2024

DNA DAMAGE, MODIFICATION AND REPAIR

Xianhao Zhou
University of Rhode Island, zhouxianhao@outlook.com

DNA DAMAGE, MODIFICATION AND REPAIR

BY

XIANHAO ZHOU

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHARMACEUTICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2024

DOCTOR OF PHILOSOPHY DISSERTATION

OF

XIANHAO ZHOU

APPROVED:

Dissertation Committee:

Major Professor Deyu Li

Bongsup Cho

Gongqin Sun

Fang Wang

Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2024

ABSTRACT

DNA is necessarily to be stable and intact since it is the repository of genetic information in each living cell. However, DNA is not inert, and it is vulnerable to the endogenous and environmental assaults including reactive oxygen species (ROS), ultraviolet (UV) light, and various carcinogens. Exposure to these DNA damaging agents is associated with various adducts ranging from small methyl or etheno adducts to bulky adducts such as acetylaminofluorene. These resulting damages, if not repaired, will lead to replication block or mutation which can eventually cause cancer or other genetic diseases.

The research covered by this dissertation falls into two main categories. The first section focuses on investigations of the guanine adduct generated by Cinnamaldehyde (CA) and its mutagenesis, and the other section discusses how 2OG/Fe(II)-dependent dioxygenases, such as AlkB, repair DNA adducts.

By developing an organic solvent-based reaction, we are able to synthesize a DNA oligonucleotide containing a CA adduct, then we developed a Liquid Chromatography based method to purify the DNA with CA adducts. The results showed that CA can form two isomers with Guanine with a 1:1 ratio distribution.

By employing a mass spectrometry-based approach, we were able to evaluate the in vivo bypass and mutagenesis of CA adducts. The findings indicate that the two isomers exhibit comparable mutation patterns, with approximately 50% of the CA-G adducts being repaired back to G in cells. The most frequent mutation involves a G to

C conversion, accounting for about 45% of the CA-G adduct mutations. Mutations to A and T are relatively rare, occurring in less than 1% of cases.

The second part (chapter 2 and 3) focuses on DNA repair mediated by AlkB. By examining the repair kinetics of AlkB enzymes for various nucleobase modifications and comparing their kinetic parameters with their repair efficiencies in vivo, we established a correlation between in vitro and in vivo repair mechanisms. We introduced a novel metric, the "final repair ratio," which incorporates both the bypass and mutation ratios. Our findings reveal a significant correlation between the in vitro catalytic efficiency (Kcat/Km) and the in vivo final repair ratio. Specifically, a higher Kcat/Km value corresponds to an increased final repair ratio. This relationship enables the prediction of in vivo repair efficiency based on in vitro Kcat/Km values, offering valuable insights for future research into the in vivo activity of AlkB family enzymes and their human homologs.

By analyzing the kinetic parameter (Kcat/Km) of AlkB with various substrates, encompassing both single-stranded and double-stranded forms, we conducted a comprehensive investigation into AlkB's substrate specificity and uncovered insights into its activity on single versus double strands. Additionally, we identified novel substrates for AlkB and assessed their activity by measuring their Kcat/Km values.

In conclusion, this dissertation introduces a systematic methodology for the synthesis, characterization, and purification of a DNA adduct, as well as for conducting in vitro and in-cell repair, bypass, and mutagenesis of DNA adducts. The techniques developed and applied in this study have broader applications, enabling the

investigation of other chemicals that may form adducts and damage DNA. Furthermore, these methodologies can facilitate research into additional DNA repair pathways and contribute new perspectives to the field of DNA chemical biology.

ACKNOWLEDGMENTS

I would like to thank my Ph.D. supervisor, Dr. Deyu Li, for his expertise, guidance, and assistance throughout my graduate career at University of Rhode Island. The professionalism and enthusiasm he passed onto me, always encouraged me to become an independent scientist equipped with curiosity, passion, and vision. I am deeply grateful for his kindness and patience teaching me all the valuable skills in exploring unknowns, writing proposals, carrying out research, and most importantly, being an ambitious researcher with the goal of contributing knowledge to improve people's health benefits. Without his help, my achievements would not have been possible.

I would like to extend my sincere gratitude to my doctoral committee members, Dr. Bongsup Cho, Dr. Gongqin Sun and Dr. Fang Wang. Whenever I need help, they were always there for me. I truly appreciate their support, insights, and advice for my research projects, and serving on my committee.

I want to express my heartiest acknowledgements to Dr. Ke Bian and Dr. Qi Tang for their guidance and inspiration in my first year of graduate study at URI. The accomplishment of this dissertation was possible because of the help I received from my cooperative and resourceful labmates. Their efforts and inputs enabled me to complete this work. I am particularly thankful to Dr. Yi-Tzai Chen, Dr. Rui Qi,

Samuel Howarth, Quentin Mylie, Zhiyuan Peng for their invaluable support over the years.

In addition, I thank Dr. Ang Cai, Dr. Al Bach and Ms. Kim Andrew at Rhode Island INBRE Laboratory for their help in training me on various research instruments and giving me advice when I was having a hard time. Their help in every aspect of my research was imperative to my completion of the Ph.D. degree.

Besides, I thank Dr. Shuai Shao, who supported me during my darkest time. It is my great fortune to have you in my life. I apologize for not being able to list more names, but I will remember all the help they offered me.

PREFACE

This dissertation was following the requirement of Manuscript Format.

Manuscript I was prepared for submission to Chemical Research in Toxicology.

Manuscript II was prepared for submission to Chemical Research in Toxicology.

Manuscript III was prepared for submission to Chemical Research in Toxicology.

The titles for these manuscripts were listed below:

I. **Synthesis, characterization and Mutagenicity study of cinnamaldehyde-induced 1, N^2-cyclized deoxyguanosine adducts in DNA**

II. **Study on correlation of DNA repair efficiency obtained from in vitro and in vivo methods**

III. **Comprehensive kinetic study of the AlkB protein on various alkyl substrates**

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER - I

CHAPTER - II

CHAPTER - II: **SUPPLIMENTARY INFORMATION**

CHAPTER - III

LIST OF FIGURES

CHAPTER - I

CHAPTER – I: SUPPLIMENTARY INFORMATION

CHAPTER – II

CHAPTER – II: SUPPLIMENTARY INFORMATION

CHAPTER – III:

CHAPTER – III: SUPPLIMENTARY INFORMATION

CHAPTER - I

Prepared for submission to Chemical Research in Toxicology.

Synthesis, Characterization and Mutagenicity Study of Cinnamaldehyde-induced 1, N^2-cyclized Deoxyguanosine Adduct in DNA

Xianhao Zhou[†], Jie Wang[†], Quentin Mylie[†], Yi-Tzai Chen[†], and Deyu Li[†]*

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

* Correspondence: Deyu Li (deyuli@uri.edu)

ABSTRACT

Electronic cigarettes have been used extensively among young people in recent years. Cinnamaldehyde, a flavoring agent in electronic cigarette, affects genome integrity and cell metabolism. Here, we report that cinnamaldehyde can induce deoxyguanosine adducts in DNA under physiological conditions. Based on the NMR data (including ^{1}H, ^{13}C, COSY, HSQC and HMBC spectra), together with LC-MS, and exonuclease digestion analyses, the adducts have been characterized as 1, N^2-cyclized deoxyguanosine derivatives. In addition, results revealed that there were 1:1 diastereomers existing in the oligonucleotide. NMR coupling constants revealed that the hydroxyl group at C-10 and the phenyl group at C-12 are in *trans*-configuration as (10*S*, 12*R*)-**2a** and (10*R*, 12*S*)-**2b**. In the in vivo test, we investigated the mutagenicity of the two isomers individually. We constructed the M13mp7(L2) viral genome, which had been site-specifically modified with a cinnamaldehyde adduct, and developed a mass spectrometry-based method to investigate the biological effects of the cinnamaldehyde-guanine lesion during DNA replication. The results indicated that the cinnamaldehyde adduct significantly blocks replication, with over 50% of the cinnamaldehyde-guanine adducts being reversible to guanine. The primary mutation observed was a guanine to cytosine substitution.

INTRODUCTION

Since its introduction in 2003, the use of electronic cigarettes (e-cigarettes) has steadily increased, becoming the most prevalent tobacco product among youth.[1] E-cigarette aerosols are produced by heating liquids typically containing nicotine, glycerin, propylene glycol, flavorings, and other ingredients.[2,3] Cinnamaldehyde (CA, Figure 1), a flavoring agent found in 51% of refill fluids, exhibits high cytotoxicity to human embryonic stem cells (hESC) and human pulmonary fibroblasts (hPF) as demonstrated in MTT assays.[4] Additionally, CA has been shown to compromise respiratory immune cell function and disrupt mitochondrial respiration and glycolysis.[5,6] Furthermore, CA is associated with DNA damage in mammalian cells, both in vivo and in vitro, which may increase the risk of cancer and other diseases.[7–9]

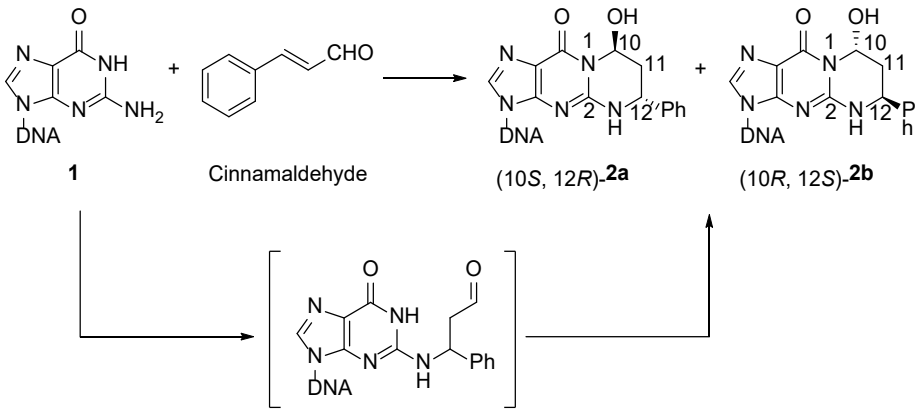


Figure 1. Structure of cinnamaldehyde and proposed mechanism of forming cyclic DNA adducts with guanine.

CA is an α,β-unsaturated compounds, which contains many carcinogens and toxic chemicals and has attracted much attention for its bioactivity in cell and tissue.[18,19] α, β-Unsaturated compounds, such as acrolein and crotonaldehyde, have two electrophilic reactive sites, the β-carbon on the conjugated double bond and the carbon of the carbonyl group. The primary genotoxicity of these compounds is presumably related to the formation of DNA adducts.[10] 2'-Deoxyguanosine (2'-dG) is the most reactive base to react with α,β-unsaturated compounds.[11–13] Mechanistically, the cyclic deoxyguanosine adducts are generated through Michael addition at the olefinic carbon by either N-1 or the exocyclic NH_2 of guanine followed by 1,2-addition at the aldehyde group by the exocyclic NH_2 or N-1.[14–17] Although DNA adduct formation has been studied extensively following human exposure to tobacco smoke[20], less attention has been focused on the chemical structure of potential CA-induced DNA adducts following e-cigarette usage. The NMR data reveal a hydroxyl group at C-10 and a phenyl group at C-12 in *trans*-configuration. Subsequently, we used the synthesized adduct to determine its in vivo toxicity. While two different adducts were generated and they cause mutation and replication block. The results show that 2 isomers induce similar mutations.

EXPERIMENT PROCEDURES

Materials. All chemicals were from Sigma-Aldrich or Fisher Scientific unless otherwise specified. Unmodified oligonucleotides used in this study were purchased from Integrated DNA Technologies.

Synthesis of 1,N^2-deoxyguanosine adducts of cinnamaldehyde. 2'-Deoxyguanosine monohydrate (50 mg, 0.18 mmol) and *trans*-cinnamaldehyde (125 mg, 0.95 mmol) were dissolved in 1.0 mL water containing 10% DMF for 12 hours at 80 ºC. The desired product was purified by silica gel column chromatograpy to yield 21 mg (30% yield) of the objective compound.

LC-MS analyses of oligonucleotide. Analyses of oligonucleotides were performed by high resolution LC-MS (AB Sciex, ABI4600). ESI was conducted by using ionspray voltage floating (ISVF) of -4.5 kV in a negative ion mode. Temperature (TEM) was set at 300 °C. Ion source gas 1 (GS1) was 15 psi; ion source gas 2 (GS2) was 15 psi; and curtain gas (CUR) was 25 psi. Declustering potential (DP) was −100 V; and collision energy was -10 V. Liquid chromatographic separation was achieved by using an Acclaim Polar Advantage II C18 column (2.1 × 250 mm; 3 μm) at a flow rate of 0.1 mL/min. Solvent A was 10 Mm Triethylammonium acetate in water, pH = 7.0, and solvent B was acetonitrile. A solvent gradient was carried out under the following conditions: 5% of B for 5 min, 5 to 20% of B over 40 min, 20 to 80% of B over 3 min, 80% B over 12 min., 80 to 5% of B over 5 min, and 5% B over 5 min. The LC column oven was set at 25 °C.

Synthesis of cinnamaldehyde-modified oligonucleotide. A solution of water/DMSO (1:1) in 8 uL was added to 13 mer oligonucleotide (20 nmol, 5'-TTTTTTGTTTTTT-3') and 4 uL cinnamaldehyde. The reaction mixture was incubated for two days at 37 ºC.

Oligonucleotide purification and characterization with LC-MS. The analyses of oligonucleotides were performed by high resolution LC-MS (AB Sciex, ABI4600). ESI was conducted by using ionspray voltage floating (ISVF) of 4.0 kV in a positive ion mode. Temperature (TEM) was set at 300 °C. Ion source gas 1 (GS1) was 35 psi; ion source gas 2 (GS2) was 35 psi; and curtain gas (CUR) was 20 psi. Declustering potential (DP) was 50 V; and collision energy was 15 V. Liquid chromatographic separation was achieved by using a Luna Omega 5 μm Polar C18 100 Å column (150 × 4.6 mm) at a flow rate of 0.2 mL/min. Solvent A was 10mM Triethylammonium acetate (TEAA) in water, and solvent B was acetonitrile. A solvent gradient was carried out under the following conditions: 5 to 15% of B over 180 min, 15 to 90% of B over 3 min, 90% B over 10 min., 90 to 5% of B over 3 min, and 5% B over 20 min. LC column oven was set at 40 °C. The UV signal at 260 nm was used to detect the oligonucleotide species together with the MS signal.

Exonuclease digestion with MALDI-TOF analysis. The modified oligonucleotides were characterized by 3′ to 5′ exonuclease digestion (SVP) and 5’ to 3’ exonuclease digestion (BSP) followed by MALDI-TOF analysis.[31] For 3′ to 5′ exonuclease digestion, 6.0 μL sample containing 150 pmol of a CA-modified oligonucleotide was used for digestion. SVP (0.01 unit) was added together with 2.0 μL of ammonium citrate (100 mM, pH 9.4). For 5′ to 3′ exonuclease digestion, 6.0 μL sample containing 150 pmol of a CA-modified oligonucleotide was used for digestion. BSP (0.01 unit) was added together with 2.0 μL water. For MALDI-TOF analyses, 1.0 μL reaction

mixture was used at a certain time point. The digestion was quenched by mixing with 1.0 μL MALDI analysis matrix (3-hydroxypicolinic acid and diammonium hydrogen citrate in a 1:1 ratio). Samples were analyzed by a Shimadzu Axima Performance MALDI-TOF mass spectrometer.

Digestion Method. A digestion mixture was prepared by adding 250 Units Benzonase, 300 mU phosphodiesterase I and 200 Units alkaline phosphatase to 5 mL Tris-HCl buffer (20mM, pH = 7.9) containing 100 mM NaCl and 20 mM $MgCl_2$. The oligonucleotide was digested by adding 50 uL digestion mixture and incubation at 37 ^{o}C for 6 h.

M13 genome construction. 13mer oligonucleotides containing lesions and two 21mer barcode-containing sequences (5'-TACCGTCGNNNCGCGCATGCA-3', and 5'-TCTCGAGTGNNNCGTCAGCAC-3') were phosphorylated, annealed with two scaffolds (Scaffold I, 5'-AAAAATGCATGCGCG-3'; Scaffold II, 5'-CACTCGAGAAAAAA-3'), and ligated into 55mer or 58mer oligonucleotide strands. The 55mer and 58mer oligonucleotides were further ligated into M13mp7(L2) single-stranded viral genome by using methods that reported previously.[15,21] The constructed genomes were purified using phenol/chloroform/isoamyl alcohol (25:24:1) extraction and recovered from ethanol precipitation. The scheme of general M13 construction

procedure is shown in Figure 2.

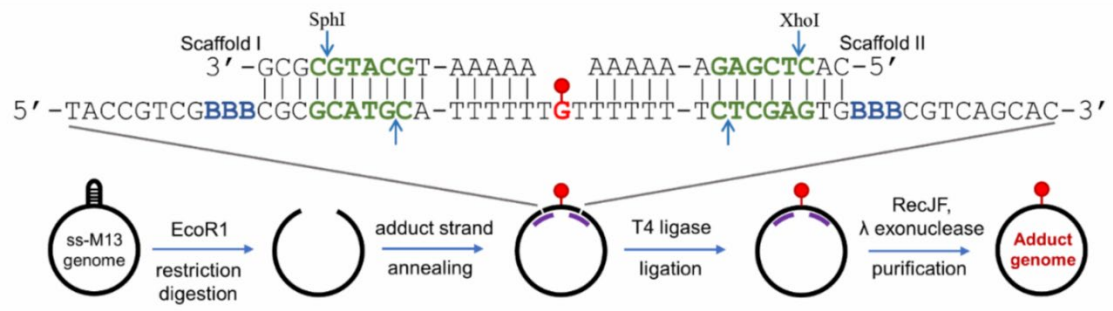


Figure 2. The Procedure of preparing the ssDNA plasmid to demonstrate the replication block and mutagenicity in E.coli. The guanine with an adduct (●) will be ligated into a ssDNA with two scaffolds.

Construction of ssDNA plasmid with site specific Cinnamaldehyde-guanine adduct. We will first construct site-specifically modified genome with Cinnamaldehyde-guanin adduct built in. Then we will evaluate the replication block and mutations of these adducts in E coli cell. The E coli system with a ss-DNA vector (Figure 1.) will provide clear and clean evidence to demonstrate the biological consequences of these adducts in cell without complication from potential repair systems. After the initial biological outcomes are identified[22-23]. Because the adducts are relatively stable within 7 day, we developed a method that allows the genome construction and vector transfection steps in the following sections to be carried out within 7 days.[24]

Mutagenicity and replication block of cinnamaldehyde-DNA adducts in cells. Upon successful vector creation, genomes with a single adduct are integrated into cells, allowed to multiply, and the offspring are subsequently isolated. A decrease in the relative yield of offspring signifies a replication obstruction caused by the adduct, ascertained through the competitive replication of the adduct bypass assay. We will also apply the restriction endonuclease and post-labeling assay to identify the kind, quantity, and genetic prerequisites for the mutagenesis provoked by the adduct. These methodologies will facilitate the assessment of the genotoxic and mutagenic potentials of diverse adducts that develop in the genomes of cells exposed to DNA damaging agents. For assays involving E. coli, HK82 cells will be utilized.

Assay for detecting replication block. This test evaluates replication hindrance by introducing a mix of genomes that comprises adduct to competitor in 30:1 ratio via transfection. PCR amplification of the offspring DNA using CRAB primers signifies whether the 30:1 ratio remains unaltered, suggesting that the adduct does not obstruct the DNA polymerase during replication. If the adduct poses a barrier to replication, the competitor will prevail, since the adduct imposes a competitive growth disadvantage on its genome when the transfection mixture is propagated in a solution. The PCR product of the non-adduct-bearing competitor contains 3 additional nucleotides compared to the adduct-bearing PCR product, which is discernible when the PCR products are cut with XhoI, and subsequently examined using HPLC-MS[26-27].

Assay for detecting mutagenicity. This procedure determines the mutation rate and the type of mutation after the adduct of interest has been processed by the cellular replication machinery. It will be applied to all adducts present in cells. The method employs PCR to amplify the region that was home to the adduct.[28] This allows the collection of mutation information from adducts replicated in virtually any E. coli or mammalian vector, as well as from a whole chromosome. By using a specific primer set, only the genome that contained the adduct is amplified; this 17mer signal band is then scrutinized by LC-MS to provide the percentage of A, G, T, and C. The mutation patterns of the associated adduct can be computed from the ratio of these four nucleotides.

RESULTS

Initially, to investigate whether cinnamaldehyde adduct could be formed in DNA, 13mer dG-containing oligonucleotide (5'-TTTTTTGTTTTTT-3') was synthesized; and the 13mer was eluted at 32.75 min in HPLC analysis (Figure 2c) and showed 1304.2095 m/z in high resolution MD analysis (Figure 2a, S1). Then, the 13mer starting material was incubated in CA solution of water/DMSO (1:1) at 37 ºC for 48 hours. In the following LC-MS analysis, a new group of peak with an m/z for a monoisotopic peak at 1348.2265 at –3 charge state was observed, consistent with the added molec ular weight of cinnamaldehyde (Figures 3b, S2). Based upon the retention time on HPLC chromatograms using a reverse-phase C18 column, the CA-modified products were eluted (around 38 min) after the 13mer unmodified starting material, suggesting that the product was less polar than the starting material. However, a relatively wide HPLC signal of the product consisted of two unresolved peaks (38.11 and 38.68 min), indicating there were possibly two isomers (Figure 3d).

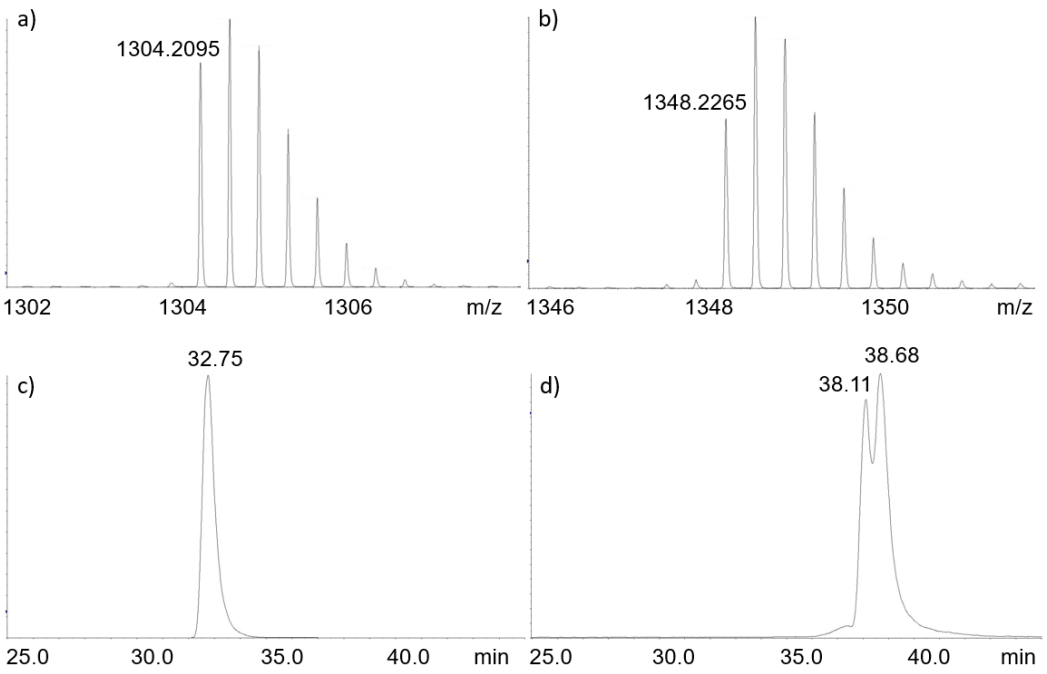


Figure 3. High resolution ESI-TOF LC-MS analyses (a and b) and HPLC chromatograms (c and d) of starting material and CA-DNA products. The observed *m/z* values represent the oligonucleotides under their -3 charge state. a) Peak envelope of 13mer starting material (5'-TTTTTTGTTTTTT-3'), calculated monoisotopic peak: 1304.2086 (-3); b) Peak envelope of 13mer CA-modified oligonucleotide (5'-TTTTTTG*TTTTTT-3', G* = CA modified guanine adduct), calculated monoisotopic peak: 1348.2278 (-3); c) HPLC profile of 13mer starting material; d) HPLC profile of CA-modified oligonucleotide.

To further confirm the location of the modified base, experiments of exonuclease digestion with MALDI-TOF analyses of CA-modified oligonucleotide were conducted. MALDI-TOF has been widely used to identify adduct location by determining mass changes during the time-course digestion.[21–23] Two exonuclease

enzymes, bovine spleen phosphodiesterase (BSP) and snake venom phosphodiesterase (SVP), were used to digest the modified oligonucleotide from 5' to 3' and 3' to 5', respectively. During exonuclease digestion, the lower masses corresponded to the 13 mer to 7 mer fragments with the same signature peak of 2225.6 Da, a finding that indicates that the cinnamaldehyde reacted in a site-specific manner with the guanine base. Both reactions were terminated by 7 mer fragment, but neither generated shorter oligonucleotide fragments (Figures 4 and Figure 5). One possible explanation for these observations is that the cinnamaldehyde-modified guanosine blocks BSP and SVP enzyme activity due to binding hindrance.

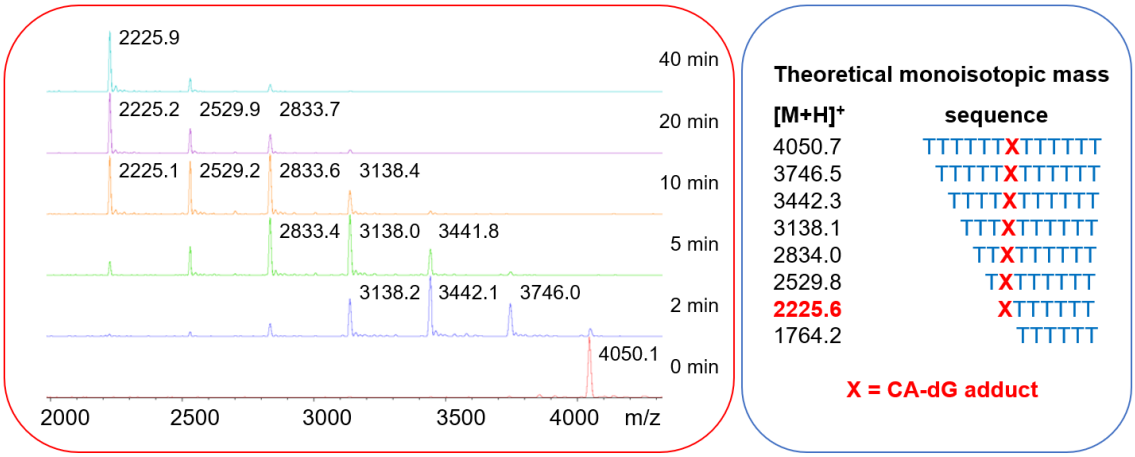


Figure 4. Time-course MALDI-TOF analysis of BSP digestion products of the 13 mer cinnamaldehyde-modified oligonucleotide. Theoretical masses are listed in the inset. The theoretical monoisotopic mass at 2225.6 is highlighted in red because it is the smallest digestion product containing the modification.

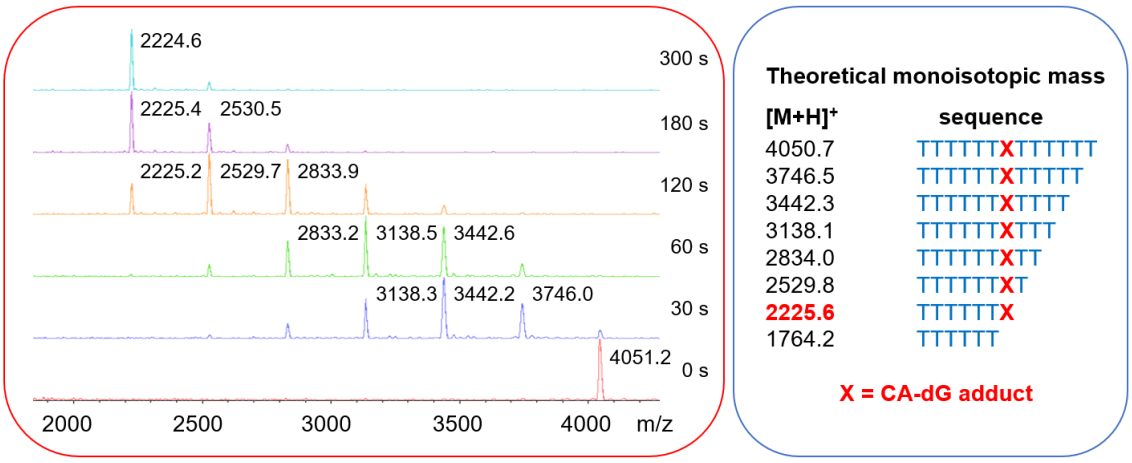
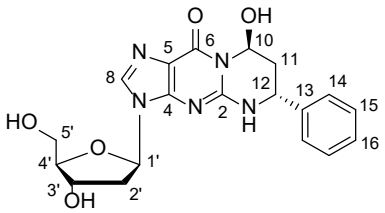


Figure 5. Time-course MALDI-TOF analysis of VSP digestion products of the 13 mer cinnamaldehyde-modified oligonucleotide. Theoretical masses are listed in the inset. The theoretical monoisotopic mass at 2224.6 is highlighted in red because it is the smallest digestion product containing the modification.

In order to study the structure of cinnamaldehyde-induced 2'-dG adduct in the oligonucleotide, CA was used to react with 2'-deoxyguanosine nucleoside to generate adducts. The ^{1}H and ^{13}C data of deoxyguanosine and synthetic **2** are shown in Table 1 (Figures S3-4, S8-9).

Table 1. ^{1}H and ^{13}C NMR assignments for compound **2** in DMSO-d_6



(10*S*, 12*R*) or (10*R*, 12*S*)-**2**

Position	δ_H	Multiplicity (*J* in Hz)	δ_C
1’	6.26	t (6.9)	84.1
2’	2.64-2.71	m	39.9
	2.33-2.37	m	
3’	4.49-4.54	m	71.4 (71.3)
4’	3.94-3.99	m	87.8 (87.8)
5’	3.77	dd (12.0, 3.6)	62.0 (62.0)
	3.70	ddd (12.0, 5.8, 4.0)	
2			151.2 (151.2)
4			149.9
5			115.9
6			156.6
8	7.95	s	136.7
10	6.40	t (2.5)	70.4
11	2.27-2.33	ddd (13.5, 4.5, 2.5)	36.3 (36.3)
	1.88-1.96	dddd (13.5, 12.4, 4.5, 2.5)	
12	4.85-4.89	m	49.6 (49.6)
13			141.0 (141.0)
14	7.44	d (7.3)	126.5 (126.4)
15	7.39	t (7.5)	128.5 (128.5)
16	7.31-7.35	m	127.9 (127.9)

The proposed structure **2** was characterized by ^{1}H-^{1}H two-dimensional (2D) correlation spectroscopy (COSY) (Figure S10), heteronuclear single quantum

coherence (HSQC) and ^{1}H-^{13}C 2D heteronuclear multiple-bond correlation (HMBC) NMR studies (Figures S11-12). The three-bond correlation through H-10/C-10/N-1/C-6 was observed in the HMBC spectrum, clearly demonstrating the regioselectivity of cinnamaldehyde (Figure 6).

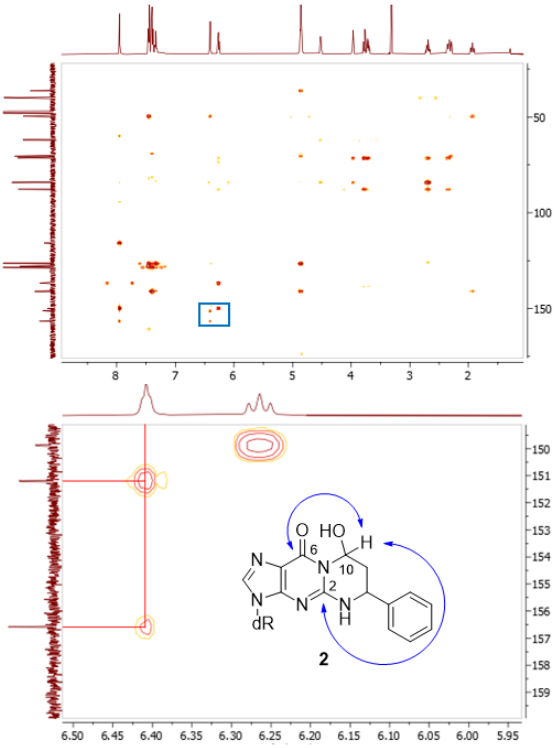


Figure 6. ^{1}H–^{13}C HMBC NMR spectra of compound **2**. The highlighted section (blue box; top) is expanded in the bottom panel.

The CA-dG adduct was identified as a 10-hydroxy-12-phenyl compound, and the possibility of the forming regioisomer, the 12-hydroxy-10-phenyl compound was

ruled out. Compound **2** was tentatively assigned as a 1,N^2-cyclized deoxyguanosine derivative.

Based on these results, there are two chiral carbons introduced at C-10 and C-12. However, the relative configuration of hydroxyl and phenyl group remained unclear because these groups could exist in *cis* or *trans*-configuration. Although only one set of signals was found in the ^{1}H NMR spectrum, a signal consisting of two peaks in the HPLC and two equal peaks for certain carbons in ^{13}C NMR spectrum indicated the presence of a pair of diastereomers. The coupling constants can aid in the identification of the stereo chemistry of these two chiral groups. The geminal coupling constant between two methylene protons at C-11 was approximately 13.5 Hz. The coupling constant of the C-10 proton to either of two C-11 methylene protons was 2.5 Hz, corresponding to an equatorial-equatorial and an equatorial-axial relationship.[12] However, the coupling constant 12.4 Hz was observed for the C-12 proton to the axial proton of C-11 methylene, indicating an axial-axial relationship. These observations are difficult to reconcile with the stereochemical relationships, as the C-10 hydroxyl and the C-12 phenyl should appear in *trans*-configuration. Based on prior work,[24] the diastereomeric peaks at 10.82 min and 11.36 min were assigned as (10*S*, 12*R*)-**2a** and (10*R*, 12*S*)-**2b**, respectively. The structure was further confirmed by mass spectrum analysis, showing a $[M+H]^+$ ion at m/z 400.1616 (400.1615 calculated for $C_{19}H_{22}N_5O_5^+$), which corresponds to an increase of 132 Da (cinnamaldehyde: C_9H_8O) from $[2'\text{-dG}+H]^+$ (Figure S13).

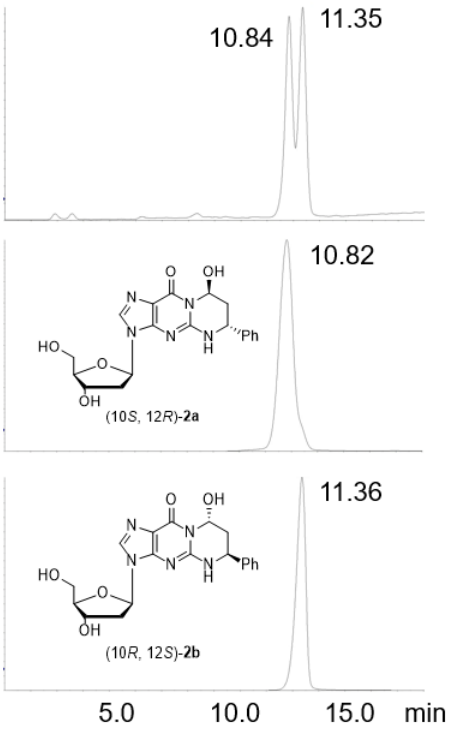


Figure 7. HPLC chromatograms of compound **2**. Top: 13 mer cinnamaldehyde-induced oligonucleotide digested by cocktail enzymes; middle: HPLC of (10*S*, 12*R*)-**2a**; bottom: HPLC of (10*R*, 12*S*)-**2b.**

Subsequently, we applied the same synthesis technique to produce a 13-mer oligonucleotide containing the CA-G adduct. This product was subsequently identified and purified using LC-MS (Figure S1-2). The chromatogram revealed the peak of the starting material at 120.5 minutes, with the first product peak emerging at 146.7 minutes (as depicted in Figure 7.B), and the second product peak appearing at 152.2 minutes (illustrated in Figure 7.C). These three peaks were distinctly separated by this

method, allowing the two target products to be individually collected. After lyophilization, we obtained two pure isomers of the CA-G oligonucleotides.

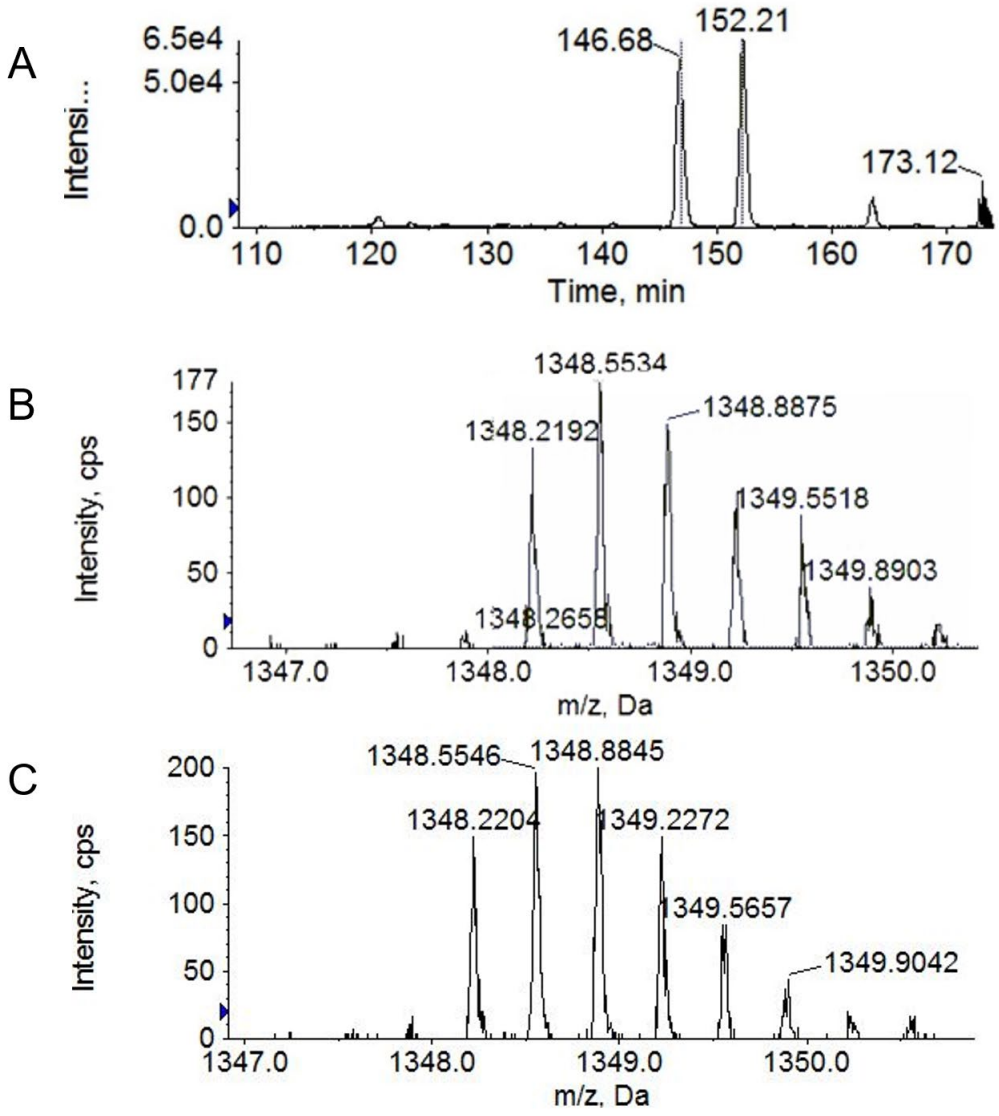


Figure 8. LCMS for the separation of 13mer CA-G containing oligo. A. the extracted ion for 13mer CA-G containing oligo. B. the MS spectrum for the peak at 146.7 min. C. MS spectrum for the peak at 152.2 min.

The stability of the adduct was evaluated, demonstrating that the 13-mer-G-CA adduct remained stable at 37°C for 7 days. Post-incubation, the 13-mer-(10S, 12R)-2a-CA-G adduct (referred to as oligo-2a) exhibited a 5% reversion to the original 13-mer G. Similarly, the 13-mer-(10S, 12R)-2b-CA-G adduct (oligo-2b) showed a 4.5% reversion. Notably, 7.9% of oligo-2a transformed into oligo-2b, and 4.2% of oligo-2b converted into oligo-2a over the same period. Additionally, a product with a mass 2 Da less than the original was observed after 7 days, with 7.4% of oligo-2a and 8.5% of oligo-2b converting to this lower mass species. These observations suggest that the CA adduct is largely stable under the tested conditions, with a small portion of the CA-G adduct reverting to G. This data enhances our understanding of the CA adduct's stability properties.

With the synthesized cinnamaldehyde-2'-dG containing oligonucleotide in hand, we turned our focus to its mutagenic potential in vivo. The 13-mer CA-adduct-containing DNA oligonucleotide was purified and divided into two distinct fractions. Each fraction was then integrated into the M13 viral genome (Figure S14-15). To assess lesion bypass capability within cells, these engineered M13 viral genomes were combined with competitor genomes at a ratio of 30:1 and introduced into E. coli strains via electroporation. Specifically, the E. coli AlkB-deficient strain HK82 was prepared as electrocompetent cells for this purpose. Following transfection, progeny phages were collected and propagated by infecting SCS110 wild-type cells.

Subsequently, M13 genomes were extracted, and the region containing the lesion was amplified via PCR using specific primers. This amplification was followed by a double digestion with XhoI and SphI endonucleases, resulting in DNA fragments of 18 or 20 bases, compared to 23 bases for the control. These fragments were analyzed using LC-ESI-TOF-MS (AB Sciex, ABI4600). The data presented are means and standard deviations from three independent experiments. To assess the efficiency of lesion bypass, we calculated the ratio of intensities between the 18/20mer fragments and the 23mer fragments. This ratio was then normalized against that from a parallel experiment using an unmodified "G" control, set as the benchmark for 100% bypass efficiency. The bypass efficiency for the CA-G adduct was 3.4% for CA-G-**2a** and 4.5% for CA-G-**2b**. These results from the bypass assay demonstrate that the CA-dG lesion significantly blocks DNA replication, irrespective of the isomer involved.

The mutagenicity of two distinct CA adduct lesions was assessed using the previously described methodology. The evaluation of the CA adducts' mutagenicity was based on LCMS results, where the mutation fraction was calculated by dividing the signal of each mutated oligonucleotide by the total signal intensity. Our findings suggest that the mutation spectra of the two isomers are not significantly different. Approximately 52% of the CA adducts are repaired back to guanine without any mutation. In terms of mutation types, the G to T mutation is the most prevalent, representing about 43% of all mutations observed (see Figure 9).

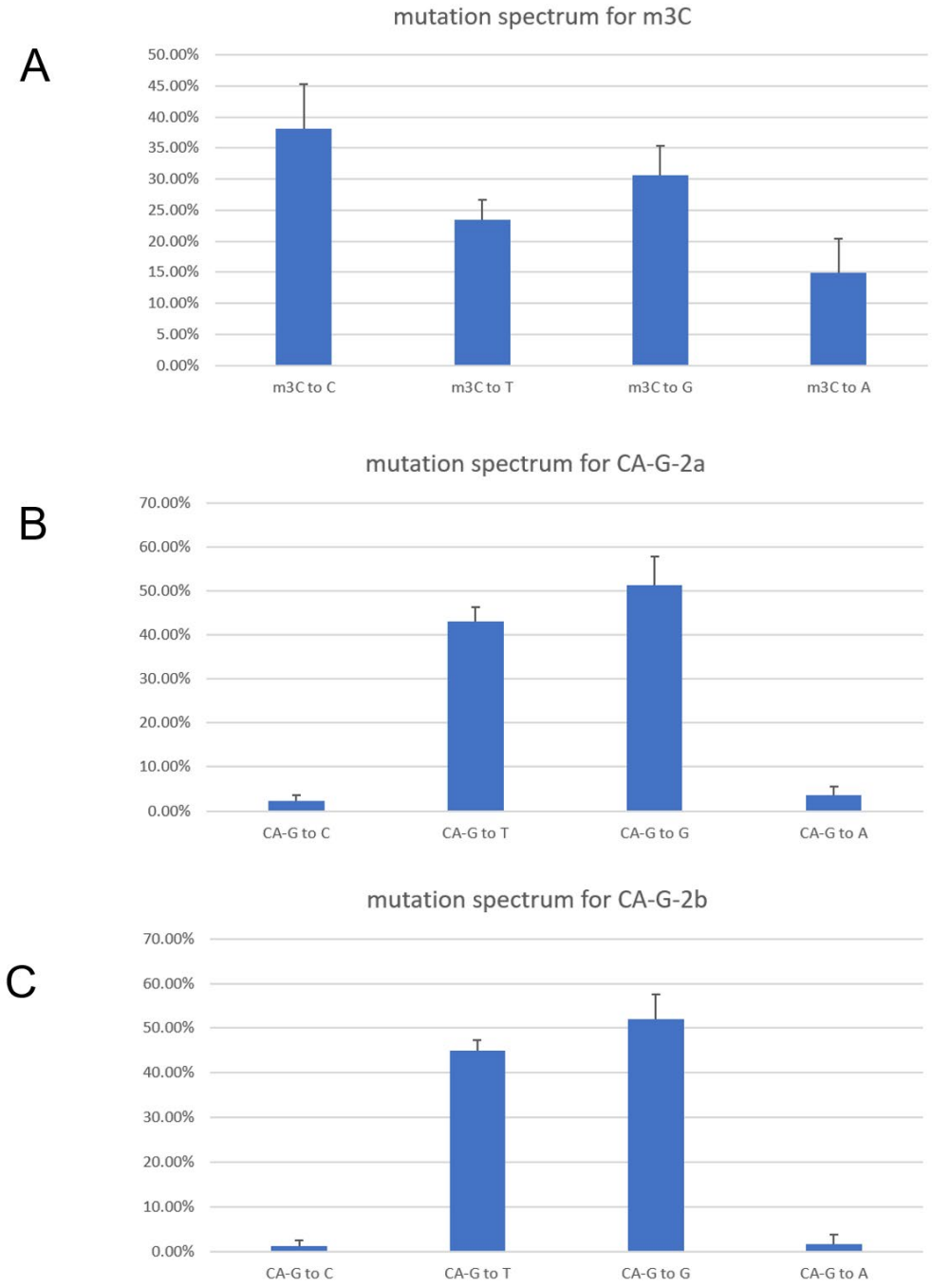


Figure 9. Mutation spectrum of different lesion. CA-G-2a (B) and CA-G-2b (C). m3C (A) used as the positive control.

DISCUSSION

Human exposure to the α,β-unsaturated compounds can occur from exposure to tobacco smoke and automobile exhaust, and from endogenous lipid peroxidation.[26] With α, β-unsaturated double bonds, these compounds can easily penetrate through cell membrane and bind with reactive cellular species. Acrolein and crotonaldehyde can react with deoxyguanosine in DNA to generate 1,N^2-cyclized deoxyguanosine adducts, such as PdG, α-OH-PdG, and γ-OH-PdG. γ-OH-PdG is more mutagenic than α-OH-PdG and predominantly induces G to T transversion.[27] In addition, these deoxyguanosine DNA adducts can be trapped by protein residues (e.g., cysteine) to generate peptide- and protein-DNA conjugates.[28,29] Similar to acrolein and crotonaldehyde, CA contains the same α, β-unsaturated bond but with a phenyl group. Even low concentrations of CA (1.7×10^{-5} M or 2.2 mg/mL) would be toxic in live cell imaging and comet assays and would show a significant decrease in cell viability in the MTT assay.[3] Interestingly, genotoxicity was not found in cells exposed to e-cigarette aerosols that were not cinnamon flavored.[30]

In conclusion, our research demonstrates that cinnamaldehyde (CA) can induce the formation of 1, N^2-cyclized deoxyguanosine adducts in oligonucleotides under physiological conditions. Structural characterization of these adducts, confirmed through NMR spectroscopy and HPLC analyses, reveals the presence of diastereomers

in an approximate 1:1 ratio. Notably, the hydroxyl group at the C-10 position and the phenyl group at C-12 are positioned in a trans configuration. Furthermore, in vivo mutagenicity assays show that these adducts can impede replication and induce mutations in E. coli. Future research should focus on examining the effects of cinnamaldehyde adducts on mutagenicity and replication bypass in human cells to further clarify their biochemical properties and biological consequences.

REFERENCES

1. Bhatnagar, A. *et al.* Electronic Cigarettes: A Policy Statement From the American Heart Association. *Circulation* **130**, 1418–1436 (2014).

2. Trtchounian, A., Williams, M. & Talbot, P. Conventional and electronic cigarettes (e-cigarettes) have different smoking characteristics. *Nicotine Tob. Res.* **12**, 905–912 (2010).

3. Behar, R. Z. *et al.* Distribution, quantification and toxicity of cinnamaldehyde in electronic cigarette refill fluids and aerosols. *Tob. Control* **25**, ii94–ii102 (2016).

4. Behar, R. Z. *et al.* Identification of toxicants in cinnamon-flavored electronic cigarette refill fluids. *Toxicol. In Vitro* **28**, 198–208 (2014).

5. Clapp, P. W. *et al.* Flavored e-cigarette liquids and cinnamaldehyde impair respiratory innate immune cell function. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **313**, L278–L292 (2017).

6. Clapp, P. W. *et al.* Cinnamaldehyde in flavored e-cigarette liquids temporarily suppresses bronchial epithelial cell ciliary motility by dysregulation of mitochondrial function. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **316**, L470–L486 (2019).

7. King, A. A. *et al.* Antimutagenicity of cinnamaldehyde and vanillin in human cells: Global gene expression and possible role of DNA damage and repair. *Mutat. Res. Mol. Mech. Mutagen.* **616**, 60–69 (2007).

8. Neudecker, T. The genetic toxicology of cinnamaldehyde. *Mutat. Res. Genet. Toxicol.* **277**, 173–185 (1992).

9. Kiwamoto, R., Ploeg, D., Rietjens, I. M. C. M. & Punt, A. Dose-dependent DNA adduct formation by cinnamaldehyde and other food-borne α,β-unsaturated aldehydes predicted by physiologically based in silico modelling. *Toxicol. In Vitro* **31**, 114–125 (2016).

10. Janzowski, C. *et al.* Unsaturated carbonyl compounds: induction of oxidative DNA damage in mammalian cells. *Mutagenesis* **18**, 465–470 (2003).

11. Hecht, S. S., Lin, Dorothy., Chuang, Julia. & Castonguay, Andre. A study of chemical carcinogenesis. 91. Reactions with deoxyguanosine of 4-

(carbethoxynitrosamino)-1-(3-pyridyl)-1-butanone, a model compound for .alpha.-hydroxylation of tobacco-specific nitrosamines. *J. Am. Chem. Soc.* **108**, 1292–1295 (1986).

12. Eder, E., Hoffman, C., Bastian, H., Deininger, C. & Scheckenbach, S. Molecular mechanisms of DNA damage initiated by alpha, beta-unsaturated carbonyl compounds as criteria for genotoxicity and mutagenicity. *Environ. Health Perspect.* **88**, 99–106 (1990).

13. Cheng, G. *et al.* Reactions of Formaldehyde Plus Acetaldehyde with Deoxyguanosine and DNA: Formation of Cyclic Deoxyguanosine Adducts and Formaldehyde Cross-Links. *Chem. Res. Toxicol.* **16**, 145–152 (2003).

14. Chung, F. L., Roy, K. & Hecht, S. S. Study of reactions of .alpha.,.beta.-unsaturated carbonyl compounds with deoxyguanosine. *J. Org. Chem.* **53**, 14–17 (1988).

15. Maekawa, M. *et al.* Identification of 4-Oxo-2-hexenal and Other Direct Mutagens Formed in Model Lipid Peroxidation Reactions as dGuo Adducts. *Chem. Res. Toxicol.* **19**, 130–138 (2006).

16. Zhang, X.-Y. & Elfarra, A. A. Characterization of the Reaction Products of 2‘-Deoxyguanosine and 1,2,3,4-Diepoxybutane after Acid Hydrolysis: Formation of Novel Guanine and Pyrimidine Adducts. *Chem. Res. Toxicol.* **17**, 521–528 (2004).

17. Huang, Y. & Johnson, F. Regioisomeric synthesis and characteristics of the alpha-hydroxy-1,N(2)-propanodeoxyguanosine. *Chem. Res. Toxicol.* **15**, 236–239 (2002).

18. Zhu, R. *et al.* Cinnamaldehyde in diabetes: A review of pharmacology, pharmacokinetics and safety. *Pharmacol. Res.* **122**, 78–89 (2017).

19. Hajinejad, M., Ghaddaripouri, M., Dabzadeh, M., Forouzanfar, F. & Sahab-Negah, S. Natural Cinnamaldehyde and Its Derivatives Ameliorate Neuroinflammatory Pathways in Neurodegenerative Diseases. *BioMed Res. Int.* **2020**, 1–9 (2020).

20. Ma, B., Stepanov, I. & Hecht, S. Recent Studies on DNA Adducts Resulting from Human Exposure to Tobacco Smoke. *Toxics* **7**, 16 (2019).

21. Jain, V. *et al.* Structural and thermodynamic insight into Escherichia coli UvrABC-mediated incision of cluster diacetylaminofluorene adducts on the NarI sequence. *Chem. Res. Toxicol.* **26**, 1251–1262 (2013).

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

23. Zhang, L. K. & Gross, M. L. Matrix-assisted laser desorption/ionization mass spectrometry methods for oligodeoxynucleotides: improvements in matrix, detection

limits, quantification, and sequencing. *J. Am. Soc. Mass Spectrom.* **11**, 854–865 (2000).

24. Rezaei, M., Harris, T. M. & Rizzo, C. J. Stereoselective synthesis of the 1,N2-deoxyguanosine adducts of cinnamaldehyde. A stereocontrolled route to deoxyguanosine adducts of α,β-unsaturated aldehydes. *Tetrahedron Lett.* **44**, 7513–7516 (2003).

25. Quinlivan, E. P. & Gregory, J. F. DNA digestion to deoxyribonucleoside: a simplified one-step procedure. *Anal. Biochem.* **373**, 383–385 (2008).

26. Facchinetti, F. *et al.* α,β-Unsaturated Aldehydes in Cigarette Smoke Release Inflammatory Mediators from Human Macrophages. *Am. J. Respir. Cell Mol. Biol.* **37**, 617–623 (2007).

27. Liu, X., Zhu, M. & Xie, J. Mutagenicity of acrolein and acrolein-induced DNA adducts. *Toxicol. Mech. Methods* **20**, 36–44 (2010).

28. Kurtz, A. J. & Lloyd, R. S. 1,N 2-Deoxyguanosine Adducts of Acrolein, Crotonaldehyde, and trans-4-Hydroxynonenal Cross-link to Peptides via Schiff Base Linkage. *J. Biol. Chem.* **278**, 5970–5976 (2003).

29. Sanchez, A. M. *et al.* Comparative Evaluation of the Bioreactivity and Mutagenic Spectra of Acrolein-Derived α-HOPdG and γ-HOPdG Regioisomeric Deoxyguanosine Adducts. *Chem. Res. Toxicol.* **16**, 1019–1028 (2003).

30. Misra, M., Leverette, R., Cooper, B., Bennett, M. & Brown, S. Comparative In Vitro Toxicity Profile of Electronic and Tobacco Cigarettes, Smokeless Tobacco and Nicotine Replacement Therapy Products: E-Liquids, Extracts and Collected Aerosols. *Int. J. Environ. Res. Public. Health* **11**, 11325–11347 (2014).

31. Tang, Q. *et al.* Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega* **2**, 8205–8212 (2017).

SUPPLEMENTARY INFORMATION

Synthesis and characterization of cinnamaldehyde-induced 1,N^2-cyclized deoxyguanosine adduct in DNA

Jie Wang, Xianhao Zhou 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;

TABLE OF CONTENTS

Figure S13. MS analysis of CA-dG adduct under positive ion mode.

Figure S14. Agarose gel for m13 plasmid linearization.

Figure S15. Agarose gel for the verification of constructed m13-lesion plasmid.

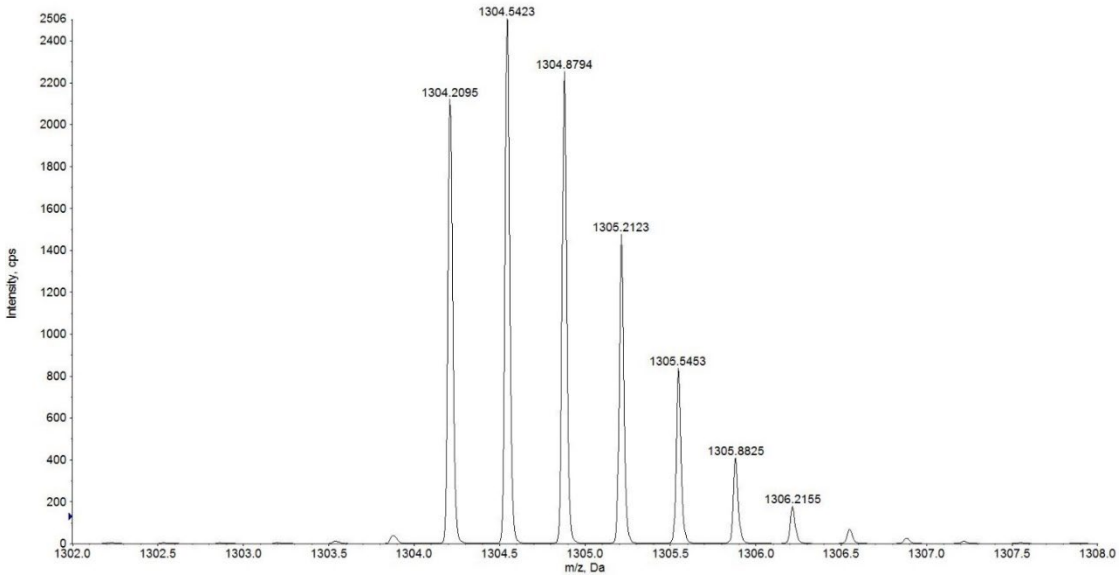


Figure S1. High resolution LC-MS analysis of 13mer oligonucleotide starting material (5'-TTTTTTGTTTTTT-3'), calculated: 1304.2086 (-3 charge), observed: 1304.2095 (-3 charge).

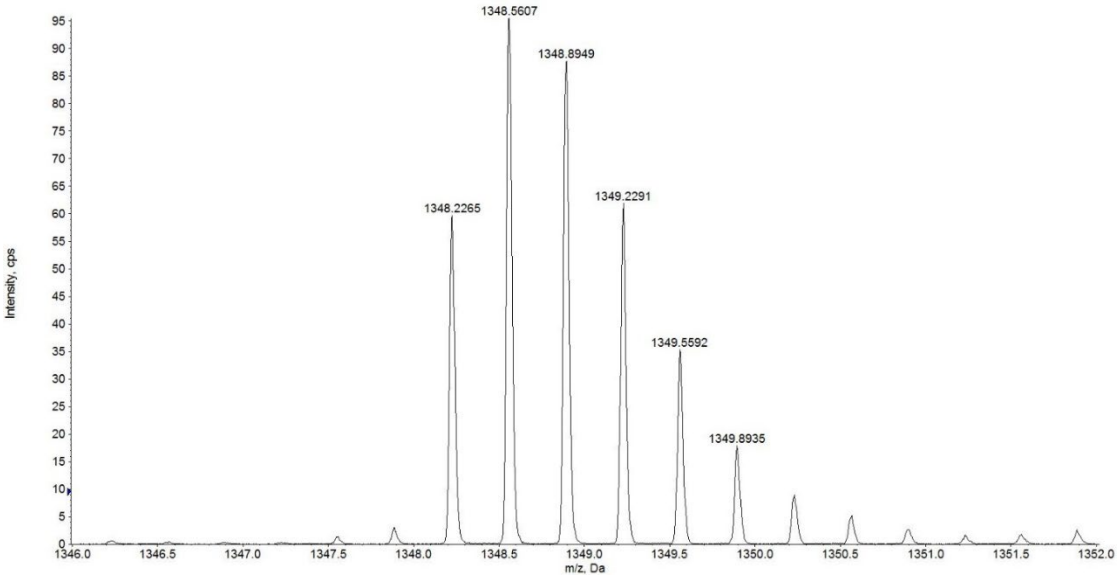


Figure S2. High resolution LC-MS analysis of 13mer CA-modified oligonucleotide, calculated: 1348.2278 (-3 charge), observed: 1348.2265 (-3 charge).

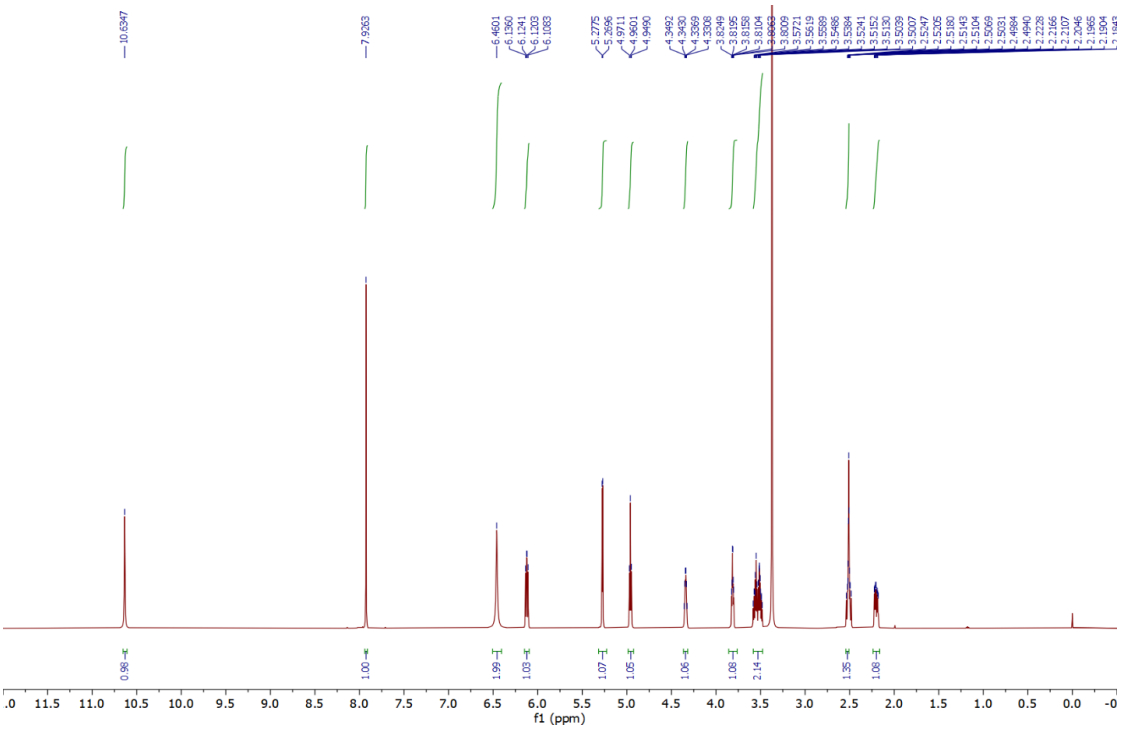


Figure S3. ^{1}H NMR spectrum of deoxyguanosine in DMSO-d_6.

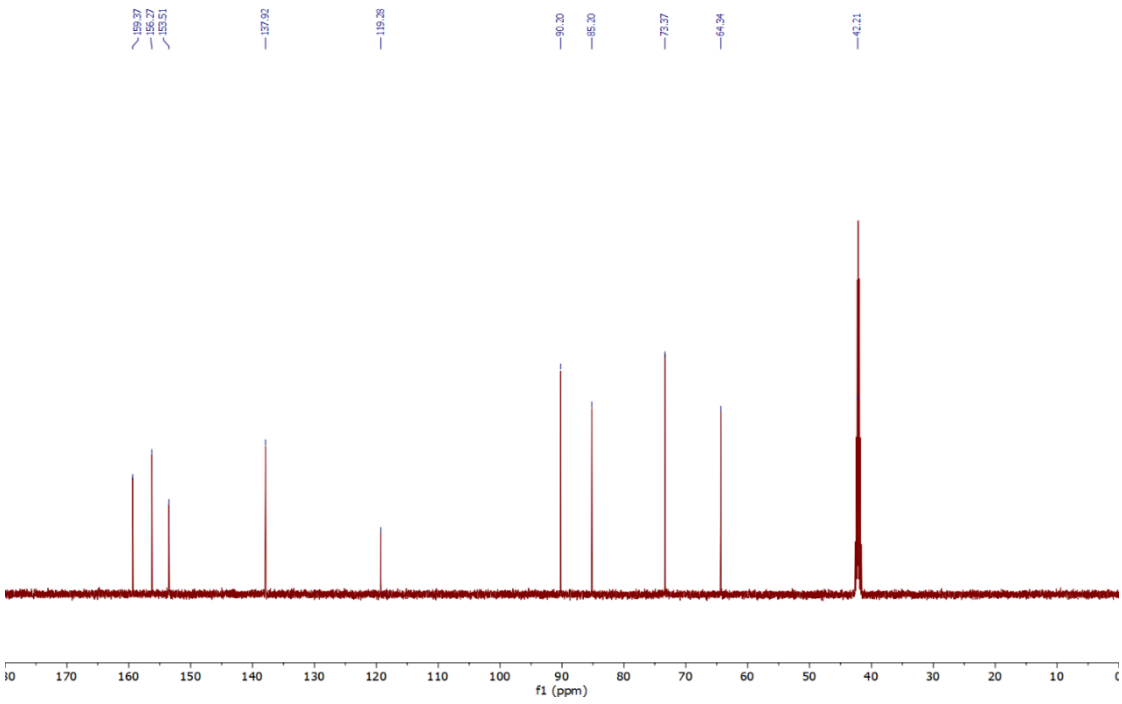


Figure S4. ^{13}C NMR spectrum of deoxyguanosine in DMSO-d_6.

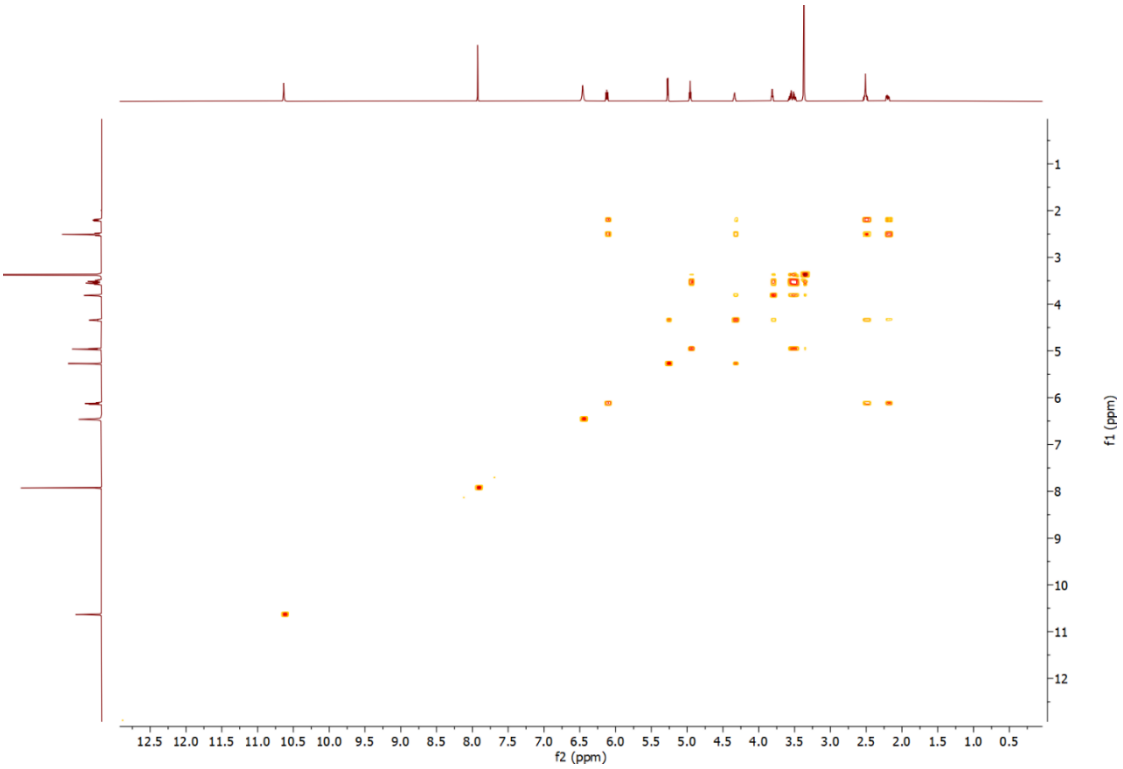


Figure S5. ^{1}H-^{1}H COSY NMR spectrum of deoxyguanosine in DMSO-d_6.

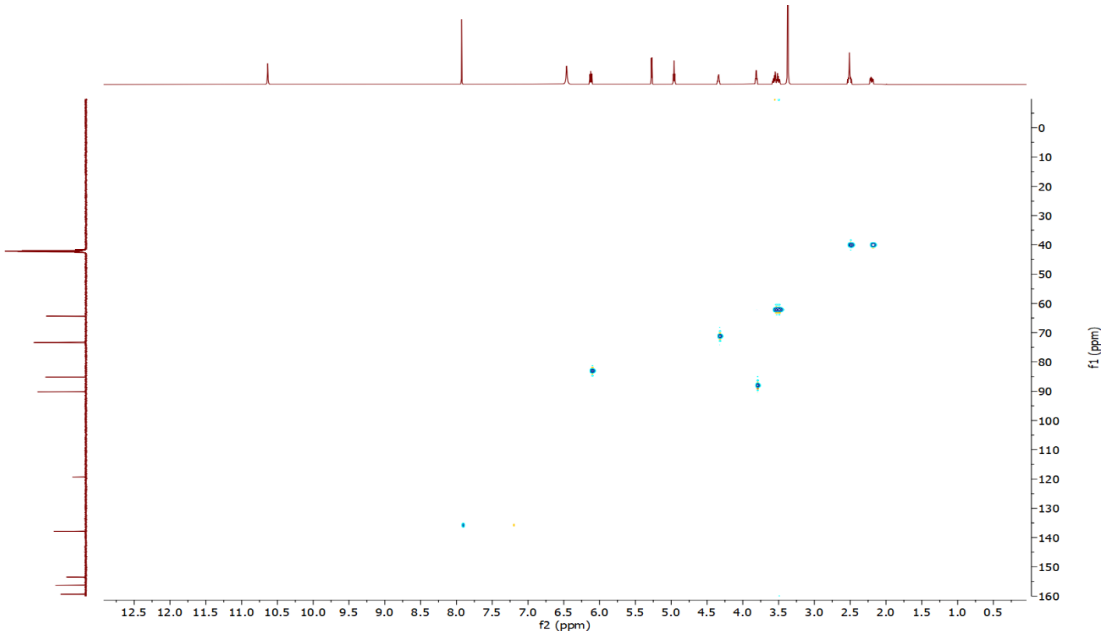


Figure S6. ^{1}H-^{13}C HSQC NMR spectrum of deoxyguanosine in DMSO-d_6.

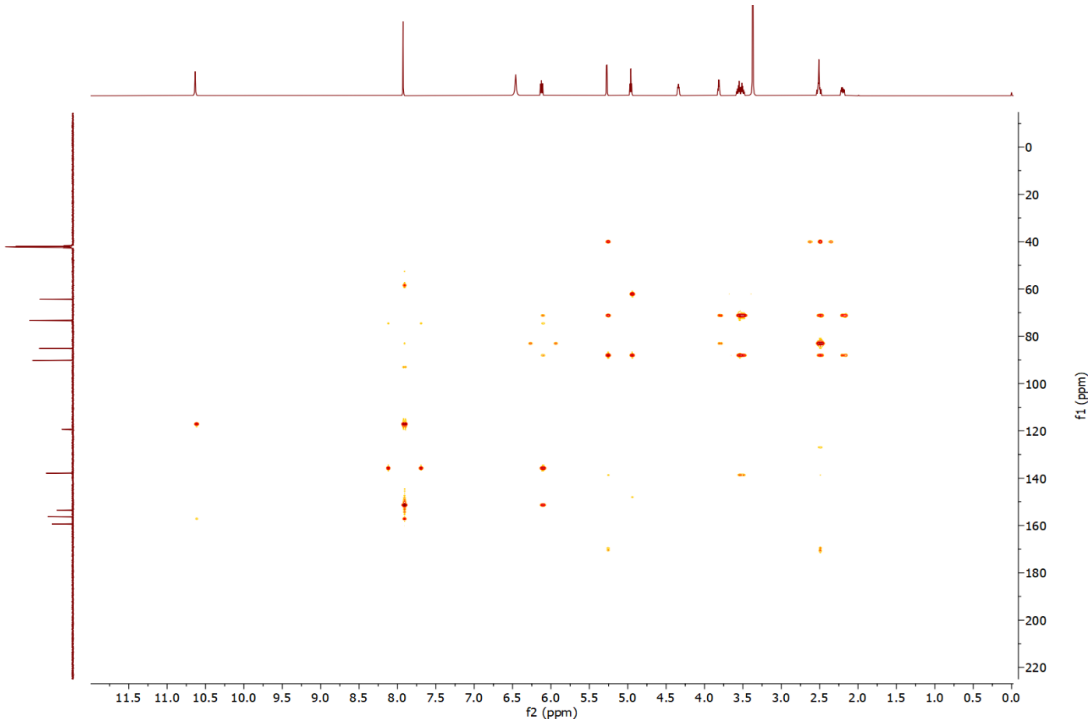


Figure S7. ^{1}H-^{13}C HMBC NMR spectrum of deoxyguanosine in DMSO-d_6.

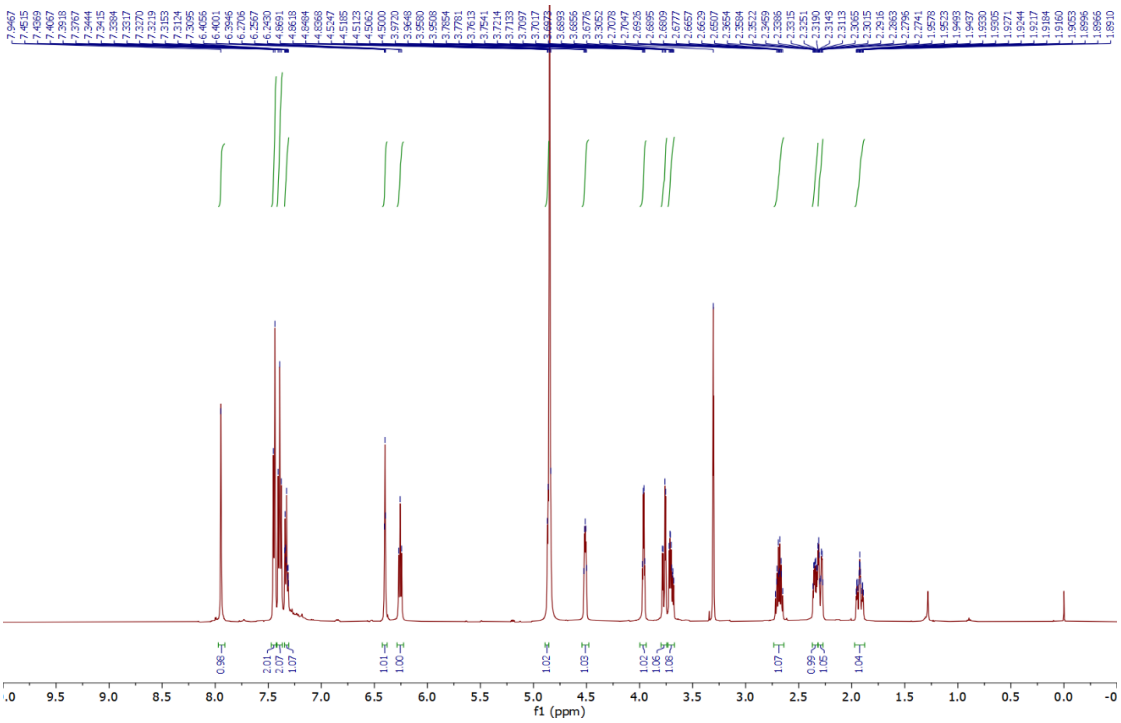


Figure S8. ^{1}H NMR spectrum of CA-dG adducts in CD_3OD.

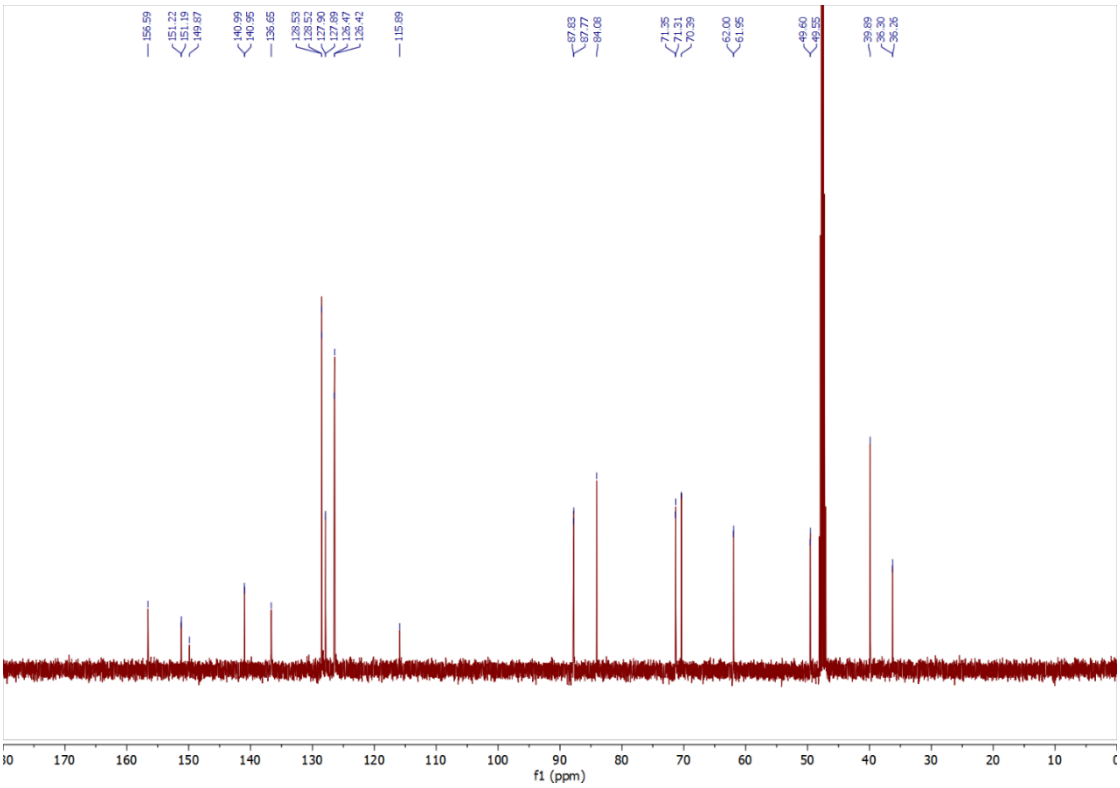


Figure S9. ^{13}C NMR spectrum of CA-dG adducts in CD_3OD.

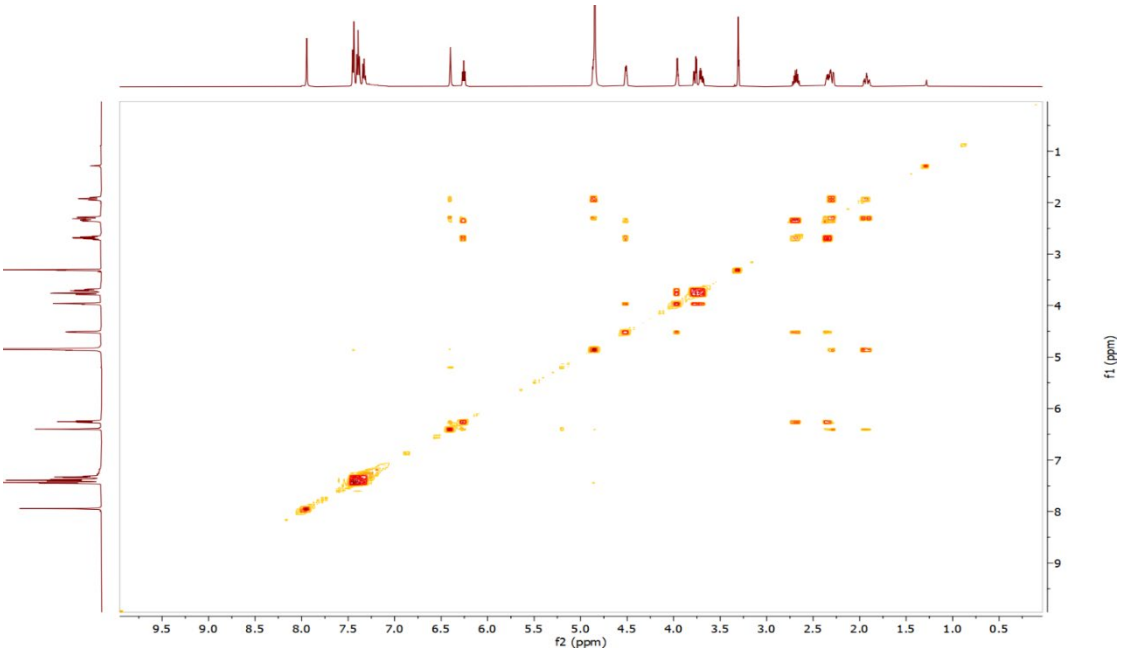


Figure S10. ^{1}H-^{1}HCOSY NMR spectrum of CA-dG adducts in CD_3OD.

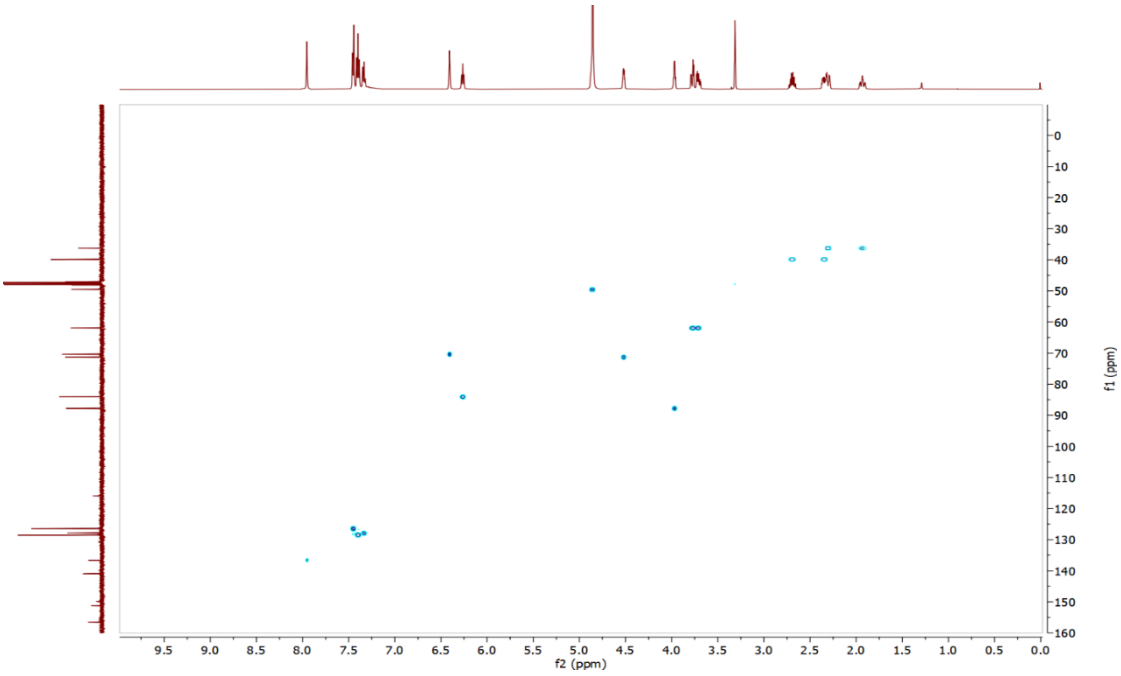


Figure S11. ^{1}H-^{13}C HSQC NMR spectrum of CA-dG adducts in CD_3OD.

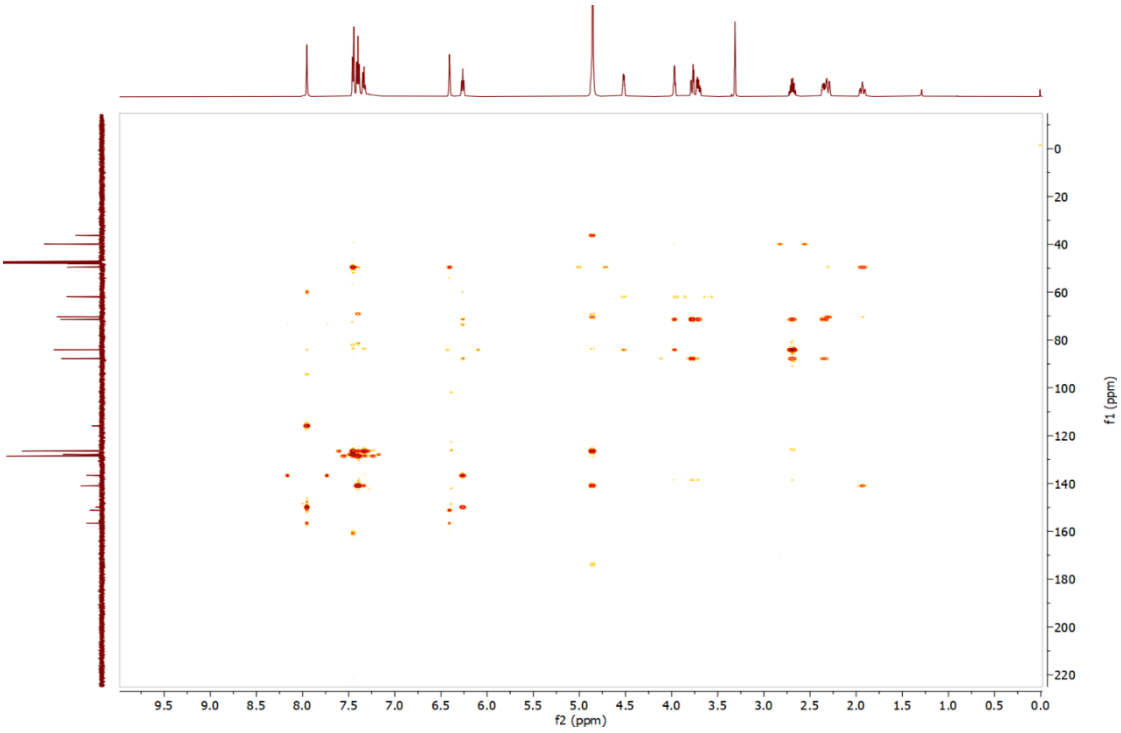


Figure S12. ^{1}H-^{13}C HMBC NMR spectrum of CA-dG adducts in CD_3OD.

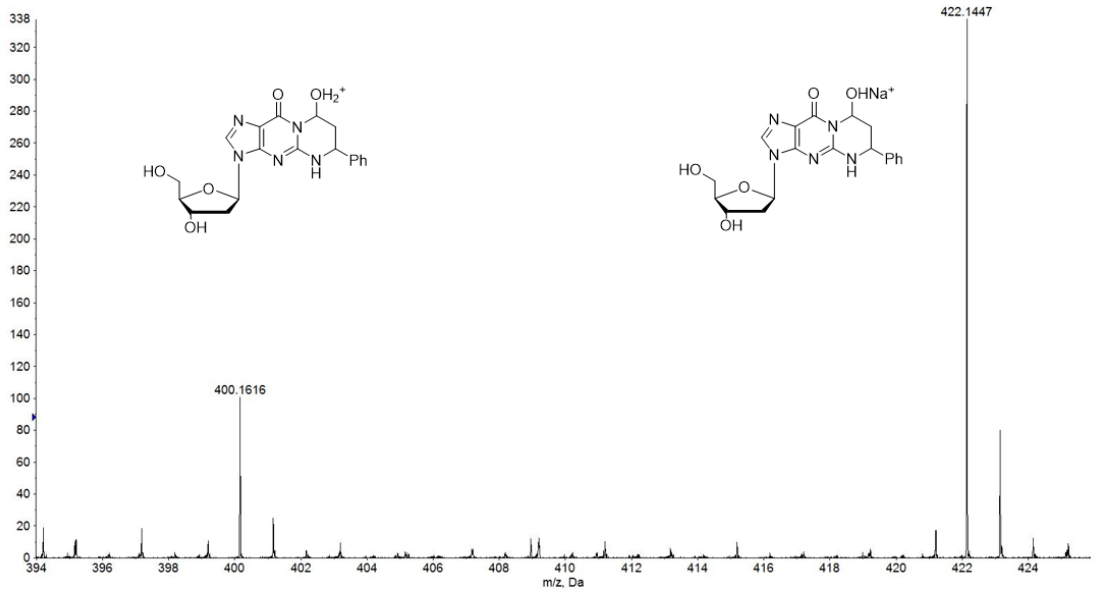
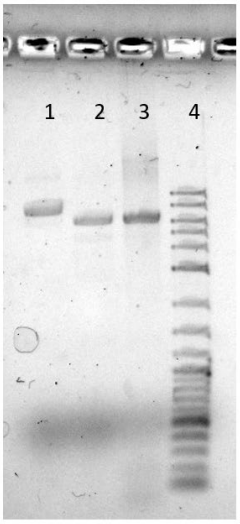
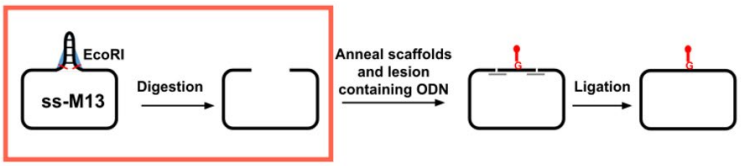


Figure S13. MS analysis of CA-dG adduct under positive ion mode, calculated for $[M+H]^+$: 400.1616; observed: 400.1616; calculated for $[M+Na]^+$: 422.1435; observed: 422.1447.





1	M13
2	linearized M13
3	ligation reaction mixture
4	ladder

Figure S14. Agarose gel for m13 plasmid linearization

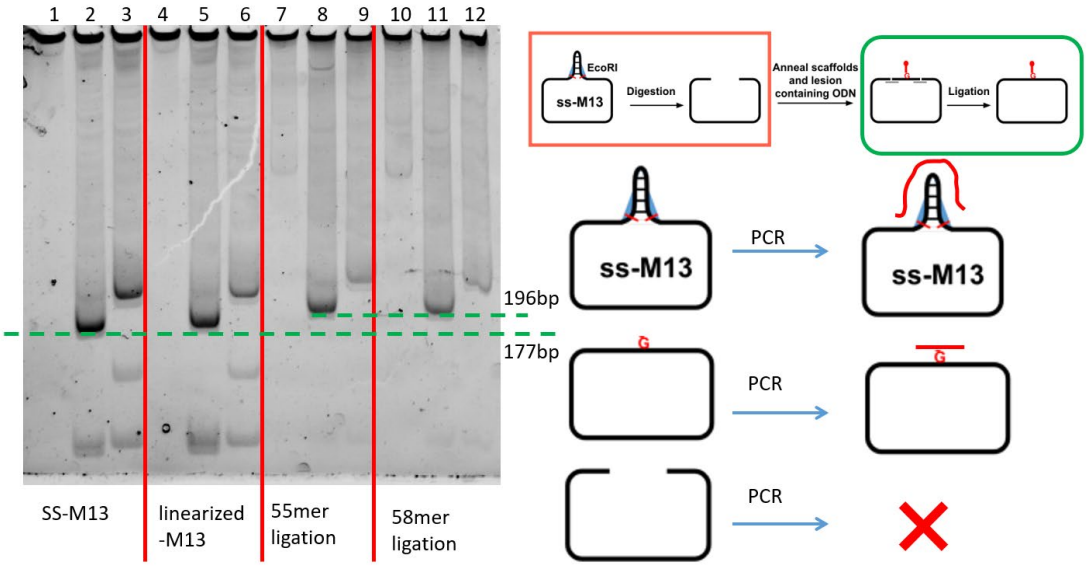


Figure S15. Agarose gel for the verification of constructed m13-lesion plasmid. PCR was utilized as the method for verify the constructed plasmid.

CHAPTER - II

Prepared for submission to Chemical Research in Toxicology.

Study on Correlation of DNA Damage Repair Efficiency Obtained in vitro and in vivo methods

Xianhao Zhou[†], Jian Ma[†], Samuel Howarth[†], Yi-Tzai Chen[†], and Deyu Li[†,*]

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

* Correspondence: Deyu Li (deyuli@uri.edu)

ABSTRACT

Methylation patterns in nucleic acids are essential for genetic integrity and epigenetic regulation, involving the complex addition and removal of methyl groups. Understanding demethylation mechanisms in vivo is crucial for genetic and epigenetic research, although studying demethylation activity within living systems poses significant challenges. This chapter focuses on oxidative demethylation by the 2-oxoglutarate (2-OG) and Fe(II)-dependent AlkB family of enzymes, specifically the variant from Escherichia coli. Although the AlkB enzyme's activity towards certain substrates has been documented in vivo, its interaction with other specific substrates remains less understood. Our research focuses on analyzing the kinetic behavior of AlkB enzymes when interacting with different alkylated substrates. We aim to determine kinetic parameters using in vitro assays and will complement these results with in vivo data from published studies. Our goal is to identify correlations that enhance our understanding of AlkB's demethylation activity in biological systems. This holistic approach is intended to enrich our knowledge of the involvement of AlkB enzymes in genetic and epigenetic modifications.

INTRODUCTION

The AlkB enzyme is an integral part of the DNA repair machinery in E. coli, with a well-established role in mitigating alkylation damage. [1-3] It operates as an Fe(II)- and α-ketoglutarate-dependent dioxygenase, reversing DNA damage through a direct reversal mechanism. Many different DNA adduct had been studied and reported. [4-8]

1-Methyladenine (m1A) is a DNA lesion caused by alkylating agents like methyl methanesulfonate (MMS), primarily affecting single-stranded DNA where the adenine's N1 atom is exposed. Although m1A is cytotoxic and disrupts DNA replication, it is only slightly mutagenic. This instability results from the Dimroth rearrangement, where the methyl group shifts from N1 to N6 on adenine. The AlkB enzyme counteract this damage through an oxidative mechanism that removes the methyl group as formaldehyde, essential for maintaining DNA replication and preventing genotoxicity. [9-13]

1-Methylguanine (m1G) is a common DNA adduct in both lab settings and living organisms, significantly hindering DNA replication. However, the repair enzyme AlkB effectively reverses m1G damage, reducing its replication-blocking and mutagenic effects. In the absence of AlkB, m1G-induced mutations are high, but with AlkB, lesion bypass increases eightfold, and mutagenicity drops from 80% to 4%. The primary mutations from m1G are G → T, G → A, and G → C, with G → T being most frequent. Additionally, the SOS response activates bypass polymerases, which

surprisingly show an anti-mutagenic effect on m1G, highlighting the intricate relationship between DNA repair and replication fidelity. [3, 17]

3-Methylcytosine (m3C) is a DNA adduct formed primarily in single-stranded DNA by SN2 alkylating agents like MMS. It's more common in single-stranded DNA because the N3 position of cytosine is more exposed. m3C stalls DNA synthesis, significantly affecting replication. The AlkB enzyme efficiently repairs m3C in vitro and in vivo, reducing its mutagenicity from about 30% to lower rates and primarily preventing C → T and C → A mutations. Without AlkB, SOS bypass polymerases can still replicate past these lesions but at a higher mutation rate. Studies indicate m3C can block replication by inhibiting DNA polymerase I, despite some bypass that introduces replication errors.[13, 20]

is a significant DNA adduct formed by formaldehyde, which is used industrially and produced naturally through the metabolism of serine, choline, and methionine. In molecular genetics, compounds like formaldehyde target the exocyclic amino group of deoxyguanosine, forming N-alkyl-dG adducts similar to those created by acetaldehyde. For example, acetaldehyde produces N2-ethyl-2′-deoxyguanosine (N2-Et-dG), detectable in human tissues and urine, particularly in individuals who consume alcohol.[23] Building on the structural similarities between acetaldehyde and formaldehyde, it is suggested that formaldehyde leads to the formation of N2-Me-dG adducts in DNA. Research using oligodeoxynucleotides with site-specific N2-Me-dG modifications showed that these adducts could induce G→A mutations during primer

extension experiments with E. coli DNA polymerase I. These results emphasize the miscoding potential of m2G and underscore the genetic risks associated with formaldehyde exposure.[24]

1, N^6-ethenoadenine (eA), a bifunctional DNA lesion, forms endogenously in humans and rodents from adenine reacting with unsaturated lipid peroxidation products and is also induced by vinyl chloride and its metabolite chloroacetaldehyde. Detected in chronically inflamed tissues, eA is linked to cancer promotion and neurological disorders due to oxidative stress. In duplex DNA, eA is mainly repaired by the base excision repair (BER) pathway, particularly through glycosylases like AAG. Recent studies have shown that the AlkB enzyme, and its human homolog ABH3, can reverse this damage in vitro via an iron-mediated reaction with α-ketoglutarate, reducing mutagenicity significantly in AlkB-proficient E. coli compared to deficient strains.[25-28]

3-Methylthymine (m3T) is a DNA lesion generated by the action of SN2 alkylating agents, such as methylmethanesulfonate (MMS), and has been identified in both in vitro and in vivo studies.[29-30] It is known to be a particularly strong obstacle to DNA replication, which only marginally improves with the induction of the SOS bypass polymerase system. AlkB has limited effectiveness on m3T, thus it remains a significant block to replication[31]. Remarkably, the fat mass and obesity-associated protein (FTO), identified as a 2-oxoglutarate-dependent nucleic acid demethylase, shows proficiency in repairing m3T in single-stranded DNA but not in double-stranded DNA.[32] In terms of its mutagenic potential, m3T leads to a high mutagenicity

rate of approximately 60% in cells that are deficient in both SOS repair and AlkB, with the mutations primarily resulting in T → A and T → C transitions. In vitro assays further reveal that m3T strongly impedes the Klenow fragment of DNA polymerase I, causing an increase in the incorporation of deoxythymidine triphosphate (dTTP) opposite the adduct on a synthetic template. Interestingly, the replication machinery tends to incorporate thymine opposite 3-ethyldeoxythymidine, although there is conflicting evidence with some studies showing exclusive incorporation of adenine. This variance highlights the nuanced challenges that different alkylated DNA lesions pose to DNA polymerases and underscores the need for diverse repair mechanisms to maintain genomic integrity.[33]

In our study, we quantified the catalytic efficiency (Kcat/Km) of AlkB in repairing various lesions using biochemical reactions and reviewed in vivo data from the literature. Our goal is to identify correlations between in vitro and in vivo findings. Establishing these correlations could allow us to use in vitro data to predict AlkB's repair efficiency in living organisms.

METHOD

Oligonucleotide Synthesis. Oligonucleotides with a length of 16 bases were synthesized, featuring the sequence 5'-GAAGACCTXGGCGTCC-3', where 'X' marks the site of the lesions. Complementary 23-mer oligonucleotides, with the sequence 5′ -CTGGGACGCCYAGGTCTTCACTG-3′ (where 'Y' corresponds to the

conventional bases A, T, C, or G), were acquired from IDT. The synthesis of these DNA sequences was carried out using solid-phase phosphoramidite chemistry on a MerMade-4 Oligonucleotide Synthesizer. Subsequent purification of the oligonucleotides was conducted using HPLC (High-Performance Liquid Chromatography) provided by Thermo Fisher Scientific, utilizing a DNAPac PA-100 Semi-Preparative column from Phenomenex. The solvents used in this process were solvent A, comprising 100 mM 1:1 triethylamine-acetic acid (TEAA) in water, and solvent B, which was pure acetonitrile. The quantification of the DNA concentration was performed using UV absorbance at 260nm, measured by a NanoDrop device. To characterize the oligonucleotides, HPLC-electrospray ionization triple quadrupole time of flight mass spectrometry (AB Sciex) was employed.

Expression and Purification of the AlkB. His-tagged AlkB protein was produced by introducing pET28a-AlkB into E. coli strains Rosetta2(DE3)pLysS and BL21(DE3)pLysS. Protein expression in these cells was triggered by adding 1mM isopropyl-β-D-thiogalactopyranoside (IPTG) at a temperature of 30 °C. Subsequent purification of the expressed protein was achieved through affinity chromatography methods. The His-tag component of the AlkB protein was then cleaved using thrombin. The thoroughly purified protein was subsequently stored at -80 °C in a previously described AlkB storage buffer for preservation.

Enzymatic Reaction. To evaluate the demethylase activity of the AlkB family enzymes on single-stranded and double-stranded DNA substrates, enzymatic assays

were conducted at a constant temperature of 37 °C, with varying time intervals for kinetic analysis. The reaction mixture contained 70.0 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H2O$, 0.93 mM α-KG, 1.86 mM ascorbic acid, and 46.5 mM HEPES, adjusted to pH 8.0. The enzymatic process was halted by introducing 10 mM EDTA and heating the solution to 95 °C for a duration of 5 minutes. In a standard procedure, purified proteins and oligonucleotides with specific DNA adducts were combined in a 20 μL reaction volume, which included all necessary cofactors. For the separation of substrates and products, particularly for 16mer m1A and A or 16mer m3C and C, the HPLC conditions were set to initiate with a 5-minute gradient of 1.5 M ammonium acetate, increasing from 50% to 65%, followed by a 2-minute phase at 70% ammonium acetate. A DNApac PA-100 column (4× 250 mm) from Thermo Scientific was employed for this purpose, with UV detection set at 260 nm. These assays were all replicated three times to ensure reliable results.

Kinetic Studies. To calculate the kinetic parameters kcat/Km for the repair reactions, initial reaction rates were measured by maintaining a consistent enzyme concentration while altering the concentration of the substrate. The reactions were timed at intervals of 0, 0.5, 1, 4, 10, and 30 minutes. Each reaction was conducted at a temperature of 37 °C and replicated three times to ensure reliability and precision in the results.

RESULTS

Oligonucleotide Synthesis and Protein Purification.

16mer oligonucleotides containing adducts were chemically synthesized with the sequence 5'-GAAGACCTXGGCGTCC-3' (X denotes the alkylated base). 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 LCMS (Table S1). The genes for E. coli AlkB was cloned into pET28a+ expression vector; the incorporation of the correct sequences was confirmed by sequencing the corresponding plasmids. The protein was then expressed in E. coli hosts, isolated and purified by affinity chromatography as described in Experimental Section.

Condition optimization for Kinetic research.

The determination of kinetic constants relies upon the initial velocities of enzymatic reactions. To ensure accuracy, it is imperative that these velocities are maintained at a moderate level. Consequently, our objective was to refine the reaction conditions to achieve a conversion rate of 30% to 50% within a 10-minute timeframe. Additionally, the repair ratio serves as an indicator of a substrate's suitability for kinetic analysis, reflecting its adequacy for precise measurement of kinetic parameters.

Table 1. Repair ratio of AlkB with different substrate in 10 min reaction

Lesion	Conc. of enzyme(uM)	Repair ratio
m3C	1	62%
m1A	5	67%

m1G	20	21%
m2G	20	4.4%
eA	2	57%
eG	20	8.7%
eC	20	6.2%
m3T	20	2.5%

Enzymatic Assay for Measuring Kinetic Constants.

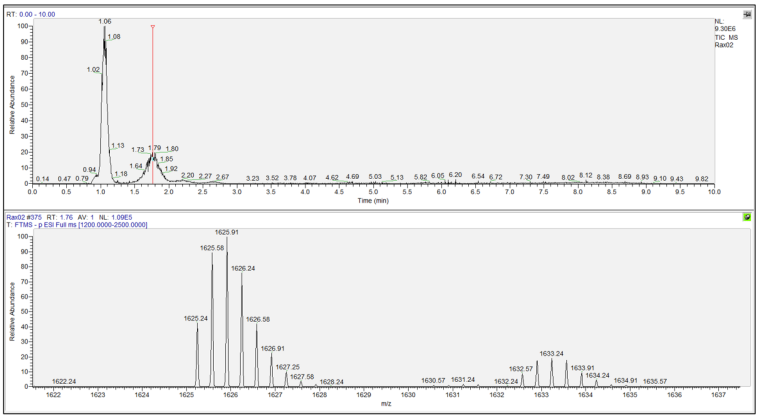
For each enzymatic reaction, the adduct-containing oligonucleotide was incubated with the necessary cofactors for the AlkB reaction: Fe(II), αKG, and ascorbic acid. Below, m3C will be used as an example to explain the LCMS analyses. For the ss-DNA reactions, the starting material 16mer m3C (Figure 1A) and product 16mer C (Figure 1B) were extracted in MS, and the amount of each was quantified by peak area integration.

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 repairing different substrates. In a typical kinetic analysis (e.g., AlkB repairing m1A), 1 μM AlkB enzyme were mixed with different concentrations of oligonucleotide substrate (2-5 μM) and the

extent of the repair reaction was quantified at different time points (see Experimental section for details). The concentrations of the product 16mer A were used to calculate the Kcat and Km. the integrated MS signal in LCMS was used to calculate the repair ratio. the concentration of product was calculated by repair ratio multiply with initial concentration of starting material. 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. All Kcat/Km data are shown in Supplementary information Figure S1 to S6.

A



B

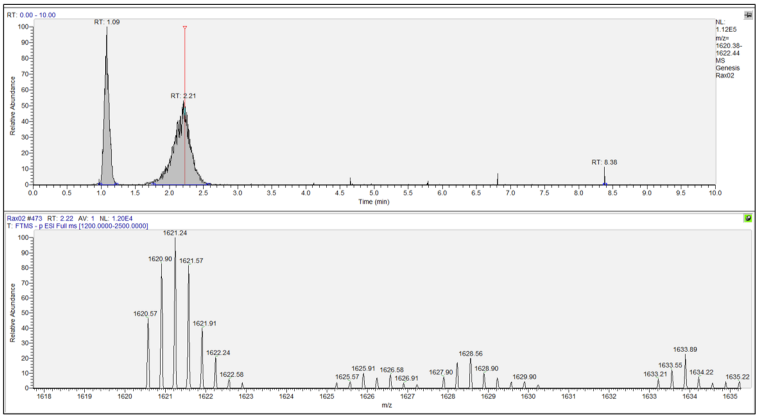


Figure 1. LCMS analysis for the AlkB repair reaction. A. the extracted ion current and MS spectrum for m3C, B. the extracted ion current and MS spectrum for C (repaired product).

For AlkB repair of m1A in ss-DNA, the kcat of was 2.5 ± 0.1 min^{-1} and the KM was 7.3 ± 0.9 μM, which are comparable to the literature reported.

The repair efficiency of m1A, m3C, eA are high, the Kcat/Km is 1.08 for m3C, 0.65 for m1A, 0.38 for eA.

The repair efficiency of m1G, m2G, eC, eG and m3T are low. The Kcat/Km for m1G was 0.037, which is considered as low activity, the eC and eG are even lower (Tabel 2). Based on previous repair efficiency result, With a 10uM enzyme, 5uM adduct oligo reaction condition, the repair ratio for m1G in 30min is 30%, while the repair of m3T and m2G are less than 5%. We considered them as low activity and did not do the Kcat/Km measurement.

Table 2. Kcat/Km for different DNA adduct as substrate repaired by AlkB.

lesion	Kcat/Km (Min^{-1}/μM)
m3C	1.08
m1A	6.1×10^{-1}
eA	4.8×10^{-1}

eG	1.5×10^{-3}
eC	7.2×10^{-3}
m1G	3.7×10^{-2}

In vivo data analysis

From reviewing previous literature, we noted that in vivo experiments measure both bypass and repair rates. The bypass rate indicates if an adduct blocks DNA replication; a low rate suggests the adduct is too bulky for the repair enzyme to handle, preventing DNA polymerase from synthesizing a new strand. A high bypass rate indicates that the repair enzyme can effectively accommodate and repair the adduct back to a normal DNA base. The repair rate assesses whether the enzyme correctly repairs the adduct, with a high rate indicating accurate restoration to the original base and low mutation likelihood, whereas a low rate suggests the repair process introduces mutations. In our study, we used the product of the bypass and repair rates as a parameter to evaluate the efficiency of DNA adduct repair.

The table data indicates that the final repair rates for m1A, m3C, and eA are high at 110%, 98.98%, and 84%, respectively, mirroring the trend observed in their Kcat/Km values. Conversely, the final repair rates for m1G and eC are significantly lower at 14.55% and 15.75%, respectively, and even lower for m3T and eG, both below 5%,

aligning with their Kcat/Km data (Figure 2). However, the in vivo data for m2G presents an anomaly, showing a high final repair rate of 97%. This contrasts sharply with its low in vitro repair efficiency, comparable to that of m3T.

Table 3. Different DNA adduct as substrate repaired in vivo

lesion	By pass ratio with AlkB	Repair ratio with AlkB	Final repair (bypass * repair)
m1A	110%	100%	110.000%
m3C	101%	98%	98.980%
m1G	15%	97%	14.550%
m3T	5%	60%	3.000%
eA	85%	99.5%	84.575%
eC	25%	63%	15.750%
eG	4.4%	81%	3.564%
m2G	99%	98%	97.020%

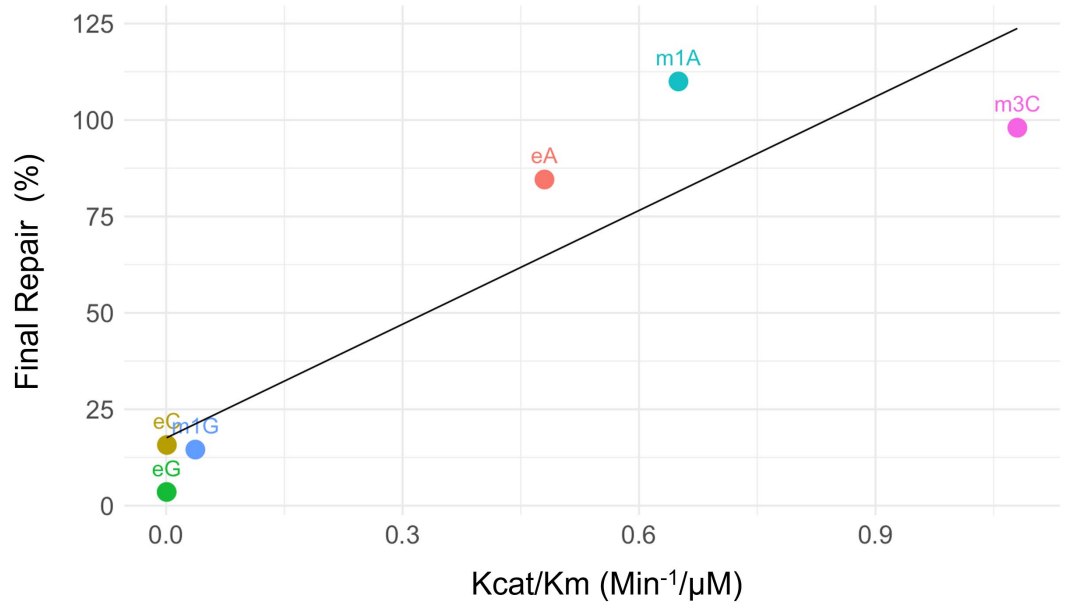


Figure 2. Correlation between Kcat/Km and Final repair ratio

DISCUSSION:

DNA repair plays a pivotal role in the field of biology and human health, with research into AlkB offering valuable insights for future studies. However, determining in vivo repair efficiency is challenging, involving complex processes such as plasmid construction, transformation, and DNA extraction for analysis. By establishing a correlation between in vitro and in vivo data, we can predict the in vivo efficacy of AlkB-mediated DNA adduct repair using in vitro catalytic efficiency (Kcat/Km) measurements.

In this study, we introduced a novel metric termed 'final repair', which accounts for both replication blockage and repair efficiency. This is a critical consideration for in vitro reactions, where two main factors influence repair efficiency. The first is the size and shape of the adduct, which may hinder its fit into the enzyme's pocket, reducing access to the active site and, consequently, repair rates. The second factor is the adduct's conformation; even if an adduct can easily enter the enzyme's pocket, its repairability depends on the proximity of the modification group to the active site. In vivo, these factors correspond to replication blockage and repair accuracy, respectively.

Notably, the adduct m2G stands out in our findings. Despite exhibiting low repair efficiency in vitro, in vivo data indicate both minimal replication blockage and high repair efficiency for m2G. This discrepancy might be explained by the ability of the m2G adduct to rotate within the double helix, avoiding disruption of Watson-Crick pairing. Consequently, m2G functions effectively in DNA without causing mutations. The in vitro results highlight the genuine activity of the enzyme in repairing adducts.

Future research should leverage the established correlation between in vitro and in vivo data for predictive analyses of in vivo activity, given the challenges of conducting in vivo assays.

REFERENCE：

1. Kataoka, H.; Yamamoto, Y.; Sekiguchi, M. J. J. o. b., A new gene (alkB) of Escherichia coli that controls sensitivity to methyl methane sulfonate. **1983,** *153* (3), 1301-1307.

2. Wang, J.; Qi, R.; Li, H.; Christov, C.; Lehnert, N.; Li, D., Genetic and Epigenetic Biomarkers Related to 2-Oxoglutarate/Fe (II)-Dependent Oxygenases and Implications for Disease and Toxicology. In *Biomarkers in Toxicology*, Springer: 2023; pp 323-349.

3. Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. J. J. o. B. C., The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. **2015,** *290* (34), 20734-20742.

4. Waheed, S. O.; Ramanan, R.; Chaturvedi, S. S.; Lehnert, N.; Schofield, C. J.; Christov, C. Z.; Karabencheva-Christova, T. G. J. A. C. S., Role of structural dynamics in selectivity and mechanism of non-heme Fe (II) and 2-oxoglutarate-dependent oxygenases involved in DNA repair. **2020,** *6* (5), 795-814.

5. Wang, B.; Usharani, D.; Li, C.; Shaik, S. J. J. o. t. A. C. S., Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. **2014,** *136* (39), 13895-13901.

6. Fang, D.; Lord, R. L.; Cisneros, G. A. J. T. J. o. P. C. B., Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. **2013,** *117* (21), 6410-6420.

7. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; De Visser, S. P. J. C. A. E. J., Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. **2014,** *20* (2), 435-446.

8. Fang, D.; Cisneros, G. A. s. J. J. o. c. t.; computation, Alternative pathway for the reaction catalyzed by DNA dealkylase AlkB from ab initio QM/MM calculations. **2014,** *10* (11), 5136-5148.

9. Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D. J. C. r. i. t., Oncometabolites d-and l-2-hydroxyglutarate inhibit the AlkB family DNA repair enzymes under physiological conditions. **2017,** *30* (4), 1102-1110.

10. Yang, T.; Cheong, A.; Mai, X.; Zou, S.; Woon, E. C. J. C. c., A methylation-switchable conformational probe for the sensitive and selective detection of RNA demethylase activity. **2016,** *52* (36), 6181-6184.

11. Zhu, C.; Yi, C. J. A. C. I. E., Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. **2014,** *53* (14), 3659-3662.

12. Yu, B.; Hunt, J. F. J. P. o. t. N. A. o. S., Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. **2009,** *106* (34), 14315-14320.

13. Roy, T. W.; Bhagwat, A. J. N. a. r., Kinetic studies of Escherichia coli AlkB using a new fluorescence-based assay for DNA demethylation. **2007,** *35* (21), e147-e147.

14. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F. J. N., Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. **2006,** *439* (7078), 879-884.

15. Falnes, P. Ø. J. N. a. r., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. **2004,** *32* (21), 6260-6267.

16. Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. J. C. r. i. t., Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. **2016,** *29* (4), 687-693.

17. Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T. J. A. C. I. E., Selective enzymatic demethylation of N2, N2-dimethylguanosine in RNA and its application in high-throughput tRNA sequencing. **2017,** *56* (18), 5017-5020.

18. Nigam, R.; Anindya, R. J. B.; communications, b. r., Escherichia coli single-stranded DNA binding protein SSB promotes AlkB-mediated DNA dealkylation repair. **2018,** *496* (2), 274-279.

19. Ergel, B.; Gill, M. L.; Brown, L.; Yu, B.; Palmer, A. G.; Hunt, J. F. J. J. o. B. C., Protein dynamics control the progression and efficiency of the catalytic reaction cycle of the Escherichia coli DNA-repair enzyme AlkB. **2014,** *289* (43), 29584-29601.

20. Karkhanina, A. A.; Mecinovic, J.; Musheev, M. U.; Krylova, S. M.; Petrov, A. P.; Hewitson, K. S.; Flashman, E.; Schofield, C. J.; Krylov, S. N. J. A. c., Direct analysis of enzyme-catalyzed DNA demethylation. **2009,** *81* (14), 5871-5875.

21. Hsiao, Y.-C.; Liu, C.-W.; Chi, L.; Yang, Y.; Lu, K. J. C. r. i. t., Effects of gut microbiome on carcinogenic DNA damage. **2020,** *33* (8), 2130-2138.

22. Yasui, M.; Matsui, S.; Ihara, M.; Laxmi, Y. S.; Shibutani, S.; Matsuda, T. J. N. A. R., Translesional synthesis on a DNA template containing N 2-methyl-2′-deoxyguanosine catalyzed by the Klenow fragment of Escherichia coli DNA polymerase I. **2001,** *29* (9), 1994-2001.

23. Lu, K.; Collins, L. B.; Ru, H.; Bermudez, E.; Swenberg, J. A. J. T. s., Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. **2010,** *116* (2), 441-451.

24. Lu, K.; Gul, H.; Upton, P. B.; Moeller, B. C.; Swenberg, J. A. J. T. s., Formation of hydroxymethyl DNA adducts in rats orally exposed to stable isotope labeled methanol. **2012,** *126* (1), 28-38.

25. Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M., AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nature Structural & Molecular Biology* **2005,** *12* (10), 855-860.

26. Baldwin, M. R.; Admiraal, S. J.; O'Brien, P. J., Transient kinetic analysis of oxidative dealkylation by the direct reversal DNA repair enzyme AlkB. *Journal of Biological Chemistry* **2020,** *295* (21), 7317-7326.

27. Shivange, G.; Kodipelli, N.; Anindya, R., 2-Hydrazinobenzothiazole-based etheno-adduct repair protocol (HERP): A method for quantitative determination of direct repair of etheno-bases. *DNA Repair* **2015,** *28*, 8-13.

28. Mishina, Y.; Yang, C.-G.; He, C., Direct Repair of the Exocyclic DNA Adduct 1,N6-Ethenoadenine by the DNA Repair AlkB Proteins. *Journal of the American Chemical Society* **2005,** *127* (42), 14594-14595.

29. Falnes, P. Ø., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Research* **2004,** *32* (21), 6260-6267.

30. Huff, A. C.; Topal, M. D., DNA damage at thymine N-3 abolishes base-pairing capacity during DNA synthesis. *Journal of Biological Chemistry* **1987,** *262* (26), 12843-12850.

31. Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D., Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chemical Research in Toxicology* **2016,** *29* (4), 687-693.

32. Jia, G.; Yang, C.-G.; Yang, S.; Jian, X.; Yi, C.; Zhou, Z.; He, C., Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Letters* **2008,** *582* (23), 3313-3319.

33. Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proceedings of the National Academy of Sciences* **2004,** *101* (39), 14051-14056.

SUPPLEMENTARY INFORMATION

Study on Correlation of DNA Damage Repair Efficiency Obtained in vitro and in vivo methods

Xianhao Zhou[†], Jian Ma[†], Samuel Howarth[†], Yi-Tzai Chen[†], and Deyu Li[†,*]

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

* Correspondence: Deyu Li (deyuli@uri.edu)

TABLE OF CONTENTS

Table S1. Theoretical m/z and experimental m/z of different oligo synthesized.

lesion	Theoretical m/z at -3	Experimental m/z at -3
m3C	1625.2807	1625.2837
m1A	1633.2844	1633.2875
m1G	1638.6161	1638.6193
m2G	1651.9515	1651.9546
eA	1636.6126	1636.6155
eG	1641.9442	1641.9476
eC	1628.6088	1628.6094
m3T	1630.2806	1630.2838

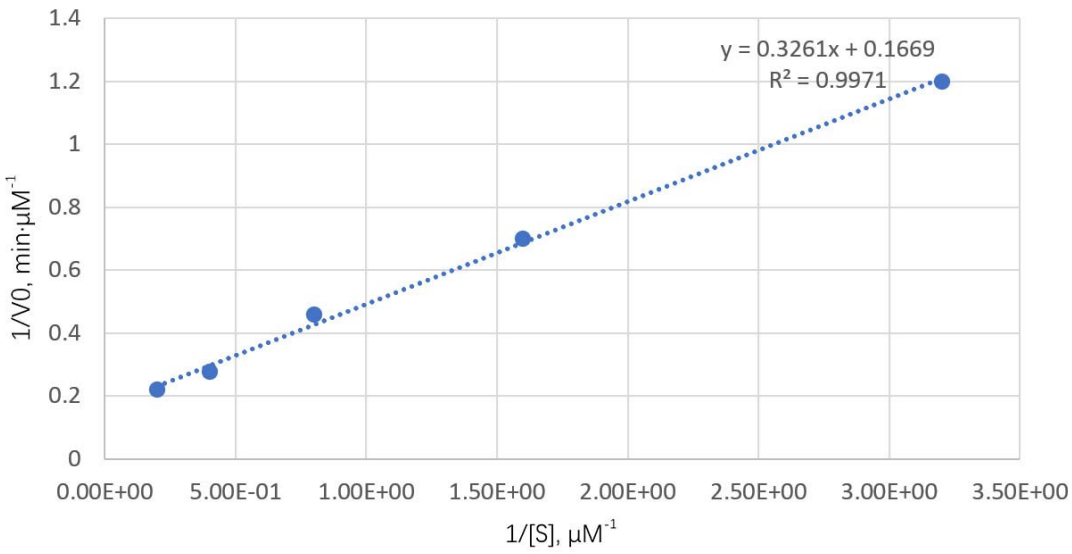


Figure S1. Lineweaver-Burk plot for m1A, the concentration of enzyme used in this reaction series is 5 uM

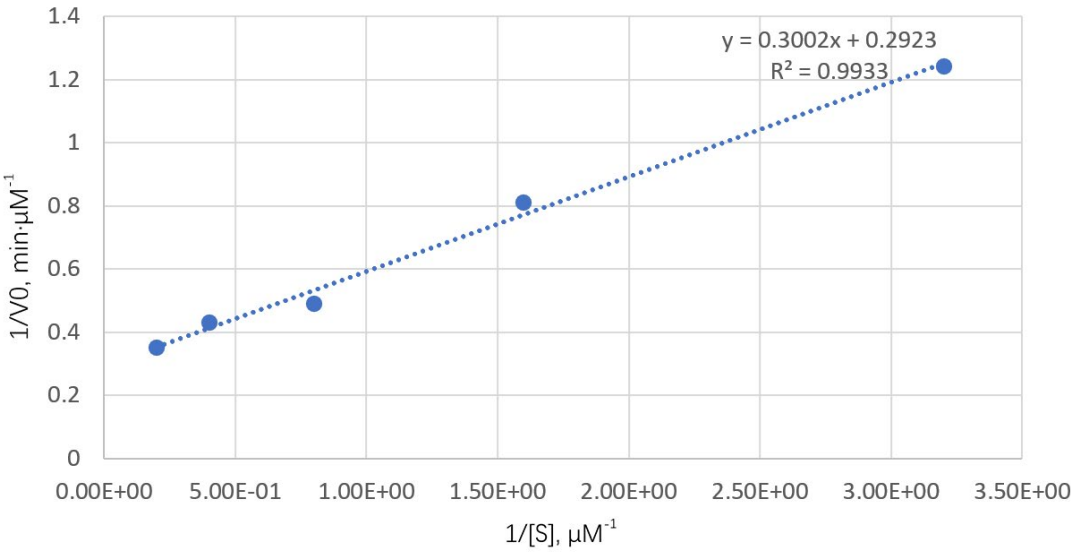


Figure S2. Lineweaver-Burk plot for m3C, the concentration of enzyme used in this reaction series is 1 uM

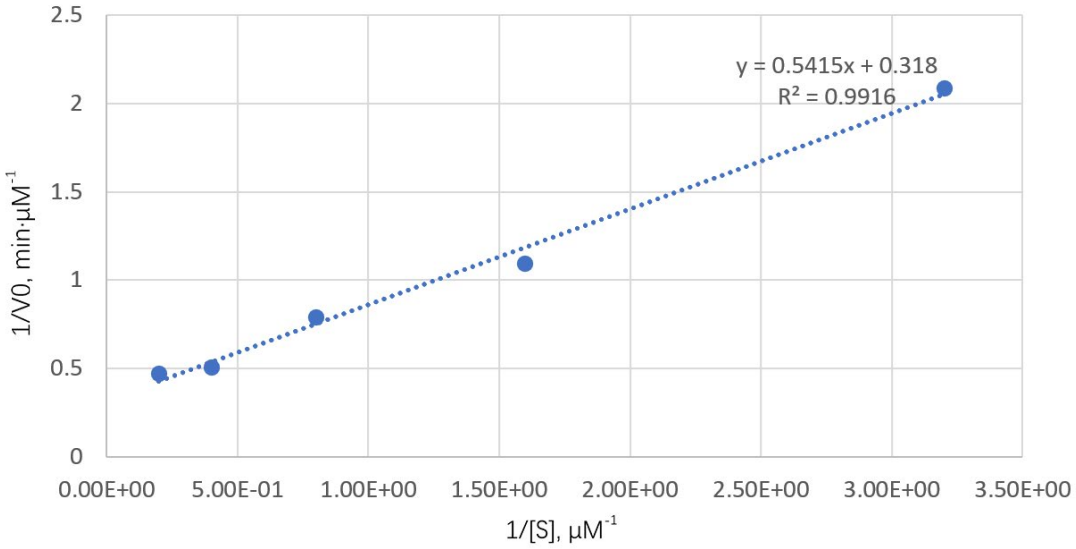


Figure S3. Lineweaver-Burk plot for eA, the concentration of enzyme used in this reaction series is 5 uM

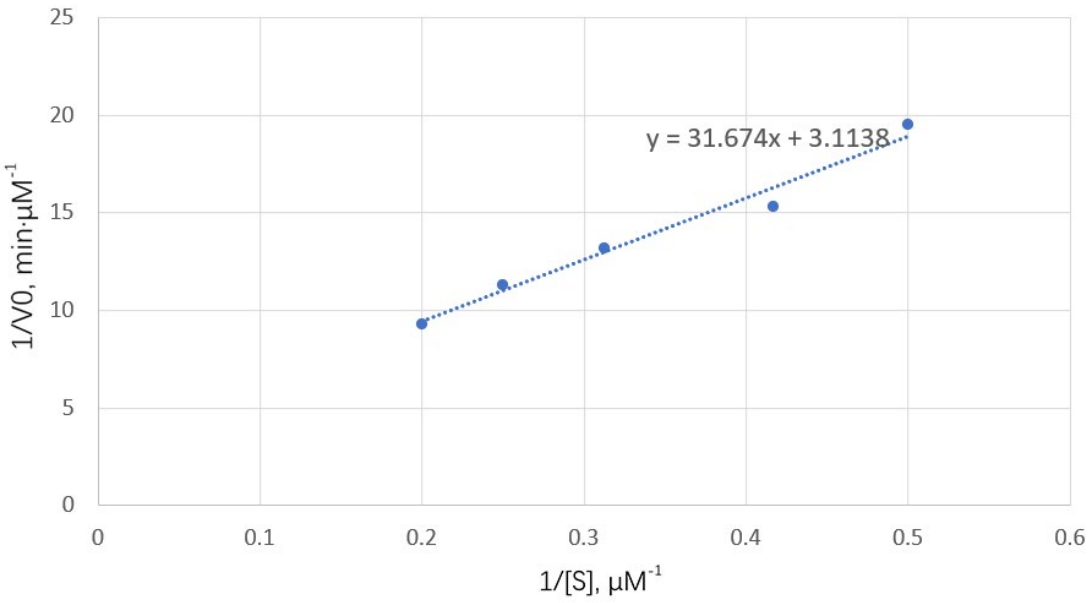


Figure S4. Lineweaver-Burk plot for eG, the concentration of enzyme used in this reaction series is 20 uM

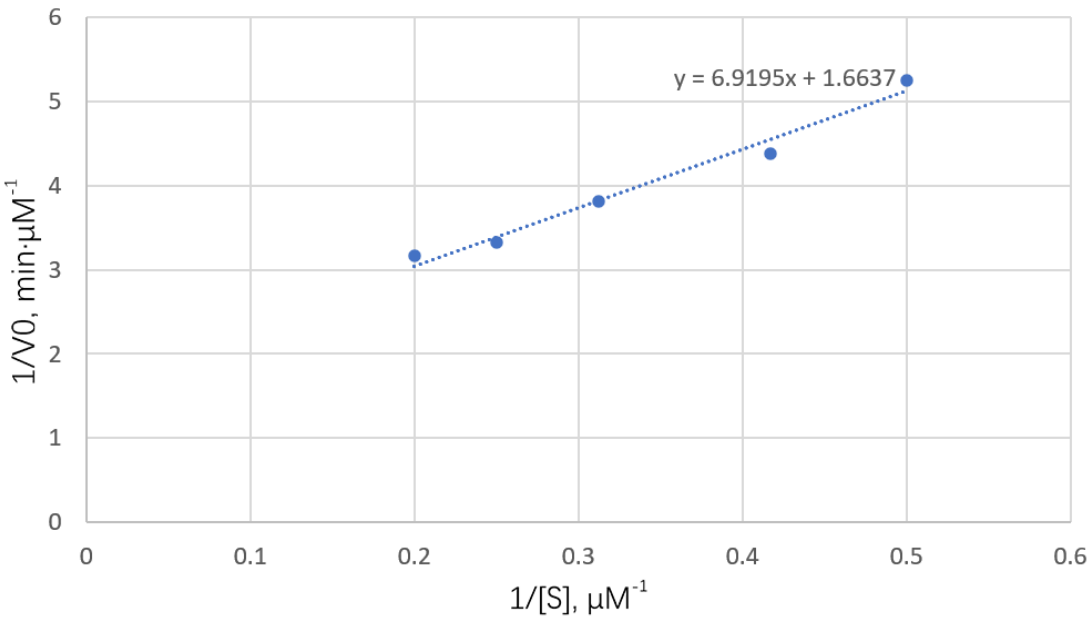


Figure S5. Lineweaver-Burk plot for eC, the concentration of enzyme used in this reaction series is 20 uM

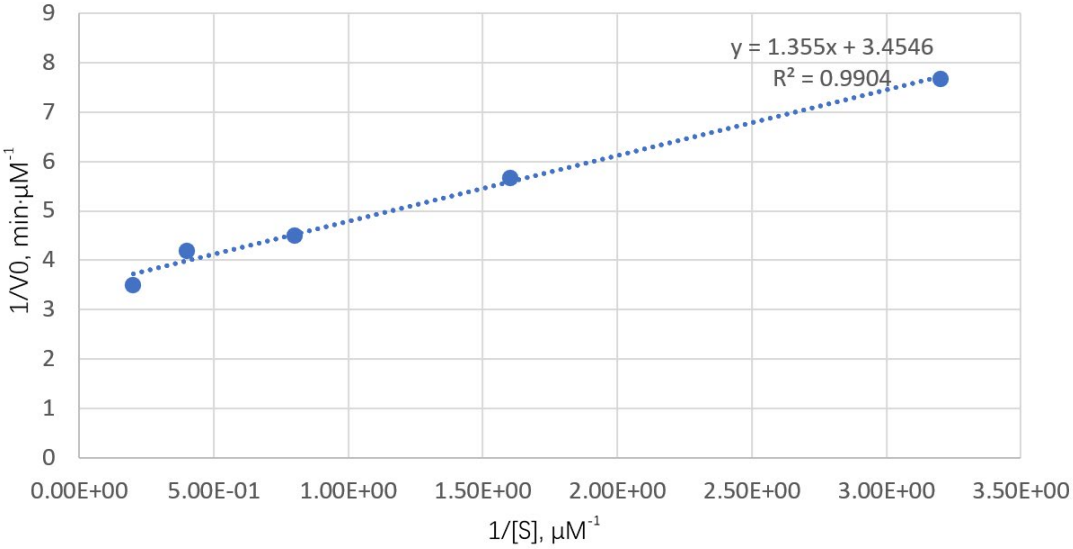


Figure S6. Lineweaver-Burk plot for m1G, the concentration of enzyme used in this reaction serious is 20 uM

.

CHAPTER - III

Prepared for submission to Chemical Research in Toxicology.

Comprehensive Kinetic Study of AlkB Protein on Various Alkyl Substrates

Xianhao Zhou[†], Zhiyuan Peng[†], Samuel Howarth[†], Evans Boateng Boakye[†], and Deyu Li[†,*]

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

* Correspondence: Deyu Li (deyuli@uri.edu)

ABSTRACT

The AlkB enzyme family, distinguished by its α-ketoglutarate/Fe(II)-dependent repair mechanism, is crucial in rectifying alkylated DNA, predominantly targeting lesions such as 1-methyladenine and 3-methylcytosine within single-stranded DNA (ss-DNA). Our research expands the known spectrum of AlkB substrates by exploring its repair efficacy on less frequently studied dimethyl adducts, notably 6,6-dimethyladenine and 4,4-dimethylcytosine, across both ss-DNA and double-stranded DNA (ds-DNA) environments. This study sheds light on AlkB's extraordinary versatility and adaptability in managing a broad array of DNA damages within varying DNA architectures. Our findings not only enhance the comprehension of AlkB's specificity towards different substrates but also highlight the intricate relationship between DNA repair mechanisms and the structural intricacies of DNA. These insights contribute significantly to our understanding of AlkB's role in the maintenance and repair of the genome, illustrating its intricate functionality in preserving genomic integrity.

INTRODUCTION

Nuclear DNA is constantly subjected to damage from both external and internal sources, leading to a range of DNA adducts within the genome. [1-3]When these adducts occur in single-stranded DNA regions, such as at replication forks, they can result in mutations or impede DNA synthesis. In a double-stranded context, adducts might alter the DNA structure, destabilizing the duplex or interfering with the binding of sequence-specific proteins like transcription factors. To counteract these detrimental effects, organisms have evolved sophisticated DNA repair mechanisms capable of addressing lesions in both single- and double-stranded DNA contexts. [4-6]Among these repair mechanisms, the AlkB protein in E. coli is known to effectively repair alkyl DNA adducts, such as 1-methyladenine (m1A) and 3-methylcytosine (m3C), preferring single-stranded substrates but also capable in double-stranded contexts.[7]

AlkB functions as a dioxygenase, utilizing an α-ketoglutarate/Fe(II)-dependent process to remove alkyl groups and restore the original DNA bases. [8] Since its discovery, various DNA adducts have been identified as AlkB substrates, including all seven N-methyl lesions that occur on the Watson-Crick base-pairing face of nucleobases. These adducts encompass m1A, m3C, m6A, N4-methylcytosine (m4C), 1-methylguanine (m1G), N2-methylguanine (m2G), and 3-methylthymine (m3T), with m6A, m4C, and m2G being exocyclic adducts that can potentially avoid disrupting base pairing. [9-17] In contrast, the methyl groups on m1A, m3C, m1G, and m3T are positioned on the nucleobase ring, likely hindering hydrogen-bond formation if not

repaired. AlkB's substrate range also extends to other adducts, including 3-ethylcytosine (e3C), N^2-ethylguanine, and several more complex lesions.[18]

In our study, we analyzed the kinetic parameters, specifically Kcat/Km, of AlkB repairing various substrates in both single- and double-stranded DNA. We synthesized and purified oligonucleotides containing these adducts and employed HPLC-MS to investigate DNA repair efficiency. Our findings indicate that substrates considered weak in single-stranded contexts exhibit higher activity in double-stranded DNA, highlighting the nuanced interplay between substrate structure, DNA conformation, and repair efficiency.

METHOD

Oligonucleotide Synthesis. Oligonucleotides with a length of 16 bases were synthesized, featuring the sequence 5'-GAAGACCTXGGCGTCC-3', where 'X' marks the site of the lesions. Complementary 23-mer oligonucleotides, with the sequence 5′-CTGGGACGCCYAGGTCTTCACTG-3′ (where 'Y' corresponds to the conventional bases A, T, C, or G), were acquired from IDT. The synthesis of these DNA sequences was carried out using solid-phase phosphoramidite chemistry on a MerMade-4 Oligonucleotide Synthesizer. Subsequent purification of the oligonucleotides was conducted using HPLC (High-Performance Liquid Chromatography) provided by Thermo Fisher Scientific, utilizing a DNAPac PA-100 Semi-Preparative column from Phenomenex. The solvents used in this process were solvent A, comprising 100 mM

1:1 triethylamine-acetic acid (TEAA) in water, and solvent B, which was pure acetonitrile. The quantification of the DNA concentration was performed using UV absorbance at 260nm, measured by a NanoDrop device. To characterize the oligonucleotides, HPLC-electrospray ionization triple quadrupole time of flight mass spectrometry (AB Sciex) was employed.

Expression and Purification of the AlkB. His-tagged AlkB protein was produced by introducing pET28a-AlkB into E. coli strains Rosetta2(DE3)pLysS and BL21(DE3)pLysS. Protein expression in these cells was triggered by adding 1mM isopropyl-β-D-thiogalactopyranoside (IPTG) at a temperature of 30 °C. Subsequent purification of the expressed protein was achieved through affinity chromatography methods. The His-tag component of the AlkB protein was then cleaved using thrombin. The thoroughly purified protein was subsequently stored at -80 °C in a previously described AlkB storage buffer for preservation.

Enzymatic Reaction. To evaluate the demethylase activity of the AlkB family enzymes on single-stranded and double-stranded DNA substrates, enzymatic assays were conducted at a constant temperature of 37 °C, with varying time intervals for kinetic analysis. The reaction mixture contained 70.0 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H2O$, 0.93 mM αKG, 1.86 mM ascorbic acid, and 46.5 mM HEPES, adjusted to pH 8.0. The enzymatic process was halted by introducing 10 mM EDTA and heating the solution to 95 °C for a duration of 5 minutes. In a standard procedure, purified proteins and oligonucleotides with specific DNA adducts were combined in a 20 μL reaction

volume, which included all necessary cofactors. For the separation of substrates and products, particularly for 16mer m1A and A or 16mer m3C and C, the HPLC conditions were set to initiate with a 5-minute gradient of 1.5 M ammonium acetate, increasing from 50% to 65%, followed by a 2-minute phase at 70% ammonium acetate. A DNApac PA-100 column (4× 250 mm) from Thermo Scientific was employed for this purpose, with UV detection set at 260 nm. These assays were all replicated three times to ensure reliable results.

Kinetic Studies. To calculate the kinetic parameters kcat/Km for the repair reactions, initial reaction rates were measured by maintaining a consistent enzyme concentration while altering the concentration of the substrate. The reactions were timed at intervals of 0, 0.5, 1, 4, 10, and 30 minutes. For some substrate with low activity, the time points are: 0, 1, 2, 5, 10, 30 minutes. Each reaction was conducted at a temperature of 37 °C and replicated three times to ensure reliability and precision in the results.

RESULTS

To test the kinetics of different substrate on AlkB, we first chemically synthesized oligonucleotides by site-specifically incorporating m1A, m3C, eA, eG, eC, m1G. We also expressed and purified the *E. coli* AlkB protein.55 Then, we performed kinetic experiments to determine the kcat and Km of the AlkB enzymes as they repair different adducts in ss-DNA and ds-DNA.

Oligonucleotide Synthesis and Protein Purification. 16mer oligonucleotides containing different DNA adduct were chemically synthesized with the sequence 5'-GAAGACCTXGGCGTCC-3' (X denotes the alkylated base). The synthesized oligos were purified by reverse phase HPLC (Figure 1). 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 (Figure S1-S3). The genes for *E. coli* AlkB were cloned into pET28a+ expression vector; the incorporation of the correct sequences was confirmed by sequencing the corresponding plasmids. The proteins were then expressed in *E. coli* hosts, isolated and purified by affinity chromatography (Figure 2).

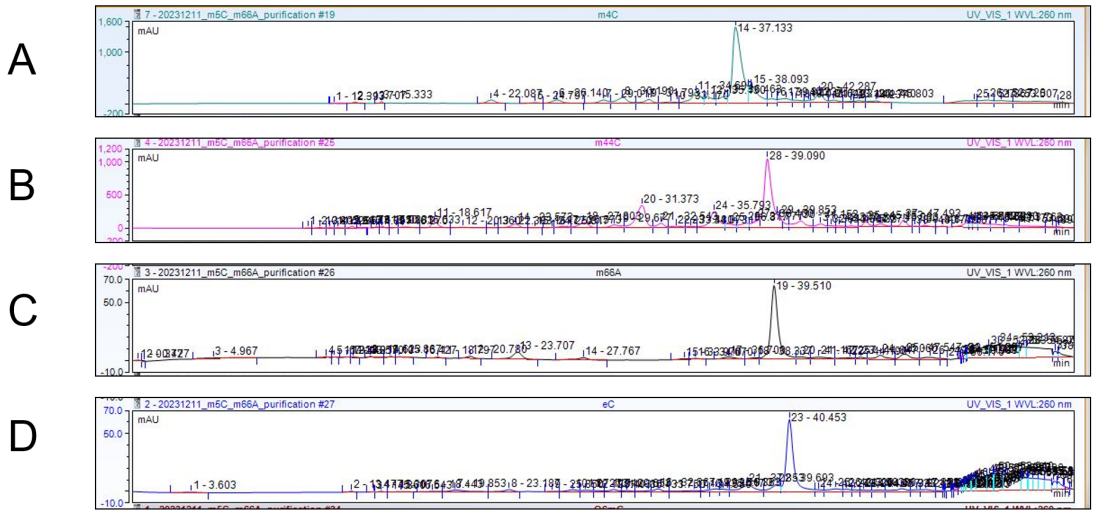


Figure 1. The HPLC purification for the synthesized Oligos with DNA adducts. (A).m4C, (B).m44C, (C). m66A, (D). eC

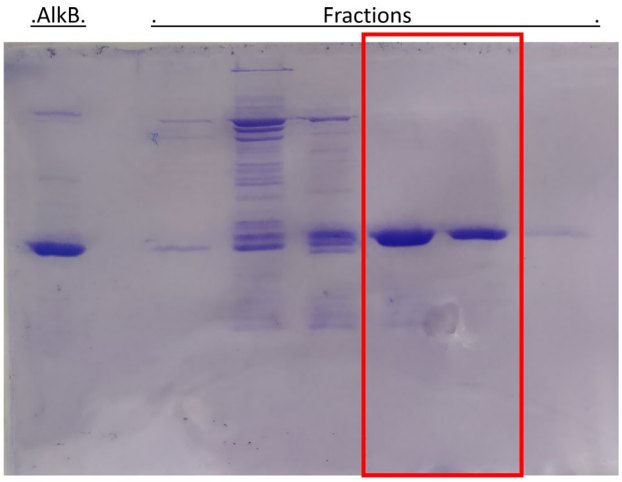


Figure 2. Protein purification with FPLC, the fractions within the red box is considered as purified AlkB.

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), αKG, and ascorbic acid. Below, m1A will be used as an example to explain the LCMS analyses. For the ss-DNA reactions, the starting material 16mer m1A (Figure 2A) and product 16mer A (Figure 2B) were extracted in MS, and the amount of each was quantified by peak area integration.

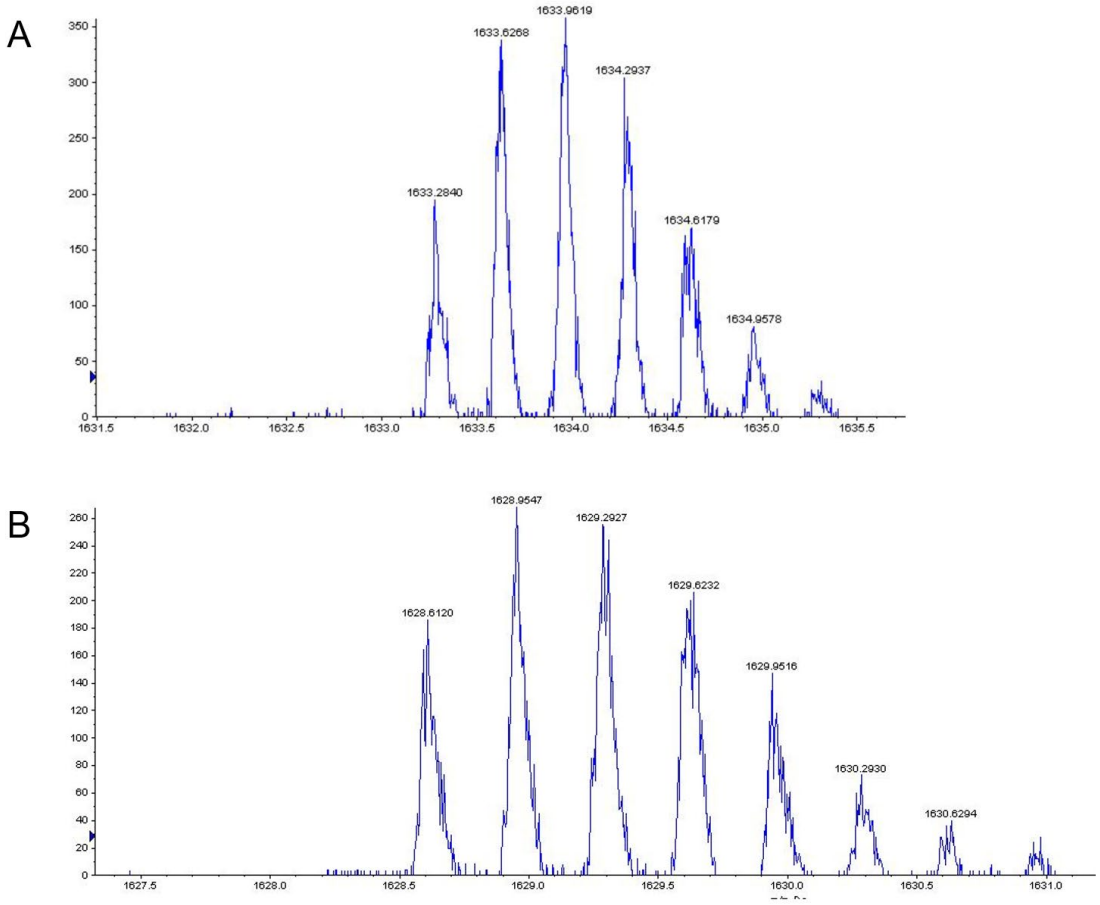


Figure 3. (A). MS signal for m1A. (B). MS signal for A (repaired from m1A).

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 repairing different substrates. In a typical kinetic analysis (e.g., AlkB repairing m1A), 1 μM AlkB enzyme were mixed with different concentrations of oligonucleotide substrate (2-5 μM) and the extent of the repair reaction was quantified at different time points (see Experimental section for details). The concentrations of the product 16mer A were used to calculate the kcat and KM. the integrated MS signal in LCMS was used

to calculate the repair ratio. the concentration of product was calculated by repair ratio multiply with initial concentration of starting material. 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.

The repair efficiency of m1A, m3C, eA are high, the Kcat/Km is 1.03 for m3C, 0.66 for m1A, 0.37 for eA. The repair efficiency of m1G, m2G, eC, eG and m3T are low. The Kcat/Km for m1G was 0.042, which is considered as low activity, however, eC and eG are even lower. The m66A shows low activity, the Kcat/Km for m66A is below 0.01, but it is still better substrate than m6A. All data shows in Table 1 and Table 2.

Tabel 1. Kcat/Km for different DNA adduct as substrate repaired by AlkB in ss-DNA.

DNA adduct	Kcat/Km ($Min^{-1}/\mu M$)
m3C	1.03
m1A	6.6×10^{-1}
eA	3.7×10^{-1}
eG	6.1×10^{-3}
eC	8.3×10^{-3}

m1G	4.2×10^{-2}
M66A	3.1×10^{-3}
M2G	2.1×10^{-3}

Tabel 2. Kcat/Km for different DNA adduct as substrate repaired by AlkB in ds-DNA.

DNA adduct	Kcat/Km ($Min^{-1}/\mu M$)
m3C	7.5×10^{-1}
m1A	4.5×10^{-1}
eA	3.2×10^{-1}
eG	9.3×10^{-3}
eC	1.1×10^{-2}
m1G	6.2×10^{-2}
m66A	8.2×10^{-2}
m2G	3.4×10^{-3}

DISCUSSION

In this study, we conducted a comprehensive kinetic analysis of the AlkB enzyme, focusing on its interaction with diverse DNA adducts across both single-stranded (ss-DNA) and double-stranded (ds-DNA) configurations. Prior research delineates a predilection of AlkB towards the repair of m1A and m3C lesions within ss-DNA, in contrast to a preference for m3T adducts within ds-DNA matrices. This study aimed to elucidate the substrate specificity and strand preference of AlkB, particularly distinguishing between high-affinity (strong) and low-affinity (weak) substrates. Our empirical data elucidate a differential strand preference: high catalytic efficiency (Kcat/Km) was observed for high-affinity substrates such as m1A and m3C in ss-DNA, whereas an elevated Kcat/Km for certain low-affinity substrates was noted in ds-DNA contexts.

The kinetic parameters derived from this study provide a robust foundation, corroborating previously observed substrate preferences of AlkB. This research also extends to the kinetic profiling of hitherto unreported substrates, including dimethylated adducts such as m66A and m44C. Intriguingly, these dimethyl adducts demonstrated a propensity for conversion to their monomethyl counterparts. This phenomenon indicates that the active site of AlkB possesses sufficient spatial accommodation to facilitate such transformations, heralding further inquiries into the structural conformation of AlkB in complex with dimethyl adducts to decipher the mechanistic underpinnings of their enhanced reactivity.

Moreover, this manuscript introduces a methodological paradigm for the quantification of kinetic parameters pertinent to DNA repair enzymes. Such methodologies hold promise for the kinetic examination of AlkB homologues in subsequent research endeavors, thereby advancing the molecular understanding of DNA repair processes across varied biological entities.

REFERENCE

1. Huff, A. C.; Topal, M. D., DNA damage at thymine N-3 abolishes base-pairing capacity during DNA synthesis. *Journal of Biological Chemistry* **1987,** *262* (26), 12843-12850.

2. Barrows, L. R.; Magee, P. N. J. C., Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro. **1982,** *3* (3), 349-351.

3. Lindahl, T.; Sedgwick, B.; Sekiguchi, M.; Nakabeppu, Y. J. A. r. o. b., Regulation and expression of the adaptive response to alkylating agents. **1988,** *57* (1), 133-157.

4. Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. J. N., AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. **2002,** *419* (6903), 178-182.

5. Mathison, B. H.; Frame, S. R.; Bogdanffy, M. S. J. I. t., DNA methylation, cell proliferation, and histopathology in rats following repeated inhalation exposure to dimethyl sulfate. **2004,** *16* (9), 581-592.

6. Rydberg, B.; Lindahl, T. J. T. E. j., Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. **1982,** *1* (2), 211-216.

7. Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D., Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chemical Research in Toxicology* **2016,** *29* (4), 687-693.

8. Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B., Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences* **2002,** *99* (26), 16660-16665.

9. Falnes, P. Ø., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Research* **2004,** *32* (21), 6260-6267.

10. Yu, B.; Hunt, J. F., Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proceedings of the National Academy of Sciences* **2009,** *106* (34), 14315-14320.

11. Lu, K.; Gul, H.; Upton, P. B.; Moeller, B. C.; Swenberg, J. A. J. T. s., Formation of hydroxymethyl DNA adducts in rats orally exposed to stable isotope labeled methanol. **2012,** *126* (1), 28-38.

12. Yasui, M.; Matsui, S.; Ihara, M.; Laxmi, Y. S.; Shibutani, S.; Matsuda, T. J. N. A. R., Translesional synthesis on a DNA template containing N 2-methyl-2′-deoxyguanosine catalyzed by the Klenow fragment of Escherichia coli DNA polymerase I. **2001,** *29* (9), 1994-2001.

13. Hsiao, Y.-C.; Liu, C.-W.; Chi, L.; Yang, Y.; Lu, K. J. C. r. i. t., Effects of gut microbiome on carcinogenic DNA damage. **2020,** *33* (8), 2130-2138.

14. Nigam, R.; Anindya, R. J. B.; communications, b. r., Escherichia coli single-stranded DNA binding protein SSB promotes AlkB-mediated DNA dealkylation repair. **2018,** *496* (2), 274-279.

15. Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D. J. C. r. i. t., Oncometabolites d-and l-2-hydroxyglutarate inhibit the AlkB family DNA repair enzymes under physiological conditions. **2017,** *30* (4), 1102-1110.

16. Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T. J. A. C. I. E., Selective enzymatic demethylation of N2, N2-dimethylguanosine in RNA and its application in high-throughput tRNA sequencing. **2017,** *56* (18), 5017-5020.

17. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; De Visser, S. P. J. C. A. E. J., Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. **2014,** *20* (2), 435-446.

18. Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. J. J. o. B. C., The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. **2015,** *290* (34), 20734-20742.

SUPPLEMENTARY INFORMATION

Comprehensive Kinetic Study of AlkB Protein on Various Alkyl Substrates

Xianhao Zhou[†], Zhiyuan Peng[†], Samuel Howarth[†], Evans Boateng Boakye[†], and Deyu Li[†,*]

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

* Correspondence: Deyu Li (deyuli@uri.edu)

TABLE OF CONTENT

Figure S5. Lineweaver-Burk plot for ds-m1A, the concentration of enzyme used in this reaction series is 5 uM

Figure S6. Lineweaver-Burk plot for ds-eA, the concentration of enzyme used in this reaction series is 5 uM

Figure S7. Lineweaver-Burk plot for ds-eG, the concentration of enzyme used in this reaction series is 20 uM.

Figure S8. Lineweaver-Burk plot for ds-eC, the concentration of enzyme used in this reaction series is 20 uM.

Figure S9. Lineweaver-Burk plot for ds-m66A, the concentration of enzyme used in this reaction series is 20 uM.

Figure S10. Lineweaver-Burk plot for ds-m1G, the concentration of enzyme used in this reaction series is 20 uM.

Figure S11. Lineweaver-Burk plot for ds-m1G, the concentration of enzyme used in this reaction series is 20 uM ds-m2G

Figure S12. Lineweaver-Burk plot for ss-m3C, the concentration of enzyme used in reaction series is 1 uM

Figure S13. Lineweaver-Burk plot for ss-m1A, the concentration of enzyme used in reaction series is 5 uM

Figure S14. Lineweaver-Burk plot for ss-eA, the concentration of enzyme used in reaction series is 5 uM

Figure S15. Lineweaver-Burk plot for ss-eG, the concentration of enzyme used in reaction series is 20 uM

Figure S16. Lineweaver-Burk plot for ss-eC, the concentration of enzyme used in reaction series is 20 uM

Figure S17. Lineweaver-Burk plot for ss-m66A, the concentration of enzyme used in reaction series is 20 uM.

Figure S18. Lineweaver-Burk plot for ss-m1G, the concentration of enzyme used in reaction series is 20 uM.

Figure S19. Lineweaver-Burk plot for ss-m2G, the concentration of enzyme used in reaction series is 20 uM

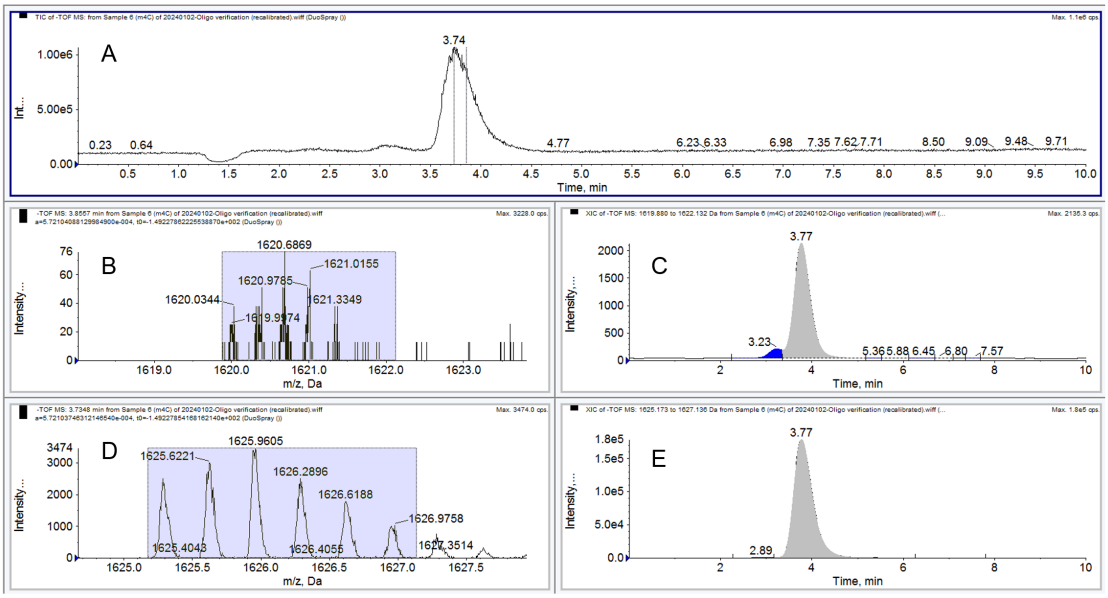


Figure S1. Verification of the synthesized purified m4C Oligo. (A). TIC for the 16mer m4C oligo. (B). MS spectrum for 16mer oligo with C in X position (major impurity). (C). Integrated signal for the impurity. (D). MS spectrum for 16mer oligo with m4C in X position (target product). (E). Integrated signal for the target product.

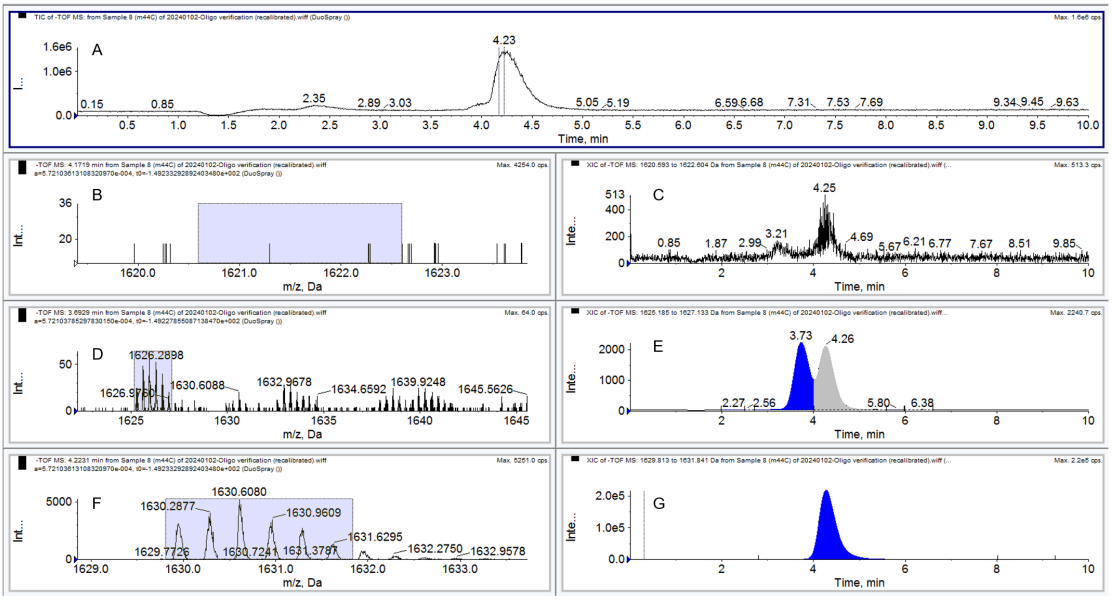


Figure S2. Verification of the synthesized purified m44C Oligo. (A). TIC for the 16mer m44C oligo. (B). MS spectrum for 16mer oligo with C in X position. (C). Integrated signal for the impurity showing in (B). (D). MS spectrum for 16mer oligo with m4C in X position (major impurity). (E). Integrated signal for the impurity showing in (D). (F). MS spectrum for 16mer oligo with m44C in X position (target product). (G). Integrated signal for the target product.

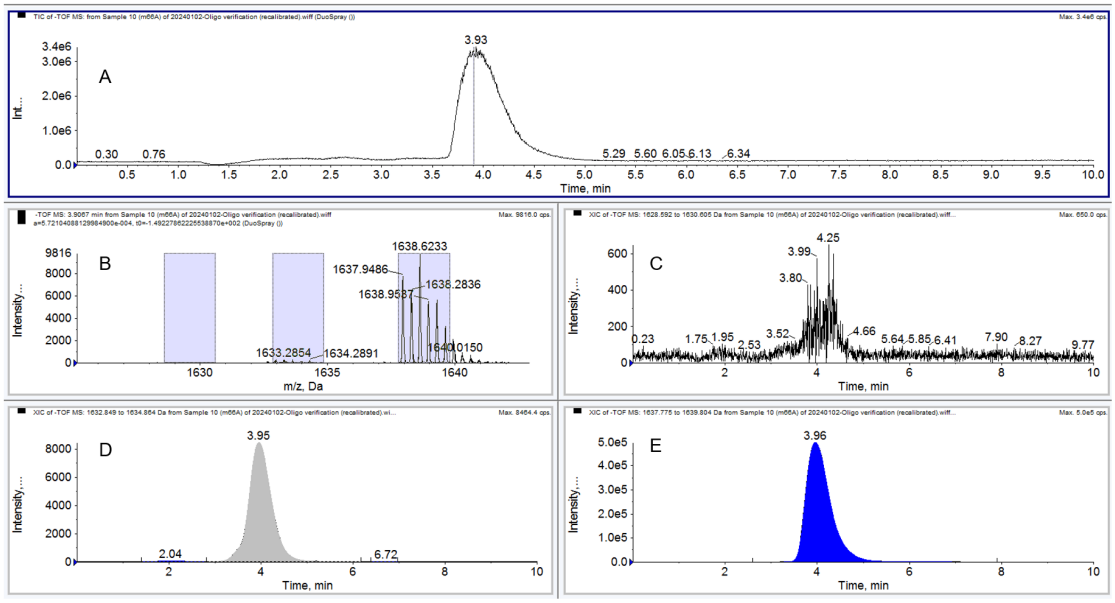


Figure S3. Verification of the synthesized purified m66A Oligo. (A) TIC of the 16mer m66A oligo. (B). MS spectrum for 16mer oligo with A (blue box in the left), m6A (blue box in the middle) and m66A (blue box in the right) in X position. (C). Integrated signal for the 16mer oligo with A impurity. (D). Integrated signal for the 16mer oligo with m6A impurity. (E). Integrated signal for the 16mer oligo with m66A (target product).

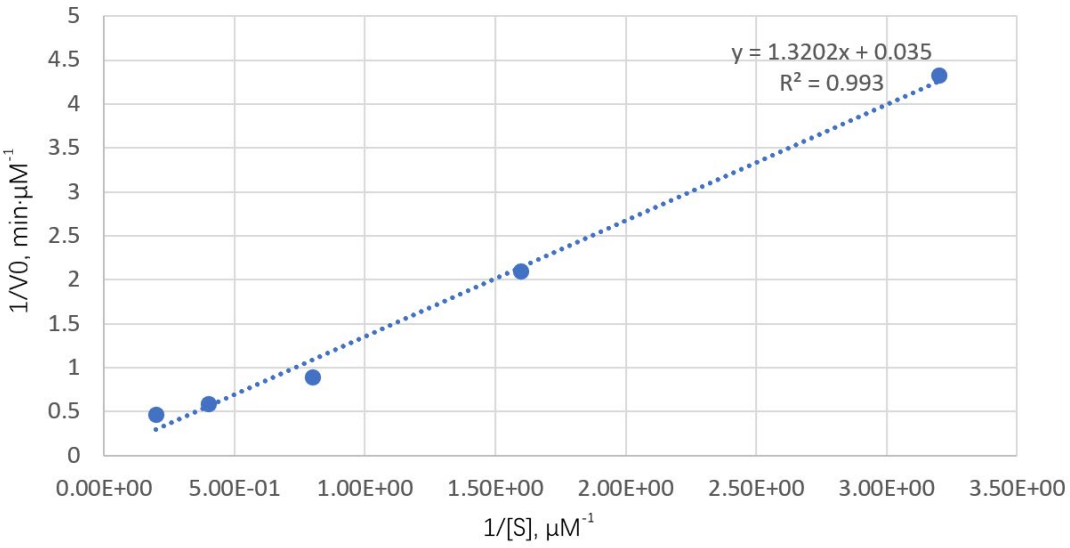


Figure S4. Lineweaver-Burk plot for ds-m3C, the concentration of enzyme used in this reaction series is 1 uM

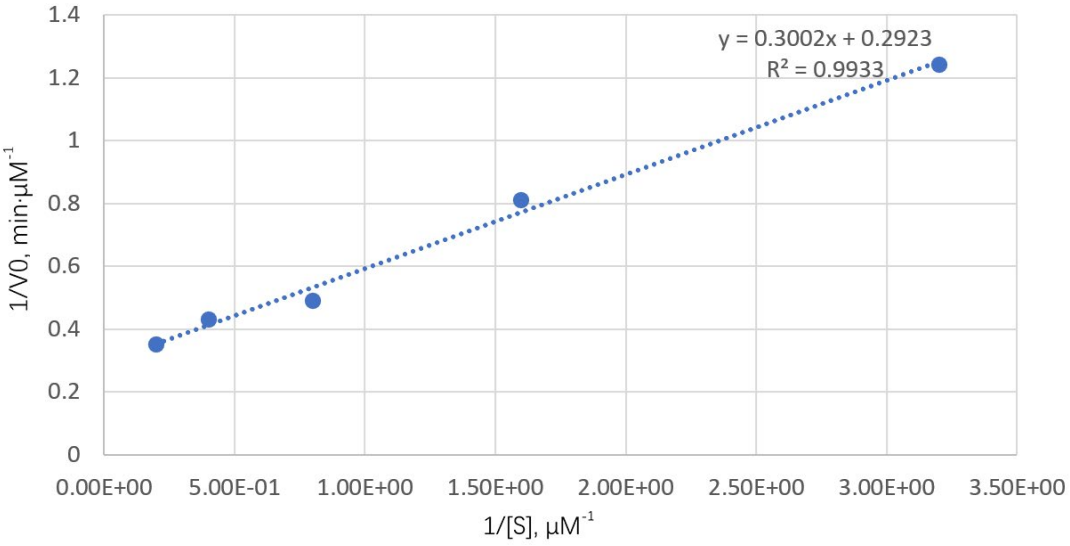


Figure S5. Lineweaver-Burk plot for ds-m1A, the concentration of enzyme used in this reaction series is 5 uM

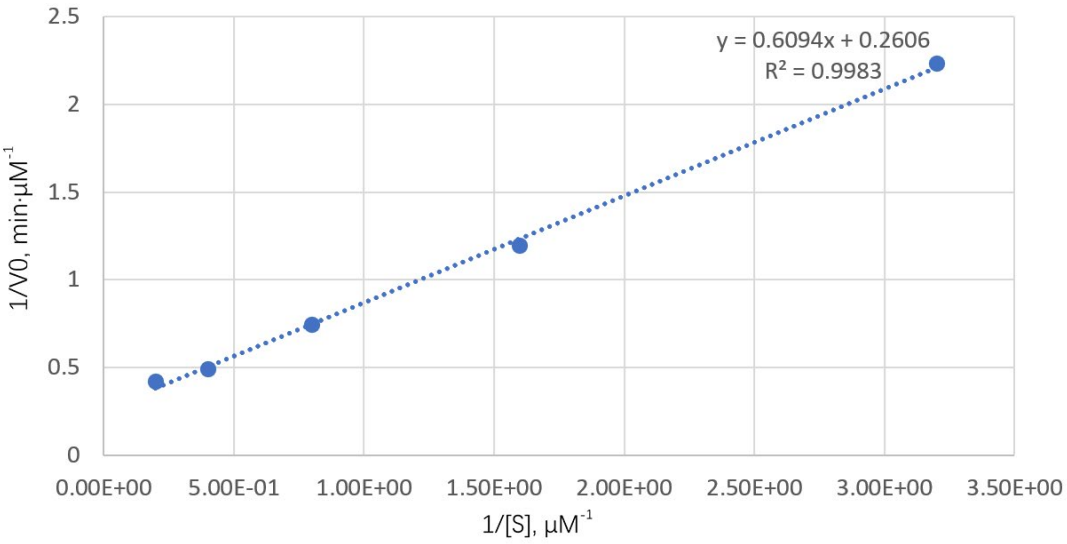


Figure S6. Lineweaver-Burk plot for ds-eA, the concentration of enzyme used in this reaction series is 5 uM

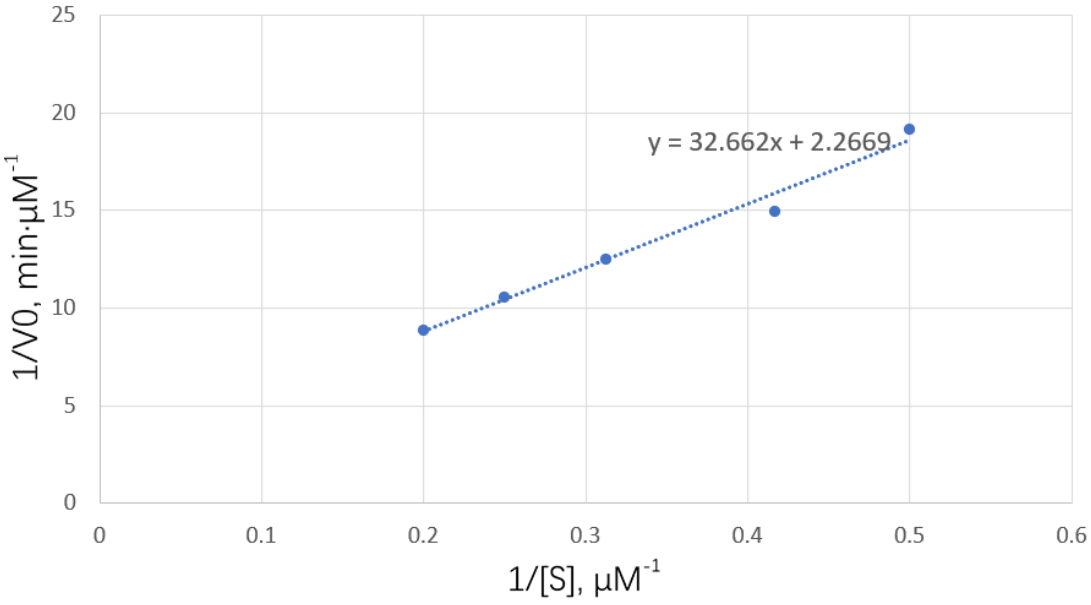


Figure S7. Lineweaver-Burk plot for ds-eG, the concentration of enzyme used in this reaction series is 20 uM.

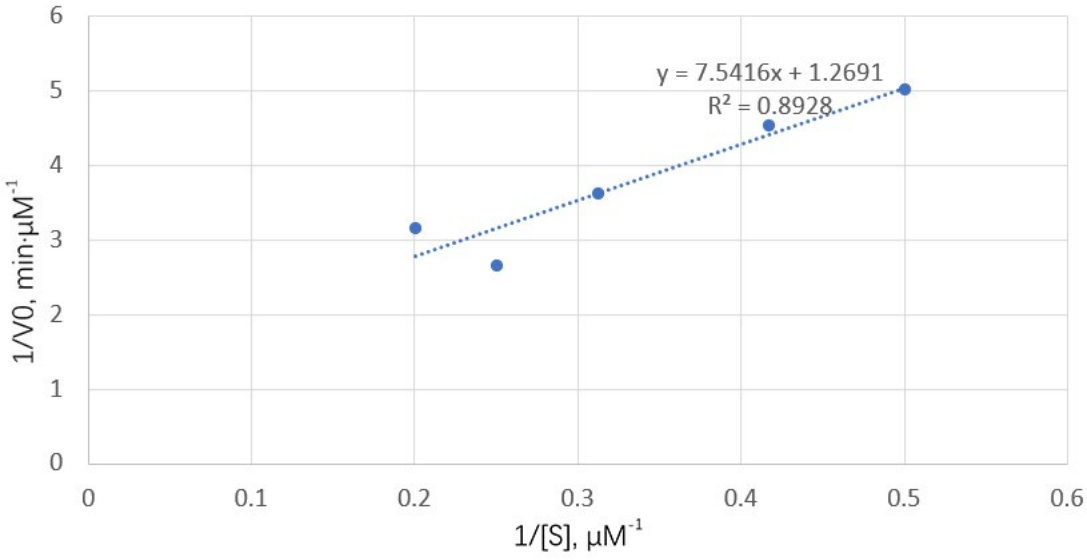


Figure S8. Lineweaver-Burk plot for ds-eC, the concentration of enzyme used in this reaction series is 20 uM.

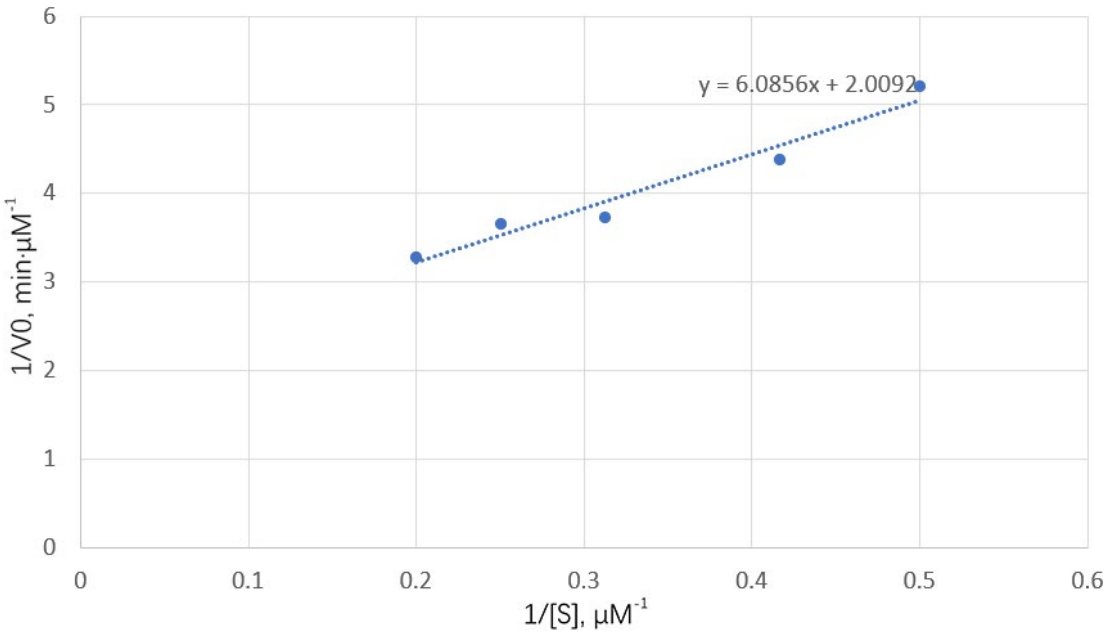


Figure S9. Lineweaver-Burk plot for ds-m66A, the concentration of enzyme used in this reaction series is 20 uM.

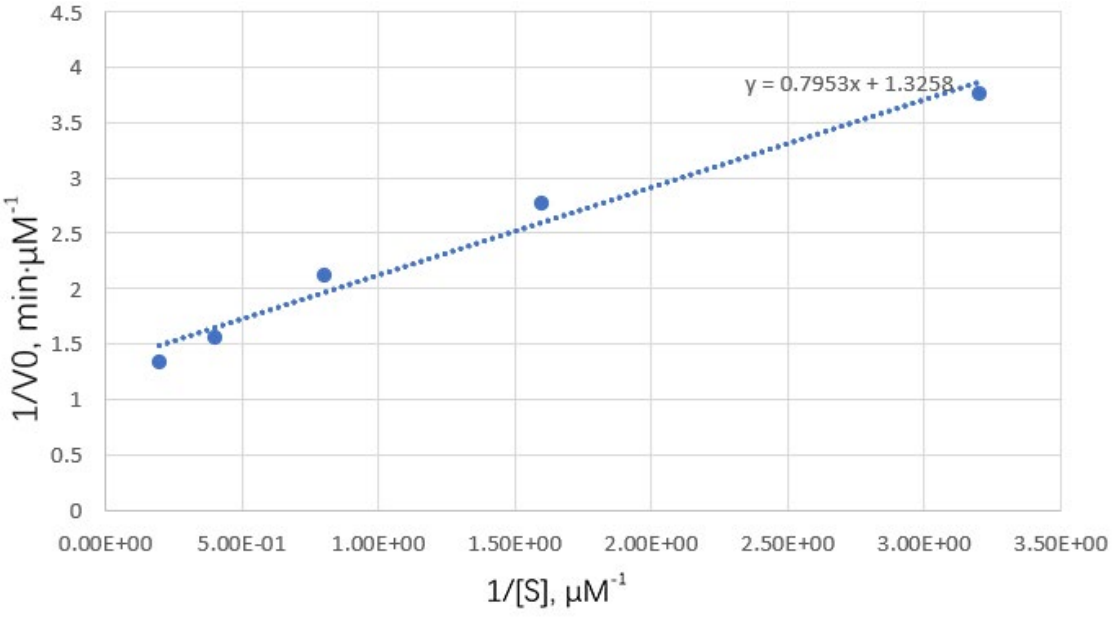


Figure S10. Lineweaver-Burk plot for ds-m1G, the concentration of enzyme used in this reaction series is 20 uM.

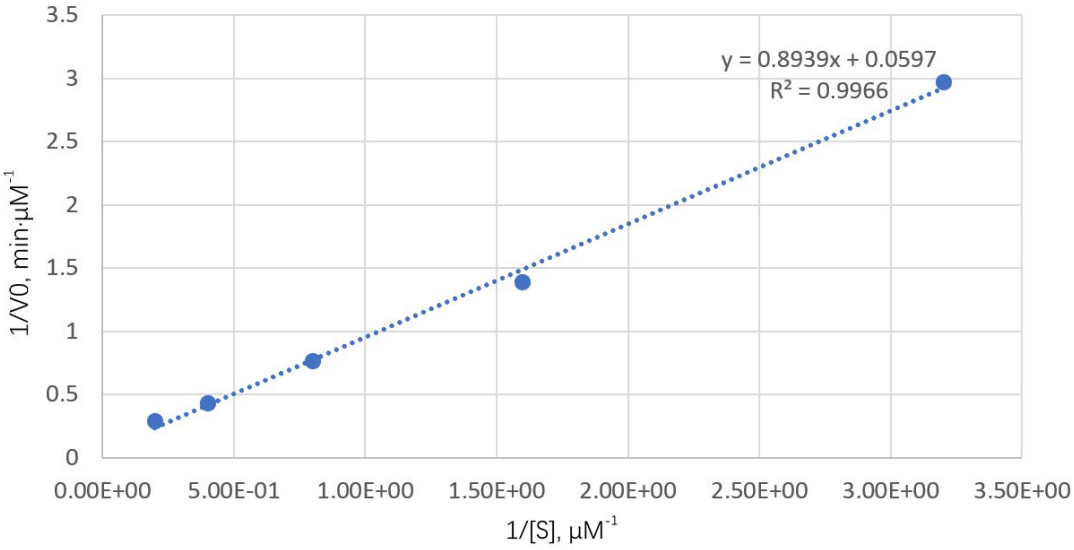


Figure S11. Lineweaver-Burk plot for ss-m3C, the concentration of enzyme used in reaction series is 1 uM

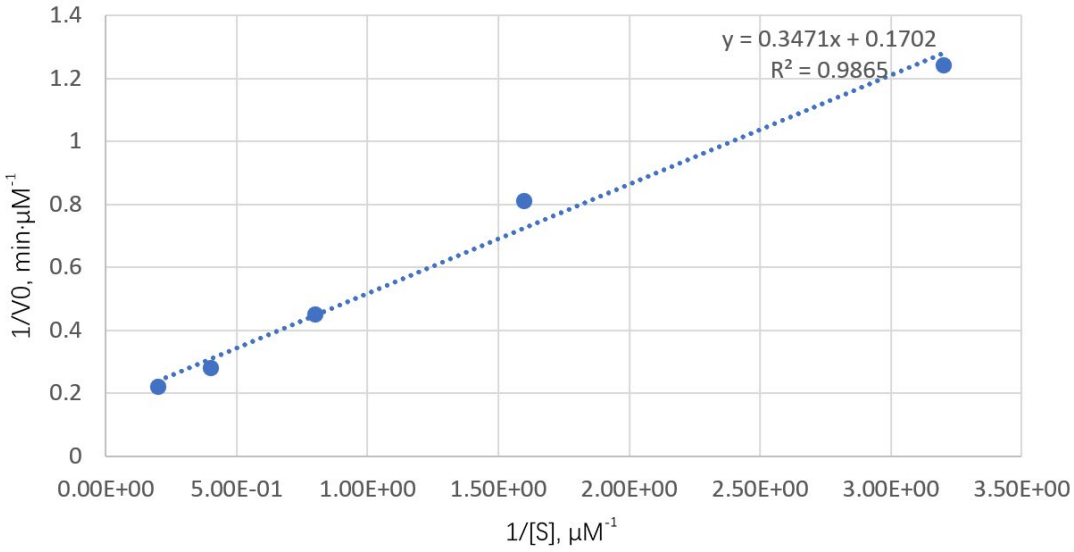


Figure S12. Lineweaver-Burk plot for ss-m1A, the concentration of enzyme used in reaction series is 5 uM

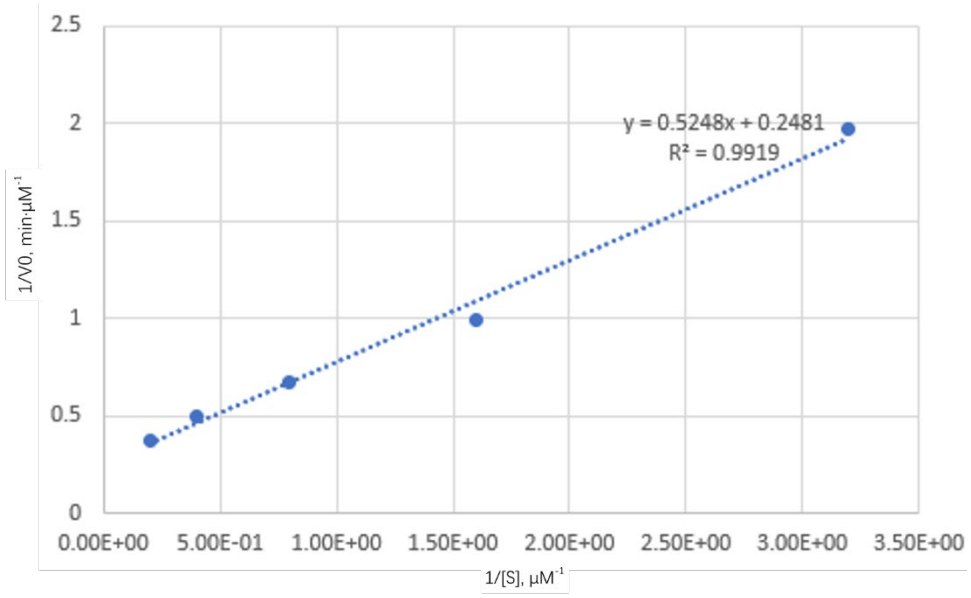


Figure S13. Lineweaver-Burk plot for ss-eA, the concentration of enzyme used in reaction series is 5 uM

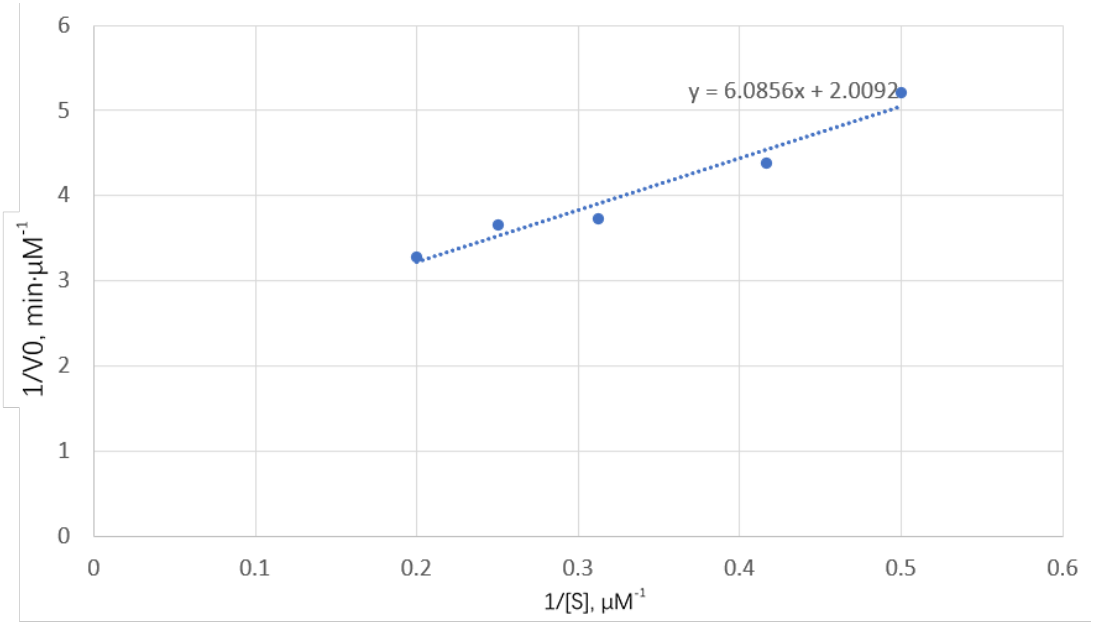


Figure S14. Lineweaver-Burk plot for ss-eG, the concentration of enzyme used in reaction series is 20 uM

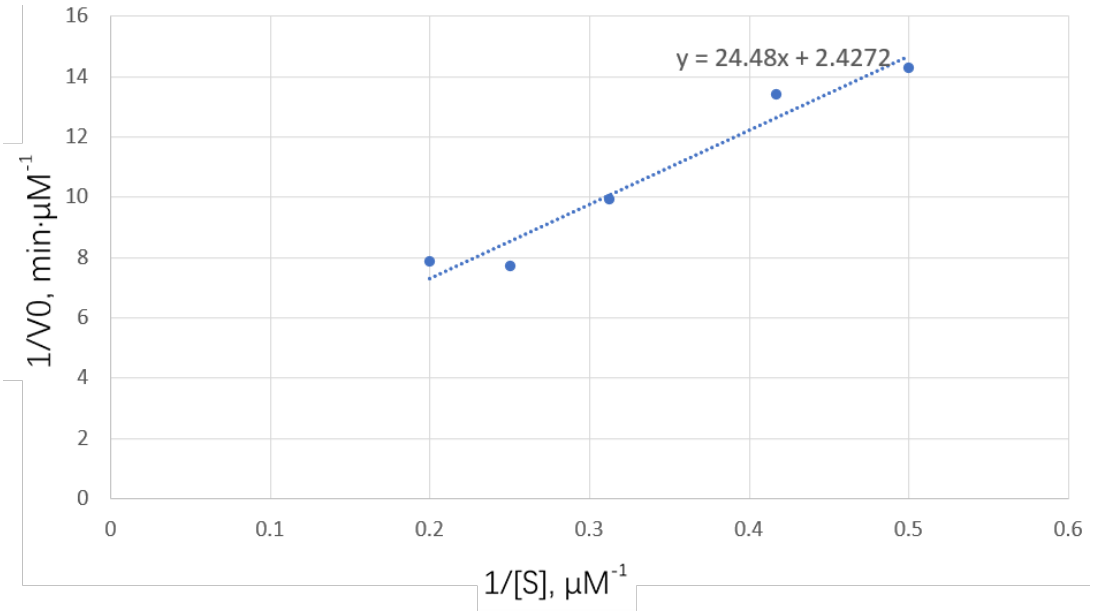


Figure S15. Lineweaver-Burk plot for ss-eC, the concentration of enzyme used in reaction series is 20 uM

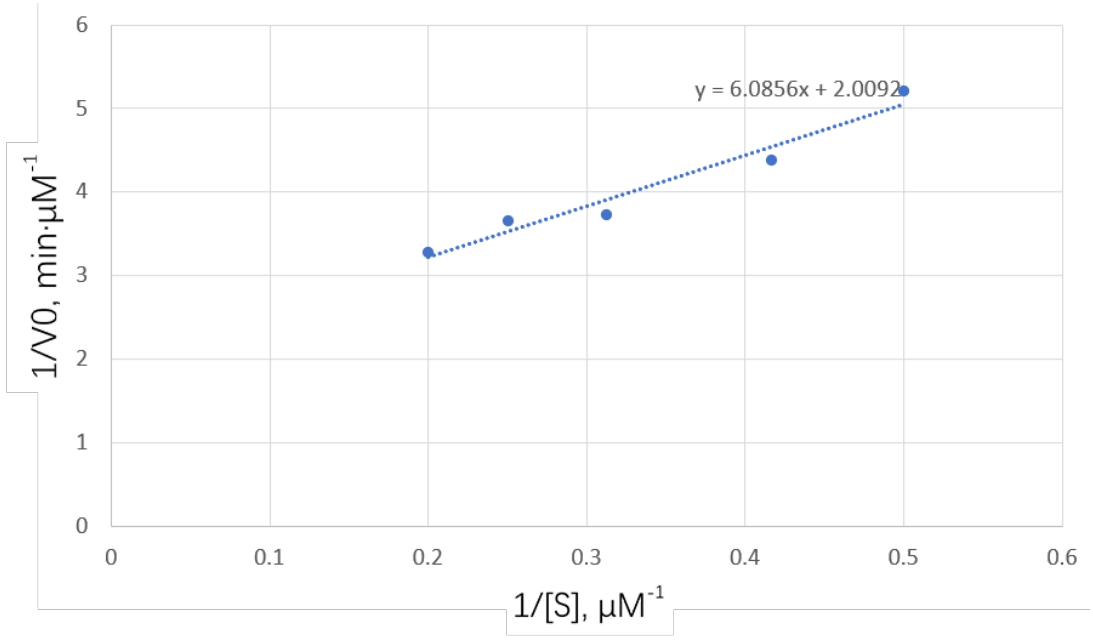


Figure S16. Lineweaver-Burk plot for ss-m66A, the concentration of enzyme used in reaction series is 20 uM.

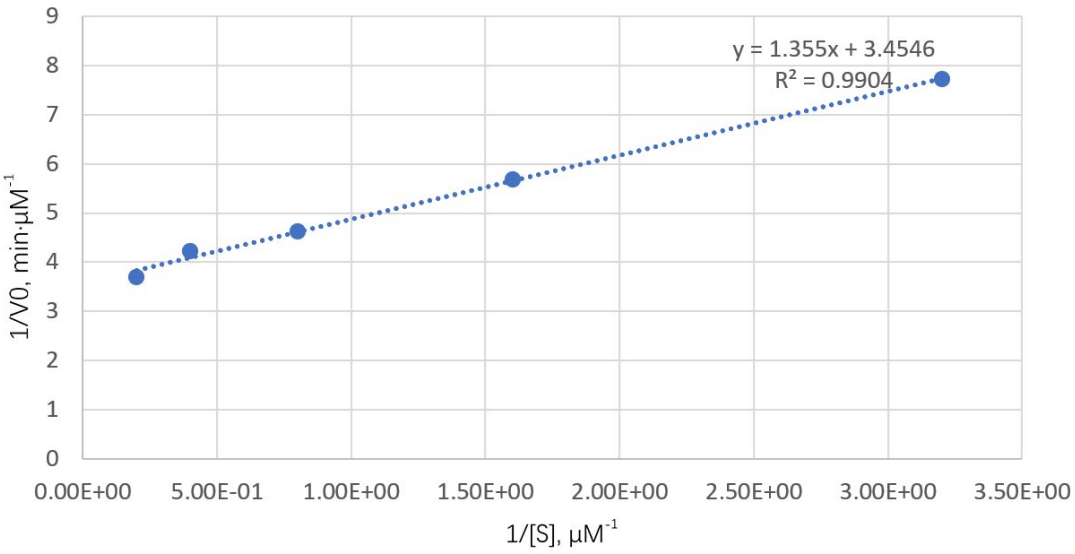


Figure S17. Lineweaver-Burk plot for ss-m1G, the concentration of enzyme used in reaction series is 20 uM.