BIOCHEMICAL STUDIES OF REPLICATION-DEPENDENT DNA REPAIR BY ENDONUCLEASE VIII-LIKE 3

By

Alyssa Aurora Rodriguez

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

June 30, 2021

Nashville, Tennessee

Approved:

Katherine L. Friedman, Ph.D.

Brandt F. Eichman, Ph.D.

Walter J. Chazin, Ph.D.

Yi Ren, Ph.D.

Jared Nordman, Ph.D.

To my querido Abuelito Jose C. Lopez who was the most hardworking and loving grandfather in the world. Thank you for your daily sacrifice for our family.

To my loving family who supported and inspired me every day to do and be better. You are my foundation in everything and I am truly blessed. My successes I share equally with you all.

Acknowledgements

I first want to thank my family for all their support and love. Mom and Dad thank you for always supporting me in all my ambitious plans, goals, and dreams. Thank you for always encouraging my learning and allowing my research to take me to institutions all across the world. Thank you for providing me with space and emotional, mental, and spiritual support to discover my sense of self and define my core values. Thank you for teaching me life skills to live independently and contribute to society. To my grandparents who have passed from this earth and are living, thank you for making the sacrifices in your life to come to this country for the betterment of your family and future children. Thank you for exemplifying what hard work and passion in your life looks like. I am your legacy and will continue to honor you all in my life. To my sisters, Sarina and Marisa thank you for all your patience and love though my studies over the past few decades. Thank you for always making me laugh and making me feel like the little kids we were every time I come home to the ranch. To my extended family across the country and in Mexico, my deepest appreciation for your prayers, love, and support. All your kind words have motivated me to continue in my studies. To all my friends who have become like family throughout my primary education all the way to graduate school, thank you for all the adventures. We have laughed, cried, rejoiced, and prayed together through a truly holistic experience of life. Thank you for always being flexible with my unpredictable lab schedule and always bringing your whole self to our gatherings. I will miss you all, as you have become my family in Nashville.

Thank you to my advisor Dr. Brandt F. Eichman for your patience, flexibility, and creativity in working with me as a graduate student in your lab. Thank you for your

availability to share data and brainstorm ideas as very visual thinkers. I appreciate and respect your professionalism in the scientific community. Thank you for always inviting my new ideas and supporting my career goals. You have been an exceptional example of a principal investigator and how I hope to run my lab as a mentor, educator, and researcher one day. To my committee, thank you for your time, energy, suggestions, feedback, and for challenging me to reach my upmost potential. I am also very appreciative to the Eichman lab members who made it easy to come to lab every day, including most weekends. The focused and hardworking atmosphere encouraged me to join the lab and excel. Everyone was open to brainstorm projects, ideas, and troubleshooting issues. I also would like to thank lab members and collaborators who directly worked on the NEIL3 project or contributed reagents, substrate, and protein.

I would also like to thank Beth Bowman from the IGP program for seeing my potential and first introducing me to Vanderbilt University as a place for my graduate school education. To Roger Chalkley, Linda Sealy, and Christina Keeton from the IMSD program, thank you for your countless hours of work and support to bring diverse students and perspectives to the science community at Vanderbilt University. Thank you for the first year of NIH funding through the IMSD program. You have changed so many lives and empowered so many students to change the world. I would also like to thank the Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS) for designing a professional society where Chicano/Hispanic scientists feel welcomed in the science community and inviting us to bring our multi-cultural identities to science. I loved all the SACNAS conference I have participated in San Antonio, Texas and Honolulu, Hawaii to present my research, meet potential graduate students, and

expand my professional development. Some of my fondest memories and travels are with the Vanderbilt SACNAS chapter members, thank you all. Thank you to the Center for Structural Biology and Molecular Biology Training Program and National Institutes of Health training grant for building a community of structural biologists, faculty and trainees at Vanderbilt University and for the funding during my second year in graduate school. My past three years of funding have come from the National Science Foundation, and I thank the Graduate Research Fellowship Program for their financial support for my research, educational, and professional development. I would also like to thank the science communication platform BioRender.com for my figures that I designed and created. Due to the web-based software I was able to upgrade my simple pencil-etched designs into beautiful and professional figures to share my research with fellow scientists and the general public.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5'-dRP: 5'-deoxyribose phosphate

5-OHC: 5-hydroxycytosine

5-OHU: 5-hydoxyuracil

6-4PPS: pyrimidine-(6-4)-pyrimidine

8-oxoA: 7,8-dihydro-8-oxoadenine

8-oxoG: 7,8-dihydro-8-oxoguanine

AA-ICL: acetaldehyde interstrand crosslink

AP: apurinic/apyrimidinic/abasic

APE1: AP endonuclease 1

APE2: AP endonuclease 2

AP-ICL: abasic interstrand crosslink

APTX: aprataxin

ATR: ataxia telangiectasia and Rad3-related protein

ATRIP: ATR-interacting protein

AZB: azinomycin B

BER: base excision repair

BLM: Bloom syndrome protein

βME: β-mercaptoethanol

BSA: bovine serum albumin

BTR: Bloom's syndrome complex

CCNA2: cyclin A2

CCNB2: cyclin B2

CDC45: cell division cycle 45

CHK1: checkpoint kinase 1

CG: cytosine glycol

CMG: CDC45-MCM2-7-GINS

CPD: cyclobutane pyrimidine dimer

Cryo-EM: cryo-electron microscopy

CS: Cockayne syndrome

CSB: Cockayne syndrome protein B

CUL4A: cullin 4A

dA-AP ICL: deoxyadenine abasic interstrand crosslink

dC-AP ICL: deoxycytosine abasic interstrand crosslink

DDB1: DNA damage-binding protein 1

DDR: DNA damage response

dG-AP ICL: deoxyguanine abasic interstrand crosslink

DHT: dihydrothymine

DHU: dihydrouracil

DNA: deoxyribonucleic acid

dN-AP ICL: deoxynucleotide abasic interstrand crosslink

DPC: DNA protein crosslink

dsDNA: double stranded DNA

DUB: deubiquitinating/deubiquitylating enzyme

E. coil: *Escherichia coli*

EEP: exonuclease/endonuclease/phosphodiesterase

EMSA: electrophoretic mobility shift assay

Endo III: endonuclease 3

ERCC1: excision repair cross-complementation group 1

ERCC4: excision repair cross-complementation group 4

ERI2: exoribonuclease family member 2

FA: Fanconi anemia

FAAP20: Fanconi anemia associated protein 20

FAAP24: Fanconi anemia associated protein 24

FAAP100: Fanconi anemia core complex associated protein 100

FANCA: Fanconi anemia protein A

FANCB: Fanconi anemia protein B

FANCC: Fanconi anemia protein C

FANCD: Fanconi anemia protein D

FANCE: Fanconi anemia protein E

FANCF: Fanconi anemia protein F

FANCG: Fanconi anemia protein G

FANCI: Fanconi anemia protein I

FANCL: Fanconi anemia protein L

FANCM: Fanconi anemia protein M

FANCV: Fanconi anemia protein V

FapyA: 4,6-diamino-5-formamidopyrimidine

FapyG: 2,6-diamino-4-hydroxy-5-formamidopyrimidine

FEN1: flap endonuclease 1

FlAsH: fluorescein arsenical hairpin

Fpg: formamidopyrimidine DNA glycosylase

FRET: fluorescence resonance energy transfer

GBM: glioblastoma multiform

GD: glycosylase domain

GG-NER: global genomic NER

Gh: guanidinohydantoin

GRF: glycine-arginine-x-phenylalanine

GRF-ZF: GRF zinc finger

GSH: glutathione

GST: glutathione-s-transferase

H_2O_2: hydrogen peroxide

H2TH: helix-two-turns-helix

hFPG2: human formamidopyrimidine 2 also known as NEIL3

HhH: helix-hairpin-helix

HLTF: helicase-like transcription factor

HMCES: 5-hydroxymethylcytosine binding, embryonic stem cell-specific

hNEIL3: human NEIL3

HOIL-1L: Heme-oxidized IRP2 ubiquitin ligase 1

HR: homologous recombination

ICL: interstrand crosslink

iPOND: isolation of proteins on nascent DNA

JNK: Jun N-terminal kinase

kDa: kilodalton

Lig1: ligase 1

Lig3: ligase 3

LUAD: lung adenocarcinoma

LUBAC: linear ubiquitin chain assembly complex

MMR: mismatch repair

Mms2: ubiquitin conjugating enzyme

MCM: mini chromosome maintenance

MHF1: FANCM interacting histone-fold protein 1

MHF2: FANCM interacting histone-fold protein 2

mNEIL3: mouse NEIL3

Nei: endonuclease VIII

NEIL: endonuclease VIII-like

NEIL1: endonuclease VIII-like 1

NEIL2: endonuclease VIII-like 2

NEIL3: endonuclease VIII-like 3

NEB: New England BioLabs

NER: nucleotide excision repair

NF-κB: nuclear factor κB

NHEJ: non-homologous end joining

Npl4: nuclear protein localization 4 protein

nt: nucleotide

Nth: endonuclease III

NZF: Npl4 zinc finger

O_2: superoxide

PCNA: proliferating cell nuclear antigen

PDB: protein data base

PIP: PCNA interacting protein

PMSF: phenylmethylsulfonyl fluoride

PNKP: polynucleotide kinase phosphatase

POL β: Polymerase beta

PTM: post-translational modification

Ranbp: RAN binding protein

RER: ribonucleotide excision repair

RNA: ribonucleic acid

RNAPII: RNA polymerase II

RNS: reactive nitrogen species

ROS: reactive oxygen species

RPA: replication protein A

RMSD: root-mean-square deviation

SAXS: small angle x-ray scattering

SD: standard deviation

SHG: second-harmonic generation

SMARCAL: SW1/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1

Sp: spiroiminodihydantoin

S phase: synthesis phase

ssDNA: single stranded DNA

T4 PNK: T4 DNA polynucleotide kinase

TAB: TAK1 binding protein

TAK1: transforming growth factor-β-activated kinase 1

TC-NER: transcription-coupled NER

TFIIH: transcription factor 2H

Tg: thymine glycol

THF: tetrahydrofuran

TLS: translesion synthesis

TOP1: topoisomerase 1

TOP2A: topoisomerase 2α

TOP3A: topoisomerase 3α

TOP3β: topoisomerase 3β

TRABID: tumor-necrosis factor receptor-associated factor-binding protein domain

TRAIP: TRAF interacting protein

TRF1: telomere-binding protein

TRFH: TRF homology protein

TTD: trichothiodystrophy

TTF2: transcription termination factor 2

UAF1: USP1-associated factor 1

Ub: ubiquitin

Uba1: ubiquitin-activating enzyme E1

Ubc13: E2 ubiquitin conjugating enzyme

UBD: ubiquitin binding domain

UBE2C: ubiquitin conjugating enzyme E2C

UBZ: ubiquitin binding zinc finger

UDG: uracil DNA glycosylase

Ufd1: ubiquitin fusion degradation 1

UHRF1: ubiquitin like with PHD and ring finger domains 1

USP1: ubiquitin carboxy-terminal hydrolase 1

UV-DDB2: ultraviolet light DNA-damage-binding 2

Vps36: vacuolar protein-sorting-associated protein

XP: xeroderma pigmentosum

XPA: xeroderma pigmentosum A

XPC: xeroderma pigmentosum C

XPF: xeroderma pigmentosum F

XPG: xeroderma pigmentosum G

XRCC1: x-ray repair cross complementing 1 protein

YB-1: Y-box-binding-protein 1

YTMA: 3-yatakemycinyl-2'-deoxyadenosine

ZF: zinc finger

ZGRF1: zinc GRF 1

ZR: zinc ribbon

ZRANB3: zinc finger RAN-binding domain-containing protein

Chapter 1

Introduction[1]

All genomic information is stored in DNA; therefore, its stability is of upmost importance for healthy replication and prevention of disease and aging. However, DNA is under constant attack from various DNA damaging agents with exogenous and endogenous sources. Additionally, there is a multitude of DNA repair pathways available to remove small nucleobase damaged lesions and more complex types of damage that span larger regions of DNA and even multiple DNA strands. Small DNA modifications to DNA nucleobases can be excised by DNA glycosylases in the base excision repair (BER) pathway. This dissertation will focus on the BER pathway including the sources of damage, repair mechanisms and initial damage recognition by DNA glycosylases. Additionally, I will outline the novel function of the endonuclease VIII-like (NEIL) glycosylases and describe their function in DNA repair.

Sources of DNA Damage

Exogenous sources include ultra-violet light from sunlight and ionizing radiation leading to depurination/depyrimidination or abasic (AP) sites, strand breakage, and bulky DNA lesions such as cyclobutane pyrimidine dimers (CPDs) that inhibit DNA replication by blocking RNA polymerase II (Altieri et al., 2008; Conconi and Bell, 2017).

[1] This work is published in part in Mullins, E.A., Rodriguez, A.A., Bradley, N.P., and Eichman, B.F. (2019) Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* 44, 765-781.

Environmental toxins include pesticides, toxic metals such as mercury, and cigarette smoke. Exogenous sources can also include alkylating chemotherapeutics and medicinal treatments. For example, psoralen is derived from plants and fungi and is used in the treatment of psoriasis and vitiligo (Cimino et al., 1985).

Small endogenous DNA nucleobase damage sources include alkylating agents, reactive oxygen species (ROS) and reactive nitrogen species (RNS). DNA consists of four nucleobases, purines include guanine and adenine and pyrimidines include thymine and cytosine. Minor DNA modifications such as alkylation events result in additional carbon and hydrogen atoms to DNA. Malondialdehyde and acrolein are endogenous bifunctional alkylating agents resulting from lipid peroxidation. Anti-tumor therapies such as mitomycin C and mechlorethamine are powerful DNA alkylating agents (Gates, 2009). Normal cellular processes such as cellular metabolism result in harmful oxidative by-products, such as ROS and RNS. ROS by-product molecules include free radical superoxide ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2) while RNS examples include dinitrogen trioxide (N_2O_3), nitrous acid (HNO_2), and nitric oxide (NO) (Altieri et al., 2008).

Types of DNA damage

Loss of genetic information from AP sites is cytotoxic to the cell. AP sites are abundant with as many as 10,000 AP sites per cell, per day and depurination is more common than depyrimidination (Lindahl and Nyberg, 1972). AP sites result from spontaneous or enzymatic loss of a DNA nucleobase from a weak covalent bond between the nitrogenous nucleobase and deoxyribose sugar (Altieri et al., 2008). AP sites are present in two forms, the major ring-closed acetal and minor ring-opened aldehyde

(Gates, 2009). Other forms of DNA damage include a strand cleavage event to the DNA phosphate backbone resulting in a single stranded DNA (ssDNA) damage break or two cleavage events on each of the two DNA strands leading to a double stranded DNA (dsDNA) break (Figure 1). All DNA atoms are susceptible to alkylation damage; however, the N7-guanine position is prone to alkylation damage due to its natural highly nucleophilic site. Other exocyclic nitrogen atoms on DNA nucleobases that are alkylated produce stable N6-adenine, N2-guanine, or N4-cytosine nucleobases (Gates, 2009). Oxidative sources from endogenous metabolic events result in small DNA oxidative damage including 5,6-dihydrothymine (DHT), thymine glycol (TG), cytosine glycol (CG), 5-hydroxycytosine (5-OHC), 8-oxo-7,8-dihydroguanine (8-oxoG), 7,8-dihydro-8-oxoadenine (8-oxoA), 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG), 4,6-diamino-5- formamidopyrimidine (FapyA), and the further oxidation products of 8-oxoG lesions, spiroiminodihydantoin (Sp) and guanidinohydantoin (Gh) (Altieri et al., 2008).

Bulky lesions include DNA intrastrand crosslinks and DNA interstrand crosslinks (ICLs) with the latter abbreviated as ICLs (Figure 1). DNA intrastrand crosslinks consist of a covalent bond between two nucleotides on the same strand of DNA. Intrastrand crosslinks are localized to one DNA strand as opposed to interstrand crosslinks that covalently tether two strands of DNA (Figure 1). The duplex spanning nature of ICLs block replication and transcription that require separation of the two DNA strands and halt these processes. ICL damage is detrimental to the cell if not corrected for replication and transcription to continue. DNA ICLs have exogenous as well as proposed endogenous sources.

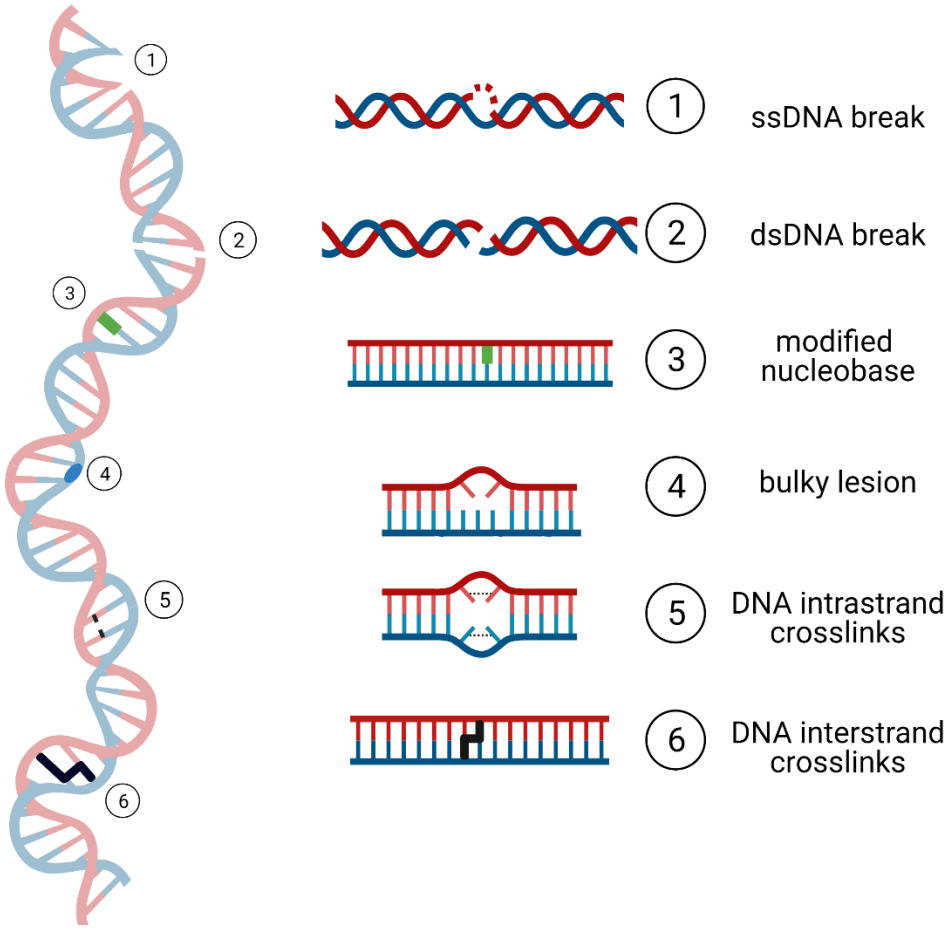

Figure 1. DNA damage is present in a variety of lesions within the genome. DNA lesions include 1) single stranded DNA (ssDNA) and 2) double stranded DNA (dsDNA) breaks. 3) Small modified nucleobases include methylated, alkylated, and oxidative damage (green square) while complex DNA damage includes 4) bulky lesions, 5) intrastrand crosslinks and 6) interstrand crosslinks.

Both cisplatin-ICLs and psoralen-ICLs were repaired in a replication-dependent manner requiring converged replication forks to stall at the ICL damage site (Semlow et al., 2016). However, the repair of these two lesions after replication fork collision differs substantially, with either incision independent or dependent pathways, and will be discussed in detail in Figure 3. Platinum compounds such as cisplatin result in intrastrand

and interstrand crosslinks. These compounds include cisplatin, carboplatin, and oxaliplatin which create a covalent bond between the N7 of two guanine nucleotides (Semlow and Walter, 2021). DNA crosslinking agents cisplatin, carboplatin, oxaliplatin, satraplatin, and picoplatin, all of which are platinum containing, have successful clinical applications for testicular, ovarian, non-small-cell lung, colorectal, prostate, and breast cancer (Deans and West, 2011).

Bifunctional alkylating agents that result in ICLs include nitrogen mustards, such as mechlorethamine, which was first introduced during World War II as a mustard gas biowarfare agent (Guainazzi and Scharer, 2010; Raschle et al., 2008). Nitrogen mustards chlorambucil, cyclophosphamide and melphalan create aziridinium ions that alkylate guanine, adenine, and cytosine DNA nucleotides to form ICLs (Semlow and Walter, 2021). DNA ICLs have been very successful in clinical applications as chemotherapeutics. Nitrogen mustards such as cyclophosphamide, melphalan, chlorambucil, and ifosfamide have been clinically utilized to treat lymphoma, myeloma, leukemia, and non-small-cell lung cancer respectively (Deans and West, 2011; Guainazzi and Scharer, 2010). Mitomycin C has been successful as a chemotherapeutic to treat esophageal and bladder cancer. However, some of these treatments result in harmful side-effects such as neutropenia and leukopenia and further investigation is necessary to increase specificity of chemotherapy delivery to cancer cells (Deans and West, 2011). Psoralen-ICLs have historically been used as a skin treatment for psoriasis, as shown by ancient Egyptians who would harvest the bishop's weed (*Ammi majus*) for psoralen and, under sunlight exposure, treat skin diseases such as psoriasis and vitiligo (Hashimoto et al., 2016; Semlow and Walter, 2021). The planar psoralen compound and other psoralen-

derivatives, trimethylpsoralen and 8-methoxypsoralen, contain furan rings. Psoralen intercalates between thymine DNA nucleotides and upon UV-irradiation between 320-400 nm forms psoralen-ICLs within duplex DNA. Psoralen is also capable of intercalating into RNA (Cimino et al., 1985).

Endogenous sources of DNA ICLs include aldehydes, ROS, AP sites, and DNA-protein crosslinks (DPC) (Housh et al., 2021). Endogenous aldehydes such a formaldehyde, acetaldehyde, and malondialdehyde have been shown to form aldehyde DNA ICLs (AA-ICL) between the exocyclic amines of purine nucleotides (Amunugama and Walter, 2020; Semlow and Walter, 2021). DPCs include DNA covalently bound to protein and also cause a bulky barrier to DNA replication preventing DNA strand separation. Interestingly, some enzymes create DPCs as a lesion protection mechanism such as the enzyme 5-hydroxymethylcytosine binding, embryonic stem cell-specific (HMCES) that protects AP sites in DNA (Amidon and Eichman, 2020). HMCES contains a SOS-response associated peptidase (SRAP) domain and *E. coli* ortholog YedK, which is similar in sequence and structure to SRAP, forms a DPC intermediate via a thiazolidine linkage between an AP site on DNA and N-terminal cysteine residue within YedK. This DPC intermediate is thought to protect the AP site from endonuclease cleavage and nucleotide incorporation by error-prone polymerases (Thompson et al., 2019).

The experimental study of ICLs is limited to the *in vitro* synthesis of model ICLs. One such class of recapitulated ICLs possibly present *in vivo* include AP-ICLs. Since AP sites are abundant in human cells, the probability of an AP site forming an ICL is high. The synthesis of these ICLs will aid in their detection in cells and ICL synthesis and optimization is under investigation (Johnson et al., 2013; Price et al., 2014). There are

currently three forms of synthesized ICLs that contain a covalent bond between an AP site and a nucleotide on the opposing strand. The deoxyguanine AP (dG-AP) ICLs consist of a covalent bond between an AP site and N2 of guanine located 3' to the AP site on the opposite DNA strand. dG-AP ICLs have been synthesized with 2-3% yield (Johnson et al., 2013). Higher DNA ICL yields of 15-70% have resulted from synthesis of a covalent bond between the AP site and N6 amino group of adenine located 3' to AP site on opposite strand, resulting in a deoxyadenine AP (dA-AP) ICL (Price et al., 2015; Price et al., 2014). Recently, a deoxycytosine AP (dC-AP) ICL was synthesized with 15% yield and shows a covalent bond between the AP site and cytosine nucleotide 3' to the AP-site on the opposite DNA strand (Varela et al., 2021). The recent publications of synthesized deoxynucleotide AP (dN-AP) ICLs are revealing the numerous AP-ICLs that may form *in vivo.* Confirmation that these ICLs exist in cells is key to understanding the repair mechanisms of these complex DNA lesions.

DNA Repair Pathways

Cells are equipped with a multitude of DNA repair pathways to excise, cleave, unhook, and bypass various types of DNA lesions. Some of these repair pathways include BER to remove small modified DNA nucleobases, mismatch repair (MMR) to fix non-complementary nucleobases, ribonucleotide excision repair (RER) to remove ribonucleic acids within DNA, double stranded break repair such as the non-homologous end joining (NHEJ) and homologous recombination (HR) pathways, nucleotide excision repair (NER) to remove bulky lesions, and ICL repair to unhook and excise ICLs. The broad range of DNA repair pathways has been published (Fagbemi et al., 2011; Gates, 2009; Kellner

and Luke, 2020; Krokan and Bjoras, 2013; Kunkel and Erie, 2005; Sancar and Sancar, 1988; Semlow and Walter, 2021; Shrivastav et al., 2008). The focus of my dissertation will be on BER, NER, and ICL repair because each repair pathway is related to the removal of both simple and complex DNA lesions, such as base modifications and DNA ICLs.

Base Excision Repair pathway

The BER pathway is conserved from prokaryotes to eukaryotes. As the first enzyme in the BER pathway, DNA glycosylases recognize and excise minimal helix-distorting damage in the form of a modified DNA nucleobase. There are two main types of DNA glycosylases, monofunctional and bifunctional. Monofunctional glycosylases perform a hydrolysis reaction which results in an oxocarbenium intermediate and cleavage of the N-glycosidic bond leaving an AP site and free nucleobase (Figure 2A, B) (Mullins et al., 2019). DNA glycosylases are product-inhibited, so they bind tightly to the AP site product. AP endonuclease 1 (APE1) cleaves the DNA-phosphate backbone 5' to the AP site creating a ssDNA break with ends containing a 3'-hydroxyl and 5'-sugar phosphate known as a 5'-deoxyribose phosphate (5'-dRP). Short and long patch BER pathways follow, depending on the presence of a nick or gap introduced into the DNA after nucleobase excision, respectively. Short patch is limited to one modified nucleobase as subject to excision. Long patch BER includes incorporation of 2-10 nucleotides by Polymerase β (POLβ) resulting in a 5' overhang that is cleaved by flap endonuclease 1

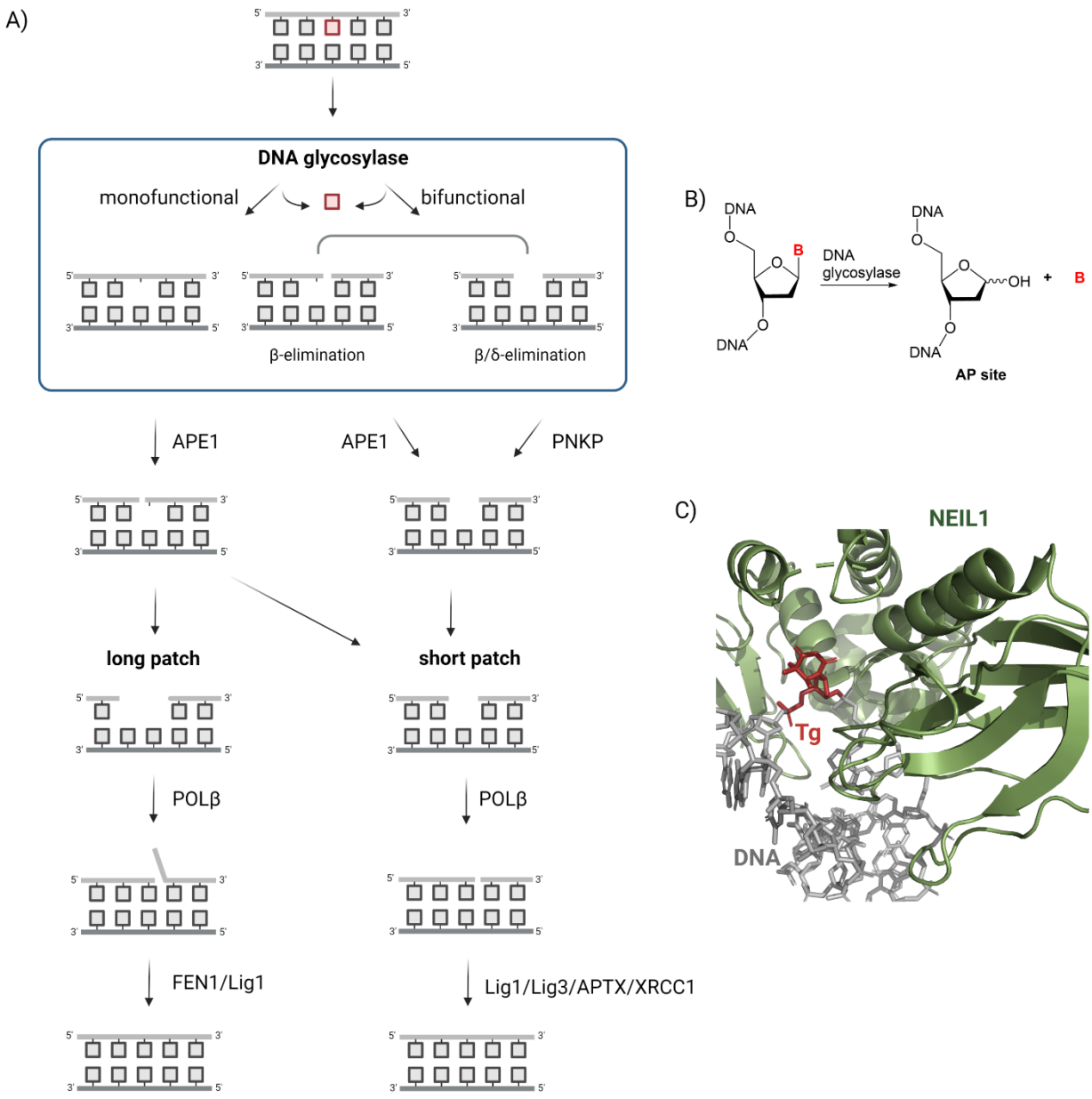
A)
DNA glycosylase
monofunctional
bifunctional
β-elimination
β/δ-elimination
APE1
APE1
PNKP
long patch
short patch
POLβ
POLβ
FEN1/Lig1
Lig1/Lig3/APTX/XRCC1
B)
DNA
B
DNA glycosylase
OH
+
B
AP site
C)
NEIL1
Tg
DNA

Figure 2. The base excision repair pathway is initiated by DNA glycosylases. A) The base excision repair (BER) pathway begins with recognition and excision of damaged nucleobase (red square) by a monofunctional or bifunctional DNA glycosylase. Monofunctional glycosylases perform excision by hydrolysis, leaving an abasic (AP) site. Bifunctional glycosylases contain additional lyase activity leaving β- or β/δ-elimination products. Long patch BER includes filling of the DNA gap by polymerase β (POLβ) followed by overhang cleavage by FEN1 and ligation of the DNA phosphate backbone by ligase 1 (Lig1). Short patch BER includes nucleotide incorporation by POLβ followed by ligation of the DNA phosphate backbone by ligase 1 or ligase 3 (Lig1, Lig3). B) A DNA glycosylase cleaves the N-glycosidic bond between the deoxyribose sugar and nucleobase producing an AP site and cleaved nucleobase. C) Crystal structure of bifunctional DNA glycosylase NEIL1 (green) in complex with duplex DNA (gray) containing oxidative lesion thymine glycol (Tg) (red) flipped into the active site (PDB 5ITY).

(FEN1) (Caldecott, 2020). The DNA backbone is sealed by DNA ligase 1 (Lig1) leaving a complete non-damaged DNA duplex (Figure 2A).

Like monofunctional glycosylases, bifunctional DNA glycosylases also cleave the N-glycosidic bond, but they contain additional β-lyase activity, cleaving the phosphodiester DNA backbone (Figure 2A). The amino group from a lysine residue or α-amino of the N-terminal amino acid acts as a nucleophile to form an iminium intermediate (Berti and McCann, 2006). Bifunctional DNA glycosylases cleave 3' and/or 5' to the AP site, resulting in β- and β/δ-elimination products respectively (Berti and McCann, 2006). Cleavage of the phosphodiester DNA backbone 3' to the AP site results in a ssDNA break or β-elimination product consisting of 3'-phospho-α,β-unsaturated aldehyde and 5'-phosphoryl group. In short patch BER, APE1 can further cleave the DNA backbone. The β/δ-elimination products result from further 5' cleavage to the α,β-unsaturated aldehyde moiety leads to a 3'- phosphate which must be removed by polynucleotide kinase/phosphatase (PNKP) in the short-patch BER pathway leaving a 3'-hydroxyl. POLβ incorporates the correct nucleotides and the DNA phosphate backbone nick is sealed by

DNA ligase 1 (Lig1) or ligase 3 (Lig 3) with scaffold protein X-ray repair cross complementing 1 protein (XRCC1) and accessory protein (APTX) (Figure 2A) (Altieri et al., 2008; Krokan and Bjoras, 2013).

DNA Glycosylases

In 1974 Tomas Lindahl discovered the first DNA glycosylase, *Escherichia coli* (*E. coli)* uracil DNA glycosylase (UDG). Lindahl was awarded the Noble Prize in Chemistry in 2015 along with chemists Paul L. Modrich and Aziz Sancar for their research on DNA repair. Cytosine deamination results in an uracil nucleobase within DNA and a DNA glycosylase was found to cleave between the deoxyribose sugar and nucleobase as the first enzyme in the BER pathway (Krokan and Bjoras, 2013). DNA glycosylases recognize DNA lesions by detecting uncommon base-stacking, base-pairing and solvation interactions (Jiang et al., 2001; Ramstein and Lavery, 1988; Roberts and Cheng, 1998; Stivers, 2004; Yang, 2006). The two-step mechanism of BER involves cleavage of the nucleobase prior to nucleophilic attack leading to an oxocarbenium intermediate. Catalytic residues required for BER include carboxylate functional groups on aspartate and glutamate and carboxamide functional groups on asparagine and glutamine. The DNA glycosylase cleaves the N-glycosidic bond between the nucleobase and C1' of the deoxyribose sugar (Mullins et al., 2019).

Historically, most DNA glycosylases were thought to follow a mechanism whereby the damaged nucleobase flips out of the DNA, via helical bending and widening, into the active site of the DNA glycosylase. The DNA glycosylase inserts intercalating residues into the position of the absent “flipped out” nucleobase to stabilize the helix. For example,

a crystal structure was determined of bifunctional DNA glycosylase endonuclease VIII-like 1 (NEIL1) bound to duplex DNA containing a flipped-out thymine glycol (Tg) nucleobase into the NEIL1 active site (Figure 2C) (Zhu et al., 2016). Typically, the active site of a DNA glycosylase can only accommodate one damaged nucleobase at most. Therefore, DNA glycosylases were traditionally thought to only excise small DNA modifications.

However, new excision mechanisms by DNA glycosylases have been discovered that do not include base-flipping. These mechanisms have been discovered in bacteria and vertebrates. Bacterial DNA glycosylase AlkD was shown to excise bulky 3-yatakemycinyl-2'-deoxyadenosine (YTMA) DNA lesions without a base flipping mechanism (Mullins et al., 2017). AlkD does not structurally contain an active site that can accommodate one flipped out nucleobase and is missing available DNA intercalating residues that would replace the flipped out DNA nucleobase. Bacterial DNA glycosylase AlkZ also performs a non-base flipping mechanism to unhook azinomycin (AZB) DNA ICLs. Mutational analysis and computational rigid-body docking identified three catalytic elements within the active site of AlkZ to bind DNA and catalyze glycosylase activity (Mullins et al., 2019; Mullins et al., 2017). These three catalytic elements included the β-hairpin, glutamine 37, and glutamine 39 of the DNA binding cleft. Vertebrate NEIL1 and endonuclease VIII-like 3 (NEIL3) have been shown to excise bulky adducts and unhook psoralen-ICLs and AP-ICLs which will be discussed further in the following sections.

Nucleotide Excision Repair

Most bulky DNA lesions are helix-distorting and removed by the NER pathway which consists of the coordination and regulation of approximately 20-30 proteins (Altieri et al., 2008; Sugitani et al., 2016). NER removes bulky adducts from DNA such as photoproducts CPD and pyrimidine-(6-4)-pyrimidine (6-4PPS). There are three diseases [xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD)] associated with NER (Berneburg and Lehmann, 2001). XP is characterized by human patients with UV-light sensitivity leading to skin cancer susceptibility. CS patients suffer from dwarfism, neurological abnormalities such as demyelination, sun sensitivity and ocular issues such as cataracts. TTD patients show growth abnormalities, brittle hair, and sun sensitivity. NER removes bulky lesions by recognition of the DNA damage, complex assembly by NER proteins, endonucleases cleave a few nucleotides away from the lesion incising 5' and 3' to the lesion, resulting in a DNA gap which is filled by DNA polymerases and the DNA phosphate backbone is sealed by a DNA ligase (Altieri et al., 2008).

In humans, the NER pathway is split into two sub-pathways, global genomic NER (GG-NER) and transcription-coupled NER (TC-NER). NER begins with recognition of the bulky lesion by xeroderma pigmentosum C (XPC) and ultraviolet light, DNA-damage-binding 2 (UV-DDB2) proteins. Mutation of the XP proteins results in the XP disease phenotype where patients are sensitive to UV light and have a high probability of developing skin cancer (Sugitani et al., 2016). The ubiquitin ligase complex made up of cullin 4A (CUL4A), DNA damage-binding protein 1 (DDB1), and regulator of cullins 1 (ROC1) ubiquitylates XPC and UV-DDB2 to increase the affinity of XPC for DNA (Tapryal

et al., 2021). To verify the presence of DNA damage, the transcription factor 2H (TFIIH) complex is recruited to the lesion and separates the duplex DNA strands next to the lesion via its helicase activity, forming a NER bubble structure. Assisting in the positioning of the endonuclease proteins are xeroderma pigmentosum A (XPA) and single stranded binding protein replication protein A (RPA). XPA-RPA and xeroderma pigmentosum G (XPG) form the pre-incision complex. Endonucleases excision repair cross-complementation group 1 (ERCC1)-xeroderma pigmentosum F (XPF) and XPG cleave 5' and 3' to the bulky DNA lesion leaving an approximately 30 nucleotide long gap in one DNA strand. DNA polymerases δ, ε, or κ fill in the gap following by ligation by ligase I or ligase IIIα (Fagbemi et al., 2011).

DNA Interstrand Crosslink Repair

The cell cycle is made up of multiple phases including interphase (G1, S, G2) and mitosis. The different phases of the cell cycle are not exempt from DNA damage. DNA replication occurs during the synthesis (S) phase in which the replisome assembles and copies the genome. Damage at the replication fork includes ICLs and DPCs that prevent DNA strand separation and halt replication when encountered by a replication fork (Ashour and Mosammaparast, 2021). ICL damage at converged replication forks has been investigated and DNA glycosylases will be discussed below with their role in DNA replication (Zhang et al., 2015). ICL repair includes a combination of pathways dependent on the type of ICL lesion and cell cycle stage. Secondary or back-up pathways are also available to repair the DNA ICL if the primary repair pathway is unsuccessful. Similar to the NER pathway, ICL repair pathways can result in two incision events. In vertebrates'

cells during S-phase, ICLs are typically resolved by the Fanconi Anemia (FA) pathway. The autosomal recessive FA disease was first observed in 1927 by Swiss pediatrician Guido Fanconi who saw a pattern in patients suffering from bone marrow failure, predisposition to cancer, anemia, birth defects and sensitivity to DNA ICL agents. (Altieri et al., 2008; Auerbach, 1995; Fiesco-Roa et al., 2019; Semlow and Walter, 2021). There are currently 22 *FANC* gene mutations that result in the FA disease. A patient is typically diagnosed with FA disease after thorough examination by a physician for phenotypic characteristics such as café au lait spots on skin, abnormal thumbs or polydactyly, short statue, and microcephaly. To determine the FA mutation, DNA crosslinking agents such as diepoxybutan (DEB) or mitomycin C (MMC) are added to a sample of patient lymphocyte culture to induce cell sensitivity, the cells are retrovirally transfected for the known FA genes which complement the cell sensitivity, and gene sequenced to confirm FA (Shimamura and Alter, 2010).

The FA pathway includes complex, regulated coordination of crosstalk between several FANC proteins and proteins involved in NER, translesion synthesis (TLS), and HR repair pathways. The FA pathway begins with recognition of the ICL lesion on DNA by Fanconi anemia proteins (FANC) such as FANCM with histone fold proteins FANCM interacting histone-fold protein 1-2 (MHF1-MHF2). FANCM signals the recruitment of the FANCD2-FANCI complex to DNA and recruits the FA protein complex made up of FANCA, B, C, E, F, G, L, M, and Fanconi anemia core complex associated protein 100 (FAAP100), FA-associated protein 20 (FAAP20), and FA-associated protein 24 (FAAP24) (Ashour and Mosammaparast, 2021; Ceccaldi et al., 2016; Deans and West, 2011; Rodriguez and D'Andrea, 2017). The E3 ubiquitin ligase FANCL within the FA core

complex activates the ataxia telangiectasia and RAD3-related (ATR) checkpoint response that monoubiquitylates FANCD2-FANCI. The ubiquitylated FANCD2-FANCI complex recruits scaffolding protein SLX4/FANCP for incision by endonuclease FANCQ /XPF/ERCC4 producing a DNA adduct and double stranded DNA break. TLS follows with TLS-specific polymerases that use the cognate lesion as a template for nucleotide incorporation (Powers and Washington, 2018). For the FA pathway, the TLS components include non-traditional DNA polymerase REV1, Pol ζ accessory and catalytic subunits REV7/FANCV and REV3 respectively (Rodriguez and D'Andrea, 2017). The Bloom's syndrome complex (BTR) consisting of RMI2, Bloom syndrome protein (BLM), RMI1, topoisomerase 3α (TOP3A) is recruited and signals HR associated proteins to repair the double stranded break. HR includes 5' resection at the double stranded break leaving 3' ssDNA which is annealed by Rad51 to a homologous DNA sequence, followed by strand exchange leading to a D-loop intermediate and DNA synthesis (Prado, 2018). The FA pathway concludes with deubiquitylation of FANCD-FANCI by deubiquitylating enzyme ubiquitin carboxy-terminal hydrolase 1 (USP1) and USP1-associated factor 1 (UAF1) (Ashour and Mosammaparast, 2021; Ceccaldi et al., 2016; Deans and West, 2011).

In 2016, the DNA glycosylase NEIL3 was shown to unhook ICLs *in vitro,* shedding light on the first mammalian glycosylase involved with ICL repair. NEIL3 unhooks psoralen-ICLs and dA-AP ICLs at the non-native glycosidic bond leaving an AP site and monoadduct in an incision independent manner, compared to the FA pathway that contains dual incision events (Semlow et al., 2016). In 2019, NEIL3 was shown to unhook ICLs from a converging replication fork in *Xenopus* egg extracts and this incision-

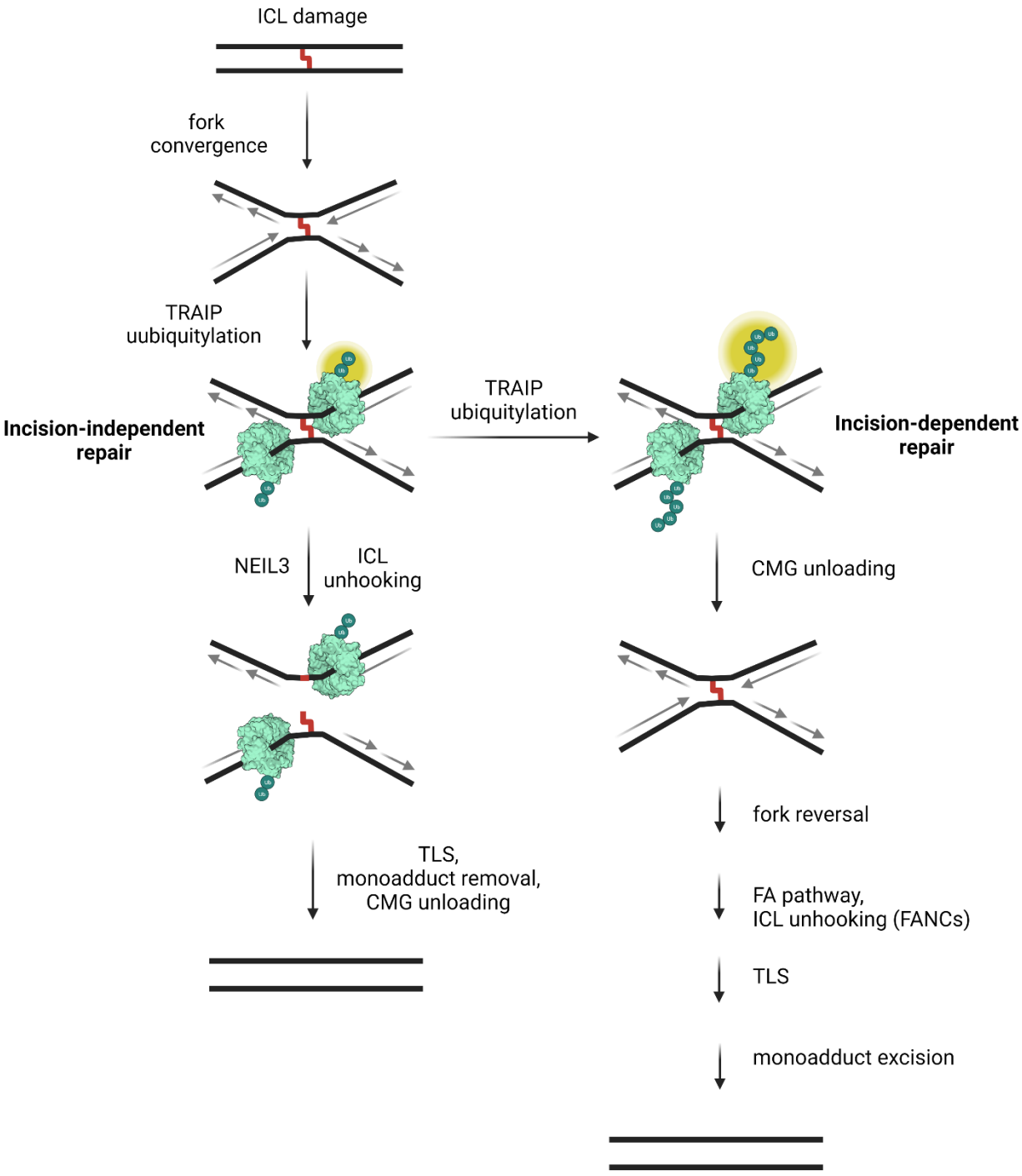
ICL damage
fork convergence
TRAIP uubiquitylation
Incision-independent repair
TRAIP ubiquitylation
Incision-dependent repair
NEIL3
ICL unhooking
CMG unloading
TLS, monoadduct removal, CMG unloading
fork reversal
FA pathway, ICL unhooking (FANCs)
TLS
monoadduct excision

Figure 3. Replication-coupled interstrand crosslink repair. ICL DNA damage covalently links two strands of duplex DNA, leading to replication-dependent fork convergence. TRAIP E3 ubiquitin ligase ubiquitylates replicative CMG helicase (blue, PDB 6SKO) with short or long chains of polyubiquitin (blue circles). The first pathway for ICL removal is the incision-independent repair pathway whereby DNA glycosylase NEIL3 unhooks the ICL resulting in an AP site and monoadduct (red). The lesion is bypassed by TLS, monoadduct removed and CMG unloaded from the DNA. Longer polyubiquitin chains by TRAIP leads to incision-dependent repair consisting of CMG unloading, fork reversal, the FA pathway consisting of the coordination of dozens of FANC proteins, TLS, and monoadduct excision. Data information shown in the figure is from published work (Semlow and Walter, 2021; Wu et al., 2019).

independent NEIL3 pathway was preferred over the dual incision FA pathway (Wu et al., 2019). TRAF interacting protein (TRAIP) is the E3 ubiquitin ligase that adds ubiquitin chains to the mini chromosome maintenance 2 (MCM2), MCM3, MCM4, MCM6, MCM7, and CDC45 subunits of replicative (CDC45-MCM2-7-GINS) CMG helicase and dictates the ICL repair pathway, either NEIL3 or FA. Shorter ubiquitin chains on CMG signal ICL repair via the NEIL3 pathway while longer ubiquitin chains lead to the FA pathway (Wu et al., 2019) (Figure 3). The E2-ubiquitin conjugation enzyme is unknown as well as how this enzyme distinguishes between short or long ubiquitin chains. TRAIP travels with the replisome and ubiquitylates CMG after fork convergence at an ICL lesion. CMG does not need to be unloaded from the leading strand of the converged forks for ICL repair to progress (Semlow et al., 2016).

Endonuclease VIII-like (NEILs) DNA Glycosylases

Endonuclease VIII-like 1 (NEIL1)

DNA glycosylases that excise oxidized damage include two structural families, the helix-hairpin-helix (HhH) family and the helix-two-turns-helix (H2TH) family. HhH family members include *E. coli* endonuclease III (Nth) which excises oxidative lesions. H2HT family members include *E. coli* endonuclease VIII (Nei), and formamidopyrimidine (Fpg). Three Fpg/Nei homologs were identified in vertebrates and named endonuclease VIII-like (NEIL) 1,2, and 3 (Bandaru et al., 2002; Hazra et al., 2002; Morland et al., 2002; Wallace et al., 2003). NEIL1 expression levels are cell cycle dependent with highest expression during S phase (Hazra et al., 2002; Hegde et al., 2013). Non-peer reviewed mouse studies have shown that NEIL1 has a role in memory and anxiety regulation (Hildrestrand, 2021).

NEIL1 was discovered as the mammalian homolog to Nth, a HhH family DNA glycosylase with activity toward Tg lesions, and in $Nth^{-/-}$ knock-out mouse cell extracts endogenous glycosylase activity toward Tg or 5-OHU nucleobase modifications was shown (Takao et al., 2002). Since Nth was not present in the cell extracts, the resulting DNA glycosylase NEIL1 was identified and glycosylase activity confirmed with recombinant purified NEIL1. Other oxidative DNA lesions that recombinant bifunctional NEIL1 has shown to excise include Tg, Sp, Gh, DHT, dihydrouracil (DHU), FapyG, FapyA, 5-OHU, 8-oxoG, and AP sites within duplex DNA, bubbles, forks, and single stranded DNA (Hazra et al., 2002; Liu et al., 2013a; Parsons et al., 2007; Rosenquist et al., 2003). NEIL1 has also shown activity toward duplex DNA containing nitrogen mustard-Fapy and

aflatoxin-Fapy adducts *in vitro* and 3-stranded DNA with psoralen-ICL adduct (Couve et al., 2009; McNeill et al., 2013; Minko et al., 2019).

Endonuclease VIII-like 2 (NEIL2)

Bifunctional DNA glycosylase NEIL2 associates with a multitude of cellular processes including genome maintenance and repair and cellular inflammatory response (Sarker et al., 2021). NEIL2 removes oxidative damage from ssDNA and DNA intermediates such as transcription bubbles and its expression is cell cycle independent. NEIL2 interacts with transcription associated proteins such as RNA polymerase II (RNAPII), transcription factor Y-box-binding-protein 1 (YB-1) and Cockayne syndrome protein B (CSB) (Das et al., 2007; Sarker et al., 2021). Therefore, NEIL2 has a speculated role in the NER sub-pathway of transcription-couple repair (Sarker et al., 2021). NEIL2 excises oxidative lesions Sp, Gh, DHT, DHU, 5-OHU, 5-OHC, and Tg from bubble, fork, ssDNA and dsDNA (Liu et al., 2013a). Non-peer reviewed studies have shown early links between NEIL2 and memory and anxiety in mouse models (Hildrestrand, 2021). Neil1$^{-/-}$ and Neil2$^{-/-}$ knock out mice showed hyperactivity, decreased anxiety and increased learning behavioral phenotype after locomotor activity tests.

Endonuclease VIII-like 3 (NEIL3)

NEIL3 was previously known as Formamidopyrimidine 2 (hFPG2) due to homology to the Fpg DNA glycosylases (Morland et al., 2002). NEIL3 is expressed in the thymus, spleen, and bone marrow, and brain in mice and spleen, testis, and ovaries in humans (Hildrestrand et al., 2009; Liu et al., 2013a; Morland et al., 2002; Torisu et al., 2005).

Human NEIL3 expression was not detected by Northern blot in the brain, however a more sensitive detection method may be required to identify low levels of NEIL3 expression (Rolseth et al., 2008). NEIL3 was detected in mouse brains in progenitor cells with NEIL3 expression decreasing with age and limited to early development of less than 2 weeks of life in mouse studies (Rolseth et al., 2008). NEIL3 shows overexpression in many human cancer types including brain, urinary bladder, breast, lung, and prostate cancer (Figure 4) (Shinmura et al., 2016; Tran et al., 2020). Not surprisingly in mice NEIL3 has shown high expression in the brain where the environment is abundant in oxidative damage from the presence of ROS (Krokan and Bjoras, 2013). Likewise, NEIL3 is highly expressed in glioblastoma multiform (GBM) in humans where ROS are common and a study with 8,662 patients showed 89% of GBM patients had high NEIL3 expression levels (Tran et al., 2020). NEIL3 has shown a role in spatial stability of neurons involved with learning and memory in a non-peer reviewed study (Kunath, 2021). Hippocampal neuronal place cells contain spatial information in the form of patterned neuron firing. Neil3$^{-/-}$ in neurons from mice showed consistent remapping of neuronal firing but show an inability to fire new maps, as evidenced by spatial exploration tests in mice. This study shows a new NEIL3 spatial recognition role stemmed from neurons affecting environment recognition and exploration at the organismal level (Kunath, 2021).

NEIL3 has also shown a role in atherogenesis, pulmonary function, myocardial infarction, ischemic stroke, auto-immunity, and Neil3 is downregulated in endometriosis

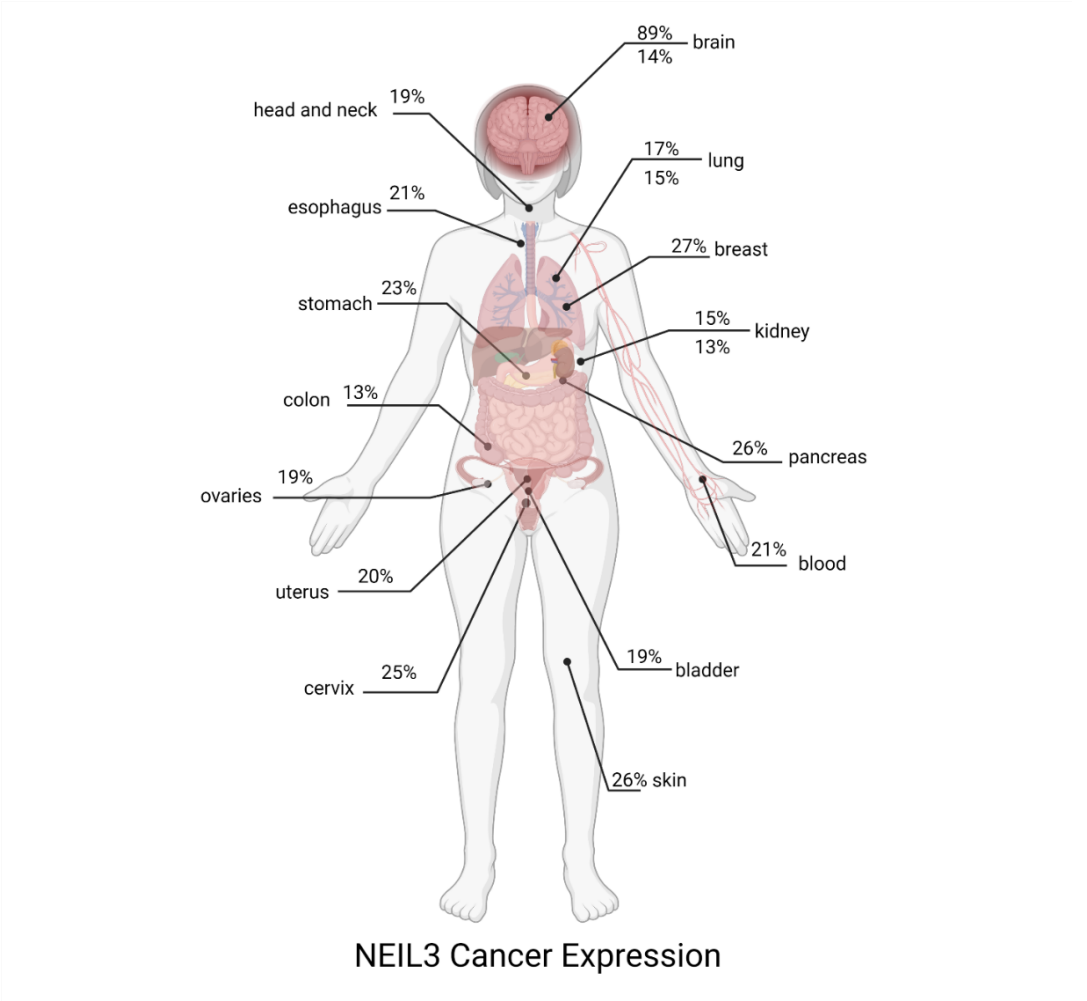

Figure 4. NEIL3 is highly expressed in many cancer types. Percentages show NEIL3 expression from 8,662 cancer patients. Overlapping cancer types in the same tissue/organ are shown as two numbers. Cancer types include brain (glioblastoma multiform, 89%, low grade glioma, 14%), lung (squamous cell carcinoma, 17%, lung adenocarcinoma, 15%), breast (invasive carcinoma 27%), kidney (renal papillary cell carcinoma, 15%, renal clear cell carcinoma, 13%), pancreas (adenocarcinoma, 26%), blood (acute myeloid leukemia, 21%), bladder (urothelial carcinoma, 19%), skin (cutaneous melanoma, 26%), cervix (squamous cell carcinoma, 25%), uterus (corpus endometrial carcinoma, 20%), ovaries (serous cystadenocarcinoma, 19%), colon (adenocarcinoma, 13%), stomach (adenocarcinoma, 23%), esophagus (carcinoma, 21%), head and neck (squamous cell, 19%). All data is from cBioPortal and publication (Tran et al., 2020).

and chlamydia (Chakraborty et al., 2015; de Sousa et al., 2017; Ehlers et al., 2016; He et al., 2016; Jalland et al., 2016; Li et al., 2018; Massaad et al., 2016; Olsen et al., 2017; Poli-Neto et al., 2021a; Quiles-Jimenez et al., 2021; Rognlien et al., 2015; Rumsey et al., 2017; Skarpengland et al., 2015; Stratigopoulou et al., 2020; Tangye, 2016; Yang et al., 2016). The substrate specificity of NEIL3 includes most oxidative lesions that can be excised by NEIL1 and NEIL2 except that NEIL3 has preference for the further oxidation products of 8-oxoG, the hydantoin lesions, over 8-oxoG (Figure 5) (Imani Nejad et al., 2020; Liu et al., 2010; Liu et al., 2012; Liu et al., 2013a; Liu et al., 2013b; Morland et al., 2002; Semlow et al., 2016; Takao et al., 2009). NEIL3 is highly expressed in proliferating cells such as stem cells and cells undergoing DNA synthesis.

Concurrently, NEIL3 is present at the replication fork as shown by isolation of proteins on nascent DNA (iPOND) and supported by the observation that replication fork progression is decreased in the absence of NEIL3 (Klattenhoff et al., 2017). NEIL3 is the first vertebrate DNA glycosylase shown to unhook DNA ICLs such as psoralen-ICLs and AP-ICLs without the FA proteins (Figure 5) (Semlow et al., 2016; Wu et al., 2019). This astonishing result has led to multiple publications investigating the NEIL3 repair mechanism (Li et al., 2020; Semlow et al., 2016; Wu et al., 2019). Chapter 2 investigates the substrate specificity of the NEIL3 glycosylase domain (GD) toward model replication forks containing a DNA ICL.

Structural comparison of the NEIL proteins

NEIL1,2, and 3 show overlapping domain architectures and structural similarities in addition to unique substrate specificity and function. All the current NEILs have a N-

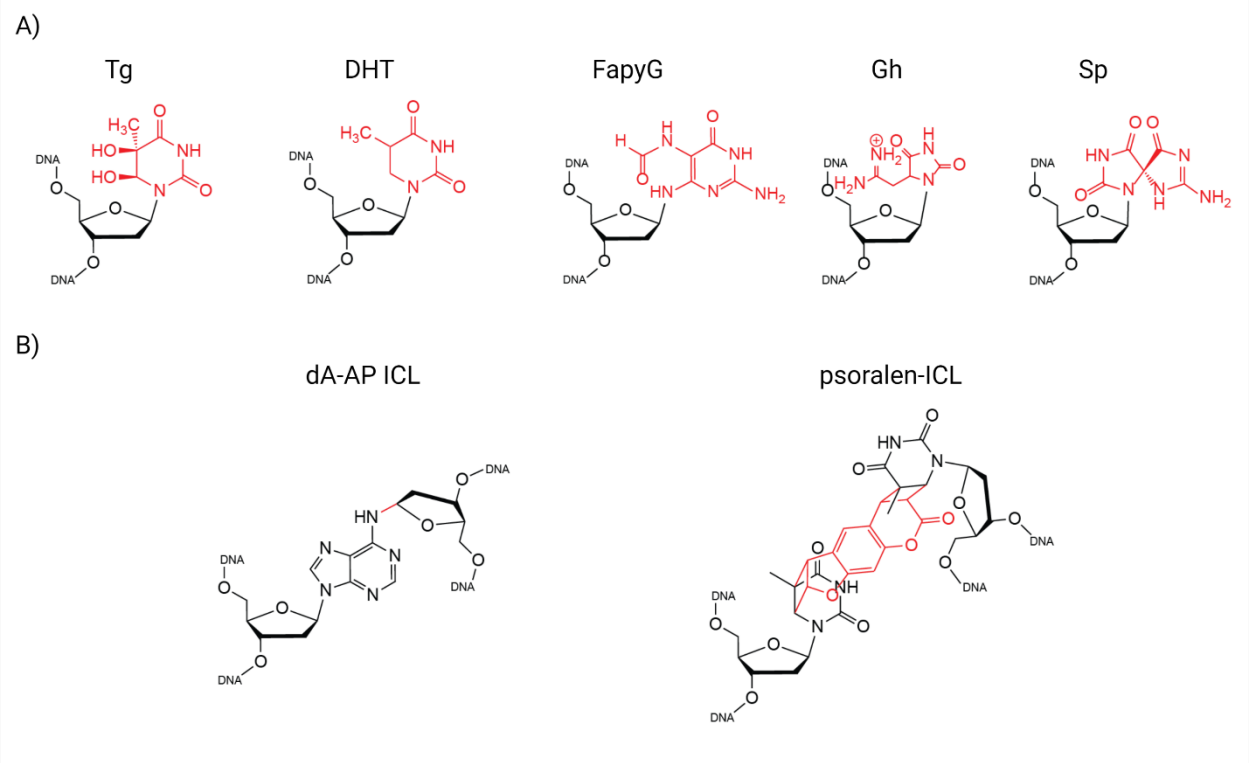

Figure 5. NEIL3 substrate specificity of oxidative lesions and DNA interstrand crosslinks. NEIL3 shows substrate specificity toward oxidative DNA nucleobases (red) including A) thymine glycol (Tg), dihydrothymine (DHT), formamidopyrimidine guanine (FapyG), 8-oxo guanine further oxidation products guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp). NEIL3 unhooks complex DNA-interstrand crosslinks including B) dA-AP ICLs and psoralen-ICLs.

terminal glycosylase domain (GD) that harbors the catalytic activity, a H2TH helix motif within the GD domain, followed by a Nei zinc finger (Nei-ZF). NEIL1,2,3 and homologous bacterial Nei enzymes contain the first five N-terminal residues MPEGP, except NEIL3 which contains a N-terminal valine at position 2, instead of a proline. NEIL1 is 390 amino acids, NEIL2 is 332 amino acids, and NEIL3 is the largest of the NEIL orthologs with 605 amino acids, due to an extended C-terminus containing multiple zinc fingers (Figure 6A). Zinc fingers are DNA binding motifs that can assist in protein folding by coordination to a

zinc metal ion. NEIL3 has an internal Npl4 zinc finger (NZF) that recruits NEIL3 to the replication fork via interaction with ubiquitinated replicative CMG helicase (Wu et al., 2019). NEIL3 also contains tandem GRxF-zinc fingers (GRF-ZFs) on its extreme C-terminus. The GRF-ZF domain binds to ssDNA and its functional role and structural features are addressed in Chapter 3.

X-ray crystallography is a powerful technique to determine the high-resolution structure of biomacromolecules. There are currently structures available of human NEIL1 lacking the C-terminal 95 amino acids bound to duplex DNA containing a Tg nucleobase modification, the open conformation of *Monodelphis domestica* (short-tailed opossum) NEIL2, and N-terminal GD of *Mus Musculus* (mouse) NEIL3 (Figure 6B) (Eckenroth et al., 2021; Liu et al., 2013b; Zhu et al., 2016). All three NEIL orthologs share a similar structure of N-terminal α-helix followed by a four antiparallel β-sheet sandwich. NEIL1 and NEIL3 were crystallized in a closed conformation with the active site between the N- and C-termini, adjacent to the interdomain hinge, while NEIL2 is crystallized in an open conformation showing flexibility and potential protein-binding interfaces. For NEIL1, the active site intercalating loop is within the β-sheet sandwich and inserts into the DNA to replace the flipped out Tg nucleobase. The intercalating residues for NEIL1 include methionine-81, arginine-118, and phenylalanine-120. Uniquely to other NEIL orthologs, NEIL2's open conformation consists of two inserted disordered loops that are not present in NEIL1 or NEIL3 (Eckenroth et al., 2021). The smaller of the inserted loops is predicted

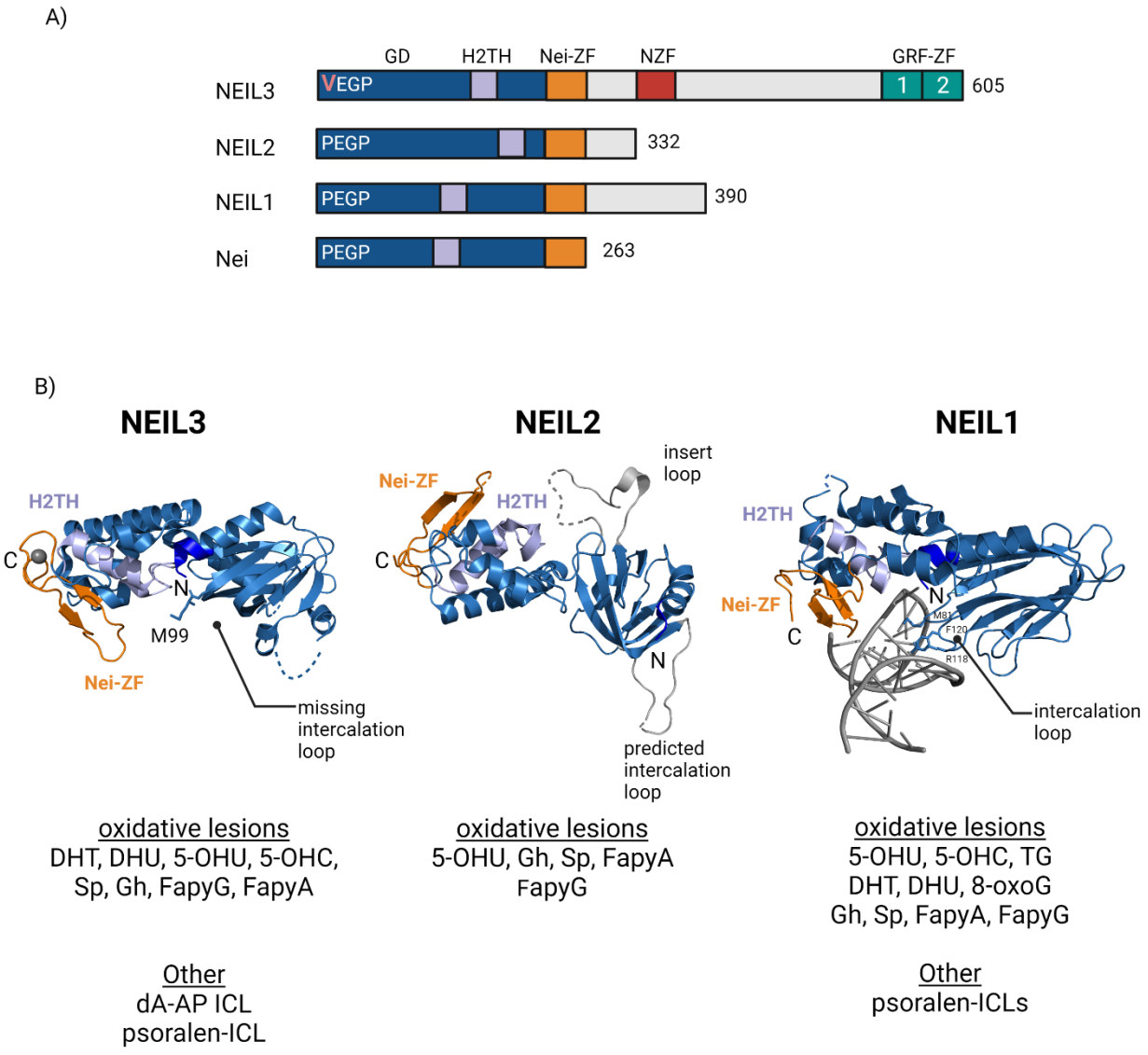

Figure 6. Domain architecture, structural features, and substrate specificity of NEIL orthologs and homolog. A) Schematic of domain map of endonuclease VIII-like (NEIL) orthologs NEIL1,2,3 and homolog endonuclease VIII (Nei) consisting of glycosylase domain (GD, dark blue) with N-terminal residues shown. Other domains include helix-2-turns-helix (HT2T, purple) motif, Nei-zinc finger (Nei-ZF, orange), Npl4-ZF (NZF, red), and two GRxF zinc-fingers (GRF, teal). B) Cartoon representation of mouse NEIL3 glycosylase domain (PDB: 3W0F), *Monodelphis domestica* NEIL2 (PDB: 6VJI), and human NEIL1 bound to duplex DNA with thymine glycol nucleobase modification (PDB: 5ITY). Intercalation residues are shown as side chains for NEIL1. Structural elements are highlighted with same color scheme shows in A.

to be the intercalation loop for NEIL2 with leucine-141 as one of the three intercalating residues, however two other intercalating residues are missing.

The larger loop is thought to interact with other DNA repair protein, many of which have not been characterized yet. NEIL3 is thought to perform a non-base flipping mechanism, unlike many other DNA glycosylases, because it is missing the intercalation loop found in NEIL1 and NEIL2. NEIL3 is also missing two of the three void filling residues (similar to NEIL2), but does possess one intercalating residue, methionine-99 (Figure 6) (Liu et al., 2013b). The NEIL3 positively charged active site accommodates ssDNA and computational models attempting to dock duplex DNA show defined clash sites that disfavor the second DNA strand (Liu et al., 2013b). Glycosylase activity assays have confirmed the ssDNA preference of NEIL3-GD over duplex DNA (Liu et al., 2010). The possibility of a non-base flipping mechanism may show how NEIL3 is able to excise a wide variety of DNA damage from small base lesions to ICLs.

All NEIL orthologs contain an interdomain hinge followed by H2HT motif α-helix and short Nei-ZF motif (Figure 6). NEIL2 shows an extended interdomain linker that may assist in the torsional twist and flexibility necessary for catalysis of transcription-coupled repair DNA intermediates or protein-protein interactions mediated by the additional NEIL2 inserted loop. NEIL1 contains a “zinc-less” zinc finger due to the absence of zinc coordination residues and zinc atom (Zhu et al., 2016). Structural details of the NEIL3-GD have provided evidence for its unique ssDNA specificity. The determined x-ray crystal structure of NEIL3-GD consists of two domains connected by a short linker (Liu et al., 2013b). The N-terminus harbors the active site which contains a N-terminal valine residue

(proline in NEIL1 and NEIL2 orthologs) after the initiator methionine is removed which acts as a nucleophile to attack the C1' on the deoxyribose sugar (Liu et al., 2013a).

This dissertation will include a summary of my NEIL3 published work investigating the biochemical studies of replication-dependent DNA repair by NEIL3. Chapter 2 will reveal the unhooking activity of DNA ICL fork structures by NEIL3. Chapter 3 will address the autoinhibitory role of the C-terminal GRF-ZF domain of NEIL3. Chapter 4 consists of a discussion of the implications of my work and future directions of the NEIL3 project. All references are listed at the end of the dissertation and publication references in which I am a co-author are marked at the beginning of each chapter.

Chapter 2

Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3[2]

Introduction

Like other mammalian glycosylases, NEIL3 is able to excise small DNA nucleobase modifications such as oxidative damage (Liu et al., 2013a). The NEIL3 contains an N-terminal glycosylase domain (GD) that contains all the residues necessary to excise oxidative DNA damage. The NEIL3-GD includes the H2TH and zinc finger motif and excludes 324 amino acids from the NEIL3 C-terminus (Liu et al., 2010). The NEIL3-GD was originally studied in preference to the NEIL3 full length (NEIL3-FL) protein due to its high stability throughout the purification process. A methionine aminopeptidase enzyme that cleaves the N-terminal initiator methionine was co-expressed with NEIL3 to optimize methionine processing and expression of active NEIL3-FL and NEIL3-GD (Liu et al., 2012). The NEIL3-GD has preference toward lesions within ssDNA or substrates containing ssDNA architecture, such as DNA bubbles that mimic transcription bubbles and fork-like structures that mimic DNA replication intermediates, over duplex DNA (Imani Nejad et al., 2020; Liu et al., 2010; Liu et al., 2013a; Liu et al., 2013b). The NEIL3-GD excises small oxidative base modifications in addition to further oxidation products of 8-

[2]This work is published in part in Imani Nejad, M., Housh, K., Rodriguez, A.A., Haldar, T., Kathe, S., Wallace, S.S., Eichman, B.F., and Gates K.S. (2020) Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3. *DNA Repair (Amst.)* 86, 102752. I generated figures 16 and 17, prepared and characterized NEIL3-GD enzymes, designed and conducted experiments, interpreted data, and contributed to writing of the publication.

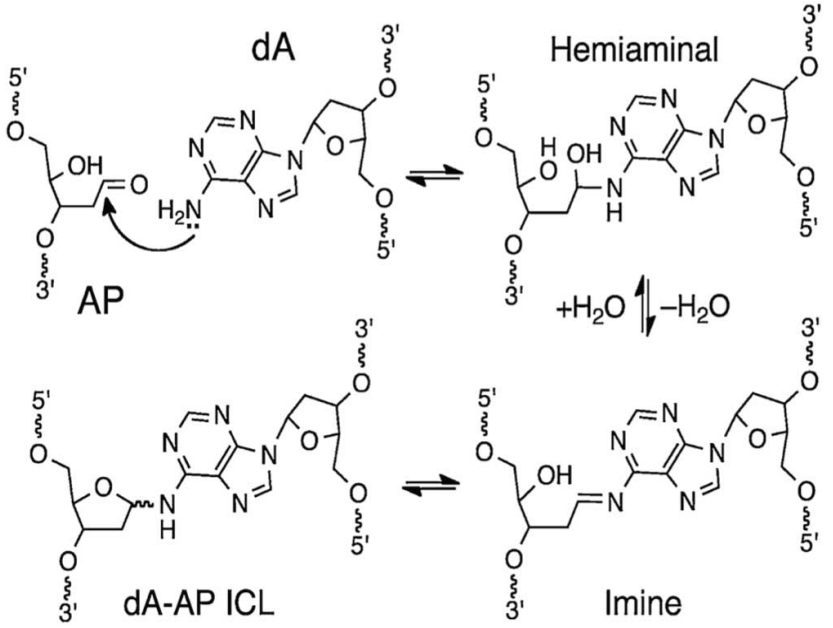

Figure 7. Formation of dA-AP ICL. The adenine (dA) residue on one DNA strand reacts with an open abasic (AP) site on the opposite DNA strand forming a hemiaminal and imine intermediate leading to formation of dA-AP ICL.

oxoG like Sp and Gh (Krokeide et al., 2013; Liu et al., 2013b). Additionally, NEIL3 excises DNA lesions such as Gh, FapyG and FapyA and has a low preference toward 8-oxoG (Liu et al., 2010).

Xenopus NEIL3-FL has the ability to unhook the interstrand covalent linkage between an AP site and adenine nucleotide (Semlow et al., 2016). Since abasic sites are predominant *in vivo*, the spontaneous formation of dA-AP ICLs from aldehydes has been hypothesized. The dA-AP ICL consists of an abasic site covalently bound to a DNA nucleobase. An AP site is in equilibrium between a cyclic hemiacetal and ring-open aldehyde that contains an aldehyde functional group and can react with adenine residues

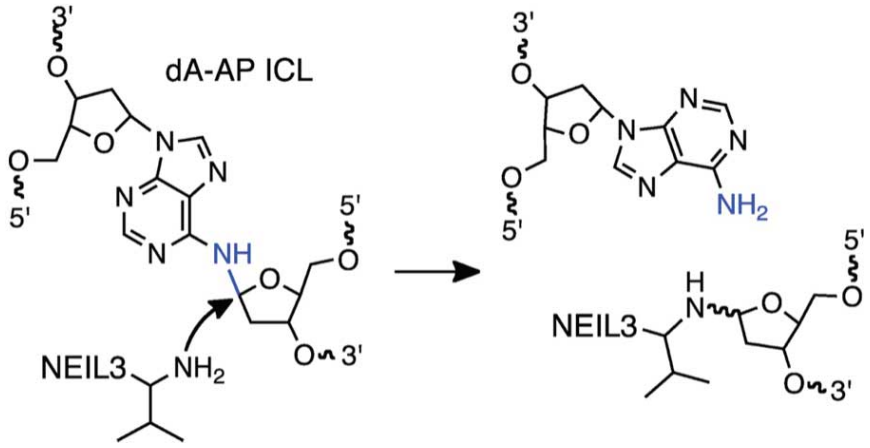

Figure 8. NEIL3 unhooks dA-AP ICL. NEIL3 cleaves the non-native glycosidic bond between N6 of adenine (blue) and C1 of AP site. Valine 2 of NEIL3 acts as nucleophile to attack C1 of AP site.

in the undamaged strand of duplex DNA (Figure 7). Two types of AP-ICLs have been synthesized with a covalent bond between an AP site and adenine or guanine nucleobase on the opposite strand (Johnson et al., 2013; Price et al., 2015; Price et al., 2014). The dA-AP ICL is stable for days and can be produced in yields of 15-70 % under physiologically relevant conditions (Price et al., 2015; Price et al., 2014). NEIL3 cleaves the non-native glycosidic bond between the N^6 of adenine and C^1 of the AP site resulting in an unhooked ICL (Figure 8).

NEIL3 has been described as a glycosylase with bifunctional lyase activity, but there is contradiction over the presence of lyase activity (Liu et al., 2010; Takao et al., 2009). The valine at position two in NEIL3 acts as a nucleophile to form the Schiff base intermediate (Liu et al., 2010). In favor of NEIL3 containing bifunctional activity, lyase assays have shown that NEIL3 cleaves DNA substrates with oxidative damage resulting in majority β-elimination with α,β-unsaturated aldehyde product. Interestingly, NEIL3-FL

shows two bifunctional glycosylase products such as β-elimination and β,δ-elimination products (Liu et al., 2010).

The NEIL3-GD published structural and biochemical data has begun to answer some early questions of NEIL3 substrate specificity and led to a hypothesis on the nature of optimal DNA substrate which contains single-stranded DNA character such as a lesion within single-stranded DNA, bubble structure transcription intermediates, and model replication forks. Understanding the detailed substrate preference of NEIL3 can help shed light on the ICL repair mechanism that underlies biological disease such as cancer predisposition and aging. NEIL3 has been shown to remove DNA damage from single stranded DNA substrates as well as bubbles and splayed arms that mimic DNA at a replication fork (Liu et al., 2010). However, the details of these substrates have not been investigated in regards to the location of the damaged nucleobase, length, and polarity of the replication fork. Here, we investigate these questions about NEIL3 to further uncover its unique activity and role in DNA repair and replication. Our results show that NEIL3 unhooks dA-AP ICLs and excises oxidized monoadducts on the leading template strand on model replication fork DNA substrates. Base pairing adjacent to the ICL inhibits activity and unhooking activity is unique to NEIL3 and not other BER enzymes including APE1, Fpg, Endo III, and NEIL1.

Results

The glycosylase domain of NEIL3 excises dA-AP ICLs in DNA fork structure

Mus musculus NEIL3-GD was purified as described in the methods section. Briefly, ICL unhooking assays were performed *in vitro* with dA-AP ICLs containing substrates and recombinant NEIL3-GD. The substrates included a duplex DNA and splayed arm/fork substrates that mimic DNA structures at a replication fork. The duplex substrate contained 38 base pairs and a dA-AP ICL at the center. The splayed arm/fork substrate consisted of a 37 base oligonucleotide containing a radiolabeled P^{32} on the 5′ end, 21 base pairs of duplex DNA followed by an abasic site with two base pairs of duplex DNA and 15 nucleotides of single stranded DNA. The reaction was incubated at 37°C for 2 hours in 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, and 100 µg/mL BSA. NEIL3 was shown to have no activity toward duplex substrate containing an dA-AP ICL (Figure 9).

NEIL3-GD unhooked and nicked the fork substrate containing the AP-ICL substrate to result in a 21 nucleotide long product (Figure 9, lane 3). A time course up to 27 hours showed dA-AP ICL incision and unhooking by NEIL3-GD on a forked substrate (Figure 10). After dA-AP ICL cleavage, strand separation occurred of 39 nucleotides, followed by lyase activity with cleavage 3′ (β elimination) and 5′ (β, δ elimination) to the AP site, resulting in 17 nucleotide and 21 nucleotide DNA fragments (Figure 11). Approximately 90% of these cleavage products were produced after 27 hours. Additionally, NEIL3-GD cleaves the dA-AP ICL leaving a single stranded DNA strand,

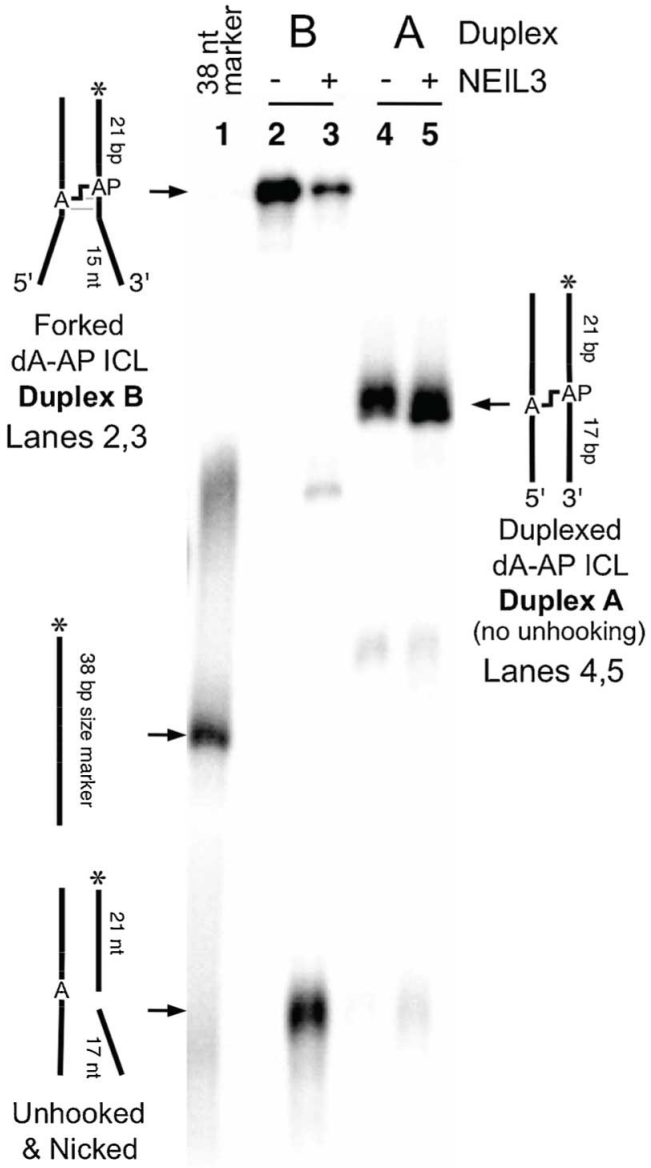

Figure 9. NEIL3 unhooks dA-AP ICL from fork DNA substrate and not duplex DNA substrate. DNA substrates were 5'-^{32}P-labeled and cross-linked duplexes were incubated with NEIL3 in 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM EDTA, 1 mM DTT and 100 µg/mL BSA at 37°C for 2 hours. Formamide loading buffer was added and samples loaded on 20% denaturing polyacrylamide gel. After electrophoresis ^{32}P-labeled oligonucleotides were visualized by phosphorimager analysis.

regardless of the location of the radiolabel (Figure 11). A radiolabel was added to the 5' (duplex B) or 3' end (duplex B') of DNA oligonucleotide fork substrate and incubated with

NEIL3-GD in 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 100 μg/mL BSA at 37°C for 2 hours.

Base pairing next to the AP-ICL lesion inhibits unhooking by NEIL3

To determine if the base pairs adjacent to the dA-AP ICL lesion affect the NEIL3-GD unhooking activity, we synthesized DNA fork substrates with various lengths of duplex DNA between the dA-AP ICL and single stranded region of the substrate. Fork substrates consisted of 21 nucleotides of duplex DNA on the 5′ of the AP site. The 3′ side of the AP site was designed to contain 2-8 base pairs of duplex DNA up to the ssDNA (Figure 12). ICL unhooking assays were completed with NEIL3-GD and leading fork substrates and cleavage products were quantified (see Materials and Methods). NEIL3-GD had the highest unhooking activity toward forks where the dA-AP ICL was flanked by the fewest base pairs of duplex DNA (2 base pairs) and had least activity when an 8 base pair duplex region separated the dA-AP ICL from the splayed arm. Also, inhibitory effects were found in fork substrates that contained six or more base pairs 3′ to the AP site and 5′ to the single stranded DNA junction.

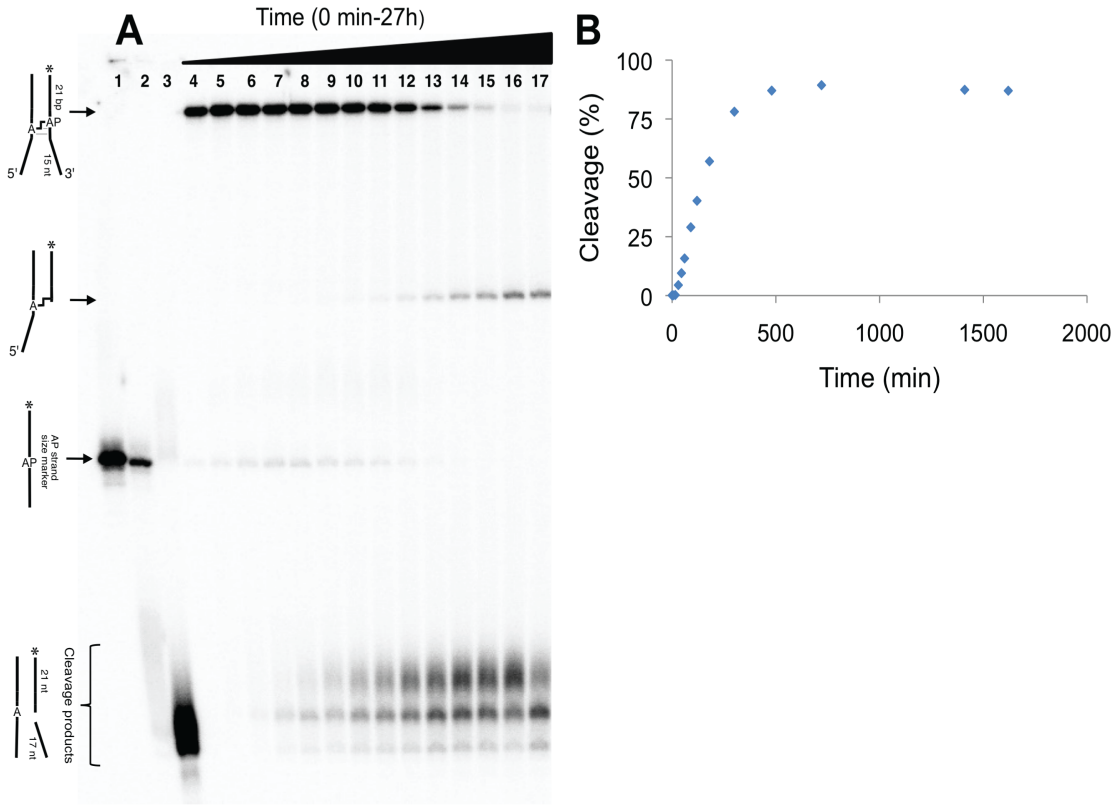

Figure 10. dA-AP ICL unhooking activity by NEIL3 over time. Purified fork cross-link B was incubated in NEIL3 buffer with 250 nM NEIL3-GD at 37°C. At timepoints 0-27 hours aliquots were removed from reaction sample and frozen prior to gel analysis. ^{32}P-labeled oligonucleotides were resolved on 20% denaturing gel and the radioactivity was quantified by phosphorimager analysis. A) Phosphorimage of gel. Lane 1: 39 nt ssDNA marker containing dU; Lane 2: Corresponding 39 nt ssDNA oligonucleotide containing AP-site; Lane 3: AP-oligonucleotide cleaved by 0.1 M piperidine for 30 minutes at 90°C to yield 21 nt fragment, Lanes 4-17: duplex B and NEIL3-GD incubated for 0, 5, 15, 30, 45, 60, 90, 120, 180, 300, 480, 720, 1410, 1620 minutes. Second band down on gel maybe an interstrand crosslink derived from the enzyme-bound 3'-alkenal elimination product (3'-dRP group) with nucleobase on opposing DNA strand. Characterization from this type of crosslink derived from β-elimination at an AP site in duplex DNA (Yang et al., 2017). The major NEIL3 unhooking dA-AP ICL cleavage product (lanes 7-17) migrates slower than the 3'-dRP and 3'-phosphate cleavage products generated by piperidine workup of AP oligonucleotide (lane 3). Control experiments showed that the unhooking cleavage product results from β-elimination to generate 3'-alkenal cleavage product, followed by conjugate addition of DTT in assay buffer. Conjugate addition of thiols to 3'-alkenal product of strand cleavage at AP site has been shown previously (Bailly and Verly, 1988). We observe a mixture of DTT-adduct, 3'-alkenal, and 3'-phosphate cleavage products that evolve with time. Under these single-turnover conditions the half-life of the unhooking reaction is 130 minutes.

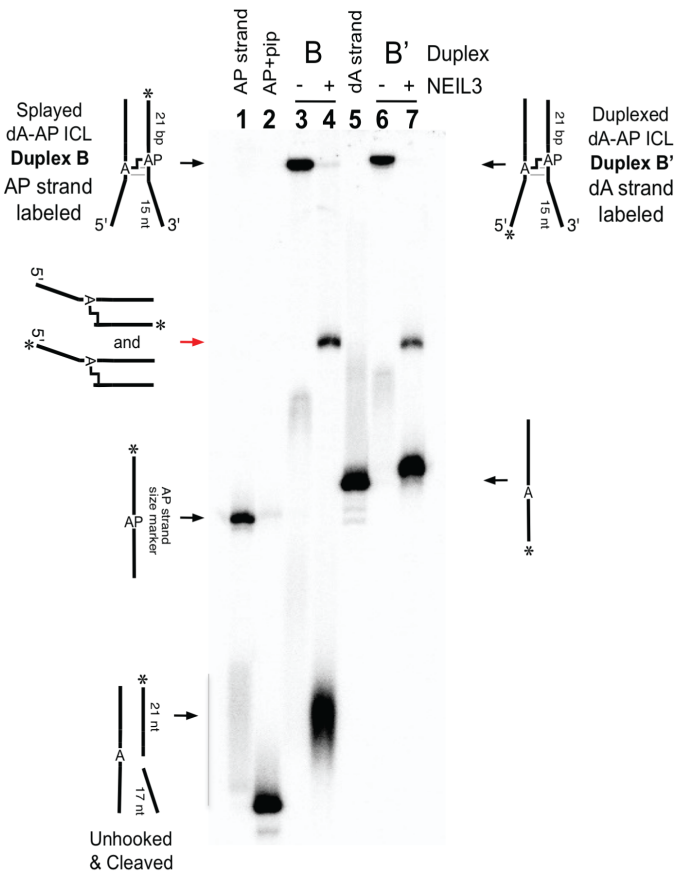

Figure 11. NEIL3-GD selectively unhooks dA-AP ICL in fork substrate to release a full-length dA-containing strand. Fork substrates were ^{32}P-labeled on 5' or 3' end and incubated with NEIL3-GD at 37°C for 2 hours. Cleavage products include NEIL3 unhooking of dA-AP ICL product (lane 4) and 3'-phosphate cleavage product from piperidine workup of the AP oligonucleotide (lane 2). The slight shift in dA strand in lane 5 compared to lane 7 is due to sample loaded in water. Red arrow indicates an intermediate that maybe an ICL derived from the reaction of the enzyme-bound 3'-alkenal eliminate product (3'-dRP group) with nucleobase on opposing strand.

NEIL3, but not other base excision enzymes including APE1, Fpg, Endo III, and NEIL1, unhook the dA-AP ICL in fork structures

To identify if dA-AP ICL unhooking was unique to NEIL3, ICL unhooking assays were completed with downstream BER enzymes including APE1, Fpg, endonuclease 3 (Endo III), and human NEIL1 with DNA substrate containing dA-AP ICL on leading template strand (Figure 13). APE1 is the second enzyme in the BER pathway that cleaves 5′ to the AP site for those glycosylases lacking lyase activity. Fpg and Endo III are the bacterial homologs of NEIL3 that remove oxidative damage. Human NEIL1 is an ortholog to NEIL3 with preference toward lesions within ssDNA and contains a zinc-less zinc finger (Doublie et al., 2004).

Base excision repair enzymes were purified as outlined in the methods section. Briefly, BER enzymes were incubated at 37°C for 24 hours with fork substrates containing an AP site on the leading template strand and adenine on the opposite strand but lacking a cross-link, as a control. There was no activity from the BER enzymes toward leading fork substrates containing dA-Ap ICL (Figure 13). The unhooking yields of BER enzymes was 3 ± 1% for APE1, 50 ± 2% for spermine, 8 ± 1% for Endo III, 15 ± 2% for Fpg, 4 ± 1% for NEIL1, and 95 ± 3% for NEIL3. All BER enzymes showed activity toward uncross-linked leading fork substrate (Figure 14). Spermine was added as a positive control to show a chemically cleaved ICL.

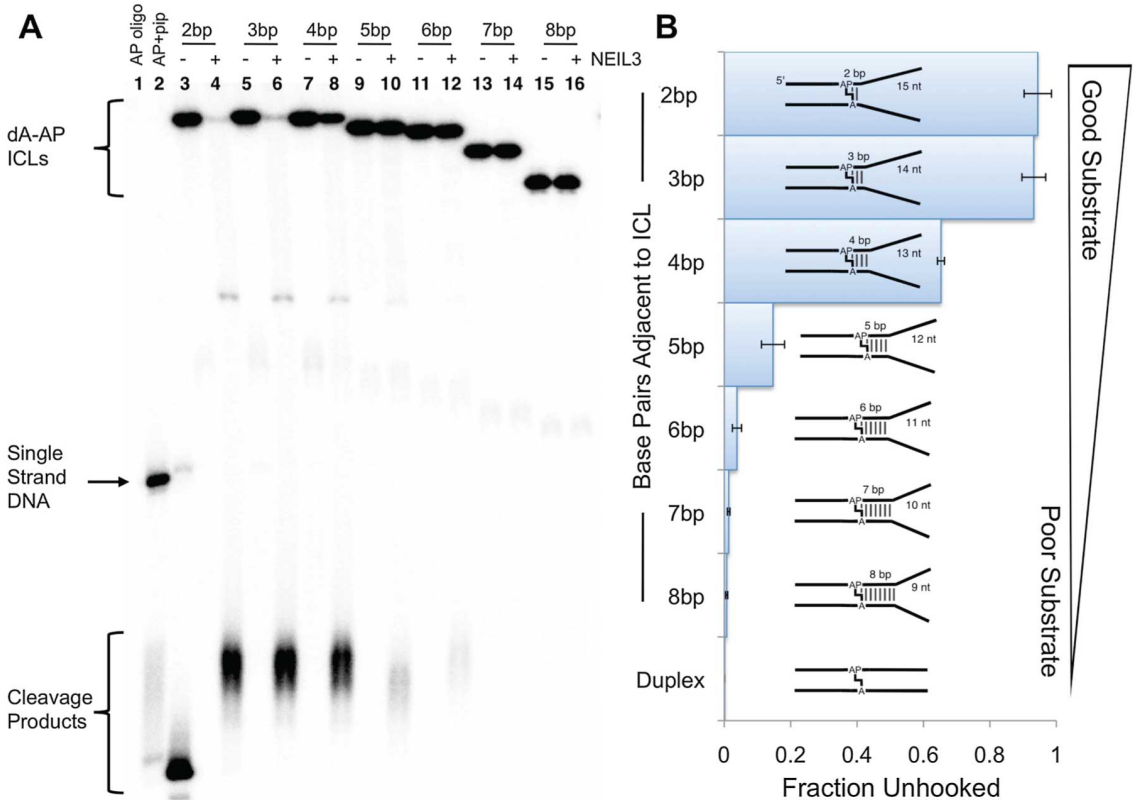

Figure 12. Base pairing adjacent to the dA-AP ICL inhibits NEIL3-GD unhooking activity. Fork substrates consist of 21 nt of paired duplex region on the 5'-side of the AP site and 17 nt, with various numbers of base pairs, on the 3'-side of the AP site. A) Purified ^{32}P-labeled dA-AP ICL forks substrates were incubated with NEIL3-GD at 37°C for 2 hours. B) NEIL3-GD substrate preference for fork substrates with base pair adjacent to ICL varied. Data for unhooking reactions is average ± S.D.

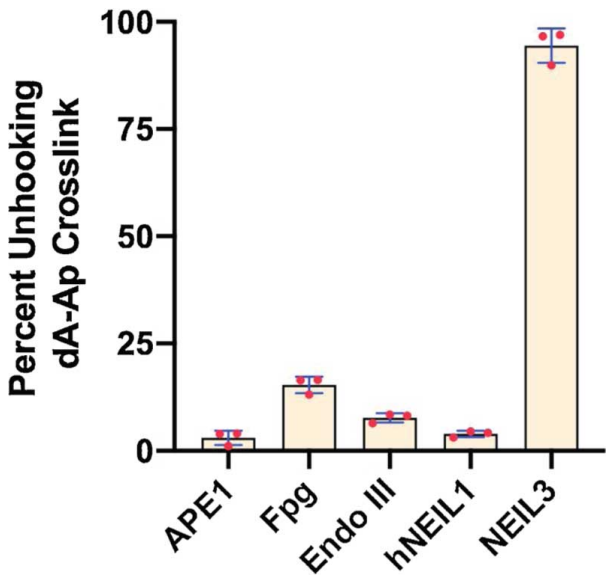

Figure 13. Base excision repair enzymes other than NEIL3-GD do not effectively unhook dA-AP ICL in fork substrates. The fork substrate was incubated with BER enzymes APE1, Fpg, Endo III, hNEIL1, and hNEIL3 at 37°C for 24 hours. Data represents average ± S.D. from electrophoretic analysis of unhooking reactions.

NEIL3 unhooks AP-ICL on leading template fork substrate

Next, we asked if NEIL3-GD had a preference toward an AP lesion on the template leading versus lagging strand which we label as leading or lagging fork respectively. This information would provide context as to where NEIL3 could be active within the replisome. Both substrates contained the AP site on the top strand of the fork and polarity was introduced with location of splayed single stranded DNA. Interestingly, NEIL3-GD only showed unhooking activity toward a leading fork substrate with AP site on the leading template strand and no activity toward the lagging fork substrate (Figure 15).

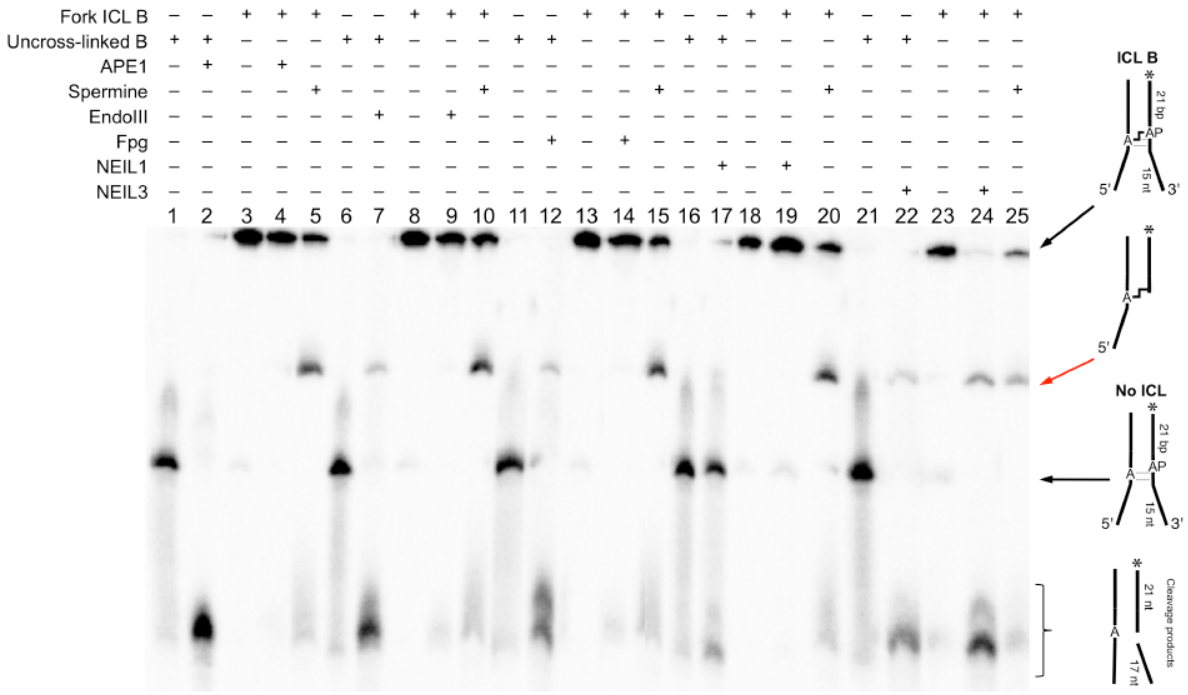

Figure 14. Base excision enzymes besides NEIL3-GD do not effectively unhook dA-AP ICL in fork substrates. Fork substrates with dA-AP ICL were incubated with BER enzymes at 37°C for 24 hours. Positive control included uncross-linked fork B and all BER enzymes showed endonuclease or lyase activity on this fork AP-site containing substrate. Red arrow labels ICL derived from 3'-dRP group and nucleobase on opposite strand.

To distinguish if there was a binding preference of leading over lagging fork DNA substrate we inquired on the ability of NEIL3 to bind to fork substrates containing an dA-AP ICL. Fluorescence anisotropy was utilized to assess binding affinity with NEIL3-GD catalytically dead mutant NEIL3-GD E3Q which retains DNA binding capabilities and dA-AP ICL substrates containing 6-carboxyfluorescein (FAM) fluorophore (Materials and Methods). NEIL3-GD E3Q has been utilized in other published binding assays to quantify binding and not catalytic function (Liu et al., 2013b). NEIL3-GD E3Q was added in excess

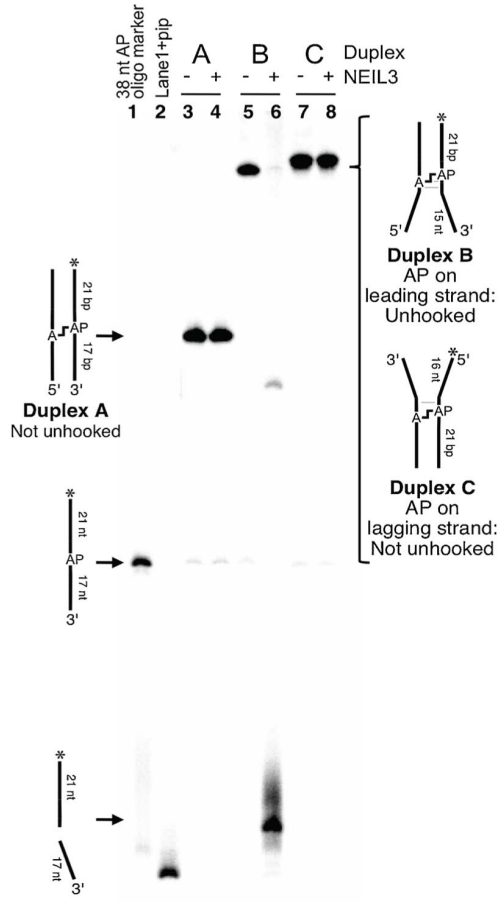

Figure 15. NEIL3-GD selectively unhooks the dA-AP ICL located at the duplex/single-strand junction of fork substrate. NEIL3-GD unhooks a dA-AP ICL containing where the AP site resides on the leading template strand of a model replication fork.

of DNA substrates. NEIL3-GD E3Q had equivalent affinity toward both lead and lag forks containing a dA-AP ICL lesion (Figure 16). The binding affinity (K_d) was 0.5 +/- 0.1 μM for leading fork substrate and 0.4 +/- 0.2 μM for lagging fork substrate. There was no binding to the FAM fluorophore as the total intensity remained consistent throughout the experiment. Therefore, the activity toward the lead AP-ICL fork is not due to DNA binding preference. To confirm the preference for lead over lag fork DNA substrate, NEIL3-GD base excision activity was monitored with DNA substrate containing a dA-AP ICL on the

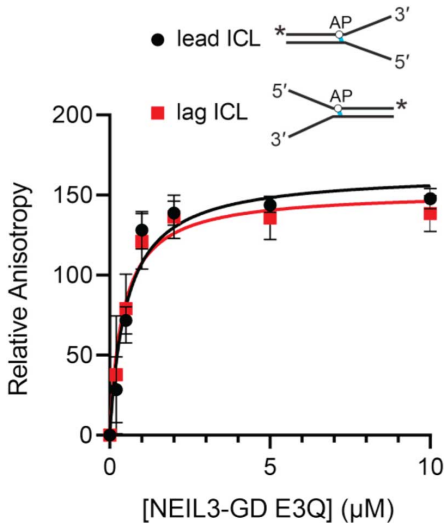

Figure 16. NEIL3-GD has same affinity for leading and lagging fork substrates containing dA-AP ICL. DNA binding curves with FAM-labeled dA-AP ICL substrate titrated with catalytically-inactive E3Q mutant of NEIL3-GD, monitored via fluorescence anisotropy. The plot shows average ± S.D. for three independent measurements. In the fork schematics, the blue line indicates the position of the ICL and the asterisk denotes the position of the FAM label.

leading or lagging template strand. The substrates included 25mer oligonucleotide containing DHT monoadduct on the top strand within single stranded, DHT on the leading template strand, DHT on the lagging template strand, and duplex DNA (Materials and Methods section). Additionally, a negative control of catalytically dead mutant, NEIL3-GD K82A, was also assayed for base excision activity with single stranded substrate containing DHT modification. Similar to ICL unhooking assays with NEIL3-GD, the base excision activity was highest in ssDNA and lead fork substrate followed by lag fork and little to no activity toward dsDNA. NEIL3-GD K82A did not show activity toward ssDNA with DHT (Figure 17). Therefore NEIL3-GD shows preference toward lead fork substrates containing dA-AP ICL and DHT monoadduct.

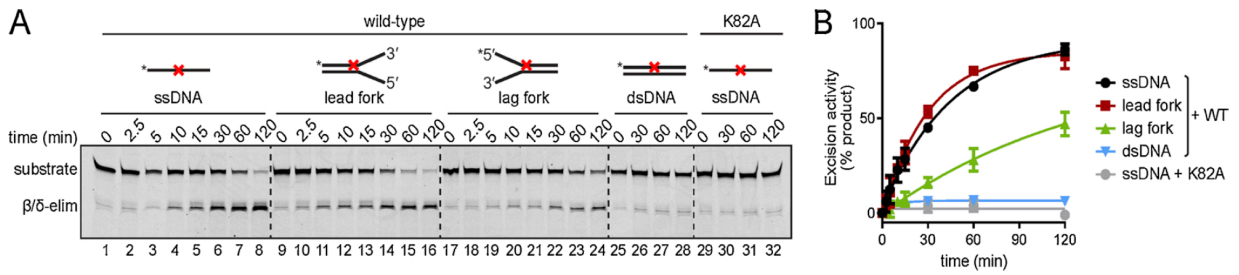

Figure 17. NEIL3-GD preferentially removes dihydrothymine (DHT) monoadduct from the leading template strand of a fork substrate. A) Representative polyacrylamide gel showing the time-course for NEIL3-GD activity against DHT (red X)-containing structures. Lanes 1-28, wild-type mNEIL3-GD; lanes 29-32, K82 inactive mutant control. B) Quantification of data from three independent experiments (average ± S.D.)

Discussion

The DNA glycosylase NEIL3 has shown the ability to excise small base modifications and more complex DNA interstrand crosslink lesions. Originally, DNA glycosylases were thought to excise DNA damage only by a “base flipping” mechanism in which the damaged nucleobase is inserted into the glycosylase active site (Mullins et al., 2019). Also, DNA glycosylases were thought to only excise small nucleobase modifications. However, a new branch of glycosylases has been discovered that do not function through the conserved “base flipping mechanism,” such as bacterial glycosylase homologs YTkR2 and AlkD that remove DNA bulky adducts such as (YTMA) and pyridyloxobuty adducts. AlkZ was shown to excise (AZB) DNA ICLs (Mullins et al., 2019; Mullins et al., 2017; Mullins et al., 2015). Within the vertebrate DNA glycosylase family, NEIL3 has been hypothesized to utilize a non-base flipping mechanism due to structural observations and its ability to excise complex DNA lesions such as psoralen and dA-AP ICLs (Imani Nejad et al., 2020; Semlow et al., 2016; Wu et al., 2019). Most surprisingly is

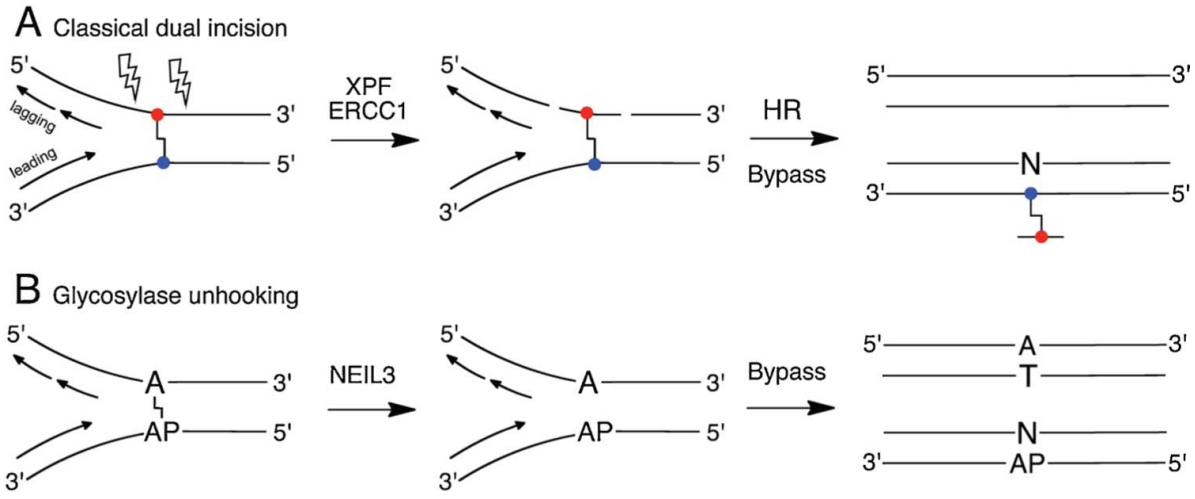

Figure 18. Mechanism of ICL repair. A) Classical ICL unhooking required dual incisions by NER proteins XPF and ERCC1 on either side of the ICL followed by HR and translesion synthesis (bypass). B) Unhooking mechanism of dA-Ap ICL by NEIL3 that avoids DNA double strand breaks.

that, due to the nature of DNA glycosylases resulting in a single DNA break, NEIL3 avoids the classical dual incision event common to the NER, FA, and HR pathways to unhook complex ICL lesions (Figure 18).

The majority of DNA glycosylases remove small base DNA modifications from duplex DNA. However, NEIL1 and NEIL3 have been shown to remove DNA lesions from single stranded DNA, forks, and bubble structures. Interestingly, NEIL3 is the only NEIL ortholog to show preference toward single stranded DNA containing a DNA lesion, over duplex DNA (Liu et al., 2013a). Yet, studies with NEIL3 and DNA substrates containing single and duplex architectures have been limited (Liu et al., 2010). Splayed arm substrates or fork-like structures are ideal due to their single and duplex characteristics that model a replication fork. NEIL3 has been identified at replication forks from iPOND analysis (Klattenhoff et al., 2017). Lesion site specificity has been investigated by NEIL3 with model replication fork substrates. Small DNA nucleobase lesions were tested within

duplex DNA, single stranded DNA or at the DNA junction of the model replication fork (Albelazi et al., 2019). NEIL3 showed greatest excision activity toward Tg and 5-OHU lesions within ssDNA or on the splayed arm consisting of single stranded DNA (Albelazi et al., 2019).

Even though published data is available that shows NEIL3 can remove oxidative DNA damage within DNA fork substrates (Albelazi et al., 2019), larger DNA damage complexes, such as ICLs, have not been investigated in model replication forks. NEIL3 was the first vertebrate glycosylase shown to unhook ICL lesions in *Xenopus* nuclear egg extracts (Semlow et al., 2016). The dA-AP ICL has a covalent bond between an AP site and adenine DNA nucleotide and has been synthesized in up to 70% yield for biochemical studies. NEIL3 was shown to unhook the dA-AP ICL by cleaving the non-native glycosidic bond between the deoxyribose sugar and N6-adenine.

I endeavored to identify if NEIL3 had a preference toward replication fork substrates with a small oxidative lesion or larger ICL on the leading or lagging template DNA strand. We confirmed that NEIL3 could excise lesions from fork DNA (Albelazi et al., 2019). We showed that recombinantly purified NEIL3-GD was able to excise a dA-AP ICL and DHT lesion from fork DNA substrates and not duplex DNA, with preference toward a leading fork substrate (Figure 18). We see that NEIL3 shows ICL unhooking and AP-lyase products which includes cleavage 5' and/or 3' of the AP site. Interestingly the bifunctional activity of NEIL3 was seen in DNA/RNA duplex substrates with dA-AP ICL within 3 nt of ssDNA and not within 24 nt of ssDNA (Semlow et al., 2016). It is possible that lyase activity of NEIL3 is present with ICL substrates within a longer ssDNA strand.

We showed that NEIL3-GD has preference toward DNA fork substrates with 2 base pairs next to the dA-AP ICL relative to the splayed arm junction and the least preference toward 8 base pairs adjacent to the ICL. NEIL3 has the highest preference toward lesions within ssDNA (Liu et al., 2013b), therefore it is not surprising that forks that contain the lesion closest to the splayed arm containing ssDNA would also be preferred. This may be due to NEIL3 binding to ssDNA as an anchor point first and then sliding along the substrate to find the dA-AP ICL. Likewise, the dA-AP ICL within the duplex DNA may hinder NEIL3 from identifying the lesion due to the inability of substrate contact within the active site. Structural information supports why NEIL3 cannot excise lesions from within duplex DNA. The active site region is positively charged to accommodate ssDNA, while the area positioned for the second DNA strand is abundant with negatively charged residues that would clash with duplex DNA (Liu et al., 2013b).

We showed that NEIL3-GD has unhooking activity toward a dA-AP ICL on the leading template strand but not on the lagging template strand. Interestingly, when we replaced a dA-AP ICL with DHT, NEIL3-GD was able to excise forks with DHT on the leading or lagging template strand, with preference toward the leading fork. Even though our DNA binding studies do not show a difference in NEIL3-GD binding to leading or lagging fork structures, FA is unable to identify where NEIL3 is binding to the substrate since these are bulk solution experiments. Therefore, further investigation is required to determine the location of DNA binding for example by site-directed mutagenesis of the predicted DNA binding location on NEIL3-GD. We identified that the ICL unhooking activity is unique to NEIL3 and is not catalyzed by several other base excision repair

enzymes including APE1, Fpg, Endo III, and human NEIL1. Therefore AP-lyase or AP-endonuclease activity was not sufficient to unhook the ICL.

Bacterial Fpg/Nei homologs show a similar structure to the NEIL3-GD. A NEIL3-GD crystal structure shows a two-domain architecture with active site at the N-terminus followed by α-helix, two layered β-sandwich, and four α-helices including the helix-2-turn helix motif and zinc finger motif (Liu et al., 2013b). The two domains are connected by an interdomain hinge. Unique to NEIL3 is the interdomain hinge loop (αB) which is perpendicular, compared to Fpg/Nei structures. The hinge is stabilized by hydrogen bonding and salt bridges, making NEIL3-GD more rigid than other Fpg/Nei homologs. The NEIL3-GD structure reveals why NEIL3 prefers ssDNA over dsDNA. Fpg/Nei glycosylases contain a positively charged region along the active site that binds one of the two DNA strands containing the damaged lesion. NEIL3 also contains a positively charged region, through the active site or the thinnest part of the protein, that could bind to a single strand of negatively charged DNA.

Similarly, the Fpg/Nei glycosylases contain a second positively charged patch with basic residues that contact the phosphate backbone of the duplex DNA strand. However, the DNA binding patch in NEIL3 is negatively charged and is shown to clash with modeled duplex DNA (Liu et al., 2013b). It is hypothesized that NEIL3 does not have duplex DNA preference because this glycosylase is missing the wedge residues that typically stabilize the second strand of DNA (Liu et al., 2013b). This NEIL3-GD structure provides evidence as to NEIL3 does not bind duplex DNA, however more structures need to be determined of NEIL3 bound to a DNA substrate.

Structural details of the DNA glycosylases can provide reason into their substrate specificity. The majority of DNA glycosylases are known to “flip” the damaged DNA nucleobase into the active site and contain a catalytic triad of void filling residues that intercalate into the DNA to substitute for the flipped out DNA nucleobase. These intercalating residues are found in the intercalation loop of Fpg/Nei glycosylases. Methionine 99 is one of the intercalating residues found in the intercalation loop of Fpg/Nei glycosylases and NEIL3. The two additional intercalating residues in the triad are typically found in a neighboring beta-loop in Fpg/Nei glycosylases. However, NEIL3 has a truncated beta-loop missing two of the three void filling residues, phenylalanine and arginine.

On the front face of NEIL3, there is a truncated 8-oxoG capping αF- β9 loop between the H2TH and zinc finger motif at the C-terminus (Liu et al., 2013b). The 8-oxoG capping loop of Fpg/Nei glycosylases intercalates into DNA with an 8-oxoG lesion. But the 8-oxoG capping loop in NEIL3 is shortened so that is does not contact DNA and provides structural evidence as to why NEIL3 does not remove the 8-oxoG DNA lesion. It is unknown how this loop if at all interacts with the dA-AP ICL which has shown to be in the ring closed β-configuration via NMR (Kellum et al., 2021). The structural snapshot of NEIL3 bound to a DNA substrate remains unanswered and a future structural model will open many more questions on how this DNA glycosylase can unhook ICLs and a broad range of DNA oxidative lesions.

This work characterizes the N-terminal NEIL3-GD yet there is still much to be learned about the C-terminus that contains multiple zinc fingers. The internal NZF zinc finger within NEIL3-FL in *Xenopus* egg extracts was shown to bind ubiquitinated CMG

helicase and aid in the recruitment of NEIL3 to the replication fork (Wu et al., 2019). The extreme C-terminus of NEIL3 contains two tandem GRF-ZF motifs that bind ssDNA and may orient NEIL3 at the replication fork (Wu et al., 2019). Even though we show that the NEIL3-GD acts independently as an unhooking enzyme, additional investigation of the C-terminus of NEIL3 is required to understand the function of NEIL3-FL at a replication fork.

Chapter 3

An autoinhibitory role for the GRF zinc finger domain of DNA glycosylase NEIL3[3]

Introduction

Zinc fingers (ZF) are abundant in many proteins and have a multitude of functions like binding DNA. Zinc fingers were first characterized in transcription factor IIIA as 30-amino acid motifs that contained fingers that would “grip” DNA (Klug and Schwabe, 1995). Zinc fingers contain a zinc ion that is coordinated by four cysteine or histidine residues in a tetrahedral configuration. Zinc fingers can be further classified based on their conserved amino acid sequence. The GRxF zinc finger (GRF-ZF) is named after the three conserved residues found in the domain: glycine, arginine, and phenylalanine. These GRF-ZF domains have been identified in several human proteins involved in DNA damage response and transcriptional activity, such as transcription termination factor 2 (TTF2), exoribonuclease family member 2 (ERI2), zinc finger GRF-type containing 1 (ZGRF1), (TOP3A), and NEIL3 (Wallace et al., 2017).

NEIL3 is a DNA repair enzyme that contains glycosylase activity catalyzed by the glycosylase domain at the N-terminus. The glycosylase domain excises oxidized DNA lesions and unhooks interstrand crosslinks containing psoralen or an AP site (Imani Nejad et al., 2020; Semlow et al., 2016). The C-terminus of NEIL3 contains multiple unique zinc

[3] This work is published in Rodriguez, A.A., Wojtaszek, J.L., Greer, B.H., Haldar, T., Gates, K.S and Eichman, B.F. (2020) An autoinhibitory role for the GRF zinc finger domain of DNA glycosylase NEIL3. *J. Biol. Chem.* 295, 15566-15575.

fingers compared to other NEIL orthologs (Liu et al., 2013a). NEIL3 contains an internal NPL4/RAN binding protein (Ranbp) zinc finger and two tandem GRF-ZF motifs at its C-terminus.

The Npl4/Ranbp zinc finger and GRF-ZF domain are necessary for NEIL3 recruitment to the replication fork in *Xenopus* nuclear extracts (Wu et al., 2019). Truncated NEIL3 lacking the C-terminal zinc fingers was unable to unhook and bind circular plasmid DNA containing a dA-AP ICL lesion (Wu et al., 2019). The dA-AP ICL contains an AP site on one DNA stand covalently bound to an adenine residue on the opposite strand. These lesions have been postulated to exist *in vivo* due to the high abundance of endogenous AP sites and aldehydes present within the cell (Housh et al., 2021; Price et al., 2014).

Other enzymes that contain a GRF-ZF domain and GRF-zinc ribbon motifs include apurinic/apyrimidinic-endonuclease 2 (APE2) and topoisomerase 1 (TOP1). APE2 consists of an N-terminal exonuclease/endonuclease/phosphodiesterase (EEP) domain that harbors weak exonuclease and phosphodiesterase DNA resection activity in the 3' to 5' direction that is stimulated by proliferating cell nuclear antigen (PCNA) (Burkovics et al., 2006; Tell et al., 2009; Wilson and Barsky, 2001). APE2 has an internal PCNA-interacting motif (PIP) and C-terminal GRF-ZF domain. Proteolysis by chymotrypsin digestion identified a stable EEP domain and GRF-ZF domain (Wallace et al., 2017). The structure of the GRF-ZF core of APE2 was determined by x-ray crystallography and comprised a N-terminal helix followed by polyproline helix hinge connected to the GRF-ZF. The GRF-ZF domain consists of three anti-parallel beta sheets resembling a claw-like structure. The beta sheets consist of a loop on their ends that contain a zinc ion

coordinated by three cysteines and one histidine (Wallace et al., 2017). The concave surface of the GRF-ZF contains positively charged residues. DNA binding assays via electrophoretic mobility shift assays (EMSAs) and fluorescence polarization showed that APE2 had highest affinity toward DNA substrates with single stranded regions (Wallace et al., 2017).

Additionally, NMR studies with the APE2 GRF-ZF domain showed significant chemical perturbations along the front cleft (R473, K476, K477) when ssDNA was introduced into solution with APE2 GRF-ZF. When the residues within the positively-charged cleft were mutated to negatively-charged residues, DNA binding to ssDNA decreased, suggesting that ssDNA does bind along the cleft. To identify if the GRF-ZF enhances the catalytic activity of APE, stimulated PCNA nuclease activity assays were completed with APE2 GRF-ZF mutants and DNA substrate consisting of a 3' recessed overhang end. PCNA is required for 3'-5' nuclease activity by APE2. APE2 GRF-ZF mutant R502E showed a significant decrease in the generation of cleavage products showing that loss of DNA binding in the GRF-ZF decreases the exonuclease activity of APE2 (Wallace et al., 2017). The GRF-ZF of APE2 was necessary in the recruitment of the replication-dependent chromatin-binding proteins such as replication protein RPA32, ATR, ATR-interacting protein (ATRIP), and Rad9, which are involved with the oxidative stress-induced ATR-Chk1 checkpoint. None of these proteins were recruited when the GRF-ZF of APE2 was truncated in a *Xenopus* cell free system (Wallace et al., 2017). Therefore, the GRF-ZF domain of APE2 might assist in recruitment of the ATR-CHk1 proteins.

Bacterial Top1 is a DNA-binding enzyme of the type 1A topoisomerase family that unknots hyper negative supercoiled DNA by incision of one DNA strand (Tan et al., 2015). Top1 can remove deleterious RNA-loops that block DNA transcription and replication. The N-terminus of Top1 contains cleavage function to produce a ssDNA break. The C-terminus contains five DNA-binding motifs or ribbons. The zinc ribbons are DNA-binding motifs that coordinate one zinc ion by four cysteines. *Escherichia coli* Top1 contains five zinc ribbons located on the C-terminus of the protein and these zinc ribbons are found in homologous eukaryotic type 1A topoisomerase, TOP3α and TOP3β. Homologous Top1 enzymes contain at least 3 zinc ribbons on the C-terminus.

The structure of full-length *E.coli* Top1 was determined with the C-terminus consisting of five zinc-ribbons bound to ssDNA of 11 nucleotides in length (Tan et al., 2015). In the crystal structure, 4 out of the 5 C-terminal zinc ribbons are bound to ssDNA (3 nucleotides) and 3 out of the 5 zinc ribbons contains a coordinated zinc ion. Each zinc ribbon consists of 3 anti-parallel beta strands with an aromatic residue that contributes to base-stacking interaction with DNA. Top1 relaxation activity decreased when the base-stacking aromatic residue in the zinc ribbon was mutated to a non-aromatic residue (Tan et al., 2015). Therefore, the zinc ribbons are necessary for complete relaxation activity of Top1.

Currently there is little understanding of the biochemical and structural modality of the GRF-ZF of NEIL3. Even though published data has shown that the GRF-ZFs of NEIL3 are necessary for NEIL3 recruitment to the replication fork (Wu et al., 2019), it is unknown how the GRF-ZFs regulate NEIL3 activity when at the site of damage at a replication fork. To do this would require an *in vitro* system capable of monitoring NEIL3 base excision

and ICL unhooking activity after recruitment to a converged replication fork. Additionally, there is no structural information of the GRF-ZF domain of NEIL3. Even though other GRF-ZF domain structures are available in the protein data base (PDB), it is unclear if the NEIL3 GRF-ZFs would have a similar structure, especially since NEIL3 has unique substrate specificity toward AP-ICLs (Semlow and Walter, 2021). Structural information would provide a snapshot of how the GRF-ZF domain is interacting with ssDNA, where the DNA-binding region is located and the specific DNA-binding residues. Structural information and biochemical studies would elucidate the mechanism of GRF-ZF modulation and provide insight on how NEIL3 can remove DNA damage during replication. In the following section I will provide structure information into the DNA binding nature of the NEIL3 GRF-ZF domain. Additionally, I will show a novel inhibitory function of the GRF-ZF domain toward base excision and ICL unhooking activity of NEIL3.

Results

GRF zinc fingers bind ssDNA and fork-like structures

NEIL3 contains a C-terminal extension consisting of several zinc fingers. The extreme C-terminus contains two tandem GRF-ZFs and it was unknown why there were two tandem GRF-ZF motifs versus one as seen in other proteins (Wallace et al., 2017). It has also been shown that the individual GRF-ZFs bind ssDNA with higher affinity than duplex DNA, yet the tandem GRF-ZF motifs have not been investigated (Wu et al., 2019). Additionally, very little biochemical and structural information is known about the NEIL3 GRF-ZF domain in tandem.

The *Mus musculus* NEIL3 GRF1, GRF2, and tandem GRF1 and GRF2 (herein referred to as GRF12) were purified (Figure 19) with N-terminal glutathione (GST) tag and DNA binding was quantified by EMSAs. DNA substrates that were tested included ssDNA (40 nucleotides), duplex DNA (40 base pairs), and splayed arms (forks) consisting of 20 base pairs of duplex DNA and 20 nucleotides of ssDNA. All of the GRF-ZF constructs (GRF1, GRF2, GRF12) bound to ssDNA which is consistent with other publications (Figure 19) (Ha et al., 2020; Wu et al., 2019).

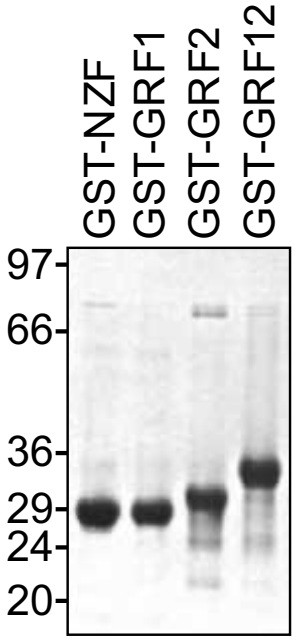

Figure 19. Coomassie-stained SDS-PAGE gel of purified GST proteins. Proteins were used in Figure 20 and Figure 22 G,H.

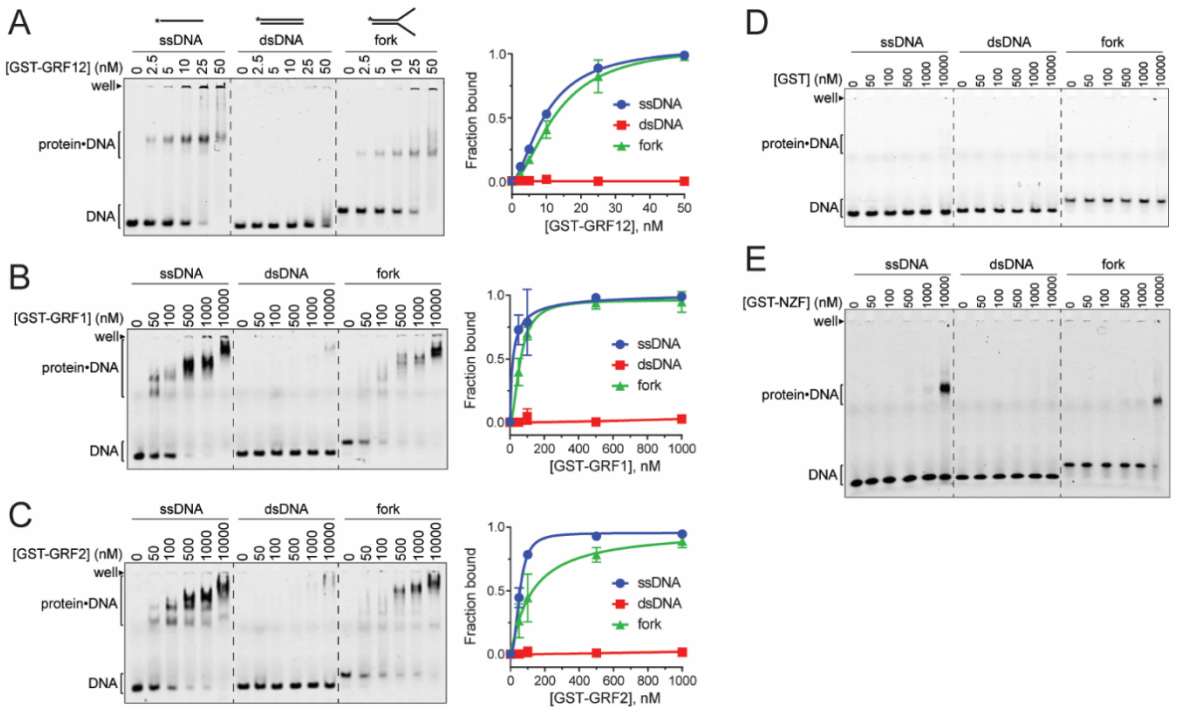

Figure 20. NEIL3 GRF motifs bind ssDNA. A-E) EMSAs for mNEIL3 ZF motifs binding to ssDNA, dsDNA, and fork substrates. Gels A-C are quantified in the plots on right. The data are average ± S.D. for n=3 replicates. A, GST-GRF12. B, GST-GRF1. C, GST-GRF2. D, GST control. E, GST-NZF. Data acquired by Briana Greer.

None of the GRF constructs bound to duplex DNA. All GRF-ZF constructs bound fork substrates similar to ssDNA. The tandem GRF12 shows 5-10 fold tighter binding (apparent K_d = 10 nM) to ssDNA and forks compared to individual GRF-ZF motifs (apparent K_d = 50-100 nM). Interestingly, GRF12 showed one protein-DNA complex band, whereas the individual GRF1 or GRF2 motif had multiple protein-DNA complex bands due to non-specific binding from GRF concentrations of 0.5-10 μM (Figure 20A). Additionally, GRF2 showed a high DNA binding affinity to ssDNA over the fork substrate. The GST tag was tested for DNA binding as a negative control and no ssDNA binding was observed (Figure 20D). Additionally, the internal NZF was purified and DNA binding

was also tested via EMSAs but did not show appreciable affinity to DNA (Figure 20E). In summary, the DNA binding results with NEIL3 GRF-ZFs show affinity to ssDNA and fork DNA of 40 nucleotides in length. Additionally, two tandem GRF-ZF motifs increase the DNA binding affinity toward ssDNA versus the individual GRF-ZF motifs.

Structural basis for ssDNA binding by the NEIL3 GRF domain

To understand the GRF-ZF that binds ssDNA, we utilized x-ray crystallography to determine the first 2.6 Å crystal structure of the human NEIL3 GRF12 domain using single-wavelength anomalous dispersion at the zinc edge (Table 1, Figure 21,22). The asymmetric unit contains two GRF12 protomers with R_{work} of 22.7% and R_{free} of 26.6%. Other crystallographic model statistics are listed in Table 1. Stereoimages show the refined model superimposed on the electron density (Figure 20). The final butterfly-like model contains two GRF motifs linked in a head-to-tail fashion. In between the tandem GRF-ZFs there are three amino acids (aspartate, leucine, serine) that connect GRF1 to GRF2. Each GRF-ZF motif coordinates one zinc atom by three cysteines and one histidine residue within the metal-binding loops on opposite corners (Figure 22A).

Table 1. Data collection and refinement statistics. The data were generated from a single crystal. Data collected by Jessica Wojtaszek.

	Zn-SAD	Native
Data collection		
Space group	I4	I4
Cell dimensions		
a, *b*, *c* (Å)	93.608, 93.608, 63.718	93.494, 93.494, 63.646
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Wavelength	1.27059	1.0000
Resolution (Å)[a]	2.80 (2.85–2.80)	2.60 (2.64–2.60)
R_{sym}[a]	0.056 (0.727)	0.038 (0.656)
R_{meas}[a]	0.060 (0.799)	0.043 (0.742)
R_{pim}[a]	0.022 (0.326)	0.020 (0.344)
$CC_{1/2}$	0.810	0.840
$I/\sigma I$[a]	30.8 (2.2)	31.4 (2.0)
Completeness (%)[a]	99.9 (99.7)	99.8 (100.0)
Redundancy[a]	7.4 (5.9)	4.5 (4.6)
Refinement		
Resolution (Å)[a]		34.95–2.60 (2.70–2.60)
No. reflections[a]		7550 (469)
R_{work}/R_{free}[a]		0.227/0.266 (0.271/0.306)
No. atoms		
Protein		1332
Solvent		0
B-factors		
Protein		51.9
Solvent		
RMSDs		
Bond lengths (Å)		0.003
Bond angles (°)		0.675
Ramachandran plot (%)		
Favored		89.9
Allowed		10.1
Outliers		0

[a]Numbers in parentheses refer to data in the highest-resolution shell.

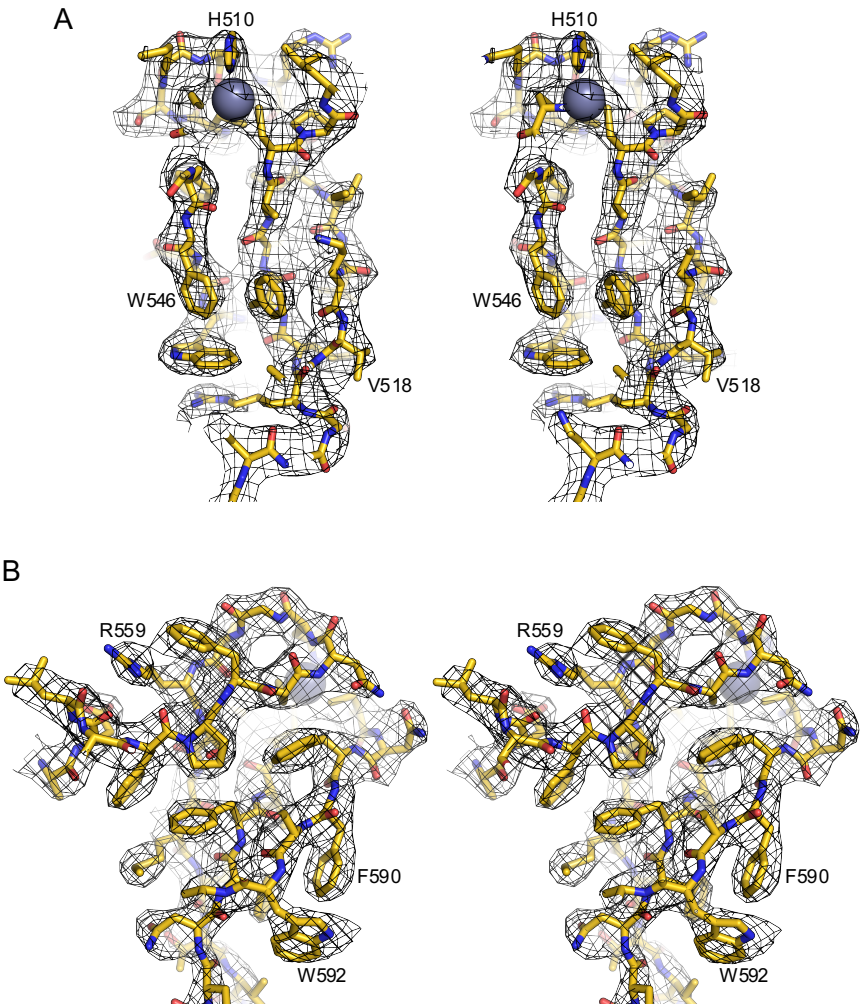

Figure 21. Stereoimages of representative regions of the refined model superimposed onto 2Fobs-Fcalc electron density (contoured to 1σ). Labeled residues are for orientation purposes only. A) GRF1. B) GRF2.

The GRF-ZF domain contains two virtually identical GRF-ZF motifs (Figure 22D). Each GRF motif forms a crescent or claw-like structure consisting of three anti-parallel beta strands with one end encompassing the zinc coordination site and the other end made up of a 7-10 residue loop. The concave, front face and outward surface of the GRF-ZF domain contains basic residues, many of which have been identified as DNA binding

residues in *Xenopus* NEIL3 and APE2 (Wallace et al., 2017; Wu et al., 2019). This positively charged concave face creates a continuous surface area across the GRF-ZF motifs, as demonstrated in the electrostatic potential map, and likely binds ssDNA (Figure 22B). The convex or back face of the GRF-ZF domain is neutral with no overall charge as observed in the concave face. Interestingly, the two protomers are in different conformational states (Figure 22C).

As demonstrated by the overlap of the two protomers, one protomer (yellow) is stretched open with a bend at the linker between the two GRF-ZF motifs, leading to a flexible and open conformation. The two different protomers show the inherent flexibility within the GRF-ZF domain to accommodate various ssDNA conformations. The two NEIL3 GRF-ZF motifs are very similar, almost identical, and show high similarity to the GRF-ZF in xAPE2 with a RMSD of 0.86 Å (GRF1) and 0.60 Å (GRF2) for all atoms. DNA binding residues previously identified in xAPE2 as Arg^{473}, Phe^{486}, and Lys^{477} correspond to Arg^{517}, Phe^{530}, and Lys^{521} in NEIL3 (Figure 22E) (Wallace et al., 2017). These DNA binding residues are located in the positive concave front face of NEIL3 GRF-ZF domain and overlap well with the DNA binding residues in xAPE2 as shown in the sequence-based sequence alignment (Figure 22E,F). To confirm these were DNA binding residues for NEIL3, charge reversal mutants were created by changing positively charged residues to negatively charged glutamate.

Solution-based DNA binding assays were performed via fluorescence anisotropy (Materials and Methods) to quantitatively measure the binding of GRF-ZF mutants to 40 nucleotide long ssDNA with 5' FAM fluorophore. Mutants were tested in individual GRF-ZF motifs with N-terminal GST tag and tandem GRF-ZF domain without GST tag. Wild-

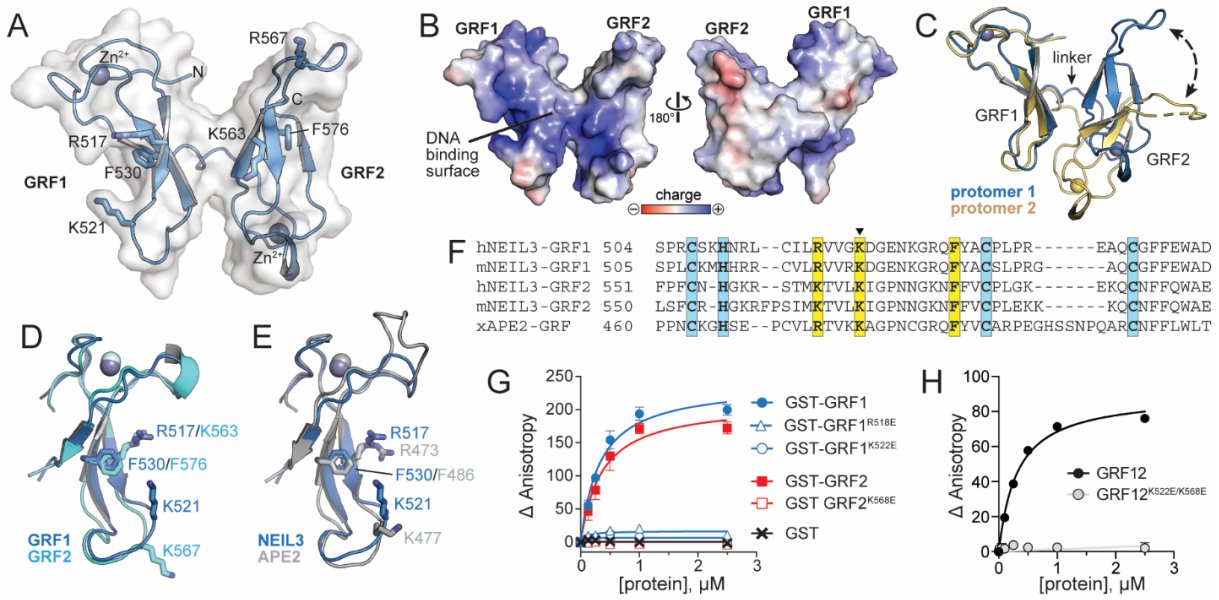

Figure 22. Crystal structure of hNEIL3 GRF domain. A) Structure of one of the two protomers in the asymmetric unit. The molecular surface is shown in white and the positions of several conserved DNA binding residues are indicated. B) the two faces of the GRF12 structure, colored by electrostatic potential. The image on the left is in the same orientation as in A. C) the two GRF12 molecules in the asymmetric unit, superimposed by GRF1. The difference in relative position of GRF2 is highlighted by a dashed arrow. D) superposition of GRF1 and GRF2. E) superposition of GRF motifs from hNEIL3 (blue) and xAPE2 (silver, PDB ID 5U6Z). F) structure- based sequence alignment of hNEIL3 and xAPE2 GRF structures together with aligned sequences from mNEIL3. Zn^{2+}-coordinating and DNA-binding residues are highlighted blue and yellow, respectively. G,H) DNA binding of individual mNEIL3 GRF motifs as GST fusion proteins (G) and the tandem, untagged GRF12 domain (H). The data are average ± S.D. (n=3).

type GST-GRF1 and GST-GRF2 bound to ssDNA with K_d of 0.3 ± 0.04 and 0.4 ± 0.06 μM respectively (Figure 22G). The negative control of GST tag showed no DNA binding. We showed that all glutamate mutants in individual GRF-ZF motifs including *Mus musculus* NEIL3 Arg^{518} or Lys^{522} which correlates to Arg^{517} and Lys^{521} in human NEIL3 respectively, showed ablation of DNA binding (Figure 22G). Published data showed that the K522E mutant in mNEIL3 GRF-ZF1 also shows ablation in corresponding mutant

K500E in xNEIL3 MBP-GRF1 (Wu et al., 2019). The lysine to glutamate double mutant (K522E/K568E) was introduced into the GRF-ZF domain without GST tag and we also observed ablation of DNA binding (Figure 22H). Wild-type GRF12 bound to ssDNA with K_d value of 0.2 ± 0.01 μM and double mutant GRF12-K522E/K568E showed 100-fold reduced binding affinity. Therefore, this positively charged concave surface in which I have observed a decrease in affinity for three point mutations is the proposed DNA binding surface of the GRF-ZF domain.

Structural homologs to NEIL3 GRF-ZF domain were identified by a Dali search. The two structural homologs were *E. coli* Top1 which contains several zinc ribbons and *Xenopus* APE2 that contains one C-terminal GRF-ZF motif (Tan et al., 2015; Wallace et al., 2017). The proteins are similar in that NEIL3, Top1, and APE all contain GRF-ZF motif(s) or GRF-like ZF motifs on their extreme C-terminus (Figure 23A). Top1 is unique in that is contains 5 zinc ribbon (ZR) motifs with 4 of these motifs bound to ssDNA in a deposited crystal structure (Figure 23B) (Tan et al., 2015). Top1 contains repeating ZR motifs that are connected by flexible linkers and bind ssDNA along a continuous face. These 5 ZR of Top1 provide an example of how the NEIL3 GRF-ZF motifs could bind ssDNA. Since the Top1-ZRs are similar to each other, we wanted to compare the NEIL3 GRF-ZF to one of the ZRs in Top1. Top1-ZR1 was most similar to the NEIL3 GRF-ZF with an RMSD of 1.86 Å for backbone atoms (Figure 23C). A structure-based sequence alignment was performed with the NEIL3 GRF-ZFs and Top-ZR (Figure 23D).

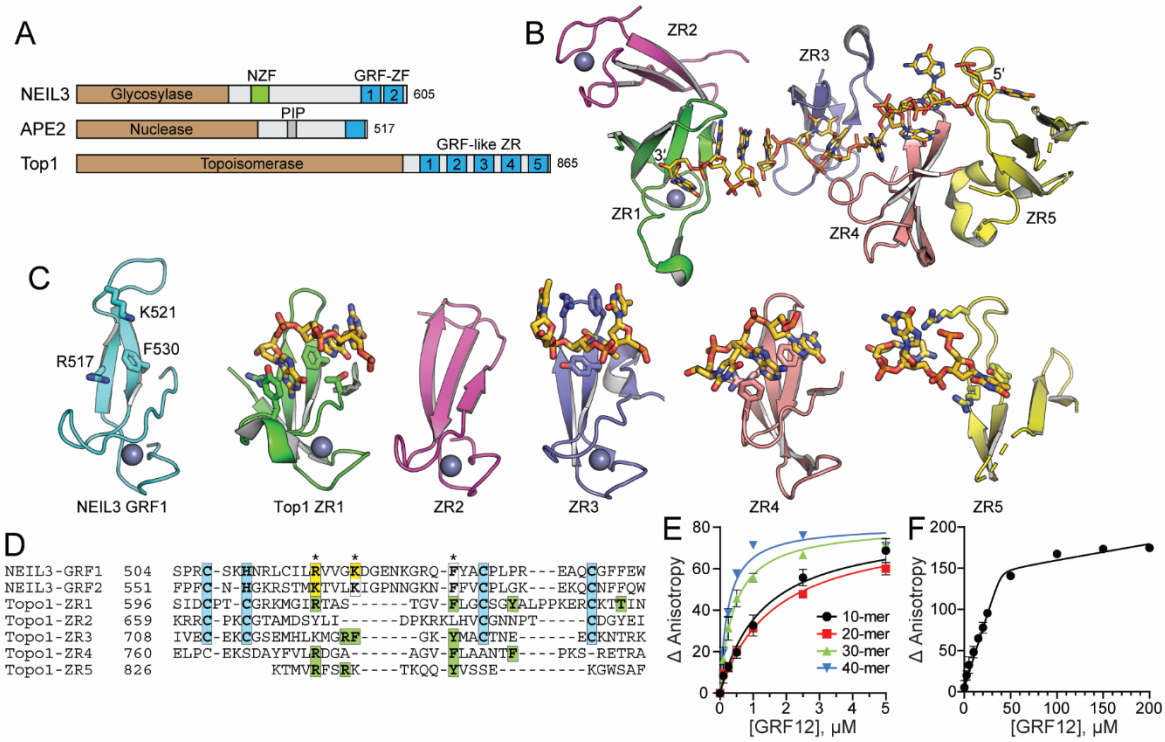

Figure 23. DNA-binding modes by GRF motifs. A) NEIL3, APE2, and *E.coli* Top1 contain GRF-ZF and GRF-like ZR motifs at their C-terminus. B) the structure of the Top1 ZR domain, with DNA shown as sticks (PDB ID 4RUL). C) Comparison of hNEIL3 GRF1 (cyan) with individual Top1 ZR motifs (colored in B). DNA-binding residues and DNA are shown as sticks. D) structure-based sequence alignment. Zn^{2+}-coordinating residues are highlighted cyan. NEIL3 residues that affect DNA binding upon mutation are highlighted yellow and DNA interacting residues in the Top1 structure are green. Asterisks mark the highest conservation among DNA-binding positions. E) DNA binding of mNEIL3 GRF12 to ssDNA of varying lengths. Total [DNA] used was 25 nM. The data are means ± S.D. (n=3). F) stoichiometry of binding of GRF12 to 40-mer ssDNA. Total [DNA]=5.05 μM (>>K_d). The inflection point in the titration curve fits to [GRF12] = 40 μM, which equals an 8:1 GRF12:DNA molar ratio.

Top1-ZR1 interacts with ssDNA via an arginine that in mNEIL3 is Arg^{517} and in hNEIL3 is Lys^{563}. Aromatic residues Phe^{616} in the beta strand of Top1 interacts with ssDNA and this residue corresponds to Phe^{530} in mNEIL3 and Phe^{576} in hNEIL3 (Figure 23C,D). The other Top1-ZRs are similar in that they contain an arginine side chain that contacts the deoxyribose-phosphate backbone of ssDNA and an aromatic residue

(phenylalanine or tyrosine) within the beta strand of the concave cleft that provides nucleobase stacking interactions (Figure 23C,D). Even though all Top1-ZRs have the ability to adopt various conformations, they configure in a similar orientation to bind 2-3 nucleotides of ssDNA along the ZR concave cleft.

We next wanted to investigate if the NEIL3 GRF-ZFs would have a similar DNA footprint to the Top1-ZRs. Since the determined NEIL3 GRF-ZF structure was in the absence of DNA, we turned to biophysical techniques such as fluorescence anisotropy to determine the DNA footprint. We determined the DNA binding of NEIL3 GRF-ZF to ssDNA from 10 to 40 nucleotides in length. NEIL3 GRF-ZF showed a correlation of high binding affinity to longer ssDNA substrates (Figure 23E). The K_d values (μM) of GRF-ZF to ssDNA were 1.8 ± 0.3 for 10mer, 1.3 ± 0.2 for 20mer, 0.4 ± 0.1 for 30mer, and 0.2 ± 0.01 for 40mer. To determine how many nucleotides of DNA were bound to the GRF-ZF domain we performed a stoichiometric experiment with excess ssDNA of 40 nucleotides. The saturation binding curve showed an inflection point (saturation) at 40 μM, which corresponds to 8 molecules of GRF-ZF bound to 40 nucleotides of ssDNA (Figure 23F). Therefore, each GRF-ZF domain would bind 5 nucleotides or each GRF-ZF motif (each GRF motif 1 or 2 is half of the GRF domain) would bind 2-3 nucleotides of ssDNA, which is consistent with the 2-3 nucleotides that each Top1-ZR binds.

The GRF domain inhibits NEIL3 glycosylase activity

Thus far the structural and biophysical results have shown how the NEIL3 GRF-ZFs could bind DNA and the optimal length of DNA substrate. However, it was unknown how the NEIL3 GRF-ZF domain affects NEIL3 activity. The GRF-ZF and ZRs from APE2

and Top1 respectively were shown to enhance catalytic activity of these enzymes (Chen et al., 2012; Tan et al., 2015; Tse-Dinh, 1991; Wallace et al., 2017; Zhang et al., 1996). Next, we tested if the GRF-ZF domain of NEIL3 would also enhance the ICL unhooking activity of full length mNEIL3. NEIL3 has the unique ability to unhook dA-AP ICLs and psoralen-ICLs and is the first vertebrate DNA glycosylase shown to be able to do so (Semlow et al., 2016; Wu et al., 2019). We performed ICL unhooking assays with AP-ICLs under single turnover conditions (Materials and Methods). Three mNEIL3 constructs were expressed and purified from *E.coli*, including NEIL3 full length (wild-type), NEIL3 with double mutant K522E/K568E (GRF^{mut}) in the GRF-ZF domain that ablates DNA binding, and NEIL3 truncated to no longer contain the GRF-ZF domain (ΔGRF) (Figure 24A). We previously published that the NEIL3-GD was unable to unhook AP-ICLs fork substrates when the AP site is on the lagging template strand (Imani Nejad et al., 2020). We found similar results with NEIL3-FL with no activity toward a DNA fork with AP site on the lagging template strand (Figure 24B). However, we observed ICL unhooking and AP lyase activity with the three NEIL3 constructs and substrate containing AP site on the leading template strand (lead fork). Surprisingly, both GRF^{mut} and ΔGRF showed higher ICL unhooking activity compared the GRF-ZF WT, suggesting the GRF-ZF plays an autoinhibitory role (Figure 24).

To confirm the GRF-ZF was solely responsible for inhibition toward NEIL3 activity we performed an *in trans* experiment with NEIL3-GD that contains glycosylase activity (Imani Nejad et al., 2020; Liu et al., 2010). The experiment set-up including the NEIL3-GD and equimolar addition of either GRF12 or $GRF12^{mut}$ (which does not bind DNA),

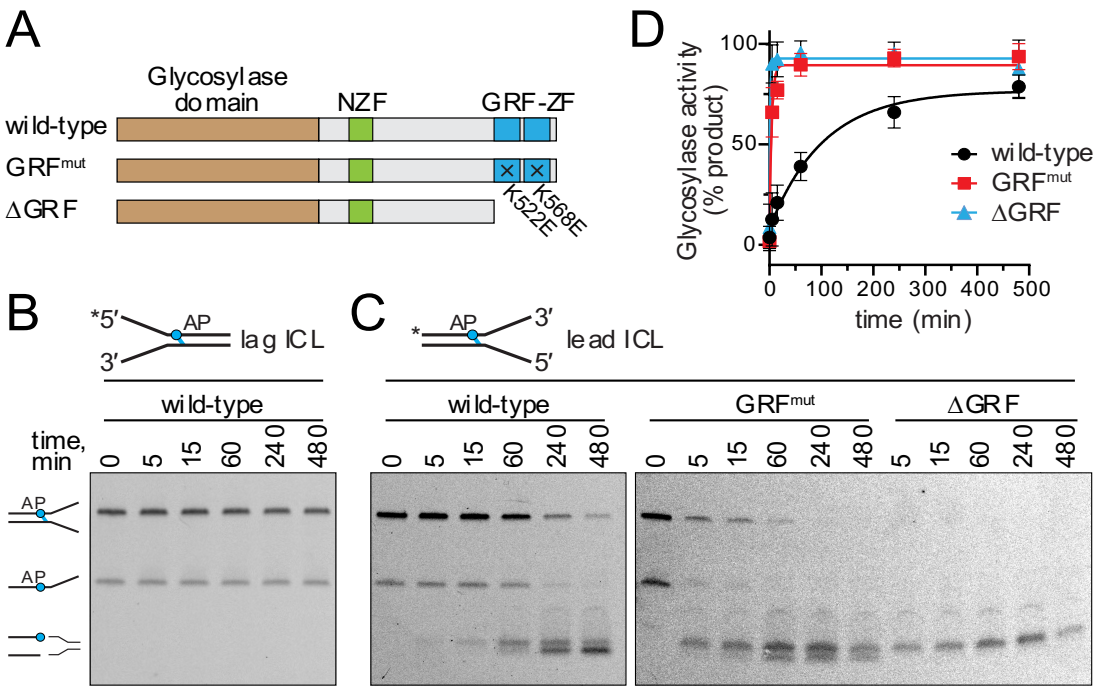

Figure 24. DNA binding by the GRF domain inhibits NEIL3 ICL unhooking activity. A) Schematic of protein used in the experiment. B) Representative gel for ICL unhooking by WT mNEIL3 against a lagging-strand dA-AP ICL fork substrate. C) ICL unhooking activity of WT, GRF^{mut}, and ΔGRF mNEIL3 proteins against the leading-strand dA-AP ICL fork. D) quantification of data shown in B and C (means ± S.D. for three independent experiments).

followed by addition of DNA substrate (Figure 25A). The DNA substrate utilized for base excision activity assays (Materials and Methods section) was small oxidized nucleobase DHT which NEIL3 has previously shown to excise (Imani Nejad et al., 2020). The DHT lesion was introduced into ssDNA (25mer) on the leading template strand of the splayed arm (lead fork) or on the lagging template strand of the splayed arm (lag fork). Addition of GRF12 decreased base excision activity over one hour with ssDNA (Figure 25 B,C, D,E). To confirm if the decrease in base excision activity was due to competition between

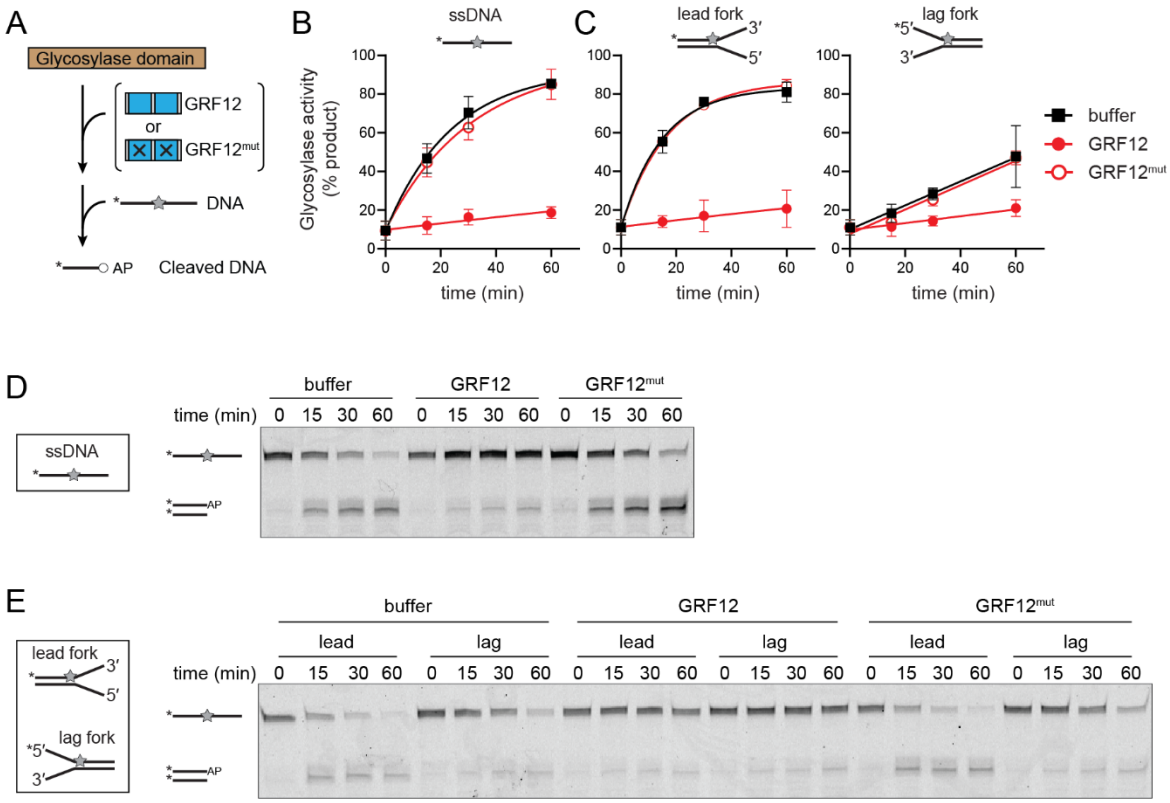

Figure 25. DNA binding by GRF inhibits NEIL3 glycosylase activity *in trans*. A) experimental design. NEIL3 glycosylase domain was mixed with either GRF12, GRF12K522E/K568E (GRF12mut), or buffer prior to incubating with either ssDNA or splayed arms/forks containing a DHT lesion. B and C) quantification of DNA glycosylase activity for ssDNA (B) and fork (C) substrates. The data are means ±S.D. (n=3). D,E) Representative gels used for data shown in B,C. Glycosylase assays for GRF12 inhibition of DHT in ssDNA (D) and forked substrates (E).

the GRF-ZF domain and NEIL3-GD, we replaced the GRF12 with GRF12^{mut} in the base excision activity assay. The addition of GRF12^{mut} did not show inhibition, but similar activity to NEIL3-GD without addition of GRF-ZF domain (Figure 25 B,C,D,E). Therefore, the GRF-ZF domain is responsible for inhibition of NEIL3 unhooking and base excision activity. Previously, we have shown that the mNEIL3-GD has weak ability to excise DHT

from a lagging fork substrate (Imani Nejad et al., 2020). We therefore tested glycosylase activity with DNA substrates containing DHT on the leading versus lagging template strands, labeled lead fork and lag fork respectively (Figure 25C,E). We observed consistent results when GRF or GRF^{mut} was added with lead or lag forks DNA substrates and glycosylase activity was monitored. The GRF-ZF domain shows consistent inhibitory role with ssDNA, lead and lag fork substrates (Figure 25B,C,D,E).

Discussion

Zinc fingers are highly abundant DNA binding motifs that can be found on the C-terminus of DNA repair enzymes such as metazoan APE2 and bacterial Top1 (Tan et al., 2015; Wallace et al., 2017). NEIL3 is a DNA glycosylase that contains two tandem GRF-ZF motifs on its C-terminus. We have shown the structure of the GRF-ZF domain of NEIL3 and that it binds ssDNA in a similar fashion to the GRF-ZF of APE2 and ZRs of Top1. Functionally, the GRF-ZF domain in NEIL3 shows an *in vitro* autoinhibitory role toward base excision and ICL unhooking activity, which is contrary to the enhancing function of other GRF-ZFs (Chen et al., 2012; Tse-Dinh, 1991; Wallace et al., 2017; Zhang et al., 1996). This autoinhibitory role could be explained through the high binding affinity of the GRF-ZF domain toward ssDNA. Also, our work and other publications have shown that NEIL3 has preference toward substrates with ssDNA architecture present (ssDNA, forks) (Ha et al., 2020; Imani Nejad et al., 2020; Liu et al., 2010; Semlow et al., 2016). One reason for NEIL3 inhibition by the GRF-ZF domain is due to competition for the ssDNA substrate. The GRF-ZF domain may be competing against the NEILGD for any available

ssDNA. When we introduced charge reversal mutants (R518E, K522E, K568E) into the GRF-ZF domain that no longer bound DNA, we rescued glycosylase activity as the NEIL3GD was able to freely bind DNA.

Interestingly, this inhibitory role of the GRF-ZF domain has also been observed *in trans* with APE1 downstream BER enzyme in the presence of ssDNA (Ha et al., 2020; Rodriguez et al., 2020). It is unclear if the inhibitory mechanism is due to the inability of APE1 to bind or recognize ssDNA when the GRF-ZF is bound to DNA or if the GRF-ZF binds to APE1 which in turn inhibits DNA binding (Ha et al., 2020). Another option was that the GRF-ZF domain interacts *in cis* with the NEIL3GD to modulate activity. Other deposited protein structures mimic ssDNA binding to APE2 and regulation by binding to ssDNA or other proteins (Chen et al., 2012; Tse-Dinh, 1991; Wallace et al., 2017; Zhang et al., 1996). We hypothesize that the unique C-terminus of NEIL3 including GRF-ZF domain provides NEIL3 with the ability to remove complex DNA lesions such as DNA ICLs in a replication-dependent manner by DNA-binding interactions and protein-protein interactions.

The C-terminal GRF-ZF of NEIL3 aids in its recruitment to an ICL-converged replication fork (Wu et al., 2019). The internal NZF domain of NEIL3 interacts with ubiquitylated CMG helicase while the GRF-ZF has been hypothesized to bind to available ssDNA on the lagging template strand of the converged fork (Wu et al., 2019). Published results showed that if the GRF-ZF domain was mutated to no longer bind DNA, NEIL3 recruitment to the site of damage at a converged fork decreased in *Xenopus* egg extracts. However, these results were not completed *in vitro*, so it maybe that the GRF-ZF domain assists in NEIL3 recruitment to damage in coordination with the internal NZF domain that

binds ubiquitylated CMG helicase (Wu et al., 2019). Our *in vitro* studies show that NEIL3 glycosylase activity increases when the GRF-ssDNA interaction is disrupted. Therefore, the GRF-ZF domain may remain bound to ssDNA at a converged fork. The location and distance of the GRF-ZF domain relative to the site of damage maybe important to either inhibit glycosylase activity when close to the NEIL3GD or allow NEIL3GD to bind DNA when the GRF-ZF is farther away from the DNA lesion. Other NEIL3-replisome contacts such as with the NEIL3-NZF and ubiquitylated CMG may assist in the positioning of the NEIL3GD and GRF-ZF domain along DNA. The DNA sequence dependence has not been rigorously investigated but multiple sequences have been investigated and show sequence independent NEIL3 GRF-ZF binding (Ha et al., 2020). Additionally, the inhibition of the GRF-ZF domain may regulate the timing of NEIL3 activity to only after recruitment to the site of damage within the converged replication fork. This would prevent unintentional DNA strand breakage during DNA replication. Interestingly, the overexpression of NEIL3 GRF-ZF domain has shown a decrease in oxidation stress (alkaline and hydrogen peroxide treatment) in *Xenopus* nuclear extracts (Ha et al., 2020). New *in vivo* studies would be required to shows the biological significance of NEIL3 GRF-ZF.

We find that there are two tandem GRF-ZF motifs to aid in ssDNA binding as tandem GRF-ZFs bind DNA more tightly than individual motifs, which is consistent with other published results (Ha et al., 2020). Additionally, the two unique protomers of the GRF-ZF domain show a conformation flexibility that may bind various ssDNA conformations and other complex DNA substrates. For example, NEIL3 has been shown to be recruited to damage beyond converged replisomes, such as telomeric damage

(Zhou et al., 2017). The GRF-ZF structural flexibility may also hint at possible protein-protein interactions. For example, the GRF-ZF2 contains putative ubiquitylation sites that may bind other unknown protein partners. The GRF-ZF domain of NEIL3 shows interaction with APE1 and PCNA but no binding is observed with individual GRF-ZF motifs (Ha et al., 2020). Full length NEIL3 has been shown to interact with PCNA, flap endonuclease 1 (FEN1), APE1, TRF1, and CMG (Ha et al., 2020; Wu et al., 2019; Zhou et al., 2017). The GRF-ZF has been shown to interact with APE1 (Ha et al., 2020). There is a possibility for more protein partners as NEIL3 has many roles in addition to DNA replication, such as cell signaling, immunity, pulmonary function, myocardial infarction, and ischemic stroke (Chakraborty et al., 2015; de Sousa et al., 2017; Ehlers et al., 2016; He et al., 2016; Jalland et al., 2016; Li et al., 2018; Liu et al., 2013a; Massaad et al., 2016; Olsen et al., 2017; Rognlien et al., 2015; Rumsey et al., 2017; Skarpengland et al., 2015; Tangye, 2016; Yang et al., 2016).

Chapter 4

Discussion and Future Directions

New details of NEIL3-GD substrate specificity

In 2006, NEIL3 was discovered in vertebrates as an ortholog of the bacterial Nei glycosylases that were shown to excise small oxidative DNA lesions (Liu et al., 2013a). It came as a surprise to the glycosylase field that NEIL3 was able to unhook more complex types of DNA damage beyond the canonical small oxidative DNA nucleobases such as psoralen and AP-ICLs (Semlow et al., 2016; Wu et al., 2019). Semlow *et al.* showed that NEIL3 could excise a psoralen-ICL and dA-AP ICL from RNA/DNA hybrid substrates containing one nucleotide on either side of the AP site. NEIL3 unhooked both of these dA-AP ICL substrates and showed lyase activity resulting in cleavage 3' to the AP site (Semlow et al., 2016). There were still many unanswered questions following the discovery that NEIL3 could unhook ICLs. The optimal make-up of the DNA substrate defined by oligonucleotide characteristics, length, and polarity were unknown. Identifying the ICL substrate specificity of NEIL3 would show where its activity would be most plausible within a cell.

We set out to investigate these details of the optimal DNA substrate of NEIL3 with DNA substrates that model a replication fork. In Chapter 2, I showed that the NEIL3 N-terminal glycosylase domain has specificity toward DNA substrates containing a dA-AP ICL on the leading template strand and base pairing up to two base pairs from the lesion. We show that NEIL3 lacks activity toward a fork substrate with the dA-AP ICL on the lagging parental strand. Additionally, the specificity of NEIL3 activity aligns with the high

expression levels of NEIL3 during S-phase when DNA replication is occurring (Zhou et al., 2017). We showed that base pairing close to the DNA lesion also affects NEIL3 specificity. NEIL3 favors DNA fork substrates with 2-base pairs adjacent to the lesion over substrates with 3-8 base pairs next to the lesion. These results support NEIL3's substrate preference of lesions within single stranded DNA and little to no activity toward lesions within duplex DNA (Liu et al., 2010; Liu et al., 2012; Liu et al., 2013b). Our work has focused on a splayed arm DNA substrate with parent template DNA strands. Future work should incorporate a three-stranded fork structure that contains model nascent strands. I hypothesize that NEIL3 would have a similar preference toward a three-strand DNA fork as it did toward a leading fork substrate, with a lesion on the top strand, and nascent DNA present on the leading strand. This hypothesis could be confirmed by glycosylase activity assay with three-stranded fork substrates containing DHT lesion and monitoring activity with NEIL3-GD present.

The optimal substrate also provides a target for x-ray crystallography as there is currently no structure available of NEIL3 bound to DNA. I conducted crystal trials with NEIL3-GD and NEIL3-GD E3Q in the presence of dA-AP ICLs, amino-purines ICLs, and ssDNA containing AP analog tetrahydrofuran (THF) and 1-azaribose, but crystal growth was unsuccessful. The NEIL3-GD crystal structure has been determined and published conditions were repeated for confirmation (Liu et al., 2013b). To attempt to reproduce crystals of NEIL3 from Liu *et al*., the NEIL3-GD was purified with high purity and yield for crystallographic studies (Materials and Methods section). No crystal formation was observed under sitting or hanging drop screening conditions. However, crystals did form in the presence of NEIL3-GD under oil conditions (Hauptman-Woodward Medical

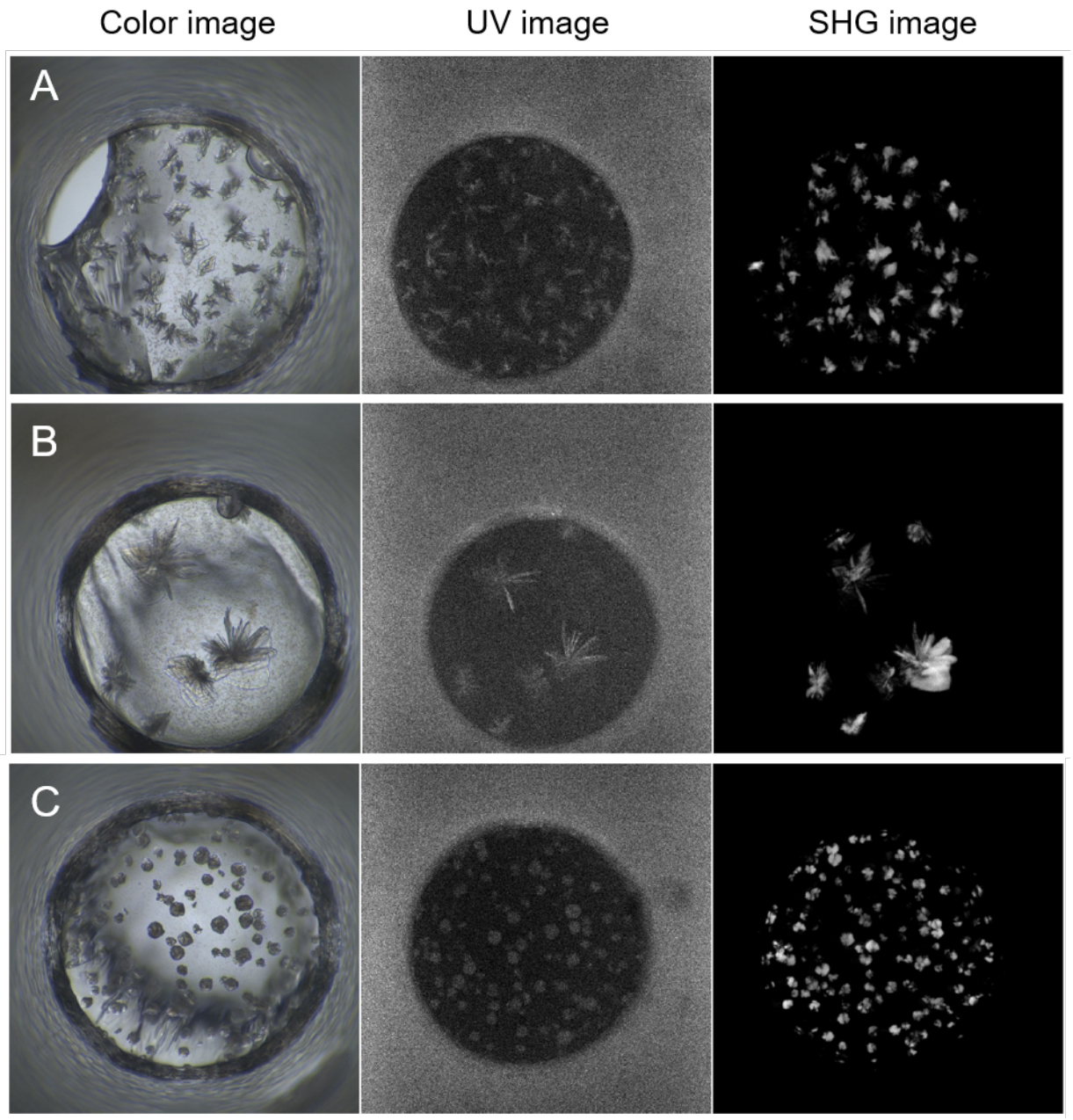

Figure 26. NEIL3-GD crystals under oil from Hauptman-Woodward Medical Research Institute High Throughput Crystallization Screen Center. Color image, UV image to detect fluorescence of protein, and second-harmonic generation (SHG) images to detect chiral crystals were taken. Images were taken after 42 days at 4°C. NEIL3-GD at 11 mg/mL in well condition A) 0.1 M ammonium thiocyanate, 0.1 M Bis-Tris Propane pH 7, 24% (w/v) PEG 20,000 B) 0.1 M ammonium nitrate, 0.1 M Tris pH 8, 20% (w/v) PEG 8000 C) 0.1 M potassium thiocyanate, 0.1 M Tris pH 8, 20% (w/v) PEG 1000.

Research Institute High Throughput Crystallization Screen Center) including 104 positive crystal formations from 1536 crystal conditions (Figure 26). Diffraction data were not collected and therefore these crystals could not be confirmed as NEIL3-GD over salt crystals.

Crystal trials of NEIL3-GD and DNA substrate were also pursued under various spare matrix screens. Drops were set-up with 1:1 ratio of NEIL3-GD and ICL substrates. The dA-AP ICL substrates consisted of oligonucleotides 12-17 in length with dA-AP ICL at the center. Additionally, dA-AP ICL mimics from a 2-aminopurine residue were utilized in crystallographic studies due to their high yield and stability (Nejad et al., 2019). Binding mutant NEIL3-GD E3Q was purified with similar purity and yield as NEIL3-GD and incorporated into crystallographic studies. Perhaps no crystals of the NEIL3-GD in complex with DNA have been published because the complex undergoes a large conformation change that inhibits the formation of a crystal lattice. Structural questions remain of the NEIL3-GD bound to DNA substrate such as an AP site, ICL, or oxidative lesion in ssDNA or fork DNA substrate form.

Another option would be to lock the complex into place with a currently unknown non-unhookable form of the ICL bound to NEIL3-GD. Another approach to reveal the molecular detail of NEIL3 bound to DNA would be to work with recombinantly purified NEIL3-FL. Since NEIL3-FL contains the catalytic glycosylase domain as well as additional zinc fingers in the C-terminus, which we have shown regulate the glycosylase function, a structure of NEIL3FL would be especially beneficial to understand how this molecular machine can unhook ICLs.

Two decades ago, x-ray crystallography was the only technique that could determine structures at sub-atomic resolution. However, the resolution revolution has brought cryo-electron microscopy (cryo-EM) to the forefront due to advances in electron detectors, imaging processing, and computational capabilities (Callaway, 2015; Kuhlbrandt, 2014). Currently, cryo-EM is limited to larger macromolecules with recently more structures determined of biomolecules below 200 kDa (Wu and Lander, 2020). NEIL3-FL is currently 68 kDa in molecular weight with C-terminal dynamic and disordered regions. To provide structure and rigidity, future structural studies may require NEIL3FL-DNA complex or NEIL3-FL to be bound to ubiquitin chains via its internal NZF domain or bound to proliferating cell nuclear antigen (PCNA) via its C-terminal GRF-ZF domain, which would increase the molecular weight to approximately 100 kDa to be a suitable candidate for cryo-EM studies.

Additionally, high resolution structural biology techniques are not required to answer questions about the general shape of native NEIL3-FL. Small angle x-ray scattering (SAXS) would be a potential structural biology technique to view the 3-dimensional envelope and shape of NEIL3-FL by detecting an x-ray diffraction pattern from a beam that passes through the protein in solution (Grawert and Svergun, 2020). The available x-ray crystal structures of NEIL3-GD and GRF-ZF domain could be modeled into the NEIL3-FL envelope to determine domain positioning, architecture, and possible conformational changes. Additionally, conformation changes of NEIL3-FL can be probed by collecting SAXS diffraction data in the presence and absence of DNA substrate. However, due to the dynamic nature of NEIL3 there may be multiple conformations present.

The NEIL3 GRF-ZF domain harbors new roles for NEIL3 glycosylase activity regulation

Unique to NEIL3 is its extended C-terminus filled with zinc fingers with unknown function and structural information. It was also unclear what role, if any, these zinc fingers provided toward the catalytic or glycosylase domain of NEIL3. Additionally, there were many unanswered questions about the structure of the disordered NEIL3 C-terminus. We successfully purified recombinant GRF1, GRF2, GRF1/2 domains of NEIL3 as well as the internal NZF domain. We showed via EMSAs that the GRF-ZF domain bound to ssDNA which was consistent with previous *in vitro* studies and binding studies with endogenously expressed NEIL3 (Ha et al., 2020; Wu et al., 2019; Zhou et al., 2017).

Additionally, we showed that the GRF-ZF domain also bound to fork substrates albeit with the same affinity as ssDNA. We determined the first crystal structure of the NEIL3 GRF-ZF domain that shows a claw-like structure with DNA binding surface along the front face of the domain. I have identified the DNA binding cleft by mutating residues along the DNA binding cleft and observing ablation of DNA binding. Future crystallographic studies will be required to expose the binding details of GRF-ZF domain with ssDNA and provide evidence to our hypothesis that the GRF-ZF domain is flexible and can accommodate various ssDNA substrate architectures. Even though the GRF-ZF domain binds tightly to DNA, a challenge of x-ray crystallography of the GRF-ZF-ssDNA complex may be that a large conformational change is present at the interdomain hinge. While such flexibility may assist in binding a large array of ssDNA conformations, it may also impede crystal formation by interfering with crystal packing.

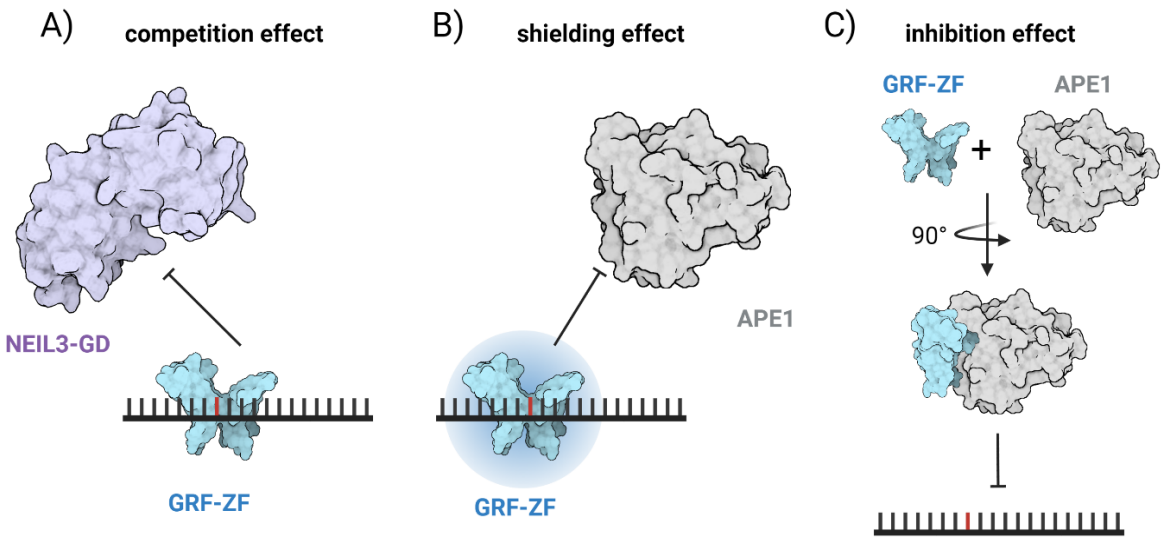

Figure 27. NEIL3 GRF-ZF domain inhibitory mechanisms. A) Competition effect results in GRF-ZF (blue, PDB 7JL5) outcompeting NEIL3-GD (purple, PDB 3W0F) for available ssDNA containing DNA lesions (red). B) The GRF-ZF domain binds ssDNA containing damage and blocks APE1 (gray, PDB 4QHD) from binding in the shielding effect. C) GRF-ZF and APE1 bind at the DNA binding interface and form a GRF-ZF-APE1 complex that cannot bind ssDNA. Competition effect data published and cited (Rodriguez et al., 2020). Shielding and inhibition effect data from separate publication (Ha et al., 2020).

Lastly, we uncover the role of the GRF-ZF domain as an autoinhibitory domain to the NEIL3-GD and confirm these results *in cis* and *in trans*. Interestingly, the NEIL3 GRF-ZF domain has also shown an inhibitory role toward APE1 activity, the second enzyme in the BER pathway (Ha et al., 2020). In Ha *et al.*, two types of effects were investigated and discussed, the shielding and inhibition effect. The shielding effect results from the GRF-ZF domain binding to the AP-site within ssDNA and blocking APE1 from binding ssDNA. Presumably, the AP site is a result of NEIL3-GD activity at the DNA damaged lesion. Our results complement this shielding effect in regards to NEIL3 as we show that

the GRF-ZF domain outcompetes the NEIL3-GD for ssDNA and blocks NEIL3-GD from binding and excising the damaged DNA lesion. The inhibition effect proposes that the GRF-ZF domain binds to APE1 through protein-protein interactions and the GRF-ZF-APE1 complex cannot bind ssDNA (Ha et al., 2020). Our structural studies of the GRF-ZF domain are in favor of the GRF-ZF-APE1 complex interacting at the positively charged cleft of NEIL3 GRF-ZF. The results of the known inhibitory effects of the NEIL3 GRF-ZF domain are shown in Figure 27 and include our proposed competition effect (Rodriguez et al., 2020) and the published shielding and inhibitory effect (Ha et al., 2020). In review, the competition effect is similar to the shielding effect in that the GRF-ZF domain inhibits binding to DNA by the NEIL3-GD and APE1 protein respectively.

In mammalian cells, capping mechanisms are present to protect the end of chromosomes from being flagged as a double stranded DNA break. The chromosome ends are protected by telomeres and telomere binding proteins that inhibit the recognition of dsDNA ends by DNA damage response (DDR) proteins (Zhou et al., 2017). NEIL3 is present at telomeres during S/G2 phase and interacts with shelterin complex protein telomeric repeat factor (TRF1). The TRF homology (TRFH) domain of TRF1 was also shown to inhibit NEIL3-GD activity in the absence of the GRF-ZF domain. From our published results where we show inhibition of NEIL3-GD by the GRF-ZF domain (Rodriguez et al., 2020), it is possible that the GRF-ZF domain interacts with TRF1 domain bound to ssDNA as a protective method for telomeres (Zhou et al., 2017).

Our results are in favor of a model that currently shows NEIL3 and the location and function of each domain bound to DNA at the replisome during ICL repair. This model originally proposed the NZF domain interacts with ubiquitinated CMG helicase to recruit

NEIL3 to the site of damage, the NEIL3-GD bound at a replication fork converged at an ICL, and the GRF-ZF domain bound to available ssDNA on the lagging parental strand (Wu et al., 2019). Our DNA binding data confirms that the GRF-ZFs bind ssDNA and NEIL3-GD binds a dA-AP ICL lesion. Even though this initial model predicted where each domain would be located on a converged fork, it lacked details of where and how each domain would bind to their protein partner or DNA substrate. Our results provide structural and biochemical details to add to this model, such as the GRF-ZF domain binding 5 nucleotides of ssDNA and that the positioning of the NEIL3 GRF-ZF domain relative to the NEIL3-GD domain could elicit a regulatory role (Figure 28).

If the GRF-ZF domain was too close to the NEIL3-GD domain, glycosylase activity would be repressed while if the GRF-ZF domain was far enough from the NEIL3GD, this might allow glycosylase activity. Future studies are required to examine the DNA length at which autoinhibition of the NEIL3-GD domain by GRF-ZF domain would be impeded. This could be done via single molecule techniques that are independent of population bulk experiments and can focus on individual biomolecules. Fluorophores can be added to NEIL3-GD, GRF-ZF domain, and site-specific DNA with Alexa fluorophores, Janelia fluorophores, and N-hydroxysuccinimide-ester respectively for all three components to be visualized via fluorescence microscopy. The NEIL3-GD can be anchored to one end of ssDNA at various lengths containing an oxidative lesion such as DHT. Fluorescence resonance energy transfer (FRET) analysis between NEIL3-GD and the GRF-ZF domain would give a relative signal if the proteins were within 2-8 nm of each other or approximately 6-24 nucleotides of DNA (Mohapatra et al., 2020). The GRF domain might position or anchor the NEIL3 domain for optimal activity at the lesion site *in cis*. It is

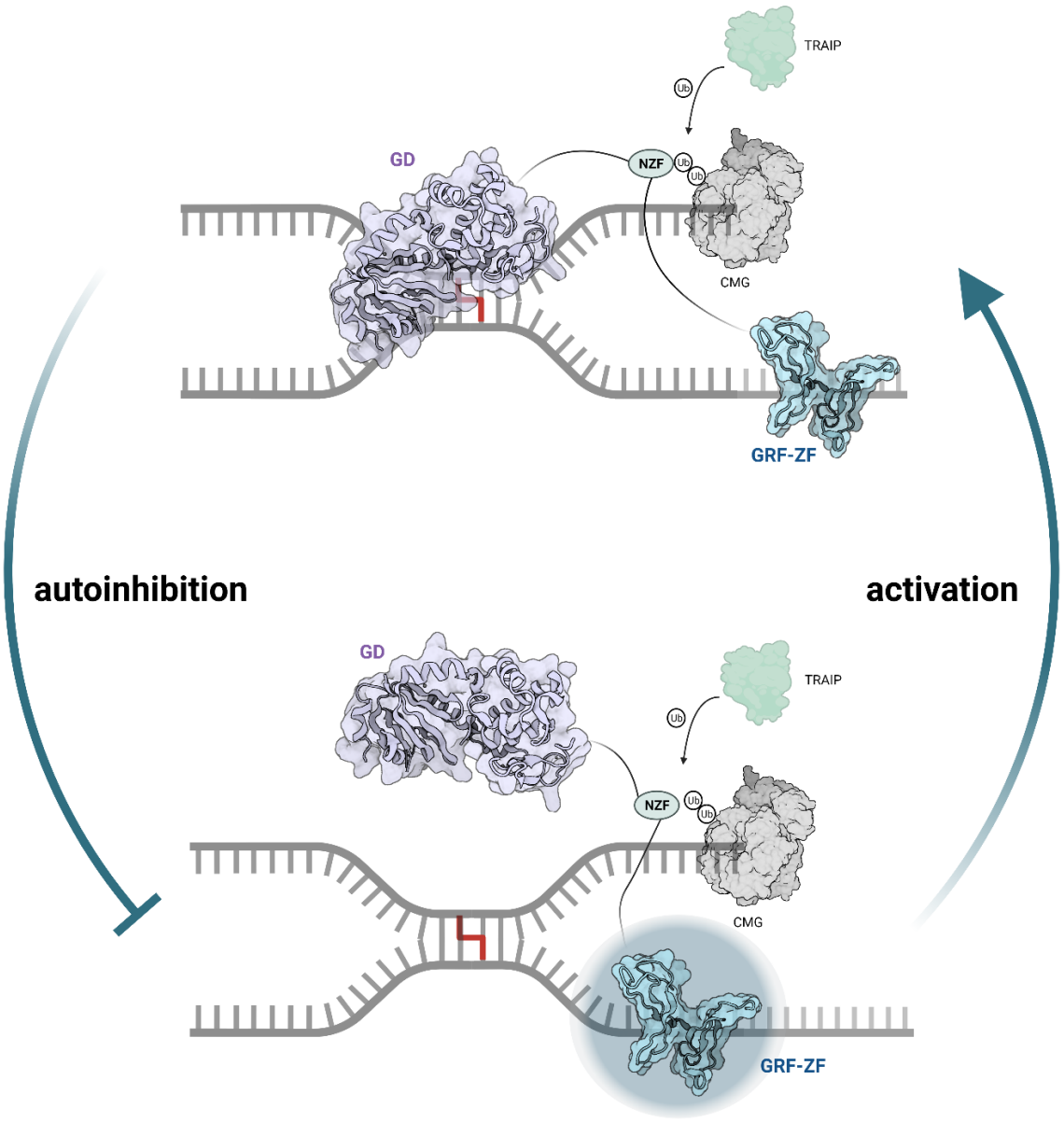

Figure 28. Regulated NEIL3 activity at a fork-converged ICL. The GRF-ZF domain on lagging parental DNA strand inhibits NEIL3-GD activity when in close proximity to the NEIL-GD. NEIL3-GD is active when the GRF-ZF domain is far away from the NEIL3-GD. ICL is shown in red at converged replication forks. E3 ubiquitin ligase TRAIP adds a short ubiquitin chain (Ub) replicative CMG helicase in gray on parental DNA leading strand. NEIL3-NZF in green binds to ubiquitinated CMG helicase. PDB ID for NEIL3-GD (3W0F), GRF-ZF (7JL5), CMG (5U8S). Proteins and model are not to scale. Updated structural details and regulatory role by GRF-ZF domain added to initial published model (Wu et al., 2019).

unknown how the GRF domain distinguishes the correct length for DNA from the NEIL3-GD. A method that allows for visualization of the excised DHT lesion in addition to the NEIL3-GD and GRF-ZF domain would provide the answers on how these two domains communicate with each other and regulate NEIL3 activity.

Our updated model also does not answer the sequence of events that occur at the converged fork. One cycle of events may include the NZF binding to ubiquitylated CMG first, followed by GRF-ZF domain binding to ssDNA on the parental lagging strand, and finally NEIL3-GD binding to the DNA lesion site within ssDNA region upon regulation by GRF-ZF domain. Detailed kinetic studies would need to recapitulate the replisome and NEIL3 assembly at the converged replication fork with ICL. E3 ubiquitin ligase TRAIP adds ubiquitin to replicative CMG helicase and distinguishes two ICL pathways, either NEIL3 or FA dependent with addition of short or long ubiquitin chain respectively, in *Xenopus* egg extracts and human cells (Li et al., 2020; Wu et al., 2019). It is unclear how many ubiquitin molecules represent a short chain and how many ubiquitin molecules are required for the FA pathway to occur. The NEIL3 NZF-ubiquitin interaction will be discussed further in the following discussion section.

The long C-terminal extension of NEIL3 also raises the question if these domains are involved with protein-protein interactions like the C-terminus of NEIL1 (Zhu et al., 2016). The GRF domain of NEIL3 has shown interactions with PCNA and contains a conserved PCNA-binding motif (QILDEEF) between the NZF and GRF-ZF domain, as identified by sequence comparison to other PCNA binding motifs (Ha et al., 2020; Liu et al., 2013a; Zhou et al., 2017). PCNA is a trimeric ring that encircles DNA, interacts with proteins via the PIP box to recruit DNA repair machinery, and activates DNA damage

response pathways via post-translational modifications (PTMs), such as ubiquitination and sumoylation (Moldovan et al., 2007; Sun and Chen, 2004). PCNA is monoubiquitinated within the error-prone translesion synthesis (TLS) DNA repair pathway and polyubiquitinated in the error-free DNA repair pathway (Sun and Chen, 2004).

Additional protein-protein interactions have been shown between NEIL3 and FEN1 and APE1 (Ha et al., 2020; Zhou et al., 2017). Identification of NEIL3 protein binding partners via mass spectrometry and nuclear cell pull-downs may reveal new NEIL3 function as a protein-protein complex. In addition to protein-protein interactions, the NEIL3 C-terminus has available sites of post-translational modifications. For example, GRF-ZF2 has ubiquitination sites at K558, K563, K567, K574, K583, K585, K600 as identified on www.phosphosite.org. Lysine 316 and 331 are also available for ubiquitin interaction. The GRF-ZF domain would not be the only domain available to interact with ubiquitin as the NZF domain of NEIL3 shows binding to ubiquitin (Wu et al., 2019). Since the GRF2 motif has more potential PTM sites than GRF-ZF1, its main role may be for protein-protein interactions (Ha et al., 2020).

Another question that remains is if there are other ICL repair enzymes similar to NEIL3 that contain a GRF-ZF domain. One such protein includes the zinc finger GRF-type containing 1 (ZGRF) also known as c4orf21. ZGRF shows 5' to 3' helicase activity and ZGRF null cells are sensitive to ICL agents mitomycin C and camptothecin (Adamson et al., 2012; Brannvoll et al., 2020; Smogorzewska et al., 2010). ZGRF contains a conserved N-terminal DUF2439 domain, an internal GRF-ZF domain, and C-terminal helicase domain. The internal ZGRF-GRF-ZF domain is 44% identical and 58% similar to the NEIL3 GRF-ZF1 motif and shares 40% identity and 57% similarity to NEIL3 GRF-ZF2.

The ZGRF is thought to be involved with ICL repair involving homologous recombination (HR) as observed by ZGRF and HR-protein RAD51 interaction (Zhou et al., 2017). Published studies also show that ZGRF can dissociation HR DNA substrate intermediate, D-loops, into ssDNA products in an ATP-dependent manner showing that ZGRF can dissociate D-loops during HR (Zhou et al., 2017). Further investigation is required to determine if the ZGRF GRF-ZF domain acts similarly to the NEIL3 GRF-ZF domain and inhibits helicase activity of ZGRF. Cell cycle dependent expression of ZGRF is also unknown and should be investigated.

NEIL3 shows high expression during S-phase and is highly expressed in replicating cells such as stem cells and cancer cells and is present at the replisome (Klattenhoff et al., 2017). We show that the NEIL3-GD has preference for DNA substrates that mimic a model replication fork, consistent with NEIL3 being present at a replication fork. High levels of RPA were observed in NEIL3 deficient cells showing the high levels of ssDNA present in the absence of NEIL3 (Klattenhoff et al., 2017). It is unknown if a physical interaction is present between the two ssDNA binding proteins, RPA and NEIL3, and if the two proteins signal and coordinate action on the presence of ssDNA damage (Klattenhoff et al., 2017). It is also unclear what unidentified proteins are present at the replisome that could be signaling DNA damage to NEIL3. Potential binding partners may include the nine hub genes that are co-expressed with NEIL3 such as topoisomerase 2α (TOP2A), ubiquitin conjugating enzyme E2C (UBE2C), and the cell-cycle proteins such as cyclin A2 (CCNA2) and cyclin B2 (CCNB2) (Zhao et al., 2021).

New ICL repair pathways

NEIL3 removes a broad range of DNA damage in the form of small oxidative lesions and more complex ICLs. An understanding of NEIL3 DNA damage preference is emerging following the discovery of new synthesized DNA lesions with potential of existing *in vivo*. A new type of endogenous acetaldehyde-ICL (AA-ICL) has been recently discovered and synthesized. AA-ICLs form from acetaldehyde, a by-product of alcohol consumption, and form a N^2-propanoguanine precursor ICL with guanine that when present within a 5'-CpG sequence forms three AA-ICL crosslinks with the N^2-amine of guanine on the opposite DNA strand (Hodskinson et al., 2020). The mechanism of AA-ICL repair is unknown, however what is known is that AA-ICLs are repaired by a non-FA pathway that requires fork convergence and avoids single and double DNA strand breaks (Amunugama and Walter, 2020; Hodskinson et al., 2020). Fork reversal is plausible yet it is unclear if the mechanism is enzyme initiated or mechanically driven.

Future studies are required to identify the fork reversal enzyme at play such as eukaryotic fork reversal enzymes zinc finger RAN-binding domain-containing protein (ZRANB3), helicase-like transcription factor (HLTF), or SW1/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1). Even though NEIL3 was not able to excise AA-ICLs (Hodskinson et al., 2020), the identity of an AA-ICL repair enzyme would add new insight into the diverse repair mechanisms of ICL repair. The ICL repair enzyme could be identified by creating a plasmid with an AA-ICL and incubating it in *Xenopus* nuclear extracts followed by pulling out the enzymes that bind to the AA-ICL containing plasmid. Mass spectrometry could identify the protein and fork reversal assays with AA-ICL containing DNA substrates and

recombinantly purified fork reversal enzyme would confirm the identity of the AA-ICL repairing enzyme. Since DNA glycosylases are emerging as enzymes that can unhook ICLs, it would be fascinating to test if an enzyme within the vertebrate DNA glycosylase family can unhook an AA-ICL.

DNA-protein crosslinks (DPCs) are another form of endogenous DNA crosslink. The human HMCES and *E.coli* YedK are capable of creating stable DPCs *in vitro* to assist in targeting proteins for degradation (Thompson et al., 2019). Structural studies have revealed the chemical nature of the DPC including a stable thiazolidine linkage between an open AP site and YedK. The DPC was formed from YedK protein, yet it is unknown if an enzyme can unhook or excise the DPC lesion. NEIL3 is a possible candidate for unhooking a crosslink of this nature. ICL unhooking assays could be performed with DPCs and NEIL3-FL to identify activity toward a DPC lesion. If activity was present, the analysis could be repeated with the NEIL3-GD to identify if the catalytic domain can remove the DPC independent of the C-terminal extension as observed with other DNA lesions (Imani Nejad et al., 2020). There are bound to be additional NEIL3 DNA substrates currently unknown at this time because the DNA lesions have not been discovered and/or synthesized yet for *in vitro* study.

Stoichiometry and specificity of ubiquitination of the NEIL3 NZF

The internal NZF domain of NEIL3 has been shown to bind polyubiquitinated replicative CMG helicase (Wu et al., 2019), however the nature of the ubiquitin state is unknown. Here I will present my preliminary results on the nature and stoichiometry ubiquitin bound to the NEIL3-NZF domain. I will include an introduction to ubiquitin

including its cellular role as a post-translation modification, how ubiquitin is applicable to DNA repair, and the biophysical ubiquitin essay that I have established. In addition, I will speculate on the binding nature of the NEIL3-NZF relative to other NZF containing proteins. I will also provide structural evidence for the proposed binding interface of the NEIL3-NZF toward ubiquitin.

Ubiquitin is the one of the most common post-translation modifications with regulatory roles in DNA repair, protein transport, cell signaling, and proteasomal degradation (Hershko and Ciechanover, 1998; Hicke, 2001; Mirsanaye et al., 2021; Oh et al., 2018; Pickart, 2001). Ubiquitylation or ubiquitination is the process whereby a monoubiquitin biomolecule, multiple monoubiquitin molecules (multimonoubiquitin) or chains of ubiquitin (polyubiquitin) are covalently bound to protein (Oh et al., 2018). Ubiquitin molecules are added in an ATP-dependent manner and sequential process. Ubiquitin activating-enzymes (E1), conjugation-enzymes (E2), and ligating-enzymes (E3) add the additional ubiquitin or ubiquitin chain (Bergink and Jentsch, 2009). Specificity and regulatory action results from the 700 E3-ubiquitin ligases in the human genome, in contrast to the two E1-activating and 40 E2-conjugating ubiquitin enzymes in humans (Ebner et al., 2017). Ubiquitination is reversible and deubiquitinases (DUBs) in humans remove ubiquitin (Ebner et al., 2017).

Polyubiquitin chains provide functional specificity and regulation, dependent on the type of ubiquitin linkage. A diubiquitin molecule, consisting of two ubiquitin molecules, can be conjugated from the C-terminus of one ubiquitin molecule via an isopeptide bond to an ε-amino group of a lysine residue in the adjacent ubiquitin. There are seven lysine amino acids in ubiquitin (K6, K11, K27, K29, K33, K48, K63) that are available for

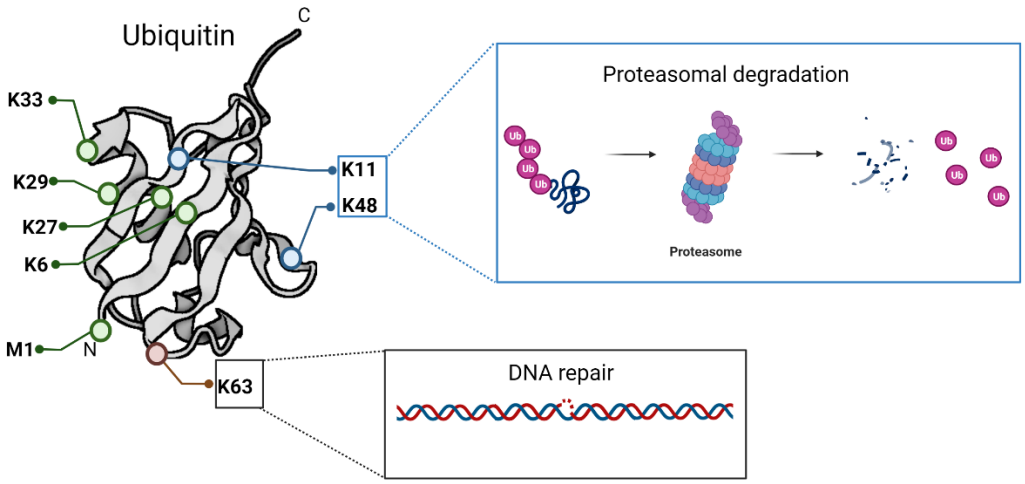

Figure 29. Lysine residues within ubiquitin available for linkage specific ubiquitination. Seven available ubiquitination sites are available via lysine residue conjugation (K6, K11, K27, K29, K33, K48, K63) and N-terminal methionine (M1). K11 and K48 (blue) are involved with proteasomal degradation, while K63 (red) has a role in DNA repair.

conjugation (Figure 29). The C-terminal residue (G76) of one ubiquitin molecule can be covalently linked via a peptide bond to the N-terminal methionine (M1) of the adjacent ubiquitin molecule to form a linear ubiquitin chain in a head-to-tail fashion (Komander et al., 2009). The C-terminus of the proximal ubiquitin binds to the amino lysine residue of the distal ubiquitin or receptor ubiquitin. M1 and K63-linked ubiquitin are involved with non-degradation processes such as DNA repair and kinase activation in the NF-κβ transcription factor pathway (Oh et al., 2018). Unique hybrid branched chains have also been observed with M1/K63 hybrid chains involved in the linear ubiquitin chain assembly complex (LUBAC) pathway (Oh et al., 2018). The K48 and K11-linked diubiquitin chains target proteins for 26S proteasomal degradation (Oh et al., 2018).

Ubiquitin is a small, stable 76 amino acid protein and contains a beta-sheet and an α-helix (Vijay-Kumar et al., 1987) (Figure 29). A hydrophobic core within ubiquitin leads to few solvent exposed residues, such as isoleucine-44 and valine-70. Even though linear, K63-linked, and K48-linked diubiquitin are identical in their overall chemical composition, each has unique structural distinctions. K48-linked ubiquitin chains show a closed conformation with hydrophobic residue (L8, I44, V70) interacting with both ubiquitin molecules.

However, the K63-linked diubiquitin shows an open, extended conformation with hydrophobic core residues exposed (Komander et al., 2009; Varadan et al., 2004). Additionally, the only site of contact between the ubiquitin molecules is at the K63-linkage site and no additional contacts between the ubiquitin molecules were present, as observed in K48-linked diubiquitin. Many proteins bind ubiquitin via a ubiquitin binding domain (UBD) such as a ubiquitin binding zinc fingers (UBZ) (Toma et al., 2015). One type of UBZ is found in the nuclear protein localization 4 protein (Npl4), also known as the nuclear pore protein RAN binding protein (Ranbp) zinc finger, which is involved in ubiquitin processing for cell signaling and protein-protein interactions (Wang et al., 2003). The short 30 amino acid long Npl4-zinc finger (NZF) was first discovered in the Npl4 protein which interacts with ubiquitin fusion degradation 1 (Ufd1) protein to create an AAA ATPase adaptor complex (Alam et al., 2004; Meyer et al., 2002; Nakielny et al., 1999; Wang et al., 2003; Yaseen and Blobel, 1999). The NZF contains consensus sequence X(4)-W-X-C-X(2)-C-X(3)-N-X(6)-C-X(2)-C-X(5) with four cysteines that coordinate one zinc ion similar to the rubredoxin knuckle fold in metalloproteins (Alam et al., 2004; Falquet et al., 2002; Wang et al., 2003).

The NZF domain is found in approximately 200 proteins such as kinase binding partners and DNA replication and repair proteins (Kulathu et al., 2009; Sato et al., 2011; Sato et al., 2009; Wang et al., 2003). The NZF motif has a conserved TF-X(10)-Φ (TF/Φ) which is characterized by a threonine and phenylalanine residue followed by ten binding protein domain (TRABID), ZRANB3, and NEIL3 (Figure 30A). Some of these NZF domains have shown biochemical, biophysical, and structural interactions with ubiquitin species *in vitro*. For example, the Npl4-NZF domain, which is a co-factor of ubiquitin associated ATPase p97, has a TF/M motif that binds to monoubiquitin (Alam et al., 2004; Mirsanaye et al., 2021). The TF/M motif was inserted into the non-ubiquitin binding NZF domain of Ranbp2 to form a chimeric protein and ubiquitin binding resulted (Alam et al., 2004).

The transforming growth factor-β-activated kinase 1 (TAK1) contains an adaptor subunit TAK1 binding protein (TAB) domain involved with the nuclear factor κB (NF-κB) and Jun N-terminal kinase (JNK) pathway associated with immune and inflammatory cellular responses (Kanayama et al., 2004; Kulathu et al., 2009; Sato et al., 2009). The NZF of TAB2/TAB3 binds K63-linked di- and triubiquitin, not linear diubiquitin, and contains a complete TF/Φ motif. Structures of the TAB proteins bound to diubiquitin show two NZF-ubiquitin (Ub) binding sites at the isoleucine-44 hydrophobic patches of ubiquitin and no interaction at the site of the K63-linkage (Kulathu et al., 2009; Sato et al., 2009). Oppositely, the HOIL1L-NZF, with additional R208-P209 dipeptide sequence, binds linear diubiquitin and also contains a complete TF/Φ motif (Sato et al., 2011). Vps36-1 is a eukaryotic yeast protein that contains a complete TF/Φ motif while TRABID and fork-

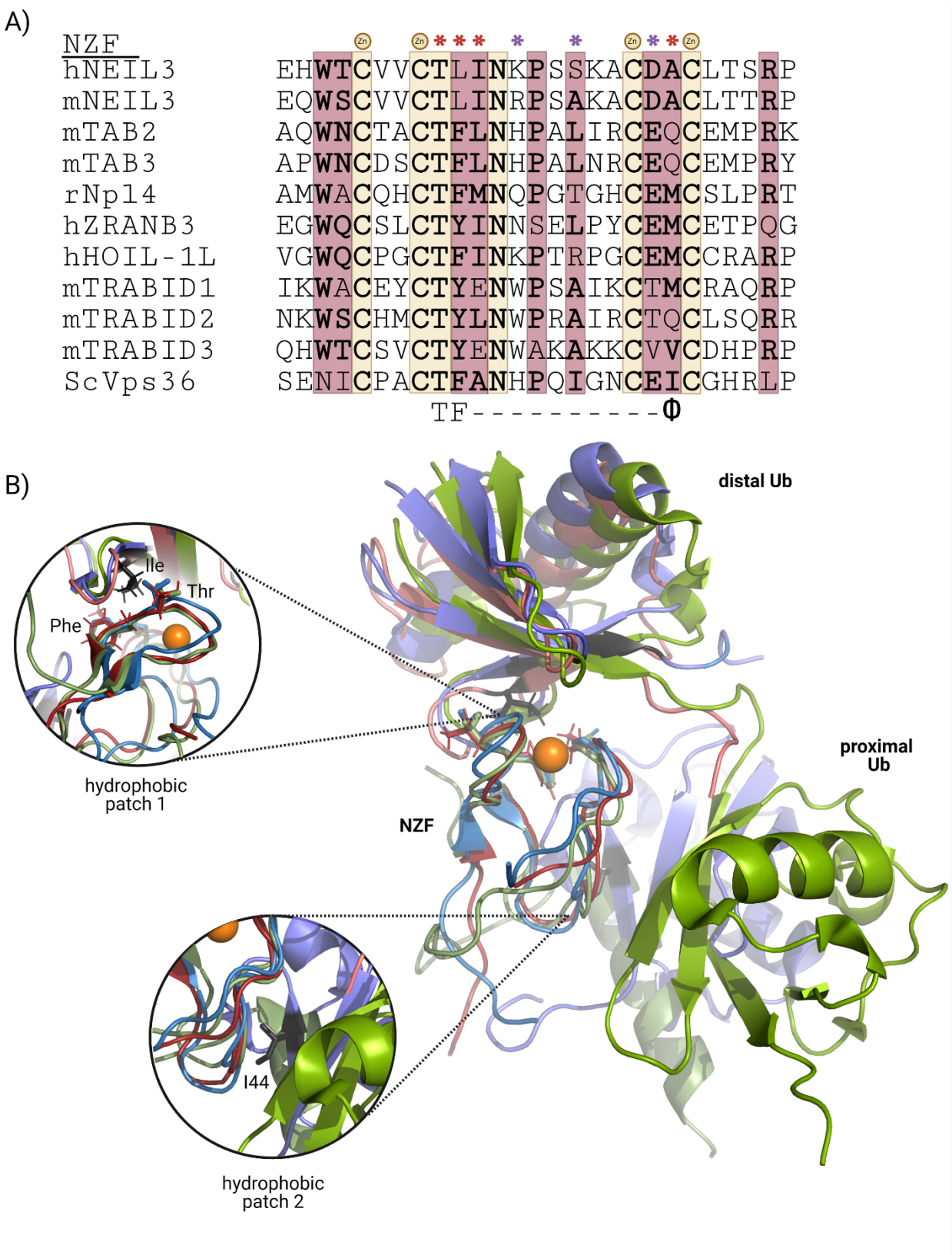
A)
NZF
hNEIL3 EHWTCVVCTLINKPSSKACDACLTSRP
mNEIL3 EQWSCVVCTLINRPSAKACDACLTTRP
mTAB2 AQWNCTACTFLNHPALIRCEQCEMPRK
mTAB3 APWNCDSCTFLNHPALNRCEQCEMPRY
rNp14 AMWACQHCTFMNQPGTGHCEMCSLPRT
hZRANB3 EGWQCSLCTYINNSELPYCEMCETPQG
hHOIL-1L VGWQCPGCTFINKPTRPGCEMCCRARP
mTRABID1 IKWACEYCTYENWPSAIKCTMCRAQRP
mTRABID2 NKWSCHMCTYLNWPRAIRCTQCLSQRR
mTRABID3 QHWTCSVCTYENWAKAKKCVVCDHPRP
ScVps36 SENICPACTFANHPQIGNCEICGHRLP
TF----------Φ
B)
distal Ub
Ile
Thr
Phe
hydrophobic patch 1
NZF
proximal Ub
I44
hydrophobic patch 2

Figure 30. Sequence and structural alignment of NZF containing proteins. A) Sequence alignments of NZF domain or motifs of human NEIL3 (hNEIL3), *Mus musculus* NEIL3 (mNEIL3), *Mus musculus* TAB2 (mTAB2), *Mus musculus* TAB3 (mTAB3), rat Npl4 (rNpl4), human ZRANB3 (hZRANB3), human HOIL-1L (hHOIL-1L), *Mus musculus* TRABIDs (mTRABID1, mTRABID2, mTRABID3), *Saccharomyces cerevisiae* Vps36 (ScVps36). Identical residues (yellow) and similar residues (rose) are highlighted. Above the residues, zinc coordinating cysteines (Zn circle above residues), distal (red asterisk) and proximal (purple asterisk) ubiquitin interactions are marked. The TF/Φ motif is shown below aligned sequences. B) Overlay of rNpl4-NZF and ubiquitin (red, PDB IQ5W), mTAB2-NZF and K63-linked diubiquitin (purple, PDB 3A9J), and hHOIL-1L-NZF and linear diubiquitin (green, PDB 3B08). Distal and proximal ubiquitin molecules are labeled. Hydrophobic patch 1 and 2 show isoleucine 44 (black) on each ubiquitin molecule. Zoomed in circle of hydrophobic patch 1 shows back view of interaction between isoleucine 44 and conserved NZF threonine (Thr) -phenylalanine (Phe) motif.

remodeler ZRANB3 contain an incomplete TF/Φ motif (Figure 30A). Sequence alignments with other NZF containing proteins suggests that ZRANB3-NZF has two ubiquitin binding sites that may bind linear, K63-linked, and/or K48-linked diubiquitin (Kulathu et al., 2009).

An overlay of the determined crystal and NMR structures are shown for the NZF containing proteins, Npl4, TAB2, and HOIL-1L bound to monoubiquitin, K63-linked ubiquitin and linear diubiquitin respectively (Figure 30B). Regardless of the linkage or substrate stoichiometry, all ubiquitin-NZF interactions occur at the hydrophobic patches surrounding isoleucine 44. Hydrophobic patch 1 shows interactions between isoleucine-44 and the conserved TF motif located in the NZF loop (Figure 30B). NEIL3 contains an internal NZF domain that binds to ubiquitylated CMG helicase. E3-ubiquitin ligase TRAIP ubiquitinates replicative CMG helicase at converged replication forks with a DNA ICL. A short chain of ubiquitin by TRAIP leads to ICL unhooking by the NEIL3 while longer ubiquitin chains result in the FA pathway to resolve the DNA ICL (Wu et al., 2019). Currently, the stoichiometry and specificity of the ubiquitin chain that TRAIP adds to

replicative CMG helicase to distinguish between the NEIL3 and FA pathway is unknown and ongoing research to answer this question is underway. NEIL3-NZF contains a TL/A motif instead of the conserved TF/Φ motif in TAB-NZF domains (Figure 30A). Additionally, there is no structural data available to show the molecular interaction between NEIL3-NZF and ubiquitin species. Future work on this project will investigate the ubiquitin binding and structural targets of the NEIL3-NZF interaction while probing the variability in the TF/Φ motif of the NEIL3-NZF.

I have introduced a short fluorescein arsenical hairpin (FlAsH) tetracysteine peptide (CCPGCC) onto the C-terminus of the NEIL3-NZF. FlAsH tags in the presence of biarsenical reagent (FlAsH-EDT_2) emit fluorescence and are powerful, non-invasive probes to monitor protein dynamics (Griffin et al., 1998; Pomorski and Krezel, 2020). I have purified the NZF domain of NEIL3 and ubiquitin species monoubiquitin (Ub) and linear diubiquitin (linear Ub_2) while Carl Schiltz purified K63-linked diubiquitin (Figure 31B). I added monoubiquitin, linear-Ub_2, and K63-linked diubiquitin to fluorescent NZF-FlAsH and monitored the binding interaction via fluorescence anisotropy (Figure 31A).

Future purification optimization is required to confirm pure K63-linked diubiquitin from ubiquitinated E2 ligase by western blot with antibodies for ubiquitin and E2 ubiquitin-conjugating enzyme Ubc13 (Materials and Methods). Preliminary anisotropy results do not show a strong preference toward one ubiquitin species as seen for the TAB2-NZF (Kulathu et al., 2009) (Figure 31C). We do not observe quenching of the FlAsH tag as shown by consistent total intensity throughout the anisotropy experiments with increasing concentration of ubiquitin species (Figure 31D). These preliminary results may

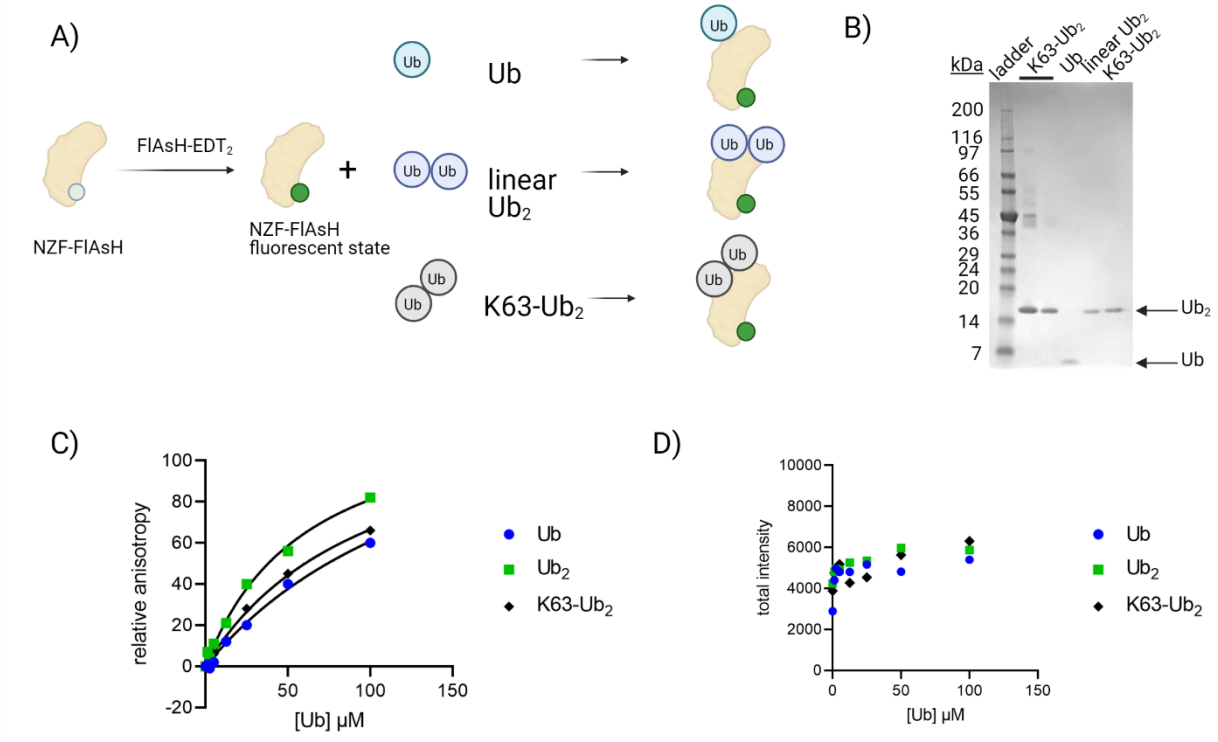

Figure 31. Ubiquitin binding assay of NEIL3-NZF. A) Experimental set-up of ubiquitin binding to NEIL3-NZF. A short FlAsH tag (CCPGCC) was added to the C-terminus of the NZF domain of NEIL3. Addition of FlAsH-EDT_2 reagent results in fluorescence. Ubiquitin species of monoubiquitin (Ub), linear diubiquitin (Ub_2), or K63-linked diubiquitin (K63-Ub_2) were added to fluorescent NEIL3-NZF and bound species were monitored via fluorescence anisotropy. B) SDS-page gel of purified monoubiquitin, linear diubiquitin and K63-linked diubiquitin (K63-Ub_2 purified by Carl Schiltz). C) Preliminary fluorescent anisotropy experiments show no strong preference toward one ubiquitin species. D) Total intensity from anisotropy experiment in C.

show that the NEIL3-NZF binds monoubiquitin in a similar configuration as NMR structure of Npl4-NZF bound to monoubiquitin, even though Npl4-NZF has a complete TF/Φ motif, while NEIL3-NZF has a TL/A motif. Regardless, the isoleucine-44 and hydrophobic patch on ubiquitin are present for direct interaction as shown in the determined crystal structure of Npl4-NZF and ubiquitin. The NEIL3-NZF structure is predicted to be nearly identical to

the current NZF domains of Npl4, TAB2, TAB3, HOIL-1L, and TRABID which show a root-mean-square deviation (RMSD) of less than 0.9 Å (Kristariyanto et al., 2015).

However, a structure of the NEIL3-NZF by x-ray crystallography would confirm if the structure is similar to the other NZF domains from proteins listed above. Regardless, there is a possibility that TRAIP ligates non-uniform heterogeneous and branched ubiquitin chains to CMG at a replication fork. Additionally, the identity of the E2-conjugating ubiquitin enzyme is unknown. Wu *et al.* has shown that ubiquitylated CMG is insensitive to DUBs recognizing M1, K6, K11, K27, K29, K33, K48, and K63 ubiquitin linkages in *Xenopus* egg extracts (Wu et al., 2019). Yet contrary results show that Mcm7 is ubiquitylated at K29 (Maric et al., 2017). The NZF1 domain of deubiquitinase TRABID was shown to bind K29-linked heterotypic polyubiquitin and the crystallographic structure shows NZF interaction with proximal and distal ubiquitin moieties and not with the K29-linked isopeptide bond (Kristariyanto et al., 2015). Thus, NEIL3-NZF may recognize a variety of heterogenous branched ubiquitin chains. Future work includes identifying how to synthesize the wide variety of ubiquitin heterogenous chains and test for ubiquitin binding via fluorescence anisotropy.

Long ubiquitin chains on CMG are recognized by the Udf-Npl4 complex which contains an NZF domain within Npl4 and initiates the FA pathway (Mirsanaye et al., 2021). Crystallography is a powerful high-resolution technique to determine the chain length and interaction of ubiquitin to the NEIL3-NZF. I have set-up multiple crystal trials with human monoubiquitin and mouse NEIL3-NZF (with N-terminal HisGST tag and without any tags) in a 1:1 molar ratio with no crystal hits yet. Future studies include increasing the molar ratio of ubiquitin to NZF and replacing the human recombinant NZF protein instead of the

mouse NZF to increase the crystal trial possibilities. Additional work is required to understand the NZF-ubiquitin interaction of NEIL3 at the replication fork in addition to how other NZFs bind ubiquitin to regulate function via biochemical assays and structural biology techniques.

NEIL3 in autoimmunity and neurological function and disease

New roles of NEIL3 in human auto-immunity and reproductive health associated with chlamydia and endometriosis have been investigated (Massaad et al., 2016; Poli-Neto et al., 2021b). NEIL3 shows an emerging role in rare autoimmune diseases, as exemplified by the D132V mutation in the NEIL3-GD (Massaad et al., 2016). The D132V mutation was identified in 3 siblings within a consanguinous family and exhibited respiratory tract infections, anemia, bronchiectasis, bleeding, and chronic diarrhea. Unfortunately, all three patients with the autosomal recessive D132V mutation in NEIL3 died before they were 20 years old and the D132V mutation is present in 2% of all healthy Middle Eastern people (Massaad et al., 2016). To my knowledge, the only study completed with cells from D132V patients showed two-fold telomere loss compared to cells from healthy siblings (Zhou et al., 2017).

Because the D132V mutation associates with a significant phenotype, it is critical to confirm if this mutation affects glycosylase activity or recruitment to the DNA lesion and to examine the structural effect of this mutation. Future investigation could include using site-directed mutagenesis to create the NEIL3-D132V mutation in NEIL3-GD and NEIL3-FL constructs, purifying, and performing a base excision activity assay with DHT lesion. Minimal to no activity compared to wild-type NEIL3-GD and NEIL3-FL would identify the

D132 residue as necessary within the active site in the shortened intercalating loop, below the beta strand with DNA intercalating residue methionine-99. A structure of NEIL3-D132V with DNA would show if valine was unable to reach the DNA lesion, providing a structural reason for a phenotype seen in humans.

NEIL3 and clinical applications

NEIL3 is highly expressed in several cancer types. Survival increases when cancer is identified in its early stages. The stage at which cancer is detected is absolutely essential and there is growing research to identify potential cancer biomarkers. Lung cancer is the leading fatal human cancer type and 45% of lung cancer is the subtype lung adenocarcinoma (LUAD) (Zhao et al., 2021). NEIL3 was identified as a potential biomarker for LUAD as well as 9 genes that are co-expressed with NEIL3. Clinical trials are necessary to confirm NEIL3 as a cancer biomarker by creating a simple diagnostic tool that can be used in a clinical setting. Currently, NEIL3 is confirmed to be present in tissue type by western blot. If a similar mechanism with NEIL3 antibody as the detection mechanism were engineered into a simple diagnostic, then early cancer detection would be possible.

Once cancer has been detected in a patient, the next step is to create a treatment plan. Our determination of the GRF-ZF domain of NEIL3 reveals a new therapeutic target for chemotherapeutics. The GRF-ZF domain is necessary for the recruitment of NEIL3 to the site of DNA damage (Wu et al., 2019). Therefore, if the GRF-ZF domain were blocked or inhibited from binding ssDNA then NEIL3 would not reach the ICL. Many of these DNA-

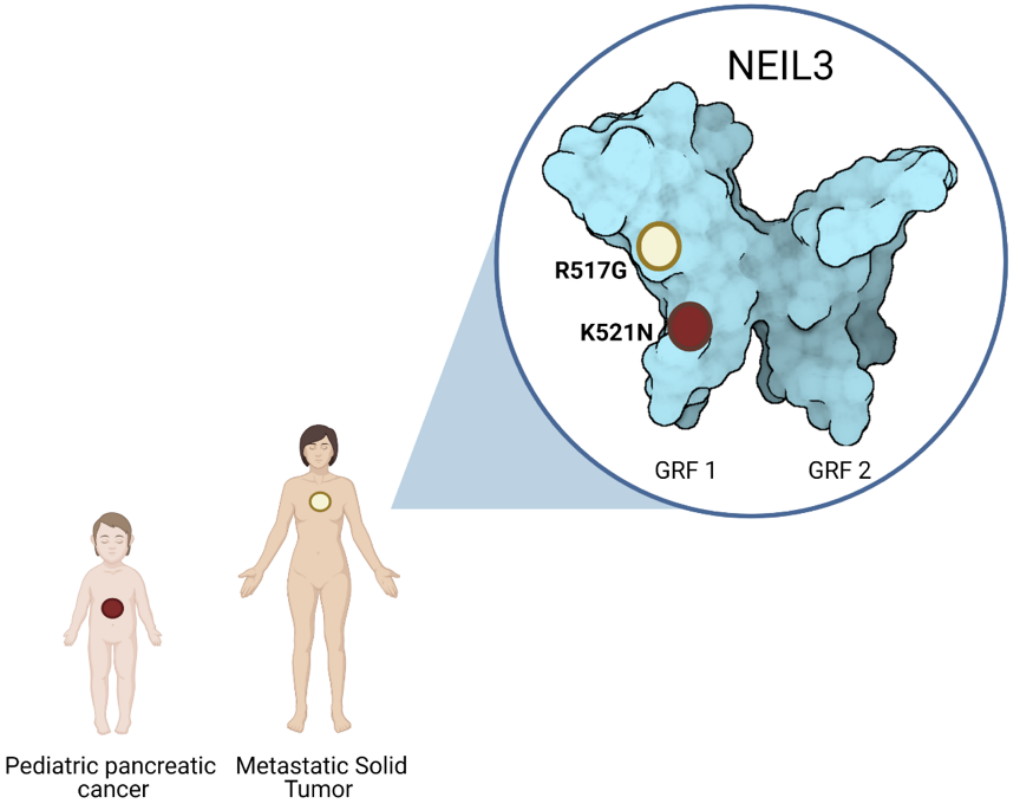

Figure 32. Clinical mutations revealed on NEIL3 GRF-ZF domain. Clinical mutations revealed on NEIL3 GRF-ZF domain. Mutations R517G and K521N are found in the DNA binding cleft of NEIL3 GRF-ZF1 within the GRF-ZF domain (PDB 7JL5) and are attributed to metastatic solid tumor and pediatric pancreatic cancer respectively.

ICLs are intentionally induced in cancer patients under the treatment of chemotherapeutics with the desire to halt replication in cancer cells. However, the chemotherapeutics are unable to specify if NEIL3 is repairing basal DNA damage in all cells, versus clinically induced damage in cancer cells. If there were a cancer cell signature that chemotherapeutics could target, to stay specific to cancer cells and not healthy cells, then the GRF-ZF domain would be an excellent domain to inhibit. Two mutants located on the NEIL3 GRF-ZF domain have shown clinical relevance and are located at R518G and K521N in NEIL3 GRF-ZF1 motif (Figure 32). These mutations have

been linked to metastatic solid tumors and pediatric pancreatic cancer in a cBioPortal analysis queried on March 10, 2021 with 118,149 samples and 114,469 patients from 303 studies (Grobner et al., 2018; Robinson et al., 2017). We have shown that both of these mutations ablate ssDNA binding when charge reversal mutants are introduced (Figure 32). The probability is high that R518G and K521N have a decreased binding affinity toward ssDNA compared to wild-type, yet EMSAs and fluorescence anisotropy can confirm these results. By inhibiting the GRF-ZF domain clinically, NEIL3 would not be recruited to DNA ICL damage induced by chemotherapeutics and cancer cells replication would be halted.

If the R517G and K521N mutants are shown to aid in cancer progression another option to decrease cancer progression may include the addition of a GRF-ZF inhibitor. The inhibitor would be small since we show that the GRF-ZF domain accommodates 5 nucleotides of ssDNA across its front cleft. Future inhibitor libraries would need to be screened against the NEIL3 GRF-ZF to identify an inhibitor with high affinity, potency, and purity.

Concluding remarks

NEIL3 is a unique DNA glycosylase that has shown regulation of human disease, from cancers, auto-immunity, and metabolic processes. This dissertation has revealed the substrate specificity of the NEIL3-GD toward model replication fork DNA substrates containing a small oxidized lesion or ICL on the leading template strand. Additionally, I have investigated the unknown NEIL3 C-terminus and biophysically identified a positively charged ssDNA binding cleft and determined the first structure of the GRF-ZF domain

showing a claw-like structure with the flexibility to bind multiple ssDNA conformation. I also identified the first autoinhibitory role of the GRF-ZF domain that may aid in regulation of glycosylase activity at a converged replication fork containing an ICL. Lastly, I have begun preliminary studies on the internal NZF domain of NEIL3 and shown initial binding to ubiquitin species to address NEIL3-NZF ubiquitin stoichiometry and specificity. My work has focused on the molecular details of NEIL3 while showing awareness of how these enzymatic mechanisms relate to published disease phenotypes. Biochemical, biophysical, and structural information have guided basic science mechanistic inquiries to all three domains in NEIL3, ultimately addressing personalized medicine applications. However, this work is only the steppingstone to investigating these molecular mechanisms of this nuclear protein involved with DNA repair and replication.

Future topics of investigation include the kinetic mechanisms and sequence of events at the replication fork. Does NEIL3 travel with the replication fork like TRAIP? What is the sequence of events after a replication fork pauses at the ICL? Does TRAIP ubiquitylation recruit NEIL3 right away or is there lag time? How does TRAIP know how long to make the ubiquitin chain on replicative CMG helicase? Are the ICL repair pathways initiated simultaneously with TRAIP ubiquitination? In terms of NEIL3 cross-talk and regulation with other proteins, what are the unidentified binding partners also expressed in S-phase? There has been little investigation into the post-replicative mechanism after ICL repair is complete. How do the NEIL3 domains (NEIL3-GD, NEIL3-NZF, and NEIL3 GRF-ZF domain) dissociate from the ICL after unhooking? I look forward to the dynamic future studies of NEIL3 and how it interacts with other proteins at the replication fork to efficiently repair diverse DNA lesions within the genome.

DNA Substrates

Table 2. DNA substrate composition from DNA binding assays and glycosylase assays.

Substrate	Length (nt)	Sequence (5′-3′)
EMSA and fluorescence anisotropy		
ssDNA	10	FAM-CTCAGGACTC
ssDNA	20	FAM-CTCAGGACTCAGTTCGTCAG
ssDNA	30	FAM-CTCAGGACTCAGTTCGTCAGCCCTTGACAG
ssDNA	40	FAM-CTCAGGACTCAGTTCGTCAGCCCTTGACAGCGATGGAAGC
dsDNA	40	1: FAM-CTCAGGACTCAGTTCGTCAGCCCTTGACAGCGATGGAAGC 2: GCTTCCATCGCTGTCAAGGGCTGACGAACTGAGTCCTGAG
Fork	40	1: FAM-CTCAGGACTCAGTTCGTCAGCCCTTGACAGCGATGGAAGC 2: CGAAGGTAGCGACAGTTCCCCTGACGAACTGAGTCCTGAG
Glycosylase assays		
dsDNA	39	1: P32-CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: AACCAGACTGCAGTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32-CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCGTTCCGGTAGCGGC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32-ATGGTCTGACGTGAATC**(AP)**TGCAACTCGCCACCAACTTCC 2: GGAAGTTGGTGGCGAGTTGC**A**AGTGGGGATGGCCTTCCC
Fork	39	1: P32-CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGTAGCGC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32-CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT

		2: CCCTTCCGGTAGCTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32- CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGTAGGTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32- CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGTAAGTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32- CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGTCAGTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: P32- CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGGCAGTTC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork Lead AP-ICL fork	39	1: FAM- CCTTCAACCACCGCTCAACTC**(AP)**TGAACTGCAGTCTGGTT 2: CCCTTCCGGTAGCGGC**A**AGAGTTGAGCGGTGGTTGAAGG
Fork	39	1: ATGGTCTGACGTGAATC**(AP)**TGCAACTCGCCACCAACTTCC-FAM 2: GGAAGTTGGTGGCGAGTTGC**A**AGTGGGGATGGCCTTCCC
ssDNA	25	FAM-GACCACTACACC**(DHT)**ATTCCTTACAAC
dsDNA	25	1: FAM-GACCACTACACC**(DHT)**ATTCCTTACAAC 2: GTTGTAAGGAATAGGTGTAGTGGTC
Lead DHT fork	25	1: FAM-GACCACTACACC**(DHT)**ATTCCTTACAAC 2: CAACATTCCTTA**A**GGTGTAGTGGTC
Lag DHT fork	25	1: FAM-GACCACTACACC**(DHT)**ATTCCTTACAAC 2: GTTGTAAGGAAT**A**CCACATCACCAG
Lag AP-ICL fork	39	1: FAM- ATGGTCTGACGTGAATC**(AP)**TGCAACTCGCCACCAACTTCC 2: GGAAGTTGGTGGCGAGTTGCA**A**GTGGGGATGGCCTTCCC
FAM, 6-carboxyfluorescein, P32, radiolabel. Non-base-paired regions are shown in red. Damaged base pairs or ICL nucleobases are bolded and underlined.		

Reagents

Reagents were purchased from the following suppliers and were of the highest purity available. Oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA) or, in the case of DHT-containing oligonucleotides, from Midland Certified Reagent Co. (Midland, TX). Uracil DNA glycosylase (UDG), and T4 DNA polynucleotide kinase (T4 PNK), formamidopyrimidine DNA glycosylase (Fpg) and endonuclease III (Nth) were from New England Biolabs (Ipswich, MA). [γ-^{32}P]-ATP (6000 Ci/mmol) was purchased from PerkinElmer. C-18 Sep-Pak cartridges were purchased from Waters (Milford, MA), and BS Poly-prep columns were obtained from BioRad (Hercules, CA). Acrylamide/bis-acrylamide 19:1 (40% Solution/Electrophoresis) was purchased from Fisher Scientific (Waltham, MA), and other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Quantification of radioactivity in polyacrylamide gels was carried out using a Personal Molecular Imager (BIORAD) with Quantity One software (v.4.6.5). Monoubiquitin for ubiquitin-binding assays was purchased from Boston Biochem (catalog U-100H).

Expression and Purifications

NEIL3FL

The mNEIL3 gene in pET30a was provided by S. Doublié (Univ. of Vermont, USA) (Liu et al., 2012). The C-terminal His_6 tag was replaced with a FLAG tag by Q5 site directed mutagenesis kit. Protein was expressed in *E. coli* BL21 RIL cells in LB media supplemented with 100 μM $ZnSO_4$ by induction with 250 μM IPTG overnight at 16°C.

Cells were lysed in 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 15% glycerol, 0.1% NP-40, 1 mM PMSF, 1 mM leupeptin, 5 mM β-mercaptoethanol (βME). Lysate was centrifuged at 20,000 x g for 30 min and supernatant collected. Anti-FLAG beads were incubated with the supernatant for 2-4 hr at 4°C. Beads were isolated by centrifugation and washed three times with 20 mM Tris-HCl pH 8.0, 600 mM NaCl, 15% glycerol, 0.05 mM TCEP, and 0.1% NP-40. Protein was eluted with 134 μM FLAG peptide in 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 15% glycerol, 0.05 mM TCEP, and 0.1% NP-40, for 15 min at 4°C.

NEIL3-GD

Nucleotides encoding mNEIL3-GD (residues 1-282) and containing a C-terminal His_6 tag was expressed from pET30a (Liu et al., 2012) in *E. coli* BL21(DE3) Star cells by auto-induction as previously described (Liu et al., 2012). Cells were lysed in Buffer A (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% glycerol, 5 mM βME) containing 20 mM imidazole, 1 mM leupeptin, and 1 mM pepstatin A. Lysate was incubated with Ni-NTA (Thermo Scientific) beads for 30 min at 4°C. Protein was eluted using a 20-500 mM imidazole gradient in Buffer A. Fractions were pooled, diluted to 75 mM NaCl in Buffer B (40 mM HEPES-NaOH pH 7.0, 10% glycerol, and 5 mM βME), and dialyzed at 4°C overnight. Protein was loaded onto a heparin-sepharose (GE Healthcare) column, washed and eluted with a NaCl gradient (0-1 M) in Buffer B. Fractions were pooled, diluted to 500 mM NaCl in Buffer B and loaded onto a HiLoad 26/600 Superdex 200 size exclusion column (GE Healthcare). Protein was eluted in 20 mM HEPES-NaOH pH 7.0, 100 mM NaCl, 1 mM TCEP, concentrated, and frozen.

Mouse NEIL3 GRF-ZF and NZF domains

Nucleotides encoding mNEIL3 GRF1 (residues 506-549), GRF2 (residues 550-595), tandem GRF12 (residues 506-595), and NZF (residues 319-353) proteins were cloned into pBG101 (Vanderbilt Center for Structural Biology), which produces an N-terminal His_6-GST fusion protein that can be cleaved by Rhinovirus 3C protease. Proteins were expressed in *E. coli* BL21 RIL cells in LB media containing 10 μM $ZnSO_4$ by induction with 250 mM IPTG overnight at 16°C. Cells were lysed in buffer containing 20 mM Tris-HCl (pH 8.0 for GRF1, pH 7.0 for GRF2, or pH 7.5 for GRF12 and NZF), 500 mM NaCl, 15% glycerol, 0.5 mM TCEP, 0.02% NP-40, 1 μM leupeptin, and 1 μM pepstatin A. Lysate was incubated with Ni-NTA (Thermo Scientific) beads for 30 min at 4°C. Beads were washed with 20 mM imidazole and protein eluted with 500 mM imidazole in Buffer C (20 mM Tris-HCl, 500 mM NaCl, 15% glycerol, 0.5 mM TCEP) supplemented with 0.02% NP-40. Imidazole was diluted to less than 200 mM and protein was incubated overnight at 4°C with glutathione (GSH)-sepharose resin (Thermo Scientific) in Buffer C/0.02% NP-40. GSH beads were washed with Buffer C/0.02% NP-40 and protein eluted with 10 mM reduced GSH in Buffer C/0.01% NP-40. Protein fractions were pooled and concentrated. At this stage the purified His_6-GST fusions were either stored for use in EMSAs or dialyzed into Buffer C and cleaved by 3C protease. The His_6-GST tag was removed with Ni-NTA in Buffer C containing 15-20 mM imidazole. GRF protein was concentrated and stored in Buffer C. GRF1-R518E, GRF1-K522E, GRF2-K568E, and GRF12-K522E-K568E (GRF^{mut}) point mutants were generated using a QuikChange (Agilent) mutagenesis kit and purified the same as the wild-type proteins. For

crystallography trials of mouse NEIL-NZF, protein was concentrated by centrifugation up to 6 mg/mL in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, and 0.5 mM TCEP.

Human NEIL3-NZF domain

Human NEIL3-NZF was mutated in plasmid from above and C-terminal FlAsH tag (CCPGCC) was added using Q5 ® site directed mutagenesis kit (NEB). Expression protocol for human NZF is the same as stated above with the following purification amendments. Lysis buffer is the same as stated above with the addition of 4 µg/mL DNase 1 and 1 mM magnesium chloride hexahydrate to decrease lysate viscosity and cell clumping. Protein was eluted from GSH column with BufferC/0.01% NP-40 and fresh 40 mM reduced glutathione-S-transferase (GST). Dialysis buffer after GSH column included buffer C with 300 mM NaCl. If desired, His_6-GST tag cleavage was followed in the protocol above.

Human GRF-ZF domain

Human NEIL3 GRF12 for crystallization (residues 501-605) was cloned into pMCSG9 by LIC cloning to produce a His_6-MBP-TEV-GRF construct. For expression, plasmid was transformed into BL21AI cells and grown in terrific broth to an OD_{600} of 0.8. Cells were induced with 20% L-arabinose, 50 mM $ZnSO_4$, and 100 mM IPTG and grown overnight at 16°C. For purification, cells were resuspended in 50 mM Tris pH 8.0, 500 mM NaCl, 1 mM TCEP, 0.1 g lysozyme, and protease inhibitor tablet (Roche). Cells were lysed and centrifuged. Lysate was added to amylose resin, flow through collected, and

resin washed with above buffer followed by 50 mM Tris pH 8.0, 1 M NaCl, and 1 mM TCEP. Salt was dropped to 150 mM NaCl and protein eluted with 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM TCEP, and 10 mM maltose. Elutions were collected, incubated with TEV protease, and subjected to heparin-sepharose chromatography with a 0.05-1.0 M NaCl gradient in 20 mM Tris pH 8 and 1 mM TCEP. Fractions were pooled, run over a HiLoad 16/600 Superdex 75 (GE Healthcare) size exclusion column in 10 mM Tris pH 8, 100 mM NaCl, and 1 mM TCEP, and protein concentrated to 10 mg/mL.

Linear diubiquitin

The linear ubiquitin gene in pET26b was provided by Shuya Fukai (The University of Tokyo, Japan) (Sato et al., 2009 EMBO). The protein was expressed into *E. coli* BL21 (DE3) cells in LB media with kanamycin. 1L cultures were grown for 3-4 hours at 37°C to an OD_{600} of 0.6. Cultures were then put on ice for 1-2 hours and induced with 0.5 mM IPTG and left to grow overnight at 16°C. Cells were harvested and pelleted for purification. Cells were lysed in lysis buffer (20 mM ammonium acetate pH 5.1 and 1 mM PMSF) and sonicated for 5 minutes total process time, 5 seconds on and 10 seconds off. Lysate was centrifuged at 20,000 RPM for 20 minutes. After centrifugation, lysate was filtered through 0.45 μm filter and pH was decreased to 4.8 with acetic acid. Solution was centrifuged again for 15 minutes at 20,000 RPM and filtered again. Ion exchange chromatography was completed with S-column (GE Healthcare 5 mL High Trap SP HP). Column was equilibrated in lysis buffer, protein was loaded, and protein was eluted with elution buffer (20 mM ammonium acetate pH 5.1, 1M NaCl). Fractions containing protein were pooled, concentrated and loaded onto size exclusion column (GE Healthcare Superdex 75

10/300) in 50 mM Tris-HCl pH 8.3 and 150 mM NaCl. Fractions containing protein were pooled, concentrated, and flash frozen for storage in 50 mM Tris-HCl pH 8.3 and 150 mM NaCl, 20% glycerol.

K63-linked diubiquitin

Published protocol was followed to form K63-linked diubiquitin (Pickart and Raasi, 2005). Briefly, ubiquitin mutants, Ub-D77 and Ub-K63R were inserted into pET3a vector with N-terminal His_6 tag with Rhinovirus 3C protease cleavage site with ampicillin resistance gene. Ubiquitin mutants were expressed in BL21(DE3) *E.coli* cells by growing at 37°C until OD_{600nm} reached 0.6. Cultures were cooled and 500 mM IPTG final concentration added to culture. Cultures were grown overnight at 16°C and pelleted the next day. Cells were lysed in lysis buffer (20 mM HEPES-NaOH pH 8, 150 mM NaCl, and 1 mM PMSF). Cells were sonicated and centrifuged to collect soluble fraction. Lysate was added to NiNTA beads and rocked at 4°C overnight. Column was washed with wash buffer (20 mM HEPES NaOH pH 8, 150 mM NaCl, 20 mM imidazole) and Ub-mutants eluted with elution buffer (20 mM HEPES-NaOH pH 8, 150 mM NaCl, 500 mM imidazole). Fractions with ubiquitin were pooled and added back to nickel column equilibrated with lysis buffer. Rhinovirus 3C protease was added to column and left overnight at 4°C. The following day the flow through was collected and contained cleaved Ub-D77 or Ub-K63R.

K63-diubiquitin reaction mixture included incubation of 50 μM total of Ub-D77 and Ub-K63R, 2 μM of E2-conjugating enzyme Ubc13/Mms2, and 0.5 μM E1-activating ubiquitin enzyme Uba1 in reaction buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP) at 37°C overnight. Reaction mixture was loaded onto Mono S 10/100

GL column (Cytiva) and ubiquitin chain was separated by cation exchange chromatography in reaction buffer A at pH 4.5, absent in $MgCl_2$ and ATP. Column was eluted with reaction buffer A with gradient up to 500 mM NaCl. Fractions containing ubiquitin mutants were pooled and separated by size exclusion chromatography on S75 10/300 GL column (GE Healthcare) in 20 mM Tris-HCl pH 8 and 150 mM NaCl. Fractions were pooled, concentrated by centrifugation, and flash frozen. Concentration was calculated on NanoDrop via A_{205nm} absorbance.

Synthesis of dA-AP ICLs

A single-stranded, uracil-containing 2'-deoxy-oligonucleotide was 5'-^{32}P-labeled using standard procedures(Sambrook et al., 1989), annealed to its complementary strand, and the resulting duplex treated with the enzyme UDG (50 units/mL, final concentration) to generate an AP-containing duplex. The UDG enzyme was removed by phenol-chloroform extraction and the DNA ethanol precipitated. For ICL generation, AP-containing duplexes were incubated in 50 mM HEPES pH 7 and 100 NaCl at 37 °C for 120 h. The DNA was ethanol precipitated, resuspended in formamide loading buffer and the slowly migrating cross-linked material was separated from uncross-linked material by electrophoresis for 10 h at 200 V on a 2 mm thick, 20% denaturing polyacrylamide gel. The splayed ICLs used in these studies were generated in approximately 30-50% yields. The DNA bands in the gel were visualized by autoradiography and the slow-moving band corresponding to cross-linked duplex was excised from the gel. The gel slice was crushed and agitated in an elution buffer composed of aqueous 200 mM NaCl and 1 mM EDTA pH 8 at 24 °C for 1 h. The gel fragments were removed by filtering through a Poly-Prep

column and the filtrate was ethanol precipitated. The resulting residue was briefly dried in a Speed-Vac concentrator and stored at -20°C. ICLs labeled with 5(6)-carboxyfluorescein (FAM) were prepared in a similar manner using a FAM-labeled dU-containing strand prepared by treatment of 5' or 3'-amino modified oligonucleotides (/5AmMC6/ or /3AmMO/ from IDT) with 5(6)-carboxyfluorescein *N*-hydroxysuccinimidyl ester, (Giusti and Adriano, 1993; Imani Nejad et al., 2020) followed by gel purification. Sequences of all DNA substrates used in these studies are shown in Table 2.

ICL Unhooking Assays and Base Excision Assays

DNA substrates containing an AP-ICL or DHT were prepared as previously described (Imani Nejad et al., 2020). Both AP-ICL unhooking and DHT excision assays were performed under single turnover conditions, which was verified by confirming that the reaction rates remained constant at higher enzyme:DNA ratios. AP-ICL unhooking reactions were performed at 25°C and contained 250 nM full-length mNEIL3 or deletion mutants and 25 nM FAM-labeled DNA in glycosylase buffer (20 mM HEPES-NaOH pH 7.0, 100 mM NaCl, 5% glycerol, 1 mM DTT, and 10 µg/mL BSA). Reactions were stopped by the addition of 10 mM EDTA/80% (v/v) formamide at 70°C for 5 min and run on precast 10% TBE/urea gels (Invitrogen) at 180V in 0.5X TBE buffer. Base excision activity of mNEIL3-GD in the presence of purified GRF12 proteins was carried out by mixing 10 µM mNEIL3-GD with 10 µM or either GRF12 or GRF12^{mut}, followed by incubation at 25°C with 50 nM FAM-labeled DHT-containing DNA in glycosylase buffer. Reactions were stopped by addition of 0.1 N NaOH and 10 mM EDTA/80% (v/v) formamide and heated

at 70°C for 5 min. Band intensities were quantified using Gel Analyzer and plotted with one-phase association exponential fit using GraphPad Prism 8.

DNA Binding Assays

Oligonucleotides used in EMSAs are shown in Table 2. His-GST fusions of mNEIL3 GRF and NZF constructs (0-10 μM) were incubated with 10 nM 6-carboxyfluorescein (FAM) labeled DNA for 30 min at 21°C in 20 mM HEPES (pH 6.6), 100 mM NaCl, 5 mM $MgCl_2$, 3% (v/v) glycerol, 0.2% (v/v) NP-40, and 0.5 mM TCEP. Samples were separated by electrophoresis on a 5% polyacrylamide/0.5X TBE gel at 200 V for 1 hr. Gels were imaged using a Typhoon Trio variable mode imager (GE Healthcare) at 532 nm excitation and 526 nm emission wavelengths. Band intensities were quantified using Gel Analyzer and plotted using GraphPad Prism 8.

DNA binding of wild-type and mutant GRF constructs were monitored by fluorescence anisotropy. For K_d determination, proteins at varying concentration were incubated with 25 nM FAM-labeled ssDNA (Table 2) in 20 mM HEPES pH 7.0, 100 mM NaCl, 5 mM $MgCl_2$, 3% glycerol, and 0.5 mM TCEP at 4°C in the dark for 30 min. To determine stoichiometry of binding, 50 nM FAM-labeled and 5 μM unlabeled 40-mer ssDNA was used so that the total DNA concentration was over 20-fold excess of the K_d. Fluorescence anisotropy data at excitation and emission wavelengths of 485 and 528 nm were collected at room temperature in 96-well plates using a BioTek Synergy H1 plate reader. Data were fit to a two-state binding model using GraphPad Prism 8. Ubiquitin binding of NEIL3-NZF and ubiquitin species (monoubiquitin, linear diubiquitin, and K63-linked diubiquitin) were monitored as described above via fluorescence anisotropy.

Ubiquitin species were incubated with 100 nM hNEIL3-NZF with C-terminal FlAsH tag in 20 mM Tris-HCl pH 8, 100 mM NaCl, 0.5 mM TCEP. Similar excitation and emission wavelengths were used as stated above.

X-ray crystallography

Crystals of hNEIL3 GRF12 were obtained by sitting drop vapor diffusion at 20°C by mixing the protein solution with 0.1 M Tris pH 8.5 and 25% PEG 6000. Crystals were cryoprotected in 20% ethylene glycol. X-ray diffraction data from a single crystal were collected at 105 K on beamline 22-ID of the Advanced Photon Source and processed with HKL2000 (Otwinowski and Minor, 1997). Phasing and refinement was carried out using the PHENIX suite of programs (Adams et al., 2010). The structure was determined by SAD phasing from 2.8-Å anomalous data collected at the Zn edge (1.27059 Å) using AutoSol and Phaser. An initial model was built using sidechain-truncated APE2 GRF residues 461-508 from PDB 5U6Z as a guide and the Zn-SAD electron density, followed by manual building of the remainder of the model in Coot (Emsley et al., 2010). The model was refined against 2.6-Å native data collected at 1.0000 Å wavelength. The final model was validated with MolProbity (Davis et al., 2007) and contained no residues in disallowed regions of the Ramachandran plot. Refinement and validation statistics are shown in Table 1. Structures were analyzed and figures made using PyMOL (Schrödinger). All software was curated by SBGrid (Morin et al., 2013). Atomic coordinates and structure factors for the hNEIL3-GRF12 crystal structure have been deposited in the Protein Data Bank under accession code 7JL5.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., and Grosse-Kunstleve, R.W. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213-221.

Adamson, B., Smogorzewska, A., Sigoillot, F.D., King, R.W., and Elledge, S.J. (2012). A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. Nat Cell Biol *14*, 318-328.

Alam, S.L., Sun, J., Payne, M., Welch, B.D., Blake, B.K., Davis, D.R., Meyer, H.H., Emr, S.D., and Sundquist, W.I. (2004). Ubiquitin interactions of NZF zinc fingers. EMBO J *23*, 1411-1421.

Albelazi, M.S., Martin, P.R., Mohammed, S., Mutti, L., Parsons, J.L., and Elder, R.H. (2019). The Biochemical Role of the Human NEIL1 and NEIL3 DNA Glycosylases on Model DNA Replication Forks. Genes (Basel) *10*.

Altieri, F., Grillo, C., Maceroni, M., and Chichiarelli, S. (2008). DNA damage and repair: from molecular mechanisms to health implications. Antioxid Redox Signal *10*, 891-937.

Amidon, K.M., and Eichman, B.F. (2020). Structural biology of DNA abasic site protection by SRAP proteins. DNA Repair (Amst) *94*, 102903.

Amunugama, R., and Walter, J.C. (2020). A new varietal of DNA interstrand crosslink repair. Cell Res *30*, 459-460.

Ashour, M.E., and Mosammaparast, N. (2021). Mechanisms of damage tolerance and

repair during DNA replication. Nucleic Acids Res *49*, 3033-3047.

Auerbach, A.D. (1995). Fanconi anemia. Dermatol Clin *13*, 41-49.

Bailly, V., and Verly, W.G. (1988). Importance of thiols in the repair mechanisms of DNA containing AP (apurinic or apyrimidinic) sites. Nucleic Acids Res *16*, 9489-9496.

Bandaru, V., Sunkara, S., Wallace, S.S., and Bond, J.P. (2002). A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. DNA Repair (Amst) *1*, 517-529.

Bergink, S., and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. Nature *458*, 461-467.

Berneburg, M., and Lehmann, A.R. (2001). Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription. Adv Genet *43*, 71-102.

Berti, P.J., and McCann, J.A. (2006). Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. Chem Rev *106*, 506-555.

Brannvoll, A., Xue, X., Kwon, Y., Kompocholi, S., Simonsen, A.K.W., Viswalingam, K.S., Gonzalez, L., Hickson, I.D., Oestergaard, V.H., Mankouri, H.W.*, et al.* (2020). The ZGRF1 Helicase Promotes Recombinational Repair of Replication-Blocking DNA Damage in Human Cells. Cell Rep *32*, 107849.

Burkovics, P., Szukacsov, V., Unk, I., and Haracska, L. (2006). Human Ape2 protein has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. Nucleic Acids Res *34*, 2508-2515.

Caldecott, K.W. (2020). Mammalian DNA base excision repair: Dancing in the moonlight. DNA Repair (Amst) *93*, 102921.

Callaway, E. (2015). The revolution will not be crystallized: a new method sweeps through structural biology. Nature *525*, 172-174.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: new players and new functions. Nat Rev Mol Cell Biol *17*, 337-349.

Chakraborty, A., Wakamiya, M., Venkova-Canova, T., Pandita, R.K., Aguilera-Aguirre, L., Sarker, A.H., Singh, D.K., Hosoki, K., Wood, T.G., Sharma, G.*, et al.* (2015). Neil2-null Mice Accumulate Oxidized DNA Bases in the Transcriptionally Active Sequences of the Genome and Are Susceptible to Innate Inflammation. J Biol Chem *290*, 24636-24648.

Chen, S.H., Wu, C.H., Plank, J.L., and Hsieh, T.S. (2012). Essential functions of C terminus of Drosophila Topoisomerase IIIalpha in double holliday junction dissolution. J Biol Chem *287*, 19346-19353.

Cimino, G.D., Gamper, H.B., Isaacs, S.T., and Hearst, J.E. (1985). Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. Annu Rev Biochem *54*, 1151-1193.

Conconi, A., and Bell, B. (2017). Molecular biology: The long and short of a DNA-damage response. Nature *545*, 165-166.

Couve, S., Mace-Aime, G., Rosselli, F., and Saparbaev, M.K. (2009). The human oxidative DNA glycosylase NEIL1 excises psoralen-induced interstrand DNA cross-links in a three-stranded DNA structure. J Biol Chem *284*, 11963-11970.

Das, S., Chattopadhyay, R., Bhakat, K.K., Boldogh, I., Kohno, K., Prasad, R., Wilson,

S.H., and Hazra, T.K. (2007). Stimulation of NEIL2-mediated oxidized base excision repair via YB-1 interaction during oxidative stress. J Biol Chem *282*, 28474-28484.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., and Richardson, J.S. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375-W383.

de Sousa, J.F., Torrieri, R., Serafim, R.B., Di Cristofaro, L.F., Escanfella, F.D., Ribeiro, R., Zanette, D.L., Paco-Larson, M.L., da Silva, W.A., Jr., Tirapelli, D.P.*, et al.* (2017). Expression signatures of DNA repair genes correlate with survival prognosis of astrocytoma patients. Tumour Biol *39*, 1010428317694552.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. Nat Rev Cancer *11*, 467-480.

Doublie, S., Bandaru, V., Bond, J.P., and Wallace, S.S. (2004). The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. Proc Natl Acad Sci U S A *101*, 10284-10289.

Ebner, P., Versteeg, G.A., and Ikeda, F. (2017). Ubiquitin enzymes in the regulation of immune responses. Crit Rev Biochem Mol Biol *52*, 425-460.

Eckenroth, B.E., Cao, V.B., Averill, A.M., Dragon, J.A., and Doublie, S. (2021). Unique Structural Features of Mammalian NEIL2 DNA Glycosylase Prime Its Activity for Diverse DNA Substrates and Environments. Structure *29*, 29-42 e24.

Ehlers, C.L., Gizer, I.R., Bizon, C., Slutske, W., Peng, Q., Schork, N.J., and Wilhelmsen,

K.C. (2016). Single nucleotide polymorphisms in the REG-CTNNA2 region of chromosome 2 and NEIL3 associated with impulsivity in a Native American sample. Genes Brain Behav *15*, 568-577.

Emsley, P., Lohkamp, B., Scott, W., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486-501.

Fagbemi, A.F., Orelli, B., and Scharer, O.D. (2011). Regulation of endonuclease activity in human nucleotide excision repair. DNA Repair (Amst) *10*, 722-729.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K., and Bairoch, A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res *30*, 235-238.

Fiesco-Roa, M.O., Giri, N., McReynolds, L.J., Best, A.F., and Alter, B.P. (2019). Genotype-phenotype associations in Fanconi anemia: A literature review. Blood Rev *37*, 100589.

Gates, K.S. (2009). An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. Chem Res Toxicol *22*, 1747-1760.

Giusti, W.G., and Adriano, T. (1993). Synthesis and characterization of 5'-fluorescent-dye-labeled oligonucleotides. PCR Methods Appl *2*, 223-227.

Grawert, T.W., and Svergun, D.I. (2020). Structural Modeling Using Solution Small-Angle X-ray Scattering (SAXS). J Mol Biol *432*, 3078-3092.

Griffin, B.A., Adams, S.R., and Tsien, R.Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. Science *281*, 269-272.

Grobner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva,

V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S.*, et al.* (2018). The landscape of genomic alterations across childhood cancers. Nature *555*, 321-327.

Guainazzi, A., and Scharer, O.D. (2010). Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. Cell Mol Life Sci *67*, 3683-3697.

Ha, A., Lin, Y., and Yan, S. (2020). A non-canonical role for the DNA glycosylase NEIL3 in suppressing APE1 endonuclease-mediated ssDNA damage. J Biol Chem *295*, 14222-14235.

Hashimoto, S., Anai, H., and Hanada, K. (2016). Mechanisms of interstrand DNA crosslink repair and human disorders. Genes Environ *38*, 9.

Hazra, T.K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y.W., Jaruga, P., Dizdaroglu, M., and Mitra, S. (2002). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. Proc Natl Acad Sci U S A *99*, 3523-3528.

He, W., Huang, P., Liu, D., Zhong, L., Yu, R., and Li, J. (2016). Polymorphism of the XRCC1 Gene Is Associated with Susceptibility and Short-Term Recovery of Ischemic Stroke. Int J Environ Res Public Health *13*.

Hegde, M.L., Hegde, P.M., Bellot, L.J., Mandal, S.M., Hazra, T.K., Li, G.M., Boldogh, I., Tomkinson, A.E., and Mitra, S. (2013). Prereplicative repair of oxidized bases in the human genome is mediated by NEIL1 DNA glycosylase together with replication proteins. Proc Natl Acad Sci U S A *110*, E3090-3099.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annu Rev Biochem *67*,

425-479.

Hicke, L. (2001). Protein regulation by monoubiquitin. Nat Rev Mol Cell Biol *2*, 195-201.

Hildrestrand, G.A., Neurauter, C.G., Diep, D.B., Castellanos, C.G., Krauss, S., Bjoras, M., and Luna, L. (2009). Expression patterns of Neil3 during embryonic brain development and neoplasia. BMC Neurosci *10*, 45.

Hildrestrand, G.A., Rolseth, V., Kunath, N., Suganthan, M., Jensen, V., Bugaj, A.M., Fernandez-Berrocal, M.S., Sikko, S.B., Vetlesen, S., Kusnierczyk, A., Olsen, A., Gutzkow, K.B., Rowe, A.D., Wang, W., Moldestad, O., Syrstad, M.D., Slupphaug, G., Eide, L., Klugland, A., Saetrom, P., Luna, L., Ye, J., Scheffler, K., Bjoras, M. (2021). NEIL1 and NEIL2 DNA glycosylases regulate anxiety and learning in a cooperative manner. bioRxiv.

Hodskinson, M.R., Bolner, A., Sato, K., Kamimae-Lanning, A.N., Rooijers, K., Witte, M., Mahesh, M., Silhan, J., Petek, M., Williams, D.M.*, et al.* (2020). Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. Nature *579*, 603-608.

Housh, K., Jha, J.S., Haldar, T., Amin, S.B.M., Islam, T., Wallace, A., Gomina, A., Guo, X., Nel, C., Wyatt, J.W.*, et al.* (2021). Formation and repair of unavoidable, endogenous interstrand cross-links in cellular DNA. DNA Repair (Amst) *98*, 103029.

Imani Nejad, M., Housh, K., Rodriguez, A.A., Haldar, T., Kathe, S., Wallace, S.S., Eichman, B.F., and Gates, K.S. (2020). Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3. DNA Repair (Amst) *86*, 102752.

Jalland, C.M., Scheffler, K., Benestad, S.L., Moldal, T., Ersdal, C., Gunnes, G.,

Suganthan, R., Bjoras, M., and Tranulis, M.A. (2016). Neil3 induced neurogenesis protects against prion disease during the clinical phase. Sci Rep *6*, 37844.

Jiang, Y.L., Kwon, K., and Stivers, J.T. (2001). Turning On uracil-DNA glycosylase using a pyrene nucleotide switch. J Biol Chem *276*, 42347-42354.

Johnson, K.M., Price, N.E., Wang, J., Fekry, M.I., Dutta, S., Seiner, D.R., Wang, Y., and Gates, K.S. (2013). On the formation and properties of interstrand DNA-DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5'-CAp sequences in duplex DNA. J Am Chem Soc *135*, 1015-1025.

Kanayama, A., Seth, R.B., Sun, L., Ea, C.K., Hong, M., Shaito, A., Chiu, Y.H., Deng, L., and Chen, Z.J. (2004). TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains. Mol Cell *15*, 535-548.

Kellner, V., and Luke, B. (2020). Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. EMBO J *39*, e102309.

Kellum, A.H., Jr., Qiu, D.Y., Voehler, M.W., Martin, W., Gates, K.S., and Stone, M.P. (2021). Structure of a Stable Interstrand DNA Cross-Link Involving a beta-N-Glycosyl Linkage Between an N(6)-dA Amino Group and an Abasic Site. Biochemistry *60*, 41-52.

Klattenhoff, A.W., Thakur, M., Chu, C.S., Ray, D., Habib, S.L., and Kidane, D. (2017). Loss of NEIL3 DNA glycosylase markedly increases replication associated double strand breaks and enhances sensitivity to ATR inhibitor in glioblastoma cells. Oncotarget *8*, 112942-112958.

Klug, A., and Schwabe, J.W. (1995). Protein motifs 5. Zinc fingers. FASEB J *9*, 597-604.

Komander, D., Reyes-Turcu, F., Licchesi, J.D., Odenwaelder, P., Wilkinson, K.D., and

Barford, D. (2009). Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. EMBO Rep *10*, 466-473.

Kristariyanto, Y.A., Abdul Rehman, S.A., Campbell, D.G., Morrice, N.A., Johnson, C., Toth, R., and Kulathu, Y. (2015). K29-selective ubiquitin binding domain reveals structural basis of specificity and heterotypic nature of k29 polyubiquitin. Mol Cell *58*, 83-94.

Krokan, H.E., and Bjoras, M. (2013). Base excision repair. Cold Spring Harb Perspect Biol *5*, a012583.

Krokeide, S.Z., Laerdahl, J.K., Salah, M., Luna, L., Cederkvist, F.H., Fleming, A.M., Burrows, C.J., Dalhus, B., and Bjoras, M. (2013). Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. DNA Repair (Amst) *12*, 1159-1164.

Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. Science *343*, 1443-1444.

Kulathu, Y., Akutsu, M., Bremm, A., Hofmann, K., and Komander, D. (2009). Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. Nat Struct Mol Biol *16*, 1328-1330.

Kunath, N., Bugai, A.M., Bigonah, P., Fernandez-Berrocal, M.S., Bjoras, M., Ye, J. (2021). DNA repair enzyme NEIL3 enables a stable neural representation of space by shaping transcription in hippocampal neurons. bioRxiv.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.

Li, N., Wang, J., Wallace, S.S., Chen, J., Zhou, J., and D'Andrea, A.D. (2020).

Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in interstrand crosslink repair. Nucleic Acids Res *48*, 3014-3028.

Li, P., Stetler, R.A., Leak, R.K., Shi, Y., Li, Y., Yu, W., Bennett, M.V.L., and Chen, J. (2018). Oxidative stress and DNA damage after cerebral ischemia: Potential therapeutic targets to repair the genome and improve stroke recovery. Neuropharmacology *134*, 208-217.

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610-3618.

Liu, M., Bandaru, V., Bond, J.P., Jaruga, P., Zhao, X., Christov, P.P., Burrows, C.J., Rizzo, C.J., Dizdaroglu, M., and Wallace, S.S. (2010). The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. Proc Natl Acad Sci U S A *107*, 4925-4930.

Liu, M., Bandaru, V., Holmes, A., Averill, A.M., Cannan, W., and Wallace, S.S. (2012). Expression and purification of active mouse and human NEIL3 proteins. Protein Expr Purif *84*, 130-139.

Liu, M., Doublie, S., and Wallace, S.S. (2013a). Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. Mutat Res *743-744*, 4-11.

Liu, M., Imamura, K., Averill, A.M., Wallace, S.S., and Doublie, S. (2013b). Structural characterization of a mouse ortholog of human NEIL3 with a marked preference for single-stranded DNA. Structure *21*, 247-256.

Maric, M., Mukherjee, P., Tatham, M.H., Hay, R., and Labib, K. (2017). Ufd1-Npl4 Recruit Cdc48 for Disassembly of Ubiquitylated CMG Helicase at the End of Chromosome Replication. Cell Rep *18*, 3033-3042.

Massaad, M.J., Zhou, J., Tsuchimoto, D., Chou, J., Jabara, H., Janssen, E., Glauzy, S., Olson, B.G., Morbach, H., Ohsumi, T.K.*, et al.* (2016). Deficiency of base excision repair enzyme NEIL3 drives increased predisposition to autoimmunity. J Clin Invest *126*, 4219-4236.

McNeill, D.R., Paramasivam, M., Baldwin, J., Huang, J., Vyjayanti, V.N., Seidman, M.M., and Wilson, D.M., 3rd (2013). NEIL1 responds and binds to psoralen-induced DNA interstrand crosslinks. J Biol Chem *288*, 12426-12436.

Meyer, H.H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. EMBO J *21*, 5645-5652.

Minko, I.G., Christov, P.P., Li, L., Stone, M.P., McCullough, A.K., and Lloyd, R.S. (2019). Processing of N(5)-substituted formamidopyrimidine DNA adducts by DNA glycosylases NEIL1 and NEIL3. DNA Repair (Amst) *73*, 49-54.

Mirsanaye, A.S., Typas, D., and Mailand, N. (2021). Ubiquitylation at Stressed Replication Forks: Mechanisms and Functions. Trends Cell Biol.

Mohapatra, S., Lin, C.T., Feng, X.A., Basu, A., and Ha, T. (2020). Single-Molecule Analysis and Engineering of DNA Motors. Chem Rev *120*, 36-78.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. Elife *2*, e01456.

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjoras, M., and Seeberg, E. (2002). Human

DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. Nucleic Acids Res *30*, 4926-4936.

Mullins, E.A., Rodriguez, A.A., Bradley, N.P., and Eichman, B.F. (2019). Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway. Trends Biochem Sci *44*, 765-781.

Mullins, E.A., Shi, R., and Eichman, B.F. (2017). Toxicity and repair of DNA adducts produced by the natural product yatakemycin. Nat Chem Biol *13*, 1002-1008.

Mullins, E.A., Shi, R., Parsons, Z.D., Yuen, P.K., David, S.S., Igarashi, Y., and Eichman, B.F. (2015). The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. Nature *527*, 254-258.

Nakielny, S., Shaikh, S., Burke, B., and Dreyfuss, G. (1999). Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain. EMBO J *18*, 1982-1995.

Nejad, M.I., Price, N.E., Haldar, T., Lewis, C., Wang, Y., and Gates, K.S. (2019). Interstrand DNA Cross-Links Derived from Reaction of a 2-Aminopurine Residue with an Abasic Site. ACS Chem Biol *14*, 1481-1489.

Oh, E., Akopian, D., and Rape, M. (2018). Principles of Ubiquitin-Dependent Signaling. Annu Rev Cell Dev Biol *34*, 137-162.

Olsen, M.B., Hildrestrand, G.A., Scheffler, K., Vinge, L.E., Alfsnes, K., Palibrk, V., Wang, J., Neurauter, C.G., Luna, L., Johansen, J.*, et al.* (2017). NEIL3-Dependent Regulation of Cardiac Fibroblast Proliferation Prevents Myocardial Rupture. Cell Rep *18*, 82-92.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in

oscillation mode. Methods Enzymol. *276*, 307-326.

Parsons, J.L., Kavli, B., Slupphaug, G., and Dianov, G.L. (2007). NEIL1 is the major DNA glycosylase that processes 5-hydroxyuracil in the proximity of a DNA single-strand break. Biochemistry *46*, 4158-4163.

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annu Rev Biochem *70*, 503-533.

Pickart, C.M., and Raasi, S. (2005). Controlled synthesis of polyubiquitin chains. Methods Enzymol *399*, 21-36.

Poli-Neto, O.B., Carlos, D., Favaretto, A.J., Rosa, E.S.J.C., Meola, J., and Tiezzi, D. (2021a). Eutopic endometrium from women with endometriosis and chlamydial endometritis share immunological cell types and DNA repair imbalance: A transcriptome meta-analytical perspective. Journal of Reproductive Immunology *145*, 103307.

Poli-Neto, O.B., Carlos, D., Favaretto, A.J., Rosa, E.S.J.C., Meola, J., and Tiezzi, D. (2021b). Eutopic endometrium from women with endometriosis and chlamydial endometritis share immunological cell types and DNA repair imbalance: A transcriptome meta-analytical perspective. J Reprod Immunol *145*, 103307.

Pomorski, A., and Krezel, A. (2020). Biarsenical fluorescent probes for multifunctional site-specific modification of proteins applicable in life sciences: an overview and future outlook. Metallomics *12*, 1179-1207.

Powers, K.T., and Washington, M.T. (2018). Eukaryotic translesion synthesis: Choosing the right tool for the job. DNA Repair (Amst) *71*, 127-134.

Prado, F. (2018). Homologous Recombination: To Fork and Beyond. Genes (Basel) *9*.

Price, N.E., Catalano, M.J., Liu, S., Wang, Y., and Gates, K.S. (2015). Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA. Nucleic Acids Res *43*, 3434-3441.

Price, N.E., Johnson, K.M., Wang, J., Fekry, M.I., Wang, Y., and Gates, K.S. (2014). Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. J Am Chem Soc *136*, 3483-3490.

Quiles-Jimenez, A., Gregersen, I., Segers, F.M., Skarpengland, T., Kroustallaki, P., Yang, K., Kong, X.Y., Lauritzen, K.H., Olsen, M.B., Karlsen, T.R.*, et al.* (2021). DNA glycosylase Neil3 regulates vascular smooth muscle cell biology during atherosclerosis development. Atherosclerosis *324*, 123-132.

Ramstein, J., and Lavery, R. (1988). Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci U S A *85*, 7231-7235.

Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Scharer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. Cell *134*, 969-980.

Roberts, R.J., and Cheng, X. (1998). Base flipping. Annu Rev Biochem *67*, 181-198.

Robinson, D.R., Wu, Y.M., Lonigro, R.J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V.*, et al.* (2017). Integrative clinical genomics of metastatic cancer. Nature *548*, 297-303.

Rodriguez, A., and D'Andrea, A. (2017). Fanconi anemia pathway. Curr Biol *27*, R986-R988.

Rodriguez, A.A., Wojtaszek, J.L., Greer, B.H., Haldar, T., Gates, K.S., Williams, R.S., and

Eichman, B.F. (2020). An autoinhibitory role for the GRF zinc finger domain of DNA glycosylase NEIL3. J Biol Chem *295*, 15566-15575.

Rognlien, A.G., Wollen, E.J., Atneosen-Asegg, M., and Saugstad, O.D. (2015). Increased expression of inflammatory genes in the neonatal mouse brain after hyperoxic reoxygenation. Pediatr Res *77*, 326-333.

Rolseth, V., Runden-Pran, E., Luna, L., McMurray, C., Bjoras, M., and Ottersen, O.P. (2008). Widespread distribution of DNA glycosylases removing oxidative DNA lesions in human and rodent brains. DNA Repair (Amst) *7*, 1578-1588.

Rosenquist, T.A., Zaika, E., Fernandes, A.S., Zharkov, D.O., Miller, H., and Grollman, A.P. (2003). The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. DNA Repair (Amst) *2*, 581-591.

Rumsey, W.L., Bolognese, B., Davis, A.B., Flamberg, P.L., Foley, J.P., Katchur, S.R., Kotzer, C.J., Osborn, R.R., and Podolin, P.L. (2017). Effects of airborne toxicants on pulmonary function and mitochondrial DNA damage in rodent lungs. Mutagenesis *32*, 343-353.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Sancar, A., and Sancar, G.B. (1988). DNA repair enzymes. Annu Rev Biochem *57*, 29-67.

Sarker, A.H., Cooper, P.K., and Hazra, T.K. (2021). DNA glycosylase NEIL2 functions in multiple cellular processes. Prog Biophys Mol Biol.

Sato, Y., Fujita, H., Yoshikawa, A., Yamashita, M., Yamagata, A., Kaiser, S.E., Iwai, K.,

and Fukai, S. (2011). Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. Proc Natl Acad Sci U S A *108*, 20520-20525.

Sato, Y., Yoshikawa, A., Yamashita, M., Yamagata, A., and Fukai, S. (2009). Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by NZF domains of TAB2 and TAB3. EMBO J *28*, 3903-3909.

Semlow, D.R., and Walter, J.C. (2021). Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. Annu Rev Biochem *90*.

Semlow, D.R., Zhang, J., Budzowska, M., Drohat, A.C., and Walter, J.C. (2016). Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. Cell *167*, 498-511 e414.

Shimamura, A., and Alter, B.P. (2010). Pathophysiology and management of inherited bone marrow failure syndromes. Blood Rev *24*, 101-122.

Shinmura, K., Kato, H., Kawanishi, Y., Igarashi, H., Goto, M., Tao, H., Inoue, Y., Nakamura, S., Misawa, K., Mineta, H.*, et al.* (2016). Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer. Oxid Med Cell Longev *2016*, 1546392.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res *18*, 134-147.

Skarpengland, T., Laugsand, L.E., Janszky, I., Luna, L., Halvorsen, B., Platou, C.G., Wang, W., Vatten, L.J., Damas, J.K., Aukrust, P.*, et al.* (2015). Genetic variants in the DNA repair gene NEIL3 and the risk of myocardial infarction in a nested case-control study. The HUNT Study. DNA Repair (Amst) *28*, 21-27.

Smogorzewska, A., Desetty, R., Saito, T.T., Schlabach, M., Lach, F.P., Sowa, M.E., Clark, A.B., Kunkel, T.A., Harper, J.W., Colaiacovo, M.P.*, et al.* (2010). A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. Mol Cell *39*, 36-47.

Stivers, J.T. (2004). Site-specific DNA damage recognition by enzyme-induced base flipping. Prog Nucleic Acid Res Mol Biol *77*, 37-65.

Stratigopoulou, M., van Dam, T.P., and Guikema, J.E.J. (2020). Base Excision Repair in the Immune System: Small DNA Lesions With Big Consequences. Front Immunol *11*, 1084.

Sugitani, N., Sivley, R.M., Perry, K.E., Capra, J.A., and Chazin, W.J. (2016). XPA: A key scaffold for human nucleotide excision repair. DNA Repair (Amst) *44*, 123-135.

Sun, L., and Chen, Z.J. (2004). The novel functions of ubiquitination in signaling. Curr Opin Cell Biol *16*, 119-126.

Takao, M., Kanno, S., Kobayashi, K., Zhang, Q.M., Yonei, S., van der Horst, G.T., and Yasui, A. (2002). A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. J Biol Chem *277*, 42205-42213.

Takao, M., Oohata, Y., Kitadokoro, K., Kobayashi, K., Iwai, S., Yasui, A., Yonei, S., and Zhang, Q.M. (2009). Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in Escherichia coli mutant. Genes Cells *14*, 261-270.

Tan, K., Zhou, Q., Cheng, B., Zhang, Z., Joachimiak, A., and Tse-Dinh, Y.C. (2015). Structural basis for suppression of hypernegative DNA supercoiling by E. coli topoisomerase I. Nucleic Acids Res *43*, 11031-11046.

Tangye, S.G. (2016). Genetic cause of immune dysregulation - one gene or two? J Clin Invest *126*, 4065-4067.

Tapryal, N., Shahabi, S., Chakraborty, A., Hosoki, K., Wakamiya, M., Sarkar, G., Sharma, G., Cardenas, V.J., Boldogh, I., Sur, S.*, et al.* (2021). Intrapulmonary administration of purified NEIL2 abrogates NF-kappaB-mediated inflammation. J Biol Chem, 100723.

Tell, G., Quadrifoglio, F., Tiribelli, C., and Kelley, M.R. (2009). The many functions of APE1/Ref-1: not only a DNA repair enzyme. Antioxid Redox Signal *11*, 601-620.

Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D., and Eichman, B.F. (2019). Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. Nat Struct Mol Biol *26*, 613-618.

Toma, A., Takahashi, T.S., Sato, Y., Yamagata, A., Goto-Ito, S., Nakada, S., Fukuto, A., Horikoshi, Y., Tashiro, S., and Fukai, S. (2015). Structural basis for ubiquitin recognition by ubiquitin-binding zinc finger of FAAP20. PLoS One *10*, e0120887.

Torisu, K., Tsuchimoto, D., Ohnishi, Y., and Nakabeppu, Y. (2005). Hematopoietic tissue-specific expression of mouse Neil3 for endonuclease VIII-like protein. J Biochem *138*, 763-772.

Tran, O.T., Tadesse, S., Chu, C., and Kidane, D. (2020). Overexpression of NEIL3 associated with altered genome and poor survival in selected types of human cancer. Tumour Biol *42*.

Tse-Dinh, Y.C. (1991). Zinc (II) coordination in Escherichia coli DNA topoisomerase I is required for cleavable complex formation with DNA. J Biol Chem *266*, 14317-14320.

Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C., and Fushman, D. (2004). Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. J Biol Chem *279*, 7055-7063.

Varela, J.G., Pierce, L.E., Guo, X., Price, N.E., Johnson, K.M., Yang, Z., Wang, Y., and Gates, K.S. (2021). Interstrand Cross-Link Formation Involving Reaction of a Mispaired Cytosine Residue with an Abasic Site in Duplex DNA. Chem Res Toxicol *34*, 1124-1132.

Vijay-Kumar, S., Bugg, C.E., and Cook, W.J. (1987). Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol *194*, 531-544.

Wallace, B.D., Berman, Z., Mueller, G.A., Lin, Y., Chang, T., Andres, S.N., Wojtaszek, J.L., DeRose, E.F., Appel, C.D., London, R.E.*, et al.* (2017). APE2 Zf-GRF facilitates 3'-5' resection of DNA damage following oxidative stress. Proc Natl Acad Sci U S A *114*, 304-309.

Wallace, S.S., Bandaru, V., Kathe, S.D., and Bond, J.P. (2003). The enigma of endonuclease VIII. DNA Repair (Amst) *2*, 441-453.

Wang, B., Alam, S.L., Meyer, H.H., Payne, M., Stemmler, T.L., Davis, D.R., and Sundquist, W.I. (2003). Structure and ubiquitin interactions of the conserved zinc finger domain of Npl4. J Biol Chem *278*, 20225-20234.

Wilson, D.M., 3rd, and Barsky, D. (2001). The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. Mutat Res *485*, 283-307.

Wu, M., and Lander, G.C. (2020). How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. Curr Opin Struct Biol *64*, 9-16.

Wu, R.A., Semlow, D.R., Kamimae-Lanning, A.N., Kochenova, O.V., Chistol, G., Hodskinson, M.R., Amunugama, R., Sparks, J.L., Wang, M., Deng, L.*, et al.* (2019). TRAIP is a master regulator of DNA interstrand crosslink repair. Nature *567*, 267-272.

Yang, L.X., Zhang, X., and Zhao, G. (2016). Ginsenoside Rd Attenuates DNA Damage by Increasing Expression of DNA Glycosylase Endonuclease VIII-like Proteins after Focal Cerebral Ischemia. Chin Med J (Engl) *129*, 1955-1962.

Yang, W. (2006). Poor base stacking at DNA lesions may initiate recognition by many repair proteins. DNA Repair (Amst) *5*, 654-666.

Yang, Z., Price, N.E., Johnson, K.M., Wang, Y., and Gates, K.S. (2017). Interstrand cross-links arising from strand breaks at true abasic sites in duplex DNA. Nucleic Acids Res *45*, 6275-6283.

Yaseen, N.R., and Blobel, G. (1999). Two distinct classes of Ran-binding sites on the nucleoporin Nup-358. Proc Natl Acad Sci U S A *96*, 5516-5521.

Zhang, H.L., Malpure, S., Li, Z., Hiasa, H., and DiGate, R.J. (1996). The role of the carboxyl-terminal amino acid residues in Escherichia coli DNA topoisomerase III-mediated catalysis. J Biol Chem *271*, 9039-9045.

Zhang, J., Dewar, J.M., Budzowska, M., Motnenko, A., Cohn, M.A., and Walter, J.C. (2015). DNA interstrand cross-link repair requires replication-fork convergence. Nat Struct Mol Biol *22*, 242-247.

Zhao, C., Liu, J., Zhou, H., Qian, X., Sun, H., Chen, X., Zheng, M., Bian, T., Liu, L., Liu, Y.*, et al.* (2021). NEIL3 may act as a potential prognostic biomarker for lung adenocarcinoma. Cancer Cell Int *21*, 228.

Zhou, J., Chan, J., Lambele, M., Yusufzai, T., Stumpff, J., Opresko, P.L., Thali, M., and Wallace, S.S. (2017). NEIL3 Repairs Telomere Damage during S Phase to Secure Chromosome Segregation at Mitosis. Cell Rep *20*, 2044-2056.

Zhu, C., Lu, L., Zhang, J., Yue, Z., Song, J., Zong, S., Liu, M., Stovicek, O., Gao, Y.Q., and Yi, C. (2016). Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair. Proc Natl Acad Sci U S A *113*, 7792-7797.