Characterization of Bacterial DNA Glycosylases Involved in Interstrand Crosslink Repair and Antibiotic Resistance

By

Noah Patrick Bradley

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

May 13, 2022

Nashville, Tennessee

Approved:

Katherine L. Friedman, Ph.D.

Brandt F. Eichman, Ph.D.

Jared T. Nordman, Ph.D.

Carmelo J. Rizzo, Ph.D.

Frederick P. Guengerich, Ph.D.

Dedication

To my dearest grandparents, Dr. Samuel and Mary Laneve, for all of your hard-work and sacrifices, and for being loving role models for our one big family.

To Dan Sharp, a compassionate and caring father, husband, brother, and son who will always be with us in our hearts. You motivate us to always be our best in life.

For my family who supported me in every aspect of my scientific and personal journey, I am forever grateful. You opened the door for all of the opportunities in my life, and I share my successes with everyone.

Acknowledgements

I would like to thank the numerous family, friends, colleagues, and collaborators for their unwavering support of my dreams, hopes, and goals- both while in graduate school and well-before. To my kind-hearted and extraordinarily hard-working parents, I love you for all you have done for me and the family, and being there for me every step of the way. Without your encouragement and guidance, life would be a much tougher journey to manage. I will always be thankful for the lessons you taught me, the sacrifices you made working hard for us, and the love shared along the way. To my sweet, loving grandparents who passed away during my graduate school journey, you are missed with all of my heart. You taught me respect and the values of putting your family first, and I cherish every moment that we got to spend together; I will continue to honor you as I move through my life and career. To my brother and sisters (Andy, Ali, Hannah, Lilly), I couldn't imagine a life without you all; your families have always been a source of inspiration and encouragement for me to persevere even through the toughest of times. As the youngest of the siblings, every one of you took care of me growing up in some form or fashion, and it has made a lasting impact on my life. To the rest of my family- my aunts, uncles, cousins, nieces & nephews, and the rest of the extended relatives who played a crucial role in my journey and progress, I have endless gratitude for the advice, support, love, joy, and bonding we have all shared together as a family. I wish I had the time and space to mention everyone individually for the position they have played in my life, but suffice it to say that I love each and every one of you for being there when it has counted the most.

To my advisor and mentor, Dr. Brandt Eichman, there are not enough words to describe the impact you have had on my scientific career and development as a burgeoning young scientist. You have always led by example, whether that pertains to designing experiments, giving professional talks, assembling data for publication, or being a collegial and collaborative scientist. You have always engendered a productive and fascinating lab environment with diverse projects and students/trainees. The freedom you provided me in my project will have a lasting impact on my scientific career, and your support has given me many significant opportunities I might never have had otherwise. If I'm fortunate enough to one day run a lab and research program with even a fraction of your success, I would be satisfied and honored. My thesis committee has also been extremely valuable and supportive of me during my graduate school trajectory. From bouncing ideas off of one another, to giving me advice on which experiments to pursue, each one of you has supplied my projects with significant benefits that pushed them to completion, and I am extremely grateful for that. To the Eichman lab as a whole, both past and present members, I cannot thank you all enough for the many facets of my training and research you have impacted. From listening to my wild ideas for projects, to training me in diverse techniques and disciplines, and troubleshooting the infinite problems that arise in working in a lab- you have all been an integral part of my work. I am particularly thankful to Dr. Elwood Mullins for his mentoring during my rotation in the Eichman lab in 2016, and for always lending an ear or eye for any exciting data during every step of my research. I am appreciative of our wonderful lab manager, Briana Greer, for her hard work and availability, and for always keeping the lab running smoothly.

My final section of acknowledgements is dedicated to all of the friends, collaborators, and colleagues I have had the exceptional privilege of meeting and working with during my exciting journey. To the numerous graduate school friends I've made, you helped create an environment which was both incredibly fun and stimulating from an academic perspective and beyond. To my core group of graduate friends (Abbie, Ari, Annah, Clare, Jacob, Joe, Kevin, Mike (Mark), Nate, and Tim), you are my friends for life and I cherish all of the memories we made together; I look forward to following all of you during your successes in your scientific journeys. To Mac & Abbie, you guys have been the best of friends to me, and I'm so thankful for a lifelong friendship with you two. To the friends I made while living in Nashville (Ben, Francesca, Josh, Maci, Tanner (and the Marions), Zach, and many more), thank you for befriending me despite a hectic schedule, and always lending an ear to my science- even if you didn't ask! I will always hold you all near and dear to my heart, and I hope to have many chances in the future to visit y'all. For the friends beyond Nashville that I've made (Alex, Bri, Britt, Cody, Chris, Dan, David, Jess, Matt, Megan, Stevie, and of course many more), I cannot thank you enough for the experiences we have shared together, and for holding my sanity in check during the ups and downs of graduate school- you guys mean more to me than you might ever know! To my collaborators, both at Vanderbilt and beyond, thank you for being incredibly supportive, patient, and rigorous with our research endeavors together; you made my successes possible in ways beyond I ever imagined. To Dr. Antonis Rokas and Jacob Steenwyk, your collegiality and expertise were essential to many aspects of my research; thank you for your support and knowledge both in the lab and beyond. For everyone I couldn't mention personally, know that in my heart I am eternally grateful for the love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Δ: heat

5'-dRP: 5'-deoxyribose phosphate

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

AMR: antimicrobial resistance

ANOVA: analysis of variances

AP: apurinic/apyrimidinic/abasic site

APE1: AP endonuclease 1

APE2: AP endonuclease 2

AP-ICL: abasic interstrand crosslink

AZB: azinomycin B

AZL: AlkZ-like

BER: base excision repair

β-ME: β-mercaptoethanol

BGC: biosynthetic gene cluster

CLB: colibactin

CPD: cyclobutane pyrimidine dimer

CRC: colorectal cancer

d7mG / 7mG: *N*7-deoxymethylguanosine

dA-AP ICL: deoxyadenosine abasic interstrand crosslink

DDR: DNA damage response/repair

DHT: dihydrothymine

DNA: deoxyribonucleic acid

DPC: DNA protein crosslink

DSB: double strand break

dsDNA: double stranded DNA

EC_{50}: 50% effective concentration

E. coli: *Escherichia coli* (*Eco*)

EMSA: electrophoretic mobility shift assay

EndoIV: endonuclease 4 (*nfo*)

ExoIII: exonuclease 3 (*xth*)

FA: Fanconi anemia

FaPyA: 4,6-diamino-5-formamidopyrimidine

FaPyG: 2,6-diamino-4-hydroxy-5-formamidopyrimidine
Fpg: formamidopyrimidine DNA glycosylase
GG-NER: global genomic NER
Gh: guanidinohydantoin
GO: gene ontology
HED: hedamycin
HGT: horizontal gene transfer
HMCES: 5-hydroxymethylcytosine binding, embryonic stem cell-specific
HN2: mechlorethamine
HPLC-MS: high-performance liquid chromatography tandem mass spectrometry
HR: homologous recombination
HRESIMS: high-resolution electrospray ionization mass spectrometry
HTH42: winged helix-turn-helix 42 superfamily
ICL: interstrand crosslink
IR: ionizing radiation
kDa: kilodalton
Lig1: ligase 1
LLD: LL-D49194α1
MA: monoadduct
MMC: mitomycin C
MMR: mismatch repair
MMS: methylmethanesulfonate
MoA: mechanism of action
Nei: endonuclease VIII
NEIL3: endonuclease VIII-like 3
NEB: New England BioLabs
NER: nucleotide excision repair
NHEJ: non-homologous end joining
NM: nitrogen mustard
NMU: *N*-methyl-*N*-nitrosourea
NRPS: non-ribosomal peptide synthetase
nt: nucleotide
ORF: open reading frame

P: 3′-phosphate
PAGE: polyacrylamide gel electrophoresis
PDB: protein data base
PKS: polyketide synthase
PMSF: phenylmethylsulfonyl fluoride
PNKP: polynucleotide kinase phosphatase
POL β: Polymerase beta
PUA: 3′-phospho-α,β-unsaturated aldehyde
RGM: resistance genome mining
RNA: ribonucleic acid
ROS: reactive oxygen species
RMSD: root-mean-square deviation
SCPCHD: spirocyclopropylcyclohexadienone family of natural products
SD: standard deviation
Spu: *Spizellomyces punctatus*
Ssa: *Streptomyces sahachiroi*
SSB: single strand break
ssDNA: single stranded DNA
SSN: sequence similarity network
STZ: streptozotocin
TC-NER: transcription-coupled NER
Tcu: *Thermomonospora curvata*
Tfu: *Thermobifida fusca*
Tg: thymine glycol
THF: tetrahydrofuran
TLS: translesion synthesis
TXNA: trioxacarcin A
UDG: uracil DNA glycosylase
UV: ultraviolet radiation
WH: winged helix
YTM: yatakemycin
YTMA: 3-yatakemycinyl-2'-deoxyadenosine
YQL: YcaQ-like

Chapter 1

Introduction[1]

DNA is a fundamental biomolecule of life which stores the genetic information of organisms and drives many of the evolutionary processes observed in nature. The accurate and faithful replication and segregation of DNA is essential for life and the propagation of species. Despite the crucial need to protect and maintain genomic integrity, DNA is subject to both physical and chemical damage from both endo- and exogenic sources in nature which impair the ability to replicate the DNA and access the genomic information for gene expression and regulation. It is therefore no surprise that organisms have evolved unique genomic and biochemical adaptations to produce antibiotic chemical agents which can modify or alter DNA (genotoxins) in competing organisms to impact fitness of a species. It is thus essential for these organisms to possess parallel, often highly specialized DNA repair pathways to combat the damage which can occur to self, and to protect the genome from damage which can modify the information contained within the sequence. This dissertation will focus on the bacterial base excision repair (BER) pathway and novel discoveries and implications in repair of highly specialized and potent bacterial genotoxins and interstrand crosslinks (ICLs). These studies emphasize the genomic, biochemical, structural, and functional properties of the AlkZ/YcaQ family of DNA glycosylases in bacteria and their role in DNA repair.

Mechanisms and Classes of DNA Damage

DNA is susceptible to chemical modification at virtually every single hetero and carbon atom within the nucleotide structure (Gates, 2009) (Fig. 1A). DNA damage can be classified by the various chemical and physical mechanisms by which nucleic acids can be altered. In general, DNA damage can fall into distinct categories whereby there is modification of the nucleobase, deoxyribose sugar, or phosphate backbone. Here, I describe the general classes and mechanisms of DNA damage arising from both exogenous and endogenous sources, and then elaborate on the specific, conserved DNA repair pathways which protect cellular life from genomic damage. I conclude the introduction with descriptions of evolutionarily specialized DNA repair pathways involving base excision repair (BER) involved in bacterial natural product antibiotic self-resistance and interstrand crosslink (ICL) repair.

Types of DNA damage

The maintenance of genome integrity is dependent on the stability of the information stored within the DNA nucleotide sequence. However, DNA contains many functionalized heteroatoms and reactive portions which can be modified by agents within the environment (Gates, 2009). DNA is composed of a double helix of nucleotides which form specific hydrogen bonding base pairs within the duplex (Fig. 1A). Common types of modifications to the nucleic acid structure include alkylation, hydrolytic deamination, or oxidation of the atoms within DNA (Gates, 2009) (Fig. 1A). Initial formation of these DNA adducts or alterations can lead to secondary DNA damage which can include single-strand breaks (SSBs), double-strand breaks (DSBs), misincorporated nucleotides during

replication, chromosomal rearrangements, insertions/deletions (indels), and ultimately cell death due to genome degradation/deregulation (Liu *et al.*, 2016; Chatterjee and Walker, 2017) (Fig. 1B). Chemical agents that covalently modify the DNA to form adducts can create a diverse set of alkylated products: small base modification (methylation), crosslinking of the nucleobases (inter- and intra-strand crosslinks), or formation of bulky, helix-distorting lesions (Gates, 2009) (Fig. 1B). In general, these various classes of DNA damage can arise by certain types of DNA damaging agents (genotoxins, metabolites), or physical processes (high-energy radiation) (Fig. 1C). The association between a compound or process that can modify DNA and the repair pathways which can recognize and fix the damage is a highly intricate and evolutionarily conserved process in nature. I will next focus on the sources of DNA damage that arises from both endogenous and exogenous sources.

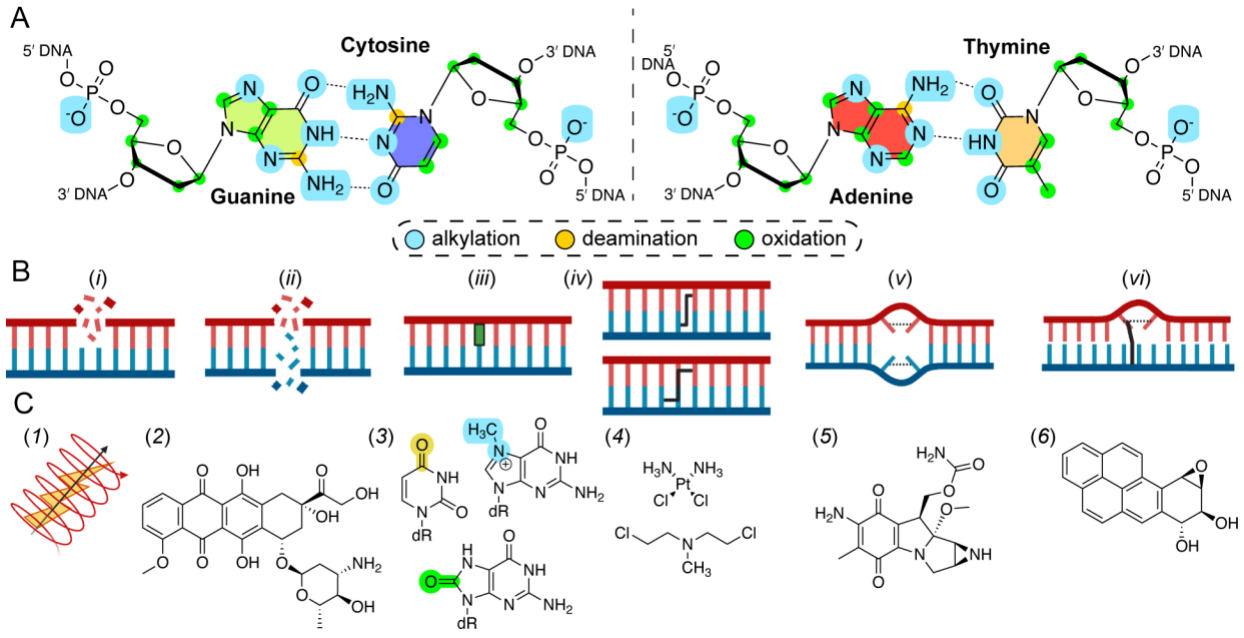


Figure 1. *Mechanisms and classes of DNA damage* (A) Chemical structure of a G:C and A:T base pair and sites/types of reactivity for each atom within DNA. Cyan-alkylation damage, tan-deamination, light green-oxidation. (B) Cartoon schematic of classes of DNA damage. From left to right: (*i*) single-strand breaks (SSBs) (*ii*) double-strand breaks (DSBs) (*iii*) base modifications (*iv*) interstrand crosslinks (ICLs); top-1,2 ICL, bottom-1,3 ICL (v) intrastrand crosslink (v*i*) bulky monoadducts (C) Chemicals, agents, or common types of DNA damage which give rise to the classes of damage in panel B. From left to right: (1) ionizing radiation (IR) (2) doxorubicin (3) deoxyuracil-tan, 7-methylguanine-cyan, 8-oxoguanine-light green (4) cisplatin-top, mechlorethamine-bottom (5) mitomycin C (6) benzo[*a*]pyrene epoxide. Figures B-C were produced with BioRender.

Sources of DNA damage

Cellular life developed and evolved in environments which are constantly under assault by physical and chemical agents that can damage DNA as well as other essential biomolecules. These sources arise from unavoidable environmental phenomena such as ionizing radiation and reactive cellular metabolites which are either byproducts of cellular metabolism, or are genetic adaptions to produce specialized defense molecules which impact an organisms fitness. Here I will describe and elaborate on natural and artificial process which can damage DNA from exogenous and endogenous sources.

Exogenous sources

An exogenous source of DNA damage would arise from a chemical or physical process that originates outside the cell where the damage occurs. Some of the most common types of exogenous DNA damage occurs from the inherent properties of the electromagnetic spectrum of light as it interacts with DNA and associated biomolecules. High-energy, low wavelength ionizing radiation (IR- X-rays, gamma rays, cosmic rays) can directly interact with the bonds in DNA, leading to their destruction and fragmentation

of the DNA through SSBs and DSBs (Sinha and Hader, 2002; Lomax *et al.*, 2013) (Fig. 2A). Ultraviolet radiation (UV) is a lower-energy wavelength of light which can excite the bonds within the DNA, leading to crosslinking of pyrimidine nucleobases to form cyclobutane pyrimidine dimers or 6-4 photoproducts (Rastogi *et al.*, 2010) (Fig. 2C). In addition to light-based exogenous sources of DNA damage, viral infections are also a source of host genomic instability, as many viruses hijack the host nucleic acid processing machinery and degrade the host genome through nuclease action (Parson and Snustad, 1975; Weitzman and Fradet-Turcotte, 2018) (Fig. 2B).

Additionally, many bacteria, fungi, and plants produce specialized secondary metabolites that target competing organisms' DNA, which can cause inhibition of processes that require accessing the information within the genome (transcription, replication, repair) (Huang *et al.*, 2021) (Fig. 2D). These natural products include the mycotoxin family of aflatoxins, plant psoralens, and topoisomerase inhibitors such as camptothecin. Because DNA damaging agents are highly efficient at inhibiting DNA replication, which can ultimately lead to cell death, it is therefore no surprise that these chemicals are used therapeutically to treat a variety of diseases, in particular cancer (Cheung-Ong *et al.*, 2013). These agents include synthetic alkylating agents such as *N*-methyl-*N*-nitrosourea derivatives (NMU-methylating agent), intercalating agents such as doxorubicin, and crosslinking agents such as the nitrogen mustard family. Rapid and efficient repair of these types exogenous damage is thus essential for chemotherapeutic resistance mechanisms, and understanding how these alterations are dealt with is an ongoing aspect of cancer biology.

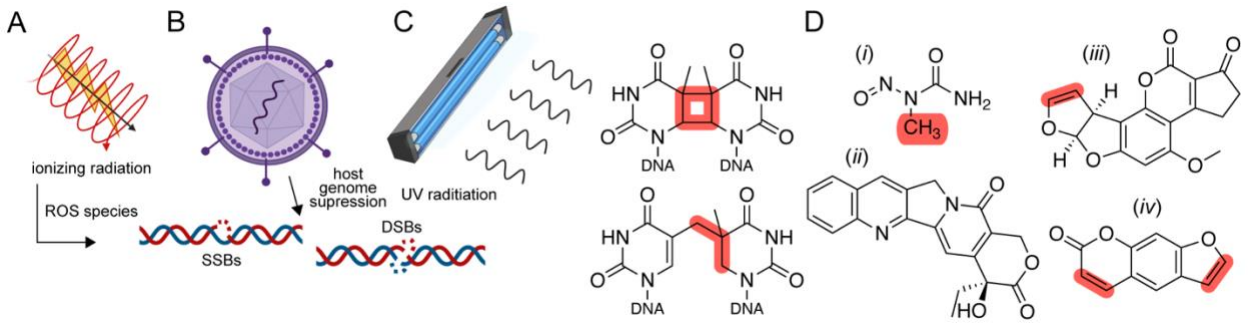


Figure 2. *Exogenous sources of DNA damage* (A) Ionizing radiation (X-rays, gamma rays, cosmic radiation) can directly damage the chemical bonds in DNA, or indirectly through the generation of reactive oxygen species (ROS), leading to single-strand and double-strand breaks. (B) Viral infections suppress the host genome and lead to degradation of the genomic DNA. (C) Ultraviolet (UV) radiation can induce the formation of cyclobutane pyrimidine dimers (CPDs-top), or 6-4 photoproducts-bottom. (D) Common chemical sources of DNA damage: (*i*) methylnitrosourea-methylating agent (*ii*) camptothecin-topoisomerase poison (*iii*) aflatoxin B_1-alkylating agent (v*i*) psoralen-ICL agent. Reactive portions/sites of DNA damage are highlighted in red. Figures A-C were produced with BioRender.

Endogenous sources

An endogenous source of DNA damage would arise from cellular, biochemical, or metabolic processes within the cell where the DNA damage is occurring (Tubbs and Nussenzweig, 2017). Many of these types of endogenous damage are unavoidable due to the conserved nature of DNA processing and metabolic requirements of cellular life. In general, these categories of damage can be classified into two types: (I) damage derived from nucleic acid-processing enzymes and (II) damage derived from reactive metabolites (Fig. 3 A-B).

The chemical and physical processing of DNA by nucleic acid-processing enzymes is essential for the replication, transcription, repair, and recombination of the genome (Hoeijmakers, 2001). However, these proteins can also lead to damage within DNA through a variety of mechanisms. DNA polymerases are involved in the replication and

repair of DNA, and together with the replication machinery faithfully duplicate the genome. Most DNA polymerases have high-fidelity when incorporating nucleotides across from the templates, however errors do arise at a low rate during replication (Bebenek and Ziuzia-Graczyk, 2018). Specialized low-fidelity translesion synthesis (TLS) polymerases can also bypass damage which occurs on the template DNA and often can incorporate the incorrect nucleotide to ensure the genome is replicated (Yang and Gao, 2018) (Fig. 3A). These errors in replication and repair lead to mutations in the genetic code, which can contribute to disease progression (Makridakis and Reichardt, 2012). Other DNA-processing enzymes such as DNA glycosylases, deoxyribonucleases, and topoisomerases can lead to strand breaks (SSBs, DSBs) and subsequent loss of genomic information (Klapacz *et al.*, 2010; Fu *et al.*, 2013; Williams *et al.*, 2019) (Fig. 3A). Due to the instability these enzymes can create, they are under tight regulation within a cellular and genetic context to minimize this damage.

In addition to these enzymatic sources of DNA damage, chemical sources of damage arise from reactive metabolites and byproducts of cellular metabolism (Fig. 3B). Alkylation of the DNA is a common type of damage, and can form from several metabolites or co-factors. Formaldehyde is a by-product of one-carbon metabolism and repair of DNA damage and can react with nucleophilic portions of the nucleobases to from monoadducts and crosslinks (Grafstrom *et al.*, 1983) (Fig. 3B). *S*-adenosylmethionine is a cofactor involved in many cellular processes involving methyl transfer reactions and contains an electrophilic *S*-methyl group that can alkylate DNA at various positions (Rydberg and Lindahl, 1982) (Fig. 3B). The metabolism and spontaneous oxidation of cellular lipids such as 4-hydroxynonenal (4-HNE) can also generate complex DNA

adducts and reactive oxygen species (ROS) such as hydroxyl radicals that can further modify DNA (Voulgaridou *et al.*, 2011) (Fig. 3B). Finally, catabolism of steroid-based hormones can generate reactive catechol quinone species, which produce ROS and lead to DNA degradation (Han and Liehr, 1995) (Fig. 3B). These diverse, inevitable sources of DNA damage pose a severe threat to cellular life, and as such, DNA repair pathways are equipped to prevent and restore these alterations to the genome.

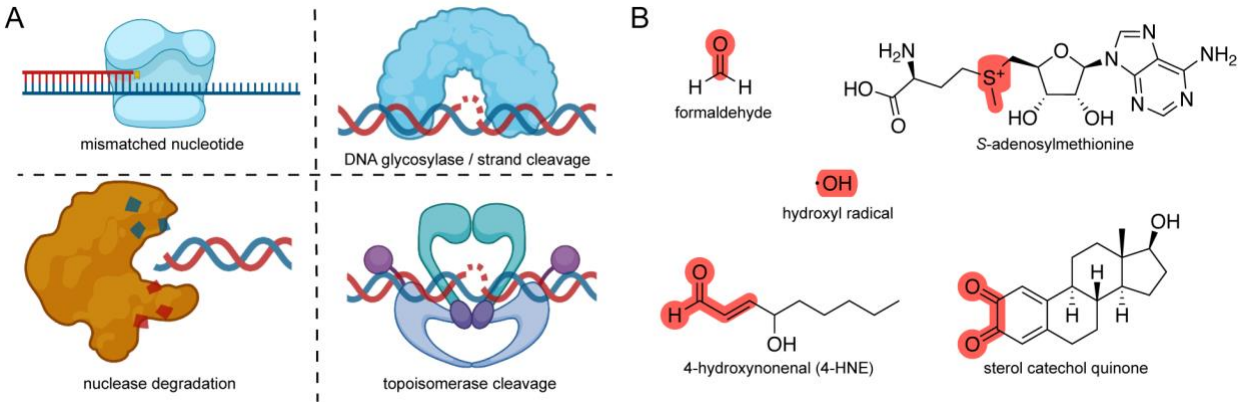


Figure 3. *Endogenous sources of DNA damage* (A) Damage arising from nucleic acid processing enzymes. From top left: incorporation of incorrect nucleotide by DNA polymerases, strand cleavage by bifunctional DNA glycosylase, nuclease degradation of DNA, and topoisomerase cleavage complex. (B) Metabolites which can chemically damage DNA. From top left: formaldehyde-one carbon metabolism/alkylation, *S*-adenosylmethionine-cofactor/alkylation, hydroxyl radical-general metabolism/oxidizer, 4-hydroxynonenal (4-HNE)-lipid peroxidation/alkylation, and sterol catechol quinones-steroid metabolism/ROS generator. Reactive portions are highlighted in red. Figure A was produced with BioRender.

DNA Repair Pathways

The totality of DNA damaging agents which arise from both endogenous and exogenous sources poses a challenge for cellular life: how can the integrity of the genome be maintained despite a constant assault of the DNA from all angles? To combat these processes, nature has evolved unique and specialized DNA repair pathways which can

recognize and repair the diverse types of damage arising in the natural world. In these next sections, I will describe and elaborate on bacterial DNA repair pathways as they are currently understood, and highlight outstanding questions in the field. The focus will be on the base excision repair (BER) pathway and DNA glycosylases, which initiate this pathway.

Base excision repair (BER) pathway

Base excision repair (BER) is a conserved DNA repair pathway from bacteria to vertebrates and canonically maintains genome integrity from damage arising from small nucleobase modifications occurring on a single strand of DNA (Krokan and Bjoras, 2013) (Fig. 4A). BER is initiated by a damage-specific DNA glycosylase (elaborated in the next section) that recognizes, binds, and catalyzes the hydrolysis of a modified nucleobase (alkylation, oxidation, deamination damage) from the phosphodeoxyribose backbone to generate an apurinic/apyrimidinic (AP/abasic) site (Mullins *et al.*, 2019). This step is the most diverse in BER, as numerous, diverse superfamilies of DNA glycosylases have been described (Brooks *et al.*, 2013). Bifunctional DNA glycosylases (elaborated later) also catalyze the elimination of the AP site. The second step of BER occurs when an AP endonuclease (ExoIII or EndoIV in bacteria) incises the AP site 5' to the lesion to nick the DNA (Mullins *et al.*, 2019). If no bifunctional activity was performed by a DNA glycosylase in step one, this nicked DNA must be end-processed through β-lyase activity of DNA polymerase I, and subsequently repair synthesis is performed by DNA Pol I. Finally, the DNA is sealed by the action of DNA ligase I to repair the DNA in an error-free process (Mullins *et al.*, 2019) (Fig 4A).

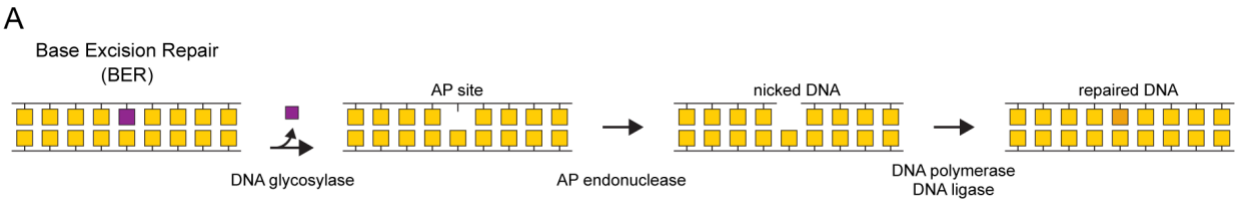


Figure 4. *General overview of the base excision repair (BER) pathway* (A) Cartoon schematic of the base excision repair pathway in bacteria. From left: a damage-specific DNA glycosylase catalyzes the hydrolysis of the damaged base (purple) from the deoxyribose backbone to generate an abasic (AP) site. AP endonucleases (ExoIII and EndoIV) incise the AP site 5' to the lesion to create a gap. This nicked DNA is end-processed to remove the deoxyribose from the backbone, and DNA polymerase I performs repair synthesis. The gap is sealed by DNA ligase I in an error-free process.

DNA glycosylases

DNA glycosylases perform the first step in the BER pathway (base excision), and are an extremely evolutionarily diverse family of DNA repair enzymes with multiple substrates, mechanisms, and structures (Fromme and Verdine, 2004; Brooks *et al.*, 2013; Mullins *et al.*, 2019). DNA glycosylases can be categorized broadly into two categories: (I) monofunctional DNA glycosylases (catalyzes only base excision), and (II) bifunctional DNA glycosylases (catalyze base excision and AP elimination). These two types of DNA glycosylases recognize and excise a variety of DNA adducts and lesions to protect the genome from mutational errors which arise when damaged DNA is replicated. In this section, I elaborate and describe the mechanistic, structural, genetic, and substrate repertoires of both monofunctional and bifunctional DNA glycosylases.

Monofunctional glycosylases

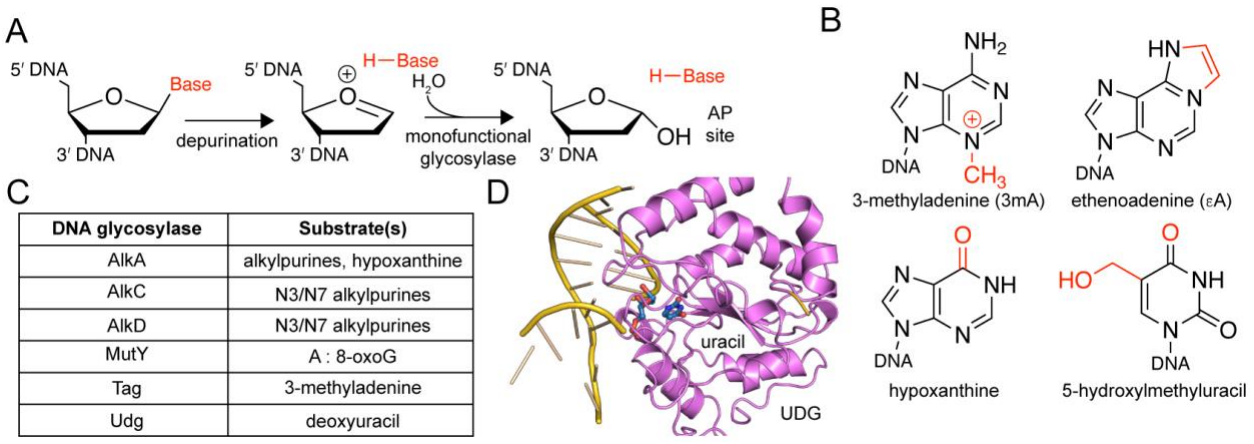
Monofunctional DNA glycosylases perform only the first step in base excision: hydrolysis of a damaged base from the deoxyribose backbone (Mullins *et al.*, 2019). The

catalytic mechanisms by which monofunctional glycosylases operate involve a dissociative S_N1 mechanism by which the enzyme first recognizes and binds to its cognate DNA damage, and then base depurination occurs (spontaneous or enzyme-catalyzed) (Drohat and Maiti, 2014) (Fig. 5A). Base depurination generates a transient oxocarbenium intermediate where the C1' anomeric carbocation is resonance stabilized by lone pairs on O4' (Fig. 5A). Monofunctional DNA glycosylases then catalyze the addition of water across the C1' anomeric carbocation (Fig. 5A) by either pre-activating a water molecule for nucleophilic attack, or through positioning of the water to have optimal access to C1' (Drohat and Maiti, 2014). In general, these enzymes use catalytic carboxamide (Asn, Gln) or carboxylate (Asp, Glu) side chains to perform the chemistry of base excision (Mullins *et al.*, 2019). A common theme for DNA adducts or lesions recognized and repaired by monofunctional glycosylases is destabilization of the *N*-glycosidic bond or base pairing ability which is dependent on the site of the nucleobase the adduct is positioned (O'Brien and Ellenberger, 2004; Drohat and Maiti, 2014) (Fig. 5B). These damages generally contain either positive charges on the nucleobase (3mA, 7mG) which allow the depurinated base to leave as a neutral group or other positions that also destabilize the base pairing ability (Drohat and Maiti, 2014) (Fig. 5B). As such, this class of glycosylases generally recognize damage that is distinct from those repaired by bifunctional glycosylases.

Due to the extreme diversity in damage that can arise to nucleobases in DNA, there are multiple superfamilies of monofunctional DNA glycosylases that recognize and repair certain types of these base lesions (Brooks *et al.*, 2013) (Fig. 5C). In bacteria, monofunctional glycosylases repair lesions derived from alkylpurine damage (AlkA, AlkC,

AlkD, Tag), deaminated cytosine (Udg), and mispaired oxidized purines (MutY) (Mullins *et al.*, 2019). These enzymes are the first step in protection against mutations and cytotoxicity of these adducts. Additional specialized glycosylase systems exist in bacteria such as the restriction endonuclease DNA glycosylase R.PabI, which recognizes and excises adenines in specific sequences to protect against phage infection (Miyazono *et al.*, 2014).

The structural mechanisms of how monofunctional DNA glycosylases recognize and catalyze base excision is a well-studied process with some general properties, although new advancements in our understanding of these enzymes have expanded the known structural properties of these proteins (Mullins *et al.*, 2019). In general, DNA glycosylases contain an active site pocket that can accommodate a flipped-out nucleotide to position within proximity to the catalytic residues, a process termed “base-flipping” (Slupphaug *et al.*, 1996; Brooks *et al.*, 2013) (Ex. uracil DNA glycosylase, Fig. 5D). To maintain base stacking within the duplex, intercalating residues are positioned within the site where the flipped-out nucleobase was. The flipped-out nucleobase is then depurinated and hydrolyzed to create an AP site, which is often tightly-bound by the enzyme (Mullins *et al.*, 2019). The structural restraints of base-flipping into a space-limited active site is the basis for why canonical DNA glycosylases cannot excise bulky or crosslinked nucleobases.



DNA glycosylase	Substrate(s)
AlkA	alkylpurines, hypoxanthine
AlkC	N3/N7 alkylpurines
AlkD	N3/N7 alkylpurines
MutY	A : 8-oxoG
Tag	3-methyladenine
Udg	deoxyuracil

Figure 5. *Structural and biochemical overview of monofunctional DNA glycosylases* (A) Chemical mechanism of monofunctional DNA glycosylase base excision. Within the enzyme active site, base depurination occurs (spontaneous or enzyme-catalyzed) to generate an oxocarbenium intermediate (middle). In the second step, the glycosylase catalyzes the addition of water across the C1' anomeric carbon, generating an AP site. (B) Chemical structures of common lesions recognized and excised by monofunctional DNA glycosylases. (C) List of common bacterial monofunctional DNA glycosylases and most prevalent lesions they excise. (D) Crystal structure of human uracil DNA glycosylase (UDG) in complex with deoxyuracil-containing DNA (PDB: 4SKN). The excised and flipped-out uracil (blue) is located within the active site pocket.

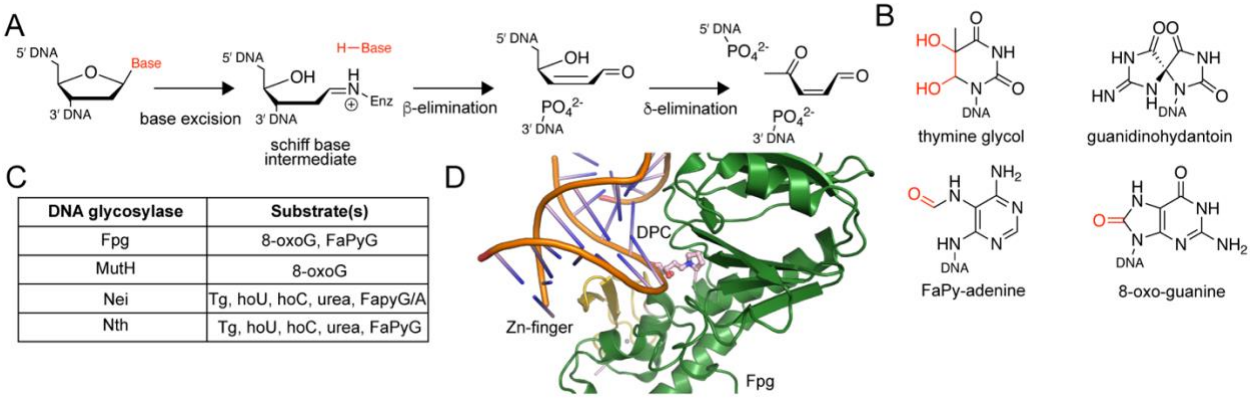
Bifunctional glycosylases

Bifunctional DNA glycosylases perform multiple steps within the BER pathway and excise and repair damage that is generally distinct to that repaired by monofunctional glycosylases (Prakash *et al.*, 2012; Mullins *et al.*, 2019). Bifunctional glycosylases first bind to their cognate lesions and perform base excision through formation of a covalent enzyme-substrate intermediate (DNA-protein crosslink; DPC) (Mullins *et al.*, 2019) (Fig. 6A). This step is generally achieved through a nucleophilic attack of a nitrogen (N) of Lys, Pro, or the N-terminal amine (Fromme *et al.*, 2004). This generates the DPC intermediate which is characterized by a Schiff base (imine) between a nitrogen and C1' (Fig. 6A). The second function of bifunctional glycosylases is lyase activity, whereby the Schiff base

DPC intermediate undergoes enzyme-catalyzed β- or β/δ-elimination to nick the DNA either 3' or 3'+5' of the AP site (Fig. 6A) and generates an α,β-unsaturated aldehyde. Because a 3'-PO_4^{2-} group is generated during this step, a free 3' hydroxyl group must be generated by a polynucleotide phosphatase/kinase (PNPK) before repair synthesis and ligation can occur (Krokan and Bjoras, 2013). The majority of DNA lesions recognized and repaired by bifunctional DNA glycosylases do not lead to significant destabilization of the *N*-glycosidic bond and are primarily derived from oxidation of the nucleobases (Prakash *et al.*, 2012) (Fig. 6B). This lack of destabilization for these adducts and lower propensity for spontaneous depurination may help explain why the two types of DNA glycosylases possess distinct catalytic mechanisms.

Bifunctional glycosylases are also highly conserved across all domains of life, but seem to have less diversity than monofunctional DNA glycosylases (Fig. 6C). In bacteria, at least four superfamilies of bifunctional enzymes exist, many of which are conserved in vertebrates (Prorok *et al.*, 2021) (Fig. 6C). These excise a diverse subset of oxidized purines and pyrimidines from the genome, but some more specialized vertebrate enzymes such as NEIL3 can catalyze the unhooking of ICLs derived from natural products and endogenous ICLs (Semlow *et al.*, 2016; Imani Nejad *et al.*, 2020; Li *et al.*, 2020). Structurally, bifunctional glycosylases operate in a manner somewhat similar to how monofunctional enzymes excise their damage in the sense that base-flipping is a common process to perform base excision (Mullins *et al.*, 2019). For example, the formamidopyrimidine DNA glycosylase (Fpg) has been covalently trapped in the enzyme-substrate intermediate to reveal the N-terminal proline residue covalently adducted to the C1' of the AP site (Fromme and Verdine, 2002) (Fig. 6D). The AP site-crosslink is rotated

into the active site of Fpg, and an intercalating residue stacks in the bases to maintain the duplex. These major differences between the classes of DNA glycosylases has led to many important discoveries for their implications in cancer, antibiotic resistance, and evolution (Jiang and Ramachandran, 2016; Mullins *et al.*, 2019).



DNA glycosylase	Substrate(s)
Fpg	8-oxoG, FaPyG
MutH	8-oxoG
Nei	Tg, hoU, hoC, urea, FapyG/A
Nth	Tg, hoU, hoC, urea, FaPyG

Figure 6. *Structural and biochemical overview of bifunctional DNA glycosylases* (A) Chemical mechanism of bifunctional DNA glycosylase base excision. Base excision occurs in the first step when the nitrogen of either a lysine, proline, or the N-terminus attacks the C1' anomeric carbon, which opens the deoxyribose ring and forms a covalent Schiff base intermediate. The second functional step that bifunctional glycosylases catalyze is nicking of the DNA backbone through β-elimination or β+δ-elimination of the 3' or 3'/5' phosphates. (B) Chemical structures of common lesions recognized and excised by bifunctional DNA glycosylases. (C) List of common bacterial bifunctional DNA glycosylases and most prevalent lesions they excise. (D) Crystal structure of *Geobacillus stearothermophilus* formamidopyrimidine DNA glycosylase (Fpg) trapped in a covalent complex with abasic site-containing DNA (PDB: 1L1Z). Fpg glycosylase domain is in green, Zn-finger motif in yellow. DNA protein crosslink (DPC) between the N-terminal proline and the ring-opened AP site is shown in pink. The covalent intermediate was trapped by reducing the Schiff base intermediate with sodium borohydride.

Nucleotide excision repair (NER) pathway

The nucleotide excision repair (NER) pathway is another highly specialized and diverse DNA repair mechanism which is highly conserved across all species (Kisker *et al.*, 2013; Schärer, 2013). NER is distinct from BER in both the types of lesions or damage it can recognize, as well as the mechanism by which the DNA is repaired. In general, the

lesions recognized by NER are bulky, helix-distorting lesions which severely disrupt the duplex structure (UV crosslinks, intrastrand crosslinks, bulky alkylators, intercalating agents) (Gillet and Schärer, 2006). In bacteria, global genome NER is initiated by recognition of the damage by the $UvrA_2B_2$ complex which recognizes distorted duplex structures (Malta *et al.*, 2007; Kisker *et al.*, 2013) (Fig. 7A). UvrC is then recruited to the site of damage, and contains dual nuclease domains which perform dual incisions on the damaged strand within the “NER bubble”. This creates a stretch of 12-13 nucleotides which is removed by the action of the DpoI/UvrD helicase complex to generate a gapped DNA (Caron *et al.*, 1985; Moolenaar *et al.*, 1995; Verhoeven *et al.*, 2002; Kisker *et al.*, 2013) (Fig. 7A). Repair synthesis by DNA polymerase I and sealing by DNA ligase I repair the gapped DNA (Van Houten *et al.*, 1988; Kisker *et al.*, 2013) (Fig. 7A). In contrast with BER, NER requires larger repair complexes and recognizes substrates which are generally recalcitrant to base excision by DNA glycosylases due to their requirement for base-flipping.

Figure 7. *Overview of the nucleotide excision repair (NER) pathway* (A) Cartoon schematic of the nucleotide excision repair pathway in bacteria. From left: a bulky, helix-distorting lesion is recognized by the UvrABC excinuclease complex which performs dual incisions of a stretch of nucleotides on one strand containing the lesion. The UvrD helicase unwinds and removes this stretch of nucleotides to generate a gapped DNA structure (middle), which is filled in through repair synthesis by a DNA polymerase and DNA ligase.

DNA interstrand crosslink (ICL) repair

DNA interstrand crosslinks (ICLs) arise when bifunctional alkylating agents covalently modify both strands within the DNA duplex, resulting in complete inhibition of strand-separating processes such as replication and transcription (Deans and West, 2011; Huang and Li, 2013). Because ICLs modify both strands within the DNA, there is no undamaged strand that would be required for normal repair synthesis, as repair on one strand still leaves a monoadduct on the other. For this reason, ICL repair in bacteria requires coordination between multiple DNA repair pathways to ensure removal of the crosslink and to prevent mutations from arising, however ICL repair is an error-prone process (Noll *et al.*, 2006). For bacterial ICL repair, the most well-studied organism is *Escherichia coli* as many of the molecular and genetic mechanisms have been elucidated here. In general, ICL repair is primarily initiated by the NER UvrABC complex whereby dual incisions are performed on one side of the crosslink, and UvrD un-anneals the 3-stranded monoadduct to create a gap (Van Houten *et al.*, 1986; Cheng *et al.*, 1991; Noll *et al.*, 2006) (Fig. 8A). This gapped 3-stranded monoadduct can undergo two repair mechanisms: (1) if no homologous template is available, translesion synthesis (TLS) can occur (generally by Pol V) across this structure in a low-fidelity process, as the adducted nucleobase cannot be faithfully read by the polymerase, or (II) homologous recombination (HR) can occur if a homologous template is available (replicated chromosome, extrachromosomal plasmid) whereby the RecABCD proteins perform homology searching, strand invasion, and replication by a DNA polymerase/ligase (Cole, 1973; Yang, 2003; Noll *et al.*, 2006) (Fig. 8A). This 3-stranded monoadduct is a suitable substrate for a second round of NER by the UvrABCD system, which will excise out a

stretch of the crosslinked nucleotides (Sladek *et al.*, 1989; Noll *et al.*, 2006) (Fig. 8A). Repair synthesis by DNA polymerase I and sealing by DNA ligase I repair the gapped DNA (Noll *et al.*, 2006) (Fig. 8A). Since ICL repair potentially requires TLS across the 3-strand monoadduct, ICL repair is not considered an error-free pathway. The primary goal of ICL repair is to remove the crosslinked bases from the genome, even if it is at the expense of introducing mutations. However, there is evidence in *E. coli* of a recombination-independent pathway for ICL repair which involved DNA polymerase II, which implicates additional DNA repair pathways in the resolution of ICLs in bacteria (Berardini, 1997, 1999). The findings of additional ICL repair enzymes and cooperation between DNA repair pathways to faithfully repair ICL damage is critical to understanding how cells maintain genomic information despite complex types of damage.

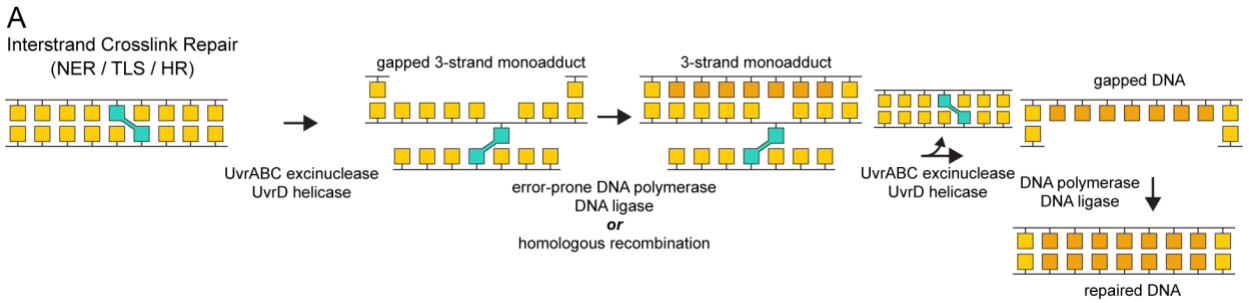


Figure 8. *Overview of bacterial DNA interstrand crosslink (ICL) repair* (A) Cartoon schematic of DNA interstrand crosslink (ICL) repair pathway in bacteria. From left: the UvrABC excinuclease complex recognizes the ICL and performs dual incisions of a stretch of nucleotides on one strand. The UvrD helicase unwinds this stretch of nucleotides to create a gapped 3-stranded monoadduct. This 3-stranded intermediate can then be bypassed by a low fidelity translesion polymerase (TLS) in an error-prone process. Alternatively, homologous recombination can be accomplished if a homologous template is available in an error-free process. Each of these steps produces a 3-stranded monoadduct (no gap), which then enters a second round of NER to liberate the crosslinked nucleotides and generates a gapped DNA (far right). This gap is filled in through repair synthesis by a DNA polymerase and DNA ligase.

Natural Product Genotoxins as a Source of DNA Damage

Bacteria are an exceptionally rich sources of secondary metabolites which affect nearly all aspects of cellular life to provide fitness, cooperation, or survival of cellular organisms (Pham *et al.*, 2019). These natural products are produced through the genetic organization of biosynthetic gene clusters (BGCs) which contain the genes necessary for the biosynthesis, regulation, transport, and resistance (if toxic) to the cognate metabolite. Bacteria produce diverse classes of natural products that modify cell signaling, cell wall biosynthesis, global metabolism, translation, and nucleic acids (Pham *et al.*, 2019; Scott and Piel, 2019). Natural products that target DNA for damage or alteration are termed genotoxins, and bacteria produce a large variety which affect all aspects of nucleic acid chemistry. In general, bacterial genotoxins can be classified as: (1) non-alkylating (no covalent modification to the nucleobase), and (2) alkylating (covalently modifies the bases within DNA). In this section, I describe the general classes and properties of both these classes of bacterial genotoxins.

Non-alkylating genotoxins

Genotoxins that do not alkylate DNA are highly diverse in their structures and chemical properties while interacting with DNA. These products can be relatively unreactive, or lead to secondary DNA damage through complex cleavage mechanisms. Non-alkylating genotoxins must be directed to DNA through some interaction with the nucleic acid structure, which is generally accomplished through intercalation into the duplex or threading within the major or minor grooves (de Almeida *et al.*, 2021). A general class of these genotoxins are strand cleavage agents (DNA molecular scissors) that do

not directly alkylate DNA but can catalyze the nicking of the DNA backbone (Fig. 9 A-B). Bleomycin is an FDA-approved cancer treatment that binds in the minor groove of DNA with its bis-thiazole moiety (Povirk *et al.*, 1979; Zhao *et al.*, 2002; Yu *et al.*, 2016) (Fig. 9A). Bleomycin is a non-heme iron polypeptide/polyketide that chelates Fe^{2+} ions to create a pseudoenzyme that reacts with molecular O_2 to generate superoxide and hydroxyl radicals (Hecht, 2000; Chen *et al.*, 2008). These ROS can abstract hydrogen atoms from the carbon atoms within the deoxyribose sugar, leading to oxidative damage to the DNA and ultimately SSBs and DSBs. Another strand cleavage agent that operates by a different mechanism is calicheamicin γ1 (Fig. 9B). Calicheamicin is a member of the enediyne family of antitumor/antibiotics and possesses a unique 10-membered ring which becomes activated for DNA attack (Nicolaou *et al.*, 1993). Calicheamicin threads the DNA minor groove and thiol-based activation of the trisulfide portion positions the enediyne system to undergo a Bergman cyclization to create a highly-reactive diradical benzene species (Zein *et al.*, 1988; Simkhada *et al.*, 2009). This diradical benzene is positioned to abstract hydrogen atoms from C1', C4', or C5' on the sugar, which then undergoes reaction with molecular O_2 to further oxidize and cleave the DNA (Walker *et al.*, 1992). Because of its potent antitumor properties, calicheamicin and other enediynes have been explored as cancer treatments (Shao, 2008).

Other types on non-alkylating genotoxins include the intercalating agents, whereby a planar aromatic ring system is suitable for insertion into the duplex to stack between the bases (Fig. 9 C-D). Intercalating agents can stabilize the duplex structure to prevent strand separation or access to the genetic information (Mukherjee and Sasikala, 2013). The intercalative properties of these agents can lead to secondary DNA damage or

inhibition of nucleic acid-processing enzymes (Beretta and Zunino, 2008). Aclacinomycin is a *Streptomyces* natural product which contains an anthracycline planar core system and aminosugars that interact within the groove of the duplex (Jensen *et al.*, 1991) (Fig. 9C). Intercalation by aclacinomycin can induce histone eviction from chromatin, disrupting the epigenomic state of the DNA (Pang *et al.*, 2013; Pang *et al.*, 2015). Doxorubicin is another *Streptomyces* natural product that is FDA-approved to treat various cancers (Rivankar, 2014) (Fig. 9D). Doxorubicin also contains an anthracycline core ring (intercalation) and aminosugars which bind within (thread) the minor groove, and leads to topoisomerase II inhibition by stabilizing the cleavage complex, thus leading to strand breakage (Frederick *et al.*, 1990; Pommier *et al.*, 2010). The diverse structural and functional properties of various types of non-alkylating bacterial genotoxins has led to tremendous advancements in the fields of discovering and developing novel antitumor and antibiotic agents.

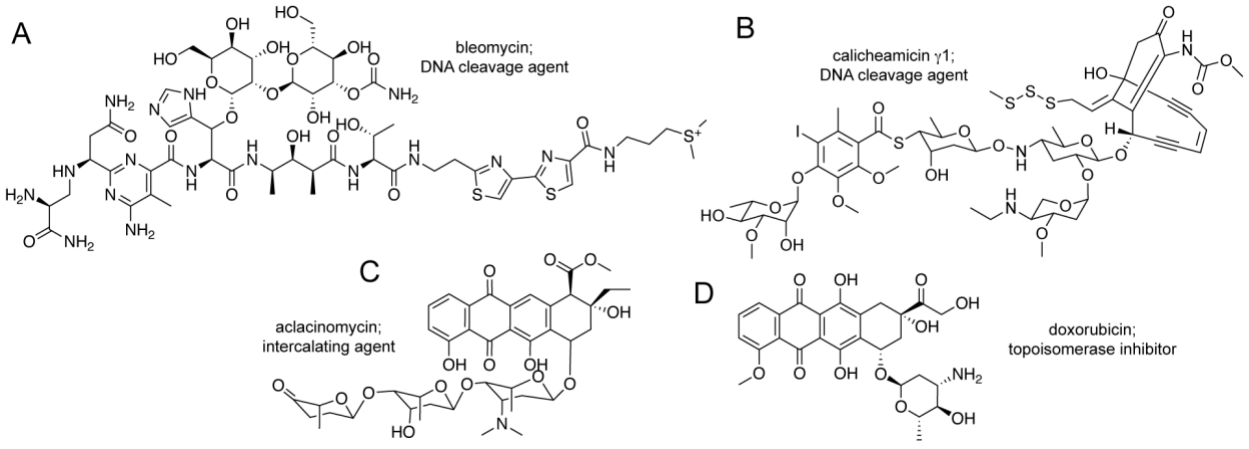


Figure 9. *Structures and properties of non-alkylating bacterial genotoxins* (A) Bleomycin is a non-heme iron-dependent DNA strand cleavage agent which generates free radical oxygen species that react with DNA. (B) Calicheamicin γ1 is an enediyne antitumor agent which binds to DNA and can abstract hydrogen atoms from the deoxyribose backbone through an intermediate diradical benzene (C) Aclacinomycin is a DNA intercalating agent that stalls replication and transcription machinery. (D) Doxorubicin is also an intercalating agent which acts as a topoisomerase inhibitor to generate SSBs and DSBs.

Alkylating genotoxins

Bacterial genotoxins that covalently modify the heteroatoms within the nucleobases are known as alkylating agents (addition of a carbon moiety). Similar to non-alkylating genotoxins, these families are known to possess highly diversified biosynthetic scaffolds and reactive groups to modify the DNA. Many of these compounds are either currently FDA-approved for certain treatments or are under investigation for their potential as therapeutics. In general, alkylating genotoxins can be classified as either: (1) monoalkylating agents (attachment to one strand of DNA), or (2) bifunctional alkylating agents (modify multiple bases of DNA to create intra- and interstrand crosslinks). Monoalkylating agents are highly diverse in their mechanisms of action and sites of DNA where alkylation takes place. Streptozotocin (STZ) is a methylating agent approved to treat pancreatic cancer and contains an *N*-methyl-*N*-nitrosourea derivative attached to glucose (Szkudelski, 2001) (Fig. 10A). STZ undergoes decomposition to form an electrophilic methyldiazonium ion that reacts at the N7 and O6 atoms of guanine and the N1/3/7 atoms of adenine to form methylated DNA adducts (Murata *et al.*, 1999). Leinamycin is a macrolactam with antitumor/antibiotic properties and contains a unique spiro-fused 1,3-dioxo-1,2-dithiolane moiety that generates DNA adducts and strand breaks (Hara *et al.*, 1990; Viswesh *et al.*, 2010) (Fig. 10A). Leinamycin binds in the minor

groove and undergoes a thiol-based activation mechanism whereby a highly-reactive 3-membered episulfonium ring system is generated, and alkylation at N7 of guanine occurs. This leinamycin-guanine adduct is rapidly depurinated to remove the guanine from the genome and generate toxic AP sites (Nooner *et al.*, 2004). A third type of monoalkylating bacterial genotoxin is derived from the spirocyclopropylcyclohexadienone (SCPCHD) family of natural products such as yatakemycin, CC-1065, and duocarmycin (Boger and Johnson, 1995; Igarashi *et al.*, 2003) (Fig. 10A). Yatakemycin (YTM) is an exceptionally potent antitumor/antibiotic that threads into the minor groove of DNA and alkylates at N3 of adenine in particular sequences through cyclopropane ring opening (Parrish *et al.*, 2003; Mullins *et al.*, 2017). YTM adducts are highly stable and prevent strand separation through stapling of the minor grooves with extensive C-H-π interactions (Mullins *et al.*, 2017).

The second, much smaller class of alkylating genotoxins in bacteria is comprised of the bifunctional alkylating agents that contain two electrophilic portions on the molecule capable of reacting with multiple nucleobases in DNA to form intrastrand (same DNA strand) or interstrand (both DNA strands) crosslinks (Schärer, 2005). ICL agents are capable of exhibiting cytotoxicity towards prokaryotes with less than 100 crosslinks per cell, making them extremely cytotoxic at low doses (Noll *et al.*, 2006). In *Streptomyces*, which are sources of a significant amount of antitumor/antibiotics used clinically, there are only two known families of bifunctional ICL agents: the mitomycins and the azinomycins (Fig. 10B). The mitomycin (MM) family is composed of mitomycin A, B, and C, with MMC being the most well-studied, and FDA-approved to treat pancreatic and stomach cancers (Tuinmann *et al.*, 2004; Lee *et al.*, 2006) (Fig. 10B). Mitomycin C is

composed of a quinone ring, a carbamate moiety, a methoxy portion, and an aziridine ring, all of which are important for the complex mechanisms of DNA crosslinking (Tomasz, 1995). The quinone portion of MMC can intercalate into the duplex, while the additional regions are positioned within the minor groove (Tomasz *et al.*, 1987; Norman *et al.*, 1990). The first step of activation occurs when the quinone portion is reduced to generate a hydroquinone (Tomasz, 1995). Oxidative elimination of methanol and reductive ring opening of the aziridine portion generates the first DNA-reactive semiquinone species (Tomasz, 1995). Alkylation then occurs at the exocyclic N2 amine of guanine, followed by oxidative elimination of the carbamate moiety to generate the second electrophilic MMC species (Tomasz, 1995). A second alkylation reaction at N2 of guanine occurs either within the same DNA strand (intrastrand crosslink) or on the opposing strand (interstrand crosslink) (Tomasz, 1995). The second family of ICL agents in bacteria is the azinomycins (A and B), which form N7-purine/N7-purine interstrand crosslinks in selective DNA sequences with their reactive epoxide and aziridine warheads (Fujiwara *et al.*, 1999; LePla *et al.*, 2005). Crosslinking by azinomycin B (AZB) begins when the methyl methoxynapthoate portion binds within the major groove of DNA to direct the warheads for alkylation (Alcaro and Coleman, 2000). The first alkylation occurs when the aziridine ring is attacked by N7 of purines, leading to monoalkylation (Salvati, 1992; Alcaro and Coleman, 2000). The second alkylation step occurs at the epoxide ring, when a second N7 on a purine in the opposing strand (1,3 spacing) opens the ring to crosslink the DNA (Salvati, 1992; Alcaro and Coleman, 2000). The N7-purine/N7-purine ICLs generated by AZB are relatively unstable and prone to depurination, which creates opposing AP sites and leads to DSBs in cells treated with AZB (Armstrong, 1992; Salvati, 1992; Fujiwara *et*

al., 1999; Kelly *et al.*, 2006). For this reason, AZB is a potent antitumor agent, but its applications are relatively understudied due to the inherent instability of the molecule. This diversity in alkylating bacterial genotoxins forms the basis for continued exploration of natural products and BGCs which produce novel, useful biomolecules.

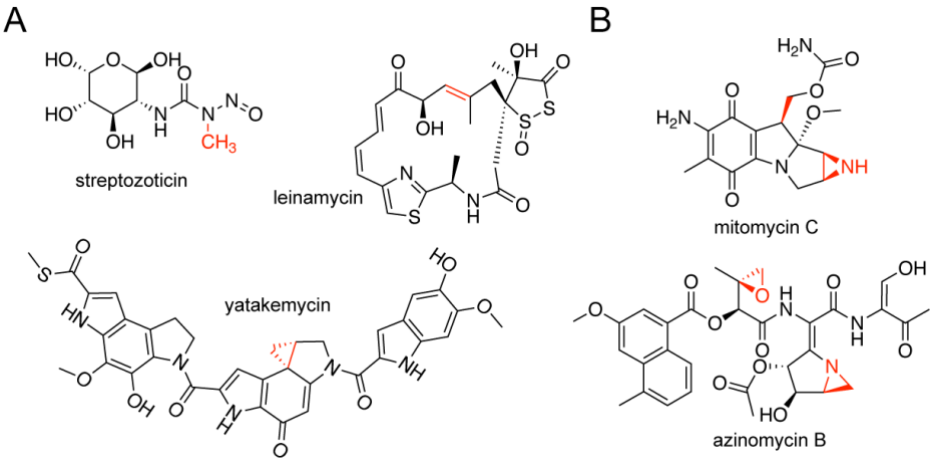


Figure 10. *Structures and properties of alkylating bacterial genotoxins* (A) Structures of monoalkylating genotoxins from bacteria. From top left: streptozotocin-methylating agent, leinamycin-N7 guanine alkylator, yatakemycin-N3 adenine alkylator. Reactive portions of the molecules are highlighted in red. (B) Structures of bifunctional alkylating agents from bacteria. From top: mitomycin C-N2 guanine-N2 guanine ICL agent, azinomycin B-N7 guanine-N7 guanine ICL agent.

Bacterial Antimicrobial Resistance Mechanisms

Bacteria that produce cytotoxic natural products must be equipped with inherent self-resistance mechanisms to combat the target of the compound(s) they produce (Hopwood, 2007). Many of these secondary metabolites are incredibly potent molecules with activities in the sub-nanomolar concentration range, so rapid and complete repair or resistance to damage that arises through exposure to these agents is paramount. In addition to self-resistance, many genetic and evolutionary processes can occur to transfer resistance mechanisms between bacteria or act through selective mutational pressures

(Boolchandani *et al.*, 2019). This phenomenon forms the basis for antimicrobial resistance (AMR) in both producing-organisms and those exposed to the specific natural product or antibiotic. In this section I will elaborate on the general molecular mechanisms of antibiotic resistance in bacteria, and expand on specific resistance mechanisms related to DNA glycosylase-mediated base excision repair (BER) of alkylating genotoxins.

General mechanisms of antibiotic resistance

Antimicrobial resistance (AMR) in bacteria is an extremely important subject to understand, as the emergence of resistant bacterial strains has posed a significant challenge for treating infections across the world (Prestinaci *et al.*, 2015). AMR can be a complex system of cooperating pathways which operate synergistically to provide cellular resistance to antibiotics, and many of these AMR phenomena are still currently being characterized at the genetic and molecular level (Blair *et al.*, 2015). To properly understand AMR in bacteria, an appreciation for the diversity of molecular targets of antibiotics is in order. Antibiotics (natural or synthetic) target critical cellular processes that when disrupted, lead to cytotoxicity and cell death (Boolchandani *et al.*, 2019). Common antibacterial targets include: folic acid synthesis (one-carbon metabolism), cell wall biosynthesis, protein translation by the 50S/30S ribosome, and nucleic acids (DNA, RNA, nucleic acid processing enzymes) (Fig. 11A). Each target of these antibiotics is generally characterized by the classes of antibiotics which targets them (sulfonamides-folate metabolism, β-lactams-cell wall metabolism, tetracyclines-protein synthesis, and quinolones-topoisomerase inhibitors) (Fig. 11A). Due to the extreme diversity in

antibiotics and their molecular targets in bacteria, nature has evolved efficient resistance mechanisms to combat the actions of these compounds.

The molecular mechanisms of AMR in bacteria fall into several general classes, although new mechanisms, pathways, and enzymes are still commonly being discovered. I will not delve deeply into the genetic mechanisms of AMR, which generally include: changes in gene expression of the target, horizontal gene transfer of resistance plasmids, duplication events of the target, and the most commonly known- mutations (point mutations, insertions, deletions) within the target protein of the antibiotic (Peterson and Kaur, 2018). Instead, here I will elaborate on the molecular mechanisms involving AMR, which generally involve enzymes, transporters, and antibiotic sequestration (Munita and Arias, 2016) (Fig. 11B). For self-resistance mechanisms, the gene encoding the resistance protein is located either within the host genome, or on an extrachromosomal plasmid (Fig. 11B). A common mechanism of host AMR is through sequestration of the drug through drug-binding proteins, or through active efflux through a transmembrane pump (Blair *et al.*, 2015) (Fig. 11B). This process controls the concentration of the drug within the cell, and limits the exposure of the target molecule from the drug. Some resistance proteins encode for enzymes which bind to and either sanitize, modify, or degrade the antibiotic which prevents it from interaction with the target (Blair *et al.*, 2015) (Fig. 11B). Additional mechanisms exist such as expression of a target protection protein which binds to the antibiotic target and prevents the interaction with the compound (Tomlinson *et al.*, 2016; Wilson *et al.*, 2020). Importantly, many of these AMR mechanisms are not sufficient on their own to provide full cellular resistance to an antibiotic, and as such many of these pathways operate simultaneously or redundantly.

Finally, AMR remains a burgeoning field where both the delicate intricacies of the system continue to be dissected, and more undiscovered mechanisms remain unexplored. These final sections are dedicated to exploring AMR mechanisms specific to BER and DNA glycosylases.

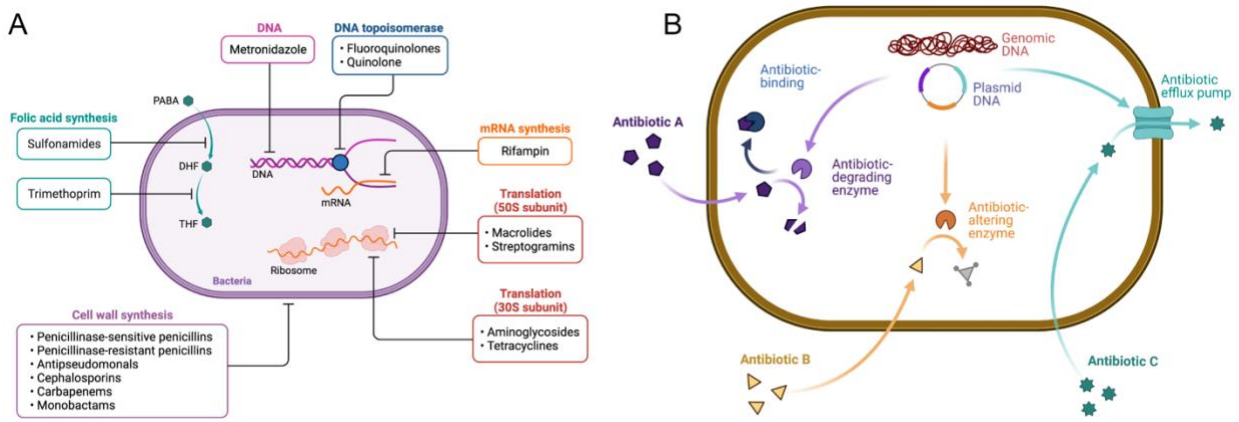


Figure 11. *Molecular targets of antibiotics and molecular mechanisms of antibiotic resistance in bacteria* (A) Cartoon diagram of common molecular targets for antimicrobial agents and representative antibiotics in each category. (B) Cartoon schematic of the major antibiotic resistance pathways in bacteria. Beginning with antibiotic A (purple), antibiotic-degrading enzymes or antibiotic-binding proteins can destroy/sequester the drug to protect the target of the antibiotic. Antibiotic B (light yellow) depicts antibiotic sanitization or deactivation through an enzyme which alters the antibiotic. Antibiotic C (blue-green) is exported through an antibiotic efflux pump to prevent it from interaction with cellular targets. Many of these mechanisms can cooperate to provide full resistance. These figures were produced in BioRender.

Antibiotic mechanisms specific to base excision repair (BER)

Until relatively recently, the association of a specific DNA repair enzyme or pathway with a particular natural product or BGC had been understudied and underappreciated. However, within the past decade, several BGCs that produce potent DNA alkylating agents have been demonstrated to contain novel DNA repair enzymes that are specific for their natural product DNA damage (Xu *et al.*, 2012; Wang *et al.*, 2016;

Ng *et al.*, 2019; Chen *et al.*, 2020). In at least two of these BGCs, the producing organisms encode for specialized DNA glycosylases that trigger the BER system to remove the natural product alkylation damage from the genome to provide self-resistance to the genotoxin (Xu *et al.*, 2012; Wang *et al.*, 2016). The structure and mechanistic properties of these enzymes, and their diverse substrate repertoire has led to significant advancements for the BER field. In these last two sub-sections, I will describe the genetics, biochemistry, and structural properties of DNA glycosylase-mediated BER, which provides self-resistance to yatakemycin (YTM) and the interstrand crosslinking agent azinomycin B (AZB).

Resistance to yatakemycin (YTM) by AlkD/YtkR2

Yatakemycin (YTM), as described in an earlier section, is a member of the SCPCHD family of antitumor/antibiotics and is an exceptionally potent DNA alkylating agent (Parrish *et al.*, 2003). YTM is produced by *Streptomyces sp.* TP-A0356 from the 36 kb YTM BGC (Fig. 12A), and has been shown to require several resistance mechanisms to allow for full biosynthesis of YTM (Xu *et al.*, 2012; Yuan *et al.*, 2017). YTM functions by first threading the tryptophan-derived subunits into the minor groove of DNA with extensive C-H-π interactions which effectively staple the strands together (Mullins *et al.*, 2017). DNA binding by SCPCHD family members has been shown through NMR studies to twist the subunits for optimal positioning of the cyclopropane warhead to undergo a reductive alkylation at N3 of adenine in specific AT-rich sequences (Bassarello *et al.*, 2003). The lone pairs on N3-adenine attack the cyclopropane ring, and these ring electrons are delocalized into the cyclohexadieneone system where reduction occurs

(Fig. 12B). This generates the YTM-Ade-DNA (YTMA-DNA) adduct which is remarkably stable and toxic (Igarashi *et al.*, 2003; Mullins *et al.*, 2017). A gene within the cluster (*ytkR2*) was shown to have homology to a related family of newly discovered DNA glycosylases, the AlkC/AlkD superfamily of HEAT-like repeat enzymes (Alseth *et al.*, 2006; Xu *et al.*, 2012). YtkR2 was shown to be essential for YTM biosynthesis and for survival of cells exposed to YTM (Xu *et al.*, 2012). Subsequent experiments demonstrated that YtkR2 functions as a DNA glycosylase that excises YTMA-DNA adducts from the genome to generate an AP site and the deglycosylated YTM-adenine nucleobase (Xu *et al.*, 2012) (Fig. 12B). The structural basis for how this DNA glycosylase can recognize and excise a bulky, helix-stapling adduct that is almost certainly refractory to base-flipping was thus a highly important avenue of research. The well-characterized homolog of YtkR2, the AlkD glycosylase from *Bacillus cereus*, was demonstrated to possess excision activity for bulky adducts and for YTMA-DNA adducts, and subsequent structural studies demonstrated that AlkD does not require base-flipping to perform base excision but rather pre-scaffolds the catalytic residues for base excision within an intact duplex (Rubinson *et al.*, 2008; Rubinson *et al.*, 2010; Mullins *et al.*, 2015). AlkD was co-crystallized with YTMA-DNA and the structure revealed a very tight association between the C-shaped HEAT-like repeats and the damaged DNA (Mullins *et al.*, 2017) (Fig. 12C). Expansion of the active site of AlkD in complex with YTMA-DNA revealed the enzyme also uses extensive C-H-π interactions from aromatic residues which allow prying open of the minor groove and hydrolysis of the adduct which is catalyzed by Asp112 (Parsons *et al.*, 2016; Mullins *et al.*, 2017) (Fig. 12C). This pre-formed active site and prying of the minor groove forms the structural basis for excision of a bulky lesion without the need for base-flipping. The

evolutionary diversity of YtkR2 homologs in SCPCHD-producing BGCs is not constrained to the YTM cluster, but has been found to play an important role in self-resistance and excision of the CC-1065 and duocarmycin analogs (Mullins *et al.*, 2021) (Fig. 11D). The mechanistic and structural insights provided by the AlkD/YtkR2 glycosylases and involvement in bulky YTMA-DNA repair have been valuable for understanding how these and other glycosylases function at the molecular and cellular level.

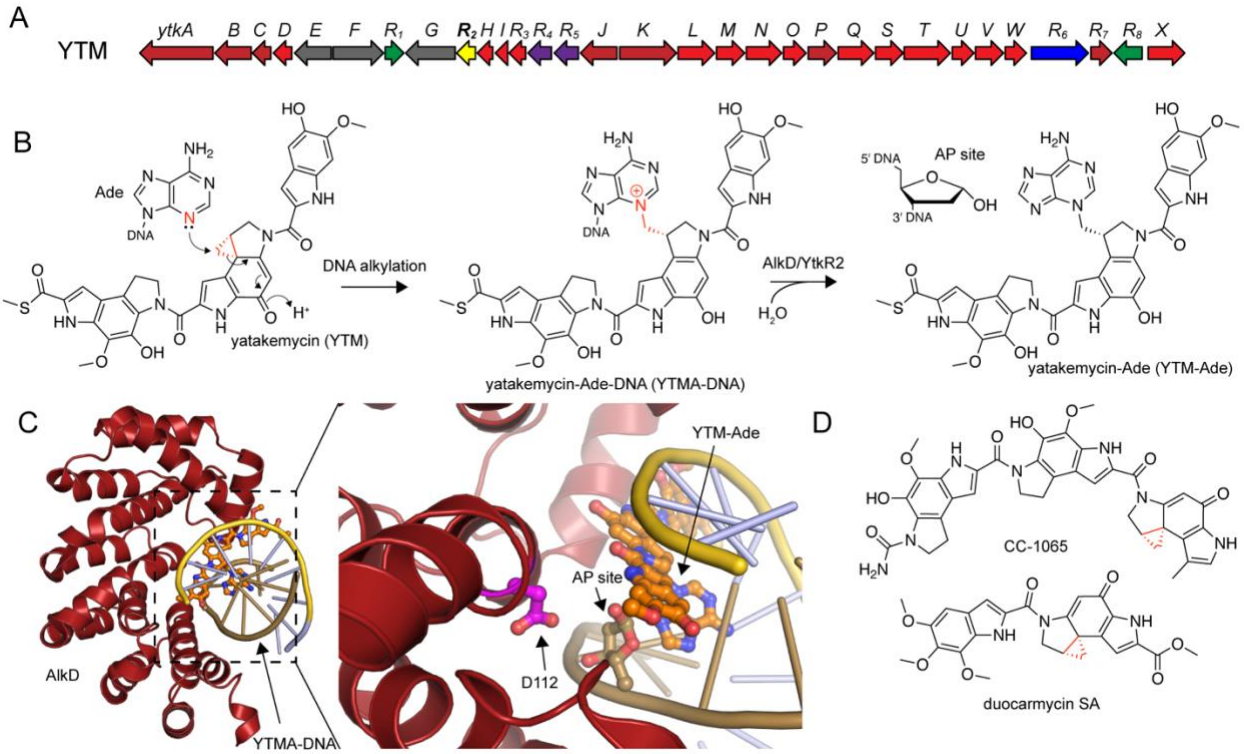


Figure 12. *Self-resistance to yatakemycin (YTM) through base excision repair* (A) Gene cluster diagram for yatakemycin (YTM) biosynthesis. Genes are colored according to their published or predicted functions. Dark red-core biosynthetic genes, light red-accessory biosynthetic genes, green-regulatory genes, blue-transport genes, purple-predicted endonucleases, yellow-YtkR2 resistance DNA glycosylase, gray-unknown. (B) Chemical reaction between YTM and N3 of adenine in AT-rich sequences. Lone pair of N3 on adenine attacks the cyclopropane ring which leads to alkylation. The YTMA-DNA adduct is recognized and excised by the host YtkR2 glycosylase, as well as the YtkR2 homolog AlkD, which liberates the YTM-Ade adduct and generates an AP site. (C) Crystal structure of AlkD in complex with YTMA-DNA ternary product (PDB: 5UUG). AlkD is composed of HEAT repeats (red) which scaffold around the YTMA-DNA. A zoom-in of the active site (inset) shows the catalytic Asp112 in proximity to the hydrolyzed YTM-Ade adduct, without base flipping into the AlkD active site. (D) Chemical structures of related spirocyclopropylcyclohexadienones which contain a YtkR2 homolog that contributes to base excision repair-mediated self-resistance.

An essential DNA glycosylase in azinomycin B (AZB) biosynthesis

The discovery of the YtkR2 and related AlkD DNA glycosylases involved in self-resistance to YTM toxicity through BER opened the possibilities of other BGCs producing genotoxins that contain unique DNA glycosylases specific to their natural products. Indeed, this was demonstrated to be the case for the azinomycin B (AZB) cluster where an essential DNA glycosylase was discovered (Wang *et al.*, 2016). AZB is produced by *Streptomyces sahachiroi* and *griseofuscus* from the 64 kb AZB cluster (Nagaoka *et al.*, 1986; Zhao *et al.*, 2008) (Fig. 13A). As described in a previous section, AZB is a non-ribosomal peptide/polyketide natural product that contains two electrophilic aziridine and epoxide warheads, which alkylate both strands of DNA to create a 1,3 N7-purine/N7-purine ICL (one base pair separates the ICL) (Armstrong, 1992; Fujiwara *et al.*, 1999; LePla *et al.*, 2005) (Fig. 13B). These AZB-ICLs are exceptionally toxic to both prokaryotes and eukaryotes and induce a strong transcriptional response involving DNA repair pathways, making AZB a candidate for antitumor use (Ishizeki *et al.*, 1987; Kelly *et al.*,

2006). While studying the biosynthesis of AZB, it was noted a gene within the cluster (*azi36*/*orf1*-eventually renamed AlkZ) contained predicted winged-helix DNA binding domains and was proposed to play a role in transcriptional regulation of the cluster (Wang *et al.*, 2016). Surprisingly, the *alkZ* gene in *S. sahachiroi* proved to be essential for the organism, as all attempts to knock it out were unsuccessful (Wang *et al.*, 2016) (Fig. 13C, top). *AlkZ* could only be deleted from the cluster through a double homologous recombination crossover, which introduced a second copy into the genome, and created an AZB-resistance strain with *alkZ* deleted in the cluster (Wang *et al.*, 2016) (Fig. 13C, middle). *AlkZ* could only be fully deleted from the genome by first deleting the essential AZB biosynthesis gene (*azi29*/*aziU3*) and then deletion of *alkZ*, which created an AZB-sensitive strain (Wang *et al.*, 2016) (Fig. 13C, bottom). Subsequent experiments demonstrated that AlkZ functioned as a DNA glycosylase that can unhook AZB-ICLs from either strand of the DNA, which generates 1,3 double AP sites, although the precise mechanism for how this is performed was unknown (Wang *et al.*, 2016) (Fig. 13B). This was the first example of a prokaryotic DNA glycosylase that can unhook DNA-ICLs and the first example of an endogenously essential DNA glycosylase in nature (Wang *et al.*, 2016). The mechanism, structure, and evolutionary breadth of AlkZ proteins were entirely uncharacterized, but it was observed that AlkZ homologs are prevalent in diverse bacterial phyla and are conserved (Wang *et al.*, 2016). AlkZ and the enormously diverse homologs of this protein made it an extremely interesting protein to characterize for its novel properties as a DNA glycosylase.

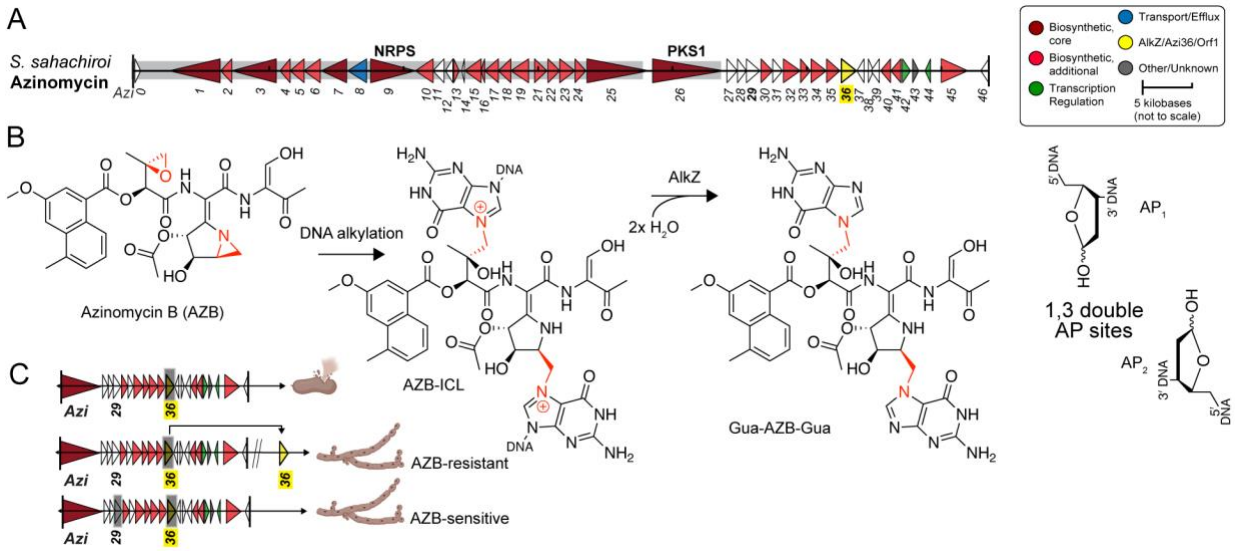


Figure 13. *AlkZ is an essential DNA glycosylase which unhooks azinomycin B (AZB) ICLs* (A) Gene cluster diagram for azinomycin B biosynthesis. Genes are colored according to the key to the right. (B) Chemical reaction of epoxide and aziridine rings (red) in AZB with N7 of guanines to create a 1,3 AZB-ICL. The essential gene encoding for AlkZ/Azi36/Orf1 catalyzes the unhooking of the AZB-ICL to create opposing abasic sites (1,3 AP sites). (C) Genetic diagram of the essential nature of AlkZ. Gene deletions are denoted with a gray box. From top: *ΔalkZ* (*Δazi36*) knockout cells in *Streptomyces sahachiroi* are lethal due to the production of AZB. If a second copy of AlkZ is inserted into the genome at a different chromosomal location, then AlkZ can be deleted within the AZB cluster to generate an AZB-resistant strain. If the biosynthesis of AZB is abrogated by first deleting *aziU3* (*azi29*), then *alkZ* (*azi36*) can be deleted to generate an AZB-sensitive strain. Figure C was produced with BioRender.

[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. Figures 4, 7, and 8 in this thesis were modeled off Figure 1 in the manuscript. I generated models and data for Figure 4 and helped write the section on AlkZ-mediated repair of AZB-ICLs. I also contributed to the editing of the manuscript as a whole.

These findings of the AlkZ DNA glycosylase involved in self-resistance to AZB-ICLs form the foundations for my thesis work in characterizing the structural, biochemical, cellular, and genetic basis of AlkZ-mediated BER of ICL agents and other genotoxic natural products. The remainder of this dissertation will focus on (mainly) published work involving AlkZ and related DNA glycosylases involved in diverse and novel roles for BER. Chapter 2 will focus on the structural and biochemical characterization of *S. sahachiroi* AlkZ and models of AZB-ICL excision. In Chapter 3, I identify a homolog of AlkZ in *E. coli* named YcaQ that is distantly-related to AlkZ and unhooks diverse ICL substrates to provide cellular resistance through a new ICL repair pathway. Chapter 4 expands on the phylogenetic and evolutionary differences between YQL and AZL proteins in *Streptomyces* and includes genome mining analyses to identify clusters associated with these proteins. Chapter 5 is a collaborative effort whereby we further apply our genome mining data from Chapter 4 to characterize four DNA glycosylases within the BGCs of trioxacarcin A (TXNA) and LLD genotoxic alkylators which perform exquisitely precise base excision of their cognate lesions and are involved in the biosynthesis and self-resistance of these natural products. Finally, I conclude with Chapter 6 where I present preliminary data involving structural modeling, crystallization efforts, genomics, and biochemistry of YcaQ and AlkZ family members. I finish Chapter 6 by expanding on the future directions and implications of this project as it continues to progress forward. All references are listed at the end of the dissertation, and publication references in which I am a co/first-author are marked at the end of each chapter with details on which experiments I conducted.

Chapter 2

Structure of a DNA Glycosylase that Unhooks Interstrand Cross-links[2]

Introduction

The chemical integrity of DNA is constantly challenged by cellular and environmental genotoxic agents that produce a diverse array of covalent nucleobase adducts. If left unrepaired these DNA lesions impair important cellular processes including replication, transcription, and cell cycle regulation, leading to mutations, chromosomal rearrangements, and genomic instability that threaten the livelihood of the organism and lead to human diseases (Jackson and Bartek, 2009). Small adducts are typically removed by base excision repair (BER), in which a lesion-specific DNA glycosylase catalyzes hydrolysis of the N-glycosidic bond of the aberrant nucleobase, liberating it from the phosphoribose backbone (Krokan and Bjoras, 2013). The resulting abasic site is nicked at the 5′ side by an apurinic/apyrimidinic (AP) endonuclease to create a free 3′-hydroxyl group necessary for synthesis of new DNA. In contrast, bulky and helix distorting lesions are removed by nucleotide excision repair (NER), which involves removal of an oligonucleotide segment containing the lesion through the action of dual nuclease incisions flanking the lesion followed by helicase removal of the damaged segment.

Interstrand crosslinks (ICLs) from a number of endogenous and environmental sources covalently tether the opposite strands of DNA and pose a major obstacle to normal DNA metabolism (Schärer, 2005; Noll *et al.*, 2006). The cytotoxicity of crosslinking agents makes them particularly effective as antitumor drugs (Rajski and Williams, 1998).

Because ICLs damage both strands of DNA, repair of ICLs is more elaborate than repair of monoadducts and involves multiple pathways (Noll *et al.*, 2006; Clauson *et al.*, 2013). In ICL repair, nuclease-dependent dual incisions on one strand flanking the lesion produce a gapped intermediate that must be filled by translesion synthesis (TLS) or homologous recombination (HR) prior to NER-associated repair of the remaining unhooked monoadduct.

Azinomycin B (AZB, Fig. 14A) is a cytotoxic, nonribosomal peptide-polyketide secondary metabolite of *Streptomyces sahachiroi* and *Streptomyces griseofuscus* (Hata *et al.*, 1954; Nagaoka *et al.*, 1986). This bifunctional alkylating agent forms ICLs *in vitro* and *in vivo* (Terawaki and Greenberg, 1966) and displays potent antibiotic and antitumor activities comparable to the chemotherapeutic mitomycin C (Shimada *et al.*, 1955; Nagaoka *et al.*, 1986; Ishizeki *et al.*, 1987). AZB ICLs are formed between the electrophilic aziridine and epoxide functional groups of AZB and the N7 nitrogens of guanosine and adenosine within d(GNC) and d(GNT) sequences (Fig. 14B) (Armstrong, 1992; Fujiwara *et al.*, 1999). Computational modeling of the AZB ICL structure is consistent with the entire AZB molecule residing in the major groove of the DNA without nucleobase intercalation by the naphthoate moiety (Alcaro and Coleman, 2000; Coleman, 2002).

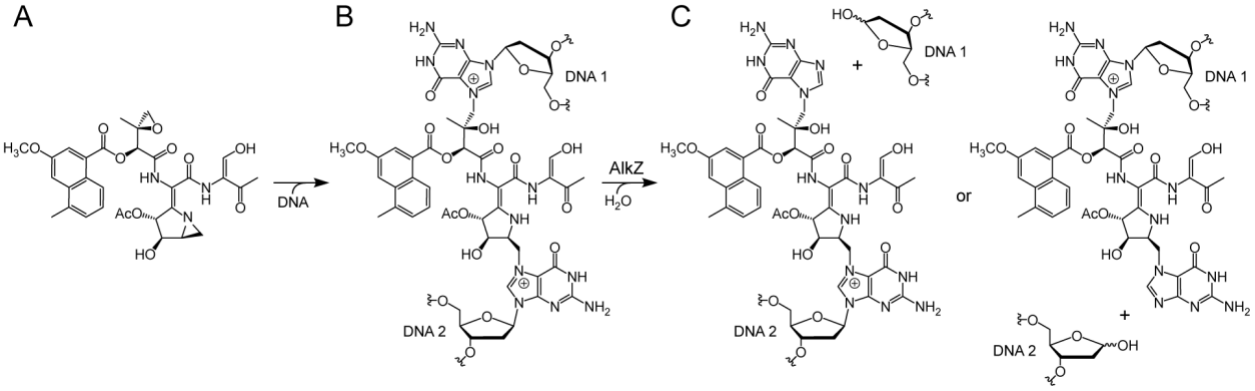


Figure 14. *Unhooking of azinomycin B interstrand crosslinks by AlkZ* (A) Chemical structure of azinomycin B (AZB). **b.** AZB crosslinks opposing purines at their N7 nitrogens via reactive aziridine and epoxide groups. **c.** AlkZ glycosylase catalyzes hydrolysis of the N-glycosidic bond of either crosslinked dG to produce two possible dG-AZB-Gua monoadducts.

Contrary to NER-associated repair of other ICLs, AZB ICLs were recently identified as the target of repair by a DNA glycosylase (Wang *et al.*, 2016). The *azi36 (orf1)* gene sits adjacent to the AZB synthesis cluster in *S. sahachiroi* (Zhao *et al.*, 2008) and provides self-resistance to cells from AZB toxicity. The Orf1 gene product unhooks AZB ICLs by cleaving the N-glycosidic bond on at least one side of the lesion to produce an abasic site that is recognized by the bacterial AP endonuclease, EndoIV (Fig. 14C) (Wang *et al.*, 2016). Orf1 belongs to the HTH_42 (Pfam 06224) superfamily of uncharacterized winged-helix containing proteins that exist in a number of pathogenic and antimicrobial producing bacteria. However, there are no structures of any HTH_42 protein and thus the molecular rationale for glycosylase repair of AZB ICLs is unknown.

The significance of this alternative, glycosylase-mediated ICL repair pathway is underscored by the recent discovery that the unrelated eukaryotic NEIL3 glycosylase unhooks psoralen and abasic site-adenine ICLs during DNA replication (Semlow *et al.*, 2016) and that human NEIL1 can excise unhooked psoralen monoadducts and ICLs (Couvé-Privat *et al.*, 2007; Couvé *et al.*, 2009). Thus, glycosylase mediated ICL repair activity has been identified in both prokaryotes and eukaryotes by two distinct proteins. However, the molecular details for how either of these enzymes can excise adducts covalently tethered to both DNA strands is unknown. To better understand the mechanisms for DNA glycosylase-mediated ICL repair, we determined the crystal

structure of the *S. sahachiroi azi36* gene product and identified the active site by mutational analysis. The structure is the defining member of the HTH_42 superfamily and an eighth structural class of DNA glycosylase. Our analysis shows how the HTH_42 architecture can support unhooking of either side of an AZB ICL or excision of an AZB-dG monoadduct. To better reflect the activity of the enzyme for base excision of alkylated DNA, we propose to rename the Azi36/Orf1 protein to AlkZ.

Results

AlkZ excises N7-methylguanine from DNA

The AlkZ protein was shown recently to act on AZB ICLs and monoadducts through hydrolysis of one or more of the N-glycosidic bonds (Wang *et al.*, 2016), but it was unclear whether the enzyme would act on adducts other than AZB. DNA glycosylases that excise alkylated DNA lesions often exhibit activity for a number of N3- and N7-methylated purines in addition to their preferred substrates (O'Brien and Ellenberger, 2004; Rubinson *et al.*, 2010). For example, *Bacillus cereus* AlkD displays robust activity for bulky minor groove N3-yatakemycinyldeoxyadenosine lesions, but also cleaves N3-methyldeoxyadenosine and N7-methyldeoxyguanosine (d7mG) (Mullins *et al.*, 2013; Mullins *et al.*, 2015). I therefore tested the ability of AlkZ to catalyze excision of N7-methylguanine (7mGua) from an oligonucleotide containing d7mG (Fig. 15A). Using a standard *in vitro* assay that measures glycosylase activity by alkaline hydrolysis of the resulting abasic site product, I found that AlkZ hydrolyzed d7mG with an observed rate (k_{obs}) of $(8.5\pm0.4)\times10^{-5}$ s^{-1} (Fig. 15B-C). This d7mG activity is modest compared to that of

other alkylpurine DNA glycosylases under similar conditions (O'Brien and Ellenberger, 2004, 2004; Parsons *et al.*, 2016), but demonstrates that an AZB adduct is not necessary for AlkZ activity and corroborates the previous data showing activity toward monoadducts (Wang *et al.*, 2016). Additionally, I found that alkaline conditions or treatment with EndoIV was required to nick the abasic DNA produced by AlkZ, indicating that AlkZ is a monofunctional DNA glycosylase that does not possess lyase activity (Fig. 15D).

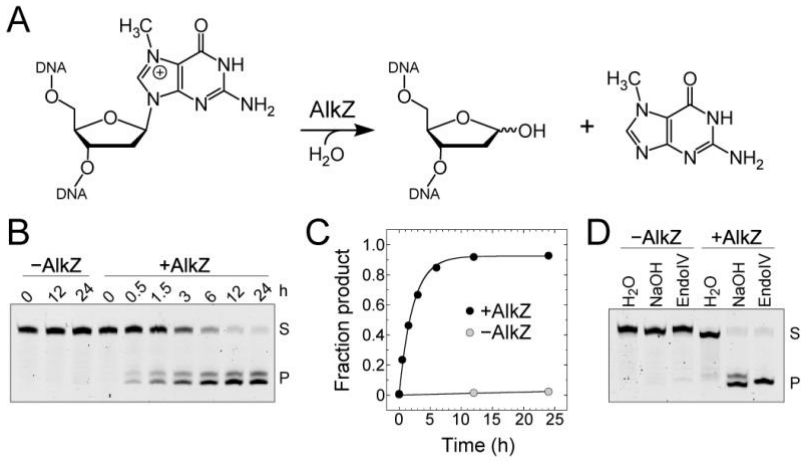


Figure 15. *Excision of N7-methylguanine by AlkZ* (A) Base excision of d7mG by AlkZ. (B) Denaturing PAGE of d7mG-DNA substrate (S) and nicked abasic-DNA product (P) after treatment without (−) or with (+) AlkZ for the specified time followed by alkaline hydrolysis. (C) Quantification of the gel shown in panel A, averaged across three independent experiments. (D) Denaturing PAGE of d7mG-DNA substrate (S) and nicked abasic-DNA product (P) after treatment without (−) or with (+) AlkZ for 24 hrs followed by treatment with either water, NaOH, or EndoIV.

AlkZ adopts a novel DNA-binding fold

HTH_42 superfamily proteins are predicted to contain two winged helix (WH) motifs, which are common helix-turn-helix (HTH) related DNA-binding elements found in transcription factors and DNA processing enzymes (Brennan, 1993). The canonical WH fold contains a three-helix bundle at the N-terminus and a three-stranded β-sheet (wing)

at the C-terminus, and is related to the helix-turn-helix sequence-specific DNA-binding motif in that α3 serves as a “recognition helix” by inserting into the DNA major groove (Gajiwala and Burley, 2000). However, there is no known structure of an HTH_42 protein, and thus it is not known how or if the putative WH motifs bind DNA. Moreover, there are no WH motifs in any of the six known DNA glycosylase protein folds (Brooks *et al.*, 2013), so the molecular details for how AlkZ acts as a DNA repair enzyme are not clear.

To understand the structural basis for DNA glycosylase-mediated ICL repair by AlkZ, we determined the crystal structure of the *S. sahachiroi* enzyme. Experimental X-ray phases were determined from single-wavelength anomalous dispersion using selenomethionyl-substituted protein (Table 3). The resulting crystallographic model, which contains all 371 residues of the protein, was refined against X-ray diffraction data extending to 2.3 Å, resulting in an R-factor of 15.7% and an R_{free} of 20.1% (Table 3). The protein adopts a cluster of three tandem WH motifs (WH1, WH2, WH3) that pack against a C-terminal β-barrel subdomain to form a C-shaped molecule (Fig. 16A-B). WH1 is centrally located and bridges WH2 and WH3 motifs. WH1 packs against WH2 with two-fold rotational symmetry with their β-wings at the interface, and against WH3 in a 90° offset head-to-tail arrangement. This cluster of WH motifs appears to be a novel protein architecture, as a search for structural homologs using the DALI server (Holm and Sander, 1993) returned no results aside from those of each individual WH (SI Fig. S2 A-C in (Mullins *et al.*, 2017)). Consistent with the established structural diversity among WH motifs, the three motifs in AlkZ show slight variations from one another (Fig. 16A-B), including a highly kinked α3 “recognition helix” in WH2 (helix αI). However, with the

exception of WH3, which lacks the short β1 strand, each WH adopts the canonical α1-β1-α2-α3-β2-β3 topology (Fig. 16B).

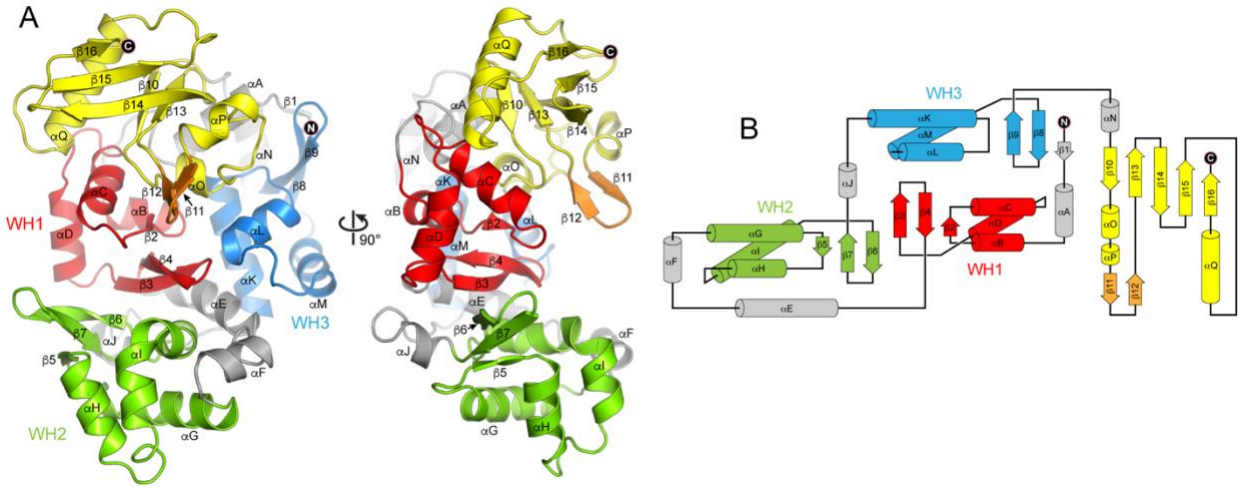


Figure 16. *Crystal structure of Streptomyces sahachiroi AlkZ* (A) Orthogonal views of the enzyme colored by subdomain. Winged helix (WH) motifs are red, green, and blue; the C-terminal domain is yellow; the putative DNA-binding β11/12 loop is orange; and connecting helices are gray. (B) Topology diagram colored as in panel A.

The most notable feature of the C-shaped structure is the concave channel that spans the width of the protein. The channel is defined by all three WH motifs and a short β-hairpin (β11/12) that protrudes from the C-terminal domain. The surface of this channel is lined with positively charged, highly conserved residues and is appropriately sized to accommodate duplex DNA (Fig. 17A-B). DNA bound along this channel would be clamped between the β11/12 hairpin on one side of the duplex and helix αI—the putative "recognition helix" of WH2—on the other side, indicating that WH2 may play a role in DNA recognition similar to that of other WH motifs. In contrast, the corresponding helices from WH1 (αD) and WH3 (αM) reside outside the central channel in positions not likely to bind DNA. Indeed, docking DNA from other WH-DNA and HTH-DNA complexes onto WH1 or WH3 positioned the DNA well outside of the putative DNA-binding channel.

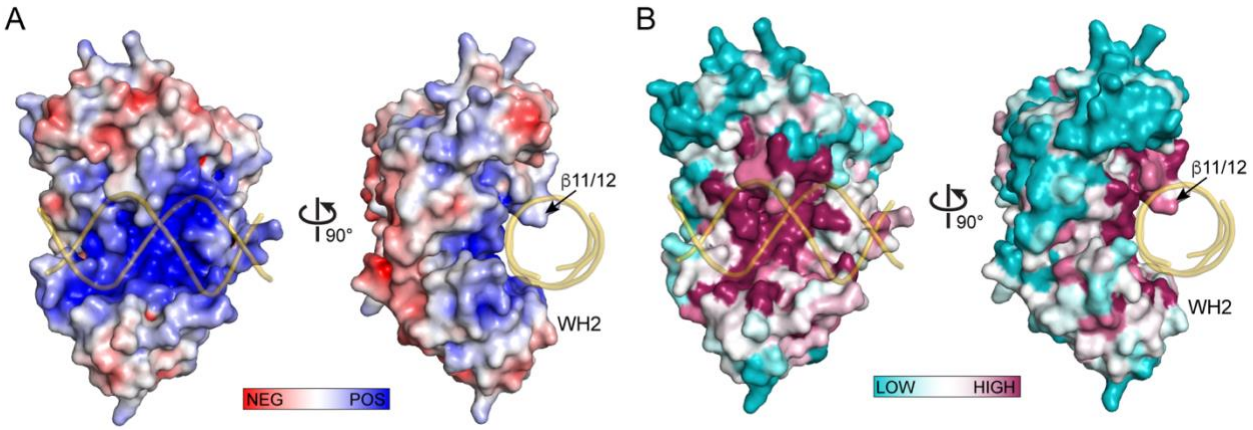


Figure 17. *Surface electrostatic potential and amino acid conservation of AlkZ* (A) Solvent-accessible surface of AlkZ colored by electrostatic potential (−7 to +7 k_BT/e_C). (B) Solvent-accessible surface of AlkZ colored by sequence conservation. B-DNA modeled against the crystal structure is shown as a transparent gold backbone trace.

Active site residues reside within the putative DNA-binding channel

To gain insight into how AlkZ properly orients an AZB lesion for catalysis, we docked a B-DNA model containing a computationally derived AZB ICL (Alcaro and Coleman, 2000) as a rigid body into the putative DNA-binding channel (Fig. 18A-B). Two nearly equivalent DNA orientations, related by the symmetric nature of the AZB ICL relative to the DNA dyad axis, optimized van der Waals and electrostatic interactions without altering the structure of the protein or the DNA. Both orientations placed the β11/12 hairpin into the minor groove directly across from the AZB ICL, and WH2 against the adduct in the major groove.

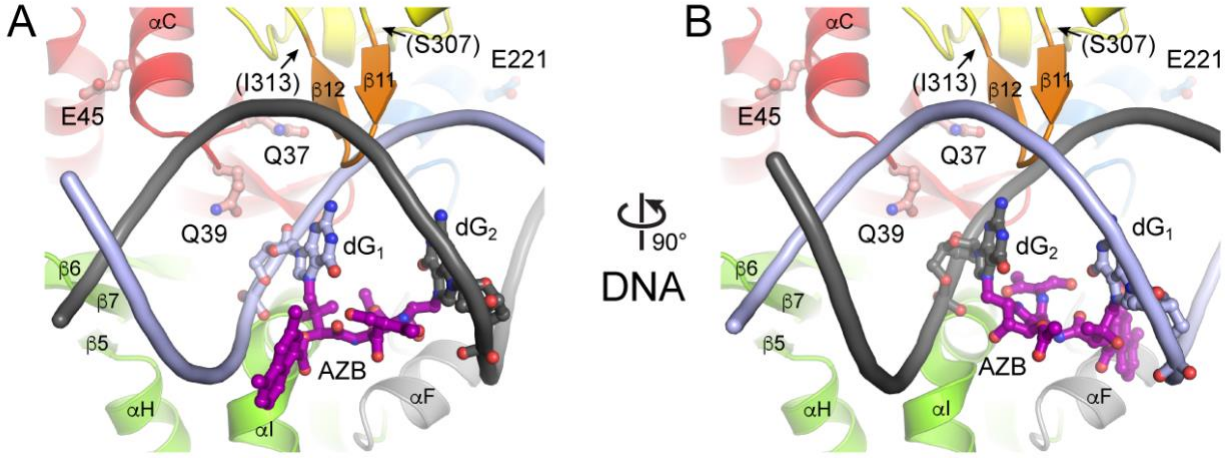


Figure 18. *Docking of AZB-ICL-DNA into AlkZ concave surface to probe for active site residues* (A-B) Orthogonal views of the AlkZ active site with AZB-ICL-DNA docked in the putative DNA-binding channel. The protein is colored as in Fig. 3, and AlkZ residues tested for their roles in base excision activity are labeled. The AZB-ICL-DNA model is shown as a slate/charcoal backbone trace with adducted guanosines (slate/charcoal) and AZB (purple) depicted in ball-and-stick representation.

We used these models to identify candidate catalytic residues among the highly conserved polar, cationic, aromatic, and hydrophobic side chains lining the concave surface (Fig. 18A-B), and thus likely in the vicinity of a bound DNA lesion. Monofunctional glycosylases generally initiate nucleophilic attack from the minor groove using carboxylate (Asp, Glu) or carboxamide (Asn, Gln) side chains (Schärer and Jiricny, 2001; Maiti *et al.*, 2012). In AlkZ, side chains from conserved glutamines Q37 and Q39 (Fig. 18A-B) point into the minor groove at either point of ICL attachment in both orientations of our docking model (Fig. 18A-B). Whereas Q37 is recessed at the rear of the channel, Q39 protrudes far enough into the minor groove to orient a water nucleophile for catalysis.

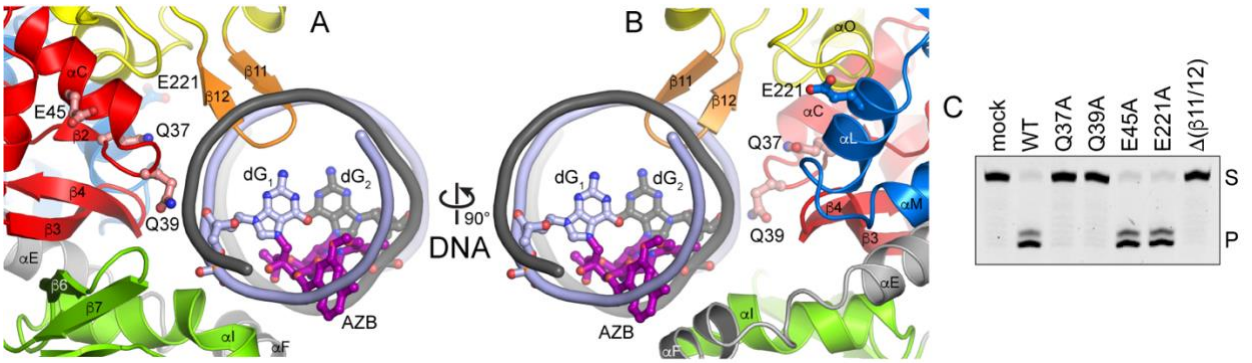


Figure 19. *Identification of the AlkZ active site residues* (A-B) Orthogonal views of AlkZ bound to AZB-ICL-DNA in a hypothetical alternate binding orientation. (C) Base excision activity of wild-type (WT) and mutant AlkZ proteins as measured by the formation of nicked abasic-DNA product (P) after incubation of protein or buffer (mock) with d7mG-DNA substrate (S) for 24 hrs.

I tested involvement of the helix penetrating QxQ motif and β11/12 hairpin in enzyme activity by substitution with alanine or glycine and measuring d7mG excision activity from purified proteins (Fig. 19C). We verified by circular dichroism that all mutant proteins were folded under the conditions of our assay (SI Fig. S3C in (Mullins *et al.*, 2017)). Both Q37A and Q39A mutants completely abrogated d7mG excision, even after 24 hours. In contrast, alanine substitution of the nearest glutamate side chains, E45 and E221, which reside just outside the putative active site on helices αC (WH1) and αL (WH3), respectively, had no effect on base excision activity. In addition to catalysis by polar residues, DNA glycosylases often stabilize a particular DNA conformation in the enzyme-substrate complex by inserting a steric wedge—either bulky side chains or a loop—into the DNA helix. In AlkZ this role is likely played by the β11/12 hairpin, which protrudes into the minor groove in our docking model (Fig. 19A-B). Consistent with this hypothesis, replacing residues 307-313 that form the β11/12 loop with two glycine residues (Δβ11/12) abrogated d7mG excision activity (Fig. 19C). These results identify

this central region of the protein as the active site, with important roles for Q37, Q39, and the β11/12 hairpin in base excision.

Discussion

This work defines the architecture of the HTH_42 family of DNA-binding proteins (Finn *et al.*, 2016) and an eighth class of DNA glycosylase (Brooks *et al.*, 2013; Miyazono *et al.*, 2014). A unique cluster of three tandem WH motifs scaffold a putative DNA-binding channel, with only one WH (WH2) positioned to participate in the canonical helix-major groove recognition used by other WH and HTH motifs (Gajiwala and Burley, 2000). The active site contains a QxQ motif and a β-hairpin essential for base excision activity and conserved among AlkZ homologs and the HTH_42 superfamily (SI Fig. S4 in (Mullins *et al.*, 2017)). Our docking models predict the β11/12 hairpin to insert into the DNA minor groove at the lesion site. This motif is likely involved in stabilizing a particular DNA conformation in the enzyme-substrate complex, similar to that observed for β-hairpins and loops in other glycosylases (Brooks *et al.*, 2013). The N-terminal glutamine (Q37) in the QxQ motif is receded in the active site cleft and thus predicted to bind the DNA backbone, consistent with histidine occupying this position in some AlkZ homologs. The C-terminal glutamine (Q39), on the other hand, likely plays a catalytic role. Our docking models predict that Q39 protrudes into the minor groove so that its carboxamide side chain is positioned to orient a water nucleophile in the manner observed for N140 in human thymine DNA glycosylase (Maiti *et al.*, 2012) and for aspartate or glutamate side chains in other glycosylases (Schärer and Jiricny, 2001). Consistent with such a catalytic

role, half of all HTH_42 proteins have an aspartate at this position, suggesting that at least some other members of this previously uncharacterized superfamily also possess DNA glycosylase activity.

AlkZ is one of two DNA glycosylases that definitively unhook an ICL (Semlow *et al.*, 2016; Wang *et al.*, 2016). The mechanism by which a DNA glycosylase unhooks an ICL is not immediately obvious since these enzymes typically capture the modified nucleoside inside the active site by extruding it from the DNA helix, and such a base flipping mechanism would be inhibited by an ICL. However, we recently discovered a non-base-flipping mechanism that enables glycosylase excision of bulky minor groove adducts and that conceivably could enable ICL excision. The bacterial AlkD glycosylase is able to recognize and cleave adenine adducts of the bulky natural product yatakemycin (YTM) without rotating the lesion from the duplex (Mullins *et al.*, 2015). Like AlkZ, AlkD adopts a C-shaped fold that engages DNA along the concave surface (Rubinson *et al.*, 2010). The AlkD active site positions a catalytic aspartate and two catalytic tryptophan side chains (Parsons *et al.*, 2016) against the deoxyribose of the alkylated adenosine, which is displaced only slightly into the minor groove of the DNA. Similarly, our AlkZ docking models show the QxQ motif to be positioned far enough into the minor groove to access the N-glycosidic bonds of the ICL without rotating the crosslinked nucleosides (Fig. 19A-B). The eukaryotic DNA glycosylase with ICL unhooking activity, NEIL3, also has an active site (Liu *et al.*, 2013) that conceivably could access a non-flipped lesion.

Previous work established that AlkZ can unhook an AZB ICL, but it was less clear if the enzyme cleaves both sides (Wang *et al.*, 2016). While full ICL excision by a DNA glycosylase would create opposing abasic sites susceptible to a deleterious double-

strand break, it would also simplify repair by eliminating the need for NER or TLS/HR pathways to remove the remaining monoadduct. Pol I and II bypass of abasic sites and major groove monoadducts (Shibutani *et al.*, 1997; Yamanaka *et al.*, 2011) suggests that BER could in principle fully process opposing abasic sites in bacteria. Several lines of evidence suggest that AlkZ can excise the entire crosslink by one of two mechanisms, either by sequentially unhooking the ICL and excising the remaining monoadduct, or by simultaneously cleaving the glycosidic bonds of both modified dG residues. First, AlkZ possesses both unhooking activity for ICLs and excision activity for monoadducts (Fig. 2 and ref. Wang *et al.*, 2016) . Second, our DNA docking models illustrate that AlkZ can bind to either side of the AZB ICL (Fig. 20A-B). Both binding orientations position the putative active site against the minor groove on either side of the ICL without obstruction from AZB.

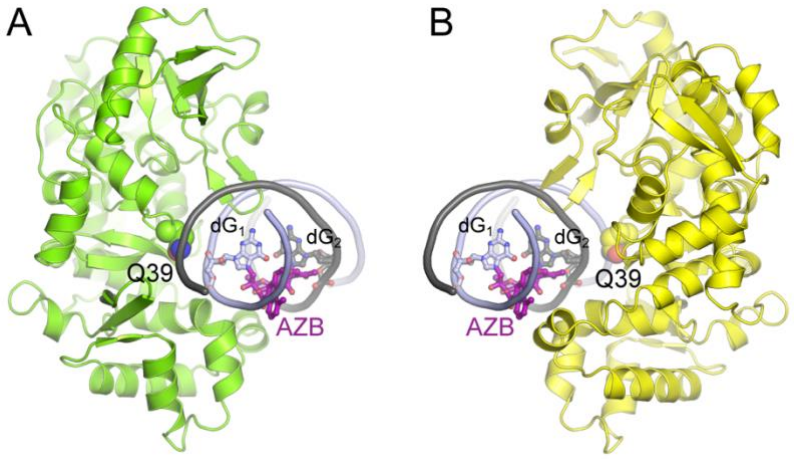


Figure 20. *Monomeric AlkZ AZB-ICL unhooking model* (A-B) Two binding orientations of AlkZ against each side of an AZB ICL. The catalytic Q39 side chain is shown in green or yellow and the AZB ICL is shown in purple.

Third, our docking models show that the two binding orientations can co-exist at the same ICL. Modelling two AlkZ molecules on opposite sides of the DNA with their

respective active sites against the two crosslinked dG residues reveals a remarkable complementarity between protein surfaces (Fig. 21A-C). While there is no evidence for dimerization of AlkZ in the presence of d7mG-DNA, dimerization may be dependent on the AZB ICL. The recent discovery of a dimeric DNA glycosylase acting as part of a restriction-modification system in thermophilic bacteria (Miyazono *et al.*, 2014) raises the possibility that AlkZ could utilize a similar strategy to simultaneously excise the entire crosslink. However, at the lower temperatures at which AlkZ functions, spontaneous formation of a double-strand break from opposing abasic sites would be less likely. Additionally, the bound AlkZ dimer may transiently shelter both abasic sites from hydrolysis until further repair by enzymes in the BER pathway occurs.

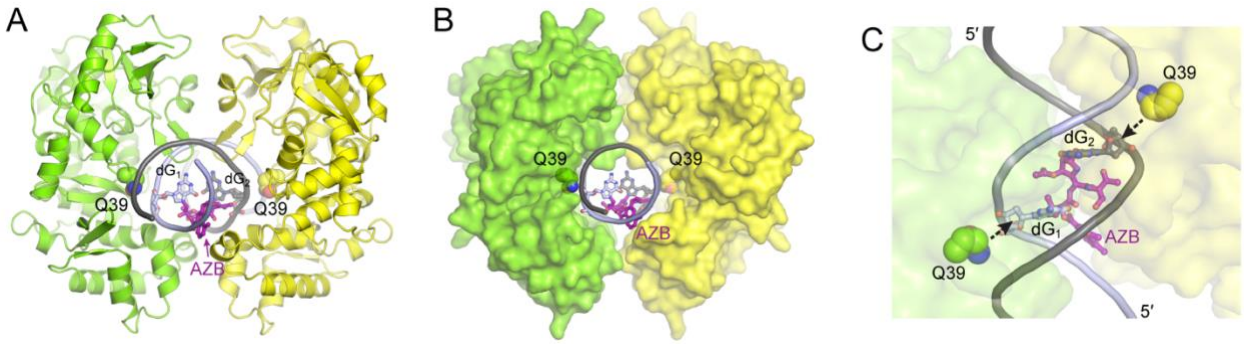


Figure 21. *Dimeric AlkZ AZB-ICL unhooking model* (A-B) Dimeric model of two AlkZ subunits bound to one AZB ICL. The proteins are shown as green and yellow cartoons (A) or solvent-accessible surfaces (B). (C) Top view of the hypothetical AlkZ dimer interface, showing simultaneous engagement of both crosslinked dG residues by the catalytic Q39 side chain.

Pathogenic and antibiotic-producing microbes have resistance mechanisms against their own toxins that are often genetically clustered with the antibiotic synthesis operon (Hubbard and Walsh, 2003; D'Costa *et al.*, 2006; Cundliffe and Demain, 2010). AlkZ is the second DNA glycosylase discovered to act as a toxin resistance mechanism

(Wang *et al.*, 2016). The AlkD homolog YtkR2 is produced in *Streptomyces* sp. TP-A0356 as a means to protect against its natural product YTM, a bulky genotoxin of the spirocyclopropylcyclohexadienone family (Igarashi *et al.*, 2003; Xu *et al.*, 2012). Similar to the AlkZ gene (*azi36/orf1*), *ytkR2* and putative BER genes are embedded within the YTM synthesis gene cluster. Although AlkZ and YtkR2 likely originated in AZB- and YTM-producing organisms, homologous proteins are evident in diverse bacteria, including human pathogens (Alseth *et al.*, 2006; Wang *et al.*, 2016). Thus, the two bacterial DNA glycosylases known to repair bulky and crosslinked DNA damage arising from potent bacterial toxins have transferred among bacterial populations. This implicates DNA repair, and alternative mechanisms thereof, in possible treatments against antibiotic resistance, and opens the possibility that other DNA repair mechanisms and genotoxic metabolites await discovery in pathogenic and antibiotic producing bacteria.

[2] This work is published in: Mullins, E.A., Warren, G.M., Bradley, N.P., and Eichman, B.F. (2017). Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4400-4405. I prepared AlkZ WT and mutant enzymes, designed and performed biochemical experiments, interpreted data, generated figures, and contributed to the editing and revision of the manuscript.

Chapter 3

Escherichia coli YcaQ is a DNA Glycosylase that Unhooks DNA Interstrand Crosslinks[3]

Introduction

DNA damage arising from a variety of endogenous and environmental agents pose a significant risk to cell viability. Interstrand DNA crosslinks (ICLs) are highly toxic DNA lesions that covalently tether the opposing strands of DNA, thereby inhibiting essential cellular processes such as DNA replication and transcription that rely on duplex unwinding (Raschle *et al.*, 2008). ICLs are produced by highly abundant endogenous metabolites and DNA repair intermediates and by a number of environmental toxins including microbial and plant natural products (e.g., mitomycin C and psoralens) that have therapeutic properties (Schärer, 2005; Noll *et al.*, 2006; Price *et al.*, 2014; Burgos-Barragan *et al.*, 2017; Hodskinson *et al.*, 2020). Because of their cytotoxicity, crosslinking agents are often potent antimicrobials and many, including nitrogen mustards, are among the most widely used drugs in cancer chemotherapy (Nagaoka *et al.*, 1986; Deans and West, 2011).

ICL repair involves unhooking the two strands by one of two known pathways, followed by repair of the resulting monoadduct (Dronkert and Kanaar, 2001; Noll *et al.*, 2006). The primary mechanism of ICL unhooking involves incisions to one strand by Fanconi anemia (FA) or nucleotide excision repair (NER) associated endonucleases (Wood, 2010; Clauson *et al.*, 2013; Hashimoto *et al.*, 2016). The resulting gap and strand break must be further processed by translesion DNA synthesis and homologous

recombination prior to repair of the monoadduct by a second round of NER. An alternative ICL repair pathway was recently discovered in both eukaryotes and prokaryotes, whereby the ICL is unhooked by DNA glycosylase cleavage of one of the N-glycosidic bonds linking the modified nucleotide to the DNA backbone, generating an abasic (AP) site on one strand but leaving the backbone intact (Semlow *et al.*, 2016; Wang *et al.*, 2016; Mullins *et al.*, 2019). DNA glycosylases typically initiate base excision repair (BER) of small monoadducts. Because their mode of lesion recognition and excision involves extruding (or flipping) a single nucleotide out of the helix, the mechanism by which they unhook ICLs tethered across strands is not understood.

Glycosylase-mediated ICL repair in bacteria was discovered in the biosynthetic gene cluster of azinomycin B (AZB), a genotoxic non-ribosomal peptide/polyketide secondary metabolite produced by the soil-dwelling microbes *Streptomyces sahachiroi* and *S. griseofuscus* (Terawaki and Greenberg, 1966). Azinomycins are bifunctional DNA alkylating agents that form ICLs between *N*7 of purines in 5'-PuNPy-3' sequences (Armstrong, 1992). AZB displays potent antibacterial activity against a variety of bacterial species (Nagaoka *et al.*, 1986), antitumor activity at lower doses than mitomycin C (Ishizeki *et al.*, 1987), and initiates transcription of DNA damage response genes after treatment (Kelly *et al.*, 2006). The AZB synthesis cluster contains a novel DNA glycosylase, AlkZ, which cleaves AZB-ICLs and *N*7-methylguanine monoadducts *in vitro* to produce an AP site that can be processed by the bacterial AP endonuclease EndoIV (Wang *et al.*, 2016; Mullins *et al.*, 2017). *AlkZ* expression is induced during production of AZB, and cells that express *alkZ* are resistant to the cytotoxic effects of AZB, even across different bacterial species (Wang *et al.*, 2016). Thus, the AlkZ glycosylase provides self-

resistance to the toxicity of these compounds in the producing organism (Wang *et al.*, 2016; Mullins *et al.*, 2019).

AlkZ belongs to the functionally uncharacterized HTH_42 superfamily of bacterial proteins known only for their arrangement of tandem winged helix-turn-helix (wHTH) motifs commonly found in transcription factors and other DNA binding proteins (Harami *et al.*, 2013). The crystal structure of AlkZ revealed a novel fold in which three wHTH motifs scaffold a C-shaped protein with a positively charged, concave surface that contains several residues important for DNA glycosylase activity (Mullins *et al.*, 2017). HTH_42 family proteins are widespread among antibiotic producers and human pathogens, including *Escherichia coli*, *Clostridium difficile*, and *Staphylococcus aureus,* among others (Wang *et al.*, 2016). Whether these proteins are bona fide AlkZ homologs that contain ICL unhooking activity is not known.

To better understand the mechanism of glycosylase-mediated ICL repair and to determine whether glycosylase ICL repair in bacteria is limited to AZB production and resistance, we compared ICL unhooking activity of *Streptomyces* AlkZ and the uncharacterized *E. coli* homolog YcaQ for different types of ICLs. Our results indicate that YcaQ is a cationic alkylpurine DNA glycosylase with robust activity for a broad range of substrates, including nitrogen mustard ICLs, whereas *Streptomyces* AlkZ is specific for AZB-ICLs. Deletion of *ycaQ* from *E. coli* increased cellular sensitivity to the crosslinking agent mechlorethamine, and this phenotype was complemented by recombinant YcaQ. Furthermore, overexpression of *ycaQ* showed a strong sensitivity to both ICL and methylating agents that was dependent on generation of toxic intermediates in the BER

pathway. Taken together, this work identifies *E. coli* YcaQ as an ICL repair glycosylase and implicates YcaQ-mediated BER as an alternative ICL repair pathway in *E. coli*.

Results

E. coli YcaQ is an alkylpurine DNA glycosylase

I first set out to determine whether other members of the HTH_42 family were bona fide AlkZ orthologs with DNA glycosylase activity. A search of the Pfam database (Finn *et al.*, 2016) revealed approximately 4,650 AlkZ orthologs, with 37% from pathogenic organisms and 21% annotated in antibiotic producing bacteria, consistent with a previous analysis (Wang *et al.*, 2016). The sequences fell into one of two clades that differ in the catalytic motif. I previously showed that the glycosylase activity of *S. sahachiroi* AlkZ is dependent on two glutamine residues within a QΦQ motif, in which Φ is a hydrophobic residue. However, about 55% of the putative orthologs have an aspartate in the third position (Fig. 22A, SI Fig. 1A in (Bradley *et al.*, 2020)). I focused on one of these QΦD sequences—the uncharacterized *ycaQ* gene found in *E. coli*, which shares 54% sequence similarity and 31% identity to *Streptomyces* AlkZ.

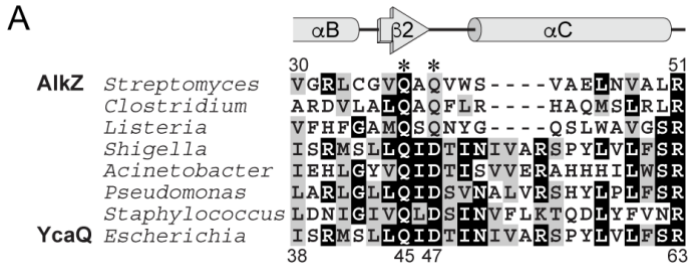


Figure 22. *Amino acid alignment of AlkZ to homologs in human pathogens* (A) Multiple sequence alignment of AlkZ/YcaQ homologs from *Streptomyces sahachiroi* and representative pathogens. Catalytic residues in AlkZ are labeled with asterisks. Secondary structures derived from the AlkZ crystal structure (PDB: 5UUJ) are shown above the alignment.

I first tested whether YcaQ has DNA glycosylase activity for a simple alkyl-DNA monoadduct using the same *N*7-deoxymethylguanosine (d7mG) substrate tested previously against AlkZ (Fig. 23A) (Mullins *et al.*, 2017). Purified YcaQ was incubated with d7mG-DNA containing a 5′-FAM label, followed by treatment with either hydroxide or the bacterial AP endonuclease, EndoIV, to nick any resulting AP sites. Compared to a no-enzyme control, YcaQ exhibited robust monofunctional base excision activity, as over 90% of the YcaQ treated substrate showed alkaline or AP endonuclease dependent cleavage (Fig. 23B). I confirmed the importance of the putative QΦD catalytic motif by showing that alanine substitution of either Gln45 or Asp47 abrogated 7mG excision (Fig. 23C). Mutation of Asp47 to asparagine did not affect base excision activity (Fig. 23C), consistent with a functional carboxamide in this position in the QΦQ family. Compared to AlkZ, YcaQ displayed much faster 7mG excision kinetics, taking the reaction to completion in under one minute (Fig. 23D).

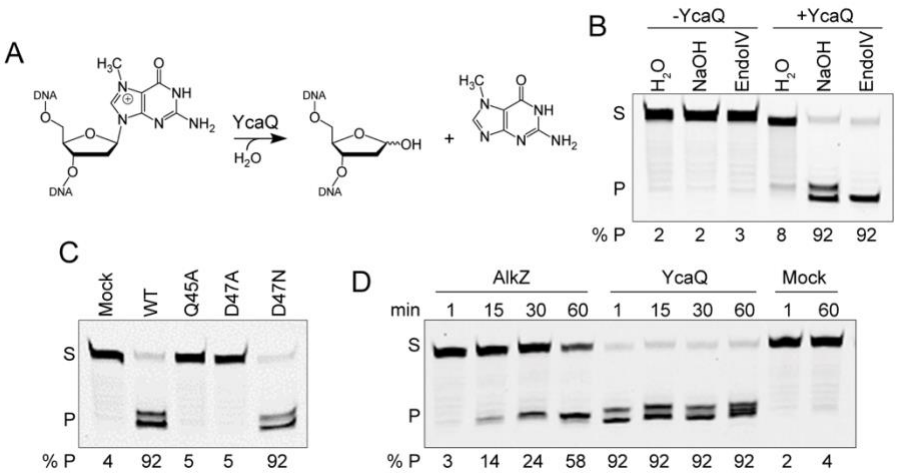


Figure 23. *E. coli YcaQ is a monofunctional DNA glycosylase specific for cationic N-alkylpurines* (A) Schematic of d7mG base excision reaction. (B-D) Denaturing PAGE of 5′-FAM labeled d7mG-DNA substrate (S) and nicked AP-DNA product (P) after treatment with enzyme or buffer (mock). AP-DNA resulting from glycosylase activity was nicked by treatment with either 0.1 M NaOH (B-C) or EndoIV (B) to produce β- and β,δ-elimination products, which are quantified below each gel. (B) d7mG-DNA was incubated with (+) or without (-) YcaQ for 1 hr followed by treatment with either water, NaOH, or EndoIV. (C) Comparison of WT and mutant YcaQ activity after 1 hr. (D) Time-dependence of d7mG excision by AlkZ and YcaQ.

I also tested wither YcaQ requires a cationic d7mG lesion for activity by converting the d7mG in our substrate to a ring-opened, neutral methylformamidopyrimidine (mFaPy) derivative (Fig. 24A). Compared to a FaPy DNA glycosylase (Fpg) control, YcaQ had no excision activity for the mFaPy adduct, even after one hour (Fig. 24A). Together these results show that YcaQ is a monofunctional DNA glycosylase from *E. coli* that can excise cationic d7mG monoadducts with rapid kinetics.

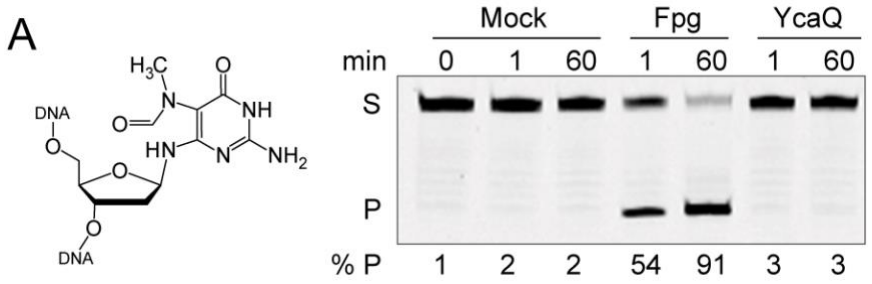


Figure 24. *YcaQ does not excise N5-methyl-FaPy adducts* (A) Structure and excision of mFaPy-dG DNA, treated with *E. coli* Fpg, YcaQ, or buffer for the specified time, followed by alkaline hydrolysis.

AlkZ and YcaQ unhook either side of an AZB-ICL

To examine ICL unhooking by the AlkZ/YcaQ family of glycosylases, I first compared their activities for AZB-ICLs. The azinomycins contain electrophilic epoxide and aziridine moieties that react with d(GCC)/d(GGC) sequences in a specific orientation to form an asymmetric ICL (Fig. 25A) (Salvati, 1992; Coleman, 2002; LePla *et al.*, 2005). AZB preferentially orients itself such that the epoxide moiety reacts with the GCC sequence and the aziridine reacts with GGC (Fig. 25B). The previous study showing AZB-ICL cleavage by AlkZ did not address if the enzyme has a preference for unhooking one side versus the other (Wang *et al.*, 2016). To better understand the mechanism of ICL unhooking, I devised a DNA substrate that would allow us to track individual strands. Pure AZB-ICL substrates for glycosylase assays were generated by incubating *Streptomyces sahachiroi* cell extracts with a 26-mer oligodeoxynucleotide duplex containing a single AZB reactive sequence, followed by denaturing PAGE purification to isolate the resulting crosslinked DNA. GCC and GGC strands were distinguished with FAM and Cy5 fluorescent labels at their 5′-ends, respectively, so that unhooking produces either FAM-AP + Cy5-monoadduct (MA) strands or FAM-MA + Cy5-AP strands (Fig. 25C). In addition, the reactive sequence is offset toward one end of the duplex so that alkaline cleavage of the resulting AP sites will generate specific FAM- or Cy5-labeled β- and β+δ-elimination products that will migrate differently on a denaturing gel (Fig. 25C).

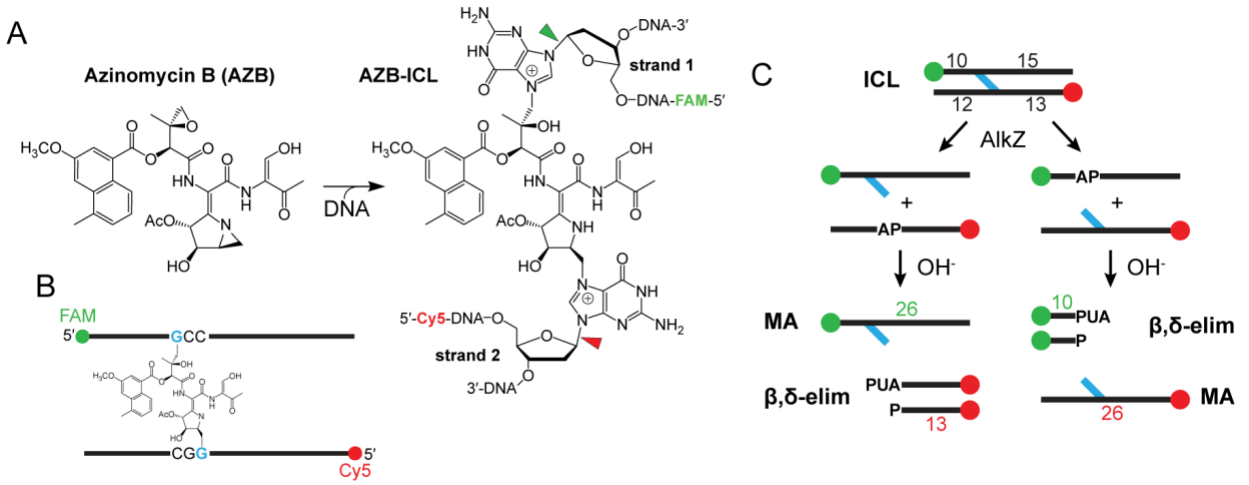


Figure 25. *ICL reaction of AZB with DNA and schematic of glycosylase unhooking assay* (A) Reaction of azinomycin B (AZB) with guanines on opposing strands to form an ICL. N-glycosidic bonds hydrolysed by AlkZ/YcaQ are highlighted with green and red arrows, corresponding to FAM and Cy5-labeled strands, respectively. (B) Schematic of the 26-mer oligonucleotide AZB-ICL substrate used in this study. The epoxide predominantly reacts with the FAM-labeled GCC strand and aziridine with the Cy5-labeled GGC strand. Green and red spheres represent FAM and Cy5 labels, respectively. (C) Schematic of the base excision assay used to monitor ICL unhooking activity. Glycosylase unhooking of ICL-DNA potentially forms monoadducts and AP sites on either strand. AP-sites are nicked with hydroxide to form shorter oligonucleotides through β,δ-elimination. ICL-, monoadduct (MA)-, and β,δ-elimination products can be separated by denaturing PAGE. PUA, 3′-phosphor-α,β-unsaturated aldehyde (β-elimination product), P, 3′-phosphate (β,δ-elimination product).

Heat treatment of the purified AZB-ICL substrate at 95°C for 5 min resulted in cleavage of both sides of the crosslink, as evidenced by both FAM- and Cy5-labeled AP and MA strands that could be resolved on the gel (Fig. 26 A, lane 3). Treatment of the heat denatured ICLs with hydroxide converted the faster migrating AP strands into nicked β/δ-elimination products and left the MA strands intact (Fig. 26A, lane 4). Alkali treatment also converted the ICL into a slower-migrating species, consistent with conversion of the double-cationic species to a neutral FaPy-ICL. Thus, AZB-ICLs display the same heat-labile depurination as other *N*7-guanine lesions (Gates *et al.*, 2004). At 25°C, the AZB-

ICL substrate is stable, showing no degradation after one hour (Fig. 26A, lanes 5-6). Incubation with AlkZ under the same condition resulted in complete (97%) ICL unhooking into elimination products from both strands, indicating that AlkZ unhooks the AZB-ICL from either side. The persistence of a modest amount of FAM-labeled alkali-resistant MA strand suggested an enzymatic preference for unhooking the GGC side of the crosslink (Fig. 26A, lane 7). Because the bases tethered by an ICL are constrained across the duplex, unhooking likely does not involve base-flipping as observed in other glycosylases. However, I found that the non-base flipping glycosylase AlkD (Mullins *et al.*, 2015; Mullins *et al.*, 2017) had no activity for the AZB-ICL substrate, indicating that ICL unhooking is not simply a product of a non-base-flipping mechanism and that recognition of AZB requires a specific type of enzyme.

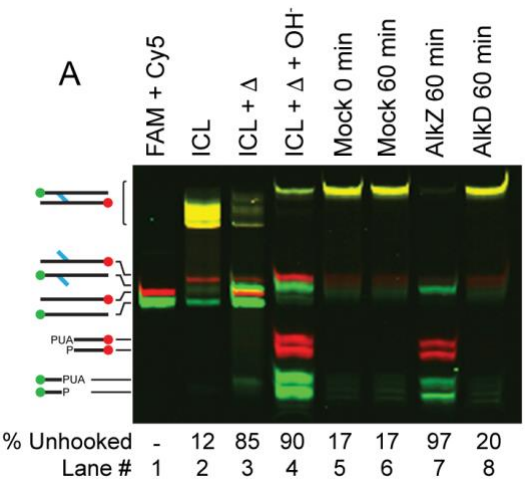


Figure 26. *AZB-ICL unhooking by AlkZ* (A) Denaturing PAGE of AZB-ICL, AZB-MA, and nicked AP-DNA products after treatment with heat (Δ), buffer (mock), or enzyme, followed by alkaline hydrolysis. The percent of β,δ-elimination product is quantified below the gel. Each image is an overlay of false-colored FAM (green) and Cy5 (red) fluorescence scans of the gels, in which yellow depicts coincident red and green intensity. Individual FAM and Cy5 imaged gels for AZB-ICLs are shown in SI Fig. S2 A-C in (Bradley *et al.*, 2020).

YcaQ also unhooked the AZB-ICL substrate to generate AP sites on both strands (Fig. 27A, lane 5), albeit in a slightly different manner than AlkZ. Interestingly, whereas the FAM-MA strand persisted in the AlkZ reaction, YcaQ generated more Cy5-MA strand (Fig. 27A, lanes 5-8), suggesting a difference in strand preference of the two enzymes. Second, alanine substitution of either residue in the catalytic QΦQ and QΦD motifs had different effects on unhooking. Mutation of QΦQ in AlkZ only partially reduced unhooking, while substitutions within YcaQ QΦD had a greater effect (Fig. 27A, lanes 3-4, 6-7), suggesting that the two enzymes rely on these motifs for catalysis to a different extent. Incidentally, the YcaQ D47N mutation had no effect on ICL unhooking activity, similar to its behavior against the d7mG monoadduct.

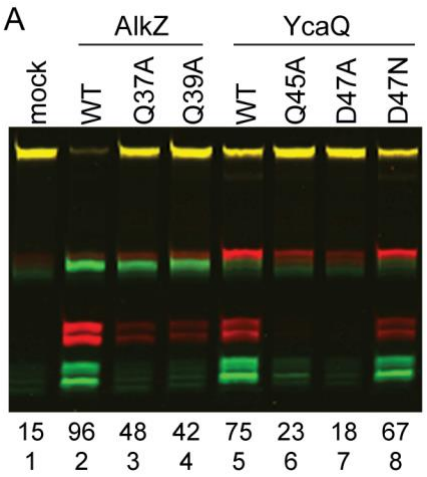


Figure 27. *AZB-ICL unhooking by AlkZ and YcaQ mutant proteins* (A) 30-min reactions between AlkZ, YcaQ, and catalytic mutants for AZB-ICL-DNA, followed by alkaline hydrolysis. The percent of β,δ-elimination product is quantified below the gel.

To further probe the apparent strand preference between AlkZ and YcaQ, we monitored the kinetics of AP and MA strand product formation over two hours under single-turnover conditions (Fig. 28A). AlkZ showed robust AZB-ICL unhooking activity (k_{cat} = 2 x 10^{-2} sec^{-1})—three orders of magnitude faster than that observed previously for

a d7mG monoadduct (k_{cat} = 8.5 x 10^{-5} sec^{-1}). YcaQ displayed 4-fold slower AZB-ICL unhooking kinetics under the same conditions (k_{cat} = 4 x 10^{-3} sec^{-1}) (Fig. 28A, lanes 14-19). Quantification of products from both FAM and Cy5 strands shows that AlkZ has a distinct preference for unhooking the GGC (Cy5) strand, as shown by the greater burst in cleaved Cy5-AP and uncleaved FAM-MA strands (Fig. 28A-B, lanes 8-13). In contrast, YcaQ had a modest preference for unhooking the GCC (FAM) strand (Fig. 28A-B, lanes 14-19), but the differences in FAM and Cy5 strand kinetics in the YcaQ reaction were not as dramatic as in AlkZ. These data show that the orthologous YcaQ and AlkZ enzymes unhook AZB-ICLs from either side, and with unique strand specificities.

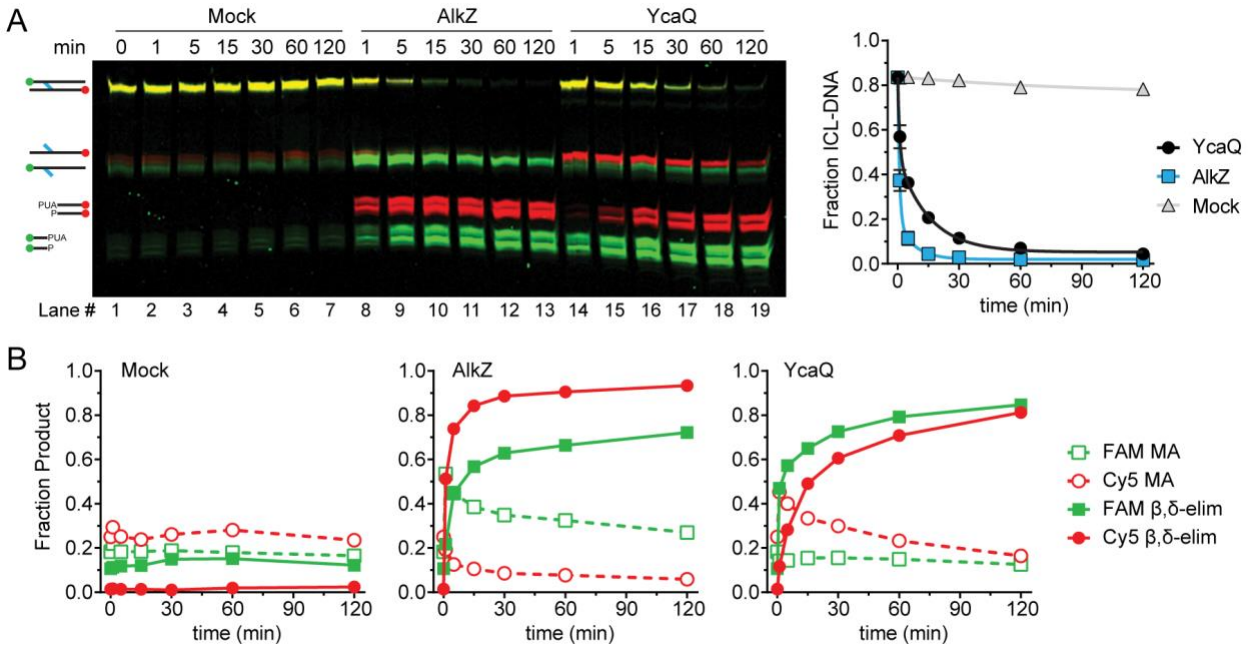


Figure 28. *Kinetics of AZB-ICL unhooking by AlkZ and YcaQ* (A) Denaturing PAGE showing time-dependence of AlkZ and YcaQ unhooking of AZB-ICLs. The fraction of ICL-DNA unhooked from three independent experiments is quantified to the right (mean ± SEM). (B) Quantification of the fraction of monoadduct (MA) and β,δ-elimination (nicked product) from the mock (left), AlkZ (middle), and YcaQ (right) reactions from the gel in panel A.

YcaQ unhooks ICLs derived from nitrogen mustards

AlkZ presumably evolved to provide self-resistance to AZB produced in soil-dwelling *S. sahachiroi*, whereas YcaQ in *E. coli* residing in the human gut would not encounter such a lesion. We therefore examined the ability of these enzymes to recognize an ICL derived from the nitrogen mustard (NM) mechlorethamine, a structurally simpler bifunctional alkylating agent than AZB (Schärer, 2005). Like AZB, mechlorethamine reacts with *N*7 of both guanines within d(GNC) sequences to form ICLs that lead to the toxicity of these compounds (Bauer, 1997) (Fig. 29A). Using a site-specific NM-ICL oligonucleotide substrate (Castano *et al.*, 2017) containing individually labeled strands (Fig. 29A), we found that YcaQ unhooked either side (Fig. 29B). We verified that this activity is specific to YcaQ, as β/δ-elimination products were was not observed from QΦD (Q45A, D47A) catalytic mutants (Fig. 29B) or from other alkylpurine DNA glycosylases, including human AAG, *E. coli* AlkA, and *Schizosaccharomyces pombe* Mag1 (Fig. 29C). Furthermore, YcaQ showed no unhooking activity toward FaPy-ICLs (SI Fig. S3 D-E in (Bradley *et al.*, 2020)), consistent with the results with mFaPy monoadducts indicating that the cationic, ring-closed guanine is required for the reaction.

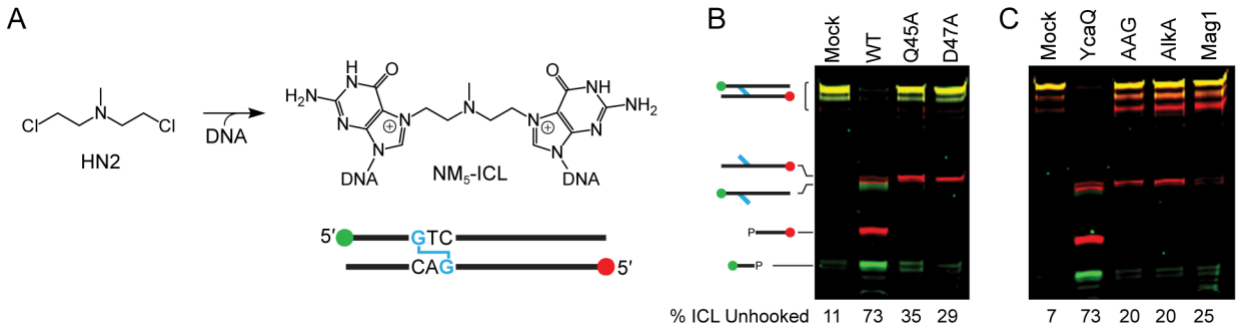


Figure 29. *Nitrogen Mustard$_5$ (NM_5)-ICL unhooking by YcaQ* (A) 5-atom nitrogen mustard (NM_5) ICL formed between mechlorethamine and guanines located on opposite DNA strands. (B-C) Denaturing PAGE of NM_5-ICL substrate, monoadduct, and nicked AP-DNA products after treatment with WT or mutant YcaQ (B), or with YcaQ, human AAGΔ83, *E. coli* AlkA, or *S. pombe* Mag1 (C) , followed by treatment with EndoIV. The percent of unhooked ICL is quantified below the gel. Images are false colored overlays of individual FAM and Cy5 scans of the gel, which are shown in SI Fig. S3 A-B, F in (Bradley *et al.*, 2020).

Most strikingly, comparison of the kinetics of NM-ICL unhooking showed that YcaQ was remarkably efficient (the reaction was complete in less than 5 minutes), whereas AlkZ showed no activity (Fig. 30A, lanes 14-19). Unlike the asymmetric unhooking observed with the AZB-ICL, we did not detect a strand preference as expected from the symmetric NM-ICL (Fig. 30A). Comparison of the activities of both enzymes for AZB- and NM-ICLs shows that AlkZ is specific for AZB, whereas YcaQ can unhook chemically diverse ICLs, with robust activity toward the ICL with less chemical functionality.

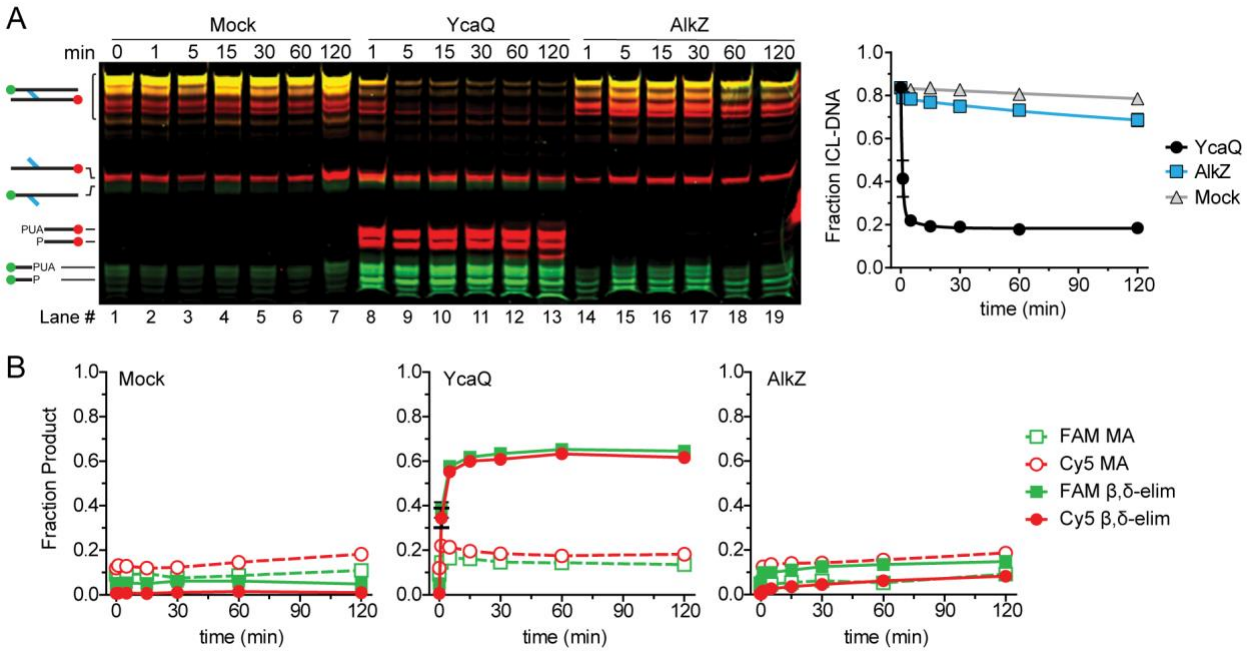


Figure 30. *Kinetics of NM_5-ICL unhooking by AlkZ and YcaQ* (A) Denaturing PAGE showing time-dependence of AlkZ and YcaQ unhooking of NM_5-ICLs with a hydroxide workup. The plot quantifies the decrease in the fraction of NM_5-ICL (mean ± SEM, n=3). (B) Quantification of the fraction of monoadduct (MA) and β,δ-elimination (nicked AP-DNA product) from the mock (left), AlkZ (right), and YcaQ (middle) reactions from the gel in panel A.

The mechanism of lesion recognition by a repair enzyme is a key question in ICL repair. Crosslinks may be located either by explicit interaction with the crosslinking molecule or by a structural perturbation of the DNA imposed by the crosslink. AZB, which has a 10-atom tether between guanines, is not expected to distort the DNA (Coleman, 2002). In contrast, the 5-atom tether in the NM-ICL has been shown to kink the double helix (Rink, 1995). To determine whether the differences in ICL specificity of AlkZ and YcaQ are the result of DNA distortion, I synthesized a NM derivative that would produce an 8-atom ICL (NM_8-ICL) when incorporated into our oligodeoxynucleotide substrate (Fig. 31A), similar to that used previously (Raschle *et al.*, 2008). The increased tether length of the NM_8-ICL should be able to crosslink the opposing guanines in B-form DNA without distorting the helix. Consistent with *N*7-alkylation at both sites, the purified NM_8-ICL was heat-labile and fully depurinated from both strands after 5 min at 95°C (Fig. 31B, lanes 2-3) but was stable at 25°C for at least 2 hours (Fig. 31B, lanes 4-10). Compared to a no-enzyme control, AlkZ exhibited a low level of unhooking activity for the NM_8-ICL substrate (Fig. 31B-C, lanes 11-16), suggesting that the steric strain present in the NM-ICL at least partially inhibits AlkZ activity. However, AlkZ's modest activity for the NM_8-ICL was still much less than that of the AZB-ICL, and thus we conclude that explicit contact with the AZB moiety is an important aspect of ICL recognition by AlkZ. YcaQ showed strong

activity toward the NM_8-ICL as it did with the NM_5-ICL (Fig. 31B-C, lanes 17-22), indicating that a kinked duplex is not required for ICL unhooking by the *E. coli* enzyme.

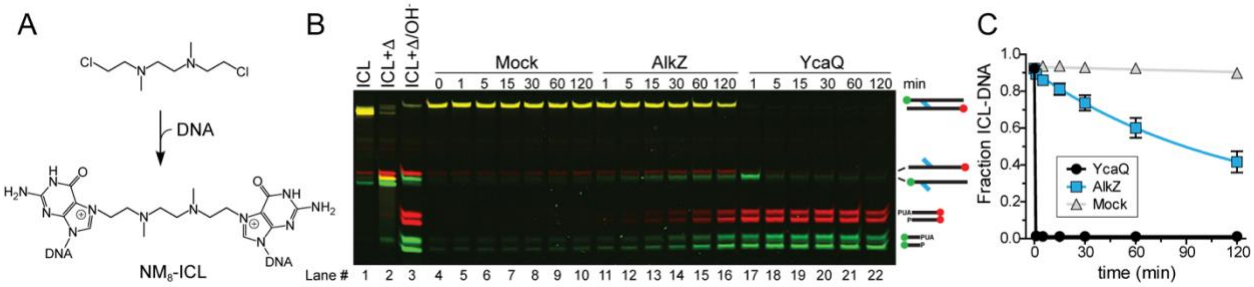


Figure 31. *Kinetics of NM_8-ICL unhooking by AlkZ and YcaQ* (A) Structure and ICL formed from 8-atom nitrogen mustard (NM_8) derivative. (B) Denaturing PAGE of the NM_8-ICL-DNA substrate and nicked abasic-DNA products after treatment with buffer (mock), AlkZ, or YcaQ for the specified time followed by alkaline hydrolysis. Heat-mediated depurination (lanes 2-3) serve as a positive control for excision products. (C) Quantification of the fraction of ICL from three separate experiments (mean ± SEM). Images are false colored overlays of individual FAM and Cy5 scans of the gel, which are shown in SI Fig. S4 A-B in (Bradley *et al.*, 2020).

YcaQ and AlkZ create opposing AP sites in vitro

Because AlkZ and YcaQ are able to act on either side of the crosslink, I asked whether they could act on both sides of the same crosslink. Such activity would generate two closely-spaced AP sites on opposite strands that in cells would potentially lead to a double-strand break (DSB). Indeed, the initial characterization of AlkZ activity toward AZB-DNA provided evidence that the enzyme produces opposing AP sites that can be subsequently cleaved by EndoIV (Wang *et al.*, 2016). To further test this, I analyzed the products EndoIV-treated AlkZ/AZB-ICL and YcaQ/NM-ICL reactions by native PAGE to quantify single- and double-nicked products. Both reactions showed at least 50% double-nicked product and a modest (10-20%) amount of monoadduct intermediate after one hour (Fig. 32A), indicating that AlkZ and YcaQ cleave both sides of the crosslinks to

generate AP sites on opposing strands. HPLC-MS analysis of the excision products of the YcaQ NM-ICL reaction confirmed the presence of an unmodified bis-guanine NM-ICL, with an observed m/z of 386.1799 Da, which is a 0.78 ppm difference from the calculated m/z of 386.1796 Da for Gua-NM_5-Gua. The difference in rates of unhooking the two sides of the crosslink shown in (Fig. 32A-B) argues that the strands are cleaved sequentially, which would proceed via an intermediate containing an AP site on one strand across from a monoadduct on the other. In support of this, we found that YcaQ is able to excise 7mG in a duplex containing an AP site on the opposite strand in the position it would reside in the unhooked NM- or AZB-ICL intermediate (Fig. 32C).

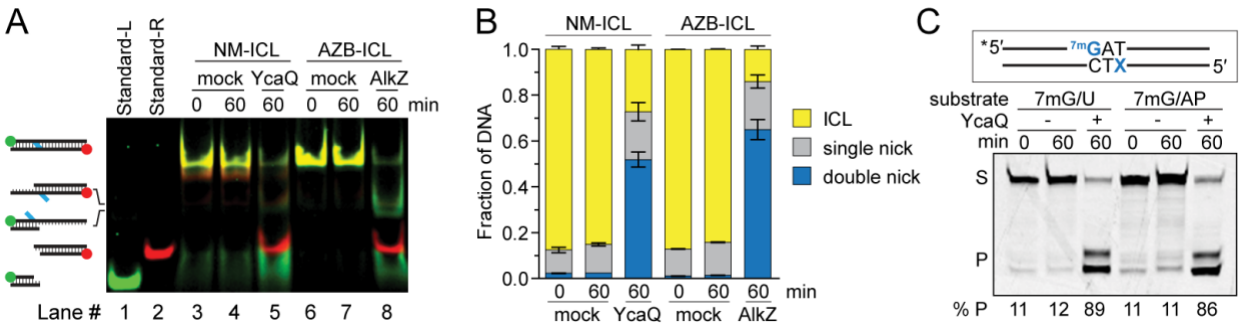


Figure 32. *AlkZ and YcaQ create opposing abasic (AP) sites during ICL unhooking* (A) Native PAGE analysis of DNA products formed from YcaQ incubation with NM_5-ICL and AlkZ incubation with AZB-ICL DNA substrates. AP sites formed from glycosylase activity were nicked by incubation with EndoIV prior to loading the DNA. Double-stranded standards for the double-nicked excision products are shown in lanes 1-2. The gel is a false-colored composite of the individual FAM- and Cy5-imaged gel, which are shown in SI Fig. S5A in (Bradley *et al.*, 2020). (B) Quantification of the total fluorescent signal for the fraction of ICL, single- and double-nicked EndoIV cleavage products from the gel in panel A. (C) Denaturing PAGE of d7mG/AP- and d7mG/dU-DNA substrates (S) and nicked AP-DNA products (P) after treatment with (+) or without (-) YcaQ followed by alkaline hydrolysis. Only the FAM-d7mG strand is visualized on the gel. The percent of β,δ-elimination products is quantified below.

Deletion or overexpression of YcaQ sensitizes E. coli to crosslinking agents

Repair of mechlorethamine derived NM-ICLs in *E. coli* is known to be initiated by the UvrABC nucleotide excision system (De Alencar *et al.*, 2005). Given the robust unhooking of NM-ICLs by YcaQ in vitro, I compared YcaQ's role in protecting *E. coli* against mechlorethamine to that of UvrA using genetic knockouts. Relative to the wild-type strain, *ΔycaQ* cells showed a modest growth defect in the presence of low levels of mechlorethamine (SI Fig. S6A in (Bradley *et al.*, 2020)). As expected from the importance of UvrA in ICL repair, *ΔuvrA* cells showed a more severe growth deficiency. This sensitivity was exacerbated in a *ΔycaQΔuvrA* double knockout (SI Fig. S6A in (Bradley *et al.*, 2020)), suggesting that YcaQ plays a minor role in ICL repair in *E. coli*. To examine this more quantitatively, we determined EC_{50} values for the deletion strains exposed to increasing doses of mechlorethamine using a colony dilution assay (Fig. A33-B). Consistent with the growth sensitivity, *ΔycaQ* cells showed a modest but significant reduction in EC_{50} compared to wild-type, and sensitivity of the *ΔycaQΔuvrA* mutant was greater than for *ΔuvrA* alone (Fig. 33A-B). Exogenous expression of *ycaQ* in the deletion strain rescued the mechlorethamine sensitivity compared to an empty vector control (Fig. 33C-D). Thus, YcaQ provides a bona fide mechanism for cellular protection against ICL toxicity, albeit to a lesser extent than the UvrABC system.

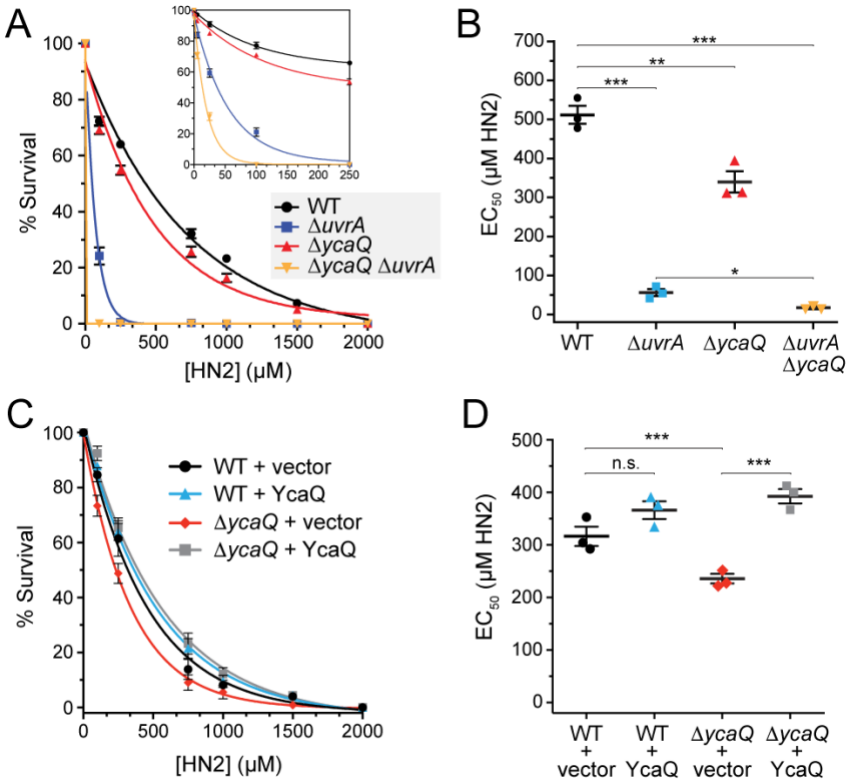


Figure 33. *Deletion of YcaQ sensitizes E. coli to mechlorethamine (HN2)* (A) Colony dilution assay for *E. coli* deletion strains exposed to increasing concentrations of mechlorethamine (HN2). Values are mean ± SEM (n=3). Percent (%) survival is relative to untreated cells. (B) EC_{50} values derived from data in panel A. Significance values were calculated using a one-way ANOVA (*, p=0.0140; **, p=0.0085; ***, p<0.0001; n.s, not significant). (C) Colony dilution assay for *E. coli* strains complimented with YcaQ or empty vector. Values are mean ± SEM (n=3). (D) Quantification of the data shown in panel D. One-way ANOVA values: ***, p=0.0004, n.s., not significant.

I also found that overexpression of *ycaQ* exhibited a dramatic increase in mechlorethamine sensitivity, with an EC_{50} value approaching that of *ΔuvrA* cells (Fig. 34A-B). This increased sensitivity was dependent on the catalytic activity of YcaQ, as overexpression of the inactive D47A mutant had no effect. Sensitivity to DNA damaging agents from DNA glycosylase overexpression is known to result from overwhelming the cells with AP sites and other toxic BER intermediates (Glassner *et al.*, 1998; Rinne *et al.*,

2004; Rinne *et al.*, 2005). Consistent with this interpretation, overexpression of both *ycaQ* and the AP endonuclease *endoIV* (*nfo*), which would help clear AP sites via BER, rescued the sensitivity (Fig. 34A-B). These results further implicate BER as a mechanism of ICL repair in bacteria.

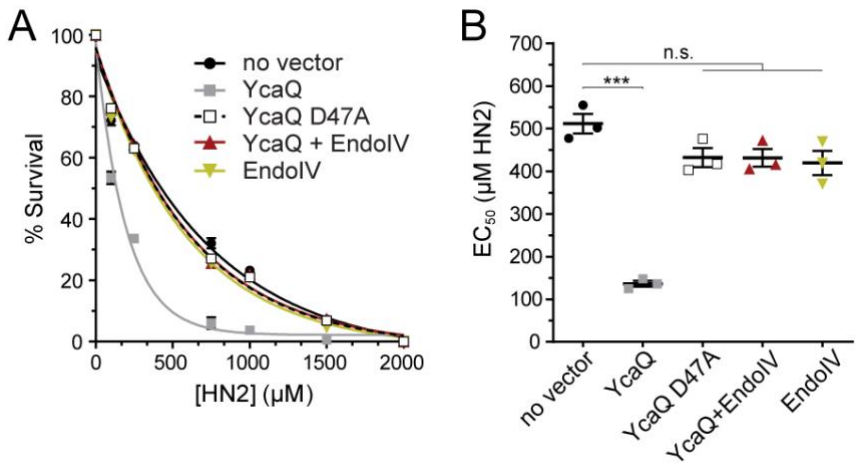


Figure 34. *Overexpression of YcaQ sensitizes E. coli to mechlorethamine* (A) Colony dilution assay showing the effect of mechlorethamine (HN2) on wild-type *E. coli* overexpressing YcaQ variants and/or EndoIV. (B) Quantification of the data shown in panel A. One-way ANOVA significance values: ***, $p<0.0001$; n.s, not significant.

Because YcaQ also displays activity toward alkyl monoadducts, I tested the effect of *ycaQ* deletion and overexpression on cells challenged with the methylating agent methylmethanesulfonate (MMS). *E. coli* contains only two other alkyl DNA glycosylases, the constitutively active Tag and inducible AlkA (Thomas *et al.*, 1982; Fernandez de Henestrosa and Barbé, 1991). I found that the *ΔycaQ* strain was no more sensitive to MMS than wild-type cells, in stark contrast to the effect of knocking out both Tag and AlkA (Fig. 35A-B, SI Fig. S6C in (Bradley *et al.*, 2020)) (Boiteux *et al.*, 1984; Nowosielska *et al.*, 2006). Overexpression of *ycaQ* increased cellular sensitivity towards MMS, although not to the same extent as observed for mechlorethamine treatment (Fig. 35A-B, SI Fig.

S6E in (Bradley *et al.*, 2020)). This sensitivity was partially rescued by overexpression of both *ycaQ* and *endoIV*, again consistent with *ycaQ* overexpression leading to generation of toxic BER intermediates.

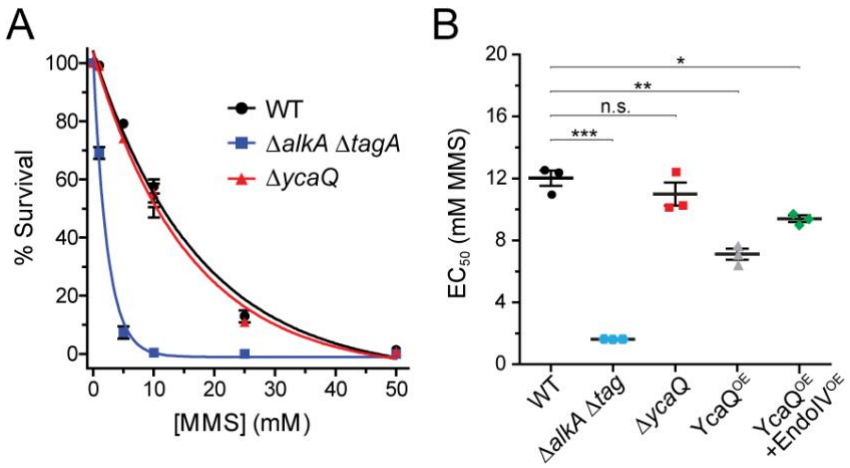


Figure 35. *Overexpression but not deletion of YcaQ sensitizes E. coli to MMS* (A) Colony dilution sensitivity of *E. coli* deletion strains (A) or wild-type *E. coli* overexpressing *ycaQ* or both *ycaQ* + *endoIV* (B) exposed to increasing concentrations of MMS. Values are mean ± SEM (n=3). (B) MMS EC_{50} values (mM) for various *E. coli* deletion (Δ) or over-expression (OE) strains. One-way ANOVA significance values: *, p=0.0088; **, p=0.0013; ***, p<0.0001; n.s, not significant.

YcaQ is constitutively expressed in cells

Prokaryotes have several inducible DNA damage responses to crosslinking and alkylating agents (Walker, 1984; Mielecki *et al.*, 2015). To test for the transcriptional induction of *ycaQ* by ICL agents, quantitative RT-PCR was performed on *E. coli* cells exposed to mechlorethamine (Fig. 36A). As part of the SOS response to crosslinking agents, the NER *uvr* genes are under the control of the *lexA* repressor. Both *lexA* and *uvrA* showed a robust increase in gene expression upon mechlorethamine treatment, whereas *ycaQ* expression remained unchanged after treatment (Fig. 36A). I also tested whether *ycaQ* is induced by the methylating agent MMS. The adaptive response genes,

ada and *alkA*, were used as positive controls and the constitutively expressed glycosylase *tag* served as a negative control. Whereas both *ada* and *alkA* showed elevated mRNA levels, I found that *ycaQ* expression is not induced by MMS (Fig. 36A). Together, these results indicate that *ycaQ* gene expression is not induced by either of the ICL or alkylating agents tested.

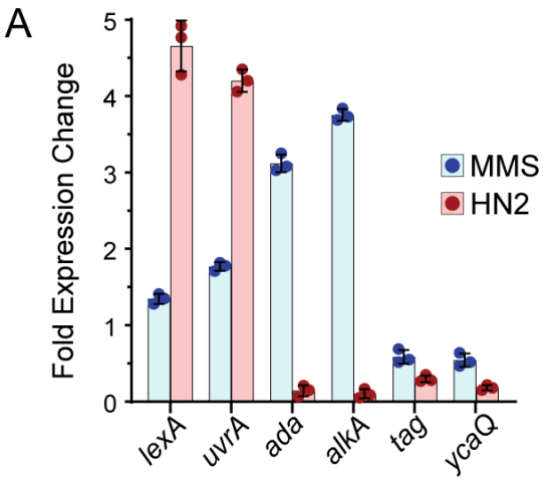


Figure 36. *YcaQ expression is not induced by HN2 or MMS in E. coli K-12 cells* (A) qRT-PCR results of DNA repair genes after treatment of *E. coli* with 5 mM MMS or 200 µM mechlorethamine (HN2) for 2 hours. Average ± SEM for three biological replicates. The fold expression change was calculated using the formula: (fold expression change) = $2^{-\Delta\Delta Ct}$, where Ct is the cycle threshold for amplification above baseline, ΔCt = Ct (gene of interest) – Ct (housekeeping gene), and $\Delta\Delta$Ct = ΔCt (treated sample) – ΔCt (untreated sample).

Discussion

YcaQ is the third alkylpurine DNA glycosylase to be identified in *E. coli* (Brooks *et al.*, 2013). Unlike Tag and AlkA, which excise only monoadducts during BER (O'Brien and Ellenberger, 2004; Metz *et al.*, 2007), YcaQ acts on crosslinked nucleobases to provide an alternative to NER-coupled ICL repair. The apparent specificity of YcaQ for cationic

alkylpurines would limit the types of ICLs to be repaired to those formed by alkylation at purine *N*3- or *N*7 positions, such as nitrogen mustards. YcaQ showed a much greater activity toward unhooking NM-ICLs than other known alkylpurine DNA glycosylases. This is somewhat contrary to a previous report that AlkA and AAG act on mustard-treated DNA, although that did not distinguish between ICL and monoadducts (Mattes *et al.*, 1996). Cells lacking *ycaQ* displayed a slight sensitivity towards mechlorethamine, but not MMS, consistent with the redundancy in repair mechanisms against methyl-DNA lesions (Morita *et al.*, 2010). Repair of NM-ICLs in *E. coli* is known to depend on NER and HR pathways, as *uvrA* and *recA* mutants show extreme sensitivity towards crosslinking agents (Berardini, 1997, 1999). The greater sensitivity of *ΔuvrA* versus *ΔycaQ* mutants to mechlorethamine shown here is consistent with NER/HR as the major ICL repair pathway for this type of lesion in *E. coli*.

A glycosylase-mediated ICL unhooking pathway could provide an error-free ICL repair mechanism by potentially bypassing the requirement for error-prone TLS across from the monoadduct (Mullins *et al.*, 2019). However, our *in vitro* data indicates that both AlkZ and YcaQ unhook crosslinks from either side to generate opposing AP sites in a 1,3 orientation that can be incised by EndoIV. In the cell, opposing AP sites in such close proximity could potentially lead to a deleterious DSB. Generation of opposing AP sites by AlkZ/YcaQ is in stark contrast to the 5mC glycosylase DEMETER in plants that removes 5mC in hemi-methylated d(5mCG/CG) islands and is inhibited by closely spaced 5mC residues on opposite strands in fully methylated $d(5mCG)_2$ sequences (Gehring *et al.*, 2006). Our data suggest a sequential mechanism of unhooking both sides of the ICL. We do not know if the enzyme dissociates and rebinds after the first unhooking event or

whether an enzyme-DNA complex persists to make the second cut, although structural modeling of the AlkZ/DNA complex shows a potential for crosstalk between two enzymes bound on either side of the crosslink (Mullins *et al.*, 2017; Mullins *et al.*, 2019). It remains to be determined in a cellular context if YcaQ forms opposing AP sites or DSBs, or if the enzyme is instead regulated to avoid a detrimental outcome. It also remains to be determined what the downstream resolution of glycosylase-mediated repair is, although previous studies suggest that *polB* (DNA pol II) may be involved in a recombination-independent mechanism for repair of NM-ICLs in *E. coli* (Berardini, 1997, 1999).

In eukaryotes, the NEIL3 glycosylase was identified as an alternate ICL pathway to FA/NER and capable of unhooking ICLs derived from abasic sites and psoralen (Semlow *et al.*, 2016; Wu *et al.*, 2019; Li *et al.*, 2020). NEIL3 and YcaQ are structurally unrelated and likely do not share the same mode of ICL recognition (Liu *et al.*, 2013; Mullins *et al.*, 2017). ICL repair by the NEIL3 glycosylase is S-phase dependent and involves convergent replication forks (Zhang *et al.*, 2015; Semlow *et al.*, 2016), and its glycosylase domain has an intrinsic specificity for DNA damage at one particular orientation of forked structures (Imani Nejad *et al.*, 2020). In contrast, the data presented here suggest that the bacterial AlkZ/YcaQ ICL glycosylase recognizes ICLs without regard to a specific DNA structure, consistent with ICL repair in bacteria occurring in the context of duplex DNA (Noll *et al.*, 2006; Huang and Li, 2013). Moreover, AlkZ's preference for an AZB-ICL is likely the result of direct contact between the protein and AZB. Our AlkZ crystal structure modeled against AZB-DNA suggests that helix αI in the second wHTH motif is in proximity to make direct contact with the AZB moiety (Fig. 37A)

(Mullins *et al.*, 2017). This region differs in sequence between the *Streptomyces* and *E. coli* enzymes, consistent with its putative role in substrate specificity (Fig. 37B).

Our results indicate that *ycaQ* expression is likely constitutive as it was not induced by either of the ICL or alkylating agents tested. Consistent with constitutive expression, *YcaQ* resides within a putative four-gene operon containing essential genes *msbA* and *lpxK* behind a σ^{70}-dependent promoter (Fig. 37C) (Keseler *et al.*, 2017). Large scale proteomic studies in *E. coli* suggest that YcaQ protein levels are among the lower 25% for abundance, indicating it may be kept at low protein concentrations in the cell (Arike *et al.*, 2012). Interestingly, the *ycaQ* operon also has a putative σ^{32}-dependent (heat shock) promotor, indicating a potential role for *ycaQ* and this operon in a global stress response (Zhao *et al.*, 2005). Additionally, high-throughput transcriptomic studies of *E. coli* under different stresses shows *ycaQ* expression is upregulated during heat and cold shock (Jozefczuk *et al.*, 2010). These data argue that YcaQ does not have a specific substrate, but instead is a mechanism to unhook diverse ICLs arising from various bifunctional alkylating agents (Fig. 37D). Nevertheless, we cannot rule out the possibility that YcaQ plays a specific role in certain strains of *E. coli* or other related bacteria that produce genotoxic secondary metabolites during a stress response. For example, *pks*$^{+}$ *E. coli* producing colibactin, an ICL agent associated with formation of colon cancer (Nougayrède *et al.*, 2006; Dalmasso *et al.*, 2014; Vizcaino and Crawford, 2015; Bossuet-Greif *et al.*, 2018; Wilson *et al.*, 2019; Xue *et al.*, 2019; Pleguezuelos-Manzano *et al.*, 2020), likely have some form of self-resistance against the toxin. YcaQ may play that role in a manner similar to that of *Streptomyces* AlkZ for azinomycin or of the AlkD/YtkR2 glycosylases for the non-covalent ICL agent yatakemycin (Mullins *et al.*, 2015; Wang *et*

al., 2016; Mullins *et al.*, 2017). Although more work is needed to understand the rationale for YcaQ involvement in ICL repair in *E. coli* and other bacteria, this work expands the role of DNA glycosylases and the BER pathway in repair of ICLs.

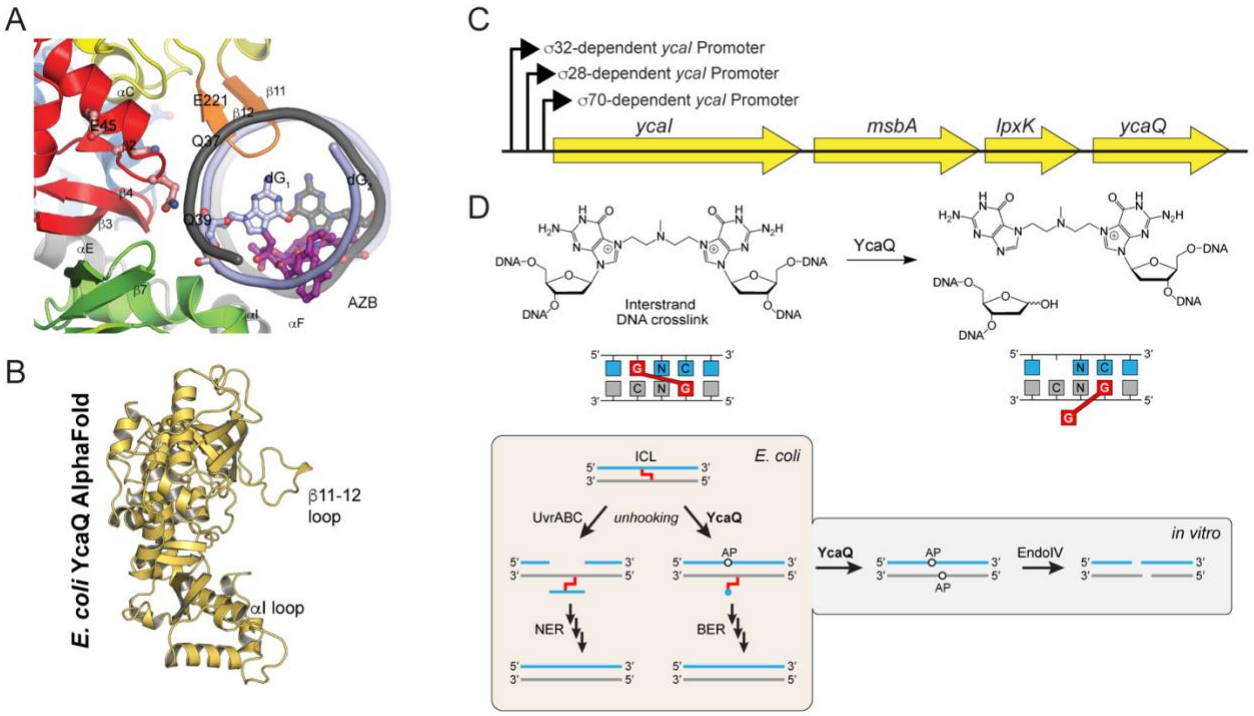


Figure 37. *Structural differences between AlkZ and YcaQ and model for ICL repair by BER in E. coli* (A) Hypothetical model of AlkZ bound to AZB-ICL-DNA (αI in green below AZB-ICL). (B) AlphaFold model (Jumper *et al.*, 2021) of *E. coli* YcaQ with αI loop and β11-12 loop labeled. (C) Predicted operon structure for YcaQ from the EcoCyc database (Keseler *et al.*, 2017). Putative δ-dependent promoters are labeled in front of the operon and are predicted based on consensus sequence similarities to known promoters. (D) Cartoon schematic for model of ICL excision repair by YcaQ in *E. coli*.

[3] This work is published in: Bradley, N.P., Washburn, L.A., Christov, P.P., Watanabe, C.M.H., and Eichman, B.F. (2020). Escherichia coli YcaQ is a DNA glycosylase that unhooks DNA interstrand crosslinks. *Nucleic Acids Res* **48**, 7005-7017. I designed and conducted all experiments, prepared the figures, and wrote manuscript.

Chapter 4

Resistance-Guided Mining of Bacterial Genotoxins Defines a Family of DNA Glycosylases[4]

Introduction

Bacteria are exceptionally rich sources of secondary metabolites, which are important for their survival and often have therapeutic value. *Streptomyces* produce 35% of all known microbial natural products and nearly 70% of all commercially useful antibiotics, with several being FDA-approved antitumor agents used as first-line cancer treatments (Demain and Sanchez, 2009; Procopio *et al.*, 2012; Jacob and Weissman, 2017; Law *et al.*, 2020). Secondary metabolites are often toxins used in ecological interactions with other organisms and can target any number of critical cellular functions (Tyc *et al.*, 2017). Natural products that damage DNA (genotoxins) form covalent or non-covalent DNA adducts that can inhibit replication and transcription, thus undermining genomic integrity through mutagenesis or cell death (Gates, 2009; Chumduri *et al.*, 2016). Consequently, genotoxins are particularly useful antineoplastic agents, as exemplified by several clinically relevant drugs including doxorubicin, bleomycin, mitomycin C, and duocarmycin analogs (Huang *et al.*, 2021).

Streptomyces produce a wide variety of DNA alkylating and oxidizing agents that have antimicrobial and antitumor properties. Spirocyclopropylcyclohexadienones (duocarmycin A and SA, yatakemycin, CC-1065) (Boger and Garbaccio, 1997; Parrish *et al.*, 2003), pluramycins (pluramycin A, hedamycin, altromycin) (Nagai *et al.*, 1967; Hansen

and Hurley, 1995; Hansen *et al.*, 1995), anthracycline glycosides (trioxacarcin A, LL-D49194α1) (Tamaoki *et al.*, 1981; Tomita *et al.*, 1981; Maiese *et al.*, 1990), and the leinamycin family (Nooner *et al.*, 2004) contain a single reactive group that covalently modifies purine nucleobases to form a broad spectrum of bulky alkyl-DNA monoadducts. *Streptomyces* also produce bifunctional alkylating agents that react with nucleobases on both DNA strands to create interstrand crosslinks (ICLs). Mitomycin C (MMC) from *S. lavendulae* crosslinks guanines at their N2 positions, and azinomycin A and B (AZA, AZB) from *S. sahachiroi* and *S. griseofuscus* crosslink purines at their N7 nitrogens (Terawaki and Greenberg, 1966). In addition to alkylating agents, several families of natural products, including bleomycins and enediynes, exert their toxicity by oxidative cleavage of DNA and RNA (Galm *et al.*, 2005).

The production of secondary metabolites in *Streptomyces* is genetically organized into biosynthetic gene clusters (BGCs), which contain the genes necessary for their biosynthesis, export, regulation, and resistance. Resistance mechanisms protect antibiotic producers from toxicity of their own natural products, and include toxin sequestration, efflux, modification, destruction, and target repair/protection (Cundliffe and Demain, 2010; Tenconi and Rigali, 2018). In the case of genotoxins, several DNA repair enzymes have been identified as target repair resistance mechanisms, including direct reversal of streptozotocin alkylation by AlkB and AGT (alkylguanine alkyltransferase) homologs (Ng *et al.*, 2019), base excision of yatakemycin-adenine adducts by the DNA glycosylase YtkR2 (Xu *et al.*, 2012; Mullins *et al.*, 2021), nucleotide excision of DNA adducts of several intercalating agents including daunorubicin (Lomovskaya *et al.*, 1996), and putative replication-coupled repair of distamycin-DNA adducts (Ma *et al.*, 2020).

The AZB BGC in *Streptomyces sahachiroi* encodes a DNA glycosylase, AlkZ, that unhooks AZB-ICLs and that provides cellular resistance against AZB toxicity (Zhao *et al.*, 2008; Wang *et al.*, 2016). ICL unhooking by AlkZ involves hydrolysis of the N-glycosidic bonds of the crosslinked deoxyguanosine residues, producing abasic (AP) sites that can be repaired by the base excision repair pathway (Mullins *et al.*, 2019). AlkZ belongs to the relatively uncharacterized HTH_42 superfamily of proteins found in antibiotic-producing and pathogenic bacteria (Wang *et al.*, 2016). The crystal structure of AlkZ revealed a unique C-shaped architecture formed by three tandem winged helix-turn-helix motifs, with two catalytically essential glutamine residues within a QΦQ motif (Φ is an aliphatic residue) located at the center of the concave surface (Mullins *et al.*, 2017). We recently characterized a second HTH_42 protein from *Escherichia coli*, YcaQ, as a DNA glycosylase that excises several types of *N*7-alkylguanine ICLs and monoadducts using a catalytic QΦD motif and that functions as a secondary pathway to nucleotide excision repair for bacterial resistance to the nitrogen mustard, mechlorethamine (Bradley *et al.*, 2020).

The targeted discovery of natural products has been employed to search for novel scaffolds in plants, fungi, and bacteria, and can be useful for identifying specific classes of compounds (Kersten and Weng, 2018; Kjaerbolling *et al.*, 2019; Belknap *et al.*, 2020). Genome mining can be used to search for unidentified BGCs through analysis of core/accessory biosynthetic genes (PKS, NRPS, tailoring enzymes), comparative/phylogeny-based mining, regulatory genes, and more recently, resistance genes (Ziemert *et al.*, 2016). Some of these resistance-based mining approaches focus on the experimental screening of antibiotic resistance, while others rely on bioinformatic

tools to identify resistance genes within clusters based on homology to known resistance genes (Thaker *et al.*, 2013; Skinnider *et al.*, 2015; Blin *et al.*, 2019; Mungan *et al.*, 2020). However, many of these resistance-based methods have not been applied in bacteria for targeted discovery.

Here, we characterized the genomic differences of the HTH_42 proteins found in 435 species of *Streptomyces* to develop additional insight into this new family of DNA repair proteins, and applied this information in resistance-guided genome mining to characterize unknown BGCs or identify new genotoxins. We found that these proteins fall into two distinct subfamilies that are delineated by amino acid sequence, genomic context, and copy number. Proteins similar to *S. sahachiroi* AlkZ (AlkZ-like, AZL) are highly variable in sequence and enriched in BGCs, many producing known genotoxic alkylating agents. I show that the AZL protein within the BGC of the known DNA alkylating agent, hedamycin (HED), is a resistance DNA glycosylase specific for HED-guanine lesions, consistent with AZL-mediated DNA repair activity as a general self-resistance mechanism to genotoxins in antibiotic producers. Moreover, we found AZL proteins in BCGs that are either uncharacterized or that produce natural products not previously known to be genotoxic, validating resistance genome mining as an approach to discover new genotoxins. In contrast, *E. coli* YcaQ like (YQL) proteins are highly conserved in sequence and genetic neighborhood and are not associated with BGCs. We show that like *E. coli* YcaQ, two YQL enzymes from Actinobacteria have weaker substrate specificity than AZL proteins, suggesting a broader role of this subfamily of HTH_42 proteins outside of antibiotic self-resistance in bacteria.

YQL and AZL proteins in Streptomyces are evolutionarily distinct

E. coli YcaQ and *S. sahachiroi* AlkZ are the only characterized members of the HTH_42 superfamily and are unique in their ability to unhook ICLs and to provide cellular resistance to crosslinking agents. Both enzymes fully unhook ICLs derived from AZB (Fig. 38A). While AlkZ is specific for AZB-ICLs and is essential to the AZB-producing organism, YcaQ unhooks a broader range of ICLs, including those derived from the simple bifunctional alkylating agent mechlorethamine (Fig. 38B), and displays robust excision activity for *N*7-methylguanine (7mG) monoadducts (Wang *et al.*, 2016; Mullins *et al.*, 2017; Bradley *et al.*, 2020). YcaQ and AlkZ belong to one of five classes of HTH_42 proteins characterized by domain organization, which accounts for >95% of all HTH_42 proteins (SI Fig. S1A in (Bradley *et al.*, 2022)). Approximately two-thirds of the known HTH_42 proteins in prokaryotes are found in Actinobacteria, with ~25% of those sequences from *Streptomycetales* (SI Fig. S1B-C in (Bradley *et al.*, 2022)). The remainder are found in several different orders of Bacteria, and a very small number (12) in Archaea.

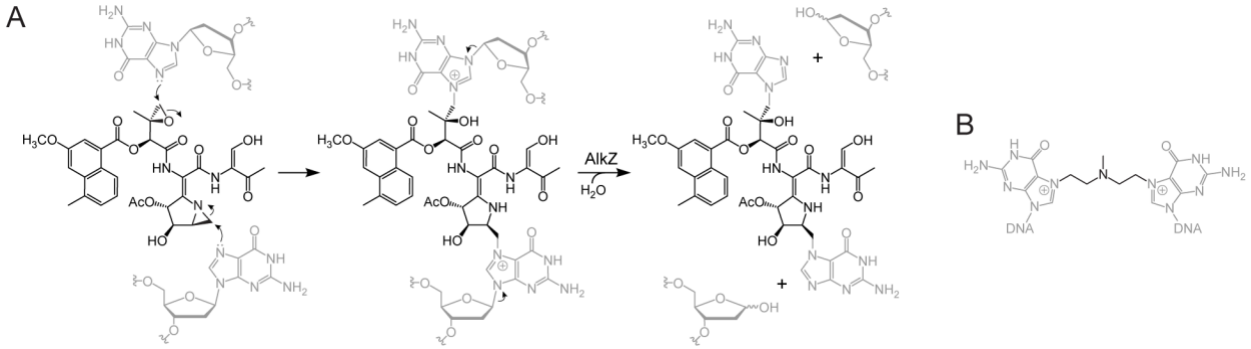


Figure 38. *Bifunctional ICL substrates of AZL and YQL enzymes and unhooking reaction by AlkZ* (A) Azinomycin B reacts with opposite strands of DNA to form an ICL, which is unhooked by AlkZ-catalyzed hydrolysis. (B) Structure of a nitrogen mustard ICL derived from mechlorethamine and unhooked by *E. coli* YcaQ.

To better understand the evolutionary and phylogenetic breadth of this superfamily in *Streptomyces*, we collected and analyzed all HTH_42 protein sequences from available genomes using a combination of BLAST searches against *Streptomyces* genomes in GenBank and HHMR protein domain searches of the BLAST hits against the Pfam database (SI Table S1 in (Bradley *et al.*, 2022)). Alignment of the 897 sequences showed that YQL and AZL proteins fall into distinct clades that represent 49% and 43% of the total number of sequences, respectively (Fig. 39A). The clades are defined in part by unique catalytic motifs QΦD (YQL) and (Q/H)ΦQ (AZL), where Φ is an aliphatic residue (Mullins *et al.*, 2017; Bradley *et al.*, 2020). YQL proteins show a high degree (>75%) of amino acid sequence conservation, whereas the AZL subfamily is more diverse, with only ~40% amino acid similarity on average. The differences in conservation are consistent with mutation rates as approximated by tip-to-root branch lengths (0.23 for YQL and 0.59 for AZL). In addition, we found that 8% of sequences do not fall into either YQL or AZL clades and contain a unique catalytic consensus sequence, HΦ(S/T)(D/E) (Fig. 39A-B). Because these sequences exhibit greater sequence similarity overall to AZL than YQL, we refer to this third homolog AZL2. Interestingly, AZL2 is more similar to YQL in its copy number and genomic location (see below), and thus is somewhat of a hybrid between AZL and YQL. I verified that proteins within the AZL2 clade contain bona fide DNA glycosylase activity, as the *S. caeruleatus* AZL2 protein excised 7mG from DNA in a similar manner to *S. sahachiroi* AlkZ (SI Fig. S1E in (Bradley *et al.*, 2022)).

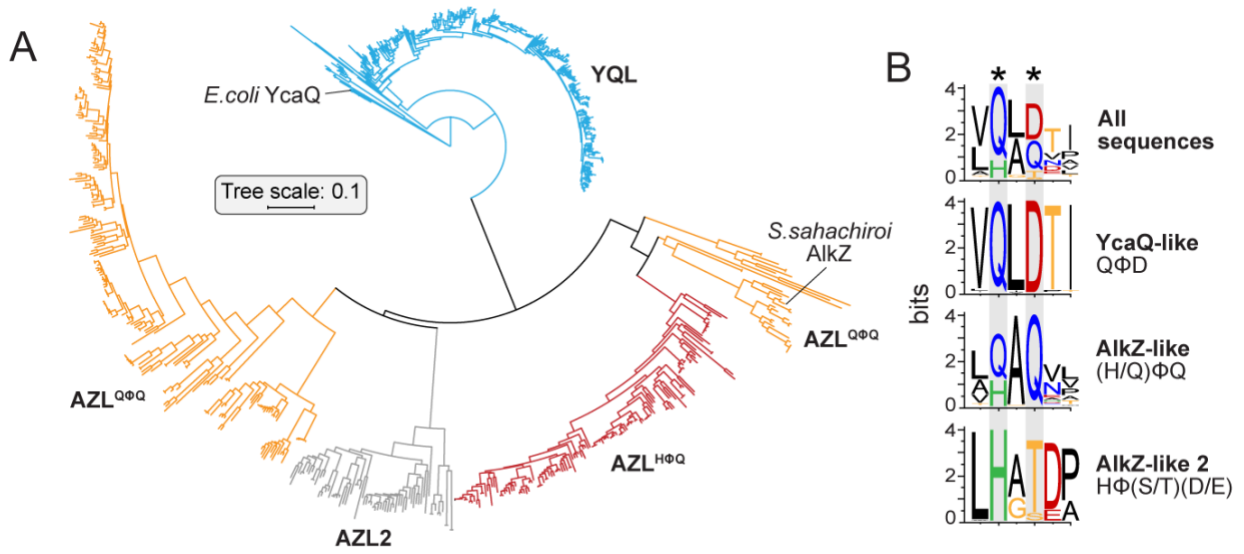


Figure 39. *Phylogenetic Organization of YQL/AZL Proteins in Streptomyces* (A) Phylogenetic tree of YcaQ-like (YQL, blue) and AlkZ-like (AZL, red/orange; AZL2, grey) *Streptomyces* proteins (n=897). The red and orange AZL clades distinguish HΦQ and QΦQ catalytic motifs. *E. coli* YcaQ and *S. sahachiroi* AlkZ proteins are labeled. (B) Sequence logos for the catalytic motifs in YQL, AZL, and AZL2 proteins. Catalytic residues are marked with asterisks. Colors correspond to side chain chemistry.

Another striking difference between the YQL and AZL families is that AZL genes are often found in multiple copies and in different combinations in many species of *Streptomyces*. The copy number differences between the different clades are significant, with the majority (90-95%) of YQL and AZL2 homologs found as a single copy and AZL mainly found in multiple (2-5) copies (Fig. 40A). The coincidence of YQL and AZL also varies. Although the most common combination is the presence of a copy of each YQL and AZL, many other combinations are observed (Fig. 40B). The number of species that contain both genes decreases as the copy number increases. For species containing either YQL or AZL (not both), the majority contain a single YQL copy, with just a few species having only AZL present. These results show that both YQL and AZL proteins

are broadly distributed across *Streptomyces* and are distinct with respect to sequence, diversity, and copy number.

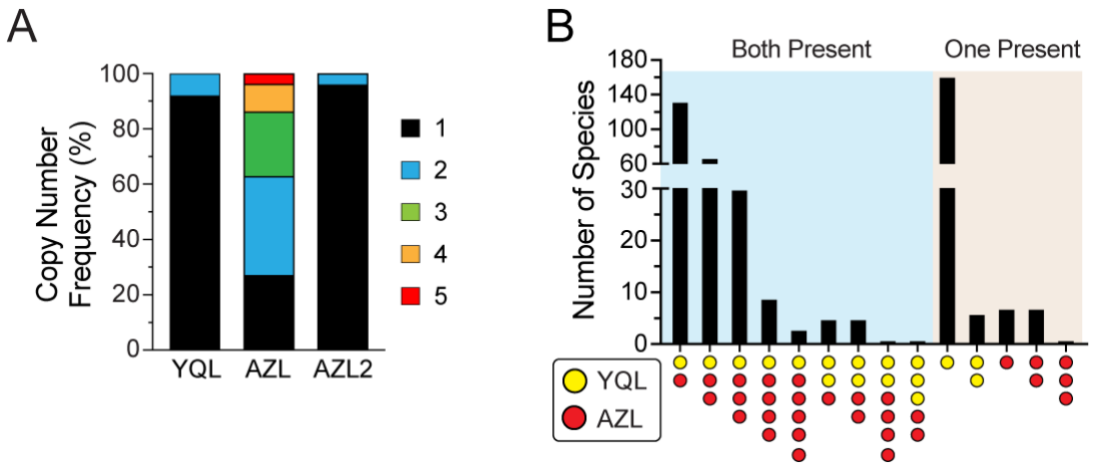


Figure 40. *Copy number and coincidence analysis of AZL/YQL proteins in Streptomyces* (A) AZL/YQL copy number frequency per *Streptomyces* genome, as a percentage of the total species analyzed (n=436 species, 897 sequences). One-way ANOVA significance (P) values of copy number variance are 0.0078 (YQL-AZL), 0.0033 (AZL-AZL2), and 0.3305 (YQL-AZL2), the latter of which is not significant. (B) YQL/AZL coincidence frequency. Blue shaded section represents species containing both YQL and AZL genes; tan shaded section represents species containing either YQL or AZL.

AZL proteins are prevalent in biosynthetic gene clusters

Given the distinct phylogeny of YQL and AZL proteins, we next examined their proximity to BGCs and characterized the identities of clusters containing a putative homolog. To perform this analysis, we identified all BGCs in the genomes of known *Streptomyces* species containing an HTH_42 protein, determined the most similar known cluster via BLAST, and extracted the distance in base pairs between the YQL/AZL gene and the nearest 3' or 5' end of each BGC (Fig. 41A, SI Table S2 in (Bradley *et al.*, 2022)). Strikingly, none of the 442 YQL genes localize to within 20 kb of the most proximal gene cluster in that organism (Fig. 41B). In contrast, AZL genes are primarily found inside or in

close genomic proximity to clusters, with an average distance of roughly 2.3 kb from the nearest BGC (compared to 25 kb for YQL). Despite their sequence similarity to AZLs, the AZL2 proteins are more like YQL in that they are also not observed within 20 kb of a BGC (SI Table S2 in (Bradley *et al.*, 2022)).

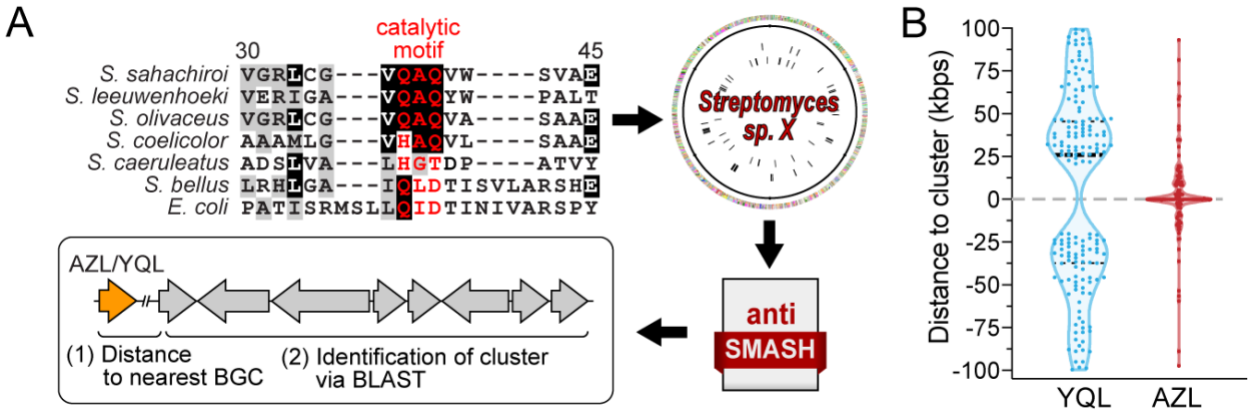


Figure 41. *Genome mining analysis and enrichment of AZL proteins in Streptomyces BGCs* (A) Schematic depicting the work flow for identification of YQL/AZL homologs in uncharacterized *Streptomyces* BGCs. Homologs were identified through the presence of the catalytic motif (red text in sequence alignment). The amino acid numbering is in relation to *S. sahachiroi* AlkZ. The corresponding *Streptomyces* genomes were input into antiSMASH, from which genomic distances between YQL/AZL genes and the nearest BGC, as well as homologous clusters were extracted. (B) Violin plot showing the distribution of distances of YQL (n=167) and AZL (n=154) homologs to the nearest BGC (in kbps; ±100 kb). The dotted line at 0 kb represents the 5' (+) / 3' (-) termini of the nearest BGC. Thick and thin dashed lines within the plot represent the median and upper/lower quartiles, respectively. The Chi-square significance (P) value between YQL and AZL data is less than 0.0001.

We found that AZL proteins are particularly enriched in uncharacterized *Streptomyces* BGCs, with 68 homologs localizing within a variety of different types of clusters (Fig. 42A-B; SI Table S3 in (Bradley *et al.*, 2022)). Almost half (n=32; 47%) localize to clusters resembling those producing known DNA damaging agents, including AZB (n=5), LL-D4919α1 (LLD, n=6), HED (n=4), ficellomycin/vazabitide A (n=5), and C-1027/leinamycin (n=2) (Terawaki and Greenberg, 1966; Reusser, 1977; Maiese *et al.*,

1990; Sugimoto *et al.*, 1990; Hansen *et al.*, 1995; Nooner *et al.*, 2004). In addition, several other clusters are related to potential DNA damaging agents on the basis of a reactive epoxide functional group in the natural product, including angucycline-like molecules (n=4) herboxidiene and asukamycin. The remaining 10 uncharacterized BGCs are related to clusters that produce macrolides/terpenes, tambromycin-like compounds, and various RiPPs/depsipeptides (Fig. 42A-B).

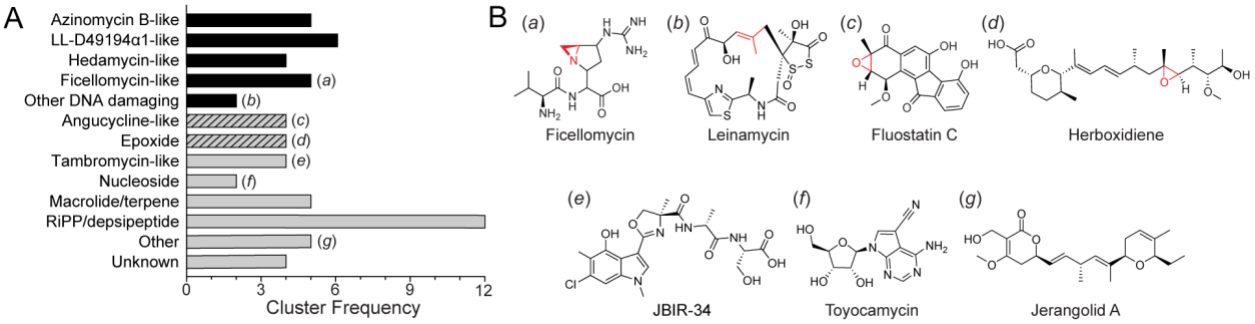


Figure 42. *Analysis of uncharacterized clusters which AZL proteins are present within* (A) Frequency of various types of BGCs in which AZL homologs were found (n=68 clusters identified). The y-axis denotes the natural product/scaffold type to which that cluster is most homologous. Black bars represent known DNA alkylators or DNA interacting metabolites, and hashed bars represent potential DNA damaging metabolites. Lowercase letters to the right of the bars correspond to structures shown in panel B. (B) Representative compounds corresponding to BGC types in panel A. Potential reactive sites are colored red. LL-D4919α1 and HED structures are shown in Fig. 44.

Bacterial genes of similar function or in a particular pathway are frequently clustered into neighborhoods or operons within the genome, and thus we investigated the nearest neighbors of *Streptomyces* YQL and AZL genes. We collected gene ontology (GO) terms describing the biological functions of the five nearest neighbors on either side of 40 YQL genes, 40 AZL genes inside BGCs, and 40 AZL genes outside BGCs, which collectively represent ~15% of the total of all homologs. Biological processes were

grouped into three categories—metabolism, signaling/cell function, and genetic information processing. Several key differences were found between the neighborhoods of AZL genes inside versus outside clusters (Fig. 43A; SI Fig. S2 in (Bradley *et al.*, 2022)). AZL genes within BGCs were more often found near terpenoid/polyketide/non-ribosomal protein synthesis and resistance/defense genes. The defense genes fell into several types: ABC transporters/permeases, α/β-fold hydrolases (VOC resistance proteins), DinB DNA-damage inducible hydrolases, and other AZL proteins. For those AZL genes found outside of BGCs, there is an abundance of neighbors involved in cell wall biosynthesis, cell cycle control, and signal transduction. In contrast, there were no significant differences between AZL neighbors involved in processing genetic information inside versus outside clusters (Fig. 43A). In contrast to the variation in the function of AZL gene neighbors, the functions of YQL neighbors (outside clusters) are nearly invariant, and are composed of a variety of different gene types with no apparent functional connection between them (Fig. 43B). The functions of many of these neighbors have not been elucidated in *Streptomyces*, but some are homologous to N-acetyltransferase, a two-component transcription factor/histidine kinase, and a DNA helicase (ComF) involved in transformation competence. Thus, both the sequences and the genomic neighborhoods of YQL proteins are relatively conserved and always found outside of BGCs, in contrast to the more variable copy number, sequence, and neighborhood of AZL genes prevalent within BGCs.

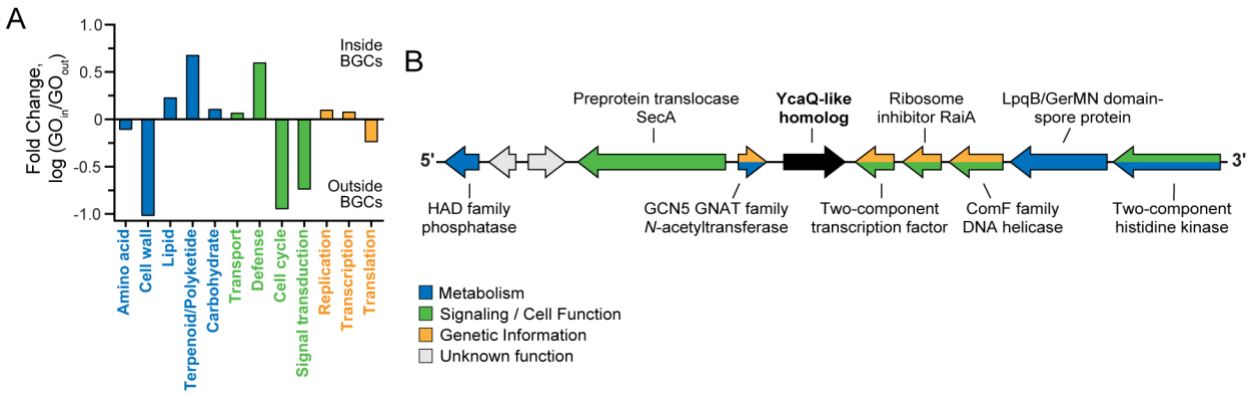


Figure 43. *Nearest neighbor gene ontology (GO) analysis of AZL and YQL proteins* (A) Nearest genes to AZL proteins found inside and outside clusters, shown as the ratio of GO terms inside and outside, and grouped by function (blue, metabolic; green, cell signaling and function; orange, genome maintenance). (B) Representative example from *Streptomyces griseoviridis* of nearest neighbor analysis for YQL homologs. Genes are colored according to function as in panel A (grey, unknown/hypothetical gene). These genes are invariant for all YQL homologs, with the exception of the outermost genes, in which only one instance of variance was observed.

Characterized BGCs containing AZL proteins

With the discovery that a significant proportion of AZL proteins reside within BGCs, we took a closer look at the nine characterized BGCs identified to contain an AlkZ homolog in the MIBiG database (SI Table S3 in (Bradley *et al.*, 2022)). Four of these produce known DNA-alkylating agents (Fig. 44A), which contain reactive epoxide moieties like AZB that are scaffolded on diverse natural product backbones (Fig. 44A). Whereas AZB is a bifunctional alkylating agent, HED, trioxacarcin A (TXNA), and LL-D49194α1 (LLD) are monofunctional alkylating agents that react with N7 of guanine in specific nucleotide sequences via their epoxide rings, and also intercalate the DNA helix via their planar ring systems (Hansen *et al.*, 1995; Pfoh *et al.*, 2008). TXNA and LLD clusters each contain two AlkZ paralogs (TxnU2/U4 and LldU1/U5), whereas the HED cluster contains one (HedH4) that resides between the two polyketide synthase genes.

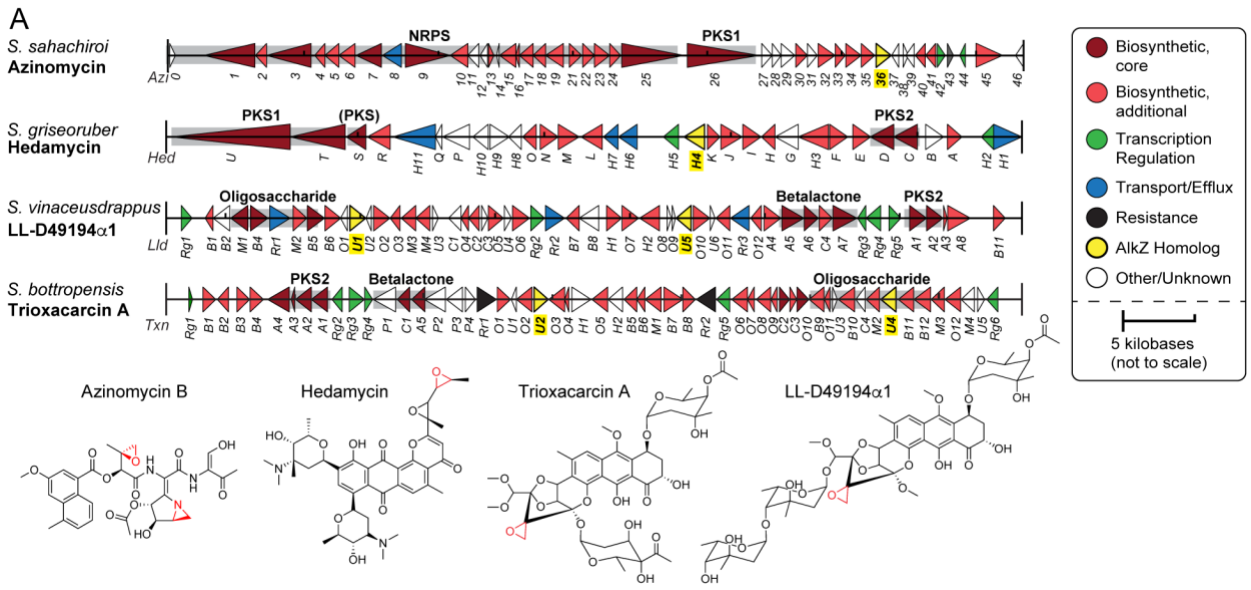


Figure 44. *AZL homologs found in alkylating genotoxin characterized BGCs* (A) Gene diagrams for AZL-containing BGCs producing DNA alkylating agents. Gene names are labeled below the cluster diagrams. The biosynthetic scaffold produced by specific genes in the cluster are labeled above the respective genes. NRPS, non-ribosomal peptide synthetase; PKS1/PKS2, type 1/2 polyketide synthase; (PKS), PKS-like. Chemical structures of the metabolites produced by each cluster are shown at the bottom of each panel. Reactive portions of the molecules are highlighted in red.

The remaining five AZL-containing clusters in MIBiG produce compounds that are not known to alkylate DNA, but that share some structural characteristics with the alkylating agents described above (Fig. 45A). Aclacinomycin contains an anthracycline core surrounded by sugars that allow it to intercalate into DNA and act as a topoisomerase I poison, potentially generating downstream DNA damage (Nitiss *et al.*, 1997). Asukamycin contains a modified PKS scaffold and an electrophilic epoxide ring and has been shown to act as both a farsenyltransferase inhibitor and a molecular glue between the UBR7 E3 ubiquitin ligase and the TP53 tumor suppressor, leading to cell death (Hara *et al.*, 1993; Isobe *et al.*, 2020). Armeniaspirol contains a unique chlorinated pyrrole and inhibits the AAA+ proteases ClpXP and ClpYQ leading to cell division arrest in Gram

positive bacteria (Labana *et al.*, 2021). The other two BGCs produce compounds of known structure but unknown function—tambromycin and JBIR-34/35 are similar NRPS compounds containing densely substituted chlorinated indole and methyloxazoline moieties (Muliandi *et al.*, 2014). The presence of AZL proteins in these clusters suggests that these compounds may be genotoxins or otherwise react with DNA, and/or that these particular AZL homologs may have a function outside of DNA repair.

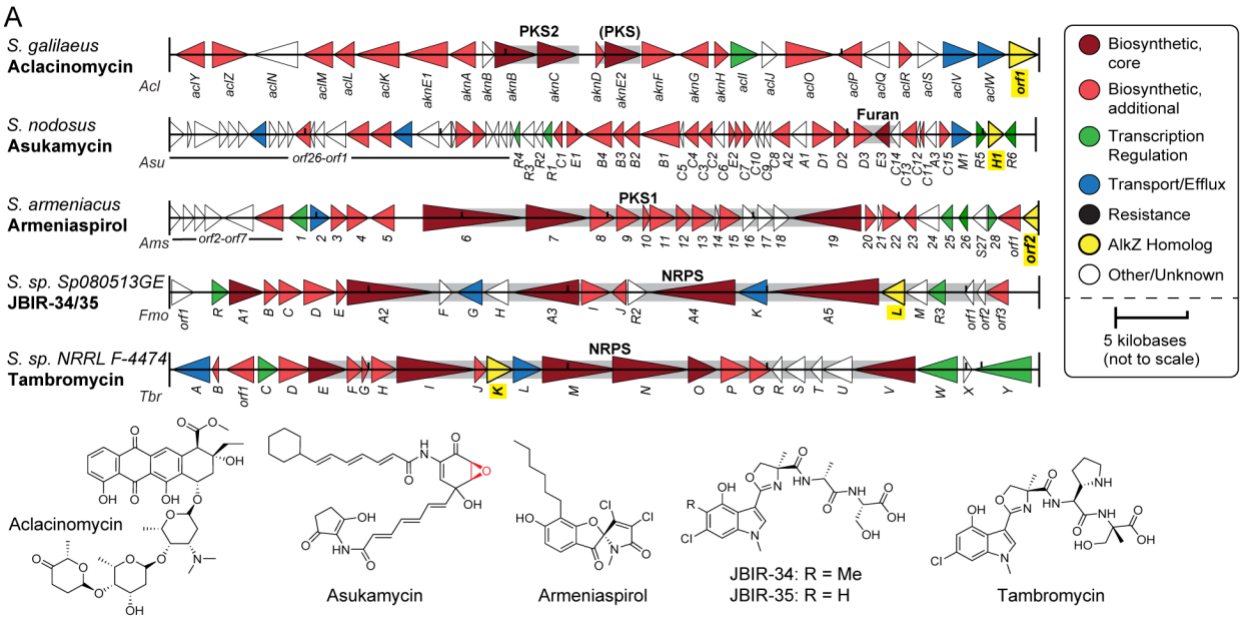


Figure 45. *AZL homologs found in non-alkylating characterized BGCs* (A) Gene diagrams for AZL-containing BGCs producing compounds not known to alkylate DNA. Gene names are labeled below the cluster diagrams. The biosynthetic scaffold produced by specific genes in the cluster are labeled above the respective genes. NRPS, non-ribosomal peptide synthetase; PKS1/PKS2, type 1/2 polyketide synthase; (PKS), PKS-like. Chemical structures of the metabolites produced by each cluster are shown at the bottom of each panel.

The AZL protein within the HED BGC is a DNA glycosylase specific for HED-DNA lesions and provides cellular resistance to HED toxicity

The *alkZ* gene embedded within the AZB BGC provides exquisite resistance to the potent cytotoxicity of this natural product (Zhao *et al.*, 2008; Wang *et al.*, 2016). To determine if AlkZ homologs other than in the AZB BGC provide self-resistance to their cognate natural products, I characterized the DNA glycosylase and cellular resistance activities of HedH4 for HED-DNA adducts. HED is a potent antibiotic/antitumor agent that induces a strong DNA damage response (Tu *et al.*, 2004). The bisepoxide side chain alkylates the N7 position of guanines in 5'-(C/T)G sequences (Fig. 46A), the highly oxidized aromatic polyketide intercalates the DNA helix, and two C-glycosidic linked aminosugars interact with the minor groove (Hansen *et al.*, 1995). I generated site-specifically labeled HED-guanosine adducts in DNA by reacting purified compound with an oligonucleotide containing a HED target sequence, d(TGTA). The HED-DNA adduct was stable relative to other *N*7-alkylguanine lesions as judged by thermal depurination (SI Fig. S3B in (Bradley *et al.*, 2022)) (Hemminki *et al.*, 1989; Bradley *et al.*, 2020). I first assessed the ability of purified HedH4 to hydrolyze HED-DNA using a gel-based glycosylase assay that monitors alkaline cleavage of the AP site product (Mullins *et al.*, 2017; Bradley *et al.*, 2020). Reaction of HedH4 with HED-DNA followed by hydroxide work-up resulted in β- and δ-elimination products consistent with production of an AP site from DNA glycosylase-mediated excision of the N-glycosidic bond of the HED-guanosine nucleotide (Fig. 46A-B). I verified the identity of the excision product as HED-guanine by HPLC/MS (Fig. 46C). To verify that the HED-guanine product was not generated by a contaminating enzyme and to examine the conservation of the catalytic QΦQ motif, I purified alanine mutants of the two glutamine residues and tested their activity under single-turnover conditions (Fig. 46D, SI Fig. S3A,C in (Bradley *et al.*, 2022)). The

calculated rate constant (k_{cat}) for wild-type HedH4 was at least 7.8 ± 0.5 min^{-1} (the reaction was complete at the earliest time point). Relative to wild-type, the Q41A mutant was at least 225-fold slower (k_{cat} = 0.04 ± 0.01 min^{-1}) and the Q43A mutant at least 10-fold slower (k_{cat} = 0.8 ± 0.2 min^{-1}), indicating that both Gln residues in the HedH4 QΦQ play a role in HED-guanine excision.

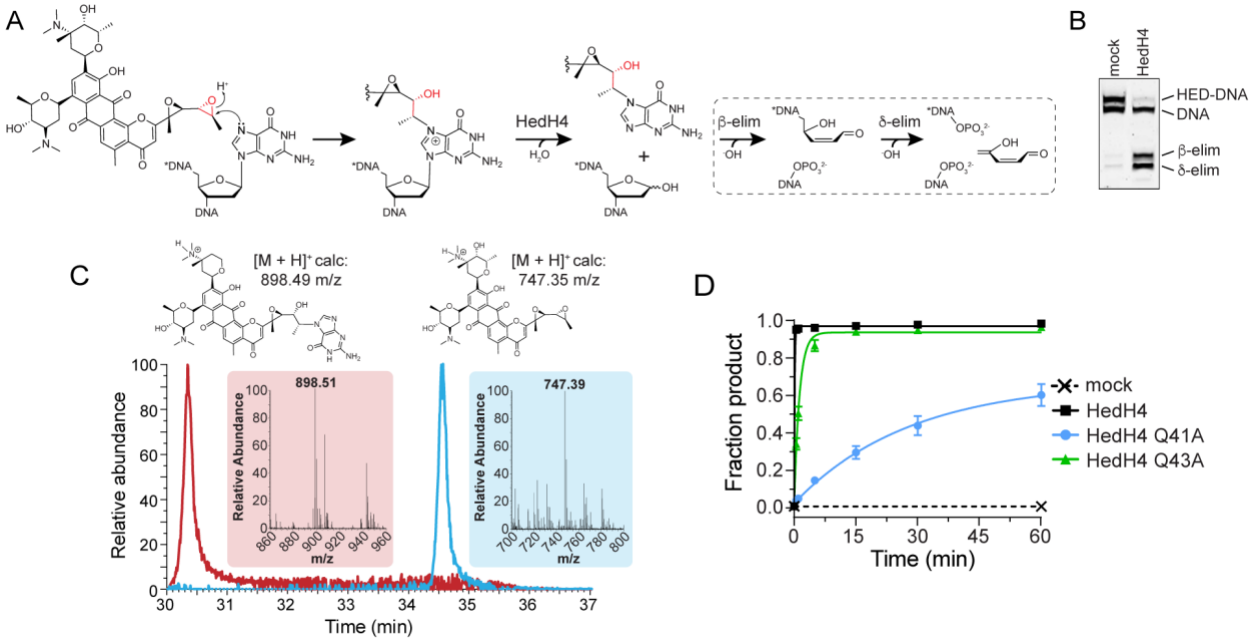


Figure 46. *HedH4 rapidly excises HED-guanine adducts from DNA* (A) HED modification of deoxyguanosine in DNA forms a HED-DNA adduct that is hydrolyzed by HedH4 to generate an abasic (AP) site in the DNA and free Hed-guanine. The reactions within the dashed line are not catalyzed by HedH4. The AP nucleotide is susceptible to base-catalyzed nicking to form shorter DNA products containing either a 3′-phospho-α,β-unsaturated aldehyde (PUA, β-elimination) or a 3′-phosphate (δ-elimination). The asterisk (*) denotes the original 5′-end of the DNA. (B) Denaturing PAGE of 5'-Cy5-labeled HED-DNA substrate and β- and δ-elimination products after treatment with enzyme or buffer (mock) for 1 hr, followed by NaOH to nick the AP site. The HED-DNA reaction only goes to ~50% completion under our reaction conditions, as shown by the two bands of equal intensity in the mock reaction. (C) HPLC-MS analysis of HED (blue) and the HED -guanine excision product from reaction of HedH4 and HED-DNA (red). Axis represents elution time (x-axis) versus relative abundance from total ion count (y-axis). Insets show mass spectra of each elution peak. (D) Wild-type and mutant HedH4 glycosylase activity for HED-DNA. Data are mean ± SD (n=3). Curves were fit to a single exponential. Representative data are shown in SI Fig. S3C in (Bradley *et al.*, 2022).

I probed specificity of HedH4 for HED-DNA adducts, first by asking whether the HED-guanosine lesion was a substrate for other bacterial alkylpurine DNA glycosylases with varying specificities. *E. coli* AlkA and YcaQ and *Bacillus cereus* AlkC and AlkD excise a relatively broad range of alkyl-DNA adducts (O'Brien and Ellenberger, 2004; Alseth *et al.*, 2006; Brooks *et al.*, 2013; Parsons *et al.*, 2016; Mullins *et al.*, 2017; Shi *et al.*, 2018; Bradley *et al.*, 2020). *S. sahachiroi* AlkZ, *S. bottropensis* TxnU2 and TxnU4, and *S. vinaceusdrappus* LldU1 and LldU5, like HedH4, are found in BGCs that produce bulky *N*7-alkyl- and intercalating DNA adducts (Fig. 44A), and each is specific for their cognate toxin (Bradley *et al.*, 2020; Chen *et al.*, 2022). Compared to HedH4, which excises 100% of the HED-guanine from DNA, none of the ten alkylpurine DNA glycosylases tested showed any appreciable activity for HED-DNA after 1 hour (Fig. 47A). Thus, the HED-DNA adduct is hydrolyzed only by the glycosylase found in the HED BGC. We next examined the ability of HedH4 to excise *N*7-alkylpurine lesions that act as substrates for other YQL and AZL enzymes. Interestingly, HedH4 showed no significant activity for the simple methyl adduct 7mG, which is removed by most alkylpurine DNA glycosylases including *E. coli* YcaQ and *S. sahachiroi* AlkZ (Fig. 47B, left). HedH4 was also unable to hydrolyze TXNA-guanosine, a substrate for TxnU4 from the TXNA BGC (Fig. 47B, right) (Chen *et al.*, 2022). I also tested the ability of HedH4 to unhook ICLs derived from AZB (Fig. 38A) and an 8-atom nitrogen mustard, NM_8 (Fig. 47C), which are substrates for *S. sahachiroi* AlkZ and *E. coli* YcaQ, respectively. Compared to AlkZ and YcaQ, HedH4 showed little to no activity for either ICL (Fig. 47D). Thus, HedH4 is highly specific for DNA adducts derived from its cognate natural product.

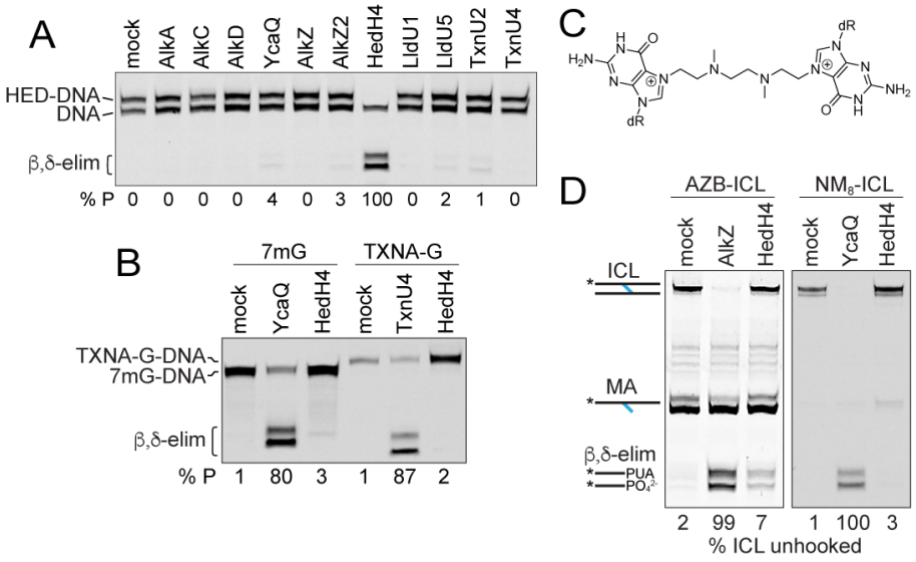


Figure 47. *HedH4 and HED-DNA specificity analysis* (A) Denaturing PAGE of HED-DNA adducts after 1 hr incubation with either buffer (mock) or bacterial alkylpurine-DNA glycosylases. (B) Denaturing PAGE of 1 hr reaction products of *E. coli* YcaQ and HedH4 with 7mG-DNA (left), and *S. bottropensis* TxnU4 and HedH4 with TXNA-DNA (right). (C) Structure of NM_8-ICL. (D) Denaturing PAGE of AZB-ICL unhooking by *S. sahachiroi* AlkZ and HedH4 (left), and NM_8-ICL unhooking by *E. coli* YcaQ and HedH4 (right). Reactions were treated with buffer (mock) or enzyme for 1 hr, followed by alkaline hydrolysis. MA, monoadduct.

I next tested if the *hedH4* gene provides heterologous resistance to HED cytotoxicity in cells. *E. coli* transformed with either vector containing *hedH4* constitutively expressed at low levels or vector alone were grown in the presence of increasing amounts of HED (SI Fig. S3D-G in (Bradley *et al.*, 2022)). HedH4 provided a modest protection against HED, as cells expressing HedH4 grew to a higher density at all HED concentrations (Fig. 48A, SI Fig. S3F-G in (Bradley *et al.*, 2022)) and had a higher IC_{50} than cells treated with vector alone (HedH4, 5.9 μM ± 0.7; vector, 3.9 μM ± 0.4). The sensitivity differences between HedH4 and vector control were more pronounced from a colony dilution assay performed under log-phase growth conditions (Fig. 48B). Cells expressing empty vector displayed an IC_{50} value of 11.1 μM ± 1.5, while cells expressing

HedH4 displayed a 4-fold reduction in sensitivity to HED (48.1 μM ± 13.8). These results indicate that HedH4 is a DNA glycosylase specific for HED-DNA adducts and provides resistance to cells exposed to the antibiotic.

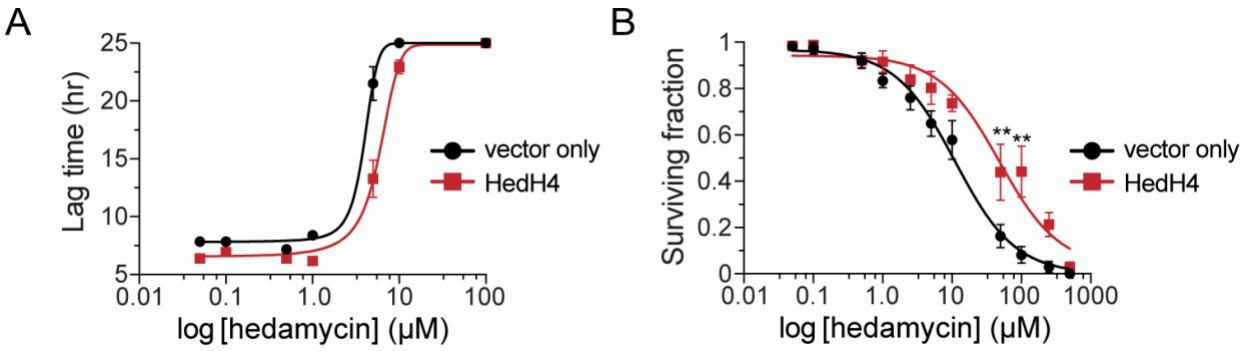


Figure 48. *HedH4 provides heterologous resistance against hedamycin toxicity* (A) HED inhibition of *E. coli* K-12 transformed with *hedH4*/pSF-OXB1 (constitutively expressed) or empty vector pSF-OXB1. The lag time is defined as the time elapsed before cells start to grow exponentially. Data are mean ± SD (n=3). Growth curves are shown in SI Fig. S3F-G in (Bradley *et al.*, 2022). Significance values were determined by unpaired t test of the mean lag time values (*: 0.05 ≤ P ≤ 0.01; ***: 0.001 ≤ P ≤ 0.0001). (J) Colony dilution assay for *E. coli* strains ±HedH4 exposed to increasing concentrations of HED for 1 hr. Surviving fraction (%) is relative to untreated cells. Values are mean ± SD (n = 3). Significance values were determined by unpaired t test of the mean sensitivity values (*: 0.05 ≤ P ≤ 0.01; **: 0.01 ≤ P ≤ 0.001).

YQL proteins from Actinobacteria hydrolyze simple N7-alkylguanosine lesions and interstrand crosslinks

We previously characterized *E. coli* YcaQ to have robust activity toward 7mG and NM-ICLs (Figs. 38B and 47C), a substrate preference distinct from AZB- and HED-specific *S. sahachiroi* AlkZ and HedH4 (Fig. 47A-B,D) (Bradley *et al.*, 2020). I therefore was interested in determining if other proteins of the YQL subfamily were functional YcaQ orthologs. I purified YQL proteins from the Actinobacteria *Thermomonospora curvata* (Tcu) and *Thermobifida fusca* (Tfu) and tested their ability to hydrolyze 7mG and unhook

NM_8-ICLs (Fig. 49A-B). Both proteins showed significant activity for both substrates, providing evidence that the YQL subfamily in general has comparable specificity for simple *N*7-alkylguanine lesions, distinguishing it biochemically from the AZL subfamily.

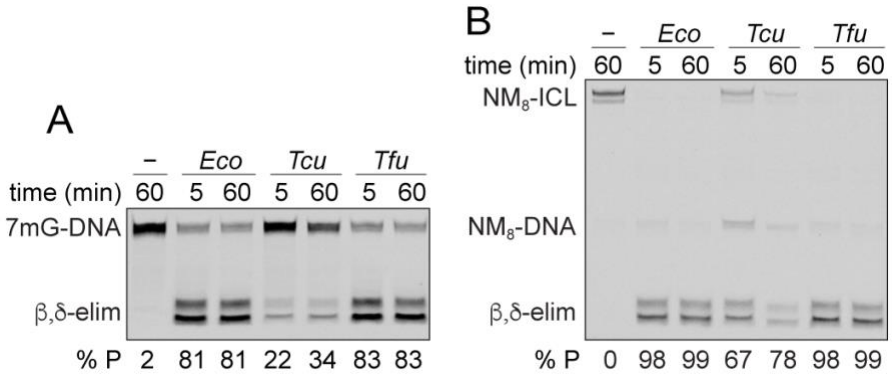


Figure 49. *YQL proteins from Actinobacteria hydrolyze simple N7-alkylguanosine lesions and interstrand crosslinks* (A,B) Denaturing PAGE of reaction products of YQL orthologs from *E. coli* (*Eco*), *Thermomonospora curvata* (*Tcu*), and *Thermobifida fusca* (*Tfu*) with 7mG-DNA (A) and NM_8-ICL (B) after 5 min and 1 hr. Lane 1 of each gel is a no-enzyme control.

Discussion

Phylogenetic characterization of the HTH_42 superfamily proteins within *Streptomyces* reveals two distinct subfamilies, YQL and AZL (the latter of which contains the AZB2 clade). Most strikingly, AZL genes, which are most prevalent in environmental microbes such as those from the phylum Actinobacteria (SI Fig. S1B in (Bradley *et al.*, 2022)), are highly enriched in BGCs. We found AZL proteins in BGCs that produce a variety of verified and putative genotoxins, with approximately one-fifth of all AZL proteins located in BGCs predicted to produce a DNA alkylating agent. I show that the AZL protein, HedH4, within the HED cluster specifically excises HED-DNA adducts and improves viability of cells grown in the presence of the compound. In a separate study, we recently

found that the two paralogs present in TXN and LLD clusters (TxnU2, TxnU4, LldU1, LldU5) are self-resistance glycosylases for these compounds (Chen *et al.*, 2022). Thus, together with the previous example from the AZB BGC (Wang *et al.*, 2016; Bradley *et al.*, 2020), there is now mounting evidence that AZL family genes have evolved largely as DNA repair self-resistance proteins against a variety of natural products. Consistent with their role in resistance, the AZL genes found inside BGCs frequently localize around a variety of other resistance genes. Moreover, the relatively high copy number and low sequence conservation of AZL proteins are consistent with increased expression or possible horizontal gene transfer events that enable these enzymes to evolve specificity for particular natural product (Hastings *et al.*, 2009). We also found AZL proteins in BGCs that by homology were not expected to produce DNA alkylators or other genotoxins. The AZL proteins in these clusters could have regulatory or protective roles outside of DNA repair. Alternatively, these clusters could have additional uncharacterized enzymes such as cytochrome P450s, sulfate adenyltransferases, or epoxidases that could convert the natural products into DNA alkylators (Thibodeaux *et al.*, 2012).

The fate of the AP sites generated by AZL enzymes is a key unanswered question regarding glycosylase-mediated self-resistance in antibiotic bacteria. While the DNA adducts of AZB and HED natural products would likely pose significant blocks to replication and transcription, their excision by AZL glycosylases also generate AP sites, which are toxic BER intermediates (Tomicic and Franekic, 1996; Posnick and Samson, 1999). Although the modest protection we observed from HedH4 overexpression in HED-challenged *E. coli* could be a result of the weak-expression promoter used, it also suggests that either the AP sites generated are poor substrates for the AP endonucleases

present in *E. coli* or that HED-DNA adducts are substrates for an alternative repair pathway. The intercalated HED-DNA adduct likely poses a unique challenge relative to other glycosylase substrates. It is likely that the HedH4-generated HED-guanine moiety remains intercalated at the AP site and requires a specialized AP endonuclease for repair. Indeed, we recently found that the excised guanine adduct of the related, intercalating natural product TXNA is a poor substrate for *E. coli* EndoIV (Chen *et al.*, 2022). More pertinent to HED biosynthesis, the producing organism *S. griseoruber* contains two copies each of ExoIII- and EndoIV-like AP endonucleases that may have evolved to incise HED AP-sites, although none are located in the *hed* BGC. In addition, the bulky HED-DNA addicts lesions are likely substrates for nucleotide excision repair pathway, which is initiated by UvrA in bacteria and has been shown to play an important role in natural product self-resistance (Lomovskaya *et al.*, 1996; Jin *et al.*, 2001; Kiakos *et al.*, 2007; Mullins *et al.*, 2017; Burby and Simmons, 2019). Indeed, within the HED BGC there is a predicted UvrA-like drug resistance protein (HedH11) that contains a partial UvrA DNA-binding domain and a conserved ABC transporter domain that could initiate NER of HED-guanosine adducts or even HED-guanine/AP-site products generated by HedH4. There are also two additional putative UvrA homologs outside of the *hed* cluster. Additionally, there are three putative transporters within the cluster—HedH7 (ABC2 type), HedH6 (DrrA-like) and HedH1 (EmrB/QacA antiporter)—which could serve to physically bind to HED and direct it out of the cell through a transmembrane transport system.

In contrast to the genotoxin-specific AZL genes, YQL and AZL2 are always found outside clusters and thus are likely to provide a more general role in protecting the genome against environmental genotoxins, similar to that shown for *E. coli* YcaQ (Bradley

et al., 2020). YQL proteins and their gene neighborhoods are very highly conserved, suggesting they play a critical role as part of a unified pathway (Rogozin *et al.*, 2002). Although that pathway is unknown, the presence of a two-component transcription factor/kinase and ComF DNA helicase within the YQL neighborhood in *Streptomyces* also hints at a signaling network for DNA uptake (Londono-Vallejo and Dubnau, 1993; Turgay *et al.*, 1998; Veening and Blokesch, 2017). Similarly, *E. coli* YcaQ is localized in a four-gene operon involved in cell wall biosynthesis and transformation competence (Bradley *et al.*, 2020). Continued exploration of the gene neighborhoods of YQL and AZL beyond *Streptomyces* will reveal a deeper understanding of the cellular roles played by these enzymes. This will be especially important for YQL, which are prevalent in human pathogens or commensal microbes (Wang *et al.*, 2016).

A small subset of HTH_42 proteins contain additional domains often associated with nucleic acid transactions (SI Fig. S1A in (Bradley *et al.*, 2022)) (Wang *et al.*, 2016). These multimodular HTH_42 proteins have been relatively understudied, although they do not appear to be associated with BGCs. Most contain an associated DEAD box helicase domain, including Lhr, a member of the helicase superfamily II (Reuven *et al.*, 1995). *Mycobacterium smegmatis* and *E. coli* Lhr have been characterized as ATP-dependent 3′→5′ ssDNA translocases with the ability to unwind RNA-DNA hybrids (Ordonez and Shuman, 2013; Warren *et al.*, 2021). Studies in *Mycobacterium tuberculosis* have demonstrated a strong transcriptional activation of *lhr* in cells exposed to MMC (Boshoff *et al.*, 2003), suggesting that Lhr may function as an RNA-DNA helicase in response to MMC-DNA crosslinks. While the structure of the C-terminal HTH_42 domain of *M. smegmatis* Lhr is similar to AlkZ, it lacks the catalytic QΦQ motif and adopts

a tetrameric structure that occludes the putative DNA binding surface (Warren *et al.*, 2021). Thus, the function of the Lhr HTH_42 domain and its interplay with the helicase core remains to be determined.

Resistance genome mining has emerged as a critical bioinformatically driven pipeline to discover novel natural products and gene clusters in several organisms (Panter *et al.*, 2018; Yan *et al.*, 2018). A key benefit of resistance genome mining is the dramatically decreased candidate pool as a result of targeted identification of gene clusters containing a resistance gene. Generally, these methods require a basic understanding of the resistance mechanisms involved. We sought to use this approach for the first time to hunt for BGCs that produce alkylating genotoxins, using prior knowledge of the DNA repair functions of *S. sahachiroi* AlkZ within the AZB cluster (Wang *et al.*, 2016; Mullins *et al.*, 2017; Bradley *et al.*, 2020). In this study, we examined 435 *Streptomyces* species for BGCs within which an AlkZ-related gene was located and found 62 uncharacterized clusters that are candidates for targeted elucidation of their products. Characterization of these orphan clusters could provide new analogs or types of DNA alkylating/damaging secondary metabolites, an important step in developing new antitumor or antibiotic treatments. This classification of YQL/AZL proteins in *Streptomyces* is an important first step in understanding their evolutionary connection to each other and to BGCs of different types, and demonstrates that targeted resistance genome mining is a viable approach to discover novel genotoxins and resistance mechanisms from uncharacterized BGCs.

[4] This work is published in: Bradley, N.P.[§], Wahl, K.L.[§], Steenwyk, J.L., Rokas, A., and Eichman, B.F. (2022). Resistance-guided mining of bacterial genotoxins defines a family of DNA glycosylases. *mBio* (*In Press*). I designed/conducted experiments, wrote, and edited the manuscript. [§]authors contributed equally.

Chapter 5

Base Excision Repair System Targeting DNA Adducts of Trioxacarcin/LL-D49194 Antibiotics for Self-Resistance[5]

Introduction

Genome stability and integrity are continually challenged by both intrinsic and extrinsic genotoxic agents that generate a diversity of DNA damage through oxidation, alkylation, or hydrolytic deamination (Gates, 2009). Among the most common forms of damage are those derived from alkylating agents, which can potentially modify any of the heteroatoms in duplex DNA. Different sites are alkylated depending on the nature of the DNA-alkylating agents. The resulting DNA damage—including single or double strand breaks, inter- or intra-strand crosslinks, base detachment and base modification—interferes with normal cellular processes, causing DNA mutations, chromosomal rearrangements and instability, which can contribute to heritable diseases and even cell death (Jackson and Bartek, 2009; Terabayashi T and Hanada, 2018). Due to their cytotoxicity, DNA damaging agents often possess certain antimicrobial or antitumor activities, and some of them are used extensively as drugs in cancer treatment (Verweij and Pinedo, 1990; de Wit *et al.*, 1997; Deans and West, 2011; Brulikova *et al.*, 2012; Singh *et al.*, 2018).

In the cell, DNA damage is repaired by several highly conserved pathways (Jackson and Bartek, 2009). Alkylated DNA is eliminated from the genome predominantly by direct reversal, base excision repair (BER), or nucleotide excision repair (NER)

pathways (Fromme and Verdine, 2004; Truglio *et al.*, 2006; Hitomi *et al.*, 2007; Nouspikel, 2009; Yi and He, 2013). Direct reversal enzymes (e.g., alkylguanine DNA alkyltransferases and AlkB-family dioxygenases) extract alkyl substituents from the nucleobase to leave the nucleotide and DNA backbone intact, and can remove not only small base modifications, but also inter-strand DNA crosslinks and bulky exocyclic DNA adducts (Pegg and Byers, 1992; Wang *et al.*, 1997; Fang *et al.*, 2008). BER also removes mainly small but also some bulky and crosslinked adducts (Krokan and Bjoras, 2013; Schermerhorn and Delaney, 2014; Mullins *et al.*, 2019), and is initiated by DNA glycosylases that liberate a single modified nucleobase from the DNA backbone through hydrolysis of the N-glycosidic bond (Stivers and Jiang, 2003; Brooks *et al.*, 2013; Drohat and Maiti, 2014; Drohat and Coey, 2016; Mullins *et al.*, 2019). This reaction forms an apurinic/apyrimidinic (AP, or abasic) site that is then incised by an AP endonuclease (e.g., Exonuclease III (*Xth*) or Endonuclease IV (EndoIV, *Nfo*) in bacteria), generating a gap in the DNA backbone. In contrast, the NER pathway removes bulky or duplex-distorting lesions by endonuclease-catalyzed incisions that isolate a lesion-containing DNA oligonucleotide (Truglio *et al.*, 2006; Wood, 2010). DNA gaps generated in BER and NER are processed, filled, and sealed by the action of a DNA polymerase and DNA ligase.

Recent studies of self-resistance mechanisms against genotoxic natural products revealed that several unrelated glycosylases participate in removing bulky adducts (Xu *et al.*, 2012; Wang *et al.*, 2016). Among them, the DNA glycosylase AlkZ, derived from *Streptomyces sahachiroi* and which resides within the biosynthetic gene cluster (BGC) of the natural product azinomycin B (AZB), repairs interstrand crosslink (ICL) damage generated by AZB (Wang *et al.*, 2016; Mullins *et al.*, 2017; Bradley *et al.*, 2020). AZB is a

bifunctional alkylating agent that forms ICLs in the major groove by linking the N7 nitrogens of purines in the duplex DNA sequence 5′-d(PuNPy)-3′ (Armstrong, 1992). AlkZ unhooks AZB-ICLs by cleaving the N-glycosidic bonds of both modified nucleotides, resulting in AP sites that can be processed by the BER pathway (Wang *et al.*, 2016; Mullins *et al.*, 2019). The crystal structure revealed that AlkZ adopts a C-shaped structure in which the concave channel contains a QΦQ motif essential for catalytic activity and a β-hairpin predicted to contact the lesion in the minor groove (Mullins *et al.*, 2017). AlkZ belongs to the uncharacterized HTH_42 superfamily of proteins widespread in antibiotic producers and pathogenic bacteria (Wang *et al.*, 2016). To date, the only other bacterial DNA glycosylase characterized as an ICL glycosylase is another HTH_42 protein, *Escherichia coli* YcaQ, which has a relaxed specificity relative to *S. sahachiroi* AlkZ and can cleave *N*7-linked nitrogen mustard (NM) ICLs and *N*7-methyl-2′-deoxyguanosine (7mG) monoadducts (Bradley *et al.*, 2020).

Trioxacarcins (TXNs) are densely oxygenated, polycyclic aromatic, and structurally complex natural products with potent cytotoxicity (Fig. 50A) (Tamaoki *et al.*, 1981; Tomita *et al.*, 1981; Maiese *et al.*, 1990; Maskey *et al.*, 2004). Trioxacarcin A (TXNA) and LL-D49194 (LLD), two of the most representative compounds in the TXN family, intercalate the base pairs of DNA and have reactive epoxide moieties that covalently alkylate the N7 of guanine in d(GT) dinucleotides, forming stable DNA lesions that impair normal cellular processes (Fitzner *et al.*, 2008; Pfoh *et al.*, 2008). Consequently, TXNA and LLD exhibit remarkable antimalarial, antibacterial, and antitumor activity (Tamaoki *et al.*, 1981; Maiese *et al.*, 1990; Maskey *et al.*, 2004). The TXNA analog gutingimycin (Fig. 50B), which contains a TXN skeleton and a guanine

(Gua) group, has been isolated from the fermentation broth of a marine *Streptomycete* (Maskey *et al.*, 2004). Given that TXNA and LLD are alkylating agents that selectively modify deoxyguanosine (G) to form DNA adducts, we speculated that the biosynthetic pathways of the two natural products should contain DNA glycosylases responsible for cleaving TXNA/LLD-DNA, in which gutingimycin and LLD-Gua are the resulting products (Fig. 50C). Therefore, we became interested in the DNA damage repair mechanism targeting TXNs family of DNA alkylating agents.

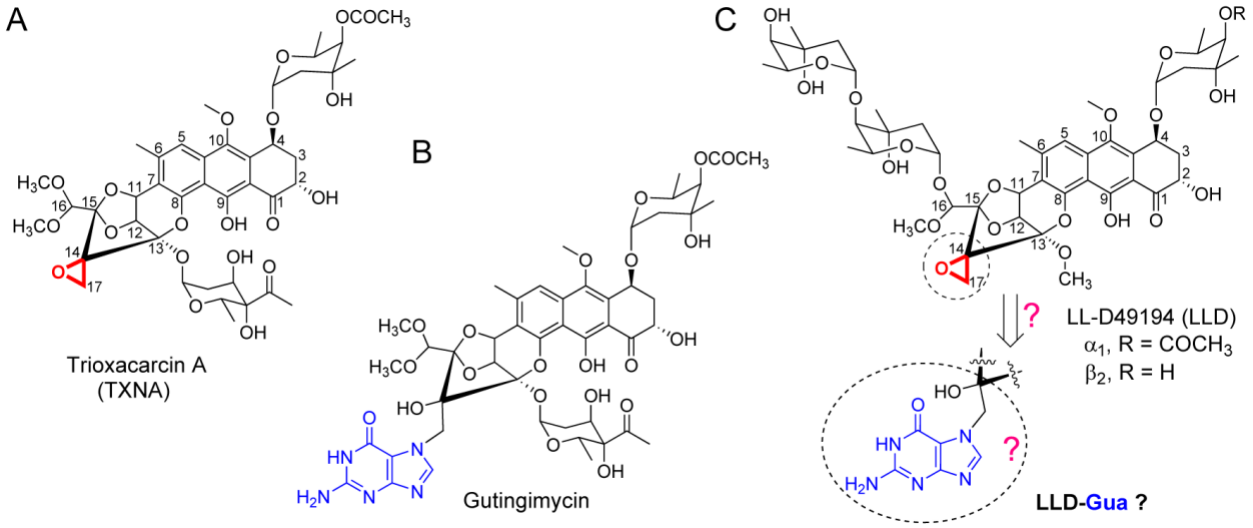


Figure 50. *Structures of TXNs family compounds and related metabolites* (A-C) Structures of trioxacarcin A (TXNA) (A), gutingimycin (B), LL-D49194 (LLD) and LLD-guanine (LLD-Gua) (C). Reactive epoxide moieties are highlighted in red. Guanine nucleobases are highlighted in blue.

Herein, we report four DNA glycosylases identified from the TXNs BGC, in which TxnU2/U4 (GenBank accession numbers AKT74276 and AKT74302) are derived from the TXN BGC (*txn*, GenBank accession number KP410250) and LldU1/U5 (GenBank accession numbers QDQ37873 and QDQ37896) originate from the LLD BGC (*lld*, GenBank accession number MK501817). TxnU2/4 and LldU1/5 belong to the HTH_42

superfamily and are monofunctional DNA glycosylases that excise TXNA- and LLD-DNA adducts, in which TxnU4 and LldU1 play the major roles in toxin resistance. Interestingly, TxnU4 and LldU1 cannot excise *N*7-methyl or crosslinked G adducts like their homologs AlkZ and YcaQ (Mullins *et al.*, 2017; Bradley *et al.*, 2020), nor can TXNA-DNA lesions be excised by any other alkylpurine DNA glycosylase. Moreover, relative to AlkZ, TxnU4 and LldU1 have a unique catalytic motif that process TXNA- and LLD-DNA lesions differently and that may explain the redundancy for two paralogs in each *txn* and *lld* biosynthetic gene cluster. I also show that AP sites derived from TXNA-DNA excision are processed less efficiently than those generated from 7mG depurination, suggesting that the product of TXNA-DNA excision requires a specialized mechanism for repair.

Results

Self-resistance determinants TxnU2/U4 and LldU1/U5 are closely related to TXNs production

Previously, we identified the BGCs of TXNA (*txn*) and LLD (*lld*) and characterized their partial biosynthetic pathways including starter unit and tailoring steps (Yang *et al.*, 2015; Zhang *et al.*, 2015; Hou *et al.*, 2018; Dong *et al.*, 2019; Shen *et al.*, 2019; Yin *et al.*, 2020), but the function of many of the proteins encoded in their BGCs are unknown. To study the repair mechanism of DNA damage arising from TXNs family of alkylating agents, we first investigated all proteins encoded within and adjacent to the TXNA and LLD BGCs (Zhang *et al.*, 2015; Dong *et al.*, 2019). BLASTP analysis showed that TxnU2/U4 derived from *TXNA a*nd LldU1/U5 derived from *lld* belong to the HTH_42

superfamily and exhibit homology to the DNA glycosylase AlkZ with low sequence identity (26-33%) and similarity (39-46%) (Fig. 51A, C). AlkZ is found within the AZB BGC and has been reported to be an essential resistance protein in AZB biosynthesis by unhooking AZB-ICLs, which would trigger the BER pathway (Fig. 51B) (Wang *et al.*, 2016). We therefore speculated that TxnU2/U4 and LldU1/U5 could confer resistance to TXNA and LLD for self-protection in the producer.

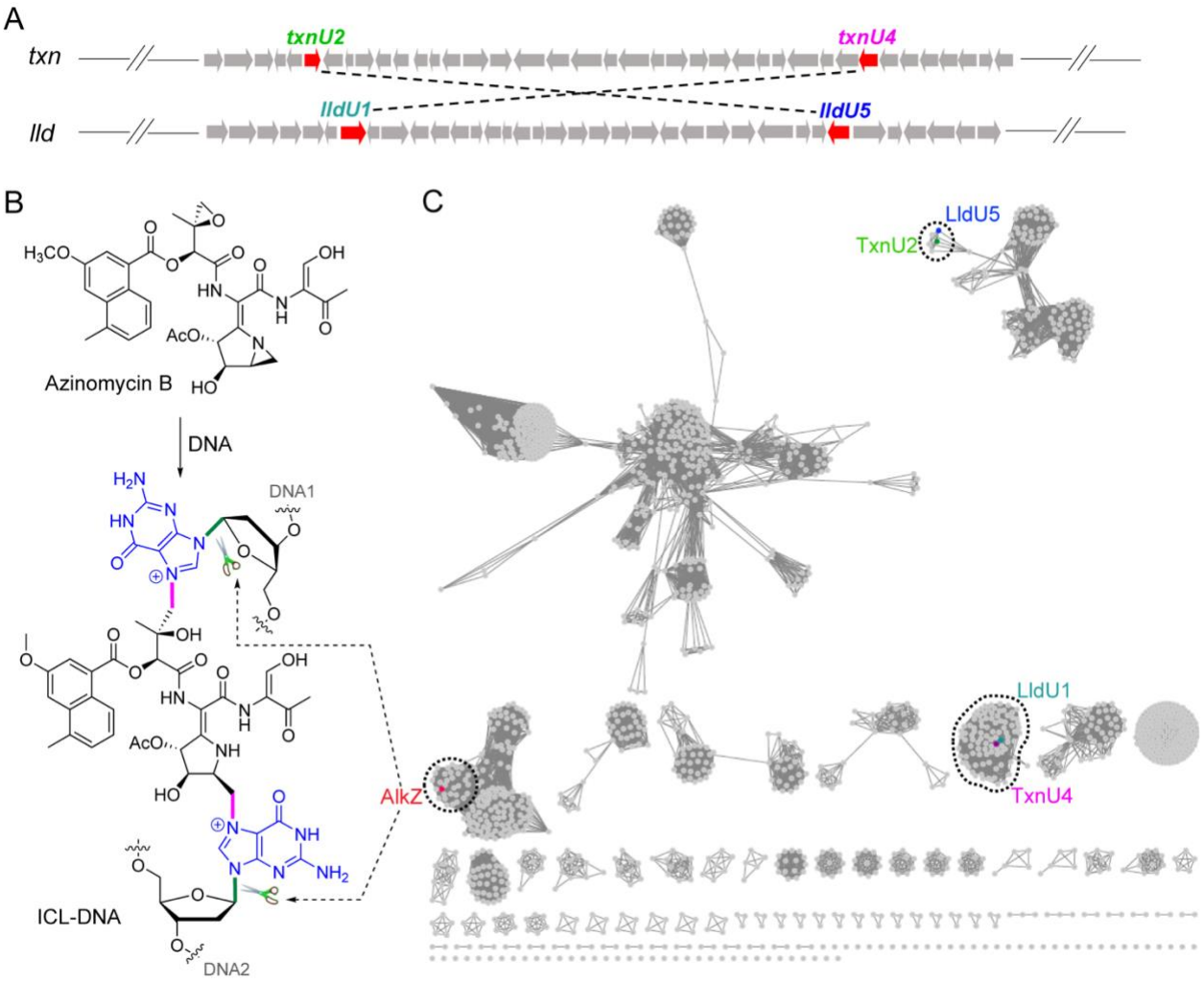


Figure 51. *Genomic analysis of self-resistance determinants TxnU2/U4 and LldU1/U5* (A) Biosynthetic gene clusters (BGC) containing genes encoding HTH_42 superfamily proteins TxnU2/U4 and LldU1/U5. The cluster *txn* is a BGC of TXNs, *lld* is responsible for the BGC of LLD. The two genes connected with dashed lines encode homologous proteins; TxnU2 shares 83% amino acid sequence identity and 90% similarity with LldU5, and TxnU4 shares 71% amino acid sequence identity and 82% similarity with LldU1. (B) Base excision of AZB-ICL-DNA by AlkZ. (C) Sequence similarity network (SSN) analysis of homologous proteins TxnU2/U4 and LldU1/U5. The SSN was constructed by the online Enzyme Function Initiative-Enzyme Similarity Tool with an alignment score threshold of 110. The proteins TxnU2/U4, LldU1/U5 and AlkZ were located in three different clades.

To understand the function of these four proteins, the genes *txnU2*/*U4* from the TXNA producer *S. bottropensis* NRRL 12051 and *lldU1*/*U4* from the LLD producer *S. vinaceusdrappus* NRRL 15735 were deleted (SI Fig. S1A-D in (Chen *et al.*, 2022)), and the yield of compounds in these resulting mutants and wild-type (WT) strains were determined by LC-MS. Compared to the WT strain, the production of TXNA in gene deletion mutant strains Δ*txnU2* and Δ*txnU4* was respectively remarkably reduced 72% and 82%, and the yield of LLD in Δ*lldU1* and Δ*lldU5* was also obviously decreased 85% and 80%, respectively, suggesting the genes *txnU2*/*txnU4* and *lldU1*/*lldU5* are involved in compound biosynthesis and are closely related to the efficiency of TXNA and LLD production, respectively (Fig. 52A-B).

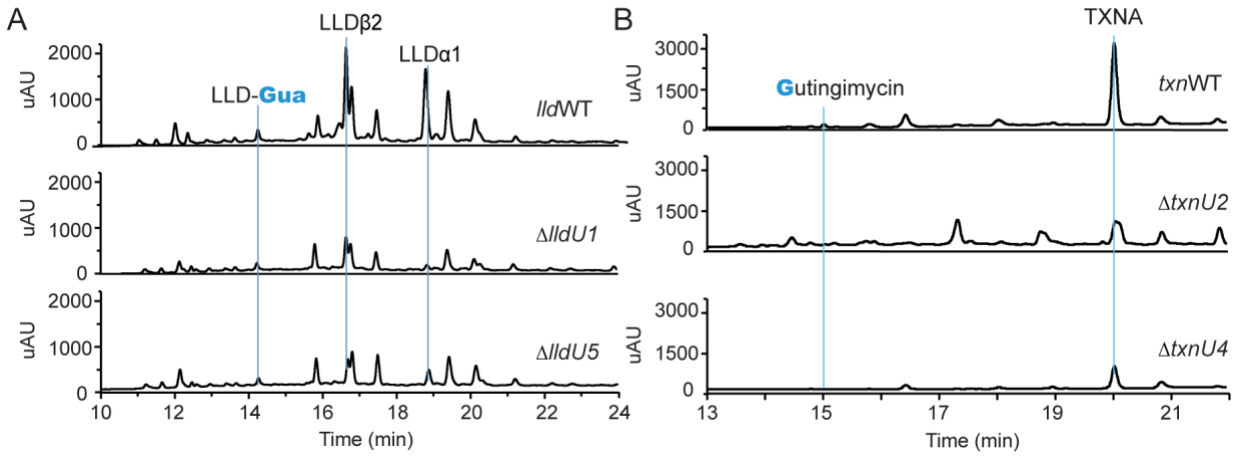


Figure 52. *LldU1/5 and TxnU2/4 are associated with TXNA and LLD biosynthesis* (A) HPLC analysis of extracts from *S. bottropensis* NRRL 12051 wild-type (*txn*WT) and mutant strains, Δ*txnU2* and Δ*txnU4*, at 400 nm absorbance. (B) HPLC profiles of extracts from *S. vinaceusdrappus* NRRL 15735 wild-type (*lld*WT) and mutant strains, Δ*lldU1* and Δ*lldU5*, at 400 nm absorbance.

To follow up this finding and further identify the *in vivo* function of the four proteins, the effect of *txnU2*/*txnU4*, *lldU1*/*lldU5* deletion and overexpression on cells challenged with TXNs was tested. Disc diffusion tests indicated that gene deletion mutants Δ*txnU4*, Δ*lldU1* and Δ*lldU5* exhibited notable sensitivity to both TXNA and LLD, but mutant Δ*txnU2* was no more sensitive to either TXNA or LLD than the WT strain (Fig. 53A-B).

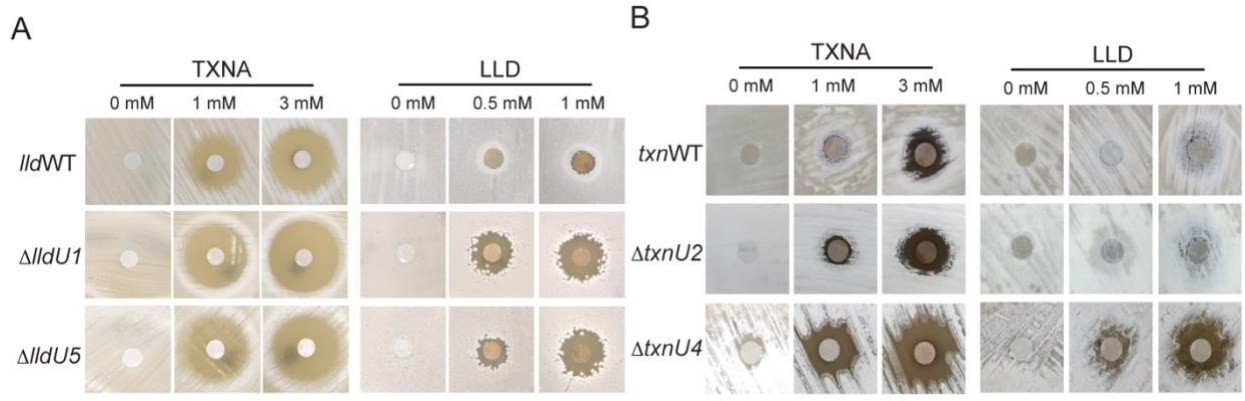


Figure 53. *In vivo characterization of the self-resistance determinants related to LLD and TXNA* (A-B) The effect of *lldU1*/*lldU5* (A) and *txnU2*/*txnU4* (B) deletion on cells challenged with increasing concentrations of TXNA (left) and LLD (right) was tested by a disc diffusion assay. Filter paper discs spotted with different concentrations of TXNA or LLD were laid on the MS plate pre-inoculated with wild type or mutant strains. After incubation at 30°C for 36 hr resistance levels to TXNA or LLD were determined by the zone of inhibition.

Overexpression of *txnU2*/*txnU4* and *lldU1*/*lldU5* in *S. lividans* 1326, a TXNs-sensitive strain, increased cellular viability towards both TXNA and LLD (Fig. 54A). Moreover, consistent with the growth viability in *Streptomyces*, the survival ratio of *E. coli* BL21 that overexpressed *txnU4* or *lldU1* against TXNA was significantly higher than control cells, while *txnU2* overexpression was weakly protective, and there was no effect for *lldU5* overexpression (Fig. 54B). Together, these results show that TxnU2/U4 and LldU1/U5 are self-resistance determinants in TXNA and LLD producers, and among them TxnU4 and LldU1 display the major roles.

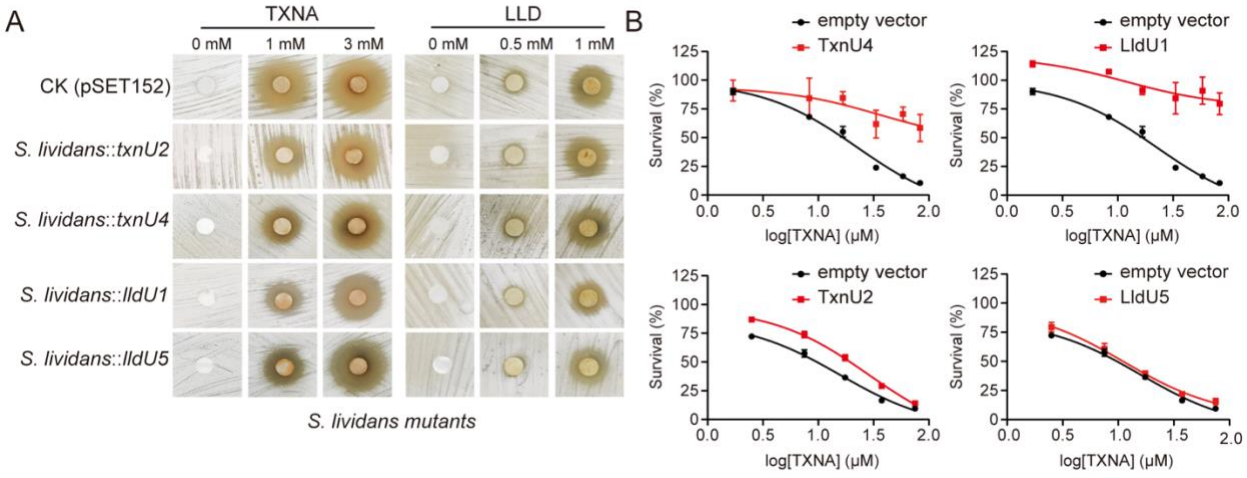


Figure 54. *Overexpression of TxnU2/TxnU4 and LldU1/LldU5 confer resistance to heterologous hosts against TXNA and LLD* (A) Disc diffusion test assay to determine the antibiotic sensitivity of heterologous expression strains *S. lividans*::pSET152, *S. lividans*::txnU2, *S. lividans*::txnU4, *S. lividans*::lldU1 and *S. lividans*::lldU5 to TXNA (left) and LLD (right). (B) TXNA inhibition of *E. coli* BL21 cells transformed with protein overexpression plasmid txnU2-pET28a, txnU4-pET28a, lldU1-pET28a, lldU5-pET28a or empty vector pET28a alone. Data are mean ± SD (n=3).

TxnU2/U4 and LldU1/U5 are DNA glycosylases that excise TXNA- and LLD-DNA adducts

To determine if TxnU2, TxnU4, LldU1, and LldU5 are DNA glycosylases capable of excising TXNA- and LLD-Gua adducts from DNA, an 8-bp oligodeoxynucleotide duplex d(AACCGGTT) was designed based on a previous report to generate site-specific TXNA/LLD adducts (Fig. 55A) (Fitzner *et al.*, 2008; Pfoh *et al.*, 2008). The reaction products of TXNA- and LLD-Gua excision by TxnU and LldU enzymes were detected by HPLC-MS and with 271 and 400 nm UV light (Fig. 55B-C, SI Fig. S2 A-B in (Chen *et al.*, 2022)). After treatment with TXNA, two new peaks appeared at 18.2 and 18.9 minutes. The *m/z* values of the two peaks were 1644, which was consistent with that of the $[M+2H]^{2+}$ ion of the monoalkylated adduct generated by covalent binding of one molecule TXNA to either G within the duplex d(AACCGGTT) (SI Fig. S2A in (Chen *et al.*, 2022)). Given the previous sequence selectivity studies showing that TXNA reacts preferentially with the DNA sequence 5′-GT (Fitzner *et al.*, 2008; Pfoh *et al.*, 2008), we supposed that the product with the later retention time (18.9 minutes) and larger peak area is 5′-AACCG(TXNA-G)TT-3′, and the other peak at 18.2 minutes is 5′-AACC(TXNA-G)GTT-3′. As TxnU2 or TxnU4 was added, the amount of the two adducts decreased, and a new peak with *m/z* 1028 appeared, which was supposedly the excision product of TxnU2 and TxnU4 (Fig. 55B). The molecular weight of the product is equal to that of gutingimycin,

which contains a TXN skeleton and a Gua nucleobase. In addition, the molecular formula $C_{47}H_{57}O_{21}N_5$ determined by HRESIMS ($[M+H]^+$ *m/z* 1028.53) and the fragments detected by tandem-MS were consistent with gutingimycin (SI Fig. S3A in (Chen *et al.*, 2022)), confirming that TxnU2 and TxnU4 are able to catalyze excision of TXNA-Gua adducts from DNA. An extended time course indicated that TxnU4 preferentially cleaved the 5′-AACCG(TXNA-G)TT-3′ among the two alkylated products (SI Fig. S4 in (Chen *et al.*, 2022)). Likewise, under the same experimental condition, the two alkylation products arising from LLD were excised by LldU1, forming a new compound with *m/z* 1102 in the mass spectra, whereas LldU5 showed no activity (Fig. 55C, SI Fig. S2B in (Chen *et al.*, 2022)). HRESIMS data ($[M+H]^+$ *m/z* 1102.43, calcd for $C_{51}H_{68}O_{22}N_5$) and tandem-MS analysis indicated that the excision product is LLD-Gua (SI Fig. S3B in (Chen *et al.*, 2022)), suggesting LldU1 is capable of excising LLD-G adducts from DNA.

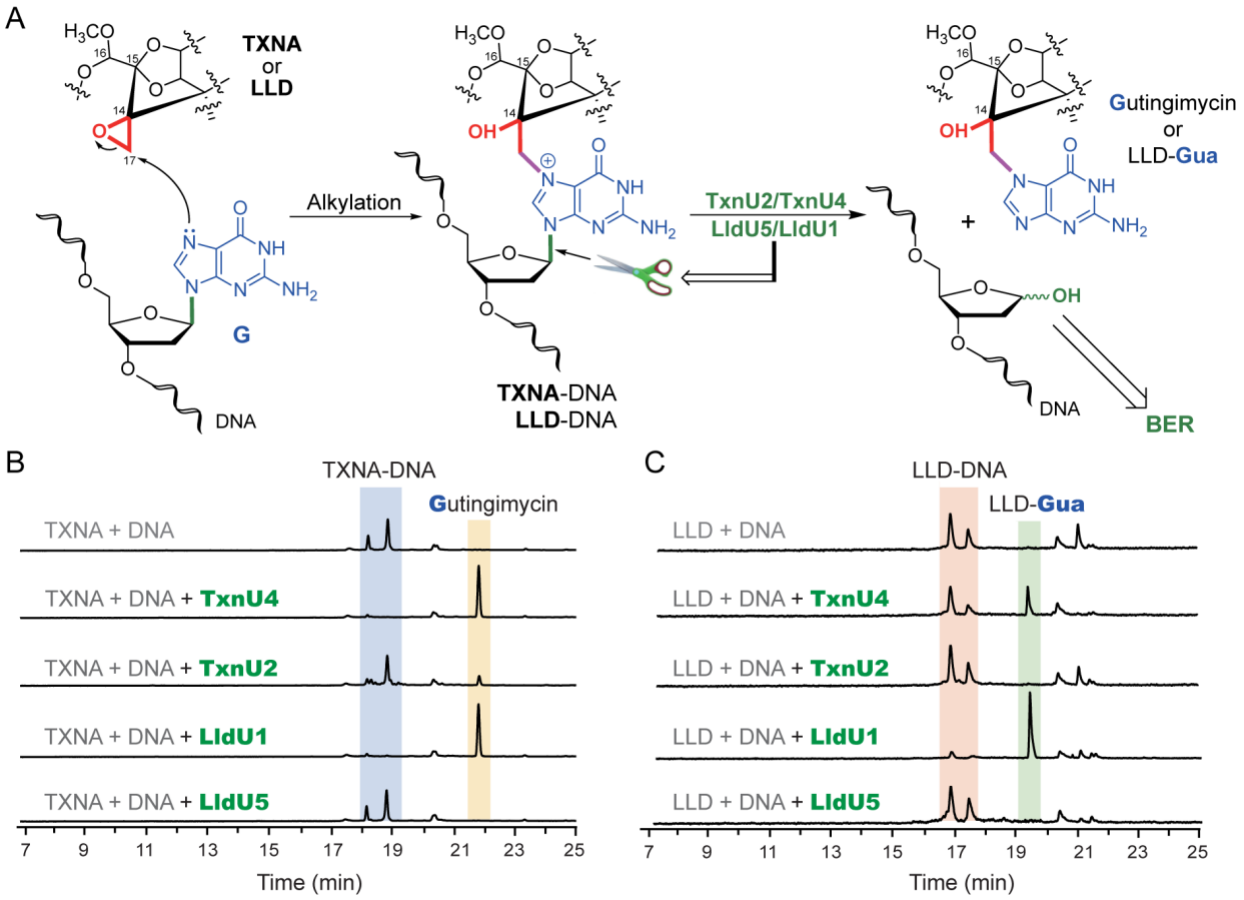


Figure 55. *Base excision of TXNA and LLD by TxnU2/4 and LldU1/5 by HPLC-MS* (A) Chemical reaction between the epoxide moiety of TXNA or LLD and N7 of G in DNA. DNA glycosylases catalyze the hydrolysis of the N-glycosidic bond to liberate the alkylguanine adduct, generating an AP site in the DNA. (B,C) LC-MS analysis of the cleavage products of TxnU2, TxnU4, LldU1 and LldU5 reaction with TXNA-DNA (B) and LLD- DNA (C). An 8-bp oligodeoxynucleotide duplex d(AACCGGTT) was pre-incubated with TXNA or LLD at 16°C for 2 hr, followed by treatment with enzymes TxnU2, TxnU4, LldU1 and LldU5 for 2 hr. The reaction mixtures were analyzed by LC-MS at 400 nm absorbance.

For further confirmation, an *in vitro* gel-based assay was performed to quantify the β- and δ-elimination products generated by alkaline hydrolysis of the AP site product of base excision (Fig. 56A) (Mullins *et al.*, 2017). We verified that the amount of product observed in this assay was not influenced by the use of NaOH to cleave glycosylase generated AP sites, as similar results were obtained with piperidine (SI Fig. S5A-B in

(Chen *et al.*, 2022)). Purified enzymes were incubated with either TXNA- or LLD-DNA substrates for 30 min under single turnover conditions. We found that all four enzymes produced a significant amount of product as compared to a no-enzyme control (Fig. 56B-C). The weaker activity of LldU5 relative to the other three enzymes (Fig. 56C) is likely the result of poor protein solubility observed during expression and purification. Single-turnover kinetic analysis showed that TXNA-Gua excision by TxnU4 (k_{st} = 4.6 min^{-1}) is approximately 4 times faster than *S. sahachiroi* AlkZ and *E. coli* YcaQ activity toward AZB-ICL (k_{st} = 1.2 min^{-1}) and NM-ICL (k_{st} = 1.1 min^{-1}) substrates, respectively (Bradley *et al.*, 2020) (Fig. 56D, SI Fig. S5D in (Chen *et al.*, 2022)). The enzyme also efficiently turns over (k_{mt} = 0.3 min^{-1}) and shows no observable product inhibition, as evidenced by multiple-turnover kinetics (Fig. 56D, SI Fig. S5D in (Chen *et al.*, 2022)). Thus, these enzymes excise TXNs lesions rapidly and efficiently relative to their distant orthologs. Moreover, the *in vitro* excision activities of TxnU2/U4 and LldU1/U5 were further confirmed by the detection of excision products in gene deletion mutant strains (Fig. 56A-B, SI Fig. S2A-B in (Chen *et al.*, 2022)). Compared to the WT strain, the production of LLD-Gua in gene deletion mutant strains Δ*lldU1* and Δ*lldU5* was respectively reduced 43% and 30%, and the yield of gutingimycin in Δ*txnU2* and Δ*txnU4* was also respectively decreased 95% and 99%, suggesting the glycosylases TxnU2/U4 and LldU1/U5 are functional *in vivo*.

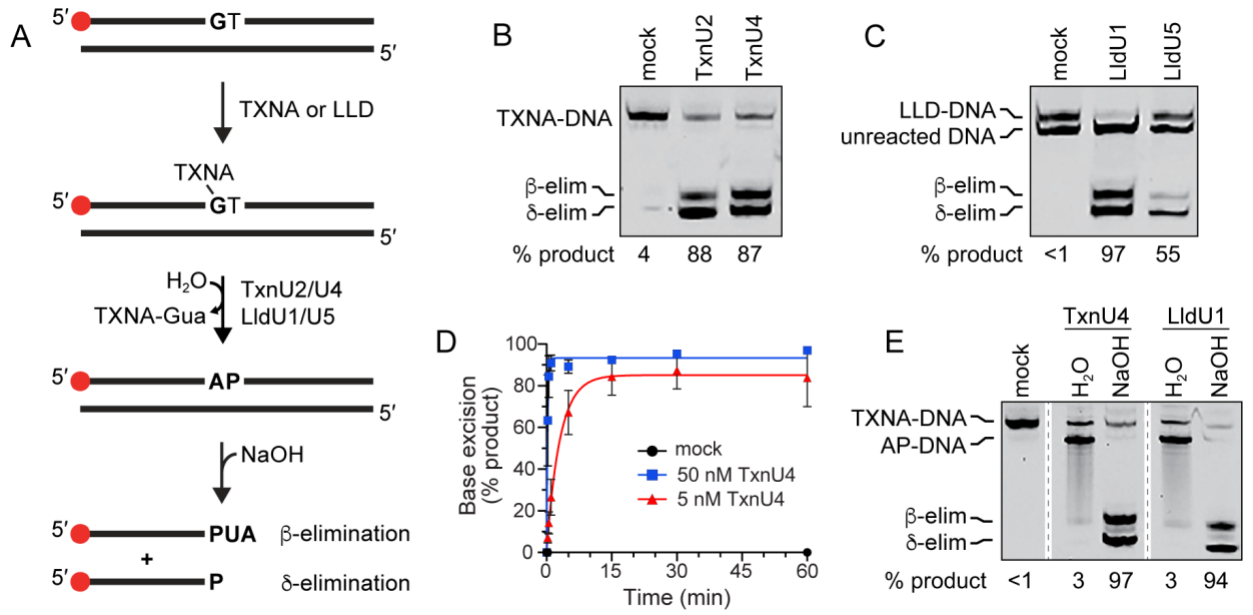


Figure 56. *TxnU2/4 and LldU1/5 are monofunctional DNA glycosylases which rapidly excise their cognate lesions* (A) Schematic of the base excision assay performed in panels B-E. DNA containing a centrally located GT dinucleotide and a 5'-Cy5-label (red circle) is incubated with TXNA or LLD to form the substrate. Incubation with TxnU2/U4 or LldU1/U5 generates an AP site, which is cleaved with hydroxide to generate β- and δ-elimination products. PUA, 3′-phospho-α,β-unsaturated aldehyde; P, 3′-phosphate. (B,C) Denaturing PAGE of TXNA-DNA (B) and LLD-DNA (C) reactions after treatment with enzyme or buffer (mock) for 30 min. Formation of the LLD-DNA substrate only went to ~50% completion, with unreacted DNA migrating faster on the gel. Substrate and product DNA migrate as expected for their sizes, as judged by their relative position to bromophenol blue and xylene cyanol tracking dyes, (SI Fig. S5C in (Chen *et al.*, 2022)). (D) Single- (blue) and multiple-turnover (red) excision kinetics of TxnU4 against TXNA-DNA. 50 nM TXNA-DNA was incubated with buffer (mock), 50 nM TxnU4 (1:1 protein:DNA), or 5 nM TxnU4 (1:10 protein:DNA). Data are mean ± SD (n=3). A representative gel from which the data were quantified is shown in SI Fig. S5D in (Chen *et al.*, 2022). (E) Denaturing PAGE of TXNA-DNA adducts after 30-min incubation with TxnU4 or LldU1, followed by work-up with either H_2O or NaOH.

Monofunctional glycosylases catalyze only hydrolysis of the N-glycosidic bond, whereas bifunctional glycosylases also nick the backbone to generate β- and δ-elimination products. Based on our previous functional analysis of the homolog AlkZ, we hypothesized that TxnU and LldU enzymes were monofunctional. Indeed, similar to AlkZ, NaOH was required to nick the AP-DNA product formed by TxnU4 and LldU1 (Fig. 56E).

Treating the reacted TXN-DNA with water preserved the AP site, while treatment with hydroxide cleaved the AP site to generate β- and δ- elimination products. These results indicate that the TxnU and LldU enzymes are monofunctional glycosylases and do not contain intrinsic DNA lyase activity.

TxnU4 and LldU1 remove TXNs-guanine adducts with a similar but distinct catalytic motif relative to AlkZ

The active sites of all monofunctional DNA glycosylases contain catalytic carboxyl (Asp, Glu) or carboxamide (Asn, Gln) residues that promote base excision by electrostatically stabilizing the positive charge that develops on the deoxyribose as the glycosidic bond is broken, and by deprotonating or positioning a water molecule for nucleophilic attack of the anomeric C1′ carbon (Stivers and Jiang, 2003; Brooks *et al.*, 2013; Drohat and Maiti, 2014; Drohat and Coey, 2016; Mullins *et al.*, 2019). We previously showed that the TxnU/LldU homolog AlkZ contains a catalytic QΦQ motif (Φ is a small aliphatic residue) (Fig. 57A), and that mutation of either flanking glutamine abrogates base excision of monoadducts and severely reduces ICL unhooking activities (Mullins *et al.*, 2017; Bradley *et al.*, 2020). Based on a rigid-body docking model of AlkZ in complex with AZB-DNA (Mullins *et al.*, 2017), the C-terminal glutamine side chain is likely within proximity to the lesion deoxyribose to position a catalytic water molecule (Fig. 7A-C). Although the N-terminal glutamine is more recessed and contacts the DNA backbone of a neighboring nucleotide, a slight rotation of the DNA around the helical axis in our docking model would position this residue for catalysis on the adducted nucleotide, and thus either residue theoretically can play a catalytic role in base excision.

Like AlkZ, TxnU2 and LldU5 contain a QΦQ motif, whereas TxnU4 and LldU1 contain a histidine residue (H43) in the first position (Fig. 57A). Both QΦQ and HΦQ motifs are predicted to reside in the same location as those observed in AlkZ (Figs. 57A, 62C), and the His imidazole should be able to perform the same catalytic function as described above for carboxylate and carboxamide side chains. We examined the functional role of the HΦQ motifs in TxnU4 and LldU1 by purifying H43A and Q45A mutants and measuring TXNA-DNA and LLD-DNA excision activity. Wild-type TxnU4 removed 94% of the TXNA-DNA adduct after 30 seconds. At this same short time point, the TxnU4 H43A mutant showed no activity, whereas substitution of Gln45 with alanine had no effect on TxnU4 activity (Fig. 57B). Interestingly, we found the exact opposite effect of H43 and Q45 residues in LldU1 tested against an LLD-DNA substrate; LldU1 H43A had no effect compared to wild-type, whereas Lld1 Q45A showed no activity (Fig. 57C). We also tested the activity of the QΦQ motif in LldU5; alanine substitution of either glutamine abrogated activity compared to the wild-type enzyme (Fig. 57D), similar to that shown for AlkZ (Mullins *et al.*, 2017). We were unable to test the activity of TxnU2 mutants because the proteins were unstable and not amenable to purification. These results indicate that the *in vitro* activity we observe from purified protein is not the result of a contaminating activity in our protein preparations, and suggest that either the histidine or glutamine residues within TxnU4 and LldU1 HΦQ motifs are catalytic, and that they engage TXNA-G and LLG-G lesions differently.

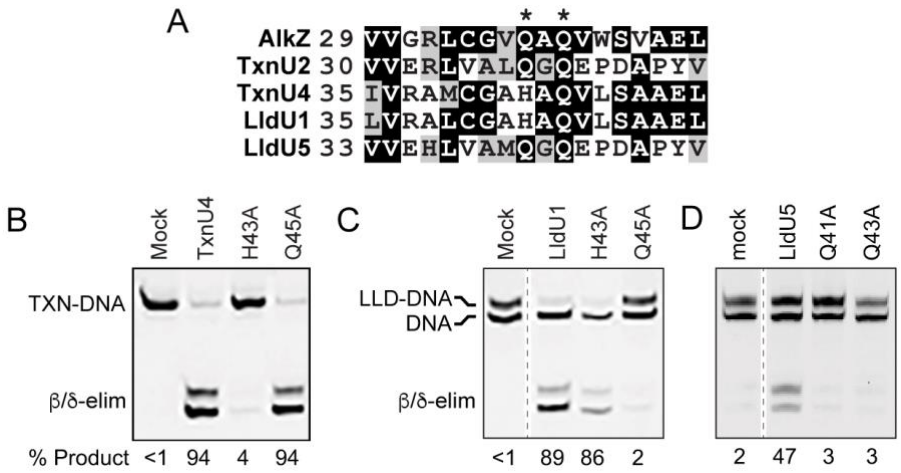


Figure 57. *Mutational analysis of excision activity for LldU1/5 and TxnU4* (A) Sequence alignment of the catalytic residues in *S. sahachiroi* AlkZ and TxnU2/U4 and LldU1/U5. Denaturing PAGE of TxnU4/TXNA-DNA (B) and LldU1/LLD-DNA (C) Single-turnover reactions containing 1 μM protein and 50 nM DNA. WT and mutant proteins were incubated with substrates for 30 sec. (D) Single turnover reactions between LldU5 enzymes and LLD-DNA were carried out for 96 hr.

TXNs form stable DNA adducts that are specifically excised by TxnU and LldU glycosylases

*N*7-alkyl-2′-deoxyguanosine adducts (e.g., 7mG) are generally thermally unstable and prone to depurination (Hemminki *et al.*, 1989). I therefore explored the stability of TXNs-DNA adducts. Heating the TXNA-DNA to 95°C for 5 minutes, followed by either water or hydroxide workup, led to depurination of only 32% of the adduct (Fig. 58A). In contrast, our previous studies show 90% depurination of *N*7-linked NM- and AZB-ICLs under the same conditions (Bradley *et al.*, 2020) suggesting that TXNA-DNA adducts are more stable than other *N*7-alkyl lesions. To test this, I directly compared the stabilities of TXNA-DNA and 7mG-DNA adducts by monitoring their spontaneous depurination rates at 37°C over a period of 7 days. I found that the TXNA-G *N*-glycosidic bond is at least 5 times more stable than that of 7mG (Fig. 58B, SI Fig. S5E in (Chen *et al.*, 2022)). Thus,

relative to 7mG, TXNA adducts are more resistant to spontaneous depurination, which may be an important property for TXNs toxicity.

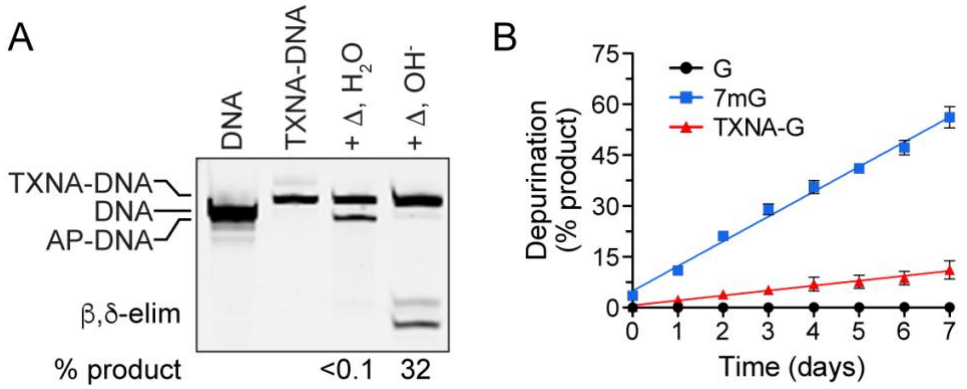


Figure 58. *TXNA forms highly stable DNA adducts that are resistant to depurination* (A) Denaturing PAGE of TXNA-DNA adducts after thermal depurination. TXNA-DNA was heated (Δ) to 95°C for 5 min, followed by treatment with either water or NaOH. (B) Kinetics of spontaneous depurination of TNXA in DNA as compared with 7mG or unmodified G. Data are mean ± SD (n=3). Half-lives derived from linear regression of the data are 6.1 ± 0.3 days (7mG) and 33.9 ± 8.1 days (TXNA). A representative gel from which this data was quantified is shown in SI Fig. S5E in (Chen *et al.*, 2022).

I next tested the ability of other bacterial alkylpurine DNA glycosylases to excise TXNA-Gua from DNA. These glycosylases, which include *E. coli* AlkA and YcaQ, *Bacillus cereus* AlkC and AlkD, and *S. sahachiroi* AlkZ, have widely varying substrate specificities in addition to their ability to excise 7mG (O'Brien and Ellenberger, 2004; Alseth *et al.*, 2006; Brooks *et al.*, 2013; Parsons *et al.*, 2016; Mullins *et al.*, 2017; Shi *et al.*, 2018; Bradley *et al.*, 2020). Under the experimental conditions tested, we were unable to detect TXNA excision products from any of these glycosylases (Fig. 59A), indicating that recognition of the TXNA lesion is confined to a glycosylase found in a TXNs BGC. I also compared the cross-reactivity of the TxnU and LldU enzymes by testing the ability of TxnU2 and TxnU4 to excise LLD adducts and of LldU1 and LldU5 to excise TXNA

adducts, and found that both TxnU4 and LldU1 are capable of excising both TXNA and LLD adducts (Fig. 59B), consistent with our results from HPLC analysis (Fig. 55B-C).

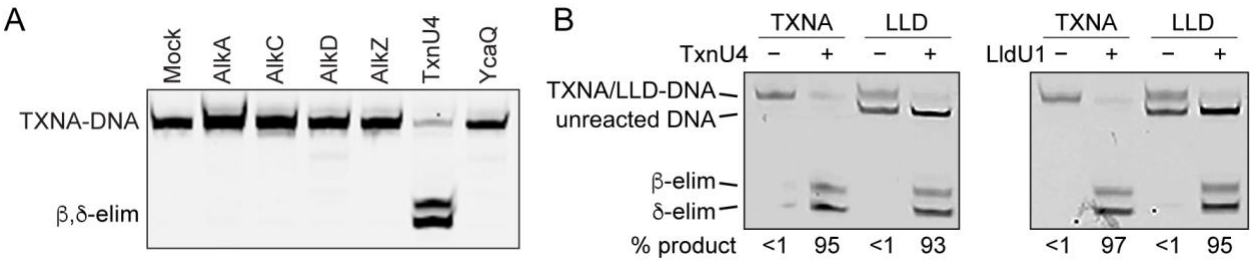


Figure 59. *LLD and TXNA DNA adducts are excised only by LldU/TxnU glycosylases* (A) Denaturing PAGE of TXNA-DNA adducts after 1-hr incubation with either buffer (mock) or bacterial alkyl-DNA glycosylases. (B) TxnU4 can excise LLD-DNA and LldU1 can excise TXN-G-DNA adducts.

Given the efficient activity of TxnU4 for TXNA lesions (Fig. 56D), we were interested in determining whether TxnU and LldU could cleave other, less stable *N*7-alkyl-DNA adducts. I previously found that *E. coli* YcaQ readily excises 7mG (Fig. 60A) and unhooks NM-ICLs generated from reaction of DNA with mechlorethamine (Fig. 60D) (Mullins *et al.*, 2017; Bradley *et al.*, 2020). To our surprise, in contrast to YcaQ, neither TxnU4 nor LldU1 showed any significant activity toward 7mG (Fig. 60A) or a NM-ICL (Fig. 60D) after 30 min, despite the lower stability of these lesions relative to TXNs adducts. The inability of TxnU4 to act on these less stable *N*7-alkyl adducts and of other alkylpurine DNA glycosylases to process TXNA-DNA indicate that the TxnU/LldU enzymes are highly specific for their cognate natural products, and suggests that the enzymes likely recognize a specific feature of the TXNs-DNA substrates either directly through interaction with the compound or indirectly through the structural distortion to the DNA imposed by the intercalated adduct (Fig. 60A-C).

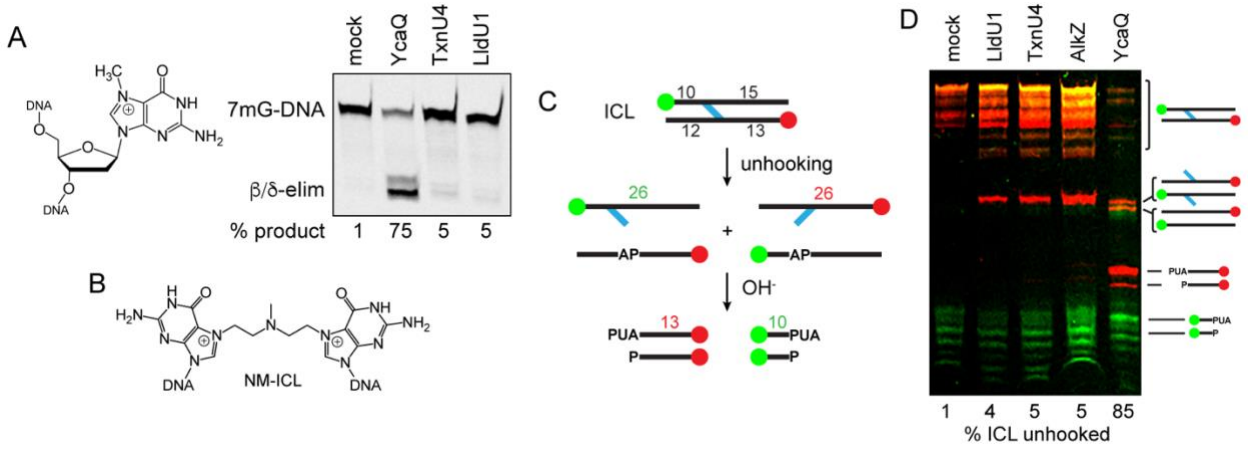


Figure 60. *LldU1 and TxnU4 cannot remove 7mG or unhook NM$_5$-ICLs* (A) Denaturing PAGE of 30-min reaction products of *E. coli* YcaQ and *Streptomyces* TxnU4 and LldU1 with 7mG-DNA. (B) Structure of nitrogen mustard (NM$_5$)-ICL produced by reaction of mechlorethamine with guanines on opposite DNA strands. (C) Schematic of ICL unhooking reactions. Strands are 5′-labeled with either FAM (green) or Cy5 (red). Unhooking by a glycosylase produces single stands containing either monoadducts or AP-sites, the latter which are susceptible to nicking by hydroxide. (D) Denaturing PAGE of NM-ICL unhooking reactions after treatment with buffer (mock) or enzyme for 30 min, followed by alkaline hydrolysis. The percent of β/δ-elimination products is quantified below the gel. Each image is an overlay of false-colored FAM (green) and Cy5 (red) fluorescence scans of the gels, in which yellow depicts coincident red and green intensity.

AP sites generated from TxnU4 cleavage of TXNA-DNA are inefficiently processed by EndoIV

The AP site product of DNA glycosylase activity is a toxic intermediate of the BER pathway, and thus must be efficiently incised by an AP endonuclease for completion of the pathway. I therefore investigated the efficiency with which a bacterial AP endonuclease could act on the product of the TxnU4/TXNA-DNA reaction. When comparing various methods to cleave TxnU4-generated AP sites in our gel-based assay, I noticed that *E. coli* EndoIV did not fully incise the AP-DNA created by TxnU4 (SI Fig. S5B in (Chen *et al.*, 2022)). The EndoIV reaction was carried out under the same

conditions that show 100% incision activity from AP sites generated by AlkZ or YcaQ excision of 7mG (Mullins *et al.*, 2017; Bradley *et al.*, 2020), suggesting that the product of the TxnU4/TXNA-DNA reaction inhibits the AP endonuclease. I therefore followed up on this result by comparing the kinetics of EndoIV cleavage of AP sites generated by TxnU4/TXNA-DNA and YcaQ/7mG-DNA reactions (Fig. 61A-B). I wished to examine AP site processing without interference from residual glycosylase bound to either substrate or product DNA. Therefore, AP sites were generated under conditions that allow for completion of the glycosylase reaction with sub-saturating concentrations of protein with respect to DNA. I found that EndoIV incision of AP sites formed by YcaQ/7mG-DNA are rapidly and fully incised (k_{obs} = 2.8 min^{-1}) within 5 min (Fig. 61A-B). In contrast, EndoIV incision of AP sites generated from TxnU4/TXNA-DNA showed biphasic kinetics. The first phase is consistent with the first enzymatic under our experimental conditions, and showed similar kinetics (k_{fast} = 2.0 min^{-1}) as EndoIV activity on 7mG-produced AP sites. However, the second phase (i.e., subsequent turnovers) was 200-fold slower (k_{slow} = 0.02 min^{-1}), suggesting that *E. coli* EndoIV is product inhibited when processing TXNA-generated AP sites. More importantly, the difference in EndoIV processing of TXNA and 7mG excision products indicates a difference in AP sites generated from the two lesions, the most likely rationale for which is that gutingimycin (TXNA-Gua) remains intercalated in the DNA after glycosylase excision. These data show that the AP-DNA/TXNA-Gua product poses a challenge for processing by *E. coli* IV, and suggests that a specialized AP endonuclease may be required for efficient BER of these lesions.

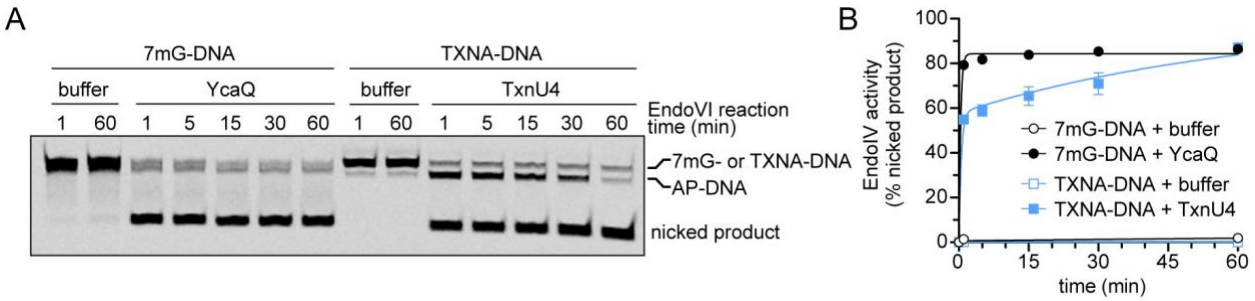


Figure 61. *AP sites generated from TxnU4 action on TXNA-DNA are incised inefficiently by EndoIV* (A) Representative denaturing PAGE of EndoIV incision of AP-DNA generated from YcaQ excision of 7mG or TxnU4 excision of TXNA-G. 50 nM 7mG- or TXNA-DNA was incubated with either buffer or 5 nM YcaQ or TxnU4 for 2 hr at 25°C to generate AP sites, followed by addition of EndoIV at a final concentration of 17 nM EndoIV and 40 nM DNA. EndoIV reactions were incubated at 37°C for the specified times prior to denaturing and electrophoresis. (B) Quantification of the gel in panel A. Data are mean ± SD (n=3). 7mG data were fit to a one-phase exponential (k = 2.8 min^{-1}, R^2 = 0.9975), and TXNA data were fit to a biphasic exponential (k_{fast} = 2.0 min^{-1}, k_{slow} = 0.02 min^{-1}, R^2 = 0.9755).

Discussion

In this study, HTH_42 superfamily proteins TxnU2/U4 and LldU1/U5 were discovered to provide cellular resistance to TXNA and LLD toxicity, respectively, providing an explanation for the evolutionary function of these proteins within the BGC of each antibiotic. Sequence (BLASTP) and structural (AlphaFold) analyses show that TxnU and LldU share homology with AlkZ and YcaQ (Fig. 62A-B), and the *in vitro* enzymatic activity confirms that like AlkZ/YcaQ, both TxnU and LldU are monofunctional DNA glycosylases acting on *N*7-alkylguanine adducts (Wang *et al.*, 2016; Mullins *et al.*, 2017; Bradley *et al.*, 2020). However, the TxnU/LldU enzymes differ from their HTH_42 homologs—and other alkylpurine DNA glycosylases—with respect to substrate specificity, catalytic machinery, and genomic context.

In terms of specificity, most alkylpurine DNA glycosylases hydrolyze 7mG in addition to their major substrates (O'Brien and Ellenberger, 2004; Alseth *et al.*, 2006; Brooks *et al.*, 2013; Parsons *et al.*, 2016; Mullins *et al.*, 2017; Shi *et al.*, 2018; Bradley *et al.*, 2020). Interestingly, despite the lower stability of the 7mG N-glycosidic bond, TxnU4 and LldU1 did not exhibit 7mG activity, indicating that TxnU/LldU specifically recognize TXNA-G and LLD-G as opposed to the instability in the N-glycosidic bond generated by substitution of guanine at N7 (Gates, 2009). Similarly, the TXNs-DNA lesions did not appear to be substrates for the other alkylpurine DNA glycosylases, including AlkZ/YcaQ and YtkR2/AlkD, which also act on bulky lesions (Xu *et al.*, 2012; Mullins *et al.*, 2015; Wang *et al.*, 2016; Mullins *et al.*, 2017; Bradley *et al.*, 2020). The lack of activity of TxnU/LldU for less stable *N*7-alkylguanine adducts and the inability of other glycosylases to hydrolyze TXNA-G indicate that TxnU and LldU are highly specific for their own natural products. The most significant differences between LLD/TXN-G and other known *N*7-alkylpurine glycosylase substrates are their ability to intercalate into the DNA base stack and their sugar substituents (Fig. 63A-C). Based on the TXNA-DNA crystal structure, TXNA intercalates the d(GT/AC) base step and forms hydrogen bonds with the duplex DNA through the two sugar moieties, leading to the 4-sugar in the minor groove and the 13-sugar residing in the major groove (Pfoh *et al.*, 2008). In addition, TXNA extrudes the base near the 3′ end of the alkylating site out of the helix, leading to an increased helical twist (Pfoh *et al.*, 2008).

To our knowledge, TxnU and LldU are the only DNA glycosylases identified with activity toward intercalated DNA substrates. An AlkZ-derived homology model of TxnU4 docked against the TXNA-DNA crystal structure provides a rationale for this specificity

(Fig. 63A-C). Our previous work predicted that AlkZ employs two important secondary structural elements to engage the DNA substrate from opposite faces of the DNA—the β11/12-hairpin is posited to contact the lesion in the minor groove, and helix αI is predicted to make direct contact to the AZB compound from the major groove side (Mullins *et al.*, 2017; Bradley *et al.*, 2020) (Figs. 19A-B, 63A-C). Because the TXN compounds intercalate both strands of DNA, they protrude from both major and minor groove sides. Consequently, helix αI and the β11/12-hairpin likely contact TXNs from both grooves, with helix αI recognizing the C13- or C16-modified sugars on one end and the β11/12-hairpin recognizing C-4 modified sugar on the other end. Interestingly, the sequences and predicted structures of these two recognition elements are not conserved between AlkZ and TxnU/LldU (Figs. 62A-B, 63A-C), consistent with their predicted roles in recognition of two different classes of natural products.

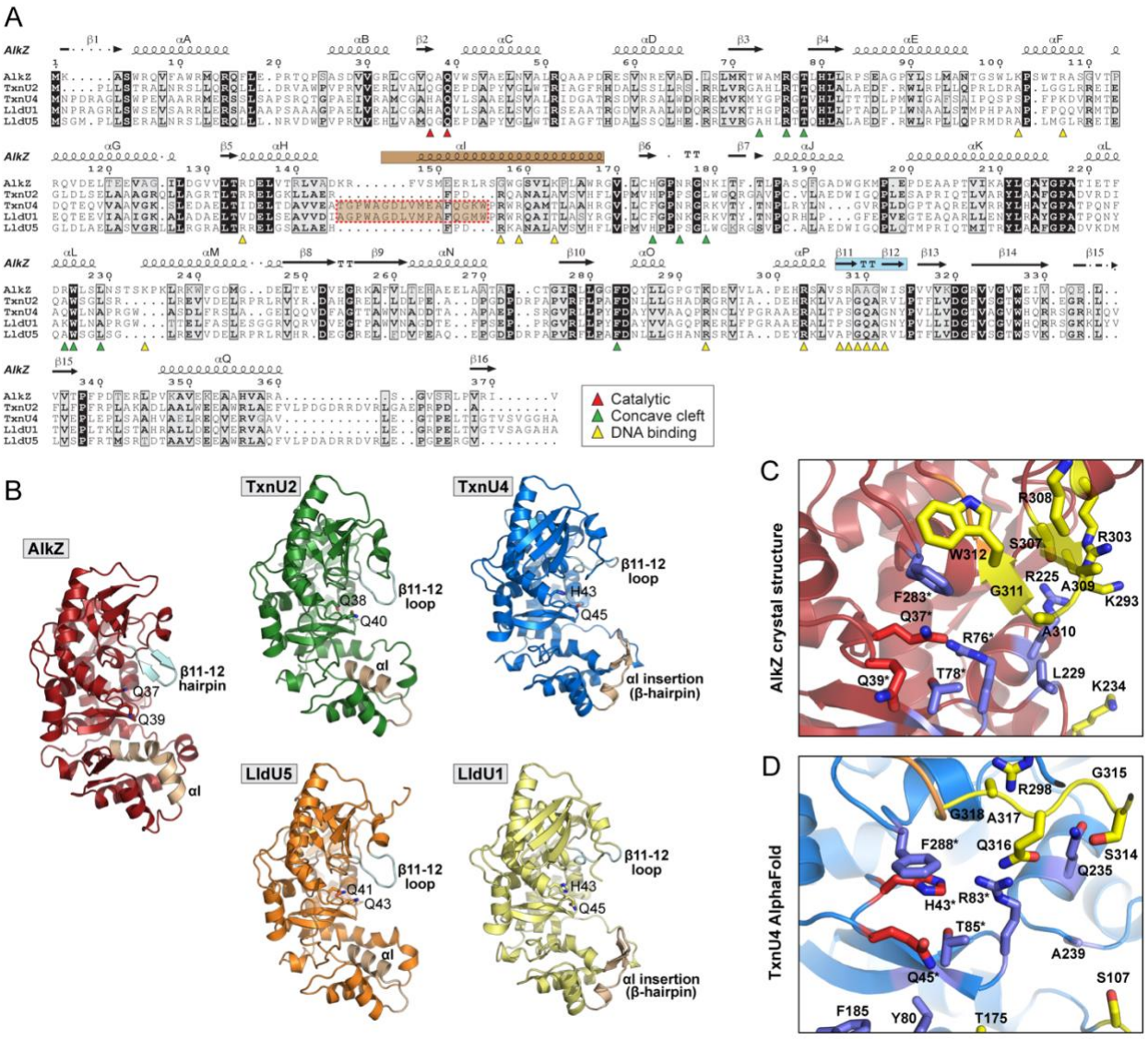


Figure 62. *Conservation and structural models of TxnU/LldU and AlkZ enzymes* (A) Sequence alignment of AlkZ, TxnU2/U4, and LldU1/U5 generated using T-Coffee Expresso (Notredame *et al.*, 2000) and ESPript 3.0 (Robert and Gouet, 2014). Secondary structure above the sequences is derived from the AlkZ crystal structure (PDB ID 5UUJ (Mullins *et al.*, 2017). Helix αI and the β11/12-hairpin predicted to contact the DNA lesion are colored brown and cyan, respectively. The dashed red box highlights the helix αI insertion unique to TxnU4 and LldU1. Triangles below the sequences indicate catalytic (red) or DNA binding (yellow) residues, and residues residing within the concave cleft (green). (B) Comparison of AlkZ crystal structure (PDB ID 5UUJ) to TxnU/LldU models generated by AlphaFold2 (Jumper *et al.*, 2021). Catalytic QΦQ and HΦQ residues are in ball-and-stick, and putative lesion-binding secondary structural elements are highlighted brown and cyan as in panel A. (C,D) Active sites of AlkZ (C) and TxnU4 (D). Catalytic residues are red, β11-12-loop residues are yellow, and residues protruding into the DNA binding channel are blue.

Regarding catalysis, the AlkZ/YcaQ/TxnU/LldU family of HTH_42 enzymes act on crosslinked or intercalated substates that are not likely to be extruded from the DNA, as observed for base-flipping glycosylases including human AAG and bacterial AlkA (Brooks *et al.*, 2013; Mullins *et al.*, 2019). Consistently, the HTH_42 enzymes, like their non-base-flipping counterparts YtkR2/AlkD, do not contain residues that would intercalate the DNA helix to stabilize an extruded nucleobase in the active site, nor do they contain a nucleobase binding pocket within the active site (Rubinson *et al.*, 2010; Mullins *et al.*, 2015). Instead, the catalytic residues are pre-organized to contact the target N-glycosidic bond within an intact DNA duplex (Mullins *et al.*, 2017) (Figs. 62C, 63A-C). We previously showed that the catalytic motifs of the HTH_42 superfamily are divided into QΦQ and QΦD types (Bradley *et al.*, 2020). Sequence similarity network (SSN) analysis showed that the five proteins—AlkZ, LldU1/U5 and TxnU2/U4—are located in three different clades, in which TxnU2 and LldU5 are clustered into one clade, TxnU4 and LldU1 are clustered into another, and AlkZ clustered in a third (Fig. 51C). The catalytic motif of TxnU2 and LldU5 is the same as AlkZ and belongs to the QΦQ type. However, the catalytic motifs of TxnU4 and LldU1 belong to neither QΦQ nor QΦD, but instead contain an HΦQ motif (Figs. 57A, 62A). Our structural models predict the HΦQ side chains to be in the same locations as those in AlkZ QΦQ, and thus either could reside close enough to the target TXNA-G or LLD-G nucleotide to catalyze hydrolysis (Figs. 62D, 63A-C) (Mullins *et al.*, 2017). Interestingly, however, our mutational analysis revealed that HΦQ behaves differently than QΦQ and QΦD in two respects. First, mutation of only one residue affected base excision, in contrast to QΦQ (AlkZ) and YcaQ (QΦD), in which

mutation of either residue within the motif affects base excision activity (Mullins *et al.*, 2017; Bradley *et al.*, 2020). Second, the two HΦQ motifs in TxnU4 and LldU1 have different effects for TXNA- and LLD-G adducts, respectively; the histidine in TxnU4 had the greater effect on excision of gutingimycin and the glutamine in LldU1 had the greater effect on LLD-G excision. The cross-reactivity of TxnU4 and LldU1 against TXNs and their high sequence similarity suggests that the two glycosylases have similar substrate recognition pockets, and thus the different effects of their His and Gln mutants most likely stem from the manner in which TXNA-G and LLD-G lesions are positioned within the active site (Fig. 63D). These compounds are distinguished by the sugar substituents at position 13 (TXNA) and 16 (LLD) (Fig. 63A, C), which reside in the major groove and thus likely are contacted by helix αI as described above (Fig. 63A-C). Interestingly, TxnU4 and LldU1 contain a 10-15-amino acid insertion in helix αI that the AlphaFold model predicts forms a β-hairpin (Figs. 62A-B, 63A-C). Steric interaction from this helix αI insertion with the unique 13- and 16-sugar substituents in the major groove would displace the TXNA- and LLD-DNAs differently, placing the target deoxyriboses of TXNA-G and LLD-G in proximity to His43 and Gln45, respectively (Fig. 63D). Thus, although TxnU4 and LldU1 share the same catalytic motif, the insertion in the predicted drug-binding αI-helix and the differences in sugar moieties in TXNs may alter how the two proteins engage their substrates. Consistent with this rationale, neither LldU5, AlkZ, nor YcaQ contain the αI helix insertion, and none of these show a preferential catalytic residue within QΦQ or QΦD motifs (Mullins *et al.*, 2017; Bradley *et al.*, 2020).

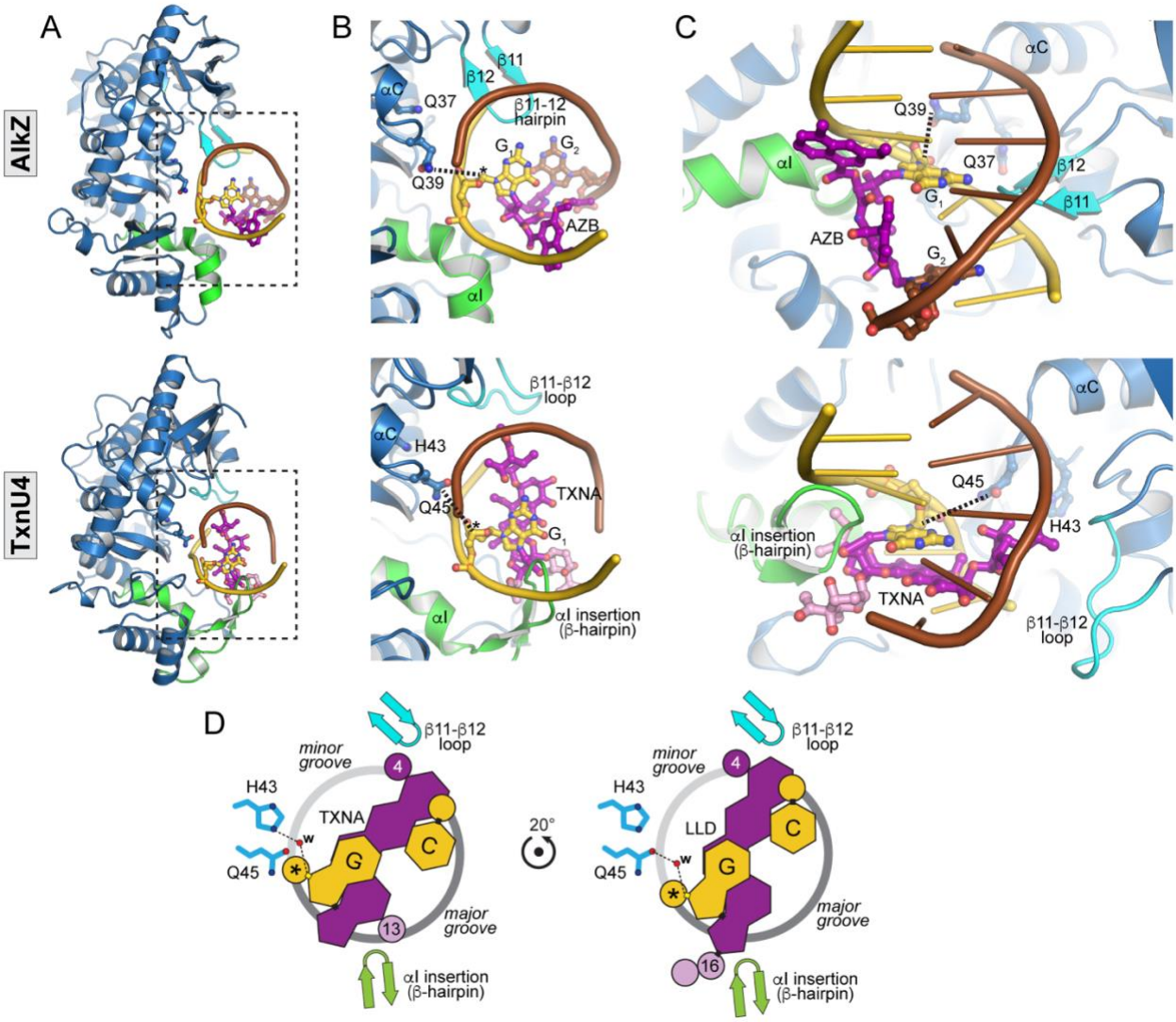

A
B
C
AlkZ
TxnU4
αC
β11
β12
Q37
β11-12
hairpin
G1
G2
Q39
AZB
αI
β11-β12
loop
H43
TXNA
Q45
αI insertion
(β-hairpin)
D
minor
groove
major
groove
4
13
16
20°
LLD
C
G
w

Figure 63. *DNA binding models of AlkZ and TxnU4* (A-C) The AlkZ crystal structure (PDB ID 5UUJ) docked against an AZB-ICL-DNA model (Alcaro and Coleman, 2000) is on top and the TxnU4 AlphaFold2 model docked against the TXNA-DNA crystal structure (Pfoh *et al.*, 2008) is on bottom. (A) DNA is docked into the concave cleft containing active site residues (ball-and-stick) and secondary structural elements (helix αI, green; β11-12 hairpin, cyan) predicted to contact the DNA lesion (Mullins *et al.*, 2017). (B) Zoom-in of region outlined by the dashed box in panel A. AZB and TXNA are colored purple. Sugar substituents on TXNA that differ between LLD are colored pink. Guanine residues covalently attached to the natural products are shown in ball-and-stick and C1′ of the adducted nucleotide is labeled with an asterisk. Dashed lines represent putative water-mediated contacts between the C-terminal glutamine of the QΦQ and HΦQ catalytic motifs and the target C1′. In both models, slight rotation of the DNA around the helical axis would place the adducted nucleotide in proximity to Q37 (AlkZ) or H43 (TxnU4). (C) Side view looking into the concave cleft. Helix αI (green) protrudes into the major groove side of the lesion and is predicted to interact with AZB and TXNA directly. The β11-12 hairpin in AlkZ and corresponding loop in the TxnU4 model are positioned on the minor groove side of the lesion. Because TXNA (and LLD) intercalate the DNA and span both major and minor grooves, the β11-12 loop likely interacts with the intercalated compounds directly. (D) Schematic of TxnU4/TXNA-DNA (left) and LldU1/LLD-DNA (right) complexes, viewed down the helical axis and colored as in panels A-C. The grey circle represents the DNA backbone, with major and minor grooves colored dark and light grey, respectively. Sugar moieties are depicted as circles and labeled numerically with their position in TXN. Unique sugar substituents between TXNA and LLD are pink. The catalytic water is labeled “w”. TXNA-DNA is rotated 20° clockwise relative to LLD-DNA, which places the target N-glycosidic bond (yellow rectangle) within water-mediated hydrogen bonding distance to H43. The DNA in each complex is likely held in place by interactions between helix αI insertion and the unique sugar substituents on TXNA and LLD.

A growing number of specialized DNA glycosylases produced from the BGCs of genotoxic secondary metabolites have been determined, including those involved in self-resistance to AZB and yatakemycin/CC-1065 (Xu *et al.*, 2012; Wang *et al.*, 2016; Mullins *et al.*, 2021). Our cellular resistance/sensitivity assays demonstrate *txnU2/4* and *lldU1/5* are key determinants in self-resistance to TXNA/LLD. The presence of multiple copies of these DNA glycosylases is unique to the *txn* and *lld* BGCs, and may provide redundancy to ensure repair of the highly genotoxic TXN metabolites, in contrast to the lethality of AlkZ knockouts in azinomycin B-producing *S. sahachiroi* (Wang *et al.*, 2016). Based on

our finding that TxnU4 and LldU1 play the major roles in toxin resistance, it is interesting to speculate that TxnU2 and LldU5 play more secondary roles, such as removing lesions formed by TXN derivatives generated either from an alternative biosynthetic route or from catabolism of TXNA/LLD.

The subsequent BER steps necessary for repair of DNA lesions generated from secondary metabolites, and the roles of other pathways (e.g., NER) are remaining questions. Regarding BER, our finding that *E. coli* EndoIV processed TXNs AP-sites less efficiently than 7mG-derived AP-sites suggests that specialized nucleases act on the AP-DNA/TXN-Gua product, as predicted for the putative *ytkR4* and *ytkR5* nucleases located within yatakemycin BGC (Huang *et al.*, 2012; Xu *et al.*, 2012; Mullins *et al.*, 2017). Although there do not appear to be any nucleases within the *txn*/*lld* clusters, genomic analysis reveals both ExoIII and EndoIV orthologs in TXNA/LLD producing strains (and two ExoIII paralogs in the case of *S. bottropensis*). Given the bulky, helix-distorting nature of these compounds, it is also likely that NER or other pathways play a role in their repair, as previously shown for yatakemycin-family and NM-ICL-DNA lesions (Selby and Sancar, 1988; Jin *et al.*, 2001; Kiakos *et al.*, 2007; Mullins *et al.*, 2017; Bradley *et al.*, 2020). Indeed, *S. vinaceusdrappus* and *S. bottropensis* contain one and three UvrA paralogs, respectively. It is also possible that TXN-DNA lesions are recognized by other enzymes outside of BER or NER, as reported for the structure-specific AziN nuclease within the AZB BGC (Chen *et al.*, 2020). More work is needed to elucidate the full landscape of cellular mechanisms of repair of these unique DNA damaging agents. Taken together, this work characterizes a unique family of DNA glycosylases from the HTH_42 superfamily that act on heavily functionalized, intercalated DNA adducts, and provides

further evidence for that DNA glycosylases residing in BGCs have evolved an exquisite specificity for aberrant nucleotides formed by their cognate genotoxic natural products.

[5] This work is published in: Chen, X.[§], Bradley, N.P.[§], Lu, W.[§], Wahl, K.L., Zhang, M., Yuan, H., Hou, X.-F., Eichman, B.F.*, and Tang, G.-L.* (2022). Base excision repair system targeting DNA adducts of trioxacarcin/LL-D49194 antibiotics for self-resistance. *Nucleic Acids Res*. I designed and conducted experiments in Figures 4-7, wrote and edited the manuscript, and performed the revisions. [§]authors contributed equally; *co-corresponding authors.

Chapter 6

Concluding Discussions and Future Directions

Introduction

Fifteen years ago, DNA glycosylases were believed to only have the ability to excise small, single-base modifications from the genome due to structural constraints within the enzymes. It was thought that all superfamilies of DNA glycosylases used a conserved nucleotide flipping mechanism, and that their structural and evolutionary diversity was well understood (Schärer and Jiricny, 2001). One thing that is certain in biology is that the diversity we have observed so far in nature has merely scratched the surface of the true reservoir of evolutionary breadth. Fast forward to today (2022), and we have had the exceptional opportunity to observe the discovery of extremely diverse and specialized DNA glycosylases with involvement in many aspects of biology (Mullins *et al.*, 2019). From novel structural mechanisms of base excision, to new and diverse DNA lesions, and even involvement in additional conserved DNA repair pathways- DNA glycosylases have remained an important and interesting family of enzymes to study. In this final section, I will present preliminary data involving structural modeling of AZL and YQL proteins, diverse evolutionary roles of AlkZ proteins through genetic studies, and new substrates of proteins in the YQL family and potential implications in human health and disease. I will synthesize my work and thoughts at the end and discuss important future directions for these projects from structural, functional, and genetic perspectives.

Preliminary Results

Structural models for natural product genotoxin and ICL repair by AZL/YQL glycosylases

A major question which has emerged during the research conducted in this thesis pertains to structural interactions of AZL and YQL glycosylases with their cognate DNA damage- whether that be a natural product genotoxin or interstrand crosslinks of various types. This is an important process to understand, as it will provide detailed mechanistic insights into this superfamily of enzymes and answer key outstanding questions such as: how do these proteins recognize their lesions and discriminate against undamaged DNA?, how do these enzymes catalyze base excision of substrates which are recalcitrant to nucleotide flipping (ICLs, bulky intercalating agents)?, and how is the specificity or promiscuity of these enzymes defined at the atomic level? In this section, I will attempt to generate hypotheses for these questions based on structural modeling of AZL and YQL proteins with different DNA substrates, and describe efforts towards determining a high-resolution structure of a YQL homolog in complex with a DNA substrate that mimics the product of the glycosylase reaction (THF- tetrahydrofuran).

The crystal structure of AlkZ in a form unbound to a DNA ligand has provided us with tremendous information regarding the overall 3-D organization of this enzyme and some of the residues/regions of the protein involved in the catalysis of base excision (Mullins *et al.*, 2017) (Fig. 64A, left). The three tandem winged helix-turn-helix motifs scaffold into a C-shaped structure with the catalytic residues pre-positioned for base excision activity. YcaQ from *E. coli* can remove the same AZB-ICLs that AlkZ does, but also has a much broader substrate repertoire including nitrogen mustard crosslinks and

monoadducts (Bradley *et al.*, 2020). AZL and YQL enzymes reside within the same superfamily, but represent distinct subfamilies within this clade, which signifies that despite being homologs to one another, they perform unique functions and activities in a cellular context (Bradley *et al.*, 2022). Unfortunately, only very small amounts of *E. coli* YcaQ can be expressed and purified for biochemical studies. Due to this poor solubility and yield, YcaQ from *E. coli* is not the ideal candidate for structural studies by X-ray crystallography at this time. However, we are still able to generate models of YcaQ's structure through homology modeling with AlphaFold (Jumper *et al.*, 2021) (Fig. 64A, right). From this model, YcaQ is predicted to have a similar overall architecture to *S. sahachiroi* AlkZ, but with some potentially key differences in important regions of the protein. YcaQ is expected to organize the tandem winged helix motifs into a C-shaped structure, but the key features present in AlkZ seem to be different in the YcaQ model (Fig. 64A). Specifically, the β11-12 hairpin of AlkZ is modeled with fairly high confidence in YcaQ as loop structure, although some degree of alpha helical content might be present (Fig. 64A). The αI helix in AlkZ which is predicted to interact with the AZB moiety in the major groove is modeled in YcaQ as an extended disordered loop, but the confidence in the model at this specific region is relatively low from the AlphaFold statistics (Fig. 64A).

Using this AlphaFold model of YcaQ, I can propose hypotheses about how this enzyme is recognizing and performing base excision of crosslinked nucleotides. A model of a hydrazine NM_5-ICL was docked onto the AlkZ-AZB-ICL DNA binding model using the crosslinked guanines as an anchor point, and the YcaQ AlphaFold model was aligned onto the AlkZ structure to gain insights into the possible mechanisms of DNA binding, lesion recognition, and catalysis (Fig. 64B). Alignment of this NM_5-ICL onto the YcaQ

model shows a high surface complementarity between the kinked DNA structure and the active site surface of the enzyme, as NM_5-ICLs are expected to distort the DNA duplex (Bauer, 1997). This concave surface that contains the active site residues is lined with putative DNA binding residues such as positively charged Lys and Arg side chains, including on the loops surrounding the cleft (Fig. 64B). More detailed analysis of the active site of YcaQ in this NM_5-ICL DNA binding model shows the 11-12 loop on YcaQ is predicted to insert into the minor groove on the top of the DNA, and highly conserved aromatic residues lie within this cleft and surround the catalytic Q45 and D47 residues (Fig. 64C). These aromatic residues (W85, W112, and Y314) are highly conserved amongst YQL enzymes, and could serve to either scaffold the active site, or possibly play a catalytic role in the chemistry of base excision (Fig. 64C). Tyr314 for example, is positioned in relative orientation to the catalytic residues already characterized, and could possibly help orient/activate a water molecule during the hydrolysis step; mutation of these residues and testing of excision activity are thus important for understanding the architecture and mechanism of these enzymes. Importantly, the symmetric NM_5-ICL can be rotated 180° along the long axis of the DNA to position the crosslink for unhooking by YcaQ in either orientation, which is what we observed through my biochemical studies (Bradley *et al.*, 2020). In summary, the mechanistic details for how these glycosylases function is still actively being pursued in high-resolution, and a structure of either AZL or YQL proteins in complex with DNA would provide useful information to answer these important questions.

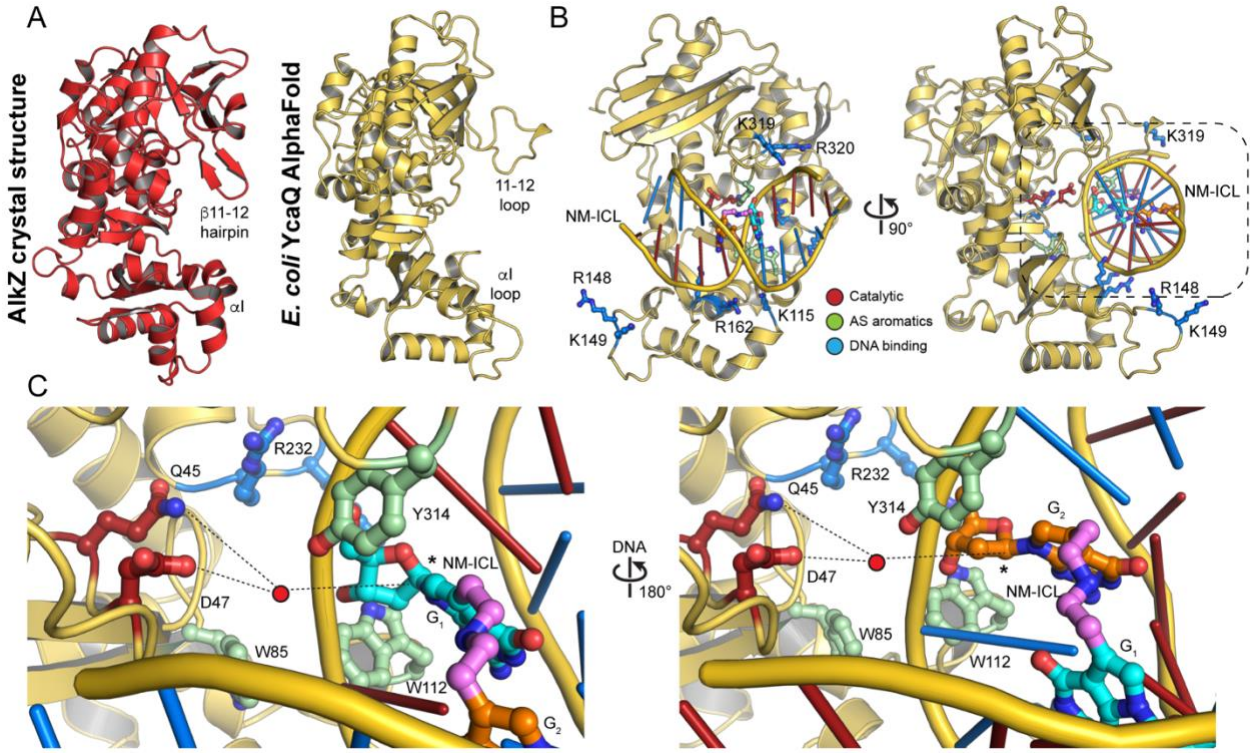


Figure 64. *E. coli YcaQ AlphaFold binding model with NM_5-ICL-DNA* (A) Structural comparison between the *Streptomyces sahachiroi* AlkZ crystal structure (PDB: 5UUJ) (red) and an AlphaFold homology model of *E. coli* YcaQ (Jumper *et al.*, 2021). Secondary structure around the αI helix and β11-12 hairpin in AlkZ are labeled. (B) DNA binding model of *E. coli* YcaQ AlphaFold model with a hydrazine NM_5-ICL model. Putative DNA binding residues are colored in blue, catalytic residues in red, and active site (AS) aromatic residues in green. (C) Zoom in of the YcaQ active site (dotted lines in panel B) with highly conserved residues shown in ball-and-stick and labeled. Dashed lines represent hydrogen bonding; catalytic water molecule is in red. DNA is rotated 180° perpendicular to the long axis between the left and right panels.

Additional models for how AZL enzymes can recognize and excise a diverse range of natural product genotoxin DNA damage can be explored to gain insight towards the incredible specificity and efficacy of these enzymes. In our bioinformatics project involving genome mining of bacterial genotoxins, we identified and characterized the AlkZ homolog in the hedamycin BGC (HedH4) as a glycosylase with exquisite specificity towards HED-guanine DNA adducts (Bradley *et al.*, 2022). Hedamycin forms complex, multivalent DNA

damage whereby the aromatic anthracycline portion intercalates into the GC base step (5' the adducted guanine), the aminosugars thread into the minor groove on the opposing side of the duplex, and the diepoxybutane warhead is perfectly positioned for alkylation at N7 guanine (Hansen *et al.*, 1995) (Fig. 65A). The *N*-glycosidic bond connecting N9 of the HED-guanine to C1' of the deoxyribose heavily shields this bond from spontaneous/thermal hydrolysis, and the intercalation of the hedamycin would be expected to allow the reformation of the *N*-glycosidic bond if spontaneous depurination did occur (Hansen *et al.*, 1995; Drohat and Maiti, 2014; Bradley *et al.*, 2022) (Fig. 65A). This intimate association of the hedamycin with the DNA, and the specificity of HedH4 for only removing HED-guanine suggests there are unique structural features of HedH4 for HED-DNA recognition and excision. To investigate this, I again generated AlphaFold homology models of HedH4 and aligned this to the HED-DNA NMR structure in a similar process as has been done for the other DNA binding models (Hansen *et al.*, 1995; Mullins *et al.*, 2017; Chen *et al.*, 2022) (Fig. 65B-C). The HED-DNA can be theoretically positioned in the HedH4 binding cleft so that the aminosugars which line the minor groove are facing upwards (Fig. 65B) or downwards (Fig. 65C), however the models are not suitable for discriminating the orientation of the HED-DNA. Just like with our other models with YcaQ, TxnU2/4, and LldU1/5, the secondary structure of regions surrounding the active site cleft are the points of greatest predicted differences between the proteins (Fig. 65B-C). The β11-12 hairpin in AlkZ is modeled (with high confidence) as an α-helix in HedH4, which is different than what was predicted for any of the other proteins (Fig. 65B-C, orange). The αI portion of HedH4 is also modeled as a kinked helix (just like in the AlkZ crystal structure), which suggest perhaps a similar mechanism for recognition of the

specific natural product in this region (Fig. 65B-C, red). A surface filling representation of HedH4 shows an extremely high degree of complementarity between the cleft of the enzyme, and the structure of the HED-DNA (Fig. 65B-C). This suggests a steric-based rationale for why HedH4 (and other AZL enzymes) are specific for their BGC genotoxin-glycosylase pair; the secondary structure surrounding the active site cleft selects for the type of damage which can be bound and excised. This of course does not eliminate the possibility of specific residues interacting with the drug-DNA complex, but mutational studies involving the swapping of the 11-12 region and αI between these AZL glycosylases will help inform the specificity. Currently, structural efforts are underway to co-crystallize several AZL enzymes with their cognate natural product DNA damage, and high-resolution atomic structures of this would be monumental in the investigation of these enzymes unique specificity and activity.

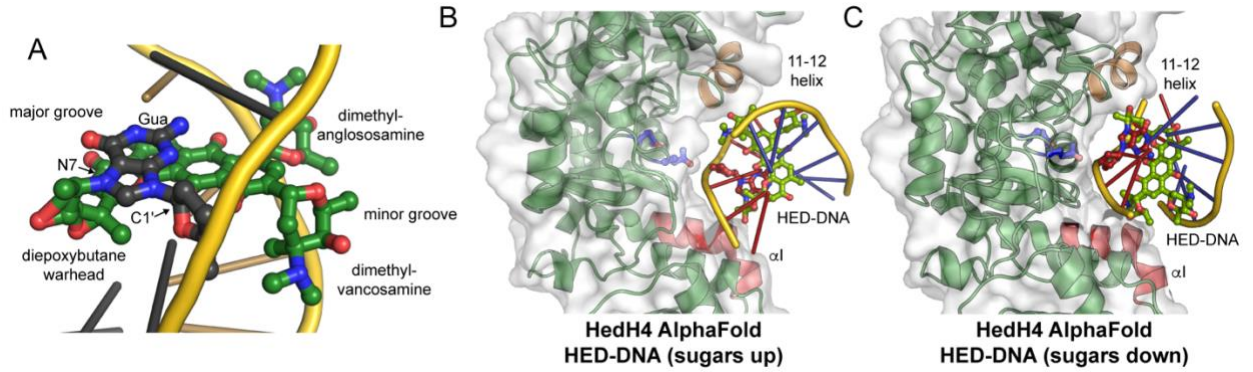


Figure 65. *Streptomyces griseoruber HedH4 AlphaFold binding model with HED-DNA* (A) NMR solution structure of hedamycin (green) in complex with DNA (PDB: 1JHI). Sugar substituents are located in the minor groove opposite the adducted guanine. The anthracycline core is intercalated 5' to the guanine, and the diepoxybutane warhead is alkylated at N7-guanine (black). (B) Structural modeling of a HedH4 (green) AlphaFold homology model (Jumper *et al.*, 2021) in complex with HED-DNA (sugars up). HedH4 is shown as surface representation. The 11-12 predicted helix is colored in orange, and αI is colored in red. (C) Structural modeling of a HedH4 AlphaFold homology model in complex with HED-DNA (sugars down). Colors are the same as in panel B.

Crystallization of Thermobifida fusca YQL with THF/7mG-DNA

The relatively poor expression and solubility of *E. coli* YcaQ limits the use of X-ray crystallography to study this protein at high-resolution. Since one of the major goals of this project is to characterize at the atomic level how these proteins recognize and perform base excision of unique substrates, we therefore screened through homologs of YcaQ (YQL proteins) to find suitable candidates for crystallography. We identified two Actinobacterial YQL proteins from *Thermobifida fusca* (*Tfu*; 34% identical, 50% similar) and *Thermomonospora curvata* (*Tcu*; 32% identical, 48% similar) which had good protein expression and solubility, and characterized their glycosylase activity towards 7mG and NM_8-ICLs (Bradley *et al.*, 2022) (Fig. 49A-B). A sequence alignment between *E. coli* YcaQ and the YQL glycosylase from *Tfu* shows the relatively high degree of similarity between the proteins (50%), and highlights the significant regions where insertions or deletions occur between the enzymes (Fig. 66A). The YQL glycosylase from *Tfu* displayed the best solubility and protein qualities, and I was able to purify suitable amount of YQL^{Tfu} for co-crystallization studies with a product-mimic DNA substrate (Fig. 66B-C). The product-mimic we choose to use is an abasic (AP) site mimetic comprised of a tetrahydrofuran (THF) and an intercalated 7-methylguanine free nucleobase (Fig. 66B). THF mimics the AP site structurally, but lacks the C1' hydroxyl group which allows for deoxyribose ring opening and strand cleavage. Generally, DNA glycosylases will bind tightly to the THF site to form a ternary product complex (Mullins *et al.*, 2015). The inclusion of the free 7mG nucleobase ensures that base stacking is maintained within the duplex, as it is unlikely these YQL enzymes use a base-flipping

mechanism or contain intercalating residues. To test for binding affinity of YQL^{Tfu} to THF/7mG-DNA, a fluorescence anisotropy binding assay was performed with YQL^{Tfu}, AlkA from *E. coli*, and *S. sahachiroi* AlkZ (Fig. 66D). As known previously, AlkA^{Eco} can bind with relatively high affinity to THF-DNA (low μM in this experiment), but AlkZ^{Ssa} showed nearly no binding to this THF/7mG-DNA, which indicates it may have relatively low affinity towards AP sites (Fig. 66D). YQL^{Tfu}, however, displayed the highest affinity for the THF/7mG substrate in this fluorescence-based assay, with a K_d in approximately the low-mid nM range (Fig. 66D). Next, I was interested in crystallizing the ternary complex between YQL^{Tfu} and a THF/7mG-DNA. For this I screened a library of THF/7mG oligonucleotides with varying lengths and overhangs, and obtained needle-like crystals with a blunt end ds18-mer THF/7mG-DNA (Fig. 66E). Crystals were grown in optimization trays to improve crystal morphology, and several candidate crystals were looped, cryoprotected, and frozen for data collection at the APS synchrotron. Preliminary diffraction indicated that indeed there was protein present within the crystals, but the diffraction was too weak and not at high resolution to collect a full dataset for structure determination (Fig. 66F). Moving forward, these crystals will be optimized through seeding experiments (both micro- and macro-seeding), and by optimizing the sequence of YQL^{Tfu} (remove any disordered regions, decrease surface entropy) so that a high-resolution structure of YQL^{Tfu} in complex with THF/7mG-DNA can be achieved.

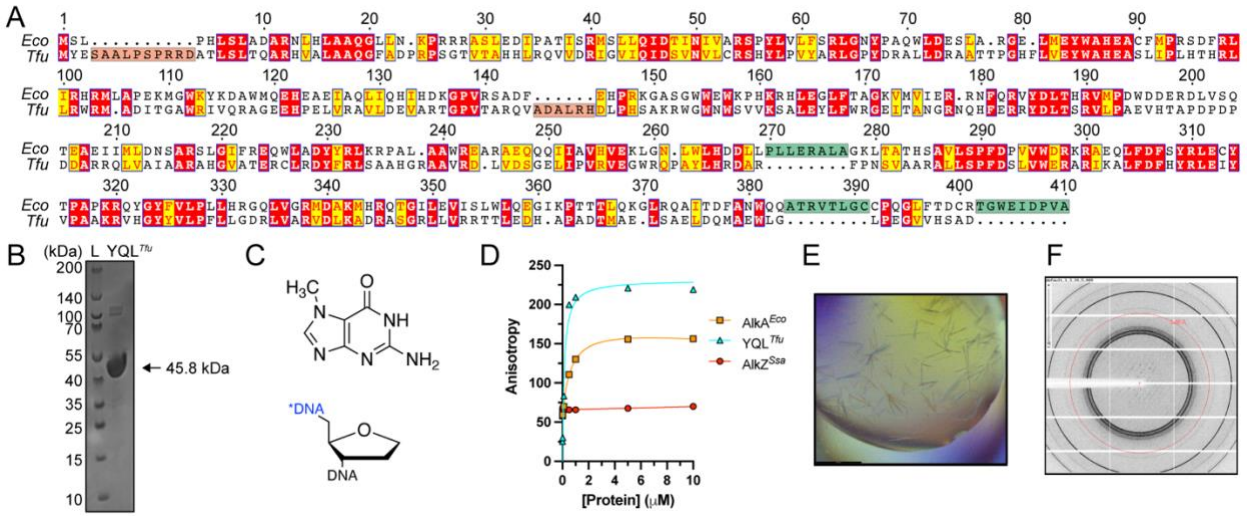


Figure 66. *Sequence alignment, DNA binding, and crystallization of Thermobifida fusca YQL with THF/7mG-DNA* (A) Amino acid sequence alignment between *Escherichia coli* YcaQ (top) and *Thermobifida fusca* YQL (bottom). Numbering is in references to *E. coli* YcaQ. Red-shaded boxes are identical residues; yellow-shaded boxes are similar residues. Insertions in *E. coli* are boxed in green; insertions in *T. fusca* are boxed in salmon. (B) SDS-PAGE gel of purified YQL from *T. fusca*. (C) Chemical structure of a tetrahydrofuran (THF)/7mG-containing DNA used for binding (blue FAM 5' label) and crystallization experiments. (D) Fluorescence anisotropy DNA binding of *E. coli* AlkA (orange), *T. fusca* YQL (cyan), and *S. sahachiroi* AlkZ (red) to a FAM-labeled oligo containing a central THF/7mG nucleobase. (E) Crystals of *T. fusca* YQL with an 18-mer THF/7mG oligonucleotide substrate. Crystals were grown in: 10% w/v PEG 8K, 0.1 M MES pH 6.0, 0.1 M NaCl, and 0.05 M $CaCl_2$. (F) Preliminary X-ray diffraction of crystals of *T. fusca* YQL with an 18-mer THF/7mG DNA.

Horizontal gene transfer of AZL into chytrid fungal eukaryotes

During the bioinformatic analysis of AZL proteins in *Streptomyces*, I noticed a series of AZL sequences distributed amongst a small subset of Eukaryotic species, specifically within several chytrid fungi. These sequences come from five *Spizellomyces*, *Powellomyces*, and *Piromyces* species within the Chytridiomycota phylum, and contain the catalytic residues found in *S. sahachiroi* AlkZ (Fig. 67A). Chytrid fungi are zoosporic organisms which are highly diverse in their habitat, life cycle, and ecology. These AZL protein sequences range from 24-28% identity to $AlkZ^{Ssa}$, and all are intron-less genes

within the fungal species. The finding of AZL sequences exclusively within these fungal species signifies that this may be an inter-domain horizontal gene transfer (HGT) event between an ancient bacterial species and a common ancestor of chytrid fungi. Consistent with an HGT event into these fungal species, some of the further diverged fungal species contain an extended N-terminus which contains a strong predicted nuclear localization signal (NLS), likely for directed localization of the AZL protein into the Eukaryotic nucleus (Fig. 67B). A taxonomic assessment of the fungal species which contain an AZL protein shows that a Chytridiomycota ancestor likely obtained the AZL gene though an HGT event ca. 800 MYA, and it was retained in several species while the phylum diverged (Fig. 67C). The *Piromyces* genus did not evolve an NLS sequence on the AZL protein, while *Spizellomyces* and *Powellomyces* do possess this NLS, which apparently appeared ca. 500 MYA (Fig. 67C). The relatively small size of these fungal AZL proteins (~46 kDa) may have originally allowed for passive diffusion through the nuclear pore complex, and later a NLS was added to increase the efficiency of transport. A phylogenetic assessment of *Spizellomyces punctatus* (*Spu*) AlkZ in relation to the top 50 BLASTp hits supports a HGT event, as all the fungal AZL sequences are most highly related to gene sequences of HTH42 superfamily proteins within the Actinobacteria, Chloroflexi, and Proteobacteria phyla (Fig. 67D). This apparent identification of an inter-domain HGT event between a bacterial species which contained an ancient AZL gene and an ancestor of a chytrid fungal species signifies the importance of this self-resistance DNA glycosylase in the chemical warfare which occurs in soil-dwelling microbes such as *Streptomyces* and these chytrid fungi, which could be how this HGT event occurred (shared habitat).

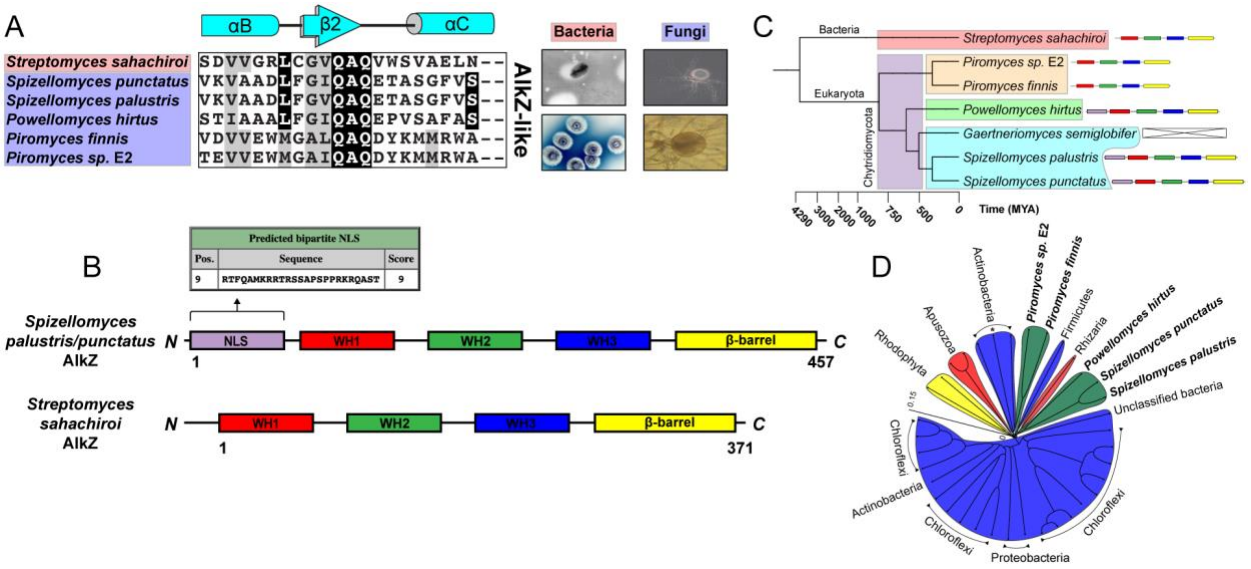


Figure 67. *Phylogenetic and taxonomic evidence of AlkZ horizontal gene transfer into chytrid fungal eukaryotes* (A) Sequence alignment between the catalytic region of *Streptomyces sahachiroi* AlkZ and chytrid fungal AlkZ sequences identified in Pfam. Secondary structure of AlkZ is shown above the alignment. (B) Motif diagram comparing fungal (top) and bacterial AlkZ (bottom) proteins. Three of the five fungal AlkZ proteins contain a strong N-terminal nuclear localization signal (NLS; lavender) as mapped by cNLS Mapper (Kosugi *et al.*, 2009). Individual winged-helix (WH) domains are colored in red (1), green (2), and blue (3); C-terminal extension is in yellow. (C) Taxonomy of fungal species containing an AlkZ ortholog (except *Gaertneriomyces*) with respect to *Streptomyces sahachiroi* (outgroup). Time (MYA) is on the X-axis. Red: bacterial phyla, purple: Chytridiomycota phylum, tan: *Neocallimastigaceae* family, green: *Powellomycetaceae* family, cyan: *Spizellomycetaceae* family. Motif arrangement of AlkZ in these genomes is diagrammed to the right of the tree. (D) Phylogeny of *Spizellomyces punctatus* AlkZ proteins in alignment with the top 50 tBLASTn hits. Tree scale is in amino acid substitutions per site. Asterisk denotes *Streptomyces sahachiroi* AlkZ. Blue: bacterial phyla, yellow: *Rhodophyta* phylum (red algae), red: protist phyla, green: fungal phyla.

Having established a HGT event of these AZL proteins into these fungal species, the next goal was to test their glycosylase activity *in vitro* to establish their substrate preferences, if any. I choose to analyze AlkZ from *Spizellomyces punctatus* (*Spu*), whereby I deleted the residues at the N-terminus which were predicted to be the NLS motif, and purified the protein for biochemical studies (Fig. 68A). AlkZ^{Spu} was then tested for base excision activity against a panel of *N*7-alkyl DNA substituents derived from a

variety of DNA damaging agents (monoadducts, ICLs, bulky intercalators) (Fig. 68B). From this excision assay, AlkZ^{Spu} appears to have robust activity for removing 7mG and unhooking ICLs derived from NM_5- and NM_8-ICLs, but low activity for the intercalating alkylators (HED, LLD, TXNA) (Fig. 68B). This activity for the nitrogen mustard ICLs is higher than it was for *S. sahachiroi* AlkZ, which could signify a diversification of substrates for this AZL protein after the HGT event. It will of course be of interest to determine the structure of AlkZ^{Spu} to see if there are any major differences between it and AlkZ^{Ssa}, as this may provide some insight into the differences in activities. Finally, now that several of these chytrid fungal species which contain an AZL protein are being sequenced and genetic manipulation systems are being developed to study their biology, the involvement of these AZL proteins in genotoxin resistance can be studied at an organismal level (Russ *et al.*, 2016; Medina *et al.*, 2020).

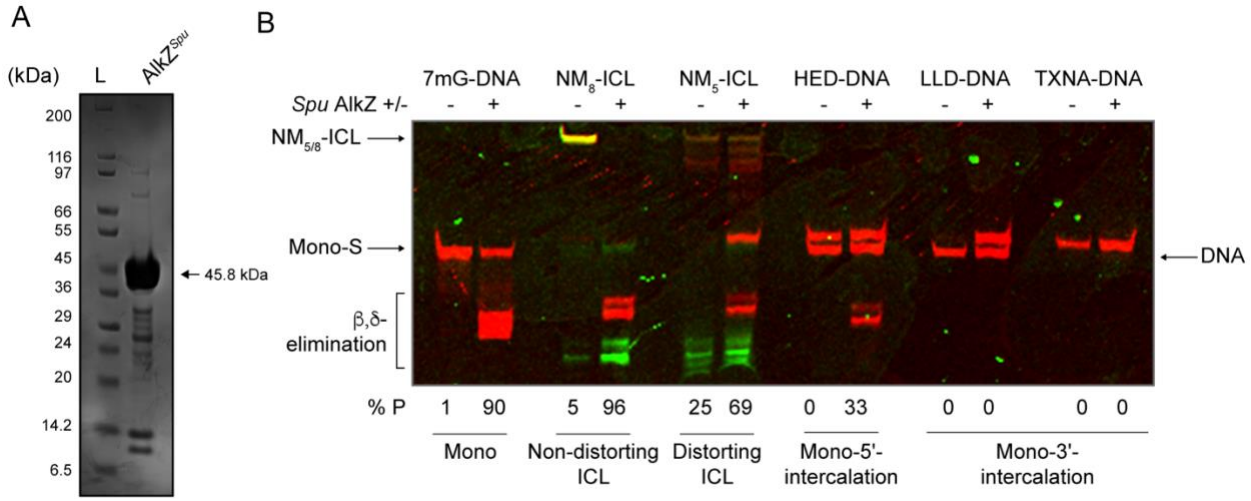


Figure 68. *Purification and base excision activity of Spizellomyces punctatus AlkZ ΔN* (A) SDS-PAGE gel of purified AlkZ (ΔN-terminus) from the fungal species *Spizellomyces punctatus*. (B) Denaturing PAGE gel of various alkylated DNA substrates tested against *Spu* AlkZ (after 6 hr incubation). From left: 7mG-DNA (simple monoadduct), NM_8-ICL (non-distorting ICL), NM_5-ICL (kinked ICL), HED-DNA (5' intercalator), and LLD/TXNA-DNA (3' intercalators). Mono-S refers to monoadduct substrates; DNA label is unreacted. Gel is false colored with FAM (green) and Cy5 (red) channels. % product (%P) is labeled below each lane.

Genome mining of AZL homologs in all bacterial phyla

While bacteria in the genus *Streptomyces* are a rich and abundant source of natural products which have important therapeutic potential in virtually every aspect of human health and disease, there are still many secondary metabolites which are produced beyond *Streptomyces* that have importance. For this reason, I am also interested in mining the available genomes of all bacterial phyla for AZL homologs and their proximity to novel BGCs, with the ultimate goal of identifying new types of genotoxins. Similar to our previous genome mining approach in *Streptomyces*, I performed preliminary searches in the Actinobacterial phylum (taxid:201174) of several species which contained an AZL protein, and obtained several hits (Fig. 69A-D). None of the BGCs are of characterized origin, so the products which are predicted from these clusters are based off cluster BLAST analyses. Azicemicin A, for example, contains a planar angucycline ring system and an *N*-methylaziridine three-membered ring, which could be directed to DNA for alkylation, although the mechanism of action for this compound is also unknown (Tsuchida *et al.*, 1995) (Fig. 69A). Simocyclinone D8 is an aminocoumarin-tetraene-polyketide which is an inhibitor of DNA gyrase and an inhibitor of the SimR transcription factor, and contains an epoxide ring which could have the potential to alkylate DNA, although this has not been tested or observed (Flatman *et al.*, 2005; Buttner *et al.*, 2018) (Fig. 69B). Thailanstatin A also contains an epoxide ring and is an NRPS/polyketide antitumor/antibiotic which exerts its mechanism of action by inhibition of the splicing machinery in eukaryotic cells, leading to its cytotoxicity (Kaida *et al.*, 2007) (Fig. 69C). Finally, LLD family analogs can be discovered through this process,

as this family of antibiotics are apparently well-dispersed amongst Actinobacterial species (Fig. 69D). These compounds such as azicemicin A are targets for searching for a mechanism of action (MoA) such as through DNA alkylation since they currently do not have a known MoA. These newly identified biosynthetic gene clusters are rich sources of genetic information, whereby the whole cluster can be cloned and inserted into expression vectors to allow for the targeted identification of novel genotoxins and potential therapeutics.

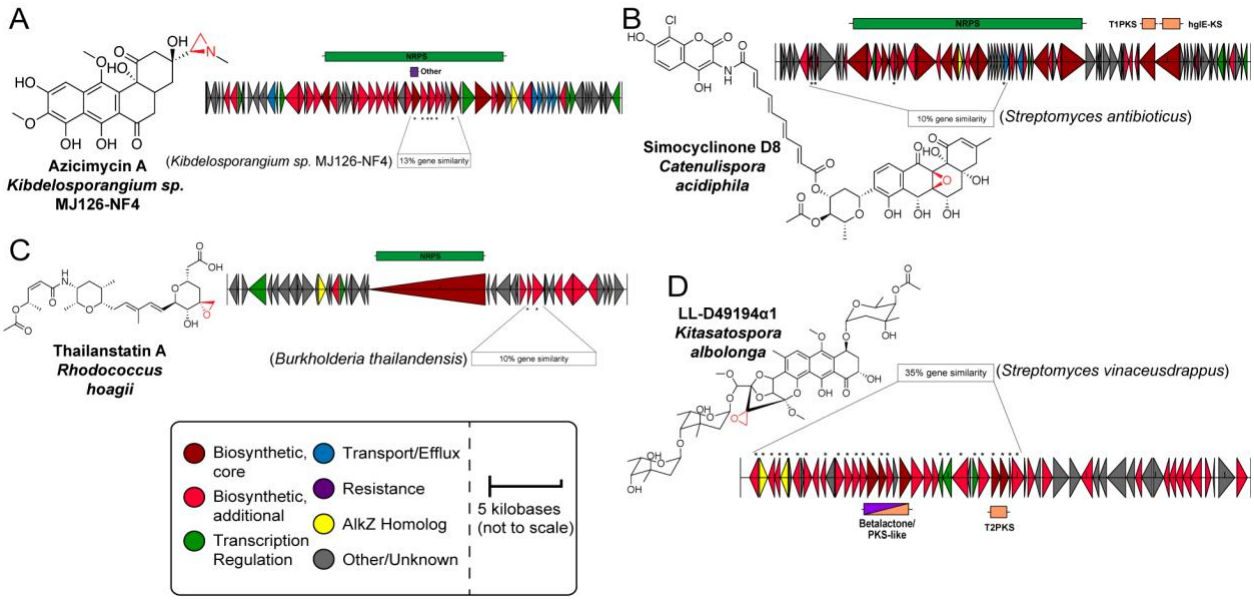


Figure 69. *Actinobacterial BGCs beyond Streptomyces containing AlkZ homologs* (A-D) Gene diagrams for AlkZ-containing BGCs in Actinobacterial genomes. The biosynthetic scaffold produced by specific genes in the cluster are labeled above the respective genes. NRPS, non-ribosomal peptide synthetase; PKS1/PKS2, type 1/2 polyketide synthase; (PKS), PKS-like. Chemical structures of the metabolites produced by each cluster are shown at the bottom of each panel. Potential sites of reactivity are highlighted in red. % similarity is in reference to the producing strain.

Expanding the substrate repertoire of E. coli YcaQ

One of the major questions which has emerged from the discovery of a DNA glycosylase in *E. coli* (YcaQ) which can recognize and remove *N*7-Gua/*N*7-Gua

interstrand crosslinks from DNA is a significant one: what could the endogenous or preferred substrate of the YcaQ enzymes be? This is an important question because if *E. coli* are being exposed to genotoxins which form ICLs, then it is likely that humans can be exposed to the same agents, as *E. coli* and related bacteria reside in our digestive tract as normally commensalistic microbes (Conway and Cohen, 2015). As it turns out, within the past 15 years or so, the appreciation for the contribution of the gut microbiota to the progression of colorectal cancer (CRC) has advanced tremendously, and we now understand that *E. coli* can produce agents which create ICLs in human gut epithelia and enhance the transformation of CRC tumor formation (Bonnet *et al.*, 2014). While microorganisms such as *E. coli* are generally not known to produce a large array of secondary metabolites from BGCs, a cluster found in strains associated with the progression of CRC was found to produce the natural product genotoxin, colibactin, which forms ICLs in infected human cells (Bossuet-Greif *et al.*, 2018; Xue *et al.*, 2019) (Fig. 70A).

The chemical structure of colibactin (CLB) was only elucidated within the past few years, and the mechanism, sequence specificity, and consequences of DNA crosslinking by CLB are still under intense investigation (Cuevas-Ramos *et al.*, 2010; Xue *et al.*, 2019). Colibactin is a hybrid non-ribosomal peptide/polyketide secondary metabolite produced by *pks*$^{+}$ *E. coli* and related bacterial species from the CLB gene cluster, which is located on a highly-mobile DNA element (Putze *et al.*, 2009; Fais *et al.*, 2018). Colibactin can bind within the minor groove of AT-rich regions of the genome, and proceeds through alkylation at N3 of adenines by opening of the two cyclopropane warheads in CLB; the CLB-ICLs can become oxidized to create further modified DNA damage (Wilson *et al.*,

2019; Xue *et al.*, 2019) (Fig. 70A). Like many forms of initial DNA damage, ICLs derived from colibactin are prone to further transformation into complex damage such as double strand breaks (DSBs). CLB-ICLs, because they place a positive charge on the adenine nucleobase, are prone to spontaneous depurination which would create dual AP sites within the genome (Xue *et al.*, 2020) (Fig. 70B). Abasic (AP) sites are highly toxic to the cell, and can undergo multiple pathways of either repair, or further damage (Fig. 70B). AP sites can of course enter the BER pathway though AP endonuclease (1 or 2 in humans) incision, but incision on both strands would create a DSB (Fig. 70B, *i*). AP sites can also be nicked spontaneously through reactive cellular metabolites, which catalyze the elimination of the 3' phosphate, leading to strand breakage (Fig. 70B, *ii*). AP sites can react with suitably placed exocyclic amines of nucleobases (particularly N2 of adenine) to create a different ICL, dA-AP-ICLs, which can be unhooked by the NEIL3 DNA glycosylase during replication (Price *et al.*, 2014; Imani Nejad *et al.*, 2020) (Fig. 70B, *iii*). Finally, AP sites can undergo DNA-protein crosslink formation with the AP site protection protein, HMCES, which forms a thiazolidine ring with the N-terminal cysteine residue to prevent strand breakage or TLS mutagenic bypass (Mohni *et al.*, 2019; Thompson *et al.*, 2019) (Fig. 70B, *iv*). The ultimate consequences of colibactin-induced DNA damage in humans is significant: the progression of CRC formation and tumorigenesis within the human gut. In general, *E. coli* will only produce colibactin under chronic inflammatory stress, which causes signature DSBs to form in the human genome (AT-rich regions) and activation of DNA repair pathways (Fig. 70C). This eventually leads to chronic gut inflammation, and ultimately the progression of CRC through a variety of cellular and

molecular mechanisms (Dziubanska-Kusibab *et al.*, 2020; Li *et al.*, 2020; Pleguezuelos-Manzano *et al.*, 2020) (Fig. 70C).

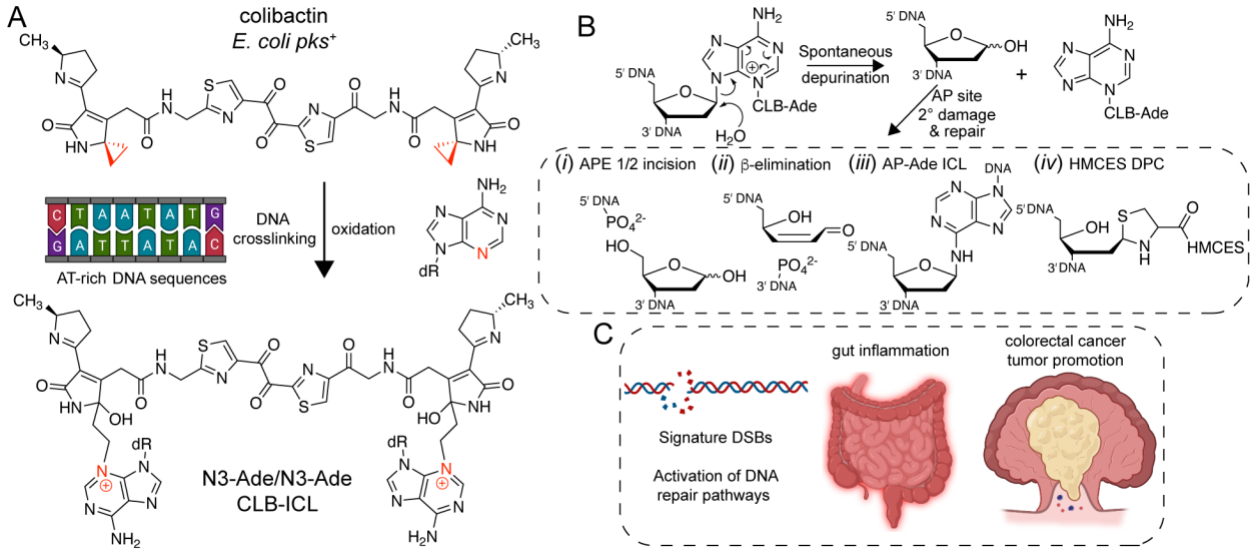


Figure 70. *Mechanism of colibactin (CLB) crosslinking and downstream effects of colibactin-induced DNA damage in progression of colorectal cancer (CRC)* (A) Chemical structure of colibactin from *pks*$^+$ *Escherichia coli* strains. Reactive cyclopropane rings are highlighted in red. Colibactin alkylates AT-rich DNA sequences at N3 of adenine (red) to created N3-Ade/N3-Ade CLB-ICLs. (B) Secondary damage caused by CLB-ICLs through spontaneous depurination of N3-Ade CLB-ICLs. The N-glycosidic bond of N3-Ade DNA adducts are destabilized by the positive charge on the nucleobase. Depurination of N3-Ade adducts generates an AP site which can enter DNA repair/protection pathways, or lead to secondary (2°) DNA damage through: (*i*) AP endonuclease incision, (*ii*) β-elimination catalyzed by base, (*iii*) formation of AP-Ade ICLs through reaction of N2-Ade, or (*iv*) AP site protection through DPC formation with HMCES. (C) Colibactin-induced DNA damage promotes signature double strand breaks (DSBs) in the genome and activation of DNA repair and inflammatory pathways in the human gut. Prolonged exposure to *pks*$^+$ *E. coli* can eventually lead to colorectal cancer progression through tumor promotion.

The discovery and appreciation for the impact of the gut microbiota and its role in human health and disease has spawned significant questions and research topics within the last decade. In relation to colibactin and the crosslinks it forms in DNA, the obvious question is how to bacteria repair and/or resist the potent damage created by exposure to colibactin and related metabolites? It is tempting to postulate that the YcaQ enzyme

system we have identified and characterized as an ICL glycosylase could play a role in repair of CLB-ICLs through a self-resistance mechanism, similar to AlkZ and AZB-ICLs (Wang *et al.*, 2016; Bradley *et al.*, 2020). Since the known damage caused by colibactin is at N3 of adenine, however, and I have not fully characterized the substrates of the YcaQ family, the data I have obtained are preliminary in origin, but provide a solid starting point for investigation. The biosynthesis of colibactin is still under investigation so that enough can be obtained for various studies, but for now cannot be obtained in high enough yields (Vizcaino and Crawford, 2015; Xue *et al.*, 2019). As an alternative to the true colibactin which is highly difficult to obtain, a synthetic “left-half” of colibactin (CLB-15a) has been synthesized by a collaborator which can be used for biochemical studies (Healy *et al.*, 2016) (Fig. 71A). CLB-15a contains the bis-thiazole region and a positively-charged dimethylamino group to direct the single cyclopropane warhead for alkylation, although the sequence specificity is not entirely known (Fig. 71A). To test for base excision activity of *E. coli* YcaQ against DNA adducts generated by CLB-15a, a fluorescent GC-rich oligo with a central adenine residue was reacted with CLB-15a in the hopes of creating a single adduct on the adenine (Fig. 71A). This, however, was not what was observed, and a mixture of alkylation at the central adenine (N3), and the flanking guanines (likely N7) was seen; importantly, the N7-guanine adducts can be chemically converted to formamidopyrimidine (FaPy) adducts with base treatment, which is important for stability and testing base excision (Fig. 71A). When these CLB-15a DNA adducts were tested for thermal stability, a mixture of alkali-labile products were generated, consistent with mixed alkylation at the central adenine and flanking guanines (Fig. 71B, lane 3). Incubation of these native CLB-15a DNA adducts with YcaQ generated alkali-labile AP

sites at the ± 3 flanking guanines in the sequence, but not at the central adenine (Fig. 71B, lane 5). Incubation with the other alkylpurine DNA glycosylases in *E. coli* (AlkA/Tag), or the YcaQ D47A catalytic mutant did not lead to any significant product formation (Fig. 71B). Consistent with the formation of N7-guanine adducts, conversion of the CLB-15a DNA to FaPy derivatives completely abolished the heat-lability and YcaQ excision activity of these adducts (Fig. 71B, right). These results together suggest that colibactin may form additional types of DNA damage, such as at N7 of guanine in addition to N3 adenine, and that YcaQ is capable of excising DNA adducts created by the colibactin monoalkylating derivative of CLB-15a.

Although full-length colibactin has yet to be obtained in significant amounts for biochemical and cellular studies, there are techniques available to generate CLB-ICLs in plasmids which are co-cultured with *pks*$^{+}$ *E. coli* strains. These CLB-ICLs have been provided through collaboration, and a preliminary unhooking assay with YcaQ has been under development (Fig. 71C-D). YcaQ is first incubated with the plasmid CLB-ICLs, and next with EndoIV; if incubation with YcaQ generates AP sites, then EndoIV will nick these and lead to increased relaxation of the DNA (Fig. 71C). Under a single time-point, I can observe that incubation with YcaQ leads to an increased proportion of nicked CLB-ICL DNA, whereas with mock or D47A this is not significantly observed (Fig. 71D, left). If I take a time course of this reaction, I can see a time-dependent nicking of the CLB-ICL plasmid DNA when incubated with YcaQ WT (Fig. 71D, right). These results suggest that YcaQ may play a cellular role in self-resistance to colibactin-induced DNA damage, although more studies are required to elucidate the mechanism and verify the types of colibactin damaged which is recognized and repaired by YcaQ.

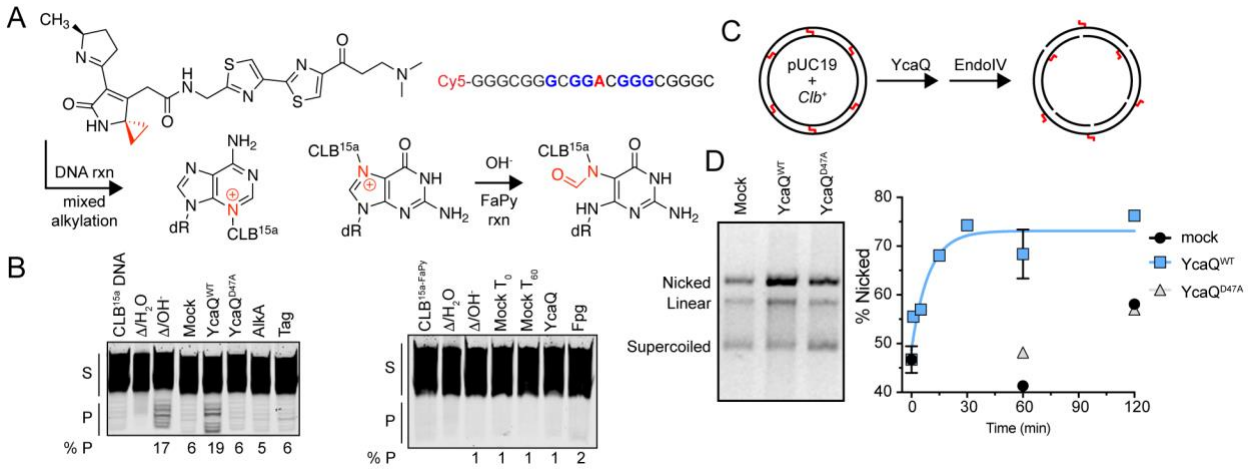


Figure 71. *Excision of a colibactin monoadduct by E. coli YcaQ and CLB-plasmid nicking* (A) Structure of the colibactin (CLB) 15a monoalkylator with reactive cyclopropane ring highlighted in red, and Cy5-DNA used in glycosylase reactions. Sites of observed reactivity are highlighted in red (adenine) or blue (guanines). Reaction of CLB^{15a} with this DNA sequence generated mixed alkylation products at the central guanines and the single adenine. Reaction of this mixed DNA alkylation with hydroxide converts N7-guanine to the N5-substituted FaPy lesion. (B) Base excision assays with native CLB^{15a}-DNA (left) and hydroxide-treated CLB^{15a}-DNA (FaPy; right). Reactions were incubated for 1 hour at 25°C before hydroxide work-up and denaturing PAGE. % P is the fold change of nicked DNA (P) intensity in the mock lane (first lanes) compared to the reaction lanes. (C) Cartoon diagram of the pUC19 Clb^+ ICL plasmid and the unhooking assay performed with *E. coli* YcaQ and EndoIV. Unhooking of ICLs by YcaQ would generate AP sites which could be incised by EndoIV to create nicked (relaxed) DNA. (D) Representative agarose gel of CLB-ICL plasmid unhooking assay with YcaQ WT and D47A catalytic mutant (left), and time course kinetics of nicked DNA generated by incubation with YcaQ WT (right).

Final Conclusions and Future Directions

The insights gained into bacterial DNA interstrand crosslink (ICL) repair and self-resistance to natural product genotoxins through our investigations of the AZL and YQL family of DNA glycosylases has provided significant avenues of research to continue exploring through many experimental facets. Structurally, it remains to be understood exactly how these evolutionarily diverse enzymes can recognize and excise a variety of highly complex and functionalized DNA damage which is normally repaired through other

coordinated DNA repair pathways. While it has not been experimentally established that these enzymes do not require a base flipping mechanism for excision, this is the most likely mechanism for excision of crosslinked or intercalated DNA damage. Establishing how these glycosylases can perform this reaction will be valuable for understanding the general principles of how DNA glycosylases can function. From a genetic perspective, the distribution of AZL and YQL genes in bacteria is still being worked out to understand their association with biosynthetic gene clusters (AZL) and gene neighborhoods/operons associated with their functions (YQL). The operon which YcaQ is part of in *E. coli* is conserved amongst related Proteobacteria, and understanding how YcaQ functions in this genetic context will be important for understanding any broader roles it may play. Finally, the cellular and functional roles of YQL enzymes in bacteria should be examined for their contribution to pathogenesis and suppression of mutations. Since YQL proteins are abundant in pathogenic microbes, examining if this enzyme family plays any role in the fidelity of the infection process will be crucial to understand in the age of antimicrobial resistance (AMR). Additionally, the investigation of if these proteins play a role in suppression of mutations, or have any impact on the competence of transformation uptake (as many YQL genes are associated with transformation genes) will be an essential avenue to explore.

The AZL and YQL family of DNA glycosylase evolved over a billion years ago, and have been split into subfamilies of highly diverse (AZL) and conserved (YQL) genes with highly varied substrates and apparent functions. The discovery of these proteins over the last five years has merely scratched the surface of their complexity, and I anticipate many great and fascinating breakthroughs from studying these curiously complex enzymes.

DNA Substrates

Table 1. Oligodeoxynucleotides used in these studies

Oligo Name [a]	Sequence (5′→3′) [b]	Use	Length (nts)
PCR Primers and Cloning			
AlkZ_FP_BamHI	TCATCAGGATCCATGAAGGCCAGCTGGAGA	cloning, gDNA→pBG103	30
AlkZ_RP_XhoI	CCGGTGCGCATCGTCTGACTCGAGTGATGA	cloning, gDNA→pBG103	30
YcaQ_FP_BamHI	TCATCAGGATCCATGTCGCTGCCGCACCTC	cloning, gDNA→pBG103	30
YcaQ_RP_SalI	TATACCGTCGACTTATGCGACGGGGTCTAT	cloning, gDNA→pBG103	30
YcaQ_FP_NcoI	TGATGACCATGGATGTCGCTGCCGCACCTC	cloning, gDNA→pSF-OXB11	30
YcaQ_RP_XbaI	TGATGATCTAGATTATGCGACGGGGTCTAT	cloning, gDNA→pSF-OXB11	30
EndoIV_FP_BamHI	TGATGAGGATCCATGAAATACATTGGAGCG	cloning, gDNA→pHD116	30
EndoIV_RP_SalI	TGATGAGTCGACTCAGGCTACCGCTTTTTC	cloning, gDNA→pHD116	30
HedH4_FP_NcoI	TGATGACCATGGATGCAATTGGGATTAACA	cloning, pBG102→pSF-OXB1	30
HedH4_RP_XbaI	TGATGATCTAGATTAACGCCAGCCGTCACG	cloning, pBG102→pSF-OXB1	30
AlkZ_Q37A_FP	CTGTGGCGTGGCGGCACAGGTCTG	mutational primer	24
AlkZ_Q37A_RP	AGTCTGCCGACGACGTCA	mutational primer	18
AlkZ_Q39A_FP	CGTGCAGGCAGCGGTCTGGAGTGTC	mutational primer	25
AlkZ_Q39A_RP	CCACAGAGTCTGCCGACG	mutational primer	18
AlkZ_E45A_FP	GAGTGTCGCCGCGCTGAACGTCG	mutational primer	23
AlkZ_E45A_RP	CAGACCTGTGCCTGCACG	mutational primer	18
AlkZ_E221A_FP	TGCGACGATCGCGACCTTCGACCGC	mutational primer	25
AlkZ_E221A_RP	GGGCCGTACGCGCCCAGG	mutational primer	18
AlkZ_Δ(β11/12)_FP	GGCGGCTCCCCCGTGGTCGTGAAG	mutational primer	24
AlkZ_Δ(β11/12)_RP	CACCGCGGAGCGGTGTTC	mutational primer	18
YcaQ_Q45A_FP	CTTGCTGGCGATCGATACCA	mutational primer	20
YcaQ_Q45A_RP	GACATGCGGGAGATCGTTGC	mutational primer	20
YcaQ_D47A_FP	GCAAATCGCGACCATCAATA	mutational primer	20
YcaQ_D47A_RP	AGCAAGGACATGCGGGAGAT	mutational primer	20
YcaQ_D47N_FP	GCAAATCAACACCATCAATATTG	mutational primer	23
YcaQ_D47N_RP	AGCAAGGACATGCGGGAG	mutational primer	18
HedH4_Q41A_FP	AGCTGGCCTAGCGGCGCAGGACC	mutational primer	23
HedH4_Q41A_RP	AACAAGGTAATAACTTCCAGC	mutational primer	21

Oligo Name [a]	Sequence (5′→3′) [b]	Use	Length (nts)
HedH4_Q43A_FP	CCTACAAGCGGCGGACCCGGAGC	mutational primer	23
HedH4_Q43A_RP	CCAGCTAACAAGGTAATAACTTC	mutational primer	23
LldU1_H43A_FP	GTGCGGCGCGGCGGCGCAGGTTCTG	mutational primer	25
LldU1_H43A_RP	AGCGCACGCACCAGTTCC	mutational primer	18
LldU1_Q45A_FP	CGCGCACGCGGCGGTTCTGAGCGCG	mutational primer	25
LldU1_Q45A_RP	CCGCACAGCGCACGCACC	mutational primer	18
LldU5_Q41A_FP	GGTTGCGATGGCGGGTCAAGAACCG	mutational primer	25
LldU5_Q41A_RP	AGGTGCTCAACCACACGC	mutational primer	18
LldU5_Q43A_FP	GATGCAGGGTGCGGAACCGGATGCG	mutational primer	25
LldU5_Q43A_RP	GCAACCAGGTGCTCAAC	mutational primer	17
TxnU4_H43A_FP	GTGCGGTGCGGCGGCGCAAGTTC	mutational primer	23
TxnU4_H43A_RP	ATCGCACGCACAATTTCC	mutational primer	18
TxnU4_Q45A_FP	TGCGCATGCGGCGGTTCTGAGCGCG	mutational primer	25
TxnU4_Q45A_RP	CCGCACATCGCACGCACA	mutational primer	18
ΔycaQ_Kan_FP	CTTCTGGCAACTAGTTACGCCGCGGCAGCGT TCGATTGATGGAGTCATGAGTGTAGGCTGGA GCTGCTTC	*ycaQ* deletion primer	70
ΔycaQ_Kan_RP	GTTCCTCCAGATGGATCGGGTTATGAATGCAT AAATCTTATCATAATCATCATATGAATATCCTC CTTAG	*ycaQ* deletion primer	70
ΔtxnU2_FP [c]	GGCTGCGGACAAGGGTTCGAGGGGCTCAAC CGTCACCTGATTCCGGGGATCCGTCGACC	*txnU2* deletion primer	59
ΔtxnU2_RP [c]	ACCTACACCAATCCGCTGCGCTACCTGAACG GCGTGCGCTGTAGGCTGGAGCTGCTTC	*txnU2* deletion primer	58
ΔtxnU4_FP [c]	GGCTGCGGACAAGGGTTCGAGGGGCTCAAC CGTCACCTGATTCCGGGGATCCGTCGACC	*txnU4* deletion primer	59
ΔtxnU4_RP [c]	ACCTACACCAATCCGCTGCGCTACCTGAACG GCGTGCGCTGTAGGCTGGAGCTGCTTC	*txnU4* deletion primer	58
ΔlldU1_LFP [d]	AAAAAGCTTGTGTCCTGGCACGAACGAC	*lldU1* deletion primer (left)	28
ΔlldU1_LRP [d]	AAATCTAGAGCCGTGCGTCTTCACCAG	*lldU1* deletion primer (left)	27
ΔlldU1_RFP [d]	AAATCTAGACTCACGAAACTCTCCACGGC	*lldU1* deletion primer (right)	29
ΔlldU1_RRP [d]	AAAGAATTCCGAGCAAGGTCAGCGGAT	*lldU1* deletion primer (right)	27
ΔlldU5_LFP [d]	AAAAAGCTTGGCGTAGCCGACGAAACT	*lldU5* deletion primer (left)	27
ΔlldU5_LRP [d]	AAATCTAGAATGTCGAGGACCGACACG	*lldU5* deletion primer (left)	27
ΔlldU5_RFP [d]	AAATCTAGACTGGGTGGTGCGCAAATG	*lldU5* deletion primer (right)	27
ΔlldU5_RRP [d]	AAAGAATTCGACCGGCTCCGTGATACG	*lldU5* deletion primer (right)	27
Ada_FP	AACGACCCAGCTCACAAT	qPCR primer	18
Ada_RP	GCATGACGGCTAAACAATTCC	qPCR primer	21
AlkA_FP	CCCGGTATTGTCATTGGTAAGG	qPCR primer	22
AlkA_RP	CAGACCCGCAGGCATTAAA	qPCR primer	19
GapA_FP	CGGTACCGTTGAAGTGAAAGA	qPCR primer	21
GapA_RP	ACTTCGTCCCATTTCAGGTTAG	qPCR primer	22
LexA_FP	CTGTTGCAGGAAGAGGAAGAA	qPCR primer	21
LexA_RP	GGAAGGATCGACCTGATAATGAC	qPCR primer	22

Oligo Name [a]	Sequence (5′→3′) [b]	Use	Length (nts)
UvrA_FP	CGCACGGACGATTACTGATAAA	qPCR primer	22
UvrA_RP	CGAAGGCGTGCATAATA	qPCR primer	17
Tag_FP	GAGTCAGGACCCGCTTTATATT	qPCR primer	22
Tag_RP	GGACGGTGATCCACGATAAT	qPCR primer	20
YcaQ_FP	AGGTGATGGTGATTGAACGG	qPCR primer	20
YcaQ_RP	ATGATTTCTGCTTCTGTTTGCG	qPCR primer	22
Base Excision Assays			
7mG_Top [e]	FAM-CACCACTACACC(7mG)ATTCCTTACAAC	base excision	25
7mG_Top [e]	Cy5-CACCACTACACC(7mG)ATTCCTTACAAC	base excision	25
7mG_Bottom	GTTGTAAGGAATCGGTGTAGTGGTG	base excision	25
AZB_Top	FAM-AAAAATAAAA(G)CCAAATAAAAATAAA	base excision	26
AZB_Bottom	Cy5-TTTATTTTTATTT(G)GCTTTTATTTTT	base excision	26
NM_Top	FAM-AAAAATAAAA(G)TCAAATAAAAATAAA	base excision	26
NM-Bottom	Cy5-TTTATTTTTATTT(G)ACTTTTATTTTT	base excision	26
AP_DNA	GTTGTAAGGAUTCGGTGTAGTGGTG	base excision	25
AP_DNA_FAM	FAM-GTTGTAAGGA(U)TCGGTGTAGTGGTG	base excision	25
Hed_Top	Cy5-AATATTAATAAT(G)TAATTTAAATTA	base excision	25
Hed_Bottom	TAATTTAAATTACATTATTAATATT	base excision	25
TXN/LLD_HPLC	AACCG(G)TT	base excision	8
TXN/LLD_Top	Cy5-ATCGAATCCAAT(G)TCTAAGGAATTCT	base excision	26
TXN/LLD_Bottom	AGAATTCCTTAGACATTGGATTCGAT	base excision	26
CLB_Top	Cy5-GGGCGGGCGG(A)CGGGCGGGC	base excision	20
CLB_Bottom	GCCCGCCCGTCCGCCCGCCC	base excision	20
THF_25_Top [f]	FAM-ACGTGGCACTTG(THF)GCAACTCGACTG	DNA anisotropy	25
THF_25_Bottom	CAGTCGAGTTCGACAAGTGCCACGT	DNA anisotropy	25
THF_18_Top [f]	TGAGTCGT(THF)GATGACCAC	YQL^{Tfu} crystallization	18
THF_18_Bottom	GTGGTCATCCACGACTCA	YQL^{Tfu} crystallization	18

[a] Oligodeoxynucleotides were dissolved in TE buffer (10 mM Tris•HCl pH 8.0, 1 mM EDTA) and stored at -20°C (unless noted otherwise in methods). Cy5/FAM oligonucleotides were stored in the dark.

[b] The underlined G/A is the target site of alkylation. FAM, 6-carboxyfluorescein; Cy5, cyanine 5.

[c] The gene interruption mutant strains of *txnU2* and *txnU4* were constructed with REDIRECT technology as previously described (Gust *et al.*, 2003; Zhang *et al.*, 2015).

[d] The in-frame deletion mutants of gene *lldU1* and *lldU5* were constructed by using the temperature-sensitive plasmid pKC1139, and were introduced into the *S. vinaceusdrappus* NRRL 15735 using *E. coli-Streptomyces* intergeneric conjugation.

[e] The blue text in 7mG_Top denotes nucleotides that are incorporated enzymatically

[f] THF: tetrahydrofuran

Table 2. Cellular strains and plasmids used in these studies

Strain/Plasmid Name	Characteristics	Source or Reference
Strains		
Streptomyces sahachiroi	Azinomycins A & B producing strain (wild type)	ATCC 33158
Escherichia coli K-12 BW25113 WT	Wild type *E. coli* strain for cellular studies	Keio Collection; GE
E. coli K-12 Δ*uvrA*	Δ*uvrA E. coli* strain for cellular studies; ::Kan^{Flp}	Keio Collection; GE
E.coli K-12 Δ*ycaQ*	Δ*ycaQ E. coli* strain for cellular studies; ::Kan^{R}	Keio Collection; GE
E. coli K-12 Δ*uvrA*/Δ*ycaQ*	Δ*uvrA*/Δ*ycaQ E. coli* strain for cellular studies; ::Kan^{R}	(Bradley *et al.*, 2020)
E. coli K-12 Δ*tag*/Δ*alkA*	Δ*tag*/Δ*alkA E. coli* strain for cellular studies; ::Kan^{R}	Pat O'Brien lab
E. coli DH5α	*E. coli* strain for cloning and genetic manipulation	NEB
E. coli Tuner (DE3)	*E. coli* strain for protein overexpression; Δ*lacZY*	Vanderbilt CSB
S. bottropensis	Trioxacarcin A producing strain (wild type)	NRRL 12051
Δ*txnU2*	In-frame deletion mutant of *txnU2*	(Chen *et al.*, 2022)
Δ*txnU4*	In-frame deletion mutant of *txnU4*	(Chen *et al.*, 2022)
S. vinaceusdrappus	LL-D49194α1and LL-D49194β2 producing strain (wild type)	NRRL 15735
Δ*lldU1*	In-frame deletion mutant of *lldU1*	(Chen *et al.*, 2022)
Δ*lldU5*	In-frame deletion mutant of *lldU5*	(Chen *et al.*, 2022)
S. lividans 1326	Trioxacarcin A, LL-D49194α1and LL-D49194β2 sensitive strain, used for the heterologous expression of the genes *txnU2*, *txnU4*, *lldU1* and *lldU5*	NRRL 16148
S. lividans::pSET152	Heterologous expression strain with empty vector, served as negative control strain	(Chen *et al.*, 2022)
E. coli BL21 (DE3)	Host strain for protein overexpression	Invitrogen
Plasmids		
pBG102	Protein expression vector with N-terminal His_6-SUMO tag; Kan^{R}	Vanderbilt Center for Structural Biology (CSB)
pBG103	Protein expression vector with N-terminal His_6-SUMO tag; Kan^{R}	Vanderbilt CSB
pET28a	Protein expression vector with N-terminal His_6 tag; Kan^{R}	Novagen
pHD116	Protein expression vector with N-terminal His_6 tag; Amp^{R}	Vanderbilt CSB
pCP20	Temperature-sensitive plasmid expressing FLP recombinase; Amp^{R}	Addgene
pKM208	Temperature-sensitive plasmid expressing λ-Red (Beta, Exo, Gam) recombineering system; Kan^{R}	Addgene
pSF-OXB1	Weak strength bacterial vector; constitutive expression from modified *araBAD (ΔaraC)* promoter system; Kan^{R}	Millipore-Sigma
pSF-OXB11	Intermediate strength vector; constitutive expression from modified *recA (ΔlexA)* promoter system; Kan^{R}	Millipore-Sigma
pTG5001	Fosmid vector containing TXN gene cluster	(Chen *et al.*, 2022)
pIJ773	Vector containing the apramycin resistance gene *aac(3)IV* and *oriT*; Apr^{R}	(Chen *et al.*, 2022)
pKC1139	Streptomyces temperature-sensitive plasmid for gene deletion; Apr^{R}	(Chen *et al.*, 2022)

Table 3. X-ray data collection and refinement statistics for *S. sahachiroi* AlkZ

Data collection	
Space group	$P6_5$
Cell dimensions	
a, *b*, *c* (Å)	78.92, 78.92, 139.40
α, *β*, *γ* (°)	90.00, 90.00, 120.00
Resolution (Å)	50.00–2.30 (2.38–2.30) [a]
R_{sym}	0.118 (0.481)
Avg. $I/\sigma I$	29.1 (7.0)
Completeness (%)	96.2 (100.0)
Redundancy	12.3 (12.6)
Wilson *B*-factor (Å^2)	21.8
Refinement	
Resolution (Å)	39.47–2.30 (2.42–2.30)
No. reflections	20,828 (3,014)
R_{work}	0.157 (0.170)
R_{free} [b]	0.201 (0.220)
No. atoms [c]	
Protein	2,882
Water	245
Avg. *B*-factors [c, d] (Å^2)	
Protein	25.0
Water	27.6
Refinement	
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.769
Ramachandran distribution (%)	
Favored	98.1
Allowed	1.9
Disallowed	0.0

[a] Statistics for the highest resolution shell are shown in parentheses.

[b] R_{free} was determined from the 5% of reflections excluded from refinement.

[c] Riding hydrogens were not included in no. atoms or avg. *B*-factors.

[d] Equivalent isotropic *B*-factors were calculated in conjunction with TLS-derived anisotropic *B*-factors.

Table 4. AlkZ homologs found in characterized biosynthetic gene clusters (BGCs)

Organism	Protein ID	% I/S to AlkZ	Distance to BGC (kbps/ORFs)	Cluster (% similarity)	Gene Name	Function
S. armeniacus	AXK36599.1	25/37	0/0	Armeniaspirol (100%)	*orf2*	Unknown
S. bottropensis	AKT74276.1	30/46	0/0	Trioxacarcin (100%)	*txnU 2*	DNA alkylating agent
S. bottropensis	AKT74302.1	31/45	0/0	Trioxacarcin (100%)	*txnU 4*	DNA alkylating agent
S. galilaeus	QEU68884.1	45/58	0.5/1	Aclacinomycin/Aklavinone (100%)	*orf1* [a]	Topoisomerase inhibitor
S. griseoruber	AAP85352.1	22/37	0/0	Hedamycin (100%)	*hedH4*	DNA alkylating agent
S. nodosus	ADI58656.1	24/36	0/0	Asukamycin (100%)	*asuH 1*	Farsenyltransferase inhibitor/UBR7 alkylator
S. sahachiroi	ABY83174.1	100/100	0/0	Azinomycin B (100%)	*alkZ*	DNA ICL agent
S. sp. NRRL F-4474	WP_030845410.1	29/43	0/0	Tambromycin (100%)	*tbrK*	Unknown
S. sp. Sp080513GE-23	BAP16700.1	30/42	0/0	JBIR-34/35 (100%)	*fmoL*	Unknown
S. vinaceusdrappus	QDQ37873.1	32/46	0/0	LL-D49194α1 (100%)	*lldU1*	DNA alkylating agent
S. vinaceusdrappus	QDQ37896.1	25/39	0/0	LL-D49194α1 (100%)	*lldU5*	DNA alkylating agent

[a] orf1 is located at the extreme 3′ terminus of gene cluster sequence (partial sequence of N-terminus)

Table 5. AlkZ homologs found in uncharacterized biosynthetic gene clusters (BGCs)

Organism	Protein ID	% I/S to AlkZ	Distance to BGC (kbps/ORFs)	Cluster BLAST (% similarity) [a]
S. abyssomicinicus	WP_164540990.1	52/64	-0.6/1	No BGC identified
S. alboflavus	ARX89312.1	33/45	-0.6/1	Toyocamycin (40%)
S. agglomeratus	WP_069933964.1	26/39	0/0	Meilingmycin (3%)
S. angustmyceticus	WP_086720958.1	25/37	0/0	LL-D49194α1 (35%)
S. aurantiacus	WP_037658687.1	26/39	0/0	Oxazolomycin (9%)
S. autolyticus	WP_079256125.1	28/39	0.1/1	Meilingmycin (3%)
S. avicenniae	WP_052850112.1	30/40	2.1/2	Ficellomycin (10%)
S. barkulensis	WP_101258522.1	28/42	0/0	Fluostatin (27%)
S. bobili	WP_086767611.1	27/37	0/0	LL-D49194α1 (71%)
S. caeruleatus	WP_062716690.1	26/40	0/0	Daptomycin (3%)
S. candidus	WP_185031354.1	32/48	0/0	LL-D49194α1 (39%)
S. candidus	WP_185031360.1	25/38	0/0	LL-D49194α1 (39%)
S. caniferus	WP_159474002.1	26/38	19.2/5	LL-D49194α1 (35%)
S. caniferus	WP_159474018.1	32/46	13.1/5	LL-D49194α1 (35%)
S. caniferus	WP_159475850.1	25/34	0/0	LL-D49194α1 (35%)
S. chartreusis	WP_150499220.1	26/39	0/0	Herboxidiene (7%)
S. chilikensis	WP_166020931.1	54/65	0/0	Antimycin (26%)
S. cinereoruber	WP_152370881.1	50/64	-10.1/4	Ficellomycin (3%)
S. corynorhini	WP_114626433.1	26/41	0/0	Hedamycin (6%)
S. davaonensis	WP_015656894.1	26/38	0/0	Stigmatellin (10%)
S. davawensis	CCK26500.1	25/36	0/0	Jerangolid A (9%)
S. decoyicus	WP_107083370.1	26/40	0/0	LL-D49194α1 (43%)
S. diastatochromogenes	WP_094219861.1	24/37	0/0	Asukamycin (6%)
S. flavidovirens	WP_051338381.1	93/96	0/0	Azinomycin B (78%)
S. filamentosus	KAA6211572.1	50/63	0/0	Azinomycin B (51%)
S. galilaeus	WP_150473384.1	26/37	0/0	Cinerubin B (100%)
S. graminilatus	WP_055524427.1	28/43	0/0	Clavams (12%)
S. griseoviridis	WP_127177155.1	25/39	0/0	LL-D49194α1 (34%)
S. inusitatus	GGZ61434.1	26/38	0/0	AmfS (100%)
S. janthinus	GGS78276.1	27/41	0/0	No BGC identified
S. jeddahensis	WP_067285245.1	25/37	0/0	Asukamycin (3%)
S. kebangsaanensis	WP_073947318.1	49/63	0/0	Azinomycin B (46%)
S. laurentii	BAU84011.1	25/40	0/0	Hedamycin (34%)
S. lavendulae	WP_030237597.1	31/42	0/0	Valinomycin (11%)
S. leeuwenhoekii	WP_078648136.1	27/40	0/0	Leinamycin (18%)
S. lunaelactis	WP_108153375.1	25/40	0/0	ECO-02301 (46%)
S. lunaelactis	WP_108154250.1	26/40	0/0	No BGC identified
S. lushanensis	WP_066947886.1	31/46	15.9/4	Hedamycin (34%)
S. malaysiense	WP_167377699.1	31/42	0/0	JBIR-34/35 (70%)
S. netropsis	WP_165451496.1	37/52	0/0	C-1027 (6%)
S. netropsis	WP_130879439.1	37/52	-0.5/1	Guadinomine (7%)
S. netropsis	WP_130878334.1	27/39	0/0	Vazabitide A (28%)
S. nigra	WP_108710889.1	26/39	0/0	Grincamycin (5%)
S. nojiriensis	GGR83661.1	30/43	0/0	Tambromycin (96%)
S. olivaceus	CP016795.1	60/73	0/0	Azinomycin B (55%)

Organism	Protein ID	% I/S to AlkZ	Distance to BGC (kbps/ORFs)	Cluster BLAST (% similarity)
S. pini	WP_093852302.1	27/42	0/0	LL-D49194α1 (30%)
S. populi	WP_103550573.1	28/42	0/0	Enduracidin (6%)
S. puniciscabiei	WP_159045481.1	30/45	0/0	JBIR-34/35 (79%)
S. radiopugnans	WP_093657592.1	24/38	0/0	LL-D49194α1 (43%)
S. rimosus	QGY71233.1	51/64	-0.2/1	Daptomycin (22%)
S. roseochromogenus	WP_023549546.1	25/37	0/0	Hedamycin (28%)
S. rubidus	WP_075018101.1	27/40	0/0	Hopene (69%)
S. scabrispora	WP_078980627.1	25/37	0/0	Daptomycin (4%)
S. scabrispora	WP_078974087.1	26/39	0/0	No BGC identified
S. silvensis	WP_079086777.1	31/44	0/0	Skyllamycin (10%)
S. subrutilus	WP_069918639.1	31/42	0/0	SapB (100%)
S. subrutilus	WP_069923306.1	53/68	0/0	Stenothricin (13%)
S. tateyamensis	WP_110670119.1	27/40	0/0	Salinomycin (28%)
S. toyocaensis	WP_037926092.1	27/41	0.4/1	Diazepinomicin (60%)
S. triticagri	WP_128559347.1	27/42	0/0	SCO-2138 (35%)
S. uncialis	WP_073792593.1	26/40	-12.3/5	Vazabitide A (4%)
S. violarus	WP_184587254.1	26/41	0/0	Hygrocin A (6%)
S. violascens	GGU18793.1	26/39	-1.6/1	A201A (6%)
S. violens	WP_051831071.1	76/85	0/0	Azinomycin B (76%)
S. viridosporus	WP_004992266.1	26/37	0/0	SapB (75%)
S. wuyuanensis	WP_093654700.1	28/41	0/0	Tambromycin (92%)
S. xinghaiensis	WP_019711508.1	25/45	0/0	LL-D49194α1 (37%)
S. zhaozhouensis	WP_097230763.1	28/41	0/0	Daptomycin (9%)

[a] *Cluster BLAST* (column E) is the most similar BGC as determined by cluster BLAST analysis (*% similarity* is the percentage of genes in uncharacterized BGC that have homology to genes in the known similar BGC).

Reagents

Reagents were purchased or obtained from the following suppliers or labs, and were of the highest purity available. Oligonucleotides were purchased from Integrated DNA Technologies (IDT) (Table 1). Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich, and all enzymes were purchased from New England Biolabs (NEB). *E. coli* AlkA was generously provided by Pat O'Brien (University of Michigan). Human AAGΔ83, *Schizosaccharomyces pombe* Mag1, and *Bacillus cereus* AlkC/AlkD were purified as previously described (O'Brien and Ellenberger, 2004; Rubinson *et al.*, 2008; Adhikary and Eichman, 2011; Shi *et al.*, 2018). Hedamycin was obtained from the National Cancer Institute's Developmental Therapeutic Program (NCI DTP) Open Compound Repository (NSC 70929).

Protein Expression and Purifications

Streptomyces sahachiroi AlkZ wild-type and mutants

The *azi36* gene was amplified by PCR from genomic DNA isolated from *S. sahachiroi* and cloned into pBG102. His_6-SUMO-AlkZ was overexpressed in *E. coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 100 µg/mL kanamycin and 50 µM IPTG. Cells were lysed in buffer A with gentle sonication and cell debris removed by centrifugation at 45,000 ×g and 4°C for 30 min. Clarified lysate was passed over 5 mL of Ni-NTA agarose resin equilibrated in buffer A (50 mM Tris•HCl pH 8.5, 500 mM NaCl, and 10% glycerol). The column was washed with 100 mL of 20 mM imidazole/buffer A and protein eluted in 50 mL of 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA and 1 mM TCEP prior to incubation with ~0.5 mg

of Rhinovirus 3C (PreScission) protease at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B (50 mM Tris•HCl pH 8.5, 10% (v/v) glycerol, 1 mM TCEP, and 0.1 mM EDTA). Diluted protein was loaded onto a 5-mL heparin Sepharose column, washed with 25 mL of 50 mM NaCl/buffer B, and eluted with 50 mL of a 50-1000 mM NaCl/buffer B linear gradient. Fractions were pooled and repassed over Ni-NTA agarose resin equilibrated in buffer A. Flow-through and protein that eluted with 50 mL of 20 mM imidazole/buffer A were pooled, concentrated and filtered, and passed over Superdex 200 size exclusion resin equilibrated in buffer C (20 mM Tris•HCl pH 8.5, 100 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, and 0.1 mM EDTA). Protein was eluted with 1.5 CV of buffer C, concentrated to 6 mg/mL, flash-frozen in liquid nitrogen, and stored at −80°C.

AlkZ mutants were generated using the Q5 mutagenesis kit (New England BioLabs). The Δ(β11/12) mutant was generated by replacing residues 307-SRAAGWI-313 with two glycine residues. Mutant proteins were overexpressed and purified the same as wild-type.

Selenomethionyl-substituted AlkZ (SeMet-AlkZ) was overexpressed and purified the same as the wild-type protein, except that protein was overexpressed in M9 medium supplemented with 0.4% (w/v) dextrose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mg/L thiamine, 70 mg/L selenomethionine, 50 mg/L leucine, 50 mg/L isoleucine, 50 mg/L valine, 100 mg/L phenylalanine, 100 mg/L lysine, 100 mg/L threonine, 100 µg/mL kanamycin, and 50 µM IPTG.

Escherichia coli YcaQ wild-type and mutants

The *ycaQ* gene was amplified by PCR from genomic DNA isolated from *E. coli* K-12 MG1655 cells and cloned into pBG103. His_6-SUMO-YcaQ was overexpressed in *E. coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 30 μg/mL kanamycin and 50 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were lysed in buffer A with sonication and cell debris removed by centrifugation at 45,000 × g at 4°C for 30 min. Clarified lysate was passed over Ni-NTA agarose equilibrated in buffer A [50 mM Tris•HCl pH 8.0, 500 mM NaCl, 20 mM imidazole and 10% (vol/vol) glycerol] and protein eluted in 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA and 1 mM TCEP before incubation with ~0.5 mg of Rhinovirus 3C protease (PreScission) at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B [50 mM Tris•HCl pH 8.0, 10% (vol/vol) glycerol, 1 mM TCEP, and 0.1 mM EDTA] and purified by heparin sepharose using a 0–1 M NaCl/buffer B linear gradient. Fractions were pooled and repassed over Ni-NTA agarose in buffer A, concentrated and filtered, and passed over Superdex 200 size exclusion resin equilibrated in buffer C [20 mM Tris•HCl pH 8.0, 100 mM NaCl, 5% (vol/vol) glycerol, 1 mM TCEP, and 0.1 mM EDTA]. Protein was concentrated to 10 μM, flash frozen in liquid nitrogen, and stored at −80°C. YcaQ mutants were generated using the Q5 Mutagenesis Kit (New England BioLabs). Mutant proteins were overexpressed and purified the same as wild-type.

Streptomyces griseoruber HedH4 wild-type and mutants

Streptomyces griseoruber hedH4 was codon optimized and synthesized by GenScript and cloned into pBG102. The N-terminal His_6-SUMO fusion protein was

overexpressed in *Escherichia coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 30 μg/mL kanamycin and 50 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were lysed by sonication in buffer A supplemented with 1 mM PMSF and 1 mM TCEP and cell debris was removed by centrifugation at 45,000 × g at 4°C for 30 min. Clarified lysate was passed over Ni-NTA agarose equilibrated in buffer A (50 mM Tris•HCl pH 8.5, 500 mM NaCl, 25 mM imidazole, and 10% (vol/vol) glycerol) and protein eluted in 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA and 1 mM tris(2-carboxyethyl)phosphine (TCEP) before incubation with 0.5 mg Rhinovirus 3C protease (PreScission) at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B (50 mM Tris•HCl pH 8.5, 10% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA) and purified by heparin sepharose using a 0–1 M NaCl/buffer B linear gradient. Fractions were pooled and passed over Ni-NTA agarose in buffer A, concentrated and filtered, and buffer exchanged into buffer C (20 mM Tris•HCl pH 8.5, 100 mM NaCl, 5% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA). Protein was concentrated to 4 mg/mL, flash-frozen in liquid nitrogen, and stored at −80°C. Mutant protein expression vectors were generated using the Q5 Mutagenesis Kit (New England BioLabs) and proteins overexpressed and purified the same as wild-type.

Streptomyces caeruleatus AZL2 wild-type

Streptomyces caeruleatus AZL2 was codon optimized and synthesized by GenScript and cloned into pBG102. The protein expression, purification, and storage of *S. caeruleatus* AZL2 was the same as *S. griseoruber* HedH4.

Streptomyces bottropensis TxnU2 and TxnU4 wild-type and mutants

The *txnU2 and txnU4* genes were codon optimized and synthesized by GenScript and cloned into pBG102. N-terminal His_6-SUMO proteins were overexpressed in *E. coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 30 μg/mL kanamycin and 50 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were lysed in buffer A with sonication and cell debris removed by centrifugation at 45,000 × g at 4°C for 30 min. Clarified lysate was passed over Ni-NTA agarose equilibrated in buffer A (50 mM Tris•HCl pH 8.5, 500 mM NaCl, 20 mM imidazole, and 10% (vol/vol) glycerol) and protein eluted in 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 1 mM dithiothreitol (DTT) before incubation with 0.5 mg of Rhinovirus 3C (PreScission) protease and 0.5 mg of yeast ubiquitin-like-specific protease 1 (Ulp1) at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B (50 mM Tris•HCl pH 8.5, 10% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA) and purified by heparin sepharose using a 0–1 M NaCl/buffer B linear gradient. Fractions were pooled and repassed over Ni-NTA agarose in buffer A, concentrated and filtered, and buffer exchanged into buffer C (20 mM Tris•HCl pH 8.5, 100 mM NaCl, 5% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA). Proteins were concentrated to 100 μM, flash-frozen in liquid nitrogen, and stored at −80°C. For purification of TxnU2, buffers A and B were supplemented with 0.02% NP-40 (v/v) and buffer C was supplemented with 0.01% NP-40. Proteins used in HPLC analysis did not contain NP-40. *TxnU2/4* mutants were generated using the Q5 Mutagenesis Kit (New England BioLabs). Mutant proteins were overexpressed and purified the same as WT.

Streptomyces vinaceusdrappus LldU1 and LldU5 wild-type and mutants

The *lldU1 and lldU5* genes were synthesized by GenScript and cloned into pBG102. The protein expression, purification, and storage of AlkZ2 was the same as *S. bottropensis* TxnU2 and TxnU4. For purification of LldU5, buffers A and B were supplemented with 0.02% NP-40 (v/v) and buffer C was supplemented with 0.01% NP-40 (v/v). Proteins used in HPLC analysis did not contain NP-40. *LldU1/5* mutants were generated using the Q5 Mutagenesis Kit (New England BioLabs). Mutant proteins were overexpressed and purified the same as WT.

Thermobifida fusca and Thermomonospora curvata YQL wild-type

Thermobifida fusca and Thermomonospora curvata YQL wild type were codon optimized and synthesized by GenScript and cloned into pBG102. His_6-SUMO-$YQL^{Tfu/Tcu}$ were overexpressed in *E. coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 30 μg/mL kanamycin and 50 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were lysed with sonication in buffer A supplemented with 1 mM PMSF and 1 mM TCEP and cell debris was removed by centrifugation at 45,000 × g at 4°C for 30 min. Clarified lysate was passed over Ni-NTA agarose equilibrated in buffer A [50 mM HEPES•NaOH pH 8.0, 500 mM NaCl, 25 mM imidazole and 10% (vol/vol) glycerol] and protein eluted in 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA and 1 mM tris(2-carboxyethyl)phosphine (TCEP) before incubation with ~0.5 mg of Rhinovirus 3C protease (PreScission) at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B

[50 mM HEPES•NaOH 8.0, 10% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA] and purified by heparin sepharose using a 0–1 M NaCl/buffer B linear gradient. Fractions were concentrated and passed over Superdex 200 size exclusion resin equilibrated in buffer C [25 mM HEPES•NaOH pH 8.0, 200 mM NaCl, 5% (vol/vol) glycerol, 1 mM TCEP, and 0.1 mM EDTA]. Protein was concentrated to 20 μM, flash frozen in liquid nitrogen, and stored at −80°C.

Spizellomyces punctatus ΔN-term

Spizellomyces punctatus a*lkZ* wild type was codon optimized and synthesized by GenScript and cloned into pBG102. Amino acids F2-I47 were deleted using the Q5 Mutagenesis Kit to generate ΔN-AlkZ^{Spu}. His$_6$-SUMO-ΔN-AlkZ^{Spu} was overexpressed in *E. coli* Tuner (DE3) cells at 16°C for 18 h in LB medium supplemented with 30 μg/mL kanamycin and 50 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were lysed with sonication and cell debris removed by centrifugation at 45,000 × g at 4°C for 30 min. Clarified lysate was passed over Ni-NTA agarose equilibrated in buffer A [50 mM HEPES•NaOH pH 8.0, 500 mM NaCl, 25 mM imidazole and 10% (vol/vol) glycerol] and protein eluted in 250 mM imidazole/buffer A. Protein fractions were pooled and supplemented with 0.1 mM EDTA and 1 mM tris(2-carboxyethyl)phosphine (TCEP) before incubation with ~0.5 mg of Rhinovirus 3C protease (PreScission) at 4°C overnight. Cleaved protein was diluted 10-fold in buffer B [50 mM HEPES•NaOH 8.0, 10% (vol/vol) glycerol, 0.1 mM TCEP, and 0.1 mM EDTA] and purified by heparin sepharose using a 0–1 M NaCl/buffer B linear gradient. Fractions were pooled and repassed over Ni-NTA agarose in buffer A, concentrated and filtered, and passed over Superdex 200 size

exclusion resin equilibrated in buffer C [25 mM HEPES•NaOH pH 8.0, 200 mM NaCl, 5% (vol/vol) glycerol, 1 mM TCEP, and 0.1 mM EDTA]. Protein was concentrated to 100 μM, flash frozen in liquid nitrogen, and stored at −80°C.

Base Excision Assays

N7-deoxymethylguanosine (d7mG) DNA preparation and assays

DNA substrates containing a single d7mG lesion and a 5'-6-carboxyfluorescein (FAM) or 5'-Cyanine (Cy5) label within the sequence (FAM/Cy5)-d(CACCACTACACC(7mG)ATTCCTTACAAC)/d(GTTGTAAGGAATCGGTGTAGTGGTG) (Table 1) were prepared as described previously (Mullins *et al.*, 2013). The *N*5-substitutied formidopyrimidine (*N*5-FaPy) derivative of 7mG-DNA was generated by treating 500 nM DNA with 100 mM NaOH at 37°C for 30 min, followed by re-adjusting to pH 7 with 100 mM HCl. The substrate to test excision of d7mG in proximity to an AP site was prepared by incubating unlabeled DNA with a sequence d(GTTGTAAGGA(U)TCGGTGTAGTGGTC) (Table 1) complementary to the FAM-labeled d7mG oligo with 1 U of UDG at 37°C for 30 min (SI Fig. S5B in (Bradley *et al.*, 2020). AP-DNA was phenol-chloroform extracted and annealed to 12-fold molar excess of the complementary d7mG-DNA.

In each glycosylase reaction, 1 μM enzyme was incubated with 100 nM (Cy5/FAM)-DNA in glycosylase buffer [50 mM HEPES pH 8.5, 100 mM KCl, 10 mM EDTA, and 10% (vol/vol) glycerol] at 25°C. At various time points, 4-μL aliquots were added to 1 μL of either 1M NaOH and heated at 70°C for 2 min or 83 nM EndoIV and incubated at 37°C for 5 min. To test for AP lyase activity, the 4-μL aliquots were added to

1 µL glycosylase buffer and incubated at 70°C for 2 min. Samples were then denatured at 70°C for 5 min in 5 mM EDTA pH 8.0, 80% (wt/vol) formamide, 1 mg/mL blue dextran, and electrophoresed on a 20% (wt/vol) acrylamide/8 M urea sequencing gel at 40 watts for 1.5 hrs in 0.5 × TBE buffer (45 mM Tris, 45 mM borate, and 1 mM EDTA pH 8.0). Gels were imaged on a Typhoon Trio variable mode imager (GE Healthcare) using 488-nm excitation and 526-nm emission fluorescence, and bands were quantified with ImageQuant (GE Healthcare). All excision assays were performed in triplicate.

Azinomycin B (AZB) ICL-DNA preparation and assays

Azinomycin B was obtained using the fermentation protocol established previously (Kelly *et al.*, 2008). Briefly, *Streptomyces sahachiroi* cultures were centrifuged at 7000 rpm at 4°C, cell pellets discarded, and the medium extracted with an equal volume of dichloromethane. The organic layer was collected, dried over anhydrous magnesium sulfate, and concentrated *in vacuo*. The resulting crude extract was stored at -80°C and supplemented with 10% methanol prior to use. The presence of the AZB moiety was verified by HPLC-MS. AZB-ICLs were generated by annealing DNA with the sequence FAM-d(AAAAATAAAA(G)CCAAATAAAAATAAA) to the complementary oligo containing a 5'-Cyanine (Cy5) label (Table 1). DNA (100 µM duplex) was incubated with 10 mg crude extract at 4°C on ice for 24 hrs in the dark. The DNA was filtered through a 0.22 µm filter to remove debris from the extract and desalted using Illustra MicroSpin G25 columns (GE Healthcare) equilibrated in TE buffer (10 mM Tris•HCl pH 8.0, 1 mM EDTA). AZB-DNA was purified by denaturing PAGE. Briefly, DNA was denatured at 55°C for 2 min in 5 mM EDTA pH 8.0, 80% (wt/vol) formamide, 0.5 mg/mL orange G and electrophoresed on a

15% (wt/vol) acrylamide/8 M urea denaturing gel pre-run in 0.5 × TBE buffer. FAM (488 nm excitation, 526 nm emission) and Cy5 (633 nm excitation, 670 nm emission) fluorescence were detected from the gel using a Typhoon Trio variable mode imager (GE Healthcare). The band with the slowest electrophoretic mobility that contained both the FAM and Cy5 fluorescence was cut out and placed in 3,500 MWCO dialysis tubing (Thermo Fischer Scientific) with 2 mL of 0.5 × TBE, and the DNA electroeluted from the gel at 120 volts for 1 hr. The ICL-DNA were concentrated to 4 μM using an Amicon Ultra-3K filter (3,500 rpm, 4°C, 45 min), buffer exchanged into TE buffer, aliquoted, and stored at -80°C.

Base excision unhooking of AZB-ICLs used a modified version of the d7mG protocol. Positive controls for ICLs involved heating the lesions to 95°C for 5 min to fully depurinate the DNA, followed by work-up with either water or NaOH to nick the backbone. To perform β/δ-elimination of AP-DNA products, samples were incubated with either 0.2 M NaOH at 55°C for 2 min or 17 nM EndoIV at 37°C for 5 min. Samples were denatured at 55°C in loading buffer for 5 min prior to electrophoresis. Gels were imaged for both FAM and Cy5 fluorescence and quantified individually. ICL unhooking was quantification by summing the individual FAM and Cy5 fluorescence intensities. The raw gels were artificially colored using Adobe Photoshop and overlaid using ImageJ software. Native PAGE analysis of AZB-ICL excision products were carried out by suspending reaction mixtures in 2× native loading buffer [30% (vol/vol) glycerol, 4 mg/mL blue dextran, 400 mM NaCl, 100 mM Tris•HCl (pH 8.0), and 10 mM EDTA] and running on an 15% (wt/vol) acrylamide sequencing gel at 5 W for 3 hrs. Double-stranded DNA standards were annealed in SSC buffer (300 mM NaCl and 30 mM trisodium citrate pH 7.0).

Nitrogen mustard (NM) ICL-DNA preparation and assays

NM-ICL DNA substrates were generated and purified similar to an established protocol (Castano *et al.*, 2017). DNA with the sequence FAM-d(AAAAATAAAA(G)TCAAATAAAAATAAA) (Table 1) was annealed to the complementary Cy5-oligodeoxynucleotide at 100 μM duplex in 40 mM sodium cacodylate (pH 7.0). Mechlorethamine was purchase from Sigma Aldrich; NM_8 compound was synthesized and purified by the Vanderbilt Molecular Design and Synthesis Center. Crosslinks were generated by incubating the DNA with 300 μM mechlorethamine•HCl or NM_8 compound for 3 hrs at 37°C while shaking in the dark. Unreacted drug was removed using a G25 spin column equilibrated in TE buffer, and NM-ICL-DNA gel purified the same as AZB-DNA. The *N*5-substitutied formidopyrimidine (*N*5-FaPy) derivative of NM_5-ICL-DNA was generated the same as for d7mG.

Base excision unhooking of NM-ICLs were conducted the same as AZB-ICLs. Native PAGE analysis of NM-ICL excision products were the same as AZB-ICLs.

Hedamycin (HED) DNA preparation

DNA substrates containing a single hedamycin-guanosine adduct were prepared by annealing DNA with the sequence 5'-Cy5-d(AATATTAATAAT(G)TAATTTAAATTA) to the complementary unlabeled oligo. Hedamycin was dissolved in DMSO to a concentration of 5 mM. 100 μM DNA was incubated with 200 μM hedamycin in 10% methanol and 20% DMSO at 4°C on ice in the dark for 24 hr. Unreacted drug was

removed using an Illustra G-25 spin column (GE Healthcare) equilibrated in TE buffer (10 mM Tris•HCl pH 8.0, 1 mM EDTA pH 8.0), and the DNA was stored at -80°C.

Denaturing PAGE analysis of HedH4 HED-DNA excision

In each glycosylase reaction, 1 μM enzyme was incubated with 50 nM DNA in glycosylase buffer (50 mM HEPES pH 8.5, 100 mM KCl, 1 mM EDTA, and 10% (vol/vol) glycerol) at 25°C. At various time points, 4-μL aliquots were added to 1 μL of 1M NaOH and heated at 70°C for 2 min. Samples were denatured at 70°C for 5 min in 5 mM EDTA pH 8.0, 80% (wt/vol) formamide, and 1 mg/mL blue dextran prior to electrophoresis on a 20% (wt/vol) acrylamide/8 M urea sequencing gel at 40 watts for 1 hr in 0.5 × TBE buffer (45 mM Tris, 45 mM borate, and 1 mM EDTA pH 8.0). Gels were imaged on a Typhoon Trio variable mode imager (GE Healthcare) using 633-nm excitation/670-nm emission fluorescence for Cy5, and bands were quantified with ImageQuant (GE Healthcare). All excision assays were performed in triplicate. Unreacted HED-DNA was not included in quantifications.

HPLC-MS analysis of hedamycin and hedamycin-guanine

HPLC was performed on an Agilent Series 1100 system equipped with an analytical SymmetryShield RP-C18 column (3.5 μm, 4.6 mm × 7.5 mm, 100 Å pore size) and using a linear gradient from 90% buffer A (10 mM ammonium formate) / 10% buffer B (100% methanol) to 100% B over 40 min and a flow rate of 0.4 mL/min. Hedamycin was diluted to 50 μM in 10% methanol and stored on ice prior to HPLC injection. To analyze the product of HedH4 activity, Hed-DNA was diluted to 10 μM in glycosylase

buffer and reacted with 50 μM HedH4 for 1 hr at room temperature before injection. Mass spectrometry was performed with an LTQ Orbitrap XL Hybrid FT Mass Spectrometer (Thermo Fisher Scientific) in positive ion mode from 300-1000 m/z.

Trioxacarcin A (TXNA) and LL-D49149α1 (LLD) DNA preparation

The TXNA- and LLD-DNA substrates for HPLC analysis, which contained two lesions per duplex, were prepared by annealing the 8-bp self-complementary strand 5′-AACCG(G)TT-3′ (Pfoh *et al.*, 2008), followed by incubation of 50 μM DNA with 100 μM TXNA or LLD in PBS buffer (pH 7.0) at 16°C for 2 hr. TXNA- and LLD-DNA substrates used in gel-based assays contained a single TXNA-G or LLD-G adduct and a 5'-Cyanine 5 (Cy5) label, and were prepared by annealing the strand containing the TXNA/LLD target sequence to the complementary unlabeled oligo (Table 1), followed by incubation of 100 μM DNA with 200 μM TXNA or LLD in 10% methanol and 20% DMSO at 4°C on ice in the dark for 36 hr. Unreacted drug was removed using a G-25 spin column equilibrated in TE buffer (pH 8.0), and the DNA was stored at -80°C.

HPLC analysis of TXNA-G (gutingimycin) and LLD-D excision

A 50 μL reaction containing 50 μM TXNA- or LLD-DNA, 20 μM protein, and buffer (100 mM Na_2HPO_4, 100mM NaH_2PO_4, 500 mM NaCl, pH 7.0) was incubated at 16°C for 2 hr. The reaction mixtures were quenched with 30 μL methanol and analyzed by LC-MS at 400 nm absorbance. TXNA-Gua (gutingimycin), $[M+H]^+$ ion with *m/z* 1028.53; LLD-G, $[M+H]^+$ ion with *m/z* 1102.43.

Denaturing PAGE analysis of LldU1/5 and TxnU2/4 LLD/TXNA excision

Glycosylase reactions were performed with 50 nM DNA in glycosylase buffer (50 mM HEPES pH 8.5, 100 mM KCl, 1 mM EDTA, and 10% (vol/vol) glycerol) at 25°C. Single-timepoint reactions shown in (Figs. 56B-C, E, 57B-D, 59A-B, 60A, D) were performed with 1 μM enzyme for either 30 sec, 30 min, or 96 hr, as indicated in each figure legend. Single- and multiple-turnover kinetics reactions shown in (Fig. 56D) were performed with 50 nM (single turnover) or 5 nM (multiple turnover) TxnU4 and 50 nM Cy5-labeled TXNA-DNA. Thermal depurination controls shown in (Fig. 58A) were conducted at 95°C for 5 min. Enzyme and mock reactions involving TXNA, LLD, and 7mG monoadducts were quenched by adding 1 μL of 1 M NaOH to a 4-μL reaction aliquot and heating at 70°C for 2 min. Samples were denatured by addition of 5 μL loading buffer containing 5 mM EDTA pH 8.0, 80% (wt/vol) formamide, and 1 mg/mL blue dextran, and incubating at 70°C for 5 min. Samples were electrophoresed on a 20% (wt/vol) acrylamide/8 M urea sequencing gel at 40 W for 1.5 hr in 0.5× TBE buffer (45 mM Tris, 45 mM borate, and 1 mM EDTA pH 8.0). Gels were imaged on a Typhoon Trio variable mode imager (GE Healthcare) for Cy5 fluorescence (633 nm excitation, 670 nm emission), and bands were quantified with ImageQuant (GE Healthcare). Percent product was calculated as the percent of both β- and δ-elimination bands divided by the total intensity of substrate and β/δ-elimination bands. Unreacted DNA in LLD-DNA reactions was not included in the calculation of percent product. NM-ICLs reactions were performed the same as monoadducts, but were quenched and denatured at 55°C prior to electrophoresis. Gels were imaged for both FAM (488 nm excitation, 526 nm emission) and Cy5 fluorescence and artificially colored (FAM, green; Cy5, red) using Adobe

Photoshop and overlaid using ImageJ software as previously described (Bradley *et al.*, 2020). All excision assays were performed in triplicate. Unreacted LLD-DNA was not included in the quantifications.

Spontaneous depurination of dG, d7mG, and dTXNA-G from DNA

Non-enzymatic depurination of G, 7mG, and TXNA-G were conducted at 37°C in glycosylase buffer using 50 nM DNA, with the same Cy5-oligodeoxynucleotides described above. The G-DNA oligo was the same as that used to make the TXNA-G oligo. Samples were quenched and products quantified the same as the enzymatic reactions described above.

EndoIV abasic (AP) site incision kinetics

AP-DNA substrates were generated by incubation of 5 nM YcaQ or TxnU4 with 50 nM Cy5-(TXNA/7mG)-DNA in glycosylase buffer for 2 hr at 25°C. EndoIV incision reactions were performed by adding 6 µL of 83 nM EndoIV (17 nM final concentration) to a 24-µL glycosylase reaction aliquot and incubating at 37°C. Reactions were heated at 70°C for 5 min with 5 µL of formamide/blue dextran loading buffer and electrophoresed and imaged as above. Curve fitting was performed in Prism 9 using a single exponential one-phase association for 7mG-AP site incision and an exponential two-phase association for TXNA-G-AP sites.

Colibactin plasmid unhooking and colibactin-15a DNA preparation and excision

The colibactin 15a monoalkylating derivative was supplied by the Seth Herzon lab (Yale University) and prepared as in (Healy *et al.*, 2016). CLB^{15a} was dissolved in distilled water to a concentration of 5 mM and stored at -80°C. CLB^{15a}-DNA adducts were prepared by annealing the Cy5 strand containing the CLB reactive sequence (central adenine) to the complementary unlabeled oligo (Table 1), followed by incubation of 100 μM DNA with 500 μM CLB^{15a} in TE buffer (pH 8.0) at 37°C in the dark for 3 hrs. Unreacted drug was removed using a G-25 spin column equilibrated in TE buffer (pH 8.0), and the DNA was stored at -80°C. The *N*5-substitutied formidopyrimidine (*N*5-FaPy) derivative of CLB^{15a}-DNA was generated by treating 500 nM DNA with 100 mM NaOH at 37°C for 30 min, followed by re-adjusting to pH 7.0 with 100 mM HCl. Base excision assays using (FaPy)-CLB^{15a}-DNA were performed identical to previous assays described in terms of hydroxide work-up conditions. The concentration of various proteins in these reactions was 1 μM, and the DNA concentration was slightly higher than normal at 500 nM. DNA were electrophoresed and gels imaged/quantified in the same manner as described previously.

CLB-ICL plasmids (derived from the pUC19 plasmid; NEB) were supplied by the Jason Crawford lab (Yale University), and were prepared as in (Xue *et al.*, 2019). Plasmids were stored at -20°C. For CLB-plasmid ICL unhooking assays, 10 ng/μL of pUC19 CLB-ICL was incubated with 1 μM *E. coli* YcaQ in glycosylase buffer for the specified time at 37°C. At each time point, 50 ng of reaction mixture (5 μL) was removed an incubated with 1 μL of *E. coli* EndoIV (14 nM) for 5 min at 37°C. EndoIV reactions were combined with 6 μL of 2X blue dextran formamide loading buffer, and denatured at 70°C

for 5 min. Reactions were electrophoresed on a 1% TAE-agarose native gel at 100 volts for 1.5 hrs, and subsequently stained with SYBR-Gold stain (1:10,000 dilution) at 25°C for 2 hrs. Gels were visualized on a BioRad gel imaging system using ultraviolet detection, and bands (nicked, linear, supercoiled) were quantified with ImageQuant (GE Healthcare). Time course data was plotted using Prism 9.0 software and fit with a one-phase association.

Cellular Assays

E. coli growth curves in mechlorethamine (HN2) and MMS

Genetic knockouts of *ycaQ* and *uvrA* were obtained from the *E. coli* Keio knockout collection (Dharmacon, GE Healthcare) that contained a kanamycin resistance cassette in place of the endogenous gene (Table 2). The kanamycin resistance cassettes were removed using a FLP recombinase expressed on a temperature-sensitive pCP20 plasmid (Amp^R). *ΔuvrA*/*ΔycaQ E. coli* were generated by recombineering through knockout of *ycaQ* in l-Red competent *ΔuvrA* cells (λ-Red carried on temperature-sensitive pKN208 plasmid- Kan^R). To generate *E. coli* growth curves, overnight cultures were diluted to 0.01 OD_{600} in LB supplemented with either 0 μM or 33 μM mechlorethamine, or 5 mM MMS in a 96-well flat-bottom plate. The plate was incubated at 30°C with shaking for 11 hr and cell density was measured at 600 nm every 20 min using a Bio-Tek Synergy 2 microplate reader.

E. coli survival in MMS and HN2 by colony dilution

E. coli survival curves after MMS or mechlorethamine treatment were performed using a colony dilution assay. *YcaQ* was overexpressed from a modified pBG103 (Kan^R) vector and *nfo* (EndoIV) was overexpressed from a modified pHD116 (Amp^R) vector, and expression confirmed by SDS-PAGE of cell lysates after IPTG induction (SI Fig. S6B in (Bradley *et al.*, 2020)). A saturated overnight culture from a single colony was diluted to 0.01 OD_{600} in fresh LB media and grown to 0.4 OD_{600} at 37°C, IPTG added to 100 μM, and the cells incubated at 37°C for 1 h. The cells were transferred to fresh media and treated with various concentrations of either MMS or mechlorethamine for 2 hr at 37°C. Treated cells were transferred to fresh media, serially diluted by a factor of 10^{-5} or 10^{-6} in LB media, and 100 μL of diluted cells were plated on LB agar plates and grown at 37°C overnight. Colonies were counted the next morning and the CFU/mL of culture was determined. The percent survival was calculated as $CFU/mL_{Treated}/CFU/mL_{Untreated}$. Curves were fit to single exponential and EC_{50} values determined by the half-time, $t_{1/2}$. For the genetic complementation experiments, *ycaQ* was expressed from a pSF-OXB11 (Kan^R) plasmid allows for constitutive gene expression at intermediate levels, and LB media was supplemented with 30 μg/mL kanamycin to retain the plasmid (Table 2).

Detection of gene expression by quantitative RT-PCR in E. coli

Saturated *E. coli* K-12 cultures from a single colony were diluted to 0.01 OD_{600} in LB media and grown at 37°C until an OD_{600} of 0.5, after which MMS (5 mM) or mechlorethamine (200 μM) was added and cultures grown at 37°C for an additional 2 hr. Cultures were lysed with TRIzol reagent and the RNA extracted with phenol-chloroform

and precipitated with isopropanol/ethanol as in (Hay, 2017). Residual genomic DNA was removed from the RNA by treatment with DNase I (New England BioLabs). The RNA was re-extracted from the reaction mixture and quantified by absorbance at 260 and 280 nm. The specificity of the primers and quality of the RNA was verified by agarose gel analysis of RT-PCR products (SI Fig. S6F in (Bradley *et al.*, 2020)). cDNA synthesis and qPCR were performed in a single step reaction using the iTaq Universal SYBR Green One-Step Kit (Bio-Rad) on a BioRad CFX-96 real-time PCR thermal cycler. The housekeeping gene used was *gapA* (*Eco* GAPDH). The results from the qPCR experiments were performed on three biologically replicated RNA extractions from both MMS and mechlorethamine treatments. The fold expression change was calculated using the formula: (fold expression change) = $2^{-\Delta\Delta Ct}$, where Ct is the cycle threshold for amplification above baseline, ΔCt = Ct (gene of interest) – Ct (housekeeping gene), and $\Delta\Delta Ct$ = ΔCt (treated sample) – ΔCt (untreated sample).

Hedamycin resistance in E. coli by growth curves and colony dilution

The *hedH4* wild-type gene was sub-cloned from pBG102 into pSF-OXB1 using NcoI and XbaI restriction sites. The pSF-OXB1 vector contains a kanamycin resistance gene and allows for constitutive low-level expression from a modified AraBAD promoter. pSF-OXB1 and HedH4/pSF-OXB1 were transformed into *E. coli* K-12 cells (Table 2). Cloning of *hedH4* was confirmed by sequencing, restriction digest using NcoI-HF/XbaI (SI Fig. S3D in (Bradley *et al.*, 2022)), and colony PCR of K-12 transformants using the HedH4 NcoI and XbaI primers (SI Fig. S3E in (Bradley *et al.*, 2022), Table 1). Cultures were grown at 37°C in LB media supplemented with 30 µg/mL Kan. Growth curves were

generated by diluting overnight cultures to 0.01 OD_{600} in LB/Kan supplemented with 0 nM-100 µM hedamycin in a 96-well flat-bottom plate. The plate was incubated at 30°C with shaking for 24 hrs and cell density was measured at 600 nm every 20 min using a Bio-Tek Synergy 2 microplate reader. IC_{50} values were determined from a fit to the equation, Lag time = Min_{lag} + (Max_{lag}-Min_{lag})/(1+(IC_{50}/[Hed])). Growths were performed in triplicate.

E. coli survival curves after hedamycin treatment were performed using a colony dilution assay. A saturated overnight LB/Kan culture from a single colony was diluted to 0.01 OD_{600} in 1 mL fresh LB/Kan media and grown to 0.6 OD_{600} at 37°C. The cells were treated with various concentrations of hedamycin for 1 h at 37°C. Treated cells were transferred to fresh LB/Kan media, serially diluted by 10^{-6} in LB/Kan media, and 100 µL of diluted cells were plated on LB/Kan agar plates and grown at 37°C overnight. Colonies were counted the next morning and the CFU/mL culture was determined. The percent survival was calculated as $CFU/mL_{Treated}$ / $CFU/mL_{Untreated}$. Curves were plotted on a logarithmic scale and IC_{50} values determined by non-linear regression fits to the data. Growths were performed in triplicate.

Fermentation and isolation of TXNA and LLD

For TXNA production, *S. bottropensis* NRRL 12051 and its relative mutant strains were cultivated as previously reported (Gust *et al.*, 2003). After fermentation in SYG medium (soluble starch 60 g/L, glucose 10/L, yeast extract 10/L, NaCl g/L, $MgSO_4 \cdot 7H_2O$ 1 g/L, KH_2PO_4 1 g/L, $CuSO_4 \cdot 5H_2O$ 70 mg/L, $FeSO_4 \cdot 7H_2O$ 10 mg/L, $MnCl_2 \cdot 4H_2O$ 8 mg/L, $ZnSO_4 \cdot 7H_2O$ 2 mg/L, $CoCl_2 \cdot 7H_2O$ 6 µg/L, HP20 30g/L) for 5 days, the TXNA was isolated

and detected as described (Zhang *et al.*, 2015). The fermentation and isolation of LLD was similar to TXNA (Dong *et al.*, 2019). *S. vinaceusdrappus* NRRL 15735 and those mutants were cultivated in SYG medium for 10 days, and then isolated and detected by HPLC. HPLC analysis was performed on an Acclaim 120 C18 column (5 μm, 4.6 × 250 mm) at a flow rate of 1.0 mL/min and a linear gradient program: 0-5 min, 10% phase B (0.1% formic acid in CH_3CN); 5-24 min, solvent B gradient from 10 to 90% followed with 90% B at 24-26 min; 26-27 min, gradient from 90 to 10% B; 27-31 min, constant 10% B. Phase A is 0.1% formic acid in H_2O. TXNA/LLD-related compounds were determined by measuring UV absorbance at 400 nm using an Agilent 1200 series system (Zhang *et al.*, 2015; Dong *et al.*, 2019). LC-MS was carried out on a ThermoFisher LTQ XL under the same conditions.

Cellular TXNA and LLD self-resistance assays

The inhibition zones of *Streptomyces* were performed by a disc diffusion assay. Specifically, filter paper discs spotted with different concentrations of TXNA or LLD were laid on the MS plate (20 g/L soybean meal, 20 g/L mannitol, 20g/L agar, pH 7.2), which were pre-inoculated with wild-type strains *S. bottropensis* NRRL 12051 (*txn*WT), *S. vinaceusdrappus* NRRL 15735 (*lld*WT), the gene mutant strains, Δ*txnU2*, Δ*txnU4*, Δ*lldU1*, Δ*lldU5* or heterologous expression strains *S. lividans*::pSET152, *S. lividans*::*txnU2*, *S. lividans*::*txnU4*, *S. lividans*::*lldU1*, *S. lividans*::*lldU5* (Table 2). After incubation at 30°C for 36 hr, resistance levels to TXNA or LLD were determined by the zone of inhibition.

Heterologous survival assays were conducted in *E. coli*. BL21 cells transformed with protein overexpression plasmid *txnU2*-pET28a, *txnU4*-pET28a, *lldU1*-pET28a, *lldU5*-

pET28a or empty vector pET28a alone, and were grown overnight at 37°C in LB medium containing 50 µg/mL kanamycin (Kan). The overnight cultures were then transferred to fresh LB medium supplemented with 50 µg/mL Kan and incubated at 30°C. When the OD_{600} reached 0.6, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. After growing at 16°C for 2 hr, cells were diluted to 0.01 OD_{600} in 2 mL fresh LB supplemented with Kan and IPTG. The dilutions were treated with various concentrations of TXNA for 12 hr at 30°C and cell density was measured by OD_{600}. The surviving fraction (%) was calculated as $(OD_{600Treated}/OD_{600Untreated})*100$. The data were fit by non-linear regression and plotted using GraphPad 8.0 software.

Protein X-Ray Crystallography and AlphaFold Modelling

Crystallization, X-ray data collection, and refinement of S. sahachiroi AlkZ

SeMet-AlkZ was crystallized by sitting drop vapor diffusion at 21°C by mixing equal volumes of 2 mg/mL protein and reservoir solution containing 18% (w/v) PEG 8000, 10 mM Tris•HCl pH 7.5, and 10 mM $MgCl_2$. Crystals were harvested four days after setting the drops, cryo-protected in reservoir solution supplemented with 15% (v/v) glycerol, and flash-frozen in liquid nitrogen. X-ray diffraction data from a single frozen crystal were collected at a wavelength of 0.97857 Å at the Advanced Photon Source Beamline 21-ID-G at Argonne National Laboratory and processed with HKL2000 (Otwinowski and Minor, 1997). Phasing and refinement were carried out using the PHENIX suite of programs (Adams *et al.*, 2010). Phases were calculated from the positions of eight Se atoms and modified by solvent flattening using PHENIX AutoSol. A partial model consisting of 347 of 371 SeMet-AlkZ residues was built automatically by PHENIX AutoBuild. The remaining

residues were built manually in Coot (Emsley *et al.*, 2010). After adding riding hydrogens with PHENIX ReadySet, atomic coordinates and TLS-derived anisotropic *B*-factors were refined using PHENIX Refine. The final model was validated with MolProbity (Davis *et al.*, 2007) and contained no residues in the disallowed regions of the Ramachandran plot. Refinement and validation statistics are included in (Table 3). Atomic coordinates and structure factors were deposited in the Protein Data Bank under accession code 5UUJ. The AlkZ-DNA model was prepared by extending the previously determined AZB ICL DNA model (Alcaro and Coleman, 2000) with three B-form base pairs on each end, followed by manual docking onto the AlkZ structure using PyMOL. All structural biology software was curated by SBGrid (Morin *et al.*, 2013).

AlphaFold structural modelling of LldU1/5 and TxnU2/4 and TXNA-DNA

The model of TxnU4 bound to TXNA-DNA was based on our previous model of AlkZ docked against an AZB-ICL-DNA (Mullins *et al.*, 2017). A TxnU4 homology model calculated using AlphaFold2 (Jumper *et al.*, 2021) was superimposed onto AlkZ. The crystal structure of TXNA-DNA was superimposed onto the AZB-ICL-DNA model using the adducted guanine as an anchor point. Protein and DNA were treated as rigid bodies, and no energy minimization was performed. All modeling was performed in PyMOL v2.3.2 (Schrödinger, Inc.).

AlphaFold structural modelling of E. coli YcaQ and NM_5-ICL-DNA

The model of *E. coli* YcaQ bound to a kinked hydrazine NM_5-ICL-DNA was based on our previous model of AlkZ docked against an AZB-ICL-DNA (Mullins *et al.*, 2017). A

YcaQ homology model calculated using AlphaFold2 (Jumper *et al.*, 2021) was superimposed onto AlkZ. The model of the hydrazine NM_5-ICL (provided to us by Orlando Schärer; Ulsan National Institute of Science and Technology, South Korea) was superimposed onto the AZB-ICL-DNA model using the adducted guanine as an anchor point. Protein and DNA were treated as rigid bodies, and no energy minimization was performed. All modeling was performed in PyMOL v2.3.2 (Schrödinger, Inc.).

AlphaFold structural modelling of S. griseoruber HedH4 and HED-DNA

The model of HedH4 bound to HED-DNA was based on our previous model of AlkZ docked against an AZB-ICL-DNA (Mullins *et al.*, 2017). A HedH4 homology model calculated using AlphaFold2 (Jumper *et al.*, 2021) was superimposed onto AlkZ. The NMR structure of HED-DNA (PDB: 1JHI) was superimposed onto the AZB-ICL-DNA model using the adducted guanine as an anchor point. Protein and DNA were treated as rigid bodies, and no energy minimization was performed. All modeling was performed in PyMOL v2.3.2 (Schrödinger, Inc.).

THF/7mG DNA preparation and Tfu YQL DNA binding by fluorescence anisotropy

DNA substrates containing a central tetrahydrofuran (THF) spacer and an intercalated 7-methylguanine (7mG) nucleobase were prepared by annealing the FAM-labeled THF_25 to the complementary DNA strand (Table 1) at 200 μM in the presence of a saturated solution of 7mG free base (5 mM prepared in distilled water) in SSC buffer (described above). The solution was heated to 85°C and slowly cooled to room temperature (25°C). Excess undissolved 7mG was removed by centrifugation, and the

DNA was store at -20°C. DNA binding of AlkA^{Eco}, YQL^{Tfu}, and AlkZ^{Ssa} to THF/7mG-DNA was monitored by fluorescence anisotropy. Proteins at varying concentration were incubated with 25 nM FAM-labeled THF/7mG-DNA (Table 1) in glycosylase buffer at 4°C in the dark for 30 min. Data was collected at excitation and emission wavelengths of 485 and 528 nm and were collected at room temperature (25°C) in 96-well plates using a BioTek Synergy H1 plate reader. Data were fit to a one-state binding model using GraphPad Prism 9. Fluorescence quenching of the FAM fluorophore was monitored by plotting total fluorescence ($I_T = I_{\parallel} + 2I_{\perp}$) versus protein concentration, and none was observed.

Crystallization of Tfu YQL with THF/7mG-DNA and preliminary X-ray diffraction

The protein-DNA complex of YQL from *Thermobifida fusca* with THF/7mG-DNA was prepared by combining 10 μM (0.5 mgs/mL) YQL^{Tfu} with 15 μM THF/7mG 18-mer blunt ended DNA with the sequence [5'-d(TGAGTCGT(THF)GATGACCAC)/5'-d(GTGGTCATCCACGACTCA)] (Table 1) and incubating at 4°C for 30 min on ice (diluent: SEC buffer for YQL^{Tfu}). Reactions were then concentrated to 4 mgs/mL in a 10K MWCO spin concentrator, spin-filtered through a 0.2 μm filter, and centrifuged at 20,000 ×g for 10 min at 4°C. Complexes were crystallized using the sitting-drop vapor-diffusion method. Drops were prepared by mixing 2 μL of 4 mgs/mL YQL^{Tfu} + THF/7mG-DNA (120 μM DNA; 1.5 × [protein]) with 2 μL reservoir solution (10% w/v PEG 8K, 0.1 M MES pH 6.0, 0.1 M NaCl, and 0.05 M $CaCl_2$). All drops were equilibrated at 21°C against 500 μL of reservoir solution. After several days, small, needle-like crystal appeared, and a sub-microscopic seed stock of these crystals were prepared via seed bead vortexing in reservoir solution

supplemented with 15% w/v PEG 8K. Complexes were crystallized again via hanging drop vapor diffusion by mixing 2 µL of 3 mgs/mL YQL^{Tfu} + THF/7mG-DNA (120 µM DNA; 1.5 × [protein]) with 1.5 µL reservoir solution (6% w/v PEG 8K, 0.1 M MES pH 6.0, 0.1 M NaCl, and 0.05 M $CaCl_2$) and 0.5 µL of 10^{-4} dilution of seed stock. All drops were equilibrated at 21°C against 500 µL of reservoir solution. After several days, crystals were collected, briefly soaked in reservoir solution supplemented with 15% (v/v) glycerol, and flash-cooled in liquid nitrogen. Preliminary X-ray diffraction data was collected on beamline 21-ID-D (λ = 1.00324 Å) at the Advanced Photon Source.

Phylogenetic and Bioinformatic Analyses

Taxonomy and phylogeny of Streptomyces HTH_42 proteins

To identify HTH_42 proteins in *Streptomyces*, the protein sequences for YcaQ (GenBank accession number QHB65847.1) and AlkZ (GenBank accession number ABY83174.1) were used for tBLASTn and BLASTp searches (BLAST+ v2.11.0) against all *Streptomyces* genomes (taxid:1883). Searches were run with the BLOSUM62 matrix, 1000 maximum target sequences, and 0.05 threshold using an e-value and identity cutoff of 10^{-4} and 25%, respectively. All hits were verified for the presence of the (H/Q)Φ(D/Q) catalytic motif, during which the (H/Q)Φ(S/T)(D/E) (AZL2) variant was identified. Truncated genes, poor sequence quality genes, and pseudogenes were eliminated. Additional sequences were obtained by searching the Pfam database v33.1 (El-Gebali *et al.*, 2019) for *Streptomyces* HTH_42 superfamily members (PF06224). Sequences from Pfam were sorted according to their domain classes (SI Fig. S1A in (Bradley *et al.*, 2022)), and only sequences from Class 1 with >75% coverage were included. Protein sequences

were aligned using EMBL-EBI Clustal OmegaW or MAFFT v7 using default parameters (Katoh *et al.*, 2019; Madeira *et al.*, 2019). The evolutionary history of YQL/AZL sequences were reconstructed using IQTREE2 with default settings (Minh *et al.*, 2020), and the phylogenetic tree was assembled with the Interactive Tree of Life (v5) phylogeny display tool (Letunic and Bork, 2019). Sequence logos were generated with WebLogo v2.8.2 (Crooks *et al.*, 2004). The copy number frequency and coincidence of YQL/AZL in the same genome was determined by manually counting the number and identity of homologs in each species. A list of all YQL/AZL/AZL2 proteins and *Streptomyces* genomes analyzed in this study can be found in SI Table S1 in (Bradley *et al.*, 2022).

Identification of AZL homologs in known biosynthetic gene clusters

To find AZL proteins in verified and/or published BGCs, we searched MIBiG v2.0 for the AZB BGC (BGC0000960) from *S. sahachiroi* (Zhao *et al.*, 2008; Kautsar *et al.*, 2020), followed by an iterative search using the *MIBiG Hits* function until no more hits were obtained. The homologs TxnU2 and TxnU4 were identified from the initial BLAST search within the deposited NCBI trioxacarcin BGC sequence (Zhang *et al.*, 2015). The homolog within the aclacinomycin BGC was also identified in the initial BLAST search as appearing in proximity to aclacinomycin biosynthesis genes. Closer inspection of the published sequence for the aclacinomycin BGC (GenBank accession number AB008466.1) revealed an AZL protein (Orf1) located immediately 3' of the cluster (Chung *et al.*, 2002). A detailed list of the AZL proteins in known BGCs can be found in Table 4.

Identification of AZL homologs in uncharacterized biosynthetic gene clusters

To determine the physical distance in base pairs between the genomic coordinates of AZL proteins and those of BGCs present in the genome assemblies of *Streptomyces* (average number of scaffolds: 96.30; minimum: 1; maximum: 1,956), we first predicted the BGCs in each genome using antiSMASH v5.1.0 (Blin *et al.*, 2019) with the *taxon* parameter set to *bacteria*. Using the BGC sequences identified from antiSMASH and AZL sequences, a custom python script using Biopython (Cock *et al.*, 2009) determined the shortest base pair distance between the physical location of the YQL/AZL gene and the location of the nearest BGC on the same scaffold (less than 2 Mbps away). To be considered within a BGC, the homolog had to be observed within 5 genes or 2 kb of the nearest cluster. Known Cluster BLAST was performed within antiSMASH to determine the most similar BGC to the unknown clusters, and the result with the highest percentage of similar genes was recorded as the most similar cluster. A detailed list of the genome information, cluster IDs, and the closest 3' and/or 5' BGC can be found in SI Table S2 in (Bradley *et al.*, 2022). A detailed list of the AZL proteins in uncharacterized BGCs can be found in Table 5. The mining of AZL homolog in Actinobacterial BGCs beyond *Streptomyces* was performed in a manner similar to that described above (Fig. 69A-D).

Gene ontology analysis for selected Streptomyces YQL and AZL homologs

To identify GO terms for nearest neighbors identified through BLAST, Pfam, and MIBiG searches, we randomly chose 40 homologs each of AZL inside BGCs, AZL outside BGCs, and YQL, which represent ~10% of the sequences for each. Amino acid sequences for the five genes on both sides of the YQL/AZL genes were downloaded from

the NCBI database, for a total of 400 neighbors for each of the three classes. Cellular functions of any already annotated genes in the NCBI database were identified and recorded. The downloaded sequences were then run through the GhostKOALA (v2.2) and eggNOG (v5.0) GO annotation databases (Kanehisa *et al.*, 2016; Huerta-Cepas *et al.*, 2019). After known GO terms for all gene neighbors were identified, proteins were categorized by biological processes and molecular functions, and the values for these terms were used to create the GO term distributions. Proteins that had multiple GO terms associated with them were counted into each class of terms. A list of all proteins and their annotated GO terms can be found in (SI Table S5-S6 in (Bradley *et al.*, 2022)).

Sequence similarity network (SSN) analysis for LldU and TxnU proteins

15,119 homologous proteins of AlkZ were obtained from the InterPro website (Mitchell *et al.*, 2015) by using AlkZ as the query. Sequences were then clustered by CD-HIT Suite (Huang *et al.*, 2010) with 53% sequence identity threshold. The representatives of the resulting clusters and TxnU2, TxnU4, LldU1, LidU5, AlkZ were used for construction of SSN by the online Enzyme Function Initiative-Enzyme Similarity Tool (Gerlt *et al.*, 2015) with an alignment score threshold of 110. Cytoscape software was used to view the sequence similarity networks.

Phylogenetic and taxonomic identification of horizontal gene transfer of AlkZ into chytrid fungi

Through bioinformatic searches of the HTH 42 superfamily within the Pfam database (El-Gebali *et al.*, 2019), a small subset of AlkZ-like protein sequences were

observed within a small number of fungal eukaryote species (specifically chytrid fungi). These sequences were downloaded from Pfam and aligned to *Streptomyces sahachiroi* AlkZ using the T-Coffee alignment software (Notredame *et al.*, 2000). Analysis of the N-terminal localization sequence (NLS) was performed using the cNLS mapper software (Kosugi *et al.*, 2009). The taxonomic analysis of eukaryotic species was performed by first downloading the tree elements from the NCBI common tree taxonomy browser (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi), and generating the tree file (Newick) using PhyloT v2 (https://phylot.biobyte.de/index.cgi). The taxonomic tree was visualized and modified with iTOL v4 (Letunic and Bork, 2019). The proximal phylogeny of *Spizellomyces punctatus* AlkZ was generated through alignment of the top 50 BLASTp hits using T-Coffee alignment and the iTOL visualization software.

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* **66**, 213-221.

Adhikary, S., and Eichman, B.F. (2011). Analysis of Substrate Specificity of *Schizosaccharomyces Pombe* Mag1 Alkylpurine DNA Glycosylase. *EMBO Rep* **12**, 1286-1292.

Alcaro, S., and Coleman, R.S. (2000). A molecular model for DNA cross-linking by the antitumor agent azinomycin B. *J Med Chem* **43**, 2783-2788.

Alseth, I., Rognes, T., Lindbäck, T., Solberg, I., Robertsen, K., Kristiansen, K.I., Mainieri, D., Lillehagen, L., Kolstø, A.B., and Bjørås, M. (2006). A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol Microbiol* **59**, 1602-1609.

Arike, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012). Comparison and Applications of Label-Free Absolute Proteome Quantification Methods on *Escherichia Coli*. *J Proteomics* **75**, 5437-5448.

Armstrong, R.W., M. E. Salvati, and M. Nguyen (1992). Novel Interstrand Cross-Links Induced by the Antitumor Antibiotic Carzinophilin/Azinomycin B. *J Am Chem Soc* **114**, 3144-3145.

Bassarello, C., Cimino, P., Bifulco, G., Boger, D.L., Smith, J.A., Chazin, W.J., and Gomez-Paloma, L. (2003). NMR structure of the (+)-CPI-indole/d(GACTAATTGAC)-d(GTCAATTAGTC) covalent complex. *Chembiochem* **4**, 1188-1193.

Bauer, G.B., and L. F. Povirk (1997). Specificity and Kinetics of Interstrand and Intrastrand Bifunctional Alkylation by Nitrogen Mustards at a G-G-C Sequence. *Nucleic Acids Res* **25**, 1211-1218.

Bebenek, A., and Ziuzia-Graczyk, I. (2018). Fidelity of DNA replication-a matter of proofreading. *Curr Genet* **64**, 985-996.

Belknap, K.C., Park, C.J., Barth, B.M., and Andam, C.P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci Rep* **10**, 2003.

Berardini, M., P. L. Foster, and E. L. Loechler (1999). DNA Polymerase II (*polB*) Is Involved in a New DNA Repair Pathway for DNA Interstrand Cross-Links in *Escherichia coli*. *J Bacteriol Res* **191**, 2878-2882.

Berardini, M., W. Mackay, and E. L. Loechler (1997). Evidence for a Recombination-Independent Pathway for the Repair of DNA Interstrand Cross-Links Based on a Site-Specific Study with Nitrogen Mustard. *Biochemistry* **36**, 3506-3513.

Beretta, G.L., and Zunino, F. (2008). Molecular mechanisms of anthracycline activity. *Top Curr Chem* **283**, 1-19.

Blair, J.M., Webber, M.A., Baylay, A.J., Ogbolu, D.O., and Piddock, L.J. (2015). Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol* **13**, 42-51.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87.

Boger, D.L., and Garbaccio, R.M. (1997). Catalysis of the CC-1065 and Duocarmycin DNA Alkylation Reaction: DNA Binding Induced Conformational Change in the Agent Results in Activation *Bioorg Med Chem* **5**, 263-276.

Boger, D.L., and Johnson, D.S. (1995). CC-1065 and the duocarmycins: unraveling the keys to a new class of naturally derived DNA alkylating agents. *Proc Natl Acad Sci U S A* **92**, 3642-3649.

Boiteux, S., Huisman, O., and Laval, J. (1984). 3-Methyladenine Residues in DNA Induce the SOS Function *sfiA* in *Escherichia coli*. *EMBO J* **3**, 2569-2573.

Bonnet, M., Buc, E., Sauvanet, P., Darcha, C., Dubois, D., Pereira, B., Dechelotte, P., Bonnet, R., Pezet, D., and Darfeuille-Michaud, A. (2014). Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clin Cancer Res* **20**, 859-867.

Boolchandani, M., D'Souza, A.W., and Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20**, 356-370.

Boshoff, H.I., Reed, M.B., Barry, C.E., 3rd, and Mizrahi, V. (2003). DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* **113**, 183-193.

Bossuet-Greif, N., Vignard, J., Taieb, F., Mirey, G., Dubois, D., Petit, C., Oswald, E., and Nougayrede, J.P. (2018). The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *MBio* **9**, e02393-02317.

Bradley, N.P., Wahl, K.L., Steenwyk, J.L., Rokas, A., and Eichman, B.F. (2022). Resistance-guided mining of bacterial genotoxins defines a family of DNA glycosylases. *MBio (In Press)*.

Bradley, N.P., Washburn, L.A., Christov, P.P., Watanabe, C.M.H., and Eichman, B.F. (2020). *Escherichia coli* YcaQ is a DNA glycosylase that unhooks DNA interstrand crosslinks. *Nucleic Acids Res* **48**, 7005-7017.

Brennan, R.G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* **74**, 773-776.

Brooks, S.C., Adhikary, S., Rubinson, E.H., and Eichman, B.F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochim Biophys Acta* **1834**, 247-271.

Brulikova, L., Hlavac, J., and Hradil, P. (2012). DNA interstrand cross-linking agents and their chemotherapeutic potential. *Curr Med Chem* **19**, 364-385.

Burby, P.E., and Simmons, L.A. (2019). A bacterial DNA repair pathway specific to a natural antibiotic. *Mol Microbiol* **111**, 338-353.

Burgos-Barragan, G., Wit, N., Meiser, J., Dingler, F.A., Pietzke, M., Mulderrig, L., Pontel, L.B., Rosado, I.V., Brewer, T.F., Cordell, R.L.*, et al.* (2017). Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature* **548**, 549-554.

Buttner, M.J., Schafer, M., Lawson, D.M., and Maxwell, A. (2018). Structural insights into simocyclinone as an antibiotic, effector ligand and substrate. *FEMS Microbiol Rev* **42**.

Caron, P.R., Kushner, S.R., and Grossman, L. (1985). Involvement of helicase II (*uvrD* gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex. *Proc Natl Acad Sci U S A* **82**, 4925-4929.

Castano, A., Roy, U., and Schärer, O.D. (2017). Preparation of Stable Nitrogen Mustard DNA Interstrand Cross-Link Analogs for Biochemical and Cell Biological Studies. *Methods Enzymol* **591**, 415-431.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**, 235-263.

Chen, J., Ghorai, M.K., Kenney, G., and Stubbe, J. (2008). Mechanistic studies on bleomycin-mediated DNA damage: multiple binding modes can result in double-stranded DNA cleavage. *Nucleic Acids Res* **36**, 3781-3790.

Chen, X., Bradley, N.P., Lub, W., Wahl, K.L., Zhang, M., Yuan, H., Hou, X.F., Eichman, B.F., and Tang, G.L. (2022). Base excision repair system targeting DNA adducts of antibiotics trioxacarcin/LL-D49194 for self-resistance. *Nucleic Acids Res (In Press)*.

Chen, X., Sun, Y., Wang, S., Ying, K., Xiao, L., Liu, K., Zuo, X., and He, J. (2020). Identification of a novel structure-specific endonuclease AziN that contributes to the repair of azinomycin B-mediated DNA interstrand crosslinks. *Nucleic Acids Res* **48**, 709-718.

Cheng, S., Sancar, A., and Hearst, J.E. (1991). RecA-dependent incision of psoralen-crosslinked DNA by (A)BC excinuclease. *Nucleic Acids Res* **19**, 657-663.

Cheung-Ong, K., Giaever, G., and Nislow, C. (2013). DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology. *Chem Biol* **20**, 648-659.

Chumduri, C., Gurumurthy, R.K., Zietlow, R., and Meyer, T.F. (2016). Subversion of host genome integrity by bacterial pathogens. *Nat Rev Mol Cell Biol* **17**, 659-673.

Chung, J.Y., Fujii, I., Harada, S., Sankawa, U., and Ebizuka, Y. (2002). Expression, purification, and characterization of AknX anthrone oxygenase, which is involved in aklavinone biosynthesis in *Streptomyces galilaeus*. *J Bacteriol* **184**, 6115-6122.

Clauson, C., Schärer, O.D., and Niedernhofer, L. (2013). Advances in Understanding the Complex Mechanisms Of DNA Interstrand Cross-Link Repair. *Cold Spring Harb Perspect Biol* **5**, a012732.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B.*, et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423.

Cole, R.S. (1973). Repair of DNA containing interstrand crosslinks in *Escherichia coli*: sequential excision and recombination. *Proc Natl Acad Sci U S A* **70**, 1064-1068.

Coleman, R.S., R. J. Perez, C. H. Burk, and A. Navarro (2002). Studies on the Mechanism of Action of Azinomycin B: Definition of Regioselectivity and Sequence Selectivity of DNA Cross-Link Formation and Clarification of the Role of the Naphthoate. *J Am Chem Soc* **124**, 13008-13017.

Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbiol Spectr* **3**.

Couvé, 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.

Couvé-Privat, S., Mace, G., Rosselli, F., and Saparbaev, M.K. (2007). Psoralen-induced DNA adducts are substrates for the base excision repair pathway in human cells. *Nucleic Acids Res* **35**, 5672-5682.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190.

Cuevas-Ramos, G., Petit, C.R., Marcq, I., Boury, M., Oswald, E., and Nougayrede, J.P. (2010). *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A* **107**, 11537-11542.

Cundliffe, E., and Demain, A.L. (2010). Avoidance of suicide in antibiotic-producing microbes. *J Ind Microbiol Biotechnol* **37**, 643-672.

D'Costa, V.M., McGrann, K.M., Hughes, D.W., and Wright, G.D. (2006). Sampling the antibiotic resistome. *Science* **311**, 374-377.

Dalmasso, G., Cougnoux, A., Delmas, J., Darfeuille-Michaud, A., and Bonnet, R. (2014). The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. *Gut Microbes* **5**, 675-680.

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 Alencar, T.A., Leitao, A.C., and Lage, C. (2005). Nitrogen mustard- and half-mustard-induced damage in *Escherichia coli* requires different DNA repair pathways. *Mutat Res* **582**, 105-115.

de Almeida, L.C., Calil, F.A., Machado-Neto, J.A., and Costa-Lotufo, L.V. (2021). DNA damaging agents and DNA repair: From carcinogenesis to cancer therapy. *Cancer Genet* **252-253**, 6-24.

de Wit, R., Stoter, G., Kaye, S.B., Sleijfer, D.T., Jones, W.G., ten Bokkel Huinink, W.W., Rea, L.A., Collette, L., and Sylvester, R. (1997). Importance of bleomycin in combination chemotherapy for good-prognosis testicular nonseminoma: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group. *Am J Clin Oncol* **15**, 1837-1843.

Deans, A.J., and West, S.C. (2011). DNA Interstrand Crosslink Repair and Cancer. *Nat Rev Cancer* **11**, 467-480.

Demain, A.L., and Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *J Antibiot (Tokyo)* **62**, 5-16.

Dong, L., Shen, Y., Hou, X.-F., Li, W.-J., and Tang, G.-L. (2019). Discovery of druggability-improved analogues by investigation of the LL-D49194α1 biosynthetic pathway. *Org Lett* **21**, 2322-2325.

Drohat, A.C., and Coey, C.T. (2016). Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA. *Chem Rev* **116**, 12711-12729.

Drohat, A.C., and Maiti, A. (2014). Mechanisms for enzymatic cleavage of the *N*-glycosidic bond in DNA. *Org Biomol Chem* **12**, 8367-8378.

Dronkert, M.L.G., and Kanaar, R. (2001). Repair of DNA interstrand cross-links. *Mutat Res* **486**, 217-247.

Dziubanska-Kusibab, P.J., Berger, H., Battistini, F., Bouwman, B.A.M., Iftekhar, A., Katainen, R., Cajuso, T., Crosetto, N., Orozco, M., Aaltonen, L.A.*, et al.* (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med* **26**, 1063-1069.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A.*, et al.* (2019). The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432.

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

Fais, T., Delmas, J., Barnich, N., Bonnet, R., and Dalmasso, G. (2018). Colibactin: More Than a New Bacterial Toxin. *Toxins (Basel)* **10**.

Fang, Q., Noronha, A.M., Murphy, S.P., Wilds, C.J., Tubbs, J.L., Tainer, J.A., Chowdhury, G., Guengerich, F.P., and Pegg, A.E. (2008). Repair of *O*6-G-alkyl-*O*6-G interstrand cross-links by human *O*6-alkylguanine-DNA alkyltransferase. *Biochemistry* **47**, 10892-10903.

Fernandez de Henestrosa, A.R., and Barbé, J. (1991). Induction of the *alkA* Gene of *Escherichia coli* in Gram-Negative Bacteria. *J Bacteriol Res* **173**, 7736-7740.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A.*, et al.* (2016). The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res* **44**, D279-285.

Fitzner, A., Frauendorf, H., Laatsch, H., and Diederichsen, U. (2008). Formation of gutingimycin: analytical investigation of trioxacarcin A-mediated alkylation of dsDNA. *Anal Bioanal Chem* **390**, 1139-1147.

Flatman, R.H., Howells, A.J., Heide, L., Fiedler, H.P., and Maxwell, A. (2005). Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob Agents Chemother* **49**, 1093-1100.

Frederick, C.A., Williams, L.D., Ughetto, G., van der Marel, G.A., van Boom, J.H., Rich, A., and Wang, A.H. (1990). Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry* **29**, 2538-2549.

Fromme, J.C., Banerjee, A., and Verdine, G.L. (2004). DNA glycosylase recognition and catalysis. *Curr Opin Struct Biol* **14**, 43-49.

Fromme, J.C., and Verdine, G.L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Biol* **9**, 544-552.

Fromme, J.C., and Verdine, G.L. (2004). Base excision repair. *Adv Protein Chem Struct Biol* **69**, 1-41.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826.

Fujiwara, T., Saito, I., and Sugiyama, H. (1999). Highly efficient DNA interstrand crosslinking induced by an antitumor antibiotic, carzinophilin. *Tetrahedron Lett* **40**, 315-318.

Gajiwala, K.S., and Burley, S.K. (2000). Winged helix proteins. *Curr Opin Struct Biol* **10**, 110-116.

Galm, U., Hager, M.H., Van Lanen, S.G., Ju, J., Thorson, J.S., and Shen, B. (2005). Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem Rev* **105**, 739-758.

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.

Gates, K.S., Nooner, T., and Dutta, S. (2004). Biologically Relevant Chemical Reactions of *N*7-Alkylguanine Residues in DNA. *Chem Res Toxicol* **17**, 840-852.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495-506.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., and Whalen, K.L. (2015). Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta Gen Subj* **1854**, 1019-1037.

Gillet, L.C., and Schärer, O.D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* **106**, 253-276.

Glassner, B.J., Rasmussen, L.J., Najarian, M.T., Posnick, L.M., and Leona D. Samson, L.D. (1998). Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proc Natl Acad Sci U S A* **95**, 9997–10002.

Grafstrom, R.C., Fornace, A.J., Jr., Autrup, H., Lechner, J.F., and Harris, C.C. (1983). Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. *Science* **220**, 216-218.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* **100**, 1541-1546.

Han, X., and Liehr, J.G. (1995). Microsome-mediated 8-hydroxylation of guanine bases of DNA by steroid estrogens: correlation of DNA damage by free radicals with metabolic activation to quinones. *Carcinogenesis* **16**, 2571-2574.

Hansen, M., and Hurley, L. (1995). Altromycin B Threads the DNA Helix Interacting with Both the Major and the Minor Grooves To Position Itself for Site-Directed Alkylation of Guanine N7. *J Am Chem Soc* **117**, 2421-2429.

Hansen, M., Yun , S., and Hurley, L. (1995). Hedamycin intercalates the DNA helix and, through carbohydrate-mediated recognition in the minor groove, directs *N*-alkylation of guanine in the major groove in a sequence-specific manner *Chem Biol* **2**, 229-240.

Hara, M., Akasaka, K., Akinaga, S., Okabe, M., Nakano, H., Gomez, R., Wood, D., Uh, M., and Tamanoi, F. (1993). Identification of Ras farnesyltransferase inhibitors by microbial screening. *Proc Natl Acad Sci U S A* **90**, 2281-2285.

Hara, M., Saitoh, Y., and Nakano, H. (1990). DNA strand scission by the novel antitumor antibiotic leinamycin. *Biochemistry* **29**, 5676-5681.

Harami, G.M., Gyimesi, M., and Kovacs, M. (2013). From Keys to Bulldozers: Expanding Roles for Winged Helix Domains in Nucleic-Acid-Binding Proteins. *Trends Biochem Sci* **38**, 364-371.

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

Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* **10**, 551-564.

Hata, T., Koga, F., Sano, Y., Kanamori, K., Matsumae, A., Sugawara, R., Hoshi, T., Shima, T., Ito, S., and Tomizawa, S. (1954). Carzinophilin, a new tumor inhibitory substance produced by *Streptomyces*. *J Antibiot (Tokyo)* **7**, 107-112.

Hay, M., Y. M. Li, and Y. Ma (2017). RNA Extraction Of *Escherichia Coli* Grown in Lysogeny Broth for use in RT-qPCR. *JEMI Methods* **1**, 1-6.

Healy, A.R., Nikolayevskiy, H., Patel, J.R., Crawford, J.M., and Herzon, S.B. (2016). A Mechanistic Model for Colibactin-Induced Genotoxicity. *J Am Chem Soc* **138**, 15563-15570.

Hecht, S.M. (2000). Bleomycin: new perspectives on the mechanism of action. *J Nat Prod* **63**, 158-168.

Hemminki, K., Peltonen, K., and Vodicka, P. (1989). Depurination from DNA of 7-methylguanine, 7-(2-aminoethyl)-guanine and ring-opened 7-methylguanines. *Chem Biol Interact* **70**, 289-303.

Hitomi, K., Iwai, S., and Tainer, J.A. (2007). The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)* **6**, 410-428.

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.

Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-138.

Hopwood, D.A. (2007). How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol Microbiol* **63**, 937-940.

Hou, X.F., Song, Y.J., Zhang, M., Lan, W., Meng, S., Wang, C., Pan, H.X., Cao, C., and Tang, G.L. (2018). Enzymology of anthraquinone-γ-pyrone ring formation in complex aromatic polyketide biosynthesis. *Angew Chem Int Ed* **57**, 13475-13479.

Huang, M., Lu, J.J., and Ding, J. (2021). Natural Products in Cancer Therapy: Past, Present and Future. *Nat Prod Bioprospect* **11**, 5-13.

Huang, W., Xu, H., Li, Y., Zhang, F., Chen, X.Y., He, Q.L., Igarashi, Y., and Tang, G.L. (2012). Characterization of yatakemycin gene cluster revealing a radical *S*-adenosylmethionine dependent methyltransferase and highlighting spirocyclopropane biosynthesis. *J Am Chem Soc* **134**, 8831-8840.

Huang, Y., and Li, L. (2013). DNA Crosslinking Damage and Cancer - A Tale Of Friend and Foe. *Transl Cancer Res* **2**, 144-154.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680-682.

Hubbard, B.K., and Walsh, C.T. (2003). Vancomycin assembly: nature's way. *Angew Chem Int Ed* **42**, 730-765.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J.*, et al.* (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-D314.

Igarashi, Y., Futamata, K., Fujita, T., Sekine, A., Senda, H., Naoki, H., and Furumai, T. (2003). Yatakemycin, a novel antifungal antibiotic produced by *Streptomyces sp.* TP-A0356. *J Antibiot (Tokyo)* **56**, 107-113.

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.

Ishizeki, S., Ohtsuka, M., Irinoda, K., Kukita, K., Nagaoka, K., and Nakashima, T. (1987). Azinomycins A and B, new antitumor antibiotics. III. Antitumor activity. *J Antibiot (Tokyo)* **40**, 60-65.

Isobe, Y., Okumura, M., McGregor, L.M., Brittain, S.M., Jones, M.D., Liang, X., White, R., Forrester, W., McKenna, J.M., Tallarico, J.A.*, et al.* (2020). Manumycin polyketides act as molecular glues between UBR7 and P53. *Nat Chem Biol* **16**, 1189-1198.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078.

Jacob, C., and Weissman, K.J. (2017). Unpackaging the Roles of *Streptomyces* Natural Products. *Cell Chem Biol* **24**, 1194-1195.

Jensen, P.B., Jensen, P.S., Demant, E.J., Friche, E., Sorensen, B.S., Sehested, M., Wassermann, K., Vindelov, L., Westergaard, O., and Hansen, H.H. (1991). Antagonistic effect of aclarubicin on daunorubicin-induced cytotoxicity in human small cell lung cancer cells: relationship to DNA integrity and topoisomerase II. *Cancer Res* **51**, 5093-5099.

Jiang, S.Y., and Ramachandran, S. (2016). Expansion Mechanisms and Evolutionary History on Genes Encoding DNA Glycosylases and Their Involvement in Stress and Hormone Signaling. *Genome Biol Evol* **8**, 1165-1184.

Jin, S.G., Choi, J.H., Ahn, B., O'Connor, T.R., Mar, W., and Lee, C.S. (2001). Excision repair of adozelesin-*N*3 adenine adduct by 3-methyladenine-DNA glycosylases and UvrABC nuclease. *Mol Cells* **11**, 41-47.

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., Selbig, J., and Willmitzer, L. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* **6**, 364.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A.*, et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H.*, et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol* **3**, 576-583.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-731.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**, 1160-1166.

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Munoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V.*, et al.* (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* **48**, D454-D458.

Kelly, G.T., Liu, C., Smith, R., 3rd, Coleman, R.S., and Watanabe, C.M. (2006). Cellular Effects Induced by the Antitumor Agent Azinomycin B. *Chem Biol* **13**, 485-492.

Kelly, G.T., Sharma, V., and Watanabe, C.M. (2008). An Improved Method for Culturing *Streptomyces sahachiroi*: Biosynthetic Origin of the Enol Fragment of Azinomycin B. *Bioorg Chem* **36**, 4-15.

Kersten, R.D., and Weng, J.K. (2018). Gene-guided discovery and engineering of branched cyclic peptides in plants. *Proc Natl Acad Sci U S A* **115**, E10961-E10969.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M.*, et al.* (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543-D550.

Kiakos, K., Sato, A., Asao, T., McHugh, P.J., Lee, M., and Hartley, J.A. (2007). DNA sequence selective adenine alkylation, mechanism of adduct repair, and *in vivo* antitumor activity of the novel achiral *seco*-amino-cyclopropylbenz[*e*]indolone analogue of duocarmycin AS-I-145. *Mol Cancer Ther* **6**, 2708-2718.

Kisker, C., Kuper, J., and Van Houten, B. (2013). Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* **5**, a012591.

Kjaerbolling, I., Vesth, T., and Andersen, M.R. (2019). Resistance Gene-Directed Genome Mining of 50 *Aspergillus* Species. *mSystems* **4**.

Klapacz, J., Lingaraju, G.M., Guo, H.H., Shah, D., Moar-Shoshani, A., Loeb, L.A., and Samson, L.D. (2010). Frameshift mutagenesis and microsatellite instability induced by human alkyladenine DNA glycosylase. *Mol Cell* **37**, 843-853.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106**, 10171-10176.

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

Labana, P., Dornan, M.H., Lafreniere, M., Czarny, T.L., Brown, E.D., Pezacki, J.P., and Boddy, C.N. (2021). Armeniaspirols inhibit the AAA+ proteases ClpXP and ClpYQ leading to cell division arrest in Gram-positive bacteria. *Cell Chem Biol*.

Law, J.W., Law, L.N., Letchumanan, V., Tan, L.T., Wong, S.H., Chan, K.G., Ab Mutalib, N.S., and Lee, L.H. (2020). Anticancer Drug Discovery from Microbial Sources: The Unique Mangrove *Streptomycetes*. *Molecules* **25**.

Lee, Y.J., Park, S.J., Ciccone, S.L., Kim, C.R., and Lee, S.H. (2006). An *in vivo* analysis of MMC-induced DNA damage and its repair. *Carcinogenesis* **27**, 446-453.

LePla, R.C., Landreau, C.A., Shipman, M., and Jones, G.D. (2005). On the Origin of the DNA Sequence Selectivity of the Azinomycins. *Org Biomol Chem* **3**, 1174-1175.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259.

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*.

Liu, B., Xue, Q., Tang, Y., Cao, J., Guengerich, F.P., and Zhang, H. (2016). Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat Res Rev Mutat Res* **768**, 53-67.

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

Lomax, M.E., Folkes, L.K., and O'Neill, P. (2013). Biological consequences of radiation-induced DNA damage: relevance to radiotherapy. *Clin Oncol (R Coll Radiol)* **25**, 578-585.

Lomovskaya, N., Hong, S.K., Kim, S.U., Fonstein, L., Furuya, K., and Hutchinson, R.C. (1996). The *Streptomyces peucetius drrC* gene encodes a UvrA-like protein involved in daunorubicin resistance and production. *J Bacteriol* **178**, 3238-3245.

Londono-Vallejo, J.A., and Dubnau, D. (1993). *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol Microbiol* **9**, 119-131.

Ma, L., Sun, S., Yuan, Z., Deng, Z., Tang, Y., and Yu, Y. (2020). Three putative DNA replication/repair elements encoding genes confer self-resistance to distamycin in *Streptomyces netropsis*. *Acta Biochim Biophys Sin (Shanghai)* **52**, 91-96.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D.*, et al.* (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641.

Maiese, W.M., Labeda, D.P., Korshalla, J., Kuck, N., Fantini, A.A., Wildey, M.J., Thomas, J., and Greenstein, M. (1990). LL-D49194 antibiotics, a novel family of antitumor

agents: taxonomy, fermentation and biological properties. *J Antibiot (Tokyo)* **43**, 253-258.

Maiti, A., Noon, M.S., MacKerell, A.D., Jr., Pozharski, E., and Drohat, A.C. (2012). Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc Natl Acad Sci U S A* **109**, 8091-8096.

Makridakis, N.M., and Reichardt, J.K. (2012). Translesion DNA polymerases and cancer. *Front Genet* **3**, 174.

Malta, E., Moolenaar, G.F., and Goosen, N. (2007). Dynamics of the UvrABC nucleotide excision repair proteins analyzed by fluorescence resonance energy transfer. *Biochemistry* **46**, 9080-9088.

Maskey, R.P., Helmke, E., Kayser, O., Fiebig, H.H., Maier, A., Busche, A., and Laatsch, H. (2004). Anti-cancer and antibacterial trioxacarcins with high anti-malaria activity from a marine *Streptomycete* and their absolute stereochemistry. *J Antibiot* **57**, 771-779.

Maskey, R.P., Sevvana, M., Uson, I., Helmke, E., and Laatsch, H. (2004). Gutingimycin: a highly complex metabolite from a marine *Streptomycete*. *Angew Chem Int Ed Engl* **43**, 1281-1283.

Mattes, W.B., Lee, C., Laval, J., and O'Connor, T.R. (1996). Excision of DNA adducts of nitrogen mustards by bacterial and mammalian 3-methyladenine-DNA glycosylases. *Carcinogenesis* **17**, 643-648.

Medina, E.M., Robinson, K.A., Bellingham-Johnstun, K., Ianiri, G., Laplante, C., Fritz-Laylin, L.K., and Buchler, N.E. (2020). Genetic transformation of *Spizellomyces*

punctatus, a resource for studying chytrid biology and evolutionary cell biology. *Elife* **9**.

Metz, A.H., Hollis, T., and Eichman, B.F. (2007). DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J* **26**, 2411-2420.

Mielecki, D., Wrzesinski, M., and Grzesiuk, E. (2015). Inducible repair of alkylated DNA in microorganisms. *Mutat Res Rev Mutat Res* **763**, 294-305.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S.*, et al.* (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213-D221.

Miyazono, K., Furuta, Y., Watanabe-Matsui, M., Miyakawa, T., Ito, T., Kobayashi, I., and Tanokura, M. (2014). A sequence-specific DNA glycosylase mediates restriction-modification in *Pyrococcus abyssi*. *Nat Commun* **5**, 3178.

Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P.M., and Cortez, D. (2019). HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **176**, 144-153 e113.

Moolenaar, G.F., Franken, K.L., Dijkstra, D.M., Thomas-Oates, J.E., Visse, R., van de Putte, P., and Goosen, N. (1995). The C-terminal region of the UvrB protein of

Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *J Biol Chem* **270**, 30508-30515.

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.

Morita, R., Nakane, S., Shimada, A., Inoue, M., Iino, H., Wakamatsu, T., Fukui, K., Nakagawa, N., Masui, R., and Kuramitsu, S. (2010). Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems. *J Nucleic Acids* **2010**, 179594.

Mukherjee, A., and Sasikala, W.D. (2013). Drug-DNA intercalation: from discovery to the molecular mechanism. *Adv Protein Chem Struct Biol* **92**, 1-62.

Muliandi, A., Katsuyama, Y., Sone, K., Izumikawa, M., Moriya, T., Hashimoto, J., Kozone, I., Takagi, M., Shin-ya, K., and Ohnishi, Y. (2014). Biosynthesis of the 4-methyloxazoline-containing nonribosomal peptides, JBIR-34 and -35, in *Streptomyces sp.* Sp080513GE-23. *Chem Biol* **21**, 923-934.

Mullins, E.A., Dorival, J., Tang, G.L., Boger, D.L., and Eichman, B.F. (2021). Structural evolution of a DNA repair self-resistance mechanism targeting genotoxic secondary metabolites. *Nat Commun* **12**, 6942.

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., Rubinson, E.H., Pereira, K.N., Calcutt, M.W., Christov, P.P., and Eichman, B.F. (2013). An HPLC-tandem mass spectrometry method for simultaneous detection of alkylated base excision repair products. *Methods* **64**, 59-66.

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.

Mullins, E.A., Warren, G.M., Bradley, N.P., and Eichman, B.F. (2017). Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc Natl Acad Sci USA* **114**, 4400-4405.

Mungan, M.D., Alanjary, M., Blin, K., Weber, T., Medema, M.H., and Ziemert, N. (2020). ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res* **48**, W546-W552.

Munita, J.M., and Arias, C.A. (2016). Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **4**.

Murata, M., Takahashi, A., Saito, I., and Kawanishi, S. (1999). Site-specific DNA methylation and apoptosis: induction by diabetogenic streptozotocin. *Biochem Pharmacol* **57**, 881-887.

Nagai, K., Yamaki, H., Tanaka, N., and Umezwa, H. (1967). Inhibition by pluramycin A of nucleic acid biosynthesis. *J Biochem* **62**, 321-327.

Nagaoka, K., Matsumoto, M., Oono, J., Yokoi, K., Ishizeki, S., and Nakashima, T. (1986). Azinomycins A and B, new antitumor antibiotics. I. Producing organism, fermentation, isolation, and characterization. *J Antibiot (Tokyo)* **39**, 1527-1532.

Ng, T.L., Rohac, R., Mitchell, A.J., Boal, A.K., and Balskus, E.P. (2019). An *N*-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **566**, 94-99.

Nicolaou, K.C., Smith, A.L., and Yue, E.W. (1993). Chemistry and biology of natural and designed enediynes. *Proc Natl Acad Sci U S A* **90**, 5881-5888.

Nitiss, J.L., Pourquier, P., and Pommier, Y. (1997). Aclacinomycin A Stabilizes Topoisomerase I Covalent Complexes. *Cancer Res* **57**, 4564-4569.

Noll, D.M., Mason, T.M., and Miller, P.S. (2006). Formation and Repair of Interstrand Cross-Links in DNA. *Chem Rev* **106**, 277-301.

Nooner, T., Dutta, S., and Gates, K.S. (2004). Chemical Properties of the Leinamycin-Guanine Adduct in DNA. *Chem Res Toxicol* **17**, 942-949.

Norman, D., Live, D., Sastry, M., Lipman, R., Hingerty, B.E., Tomasz, M., Broyde, S., and Patel, D.J. (1990). NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(T-A-C-G-T-A).d(T-A-C-G-T-A) duplex. *Biochemistry* **29**, 2861-2875.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217.

Nougayrède, J., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006). *Escherichia coli* Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science* **313**, 848-851.

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cell Mol Life Sci* **66**, 994-1009.

Nowosielska, A., Smith, S.A., Engelward, B.P., and Marinus, M.G. (2006). Homologous Recombination Prevents Methylation-Induced Toxicity in *Escherichia coli*. *Nucleic Acids Res* **34**, 2258-2268.

O'Brien, P.J., and Ellenberger, T. (2004). Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J Biol Chem* **279**, 9750-9757.

O'Brien, P.J., and Ellenberger, T. (2004). The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J Biol Chem* **279**, 26876-26884.

Ordonez, H., and Shuman, S. (2013). *Mycobacterium smegmatis* Lhr Is a DNA-dependent ATPase and a 3'-to-5' DNA translocase and helicase that prefers to unwind 3'-tailed RNA:DNA hybrids. *J Biol Chem* **288**, 14125-14134.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326.

Pang, B., de Jong, J., Qiao, X., Wessels, L.F., and Neefjes, J. (2015). Chemical profiling of the genome with anti-cancer drugs defines target specificities. *Nat Chem Biol* **11**, 472-480.

Pang, B., Qiao, X., Janssen, L., Velds, A., Groothuis, T., Kerkhoven, R., Nieuwland, M., Ovaa, H., Rottenberg, S., van Tellingen, O.*, et al.* (2013). Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat Commun* **4**, 1908.

Panter, F., Krug, D., Baumann, S., and Muller, R. (2018). Self-resistance guided genome mining uncovers new topoisomerase inhibitors from *myxobacteria*. *Chem Sci* **9**, 4898-4908.

Parrish, J.P., Kastrinsky, D.B., Wolkenberg, S.E., Igarashi, Y., and Boger, D.L. (2003). DNA Alkylation Properties of Yatakemycin. *J Am Chem Soc* **125**, 10971-10976.

Parson, K.A., and Snustad, D.P. (1975). Host DNA degradation after infection of *Escherichia coli* with bacteriophage T4: dependence of the alternate pathway of degradation which occurs in the absence of both T4 endonuclease II and nuclear disruption on T4 endonuclease IV. *J Virol* **15**, 221-224.

Parsons, Z.D., Bland, J.M., Mullins, E.A., and Eichman, B.F. (2016). A Catalytic Role for C-H/π Interactions in Base Excision Repair by *Bacillus cereus* DNA Glycosylase AlkD. *J Am Chem Soc* **138**, 11485-11488.

Pegg, A.E., and Byers, T.L. (1992). Repair of DNA containing *O*6-alkylguanine. *FASEB J* **6**, 2302-2310.

Peterson, E., and Kaur, P. (2018). Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Front Microbiol* **9**, 2928.

Pfoh, R., Laatsch, H., and Sheldrick, G.M. (2008). Crystal structure of trioxacarcin A covalently bound to DNA. *Nucleic Acids Res* **36**, 3508-3514.

Pham, J.V., Yilma, M.A., Feliz, A., Majid, M.T., Maffetone, N., Walker, J.R., Kim, E., Cho, H.J., Reynolds, J.M., Song, M.C*., et al.* (2019). A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Front Microbiol* **10**, 1404.

Pleguezuelos-Manzano, C., Puschhof, J., Huber, A.R., van Hoeck, A., Wood, H.M., Nomburg, J., Gurjao, C., Manders, F., Dalmasso, G., Stege, P.B*., et al.* (2020). Mutational signature in colorectal cancer caused by genotoxic pks(+) *E. coli*. *Nature* **580**, 269–273.

Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* **17**, 421-433.

Posnick, L.M., and Samson, L.D. (1999). Imbalanced base excision repair increases spontaneous mutation and alkylation sensitivity in *Escherichia coli*. *J Bacteriol* **181**, 6763-6771.

Povirk, L.F., Hogan, M., and Dattagupta, N. (1979). Binding of bleomycin to DNA: intercalation of the bithiazole rings. *Biochemistry* **18**, 96-101.

Prakash, A., Doublie, S., and Wallace, S.S. (2012). The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. *Prog Mol Biol Transl Sci* **110**, 71-91.

Prestinaci, F., Pezzotti, P., and Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathog Glob Health* **109**, 309-318.

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.

Procopio, R.E., Silva, I.R., Martins, M.K., Azevedo, J.L., and Araujo, J.M. (2012). Antibiotics produced by *Streptomyces*. *Braz J Infect Dis* **16**, 466-471.

Prorok, P., Grin, I.R., Matkarimov, B.T., Ishchenko, A.A., Laval, J., Zharkov, D.O., and Saparbaev, M. (2021). Evolutionary Origins of DNA Repair Pathways: Role of Oxygen Catastrophe in the Emergence of DNA Glycosylases. *Cells* **10**.

Putze, J., Hennequin, C., Nougayrede, J.P., Zhang, W., Homburg, S., Karch, H., Bringer, M.A., Fayolle, C., Carniel, E., Rabsch, W.*, et al.* (2009). Genetic structure and distribution of the colibactin genomic island among members of the family *Enterobacteriaceae*. *Infect Immun* **77**, 4696-4703.

Rajski, S.R., and Williams, R.M. (1998). DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev* **98**, 2723-2796.

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.

Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B., and Sinha, R.P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids* **2010**, 592980.

Reusser, F. (1977). Ficellomycin and feldamycin; inhibitors of bacterial semiconservative DNA replication. *Biochemistry* **16**, 3406-3412.

Reuven, N.B., Koonin, E.V., Rudd, K.E., and Deutscher, M.P. (1995). The gene for the longest known *Escherichia coli* protein is a member of helicase superfamily II. *J Bacteriol* **177**, 5393-5400.

Rink, S.M., and P. B. Hopkins (1995). A Mechlorethamine-Induced DNA Interstrand Cross-Link Bends Duplex DNA. *Biochemistry* **34**, 1439-1444.

Rinne, M., Caldwell, D., and Kelley, M.R. (2004). Transient adenoviral *N*-methylpurine DNA glycosylase overexpression imparts chemotherapeutic sensitivity to human breast cancer cells. *Mol Cancer Ther* **3**, 955-967.

Rinne, M.L., He, Y., Pachkowski, B.F., Nakamura, J., and Kelley, M.R. (2005). *N*-methylpurine DNA glycosylase overexpression increases alkylation sensitivity by rapidly removing non-toxic 7-methylguanine adducts. *Nucleic Acids Res* **33**, 2859-2867.

Rivankar, S. (2014). An overview of doxorubicin formulations in cancer therapy. *J Cancer Res Ther* **10**, 853-858.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324.

Rogozin, I.A., Makarova, K.S., Murvai, J., Czabarka, E., Wolf, W.S., Tatusov, R.L., Szekely, L.A., and Koonin, E.V. (2002). Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* **30**, 2212-2223.

Rubinson, E.H., Gowda, A.S., Spratt, T.E., Gold, B., and Eichman, B.F. (2010). An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**, 406-411.

Rubinson, E.H., Metz, A.H., O'Quin, J., and Eichman, B.F. (2008). A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J Mol Biol* **381**, 13-23.

Russ, C., Lang, B.F., Chen, Z., Gujja, S., Shea, T., Zeng, Q., Young, S., Cuomo, C.A., and Nusbaum, C. (2016). Genome Sequence of *Spizellomyces punctatus*. *Genome Announc* **4**.

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

Salvati, M.E., E. J. Moran, and R. W. Armstrong (1992). Simplified Method for the Isolation of Thermally Labile Drug-DNA Adducts: Characterization of Chlorambucil B and Carzinophilin/Azinomycin Alkylation Products *Tetrahedron Lett* **33**, 3711-3714.

Schärer, O.D. (2005). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* **6**, 27-32.

Schärer, O.D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol* **5**, a012609.

Schärer, O.D., and Jiricny, J. (2001). Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays* **23**, 270-281.

Schermerhorn, K.M., and Delaney, S. (2014). A chemical and kinetic perspective on base excision repair of DNA. *Acc Chem Res* **47**, 1238-1246.

Scott, T.A., and Piel, J. (2019). The hidden enzymology of bacterial natural product biosynthesis. *Nat Rev Chem* **3**, 404-425.

Selby, C.P., and Sancar, A. (1988). ABC excinuclease incises both 5' and 3' to the CC-1065-DNA adduct and its incision activity is stimulated by DNA helicase II and DNA polymerase I. *Biochemistry* **27**, 7184-7188.

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.

Shao, R.G. (2008). Pharmacology and therapeutic applications of enediyne antitumor antibiotics. *Curr Mol Pharmacol* **1**, 50-60.

Shen, Y., Nie, Q.Y., Yin, Y., Pan, H.X., Xu, B., and Tang, G.L. (2019). Production of a trioxacarcin analog by introducing a C-3 dehydratase into deoxysugar biosynthesis. *Acta Biochim Biophys Sin (Shanghai)* **51**, 539-541.

Shi, R., Mullins, E.A., Shen, X.X., Lay, K.T., Yuen, P.K., David, S.S., Rokas, A., and Eichman, B.F. (2018). Selective base excision repair of DNA damage by the non-base-flipping DNA glycosylase AlkC. *EMBO Journal* **37**, 63-74.

Shibutani, S., Takeshita, M., and Grollman, A.P. (1997). Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J Biol Chem* **272**, 13916-13922.

Shimada, N., Uekusa, M., Denda, T., Ishii, Y., Iizuka, T., Sato, Y., Hatori, T., Fukui, M., and Sudo, M. (1955). Clinical studies of carzinophilin, an antitumor substance. *J Antibiot (Tokyo)* **8**, 67-76.

Simkhada, D., Oh, T.J., Kim, E.M., Yoo, J.C., and Sohng, J.K. (2009). Cloning and characterization of CalS7 from *Micromonospora echinospora sp. calichensis* as a glucose-1-phosphate nucleotidyltransferase. *Biotechnol Lett* **31**, 147-153.

Singh, R.K., Kumar, S., Prasad, D.N., and Bhardwaj, T.R. (2018). Therapeutic journery of nitrogen mustard as alkylating anticancer agents: historic to future perspectives. *Eur J Med Chem* **151**, 401-433.

Sinha, R.P., and Hader, D.P. (2002). UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* **1**, 225-236.

Skinnider, M.A., Dejong, C.A., Rees, P.N., Johnston, C.W., Li, H., Webster, A.L., Wyatt, M.A., and Magarvey, N.A. (2015). Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res* **43**, 9645-9662.

Sladek, F.M., Munn, M.M., Rupp, W.D., and Howard-Flanders, P. (1989). *In vitro* repair of psoralen-DNA cross-links by RecA, UvrABC, and the 5'-exonuclease of DNA polymerase I. *J Biol Chem* **264**, 6755-6765.

Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E., and Tainer, J.A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92.

Stivers, J.T., and Jiang, Y.L. (2003). A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem Rev* **103**, 2729-2759.

Sugimoto, Y., Otani, T., Oie, S., Wierzba, K., and Yamada, Y. (1990). Mechanism of action of a new macromolecular antitumor antibiotic, C-1027. *J Antibiot (Tokyo)* **43**, 417-421.

Szkudelski, T. (2001). The mechanism of alloxan and streptozotocin action in B cells of the rat pancreas. *Physiol Res* **50**, 537-546.

Tamaoki, T., Shirahata, K., Iida, T., and Tomita, F. (1981). Trioxacarcins, novel antitumor antibiotics. II. Isolation, physico-chemical properties and mode of action. *J Antibiot* **34**, 1525.

Tenconi, E., and Rigali, S. (2018). Self-resistance mechanisms to DNA-damaging antitumor antibiotics in Actinobacteria. *Curr Opin Microbiol* **45**, 100-108.

Terabayashi T, and Hanada, K. (2018). Genome instability syndromes caused by impaired DNA repair and aberrant DNA damage responses. *Cell Biol Toxicol* **34**, 337-350.

Terawaki, A., and Greenberg, J. (1966). Effect of carzinophillin on bacterial deoxyribonucleic acid: formation of inter-strand cross-links in deoxyribonucleic acid and their disappearance during post-treatment incubation. *Nature* **209**, 481-484.

Thaker, M.N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A.M., Medina, R., and Wright, G.D. (2013). Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat Biotechnol* **31**, 922-927.

Thibodeaux, C.J., Chang, W.C., and Liu, H.W. (2012). Enzymatic chemistry of cyclopropane, epoxide, and aziridine biosynthesis. *Chem Rev* **112**, 1681-1709.

Thomas, L., Yang, C., and Goldthwait, D.A. (1982). Two DNA Glycosylases in *Escherichia coli* Which Release Primarily 3-Methyladenine. *Biochemistry* **21**, 1162-1169.

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.

Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem Biol* **2**, 575-579.

Tomasz, M., Lipman, R., Chowdary, D., Pawlak, J., Verdine, G.L., and Nakanishi, K. (1987). Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science* **235**, 1204-1208.

Tomicic, M., and Franekic, J. (1996). Effect of overexpression of *E. coli* 3-methyladenine-DNA glycosylase I (Tag) on survival and mutation induction in *Salmonella typhimurium*. *Mutat Res* **358**, 81-87.

Tomita, F., Tamaoki, T., Morimoto, M., and Fujimoto, K. (1981). Trioxacarcins, novel antitumor antibiotics. I. Producing organism, fermentation and biological activities. *J Antibiot* **34**, 1519-1524.

Tomlinson, J.H., Thompson, G.S., Kalverda, A.P., Zhuravleva, A., and O'Neill, A.J. (2016). A target-protection mechanism of antibiotic resistance at atomic resolution: insights into FusB-type fusidic acid resistance. *Sci Rep* **6**, 19524.

Truglio, J.J., Croteau, D.L., Van Houten, B., and Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* **106**, 233-252.

Tsuchida, T., Sawa, R., Takahashi, Y., Iinuma, H., Sawa, T., Naganawa, H., and Takeuchi, T. (1995). Azicemicins A and B, new antimicrobial agents produced by *Amycolatopsis*. *J Antibiot (Tokyo)* **48**, 1148-1152.

Tu, L.C., Melendy, T., and Beerman, T.A. (2004). DNA damage responses triggered by a highly cytotoxic monofunctional DNA alkylator, hedamycin, a pluramycin antitumor antibiotic. *Mol Cancer Ther* **3**, 577-585.

Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656.

Tuinmann, G., Hegewisch-Becker, S., Zschaber, R., Kehr, A., Schulz, J., and Hossfeld, D.K. (2004). Gemcitabine and mitomycin C in advanced pancreatic cancer: a single-institution experience. *Anti-cancer drugs* **15**, 575-579.

Turgay, K., Hahn, J., Burghoorn, J., and Dubnau, D. (1998). Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J* **17**, 6730-6738.

Tyc, O., Song, C., Dickschat, J.S., Vos, M., and Garbeva, P. (2017). The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol* **25**, 280-292.

Van Houten, B., Gamper, H., Hearst, J.E., and Sancar, A. (1988). Analysis of sequential steps of nucleotide excision repair in *Escherichia coli* using synthetic substrates containing single psoralen adducts. *J Biol Chem* **263**, 16553-16560.

Van Houten, B., Gamper, H., Holbrook, S.R., Hearst, J.E., and Sancar, A. (1986). Action mechanism of ABC excision nuclease on a DNA substrate containing a psoralen crosslink at a defined position. *Proc Natl Acad Sci U S A* **83**, 8077-8081.

Veening, J.W., and Blokesch, M. (2017). Interbacterial predation as a strategy for DNA acquisition in naturally competent bacteria. *Nat Rev Microbiol* **15**, 629.

Verhoeven, E.E., van Kesteren, M., Turner, J.J., van der Marel, G.A., van Boom, J.H., Moolenaar, G.F., and Goosen, N. (2002). The C-terminal region of *Escherichia coli* UvrC contributes to the flexibility of the UvrABC nucleotide excision repair system. *Nucleic Acids Res* **30**, 2492-2500.

Verweij, J., and Pinedo, H.M. (1990). Mitomycin C: mechanism of action, usefulness and limitations. *Anti-cancer drugs* **1**, 5-13.

Viswesh, V., Gates, K., and Sun, D. (2010). Characterization of DNA damage induced by a natural product antitumor antibiotic leinamycin in human cancer cells. *Chem Res Toxicol* **23**, 99-107.

Vizcaino, M.I., and Crawford, J.M. (2015). The Colibactin Warhead Crosslinks DNA. *Nat Chem* **7**, 411-417.

Voulgaridou, G.P., Anestopoulos, I., Franco, R., Panayiotidis, M.I., and Pappa, A. (2011). DNA damage induced by endogenous aldehydes: current state of knowledge. *Mutat Res* **711**, 13-27.

Walker, G.C. (1984). Mutagenesis and Inducible Responses to Deoxyribonucleic Acid Damage in *Escherichia coli*. *Microbiol Rev* **48**, 60-93.

Walker, S., Landovitz, R., Ding, W.D., Ellestad, G.A., and Kahne, D. (1992). Cleavage behavior of calicheamicin gamma 1 and calicheamicin T. *Proc Natl Acad Sci U S A* **89**, 4608-4612.

Wang, L., Spratt, T.E., Liu, X.-K., Hecht, S.S., Pegg, A.E., and Peterson, L.A. (1997). Pyridyloxobutyl adduct *O*6-[4-Oxo-4-(3-pyridyl)butyl]guanine is present in 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone-treated DNA and is a substrate for *O*6-alkylguanine-DNA alkyltransferase. *Chem Res Toxicol* **10**, 562-567.

Wang, S., Liu, K., Xiao, L., Yang, L., Li, H., Zhang, F., Lei, L., Li, S., Feng, X., Li, A.*, et al.* (2016). Characterization of a novel DNA glycosylase from *S. sahachiroi* involved in the reduction and repair of azinomycin B induced DNA damage. *Nucleic Acids Res* **44**, 187-197.

Warren, G.M., Wang, J., Patel, D.J., and Shuman, S. (2021). Oligomeric quaternary structure of *Escherichia coli* and *Mycobacterium smegmatis* Lhr helicases is nucleated by a novel C-terminal domain composed of five winged-helix modules. *Nucleic Acids Res* **49**, 3876-3887.

Weitzman, M.D., and Fradet-Turcotte, A. (2018). Virus DNA Replication and the Host DNA Damage Response. *Annu Rev Virol* **5**, 141-164.

Williams, J.S., Lujan, S.A., Zhou, Z.X., Burkholder, A.B., Clark, A.B., Fargo, D.C., and Kunkel, T.A. (2019). Genome-wide mutagenesis resulting from topoisomerase 1-processing of unrepaired ribonucleotides in DNA. *DNA Repair (Amst)* **84**, 102641.

Wilson, D.N., Hauryliuk, V., Atkinson, G.C., and O'Neill, A.J. (2020). Target protection as a key antibiotic resistance mechanism. *Nat Rev Microbiol* **18**, 637-648.

Wilson, M.R., Jiang, Y., Villalta, P.W., Stornetta, A., Boudreau, P.D., Carra, A., Brennan, C.A., Chun, E., Ngo, L., Samson, L.D.*, et al.* (2019). The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**.

Wood, R.D. (2010). Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ Mol Mutagen* **51**, 520-526.

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.

Xu, H., Huang, W., He, Q.L., Zhao, Z.X., Zhang, F., Wang, R., Kang, J., and Tang, G.L. (2012). Self-resistance to an antitumor antibiotic: a DNA glycosylase triggers the base-excision repair system in yatakemycin biosynthesis. *Angew Chem Int Ed Engl* **51**, 10532-10536.

Xue, M., Kim, C.S., Healy, A.R., Wernke, K.M., Wang, X., Frischling, M.C., Shine, E.E., Wang, W., Herzon, S.B., and Crawford, J.M. (2019). Structure elicidation of colibactin and its DNA cross-links. *Science*, 1-13.

Xue, M., Wernke, K.M., and Herzon, S.B. (2020). Depurination of Colibactin-Derived Interstrand Cross-Links. *Biochemistry* **59**, 892-900.

Yamanaka, K., Minko, I.G., Finkel, S.E., Goodman, M.F., and Lloyd, R.S. (2011). Role of high-fidelity *Escherichia coli* DNA polymerase I in replication bypass of a deoxyadenosine DNA-peptide cross-link. *J Bacteriol* **193**, 3815-3821.

Yan, Y., Liu, Q., Zang, X., Yuan, S., Bat-Erdene, U., Nguyen, C., Gan, J., Zhou, J., Jacobsen, S.E., and Tang, Y. (2018). Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. *Nature* **559**, 415-418.

Yang, K., Qi, L.H., Zhang, M., Hou, X.F., Pan, H.X., Tang, G.L., Wang, W., and Yuan, H. (2015). The SARP family regulator *txn9* and two-component response regulator *txn11* are key activators for trioxacarcin biosynthesis in *Streptomyces bottropensis*. *Curr Microbiol* **71**, 458-464.

Yang, W. (2003). Damage repair DNA polymerases Y. *Curr Opin Struct Biol* **13**, 23-30.

Yang, W., and Gao, Y. (2018). Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu Rev Biochem* **87**, 239-261.

Yi, C., and He, C. (2013). DNA repair by reversal of DNA damage. *Cold Spring Harb Perspect Biol* **5**, a012575.

Yin, Y., Shen, Y., Meng, S., Zhang, M., Pan, H.X., and Tang, G.L. (2020). Characterization of a membrane-bound *O*-acetyltransferase involved in trioxacarcin biosynthesis offers insights into its catalytic mechanism. *Chin J Chem* **38**, 1607-1611.

Yu, Z., Yan, B., Gao, L., Dong, C., Zhong, J., M, D.O., Nguyen, B., Seong Lee, S., Hu, X., and Liang, F. (2016). Targeted Delivery of Bleomycin: A Comprehensive Anticancer Review. *Curr Cancer Drug Targets* **16**, 509-521.

Yuan, H., Zhang, J., Cai, Y., Wu, S., Yang, K., Chan, H.C.S., Huang, W., Jin, W.B., Li, Y., Yin, Y.*, et al.* (2017). GyrI-like proteins catalyze cyclopropanoid hydrolysis to confer cellular protection. *Nat Commun* **8**, 1485.

Zein, N., Sinha, A.M., McGahren, W.J., and Ellestad, G.A. (1988). Calicheamicin gamma 1: an antitumor antibiotic that cleaves double-stranded DNA site specifically. *Science* **240**, 1198-1201.

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.

Zhang, M., Hou, X.F., Qi, L.H., Yin, Y., Li, Q., Pan, H.X., Chen, X.Y., and Tang, G.L. (2015). Biosynthesis of trioxacarcin revealing a different starter unit and complex tailoring steps for type II polyketide synthase. *Chem Sci* **6**, 3440-3447.

Zhao, C., Xia, C., Mao, Q., Forsterling, H., DeRose, E., Antholine, W.E., Subczynski, W.K., and Petering, D.H. (2002). Structures of HO(2)-Co(III)bleomycin A(2) bound to d(GAGCTC)(2) and d(GGAAGCTTCC)(2): structure-reactivity relationships of Co and Fe bleomycins. *J Inorg Biochem* **91**, 259-268.

Zhao, K., Liu, M., and Burgess, R.R. (2005). The Global Transcriptional Response of *Escherichia coli* to Induced σ^{32} Protein Involves σ^{32} Regulon Activation Followed by Inactivation and Degradation of σ^{32} *in vivo J Biol Chem* **280**, 17758-17768.

Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G.*, et al.* (2008). Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem Biol* **15**, 693-705.

Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in microbes - a review. *Nat Prod Rep* **33**, 988-1005.