University of Rhode Island

DigitalCommons@URI

Open Access Dissertations

2019

DNA damage, repair and mutational spectrum

Ke Bian

University of Rhode Island, bianke93@gmail.com

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

---

Recommended Citation

Bian, Ke, "DNA damage, repair and mutational spectrum" (2019). *Open Access Dissertations*. Paper 850.
https://digitalcommons.uri.edu/oa_diss/850

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

DNA DAMAGE, REPAIR, AND MUTATIONAL SPECTRUM

BY

KE BIAN

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHARMACEUTICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2019

DOCTOR OF PHILOSOPHY DISSERTATION
OF
KE BIAN

APPROVED:

Dissertation Committee:

Major Professor   Deyu Li

Bongsup Cho

Gongqin Sun

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2019

ABSTRACT

The integrity and stability of DNA is essential to life since it stores genetic information in every living cell. Chemicals from the environment will assault DNA to form various types of DNA damage, ranging from small covalent crosslinks between neighboring DNA bases as seen in cyclobutane pyrimidine dimers, to big bulky adducts derived from benzo[a]pyrene. This resultant damage will lead to replication block and mutation if remain unrepaired and will eventually cause cancer or other genetic diseases. The work presented in this dissertation has illustrated the important role of the AlkB family DNA repair enzymes in cancer and Wilson's Disease. In addition, we discovered these enzymes can modify epigenetic markers that affect DNA regulation. We also studied sequence-dependent conformational heterogeneity of aminobiphenyl adduct on DNA replication.

The AlkB family DNA repair enzyme is a family of α -ketoglutarate (α KG)- and non-heme iron-dependent dioxygenases. Among all the homologs in this family, human ALKBH2 and ALKBH3, and E. coli AlkB have been proved to be the major enzymes that directly remove the alkyl adducts from alkylated DNA bases like 3-methylcytosine (3mC) and 1-methyladenine (1mA). These DNA adducts will cause strong replication block and mutagenicity in cell if AlkB enzymes are suppressed by toxicants. Cancer-associated mutations often lead to perturbed cellular energy metabolism and accumulation of potentially harmful oncometabolites. Chiral molecule 2-hydroxyglutarate (2HG) and its two stereoisomers (D- and L-2HG) have been demonstrated to competitively inhibit several α KG- and iron-dependent dioxygenases, including ALKBH2 and ALKBH3. In this work, we carried out detailed kinetic analyses

of DNA repair reactions catalyzed by ALKBH2, ALKBH3 and the bacterial AlkB in the presence of D- and L-2HG in both double and single stranded DNA contexts. We not only determined kinetic parameters of inhibition, including k_{cat} , K_M , and K_i , but also correlated the relative concentrations of 2HG and α KG previously measured in tumor cells with the inhibitory effect of 2HG on the AlkB family enzymes. Both D- and L-2HG significantly inhibited the human DNA repair enzymes ALKBH2 and ALKBH3 under pathologically relevant concentrations (73-88% for D-2HG and 31-58% for L-2HG inhibition). This work provides a new perspective that the elevation of either D- or L-2HG in cancer cells may contribute to an increased mutation rate by inhibiting the DNA repair carried out by the AlkB family enzymes and thus exacerbate the genesis and progression of tumors.

Another type of inhibitor of AlkB is toxic metals, such as, copper. Disturbed metabolism of copper ions can cause diseases, such as Wilson's disease (WD). In this work, we investigated the inhibitory effect of Cu(II) ion on the AlkB family DNA repair enzymes, include human ALKBH2, ALKBH3, and *E. coli* AlkB proteins. None of the three proteins were significantly inhibited under normal cellular copper concentrations. But under WD related condition, we observed the activities of all three enzymes were strongly suppressed (inhibition from 95.2 to 100.0%). We also noted the repair efficiency under ds-DNA condition is less susceptible than ss-DNA to the inhibition.

AlkB can repair many alkylated DNA bases including 3mC, 1mA, 3-methylthymine, 1-methylguanine, ethenoadenine and ethenocytosine. But in this work, we found a new DNA base substrate for AlkB, 5-Methylcytosine (5mC). 5mC in DNA CpG islands is an important epigenetic biomarker for mammalian gene regulation. It is

oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) family enzymes, which are also α -KG/Fe(II)-dependent dioxygenases. In this work, we demonstrate that the epigenetic marker 5mC is biochemically modified to 5hmC, 5fC, and 5caC by ALKBH2, ALKBH3, and AlkB. Theoretical calculations indicate that these enzymes may bind 5mC in the syn-conformation, placing the methyl group comparable to 3-methylcytosine, the prototypic substrate of AlkB. This is the first demonstration of the AlkB proteins to oxidize a methyl group attached to carbon, instead of nitrogen, on a DNA base. These observations suggest a broader role in epigenetics for these DNA repair proteins.

Besides alkyl DNA adducts, there are bulky DNA adducts existing in human. Bulky organic carcinogens are activated *in vivo* and subsequently react with nucleobases of cellular DNA to produce adducts. Some of these DNA adducts exist in multiple conformations that are slowly interconverted to one another. Different conformations could contribute to different mutagenic and repair outcomes. Unfortunately, studies on the conformation-specific inhibition of replication, which is more relevant to cell survival, are scarce; this is presumably due to difficulties in studying the structural dynamics of DNA lesions at the replication fork. It is challenging to capture the exact nature of replication inhibition by traditional end-point assays, since they usually detect either the ensemble of consequences of all the conformers or the culmination of all cellular behaviors, such as mutagenicity or survival rate. One article reported an unusual sequence-dependent conformational heterogeneity involving FAPB-modified (4'-fluoro-4-aminobiphenyl) DNA under different sequence contexts. There were 67% B-

type (B) conformation and 33% stacked (S) conformation in $TG_1^*G_2T$ sequence; whereas, 100% B conformation was observed in $TG_1G_2^*T$ sequence. In this study, we applied primer extension assay to compare the inhibition models between these two FABP-modified DNA sequences. We utilized a combination of surface plasmon resonance (SPR) and HPLC-based steady-state kinetics to reveal the differences in terms of binding affinity and inhibition with polymerase between these two conformers (67%B:33%S and 100%B). The conformational heterogeneities from these two sequences lead to different types of inhibition on replication.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to Prof. Deyu Li, my advisor of PhD program in University of Rhode Island. This work wouldn't have been possible without his professional guidance and endless support. He eased the way for me to start my research in medicinal chemistry, especially DNA damage and repair, which apparently should be a big challenge for me when I join the lab at beginning. I feel great thankful to his patience on teaching and energy on brainstorming. These helped me turning countless inspirations into innovative projects. He is not only my mentor in research, but also the advisor in my career. My future career will be greatly benefited from all his advice.

Besides, it is a great pleasure to acknowledge my deepest thanks to Prof. Bongsup Cho for his cutting-edge ideas and comprehensive comments on every group meeting and journal club. His rigorous attitude to research encouraged me to devote myself to research wholeheartedly. I would also like to express my sincere gratitude to Prof. Gongqin Sun for his patience on teaching me the fundamental knowledge of biochemistry, which helped me design and build up this work. I am very thankful to Prof. Xuerong Wen for agreeing to serve on my dissertation committee as the chair to organize the comprehensive exam and defense.

I must express my gratitude to my lab mates for their long-term support. This work would not have been possible without the contribution of Dr. Fangyi Chen and Qi Tang, who spent numerous efforts and time together with me on this work. I am also thankful to my other lab mates that include Zach Humulock, Rui Qi, Dr. Yi-Tzai Chen, Samuel

Howarth, Xianhao Zhou, and Quentin Mylie for making my experience in the lab exciting and fun.

I am also grateful to my collaborators who have made enormous contribution to this work. Dr. Ang Cai, Alicia Crisalli, and Rachel Carley from Dr. Bongsup Cho's lab in University of Rhode Island shared many intelligent ideas and gave me huge support on DNA modification in this work. Dr. Stefan Lenz from Dr. Stacey Wetmore's lab in University of Lethbridge provided me tremendous help on computational modeling in this work. Dr. Bogdan Fedeles, Dr. Vipender Singh, Dr. Marco Jost, Dr. Catherine Drennan from Dr. John Essigmann's lab in Massachusetts Institute of Technology brought me the innovative thoughts on designing cellular experiments. Dr. Nimrat Chatterjee and Dr. Faye-Marie Vassel from Dr. Graham Walker's lab in Massachusetts Institute of Technology gave me big technical support on cellular work.

Last but not least, I would like to express my deepest gratitude to my family and friends. This work would not have been possible without their warm love and endless support.

PREFACE

This dissertation was following the requirements of Manuscript Format. Manuscript I was published in Chemical Research in Toxicology, April 2017. Manuscript II was published in Chemical Research in Toxicology, October 2017. Manuscript III was prepared to submit to Nucleic Acids Research by January 2019. Manuscript II was prepared to submit to Molecules by February 2019.

The titles for these manuscripts were listed below:

I. Oncometabolites d- and l-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions

II. Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition

III. DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

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

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	vi
PREFACE.....	viii
TABLE OF CONTENTS.....	ix
LIST OF TABLES	xi
LIST OF FIGURES	xiv
MANUSCRIPT I.....	1
TOC FIGURE	3
ABSTRACT	4
KEYWORDS	5
INTRODUCTION	6
EXPERIMENTAL PROCEDURES	11
RESULTS	14
DISCUSSION	25
REFERENCES.....	30
SUPPORTING INFORMATION	40
MANUSCRIPT II	61
ABSTRACT	63
MAIN MANUSCRIPT	64
REFERENCES.....	73
SUPPORTING INFORMATION	76
MANUSCRIPT III.....	84

ABSTRACT	87
INTRODUCTION	88
MATERIAL AND METHODS	91
RESULTS AND DISCUSSION	97
REFERENCES.....	106
SUPPORTING INFORMATION	113
MANUSCRIPT IV	152
ABSTRACT	155
KEYWORDS	156
INTRODUCTION	157
RESULTS	160
DISCUSSION	172
MATERIALS AND METHODS.....	176
CONCLUSIONS.....	180
REFERENCES.....	182
SUPPORTING INFORMATION	189

LIST OF TABLES

TABLE	PAGE
MANUSCRIPT I	
Table 1. Kinetic constants of α KG as a substrate on ALKBH2, ALKBH3 and AlkB repair reactions.	18
Table 2. Inhibition constant (K_i) of D-2HG, L-2HG and N-OG.	20
Table 3. Inhibition ratio of D-2HG (373 fold to α KG) and L-2HG (28 fold to α KG) on ALKBH2 and ALKBH3.	22
Table S1. Calculated and observed molecular weight (MW) and m/z value of oligonucleotides used in the enzymatic reactions.	44
Table S2. Initial rate for kinetic studies of ALKBH2, ALKBH3 and AlkB on α KG as a substrate under different conditions.	45
Table S3. Initial rate for kinetic studies of ALKBH2 and ALKBH3 on DNA adducts as substrates.	46
Table S4. Kinetic constants of ALKBH2 and ALKBH3 reactions on DNA adducts as substrates.	47
Table S5. 2-HG inhibition on ALKBH2 demethylation of ss- and ds-m1A.	48
Table S6. Addition of α KG reverses the inhibitory effect of 2HG toward ALKBH2, ALKBH3 and AlkB.	50
Table S7: IC ₅₀ (50% inhibition concentration) of L-2HG, D-2HG and N-OG on ALKBH2, ALKBH3 and AlkB.	50
MANUSCRIPT II	
Table 1. IC ₅₀ values of Cu(II) ion inhibiting the AlkB family enzymes.	69

Table S1. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes in ss- and ds-DNA.	80

Table S2. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes under normal and Wilson's Disease conditions.	80

Manuscript III

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides.	145

Table S2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.	146

Table S3. Summary of important hydrogen bonds formed during MD simulations of <i>anti</i> -3mC, <i>anti</i> -5mC, <i>syn</i> -5mC, <i>syn</i> -5hmC, and <i>syn</i> -5fC in the ALKBH2 complex.	147

Table S4. Summary of important hydrogen bonds formed during MD simulations of <i>anti</i> -3mC, <i>anti</i> -5mC, <i>syn</i> -5mC, <i>syn</i> -5hmC, and <i>syn</i> -5fC in the AlkB complex.	148

Table S5. Summary of important distances and dihedral angles adopted during MD simulations of <i>anti</i> -3mC, <i>anti</i> -5mC, <i>syn</i> -5mC, <i>syn</i> -5hmC, and <i>syn</i> -5fC in the ALKBH2 complex.	149

Table S6. Summary of important distances and dihedral angles adopted during MD simulations of <i>anti</i> -3mC, <i>anti</i> -5mC, <i>syn</i> -5mC, <i>syn</i> -5hmC, and <i>syn</i> -5fC in the AlkB complex.	149

Table S7. Time to half maximum conversion ($T_{1/2}$, min) for the oxidation of different alkyl DNA modifications by the three AlkB family enzymes.	150

Manuscript IV

Table 1. Steady-state kinetic parameters for insertion of dCTP opposite unmodified and	

FABP–dG adduct with Kf-exo [−]	166
Table 2. SPR binding affinities, K_D (nM) of unmodified (TG ₁ G ₂ T) and FABP–dG adducts (TG ₁ [FABP]G ₂ T and TG ₁ G ₂ [FABP]T) with Kf-exo [−] (steady-state affinity analysis) in the binary and ternary systems..	169
Table S1. Binding net stabilization energy of unmodified and FABP-adducts with Kf-exo [−] in binary system (1:1 binding).	193

LIST OF FIGURES

FIGURE	PAGE
MANUSCRIPT I	
Figure 1. a) Repair mechanism of the AlkB family enzymes on alkyl DNA lesions. Adduct m1A is used here as an example to show the steps of enzymatic catalysis. b) The generation of D- and L-2HG and mechanisms of inhibition to the AlkB family DNA repair enzymes.	8
Figure 2. HPLC analyses of the DNA repair reactions.	16
Figure 3. Inhibition of ALKBH2 repair of m1A in ss- and ds-DNA by D- and L-2HG.	21
Figure 4. Addition of α KG reverses the inhibitory effect of 2-HG toward ALKBH2 repair of m1A.	24
Figure S1. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by ALKBH2.	51
Figure S2. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by ALKBH3.	52
Figure S3. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by AlkB.	53
Figure S4. Steady-state kinetic studies probing the influence of adducts in the demethylation reactions catalyzed by ALKBH2 and ALKBH3.	54
Figure S5. The repair percentage of AlkB on ss-m1A under various concentrations of α KG.	55

Figure S6. Inhibition of ALKBH2-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.....	55
Figure S7. Inhibition of ALKBH2-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.....	56
Figure S8. Inhibition of ALKBH2-mediated ds-m1A repair by D-2HG, L-2HG and N-OG.....	56
Figure S9. Inhibition of ALKBH2-mediated ds-m3C repair by D-2HG, L-2HG and N-OG.....	57
Figure S10. Inhibition of ALKBH3-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.....	57
Figure S11. Inhibition of ALKBH3-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.....	58
Figure S12. Inhibition of AlkB-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.	58
Figure S13. Inhibition of AlkB-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.	59
Figure S14. Inhibition of AlkB-mediated ds-m1A repair by D-2HG, L-2HG and N-OG.	59
Figure S15. Inhibition of AlkB-mediated ds-m3C repair by D-2HG, L-2HG and N-OG.	60

MANUSCRIPT II

Figure 1. Repair mechanism of the AlkB family enzymes on alkyl DNA lesions and the inhibition of repair by Cu(II) ion.	66

Figure 2. IC ₅₀ measurements of Cu(II) ion on the repair of m3C by AlkBH2, AlkBH3, and AlkB under ss- and ds-DNA conditions.	68
Figure S1. HPLC analyses of the DNA repair reactions.	78
Figure S2. Dose-response curve of Fe(II) ions to the AlkB repair activity on m3C. .	79
MANUSCRIPT III	
Figure 1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC.	89
Figure 2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products.....	99
Figure 3. Product distribution and strand preference (in univariate scatterplot) for the oxidation of 5mC by the AlkB family enzymes.	101
Figure 4. Representative molecular dynamics structures of the ALKBH2 (a-c) or AlkB (d-f) complex bound to <i>anti</i> -5mC (a,d), <i>syn</i> -5mC (b,e), or <i>anti</i> -3mC (c,f).	103
Figure S1. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures.....	123
Figure S2. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides.	124
Figure S3. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode.	125
Figure S4. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode.	126
Figure S5. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode.....	127

Figure S6. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode.	128
Figure S7. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode.	129
Figure S8. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode.	130
Figure S9. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode.	131
Figure S10. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode.	132
Figure S11. MS analyses of trypsin digested proteins including the wild type and variant AlkB enzymes.	133
Figure S12. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and variant proteins.	134
Figure S13. Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to 5mC.	135
Figure S14. Distribution of active site water (red spheres) during MD simulations of AlkB bound to 5mC.	136
Figure S15. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB–5mC and ALKBH2–5mC complexes onto a crystal structure of TET2 bound to 5mC-containing DNA.	137
Figure S16. Distribution of active site water during MD simulations of ALKBH2 bound to <i>anti</i> -3mC and overlay of MD representative structures of ALKBH2 bound to <i>syn</i> -	

5mC and <i>anti</i> -3mC.....	138
Figure S17. Distribution of active site water during MD simulations of AlkB bound to <i>anti</i> -3mC and overlay of MD representative structures of AlkB bound to <i>syn</i> -5mC and <i>anti</i> -3mC.	139
Figure S18. Structures and chemical numbering of 3mC and 5mC.....	140
Figure S19. Structures and chemical numbering of <i>anti</i> -3mC, <i>syn</i> -5mC, <i>syn</i> -5hmC, and <i>syn</i> -5fC.	140
Figure S20. Distribution of active site water during MD simulations of ALKBH2 bound to <i>syn</i> -5hmC, ALKBH2 bound to <i>syn</i> -5fC, AlkB bound to <i>syn</i> -5hmC, or AlkB bound to <i>syn</i> -5fC.	141
Figure S21. Simulations of ALKBH2 or AlkB bound to 5hmC.....	142
Figure S22. <i>syn</i> -5fC Bound by ALKBH2 or AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety.	143
Figure S23. Different alkyl DNA modifications oxidized by the three AlkB family enzymes at various time points.	144

MANUSCRIPT IV

Figure 1. (a) Two major conformational views of dG-C8-FABP [<i>N</i> -(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl]: B-, and S-conformers. FABP-red, modified dG-blue, complementary C-orange. (b) The chemical structure of dG-C8-FABP.	158
Figure 2. (a) Hairpin template–primer oligonucleotide constructs of 85-mer G ₁ and 84-mer G ₂ for Kf-exo [−] , H denotes 3'-dideoxy-nucleotide; (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) <i>K_a</i>	

(K_a -modified / K_a -control) and K_d (K_d -modified / K_d -control) 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.	161
Figure 3. DNA replication models for (a) FABP-modified G_1^* (TG*GT) templates and (b) FABP-modified G_2^* (TGG*T) templates. n: lesion site. Lineweaver–Burk model of DNA synthesis catalyzed by Kf-exo [−] at (c) 10-mer, G_1 -FABP and (d) 9-mer G_2 -FABP lesion sites with dCTP at steady-state.	162
Figure 4. Sensorgrams of the ternary Kf-exo [−] complexed with (a) 85-mer control, (b) 85-mer TG ₁ [FABP]G ₂ T, (c) 84-mer control and (d) 84-mer TG ₁ G ₂ [FABP]T sequences in the presence of dNTPs (1:1 binding fitted curves are overlaid as black lines). ...	170
Figure S1. HPLC profile of FABP-modified bio-31mer TGGT template and MALDI-TOF characterization of peak 1 and peak 2.	191
Figure S2. Denaturing gel (15%) profiles of 84-mer and 85-mer ligated oligonucleotides, 31-mer, 53-mer and 54-mer non-ligated oligonucleotides.	192
Figure S3. dATP insertion efficiency results of (a) 8-mer to 9-mer and (b) 11-mer to 12-mer in control, G_1^* G ₂ T-FABP and G_1 G ₂ *T-FABP, respectively, at 10 min.	192

Published in Chemical Research in Toxicology, April 2017

**Oncometabolites d- and l-2-Hydroxyglutarate Inhibit the AlkB Family DNA
Repair Enzymes under Physiological Conditions**

Fangyi Chen^{†♦}, Ke Bian^{†♦}, Qi Tang[†], Bogdan I. Fedeles[‡], Vipender Singh^{‡§}, Zachary
T. Humulock[†], John M. Essigmann[‡], and Deyu Li^{*†}

♦F.C. and K.B. contributed equally to this work.

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

[‡]Department of Biological Engineering, Department of Chemistry, and Center for
Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, United States

*Corresponding Author: Deyu Li

Department of Biomedical and Pharmaceutical Sciences

University of Rhode Island

495N Avedisian Hall, 7 Greenhouse Rd

Kingston, Rhode Island 02881, United States

Phone: +1-401-874-9361

Email address: deyuli@uri.edu

**Oncometabolites D- and L-2-hydroxyglutarate Inhibit the AlkB Family DNA
Repair Enzymes under Physiological Conditions**

Fangyi Chen,^{†,‡} Ke Bian,^{†,‡} Qi Tang,[†] Bogdan I. Fedeles,^{⊥,§,‡} Vipender Singh,^{⊥,§,‡,Δ}
Zachary T. Humulock,[†] John M. Essigmann,^{⊥,§,‡} and Deyu Li^{*,†}

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

Departments of [⊥]Biological Engineering, [§]Chemistry, and [‡]Center for Environmental
Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts
02139, United States.

Present Address: ^ΔNovartis Institute of Biomedical Research, Cambridge,
Massachusetts 02139, United States

Corresponding Author: *E-mail: deyuli@uri.edu

[‡]F.C. and K.B. contributed equally to this work.

TOC FIGURE

ABSTRACT

Cancer-associated mutations often lead to perturbed cellular energy metabolism and accumulation of potentially harmful oncometabolites. One example is the chiral molecule 2-hydroxyglutarate (2HG); its two stereoisomers (D- and L-2HG) have been found with abnormally high concentrations in tumors featuring anomalous metabolic pathways. 2HG has been demonstrated to competitively inhibit several α -ketoglutarate (α KG)- and non-heme iron-dependent dioxygenases, including some of the AlkB family DNA repair enzymes, such as ALKBH2 and ALKBH3. However, previous studies have only provided the IC₅₀ values of D-2HG on the enzymes and the results have not been correlated to physiologically relevant concentrations of 2HG and α KG in cancer cells. In this work, we carried out detailed kinetic analyses of DNA repair reactions catalyzed by ALKBH2, ALKBH3 and the bacterial AlkB in the presence of D- and L-2HG in both double and single stranded DNA contexts. We determined kinetic parameters of inhibition, including k_{cat} , K_M , and K_i . We also correlated the relative concentrations of 2HG and α KG previously measured in tumor cells with the inhibitory effect of 2HG on the AlkB family enzymes. Both D- and L-2HG significantly inhibited the human DNA repair enzymes ALKBH2 and ALKBH3 under pathologically relevant concentrations (73-88% for D-2HG and 31-58% for L-2HG inhibition). This work provides a new perspective that the elevation of either D- or L-2HG in cancer cells may contribute to an increased mutation rate by inhibiting the DNA repair carried out by the AlkB family enzymes and thus exacerbate the genesis and progression of tumors.

KEYWORDS. 2-hydroxyglutarate, oncometabolite, ALKBH2, ALKBH3, AlkB, DNA repair inhibition, tumorigenesis

INTRODUCTION

Mutations in the isocitrate dehydrogenases 1 and 2 (IDH1 and IDH2) are frequently found in >75% human low grade glioma, secondary glioblastoma, cartilaginous tumor and >20% of acute myeloid leukemia.¹⁻⁵ Tumor-derived mutant forms of IDH catalyze the NAD-dependent dehydrogenation of α -ketoglutarate (α KG) to D-2-hydroxyglutarate (D-2HG), a function that supplants the physiological activity of IDH, which entails reductive decarboxylation of isocitrate to α KG (Figure 1b).⁶⁻¹³ L-2HG, the stereoisomer of D-2HG, has been identified as an oncometabolite with elevated concentrations in renal cell carcinoma^{9,10} neurodegenerative disorders,^{14,15} and in tissues under oxygen limitation or hypoxic conditions.^{11,12} The elevation of L-2HG under such conditions is key from either loss of expression of L-2HG dehydrogenase or promiscuous substrate utilization by lactate dehydrogenase A and malate dehydrogenases 1 and 2.^{10,12} Both D-2HG (R-2HG) and L-2HG (S-2HG) and several other molecules have been identified as oncometabolites because their accumulations in different tumor cells are originated from dysregulated energy metabolism pathways and metabolic imbalance.¹⁶⁻²⁰ Because of their structural similarity to α KG, both D- and L-2HG could compete with α KG and inhibit enzymatic processes that use α KG as a substrate. One important class of enzymes that utilizes α KG—and thus potentially susceptible to inhibition by 2HG—is the non-heme iron- and α KG-dependent dioxygenases, a family of enzymes with over 80 different members.²¹⁻²⁴ 2HG has been demonstrated to inhibit several α KG-dependent enzymes, such as histone demethylases, prolyl hydroxylases, the TET family 5-methylcytosine (5mC) hydroxylases,⁷ and some of the AlkB family DNA repair enzymes, such as ALKBH2 and ALKBH3.^{8,25} However,

in the case of the AlkB proteins, previous *in vitro* studies have only provided the IC₅₀ values of 2HG on the enzymes and the results have not been extrapolated to physiologically relevant concentrations of 2HG and αKG in cancer cells. While Wang et al. demonstrated that the accumulation of DNA damage in cells producing high levels of D-2HG is consistent with inhibition of ALKBH2 and/or ALKBH3,²⁵ the detailed mechanism of inhibition of the AlkB enzymes by 2HG has not been reported. A careful study of the inhibitory effect of both D- and L-2HG on AlkB repair enzymes is needed to quantify the extent of inhibition of the direct reversal DNA repair pathways; perturbations in these pathways would lead to unrepaired mutagenic DNA lesions, which would cause mutations that can accelerate tumor progression or enable metastatic growth. Such a study would also facilitate the identification of druggable targets related to the AlkB enzymes because many alkylating chemotherapeutic agents generate DNA adducts that are repaired by this family of repair enzymes.²⁶ Inhibition of these enzymes would thus afford a clinical benefit in anti-tumor regimens.

Figure 1. a) Repair mechanism of the AlkB family enzymes on alkyl DNA lesions. Adduct m1A is used here as an example to show the steps of enzymatic catalysis. b) The generation of D- and L-2HG and mechanisms of inhibition to the AlkB family DNA repair enzymes.

The *Escherichia coli* AlkB protein was discovered to be an α KG/Fe(II)-dependent dioxygenase that oxidizes the alkyl groups in DNA adducts formed by alkylation agents, ultimately restoring the undamaged DNA bases (Figure 1a).^{27–29} Nine human homologs of AlkB have been identified as ALKBH1–8 and FTO.^{29,30} Among these homologs, ALKBH2^{31–36} and ALKBH3^{37–40} have been identified as major DNA repair enzymes for repairing small alkyl DNA lesions. Since the initial discovery of the catalytic

mechanism of AlkB in 2002,^{27,28} a range of alkyl adducts have been identified as substrates for AlkB, ALKBH2 and ALKBH3, both *in vitro* and *in vivo*.²⁴ The adducts include all of the seven N-methyl adducts occurring at the Watson-Crick (W-C) base-pairing face of the four nucleobases.⁴¹ The seven adducts include 3-methylcytosine (m3C), *N*⁴-methylcytosine, 1-methyladenine (m1A), *N*⁶-methyladenine, 3-methylthymine, 1-methylguanine, and *N*²-methylguanine. AlkB has also been reported to repair other DNA adducts, such as 1,*N*⁶-ethenoadenine, 1,*N*⁶-ethanoadenine, 3,*N*⁴-ethenocytosine, 3-ethylcytosine, 1,*N*²-ethenoguanine, 3,*N*⁴- α -hydroxyethanocytosine, 3,*N*⁴- α -hydroxypropanocytosine, *N*²-ethylguanine, *N*²-tetrahydrofuran-2-yl-methylguanine, *N*²-furan-2-yl-methylguanine, malondialdehydeguanine, α -hydroxypropanoguanine, and γ -hydroxypropanoguanine.^{41–49} The repair efficiency and substrate scope of the AlkB family enzymes have been recently reviewed in detail.^{24,30,50–54}

In this work, we carried out kinetic analyses of DNA repair reactions catalyzed by ALKBH2, ALKBH3 and AlkB in the presence of the inhibitors D- and L-2HG. Oligonucleotides containing the methylated bases m1A and m3C were selected as substrates for the repair reactions because they are most efficiently repaired by these three enzymes.²⁴ For each substrate/enzyme/inhibitor combination, we determined a complete panel of kinetic parameters, (k_{cat} , K_M , k_{cat}/K_M , K_i), and correlated the relative concentrations of 2HG and α KG found in tumor cells with the inhibitory effect of 2HG on the AlkB family enzymes. Because ALKBH2 preferentially repairs adducts in double stranded (ds) DNA and ALKBH3 prefers to repair lesions in a single stranded (ss) DNA context, we tested the three repair enzymes with both ds and ss DNA

substrates. This is the first report of 2HG inhibition of AlkB family DNA repair evaluated in both single- and double- stranded DNA. We also developed an HPLC-based method to study DNA repair in the ds-DNA. The results showed that both ALKBH2 and ALKBH3, the major mammalian direct reversal repair enzymes for alkylated DNA damage, were significantly inhibited by D- and L-2HG under pathophysiologically relevant conditions.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis. Sixteen-mer oligonucleotides were synthesized with the sequence 5'-GAAGACCTXGGCGTCC-3' containing the lesions m1A and m3C at the X position.^{43,55} The complementary 23mer oligonucleotides were synthesized with the sequence of 5'-CTGGGACGCCYAGGTCTTCACTG-3', where Y represents the position incorporating the regular bases T or G, and named as 23mer-Tcp or 23mer-Gcp. The 23mer oligonucleotides complementary to 23mer-Tcp and 23mer-Gcp were also synthesized, with the sequence 5'-CAGTGAAGACCTZGGCGTCCCAG-3', where Z was the regular base A or C, named as 23mer-A or 23mer-C. All DNA syntheses employed solid-phase phosphoramidite chemistry performed on a MerMade-4 Oligonucleotide Synthesizer.⁵⁵ The oligonucleotides were purified by HPLC (Thermo Fisher Scientific) on a DNAPac PA-100 Semi-Preparative column (Phenomenex). Solvent A was 100 mM 1:1 triethylamine-acetic acid (TEAA) in water and solvent B was 100% acetonitrile. The concentration of DNA was determined by UV absorbance at 260nm by NanoDrop. The oligonucleotides were characterized by HPLC-electrospray ionization triple quadrupole time of flight mass spectrometry (AB Sciex) (Table S1).

Expression and Purification of the AlkB, ALKBH2 and ALKBH3 Proteins. ALKBH2, ALKBH3 and AlkB were expressed and purified as described and shown previously.^{47,55} In the following section, AlkB is used as an example to illustrate the purification protocols. Briefly, His-tagged AlkB was obtained by transforming pET28a-AlkB into *E. coli* Rosetta2(DE3)pLysS or BL21(DE3)pLysS cells and protein expression was induced by the addition of 1mM isopropyl-b-D-thiogalactopyranoside

(IPTG) at 37 °C (37 °C for ALKBH2 and 30 °C for ALKBH3). The expressed protein was purified by affinity chromatography. Thrombin was used to digest His-tag containing AlkB protein. The final purified protein was stored -80 °C in AlkB storage buffer as previously described.⁵⁵

Enzymatic Reaction. To assay the AlkB family demethylase activity toward the two substrates in ss- and ds-DNA, the enzymatic reactions were performed at 37 °C at different time points for the kinetic study of the AlkB reaction in buffer [70.0 μ M $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, 0.93 mM α KG (0.1 mM for 2-HG inhibition assay), 1.86 mM ascorbic acid and 46.5 mM HEPES (pH 8.0)]. The reactions were stopped by adding 10 mM EDTA followed by heating at 95 °C for 5 min. Typically, the purified proteins were incubated with oligonucleotides containing DNA adducts in the presence of all cofactors in a 20 μ L reaction volume. In order to separate substrate and product, 16mer m1A and A or 16mer m3C and C, the HPLC condition started with a 5 min gradient of 1.5 M ammonium acetate from 50% to 65%, followed by 2 min 70% ammonium acetate. The column was DNAPac PA-100 (4 $\times$ 250 mm) (Thermo Scientific). The UV detection wavelength was at 260 nm. Each reaction was carried out in triplicate.

For the double-stranded DNA substrates, 1.5 equivalents of the 23mer complementary oligonucleotides, 23mer-Tcp or 23mer-Gcp, were annealed with 16mer oligos by heating the mixture at 80 °C for 10 min and then cooling down to room temperature with the rate of 1 °C/10 s. The post-reaction treatments were similar to those for the ss-DNA reactions, except 1.75 equivalents of 23mer-A and 23mer-C were added together with 10 mM EDTA followed by heating up to 95 °C for 10 min and then cooled down

to room temperature with the same rate as used for annealing. The quantification method was the same as described above. Each reaction was carried out in triplicate.

Kinetic Studies. To determine K_M , k_{cat} and K_i values for the repair reactions, initial rates were obtained by keeping the DNA substrate and enzyme concentration constant and varying αKG concentration with or without various concentrations of D- or L-2HG (0, 1.0, 3.0, 5.0, 7.0, 9.0, 37.3 mM). All reactions were performed at 37 °C in triplicate and the data were analyzed by GraphPad Prism 5 with the Michaelis-Menten kinetics model. The inhibition curves were fit to the equation:
$$1/V_0 = 1/V_{max} + (K_M/V_{max}) \times (1 + [I]/K_i) \times (1/[S]).$$

RESULTS

To test the inhibitory effect of D- and L-2HG on the AlkB family enzymes, we first chemically synthesized oligonucleotides by site-specifically incorporating m1A and m3C, the major substrates of the AlkB family enzymes.^{24,43} We also expressed and purified recombinant human ALKBH2 and ALKBH3 proteins, and the *E. coli* AlkB protein.⁵⁵ Then, we performed kinetic experiments to determine the k_{cat} and K_M of the three enzymes as they repair the two adducts in both ds- and ss-DNA. After that, we measured the K_i of D- and L-2HG on the repair reactions, and finally evaluated the inhibitory effect of the oncometabolites in the concentration range reported to occur in certain human cancers.

Oligonucleotide Synthesis and Protein Purification. Two 16mer oligonucleotides containing m1A and m3C were chemically synthesized with the sequence 5'-GAAGACCTXGGCGTCC-3' (X denotes the alkylated base).⁴³ After HPLC purification, the identity of the oligonucleotides was confirmed by comparing the theoretical m/z of the oligonucleotides with the observed m/z from high resolution LC-MS (Table S1). The genes for *E. coli* AlkB and its human homologs ALKBH2 and ALKBH3 were cloned into pET28a+ expression vector; the incorporation of the correct sequences was confirmed by sequencing the corresponding plasmids. The three proteins were then expressed in *E. coli* hosts, isolated and purified by affinity chromatography as described in Experimental Section.⁵⁵

Enzymatic Assay for Measuring Kinetic Constants. For each enzymatic reaction, the adduct-containing oligonucleotide was incubated with the necessary cofactors for the AlkB reaction: Fe(II), α KG, and ascorbic acid (see Experimental Section) in either

ss- or ds-DNA. Below, m1A will be used as an example to explain the HPLC analyses. For the ss-DNA reactions, the starting material 16mer m1A (1.5 min in Figure 2a) and product 16mer A (2.8 min in Figure 2b) were well separated by anion exchange HPLC, and the amount of each was quantified by reference to standard curves. For ds-DNA repair reactions, we initially used a 16mer complementary oligonucleotide. However, the dsDNA of starting material (16mer complementary plus 16mer m1A) and the dsDNA of product (16mer complementary plus 16mer A) could not be fully separated under various HPLC conditions, thus making the quantification of reactions challenging. Therefore, we adopted a longer complementary oligonucleotide (23mer Tcp, 5.6 min in Figure 2c), which provided a similar repair efficiency as the 16mer complementary oligonucleotide. In the analysis of the 23mer reaction, the dsDNA of starting material (23mer Tcp plus 16mer m1A, 7.5min in Figure 2c) and the dsDNA of product (23mer Tcp plus 16mer A, 7.7min in Figure 2c) still could not be fully separated under the HPLC condition. Consequently, we designed another 23mer oligonucleotide that was fully complementary to 23mer Tcp (23mer A, 5.5min in Figure 2d). After the dsDNA reaction with 23mer Tcp, 23mer A was added to the reaction mixture, and the mixture was heated to 80 °C for 10 min and then slowly cooled down to room temperature. The addition of 23mer A allowed the 23mer Tcp formed perfect dsDNA with 23mer A (9.1 min in Figure 2d), thus releasing 16mer m1A and 16mer A from their previous complementarity with 23mer Tcp. Under these conditions, the 16mer m1A and 16mer A in the dsDNA repair reaction were well separated and quantified by the HPLC analyses (Figure 2d). A similar analytical strategy was successfully applied to m3C dsDNA repair reactions.

Figure 2. HPLC analyses of the DNA repair reactions. a) Starting material 16mer oligonucleotide containing m1A at the lesion site in ss-DNA reaction. b) ss-Product 16mer oligonucleotide containing A at the “lesion site” in the ss-DNA reaction. c) ds-DNA reaction products of 16mer m1A with 23mer Tcp. The mixture containing ss-16mer m1A, ss-16mer A, ds-16mer m1A:23mer Tcp, and ds-16mer A:23mer Tcp. The latter two species were not fully separable by HPLC. d) ds-DNA reaction products of 16mer m1A with 23mer Tcp and additional 23mer A, which is fully complementary to 23mer Tcp. The duplex of 23mer Tcp:23mer A was eluted as ds-DNA, thus releasing ss-16mer m1A and ss-16mer A for quantification.

Kinetic Analyses. After setting up a reliable procedure to quantify the conversion of the repair reactions, we carried out systematic kinetic analyses of the AlkB family enzymes repairing m1A and m3C. Because the purpose of this work was to measure the possible inhibition of D- and L-2HG on the repair reactions catalyzed by α KG-dependent AlkB family enzymes, the kinetic parameters of α KG in the repair reactions

were first measured. In a typical kinetic analysis (e.g., ALKBH2 repairing m1A), 5 μ M of oligonucleotide substrate, and 0.2 μ M ALKBH2 enzyme were mixed with different concentrations of α KG (5.0-70.0 μ M) and the extent of the repair reaction was quantified at different time points (see Experimental section for details). Because the repair of one molecule m1A to A requires the conversion of one molecule of α KG to succinate (Figure 1), the concentrations of the product 16mer A were used to calculate the k_{cat} and K_{M} of α KG. To ensure that the kinetic parameters reflect initial velocity, the DNA and enzyme concentrations were optimized to make sure the conversion of the repair reactions was less than 20%. All reactions were carried out in triplicate.

For ALKBH2 repair of m1A in ds-DNA (Table 1, Table S2, and Figure S1), the k_{cat} of α KG was $2.5 \pm 0.1 \text{ min}^{-1}$ and the K_{M} was $7.3 \pm 0.9 \mu\text{M}$, which are comparable to the literature reported kinetics parameters of other α KG dependent enzymes.^{7,25,56–61} The $k_{\text{cat}}/K_{\text{M}}$ value of ds-repair reaction ($0.34 \text{ min}^{-1} \cdot \mu\text{M}^{-1}$) shows that the repair was more efficient than in ss-DNA ($0.28 \text{ min}^{-1} \cdot \mu\text{M}^{-1}$), which agrees with the literature on the reported strand preference of ALKBH2.⁵⁹ The kinetic data of ALKBH2 repair of m3C showed a similar trend (Table 1, Table S2, Table S4 and Figure S2). In contrast to ALKBH2's preference for ds-DNA substrates, the kinetic parameters of ALKBH3 repair of ds-DNA substrates could not be measured due to the low conversion ratio even with very high enzyme loading, such as 5.0 μ M of ALKBH3 to 5.0 μ M substrate. Conversely, ALKBH3 could efficiently repair both DNA adducts in ss-DNA (Table 1). These results confirm the previously reported preference of ALKBH3's repair of ss-DNA substrates.^{48,55} The kinetic factors of the *E. coli* AlkB protein were also measured and the k_{cat} and K_{M} values agreed well with the literature reported k_{cat} and K_{M} of the

reactions (Table 1, Table S2 and S3, and Figure S3 and S4).^{59,60} The k_{cat}/K_M values of AlkB repair confirm that the enzyme prefers to repair m1A and m3C in ss-DNA as compared with ds-DNA.⁵⁵

Table 1. Kinetic constants of α KG as a substrate on ALKBH2, ALKBH3 and AlkB repair reactions.

Enzyme	Condition	K_M [μ M]	k_{cat} [min^{-1}]	k_{cat}/K_M [$\text{min}^{-1} \cdot \mu\text{M}^{-1}$]
ALKBH2	ss-m1A	4.1 ± 0.9	1.1 ± 0.1	0.28
	ds-m1A	7.3 ± 0.9	2.5 ± 0.1	0.34
	ss-m3C	1.4 ± 0.2	1.7 ± 0.1	1.20
	ds-m3C	1.9 ± 0.4	2.6 ± 0.1	1.34
ALKBH3	ss-m1A	2.3 ± 0.1	1.2 ± 0.0	0.51
	ss-m3C	1.9 ± 0.4	1.7 ± 0.0	0.87
AlkB	ss-m1A	7.1 ± 1.1	4.2 ± 0.2	0.59
	ds-m1A	12.7 ± 1.3	4.8 ± 0.2	0.38
	ss-m3C	19.9 ± 1.3	24.5 ± 0.7	1.23
	ds-m3C	10.8 ± 1.9	8.2 ± 0.4	0.76

2HG Inhibition of the DNA Repair Reactions Catalyzed by the AlkB Family

Enzymes. With reliable k_{cat} and K_M parameters of the three enzymes, we set out to measure the K_i values of D-2HG and L-2HG together with a positive control, N-oxalylglycine (N-OG), a commonly used inhibitor of α KG dependent enzymes.⁷ Because D-2HG and L-2HG are chiral molecules, polarimetry (P-2000 Digital Polarimeter, JASCO Inc.) was utilized to measure their optical activity in a 1 decimeter cell. The optical rotation of D-2HG was $+9.5^\circ$ ($c = 1.0$, 0.1M NaOH), which agreed

well to the value provided by the commercial source ($[\alpha]/D +8.5\pm1.5^\circ$, $c = 1.0$ in NaOH, Sigma-Aldrich Co LLC.). Similarly, the optical rotation of L-2HG was -8.0° , which was consistent with the reported $-8.5\pm1.5^\circ$ value. These values confirm the chirality and purity of the two enantiomers.

For the inhibition of the ALKBH2 repair reaction on m1A in ds- and ss-DNA, the K_i values for D-2HG are $280 \pm 61 \mu\text{M}$ and $405 \pm 61 \mu\text{M}$, respectively (Table 2, Figure 3a and 3b, Table S5). These data indicate that D-2HG has a stronger binding affinity for the complex of ALKBH2 with ds-DNA than ss-DNA. For L-2HG reactions, the K_i values are similar but smaller (stronger inhibition) than with the corresponding D-2HG reactions (Table 2, Figure 3d and 3e, Table S5). The K_i values of N-OG show much stronger inhibition (with about 10 times more potency, Table 2) of all repair reactions with the K_i values ranging from 6 to 40 μM . For the inhibition of ALKBH3, a similar trend was observed for each individual reaction for the K_i values: D-2HG > L-2HG > N-OG (Table 2). For the inhibition of AlkB-catalyzed reactions, there is no clear trend in the inhibitory potency between the D- and L-2HG; N-OG, however, is a stronger inhibitor than either of 2HG isomers. We also measured the IC_{50} of D- and L-2HG on the three enzymes (Table S7); in general, the IC_{50} values correlate well with the K_i values.

Table 2: Inhibition constant (K_i) of D-2HG, L-2HG and N-OG. The individual inhibition reactions were depicted in Figure S6 to S15.

Enzyme	Condition	K_i [μ M]		
		D-2HG	L-2HG	N-OG
ALKBH2	ss-m1A	405 ± 61	275 ± 41	30 ± 7
	ds-m1A	280 ± 61	180 ± 36	16 ± 5
	ss-m3C	152 ± 13	64 ± 3	40 ± 7
	ds-m3C	79 ± 11	76 ± 11	7 ± 2
ALKBH3	ss-m1A	545 ± 77	185 ± 23	27 ± 4
	ss-m3C	490 ± 46	228 ± 16	37 ± 3
AlkB	ss-m1A	571 ± 166	337 ± 99	0.4 ± 0.1
	ds-m1A	529 ± 126	598 ± 173	2.0 ± 1.3
	ss-m3C	447 ± 113	276 ± 111	0.4 ± 0.1
	ds-m3C	230 ± 55	308 ± 108	0.2 ± 0.0

Figure 3. Inhibition of ALKBH2 repair of m1A in ss- and ds-DNA by D- and L-2HG. a) Inhibition of D-2HG on m1A repair in ss-DNA. b) Inhibition of D-2HG on m1A repair in ds-DNA. c) Inhibition of D-2HG on m1A repair under D-2HG: α KG = 373:1 ratio conditions. d) Inhibition of L-2HG on m1A repair in ss-DNA. e) Inhibition of L-2HG on m1A repair in ds-DNA. f) Inhibition of L-2HG on m1A repair under L-2HG: α KG = 28:1 ratio conditions.

To make our experiments more relevant with regard to the anticipated cellular concentrations of metabolites/oncometabolites observed in human tumors, we also

evaluated the extent inhibition of the ALKBH2 and ALKBH3 repair reactions by varying the ratios of D- or L-2HG to α KG. For D-2HG inhibition, we tested a ratio of concentrations for D-2HG: α KG = 373:1, which was observed in glioma patients with IDH mutations (detailed information see the Discussion section).⁶ The concentration of α KG was fixed at 100 μ M to make sure that the kinetic analyses reflected steady state catalysis (Figure S5). We found that the repair efficiencies of ALKBH2 and ALKBH3 were 73-88% inhibited under such conditions, (Figure 3c, Table 3). For L-2HG inhibition, we tested a ratio of L-2HG: α KG= 28:1, which was reported in patients with kidney cancers (see Discussion section).¹⁰ We found 48-58% of ALKBH2 and 31-40% of ALKBH3's activity was inhibited under this condition. These results suggest that the strong inhibition on DNA repair observed in the *in vitro* experiments may also occur in tumor cells of cancer patients.

Table 3. Inhibition ratio of D-2HG (373 fold to α KG) and L-2HG (28 fold to α KG) on ALKBH2 and ALKBH3.

Enzyme	Condition	% Inhibition of 373-fold D-2H to α KG	% Inhibition of 28-fold L-2HG to α KG
ALKBH2	ss-m1A	86	53
	ds-m1A	77	57
	ss-m3C	88	48
	ds-m3C	88	58
ALKBH3	ss-m1A	80	40
	ds-m1A	81	37
	ss-m3C	73	32
	ds-m3C	87	31

α KG Recovery of 2HG's Inhibition on the Repair Enzymes. Because 2-HG and α KG are structurally similar, researchers hypothesized that 2HG is able to replace α KG in the active site of α KG dependent enzymes and competitively inhibit their enzymatic activities.⁷ Crystal structures of histone demethylases show that D-2HG binds to the same site as α KG in the catalytic center.⁷ We tested the competition between 2HG and α KG in the DNA repair reactions. Using ALKBH2 repair of m1A as an example, the repair ratios without adding 2HG were controlled to be around 60% under different α KG concentrations (0.1, 0.5 and 1.0 mM, Figure 4 and Table S6). For the inhibition reactions, D-2HG was added at a fixed concentration (10 mM) in the reaction mixture, which contained ALKBH2 and necessary cofactors. Then, different concentrations of α KG were added and mixed. After that the reaction was initiated by adding the oligonucleotide substrates. When 0.1 mM α KG was present, the conversion decreased to 22%. When 0.5 mM and 1.0 mM α KG were added, the repair ratio increased to 35% and 38%, respectively (Figure 4 and Table S6). This observed trend of reactivity recovery is consistent with the notion that D-2HG acts as a competitive inhibitor in the α KG-dependent DNA repair reactions.⁷ Similar recovery patterns were observed for all other D- and L-2HG inhibition reactions on all three enzymes (Figure 4 and Table S6).

Figure 4. Addition of α KG reverses the inhibitory effect of 2-HG toward ALKBH2 repair of m1A. Different concentrations of α KG were added to a fixed concentration of 2HG (10 mM) to recover the repair of m1A by a) ALKBH2, b) ALKBH3 and c) AlkB.

DISCUSSION

Biological Implications of 2HG Inhibition of DNA Repair Enzymes. In the current study, we have shown that both D- and L-enantiomers of the oncometabolite 2HG can significantly inhibit the human DNA repair enzymes ALKBH2 and ALKBH3 under physiologically relevant concentrations. The concentrations of D-2HG and α KG on average in glioma cells are 15.5 μ mol/g, and 0.0415 μ mol/g, respectively, which correspond to a concentration ratio between D-2HG and α KG of 373 to 1.^{6,7} Under this ratio condition, the repair activities of ALKBH2 and ALKBH3 were 73-88% inhibited (Table 3). The concentrations of L-2HG and α KG on average in kidney cancer cells are 1.15 μ mol/g and 0.0484 μ mol/g, respectively, which corresponds to a concentration ratio between L-2HG and α KG of 28 to 1.¹⁰ Under this ratio condition, ALKBH2 and ALKBH3's repair activities are 31-58% inhibited (Table 3). Although the relative concentration of L-2HG (1.15 μ mol/g) is more than 10 times lower than D-2HG (15.5 μ mol/g), the ALKBH2 and ALKBH3 enzymes are still soundly inhibited by L-2HG partially due to the higher binding affinity of L-2HG (i.e., lower K_i) than D-2HG (Table 2). The extent of inhibition in both cases was measured when the concentration of α KG was fixed at 100 μ M, to ensure steady state catalysis. However, at lower concentrations of α KG, (i.e., 50 or 20 μ M), the efficiency of adduct repair decreased even further. The cellular concentrations of α KG are typically around 40 to 50 μ M (0.0415 and 0.0484 μ mol/g or mM) in cancer patients,^{6,10} which are near to the 50-100 μ M range used in our experiments. Our data also show that, consistent with competitive inhibition of 2HG, the inhibition activity in the repair reaction reflects primarily the ratio between 2HG and α KG. ALKBH2 and ALKBH3 are enzymes that repair alkyl DNA damage;

hence, inhibition of DNA repair leads to alkylation product accumulation, less cellular survival, and increased mutations, which affect the resistance/sensitivity balance to alkylating chemotherapeutics. The elevation of both D- and L-2HG in cancer cells may contribute to the increased mutation rate and exacerbate tumorigenesis and progression.

Strand Preference of the Three Repair Enzymes. According to the literature, ALKBH2 prefers to repair m1A and m3C in ds-DNA, whereas ALKBH3 and AlkB prefer to repair those adducts in ss-DNA.^{24,30,55} We tested the repair activity in both ss-DNA and ds-DNA substrates in this study. The experimental results reported in this paper provide a strong kinetic basis for the previous observations. For ALKBH2, the k_{cat}/K_M values of ds-repair are higher than the repair in ss-DNA (Table 1). By contrast, the k_{cat}/K_M values of AlkB repair are higher for ss-DNA substrates than for ds-DNA substrates (Table 1). For ALKBH3, we were only able to measure the kinetic parameters for ss-repair, as the ds-repair reactions were too inefficient to evaluate. These results agree with and add quantitative detail to previous observations that ALKBH3 strongly prefers to repair adducts in ss-DNA.

Other α KG/Fe(II)-Dependent Enzymes may be Inhibited by Oncometabolites. There are about 80 proteins in the α KG/Fe(II)-dependent enzyme family, including jmjc, prolyl hydroxylase, TET, and the AlkB family enzymes.²¹⁻²³ Studies have demonstrated that D- and L-2HG inhibit jmjc and TET family proteins.^{7,62} In addition to 2HG, intermediates in the TCA cycle such as succinate and fumarate have also been found to exhibit higher-than-normal concentrations in different cancer cells (Figure 1b).¹⁸ Given their structural similarities to α KG and 2HG, these metabolites could also perturb α KG-dependent enzymatic activities in the cell, especially DNA repair processes that are

related to the AlkB family enzymes. Systematic studies are needed to explore these possibilities and correlate these biochemical results with clinical observations. These studies are also pivotal for the design and development of therapeutic agents that target the abnormal metabolic pathways of cancer.

ASSOCIATED CONTENT

Supporting Information

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

Tables of molecular weights and m/z values, initial rate for kinetic studies, and IC₅₀ values of inhibition reactions; figures for steady-state kinetics, repair percentage of reactions, and inhibition curves of different repair enzymes (PDF)

AUTHOR INFORMATION

Corresponding Author

Mailing address:

7 Green House Road, Pharmacy Building, Room 495 N, Kingston, Rhode Island 02881, United States.

Phone number: (401) 874-9361

Email address: deyuli@uri.edu

Author Contributions

[‡]F.C. and B.K. contributed equally to this work.

Funding Sources

This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 2 P20 GM103430, and a Medical Research Funds grant from the Rhode Island foundation (to D. L.). This work was also supported, in whole or in part, by National

Institutes of Health Grants P01 CA26731, R01 CA080024, and P30 ES002109 (to J.M.E.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its directors Prof. Zahir Shaikh and Prof. David Rowley, and staff Dr. Al Bach, Ms. Kim Andrews and Ms. Patricia Murray for their kind help. We also want to thank Prof. Bongsup Cho, Prof. Roberta King, Mr. Aram Babcock, and Mr. Ang Cai for their support and helpful discussions.

ABBREVIATIONS

m1A, 1-methyladenine; m3C, 3-methylcytosine; ESI, electrospray ionization; TOF, time-of-flight, MS, mass spectrometry; ss, single stranded; ds, double stranded.

REFERENCES

- (1) Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G. J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K. W., Velculescu, V. E., Vogelstein, B., and Bigner, D. D. (2009) IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360, 765–773.
- (2) Parsons, D. W., Jones, S., Zhang, X., Lin, J. C.-H., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K. N., Shinjo, S. M. O., Yan, H., Riggins, G. J., Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E., and Kinzler, K. W. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807–1812.
- (3) Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Locke, D. P., Magrini, V. J., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., McMichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X., Lin, L., Schmidt, H., Tang, Y., Haipek, C., Wiechert, M. E., Ivy, J. V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath, S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., DiPersio, J. F., Wilson, R. K., and Ley, T. J. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* 361, 1058–1066.

- (4) Yang, H., Ye, D., Guan, K.-L., and Xiong, Y. (2012) IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 18, 5562–5571.
- (5) Cairns, R. A., and Mak, T. W. (2013) Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. *Cancer Discov.* 3, 730–741.
- (6) Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liao, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744.
- (7) Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., Liu, L., Jiang, W., Liu, J., Zhang, J., Wang, B., Frye, S., Zhang, Y., Xu, Y., Lei, Q., Guan, K.-L., Zhao, S., and Xiong, Y. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30.
- (8) Chowdhury, R., Yeoh, K. K., Tian, Y.-M., Hillringhaus, L., Bagg, E. A., Rose, N. R., Leung, I. K. H., Li, X. S., Woon, E. C. Y., Yang, M., McDonough, M. A., King, O. N., Clifton, I. J., Klose, R. J., Claridge, T. D. W., Ratcliffe, P. J., Schofield, C. J., and Kawamura, A. (2011) The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* 12, 463–469.
- (9) Shim, E.-H., and Sudarshan, S. (2015) Another small molecule in the oncometabolite mix: L-2-Hydroxyglutarate in kidney cancer. *Oncoscience* 2, 483–486.

- (10) Shim, E.-H., Livi, C. B., Rakheja, D., Tan, J., Benson, D., Parekh, V., Kho, E.-Y., Ghosh, A. P., Kirkman, R., Velu, S., Dutta, S., Chenna, B., Rea, S. L., Mishur, R. J., Li, Q., Johnson-Pais, T. L., Guo, L., Bae, S., Wei, S., Block, K., and Sudarshan, S. (2014) L-2-Hydroxyglutarate: an epigenetic modifier and putative oncometabolite in renal cancer. *Cancer Discov.* 4, 1290–1298.
- (11) Oldham, W. M., Clish, C. B., Yang, Y., and Loscalzo, J. (2015) Hypoxia-mediated increases in L-2-hydroxyglutarate coordinate the metabolic response to reductive stress. *Cell Metab.* 22, 291–303.
- (12) Intlekofer, A. M., Dematteo, R. G., Venneti, S., Finley, L. W. S., Lu, C., Judkins, A. R., Rustenburg, A. S., Grinaway, P. B., Chodera, J. D., Cross, J. R., and Thompson, C. B. (2015) Hypoxia induces production of L-2-hydroxyglutarate. *Cell Metab.* 22, 304–311.
- (13) Armitage, E. G., Kotze, H. L., Allwood, J. W., Dunn, W. B., Goodacre, R., and Williams, K. J. (2015) Metabolic profiling reveals potential metabolic markers associated with hypoxia inducible factor-mediated signalling in hypoxic cancer cells. *Sci. Rep.* 5, 15649.
- (14) Worth, A. J., Gillespie, K. P., Mesaros, C., Guo, L., Basu, S. S., Snyder, N. W., and Blair, I. A. (2015) Rotenone stereospecifically increases (S)-2-hydroxyglutarate in SH-SY5Y neuronal cells. *Chem. Res. Toxicol.* 28, 948–954.
- (15) Vatrinet, R., Leone, G., De Luise, M., Girolimetti, G., Vidone, M., Gasparre, G., and Porcelli, A. M. (2017) The α -ketoglutarate dehydrogenase complex in cancer metabolic plasticity. *Cancer Metab.* 5, 3.

- (16) Pavlova, N. N., and Thompson, C. B. (2016) The emerging hallmarks of cancer metabolism. *Cell Metab.* 23, 27–47.
- (17) Hirschey, M. D., DeBerardinis, R. J., Diehl, A. M. E., Drew, J. E., Frezza, C., Green, M. F., Jones, L. W., Ko, Y. H., Le, A., Lea, M. A., Locasale, J. W., Longo, V. D., Lyssiotis, C. A., McDonnell, E., Mehrmohamadi, M., Michelotti, G., Muralidhar, V., Murphy, M. P., Pedersen, P. L., Poore, B., Raffaghello, L., Rathmell, J. C., Sivanand, S., Vander Heiden, M. G., Wellen, K. E., and Target Validation Team. (2015) Dysregulated metabolism contributes to oncogenesis. *Semin. Cancer Biol.* 35, S129–150.
- (18) Nowicki, S., and Gottlieb, E. (2015) Oncometabolites: tailoring our genes. *FEBS J.* 282, 2796–2805.
- (19) Gaude, E., and Frezza, C. (2014) Defects in mitochondrial metabolism and cancer. *Cancer Metab.* 2, 10.
- (20) Su, X., Wellen, K. E., and Rabinowitz, J. D. (2016) Metabolic control of methylation and acetylation. *Curr. Opin. Chem. Biol.* 30, 52–60.
- (21) Markolovic, S., Wilkins, S. E., and Schofield, C. J. (2015) Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20712–20722.
- (22) Hausinger, R. P., and Schofield, C. J. (Eds.). (2015) 2-Oxoglutarate-dependent oxygenases. Royal Society of Chemistry, Cambridge, UK.
- (23) Martinez, S., and Hausinger, R. P. (2015) Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20702–20711.

- (24) Fedeles, B. I., Singh, V., Delaney, J. C., Li, D., and Essigmann, J. M. (2015) The AlkB family of Fe(II)/ α -ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.* 290, 20734–20742.
- (25) Wang, P., Wu, J., Ma, S., Zhang, L., Yao, J., Hoadley, K. A., Wilkerson, M. D., Perou, C. M., Guan, K.-L., Ye, D., and Xiong, Y. (2015) Oncometabolite D-2-hydroxyglutarate inhibits ALKBH DNA repair enzymes and sensitizes IDH mutant cells to alkylating agents. *Cell Rep.* 13, 2353–2361.
- (26) Johannessen, T.-C. A., Prestegarden, L., Grudic, A., Hegi, M. E., Tysnes, B. B., and Bjerkvig, R. (2013) The DNA repair protein ALKBH2 mediates temozolomide resistance in human glioblastoma cells. *Neuro-Oncol.* 15, 269–278.
- (27) Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419, 174–178.
- (28) Falnes, P. Ø., Johansen, R. F., and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419, 178–182.
- (29) Aravind, L., and Koonin, E. V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2, RESEARCH0007.
- (30) Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., and Lindahl, T. (2007) Repair of alkylated DNA: recent advances. *DNA Repair* 6, 429–442.
- (31) Fu, D., and Samson, L. D. (2012) Direct repair of 3,N(4)-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase. *DNA Repair* 11, 46–52.

- (32) Li, P., Gao, S., Wang, L., Yu, F., Li, J., Wang, C., Li, J., and Wong, J. (2013) ABH2 couples regulation of ribosomal DNA transcription with DNA alkylation repair. *Cell Rep.* 4, 817–829.
- (33) Chen, B., Liu, H., Sun, X., and Yang, C.-G. (2010) Mechanistic insight into the recognition of single-stranded and double-stranded DNA substrates by ABH2 and ABH3. *Mol. Biosyst.* 6, 2143–2149.
- (34) Yang, C.-G., Yi, C., Duguid, E. M., Sullivan, C. T., Jian, X., Rice, P. A., and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* 452, 961–965.
- (35) Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C.-G., Cui, Q., and He, C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* 468, 330–333.
- (36) Nay, S. L., Lee, D.-H., Bates, S. E., and O'Connor, T. R. (2012) Alkbh2 protects against lethality and mutation in primary mouse embryonic fibroblasts. *DNA Repair* 11, 502–510.
- (37) Dango, S., Mosammaparast, N., Sowa, M. E., Xiong, L.-J., Wu, F., Park, K., Rubin, M., Gygi, S., Harper, J. W., and Shi, Y. (2011) DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. *Mol. Cell* 44, 373–384.
- (38) Koike, K., Ueda, Y., Hase, H., Kitae, K., Fusamae, Y., Masai, S., Inagaki, T., Saigo, Y., Hirasawa, S., Nakajima, K., Ohshio, I., Makino, Y., Konishi, N., Yamamoto, H., and Tsujikawa, K. (2012) Anti-tumor effect of AlkB homolog 3 knockdown in hormone- independent prostate cancer cells. *Curr. Cancer Drug Targets* 12, 847–856.

- (39) Calvo, J. A., Meira, L. B., Lee, C.-Y. I., Moroski-Erkul, C. A., Abolhassani, N., Taghizadeh, K., Eichinger, L. W., Muthupalani, S., Nordstrand, L. M., Klungland, A., and Samson, L. D. (2012) DNA repair is indispensable for survival after acute inflammation. *J. Clin. Invest.* 122, 2680–2689.
- (40) Sundheim, O., Vågbø, C. B., Bjørås, M., Sousa, M. M. L., Talstad, V., Aas, P. A., Drabløs, F., Krokan, H. E., Tainer, J. A., and Slupphaug, G. (2006) Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* 25, 3389–3397.
- (41) Li, D., Fedeles, B. I., Shrivastav, N., Delaney, J. C., Yang, X., Wong, C., Drennan, C. L., and Essigmann, J. M. (2013) Removal of N-alkyl modifications from N(2)-alkylguanine and N(4)-alkylcytosine in DNA by the adaptive response protein AlkB. *Chem. Res. Toxicol.* 26, 1182–1187.
- (42) Maciejewska, A. M., Poznanski, J., Kaczmarek, Z., Krowisz, B., Nieminiusz, J., Polkowska-Nowakowska, A., Grzesiuk, E., and Kusmierek, J. T. (2013) AlkB dioxygenase preferentially repairs protonated substrates: specificity against exocyclic adducts and molecular mechanism of action. *J. Biol. Chem.* 288, 432–441.
- (43) Delaney, J. C., and Essigmann, J. M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14051–14056.
- (44) Delaney, J. C., Smeester, L., Wong, C., Frick, L. E., Taghizadeh, K., Wishnok, J. S., Drennan, C. L., Samson, L. D., and Essigmann, J. M. (2005) AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* 12, 855–860.

- (45) Frick, L. E., Delaney, J. C., Wong, C., Drennan, C. L., and Essigmann, J. M. (2007) Alleviation of 1,N6-ethanoadenine genotoxicity by the *Escherichia coli* adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U. S. A.* 104, 755–760.
- (46) Li, D., Delaney, J. C., Page, C. M., Chen, A. S., Wong, C., Drennan, C. L., and Essigmann, J. M. (2010) Repair of DNA alkylation damage by the *Escherichia coli* adaptive response protein AlkB as studied by ESI-TOF mass spectrometry. *J. Nucleic Acids* 2010, 369434.
- (47) Li, D., Delaney, J. C., Page, C. M., Yang, X., Chen, A. S., Wong, C., Drennan, C. L., and Essigmann, J. M. (2012) Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the *Escherichia coli* adaptive response protein AlkB. *J. Am. Chem. Soc.* 134, 8896–8901.
- (48) Singh, V., Fedeles, B. I., Li, D., Delaney, J. C., Kozekov, I. D., Kozekova, A., Marnett, L. J., Rizzo, C. J., and Essigmann, J. M. (2014) Mechanism of repair of acrolein- and malondialdehyde-derived exocyclic guanine adducts by the α -ketoglutarate/Fe(II) dioxygenase AlkB. *Chem. Res. Toxicol.* 27, 1619–1631.
- (49) Chang, S., Fedeles, B. I., Wu, J., Delaney, J. C., Li, D., Zhao, L., Christov, P. P., Yau, E., Singh, V., Jost, M., Drennan, C. L., Marnett, L. J., Rizzo, C. J., Levine, S. S., Guengerich, F. P., and Essigmann, J. M. (2015) Next-generation sequencing reveals the biological significance of the N(2),3-ethenoguanine lesion *in vivo*. *Nucleic Acids Res.* 43, 5489–5500.
- (50) Yi, C., and He, C. (2013) DNA repair by reversal of DNA damage. *Cold Spring Harb. Perspect. Biol.* 5, a012575.

- (51) Ougland, R., Rognes, T., Klungland, A., and Larsen, E. (2015) Non-homologous functions of the AlkB homologs. *J. Mol. Cell Biol.* 7, 494–504.
- (52) Fu, D., Calvo, J. A., and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* 12, 104–120.
- (53) Shrivastav, N., Li, D., and Essigmann, J. M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* 31, 59–70.
- (54) Zheng, G., Fu, Y., and He, C. (2014) Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.* 114, 4602–4620.
- (55) Chen, F., Tang, Q., Bian, K., Humulock, Z. T., Yang, X., Jost, M., Drennan, C. L., Essigmann, J. M., and Li, D. (2016) Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.* 29, 687–693.
- (56) Zhu, C., and Yi, C. (2014) Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. *Angew. Chem. Int. Ed Engl.* 53, 3659–3662.
- (57) Ehrismann, D., Flashman, E., Genn, D. N., Mathioudakis, N., Hewitson, K. S., Ratcliffe, P. J., and Schofield, C. J. (2007) Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem. J.* 401, 227–234.
- (58) Ergel, B., Gill, M. L., Brown, L., Yu, B., Palmer, A. G., and Hunt, J. F. (2014) Protein dynamics control the progression and efficiency of the catalytic reaction cycle of the *Escherichia coli* DNA-repair enzyme AlkB. *J. Biol. Chem.* 289, 29584–29601.

- (59) Lee, D.-H., Jin, S.-G., Cai, S., Chen, Y., Pfeifer, G. P., and O'Connor, T. R. (2005) Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J. Biol. Chem.* 280, 39448–39459.
- (60) Roy, T. W., and Bhagwat, A. S. (2007) Kinetic studies of *Escherichia coli* AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res.* 35, e147.
- (61) Yu, B., and Hunt, J. F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14315–14320.
- (62) Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H. F., Tallman, M. S., Sun, Z., Wolniak, K., Peeters, J. K., Liu, W., Choe, S. E., Fantin, V. R., Paietta, E., Löwenberg, B., Licht, J. D., Godley, L. A., Delwel, R., Valk, P. J. M., Thompson, C. B., Levine, R. L., and Melnick, A. (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.

SUPPORTING INFORMATION

Oncometabolites D- and L-2-hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions

Fangyi Chen,^{†,‡} Ke Bian,^{†,‡} Qi Tang,[†] Bogdan I. Fedeles,^{⊥,§,‡} Vipender Singh,^{⊥,§,‡,Δ}
Zachary T. Humulock,[†] John M. Essigmann,^{⊥,§,‡} and Deyu Li^{*,†}

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

Departments of [⊥]Biological Engineering, [§]Chemistry, and [‡]Center for Environmental
Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts
02139, United States.

Present Address: ^ΔNovartis Institute of Biomedical Research, Cambridge,
Massachusetts 02139, United States.

Corresponding Author: *E-mail: deyuli@uri.edu

[‡]F.C. and K.B. contributed equally to this work.

TABLE OF CONTENTS

Table S1. Calculated and observed molecular weight (MW) and m/z value of oligonucleotides used in the enzymatic reactions.

Table S2. Initial rate for kinetic studies of ALKBH2, ALKBH3 and AlkB on α KG as a substrate under different conditions.

Table S3. Initial rate for kinetic studies of ALKBH2 and ALKBH3 on DNA adducts as substrates.

Table S4. Kinetic constants of ALKBH2 and ALKBH3 reactions on DNA adducts as substrates.

Table S5. 2-HG inhibition on ALKBH2 demethylation of ss- and ds-m1A.

Table S6. Addition of α KG reverses the inhibitory effect of 2HG toward ALKBH2, ALKBH3 and AlkB.

Table S7: IC₅₀ (50% inhibition concentration) of L-2HG, D-2HG and N-OG on ALKBH2, ALKBH3 and AlkB.

Figure S1. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by ALKBH2.

Figure S2. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by ALKBH3.

Figure S3. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by AlkB.

Figure S4. Steady-state kinetic studies probing the influence of adducts in the demethylation reactions catalyzed by ALKBH2 and ALKBH3.

Figure S5. The repair percentage of AlkB on ss-m1A under various concentrations of α KG.

Figure S6. Inhibition of ALKBH2-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.

Figure S7. Inhibition of ALKBH2-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.

Figure S8. Inhibition of ALKBH2-mediated ds-m1A repair by D-2HG, L-2HG and N-OG.

Figure S9. Inhibition of ALKBH2-mediated ds-m3C repair by D-2HG, L-2HG and N-OG.

Figure S10. Inhibition of ALKBH3-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.

Figure S11. Inhibition of ALKBH3-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.

Figure S8. Inhibition of ALKBH2-mediated ds-m1A repair by D-2HG, L-2HG and N-OG.

Figure S9. Inhibition of ALKBH2-mediated ds-m3C repair by D-2HG, L-2HG and N-OG.

Figure S10. Inhibition of ALKBH3-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.

Figure S11. Inhibition of ALKBH3-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.

Figure S12. Inhibition of AlkB-mediated ss-m1A repair by D-2HG, L-2HG and N-OG.

Figure S13. Inhibition of AlkB-mediated ss-m³C repair by D-2HG, L-2HG and N-OG.

Figure S14. Inhibition of AlkB-mediated ds-m¹A repair by D-2HG, L-2HG and N-OG.

Figure S15. Inhibition of AlkB-mediated ds-m³C repair by D-2HG, L-2HG and N-OG.

Table S1. Calculated and observed molecular weight (MW) and m/z value of oligonucleotides used in the enzymatic reactions. The sequence of the 16mer oligonucleotides was 5'-GAAGACCTXGGCGTCC-3', where X indicates m1A or m3C. The sequence of the complementary 23mer oligonucleotides was 5'-CTGGGACGCCYAGGTCTTCACTG-3', where Y represents the position opposite the lesion site and contains the canonical bases T and G; these molecules were named 23-Tcp and 23-Gcp, respectively. Additionally, 23mer oligonucleotides complementary to 23-Tcp and 23-Gcp were also synthesized with the sequence 5'-CAGTGAAGACCTZGGCGTCCCAG-3', where Z denotes the regular bases A and C, and named 23-A and 23-C, respectively.

Oligonucleotide	MW (calculated) of neutral species	m/z (calculate) -4 charge peak	m/z (observed) -4 charge peak
16mer m1A	4902.88	1224.71	1224.71
16mer m3C	4878.87	1218.71	1218.70
Oligonucleotide	MW (calculated) of neutral species	m/z (calculate) -5 charge peak	m/z (observed) -5 charge peak
23mer Tcp	7028.19	1404.63	1404.66
23mer Gcp	7053.19	1409.63	1409.66
23mer A	7055.22	1410.04	1410.06
23mer C	7031.21	1405.23	1405.26

Table S2. Initial rate for kinetic studies of ALKBH2, ALKBH3 and AlkB on α KG as a substrate under different conditions. The reaction rate is in $\mu\text{M}/\text{min}$.

α KG (μM)	ALKBH2 reaction Rate		α KG (μM)	ALKBH2 reaction Rate	
	ss-m1A	ds-m1A		ss-m3C	ds-m3C
5.0	0.10 ± 0.00	0.16 ± 0.01	1.0	0.14 ± 0.00	0.10 ± 0.01
10.0	0.11 ± 0.01	0.21 ± 0.03	3.0	0.22 ± 0.00	0.15 ± 0.01
20.0	0.14 ± 0.01	0.26 ± 0.03	5.0	0.29 ± 0.02	0.18 ± 0.03
30.0	0.15 ± 0.00	0.30 ± 0.02	10.0	0.30 ± 0.02	0.21 ± 0.02
50.0	0.16 ± 0.02	0.32 ± 0.03	20.0	0.32 ± 0.01	0.23 ± 0.02
70.0		0.35 ± 0.03	30.0	0.32 ± 0.03	0.25 ± 0.02
α KG (μM)	ALKBH3 reaction rate				
	ss-m1A	ss-m3C			
5	0.20 ± 0.02	0.24 ± 0.00			
10	0.24 ± 0.01	0.29 ± 0.01			
20	0.26 ± 0.01	0.30 ± 0.01			
30	0.28 ± 0.01	0.30 ± 0.02			
50	0.28 ± 0.01	0.32 ± 0.01			
70	0.28 ± 0.02	0.33 ± 0.01			
α KG (μM)	AlkB reaction rate				
	ss-m1A	ds-m1A	ss-m3C	ds-m3C	
5.0	0.17 ± 0.01	0.20 ± 0.02	0.31 ± 0.03	0.46 ± 0.04	
10.0	0.25 ± 0.01	0.32 ± 0.05	0.49 ± 0.05	0.53 ± 0.04	
20.0	0.30 ± 0.03	0.44 ± 0.02	0.75 ± 0.09	0.78 ± 0.07	
30.0	0.35 ± 0.02	0.51 ± 0.02	0.87 ± 0.13	0.92 ± 0.04	
50.0	0.34 ± 0.04	0.60 ± 0.01	1.04 ± 0.10	0.99 ± 0.09	
70.0	0.39 ± 0.04	0.59 ± 0.04	1.17 ± 0.12	1.09 ± 0.06	

Table S3. Initial rate for kinetic studies of ALKBH2 and ALKBH3 on DNA adducts as substrates.

ALKBH2 reaction rate							
ss-m1A (μM)	V0 ($\mu\text{M}/\text{min}$)	ds-m1A (μM)	V0 ($\mu\text{M}/\text{min}$)	ss-m3C (μM)	V0 ($\mu\text{M}/\text{min}$)	ds-m3C (μM)	V0 ($\mu\text{M}/\text{min}$)
7.5	0.31 ± 0.02	5	0.26 ± 0.05	2.5	0.18 ± 0.01	2.5	0.22 ± 0.02
10	0.36 ± 0.03	7.5	0.34 ± 0.03	3.5	0.20 ± 0.01	5	0.25 ± 0.02
12.5	0.49 ± 0.02	10	0.41 ± 0.04	4.5	0.24 ± 0.02	7.5	0.30 ± 0.01
15	0.53 ± 0.05	12.5	0.47 ± 0.04	5.5	0.25 ± 0.02	10	0.37 ± 0.04
17.5	0.53 ± 0.08	15	0.52 ± 0.10	6.5	0.27 ± 0.02	12.5	0.41 ± 0.02
20	0.66 ± 0.03	17.5	0.53 ± 0.09	7.5	0.31 ± 0.04	15	0.42 ± 0.03
ALKBH3 reaction rate							
ss-m1A (μM)	V0 ($\mu\text{M}/\text{min}$)	ss-m3C (μM)	V0 ($\mu\text{M}/\text{min}$)				
7.5	0.30 ± 0.02	5	0.22 ± 0.01				
10	0.33 ± 0.02	7.5	0.24 ± 0.01				
12.5	0.35 ± 0.03	10	0.25 ± 0.01				
15	0.36 ± 0.02	12.5	0.27 ± 0.01				
17.5	0.37 ± 0.02	15	0.26 ± 0.01				
20	0.38 ± 0.03	17.5	0.28 ± 0.01				

Table S4. Kinetic constants of ALKBH2 and ALKBH3 reactions on DNA adducts as substrates.

Enzyme	Substrate	K_M [μM]	k_{cat} [min^{-1}]	k_{cat} / K_M [$\text{min}^{-1} \cdot \mu\text{M}^{-1}$]
ALKBH2	ss-m1A	35.9 ± 18.2	8.9 ± 3.2	0.25
	ds-m1A	12.6 ± 1.4	9.3 ± 0.5	0.74
	ss-m3C	4.4 ± 1.1	2.4 ± 0.3	0.53
	ds-m3C	5.0 ± 1.5	5.5 ± 0.6	1.10
ALKBH3	ss-m1A	3.5 ± 0.2	1.8 ± 0.0	0.52
	ss-m3C	1.9 ± 0.5	1.2 ± 0.1	1.03

Table S5. 2-HG inhibition on ALKBH2 demethylation of ss- and ds-m1A.**D-2HG** inhibition on ALKBH2 demethylation of **ss-m1A**.

Product D-2HG /μM /mM Time/min	0.0	1.0	3.0	5.0	9.0
3	1.23 ± 0.07	1.07 ± 0.06	0.80 ± 0.03	0.61 ± 0.05	0.47 ± 0.04
6	1.43 ± 0.07	1.22 ± 0.07	0.93 ± 0.05	0.69 ± 0.05	0.55 ± 0.02
9	1.50 ± 0.06	1.29 ± 0.07	0.99 ± 0.04	0.70 ± 0.02	0.57 ± 0.04
12	1.67 ± 0.06	1.44 ± 0.05	1.11 ± 0.04	0.82 ± 0.04	0.63 ± 0.03
15	1.78 ± 0.08	1.51 ± 0.08	1.18 ± 0.03	0.89 ± 0.04	0.70 ± 0.03

D-2HG inhibition on ALKBH2 demethylation of **ds-m1A**

Product D-2HG /μM /mM Time/min	0.0	1.0	3.0	5.0	9.0
5	1.54 ± 0.04	1.20 ± 0.08	0.85 ± 0.02	0.59 ± 0.07	0.37 ± 0.15
8	1.68 ± 0.11	1.43 ± 0.03	0.96 ± 0.02	0.74 ± 0.12	0.44 ± 0.10
11	1.67 ± 0.01	1.55 ± 0.19	1.06 ± 0.22	0.81 ± 0.13	0.52 ± 0.02
14	1.85 ± 0.01	1.59 ± 0.13	1.08 ± 0.11	0.82 ± 0.11	0.52 ± 0.06

D-2HG inhibition on ALKBH2 demethylation of **ss-m1A** and **ds-m1A** under 373 fold to αKG condition.

Repair ratio % D-2HG /mM Substrate	0.0	9.0	37.3
ss-m1A	93.9 ± 6.4	37.2 ± 1.5	12.7 ± 0.1
ds-m1A	97.7 ± 1.6	37.8 ± 2.8	22.3 ± 0.8

L-2HG inhibition on ALKBH2 demethylation of ss-m1A

Product Time/min L-2HG /μM /mM	0.0	1.0	3.0	5.0	9.0
3	1.34 ± 0.05	0.82 ± 0.03	0.54 ± 0.02	0.43 ± 0.03	0.34 ± 0.03
6	1.61 ± 0.05	1.04 ± 0.03	0.67 ± 0.01	0.53 ± 0.05	0.44 ± 0.00
9	1.62 ± 0.04	1.13 ± 0.08	0.72 ± 0.02	0.59 ± 0.02	0.46 ± 0.01
12	1.84 ± 0.09	1.28 ± 0.07	0.85 ± 0.02	0.68 ± 0.01	0.53 ± 0.01
15	1.96 ± 0.09	1.33 ± 0.07	0.92 ± 0.04	0.73 ± 0.02	0.54 ± 0.03

L-2HG inhibition on ALKBH2 demethylation of ds-m1A

Product Time/min L-2HG /μM /mM	0.0	1.0	3.0	5.0	9.0
5	2.21 ± 0.14	1.19 ± 0.09	0.67 ± 0.07	0.48 ± 0.05	0.38 ± 0.04
8	2.56 ± 0.13	1.74 ± 0.03	0.95 ± 0.03	0.61 ± 0.06	0.43 ± 0.02
11	2.84 ± 0.22	2.05 ± 0.12	1.12 ± 0.07	0.79 ± 0.08	0.51 ± 0.03
14	2.94 ± 0.03	2.18 ± 0.08	1.04 ± 0.31	0.85 ± 0.03	0.54 ± 0.03

L-2HG inhibition on ALKBH2 demethylation of ss-m1A and ds-m1A under 28 fold to αKG condition.

Repair ratio % Substrate L-2HG /mM	0.0	1.0	2.8
ss-m1A	80.8 ± 0.9	55.1 ± 2.9	38.3 ± 0.5
ds-m1A	85.0 ± 1.2	65.1 ± 2.0	37.0 ± 0.5

Table S6. Addition of α KG reverses the inhibitory effect of 2HG toward ALKBH2, ALKBH3 and AlkB. Both D-2HG and L-2HG concentrations were fixed at 10.0 mM.

Repair ratio% α KG	Group	ALKBH2			ALKBH3			AlkB		
		No inhibitor	D-2HG	L-2HG	No inhibitor	D-2HG	L-2HG	No inhibitor	D-2HG	L-2HG
0.1 mM		61.5 $\pm$ 1.5	22.0 $\pm$ 0.7	10.9 $\pm$ 0.5	60.0 $\pm$ 1.3	37.0 $\pm$ 0.6	20.9 $\pm$ 1.1	66.0 $\pm$ 0.2	24.2 $\pm$ 0.4	30.2 $\pm$ 1.5
0.5 mM		66.0 $\pm$ 0.6	35.1 $\pm$ 0.8	21.5 $\pm$ 0.6	59.4 $\pm$ 0.1	43.3 $\pm$ 0.9	34.2 $\pm$ 1.9	65.8 $\pm$ 0.3	29.2 $\pm$ 1.1	36.9 $\pm$ 1.7
1.0 mM		63.9 $\pm$ 0.5	38.1 $\pm$ 1.1	25.5 $\pm$ 0.4	57.9 $\pm$ 1.6	46.6 $\pm$ 2.2	39.2 $\pm$ 0.6	65.0 $\pm$ 0.3	29.0 $\pm$ 0.7	40.2 $\pm$ 2.6

Table S7: IC₅₀ (50% inhibition concentration) of L-2HG, D-2HG and N-OG on ALKBH2, ALKBH3 and AlkB.

Enzyme	Adduct	IC ₅₀ [mM]		
		D-2HG	L-2HG	N-OG
ALKBH2	ss-m1A	10.3 $\pm$ 1.6	7.0 $\pm$ 1.1	0.8 $\pm$ 0.2
	ds-m1A	4.7 $\pm$ 0.8	2.7 $\pm$ 0.5	0.2 $\pm$ 0.1
	ss-m3C	10.9 $\pm$ 1.0	4.6 $\pm$ 0.2	2.9 $\pm$ 0.5
	ds-m3C	4.2 $\pm$ 0.6	4.1 $\pm$ 0.6	0.4 $\pm$ 0.1
ALKBH3	ss-m1A	24.3 $\pm$ 3.5	8.3 $\pm$ 1.0	1.2 $\pm$ 0.2
	ss-m3C	26.4 $\pm$ 2.5	12.3 $\pm$ 0.8	2.0 $\pm$ 0.2
AlkB	ss-m1A	8.6 $\pm$ 2.5	5.1 $\pm$ 1.5	6.7E-03 $\pm$ 1.5E-03
	ds-m1A	4.7 $\pm$ 1.1	5.3 $\pm$ 1.5	1.8E-02 $\pm$ 1.1E-02
	ss-m3C	2.7 $\pm$ 0.7	1.7 $\pm$ 0.7	2.3E-03 $\pm$ 0.6E-03
	ds-m3C	2.4 $\pm$ 0.6	3.2 $\pm$ 1.1	1.7E-03 $\pm$ 0.3E-03

Figure S1. Steady-state kinetic studies probing the influence of αKG on adduct demethylation reactions catalyzed by ALKBH2. Data are in Table S2.

Figure S2. Steady-state kinetic studies probing the influence of αKG on adduct demethylation reactions catalyzed by ALKBH3. Data are in Table S2.

Figure S3. Steady-state kinetic studies probing the influence of α KG on adduct demethylation reactions catalyzed by AlkB. Data are in Table S2.

Figure S4. Steady-state kinetic studies probing the influence of adducts in the demethylation reactions catalyzed by ALKBH2 and ALKBH3. Data are in Table S3.

Figure S5. The repair percentage of AlkB on ss-m1A under various concentrations of αKG. Y-axis represents the percentage conversion of starting material m1A to product A.

Figure S6. Inhibition of ALKBH2-mediated ss-m1A repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S7. Inhibition of ALKBH2-mediated ss-m3C repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S8. Inhibition of ALKBH2-mediated ds-m1A repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S9. Inhibition of ALKBH2-mediated ds-m3C repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S10. Inhibition of ALKBH3-mediated ss-m1A repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S11. Inhibition of ALKBH3-mediated ss-m3C repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S12. Inhibition of AlkB-mediated ss-m1A repair by D-2HG, L-2HG and N-OG. Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S13. Inhibition of AlkB-mediated ss-m3C repair by D-2HG, L-2HG and N-OG.

Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S14. Inhibition of AlkB-mediated ds-m1A repair by D-2HG, L-2HG and N-OG.

Left: D-2HG, middle: L-2HG and right: N-OG.

Figure S15. Inhibition of AlkB-mediated ds-m3C repair by D-2HG, L-2HG and N-OG.

Left: D-2HG, middle: L-2HG and right: N-OG.

Published in Chemical Research in Toxicology, October 2017

**Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease
Condition**

Ke Bian^{†‡}, Fangyi Chen^{†‡}, Zachary T. Humulock^{†‡}, Qi Tang[†], and Deyu Li^{*†}

[‡]K.B., F.C. and Z.H. contributed equally to this work.

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

*Corresponding Author: Deyu Li

Department of Biomedical and Pharmaceutical Sciences

University of Rhode Island

495N Avedisian Hall, 7 Greenhouse Rd

Kingston, Rhode Island 02881, United States

Phone: +1-401-874-9361

Email address: deyuli@uri.edu

Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition

Ke Bian,^{†,±} Fangyi Chen,^{†,±} Zachary T. Humulock,^{†,±} Qi Tang,[†] and Deyu Li^{*,†}

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

KEYWORDS: *Copper, Wilson's disease, ALKBH2, ALKBH3, AlkB*

ABSTRACT

Disturbed metabolism of copper ions can cause diseases, such as Wilson's disease (WD). In this work, we investigated the inhibitory effect of Cu(II) ion on the AlkB family DNA repair enzymes, which are members of the Fe(II)/ α KG-dependent dioxygenase and includes human ALKBH2, ALKBH3, and *E. coli* AlkB proteins. None of the three proteins were significantly inhibited under normal cellular copper concentrations. But under WD related condition, we observed the activities of all three enzymes were strongly suppressed (inhibition from 95.2 to 100.0%). We also noted the repair efficiency under ds-DNA condition is less susceptible than ss-DNA to the inhibition.

Copper is an essential yet toxic metal for human physiological functions and the cellular concentrations of its ions, such as Cu(II) and Cu(I) ions, are tightly regulated. If the concentrations of these ions are dysregulated, certain diseases, such as Wilson's disease (WD), may occur.^{1,2} For patients under WD condition, the level of copper in human liver samples raises as much as 249.9 ng/mg (dry weight), which is significantly higher (more than 10 fold) than the level in normal cells (21.5 ng/mg).¹ These elevated copper levels create an inflammatory cellular environment in liver tissue cells and damage their lipid bilayers, thus forming a variety of DNA adducts. Nair et. al have reported the ethenoadenine (eA) and ethenocytosine (eC) DNA lesions are significantly increased under WD condition.^{1,3} These etheno DNA lesions are substrates of the AlkB family DNA repair enzymes, such as the human homologs ALKBH2 and 3, which use an Fe(II)/alpha-ketoglutarate (α KG)-dependent mechanism to oxidize the abnormal alkyl groups, ultimately restoring the undamaged DNA bases.⁴ Previously, the copper(II) ion has been reported to replace the Fe(II) ion in AlkB in a spectroscopic study, and to inhibit several Fe(II)/ α KG family enzymes, such as JMJD2A, JMJD2E and PHD2.⁵⁻⁷ Several divalent ions, such as Ni²⁺, Co²⁺ and Mn²⁺, have been demonstrated to inhibit other enzymes in this family, such as C-P4H, PLOD1, FIH, TauD, JMJD1A and ALKBH2.^{5,8} It is reasonable to predict, then, that the high level of copper ions accumulated under WD condition inhibits the DNA repair capacity of both ALKBH2 and 3. Therefore the elevated level of copper ions found in WD condition induces damages and increases the number of etheno adducts within the cell in a "two-fold" manner; it not only induces alkyl DNA damage through lipid peroxidation but also suppresses the repair of those adducts from the AlkB family DNA repair enzymes.

The AlkB protein in *E. coli* was discovered to be an Fe(II)/ α KG- dependent dioxygenase that repairs various alkyl DNA lesions (Figure 1).^{9,10} Nine human homologs of AlkB have been identified as ALKBH1-8 and FTO. Among these homologs, ALKBH2 and ALKBH3 have been characterized as DNA repair proteins.^{11,12} A variety of alkyl DNA lesions have been proven as substrates for AlkB, ALKBH2 and ALKBH3, both *in vitro* and *in vivo*. The lesions include methyl adducts, like 3-methylcytosine (m3C) and 1-methyladenine, etheno adducts, such as eA and eC, and other lesions with complex alkyl modifications to the DNA bases.¹² Among those adducts, m3C and m1A have been reported as the best substrates for the AlkB family enzymes.^{11,12} Previous *in vitro* studies on the inhibition of other Fe(II)/ α KG-dependent enzymes by Cu(II) ion have only provided the IC₅₀ values and the results have not been extrapolated to physiologically and pathologically relevant concentrations.⁶ In this work, for the first time we not only measured the IC₅₀ values of Cu(II) ion on the ABH2, ABH3, and AlkB proteins but also tested the inhibitory effects under normal cellular and Wilson's disease condition.

Figure 1. Repair mechanism of the AlkB family enzymes on alkyl DNA lesions and the inhibition of repair by Cu(II) ion. Adduct m3C is used here as an example to show the enzymatic catalysis.

To test the inhibitory effect of Cu(II) ion on the AlkB family enzymes, we first site-specifically synthesized a 16mer oligonucleotide containing m3C, the major substrate of the AlkB family enzymes.^{12,13} We expressed and purified recombinant human ALKBH2 and ALKBH3 proteins, and the *E. coli* AlkB protein.¹⁴ Then, we evaluated the repair efficiency of the enzymes (Figure S1) and the inhibitory effect of Cu(II) ion under different conditions. It has been reported that the cellular concentrations of “free” copper and iron ions are very limited. Under normal cellular conditions, the concentration of free copper ions is 1.5 μM , which is 1.5 fold greater than free iron ions (1.0 μM). For the WD patients, the level of free copper ion is increased to 24.4 μM .¹⁵ In this work, we tested different Fe(II) ion concentrations and selected 5.0 μM to ensure the maximal efficiency of repair by the AlkB family enzymes (Figure S2); this concentration was similar to the Fe(II) ion concentrations used in other inhibition studies.⁶ The inhibitory effects of Cu(II) ion were tested from 0.0 to 100.0 μM , among which 7.5 μM of Cu(II) was used to mimic normal cellular condition. Though the cellular levels of copper ion under WD condition vary considerably from different

reports, all of the levels are more than 10 fold higher than the normal cellular concentration.^{1,2,15} To make this work rigorous and relevant to pathological disease condition, we used 75.0 μM (10 fold) of copper ion to represent the WD condition.

For each enzymatic reaction, the m3C-containing oligonucleotides were incubated with the necessary cofactors for the AlkB reaction: Fe(II) ion, αKG , and ascorbic acid (see Experimental Section in Supporting Information) in either ss- or ds-DNA. In a typical inhibition test (e.g., ALKBH2 repairing m3C), 5.0 μM of oligonucleotide substrate, and 2.0 μM ALKBH2 enzyme were mixed with a fixed concentration of Fe(II) ion (5.0 μM) plus different concentrations of Cu(II) ion, and the extent of the repair reaction was quantified (see Experimental Section for details). Because the AlkB family enzymes are able to repair DNA lesions in both ss- and ds-DNA, the repair reactions of m3C were carried out under both conditions.

For the Cu(II) ion inhibition of the ALKBH2 repair reaction on m3C in ss- and ds-DNA (Figure 2 and Table S1), the IC_{50} values are 14.8 μM and 54.0 μM , respectively (Table 1). The IC_{50} values for AlkBH3 and AlkB were also measured (Table 1). From the IC_{50} tests, we found the IC_{50} values are higher in ds-DNA condition than ss-DNA condition for all three enzymes, despite the preference of AlkBH3 and AlkB to repair m3C in ss-DNA and AlkBH2 preferring repair in ds-DNA.¹² These results shows the repair under ds condition is less susceptible than ss condition to the inhibition from Cu(II) ion. One possibility is that copper ion is harder to access to the active site of an enzyme and replace its iron ion due to the more crowded environment under ds-DNA binding condition.

Figure 2. IC₅₀ measurements of Cu(II) ion on the repair of m3C by AlkBH2, AlkBH3, and AlkB under ss- and ds-DNA conditions.

Table 1. IC₅₀ values of Cu(II) ion inhibiting the AlkB family enzymes. The repair reactions of m3C were carried out under both ss- and ds-DNA conditions.

Protein	Substrate	IC ₅₀ (μM)
ALKBH2	ss-DNA	14.8
	ds-DNA	54.0
ALKBH3	ss-DNA	23.8
	ds-DNA	28.4
AlkB	ss-DNA	10.3
	ds-DNA	29.3

We then tested the inhibition of Cu(II) ion under normal physiological condition with the concentration at 7.5 μM. We found that none of the three proteins were significantly inhibited under this concentration, with the highest inhibition at 7.0% for ALKBH3 repairing m3C in ds-DNA. (Table S2 in SI). These results indicated that the repair efficiency of the AlkB family enzymes are barely affected by Cu(II) ions under normal cellular conditions. Next, we tested the inhibition of Cu(II) ions under WD condition with the concentration of Cu(II) ion at 75.0 μM. We found all three enzymes were strongly inhibited under WD condition (inhibition from 95.2 to 100.0%, Table S2).

In this work, we used m3C, the best substrate of AlkB, to demonstrate the AlkB family enzymes are strongly inhibited by high concentration of Cu(II) ion under WD condition. Other substrates, such as the etheno DNA lesions will be tested both *in vitro* and *in vivo*

in the future, which will provide a new explanation to the previously observed elevation of etheno adducts in Wilson's disease.¹ Considering that there are about eighty Fe(II)/ α KG-dependent enzymes in human body,¹¹ it is highly possible that some of those enzymes could be inhibited in a similar mechanism by the high concentrations of the copper ion under Wilson's disease condition.

ASSOCIATED CONTENT

Supporting Information

Two supplementary tables, two supplementary figures and Experimental Procedures (in PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: deyuli@uri.edu

Author Contributions

[±] K. B., F. C., and Z. H. contributed equally to this work.

Funding Sources

This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 2 P20 GM103430, and a Medical Research Funds grant from the Rhode Island foundation (to D. L.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its directors Prof. Zahir Shaikh and Prof. Bongsup Cho, and staff Dr. Al Bach, Ms. Kim Andrews and Ms. Patricia Murray for their kind help. We also want to thank Mr. Michael Vittori and Ms. Kerri Bradshaw for their support and helpful discussions.

ABBREVIATIONS

m3C, 3-methylcytosine; ss, single stranded; ds, double stranded; WD, Wilson's disease;
 α KG, alpha-ketoglutarate; eA, ethenoadenine; eC, ethenocytosine.

REFERENCES

- (1) Nair, J., Carmichael, P. L., Fernando, R. C., Phillips, D. H., Strain, A. J., and Bartsch, H. (1998) Lipid peroxidation-induced etheno-DNA adducts in the liver of patients with the genetic metal storage disorders Wilson's disease and primary hemochromatosis. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* 7, 435–440.
- (2) Wu, F., Wang, J., Pu, C., Qiao, L., and Jiang, C. (2015) Wilson's disease: a comprehensive review of the molecular mechanisms. *Int. J. Mol. Sci.* 16, 6419–6431.
- (3) Nair, J., Sone, H., Nagao, M., Barbin, A., and Bartsch, H. (1996) Copper-dependent formation of miscoding etheno-DNA adducts in the liver of Long Evans cinnamon (LEC) rats developing hereditary hepatitis and hepatocellular carcinoma. *Cancer Res.* 56, 1267–1271.
- (4) Delaney, J. C., Smeester, L., Wong, C., Frick, L. E., Taghizadeh, K., Wishnok, J. S., Drennan, C. L., Samson, L. D., and Essigmann, J. M. (2005) AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* 12, 855–860.
- (5) Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A., and Schofield, C. J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* 40, 4364–4397.
- (6) Sekirnik, R., Rose, N. R., Mecinović, J., and Schofield, C. J. (2010) 2-Oxoglutarate oxygenases are inhibited by a range of transition metals. *Met. Integr. Biometal Sci.* 2, 397–399.

- (7) Bleijlevens, B., Shivarattan, T., Sedgwick, B., Rigby, S. E. J., and Matthews, S. J. (2007) Replacement of non-heme Fe(II) with Cu(II) in the alpha-ketoglutarate dependent DNA repair enzyme AlkB: spectroscopic characterization of the active site. *J. Inorg. Biochem.* 101, 1043–1048.
- (8) Chen, H., Giri, N. C., Zhang, R., Yamane, K., Zhang, Y., Maroney, M., and Costa, M. (2010) Nickel ions inhibit histone demethylase JMJD1A and DNA repair enzyme ABH2 by replacing the ferrous iron in the catalytic centers. *J. Biol. Chem.* 285, 7374–7383.
- (9) Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419, 174–178.
- (10) Falnes, P. Ø., Johansen, R. F., and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419, 178–182.
- (11) Hausinger, R. P., and Schofield, C. J. (Eds.). (2015) 2-Oxoglutarate-dependent oxygenases. Royal Society of Chemistry, Cambridge, UK.
- (12) Fedeles, B. I., Singh, V., Delaney, J. C., Li, D., and Essigmann, J. M. (2015) The AlkB family of Fe(II)/ α -ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.* 290, 20734–20742.
- (13) Delaney, J. C., and Essigmann, J. M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14051–14056.
- (14) Chen, F., Tang, Q., Bian, K., Humulock, Z. T., Yang, X., Jost, M., Drennan, C. L., Essigmann, J. M., and Li, D. (2016) Adaptive response enzyme AlkB preferentially

repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.* 29, 687–693.

(15) McMillin, G. A., Travis, J. J., and Hunt, J. W. (2009) Direct measurement of free copper in serum or plasma ultrafiltrate. *Am. J. Clin. Pathol.* 131, 160–165.

SUPPORTING INFORMATION

Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition

Ke Bian,^{†,±} Fangyi Chen,^{†,±} Zachary T. Humulock,^{†,±} Qi Tang,[†] and Deyu Li^{*,†}

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

Corresponding Author: *E-mail: deyuli@uri.edu

[±] K.B., F.C., and Z. H. contributed equally to this work.

TABLE OF CONTENTS

Figure S1. HPLC analyses of the DNA repair reactions.

Figure S2. Dose-response curve of Fe(II) ions to the AlkB repair activity on m3C.

Table S1. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes in ss- and ds-DNA.

Table S2. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes under normal and Wilson's Disease conditions.

Experimental Section

Figure S1. HPLC analyses of the DNA repair reactions. Top: Starting material ss-16mer oligonucleotide containing m3C at the lesion site. Middle: Product ss-16mer oligonucleotide containing C at the “lesion site”. Bottom: Reaction mixture of starting material 16mer m3C and product 16mer C.

Figure S2. Dose-response curve of Fe(II) ions to the AlkB repair activity on m3C. In each reaction, 5.0 μM of oligonucleotide substrate, and 2.0 μM AlkB enzyme were mixed with different concentrations of Fe(II) ion, and the extent of the repair reaction was quantified.

Table S1. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes in ss- and ds-DNA. The data indicate the percentiles of repair. Each reaction was performed in triplicate.

Cu(II)/ μ M	ALKBH2		ALKBH3		AlkB	
	ss-DNA	ds-DNA	ss-DNA	ds-DNA	ss-DNA	ds-DNA
0.0	92.6 $\pm$ 3.5	100.0 $\pm$ 0	100.0 $\pm$ 0	77.6 $\pm$ 3.1	100.0 $\pm$ 0	100.0 $\pm$ 0
5.0	92.1 $\pm$ 2.4	97.9 $\pm$ 3.7	100.0 $\pm$ 0	75.5 $\pm$ 2.1	99.7 $\pm$ 0.1	100.0 $\pm$ 0
7.5	88.5 $\pm$ 0.6	98.1 $\pm$ 3.2	100.0 $\pm$ 0	72.2 $\pm$ 0.5	98.3 $\pm$ 0.5	100.0 $\pm$ 0
10.0	75.9 $\pm$ 0.5	99.0 $\pm$ 1.7	100.0 $\pm$ 0	66.4 $\pm$ 3.6	54.8 $\pm$ 3.6	100.0 $\pm$ 0
20.0	6.5 $\pm$ 3.1	98.5 $\pm$ 1.4	74.7 $\pm$ 5.5	51.9 $\pm$ 3.1	2.6 $\pm$ 0.3	98.9 $\pm$ 1.0
30.0	1.0 $\pm$ 0.3	90.1 $\pm$ 3.2	8.7 $\pm$ 0.8	35.5 $\pm$ 1.7	0.6 $\pm$ 0	45.3 $\pm$ 2.3
40.0	0 $\pm$ 0	76.5 $\pm$ 2.7	1.2 $\pm$ 1.1	25.7 $\pm$ 1.7	0 $\pm$ 0	19.1 $\pm$ 1.5
50.0	0 $\pm$ 0	54.4 $\pm$ 6	0.6 $\pm$ 1.1	16.7 $\pm$ 2.7	0 $\pm$ 0	8.4 $\pm$ 0.7
75.0	0 $\pm$ 0	4.0 $\pm$ 1.4	0 $\pm$ 0	3.7 $\pm$ 1.2	0 $\pm$ 0	2.0 $\pm$ 0.6
100.0	0 $\pm$ 0	0.4 $\pm$ 0.6	0 $\pm$ 0	1.7 $\pm$ 0.2	0 $\pm$ 0	0 $\pm$ 0

Table S2. Cu(II) ion inhibition on the demethylation of m3C by the AlkB family enzymes under normal and Wilson's Disease conditions. The data indicate the percentiles of inhibition. Each reaction was performed in triplicate.

	Cu/Fe ratio	ALKBH2		ALKBH3		AlkB	
		ss-DNA	ds-DNA	ss-DNA	ds-DNA	ss-DNA	ds-DNA
Normal	1.5	4.4	1.9	0.0	7.0	1.7	0.0
Wilson's Disease	15.0	100.0	96.0	100.0	95.2	100.0	98.0

Experimental Section

Oligonucleotide Synthesis. The synthesis of oligonucleotides (ODNs) were described previously.^{1,2} Briefly, the 16mer DNA oligonucleotide with the sequence 5'-GAAGACCTXGGCGTCC-3' (X represents the 3-methylcytosine), the complementary 23merGcp 5'-CTGGGACGCCGAGGTCTTCACTG-3' and 23merC 5'-CAGTGAAGACCTCGGCGTCCCAG-3', were synthesized by MerMade-4 oligonucleotide synthesizer. All ODNs were purified by HPLC (Thermo Fisher Scientific) with a DNAPac PA-100 Semi-Preparative column (Thermo Fisher Scientific). Solvent A of 100 mM mixture (1:1) of triethylamine-acetic acid in water and solvent B 100% acetonitrile were employed as mobile phases. The concentration of DNA was determined by NanoDrop (Thermo Fisher Scientific) with UV absorbance at 260 nm. The ODNs were characterized by HPLC–electrospray ionization triple quadrupole time-of-flight mass spectrometry (AB Sciex).

Expression and Purification of the AlkB, ALKBH2, and ALKBH3 Proteins. The expression and purification of AlkB, ALKBH2, and ALKBH3 were described previously.^{1,2} Briefly, His-tagged AlkB was obtained by transforming pET-28a(+)-AlkB into *E. coli* Rosetta2(DE3)pLysS (BL21(DE3)pLysS for ALKBH2 and ALKBH3) cells, followed by adding 1 mM isopropyl β -D-thiogalactopyranoside (IPTG) to induce the expression at 37 °C (37 °C for ALKBH2 and 30 °C for ALKBH3). The expressed proteins were purified by FPLC (GE healthcare) with an affinity chromatography column. The His-tag was removed by thrombin and the final purification step used ion-exchange column chromatography.

Enzymatic Reaction. The AlkB family demethylase activity reactions on ss- and ds-DNA were described previously.^{1,2} All reactions were performed at 37°C in reaction buffer [5 µM Fe(NH₄)₂(SO₄)₂, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM HEPES (pH 8.0)] for 1 hour. Varying concentrations of copper (II) chloride (0.0, 5.0, 7.5, 10.0, 20.0, 30.0, 40.0, 50.0, 75.0, and 100.0 µM) were used for the inhibition tests. 16mer-m3C (ss- or ds-DNA) was pre-mixed with reaction buffer in a concentration of 5.0 µM. The reactions were initiated by adding 0.2 µM AlkB (2.0 µM for ALKBH2 and ALKBH3). The reactions were stopped by adding 10.0 mM EDTA followed by heating to 95 °C for 5 min. For reactions with ds-DNA, most steps were identical to ss-DNA reactions, except 16mer m3C containing ODN was annealed to 23mer complimentary strand (23mer-Gcp) before reaction. The reaction was terminated by adding excessive amount of 23merC ODN, which was the perfect complement to 23mer-Gcp, with 10 mM EDTA, then heated up to 95°C and slowly cooled down to room temperature. Each reaction was performed in triplicate. All the results of reaction were analyzed by HPLC.

HPLC conditions. All reaction samples were quantified by DNAPac PA-100 column (4 mm × 250 mm, Thermo Scientific) with isocratic 60% mobile B, 1.5 M ammonium acetate, under a constant flow rate of 1.0 mL/min. Mobile A was water. The UV detection wavelength was 260 nm.

REFERENCES

- (1) Chen, F., Tang, Q., Bian, K., Humulock, Z. T., Yang, X., Jost, M., Drennan, C. L., Essigmann, J. M., and Li, D. (2016) Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.* 29, 687–693.
- (2) Chen, F., Bian, K., Tang, Q., Fedeles, B. I., Singh, V., Humulock, Z. T., Essigmann, J. M., and Li, D. (2017) Oncometabolites d- and l-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions. *Chem. Res. Toxicol.* 30, 1102–1110.

Prepared to submit to Nucleic Acids Research, March 2019

DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

Ke Bian,^{1,±} Stefan A.P. Lenz,^{2,±} Qi Tang,¹ Fangyi Chen,¹ Rui Qi,¹ Marco Jost,³
Catherine L. Drennan,^{3,4,6,7} John M. Essigmann,^{3,5,6} Stacey D. Wetmore,^{2,*} and Deyu
Li^{1,*}

[±]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

¹Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy,
University of Rhode Island, Kingston, Rhode Island 02881, United States

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

³Department of Chemistry, ⁴Department of Biology, ⁵Department of Biological
Engineering, ⁶Center for Environmental Health Sciences, and ⁷Howard Hughes
Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts
02139, United States

*Corresponding Authors: Deyu Li

Department of Biomedical and Pharmaceutical Sciences
University of Rhode Island

495N Avedisian Hall, 7 Greenhouse Rd
Kingston, Rhode Island 02881, United States
Phone: +1 (401) 874-9361
Email address: deyuli@uri.edu

Stacey Wetmore
Department of Chemistry and Biochemistry
University of Lethbridge
4401 University Drive West
Lethbridge, Alberta T1K 3M4, Canada
Phone: +1 (403) 329-2323
Email address: stacey.wetmore@uleth.ca

DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

Ke Bian,^{1,±} Stefan A.P. Lenz,^{2,±} Qi Tang,¹ Fangyi Chen,¹ Rui Qi,¹ Marco Jost,³ Catherine L. Drennan,^{3,4,6,7} John M. Essigmann,^{3,5,6} Stacey D. Wetmore,^{2,*} and Deyu Li^{1,*}

¹Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States

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

³Department of Chemistry, ⁴Department of Biology, ⁵Department of Biological Engineering, ⁶Center for Environmental Health Sciences, and ⁷Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[±]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

*To whom correspondence should be addressed. Deyu Li, Tel: +1(401) 874-9361; Email: deyuli@uri.edu. Correspondence may also be addressed to Stacey D. Wetmore, Tel: +1(403) 329-2323; Fax: +1(403) 329-2057; Email: stacey.wetmore@uleth.ca.

Present Address: [Fangyi Chen], Xiamen University, 4221 Xiang An South Road, Xiang An District, Xiamen, Fujian 361102, P.R. China. [Marco Jost], Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94158, United States.

ABSTRACT

5-Methylcytosine (5mC) in DNA CpG islands is an important epigenetic biomarker for mammalian gene regulation. It is oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) family enzymes, which are α -ketoglutarate (α -KG)/Fe(II)-dependent dioxygenases. In this work, we demonstrate that the epigenetic marker 5mC is biochemically modified to 5hmC, 5fC, and 5caC by another class of α -KG/Fe(II)-dependent proteins – the DNA repair enzymes in the AlkB family, which include human ALKBH2, ALKBH3, and AlkB in *E. coli*. Theoretical calculations indicate that these enzymes may bind 5mC in the syn-conformation, placing the methyl group comparable to 3-methylcytosine, the prototypic substrate of AlkB. This is the first demonstration of the AlkB proteins to oxidize a methyl group attached to carbon, instead of nitrogen, on a DNA base. These observations suggest a broader role in epigenetics for these DNA repair proteins.

INTRODUCTION

In mammals, methylation at the 5-position of cytosine (5-methylcytosine, 5mC, Figure 1) is the major form of DNA modification and occurs mainly on CpG dinucleotide sites.(1–3) This methylation is achieved and maintained by *S*-adenosylmethionine-dependent methyltransferases.(4) The reverse process termed demethylation of 5mC is first carried out by the ten-eleven translocation proteins (TET1, 2, and 3) through iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 1). Subsequently, the oxidation products 5fC and 5caC are removed and repaired back to unmethylated cytosine by thymine DNA glycosylase coupled with base excision repair.(1, 5–8) The 5mC modification and its three oxidative derivatives play important roles in epigenetic regulations and cell development: they function in transcriptional regulation, gene silencing and reprogramming.(1, 5) The misregulations of 5mC and derivatives have been associated with cancer and other diseases.(1, 9, 10)

The TET family enzymes belong to the α -ketoglutarate (α -KG)/Fe(II)-dependent dioxygenases; they have been extensively studied in the last decade for their biological functions, biochemical activities and structural features.(1, 5, 6, 11) In the structures of the TET enzymes, the highly conserved N-terminal β -hairpin-like element for DNA-base recognition and the C-terminal catalytic domain are also found in another family of α -KG/Fe(II)-dependent nucleic acid-modifying dioxygenases, the AlkB family proteins.(5, 12, 13) In this work, we show that the epigenetic modulator 5mC is modified *in vitro* to 5hmC, 5fC, and 5caC by the DNA repair enzymes in the AlkB family, including human ALKBH2, ALKBH3, and its *Escherichia coli* (*E. coli*) homolog AlkB

(Figure 1). The results suggest that these DNA repair enzymes may play a wider role in epigenetic modulation.

Figure 1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC.

Different homologs of the *E. coli* AlkB protein exist in prokaryotic and eukaryotic species; nine homologs exist in human cells (ALKBH1-8 and FTO).(12, 14) Among the nine homologs, ALKBH2 and ALKBH3 are DNA repair enzymes that protect the informational integrity of the genome.(5, 14–18) They use an α -KG/Fe(II)-dependent mechanism to oxidize aberrant alkyl groups, ultimately restoring the undamaged DNA bases.(14, 17, 18) The reported substrate scope of AlkB, ALKBH2 and ALKBH3 includes 3-methylcytosine (3mC, Figure 1), 1-methyladenine (1mA), 3-methylthymine (3mT) and 1-methylguanine (1mG), as well as other nitrogen-attached methyl lesions occurring at the Watson-Crick base pairing interface of DNA bases.(15, 16, 18–20)

Key structural information about the AlkB family enzymes has been obtained from crystal structures of AlkB and ALKBH2 bound to lesion-containing DNA, which reveal that their active sites share several characteristics. Specifically, the AlkB (ALKBH2 in

brackets) complex contains a metal center Fe(II) in the wild-type enzymes) coordinated to H131 (H171), H187 (H236), D133 (D173), α -KG, and molecular oxygen.(13, 21) Repair of 3mC or 1mA has been proposed to be further enhanced by interactions between D135 of AlkB (E175 of ALKBH2) and the exocyclic amino groups of the lesion.(22) Interestingly, all crystal structures of the AlkB enzymes indicate the lesions are bound in the *anti* glycosidic bond conformation (e.g., for 3mC: χ ; $\angle(\text{O4}'\text{C1}'\text{N1C2}) = 180^\circ \pm 90^\circ$, Figure S18a.)

Although both 5mC and 3mC carry a methyl modification, the methyl groups are on the opposite sides of the pyrimidine ring. It is reasonable to predict that members of the AlkB enzyme family may be able to oxidize 5mC if the methyl group can be positioned near the catalytic center. The *in vitro* experimental results demonstrated here reveal that the AlkB enzymes can not only repair DNA lesions, such as 3mC, but also modify the epigenetic biomarker 5mC and generate its oxidative derivatives. Our theoretical calculations suggest that AlkB enzymes bind 5mC in the *syn* glycosidic conformation ($\chi = 0^\circ \pm 90^\circ$, Figure S18b) to align the 5-methyl moiety for oxidation, which is similar to how the TET family enzymes bind 5mC. This paper is the first work to demonstrate the ability of the AlkB family enzymes to oxidize a methyl group that is attached to carbon, instead of nitrogen, on a DNA base.

MATERIAL AND METHODS

Synthesis of oligonucleotides containing 5mC and other modifications. All oligonucleotides used in this study were synthesized by solid-phase synthesis.(23–26) The 5mC and other phosphoramidites was purchased from Glen Research. Synthetic oligonucleotides were purified by reverse-phase HPLC and identified by electrospray ionization mass spectrometry (ESI-MS).

Protein expression and purification. The expression and purification of ALKBH2, ALKBH3 and AlkB proteins were described by previous published papers.(23, 24) ALKBH2 and ALKBH3 in storage buffer containing 50 mM N-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid (TAPS), 300 mM NaCl, 10% glycerol, and 1 mM 2-mercaptoethanol, pH 8.0, AlkB in similar storage buffer (10 mM Tris, 100 mM NaCl, 1 mM 2-mercaptoethanol, 10% glycerol, pH 8.0), were all stored at -80 °C.

Enzymatic reaction. The reactions were performed based on previous published procedures.(23, 24) A 16mer oligonucleotide containing one 5mC with sequence of 5'-GAAGACCT-5mC-GGCGTCC-3' was used as the substrate. One hundred pmol 16mer 5mC was mixed into reaction buffer (5 μ M Fe(NH₄)₂(SO₄)₂, 0.93 mM α -ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM TAPS, pH 7.0) in a total volume of 16 μ L. The reactions were started by adding proteins (170 pmol AlkB, 140 pmol ALKBH2 or 128 pmol ALKBH3), and incubated at 37 °C for 1 hour. EDTA was employed to quench the reactions and the reaction mixture was immediately heated up to 80 °C for 5min. A 16/23mer duplex DNA was preannealed as the substrate for dsDNA reaction.(24, 25) The 23mer oligonucleotide is complementary to the 16mer sequence plus 7 nucleotides

longer with the sequence of 5'-CTGGGACGCCGAGGTCTTCACTG-3'. The rest steps of dsDNA reaction were the same as ssDNA reaction described above. All reactions were carried out in triplicates.

Oligonucleotide digestion. The procedures of oligonucleotide digestion to deoxyribonucleoside were adopted from published procedure.(27) A digestion mixture was premade by adding 250 Units Benzonase (Sigma-Aldrich, MO), 300 mUnits phosphodiesterase I (Sigma-Aldrich, MO) and 200 Units alkaline phosphatase (Sigma-Aldrich, MO) to 5 mL Tris-HCl buffer (20 mM, pH 7.9) containing 100 mM NaCl and 20 mM MgCl₂. Reaction of oligonucleotide containing 5mC with AlkB was quenched by heating up to 80°C for 5 min, then digested by adding 50 µL digestion mixture and incubating at 37°C for 6 hours. The nucleoside products were analyzed by HPLC-MS. Samples were chromatographed on a Luna Omega Polar C18 column (150 X 4.6 mm, 5µm, 100Å, Phenomenex, CA) eluted at 1 mL/min with an acetonitrile gradient (1%-15%). ESI triple quadrupole time of flight mass spectrometry was conducted to detect nucleoside signals in the negative ion mode. Standard deoxyribonucleosides (including dA, dT, dC, dG, 5mdC, 5hmdC, 5fdC and 5cadC) were purchased from Berry & Associates (MI).

Protein digestion. The trypsin digestion of wide type AlkB and its variants were set up according to Trypsin-ultra™ kit (New England Biolabs, MA). Ten µg of AlkB proteins were mixed with 250 ng trypsin in trypsin-ultra buffer, incubating at 37 °C overnight. The protein fragments were analyzed by HPLC-MS.(24, 26) Standard protein fragments (including LFYHGIQPLK and LSLHQDK sequences) were purchased from New England Peptide (MA).

Preparing catalytically inactive variants of AlkB. The AlkB H131A, D133A, and H187A variants were generated by QuikChange Mutagenesis (Agilent), using pET24a-AlkB Δ N11 as the PCR template and primer pairs encoding the desired mutations. All three variants were purified essentially as described.(28) Briefly, One Shot BL21 Star (DE3)pLysS *E. coli* cells (Invitrogen) transformed with an AlkB variant construct were grown at 37 °C until OD₆₀₀ had reached ~0.4, at which point the temperature was lowered to 30 °C and protein production was induced by addition of 1 mM IPTG. Cells were harvested after 4 h and stored at –80 °C until use. For purification, cell pellets were resuspended in lysis buffer (10 mM Tris, pH 7.3, 300 mM NaCl, 2 mM CaCl₂, 10 mM MgCl₂, 5% (v/v) glycerol, 1 mM 2-mercaptoethanol) and lysed by sonication. After clarification by centrifugation, the lysate was loaded onto a Ni-NTA column (Qiagen), the column was washed twice with lysis buffer supplemented with 10 mM and 20 mM imidazole, and bound protein was eluted with lysis buffer supplemented with 70 mM and 250 mM imidazole. Elution fractions containing AlkB, as assessed by SDS-PAGE, were combined and dialyzed for 16 h against 50 mM 2-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]ethanesulfonic acid (TES), pH 7.1 with or without 20 mM NaCl, and loaded onto a 5 mL HiTrap SP cation exchange column (GE Healthcare). Bound AlkB was eluted with a linear gradient of 0.02 – 1 M NaCl over 12 column volumes (60 mL). Fractions containing pure AlkB were pooled and purity was established by SDS-PAGE.

Computational methodology. X-ray crystal structures of AlkB (PDB ID: 3O1S)(22) and ALKBH2 (PDB ID: 3RZJ)(21) bound to lesion-containing DNA were chosen as initial structures. The 3O1S crystal structure contains AlkB bound to an oxidized 3mC

intermediate-containing DNA, Fe(II) ion, and succinate. The oxidized intermediate was reverted to 3mC, and a water molecule bound to Fe(II) was modelled as an oxo ligand to generate the Fe(IV)–oxo complex. To generate the active Fe(IV)–oxo ALKBH2 complex, the cofactors (Fe(IV), succinate, and oxo ligand) and metal-binding amino acids (H131, D133, and H187) of the AlkB complex were super imposed onto the cofactors (Mn(II) and α -KG) and metal-binding amino acids (H171, D173, and H236) of the ALKBH2 complex. Hydrogen atoms were assigned using the tLEaP AMBER utility.(29) Protonation states of ionizable amino acids were initially assigned using H++(30) and adjusted using chemical intuition. For AlkB, H66, H97, H172, and H197 were assigned epsilon protonation, while H72, H131, and H187 were assigned delta protonation. For ALKBH2, H55, H59, H106, H144, and H228 were assigned epsilon protonation, H199 and H220 were assigned delta protonation, and H167 was modelled as cationic. All crosslinks induced to facilitate crystallization between DNA and protein were removed for both complexes. The G169C, C67S, C165S, and C192S mutations to ALKBH2 were reverted to generate the wild-type enzyme, and the overhanging DNA ends (residues 259, 260, and 284) were truncated. To generate the AlkB–5mC or ALKBH2–5mC complexes, the *anti* and *syn* conformations of the 5mC nucleotide were overlaid onto the 3mC substrate in the ALKBH2- or AlkB-DNA complex using chemical intuition to minimize steric clashes between the bound substrate and active site amino acids.

Each complex was modelled using the AMBER parm14SB force field. Parameters assigned to the nonstandard 5mC and 3mC nucleotides were supplemented with GAFF parameters, and Restrained Electrostatic Potential (RESP) charges. For Fe(IV) and the

iron-ligating residues, the Metal Center Parameter Builder(31) was used to assign RESP charges, and bonding, angle, dihedral and non-bonding parameters based on B3LYP/LANL2DZ (Fe(IV)) and B3LYP/6-31G(d,p) (H, C, N, and O) optimized structures (Gaussian 09 revision D.01)(32) of the iron-binding site using the Seminario method.(33) The complexes were neutralized with Na⁺ counter ions, and solvated in a water box such that at least 10.0 Å of water exists between the DNA–protein complex and water box boundary.

All minimization, heating, equilibration, and production steps were performed with the GPU-accelerated PMEMD module available in AMBER 14 or AMBER 16.(34–36) For each step, the particle mesh Ewald method was used to employ a 10.0 Å electrostatic cutoff. To minimize each system, 1000 cycles of steepest descent (SD) and subsequent 1000 cycles of conjugate gradient (CG) minimization were first applied to the solvent and Na⁺ counter ions, followed by the complex hydrogen atoms, and finally the complex heavy/hydrogen atoms. For the final minimization step, the entire system was subjected to 1000 cycles of SD minimization followed by 2000 cycles of CG minimization. Subsequently, each system was heated to 310 K over 120 ps with a 25.0 kcal mol⁻¹ Å⁻² restraint placed on the solute using the Langevin thermostat with a time step of 1 fs under NVT conditions. The 25.0 kcal mol⁻¹ Å⁻² restraint was reduced to 0.0 kcal mol⁻¹ Å⁻² over 100 ps under constant temperature and pressure conditions (Berendsen barostat) using a 2 fs time step and SHAKE to constrain bonds involving hydrogen, which was followed by 1 ns of unrestrained equilibration. Several pre-production simulations were performed for each system to ensure adequate sampling and accommodation of each substrate within the active site of ALKBH2 or AlkB.

Representative structures (clustering methodology below) were chosen from these pre-production simulations to initiate the final 100 ns production simulations. Coordinates were saved from each simulation every 5 ps and analyzed over the same interval.

The CPPTRAJ(37) module was utilized for analysis of each trajectory. The occupancy of hydrogen bonds is reported for the duration of the simulation that the heavy atom distance is $< 3.4 \text{ \AA}$ and the hydrogen-bonding angle is $> 120^\circ$. To examine the placement of water within the active site, a three dimensional $20 \text{ \AA}^3$ grid centered on the bound substrate was generated. Dark red spheres are shown on representative structures and reflect an oxygen atom of a water molecule located in a $0.25 \text{ \AA}^3$ grid space for at least 40% of the simulation. The pre-production and production trajectories were clustered according to the positions of key active site residues (bound substrate, W69, H131, D135 and E136 for AlkB, and bound substrate, F124, H171, D174 and E175 for ALKBH2). A hierarchical agglomerative clustering methodology was used to obtain a $3.0 \text{ \AA}$ minimum cutoff between clusters or four clusters were obtained. While the representative structures are static snapshots, we report key structural parameters over the entire trajectory to ensure that conclusions are representative of each simulation trajectory. All distances are reported in $\text{\AA}$ and all angles are reported in degrees.

RESULTS AND DISCUSSION

A 16mer DNA oligonucleotide (5'-GAAGACCTXGGCGTCC-3', X = 5mC) containing 5mC in a CpG dinucleotide context was prepared through solid phase DNA synthesis with the phosphoramidite of 5mC.(23–26) High resolution electrospray ionization time-of-flight (ESI-TOF) MS analysis of the oligonucleotide exhibited an m/z of 1625.281 at its -3 charge state, which is in good agreement with the theoretical m/z 1625.281 expected of the product oligonucleotide (Figure 2a and Table S1). In the same -3 charge envelope, we also observed the ions of 5mC + Na⁺ (1632.608) and 5mC + K⁺ (1637.928). The observed m/z values of these species are consistent with the corresponding calculated m/z values (Table S1).

Previously, our lab has purified the three AlkB family enzymes mentioned above and tested their repair efficiency for 3mC, 1mA, 3mT, and 1mG in both ss- and ds-DNA.(23, 24) Similar procedures were adopted for the modification reactions of 5mC. The reaction conditions and the oligonucleotide-enzyme ratios are similar to those observed in the conversion of 5mC to its oxidative derivatives by the TET proteins reported in the literature.(6–8) For each enzyme, experiments were conducted in triplicate both in the presence and absence of the enzyme with all necessary cofactors at 37°C under both ds- and ss-DNA conditions, and the reaction products were analyzed by high resolution MS to ensure the differentiation of reaction products that have very similar m/z values.(24)

First, we tried to identify the new oxidative products that appeared after the enzymatic reactions. The three oxidative products, 5hmC, 5fC, and 5caC, all formed in the reactions with all three enzymes, but every enzyme had a preference to generate a

certain oxidative derivative. Below we use typical examples to demonstrate the formation of a certain product. The MS results of ALKBH2 oxidizing 5mC in ds-DNA (Figure 2b) showed a new oligonucleotide species that has an m/z for the monoisotopic peak at 1630.615 at -3 charge state, which corresponds well to the theoretical m/z value of the 16mer oligonucleotide containing 5hmC (1630.612 calculated, Table S1). In the reaction of 5mC with ALKBH3 in ss-DNA, another oligonucleotide envelope appeared at m/z 1629.951 (Figure 2c), which agrees with the 5fC base in the 16mer oligonucleotide (1629.940 calculated). For the oxidation of 5mC by the AlkB protein in ss-DNA, we observed the peak envelopes of 5hmC (1630.603), 5fC (1629.933) together with a new species with an m/z value of 1642.604, which is close to the 16mer oligonucleotide containing the sodium salt of 5caC (1642.600 calculated, Figure 2d and Table S1). Other oligonucleotide species containing metal ions, such as Na^+ and K^+ , were also observed. The complete assignments of the major species generated from MS analyses are summarized in Figure S1 and Table S1.

Figure 2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products. The observed m/z values represent the oligonucleotides under their -3 charge state. a) 5mC; b) 5mC (in ds-DNA) + ALKBH2; c) 5mC (in ss-DNA) + ALKBH3; and d) 5mC (in ss-DNA) + AlkB.

The product oligonucleotides were digested into single nucleosides, analyzed by LC-MS, and compared with standard nucleosides to confirm the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC, and 5caC (see the Product oligonucleotides analyses section in SI, Figure S2 to S10). Also, to make sure the oxidations were carried out by AlkB and its homologs, we generated the catalytically inactive protein variants of AlkB: H131A, D133A, and H187A (Figure S11). The three substituted amino acids in AlkB are the key residues that coordinate the Fe(II) ion.(13,

38) The sequences of wild type and variant proteins were confirmed by trypsin digestion with MS analyses (Figure S11 and Table S2). None of the AlkB variants showed any detectable oxidative product when reacted with 5mC (Figure S12); these observations suggest that the oxidative modifications were carried out by AlkB and not by a contaminating enzyme.

For the oxidation of 5mC, the formation of products 5hmC, 5fC, and 5caC had different distribution patterns for the three enzymes; and the three enzymes had different preferences for oxidation in ss- or ds-DNA reactions (Figure 3 and the Product distribution for the oxidation of 5mC section in SI). The overall activities of the three AlkB proteins on oxidizing 5mC are similar to the proficiencies of the TET enzymes on modifying 5mC reported in the literature(6–8) and are generally consistent with theoretical calculations that report higher barriers for each successive oxidation steps by TET2.(39) The conversion of 5mC to the corresponding oxidative intermediates by the AlkB proteins were comparable to their repair of other known substrates (Modifications on different alkyl substrates by the AlkB enzymes section in SI). In all of the enzymatic reactions, we only observed the oxidation of 5mC, but not the thymine DNA base, which naturally has a 5-methyl group. The same finding was reported for the TET family enzymes.(8)

Figure 3. Product distribution and strand preference (in univariate scatterplot) for the oxidation of 5mC by the AlkB family enzymes. Reaction products (5hmC, 5fC and 5caC) generated from the reactions of 5mC with the AlkB family enzymes in a) ds-DNA and b) ss-DNA. c) Total product percentage from reactions of the AlkB family enzymes oxidizing 5mC in ds- and ss-DNA.

To probe the molecular basis by which the AlkB enzymes are able to oxidize 5mC, we performed molecular dynamics (MD) simulations to examine how 5mC is accommodated in the active sites of ALKBH2 and AlkB (Figure 4). We found that the 5-methyl of *anti*-5mC in our model is far from the Fe(IV)–oxo moiety (~ 5.3 Å for ALKBH2 in Figure 4a and ~ 7.9 Å for AlkB in Figure 4d; also in Figures S13-S14, and Tables S3-S6). In contrast, the distances between the Fe(IV)–oxo moiety and the 5-methyl group in the models with *syn*-5mC bound to ALKBH2 or AlkB are much shorter (~ 3.8 Å for ALKBH2 in Figure 4b and ~ 3.6 Å for AlkB in Figure 4e) and similar to the distances in the structures of ALKBH2 or AlkB bound to their prototypic substrate 3mC in the *anti*-conformation (~ 3.3 Å for ALKBH2 in Figure 4c and ~ 3.3 Å for AlkB in Figure 4f.; See also the Simulations of ALKBH2 and AlkB bound to 3mC, 5hmC, and 5fC section in SI). With *syn*-5mC bound to ALKBH2 and AlkB, hydrogen bonds appear possible between the N^4 amino group of 5mC and active site residues. Specifically, *syn*-5mC interacts with Asp and Glu residues (D174/E175 for ALKBH2, and D135/E136 for AlkB) through water (Tables S3-S4 and Figures S13-S14), as well as the Y122 hydroxy group in ALKBH2 (Table S3). These interactions likely facilitate oxidative catalysis by positioning the C5 methyl group near the Fe(IV)–oxo moiety ($\sim 3.6 - 3.8$ Å; Tables S5-S6).

Interestingly, a crystal structure of TET2 co-crystallized with 5mC-containing DNA reveals *syn*-5mC in the active site,(40) and the reported χ torsion angle is consistent with that predicted for *syn*-5mC in ALKBH2/AlkB (Figure S15 and Tables S5-S6). Based on our combined experimental and theoretical data, we propose that the AlkB family enzymes are able to oxidize 5mC bound only in the *syn*-conformation.

Figure 4. Representative molecular dynamics structures of the ALKBH2 (a-c) or AlkB (d-f) complex bound to *anti*-5mC (a,d), *syn*-5mC (b,e), or *anti*-3mC (c,f). The distance between the oxo-moiety and methyl groups is highlighted with dashed lines.

To provide insight into the variable activities of ALKBH2 and AlkB for subsequent nucleobase oxidation (Figure 3), MD simulations were performed on *syn*-5hmC and 5fC bound in the active sites. For 5hmC, the C5 substituent is further from the Fe(IV)–oxo moiety for ALKBH2 (~4.5 Å, Figure S16 and Tables S5) compared to AlkB (3.4 Å, Figure S17 and Tables S6), which is consistent with the relative low abundance of the 5fC product for ALKBH2 (Figures 3, S19 to S21 and Tables S3 to S6). For 5fC, increased flexibility of the bound nucleobase may permit enhanced catalysis and generate more 5caC for ALKBH2 comparing to AlkB (Figures 3a, 3b, and S22). In addition to the insights provided by the MD simulations (see SI for detailed discussions of these calculations), several other factors could influence the product distributions including DNA binding, oxidative reactivity and different base flipping mechanisms used by each enzyme.

In this paper, we demonstrated the *in vitro* oxidative modification of 5mC to 5hmC, 5fC, and 5caC by the three AlkB DNA repair enzymes. Thus, the AlkB proteins are not only able to repair DNA adducts, such as 3mC and 3mT, but also can edit the epigenetic modification 5mC and generate the corresponding oxidative derivatives. These observations suggest a possible connection between DNA repair and epigenetic gene modulation. Future investigation includes analyzing the kinetic parameters of the three AlkB enzymes acting on 5mC, confirming the oxidation of 5mC by the AlkB enzymes in cell, and probing whether other α -KG/Fe(II)-dependent dioxygenases can oxidize 5mC.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

The authors want to thank the RI-INBRE program, its directors Prof. Zahir Shaikh and Prof. Bongsup Cho, and staff Dr. Al Bach, Kim Andrews, and Patricia Murray for their kind help. We also want to thank Dr. Ang Cai, Dr. Jeremy Setser, Ms. Kerri Bradshaw, and Mr. Michael Vittori for helpful contributions.

FUNDING

This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM103430. This work was also supported by National Institutes of Health under grant numbers R15 CA213042 and R01 ES028865 (to D.L.). This work was also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, Discovery 2016-04568, S.D.W., and CGS-D/MSFSS, S.A.P.L.), the Canada Foundation for Innovation (22770, S.D.W.), and Alberta Innovates–Technology Futures (AI-TF, S.A.P.L.). This work was also supported by National Cancer Institute or National Institute of Environmental Health Sciences grants P01 CA26731, R01 CA080024, and P30 ES002109 (to J.M.E.). C.L.D. is an HHMI investigator.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
2. Smith,Z.D. and Meissner,A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
3. Li,E. and Zhang,Y. (2014) DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, a019133.
4. Law,J.A. and Jacobsen,S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
5. Hausinger,R.P. and Schofield,C.J. eds. (2015) 2-Oxoglutarate-dependent oxygenases Royal Society of Chemistry, Cambridge, UK.
6. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L., *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
7. Ito,S., D'Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
8. Zhang,L., Chen,W., Iyer,L.M., Hu,J., Wang,G., Fu,Y., Yu,M., Dai,Q., Aravind,L. and He,C. (2014) A TET homologue protein from *Coprinopsis cinerea* (CcTET) that biochemically converts 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.*, **136**, 4801–4804.

9. Klungland,A. and Robertson,A.B. (2017) Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radic. Biol. Med.*, **107**, 62–68.
10. Cimmino,L. and Aifantis,I. (2017) Alternative roles for oxidized mCs and TETs. *Curr. Opin. Genet. Dev.*, **42**, 1–7.
11. Hashimoto,H., Zhang,X., Vertino,P.M. and Cheng,X. (2015) The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations. *J. Biol. Chem.*, **290**, 20723–20733.
12. Aravind,L. and Koonin,E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.*, **2**, RESEARCH0007.
13. Yu,B., Edstrom,W.C., Benach,J., Hamuro,Y., Weber,P.C., Gibney,B.R. and Hunt,J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879–884.
14. Sedgwick,B., Bates,P.A., Paik,J., Jacobs,S.C. and Lindahl,T. (2007) Repair of alkylated DNA: recent advances. *DNA Repair*, **6**, 429–442.
15. Trewick,S.C., Henshaw,T.F., Hausinger,R.P., Lindahl,T. and Sedgwick,B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature*, **419**, 174–178.
16. Falnes,P.Ø., Johansen,R.F. and Seeberg,E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature*, **419**, 178–182.
17. Zheng,G., Fu,Y. and He,C. (2014) Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.*, **114**, 4602–4620.

18. Fedeles,B.I., Singh,V., Delaney,J.C., Li,D. and Essigmann,J.M. (2015) The AlkB family of Fe(II)/ α -ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.*, **290**, 20734–20742.
19. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 14051–14056.
20. Li,D., Fedeles,B.I., Shrivastav,N., Delaney,J.C., Yang,X., Wong,C., Drennan,C.L. and Essigmann,J.M. (2013) Removal of N-alkyl modifications from N(2)-alkylguanine and N(4)-alkylcytosine in DNA by the adaptive response protein AlkB. *Chem. Res. Toxicol.*, **26**, 1182–1187.
21. Yi,C., Chen,B., Qi,B., Zhang,W., Jia,G., Zhang,L., Li,C.J., Dinner,A.R., Yang,C.-G. and He,C. (2012) Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.*, **19**, 671–676.
22. Yi,C., Jia,G., Hou,G., Dai,Q., Zhang,W., Zheng,G., Jian,X., Yang,C.-G., Cui,Q. and He,C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330–333.
23. Chen,F., Tang,Q., Bian,K., Humulock,Z.T., Yang,X., Jost,M., Drennan,C.L., Essigmann,J.M. and Li,D. (2016) Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.*, **29**, 687–693.
24. Chen,F., Bian,K., Tang,Q., Fedeles,B.I., Singh,V., Humulock,Z.T., Essigmann,J.M. and Li,D. (2017) Oncometabolites d- and l-2-Hydroxyglutarate Inhibit the AlkB Family

- DNA Repair Enzymes under Physiological Conditions. *Chem. Res. Toxicol.*, **30**, 1102–1110.
25. Bian,K., Chen,F., Humulock,Z.T., Tang,Q. and Li,D. (2017) Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson’s Disease Condition. *Chem. Res. Toxicol.*, **30**, 1794–1796.
26. Tang,Q., Cai,A., Bian,K., Chen,F., Delaney,J.C., Adusumalli,S., Bach,A.C., Akhlaghi,F., Cho,B.P. and Li,D. (2017) Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega*, **2**, 8205–8212.
27. Quinlivan,E.P. and Gregory,J.F. (2008) DNA digestion to deoxyribonucleoside: a simplified one-step procedure. *Anal. Biochem.*, **373**, 383–385.
28. Frick,L.E., Delaney,J.C., Wong,C., Drennan,C.L. and Essigmann,J.M. (2007) Alleviation of 1,N6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 755–760.
29. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo,R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; and Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. (2016) AmberTools16 University of California: San Francisco.

30. Anandakrishnan,R., Aguilar,B. and Onufriev,A.V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, **40**, W537-541.
31. Li,P. and Merz,K.M. (2016) MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.*, **56**, 599–604.
32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.;, Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.;, Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.;, Throssell, K.; Montgomery, J. A. Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, and J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. (2016) Gaussian 09, Revision D.01 Gaussian, Inc.: Wallingford CT,.
33. Seminario,J.M. (1996) Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.*, **60**, 1271–1277.
34. Salomon-Ferrer,R., Götz,A.W., Poole,D., Le Grand,S. and Walker,R.C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.

35. Götz,A.W., Williamson,M.J., Xu,D., Poole,D., Le Grand,S. and Walker,R.C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.*, **8**, 1542–1555.
36. Le Grand,S., Götz,A.W. and Walker,R.C. (2013) SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.*, **184**, 374–380.
37. Roe,D.R. and Cheatham,T.E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
38. Yu,B. and Hunt,J.F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 14315–14320.
39. Lu,J., Hu,L., Cheng,J., Fang,D., Wang,C., Yu,K., Jiang,H., Cui,Q., Xu,Y. and Luo,C. (2016) A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys. Chem. Chem. Phys. PCCP*, **18**, 4728–4738.
40. Hu,L., Li,Z., Cheng,J., Rao,Q., Gong,W., Liu,M., Shi,Y.G., Zhu,J., Wang,P. and Xu,Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555.

TABLE AND FIGURES LEGENDS

Figure 1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC.

Figure 2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products. The observed m/z values represent the oligonucleotides under their -3 charge state. a) 5mC; b) 5mC (in ds-DNA) + ALKBH2; c) 5mC (in ss-DNA) + ALKBH3; and d) 5mC (in ss-DNA) + AlkB.

Figure 3. Product distribution and strand preference (in univariate scatterplot) for the oxidation of 5mC by the AlkB family enzymes. Reaction products (5hmC, 5fC and 5caC) generated from the reactions of 5mC with the AlkB family enzymes in a) ds-DNA and b) ss-DNA. c) Total product percentage from reactions of the AlkB family enzymes oxidizing 5mC in ds- and ss-DNA.

Figure 4. Representative molecular dynamics structures of the ALKBH2 (a-c) or AlkB (d-f) complex bound to *anti*-5mC (a,d), *syn*-5mC (b,e), or *anti*-3mC (c,f). The distance between the oxo-moiety and methyl groups is highlighted with dashed lines.

SUPPORTING INFORMATION

DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

Ke Bian,^{1,±} Stefan A.P. Lenz,^{2,±} Qi Tang,¹ Fangyi Chen,¹ Rui Qi,¹ Marco Jost,³ Catherine L. Drennan,^{3,4,6,7} John M. Essigmann,^{3,5,6} Stacey D. Wetmore,^{2,*} and Deyu Li^{1,*}

¹Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States

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

³Department of Chemistry, ⁴Department of Biology, ⁵Department of Biological Engineering, ⁶Center for Environmental Health Sciences, and ⁷Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[±]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

^{*}To whom correspondence should be addressed. Deyu Li, Tel: +1(401)874-9361; Fax: +1(401)874-9361; Email: deyuli@uri.edu. Correspondence may also be addressed to Stacey D. Wetmore, Tel: +1(403) 329-2323; fax: +1(403) 329-2057; Email: stacey.wetmore@uleth.ca.

Present Address: [Fangyi Chen], Xiamen University, 4221 Xiang An South Road, Xiang An District, Xiamen, Fujian 361102, P.R. China. [Marco Jost], Department of

Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94158, United States.

TABLE OF CONTENTS

Product oligonucleotides analyses

Product distribution for the oxidation of 5mC

Modifications on different alkyl substrates by the AlkB enzymes

Simulations of ALKBH2 and AlkB bound to 3mC, 5mC, 5hmC, and 5fC

Figure S1. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures.

Figure S2. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides.

Figure S3. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode.

Figure S4. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode.

Figure S5. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode.

Figure S6. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode.

Figure S7. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode.

Figure S8. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode.

Figure S9. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode.

Figure S10. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode.

Figure S11. MS analyses of trypsin digested proteins including the wild type and variant AlkB enzymes.

Figure S12. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and variant proteins.

Figure S13. Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to 5mC.

Figure S14. Distribution of active site water (red spheres) during MD simulations of AlkB bound to 5mC.

Figure S15. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB–5mC and ALKBH2–5mC complexes onto a crystal structure of TET2 bound to 5mC-containing DNA.

Figure S16. Distribution of active site water during MD simulations of ALKBH2 bound to *anti*-3mC and overlay of MD representative structures of ALKBH2 bound to *syn*-5mC and *anti*-3mC.

Figure S17. Distribution of active site water during MD simulations of AlkB bound to *anti*-3mC and overlay of MD representative structures of AlkB bound to *syn*-5mC and *anti*-3mC.

Figure S18. Structures and chemical numbering of 3mC and 5mC.

Figure S19. Structures and chemical numbering of *anti*-3mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC.

Figure S20. Distribution of active site water during MD simulations of ALKBH2 bound to *syn*-5hmC, ALKBH2 bound to *syn*-5fC, AlkB bound to *syn*-5hmC, or AlkB bound to *syn*-5fC.

Figure S21. Simulations of ALKBH2 or AlkB bound to 5hmC.

Figure S22. *syn*-5fC Bound by ALKBH2 or AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety.

Figure S23. Different alkyl DNA modifications oxidized by the three AlkB family enzymes at various time points.

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides.

Table S2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.

Table S3. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

Table S4. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

Table S5. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

Table S6. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

Table S7. Time to half maximum conversion ($T_{1/2}$, min) for the oxidation of different alkyl DNA modifications by the three AlkB family enzymes.

Product oligonucleotides analyses

It was important to assure that the observed new oligonucleotide species indeed contain the proposed oxidative products. The product oligonucleotides were digested with benzonase into single nucleosides, which were analyzed by LC-MS and compared with standard nucleosides for retention time and molecular weight. In the digestion of the product generated from the AlkB reaction, a nucleoside was eluted out at 7.3 min in HPLC (Figure S2) and had an observed m/z 240.2386 in the negative mode of MS analysis (theoretical m/z 240.2392, Figure S4); these data agree well with the analysis of the standard sample of 5mC nucleoside (7.3 min in HPLC and 240.2385 in m/z , Figure S3). Another nucleoside from digestion had a retention time at 4.9 min in HPLC (Figure S2) and had an m/z 256.2387 in the MS analysis (Figure S6); these data are comparable to the standard sample of 5hmC nucleoside (5.0 min in HPLC and 256.2409 in m/z , theoretical m/z 256.2382, Figures S2 and S5). Similarly, nucleosides 5fC and 5caC were also discovered in the mixture of the digested reaction product and their identities were confirmed by comparing to standard nucleosides (Figures S2, S7 to S10). These observations support that the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC, and 5caC.

Product distribution for the oxidation of 5mC

For the oxidation of 5mC, the formation of products 5hmC, 5fC, and 5caC had different distribution patterns for the three enzymes; and the three enzymes had different preferences of oxidation in ss- or ds-DNA reactions (Figure 3). The relative amount of each product in the final reaction mixture was quantified according to the abundance of

its corresponding ion at -3 charge state in the MS analysis. The product oligonucleotides containing 5hmC was the dominant species for the reactions of ALKBH2 and 3 in both ds- and ss-DNA (Figures 3a and 3b). For example, 5hmC represents 78.7% of the total amount of the three oxidative products in the reaction of ALKBH2 with 5mC in ds-DNA; 5fC and 5caC are quantified as 10.9% and 10.4% correspondingly (Figure 3a). For the oxidation of 5mC by AlkB, 5hmC is the major species (51.2%) in ds-DNA reaction; however, 5fC is the most abundant species (65.3%) in ss-DNA reaction (Figures 3a and 3b). We were also able to quantify the yield of the oxidative products by comparing them to the total oligonucleotide species. The ratios of the oxidative products show ALKBH2 prefers to oxidize 5mC in ds-DNA (8.7%) over in ss-DNA (5.5%), and ALKBH3 prefers to modify 5mC in ss-DNA (22.1%) over in ds-DNA (4.2%) (Figure 3c). These results support the strand preference in repair reactions of ALKBH2 (preferring ds-DNA) and ALKBH3 (preferring ss-DNA) reported in the literature.⁽¹⁾ For AlkB, the enzyme oxidized 5mC in ss-DNA (23.5%) more efficiently than in ds-DNA (13.3%) (Figure 3c); these results are similar to the preference of AlkB repairing 3mC in ss-DNA.⁽²⁾

Modifications on different alkyl substrates by the AlkB enzymes

We compared the conversion efficiency of 5mC by the three AlkB family enzymes to other known methyl substrates of the AlkB proteins. Those substrates include 3mC, 1mA, 3mT, and 1mG; and all of them have been demonstrated to be repaired by AlkB both *in vitro* and *in vivo*.^(1, 3) The results show the conversion of 5mC to the corresponding oxidative products are comparable to the demethylation of 3mT and

1mG; and the reactions of these three methyl modifications are slower than 3mC and 1mA (Figure S23 and Table S7). For example, $T_{1/2}$ of the reactions for 5mC (16.9 min) is slightly shorter than 3mT (17.4 min) but longer than 1mG (1.7 min) in the AlkB reactions. For the ALKBH2 reactions, $T_{1/2}$ for 5mC are comparable to 3mT and 1mG (all between 16.4 to 18.7 min). For the ALKBH3 protein, 5mC is oxidized faster than 3mT and 1mG with about 5 minutes shorter in $T_{1/2}$ than the other two (Figure S23 and Table S7). Those biochemical results may indicate similar modification could happen in cell: the modification of 5mC is comparable to the weaker AlkB substrates 3mT and 1mG but less efficiently than the stronger substrates 3mC and 1mA.(3)

Simulations of AlkBH2 and AlkB bound to 3mC, 5mC, 5hmC, and 5fC

To validate that the simulated structures of the ALKBH2 or AlkB complexes are consistent with the observed catalytic activity, we examined the active site conformations adopted upon binding of both enzymes to *anti*-3mC-, *syn*-5mC-, *syn*-5hmC-, or *syn*-5fC-containing DNA. Similar DNA–protein interactions form when 3mC and 5mC are bound. Specifically, the 3mC N⁴ amino group forms direct hydrogen bonds with the D174 or E175 sidechains when bound to ALKBH2 (Figure S16b and Table S3), and direct or water-mediated hydrogen bonds with D135 and E136 when bound to AlkB (Table S4 and Figure S17). As a result, the N3 methyl group of 3mC occupies an equivalent active site position as the C5 methyl group of 5mC for both enzymes (Figures S16b and S17b), with a distance between the Fe(IV)–oxo and the 3mC methyl moiety of ~ 3.3 Å for both enzymes (Tables S5–S6), and the *syn*-5mC moiety of ~3.6 Å for AlkB and 3.8 Å for ALKBH2. Thus, the simulation data is

consistent with the proposal that the AlkB family of enzymes is able to oxidize 5mC in the *syn* orientation.

Similar to ALKBH2 or AlkB bound to 5mC, *syn*-5hmC is stabilized by direct or water-mediated hydrogen bonds between the N⁴ amino group of the nucleobase and the carboxylate moieties of D174 and E175 (D135 and E136 in AlkB; Tables S3-S4 and Figure S20). For 5hmC bound by ALKBH2 or AlkB, hydrogen bonds are formed between the C5 substituent hydroxy group and the Fe(IV)-oxo moiety (occupancy = 34% for ALKBH2 and 52% for AlkB; Tables S3-S4 and Figure S21). In ALKBH2, an additional hydrogen bond is formed between the C5 substituent hydroxy group and D173 (59%, Table S3), which results in one orientation of the C5 substituent throughout the simulation (Figure S21a). In contrast, the C5 substituent hydroxy group does not interact with D133 in AlkB, which results in two conformations of 5hmC within the active site (Figure S21b-d). More importantly, the C5 substituent is at an optimum distance from the Fe(IV)-oxo moiety in AlkB (~3.4 Å), which matches the prototypic substrate (3mC; ~3.3 Å), while the equivalent distance is longer in the ALKBH2–5hmC complex (~4.5 Å; Tables S5-S6 and Figure 3), which would impede catalysis. This helps explain the observed higher abundance of 5fC for AlkB compared to ALKBH2 catalyzed oxidation on 5hmC, although several other factors could also be significant such as DNA binding and unique base flipping mechanism for each enzyme.

As discussed for 5mC and 5hmC, the *syn*-conformation of 5fC is stabilized by hydrogen bonds between the N⁴ amino group of the nucleobase and carboxylate sidechains of active site residues (Tables S3-S4 and Figure S20). When bound to either ALKBH2 or AlkB, 5fC is planar due to an intramolecular hydrogen bond between the

N⁴ amino group and the carbonyl of the C5 substituent. A hydrogen bond also forms between the carbonyl of the C5 substituent and an active site arginine, which is notably more persistent for AlkB (R210, occupancy = 100.0%) than ALKBH2 (R254, occupancy = 29.2%, Tables S3-S4 and Figure S20). As a result, the distance between the hydrogen atom of the C5 substituent and the Fe(IV)-oxo group is longer for AlkB (~3.8 Å) compared to ALKBH2 (3.3 Å; Tables S5-S6 and Figure S22). Although the difference in distance for 5fC bound by ALKBH2 and AlkB is not as significant as observed for 5hmC, QM/MM studies on TET2-catalyzed oxidation of 5mC, 5hmC, and 5fC reveal that the initial hydrogen atom abstraction step is rate limiting, with the barrier increasing as 5mC < 5hmC < 5fC.⁽⁴⁾ This suggests that the position of the substituent relative to the Fe(IV)-oxo group is even more crucial in the case of 5fC. Thus, our predicted structures correlate with the lower abundance of 5caC relative to 5fC for AlkB compared to ALKBH2 (Figure 3). Nevertheless, as discussed for 5hmC, nucleotide recognition and the base flipping mechanism could also play different roles in oxidative conversion of 5fC to 5caC.

Figure S1. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table S1.

Figure S2. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides.

Figure S3. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S4. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S5. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S6. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S7. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode.

The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S8. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S9. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

Figure S10. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

AlkB sequence		
MLDLFADAEP	WQEPLAAGAV	ILRRFAFNAA
MTNCGHLGWT	THRQGYLYSP	IDPQTNKPWP
RYAPGAKLSL	HQDKDEPDLR	APIVSVSLGL
ESRLFYHGIQ	PLKAGFHPLT	IDCRYNLTFR
		QAGKKE
Peptide fragments from trypsin digestion		
Mass	Position	Peptide sequence
3415.6230	74-102	QGYLYSPIDPQTNKPWPAMPQSFHNLQCR
2835.2713	48-73	QMVTPGGYTMSVAMTNCGHLGWTTHR
2525.2988	1-23	MLDLFADAEPWQEPLAAGAV ILR
2013.1662	141-160	APIVSVSLGLPAIFQFGGLK
1992.9363	103-121	AATAAGYPDFQPDACLINR
1764.9158	168-183	LLLEHGDVVVWGGESR
1347.6419	36-47	DINDVASQSPFR
1278.6720	25-35	FAFNAAEQLIR
1228.6022	194-204	AGFHPLTIDCR
1214.6811	184-193	LFYHGIQPLK
839.4501	128-134	LSLHQDK
812.4181	205-210	YNLTFR
743.3450	135-140	DEPDLR
605.3173	122-127	YAPGAK
585.3122	162-166	NDPLK

Figure S11. MS analyses of trypsin digested proteins including the wild type and variant AlkB enzymes.

Figure S12. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and variant proteins. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table S1.

Figure S13. Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to a) anti-5mC or b) syn-5mC. c) ALKBH2 bound to syn-5mC highlighting key hydrogen-bonding interactions. d) Overlay of MD representative structures of ALKBH2 bound to anti- (blue) or syn- (gray) 5mC.

Figure S14. Distribution of active site water (red spheres) during MD simulations of AlkB bound to a) anti-5mC or b) syn-5mC. c) AlkB bound to syn-5mC highlighting key water mediated hydrogen-bonding interactions. d) Overlay of MD representative structures of AlkB bound to anti- (cyan) or syn- (blue) 5mC.

Figure S15. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB-5mC (blue) and ALKBH2-5mC (gray) complexes onto a crystal structure of TET2 bound to 5mC-containing DNA, Fe(II), and α -KG (teal; PDB ID: 4NM6),(5) highlighting the similarity in the 5mC glycosidic bond conformation (χ).

Figure S16. Distribution of active site water during MD simulations of ALKBH2 bound to anti-3mC and overlay of MD representative structures of ALKBH2 bound to syn-5mC and anti-3mC. a) Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to anti-3mC. b) ALKBH2 bound to anti-3mC, highlighting key hydrogen-bonding interactions. c) Overlay of MD representative structures of ALKBH2 bound to syn-5mC (gray) and anti-3mC (blue).

Figure S17. Distribution of active site water during MD simulations of AlkB bound to anti-3mC and overlay of MD representative structures of AlkB bound to syn-5mC and anti-3mC. a) Distribution of active site water (red spheres) during MD simulations of AlkB bound to anti-3mC. b) AlkB bound to anti-3mC, highlighting key water mediated hydrogen-bonding interactions. c) Overlay of MD representative structures of AlkB bound to syn-5mC (blue) and anti-3mC (cyan).

a)

b)

Figure S18. Structures and chemical numbering of a) 3mC and b) 5mC. The glycosidic bond orientation is defined by dihedral angle χ (blue, $\angle O4'C1'N1C2$) as either anti (left, $180 \pm 90^\circ$) or syn (right, $0 \pm 90^\circ$).

Figure S19. Structures and chemical numbering of a) anti-3mC, b) syn-5mC, c) syn-5hmC, and d) syn-5fC. The glycosidic bond orientation is defined by the χ dihedral angle (blue, $\angle(O4'C1'N1C2)$) as anti ($180 \pm 90^\circ$) or syn ($0 \pm 90^\circ$).

Figure S20. Distribution of active site water (red spheres) during MD simulations of a) ALKBH2 bound to syn-5hmC, b) ALKBH2 bound to syn-5fC, c) AlkB bound to syn-5hmC, or d) AlkB bound to syn-5fC. Key hydrogen-bonding interactions are highlighted with dashed lines.

Figure S21. Simulations of ALKBH2 or AlkB bound to 5hmC. a) ALKBH2 bound to syn-5hmC, and AlkB bound to syn-5hmC with b) θ ($\angle(\text{C4C5C7O7})$) $\sim -132^\circ$ and c) $\theta \sim 50^\circ$, highlighting different orientations of the C5 substituent. d) Histogram displaying the occupancy of θ (degree) during simulations of ALKBH2 or AlkB bound to 5hmC.

Figure S22. syn-5fC Bound by a) ALKBH2 or b) AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety.

Figure S23. Different alkyl DNA modifications oxidized by the three AlkB family enzymes at various time points. Top: AlkB; middle: ALKBH2; and bottom: ALKBH3.

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. The sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of 5mC or oxidized bases.

modification	MW (calculated) of neutral species	m/z (calculated) -3 charge peak	m/z (observed) -3 charge peak			
			5mC control	ALKBH2	ALKBH3	AlkB
5mC	4878.866	1625.281	1625.281	1625.279	1625.285	1625.276
5mC + Na⁺	4900.848	1632.608	1632.608	1632.616	1632.613	1632.608
5mC + K⁺	4916.822	1637.933	1637.928	1637.941	1637.942	1637.931
5mC + 2Na⁺	4922.830	1639.935	1639.941	1639.937	1639.938	1639.931
5mC + Na⁺ + K⁺	4938.803	1645.260	1645.267	1645.273	1645.261	1645.263
5hmC	4894.861	1630.612	-	1630.615	1630.607	1630.603
5hmC + Na⁺	4916.843	1637.940	-	1637.941	1637.942	1637.931
5hmC + K⁺	4932.816	1643.264	-	1643.265	1643.259	1643.262
5fC	4892.845	1629.940	-	1629.936	1629.951	1629.933
5fC + Na⁺	4914.827	1637.268	-	1637.288	1637.285	1637.257
5fC + 2Na⁺	4936.809	1644.595	-	1644.605	1644.588	1644.598
5caC + Na⁺	4930.814	1642.600	-	1642.616	1642.611	1642.604

Table S2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.

peptide sequence	MW	m/z	m/z (observed) +1 charge peak			
	(calculated)	(calculated)	AlkB	AlkB	AlkB	AlkB
	of neutral species	+1 charge peak	wt	D133A	H131A	H187A
LSLHQDK	839.450	840.457	840.470	-	-	840.470
LSLHQAK	795.460	796.468	-	796.480	-	-
LSLAQDK	773.428	774.436	-	-	774.449	-
LFYHGIQPLK	1214.681	1215.688	1215.706	1215.706	1215.706	-
LFYAGIQPLK	1148.659	1149.667	-	-	-	1149.681

Table S3. Summary of important hydrogen bonds formed during MD simulations of anti-3mC, anti-5mC, syn-5mC, syn-5hmC, and syn-5fC in the ALKBH2 complex.

	<i>syn-5fC</i>			<i>syn-5hmC</i>			<i>syn-5mC</i>			<i>anti-5mC</i>			<i>3mC</i>		
	Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}	
E175(Oδ)···mC(N4H)	2.9 (160)	100.00%		2.9 (151)	83.20%		3.0 (145)	0.30%		N/A	0.00%		2.8 (159)	99.10%	
D174(Oδ)···mC(N4H)	N/A	0.00%		2.9 (164)	32.00%		N/A	0.00%		N/A	0.00%		2.8 (166)	94.30%	
Y122(O_H)···mC(N4H)	3.1 (136)	26.20%		3.1 (141)	5.00%		3.0 (146)	90.00%		3.1 (145)	15.90%		N/A	0.00%	
H₂O(O)···mC(N4H)	3.2 (126)	0.00%		3.1 (151)	>100.0%		3.1 (137)	72.20%		3.1 (146)	78.90%		3.1 (138)	55.30%	
mC(O2)···H₂O(OH)	2.8 (151)	56.10%		2.8 (150)	>100.0%		2.9 (154)	94.30%		3.1 (143)	4.80%		3.2 (128)	0.10%	
mC(N3)···H₂O(OH)	3.2 (138)	0.90%		3.0 (146)	30.00%		3.1 (147)	62.50%		3.0 (154)	57.60%		N/A	0.00%	
E175(Oδ)···Y122(OH)	2.9 (154)	41.90%		2.9 (163)	100.00%		2.8 (164)	78.70%		2.8 (163)	0.20%		2.7 (165)	99.00%	
H₂O(O)···Y122(OH)	N/A	0.00%		3.1 (129)	10.20%		3.2 (129)	17.70%		3.0 (151)	>100%		N/A	0.00%	
E175(Oε)···H₂O(OH)	2.7 (159)	>100.0%		2.7 (162)	>100.0%		2.7 (160)	>100%		2.8 (159)	>100%		N/A	0.00%	
D174(Oδ)···H₂O(OH)	2.8 (152)	>100.0%		2.8 (157)	>100.0%		2.8 (160)	>100%		2.8 (157)	>100%		2.8 (158)	>100%	
mC(O5)···R254(NηH)	3.1 (132)	29.20%		3.1 (140)	37.00%		N/A	N/A		N/A	N/A		N/A	N/A	
Fe(IV)-oxo(O)···mC(O5H)	N/A	N/A		2.9 (144)	33.90%		N/A	N/A		N/A	N/A		N/A	N/A	
D173(Oδ)···mC(O5H)	N/A	N/A		2.7 (152)	58.50%		N/A	N/A		N/A	N/A		N/A	N/A	

^aPercent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. ^bGreater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site. ^cHydrogen bonding distance in Å. ^dAngle in ° (in parentheses).

Table S4. Summary of important hydrogen bonds formed during MD simulations of anti-3mC, anti-5mC, syn-5mC, syn-5hmC, and syn-5fC in the AlkB complex.

	<i>syn-5fC</i>			<i>syn-5hmC</i>			<i>syn-5mC</i>			<i>anti-5mC</i>			<i>3mC</i>		
	Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}		Dist (Å) ^{c,d}	% ^{a,b}	
D133(Oδ)···mC(N4H)	3.2 (132)	0.20%		3.2 (132)	1.60%		3.1 (143)	11.80%		3.2 (135)	0.60%		2.9 (147)	42.60%	
H₂O(O)···mC(N4H)	3.1 (137)	>100.0%		3.1 (138)	>100.0%		3.1 (145)	>100%		3.1 (142)	>100%		3.1 (150)	>100%	
mC(O2)···H₂O(OH)	3.0 (143)	57.40%		3.0 (141)	72.90%		2.9 (152)	>100%		2.9 (154)	>100%		3.0 (150)	32.90%	
mC(N3)···H₂O(OH)	2.9 (155)	>100.0%		3.0 (156)	>100.0%		3.0 (152)	95.50%		2.9 (150)	91.00%		N/A	0.00%	
E136(Oδ)···Y78(OH)	2.7 (164)	96.30%		2.7 (165)	65.70%		2.7 (162)	36.00%		2.7 (165)	58.50%		2.6 (163)	62.30%	
H₂O(O)···Y78(OH)	3.1 (141)	43.70%		3.1 (135)	42.70%		3.1 (145)	5.60%		3.1 (137)	25.80%		3.1 (150)	88.90%	
E136(Oε)···H₂O(OH)	2.7 (162)	>100.0%		2.8 (158)	97.80%		2.9 (152)	>100%		2.7 (160)	>100%		2.8 (158)	>100%	
D135(Oδ)···H₂O(OH)	2.8 (157)	>100.0%		2.8 (157)	>100.0%		2.8 (157)	>100%		2.8 (158)	>100%		2.8 (158)	>100%	
mC(O5)···R210(NηH)	2.8 (155)	100.00%		2.9 (160)	35.10%		N/A	N/A		N/A	N/A		N/A	N/A	
Fe(IV)-oxo(O)···mC(O5H)	N/A	N/A		2.8 (152)	51.70%		N/A	N/A		N/A	N/A		N/A	N/A	
D133(Oδ)···mC(O5H)	N/A	N/A		N/A	0.00%		N/A	N/A		N/A	N/A		N/A	N/A	

^aPercent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. ^bGreater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site. ^cHydrogen-bonding distance in Å. ^dAngle in ° (in parentheses).

Table S5. Summary of important distances and dihedral angles adopted during MD simulations of anti-3mC, anti-5mC, syn-5mC, syn-5hmC, and syn-5fC in the ALKBH2 complex.

	3mC	<i>anti</i>-5mC	<i>syn</i>-5mC	<i>syn</i>-5hmC	<i>syn</i>-5fC
	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Fe(IV)–oxo$\cdots$C(C7)^a	3.3 (0.2)	5.3 (0.5)	3.8 (0.3)	4.5 (0.8)	3.2 (0.4)
Fe(IV)–oxo$\cdots$C(H7¹)^a	3.2 (0.7)	5.5 (0.8)	3.6 (0.6)	4.7 (1.1)	3.3 (0.4)
Fe(IV)–oxo$\cdots$C(H7²)^a	3.2 (0.6)	5.5 (0.8)	3.6 (0.6)	6.9 (0.6)	N/A
Fe(IV)–oxo$\cdots$C(H7³)^a	3.2 (0.7)	5.5 (0.8)	3.6 (0.6)	3.2 (1.4)	N/A
C(χ)^b	–147 (15)	–143 (19)	49 (11)	–9 (17)	–19 (19)

^aModification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). ^b χ of bound pyrimidine defined as $\angle(\text{O4}'\text{C1}'\text{N1C2})$ (Figures S18 and S19).

Table S6. Summary of important distances and dihedral angles adopted during MD simulations of anti-3mC, anti-5mC, syn-5mC, syn-5hmC, and syn-5fC in the AlkB complex.

	3mC	<i>anti</i>-5mC	<i>syn</i>-5mC	<i>syn</i>-5hmC	<i>syn</i>-5fC
	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Fe(IV)–oxo$\cdots$C(C7)^a	3.3 (0.2)	7.9 (2.0)	3.6 (0.4)	3.4 (0.2)	3.5 (0.4)
Fe(IV)–oxo$\cdots$C(H7¹)^a	3.5 (0.7)	8.2 (2.0)	3.6 (0.7)	3.3 (0.6)	3.8 (0.5)
Fe(IV)–oxo$\cdots$C(H7²)^a	3.4 (0.7)	8.2 (2.0)	3.5 (0.7)	5.8 (0.4)	N/A
Fe(IV)–oxo$\cdots$C(H7³)^a	3.4 (0.7)	8.2 (2.0)	3.6 (0.7)	3.3 (1.5)	N/A
C(χ)^b	–128 (11)	–116 (13)	32 (17)	44 (18)	51 (10)

^aModification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). ^b χ of bound pyrimidine defined as $\angle(\text{O4}'\text{C1}'\text{N1C2})$ (Figures S18 and S19).

Table S7. Time to half maximum conversion (T1/2, min) for the oxidation of different alkyl DNA modifications by the three AlkB family enzymes.

	1mA	3mC	5mC	3mT	1mG
AlkB	< 1.0	< 1.0	16.9	17.4	1.7
ALKBH2	< 1.0	< 1.0	17.2	18.7	16.4
ALKBH3	< 1.0	< 1.0	25.2	30.0	30.7

REFERENCES

1. Chen,F., Tang,Q., Bian,K., Humulock,Z.T., Yang,X., Jost,M., Drennan,C.L., Essigmann,J.M. and Li,D. (2016) Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.*, **29**, 687–693.
2. Fedeles,B.I., Singh,V., Delaney,J.C., Li,D. and Essigmann,J.M. (2015) The AlkB family of Fe(II)/ α -ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.*, **290**, 20734–20742.
3. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 14051–14056.
4. Lu,J., Hu,L., Cheng,J., Fang,D., Wang,C., Yu,K., Jiang,H., Cui,Q., Xu,Y. and Luo,C. (2016) A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys. Chem. Chem. Phys. PCCP*, **18**, 4728–4738.
5. Hu,L., Li,Z., Cheng,J., Rao,Q., Gong,W., Liu,M., Shi,Y.G., Zhu,J., Wang,P. and Xu,Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555.

Prepared to submit to *Molecules*, March 2019

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

Ang Cai[±], Ke Bian[±], Fangyi Chen, Qi Tang, Rachel Carley, Deyu Li^{*}, and Bongsup Cho^{*}

[±]A.C. and K.B. contributed equally to this work.

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

^{*}Corresponding Authors: Bongsup Cho

Department of Biomedical and Pharmaceutical Sciences

University of Rhode Island

495E Avedisian Hall, 7 Greenhouse Rd

Kingston, Rhode Island 02881, United States

Phone: +1 (401) 874-5024

Email address: bcho@uri.edu

Deyu Li

Department of Biomedical and Pharmaceutical Sciences

University of Rhode Island

495N Avedisian Hall, 7 Greenhouse Rd

Kingston, Rhode Island 02881, United States

Phone: +1 (401) 874-9361

Email address: deyuli@uri.edu

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

Ang Cai, Ke Bian, Fangyi Chen, Qi Tang, Rachel Carley, Deyu Li,* and Bongsup Cho*

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

*Co-correspondence: bcho@uri.edu; Tel: +1 401-874-5024; deyuli@uri.edu; Tel: +1 401-874-9361

ABSTRACT

Bulky organic carcinogens are activated *in vivo* and subsequently react with nucleobases of cellular DNA to produce adducts. Some of these DNA adducts exist in multiple conformations that are slowly interconverted to one another. Different conformations have been implicated in different mutagenic and repair outcomes. However, studies on the conformation-specific inhibition of replication, which is more relevant to cell survival, are scarce, presumably due to the structural dynamics of DNA lesions at the replication fork. It is difficult to capture the exact nature of replication inhibition by existing end-point assays, which usually detect either the ensemble of consequences of all the conformers or the culmination of all cellular behaviors, such as mutagenicity or survival rate. We previously reported very unusual sequence-dependent conformational heterogeneities involving FABP-modified DNA under different sequence contexts (TG₁*G₂T [67%B:33%S] and TG₁G₂*T [100%B], G*, *N*-(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl) (Cai et al. *Nucleic Acids Research*, 46, 6356-6370 (2018)). In the present study, we attempted to correlate the *in vitro* inhibition of polymerase activity to different conformations from the single FABP-modified DNA lesion. We utilized a combination of surface plasmon resonance (SPR) and HPLC-based steady-state kinetics to reveal the differences in terms of binding affinity and inhibition with polymerase between these two conformers (67%B:33%S and 100%B).

KEYWORDS

4-aminobiphenyl; bulky DNA lesion, conformational heterogeneity, surface plasma resonance (SPR) binding kinetics; steady state enzyme kinetics, Klenow fragment

INTRODUCTION

The human genome is under constant assault from the environment. DNA-damage-induced mutations are known to trigger chemical carcinogenesis [1-5]. Therefore, understanding the biological responses of cells to DNA mutations is important. Arylamines belong to a notorious group of environmental mutagens/carcinogens that produce bulky DNA adducts *in vivo* [6-10]. They are known to adopt different unique conformational motifs: the major groove B-type (B), stacked (S) (Figure 1a) [11], or the minor groove wedge (W) [12]. Consequently, these bulky-lesion induced conformational heterogeneities complicate mutational and repair outcomes [9-11, 13-15].

4-Aminobiphenyl (ABP) is a major etiological agent of human bladder carcinogen and a potent urinary-bladder carcinogen in experimental animals. As such, commercial production of ABP is banned, however, exposure to ABP can still take place from cigarette smoke. ABP is activated by cellular *N*-acetyltransferase to produce dG-C8-substituted adduct as a major DNA lesion (dG-C8-ABP, Figure 1b) [12, 16-20]. The majority of human bladder cancer has a mutation in the p53 gene. Compared with other cancers, the ABP-induced mutations are more evenly distributed along the p53 gene and the mutation hotspots occur at both CpG, such as codons 175, 248, and 273, and non-CpG sites, such as codons 280 and 285, the latter two being unique mutational hotspots for bladder and other urinary tract cancers [21]. The major induced mutation was is G to T transversion mutation. Translesion synthesis (TLS) over dG-C8-ABP in two different sequences (CCG*GAGGC and CCGGAG*GCC, G*=dG-C8-ABP), which represent codon 248 and 249 sequences of the human p53 tumor suppressor gene,

respectively, has confirmed that codon 248 is both hot spot of ABP adduct formation and G to T mutation [22]. These results suggest that the efficiency of TLS over dG-C8-ABP is influenced by the surrounding DNA sequences. The structurally similar liver carcinogens 2-aminofluorene (AF) and *N*-2-acetylaminofluorene (AAF) produce similar C8-dG adducts. AF could be error-free by correctly base pairing with an incoming dC, whereas AAF blocks the replication process and requires recruitment of lesion bypass polymerases for TLS. In *Escherichia coli*, the TLS of AF produces point and frameshift mutations, whereas bypass of the AAF lesion is frequently accompanied by a frameshift mutation.

Figure 1. (a) Two major conformational views of dG-C8-FABP [*N*-(2'-deoxyguanosin-8-yl)-4'-fluoro-4-aminobiphenyl]: B-, and S-conformers. FABP-red, modified dG-blue, complementary C-orange. (b) The chemical structure of dG-C8-FABP.

Our previous structural studies showed that fully paired duplexes containing the fluorine containing model 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 TG₁*G₂T (G*=FABP) sequence context at 25 °C [23]. Meanwhile, the same lesion in the TG₁G₂*T context exhibited exclusively the B-type conformation under identical experimental condition. When a replication polymerase

encounters a bulky DNA lesion, the polymerase is likely to stall thus stopping DNA synthesis. The lesion also influences whether the DNA replication will be error-free or error-prone. The aforementioned striking sequence effect in buffer systems warrants systematic studies in the presence of a polymerase. Vaidyanathan et al. [24] probed the sequence effect of dG-C8-AF (AF, aminofluorene), a structural analog of ABP, on nucleotide insertion efficiencies catalyzed by the Klenow fragment (Kf-exo⁻) on TG*A and CG*A sequences [25-26]. They found that the S conformer of CG*A thermodynamically favors the insertion of mutagenic A over non-mutagenic C at the lesion site. Xu et al. reported that Kf-exo⁻ is a strong binder to template/primer junctions, but with minimal nucleotide selectivity against modified DNA [27]. The sequence-dependent conformational heterogeneity may play an important role in DNA replication and mutation. We hypothesize that different conformations lead to different extents of polymerase binding affinities, kinetic behaviors, and replication blocks, ultimately resulting in complex toxic and mutational outcomes. Here we conducted surface plasmon resonance (SPR) binding experiments and steady-state enzyme kinetics *in vitro* to probe the effect of ABP-induced conformational heterogeneity on DNA replication.

RESULTS

DNA Sequence Systems

For SPR binding experiments, we constructed two biotinylated hairpin-based template–primer strands (Figure 2a), 85-mer G_1^* adduct and 84-mer G_2^* adduct ($G^* = \text{dG-C8-FABP}$). Specifically, two FABP-modified–biotin–31-mer DNA sequences ($TG_1^*G_2T$ and $TG_1G_2^*T$) were purified by HPLC and characterized individually by enzyme digestion/MALDI-MS [27-30] (Figure S1). A 54-mer hairpin DNA was annealed and ligated to the biotin–31-mer $TG_1^*G_2T$ oligonucleotide, whereas a 53-mer hairpin DNA was ligated to the biotin–31-mer $TG_1G_2^*T$ oligonucleotide. The hairpin structures were used to improve the thermal stability of the oligonucleotide duplex during SPR experiments and the lesion G_1^* and G_2^* adduct were placed at the 21st and 22nd bases, respectively, to avoid clash between the polymerase protein and the gold chip surface [27, 31-32].

For HPLC-based 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) (Figure 3a and 3b) [23, 33-35].

Figure 2. (a) Hairpin template–primer oligonucleotide constructs of 85-mer G₁ and 84-mer G₂ for Kf-exo[−], H denotes 3'-dideoxy-nucleotide; (b) Sensorgrams of binary complexes of Kf-exo[−] with 85-mer and 84-mer unmodified control, G₁-FABP and G₂-FABP modified sequences (1:1 binding fitted curves are overlaid as black lines); (c) K_a (K_a-modified / K_a-control) and K_d (K_d-modified / K_d-control) ratio, G₁ and G₂ represent the ratio of 85-mer FABP-G₁/85-mer-control and 84-mer FABP-G₂/84-mer-control, respectively.

Figure 3. DNA replication models for (a) FABP-modified G_1^* (TG*GT) templates and (b) FABP-modified G_2^* (TGG*T) templates. n: lesion site. Lineweaver-Burk model of DNA synthesis catalyzed by Kf-exo⁻ at (c) 10-mer, G_1 -FABP and (d) 9-mer G_2 -FABP lesion sites with dCTP at steady-state. The G_1 -FABP against with Kf-exo⁻ adopted a competitive inhibitor model, whereas the behavior of $TG_1G_2^*T$ showed as non-competitive (see text).

Oligonucleotide Characterization by MALDI-TOF MS (Matrix-Assisted Laser Desorption/ Ionization-Time of Flight Mass Spectrometry)

The FABP-modified biotin–31-mer oligonucleotides were characterized by exonuclease enzyme digestion followed by MALDI-TOF MS analysis in accordance with the published procedures [27, 36]. In the present case, 5'-3' exonuclease digestion on DNA was difficult to carry out due to the binding hindrance of 5'-biotin motif to the enzyme. Figure S1b 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 from 6 min to 10 min was assigned to the G_1^* -FABP-modified 21-mer. These results confirm the first eluting peak (peak 1) from the HPLC profile (Figure S1a) is biotin–31-mer $TG_1^*G_2T$. Figure S1c presents the MALDI-TOF MS spectra of the peak 2 on HPLC 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 digestion stalled from 4-10 min at m/z 7,116, which corresponds to the 22-mer fragment containing the FABP-modified guanine (theoretical $m/z = 7,116$). These results confirm that peak 2 is G_2^* -FABP-modified biotin–31-mer.

The 84- and 85-mer biotinylated oligonucleotides were purified and identified by 15% denaturing polyacrylamide gel (Hercules, CA) [27]. Figure S2 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 correspondingly. The ligated and purified 85-mer control/ G_1^* and 84-mer control/ G_2^* were used for SPR experiments.

HPLC-based Steady-state Kinetics

We conducted steady-state experiments to investigate the impact of conformational heterogeneity on nucleotide insertion kinetics [37]. The *E. coli* exonuclease-deficient Kf-exo[−] was used for single-nucleotide incorporation. Although the modified base could pair with dCTP to complete the primer extension reaction induced by Kf-exo[−], the reaction efficiency was much poorer than the regular DNA template. The change of efficiency is represented by the enzyme kinetic parameters, K_m and k_{cat} , and the results are summarized in Table 1. The bulky C8 adduct on guanine does not directly block the Watson-Crick base pairing, but it could either physically interfere with the dNTP binding pocket in Kf-exo[−] when the FABP-G holds an “S” conformation; or distort Kf-exo[−] structure in the ternary complexes and influence the geometry at the active site of forming phosphodiester bonds when FABP-G holds the “B” conformation (Figure 1a). In both scenarios, the bulky adduct acts as an inhibitor, but in two different ways (Figure 3). In order to apply the inhibition kinetic model, the whole primer extension assay was performed by maintaining the concentration of inhibitor (FABP-containing DNA duplex), and varying the concentration of substrate, dNTP (dCTP or dATP). 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 (Figure 3).

dCTP Incorporation: Specifically, adding dCTP to the n-1 (16/10-mer) G_1^* adduct elongated a single 11-mer (Figure 3a), whereas the n-1 (16-/9-mer) G_2^* adduct produced 10- and 11-mer (Figure 3b). The formation of 10- and 11-mer mixture from the G_2^* adduct sequence was difficult to quantify; thus, the reduction of the starting material 9-mer was employed to determine the kinetic parameters for both G_1^* and G_2^* reactions.

The n-1 (16/9-mer) G_2^* adduct shows a K_m value (5.7 μM) similar to that of the control (5.8 μM); such similarity indicates an almost equal affinity (Table 1). However, the insertion efficiency f_{ins} of dCTP opposite $-G_2^*$ is 4-fold lower than that of the unmodified control (0.24 :1) (Table 1). By contrast, the K_m of the G_1^* adduct shows a lower affinity (23.4 μM) than that of the control (7.4 μM). The insertion efficiency f_{ins} in G_1^* is three times lower than the control (0.33 to 1). These two forms of enzyme inhibition can be clearly shown by the Lineweaver-Burk plot. As shown by the Lineweaver–Burke enzyme inhibition model (Figure 3c), the (16/10-mer) G_1^* adduct and its control intersect on the Y-axis and a competitive inhibition behavior is suggested. By contrast, the n-1 16/9-mer G_2^* adduct and its control merge at the x-axis. This behavior can be explained as non-competitive inhibition (Figure 3d). 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 from the HPLC profile. This result was achieved because the bulky FABP on the G_2 at the lesion site (n) does not block the replication from n-1 to n+1.

Table 1. Steady-state kinetic parameters for insertion of dCTP opposite unmodified and FABP–dG adduct with Kf-exo[−]

Sequence	K _m (μM)	k _{cat} (min ^{−1})	k _{cat} /K _m (μM ^{−1} ·min ^{−1})	*f _{ins}
FABP-G ₁	23.4 ± 0.01	29.7 ± 0.63	1.27	0.33
Control-G ₁	7.4 ± 0.10	28.8 ± 0.96	3.87	1.00
FABP-G ₂	5.7 ± 0.01	5.5 ± 0.11	0.97	0.24
Control-G ₂	5.8 ± 0.11	23.2 ± 0.70	4.00	1.00

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

dATP Incorporation: The long-range lesion effect was examined by initiating the primer elongation with a 16/8-mer template/primer that produces n-2 for G₂* (Figure 3b) and n-3 for G₁* (Figure 3a). Figure S3a presents the results at 10 min. In the control, ~70% of the 8-mer was extended to 9-mer. With adduct on G₁, ~40% of the 8-mer was converted to 9-mer, indicating a moderate pre-lesion effect at G₁*. However, G₂* blocked ~80% of the 8-mer converting to 9-mer, which indicates a much stronger pre-lesion effect than that of G₁*. These results are not surprising because the 8-mer primer is closer to the G₂* adduct than the G₁* adduct. In the 11-mer (Figure S3b), unmodified control exhibited 90% of dATP insertion. However, only ~20% of the 11-mer was elongated to 12-mer opposite the G₁* 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 for G₂*. These results suggest significant retardation of insertion close to the lesion site.

Kf-exo[−] SPR binding Kinetics

Binary System—Figure 2b shows the sensorgrams for the binary binding between Kf-exo[−] and unmodified controls (85-mer/84-mer TGGT) or modified sequences (85-mer TG₁*G₂T/84-mer TG₁G₂*T). These sequences represent replication fork for the G₁* and G₂*, respectively. Kf-exo[−] shows a much stronger binding affinity (K_D) to the modified 85-mer TG₁*G₂T (8.3-fold) and 84-mer (5.6-fold) TG₁G₂*T sequences relative to their unmodified controls, 85-mer and 84-mer TGGT, respectively (Table 2). The differences are more striking in the dissociation rates (85-mer TGGT k_d : 0.84 s^{−1} vs. 85-mer TG₁*G₂T k_d : 0.17 s^{−1}; 84-mer TGGT k_d : 0.03 s^{−1} vs. 84-mer TG₁G₂*T k_d : 0.009 s^{−1}). There is a substantial difference in K_d ratio (K_d -modified/ K_d -control) between G₁ and G₂ (Figure 2c). By contrast, the K_a ratios of G₁ and G₂ are close to each other, indicating similar association rates. The stabilization energies here specifically measure the energies of forming hydrogen bonds between Kf-exo[−] and DNA sequences [38]. For the binary system the net stabilization energies are positive (Table S1) which indicate the hydrogen bonds that form between Kf-exo[−] and modified sequences (85-mer TG₁*G₂T/84-mer TG₁G₂*T) are stronger than their unmodified controls (85-mer/84-mer TGGT). It is interesting to note that 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 TG₁G₂*T is 12.2-fold greater than that of the 85-mer TG₁*G₂T.

Ternary System— dNTP was added to form Klenow ternary complexes. Kf-exo[−] bound tightly to the unmodified controls when the correct dCTP was introduced in the ternary system (Figure 4). In the unmodified 85-mer, the binding affinities of dATP, dGTP, and dTTP to the 84-mer controls are reduced by 315-, 284-, and 89-fold,

respectively, relative to that of dCTP (Table 2). However, nucleotide selectivity is significantly decreased for the ternary complexes with FABP-dG adduct. The K_D value for the 85-mer TG₁*G₂T is only reduced by ~5-fold in dATP, dGTP, and dTTP than that of dCTP. Similar results were obtained for the 84-mer G₂* adduct, where binding tightness is reduced by ~2-fold in dATP, dGTP, and dTTP compared with that of dCTP. Moreover, as in the binary system, slower dissociation rates were observed for the unmodified and G₂* modified 84-mers. The K_D values of dCTP, dATP, dGTP, and dTTP in the unmodified and G₂* modified 84-mers are smaller than those of the unmodified 85-mer (5- and 13-fold smaller in dCTP and dTTP) and G₁* modified 85-mer TG₁*G₂T (4-, 13-, 10-, and 12-fold smaller in dCTP, dATP, dGTP, and dTTP, respectively).

Table 2. SPR binding affinities, K_D (nM) of unmodified (TG₁G₂T) and FABP–dG adducts (TG₁[FABP]G₂T and TG₁G₂[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.

Sequence	Binary			K_D of Ternary (nM)			
	k_a (1/Ms) $\times 10^7$	k_d (1/s)	K_D (nM)	dCTP	dATP	dGTP	dTTP
85-mer TG₁*G₂T	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 TG₁G₂*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)

[#]standard deviation

(a)

(b)

(c)

(d)

Figure 4. Sensorgrams of the 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).

DISCUSSION

We previously conducted [23] systematic spectroscopic, thermodynamic, and chip binding (^{19}F -NMR, CD, DSC, and SPR) studies for the extension of 16-mer $\text{TG}_1^*\text{G}_2\text{T}$ and $\text{TG}_1\text{G}_2^*\text{T}$ ($\text{G}^*=\text{FABP}$) sequences. These protein free model systems mimic a translesion synthesis of the bulky FABP lesion in two very distinctive sequence-dependent conformational heterogeneities at G_1 (67%B:33%S) and G_2 (100%B). The results indicate that the sequence-dependent conformational complexities appear to persist at various elongation positions, including the ss/ds junction. The B-conformer is a major thermodynamic stabilizer in duplex settings, whereas the S-conformer is a destabilizer. However, the opposite is the case for adduct at the ss/ds junction. In particular, the S-conformation promotes lesion stacking with nascent base pairs at the replication fork. SPR results reveal that the S-conformation increases the binding affinity with the complementary strands in the order of $\text{G}_1^* > \text{G}_2^*$. In the present work, we examined the effects of these unusual sequence effects on Klenow polymerase binding (binary and ternary) and nucleotide insertion kinetics.

Improvement on Model Hairpin Oligonucleotide Construction

We previously reported a construction of biotinylated hairpin-based template–primer strand for DNA–polymerase SPR binding studies [27, 39-40]. The general strategy was to ligate a biotinylated arylamine-modified 31-mer sequence with 52-mer hairpin DNA to form an 83-mer hairpin. This was then followed by addition of a ddNTP to the 3'-end to prevent potential primer elongation. However, the ddNTP addition step was low yielding and the purification of ligated products was difficult. In the present

study, we succeeded in direct ligation of a biotinylated modified 31-mer with ddNTP-containing 53- and 54-mer hairpin DNA. This process improved the yield significantly (from ~10% to ~ 30%), and the products were readily separated by HPLC. Also, previously, individual dNTPs (100 μ M) were mixed with varying concentrations of Kf-exo⁻ in sample buffers and injected over the surface without any dNTP in the running buffer. This would create a complication resulting in two variables, concentrations of dNTPs and Kf-exo⁻ [39]. In the present work, we circumvented the complication by adding individual dNTP in running buffer and introducing Kf-exo⁻ directly onto the chip surface. This system has only one variable, thus increasing the system stability and accuracy for the translesion synthesis.

Lesion and Sequence Effects on SPR Binding Affinities and Kinetics

Tight binding of Kf-exo⁻ with the unmodified dG control was observed when the correct nucleotide dCTP was presented in both 85-mer and 84-mer. This result is in a good agreement with expectation and shows high nucleotide selectivity (K_D dCTP $\ll$ K_D dTTP $<$ K_D dATP $\sim$ K_D dGTP). Meanwhile, a remarkably 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 G₁* and 84-mer G₂* interactions are 8.3 and 5.6 fold higher than the corresponding unmodified controls, respectively. Adduct-induced tighter binding affinity with Kf-exo⁻ has been reported [27, 41] 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 is low. In particular, the K_D of the dCTP at the opposite G₁* and G₂* are only ~5- and ~2-fold tighter than the incorrect dATP,

dGTP, and dTTP, respectively. This is a strong lesion effect. We observed that the usual 1:1 model SPR simulation did not provide a clean fit of dCTP for G_1^* in the ternary system. It is possible that the present DNA adduct-Kf complex exist in multiple stages due to the FABP-induced S/B conformational heterogeneity at G_1^* replication fork. However, in most cases conformational changes in protein are much faster than SPR time scale and thus additional studies would be necessary to confirm these possibilities.

Our HPLC-based steady-state kinetics data indicate competitive inhibition for the S/B-conformer replication fork at G_1^* . In other words, the S/B conformational heterogeneity at G_1^* is inhibitory to replication probably due to some unfavorable clash 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 dG. However, the exclusively B-conformeric G_2^* accommodates well for the incoming dCTP, resulting in non-competitive inhibition. The equal affinity (K_m) between G_2^* and its unmodified dG reveals that the B-conformation in the major groove may not interfere with the insertion of the correct dCTP. This finding might also explain the greater K_D of dCTP at G_1^* over G_2^* of the SPR results and indicates a weaker binding affinity for G_1^* over G_2^* . Alternatively, the B-conformer may not interfere with the Watson–Crick base pair, but it may alter the native conformations or hinder the formation of phosphodiester bond. The S-conformer portion of G_1^* cannot provide a proper Watson–Crick base pair for replication and it may need to convert back to the B-conformation to be replicated properly. Thus, the S-conformer may function as a competitive inhibitor for 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 necessary step in the DNA polymerase proofreading process as well.

MATERIALS AND METHODS

Model Oligonucleotide DNA sequences.

FABP-modified 16-mer templates were prepared in accordance with the published procedures [42]: Two mono-adducts (d(5'-CTTCTG₁*G₂TCCTCATTC-3')) (G* = FABP) (G₁* adduct) and (d(5'-CTTCTG₁G₂*TCCTCATTC-3')) (G₂* 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 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 sequence, 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 mobile phase (Figure S1a).

HPLC-based Steady-State Kinetic Analysis

The extension efficiency of Kf-exo⁻ polymerase opposite G₁* and G₂* adducts was determined by steady-state kinetic experiments as described by published procedures [27, 43-45]. 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 initiate nucleotide insertion. All reactions were quenched by adding 10mM EDTA 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 with a 1.5 M ammonium acetate gradient of 40-60% in water for 6.5 min and quantified by the absorbance of UV at 260 nm. All reactions were performed at 22 °C in triplicate, and the results were analyzed with a GraphPad Prism 5 software by using the Lineweaver–Burk model. The inhibition curves were fitted to the equation $1/V = 1/V_{\max} + (K_M/V_{\max})/[S_{\text{dCTP}}]$. The relative insertion efficiency f_{ins} was obtained as $(k_{\text{cat}}/K_M)_{\text{modified}}/(k_{\text{cat}}/K_M)_{\text{unmodified}}$.

SPR Measurements

FABP-modified Hairpin Template/Primer Constructs. 5'-Biotinylated 31-mer containing dG-C8-FABP in the “-TG₁G₂T-” sequence context was used in SPR analysis following the reported procedures [27, 39-40]. The two FABP-modified biotinylated 31-mer G₁* and G₂* adducts were separated by RP-HPLC and characterized by MALDI-TOF MS [36]. The 84-mer and 85-mer hairpin–template–primer were prepared by following the reported protocols [27, 39]. Briefly, two different lengths of hairpin DNA sequences (53- and 54-mer) were phosphorylated at their 5'-ends, but their 3'-ends were modified with ddA and ddC, respectively, to prevent further primer elongation (Figure 2a). The biotinylated 31-mer modified G₁* 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 oligonucleotide was purified in 15% denaturing polyacrylamide gel and extracted using the electroelution method followed by desalting by again using G-25 spin columns. The corresponding biotinylated 85-mer was finally purified by RP-HPLC. G₂* 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 S2).

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 [27, 39]. 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 three times, and the system was equilibrated with running buffer for 2 h. The 84- and 85-mer unmodified and modified G₁* and G₂* 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 by SPR.

The SPR system was first 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 cell 2 to 4 (cell 1 as blank reference). Surface testing, regeneration buffer scouting, and mass transport limitation test were conducted prior to kinetic experiments following previous reports [27, 39]. 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 previously [27]. Briefly, the 1× HBS-P running buffer (containing 100 µg/mL bovine serum albumin [BSA] and 5 mM MgCl₂) 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 MgCl₂, 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 provided by the manufacturer (GE Healthcare).

CONCLUSIONS

In this paper, we conducted SPR and HPLC-based steady-state kinetics studies to probe adduct-induced conformation-dependent replication block during translesion synthesis (TLS). The FABP-modified DNA adduct adopts a mixture of B and S conformations in the $TG_1^*G_2T$ (67%B:33%S), but shows exclusively B conformation in the $TG_1G_2^*T$ sequence context [23]. According to the present Lineweaver–Burk enzyme inhibition model, the S/B-conformeric mix G_1^* adduct exhibits a competitive-inhibition, whereas the B-conformeric G_2^* adduct behaves as a non-competitive inhibitor in the nucleotide insertion step. These results indicate that the S-conformer may not be able to accommodate the incoming dCTP and exhibits a competitive behavior with incoming dNTPs, thus blocking replication. By contrast, the exclusive B-conformer G_2^* does not interfere with the Watson-Crick base pairing, resulting in a proper dCTP insertion. As such, the B-conformer shows a non-competitive-inhibition behavior. The SPR binding results implicate an adduct-induced tight binding with Kf-exo[−] in a binary system. In the ternary system, nucleotide selectivity decreases when G_1 and G_2 are modified by FABP. From these experiments, we observed the effect of conformational heterogeneity induced by bulky lesion on replication block.

It has been reported that compared with other cancers, the ABP-induced mutations are more evenly distributed along the p53 gene and the mutation hotspots occur through the genome with the major mutation being G to T transversion [21-22]. TLS over dG-C8-ABP in two different sequences (CCG*GAGGC and CCGGAG*GCC, G*=dG-C8-ABP), which represent codon 248 and 249 sequences of the human p53 tumor suppressor gene, respectively, has confirmed that codon 248 is a hot spot for adduct

formation and G to T mutation. These results suggest that the efficiency of TLS over dG-C8-ABP is affected by the surrounding DNA sequences of the ABP lesion, consequently the B/S conformational heterogeneity as described here. Elucidation of conformation-specific bypass, mutational and repair processes over the ABP adducts in cell should clarify the molecular mechanisms underlying ABP-induced mutagenesis and carcinogenesis.

In the present paper, we demonstrated the combination of SPR binding and HPLC steady-state kinetics as a power tool in investigating FAPB-induced conformational heterogeneity in TLS. This approach can be applied to studying other bulky DNA adducts.

Funding: “This research was funded by National Institutes of Health CA098296 (B.P.C.), [R21ES028384](#) (B.P.C.), and R15 CA213042 (D.L.), and R01 ES028865 (to D.L.).

Acknowledgments: “This research was supported in part by RI-INBRE Research Core Facility supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health [P20GM103430].

Conflicts of Interest: “The authors declare no conflict of interest.” “The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.”

REFERENCES

1. Voulgaridou, G.-P.; Anestopoulos, I.; Franco, R.; Panayiotidis, M. I.; Pappa, A., DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **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. Barnes, J. L.; Zubair, M.; John, K.; Poirier, M. C.; Martin, F. L., Carcinogens and DNA damage. *Biochemical Society transactions* **2018**, *46* (5), 1213-1224.
4. Loeb, L. A.; Harris, C. C., Advances in Chemical Carcinogenesis: A Historical Review and Prospective. *Cancer Research* **2008**, *68* (17), 6863.
5. Poirier, M. C., Chemical-induced DNA damage and human cancer risk. *Nature Reviews Cancer* **2004**, *4*, 630.
6. 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.
7. Hein, D. W.; Doll, M. A.; Rustan, T. D.; Gray, K.; Feng, Y.; Ferguson, R. J.; Grant, D. M., Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis* **1993**, *14* (8), 1633-1638.
8. Fretland, A. J.; Doll, M. A.; Zhu, Y.; Smith, L.; Leff, M. A.; Hein, D. W., Effect of nucleotide substitutions in N-acetyltransferase-1 on N-acetylation (deactivation) and O-acetylation (activation) of arylamine carcinogens: implications for cancer predisposition. *Cancer detection and prevention* **2002**, *26* (1), 10-14.

9. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. *Biochemistry* **1994**, *33* (6), 1373-1384.
10. Cho, B. P., Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *Journal of Environmental Science and Health, Part C* **2004**, *22* (2), 57-90.
11. 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.
12. 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.
13. Zhou, L.; Rajabzadeh, M.; Traficante, D. D.; Cho, B. P., Conformational heterogeneity of arylamine-modified DNA: 19F NMR evidence. *Journal of the American Chemical Society* **1997**, *119* (23), 5384-5389.
14. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR structural studies of a 15-mer DNA sequence from a ras protooncogene, modified at the first base of codon 61 with the carcinogen 4-aminobiphenyl. *Biochemistry* **1992**, *31* (40), 9587-9602.
15. Liang, F.; Meneni, S.; Cho, B. P., Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene– DNA adducts. *Chemical research in toxicology* **2006**, *19* (8), 1040-1043.

16. 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.
17. 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.
18. Doerge, D. R.; Churchwell, M. I.; Marques, M. M.; Beland, F. A., Quantitative analysis of 4-aminobiphenyl-C8-deoxyguanosyl DNA adducts produced in vitro and in vivo using HPLC–ES-MS. *Carcinogenesis* **1999**, 20 (6), 1055-1061.
19. Beland, F. A.; Doerge, D. R.; Churchwell, M. I.; Poirier, M. C.; Schoket, B.; Marques, M. M., Synthesis, Characterization, and Quantitation of a 4-Aminobiphenyl–DNA Adduct Standard. *Chemical research in toxicology* **1999**, 12 (1), 68-77.
20. Cho, B., Structure-function characteristics of aromatic amine-DNA adducts. *The Chemical biology of DNA damage* **2010**, 217-238.
21. Feng, Z.; Hu, W.; Rom, W. N.; Beland, F. A.; Tang, M.-s., 4-aminobiphenyl is a major etiological agent of human bladder cancer: evidence from its DNA binding spectrum in human p53 gene. *Carcinogenesis* **2002**, 23 (10), 1721-1727.
22. Yagi, T.; Fujikawa, Y.; Sawai, T.; Takamura-Enya, T.; Ito-Harashima, S.; Kawanishi, M., Error-prone and error-free translesion DNA synthesis over site-specifically created DNA adducts of aryl hydrocarbons (3-nitrobenzanthrone and 4-aminobiphenyl). *Toxicological research* **2017**, 33 (4), 265.

23. Cai, A.; Wilson, K. A.; Patnaik, S.; Wetmore, S. D.; Cho, B. P., DNA base sequence effects on bulky lesion-induced conformational heterogeneity during DNA replication. *Nucleic Acids Res* **2018**, *46* (12), 6356-6370.
24. 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.
25. Eger, B. T.; Kuchta, R. D.; Carroll, S. S.; Benkovic, P. A.; Dahlberg, M. E.; Joyce, C. M.; Benkovic, S. J., Mechanism of DNA replication fidelity for three mutants of DNA polymerase I: Klenow fragment KF (exo+), KF (polA5), and KF (exo-). *Biochemistry* **1991**, *30* (5), 1441-1448.
26. Benner, S.; Chen, R. J.; Wilson, N. A.; Abu-Shumays, R.; Hurt, N.; Lieberman, K. R.; Deamer, D. W.; Dunbar, W. B.; Akeson, M., Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nature nanotechnology* **2007**, *2* (11), 718.
27. 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.
28. Sandineni, A.; Lin, B.; MacKerell Jr, A. D.; Cho, B. P., Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis. *Chemical research in toxicology* **2013**, *26* (6), 937-951.

29. Xu, L.; Cho, B. P., Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the Nar I Mutational Hotspot. *Chemical research in toxicology* **2016**, *29* (2), 213-226.
30. Suzuki, N.; Yasui, M.; Santosh Laxmi, Y.; Ohmori, H.; Hanaoka, F.; Shibutani, S., Translesion synthesis past equine estrogen-derived 2'-deoxycytidine DNA adducts by human DNA polymerases η and κ . *Biochemistry* **2004**, *43* (35), 11312-11320.
31. Sakamoto, K.; Gouzu, H.; Komiya, K.; Kiga, D.; Yokoyama, S.; Yokomori, T.; Hagiya, M., Molecular computation by DNA hairpin formation. *Science* **2000**, *288* (5469), 1223-1226.
32. LaBean, T. H.; Yan, H.; Kopatsch, J.; Liu, F.; Winfree, E.; Reif, J. H.; Seeman, N. C., Construction, analysis, ligation, and self-assembly of DNA triple crossover complexes. *Journal of the American Chemical Society* **2000**, *122* (9), 1848-1860.
33. Vastenhout, K. J.; Tornberg, R. H.; Johnson, A. L.; Amolins, M. W.; Mays, J. R., High-performance liquid chromatography-based method to evaluate kinetics of glucosinolate hydrolysis by *Sinapis alba* myrosinase. *Analytical biochemistry* **2014**, *465*, 105-113.
34. Klingaman, C. A.; Wagner, M. J.; Brown, J. R.; Klecker, J. B.; Pauley, E. H.; Noldner, C. J.; Mays, J. R., Synthesis and spectral characterization of 2,2-diphenylethyl glucosinolate and HPLC-based reaction progress curve data for the enzymatic hydrolysis of glucosinolates by *Sinapis alba* myrosinase. *Data in Brief* **2017**, *10*, 151-181.

35. Miller, H.; Grollman, A. P., Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry* **1997**, 36 (49), 15336-15342.
36. 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.
37. Vrzheschch, 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.
38. Kuriyan, J.; Konforti, B.; Wemmer, D., *The molecules of life: Physical and chemical principles*. Garland Science: 2012.
39. 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.
40. 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.
41. 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.

42. 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.
43. 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.
44. 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.
45. Krüger, S.; Meier, C., Synthesis of Site-Specific Damaged DNA Strands by 8-(Acetylaryl amino)-2'-deoxyguanosine Adducts and Effects on Various DNA Polymerases. *European Journal of Organic Chemistry* **2013**, *2013* (6), 1158-1169.

SUPPLEMENTAL INFORMATION

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

Ang Cai, Ke Bian, Fangyi Chen, Qi Tang, Rachel Carley, Deyu Li,* and Bongsup Cho*

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

*Co-correspondence: bcho@uri.edu, Tel: +1 401-874-5024; deyuli@uri.edu, Tel: +1 401-874-9361

TABLE OF CONTENTS

Figure S1. HPLC profile of FABP-modified bio-31mer TGGT template and MALDI-TOF characterization of peak 1 and peak 2.

Figure S2. Denaturing gel (15%) profiles of 84-mer and 85-mer ligated oligonucleotides, 31-mer, 53-mer and 54-mer non-ligated oligonucleotides.

Figure S3. dATP insertion efficiency results of (a) 8-mer to 9-mer and (b) 11-mer to 12-mer in control, G1*G2T-FABP and G1G2*T-FABP, respectively, at 10 min.

Table S1. Binding net stabilization energy of unmodified and FABP-adducts with Kf-exo- in binary system (1:1 binding).

Figure S1. HPLC profile of FABP-modified bio-31mer TGGT template and MALDI-TOF characterization of peak 1 and peak 2. (a) HPLC chromatogram of a reaction mixture of biotinylated-31-mer sequence (5'-Bio-CCTCTTCCCTCACCTCTTC TG₁G₂TCCTCATTC-3') with an activated **FABP** (*N*-Acetoxy-*N*-(trifluoroacetyl)-4'-fluoro-4-aminobiphenyl) and photodiode array UV spectra of unmodified and monoadducts. (b) and (c): MALDI-TOF mass spectra of **FABP** modified biotinylated-31mer-TGGT. (b) 3'-Exonuclease digestions of **peak 1** at 0, 1, 4, 6, 7, 8 and 10min. (c) 3'-Exonuclease digestions 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'-exonuclease digestion.

Figure S2. Denaturing gel (15%) profiles of 84-mer and 85-mer ligated oligonucleotides, 31-mer, 53-mer and 54-mer non-ligated oligonucleotides.

(a)

(b)

Figure S3. dATP Insertion efficiency 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, at 10 min (see Figure 3 for sequences)

Table S1. Binding net stabilization energy of unmodified and FABP-adducts with Kf-exo⁻ in binary system (1:1 binding).

Sequence	k_d (1/s)	*Net stabilization Energy (kcal/mol)
85-mer control	0.84 (0.004)	0
85-mer TG ₁ [FABP]G ₂ T	0.17 (0.002)	0.95
84-mer control	0.03 (0.002)	0
84-mer TG ₁ G ₂ [FABP]T	0.009 (0.0001)	0.71
*Net stabilization energy (kcal/mol)= $-RT\ln(k_d)_{\text{modified}} - [-RT\ln(k_d)_{\text{unmodified}}$		