STRUCTURE AND FUNCTION STUDIES OF MICROBIAL CONJUGATIVE DNA TRANSFER & GI DRUG REACTIVATION PROCESSES

Rebecca Mae Pollet

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Biochemistry and Biophysics Department in the School of Medicine.

Chapel Hill
2016

Approved by:

Matthew R Redinbo

Kevin Slep

Gary Pielak

Saskia Neher

Hengming Ke

Anthony Richardson

ABSTRACT

Rebecca Mae Pollet: Structure and Function Studies of Microbial Conjugative DNA Transfer & GI Drug Reactivation Processes
(Under the direction of Matthew R Redinbo)

Antimicrobial resistance in *Staphylococcus aureus* presents an increasing threat to human health. This resistance is often encoded on mobile plasmids, such as pSK41; however, the transfer mechanism of these plasmids is not well understood. In this study, we examine key protein features of the relaxase enzyme, NES, which initiates and terminates the transfer of the multidrug resistance plasmid pSK41. This work establishes that both a novel C-terminal domain and two loops of the NES protein, hairpin loops 1 and 2, are essential for proper DNA cleavage and religation by the full 665-residue NES protein *in vitro.* Second, we show that NES is able to bind, cleave, and religate the *oriT* sequences of non-conjugative plasmids pSK156 and pCA347. These data indicate that the conjugative relaxase *in trans* mechanism recently described for the pWBG749 family of plasmids also applies to the pSK41 family of plasmids, further heightening the potential significance of this mechanism in the horizontal transfer of staphylococcal plasmids. Finally, we use the knowledge of important NES features to design polyamide inhibitors that disrupt key protein-DNA interactions and chelator fragments that target a required coordinated metal ion. The efficacy of these inhibitors

suggests that disrupting NES function may be a viable option for disrupting conjugative plasmid transfer.

β-glucuronidase (GUS) enzymes are responsible for the severe drug toxicity associated with the chemotherapy drug irinotecan. Previous characterization of *E. coli* GUS identified inhibitors whose efficacy is dependent on interaction with an active site loop termed Loop 1. Analysis of the Human Microbiome Project sequencing data establishes a catalog of GUS sequences present in the human gastrointestinal tract. Sequences in this catalog are classified according to the presence and size of an active site loop and we identify at least one sequence from each class for further characterization. Characterization confirms the β-glucuronidase activity of seven new enzymes and establishes the ability of six GUS enzymes to process SN-38G. We also confirm the importance of pH for optimal enzyme function. This data expands current understanding of the β-glucuronidase enzyme family and provides additional GUS enzymes against which to optimize inhibitors to prevent SN-38G toxicity.

To two real life superheroes:

my grandfather, Don “Pawpaw” Bedford, and my dad, Keith Bedford.

Both of you have shown me what it means to work hard and to love your family.

ACKNOWLEDGEMENTS

It has truly taken a village to complete this work. I must begin by thanking our collaborators- Neville Firth (University of Sydney, Australia), Josh Ramsay (Curtin University, Australia), and Anthony Richardson (UNC/University of Pittsburgh) for the NES project and Raad Gharaibeh (UNC-Charlotte) and Tope Keku (UNC) for the GUS project. Their assistance in this work has added more depth than I could ever have hoped to impart alone. The University of North Carolina at Chapel Hill has provided a fantastic training environment and several labs have opened their doors to me during my time here. The labs of Gary Pielak, David Lawrence, Saskia Neher, William Zamboni, Eric Brustad, Bo Li, and Dorothy Erie have shared equipment, knowledge, and a critical eye for evaluating data far beyond what was required of both the students and professors to aid in my research. The students and staff of UNC's BBSP, IMSD, and TIBBS programs have provided valuable companionship and training throughout my graduate career. I am truly indebted to Ashalla Freeman and the IMSD family for your constant support even when I didn't know I needed it.

The Redinbo lab has been a wonderful place to work, learn, and grow friendships. It has been a pleasure to work with each of you. Bret, I have been fortunate to be able to both begin and end my time in lab working with you. We spent my entire rotation expressing and purifying protein and while it wasn't always exciting, it was an

accurate representation of the rest of my graduate career. Aadra, Bill, Mike, Sam, and Kristen, thank you for your advice on all aspects of the GUS project and your patience for the many times you have been left waiting for my data, protocol, or time. I can only begin to imagine the great things each of you are going to do and where this project will go. Julianne, our time in the Redinbo lab has been an ever-changing adventure. Thank you for your friendship and for constantly expanding the types of research to which I have been exposed.

I thank all the former members of the Redinbo machine who helped build the lab as well as these projects. I owe special recognition to Monica for her guidance and patience when I was a brand new lab member; Michael Johnson for his mentoring, friendship, and constant support from afar; Krystle for her friendship and advice on research and teaching endeavors; and Laurie for her attention to my kitties and patience as I learned crystallography.

This body of work would be much smaller without the assistance of rotation and undergraduate students that I have had the joy of working with. Edhriz and Ben, no matter how small you feel your contribution to the GUS story is, know you greatly impacted my approach to this project and it was a privilege to work with you even for a short time. James and Jeff were the first undergraduates I worked with in the Redinbo lab and we learned together while tackling the NES project. I am so grateful for the hard work both of you contributed to that project and I know that you will go far in life. James, thank you for your friendship and your continued dedication and patience as we slogged through the messy details of GUS enzymes that continues to inform my current work. Erin, Loraine, Kristen, and Rohini: I am so grateful for your hard work and patience on

assays that didn't always replicate the way we expected. You have pushed me to clarify my understanding of assays, our enzymes, and to refine my mentorship skills. Finally, Emma - this work would not have happened without you. You changed the course of my research and I cannot begin to thank you for the countless hours you spent in lab. But more than that, I am so thankful for your friendship; you gave me the true Carolina student experience and always have an open ear to talk about science or our families. Emory is not ready for you.

I am incredibly grateful to Matt for the opportunity to work in his laboratory. Despite our jokes about your time on sabbatical, you have provided support, advice, encouragement, and, perhaps most important, a great example of how to run a successful laboratory. I greatly respect your love of science and drive to do something of impact rather than what is easy. You are not just my research advisor, you are a mentor for my life and career.

My work in the lab has been constantly supported by my friends and family outside the lab. To the friends I've met through the Beta Chi chapter of Kappa Delta, Triangle KD alumnae chapter, that I've stayed in contact with from Muldrow, TU, and DBS, and that I've met through my work at UNC, please know your interest in my work and patience in my crazy schedule has been greatly appreciated. Finally, I thank my family for their patience, encouragement, love, and support. I thank my grandparents, aunts, uncles, and cousins for understanding when I'm not at family events and being proud to tell people you know that I do science research. Mom and Dad, thank you for always believing in me. I hope my accomplishments can begin to thank you for all of your hard work to get me here. Bailey, you'll be here soon enough! Thank you for

understanding my love of science and never hesitating to ask about new research you read about. Cynthia deserves special thanks for going to coffee shops with me to work, learning enough science to understand what I'm talking about, and always making sure I'm having fun. I'm not sure this dissertation would have been written without your support. Finally, and most importantly, I thank my husband Lucas for his patience, support, and his efforts to learn about my work. Thank you for going to lab to start overnights and telling me when it's time to stop working. I like doing life with you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

%	percent
°C	degrees Celsius
±	plus or minus
<	less than
>	greater than
≥	greater than or equal to
6FAM/6-FAM/FAM	fluorescein
A	adenine
Å	angstrom
ACIDO	Acidobacteria *Terriglobus roseus* β-glucuronidase
ACTIN	Actinobacteria *Corynebacterium massiliense* β-glucuronidase
AI	arabinose inducible
Ala	alanine
Ap/Ap^R	ampicillin/ampicillin resistance
APS	ammonium persulfate
ArchIa	Archaea *Ignisphaera aggregans* β-glucuronidase
B_dorei	*Bacteroides dorei* β-glucuronidase
B_fragilis	*Bacteroides fragilis* β-glucuronidase
B_ovatus	*Bacteroides ovatus* β-glucuronidase
B_uniformis	*Bacteroides uniformis* β-glucuronidase
B. dorei	*Bacteroides dorei*

B. fragilis	*Bacteroides fragilis*
B. ovatus	*Bacteroides ovatus*
B. uniformis	*Bacteroides uniformis*
BactNk	Bacteroides *Niastella koreensis* β-glucuronidase
*Bd*GUS	*Bacteroides dorei* β-glucuronidase
*Bf*GUS	*Bacteroides fragilis* β-glucuronidase
BHI	Brain Heart Infusion Broth
BLASTN	Basic Local Alignment Search Tool for Nucleotide
BLASTp	Basic Local Alignment Search Tool for Protein
*Bo*GUS	*Bacteroides ovatus* β-glucuronidase
BovineGUS	*Bos taurus* β-glucuronidase
BSA	bovine serum albumin
*Bt*GUS	*Bos taurus* β-glucuronidase
*Bu*GUS	*Bacteroides uniformis* β-glucuronidase
C_perfringens	*Clostridium perfringens* β-glucuronidase
C-terminal, C-term	carboxyl terminal
C, Cyt	cytosine
C. perfringens	*Clostridium perfringens*
$CaCl_2$	calcium chloride
CDC	Center for Disease Control
CFL	chelator fragment library
Cl	chloride
cm	centimeter

Cm/Cm^R	chloramphenicol/chloramphenicol resistance
CPD	cysteine protease domain
*Cp*GUS	*Clostridium perfringens* β-glucuronidase
CPT	conjugative plasmid transfer
CPT-11	irinotecan, 7-ethyl-10-[4-(1-piperidino)-1-piperidino]
CRISPR	clustered regularly interspaced short palindromic repeats
CV	column volume
DICTY	Dictyoglomi *Dictyoglomus turgidum* β-glucuronidase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
DTT	D,L-dithiothreitol
E	glutamic acid
E_coli	*Escherichia coli* β-glucuronidase
E_eligens	*Eubacterium eligens* β-glucuronidase
E. coli	*Escherichia coli*
E. eligens	*Eubacterium eligens*
E. lenta	*Eggerthella lenta*
*Ec*GUS	*Escherichia coli* β-glucuronidase
EDTA	Ethylenediaminetetraacetic acid
*Ee*GUS	*Eubacterium eligens* β-glucuronidase
Em	emission
EMSA	electrophoretic mobility shift assay

Ex	excitation
f	average fluorescence anisotropy signal
F_prausnitzii	*Faecalibacterium prausnitzii* β-glucuronidase
F. prausnitzii	*Faecalibacterium prausnitzii*
FBDD	fragment-based drug design
FDA	United States Food and Drug Administration
FOLFIRI	folinic acid, fluorouracil, irinotecan
FOLFIRINOX	folinic acid, fluorouracil, irinotecan, oxaliplatin
*Fp*GUS	*Faecalibacterium prausnitzii* β-glucuronidase
Fs^R	fusidic acid resistance
Fu	fluorescence unit(s)
FUNGI	Eukaryota *Aspergillus niger* β-glucuronidase
g	g-force, gravitational acceleration
G, Gly	glycine
G, Gua	guanine
GH2	glycoside hydrolase family 2
GI	gastrointestinal
Gm/Gm^R	gentamycin/gentamycin resistance
GUS	β-glucuronidase
h	hour(s)
H_sapiens	*Homo sapiens* β-glucuronidase
H, His	histidine
H. sapiens	*Homo sapiens*

H^+	hydrogen ion
H_2O	water
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGT	horizontal gene transfer
His_6	6 x histidine affinity tag
HMGC	HMP Clustered Gene Indices
HMGI	HMP Gene Indices
HMP	Human Microbiome Project
HPLC	high-performance liquid chromatography
*Hs*GUS	*Homo sapiens* β-glucuronidase
HTS	high-throughput screen
HUH	histidine residue-hydrophobic residue-histidine residue
I.V.	intravenous
Im	imidazole
Inh 1	Inhibitor 1
$InsP_6$	inositol hexakisphosphate
IPTG	isopropyl-β-D-thiogalactopyranoside
IR2	inverted repeats of the pWBG749 family *oriT*
IRI	irinotecan
K	K_D
K, Lys	lysine
k_{cat}	rate of enzyme turnover

k_{cat}/K_M	catalytic efficiency
KCl	potassium chloride
K_D	dissociation constant
KH_2PO_4	potassium phosphate
K_i	dissociation constant for the inhibitor
K_M	Michaelis-Menten constant
Km^R	kanamycin resistance
KO	knock out
L	liter
L-DOPA	levodopa, L-3,4-dihydroxyphenylalanine
L, Leu	leucine
L. rhamnosus	*Lactobacillus rhamnosus*
L1	Loop 1
L2	Loop 2
LB	lysogeny broth
LIC	ligation independent cloning
*Lr*GUS	*Lactobacillus rhamnosus* β-glucuronidase
M	methionine
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight
max	average fluorescence anisotropy signal of sample at saturating concentration of protein
MBP	maltose binding protein
mg	milligram
$MgCl_2$	magnesium chloride

MIC	minimum inhibitory concentration
min	minute
min	average fluorescence anisotropy signal of no protein control
mL	milliliter
mL1	Mini Loop 1
mL1mL2	Mini Loop 1 Mini Loop 2
mL2	Mini Loop 2
mM	millimolar
MOB	mobility enzymes
mob	mobility genes
MRSA	methicillin resistant *Staphylococcus aureus*
MSA	multiple sequence alignment
MSP	multiple species
n	the total number of reads in that HMP sequencing sample
N	the total number of HMP sequencing samples
N	total number of GUS sequences in each individual
N-terminal	amino terminal
N, Asn	asparagine
NaCl	sodium chloride
NAD^+	nicotinamide adenine dinucleotide- oxidized form
NADH	nicotinamide adenine dinucleotide- reduced form
Nb/Nb^R	novobiocin/novobiocin resistance
NCBI	National Center for Biotechnology Information

NES	nicking enzyme of *Staphylococcus*
nes	nicking enzyme of *Staphylococcus* gene
Ni	nickel
ni	no inhibition
Ni^{2+}	nickel cation
NIH	National Institutes of Health
nL	No Loop
nM	nanomolar
nm	nanometer
Nm^R	neomycin resistance
NSAID	nonsteroidal anti-inflammatory drug
NxKG	asparagine residue, any residue, lysine residue, glycine residue
OD_{600}	optical density at 600 nm wavelength
oligos	oligonucleotides
Orf	gene open reading frame
oriT	origin of transfer
p	p-value
P_merdae	*Parabacteroides merdae* β-glucuronidase
P, Pro	proline
P. merdae	*Parabacteroides merdae*
PAGE	polyacrylamide gel
PCR	polymerase chain reaction
PDB	Protein Data Bank

PDB ID	PDB identification code
PEG	polyethylene glycol
pH	negative log (base 10) of the molar concentration of hydronium ions
Phe	phenylalanine
*Pm*GUS	*Parabacteroides merdae* β-glucuronidase
PNP	*p*-nitrophenyl
PNPG	*p*-nitrophenyl glucuronide
PROT	Proteobacteria *Vibrio harveyi* β-glucuronidase
Py	pyrrole
R	arginine
RC	read count for a gene in a particular HMP sequencing sample
RCSB	Research Collaboratory for Structural Bioinformatics
Rf^R	rifampin resistance
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
S_agalactiae	*Streptococcus agalactiae* β-glucuronidase
s^{-1}	per second
$s^{-1}mM^{-1}$	per second per millimolar
S. agalactiae	*Streptococcus agalactiae*
S. aureus	*Staphylococcus aureus*
*Sa*GUS	*Streptococcus agalactiae* β-glucuronidase
SDS	sodium dodecyl sulfate
SEM	standard error of mean

Ser	serine
∑x	the total number of reads in all HMP sequencing samples
Sm/Sm^R	streptomycin/streptomycin resistance
SN-38	7-ethyl-10-hydroxyl-camptothecin
SN-38-glucuronide	7-ethyl-10-hydroxyl-camptothecin glucuronide
SN-38G	7-ethyl-10-hydroxyl-camptothecin glucuronide
S_N2	bimolecular nucleophilic substitution
sp.	species
T	Thymine
T	total DNA concentration
T-strand	transfer strand
T4CP	type 4 coupling protein
T4SS	type four secretion system
TAE	tris base, acetic acid, EDTA
TBE	tris, borate, EDTA
Tb^R	tobramycin resistance
TCEP	tris(2-carboxyethyl)phosphine
TEMED	tetramethylethylenediamine
TherTn	Thermatogae *Thermotoga naphthophila* β-glucuronidase
T_m	melting temperature
tra	conjugative transfer gene
Tris	tris(hydroxymethyl)aminomethane
Udh	uronate dehydrogenase enzyme

UDP	uridine 5'-diphosphate
UGT	UDP-glucuronosyltransferase
UNC	University of North Carolina at Chapel Hill
UNC-CH	University of North Carolina at Chapel Hill
UV	ultraviolet
vol/vol, v/v	volume per volume
VRSA	vancomycin resistant *S. aureus*
vs	versus
W	tryptophan
WT	wild type
x	total protein concentration
Y, Tyr	tyrosine
α	alpha
β	beta
Δ	deletion
ΔL1	Hairpin Loop 1 deletion
ΔL1ΔL2	Hairpin Loop 1 Loop 2 deletion
ΔL2	Hairpin Loop 2 deletion
λ	wavelength
µg	microgram
µL	microliter
µM	micromolar
µm	micrometer

CHAPTER 1: INTRODUCTION TO CONJUGATIVE PLASMID TRANSFER OF pSKS41 IN *STAPHYLOCOCCUS AUREUS*

Staphylococcus aureus and Antibiotic Resistance

Staphylococcus aureus is a gram-positive bacterium that is found as a commensal in the nose, respiratory tract, gastrointestinal tract, and on the skin of humans (1). Despite this potentially beneficial role in the human microbiota, *S. aureus* is best known as a pathogen that causes skin and respiratory infections. While these infections are normally treatable, infections resistant to antibiotic treatment can enter the bloodstream and cause sepsis and death. Since the introduction of penicillin and other antibiotic compounds, resistant strains of *S. aureus* have been appearing in human infections. It took just two years after the introduction of methicillin in the clinic for methicillin-resistant *Staphylococcus* to begin to present a barrier to treatment (2). In 2011, methicillin-resistant *S. aureus* (MRSA) caused an estimated 80,461 infections in the United States leading to 11,285 deaths (2). Staphylococcal bacteria are a leading cause of healthcare-associated infections leading the Center for Disease Control (CDC) to classify MRSA as a serious threat (2).

In addition to methicillin, *S. aureus* has been found to be resistant to a wide variety of compounds including antiseptics, antibiotics, and disinfectants. Antibiotic

resistances found in *S. aureus* can include oxacillin, erythromycin, levofloxacin, clindamycin, mupirocin, doxycycline, gentamicin, trimethoprim-sulfamethoxazole, and tetracycline, which when found in the same strain can leave only harsh compounds available for treatment (3). It has also been shown that MRSA can obtain resistance to vancomycin, the current front-line treatment, leading to infections that are resistant to almost all approved antibiotics (4-7). Despite the rise of antibiotic resistance in *S. aureus* and other bacteria, the number of new antibiotics approved by the FDA has consistently decreased over the past two decades.

Even as new antibiotics are introduced, it is expected that resistance will arise quickly as seen in the past. Most antibiotics to date have been discovered as or derived from small molecules produced by bacteria or plants. In these natural environments, these small molecules function as bacteriostatic or bacteriotoxic agents in order to regulate the surrounding bacterial community, much like how they function as antibiotics relevant to human health. This means the bacteria producing the compound or any bacteria that persist in this natural environment must maintain a resistance element to avoid toxicity from the compound. In fact, environmental bacteria including those preserved for hundreds of years in permafrost, isolated caves, and ancient human specimens contain resistance genes (8). Because resistance to a compound similar to any new antibiotics almost certainly already exists in nature, it is likely only a matter of time until that resistance element moves into pathogenic bacteria.

Due to both the currently existing antibiotic resistance and the high potential for the rise of future resistances, this work set out to better understand the

mechanisms by which antibiotic resistance is spread between bacteria. We hope that characterization of these processes will lead to drugs that specifically target antibiotic resistant bacterial communities or prevent the spread of antibiotic resistance within bacterial communities.

Conjugative Plasmid Transfer

Bacteria become resistant to the effects of antibiotics and other small molecules by obtaining genetic elements that encode for proteins that allow the bacteria to breakdown or transport the lethal substance out of their cell or encode modified targets with reduced antibiotic drug-reactivity. These resistance genetic elements can be incorporated into the chromosome of a bacterium or may reside on extra-chromosomal elements such as gene cassettes, transposons, bacteriophages, and single- or double-stranded DNA plasmids. Bacteria can transfer their genetic material, including these resistance elements, through two routes: vertical or horizontal gene transfer. Vertical gene transfer takes place between a mother and her daughter cells via duplication of the chromosomal and extra-chromosomal genetic material and distribution of a complete copy to each daughter cell. Horizontal gene transfer (HGT) is the transmission between mature, parent bacteria and involves only extra-chromosomal elements. HGT takes many forms including scavenging of DNA by naturally competent bacterial cells, bacteriophage facilitated transfer, and regulated transfer for single- and double-stranded plasmids.

Most of the resistance genetic elements in *S. aureus* are found on double-stranded DNA (dsDNA) plasmids or are incorporated into the chromosome after

movement through a dsDNA plasmid; therefore, this work will focus on the transfer mechanism of these plasmids between bacterium. Double-stranded DNA plasmids can be beneficial to the larger bacterial community as many of these plasmids are readily involved in HGT, allowing an entire population of bacteria to quickly acquire antibiotic resistance (9). HGT of dsDNA plasmids is called conjugative plasmid transfer (CPT). During CPT, a donor bacterium transfers one strand of a double-stranded DNA plasmid to a recipient bacterium (Figure 1.1).

There are several sets of machinery required for this process that may be expressed on the plasmid being transferred or on a co-resident plasmid in the donor bacterium. This machinery includes the mobility (MOB) set of enzymes that is responsible for the replication of the plasmid as well as genes for a type 4 secretion system (T4SS) and type 4 coupling protein (T4CP) that facilitate transfer of the plasmid (10). The MOB set of proteins forms a large complex called the relaxosome. The relaxosome complex includes the highly conserved relaxase (NES in Figure 1.1) as well as the replication machinery of the plasmid. The T4SS and T4CP form a secretion system through which DNA travels to the recipient cell (blue shaded box in Figure 1.1).

After relaxosome formation, the relaxase initiates CPT by binding to the *origin of transfer* (*oriT*) of the plasmid and creating a single-stranded break at the *nic* site in the strand to be transferred (T-strand, red in Figure 1.1) (9, 11). A DNA helicase in the relaxosome then separates the T-strand from the parent strand, beginning at the *nic* site. The free T-strand then begins to travel into the recipient cell through the T4CP/T4SS complex. The exact mechanisms involved in this process are unclear;

some data indicate that the covalently bound relaxase also travels into the recipient bacterium where the relaxase can then ligate the T-strand back together (12). A DNA polymerase within the donor bacterium replaces the T-strand as it is transferred and a DNA polymerase within the recipient bacterium synthesizes a copy of the T-strand so that the process culminates with two bacteria each containing a full, double-stranded plasmid that can confer antibiotic resistance.

Types of Plasmids

Plasmids that undergo CPT are divided into three classes based on the presence or absence of the MOB, T4CP, and T4SS gene clusters (Figure 1.2). The best studied type of plasmid is the conjugative plasmid which encodes all of the machinery needed to replicate and transmit itself including the MOB, T4CP, and T4SS clusters (Figure 1.2, 10). Mobilizable plasmids do not encode all of the machinery needed for their transmission; they express only the MOB proteins required for replication but do not encode their own T4SS and therefore must use the conjugation machinery expressed by a conjugative plasmid co-resident in the same bacterium (10). A small number of mobilizable plasmids encode their own T4CP but most utilize the T4CP from the co-resident plasmid expressing the T4SS.

There is also a third type of plasmid that does not encode its own replication or conjugative machinery. Until recently it was assumed these plasmids were transferred via mechanisms besides conjugative plasmid transfer such as bacteriophage. However, O'Brien *et al.* showed that these plasmids can partake in CPT via a relaxase-*in trans* conjugative transfer mechanism in which they take

advantage of both the replication (MOB) and conjugation (T4CP/T4SS) machinery of a co-resident plasmid (Figure 1.2,13, 14). For this relaxase-*in trans* transfer, it is important that the MOB proteins of a conjugative or mobilizable plasmid co-resident in the bacterium match an *oriT* found on the *in trans* plasmid. Therefore, understanding of MOB proteins is important for characterization of CPT of all three plasmid classes.

The relaxase is the only highly conserved member of the MOB family of proteins and is used to identify and classify conjugative and mobilizable plasmids (Figure 1.2). A relaxase enzyme is defined by the presence of two motifs (15). The first motif is the Y motif which consists of one or more tyrosines that are responsible for plasmid nicking and ligation. This conserved tyrosine (Y25 in Figure 1.3) initiates a nucleophilic substitution-type (S_N2) attack on the scissile phosphate linking two nucleotides at the *nic* site (16). This generates a free 3' hydroxyl and a 5' covalent phosphotyrosine bond. The reaction is reversible with the 3' hydroxyl acting as a nucleophile to ligate the strand together in the recipient cell.

The second motif, the HUH motif, consists of a histidine residue, hydrophobic residue, and histidine residue and is the mostly highly conserved relaxase motif. Often a third histidine is found slightly upstream from the HUH motif and forms a triad responsible for coordinating a metal cation in the relaxase active site (H123, H131, and H133 in Figure 1.3). This divalent metal ion is required for DNA cleavage, and several metals including nickel, manganese, zinc, and copper are sufficient for proper function (16-18). As shown in Figure 1.3, the metal coordinates one oxygen of the scissile phosphate at which the DNA cleavage will occur (16, 19).

Based on the sequence of their respective relaxase and in particular, the sequence in and around the Y and HUH motifs, conjugative and mobilizable plasmids have been classified into six families. These families are MOB_F, MOB_H, MOB_C, MOB_Q, MOB_P, and MOB_V (Figure 1.2, 20, 21). The relaxases of the MOB_H and MOB_C families have a different architecture and may not contain regions that resemble the Y and HUH motifs (15, 22). The MOB_F family is the best characterized of the six families and contains relaxases with two to three tyrosines in the Y motif (15). The MOB_Q, MOB_P, and MOB_V plasmid families encode relaxases containing only one tyrosine in the Y motif (15). Because the relaxase is responsible for the initiation and termination of CPT, understanding of these proteins could lead to development of novel antibiotics targeting resistance-spreading bacterial infections. Indeed, it has already been shown that inhibiting the relaxase can disrupt propagation of the host plasmid (23).

Introduction to *S. aureus* plasmid pSK41

pSK41 is a 46,445 nucleotide, conjugative, multiresistance plasmid from *S. aureus* (24). Related plasmids were first detected in the mid-1970s; the transfer region of pSK41 was analyzed by Firth and colleagues in 1993 and the complete sequence of pSK41 was published in 1998 (24, 25). pSK41 is now considered the prototype plasmid for this family.

pSK41 family plasmids have been found to harbor resistance determinants for antiseptics, disinfectants, and many antibiotics including tetracycline, gentamicin, tobramycin, kanamycin, neomycin, paromomycin, quaternary ammonium

compounds, trimethoprim, mupirocin, macrolides, lincosamides, streptogramins, bleomycin, and β-lactams (3, 24-26). Of special note is the incorporation of vancomycin resistance into pSK41 to form the new plasmid pLW1043 (4). pSK41 has also been shown to contribute to the spread of antibiotic resistance through facilitating the transfer of mobilizable plasmids such as pC221 and pSK639 (27, 28). Because pSK41 and related plasmids confer such a wide range of resistances, it is an important target for understanding and modulating the spread of antibiotic resistance in *S. aureus.*

Introduction to Nicking Enzyme of *Staphylococcus* (NES)

The relaxase of pSK41 is named nicking enzyme of staphylococci (NES). This gene and its *oriT* target were first identified on the pGO1 plasmid by Climo *et al.* in 1996 (29). The *nes* gene was confirmed in pSK41 by Berg and colleagues (24). NES is a single-tyrosine relaxase in the MOB_Q family and is considered the prototype relaxase for this family.

Further characterization of the NES protein was carried out by the Redinbo lab in order to characterize the catalytically active tyrosine (residue 25) and *nic* site of the *oriT* (Figure 1.3,17). Edwards *et al*. also solved the crystal structure of the N-terminal relaxase domain of NES bound to its *oriT* DNA (17). This was the first structure of a single tyrosine relaxase bound to its cognate DNA. Previously solved crystal structures of relaxases in complex with their *oriT* DNA showed that the DNA bound in a cleft next to a “thumb” motif of the protein that then closed over the DNA to hold the DNA in place (30). NES, however, does not contain a thumb motif and

instead has two unique features to properly position the *oriT* DNA (17). The first feature is two protein loops (loop 1 and loop 2 in Figure 1.4A) that extend into the major and minor groove of the DNA hairpin formed by the *oriT* DNA. The second feature is nucleotide guanine-26 (magenta in Figure 1.4A) of the *oriT* that is flipped in the opposite direction from the surrounding nucleotides and makes significant contacts with amino acid side chains of NES.

Relaxases from the MOB_Q, MOB_P, and MOB_V families including NES are two domain proteins with the N-terminal domain containing the relaxase motif. The C-terminal domain is variable and often functions as a helicase, primase, or DNA-binding domain that contributes to the DNA replication process (15). Interestingly, the NES C-terminal domain shows no similarity to previously characterized C-terminal domains of relaxases or other relaxosome proteins. The C-terminal domain of NES was also crystallized and its structure solved; however, both this C-terminal domain structure and the SAXS envelope of the full-length protein did not give any suggestion of the potential function of this domain (Figure 1.4B,17).

Biochemical assays for DNA binding and cleavage conducted with the relaxase domain established that the protein loops 1 and 2 are important for proper activity and explored the importance of variation and length of the *oriT* (17). In addition, *S. aureus* conjugation assays showed that both the C-terminal domain and loop 1 and loop 2 of the relaxase are essential for proper conjugation of pSK41 (17). Therefore, this work characterizes at what point in CPT these protein features play an important role in NES function.

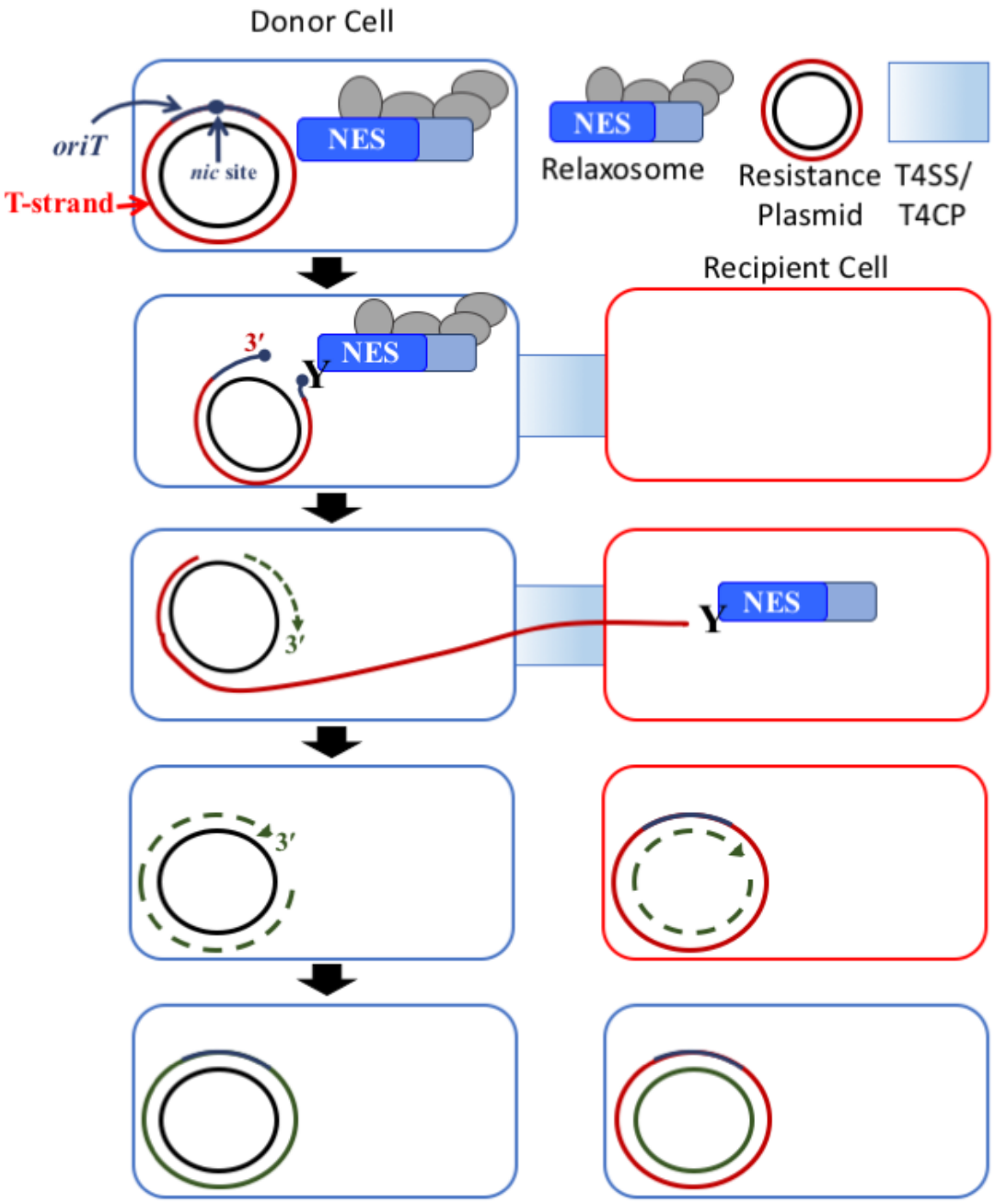


Figure 1.1. Mechanism of Conjugative Plasmid Transfer
Conjugative Plasmid Transfer (CPT) is initiated when the relaxase (NES) in the relaxosome complex binds to the *oriT* (blue) on the transfer strand (red, T-strand) in order to cleave the plasmids at the nic site (blue dot). The donor cell (blue) and recipient cell (red) are then connected via the type 4 secretion system (T4SS) and Type 4 coupling protein (T4CP) to allow the relaxase (NES) and T-strand to move into the recipient cell. Each of the original strands of the resistance plasmid is replicated such that both the donor and recipient cell end with a complete double stranded plasmid.

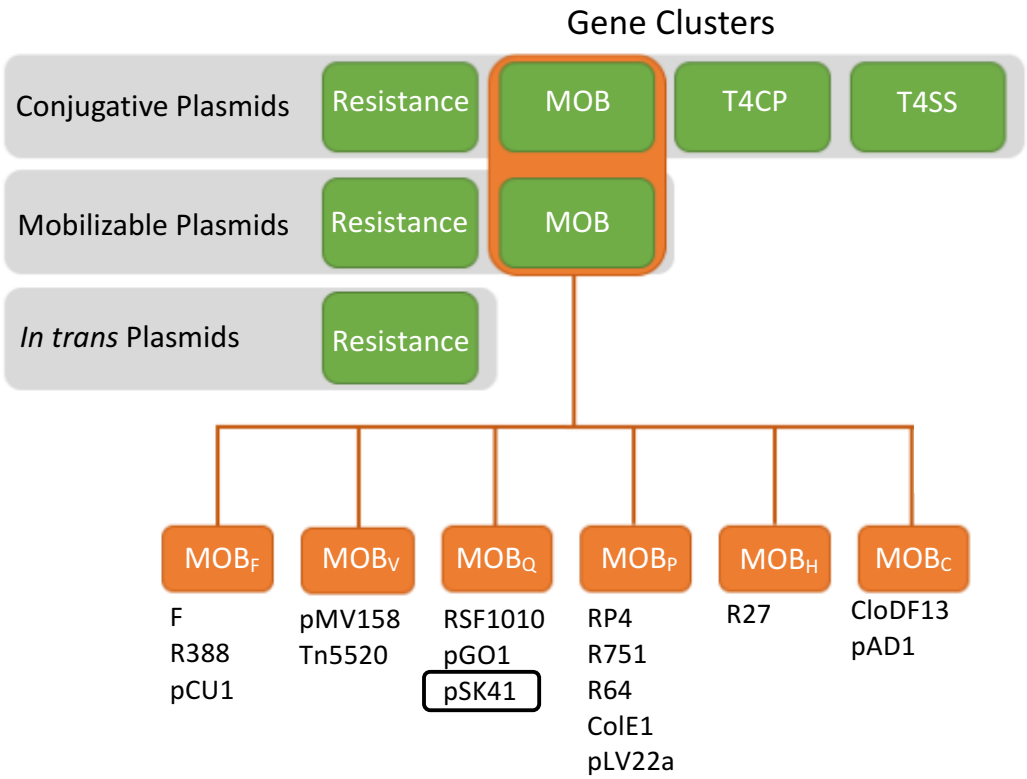


Figure 1.2. Classification of Plasmids and MOB Genes
Bacterial plasmids are classified according to the genes encoded on the plasmid. Conjugative plasmids encode resistance elements, mobilization (MOB) genes, a type 4 coupling protein (T4CP), and a type 4 secretion system (T4SS). Mobilizable plasmids encode only resistance elements and MOB genes while *in trans* plasmids encode on resistance elements. The MOB genes on conjugative and mobilizable plasmids are in turn classified as MOB_F, MOB_H, MOB_C, MOB_Q, MOB_P, AND MOB_V according to the motifs found in their relaxase gene.

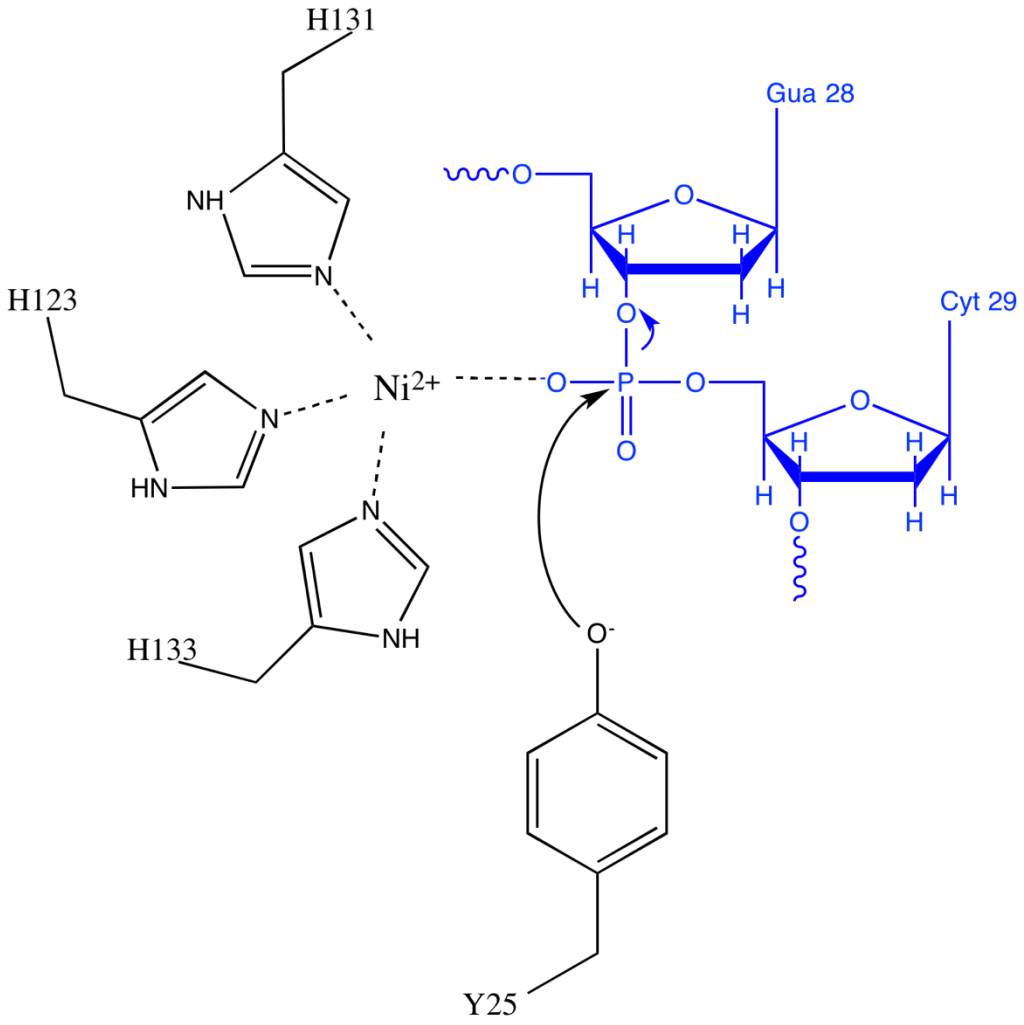


Figure 1.3. Proposed Mechanism of DNA Cleavage by NES
The tyrosine residue of the relaxase (Y25 in NES) initiates a nucleophilic substitution-type attack on the scissile phosphate of the *nic site* of the transfer strand of the plasmid to be transferred. In pSK41 this is between a guanine and cytosine (nucleotides 28 and 29 in the oligonucleotides used throughout this work). This attack creates a free 3' hydroxyl and the 5' phosphate becomes covalently attached to the tyrosine of the relaxase. The metal coordinated by the HUH motif (H123, H131, H133) is shown as Ni^{2+} here, as has been shown in purified NES; however, other divalent metals are likely the predominant cofactor *in vivo* for NES and other relaxases.

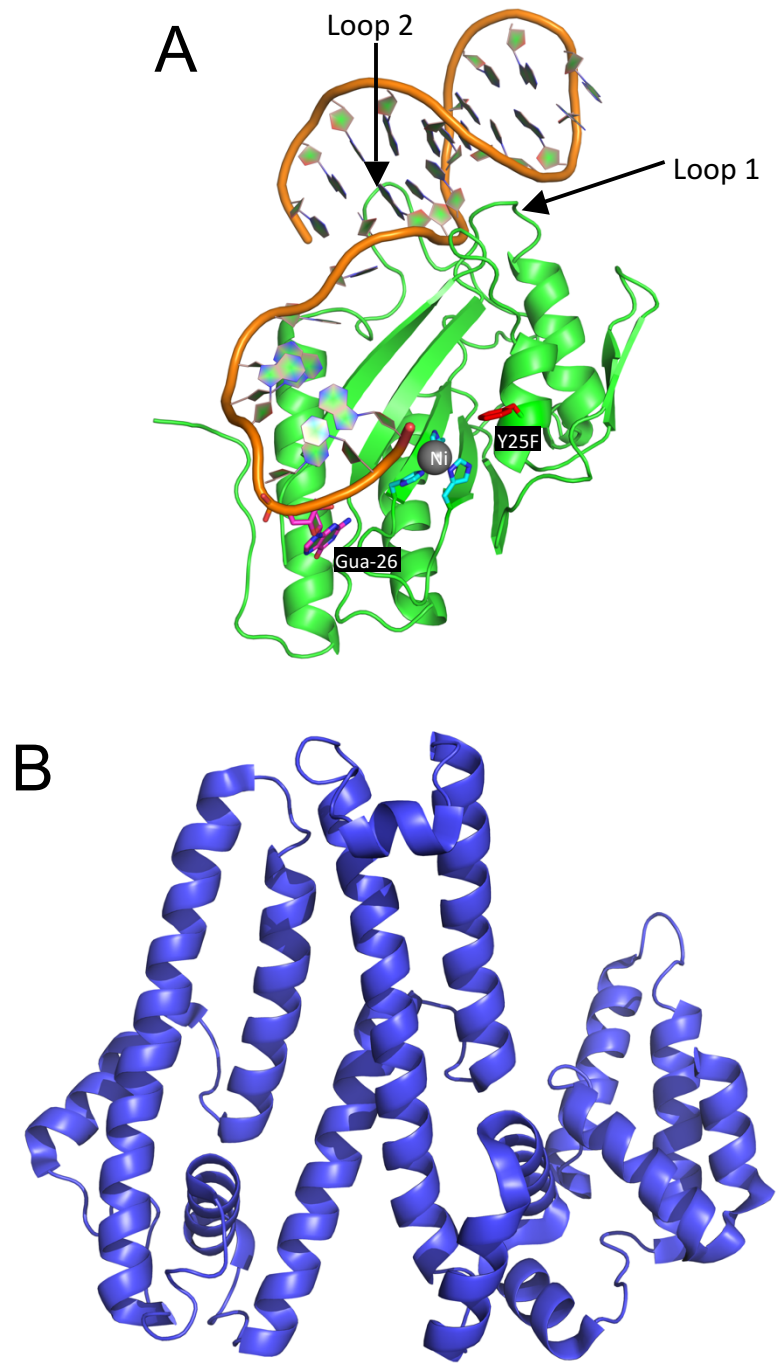


Figure 1.4. NES Structure from *S. aureus* pSK41
A. 2.9 Å crystal structure of the NES relaxase domain (residues 1-195) in complex with a 30-nucleotide DNA sequence. PDB ID: 4HT4
B. 3.0 Å crystal structure of the NES C-terminal domain (residues 254-593). PDB ID: 4HTE

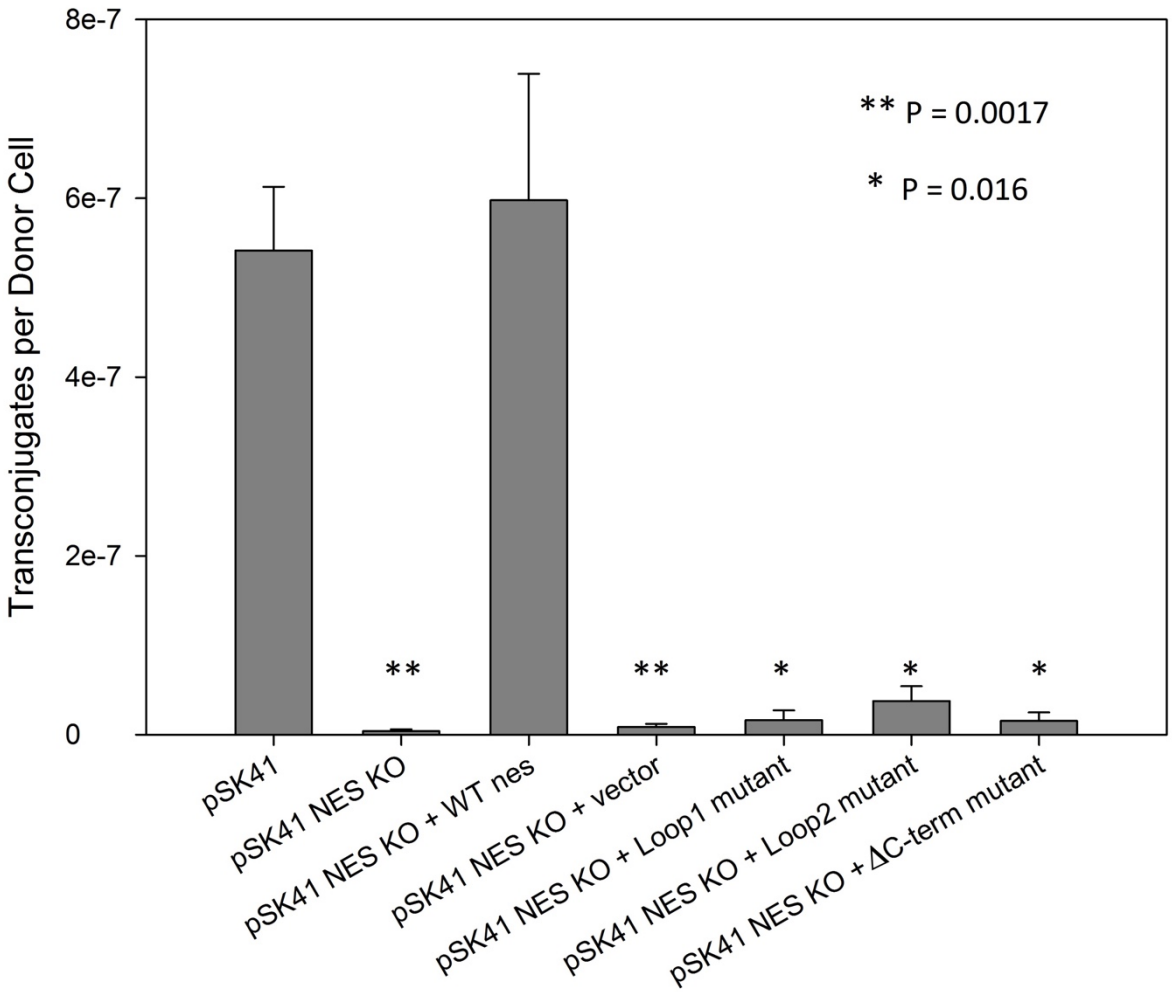


Figure 1.5. Effect of NES Mutants on Conjugation in *S aureus*
Conjugation of pSK41 in *S. aureus* and the effect of *nes* deletion (KO) or complementation with WT and designed NES loop 1, loop 2, and C-terminal deletion (ΔC-term) mutants. Originally published in Edwards *et al.* 2013 (17).

REFERENCES

1. **Brown AF**, **Leech JM**, **Rogers TR**, **McLoughlin RM**. 2014. Staphylococcus aureus Colonization: Modulation of Host Immune Response and Impact on Human Vaccine Design. Front Immunol **4**:507.

2. **Centers for Disease Control and Prevention (CDC)**. 2013. Antibiotic Resistance Threats in the United States, 2013. cdcgov.

3. **McDougal LK**, **Fosheim GE**, **Nicholson A**, **Bulens SN**, **Limbago BM**, **Shearer JES**, **Summers AO**, **Patel JB**. 2010. Emergence of resistance among USA300 methicillin-resistant Staphylococcus aureus isolates causing invasive disease in the United States. Antimicrobial Agents and Chemotherapy **54**:3804–3811.

4. **Chang S**, **Sievert DM**, **Hageman JC**, **Boulton ML**, **Tenover FC**, **Downes FP**, **Shah S**, **Rudrik JT**, **Pupp GR**, **Brown WJ**, **Cardo D**, **Fridkin SK**, **Vancomycin-Resistant Staphylococcus aureus Investigative Team**. 2003. Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. N Engl J Med **348**:1342–1347.

5. **Sievert DM**, **Rudrik JT**, **Patel JB**, **McDonald LC**, **Wilkins MJ**, **Hageman JC**. 2008. Vancomycin-resistant Staphylococcus aureus in the United States, 2002-2006. Clin Infect Dis **46**:668–674.

6. **Weigel LM**, **Clewell DB**, **Gill SR**, **Clark NC**, **McDougal LK**, **Flannagan SE**, **Kolonay JF**, **Shetty J**, **Killgore GE**, **Tenover FC**. 2003. Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. Science **302**:1569–1571.

7. **Zhu W**, **Clark N**, **Patel JB**. 2013. pSK41-Like Plasmid Is Necessary for Inc18-Like *vanA* Plasmid Transfer from Enterococcus faecalis to Staphylococcus aureus *In Vitro*. Antimicrobial Agents and Chemotherapy **57**:212.

8. **Perry J**, **Waglechner N**, **Wright G**. 2016. The Prehistory of Antibiotic Resistance. Cold Spring Harb Perspect Med **6**.

9. **Chen I**, **Christie PJ**, **Dubnau D**. 2005. The ins and outs of DNA transfer in bacteria. Science **310**:1456–1460.

10. **Smillie C**, **Garcillán-Barcia MP**, **Francia MV**, **Rocha EPC**, **la Cruz de F**. 2010. Mobility of plasmids. Microbiol Mol Biol Rev **74**:434–452.

11. **Grohmann E**, **Muth G**, **Espinosa M**. 2003. Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev **67**:277–301.

12. **Dostál L**, **Shao S**, **Schildbach JF**. 2011. Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Research **39**:2658–2670.

13. **O'Brien FG**, **Ramsay JP**, **Monecke S**, **Coombs GW**, **Robinson OJ**, **Htet Z**, **Alshaikh FAM**, **Grubb WB**. 2014. Staphylococcus aureus plasmids without mobilization genes are mobilized by a novel conjugative plasmid from community isolates. Journal of Antimicrobial Chemotherapy **70**:649–652.

14. **O'Brien FG**, **Eto KY**, **Murphy RJT**, **Fairhurst HM**, **Coombs GW**, **Grubb WB**, **Ramsay JP**. 2015. Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in Staphylococcus aureus. Nucleic Acids Research **43**:7971–7983.

15. **Chandler M**, **la Cruz de F**, **Dyda F**, **Hickman AB**, **Moncalián G**, **Ton-Hoang B**. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol **11**:525–538.

16. **Boer R**, **Russi S**, **Guasch A**, **Lucas M**, **Blanco AG**, **Pérez-Luque R**, **Coll M**, **la Cruz de F**. 2006. Unveiling the molecular mechanism of a conjugative relaxase: The structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. Journal of Molecular Biology **358**:857–869.

17. **Edwards JS**, **Betts L**, **Frazier ML**, **Pollet RM**, **Kwong SM**, **Walton WG**, **Ballentine WK**, **Huang JJ**, **Habibi S**, **Del Campo M**, **Meier JL**, **Dervan PB**, **Firth N**, **Redinbo MR**. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **110**:2804–2809.

18. **Xia S**, **Robertus JD**. 2009. Effect of divalent ions on the minimal relaxase domain of MobA. Arch Biochem Biophys **488**:42–47.

19. **Larkin C**, **Haft RJF**, **Harley MJ**, **Traxler B**, **Schildbach JF**. 2007. Roles of active site residues and the HUH motif of the F plasmid TraI relaxase. J Biol Chem **282**:33707–33713.

20. **Francia MV**, **Varsaki A**, **Garcillán-Barcia MP**, **Latorre A**, **Drainas C**, **la Cruz de F**. 2004. A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol Rev **28**:79–100.

21. **Garcillán-Barcia MP**, **Francia MV**, **la Cruz de F**. 2009. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev **33**:657–687.

22. **Francia MV**, **Clewell DB**, **la Cruz de F**, **Moncalián G**. 2013. Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America **110**:13606–13611.

23. **Lujan SA**, **Guogas LM**, **Ragonese H**, **Matson SW**, **Redinbo MR**. 2007. Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proceedings of the National Academy of Sciences of the United States of America **104**:12282–12287.

24. **Berg T**, **Firth N**, **Apisiridej S**, **Hettiaratchi A**, **Leelaporn A**, **Skurray RA**. 1998. Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. J Bacteriol **180**:4350–4359.

25. **Firth N**, **Ridgway KP**, **Byrne ME**, **Fink PD**, **Johnson L**, **Paulsen IT**, **Skurray RA**. 1993. Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41. Structure **136**:13–25.

26. **Liu MA**, **Kwong SM**, **Jensen SO**, **Brzoska AJ**, **Firth N**. 2013. Biology of the staphylococcal conjugative multiresistance plasmid pSK41. Plasmid **70**:42–51.

27. **Caryl JA**, **Thomas CD**. 2006. Investigating the basis of substrate recognition in the pC221 relaxosome. Mol Microbiol **60**:1302–1318.

28. **Apisiridej S**, **Leelaporn A**, **Scaramuzzi CD**, **Skurray RA**, **Firth N**. 1997. Molecular analysis of a mobilizable theta-mode trimethoprim resistance plasmid from coagulase-negative staphylococci. Plasmid **38**:13–24.

29. **Climo MW**, **Sharma VK**, **Archer GL**. 1996. Identification and characterization of the origin of conjugative transfer (oriT) and a gene (nes) encoding a single-stranded endonuclease on the staphylococcal plasmid pGO1. J Bacteriol **178**:4975–4983.

30. **Larkin C**, **Datta S**, **Harley MJ**, **Anderson BJ**, **Ebie A**, **Hargreaves V**, **Schildbach JF**. 2005. Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. Structure **13**:1533–1544.

CHAPTER 2: PROCESSING OF NONCONJUGATIVE RESISTANCE PLASMIDS BY CONJUGATIVE NICKING ENZYME OF STAPHYLOCOCCI[1]

Introduction

Antimicrobial resistant strains of *Staphylococcus aureus* are a growing concern for hospital- and community-acquired infections. Most *S. aureus* bacteria examined clinically harbor at least one plasmid that encodes for antimicrobial resistance, and many plasmids carry multiple antimicrobial resistance determinants. The pSK41 family of plasmids is made up of large, low-copy-number, conjugative plasmids for which pSK41 is used as a prototype for characterization (1-4). These plasmids carry a variety of antimicrobial resistance determinants, including those against aminoglycosides, penicillins, tetracycline, bleomycin, trimethoprim, macrolides, lincosamides, mupirocin, antiseptics, and disinfectants (2, 4-9). This family of plasmids also played a key role in the rise of vancomycin resistant *S. aureus* (VRSA) (7, 8, 10). In addition to the antimicrobial resistance, they also carry transfer (*tra*) genes encoding the proteins necessary to conduct conjugative plasmid transfer that spread these plasmids among *S. aureus* and other gram-positive bacteria (6, 7, 10, 11).

[1] This chapter adapted from the previously published work Pollet RM, Ingle JD, Hymes JP, Eakes TC, Yui Eto K, Kwong SM, Ramsay JP, Firth N, Redinbo MR. 2016. Processing of Nonconjugative Resistance Plasmids by Conjugation Nicking Enzyme of Staphylococci. Journal of Bacteriology 198:888-897.

One of the proteins essential for conjugative plasmid transfer is the relaxase enzyme. A relaxase is responsible for initiation and completion of the transfer process as it cleaves one strand of the double-stranded plasmid to begin transfer, and then ligates that strand back together to complete transfer (8, 12-15). There are two classes of relaxases: multi-tyrosine relaxases that use a “thumb” motif to position the plasmid DNA for processing, and single-tyrosine relaxases which lack this thumb motif (9, 16). The relaxase of pSK41 is termed NES, nicking enzyme in *S. aureus*, and is a single-tyrosine relaxase (1-6, 8, 9). NES contains a relaxase N-terminal 220 residues and a C-terminal 350 residues necessary for *in vivo* function but via an uncertain mechanism (5, 7, 8, 10). Here we determine in which steps of conjugation the C-terminal domain plays an important role.

The crystal structure of the relaxase domain of NES was the first of a single-tyrosine relaxase bound to its target DNA, allowing for more detailed characterization of the protein-DNA interactions than previously possible (6-8, 11). This structure revealed two sets of important protein-DNA interactions. The first is that the “thumb” used by multi-tyrosine relaxases to position the DNA appears to be replaced by 12 protein-DNA contacts including a buried nucleotide 3 bases upstream of the *nic* site that places the DNA in the correct position to be nicked by the single, catalytically active tyrosine. The second set of protein-DNA interactions unique to the NES-DNA complex is composed of two protein loops, termed Hairpin Loop 1 and 2, that surround the DNA hairpin formed upstream of the *nic* site (Figure 2.1A and B). NES Hairpin Loop 1, shown in yellow in Figure 2.1A and B, forms two base specific contacts with the minor groove of the *oriT* DNA and one contact to the

phosphate backbone. NES Hairpin Loop 2, green in Figure 2.1A and B, contacts the DNA more extensively, with six base specific interactions and four phosphate contacts in the major groove of the DNA hairpin. Edwards *et al.* previously showed *in vitro* that these loops disrupt DNA cleavage by the relaxase domain alone and *in vivo* that full-length NES protein lacking these loops was not able to facilitate plasmid transfer (7, 8, 10, 14). However, this important protein-DNA interaction had not been characterized *in vitro* in the context of the full-length 665-residue NES protein and we set out to determine in which steps of conjugation this interaction plays a role.

pSK41-like conjugative plasmids have been shown to mobilize several smaller co-resident plasmids such as pC221 and pSK639, which encode their own *mob* genes (9, 12, 13, 15, 16). Recently, O'Brien *et al.* showed that another staphylococcal conjugative plasmid pWBG749, which is unrelated to pSK41, can facilitate the mobilization of other plasmids that lack *mob* genes (17). They demonstrated that this transfer is facilitated by origin-of-transfer sequences on the non-conjugative plasmids that mimic the pWBG749 origin-of-transfer sequence, suggesting a conjugative relaxase-*in trans* mechanism (18). We have identified sequences similar to the pSK41 origin of transfer on numerous non-conjugative staphylococcal resistance plasmids (Appendices 1 and 2), raising the possibility that pSK41 family plasmids might likewise facilitate mobilization of other plasmids by an analogous relaxase-*in trans* mechanism mediated by NES. To investigate this possibility, we have characterized the pSK41 *oriT* mimic sequences from two divergent non-conjugative plasmids. The first plasmid, pSK156, was isolated from a clinical strain in 1951 and is the earliest known multidrug efflux-encoding plasmid

(19). The second plasmid, pCA347, was first sequenced in 2013 after isolation from a USA600 methicillin-resistant strain of *S. aureus* and encodes resistance to penicillin and heavy metals (20). Importantly, the variation in the origin-of-transfer sequence of pSK41 and the mimics of pSK156 and pCA347 is in the hairpin region of the DNA (Figure 2.4A). Based on these observations we sought to explore the ability of NES to bind to and process putative *oriT* regions from pSK156 and pCA347 and the ability of pSK41 to facilitate transfer of plasmids containing these putative *oriT* regions in order to examine the potential for mobilization of plasmids containing pSK41 *oriT* mimics in staphylococci.

Materials and Methods

Cloning, Expression and Purification of the Relaxase Domain of NES

The relaxase domain of NES was previously cloned into the cysteine protease domain (CPD) fusion protein expression system developed by Shen *et al*. and optimized by this lab (7, 21). Sequence confirmed constructs were transformed into *Escherichia coli* BL21 (DE3) AI cells and grown in 1.5 L of lysogeny broth (LB), in the presence of ampicillin, at 37 °C, with shaking, to an optical density of 0.6–0.8. An L-Arabinose solution was added to the growth at a final concentration of 0.2% (vol/vol), and the temperature was reduced to 18 °C for 30 min. Protein expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) and cells were allowed to grow for 16 h. The resulting growths were spun at 4,500 × g for 30 min at 4 °C and pellets were stored at −80 °C. Individual pellets were resuspended in Buffer A [500 mM NaCl, 25 mM imidazole, 20 mM potassium phosphate buffer,

pH 7.4, and 0.02% (vol/vol) sodium azide] along with protease inhibitor tablets (Roche), DNase, and lysozyme. The slurry was lysed using a Fisher Scientific Sonic Dismembrator. The lysed cells were then spun at 18,500 × g for 1 h. The supernatant was filtered and flowed over a 5 mL HisTrap column (GE Healthcare). The column and bound protein was then washed with Buffer A followed by incubation with 1 CV of 2 mM inositol hexakisphosphate ($InsP_6$) for 3 hours at 4 °C. The cleaved protein was washed off the column with Buffer A and flowed directly onto a Superdex 200 column (GE Healthcare) preequilibrated in sizing buffer [300 mM NaCl, 50 mM Tris buffer, pH 7.4, and 0.02% (vol/vol) sodium azide]. Fractions containing protein as assessed by UV absorbance were analyzed via SDS-PAGE and fractions containing >95% pure protein were combined and stored at -80 °C.

Cloning, Expression and Purification of Full-Length NES

Wild-type full-length NES was previously cloned into the cysteine protease domain (CPD) fusion protein expression system developed by Shen *et al*. and optimized by this lab (7, 21). Loop deletion mutants were made through site directed mutagenesis to remove Hairpin Loop 1 (residues 77 to 82) and Hairpin Loop 2 (residues 150 to 157) and replace each with a linker composed of one glycine and one serine. Cleavage inactive mutants used in binding studies were made by replacing the tyrosine at amino acid position 25 with a phenylalanine. The resulting plasmids were transformed into *Escherichia coli* BL21 (DE3) AI cells and grown in 1.5 L of lysogeny broth (LB) in the presence of 0.1 mg/ml ampicillin at 37°C with shaking. At an optical density of 0.6-0.8, an L-Arabinose solution was added at a final concentration of 0.2% (vol/vol) and the temperature reduced to 18°C. After 30

minutes, protein expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) and cells were allowed to grow for 16 hours. The cells were pelleted and stored at -80°C. Individual cell pellets were resuspended in Buffer A [500 mM NaCl, 20 mM KH_2PO_4 pH 7.4, 25 mM Imidazole, 0.02% (v/v) sodium azide] along with protease inhibitor tablets (Roche), DNase, and lysozyme. The mixture was sonicated and then clarified via centrifugation. The supernatant was filtered and loaded onto a 5 mL HisTrap column (GE Healthcare). The CPD expression system contains a His_6 tag in addition to the CPD tag, which has self-cleavage abilities in the presence of inositol hexakisphosphate ($InsP_6$). Therefore, after the His-CPD-NES fusion protein was bound to the column via the His_6 tag, the column was washed with 2 column volumes (CV) Buffer A and then incubated with 2 mM $InsP_6$ for 3 hours at 4°C. The NES protein was then eluted off the column in Buffer A while the His_6 and CPD tags remained bound to the column. The NES protein was then passed over a Superdex 200 column (GE Healthcare) pre-equilibrated in sizing buffer [25 mM HEPES pH 7.4, 300 mM NaCl, 0.02% (v/v) sodium azide]. Purity of each fraction was assessed by SDS-PAGE gel and fractions containing >95% pure protein were combined and concentrated to approximately 1.2 mg/ml.

DNA Binding Studies

5'-end 6-FAM labeled DNA oligos were ordered from Integrated DNA Technologies, resuspended in annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 0.05 mM EDTA), and hairpin was formed by heating the oligo to 98°C for one minute and then cooling the solution by 3°C per second. The dissociation constant of

binding was calculated using fluorescence anisotropy as described in Edwards *et al.* (7). Briefly, protein was serially diluted into a buffer of 100 mM NaCl, 0.1 mg/ml BSA, 5 mM Magnesium Acetate, 25 mM Tris Acetate, pH 7.5 to give 40 µL at final protein concentrations ranging from 0 to 0.5 µM. Assays were conducted in a 384-well black assay plate (Costar) allowing for 16 concentrations of protein. 10 µL of the DNA probe was added to the 40 µL protein solution resulting in a final concentration of DNA of 50 nM in a total volume of 50 µL in each well. Fluorescence anisotropy of the fluorescein-labeled DNA was observed via excitation at 485 nm and emission at 520 nm using a PHERAstar plate reader (BMG Labtech). Measurements were made in triplicate and reported values are the average of three separate triplicate runs. Data were plotted as average fluorescence anisotropy as a function of protein concentration using Graphpad PRISM v6.05 (Graphpad, 2014). The following equation was employed to fit the data and to calculate the K_D for the substrate:

$$f = \min + (\max - \min)\frac{\left\{(T + x + K) - \left[(-T - x - K)^2 - 4Tx\right]^{\frac{1}{2}}\right\}}{2T}$$

where f is average fluorescence anisotropy signal; T, total DNA concentration (set to 50 nM); x, total protein concentration; K, K_D; min, average fluorescence anisotropy signal of no protein control; and max, average fluorescence anisotropy signal of sample at saturating concentration of protein. A single binding site was assumed and standard error is reported for each measurement. The reported values are an average of at least 5 independent experiments.

DNA Cleavage Assays

The same 5'-end 6-FAM labeled DNA oligos used for the DNA binding studies were used to measure equilibrium DNA cleavage via polyacrylamide gel electrophoresis (PAGE) gels. Each 10 µL reaction contained 1.52 µM NES protein (relaxase or full-length constructs), 1 µM DNA substrate, and EMSA buffer (50 mM NaCl, 20 mM Tris, pH 7.4, 0.02% (vol/vol) sodium azide). The reaction was incubated at 37°C for 1 hour and quenched by the addition of 2X running buffer (0.01% xylene cyanol, 0.01% Bromophenol Blue, 85% formamide, 20 mM EDTA, 2X TAE, 0.2% SDS). The resulting 20 µL reactions were run through a denaturing 16% polyacrylamide gel [35 mL 16% acrylamide gel stock (8 M urea, 16% polyacrylamide/bisacrylamide, 1X TBE), 300 µL 10% ammonium persulfate (APS), 33 µL tetramethylethylenediamine (TEMED)] in 1X TBE running buffer to separate cleaved product DNA from the substrate. Using the fluorescein tag, oligos were visualized using a VersaDoc Imaging System, 4400 MP (BioRad) and the QuantityOne software (BioRad). ImageJ 1.45s software was used to quantify band intensities and the percent cleavage product formation was calculated as a percentage of the product band intensity divided by the product plus the substrate band intensities. The average of at least six individual cleavage experiments are presented.

DNA Strand Transfer Assays

DNA strand transfer assays were performed similarly to DNA cleavage assays except two pieces of DNA were used. The first piece of DNA was an unlabeled substrate of the same sequence and length as the oligos used in the DNA

binding and cleavage studies (red DNA in Figure 2.3B). The second piece of DNA was a 5'-end 6-FAM labeled DNA oligo of the same sequence as the unlabeled substrate, but ending at the NES cleavage site (black DNA in Figure 2.3B). Each 10 µL reaction contained 1.52 µM NES protein (relaxase or full-length constructs), 1 µM unlabeled DNA substrate, 1 µM labeled DNA substrate, and EMSA buffer. The reaction was incubated, run, and analyzed as in the DNA cleavage assays. Percent strand transfer was calculated as a percentage of the product band intensity divided by the product plus the labeled substrate band intensities. The unlabeled DNA substrate was not visualized or quantified. The average of at least six individual cleavage experiments are presented.

Structure Modeling

The NES relaxase domain-DNA complex structure reported previously (Edwards *et al.*; RCSB accession code 4HT4) was employed for Figures 2.1A-B and 2.4B-C. (7). For Figures 2.4D and E in which pSK156 and pCA347 were modeled in place of the original pSK41 DNA, *Coot* was used to mutate each DNA residue, and the final figures were rendered in PyMol (22, 23).

Plasmid Sequence Analysis

The plasmid database compiled and analyzed for pWBG749 family *oriT* sequences by O'Brien *et al.* was analyzed for *oriT* sequences similar to that of pSK41 (18). The online interface of BLASTN was used to search these plasmids for the sequence ATAAGTGCGCCCTTACGGGATTTAAC from the pSK41 *oriT* and each sequence with a match was manually inspected for an adjacent DNA hairpin sequence (24). Plasmids were then grouped according to varying sequences found

in the DNA hairpin. Plasmids determined to carry potential pSK41 *oriT*-mimics were then searched for the NES relaxase gene (Accession Code: NC_005024.1, nucleotides 8115 to 10112) to determine if the plasmid is a conjugative plasmid.

Bacterial Strains, Plasmids, Growth and Assay Conditions

Strains and plasmids used are listed in Table 2.1. *E. coli* and *S. aureus* were cultured at 37°C on LB agar or in liquid LB medium with aeration (200 rpm). When required, growth medium was supplemented with antibiotics at the following concentrations: ampicillin (Ap) 100 µg/mL; chloramphenicol (Cm) 10 µg/mL; gentamicin (Gm) 20 µg/mL; novobiocin (Nb) 5 µg/mL; streptomycin (Sm) 50 µg/mL.

DNA fragments encompassing *oriT* regions were synthesised as GeneArt Strings (Figure 2.6; Life Technologies) and cloned into *Hin*dIII and/or *Bam*HI sites of the pSK1-based *S. aureus*/*E. coli* shuttle vector pSK5632. The insert integrity was verified by sequencing. pSK5632 constructs were introduced into the restriction-deficient *S. aureus* strain RN4220 by electroporation. pSK41 was introduced into each resulting strain by conjugation with strain SK5428 and resulting Cm^R/Gm^R-transconjugants were used as donors in mobilization experiments. Mobilization assays were conducted in BHI liquid medium (Sigma Aldrich) containing 40% (final) polyethylene glycol (PEG) as described previously (O'Brien *et al.*, 2015). The WBG4515 strain was used as a recipient and Sm/Nb was used to select against donors. Transconjugants were isolated on media additionally carrying either Gm (for pSK41) or Cm (for pSK5632).

Results

Role of the C-terminal Domain of NES

As discussed in Chapter 1, the C-terminal domain of NES is essential for function *in vivo;* Edwards *et al.* showed that a NES mutant lacking its C-terminal domain was not able to successfully facilitate conjugation of pSK41 (25). We set out to characterize the role of the C-terminal domain *in vitro* to determine in which step of conjugation it plays a significant role. To do so we compared the activity of the relaxase domain (1-220) of NES and the full-length (1-665) NES in both cleavage and strand transfer assays. In each case we also tested a range of oligonucleotide lengths as this has been previously shown to effect NES activity but it is unclear which lengths are biologically relevant (25). The oligonucleotides used are shown in Figure 2.1C.

In the cleavage assay, NES 1-220 and NES 1-665 showed differential activity ($p<0.005$) against OriTHP30 and OriTHP35 (Figure 2.2A). NES 1-665 was able to process OriTHP30 at a higher rate than NES 1-200; however, NES 1-220 was able to process OriTHP35 at a higher rate. In contrast, NES 1-220 and NES 1-665 did not show differential activity against OriTHP40 and OriTHP45 leading us to conclude that the C-terminal domain does not play a major role in the cleavage activity of NES.

As shown in Figure 2.2B, NES 1-220 and NES 1-665 did not show differential strand transfer activity against OriTHP30, OriTHP35, or OriTHP40. Interestingly, NES 1-220 showed significantly higher activity against OriTHP45 than NES 1-665.

We conclude that the C-terminal domain is important for proper regulation of the strand transfer activity of NES.

Characterization of NES Hairpin Loop 1 and 2

The crystal structure of the relaxase domain of NES in complex with pSK41 *oriT* DNA hairpin, reported previously by Edwards *et al.*, revealed two features unique to this class of relaxase: two protein loops, termed Hairpin Loop 1 and 2 (Figure 2.1A and B), that clamp around the hairpin duplex of the *oriT* DNA (7). These contacts are unique to this class of relaxase compared to those observed with the longer, multi-tyrosine relaxases like F TraI, and they have yet to be characterized in the context of the full-length protein. Hence, we sought to determine the impact deleting these unique loops would have on NES functions *in vitro*. Hairpin Loop 1 deletion (ΔL1), Hairpin Loop 2 deletion (ΔL2), or double-deletion (ΔL1ΔL2) forms of full-length NES protein were created using site-directed mutagenesis in which the loops were replaced with Gly-Ser linkers. The proteins were expressed recombinantly in *E. coli* and purified to homogeneity. DNA binding, cleavage, and strand transfer assays were conducted using DNA oligonucleotides similar to that employed in the complex presented in the crystal structure and possessing the same sequence as the origin of transfer (*oriT*) of NES conjugated plasmid pSK41 (Figure 2.1C).

For DNA binding studies, these variant proteins contained an active site Tyr-25-Phe mutation. Previous evidence suggested NES relaxase activity is dependent on oligonucleotide length (Figure 2.2, 7) so varying lengths of the *oriT* were used to verify this trend and named as in Figure 2.1C. Longer oligonucleotides should better

mimic *in vivo* plasmid transfer. As shown in Figure 2.3A, the ΔL1 form of full-length NES exhibited increased DNA binding ($p<0.005$) compared to wild-type NES on the OriTHP35 and OriTHP40 oligonucleotides. ΔL2 NES did not demonstrate significantly different DNA binding on any oligo. In contrast, ΔL1ΔL2 NES showed significantly increased DNA binding ($p=0.0002$) on the shortest oligo tested, OriTHP30, but decreased binding ($p<0.005$) on OriTHP35. For the longest oligonucleotide tested, OriTHP45, no difference in binding was observed for any variant proteins compared to wild-type NES. Thus, we conclude that eliminating Hairpin Loops 1 or 2 from full-length NES can alter DNA binding *in vitro* in an oligonucleotide length-dependent manner.

DNA cleavage and strand transfer assays were conducted as described in Figure 2.3B. The cleavage assay mimics the cleavage of the plasmid *oriT* to produce the single strand transferred during conjugation. On OriTHP30, only ΔL1 NES exhibited a significant ($p<0.0001$) difference in DNA cleavage, in this case a reduction, relative to wild-type NES (Figure 2.3C). On the longer OriTHP35, OriTHP40 and OriTHP45 oligos, all three variant proteins (ΔL1, ΔL2, ΔL1ΔL2) demonstrated statistically significant ($p<0.0001$) increases in levels of DNA cleavage relative to wild-type NES. For these longer oligos, wild-type cleavage was observed at ~4%, while the variant proteins exhibited 2- to 7-fold increases in cleavage. We conclude that eliminating the DNA hairpin-associating loops from NES increases DNA cleavage by the enzyme.

A more dramatic effect was observed in examining DNA strand transfer by the variant full-length forms of NES. The strand transfer assay measured the

ligation of a portion of DNA covalently linked to NES following cleavage to a new piece of DNA containing the hairpin characteristic of the *oriT* (Figure 2.3B). This mimics the ligation step of conjugation that ends plasmid transfer. For OriTHP30, all NES variants (ΔL1, ΔL2, and ΔL1ΔL2) showed 5- to 15-fold increases in DNA strand transfer relative to wild-type NES (Figure 2.3D). For OriTHPs 35, 40 and 45, the increases were even larger – 25% to nearly 50% of the substrate oligos provided to the NES variants were processed to strand transfer, while ~5% of the oligos formed strand transfer products with wild-type NES. Thus, eliminating the DNA hairpin contacting loops of NES produces significant and dramatic increases in the level of DNA strand transfer *in vitro* compared to wild-type NES. It can be concluded that the Hairpin Loop 1 and Loop 2 regions of NES play an important role, particularly on longer DNA substrates more relevant to transfer *in vivo*, in limiting the level of DNA religation during conjugation.

Modeling of NES Bound to pSK156 and pCA347

We next sought to determine if related DNA sequences from other plasmids might serve as substrates for pSK41 NES. We examined the *S. aureus* plasmids of known sequence and selected two with sequences identical to the pSK41 *oriT* cleavage site. These two plasmids, pSK156 and pCA347, exhibited the same sequence as pSK41 in the 20 nucleotide region from the predicted hairpin through the *nic* site, but deviated somewhat from the DNA hairpin region of the pSK41 *oriT* (Figure 2.4A). We predicted based on modeling that pSK156 and pCA347 might each form 8 base pair DNA hairpins with a one-nucleotide bulge; by contrast, pSK41 is known from crystal structure to form a 7 base pair DNA hairpin with no bulge

(Figure 2.4A and B). Interestingly, within the predicted DNA hairpins of pSK156 and pCA347, nucleotides at the base of the DNA hairpin (G_3, C_{17} and G_{18}) are conserved with the sequence of pSK41 (Figure 2.4A, C, D and E). Furthermore, we noted that the 8 base pair hairpins predicted for pSK156 and pCA347 are nearly identical in sequence to each other (Figure 2.4A, D and E).

We next modeled the pSK156 and pCA347 DNA sequences into the pSK41 NES relaxase domain-DNA hairpin complex crystal structure. For reference, Figure 2.4B shows the NES relaxase domain in complex with the pSK41 DNA hairpin, highlighting the interactions between the protein and DNA; the boxed region contains all the protein contacts with the DNA hairpin and will remain the focus of the pSK156 and pCA347 models. As shown in Figure 2.4C, NES makes base specific contacts with the pSK41 DNA hairpin at C_4 via N154, T_{16} via R78, C_{17} via G153 and N154, G_{18} via R151, and C_{19} via Y156. All but one contact to pSK156 and two contacts to pCA347, along with six phosphate contacts, are maintained in the models despite the changes in the DNA sequences between these plasmids and pSK41 (Figures 2.4D and E). Because C_{17} and G_{18} are conserved in both pSK156 and pCA347, the contacts via G153, N154, and R151 are maintained. The cytosine at position 19 in pSK41 is replaced by a thymine in pSK156 and pCA347; however, the para-oxygen of thymine appears capable of receiving a hydrogen bond from Y156 of NES. Position 16 of pSK156 and pCA347 contains an adenine rather than the thymine found in pSK41. The ring nitrogen of adenine appears capable in our models of receiving the same hydrogen bond from R78 as the oxygen of thymine; however, while the thymine oxygen can form two hydrogen bonds, the adenine

nitrogen can form only one. Thus, in spite of sequence differences between pSK41 and these other two *S. aureus* plasmids, contacts between NES and the predicted *oriT*s of all three plasmids are largely maintained.

An additional contact is predicted to be lost between NES and pCA347. While a cytosine is conserved in the same positions in pSK41 (position -4) and pSK156 (position -2), it is a thymine in pCA347 (T_2; Figure 2.4E). In pSK41 and pSK156 the amine group of C_4 donates a hydrogen bond to the oxygen of N154; however, the para-oxygen of thymine cannot form the same interaction. It is possible that the asparagine side chain could rotate to allow the thymine oxygen to receive a hydrogen bond from the N154 side chain amine. In doing so, though, this side chain would lose an interaction with C_{17}. Despite this potential change, five base-specific contacts and six phosphate contacts are maintained in our models between NES and the sequences of plasmids pSK156 and pCA347 in this region. Thus, we hypothesize that NES is capable of binding to and utilizing these potential *oriT* regions of pSK156 and pCA347 as substrates.

Characterization of NES Processing of pSK156 and pCA347

We next analyzed the ability of pSK41 NES to process the potential *oriT* regions of pSK156 and pCA347 by measuring the protein's ability to employ these DNAs for binding, cleavage and strand transfer. For DNA binding studies, wild-type full-length NES with an active site Y25F mutation was employed along with OriTHP40-like forms of the pSK41, pSK156, and pCA347 (Figure 2.4A). The OriTHP40-like form alone was analyzed as longer oligonucleotides have been shown to be important for the regulatory function of the C-terminal domain (Figure

2.2, 7) but significant differences between OriTHP40 and 45 were not seen in assays with the NES loop deletion protein mutants. NES bound the *oriT* mimic regions of pSK156 and pCA347 but less well compared to its binding of the pSK41 *oriT* (Figure 2.5A). The K_D of NES binding to pSK41 is 19.3 ± 3 nM; in contrast, NES binds to pSK156 and pCA347 3- and 9-fold weaker, with K_Ds of 55.8 ± 9 nM and 175 ± 30 nM, respectively. While the loss of one or two hydrogen bonds is not sufficient to explain this decrease in binding affinity, it is interesting that the changes in binding affinity reflect the degree of change in sequence and interactions seen in our models.

DNA cleavage and strand transfer assays with full-length wild-type pSK41 NES and the OriTHP40-like regions of pSK156 and pCA347 were conducted as described in Figure 2.3B. For DNA cleavage, pSK156 exhibited the same level of activity as pSK41, while pCA347 showed significantly decreased cleavage by NES ($p<0.0001$, Figure 2.5B). However, as the cleavage process is dependent on NES first binding the DNA, this reduction in cleavage may result from the decrease in binding seen in Figure 2.5A for pCA347. For DNA strand transfer, both pSK156 and pCA347 showed significantly increased DNA strand transfer, with 3- and 7-fold increases in strand transfer for pSK156 and pCA347, respectively, relative to pSK41 (Figure 2.5C). Taken together, these DNA binding, cleavage, and strand transfer data reveal that NES is capable of processing pSK156 and pCA347 *oriT*-like sites but does so at lower efficiency than at its cognate site.

Relaxase-*in trans* Mobilization by pSK41 *In Vivo*

To investigate the ability of pSK41 to facilitate relaxase-*in trans* mobilization, the *oriT*-like sites corresponding to pSK156 and pCA347, and the pSK41 *oriT* sequence itself were synthesized and cloned into the non-mobilizable shuttle vector pSK5632 (26) to generate the plasmids pSK6881, pSK6879 and pSK6877, respectively; the DNA fragments cloned are shown in Figure 2.6. These new plasmid constructs and pSK5632 were electroporated into *S. aureus* strain RN4220 and pSK41 was subsequently introduced via conjugation. These strains were then used as donors in mobilization assays with the recipient strain *S. aureus* WBG4515. As shown in Table 2.2, pSK41 was found to mobilize pSK6877, containing its own cognate *oriT* sequence, at a frequency of 2.9×10^{-5}, approximately five-fold lower than pSK41 itself transferred in the same assay (1.4×10^{-4}). However, despite repeated efforts, mobilization of the plasmids containing the pSK156 or pCA347 *oriT* mimics was never detected. These results demonstrate that pSK41-encoded NES can mediate *in trans* mobilization of a plasmid containing a copy of its own *oriT* site, but suggest its activity on the variant *oriT*-like sites from pSK156 and pCA347 is inadequate to facilitate plasmid transfer *in vivo*. As discussed below, an accessory protein may be required to complete relaxase-*in trans* transfer *in vivo*.

Discussion

Transfer of conjugative and mobilizable plasmids is a major route by which antimicrobial resistance propagates, but the lack of details about the mechanism of this process impedes efforts to slow or prevent the spread of such resistance. We

focus on the mechanism of action of the NES relaxase enzyme encoded by pSK41 and related plasmids from staphylococci. NES is a two domain protein where the C-terminal domain is distinct from that which is normally paired with a relaxase domain. Rather than possessing a function important to the conjugation process such as a helicase, NES does not seem to possess any catalytic activity but was previously shown to be essential for successful conjugation (25). Here we show that the C-terminal domain is important for the strand-transfer (ligation) function but not cleavage activity of NES (Figure 2.2). It is likely that the increase in DNA strand transfer causes DNA to be ligated before transfer is complete. It has been suggested that the ligation action of single-tyrosine relaxases such as NES requires homo-dimerization, which could be mediated by the C-terminal domain. Alternatively, although we have not measured the DNA binding ability of the C-terminal domain, it could play a role in DNA sequence discrimination to ensure ligation only occurs to the 3' end of the T-strand rather than any DNA blunt end it encounters.

Formation of a DNA hairpin in the pSK41 *oriT* and the importance of the associated NES Hairpin Loop 1 and 2 had been suggested previously (7). We demonstrate here that NES Hairpin Loops 1 and 2 are important for proper DNA cleavage and strand transfer (Figure 2.3C and D) but not for DNA binding (Figure 2.3A). The large increase in DNA strand transfer causing DNA to be ligated before transfer is complete is likely the biggest contributor to the large reduction of plasmid transfer seen when either Hairpin Loop 1 or Loop 2 are eliminated from the encoded NES enzyme (7). It is also likely that accessory proteins in the pSK41 relaxosome

complex with NES through interactions with the NES Hairpin Loop 1 and 2, amplifying the effect of loss of these protein features.

Because relaxases are essential for transfer, share many common features, and are unique to the conjugative plasmid system, they represent a novel therapeutic target for decreasing the spread of antimicrobial resistance to allow current antimicrobial compounds to maintain efficacy. As explored previously and continued in Chapter 3, there are two potential sites of disruption common to relaxases: the metal binding site and specific protein-DNA interactions (7, 27). These results validate the NES Hairpin Loop 1 and 2 DNA interactions as a target site for such therapeutics. By disrupting the specific protein-DNA interactions in the NES Hairpin Loops, a molecule such as a sequence-specific polyamide could specifically disrupt cleavage and religation during pSK41 conjugation (7). As there seems to be some sequence conservation at the base of the DNA hairpin, this inhibitor molecule could target mobilizable plasmids in addition to the conjugative plasmid. Interestingly, there is a biological example of relaxase interference from *Staphylococcus epidermidis* strains carrying a CRISPR spacer that matches the *nes* gene of pSK41 and limits conjugative transfer (28). Targeted disruption of conjugation after initiation of the process and formation of the mating pore could cause cell death specifically in conjugative plasmid containing bacteria. This targeted approach to bactericidal compounds is desirable as we learn more about the importance of the human skin microbiome (29).

Despite the importance of the protein-DNA interactions at the DNA hairpin, we show that there is some flexibility in the DNA hairpin sequence allowing

sequences from pSK156 and pCA347 to be processed by NES. The origin-of-transfer-mimic sequences of pSK156 and pCA347 maintain all but one or two protein-DNA contacts, respectively, and are able to be bound, cleaved, and ligated by NES, although with altered efficiency. We were therefore somewhat surprised to find that plasmid constructs containing pCA347 or pSK156 *oriT* mimics could not be mobilized from cells harboring pSK41 co-resident, in contrast to a pSK41 *oriT* construct. However, the analogous relaxase-*in trans* mobilization phenomena recently described for the distinct pWBG749-like conjugative plasmids provide a precedent that likely explains this apparent paradox. Namely, pWBG749 *oriT*-like sequences exist as sub-types differentiated by sequence divergence in an inverted repeat (IR2) located adjacent to the *nic* site-containing core sequence (18). This results in specificity between various mobilizable plasmids and particular pWBG749-like conjugative plasmids. Thus, pWBG749 can mobilize plasmids with a pWBG749-like *oriT* of sub-type OT49 but not those carrying an OT45 sub-type, which instead can be mobilized by pWBG749-like conjugative plasmids that possess a cognate OT45 sub-type *oriT* (18). Despite this, pWBG749 was able to stimulate recombination between OT49- and OT45-type *oriT* sequences carried on the same mobilizable plasmid, indicating that the pWBG749 relaxosome could recognize the OT45-type *oriT* even though it can't mediate transfer of that sub-type (18). By analogy, it would seem plausible that the pSK156 and pCA347 *oriT*-mimics examined here represent sub-types of pSK41-like *oriT*s that can be recognized by NES but cannot be mobilized by the pSK41 relaxosome. In the case of the pWBG749 system, it has been shown that specificity for IR2 sub-types is dictated by

a small putative DNA-binding accessory protein, SmpO, encoded adjacent to the *oriT* on pWBG749, rather than the relaxase SmpP (18), The involvement of accessory proteins in the pSK41 relaxosome is yet to be established.

Importantly, the scenario proposed above implies the existence of pSK41-like conjugative plasmids with variant *oriT* sequences that would be capable of mobilizing plasmids such as pSK156, pCA347 and other plasmids listed in Appendix 2. Although no such variant pSK41-like plasmids have been detected to date, the presence of variant pSK41-like *oriT* mimic sequences on one fifth of all sequenced staphylococcal plasmids (not including pSK41-like plasmids themselves) makes the whereabouts of such a reservoir an important question, since it is clearly influencing the evolution of plasmids in clinical staphylococci.

Interestingly, the origin-of-transfer-mimic sequence found in pCA347 is identical to a sequence found in pWBG757, a plasmid that could not be mobilized with pWB749 in the studies by O'Brien and colleagues (17). Comparative data such as this may allow us to classify mobilizable plasmids into families related to relaxase(s) used for trans-mobilization. The origin-of-transfer-mimic sequence of pSK156 is also found in plasmid pWB747, which could be mobilized by pWBG749, suggesting pWBG747 harbors two distinct origin-of-transfer sequences to maximize its ability to be transferred (17, 18).

We searched the sequenced staphylococcal plasmids for other pSK41 *oriT*-like sequences and found 85 sequences from 83 different plasmids, including 14 pSK41-family conjugative plasmids (Appendix 1). This represents 23.6% of *Staphylococcus* plasmids which is significantly lower than the 53% of plasmids O'Brien *et al.*

identified as harboring the pWBG749 *oriT*-like sequence (18). However, the set identified here includes 26 plasmids that do not have a pWBG749 *oriT* sequence, suggesting NES is an important factor in the relaxase-*in trans* conjugation mechanism.

Analysis of all 85 sequences shows that *oriT* sequences identical to that of pSK41 are only evident on plasmids which encode their own NES protein. On other plasmids, the pCA347 and pSK156 hairpin sequences with their one-nucleotide difference are by far the most common *oriT* mimic, representing 73% of the sequences (Appendix 2). The other two major pSK41 *oriT*-mimic types are significantly different in sequence but are still predicted to form a DNA hairpin which will allow for most of the NES protein-DNA interactions seen with the pSK41 *oriT* to be maintained, again suggesting that these specific protein-DNA interactions may be a potential therapeutic target. It is likely that that NES proteins are capable of acting on a wide range of non-conjugative staphylococcal plasmids that contain an *oriT* mimic sequence, ranging from the oldest known multidrug resistance plasmid pSK156 to prevalent contemporary plasmids such as pMW2 and pUSA300HOUMR. The results described here imply that the recently described relaxase *in-trans* mechanism of mobilization extends beyond pWBG749-like conjugative plasmids to the clinically more prevalent pSK41-like plasmids, thereby further increasing the proportion of staphylococcal plasmids that are potentially mobilizable. These observations lend further weight to the recent proposal that relaxase-*in trans* mobilization represents a significant driver of horizontal transfer in *Staphylococci* (18).

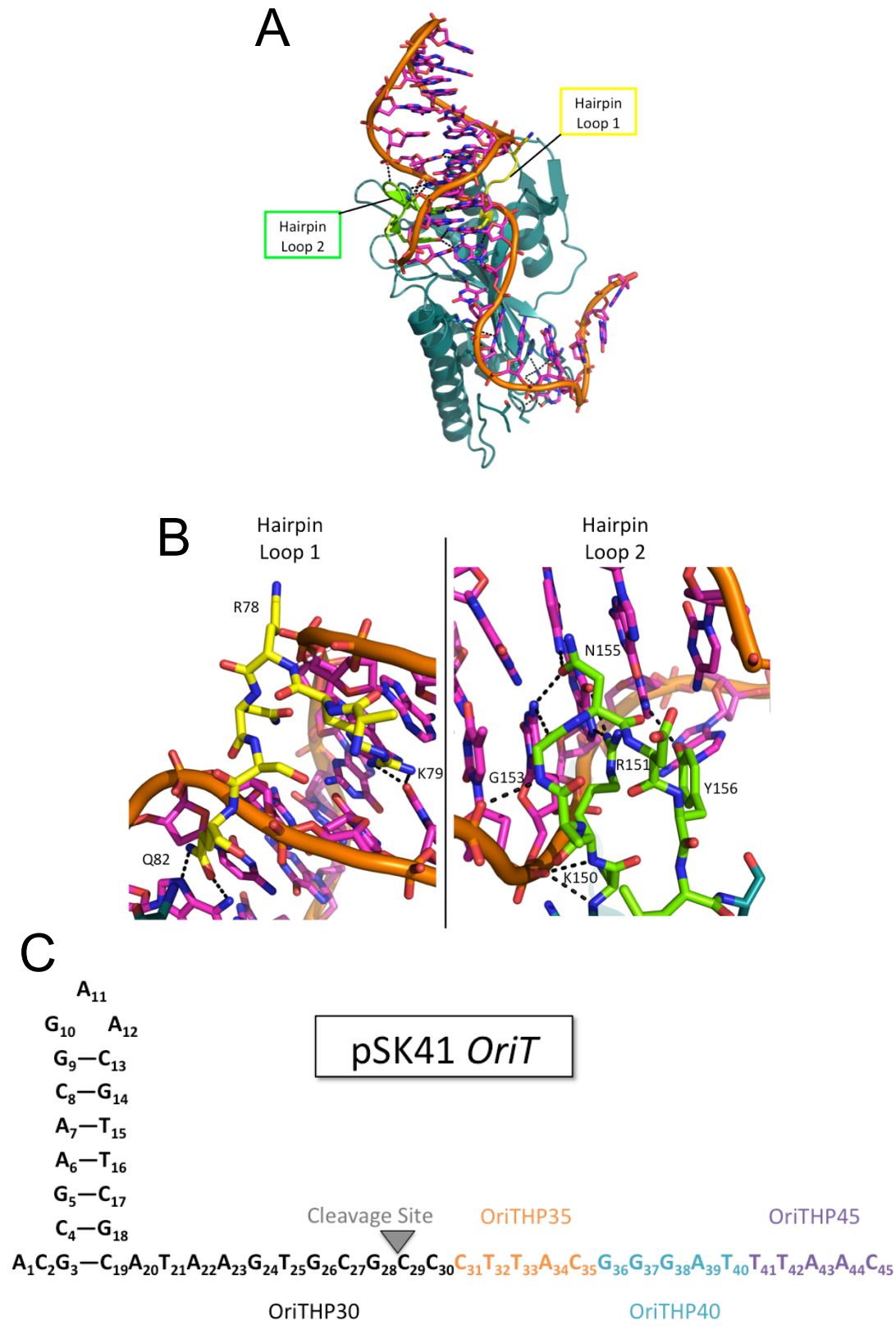

A
Hairpin Loop 1
Hairpin Loop 2
B
Hairpin Loop 1
Hairpin Loop 2
R78
K79
Q82
N155
R151
G153
Y156
K150
C
A11
G10 A12
G9—C13
C8—G14
A7—T15
A6—T16
G5—C17
C4—G18
pSK41 OriT
Cleavage Site
OriTHP35
OriTHP45
A1C2G3—C19A20T21A22A23G24T25G26C27G28C29C30C31T32T33A34C35G36G37G38A39T40T41T42A43A44C45
OriTHP30
OriTHP40

Figure 2.1. Structure of NES Relaxase Domain and pSK41 *oriT*
A. Structure of the NES relaxase domain in complex with DNA from the pSK41 origin of transfer (4HT4, 13). The NES Hairpin Loop 1 is shown in yellow and NES Hairpin Loop 2 in green.
B. Hairpin Loop 1 (yellow) of NES binds in the minor groove of the DNA hairpin formed by the pSK41 origin of transfer while Hairpin Loop 2 (green) binds to the major groove.
C. Schematic of the pSK41 *oriT* and the oligonucleotides used in these studies. When only black portion is used, the oligo is referred to as OriTHP30. When the sequence is extended to include the orange portion it is referred to as OriTHP35, the teal as OriTHP40, and the purple as OriTHP45.

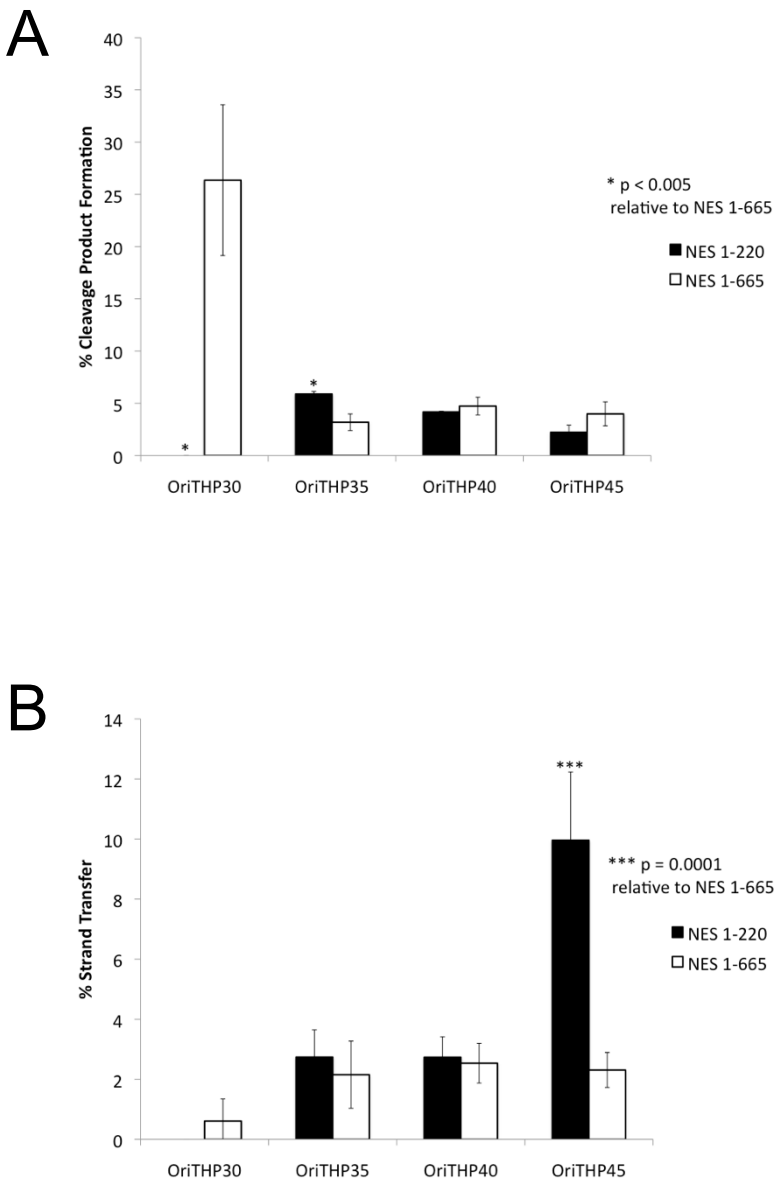


Figure 2.2. Cleavage and Strand Transfer by Relaxase Domain and Full-length NES.

A. Effect of oligonucleotide length on cleavage by NES 1-220 (relaxase domain) and NES 1-665 (full-length) protein.
B. Effect of oligonucleotide length on strand transfer by NES 1-220 (relaxase domain) and NES 1-665 (full-length) protein.

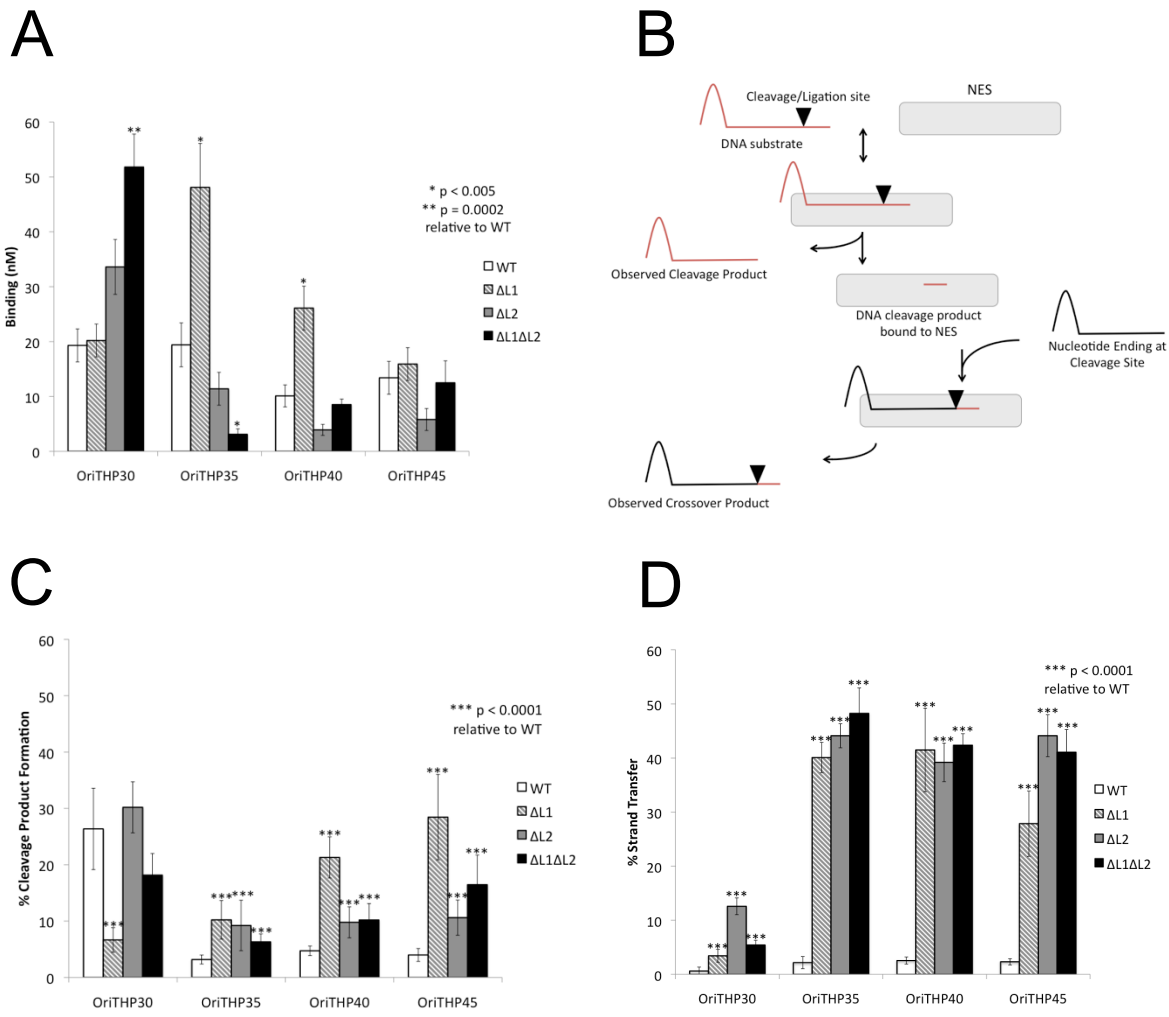
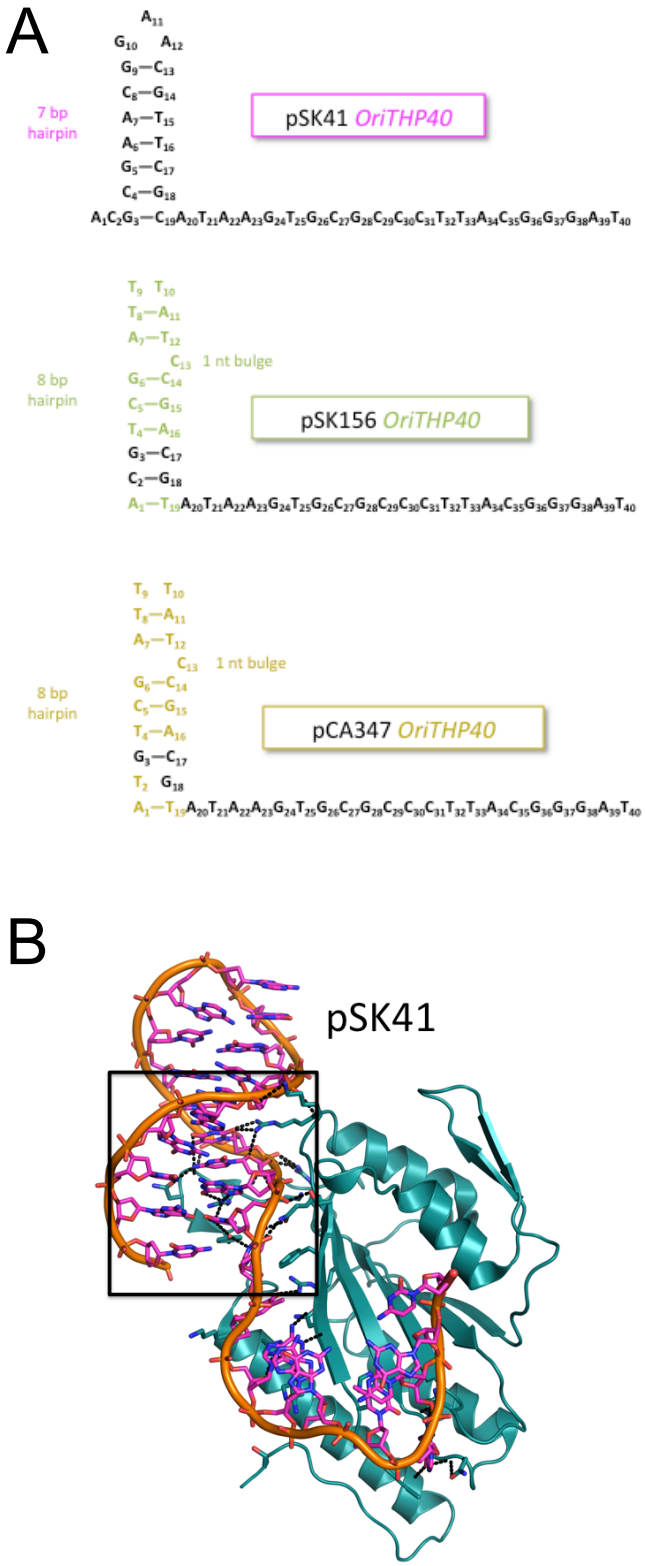
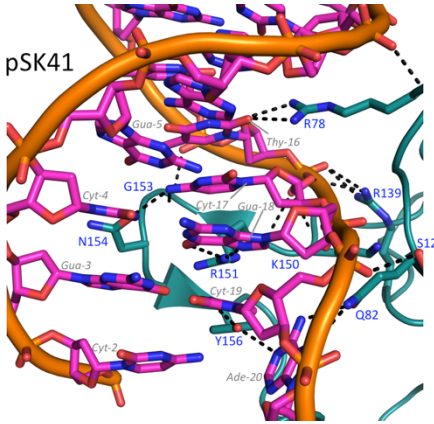
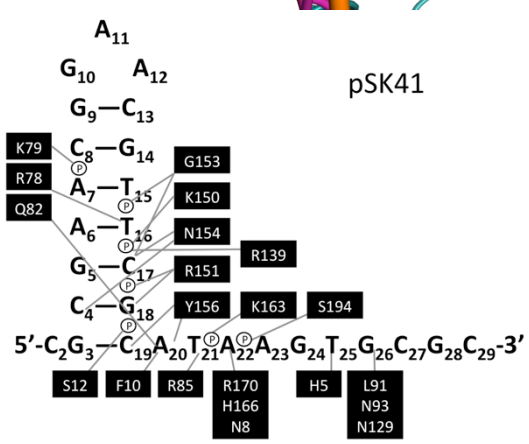


Figure 2.3. Functional Analysis of NES Loop Deletion Mutants
A. K_D ± standard deviation of DNA binding measured by fluorescence anisotropy for the indicated pSK41 oligonucleotides and NES mutants.
B. Schematic of the DNA cleavage and strand transfer assays. DNA cleavage assays involve only the red DNA substrate labeled with a 5' 6-FAM. DNA strand transfer assays involved both the red and black DNA substrate with the red substrate being unlabeled and the black substrate being 5' 6-FAM labeled.
C. Functional analysis of cleavage activity of NES mutants on varying pSK41 oligonucleotides. Wild-type data is the same as that presented in Figure 2.2A.
D. Functional analysis of strand transfer activity of NES mutants on varying pSK41 oligonucleotides to mimic religation during conjugative plasmid transfer. Wild-type data is the same as that presented in Figure 2.2B.

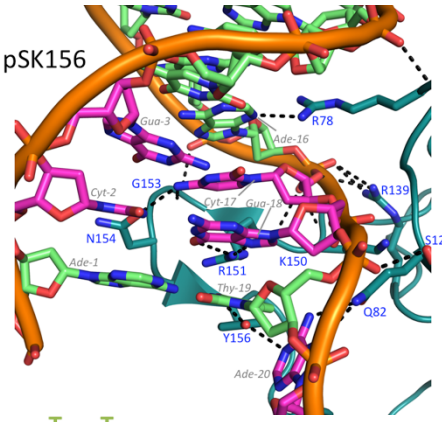

A
7 bp hairpin
pSK41 OriTHP40
8 bp hairpin
pSK156 OriTHP40
1 nt bulge
8 bp hairpin
pCA347 OriTHP40
1 nt bulge
B
pSK41

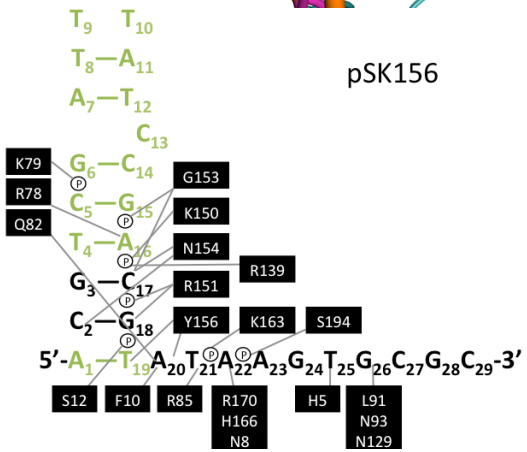
C


pSK41
R78
Gua-5
Thy-16
G153
Cyt-4
Cyt-17
Gua-18
R139
N154
S12
Gua-3
R151
K150
Cyt-19
Q82
Y156
Cyt-2
Ade-20

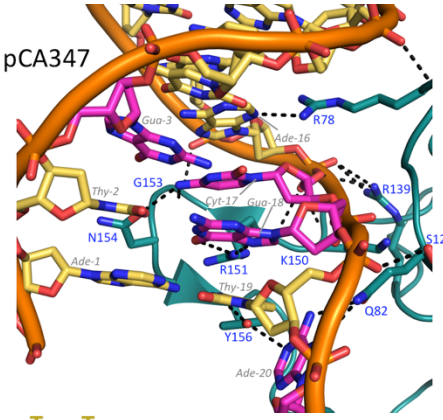

A11
G10 A12
G9—C13
pSK41
K79
R78
Q82
C8—G14
A7—T15
A6—T16
G5—C17
C4—G18
G153
K150
N154
R139
R151
Y156
K163
S194
5'-C2G3—C19A20T21A22A23G24T25G26C27G28C29-3'
S12
F10
R85
R170
H166
N8
H5
L91
N93
N129

D


pSK156
R78
Gua-3
Ade-16
G153
Cyt-2
Cyt-17
Gua-18
R139
N154
S12
Ade-1
R151
K150
Thy-19
Q82
Y156
Ade-20


T9 T10
T8—A11
A7—T12
C13
pSK156
K79
R78
Q82
G6—C14
C5—G15
T4—A16
G3—C17
C2—G18
G153
K150
N154
R139
R151
Y156
K163
S194
5'-A1—T19A20T21A22A23G24T25G26C27G28C29-3'
S12
F10
R85
R170
H166
N8
H5
L91
N93
N129

E


pCA347
R78
Gua-3
Ade-16
Thy-2
G153
Cyt-17
Gua-18
R139
N154
S12
Ade-1
R151
K150
Thy-19
Q82
Y156
Ade-20

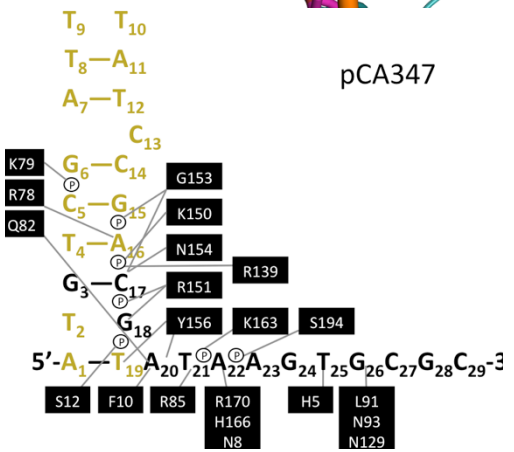
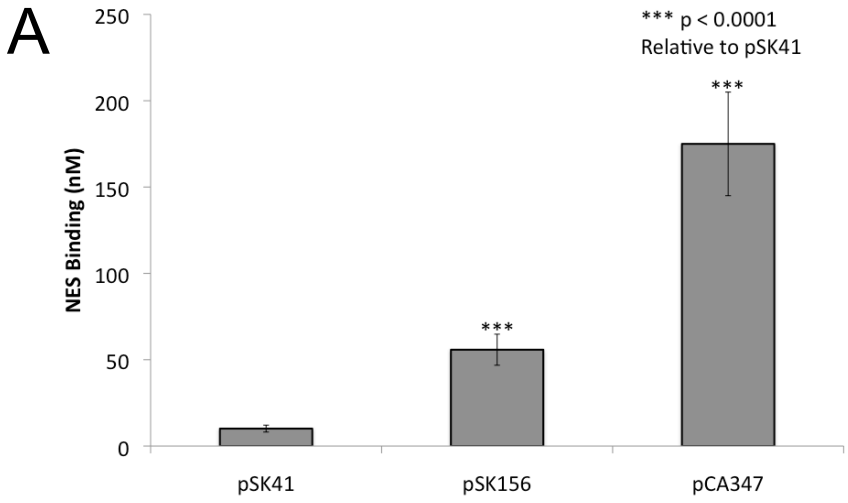
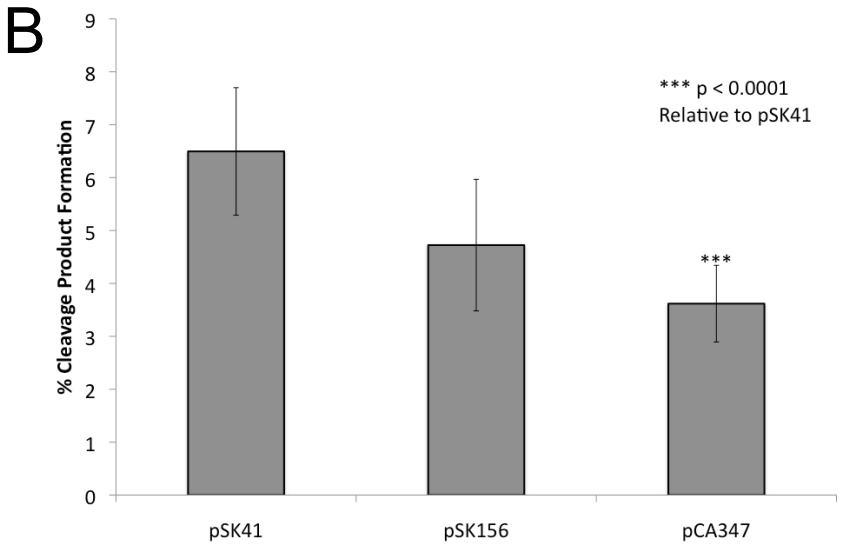

T9 T10
T8—A11
A7—T12
C13
pCA347
K79
R78
Q82
G6—C14
C5—G15
T4—A16
G3—C17
T2
G18
G153
K150
N154
R139
R151
Y156
K163
S194
5'-A1—T19A20T21A22A23G24T25G26C27G28C29-3
S12
F10
R85
R170
H166
N8
H5
L91
N93
N129

Figure 2.4. Modeled Structure of the pSK41, pSK156, and pCA347 *oriT*s
A. Schematic of the pSK41, pSK156, and pCA347 oligonucleotides used in these studies. Colored nucleotides indicate a difference in sequences from the pSK41 *oriT*.
B. The relaxase domain of NES in complex with the pSK41 *oriT*. The box shows the region focused on for Figure 2.4C, D and E.
C. Contacts between the NES relaxase domain Hairpin Loop 1 and 2 amino acids and the pSK41 *oriT* nucleotide.
D. Contacts between the NES relaxase domain Hairpin Loop 1 and 2 amino acids and the modeled pSK156 *oriT* nucleotide. Green nucleotides differ from the pSK41 *oriT*.
E. Contacts between the NES relaxase domain Hairpin Loop 1 and 2 amino acids and the modeled pCA347 *oriT* nucleotide. Gold nucleotides differ from the pSK41 *oriT*.


A
*** p < 0.0001
Relative to pSK41
NES Binding (nM)
250
200
150
100
50
0
***
***
pSK41
pSK156
pCA347


B
*** p < 0.0001
Relative to pSK41
% Cleavage Product Formation
9
8
7
6
5
4
3
2
1
0
***
pSK41
pSK156
pCA347

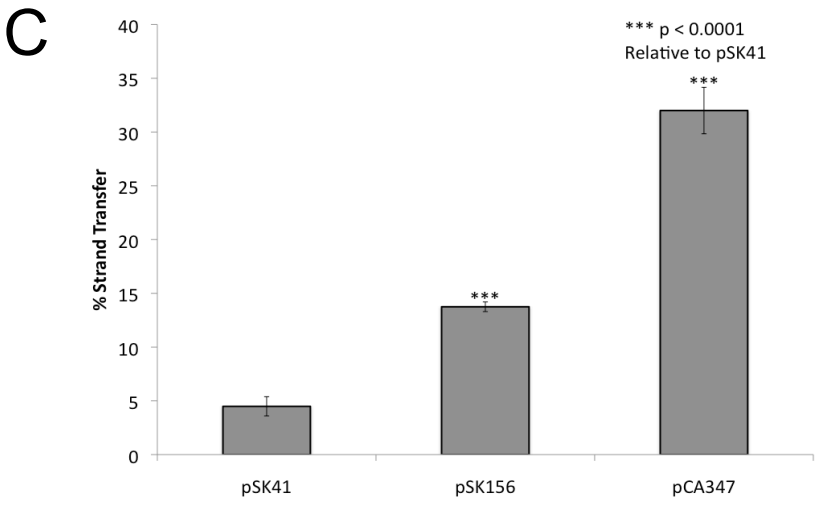

C
*** p < 0.0001
Relative to pSK41
% Strand Transfer
40
35
30
25
20
15
10
5
0
***
***
pSK41
pSK156
pCA347

Figure 2.5. NES Processing of pSK41, pSK156, and pCA347 *oriT* Oligonucleotides

A. K_D ± standard deviation of DNA binding measured by fluorescence anisotropy for the pSK41, pSK156, and pCA347 *oriT*s. The pSK41 data is the WT NES data as presented in Figure 2.3A.

B. Cleavage activity of the pSK41-encoded NES protein on the pSK41, pSK156, and pCA347 *oriT*s.

C. Strand transfer activity of the pSK41-encoded NES protein on the pSK41, pSK156, and pCA347 *oriT*s.

pSK41 *oriT*-string

cccaagcttAGC**ACGCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGAT**TTAACTAGAT
TATAACGGTAAAACTTGAATCTGTCAAACGAAggatccgc

pSK156 *oriT*-string

gatcggatccTTTTTCTTTCGACGCCGTATCGTCGATGTAATTCAAAAAAGTTATGGGCTAT
AAATCTACATCACTTTTTCAAGAATGTAGTAGCAATATTCAACGCAAATTAATTTTTATAAC
TGCTCGGAATATCTCAAGCCGTTTCTTTAATTTTGAAATAAAAAAAATCGACGAAGGTCGAT
TACGTTTTTGTAC**ACGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGAT**TTAACTAGA
TTATAACGACGAATTTTAGACCTGTAAAGCAATggtaccaagcttggatccgatc

pCA347 *oriT*-string

gatcggatccTTTTTCTTTCGACGCCGTATCGTCGATGTAATTCAAAAAAGTTATGGGCTAT
AAATCTACATCACTTTTTCAAGAATGTAGTAGCAATATTCAACGCAAATTAATTTTTATAAC
TGCTCGGAATATCTCAAGCCGTTTCTTTAATTTTGAAATAAAAAAATCGACGAAGGTCGATT
ACGTTTTTGCAC**ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGAT**TTAACTAGAT
TATAACGACGAATTTTAGACCTGTAAAGCAATggtaccaagcttggatccgatc

Figure 2.6. Nucleotide Sequences for *in vivo* Transfer Assays
Nucleotide sequences of the GeneArt Strings (Invitrogen) encompassing the *oriT* regions of plasmids pSK41 (GenBank Acc. AF051917, nt 10189-10273), pSK156 (GenBank Acc. GQ900448, nt 13396-13126) and pCA347 (GenBank Acc. CP006045, nt 3825-4094). Sequences in lower case were added to the DNA fragments to facilitate cloning and the *Bam*HI and *Hin*dIII restriction sites used are underlined. Sequence in bold correlates to the oligonucleotide *oriTs* shown in Fig. 2.4A.

Table 2.1. Bacterial Strains and Plasmids Used in This Study

Strain or plasmid	Description[a]	Reference or source
Escherichia coli		
DH5α	F⁻ *endA hsdR17 supE44 thi-1 λ⁻ recA1 gyrA96 relA1 ϕ80 dlacZΔ*M15	Bethesda Research Laboratories
BL21(DE3)	F⁻ *ompT hsdSB* (r_B^- m_B^-) *dcm gal λ*(DE3)	Novagen
Staphylococcus aureus		
RN4220	Restriction-deficient derivative of NCTC8325-4	Kreiswirth *et al.*, 1983
SK5428	SK982 harboring pSK41	Lyon *et al.*, 1984
WBG541	Sm^R/Nb^R derivative of NCTC8325-4	Townsend *et al.*, 1983
WBG4515	Fs^R/Rf^R derivative of NCTC8325-4	Townsend *et al.*, 1983
Plasmids		
pSK41	Gm^R, Tb^R, Km^R, Nm^R, *tra*⁺, conjugative multiresistance plasmid	Berg *et al.*, 1998
pSK5632	pSK1-based *S. aureus* (Cm^R)/*E. coli* (Ap^R) shuttle vector	Grkovic *et al.*, 2003
pSK6877	pSK41 *oriT*-string cloned into *Hin*dIII and *Bam*HI of pSK5632	This study
pSK6879	pCA347 *oriT*-string cloned into *Bam*HI of pSK5632	This study
pSK6881	pSK156 *oriT*-string cloned into *Bam*HI of pSK5632	This study

[a] Sm^R, streptomycin resistance; Nb^R, novobiocin resistance; Fs^R, fusidic acid resistance; Rf^R, rifampin resistance; Gm^R, gentamycin resistance; Tb^R, tobramycin resistance; Km^R, kanamycin resistance; Nm^R, neomycin resistance; Cm^R, chloramphenicol resistance; Ap^R, ampicillin resistance; *tra*⁺ conjugative transfer genes.

Table 2.2. Relaxase-*in trans* Mobilization of Plasmids Containing *oriT* Sites

		Transfer frequency*	
Plasmid	***oriT* site**	**pSK41**	**pSK5632 derivative**
pSK5632	none	9.9×10^{-5}	Not detected
pSK6877	pSK41	1.4×10^{-4}	2.9×10^{-5}
pSK6879	pCA347	6.6×10^{-5}	Not detected
pSK6881	pSK156	8.7×10^{-5}	Not detected

* Transfer frequencies are presented as per-donor frequencies and are the average of three experiments.

REFERENCES

1. **Liu MA**, **Kwong SM**, **Jensen SO**, **Brzoska AJ**, **Firth N**. 2013. Biology of the staphylococcal conjugative multiresistance plasmid pSK41. Plasmid **70**:42–51.

2. **Pérez-Roth E**, **Kwong SM**, **Alcoba-Florez J**, **Firth N**, **Méndez-Alvarez S**. 2010. Complete nucleotide sequence and comparative analysis of pPR9, a 41.7-kilobase conjugative staphylococcal multiresistance plasmid conferring high-level mupirocin resistance. Antimicrobial Agents and Chemotherapy **54**:2252–2257.

3. **Climo MW**, **Sharma VK**, **Archer GL**. 1996. Identification and characterization of the origin of conjugative transfer (oriT) and a gene (nes) encoding a single-stranded endonuclease on the staphylococcal plasmid pGO1. J Bacteriol **178**:4975–4983.

4. **McDougal LK**, **Fosheim GE**, **Nicholson A**, **Bulens SN**, **Limbago BM**, **Shearer JES**, **Summers AO**, **Patel JB**. 2010. Emergence of resistance among USA300 methicillin-resistant Staphylococcus aureus isolates causing invasive disease in the United States. Antimicrobial Agents and Chemotherapy **54**:3804–3811.

5. **Shearer JES**, **Wireman J**, **Hostetler J**, **Forberger H**, **Borman J**, **Gill J**, **Sanchez S**, **Mankin A**, **Lamarre J**, **Lindsay JA**, **Bayles K**, **Nicholson A**, **O'Brien F**, **Jensen SO**, **Firth N**, **Skurray RA**, **Summers AO**. 2011. Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. G3 (Bethesda) **1**:581–591.

6. **Berg T**, **Firth N**, **Apisiridej S**, **Hettiaratchi A**, **Leelaporn A**, **Skurray RA**. 1998. Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. J Bacteriol **180**:4350–4359.

7. **Edwards JS**, **Betts L**, **Frazier ML**, **Pollet RM**, **Kwong SM**, **Walton WG**, **Ballentine WK**, **Huang JJ**, **Habibi S**, **Del Campo M**, **Meier JL**, **Dervan PB**, **Firth N**, **Redinbo MR**. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **110**:2804–2809.

8. **Weigel LM**, **Clewell DB**, **Gill SR**, **Clark NC**, **McDougal LK**, **Flannagan SE**, **Kolonay JF**, **Shetty J**, **Killgore GE**, **Tenover FC**. 2003. Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. Science **302**:1569–1571.

9. **Diep BA**, **Gill SR**, **Chang RF**, **Phan TH**, **Chen JH**, **Davidson MG**, **Lin F**, **Lin J**, **Carleton HA**, **Mongodin EF**, **Sensabaugh GF**, **Perdreau-Remington F**. 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus. Lancet **367**:731–739.

10. **Zhu W**, **Clark N**, **Patel JB**. 2013. pSK41-Like Plasmid Is Necessary for Inc18-Like *vanA* Plasmid Transfer from Enterococcus faecalis to Staphylococcus aureus *In Vitro*. Antimicrobial Agents and Chemotherapy **57**:212.

11. **Grohmann E**, **Muth G**, **Espinosa M**. 2003. Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev **67**:277–301.

12. **Projan SJ**, **Archer GL**. 1989. Mobilization of the relaxable Staphylococcus aureus plasmid pC221 by the conjugative plasmid pGO1 involves three pC221 loci. J Bacteriol **171**:1841–1845.

13. **Apisiridej S**, **Leelaporn A**, **Scaramuzzi CD**, **Skurray RA**, **Firth N**. 1997. Molecular analysis of a mobilizable theta-mode trimethoprim resistance plasmid from coagulase-negative staphylococci. Plasmid **38**:13–24.

14. **Smillie C**, **Garcillán-Barcia MP**, **Francia MV**, **Rocha EPC**, **la Cruz de F**. 2010. Mobility of plasmids. Microbiol Mol Biol Rev **74**:434–452.

15. **Caryl JA**, **Thomas CD**. 2006. Investigating the basis of substrate recognition in the pC221 relaxosome. Mol Microbiol **60**:1302–1318.

16. **Wong JJW**, **Lu J**, **Glover JNM**. 2012. Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. Mol Microbiol **85**:602–617.

17. **O'Brien FG**, **Ramsay JP**, **Monecke S**, **Coombs GW**, **Robinson OJ**, **Htet Z**, **Alshaikh FAM**, **Grubb WB**. 2014. Staphylococcus aureus plasmids without mobilization genes are mobilized by a novel conjugative plasmid from community isolates. Journal of Antimicrobial Chemotherapy **70**:649–652.

18. **O'Brien FG**, **Eto KY**, **Murphy RJT**, **Fairhurst HM**, **Coombs GW**, **Grubb WB**, **Ramsay JP**. 2015. Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in Staphylococcus aureus. Nucleic Acids Research **43**:7971–7983.

19. **Paulsen IT**, **Brown MH**, **Skurray RA**. 1998. Characterization of the earliest known Staphylococcus aureus plasmid encoding a multidrug efflux system. J Bacteriol **180**:3477–3479.

20. **Stegger M**, **Driebe EM**, **Roe C**, **Lemmer D**, **Bowers JR**, **Engelthaler DM**, **Keim P**, **Andersen PS**. 2013. Genome Sequence of Staphylococcus aureus Strain CA-347, a USA600 Methicillin-Resistant Isolate. Genome Announc **1**.

21. **Shen A**, **Lupardus PJ**, **Morell M**, **Ponder EL**, **Sadaghiani AM**, **Garcia KC**, **Bogyo M**. 2009. Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. PLoS ONE **4**:e8119.

22. **Emsley P**, **Lohkamp B**, **Scott WG**, **Cowtan K**. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66**:486–501.

23. The PyMOL Molecular Graphics System, Version 1.5.0.5, Schrödinger, LLC.

24. **Johnson M**, **Zaretskaya I**, **Raytselis Y**, **Merezhuk Y**, **McGinnis S**, **Madden TL**. 2008. NCBI BLAST: a better web interface. Nucleic Acids Research **36**:W5–9.

25. **Edwards JS**, **Betts L**, **Frazier ML**, **Pollet RM**, **Kwong SM**, **Walton WG**, **Ballentine WK**, **Huang JJ**, **Habibi S**, **Del Campo M**, **Meier JL**, **Dervan PB**, **Firth N**, **Redinbo MR**. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **110**:2804–2809.

26. **Grkovic S**, **Brown MH**, **Hardie KM**, **Firth N**, **Skurray RA**. 2003. Stable low-copy-number Staphylococcus aureus shuttle vectors. Microbiology (Reading, Engl) **149**:785–794.

27. **Lujan SA**, **Guogas LM**, **Ragonese H**, **Matson SW**, **Redinbo MR**. 2007. Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proceedings of the National Academy of Sciences of the United States of America **104**:12282–12287.

28. **Marraffini LA**, **Sontheimer EJ**. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322**:1843–1845.

29. **Schommer NN**, **Gallo RL**. 2013. Structure and function of the human skin microbiome. Trends Microbiol **21**:660–668.

CHAPTER 3: INHIBITION OF NICKING ENZYME IN STAPHYLOCOCCI (NES)

Introduction

Conjugative plasmid transfer (CPT) is the predominant mechanism through which antibiotic resistance spreads in human pathogens (1). There is a great need to understand the mechanism of CPT in different bacterial and plasmid systems and use that knowledge to design ways to specifically target resistant bacteria or prevent the future spread of that resistance. As the initiator of CPT and a highly conserved element, the relaxase is an attractive target for such work (1). Indeed, inhibitors for relaxase proteins have been previously characterized, although they have not shown the desired efficacy *in vivo* (2).

The enzyme of interest is NES, the relaxase on the plasmid pSK41. We hope that by targeting the relaxase, we might design inhibitors that have one of two effects. First, by inhibiting the relaxase and disrupting CPT, the inhibitor might have a specific antibiotic effect against bacteria containing the relaxase-encoding plasmid. CPT is a tightly regulated process that involves opening a pore (the type 4 secretion system discussed in Chapter 1) between two bacterial cells as well as modulating the form of resistance-encoding DNA (double-stranded plasmid to single-stranded T-strand back to a double-stranded plasmid as described in Chapter 1). Thus, disruption of this process and specifically, DNA processing is likely to trigger key

regulatory mechanisms such as those responding to increased amounts of free DNA ends and may lead to cell death. Second, if the inhibitor does not have an antibiotic effect, by disrupting CPT, we might eliminate plasmid-encoded resistances from that bacterial population. Since many humans are permanently colonized with *S. aureus*, eliminating resistances from their infections and/or colonizing microbes would decrease the likelihood of future resistant infections and transfer of resistance to other colonizers (3). The benefit to this mechanism is that bacteria are unlikely to develop resistance to such treatments quickly.

NES is a single-tyrosine relaxase that uses several key features to bind and cleave the DNA found in the origin-of-transfer (*oriT*) of pSK41. The first feature is two protein loops, Hairpin Loop 1 and 2, that form contacts with the DNA hairpin formed upstream of the *nic* site in the *oriT.* The role of these protein loops in NES function is analyzed in Chapter 2 and they are shown to be essential for proper cleavage and ligation of DNA. Therefore, we set out to characterize pyrrole-imidazole (Py-Im) polyamides inhibitors which can be designed to bind GC-rich sequences with high affinity in order to disrupt protein-DNA interactions in this region (4, 5).

Another important region for NES function is the HUH-motif that binds a divalent metal ion and is important for coordination of the phosphate backbone of the *oriT* DNA prior to cleavage (Chapter 1). Edwards *et al.* showed that without a metal bound in this position, NES is unable to cleave DNA, suggesting that a small molecule that could specifically disrupt this metal coordination would inhibit NES function and CPT (6). The Cohen laboratory at University of California, San Diego,

has compiled a chelator fragment library (CFL) based on a variety of metal binding groups in order to facilitate fragment-based drug design (FBDD) against metalloenzymes such as NES (7). This approach is ideal for this system as we have not been able to develop a high-throughput screening (HTS) assay for relaxase activity and FBDD is a more efficient exploration of chemically diverse space that can lead to higher ligand efficiencies than traditional HTS methods (7). The refined library CFL-1.1 contains 96 metal chelators with two to four donor atoms for metal binding including picolinic acids, hydroxyquinolones, pyrimidines, hydroxypyrones, hydroxypyridinones, salicylic acids, hydroxamic acids, and sulfonamides (Figure 3.3, 7, 8). We set out to characterize these compounds and their derivatives for the ability to disrupt the metal coordination of the HUH-motif of NES in order to disrupt NES function and therefore conjugative plasmid transfer.

Materials and Methods[2]

Polyamide Synthesis

Polyamides were synthesized using microwave-assisted solid-phase synthesis on oxime resin as reported previously (9). Following cleavage from oxime resin, compounds were purified by preparative HPLC. The identity and purity of each polyamide were confirmed by MALDI-TOF and analytical HPLC analysis. Match Polyamide 1 ImβImPy-(R)α-NH2γ-PyβImPy-(+): MS (MALDI-TOF) calculated for

[2] The Polyamide Synthesis, DNA Duplex Melting Temperature Analysis, and Polyamide Inhibition sections were previously published as a portion of the manuscript Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM, Walton WG, Ballentine WK, Huang JJ, Habibi S, Del Campo M, Meier JL, Dervan PB, Firth N, Redinbo MR. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America 110:2804–2809.

C50H70N21O19 [M+H]+ 1108.6, found 1108.9. Mismatch Polyamide 2 ImImImPy-(R)α-NH2γ-PyPyPyPy-(+): MS (MALDI-TOF) calculated for C56H72N23O9 [M+H]+ 1210.6, found 1209.9. Polyamides were suspended in 100% DMSO at 10 mM and stored at -80°C.

DNA Duplex Melting Temperature Analysis

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. An aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 at pH 7.0 was used as an analysis buffer. Oligonucleotides (0.1-mM stock solutions dissolved in 10 mM Tris·Cl, 0.1 mM EDTA, pH 8.0) were purchased from Integrated DNA Technologies. DNA duplexes and hairpin polyamides (1 or 2) were mixed to a final concentration of 1 μM and 1.2 μM, respectively, for each experiment. Before analysis, samples were heated to 90 °C and cooled to a starting temperature of 25 °C with a heating rate of 5 °C/min for each ramp. Denaturation profiles were recorded at λ = 260 nm from 25 °C to 90 °C with a heating rate of 0.5 °C/min. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile and represent the average of two independent measurements. All melting temperature shifts (ΔT_m) are calculated relative to a standardized naked control oligonucleotide.

Polyamide Inhibition

DNA cleavage assays were conducted *in vitro* with either the relaxase domain of NES (1–220) or the full-length (1-665) enzyme as described in Chapter 2. For polyamide inhibition, 2.5 μL of 20 μM OriTHP37 DNA substrate was incubated with

2.5 µL of polyamide at 20x the reaction concentration or 100% DMSO for 15 min. Each reaction contained 1 µL of the incubated DNA mixture in place of the 1 µL of DNA alone previously used.

Chelator Fragment Library Synthesis

Most of the fragments in the CFL-1.1 were obtained from commercial sources (Sigma Aldrich, Acros). The remaining compounds (B10, D1, D2, D3, D5, D8, D9, D10, D11, D12, E3, E4, E5, E7, E8, E9, E10, E11, E12, and G6) were prepared as described in Agrawal *et al.* and Jacobsen *et al.* (7, 8). The 4HPT1 derivative library was synthesized as described in Agrawal *et al.* and Garner *et al.* using a microwave-assisted synthetic procedure (7, 10). Each compound was suspended in 100% DMSO to a concentration of 50 mM (5mM for C10 and H1) and stored at 4°C.

Chelator Fragment Library Inhibition

DNA cleavage assays were conducted with full-length NES similarly as described in Chapter 2. Briefly, 1 µL 15.2 µM NES protein and 1 µL chelator at 10x the final chelator concentration or 100% DMSO were incubated in 7 µL EMSA buffer at room temperature for 15 minutes. 1 µL of 10 µM 5'-FAM labeled DNA substrate (OriTHP37- 5'-ACGCGAACGGAACGTTCGCATAAGTGCGCCCTTACGG-3') was then added and the resulting reaction incubated at 37°C for 1 hour and quenched by the addition of 2X running buffer (0.01% xylene cyanol, 0.01% Bromophenol Blue, 85% formamide, 20 mM EDTA, 2X TAE, 0.2% SDS). The resulting 20 µL reactions were run through a denaturing 16% polyacrylamide gel.

Results

Polyamide Inhibition[3]

The dependence of conjugative plasmid transfer of pSK41 on intact protein Loops 1 and 2 suggests the potential for inhibiting relaxase function and limiting the spread of antibiotic resistance genes through disruption of those NES-DNA interactions. This led us to investigate inhibition of NES by Py-Im polyamides, a class of sequence-specific minor-groove-binding compounds that can be programmed to bind GC-rich sequences with high affinity (Figure 3.1A, 11). Match polyamide 1 was designed to bind selectively to the 5'- GCGAA-3' sequence contacted by the essential Loop 1 of NES (see Chapter 2, Figure 3.1A). As a control the activity of this compound was compared to a mismatch polyamide 2, which shows lower affinity for the 5'-GCGAA-3' sequence (Figure 3.1A and B). The DNA cleavage activity of the relaxase 1-220 region of NES was eliminated by match polyamide 1 at 50 μM (Figure 3.2A). Interestingly, 2.5 to 50 μM concentrations of the mismatch control polyamide 2 caused a significant increase in product formation by the relaxase (Figure 3.2A). This result is similar to those obtained when the Loop 1 and 2 regions of the relaxase were deleted (6). Taken together, these data indicate that, for the isolated relaxase domain, disrupting the protein-DNA interactions around the DNA hairpin can significantly shift the enzyme's cleavage-religation equilibrium. This effect, however, is not seen when intact, 1-665 NES is used.

[3] The Polyamide Inhibition section was previously published as a portion of the manuscript Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM, Walton WG, Ballentine WK, Huang JJ, Habibi S, Del Campo M, Meier JL, Dervan PB, Firth N, Redinbo MR. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America 110:2804–2809.

With full-length 1-665 NES, mismatch polyamide 2 showed no significant effect, while the match polyamide 1 significantly reduced DNA cleavage activity at 25 μM and eliminated activity at 50 μM (Figure 3.2B). The difference between the relaxase and full-length results is likely a result of the C-terminal domain regulation as explored in Chapter 2. A closer examination of NES inhibition by concentrations of match polyamide 1 between 5 and 50 μM produced inhibition curves and IC_{50} values of 18 and 21 μM for the full-length and relaxase regions of NES, respectively (Figure 3.2C). Thus, full-length NES is inhibited by a polyamide targeted to the hairpin minor groove DNA sequence and is resistant to the effects of the non-specific polyamide. Taken together, these results confirm that full-length NES acts in a manner distinct from the isolated relaxase with respect to catalytic activity, as well as confirming the importance of the DNA hairpin-NES Loop 1 interaction explored in Chapter 2. The data also suggest that with further development, small molecules capable of inhibiting NES-DNA interactions could disrupt NES function to prevent antibiotic resistance transfer.

Fragment-Based Chelator Inhibition

Chelator Fragment Library-1.1

Previous data both with NES and other relaxases have shown that the metal coordinated by the HUH motif is essential for relaxase nicking (6, 12, 13). Therefore, we set out to identify chelator based inhibitors of NES from the Chelator Fragment Library-1.1 (CFL-1.1) introduced in Agrawal *et al.* and refined in Jacobsen *et al* (7, 8). The Cohen lab provided CFL-1.1, a 96 compound library that was screened at 100 μM for effects on NES cleavage to identify the most promising candidates

(Figure 3.3, 3.4). Previous screens with this library have been conducted at 1 mM but even with the more stringent requirement of 60% inhibition at 100 μM, we identified 17 hits for a 17.7% hit rate.

Compounds in this library are arranged into lettered sets which share the same core structure, usually an aromatic ring component (Figure 3.3). At 100 μM, the A compounds, picolinic acids, showed varying efficacy ranging from no significant effect (A2, A3, A4, A6, A7, A9, A11, A12) to 100% inhibition (A8, Figure 3.4A). Compounds A1 (71% inhibition), A5 (61% inhibition), A8 (100% inhibition), and A10 (62% inhibition) were chosen for further characterization. The B compounds, quinolones, also showed varying efficacy ranging from no significant effect (B5, B8, B9, B11) to a 75% reduction in activity (B12, Figure 3.4B). Compounds B2 (73% inhibition), B10 (66% inhibition), and B12 (75% inhibition) were chosen for further characterization. While two C compounds, pyrimidines, showed statistically significant decreases in NES activity, only one compound showed efficacy above the 60% inhibition threshold for future analysis (Figure 3.4C). Inhibitor C5 showed a 79% reduction in enzyme activity and moved forward for further characterization. None of the D compounds, hydroxypyrones, exhibited efficacy above the 60% inhibition threshold; the most potent compound, D2, showed on a 52% reduction in enzyme activity (Figure 3.4D). The E compounds, hydroxypyridinones, showed varying efficacy ranging from no effect (E1, E3, E4, E5, E6, E8, E9, E10) to an 87% reduction in activity (E7, Figure 3.4E). Compounds E2 (85% inhibition), E7 (87% inhibition), E11 (60% inhibition), and E12 (82% inhibition) were selected for further characterization. The F compounds, salicylic acids, showed

very poor efficacy with only F9 showing a decrease in activity greater than the DMSO control ($p=0.0061$) and none showing a reduction in activity of 60% or greater (Figure 3.4F). The G compounds, miscellaneous structures, were the most effective class with G3 showing 100% inhibition, G6 and G7 showing 93% and 91% inhibition respectively, and G10 showing a 65% reduction in activity (Figure 3.4G). Finally, the H compounds, also miscellaneous, showed poor activity with only H1 (62% inhibition) showing statistically significant reduction in NES activity beyond the effects of DMSO ($p=0.0095$, Figure 3.4H).

The compounds chosen for future study were screened at 0.1, 1, 10 and 100 μM. Compounds A1, A5, A8, and A10 are shown in Figure 3.5A. Only Inhibitor A8 showed efficacy at 10 μM with a 48% reduction in activity that was not statistically significant compared to DMSO treatment. Compounds B2, B10, B12, and C5 did not show efficacy at 10 μM or lower concentrations (Figure 3.5B). B10 did show a slight decrease in activity but did not reach significance. Inhibitor E2 exhibited a 77% reduction in activity at 10 μM as well as small but significant effects at 1 and 0.1 μM (Figure 3.5C). This prompted further characterization of this backbone via derivatization. Despite E7's 87% reduction in NES activity at 100 μM, it showed only a 57% reduction at 10 μM and no significant effect at 0.1 and 1 μM (Figure 3.5C). However, because of its backbone similarity to E2 and efficacy at 10 μM, we chose to move forward with derivatization of this compound as well. E11 and E12 did not show efficacy in reducing enzyme activity at 10 μM and therefore were not considered for further characterization (Figure 3.5C). Compounds G3, G6, and G7 were the most promising compounds screened; they showed a 93%, 79% and 80%

reduction in activity, respectively (Figure 3.5D). G3 and G6 showed modest but significant effects at 1 μM and only G7 exhibited an effect, although not significant, at 0.1 μM. Despite these results, G3, G6, and G7 were not considered for further derivatization as previous work has shown them to be a general chelator of metals from all protein. Therefore, without attaching an additional fragment to target the molecule to NES, treatment with these compounds would cause major side effects both for the *S. aureus* being treated as well as the infected host. Compounds G10 and H1 did not exhibit significant inhibitory effects at 10 μM and therefore were not considered for further characterization.

HPT1 Library

Derivatives of E7 make up the HPT1 library shown in Figure 3.6. To test for increased efficacy, these 7 compounds were screened in NES cleavage assays at 0.1, 1, 10, and 100 μM. The 4HPT1.C7 compound showed increased efficacy at 100 μM where it eliminated NES cleavage activity (Figure 3.6B). However, while NES cleavage activity was significantly reduced at 0.1, 1, and 10 μM, it did not show a dose dependent reduction suggesting 4HPT1.C7 is not an effective inhibitor. 4HPT1.F11 was more effective in decreasing NES cleavage activity at 100 μM than the parent E7 compound; however, it did not show any effect on activity at 10 μM and caused an increase in activity at 0.1 and 1 μM. Larger derivatives 4HPT1.C10, 4HPT1.E6, 4HPT1.F1, 4HPT1.F3, and 4HPT1.F9 showed decreased efficacy at 100 μM as compared to E7, with 4HPT1.F1 having no effect on NES cleavage (Figure 3.6B and C). We conclude that derivatization of E7 will not increase the efficacy of inhibitors against NES activity.

DMD Library

The DMD library is composed of derivatives of the E2 compound with the structures and names as shown in Figure 3.7A. We screened these 5 derivatives at 0.1, 1, 10, 50, and 100 μM for increased efficacy as compared to the parent E2 compound. All compounds in the DMD library eliminated NES activity at 100 μM and were moderately more effective at 50 μM, although the reduction in activity beyond that of E2 did not reach significance (Figure 3.7B). All of the DMD compounds maintained efficacy at 10 μM but only DMD-11 was more effective than E2, although again not reaching significance (Figure 3.7B). At 0.1 and 1 μM, none of the DMD library compounds showed a significant reduction in activity. However, the modest reduction in activity caused by DMD-14 was significant when compared to E2 and DMD-13 showed a significant increase in activity. The efficacy of these CFL compounds, especially DMD-11, supports previous data about the importance of the bound metal to relaxase function and is proof of principle that a chelating inhibitor is an effective modulator of relaxase activity. With further development such chelator fragments paired with other inhibitory or targeting fragments could be capable of selectivity inhibiting NES function to prevent antibiotic resistance transfer.

Discussion

Antibiotic resistance is largely spread among bacterial populations by conjugative plasmid transfer (14, 15). This transfer is facilitated by the relaxase enzyme which is often, but not always, encoded on the plasmid to be transferred. We focused on the NES relaxase enzyme encoded on the plasmid pSK41. Previous

characterization of this enzyme identified several regions essential for proper function of NES (Chapter 2, 6).

One essential region is two protein loops (Loop 1 and Loop 2) that bind to the minor and major groove, respectively, of the hairpin formed by the pSK41 *oriT* DNA. We establish that disruption of the minor groove contacts with a polyamide inhibits NES activity *in vitro* (Figure 3.2). Minor-groove-targeted Py-Im match polyamide 1, designed to bind to the GC-rich DNA minor groove of the NES substrate hairpin, demonstrated IC_{50} values of 18 and 21 μM against the full-length and relaxase domain of NES, respectively, with respect to DNA cleavage. While it is possible that optimization of polyamide architecture and functionalization will yield organism-specific uptake in *S. aureus*, these compounds may be able to bind similar DNA sites throughout the genome resulting in off-target effects. Indeed, we found that the minimum inhibitory concentration (MIC) of match polyamide 1 is 16 μM for cultured *S. aureus*.

A second region essential for NES function is the HUH-motif and the bound divalent metal. We show that small molecule chelators successfully inhibit NES cleavage activity *in vitro* (Figure 3.4-3.7). Using CFL-1.1 we found 17 small molecules that show a 60% or greater reduction in NES activity at 100 μM (Figure 3.4). Of these 17 hits, six compounds (A8, E2, E7, G3, G6, G7) also showed significant effects at 10 μM and were considered for further optimization. Due to their similarity, E2 and E7 were derivatized and screened for increased efficacy. E7 derivatives did not show increased, dose-dependent efficacy. In addition, E7 and related compounds have been found to be potent inhibitors of other

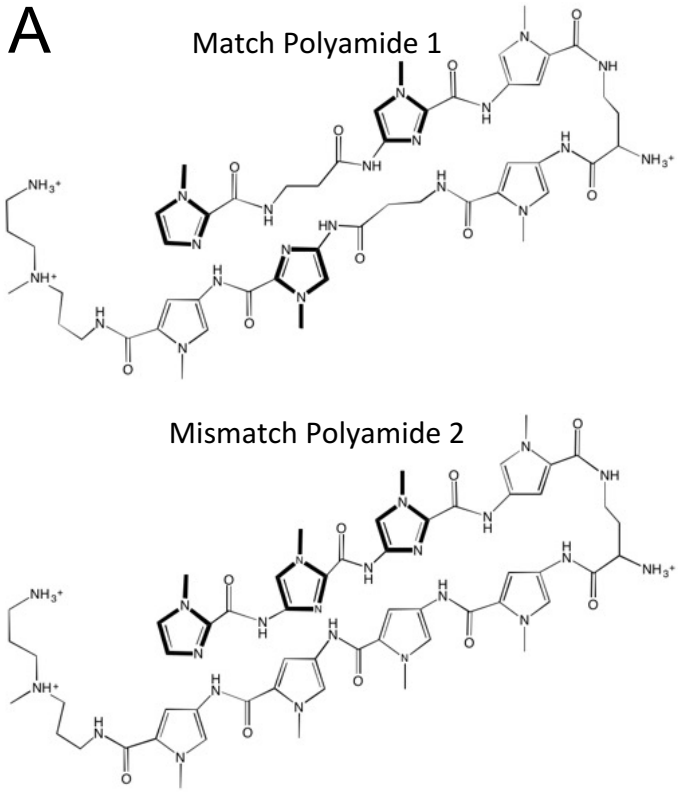
metalloenzymes, suggesting it would have many off target effects *in vivo* (7, 8, 10). The best E2 derivative compound, DMD-11, has an estimated IC_{50} of 10 μM with respect to DNA cleavage. Further derivatization and screens for selectivity are needed to improve and verify the efficacy of this compound class. Despite concerns with other compounds tested against NES, compounds from CFL-1.1 have been shown to have high selectivity for their target proteins suggesting it is possible to develop E2 into a NES selective molecule (16).

One path for optimization via fragment based drug design is to join a chelator fragment with an additional inhibitory element. Another essential region for DNA binding by NES is the guanine-26 nucleotide base that is buried in a cavity where it makes two hydrogen-bonding contacts with NES residues. This cavity is approximately 10 Å from the metal bound at the HUH-motif; thus, we hypothesize that a purine nucleotide or mimic linked to an optimized chelator fragment would yield a selective and efficient inhibitor of NES activity. A co-crystal structure or docking model of NES bound to E2 or DMD-11 would assist both with optimization of the metal binding group as well as designing the linker to a purine mimic. Thus far, crystal trials of the relaxase domain in complex with a shorter *oriT* DNA oligonucleotide and chelator fragment hits have yielded crystals with poor diffraction, the best diffraction being to 8 Å.

Taken together, the data presented in this chapter confirm that inhibition of NES activity is possible via several different protein sites. Indeed, targeting NES activity as a means to limit DNA transfer has been seen in *Staphylococcus epidermis* strain RP62a which contains a functional CRISPR spacer targeting the

nes gene (17). This, in combination with the *in vivo* conjugative transfer data with mutant forms of NES presented in Chapter 1 and Edwards *et al.,* suggests that both the polyamide and chelator inhibitors would serve as inhibitors of DNA transfer via their impact on NES function.

A



B

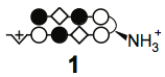
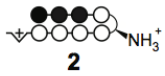
Polyamide	duplex 5'-TGGT **GCGAA** CCT-3' T_m / °C	ΔT_m / °C
—	59.2 (±1.3)	—
 **1**	68.6 (±0.7)	9.3
 **2**	62.2 (±0.7)	2.9

Figure 3.1. Polyamide Structure
A. Structures of the Match Polyamide 1 and Mismatch Polyamide 2
B. DNA thermal melting with polyamides. Shown is melting temperature (T_m) of the DNA duplex alone and with the Match and Mismatch Polyamides (1 and 2, respectively). The change in T_m is also indicated. One strand of the duplex oligo is shown, with the target sequence depicted in blue.
This data was originally published in Edwards *et al.* (6)

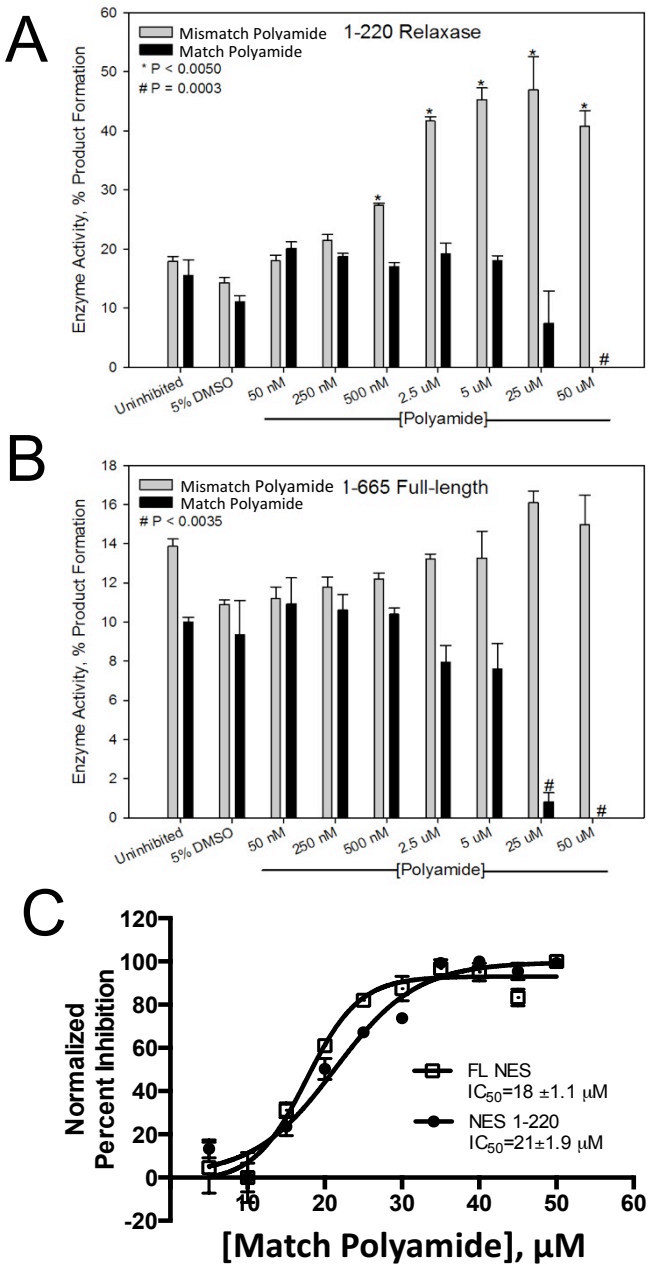


Figure 3.2 Polyamide Impact on NES Activity
A. Inhibition of NES relaxase *in vitro* with Match Polyamide 1 (black) and Mismatch Polyamide 2 (grey). Polyamides were dissolved in 5% DMSO.
B. Inhibition of Full-length NES *in vitro* with Match Polyamide 1 (black) and Mismatch Polyamide 2 (grey). Polyamides were dissolved in 5% DMSO.
C. Eighteen- and 21-μM IC50 values of the Match Polyamide 1 for the full-length and relaxase forms of NES, respectively.
This data was originally published in Edwards *et al.* (6)

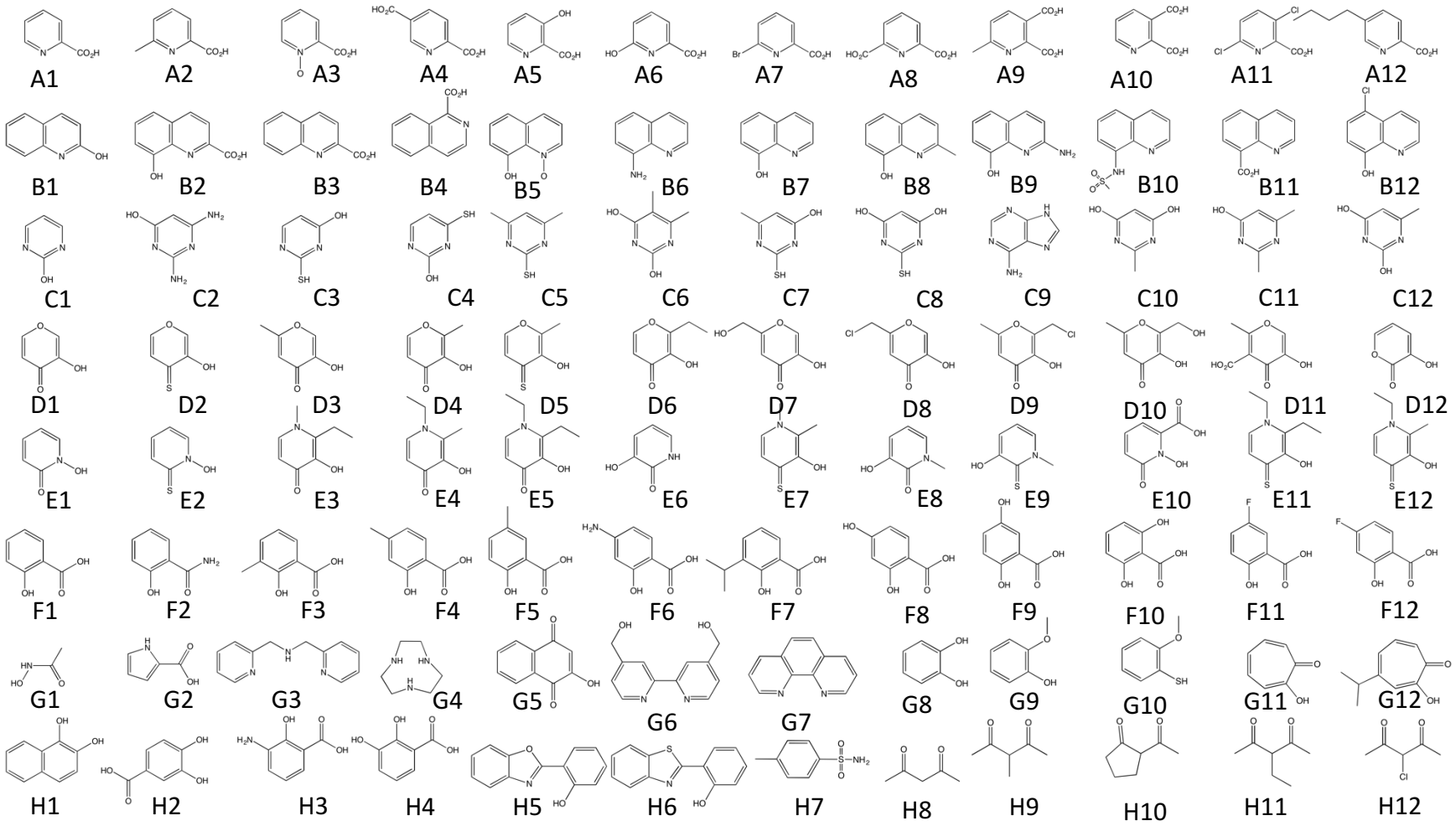
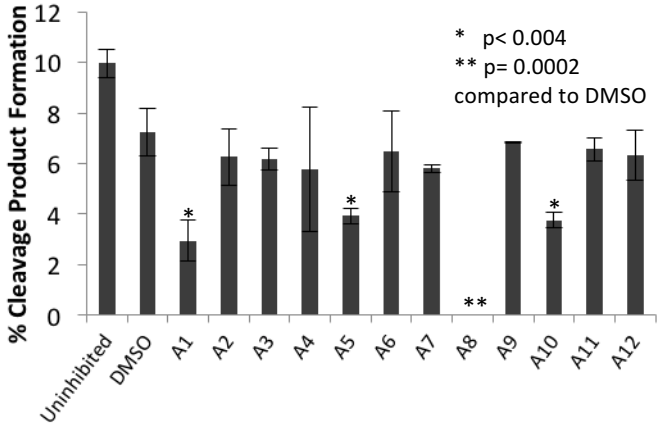
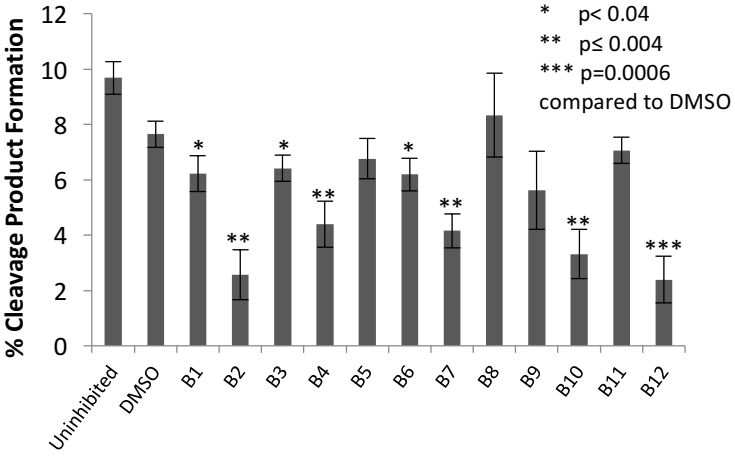


Figure 3.3. Chelator Fragment Library- 1.1 (CFL-1.1)
Structures of all fragments in CFL-1.1. Each row corresponds to a metal binding group class: picolinic acids (A), quinolones (B), pyrimidines (C), hydroxypyrones (D), hydroxypyridinones (E), and salicylic acids (F). Compounds in row G and H are miscellaneous.

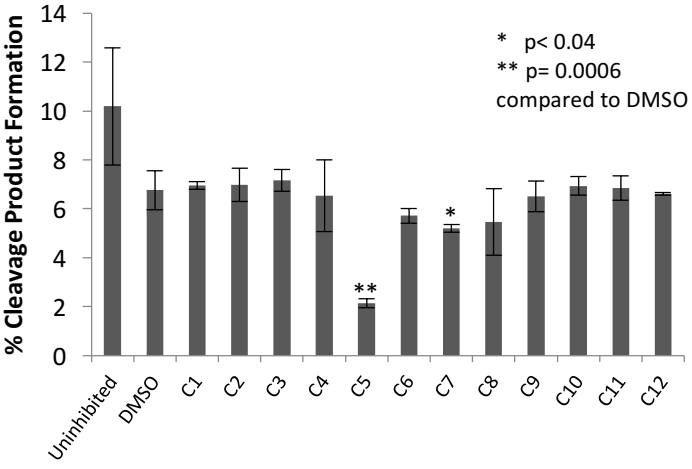
A


% Cleavage Product Formation
0
2
4
6
8
10
12
* p< 0.004
** p= 0.0002
compared to DMSO
Uninhibited
DMSO
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12

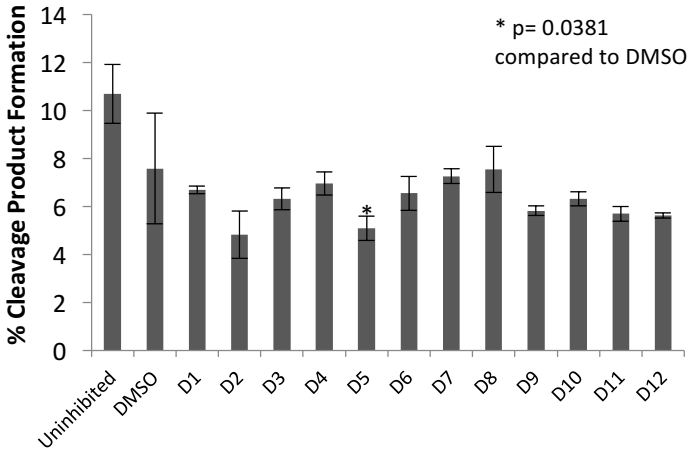
B


% Cleavage Product Formation
0
2
4
6
8
10
12
* p< 0.04
** p≤ 0.004
*** p=0.0006
compared to DMSO
Uninhibited
DMSO
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B12

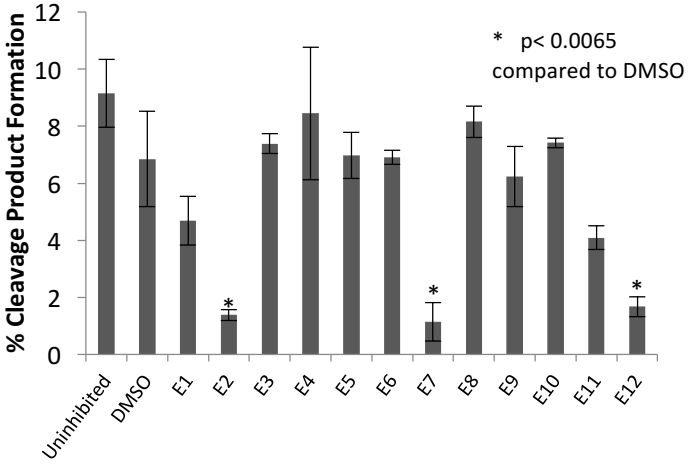
C


% Cleavage Product Formation
0
2
4
6
8
10
12
14
* p< 0.04
** p= 0.0006
compared to DMSO
Uninhibited
DMSO
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12

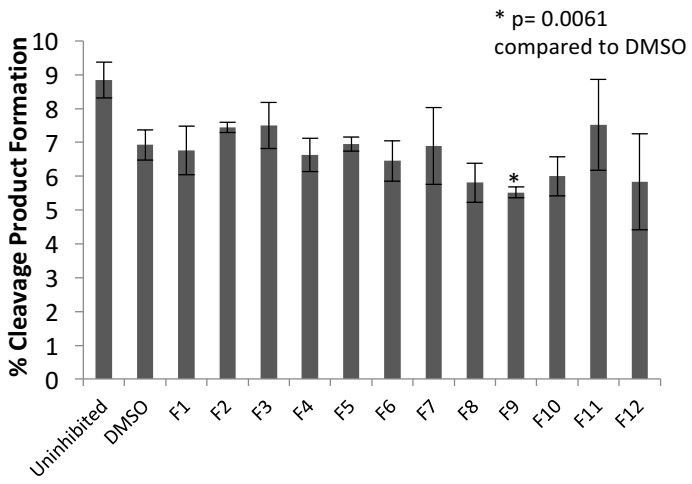
D


% Cleavage Product Formation
14
12
10
8
6
4
2
0
* p= 0.0381
compared to DMSO
*
Uninhibited
DMSO
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12

E


% Cleavage Product Formation
12
10
8
6
4
2
0
* p< 0.0065
compared to DMSO
*
*
*
Uninhibited
DMSO
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11
E12

F


% Cleavage Product Formation
10
9
8
7
6
5
4
3
2
1
0
* p= 0.0061
compared to DMSO
*
Uninhibited
DMSO
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12

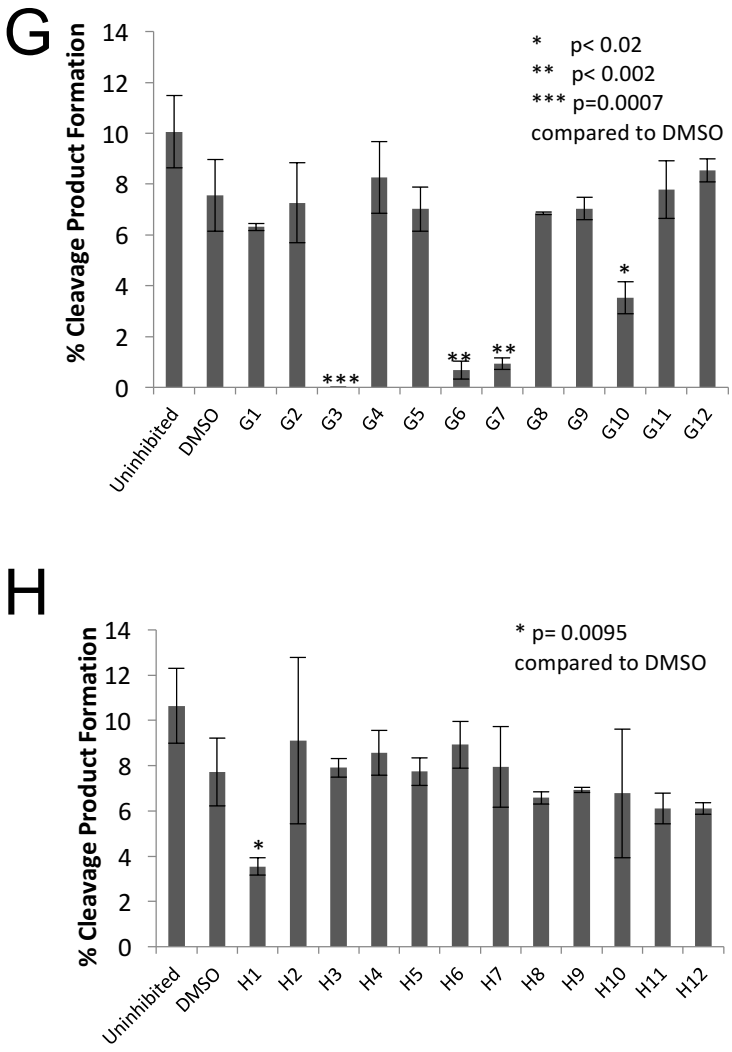
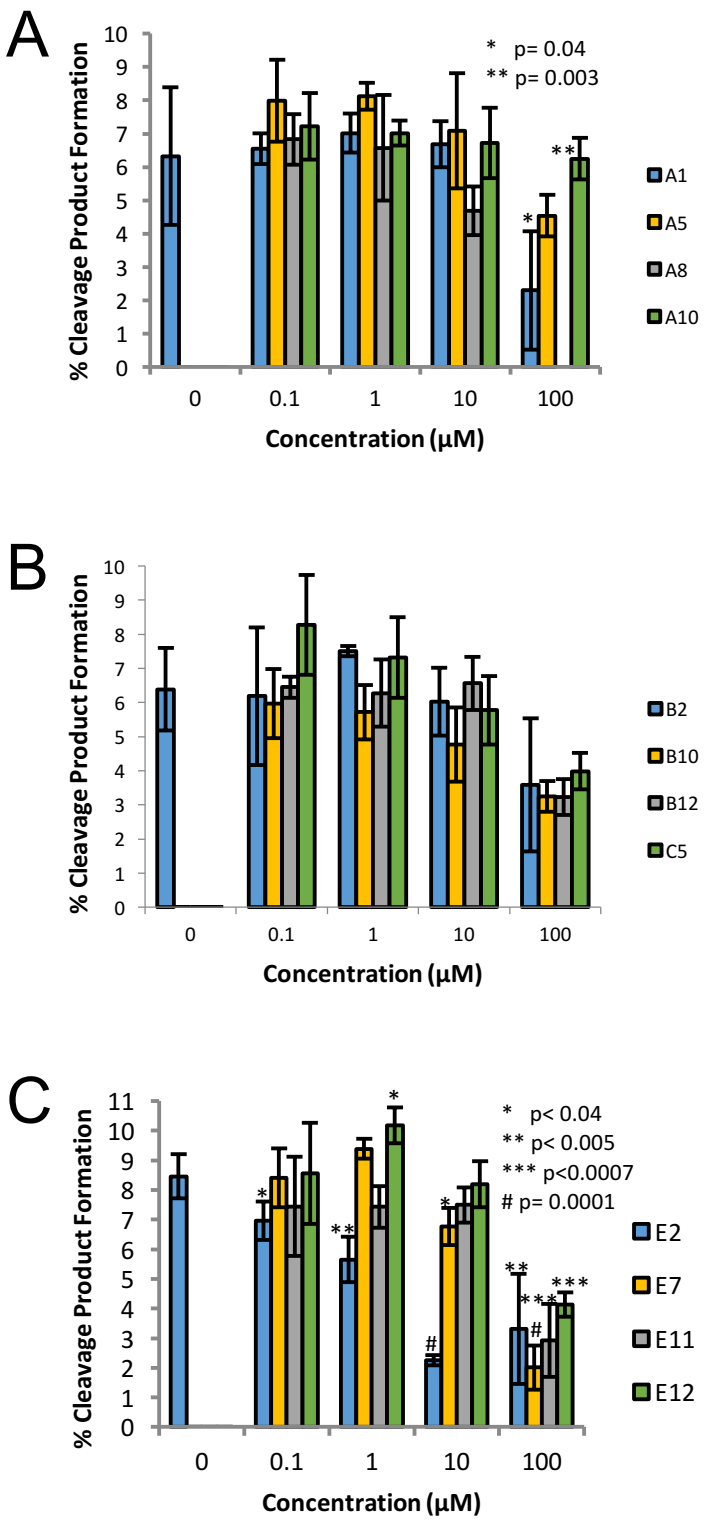


Figure 3.4. 100μM Screen of CFL-1.1
A. Effect of compounds A1-A12 on NES cleavage product formation.
B. Effect of compounds B1-B12 on NES cleavage product formation.
C. Effect of compounds C1-C12 on NES cleavage product formation.
D. Effect of compounds D1-D12 on NES cleavage product formation.
E. Effect of compounds E1-E12 on NES cleavage product formation.
F. Effect of compounds F1-F12 on NES cleavage product formation.
G. Effect of compounds G1-G12 on NES cleavage product formation.
H. Effect of compounds H1-H12 on NES cleavage product formation.


A
% Cleavage Product Formation
Concentration (µM)
0
0.1
1
10
100
* p= 0.04
** p= 0.003
A1
A5
A8
A10
B
% Cleavage Product Formation
Concentration (µM)
B2
B10
B12
C5
C
% Cleavage Product Formation
Concentration (µM)
* p< 0.04
** p< 0.005
*** p<0.0007
# p= 0.0001
E2
E7
E11
E12

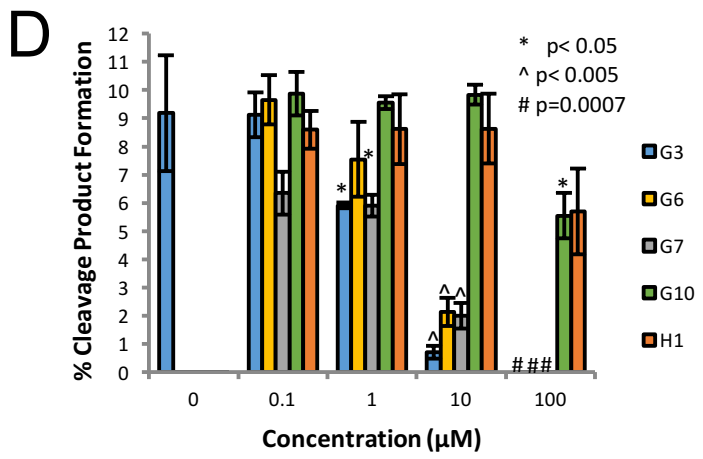
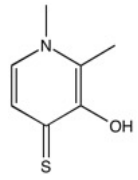
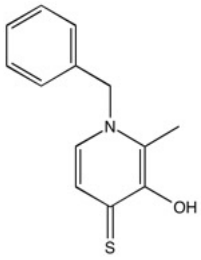


Figure 3.5. Chelator Impact on NES Cleavage Product Formation
A. Effect of compounds A1, A5, A8, and A10 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. 0 µM data is DMSO control and p-values are calculated against this control.
B. Effect of compounds B2, B10, B12, and C5 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. 0 µM data is DMSO control and p-values are calculated against this control.
C. Effect of compounds E2, E7, E11, and E12 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. 0 µM data is DMSO control and p-values are calculated against this control.
D. Effect of compounds G3, G6, G7, G10, and H1 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. 0 µM data is DMSO control and p-values are calculated against this control.

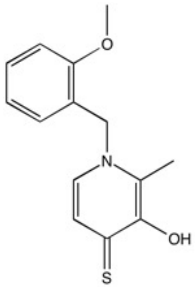
A



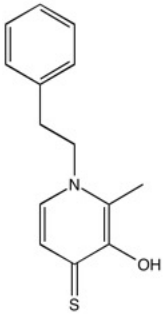
E7



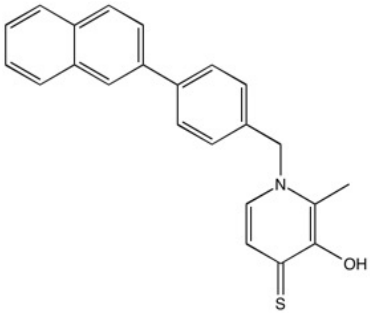
4HPT1.C7



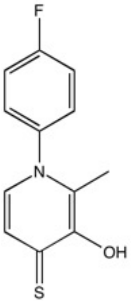
4HPT1.C10



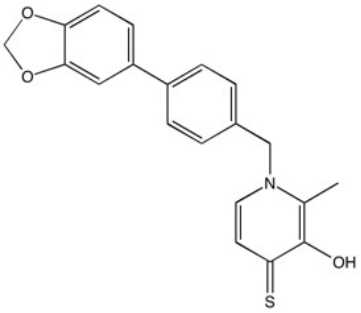
4HPT1.E6



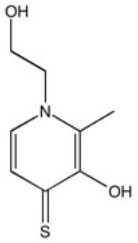
4HPT1.F1



4HPT1.F3



4HPT1.F9



4HPT1.F11

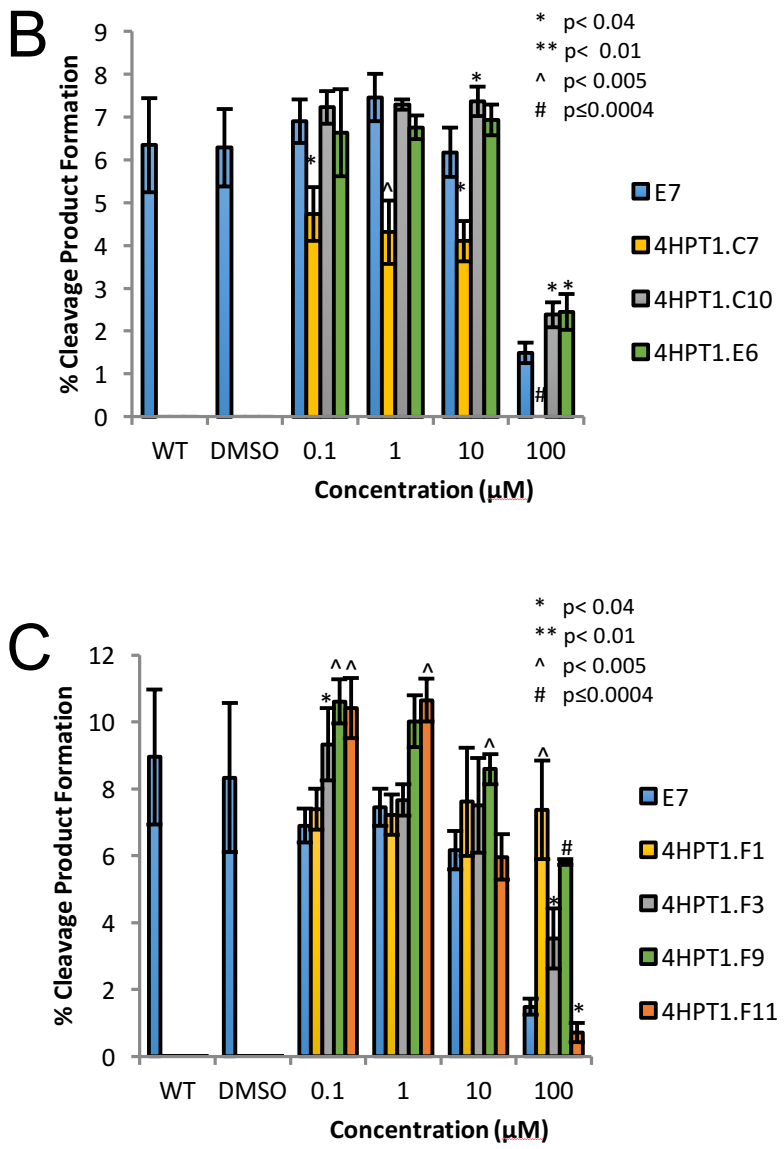


Figure 3.6. HPT1 E7 Derivative Library
A. Structures of E7 derivatives for the HPT1 Library
B. Effect of compounds 4HPT1.C7, 4HPT1.C10, and 4HPT1.E6 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. Indicated p-values are calculated with reference to the E7 reaction.
C. Effect of compounds 4HPT1.F1, 4HPT1.F3, 4HPT1.F9, 4HPT1.F11 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. Indicated p-values are calculated with reference to the E7 reaction.

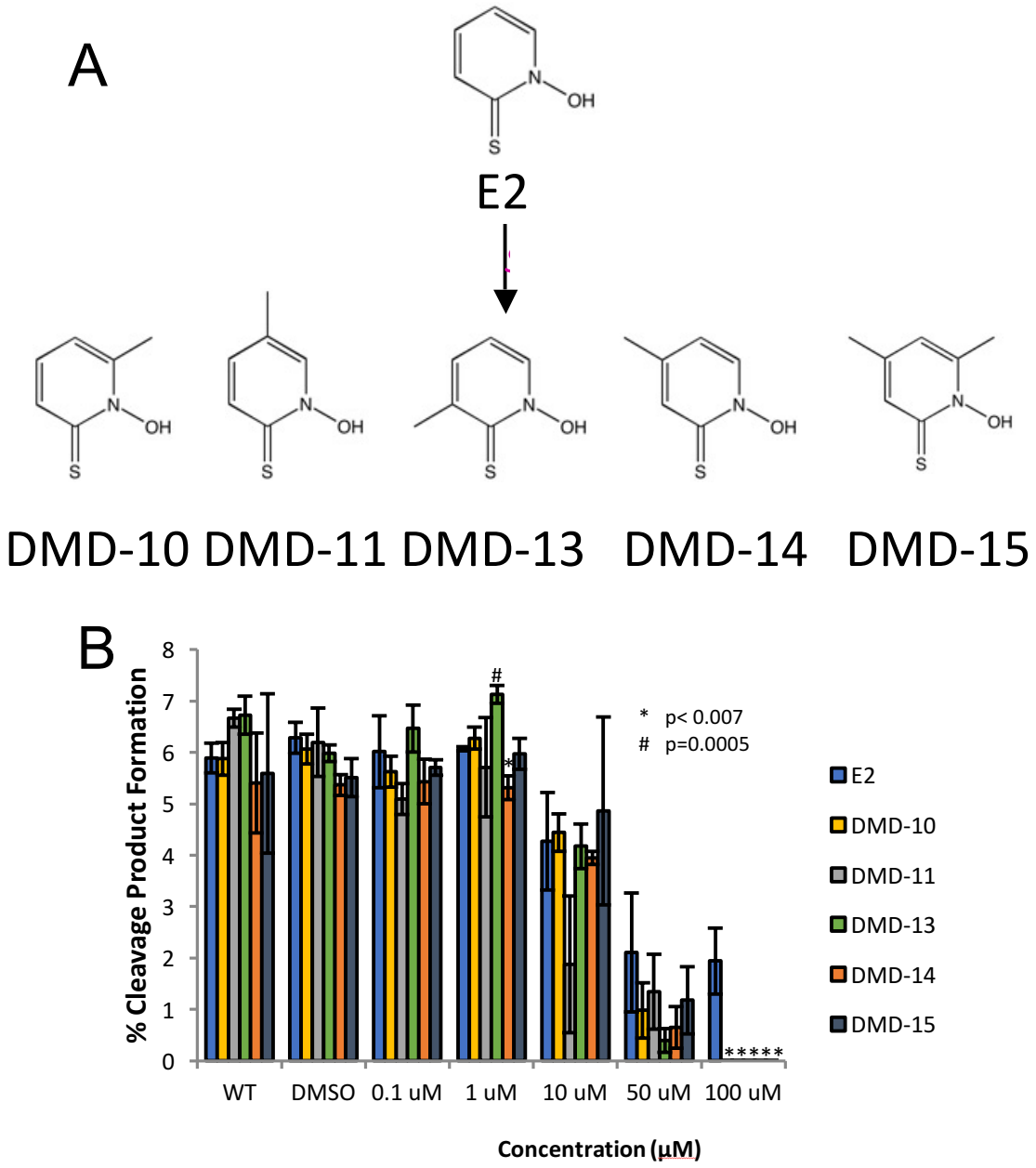


Figure 3.7. HPT1 E2 Derivative Library
A. Structures of E2 derivatives for the DMD Library
B. Effect of compounds DMD-10, DMD-11, DMD-13, DMD-14, and DMD-15 at 0.1, 1, 10, and 100 µM on NES cleavage product formation. Indicated p-values are calculated with reference to the E2 reaction.

REFERENCES

1. **Baquero F**, **Coque TM**, **la Cruz de F**. 2011. Ecology and evolution as targets: the need for novel eco-evo drugs and strategies to fight antibiotic resistance. Antimicrobial Agents and Chemotherapy **55**:3649–3660.

2. **Lujan SA**, **Guogas LM**, **Ragonese H**, **Matson SW**, **Redinbo MR**. 2007. Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proceedings of the National Academy of Sciences of the United States of America **104**:12282–12287.

3. **Otto M**. 2013. Blocking the spread of resistance. Sci Transl Med **5**:184fs17.

4. **Kielkopf CL**, **White S**, **Szewczyk JW**, **Turner JM**, **Baird EE**, **Dervan PB**, **Rees DC**. 1998. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. Science **282**:111–115.

5. **Pilch DS**, **Poklar N**, **Gelfand CA**, **Law SM**, **Breslauer KJ**, **Baird EE**, **Dervan PB**. 1996. Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. Proceedings of the National Academy of Sciences of the United States of America **93**:8306–8311.

6. **Edwards JS**, **Betts L**, **Frazier ML**, **Pollet RM**, **Kwong SM**, **Walton WG**, **Ballentine WK**, **Huang JJ**, **Habibi S**, **Del Campo M**, **Meier JL**, **Dervan PB**, **Firth N**, **Redinbo MR**. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **110**:2804–2809.

7. **Agrawal A**, **Johnson SL**, **Jacobsen JA**, **Miller MT**, **Chen L-H**, **Pellecchia M**, **Cohen SM**. 2010. Chelator fragment libraries for targeting metalloproteinases. ChemMedChem **5**:195–199.

8. **Jacobsen JA**, **Fullagar JL**, **Miller MT**, **Cohen SM**. 2011. Identifying Chelators for Metalloprotein Inhibitors Using a Fragment-Based Approach **54**:591–602.

9. **Li BC**, **Montgomery DC**, **Puckett JW**, **Dervan PB**. 2012. Synthesis of Cyclic Py-Im Polyamide Libraries. J Org Chem.

10. **Garner AL**, **Struss AK**, **Fullagar JL**, **Agrawal A**, **Moreno AY**, **Cohen SM**, **Janda KD**. 2012. 3-Hydroxy-1-alkyl-2-methylpyridine-4(1 H)-thiones: Inhibition of the Pseudomonas aeruginosaVirulence Factor LasB. ACS Med Chem Lett **3**:668–672.

11. **Dervan PB**, **Edelson BS**. 2003. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. Curr Opin Struct Biol **13**:284–299.

12. **Nash RP**, **Habibi S**, **Cheng Y**, **Lujan SA**, **Redinbo MR**. 2010. The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. Nucleic Acids Research **38**:5929–5943.

13. **Xia S**, **Robertus JD**. 2009. Effect of divalent ions on the minimal relaxase domain of MobA. Arch Biochem Biophys **488**:42–47.

14. **Smillie C**, **Garcillán-Barcia MP**, **Francia MV**, **Rocha EPC**, **la Cruz de F**. 2010. Mobility of plasmids. Microbiol Mol Biol Rev **74**:434–452.

15. **Wong JJW**, **Lu J**, **Glover JNM**. 2012. Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. Mol Microbiol **85**:602–617.

16. **Fullagar JL**, **Garner AL**, **Struss AK**, **Day JA**, **Martin DP**, **Yu J**, **Cai X**, **Janda KD**, **Cohen SM**. 2013. Antagonism of a zinc metalloprotease using a unique metal-chelating scaffold: tropolones as inhibitors of P. aeruginosa elastase. Chem Commun (Camb) **49**:3197–3199.

17. **Marraffini LA**, **Sontheimer EJ**. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322**:1843–1845.

CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS FOR CHARACTERIZATION OF pSK41 AND ITS RELAXASE, NES

Conclusions

The Redinbo lab has had a long-lasting interest in conjugative plasmid transfer (CPT) and the previous three chapters present work characterizing the plasmid pSK41 and its relaxase, NES. Specifically, initial characterization conducted with the relaxase domain of NES by Jon Edwards has been extended to characterization of the full-length protein. As discussed in Chapter 2, the C-terminal domain of NES functions as a regulator of the relaxase domain as opposed to the independent function of most relaxase-associated C-terminal domains. In addition, NES protein Hairpin Loops 1 and 2 are established as important features for proper DNA cleavage and religation, but not DNA binding. Having confirmed the importance of this protein-DNA interaction, the importance of the DNA sequence that these protein loops contact was then analyzed and it was noted that NES is able to process variant hairpins found in the *oriT*s of the plasmids pSK156 and pCA347 *in vitro*. While this relaxase-*in trans* mechanism had previously been shown for the pWBG749 family of plasmids in *S. aureus*, this work suggests that other plasmid families including pSK41 likely utilize the same system to facilitate conjugation of plasmids that were previously considered non-conjugative. Chapter 3 takes

advantage of protein regions that were noted as important in Chapter 1 and 2 in order to design inhibitors of NES activity. We hope this work will lead to drugs that will decrease the spread of antibiotic resistance in *S. aureus* communities. These experiments have also spurred several additional questions about how NES facilitates conjugative plasmid transfer not only of pSK41, but also of pSK156 and pCA347. These additional hypotheses are explored below.

Future Directions

Formation of the DNA Hairpin

Chapter 2 establishes that the protein-DNA contacts formed to the DNA hairpin just upstream of the *nic* site of the *oriT* are important for proper DNA cleavage and religation (strand transfer) by NES (Figure 2.3). However, it has not been established when during CPT the inverted repeats of the *oriT* form this DNA hairpin *in vivo*. While this is difficult to directly measure *in vivo*, NES activity *in vitro* on DNA with and without the hairpin would further establish when the hairpin is essential and therefore, when it is likely to form. An ideal experiment would be to compare binding, cleavage, and strand transfer rates between identical linear and annealed, hairpin forms of the *oriT* sequence. Because hairpin formation is favorable when the DNA is not in its double-stranded form, it may be difficult to obtain a clean linear portion so an alternative approach is to test oligonucleotides with a disrupted DNA hairpin (altered bases in the inverted repeats), no hairpin (a short oligonucleotide that starts directly after the hairpin), and with only the second inverted repeat closest to the *nic* site (Table 4.1). Edwards *et al.* showed that oligonucleotides with shortened or eliminated first inverted repeats (the inverted

repeat furthest from the *nic* site) were not able to out-compete binding of the full *oriT* sequence to the relaxase domain of NES (1). However, an oligonucleotide with a scrambled first inverted repeat was able to bind to the relaxase domain, even in the presence of the wild-type *oriT* oligonucleotides. If a similar trend holds true in direct binding experiments with the full-length NES protein, it would suggest that hairpin formation is not required for binding of NES and initiation of CPT but that it is important that the full length of both inverted repeats be exposed in a single-stranded form.

Similar cleavage and strand transfer experiments will characterize the importance of hairpin formation in those steps of CPT as well. Preliminary cleavage studies of oligonucleotides with a disrupted DNA hairpin (altered bases in the inverted repeats), no hairpin (a short oligonucleotide that starts directly after the hairpin), and with only the second inverted repeat (half hairpin) suggest that NES is able to cleave these oligonucleotides but with increased rates as seen in the loop-deletion mutations of the NES enzyme (Figure 4.1). However, cleavage rates are necessarily dependent on binding so it is also important to know binding affinities for each oligonucleotide before drawing conclusions. A full analysis of binding, cleavage, and strand transfer will help form a model for hairpin formation upon which further experiments can build. As we learn more about pSK41 and CPT, it will also be important to consider how other MOB proteins might influence hairpin formation.

Characterization of the pSK41 Relaxosome

Chapter 2 demonstrates that *in vitro* NES is able to bind, cleave, and ligate *oriT* DNA from plasmids pSK156 and pCA347 (Figure 2.5). However, in *in vivo*

conjugation assays, NES and the other MOB genes encoded on pSK41 were not able to facilitate transfer of plasmids containing the pSK156 or pCA347 *oriT* (Chapter 2 Table 2.2). We hypothesize that this is because an accessory protein is required that differentiates between different sub-types in the pSK41 family and this accessory protein encoded on pSK41 is incongruent with the pSK156 and pCA347 *oriT*s. This relaxase-*in trans* conjugation system was first shown for the plasmid pWBG749 and O'Brien *et al.* showed that a small putative DNA-binding accessory protein, SmpO, is responsible for differentiating sub-types of plasmids containing pWBG749-like *oriT* sequences (2, 3). NES is the only protein in the pSK41 relaxosome that has been characterized so it is difficult to hypothesize what this accessory protein might be for the pSK41 system. Homolog models and preliminary work by the Redinbo lab have identified pSK41 proteins Orf86 and Orf90 as putative DNA binding proteins so they are candidates for this accessory protein but there are likely other candidates in the Orf and Tra regions of pSK41. Systematic characterization of the MOB genes of pSK41 will not only aid in identifying this accessory protein, but will also reveal additional information about the conjugation process.

Optimization of Chelator Fragment Inhibitors

Data presented in Chapter 3 characterizes preliminary NES inhibitors from a chelator fragment library and proposed optimization of these inhibitors via attachment of a guanine-mimic to block binding of the uniquely positioned nucleotide Gua-26 in the pSK41 *oriT*. Edwards *et al.* showed that *oriT* DNA with an abasic site at position 26 was cleaved at an increased rate by the NES relaxase domain and the

structure revealed the pocket in which this nucleotide specifically binds (Figure 4.2, 1). Further basic characterization of this DNA-protein interaction in the context of the full-length protein will provide insight on how to optimize a small molecule to bind in this pocket. Binding, cleavage, and strand transfer assays with DNA containing an abasic site in place of Gua-26 with the full-length NES protein will serve as a proof of principle that disruption of this interaction will have a significant effect on NES function. If we are able to identify a small molecule that targets this position, the effects will likely mimic the NES activity seen with the abasic site DNA. The chelating inhibitors could even be tested in this system to further explore the combination approach before synthesis of new small molecules.

There are also several mutations that would produce clashes between the newly introduced side chain and the guanine nucleotide. M3W, G90L, and N93M mutations will introduce larger side chains that will likely disrupt the binding pocket into which the guanine nucleotide normally sites. The clashed formed by M3W is shown in Figure 4.3. This will again serve as a proof of principle that disruption of proper guanine binding will have a significant effect on NES function. In addition, it will provide information on the flexibility of this binding site and parts of the binding pocket that it will be beneficial to disrupt.

An effective inhibitor of the Gua-26 binding pocket will need to have a high binding affinity in order to bind prior to DNA binding or outcompete the Gua-26 nucleotide. As shown in Figure 4.2, there are several contacts formed between NES and the Gua-26 nucleotide. While most of the contacts are to backbone oxygens, the impact of the hydrogen bond to N129 could be explored by mutation to leucine or

alanine. Unfortunately, most of the other surrounding amino acid side chains are not amendable to forming hydrogen bonds so most optimization will depend on optimizing backbone contacts.

An additional consideration for the chelator fragment library is specificity of the chelators for different metals. While we have confirmed the metal bound in our purified NES enzyme is nickel, it is unclear if this is physiologically relevant or a consequence of purification on a nickel column (1). Therefore, chelators that are effective against a nickel-bound NES *in vitro* may show differential effectiveness *in vivo.* It is possible to chelate out the nickel bound to NES with EDTA; therefore, it should be possible to replace this nickel with other metals that facilitate NES activity such as manganese and copper in order to analyze the effect of the chelators against varying metal cofactors (1).

Variation in NES Genes

Although pSK41 is the prototype plasmids for the MOB_Q class and NES is the best characterized single-tyrosine relaxase, there are several other plasmids that contain an NES gene and some variation is seen in this gene across plasmids. For example, the conjugative transfer regions of the pGO1 and pSK41 plasmids are considered to be essentially identical; however, there is a 24-nucleotide segment in the *nes* gene that differs between the two plasmids (4). This results in an 8 amino acid sequence that is different between the NES enzymes expressed by each plasmid and could have significant implications on the function of the enzyme. It has been suggested that the *nic* site of pGO1 and pSK41 differs by one nucleotide position despite the *oriT* being the same sequence. Perhaps this 8 amino acid

difference between the NES enzymes could explain this difference and characterization of that change would further understanding of NES DNA binding function.

Additionally, several plasmids that contain a *nes* gene have frame-shift mutations in the gene that result in a truncated enzyme. Truncations have been seen that result in 303, 482, or 555 amino acid proteins rather than the 665 residue full-length enzyme characterized here. It is unclear if these truncated enzymes are functional and able to facilitate CPT. Notably, the frame shift mutation resulting in a 303-residue enzyme is found on the pUSA03 plasmid in the MRSA-USA300 strain. This is the C-terminal mutant that was not able to facilitate transfer of pSK41 *in vivo* in Edwards *et al*. but it is possible that other changes in the pUSA03 plasmid allow the shorter enzyme to be sufficient for transfer (Chapter 1 Figure 1.5) (1).

Conjugative plasmids that encode *nes* genes or plasmids that move via a relaxase *in-trans* mechanism represent a large reservoir of antibiotic resistance genes. Understanding of this highly conserved enzyme that is responsible for initiation via DNA cleavage and termination via DNA ligation is key to understanding the mechanism of conjugative plasmid transfer and developing therapeutics to limit spread of genes encoding for antibiotic resistance. While work in this thesis has expanded understanding of the function and potential for inhibition of NES, there are also many questions left to answer about this enzyme across plasmids and many ways to optimize inhibition.

Table 4.1. Altered Hairpin Oligonucleotides

Oligo Name	Sequence (Hairpin Underlined)
OriTHP40	ACGCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGAT
Disrupted Hairpin	ACATAGGTAAAAATGGATAATAAGTGCGCCCTTACGGGAT
No Hairpin	ATAAGTGCGCCCTTACGGGAT
Half Hairpin	AACGTTCGCATAAGTGCGCCCTTACGGGAT

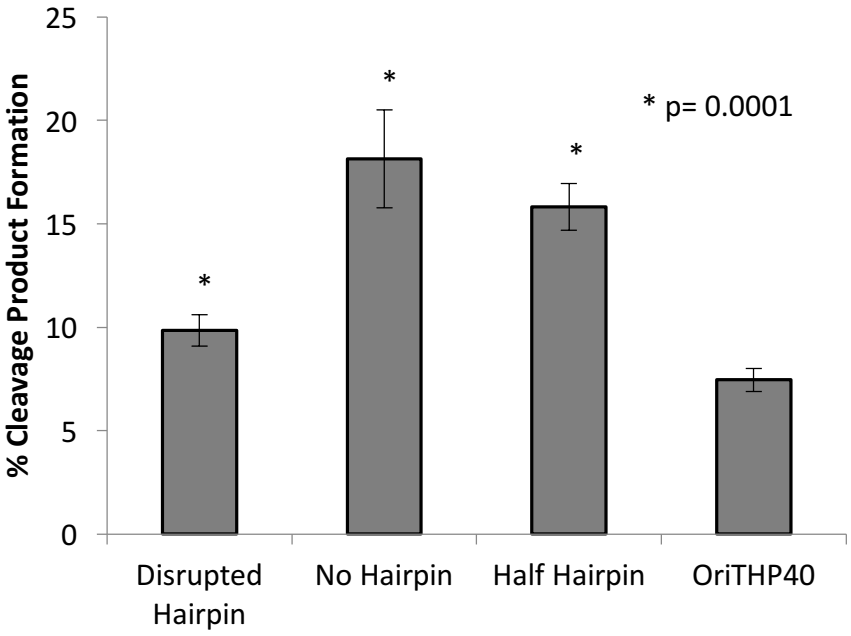


Figure 4.1. NES Cleavage of Altered Hairpin Oligonucleotides
NES cleavage activity on oligonucleotides containing the wild-type sequence (OriTHP40), a disrupted hairpin, beginning after the hairpin (no hairpin), or half of the hairpin sequence from the *oriT* of pSK41.

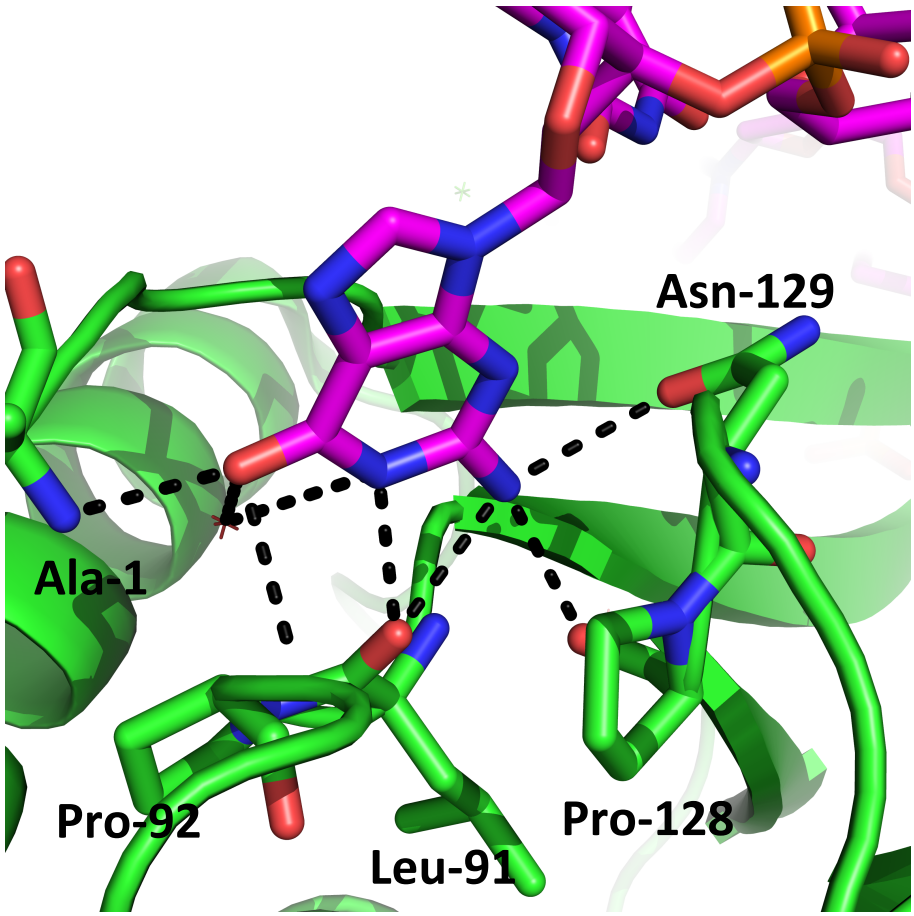


Figure 4.2. Gua-26 Binding Pocket
Guanine-26 (magenta) in the pSK41 *oriT* is coordinated by the NES protein backbone and the asparagine-129 side chain (green). PDB ID: 4HT4

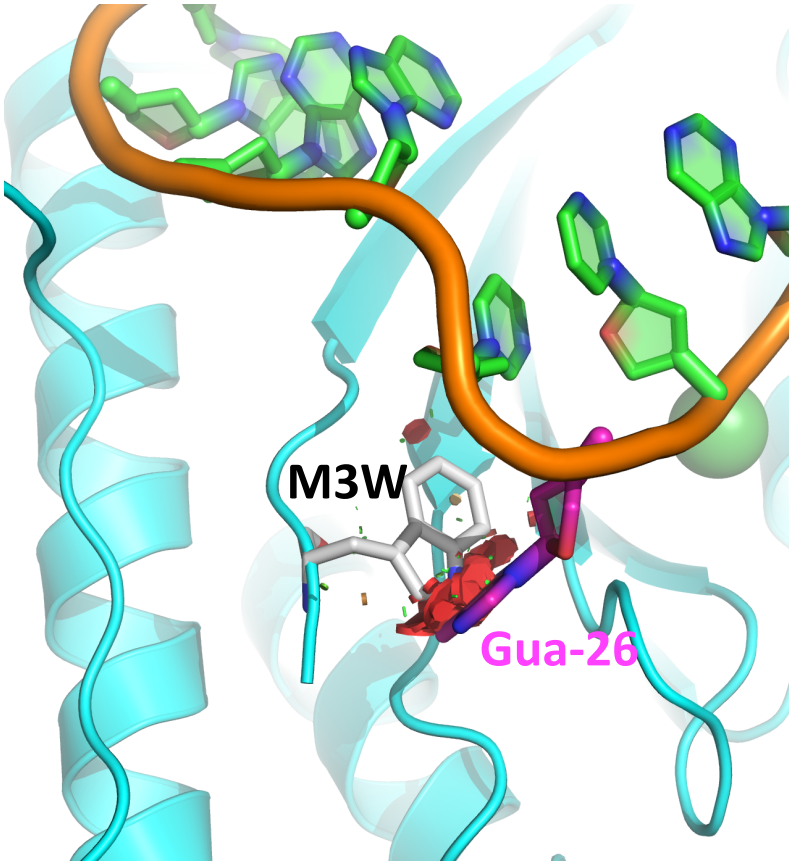


Figure 4.3. M3W Mutation Effects
Guanine-26 (magenta) in the pSK41 *oriT* clashes (red) with the M3W mutation (white) of the NES relaxase domain (cyan). PDB ID: 4HT4

REFERENCES

1. **Edwards JS**, **Betts L**, **Frazier ML**, **Pollet RM**, **Kwong SM**, **Walton WG**, **Ballentine WK**, **Huang JJ**, **Habibi S**, **Del Campo M**, **Meier JL**, **Dervan PB**, **Firth N**, **Redinbo MR**. 2013. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **110**:2804–2809.

2. **O'Brien FG**, **Ramsay JP**, **Monecke S**, **Coombs GW**, **Robinson OJ**, **Htet Z**, **Alshaikh FAM**, **Grubb WB**. 2014. Staphylococcus aureus plasmids without mobilization genes are mobilized by a novel conjugative plasmid from community isolates. Journal of Antimicrobial Chemotherapy **70**:649–652.

3. **O'Brien FG**, **Eto KY**, **Murphy RJT**, **Fairhurst HM**, **Coombs GW**, **Grubb WB**, **Ramsay JP**. 2015. Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in Staphylococcus aureus. Nucleic Acids Research **43**:7971–7983.

4. **Berg T**, **Firth N**, **Apisiridej S**, **Hettiaratchi A**, **Leelaporn A**, **Skurray RA**. 1998. Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. J Bacteriol **180**:4350–4359.

CHAPTER 5: INTRODUCTION TO β-GLUCURONIDASE- MICROBIAL ENZYME RESPONSIBLE FOR GI DRUG REACTIVATION

Microbiome

It has long been appreciated that many bacteria make the human body their home. Recent technological advances have greatly expanded our appreciation of the vast number and types of bacteria that are found on and in the human body and given rise to the human microbiome field of study. The term microbiota refers to the symbiotic bacteria that live in and on the human bodies. A recent estimate calculated that each of person maintains approximately equal numbers of human and bacterial cells with the bacterial cells concentrated largely on the skin and on the mucus membranes of the mouth and gastrointestinal tract (GI) (1). The microbiome is the collection of genomes associated with the microbiota. Because of the large variability between bacterial genomes and quick reproduction time of bacteria, the microbiome contains many more unique genes and evolves much more quickly than the human genome. While bacteria have been studied for hundreds of years, only recently has it become feasible to sequence and compile large numbers of bacterial genomes in order to explore the variations between strains and genes that are non-essential or do not impart a strong phenotype.

To better understand this vast bacterial population, the National Institutes of Health (NIH) began the Human Microbiome Project (HMP) to establish resources that would enable research to characterize the human microbiome and its role in human health and disease (2). This included collecting and sequencing microbial samples from healthy individuals across several different body sites including nasal passages, oral cavity, skin, gastrointestinal tract, and urogenital tract (2). In addition to the whole-genome sequencing metagenomics studies conducted with these microbial samples, the HMP is also working to culture bacteria from these samples in order to establish 3,000 reference strains with complete genome sequences.

Completion of the HMP and the on-going efforts of other microbiome sequencing initiatives such as the American Gut Project have highlighted the complexity and diversity of the human microbiota. Samples vary widely between individuals, likely reflecting the highly variable environments in which we live, genetic differences, and various foods and drugs that we consume (2). Samples also show significant diversity when the same individual is sampled over time (3). Only healthy individuals were sampled for the HMP in hopes of establishing a common, healthy baseline to which the microbiota associated with disease could be compared. Subsequent comparisons have suggested links between microbiota dysbiosis and autism spectrum disorders, metabolic disorders, irritable bowel syndrome, diabetes, and cardiovascular disease (4-6). These associations emphasize the importance of maintaining a healthy microbiota and limiting the use of broad-spectrum antibiotics, especially in early childhood, as this significantly alters the diversity of the GI microbiota (7, 8). More recent studies have focused on determining the mechanism

by which changes in the microbiota lead to disease, examining both which individual bacteria are responsible for the effect as well as characterizing molecular processes those bacteria facilitate by narrowing in on a particular gene or gene cluster. This has allowed for an increased appreciation of the interactions between mammalian tissues and the microbiota (9).

Glycoside Hydrolases

One microbial enzyme class that is of particular interest is glycoside hydrolases. Glycoside hydrolases are expressed by both prokaryotic and eukaryotic cells and are responsible for hydrolyzing the glycosidic bond between carbohydrate sugars or between a sugar and a non-carbohydrate moiety.

The average human diet contains a vast array of both simple and complex carbohydrates that include glycosidic sugars. Glycosidic sugars come in a variety of discrete but related structures and can also be derivatized with the attachment of a functional group such as sulfate. These varying sugars can be arranged through a variety of different bonds in an almost endless library of carbohydrate molecules that must be broken down in order to be used for both human and bacterial energy production. The 135 families of glycoside hydrolases are the primary enzymes responsible for degradation of these large carbohydrate molecules. Many of these enzymes have strict specificity for one glycosidic sugar (e.g. glucuronate vs galactose) as well as the bond linking the sugar to the next molecule (e.g. alpha- vs beta-linkage); therefore, it is not surprising that both mammals and bacteria express a wide range of glycoside hydrolase enzymes (10, 11).

β-Glucuronidase[4]

β-Glucuronidase, or GUS, enzymes are members of the glycoside hydrolase family 2 (GH2). These enzymes are found in plants, animals, and bacteria in order to cleave a beta-linked glucuronide moiety from a larger compound. As with other glycoside hydrolase enzymes, the hydrolysis of the glycosidic bond by GUS enzymes is catalyzed by two amino acid residues, usually glutamic acids. The proposed mechanism for the *E. coli* GUS is shown in Figure 5.1. This proposed mechanism is based on similarity to the lysozyme and β-galactosidase mechanisms that have been more extensively studied (12).

The human GUS enzyme is found in lysosomes where it hydrolyzes glucuronic acid from the non-reducing end of glycosaminoglycans such as chondroitin sulfate and hyaluronic acid (13). The structure of human GUS was reported in 1996 and characterized at high resolution in 2013 (13, 14). This is an essential enzyme in humans and deficiencies resulting from a mutation in the gene lead to a build-up of non-hydrolyzed glycosaminoglycans referred to as Sly syndrome (15).

Microbial GUS enzymes allow the encoding microbes to scavenge for a carbon-based energy source. There are a wide variety of compounds containing glucuronide moieties that enter the GI including glycosaminoglycans and small molecules that are conjugated to a single glucuronide sugar through phase II drug

[4] This section adapted from the previously published work Wallace BD, Roberts AB, Pollet RM, Ingle JD, Biernat KA, Pellock SJ, Kumar Benkatesh M, Guthrie L, O'Neal SK, Robinson SJ, Dollinger M, Figueroa E, McShane SR, Cohen RD, Jin J, Frye SV, Zamboni WC, Pepe-Ranney C, Mani S, Kelly L, Redinbo MR. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chemistry & Biology 22(9):1238-1249.

metabolism in the liver; however, it is still unclear what the native substrate is for these enzymes. Microbial GUS activity was first characterized in *E. coli* and *E. coli* GUS (*Ec*GUS) is often used as a reporter gene to monitor gene expression in plant, mammalian, and even bacterial cells via a robust, colorimetric assay (16, 17).

*Ec*GUS is only essential for *E. coli* bacteria when glucuronic acid is the only available carbon source (17). The structure of *Ec*GUS was determined in 2010 by the Redinbo lab and showed a functional tetramer that has been confirmed by gel filtration chromatography and multi-angle light scattering (Figure 5.2) (18, 19). In order to expand the understanding of microbial enzymes, the GUS enzymes from *Streptococcus agalactiae* (*Sa*GUS) and *Clostridium perfringens* (*Cp*GUS) have also been characterized (19). Structures of *Sa*GUS and *Cp*GUS exhibit overall structures similar to the *Ec*GUS and human GUS structures reported previously (Figure 5.2) (13, 14, 18).

Using the *Ec*GUS, *Sa*GUS, and *Cp*GUS crystal structures as a guide, we were able to define functionally relevant sequence motifs. GUS enzymes align ether-linked glucuronides for hydrolysis using hydrophobic, hydrogen bonding, and electrostatic interactions with the glucuronic acid sugar moiety. The asparagine and lysine residues that contact the carboxylic acid group unique to glucuronic acid relative to galactose are particularly critical and we have denoted this as the N-K motif (N566/567/563, K568/569/565 in Figure 5.3A). This motif is completely conserved even in more distantly related GUS enzymes from Archaeal and Thermotoga sources, as well as Bacteroides species and a gene obtained from an uncultured sample whose gene product, H11G11, was demonstrated to have GUS

activity (Figure 5.3B) (20). This motif was previously noted when Matsumura and Ellington evolved glucuronidase activity from galactosidase enzymes (21). In addition to the N-K motif, a tyrosine residue located ~4 Å from the glucuronic acid binding site appears to ensure by steric occlusion that only beta-linked substrates are processed by the two catalytic glutamic acid residues of GUS enzymes; an alpha-linkage would clash with the aromatic ring at this position (Y468 in Figure 5.3A).

Using the N-K motif as a marker for GUS enzymes within the GH2 family, we were able to identify a previously solved structure of a *Bacteroides fragilis* enzyme (RCSB: 3CMG) as a GUS enzyme rather than its originally annotated beta-galactosidase label (Figure 5.2). As shown in Figure 5.3B, this *B. fragilis* GUS (*Bf*GUS) maintains the N-K motif. Indeed, *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Bf*GUS were shown to have activity against the glucuronidated substrate *p*-nitrophenyl glucuronide (PNPG) (Table 5.1). Thus, we propose that the N-K and Y motifs should join the catalytic glutamic acid residues as structural features considered essential for GUS activity and defining GUS sequences.

The *Ec*GUS, *Sa*GUS, and *Cp*GUS enzymes have a unique motif relative to the human GUS structure: they contain a “bacterial loop” that folds over the active site (Figure 5.4A). By contrast, human GUS lacks this loop. The bacterial loops are highly divergent in sequence, sharing only 11-36% identity between *Ec*GUS, *Cp*GUS, and *Sa*GUS (Figure 5.4B). In human GUS, this region is replaced by a single residue, P415. Sequence analysis of bacterial loops in a set of microbial GUS enzymes from the NCBI database reveals that some of the enzymes contain the

loop and some do not (Figure 5.4B). *Bf*GUS seems to be an intermediate between a full bacteria loop like that seen in *Ec*GUS, *Sa*GUS, and *Cp*GUS and no loop as in human GUS.

These observations support the conclusion that the GI microbiome maintains an array of enzyme orthologs capable of scavenging glucuronic acid from a range of chemically distinct substrates. Diverse endobiotic, food, and xenobiotic glucuronides are expected to be regularly delivered to the GI tract, and thus the microbial species therein are poised to utilize these diverse sources of carbon. The addition of *Bf*GUS to the array of characterized GUS enzymes has emphasized the limited scope of the current understanding of the sequence and structure of GUS enzymes. The following chapters begin to expand current understanding of the GUS enzyme family by defining a catalog of GUS sequences found in the human microbiome and characterizing some of those unique enzymes.

Irinotecan[5]

Irinotecan (CPT-11) is a potent anticancer drug used to treat solid tumors and other malignancies (22-24). It is an essential component of FOLFIRI (where IRI indicates irinotecan), which is widely used for late-stage colorectal cancer (25, 26), and FOLFIRINOX, which was recently recommended as the first-line regimen for pancreatic cancer (27). According to the Food and Drug Administration package

[5] This section adapted from the previously published work Wallace BD, Roberts AB, Pollet RM, Ingle JD, Biernat KA, Pellock SJ, Kumar Benkatesh M, Guthrie L, O'Neal SK, Robinson SJ, Dollinger M, Figueroa E, McShane SR, Cohen RD, Jin J, Frye SV, Zamboni WC, Pepe-Ranney C, Mani S, Kelly L, Redinbo MR. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chemistry & Biology 22(9):1238-1249.

insert for irinotecan, up to 88% of patients experience diarrhea and 31% show grade 3-4 diarrhea, which requires significant medical intervention including dose reductions or treatment termination (28-30).

As shown in Figure 5.5, irinotecan is a prodrug that is activated by human carboxylestrases in the liver during phase I drug metabolism (30, 31). This active metabolite, SN-38, is a topoisomerase I poison and leads to irreversible double-strand breaks in the DNA of targeted cells (32). SN-38 is further processed by UDP-glucuronosyltransferases (UGT), in particular UGT1A1, to become the inactivated glucuronide conjugate SN-38G, which is sent to the GI tract for elimination (28-30). In the GI tract, bacterial GUS enzymes remove the glucuronic acid and thus reactivate the topoisomerase I poison SN-38, which causes GI damage and the diarrhea that is dose-limiting for irinotecan (33-36).

Several strategies have been tested to decrease the toxicity caused by irinotecan. Reduction in irinotecan dosing and administration of anti-motility medication such as loperamide are the most basic approaches for patients experiencing severe diarrhea (37). Antibiotics to target the host microbiota have been administered prior to irinotecan treatment; however, as discussed above, the microbiota plays an important role in many biological functions and disruption of those functions has significant effects, including increasing the risk of pathogenic infection in these cancer patients who are often already immune-compromised (31). Attempts to produce SN-38 analogues that cannot be glucuronidated have been pursued, but results in the clinic have been slow to materialize. Low-potency GUS inhibitors have also been explored and one tested in rats in 2004 showed limited

efficacy in reducing the GI toxicity associated with irinotecan. Therefore, the Redinbo lab set out to design high-potency inhibitors targeted to bacteria GUS enzymes in order to alleviate this irinotecan-induced diarrhea.

GUS Inhibitors

In 2010, Wallace *et al.* presented four potent and novel inhibitors of bacterial GUS enzymes identified by a 10,000-compound high-throughput screen using *Ec*GUS and showed that they significantly reduced the GI damage caused by irinotecan in mice (Inhibitors 1-4 in Table 5.2) (38, 39). An additional four novel inhibitors were explored by Roberts *et al.* in 2013 (Inhibitors 5-8 in Table 5.2). These inhibitors are non-toxic to both bacterial and mammalian cells and are selective for bacterial GUS enzymes containing the bacterial loop by more than 1,000-fold over the human enzyme orthologue (18, 40). We pinpointed the basis for the selectivity to a loop present in bacterial GUS enzymes that is missing from the human enzyme ortholog; the inhibitors bind to this loop, making them ineffective against mammalian GUS enzymes (18). In an attempt to optimize the efficacy of Inhibitor 1, compounds R1 and R3 were designed using knowledge gained through *Ec*GUS-inhibitor structures.

These compounds were also tested against *Sa*GUS, *Cp*GUS, and *Bf*GUS. All characterized compounds inhibit *Ec*GUS with K_i values below 2 μM. However, the compounds exhibit markedly different effects toward *Cp*GUS and *Sa*GUS with some compounds showing efficacy against both enzymes in addition to *Ec*GUS and other compounds showing much weaker or no efficacy against the new enzymes despite

the presence of the bacterial loop (Table 5.2). The smaller bacterial loop of *Bf*GUS is not sufficient for the activity of these compounds and *Bf*GUS is not subject to inhibition at up to 100 μM inhibitor concentrations (Table 5.2). This result is reminiscent of the no loop, mammalian GUS enzyme from *Bos taurus* (bovine, *Bt*GUS), which was also not subject to inhibition (Table 5.2). These results confirm, as seen in crystal structures, that a full bacterial loop is essential for efficacy of these inhibitors in addition for their selectivity relative to the mammalian enzymes. Despite the ineffectiveness against a subset of bacterial enzymes such as *Bf*GUS, Inhibitors 1, 5, and R1 have been shown to be effective in reducing CPT-11 induced diarrhea in mice; the decrease in diarrhea in mice following Inhibitor 1 treatment is shown in Figure 5.6A.

There were concerns that these inhibitors of microbial GUS activity would affect irinotecan pharmacokinetics and therefore decrease efficacy in patients. To address this, the Redinbo lab measured circulating plasma levels of CPT-11 (irinotecan) with and without Inhibitor 1 treatment. As shown in Figure 5.6B, Inhibitor 1 was effective at reducing the percentage of mice with CPT-11 induced diarrhea, but the circulating levels of CPT-11, SN-38, and SN-38G were not affected by Inhibitor 1 treatment. This suggests that SN-38 formed from SN-38G cleavage by bacterial GUS enzymes is not reabsorbed into the plasma and therefore, inhibitor treatment does not significantly disrupt the circulating plasma levels of this drug and should not impact tumor response to CPT-11.

With these very promising results, the Redinbo lab plans to continue to optimize these inhibitors. However, to ensure we are targeting the GUS enzymes

primarily responsible for SN-38G processing and optimizing the inhibitors against those enzymes, we need to understand what GUS enzymes are present in the GI and which interact with SN-38G. The following chapters contain the initial steps to catalog and characterize GUS enzymes from the human microbiome. These results will be coupled with microbiome sequencing studies in mice to inform future work to optimize GUS inhibitors.

Table 5.1. Catalytic Activity of GUS
Data are presented as the average over >3 experiments ± SEM for *Escherichia coli* (*Ec*GUS), *Clostridium perfringens* (*Cp*GUS), *Streptococcus agalactiae* (*Sa*GUS), and *Bacteroides fragilis* (*Bf*GUS) β-Glucuronidases. k_{cat}, catalytic rate; K_M, Michaelis constant; k_{cat}/K_M, catalytic efficiency. Originally published in Wallace *et al.* (2015) (19)

Enzyme	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
*Ec*GUS	120 ± 12	0.13 ± 0.01	920 ± 160
*Sa*GUS	80 ± 2	0.36 ± 0.03	222 ± 24
*Cp*GUS	2.6 ± 0.6	1.1 ± 0.2	2.4 ± 1
*Bf*GUS	18 ± 1	1.9 ± 0.3	9.5 ± 2

Table 5.2. *In vitro* GUS Inhibition, K_i (μM)
Enzyme inhibition properties of small molecule inhibitors against representative GUS enzymes. Grey data from Roberts *et al.* (2013), Wallace *et al.* (2010) (18, 40). Black data from Wallace *et al.* (2015) (19) *ni*: no inhibition. Data are presented as the average over >3 experiments ± SEM for *Escherichia coli* (*Ec*GUS), *Clostridium perfringens* (*Cp*GUS), *Streptococcus agalactiae* (*Sa*GUS), *Bacteroides fragilis* (*Bf*GUS), and *Bos taurus* (bovine, *Bt*GUS) β-Glucuronidases.

Inhibitor	*Cp*GUS	*Sa*GUS	*Ec*GUS	*Bf*GUS	*Bt*GUS (mammalian)
1	0.97 ± 0.1	1.4 ± 0.4	0.16 ± 0.01	*ni*	*ni*
R1	*ni*	*ni*	1.9 ± 0.5	*ni*	
R3	*ni*	*ni*	0.61 ± 0.2	*ni*	
2	1.1 ± 0.5	3.0 ± 1	0.21 ± 0.03	*ni*	*ni*
3	7.8 ± 0.9	11 ± 3	0.68 ± 0.08	*ni*	*ni*
4	24 ± 3	36 ± 5	1.4 ± 0.2	*ni*	*ni*
5	0.54 ± 0.2	0.81 ± 0.2	0.22 ± 0.04	*ni*	*ni*
6	6.1 ± 2	2.8 ± 0.3	0.67 ± 0.03	*ni*	*ni*
7	*ni*	*ni*	1.9 ± 0.02	*ni*	*ni*
8	*ni*	*ni*	0.96 ± 0.03	*ni*	*ni*

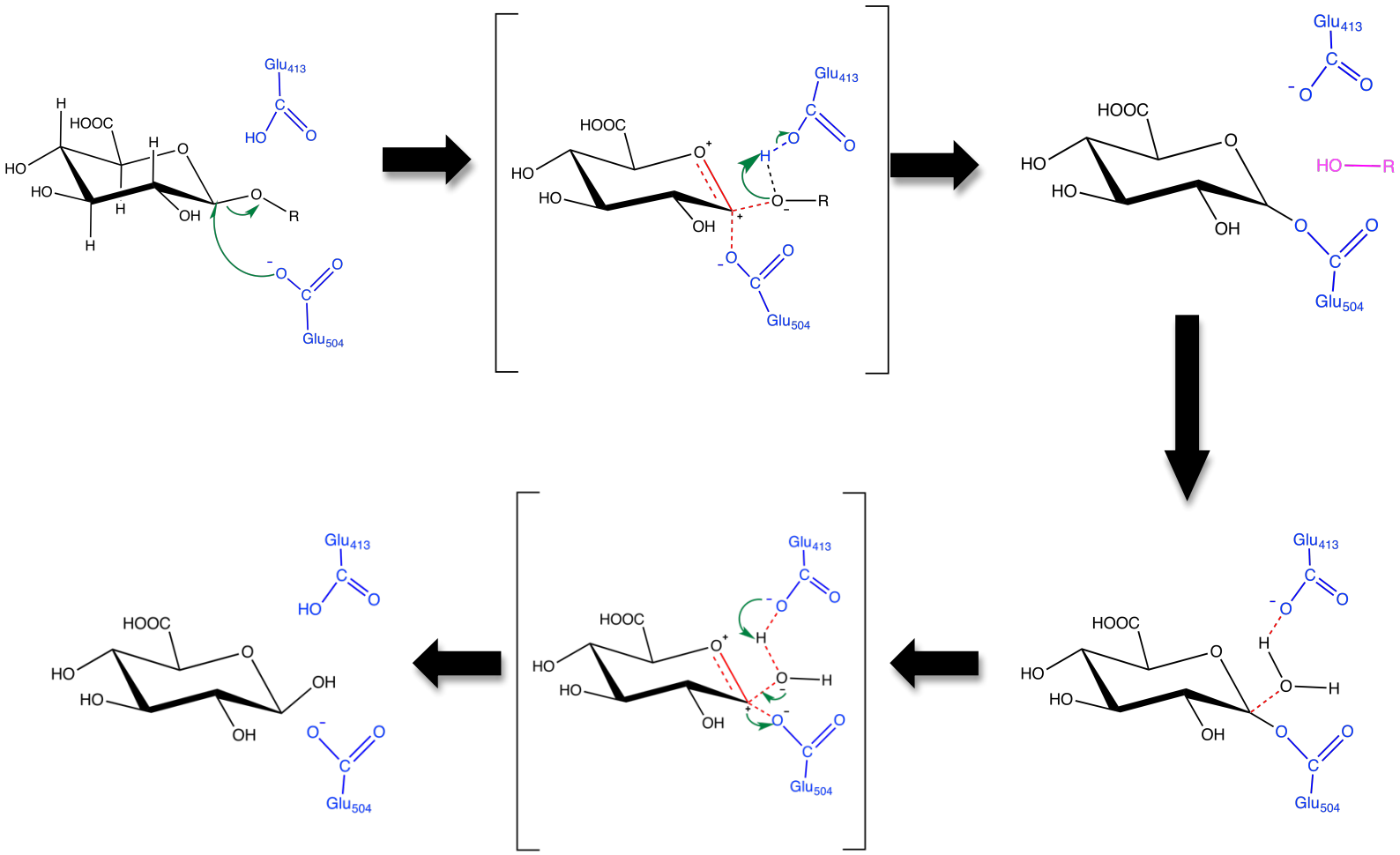

Glu413
Glu504
HOOC
HO
OH
R
H

Figure 5.1. Enzymatic Mechanism of *E. coli* GUS
Glutamic acid residues 413 and 504 are hypothesized to serve as the key catalytic residues in order to process a generic glucuronide-linked substrate.

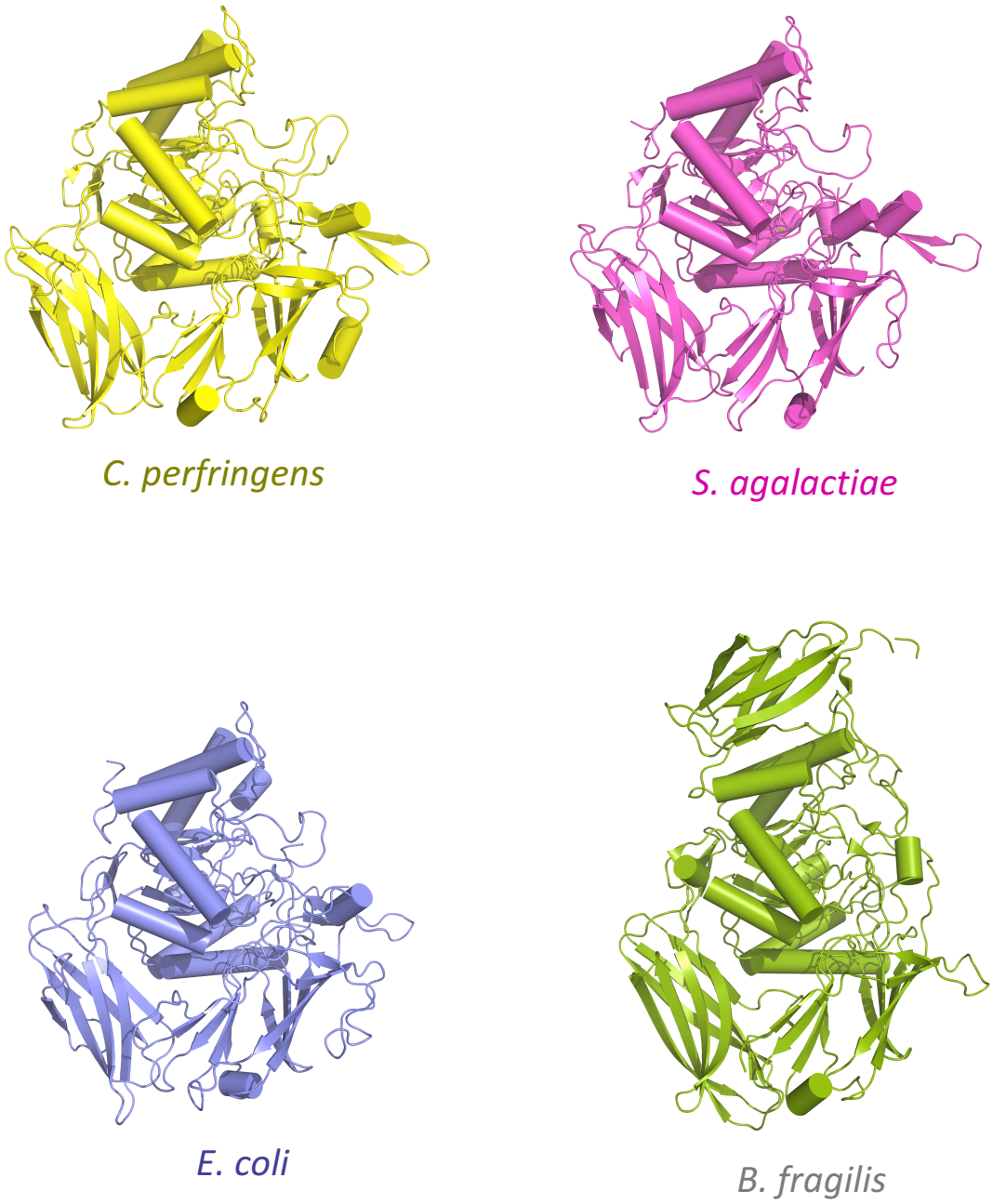
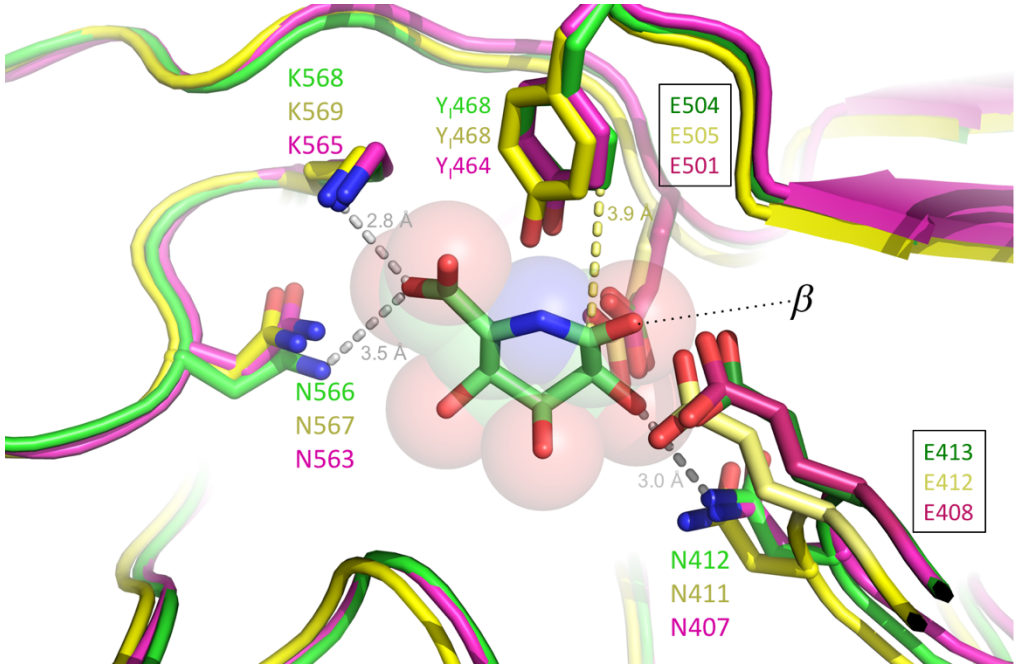


Figure 5.2. Structure of Microbial GUS Enzymes
Crystal structures of bacterial GUS enzymes from *Clostridium perfringens* (yellow, *Cp*GUS, PDB ID:4JKM), *Streptococcus agalactiae* (magenta, *Sa*GUS, PDB ID:4JKL), *Escherichia coli* (purple, *Ec*GUS, PDB ID:3K46), and *Bacteroides fragilis* (green, *Bf*GUS, RCSB:3CMG).

A



B

N-K Motif

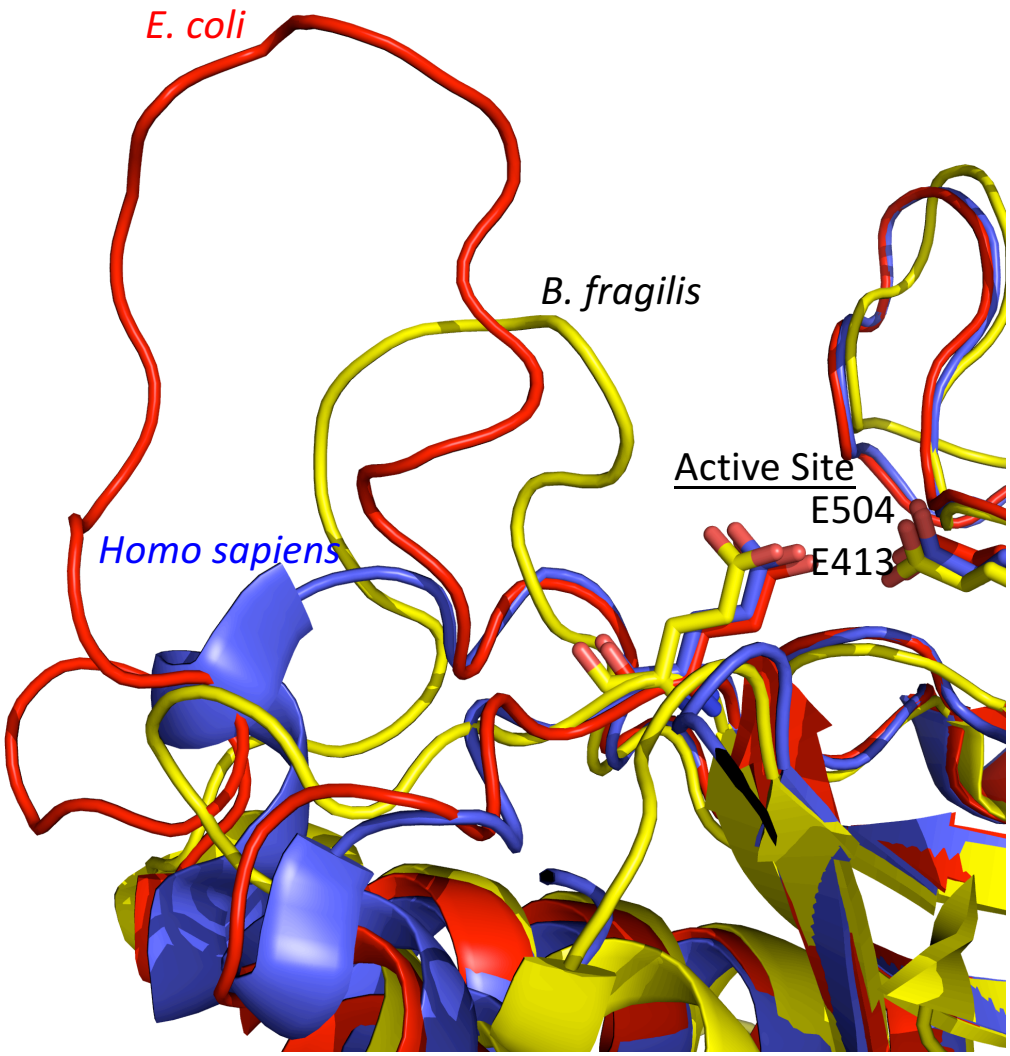
*Cp*GUS	559-QGIIRVQG	NKK-569
*Sa*GUS	555-LMILRVQG	NHK-565
***Ec*GUS**	558-QGILRVGG	**NKK**-568
*Hs*GUS	QSPTRVLG	NKK
Arch*Ia*	QNHWRTML	NRK
Ther*Tn*	QNVRRPIL	NHK
Bact*Nk*	NPVYQQGY	NRK
BfGUS	TEGDRPGI	NDK
H11G11	SEGGENGM	NHK

Figure 5.3. N-K and Y Motif

A. Superposition of the active sites of *Cp*GUS (yellow), *Sa*GUS (magenta), and *Ec*GUS green) in complex with the glucuronic acid mimic glucaro-ð-lactam, showing the binding of the ligand's carboxylate moiety (left) and indicating that the proximity of the conserved tyrosine indicated (Y_I468/464) favors the alignment of a β linkage (right) for catalysis. The labels for the two catalytic glutamic acid residues of each enzyme are boxed.

B. The Asn (N) and Lys (K) residues create the N-K loop conserved in β-glucuronidases (GUS). GUS enzymes listed those characterized by the Redinbo lab (*Cp*GUS, *Sa*GUS, *Ec*GUS, and *Bf*GUS) as well as those from mammalian (*Homo sapiens, Hs*GUS), Archaeal (Arch*Ia*), Thermatogae (Therm*Tn*), and other sources. The enzymes for which a crystal structure is available are underlined. ArchIa, Archaea Ignisphaera aggregans, WP_013302863; TherTn, Thermotogae Thermotoga naphthophila, ADA67771; BactNk, Bacteroidetes Niastella koreensis, AEV98753; H11G11, 745-residue protein from uncultured bacterium.

Originally published in Wallace *et al.* (2015) (19)


E. coli
B. fragilis
Homo sapiens
Active Site
E504
E413

B

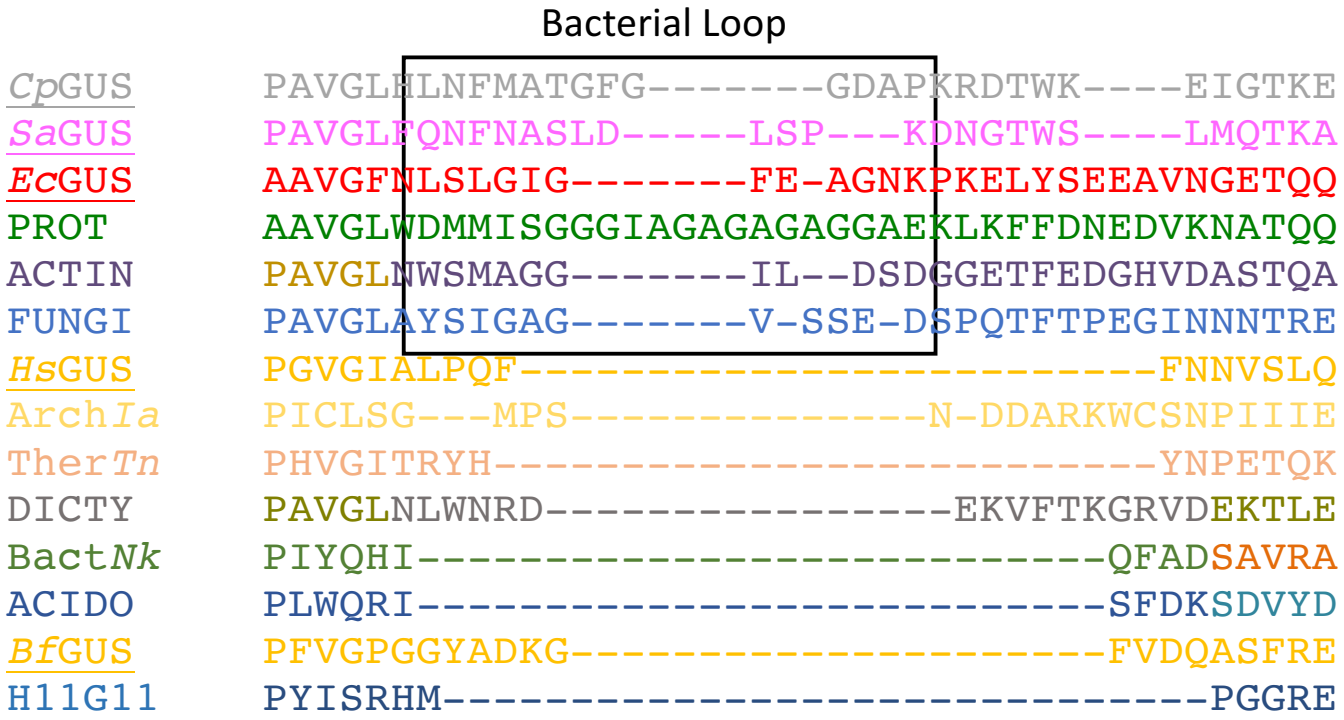

Bacterial Loop
CpGUS PAVGLHLNFMATGFG-------GDAPKRDTWK----EIGTKE
SaGUS PAVGLFQNFNASLD-----LSP---KDNGTWS----LMQTKA
EcGUS AAVGFNLSLGIG-------FE-AGNKPKELYSEEAVNGETQQ
PROT AAVGLWDMMISGGGIAGAGAGAGGAEKLKFFDNEDVKNATQQ
ACTIN PAVGLNWSMAGG-------IL--DSDGGETFEDGHVDASTQA
FUNGI PAVGLAYSIGAG-------V-SSE-DSPQTFTPEGINNNTRE
HsGUS PGVGIALPQF----------FNNVSLQ
ArchIa PICLSG---MPS----------N-DDARKWCSNPIIIE
TherTn PHVGITRYH----------YNPETQK
DICTY PAVGLNLWNRD----------EKVFTKGRVDEKTLE
BactNk PIYQHI----------QFADSAVRA
ACIDO PLWQRI----------SFDKSDVYD
BfGUS PFVGPGGYADKG----------FVDQASFRE
H11G11 PYISRHM----------PGGRE

Figure 5.4. Presence of the Bacterial Loop in Representative GUS Enzymes
A. The representative bacterial loop containing GUS enzyme *Ec*GUS (red) and smaller bacterial loop containing GUS enzymes *Bf*GUS (yellow) superimposed on the structure of human GUS (purple) reveals the lack of a bacterial loop in the mammalian enzyme and the size difference between loops of *Ec*GUS and *Bf*GUS.
B. The bacterial loop, boxed, is maintained in several GUS enzymes including those from several fungi, but is missing from other enzymes including some from bacterial sources such as H11G11. Enzymes such as *Bf*GUS and DICTY have a smaller bacterial loop than that of *Sa*GUS, *Cp*GUS, and *Ec*GUS. The enzymes for which a crystal structure is available are underlined. PROT, Proteobacteria Vibrio harveyi, WP_005434141; ACTIN, Actinobacteria Corynebacterium massiliense, WP_022863751; FUNGI, Eukaryota Aspergillus niger, CAK36851; ArchIa, Archaea Ignisphaera aggregans, WP_013302863; TherTn, Thermotogae Thermotoga naphthophila, ADA67771; DICTY, Dictyoglomi Dictyoglomus turgidum, ACK42813; BactNk, Bacteroidetes Niastella koreensis, AEV98753; ACIDO, Acidobacteriua Terriglobus roseus, WP_01478374;H11G11, 745-residue protein from uncultured
Adapted from figure originally published in Wallace *et al.* (2015) (19)

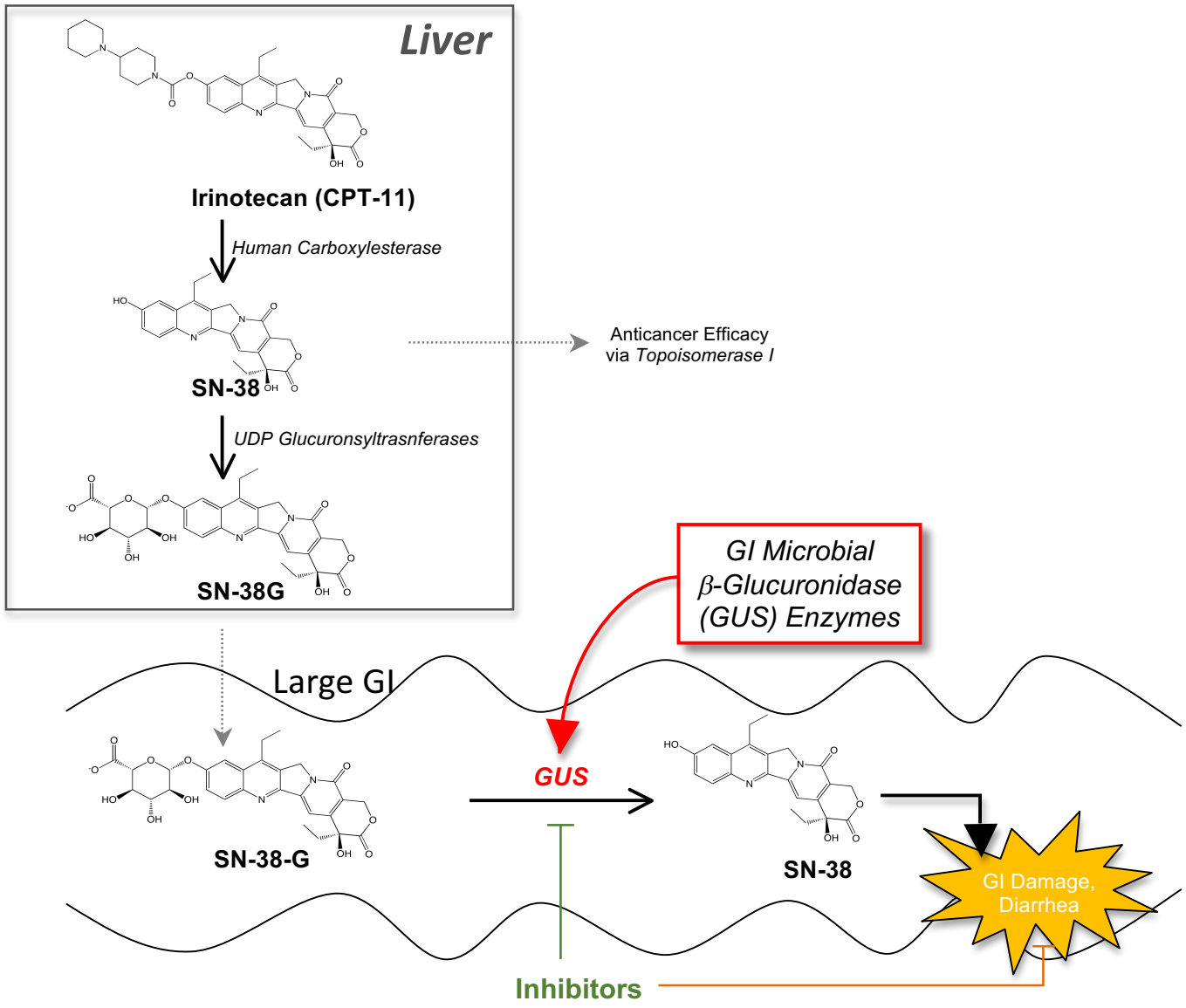

Liver
Irinotecan (CPT-11)
Human Carboxylesterase
SN-38
Anticancer Efficacy
via Topoisomerase I
UDP Glucuronsyltrasnferases
SN-38G
GI Microbial
β-Glucuronidase
(GUS) Enzymes
Large GI
SN-38-G
GUS
SN-38
GI Damage,
Diarrhea
Inhibitors

Figure 5.5. Metabolic Pathway of CPT-11
Irinotecan (CPT-11) is administered via I.V. injection to patients and then goes through the following metabolic pathways to reach the tumor and eventually be eliminated from the body. However, GI microbial GUS enzymes interfere with this elimination process causing GI damage and severe diarrhea in patients. The Redinbo lab works to discover potent inhibitors to these GUS enzymes in order to eliminate the GI damage and diarrhea caused by reactivated SN-38.

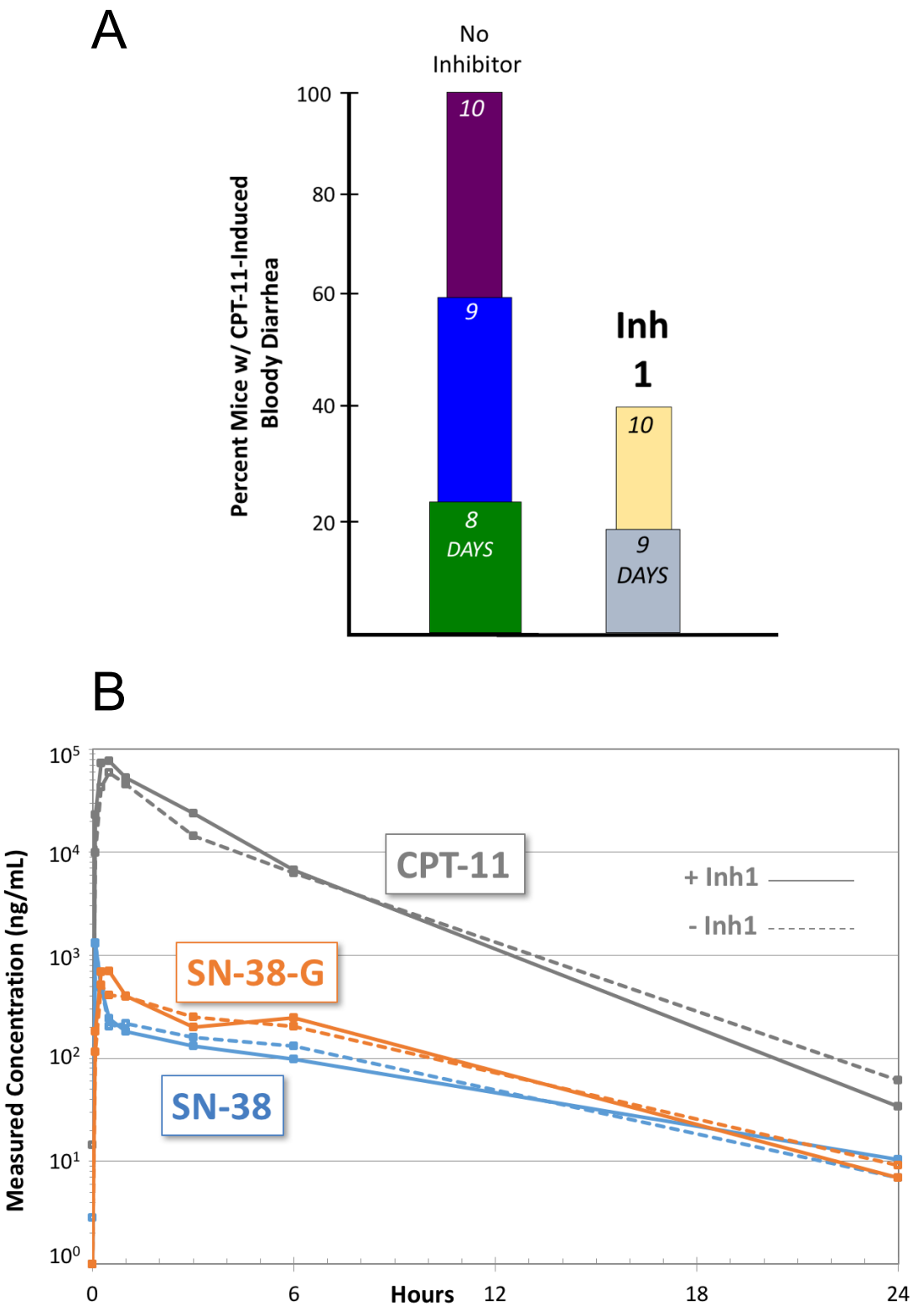

A
No Inhibitor
Percent Mice w/ CPT-11-Induced Bloody Diarrhea
100
80
60
40
20
10
9
8 DAYS
Inh 1
10
9 DAYS
B
Measured Concentration (ng/mL)
10^5
10^4
10^3
10^2
10^1
10^0
CPT-11
+ Inh1
- Inh1
SN-38-G
SN-38
0
6
Hours
12
18
24

Figure 5.6. Microbial GUS Inhibitors in Mice Treated with CPT-11
A. CPT-11 (irinotecan) produced delayed diarrhea in 25%, 60%, and 100% of mice in 8, 9, and 10 days, respectively. Oral delivery of Inhibitor 1 eliminates the appearance of diarrhea on day 8 and reduces the number of mice that experience diarrhea on days 9 and 10.
B. The circulating plasma levels of CPT-11 (irinotecan), its active metabolite SN-38, or its inactive glucuronide conjugate SN-38-G are unaffected by the oral delivery of the microbial GUS Inhibitor 1 (Inh1).
Adapted from figure originally published in Wallace *et al.* (2015) (19)

REFERENCES

1. **Sender R**, **Fuchs S**, **Milo R**. 2016. Revised estimates for the number of human and bacteria cells in the bodybioRxiv. BioRxiv.org.

2. **Human Microbiome Project Consortium**. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**:207–214.

3. **Lozupone CA**, **Stombaugh JI**, **Gordon JI**, **Jansson JK**, **Knight R**. 2012. Diversity, stability and resilience of the human gut microbiota. Nature **489**:220–230.

4. **Rosenfeld CS**. 2015. Microbiome Disturbances and Autism Spectrum Disorders. Drug Metab Dispos **43**:1557–1571.

5. **Wang Z**, **Klipfell E**, **Bennett BJ**, **Koeth R**, **Levison BS**, **Dugar B**, **Feldstein AE**, **Britt EB**, **Fu X**, **Chung Y-M**, **Wu Y**, **Schauer P**, **Smith JD**, **Allayee H**, **Tang WHW**, **DiDonato JA**, **Lusis AJ**, **Hazen SL**. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature **472**:57–63.

6. **Kinross JM**, **Darzi AW**, **Nicholson JK**. 2011. Gut microbiome-host interactions in health and disease. Genome Med **3**:14.

7. **Bäckhed F**, **Fraser CM**, **Ringel Y**, **Sanders ME**, **Sartor RB**, **Sherman PM**, **Versalovic J**, **Young V**, **Finlay BB**. 2012. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. Cell Host Microbe **12**:611–622.

8. **Schulfer A**, **Blaser MJ**. 2015. Risks of Antibiotic Exposures Early in Life on the Developing Microbiome. PLoS Pathog **11**:e1004903.

9. **Shanahan F**. 2013. The colonic microbiota in health and disease. Current Opinion in Gastroenterology **29**:49–54.

10. **Pollet A**, **Delcour JA**, **Courtin CM**. 2010. Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. Crit Rev Biotechnol **30**:176–191.

11. **Juge N**, **Tailford L**, **Owen CD**. 2016. Sialidases from gut bacteria: a mini-review. Biochem Soc Trans **44**:166–175.

12. **Juers DH**, **Heightman TD**, **Vasella A**, **McCarter JD**, **Mackenzie L**, **Withers SG**, **Matthews BW**. 2001. A structural view of the action of Escherichia coli (lacZ) beta-galactosidase. Biochemistry **40**:14781–14794.

13. **Jain S**, **Drendel WB**, **Chen ZW**, **Mathews FS**, **Sly WS**, **Grubb JH**. 1996. Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. Nat Struct Biol **3**:375–381.

14. **Hassan MI**, **Waheed A**, **Grubb JH**, **Klei HE**, **Korolev S**, **Sly WS**. 2013. High resolution crystal structure of human β-glucuronidase reveals structural basis of lysosome targeting. PLoS ONE **8**:e79687.

15. **Sly WS**, **Quinton BA**, **McAlister WH**, **Rimoin DL**. 1973. Beta glucuronidase deficiency: report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. J Pediatr **82**:249–257.

16. **Jefferson RA**, **Burgess SM**, **Hirsh D**. 1986. beta-Glucuronidase from Escherichia coli as a gene-fusion marker. Proceedings of the National Academy of Sciences of the United States of America **83**:8447–8451.

17. **Novel G**, **Novel M**. 1973. [Mutants of E. coli K 12 unable to grow on methyl-beta-D-glucuronide: map l ocation of uid A. locus of the structural gene of beta-D-glucuronidase]. Mol Gen Genet **120**:319–335.

18. **Wallace BD**, **Wang H**, **Lane KT**, **Scott JE**, **Orans J**, **Koo JS**, **Venkatesh M**, **Jobin C**, **Yeh L-A**, **Mani S**, **Redinbo MR**. 2010. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science **330**:831–835.

19. **Wallace BD**, **Roberts AB**, **Pollet RM**, **Ingle JD**, **Biernat KA**, **Pellock SJ**, **Venkatesh MK**, **Guthrie L**, **O'Neal SK**, **Robinson SJ**, **Dollinger M**, **Figueroa E**, **McShane SR**, **Cohen RD**, **Jin J**, **Frye SV**, **Zamboni WC**, **Pepe-Ranney C**, **Mani S**, **Kelly L**, **Redinbo MR**. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem Biol **22**:1238–1249.

20. **Gloux K**, **Berteau O**, **oumami El H**, **Béguet F**, **Leclerc M**, **Dore J**. 2011. A metagenomic β-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. Proceedings of the National Academy of Sciences of the United States of America **108**:4539–4546.

21. **Matsumura I**, **Ellington AD**. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. Journal of Molecular Biology **305**:331–339.

22. **Cunningham D**, **Pyrhönen S**, **James RD**, **Punt CJ**, **Hickish TF**, **Heikkila R**, **Johannesen TB**, **Starkhammar H**, **Topham CA**, **Awad L**, **Jacques C**, **Herait P**. 1998. Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. Lancet **352**:1413–1418.

23. **Douillard JY**, **Cunningham D**, **Roth AD**, **Navarro M**, **James RD**, **Karasek P**, **Jandik P**, **Iveson T**, **Carmichael J**, **Alakl M**, **Gruia G**, **Awad L**, **Rougier P**. 2000. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. Lancet **355**:1041–1047.

24. **Hurwitz H**, **Fehrenbacher L**, **Novotny W**, **Cartwright T**, **Hainsworth J**, **Heim W**, **Berlin J**, **Baron A**, **Griffing S**, **Holmgren E**, **Ferrara N**, **Fyfe G**, **Rogers B**, **Ross R**, **Kabbinavar F**. 2004. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med **350**:2335–2342.

25. **Hebbar M**, **Ychou M**, **Ducreux M**. 2009. Current place of high-dose irinotecan chemotherapy in patients with metastatic colorectal cancer. J Cancer Res Clin Oncol **135**:749–752.

26. **Kelly H**, **Goldberg RM**. 2005. Systemic therapy for metastatic colorectal cancer: current options, current evidence. J Clin Oncol **23**:4553–4560.

27. **Conroy T**, **Desseigne F**, **Ychou M**, **Bouché O**, **Guimbaud R**, **Bécouarn Y**, **Adenis A**, **Raoul J-L**, **Gourgou-Bourgade S**, **la Fouchardière de C**, **Bennouna J**, **Bachet J-B**, **Khemissa-Akouz F**, **Péré-Vergé D**, **Delbaldo C**, **Assenat E**, **Chauffert B**, **Michel P**, **Montoto-Grillot C**, **Ducreux M**, **Groupe Tumeurs Digestives of Unicancer**, **PRODIGE Intergroup**. 2011. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med **364**:1817–1825.

28. **Chabot GG**. 1996. Clinical Pharmacology and Pharmacodynamics of Irinotecan. Annals of the New York Academy of Sciences **803**:164–172.

29. **Ma MK**, **McLeod HL**. 2003. Lessons learned from the irinotecan metabolic pathway. Curr Med Chem **10**:41–49.

30. **Mathijssen RH**, **van Alphen RJ**, **Verweij J**, **Loos WJ**, **Nooter K**, **Stoter G**, **Sparreboom A**. 2001. Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). Clin Cancer Res **7**:2182–2194.

31. **Alimonti A**, **Gelibter A**, **Pavese I**, **Satta F**, **Cognetti F**, **Ferretti G**, **Rasio D**, **Vecchione A**, **Di Palma M**. 2004. New approaches to prevent intestinal toxicity of irinotecan-based regimens. Cancer Treat Rev **30**:555–562.

32. **Pommier Y**. 2006. Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer **6**:789–802.

33. **Gupta E**, **Lestingi TM**, **Mick R**, **Ramirez J**, **Vokes EE**, **Ratain MJ**. 1994. Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. Cancer Res **54**:3723–3725.

34. **Pérez-Roth E**, **Kwong SM**, **Alcoba-Florez J**, **Firth N**, **Méndez-Alvarez S**. 2010. Complete nucleotide sequence and comparative analysis of pPR9, a 41.7-kilobase conjugative staphylococcal multiresistance plasmid conferring high-level mupirocin resistance. Antimicrobial Agents and Chemotherapy **54**:2252–2257.

35. **Dranitsaris G**, **Maroun J**, **Shah A**. 2005. Estimating the cost of illness in

colorectal cancer patients who were hospitalized for severe chemotherapy-induced diarrhea. Can J Gastroenterol **19**:83–87.

36. **Stein A**, **Voigt W**, **Jordan K**. 2010. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol **2**:51–63.

37. **Abigerges D**, **Armand J-P**, **Chabot GG**, **Da Costa L**, **Fadel E**, **Cote C**, **Hérait P**, **Gandia D**. 1994. Irinotecan (CPT-11) High-Dose Escalation Using Intensive High-Dose Loperamide to Control Diarrhea. Journal of the National Cancer Institute **86**:446–449.

38. **Wallace BD**, **Wang H**, **Lane KT**, **Scott JE**, **Orans J**, **Koo JS**, **Venkatesh M**, **Jobin C**, **Yeh L-A**, **Mani S**, **Redinbo MR**. 2010. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science **330**:831–835.

39. **Ahmad S**, **Hughes MA**, **Lane KT**, **Redinbo MR**, **Yeh L-A**, **Scott JE**. 2011. A High Throughput Assay for Discovery of Bacterial β-Glucuronidase Inhibitors. Curr Chem Genomics **5**:13–20.

40. **Roberts AB**, **Wallace BD**, **Venkatesh MK**, **Mani S**, **Redinbo MR**. 2013. Molecular Insights into Microbial β-Glucuronidase Inhibition to Abrogate CPT-11 Toxicity. Molecular Pharmacology **84**:208–217.

CHAPTER 6: DEFINING THE β-GLUCURONIDASE ENZYME FAMILY IN THE HUMAN GASTROINTESTINAL TRACT

Introduction

The human gastrointestinal (GI) microbiota is a vast ecosystem of bacteria that are constantly interacting with each other, the human body, and many of the substances that enter the human body. The GI microbiota has been shown to interact with a wide range of small molecules, both produced by the human body and from outside the body. For instance, conjugated bile acids that are produced in the human liver can be de-conjugated by the GI microbiota which can lead to bacterial overgrowth and inflammation (1). In addition, a large part of food digestion and absorption takes place in the GI and the microbiota play a key role in breaking down the polysaccharides that humans lack the enzymes to process (2). The microbiota can also activate pro-drugs as seen with sulfasalazine, inactivate drug molecules such as in the processing of digoxin by *E. lenta,* or bind drugs, preventing the drug from reaching its intended target as seen in L-DOPA treatment (3). We are interested in such microbiota-drug interactions, especially microbiota processing of drugs that have been inactivated in the liver and are then excreted via the intestine.

The chemotherapy drug irinotecan (CPT-11) is a prodrug that is activated through phase I drug metabolism to the compound SN-38 (4, 5). SN-38 inhibits

topoisomerase I and is used as a treatment of solid tumors including colorectal and pancreatic cancer (6). SN-38 then returns to the liver where a glucuronide sugar is attached by the UGT enzymes of phase II drug metabolism (4, 7, 8). The SN-38-glucuronide (SN-38G) compound is more soluble and enters the intestines in order to be excreted from the body. β-glucuronidase (GUS) enzymes expressed by the GI microbiota can cleave the glucuronide sugar off SN-38G resulting in active SN-38. SN-38 causes cell death of the intestinal epithelium leading to delayed-onset diarrhea that often leads to patients receiving lower doses of irinotecan or being taken off this effective chemotherapy completely (9-12). In 2010 the Redinbo lab showed that inhibitors designed against the purified *E. coli* GUS reduce the number of bloody diarrheal events in mice co-treated with irinotecan and inhibitor (13).

Since 2010 the Redinbo lab has characterized the GUS enzyme from *E. coli* (*Ec*GUS), *Streptococcus agalactiae* (*Sa*GUS), *Clostridium perfringens* (*Cp*GUS), and *Bacteroides fragilis* (*Bf*GUS) and eight small molecule inhibitors (14, 15). This initial characterization included solving the structures of *Ec*GUS, *Sa*GUS, and *Cp*GUS and characterization of a previously solved structure (PDB ID: 3CMG) as a β-glucuronidase (*Bf*GUS) (14). This allowed us to identify key conserved residues in the active site including two catalytic glutamic acids, an NxKG motif that interacts with the carboxylic acid group of glucuronic acid, an asparagine residue next to the first catalytic glutamic acid, and a tyrosine defined as part of the Y loop in Wallace *et al.* (14). Despite previous characterization of these enzymes, little evidence exists about the abundance of these enzymes or even the associated bacteria in the human GI. In order to understand which bacteria in the human GI express GUS and

the relative abundances of GUS genes, we set out to analyze the Human Microbiome Project (HMP) sequencing repository for GUS genes.

The HMP sequencing repository is part of the NIH-funded Human Microbiome Project Consortium that set out to develop resources and methods to manage high-throughput metagenomics data in order to characterize the microbial communities obtained from sampling a large number of healthy individuals across a wide range of body sites (16). There are over 11,000 samples from 300 healthy adults across 18 specific body sites (16). We chose to analyze sequences obtained from stool samples for which 139 individuals were sampled. To facilitate analysis, the metagenomics sequencing data from these samples has undergone gene prediction and annotation. Therefore, we were able to use the HMGI- Gene Indices and HMGC- Clustered Gene Indices catalogs to search for GUS genes. As shown in Figure 6.1, we screened genes from stool samples in these catalogs for the key active site residues we had previously characterized in order to identify genes encoding for GUS enzymes. We then classified those identified GUS sequences according to a highly variable loop region just outside of the active site. The Redinbo lab previously identified the bacterial loop found in *Ec*GUS, *Sa*GUS, and *Cp*GUS as a novel feature that allowed for the discovery of small molecule inhibitors that are selective for these bacterial GUS enzymes over the human ortholog (13, 14). We have re-named the bacterial loop Loop 1 as we identified a second loop, Loop 2, that is predicted to occupy a similar position at the entrance of the active site of these enzymes. We classified each of the identified GUS sequences based on the size and type of active site loop they contain. There are six loop classes: no loop, Loop 1,

Mini Loop 1, Loop 2, Mini Loop 2, and Mini Loop 1 Mini Loop 2. Previously characterized enzymes *Ec*GUS, *Sa*GUS, and *Cp*GUS fall into the Loop 1 class while *Bf*GUS is a Mini Loop 1 enzyme. This analysis showed that the human GI microbiota expresses many GUS enzymes with a much wider sequence diversity that has previously been explored.

Materials and Methods

Human Microbiome Project (HMP) data

Protein sequences for stool samples from the HMP Clustered genes- HMGC (http://hmpdacc.org/HMGC) and Gene Indices – HMGI (http://hmpdacc.org/HMGI/) were downloaded as FASTA files and used as a database for subsequent NCBI BLASTp searches. Microbial assemblies, annotations and reads for stool samples were obtained from http://hmpdacc.org/HMASM/ and used for abundance calculations.

HMP β-glucuronidase Identification

In order to identify β-glucuronidases (GUS) in the HMP stool proteins, we used *Escherichia coli* β-glucuronidase (NCBI Accession: NP_416134.1), *Clostridium perfringens* β-glucuronidase (NCBI Accession: WP_003467686.1), *Streptococcus agalactiae* β-glucuronidase (NCBI Accession: WP_000966715.1) and *Bacteroides fragilis* (NCBI Accession: 3CMG) in a pair-wise alignment approach. Briefly, each HMP stool protein was aligned to the above proteins using NCBI BLASTp. We considered alignments with an E-value < 0.05 and percent identity ≥25 to be valid hits. Only those valid hits were then checked for the presence of NxKG motif,

catalytic E residues and N and Y motifs corresponding to GUS positions shown in Table 6.1. We classified HMP proteins that satisfied all the above conditions as β-glucuronidase.

HMP β-glucuronidase Loop Classification

HMP GUS proteins identified above were subjected to a multiple sequence alignment (MSA) using Clustal Omega along with selected model GUS sequences (*E. coli* (NCBI Accession: NP_416134.1), *Streptococcus agalactiae* (NCBI Accession: WP_000966715.1), *Clostridium perfringens* (NCBI Accession: WP_003467686.1), *Homo sapiens* (NCBI Accession: NP_000172.2), H11G11-BG from uncultured bacterium (NCBI Accession: CBJ55484.1) and *Bacteroides fragilis* (NCBI Accession: 3CMG)).

The MSA alignment was examined for the presence of two loop regions. The first (loop 1) corresponds to *E. coli* GUS positions 356 to 380 and the second (loop 2) corresponds to *E. coli* GUS positions 416 to 419. HMP GUS proteins were then categorized into six categories based on the number of residues present in those two regions after excluding gap characters as shown in Table 6.2.

Gene Abundance Calculations

The trimmed and duplicate-marked reads for each stool sample were aligned to its own assembly using bowtie2 (v.2.2.5) (17). Alignments were stored in bam files and gene specific counts were obtained by running the program featurecount from the subread package (v.1.4.6-p2-Linux-x86_64) on each bam file (18). To account for differences in read numbers between samples, we normalized and log transformed the counts as shown below:

$$log_{10}\left(\frac{RC}{n} \times \frac{\sum x}{N} + 1\right)$$

where RC is the read count for a gene in a particular sample, *n* is the total number of reads in that sample, the sum of *x* is the total number of reads in all samples and N is the total number of samples.

Results

HMGC- Clustered Gene Indices GUS Enzymes

The HMGC- Clustered Gene Indices is a non-redundant catalog of bacterial genes found during the Human Microbiome Project sequencing initiative grouped by body site. Genes from the HMP gene indices (HMGI discussed below) were compared at a 95% identity cut-off and redundant sequences were removed. The translated protein sequences from this database were compared to previously characterized GUS proteins to identify sequences possessing the conserved NxKG motif, catalytic E residues and N and Y motifs discussed in Chapter 5 and shown in Table 6.1. This analysis identified 293 unique GUS sequences, most of which have never been characterized as GUS enzymes (Figure 6.2A).

This non-redundant set of sequences gave us a manageable dataset in order to characterize important features of this enzyme class. One feature this dataset uncovered is a wide range in protein sequence length. As shown in Figure 6.2B, about one third of these GUS sequences are longer than the four previously characterized GUS enzymes. This is interesting as we previously established that both *Ec*GUS and *Bf*GUS form tetramers in solution but models of the structures of

these longer sequences suggest that this extra length may interfere with that tetramer formation (14). We are currently working to determine which bacterial species each of these GUS sequences was sequenced from in order to establish a database that will allow for easy prediction of the GUS composition of a bacterial community characterized only by 16S rRNA sequencing.

HMGI- Gene Indices GUS Enzymes

Stool samples were taken from 139 individuals and the genes found via whole genome sequencing of bacteria in those samples are found in the HMGI- Gene Indices database of the Human Microbiome Project (HMP). Each protein in these sequences was aligned to the previously characterized GUS proteins (*Ec*GUS, *Sa*GUS, *Cp*GUS, and *Bf*GUS) and analyzed for the presence of NxKG motif, catalytic E residues and N and Y motifs as shown in Table 6.1. This resulted in the identification of 3,013 GUS proteins as shown in Figure 6.3A. This dataset represents the first compilation of the GUS proteins found in the human gastrointestinal microbiome from sequencing data.

This dataset allowed us to analyze details about the number and type of GUS proteins each individual possessed. As shown in Figure 6.3B, there is a wide range of the number of GUS enzymes found in the stool samples of each individual with the minimum number being 4 and the maximum number being 40. This mimics the high diversity seen in other characterizations of the human GI microbiome.

Establishment of GUS Loop Characterizations

As described in Chapter 5, we have previously established that a flexible loop outside of the GUS active site is important for selective inhibition of these enzymes;

this loop was previously termed the bacterial loop but will now be referred to as Loop 1. Using alignments of the 293 GUS sequences identified from the HMGC database we were able to identify 3 distinct categories of enzymes with respect to Loop 1. The first category is made up of GUS enzymes that contain no loop in this region and resemble the tight turn seen in the *Homo sapiens* GUS enzyme (H_sapiens, P_merdae, H11G11-BG, B_uniformis, and B_dorei sequences in Figure 6.4A). The second category is made up of enzymes that contain a full loop of 15 or more residues in this region that resembles that previously characterized in *Ec*GUS, *Cp*GUS, and *Sa*GUS (Loop 1 or L1 category, red sequences in Figure 6.4A). The third category is made up of enzymes that fall in between the first two categories and have a smaller loop between 10 and 15 residues (B_fragilis, green in Figure 6.4A). We have termed this Mini Loop 1 (mL1) due to its smaller size. This characterization left 75.8% of the sequences in the no loop category, prompting us to look more closely at the active site of this category. Approximately 35 residues downstream from Loop 1, we identified an additional loop region that is predicted to also extend over the active site in a similar manner to Loop 1. We termed this additional loop region Loop 2. Just as in the Loop 1 region, we found enzymes with a full loop of 12 residues or more (Loop 2 or L2 category, B_uniformis in Figure 6.4A) as well as enzymes with a shorter loop of 9 to 11 residues (mini Loop 2 or mL2 category, P_merdae in Figure 6.4A). Interestingly, enzymes that contain a mini Loop 1 may also contain a mini Loop 2 (mini Loop 1 mini Loop 2 or mL1mL2 category, B_ovatus in Figure 6.4A) but the full Loop 1 or Loop 2 were not seen in combination together or with a mini loop in the opposing position. We hypothesize that a large flexible

region in both the Loop 1 and Loop 2 positions is not favorable due to the many clashes that would occur as the two loops moved. The two loops in combination might also hinder function of the enzyme by blocking access of a substrate trying to enter the active site or preventing the products from diffusing out of the active site. There were also enzymes that did not contain a loop in either of these positions (no Loop or nL category, B_dorei in Figure 6.4A) which results in bacterial enzymes that are similar to the *Homo sapiens* and other mammalian GUS enzymes that have been characterized. The full alignment of these enzymes is shown in Appendix 3.

Using the criteria set out above and in Table 6.2, the HMGC (293 sequences) and HMGI (3013 sequences) datasets were classified into these six loop classifications (Figure 6.4 B and C). The distribution across loop classes was similar for each dataset. Because the HMGI dataset includes information on GUS content in each individual, we moved forward with analysis of the HMGI GUS dataset. Of the 3013 sequences identified as GUS sequences, 144 did not have sufficient coverage in the area of the loops in order to determine their loop classification. Forty-eight percent of individuals in the database possessed at least one of these sequences (Figure 6.5A). The 144 unclassified sequences represent 3.8% of all identified GUS sequences. Previously characterized *Ec*GUS, *Cp*GUS, and *Sa*GUS fall into the L1 class. 67% of individuals sampled possessed at least one L1 GUS sequence with an average 1.8 ± 1 L1 GUS sequences per individual (Figure 6.5B). The L1 class contains 165 sequences which is 5.5% of all identified GUS sequences. The previously characterized *Bf*GUS is a mL1 enzyme. 98% of individuals possessed at least one mL1 GUS sequence with an average 4.5 ± 2.2 GUS sequences per

individual (Figure 6.5C). The 615 mL1 GUS sequences represent 20.4% of all identified GUS sequences. Three hundred and fifty-one identified sequences were classified as L2 GUS sequences which is 11.6% of all identified sequences. 91% of individuals had at least one L2 GUS sequence with an average 2.7 ± 1.5 GUS sequences per individual (Figure 6.5D). The mL2 class represents 6.9% of all identified GUS sequences with 208 sequences identified from the HMGI database. 80% of all individuals possessed at least one mL2 GUS sequences with an average 1.9 ± 0.8 mL2 GUS sequences per individual (Figure 6.5E). The mL1mL2 class is the least well represented in this sample with just 28% of individuals having at least one mL1mL2 GUS sequence and an average of 1.1 ± 0.2 mL1mL2 GUS sequences per individual (Figure 6.5F). From the HMGI database, we identified only 36 mL1mL2 GUS sequences which represents just 1.2% of all identified GUS sequences. Finally, the no loop class represents the largest percentage (49.6%) of the identified GUS sequences with 1494 GUS sequences. All individuals possessed a nL GUS sequence. There was an average of 11 ± 5 nL GUS sequences per individual (Figure 6.5G).

GUS-Containing Bacteria

As mentioned above, we are interested in establishing which bacteria are likely to possess each identified GUS sequence. For each of the 139 individuals sampled, we identified the most abundant sequence (via read counts) in each loop class and matched the sequence to a deposited reference genome in NCBI. This allowed us to identify the organism from which the sequence likely originated and therefore the phyla distribution of GUS sequence containing bacteria for each loop

class. Because the 3013 GUS sequences is a redundant set of sequences found across the 139 individuals, several of the most abundant sequences in each loop class were the same across individuals. For each loop class we present the top 5 most common abundant sequences with the number of individuals for which each was the most abundant sequence of that loop class in the count column. The majority of L1 GUS sequences could not be matched to a deposited reference genome. Of the sequences that were identified, 42 matched to bacteria from the Firmicutes phylum and one was from the Proteobacteria phylum. The most common sequences in the L1 class were from *Faecalibacterium prausnitzii* and *Eubacterium eligens*. These GUS sequences were chosen to be expressed, purified, and characterized in the lab as described in Chapter 7. The mL1 class of sequences was found predominately in bacteria from the Bacterodetes phylum with a small number from the Firmicutes phyla. Although *Bacteroides fragilis* was not an abundant sequence seen in this analysis, we chose to move forward with characterization with *Bf*GUS that has already expressed in the Redinbo lab. All identified L2 GUS sequences are from the Bacteroidetes phyla. The most common sequences were associated with multiple species (MSP) of *Bacteroidetes* and *Bacteroides uniformis*. The sequence associated with *Bacteroides uniformis* was chosen for characterization in the lab. The mL2 GUS sequences were also predominately found in bacteria from the Bacteroidetes phylum. Uniquely, there were also sequences found in bacteria of the Verrucomicrobia phylum. The most common sequence was from *Parabacteroides merdae* and was chosen for further characterization. The sequences in the mL1mL2 class that were able to be matched to reference genomes

were found exclusively in bacteria from the Bacteroidetes phyla. Two different sequences, each seen in 12 individuals, were associated with *Bacteroides ovatus* and one of these sequences was chosen for further characterization. Finally, 65.7% of matched sequences from the nL class were found in bacteria from the Bacteroidetes phylum. The remaining 34.3% of matched sequences were from the Firmicutes phylum. The most common sequences were found in *Bacteroides uniformis* and *Bacteroides dorei*. The sequence from *Bacteroides dorei* was chosen for further characterization in the Redinbo lab. This analysis of the Human Microbiome Project stool sample sequencing database provides the first in-depth look at the GUS sequences contained in the human microbiome. This dataset gives us a better appreciation of the variation in GUS sequence and has allowed us to choose representative proteins to further characterize the function of this enzyme class as discussed in Chapter 7.

Discussion

Bacterial GUS enzymes have been established as the cause of drug toxicity of glucuronidated compounds such as irinotecan and NSAIDs. However, previous efforts to compile a catalog of GUS sequences found in the human GI microbiota have depended on the automated annotation of genes which we have previously shown is imprecise and often confuses β-glucuronidase and β-galactosidase enzymes (14). Instead we chose to classify genes as a GUS based on the presence of key, conserved active site residues in a one-to-one sequence alignment with the previously discussed, well-characterized bacterial GUS enzymes. This resulted in a

much more complete catalog of 293 non-redundant GUS sequences identified from the HMP Clustered Gene Indices Database. Identifying GUS sequences from the HMP Gene Indices Database also allowed us to examine the variability in GUS sequences across 139 individuals. Although the vast variability of the microbiota from person to person has been shown in analyses of other microbiome characteristics, this is the first evidence to suggest that the number and type of GUS enzymes expressed by each individual is highly variable. This may contribute to the varying severity of drug toxicity seen with drugs such as irinotecan.

This new dataset of GUS sequences allowed us to identify features of this enzyme class that have not previously been explored. For instance, almost one third of these GUS sequences are longer than previously characterized bacterial enzymes. The extra length forms an uncharacterized C-terminal domain that likely disrupts tetramer formation and may alter the function of these enzymes. In addition, we were able to expand understanding of the previously identified bacterial loop at the entrance of the active site. In addition to this bacterial loop, now termed Loop 1, we identified a second loop region (Loop 2) that is predicted to extend over the entrance of the active site similar to Loop 1. By analyzing the absence or length of the loop in both the Loop 1 and Loop 2 regions, we were able to break the larger GUS sequence datasets into six categories for further characterization. The loop classes are Loop 1 (L1), Mini Loop 1 (mL1), Loop 2 (L2), Mini Loop 2 (mL2), Mini Loop 1 Mini Loop 2 (mL1mL2), and no Loop (nL) with loop lengths as defined in Table 6.2. Because these loops sit at the entrance of the active site, we hypothesize that the presence and size of the loop may impact the size of substrate the enzyme

is able to process. We have also previously shown that Loop 1 is important for the selectivity of GUS inhibitors over *H. sapiens* GUS while Mini Loop 1 is not sufficient to facilitate efficacy of the inhibitors (13, 14). Therefore, we are interested in the type of GUS sequences found in each individual. Indeed, we found variability in the presence of some loop classes as well as the number of sequences in each class that individuals have. Most individuals possessed a few different sequences in the L1, mL1 L2, and mL2 classes suggesting that these enzymes are a regular member of the GI community; however, the activity of various sequences within each class may be redundant such that each individual needs only one sequence of each type. All individuals possessed at least one nL GUS and the average nL GUS sequence content was 11 ± 5 nL GUS sequences. This suggests that each nL GUS enzyme may be selective for a specific substrate so that each individual maintains several nL GUS enzymes in order to process varying but related substrates. Alternatively, only 28% of individuals contained a mL1mL2 GUS sequence and most of those individuals contained only one mL1mL2 sequence. While all individuals sampled were considered healthy, we do not have details about significant variations in their diet or genetics that might cause significant changes in the glucuronidated compounds present in their GI that are degraded by mL1mL2 GUS enzymes. Despite this interesting distribution of total number of GUS sequences and number of sequences in each loop class, this data does not take into account the abundance or expression of these sequences. It is well known that some bacteria are much more abundant in the microbiota and it is also possible that some bacteria contain more than one GUS sequence, either of different sequence, different class, or

duplication of a gene. Future analysis will account for abundance of each gene by using normalized read counts.

We are also interested in which bacteria express each of the identified GUS sequences. We were able to match the most abundant sequence of each loop class from each of the 139 individuals to deposited reference genomes in order to assign the taxonomy of the expressing bacteria. Looking at the phyla level yielded interesting trends in distribution of GUS sequences of each loop class. All loop classes except for L1 were dominated by bacteria of the Bacteroidetes phylum. It seems likely that the L2 and mL1mL2 GUS sequences are found exclusively in the Bacteroidetes phylum. The mL2 class contained the only identified GUS sequences from the Verrucomicrobia phylum. The mL1 and nL classes include sequences from the Firmicutes phylum in addition to Bacteroidetes but sequences from Firmicutes seem to be less common. The L1 class presented a unique taxonomic distribution; although over half of the sequences could not be matched to a deposited reference genome, the remaining sequences were exclusively from the Firmicutes and Proteobacteria phyla. This is the only class that included a sequence from the Proteobacteria phylum. While previous characterization has focused on L1 class enzymes from the Firmicutes (*Sa*GUS, *Cp*GUS) and Proteobacteria (*Ec*GUS) phyla, this new dataset shows that there is a wide range of GUS sequence as well as bacterial diversity that remains to be characterized in order to better understand the GUS family of enzymes. Indeed, we have chosen sequences from each of the loop classes in order to characterize the structure and function of these novel GUS

enzymes. An exploration of the function of these enzymes is presented in Chapter 7 of this thesis.

This exploration of the HMP sequence repository establishes the first catalog of GUS sequences based on homology to characterized bacterial enzymes. This new information about the diversity of sequences included in this enzyme class will lead to better understanding of the function of these enzymes in the human GI microbiota and their impact on human health.

Bacterium	NxKG motif positions	Catalytic E residues positions	N and Y motifs positions
Escherichia coli	566, 568, 569	413, 504	412, 468
Clostridium perfringens	567, 569, 570	412, 505	411, 468
Streptococcus agalactiae	563, 565, 566	408, 501	407, 464
Bacteroides fragilis	547, 549, 550	395, 476	394, 445

Table 6.1. Key active site residues used for β-glucuronidase Identification.

Category	Residues in Loop 1 region- *Ec*GUS 356-380	Residues in Loop 2 region- *Ec*GUS 416-419
No Loop (nL)	< 10	< 9
Loop 1 (L1)	> 15	< 9
Mini Loop 1 (mL1)	≤ 15 and ≥ 10	< 9
Loop 2 (L2)	< 10	≥ 12
Mini Loop 2 (mL2)	< 10	≥ 9 and < 12
MiniLoop 1 + MiniLoop 2 (mL1mL2)	≤ 15 and ≥ 10	≥ 9 and < 12

Table 6.2. Criteria used for β-glucuronidase Loop Classification

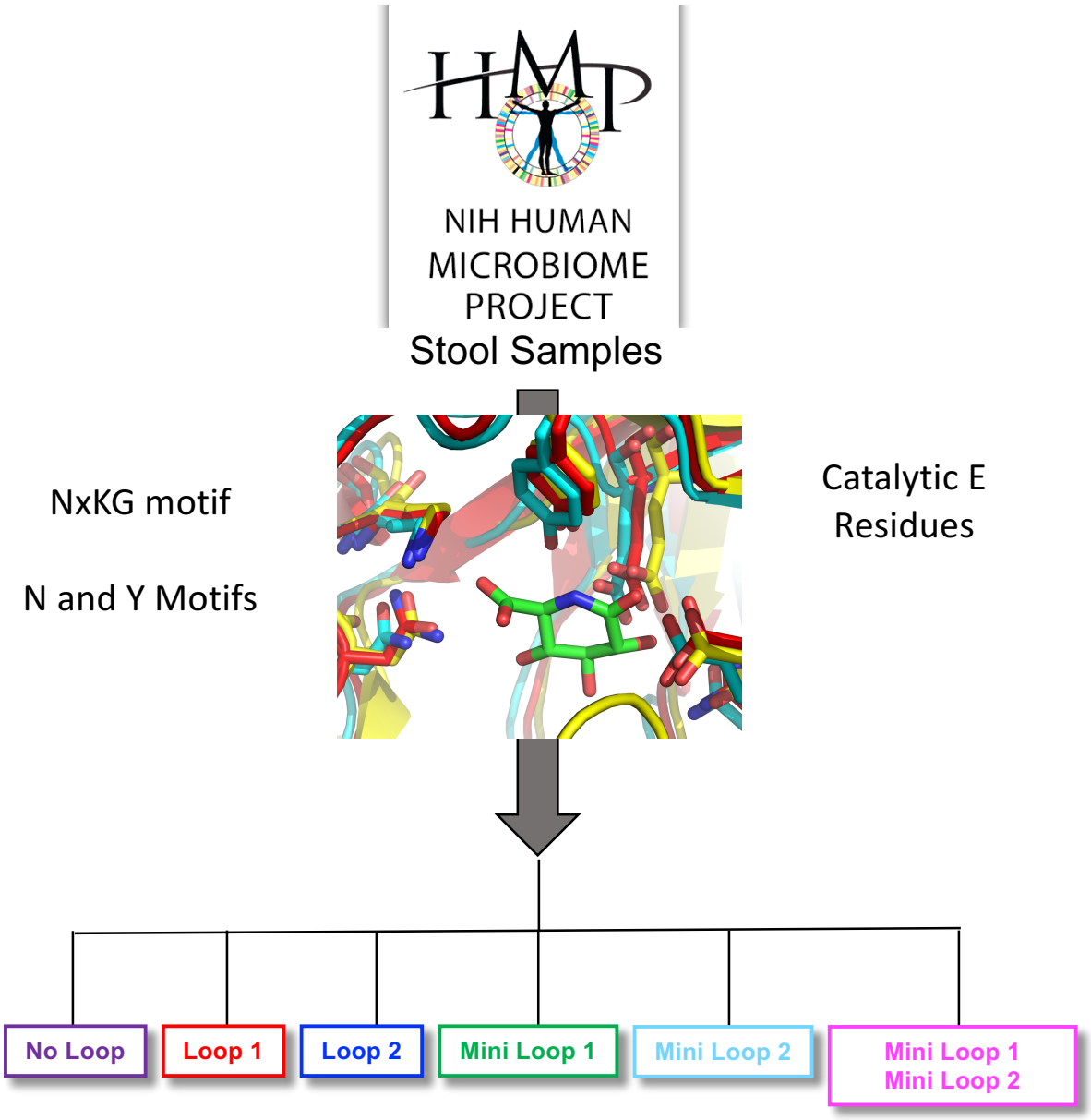


Figure 6.1. Workflow for Identification and Classification of Novel GUS Sequences
Protein sequences for stool samples from the HMP Clustered Genes (HMGC) or the HMP Gene Indices (HMGI) were aligned to previously characterized GUS sequences in order to identify sequences containing the key active site residues. These novel GUS sequences were then classified according to the loop features they contain.

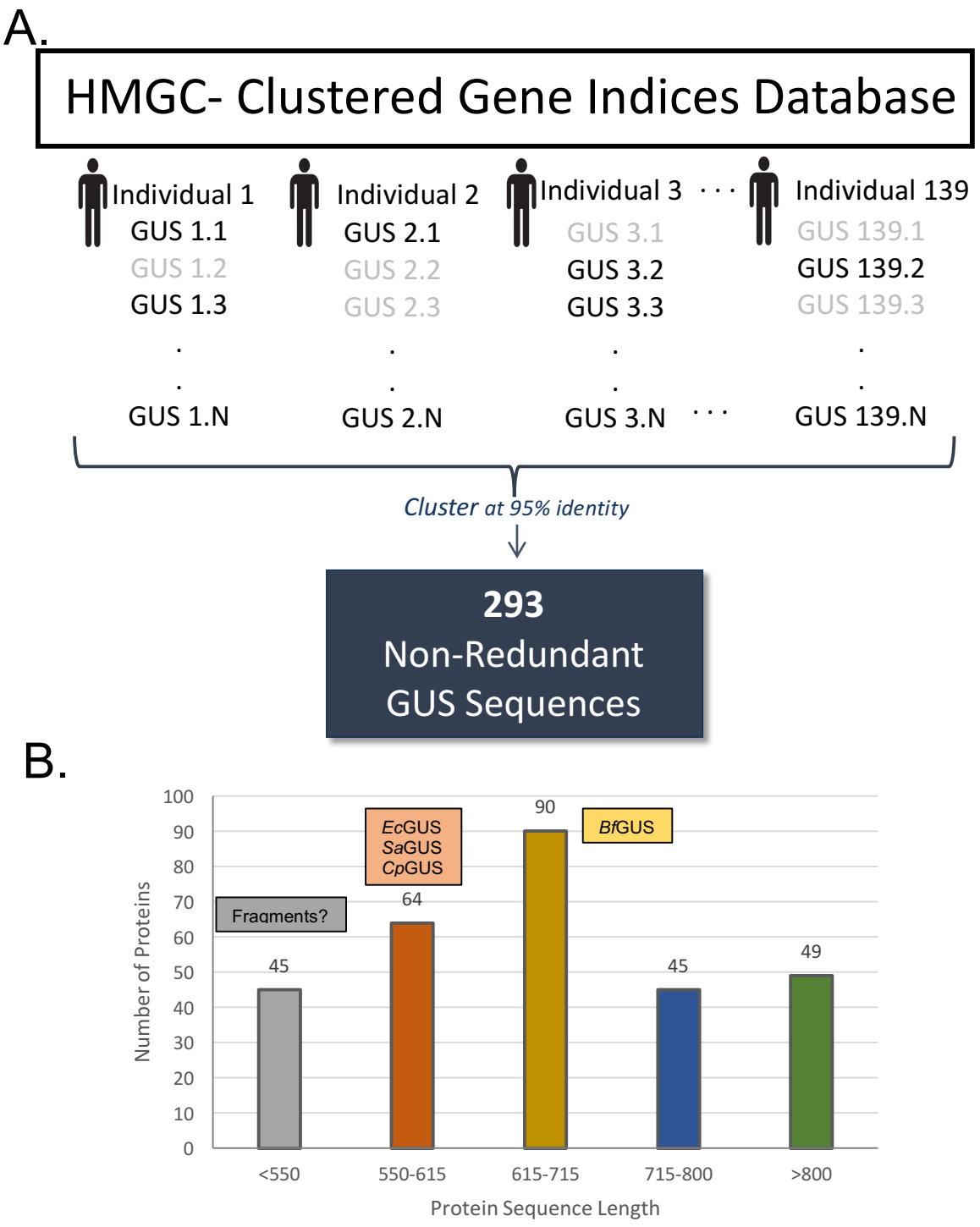
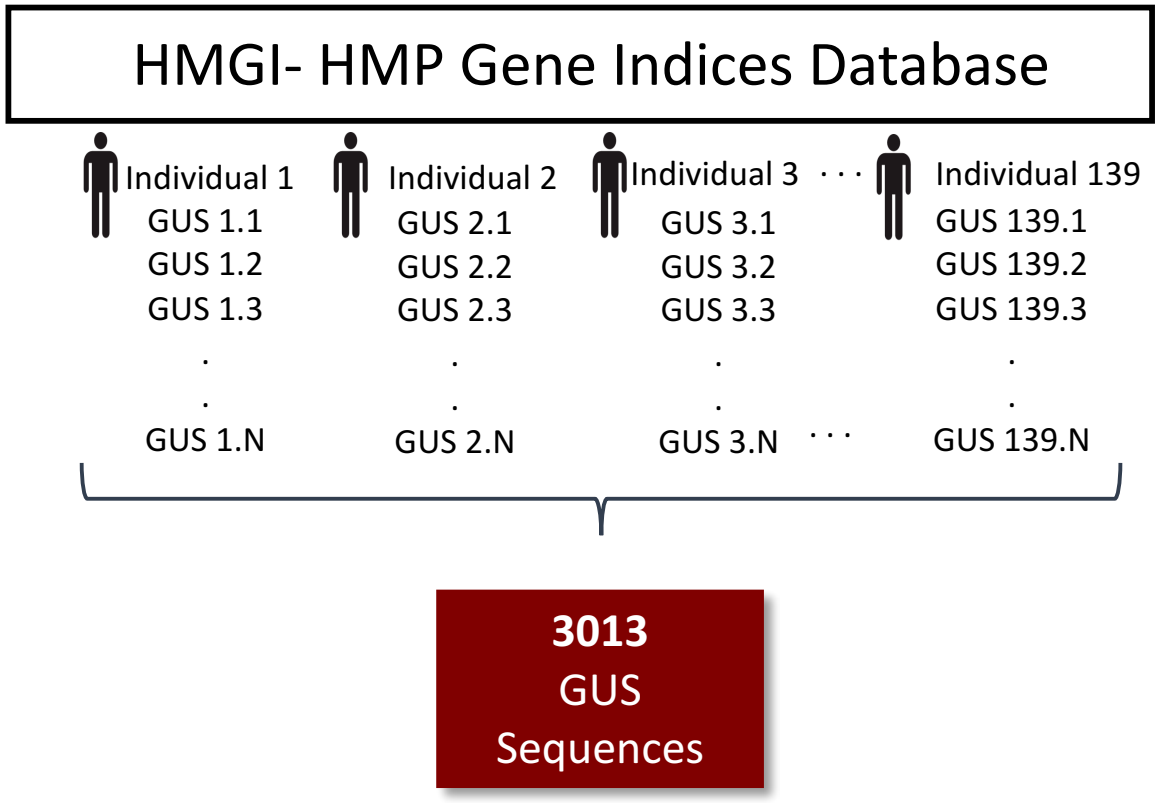


Figure 6.2. Identification of Novel GUS Sequences from HMGC- HMP Clustered Gene Indices Database

A. 293 Non-Redundant Novel GUS Sequences were identified from the HMGC database.

B. Distribution of protein sequence length of the 293 Novel GUS Sequences.

A.


HMGI- HMP Gene Indices Database
Individual 1
GUS 1.1
GUS 1.2
GUS 1.3
GUS 1.N
Individual 2
GUS 2.1
GUS 2.2
GUS 2.3
GUS 2.N
Individual 3
GUS 3.1
GUS 3.2
GUS 3.3
GUS 3.N
Individual 139
GUS 139.1
GUS 139.2
GUS 139.3
GUS 139.N
3013
GUS
Sequences

B.

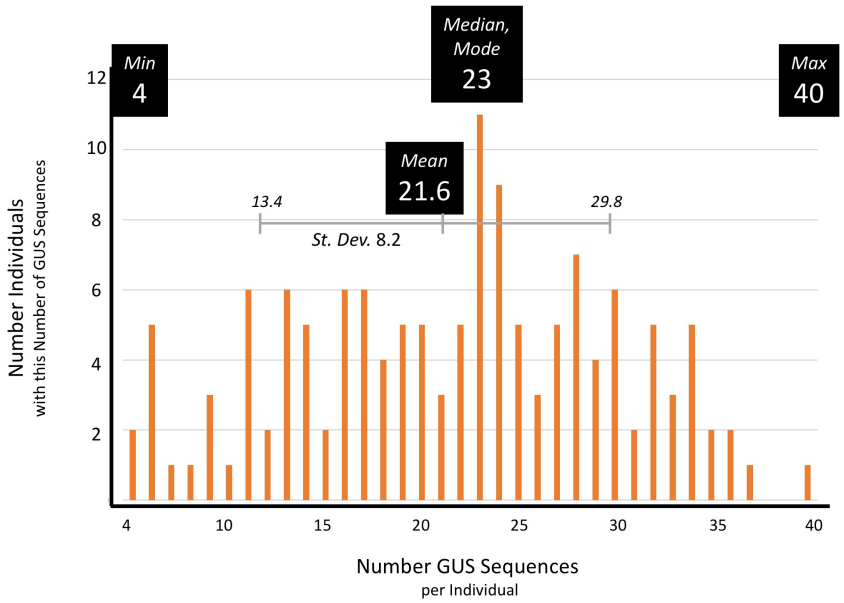
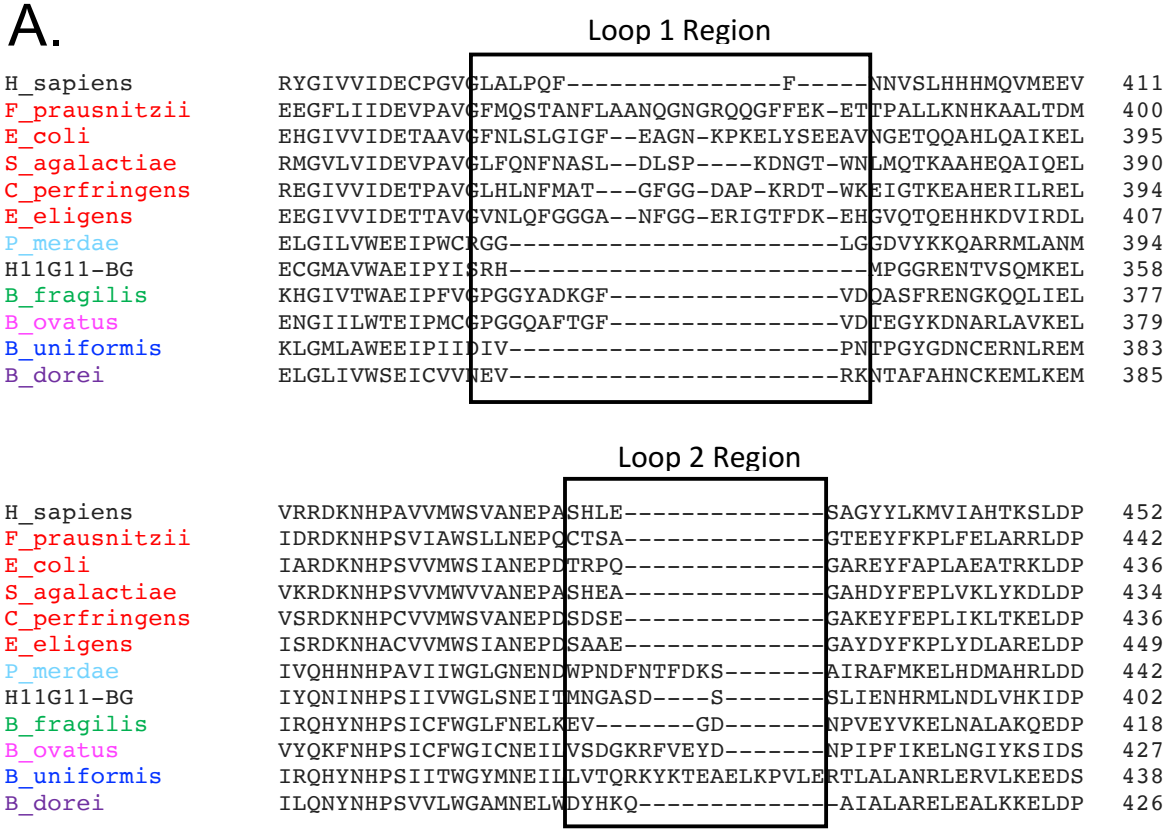

Min
4
Median,
Mode
23
Max
40
Mean
21.6
13.4
29.8
St. Dev. 8.2
Number Individuals
with this Number of GUS Sequences
Number GUS Sequences
per Individual

Figure 6.3. Identification of Novel GUS Sequences from HMGI- HMP Gene Indices Database

A. 3013 GUS Sequences were identified from the HMGI database. N is the total number of GUS sequences in each individual. 139 individuals were sampled.

B. Number of GUS Sequences found in each individual in the HMGI database.

A.


Loop 1 Region
H_sapiens RYGIVVIDECPGVGLALPQF---------------F-----NNVSLHHHMQVMEEV 411
F_prausnitzii EEGFLIIDEVPAVGFMQSTANFLAANQGNGRQQGFFEK-ETTPALLKNHKAALTDM 400
E_coli EHGIVVIDETAAVGFNLSLGIGF--EAGN-KPKELYSEEAVNGETQQAHLQAIKEL 395
S_agalactiae RMGVLVIDEVPAVGLFQNFNASL--DLSP----KDNGT-WNLMQTKAAHEQAIQEL 390
C_perfringens REGIVVIDETPAVGLHLNFMAT---GFGG-DAP-KRDT-WKEIGTKEAHERILREL 394
E_eligens EEGIVVIDETTAVGVNLQFGGGA--NFGG-ERIGTFDK-EHGVQTQEHHKDVIRDL 407
P_merdae ELGILVWEEIPWCRGG----------LGGDVYKKQARRMLANM 394
H11G11-BG ECGMAVWAEIPYISRH----------MPGGRENTVSQMKEL 358
B_fragilis KHGIVTWAEIPFVGPGGYADKGF----------VDQASFRENGKQQLIEL 377
B_ovatus ENGIILWTEIPMCGPGGQAFTGF----------VDTEGYKDNARLAVKEL 379
B_uniformis KLGMLAWEEIPIIDIV----------PNTPGYGDNCERNLREM 383
B_dorei ELGLIVWSEICVVNEV----------RKNTAFAHNCKEMLKEM 385
Loop 2 Region
H_sapiens VRRDKNHPAVVMWSVANEPASHLE----------SAGYYLKMVIAHTKSLDP 452
F_prausnitzii IDRDKNHPSVIAWSLLNEPQCTSA----------GTEEYFKPLFELARRLDP 442
E_coli IARDKNHPSVVMWSIANEPDTRPQ----------GAREYFAPLAEATRKLDP 436
S_agalactiae VKRDKNHPSVVMWVVANEPASHEA----------GAHDYFEPLVKLYKDLDP 434
C_perfringens VSRDKNHPCVVMWSVANEPDSDSE----------GAKEYFEPLIKLTKELDP 436
E_eligens ISRDKNHACVVMWSIANEPDSAAE----------GAYDYFKPLYDLARELDP 449
P_merdae IVQHHNHPAVIIWGLGNENDWPNDFNTFDKS-------AIRAFMKELHDMAHRLDD 442
H11G11-BG IYQNINHPSIIVWGLSNEITMNGASD----S-------SLIENHRMLNDLVHKIDP 402
B_fragilis IRQHYNHPSICFWGLFNELKEV-------GD-------NPVEYVKELNALAKQEDP 418
B_ovatus VYQKFNHPSICFWGICNEILVSDGKRFVEYD-------NPIPFIKELNGIYKSIDS 427
B_uniformis IRQHYNHPSIITWGYMNEILLVTQRKYKTEAELKPVLERTLALANRLERVLKEEDS 438
B_dorei ILQNYNHPSVVLWGAMNELWDYHKQ----------AIALARELEALKKELDP 426

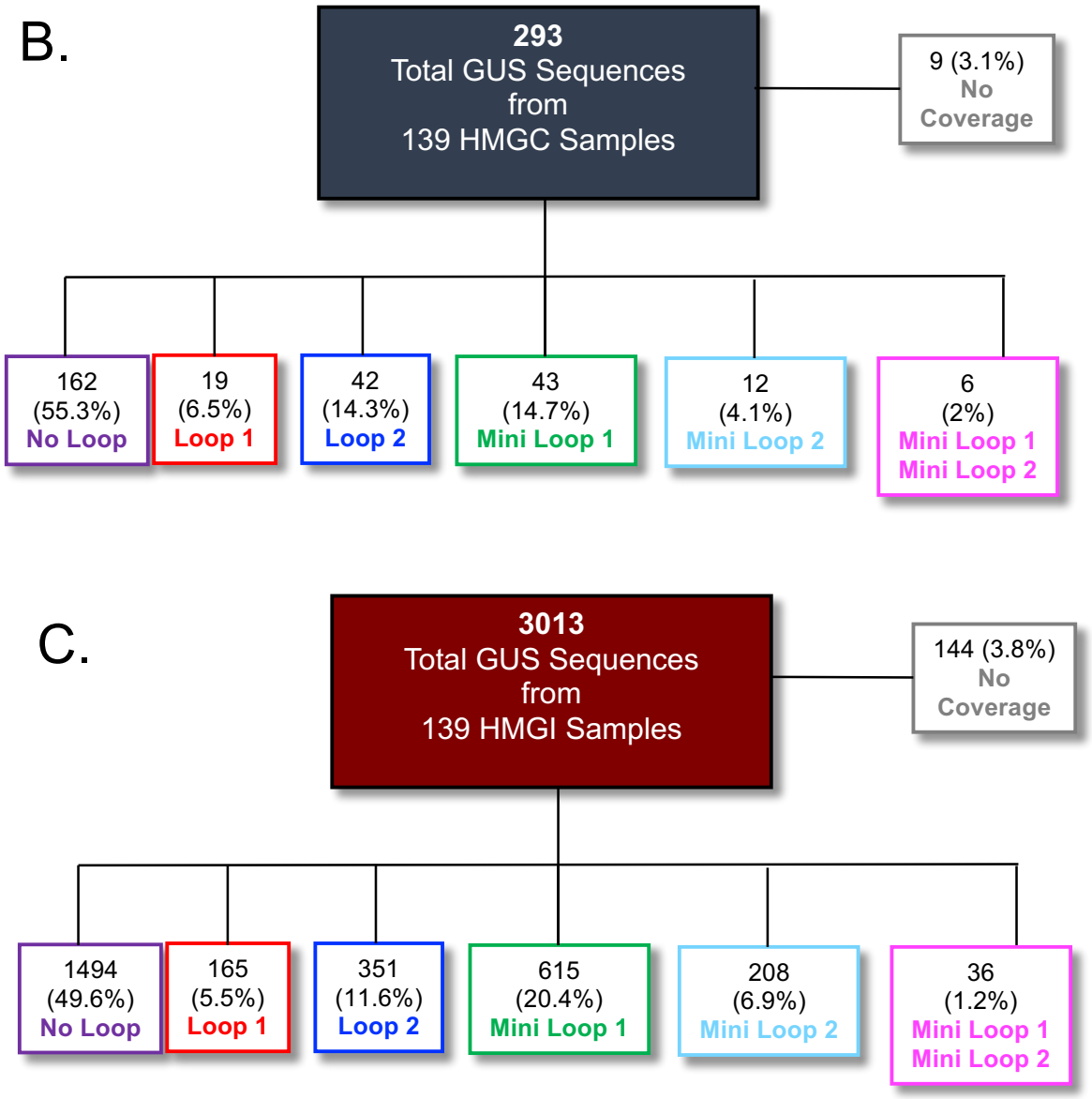
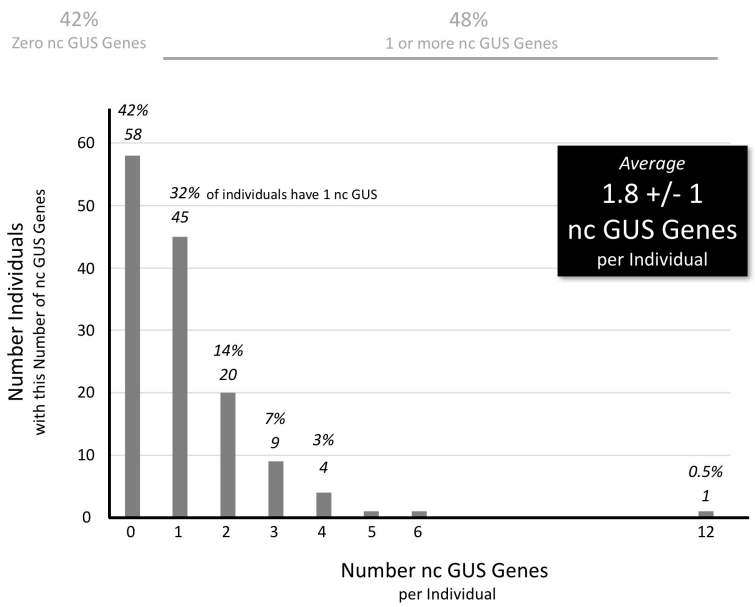
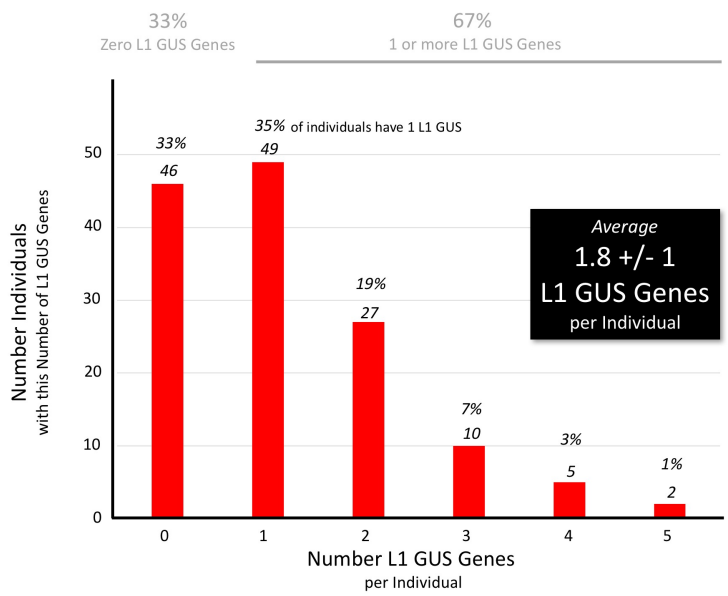


Figure 6.4. Classification of Novel GUS Sequences by Loop
A. Example of multiple sequence alignment used to assign loop category according to the specifications set out in Table 6.2. The bacterium from which each sequence was taken is color-coded for the assigned loop class. Purple- no loop, red-Loop 1, royal blue- Loop 2, green- Mini Loop 1, light blue- Mini Loop 2, magenta- Mini Loop 1 Mini Loop 2.
B. Distribution of sequences in each loop class of the 293 Non-redundant Novel GUS Sequences found in the HMGC database.
C. Distribution of sequences in each loop class of the 3013 GUS Sequences found in the HMGI database.

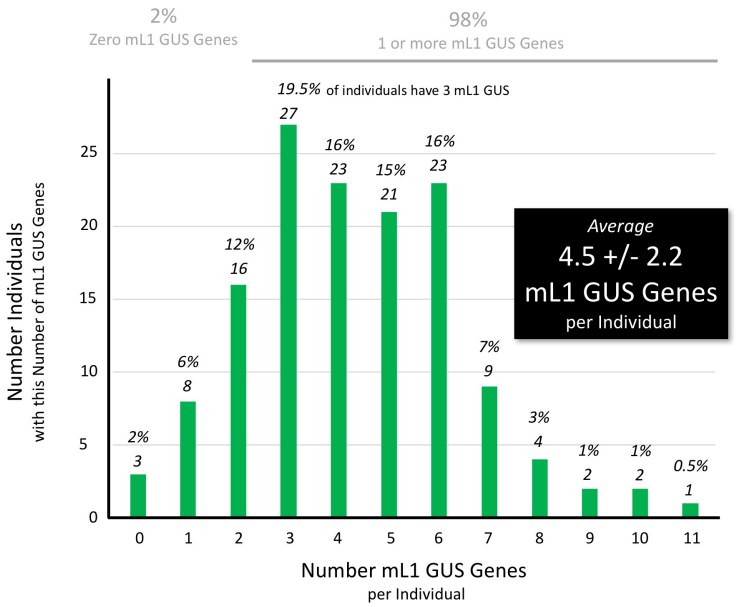
A.


42%
Zero nc GUS Genes
48%
1 or more nc GUS Genes
42%
58
32% of individuals have 1 nc GUS
45
14%
20
7%
9
3%
4
0.5%
1
Average
1.8 +/- 1
nc GUS Genes
per Individual
Number Individuals
with this Number of nc GUS Genes
0
10
20
30
40
50
60
0
1
2
3
4
5
6
12
Number nc GUS Genes
per Individual

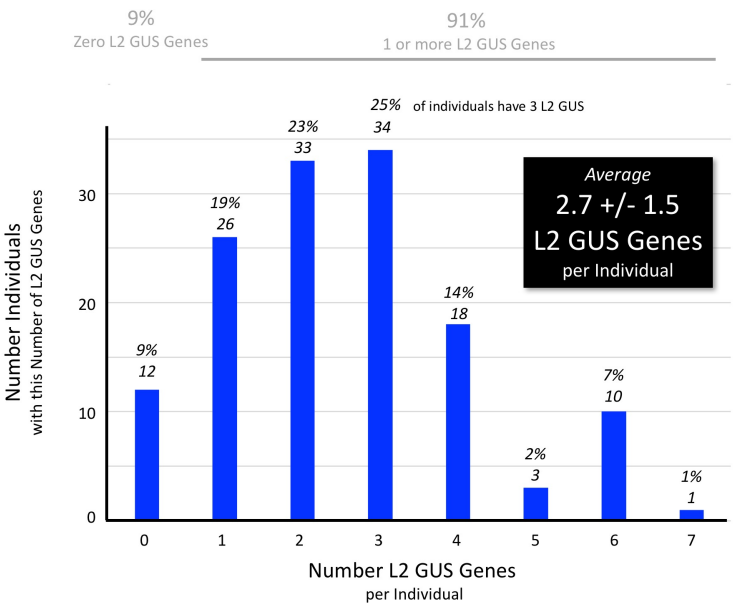
B.


33%
Zero L1 GUS Genes
67%
1 or more L1 GUS Genes
33%
46
35% of individuals have 1 L1 GUS
49
19%
27
7%
10
3%
5
1%
2
Average
1.8 +/- 1
L1 GUS Genes
per Individual
Number Individuals
with this Number of L1 GUS Genes
0
10
20
30
40
50
0
1
2
3
4
5
Number L1 GUS Genes
per Individual

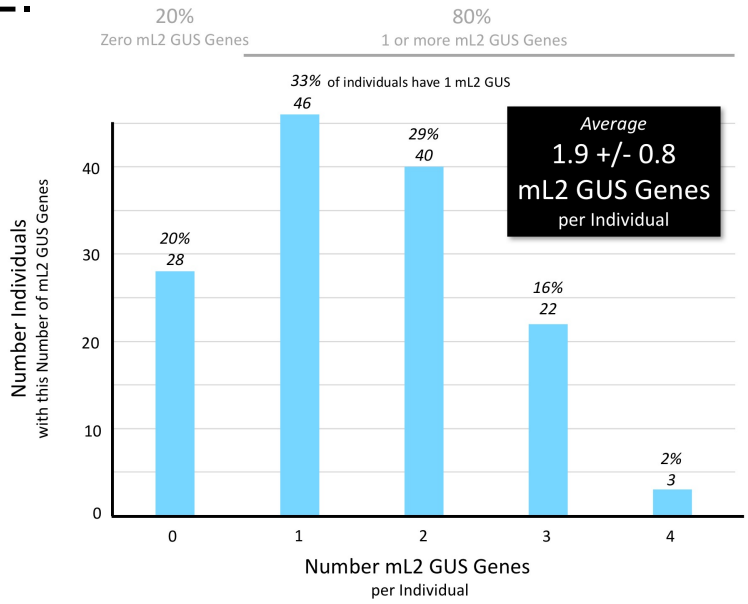
C.


2%
Zero mL1 GUS Genes
98%
1 or more mL1 GUS Genes
19.5% of individuals have 3 mL1 GUS
Number Individuals
with this Number of mL1 GUS Genes
Average
4.5 +/- 2.2
mL1 GUS Genes
per Individual
2%
3
6%
8
12%
16
27
16%
23
15%
21
16%
23
7%
9
3%
4
1%
2
1%
2
0.5%
1
0
5
10
15
20
25
0 1 2 3 4 5 6 7 8 9 10 11
Number mL1 GUS Genes
per Individual

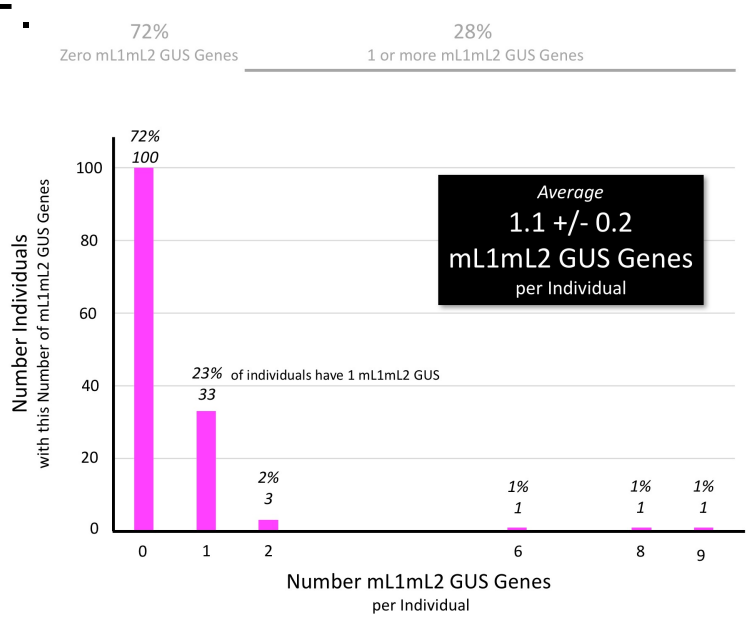
D.


9%
Zero L2 GUS Genes
91%
1 or more L2 GUS Genes
25% of individuals have 3 L2 GUS
Number Individuals
with this Number of L2 GUS Genes
Average
2.7 +/- 1.5
L2 GUS Genes
per Individual
9%
12
19%
26
23%
33
25%
34
14%
18
2%
3
7%
10
1%
1
0
10
20
30
0 1 2 3 4 5 6 7
Number L2 GUS Genes
per Individual

E.


20%
Zero mL2 GUS Genes
80%
1 or more mL2 GUS Genes
33% of individuals have 1 mL2 GUS
46
29%
40
20%
28
16%
22
2%
3
Average
1.9 +/- 0.8
mL2 GUS Genes
per Individual
Number Individuals
with this Number of mL2 GUS Genes
0
10
20
30
40
0
1
2
3
4
Number mL2 GUS Genes
per Individual

F.


72%
Zero mL1mL2 GUS Genes
28%
1 or more mL1mL2 GUS Genes
72%
100
23% of individuals have 1 mL1mL2 GUS
33
2%
3
1%
1
1%
1
1%
1
Average
1.1 +/- 0.2
mL1mL2 GUS Genes
per Individual
Number Individuals
with this Number of mL1mL2 GUS Genes
0
20
40
60
80
100
0
1
2
6
8
9
Number mL1mL2 GUS Genes
per Individual

G.

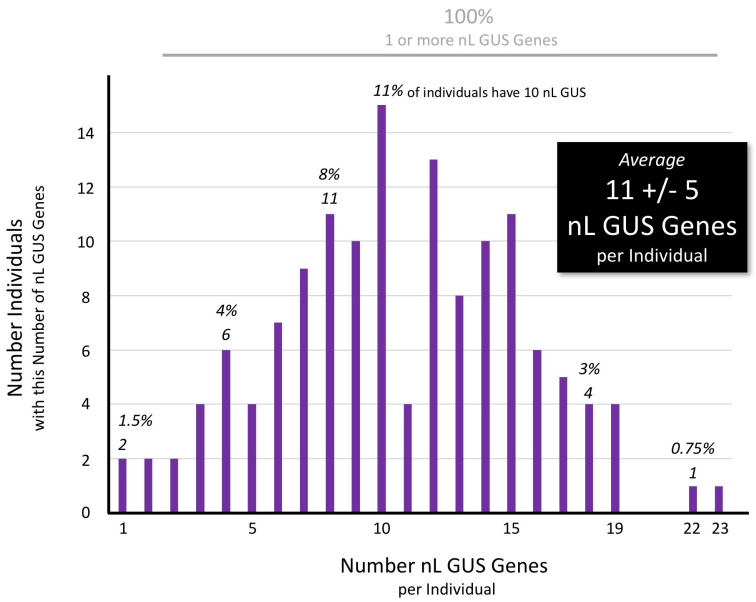
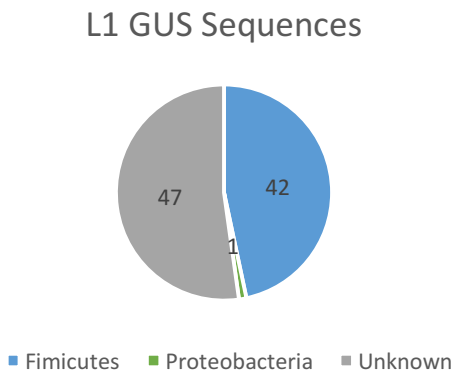


Figure 6.5. Number of GUS Sequences of Each Loop Class Found per Individual

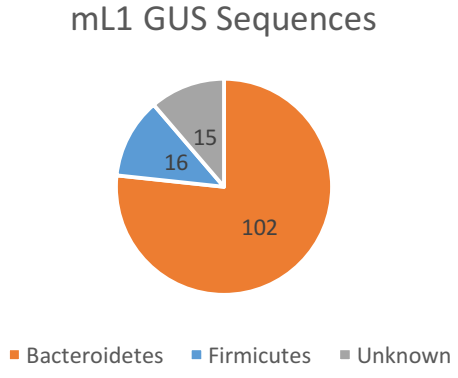
A. Number of no coverage (nc) GUS Genes per individual.
B. Number of Loop 1 (L1) GUS Genes per individual.
C. Number of Mini Loop 1 (mL1) GUS Genes per individual.
D. Number of Loop 2 (L2) GUS Genes per individual.
E. Number of Mini Loop 2 (mL2) GUS Genes per individual.
F. Number of Mini Loop 1 Mini Loop 2 (mL1mL2) GUS Genes per individual.
G. Number of No Loop (nL) GUS Genes per individual.

A.



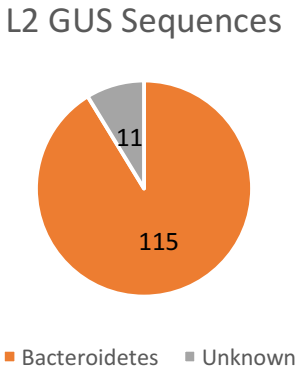
Name	Phylum	Count
Faecalibacterium prausnitzii	**Firmicutes**	**13**
Eubacterium eligens CAG:72	**Firmicutes**	**11**
Eubacterium sp. CAG:76	Firmicutes	7
Eubacterium sp. CAG:38	Firmicutes	5
Eubacterium eligens	Firmicutes	4
Uncultured bacterium (fragment)	Unknown	4

B.



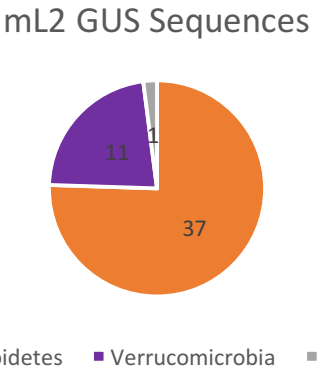
Name	Phylum	Count
Bacteroides vulgatus	Bacteroidetes	11
Bacteroides vulgatus	Bacteroidetes	11
Parabacteroides merdae	Bacteroidetes	9
Bacteroides sp. CAG:98	Bacteroidetes	8
Bacteroides dorei	Bacteroidetes	7

C.



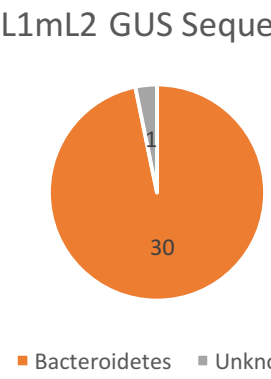
Name	Phylum	Count
Bacteroides MSP	Bacteroidetes	15
Bacteroides uniformis	**Bacteroidetes**	**11**
Bacteroides vulgatus	Bacteroidetes	9
Bacteroides sp. 3_1_23	Bacteroidetes	9
Bacteroides uniformis	Bacteroidetes	8

D.



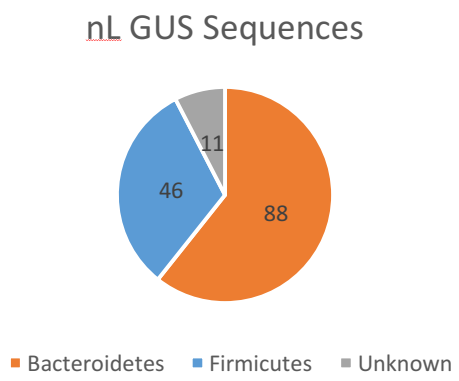
Name	Phylum	Count
Parabacteroides merdae	**Bacteroidetes**	**38**
Bacteroides MSP	Bacteroidetes	33
Akkermansia sp. CAG:344	Verrucomicrobia	10
Bacteroides cellulosilyticus	Bacteroidetes	10
Bacteroides cellulosilyticus	Bacteroidetes	6

E.



Name	Phylum	Count
Bacteroides ovatus	**Bacteroidetes**	**12**
Bacteroides ovatus	Bacteroidetes	12
Bacteroides 754	Bacteroidetes	4
Bacteroides MSP	Bacteroidetes	2

F.



Name	Phylum	Count
Bacteroides uniformis	Bacteroidetes	24
Bacteroides dorei	**Bacteroidetes**	**19**
Bacteroides vulgatus	Bacteroidetes	9
Parabacteroides	Bacteroidetes	8
Bacteroides	Bacteroidetes	7

Figure 6.6. Predicted Bacteria Containing GUS Sequence of Each Loop Class
For each of individual, the most abundant GUS sequence in each loop class was matched to a deposited reference genome using BLASTp. The matched organism, phylum of that organism, and number of individuals for which each sequence was the most abundant sequence of that loop class (count) is presented for the top five most common sequences for the Loop 1 (A), mini Loop 1 (B), Loop 2 (C), mini Loop 2 (D), mini Loop 1 mini Loop 2 (E), and no loop (F) classes. Bolded sequences were chosen for further characterization.

REFERENCES

1. **Ridlon JM**, **Kang DJ**, **Hylemon PB**, **Bajaj JS**. 2014. Bile acids and the gut microbiome. Current Opinion in Gastroenterology **30**:332–338.

2. **Koropatkin NM**, **Cameron EA**, **Martens EC**. 2012. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol **10**:323–335.

3. **Kim D-H**. 2015. Gut Microbiota-Mediated Drug-Antibiotic Interactions. Drug Metab Dispos **43**:1581–1589.

4. **Mathijssen RH**, **van Alphen RJ**, **Verweij J**, **Loos WJ**, **Nooter K**, **Stoter G**, **Sparreboom A**. 2001. Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). Clin Cancer Res **7**:2182–2194.

5. **Alimonti A**, **Gelibter A**, **Pavese I**, **Satta F**, **Cognetti F**, **Ferretti G**, **Rasio D**, **Vecchione A**, **Di Palma M**. 2004. New approaches to prevent intestinal toxicity of irinotecan-based regimens. Cancer Treat Rev **30**:555–562.

6. **Pommier Y**. 2006. Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer **6**:789–802.

7. **Chabot GG**. 1996. Clinical Pharmacology and Pharmacodynamics of Irinotecan. Annals of the New York Academy of Sciences **803**:164–172.

8. **Ma MK**, **McLeod HL**. 2003. Lessons learned from the irinotecan metabolic pathway. Curr Med Chem **10**:41–49.

9. **Gupta E**, **Lestingi TM**, **Mick R**, **Ramirez J**, **Vokes EE**, **Ratain MJ**. 1994. Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. Cancer Res **54**:3723–3725.

10. **Pérez-Roth E**, **Kwong SM**, **Alcoba-Florez J**, **Firth N**, **Méndez-Alvarez S**. 2010. Complete nucleotide sequence and comparative analysis of pPR9, a 41.7-kilobase conjugative staphylococcal multiresistance plasmid conferring high-level mupirocin resistance. Antimicrobial Agents and Chemotherapy **54**:2252–2257.

11. **Dranitsaris G**, **Maroun J**, **Shah A**. 2005. Estimating the cost of illness in colorectal cancer patients who were hospitalized for severe chemotherapy-induced diarrhea. Can J Gastroenterol **19**:83–87.

12. **Stein A**, **Voigt W**, **Jordan K**. 2010. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol **2**:51–63.

13. **Wallace BD**, **Wang H**, **Lane KT**, **Scott JE**, **Orans J**, **Koo JS**, **Venkatesh M**, **Jobin C**, **Yeh L-A**, **Mani S**, **Redinbo MR**. 2010. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science **330**:831–835.

14. **Wallace BD**, **Roberts AB**, **Pollet RM**, **Ingle JD**, **Biernat KA**, **Pellock SJ**, **Venkatesh MK**, **Guthrie L**, **O'Neal SK**, **Robinson SJ**, **Dollinger M**, **Figueroa E**, **McShane SR**, **Cohen RD**, **Jin J**, **Frye SV**, **Zamboni WC**, **Pepe-Ranney C**, **Mani S**, **Kelly L**, **Redinbo MR**. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem Biol **22**:1238–1249.

15. **Roberts AB**, **Wallace BD**, **Venkatesh MK**, **Mani S**, **Redinbo MR**. 2013. Molecular Insights into Microbial β-Glucuronidase Inhibition to Abrogate CPT-11 Toxicity. Molecular Pharmacology **84**:208–217.

16. **Human Microbiome Project Consortium**. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**:207–214.

17. **Langmead B**, **Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods **9**:357–359.

18. **Liao Y**, **Smyth GK**, **Shi W**. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923–930.

CHAPTER 7: FUNCTIONAL CHARACTERIZATION OF β-GLUCURONIDASE ENZYMES IDENTIFIED IN THE HUMAN GASTROINTESTINAL TRACT

Introduction

Changes in the human gut microbiota have been linked to many disorders including obesity, inflammatory bowel disease, and intensity of *Clostridium difficile* infection (1, 2). The activity of specific microbes in the gut has also been linked to the development of heart disease and varied response to drug therapy (3, 4). One drug that has varying effects because of the microbes in the intestine is the chemotherapy irinotecan (CPT-11). The active form of irinotecan, SN-38, is inactivated in the liver during Phase II drug metabolism by the addition of a glucuronide sugar to form SN-38-glucuronide (SN-38G) (5-7). SN-38G is then excreted from the body through the intestines. However, the β-glucuronidase (GUS) enzymes expressed by many microbes in the intestine can cleave the glucuronide sugar off SN-38G, resulting in active SN-38 that kills the intestinal epithelium and leads to dose-limiting diarrhea (8-11).

To better understand the activity of the GUS enzymes responsible for this drug toxicity, the Redinbo lab characterized the GUS enzyme from *E. coli* (*Ec*GUS) and identified small molecule inhibitors that have been shown to be effective in decreasing the side effects of irinotecan in mice (12, 13). Acknowledging that the

microbiota is made up of a wide range of bacteria, we also characterized the GUS enzymes from *Streptococcus agalactiae* (*Sa*GUS), *Clostridium perfringens* (*Cp*GUS), and *Bacteroides fragilis* (*Bf*GUS) (14). Characterization of these enzymes with inhibitors revealed a flexible loop outside the active site referred to as Loop 1 (L1) that is essential for the efficacy of selective GUS inhibitors. The structures of these four enzymes revealed key features of GUS enzymes including conserved residues in the active site (14).

These key GUS features were used to compile a catalog of all GUS sequences found in the Human Microbiome Project (HMP) sequencing repository as described in Chapter 6. We identified 3,013 GUS sequences from the stool samples of 139 individuals which we were then able to break down into classes according to the presence, size, and location of a flexible loop outside the active site. As mentioned above, such a loop (L1) had previously been characterized but analysis of this GUS sequence catalog revealed a second loop (Loop 2, L2) found just downstream in the GUS sequence and predicted to occupy a similar position to L1. These loops can be completely absent or can be smaller in size (mini) than the previously characterized L1 found in *Ec*GUS, *Sa*GUS, and *Cp*GUS. This gave us six loop classes: Loop 1 (L1), mini Loop 1 (mL1), Loop 2 (L2), mini Loop 2 (mL2), mini Loop1 mini Loop 2 (mL1mL2), and no loop (nL). We chose sequences from each class to express, purify, and characterize in order to further expand understanding of GUS enzyme function. These include the GUS enzymes from *Eubacterium eligens* (*Ee*GUS, L1), *Faecalibacterium prausnitizii* (*Fp*GUS, L1), *Bacteroides uniformis* (*Bu*GUS, L2), *Parabacteroides merdae* (*Pm*GUS, mL2), *Bacteroides ovatus*

(*Bo*GUS, mL1mL2), and *Bacteroides dorei* (*Bd*GUS, nL). The full sequence for each of these enzymes is presented in Appendix 3 in the alignment used to define their loop class.

The following data confirm each of these GUS sequences as an active GUS enzyme and show initial characterization of the parameters under which the enzymes show optimal function. Because interest in this enzyme class originally stemmed from the ability to process SN-38G in the human intestine, we present the first evidence that a wide range of GUS enzymes are able to process the compound SN-38G. Finally, initial characterization of the GUS enzyme from *Lactobacillus rhamnosus* (LrGUS) is presented. This enzyme was not identified through analysis of the HMP, but rather by culturing bacteria found in a colon polyp biopsy sample. While this work is still ongoing, it greatly expands the current understanding of bacteria GUS enzymes and gives a better picture of the vast amount of chemistry these enzymes are responsible for in the human intestine.

Materials and Methods

Cloning of GUS Enzymes

Clostridium perfringens GUS

The *C. perfringens* β-glucuronidase (*Cp*GUS) gene was amplified from the pLIC-MBP expression vector used in Wallace *et al.* and inserted into the pLIC-His vector using the primers: Forward, 5'-TACTCCAATCCAATGCGATGTTATATCCAA TAATTACAGAATCAAGACAATTAATAGAC-3'; Reverse, 5'-TTATCCACTTCCAATG CGCTATTTTTTGTATCCAAATTCCGGTATATTTGTCCATCTTTC-3' (12).

***Eubacterium eligens* GUS**

The codon-optimized *E. eligens* β-glucuronidase (*Ee*GUS) gene was purchases from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'-TACTTCCAATCCAATGCGATGCTGTACCCTGTCCTGACTCAGTCTCG-3'; Reverse, 5'-TTATCCACTTCCAATGCGCTATTTGGTTTTGTAGCCAAATTCCGGAATGGTGGACCAG-3'.

***Faecalibacterium prausnitizii* GUS**

The *F. prausnitizii* β-glucuronidase (*Fp*GUS) gene was purchased from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'-TACTTCCAATCCAATGCGATGAACCGTAGCCTGCTGTACCCTCGTG-3'; Reverse, 5'- TATCCACTTCCAATGCGCTTTTTTTGCGTTTTTTGAAGTCAACCG-3'.

***Bacteroides uniformis* GUS**

The full-length *B. uniformis* β-glucuronidase (*Bu*GUS) gene was purchased from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'- TACTTCCAATCCAATGCGATGGAGCGTGAGAAGAACACCCTGCCTCAAAAG-3'; Reverse, 5'-TTATCCACTTCCAATGCGCTAATAAATGTTGCGCAGTTTAATGCCATTCAGGAAACACG-3'. The mature gene lacking the signal peptide was also cloned into the pLIC-His vector using the same reverse primer and the forward primer 5'- TACTTCCAATCCAATGCGCAACGTCAAACTCAAACCATCAACGACTCCTGGAAG-3'.

***Parabacteroides merdae* GUS**

The full-length *P. merdae* β-glucuronidase (*Pm*GUS) gene was purchased from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'-TACTTCCAATCCAATGCGATGAAA TACCTGTTCGTCGCTTGCCTGCTGTG-3'; Reverse, 5'-TTATCCACTTCCAATGCG CTATTTCGGGCTTTTCAGTTCCAGAAACGCGGTC-3'.

***Bacteroides ovatus* GUS**

The full-length *B. ovatus* β-glucuronidase (*Bo*GUS) gene was purchased from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'- TACTTCCAATCCAATGCGATGAAGAA CCGCATCATTATTCTGTGCCTGGTATGCCTGTG-3'; Reverse, 5'- TTATCCACTTC CAATGCGCTATTTAATGCAGTACCATTCGCAGGTGTCGG-3'. The mature gene lacking the signal peptide was also cloned into the pLIC-His vector using the primers: Forward, 5'- TACTTCCAATCCAATGCGCAGGAAACTTCTCCGCGTA CTATCTTCTCTCTGAAC-3'; Reverse, 5'-TTATCCACTTCCAATGCGCTATTTA ATGCAGTACCATTCGCAGGTGTCGGACAG-3'.

***Bacteroides dorei* GUS**

The full-length *B. dorei* β-glucuronidase (*Bd*GUS) gene was purchased from Biobasic in the pUC57 vector. The gene was amplified and inserted into the pLIC-His vector using the primers: Forward, 5'-TACTTCCAATCCAATGCGATGAAGCGTT TCGCAGGTTGGCTGCTGTTCTTC-3'; Reverse, 5'-TTATCCACTTCCAATGCGCT

AGCGAATTTTTTTAACTTTCAGGCCGCTAATTACGCCCTG-3'. The mature gene lacking the signal peptide was also cloned into the pLIC-His vector using the same reverse primer and the forward primer 5'-TACTTCCAATCCAATGCGAGCGAGAT CAGCATCACTGACTCTTGGAAGTATAAG-3'.

Lactobacillus rhamnosus GUS

Lactobacillus rhamnosus was isolated from colon polyp biopsies and cultured by the Keku lab at the University of North Carolina at Chapel Hill (UNC-CH). The bacteria were cultured in clostridial media (Becton & Dickinson) for 48 hours in an anaerobic chamber at 37°C. 1 ml of bacterial culture was pelleted and the resulting cell pellet used for genomic DNA isolation. Genomic DNA was isolated using the GeneJet Genomic DNA Purification Kit (Thermo Scientific).

The β-glucuronidase gene from *L. rhamnosus* (*Lr*GUS) was amplified via PCR using the primers: Forward, 5'-TACTTCCAATCCAATGCGATGGAGACATC GTTGTTATACCCAGTGAC-3'; Reverse, 5'-TTATCCACTTCCAATGCGCTCTTT GCTTTATAATCCAGCGGCAGCTTATTC-3'. The gene was then inserted into the pLIC-His vector using ligation-independent cloning (Sondek Lab, UNC-CH).

Expression and Purification of GUS Enzymes

E. coli GUS

E. coli β-glucuronidase (*Ec*GUS) was expressed and purified as previously described (12).

Streptococcus agalactiae GUS

S. agalactiae β-glucuronidase (*Sa*GUS) was expressed and purified as previously described (14).

***Cp*GUS**

The newly cloned His-tagged protein was expressed and purified as described for the MBP-tagged *Cp*GUS in Wallace *et al.* (14).

***Bf*GUS**

The *Bf*GUS gene in the pSGX3 plasmid was transformed into BL21 DE3 Gold cells for enzyme expression. Cells were grown in the presence of kanamycin in LB medium with vigorous shaking at 37ºC to an OD_{600} of 0.5, at which point the temperature was reduced to 18ºC. At OD_{600}=0.8, protein expression was induced by the addition of 0.1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) and incubation continued overnight. Cells were collected by centrifugation at 4500xg for 20 min at 4ºC in a Sorvall (model RC-3B) swinging bucket centrifuge. Cell pellets were resuspended in 50 mL Buffer A (50 mM Tris HCl pH 7.8, 50 mM Imidazole, 10 mM methionine, 500 mM NaCl, 1 mM DTT, 5% glycerol), DNase, lysozyme, and a Roche complete-EDTA free protease inhibitor tablet. Resuspended cells were sonicated and clarified via centrifugation at 14,500xg for 60 min in a Sorvall (model RC-5B). The lysate was flowed over a Ni-NTA HP column (GE Healthcare) loaded onto the Aktaxpress FPLC system (Amersham Bioscience) and washed with Buffer A. Protein was eluted with Buffer B (50 mM Tris HCl pH 7.8, 500 mM Imidazole, 10 mM methionine, 500 mM NaCl, 1 mM DTT, 5% glycerol). Fractions containing the protein of interest were combined and passed over a HiLoad™ 16/60 Superdex™ 200 gel filtration column. Protein was eluted in S200 Buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 10 mM methionine, 1mM DTT, 5% glycerol). Fractions were analyzed by SDS-

PAGE and those with >95% purity were combined and concentrated for long-term storage at -80ºC.

_Ee_GUS, _Fp_GUS, _Bu_GUS, _Pm_GUS, _Bo_GUS, and _Bd_GUS

The corresponding plasmids described above were each transformed into BL21-DE3 AI competent cells (Invitrogen) for enzyme expression. Cells were grown in LB medium to an OD_{600} of 0.8 in the presence of ampicillin with vigorous shaking at 37ºC. Transcription of the T7 RNA polymerase was induced by addition of arabinose to a final concentration of 0.2% (v/v) and the temperature was reduced to 18ºC. Thirty minutes later protein expression was induced by the addition of 100 µM IPTG and further incubation overnight. Cells were collected, lysed, and purified as described above for *Bf*GUS with the following buffers: Buffer A- 25 mM HEPES pH 7.4, 50 mM Imidazole, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP; Buffer B- 25 mM HEPES pH 7.4, 250 mM Imidazole, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP; S200 Buffer- 20 mM HEPES pH 7.4, 100 mM NaCl, 5% glycerol, 0.5 mM TCEP. Protein with >95% purity was stored at -80ºC for use in assays.

PNPG Assay

Para-nitrophenyl glucuronide (PNPG) was purchased as a solid and suspended in water to a concentration of 50 mM. Reactions were conducted in 96-well, black, clear-bottom assay plates (Costar, Tewksbury MA) at 37ºC with 50 µL in total volume. The reaction consisted of 10 µL assay buffer (100 mM HEPES, 250 mM NaCl, pH 7.4), 10 µL enzyme, and 30 µL of substrate. The product formation was measured via absorbance at 405 nm using a PHERAstar *Plus* microplate reader (BMG Labtech, Ortenberg, Germany). To determine kinetic values, initial velocities

were determined for multiple substrate concentrations and Michaelis-Menten kinetics use to calculate K_M, k_{cat}, and catalytic efficiency. For pH screening, reactions were conducted using assay buffer at pH 4, 4.5, 5, 5.5, 6, 6.5, 7, and 7.4. At each time point, the reaction was quenched with 100 µL of 0.2 M sodium carbonate. When all reactions were quenched, product formation was measured.

SN-38G HPLC

Each reaction consisted of 500 nM bacterial GUS enzyme or 2 µM BovineGUS enzyme and 10 µM SN-38G substrate in a total volume of 250 µL. Reactions were incubated at room temperature and a 20 µL sample taken at each time point and loaded directly on to the HPLC column.

GUS enzyme processing of SN-38G was analyzed by separation of SN-38 and SN-38G via HPLC. Buffer was flowed at 0.5 ml/min over a Phenomenex Luna 5 µm C18(2) reverse phase column. Buffer A consisted of 20 mM ammonium acetate pH 3.5 and Buffer B was 100% acetonitrile. SN38 and SN38G was visualized using UV detection with excitation at 368 nm and emission at 515 nm. Buffers were mixed as follows:

Time	% B
1 min	10%
3 min	30%
12 min	34%
14 min	34 %
15 min	35%
19 min	100%
22 min	10%
22.5 min	STOP

SN-38G eluted at approximately 8 minutes while SN-38 eluted at 13.3 minutes. The area under the curve was calculated for each peak and a rate used to calculate the

percent signal in the SN-38G peak. The percent SN-38G is reported for each time point.

SN-38G Assay

SN-38-glucuronide (SN-38G) was purchased as a solid from Toronto Research Chemicals-Canada and suspended in 100% DMSO to a concentration of 5 mM. Reactions were conducted in 96-well, half-area, black, clear-bottom assay plates (Costar, Tewksbury MA) at 37°C with 50 µL in total volume. The reaction consisted of 10 µL assay buffer (100 mM HEPES, 250 mM NaCl, pH 7.4), 10 µL enzyme, and 30 µL of substrate. The disappearance of substrate as it is processed by GUS was measured via Ex 375 nm Em 415 nm fluorescence. Initial velocities (Fu/sec) were determined for multiple substrate concentrations.

Results

GUS Activity

Analysis of the Human Microbiome Project (HMP) sequencing repository as described in Chapter 6 allowed for the identification of novel GUS sequences from which we chose six representative proteins to characterize in addition to the four previously characterized enzymes. These new proteins will also allow us to explore the activity of each of the newly defined loop classes. The Loop 1 (L1) class is represented by the GUS enzymes from *E. coli* (*Ec*GUS), *S. agalactiae* (*Sa*GUS), *C. perfringens* (*Cp*GUS), *E. eligens* (*Ee*GUS), and *F. prausnitizii* (*Fp*GUS, L1). The mini Loop 1 (mL1) class is represented by the previously characterized *B. fragilis* (*Bf*GUS). The Loop 2 (L2) class is represented by *B. uniformis* (*Bu*GUS) and the

mini Loop 2 (mL2) class by *P. merdae* (*Pm*GUS). *B. ovatus* (*Bo*GUS) serves as the representative mini Loop 1 mini Loop 2 (mL1mL2) enzyme and *B. dorei* (*Bd*GUS) is the no loop (nL) enzyme. We first set out to confirm activity of each purified enzyme using the substrate *p*-nitrophenyl glucuronide (PNPG). PNPG is a synthetic substrate with no physiological relevance but provides a quick and robust assay for visualizing GUS activity. As the four previously characterized enzymes had substantial activity at pH 7.4, this reaction condition was used initial characterization of the new enzymes as it is unclear where in the intestines these bacteria and their GUS may exist and therefore what the physiologically relevant pH may be.

The Redinbo lab has previously characterized Loop 1 (L1) enzymes *Ec*GUS, *Sa*GUS, and the MBP-tagged version of *Cp*GUS and the kinetic parameters of these enzymes is presented for comparison (12, 14). From the novel GUS sequences identified from the HMP sequencing repository, we identified *Ee*GUS and *Fp*GUS as sequences that are often the most abundant L1 GUS in individuals. We were able to confirm that these enzymes are able to process PNPG. *Ee*GUS showed moderate but still substantial activity as compared to the efficient *Ec*GUS and *Sa*GUS enzymes with a k_{cat} of 41 ± 3 s^{-1}, K_M of 0.49 ± 0.09 mM, and a catalytic efficiency of 83.7 $s^{-1}mM^{-1}$ (Table 7.1). *Fp*GUS showed poor activity; while an increase in PNP could be seen, the production was too slow to be quantified in this assay. *Bf*GUS has been previously characterized and is considered the representative mL1 enzyme (14). Representative L2 enzyme *Bu*GUS showed a catalytic efficiency similar to that of *Bf*GUS (9 $s^{-1}mM^{-1}$ and 9.5 $s^{-1}mM^{-1}$ respectably); however, this is a result of varying k_{cat} and K_M values. *Bf*GUS displays a k_{cat} of 18 ± 1 s^{-1} and a K_M of

1.9 ± 0.3 mM as compared to a k_{cat} of 4.5 ± 0.007 s^{-1} and a K_M of 0.5 ± 0.03 mM for *Bu*GUS (Table 7.1). Both mL2 enzyme *Pm*GUS and mL1mL2 enzyme *Bo*GUS showed activity that was too slow to quantify in this assay. However, we did see significant activity over no protein controls that indicates that these enzymes do function as GUS enzymes. Finally, no loop (nL) enzyme *Bd*GUS exhibited poor but quantifiable activity with a k_{cat} of 21 ± 2 s^{-1}, K_M of 2.7 ± 0.5 mM, and a catalytic efficiency of 7.7 $s^{-1}mM^{-1}$ (Table 7.1). This data confirms that each of these newly expressed GUS enzymes is able to process a glucuronidated substrate and validates the selection criteria used to identify GUS sequences in Chapter 6. However, the poor activity seen with many of these enzymes suggests that additional optimization of purification procedures or assay conditions may be beneficial.

SN-38G Processing

Much of the interest in GUS enzymes has been driven by the discovery that they are responsible for cleaving the glucuronide sugar off inactivate SN-38G that excreted through the intestines. This cleavage results in active SN-38 in the intestines that leads to epithelial cell death and significant drug toxicity for many patients. With this as the focus, we set out to characterize the ability of each of the GUS enzymes to process SN-38G.

Previous quantification of SN-38G to SN-38 ratios and, indeed, the work that established GUS enzymes as the driver of this drug toxicity has been carried out using HPLC quantification of SN-38G and SN-38 presence in samples. We set out to use this technique to establish whether these GUS enzymes of interest are able to

process SN-38G. For each reaction the GUS enzyme was incubated with SN-38G and samples were analyzed for the SN-38G:SN-38 ratio at various time points. The GUS from *Bos taurus* (BovineGUS) was used as a negative control as it has previously been shown that mammalian GUS enzymes have poor activity against SN-38G. Indeed, we found that over 67.5 minutes, BovineGUS was able to process only 20% of the SN-38G into SN-38 (Figure 7.1). Alternatively, *Ec*GUS, *Sa*GUS, and *Bf*GUS were all able to process SN-38G very efficiently with only trace amounts remaining after 22.5 minutes (Figure 7.1). While it had been previously established that *Ec*GUS can process SN-38G, this is the first evidence that *Sa*GUS and *Bf*GUS can also process this substrate.

Because HPLC is a time and labor-intensive assay, we developed a fluorescence-based assay to more quickly quantify GUS activity against SN-38G. As shown in Figure 7.2A, SN-38G and SN-38 exhibit different emission spectra when excited at 375 nm. This allows for selective visualization of just one compound at the peak of their emission profile. Because SN-38 has poor solubility in the aqueous solutions needed to maintain enzyme activity, we chose to measure the loss of SN-38G signal at Ex 375 nm Em 415 nm. The activity of each of the GUS enzymes was measured against varying concentrations of SN-38G and the rates of SN-38G cleavage are shown in Figure 7.2B. Here higher rates of SN-38G cleavage mean more efficient processing of SN-38G. With more replicates and controls, this data will lead to kinetic values like those established in the PNPG assay. These preliminary data mimic what was seen in the HPLC analysis of SN-38G: *Ec*GUS, *Sa*GUS, and *Bf*GUS are able to efficiently process SN-38G. Because this assay

allows for visualization of SN-38G cleavage at much shorter time points (seconds versus minutes) and at varying SN-38G concentrations, we are able to tease out the relative activities of these enzymes. *Ec*GUS is the most efficient at processing SN-38G of previously characterized enzymes while *Sa*GUS and *Bf*GUS show similar, less efficient processing. Interestingly, L1 GUS *Ee*GUS showed much more efficient SN-38G processing than *Ec*GUS. This solidified the L1 class as the dominant processors of SN-38G. No loop GUS *Bd*GUS showed an ability to process SN-38G similar to *Bf*GUS. *Bo*GUS, a mL1mL2 enzyme, was also able to process SN-38G but with low efficiency. Alternatively, no cleavage of SN-38G was seen with L2 enzyme *Bu*GUS and mL2 enzyme *Pm*GUS. While future characterization will give a much more complete picture of cleavage SN-38G by GUS enzymes, this analysis is the first evidence that enzymes within the GUS family may have substrate specificity.

GUS Activity at Varying pH

The initial characterization of the GUS enzymes presented above lead us to realize there is much about this enzyme family that has not been previously explored. Upon closer analysis we saw that the attachment of an MBP-tag to the GUS enzyme as in the *Cp*GUS construct impedes the activity of this enzyme even though the MBP tag was required for crystallization. We cloned *Cp*GUS into the pLIC-His vector to eliminate the MBP-tag and moved forward with characterization of this construct. In addition, we found that *Bu*GUS, *Pm*GUS, *Bo*GUS, and *Bd*GUS each contain an approximately 20 amino acid signal sequence at the N-terminal of each protein sequence that is predicted to be cleaved before the protein is transported to the periplasmic space of these gram-negative bacteria. In order to

produce the form of the enzyme that is most likely to interact with substrates, we cloned these genes without the signal sequence portion. Finally, as mentioned above, it is unclear where the bacteria expressing these GUS enzymes may be found in the intestine and therefore what pH the enzymes are likely to function in. GUS enzymes purified from other environments have previously been shown to have optimal activity at pH values much lower than 7.4 (15, 16). Therefore, we set out to test the activity of these optimized enzyme constructs against PNPG at a range of pH values.

*Ec*GUS and *Sa*GUS show similar activity at pH values ranging from 5 to 7.4 (Figure 7.3A, B). While both enzymes show almost no activity at pH 4, *Ec*GUS exhibits moderate activity at pH 4.5 while *Sa*GUS has poor activity at the same pH. The His-tagged construct of *Cp*GUS shows high activity at pH values from 5.5 to 7.4 and only slightly lower activity at pH 5 (Figure 7.3C). Similar to *Sa*GUS, *Cp*GUS has almost no activity at pH 4 and very poor activity at pH 4.5. *Ee*GUS exhibits activity at all pH values tested but has optimal function from pH 4.5 to 6.5 (Figure 7.3D). Significant but lower activity is seen at pH 4, 7, and 7.4. Future characterization of *Ee*GUS will be conducted at pH 6.5. *Fp*GUS exhibits optimal activity from pH 4.5 to 5.5 (Figure 7.3E). A moderate decrease in activity is seen at pH 4, 6, 6.5, and 7. At pH 7.4, activity is greatly attenuated as seen in initial kinetic activity assays. Future characterization of *Fp*GUS activity will be conducted at pH 5. *Bu*GUS shows optimal function at pH values 4.5 to 7 (Figure 7.3F). Activity is still strong but slightly less robust at pH 7.4 while activity is very poor at pH 4. A pH of 6.5 will be used for future *Bu*GUS assays. These results highlight the differences between each of these GUS

enzymes and suggest that they may function in varying environments within the intestine. We predict that removal of the signal peptide and optimization of the assay pH will allow us to quantify activity for the *Pm*GUS and *Bo*GUS in addition to providing important information about how each of these enzymes functions.

Lactobacillus rhamnosus GUS

While characterization of GUS enzymes has focused primarily on those identified in the HMP sequencing analysis described in Chapter 7, the GUS enzyme from *Lactobacillus rhamnosus* (*Lr*GUS) has also been identified as an enzyme of interest. This enzyme was identified as an interesting target after *L. rhamnosus* was isolated from colon polyp biopsies and cultured by the Keku lab at UNC-CH. After observing very poor activity against PNPG and no activity against SN-38G (data not shown), a pH screen was conducted to find the optimal pH for activity assays. As shown in Figure 7.4, *Lr*GUS is indeed inactive at pH 7.4 but shows increasing activity as the assay conditions become more acidic. The optimal activity of *Lr*GUS is seen from pH 4.5 to 5.5 in this assay. This is congruent with a previous analysis of function of partially purified LrGUS that found optimal activity at pH 4.5 with similar activity at pH 3.5 and 5 (17). Further characterization of *Lr*GUS will be conducted at pH 4.5. Characterization of *Lr*GUS extends this catalog of GUS enzymes from enzymes identified from healthy individuals to enzymes from a dysbiotic environment.

Discussion

The role of human GI microbiota-associated GUS enzymes in the drug toxicity of irinotecan and other drugs is well established. However, only a small number of these enzymes have been characterized. The data presented here shows the wide variety of GUS sequences identified from the HMP in Chapter 6 are functional GUS enzymes that are able to process the glucuronidated substrate PNPG with varying affinities (Table 1). While this initial activity validates the rubric for identifying GUS sequences, further characterization shows that we still have much to learn about the activity of these enzymes. Figure 7.4 shows that despite many similarities in sequence, especially in the active site, these enzymes have optimal function at varying pH ranges. L1 enzymes *Ec*GUS, *Sa*GUS, *Cp*GUS, *Ee*GUS, and *Fp*GUS are expected to be cytosolic. Alternatively, *Bu*GUS and the enzymes characterized from the mL1, L2, mL2, mL1mL2, and nL classes have a N-terminal signal sequence that we predict sends these enzymes to the periplasmic space of gram-negative bacteria. The cytosol is expected to be buffered; however, the pH of the surrounding environment has been shown to have a large effect on the internal pH of bacteria and some bacteria such as *Lactobacillus* are known to maintain a slightly more acidic cytosol. In contrast, the pH of the periplasmic space is likely highly dependent on the pH of the surrounding environment. Little work has been done to date to characterize the distribution of various bacterial species along the intestine, but we hypothesize that the optimal pH range of the function of a bacteria's GUS enzyme will match the pH of the locations in which that bacteria is found within the intestine. For instance, enzymes such as *Ec*GUS and *Sa*GUS have

optimal function at a wide range of pH values and would therefore function regardless of where their host bacterium was in the intestine. Alternatively, *Fp*GUS has a limited range of pH values at which it shows optimal function. We know the host bacteria *F. prausnitzii* can make its way to the end of the GI as this sequence was identified in stool samples, but we hypothesize that the bacteria are dominate higher in the GI where the pH is more acidic. Further characterization of the optimal function of these enzymes and their host bacteria will allow us to further refine this hypothesis and understand the diverse ecosystems of the human GI.

Other members of the Redinbo lab are focused on optimizing the previously discovered GUS inhibitors in order to eliminate GUS-mediated irinotecan drug toxicity. Therefore, it is important for us to understand which GUS enzymes are responsible for the reactivation of SN-38 in the GI. Figure 7.2 and 7.3 present the first in-depth analysis of this reaction with purified GUS enzymes. It had been previously established that purified *Ec*GUS is able to efficiently process SN-38G, but this data shows that the reaction proceeds much more quickly than the 1-hour time point previously analyzed (12). Because of the similarity between the L1 enzymes, it was not surprising that *Sa*GUS and *Ee*GUS are also able to process SN-38G well. Previous characterization of the active site loop L1 had established its role in the efficacy of the GUS inhibitors, but it was still unclear if L1 or any variation of this active site loop played a role in substrate specificity. We hypothesized that L1, and perhaps L2, would act as a gate over the active site in order to hold in small substrates such as PNPG and SN-38G. Therefore, the activity of *Bf*GUS seen in the HPLC analysis and refined in the fluorescence assay was a surprise that does not

support this hypothesis. The relatively high activity of both *Bf*GUS and *Bd*GUS against SN-38G have lead us to explore other features in the active site that may play an important role in substrate specificity.

Finally, initial characterization of the *Lr*GUS enzyme is presented. All sequences identified in the HMP were from the stool samples of healthy donors, which represents a limited snapshot of the human microbiota. We identified the *Lr*GUS enzyme from cultured bacteria from a colon polyp biopsy. While stool samples contain primarily mobile bacteria that reside in the lumen of the intestine, bacteria cultured from this biopsy are more likely to be adherent to the mucus layer of the intestine. In addition, there is evidence that dysbiosis of the microbiota can lead to colon cancer, so bacteria cultured from potentially cancerous polyps may play a role in carcinogenesis. As we expand the current understanding of GUS enzymes, we would also like to expand the sources from which characterized sequences come and this is the first effort in this direction.

This chapter presents the first in-depth analysis of the function of a broad range of GUS enzymes. While bacterial GUS enzymes including *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Bf*GUS have been characterized previously, this focus on GUS sequences derived from human stool or colon biopsy samples allows closer characterization of the chemistry catalyzed by these enzymes in the human GI. This initial analysis has shown that while previously characterized enzymes are highly active against glucuronidated substrates including PNPG and SN-38G, there is a wide range of GUS enzymes that are also active against those substrates and likely play a significant role in the human GI.

Enzyme	Loop Class	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
*Ec*GUS	L1	120 ± 12	0.13 ± 0.01	920 ± 160
*Sa*GUS	L1	80 ± 2	0.36 ± 0.03	222 ± 24
MBP-*Cp*GUS	L1	2.6 ± 0.6	1.1 ± 0.2	2.4 ± 1
*Ee*GUS	L1	41 ± 3	0.49 ± 0.09	83.7 ± 21
*Fp*GUS	L1	unquantifiable		
*Bf*GUS	mL1	18 ± 1	1.9 ± 0.3	9.5 ± 2
*Bu*GUS	L2	4.5 ± 0.07	0.50 ± 0.03	9 ± 0.7
*Pm*GUS	mL2	unquantifiable		
*Bo*GUS	mL1mL2	unquantifiable		
*Bd*GUS	nL	21 ± 2	2.7 ± 0.5	7.7 ± 2

Table 7.1. Catalytic Activity of β-glucuronidase Enzymes at pH 7.4
Data are presented as the average over 3 experiments ± SEM. k_{cat}, catalytic rate; K_M, Michaelis constant; k_{cat}/K_M, catalytic efficiency.

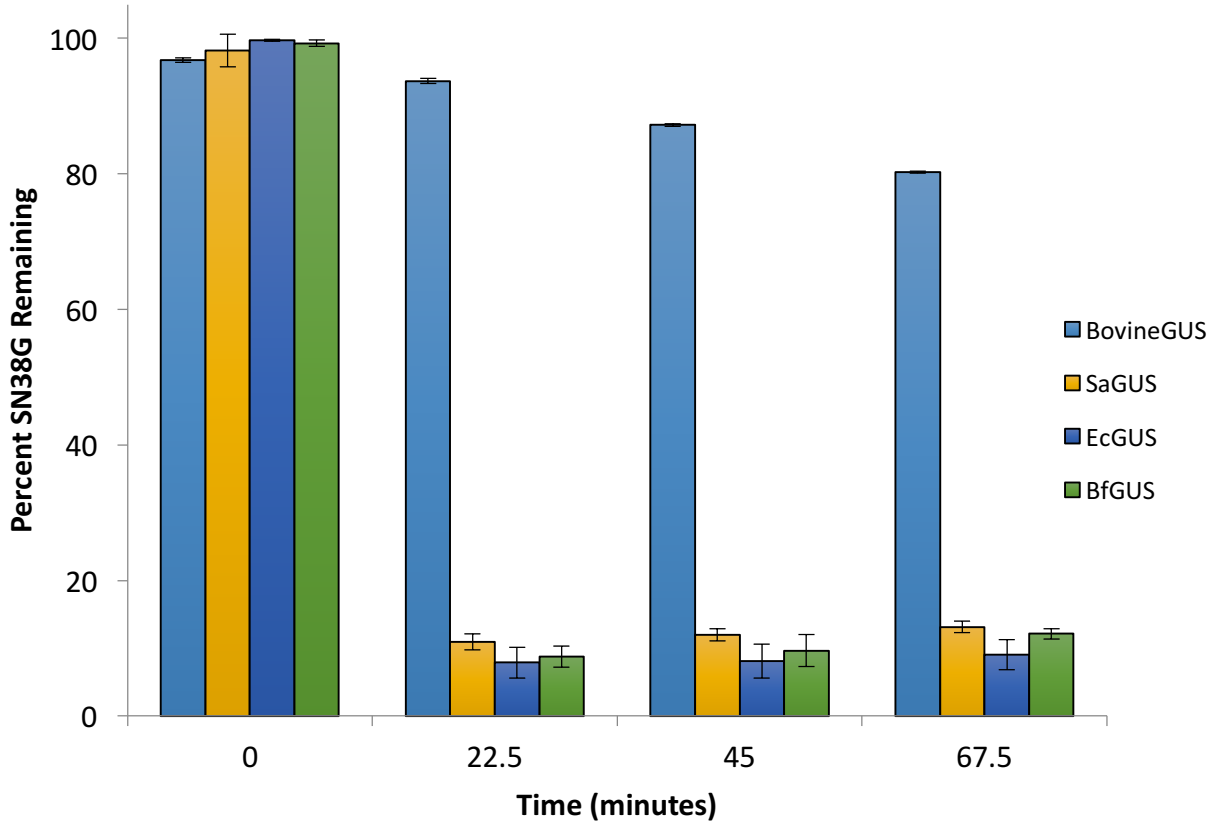


Figure 7.1. Analysis of β-glucuronidase Activity Against SN-38G via HPLC
HPLC was used to quantify the amount of SN-38G remaining after incubation with the β-glucuronidase from *Bos taurus* (BovineGUS), *S. agalactiae* (*Sa*GUS), *C. perfringens* (*Cp*GUS), and *B. fragilis* (*Bf*GUS) over time.

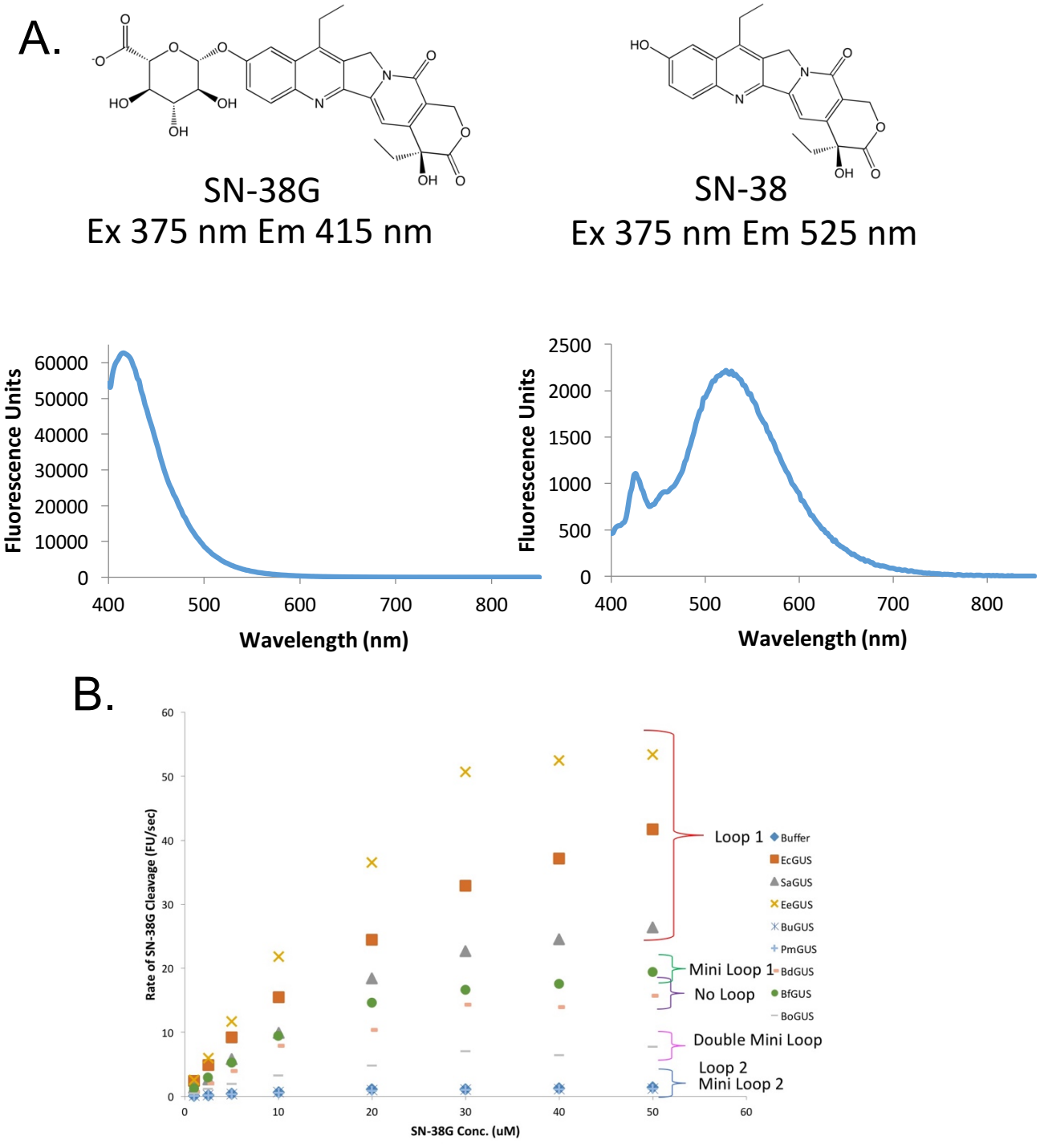


Figure 7.2. β-glucuronidase Activity Against SN-38G via Fluorescence
A. Emission wavelength scan of SN-38G and SN-38 after excitation at 375 nm.
B. Rate of SN-38G cleavage as measured by fluorescence at varying SN-38G concentrations by β-glucuronidases identified from the Human Microbiome Project.

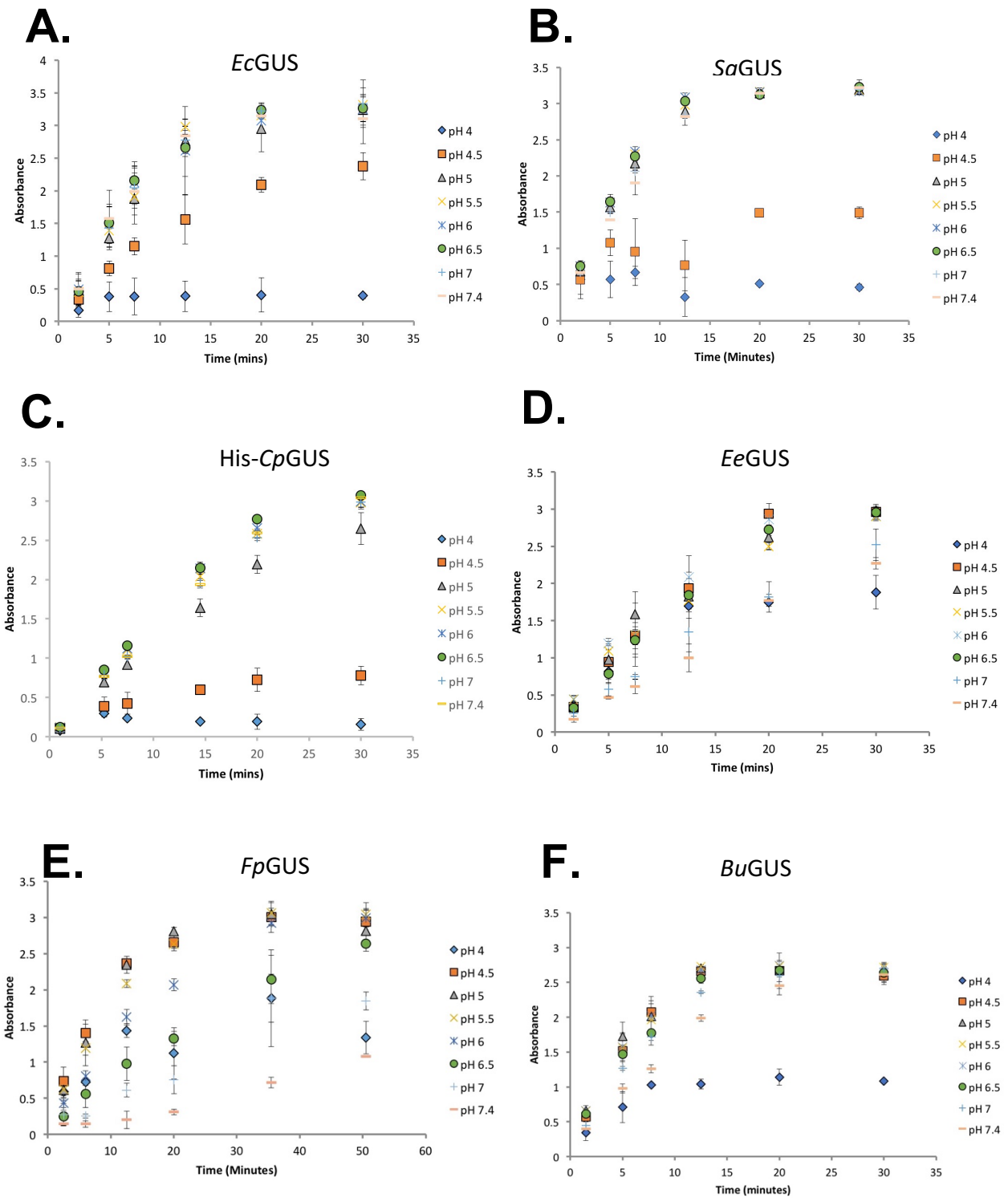


Figure 7.3. Catalytic Activity of β-glucuronidase Enzymes at Varying pH
Activity of *Ec*GUS (A), *Sa*GUS (B), *Cp*GUS (His-tagged construct, C), *Ee*GUS (D), *Fp*GUS (E), and *Bu*GUS (F) against a single PNPG concentration over time at pH values 4, 4.5, 5, 5.5, 6, 6.5, 7, and 7.4.

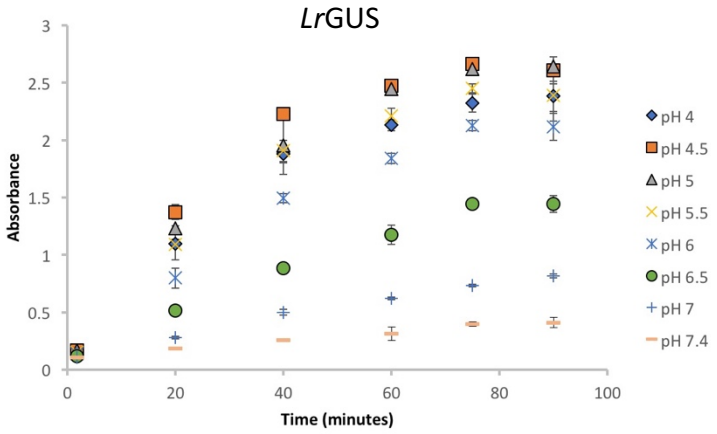


Figure 7.4. Catalytic Activity of *L. rhamnosus* β-glucuronidase at Varying pH
L. rhamnosus GUS (*Lr*GUS) shows varying function against the substrate PNPG at pH values ranging from 4.5 to 7.4.

REFERENCES

1. **Kinross JM**, **Darzi AW**, **Nicholson JK**. 2011. Gut microbiome-host interactions in health and disease. Genome Med **3**:14.

2. **Bäckhed F**, **Fraser CM**, **Ringel Y**, **Sanders ME**, **Sartor RB**, **Sherman PM**, **Versalovic J**, **Young V**, **Finlay BB**. 2012. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. Cell Host Microbe **12**:611–622.

3. **Wang Z**, **Klipfell E**, **Bennett BJ**, **Koeth R**, **Levison BS**, **Dugar B**, **Feldstein AE**, **Britt EB**, **Fu X**, **Chung Y-M**, **Wu Y**, **Schauer P**, **Smith JD**, **Allayee H**, **Tang WHW**, **DiDonato JA**, **Lusis AJ**, **Hazen SL**. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature **472**:57–63.

4. **Kim D-H**. 2015. Gut Microbiota-Mediated Drug-Antibiotic Interactions. Drug Metab Dispos **43**:1581–1589.

5. **Chabot GG**. 1996. Clinical Pharmacology and Pharmacodynamics of Irinotecan. Annals of the New York Academy of Sciences **803**:164–172.

6. **Ma MK**, **McLeod HL**. 2003. Lessons learned from the irinotecan metabolic pathway. Curr Med Chem **10**:41–49.

7. **Mathijssen RH**, **van Alphen RJ**, **Verweij J**, **Loos WJ**, **Nooter K**, **Stoter G**, **Sparreboom A**. 2001. Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). Clin Cancer Res **7**:2182–2194.

8. **Gupta E**, **Lestingi TM**, **Mick R**, **Ramirez J**, **Vokes EE**, **Ratain MJ**. 1994. Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. Cancer Res **54**:3723–3725.

9. **Pérez-Roth E**, **Kwong SM**, **Alcoba-Florez J**, **Firth N**, **Méndez-Alvarez S**. 2010. Complete nucleotide sequence and comparative analysis of pPR9, a 41.7-kilobase conjugative staphylococcal multiresistance plasmid conferring high-level mupirocin resistance. Antimicrobial Agents and Chemotherapy **54**:2252–2257.

10. **Dranitsaris G**, **Maroun J**, **Shah A**. 2005. Estimating the cost of illness in colorectal cancer patients who were hospitalized for severe chemotherapy-induced diarrhea. Can J Gastroenterol **19**:83–87.

11. **Stein A**, **Voigt W**, **Jordan K**. 2010. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol **2**:51–63.

12. **Wallace BD**, **Wang H**, **Lane KT**, **Scott JE**, **Orans J**, **Koo JS**, **Venkatesh M**, **Jobin C**, **Yeh L-A**, **Mani S**, **Redinbo MR**. 2010. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science **330**:831–835.

13. **Roberts AB**, **Wallace BD**, **Venkatesh MK**, **Mani S**, **Redinbo MR**. 2013. Molecular Insights into Microbial β-Glucuronidase Inhibition to Abrogate CPT-11 Toxicity. Molecular Pharmacology **84**:208–217.

14. **Wallace BD**, **Roberts AB**, **Pollet RM**, **Ingle JD**, **Biernat KA**, **Pellock SJ**, **Venkatesh MK**, **Guthrie L**, **O'Neal SK**, **Robinson SJ**, **Dollinger M**, **Figueroa E**, **McShane SR**, **Cohen RD**, **Jin J**, **Frye SV**, **Zamboni WC**, **Pepe-Ranney C**, **Mani S**, **Kelly L**, **Redinbo MR**. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem Biol **22**:1238–1249.

15. **Russell WM**, **Klaenhammer TR**. 2001. Identification and cloning of gusA, encoding a new beta-glucuronidase from Lactobacillus gasseri ADH. Appl Environ Microbiol **67**:1253–1261.

16. **Callanan MJ**, **Russell WM**, **Klaenhammer TR**. 2007. Modification of Lactobacillus β-glucuronidase activity by random mutagenesis. Gene **389**:122–127.

17. **Pham PL**, **Dupont I**, **Roy D**, **Lapointe G**, **Cerning J**. 2000. Production of exopolysaccharide by Lactobacillus rhamnosus R and analysis of its enzymatic degradation during prolonged fermentation. Appl Environ Microbiol **66**:2302–2310.

CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS FOR CHARACTERIZATION OF β-GLUCURONIDASE AND GI DRUG REACTIVATION

Conclusions

β-glucuronidase (GUS) enzymes encoded by the human GI microbiota are responsible for the severe drug toxicity associated with the chemotherapy drug irinotecan (CPT-11). Initial work on this process by the Redinbo lab greatly expanded understanding of the *E. coli*-type GUS proteins and characterized inhibitors that are effective against mice models of irinotecan-induced drug toxicity. The work presented here compiles a catalog of GUS sequences mined from the Human Microbiome Project (HMP) sequencing repository. This catalog gives a new appreciation of the breadth of GUS sequences found within the human microbiome and suggests that this enzyme class may play a larger role in the intestinal ecosystem than previously explored. Previous efforts to catalog GUS proteins have depended on gene annotations, which we have shown are often incorrect due to the high sequence similarity, especially in the active site, between various glycoside hydrolases such as β-glucuronidases and β-galactosidases (1). This approach using one-to-one sequence alignments to identify the presence or absence of key conserved residues in all identified genes in the HMP sequences repository is a

much more robust approach that can be used to catalog any well characterized gene of interest.

This large catalog of GUS proteins is broken into smaller classes of sequences based on the presence or absence and size of two similar active site loops. Representative enzymes from each loop class were chosen and characterized as presented in Chapter 7. The seven new GUS enzymes presented here more than double the size of the catalog of microbial GUS proteins and allows us to explore the diverse function of this enzyme family. While this work has given us unique insight into GUS enzymes, it has also opened many new avenues for further research, some of which are explored below.

Future Directions

Further Characterization of the HMP GUS Sequence Catalogues

As described in Chapter 6, the analysis of the HMP sequencing repository provided two datasets of GUS sequences. The first dataset from the HMGC-Clustered Gene Indices database is a non-redundant catalog of GUS genes. While initial analysis of this dataset has shown an interesting distribution of sequence lengths (Figure 6.2 of Chapter 6), there is still much analysis to be done with these sequences. The first goal is to match each sequence in this dataset to one of the reference genomes of bacteria cultured as part of the HMP or deposited in the NCBI Reference Sequence Database. This will allow us to associate each GUS sequence with a bacteria or closely related set of bacteria that the sequence was likely isolated from in the HMP fecal samples and compile a list of GUS-containing bacteria. This

smaller, non-redundant set also allows for better understanding of which enzymes contain features such as the extended C-terminal domain and signal peptide and if these features cluster with the active site loops.

The second dataset from the HMGI- Gene Indices is a redundant set of GUS sequences that can be grouped by which of the 139 individuals they were obtained from. Thus far we have analyzed the total number of GUS sequences each individual possessed and which loop class those sequences fell into. However, we have not taken into account the abundance of a sequence in each individual. It is likely that some bacteria and therefore their associated GUS are more abundant in the gut and therefore contribute more significantly to the intestinal environment. These abundant GUS sequences may be similar across individuals or may vary widely.

Structure and Function of New GUS Enzymes

While we have identified the active site loops as an interesting variable region of this enzyme, continued functional and structural characterization of these enzymes will likely lead to several additional regions of interest. A quick glance at the sequence alignment of these enzymes presented in Appendix 3 already reveals small insertions and deletions; additional characterization may match those sequence changes to interesting functional changes. It is also likely that we will be able to link the variability in optimal pH conditions shown in Chapter 7 to changes in just a few amino acids within the active site of these enzymes.

SN-38G Processing

Chapter 7 includes the first evidence that GUS enzymes in addition to *Ec*GUS can process SN-38G, including enzymes that cannot be inhibited by the well-characterized inhibitors produced by the Redinbo lab. Despite not inhibiting GUS enzymes that process SN-38G, these inhibitors are effective in reducing the irinotecan-induced diarrhea in mice. This suggests we still have much to learn about GUS activity *in vivo* and which enzymes encounter SN-38G. Continued characterization of the bacteria that encode GUS sequences from the HMP analysis, will lead to better understanding of the regulatory mechanisms that may impact GUS expression and transport of glucuronidated compounds. The Redinbo lab is currently characterizing the GUS gene repressor (GusR) found in *E. coli* and a small number of related bacteria that allows for increased GUS expression when glucuronidated compounds bind GusR. There is also a known glucuronide transporter in the same operon in *E. coli,* but this transporter as well as the transporters that are likely present in other bacteria have not been characterized. It is possible that these transporters show substrate selectivity such that many of these GUS enzymes rarely come in contact with SN-38G. In addition to these molecular-level details, members of the Redinbo lab are also conducting mice experiments to measure the effect of irinotecan and inhibitor treatment on the microbiota. A better understanding of the complex environment in which GUS enzymes interact with SN-38G will allow for further optimization of previously characterized inhibitors and improve the efficacy against drug toxicity.

A related area of exploration is characterization of why some patients experience severe diarrhea after irinotecan treatment while other patients have minimal side effects. We hypothesize that patients that experience severe diarrhea possess significantly more of the bacteria and GUS enzymes that are able to process SN-38G. We are interested in comparing microbiome samples from patients treated with irinotecan but with varying degrees of drug toxicity severity in order to correlate variation in the microbiota with severity of drug toxicity. If high-processing SN-38G bacteria are identified through molecular approaches, mouse studies, or patient data, it may be possible to design a screening tool to determine which patients are likely to react poorly to irinotecan treatment and would be ideal candidates for GUS inhibitor treatment.

Other Glucuronidated Compounds

Both drugs and endogenous compounds are glucuronidated by UGT enzymes in liver. Although this work focuses on SN-38G, many other glucuronidated compounds are also processed by GUS enzymes in the GI. In fact, the Redinbo lab and collaborators have shown that GUS inhibitors are effective in reducing the size and number of small intestinal ulcers caused by non-steroidal anti-inflammatory drugs (NSAIDs) (2, 3). Many NSAIDs including ketoprofen, indomethacin, and diclofenac are inactivated and eliminated through glucuronidation as described for SN-38G in Chapter 5 (4). Two thirds of patients show drug-induced small intestine lesions after both chronic (>3 months) and short-term (<1 week) NSAID use (4). While the mechanism of small intestinal damage is unclear, the efficacy of GUS inhibitors confirms that the damage is GUS mediated (2, 3). Because of the success

of these inhibitors in these mouse studies, we are interested in characterizing NSAID-glucuronide processing with purified GUS enzymes. While characterization can and has been conducted using HPLC, that is a time- and effort-intensive approach that could be replaced by an absorbance or fluorescence based assay that could be conducted quickly in a plate reader. While some compounds can be analyzed directly as described for SN-38G in Chapter 7, not all compounds will show varying absorbance for fluorescence profiles with and without the glucuronide sugar. Therefore, an additional requirement for this assay is that it will function with a variety of glucuronidated compounds.

In order to satisfy these two requirements, we have designed a coupled assay during which the free glucuronide sugar generated by the GUS enzyme is utilized in a second reaction that generates a convenient read-out. Uronate dehydrogenase (Udh) is an enzyme that processes glucuronic acid using NAD^+ as a cofactor that can be used as a read-out. The basic assay scheme is shown in Figure 8.1. Because Udh uses NAD^+ as a cofactor and produces one molecule of NADH for each molecule of glucuronic acid, an observed increase in NADH signal via absorption at 340 nm correlates to the increase in glucuronic acid produced by the GUS enzyme included in the assay. As long as the catalytic activity and concentration of Udh in the assay is sufficiently high, the glucuronic acid produced by GUS enzyme(s) should be the limiting reagent. Therefore, the Udh kinetics observed via NADH production should mimic the kinetics of glucuronic acid production by the GUS enzyme. Initial work in the Redinbo lab has shown this assay scheme to be an effective way to measure glucuronic acid production using both

purified *Agrobacterium tumefaciens* Udh and the D-Glucuronic/D-Galacturonic Acid Assay Kit from Megazyme. This assay will allow us to explore a much wider variety of glucuronidated compounds ranging from drugs such as NSAIDs to endogenous compounds such as steroid hormones. Further understanding of GUS function will allow us to identify the native substrates of this enzyme family as well as explore the ways in which bacteria have adapted to the environment of the human gut. We also hope that we will be able to optimize the initial GUS inhibitors to be effective against a wide range of GUS mediated intestinal damage.

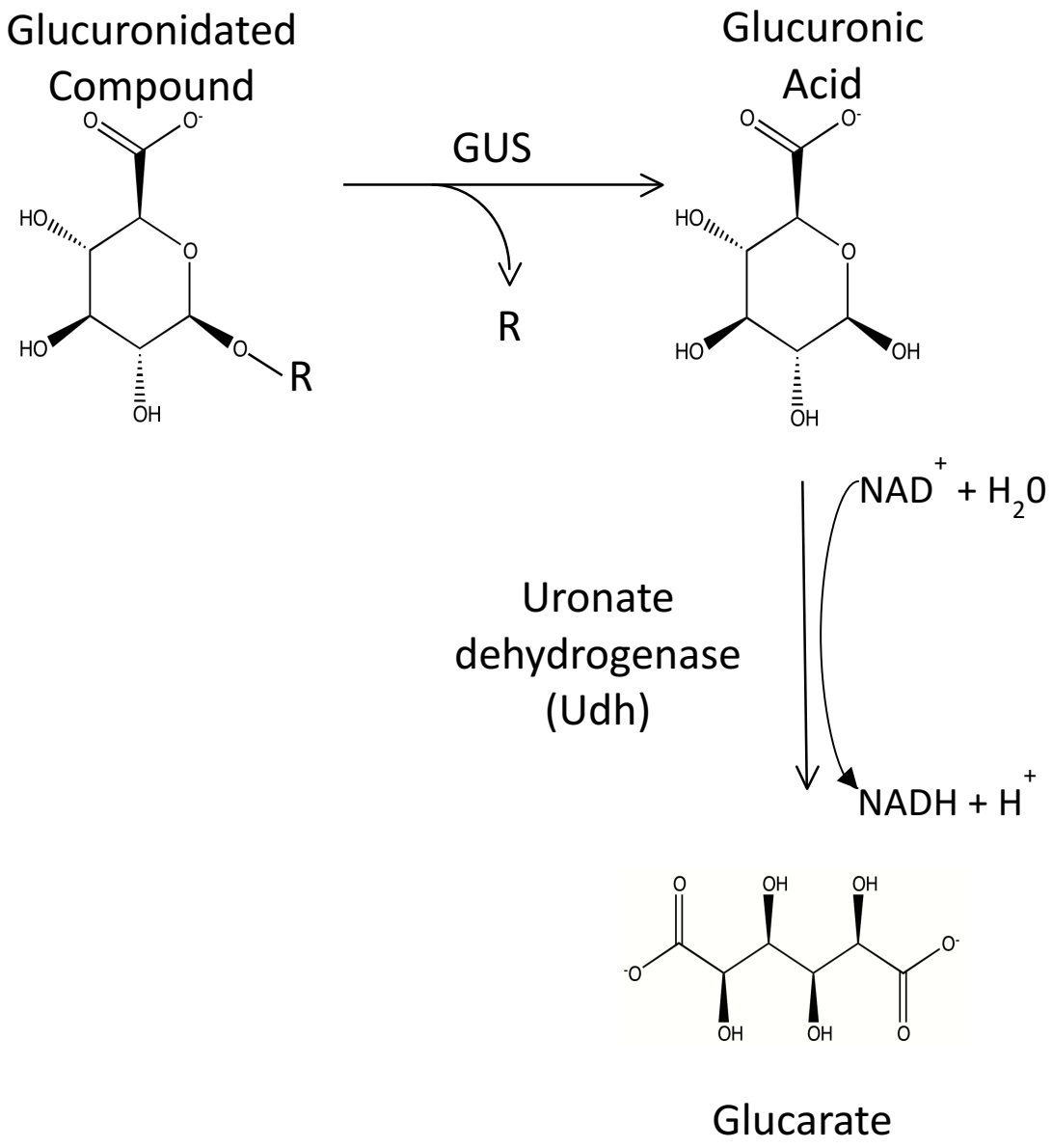


Figure 8.1 Uronate dehydrogenase Coupled Assay Scheme

REFERENCES

1. **Wallace BD**, **Roberts AB**, **Pollet RM**, **Ingle JD**, **Biernat KA**, **Pellock SJ**, **Venkatesh MK**, **Guthrie L**, **O'Neal SK**, **Robinson SJ**, **Dollinger M**, **Figueroa E**, **McShane SR**, **Cohen RD**, **Jin J**, **Frye SV**, **Zamboni WC**, **Pepe-Ranney C**, **Mani S**, **Kelly L**, **Redinbo MR**. 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem Biol **22**:1238–1249.

2. **LoGuidice A**, **Wallace BD**, **Bendel L**, **Redinbo MR**, **Boelsterli UA**. 2012. Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. J Pharmacol Exp Ther **341**:447–454.

3. **Saitta KS**, **Zhang C**, **Lee KK**, **Fujimoto K**, **Redinbo MR**, **Boelsterli UA**. 2014. Bacterial β-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: mode of action and pharmacokinetics. Xenobiotica **44**:28–35.

4. **Boelsterli UA**, **Redinbo MR**, **Saitta KS**. 2013. Multiple NSAID-induced hits injure the small intestine: underlying mechanisms and novel strategies. Toxicol Sci **131**:654–667.

APPENDIX 1: pWBG749 OR pSK41 ORIGIN-OF-TRANSFER MIMC SEQUENCE CONTENT

pWBG749 or pSK41 origin-of-transfer mimc sequence content of *Staphylococcal* plasmids as determined by BLASTN. Adapted from O'Brien et al 2015 (Nucleic Acids Research 43:7971-7983).

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pSA1308	NC_007928.1	TRUE	TRUE
aureus	pWBG1773	NC_010616.1	TRUE	FALSE
aureus	pKH3	NC_005020.1	TRUE	FALSE
aureus	pLGA251	NC_017348.1	TRUE	FALSE
aureus	pKH12	NC_010687.1	TRUE	FALSE
aureus	pTW20_2	FN433598.1	TRUE	FALSE
aureus	SAP070A	GQ900423.1	TRUE	FALSE
aureus	SAP104A	GQ900450.2	TRUE	FALSE
aureus	pKH14	NC_010428.1	TRUE	FALSE
aureus	p18807-P01	CP002136.1	TRUE	FALSE
aureus	p18808-P01	CP002138.1	TRUE	FALSE
aureus	p18811-P01	CP002144.1	TRUE	FALSE
aureus	pUSA01	NC_007790.1	TRUE	FALSE
aureus	SAP046B	GQ900404.1	TRUE	FALSE
aureus	SAP049B	GQ900408.1	TRUE	FALSE
aureus	SAP051B	GQ900411.1	TRUE	FALSE
aureus	pKH18	NC_010231.1	TRUE	TRUE
aureus	pC55s	AY048756.1	TRUE	TRUE
aureus	pNewBould305	NZ_AKYW01000028.1	TRUE	TRUE
aureus	pWBG760	GQ900473.1	TRUE	FALSE
aureus	USA300_TCH959	NZ_AASB02000125.1	TRUE	TRUE
aureus	ATCC 51811	NZ_ADVP01000034.1	TRUE	TRUE
aureus	SAP105B	NC_013378.1	TRUE	FALSE
aureus	pBORa53	NC_013550.1	TRUE	TRUE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	A8796	NZ_ADJJ01000025.1	TRUE	TRUE
aureus	TCH70_2	NZ_ACHH02000018.1	TRUE	FALSE
aureus	pSA268	NC_023278.1	TRUE	TRUE
aureus	pUSA300HOUMS	NC_010066.1	TRUE	FALSE
aureus	55/2053	NC_022126.1	TRUE	FALSE
aureus	SAP055A	GQ900414.1	TRUE	FALSE
aureus	pSAS	BX571858.1	TRUE	TRUE
aureus	pWBG750	GQ900392.1	TRUE	TRUE
aureus	SAP072A	GQ900424.1	TRUE	TRUE
aureus	pMW2	NC_005011.1	TRUE	TRUE
aureus	SAP053A	GQ900413.1	TRUE	TRUE
aureus	p21	NC_002517.1	TRUE	TRUE
aureus	SAP073A	GQ900425.1	TRUE	TRUE
aureus	pSaa6159	CP002115.1	TRUE	TRUE
aureus	pWBG757	GQ900397.1	TRUE	TRUE
aureus	pWBG763	GQ900467.1	TRUE	TRUE
aureus	65-1322	NZ_GG700565.1	TRUE	FALSE
aureus	SAP058A	GQ900479.1	TRUE	FALSE
aureus	pSK62	GQ900446.1	TRUE	FALSE
aureus	pSA1379	NC_007931.1	TRUE	TRUE
aureus	SAP060A	GQ900416.1	TRUE	FALSE
aureus	A8115_con_1.21	NZ_ACKG01000021.1	TRUE	FALSE
aureus	SAP051A	GQ900410.1	TRUE	TRUE
aureus	M876	NZ_GG700623.1	TRUE	FALSE
aureus	E1410	NZ_GG700609.1	TRUE	FALSE
aureus	A8117	NZ_ACYO01000019.1	TRUE	FALSE
aureus	M1015	NZ_GG749015.1	TRUE	FALSE
aureus	M809	NZ_GG749324.1	TRUE	FALSE
aureus	pSK76	GQ900444.1	TRUE	FALSE
aureus	C101_1.21	NZ_GG730140.1	TRUE	FALSE
aureus	WBG10049	NZ_GG730219.1	TRUE	FALSE
aureus	pWBG756	GQ900472.1	TRUE	FALSE
aureus	TCH60_unnamed	NC_017345.1	TRUE	FALSE
aureus	58-424_3	NZ_GG749073.1	TRUE	FALSE
aureus	A9635	NZ_ACKI01000002.1	TRUE	TRUE
aureus	pCA347	NC_021552.1	TRUE	TRUE
aureus	pN315	NC_003140.1	TRUE	TRUE
aureus	pWBG752	GQ900394.1	TRUE	TRUE
aureus	pST75	NC_016942.1	TRUE	FALSE
aureus	SAP049A	GQ900407.1	TRUE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pWBG755	GQ900471.1	TRUE	FALSE
aureus	pPM1	NC_019148.1	TRUE	TRUE
aureus	SAP063A	GQ900418.1	TRUE	FALSE
aureus	SAP059A	GQ900480.1	TRUE	FALSE
aureus	SAP065B	GQ900481.1	TRUE	FALSE
aureus	pWBG761	GQ900474.1	TRUE	FALSE
aureus	SAP071A	GQ900485.1	TRUE	FALSE
aureus	p18807-P03	CP002135.1	TRUE	TRUE
aureus	pUSA300HOUMR	NC_010063.1	TRUE	TRUE
aureus	p18808-P03	CP002137.1	TRUE	TRUE
aureus	p18810-P03	NC_018963.1	TRUE	TRUE
aureus	SAP050A	GQ900409.1	TRUE	TRUE
aureus	p18806-P03	CP002134.1	TRUE	TRUE
aureus	pLAC-P03	CP002149.1	TRUE	TRUE
aureus	SAP015A	GQ900380.1	TRUE	TRUE
aureus	SAP046A	GQ900403.1	TRUE	TRUE
aureus	p18805-03	CP002132.1	TRUE	TRUE
aureus	p18811-03	CP002143.1	TRUE	TRUE
aureus	SAP056A	GQ900478.1	TRUE	TRUE
aureus	M0408_conjugative	AIWO01000029	TRUE	FALSE
aureus	SAP012A	GQ900377.1	TRUE	FALSE
aureus	A9781	NZ_ACKL01000034.1	TRUE	FALSE
aureus	pWBG744	GQ900398.1	TRUE	FALSE
aureus	SAP048A	GQ900406.1	TRUE	FALSE
aureus	SAP074A	GQ900426.1	TRUE	FALSE
aureus	pLUH02	FR714929.1	TRUE	FALSE
aureus	pZ172_1	NC_022610.1	TRUE	FALSE
aureus	p19231-P03	CP002147.1	TRUE	FALSE
aureus	SAP019A	GQ900385.1	TRUE	FALSE
aureus	pSK67	NC_019010.1	TRUE	FALSE
aureus	A8819	NZ_ADJK01000020.1	TRUE	FALSE
aureus	pWBG759	GQ900401.1	TRUE	TRUE
aureus	p18809-P04	CP002146.1	TRUE	TRUE
aureus	SAP047A	NC_013331.1	TRUE	TRUE
aureus	pI258	GQ900378.1	TRUE	FALSE
aureus	SAP104C	GQ900498.1	TRUE	FALSE
aureus	SAP070B	GQ900484.1	TRUE	FALSE
aureus	pSK59	GQ900488.1	TRUE	FALSE
aureus	pTW20_1	FN433597.1	TRUE	FALSE
aureus	SAP027A	GQ900388.1	TRUE	TRUE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pSJH101	NC_009619.1	TRUE	TRUE
aureus	pSJH901	NC_009477.1	TRUE	TRUE
aureus	SAP064A	GQ900419.1	TRUE	FALSE
aureus	pSK57	GQ900493.1	TRUE	FALSE
aureus	pSK80	GQ900492.1	TRUE	FALSE
aureus	SAP066A	GQ900482.1	TRUE	FALSE
aureus	SAP101A	GQ900495.1	TRUE	FALSE
aureus	pI3T3	NC_020565.1	TRUE	FALSE
aureus	SAP052A	GQ900412.1	TRUE	TRUE
aureus	SAP017A	GQ900382.1	TRUE	TRUE
aureus	pI6T6	NC_020567.1	TRUE	FALSE
aureus	pCM05	GQ900387.1	TRUE	FALSE
aureus	pWBG747	GQ900399.1	TRUE	TRUE
aureus	pWBG746	GQ900390.1	TRUE	TRUE
aureus	D139	GG730186.1	TRUE	FALSE
aureus	pSK64	GQ915268.1	TRUE	FALSE
aureus	pEDINA	NC_010077.1	TRUE	TRUE
aureus	SAP076A	GQ900427.1	TRUE	FALSE
aureus	pSK60	GQ915267.1	TRUE	FALSE
aureus	pSK74	GQ915266.1	TRUE	FALSE
aureus	pTZ2162	NC_010419.1	TRUE	TRUE
aureus	SAP078A	GQ900430.1	TRUE	FALSE
aureus	SAP077A	GQ900428.1	TRUE	FALSE
aureus	pSK79	GQ900489.1	TRUE	FALSE
aureus	SAP102A	GQ900496.1	TRUE	FALSE
aureus	pSK21	GQ900490.1	TRUE	FALSE
aureus	SAP075A	GQ900486.1	TRUE	FALSE
aureus	SAP067A	GQ900483.1	TRUE	FALSE
aureus	SAP054A	GQ900477.1	TRUE	FALSE
aureus	pWBG749	GQ900391.1	TRUE	FALSE
aureus	pWBG745	NC_013325.1	TRUE	FALSE
aureus	pETB	NC_003265.1	TRUE	FALSE
aureus	pSK53	GQ915270.1	TRUE	FALSE
aureus	pV030-8	NC_010279.1	TRUE	TRUE
aureus	SAP057A	GQ900415.1	TRUE	FALSE
aureus	pSK23	GQ900491.1	TRUE	FALSE
aureus	SAP082A	GQ900434.1	TRUE	TRUE
aureus	pWBG748	GQ915265.1	TRUE	FALSE
aureus	pSK156	GQ900448.1	TRUE	TRUE
aureus	SAP079A	GQ900432.1	TRUE	TRUE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pGO1	NC_012547.1	TRUE	TRUE
aureus	pWBG762	GQ900475.1	TRUE	TRUE
aureus	pLW043	NC_005054.1	TRUE	TRUE
epidermis	SAP045A	NC_013374.1	TRUE	FALSE
epidermis	SAP106A	NC_013379.2	TRUE	FALSE
epidermis	SAP016A	NC_013372.1	TRUE	FALSE
epidermis	Epidermis_conjugative	AHLC01000011.1	TRUE	FALSE
simulans	pACK5	NC_015176.1	TRUE	FALSE
simulans	pACK2	NC_015173.2	TRUE	FALSE
warneri	pPI-2	NC_005208.1	TRUE	FALSE
arlettae	pSS-03	NC_016054.1	FALSE	FALSE
aureus	pSN2	NC_005565.1	FALSE	FALSE
aureus	SAP093B	NC_013310.1	FALSE	FALSE
aureus	pDLK3	NC_013969.1	FALSE	FALSE
aureus	pAVY	NC_013451.1	FALSE	FALSE
aureus	pSK6	NC_001995.1	FALSE	FALSE
aureus	SAP104B	GQ900451.2	FALSE	FALSE
aureus	pSK3	NC_001994.1	FALSE	FALSE
aureus	pNVH99	AJ296103.1	FALSE	FALSE
aureus	pLUH01	NC_017346.1	FALSE	FALSE
aureus	pWBG754	GQ900396.1	FALSE	FALSE
aureus	TCH70_1	NZ_ACHH02000015.1	FALSE	FALSE
aureus	pS1e	NZ_AUPS01000034.1	FALSE	FALSE
aureus	A6300	NZ_ACKF01000018.1	FALSE	FALSE
aureus	p9b	NC_019143.1	FALSE	FALSE
aureus	SAP087A	GQ900439.1	FALSE	FALSE
aureus	pWBG764	GQ900468.1	FALSE	FALSE
aureus	A9765	NZ_ACSN01000068.1	FALSE	FALSE
aureus	A9719	NZ_ACKJ01000014.1	FALSE	FALSE
aureus	pDLK1	NC_019139.1	FALSE	FALSE
aureus	pKH20	NC_010686.1	FALSE	FALSE
aureus	SAP078B	GQ900431.1	FALSE	FALSE
aureus	pKH8	U50077.1	FALSE	FALSE
aureus	Cn1	NC_022228.1	FALSE	FALSE
aureus	CF-Marseille	NZ_CABA01000093.1	FALSE	FALSE
aureus	p19321-P01	NC_018969.1	FALSE	FALSE
aureus	pE5	M17990.1	FALSE	FALSE
aureus	pKH19	NC_010685.1	FALSE	FALSE
aureus	pWBG738	NC_007209.1	FALSE	FALSE
aureus	pWBG751	GQ900393.1	FALSE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pT48	NC_001395.1	FALSE	FALSE
aureus	TCH130	NZ_ACHD01000266.1	FALSE	FALSE
aureus	TCH959	NZ_AASB02000192.1	FALSE	FALSE
aureus	pKH4	U81980.1	FALSE	FALSE
aureus	pKH21	NC_010684.1	FALSE	TRUE
aureus	A5948_3	NZ_ACKD01000027.1	FALSE	FALSE
aureus	pS1d	NZ_AUPS01000033.1	FALSE	TRUE
aureus	pNVH01	NC_004562.1	FALSE	TRUE
aureus	pUR5425	NC_019146.1	FALSE	TRUE
aureus	pBMSa1	AY541446.1	FALSE	TRUE
aureus	pKH15	NC_010427.1	FALSE	FALSE
aureus	pBMb9393	NC_021657.1	FALSE	FALSE
aureus	pDLK2	GU562625.1	FALSE	FALSE
aureus	SAP065A	GQ900420.1	FALSE	FALSE
aureus	pKH13	NC_010426.1	FALSE	FALSE
aureus	pC194	NC_002013.1	FALSE	FALSE
aureus	pS0385-3	AM990995.1	FALSE	TRUE
aureus	pE194	NC_005908.1	FALSE	FALSE
aureus	pSBK203	U35036.1	FALSE	FALSE
aureus	pNS1	M16217.1	FALSE	FALSE
aureus	pS1c	NZ_AUPS01000031.1	FALSE	FALSE
aureus	SAP085B	GQ900438.1	FALSE	FALSE
aureus	pKH7	NC_002096.1	FALSE	FALSE
aureus	C101_1.1	NZ_GG730120.1	FALSE	FALSE
aureus	M899	NZ_GG730190.1	FALSE	FALSE
aureus	pS0385-2	AM990994.1	FALSE	FALSE
aureus	pS123b	NZ_AUPU01000024.1	FALSE	FALSE
aureus	pS194	NC_005564.1	FALSE	FALSE
aureus	pKH6	U38428.1	FALSE	FALSE
aureus	pUSA02	NC_007791.1	FALSE	FALSE
aureus	SAP085A	GQ900437.1	FALSE	FALSE
aureus	SAP093A	GQ900441.1	FALSE	FALSE
aureus	SAP094A	GQ900443.1	FALSE	FALSE
aureus	SAP095A	GQ900445.1	FALSE	FALSE
aureus	pT181	NC_006629.2	FALSE	FALSE
aureus	pKH17	NC_010284.1	FALSE	FALSE
aureus	pKH16	NC_010262.1	FALSE	FALSE
aureus	ATCC BAA-39	NZ_AEEK01000040.1	FALSE	FALSE
aureus	A8115_con_1.2	NZ_ACKG01000002.1	FALSE	FALSE
aureus	SAP060B	GQ900417.1	FALSE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	MR1_139	NZ_ACZQ01000139.1	FALSE	FALSE
aureus	JKD6009	NZ_ABSA01000025.1	FALSE	FALSE
aureus	pUB110	NC_001384.1	FALSE	FALSE
aureus	A10102	NZ_ACSO01000036.1	FALSE	FALSE
aureus	A6300_2	NZ_ACKF01000048.1	FALSE	FALSE
aureus	A6244	NZ_ACKE01000028.1	FALSE	FALSE
aureus	A9763	NZ_ACKK01000035.1	FALSE	FALSE
aureus	pC221	NC_006977.1	FALSE	FALSE
aureus	pTZ4	NC_010111.1	FALSE	FALSE
aureus	A5937	NZ_ACKC01000002.1	FALSE	FALSE
aureus	CF-Marseille_2	NZ_CABA01000045.1	FALSE	FALSE
aureus	SAP084A	GQ900436.1	FALSE	FALSE
aureus	SAP089A	NC_013308.1	FALSE	FALSE
aureus	MR1_138	NZ_ACZQ01000138.1	FALSE	FALSE
aureus	pC223	AY355285.1	FALSE	FALSE
aureus	pKKS49	NC_019149.1	FALSE	FALSE
aureus	pS1b	NZ_AUPS01000028.1	FALSE	FALSE
aureus	A9754	NZ_ADJI01000035.1	FALSE	FALSE
aureus	58-424_1	NZ_GG749054.1	FALSE	FALSE
aureus	pS0385-1	AM990993.1	FALSE	FALSE
aureus	pCPS49	NC_019142.1	FALSE	FALSE
aureus	pS1a	NZ_AUPS01000027.1	FALSE	FALSE
aureus	pS94a	NZ_AUPW01000021.1	FALSE	FALSE
aureus	pVGA	NC_011605.1	FALSE	FALSE
aureus	pCPS32	NC_019141.1	FALSE	FALSE
aureus	pJ3358	NC_001763.1	FALSE	FALSE
aureus	pUR3912	NC_020183.2	FALSE	FALSE
aureus	pKKS627	NC_014156.1	FALSE	FALSE
aureus	pSA8589	NC_021230.1	FALSE	FALSE
aureus	pMSA16	NC_019144.1	FALSE	FALSE
aureus	pUR4128	NC_019147.1	FALSE	FALSE
aureus	pUR2355	NC_019145.1	FALSE	FALSE
aureus	pRJ6	NC_011522.1	FALSE	FALSE
aureus	pS130a	NZ_AUPT01000023.1	FALSE	FALSE
aureus	pSK17	GQ900513.1	FALSE	FALSE
aureus	pRJ9	AF447813.1	FALSE	FALSE
aureus	SAP077B	NC_013341.1	FALSE	FALSE
aureus	pKKS832	FN377602.2	FALSE	FALSE
aureus	pS123a	NZ_AUPU01000021.1	FALSE	FALSE
aureus	SAP099B	GQ900449.1	FALSE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
aureus	pAVX	NC_013453.1	FALSE	FALSE
aureus	pCH91	NC_020227.1	FALSE	FALSE
aureus	pUB101	NC_005127.1	FALSE	FALSE
aureus	SAP015B	GQ900500.1	FALSE	FALSE
aureus	p11819-97	NC_017350.1	FALSE	FALSE
aureus	VRSAp	NC_002774.1	FALSE	FALSE
aureus	SAP105A	NC_013377.1	FALSE	TRUE
aureus	68-397	NZ_GG700590.1	FALSE	FALSE
aureus	pSK77	GQ900494.1	FALSE	FALSE
aureus	pSK1	NC_014369.1	FALSE	FALSE
aureus	SAP103A	GQ900497.1	FALSE	FALSE
aureus	pWBG753	GQ900395.1	FALSE	FALSE
aureus	p18813-P03	NC_018967.1	FALSE	TRUE
aureus	pUSA03	NC_007792.1	FALSE	TRUE
aureus	pI5S5	NC_020535.1	FALSE	TRUE
aureus	pWBG758	NC_013329.1	FALSE	FALSE
aureus	pSA737	NC_021076.1	FALSE	FALSE
aureus	pPR9	NC_013653.1	FALSE	TRUE
aureus	SAP069A	GQ900422.1	FALSE	TRUE
aureus	SAP080A	GQ900433.1	FALSE	TRUE
aureus	pSK41	AF051917.1	FALSE	TRUE
aureus	SAP014A	GQ900379.1	FALSE	TRUE
aureus	SAP068A	GQ900421.1	FALSE	TRUE
aureus	pETB	NC_022598.1	FALSE	TRUE
chromogenes	pLNU8	NC_008352.1	FALSE	FALSE
chromogenes	pLNU4	NC_007771.1	FALSE	TRUE
chromogenes	pLNU9	NC_008354.1	FALSE	FALSE
epidermis	pBE131	NC_001390.1	FALSE	FALSE
epidermis	pSepCH	NC_003969.1	FALSE	FALSE
epidermis	pSK108	NC_013395.1	FALSE	FALSE
epidermis	SAP108D	NC_013392.2	FALSE	FALSE
epidermis	pLNU6	NC_008356.1	FALSE	TRUE
epidermis	pSE-12228-01	NC_005008.1	FALSE	FALSE
epidermis	SAP108C	NC_013391.1	FALSE	FALSE
epidermis	pSK103	NC_013394.1	FALSE	FALSE
epidermis	pSE-12228-02	NC_005007.1	FALSE	FALSE
epidermis	SAP108B	NC_013390.1	FALSE	FALSE
epidermis	pSE-12228-06	NC_005003.1	FALSE	FALSE
epidermis	SAP110B	NC_013384.1	FALSE	FALSE
epidermis	pUR3937	NC_019304.1	FALSE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
epidermis	pUR3036	NC_019303.1	FALSE	FALSE
epidermis	pSE-12228-03	NC_005006.1	FALSE	FALSE
epidermis	pSK639	NC_005566.1	FALSE	FALSE
epidermis	SAP106B	NC_013380.1	FALSE	FALSE
epidermis	SAP107B	NC_013382.1	FALSE	FALSE
epidermis	pSE-12228-04	NC_005005.1	FALSE	FALSE
epidermis	pSK105	NC_013393.1	FALSE	FALSE
epidermis	pSE-12228-05	NC_005004.1	FALSE	FALSE
epidermis	SAP108A	NC_013389.1	FALSE	TRUE
epidermis	SAP110A	NC_013383.1	FALSE	FALSE
epidermis	pSERP	NC_006663.1	FALSE	FALSE
epidermis	pSWS47	NC_022618.1	FALSE	FALSE
epidermis	SAP107A	NC_013381.1	FALSE	FALSE
equorum	pSEQU3	NZ_AVBD01000026.1	FALSE	FALSE
equorum	pSEQU2	NZ_AVBD01000024.1	FALSE	FALSE
equorum	pSEQU1	NZ_AVBD01000023.1	FALSE	FALSE
haemolyticus	pSHaeA	NC_007169.1	FALSE	FALSE
haemolyticus	pSHaeB	NC_007170.1	FALSE	FALSE
haemolyticus	pLNU3	NC_007770.1	FALSE	TRUE
haemolyticus	pLNU7	NC_008353.1	FALSE	TRUE
haemolyticus	pSHaeC	NC_007171.1	FALSE	FALSE
hyicus	hyicus_unnamed	NC_016137.1	FALSE	FALSE
hyicus	p7313178-1	NC_016140.1	FALSE	FALSE
hyicus	p9811071-1	NC_016139.1	FALSE	FALSE
hyicus	pKKS966	NC_015171.1	FALSE	FALSE
hyicus	pSTE1	NC_020237.1	FALSE	FALSE
lentus	pSTE2	NC_006871.1	FALSE	FALSE
lugdunensis	pLUG10	NC_002093.1	FALSE	FALSE
pasteuri	pSP187	NC_007167.1	FALSE	FALSE
saprophyticus	pSES22	NC_007621.1	FALSE	FALSE
saprophyticus	pSSP2	NC_007352.1	FALSE	TRUE
saprophyticus	pSSAP2	NC_016643.1	FALSE	FALSE
saprophyticus	pSSP1	NC_007351.1	FALSE	TRUE
saprophyticus	pSSAP1	NC_015432.1	FALSE	FALSE
sciuri	pC194-like	NC_010626.1	FALSE	FALSE
sciuri	pACK6	NC_006974.1	FALSE	FALSE
sciuri	pSCFS1	NC_005076.1	FALSE	FALSE
simulans	pLNU5	NC_008351.1	FALSE	TRUE
simulans	pLNU2	NC_007769.1	FALSE	TRUE
simulans	pACK4	NC_013033.1	FALSE	FALSE

Species	Plasmid name	Accession	pWBG749 OriT?	pSK41 OriT?
simulans	pACK3	NC_013945.1	FALSE	FALSE
simulans	pACK1	NC_013944.1	FALSE	FALSE
sp. 693-2	SAP008A	NC_013371.1	FALSE	FALSE
sp. 693-2	pLEW6932	NC_009130.1	FALSE	FALSE
sp. CDC25	SAP018B	NC_013388.1	FALSE	FALSE
sp. CDC25	SAP018A	NC_013387.1	FALSE	FALSE
sp. CDC3	SAP020A	NC_013373.1	FALSE	FALSE
warneri	pvSw7	NC_020269.1	FALSE	FALSE
warneri	pvSw3	NC_020265.1	FALSE	FALSE
warneri	pSZ4	NC_020165.1	FALSE	FALSE
warneri	pvSw6	NC_020268.1	FALSE	FALSE
warneri	pvSw2	NC_020264.1	FALSE	FALSE
warneri	pvSw1	NC_020274.1	FALSE	FALSE
warneri	pvSw5	NC_020267.1	FALSE	FALSE
warneri	pvSw4	NC_020266.1	FALSE	FALSE
warneri	pPI-1	NC_005207.3	FALSE	FALSE

APPENDIX 2: PLASMIDS CONTAINING pSK41 ORIGIN-OF-TRANSFER MIMIC SEQUENCES

Details on plasmids containing pSK41 origin-of-transfer mimic sequences as noted in Appendix 1.

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	NC_007928.1		pSA1308	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_010231.1	EU333812.1	pKH18	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	AY048756.1		pC55s	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NZ_AKYW01000028.1		pNewBould 305	pSA1308	A	Genome	FALSE	FALSE	MINUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_013550.1		pBORa53	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_013331.1		SAP047A	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_019146.1		pUR5425	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
haemo-lyticus	NC_007770.1		pLNU3	pSA1308	A	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
sapro-phyticus	NC_007351.1		pSSP1	pSA1308	A	TRUE	FALSE	FALSE	MINUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900412.1		SAP052A	pSA1308	B	TRUE	FALSE	TRUE	MINUS	GCGTTCTTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	AM990995.1		pS0385-3	pSA1308	B	TRUE	FALSE	FALSE	MINUS	GCGTTCTTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
epi-dermis	NC_013389.1		SAP108A	pSA1308	B	TRUE	FALSE	FALSE	MINUS	GCGTTCTTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC
haemo-lyticus	NC_008353.1		pLNU7	pSA1308	B	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTTAGAACGCATAAGTGCGCCCTTACGGGATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	AY541446.1		pBMSa1	pSA1308	C	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTTAGAACGTATAAGTGCGCCCTTACGGGATTTAAC
chromo-genes	NC_007771.1		pLNU4	pSA1308	D	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGCATAAGTGCGCCCTTACGGGAATTAAC
simulans	NC_007769.1		pLNU2	pSA1308	CD	TRUE	FALSE	FALSE	PLUS	GCGTTCTTTTAGAACGTATAAGTGCGCCCTTACGGGAATTAAC
aureus	NZ_AASB02000125.1	Unknown	USA300_ TCH959	pCA347/ pSK156	A	Genome	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	NZ_ADVP01000034.1	Unknown	ATCC 51811	pCA347/ pSK156	A	Genome	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	NZ_ADJJ01000025.1		A8796	pCA347/ pSK156	A	Genome	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	BX571858.1		pSAS	pCA347/ pSK156	A	Genome	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900392.1		pWBG750	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900424.1		SAP072A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_005011.1		pMW2	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900413.1		SAP053A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_002517.1		p21	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900425.1		SAP073A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC
aureus	CP002115.1		pSaa6159	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGTATAAGTGCGCCCTTACGGGATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	GQ900397.1		pWBG757	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900467.1		pWBG763	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_007931.1		pSA1379	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900410.1		SAP051A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NZ_ACKI01000002.1		A9635	pCA347/ pSK156	A	Genome	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_021552.1		pCA347	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_003140.1		pN315	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900394.1		pWBG752	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002135.1		p18807-P03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_010063.1		pUSA300 HOUMR	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002137.1		p18808-P03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_018963.1		p18810-P03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900409.1		SAP050A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002134.1		p18806-P03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	CP002149.1		pLAC-P03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900380.1		SAP015A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900403.1		SAP046A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002132.1		p18805-03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002143.1		p18811-03	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	CP002146.1		p18809-P04	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900388.1		SAP027A	pCA347/ pSK156	A	TRUE	FALSE	FALSE	PLUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_009619.1		pSJH101	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_009477.1		pSJH901	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900412.1		SAP052A	pCA347/ pSK156	A	TRUE	FALSE	TRUE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_010419.1		pTZ2162	pCA347/ pSK156	A	TRUE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900475.1		pWBG762	pCA347/ pSK156	A	FALSE	FALSE	FALSE	MINUS	ATGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_023278.1		pSA268	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_019148.1		pPM1	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	GQ900478.1		SAP056A	pCA347/ pSK156	B	FALSE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900401.1		pWBG759	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900382.1		SAP017A	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900399.1		pWBG747	pCA347/ pSK156	B	TRUE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900390.1		pWBG746	pCA347/ pSK156	B	TRUE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_010077.1		pEDINA	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	GQ900448.1		pSK156	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_010684.1		pKH21	pCA347/ pSK156	B	TRUE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NC_013377.1		SAP105A	pCA347/ pSK156	B	TRUE	FALSE	FALSE	MINUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
epi-dermis	NC_008356.1		pLNU6	pCA347/ pSK156	B	TRUE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
pLNU5	NC_008351.1		pLNU5	pCA347/ pSK156	B	TRUE	FALSE	FALSE	PLUS	ACGTCGATTTATCCGACGT ATAAGTGCGCCCTTACGG GATTTAAC
aureus	NZ_AUPS01000033.1		pS1d	pCA347/ pSK156	C	Genome	FALSE	FALSE	MINUS	CACGTCGATTTATCCGACG TGTAAGTGCGCCCTTACG GGATTTAAC
aureus	NC_004562.1		pNVH01	pCA347/ pSK156	C	TRUE	FALSE	FALSE	MINUS	CACGTCGATTTATCCGACG TGTAAGTGCGCCCTTACG GGATTTAAC
aureus	NC_022598.1		pETB	pETB	A	TRUE	FALSE	TRUE	MINUS	ACCGTTTGGAACAAACGTA TAAGTGCGCCCTTACGGG ATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	NC_022598.1		pETB	pETB	B	TRUE	FALSE	TRUE	PLUS	AGCGTTTGGAACAAACGTATAAGTGCGCCCTTACGGGAGTTAA
sapro-phyticus	NC_007352.1		pSSP2	pETB	C	TRUE	FALSE	FALSE	PLUS	CGTTTGGAGCAAACGTGTAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_010279.1		pV030-8	pV030-8	A	TRUE	TRUE	FALSE	PLUS	TACGTTTGGAACAAACGTGTAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_018967.1		p18813-P03	pV030-8	A	TRUE	TRUE	FALSE	PLUS	TACGTTTGGAACAAACGTGTAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900434.1		SAP082A	pSK41	A	TRUE	TRUE	FALSE	MINUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900432.1		SAP079A	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_012547.1		pGO1	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_005054.1		pLW043	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_007792.1		pUSA03	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_020535.1		pI5S5	pSK41	B	TRUE	TRUE	FALSE	PLUS	GCGAACGTAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	NC_013653.1		pPR9	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900422.1		SAP069A	pSK41	A	TRUE	TRUE	FALSE	MINUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	GQ900433.1		SAP080A	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
aureus	AF051917.1		pSK41	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC

Species	Ref	Updated Ref	Plasmid_ name	OriT_ type	Sequence _variant	Complete Seq?	NES-type gene?	Multiple 41 OriTs?	Strand	Sequence
aureus	GQ900379.1		SAP014A	pSK41	A	TRUE	TRUE	FALSE	PLUS	GCGAACGGAACGTTCGCA TAAGTGCGCCCTTACGGG ATTTAAC
aureus	GQ900421.1		SAP068A	pSK41	A	TRUE	TRUE	FALSE	MINUS	GCGAACGGAACGTTCGCA TAAGTGCGCCCTTACGGG ATTTAAC

APPENDIX 3: MULTIPLE-SEQUENCE ALIGNMENT OF β-GLUCURONIDASE SEQUENCES OF INTEREST

The following multiple sequence alignment was used to define the loop class of each GUS sequence according to the location and length parameters described in Table 6.2.

```
CLUSTAL O(1.2.1) multiple sequence alignment

H_sapiens           MARGS---AVAWAALGPLLWGCALGLQGGMLYPQESPSRECKELDGLWSFRADFSDNRRR
F_prausnitzii       -------------------------MNRSLLYPRATTTRRLIGLDGMWRFSFD---PESK
E_coli              ----------------------------MLRPVETPTREIKKLDGLWAFSLD---RENC
S_agalactiae        ----------------------------MLYPLLTKTRNTYDLGGIWNFKLG---EH-N
C_perfringens       ----------------------------MLYPIITESRQLIDLSGIWKFKLN---EG-N
E_eligens           ----------------------------MLYPVLTQSRLLSDLSGVWDFKLD---NG-K
P_merdae            ---------MKYLFVACLL--CLSV---LSAVAKV-PAMNKIRLTNNWEYLKG---DLGG
H11G11-BG           -------------------------------------MREVININKNWLFSKK---EQPV
B_Fragilis          ----------------------------------MSLRQDILLNNNWNFRFS---HQVQ
B_ovatus            --------MKNRIIILCLVCLCLVN---IGLFAQETSPRTIFSLNEGWECRPI---TTVN
B_uniformis         MEREKNTLPQKACHWMAAVIISLFV------LPPVHAQRQTQTINDSWKFLKG---ECTA
B_dorei             --------MKRFAGWLLFFWGC-----------ICCICASEISITDSWKYKAE---NDER

H_sapiens           GFEEQWYRRP-------LWESGPTVDMPVPSSFNDI-SQDWRLRHFVGWVWYEREVILPE
F_prausnitzii       GVEAGWA-LD-------LPS---SLSMPVPASFCDL-FTDKASREYCGDFWYETSFFVPA
E_coli              GIDQRWWESA-------LQE---SRAIAVPGSFNDQ-FADADIRNYAGNVWYQREVFIPK
S_agalactiae        PN------EL-------LPS---DEVMVIPTSFNDL-MVSKEKRDYIGDFWYEKVIEVPK
C_perfringens       GLTEELSKAP-------LED---TIEMAVPSSYNDL-VESQEVRDHVGWVWYERNFTIPK
E_eligens           GFEEKWYEKP-------LKD---ADTMPVPASYNDL-KEGTDFRDHYGWVFYQRNISVPE
P_merdae            -----IWEAVRPAAPGSSEAVPIWQPVTLPHCFNAEDAVDPDVNYYEGPGWYKTLLAIDN
H11G11-BG           -------------P---KTLPEDWESVNLPHTWNGTDGQDGGNDYYRGKCCYVKLLKKAD
B_Fragilis          ----------------------GDTRRVDLPHTWNAQDALAGKIDYKRGIGNYEKALYIRP
B_ovatus            -------------------RKAPFTPVTIPHTWNTS-YIEGTTLYERKMMVYQRPLVVTK
B_uniformis         -------------AADSAFDDSKWTSIHLPHTWNTDAYT--EKDYYRGTGWYRRQLTLPQ
B_dorei             -------------FSSMDWNDSDWVTVDLPHTWNAGDVIDEQRGYRRGISWYRKKLFIPS

H_sapiens           RWTQDLRTRVVLRIGSAHSYAIVWVNGVDTLEHEGGYLPFEADISNLVQV----------
F_prausnitzii       EWSG---WDIVLRFGSVTHRARVFVNGVEVAQHEGGFLPFDATVTNIVRY----------
E_coli              GWAG---QRIVLRFDAVTHYGKVWVNNQEVMEHQGGYTPFEADVTPYVIA----------
S_agalactiae        VSED---EEMVLRFGSVTHQAKIYVDGVLVGEHKGGFTPFEVLVPECKYN----------
C_perfringens       TLLN---ERIVLRFGSATHEAKVYLNGELLVEHKGGFTPFEAEINDLLVS----------
E_eligens           YVKS---QRIVLRCAAVTHYAMIYLNGKLICEHKGGFLPFEVELNDDLQD----------
P_merdae            PYRN---GRIVLDFDGAGQKTDVYVYTTHVGSHVGGYDSWNVDITDAVKAFLGSKDAERF
H11G11-BG           LGEK---PVHYIQFDGVNSSAEVWWNGEKIGSHDGGYSAFRVRIPEISD-----------
B_Fragilis          EWKG---KRLFLRFDGVNSIADVFINRKHIGEHRGGYGAFIFEITDLVKY----------
B_ovatus            AMKN---KRLFLYFEGVNSAAQVFMNRRTVGEHLGGYTAFCIEITDEVKE----------
B_uniformis         GWKE---KQIILRLDAAGKSATIYINGKNVGEHAGGYTACSFNITPFLSF----------
B_dorei             EARD---KKITLRFDGVASKADVYLNGKLLKTHLGAYTAFGVDITDICEV----------
```

```
H_sapiens          GPLPSRLRITIAINNTLTP-TTLPPGTIQYLTDT-----SKYPKGYFVQNTYFDFFNYAG
F_prausnitzii      NQ---FNKLSVLANNELSE-TMLPAGTTCTLA-----------DGRKIAAPYFDFYNYAG
E_coli             GK---SVRITVCVNNELNW-QTIPPGMVIT-----------DENGKKKQSYFHDFFNYAG
S_agalactiae       NE---KIKVSICANNVLDY-TTLPVGNYSEII-------QEDGSIKKKVRENFDFFNYAG
C_perfringens      G----DNRLTVAVNNIIDE-TTLPVGLVKEV--------EVDGKKVIKNSVNFDFFNYAG
E_eligens          G----DNLLTIAVNNVIDY-TTLPVGGKANMMSGMMGGMGAGASDKPQNNPNFDFFNYCG
P_merdae           KG---KVPLSIRCDNSRDL-EMIPS------------------------DLADFNIYGG
H11G11-BG          -----ENILTVYADNSPN--DTVYP------------------------QVADFTFYGG
B_Fragilis         GE---KNSVLVRANNGEQ--LDIMP------------------------LVGDFNFYGG
B_ovatus           GE---NL-LEVWASNAYR--TDILP------------------------VSGDFNVNGG
B_uniformis        DT---PNTLAVCVDNAR---QDIAP------------------------ISGDFTFFGG
B_dorei            GK---ENLLAVKVDNSSSLGEILPP------------------------VSGDFSIFGG

H_sapiens          LQRSVLLYTTPTTYIDDITVTTSVEQ----DS----GLVNYQIS--VKGSN--LFKLEVR
F_prausnitzii      IHRPVWLMALPKERVLDYSTRYRLTE----TG----AEIDYTVS--TNGP----HPVTVE
E_coli             IHRSVMLYTTPNTWVDDITVVTHVAQDCN--H----ASVDWQVV--ANG------DVSVE
S_agalactiae       VHRPLKLMIRPKNHIFDITITSRLSDDLQ--S----ADLHFLVE--TNQK---VDEVRIS
C_perfringens      IHRPVKIYTTPKSYIEDITIVTDFKE----NN----GYVNYEVQ--AVGK----CNIKVT
E_eligens          ITRPVKIYTTPETYINDITVTADIDFTKEEPS----AVLNYNVE--IKGKDYNNITCKVE
P_merdae           LYRYLNLVYLPEVSFEQIHLESSLSSNL------KEGILKVKTSFYNP-EDIRKADVTVS
H11G11-BG          IYRDVTVIGVDESHFDLEFYGSSGIMITPKVSG-----LSAAVNITARVTNPQDCSVRFV
B_Fragilis         IYRDVHLLITDETCISPLDYASPGVYLVQEVVSPQEAKVCAKVNLSNRA-ADGTAELQVL
B_ovatus           IHRPCHLIVTGQDCISPLFYASPGVFIHQENISKTVADVNVETHLSLKN-KKQGLRLKTT
B_uniformis        IYRDVWLTAVPNQHFNLTNHGSDGLFISTPQVSEEQATLSIRGEVKNDAPEKATLELTHT
B_dorei            IYRRVFLQWTEKVHFVTEPYAAVPVRIQTPEVSVSEASMQIVAFLKNDFTDTKHVHVNVF

H_sapiens          LLDAENKVVANGT-------G-----TQGQLKVPGVSLWWPYLMHERPAYLYSLEVQLTA
F_prausnitzii      LYDGT-TRVAESS-------G-----TTGTLVVKNARLWNVH-----AAYLYDLVIRIHE
E_coli             LRDADQQVVATGQ-------G-----TSGTLQVVNPHLWQPG-----EGYLYELCVTAKS
S_agalactiae       VFDEDNKLVGE---------T-----KDSRLFLSDVHLWEVL-----NAYLYTARVEIFV
C_perfringens      IIDEENNIVAEGE-------G-----KEGKLTINNVHLWEPM-----NAYLYKLKVELLD
E_eligens          LFDEEGTKLSETE-------G-----SEGTFEISNVRLWQPL-----NAYLYKIKVTAGQ
P_merdae           VYDVDRKPVFSKTLEGILPLG-DQ--LLAKMKIKNPVLWDV-----DVPQLYTCELTVKT
H11G11-BG          VTDADKKPVGEK-------NV-DASDGKTVIEIENAHLWNGT----QDPYLYSLTAELL-
B_Fragilis         VTDGTKVICKESRNVS-LKQG-ADILEQLPLLIQKPRLWNGC----EDPFMYQVSISLH-
B_ovatus           VADADNKTV----ASN-EVEV-SDVIVKQPMKIHRPILWDGK----KNPYLYTVTVELY-
B_uniformis        IYRPDGTLLQTLKKNIQLKAGETYAFSNEATPVLKPELWTP-----ETPRLYRVETTLRN
B_dorei            LCDEMNRIVKEKQLKLKLIPGRKYPISTSVGRIENPHLWSP-----ELPYLYTVKVQVCD
```

```
H_sapiens          QTSLGPVSDFYTLPVGIRTVAVT-KSQFLINGKPFYFHGVNKHEDADIRGKGFDWPLLVK
F_prausnitzii      GS---AVVDEYLDRIGIRTFEIR-HGRFLLNGSPVYLRGFGRHEDADIRGRGLDLPTVKR
E_coli             QT----ECDIYPLRVGIRSVAVK-GEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVH
S_agalactiae       DN---QLQDVYEENFGLREIEVT-NGQFLLNRKPIYFKGFGKHEDTFINGRGLNEAANLM
C_perfringens      DE---EIIDTYFEEFGVRTVEVK-DGKFLINNKPFYFKGFGKHEDSYVNGRGINEAINIK
E_eligens          --------DVYTLPYGVRSVRVD-GTKFLINEKPFYFKGYGKHEDTFPNGRGINLPMNTK
P_merdae           ----PDQTFTTEERFGFRHTEFKDKGPFFLNGKRLLLRGTHRHEDHAGVAQAMTEDMMRR
H11G11-BG          -K-DGEKTDEISVRFGCRSFSIDPQKGFILNGKPCPLRGVSRHQDRPGIGNALTEKEHRE
B_Fragilis         -K-DGKQIDSVTQPLGLRYYHTDPDKGFFLNGKHLPLHGVCRHQDRAEVGNALRPQHHEE
B_ovatus           -D-GNLLKDRMVQRTGFRYFSVDHEKGFFLNGEYLNLYGFCRHEDAVGRASALLPEDYRM
B_uniformis        RK-TKTLLDQSNHYTAFRWFRFDGDEGFFLNGKPYKLRGICRHQDQKPIGPALTDEMHRR
B_dorei            AK-NGEMYQEVISPVGFRWFSVD-KTGFYLNGKYLKLRGAARHQDYAGLGTAIPVEMNRR

H_sapiens          DFNLLRWLGANAFRTSHYPYAEEVMQMCDRYGIVVIDECPGVGLALPQF-----------
F_prausnitzii      DFELMKWIGANCFRTSHYPYAEEIYQMADEEGFLIIDEVPAVGFMQSTANFLAANQGNGR
E_coli             DHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVVIDETAAVGFNLSLGIGF--EAGN-K
S_agalactiae       DLNLLKDMGANSFRTSHYPYSEEMMRLADRMGVLVIDEVPAVGLFQNFNASL--DLSP--
C_perfringens      DFNLMKWIGANSFRTSHYPYSEEIMRLADREGIVVIDETPAVGLHLNFMAT---GFGG-D
E_eligens          DISIMKWQHANSFRTSHYPYSEEMMRLCDEEGIVVIDETTAVGVNLQFGGGA--NFGG-E
P_merdae           EMRMMKDMGVNFIRLGHYQQSEIILDLCDELGILVWEEIPWCRGG---------------
H11G11-BG          DMDLICELGANTIRLAHYQHSRVFYDLCDECGMAVWAEIPYISRH---------------
B_Fragilis         DVALMREMGVNAIRLAHYPQATYMYDLMDKHGIVTWAEIPFVGPGGYADKGF--------
B_ovatus           DMELIKESGATAMRLAHYPHAEPMYDLSDENGIILWTEIPMCGPGGQAFTGF--------
B_uniformis        DFLLMKEMGANFIRISHYPQDDALLEMCDKLGMLAWEEIPIIDIV---------------
B_dorei            DMRLLKEMGANFVRISHYPQDPEIYRACDELGLIVWSEICVVNEV---------------

H_sapiens          ----F-----NNVSLHHHMQVMEEVVRRDKNHPAVVMWSVANEPASHLE-----------
F_prausnitzii      QQGFFEK-ETTPALLKNHKAALTDMIDRDKNHPSVIAWSLLNEPQCTSA-----------
E_coli             PKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMWSIANEPDTRPQ-----------
S_agalactiae       --KDNGT-WNLMQTKAAHEQAIQELVKRDKNHPSVVMWVVANEPASHEA-----------
C_perfringens      AP-KRDT-WKEIGTKEAHERILRELVSRDKNHPCVVMWSVANEPDSDSE-----------
E_eligens          RIGTFDK-EHGVQTQEHHKDVIRDLISRDKNHACVVMWSIANEPDSAAE-----------
P_merdae           --------LGGDVYKKQARRMLANMIVQHHNHPAVIIWGLGNENDWPNDFNTFDKS----
H11G11-BG          ----------MPGGRENTVSQMKELIYQNINHPSIIVWGLSNEITMNGASD----S----
B_Fragilis         --------VDQASFRENGKQQLIELIRQHYNHPSICFWGLFNELKEV-------GD----
B_ovatus           --------VDTEGYKDNARLAVKELVYQKFNHPSICFWGICNEILVSDGKRFVEYD----
B_uniformis        --------PNTPGYGDNCERNLREMIRQHYNHPSIITWGYMNEILLVTQRKYKTEAELKP
B_dorei            --------RKNTAFAHNCKEMLKEMILQNYNHPSVVLWGAMNELWDYHKQ----------
```

```
H_sapiens          ---SAGYYLKMVIAHTKSLDPS-RPVTFVSNS-------NYAADKGAPYVDVICLNSYYS
F_prausnitzii      ---GTEEYFKPLFELARRLDPQKRPRTYTVLMTS-----LPDTSKGQRFADFVSLNRYYG
E_coli             ---GAREYFAPLAEATRKLDPT-RPITCVNVMFC-----DAHTDTISDLFDVLCLNRYYG
S_agalactiae       ---GAHDYFEPLVKLYKDLDPQKRPVTLVNILMA-----TPDRDQVMDLVDVVCLNRYYG
C_perfringens      ---GAKEYFEPLIKLTKELDPQKRPVTVVTYLMS-----TPDRCKVGDIVDVLCLNRYYG
E_eligens          ---GAYDYFKPLYDLARELDPQKRPCTLVSVQGT-----TADTDCSSQLSDVICLNRYYG
P_merdae           ---AIRAFMKELHDMAHRLDDT-RMTAIRRCEFC------------NDIVDVYSPSIWAG
H11G11-BG          ---SLIENHRMLNDLVHKIDPT-RPTTIAVLSMC------DPGEEYVRIPDVLSYNHYFG
B_Fragilis         ---NPVEYVKELNALAKQEDPT-RPTTS---ASN------QDGNL-NFITENIAWNRYDG
B_ovatus           ---NPIPFIKELNGIYKSIDSS-RLTAL---ATC------VDQSYYLGCSDLIAWNKYFG
B_uniformis        VLERTLALANRLERVLKEEDST-RISTMAFHGSN-----SYNETGLSKITDIVGWNLYQG
B_dorei            ----AIALARELEALKKELDPY-RLSCVAFHAFTWEKPYTQSSKEMFSISDVNGVNVYES

H_sapiens          WYHDYG-HLELIQLQLATQFENWYKKY-QKPIIQSEYGAETIAGFHQDPPL---------
F_prausnitzii      WYVLGGAGLADAEAAFHHEMDGWAKVLHGRPLIFTEYGTDNLSGAHKLPSV---------
E_coli             WYVQSG-DLETAEKVLEKELLAWQEKL-HQPIIITEYGVDTLAGLHSMYTD---------
S_agalactiae       WYVDHG-DLTNAEVGIRKELLEWQDKFPDKPIIITEYGADTLPGLHSTWNI---------
C_perfringens      WYVAGG-DLEEAKRMLEDELKGWEERCPKTPIMFTEYGADTVAGLHDTVPV---------
E_eligens          WYFGGP-DLEVSEIGLRKELSDWGKL--GKPVMFTEYGADTVSGLHDTTSV---------
P_merdae           WYRGVFT-------DYKSISEQEMQKV--KHFLHVEWGGDSHARRHSEDAFYNLKNIEAG
H11G11-BG          WYGGKTD-------MYGPWFDKFHKKYPDRAVGMSEYGCEALNWHTS-DPQ---------
B_Fragilis         WYGSTPK-------TLATFLDRTHKKHPELRIGISEYGAGASIYHQQ-DSL---------
B_ovatus           WYKDAAP-------SASKFFDDCRDSSKGIPVGVSEYGGGASINHHQ-WPL---------
B_uniformis        WYGGDLT-------GFEKFLAQQHQNHPTHPMIVSEYGAGSDKRLHSLHPR---------
B_dorei            WYQGDSA-------TIAPMFDKFCSYSTAKPRFLSEFGAGSDERIHSYTPR---------

H_sapiens          ------------------------MFTEEYQKSLLEQYHLGLDQKRRKYVVGELIWNFAD
F_prausnitzii      ------------------------MWSAEYQNEYLEMTHAVFDH--YDFVQGELVWNFAD
E_coli             ------------------------MWSEEYQCAWLDMYHRVFDR--VSAVVGEQVWNFAD
S_agalactiae       ------------------------PYTEEFQCDFYEMSHRVFDG--IPNLVGEQVWNFAD
C_perfringens      ------------------------MFTEEYQVEYYKANHEVMDK--CKNFVGEQVWNFAD
E_eligens          ------------------------MYTEEYQVEYYEMNNKVFDE--FDFVVGEQAWNFAD
P_merdae           KGGDERAGDASLYGGVPRASRDGD-WSESYVVRLIDWHLKEQET--MPWLTGTAYWPFKD
H11G11-BG          ------QGD----------------YTEEYQAKYHEDVIRQIAV--RPWLWSTHVWNMFD
B_Fragilis         ------KQP----------SASGWWHPENWQTYYHMENWKIIAE--RPFVWGTFVWNMFD
B_ovatus           ------AME---------DRSDSHFHPEEAQTFCHEGNWESFAK--RPYLWAKFIWVFAD
B_uniformis        ------AFD----------------FSIEYQQKYLEHYLPVLED--TPYICGGTHWNFID
B_dorei            ------TFD----------------FTPEFQLDFNRRYINEMEK--RPDYIGYSIWNLVD
```

```
H_sapiens          FMTEQSPT--RVLGNKKGIFTRQRQ-PKSAAFLLRERYWK-IANETRYPHSVAKS-----
F_prausnitzii      FQTTEGIL--RVDGNKKGIFTRQRQ-PKDAAYLFRKRWTT-LPVDFKKRKK---------
E_coli             FATSQGIL--RVGGNKKGIFTRDRK-PKSAAFLLQKRWTG-MN-FGEKPQQGGKQ-----
S_agalactiae       FETNLMIL--RVQGNHKGLFSRNRQ-PKQVVKEFKKRWMT-IPHYHNKKNSVK-------
C_perfringens      FATSQGII--RVQGNKKGIFTRERK-PKMIAHSLRERWTN-IPEFGYKK-----------
E_eligens          FATSQSLL--RVQGNKKGLFTRDRK-PKMVAHYFRNRWST-IPEFGYKTK----------
P_merdae           FSTPVRPDNPVPYVNQKGVVERDFT-PKESYYVFQSYWT-EKPMIHIYGHTWPVRWGGKD
H11G11-BG          FAADARSEGGENGMNHKGLVTFDRKYKKDSFYAYKA-WLSDEPFVHICGKRYIDRPESM-
B_Fragilis         FGAAHRTEGDRPGINDKGLVTFDRKVRKDAFYFYKANWNKQEPMIYLAEKRCRLRYQPE-
B_ovatus           FPSYMRQEGEKDGYNDKGLVTHDRKTKKDAFYFYKANWNP-EPMIYITSRRFTKRDNPK-
B_uniformis        FSSALRDES-MPRINNKGLVYADRT-PKDVYHYYQAAWRKDIPVLHIASRDWTDRAGVQQ
B_dorei            FQVDGRGDS-KPNLNQKGMLTEDRR-KKEIYYYCQARWS-DIPMIHIAGADWTKRVEICD

H_sapiens          -------------QCLENSLFT---------------------------------------
F_prausnitzii      ------------------------------------------------------------
E_coli             ------------------------------------------------------------
S_agalactiae       ------------------------------------------------------------
C_perfringens      ------------------------------------------------------------
E_eligens          ------------------------------------------------------------
P_merdae           DR----KEILVYSNCDEVELFVNGVSQGVKRRNSQDYPAAGLRWNCV-YQEGMNEIRAVG
H11G11-BG          ------TSVTVYTNEPSVELFANGKSLGVQKRGEFPF----FYFSVP--NEGETVLTAKA
B_Fragilis         ------QTFMAFTTAPEAELFVNGVSCGKQKADTYST----VVWKNVKLTSGENIIRVT-
B_ovatus           ------TDIKVFTNLKEATLYINNRKIGTMKPDEMNR----VIWKDIRLNDGRNIICVE-
B_uniformis        GNAPVYLPVKIYTNLSEVELFIDGISLGKQKTENYTA----TF-EVP-FSNRNPFLFAQG
B_dorei            DSINV-RKISVFSNQKTVELIHNGKSLGVREVVNGEA----VF-AVP-FINGENLLDARS

H_sapiens          ------------------------------------------------------------
F_prausnitzii      ------------------------------------------------------------
E_coli             ------------------------------------------------------------
S_agalactiae       ------------------------------------------------------------
C_perfringens      ------------------------------------------------------------
E_eligens          ------------------------------------------------------------
P_merdae           VKKKEKKEVSDVIRQE----------------------------------------------
H11G11-BG          GDC------TDESRIRKVDKANPDY----------VLQEEG-AVI---------------
B_Fragilis         -------------------TPGKKP----------LT---D-EVT---------------
B_ovatus           -------------------GKNGKG----------LL---S-DTC---------------
B_uniformis        NYQ--GKTVQDGLRINF--TPIPACLDANNLKGLELAVNVG-SQCFFTSDESQLTWLPDQ
B_dorei            GA------LSDRLKIQM--KLLSSRLTDSDVLLDGLCINLGQEHCYFIDPQLQEIWIPDK
```

```
H_sapiens          ------------------------------------------------------------
F_prausnitzii      ------------------------------------------------------------
E_coli             ------------------------------------------------------------
S_agalactiae       ------------------------------------------------------------
C_perfringens      ------------------------------------------------------------
E_eligens          ------------------------------------------------------------
P_merdae           -YQTAKWDKEAACQVSL-------------LSE---EGDTALVQVQLIDKNGIRCLSSKK
H11G11-BG          -----NWFEIETPPGYMS---VND-----TIGDILATAKGKLLALKI---LKMVRANMKK
B_Fragilis         -----VEYKEDREGHHHH---HH--------------------------------------
B_ovatus           -----EWYCIK----------------------------------------------------
B_uniformis        PYAAGSWGYIGGKEGT-------------AQTEIQNTADGPLFQTLR---NEIEGY----
B_dorei            PYTKGSWGYMDGKPFNSWPGSSHDGVRYGVGADIKNTFLEPLFQTFL---IGTTCY----

H_sapiens          ------------------------------------------------------------
F_prausnitzii      ------------------------------------------------------------
E_coli             ------------------------------------------------------------
S_agalactiae       ------------------------------------------------------------
C_perfringens      ------------------------------------------------------------
E_eligens          ------------------------------------------------------------
P_merdae           QITFEIAGDG---------SLICNLGTSTGSRKVQAYNGRALI-------RIKR----NE
H11G11-BG          NKGKSTGGMADMAKGMKINKSIIEMGKGFSVKRVCMMAGGLFTKEQILEINA--------
B_Fragilis         ------------------------------------------------------------
B_ovatus           ------------------------------------------------------------
B_uniformis        --------RFDAP------QGVYEIE---------LLFTDIFRRNAGIAYQLDRNGQQEN
B_dorei            --------RLDVP------DGVYEIG---------FYFTEPFSKDERK--NIVRTGVSAE

H_sapiens          ------------------------------------------------------------
F_prausnitzii      ------------------------------------------------------------
E_coli             ------------------------------------------------------------
S_agalactiae       ------------------------------------------------------------
C_perfringens      ------------------------------------------------------------
E_eligens          ------------------------------------------------------------
P_merdae           GNSVVA--------------VKSEGLPTAFLELKS--------------PK---------
H11G11-BG          ----------------SLNKIKKK--------------------------------------
B_Fragilis         ------------------------------------------------------------
B_ovatus           ------------------------------------------------------------
B_uniformis        RESTFGISINGEVVEESLSPCKESGYFRALRKKYYI-TNDKEYIDIRFHSTSGTCFLNGI
B_dorei            GQRVFDVSVNGEKLIDSLNLADSYGEQTAVVKTLVVNVRNHEGLEILLSPQKGQGVISGL
```

```
H_sapiens          ------
F_prausnitzii      ------
E_coli             ------
S_agalactiae       ------
C_perfringens      ------
E_eligens          ------
P_merdae           ------
H11G11-BG          ---QK-
B_Fragilis         ------
B_ovatus           ------
B_uniformis        KLRNIY
B_dorei            KVKKIR
```