvent exchange. This finding is particularly true for n+3 and n+6 cases in both dC and dT series. Similar imino proton spectral transitions were obtained with the FABP-modified TLS system (Figures 37 and 38). In summary, the imino proton NMR results generally support the sequence dependent conformational heterogeneity observed in the ^{19}F NMR experiments.

Surface plasmon resonance

We used surface plasmon resonance (SPR) to examine the binding interactions of the modified 16-mer templates as a function of primer length during TLS involvement in the formation of -2 bulge structures. The FAAF/FABP modified sequences were either full length or -2 SMI duplexes in the dC and dT series. Figures 44 and 45 show the SPR set up used in the present study. The procedure is similar to that reported previously,[38] which involves modified biotinylated 16-mer sequence on a streptavidin-coated carboxymethylated surface with addition of various primers as flow-through analytes.

Biotinylated *Nar*I 16-mer sequence: FAAF/FABP modification was performed as usual with 5'-biotin labeled 16-mer sequence to prepare the TLS samples used for NMR/CD and thermodynamic experiments. The HPLC profiles of the biotin-16-mers are expected to differ because of biotin's hydrophobicity. No HPLC separation of adducted biotinylated oligonucleotides has been reported in the literature. After repeated attempts, we found a system (see the Materials and Methods) that allowed separation of FAAF modified biotinylated sequences in the dC and dT series in less than 20 and 35 min, respectively (Figures 40a and 40b). Three mono-FAAF-modified oligos were separated in the 12 min to 17 min range. Mono-FABP modified oligos were similarly separated in 45 -55 min (Figure 41).

The modified biotinylated 16-mer sequence templates were characterized by 3'-5' exonuclease digestion followed by MALDI-TOF. Figure 42a shows the spectra obtained from peak 2 of the dC series, which displayed the parent ion at 5424 m/z before digestion and major fragments at 3625 and 3336 m/z upon digestion. These findings indicated FAAF modification at G_3. Figure 42b shows the spectra obtained for peak 2 in the dT series, which exhibited the parent ion at 5439 m/z and major persistent fragments at 3625

and 3336 m/z. The results indicated FAAF modification at G_3. FABP modified dC/dT series was similarly characterized. Figure 43a shows the following spectra of the dC series: the parent ion at 5370 m/z before digestion and a persistent peak at 3282 m/z, which indicated G_3 modification. Figure 43b shows the following spectra of the FABP modified dT series: the parent ion at 5385 m/z and a persistent fragment ion 3283 m/z, which indicated G_3 modification.

SPR setup: The SPR experiments aimed to measure real-time association between template and complementary strands in the absence of a polymerase The FABP/FAAF-G_3 modified 16-mer *Nar*I 16-mer sequences characterized above were individually coated on the streptavidin pre-immobilized chip. Binding strengths were measured by injecting primers of different lengths (Figures 44 and 45). Each elongation required different concentrations of complementary strands to achieve steady-state associations: (n-1) 25 nM, (n) 50 nM, (n+1) 75 nM, (n+2) 100 nM, (n+3) (n+6) (n+8) 150 nM. After reaching a steady state, primers were washed off by a running buffer, and the system was regenerated by NaOH addition. Figure 44 shows the sensorgrams obtained from the FAAF-modified TLS systems either in the fully extended (top) or -2 SMI (bottom) as a function of primer length. Measurements were conducted in both dC (left) and dT (right) series. A typical binding affinity of KD (ka/kd) kinetics could not be applied in the present case because the association rate constant (ka) is concentration dependent, yet we used different concentrations for each length primer. DNA strand binding is also not an amenable traditional KD designed for weak macromolecular bindings. However, all the experimental conditions were kept identical for each primer. Thus, the response units (RUs) and more importantly, the dissociation rate constant (kd)

could be used to compare the binding strengths during a simulated TLS. The results should provide the extent of lesion effect during the TLS. As such, we conducted fitting of kd curves (Figures 48 and 49) using Scrubber (BioLogic Software). The resultant kd values are summarized in Tables 7 and 8.

Unmodified control models: In the fully extended model (Figures 44a and 44b), minimal changes occurred with increasing primer lengths at n-1 (16-/7-mer) and n (16-/8-mer), but a significant increment of RU values began from n+1 (16-/9-mer) onwards. In contrast with -2 SMI models (Figures 44e and 44f), abrupt changes in RU intensities (20 RU to 35 RU) and faster dissociations were observed between n (16-/8-mer) and n+2 (16-/10-mer) steps. These results support formation of a -2 bulge structure during the TLS steps. However, the bulge, G_3C or CG_3, formed for the unmodified sequences is unknown. In both full and -2 SMI scenarios, the nature of the 'N' base exhibited a minimal effect on binding characteristic, i.e., comparison between dC (Figures 44a, 44c, 44e and 44g) and dT (Figures 44b, 44d, 44f, and 44h) series.

***FAAF modified fully extended TLS*:** The FAAF-modified full length dC series (Figure 44c) exhibited a gradual increment of RU during the TLS, i.e., n-1 to n+8. Faster dissociation rates compared with the unmodified controls, particularly on n+1 to n+3, indicated weak binding strength around the lesion. Notably, a significant increase of RU (up to 20 RU) occurred at the lesion site n. By contrast, in the full-length dT series (Figure 44d), the RU intensities were significantly suppressed throughout TLS, with the effect much greater in the bulge area (n-1 to n+2). The dissociation rates up to n+3 were also much faster than those of the unmodified controls.

FAAF modified -2 SMI TLS: The FAAF-modified -2 SMI models (Figures 44g and 44h)

showed fairly gradual increment of RU with increasing primer lengths. The overall RU intensities were greater for dC (Figure 44g) over the dT (Figure 44h) series. Interestingly, the overall sensorgram patterns of the FAAF -2 SMI are quite similar to those of the unmodified fully extended (Figures 44a and 44b). Significant increases also occurred in dissociations rates. These results suggested the strengthening of the template-primer binding affinities by the FAAF-lesion at G_3. For example, the binding strength of FAAF-modified SMI at n+1 (16-/9-mer, pink) was increased by 1.92-fold (2.67/1.39) and 3.30-fold (35/10.6) relative to the unmodified SMI controls for the dC and dT series, respectively. The lesion effect persisted at n+2 (30.3-fold, 3/0.099) (26.5-fold, 14.8/0.559), n+3 (7.58-fold, 0.311/0.041) (46.2-fold, 8.08/0.175), and n+6 (7.8-fold, 0.146/0.0188) (3.2-fold, 0.0898/0.0279) for the dC and dT series.

FABP model series. We conducted an identical set of SPR experiments as above except that the lesion was switched to FABP. The two model systems exhibited a basically similar SPR binding and dissociation characteristics on both unmodified controls (Figure 45a, 45b, 45e, and 45f) and FABP modified fully extended (Figures 45c and 45d) and -2 SMI (Figure 45g and 45h) models. However, the RU intensities of the n to n+2 were notably suppressed by FAAF compared with FABP in the dT series (Figure 44d vs. Figure 45d and Figure 44h vs. Figure 45h) for full length extended and -2 SMI, respectively. Table 8 shows the binding strength for FABP modified -2 SMI that increased relative to the following unmodified controls: 1.21-fold (2.22/1.83) and 3.31-fold (33/9.98) at n+1; (209-fold, 2.93/0.014) (39.6-fold 15.9/0.4017) at n+2; (6.57-fold, 0.4207/0.064) (58.3-fold, 7.46/0.128) at n+3; (6.8-fold, 0.1914/0.02808) (2.63-fold, 0.0792/0.0301) at n+6.

Figure 46 overlays the steady-state normalized sensorgrams derived from four TLS models: unmodified full and -2 SMI; FAAF-modified full length and -2 SMI. The binding strength can be gleaned directly from the shapes of these dissociate curves. The binding strength at n+1 were in the order of unmod full > FAAF -2 SMI > FAAF full ~ unmod -2 SMI; FAAF -2 SMI > unmod full > FAAF full > unmod -2 SMI at n+2; FAAF -2 SMI > unmod full > unmod -2 SMI > FAAF full at n+3; FAAF -2del ~ unmod full ~ FAAF -2 SMI > unmod -2 SMI at n+8/n+6. In the dT series, the unmodified full model was most stable: unmod full > FAAF full > FAAF -2 SMI > unmod -2 SMI at n+1; unmod full > FAAF -2 SMI > FAAF full > unmod -2 SMI at n+2; unmod full > FAAF -2 SMI > FAAF full > unmod -2 SMI at n+3; unmod full ~ FAAF full > FAAF -2 SMI > unmod -2 SMI at n+8/n+6.

Figure 47 overlays the steady-state normalized sensorgrams derived from four TLS models: unmodified full and -2 SMI; FABP-modified full length and -2 SMI. In the dC series, the binding strength at n+1 were in the order of unmod full > FABP -2 SMI > FABP full = unmod -2 SMI; unmod full = FABP -2 SMI > FABP full > unmod -2 SMI at n+2; unmod full ~ FABP -2 SMI > FABP full > unmod -2 SMI at n+3; unmod full ~ FABP full ~ FABP -2 SMI > unmod -2 SMI at n+8/n+6. In the dT series, the unmod full > FABP full > FABP -2 SMI > unmod full at n+1; unmod full > FABP -2 SMI > FABP full > unmod -2 SMI at n+2; unmod full > FABP -2 SMI > FABP full > unmod -2 SMI at n+3; unmod full ~ FABP full ~ FABP -2 SMI ~ unmod -2 SMI at n+8/n+6.

Discussion

We conducted a series of systematic structural studies to probe the conformational

mechanisms of arylamine-induced -2 frameshift mutation frequently observed in the *E. coli Nar*I sequence (5'---CCGGCG*C*N*---3'; *N*= dC and dT) during translesion synthesis (TLS). We used two well-characterized fluorinated DNA lesions, FAAF and FABP, as models for carcinogen 2-aminofluorene and 4-aminobyphenyl. The objective of the present study was to determine the conformational consequences of lesion structures (size, bulkiness, and overall topology) and the 3'-next flanking base sequence *N* (dC or dT) in generating a -2 slipped mutagenic intermediate (SMI), the bulge structure responsible for AAF-induced -2 frameshift mutagenesis. We previously showed that the bulky *N*-acetylated and planar AAF utilizes a mixture of S/B/W-conformations, whereas the less bulky *N*-deacetylated and non-planar ABP exists mostly in B-type conformation.[4, 11] The AAF-induced conformational heterogeneity was largely dependent on the nature of flanking bases around the lesion (NG*N sequence context), which in turn led to different mutational and repair outcomes. Earlier, we showed that the AF-modified -2 SMI 12-mer duplex with N=C ($CTCG_1G_2CG_3$*C***C**ATC) adopts exclusively an "inserted" stacked S conformer (S-SMI), whereas the same duplex with N=T ($CTCG_1G_2CG_3$*C***T**ATC) exists in a mixture of S-SMI and the "solvent exposed" B-type B-SMI conformers.[28] These results explain why the unusual frameshift vulnerability of the AF lesion at G_3 is dictated by the nature of the next flanking base ***N*** (C >> T). However, the detailed conformational mechanisms of SMI formation have yet to be elucidated, which is the subject of the present study.

We hypothesized that the conformational, thermodynamic, and binding stabilities of -2 SMI are critical factors to determine the efficacy of frameshift mutations in the *Nar*I sequence context. In this study, we examined the conformational details of how a bulky

lesion favors a certain specific type of bulge structure during TLS. As mentioned in the Results section, the Streisinger-based -2 bulge formation in the *NarI* sequence would allow two possible -2 SMI structures, G_3*C or CG_3* looped out, albeit producing an identical -2 deletion daughter strand. Basing on the NMR and fluorescence results, we selected the G_3*C and CG_3* models for FAAF or FABP, respectively (Figure 2). We utilized the fluorine-tagged FAAF and FABP as model lesions for AAF and ABP to obtain dynamic ^{19}F NMR, which allows the measurement of conformational heterogeneity. FAAF represents an *N*-acetylated 2-aminofluorene and is therefore rigid, bulky, and coplanar. By contrast, the *N*-deacetylated FABP is conformationally flexible, less bulky, and nonplanar. We performed in two different 3'-next flanking base sequence contexts for the dC and dT series [CTCTC$G_1G_2CG_3CN$ATCAC-3' N =C: dC series or N=T: dT series; G_3*=FAAF or FABP] (Figure 2b, c). The choice for these sequences was based on previous mutation studies, which indicated that the bulky lesion AAF at G_3 position induces -2 deletion mutations at the highest frequency, and their propensity is modulated by the nature of the nucleotide in the *N* position (C >> T). We acquired a combination of biophysical parameters to elucidate the conformational mechanism for the formation of -2 SMI in a chemically simulated TLS. Figure 4 presents a model for FAAF and FABP-induced -2 frameshift mutagenesis on the basis of our findings in this study. This schematic diagram shows the progression of lesion-induced conformational heterogeneity during a simulated TLS (n-1 to n+6) (Figure 39).

Overall TLS: At n-1, the modified G_3* at the ds/ss replication fork is likely to exist in a mixture of *syn* and *anti*-glycosidic and closely related conformations. This phenomenon is illustrated in Figure 39 in which conformationally flexible lesions (FAAF or FABP)

are shown in multiple red-dotted ovals (labeled A). This is supported by a complex combination of sharp and broad signals observed in the ^{19}F NMR spectra (Figures 30, 31, 35, and 36). Upon addition of the correct C, the modified G_3* will produce an unstable G_3*: C base pair at the lesion site n (B). The presumed G_3*: C pair (B) is expected to be less stable than the regular Watson-Crick-based G:C pairing because the bulky arylamine at the C8 of dG tends to favor a *syn*-glycosidic conformation. Such instability and heterogeneity can cause a polymerase to stall, so a slippage occurs, or the DNA synthesis could be completely blocked for recruitment of bypass polymerases. The ^{19}F NMR data support the presence of a complex conformational heterogeneity at both pre (n-1) and lesion (n) sites.

The above-mentioned G_3*: C pair undergoes two different slippage pathways via single-base 'C' or two-base 'CG' from the 3'-terminal of the 8-mer primer, which yields CG_3* and G_3*C -2 bulge SMI structures C and D, respectively (Figure 39). As mentioned earlier, G_3*C SMI (D) is preferred by FAAF, whereas CG_3* SMI is favored by FABP. As evidenced by NMR, in both cases, the formation of -2 bulge significantly shapes up at n+1 (C and E for 'C' and 'CG' slippage, respectively) (see Figures 30, 31, 35 and 36). The conformational flexibility and instability involving a bulge formation at n-1 ~ n+1 sites are reinforced by little thermal (ΔTm) and thermodynamic (ΔΔG) changes, a lower surface resonance (SPR) response units (RU), and faster dissociation rates (kd) relative to the unmodified controls (Figure 20–26, Tables 1–5 and Tables 7-8).

Conformation rigidity improves as the length of the primer increases, which is again evidenced by the dynamic ^{19}F NMR spectra. The continued progressive TLS from n+1 produced two very different -2 SMI conformers at n+3 and ultimately in the fully

matured n+6 (F and G). They are the B-SMI conformer (F1 and G1 for FABP and FAAF, respectively) in which the lesion is solvent exposed as in the B-type conformation, or S-SMI (F2 and G2 for FABP and FAAF, respectively) in which the lesion is solvent protected and inserted/stacked with the presumed *syn*-glycosidic G_3*. Figure 39 inset illustrates the progressive nature of lesion conformational rigidity during the simulated TLS, i.e., light dotted > solid dotted > solid lines as a function of primer lengths. For example, the B-SMI (light dotted) is likely to be conformationally more flexible than the base stacked S-SMI (solid dotted) at n+1(as in C and E), and the development continues to improve the rigidity for B- and S-SMI at fully paired n+6 duplexes (solid dotted and solid lines in F and G, respectively). The multiplicity of light dotted ovals indicates conformational flexibility within the *syn* or *anti*-glycosidic conformers. The progressive nature of bulge stability is also supported by appropriate variances in Tm and ΔH values (UV-melting and DSC), as well as the SPR binding characteristics (RU) and dissociation rates (kd) during bulge formation (Figure 20–26, Tables 1–5 and Table 7-8).

Lesion Effect: The structures of FABP and FAAF are generally similar in that both are C8-substituted dG lesions, but they differ in two major ways: 1) FABP lacks a bridging methylene carbon, so it is less coplanar than FAAF, and 2) FAAF is *N*-acetylated and is thus steric near the adduction point and perturbs the DNA helix. We found that FABP prefers, on average, B-SMI (~90% B) over S-SMI (5-10%) regardless of the nature of the 3'-flanking base N (dC or dT). This is in contrast to FAAF, which showed a mixture of B-SMI (59% and 86%) and S-SMI (41% and 14%) for the dC and dT series, respectively. These results indicate the importance of the relative nonplanarity of FABP over FAAF in producing a great amount of S-SMI. Our ^{19}F NMR data are in

general agreement with those of Milhe's ^{1}H NMR study of an AAF-modified *Nar*I-based 11/9-mer duplex (5'-ACCGGCG*CCACA-3')(5'-TGTG--GCCGGT-3'), which showed 80% of syn-modified dG* S-SMI conformation.[36] A similar study by Mao et al.[5] showed an exclusive presence of stacked S-SMI conformation for *N*-deacetylated AF in the *Nar*I-based 12/10-mer duplex (5'-CTCGGCG*CCATC-3') (5'-GATGG--CCGAG-3'). The latter appears to be in direct contrast to that observed for the similarly *N*-deacetylated ABP lesion in the present study. No NMR structures are available yet on ABP in the -2 SMI duplex, so a direct comparison is not possible. Nonetheless, our results support the importance of lesion coplanarity in producing S-SMI. The significant reduction in the population of the stacked *syn*-G* in FAAF (14%–41%) over FABP (~90%) in the *Nar*I-based -2 deletion duplexes indicates the importance of the '*N*-acetyl factor' and lesion coplanarity in producing -2 bulge S-SMI.

FAAF and FABP-modified -2 SMI at n+6 showed consistently greater thermal and thermodynamic stabilities relative to the fully paired counterparts (Tm 10.5 °C to 11.7 °C and ΔΔG −2.6 to −3.2 kcal/mole for FAAF and Tm 7.5 °C to 8.0 °C and ΔΔG −1.1~ −2. kcal/mole for FABP). This is in contrast to the consistent decreases in thermal and thermodynamic stabilities observed for the corresponding fully paired complementary duplexes (Tables 1–4) examined in the present and previous studies. The thermal stability increases as the bulge formation matures from n+1 to n+6, and the trend persists for both FABP and FAAF throughout TLS, as illustrated in Figure 39 (Tables 1–4). A gradual increase in thermal stability, however, seems to be inconsistent with the striking conformational differences in B- and S-SMI observed between FAAF and FABP. The result indicates that lesion stacking and bulge formations are both important factors

that contribute to the stability of -2 bulge duplexes.

Sequence Effect on Bulge Formation: Another interesting finding is the effect of the 3'-next flanking N base (dC *vs.* dT series, Figure 39) on bulge formation, i.e., FABP at n+6 (F) prefers, on average, B-SMI (86% and 94% B) over S-SMI (14% and 6%) for the dC and dT series, respectively. In other words, no discernible difference in conformational population was observed for FABP between the dC and dT series. However, that was not the case for FAAF, which showed a significant S/B-population difference between the two series: 59%: 41% of B- and S-SMI for the dC series and 86%:14% of B- and S-SMI for the dT series. These results are consistent with our proposed model (Figure 39), which contends the importance of the "lesion coplanarity" and "*N*-acetyl" factor. For example, the planar and hydrophobic lesion in the FAAF-induced S-SMI (G2) maintains direct molecular interactions with the N-N' base pair in the bulge structure. However, no such contacts are likely for the FABP-induced CG_3* bulge S-SMI (F2). As a result, the N-N' base pair has a much greater influence on the S-SMI population (41% and 14% for the dC and dT series) of the CG_3* bulge FAAF pathway (G) than that (10% and 5% for the dC and dT series) of the G_3*C bulge pathway FABP (F). The hydrogen bond strength of the N-N' base pair clearly plays a role, so the three hydrogen-bonded C: G at N position enables a stacking interaction with the lesion better than the two hydrogen-bonded T:A base pairs. Furthermore, the stability of the N-N' base pair helps docking and stacking of the bulky hydrophobic lesion into the small pocket of -2 SMI. The calorimetric data (DSC) show a consistent increase in Tm, □H, and □G for the dC series over the dT ones (See Figure 26, Table 5). The DSC results are also consistent with SPR binding affinity data, which exhibited consistently high RU

intensities and slow dissociation rates (kd) throughout TLS in both FAAF and FABP. The effect of the 3'-next flanking base sequence is also reinforced by CD spectra (Figure 27), in which the FAAF-dC series showed a gradual increase in ellipticity at 270 nm during n-1 → n → n+1 progression. The CD results indicated a progressive strengthening of base stacking. However, no such change was noted in the dT series. A similar set of results was obtained for FABP. Taken together, the CD results indicate the effect of the 3'-ext flanking base on lesion stacking within the -2 bulge structures.

'C' versus 'CG' Streinger Slippage: Finally, our study sheds some light on a fundamental question as to why FAAF and FABP undergo unique slippage during TLS ('CG' and 'C' for the B→C and B→D pathways, respectively). As detailed in the Results section above, the available ^{1}H NMR, mass, and fluorescence results facilitated the CG_3* and G_3*C -2 bulge as the most likely scenarios for the respective FABP- and FAAF-induced TLS pathways. The conformational stability of the G*: C pair at the lesion site n (B) appears to be a major determining factor. In fully paired complementary duplexes, the bulky *N*-acetylated FAAF prefers *syn*-glycosidic G* conformation, whereas FABP is *N*-deacetylated and conformationally flexible, thus exists mostly in the *anti*-glycosidic B-type conformer. A similar conformational preference is expected at the ss/ds replication fork. As a result, *syn*-FAAF-G*: C is expected to produce great conformational instability, which triggers a two-base ('CG') slippage. By contrast, the flexible *anti*-FABP-G*: C is less of a trigger; it induces a one-base 'C' slippage. Once the slippage pathway is defined, subsequent primer elongation continues to produce respective -2 bulge structures (F and G), each resulting in a mixture of B- and S-SMI conformations.

We previously studied the effect of the 3'-next flanking base (N= dC or dT) by

using FAF-modified *Nar*I-based 12/10-mer -2 deletion [(5'-CTCGGCG*C***N**ATC-3') (5'-GATNGCCGAG-3')] duplexes. The results showed that the *Nar*I-dC/-2 deletion duplex exhibits mostly an S-SMI conformer, whereas the *Nar*I-dT/-2 deletion duplex exists as a mixture of the S- and B-SMI conformers. Schorr and Carell have conducted an elegant primer extension study coupled with MALDI-TOF[26] mass spectrometry to show that frameshift formation is triggered by the unstable base pairing of the AAF lesion with the correct incoming dC. Such configurations have been observed in both replicative and bypass polymerases, and the stability of bulged-out structures and the subsequent elongation determines the propensity for frameshift mutagenesis. To this end, we conducted studies on FAAF-modified *Nar*I-sequence corresponding to −1, −2, and −3 deletion duplexes.[27] These SMIs existed in a mixture of B- and S-SMI conformers, with the population of the S conformer and the thermodynamic stability in the order of −1 > −2 > −3. The results indicate the good stability of S-SMI, which supports the results of the aforementioned work of Schorr and Carell, as well as emphasizes the importance of SMI stability for frameshift mutations.

In summary, we presented a conformational TLS model for arylamine-induced -2 frameshift mutagenesis in the *E. coli Nar*I mutational hot spot sequence. Figure 39 shows a cartoon depiction of the conformational details of the proposed model, which are based on a combination of systematic spectroscopic (^{19}F NMR/CD), thermodynamic (UV-melting/DSC), and affinity binding (SPR) data. Our findings indicate that the Streisinger-based -2 bulge formation is triggered by several factors, including the adduct structure and conformation at the replication fork, as well as the nature of base sequences

surrounding the lesion site. The extent of conformational stability of the G_3*: C pair determines the nature of a slippage ('CG' vs. 'C'), and subsequent primer elongation yields the respective -2 CG_3* or G_3*C bulge structures. Each bulge structure exists in a mixture of B-SMI and S-SMI, in which the bulky lesion is located outside the bulge ("solvent accessible") and inserted into the bulge ("solvent protected"). B-/S-SMI population ratios are dependent on various structural factors, such as the size, bulkiness ('*N*-acetyl), coplanarity, and overall topology of a lesion, as well as the 3'-base sequence (N) next flanking to the lesion site.

Reference

[1] Luch A. Nature and nurture - lessons from chemical carcinogenesis. Nature reviews Cancer. 2005;5:113-25.

[2] Shibutani S, Grollman AP. Molecular mechanisms of mutagenesis by aromatic amines and amides. Mutation research. 1997;376:71-8.

[3] Heflich RH, Neft RE. Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. Mutation research. 1994;318:73-114.

[4] Patnaik S, Cho BP. Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. Chemical research in toxicology. 2010;23:1650-2.

[5] Patel DJ, Mao B, Gu Z, Hingerty BE, Gorin A, Basu AK, et al. Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance. Chemical research in toxicology. 1998;11:391-407.

[6] Hsu GW, Kiefer JR, Burnouf D, Becherel OJ, Fuchs RP, Beese LS. Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. The Journal of biological chemistry. 2004;279:50280-5.

[7] Dutta S, Li Y, Johnson D, Dzantiev L, Richardson CC, Romano LJ, et al. Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis. Proceedings of the National Academy of Sciences of the United States of America. 2004;101:16186-91.

[8] Shibutani S, Suzuki N, Tan X, Johnson F, Grollman AP. Influence of flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells. Biochemistry. 2001;40:3717-22.

[9] Grollman AP, Shibutani S. Mutagenic specificity of chemical carcinogens as determined by studies of single DNA adducts. IARC scientific publications. 1994:385-97.

[10] Meneni SR, Shell SM, Gao L, Jurecka P, Lee W, Sponer J, et al. Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. Biochemistry. 2007;46:11263-78.

[11] Jain V, Hilton B, Patnaik S, Zou Y, Chiarelli MP, Cho BP. Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. Nucleic acids research. 2012;40:3939-51.

[12] Cho BP. Dynamic conformational heterogeneities of carcinogen-DNA adducts and

their mutagenic relevance. Journal of environmental science and health Part C, Environmental carcinogenesis & ecotoxicology reviews. 2004;22:57-90.

[13] Meneni S, Liang F, Cho BP. Examination of the long-range effects of aminofluorene-induced conformational heterogeneity and its relevance to the mechanism of translesional DNA synthesis. Journal of molecular biology. 2007;366:1387-400.

[14] Meneni S, Shell SM, Zou Y, Cho BP. Conformation-specific recognition of carcinogen-DNA adduct in escherichia coli nucleotide excision repair. Chemical research in toxicology. 2007;20:6-10.

[15] Vaidyanathan VG, Liang F, Beard WA, Shock DD, Wilson SH, Cho BP. Insights into the conformation of aminofluorene-deoxyguanine adduct in a DNA polymerase active site. The Journal of biological chemistry. 2013;288:23573-85.

[16] Rechkoblit O, Kolbanovskiy A, Malinina L, Geacintov NE, Broyde S, Patel DJ. Mechanism of error-free and semitargeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4. Nature structural & molecular biology. 2010;17:379-88.

[17] Cho BP, Beland FA, Marques MM. NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. Biochemistry. 1994;33:1373-84.

[18] Broschard TH, Koffel-Schwartz N, Fuchs RP. Sequence-dependent modulation of frameshift mutagenesis at NarI-derived mutation hot spots. Journal of molecular biology. 1999;288:191-9.

[19] Fuchs RP, Fujii S. Translesion synthesis in Escherichia coli: lessons from the NarI mutation hot spot. DNA repair. 2007;6:1032-41.

[20] Koffel-Schwartz N, Fuchs RP. Sequence determinants for -2 frameshift mutagenesis

at NarI-derived hot spots. Journal of molecular biology. 1995;252:507-13.

[21] Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, et al. Frameshift mutations and the genetic code. Cold Spring Harbor symposia on quantitative biology. 1966;31:77-84.

[22] Garcia-Diaz M, Kunkel TA. Mechanism of a genetic glissando: structural biology of indel mutations. Trends in biochemical sciences. 2006;31:206-14.

[23] Schorr S, Schneider S, Lammens K, Hopfner KP, Carell T. Mechanism of replication blocking and bypass of Y-family polymerase {eta} by bulky acetylaminofluorene DNA adducts. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:20720-5.

[24] Kirouac KN, Basu AK, Ling H. Replication of a carcinogenic nitropyrene DNA lesion by human Y-family DNA polymerase. Nucleic acids research. 2013;41:2060-71.

[25] Gill JP, Romano LJ. Mechanism for N-acetyl-2-aminofluorene-induced frameshift mutagenesis by Escherichia coli DNA polymerase I (Klenow fragment). Biochemistry. 2005;44:15387-95.

[26] Schorr S, Carell T. Mechanism of acetylaminofluorene-dG induced frameshifting by polymerase eta. Chembiochem : a European journal of chemical biology. 2010;11:2534-7.

[27] Sandineni A, Lin B, MacKerell AD, Jr., Cho BP. Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis. Chemical research in toxicology. 2013;26:937-51.

[28] Jain N, Li Y, Zhang L, Meneni SR, Cho BP. Probing the sequence effects on NarI-induced -2 frameshift mutagenesis by dynamic 19F NMR, UV, and CD spectroscopy. Biochemistry. 2007;46:13310-21.

[29] Cho BP, Zhou L. Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by 19F NMR spectroscopy. Biochemistry. 1999;38:7572-83.

[30] Meneni SR, D'Mello R, Norigian G, Baker G, Gao L, Chiarelli MP, et al. Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. Nucleic acids research. 2006;34:755-63.

[31] Chakrabarti MC, Schwarz FP. Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry. Nucleic acids research. 1999;27:4801-6.

[32] Jain N, Meneni S, Jain V, Cho BP. Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. Nucleic acids research. 2009;37:1628-37.

[33] Vaidyanathan VG, Xu L, Cho BP. Binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo(-)) and various arylamine DNA lesions characterized by surface plasmon resonance. Chemical research in toxicology. 2012;25:1568-70.

[34] Hoffmann GR, Fuchs RP. Mechanisms of frameshift mutations: insight from aromatic amines. Chemical research in toxicology. 1997;10:347-59.

[35] Mao B, Gu Z, Gorin A, Hingerty BE, Broyde S, Patel DJ. Solution structure of the aminofluorene-stacked conformer of the syn [AF]-C8-dG adduct positioned at a template-primer junction. Biochemistry. 1997;36:14491-501.

[36] Milhe C, Dhalluin C, Fuchs RP, Lefevre JF. NMR evidence of the stabilisation by the carcinogen N-2-acetylaminofluorene of a frameshift mutagenesis intermediate. Nucleic acids research. 1994;22:4646-52.

[37] Liang F, Cho BP. Probing the thermodynamics of aminofluorene-induced translesion DNA synthesis by differential scanning calorimetry. Journal of the American Chemical Society. 2007;129:12108-9.

[38] Jain V, Vaidyanathan VG, Patnaik S, Gopal S, Cho BP. Conformational insights into the lesion and sequence effects for arylamine-induced translesion DNA synthesis: 19F NMR, surface plasmon resonance, and primer kinetic studies. Biochemistry. 2014;53:4059-71.

Figure legends

Figure 1: (a) Chemical structures of FAF/FAAF/FABP modified guanines (b) major groove views of prototype B-, S- and W- conformers of arylamine-DNA in CPK model. Color code: DNA duplex, gray; arylamine lesion, red; modified-dG, cyan; dC opposite the lesion site, green.

Figure 2: Proposed translesion synthesis (TLS) models for FAAF and FABP of *Nar*I dC/dT sequence. (A) full length extended model with full length primers (B) FAAF modified slipped mutagenic model with G_3C -2 deletion primers (C) FABP modified slipped mutagenic model with CG_3 -2 deletion primers. The red guanine G_3 position was modified by FAAF/FABP adduct, whereas unmodified guanine as control. The blue base in the template can be C or T, named as dC or dT series, respectively. The blue base in the primers is G or A which pairs with C or T.

Figure 3: (a) Slippage model cited from *Hoffmann, G. and Fuchs, R. P. Chemical Research in Toxicology 1997* (b) Slippage model for the -2 frameshift mutation by FAAF/FABP adduct on the hot spot *Nar*I sequence (5'-GGCGCN-3').

Figure 4: Proposed mechanism of -2 deletion bulge formation of AAF/AF/ABP modified *Nar*I dC/dT series.

Figure 5: (a) Chromatogram profile of the reaction mixture of FAAF modified 16-mer *Nar*I sequence. The mono-(G_1, G_2, G_3), di- and tri- FAAF adducts eluted at the 11-14, 15-18 and 19 min were purified by reverse-phase HPLC using C18 column and characterized by MALDI-TOF (b) Photodiode array UV/Vis spectra of seven peaks, in which the intensity of the 300-325 nm shoulders indicate the number of the adducts: mono-, di, and tri-FAAF adducts.

Figure 6: FAAF modified *Nar*I dC sequence chromatogram profiles from reaction mixture. (a) 25 min gradient method developed in the present project (b) 90 min method was used in previous paper (Jain et al., *Nucleic Acids Research*, 2012, Vol. 40, 3939-3951).

Figure 7: (a) Chromatogram profile of the reaction mixture of FABP modified 16-mer *Nar*I sequence. The mono-(G_1, G_2, G_3), di- and tri- FABP adducts eluted at 19-24, 34-38 and 42 min were purified by reverse-phase HPLC using clarity column and characterized by MALDI-TOF (b) Photodiode array UV/Vis spectra of seven peaks. The shoulder intensity at 300-325 nm indicate the number of the adducts: mono-, di, and tri-FABP adducts.

Figure 8: MALDI spectra of 3' and 5' enzyme digestions of FAAF dC peak 1 sample. Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3' digestion profiles of 5017 m/z at 0s corresponds to the FAAF modified 16-mer dC template. 2310 and 1981 m/z correspond to the modified lesion site of G_1 (b) 5' digestion profiles of 5017 m/z ion at 0 s shows the whole sequence and 3832 m/z peak corresponds to the fragment near the lesion G_1. Both 3' and 5' digestions show peak 1 as G_1.

Figure 9: MALDI spectra of 3' and 5' enzyme digestions of FAAF dC peak 2 sample. Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3' digestion profiles of 5017 m/z at 0s corresponds to the FAAF modified 16-mer dC template. 3218 and 2929 m/z correspond to the modified lesion site of G_3 (b) 5' digestion profiles of 5017 m/z ion at 0 s shows the whole sequence and 2883 m/z peak corresponds to the fragment near the lesion G_3, 2594 m/z peak indicates the G_3 lesion site.

Both 3’ and 5’ digestions show peak 2 as G_3.

Figure 10: MALDI spectra of 3’ and 5’ enzyme digestions of FAAF dC peak 3 sample. Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 5017 m/z at 0s corresponds to the FAAF modified 16-mer dC template. 2599 and 2310 m/z correspond to the modified lesion site of G_2 (b) 5’ digestion profiles of 5017 m/z ion at 0 s shows the whole sequence and 3542 m/z peak corresponds to the fragment near the lesion G_2, 3214 shows the G_2 lesion site. Both 3’ and 5’ digestions show peak 3 as G_2.

Figure 11: MALDI spectra of 3’ and 5’ enzyme digestions of FAAF dT peak 1 sample. Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 5031 m/z at 0s corresponds to the FAAF modified 16-mer dT template. 2310 and 1981 m/z correspond to the modified lesion site of G_1 (b) 5’ digestion profiles of 5031 m/z ion at 0 s shows the whole sequence and 3844 m/z peak corresponds to the fragment near the lesion G_1. Both 3’ and 5’ digestions show peak 1 as G_1.

Figure 12: MALDI spectra of 3’ and 5’ enzyme digestions of FAAF dT peak 2 sample. Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 5031 m/z at 0s corresponds to the FAAF modified 16-mer dT template. 3218 and 2928 m/z correspond to the modified lesion site of G_3 (b) 5’ digestion profiles of 5031 m/z ion at 0 s shows the whole sequence and 2896 m/z peak corresponds to the fragment near the lesion G_3. Both 3’ and 5’ digestions show peak 2 as G_3.

Figure 13: MALDI spectra of 3’ and 5’ enzyme digestions of FAAF dT peak 3 sample.

Molecular weight of DNA fragments of FAAF modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 5031 m/z at 0s corresponds to the FAAF modified 16-mer dT template. 2599 and 2310 m/z correspond to the modified lesion site of G_2 (b) 5’ digestion profiles of 5031 m/z ion at 0 s shows the whole sequence and 3844 , 3555 m/z peaks correspond to the fragment near the lesion G_2. Both 3’ and 5’ digestions show peak 3 as G_2.

Figure 14: MALDI spectra of 3’ and 5’ enzyme digestions of FABP dC peak 1 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 4963 m/z at 0s corresponds to the FABP modified 16-mer dC template. 1929 m/z corresponds to the modified lesion site of G_1 (b) 5’ digestion profiles of 4963 m/z ion at 0 s shows the whole sequence and 3776 m/z peak corresponds to the fragment near the lesion G_1, 3487 m/z peak indicates the G_1 lesion site. Both 3’ and 5’ digestions show peak 1 as G_1.

Figure 15: MALDI spectra of 3’ and 5’ enzyme digestions of FABP dC peak 2 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 4963 m/z at 0s corresponds to the FABP modified 16-mer dC template. 2876 m/z corresponds to the modified lesion site of G_3 (b) 5’ digestion profiles of 4963 m/z ion at 0 s shows the whole sequence and 2830 m/z peak corresponds to the fragment near the lesion G_3, 2540 m/z peak indicates the G_3 lesion site. Both 3’ and 5’ digestions show peak 2 as G_3.

Figure 16: MALDI spectra of 3’ and 5’ enzyme digestions of FABP dC peak 3 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3’ digestion profiles of 4963 m/z at 0s corresponds to the FABP modified 16-

mer dC template. 2258 m/z corresponds to the modified lesion site of G_2 (b) 5' digestion profiles of 4963 m/z ion at 0 s shows the whole sequence and 3776 m/z peak corresponds to the fragment near the lesion G_2, 3159 m/z peak indicates the G_2 lesion site. Both 3' and 5' digestions show peak 3 as G_2.

Figure 17: MALDI spectra of 3' and 5' enzyme digestions of FABP dT peak 1 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3' digestion profiles of 4980 m/z at 0s corresponds to the FABP modified 16-mer dT template. 1930 m/z corresponds to the modified lesion site of G_1 (b) 5' digestion profiles of 4980 m/z ion at 0 s shows the whole sequence and 3504 m/z peak indicates the G_1 lesion site. Both 3' and 5' digestions show peak 1 as G_1.

Figure 18: MALDI spectra of 3' and 5' enzyme digestions of FABP dT peak 2 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3' digestion profiles of 4980 m/z at 0s corresponds to the FABP modified 16-mer dT template. 2878 m/z corresponds to the modified lesion site of G_3 (b) 5' digestion profiles of 4980 m/z ion at 0 s shows the whole sequence and 2844 m/z peak corresponds to the fragment near the lesion G_3. Both 3' and 5' digestions show peak 2 as G_3.

Figure 19: MALDI spectra of 3' and 5' enzyme digestions of FABP dT peak 3 sample. Molecular weight of DNA fragments of FABP modified fragments listed in the inset boxes. (a) 3' digestion profiles of 4980 m/z at 0s corresponds to the FABP modified 16-mer dT template. 2259 m/z corresponds to the modified lesion site of G_2 (b) 5' digestion profiles of 4980 m/z ion at 0 s shows the whole sequence and 3505 m/z peak corresponds to the fragment near the lesion G2, 3175 m/z peak indicates the G_2 lesion site. Both 3' and 5' digestions show peak 3 as G_2.

Figure 20: UV thermal melting curves for four TLS models of FAAF modified sequences. (a) dC series (b) dT series.

Figure 21: UV thermal melting curves for four TLS models of FABP modified sequences. (a) dC series (b) dT series.

Figure 22: Thermal and thermodynamic parameters from UV overlay of FAAF dC sequence based on the increment of primers, left side is the comparison of sequence with full length primer and right side is comparison of sequence with -2 deletion primers; (a) comparison of melting temperature (b)comparison of -ΔG change. Blue is unmodified control and red is FAAF modified.

Figure 23: Thermal and thermodynamic parameters from UV overlay of FAAF dT sequence based on the increment of primers, left side is the comparison of sequence with full length primer and right side is comparison of sequence with -2 deletion primers (a)comparison of melting temperature; (b)comparison of -ΔG change. Blue is unmodified control and red is FAAF modified.

Figure 24: Thermal and thermodynamic parameters from UV overlay of FABP dC sequence based on the increment of primers, left side is the comparison of sequence with full length primer and right side is comparison of sequence with -2 deletion primers; (a) comparison of melting temperature (b) comparison of -ΔG change. Blue is unmodified control and red is FABP modified.

Figure 25: Thermal and thermodynamic parameters from UV overlay of FABP dT sequence based on the increment of primers, left side is the comparison of sequence with full length primer and right side is comparison of sequence with -2 deletion primers: (a) comparison of melting temperature (b) comparison of -ΔG change. Blue is unmodified

control and red is FABP modified.

Figure 26: DSC curves of FAAF recorded from 15 °C to 85 °C. (a) dC unmodified template with -2 deletion primers (b) dC G_3 FAAF modified sequence with -2 deletion primers (c) dT unmodified template with -2 deletion primers (d) dT G_3 FAAF modified sequence with -2 deletion primers.

Figure 27: CD spectral overlays of G_3-FAAF/FABP-modified sequence in three -2 deletion duplex forms. (a) dC and (b) dT with primers of n-1, n, n+1 at 25ºC. Green dot: with n primer; blue line: with n-1 primer; red dot: with n+1 -2 deletion primer.

Figure 28: CD spectral overlays of G_3-FAAF-modified sequence (red) with unmodified sequence control (blue) in -2 deletion models. (a) dC and (b) dT with primers of n-1, n, n+1, n+2,n+3 and n+6 at 25ºC.

Figure 29: CD spectral overlays of G_3-FABP-modified sequence (red) with unmodified sequence control (blue) in -2 deletion models. (a) dC and (b) dT with primers of n-1, n, n+1, n+2,n+3 and n+6 at 25ºC.

Figure 30: Dynamic ^{19}F NMR spectra of dC G_3- FAAF template paired with -2 del primers (n-1, n, n+1, n+3, n+6) from 5 to 70 °C.

Figure 31: Dynamic ^{19}F NMR spectra of dT G_3- FAAF template paired with -2 del primers (n-1, n, n+1, n+3, n+6) from 5 to 70 °C.

Figure 32: Imino proton NMR spectra of dC G_3- FAAF template paired with -2 del primers (n-1, n, n+1, n+3, n+6) from 5 to 60 °C.

Figure 33: Imino proton NMR spectra of dT G_3- FAAF template paired with -2 del primers at (n-1, n, n+1, n+3, n+6) from 5 to 60 °C.

Figure 34: Simulation of FAAF modified dC/dT duplexes from n-1 to n+6 at 20 °C.

Conformer populations show in %.

Figure 35: Dynamic ^{19}F NMR of FABP modified G_3 of dC series along with -2 deletion primers from 5 to 70 °C.

Figure 36: Dynamic 19F NMR of FABP modified G_3 of dT series along with -2 deletion primers from 5 to 60 °C.

Figure 37: Imino proton NMR sepctra of FABP modified G_3 of dC series along with -2 deletion primers from 5 to 60 °C.

Figure 38: Imino proton NMR sepctra of FABP modified G_3 of dT series along with -2 deletion primers from 5 to 60 °C.

Figure 39: Mechanism of FAAF/FABP modified *Nar*I sequence forming the bulge structure during the TLS.

Figure 40: HPLC chromatography profiles of FAAF modified 5'-biotin-*Nar*I-sequence. (a) dC sequence; mono-adducts eluted between 12-14 min (b) dT sequence mixture, mono-adducts eluted between 13-18 min.

Figure 41: HPLC chromatography profile of FABP modified 5'-biotinylated dC sequence. Mono-adducts eluted between 45-53 mins.

Figure 42: 3' SVP digestion of FAAF modified biotin dC/dT monoadduct Peak 2. (a) dC sequence, 5424 m/z ion at 0 s corresponds to the FAAF modified 5'-Biotin-16-mer dC template. The 3915 and 3625 m/z peaks correspond to the fragments near the lesion; the digestion stopped at 3336 m/z peak shows the G_3 modified site (b) dT sequence, 5439 m/z at 0 s corresponds to FAAF modified 5'-biotin-16-mer dT template. The 3625 and 3336 m/z peaks suggest the G_3 modified site.

Figure 43: 3' SVP digestion of FABP modified biotin dC/dT monoadduct Peak 2. (a) dC

sequence, 5370 m/z ion at 0 s corresponds to the FABP modified 5'-Biotin-16-mer dC template. The 3282 m/z peak corresponds to the fragments at G_3 modified site. (b) dT sequence, 5383 m/z at 0s corresponds to FABP modified 5'-biotin-16-mer dT template. The 3283 m/z peak suggests the G_3 modified site.

Figure 44: SPR sensorgrams of FAAF four stimulated models from n-1 to n+8/n+6 in dC/dT series. (a) dC unmodified in full length model (b) dT unmodified in full length model (c) dC FAAF modified in full length model (d) dT FAAF modified in full length model (e) dC unmodified in -2 SMI model (f) dT unmodified in -2 SMI model (g) dC FAAF modified in -2 SMI model (h) dT FAAF modified in -2 SMI model.

Figure 45: SPR sensorgrams of FABP four stimulated models from n-1 to n+8/n+6 in dC/dT series. (a) dC unmodified in full length model (b) dT unmodified in full length model (c) dC FABP modified in full length model (d) dT FABP modified in full length model (e) dC unmodified in -2 SMI model (f) dT unmodified in -2 SMI model (g) dC FABP modified in -2 SMI model (h) dT FABP modified in -2 SMI model.

Figure 46: Normalized SPR sensorgrams of FAAF modified four stimulated models at n, n+1, n+2, n+3 and n+8/n+6 position in (a) dC series; (b) dT series.

Figure 47: Normalized SPR sensorgrams of FABP modified four stimulated models at n, n+1, n+2, n+3 and n+8/n+6 position in (a) dC series; (b) dT series.

Figure 48: Dissociate rate constant (kd) simulated SPR sensorgrams of four different models with FAAF fitted by scrubber. Red lines are fitted and black is raw data. (a) dC series (b) dT series.

Figure 49: Dissociate rate constant (kd) simulated SPR sensorgrams of four different models with FABP fitted by scrubber. Red lines are fitted and black is raw data. (a) dC

series (b) dT series.

Table legends

Table 1: Thermal and thermodynamic parameters of G_3- FAAF-modified dC duplexes from UV melting

Table 2: Thermal and thermodynamic parameters of G_3- FAAF-modified dT duplexes from UV melting

Table 3: Thermal and thermodynamic parameters of G_3- FABP-modified dC duplexes from UV melting

Table 4: Thermal and thermodynamic parameters of G_3- FABP-modified dT duplexes from UV melting

Table 5: Thermal and thermodynamic parameters of G_3- FAAF-modified dC/dT duplexes from DSC

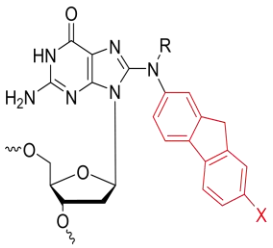
Table 6: Blue shift comparison between FAAF/FABP modified sequence and unmodified control in -2 SMI model

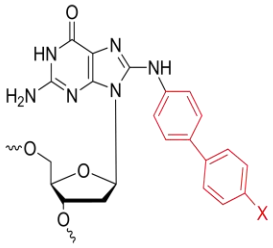
Table 7: The dissociate rate constant (k_d, s^{-1}) of individual primer in FAAF modified sequence

Table 8: The dissociate rate constant (k_d, s^{-1}) of individual primer in FABP modified sequence

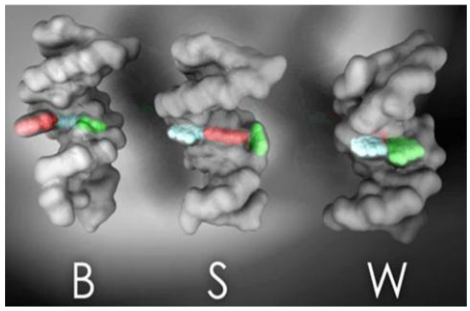
Figure 1

(a)





(b)



R=H, X=H; dG-C8-AF (AF)
R=Ac, X=H; dG-C8-AAF (AAF) **X=H; dG-C8-ABP (ABP)**
R=H, X=F; dG-C8-FAF (FAF) **X=F; dG-C8-FABP (FABP)**
R=Ac, X=F; dG-C8-FAAF (FAAF)

Figure 2

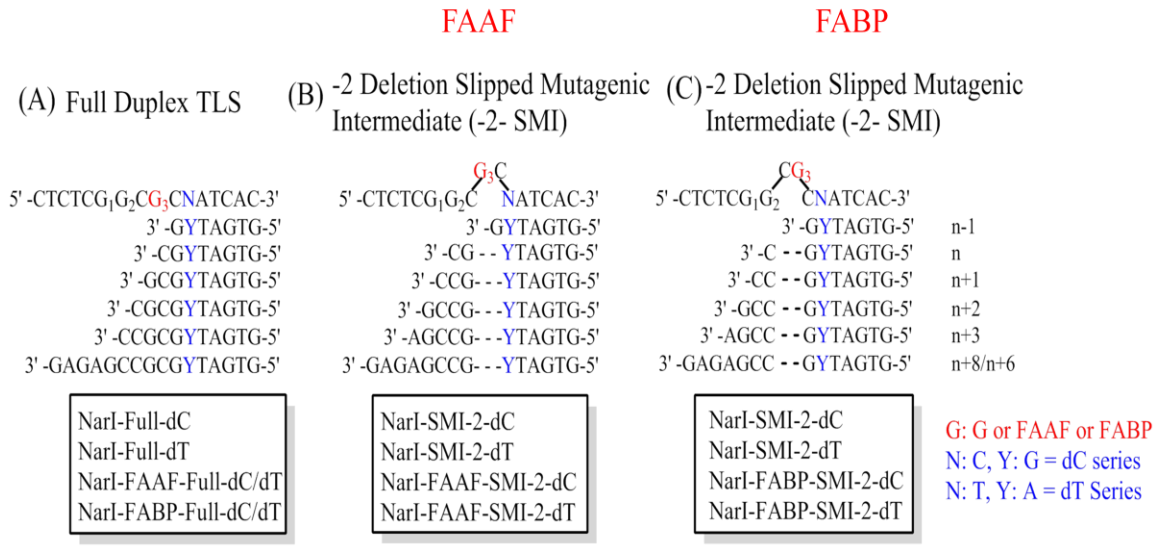

FAAF
FABP
(A) Full Duplex TLS
(B) -2 Deletion Slipped Mutagenic Intermediate (-2- SMI)
(C) -2 Deletion Slipped Mutagenic Intermediate (-2- SMI)
5' -CTCTCG1G2CG3CNATCAC-3'
3' -GYTAGTG-5'
3' -CGYTAGTG-5'
3' -GCGYTAGTG-5'
3' -CGCGYTAGTG-5'
3' -CCGCGYTAGTG-5'
3' -GAGAGCCGCGYTAGTG-5'
NarI-Full-dC
NarI-Full-dT
NarI-FAAF-Full-dC/dT
NarI-FABP-Full-dC/dT
5' -CTCTCG1G2C NATCAC-3'
3' -GYTAGTG-5'
3' -CG - - YTAGTG-5'
3' -CCG- - -YTAGTG-5'
3' -GCCG- - -YTAGTG-5'
3' -AGCCG- - -YTAGTG-5'
3' -GAGAGCCG- - -YTAGTG-5'
NarI-SMI-2-dC
NarI-SMI-2-dT
NarI-FAAF-SMI-2-dC
NarI-FAAF-SMI-2-dT
5' -CTCTCG1G2 CNATCAC-3'
3' -GYTAGTG-5'
3' -C - -GYTAGTG-5'
3' -CC - -GYTAGTG-5'
3' -GCC - -GYTAGTG-5'
3' -AGCC - -GYTAGTG-5'
3' -GAGAGCC - -GYTAGTG-5'
NarI-SMI-2-dC
NarI-SMI-2-dT
NarI-FABP-SMI-2-dC
NarI-FABP-SMI-2-dT
n-1
n
n+1
n+2
n+3
n+8/n+6
G: G or FAAF or FABP
N: C, Y: G = dC series
N: T, Y: A = dT Series

Figure 3

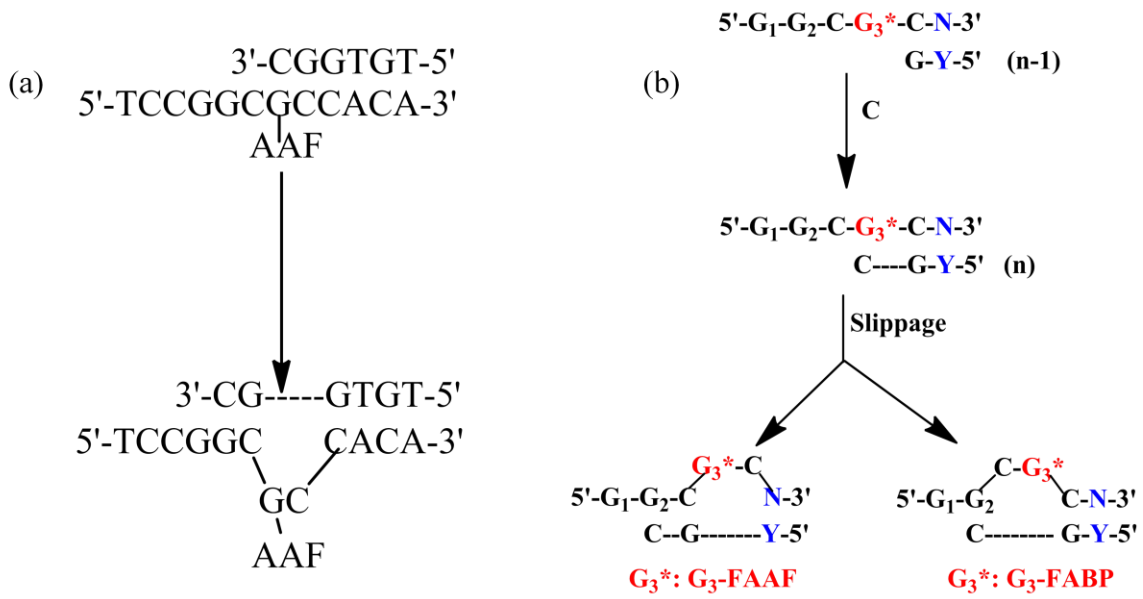

(a)
3'-CGGTGT-5'
5'-TCCGGCGCCACA-3'
AAF
3'-CG-----GTGT-5'
5'-TCCGGC CACA-3'
GC
AAF
(b)
5'-G1-G2-C-G3*-C-N-3'
G-Y-5' (n-1)
C
5'-G1-G2-C-G3*-C-N-3'
C----G-Y-5' (n)
Slippage
G3*-C
5'-G1-G2-C N-3'
C--G-------Y-5'
G3*: G3-FAAF
C-G3*
5'-G1-G2 C-N-3'
C-------- G-Y-5'
G3*: G3-FABP

Figure 4

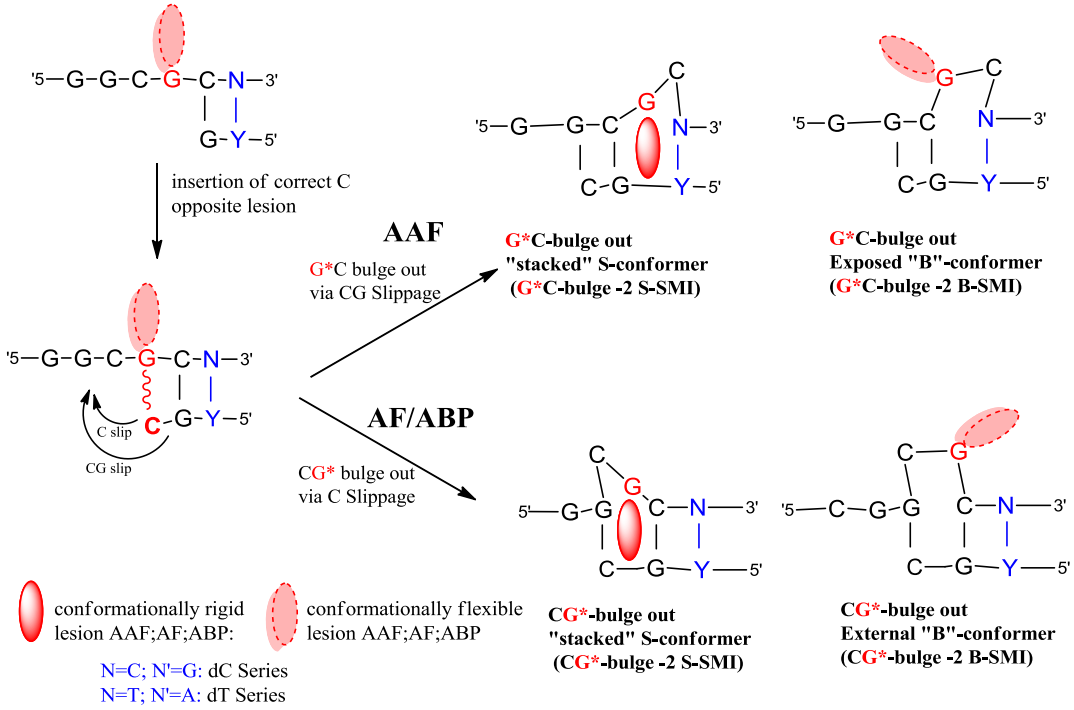

'5—G–G–C–G–C–N—3'
G–Y—5'
insertion of correct C
opposite lesion
'5—G–G–C–G–C–N—3'
C slip
C
G–Y—5'
CG slip
AAF
G*C bulge out
via CG Slippage
G*C-bulge out
"stacked" S-conformer
(G*C-bulge -2 S-SMI)
G*C-bulge out
Exposed "B"-conformer
(G*C-bulge -2 B-SMI)
AF/ABP
CG* bulge out
via C Slippage
CG*-bulge out
"stacked" S-conformer
(CG*-bulge -2 S-SMI)
CG*-bulge out
External "B"-conformer
(CG*-bulge -2 B-SMI)
conformationally rigid
lesion AAF;AF;ABP:
conformationally flexible
lesion AAF;AF;ABP
N=C; N'=G: dC Series
N=T; N'=A: dT Series

Figure 5

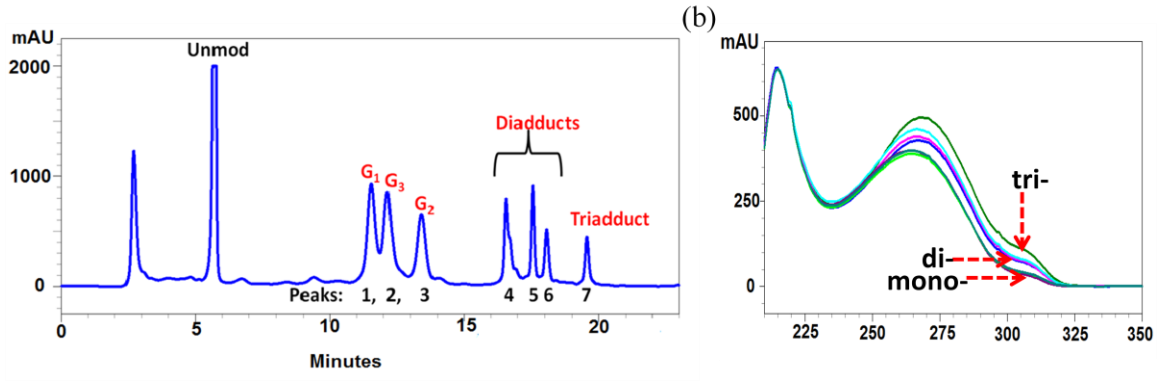

(b)
mAU
2000
1000
0
Unmod
G1 G3
G2
Diadducts
Triadduct
Peaks: 1, 2, 3 4 5 6 7
0 5 10 15 20
Minutes
mAU
500
250
0
tri-
di-
mono-
225 250 275 300 325 350

Figure 6

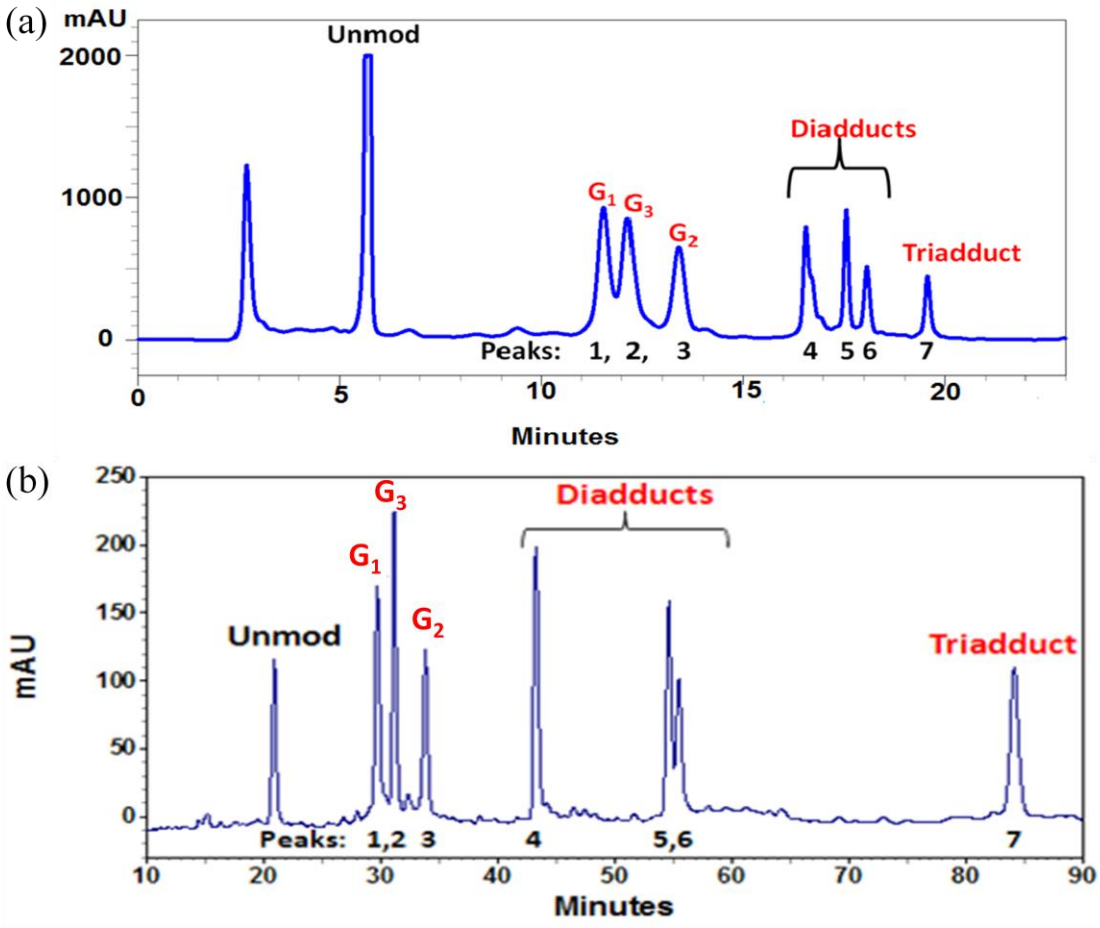

(a)
mAU
2000
1000
0
Unmod
Diadducts
G1
G3
G2
Triadduct
Peaks: 1, 2, 3
4 5 6 7
0
5
10
15
20
Minutes
(b)
250
200
150
100
50
0
mAU
G3
G1
G2
Unmod
Diadducts
Triadduct
Peaks: 1,2 3
4
5,6
7
10
20
30
40
50
60
70
80
90
Minutes

Figure 7

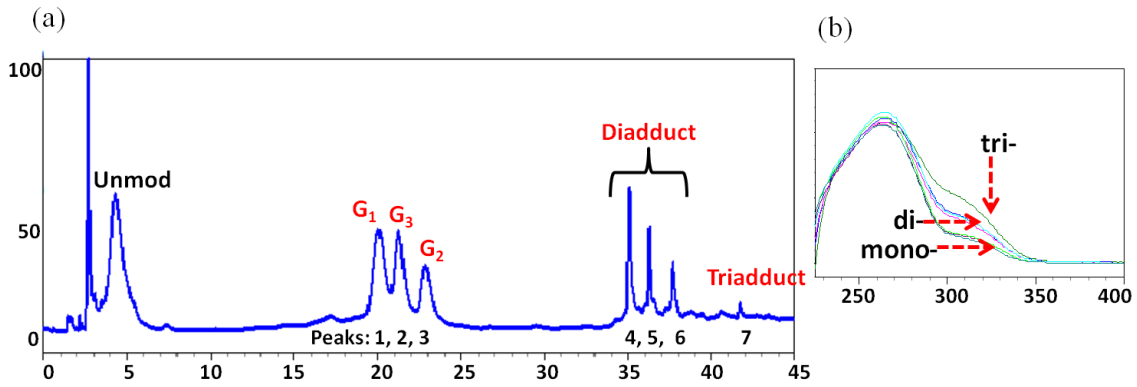

(a)
100
50
0
Unmod
G1
G3
G2
Diadduct
Triadduct
Peaks: 1, 2, 3
4, 5, 6
7
0
5
10
15
20
25
30
35
40
45
(b)
tri-
di-
mono-
250
300
350
400

Figure 8

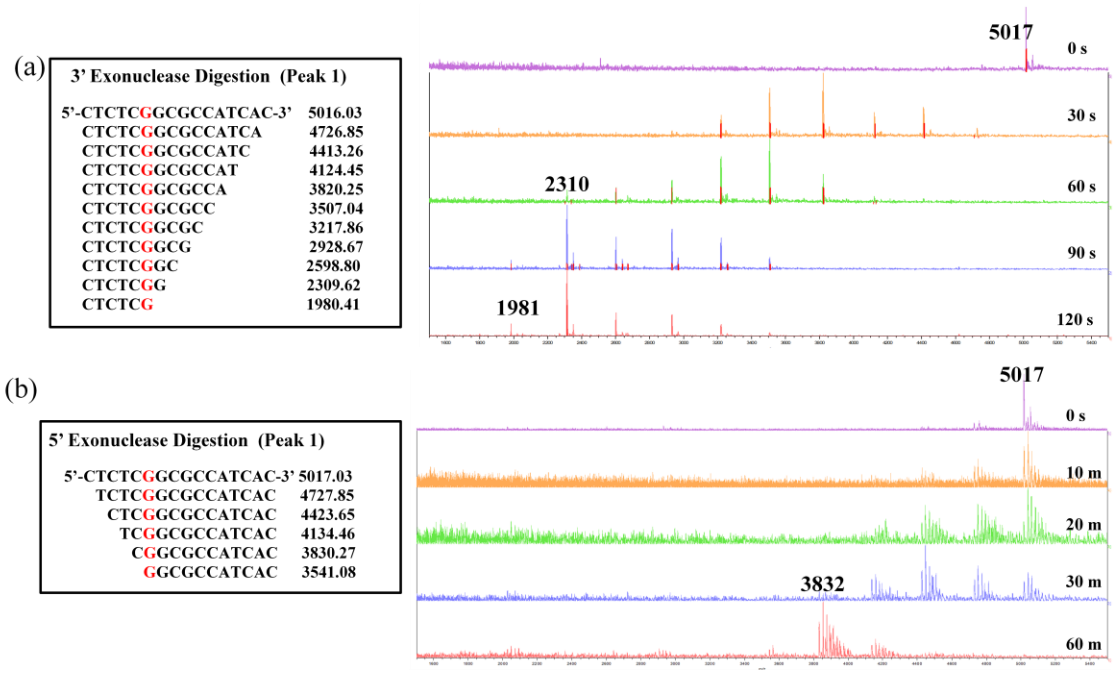
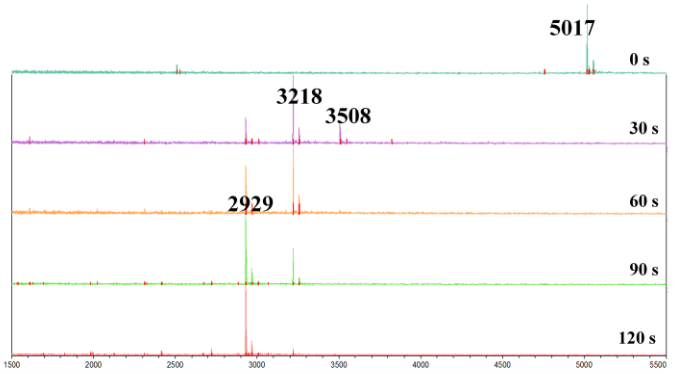

(a)
3' Exonuclease Digestion (Peak 1)
5'-CTCTCGGCGCCATCAC-3' 5016.03
CTCTCGGCGCCATCA 4726.85
CTCTCGGCGCCATC 4413.26
CTCTCGGCGCCAT 4124.45
CTCTCGGCGCCA 3820.25
CTCTCGGCGCC 3507.04
CTCTCGGCGC 3217.86
CTCTCGGCG 2928.67
CTCTCGGC 2598.80
CTCTCGG 2309.62
CTCTCG 1980.41
5017
0 s
30 s
2310
60 s
90 s
1981
120 s
(b)
5' Exonuclease Digestion (Peak 1)
5'-CTCTCGGCGCCATCAC-3' 5017.03
TCTCGGCGCCATCAC 4727.85
CTCGGCGCCATCAC 4423.65
TCGGCGCCATCAC 4134.46
CGGCGCCATCAC 3830.27
GGCGCCATCAC 3541.08
5017
0 s
10 m
20 m
3832
30 m
60 m

Figure 9

(a)

3' Exonuclease Digestion (Peak 2)	
5'-CTCTCGGCGCCATCAC-3'	5016.03
CTCTCGGCGCCATCA	4726.85
CTCTCGGCGCCATC	4413.26
CTCTCGGCGCCAT	4124.45
CTCTCGGCGCCA	3820.25
CTCTCGGCGCC	3507.04
CTCTCGGCGC	3217.86
CTCTCGGCG	2928.67



(b)

5' Exonuclease Digestion (Peak 2)	
5'-CTCTCGGCGCCATCAC-3'	5017.03
TCTCGGCGCCATCAC	4727.85
CTCGGCGCCATCAC	4423.65
TCGGCGCCATCAC	4134.46
CGGCGCCATCAC	3830.27
GGCGCCATCAC	3541.08
GCGCCATCAC	3211.85
CGCCATCAC	2882.66
GCCATCAC	2593.48

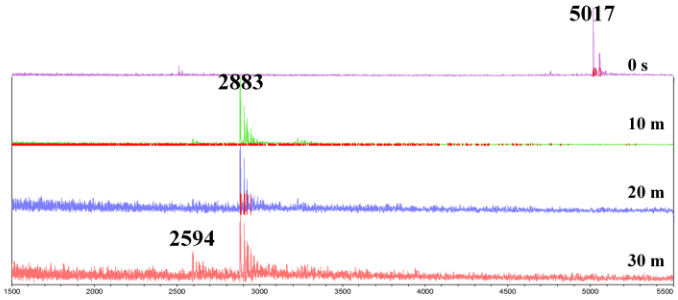


Figure 10

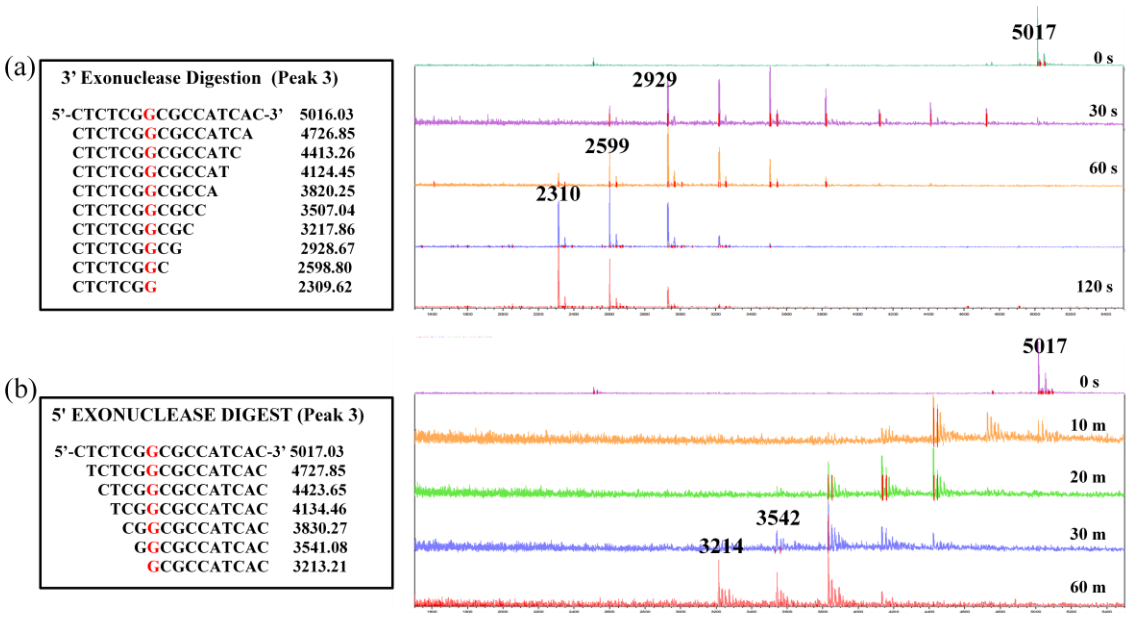
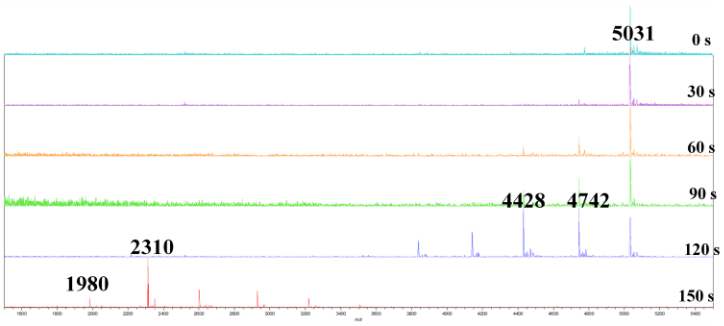

(a)
3' Exonuclease Digestion (Peak 3)
5'-CTCTCGGCGCCATCAC-3' 5016.03
CTCTCGGCGCCATCA 4726.85
CTCTCGGCGCCATC 4413.26
CTCTCGGCGCCAT 4124.45
CTCTCGGCGCCA 3820.25
CTCTCGGCGCC 3507.04
CTCTCGGCGC 3217.86
CTCTCGGCG 2928.67
CTCTCGGC 2598.80
CTCTCGG 2309.62
5017
0 s
2929
30 s
2599
60 s
2310
120 s
(b)
5' EXONUCLEASE DIGEST (Peak 3)
5'-CTCTCGGCGCCATCAC-3' 5017.03
TCTCGGCGCCATCAC 4727.85
CTCGGCGCCATCAC 4423.65
TCGGCGCCATCAC 4134.46
CGGCGCCATCAC 3830.27
GGCGCCATCAC 3541.08
GCGCCATCAC 3213.21
5017
0 s
10 m
20 m
3542
3214
30 m
60 m

Figure 11

(a)

3' EXONUCLEASE DIGEST (Peak 1)	
5'-CTCTCGGCGCTATCAC-3'	5030.37
5'-CTCTCGGCGCTATCA	4741.19
5'-CTCTCGGCGCTATC	4427.98
5'-CTCTCGGCGCTAT	4138.80
5'-CTCTCGGCGCTA	3834.60
5'-CTCTCGGCGCT	3521.39
5'-CTCTCGGCGC	3217.20
5'-CTCTCGGCG	2928.01
5'-CTCTCGGC	2598.80
5'-CTCTCGG	2309.62
5'-CTCTCG	1980.41



(b)

5' Exonuclease Digestion (Peak 1)	
5'-CTCTCGGCGCTATCAC-3'	5030.03
TCTCGGCGCTATCAC	4739.85
CTCGGCGCTATCAC	4435.65
TCGGCGCTATCAC	4146.46
CGGCGCTATCAC	3842.27
GGCGCTATCAC	3553.08

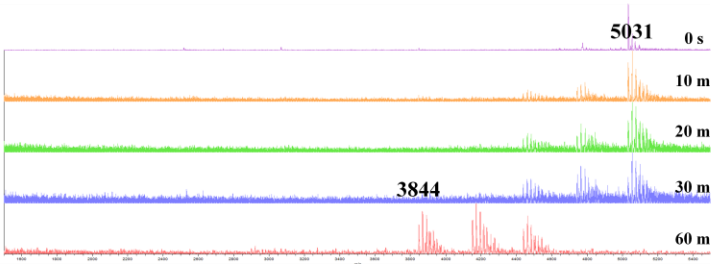
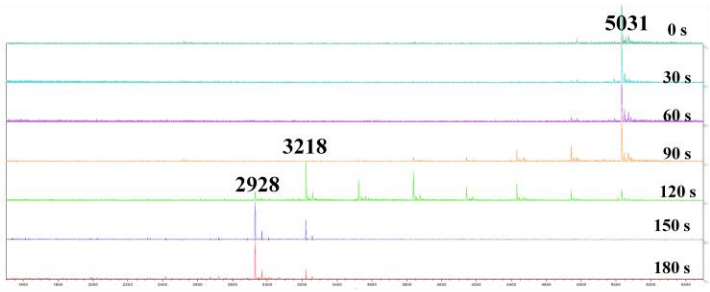


Figure 12

(a)

3' Exonuclease Digestion (Peak 2)	
5'-CTCTCGGCGCTATCAC-3'	5030.03
CTCTCGGCGCTATCA	4740.85
CTCTCGGCGCTATC	4427.26
CTCTCGGCGCTAT	4138.45
CTCTCGGCGCTA	3834.25
CTCTCGGCGCT	3521.04
CTCTCGGCGC	3217.86
CTCTCGGCG	2928.67



(b)

5' Exonuclease Digestion (Peak2)	
5'-CTCTCGGCGCTATCAC-3'	5030.03
TCTCGGCGCTATCAC	4740.85
CTCGGCGCTATCAC	4436.65
TCGGCGCTATCAC	4147.46
CGGCGCTATCAC	3843.27
GGCGCTATCAC	3554.08
GCGCTATCAC	3224.85
CGCTATCAC	2895.66
GCTATCAC	2606.48

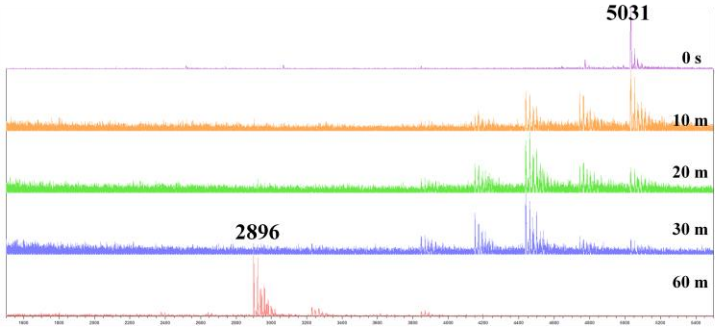


Figure 13

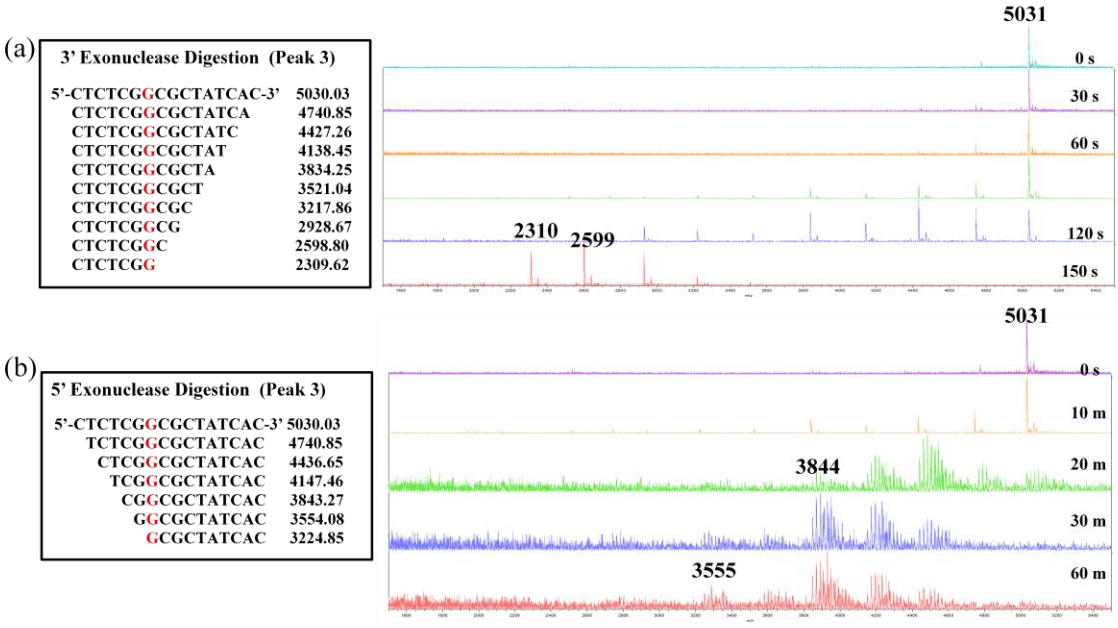

(a)
3' Exonuclease Digestion (Peak 3)
5'-CTCTCGGCGCTATCAC-3' 5030.03
CTCTCGGCGCTATCA 4740.85
CTCTCGGCGCTATC 4427.26
CTCTCGGCGCTAT 4138.45
CTCTCGGCGCTA 3834.25
CTCTCGGCGCT 3521.04
CTCTCGGCGC 3217.86
CTCTCGGCG 2928.67
CTCTCGGC 2598.80
CTCTCGG 2309.62
5031
0 s
30 s
60 s
2310
2599
120 s
150 s
(b)
5' Exonuclease Digestion (Peak 3)
5'-CTCTCGGCGCTATCAC-3' 5030.03
TCTCGGCGCTATCAC 4740.85
CTCGGCGCTATCAC 4436.65
TCGGCGCTATCAC 4147.46
CGGCGCTATCAC 3843.27
GGCGCTATCAC 3554.08
GCGCTATCAC 3224.85
5031
0 s
10 m
3844
20 m
30 m
3555
60 m

Figure 14

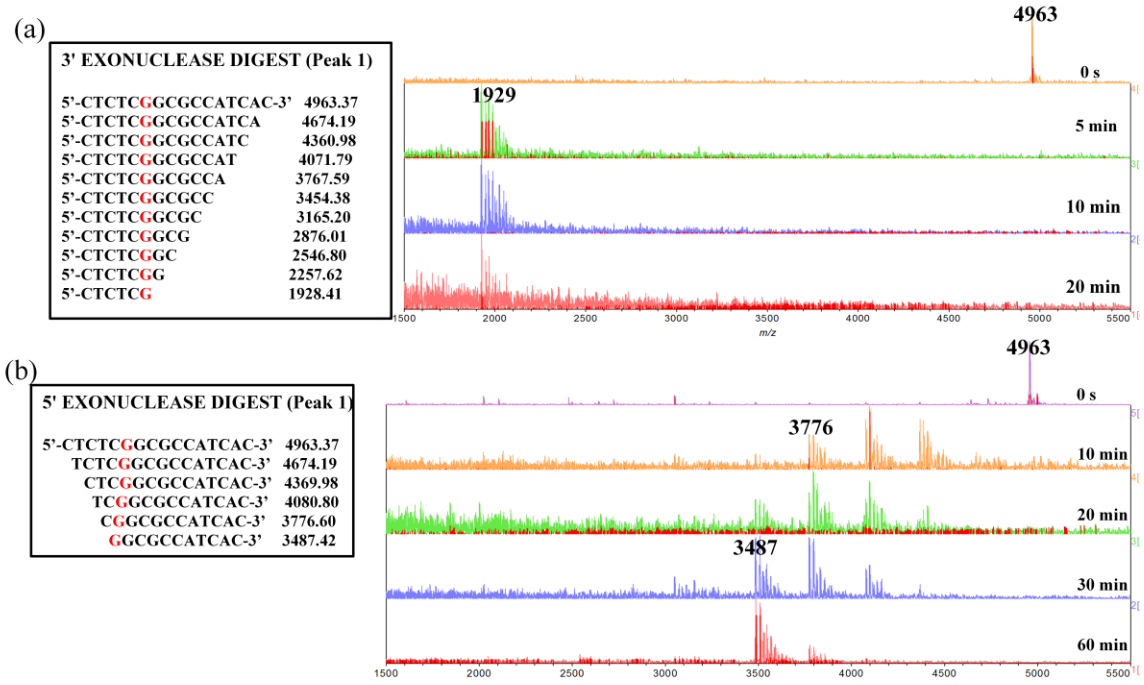
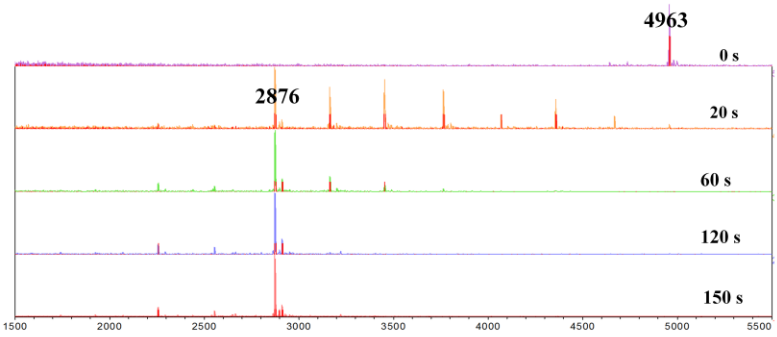

(a)
3' EXONUCLEASE DIGEST (Peak 1)
5'-CTCTCGGCGCCATCAC-3' 4963.37
5'-CTCTCGGCGCCATCA 4674.19
5'-CTCTCGGCGCCATC 4360.98
5'-CTCTCGGCGCCAT 4071.79
5'-CTCTCGGCGCCA 3767.59
5'-CTCTCGGCGCC 3454.38
5'-CTCTCGGCGC 3165.20
5'-CTCTCGGCG 2876.01
5'-CTCTCGGC 2546.80
5'-CTCTCGG 2257.62
5'-CTCTCG 1928.41
4963
1929
0 s
5 min
10 min
20 min
m/z
(b)
5' EXONUCLEASE DIGEST (Peak 1)
5'-CTCTCGGCGCCATCAC-3' 4963.37
TCTCGGCGCCATCAC-3' 4674.19
CTCGGCGCCATCAC-3' 4369.98
TCGGCGCCATCAC-3' 4080.80
CGGCGCCATCAC-3' 3776.60
GGCGCCATCAC-3' 3487.42
4963
3776
3487
0 s
10 min
20 min
30 min
60 min

Figure 15

(a)

3' EXONUCLEASE DIGEST (Peak 2)

5'-CTCTCGGCGCCATCAC-3'	4963.37
5'-CTCTCGGCGCCATCA	4674.19
5'-CTCTCGGCGCCATC	4360.98
5'-CTCTCGGCGCCAT	4071.79
5'-CTCTCGGCGCCA	3767.59
5'-CTCTCGGCGCC	3454.38
5'-CTCTCGGCGC	3165.20
5'-CTCTCGGCG	2876.01



5' EXONUCLEASE DIGEST (Peak 2)

5'-CTCTCGGCGCCATCAC-3'	4963.37
TCTCGGCGCCATCAC-3'	4674.19
CTCGGCGCCATCAC-3'	4369.98
TCGGCGCCATCAC-3'	4080.80
CGGCGCCATCAC-3'	3776.60
GGCGCCATCAC-3'	3487.42
GCGCCATCAC-3'	3158.21
CGCCATCAC-3'	2829.00
GCCATCAC-3'	2539.82

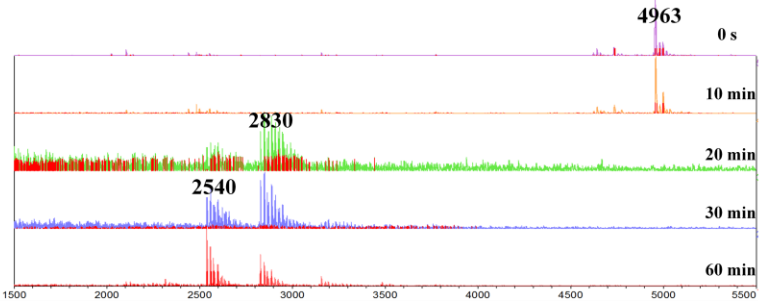
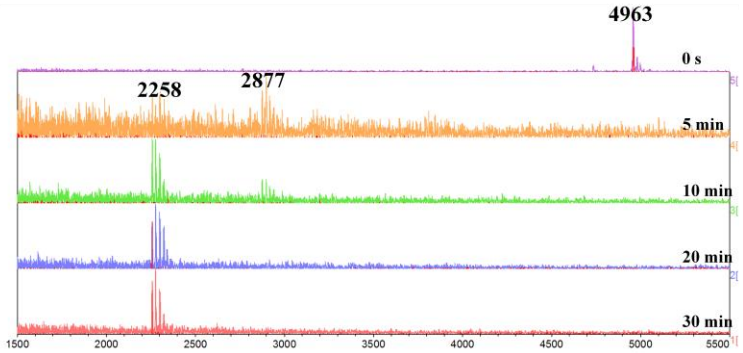


Figure 16

(a)

3' EXONUCLEASE DIGEST (Peak 3)	
5'-CTCTCGGCGCCATCAC-3'	4963.37
5'-CTCTCGGCGCCATCA	4674.19
5'-CTCTCGGCGCCATC	4360.98
5'-CTCTCGGCGCCAT	4071.79
5'-CTCTCGGCGCCA	3767.59
5'-CTCTCGGCGCC	3454.38
5'-CTCTCGGCGC	3165.20
5'-CTCTCGGCG	2876.01
5'-CTCTCGGC	2546.80
5'-CTCTCGG	2257.62



(b)

5' EXONUCLEASE DIGEST (Peak 3)	
5'-CTCTCGGCGCCATCAC-3'	4963.37
TCTCGGCGCCATCAC-3'	4674.19
CTCGGCGCCATCAC-3'	4369.98
TCGGCGCCATCAC-3'	4080.80
CGGCGCCATCAC-3'	3776.60
GGCGCCATCAC-3'	3487.42
GCGCCATCAC-3'	3158.21

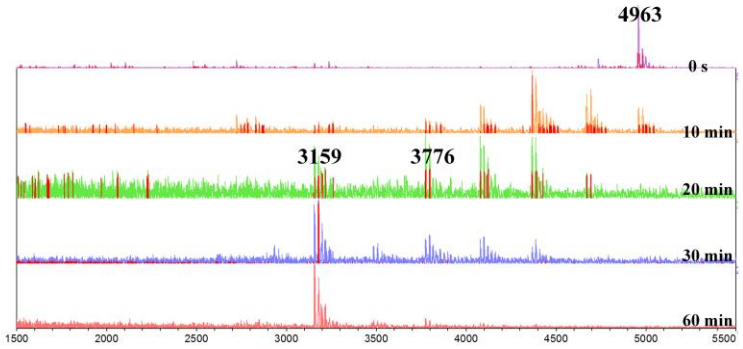
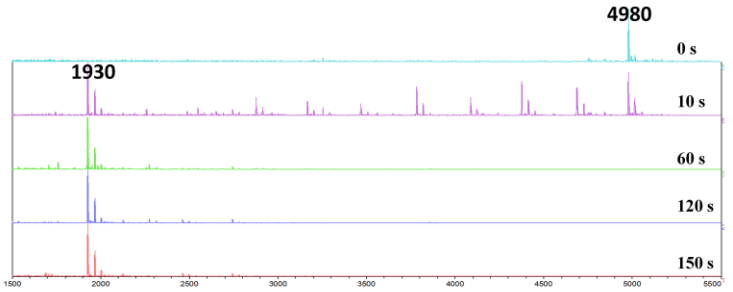


Figure 17

(a)

3' EXONUCLEASE DIGEST (Peak 1)	
5'-CTCTCGGCGCTATCAC-3'	4978.37
5'-CTCTCGGCGCTATCA	4689.19
5'-CTCTCGGCGCTATC	4375.98
5'-CTCTCGGCGCTAT	4086.80
5'-CTCTCGGCGCTA	3782.60
5'-CTCTCGGCGCT	3469.39
5'-CTCTCGGCGC	3165.20
5'-CTCTCGGCG	2876.01
5'-CTCTCGGC	2546.80
5'-CTCTCGG	2257.62
5'-CTCTCG	1928.41



(b)

5' Exonuclease Digestion (Peak 1)	
5'-CTCTCGGCGCTATCAC-3'	4978.03
TCTCGGCGCTATCAC	4687.85
CTCGGCGCTATCAC	4383.65
TCGGCGCTATCAC	4094.46
CGGCGCTATCAC	3790.27
GGCGCTATCAC	3501.08

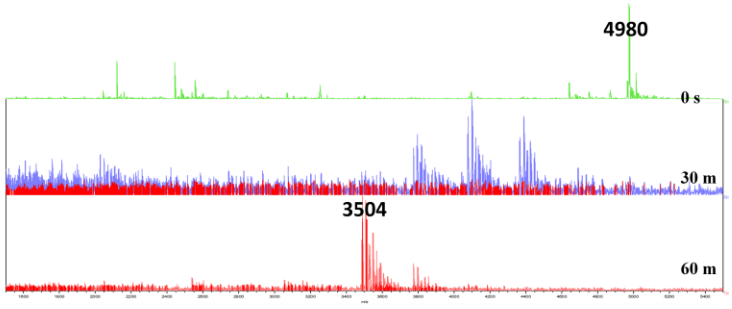


Figure 18

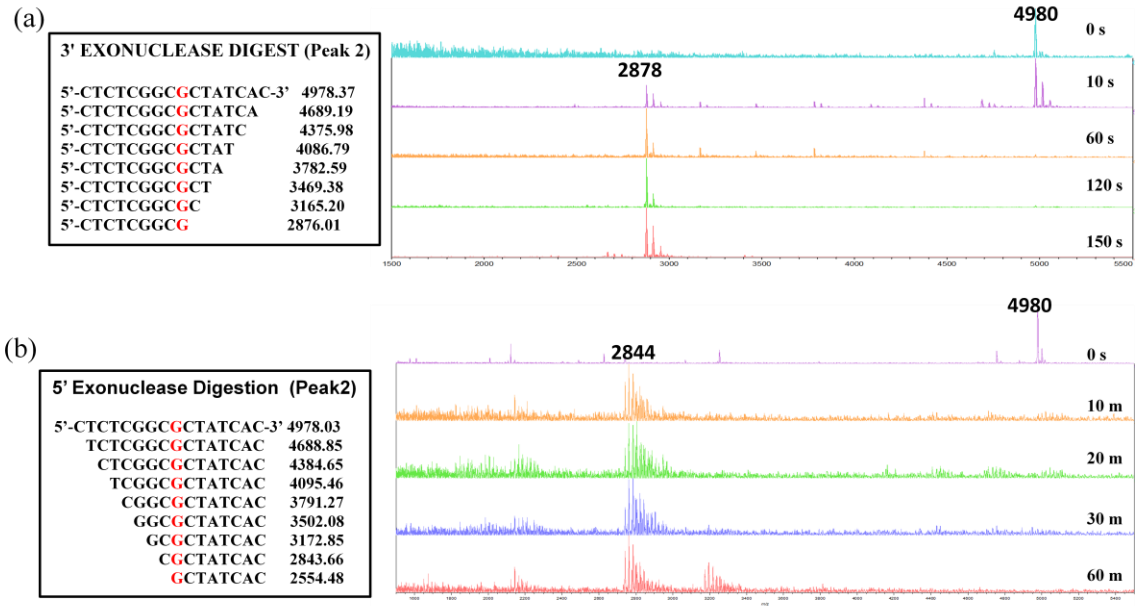
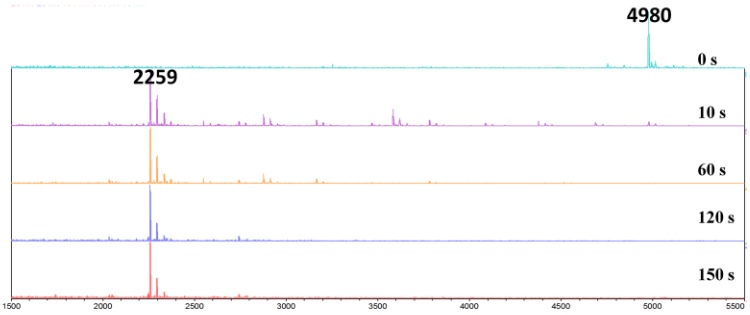

(a)
3' EXONUCLEASE DIGEST (Peak 2)
5'-CTCTCGGCGCTATCAC-3' 4978.37
5'-CTCTCGGCGCTATCA 4689.19
5'-CTCTCGGCGCTATC 4375.98
5'-CTCTCGGCGCTAT 4086.79
5'-CTCTCGGCGCTA 3782.59
5'-CTCTCGGCGCT 3469.38
5'-CTCTCGGCGC 3165.20
5'-CTCTCGGCG 2876.01
4980
2878
0 s
10 s
60 s
120 s
150 s
1500
2000
2500
3000
3500
4000
4500
5000
5500
(b)
5' Exonuclease Digestion (Peak2)
5'-CTCTCGGCGCTATCAC-3' 4978.03
TCTCGGCGCTATCAC 4688.85
CTCGGCGCTATCAC 4384.65
TCGGCGCTATCAC 4095.46
CGGCGCTATCAC 3791.27
GGCGCTATCAC 3502.08
GCGCTATCAC 3172.85
CGCTATCAC 2843.66
GCTATCAC 2554.48
4980
2844
0 s
10 m
20 m
30 m
60 m

Figure 19

(a)

3' EXONUCLEASE DIGEST (Peak 3)	
5'-CTCTCGGCGCTATCAC-3'	4978.37
5'-CTCTCGGCGCTATCA	4689.19
5'-CTCTCGGCGCTATC	4375.98
5'-CTCTCGGCGCTAT	4086.79
5'-CTCTCGGCGCTA	3782.59
5'-CTCTCGGCGCT	3469.38
5'-CTCTCGGCGC	3165.20
5'-CTCTCGGCG	2876.01
5'-CTCTCGGC	2546.80
5'-CTCTCGG	2257.62



(b)

5' Exonuclease Digestion (Peak 3)	
5'-CTCTCGGCGCTATCAC-3'	4978.03
TCTCGGCGCTATCAC	4688.85
CTCGGCGCTATCAC	4384.65
TCGGCGCTATCAC	4095.46
CGGCGCTATCAC	3791.27
GGCGCTATCAC	3502.08
GCGCTATCAC	3172.85

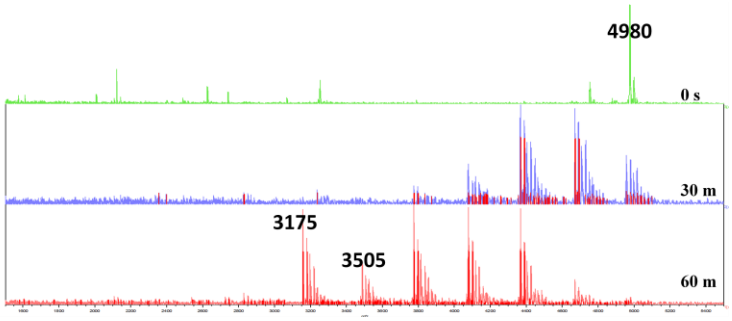


Figure 20

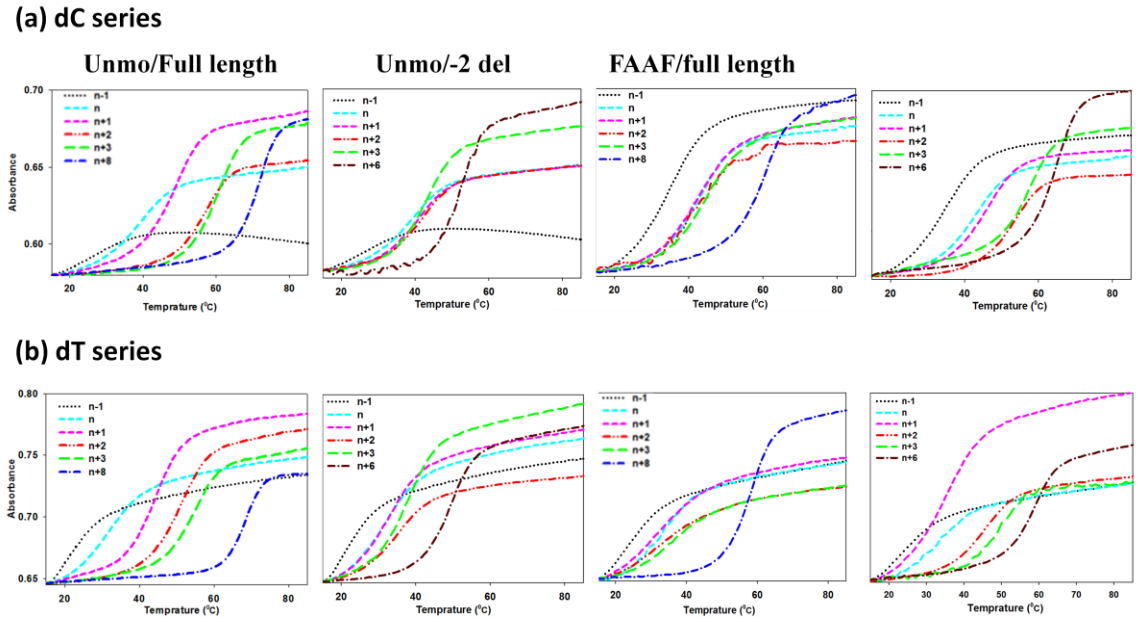

(a) dC series
Unmo/Full length
Unmo/-2 del
FAAF/full length
(b) dT series
Absorbance
Temprature (°c)
n-1
n
n+1
n+2
n+3
n+6
n+8

Figure 21

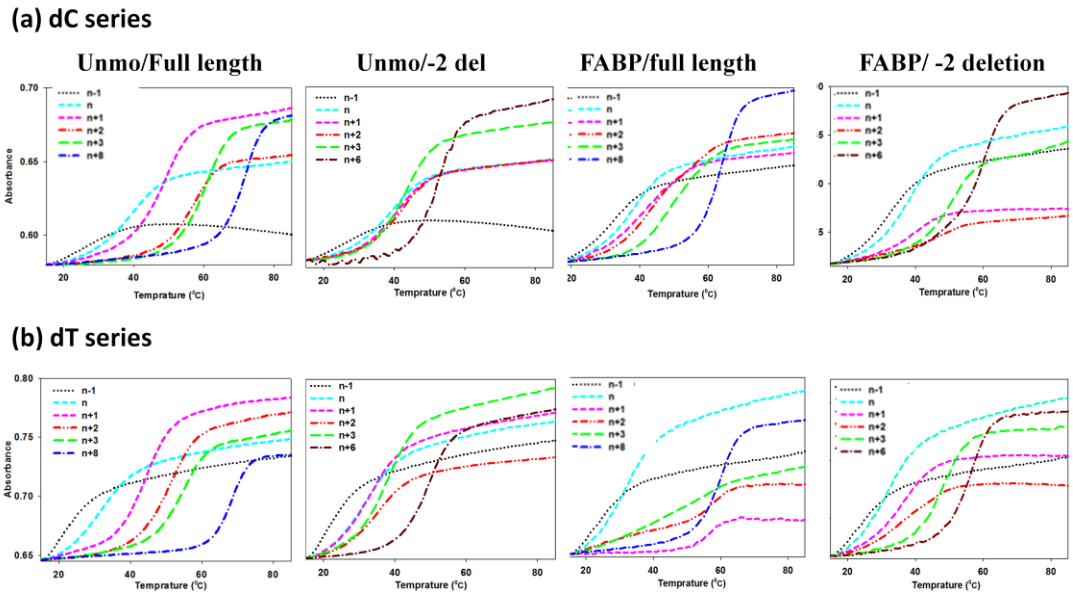

(a) dC series
Unmo/Full length
Unmo/-2 del
FABP/full length
FABP/ -2 deletion
(b) dT series
Absorbance
Temprature (°c)
n-1
n
n+1
n+2
n+3
n+6
n+8

Figure 22

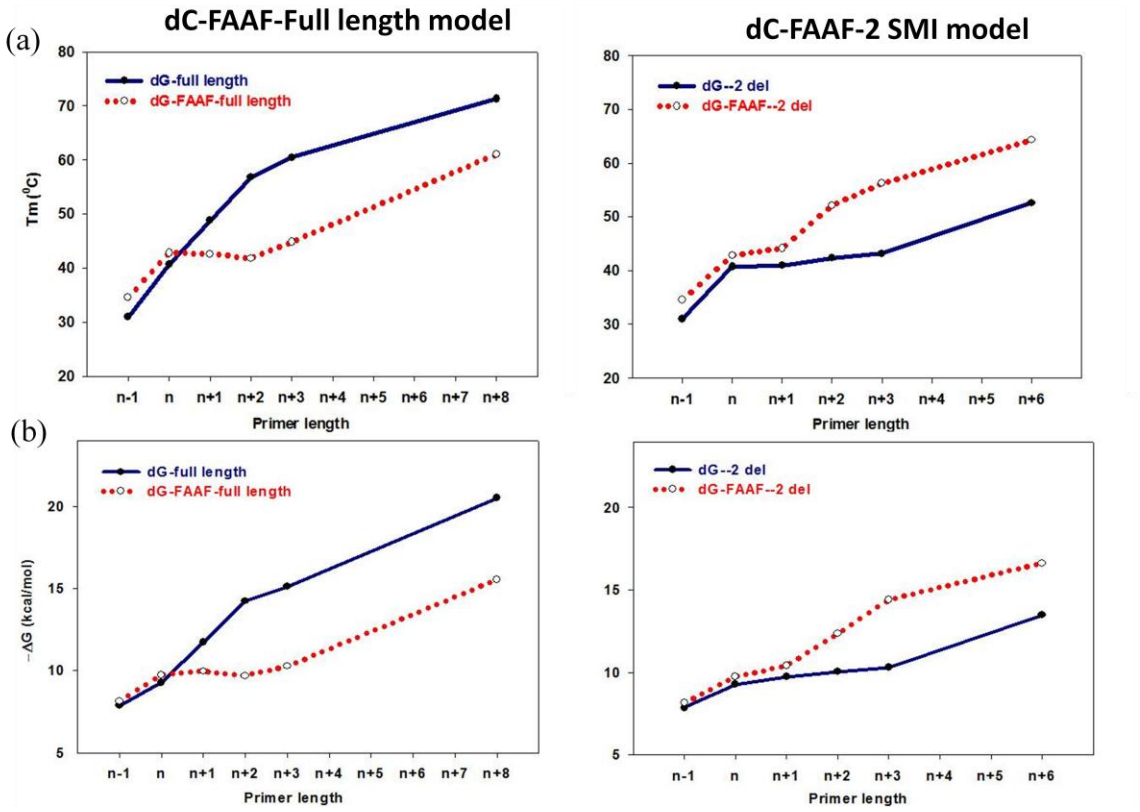

(a)
dC-FAAF-Full length model
dG-full length
dG-FAAF-full length
Tm (°C)
Primer length
dC-FAAF-2 SMI model
dG--2 del
dG-FAAF--2 del
(b)
−ΔG (kcal/mol)
n-1 n n+1 n+2 n+3 n+4 n+5 n+6 n+7 n+8

Figure 23

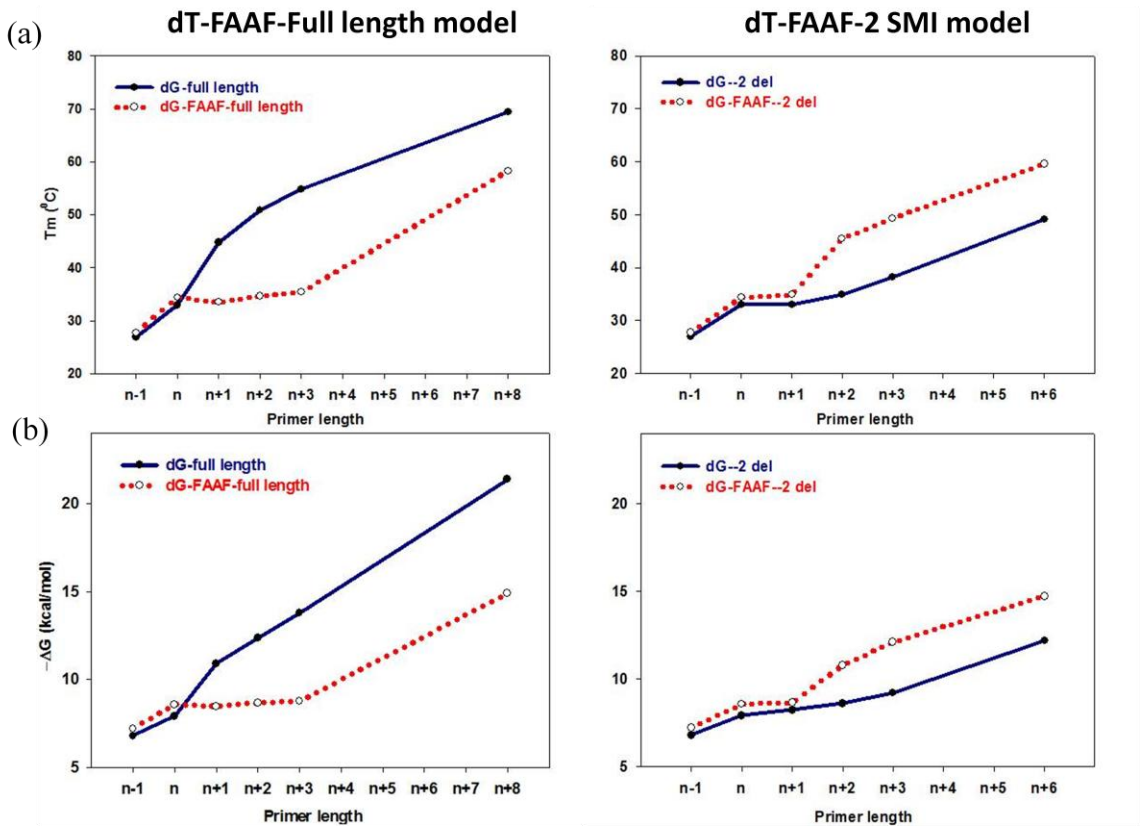

(a)
dT-FAAF-Full length model
dT-FAAF-2 SMI model
dG-full length
dG-FAAF-full length
dG--2 del
dG-FAAF--2 del
Tm (°C)
Primer length
(b)
−ΔG (kcal/mol)
Primer length

Figure 24

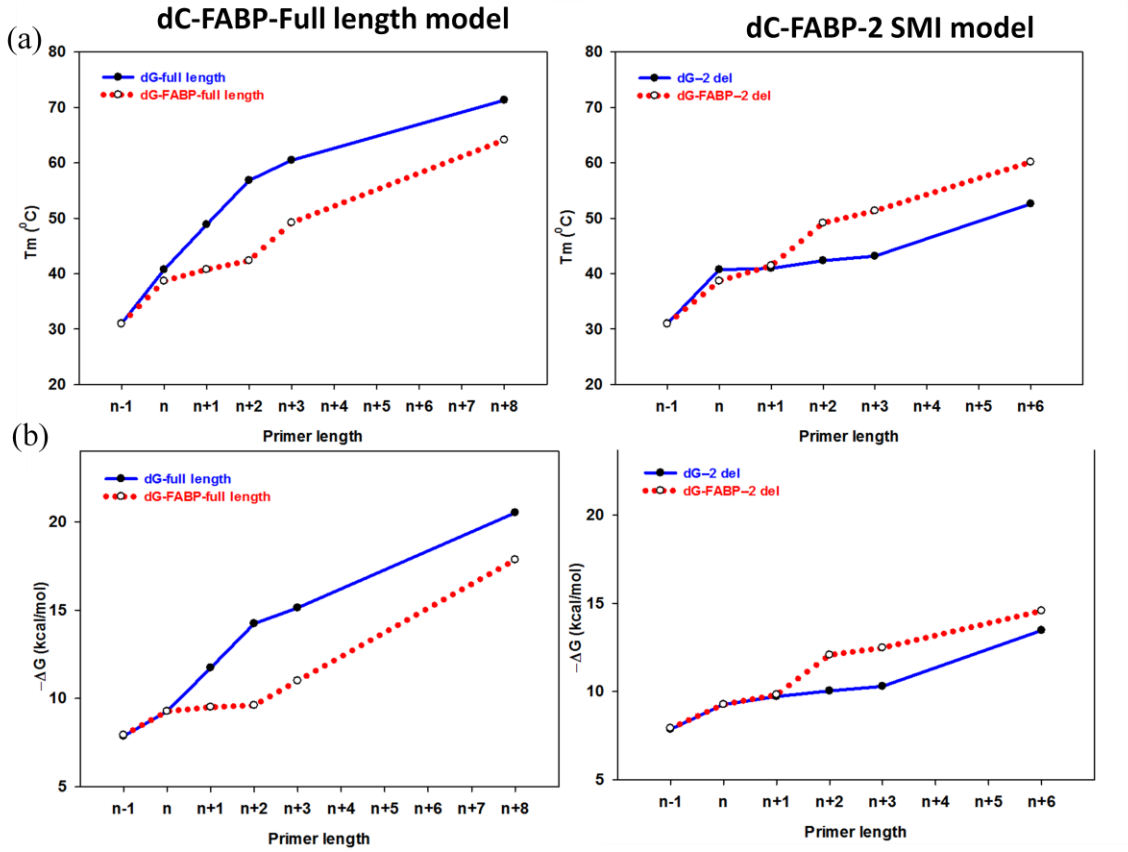

(a)
dC-FABP-Full length model
dG-full length
dG-FABP-full length
Tm (°C)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6 n+7 n+8
dC-FABP-2 SMI model
dG–2 del
dG-FABP–2 del
Tm (°C)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6
(b)
dG-full length
dG-FABP-full length
–ΔG (kcal/mol)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6 n+7 n+8
dG–2 del
dG-FABP–2 del
–ΔG (kcal/mol)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6

Figure 25

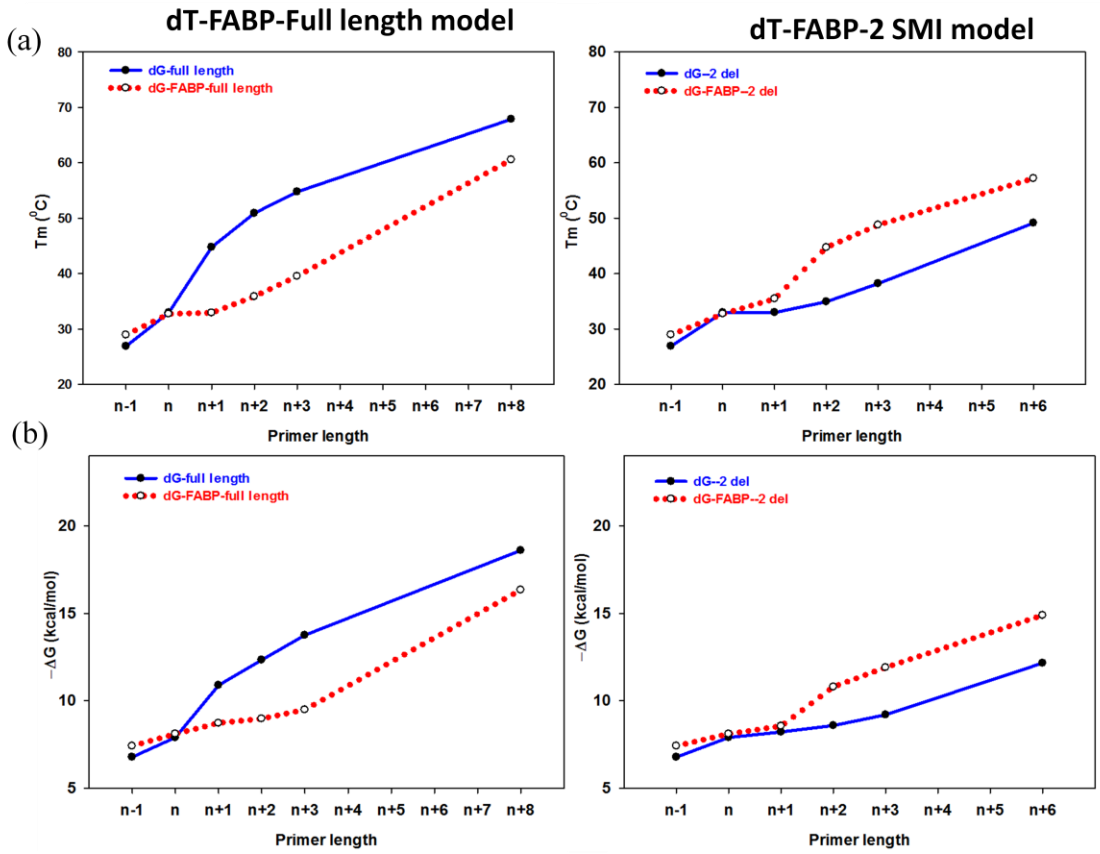

(a)
dT-FABP-Full length model
dG-full length
dG-FABP-full length
Tm (°C)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6 n+7 n+8
dT-FABP-2 SMI model
dG--2 del
dG-FABP--2 del
Tm (°C)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6
(b)
dG-full length
dG-FABP-full length
−ΔG (kcal/mol)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6 n+7 n+8
dG--2 del
dG-FABP--2 del
−ΔG (kcal/mol)
Primer length
n-1 n n+1 n+2 n+3 n+4 n+5 n+6

Figure 26

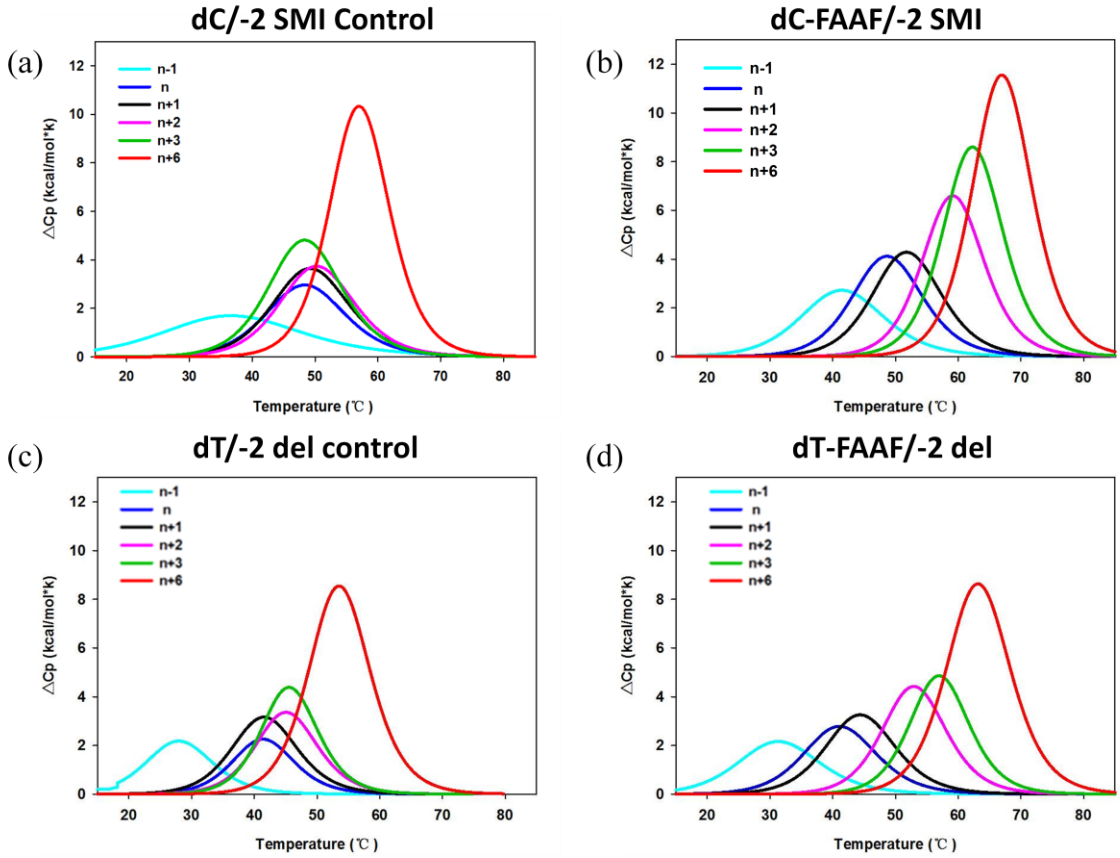

(a)
dC/-2 SMI Control
(b)
dC-FAAF/-2 SMI
(c)
dT/-2 del control
(d)
dT-FAAF/-2 del
n-1
n
n+1
n+2
n+3
n+6
△Cp (kcal/mol*K)
Temperature (℃)

Figure 27

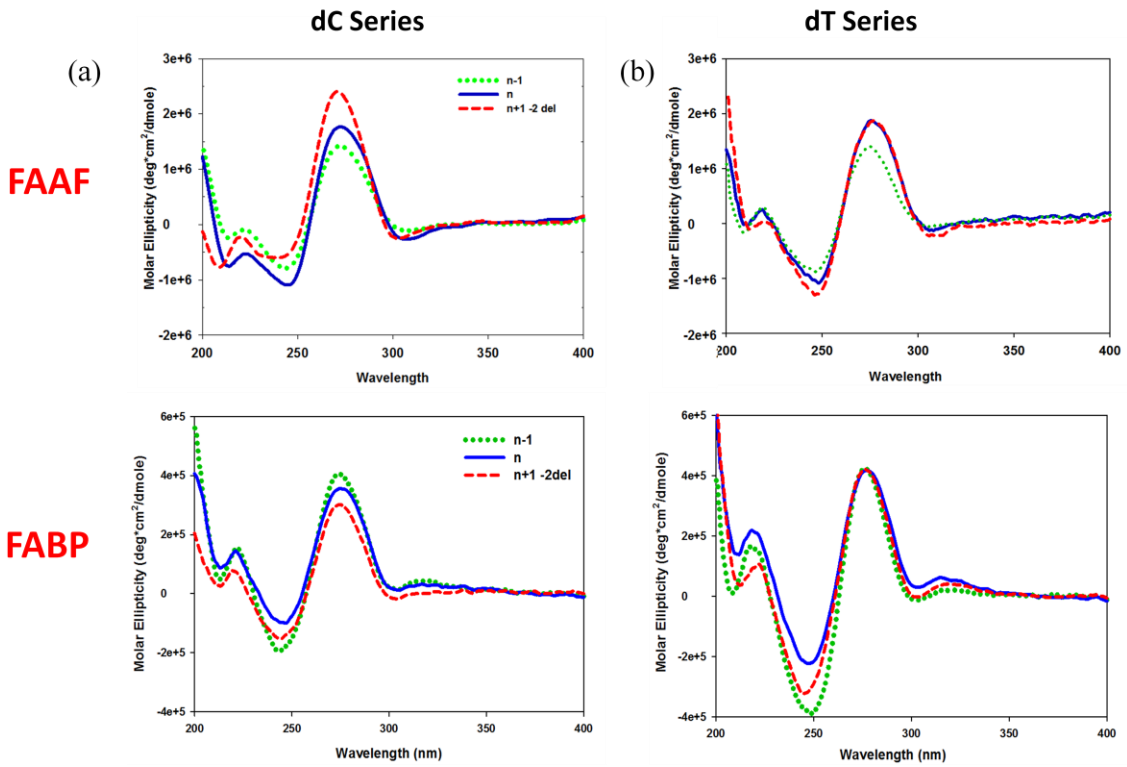

dC Series
dT Series
(a)
(b)
FAAF
FABP
n-1
n
n+1 -2 del
n+1 -2del
Molar Ellipticity (deg*cm²/dmole)
Wavelength
Wavelength (nm)

Figure 28

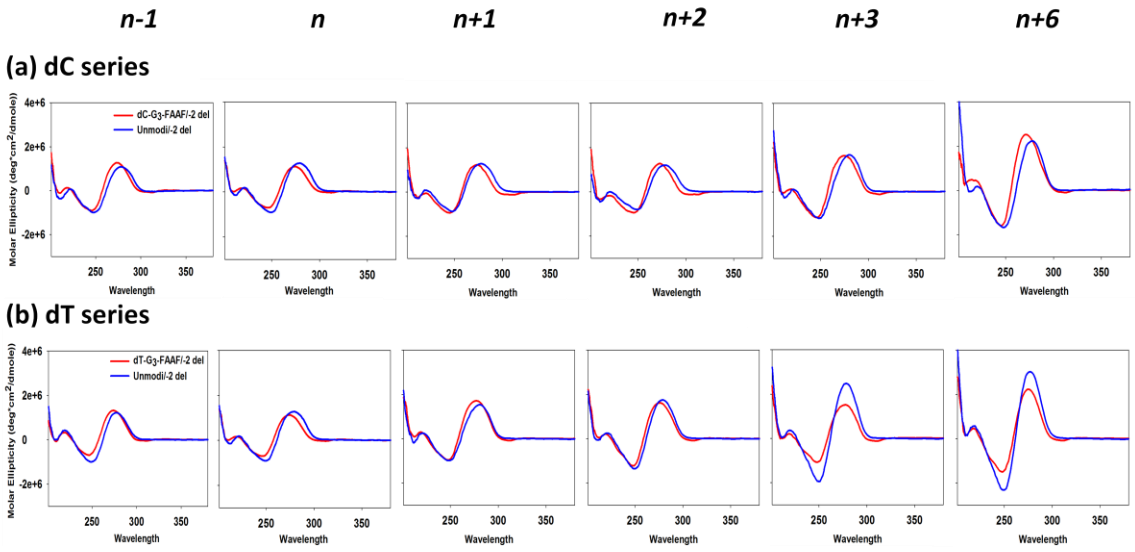

n-1
n
n+1
n+2
n+3
n+6
(a) dC series
dC-G3-FAAF/-2 del
Unmodi/-2 del
(b) dT series
dT-G3-FAAF/-2 del
Unmodi/-2 del
Molar Ellipticity (deg*cm2/dmole)
4e+6
2e+6
0
-2e+6
250
300
350
Wavelength

Figure 29

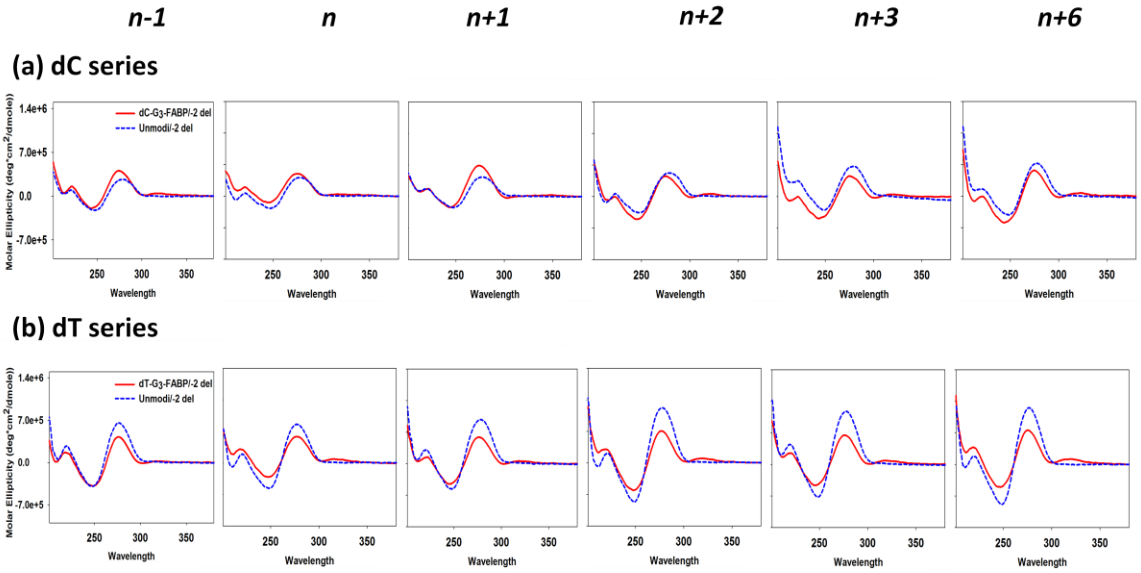

n-1
n
n+1
n+2
n+3
n+6
(a) dC series
dC-G3-FABP/-2 del
Unmodil/-2 del
Molar Ellipticity (deg*cm2/dmole)
1.4e+6
7.0e+5
0.0
-7.0e+5
250
300
350
Wavelength
(b) dT series
dT-G3-FABP/-2 del
Unmodil/-2 del

Figure 30

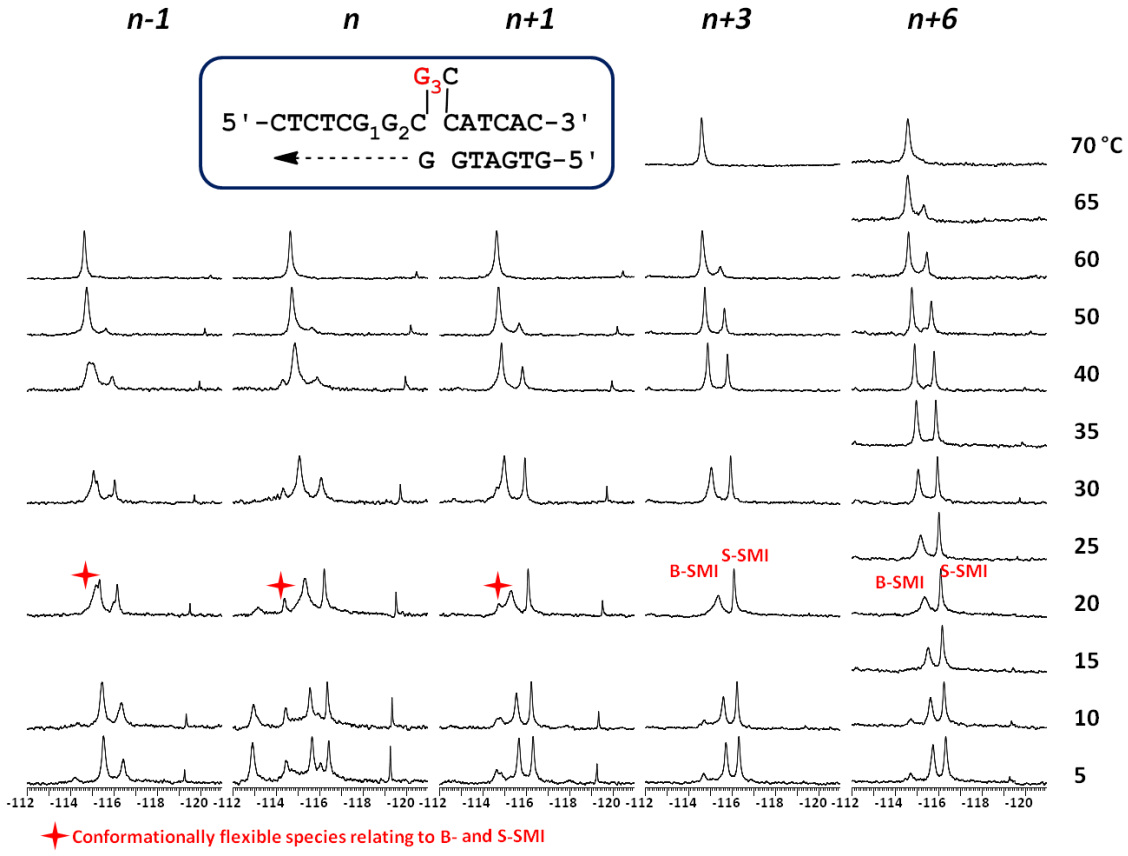

n-1
n
n+1
n+3
n+6
G3C
5'-CTCTCG1G2C CATCAC-3'
G GTAGTG-5'
70 °C
65
60
50
40
35
30
25
20
15
10
5
B-SMI
S-SMI
B-SMI
S-SMI
-112 -114 -116 -118 -120
Conformationally flexible species relating to B- and S-SMI

Figure 31

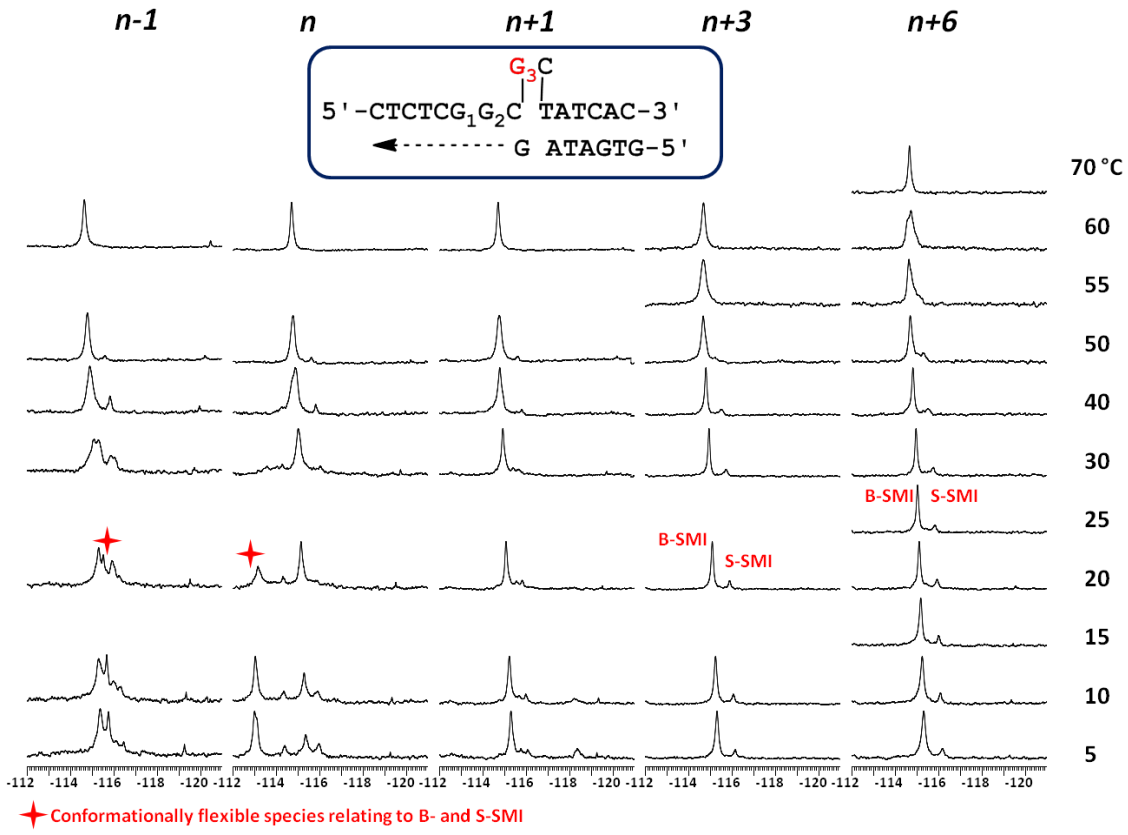

n-1
n
n+1
n+3
n+6
G3C
5'-CTCTCG1G2C TATCAC-3'
G ATAGTG-5'
70 °C
60
55
50
40
30
25
20
15
10
5
B-SMI
S-SMI
B-SMI
S-SMI
-112 -114 -116 -118 -120
Conformationally flexible species relating to B- and S-SMI

Figure 32

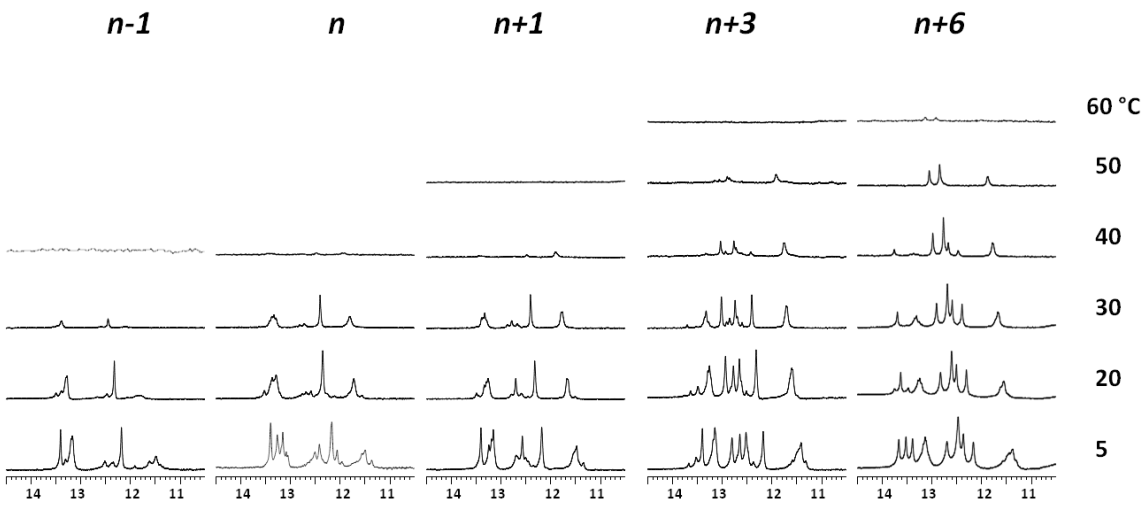


Figure 33

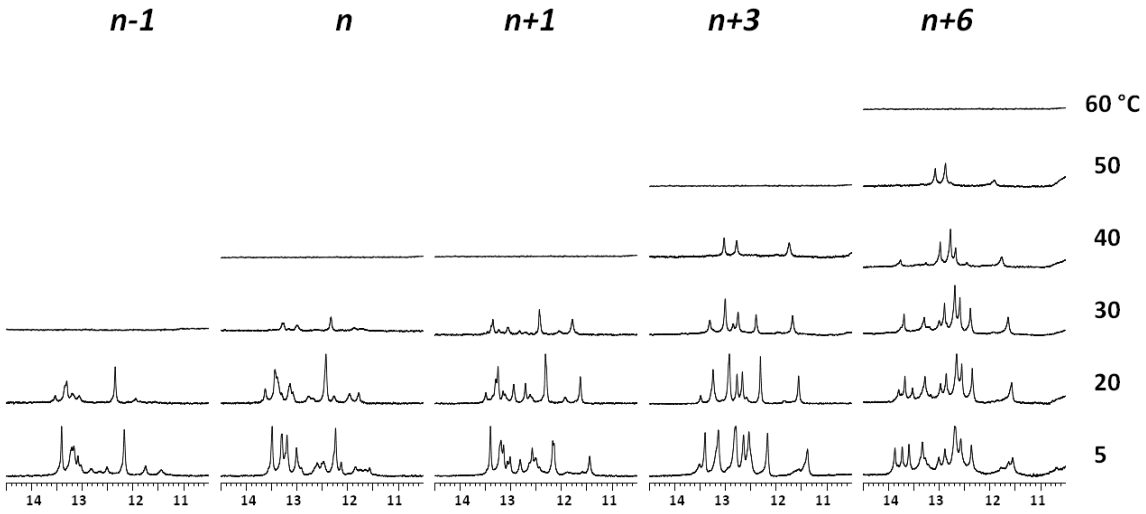


Figure 34

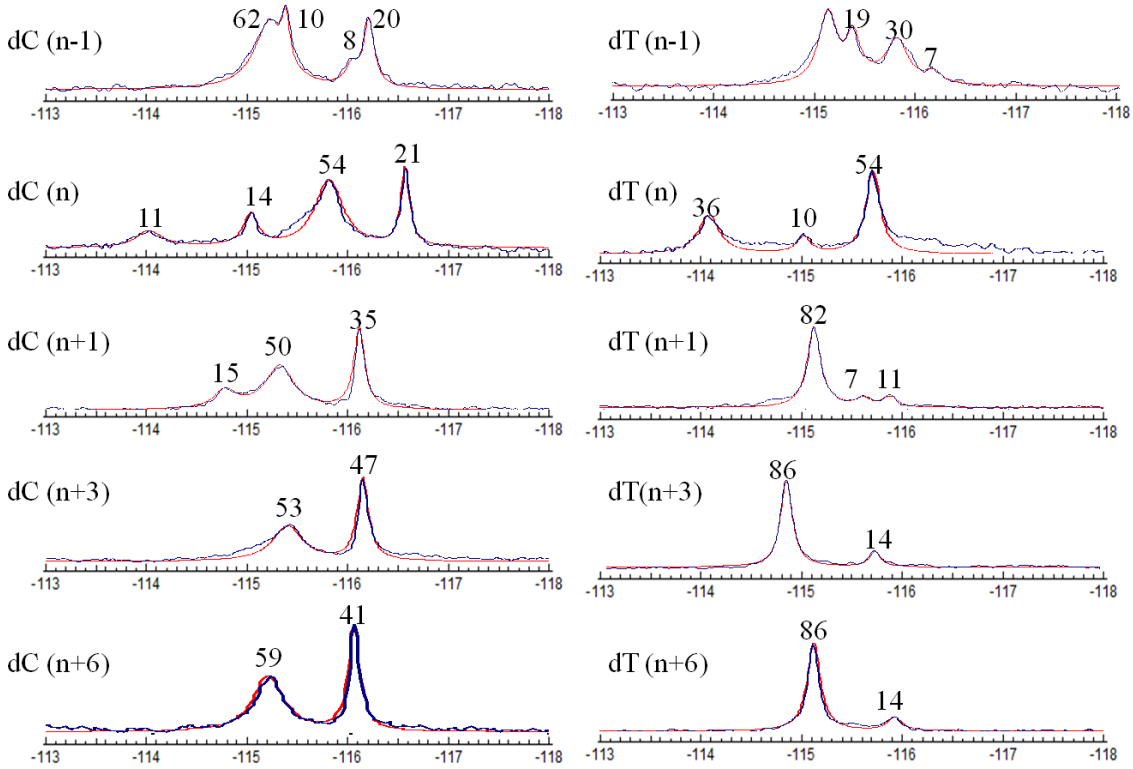

dC (n-1)
62
10
8
20
dT (n-1)
19
30
7
dC (n)
11
14
54
21
dT (n)
36
10
54
dC (n+1)
15
50
35
dT (n+1)
7
11
82
dC (n+3)
53
47
dT(n+3)
14
86
dC (n+6)
59
41
dT (n+6)
14
86
-113
-114
-115
-116
-117
-118

Figure 35

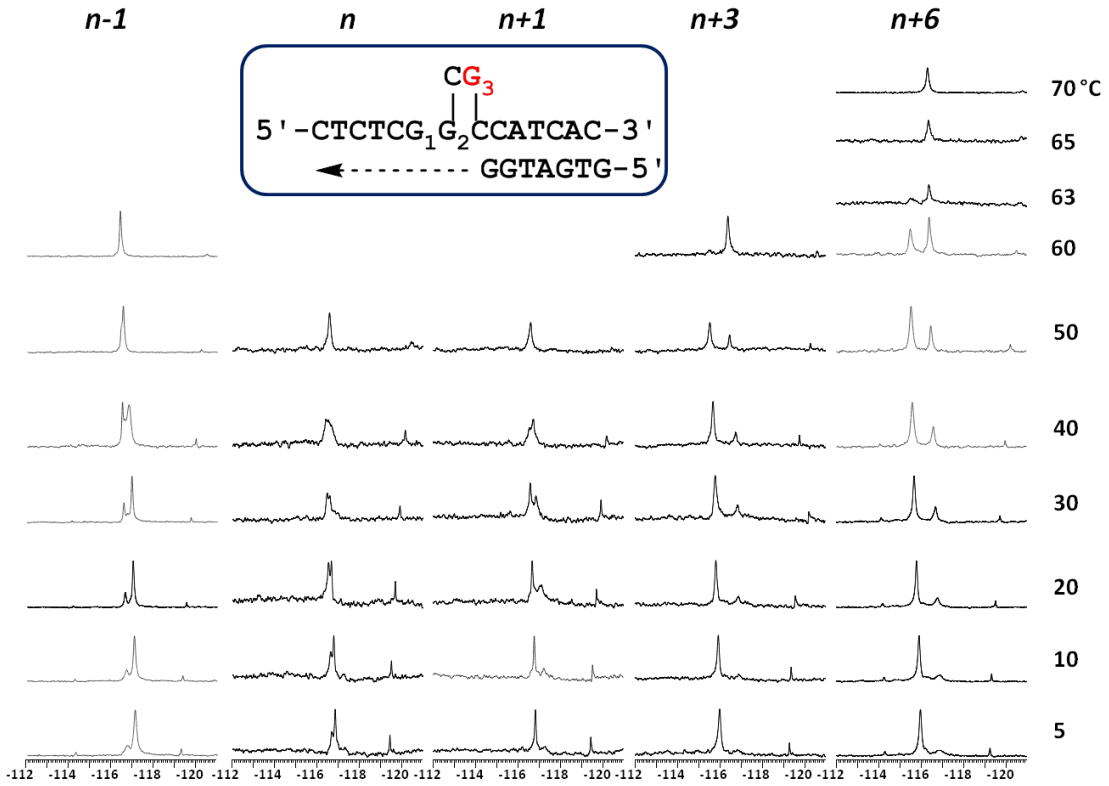

n-1
n
n+1
n+3
n+6
CG3
5'-CTCTCG1G2CCATCAC-3'
GGTAGTG-5'
70 °C
65
63
60
50
40
30
20
10
5
-112 -114 -116 -118 -120

Figure 36

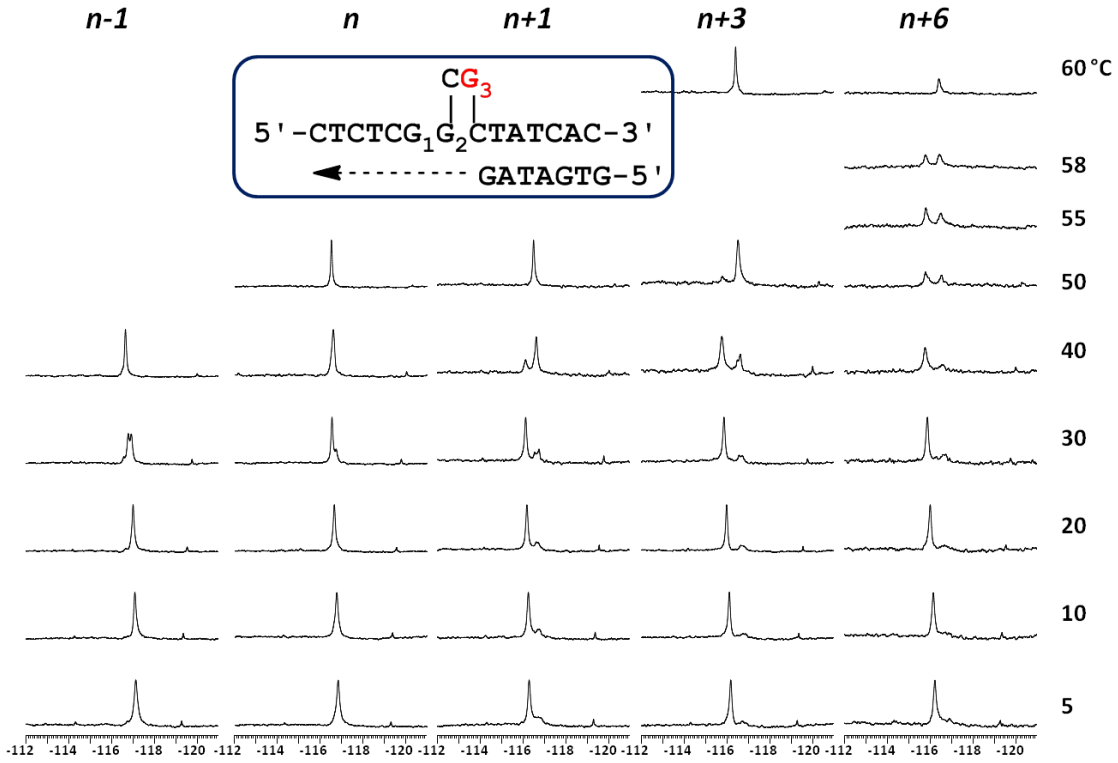

n-1
n
n+1
n+3
n+6
CG3
5'-CTCTCG1G2CTATCAC-3'
GATAGTG-5'
60 °C
58
55
50
40
30
20
10
5
-112 -114 -116 -118 -120

Figure 37

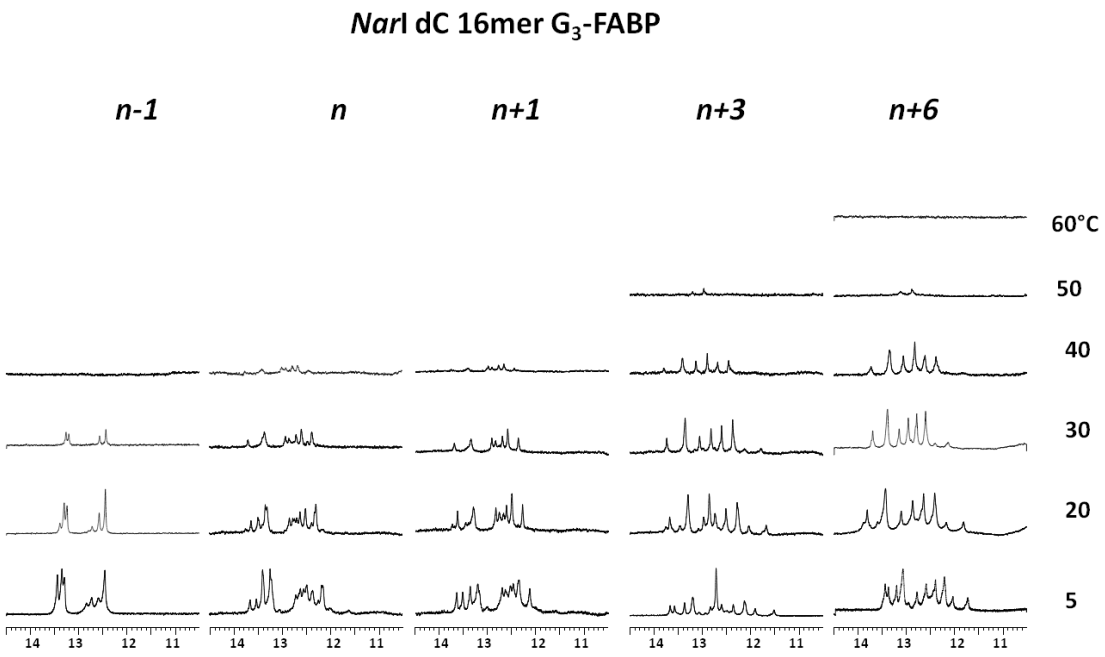


Figure 38

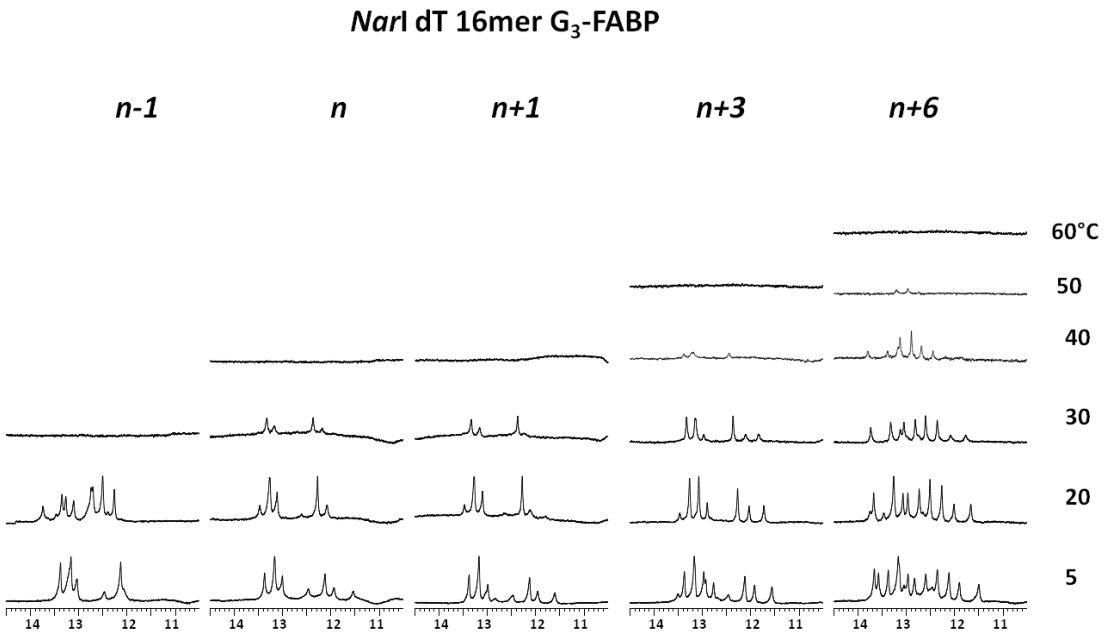


Figure 39

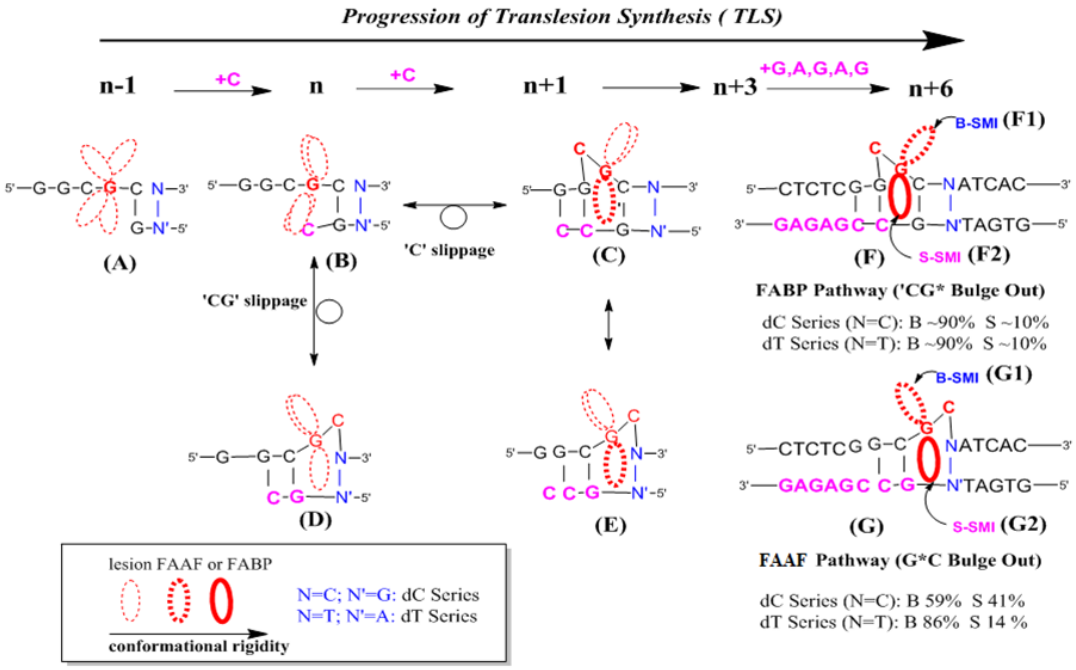

Progression of Translesion Synthesis (TLS)
n-1
+C
n
+C
n+1
n+3
+G,A,G,A,G
n+6
(A)
(B)
'C' slippage
(C)
'CG' slippage
(D)
(E)
B-SMI (F1)
(F)
S-SMI (F2)
FABP Pathway ('CG* Bulge Out)
dC Series (N=C): B ~90% S ~10%
dT Series (N=T): B ~90% S ~10%
B-SMI (G1)
(G)
S-SMI (G2)
FAAF Pathway (G*C Bulge Out)
dC Series (N=C): B 59% S 41%
dT Series (N=T): B 86% S 14 %
lesion FAAF or FABP
N=C; N'=G: dC Series
N=T; N'=A: dT Series
conformational rigidity

Figure 40

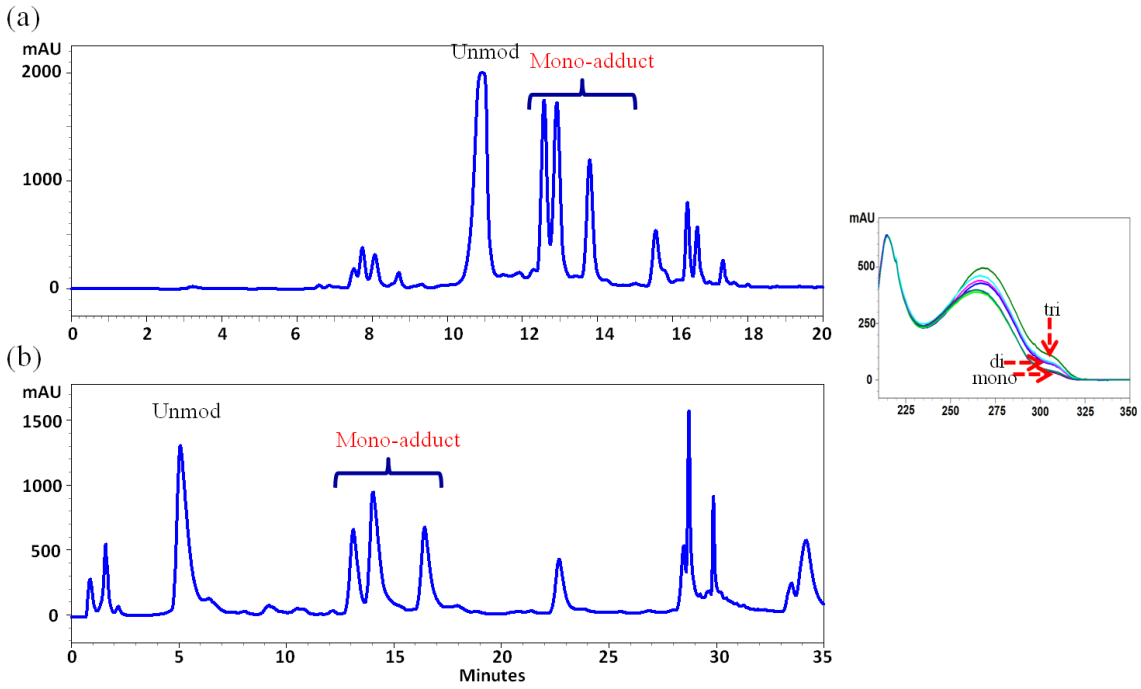

(a)
mAU
2000
1000
0
Unmod
Mono-adduct
0
2
4
6
8
10
12
14
16
18
20
mAU
500
250
0
tri
di
mono
225
250
275
300
325
350
(b)
mAU
1500
1000
500
0
Unmod
Mono-adduct
0
5
10
15
20
25
30
35
Minutes

Figure 41

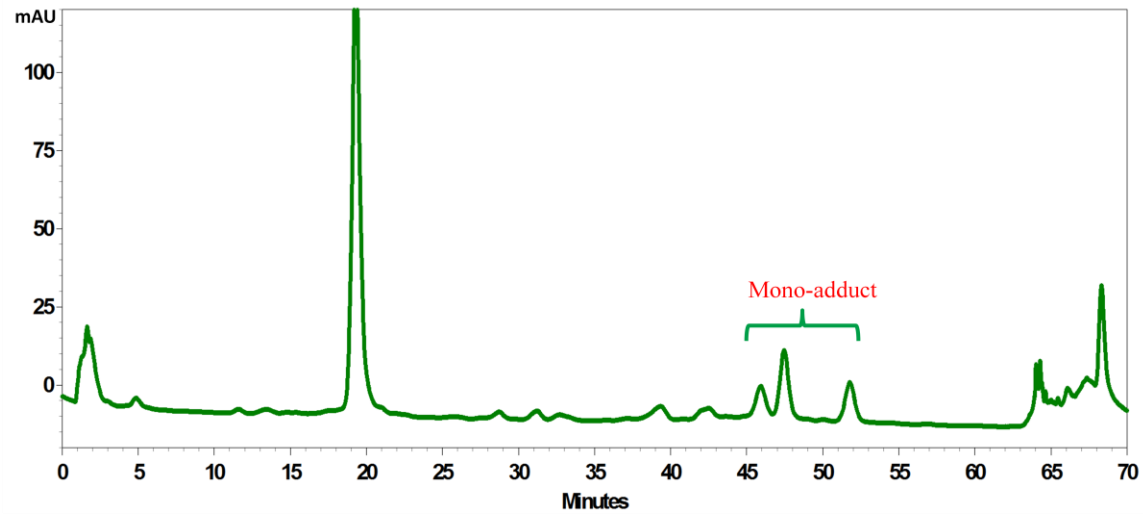
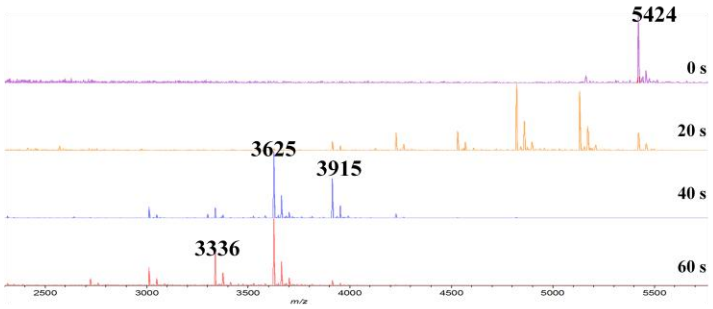

mAU
100
75
50
25
0
Mono-adduct
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
Minutes

Figure 42

(a)

3' EXONUCLEASE DIGEST (dC Peak 2)	
5'-Bio-CTCTCGGCGCCATCAC-3'	5421
5'-Bio-CTCTCGGCGCCATCA	5131
5'-Bio-CTCTCGGCGCCATC	4818
5'-Bio-CTCTCGGCGCCAT	4529
5'-Bio-CTCTCGGCGCCA	4225
5'-Bio-CTCTCGGCGCC	3912
5'-Bio-CTCTCGGCGC	3622
5'-Bio-CTCTCGGCG	3333



(b)

3' EXONUCLEASE DIGEST (dT Peak 2)	
5'-Bio-CTCTCGGCGCTATCAC-3'	5436.16
5'-Bio-CTCTCGGCGCTATCA	5146.98
5'-Bio-CTCTCGGCGCTATC	4833.77
5'-Bio-CTCTCGGCGCTAT	4544.58
5'-Bio-CTCTCGGCGCTA	4250.38
5'-Bio-CTCTCGGCGCT	3927.17
5'-Bio-CTCTCGGCGC	3622.98
5'-Bio-CTCTCGGCG	3333.80

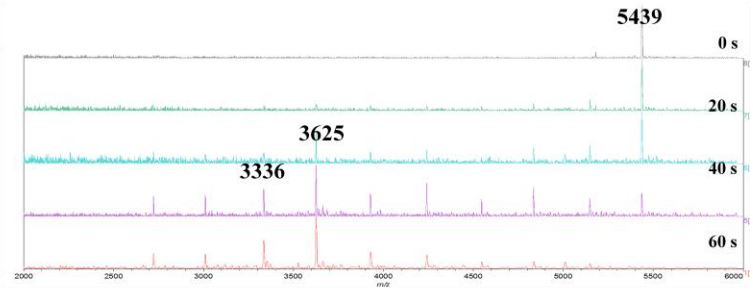


Figure 43

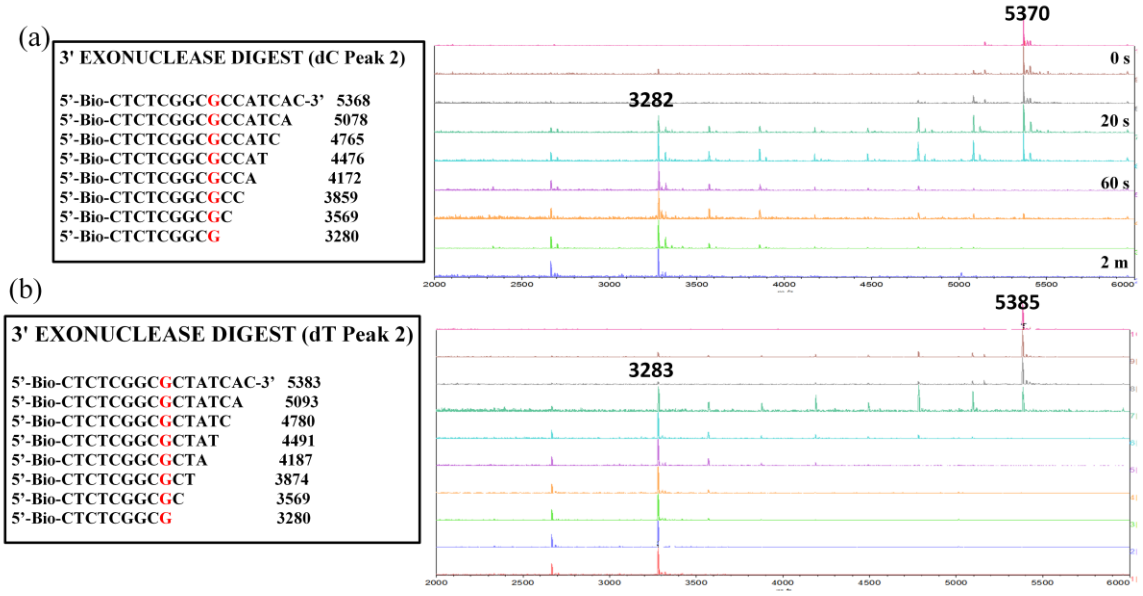

(a)
3' EXONUCLEASE DIGEST (dC Peak 2)
5'-Bio-CTCTCGGCGCCATCAC-3' 5368
5'-Bio-CTCTCGGCGCCATCA 5078
5'-Bio-CTCTCGGCGCCATC 4765
5'-Bio-CTCTCGGCGCCAT 4476
5'-Bio-CTCTCGGCGCCA 4172
5'-Bio-CTCTCGGCGCC 3859
5'-Bio-CTCTCGGCGC 3569
5'-Bio-CTCTCGGCG 3280
5370
3282
0 s
20 s
60 s
2 m
2000
2500
3000
3500
4000
4500
5000
5500
6000
(b)
3' EXONUCLEASE DIGEST (dT Peak 2)
5'-Bio-CTCTCGGCGCTATCAC-3' 5383
5'-Bio-CTCTCGGCGCTATCA 5093
5'-Bio-CTCTCGGCGCTATC 4780
5'-Bio-CTCTCGGCGCTAT 4491
5'-Bio-CTCTCGGCGCTA 4187
5'-Bio-CTCTCGGCGCT 3874
5'-Bio-CTCTCGGCGC 3569
5'-Bio-CTCTCGGCG 3280
5385
3283
2000
2500
3000
3500
4000
4500
5000
5500
6000

Figure 44

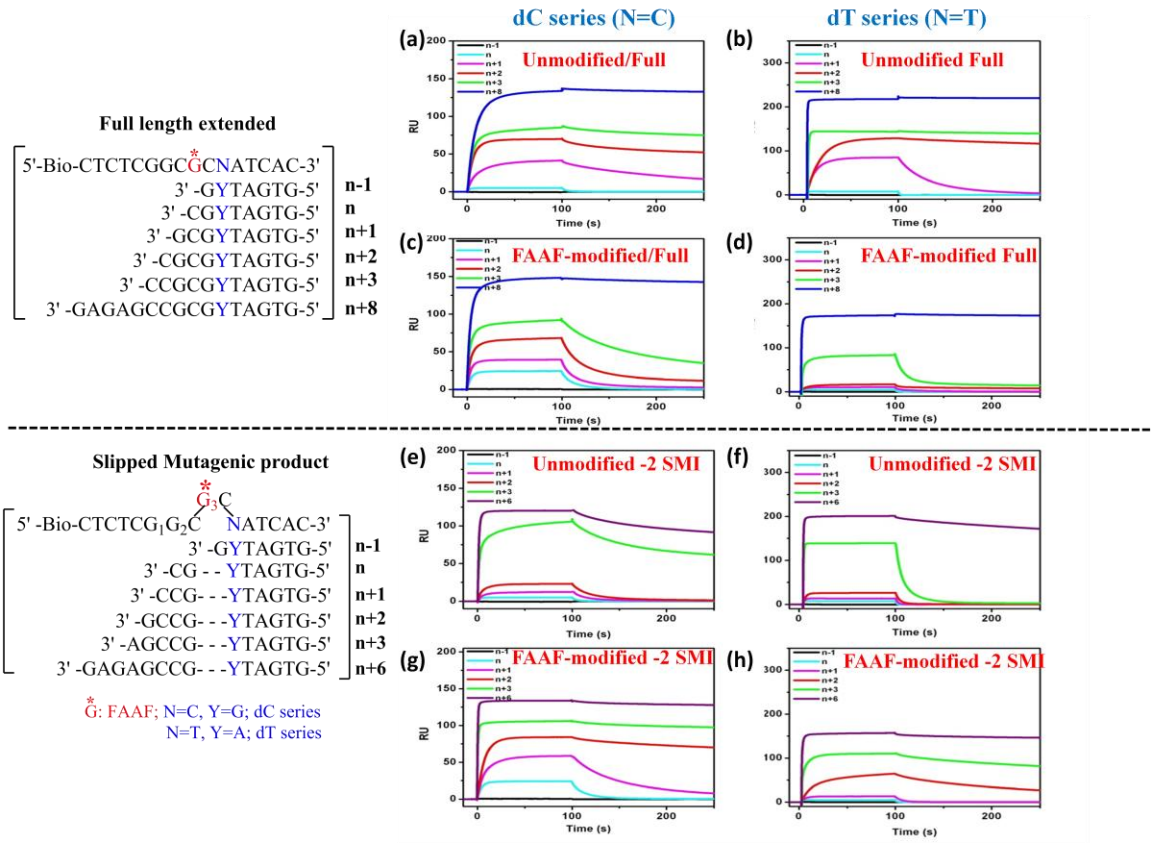

dC series (N=C)
dT series (N=T)
Full length extended
5'-Bio-CTCTCGGC*GCNATCAC-3'
3' -GYTAGTG-5' n-1
3' -CGYTAGTG-5' n
3' -GCGYTAGTG-5' n+1
3' -CGCGYTAGTG-5' n+2
3' -CCGCGYTAGTG-5' n+3
3' -GAGAGCCGCGYTAGTG-5' n+8
(a) Unmodified/Full
(b) Unmodified Full
(c) FAAF-modified/Full
(d) FAAF-modified Full
Slipped Mutagenic product
5' -Bio-CTCTCG1G2C G3C NATCAC-3'
3' -GYTAGTG-5' n-1
3' -CG - - YTAGTG-5' n
3' -CCG- - -YTAGTG-5' n+1
3' -GCCG- - -YTAGTG-5' n+2
3' -AGCCG- - -YTAGTG-5' n+3
3' -GAGAGCCG- - -YTAGTG-5' n+6
*G: FAAF; N=C, Y=G; dC series
N=T, Y=A; dT series
(e) Unmodified -2 SMI
(f) Unmodified -2 SMI
(g) FAAF-modified -2 SMI
(h) FAAF-modified -2 SMI
RU
Time (s)

Figure 45

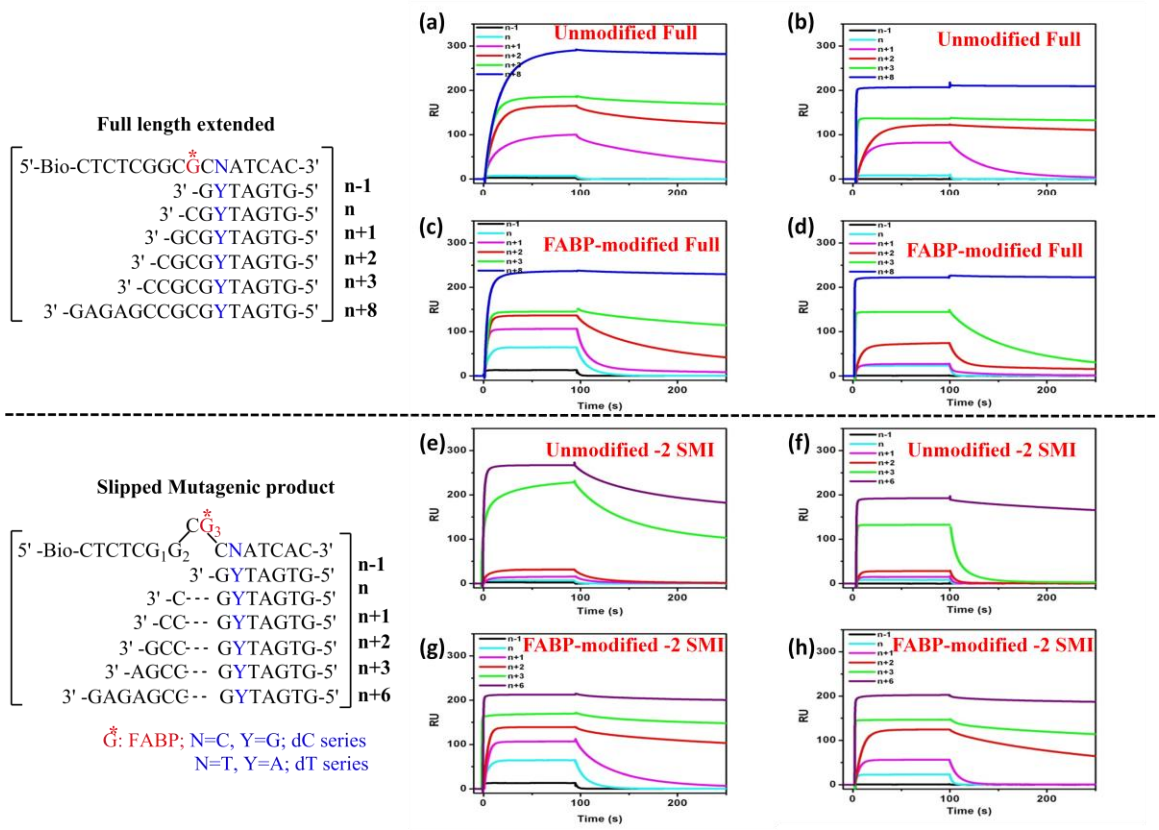

Full length extended
5'-Bio-CTCTCGGCĠCNATCAC-3'
3' -GYTAGTG-5'
3' -CGYTAGTG-5'
3' -GCGYTAGTG-5'
3' -CGCGYTAGTG-5'
3' -CCGCGYTAGTG-5'
3' -GAGAGCCGCGYTAGTG-5'
n-1
n
n+1
n+2
n+3
n+8
(a) Unmodified Full
(b) Unmodified Full
(c) FABP-modified Full
(d) FABP-modified Full
Slipped Mutagenic product
5' -Bio-CTCTCG1G2 CG3 CNATCAC-3'
3' -GYTAGTG-5'
3' -C--- GYTAGTG-5'
3' -CC--- GYTAGTG-5'
3' -GCC--- GYTAGTG-5'
3' -AGCC--- GYTAGTG-5'
3' -GAGAGCC--- GYTAGTG-5'
n-1
n
n+1
n+2
n+3
n+6
Ġ: FABP; N=C, Y=G; dC series
N=T, Y=A; dT series
(e) Unmodified -2 SMI
(f) Unmodified -2 SMI
(g) FABP-modified -2 SMI
(h) FABP-modified -2 SMI
RU
Time (s)

Figure 46

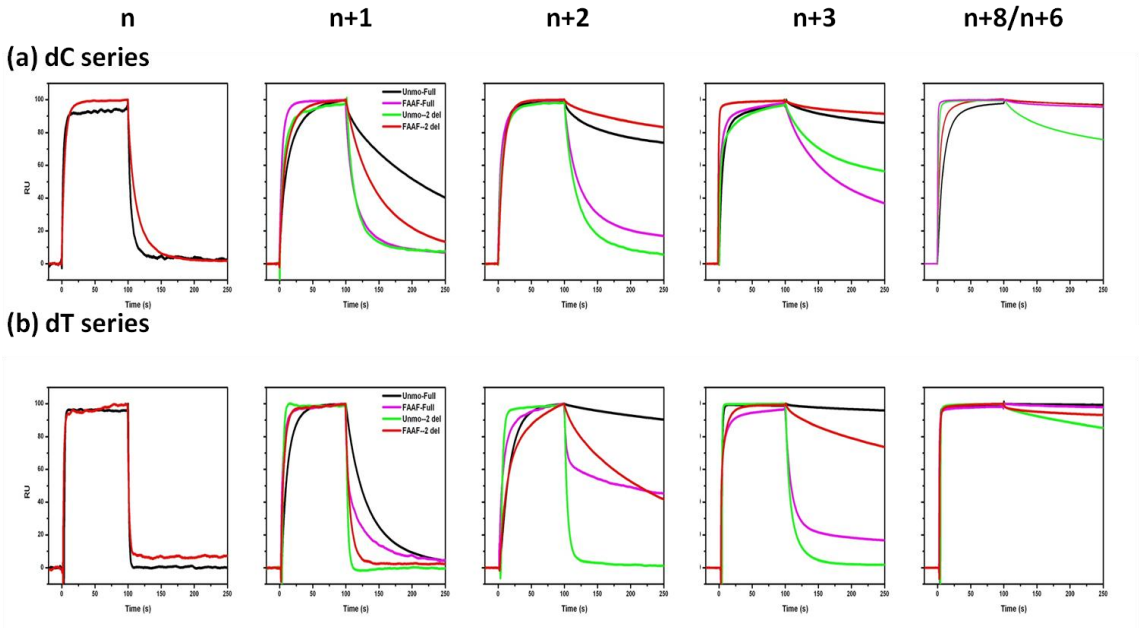

n
n+1
n+2
n+3
n+8/n+6
(a) dC series
(b) dT series
Unmo-Full
FAAF-Full
Unmo--2 del
FAAF--2 del
RU
Time (s)

Figure 47

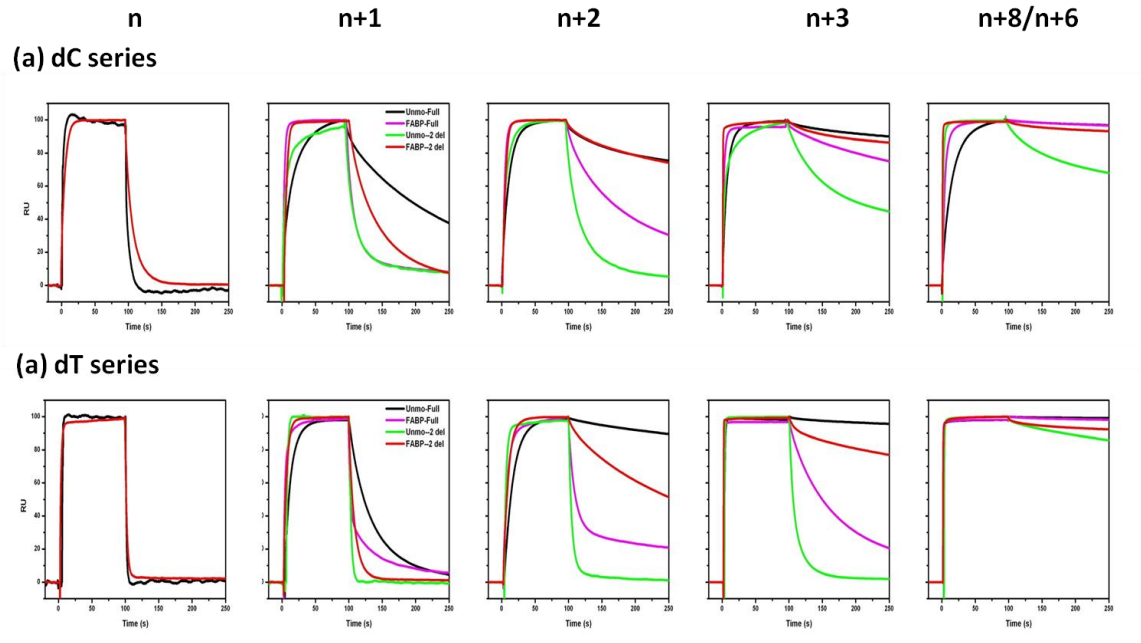

n
n+1
n+2
n+3
n+8/n+6
(a) dC series
(a) dT series
Unmo-Full
FABP-Full
Unmo--2 del
FABP--2 del
RU
Time (s)

Figure 48

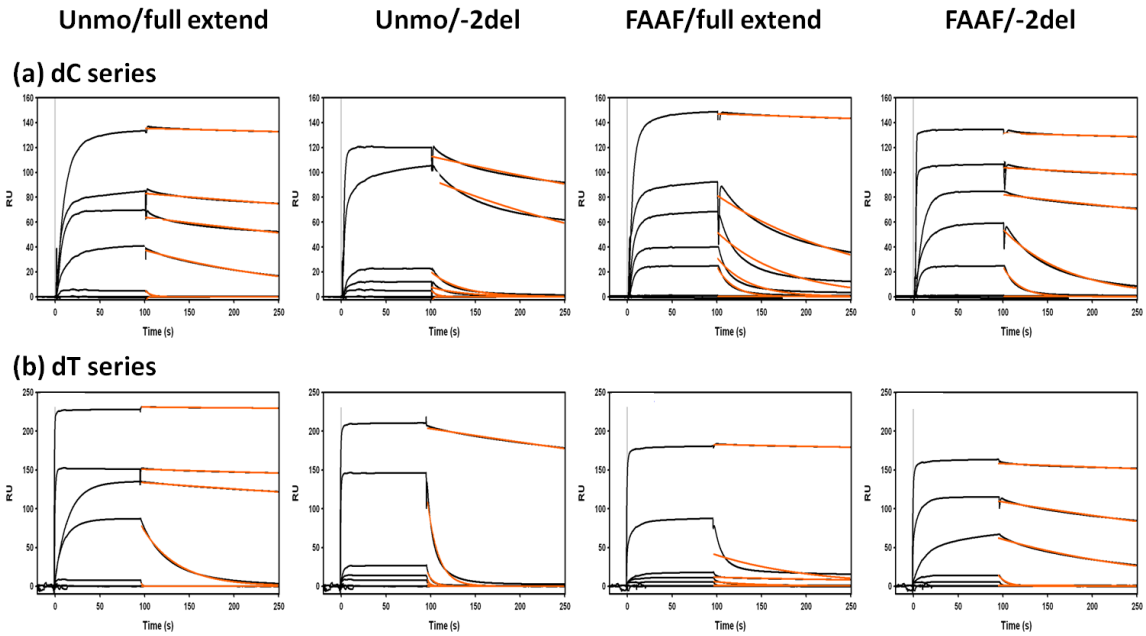

Unmo/full extend
Unmo/-2del
FAAF/full extend
FAAF/-2del
(a) dC series
(b) dT series
RU
Time (s)

Figure 49

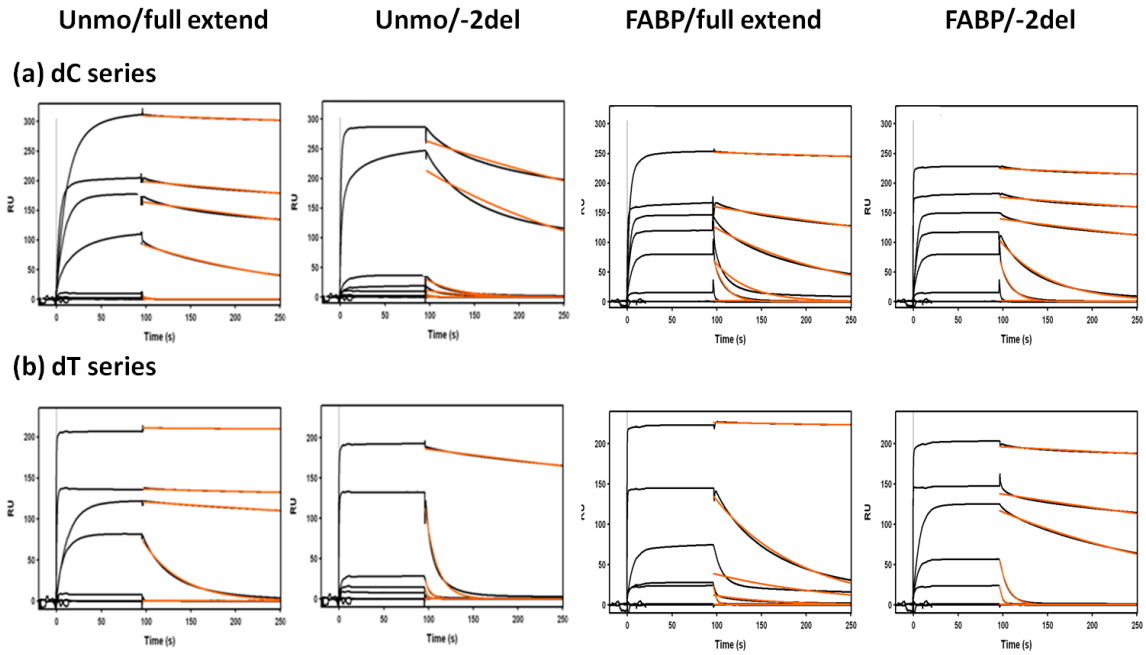

Unmo/full extend
Unmo/-2del
FABP/full extend
FABP/-2del
(a) dC series
(b) dT series
RU
Time (s)

Table 1

5'-CTCTCGGCG$_3$CCATCAC-3'						
	-ΔG[b] (kcal/mol)	**-ΔH[b] (kcal/mol)**	***T*m[c] (°C)**	**ΔΔG[d] (kcal/mol)**	**ΔΔH[e] (kcal/mol)**	**Δ*T*m[f] (°C)**
-2 deletion						
n-1	8.14 (7.87)	55.75 (38.16)	34.58 (31.00)	-0.27	-17.59	3.58
n	9.75 (9.27)	64.78 (57.40)	42.82 (40.72)	-0.48	-7.38	2.10
n+1	10.40 (9.72)	64.87 (61.06)	44.16 (40.95)	-0.68	-3.81	3.21
n+2	12.36 (10.04)	75.42 (64.32)	52.09 (42.35)	-2.32	-11.10	9.74
n+3	14.40 (10.30)	93.64 (68.81)	56.23 (43.17)	-4.10	-24.83	13.06
n+6	16.63 (13.47)	99.54 (94.28)	64.31 (52.60)	-3.16	-5.26	11.70
Full length						
n+1	9.97 (11.74)	56.36 (75.84)	42.65 (48.85)	1.77	19.48	-6.20
n+2	9.70 (14.24)	48.87 (88.07)	41.80 (56.83)	4.54	39.20	-15.03
n+3	10.27 (15.14)	53.49 (87.89)	44.85 (60.46)	4.87	34.40	-15.61
n+8	15.54 (20.52)	90.04 (115.97)	61.03 (71.30)	4.98	25.93	-10.27

[a]Figure 1 shows structure and sequence details (G* = FAAF-adduct).
[b]The average standard deviations for -ΔG, -ΔH, and *T*m are ± 0.13, ± 4.90, and ± 0.30, respectively.
[c]T_m values at 32 mM taken from the $1/T_m$ - $\ln C_t/4$ Meltwin plots .
[d] Δ ΔG = ΔG (modified duplex) - ΔG (control duplex).
[e] Δ ΔH = ΔH (modified duplex) - ΔH (control duplex).
[f] $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Table 2

5'-CTCTCGGCG$_3$CTATCAC-3'						
	-ΔG[b] (kcal/mol)	**-ΔH[ob] (kcal/mol)**	***T*m[b, c] (°C)**	**ΔΔG[d] (kcal/mol)**	**ΔΔH[e] (kcal/mol)**	**Δ*T*m[f] (°C)**
-2 deletion						
n-1	7.20 (6.79)	56.34 (63.79)	27.73 (26.88)	-0.41	7.45	0.85
n	8.56 (7.91)	54.47 (52.13)	34.36 (32.95)	-0.65	-2.34	1.41
n+1	8.62 (8.22)	53.21 (55.31)	34.85 (32.96)	-0.40	2.10	1.89
n+2	10.77 (8.59)	69.42 (53.59)	45.45 (34.92)	-2.18	-15.83	10.53
n+3	12.08 (9.20)	81.81 (66.15)	49.32 (38.19)	-2.88	-15.66	11.13
n+6	14.71 (12.16)	90.14 (85.27)	59.60 (49.12)	-2.55	-4.87	10.48
Full length						
n+1	8.47 (10.89)	41.68 (78.81)	33.48 (44.77)	2.42	37.13	-11.29
n+2	8.67 (12.34)	39.41 (78.75)	34.64 (50.90)	3.67	39.34	-16.26
n+3	8.74 (13.76)	42.00 (88.71)	35.36 (54.76)	5.02	46.71	-19.40
n+8	14.89 (18.60)	92.52 (107.75)	58.26 (67.87)	3.71	15.23	-9.61

[a]Figure 1 shows structure and sequence details (G* = FAAF-adduct).
[b]The average standard deviations for -ΔG, -ΔH, and *T*m are ± 0.16, ± 3.40, and ± 0.30 respectively.
[c]T_m values at 32 mM taken from the $1/T_m$ - $\ln C_t/4$ Meltwin plots .
[d] Δ ΔG = ΔG (modified duplex) - ΔG (control duplex).
[e] Δ Δ H = ΔH (modified duplex) - ΔH (control duplex).
[f] $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Table 3

5'-CTCTCGGCG$_3$CCATCAC-3'

	-ΔG°[b] (kcal/mol)	-ΔH°[b] (kcal/mol)	*T*m[c] (°C)	ΔΔG°[d] (kcal/mol)	ΔΔH°[e] (kcal/mol)	Δ*T*m[f] (°C)
-2 deletion						
n-1	7.92 (7.87)	57.42 (38.16)	30.93 (31.00)	-0.05	-19.26	-0.07
n	9.27 (9.27)	58.31 (57.40)	38.67 (40.72)	-0.00	-0.91	-2.05
n+1	9.82 (9.72)	60.42 (61.06)	41.43 (40.95)	-0.10	0.64	0.48
n+2	12.07 (10.04)	78.24 (64.32)	49.14 (42.35)	-2.03	-13.92	6.79
n+3	12.48 (10.30)	78.52 (68.81)	51.34 (43.17)	-2.18	-9.71	8.17
n+6	14.57 (13.47)	80.85 (94.28)	60.14 (52.60)	-1.10	13.43	7.54
Full length						
n+1	9.51 (11.74)	41.18 (75.84)	40.76 (48.85)	1.77	19.48	-8.09
n+2	9.61 (14.24)	35.20 (88.07)	42.36 (56.83)	4.54	39.20	-14.47
n+3	11.00 (15.14)	52.13 (87.89)	49.20 (60.46)	4.87	34.40	-11.26
n+8	17.85 (20.52)	110.60 (115.97)	64.13 (71.30)	4.98	25.93	-7.17

[a]Figure 1 shows structure and sequence details (G* = FABP-adduct).
[b]The average standard deviations for -DG° , -DH° , and *T*m are ± 0.13, ± 4.90, and ± 0.30, respectively.
[c]T_m values at 32 mM taken from the $1/T_m$ - $\ln C_t/4$ Meltwin plots .
[d]DDG = DG° (modified duplex) - DG° (control duplex).
[e]DDH = DH° (modified duplex) - DH° (control duplex).
[f]D*T*m = *T*m (modified duplex) - *T*m (control duplex).

Table 4

5'-CTCTCGGCG$_3$CTATCAC-3'						
	-ΔG°[c] (kcal/mol)	**-ΔH°[c] (kcal/mol)**	***T*m[c,d] (°C)**	**ΔΔG°[c,e] (kcal/mol)**	**ΔΔH°[c,f] (kcal/mol)**	**Δ*T*m[c,g] (°C)**
-2 deletion						
n-1	7.42 (6.79)	53.38 (63.79)	28.93 (26.88)	-0.63	10.41	2.05
n	8.11 (7.91)	52.93 (52.13)	32.71 (32.95)	-0.20	-0.80	-0.24
n+1	8.56 (8.22)	53.94 (55.31)	35.44 (32.96)	-0.34	1.37	2.48
n+2	10.80 (8.59)	69.58 (53.59)	44.70 (34.92)	-2.21	-15.99	9.78
n+3	11.91 (9.20)	83.83 (66.15)	48.76 (38.19)	-2.71	-17.68	10.57
n+6	14.89(12.16)	98.96 (85.27)	57.16 (49.12)	-2.73	-13.69	8.04
Full length						
n+1	8.73 (10.89)	32.84 (78.81)	32.91 (44.77)	2.16	45.97	-11.86
n+2	8.98 (12.34)	40.47 (78.75)	35.86 (50.90)	3.36	38.28	-15.04
n+3	9.49 (13.76)	45.20 (88.71)	39.51 (54.76)	4.27	43.51	-15.25
n+8	16.34 (18.60)	106.23 (107.75)	60.55 (67.87)	2.26	1.52	-7.32

[a]Figure 1 shows structure and sequence details (G* = FABP-adduct).
[b]The average standard deviations for -DG° , -DH° , and *T*m are ± 0.16, ± 3.40, and ± 0..30 respectively.
[c]T_m values at 32 mM taken from the $1/T_m$ - $\ln C_t/4$ Meltwin plots .
[d]DDG = DG° (modified duplex) - DG° (control duplex).
[e]DDH = DH° (modified duplex) - DH° (control duplex).
[f]D*T*m = *T*m (modified duplex) - *T*m (control duplex).

Table 5

5'-CTCTCGGCG$_3$CCATCAC-3'						
-2 deletion	-ΔG°$_{37°C}$ (kcal/mol)	-ΔH° (kcal/mol)	*T*m[a] (°C)	ΔΔG°$_{37°C}$[b] (kcal/mol)	ΔΔH°[c] (kcal/mol)	Δ*T*m[d] (°C)
n-1	6.86 (5.84)	48.20 (45.00)	41.90 (35.10)	-1.02	-3.20	6.80
n	8.40 (7.85)	62.00 (48.50)	48.80 (48.50)	-0.55	-13.30	3.30
n+1	8.96 (8.31)	62.30 (57.50)	51.90 (49.30)	-0.65	-4.80	2.60
n+2	11.91 (8.92)	86.60 (68.10)	59.20 (50.30)	-2.99	-18.50	8.90
n+3	14.52 (9.20)	110.80 (85.40)	62.40 (48.60)	-5.32	-25.40	13.80
n+6	19.20 (13.57)	147.50 (120.90)	67.10 (57.30)	-5.63	-26.60	9.80
5'-CTCTCGGCG$_3$C*T*ATCAC-3'						
	-ΔG°$_{37°C}$ (kcal/mol)	-ΔH° (kcal/mol)	*T*m[a] (°C)	ΔΔG°$_{37°C}$[b] (kcal/mol)	ΔΔH°[c] (kcal/mol)	Δ*T*m[d] (°C)
n-1	5.54 (5.37)	37.30 (30.10)	32.30 (29.60)	-0.17	-7.20	2.70
n	6.68 (6.49)	42.50 (29.30)	41.20 (41.10)	-0.19	-13.20	0.10
n+1	7.23 (6.77)	47.40 (43.90)	44.50 (41.70)	-0.46	-3.50	2.80
n+2	8.95 (7.25)	57.40 (44.12)	53.10 (45.20)	-1.70	-13.28	7.90
n+3	9.79 (7.50)	60.30 (52.90)	57.10 (45.30)	-2.29	-7.40	11.80
n+6	15.40 (12.00)	118.50 (109.90)	63.30 (54.50)	-3.40	-8.60	8.80

[a] T_m value is the maximum point of the DSC curve.
[b] $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).
[c] $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex).
[d] $\Delta T_m = T_m$ (modified duplex) - T_m (control duplex).

Table 6

	FAAF		FABP	
	Blue Shift Δ^a(G*- G)(nm)		Blue Shift Δ(G*- G)(nm)	
	dC	dT	dC	dT
n-1	6	4	4	1
n	6	2	3	1
n+1	5	3	2	2
n+2	5	3	5	1
n+3	5	1	5	2
n+6	7	1	3	1

Δ^a: Difference in the wavelength of positive band between adduct modified and control.

Table 7

Position	Sequence	dG/full ($\times 10^{-2}$)	dG/-2 del ($\times 10^{-2}$)	FAAF/full ($\times 10^{-2}$)	FAAF/-2 del ($\times 10^{-2}$)
n-1	dC	80.0	80.0	0	0
	dT	4.00	4.00	0	0
n	dC	16.6	16.6	5.56	5.56
	dT	85.0	85.0	0.500	0.500
n+1	dC	0.553	2.67	1.38	1.39
	dT	2.53	35.0	2.31	10.6
n+2	dC	0.144	3.00	0.101	0.099
	dT	0.0622	14.8	0.270	0.559
n+3	dC	0.0702	0.311	0.0371	0.0410
	dT	0.0224	8.08	1.04	0.175
n+6	dC		0.146		0.0188
	dT		0.0898		0.0279
n+8	dC	0.0129		0.0133	
	dT	0.00421		0.0124	

* Errors are within the limit of 5%

Table 8

Position	Sequence	dG/full ($\times 10^{-2}$)	dG/-2 del ($\times 10^{-2}$)	FABP/full ($\times 10^{-2}$)	FABP/-2 del ($\times 10^{-2}$)
n-1	dC	9	9	0.8	0.8
	dT	7	7	0	0
n	dC	17	17	7.68	7.68
	dT	1	1	27.8	27.8
n+1	dC	0.5534	2.22	2.67	1.83
	dT	2.37	33	1.55	9.98
n+2	dC	0.1328	2.93	0.68	0.014
	dT	0.0613	15.9	0.744	0.4017
n+3	dC	0.0696	0.4207	0.1507	0.06416
	dT	0.0197	7.46	1.045	0.128
n+6	dC		0.1914		0.02808
	dT		0.0792		0.0301
n+8	dC	0.0165		0.0183	
	dT	0.0036		0.0082	

* Errors are within the limit of 5%

Appendix

Published in ***Nature Protocol Exchange***, 2013, doi:10.1038/protex.2013.054

Binding kinetics of DNA-protein interaction using surface plasmon resonance

V.G. Vaidyanathan, L. Xu and Bongsup P. Cho*

Surface plasmon resonance (SPR) has been used extensively in the field of DNA/DNA, DNA/protein, and small molecule protein/DNA interactions. However, there have been growing concerns with regard to the proper designing of experiments and the quality of analysis and reporting of SPR results *(1)*. Here we describe a protocol that is designed to address some of those issues. It encompasses procedural steps beginning with immobilization of streptavidin on CM5 chips to the final step of data reporting on DNA-polymerase interaction binding kinetics. In evaluating the protocol, we carried out experiments using a simple methodology developed in our laboratory, taking advantage of the high sensitivity and superior signal-to-noise ratio of Biacore T200. We probed the binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo$^-$) and various arylamine DNA lesions. We employed unmodified and carcinogen-modified oligonucleotides in the presence and absence of dNTPs. The total time required to carry out the method to completion is between one and two weeks, approximately two days for the SPR binding assays and one week for synthesis, purification, and characterization of modified oligonucleotides. Though the protocol presented here is meant for Biacore T100 or T200 model, the overall methodology can be applied for other instruments also.

Reagents

CM5 sensor S chip (Research grade, cat. No. BR-1005-30)

HBS-EP+ (10X containing 0.1 M HEPES, 1.5 M NaCl, 30 mM EDTA and 0.5% v/v Surfactant P20) (GE Healthcare, cat. No. BR-1001-88)

Streptavidin (Piercenet, cat. No. 21125)

HBS-P+ (10X containing 0.1 M HEPES, 1.5 M NaCl, and 0.5% v/v Surfactant P20)(cat. No. BR-1003-68)

Formamide (Sigma-Aldrich, cat. No. F7508)

Amine coupling kit (GE Healthcare, cat. No. BR-1000-50)

Bromophenol blue (Sigma-Aldrich, cat. No. B0126)

EDTA (EMD Biochemicals, cat. No. 4055-100ML)

Tris/NaCl (Fisher Scientific, cat. No. BP2478-500)

T4 DNA ligase and ligase buffer (New England BioLabs, cat. No. M0202S)

Sigmacote (Sigma Aldrich, cat. No. SL2)

Bovine Serum from Albumin (Sigma, cat. No. A9418)

40%, 19:1 Acrylamide/Bis (Bio-Rad, cat. No. 161-0144)

TBE (Promega, cat. No. V4251)

Urea (Fisher Scientific, cat. No. 104924)

Ammonium persulfate (APS) (Sigma, cat. No. A3426)

TEMED (Fisher BioReagents, cat. No. BP150-100)

10% glycerol (Sigma-Aldrich, cat. No. G5516)

ddTTP (GE Healthcare, cat.No. 27-2045-01)

Magnesium chloride (Fisher Scientific, cat. No. M8266)

Sodium acetate (Sigma-Aldrich, cat. No. S8750)

3-Hydroxypicolinic acid (3-HPA) (Fluka analytical, cat. No. 56197)

Ammonium citrate dibasic (MP, Biomedicals, cat. No. 152494)

Tris/EDTA (Fisher Scientific, cat. No. BP2475-1)

n-Butanol (ACROS, cat. No. 42349-0010)

Phenol: Chloroform: Isoamyl acohol (25:24:1, v/v) (Invitrogen, cat No. 15593-031)

Chloroform (Pharmco-AAPER, cat No. 309000000)

DNA (Eurofins)

Kf-exo$^-$ (gift from Dr. Catherine Joyce at Yale University)

Equipments

Biacore T200 SPR instrument (GE Healthcare)

MALDI-TOF spectrometer (Axima Performance, Shimadzu Biotech)

Sequencing gel apparatus (Bio-Rad)

Centrifuge (Eppendorf, 5414 D)

Speedvac (ThermoSavant, model: SPD 2010-220)

HPLC instrument (Hitachi LaChrome Elite L2400 series)

Spectrophotometer (Eppendorf)

Dry bath (Isotemp, Fisher Scientific)

Procedure (Duration: 5 days)

Sample preparation

Day 1:

A. Preparation of 5′-Biotin-DNA-83 mer Ligation

1. DNA annealing: Mix 5′-Biotin-DNA 31 mer (unmodified or modified) and 52 mer hairpin DNA (1:1.5) ratio in 10 mM Tris/50 mM NaCl buffer and heat to 95°C for 5 min and cool down slowly to room temperature (approx. 2-3 h).

2. Dry the sample in Speedvac and dissolve it in 25 μL deionized water and desalt it using Illustra G-25 spin columns.

3. To the desalted solution, add 3 μL T4 DNA ligase buffer (10 x), add T4 DNA ligase 2.5 μL (2000 U/μL) and 19.5 μL deionized water and incubate at 20°C for 16 h.

4. Centrifuge and add 20 μL loading dye (consists of 50 μL 0.5 M EDTA/ 950 μL formamide), heat it to 95°C for 5 min; cool it down using ice-bath.

B. Purification of oligonucleotides (83 mer) by using 10% denaturing gel

5. Mix 40% acrylamine/ Bis 17.5 mL, 10 x TBE 7mL, urea 29 g in 39.5 mL deionized water in a conical flask and dissolve the mixture.

6. Wipe either outer or inner plate with Sigmacote. (**Critical step**: Don't wipe Sigmacote on both plates).

7. Setup the glass plates, cast the gel after adding 200 μL APS (30% w/v) and 100 μL TEMED to the acrylamide solution (step 5) and leave it for 30-45 min to solidify.

8. After removing the comb, flush the wells with the running buffer (1 x TBE) to remove the residual urea.

9. Pre-run the gel at 2,000 V for 30 min.

10. Load the DNA samples and run the gel at 2,000 V for 2-3 h.

11. After completion of the run, cool down the gel with cold water and pry the gel plates quickly.

12. Cover the gel with saran wrap, peel the gel and expose over the TLC plate.

13. Cut desired ligated oligonucleotide bands by exposing under short wavelength UV and transfer to a microcentrifuge (1.5 mL).

14. Crush the gel using micropipette tip.

15. Add 1 mL 1x TE buffer and keep in the -80° C refrigerator for 10 min, heat it at 95° C for 5 min; centrifuge and collect the supernatant.

16. Repeat step 15 for three times and pool the supernatant into one.

17. Add 1 mL 1 x TE buffer to the crushed gel, incubate at 37° C overnight and centrifuge and merge the supernatant with step 16.

Day 2:

18. Filter the pooled solution using 0.2 μm filter.

19. Reduce the volume to 0.2 mL by extracting with n-butanol.

20. Add 200 μL Phenol: Chloroform: Isoamyl alcohol (25:24:1, v/v), vortex, centrifuge and discard the organic layer.

21. To the aqueous solution add 200 μL chloroform and vortex, discard the organic layer.

22. Add 20 μL sodium acetate (pH 5.2, 3 M), 80 μL deionized water and 1.2 mL 100% ethanol, freeze it in -80° C for 30 min.

23. Centrifuge the sample at 13,000 rpm for 30 min and remove the supernatant.

24. Add 100 μL 70 % ethanol, centrifuge for 5 min, remove the supernatant and dry it in speedvac.

25. Dissolve the white precipitate in 25 μL deionized water and desalt it using spin column.

C. Preparation of 5′-Biotin-DNA-84 mer

26. Mix 1 μL Klenow fragment-exo$^-$ (Kf-exo$^-$) with 4 μL dilution buffer (50 mM Tris/ 10% glycerol/ 100 μg/ml BSA), 1 μL ddTTP (100 mM), 10 μL $MgCl_2$ (5 mM), Tris (50 mM) to 5′-Biotin-DNA 83 mer (in 10:1 ratio, Kf-exo$^-$: DNA), incubate at 37° C overnight.

Day 3:

27. Repeat steps 20-25.

28. Purify the oligonucleotides using RP-HPLC, Clarity column (pore size 3 μm, Oligo-RP 50 × 4.6 mm, cat. No. 00B-4441-E0) in the mobile phase (ammonium acetate and acetonitrile), linear gradient: 3% acetonitrile increase to 7% in 5 min, 17% acetonitrile in 20 min, 22% acetonitrile in 25 min.

29. Lyophilize the samples and measure the OD at 260 nm.

D. Characterization of oligonucleotides using MALDI-TOF

Setting up Calibration file for linear negative mode (for MW > 10,000 Da)

30. HPLC purified 52 mer hairpin DNA (MW 15,161 Da), 80 mer (MW 24,293 Da), 90 mer (MW 27,431 Da) and 100 mer DNA (MW 30,496 Da) are used as calibration standards.

31. Prepare the standard MALDI samples by mixing 1 μL of standard (100 pmol) with 1 μL 3-HPA (50 mg/mL in acetonitrile: water 1:1 v/v) and 1 μL ammonium citrate dibasic (50 mg/mL, water).

32. Spot the standards (1 μL) on MALDI steel plate (model DE 1580 TA).

33. Dry the sample spots and insert the plate in the MALDI instrument.

34. Choose the linear negative tuning mode, molecule range 5,000-32,000, firing

power 120, profiles 200, and shots 100, pulsed extraction optimized at 30,000 Da.

35. In the calibration window, enter 4 standards' mass and name.

36. Fire one standard a time, place the cursor to the required peak and update in the calibration window.

37. Repeat this step to finish the rest of the standards, and click the “Calibrate” button twice.

38. Save the calibration method in the calibration files.

For characterization of 31, 83, 84 mer 5’-Biotin-DNA

39. Mix 100 pmol oligonucleotide with 1μL 3-HPA and 1 μL ammonium citrate dibasic; spot it on MALDI plate.

40. Choose linear negative mode, molecule range 5000-30,000, firing power 100-120, profiles 200, shots 100, pulsed extraction optimized at 30,000 Da.

41. Load the linear negative calibration profile.

42. Start firing 83 and 84 mer samples.

43. For 31 mer DNA (MW < 10,000 Da), linear negative mode is not applicable because of large signal to noise ratios, reflectron positive mode and peptide calibration profile can be used.

44. In the peak processing part, advanced scenario is used, along with 1 channel peak width, average smoothing method, 20 channels smoothing filter width, subtract the baseline, 80 channels of baseline filter width, 25 % Centroid threshold peak detection method, double threshold, 1 mass range.

Day 4

Step 1: Immobilization of Streptavidin

Open > New Wizard Template > Immobilization

45. Select Chip type CM5.

46. Check immobilize flow cells (1, 2) or (1, 2, 3, 4) (keeping 1, 3 as blank and 2,4 are samples).

47. **Flow cell 1:** Method: Amine; Check specify contact time and flow rate: Contact time: 420 s; Flow rate: 10 μL/min.

48. **Flow cell 2:** Method: Amine; Ligand: Streptavidin; Dilute ligand: Uncheck (if it is already diluted); Check specify contact time and flow rate: Contact time: 420 s; Flow rate: 10 μL/min.

49. Prime before run (check if it is not primed before).

50. Analysis temperature: 25 ° C.

51. Sample compartment temperature: 25 ° C.

52. For immobilizing flow cell 1: EDC: 89 μL; NHS: 89 μL; Empty vial; Ethanolamine: 129 μL.

53. For immobilizing flow cell 2: EDC: 89 μL; NHS: 89 μL; Empty vial; Ethanolamine: 129 μL; Streptavidin: 98 μL.

54. Choose menu >Automatic positioning > Pooling > Auto.

55. Keep running buffer in left tray and insert buffer tubing A (In this step, running buffer: 100 mL 1 x HBS-EP$^+$ buffer; but varies in DNA binding kinetics).

56. Keep fresh deionized water (200 mL) in right tray.

57. Empty the waste bottle.

58. Save the wizard (save as ….).

Step 2: DNA coating

59. Choose Run > Manual run > select the flow path: 1, 2.

60. Flow rate: 2 μL/min and select the appropriate rack.

61. Inject 50 mM NaOH 60 s pulse for 5 times till the drop in response unit before and after injection of NaOH lies between 10 and 20 RU.

62. Inject 1 x HBS-EP^+ buffer for 3 times (1 min pulse).

63. Leave the chip for 30 min to 1 h depending on the baseline drift.

64. Select the channel to flow cell 2 (**Critical step:** Don't forget to change the flow cell to 2 otherwise biotin-DNA will be coated in flow cell 1 also and it is difficult to remove the biotin-DNA).

65. **Critical step**: Inject biotinylated DNA (0.25 or 0.3 nM) for 1 min and stop the injection after 30 s.

66. **Critical step**: Check the rise in the response unit. If it goes beyond 5 RU with in 30 s, dilute the sample.

67. **Critical step**: Increase in response should be between 0.5 and 3 RU. Leave it for 15 min to see any drift in baseline.

68. Change the buffer to 1 x HBS-P^+/ 100 ug/mL BSA/ 5 mM $MgCl_2$. Prime the system.

69. To ensure the hairpin-oligonucleotide contains 5'-dideoxy base, inject the sample containing Kf-exo^-+ 100 mM ddTTP + 1 x HBS-P^+/ BSA/ 5 mM $MgCl_2$ buffer for 5 min.

70. Inject 0.05% SDS for 240 s (2 μL/min flow rate) and inject running buffer for 5 min. Now the surface is ready for further studies.

Step 3: Regeneration scouting

71. Select the flow path and chip type.

72. Choose number of regeneration buffer set (either 1 or 2).

73. Run conditioning cycles with buffer (1 x HBS-P^+/ BSA/ 5 mM $MgCl_2$ for 30 s and 3 injections).

74. Sample name (Prepare Kf-exo- (5 nM)).

75. Contact time: 30 s and flow rate: 100 μL/min.

76. Scouting parameters: Flow rate: 100 μL/min; Contact time: 30 s; Stabilization period; 300 s; Number of conditions: 3; Number of cycles for each condition: 5 ; Lock: contact time; Provide names for each regeneration buffer: (0.1% SDS; 0.05% SDS; 1M NaCl in this case).

77. Repeat steps 54-58. (**Critical step:** Check the binding response and baseline drift of all the cycles.) (**Critical step:** Running buffer: 1 x HBS-P^+/ 100 ug/mL BSA/ 5 mM $MgCl_2$)

Step 4: Surface performance

78. Repeat steps 72-75. (Flow rate: 100 μL/min; Contact time: 30 s; Stabilization period: 300 s; Number of conditions: 3; Number of cycles: 20)

79. Select the best regeneration buffer from previous assay (regeneration buffer scouting).

80. Repeat steps 54-58. (**Critical step:** Check the binding response and baseline drift of all the cycles.)

Step 5: Mass transport

81. Repeat steps 72-75.

82. Choose regeneration: Solution: 0.05% SDS (in this case); Contact time: 30 s; Flow rate: 100 μL/min and Stabilization period: 300 s.

83. Input Sample Id; (either one or more samples of different concentrations). Rate should be independent of flow rate (Figure 1).

84. Repeat steps 54-58.

Day 5:

Step 6: Kinetics

85. Repeat steps 72-75.

86. Injection parameters: Contact time: 30 s; Flow rate: 100 μL/min; Dissociation time: 60 s; Regeneration solution: 0.05% SDS; Contact time: 30 s; Flow rate: 100 μL/min; Stabilization period: 300 s.

87. Input sample id; concentration; molecular weight etc. (**Critical Step:** At least each concentration of analyte should be injected in duplicate or triplicate and in random.)

88. Repeat steps: 54-58.

Step 7: BIAevaluation

89. Choose kinetics/Affinity > Surface bound.

90. Select the curves to fit.

91. Zoom the curves to remove the spikes by right click and drag.

92. Select kinetics or affinity to fit the data.

93. Select the model to fit. (start with 1:1)

94. Check the kinetic data in tools.

95. **Critical Step:** As the modification factor (M) sliding bar varies, blue and red lines (rate constants increase or decrease) should vary. Otherwise data is limited by mass transport (Figure 2).

96. **Critical Steps: Check the following parameters**

i. How well does the fitted curve overlay with the experimental data.

ii. Does the random injection of same concentration of analyte overlay.

iii. Check the residual range (between the green lines in Biaevaluation software).

iv. Does χ^2 fall within 1% of highest signal response.

v. Does k_a and k_d values fall within instrument specification and check whether it makes any biological significance.Make sure T values are significant. For k_a and k_d, T values should be higher and k_t values, it should be as minimum as possible.

vi. Mass transport limitation: Check whether data is limited by mass transport (step 96).

vii. Check the U value (this feature present in Biacore T200 not in T100).

Step 8: BIAsimulation

97. Once k_a and k_d values are determined, input these parameters in BIAsimulation Basic kinetics module.

98. Compare the curves between simulated and experimental curves (Figure 3).

Step 9: Preparation of reports

99. The fitted curves can be plotted by exporting the file in ASCII format by right-click over the curves and imported it in any plotting software.

Anticipated results

Due to the high sensitivity of Biacore T200, the DNA coated on the surface and polymerase used in this study was as low as 0.7-3.5 RU and 10 nM, respectively. The amount of DNA and polymerase required for this assay is 20-100 fold lower than that required by previously reported methods*(2)*. With low DNA concentration potentially confounding complexities of mass transport limitation could be minimized and possibly

avoided. In this protocol, hairpin-DNA was used to achieve additional stability as well as to overcome the likelihood of presence of single-stranded template alone which may complicate in obtaining accurate kinetics parameters.

References

1. Rich, R. L., and Myszka, D. G. (2008) Survey of the year 2007 commercial optical biosensor literature, *Journal of molecular recognition : JMR 21*, 355-400.

2. Delagoutte, E., and Von Hippel, P. H. (2003) Function and assembly of the bacteriophage T4 DNA replication complex: interactions of the T4 polymerase with various model DNA constructs, *J Biol Chem 278*, 25435-25447.

Associated Publication

V.G. Vaidyanathan, L. Xu and B. P. Cho (**2012**) *Binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo−) and various arylamine DNA lesions characterized by surface plasmon resonance*

Chem. Res. Toxicol., 25, 1568-1570.

Author Information

V.G. Vaidyanathan,[⊥] Lifang Xu, and Bongsup P. Cho*

Affiliation: Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI 02881 USA

[⊥] Current address: Chemical Lab, Central Leather Research Institute, Adyar, Chennai 600020. India Email: vaidychem@yahoo.com

*Corresponding Author: Phone: +1 401 874 5024; Fax: +1 401 874 5766

E-mail: bcho@uri.edu.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

The authors thank Dr. Paul Belcher (GE Healthcare) for his valuable inputs. This research is supported by NCI/NIH (CA098296) and NCRR/NIH (P20 RR016457).

Figure legends

Figure 1: Effect of mass transport limitation. (a) Rate varies with the flow rate (5, 15, 75 μL/min) of Kf-exo$^-$ due to high DNA surface density (b) Rate is independent of flow rate.

Figure 2: Binding kinetics of polymerase to DNA affected by mass transport. Red circles show the modification factor M at maximum value 10. The original data is in black; the blue curves are simulated k_a and k_d multiplied by M; the red show the simulated k_a and k_d divided by M. The divergence of red and blue curves will be observed in no mass transfer case. (a) and (b) kinetics data completely affected by mass transfer as the modification factor varies (c) No mass transfer.

Figure 3: Binding kinetics of polymerase with DNA. (a) Experimental and fitted data in black and red, respectively. (b) Simulated data for various concentrations using the k_a and k_d values (k_a: 9.2×10^7 $M^{-1}s^{-1}$; k_d: 0.12 s^{-1}

Figure 1:

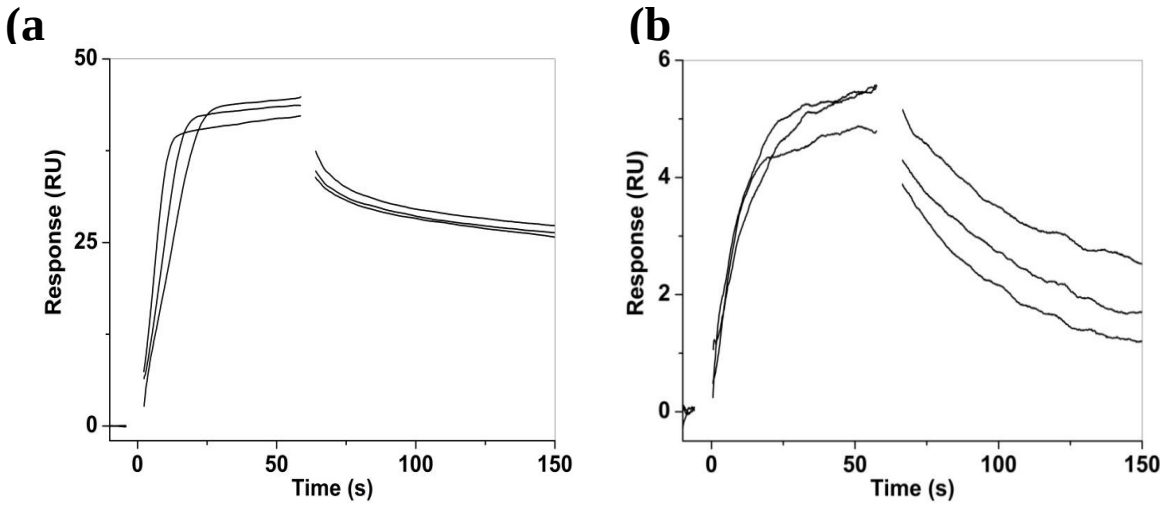

(a
50
25
0
Response (RU)
0
50
100
150
Time (s)
(b
6
4
2
0
Response (RU)
0
50
100
150
Time (s)

Figure 2:

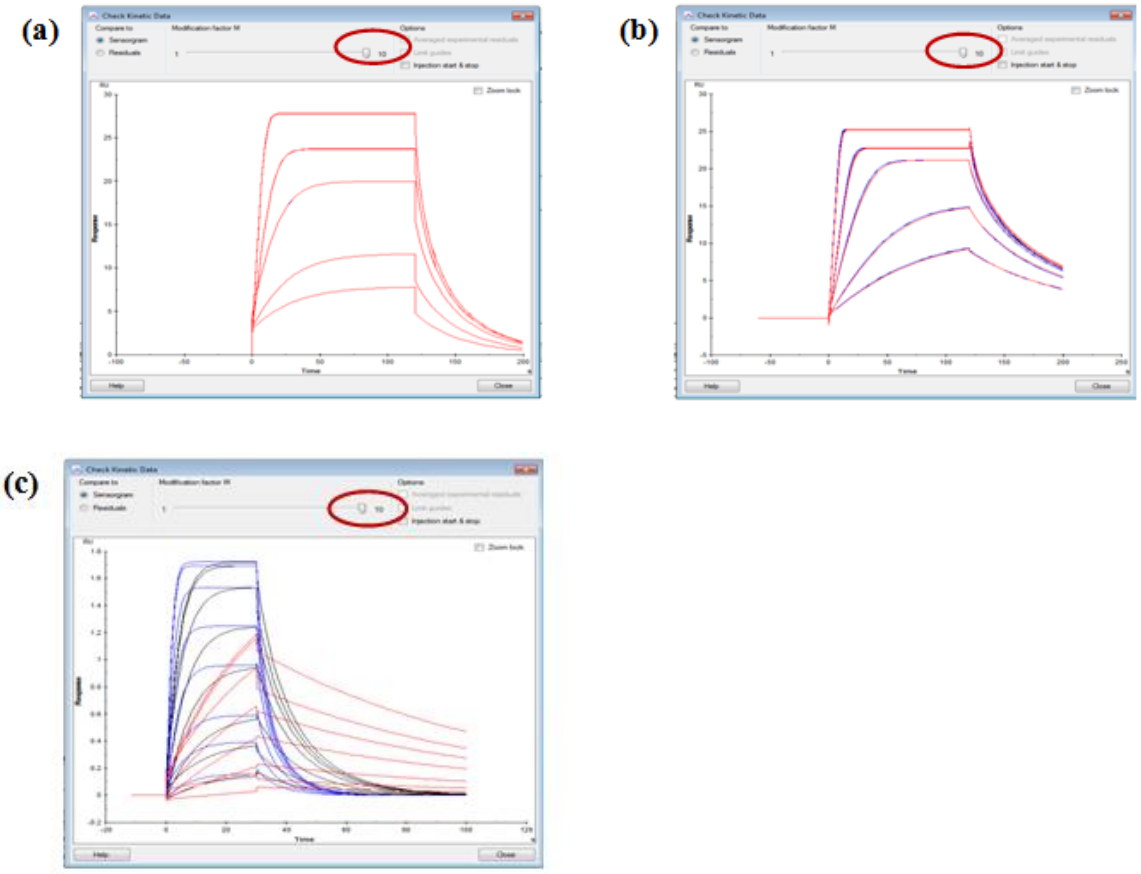
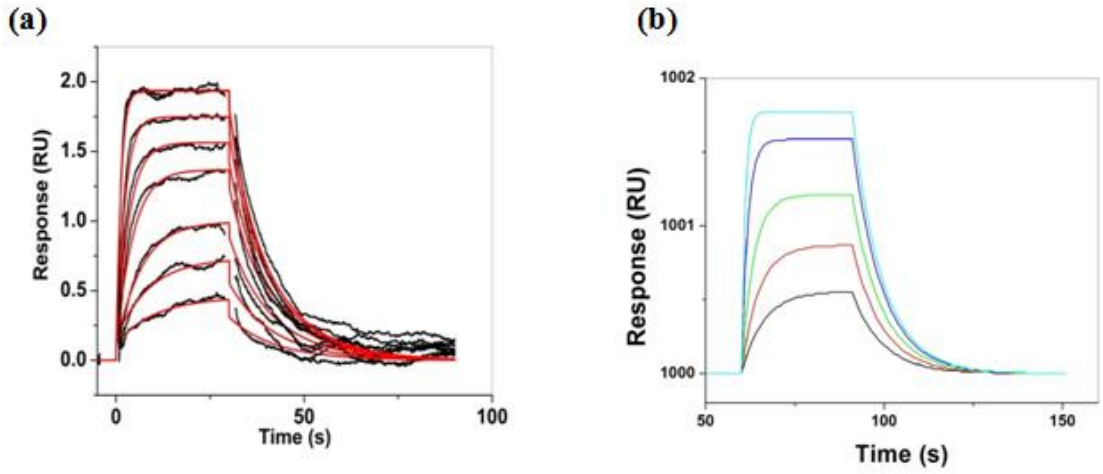


Figure 3:


(a)
Response (RU)
2.0
1.5
1.0
0.5
0.0
0
50
100
Time (s)
(b)
Response (RU)
1002
1001
1000
50
100
150
Time (s)