2017

Conformational Insights Into Aminobiphenyl-Modified DNA: Implications for Mutation and Repair

Ang Cai
University of Rhode Island, angcai@uri.edu

CONFORMATIONAL INSIGHTS INTO AMINOBIPHENYL-MODIFIED DNA: IMPLICATIONS FOR MUTATION AND REPAIR

BY

ANG CAI

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

2017

DOCTOR OF PHILOSOPHY DISSERTATION

OF

ANG CAI

APPROVED:

Dissertation Committee

Major Professor Bongsup Cho

Roberta King

Joel A. Dain

Deyu Li

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2017

ABSTRACT

Cancer is a major public health problem worldwide. An estimated 1.6 million new cases of cancer will be diagnosed in the United States in 2017. Most cancers are assumed to develop from the initiation of chemically induced DNA damages. Chemical carcinogens are genotoxic and they undergo enzymatic activation into reactive electrophilic species to form covalent bonds with the electron-rich DNA that form adducts. Aromatic amines are one of the most examined chemical carcinogens, and once activated, these amines attack the C-8 position of guanine to form C8-dG adducts (ABP [N-(2′-deoxyguanosin-8-yl)-2-aminobiphenyl], AF [N-(2′-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [N-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene]); these formations result in DNA lesions. These bulky DNA lesions at the replication fork can induce a conformational heterogeneity, which complicates mutational outcomes. These outcomes are due to a single DNA adduct that can adopt multiple conformations at the replication fork, and a polymerase may process each conformation in different ways. Studies have shown differences in the mutational consequences of arylamine adducts in different DNA sequences. AF, AAF, and ABP all adopt anti B-DNA , syn stacked (S), and wedge (W) DNA conformational heterogeneity depending on sequence contexts. A replication polymerase is likely to stall or completely stop DNA synthesis when it encounters a bulky DNA lesion; this outcome allows recruitment of various bypass polymerases. This move will determine whether the synthesis is to be error free or error prone. Understanding the processes involving replication polymerases in arylamine-induced translesion synthesis is important; these processes are implicated in mutation and repair.

Our working hypothesis is that the sequence-dependent arylamine-conformational heterogeneity is an important consideration for its mutational and repair outcomes. In **Manuscript I** (to be submitted to *Journal of Chemical Research in Toxicology*), we explored the arylamine-conformational heterogeneity. We previously reported a 3′-next flanking sequence effect [G*CT vs. G*CA; G*, FABP, N-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl; FAF, N-(2′-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene] on arylamine-DNA lesions in relation to DNA replication and repair. In the present study, we present an unusual sequence effect on a 16-mer fully paired DNA duplex 5′-CTTCTG$_1$G$_2$TCCTCATTC-3′, in which the same lesion modification at G$_1$ and G$_2$ revealed a dramatic difference in S (base-displaced stacked)/B (major groove binding B-type) conformational heterogeneity. We obtained conformational (^{19}F-NMR/CD), calorimetric (DSC and ITC), and binding (SPR) data for a set of four 16-mer modified templates (FABP vs. FAF and TG$_1$*G$_2$T vs. TG$_1$G$_2$*T) during primer elongation. Dynamic ^{19}F-NMR results indicate that FABP modification at G$_1$ and G$_2$ resulted in 67%:33% of B:S and 100% B population ratios, respectively, and the planar FAF in the same sequence contexts exhibited 25%:75% B:S and 83%:17% B:S, respectively. A significant increase was observed in S-conformation (0% to 17%) going from a twisted nonplanar FABP to a planar FAF. Calorimetric (DSC and ITC) and UV-melting data revealed that the B-conformer is a thermodynamic stabilizer in duplex settings, and the S-conformer is a destabilizer. The situation is reversed during the primer elongation process that involves ss/ds duplexes, i.e., S-conformation promotes lesion stacking at the ss/ds replication fork. This finding is in agreement with SPR binding results, which showed that lesion stacking increases the binding affinity of the complementary strands

in the order of FABP > FAF and G*GT > GG*T. Surprisingly, our data revealed that the twisted biphenyl is a better stacker at ss/ds junction than the coplanar FAF, possibly due to its ring flexibility that helps accommodate the replication fork structures. This finding is not the case for fully paired duplexes in which base-displaced stacking is enhanced by lesion planarity. The extent of primer dissociation rates across the lesion was influenced by the DNA sequences and the nature of the lesion (FAF vs. FABP). These results provide conformational insights into the important role of ss/ds sequence effects in modulating bulky lesion-induced translesion DNA synthesis. **Manuscript II** (to be submitted to *Journal of Nature Structural & Molecular Biology*) is the extension of manuscript I. Bulky organic carcinogens are usually activated *in vivo* and form DNA adducts. Some of these DNA adducts exist in multiple conformations that are slowly interconverted to one another. For example, the dG-C8-lesion of the human bladder carcinogen 4-aminobiphenyl (ABP) has been reported to exist in a sequence-dependent mixture of B/S conformations. Different conformations have been shown to lead to different kinetic capacities, binding capacities, and ultimately different translesion synthesis and transcription capacities. Although rare, conformation-dependent mutagenic and repair patterns have been reported. For example, the bulky N-acetylaminofluorene at G_3 in the *Nar*I sequence (5′-$G_1G_2CG_3CN$-3′) induces a high frequency of −2 frameshift mutations, whereas the adducts on G_1 and G_2 are not mutagenic. Results show why mutagenic hotspots exist in the genome. However, the conformation-specific replication block, which is more relevant to replication, has not been reported. Such research deficiency may be due to the dynamicity of the replication block process. As such, the replication block becomes difficult to capture by current

end-point assays, which usually detect either the ensemble of consequence of all the conformers or the culmination of all cellular behaviors, such as mutagenicity or survival rate. In Manuscript I, we showed very unusual sequence-dependent conformational heterogeneities involving FABP-modified DNA under different sequence contexts (TG_1*G_2T [67%B:33%S] and TG_1G_2*T [100%B], G*, FABP, N-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl). Different biological consequences are believed to be linked to different adducts. In the present study, we attempted to correlate the replication block to different conformations from the single FABP-modified DNA lesion. We utilized a combination of biophysical (SPR and steady-state kinetics) tools to reveal the differences *in vitro*. The results were discussed in terms of conformational differences.

Manuscript III (has been accepted by *Journal of ACS Omega*) is a technical paper. We focused on the oligonucleotides. Oligonucleotides serve as important tools for biological, chemical, and medical research. The preparation of oligonucleotides through automated solid-phase synthesis is well established. However, identification of byproducts generated from DNA synthesis, especially from oligonucleotides containing site-specific modifications, is sometimes challenging. Typical HPLC, Mass Spectrometry (MS), and gel electrophoresis methods alone are not sufficient for characterizing unexpected byproducts, especially for those having identical or very similar molecular weight (MW) to the products. We used a rigorous quality control procedure to characterize byproducts generated during oligonucleotide syntheses: 1) purify oligonucleotide by different HPLC systems; 2) determine the exact MW by high resolution MS; 3) locate modification position by MS/MS or exonuclease digestion with MALDI-TOF analysis; and 4) conduct, where applicable, enzymatic assays. We applied

these steps to characterize byproducts in the syntheses of oligonucleotides containing important methyl DNA adducts 1-methyladenine (m1A) and 3-methylcytosine (m3C). In m1A synthesis, we differentiated a regioisomeric byproduct 6-methyladenine, which possesses identical MW to m1A. As for m3C, we identified a deamination byproduct 3-methyluracil, which is only 1 Dalton greater than m3C in the ~ 4,900 Dalton context. The detection of these byproducts would be very challenging if the abovementioned procedure were not adopted.

ACKNOWLEDGEMENTS

Many people were involved in this dissertation process and there are many people to thank. First, I would like to express my sincere gratitude to my advisor Dr. Bongsup Cho for his guidance, understanding, patience, and most importantly, his encouragement during my studies at the University of Rhode Island. I am very lucky to have such an outstanding mentor who gave me the power to work with such independence on my research. For everything you've done for me, Dr. Cho, I cannot thank you enough.

I would also like to thank my committee members, Dr. Roberta King, Dr. Joel Dain, Dr. Deyu Li, and the defense chair, Dr. Ying Zhang, for their valuable contribution and feedback on my dissertation. In particular, I would like to thank my collaborative lab professors, Dr. Deyu Li and Dr. Joel Dain, for their exceptional and valuable suggestions on my projects. Additionally, I am grateful for the time spent with Dr. Cho's lab members, Dr. Lifang Xu, Matthew Blake, Dr. Vaidyanathan Ganesan, Dr. Sathyaraj Gopal and Dr. Vipin Jain. It was a pleasure to work with these fine scientists. Their guidance and patience made it a smooth transition into Dr. Cho's lab for me.

Many thanks goes to all of my friends especially Dr. Hang Ma, Qi Tang, Ke Bian. I thank for their help, support and the wonderful time we had together at the lab. The lab environment could somewhat be stressful, but not when they are around. Their humor and entertainment is contagious. A good laugh was all I needed to get me back to work.

Finally, and most importantly, I would like to thank my family. I owe an obvious debt of gratitude to my parents, Qisheng and Yufeng, my wife, Ai Chang and my sister, Jie

for their love, devotion, constant encouragement and understanding during this journey at URI. It would have been impossible without their support.

PREFACE

This dissertation has been written in manuscript style format according to the University of Rhode Island "Guidelines for the Format of Theses and Dissertations".

MANUSCRIPT I: DNA Base Sequence Effects on Bulky Lesion-Induced Conformational Heterogeneity during Simulated Translesion Synthesis.

This manuscript will be submitted to Chemical Research in Toxicology.

MANUSCRIPT II: Probing the Effect of Bulky Lesion-Induced Replication Fork Conformational Heterogeneity Using 4-Aminobiphenyl-Modified DNA.

This manuscript will be submitted to Journal of Nature Structural & Molecular Biology.

MANUSCRIPT III: Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine.

This manuscript has been accepted by Journal of ACS Omega.

TABLE OF CONTENTS

ABSTRACT ..... **ii**

ACKNOWLEDGEMENTS ..... **vii**

PREFACE ..... **ix**

TABLE OF CONTENTS ..... **x**

LIST OF FIGURES ..... **xi**

LIST OF TABLES ..... **xix**

MANUSCRIPT I **1**

MANUSCRIPT II **71**

MANUSCRIPT III **104**

LIST OF FIGURES

MANUSCRIPT I

Figure 1: (a) Structures of ABP [N-(2ˈ-deoxyguanosin-8-yl)-4-aminobiphenyl], AF [N-(2ˈ-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [N-(2ˈ-deoxyguanosin-8-yl)-2-acetylaminofluorene] and their fluoro models, FABP [N-(2ˈ-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl], FAF [N-(2ˈ-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene], and FAAF [N-(2ˈ-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene]. (b) Major groove views of arylamine-conformational motifs: B-, S-, and W-conformers. (color code: Arylamine- red CPK, modified dG- blue, complementary C- orange). 43

Figure 2: Translesion synthesis (TLS) models for (a) unmodified control, (b) FABP/FAF-modified G1 (G*GT) templates, and (c) FABP/FAF-modified G2(GG*T) templates..... 44

Figure 3: HPLC chromatogram of a reaction mixture of 16-mer sequence (5'-CTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FABP** (N-Acetoxy-N-(trifluoroacetyl)-4'-fluoro-4-aminobiphenyl) and photodiode array UV spectra of unmodified, mono-, di-FABP adducts. 45

Figure 4: HPLC chromatogram of a reaction mixture of 16-mer sequence (5'-CTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FAF** (N-acetoxy-N-(trifluoroacetyl)-7-fluoro-2-aminofluorene) and photodiode array UV spectra of unmodified and mono-FAF adducts. 46

Figure 5: HPLC profile, photodiode array UV spectra of mono-**FABP** adducts and MALDI-TOF mass spectra of **FABP** modified **biotinylated** 16mer-GGT for

SPR experiments: (a) HPLC chromatogram of **FABP** modified **biotinylated** 16mer-GGT. (b) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 3min 5min, 6min and 7min. (c) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 3min 5min, 6min and 7min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion. 47

Figure 6: MALDI-TOF mass spectra of **FABP** modified 16mer-GGT: (a) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 2min and 5min. (b) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 2min and 5min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion. 48

Figure 7: MALDI-TOF mass spectra of **FAF** modified 16mer-GGT: (a) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 1min, 2min, 3min, 5min and 6min. (b) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 1min, 2min, 3min, 5min and 6min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion. 49

Figure 8: Dynamic ^{19}F-NMR spectra (5-60°C) of **FABP**-modified duplexes with primer elongation. (a) FABP-modified G1 (G*GT) showing a 67% B- and 33% S- conformational mixture in full duplex. (b) FABP-modified G2 (GG*T) exhibiting 100% B-conformer in full duplex. For all of the duplexes, the ^{19}F signals around 60°C coalesced into a sharp single peak around -115ppm indicating a denatured FABP-residue. * indicates unassigned minor conformers 50

Figure 9: Dynamic ^{19}F-NMR spectra (5-60°C) of **FAF**-modified duplexes with primer elongation. (a) FAF-modified G1 (G*GT) showing a 25% B- and 75% S- conformational mixture in full duplex. (b) FAF-modified G2 (GG*T) exhibiting 83% B- and 17% S-conformer in full duplex. For all of the duplexes, the ^{19}F signals around 60°C coalesced into a sharp single peak around -117ppm indicating a denatured FAF-residue. * indicates unassigned minor conformers. .. 51

Figure 10: Deuterium isotope effect of FABP-modified 16-mer GGT full duplexes in 10% and 100% D_2O at 25°C. (a) FABP-modified G1 (G*GT) (0.18 and 0.08 ppm for B- and S-conformer, respectively). (b) FABP-modified G2 (GG*T) (0.18 ppm for B-conformer)..... 52

Figure 11: Deuterium isotope effect of FAF-modified 16-mer GGT full duplexes in 10% and 100% D_2O at 25°C. (a) FAF-modified G1 (G*GT) (0.19 and 0.04 ppm for B- and S-conformer, respectively). (b) FAF-modified G2 (GG*T) (0.18 and 0.04 ppm for B- and S-conformer, respectively) 53

Figure 12: CD spectral overlays of primer elongation (9 mer to 16 mer) for **FABP** modified 16-mer GGT at 25°C. (a) unmodified control, (b) FABP modified G1 (G*GT), (c) FABP modified G2 (GG*T), (d) Full duplex overlays of unmodified and FABP G1/G2-modified sequences..... 54

Figure 13: CD spectral overlays of primer elongation (9 mer to 16 mer) for **FAF** modified 16-mer GGT at 25°C. (a) unmodified control, (b) FAF modified G1 (G*GT), (c) FAF modified G2 (GG*T), (d) Full duplex overlays of unmodified and FAF G1/G2-modified sequences..... 55

Figure 14: (a) UV melting (T_m) of **FABP** modified 16-mer GGT with primer elongation (9 mer to 16 mer). (b) UV melting (T_m) of **FAF** modified 16-mer GGT with primer elongation (9 mer to 16 mer). Unmodified control is blue line, FABP/FAF-modified G1 (G*GT) is shown as red dash line, and FABP/FAF-modified G2 (GG*T) is shown as yellow dash line. 56

Figure 15: DSC heat capacity ΔC_p *vs* temperature (15-85 °C) for primer elongation (9 mer to 16 mer) of the (a) control, (b) **FABP** modified G1 (G*GT) and (c) **FABP** modified G2 (GG*T). (d) Full duplex overlays of unmodified and **FABP** G1/G2-modified sequences. FABP-G1 and FABP-G2 show smaller T_m and ΔH than the control/unmodified in full duplex exhibiting that S-conformer (FABP-G1, G*GT) might produce thermodynamic destabilization in full duplex. 57

Figure 16: ITC binding thermograms of duplex formation for (a) FABP- and (b) FAF- modified 16 mer GGT full duplexes at 25°C. 58

Figure 17: SPR sensorgrams of unmodified and FABP/FAF-modified TLS: (a) Unmodified and FABP-modified G1 (G*GT) and G2 (GG*T), (b) Unmodified and FAF-modified G1 (G*GT) and G2 (GG*T) 59

Figure 18: Normalized SPR sensorgrams of (a) FABP- and (b) FAF- modified 16 mer GGT sequences with primer elongation (9 mer to 16 mer). Slower dissociation shows better stability of the primer−template duplexes. 60

Figure 19: Normalized SPR sensorgrams of FABP-modified 16 mer GGT sequences with 9-mer and 10-mer primers, and progression of lesion conformational rigidity. (a) 9-mer and (b) 10-mer. Slower dissociation shows better stability of the primer−template duplexes. 61

Figure 20: *syn*- and *anti*-FABP-modified dG and their energy barrier from nucleoside MD (Modeling Dynamic) simulation. 62

Figure 21: FABP-dG DNA helices structures from MD simulation. (a) G_1 position, (b) G_2 position. B conformation is preferred at both the G1 and G2 positions..... 63

MANUSCRIPT II

Figure 1: (a) The views of FABP-conformational motifs: B-, and S-conformers. ABP-red, modified dG-blue, complementary C-orange. (b) dG-C8-FABP 94

Figure 2: (a) HPLC chromatogram of a reaction mixture of biotinylated-31-mer sequence (5'-Bio-CCTCTTCCCTCACCTCTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FABP** (N-Acetoxy-N-(trifluoroacetyl)-4'-fluoro-4-aminobiphenyl) and photodiode array UV spectra of unmodified, mono-adducts. (b) MALDI-TOF mass spectra of **FABP** modified biotinylated-31mer-TGGT: (a) 3'-enzyme digestion of **peak 1** at 0, 1, 4, 6, 7, 8 and 10min. (b) 3'-enzyme digestion of **peak 2** at 0, 1, 4, 6, 7, 8 and 10min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion 95

Figure 3: (a) Hairpin template–primer oligonucleotide constructs of 85-mer G_1 and 84-mer G_2 for Kf-exo$^-$; (b) Sensorgrams of **binary** complexes of Kf-exo$^-$ with 85-mer and 84-mer unmodified control, G_1-FABP and G_2-FABP modified sequences (1:1 binding fitted curves are overlaid as black lines); (c) K_a (K_a-modified / K_a-contol) and K_d (K_d-modified / K_d-contol) ratio, G_1 and G_2 represent the ratio of 85-mer FABP-G_1/85-mer-control and 84-mer FABP-G_2/84-mer-control 96

Figure 4: Denaturing gel (15%) profiles of 84-mer and 85-mer ligated oligonucleotides, 31-mer, 53-mer and 54-mer non-ligated oligonucleotides 97

Figure 5: Lineweaver – Burk model of the DNA Repair Reactions Catalyzed by the Kf-exo− at (a) 10-mer, G1-FABP and (b) 9-mer G2-FABP lesion sites with dCTP at steady-state. According to the model, the behavior of G1-FABP against with Kf–exo– showed as a competitive inhibitor, whereas the behavior of TG1G2*T played as non-competitive..... 98

Figure 6: dATP insertion efficiency results of (a) 8-mer to 9-mer and (b) 11-mer to 12-mer in control, TG1*G2T-FABP and TG1G2*T-FABP, respectively, in 10 min 99

Figure 7: Sensorgrams of ternary Kf-exo− complexed with (a) 85-mer control, (b) 85-mer TG1[FABP]G2T, (c) 84-mer control and (d) 84-mer TG1G2[FABP]T sequences in the presence of dNTPs (1:1 binding fitted curves are overlaid as black lines) 100

MANUSCRIPT III

Figure 1: Oligonucleotide products and byproducts studied in this work. a) the structures of adducts and byproducts from chemical syntheses and b) the alkyl adduct m1A, as an example, is repaired in the presence of the AlkB enzyme and necessary cofactors..... 129

Figure 2: Reverse phase HPLC chromatograms of 16mer DNA alkyl products and byproducts. The retention time of a modification-containing oligonucleotide is

labeled on top of the corresponding chromatogram. The modifications are: a) m6A); b) m1A; c) A; d) mixture of m3C + m3U; e) m3C; and f) m3U 130

Figure 3: High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing target modifications and byproducts. Data represent the -4 charge envelopes and the monoisotopic peak (all 12C, 14N, etc.) values are labeled above the first peak in each peak envelope. a) The oligonucleotide (containing m6A) generated from initial m1A synthesis; b) m1A; c) m3C; and d) m3U 131

Figure 4: Predicted fragmentation pattern from collision-induced-dissociation of the 16mer oligonucleotide products from the m1A and m3C syntheses. X denotes the modified nucleotides 132

Figure 5: Time-course MALDI-TOF analyses of SVP digestion products of the 16mer product oligonucleotide (containing m6A) generated from m1A synthesis. Theoretical masses are listed in the inset. The theoretical monoisotopic mass at 2745.5 is highlighted in red because it is the smallest digestion product containing the modification..... 133

Figure 6: HPLC profiles of the AlkB repair reactions on different alkyl substrates. The three chromatograms within each panel represent one set of repair reaction including the oligonucleotides of starting material, reaction mixture, and product. For example, in the panel containing chromatograms a, b, & c, a) represents the starting material m6A, b) represents the repair reaction of m6A by AlkB, and c) represents the pure product adenine synthesized separately. Chromatograms shown in a, b, & c relevant to m6A repair and in j, k, & l relevant to m3U repair were analyzed under reverse phase conditions.

Chromatograms shown in d, e, & f relevant to m1A repair and in g, h, & i relevant to m3C repair were analyzed under anion exchange conditions (see Experimental Procedures for detailed information). a) m6A; b) m6A+AlkB; c) A; d) m1A; e) m1A+AlkB; f) A; g) m3C; h) m3C+AlkB; i) C; j) m3U; k) m3U+AlkB; and l) U..... 134

LIST OF TABLES

MANUSCRIPT I

Table 1: Partial charges and atom types for FABP-dG. 65

Table 2: UV melting thermal and thermodynamic parameters of **FABP** modified 16-mer G1 (G*GT) and G2 (GG*T) duplexes. 66

Table 3: UV melting thermal and thermodynamic parameters of **FAF** modified 16-mer G1 (G*GT) and G2 (GG*T) duplexes. 67

Table 4: Thermal and Thermodynamic Parameters Derived from DSC 68

Table 5: Thermodynamic parameters C, N, K, ΔH and ΔS derived from the ITC 69

Table 6: Dissociation off-rate constants (k_d, x10^{-2}, s^{-1}) from SPR 70

MANUSCRIPT II

Table 1: Steady-state kinetic parameters for insertion of dCTP opposite unmodified and FABP−dG adduct with Kf-exo$^-$ 101

Table 2: SPR binding affinities, K_D (nM) of unmodified (TG_1G_2T) and FABP−dG adducts (TG_1[FABP]G_2T and TG_1G_2[FABP]T) with Kf-exo$^-$ (steady-state affinity analysis) in the binary and ternary systems. Association and dissociation rate constants, k_a and k_d, in binary system are listed 102

Table 3: Binding net stabilization energy of unmodified and FABP-adducts with Kf-exo$^-$ in binary system (1:1 binding) 103

MANUSCRIPT III

Table 1: Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. For m1A and m3C synthesis, the sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of the modified bases..... 135

MANUSCRIPT I

To be submitted to ***Chemical Research in Toxicology***

DNA Base Sequence Effects on Bulky Lesion-Induced Conformational Heterogeneity during Simulated Translesion Synthesis

Ang Cai[1], Katie A. Wilson[2], Stacey D. Wetmore[2] and Bongsup P. Cho[1*]

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

[2]Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta T1K 3M4, Canada

*Correspondence to Bongsup Cho:

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

Present address:

Department of Biomedical and Pharmaceutical Sciences,

University of Rhode Island

7 Greenhouse Road,

Kingston, RI, 02881, USA

ABSTRACT

DNA damage by bulky organic lesions can affect DNA replication and repair efficiencies. We previously reported a 3′-next flanking sequence effect [G*CT vs. G*CA; G*, FABP, N-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl; FAF, N-(2′-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene] on arylamine-DNA lesions in relation to DNA replication and repair[1, 2]. In the present study, we present an unusual sequence effect on a 16-mer fully paired DNA duplex 5′-CTTCTG$_1$G$_2$TCCTCATTC-3′, in which the same lesion modification at G$_1$ and G$_2$ revealed a dramatic difference in S (base-displaced stacked)/B (major groove binding B-type) conformational heterogeneity. We obtained conformational (^{19}F-NMR/CD), calorimetric (DSC and ITC), and binding (SPR) data for a set of four 16-mer modified templates (FABP vs. FAF and TG$_1$*G$_2$T vs. TG$_1$G$_2$*T) during primer elongation. Dynamic ^{19}F-NMR results indicate that FABP modification at G$_1$ and G$_2$ resulted in 67%:33% of B:S and 100% B population ratios, respectively, and the planar FAF in the same sequence contexts exhibited 25%:75% B:S and 83%:17% B:S, respectively. A significant increase was observed in S-conformation (0% to 17%) going from a twisted nonplanar FABP to a planar FAF. Calorimetric (DSC and ITC) and UV-melting data revealed that the B-conformer is a thermodynamic stabilizer in duplex settings, and the S-conformer is a destabilizer. The situation is reversed during the primer elongation process that involves ss/ds duplexes, i.e., S-conformation promotes lesion stacking at the ss/ds replication fork. This finding is in agreement with SPR binding results, which showed that lesion stacking increases the binding affinity of the complementary strands in the order of FABP > FAF and G*GT > GG*T. Surprisingly, our data revealed that the twisted biphenyl is a better stacker at

ss/ds junction than the coplanar FAF, possibly due to its ring flexibility that helps accommodate the replication fork structures. This finding is not the case for fully paired duplexes in which base-displaced stacking is enhanced by lesion planarity. The extent of primer dissociation rates across the lesion was influenced by the DNA sequences and the nature of the lesion (FAF vs. FABP). These results provide conformational insights into the important role of ss/ds sequence effects in modulating bulky lesion-induced translesion DNA synthesis.

INTRODUCTION

Abbreviation: ABP: N-(2′-deoxyguanosin-8-yl)-4-aminobiphenyl; AF: N-(2′-deoxyguanosin-8-yl)-2-aminofluorene; AAF: N-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene; ^{19}FNMR: fluorine nuclear magnetic resonance.

When a replication polymerase encounters a DNA lesion, DNA synthesis pauses or stops to allow the recruitment of bypass polymerases. This in turn will affect whether the synthesis becomes error free or error prone. Some bulky DNA lesions at the single-strand (ss)/double-strand (ds) replication fork can exist in multiple conformations, which can affect mutational outcomes. In summary, a single bulky adduct can induce conformational heterogeneity at the ss/ds replication fork, wherein a polymerase may process each of the conformations in a uniquely different way. A similar conformational argument can be made for DNA duplexes, which is the usual substrate for nucleotide excision repair. The latter removes various bulky lesions, such as arylamines, aflatoxins, and aristolochic acids.

Arylamines are carcinogens derived from cigarette smoke, incomplete combustion of diesel exhaust, and overcooked meat among others[3, 4]. These aromatic chemicals per se do not directly interact with DNA; however, they are metabolized in vivo by cytochrome P450 and subsequently activated by cellular enzymes, such as N-acetyltrasferase or sulfotransferase to form DNA adducts[5, 6]. For example, the well-known human bladder carcinogen 4-aminobiphenyl (ABP) is activated to produce C8-substituted dG-ABP as a major DNA lesion[7 8]. The bulkier 2-aminofluorene (AF) and N-2-acetylaminofluorene

(AAF) produce similar C8-dG adducts (Figure 1a). These bulky DNA adducts have been extensively studied and shown to adopt at least three major conformations in the duplex setting: major groove B-type (B), base-displaced stacked (S), or minor groove wedge (W) (Figure 1b), and the S/B/W-population ratios are influenced by neighboring sequence contexts[1,9,10]. In the S-conformation, the carcinogen stacks between the bases and disrupts base pairing in the double helix, whereas the B-type conformer places the carcinogen in the major groove of the DNA helix that maintains the traditional Watson–Crick base pairings. The W conformer was originally found in dA mismatch duplexes, and the hydrophobic lesion is well-placed in the narrow minor groove[11–14]. These adducts exhibit sequence-dependent S/B/W population ratios and consequently have different repair and mutational efficiencies[1, 15, 16,12–14, 17]. For example, the purine base at the 3′-flanking base favors S-conformer in the AF- and AAF-modified duplexes[14, 17 12, 18–20]. AAF at G_3 in the *Nar*I sequence (5′-$G_1G_2CG_3$CN-3′) induces a greater frequency of −2 frame shift mutations with the G_3-mutational frequency regulated by the nature of the N base[18, 21, 22]. Miller et al. reported a significant reduction in nucleotide incorporation rates at neighboring and downstream of the lesion site by Klenow fragment (Kf) exo$^-$[23]. We investigated the sequence effect of AF on nucleotide insertion efficiencies catalyzed by Kf–exo$^-$ on TG*A and CG*A translesion synthesis (TLS) systems and found the S conformer of CG*A thermodynamically favors insertion of A over C at the lesion site[19].

We previously reported an unusual 3′-flanking T effect on a random duplex sequence (5′-CCATCG*CNACC-3′; N=T or A), i.e., FABP modification at G exhibited a 40%:60% B:S in the G*CT context whereas 100% B in the G*CA sequence context. Though not

as dramatic, a similar trend was observed for the AF and AAF lesions[1]. In the present paper, another unusual ABP-induced sequence effect on a random 16-mer duplex sequence (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) was investigated. ABP-modification at G$_1$ exhibited a 67:33 ratio of B:S conformations, whereas the B conformation was adopted exclusively at the G$_2$ modification. We conducted a series of conformational [NMR, induced circular dichroism (ICD)], thermodynamic [UV-melting, differential scanning calorimetry (DSC)], and binding [surface plasmon resonance (SPR), isothermal titration calorimetry (ITC)] experiments for two 16-mer ss/ds DNA templates [(5′-CTTCTG$_1$*G$_2$TCCTCATTC-3′) and (5′-CTTCTG$_1$G$_2$*TCCTCATTC-3′); G$_1$*/G$_2$* =FABP or FAF (fluorinated ABP, AF)], which mimic a TLS (Figure 2). The experimental data showed that the B-conformer is a thermodynamic stabilizer and the S-conformer is a destabilizer in full duplex settings. However, the opposite was true for a simulated TLS that involved ss/ds sequences. Lesion stacking increased the binding affinity of the complementary strands in the order of FABP > FAF and G*GT > GG*T. The twisted nonplanar ABP was a better stacker than the planar AF in ss/ds systems, which were in contrast to the duplexes. The present study provided valuable insights into the important distinction between ss/ds and full ds sequence contexts and flanking base sequences and lesion stacking in modulating bulky lesion-induced conformational heterogeneity.

MATERIALS AND METHODS

Caution: 4-Aminobiphenyl derivatives are known human bladder carcinogens. Aminofluorene analogues are animal carcinogens; therefore, must be handled with caution.

Oligodeoxynucleotides (oligo, 10 μmol scale) in desalted form were obtained from Integrated DNA Technologies (Iowa) and were purified by using a reverse phase HPLC. The HPLC system comprised a Hitachi EZChrom Elite HPLC with a L-2450 diode array detector and a Phenomenex Luna C18 column (150 x 10 mm, 5.0 μm) (Phenomenex, Torrance, CA). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Preparation of Lesion-Modified Oligodeoxynucleotides

FABP-modified 16-mer templates (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) were prepared according to the published procedures[22]. For each lesion, we obtained two mono-adducts [d(5′-CTTCTG$_1$*G$_2$TCCTCATTC-3′)] and [d(5′-CTTCTG$_1$G$_2$*TCCTCATTC-3′)] (G*=FABP) and one di-adduct [d(5′-CTTCTG$_1$*G$_2$*TCCTCATTC-3′)]. Briefly, 5–10 mg of *N*-acetoxy-*N*-trifuoroacetyl-7-fluoro-4-aminobiphenyl dissolved in 500 μL absolute ethanol was added to 10 mL sodium citrate buffer (10 mM, pH 6.0) containing approximately 200 ods of unmodified 16-mer oligo and incubated in a 37 °C water bath shaker for 1 h. Unreacted carcinogens were removed by extraction with anhydrous ether. An acetonitrile gradient of 14%–25% in triethylammonium acetate (TEAA) buffer (100 mM, pH 7.0) for 30 min was used to separate the two mono-adduct strands from di-adduct and unreacted 16-mer oligos. Figure 3 shows the HPLC profile and photodiode array UV spectra of FABP-modified 16-mers. The combined collection of the two

mono-adduct oligos was lyophilized and stored in −80°C. FAF-modified 16-mer oligos were prepared similarly according to published procedures and the resulting HPLC profile is shown in Figure 4.

MALDI-TOF Mass Spectrometry Characterization of Mono-adducts

The FABP- and FAF-modified mono-adducted 16-mer oligos above were digested by 3′–5′ exonuclease enzyme digestion and characterized by MALDI-TOF mass spectrometry following published procedures[24]. Briefly, 1 μL (0.1 unit) of snake venom phosphodiesterase (SVP) was added to a solution of 1 μL of modified oligo (~100–200 pmol), 6 μL of ammonium citrate (100 mM, pH 9.4), and 6 μL of water for 3′–5′digestion. In each case, 1 μL of the reaction mixture was withdrawn at regular time intervals and was quenched by mixing with 1 μL of matrix (3-hydroxypicolinic acid and diammonium hydrogen citrate in a 1:1 ratio). All MALDI spectra were obtained using Shimadzu Axima Performance MALDI-TOF mass spectrometer equipped with a 70 Hz nitrogen laser in positive ion reflectron mode[12].

UV Melting

UV thermal melting experiments were performed on a Cary100 Bio UV-Vis spectrophotometer (Santa Clara, CA) equipped with a 6 × 6 multi-cell chamber (1.0 cm path length) and connected with a built-in Peltier temperature controller. Samples were prepared in the buffer (containing 0.2M NaCl, 10 mM sodium phosphate and 0.2 mM EDTA at pH 7.0) in the range of 0.8 μM to 5.6 μM. Samples were scanned forward/reverse five times during melting experiments. Thermodynamic parameters

were calculated using the program MELTWIN version 3.5 according to previous reports[22].

DSC

DSC calorimetry measurements were performed on a Nano-DSC from TA Instrument (Lindon, UT, USA) according to previous reports[25]. Solutions at 0.1 mM containing unmodified or modified templates with various primers were dissolved in a pH 7.0 buffer (0.1 M NaCl, 20 mM sodium phosphate), and the solutions were incubated at 95 °C for 5 mins and cooled at room temperature for 3 h to ensure duplex formation. After 15 min degassing, samples were loaded into the instrument cells. A buffer scan was used as control and then subtracted from the sample scan and normalized for heating rate. All samples were scanned against the control buffer from 15 °C to 85 °C at a rate of 1 °C/min by measuring five repetitions. Raw data were collected as microwatts versus temperature. ΔH is an integration of ΔCp^{ex} over temperature T. T_m was the temperature at half the peak area. Thermodynamic parameters (ΔG and ΔS) were determined according to previously described procedures[26, 27].

CD Spectra

Circular dichroism (CD) experiments were conducted on a Jasco J-810 spectropolarimeter (Easton, MD) equipped with a Peltier temperature controller. Typically, a modified template (20 μM) was annealed with an equimolar amount of a primer in 400 μL buffer (containing 0.2M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA, pH 7.0). The samples were hybridized by heating to 95 °C for 5 min and slowly

cooled to room temperature for 3 h to ensure duplex formation. Spectra were acquired every 0.2 nm from 200 nm to 400 nm at a rate of 50 nm/min for five repetitions and smoothed with 25-point adaptive algorithms provided by Jasco[14].

Dynamic ^{19}F -NMR

Approximately 0.1 mM of full and ss/ds duplex samples (Figure 2) were dissolved in 250 μL of NMR buffer (pH 7.0) containing 10% D2O/90% H2O, 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA. Solutions were filtered into a Shigemi tube through a 0.22 μm membrane filter. All ^{1}H and ^{19}F NMR experiments were performed on a Varian NMR spectrometer with a HFC probe operating at 500.0 and 476.5 MHz, respectively[9, 16, 28]. Imino proton spectra was obtained at different temperature points (5 °C to 50 °C) using phase-sensitive jump-return sequences and referenced to DSS. ^{19}F-NMR spectra were recorded with increment of 5 °C–10 °C (from 5 °C to 60 °C) in the ^{1}H-decoupled mode and referenced to C6F6 in C6D6 at −164.9 ppm. Computer line shape simulations were processed as reported previously using MestReNova 10.0.2 (Mestrelab Research, Spain)[9, 29].

SPR

SPR binding measurements were conducted using a Biacore T200 instrument (GE Healthcare, Marlborough, MA). Two isomeric 16-mer biotinylated DNA template strands (5′-biotin-CTTCTG$_1$*G$_2$TCCTCATTC-3′ and 5′-biotin-CTTCTG$_1$G$_2$*TCCTCATTC-3′; G*= FABP or FAF) were each modified with FABP or FAF[2, 20, 25, 30]. HPLC gradient conditions for biotinylated FABP-modified oligos were

16%–25% gradient on acetonitrile for 55 min in TEAA buffer (100 mM, pH 7.0) with a flow rate of 3.0 mL/min (Figure 5a). The biotin-FABP-modified sequences were characterized by MALDI-TOF-MS using 3–5′exonuclease enzyme digestion (Figure 5b–c).

Carboxymethylated CM 5 chip was activated by amine coupling kit according to published procedures[20, 25, 30]. Streptavidin (SA, 50 µg/mL dissolved in sodium acetate buffer, pH 4.5) was coated on flow cells 2, 3, and 4 after the injection of EDC/NHS mixture for 7 min. Flow cell 1 was left as blank. Ethanolamine (1 M) was injected over the surface to block unreacted esters. The unbound SA was removed by five pulses of 50 mM NaOH and three pulses of running buffer. The surface was stabilized for 20–30 min in running buffer before DNA coating. Flow cells 2, 3, and 4 were coated around 300 RU of unmodified DNA sequence (–G_1G_2T–), FABP modified (–G_1*G_2T–) and (–G_1G_2*T–) DNA sequences by the manual control mode, respectively. Samples of 9- to 16-mer primer sequence were prepared separately in HBS-P+ buffer (10 mM Hepes, 150 mM NaCl, 0.05% surfactant P20 at pH 7.4) and injected over the chip surface with 100 s contact time and 360 s dissociation time at a flow rate of 15 µL/min at 25 °C.

Due to distinct molecule weights, the primers (9- to 16-mer) were used at different concentrations to gain steady-state associations: 300 nM (9-mer), 500 nM (10-mer), 500 nM (11-mer), and 750 nM (12-, 13-, and 16mer). After the system reached a steady state, excess primers were washed off by the HBS-P+ buffer, and the system was regenerated by adding NaOH (50 nM, 30 s). We were unable to measure the usual K_D (k_d/k_a) kinetics

of duplex formation for each elongation. Instead, we used dissociation rates (k_d), which are concentration-independent, to estimate binding strengths for complementary strands.

Dissociation constants (k_d) were analyzed in k_d-alone fitting mode by Scrubber software (version 2.0, Myszka and collaborators, BioLogic Software). The goodness of the fit was determined from the residual standard deviation.

ITC

VP-ITC microcalorimeter from MicroCal Inc. (Northampton, MA) was performed to measure full duplex formation of unmodified and modified templates (2.5 μM) with their complementary (25 μM) sequences prepared in 10 mM phosphate buffer (pH 7.0), containing 0.2 M NaCl and 0.2 mM EDTA. The samples were degassed thoroughly under vacuum using ThermoVac (MicroCal, Northampton, MA) for 20 mins. ITC experiments consisted of 25 consecutive injections and each injection involves 10 μL of complementary sequence with injection time 10 s and 240 s intervals between each injection with 307 rpm stirring at 25 °C. The integrated and normalized areas of raw data were plotted as ΔH of injectant against the molar ratio. The titration data were fitted using a single-site binding model provided in ORIGIN7.0 software supplied by the manufacturer. ΔH, ΔS, K, and N (reaction stoichiometry) of 16-mer full duplex formation were obtained[31, 32].

Modeling Dynamic Simulation

The initial undamaged helix (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) was generated using the NAB module of AMBER 14[33]. The helix was modified to position the FABP-dG adduct at G_1 or G_2 such that the lesion was paired opposite dC. At each position, three orientations of the lesion site were considered to generate the B, S, and W adducted DNA conformations. Resulting duplexes were neutralized with Na^+ ions and solvated in a TIP3P octahedral water box with sides 8.0 Å from the solute. All natural nucleotides were modeled with the AMBER ff14SB force field[34, 35]. FABP-dG parameters were assigned according to the GAFF[36] and AMBER ff14SB force fields using ANTECHAMBER 1.4 (Table 1)[37], and partial charges were developed using RESP charge fitting from a HF/6-31G(d) calculation by using R.E.D.v.III.4 program (Table 1)[38].

Each system was initially minimized with a 500 kcal/mol Å^2 constraint on the DNA for 500 steps of steepest descent and 500 steps of conjugate gradient minimization. Subsequently, 1,000 steps of unrestrained steepest descent and 1,500 steps of unrestrained conjugate gradient minimization were performed on the entire system. Each system was then heated over 20 ps at constant volume to 300 K. Heating was performed using a Langevin thermostat (γ = 1.0) and a 10 kcal/mol Å^2 constraint on the DNA. Subsequently, each system was simulated for 100 ns at 300 K using a 2 fs time step. To prevent unnatural unravelling of the DNA helices throughout the MD simulations, a 25 kcal/mol Å^2 harmonic distance restraint was used to maintain the heavy atom–heavy atom distance in the Watson–Crick hydrogen bonds for the terminal G:C base pairs between 2.0 Å and 4.0 Å. During simulations, the SHAKE algorithm was used to constrain covalent bonds involving hydrogen, and the periodic boundary

condition was applied. The particle mesh Ewald algorithm was used for long-range electrostatic interactions, with a nonbonded cutoff of 8.0 Å. All simulations and analysis were completed with Amber 14 using the pmemd and cpptraj modules, respectively[33].

RESULTS

Experimental DNA Sequences. Chemical structures of ABP and AF and their fluorine analogues, FABP and FAF, are shown in Figure 1a. We previously observed that FABP-modified 12-mer DNA duplex 5′-CTTCTG$_1$G$_2$TCCTC-3′ at G$_1$ and G$_2$ resulted in 60%:40% of B:S and 100% B population ratios, respectively. In the present study, the original 12-mer was extended to 16-mer (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) by adding "–ATTC–" to the 3'-end to improve thermal stability for the $n - 1$ and n TLS duplexes. NMR results indicate that the fully paired FABP modified 16-mer duplexes exhibited 67%:33% of B:S and 100% B population ratios for G$_1$ and G$_2$ modifications, which are in good agreement with those observed above for the original FABP-modified 12-mer duplex.

As shown in Figure 2, the two 16-mer template strands (TG$_1$*G$_2$T- or TG$_1$G$_2$*T-: G* = FABP or FAF) were each annealed with various lengths of complementary strands (9- to 16-mer) to generate two discrete model TLS systems ($n - 2$, $n - 1$, n, $n + 1$, $n + 2$, $n + 5$ for -G$_1$*G$_2$T- and $n - 1$, n, $n + 1$, $n + 2$, $n + 3$, $n + 6$ for -G$_1$G$_2$*T-, respectively; n is the lesion site). In the present study, we used these TLS model systems consistently for ^{19}F-NMR/CD, DSC, UV melting, and SPR experiments. Our strategy is to investigate the effects of the S/B conformational heterogeneity on thermodynamic and duplex

binding parameters. The results will help us understand the uniqueness of flanking base sequence in the TGGT context in promoting lesion-induced conformational heterogeneity.

MALDI-TOF Characterization. FABP- and FAF-modified 16-mer oligos were characterized by exonuclease enzyme digestion followed by MALDI-TOF mass spectrometry, according to published procedures[12, 24, 25]. Figure 6a–b shows the MALDI-TOF MS spectra of the 3′-5′ SVP exonuclease digestions of peaks 1 and 2 from FABP-HPLC profile at different time points (0, 0.5, 2, and 5 min). Briefly, the m/z of 4,958 at 0 min represents the control mass to charge ratio of the modified template without SVP. Within 0.5 min of 3′-5′ exonuclease digestion (Figure 6a), the lower masses showed up corresponding to the 15-mer to 6-mer fragments. The fragment at m/z = 1,943 (theoretical 1945), which was the most detectable mass between 2 min to 5 min digestion, was assigned to a 6-mer containing the FABP lesion. These results confirmed that peak 1 from the HPLC profile was the G_1*-FABP-modified sequence (5′-CTTCTG$_1$[FABP] G_2TCCTCATTC-3′, G_1*G_2T). Peak 2 from HPLC profile was characterized similarly. Figure 6b shows the control molecular ion m/z of 4,961 at 0 min. The m/z 2272 ion peak in the 0.5–5 min of digestion corresponded to the 7-mer fragment containing FABP-modified guanine. These results confirm peak 2 as the G2*-FABP modified sequence (5′-CTTCTG$_1$G$_2$[FABP] TCCTCATTC-3′, G_1G_2*T).

The FAF-modified 16-mer templates have been characterized similarly. Figure 7a–b shows the MALDI-TOF MS spectra of the 3′–5′ SVP exonuclease digestion of the FAF-

modified 16-mer oligo (peaks 1 and 2) at different time intervals. The *m/z* of 4972 (theoretical 4971) in peak 1 at 0 min (before digestion, as a control) indicated the 16mer DNA sequence containing a FAF-modified DNA base. From 0 min to 3 min, with the disappearance of the original oligonucleotide (4972) the digestion process has been observed when new lower masses showed up corresponding to the smaller fragments from 3′ cleavage (e.g., 4682, 4377, and 4073). After digestion for 3 min, the digestion stopped at *m/z* = 1955 (theoretical 1955), which represented the FAF-modified 6-mer fragment and confirmed peak 1 as G_1*-FAF modified sequence (5′-CTTCTG$_1$[FAF] G$_2$TCCTCATTC-3′, G_1*G_2T). Following the same procedure, we found the SVP digestion of peak 2 stalled at *m/z* = 2284 (theoretical 2284) after 3 min and confirmed peak 2 as G_2*-FAF modified sequence (5′-CTTCTG$_1$G$_2$[FAF] TCCTCATTC-3′, G_1G_2*T). Taken together, the SVP enzymatic digestion from the 3′-end is slowed down significantly at lesion sites of 6- and 7-mer representing G_1*G_2T and G_1G_2*T, respectively.

Dynamic ^{19}F-NMR Spectroscopy

Dynamic ^{19}F-NMR experiments were performed to measure the extent of conformational heterogeneity during primer elongation. Figures 8 and 9 show the ^{19}F-NMR spectra of the FABP- and FAF-modified duplex series as function of temperature (5 °C–60 °C). The same sequence systems were used for CD, DSC, UV-melting, and SPR experiments for comparison. Briefly, FABP-G_1*-16-mer duplex showed a mixture of S (33%) and B (67%) conformations at 25 °C (highlighted in Figure 8a). By contrast,

the corresponding FABP-G_2*-duplex exclusively revealed B-conformer (100%) with no sign of conformational heterogeneity.

FABP-G_1*G_2T

In Figure 8a, no indication of conformational heterogeneity at $n - 2$ (9-mer) and $n - 1$ (10-mer) was found, whereas the 3′-end of the primer did not reach the lesion site. A slight heterogeneity was observed at n (11-mer); however, but interestingly it disappeared at $n + 1$. At $n + 2$ (16/13-mer), however, we observed complex ^{19}F signals indicating the presence of multiple conformations. These ^{19}F signals were coalesced into a sharp single peak around −116 ppm upon around and above 50 °C; this result indicates duplex melting. The conformational pattern was steadied at fully paired 16-mer duplex ($n + 5$), which showed a ratio of 33% S and 67% B populations at 25 °C (highlighted in Figure 8a).

The ^{19}F-signals were assigned according to the general protocols based on relative chemical shifts, chemical exchange patterns, and H/D isotope effects[2]. We have investigated a number of ^{19}F-NMR spectra of arylamine-modified duplexes in various sequence settings and found that the fluorine of the S-type conformer resonates always upfield relative to that of the external binding B-type conformer and they are in slow exchange and eventually coalesce into a denatured ss. In addition, the ^{19}F tag in the B-conformation is fully exposed to the solvent in the major groove area, whereas that of the S-conformer is shielded from van der Waals interactions with neighboring base pairs. In addition, the fully exposed ^{19}F tag of the major groove binding B-conformer always

exhibits significant deuterium isotope effect (0.18 ppm) going from 10% to 100% D_2O (Figure 10 a–b). This result is in contrast to the buried ^{19}F of the stacked conformer, which shows no isotope effect (0.08 ppm) (Figure 10 a).

FABP-G_1G_2*T

G_2-modfied series shown in Figure 8b is quite different from the G_1-modified above. Some minor heterogeneities are observed on $n - 1$ (16/9-mer) and $n + 3$ (16/12-mer); in general, all duplex sequences exhibited one major signal around −117.5 ppm. This signal can be assigned as the B-like conformer because of its close proximity to the coalescent signal (ss) and also by comparison with the G_1 modified duplex above.

FAF-G_1*G_2T and FAF-G_1G_2*T

Figure 9a–b show the dynamic ^{19}F-NMR spectra for the FAF-modified series. The results indicate both FAF-G_1*G_2T and FAF-G_1G_2*T duplexes adopted more S conformation than the FABP series. FAF-G_1*G_2T had approximately 42% more S− than the FABP counterpart G_1*G_2T (75% S & 25% B) in the full duplex, whereas the S− increased from 0% to 17% in the FAF/G_1G_2*T. These results reflected the planar nature of FAF relative to FABP preferring to intercalate/stack. Moreover, the B-conformer exhibit significant deuterium isotope effect (~ 0.18 ppm) going from 10% to 100% D_2O whereas the stacked conformer shows little isotope effect (0.04 ppm) (Figure 11a–b).

ICD

Figure 12 shows CD overlays of unmodified and FABP-modified 16-mer templates with increasing primer length. The S curve consisted of +@270 nm, and the B-form DNA duplex consisted of −@250 nm. Given the conformational heterogeneity, FABP-modified TLS duplexes exhibited lesion ICD between 290 nm and 360 nm[17, 22]. This result is in contrast to the unmodified control series, which showed no such ICD. The ICD gradually increases along with the primer elongation from 9-mer to 16-mer. In the fully paired 16-mer duplex, *T* agrees with the above ^{19}F-NMR data, which showed a B/S-conformeric TG_1*G_2T and exclusively B-conformeric G_1G_2*T on fully paired 16-mer duplexes. The predominant B-conformer of G_1G_2*T highly affects the bending of DNA helix around the adduct site. G_1 and G_2 FABP-modified duplexes displayed slight blue shifts (~3 nm) relative to the unmodified controls, indicating adduct-induced DNA bending[24]. Similar results were obtained for FAF-modified series (Figure 13).

UV Melting Experiments

The UV-melting profiles of all the ss/ds duplex models were measured at the 0.8−4.0 μM range for FABP and FAF samples. The UV monophasic and sigmoidal melting curves between lnCt and T_m^{-1} ($R2 > 0.9$) confirmed that the duplexes adopted a typical helix-coil transitions.[9] Figure 14a–b shows the plots of melting (T_m) for modified duplexes as a function of primer elongation for TLS systems.

Figure 14a–b shows the melting results of FABP and FAF lesions at the G_1*G_2T (red) and G_1G_2*T (yellow) systems (red dash lines and yellow dash lines, respectively) relative to the unmodified controls (blue solid lines). Thermal and thermodynamic

parameters calculated from UV melting curves are summarized in Tables 2 and 3. For all the modified and unmodified duplexes, T_m increases with the elongation of primers. However, modified systems showed stalls around the lesion sites.

Lesion effect in the G_1*G_2T system was minimal at one and two bases before the lesion sites (ΔT_m of 0.35 °C and 0.67 °C for 9-mer and 10-mer, respectively). This finding is in contrast to that of the G_1G_2*T system, which showed minimal lesion effect at those sites (9-mer ΔT_m of 1.45 °C). However, the effect was significant when the primer elongation was opposite the lesion [11-mer and 10-mer for G_1*G_2T (ΔT_m= -2.54 °C) and G_1G_2*T (ΔT_m= -5.21 °C), respectively]. In full duplex (16-mer), G_1*G_2T showed greater negative thermodynamic effect to G_1G_2*T (ΔT_m of −7.68 °C to −6.29 °C), which can be due to the S-conformer (see ^{19}F NMR results above).

DSC

DSC experiments were performed for the FABP-modified TLS models (TLS/TG_1*G_2T vs. TLS/TG_1G_2*T). Figure 15a–d shows DSC thermogram overlays of excess heat capacity C_p^{ex} vs. temperature. T_m is the maximum point of the curves, and transition enthalpy (ΔH) is the area under the curves. DSC results are superior to UV-melting due to its independent concentration, thereby providing reliable thermodynamic parameters. DSC results are summarized in Table 4[25, 39].

Control TLS

Figure 15a shows the DSC thermograms for the control TLS system. For the natural DNA, T_m and ΔH values increased progressively with increasing primer length from 9-mer (T_m = 42.9 °C, ΔH = −71.4 kcal/mol) to 13 mer (T_m = 59.8 °C, ΔH = -87.2 kcal/mol) and increased significantly at full duplex (16 mer, T_m = 63.1 °C, ΔH = −98.5 kcal/mol).

G_1[FABP]G_2T Duplex

In FABP-G_1 profiles (Figure 15b), the 9-mer primer, which is two bases (n − 2) before the lesion (n), showed a minimal lesion effect, i.e., T_m 43.3 °C is compared with the control duplex 42.9 °C. However, 10-mer (n − 1) to 16-mer (n + 5) primers were stalled, i.e., their T_m values were clustered between 51 °C and 53 °C. The 16-mer full duplex showed a significant increase in enthalpy ΔH, indicating a full duplex formation.

G_1G_2[FABP]T Duplex

DSC patterns for the TG_1G_2*T system (Figure 15c) were similar to those of the TG_1*G_2T. Curves for 10-(n) to 13-(n + 3) mer were clustered between Tm 45 °C and 53 °C, which is lower than that of the TG_1*G_2T counterpart. The increase in ΔH was not as incremental as that of T_m. In general, the exclusively B-conformeric 16-mer TG_1G_2*T full duplex showed greater T_m and ΔH than the G_1 counterpart that contains a mixture of B and S conformers (see ^{19}F-NMR above). This finding suggests that B-conformation can be a thermodynamic destabilizer during primer elongation and yet S-conformer might produce thermodynamic destabilization in full duplex. TG_1*G_2T and TG_1G_2*T duplexes in TLS settings show consistently smaller T_m and ΔH compared with the unmodified controls (Figure 15d).

ITC for Duplex Formation

ITC is a direct solution technique and provides the nature of duplex formation through a sigmoidal curve and the respective thermodynamic parameters (ΔH, ΔG, ΔS) through plots of heat capacity (C_p) vs. molar ratios. This result indicates that the lesion destabilizes the duplex formation.

Figure 16 shows ITC thermograms of duplex formation for FABP and FAF-modified GGT sequences at 25 °C. In all cases, unmodified controls (black) showed well-defined sigmoidal curves with sharp changes in heat capacity at starting and ending points. The results indicate a typical cooperative 1:1 binding process (e.g., molar ratio of N = ~1) expected for formation of Watson–Crick-based DNA duplexes. By contrast, the ITC curves for modified full duplexes were not well defined in terms of stoichiometry and overall shape, probably due to heterogeneity. Table 5 summarizes the ITC-derived thermodynamic parameters ΔH, ΔS, K, and N for unmodified and modified sequences, for example, the binding constant order is GGT (49.9 M^{-1}) > G_1*G_2T (18.7 M^{-1}) > G_1G_2*T (5.9 M^{-1}) in FABP case. In general, the ΔH and the binding constant of the modified sequences are consistently lower than that of the controls, and the FABP shows more variances in shapes than the FAF counterpart.

SPR on Duplex Formation

SPR experiments were conducted to probe the effect of conformational heterogeneity on duplex formation and dissociation. The results were obtained in buffer-only environment without a polymerase. With different concentrations used for each complementary strand, we measured dissociation rates (k_d) to evaluate their binding strengths (see Results).

Figure 17 shows the entirety of SPR results, and Table 6 lists the respective dissociation rate constants. Figure 18 shows the overlay of sensograms for FABP and FAF lesions on the unmodified (green) and the G_1 (blue)/G_2 (purple) sequences with increasing length of complementary strands. As expected, in each complementary sequence, SPR data show gradual reduction in k_d values with elongation in unmodified and modified sequences.

By contrast, rapid duplex formation took place for all duplexes; different dissociation rates were observed for modified duplexes. In general, slower dissociation indicates greater duplex binding with thermodynamic stability. This appears to be generally true with post-lesion duplexes (16/11- to 16/16-mer). The 16/9-mer duplexes, which represents the prelesion $n - 2$ and $n - 1$ site for G_1 and G_2, respectively, showed generally faster dissociation (Figure 19a, Table 6): G_1G_2[FABP]T (7.1 s^{-1}) < G_1[FABP]G_2T (12 s^{-1}) < unmodified (16.5 s^{-1}). As for the 16/10-mer (Figure 19b), however, dissociation was dramatically slower for G_1*G_2T (0.343 s^{-1}) than the unmodified (0.515 s^{-1}) and G_1G_2*T (2.37 s^{-1}). These results are consistent with duplex stability in the order of TG_1*G_2T > unmodified > TG_1G_2*T. The 16/9-mer is one base

before ($n-1$) the G_1G_2*T, and 10/16mer is one base ($n-1$) before the G_1*G_2T duplex. The results suggest that the modified $n-1$ duplexes may form a stable S-conformer at the ss/ds junction, resulting in slow dissociation. Similar SPR patterns were observed for the FAF lesion (Figure 18b); however, general dissociation rates were slightly faster compared to the FABP lesion. This finding indicates that FABP is a better stacker in the ss/ds duplex setting.

Modeling Dynamic Simulation

With increasing size, we can pinpoint the effects of different components of the system by using a variety of different models.

FABP-dG Nucleoside: The first model we considered was the nucleoside model. The FABP-dG nucleoside has greater conformational flexibility in the *syn*-conformation (S-conformer) compared to the *anti*-conformation (B-conformation). The bulky moiety prefers a twisted conformation about the linkage between the two rings (Figure 20).

FABP-dG DNA Duplex: Figure 21 shows that the DNA helix is more stable with FABP positioned at G_2 compared with that at G_1. B conformation is preferred at G_1 and G_2 positions. At G_1, the W conformation is also stable because of interactions with the 5′ backbone. Steric prevents such an interaction at G_2. The FABP adduct does not affect the structure of the flanking base pairs.

ss/ds DNA Junction: ss/ds DNA junction is difficult to model because of the flexibility of ssDNA. Greater conformational flexibility is observed with FABP-dG is positioned at G_1. Similar to the DNA model, intercalation of the bulky moiety does not increase

the stacking energetics. However, energies were hard to compare due to the variation in structure of the ssDNA.

Summary: Twisted conformation of the bulky moiety destabilizes the S conformation. In general, the B conformation is most stable regardless of model or stand sequence. Formation of a stable W conformation is dictated by the flanking bases. Flanking bases affect the conformation adopted by FABP-dG.

DISCUSSION

Unusual Flanking Sequence Effects on the TG_1G_2T Sequence

In this study, we present a striking flanking sequence effect on a 16-mer fully paired DNA duplex 5′-CTTCTG_1G_2TCCTCATTC-3′, in which lesion modification at G_1 (TG_1*G_2T sequence) and G_2 (TG_1G_2*T sequence) yielded a dramatic difference in conformational heterogeneity. Dynamic ^{19}FNMR results indicate that FABP modification at G_1 and G_2 resulted in 67%:33% of B:S and 100% B population ratios, respectively. Modification with the planar FAF in the same sequence contexts exhibited greater S-conformer population: 25%:75% B:S and 83:17% B:S for the G_1 and G_2 position, respectively. This represents a significant increase in S-conformation from 33% to 75% and from 0% to 17% going from FABP to FAF on the G_1*G_2T and G_1G_2*T sequence context, respectively.

This example is analogous to our recent reporting on a random 11-mer DNA duplex sequences 5′-CCATCG*C<u>N</u>ACC-3′ (*N* = T or A), in which FABP modification exhibited a 40%:60% ratio of B:S and 100% B and FAF modification resulted in 10%:90%

and 34%:66% in the G*CT and G*CA sequence context, respectively[1, 2]. These 3′-next flanking sequence effects influence primer kinetics, DNA replication, and repair. Lesion stacking was primarily responsible for the greater population of the stacked S-conformer compared with that of G*CA counterparts; the S conformation decreased the binding affinity of complementary strands. These adduct-induced conformational heterogeneities are caused by 3′-next flanking sequence effect, which is different from the G_1G_2T in the present study, which represents the flanking sequence effects.

Our calorimetric results showed a greater thermodynamic stability for the B-conformeric G_2 over the S/B-conformeric G_1 in the full duplex, which indicates that the B-conformer is a thermodynamic stabilizer in fully paired duplex DNA. This finding is consistent with our modeling data, which show that DNA duplex is more stable with FABP positioned at G_2 compared with G_1, and B-conformer is preferred at G_1 and G_2 positions. The greater S-conformer population of FAF over FABP at duplex DNA is probably due to their aromatic coplanarity, i.e., the single-carbon atom methylene bridging between two aromatic rings in FAF restricts the twisting of the biphenyl and enhances the stacking capability. As shown in ITC curves (Figure 16), lesion modification has reduced enthalpic energies (~13% and ~17% average for FABP- and FAF-G_1/G_2, respectively) during duplex formation, as expected. The lesion destabilizes the duplex formation. Correlating the ITC data with conformation heterogeneity is difficult; hence, the difference in ring planarity had a major effect on the shape of binding isotherms (C), i.e., FABP has greater sequence differences compared with FAF. Modeling data also indicate that the low-energy S conformation at G_1 and G_2 is not

much different; however, a unique low-energy W-conformer at G_1 position is observed due to the interactions with the 5′ backbone. By contrast, steric clash prevents such an interaction at G_2. These results may explain the sequence-dependent isotherm fluctuation in the G_1 over G_2.

Thermodynamic Effect on Lesion Heterogeneity during Primer Elongation

We studied six discrete TLS model duplexes prepared by annealing the 16-mer TG_1*G_2T and TG_1G_2*T (G* = FABP or FAF; Figure 2) modified template with appropriate primer strands. These experiments were performed in buffer solutions without a polymerase. The main objective of these experiments is to study the effect of the lesion-induced S/B-heterogeneity on their thermal and thermodynamic stabilities during the primer elongation that mimics a TLS.

The thermodynamic results show a systemic increase in T_m and ΔH with primer elongation from 9-mer to 16-mer in unmodified control duplexes. However, FABP modification has generally lowered T_m and ΔH, as well as changed their calorimetric curve patterns for G_1*G_2T and G_1G_2*T systems. As mentioned above, for full duplexes, the exclusively B-conformeric G_2* was more thermally stable than G_1*. However, this fining is not observed during the primer elongation from 10-mer to 13-mer, in which average T_ms of G_1G_2*T is lower that of the G_1*G_2T. These results imply that the B-conformer is a thermodynamic stabilizer in the duplex setting but is a destabilizer during primer elongation. In other words, S-conformer may play as a thermodynamic stabilizer

in the TLS (as presented below), although modeling data indicate that the intercalation of the bulky moiety does not increase the stacking energetics.

The Impact of Lesion Heterogeneity on ss/ds Duplex Formation

We performed SPR experiments to investigate the effect of lesion heterogeneity on strand binding that simulates a TLS. SPR is an excellent real-time technique that allows monitoring for formation and dissociation of DNA duplexes, DNA–RNA hybridization, and triplex formation. Figure 18 shows an overlay of SPR sensorgrams as a function of various complementary strands (9-mer to 16-mer). Sensorgrams from SPR for each primer elongation and dissociation rate k_d (Table 6) was used to evaluate lesion and sequence effects. This finding is due to the different concentrations of each primer elongation; thus, conducting a global binding analysis is difficult for KD affinity values (see Results). Instead, we utilize k_d dissociation rates (Table 6) to examine lesion effects.

The binding strength for each primer elongation was increased with increasing extent of Watson–Crick hydrogen bonds. As a result, dissociation rates are steadily reduced with primer elongation. Not surprisingly, binding strength for the unmodified templates was generally greater than that of the modified ones throughout the TLS process. Initially, the effect was relatively minimal at prelesion sites ($n - 1$ and $n - 2$ for G_1* and G_2*, respectively) (Figure 19a, 9-mer). According to NMR results, the FABP-G_1* lesion with the 9-mer (16/9-mer) is likely to exist in a B/S-mixture that can be readily accommodated in the flexible $n - 2$ environment. As for G_2*, this represents $n - 1$, in which the mostly B-conformer can exist at the 16-/9-mer ss/ds junction. Small, modified

16/9-mer sequences appeared to be slightly more stabilized than the controls. However, the effect of lesion modification on dissociation was clearly seen around the lesion sites. The 16/10-mer corresponds to $n - 1$ and n for G_1* and G_2*, respectively. The effect on dissociation was in the order of $n - 1$ G_1* > control > n G_2*. The most significant lesion effect of G_1* at $n - 1$ is probably due to the presence of S-conformer population that can stabilize the ss/ds replication fork (see above). This situation is illustrated in Figure 19b. For G_2*, the primer has advanced up to the lesion site (n); hence, such lesion-induced stacking cannot attain the same stability. Modeling an ss/ds sequence was difficult and indicated a greater flexibility of the FABP lesion at G_1*; however, this sequence does not support stacking-induced stabilization at the replication fork. As a result, n G_2* was less stable than that of the unmodified control. The opposite was true for the 16/11-mer, in which the binding stability at G_1* was reduced at the lesion site presumably due to the clash with the new primer. This is not the case for the B-conformeric G_2* at $n + 1$, which did not change much other than the usual gain of binding energy through elongation. The trend remained unchanged for the rest of the elongation ($n + 2$, +3) and for the full ($n + 16$) duplex.

As shown in Figure 18, the nature of the lesion (FABP vs. FAF) had no noticeable effect on dissociation rates, except that FAF dissociates generally faster than FABP (Table 6). A similar situation was noted in our previous GCA/GCT examples[2]. Although no polymerase was involved, these SPR results are in line with the NMR/thermodynamic/modeling results given that the greater S-conformation of the G_1* sequence posed a major hindrance toward the binding affinity. In other words, the lesion

effect was propagated primarily upstream of the template strand and gradually diminished.

SUMMARY

We conducted systematic 19F NMR, CD, DSC, and SPR studies for the elongation of the TG_1*G_2T and TG_1G_2*T sequences modified by FABP and FAF. Model systems mimic a hypothetical TLS process involving the bulky structurally related arylamine lesions. DNA adducts in the TG_1*G_2T duplex exhibited greater populations of the S-conformer compared with that in the TG_1G_2*T duplex. Bulky lesions exhibited unique sequence-dependent conformational heterogeneities at various elongation positions, including the replication fork that differentially contributes to template-primer bindings and thermodynamics. We found that the B-conformer is a major thermodynamic stabilizer in the duplex settings and the S-conformer is a destabilizer. The opposite was observed during the primer elongation process involving ss/ds duplex sequences, i.e., the S-conformation promotes lesion stacking with nascent base pairs at the ss/ds replication fork. The results are supported by SPR sensograms, which revealed that the bulky adduct in the stacked conformation increases the binding affinity of the complementary strands in the order of FABP > FAF and G_1*G_2T > G_1G_2*T. Interestingly, unlike the coplanar FAF, the twisted FABP is a better stacker at ss/ds junction than in fully paired duplexes. The extent of primer dissociation rates across the lesion were influenced by the flanking DNA sequences and the stacking ability of the lesion. Future studies include examination of the effects of the conformational heterogeneity in DNA replication and repair.

REFERENCES

[1] Jain, V., Hilton, B., Lin, B., Patnaik, S., Liang, F., Darian, E., Zou, Y., MacKerell, A. D., 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 Research 41*, 869-880.

[2] 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: 19F NMR, surface plasmon resonance, and primer kinetic studies, *Biochemistry 53*, 4059-4071.

[3] Skipper, P. L., Kim, M. Y., Sun, H. L., Wogan, G. N., and Tannenbaum, S. R. (2010) Monocyclic aromatic amines as potential human carcinogens: old is new again, *Carcinogenesis 31*, 50-58.

[4] Turesky, R. J., and Le Marchand, L. (2011) Metabolism and biomarkers of heterocyclic aromatic amines in molecular epidemiology studies: lessons learned from aromatic amines, *Chem Res Toxicol 24*, 1169-1214.

[5] Hein, D. W. (1988) Acetylator genotype and arylamine-induced carcinogenesis, *Biochim Biophys Acta 948*, 37-66.

[6] Kim, D., and Guengerich, F. P. (2005) Cytochrome P450 activation of arylamines and heterocyclic amines, *Annu Rev Pharmacol Toxicol 45*, 27-49.

[7] Liang, F., and Cho, B. P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity, *Biochemistry 49*, 259-266.

[8] Zhang, Y. (2013) Understanding the Gender Disparity in Bladder Cancer Risk: The Impact of Sex Hormones and Liver on Bladder Susceptibility to Carcinogens, *Journal of Environmental Science and Health, Part C 31*, 287-304.

[9] 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.

[10] Patnaik, S., and Cho, B. P. (2010) Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity, *Chemical Research in Toxicology 23*, 1650-1652.

[11] 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.

[12] Jain, V., Hilton, B., Lin, B., Jain, A., MacKerell, A. D., Jr., Zou, Y., and Cho, B. P. (2013) Structural and thermodynamic insight into Escherichia coli UvrABC-mediated incision of cluster diacetylaminofluorene adducts on the NarI sequence, *Chem Res Toxicol 26*, 1251-1262.

[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] Meneni, S. R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M. P., and Cho, B. P. (2006) Sequence effects of aminofluorene-modified DNA duplexes:

thermodynamic and circular dichroism properties, *Nucleic Acids Res 34*, 755-763.

[15] Zou, Y., Shell, S. M., Utzat, C. D., Luo, C., Yang, Z., Geacintov, N. E., and Basu, A. K. (2003) Effects of DNA Adduct Structure and Sequence Context on Strand Opening of Repair Intermediates and Incision by UvrABC Nuclease†, *Biochemistry 42*, 12654-12661.

[16] 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 Research 37*, 1628-1637.

[17] Liang, F., Meneni, S., and Cho, B. P. (2006) Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene-DNA adducts, *Chem Res Toxicol 19*, 1040-1043.

[18] Broschard, T. H., Koffel-Schwartz, N., and Fuchs, R. P. P. (1999) Sequence-dependent modulation of frameshift mutagenesis at NarI-derived mutation hot spots, *Journal of Molecular Biology 288*, 191-199.

[19] 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.

[20] Xu, L., Vaidyanathan, V. G., and Cho, B. P. (2014) Real-time surface plasmon resonance study of biomolecular interactions between polymerase and bulky mutagenic DNA lesions, *Chem Res Toxicol 27*, 1796-1807.

[21] Koffel-Schwartz, N., and Fuchs, R. P. (1995) Sequence determinants for -2 frameshift mutagenesis at NarI-derived hot spots, *Journal of Molecular Biology 252*, 507-513.

[22] 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.

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

[24] Sandineni, A., Lin, B., MacKerell, A. D., Jr., and Cho, B. P. (2013) Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis, *Chem Res Toxicol 26*, 937-951.

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

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

[27] Chakrabarti, M. C., and Schwarz, F. P. (1999) Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry, *Nucleic Acids Res 27*, 4801-4806.

[28] Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M. P., and Cho, B. P. (2012) Conformational and thermodynamic properties modulate the nucleotide

excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence, *Nucleic Acids Research 40*, 3939-3951.

[29] Busto, N., Valladolid, J., Martínez-Alonso, M., Lozano, H. J., Jalón, F. A., Manzano, B. R., Rodríguez, A. M., Carrión, M. C., Biver, T., Leal, J. M., Espino, G., and García, B. (2013) Anticancer Activity and DNA Binding of a Bifunctional Ru(II) Arene Aqua-Complex with the 2,4-Diamino-6-(2-pyridyl)-1,3,5-triazine Ligand, *Inorganic Chemistry 52*, 9962-9974.

[30] 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.

[31] Ma, H., Wang, L., Niesen, D. B., Cai, A., Cho, B. P., Tan, W., Gu, Q., Xu, J., and Seeram, N. P. (2015) Structure Activity Related, Mechanistic, and Modeling Studies of Gallotannins containing a Glucitol-Core and alpha-Glucosidase, *RSC Adv 5*, 107904-107915.

[32] Takach, J. C., Mikulecky, P. J., and Feig, A. L. (2004) Salt-dependent heat capacity changes for RNA duplex formation, *J Am Chem Soc 126*, 6530-6531.

[33] Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., T.E. Cheatham, I., Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C.

L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., and Kollman, P. A. (2014) AMBER 14, University of California, San Francisco.

[34] Li, P., and Merz, K. M. (2014) Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions, *Journal of Chemical Theory and Computation 10*, 289-297.

[35] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *Journal of Chemical Theory and Computation 11*, 3696-3713.

[36] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules, *J. Am. Chem. Soc. 117*, 5179–5197.

[37] Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations, *J Mol Graph Model 25*, 247-260.

[38] Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993) Application of RESP Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation, *J. Am. Chem. Soc. 115*, 9620–9631.

[39] Liang, F., and Cho, B. P. (2007) Probing the thermodynamics of aminofluorene-induced translesion DNA synthesis by differential scanning calorimetry, *J Am Chem Soc 129*, 12108-12109.

FIGURE LEGENDS

Figure 1. (a) Structures of ABP [N-(2ˈ-deoxyguanosin-8-yl)-4-aminobiphenyl], AF [N-(2ˈ-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [N-(2ˈ-deoxyguanosin-8-yl)-2-acetylaminofluorene] and their fluoro models, FABP [N-(2ˈ-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl], FAF [N-(2ˈ-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene], and FAAF [N-(2ˈ-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene]. (b) Major groove views of arylamine-conformational motifs: B-, S-, and W-conformers. (color code: Arylamine- red CPK, modified dG- blue, complementary C- orange).

Figure 2. Translesion synthesis (TLS) models for (a) unmodified control, (b) FABP/FAF-modified G1 (G*GT) templates, and (c) FABP/FAF-modified G2(GG*T) templates.

Figure 3. HPLC chromatogram of a reaction mixture of 16-mer sequence (5'-CTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FABP** (N-Acetoxy-N-(trifluoroacetyl)-4'-fluoro-4-aminobiphenyl) and photodiode array UV spectra of unmodified, mono-, di-FABP adducts.

Figure 4. HPLC chromatogram of a reaction mixture of 16-mer sequence (5'-CTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FAF** (N-acetoxy-N-(trifluoroacetyl)-7-fluoro-2-aminofluorene) and photodiode array UV spectra of unmodified and mono-FAF adducts.

Figure 5. HPLC profile, photodiode array UV spectra of mono-**FABP** adducts and MALDI-TOF mass spectra of **FABP** modified **biotinylated** 16mer-GGT for SPR

experiments: (a) HPLC chromatogram of **FABP** modified **biotinylated** 16mer-GGT. (b) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 3min 5min, 6min and 7min. (c) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 3min 5min, 6min and 7min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion.

Figure 6. MALDI-TOF mass spectra of **FABP** modified 16mer-GGT: (a) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 2min and 5min. (b) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 2min and 5min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion.

Figure 7. MALDI-TOF mass spectra of **FAF** modified 16mer-GGT: (a) 3'-enzyme digestion of **peak 1** at 0min, 0.5min, 1min, 2min, 3min, 5min and 6min. (b) 3'-enzyme digestion of **peak 2** at 0min, 0.5min, 1min, 2min, 3min, 5min and 6min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion.

Figure 8. Dynamic ^{19}F-NMR spectra (5-60°C) of **FABP**-modified duplexes with primer elongation. (a) FABP-modified G1 (G*GT) showing a 67% B- and 33% S-conformational mixture in full duplex. (b) FABP-modified G2 (GG*T) exhibiting 100% B-conformer in full duplex. For all of the duplexes, the ^{19}F signals around 60°C coalesced into a sharp single peak around -115ppm indicating a denatured FABP-residue. * indicates unassigned minor conformers.

Figure 9. Dynamic ^{19}F-NMR spectra (5-60°C) of **FAF**-modified duplexes with primer elongation. (a) FAF-modified G1 (G*GT) showing a 25% B- and 75% S-conformational mixture in full duplex. (b) FAF-modified G2 (GG*T) exhibiting 83% B- and 17% S-conformer in full duplex. For all of the duplexes, the ^{19}F signals around 60°C coalesced into a sharp single peak around -117ppm indicating a denatured FAF-residue. * indicates unassigned minor conformers.

Figure 10. Deuterium isotope effect of FABP-modified 16-mer GGT full duplexes in 10% and 100% D_2O at 25°C. (a) FABP-modified G1 (G*GT) (0.18 and 0.08 ppm for B- and S-conformer, respectively). (b) FABP-modified G2 (GG*T) (0.18 ppm for B-conformer).

$\Delta\delta_S = |\delta_{S\ (100\%\ D2O)} - \delta_{S\ (10\%\ D2O)}|$; $\Delta\delta_B = |\delta_{B\ (100\%\ D2O)} - \delta_{B\ (10\%\ D2O)}|$.

Figure 11. Deuterium isotope effect of FAF-modified 16-mer GGT full duplexes in 10% and 100% D_2O at 25°C. (a) FAF-modified G1 (G*GT) (0.19 and 0.04 ppm for B- and S-conformer, respectively). (b) FAF-modified G2 (GG*T) (0.18 and 0.04 ppm for B- and S-conformer, respectively).

$\Delta\delta_S = |\delta_{S\ (100\%\ D2O)} - \delta_{S\ (10\%\ D2O)}|$; $\Delta\delta_B = |\delta_{B\ (100\%\ D2O)} - \delta_{B\ (10\%\ D2O)}|$.

Figure 12. CD spectral overlays of primer elongation (9 mer to 16 mer) for **FABP** modified 16-mer GGT at 25°C. (a) unmodified control, (b) FABP modified G1 (G*GT), (c) FABP modified G2 (GG*T), (d) Full duplex overlays of unmodified and FABP G1/G2-modified sequences.

Figure 13. CD spectral overlays of primer elongation (9 mer to 16 mer) for **FAF** modified 16-mer GGT at 25°C. (a) unmodified control, (b) FAF modified G1 (G*GT),

(c) FAF modified G2 (GG*T), (d) Full duplex overlays of unmodified and FAF G1/G2-modified sequences.

Figure 14. (a) UV melting (T_m) of **FABP** modified 16-mer GGT with primer elongation (9 mer to 16 mer). (b) UV melting (T_m) of **FAF** modified 16-mer GGT with primer elongation (9 mer to 16 mer). Unmodified control is blue line, FABP/FAF-modified G1 (G*GT) is shown as red dash line, and FABP/FAF-modified G2 (GG*T) is shown as yellow dash line.

Figure 15. DSC heat capacity ΔC_p *vs* temperature (15-85 °C) for primer elongation (9 mer to 16 mer) of the (a) control, (b) **FABP** modified G1 (G*GT) and (c) **FABP** modified G2 (GG*T). (d) Full duplex overlays of unmodified and **FABP** G1/G2-modified sequences. FABP-G1 and FABP-G2 show smaller T_m and ΔH than the control/unmodified in full duplex exhibiting that S-conformer (FABP-G1, G*GT) might produce thermodynamic destabilization in full duplex.

Figure 16. ITC binding thermograms of duplex formation for (a) FABP- and (b) FAF-modified 16 mer GGT full duplexes at 25°C.

Figure 17. SPR sensorgrams of unmodified and FABP/FAF-modified TLS: (a) Unmodified and FABP-modified G1 (G*GT) and G2 (GG*T), (b) Unmodified and FAF-modified G1 (G*GT) and G2 (GG*T).

Figure 18. Normalized SPR sensorgrams of (a) FABP- and (b) FAF- modified 16 mer GGT sequences with primer elongation (9 mer to 16 mer). Slower dissociation shows better stability of the primer−template duplexes.

Figure 19. Normalized SPR sensorgrams of FABP-modified 16 mer GGT sequences with 9-mer and 10-mer primers, and progression of lesion conformational rigidity. (a) 9-mer and (b) 10-mer. Slower dissociation shows better stability of the primer−template duplexes.

Figure 20. *syn*- and *anti*-FABP-modified dG and their energy barrier from nucleoside MD (Modeling Dynamic) simulation.

Figure 21. FABP-dG DNA helices structures from MD simulation. (a) G_1 position, (b) G_2 position. B conformation is preferred at both the G1 and G2 positions.

TABLE LEGENDS

Table 1. Partial charges and atom types for FABP-dG.

Table 2: UV melting thermal and thermodynamic parameters of **FABP** modified 16-mer G1 (G*GT) and G2 (GG*T) duplexes.

Table 3: UV melting thermal and thermodynamic parameters of **FAF** modified 16-mer G1 (G*GT) and G2 (GG*T) duplexes.

Table 4. Thermal and Thermodynamic Parameters Derived from DSC.

Table 5. Thermodynamic parameters C, N, K, ΔH and ΔS derived from the ITC.

Table 6. Dissociation off-rate constants (k_d, $x10^{-2}$, s^{-1}) from SPR.

Figure 1:

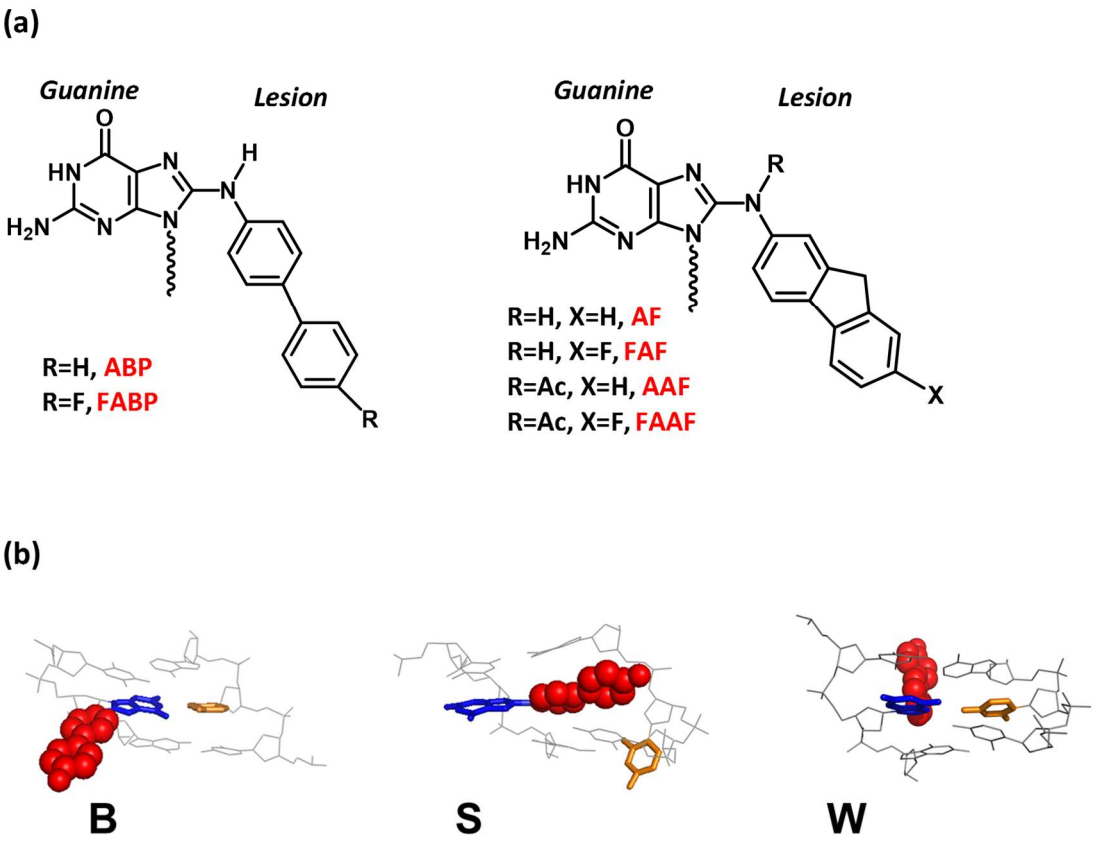

(a)
Guanine
Lesion
R=H, ABP
R=F, FABP
Guanine
Lesion
R=H, X=H, AF
R=H, X=F, FAF
R=Ac, X=H, AAF
R=Ac, X=F, FAAF
(b)
B
S
W

Figure 2:

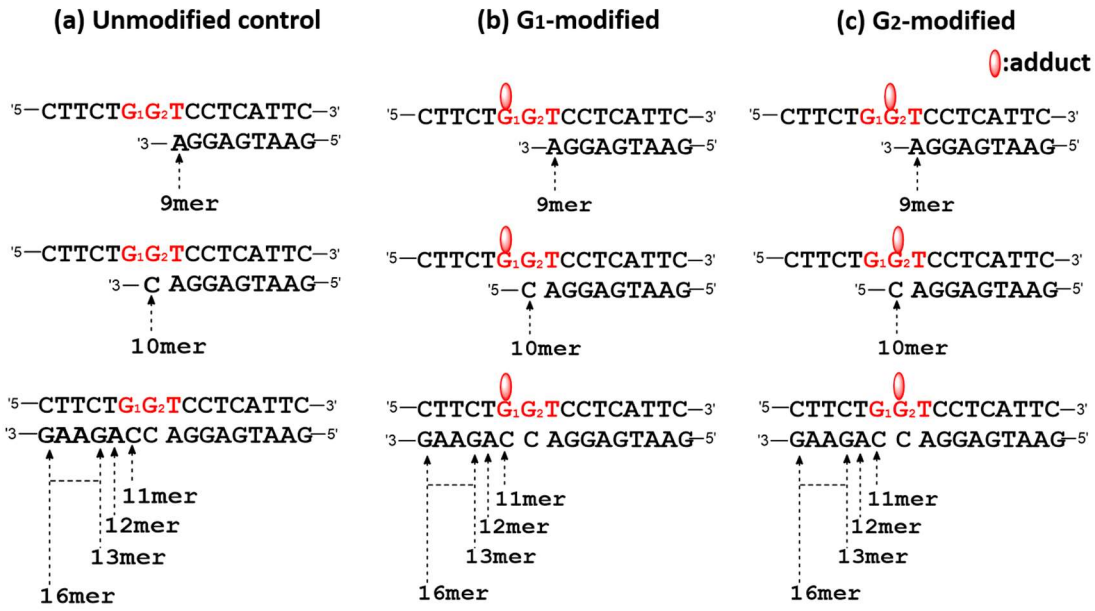

(a) Unmodified control
(b) G1-modified
(c) G2-modified
:adduct
5'—CTTCTG1G2TCCTCATTC—3'
3'—AGGAGTAAG—5'
9mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—C AGGAGTAAG—5'
10mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—GAAGACC AGGAGTAAG—5'
11mer
12mer
13mer
16mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—AGGAGTAAG—5'
9mer
5'—CTTCTG1G2TCCTCATTC—3'
5'—C AGGAGTAAG—5'
10mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—GAAGAC C AGGAGTAAG—5'
11mer
12mer
13mer
16mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—AGGAGTAAG—5'
9mer
5'—CTTCTG1G2TCCTCATTC—3'
5'—C AGGAGTAAG—5'
10mer
5'—CTTCTG1G2TCCTCATTC—3'
3'—GAAGAC C AGGAGTAAG—5'
11mer
12mer
13mer
16mer

Figure 3:

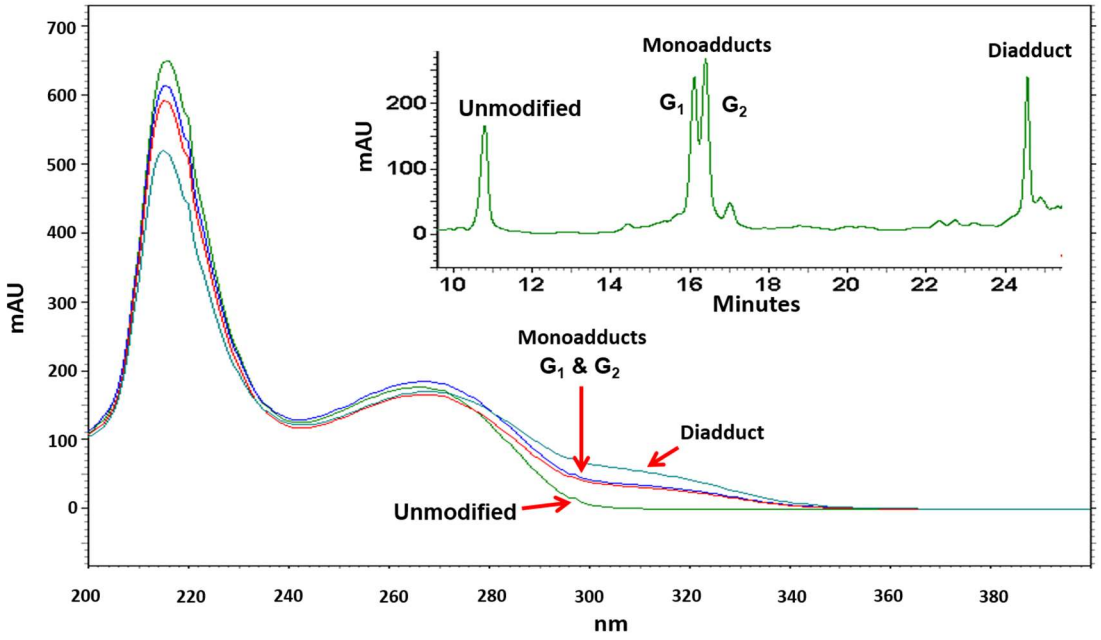

mAU
700
600
500
400
300
200
100
0
200
220
240
260
280
300
320
340
360
380
nm
Monoadducts
G1 & G2
Diadduct
Unmodified
mAU
200
100
0
Unmodified
Monoadducts
G1
G2
Diadduct
10
12
14
16
18
20
22
24
Minutes

Figure 4:

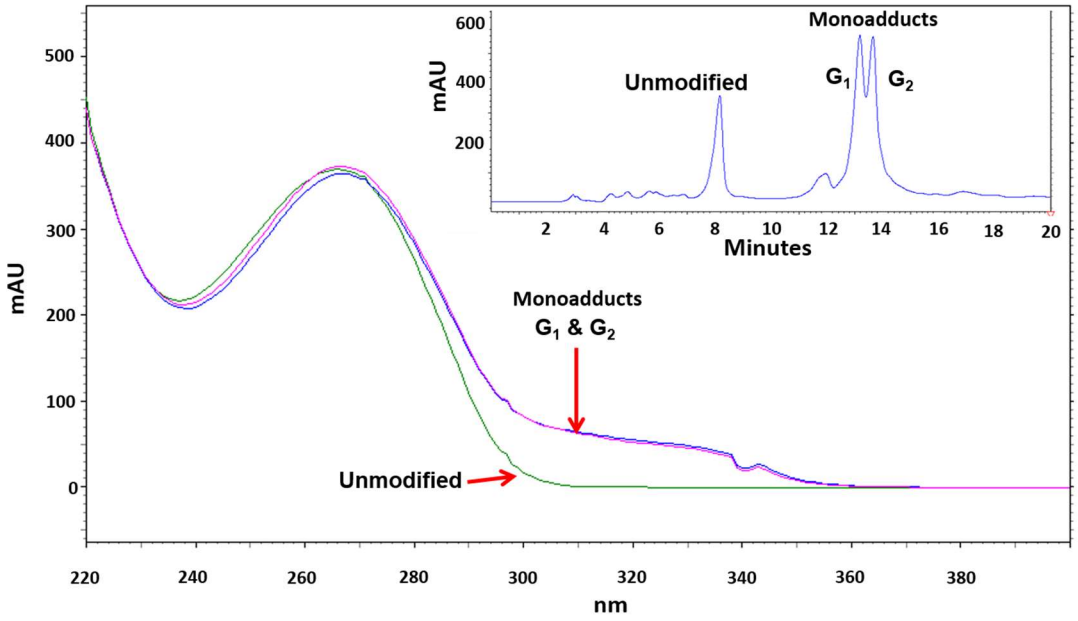
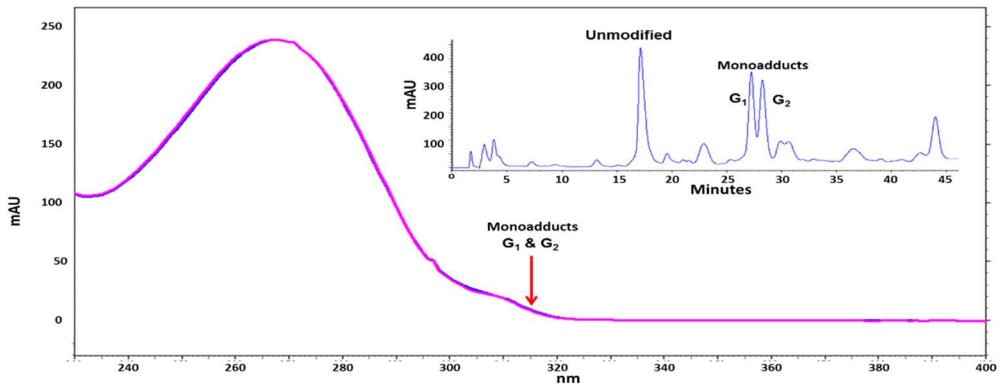

mAU
500
400
300
200
100
0
220
240
260
280
300
320
340
360
380
nm
Monoadducts
G1 & G2
Unmodified
600
400
200
mAU
Unmodified
Monoadducts
G1
G2
2
4
6
8
10
12
14
16
18
20
Minutes

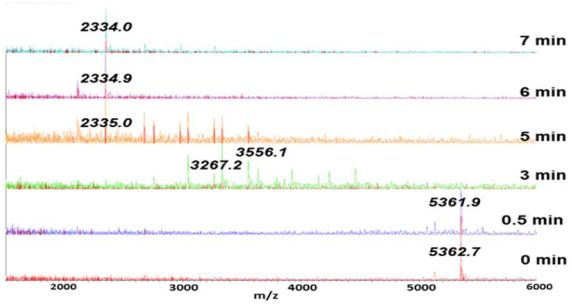
Figure 5:

(a)



(b)

mass	sequence
5363.2	Bio-CTTCTG$_1$G$_2$TCCTCATTC
5073.9	Bio-CTTCTG$_1$G$_2$TCCTCATT
4769.8	Bio-CTTCTG$_1$G$_2$TCCTCAT
4465.6	Bio-CTTCTG$_1$G$_2$TCCTCA
4152.4	Bio-CTTCTG$_1$G$_2$TCCTC
3863.2	Bio-CTTCTG$_1$G$_2$TCCT
3558.9	Bio-CTTCTG$_1$G$_2$TCC
3269.8	Bio-CTTCTG$_1$G$_2$TC
2980.6	Bio-CTTCTG$_1$G$_2$T
2676.4	Bio-CTTCTG$_1$G$_2$
2336.2	Bio-CTTCTG$_1$



(c)

mass	sequence
5363.2	Bio-CTTCTG$_1$G$_2$TCCTCATTC
5073.9	Bio-CTTCTG$_1$G$_2$TCCTCATT
4769.8	Bio-CTTCTG$_1$G$_2$TCCTCAT
4465.6	Bio-CTTCTG$_1$G$_2$TCCTCA
4152.4	Bio-CTTCTG$_1$G$_2$TCCTC
3863.2	Bio-CTTCTG$_1$G$_2$TCCT
3558.9	Bio-CTTCTG$_1$G$_2$TCC
3269.8	Bio-CTTCTG$_1$G$_2$TC
2980.6	Bio-CTTCTG$_1$G$_2$T
2676.4	Bio-CTTCTG$_1$G$_2$

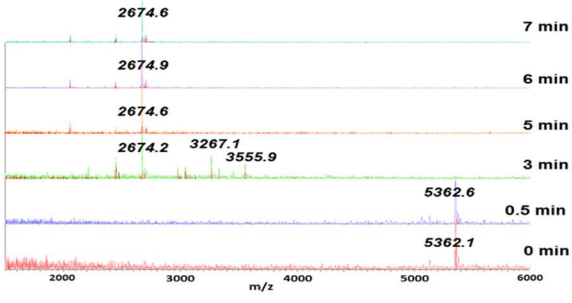
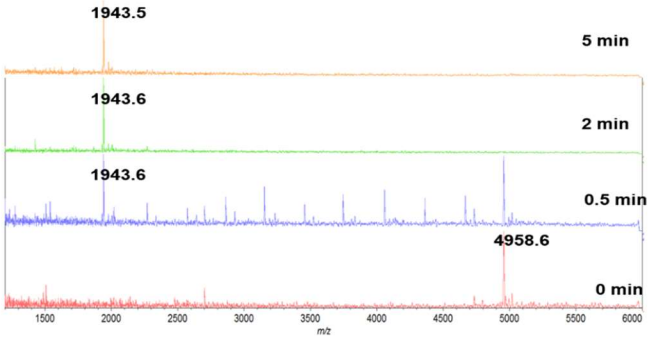


Figure 6:

(a)

Mass	Sequence
1945.285	CTTCTG$_1$
2274.495	CTTCTG$_1$G$_2$
2578.692	CTTCTG$_1$G$_2$T
2867.877	CTTCTG$_1$G$_2$TC
3157.062	CTTCTG$_1$G$_2$TCC
3461.259	CTTCTG$_1$G$_2$TCCT
3750.444	CTTCTG$_1$G$_2$TCCTC
4063.654	CTTCTG$_1$G$_2$TCCTCA
4367.851	CTTCTG$_1$G$_2$TCCTCAT
4672.048	CTTCTG$_1$G$_2$TCCTCATT
4961.233	CTTCTG$_1$G$_2$TCCTCATTC



(b)

Mass	Sequence
2274.495	CTTCTG$_1$G$_2$
2578.692	CTTCTG$_1$G$_2$T
2867.877	CTTCTG$_1$G$_2$TC
3157.062	CTTCTG$_1$G$_2$TCC
3461.259	CTTCTG$_1$G$_2$TCCT
3750.444	CTTCTG$_1$G$_2$TCCTC
4063.654	CTTCTG$_1$G$_2$TCCTCA
4367.851	CTTCTG$_1$G$_2$TCCTCAT
4672.048	CTTCTG$_1$G$_2$TCCTCATT
4961.233	CTTCTG$_1$G$_2$TCCTCATTC

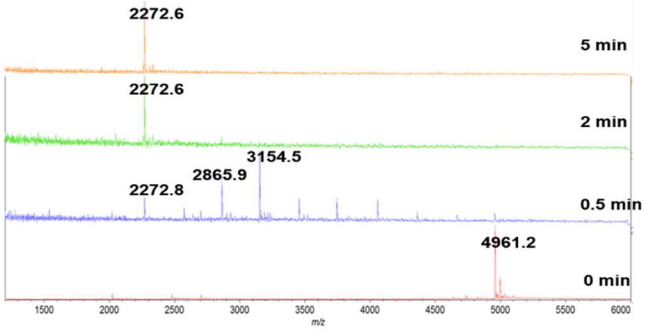
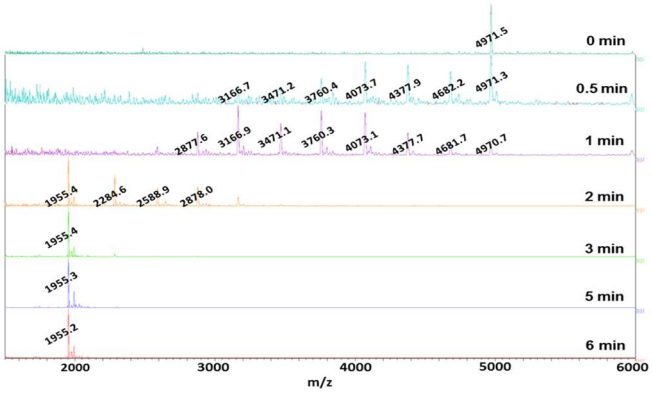


Figure 7:

(a)

Mass	Sequence
1955.085	CTTCTG$_1$
2284.295	CTTCTG$_1$G$_2$
2588.492	CTTCTG$_1$G$_2$T
2877.677	CTTCTG$_1$G$_2$TC
3166.862	CTTCTG$_1$G$_2$TCC
3471.059	CTTCTG$_1$G$_2$TCCT
3760.244	CTTCTG$_1$G$_2$TCCTC
4073.454	CTTCTG$_1$G$_2$TCCTCA
4377.651	CTTCTG$_1$G$_2$TCCTCAT
4681.848	CTTCTG$_1$G$_2$TCCTCATT
4971.033	CTTCTG$_1$G$_2$TCCTCATTC



(b)

Mass	Sequence
2284.295	CTTCTG$_1$G$_2$
2588.492	CTTCTG$_1$G$_2$T
2877.677	CTTCTG$_1$G$_2$TC
3166.862	CTTCTG$_1$G$_2$TCC
3471.059	CTTCTG$_1$G$_2$TCCT
3760.244	CTTCTG$_1$G$_2$TCCTC
4073.454	CTTCTG$_1$G$_2$TCCTCA
4377.651	CTTCTG$_1$G$_2$TCCTCAT
4681.848	CTTCTG$_1$G$_2$TCCTCATT
4971.033	CTTCTG$_1$G$_2$TCCTCATTC

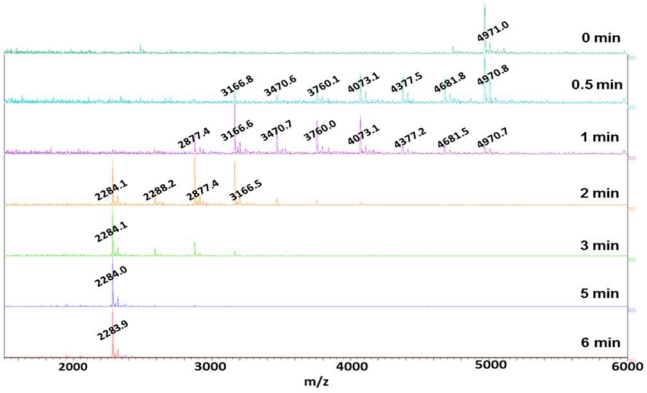
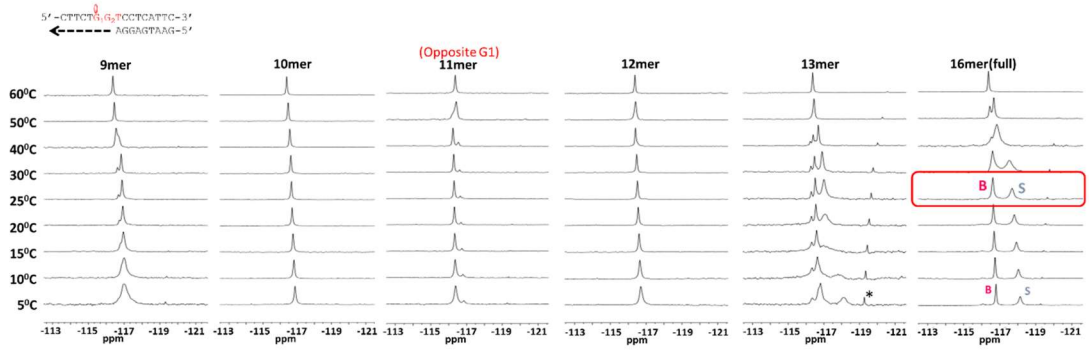


Figure 8:

(a)


5′-CTTCTG1G2TCCTCATTC-3′
AGGAGTAAG-5′
(Opposite G1)
9mer
10mer
11mer
12mer
13mer
16mer(full)
60°C
50°C
40°C
30°C
25°C
20°C
15°C
10°C
5°C
B
S
*
-113 -115 -117 -119 -121
ppm

(b)

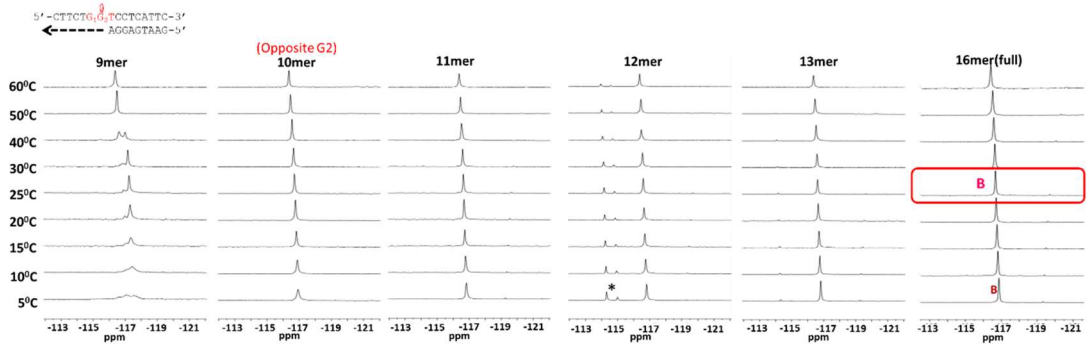
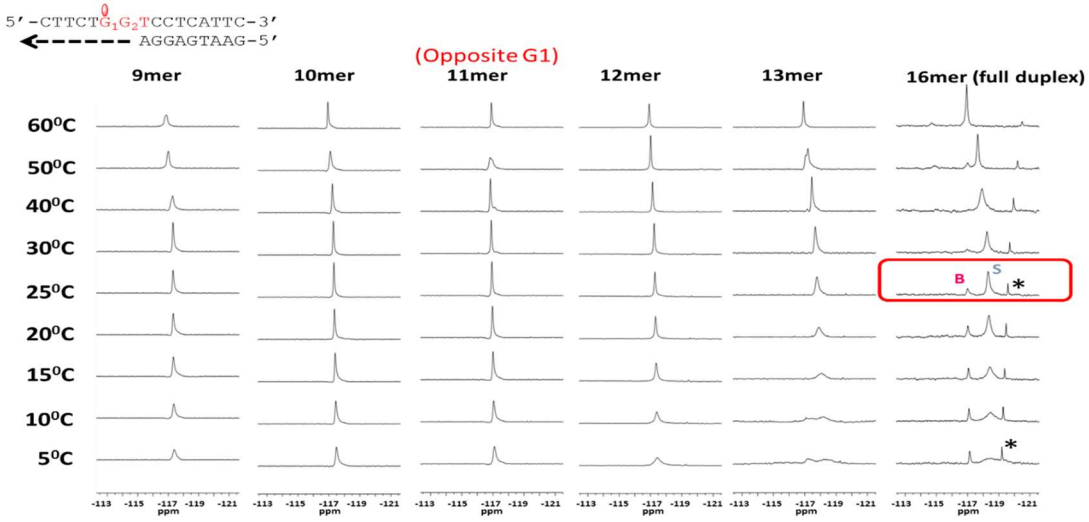

5′-CTTCTG1G2TCCTCATTC-3′
AGGAGTAAG-5′
(Opposite G2)
9mer
10mer
11mer
12mer
13mer
16mer(full)
60°C
50°C
40°C
30°C
25°C
20°C
15°C
10°C
5°C
B
*
-113 -115 -117 -119 -121
ppm

Figure 9:

(a)


5′-CTTCTG1G2TCCTCATTC-3′
AGGAGTAAG-5′
(Opposite G1)
9mer
10mer
11mer
12mer
13mer
16mer (full duplex)
60°C
50°C
40°C
30°C
25°C
20°C
15°C
10°C
5°C
B
S
*
-113 -115 -117 -119 -121
ppm

(b)

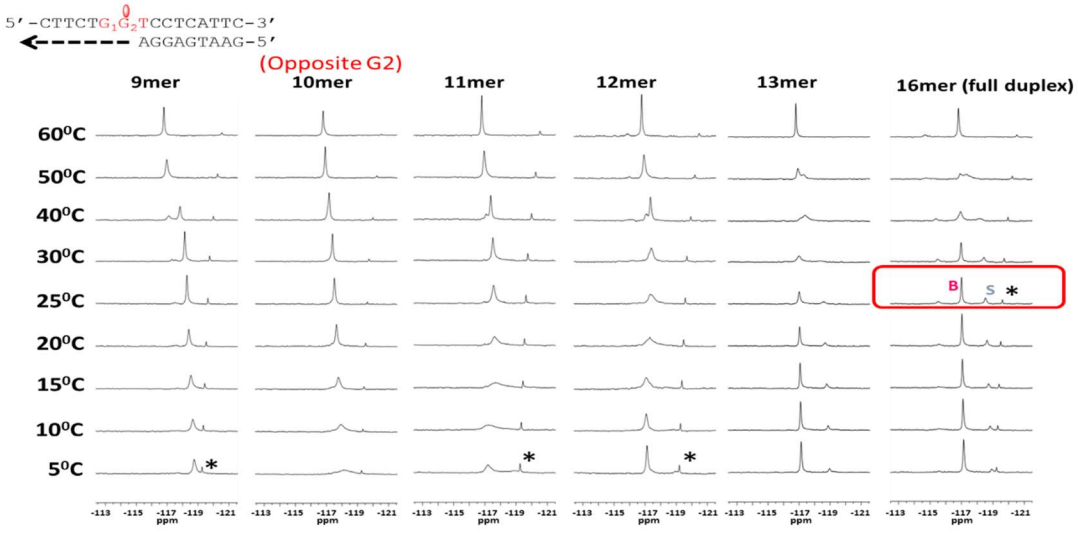

5′-CTTCTG1G2TCCTCATTC-3′
AGGAGTAAG-5′
(Opposite G2)
9mer
10mer
11mer
12mer
13mer
16mer (full duplex)
60°C
50°C
40°C
30°C
25°C
20°C
15°C
10°C
5°C
B
S
*
-113 -115 -117 -119 -121
ppm

Figure 10:

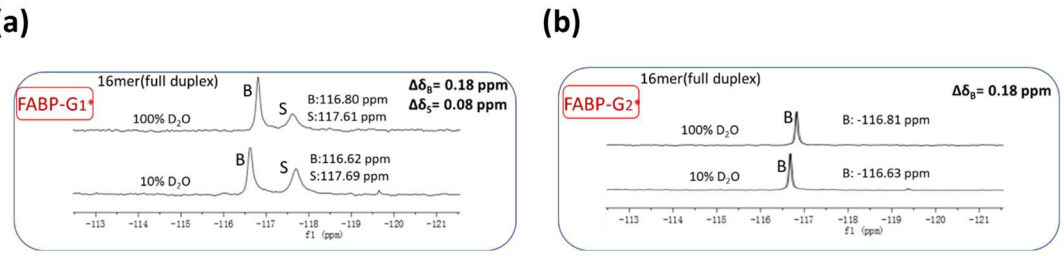

(a)
(b)
16mer(full duplex)
FABP-G1*
100% D2O
10% D2O
B
S
B:116.80 ppm
S:117.61 ppm
B:116.62 ppm
S:117.69 ppm
ΔδB= 0.18 ppm
ΔδS= 0.08 ppm
16mer(full duplex)
FABP-G2*
100% D2O
10% D2O
B: -116.81 ppm
B: -116.63 ppm
ΔδB= 0.18 ppm
f1 (ppm)

Figure 11:

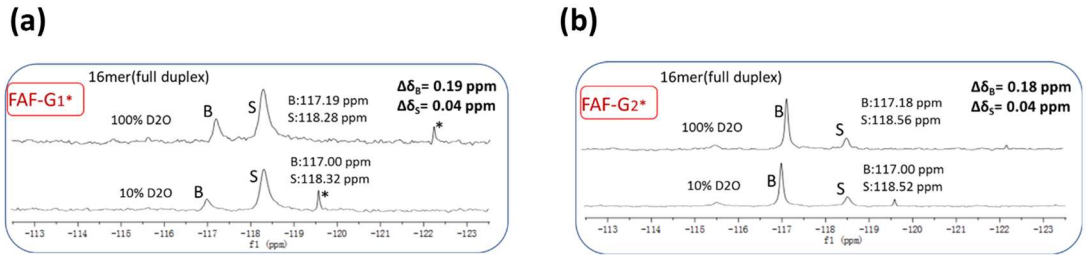

(a)
(b)
16mer(full duplex)
FAF-G1*
100% D2O
10% D2O
B
S
B:117.19 ppm
S:118.28 ppm
B:117.00 ppm
S:118.32 ppm
Δδ_B= 0.19 ppm
Δδ_S= 0.04 ppm
16mer(full duplex)
FAF-G2*
100% D2O
10% D2O
B:117.18 ppm
S:118.56 ppm
B:117.00 ppm
S:118.52 ppm
Δδ_B= 0.18 ppm
Δδ_S= 0.04 ppm
f1 (ppm)

Figure 12:

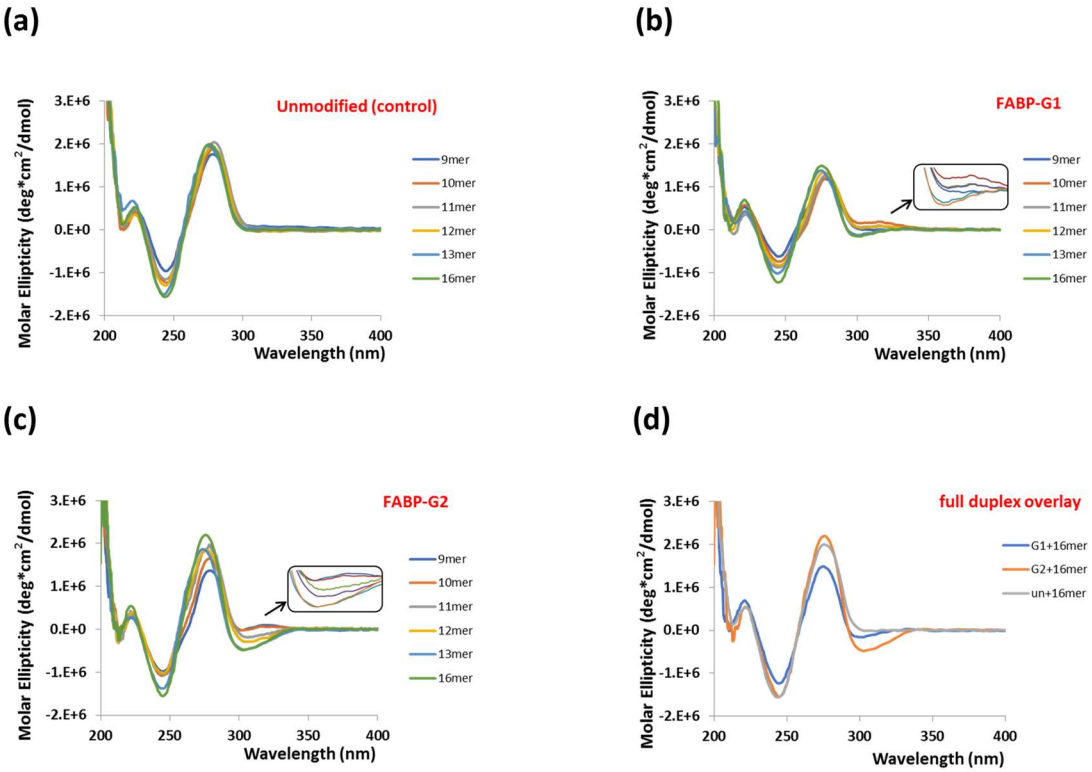

(a)
Unmodified (control)
(b)
FABP-G1
(c)
FABP-G2
(d)
full duplex overlay
Molar Ellipticity (deg*cm2/dmol)
Wavelength (nm)
9mer
10mer
11mer
12mer
13mer
16mer
G1+16mer
G2+16mer
un+16mer

Figure 13:

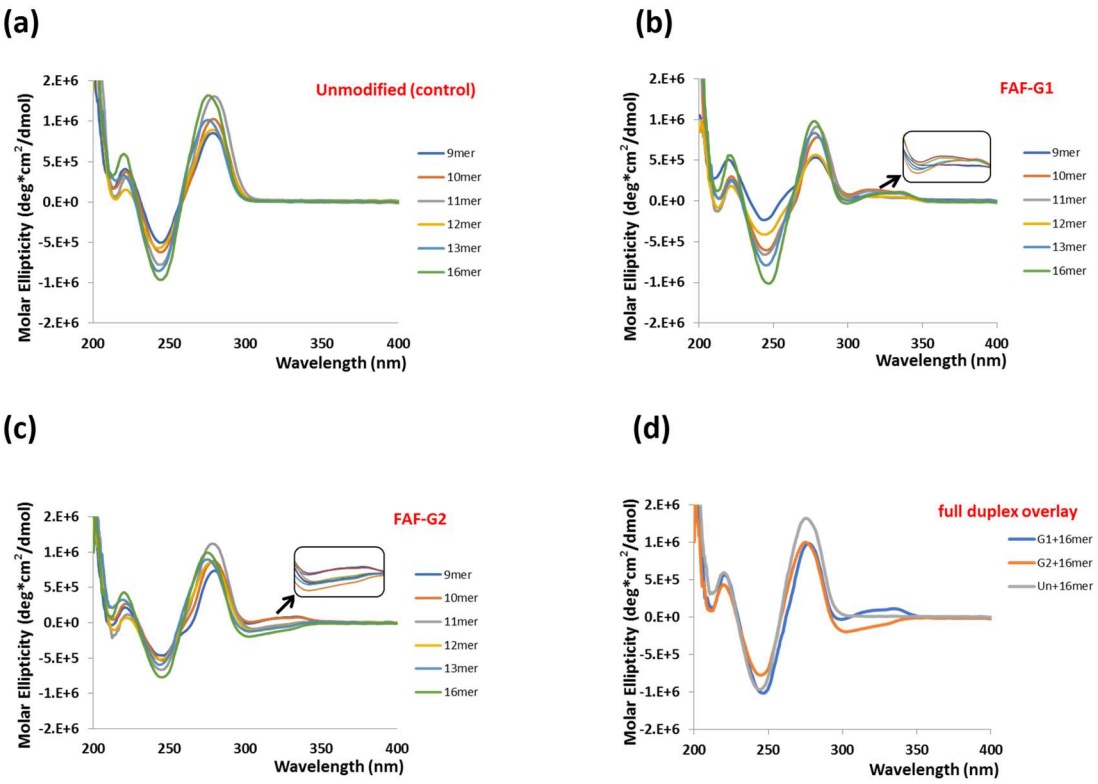
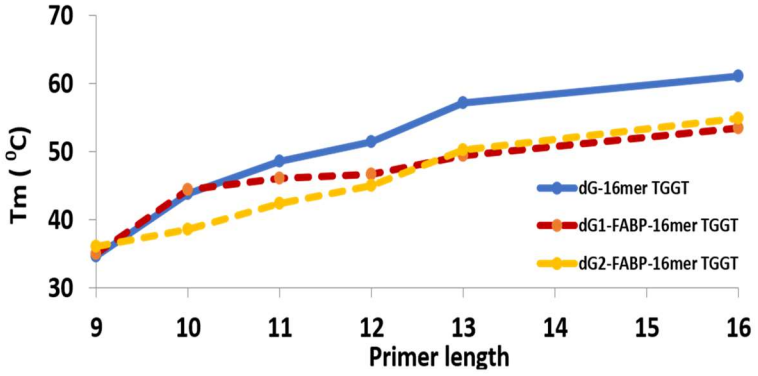

(a)
Unmodified (control)
Molar Ellipticity (deg*cm2/dmol)
2.E+6
1.E+6
5.E+5
0.E+0
-5.E+5
-1.E+6
-2.E+6
200
250
300
350
400
Wavelength (nm)
9mer
10mer
11mer
12mer
13mer
16mer
(b)
FAF-G1
(c)
FAF-G2
(d)
full duplex overlay
G1+16mer
G2+16mer
Un+16mer

Figure 14:

(a)


70
60
50
40
30
Tm (°C)
9
10
11
12
13
14
15
16
Primer length
dG-16mer TGGT
dG1-FABP-16mer TGGT
dG2-FABP-16mer TGGT

(b)

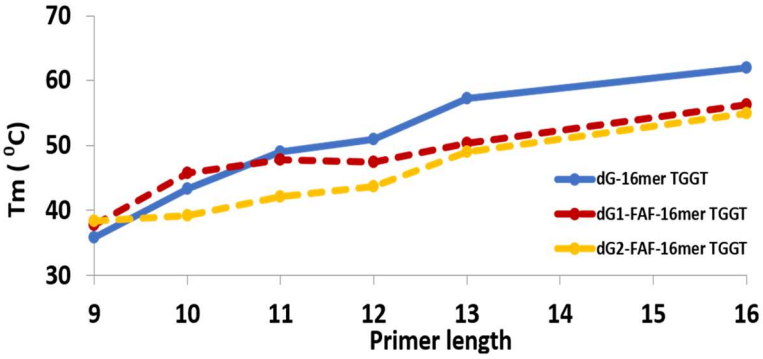

70
60
50
40
30
Tm (°C)
9
10
11
12
13
14
15
16
Primer length
dG-16mer TGGT
dG1-FAF-16mer TGGT
dG2-FAF-16mer TGGT

Figure 15:

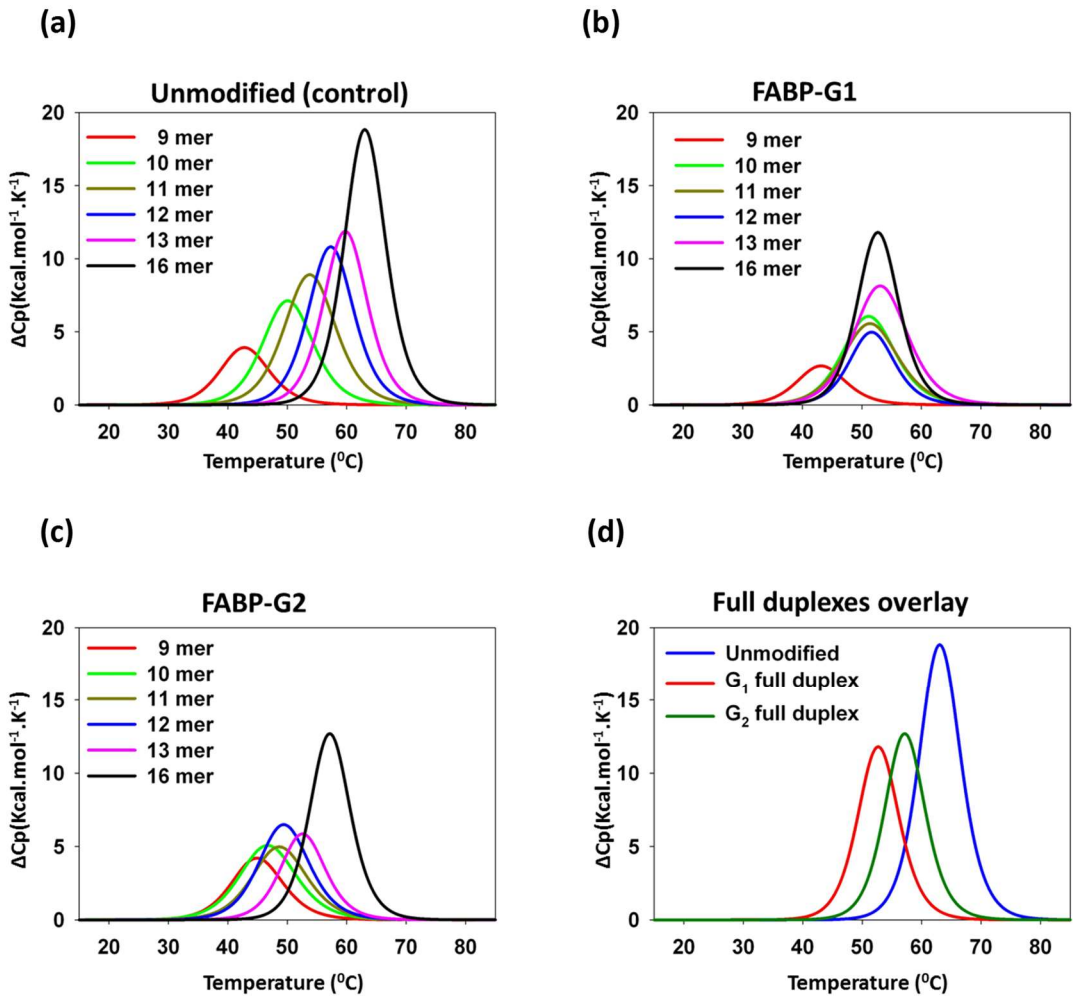

(a)
Unmodified (control)
9 mer
10 mer
11 mer
12 mer
13 mer
16 mer
ΔCp(Kcal.mol-1.K-1)
Temperature (°C)
(b)
FABP-G1
9 mer
10 mer
11 mer
12 mer
13 mer
16 mer
ΔCp(Kcal.mol-1.K-1)
Temperature (°C)
(c)
FABP-G2
9 mer
10 mer
11 mer
12 mer
13 mer
16 mer
ΔCp(Kcal.mol-1.K-1)
Temperature (°C)
(d)
Full duplexes overlay
Unmodified
G1 full duplex
G2 full duplex
ΔCp(Kcal.mol-1.K-1)
Temperature (°C)

Figure 16:

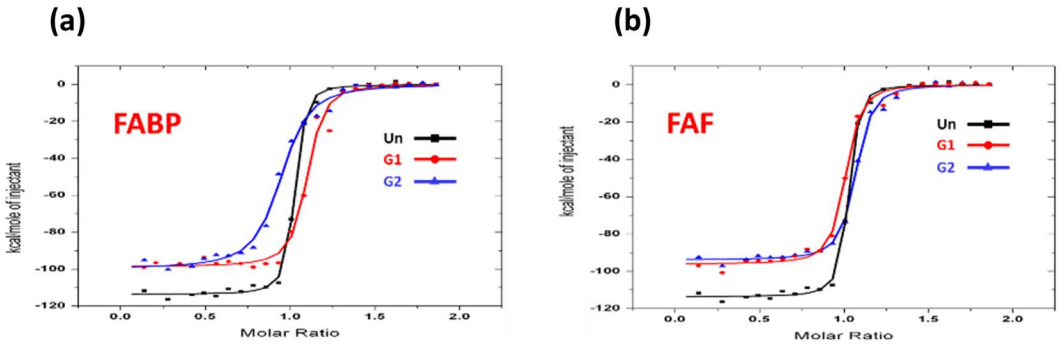
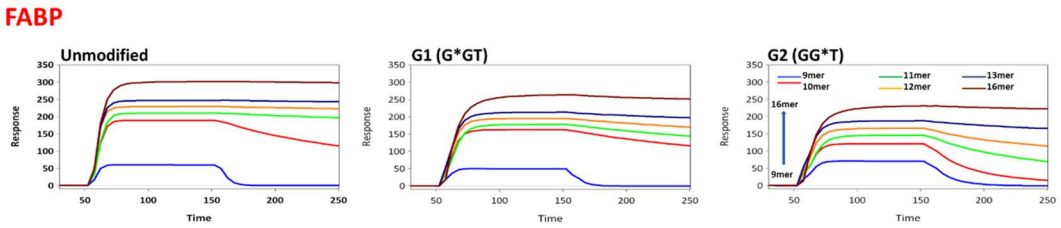

(a)
(b)
FABP
FAF
Un
G1
G2
kcal/mole of injectant
Molar Ratio
0
-20
-40
-60
-80
-100
-120
0.0
0.5
1.0
1.5
2.0

Figure 17:

(a)


FABP
Unmodified
G1 (G*GT)
G2 (GG*T)
9mer
10mer
11mer
12mer
13mer
16mer
Response
Time

(b)

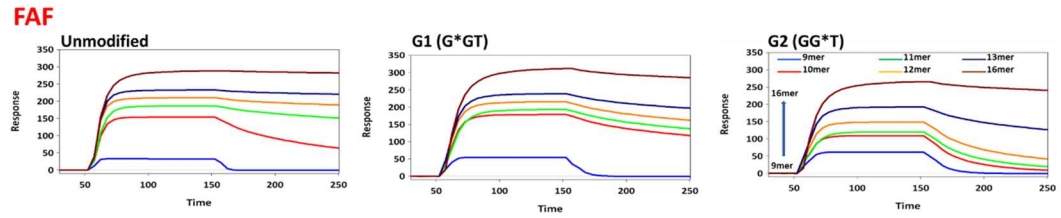
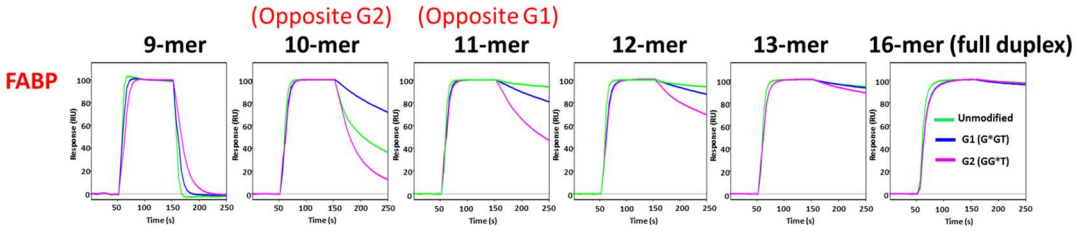

FAF
Unmodified
G1 (G*GT)
G2 (GG*T)
9mer
10mer
11mer
12mer
13mer
16mer
Response
Time

Figure 18:

(a)


(Opposite G2)
(Opposite G1)
9-mer
10-mer
11-mer
12-mer
13-mer
16-mer (full duplex)
FABP
Response (RU)
Time (s)
Unmodified
G1 (G*GT)
G2 (GG*T)

(b)

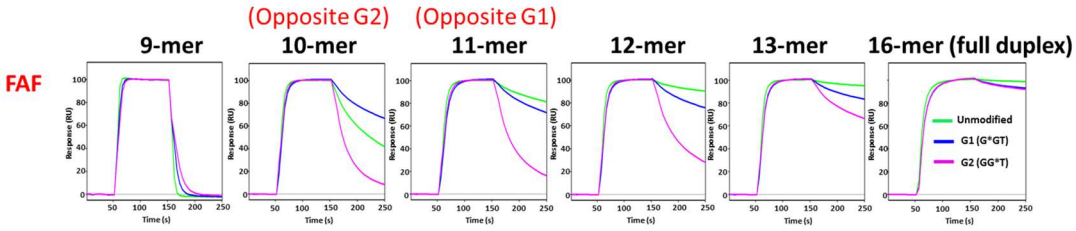

(Opposite G2)
(Opposite G1)
9-mer
10-mer
11-mer
12-mer
13-mer
16-mer (full duplex)
FAF
Response (RU)
Time (s)
Unmodified
G1 (G*GT)
G2 (GG*T)

Figure 19:

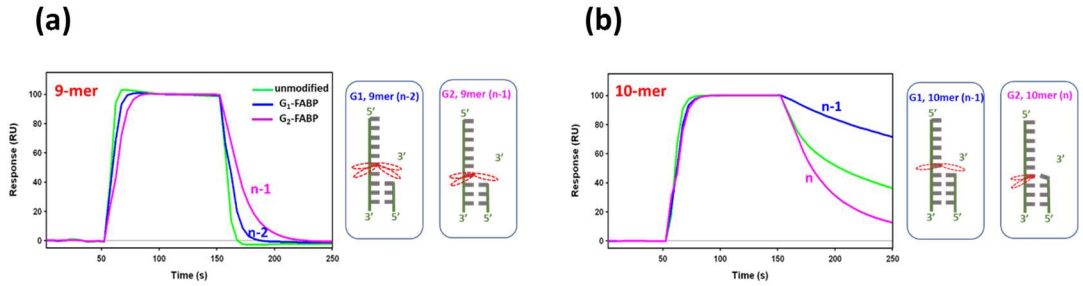

(a)
9-mer
unmodified
G1-FABP
G2-FABP
n-1
n-2
Response (RU)
Time (s)
G1, 9mer (n-2)
G2, 9mer (n-1)
(b)
10-mer
n-1
n
Response (RU)
Time (s)
G1, 10mer (n-1)
G2, 10mer (n)

Figure 20:

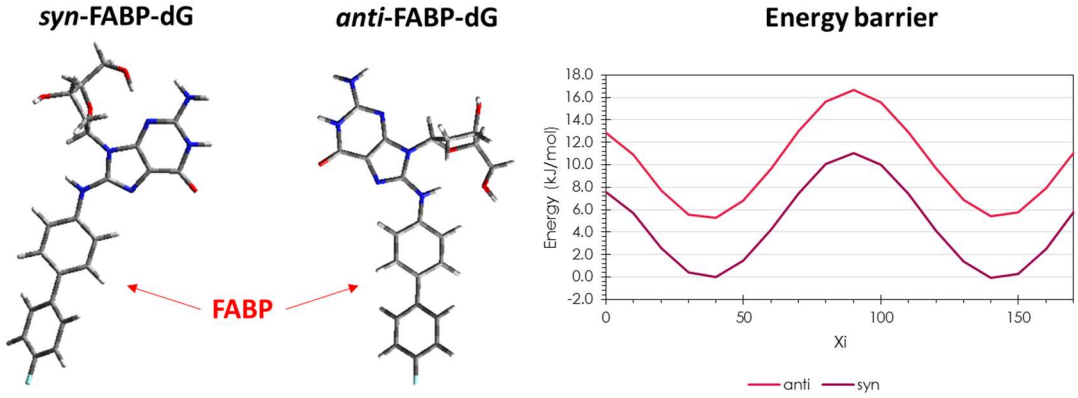
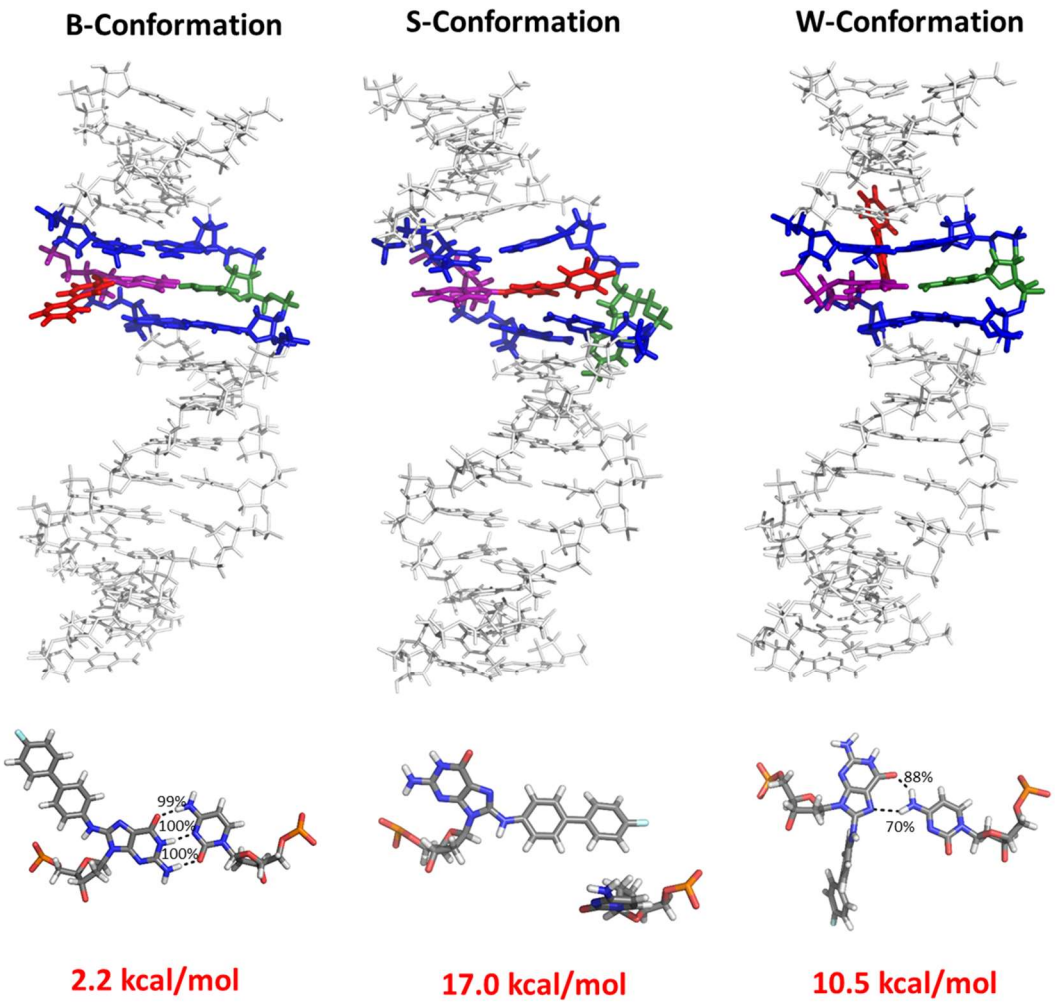

syn-FABP-dG
anti-FABP-dG
Energy barrier
FABP
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
Energy (kJ/mol)
0
50
100
150
Xi
anti
syn

Figure 21:

(a) G_1 position



(b) G_2 position

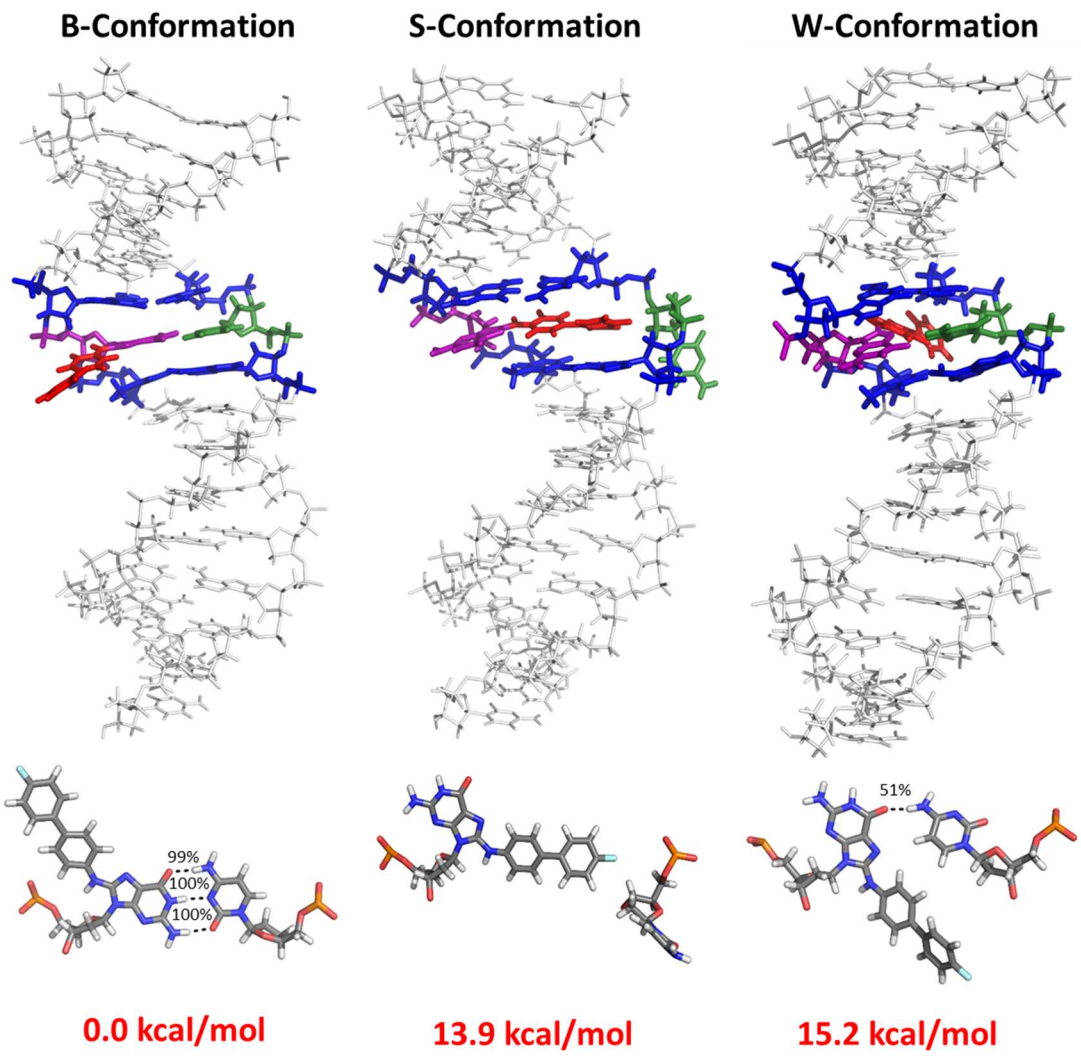


Table 1:

Atom Name	Atom Type	Charge
P	P	1.1625
O5'	OS	-0.4931
OP1	O2	-0.7634
OP2	O2	-0.7699
O3'	OS	-0.5407
C5'	CT	-0.0122
H5'1	H1	0.0791
H5'2	H1	0.0791
C4'	CT	0.1579
H4'	H1	0.1088
O4'	OS	-0.403
C1'	CT	0.1931
H1'	H2	0.0679
C3'	CT	0.1358
H3'	H1	0.0758
C2'	CT	-0.057
H2'1	HC	0.0457
H2'2	HC	0.0457
N9	N*	-0.0447
C8	CK	0.5148
N7	NB	-0.5657
C5	CB	0.1146
C6	C	0.5093
O6	O	-0.5442
N1	NA	-0.4523
H1	H	0.3452
C2	CA	0.6267
N2	N2	-0.8744
H21	H	0.3894
H22	H	0.3894
N3	NC	-0.5033
C4	CB	0.1036
N8	N2	-0.6715
H8	H	0.3746
C9	CA	0.369
C10	CA	-0.1778
H10	HA	0.1531
C11	CA	-0.3158
H11	HA	0.1754
C12	CA	-0.1902
H12	HA	0.1513
C13	CA	-0.1284
H13	HA	0.1387
C14	cp	0.0446
C15	cp	0.0704
C16	CA	-0.0807
H16	HA	0.1214
C17	CA	-0.4271
H17	HA	0.1965
C18	CA	-0.2676
H18	HA	0.1899
C19	CA	0.4444
H19	HA	-0.101
C20	CA	-0.1481
F	F	-0.0416

Table 2:

	-ΔG[a] (kcal/mol)	-ΔH (kcal/mol)	T_m[b] (°C)	ΔΔG[c] (kcal/mol)	ΔΔH[d] (kcal/mol)	ΔT_m[e] (°C)
FABP-G1						
9 mer	8.60 (8.48)	64.28 (68.69)	35.04 (34.69)	0.11	-4.41	0.35
10 mer	10.82 (10.65)	76.48 (76.11)	44.46 (43.79)	0.17	0.38	0.67
11 mer	11.27 (12.06)	78.59 (83.70)	46.12 (48.66)	-0.79	-5.11	-2.54
12 mer	11.38 (12.99)	77.40 (89.10)	46.74 (51.46)	-1.61	-11.71	-4.72
13 mer	12.38 (15.50)	87.12 (105.85)	49.42 (57.22)	-3.12	-18.73	-7.80
16 mer	14.53 (17.82)	109.16 (121.68)	53.49 (61.16)	-3.29	-12.53	-7.68
FABP-G2						
9 mer	8.86 (8.48)	59.82 (68.69)	36.14 (34.69)	0.38	-8.87	1.45
10 mer	9.34 (10.65)	59.75 (76.11)	38.59 (43.79)	-1.31	-16.36	-5.21
11 mer	10.24 (12.06)	70.69 (83.70)	42.38 (48.66)	-1.83	-13.02	-6.28
12 mer	10.93 (12.99)	75.29 (89.10)	45.08 (51.46)	-2.07	-13.82	-6.38
13 mer	12.68 (15.50)	89.23 (105.85)	50.27 (57.22)	-2.82	-16.62	-6.95
16 mer	15.00 (17.82)	109.61 (121.68)	54.88 (61.16)	-2.82	-12.08	-6.29

[a]The average standard deviations for – ΔG, -ΔH and T_m are ± 0.3, ±6.3 and 0.5, respectively.

[b]T_m values taken from $1/T_m^{-1}$ - $\ln C_t/4$ Meltwin plots.

[c]ΔΔG = ΔG (modified duplex) - ΔG (control duplex).

[d]ΔΔH = ΔH (modified duplex) - ΔH (control duplex).

[e]ΔT_m = T_m (modified duplex) - T_m (control duplex).

Table 3:

	-ΔG[a] (kcal/mol)	-ΔH (kcal/mol)	T_m[b] (°C)	ΔΔG[c] (kcal/mol)	ΔΔH[d] (kcal/mol)	ΔT_m[e] (°C)
FAF-G1						
9 mer	9.19 (8.76)	70.22 (68.63)	37.76 (35.79)	0.44	1.60	1.97
10 mer	11.05 (10.53)	74.77 (75.26)	45.68 (43.33)	0.53	-0.48	2.36
11 mer	11.76 (12.34)	81.61 (83.87)	47.80 (48.98)	-0.58	-2.26	-1.18
12 mer	11.38 (12.62)	72.99 (89.41)	47.43 (50.93)	-1.24	-16.42	-3.51
13 mer	12.59 (15.35)	86.72 (103.74)	50.32 (57.15)	-2.76	-17.02	-6.83
16 mer	16.51 (18.81)	128.03 (131.70)	56.26 (61.91)	-2.30	-3.67	-5.65
FAF-G2						
9 mer	9.31 (8.76)	65.85 (68.63)	38.42 (35.79)	0.56	-2.78	2.63
10 mer	9.44 (10.53)	58.98 (75.26)	39.19 (43.33)	-1.09	-16.28	-4.14
11 mer	10.11 (12.34)	63.52 (83.87)	42.05 (48.98)	-2.23	-20.35	-6.93
12 mer	10.51 (12.62)	70.88 (89.41)	43.65 (50.93)	-2.11	-18.53	-7.29
13 mer	12.18 (15.35)	85.06 (103.74)	48.93 (57.15)	-3.17	-18.68	-8.23
16 mer	15.12 (18.81)	111.70 (131.70)	54.91 (61.91)	-3.69	-20.00	-7.00

[a]The average standard deviations for – ΔG, -ΔH and T_m are ± 0.14, ±4.8 and 0.4, respectively.

[b]T_m values taken from $1/T_m$ - $\ln C_t/4$ Meltwin plots.

[c]ΔΔG = ΔG (modified duplex) - ΔG (control duplex).

[d]ΔΔH = ΔH (modified duplex) - ΔH (control duplex).

[e]ΔT_m = T_m (modified duplex) - T_m (control duplex).

Table 4:

sequence	−ΔH° (kcal/mol)	−ΔS° (eu)	$-\Delta G_{37\,°C}$ (kcal/mol)	Tm[a] (°C)	ΔΔH°[b] (kcal/mol)	$\Delta\Delta G_{37\,°C}$[c] (kcal/mol)	ΔTm[d] (°C)
			FABP-G1				
9 mer	71.8 (71.4)	207.3 (206.1)	7.6 (7.5)	43.3 (42.9)	-0.4	-0.1	0.4
10 mer	72.1 (72.1)	202.5 (203.4)	9.3 (9.0)	51.3 (50.0)	0	-0.3	1.3
11 mer	69.6 (74.4)	194.8 (207.9)	9.2 (9.9)	51.5 (53.8)	4.8	0.7	-2.3
12 mer	70.0 (82.0)	195.8 (228.7)	9.3 (11.1)	51.6 (56.9)	12	1.8	-5.3
13 mer	71.6 (87.2)	199.5 (242.2)	9.7 (12.1)	53.4 (59.8)	15.6	2.4	-6.4
16 mer	90.6 (98.5)	258.3 (273.4)	10.5 (13.8)	52.8 (63.1)	7.9	3.3	-10.3
			FABP-G2				
9 mer	66.5 (71.4)	189.5 (206.1)	7.8 (7.5)	45.0 (42.9)	4.9	-0.3	2.1
10 mer	64.6 (72.1)	182.1 (203.4)	8.1 (9.0)	46.9 (50.0)	7.5	0.9	-3.1
11 mer	67.8 (74.4)	191.1 (207.9)	8.6 (9.9)	48.5 (53.8)	6.6	1.3	-5.3
12 mer	74.8 (82.0)	212.1(228.7)	9.1 (11.1)	49.7 (56.9)	7.2	2	-7.2
13 mer	84.4 (87.2)	239.2 (242.2)	10.2 (12.1)	52.8 (59.8)	2.8	1.9	-7
16 mer	96.7 (98.5)	273.0 (273.4)	12.1 (13.8)	57.3 (63.1)	1.8	1.7	-5.8
Standard deviation	±3	±8	±0.3	±0.3			

[a]Tm value is the maximum point of the DSC curve.

[b]ΔΔG = ΔG (modified duplex) - ΔG (control duplex).

[c]ΔΔH = ΔH (modified duplex) - ΔH (control duplex).

[d]ΔT_m = T_m (modified duplex) - T_m (control duplex).

Table 5:

Sequence	shape of binding isotherm (C)	reaction stoichiometry (N), number of sites	binding constants (K), $[M]^{-1}$ ($x10^7$)	Enthalpy(ΔH), kcal/mol	Entropy(ΔS), cal/K▪mol
unmodified	1240	0.99±0.0024	49.9±7.58	-113.3±0.653	-340
FABP-G1	505	1.08±0.0117	18.7±8.14	-97.82±2.025	-290
FABP-G2	135	0.92±0.0083	5.9±1.15	-98.44±1.627	-295
FAF-G1	635	0.98±0.0093	25.9±11.1	-93.23±1.735	-274
FAF-G2	476	1.04±0.0043	18.3±3.01	-94.26±0.822	-278

Table 6:

		FABP			FAF		
sequences	**[C] (nM)**	**unmodified**	**G_1**	**G_2**	**unmodified**	**G_1**	**G_2**
9mer	300	16.5	12	7.1	28	13.6	9
10mer	500	0.515	0.343	2.37	0.883	0.384	2.81
11mer	500	0.072	0.216	0.769	0.185	0.321	1.96
12mer	750	0.028	0.139	0.361	0.083	0.261	1.29
13mer	750	0.02	0.073	0.122	0.041	0.169	0.382
16mer	750	0.013	0.049	0.041	0.021	0.082	0.086

MANUSCRIPT II

(to be submitted to *Journal of Nature Structural & Molecular Biology*)

Probing the Effect of Bulky Lesion-Induced Replication Fork Conformational Heterogeneity Using 4-Aminobiphenyl-Modified DNA

Ang Cai, † Ke Bian, † Bongsup P. Cho, and Deyu Li

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

Author Contributions

† A.C. and K.B. contributed equally to this work.

Corresponding Author

*E-mail: deyuli@uri.edu

Present address:

Department of Biomedical and Pharmaceutical Sciences,

University of Rhode Island

7 Greenhouse Road,

Kingston, RI, 02881, USA

ABSTRACT

Bulky organic carcinogens are usually activated *in vivo* and form DNA adducts. Some of these DNA adducts exist in multiple conformations that are slowly interconverted to one another. For example, the dG-C8-lesion of the human bladder carcinogen 4-aminobiphenyl (ABP) has been reported to exist in a sequence-dependent mixture of B/S conformations. Different conformations have been shown to lead to different kinetic capacities, binding capacities, and ultimately different translesion synthesis and transcription capacities. Although rare, conformation-dependent mutagenic and repair patterns have been reported. For example, the bulky N-acetylaminofluorene at G_3 in the *Nar*I sequence (5′-$G_1G_2CG_3CN$-3′) induces a high frequency of −2 frameshift mutations, whereas the adducts on G_1 and G_2 are not mutagenic. Results show why mutagenic hotspots exist in the genome. However, the conformation-specific replication block, which is more relevant to replication, has not been reported. Such research deficiency may be due to the dynamicity of the replication block process. As such, the replication block becomes difficult to capture by current end-point assays, which usually detect either the ensemble of consequence of all the conformers or the culmination of all cellular behaviors, such as mutagenicity or survival rate. In Manuscript I, we showed very unusual sequence-dependent conformational heterogeneities involving FABP-modified DNA under different sequence contexts (TG_1*G_2T [67%B:33%S] and TG_1G_2*T [100%B], G*, FABP, N-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl). Different biological consequences are believed to be linked to different adducts. In the present study, we attempted to correlate the replication block to different conformations from the single FABP-modified DNA lesion. We utilized a combination

of biophysical (SPR and steady-state kinetics) tools to reveal the differences *in vitro*. The results were discussed in terms of conformational differences.

INTRODUCTION

The human genome is under constant assault from the environment. DNA-damage-induced mutations are known to trigger chemical carcinogenesis[1-2]. Therefore, understanding the mechanisms of DNA mutations is important. Arylamines belong to a notorious group of environmental mutagens/carcinogens that produce bulky DNA adducts *in vivo*[3]. These lesions have been well documented to adopt the following three unique conformational motifs: the major groove B-type (B), stacked (S), or minor groove wedge (W). As such, the lesions complicate mutational and repair outcomes (Figure 1a)[4].

4-Aminobiphenyl (ABP) is a well-known human bladder carcinogen that forms preferentially dG-C8-substituted adducts (dG-C8-ABP) (Figure 1b)[5-7]. Our structural studies (Manuscript I) showed that fully paired duplexes containing dG-C8-FABP (fluorine-labeled ABP, FABP, N-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl) adopted a 67%:33% mixture of the B- and S-conformers in the TG_1*G_2T (G*=FABP) sequence context at 25 °C. Meanwhile, the same lesion in the TG_1G_2*T context exhibited exclusively the B-type conformation under identical experimental conditions. When a replication polymerase encounters a bulky DNA lesion, the polymerase likely stalls or completely stops DNA synthesis. In this case, the lesion influences whether the DNA replication will be error free or error prone. The aforementioned striking sequence effect warrants systematic polymerase binding studies. Vaidyanathan et al.[8] probed into the sequence effect of dG-C8-AF, the ABP analog, on nucleotide insertion efficiencies catalyzed by the Klenow fragment (Kf-exo$^-$) on TG*A and CG*A sequences. The

researchers found that the S conformer of CG*A thermodynamically favors the insertion of A over C at the lesion site. Xu et al.[9] reported that Kf-exo$^-$ is a strong binder to template/primer junctions, but with minimal nucleotide selectivity against modified DNA. The sequence-dependent conformational heterogeneity may play an important role in DNA replication and mutation. In the present study, we hypothesize that different conformations lead to different extents of polymerase binding affinities, kinetics, and replication blocks, ultimately resulting in complex mutational outcomes. We also conducted SPR and steady-state enzyme kinetics *in vitro* to probe into the effect of ABP-induced conformational heterogeneity.

MATERIALS AND METHODS

Model Oligonucleotide DNA sequences.

FABP-modified 16-mer templates (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) were prepared in accordance with the published procedures[10]: two mono-adducts (d(5′-CTTCTG$_1$*G$_2$TCCTCATTC-3′)) (G*=FABP) (G$_1$* adduct) and (d(5′-CTTCTG$_1$G$_2$*TCCTCATTC-3′)) (G$_2$* adduct). 5′-biotin-labeled 31-mer oligonucleotides, 5′-phosphorylated and 3′-dideoxy-A 53-mer, 5′-phosphorylated and 3′-dideoxy-C 54-mer were purchased from IDT (Integrated DNA Technologies, Inc. Coralville, IA) in desalted form and purified by reverse-phase high-performance liquid chromatography (RP-HPLC). Kf-exo$^-$ (D355A, E357A) was purchased from NEB Inc. (Ipswich, MA). All high-performance liquid chromatography (HPLC) solvents were purchased from Fisher Inc. (Pittsburgh, PA). The modified oligonucleotides were purified by HPLC (Thermo Scientific, Madison, WI) on a Phenomenex Luna C18

analytical column (100 × 46 mm, 3.0 μm) (Phenomenex, Torrance, CA). For the FABP-modified biotinylated 31-mer sequences, a 0.2 mL/min flow rate was used in a 48 min linear gradient profile (2% to 26%, v/v) acetonitrile with 100 mM TEAA buffer (pH 7.0) in a mobile phase (Figure 2a).

HPLC-based Steady-State Kinetic Analysis

The extension efficiency of Kf-exo$^-$ polymerase opposite G_1* and G_2* adducts were determined by steady-state kinetic experiments as described by published reports[11-14]. All reactions were performed using Kf-exo$^-$ (100 nM) and primers (5 μM) in NEB buffer 2 (New England Biolabs, MA) at 22 °C. The primers (8-mer to 12-mer) were pre-annealed with 6.75 μM FABP-modified or unmodified 16-mer templates by heating up to 70 °C and slowly cooling down to room temperature. Five different concentrations of dCTP or dATP (0, 12.5, 25, 50, and 100 μM) were used to initialize nucleotide insertion. All reactions were quenched by adding ethylene diamine tetraacetic acid followed by immediately denaturing at 80 °C. The initial velocity of each reaction within the steady-state range was then obtained by performing every reaction within a short time period from 5 s to 30 s. The extended and unextended primers were separated by anion-exchange HPLC at 40%–60% B, 6.5 min gradient conditions (A: water and B: 1.5 M ammonium acetate) and quantified by absorbance of UV at 260 nm. All reactions were performed at 22 °C in triplicate, and the results were analyzed through the GraphPad Prism 5 software by using the Lineweaver–Burke model. The inhibition curves were fitted to the equation $1/V = 1/V_{max} + (K_M/V_{max})/[S_{dCTP}]$. The relative insertion efficiency f_{ins} was obtained as $(k_{cat}/K_M)_{modified}/(k_{cat}/K_M)_{unmodified}$.

SPR Measurements

FABP-modified Hairpin Template/Primer Constructs. 5′-Biotinylated 31-mer containing dG-C8-FABP in the "-TG_1G_2T-" sequence context was used in SPR analysis following the reported procedures[9, 15-16]. The two FABP-modified biotinylated 31-mer G_1* and G_2* adducts were separated by RP-HPLC and characterized by MALDI-TOF mass spectrometry (MS)[17]. An 84-mer and an 85-mer hairpin–template–primer were prepared by following the reported protocols[18]. Briefly, two different lengths of hairpin DNA sequences (53- and 54-mer) were prepared. These two hairpin sequences were phosphorylated at their 5′-terminals, but their 3′-terminals were modified with ddA and ddC, respectively, to prevent further primer elongation (Figure 3a). The biotinylated 31-mer modified G_1* adduct and 54-mer hairpin were desalted by G-25 spin columns and annealed together by heating to 95 °C for 5 min and cooling to room temperature. The mixture solution was ligated in a buffer containing 4,000 units of T4 ligase enzyme for 16 h at room temperature. The resulting 85-mer oligonucleotides were purified in 15% denaturing polyacrylamide gel and extracted using the electroelution method followed by desalting by again using G-25 spin columns. The pure biotinylated 85-mer was finally purified by RP-HPLC. G_2* adduct biotinylated 84-mer and unmodified biotinylated 84- and 85-mers were also prepared similarly. All the 84- and 85-mer oligonucleotides were identified by the 15% denaturing polyacrylamide gel (Figure 4).

Immobilization of Streptavidin (SA) on CM5 S Chip and DNA Coating

SPR experiments were carried out using Biacore T200 (GE Healthcare, Piscataway, NJ). The SA via the amine coupling kit was immobilized on flow cells on the carboxymethylated dextran-coated CM5 S chip by following the reported procedures[9, 16]. After SA immobilization at around 2,000 RU on the flow cells, the chip surface was washed with 50 mM NaOH for 60 s five times to reach below 20 RU. Then, the running buffer was injected around three times, and the system was equilibrated with running buffer for ~2 h. The 84- and 85-mer unmodified and modified G_1* and G_2* biotinylated DNA sequences (2 nM) were injected over the flow cells (2 to 4) for 90-120 s to achieve 5-6 RU. The surface was stabilized with running buffer for 3 h before conducting SPR binding affinity experiments.

Real-time Kinetic Analysis

Initially, the SPR system was primed at least three times with running buffer and zero-concentration injections to condition the plasmon surface. The DNA was coated on the SA surface of flow cells 2 to 4 (cell 1 as blank reference). Surface testing, regeneration buffer scouting, and mass transport limitation test were conducted prior to kinetics experiments following previous reports[9, 16]. DNA coating around 4-5 RU did not show any impact of mass transport. The steady-state affinity analysis of Kf-exo$^-$ binding to unmodified and modified DNA was analyzed in the absence (binary) and presence of dNTPs (ternary) by varying Kf-exo$^-$ concentrations.

For the binary system, Kf-exo$^-$ was injected without dNTPs over the DNA surface in varying concentrations (0-25 nM) and repeated twice as described by previous reports[9].

Briefly, the 1× HBS-P running buffer (containing 100 μg/mL bovine serum albumin [BSA] and 5 mM $MgCl_2$) with six different concentrations Kf-exo$^-$ (0-25 nM) was injected for 30 s at a flow rate of 100 μL/min over the cells. Afterward 0.05% SDS was applied as regeneration solution at a flow rate of 100 μL/min over the surface for 30 s. The surface was stabilized with running buffer for 15 min after a regeneration step between different concentration circles.

For the ternary system, dNTPs were injected with the HBS-P running buffer for comparison. 1× HBS-P running buffer (100 μg/mL BSA, 5 mM $MgCl_2$, and 100 μM individual dNTPs) with varying concentrations of Kf-exo$^-$ was injected over the surface. For both binary and ternary systems, sensorgrams were fitted by using a 1:1 Langmuir model, and the binding affinity constants (K_D) were determined using steady-state affinity analysis in the BIA evaluation software v 1.0.

RESULTS

Model Systems

For SPR binding experiments, we constructed two biotinylated hairpin-based template–primer strands (Figure 3a), 85-mer G_1* adduct and 84-mer G_2* adduct (G* = FABP). Specifically, two FABP-modified–biotin–31-mer DNA sequences (TG_1*G_2T and TG_1G_2*T) were separated, purified, and characterized individually by enzyme digestion/MALDI-MS[9] (see below). The 54-mer hairpin DNA was annealed and ligated to the biotin–31-mer TG_1*G_2T oligonucleotide, whereas the 53-mer hairpin DNA was ligated to the biotin–31-mer TG_1G_2*T oligonucleotide. The hairpin structures were used

to improve the stability of the oligonucleotide during SPR experiments and the lesion G_1* and G_2* adduct were at the 21st and 22nd bases, respectively, to avoid clash between the polymerase protein and the gold chip surface[9].

For steady-state kinetics experiments, two 16-mer G_1* or G_2* template strands were each annealed to various lengths of complementary strands (8-mer to 11-mer) to create n-3, n-2, n-1, and n for the G_1* adduct and n-2, n-1, n, and n+1 for G_2* (n is the lesion site) (see Manuscript I).

MALDI-TOF Characterization

The FABP-modified biotin–31-mer oligos were characterized by exonuclease enzyme digestion followed by MALDI-TOF MS in accordance with the published procedures[9, 17]. In the present case, 5′-3′ exonuclease digestion was not possible because of the biotin group present at the 5′ end of the DNA. Figure 2b shows the MALDI-TOF MS spectra of the 3′-5′ snake venom phosphodiesterase (SVP) exonuclease digestions of the biotin–31-mer G_1* adduct at different time points (0, 1, 4, 6, 7, 8, and 10 min). The m/z of 9,804 (theoretical 9,803) at 0 min represents the control mass-to-charge ratio. Within 6 min of 3′-5′ exonuclease digestion, the lower masses appeared corresponding to the 27-mer to 21-mer fragments. The m/z = 6,788 (theoretical 6,787) fragment, which persisted in 6 min to 10 min was assigned to the G_1*-FABP-modified 21-mer. These results confirmed the first eluting peak (peak 1) from the HPLC profile (Figure 2b) as the biotin–31-mer TG_1*G_2T. Figure 2c presents the MALDI-TOF MS spectra of the HPLC peak 2 with 3′-5′ exonuclease digestions. The digestions were fast in the first 4 min

showing the m/z range from 9,803 to 7,116. However, the reaction stalled at m/z 7,116 from 4-10 min corresponding to the 22-mer fragment containing the FABP-modified guanine (theoretical m/z = 7,116). These results confirm peak 2 as the G_2*-FABP-modified biotin–31-mer sequence.

The 84- and 85-mer biotinylated oligonucleotides were purified and identified by 15% denaturing polyacrylamide gel (Hercules, CA). Figure 4 reveals the denaturing gel profiles of unmodified/modified 84- and 85-mer ligated oligonucleotides, biotin–31-mer, and 53-mer/54-mer non-ligated oligonucleotides. For the 85-mer control, all the biotin–31-mer and 54-mer hairpins were ligated. As for the 84-mer control, 85-mer G_1*, and 84-mer G_2*, excessive biotin–31-mer, 54-mer, and 53-mer were observed. The ligated 85-mer control/ G_1* and 84-mer control/ G_2* were employed for SPR experiments.

HPLC-based Steady-state Kinetics

We conducted steady-state experiments to investigate the impact of conformational heterogeneity on nucleotide insertion kinetics[19]. The *E. coli* exonuclease-deficient Kf-exo$^-$ was used for single-nucleotide incorporation, and the results are summarized in Table 1. dATP was added to 16/8-mer and 16/11-mer systems, whereas dCTP was added to 16/9-mer and 16/10-mer sequences.

dCTP: Specifically, adding dCTP to the n-1 (16/9-mer) G_2* adduct elongated the oligomer to 10- and 11-mer, whereas the n-1 (16/10-mer) G_1* adduct produced a single

11-mer. The 10- and 11-mer mixture from the G_2* adduct sequence was difficult to quantify; thus, reduction of the starting 9-mer was employed to determine the kinetics analysis.

The n-1 (16/9-mer) G_2* adduct showed a K_m value (5.7 μM) similar to that of the control (5.8 μM); such similarity indicates equal affinity (Table 1). However, the insertion efficiency f_{ins} of dCTP opposite -G_2* was 5-fold lower than that of the unmodified control (0.24 :1) (Table 1). By contrast, the K_m of the G_1* adduct showed a lower affinity (23.4 μM) than that of the control (7.4 μM). The insertion efficiency f_{ins} in G_1* was three times lower than the control (0.33 to 1). As shown by the Lineweaver–Burke enzyme inhibition model (Figure 5a), the (16/10-mer) G_1* adduct and its control intersected on the Y-axis. A competitive inhibition behavior was indicated. By contrast, the n-1 16/9-mer G_2* adduct and its control merged at the x-axis. This behavior can only be explained as non-competitive inhibition (Figure 5b). When the n-1 (16/9-mer) G_2* adduct elongated to 10-mer (n) and 11-mer (n+1), few 10-mer primers but most of the 11-mer primers were observed. This result was achieved because the bulky FABP on the G_2 at the lesion site (n) did not block the replication from n-1 to n+1.

dATP: The conformational and long-range lesion effect was examined by initiating the primer elongation with a 16/8-mer template/primer that produces n-2 for G_2* and n-3 for G_1*. Figure 6a presents the results at 10 min. In the control, ~70% of the 8-mer was extended to 9-mer. With adduct on G_1, ~40% of the 8-mer was converted to 9-mer. This result is a moderate pre-lesion effect at G_1*. However, G_2* blocked ~80% of the 8-mer

converting to 9-mer; this effect indicates a much stronger pre-lesion effect than that of G_1* This result is not surprising because the 8-mer primer is closer to the G_2* adduct than the G_1* adduct. In the 11-mer (Figure 6b), unmodified control exhibited 90% of dATP insertion efficiency. However, only ~20% of the 11-mer was elongated to 12-mer opposite the G_1* lesion. This finding indicates a strong post-lesion effect. Moreover, only ~55% of the 11-mer was converted to 12-mer in the n+1 G_2*. These results suggest the significant retardation of insertion rates close to the lesion site.

Kf-exo$^-$ SPR binding Kinetics

Binary System—Figure 3b shows the sensorgrams for the binary binding between Kf-exo$^-$ and unmodified controls (85-mer/84-mer TGGT) or modified sequences (85-mer TG_1*G_2T/84-mer TG_1G_2*T). These sequences represent a true replication fork for the G_1* and G_2*, respectively. Kf-exo$^-$ showed a much stronger binding to the modified 85-mer TG_1*G_2T and 84-mer TG_1G_2*T sequences than to their unmodified controls (85-mer/84-mer TGGT), respectively (Table 2). Specifically, the binding affinity (K_D) of Kf-exo$^-$ in the 85- and 84-mer was reduced by 8.3- and 5.6-fold, respectively, for the G_1* and G_2* adducts related to the controls. The differences were more striking in the dissociation rates (85-mer TGGT k_d: 0.84 s^{-1} vs. 85-mer TG_1*G_2T k_d: 0.17 s^{-1}; 84-mer TGGT k_d: 0.03 s^{-1} vs. 84-mer TG_1G_2*T k_d: 0.009 s^{-1}). The K_d ratio (K_d-modified/K_d-control) between G_1 and G_2 was (20%:30%) showing 1.5 times difference (Figure 3c). In contrast to K_d, the K_a ratios of G_1 and G_2 were close to each other. This closeness indicates similar association rates. The net stabilization energies for the binary system were positive (Table 3). Although only one base length (dC) differed between the 84-

mer and 85-mer, the binding affinity of the unmodified 84-mer was 18.2-fold tighter than that of the 85-mer. Similarly, the K_D of the 84-mer TG_1G_2*T was 12.2-fold greater than that of the 85-mer TG_1*G_2T.

Ternary System— Kf-exo$^-$ bound tightly to the unmodified controls when the correct dCTP was introduced in the ternary system (Figure 7). In the unmodified 85-mer, the binding affinities of dATP and dGTP were exceedingly weak to be measured, and the K_D for the dTTP was reduced by 234-fold relative to the correct dCTP. The binding affinities of the dATP, dGTP, and dTTP to the 84-mer controls were reduced by 315, 284, and 89 fold, respectively, relative to that of the dCTP (Table 2). However, nucleotide selectivity was extremely low for the modified ternary complexes. The K_D value for the 85-mer TG_1*G_2T was only reduced by ~5-fold in the dATP, dGTP, and dTTP than that of dCTP. Similar results were obtained for the 84-mer G_2* adduct, where binding tightness was reduced by ~2-fold in the dATP, dGTP, and dTTP compared with that of dCTP. Moreover, as in the binary system, slower dissociation rates were observed for the unmodified and G_2* modified 84-mers. The K_D values of dCTP, dATP, dGTP, and dTTP in the unmodified and G_2* modified 84-mers were smaller than those of the unmodified 85-mer (5- and 13-fold smaller in dCTP and dTTP) and G_1* modified 85-mer TG_1*G_2T (4-, 13-, 10-, and 12-fold smaller in dCTP, dATP, dGTP, and dTTP, respectively).

DISCUSSION

In Manuscript I, we conducted systematic ^{19}F-NMR, CD, DSC, and SPR studies for the elongation of the 16-mer G_1*G_2T and G_1G_2*T (G*=FABP) template sequences. These model systems mimic a TLS process involving the bulky FABP lesion in two different conformational states, that is, G_1 and G_2 modification resulted in 67%:33% of B:S and 100% B population ratios, respectively. These sequence-dependent conformational complexities appeared to persist at various elongation positions, including the replication fork. The B-conformer is a major thermodynamic stabilizer in duplex settings, whereas the S-conformer is a destabilizer. The opposite was the case for ss/ds duplex sequences. In particular, the S-conformation promotes lesion stacking with nascent base pairs at the replication fork. SPR results reveal that the S-conformation increased the binding affinity with the complementary strands in the order of G_1* > G_2*. In the present work, we examined the effects of these unusual sequence effects on polymerase binding (binary and ternary) and nucleotide insertion kinetics.

Improvement in Model Hairpin Oligonucleotide Construction

We previously reported a construction of biotinylated hairpin-based template–primer strand for DNA–polymerase SPR binding studies.[9, 15-16] The general strategy is to ligate a biotinylated arylamine-modified 31-mer sequence with 52-mer hairpin DNA to form an 83-mer hairpin, followed by adding a ddNTP to the 3′-end to prevent primer elongation. However, ddNTP addition is low yielding, and the separation of ligated products is difficult. In the current study, successfully directly ligated a biotinylated modified 31-mer with ddNTP-containing 53- and 54-mer hairpin DNA. This process improved the yield (from ~10% to ~ 30%), and the products were readily separated.

Previously, individual dNTPs (100 μM) were mixed with varying concentrations of Kf-exo⁻ in sample buffer and injected over the surface without any dNTP in the running buffer. In this experiment, two variables were involved, that is, dNTPs and Kf–exo⁻[16]. However, in the updated buffer system, each dNTP was prepared in running buffer. Thus, only one variable, Kf–exo⁻, was introduced onto the chip surface. This change increased the system stability and accuracy in simulating a TLS system.

Lesion and Sequence Effects on Binding Affinities and Kinetics

Tight binding of Kf–exo⁻ with native unmodified dG was observed when the correct nucleotide dCTP was presented in both 85-mer and 84-mer. This result strongly agreed with our expectation that showed high nucleotide selectivity ($K_{D\ dCTP} \ll K_{D\ dTTP} < K_{D\ dATP} \sim K_{D\ dGTP}$). Meanwhile, an exceptionally tighter binding of Kf–exo⁻ was found for the FABP-modified sample relative to the control, i.e., the K_D of the 85-mer G_1* and 84-mer G_2* interactions were 8.3 and 5.6 fold higher than the corresponding native controls, respectively. Adduct-induced tighter binding affinity with Kf–exo⁻ has been reported[9, 20] and may be due to the interactions between the bulky FABP and the nearby hydrophobic amino-acid residues in the active site of the Kf-exo⁻. In the ternary system, nucleotide selectivity was low. In particular, the K_D of the dCTP at the opposite G_1* and G_2* was only ~5- and ~2-fold tighter than the incorrect dATP, dGTP, and dTTP, respectively. This result is a strong lesion effect. We observed that the usual 1:1 model fittings did not achieve a clean fit of dCTP for G_1* in the ternary system. This result might be due to the S/B conformational heterogeneity observed at G_1*. Our steady-state kinetics results indicated the S/B-conformeric G_1* was involved in competitive

inhibition. In other words, the S/B conformation at G_1* was not as accommodating possibly because of some unfavorable interaction of the bulky FABP lesion with the incoming dCTP in a competitive manner. The unfavorable interactions may be caused by the competition of S-conformer with dCTP opposite of the guanine. However, the B-conformeric G_2* accommodated an incoming nucleotide, and this situation resulted in non-competitive inhibition. The equal affinity (K_m) between G_2* and its control revealed that the B-conformation in the major grove may not affect the dCTP insertion process. This finding might explain the greater K_D of dCTP at G_1* over G_2* and indicated a weaker binding affinity. As another possibility, the B-conformer did not interfere with the Watson–Crick (W–C) base pair, but the adduct in the B-conformation may have altered the native conformations or retarded the formation of phosphate diester bond. G_2* operated as a non-competitive inhibitor. Meanwhile, G_1* is a B- and S-conformeric mixture. The S-conformer cannot provide a perfect W–C base pair for regulation and replication. The S-conformation may need to convert back to the B-conformation to be replicated. Thus, the S-conformer may function as a competitive inhibitor for replication and block the replication. We observed a competitive-inhibition model for G_1* but a non-competitive-inhibition model for G_2*. These types of intermediate interactions may be a requisite step in the DNA polymerase proofreading mechanism.

SUMMARY

We carried out SPR and steady-state kinetics *in vitro* studies for the replication block with different conformations. In Manuscript I, the FABP-modified DNA achieved a

mixture of B and S conformations in the TG_1*G_2T (67%B:33%S) and exclusively B in the TG_1G_2*T sequence context. In our present work, as indicated by the Lineweaver–Burke enzyme inhibition model, the (16/10-mer) G_1* adduct exhibited a competitive-inhibition behavior, whereas the G_2* adduct served as a non-competitive inhibitor in the nucleotide insertion kinetics. These results revealed that the S-conformer may not have accommodated incoming dCTPs and exhibited a competitive behavior with incoming nucleotides and blocked the replication. By contrast, the B-conformer did not interfere with the W–C base pairing nor ultimately affected the dCTP insertion. As such, the B-conformer showed non-competitive-inhibition behavior. The SPR findings also implicated an adduct-induced tightened binding affinity with Kf-exo$^-$ in a binary system that is consistent with our previous reports. Moreover, in a ternary system, nucleotide selectivity diminished when FABP modified G_1 and G_2. We observed the effect of bulky lesion-induced replication fork conformational heterogeneity. Future studies would include the bypass replication in bacterial and in mammalian cell systems.

REFERENCES

1. Voulgaridou, G. P.; Anestopoulos, I.; Franco, R.; Panayiotidis, M. I.; Pappa, A., DNA damage induced by endogenous aldehydes: current state of knowledge. *Mutat Res* **2011,** *711* (1-2), 13-27.
2. Lange, S. S.; Takata, K.; Wood, R. D., DNA polymerases and cancer. *Nat Rev Cancer* **2011,** *11* (2), 96-110.
3. Reid, T. M.; Lee, M. S.; King, C. M., Mutagenesis by site-specific arylamine adducts in plasmid DNA: enhancing replication of the adducted strand alters mutation frequency. *Biochemistry* **1990,** *29* (26), 6153-61.
4. Patnaik, S.; Cho, B. P., Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity. *Chemical Research in Toxicology* **2010,** *23* (11), 1650-1652.
5. Meneni, S. R.; Shell, S. M.; Gao, L.; Jurecka, P.; Lee, W.; Sponer, J.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Spectroscopic and Theoretical Insights into Sequence Effects of Aminofluorene-Induced Conformational Heterogeneity and Nucleotide Excision Repair†,∇. *Biochemistry* **2007,** *46* (40), 11263-11278.
6. Tsuneoka, Y.; Dalton, T. P.; Miller, M. L.; Clay, C. D.; Shertzer, H. G.; Talaska, G.; Medvedovic, M.; Nebert, D. W., 4-aminobiphenyl-induced liver and urinary bladder DNA adduct formation in Cyp1a2(-/-) and Cyp1a2(+/+) mice. *J Natl Cancer Inst* **2003,** *95* (16), 1227-37.
7. Swaminathan, S.; Hatcher, J. F., Identification of new DNA adducts in human bladder epithelia exposed to the proximate metabolite of 4-aminobiphenyl using 32P-postlabeling method. *Chem Biol Interact* **2002,** *139* (2), 199-213.
8. Vaidyanathan, V. G.; Cho, B. P., Sequence effects on translesion synthesis of an aminofluorene-DNA adduct: conformational, thermodynamic, and primer extension kinetic studies. *Biochemistry* **2012,** *51* (9), 1983-95.
9. Xu, L.; Vaidyanathan, V. G.; Cho, B. P., Real-Time Surface Plasmon Resonance Study of Biomolecular Interactions between Polymerase and Bulky Mutagenic DNA Lesions. *Chemical Research in Toxicology* **2014,** *27* (10), 1796-1807.
10. Jain, N.; Li, Y.; Zhang, L.; Meneni, S. R.; Cho, B. P., Probing the sequence effects on NarI-induced -2 frameshift mutagenesis by dynamic 19F NMR, UV, and CD spectroscopy. *Biochemistry* **2007,** *46* (46), 13310-21.
11. Xu, L.; Vaidyanathan, V. G.; Cho, B. P., Real-time surface plasmon resonance study of biomolecular interactions between polymerase and bulky mutagenic DNA lesions. *Chem Res Toxicol* **2014,** *27* (10), 1796-807.
12. Bian, K.; Chen, F.; Humulock, Z. T.; Tang, Q.; Li, D., Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition. *Chem Res Toxicol* **2017,** *30* (10), 1794-1796.
13. Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D., Oncometabolites d- and l-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions. *Chem Res Toxicol* **2017,** *30* (4), 1102-1110.
14. Krüger, S.; Meier, C., Synthesis of Site-Specific Damaged DNA Strands by 8-(Acetylarylamino)-2′-deoxyguanosine Adducts and Effects on Various DNA Polymerases. *European Journal of Organic Chemistry* **2013,** *2013* (6), 1158-1169.
15. Vaidyanathan, V. G.; Liang, F.; Beard, W. A.; Shock, D. D.; Wilson, S. H.; Cho, B. P., Insights into the conformation of aminofluorene-deoxyguanine adduct in a DNA polymerase active site. *J Biol Chem* **2013,** *288* (32), 23573-85.

16. Vaidyanathan, V. G.; Xu, L.; Cho, B. P., 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* (8), 1568-1570.
17. Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D., Jr.; Zou, Y.; Cho, B. P., Structural and thermodynamic insight into Escherichia coli UvrABC-mediated incision of cluster diacetylaminofluorene adducts on the NarI sequence. *Chem Res Toxicol* **2013,** *26* (8), 1251-62.
18. Vaidyanathan, V. G.; Xu, L.; Cho, B. P., Binding kinetics of DNA-protein interaction using surface plasmon resonance. **2013**.
19. Vrzheshch, P. V., Steady-state kinetics of bifunctional enzymes. Taking into account kinetic hierarchy of fast and slow catalytic cycles in a generalized model. *Biochemistry (Mosc)* **2007,** *72* (9), 936-43.
20. Dzantiev, L.; Romano, L. J., 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* **1999,** *274* (6), 3279-84.

FIGURE LEGENDS

Figure 1. (a) The views of FABP-conformational motifs: B-, and S-conformers. ABP-red, modified dG-blue, complementary C-orange. (b) dG-C8-FABP.

Figure 2. (a) HPLC chromatogram of a reaction mixture of biotinylated-31-mer sequence (5'-Bio-CCTCTTCCCTCACCTCTTCTG$_1$G$_2$TCCTCATTC-3') with an activated **FABP** (N-Acetoxy-N-(trifluoroacetyl)-4'-fluoro-4-aminobiphenyl) and photodiode array UV spectra of unmodified, mono-adducts. (b) MALDI-TOF mass spectra of **FABP** modified biotinylated-31mer-TGGT: (a) 3'-enzyme digestion of **peak 1** at 0, 1, 4, 6, 7, 8 and 10min. (b) 3'-enzyme digestion of **peak 2** at 0, 1, 4, 6, 7, 8 and 10min. Insets show the theoretical MW of the corresponding fragments that should form after 3'-enzyme digestion.

Figure 3. (a) Hairpin template–primer oligonucleotide constructs of 85-mer G$_1$ and 84-mer G$_2$ for Kf-exo$^-$; (b) Sensorgrams of **binary** complexes of Kf-exo$^-$ with 85-mer and 84-mer unmodified control, G$_1$-FABP and G$_2$-FABP modified sequences (1:1 binding fitted curves are overlaid as black lines); (c) K_a (K_a-modified / K_a-contol) and K_d (K_d-modified / K_d-contol) ratio, G$_1$ and G$_2$ represent the ratio of 85-mer FABP-G$_1$/85-mer-control and 84-mer FABP-G$_2$/84-mer-control, respectively.

Figure 4. Denaturing gel (15%) profiles of 84-mer and 85-mer ligated oligonucleotides, 31-mer, 53-mer and 54-mer non-ligated oligonucleotides.

Figure 5. Lineweaver – Burk model of the DNA Repair Reactions Catalyzed by the Kf-exo$^-$ at (a) 10-mer, G$_1$-FABP and (b) 9-mer G$_2$-FABP lesion sites with dCTP at steady-

state. According to the model, the behavior of G_1-FABP against with Kf–exo$^-$ showed as a competitive inhibitor, whereas the behavior of TG_1G_2*T played as non-competitive.

Figure 6. dATP insertion efficiency results of (a) 8-mer to 9-mer and (b) 11-mer to 12-mer in control, G_1*G_2T-FABP and G_1G_2*T-FABP, respectively, in 10 min.

Figure 7. Sensorgrams of **ternary** Kf-exo$^-$ complexed with (a) 85-mer control, (b) 85-mer TG_1[FABP]G_2T, (c) 84-mer **control** and (d) 84-mer TG_1G_2[FABP]T sequences in the presence of dNTPs (1:1 binding fitted curves are overlaid as black lines).dATP insertion efficiency results of (a) 8-mer to 9-mer and (b) 11-mer to 12-mer in control, G_1*G_2T-FABP and G_1G_2*T-FABP, respectively, in 10 min.

TABLE LEGENDS

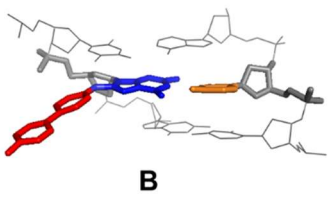
Table 1. Steady-state kinetics parameters for insertion of dCTP opposite unmodified and FABP−dG adduct with Kf-exo$^-$.

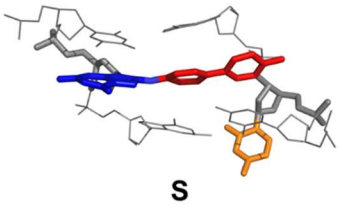
Table 2. SPR binding affinities, K_D (nM) of unmodified (TG_1G_2T) and FABP−dG adducts (TG_1[FABP]G_2T and TG_1G_2[FABP]T) with Kf-exo$^-$ (steady-state affinity analysis) in the binary and ternary systems. Association and dissociation rate constants, k_a and k_d, in binary system are listed.

Table 3. Binding net stabilization energy of unmodified and FABP-adducts with Kf-exo$^-$ in binary system (1:1 binding).

Figure 1:

(a)





(b)

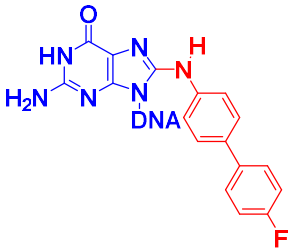
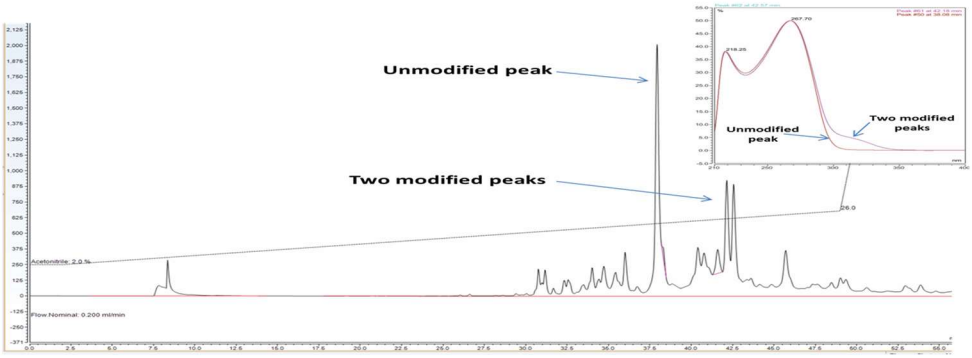


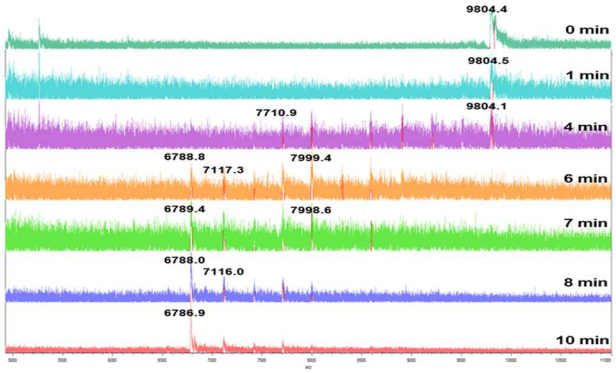
Figure 2:

(a)



(b)

mass	sequence
9803.1	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCTCATTC
9513.9	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCTCATT
9209.7	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCTCAT
8905.5	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCTCA
8592.3	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCTC
8303.1	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCCT
7998.9	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TCC
7709.7	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2TC
7420.6	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2T
7116.4	Bio-CCTCTTCCCTCACCTCTTCTG_1G_2
6787.3	Bio-CCTCTTCCCTCACCTCTTCTG_1
6271.1	Bio-CCTCTTCCCTCACCTCTTCT



(c)

mass	sequence
9803.1	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCTCATTC
9513.9	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCTCATT
9209.7	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCTCAT
8905.5	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCTCA
8592.3	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCTC
8303.1	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCCT
7998.9	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TCC
7709.7	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$TC
7420.6	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$T
7116.4	Bio-CCTCTTCCCTCACCTCTTCTG$_1G_2$
6787.3	Bio-CCTCTTCCCTCACCTCTTCTG$_1$
6271.1	Bio-CCTCTTCCCTCACCTCTTCT

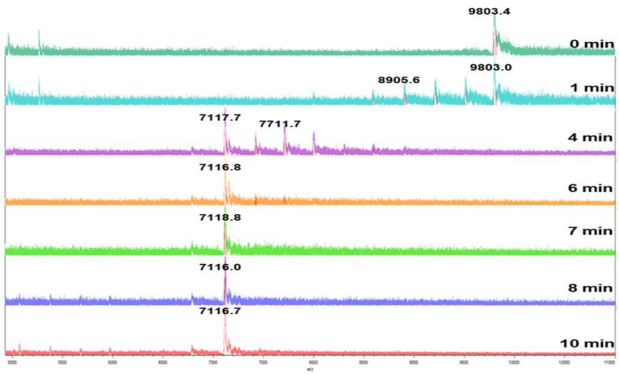
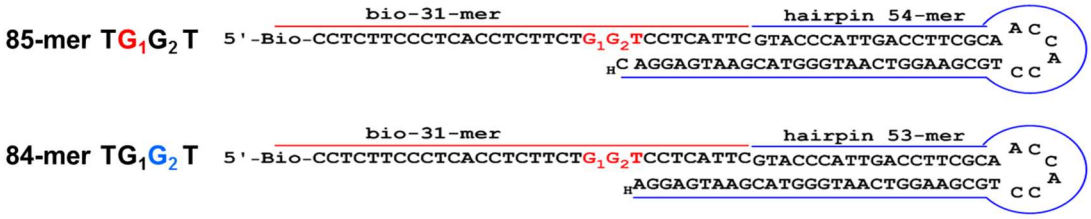
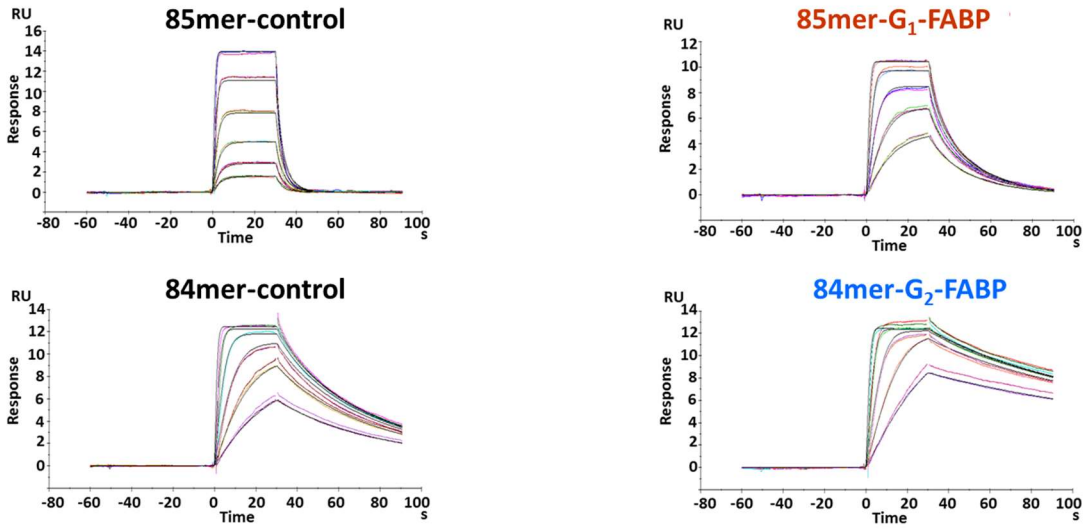


Figure 3:

(a)


85-mer TG1G2T
bio-31-mer
hairpin 54-mer
5'-Bio-CCTCTTCCCTCACCTCTTCTG1G2TCCTCATTCGTACCCATTGACCTTCGCA
AGGAGTAAGCATGGGTAACTGGAAGCGT
84-mer TG1G2T
bio-31-mer
hairpin 53-mer
5'-Bio-CCTCTTCCCTCACCTCTTCTG1G2TCCTCATTCGTACCCATTGACCTTCGCA
AGGAGTAAGCATGGGTAACTGGAAGCGT

(b)


85mer-control
85mer-G1-FABP
84mer-control
84mer-G2-FABP
RU
Response
Time
s

(c)

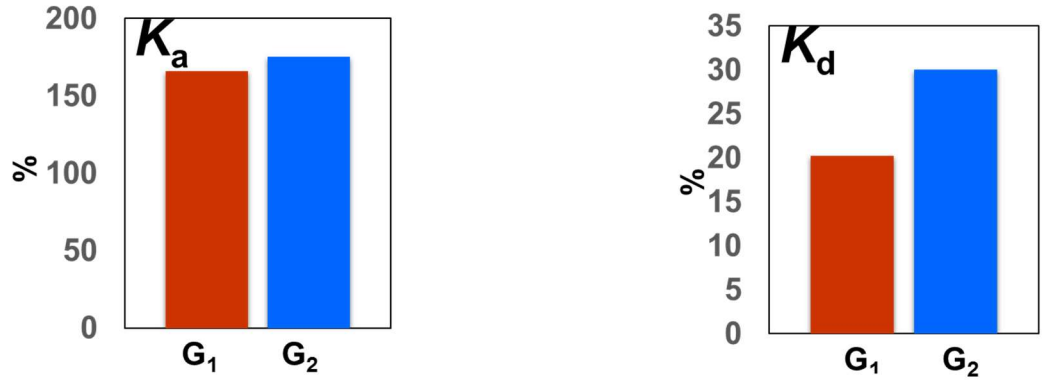

Ka
Kd
%
G1
G2

Figure 4:

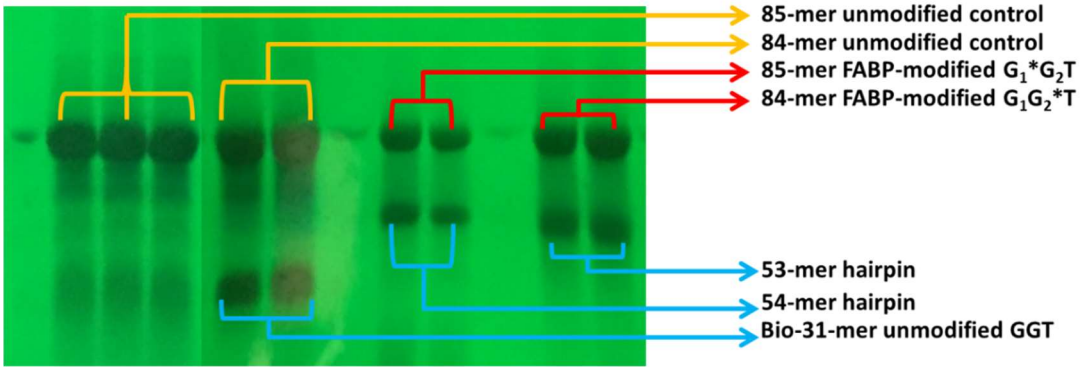
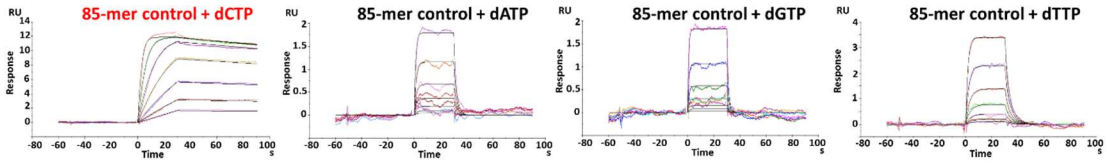
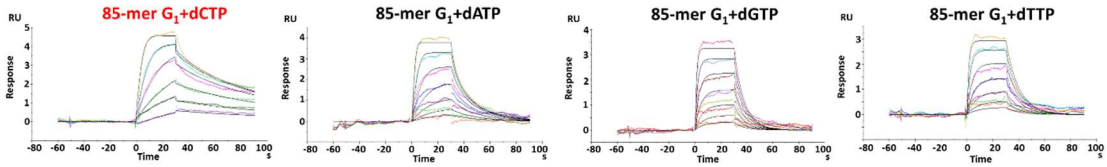

85-mer unmodified control
84-mer unmodified control
85-mer FABP-modified G_1*G_2T
84-mer FABP-modified G_1G_2*T
53-mer hairpin
54-mer hairpin
Bio-31-mer unmodified GGT

Figure 5:

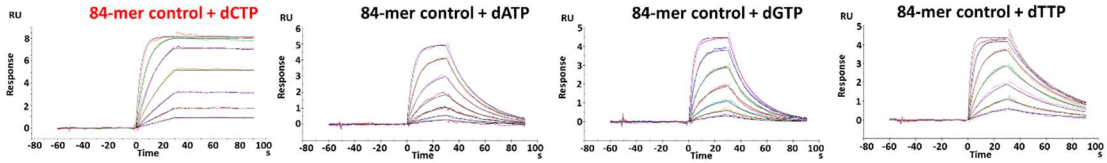
(a)



(b)



(c)



(d)

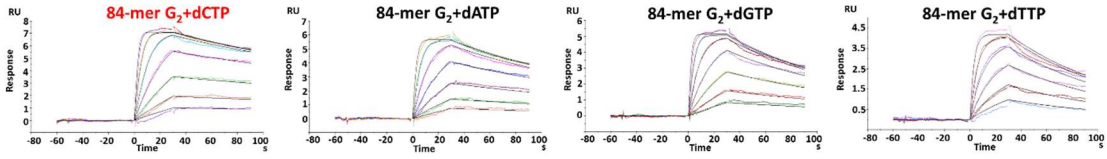


Figure 6:

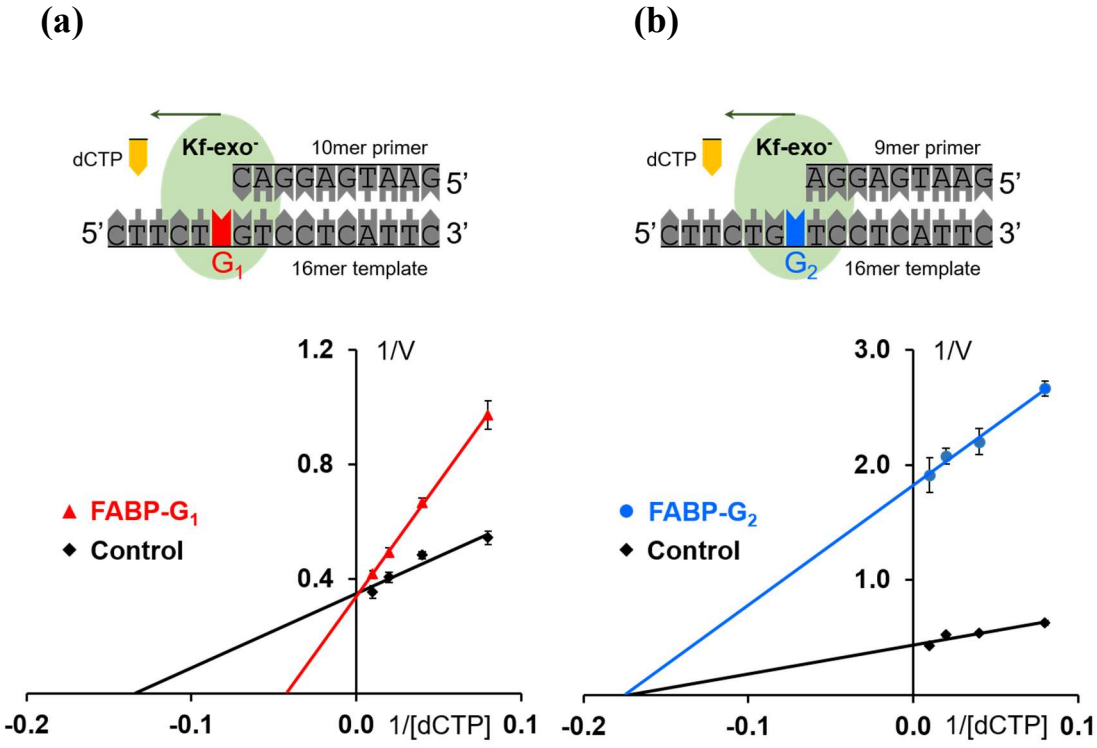

(a)
(b)
dCTP
Kf-exo-
10mer primer
CAGGAGTAAG 5'
5' CTTCTGTCCTCATTC 3'
G1
16mer template
dCTP
Kf-exo-
9mer primer
AGGAGTAAG 5'
5' CTTCTGTCCTCATTC 3'
G2
16mer template
1.2
0.8
0.4
1/V
FABP-G1
Control
-0.2
-0.1
0.0
1/[dCTP]
0.1
3.0
2.0
1.0
1/V
FABP-G2
Control
-0.2
-0.1
0.0
1/[dCTP]
0.1

Figure 7:

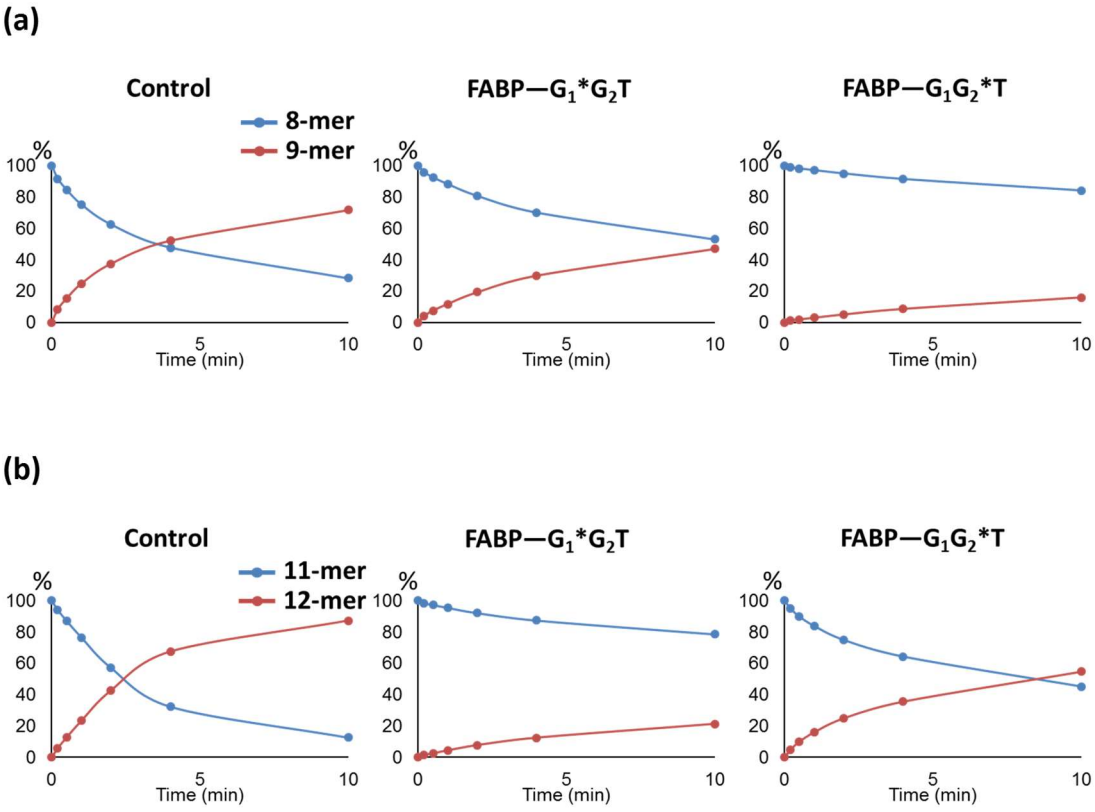

(a)
Control
FABP—G1*G2T
FABP—G1G2*T
8-mer
9-mer
%
100
80
60
40
20
0
0
5
10
Time (min)
(b)
Control
FABP—G1*G2T
FABP—G1G2*T
11-mer
12-mer
%
100
80
60
40
20
0
0
5
10
Time (min)

Table 1:

Sequence	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1} \cdot min^{-1}$)	$*f_{ins}$
FABP-G_1	23.4	29.7	1.27	0.33
Control-G_1	7.4	28.8	3.87	1.00
FABP-G_2	5.7	5.5	0.97	0.24
Control-G_2	5.8	23.2	4.00	1.00

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

Table 2:

Sequence	Binary			K_D of Ternary (nM)			
	K_a (1/Ms) x10^7	K_d (1/s)	K_D (nM)	dCTP	dATP	dGTP	dTTP
85-mer TG$_1$*G$_2$T	15.90 (0.17)	0.170 (0.002)	1.050 (0.050)	0.200 (0.060)	1.17 (0.04)	1.08 (0.07)	1.08 (0.08)
85-mer control	9.59 (0.05)	0.840 (0.004)	8.740 (0.030)	0.022 (0.001)	14.90 (5.00)	11.40 (4.20)	5.14 (0.74)
84-mer TG$_1$G$_2$*T	11.10 (0.07)	0.009 (0.000)	0.086 (0.001)	0.045 (0.006)	0.09 (0.00)	0.11 (0.00)	0.09 (0.01)
84-mer control	6.33 (0.04)	0.030 (0.002)	0.480 (0.030)	0.004 (0.000)	1.39 (0.02)	1.25 (0.01)	0.39 (0.04)

Table 3:

Sequence	k_d (1/s)	*Net stabilization Energy (*k*cal/mol)
85-mer control	0.84 (0.004)	0
85-mer TG_1[FABP]G_2T	0.17 (0.002)	0.95
84-mer control	0.03 (0.002)	0
84-mer TG_1G_2[FABP]T	0.009 (0.0001)	0.71

*Net stabilization energy (*k*cal/mol)= $-RT\ln(k_d)_{modified} - [-RT\ln(k_d)_{unmodified}]$

MANUSCRIPT III

(has been accepted by *Journal of ACS Omega*)

Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine

Qi Tang,[†,‡] Ang Cai,[†,‡] Ke Bian,[†] Fangyi Chen,[†,∥] James C. Delaney,[§] Sravani Adusumalli,[†] Alvin C. Bach, II,[†] Fatemeh Akhlaghi,[†] Bongsup P. Cho,[†] and Deyu Li[*,†]

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

[§]Visterra Inc., Cambridge, Massachusetts 02139, United States

[∥]Xiamen University, Xiamen, Fujian 361005, P. R. China

Author Contributions

[‡]Q.T. and A.C. contributed equally to this work.

Corresponding Author

*E-mail: deyuli@uri.edu

ABSTRACT

Oligonucleotides serve as important tools for biological, chemical, and medical research. The preparation of oligonucleotides through automated solid-phase synthesis is well established. However, identification of byproducts generated from DNA synthesis, especially from oligonucleotides containing site-specific modifications, is sometimes challenging. Typical HPLC, Mass Spectrometry (MS), and gel electrophoresis methods alone are not sufficient for characterizing unexpected byproducts, especially for those having identical or very similar molecular weight (MW) to the products. We used a rigorous quality control procedure to characterize byproducts generated during oligonucleotide syntheses: 1) purify oligonucleotide by different HPLC systems; 2) determine the exact MW by high resolution MS; 3) locate modification position by MS/MS or exonuclease digestion with MALDI-TOF analysis; and 4) conduct, where applicable, enzymatic assays. We applied these steps to characterize byproducts in the syntheses of oligonucleotides containing important methyl DNA adducts 1-methyladenine (m1A) and 3-methylcytosine (m3C). In m1A synthesis, we differentiated a regioisomeric byproduct 6-methyladenine, which possesses identical MW to m1A. As for m3C, we identified a deamination byproduct 3-methyluracil, which is only 1 Dalton greater than m3C in the ~ 4,900 Dalton context. The detection of these byproducts would be very challenging if the abovementioned procedure were not adopted.

INTRODUCTION

Oligonucleotides synthesized chemically are widely used as drugs and research tools in biology, chemistry, and medicine. Solid phase synthesis of DNA and RNA oligonucleotides has been well developed and automated phosphoramidite-based chemical process has become highly efficient.[1–3] Phosphoramidites of regular and some modified nucleotides are commercially available and certain oligonucleotide strands could be readily obtained from commercial sources.[4–7] Besides the great development in the synthesis of oligonucleotides, the differentiation of byproducts from the product oligonucleotides, especially on those containing site-specifically modified structures, is sometimes ignored by the end users. If those byproducts or small impurities were not identified and removed, it could have devastating consequences for the corresponding biological assays and medical treatments.[8,9] For certain instances, it is challenging to identify some byproducts generated in the synthesis and deprotection steps, especially those byproducts that have identical or very similar MW to the desired product oligonucleotide.

We have synthesized various oligonucleotides containing modified structures in the past, focusing on alkyl or aryl DNA adducts, by using solid and solution phase DNA synthesis.[10–14] In the syntheses of oligonucleotides containing 1-methyladenine (m1A) and 3-methylcytosine (m3C), we observed byproducts are either regioisomer (identical MW) to the product or have 1 Dalton in MW greater than the product in a ~ 4,900 Dalton context (Figure 1a). m1A and m3C are formed by exogenous and endogenous alkylating agents mainly in single-stranded DNA and have been detected both *in vitro*[15–21] and *in vivo*[15,20,22–29]. Both adducts are cytotoxic and block DNA replication, and are

the best substrates for the AlkB family DNA repair enzymes (Figure 1b).[30–34] m3C has also been proposed an epigenetic biomarker for cancer.[35]

Most of our DNA syntheses successfully provided target oligonucleotides. However, some oligonucleotides contained side reaction contaminants generated during the synthesis or deprotection steps, requiring further purification and identification. To that end, we applied a rigorous quality control process which entails the following steps: i) Synthesize an adduct-containing oligonucleotide from the corresponding phosphoramidite; ii) purify the product oligonucleotide by both reverse phase and anion exchange HPLC; iii) measure the exact MW of the oligonucleotide by high resolution MS (HR-MS), certain impurities can also be detected by HR-MS; iv) determine the modification position by either MS/MS or exonuclease digestion with MALDI-TOF analysis; and v) test the biological activity of the adduct-containing oligonucleotide by appropriate enzymatic assays. We are assuming that the modified nucleotide has been extensively investigated for its stability under solid-phase chemical synthesis and deprotection beforehand; however, the strategy outlined above should confirm this. Nuclear Magnetic Resonance (NMR; ^{1}H, ^{13}C, and ^{31}P) is a powerful tool for structural integrity studies of the phosphoramidite on the milligram scale.[36–38] However, NMR becomes impractical for characterizing minor impurities within the modified structures in synthetic oligonucleotides due to increasing spectral complexity, as well as the final amount of material available (e.g. on a micro-mole or less scale). Below, we report the detailed characterization of byproducts in the syntheses of oligonucleotides containing m1A and m3C. The characterization and separation of those byproducts provides confidence in the quality of oligonucleotides used in further biological experiments.

RESULTS

Identification of m6A Byproduct from m1A Synthesis. We were trying to incorporate m1A into an oligonucleotide as a substrate for the AlkB repair reaction.[34,39] A 16mer oligonucleotide containing m1A was prepared by using the commercially available phosphoramidite of m1A (Figure 1a). The final product was deprotected under standard conditions by treating the crude oligonucleotide with ammonium hydroxide at 80 ℃ for 3h. The oligonucleotide was then tested by both reverse phase and anion exchange HPLC (see Experimental Procedures for detailed conditions). The resulting chromatograms showed a single peak under both conditions (retention time 10.2 min in Figure 2a and 4.0 min in Figure S22). High resolution ESI-TOF MS analysis of the sample exhibited *m/z* at 1224.715 at its -4 charge state, which is in good agreement with the theoretical *m/z* 1224.711 expected of the product oligonucleotide (Figure 3a, S1 and Table 1).

Collision-induced-dissociation (CID) MS/MS analysis was used to determine the location of the alkyl adduct in the synthesis of m1A-containing 16mer oligonucleotide (Figure 4). The MS/MS results, presented in detail in the supporting information (Figures S5 and Tables S1), confirmed that modification occurred at the eighth position from the 3' end of the oligonucleotide and the modified base had the same MW as m1A. Analysis of 3' end fragmentations indicated that w_{15} to w_8 fragments all contain an extra methyl modification (Figure 4, Table S1 and Figure S10). On the other hand, w_7 to w_1 ions showed only standard DNA sequences without methyl modification (Table S1). These results suggested the modification is located between

w_8 and w_7, which is the eighth position from the 3' end of the 16mer oligonucleotide. From the 5' fragmentations, we observed the methyl modification in ions from (a_{15}-C) to (a_{10}-G), but no such modification from ion (a_9-X) to (a_2-A) (Figure 4, Figure S9 and Table S1). These results suggested the modification is located between (a_{10}-G) and (a_9-X), which is the ninth position from the 5' end (Figure 4). The fragmentation patterns from both the 5' and 3' ends are consistent with the proposed m1A-containing DNA sequence (Figure 4).

To further confirm the location of the modified base, we performed enzyme-digestion coupled with MALDI-TOF MS analysis.[11–13,40,41] The procedure has been widely used to confirm the location of lesion positions by determining the mass changes after partial digestion from the 3' end by the exonuclease SVP.[42–44] Figure 5 shows the positive ion MALDI-TOF MS spectra of the 3' to 5' SVP exonuclease digestion of the product oligonucleotide at different time intervals. The *m/z* of 4904.2 at 0 min (before digestion, as a control) indicated the 16mer DNA sequence contained a methylated DNA base (theoretical *m/z* = 4902.9 Dalton, herein theoretical *m/z* values are shown in parentheses after the observed *m/z* values). After 1 minute SVP digestion, while the product ion disappeared, three new lower masses appeared at *m/z* = 4615.7 (4614.8), 4326.7 (4325.8), and 4022.7 (4021.7), which correspond to the 15, 14, and 13mer fragments generated from 3' cleavage, respectively (Figure 5). The signal intensity of three new peaks were significantly increased with the disappearance of the original product oligonucleotide ion at *m/z* = 4904.2.

Further digestion resulted in a number of smaller fragments. In 3 min digestion (Figure 5 and S17), the signal of *m/z* = 3075.1 (3074.6) matches the sequence for the

fragment that was cleaved one base before the lesion. The *m/z* = 2746.4 (2745.5, containing a methyl modification), represents the digestion occurring on the 3' side of the modified adenine position (Figure 5), which also persisted in the 5 min digestion spectrum. The signal at 2419.1 (2418.5, no methyl modification) in the 3 and 5 min digestions represents the fragment generated after liberation of the deoxymethyladenosine 5' monophosphate. Digestions practically stopped at *m/z* = 1223.6/8 (1222.3), which represent the *m/z* of a 4mer nucleotide of the 5' end. Taken together, these results confirm that the modified structure is a methylated adenine and it is at the ninth position of the proposed oligonucleotide from the 5'to 3' direction.

The chromatographic and mass spectral evidence presented above indicates that the product oligonucleotide could have the MW expected for a 16mer containing the m1A adduct in the correct position. We then tested the repair of this product by the AlkB protein, which is an enzyme that repairs various alkylated DNA adducts (Figure 1b).[32,34] The result showed the repair efficiency of this oligonucleotide by AlkB was very low, even when 2.5 uM of AlkB was incubated with 5.0 uM product oligonucleotide (Figure 6b and Figure S24). This observation was in contrast to the report that m1A is a good substrate of AlkB (eg. 0.2 μM of AlkB is able to repair 5.0 μM m1A completely in 1 h).[10] We suspected that the structure in the product oligonucleotide may not be m1A, even though it had an identical MW to m1A. The reverse phase HPLC results also showed an anomalous phenomenon. For example, the positively charged m1A nucleobase[34] should be more polar than the natural base adenine and thus have a shorter retention time on a C18 column. However, the opposite was observed: adenine has a shorter retention time than the product (Figure 2a and 2c). It

has been reported that 6-methyladenine (m6A) could be generated from a Dimroth rearrangement of m1A under basic conditions(Figure 1a and Table 1).[34,45]

We therefore conducted a new batch synthesis of m1A under a milder deprotection condition (25 ℃ for 16 h). The product oligonucleotide was eluted as a single peak in both HPLC columns (Figure 2b for C18 and Figure 6d for anion exchange). The HPLC retention time in reverse phase HPLC exhibited a shorter retention time (8.5 min, Figure 2b) than the non-modified control adenine (9.0 min, Figure 2c) or the previous product (10.2 min, m6A, Figure 2a), in good agreement with the theory that m1A should be more polar than both A and m6A.

The MW of the new product was characterized by HR-MS and had the correct *m/z* of m1A at 1224.715 (1224.711) (Figure 3b, S2 and Table 1). The MS/MS results (e.g., the w_7, w_8, a_9-X, and a_{10}-G ions, Figure 4, Figure S6, S11, S12, and Table S2) demonstrated the modification is at the ninth position from the 5' end and it is a methylated adenine. The exonuclease digestion with MALDI-TOF analyses (Figure S18 and S21a) also confirmed the location and identity of adduct. For example, the 2746.1 (before adduct) and 2418.8 (after adduct) peaks at 3 min digestion in Figure S18 were present. Finally, incubation with AlkB showed an excellent repair efficiency against the new product oligonucleotide: 5.0 μM adduct was mostly repaired in 1 hr with 0.2 μM enzyme (Figure 6e).[10] Taken together, these results confirmed the newly synthesized product as a 16mer oligonucleotide containing m1A at the ninth position from the 5' end (Figure 1a).

Identification of m3U Byproduct from m3C Synthesis. We also wanted to synthesize an oligonucleotide containing m3C as the substrate for the AlkB family DNA repair

reactions within the same sequence context as for the m1A oligonucleotide (Figure 1a).[34,39] After automated synthesis and standard deprotection (80 ℃ for 3h), the HPLC chromatogram from a C18 column showed two peaks with similar intensities (9.5 min and 11.6 min, Figure 2d), which were readily separated (Figure 2e and 2f). HR-MS analysis showed that the early eluting peak (9.5 min, Figure 2e) had an *m/z* value 1218.703 at -4 charge state (Figure 3c and S3), which is in agreement with that of the *m/z* of the target m3C containing oligonucleotide (theoretical *m/z* 1218.709, Table 1). The second species at 11.6 min (Figure 2f) showed an *m/z* 1218.957 at its -4 charge state (Figure 3d and S4). Compared to the 9.5 min peak, the *m/z* value of the second species was 0.254 unit higher at the -4 charge state, indicating that the MW (4,879.850 Dalton) of the second species is 1.016 Dalton higher than the first one. This is consistent with the *m/z* value of an oligonucleotide containing m3U at its -4 charge state (1218.955, Table 1). m3U could potentially be generated from deamination of m3C under basic conditions (Figure 1a).[46] The MS/MS results of the two oligonucleotides (e.g., the w_7, w_8, a_9-X, and a_{10}-G ions, Figure 4, Figure S7, S8, S13 to S16, Table S3, and S4) demonstrated the modifications are at the ninth position from the 5' end. The exonuclease digestion with MALDI-TOF analyses (Figure S19, S20 and S21b) also confirmed the location and identity of the adducts. At 7 minute digestion, signature peaks were 2722.2 (before adduct) and 2418.8 (after adduct) for m3C (Figure S18); and 2723.3 (before adduct) and 2418.7 (after adduct) for m3U (Figure S19). The C18 HPLC retention time of the two species showed m3C containing oligonucleotide had a shorter retention time (9.5 min, Figure 2e) than m3U (11.6 min, Figure 2f). This result indicated m3C is more polar than m3U, which supports that m3C is positively charged at pH 7.0

(Figure 1a).[34] The AlkB reaction (0.2 μM) on the m3C oligonucleotide showed a good repair efficiency: 5.0 μM m3C adduct was mostly repaired to cytosine after 1 h (Figure 6h). This was not the case for m3U containing oligonucleotides under a similar condition, which showed almost no repair to m3U (Figure 6k). These observations were consistent with the observations that m3C is a strong substrate but m3U is a weak substrate for AlkB under both *in vitro* and *in vivo* conditions.[34,47]

To avoid the formation of m3U, we carried out a second batch synthesis of m3C containing oligonucleotide under a milder deprotection condition (25 ℃ for 16 h). This time we observed only the m3C containing species in the HPLC chromatogram (similar to Figure 2e). The identity of the newly synthesized m3C was confirmed by the aforementioned spectroscopic and enzymatic tests. These results indicate that the formation of the byproduct m3U was from deamination of m3C catalyzed by harsh deprotection conditions (high temperature and alkaline).[46,47] Serendipitously, the formation of m3U in the first batch of synthesis provides a new method for chemically preparing m3U-containing oligonucleotide.

DISCUSSION

Chemical synthesis of oligonucleotides is important for conducting various biological, chemical, and medical research including oligonucleotide drug development. The field of oligonucleotide synthesis has progressed such that DNA and RNA containing standard bases can be ordered from commercial sources. Most of these oligonucleotides are in high quality, but it is a good practice to perform quality control in order to confirm

the products have the correct sequences in high purity. A simple HPLC/MS analysis should be sufficient for most applications. However, a more rigorous and stringent quality control procedure[48–50] should be adopted for site-specifically modified oligonucleotides, such as epigenetic marks, DNA adducts, and drug candidates. In some cases, byproducts are generated from side chemical reactions during standard automated synthesis and deprotection steps. For this reason, specific deprotection conditions may be required for preparation of oligonucleotide containing modifications, such as epigenetic biomarkers: 5-hydroxymethyl-dC, 5-formyl-dC, and 5-carboxy-dC.[51,52] After synthesis and deprotection, it is even more important to carry out a thorough purification and characterization procedure to ensure 1) complete removal of protecting groups, 2) product having high purity, and 3) modification having the correct position and identity. In this study, we used a reliable and robust procedure to characterize byproducts from m1A and m3C syntheses.

m6A, a byproduct of m1A synthesis, exhibited identical MW and location to m1A. It is difficult to differentiate m1A and m6A without conducting bioassays and careful HPLC retention analysis. As for m3C, it was the m3U byproduct that showed a 1 Dalton difference in the ~ 4,900 Dalton context (Table 1). The two oligonucleotides could be eluted as a single species in HPLC under certain conditions, thus it is imperative to perform high resolution MS analysis and enzymatic reaction to distinguish the two products. For HPLC analysis of modified oligonucleotides, we used a combination of two different separation systems, such as reverse phase (e.g. C18) and anion exchange chromatography. For example, it is not easy to find a suitable condition for fully separating 16mer containing m1A and A with C18 columns (Figure 2b and 2c).

However, it is relatively easy to separate them with an anion exchange column, since m1A is positively charged at neutral pH but A is not (Figure 1a, 6d and 6f).[34] On the other hand, it is not easy to distinguish m6A and A under anion exchange condition (Figure S22 and S23) because of their neutrality at pH 7.0, but it is possible to separate them on the C18 column (Figure 2a and 2c).

The MW of the oligonucleotide can be determined by HR-MS, and the location of the modification can be identified by MS/MS or exonuclease digestion with MALDI-TOF analysis. To further confirm the identity of the oligonucleotide from a biological activity perspective, we recommend running enzymatic reaction on the product. In this work, we used SVP exonuclease to digest all oligonucleotides formed in the syntheses. A specific enzyme for certain types of modification could greatly help elucidate the structure of modification. So we tested the repair efficiency of m1A and m3C and their byproducts by the AlkB repair enzyme.[32,34] The AlkB protein in *E. coli* is an α-ketoglutarate/Fe(II)-dependent dioxygenase that repairs various alkyl DNA adducts, including m1A and m3C (Figure 1).[30–32,39,53–55] We found that m1A and m3C are good substrates but m6A and m3U are weak substrates of AlkB.[34,47,53] For other modified structures, it is highly recommended that a biological or enzymatic assay should be adopted for identifying the product. The reason for adopting an enzymatic test is because a byproduct (e.g. m6A) may have an identical MW as the target product (e.g. m1A), which may be hard to differentiate by LC and MS analyses, including the MS/MS and exonuclease digestion with MALDI-TOF analysis.

EXPERIMENTAL SECTION

Synthesis of Oligonucleotides Containing m1A and m3C. Oligonucleotides (16mer) with the sequence 5'-GAAGACCTXGGCGTCC-3' containing the lesions at X position were made by using solid-phase phosphoramidite chemistry on a MerMade-4 Oligonucleotide Synthesizer.[10,34,56] The phosphoramidites were purchased from ChemGenes. The final cleavage and deprotection step was usually carried out by treating the oligonucleotide with concentrated aqueous ammonium hydroxide (28%) at 80 ℃ for 3 hours. The modified cleavage and deprotection step for minimizing the byproduct formation was at 25℃ for 16 h. The concentration of an oligonucleotide was determined by measuring UV absorbance at 260 nm. The extinction coefficient (ε) of a certain adduct is calculated as its unmodified counterpart due to the slight variation between the values in the context of a 16mer DNA.

Purity Test of Oligonucleotides Containing m1A and m3C and Related Byproducts by HPLC. The purity of oligonucleotides was tested by both reverse phase (C18) and anion exchange HPLC methods. For the reverse phase test, liquid chromatographic separation was performed by using a Phenomenex Luna Semi-Preparative C18 column (9 x 250 mm, 5 μm) at a flow rate of 3 mL/min. Solvent A was 100 mM triethylammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.2 min, 2.0 to 9.0% of B over 0.3 min, 9.0 to 9.4% of B over 16 min, 9.4 to 70.0% of B over 0.5 min, 70.0% of B for 3 min, 70.0 to 2.0% of B over 0.5 min, and 2.0 % B over 4.5 min. The column oven was set at 40 ℃. The UV signal at 260 nm was used to detect the oligonucleotide absorbance. For the anion exchange LC analysis, oligonucleotides were purified and tested by using a Thermo DNAPac PA-100 anion exchange column

(4 x 250 mm, 13 μm) with solvent A as water and solvent B as 1.5 M ammonium acetate in water. A solvent gradient was carried out under the following conditions: 50.0% of B for 1 min, 50.0 to 52.0% of B over 2 min, 52.0 to 75.0% over 1 min, 75.0% of B for 2 min, 75.0 to 50.0% of B over 0.5 min, and 50.0% of B for 4.5 min. The flow rate was at 4.0 ml/min.

HPLC-MS Analysis. High resolution MS analyses of oligonucleotides were performed on AB Sciex triple quadrupole-TOF mass spectrometers (ABI4600 and ABI5600). ESI was conducted by using a needle voltage of 4.0 kV in a negative ion mode. A heated capillary was set at 600 ℃. The nebulizer gas pressure was 40 psi; the heater gas pressure was 40 psi; the curtain gas pressure was 25 psi; the declustering potential was -220 V; and the collision energy was -10 V. Liquid chromatographic separation was performed using a Phenomenex Luna C 18 column (4.6 × 100 mm; 5 μm) at a flow rate 0.4 mL/min. Solvent A was 10 mM ammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.5 min, 2.0 to 17.4% of B over 11min, 17.4 to 60.0% of B over 0.1 min, 60.0% of B for 2 min, 60.0 to 2.0% of B over 0.1 min, and 2.0 % B over 3.3 min.

MS/MS Analysis. Oligonucleotide fragmentation analyses were performed by manually injecting the oligonucleotide samples of 100 pmol/μl into AB Sciex triple quadrupole-TOF mass spectrometers (ABI4600 and ABI5600). The syringe flow rate was set at 10 μl/min. ESI was conducted by using a needle voltage of 4.5 kV in a negative ion mode. A heated capillary was set at 400 ℃. The nebulizer gas pressure was 75 psi; the heater gas pressure was 25 psi; the curtain gas pressure was 25 psi; the

declustering potential was -100 V; and the collision energy was -15 V. The parent ion *m/z* for m6A, m1A, m3C and m3U were selected as 816.14 (-6 charge), 816.14 (-6 charge), 974.77 (-5 charge) and 974.96 (-5 charge), respectively. Data analyses were performed with the AB Sciex Analyst TF software 1.7.

Exonuclease Digestion with MALDI-TOF Analysis. The modified oligonucleotides were characterized by 3'-5'exonuclease digestion followed by MALDI-TOF analysis. In general, 1.0 μL of sample containing 200-250 pmol of a modified oligonucleotide was used for digestion. SVP (0.2 unit) was added together with 6.0 μL of ammonium citrate (100 mM, pH 9.4) and 6.0 μL of water for the 3′ to 5′ digestion.[11] For the MALDI-TOF analyses, 1.0 μL of the mixture was used at a certain time point until the digestion was almost finished; the digestion was quenched by mixing with 1.0 μL of MALDI analysis matrix (3-hydroxypicolinic acid and di-ammonium hydrogen citrate in a 1:1 ratio). Samples were analyzed by a Shimadzu Axima Performance MALDI-TOF mass spectrometer using a 50Hz laser with a power setting of 90 in positive ion reflection mode with 500 shots collected.[13]

Enzymatic Reaction with the AlkB protein and HPLC Analysis. All reactions were performed at 37 ℃ for 1 h in a reaction buffer containing 70.0 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM HEPES (pH 8.0).[10] The reactions were quenched by adding 10.0 mM EDTA followed by heating at 95 ℃ for 5 min. Typically, 0.25 μM of the purified AlkB protein was incubated with 5.0 μM m6A and m1A in the presence of all cofactors in a 20 μL reaction volume. For the m3C and m3U, 0.18 μM of AlkB was incubated with 5.0 μM of

oligonucleotides. Samples were then analyzed under an HPLC condition that was able to separate the substrate and product.

To separate the starting material and product of the enzymatic reactions for the four lesions, different HPLC conditions were optimized by using either C18 or anion exchange columns. The UV detection was set at 260 nm. Specifically, (1) m6A and its repaired product A were analyzed by using a Phenomenex Luna Semi-Preparative C18 column (10 x 150 mm, 5 μm) under the following conditions: Solvent A was 100 mM triethylammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.2 min, 2.0 to 9.0% of B over 0.3 min, 9.0 to 9.4% of B over 16 min, 9.4 to 70.0% of B over 0.5 min, 70.0% of B for 3 min, 70.0 to 2.0% of B over 0.5 min, and 2.0 % B for 4.5 min. The flow rate was at 3.0 ml/min. The column oven was set at 40 ℃. This protocol was also used to analyze m3U and its repair product U. (2) m1A and its repaired product A were analyzed by using a Thermo DNAPac PA-100 anion exchange column (4 x 250 mm, 13 μm) with solvent A as water and solvent B as 1.5 M ammonium acetate in water. A solvent gradient was carried out under the following conditions: 50.0% of B for 1 min, 50.0 to 52.0% of B over 2 min, 52.0 to 75.0% over 1 min, 75.0% of B for 2 min, 75.0 to 50.0% of B over 0.5 min, and 50.0% of B for 4.5 min. The flow rate was at 4.0 ml/min. (3) m3C and its repaired product C were analyzed with a similar protocol to m1A/A analysis, except the solvent gradient was changed to the following conditions: 50.0% of B for 1 min, 50.0 to 57.0% of B over 1 min, 57.0% of B for 2 min, 57.0 to 50.0% of B for 0.5 min, then 50.0% of B for 4.5 min.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: TBD.

Tables of molecular weights and m/z values, ESI-TOF, MS/MS, and MALDI-TOF analyses (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: deyuli@uri.edu

Present Address

[§]Visterra Inc., Cambridge, Massachusetts 02139, United States

[‖]Xiamen University, Xiamen, Fujian 361005, P. R. China

Funding

The work in the paper was supported by an Institutional Development Award (IDeA) from the NIH/NIGMS under grant number 2 P20 GM103430, and a Medical Research Funds grant from the Rhode Island foundation (to D. L.). This work was in part supported by the National Institutes of Health grant CA098296 (to B.P.C.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its director Prof. Zahir Shaikh and the staff Ms. Kim Andrews and Ms. Patricia Murray for their kind help. We also want to thank Prof. John M. Essigmann, Prof. Bingfang Yan, Prof. David Rowley, and Ms. Kerri Bradshaw for helpful discussions.

REFERENCES

(1) Caruthers, M. H. *Science* **1985**, *230*, 281.

(2) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859.

(3) Matteucci, M. D.; Caruthers, M. H. *J. Am. Chem. Soc.* **1981**, *103*, 3185.

(4) Leumann, C. J. *Bioorg. Med. Chem.* **2002**, *10*, 841.

(5) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992**, *48*, 2223.

(6) Petersen, M.; Wengel, J. *Trends Biotechnol.* **2003**, *21*, 74.

(7) De Mesmaeker, A.; Altmann, K. H.; Waldner, A.; Wendeborn, S. *Curr. Opin. Struct. Biol.* **1995**, *5*, 343.

(8) *Antisense drug technology: principles, strategies, and applications*, 2nd ed.; Crooke, S. T., Ed.; CRC Press: Boca Raton, 2008.

(9) BONILLA, J. V. *Handbook of analysis of oligonucleotides and related products.*; CRC Press: Place of publication not identified, 2016.

(10) Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. *Chem. Res. Toxicol.* **2016**, *29*, 687.

(11) Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D.; Zou, Y.; Cho, B. P. *Chem. Res. Toxicol.* **2013**, *26*, 1251.

(12) Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P. *Biochemistry (Mosc.)* **2014**, *53*, 4059.

(13) Sandineni, A.; Lin, B.; MacKerell, A. D.; Cho, B. P. *Chem. Res. Toxicol.* **2013**, *26*, 937.

(14) Jain, N.; Li, Y.; Zhang, L.; Meneni, S. R.; Cho, B. P. *Biochemistry (Mosc.)* **2007**, *46*, 13310.

(15) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Springer US: Boston, MA, 1983.

(16) Beranek, D. T.; Weis, C. C.; Swenson, D. H. *Carcinogenesis* **1980**, *1*, 595.

(17) Boiteux, S.; Laval, J. *Biochimie* **1982**, *64*, 637.

(18) Gomes, J. D.; Chang, C. J. *Anal. Biochem.* **1983**, *129*, 387.

(19) Chang, C. J.; Gomes, J. D.; Byrn, S. R. *J. Org. Chem.* **1983**, *48*, 5151.

(20) Kawasaki, H.; Ninomiya, S.; Yuki, H. *Chem. Pharm. Bull. (Tokyo)* **1985**, *33*, 1170.

(21) Ashworth, D. J.; Baird, W. M.; Chang, C. J.; Ciupek, J. D.; Busch, K. L.; Cooks, R. G. *Biomed. Mass Spectrom.* **1985**, *12*, 309.

(22) Kelley, M. R.; Fishel, M. L. *Anticancer Agents Med. Chem.* **2008**, *8*, 417.

(23) Messaoudi, K.; Clavreul, A.; Lagarce, F. *Drug Discov. Today* **2015**, *20*, 899.

(24) Culp, L. A.; Dore, E.; Brown, G. M. *Arch. Biochem. Biophys.* **1970**, *136*, 73.

(25) Margison, G. P.; Margison, J. M.; Montesano, R. *Biochem. J.* **1976**, *157*, 627.

(26) Frei, J. V.; Swenson, D. H.; Warren, W.; Lawley, P. D. *Biochem. J.* **1978**, *174*, 1031.

(27) Faustman, E. M.; Goodman, J. I. *J. Pharmacol. Methods* **1980**, *4*, 305.

(28) Beranek, D. T.; Heflich, R. H.; Kodell, R. L.; Morris, S. M.; Casciano, D. A. *Mutat. Res.* **1983**, *110*, 171.

(29) Faustman-Watts, E. M.; Goodman, J. I. *Biochem. Pharmacol.* **1984**, *33*, 585.

(30) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174.

(31) Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. *Nature* **2002**, *419*, 178.

(32) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. *J. Biol. Chem.* **2015**, *290*, 20734.

(33) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. *DNA Repair* **2007**, *6*, 429.

(34) Delaney, J. C.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14051.

(35) Pataillot-Meakin, T.; Pillay, N.; Beck, S. *Epigenomics* **2016**, *8*, 451.

(36) Delaney, J. C.; Essigmann, J. M. *Chem. Res. Toxicol.* **2008**, *21*, 232.

(37) Zeng, Y.; Wang, Y. *J. Am. Chem. Soc.* **2004**, *126*, 6552.

(38) Cao, H.; Jiang, Y.; Wang, Y. *J. Am. Chem. Soc.* **2007**, *129*, 12123.

(39) Li, D.; Delaney, J. C.; Page, C. M.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. *J. Nucleic Acids* **2010**, *2010*, 369434.

(40) Gao, L.; Zhang, L.; Cho, B. P.; Chiarelli, M. P. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1147.

(41) Zhang, L. K.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 854.

(42) Redon, S.; Bombard, S.; Elizondo-Riojas, M.-A.; Chottard, J.-C. *Nucleic Acids Res.* **2003**, *31*, 1605.

(43) Wang, B.; Sayer, J. M.; Yagi, H.; Frank, H.; Seidel, A.; Jerina, D. M. *J. Am. Chem. Soc.* **2006**, *128*, 10079.

(44) Tretyakova, N.; Matter, B.; Ogdie, A.; Wishnok, J. S.; Tannenbaum, S. R. *Chem. Res. Toxicol.* **2001**, *14*, 1058.

(45) Engel, J. D. *Biochem. Biophys. Res. Commun.* **1975**, *64*, 581.

(46) Leutzinger, E. E.; Miller, P. S.; Kan, L.-S. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **1982**, *697*, 243.

(47) Frick, L. E. *The versatile E. coli adaptive response protein AlkB mitigates toxicity and mutagenicity of etheno-, ethano-, and methyl-modified bases in vivo.*; PhD Thesis: Massachusetts Institute of Technology, 2007.

(48) Basiri, B.; Bartlett, M. G. *Bioanalysis* **2014**, *6*, 1525.

(49) van Dongen, W. D.; Niessen, W. M. A. *Bioanalysis* **2011**, *3*, 541.

(50) Lin, Z. J.; Li, W.; Dai, G. *J. Pharm. Biomed. Anal.* **2007**, *44*, 330.

(51) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. *Angew. Chem. Int. Ed Engl.* **2014**, *53*, 315.

(52) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.

(53) Li, D.; Fedeles, B. I.; Shrivastav, N.; Delaney, J. C.; Yang, X.; Wong, C.; Drennan, C. L.; Essigmann, J. M. *Chem. Res. Toxicol.* **2013**, *26*, 1182.

(54) Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. *J. Am. Chem. Soc.* **2012**, *134*, 8896.

(55) Singh, V.; Fedeles, B. I.; Li, D.; Delaney, J. C.; Kozekov, I. D.; Kozekova, A.; Marnett, L. J.; Rizzo, C. J.; Essigmann, J. M. *Chem. Res. Toxicol.* **2014**, *27*, 1619.

(56) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 855.

FIGURE LEGENDS

Figure 1. Oligonucleotide products and byproducts studied in this work. a) the structures of adducts and byproducts from chemical syntheses and b) the alkyl adduct m1A, as an example, is repaired in the presence of the AlkB enzyme and necessary cofactors.

Figure 2. Reverse phase HPLC chromatograms of 16mer DNA alkyl products and byproducts. The retention time of a modification-containing oligonucleotide is labeled on top of the corresponding chromatogram. The modifications are: a) m6A); b) m1A; c) A; d) mixture of m3C + m3U; e) m3C; and f) m3U.

Figure 3. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing target modifications and byproducts. Data represent the -4 charge envelopes and the monoisotopic peak (all 12C, 14N, etc.) values are labeled above the first peak in each peak envelope. a) The oligonucleotide (containing m6A) generated from initial m1A synthesis; b) m1A; c) m3C; and d) m3U.

Figure 4. Predicted fragmentation pattern from collision-induced-dissociation of the 16mer oligonucleotide products from the m1A and m3C syntheses. X denotes the modified nucleotides.

Figure 5. Time-course MALDI-TOF analyses of SVP digestion products of the 16mer product oligonucleotide (containing m6A) generated from m1A synthesis. Theoretical masses are listed in the inset. The theoretical monoisotopic mass at 2745.5 is

highlighted in red because it is the smallest digestion product containing the modification.

Figure 6. HPLC profiles of the AlkB repair reactions on different alkyl substrates. The three chromatograms within each panel represent one set of repair reaction including the oligonucleotides of starting material, reaction mixture, and product. For example, in the panel containing chromatograms a, b, & c, a) represents the starting material m6A, b) represents the repair reaction of m6A by AlkB, and c) represents the pure product adenine synthesized separately. Chromatograms shown in a, b, & c relevant to m6A repair and in j, k, & l relevant to m3U repair were analyzed under reverse phase conditions. Chromatograms shown in d, e, & f relevant to m1A repair and in g, h, & i relevant to m3C repair were analyzed under anion exchange conditions (see Experimental Procedures for detailed information). a) m6A; b) m6A+AlkB; c) A; d) m1A; e) m1A+AlkB; f) A; g) m3C; h) m3C+AlkB; i) C; j) m3U; k) m3U+AlkB; and l) U.

TABLE LEGENDS

Table 1:

Calculated and observed monoisotopic molecular weight and *m/z* value of modified oligonucleotides. For m1A and m3C synthesis, the sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of the modified bases.

Figure 1:

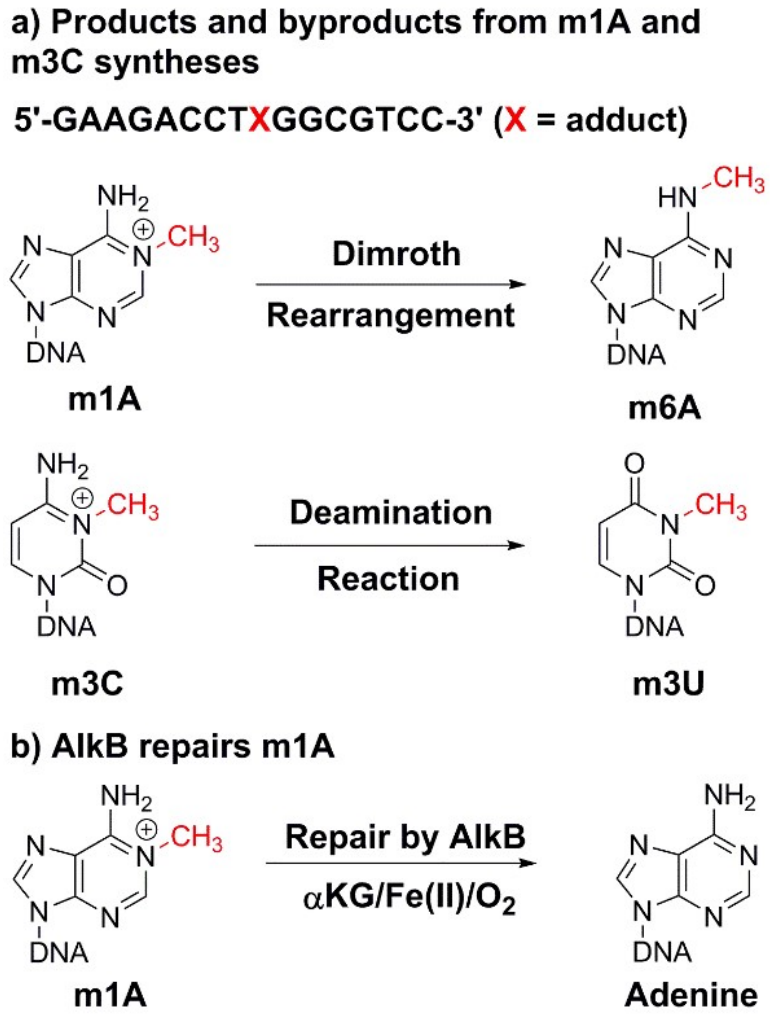

a) Products and byproducts from m1A and m3C syntheses
5'-GAAGACCTXGGCGTCC-3' (X = adduct)
Dimroth
Rearrangement
m1A
m6A
Deamination
Reaction
m3C
m3U
b) AlkB repairs m1A
Repair by AlkB
αKG/Fe(II)/O2
m1A
Adenine

Figure 2:

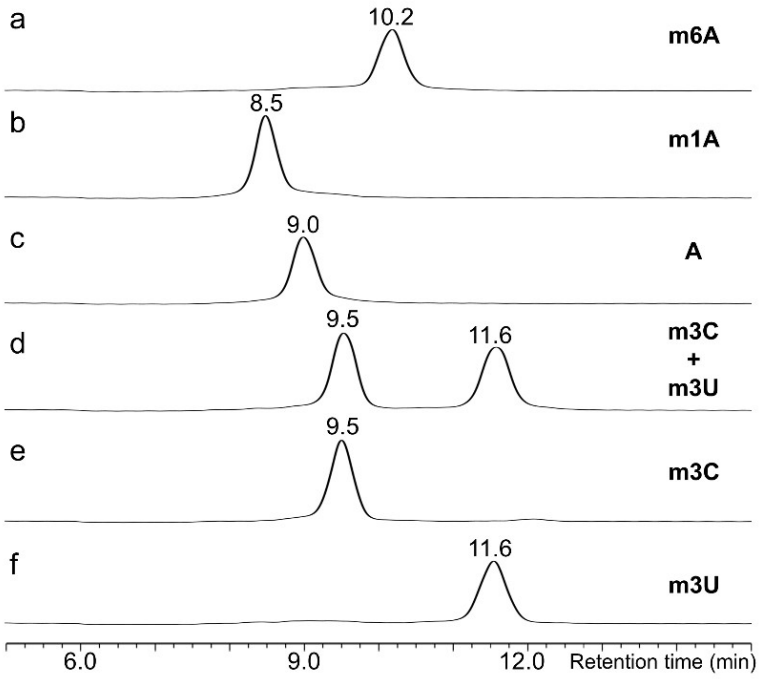

a
10.2
m6A
b
8.5
m1A
c
9.0
A
d
9.5
11.6
m3C
+
m3U
e
9.5
m3C
f
11.6
m3U
6.0
9.0
12.0
Retention time (min)

Figure 3:

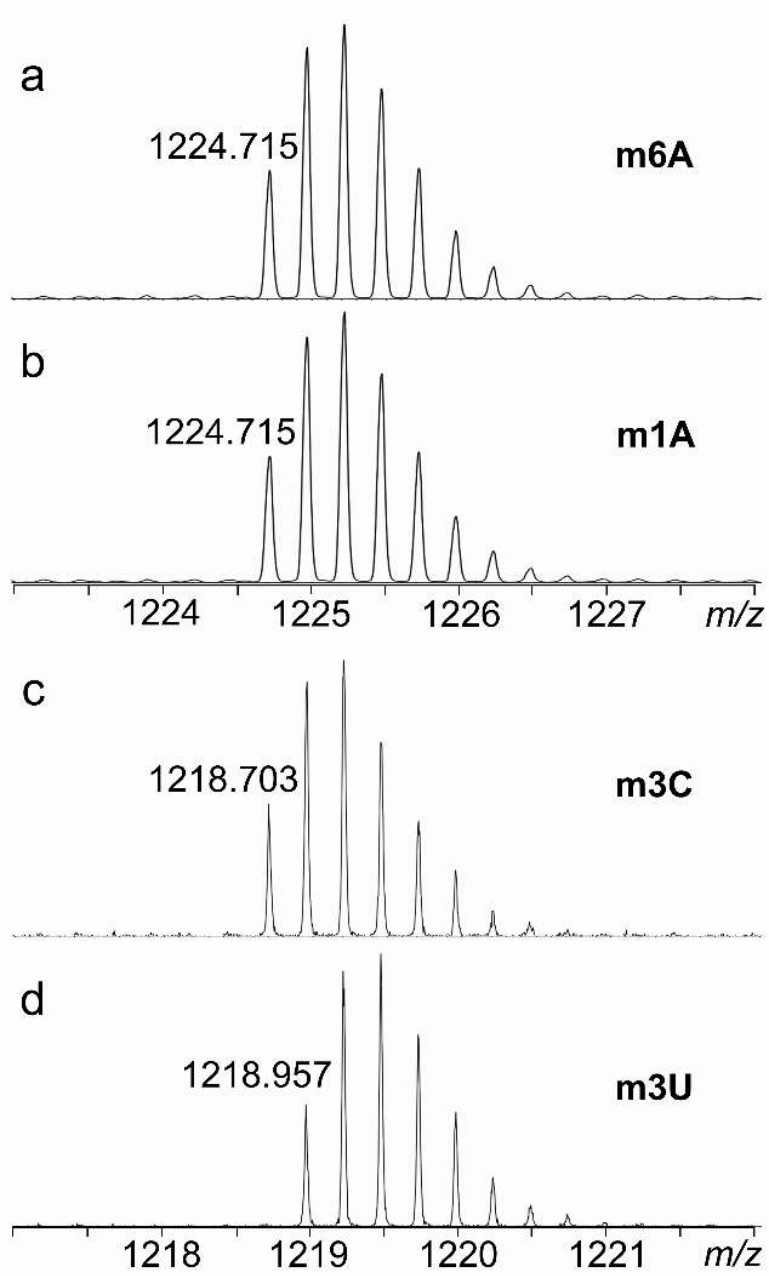

a
1224.715
m6A
b
1224.715
m1A
1224
1225
1226
1227
m/z
c
1218.703
m3C
d
1218.957
m3U
1218
1219
1220
1221
m/z

Figure 4:

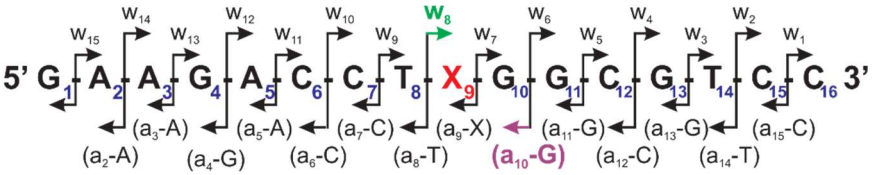

5' G A A G A C C T X G G C G T C C 3'
W15 W14 W13 W12 W11 W10 W9 W8 W7 W6 W5 W4 W3 W2 W1
(a3-A) (a5-A) (a7-C) (a9-X) (a11-G) (a13-G) (a15-C)
(a2-A) (a4-G) (a6-C) (a8-T) (a10-G) (a12-C) (a14-T)

Figure 5:

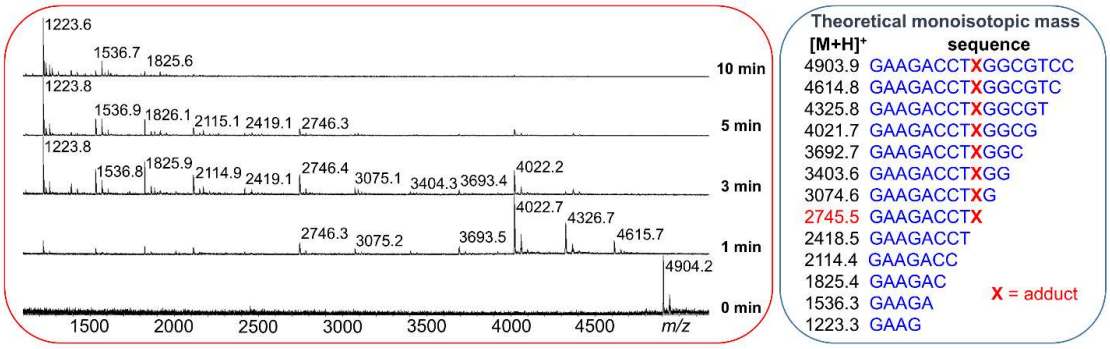

1223.6
1536.7
1825.6
10 min
1223.8
1536.9
1826.1
2115.1
2419.1
2746.3
5 min
1223.8
1536.8
1825.9
2114.9
2419.1
2746.4
3075.1
3404.3
3693.4
4022.2
3 min
4022.7
4326.7
2746.3
3075.2
3693.5
4615.7
1 min
4904.2
0 min
1500
2000
2500
3000
3500
4000
4500
m/z
Theoretical monoisotopic mass
[M+H]+
sequence
4903.9 GAAGACCTXGGCGTCC
4614.8 GAAGACCTXGGCGTC
4325.8 GAAGACCTXGGCGT
4021.7 GAAGACCTXGGCG
3692.7 GAAGACCTXGGC
3403.6 GAAGACCTXGG
3074.6 GAAGACCTXG
2745.5 GAAGACCTX
2418.5 GAAGACCT
2114.4 GAAGACC
1825.4 GAAGAC
1536.3 GAAGA
1223.3 GAAG
X = adduct

Figure 6:

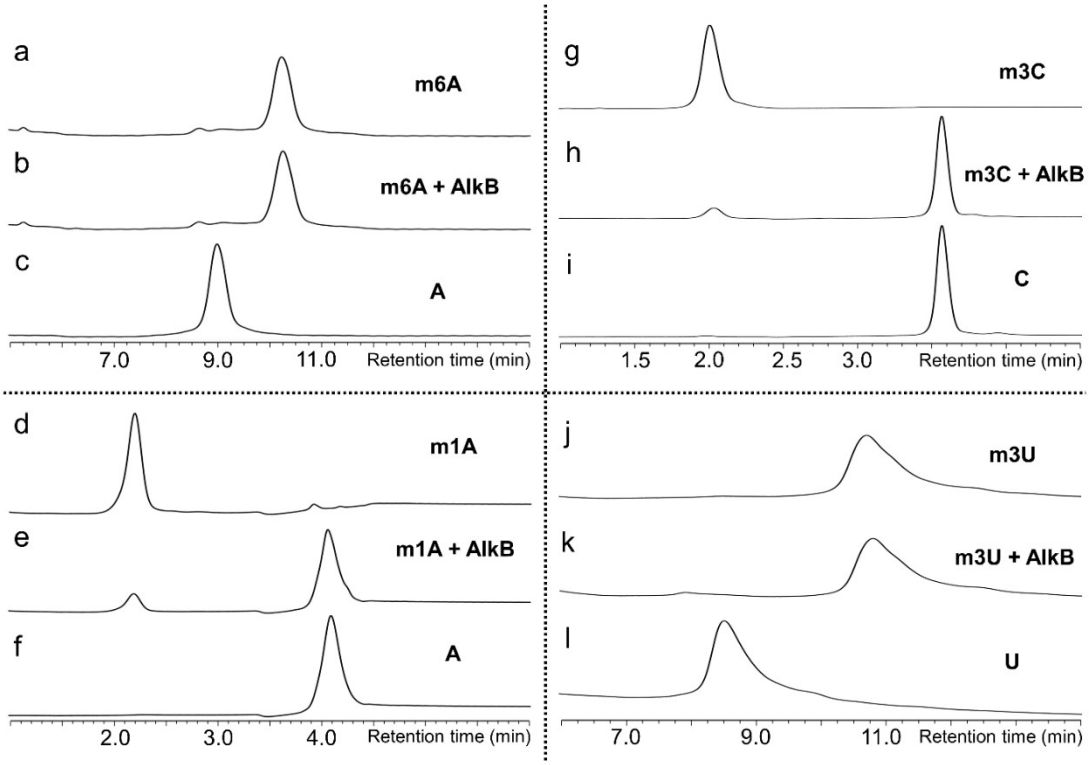

a
m6A
b
m6A + AlkB
c
A
7.0
9.0
11.0
Retention time (min)
g
m3C
h
m3C + AlkB
i
C
1.5
2.0
2.5
3.0
Retention time (min)
d
m1A
e
m1A + AlkB
f
A
2.0
3.0
4.0
Retention time (min)
j
m3U
k
m3U + AlkB
l
U
7.0
9.0
11.0
Retention time (min)

Table 1:

Lesion or base	MW (calculated) of neutral species	*m/z* (calculated) -4 charge peak	*m/z* (observed) -4 charge peak
16mer m6A	4902.877	1224.711	1224.715
16mer m1A	4902.877	1224.711	1224.715
16mer m3C	4878.866	1218.709	1218.703
16mer m3U	4879.850	1218.955	1218.957