DRUG DISCOVERY AND KINETICS IN *MYCOBACTERIUM TUBERCULOSIS*, *TRYPANOSOMA CRUZI*, AND SARS-CoV-2

A Dissertation

by

DRAKE MICHAEL MELLOTT

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank M. Raushel
Committee Members,	Thomas D. Meek
	Tadhg P. Begley
	Stephen H. Safe
Head of Department,	Joshua A. Wand

August 2021

Major Subject: Biochemistry

ABSTRACT

In the era of high-throughput drug discovery against enzyme targets, the utilization of rational drug design principles has become a path less traveled. This rational approach can provide researchers more intuitive methods to address challenging drug targets by elucidating the mechanistic details of an enzyme and designing molecules to exploit key interactions. Herein, we detail our application of synthetic chemistry, kinetic analysis, and rational drug discovery against a variety of pathogenic targets.

To guide our efforts in developing inhibitors of the cysteine protease cruzain, from *Trypanosoma cruzi*, we synthesized and kinetically characterized fluorogenic dipeptide substrates. Upon appending a fixed electrophilic warhead to our varied peptide scaffolds, we identified a linear correlation between the K_m of our substrates and K_i* of inhibitors, suggesting the binding affinity of the peptide scaffold plays a dominate role in inhibitor potency.

Beyond cruzain, we developed covalent inactivators to target the enzyme isocitrate lyase (ICL1) from *Mycobacterium tuberculosis*. Through screening succinate-based substrate analogs, we uncovered that *cis*-epoxy succinate is an uncompetitive inactivator of ICL1 that forms a covalent bond with the catalytic cysteine 191 (Cys_{191}) of the enzyme. In a different vein, we identified 5-descarboxy-5-nitro-D-isocitric acid (5-NIC) as a highly efficient mechanism-based inactivator of ICL1. Retro-aldol cleavage of 5-NIC by the enzyme forms glyoxylate and 3-nitropropionate (3-NP) in the ICL1 active site, which

allows facile formation of a thiohydroxamate adduct with Cys_{191} via reaction with the nitronate form of 3-NP.

Shifting away from ICL1, we synthesized and kinetically characterized peptide substrates containing varied lysine post-translational modifications against the sole NAD^+-dependent lysine deacylase from *Mycobacterium tuberculosis* (Rv1151c or *Mt*-Sirt). This revealed that *Mt*-Sirt has a strong specificity for long chain fatty acyl and succinyllysine substrates, thus allowing us to leverage this specificity to develop potent mechanism-based inhibitors that mimic myristoyllysine.

Finally, we identified a vinyl-sulfone-containing cysteine protease inhibitor, K777, as a potent inhibitor of SARS-CoV-2 in mammalian cells. Using activity-based protein profiling, we identified that the alkyne analog of K777 targeted human cathepsin B and L, but that only cathepsin L was able to process the SARS-CoV-2 spike protein *in vitro*.

ACKNOWLEDGEMENTS

I would like to first and foremost thank my advisor Dr. Thomas Meek for providing me the ability to pursue exciting scientific projects and my own independent ideas. These experiences have been critical in my growth and development as a scientist and as a colleague. Beyond my advisor, I would like to thank my thesis committee of Dr. Frank Raushel, Dr. Tadhg Begley, and Dr. Stephen Safe for their insightful discussions and guidance throughout this process. I would also like to thank my lab mates, past and present, for their support and assistance in designing experiments and interpreting data. In particular, I would like to thank Dr. Truc Pham, Demetrios Kostomiris, Dr. Ardala Katzfuss, and Jiyun Zhu for their assistance in both conducting experiments and providing critical insights at various stages in my graduate career. In addition, I have had the pleasure to mentor seven undergraduate students during my graduate career, and I would like to especially thank Hudson Fulcher, Mollie Rembish, and Kassie Wolf for their extensive contributions to my research and assisting in my growth as a mentor.

I would like to also thank my undergraduate research mentors from Millikin University, Drs. Anne Rammelsberg and Samuel Galewsky, for introducing me to laboratory research and allowing me to explore my scientific curiosities in an unfettered fashion. In addition, I would like to thank my research abroad mentors Dr. Paris Barnes from Millikin University, and Dr. Feng-Di Lung from Tunghai University in Taichung, Taiwan. Conducting research in a graduate lab in foreign country was key in solidifying my desire to pursue graduate school and taught me about the ability of science to transcend both cultural and language barriers.

In addition to my undergraduate mentors, I thank Dr. Gary Juneau, the owner of NMR Analysis and Consulting, a contract research organization that I worked for part-time during my undergraduate years. Dr. Juneau provided me the opportunity to experience and conduct science in the industrial sphere and was not only an excellent employer, but a friend. I would also like to thank my good friends, and fellow undergraduate chemistry classmates, Chelsea Hadsall and Brandon Barringer, PharmD., for their continued support and genuine interest in the science I conduct.

During my time as a graduate student at Texas A&M University, I have had the opportunity to meet and engage with many people who have helped me in both science and in my life outside of the lab. First and foremost, I would like many individuals from the Begley, Burgess, Gohil, Gomer, Laganowsky, Sacchettini, Raushel, and Russell Labs for their assistance in my research. Whether it was lending me a chemical, training me how to use an instrument, teaching me a new technique, or simply talking though research problems, these individuals proved invaluable in helping me grow as a scientist and carry out key experiments for my research. I owe them a debt of gratitude. In addition, I would like to thank a few important people during my time at TAMU. I would like to thank Dr. Su Tang for being a consistent source of insight and for the many inspiring late-night conversations that facilitated the development of new ideas and projects during my first few years of graduate school. Thank you to Dr. Inna Krieger for not only being a fantastic collaborator over the years, but also for her help in procuring other collaborations and assistance in experimental design. I would also like to thank Dr. Tessily Hogancamp for being such an enthusiastic colleague and generous and caring friend. In addition, I would

like to thank Dr. Seth Cory for his insight in crystallography, experimental design, and on the golf course. Thank you to Dr. Xueyun Zheng for always taking time to talk with me about my research ideas and assist me in understanding how mass spectrometry could play an integral role in advancing them. Thank you to Dr. Isita Jhulki for being a good friend and assistance in solving scientific problems. Thank you to Kristen Consalvo for not only providing me with Starbucks in exchange for my blood, but also for providing a space for us to both vent about the frustrations of graduate school. I would like particularly thank Dr. Zane Taylor for being a great friend and excellent colleague. Whether it was working through experimental issues over lunch, lending me reagents, late-night discussions in the ILSB, or finding the best food and bars at conference venues, he was always nearby and supportive.

Beyond the people who I have met during graduate school, I would also like to thank my family for their love, support, and encouragement over the years during this long and sometimes arduous process. I would like to particularly thank my sister for scientific support and strong interest in my research, but also for her out of the blue phone calls asking me how to fix an HPLC problem she encountered or what the pK_a of group on a molecule is. These calls, regardless of the situation, still always made me smile.

Last but not least, I would like to thank my girlfriend for her love, unwavering support, tolerance of my long work hours, late nights spent in lab, and listening to me talk ad nauseum about my research. She has played a vital role in continuing to spur me forward and has greatly assisted in my betterment as both a researcher and person.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Frank M. Raushel from the Department(s) of Chemistry and Biochemistry and Biophysics, Professor Tadhg P. Begley from the Department of Chemistry, and Professor Stephen H. Safe from the Department(s) of Veterinary Physiology and Pharmacology and Biochemistry and Biophysics and Professor Thomas D. Meek from the Department of Biochemistry and Biophysics.

The synthesis of the peptidomimetic vinyl heterocycle inhibitors discussed in Chapter II were produced by Dr. Bala Chenna and Dr. Linfeng Li. The kinetic characterization of these inhibitors was conducted by Dr. Linfeng Li and Dr. Xiang Zhai.

In Chapter III, Dr. Truc Pham conducted initial characterization of ICL1 with substrate analogs and *cis*-epoxy succinate. Kevin Chin also assisted with kinetics experiments for *cis*-epoxy succinate. Further, Dr. Inna Krieger resolved the high-resolution crystal structures of ICL1 bound to *cis*-epoxy succinate and conducted testing of the molecule against $mc^2$7000. Zahra Moghadamchargari conducted the protien mass spectrometry analysis of ICL1-*cis*-epoxy succinate complexes. I

In Chapter IV, Dr. Lawrence Harris and Dan Torres conducted the synthesis of 5-NIC. Dr. Inna Krieger resolved the high-resolution crystal structures of ICL1 bound to 5-NIC. Zahra Moghadamchargari conducted the protien mass spectrometry analysis of ICL1-3-NP and ICL1-5-NIC complexes.

In Chapter V, Kassie Wolf conducted cloning, expression, and purification of *Mt*-Sirt mutants. Synthesis of select thiourea inhibitors was conducted by Hudson Fulcher. In addition, Hudson Fulcher conducted kinetic analysis with the *Mt*-Sirt mutants and inhibition analysis with *Mt*-Sirt. Jiyun Zhu conducted the Alk-12 click-labeling and in-gel fluorescence.

In Chapter VI researchers from Texas A&M conducted synthesis of K777 and K777 alkyne (Dr. Bala Chenna), alkyne-azide enrichment and visualization (Jiyun Zhu), proteomic analysis of enriched samples (Dr. Klaudia Kocurek), enzyme kinetics (Demetrios Kostomiris, Dr. Linfeng Li), cloning (Dr. Ardala Katzfuss), and protein expression, and purification (Dr. Zane Taylor). In addition, researchers at University of California San Diego (Pavla Fajtová, Miriam A. Giardini, Danielle Skinner, Ken Hirata, Michael Yoon, Sungjun Beck, Aaron F. Carlin, Alex E. Clark, Laura Beretta, Vivian Hook, Anthony J. O'Donoghue, and Jair Lage de Siqueira-Neto) conducted western blotting of cathepsin L levels, proteomic analysis of cathepsin L levels from cell lysate, cell culturing of SARS-CoV-2 infected cells, and evaluation of K777 in infection assays. In addition, cell culturing of SARS-CoV-2 infected cells and evaluation of K777 in infection assays was also conducted by researchers at University of Texas Medical Branch (Chien-Te Tseng, Aleksandra Drelich, Jason Hsu, Vivian Tat) and Utah State University (Brett L. Hurst, Hong Wang), In addition, researchers from Selva Therapeutics assisted in preparing the manuscript and subsequent progression of K777 into clinical phase trials.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

This work was funded by the Department of Biochemistry and Biophysics Start-up fund and the National Institutes of Health [R21-AI127634-01].

NOMENCLATURE

$[E_t]$	Concentration of enzyme
[L]	Concentration of added inhibitor or inactivator
[S]	Substrate concentration
%	Percent
°C	Degrees Celsius
Δ	Change in/of
±	plus or minus
<	Less than
>	Greater than
~	About
≤	Less than or equal to
≥	Greater than or equal to
^{13}C NMR	Carbon-13 nuclear magnetic resonance
^{1}H NMR	Proton nuclear magnetic resonance
2-VG	2-vinyl-glyoxylate
2-VIC	2-vinyl-(2R,3S)-isocitrate
3-NHP	3-nitro-2-hydroxy-propionate
3-NIC	3-nitro isocitrate
3-NP	3-nitropropionate
3CLpro	3CL protease

(4-Pyr)Ala	4-pyridyl-alanine
5-NIC	5-descarboxy-5-nitro-D-isocitric acid
Å	Angstrom
AA	Amino acid
ABPP	Activity-based protein profiling
ACE-2	Angiotensin converting enzyme 2
ACS	Acetyl CoA synthetase
ADP	Adenosine diphosphate
ADPR	Adenosine 5′-diphosphoribose
Ala or A	Alanine
AMC	7-Amino-4-methylcoumarin
Arg or R	Arginine
Asn or N	Asparagine
Asp or D	Aspartate
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BCG	Bacillus Calmette–Guérin
Boc	tert-Butyloxycarbonyl
BSA	Bovine serum albumin
BTK	Bruton's tyrosine kinase
Caco-2	Human colon epithelium cells

Calu-3	Human lung epithelium cells
cAMP	Cyclic adenosine monophosphate
ACN	Aconitase
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
Cbz, Z	Benzyloxycarbonyl
CD_3OD or MeOD	Deuterated methanol
$CDCl_3$	Deuterated chloroform
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CID	Collision-induced dissociation
cis-EpS	(2R,3S)-epoxy-succinic acid
ClpP	Caseinolytic Mitochondrial Matrix Peptidase Proteolytic Subunit
cm	Centimeter
CO_2	Carbon dioxide
CoA	Coenzyme A
CobB	*E. coli and/or S. typhimurium* NAD-dependent deacylase
COMU	1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate

CoV-1	Coronavirus 1
CoV-2	Coronavirus 2
COVID-19	Coronavirus disease 2019
CPE	Cytopathic effect
CRISPR	Clustered regularly interspaced short palindromic repeats
CRM1	Chromosomal Maintenance 1
CS	Citrate synthase
CTSB	Cathepsin B
CTSL	Cathepsin L
$CuSO_4$	Cupric sulfate
Cys or C	Cysteine
D-IC	D-threo-isocitrate
D_2O	Deuterated water
Da	Daltons
DCM	dichloromethane
ddH2O	Sterile ultra-pure water
DEA	Diethanolamine
DIEA, DIPEA	N, N-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

DprE1	Decaprenylphosphoryl-*β*-D-ribofuranose-2′-epimerase 1
DprE2	Decaprenylphosphoryl-*β*-D-ribofuranose-2′-epimerase 2
DTT	Dithiothreitol (Cleland's Reagent)
E-64	N-[N-(L-3-trans-carboxyoxiran-2-carbonyl)-L-leucyl]-agmatine
E. coli	*Escherichia coli*
EC_{50}	Effective concentration of added inhibitor that reduces growth by 50%
EC_{90}	Concentration of added inhibitor that reduces growth by 90%
EchA19	Enoyl-CoA Hydratase
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor
em	Emission
EMR	Extended mass range
eq	Equation or Equivalents
ex	Excitation
FA	Formic Acid
FAD	Flavin adenine dinucleotide
FDA	Food and Drug Administration

Fmoc	Fluorenylmethoxycarbonyl
FRET	Förster resonance energy transfer
FUM	Fumarase
GabD1/D2	Succinate-semialdehyde dehydrogenase
GFP	Green fluorescent protein
Glc	Glycolate
Gln, Q	Glutamine
Glu, E	Glutamate
Glx	Glyoxylate
Glx-OH2	Hydrated glyoxylate
Gly, G	Glycine
GNAT	GCN5-related N-acetyltransferase
GST	Glutathione S transferase
h	Hour
H_2O	Water
HCD	Higher-energy collisional dissociation
HCl	Hydrochloric acid
HEK293	Human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	1,1,1,3,3,3-Hexafluoro-2-propanol
His or H	Histidine
His6 or H6	Hexa-histidine tag

HIV	Human immunodeficiency virus
hPhe	Homophenylalanine
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
HRV	Human rhinovirus
hSirt	Human sirtuin
Hz	Hertz
IC_{50}	Half maximal inhibitory concentration
ICD or ICDH	Isocitrate dehydrogenase
ICL	Isocitrate lyase
IDH	Isocitrate dehydrogenase
Ile or I	Isoleucine
IND	Investigational new drug
InhA	NADH-dependent enoyl acyl-carrier-protein reductase
IPTG	Isopropyl β-thiogalactoside
JAK3	Janus kinase 3
K	Kelvin
K-Pi	Potassium phosphate
K777	4-methyl-N-((S)-1-oxo-3-phenyl-1-(((S)-1-phenyl-5-(phenylsulfonyl)pentan-3-yl)amino)propan-2-yl)piperazine-1-carboxamide
k_{ac}	Rate of acylation

Kac	Lysine acetylation
k_{cat}	Rate of catalysis
k_{cat}/K_m	Specificity constant of catalysis
K_d	Apparent dissociation constant
kDa	Kilodaltons
k_{dac}	Rate of deacylation
KGD	Alpha-ketoglutarate decarboxylase
Ki	Inhibition constant
K_I	"Michaelis constant" of mechanism-based inactivation
K_i^*	Tight-binding inhibition constant
k_{inact}	Inactivation rate constant
k_{inact}/K_i	Specificity constant of inactivation
K_m	Michaelis Constant
k_{obs}	Observed rate of inactivation
k_{off}	Residence time
KOH	Potassium hydroxide
KOR	Alpha-ketoglutarate ferredoxin oxidoreductase
KRAS	Kirsten rat sarcoma viral oncogene homolog
LB	Lysogeny broth
LC-MS	Liquid chromatography–mass spectrometry
Leu or L	Leucine
LiOH	Lithium hydroxide

Lys or K	Lysine
M	Molar
m/z	Mass to charge ratio
mAbs	Milliabsorbance
MBI	Mechanism-based inactivators
Mbp	Megabase pairs
MCD	Methylcitrate dehydratase or malonyl-CoA decarboxylase
mce4	*Mycobacterium tuberculosis* Mce protein family 4
MCL	2-Methyl-isocitrate lyase
MCM	Methylmalonyl-coenzyme A mutase
MCS	Methylcitrate synthase
MDH	Malate dehydrogenase
MDR	Multidrug resistant
MeCN	Acetonitrile
MeOH	Methanol
MERS	Middle East Respiratory Syndrome
MES	2-ethanesulfonic acid
Met or M	Methionine
MEZ	Malic enzyme
mg	Milligram
Mg	Magnesium

$MgCl_2$	Magnesium chloride
MHz	Megahertz
MIC_{50}	Minimum inhibitory concentration to inhibit the growth of 50% of organisms
min	Minutes
mL	Milliliter
mM	Millimolar
MMCE	Methylmalonyl-coenzyme A epimerase
mmol	Millimoles
Mn	Manganese
mOD	Millioptical density
MoI	Multiplicity of infection
MS, MLS	Malate synthase
Mt-PAT	*Mycobacterium tuberculosis* acetyltransferase
Mt-Sirt or Rv1151c	*Mycobacterium tuberculosis* NAD-dependent lysine deacylase
Mtb	*Mycobacterium tuberculosis*
MurA	UDP-N-acetylglucosamine 1-carboxyvinyltranferase
N-5-NIC	Nitronate form of 5-descarboxy-5-nitro-D-isocitric acid
Na-Pi	Sodium phosphate
Na_2EDTA	Disodium ethylenediaminetetraacetate dihydrate

Na_2SO_4	Sodium sulfate
NaCl	Sodium Chloride
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NAM	Nicotinamide
NaOH	Sodium hydroxide
nM	Nanomolar
nm	Nanometer
NMePip	N-methyl-piperazinyl
NTA	Nitrilotriacetic acid
OADC	Oleic Acid Albumin Dextrose Catalase
ON	Overnight
p	Partition coefficient
P_1	Recognition residue 1
P_1'	Recognition residue 1 prime
P_2	Recognition residue 2
P_2'	Recognition residue 2 prime
P3N	Nitronate form of 3-nitropropionic acid
PAGE	Polyacrylamide gel electrophoresis
PAS	Para-aminosalicylic acid

PBS	Phosphate buffered saline
PCA	Pyruvate carboxylase
PCB	Penicillin-binding proteins
PCC	Propionyl-CoA carboxylase
PCK	Phosphoenolpyruvate carboxykinase
PCR	Polymerase chain reaction
PD	Pharmacodynamic
PDB	Protein Data Bank
PDHC	Pyruvate dehydrogenase complex
PDVF	Polyvinylidene fluoride
PEG4000	Polyethylene glycol 4000
PEP	Phosphoenolpyruvate
Phe or F	Phenylalanine
PK	Pyruvate kinase
PK	Pharmacokinetic
pK_a	Apparent acid dissociation constant
pK_b	Apparent base dissociation constant
Plpro	Papain-like protease
ppm	Parts per million
Pro or P	Proline
prpC	2-methylcitrate synthase
prpD	2-methylcitrate dehydratase

psi	Pound per square inch
PTM	Post-translational modification
PVHI	Peptidomimetic vinyl heterocycle inhibitors
R-	Rectus (Latin = right)
R^2	Coefficient of determination
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RPM	Revolutions per minute
RT	Room temperature
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
s or sec	Seconds
S-	Sinister (Latin = left)
S_1	Specificity binding pocket 1
S_1'	Specificity binding pocket 1 prime
S_2	Specificity binding pocket 2
S_2'	Specificity binding pocket 2 prime
SARS	Severe Acute Respiratory Syndrome
SBC	Structural Biology Center
SBDD	Structure-based drug design
SCoaS	Succinyl-CoA synthetase
SCS	Succinate synthase

SDH	Succinate dehydrogenase
SDS	Sodium dodecyl sulfate
Ser or S	Serine
SIR	Silent regulators of information
SIRT	Sirtuin enzyme
sKIE	Solvent kinetic isotope effect
S_N2	Bimolecular Nucleophilic Substitution
Succ	Succinate
t	Time
T. cruzi	*Trypanosoma cruzi*
T_3P or T3P	Propylphosphonic anhydride
TAPSO	2-Hydroxy-3-[tris(hydroxymethyl)methylamino]-1-propanesulfonic acid
TB	Tuberculosis
TBST	Tris buffered saline tween
TBTA	Tris((1-benzyl-4-triazolyl)methyl)amine
TCA	Tricarboxylic acid
TEA	Triethanolamine
TFA	Trifluoroacetic acid
Thr or T	Threonine
TLC	Thin-layer chromatography
TMPRSS2	Transmembrane protease serine-2

t_R	Retention time
TSA	Transition state analog
Tyr or Y	Tyrosine
UV	Ultraviolet
V	Volts
v/v	Volume per volume
Val, V	Valine
Vero E6	*Cercopithecus aethiops* kidney epithelial cells
v_i	Initial rate
v_i/v_o	Residual enzyme activity
V_{max}	Maximal velocity
V_o	Initial rate
wt	Weight
WT	Wild-type
wt/v	Weight per volume
Y_{max}	Upper-bound rate value at a given time
Y_{min}	Lower-bound rate value at a given time
ε	Extinction coefficient
λ_{max}	Maximal wavelength
μM	Micromolar
l_{em}	Emission wavelength
l_{ex}	Excitation wavelength

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Since the beginning of time, efficacious pathogens have had the ability to effect disease and morbidity throughout the world. For hundreds of years, the mitigation and/or elimination of these pathogens have comprised a principal goal of the scientific community and pharmaceutical industry. However, in more recent years, the rise of antibiotic resistance, limited therapeutic options for some diseases, and newly emerging pathogens have continued to challenge the scientific community to develop novel therapeutics for a large variety of diseases. In many instances, canonical approaches to drug discovery, such as phenotypic and target-based screening using large compound libraries followed by iterative structure-based drug design (SBDD) have often failed to provide tractable leads for many areas of drug discovery. Interestingly, many of the most successful therapeutic agents comprise natural products or derivatives of natural products which effect covalent bond formation with their protein or enzyme target of interest. Therefore, in the development of new molecules to target challenging enzyme classes we may be wise to take a chapter from the book of nature. Thus, in this chapter, I will briefly detail some key concepts regarding the kinetics of enzyme inhibitors and inactivators, followed by a discussion of the strategies that nature uses to form covalent bonds with enzymes and proteins, and how we as enzymologists and chemists can apply these approaches in developing new covalent enzyme inhibitors. Together, combining the clever paradigms that nature uses in drug discovery with rational drug design strategies and

mechanistic enzymology provide us a better chance of developing selective inhibitors and covalent inactivators of key pathogenic enzymes. Towards this end, my thesis research focuses on using mechanistic enzymology and rational drug design towards the development of selective covalent inhibitors to target key enzymes in *Mycobacterium tuberculosis*, *Trypanosoma cruzi*, and the recent emergent viral pathogen, SARS-CoV-2.

A. Mechanistic Enzymology: A Background on Inhibitor and Inactivator Kinetics

Enzymology is the branch of biochemistry interested in understanding how enzymes work through their relationship between structure and function.

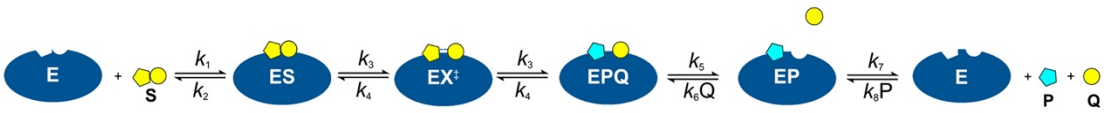


Figure 1. General Example of Unimolecular Enzymatic Catalysis

Where **E** is enzyme, **S** is substrate, **P** and **Q** are products, and **EX**‡ is the transition state. Microscopic rate constants k_1, k_2, k_3, k_4, k_5, k_6, k_7, and k_8 report on the respective rates of catalysis.

Therefore, mechanistic enzymology is a focused area of enzymology dedicated to elucidating the detailed kinetic and chemical mechanism(s) of an enzyme with respect to its interaction with binding partners (protein, small molecule substrate or inhibitor). When thinking about an enzyme, one may recall the simple lock and key illustration from biochemistry coursework. In this naïve depiction, an enzyme serves as the lock and the

substrate of the enzyme, the key. Through having the correct "notches" to interact with the enzyme, the substrate is able to bind to the enzyme and undergo transformation into product. However, this simple unimolecular chemical reaction is much more complicated than meets the eye. Take for example Figure 1 depicting the collision of substrate with enzyme. The initial binding of the substrate forms the reversible **ES** complex, often referred to as the Michaelis complex. Subsequent chemical transformation of the **ES** complex through the k_3 step results in the formation of the transient transition state (**EX**‡**)** which is rapidly (~10^{-13} sec) transformed into an **EPQ** complex. The desorption of **P** and **Q** from the enzyme (either in an ordered or random fashion) produces free enzyme (**E**) ready for another iteration of catalysis. In the direction of product formation, the maximal rate of substrate conversion to product is coined the turnover number, k_{cat}, which is equal the maximal velocity divided by the concentration of the enzyme. The binding of the substrate to the enzyme and formation of the Michaelis complex is characterized as the Michaelis constant, K_m. Together, these terms allow one to determine the specificity constant of catalysis, or the k_{cat}/K_m.

When looking closely at the scheme of enzyme-substrate complexation and subsequent product formation in Figure 1, one can begin to identify strategies that would allow the design of small molecule modulators which would inhibit the function of an enzyme. One discernable approach may be to produce structural mimics of **S**. In this case, a non-catalytically competent **S** analog, **I** (pink ligand, Figure 2), would bind to **E** in such a manner that the binding of **S** would be impeded and result in the formation of an **EI** complex. Kinetically, this competitive inhibition affects the K_m of **S** binding, that is, where

the substrate is at low or mid-range concentrations, but not when it is saturation wherein the initial velocity approaches the V_{max}.

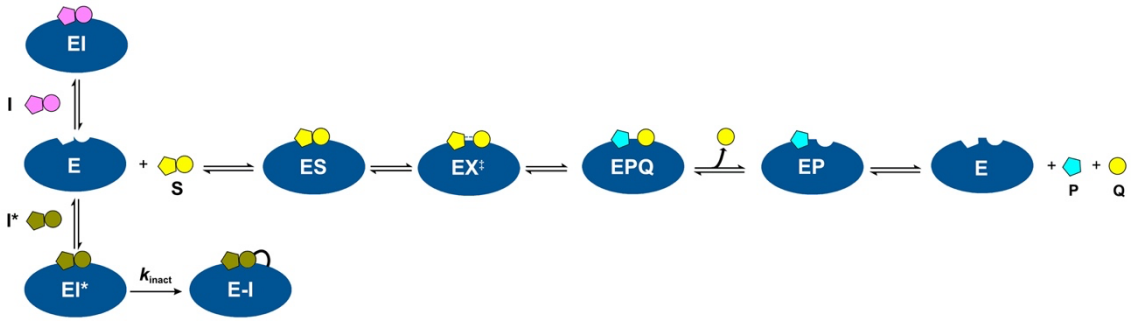


Figure 2. Competitive Inhibition and Inactivation of an Enzyme

Where the pink ligand represents a competitive inhibitor and the green ligand represents a competitive inactivator which forms a covalent adduct with the enzyme.

The design of molecules that are competitive versus the variable substrate is not limited to noncovalent molecules and can include molecules that engage in covalent bond formation with the enzyme, termed covalent inhibitors or inactivators. These molecules are in general electrophilic and allow one to exploit active site nucleophiles, such as the amino acid cysteine and serine, and develop a molecule that is structurally similar to **S** but contains a reactive group, shown here as **I*** (green ligand, Figure 2). The initial binding of **I*** to the enzyme forms a transitory **EI*** complex that affords the covalent addition of the molecule to **E** (k_{inact}) resulting in an irreversible **E-I** complex (Figure 2). In the case shown here, the inactivation event is illustrated as irreversible, but the development of

molecules that effect reversible covalent inactivation are becoming increasingly more common in the drug discovery sphere.[1-6] Another class of covalent inactivators that are competitive versus **S** are mechanism-based inactivators. This elegant class of molecules, which will be discussed in more detail later, masquerade as a substrate, but though enzymatic catalysis are converted to a reactive species which ultimately undergoes covalent inactivation of the enzyme target.[7-9] While these types of molecules can be challenging to produce, they provide an effective method to produce a selective molecule that will in theory only react with the target protein of interest due to the fact that they are only present in the active site of the target enzyme.

Beyond competitive enzyme inhibitors and inactivators, one can envision a number of other methods to target **E**. One such approach requires the design or discovery of molecules that binds to an adjacent site of the enzyme, whereby **I**, (brown diamond, Figure 3), can bind to the enzyme and form a **ESI** complex. This uncompetitive inhibition causes no reduction of the K_m of **S** and in the V_{max} as the inhibitor does not bind at the same site as **S**. A different and powerful strategy is the generation of molecules that are structural mimics of the transition state of an enzyme, coined transition state analogs (**TSA**, salmon ligand, Figure 3). This approach was initially theorized by Linus Pauling in the 1940s[10, 11] and was proposed to be a viable method to produce selective, tight-binding enzyme inhibitors. Work for nearly 80 years has indeed confirmed this initial hypothesis and led to the development of many transition state-based enzyme inhibitors, most notably for purine nucleoside phosphorylase[12] and for HIV protease, which lead to a front-line HIV therapy.[13] In addition to transition state analogs, uncompetitive inhibitors and

inactivators have the potential to provide opportunities to selectivity target an enzyme target. By definition, an uncompetitive inhibitor binds to a complex formed between the enzyme and substrate, and will result in a decrease in both the V_{max}/k_{cat} and the K_m of **S** for the enzyme.

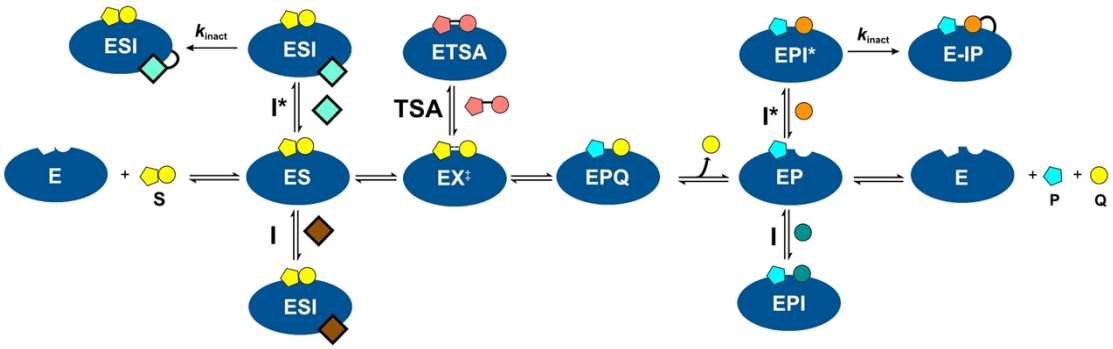


Figure 3. General Transition State Analog, and Uncompetitive Inhibitors and Inactivators.

Where the brown diamond represents an uncompetitive inhibitor, the light green diamond represents an uncompetitive inactivator, the salmon ligand represents a transition state analog (TSA), the gray circle represents an uncompetitive inhibitor and the orange circle represents an uncompetitive inactivator.

In the case of Figure 3 the binding of **I** (grey circle, Figure 3) or **I*** (orange circle, (Figure 3) to the **EP** complex to form an **EPI** or **EPI*** complex will affect uncompetitive inhibition of the enzyme (assuming ordered product release). In the case of **EPI*,** subsequent inactivation of **E** via the covalent addition to an active site residue, such as a nucleophilic cysteine,[14] will form the **E-IP** complex.

Together, through developing a strong understanding of the kinetic mechanism of an enzyme it allows one to use that information in aiding drug discovery when using a target-based approach. In many cases this rational design of molecules has the ability to result in the rapid production of possible molecular candidates and leverage the unique properties of the enzyme in a manner that inhibitors derived from high-throughput screening do not.

B. Diverse Electrophiles in Natural Products: Inspiration for Covalent Drug Discovery

In the development of small molecule modulators of enzyme activity, as detailed above, there are two major classes of molecules one generally will employ: noncovalent and covalent inhibitors/inactivators of enzymes. While both classes have their advantages and disadvantages, I will only focus on the development of covalent inhibitors/inactivators of enzymes for the remainder of this chapter. In recent years, one of the most sought-after properties of an enzyme inhibitor is its ability to engage in tight, slowly-reversible or effectively irreversible complexation with its target of intrest.[15] This property, often referred to as residence time, (the residence time is the half-life derived from k_{off}) has been shown to play a key role in the translation of *in vitro* to *in vivo* potency and selectivity of a small molecule drug more so than the static measurement of affinity (K_i, K_d, or IC_{50}) alone.[15, 16] There are multiple mechanisms by which one can produce drug-like molecules that exert prolonged target residence times, ranging from noncovalent, tight-binding molecules to the development of covalent modifiers that effect reversible or irreversible

target engagement. While there are various examples of noncovalent ligands which exert prolonged residence time on their target (methotrexate on dihydrofolate reductase,[17] DADMe-Imm-H on purine nucleoside phosphorylase,[12] and isonicotinic acyl-NAD(H) on InhA,[18] among others) it can be challenging to rationally design non-covalent molecules that exert these effects. Therefore, in general, the most facile approach to engineer enhanced residence time for a molecule is to use covalent drug design. Beyond residence time, the use of covalency (particularly when coupled with chemoproteomics)[19] serves as an important approach to identify proteins previously thought to be "undruggable." [19, 20] These undruggable targets are estimated to constitute ~85% of the human proteome,[20] and many contain poorly defined binding pockets or have proven intractable to traditional drug discovery programs.[20] While the work presented in this thesis will not focus on targeting enzymes with poorly defined binding sites, it will focus on targeting enzymes that have often proven refractory to high-throughput screening and traditional non-covalent drug design. Thus, it is important to consider alternate approaches in targeting these types of challenging enzymes, and a great place to start is by looking at how nature has targeted many enzymes and proteins.

When looking at varied isolated natural products that exert biological activity, a common trend is often revealed: covalent chemistry via electrophilic functionalities. These electrophilic "warheads" allow for efficient covalent addition to a target enzyme or protein while other portions of the structure often serve to position the warhead proximal to an enzyme nucleophile thereby allowing covalent bond formation to occur. Herein, I

give a brief, non-comprehensive, overview of these functionalities and highlight a few members key natural products that contain these reactive motifs.

In modern covalent drug discovery one of the most widely used electrophilic groups are Michael acceptors. Nature too has taken advantage of this type of electrophile in natural products and these molecules often exert potent bioactivity or reactivity. One example is the seemingly benign food additive, cinnamaldehyde which contains an α,β-unsaturated aldehyde functional group. This compound, among other electrophiles, can covalently modify the ion channel TRPA1 resulting in channel activation and thus production of a pain response.[21, 22] In addition, the synthetic incorporation of cinnamaldehyde into a peptidic scaffold as an electrophilic warhead to target the active site cysteine of protein tyrosine phosphatase resulted in the reaction of the aldehyde, in lieu of the vinyl, to form an enamine as determined by NMR.[23] Beyond this relatively simple molecule, a variety of other α,β-unsaturated moieties are found in other natural products ranging from amides (microcystin LR[24]), to ketones (curcumin[25] and zerumbone[26]), to δ-lactones (leptomycin B[27]), and even γ-lactones in molecules such as parthenolide (covalently modifies IKKβ in HeLa cells[28] and focal adhesion kinase in breast cancer cells[29]). In addition to α,β-unsaturated groups, there are other natural products which contain groups that afford Michael addition, such as a nucleoside analog containing the well-known thiol reactive maleimide group (showdomycin[30]), or methyldambullin,[31] which contains a vinyl sulfone and has been shown to exert and EC_{50} of 1.7 μM against *Trypanosoma cruzi*.[32] Further, the well-studied natural product

wortmannin contains an activated furan to which a lysine residue from phosphoinositide 3-kinase and polo-like kinase adds, resulting in the formation of a covalent complex.[33]

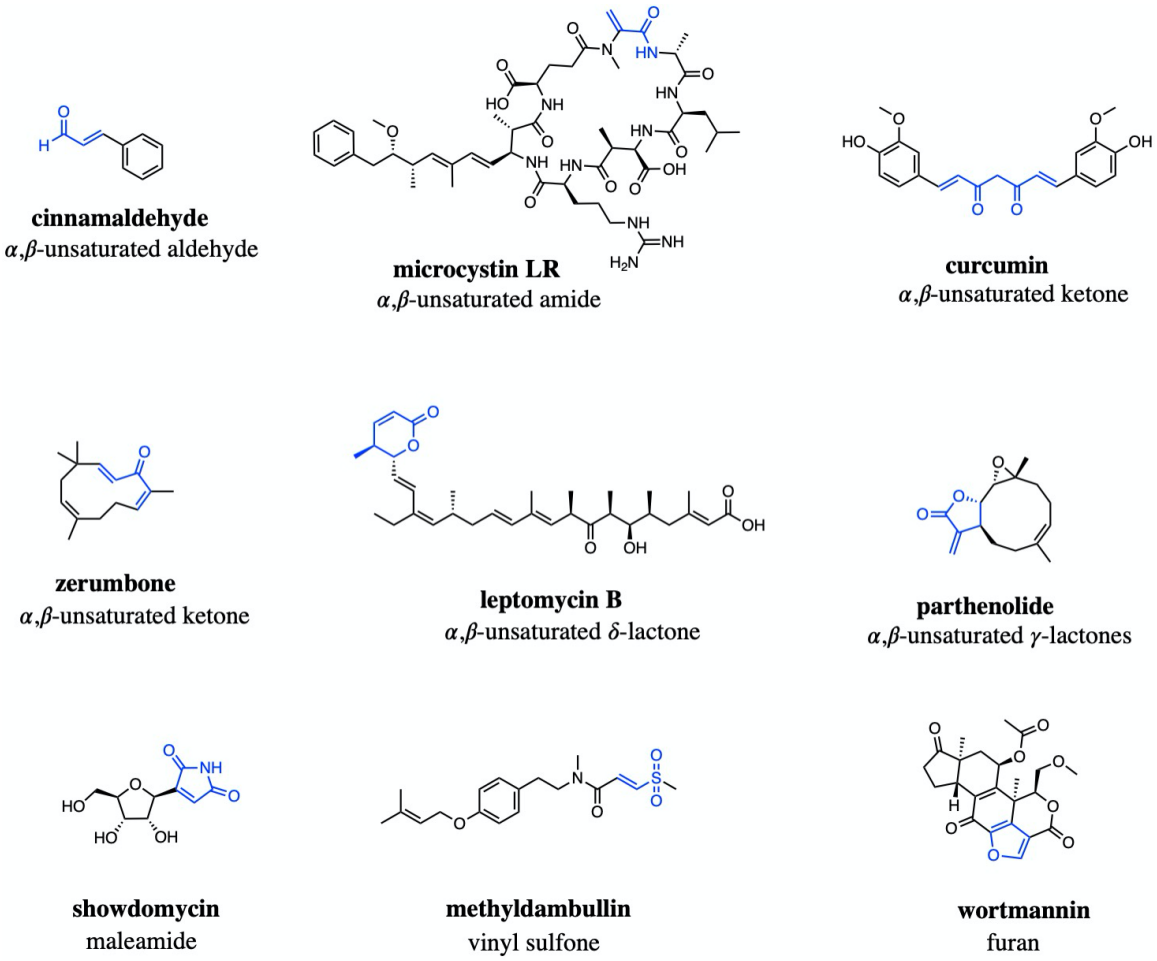


Figure 4. Natural Products Containing Michael Acceptors

Shifting away from Michael acceptors for targeting of proteins and enzymes, there are many important ring-strained electrophilic natural products which effect efficient covalent modification of various classes of enzymes. From a chemical reactivity perspective these functionalities are often weaker electrophiles than Michael-like

electrophiles and thus may result in heightened levels of selectivity. This was elegantly demonstrated by Weerapana *et al*. using activity-based protein profiling (ABPP) to screen the proteome reactivity of varied affinity labels (phenylsulfonate ester, linear epoxide, α-chloroacetamide, α,β-unsaturated ketone, and spiro-epoxide).[34] They noted that the α,β-unsaturated ketone ligand used exhibited robust modification of the proteome (304 unique peptides) at low concentration (1 μM), whereas ring strained epoxides exhibited negligible labeling and no enriched peptides up to a concentration of 20 μM.[34] While the seemingly unreactive character of these epoxide-based electrophiles on a proteomic scale is encouraging for development of selective covalent probes, there may be a few caveats regarding the high reactivity of the α,β-unsaturated ketone affinity observed by Weerapana *et al*. First is that the reactivity profile of all α,β-unsaturated ketones is not the same. The reactivity of these molecules is greatly dictated by the electron-withdrawing/donating character of groups proximal to their vinyl bonds. Therefore, it is possible to tune the reactivity of a vinyl bond in order to limit or oblate aberrant thiol reactions, as has been shown by our group[35] and others.[24] This modification of reactivity has been even taken a step further by developing Michael acceptors that undergo reversible covalent addition with enzyme targets. One such example of this is highlighted by the development of cyanoacrylamide warheads which comprise tunable reversible covalent Michael acceptors for Bruton's tyrosine kinase, one of which is now in phase III clinical trials,[36] thereby providing an approach to modulate the residence time of inhibitor on the enzyme.[5, 37] Therefore, while many α,β-unsaturated natural products do exhibit fair reactivity towards nucleophiles, an equal many have been seen to exhibit selectivity for discreet protein or

enzyme targets. Assessment of this selectivity can be achieved more efficiently now than ever before owing to approaches such as ABPP which allow for the relativity facile identification of the molecular target(s) and off-targets of a covalent inhibitor/inactivator.[38-40] Wright and Sieber provide a nice comprehensive review of the various chemoproteomics approaches used to identify the targets of natural product electrophiles which may guide this endeavor.[41] In the context of ring-strained electrophiles, this ABPP approach proved useful for analyzing the proteomic reactivity of ring strained β-lactams[42] and β-lactones.[43] These studies identified other possibly viable targets for β-lactams, such as the ATP-dependent Clp protease and a resistance associated β-lactamase[42, 44] The β-lactam warhead comprises an important electrophilic group for treatment of bacterial infections (penicillins, cephalosporins, etc.).[45] Structurally similar to the β-lactam, a γ-lactam-β-lactone bicyclic ring is found in the structure of salinosporamide A[46] (Phase III trials for glioblastoma, Celgene),[47] a potent inactivator of the chymotrypsin-like activity of the proteasome. Another type of ring-strained electrophile, an epoxide, comprises a common warhead in natural products. These weaker electrophilic warheads have seen wide use in research via E-64[48] (a pan-specific protease inactivator), and epoxomicin (selective proteasome inactivator[49]; derivative called carfilzomib is used for treatment of relapsed multiple myeloma),[50] as well as use in the clinic with the FDA-approved drug fosfomycin. Electrophiles with structural similarity to epoxides, such as aziridines and cyclopropyl groups also have been identified in natural products.

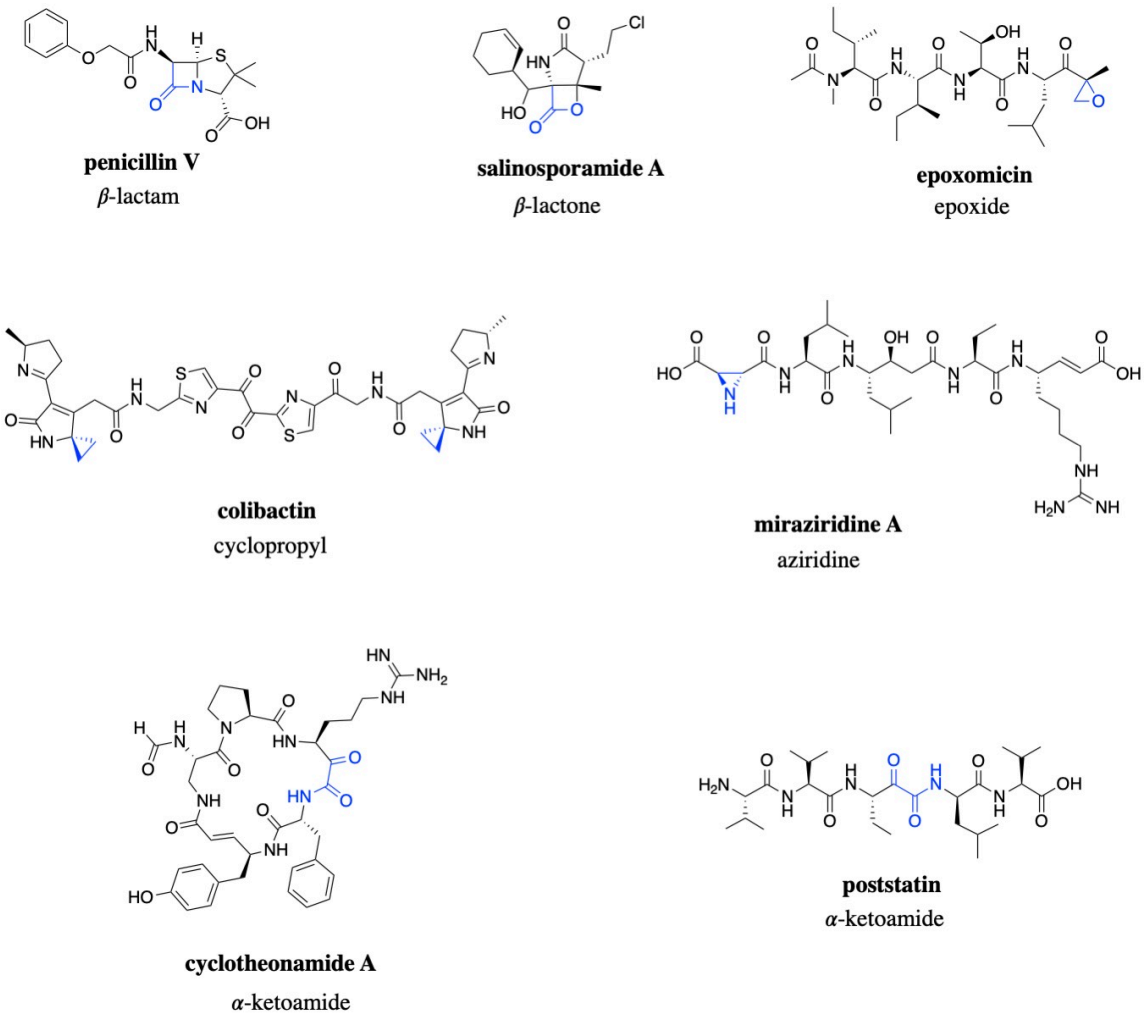


Figure 5. Electrophilic Natural Products Containing Strained Rings or α-Ketoamide Warheads

Beautiful biosynthetic work from the laboratories of Balskus[51] and Crawford[52] identified the structurally complex and challenging to isolate natural product colibactin undergoes DNA alkylation via its electrophilic cyclopropyl groups, an event that may play a role in colorectal cancer. With regard to aziridines, there are many natural products that contain

this warhead,[53] but one that is quite interesting is the molecule Miraziridine A.[54] This molecule is in effect a Swiss army knife of protease inhibitors as it has the ability to inactivate cysteine proteases with its aziridine warhead, aspartyl proteases with the amino acid statine, and an N-terminal vinyl arginine to inhibit serine, trypsin-like proteases.[54] In addition to Michael acceptors and strained rings, another group of electrophiles that need mention are the α-ketoamides. This moiety, shown in cyclotheonamide A and potstatin, comprise inhibitors of serine proteases and prolyl endopeptidase, respectively and has proved to be a popular warhead in the discovery of protease inhibitors. In fact, an α-ketoamide containing peptidic inhibitor was recently used to inhibit and solve the first crystal structure of the essential 3CLpro-like cysteine protease from SARS-CoV-2.[55]

In sum, nature has provided us as chemists and enzymologists a comprehensive molecular toolbox of tailored covalent approaches devised over millions of years of evolution. These approaches have allowed organisms to successfully defend themselves from the attacks of foreign agents by selectively targeting the weakness of these agents. Thus, we may be wise to leverage these tools and take note of key strategies employed in nature's development of selective covalent enzyme inhibitors.

C. On the Mechanism of Covalent Ligands Against Enzyme Targets

In order to put to use the diverse toolbox of electrophiles and approaches Nature has provided us to exploit enzyme and protein nucleophiles, it is important to consider how these molecules function with respect to their target(s) to aid in the design of new enzyme modulators. When discussing covalent ligands there are three major types: affinity

labelling agents,[56] quiescent affinity agents[57], and mechanism-based inactivators[7] (also known as suicide substrates).[8, 9] In this section I will provide examples of how Nature and researchers alike use quiescent affinity labels and mechanism-based inactivators to target enzymes. A quiescent affinity agent is defined as a molecule that is "quiescent" in the presence of solution-based nucleophiles at a reasonable concentration, but upon binding to an enzyme, a catalytic step, such as protonation, will activate the electrophile in the enzyme target.[58] However, the path of the enzyme inactivation for these molecules must be "aberrant" when compared to normal catalysis in order to be called a quiescent affinity label (Figure 6A,B).[58] From a kinetic view we provide two pathways in which a quiescent affinity label could function. One could envision this inactivation takes place either by the initial binding of the quiescent affinity label to the enzyme followed by inactivation (Figure 6A) or initial binding of the quiescent affinity label, a chemical step, and then inactivation (Figure 6B), as is the case with epoxide inactivators of HIV protease.[59, 60] In the case of a mechanism-based inactivator, or MBI, the molecule must be of reasonable structural similarity to that of a substrate or product of the enzyme and undergo a catalytically on-path transformation to yield a species that inactivates the enzyme prior to its release from the active site (Figure 6C).[7] Together, both quiescent affinity labels and MBIs, in contrast with affinity labelling agents like iodoacetamide,[61] and commonly used

ABPP probes,[62, 63] have a certain level of selectivity developed into their scaffolds that affords specific covalent modification of a protein or enzyme target.

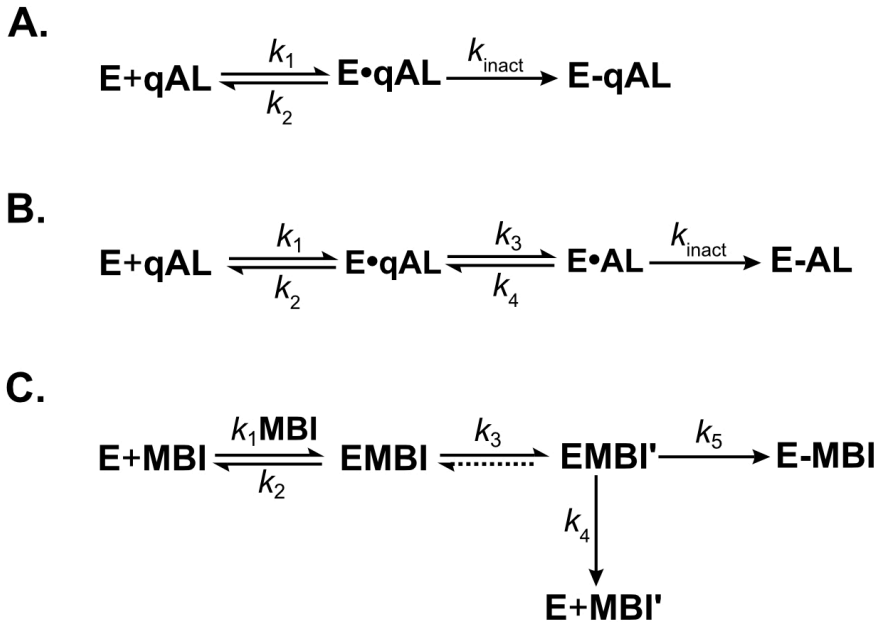


Figure 6. Kinetic Schemes for Quiescent Affinity Labeling and Mechanism-based Inactivation of an Enzyme Target

A. A kinetic scheme of a traditional quiescent affinity label which undergoes binding to an enzyme, formation of a E•qAL complex, and then undergoes covalent inactivation. **B**. This scheme details the function of a non-traditional quiescent affinity label which is more similar to a mechanism-based inactivator, which we coin a mechanism-based quiescent affinity label. In this case, the inactivator undergoes initial binding followed by a chemical step which activates the molecule for inactivation of the enzyme and results in the formation of an E-AL dead-end complex. **C.** Comprises a kinetic scheme for a mechanism-based inactivator, whereby the initial complexation of the MBI results in the formation of an EMBI complex. Subsequent catalysis of this pseudosubstrate results in the formation of a reactive EMBI′ species. This species can either be released from the enzyme active site or go on to inactivate the enzyme by the formation of a covalent adduct. The ratio of k_4/k_5 describes the partition ratio, which reports on the efficiency of the inactivation event, with the most efficient possible value being 0, demonstrating no release of the MBI′ species into solution.

When looking at the molecules that effect covalent inactivation that nature has provided, many contain built in selectivity to target specific enzymes. One such example is one of the most recognizable natural product reactive motifs, the β-lactam. This moiety is the active pharmacophore required for the antibacterial activity of the penicillins, cephalosporins, carbapenems, and monobactams. These molecules engage with the penicillin-binding proteins (PCBs), which are enzymes that carry out the transpeptidation or carboxypeptidation of the terminal D-alanine-D-alanine dipeptide on the peptidoglycan stem pentapeptide (Figure 7).[64] In the case of binding a penicillin-like molecule, the electrophilic carbonyl center of the β-lactam warhead is positioned proximal to the catalytic serine of the PCB, thus allowing for facile attack, ring opening, and the formation of a covalent complex (Figure 8). In addition, therapeutically relevant β-lactam molecules containing a C3 carboxylic acid proximal to the warhead provides nice mimicry of the C-terminal carboxylate of the D-ala-D-ala peptide. Modification of this C3 carboxylic acid precludes efficient binding and covalent bond formation with the PCBs.[42, 65, 66] One concern is the increasing resistance to β-lactam antibiotics.[67] One of the major mechanisms of this resistance occurs through the hydrolysis of the β-lactam ring by a class of enzymes known as β-lactamases which were discovered as early as 1940.[68] However, it seems that nature also identified this problematic hydrolytic enzyme and developed molecules, such as the mechanism-based inactivator clavulanic acid, to combat the degradation of β-lactam containing molecules.[69]

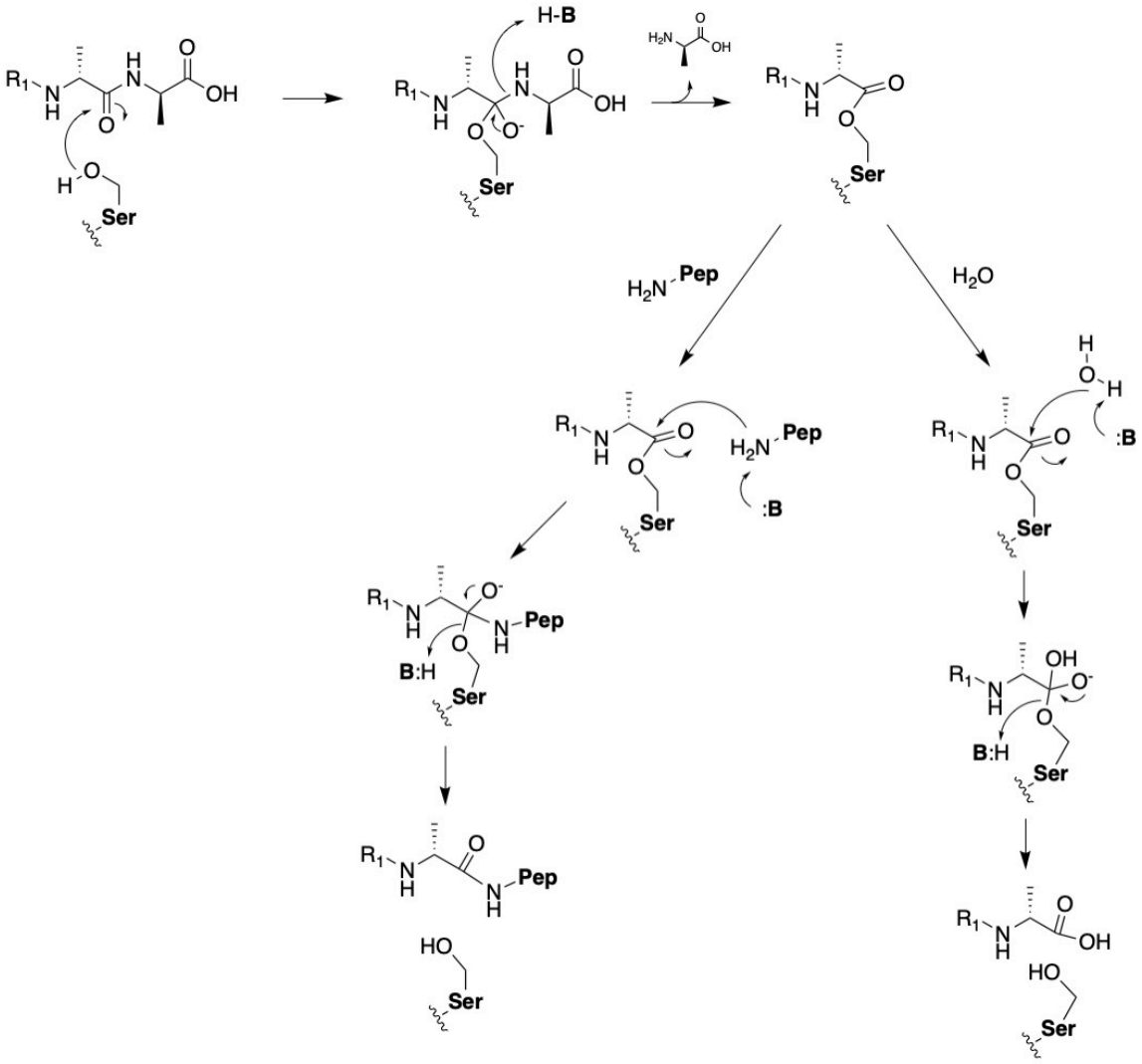


Figure 7. Chemical Mechanism of Penicillin-Binding Proteins.

Initially, the terminal D-ala-D-ala residues of the peptidoglycan pentapeptide bind to the PCB which positions its active site serine to attack the carbonyl carbon of the peptide bond. Formation of an acyl-enzyme intermediate, followed by collapse of this intermediate, results in the release of the C-terminal D-ala residue. The covalent PCB-D-ala complex can then either undergo hydrolysis to regenerate the free enzyme or transpeptidation where an amine, derived from another peptidoglycan stem pentapeptide, can add to the carbonyl-center of the enzyme bound intermediate, forming a tetrahedral intermediate which then collapses to release the transpeptidase product from the enzyme.

It took nearly 20 years from the identification of clavulanic acid to fully elucidate its mechanism of action on a β-lactamase. In a collection of beautiful work from Knowles,[70] Mobashery,[71] Chofield,[72, 73] and others it was identified that clavulanic acid undergoes a variety of chemical transformations in the active site of the β-lactamase enzyme, some of which allow for inactivation of the β-lactamase enzyme.

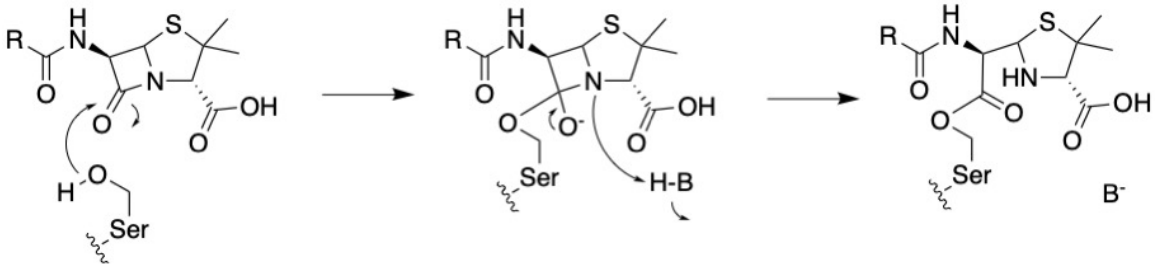


Figure 8. Mechanism of Inactivation of Penicillin-Binding Proteins by Penicillin

The active site serine of the penicillin-binding protein(s) attacks into the carbonyl of the β-lactam ring, resulting in the formation of a tetrahedral intermediate, which upon its collapse results in the formation of a stable ester bond linkage between the penicillin molecule and the protein.

Initially, clavulanic acid, akin to the mechanism observed for the penicillin-binding proteins, undergoes attack at the β-lactam ring carbonyl by an active site serine and in subsequent steps is transformed into an imine, which can tautomerize to its *cis*- and *trans*-enamine forms (Figure 9). Subsequently, these products can undergo a variety of chemical transformations ranging from the addition of a proximal active site serine residue to the imine, hydrolysis to produce an aldehyde species that can also react with the proximal

active site serine, of which these ester and ether adducts with the enzyme serine residues can be, thus regenerating the β-lactamase activity (Figure 9).[73]

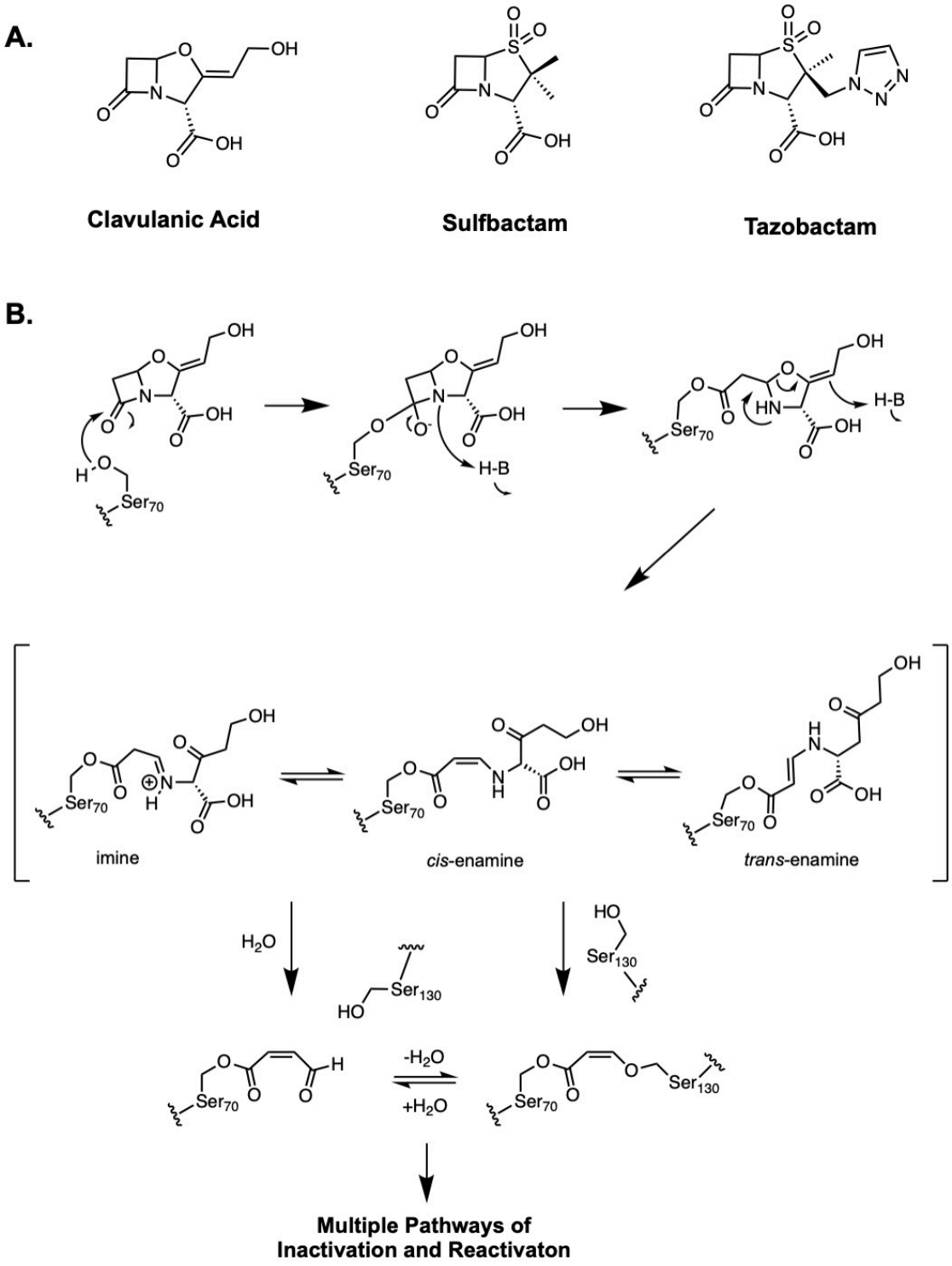


Figure 9. Mechanism-based Inactivators of β-Lactamases

A. Three known β-lactam-containing mechanism-based inactivators of β-lactamase enzymes which are used clinically. **B**. Putative mechanism of β-lactamase inactivation by clavulanic acid. For more detailed and complete mechanism please reference[73]

Only two years after to the identification of clavulanic acid, researchers at Pfizer described sulfbactam and its ability to protect β-lactam containing molecules from β-lactamases in various ampicillin-resistant bacterial strains.[74] In 1987 the synthesis of modified sulfbactams, lead to the identification of a molecule known called Tazobactam, a potent triazole containing β-lactamase inhibitor that is widely used in combination therapy with β-lactams.[75] While these molecules did inhibit β-lactamases, their partition ratio was quite poor as >100 molecules were required for 1 enzyme inactivation event,[76] demonstrating that identification of more effective MBIs to target β-lactamases may be necessary. Towards this end, researchers at AstraZeneca identified potent β-lactam-containing inhibitors which contained a diazabicyclooctane warhead, comprising the first non-β-lactam containing β-lactamase inhibitor (Figure 10A).[77] This molecule, called avibactam, is a highly efficient (a single molecule required per inactivation event[76]), slowly reversible covalent, inhibitor of a A-, C- and some D- class β-lactamases,[78] that is now used clinically to treat intra-abdominal infections in combination with the cephalosporin ceftazidime(Figure 10B).[79]

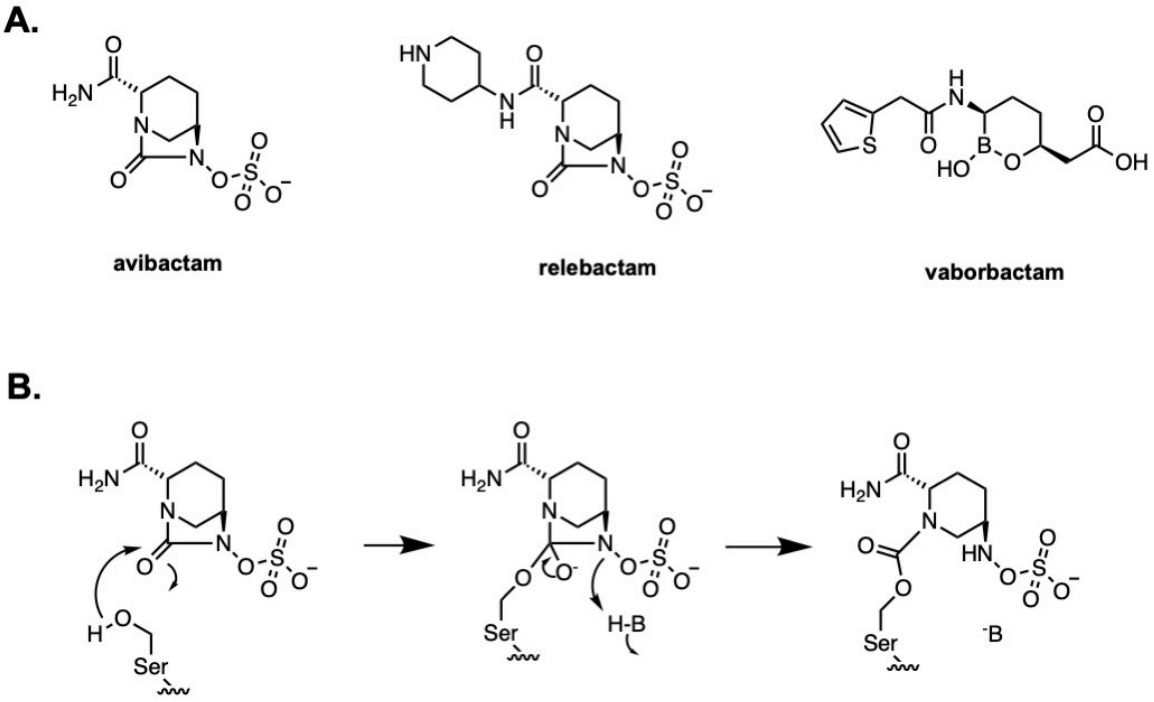


Figure 10. Non-β-Lactam-Containing β-Lactamase Inhibitors

A. Clinically used β-lactamases inhibitors containing diazabicyclooctane and boronate warheads. **B**. Proposed mechanism of a β-lactamase by avibactam that reacts with the active site a serine.

Subsequently, other non-β-lactam-containing inhibitors (vaborbactam with its boronate warhead, in clinic[80]; relebactam with its diazabicyclooctane core, in clinic)[81] have been identified as potent β-lactamase inhibitors which effect robust reversible covalent bond formation (Figure 10A). Therefore, to fight resistance against β-lactam antibiotics of interest, nearly all β-lactam therapeutic regimens now include a mechanism-based β-lactamase inhibitor such as clavulanic acid or avibactam. However, the rational drug discovery efforts against these enzymes still needs to continue as β-lactam resistant bacterial stains that contain varied β-lactamase enzymes continues to increase.[67] As of the

most recent data, over 7,100 β-lactamase enzymes have been identified in bacteria, an large increase from ~4,300 in 2019, demonstrating the magnitude of this growing problem.[82]

In addition to β-lactams, epoxide-based molecules are also effective quiescent affinity labels that can exert clinically-relevant therapeutic effects. One such antibacterial epoxide natural product, fosfomycin, targets MurA, which catalyzes the first step in glycan synthesis, converting UDP-N-acetyl glucosamine to enolpyruvyl-UDP-N-acetyl glucosamine using the co-substrate phospoenolpyruvate (PEP). Interestingly, the Michael acceptor-containing substrate PEP was shown to undergoes covalent addition with Cys_{115} of MurA and was co-purified with the enzyme.[83-85] It was observed that, fosfomycin, a structural analog of PEP, which contains an epoxide, undergoes robust inactivation of the enzyme (k_{inact} = 7.4 min^{-1}) via thiolation by the Cys_{115} thiolate.[86, 87]

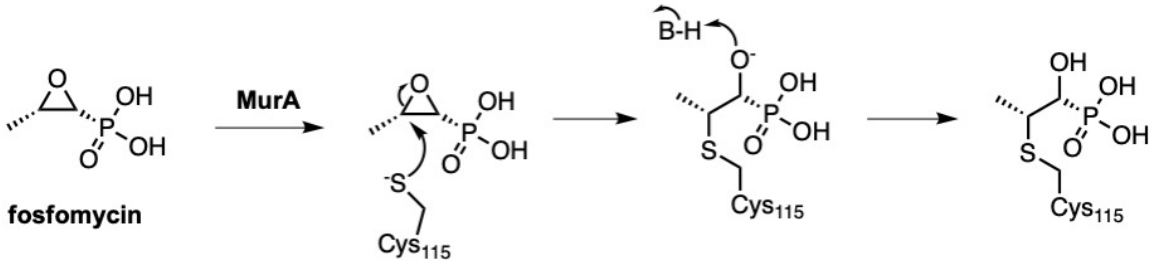


Figure 11. Putative Mechanism of Fosfomycin Inactivation of MurA

Structural evidence of this MurA inactivation reveals that thiolate addition occurs on the less sterically-hindered carbon, suggestive of an S_N2 inactivation pathway which is consistent with no pre-protonation of the epoxide ring oxygen being required (Figure

11).[88] In a similar vein, the epoxide-containing compound E-64 serves as a pan-specific irreversible cysteine protease inactivator.[48, 89-91] This molecule has seen extensive use as an active site titrant and tool compound to study the effect of cysteine proteases, such as the human cathepsins, particularly for coronavirus research.[92, 93]

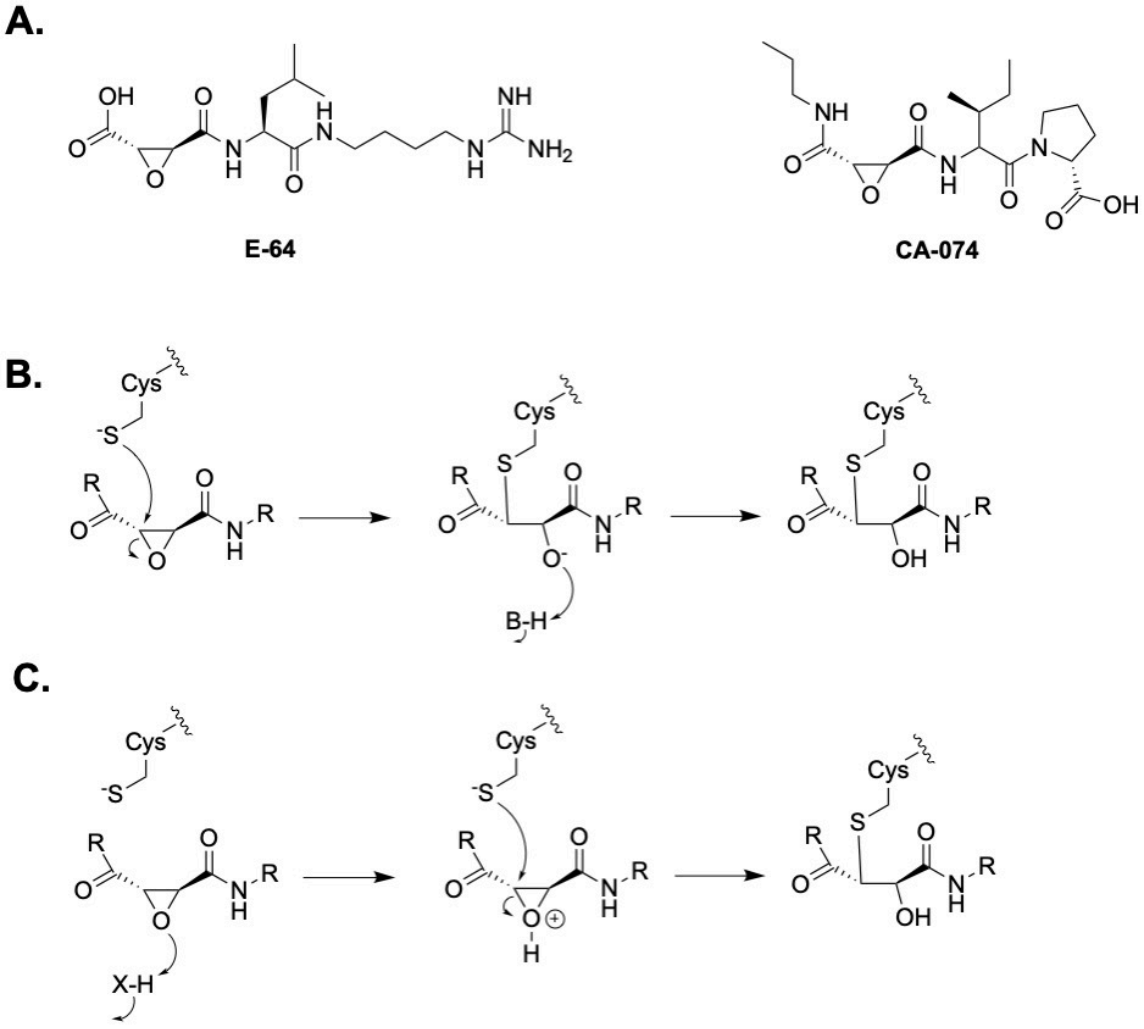


Figure 12. Structures of Epoxide Cysteine Protease Inactivators and Proposed Chemical Mechanisms of Inactivation

A. Structures of E-64 and CA-074. **B**. Mechanism for direct thiolate attack of epoxide and subsequent ring opening. **C**. Mechanism for epoxide ring protonation by an unknown proton donor (X-H) which allows for thiolate attack of the epoxide and ring opening.

While there is still some speculation regarding the precise mechanism of inactivation (Figure 12), it has been shown that chemical modification of the peptidic backbone of the E-64 scaffold allows the production of more selective protease inactivators such as CA-074, which has a high level of selectivity for human cathepsin B.[89] The conversion of this pan-specific molecule (E-64) to a highly specific inactivator (CA-074) requires important consideration into the interplay between selectivity and reactivity when designing an enzyme inactivator. In previous studies with the well-studied cysteine protease papain, Liu and Hanzlik[94] noticed a strong linear correlation ($R^2 = 0.98$) between the specificity constant of catalysis (k_{cat}/K_m) for substrates and the specificity constant of inactivation (k_{inact}/K_i) for inactivation when using inactivators with a fixed warhead and varied peptidic scaffold. This strong linear correlation suggested to the authors that the chemical mechanism for Michael addition of the vinyl warhead to Cys_{25} of papain was similar to that of substrate hydrolysis and that the binding energy from enzyme-ligand interactions may drive inactivation in a near analogous fashion to substrate hydrolysis.

Together, this situation, while not universally applicable, suggests that in order to design robust and specific quiescent affinity labels and mechanism-based inactivators one should have a detailed understanding of the key determinants of substrate binding as this information can maximize the probability of success when appending a warhead to the molecule of interest. More importantly, this highlights that optimization of specificity towards a target may allow for the installation of less reactive warheads which can undergo equally efficient inactivation compared to a more reactive warhead appended to a ligand that has less specificity for the target protein.

Beyond β-lactams and epoxides, there are various other elegant mechanisms that nature uses to inactivate enzymes with Michael acceptors such as the slow addition of the dehydroalanine residue of microcystin-LR to a non-catalytically relevant Cys_{273} residue of serine/threonine phosphatase 1A.[95] Interestingly, the elimination of this dehydroalanine residue from the inhibitor or substitution of the cysteine residue from the protein does not substantially alter the binding affinity of the molecule to the protein target, suggesting that while the addition of the Cys_{273} to the dehydroalanine is indeed an attractive approach to

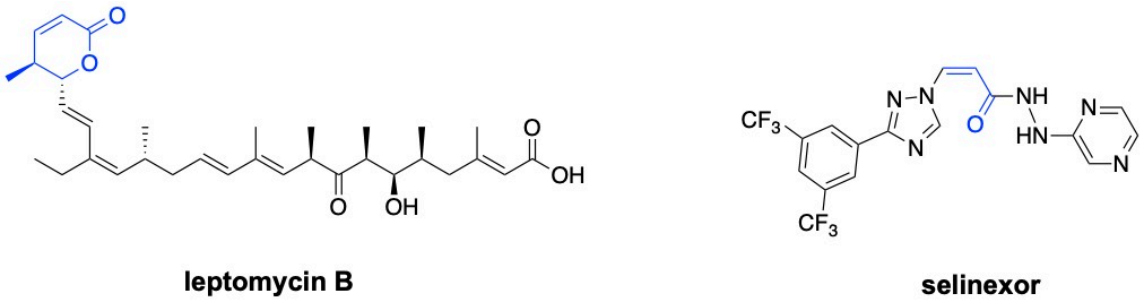


Figure 13. Structures of Leptomycin B and Selinexor

The warhead which undergoes reaction with Cys_{529} of CRM1 is labeled blue in color.

enhance residence time of the molecule, it is not wholly required for inhibition.[96] Another example of Michael addition in nature is the specific reaction of the α,β-unsaturated δ-lactone of the natural product leptomycin B with Cys_{529} of the CRM1, a conserved receptor for nuclear export signal of proteins (Figure 13).[27] In the case of leptomycin B, it binds to a channel in CRM1 used for the binding of leucine rich receptor signals, which affords inactivation via reaction with a proximal thiol (Cys_{529}) and ring opening of the α,β-unsaturated δ-lactone.[97]

This molecule, while exquisitely selective for CRM1 in cells,[27] was unsuccessful in clinical trials owing to systemic toxicity,[98] but provided inspiration for the development of molecules to inhibit CRM1. Using a mixture of docking and structure-based drug design researchers at Karyopharm were able to develop Selinexor, an α,β-unsaturated amide containing small molecule that effected covalent, slowly reversible, inactivation of CRM1 by binding to the same channel and alkylating the Cys_{529} (Figure 13).[99] This molecule has progressed to the clinic (as a combination therapy with bortezomib and dexamethasone) for the treatment of multiple myeloma and more recently for the treatment of relapsed or refractory diffuse large B-cell lymphoma.[100, 101]

This section serves to underscore a few paradigms present in many natural product electrophiles that are relevant to the design of new covalent and mechanism-based inactivators of enzymes and proteins. The defining characteristics of these molecules is their ability to mimic a substrate or product and form key noncovalent interactions to properly position an electrophilic warhead. This is highlighted by mimicry of the D-ala-D-ala dipeptide by β-lactam inhibitors of the penicillin binding proteins, the structural mimicry of PEP by fosfomycin, the selectivity induced by the peptide scaffold of E-64, and the selectivity of leptomycin B for the CRM1 channel used for binding of leucine-rich nuclear export signals. Together, these specific binding interactions allow for efficient positioning of the covalent warhead for subsequent reaction with the target nucleophile. In addition, the molecules discussed that contain β-lactam and epoxide warheads leverage the chemical mechanism of the enzyme target by exploiting its essential catalytic nucleophile. Conversely, the targeting of non-catalytically relevant nucleophiles by

leptomycin B and microcystin-LR also proved to be an effective strategy. This strategy has begun to take hold, particularly in cancer drug discovery, where many of the new classes of inhibitors use covalency. Such examples exist for epidermal growth factor receptor (EGFR), Bruton's tyrosine kinase (BTK), Kirsten rat sarcoma viral oncogene homolog (KRAS), and Janus kinase 3 (JAK3).[102, 103] Together, natural product electrophile-containing molecules can not only therapeutically viable molecules, but also

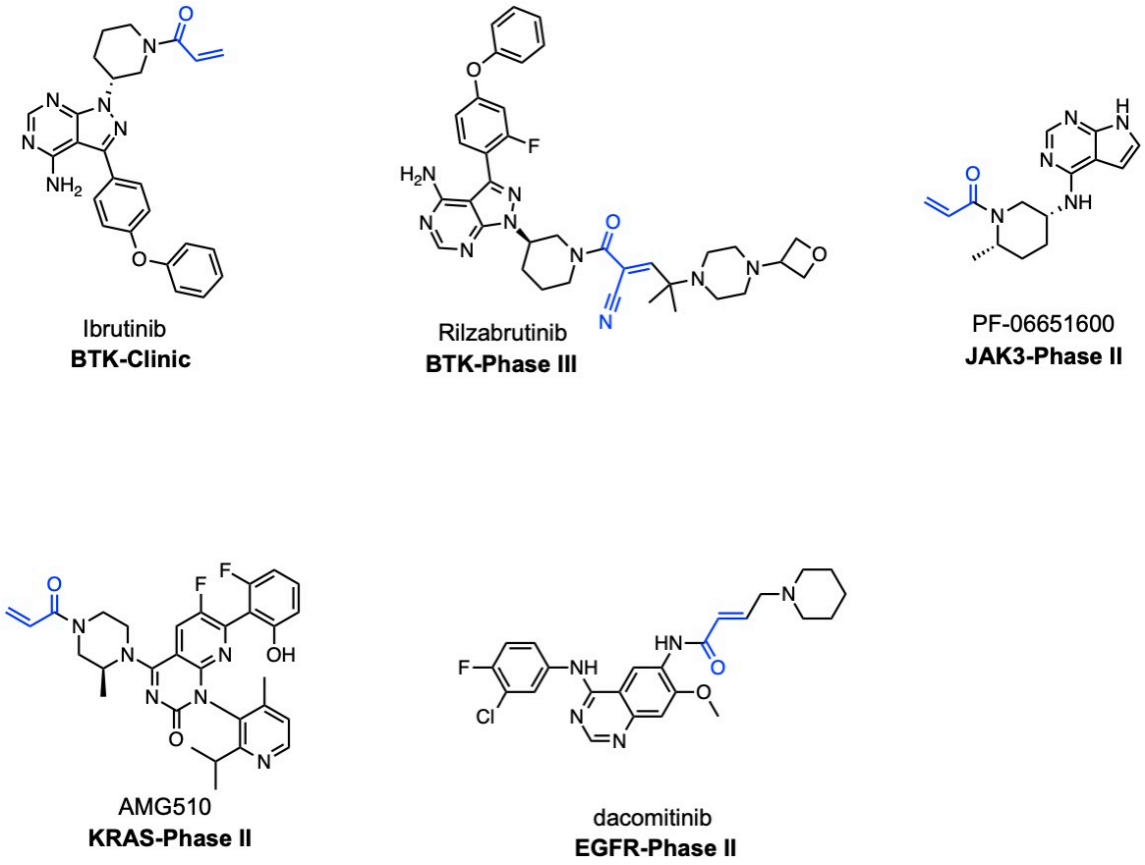


Figure 14. Examples of Covalent Inhibitors in the Clinic and in Clinical Trials for the Treatment of Cancer

The electrophilic warhead is colored in blue for each molecule.

can assist in playing a critical role in identifying new targets and methods to exploit enzymes with active site nucleophiles.

D. Summary

In sum, nature has provided us as chemists and enzymologist with key examples of how to better leverage molecules containing electrophilic warheads in the development of targeted therapeutics. Many of the molecules presented herein, while on their face range widely in electrophilic functionality and overall structure, have key conserved features which allow them to efficiently target a protein or enzyme. Often, the most important feature comprises the mimicry of a substrate or product of the enzyme or binding partner of a protein which positions the electrophile for facile modification to occur. Application of this strategy or approach, arguably pioneered by nature, to other enzyme systems by keen researchers have proven fruitful in the modern era of drug discovery. Key examples of this success come from a myriad of targets, ranging from targeting of kinases and proteasome for cancer chemotherapy using namely α,β-unsaturated carbonyls, α-ketoamides, and epoxides, to the targeting of key cysteine proteases implicated in disease (bacterial, mammalian, parasitic, viral, etc.) using warheads ranging from α,β-unsaturated groups, α-ketoamides, epoxides, halomethyl ketones, nitriles, nitro, among others. Together these approaches embolden researchers to continue to develop new approaches to selectively target enzymes and proteins with covalent warheads. Towards this end, in this thesis, I will discuss a variety of covalent drug discovery approaches we have taken in the laboratory. Chapter II focuses on better understanding the peptide recognition motif

of the parasitic cysteine protease cruzain from *Trypanosoma cruzi* to allow for the development of specific and robust reversible covalent inhibitors of this enzyme using a class of molecules known as peptidomimetic vinyl heterocycles. In Chapters III and IV we respectively discuss the development of an uncompetitive quiescent affinity label and mechanism-based inactivator against the essential *Mycobacterium tuberculosis* enzyme isocitrate lyase. In Chapter V we detail the substrate specificity characterization and development of potent mechanism-based thioamide and thiourea inhibitors against the sole lysine deacylase in *Mycobacterium tuberculosis*. In Chapter VI we detail the identification of K777, a dipeptide vinylsulfone containing Michael acceptor, as a potent and selective inhibitor of SARS-CoV-2 which has progressed to Phase II clinical trials for the treatment of COVID-19. In sum, armed with the knowledge that over 30% of marketed drugs use a covalent mechanism of action,[102] that successful drug molecules often exert prolonged residence on their primary targets,[15] and observation that covalent drug discovery has been revitalized in recent years,[104] we are optimistic that this work will serve to both aid in the development of new covalent lead molecules that may progress to the clinic and expand upon established approaches for covalent drug discovery.

CHAPTER II

SYNTHESIS AND KINETIC CHARACTRIZATION OF DIPEPTIDE SUBSTRATES FOR *TRYPANOSOMA CRUZI'S* CRUZAIN*

A. Introduction

Introduction

American trypanosomiasis, more commonly known as Chagas disease, is a deadly parasitic infection (~10,000 deaths/yr) caused by *Trypanosoma cruzi (T. cruzi)*. This parasite is endogenous to the Americas and results in 30,000 new cases each year, 8,000 of which are newborns that become infected during gestatation.[105] Although approximately 6-7 million people are estimated to be infected with *T. cruzi*, another 70 million reside in areas at risk for exposure.[105]

T. cruzi is transmitted to humans via insect vectors called triatomines, more generally referred to as kissing bugs.[106] These insects are common in parts of Mexico, Central, and South America where they have adapted to residing in or near domiciles.[107] In particular, the passage of the parasite to humans, or other mammals such as raccoons and dogs, occurs through the entry of kissing bug parasite-laden fecal matter into a bite wound or mucosal membrane.

---

*This work has been published in and is reprinted in part with permission from: Chenna, B. C., Li, L., Mellott, D. M., Zhai, X., Siqueira-Neto, J. L., Calvet Alvarez, C., Bernatchez, J. A., Desormeaux, E., Alvarez Hernandez, E., Gomez, J., McKerrow, J. H., Cruz-Reyes, J., and Meek, T. D. (2020) Peptidomimetic Vinyl Heterocyclic Inhibitors of Cruzain Effect Antitrypanosomal Activity, *Journal of Medicinal Chemistry 63*, 3298-3316.

The metacyclic trypomastigote form of the parasite (Figure 15) then enters the cells of the mammalian host and transforms into its amastigote form, which allows efficient multiplication within the infected tissue. Subsequently, the newly produced amastigotes can transform into the trypomastigote form and enter the blood stream. Once in the bloodstream, the parasite can travel through the vessels and localize in other tissues throughout the body, such as the heart. *T. cruzi* infection is often clinically silent for decades (known as the chronic phase of infection), but ultimately leads to dangerous and

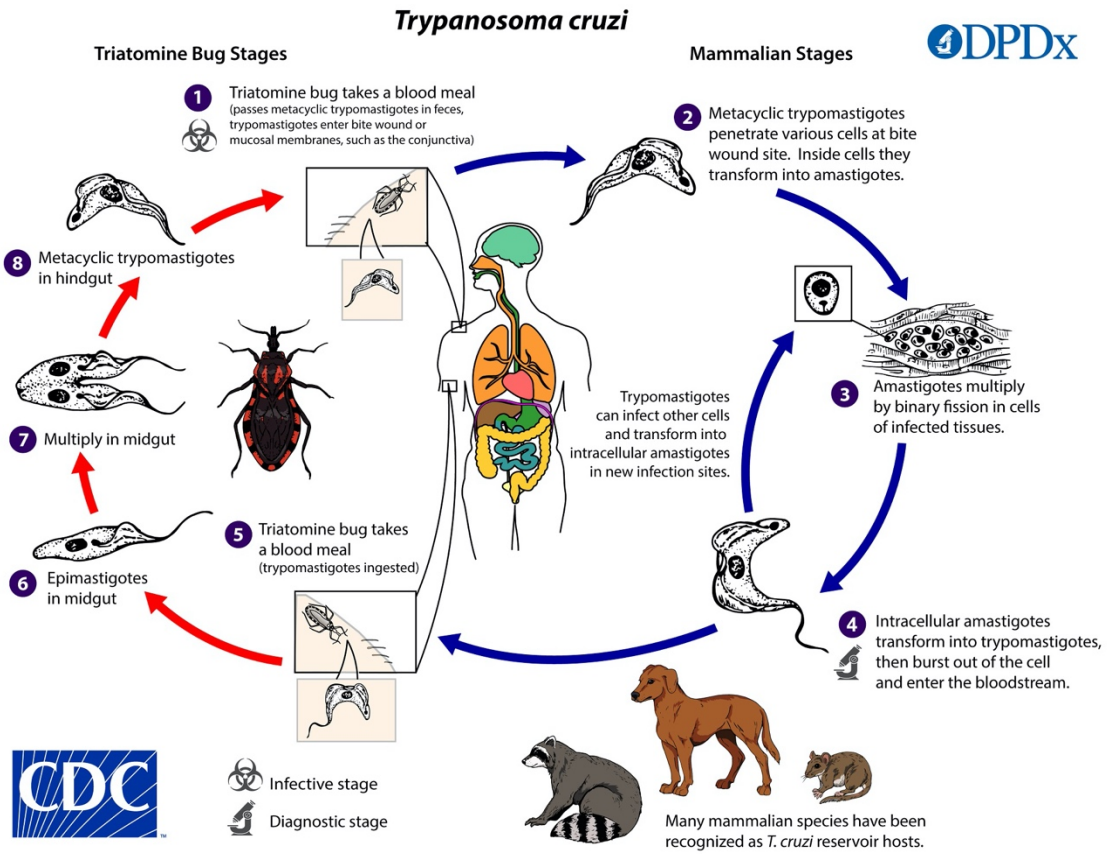


Figure 15. Life Cycle of *Trypanosoma Cruzi*

Image credit to the CDC, DPDx, and Dr. Alexander J. da Silva and Melanie Moser

potentially fatal complications of the gastrointestinal system (megaesophagus or megacolon)[108] and/or the cardiovascular system (dilated cardiomyopathy with heart failure).[109] While vector control has been fairly successful for the control and limiting of spread of triatomines,[110, 111] the therapeutic options for treatment of Chagas disease are limited.[108] Specifically, the two molecules that are approved to treat *T. cruzi* infection (benznidazole (Roche) and nifurtimox (Bayer)), are riddled with harmful side-effects, which afflict nearly 40% of patients and range in severity from gastrointestinal issues, sleep disturbance, appetite loss, weight loss, and dermatologic reactions, such as skin sensitivity.[112, 113] Moreover, these treatments have been shown to only work during the early stages of infection and are largely ineffective following ~3 weeks of infection.[114] With the treatment regimen lasting upwards of 2 months and high frequency of severe side-effects, many patients (30-50%)[115] fail to comply for the full duration of treatment. Therefore, there is an urgent need to develop new therapeutics for Chagas disease that act on all stages of the parasite's lifecycle.

One well-validated target for *T. cruzi* therapeutic intervention is the cysteine protease cruzain. This protease, a member of the family C1 clan CA cysteine proteases, has been shown to play an essential role in the ability of *T. cruzi* to carry out host infection and maintain infection of the host.[116] In addition, cruzain is also expressed in all life form stages of the parasite[117] making it an ideal target for the development of *T. cruzi* therapeutics.[118]

In general, proteases are highly sought-after drug targets for human bacterial,[119] parasitic,[120] and viral.[121-123] infectious diseases because of the often essential roles they

play in the progression of disease. These enzymes catalzye the hydrolytic cleavage of the amide bond of peptides and proteins and often employ a nucleophilic active site residue to catalyze this bond breakage. While there are four major classes of proteases (cysteine, serine, aspartic, and metalloproteases) I will be solely discussing cysteine proteases hereafter.

Cysteine proteases have a highly conserved catalytic dyad comprising a catalytic cysteine and histidine residue. Initiation of catalysis begins either by deprotonation of the catalytic cysteine by the catalytic histidine to generate a thiolate (as is the case in

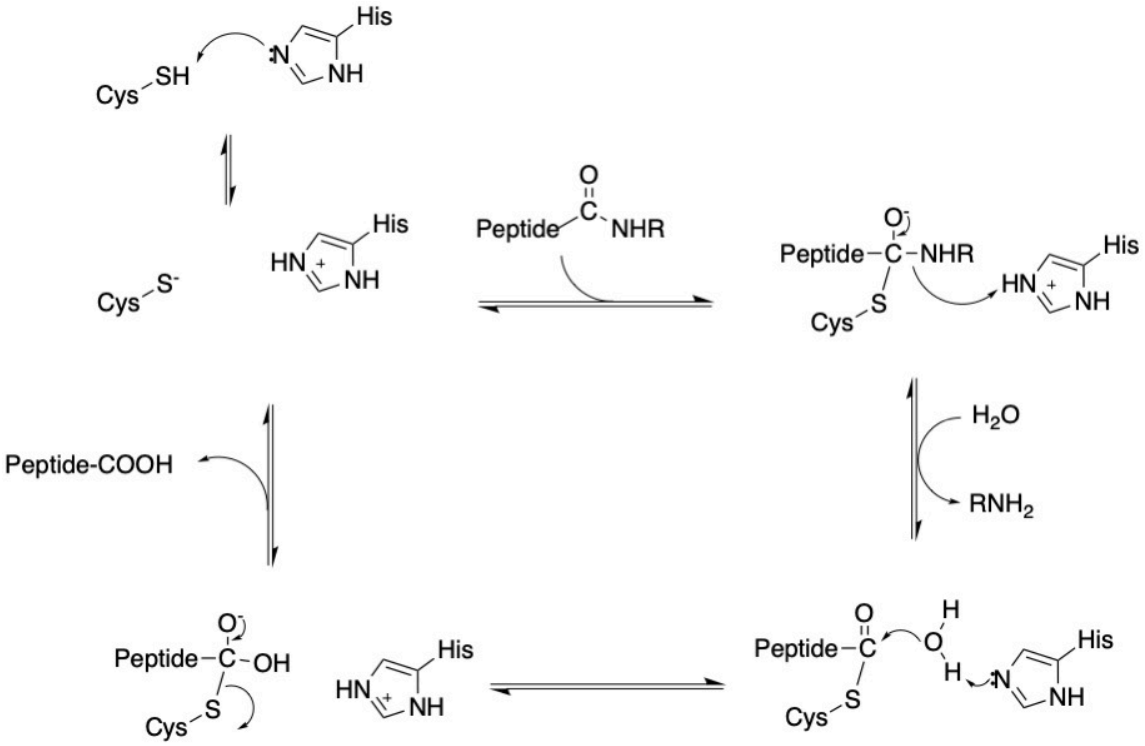


Figure 16. Chemical Mechanism of a Generic Cysteine Protease

cruzain[124]) or the enzyme may have a resting active site cysteine thiolate primed for attack onto the carbonyl oxygen of the peptide bond (Figure 16). Regardless of the route of

thiolate formation, this reactive nucleophile (pK_a = ~4.5) undergoes addition to the carbonyl center of the peptide bond to form a tetrahedral intermediate. Subsequent collapse of the tetrahedral intermediate is followed by concurrent pronation of the amine product. Then, activation of a water molecule by the active site histidine allows for its addition into the electron poor thioester-enzyme carbonyl center. Collapse of the resulting tetrahedral intermediate allows release of the peptide product and regenerates the enzyme to its thiol(ate)-imidazole/ium resting state.

One highly implemented strategy to block cysteine proteases is using peptidomimetic and non-peptidomimetic covalent inhibitors and inactivators. The design of such molecules is often predicated on the concept of installing an electrophilic "warhead" which can undergo covalent bond formation (reversible or irreversible) with the active site thiol(ate) of the protease. These warheads range from common Michael acceptors like α,β-unsaturated esters to, aldehydes, halomethyl ketones, epoxides, nitriles, and activated esters, to name a few.[125] While the determination of what warhead to employ and the design of new warheads for protease inhibitors is critical, an equally important facet is elucidation of the key determinants of binding specificity required for particular proteases. This thereby allows medicinal chemists to install a high level of selectivity into the warhead containing peptidomimetic or non-peptidomimetic scaffold.

Proteases contain specific binding pockets (S_2, S_1, S_1', S_2' etc.) that corresponds to a recognition residue (P_2, P_1, P_1', P_2'), wherein the scissile amide bond connects P_1 to P_1' (Figure 17).[126, 127] These binding pockets confer a high level of specificity for substrates, thus providing a specific recognition sequence for each enzyme.[128] Together,

the ability of scientists to produce molecules that leverage both the substate specificity conferred by the enzyme subsites and to target the active site nucleophilic thiol(ate) with electrophilic "warheads" has resulted in the successful development of potent inhibitors and inactivators against proteases.[129, 130]

In the case of the cysteine protease cruzain, the development of small molecule inhibitors and inactivators has been extensively pursued, but unfortunately none of the

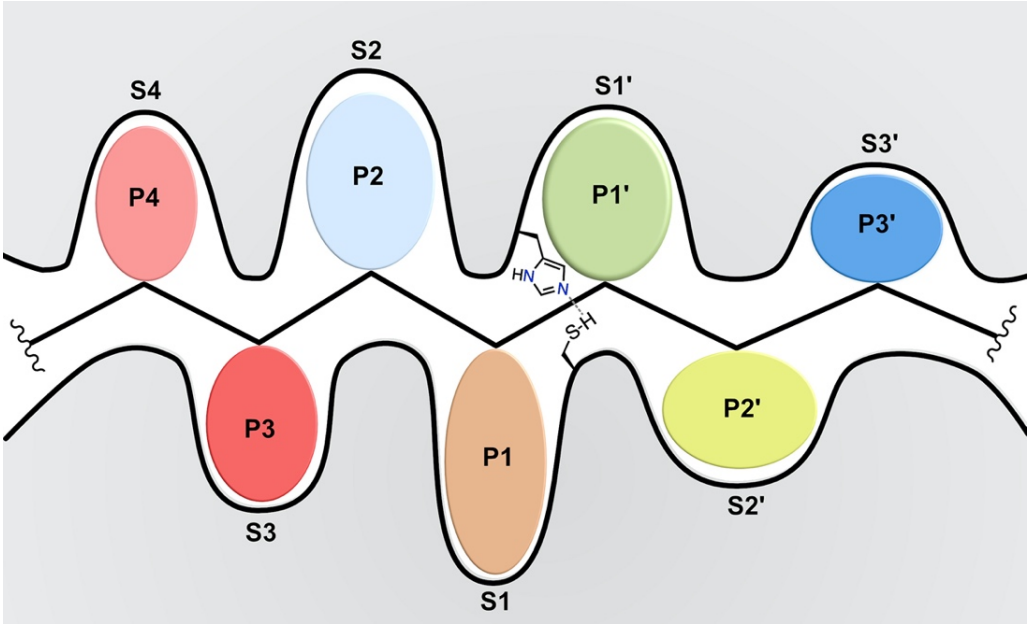


Figure 17. Generic Protease Active Site Nomenclature

Cartoon representation of the active site and binding pockets of a cysteine protease. Labeled are the position of the recognition residues (P) and binding pockets (S) in respect to Berger and Schechter nomenclature.

molecules identified to date have progressed to the clinic. These molecules have ranged in both their structure (peptidic and non-peptidic) and have also included a large variety of electrophilic warheads that allow reaction (reversible and irreversible) with the active site Cys_{25} of cruzain.[6, 35, 131-134] Albeit noncovalent competitive inhibitors of cruzain have

been identified using high throughput screening and virtual docking,[135, 136] still, the most successful molecule at targeting this enzyme is K777 (*N*-Me-pip-Phe-hPhe-VSPh), a peptidic molecule containing a vinyl sulfone warhead that undergoes irreversible thio-Michael addition with the active site Cys_{25} thiolate of cruzain.[132] Unfortunately, this molecule stumbled in preclinical trials for Chagas disease therapy as it induced vomiting in monkeys even at low dosages.[137] However, it was found that the HCl salt of the molecule acidified (pH = 3-4) the saline or water solutions it was dissolved in, likely causing the vomiting; but subsequent neutralization of the solution to neutral pH dissipated the vomiting.[137]

Of interest to us was the lack of specificity of K777 for cruzain versus human cathepsins and the irreversible nature of its covalent inactivation. It has been shown that K777 exerts nanomolar binding and robust irreversible inactivation of cathepsins B, K and particularly L.[138] This is because from a structural perspective cruzain is similar to that of human cathepsin L and exhibits similar substrate specificity to it.[139, 140] Both enzymes have a strong preference for charged residues such as arginine and lysine in their P1 site, but are also tolerant to hydrophobic groups. [139, 140] Therefore, if a molecule is a potent inhibitor or inactivator of cathepsin L it will also likely target cruzain. A recent example of this is with Gallinamide A, (2S,3S-N,N-dimethylisoleucine-Leu-Leu-Ala-vinyl-methylmethoxypyrrolinone), a natural product, which potently inactivates cathepsin L (K_i = 5 nM and $k_{inact}/K_i = 9 \times 10^5$) and tightly binds to ($IC_{50}$ = 0.3 nM) and likely inactivates cruzain.[134]

One reprieve in developing molecules for the selective targeting of cruzain may lie in the S2 subsite of the enzyme. This S2 subsite contains a glutamic acid (Glu_{208}) residue which has the ability to form hydrogen bonds with charged species and has been

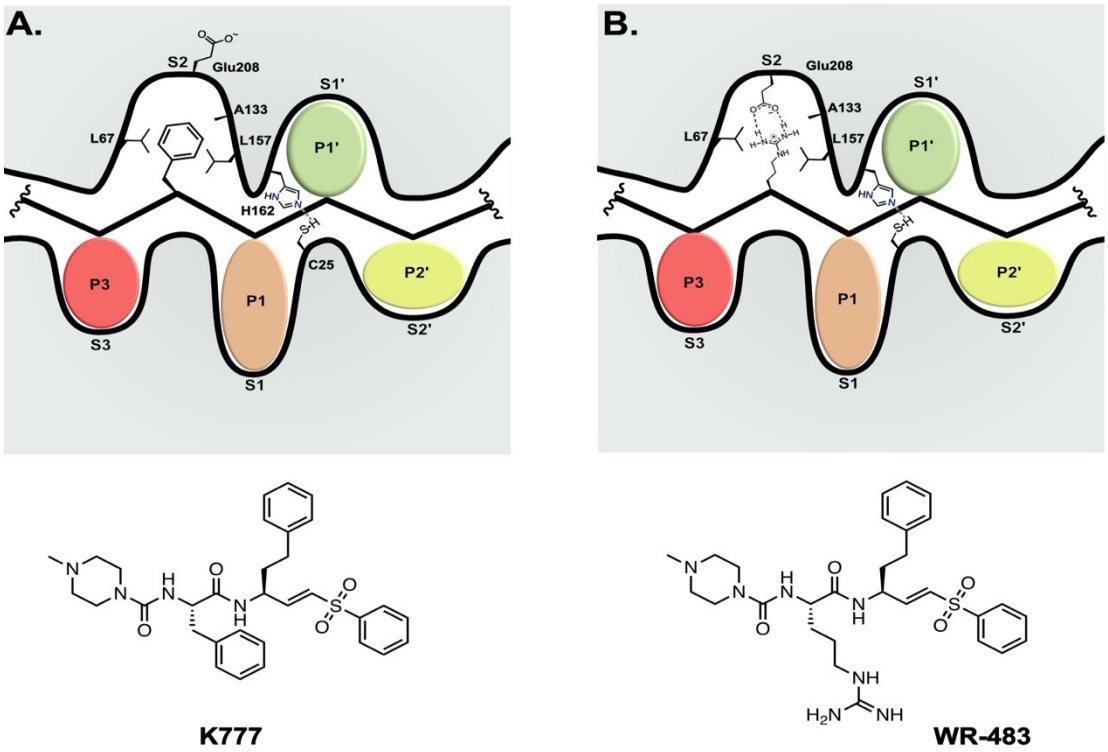


Figure 18. The S2 Subsite of Cruzain Provides Binding Specificity.

A. Cartoon binding pockets of cruzain whereby a phenylalanine binds to the S_2 subsite. Single letter amino acid codes as well as numbering of these residues correspond to the cruzain sequence in its mature form. Upon binding of a hydrophobic residue such as a phenylalanine as shown in **A.** Glu_{208} swings outwards towards solvent as the hydrophobic character of the Phe disallows interaction. **B.** Binding of an arginine in P_2 allows engagement of Glu_{208} via forms hydrogen bonds with the guanidinium group of the arginine residue.

coined a "specificity residue" (Figure 18).[141] While this Glu_{208} allows cruzain to accept both hydrophobic and basic charged amino acids, it still has a preference for hydrophobic residues owing to the predominant nonpolar character of this subsite (Figure 18).[128, 139]

Therefore, the acceptance of charged residues often results in an increased K_m, which has been postulated to partly be due to the energetic cost required to desolvate Glu_{208} when interacting with these basic residues.[141] Work from Chen *et al*. sought to modify the dipeptide scaffold of K777 to exploit this "specificity residue" by replacing the P_2 Phe of the molecule with a P_2 Arg. This substitution resulted in enhanced selectivity of the molecule (WRR-483) towards cruzain versus other human cathepsins,[142] but resulted in poor oral bioavailability and subsequent analogs failed to remedy this issue.[143] Therefore, we posit that the use of non-natural amino acids containing a mixture of hydrophobic and cationic character may assist in both minimizing the desolvation penalty of binding to the S_2 site and produce selective cruzain inhibitors. Towards this end, we sought to understand the kinetics involved in the binding and catalysis of dipeptidic molecules comprising varied polarities in the P_1 and P_2 positions and identify acceptable candidates to incorporate into inhibitor scaffolds. Herein, I detail the synthesis and characterization of a set of dipeptide substrates that assisted in deciphering an optimized peptide sequence in the development of a class of cruzain inhibitors which we coin peptidomimetic vinyl heterocycle inhibitors or PVHIs.

B. Experimental Methods

i. General Synthetic Chemistry Methods and Compound Characterization

All reagents and starting materials were obtained from commercial suppliers and used without further purification unless otherwise stated. Reactions were run under an atmosphere of nitrogen or argon and at ambient temperature unless otherwise noted. Reaction progress was monitored using thin-layer chromatography and by analysis employing an HPLC−MS (UltiMate 3000 equipped with a diode array coupled to an MSQ Plus single or ISQ-EM single quadrupole mass spectrometer, Thermo Fisher Scientific) using electrospray positive and negative ionization detectors. Reported liquid chromatography retention times (tR) were established using the following conditions: column: Phenomenex Luna 5µm C18(2) 100 Å, 4.6 mm, 50 mm; mobile phase A: water with 0.1% formic acid (v/v); mobile phase B: MeCN with 0.1% formic acid (v/v); temperature: 25°C; gradient: 0−100% B over 6 min, then a 2 min hold at 100% B; flow: 1 mL min^{-1}; and detection: MS and UV at 254, 280, 214, and 350 nm. Semi-preparative HPLC purification of compounds was performed on a Thermo Fisher Scientific UltiMate 3000 with a single wavelength detector coupled to a fraction collector. Purifications were conducted using the following conditions: column: Phenomenex Luna 5µmC18(2) 100 Å, 21.2 mm, 250 mm; mobile phase A: water with 0.1%formic acid (v/v); mobile phase B: MeCN with 0.1% formic acid (v/v); temperature: room temperature; gradient: 0−100% B over 30 min, then a 5 min hold at 100% B; flow: 20 mL min−1; and detection: UV (350 nm).$^{1}H/^{13}C$ NMR magnetic resonance spectra were obtained in $CDCl_3$,CD_3OD, or DMSO-d6 at 400 MHz/100 MHz at 298 K on a Bruker AVANCE III Nanobay console

with an Ascend magnet unless otherwise noted. The following abbreviations were utilized to describe peak patterns when appropriate: br = broad, s = singlet, d = doublet, q= quartet, t = triplet, and m = multiplet. All final compounds used for testing in assays and biological studies had purities that were determined to be >95% as evaluated by their proton NMR spectra and their HPLC/MS based on ultraviolet detection at 254 nm. Masses detected were in the range of 100−1000 Da and were detected in the positive or negative mode, depending on the ionization of the molecule.

ii. Synthetic Methods and Compound Characterization

Boc-L-homophenylalanine-AMC; tert-butyl (S)-(1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)carbamate. A round bottom flask charged with Boc-L-homophenylalanine (1.5 g, 5.64 mmol) and 7-amino-4-methylcoumarin (627.2 mg, 3.58 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (50 ml) and tetrahydrafuran (15 mL) was added followed by the addition of N,N-diisopropylethylamine (915 µL, 5.25 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (3.2 mL, 5.37 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was The reaction was allowed to react for 30 h and was monitored by TLC (3:1 hexanes:ethyl acetate) or LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~60 mL) and washed with water (25 mL, 4x) and brine (25 mL, 2x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield an off white solid. The crude amide directly used for the

next reaction without purification. LC-MS Rt: 5.59 min, *m/z* 437.3 ($[M+H]^+$, $C_{25}H_{28}N_2O_5^+$ Calcd 437.21

H_2N-L-homophenylalanine-AMC-TFA. (S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-aminium-TFA. A round bottom flask charged with Boc-L-homophenylalanine-AMC (60 mg, 0.137 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (6 mL) was added followed by the dropwise addition of trifluoroacetic acid (2 mL) at 0°C. The reaction was monitored via TLC (3:1 hexanes:ethyl acetate) and was complete following 1.5 h. The solvents were removed under reduced pressure and co-evaporated with chloroform (15 mL, 4x) and ether (15 mL, 2x) to yield a tan solid. LC-MS analysis verified the formation of the desired amine. The product was directly used for the next reaction without purification.

(S2) Cbz-L-phenylalanine-L-homophenylalanine-AMC. Benzyl ((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate. A round bottom flask charged with ^+H_3N-L-homophenylalanine-TFA (60 mg, 0.138 mmol) and Cbz-phenylalanine-OH (49.5 mg, 0.166 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (4 mL) was added followed by the addition of N,N-diisopropylethylamine (120 µL, 0.69 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (164 µL, 0.276 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min or until no starting material

remained via TLC (3:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~20 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude was purified via preparative HPLC to yield the desired amide (50.1 mg, 59% yield) as a white powder. ^{1}H NMR (400 MHz, DMSO) δ 1.91 – 2.17 (m, 2H), 2.41 (d, *J* = 1.2 Hz, 3H), 2.56 – 2.89 (m, 3H), 3.07 (dd, *J* = 13.7, 4.1 Hz, 1H), 4.30 – 4.60 (m, 2H), 4.84 – 5.11 (m, 2H), 6.28 (d, *J* = 1.5 Hz, 1H), 7.10 – 7.39 (m, 14H), 7.45 – 7.60 (m, 2H), 7.68 – 7.85 (m, 2H), 8.40 (d, *J* = 7.7 Hz, 1H), 10.46 (s, 1H). LC-MS Rt: 5.88 min, *m/z* 618.4 ($[M+H]^+$, $C_{37}H_{35}N_3O_6^+$ Calcd 618.3.

(S3) Cbz-L-leucine-L-homophenylalanine-AMC. benzyl ((S)-4-methyl-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxopentan-2-yl)carbamate. A round bottom flask charged with ^+H_3N-L-homophenylalanine • TFA (64 mg, 0.142 mmol) and Cbz-Leucine-OH (49.1 mg, 0.185 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (4 mL) and tetrahydrafuran (4 mL) was added followed by the addition of N,N-diisopropylethylamine (124 µL, 0.71 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (164 µL, 0.276 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min and no starting material remained via TLC (3:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~20 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic

layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was purified via preparative HPLC to yield (48.9 mg, 59%) as a white powder. ^{1}H NMR (400 MHz, DMSO) δ 0.90 (t, *J* = 6.9 Hz, 6H), 1.50 (t, *J* = 7.3 Hz, 2H), 1.68 (t, *J* = 7.0 Hz, 1H), 2.02 (d, *J* = 32.8 Hz, 2H), 2.41 (s, 3H), 2.54 – 2.79 (m, 3H), 4.15 (q, *J* = 7.8 Hz, 1H), 4.47 (s, 1H), 5.06 (s, 2H), 6.27 (s, 1H), 7.09 – 7.41 (m, 9H), 7.49 (t, *J* = 8.9 Hz, 2H), 7.64 – 7.86 (m, 2H), 8.25 (d, *J* = 7.7 Hz, 1H), 10.42 (s, 1H). LC-MS Rt: 5.88 min, *m/z* 584.3 ($[M+H]^+$, $C_{34}H_{37}N_3O_6^+$ Calcd 584.3

Boc-L-4-pyridylalanine-AMC. tert-butyl (S)-(1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-yl)carbamate. A round bottom flask charged with Boc-L-4-pyridylalanine-OH (325 mg, 1.22 mmol) and 7-amino-4-methylcoumarin (164.5 mg, 0.939 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (6 mL) and tetrahydrafuran (14 ml) was added followed by the addition of N,N-diisopropylethylamine (600 μL, 3.44 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (1.125 mL, 1.89 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 15 h and product formation was monitored via TLC (1% MeOH:DCM v/v) and LC-MS analysis. Solvents were removed under reduced pressure and the yellow oil was solubilized in DCM (~30 mL) and washed with water (15 mL, 4x) and brine (15 mL, 2x). The organic layer was dried over Na_2SO_4 and solvents were removed under reduced pressure. The resulting yellow solid was deemed pure enough (~94% via LC-MS, 425 mg) to proceed directly to the next step. LC-MS Rt: 2.84 min, *m/z* 424.26 ($[M+H]^+$, $C_{23}H_{25}N_3O_5^+$ Calcd 424.2

NH_2-L-4-pyridylalanine-AMC•TFA. (S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-aminium•TFA. A round bottom flask charged with Boc-L-4-pyridylalanine-AMC (150 mg, 0.353 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) was added followed by the dropwise addition of trifluoroacetic acid (3 mL) at 0°C. The reaction was monitored via TLC (1% MeOH:DCM v/v) and was complete following 1.5 h. The solvents were removed under reduced pressure and co-evaporated with chloroform and ether to yield a yellow solid. LCMS analysis verified the formation of the product. The product was directly used for the next reaction without purification.

(S5) Cbz-L-phenylalanine-L-4-pyridylalanine-AMC. Benzyl ((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate. A round bottom flask charged with $^{+}H_3N$-L-4-pyridylalanine ● TFA (154 mg, 0.352 mmol) and Cbz-phenylalanine-OH (137.3 mg, 0.459 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) and tetrahydrafuran (2 ml) was added followed by the addition of N,N-diisopropylethylamine (367.9 μL, 2.11 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (419.1 μL, 0.704 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min or until no starting material remained via TLC (1% MeOH:DCM v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~30 mL) and washed with water (15

mL, 4x) and brine (15 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. 100 mg of the crude was purified via preparative HPLC to yield (32.94 mg, 39%) as a white powder. ^{1}H NMR (400 MHz, DMSO) δ 2.33 – 2.45 (m, 3H), 2.72 (dd, *J* = 13.9, 10.5 Hz, 1H), 2.85 – 3.09 (m, 2H), 3.14 (dd, *J* = 13.9, 5.4 Hz, 1H), 4.14 – 4.37 (m, 1H), 4.68 – 4.89 (m, 1H), 4.96 (s, 2H), 6.29 (d, *J* = 1.5 Hz, 1H), 7.25 (ttd, *J* = 22.3, 14.7, 13.7, 6.4 Hz, 12H), 7.39 – 7.54 (m, 2H), 7.75 (d, *J* = 8.7 Hz, 2H), 8.25 – 8.56 (m, 3H), 10.52 (s, 1H). LC-MS Rt: 3.57 min, *m/z* 605.3 ($[M+H]^+$, $C_{35}H_{32}N_4O_6^+$ Calcd 605.23.

N-methyl-piperazine-L-phenylalanine-OH. (4-methylpiperazine-1-carbonyl)-L-phenylalanine. A round bottom flask charged with methyl (4-methylpiperazine-1-carbonyl)-L-phenylalanine (500 mg, 1.64 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous methanol (8.5 mL) and tetrahydrofuran (8.5 mL) was added followed by the addition of 0.5 M LiOH in H_20 (17 mL, 8.5 mmol). The reaction was monitored by TLC (5% MeOH:DCM v/v) and after 3.5 h the reaction was stopped and solvents were removed under reduced pressure. The reaction mixture was dissolved in acidified water (0.5 M HCl, 10 mL) to neutralize the unreacted LiOH and the solvent was removed under reduced pressure. The crude compound was then transferred to the next reaction step without purification.

(S7) N-methyl-piperazine-L-phenylalanine-L-homophenyl- alanine-AMC. 4-methyl-N-((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-

yl)amino)-1-oxo-3-phenylpropan-2-yl)piperazine-1-carboxamide. A round bottom flask charged with $^{+}H_3N$-L-homophenylalanine • TFA (105 mg, 0.313 mmol) and N-methyl-piperazine-phenylalanine-OH (160 mg, 0.55 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) was added followed by the addition of N,N-diisopropylethylamine (274 μL, 1.57 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (377 μL, 0.626 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min and no starting material remained via TLC (1:1 ethyl acetate: hexanes v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~20 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was purified via preparative HPLC to yield (21.1 mg, 11%) of a white powder. ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.94 – 2.16 (m, 6H), 2.31 (s, 4H), 2.32 – 2.41 (m, 5H), 2.42 (d, *J* = 1.2 Hz, 4H), 2.73 (t, *J* = 7.4 Hz, 2H), 2.85 – 3.04 (m, 2H), 3.25 (ddd, *J* = 12.7, 10.6, 4.9 Hz, 3H), 3.41 (ddt, *J* = 12.5, 5.6, 2.9 Hz, 3H), 4.36 (dt, *J* = 8.7, 5.3 Hz, 1H), 4.60 (td, *J* = 8.5, 4.5 Hz, 1H), 4.78 (d, *J* = 4.4 Hz, 1H), 6.20 (d, *J* = 1.4 Hz, 1H), 6.60 (d, *J* = 8.2 Hz, 1H), 7.12 – 7.40 (m, 10H), 7.52 (d, *J* = 8.5 Hz, 1H), 7.68 – 7.82 (m, 2H), 9.16 (s, 1H). LC-MS Rt: 3.29 min, *m/z* 610.1 ($[M+H]^+$, $C_{35}H_{39}N_5O_5^+$ Calcd 610.3.

Boc-L-4-pyridylalanine-L-homophenylalanine-AMC. tert-butyl ((S)-1-(((S)-1-((4-methyl-2-oxo-2H- chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-yl)carbamate. A round bottom flask charged with $^{+}H_3N$-L-

homophenylalanine • TFA (146 mg, 0.325 mmol) and Boc-L-4-pyridylalanine-OH (104 mg, 0.39 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) was added followed by the addition of N,N-diisopropylethylamine (283 μL, 1.63 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (387 μL, 0.65 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min and no starting material remained via TLC (1:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~20 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was moved directly to the next step with no purification. LC-MS Rt: 3.57 min, *m/z* 581.7 ($[M+H]^-$, $C_{33}H_{36}N_4O_6^-$ Calcd 581.3.

NH_2-L-4-pyridylalanine-L-homophenylalanine-AMC•TFA. S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-aminium • TFA. A round bottom flask charged with Boc-L-4-pyridylalanine-L-homophenylalanine-AMC (172 mg, 0.294 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) was added followed by the dropwise addition of trifluoroacetic acid (5 mL) at 0°C. The reaction was monitored via TLC (1:1 hexanes:ethyl acetate) and was complete following 1.5 h. The solvents were removed under reduced pressure and co-evaporated with chloroform (15 mL, 4x) and ether

(15 mL, 2x) to yield a tan solid. LC-MS analysis verified the formation of the desired amine. The product was directly used for the next reaction without purification.

(S8) Cbz-L-4-pyridylalanine-L-homophenylalanine-AMC. Benzyl ((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxo-3-(pyridin-4-yl)propan-2-yl)carbamate. A round bottom flask charged with NH_2- L-4-pyridylalanine-L-homophenylalanine-AMC-TFA (227 mg, 0.379 mmol; excess weight is due to residual TFA salts) was purged with N_2 gas yielding positive pressure. Anhydrous ethanol (4 mL) was added followed by the addition of a solution of sodium bicarbonate (127 mg, 1.52 mmol) in water (4 mL). Subsequently, a solution of benzyl chloroformate (107 uL, 0.76 mmol) in 1,4 dioxanes (4 mL) was added dropwise and the reaction was allowed to react overnight. Solvents were removed under reduced pressue and the reaction mixture was solubilized in dichloromethane (30 mL) and washed with H_2O (10 mL, 3x), brine (10 mL, 2x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield an orange oil. The crude amide was purified via preparative HPLC to yield (45.9 mg, 20%) as a white powder. 1H NMR (400 MHz, MeOD) δ 1.88 – 2.27 (m, 2H), 2.44 (d, *J* = 1.3 Hz, 3H), 2.68 (ddddd, *J* = 20.0, 15.3, 12.7, 7.8, 3.3 Hz, 2H), 2.96 (dd, *J* = 14.1, 9.2 Hz, 1H), 3.22 (dd, *J* = 13.9, 5.3 Hz, 1H), 4.50 (dd, *J* = 8.8, 5.3 Hz, 1H), 4.57 (dd, *J* = 9.1, 5.3 Hz, 1H), 5.05 (d, *J* = 3.5 Hz, 2H), 6.21 (d, *J* = 1.5 Hz, 1H), 7.01 – 7.37 (m, 12H), 7.47 (dd, *J* = 8.7, 2.1 Hz, 1H), 7.55 – 7.72 (m, 1H), 7.78 (d, *J* = 2.0 Hz, 1H), 8.33 (d, *J* = 5.1 Hz, 2H).LC-MS Rt: 3.79 min, *m/z* 617.5 ($[M+H]^-$, $C_{36}H_{34}N_4O_6^-$ Calcd 617.3.

Boc-L-phenylalanine-AMC. tert-butyl (S)-(1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate. A round bottom flask charged with Boc-L-phenylalanine (380 mg, 1.43 mmol) and 7-amino-4-methylcoumarin (193 mg, 1.1 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous tetrahydrafuran (30 mL) was added followed by the addition of N,N-diisopropylethylamine (958 µL, 5.5 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (1.31 mL, 2.2 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 15 h was monitored by TLC (3:1 hexanes:ethyl acetate) or LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~80 mL) and washed with water (30 mL, 4x) and brine (30 mL, 2x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was purified via preparative HPLC (363 mg, 78%) LC-MS Rt: 5.31 min, *m/z* 423.3 ($[M+H]^+$, $C_{24}H_{26}N_2O_5^+$ Calcd 423.2.

H_2N-L-phenylalanine-AMC•TFA. (S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-phenylpropan-2-aminium•TFA. A round bottom flask charged with Boc-L-phenylalanine-AMC (40 mg, 0.095 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (3 mL) was added followed by the dropwise addition of trifluoroacetic acid (1 mL) at 0°C. The reaction was monitored via TLC and was complete following 1.5 h. The solvents were removed under reduced pressure and co-evaporated with chloroform (15 mL, 4x) and ether (15 mL, 2x) to yield a tan solid. LCMS

analysis verified the formation of the desired amine. The product was directly used for the next reaction without purification.

(S9) Cbz-L-phenylalanine-L-phenylalanine-AMC. benzyl ((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-phenylpropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate. A round bottom flask charged with $^{+}H_3N$-L-phenylalanine •TFA (40 mg, 0.092 mmol) and Cbz-phenylalanine-OH (33 mg, 0.11 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (5 mL) and was added followed by the addition of N,N-diisopropylethylamine (80 μL, 0.46 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (110 μL, 0.18 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min and no starting material remained via TLC (3:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in dichloromethane (~20 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was purified via preparative HPLC to yield (34.2 mg, 62%) as a white powder. ^{1}H NMR (400 MHz, DMSO) δ 2.41 (s, 3H), 2.72 (dd, J = 13.9, 10.5 Hz, 1H), 2.87 – 3.05 (m, 2H), 3.11 (dd, J = 13.8, 5.8 Hz, 1H), 4.30 (td, J = 9.6, 9.0, 4.2 Hz, 1H), 4.73 (q, J = 7.5 Hz, 1H), 4.96 (s, 2H), 6.28 (s, 1H), 7.06 – 7.39 (m, 15H), 7.39 – 7.53 (m, 2H), 7.60 – 7.79 (m, 2H), 8.37 (d, J = 7.8 Hz, 1H), 10.47 (s, 1H). LC-MS Rt: 5.68 min, *m/z* 603.4 ($[M+H]^+$, $C_{36}H_{33}N_3O_6^+$ Calcd 604.2.

Cbz-L-arginine(Mtr)-L-homophenylalanine-AMC. Benzyl ((S)-5-(3-((4-methoxy-2,6-dimethylphenyl)sulfonyl)guanidino)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxopentan-2-yl)carbamate. A round bottom flask charged with $^{+}H_3N$-L-homophenylalanine • TFA (120 mg, 0.266 mmol) and Cbz-Arginine (Mtr)-OH (180 mg, 0.346 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (10 mL) and was added followed by the addition of N,N-diisopropylethylamine (232 µL, 1.33 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (317 µL, 0.532 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 3 h and no starting material remained via TLC (3:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in DCM (~40 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid that was moved directly to the next step. LC-MS Rt: 5.73 min, *m/z* 837.9 ($[M+H]^-$, $C_{43}H_{48}N_6O_9S^-$ Calcd 837.3.

(S10) Cbz-L-arginine-L-homophenylalanine-AMC. benzyl ((S)-5-guanidino-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-4-phenylbutan-2-yl)amino)-1-oxopentan-2-yl)carbamate. A round bottom flask charged with Cbz-Arginine (Mtr)-L-homophenylalanine-AMC (130 mg, 0.158 mmol) and phenol (450 mg, 4.8 mmol, 5% wt/v) was purged with N_2 gas yielding positive pressure. Trifluoroacetic acid (6 mL) and was added followed by allowing the reaction to proceed for 12 h. The solvent was removed under a flow of N_2 gas and the crude product was subsequently purified by preparative

HPLC to yield (34 mg, 34% yield) as a white powder. ^{1}H NMR (400 MHz, MeOD) δ 1.58 – 2.00 (m, 4H), 2.00 – 2.33 (m, 2H), 2.34 – 2.53 (m, 3H), 2.52 – 2.84 (m, 2H), 3.20 (q, *J* = 6.3, 5.8 Hz, 2H), 4.23 (t, *J* = 6.5 Hz, 1H), 4.50 (dd, *J* = 9.2, 4.9 Hz, 1H), 5.13 (s, 2H), 6.23 (t, *J* = 5.5 Hz, 1H), 6.90 – 7.41 (m, 10H), 7.47 (d, *J* = 8.7 Hz, 1H), 7.66 (dt, *J* = 18.0, 8.3 Hz, 1H), 7.84 (d, *J* = 11.3 Hz, 1H), 8.53 (s, 1H). LC-MS Rt: 3.62 min, *m/z* 625.5 ($[M+H]^-$, $C_{34}H_{38}N_6O_6^-$ Calcd 625.3.

(S11) NMe-Pip-Phe-Phe-AMC. 4-methyl-N-((S)-1-(((S)-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-3-phenylpropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)piperazine-1-carboxamide. A round bottom flask charged with $^+H_3$N-L-phenylalanine •TFA (100 mg, 0.29 mmol) and N-methyl-piperazine-phenylalanine-OH (160 mg, 0.55 mmol) was purged with N_2 gas yielding positive pressure. Anhydrous dichloromethane (15 mL) and was added followed by the addition of N,N-diisopropylethylamine (253 µL, 1.45 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (345 µL, 0.58 mmol) was added dropwise at 0°C resulting in a yellow color. The reaction was allowed to react for 45 min and no starting material remained via TLC (1:1 hexanes:ethyl acetate v/v) and LC-MS analysis. Solvents were removed and the yellow oil was solubilized in dichloromethane (~40 mL) and washed with water (10 mL, 4x) and brine (10 mL, 1x). The organic layer was dried over Na_2SO_4 and solvents removed under reduced pressure to yield a yellow solid. The crude amide was purified via preparative HPLC to yield (24.6 mg, 14%) ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.38 (d, *J* = 1.2 Hz, 3H), 2.54 (s, 3H), 2.55 – 2.78 (m, 4H), 2.81 – 3.04 (m, 1H), 3.02 – 3.28 (m, 3H), 3.32 – 3.60

(m, 4H), 4.47 (dt, J = 8.7, 5.5 Hz, 1H), 4.87 (q, J = 6.9 Hz, 1H), 5.65 (d, J = 5.7 Hz, 1H), 6.16 (d, J = 1.4 Hz, 1H), 6.83 – 7.40 (m, 15H), 7.46 (d, J = 8.7 Hz, 1H), 7.54 – 7.72 (m, 2H), 8.30 (s, 1H), 9.45 (s, 1H). LC-MS Rt: 3.15 min, *m/z* 596.2 ($[M+H]^+$, $C_{34}H_{37}N_5O_5^+$ Calcd 596.3.

iii. Enzyme Assays and Substrate Kinetics

All enzyme assays were performed at 25 °C. Initial rates of the peptidolytic reaction catalyzed by cruzain were measured by monitoring the fluorescence generated by cleavage of the dipeptide-AMC bond. Assays were conducted in 96-well plates (Greiner, flat-bottom, clear black plates) in a total volume of 250 μL, containing either 50 mM MES (pH 7.5), 50 mM TAPSO, 100 mM DEA, 1 mM CHAPS, 1 mM Na_2EDTA, 5 mM DTT and 10% DMSO (v/v) or 50 mM sodium acetate (pH 5.5), 50 mM MES, 100 mM TEA, 1 mM CHAPS, 1 mM Na_2EDTA, 5 mM DTT and 10% DMSO (v/v). Substrates were dissolved in 100% DMSO, and were then diluted 10-fold such that when added to reaction mixtures, final DMSO concentration were 10% (v/v). Reactions were initiated with addition of 25 μL of cruzain (final concentrations: 0.1 – 0.4 nM). Fluorescence was measured on either a SpectraMax M5 (Molecular Devices) or a Synergy HTX (Biotek, Wisnooki, VT) microplate reader (λ_{ex} = 360 nm, λ_{em} = 460 nm). Initial rates were determined from continuous kinetic time courses, and calculated from the earliest time points, typically at less than 10 min.

Replotting of the initial velocity data versus substrate concentration resulted in hyperbolic plots which were fitted to eq. 1:

$$v = \frac{V_{max}\,[S]}{K_m + [S]} \qquad \textbf{eq. 1}$$

Where V_{max} is the maximal velocity achieved, [S] is the molar concentration of the substrate, and K_m is the substate concentration at which the velocity is half V_{max}. Thereby allowing determination of the rate of catalysis, or the k_{cat} using eq. 2:

$$V_{max} = k_{cat}\,[E_t] \qquad \textbf{eq. 2}$$

Where, k_{cat} is the number of molecules made per enzyme per a unit of time and $[E_t]$ is the molar concentration of enzyme in the assay conditions.

C. Results and Discussion

i. Design and Synthesis of Dipeptide-AMC Substrates for Cruzain

In an effort to understand the tolerability for cruzain to bind nonnatural amino acid containing substrates which may assist in inhibitor design, we prepared a small and targeted dipeptide library for kinetic analysis. These peptides contained a C-terminal fluorophore, 7-amino-4-methylcoumarin, which upon cleavage by the protease yields a robust fluorescent signal (excitation: 360 nm, emission: 460 nm). As cruzain can tolerate both hydrophobic and hydrophilic amino acids in its P2 subsite (Figure 19) we envisioned

that the development of a substrate containing a mixture of hydrophobic and hydrophilic character may be best suited to bind to the P2 site. Therefore, we proposed that the nonnatural amino acid 4-pyridyl-alanine((4-Pyr)Ala) may comprise an ideal candidate for this P2 site. In addition, we sought to identity the kinetic behavior of the peptidic scaffold of the known potent covalent inactivator of cruzain, K777, to see how it compares to other substrates produced. We employed solution phase synthesis to produce the substrates, whereby initial coupling of the aniline of 7-amino-4-methyl coumarin to the carboxylic acid of varied N-Boc protected amino acids using T_3P (propylphosphonic anhydride) in a mixture of DCM/THF with N,N-diisopropylethylamine (DIEA) afforded robust formation of the desired Boc-AA-AMC molecule. Subsequent deprotection (30% v/v trifluoroacetic

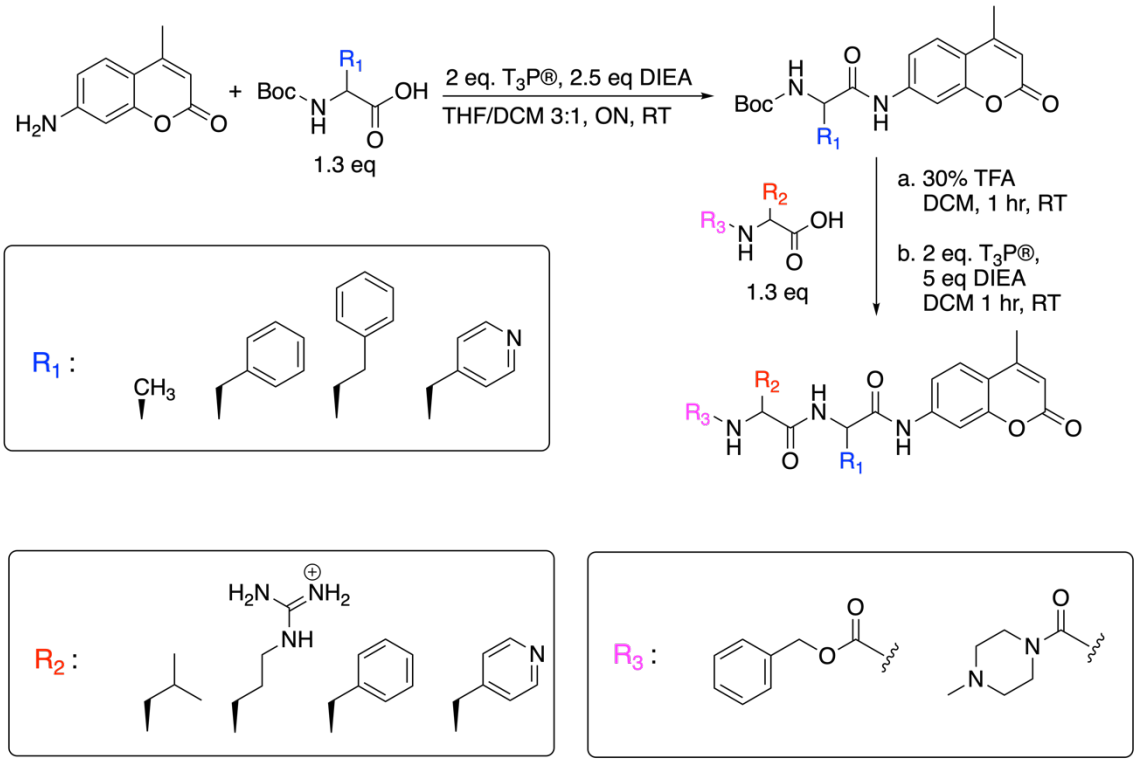


Figure 19. General Synthesis of Fluorogenic Dipeptide Substrates for Cruzain

acid in dichloromethane (DCM)) was followed by coupling the desired Cbz-AA-OH or N-MePip-AA-OH using T_3P in DCM with DIEA to result in the formation of the desired product(s) which were purified to homogeneity (>95%) with semi-preparative HPLC.

ii. Evaluation of Dipeptide Substrates

We previously conducted a detailed investigation of the catalytic mechanism of cruzain utilizing three dipeptide substrates and by using solvent kinetic isotope effects in both pre-steady state and steady state kinetic modes.[124] As with other cysteine proteases, our kinetic studies supported a double-displacement chemical mechanism in which a thioester intermediate of Cys_{25} and substrate is first formed (enzyme acylation) concurrent with the release of the amine product. This half-reaction is followed by enzyme-catalyzed hydrolysis of the thioester (enzyme deacylation) to afford the carboxylate product. We determined that for Cbz-Phe-Arg-AMC (Table 1. Kinetic Parameters of Peptide Substrates for Cruzaina) (most optimal substrate in terms of values of k_{cat} and k_{cat}/K_m) and Cbz-Arg-Arg-AMC (S6) enzyme-catalyzed deacylation comprised the slow steps of the reaction ($k_{cat} = k_{dac}$), while the rates of the acylation of active site Cys_{25} (described by k_{ac}) were 10-fold and 4-fold, respectively, more rapid than k_{cat}.[124] The parameter k_{cat}/K_m contains the individual rate constants found in the expression for k_{ac} and therefore reports on the rates of substrate binding and catalysis up to and including the acylation of active site Cys_{25}. Accordingly, this kinetic parameter comprises a useful guide for the selection of optimal dipeptide scaffolds to be incorporated into inhibitors that are meant to form reversible covalent complexes with Cys_{25} upon binding, a reaction that chemically mimics acylation.

If our dipeptidic vinyl heterocyclic inhibitors do indeed inhibit cruzain by the addition of the thiolate of Cys_{25} to their olefinic replacements for the scissile amide bond, then the values of k_{cat}/K_m (the specificity constants) for the fluorogenic substrates should provide an approximate linear correlation with values of k_{inact}/K_I (the "specificity constant" of inactivation).[7, 144]

We have kinetically characterized 12 dipeptide fluorogenic substrates of cruzain, which are of the form Cbz/NMePip-P_2-P_1-AMC, for which Cbz is benzyloxycarbonyl; NMePip is N-methyl-piperazinyl; P_2 = Phe, Leu, Arg, or 4-pyridyl-alanine((4-Pyr)Ala); P_1= Phe, hPhe (homoPhe), Arg, Ala, and (4-Pyr)Ala; and AMC is 7-amino-4-

Table 1. Kinetic Parameters of Peptide Substrates for Cruzain[a]

Substrate	Peptide	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1} s^{-1}$)	$k_{cat}/K_{m\,rel}$	$k_{cat\,rel}$
S1	Cbz-Phe-Arg-AMC	9.6 ± 0.2	0.89 ± 0.09	11 ± 1	1.0	1.00
S2	Cbz-Phe-hPhe-AMC	2.0 ± 0.2	0.26 ± 0.06	8 ± 2	0.730	0.21
S3	Cbz-Leu-hPhe-AMC	3.0 ± 0.2	0.8 ± 0.2	4 ± 1	0.36	0.21
S4	Cbz-Leu-Arg-AMC	8.4 ± 0.4	2.2 ± 0.3	3.8 ± 0.7	0.35	0.88
S5	Cbz-Phe-4-PyrAla-AMC	7.5 ± 0.5	3.7 ± 0.5	2.0 ± 0.4	0.18	0.76
S6	Cbz-Arg-Arg-AMC[b]	7.2 ± 0.1	3.7 ± 0.4	1.9 ± 0.2	0.17	0.75
S7	NMe-Pip-Phe-hPhe-AMC	4.0 ± 0.1	3.1 ± 0.3	1.3 ± 0.1	0.12	0.42
S8	Cbz-4-PyrAla-hPhe-AMC	5.3 ± 0.2	4.6 ± 0.5	1.1 ± 0.1	0.1	0.55
S9	Cbz-Phe-Phe-AMC	0.26 ± 0.01	0.34 ± 0.06	0.8 ± 0.2	0.07	0.03
S10	Cbz-Arg-hPhe-AMC	4.9 ± 0.1	6.8 ± 0.4	0.7 ± 0.04	0.06	0.51
S11	NMePip-Phe-Phe-AMC	2.8 ± 0.06	14 ± 4	0.2 ± 0.04	0.02	0.29
S12	Cbz-Phe-Ala-AMC[b]	0.89 ± 0.06	38 ± 2	0.023 ± 0.003	0.002	0.090

[a]Data were obtained at pH 7.5, 25°C, 10% (v/v) DMSO. [b]Data obtained at 2% as reported in[35]

methylcoumarin (Table 1). The largest specificity constant (k_{cat}/K_m) measured was that of Cbz-Phe-Arg-AMC (11 $\mu M^{-1}s^{-1}$, relative k_{cat}/K_m = 1.0). Substrates S2−S4 have values of k_{cat}/K_m, which are 73−35% of that of Cbz-Phe-Arg-AMC, indicating that Cbz-Phe-hPhe-AMC, Cbz-Leu-hPhe-AMC, and Cbz-Leu-Arg- AMC all comprise highly competent substrates likely effecting rapid acylation of cruzain. The poorest substrate, Cbz-Phe-Ala-

AMC (S12), demonstrated that a small side chain in the P_1 position is less preferable than more bulky hydrophobic moieties. Conversely, the dipeptide Cbz-Phe-hPhe-AMC (S2) exhibited a value of k_{cat}/K_m = 8 μM^{-1} s^{-1}, indicating that a peptide substrate with the same P_2 and P_1 amino acids as K777 is an excellent substrate of cruzain and likely involves rapid acylation. Interestingly, when the Cbz group of substrate S2 is replaced with NMePip (S7), a substrate mimic of K777, k_{cat}/K_m is 8-fold lower (1.3 μM^{-1} s^{-1}). However, its value of k_{cat} exceeded that of Cbz-Phe-hPhe-AMC (S2), suggesting that the NMePip N-terminus of S7 acts to retard cruzain acylation in comparison to the Cbz group of S2. While Cbz-Phe-Phe-AMC (S9) had a value of k_{cat}/K_m that was only 7% of Cbz-Phe-Arg-AMC, its K_m value of 0.34 μM indicated potent binding, albeit with slow turnover. This may indicate that a P_1 Phe substitution leads to favorable binding, but this side chain impedes either the acylation or deacylation step in catalysis. This was observed for dipeptide−AMC substrates for the papain-like protease, human cathepsin C.[145] We explored the viability of the (4-Pyr)Ala residue as a potential mimic of both Phe and Arg, for the latter residue when the pyridine is protonated. Cbz-Phe-(4-Pyr)Ala-AMC (S5) was found to be a good substrate (relative k_{cat}/K_m = 0.10) but with a higher value of K_m (3.7 μM) compared to substrates containing Phe and hPhe in the P_1 residue. To leverage the ability of cruzain to tolerate both charged basic and hydrophobic residues in the P_2, position, we prepared Cbz-(4-Pyr)Ala-hPhe-AMC and Cbz-Arg-hPhe-AMC (S8 and S10). These substrates displayed efficient turnover numbers (k_{cat}= 5.3 and 4.9 s^{-1}, respectively) but poor values of K_m (4.6 and 6.8 μM) in comparison to Cbz-Phe-Phe/hPhe-AMC (0.26−0.34 μM). This demonstrates that the enzyme has a strong preference for

substrates that contain a hydrophobic P_2 residue, but that the binding of basic aromatic residues to the P_2 site is preferred versus the guandinium of arginine. Together, these results indicated that dipeptide scaffolds in which P_2= Cbz-Phe or NMePip-Phe and P_1=Arg, hPhe, or (4-Pyr)Ala may be among the best to incorporate into the vinyl heterocyclic framework.

iii. Translation of Dipeptide Substrates into Inhibitors

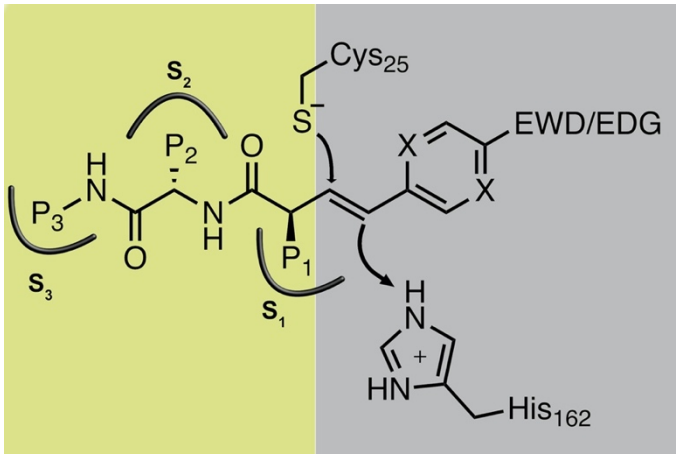


Figure 20. Two-pronged Inhibitor Design Strategy.

Gold: development of selective peptide scaffold (P_1-P_3) for binding to protease sub-sites (S_1-S_3); Grey: optimization of tuned vinyl heterocycle with electron donating and withdrawing groups.

Using the information gleaned in our dipeptide survey we produced heterocyclic inhibitors of cruzain (prepared by Dr. Bala Chenna and Dr. Linfeng Li) which comprise two portions: 1. a dipeptidic scaffold and 2. varied heterocycles adjacent to a vinyl group to tune the reactivity of the vinyl, which, in theory, should allow for tunable thio-Michael

addition with the active site Cys_{25} of cruzain (Figure 20). The discussion herein focuses solely on the dipeptidic scaffold and the role it plays in the potency (K_i^*) of vinyl heterocyclic inhibitors.

Analysis of the inhibition constant(s) for inhibitors containing a vinyl-2-pyridine warhead and varied P_1, P_2, P_3 substituents to dipeptide AMC substrates with the same P_1, P_2, and P_3 substitutions resulted in a near linear correlation (0.93 R^2) between the K_m and the K_i^* of these molecules (Figure 21). Close analysis revealed that in an analogous fashion to the substrate studies that the hydrophobic character of a P_2 residue was a key driver of potency, and this potency translated nicely from the dipeptide substrates to the inhibitors. For example, the most potent inhibitor in the class of molecules analyzed, Z-F-hF-vinyl-2-pyridine, bound a mere 1.5-fold tighter than its substrate counterpart Z-F-hF-AMC (S2) (K_i^* = 0.17 μM versus K_m = 0.26 μM). Analysis of dipeptide scaffolds containing polar P_2 moieties bound on average ~10-fold weaker than that of their hydrophobic counterparts. However, some P_2 polar substituents, such as arginine, exhibited a large decrease in the binding of the inhibitor (K_i^* = <100 μM), which, when compared to the most potent molecule analyzed here (Z-F-hF-vinyl-2-pyridine), bound <100-fold worse. Conversely, when comparing the substrate version of this inhibitor, Z-R-hF-AMC to Z-F-hF-AMC we noticed a 26-fold increase in K_m. While this change was still significant, the magnitude of change seen between the two inhibitors was much greater, which may suggest that the binding of Z-R-F-AMC to cruzain may be altered compared to the Z-R-hF-vinyl-2-pyridine inhibitor.

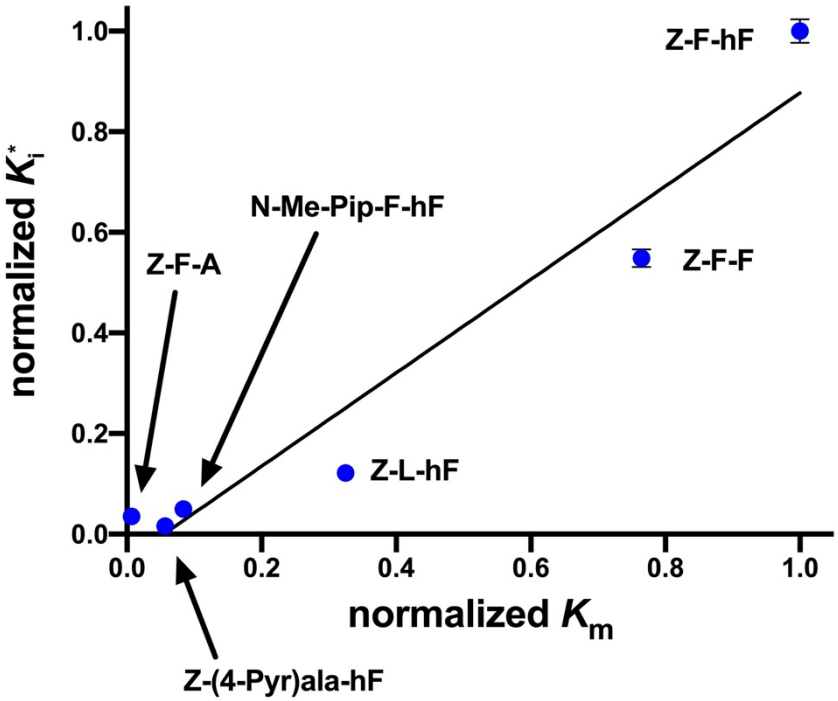


Figure 21. Comparison of K_m Versus K_i* Based on Dipeptide Scaffold

However, by in large, we saw small perturbations between the values of and K_i^* and K_m (~3-fold), suggestive that our substrate survey provided an excellent metric to guide our inhibitor design efforts. Together, this work highlights the importance of having a sound understanding of the substrate specificity of proteolytic enzymes prior to embarking on a mechanism-based drug discovery campaign.

D. Conclusions and Future Outlook

The design of mechanism-based inhibitors and inactivators for key proteolytic enzymes strongly relies on a deep understanding of not only the catalytic mechanism of the enzyme, but also substrate specificity. In the case of cruzain, the P_2 subsite of the enzyme contains a glutamate residue (Glu_{208}) that results in its tolerance for charged groups, which may lend useful in developing more selective inhibitors of this enzyme. We designed, synthesized, and characterized cruzain substrates and corresponding inhibitors containing varied S_1, S_2 and S_3 positions in an effort to further understand the substrate specificity of this enzyme and how it relates to inhibitor design. We identified that use of polar S_2 residues, even those with mixed polar and aromatic character, resulted in a marked decrease in the binding for substrates (K_m) and inhibitors (K_i^*) of cruzain. While these polar S_2 moieties may provide enhanced selectivity for cruzain, this selectivity comes at a penalty of decreased binding affinity. In addition, we also identified that the correlation between inhibitor potency and the Michalis constant for our substrates aligned nicely. Together this informed us that in order to design potent peptidic, mechanism-based inhibitors/inactivators of cruzain, or possibly other proteases, it may be ideal to initially identify and optimize a peptidic scaffold that complexes tightly to the enzyme followed by conversion of that molecule to an inhibitor or inactivator. Therefore, future studies for development of peptidic cruzain inhibitors/inactivators should focus on developing more tailored libraries of non-natural dipeptide substrates that can identify other residues that can aid in enhancement of substrate binding and also take advantage of the Glu_{208} residue

in the S_2 subsite of cruzain via use of residues with varied polarities and structural character.

While cruzain and many other proteases have been screened by libraries consisting of 10's of thousands and in some cases 100's of thousands of peptides, these studies can still miss important information. For example, in the work presented herein, the best peptide in terms of k_{cat} and k_{cat}/K_m is Z-F-R-AMC. This molecule would be easily identified as a top contender for inhibitor design in a canonical fluorescent readout assay with thousands of molecules as its catalytic rate is quite robust (9.6 sec^{-1}) and it binds to the enzyme well (0.9 μM). However, in this same assay one would not likely observe Z-F-F-AMC as a top contender because of its poor k_{cat} (0.3 sec^{-1}) although it binds very well to the enzyme target (0.34 μM). In the case presented herein, our linear correlation between K_m and K_i^* suggests that the ideal candidate for further drug elaboration would be the Z-F-F- scaffold, not the Z-F-R scaffold. Thus, it would be important to develop assays that would allow one to not only identify which peptide substrate gives the most robust signal readout, but also which binds to the enzyme the most effectively.

This would require one to develop more targeted combinatorial libraries of peptides using solid phase peptide synthesis to provide access to 100s or 1,000s of different combinations of natural and unnatural dipeptides instead of 100's of thousands. This is for two main reasons: 1. synthetic strategy and 2. assay design. With respects to synthesis, the development of combinatorial libraries is undoubtedly a powerful approach. However, when one is producing 10's or 100's of thousands of molecules the purification/characterization of each individual molecule becomes impractical, if not

impossible. While solid-phase peptide synthesis, especially of small di- and tri-peptides, generally produces a mostly pure molecule of interest, the yields (particularly isolated) are often low. This can make it challenging for one to obtain sufficient amounts of materials to conduct a kinetic study with multiple concentrations or in most cases know the precise concentration of the substrate molecule that one puts into an assay well. Further, when incubating a fixed concentration of a molecule with a fixed concentration of enzyme one will preferentially select substrates which have high rates of turnover, but not necessarily ones that bind tightly to the enzyme. Therefore, while it is important to know how catalytically robust a substrate is for a particular enzyme, we may need to employ an alternate approach to identify substrates that bind tightly to the enzyme. Ideally, this approach would entail the development of a smaller library (100-1000 substrates containing a variety of natural and unnatural amino acids) where one now has the ability to feasibly produce higher quantities of substrates and possibly validate their purity via LCMS. Together, these higher quantities and knowledge of relatively peptide purity would allow one to vary the concentrations of the substrates in the assays and therefore obtain a rapid assessment of k_{cat}, K_m, and k_{cat}/K_m for each substate. This smaller library could then be counter screened against proteases with similar substrate specificity, such as cathepsin L. This type of tailored library approach could prove powerful in the development of new and selective scaffolds for cruzain inhibitor design.

CHAPTER III CIS-EPOXYSUCCINATE IS A MECHANISM-BASED INACTIVATOR OF ISOCITRATE LYASE FROM MYCOBACTERIUM TUBERCULOSIS*

A. Introduction

i. Mycobacterium tuberculosis: Brief History of Disease and Research

At present, 1 out of every 4 people in the world are infected with *Mycobacterium tuberculosis (Mtb)* and each year there are about 10 million new cases and 1.5+ million deaths.[146] This obligate human pathogen is the causative agent of tuberculosis (TB) and has been one of the world's deadliest pathogens for thousands of years, despite the fact that only ~5-10% of infected individuals progress into active and readily contagious TB infection.[146] Modern strains of *Mtb* can be traced back nearly 15,000 years, but earlier progenitor stains likely existed as early as 3 million years ago,[147] infecting early humans and their ancestors (hominids).[148] Some of the first documentation of TB infection comes from early Egyptian art depicting tuberculosis-derived legions,[149] which was corroborated by the isolation of *Mtb* DNA and cell wall lipids from Egyptian mummies identifying the existence of *Mtb* of the spine (Pott's disease, as defined by Sir Percivall Pott in 1779[150]) in ancient Egyptian times.[151]

---

*This work has been published in and is reprinted in part with permission from: Pham, T. V., Mellott, D. M., Moghadamchargari, Z., Chen, K., Krieger, I., Laganowsky, A., Sacchettini, J. C., and Meek, T. D. (2021) Covalent Inactivation of Mycobacterium tuberculosis Isocitrate Lyase by cis-2,3-Epoxy-Succinic Acid, *ACS Chemical Biology 16*, 463-470.

For tens of thousands of years, the etiological agent of TB was unknown and went by many names ranging from "phthisis" in ancient Greece, "tabes" in ancient Rome, and "schachepheth," a word from ancient Hebrew that was used in the Biblical texts to describe TB.[152] However, two things were constant throughout these times; wild speculation on how the disease was contracted and how it could be cured. Early descriptions of TB infection included fever, sweating, coughing, and blood-stained sputum.[153] Treatments ranged from seeking fresh air, milk, and sea voyages, to more intriguing therapies such as inhaling smoke, drinking elephant urine, consuming wolf livers, or butter boiled with honey.[154]

By the late 1600's the Italian physician, Girolamo Fracastoro, proposed an early precursor to germ theory, where fomites (inanimate objects that when contaminated with an infectious agent can transmit the agent to a new host) may explain the spread of diseases such as TB.[155] In response, Italy issued edicts in the early 1700's referencing the infectious nature of TB and subsequently forbade individuals with the designated infectious diseases to be treated at public hospitals and instead sent them to designated hospitals to limit the spread of disease.[155] In the 18th century, however, with the industrial revolution sweeping Europe, there was a sharp increase in TB cases owing to overcrowded housing, poor sanitation, malnutrition, and destitute work conditions. These were prime conditions for the efficient spread of TB and allowed the mortality rate to climb to nearly 900 deaths per 100,000 people, ultimately resulting TB being the cause of one of every four deaths in Europe and North America.[152, 156] Even though thousands of years had passed since the first documented cases of TB infection, the standard of care for patient treatment had not

much changed. Principally, the best treatment comprised bed rest and fresh air in sanitoriums, a practice that lasted for another 100 years owing to a lack of viable therapeutic options.[157]

During the decline of the industrial revolution, extensive research was occurring in an attempt to both elucidate the causative agent of TB infection and identify treatment opportunities. By the late 1700's the caseous (or cheese-like) necrotic abscesses (tubercles) that were often found in the lymph nodes and lungs of those infected with TB were determined to be indicative of TB infection.[152] In a series of seminal studies by scientists, the materials inside the smallest early tubercles, or military tubercles, from deceased infected human hosts were isolated and transplanted into rabbits which lead to extensive TB infection and support for the hypothesis that TB was infectious.[152] In addition, the sputum produced by infected individuals proved to cause TB infection in guinea pigs and rabbit.[158] In one of the most well recognized and important discoveries for TB, Robert Koch, who later won a Nobel Prize for his work, stained extracted tubercule matter with an alkaline methylene blue solution which allowed visualization of rod-shaped bacilli under a microscope, demonstrating that TB was caused by a bacteria.[158, 159] In further work, he identified the presence of this bacilli in a variety of TB infected human and animal samples and also cultured the bacilli on solid media.[158] Together this work underscored the bacterial nature of TB infection and paved the way for researchers to directly study the bacilli responsible for this disease. This advance spurred on research developments in testing for TB using tuberculin skin tests[160] and the development of a vaccine from Calmette and Guérin (BCG),[161, 162] which is still used today in some

countries[163] and was recently postulated to provide protection from severe COVID-19,[164] but awaits verification from clinical trials.

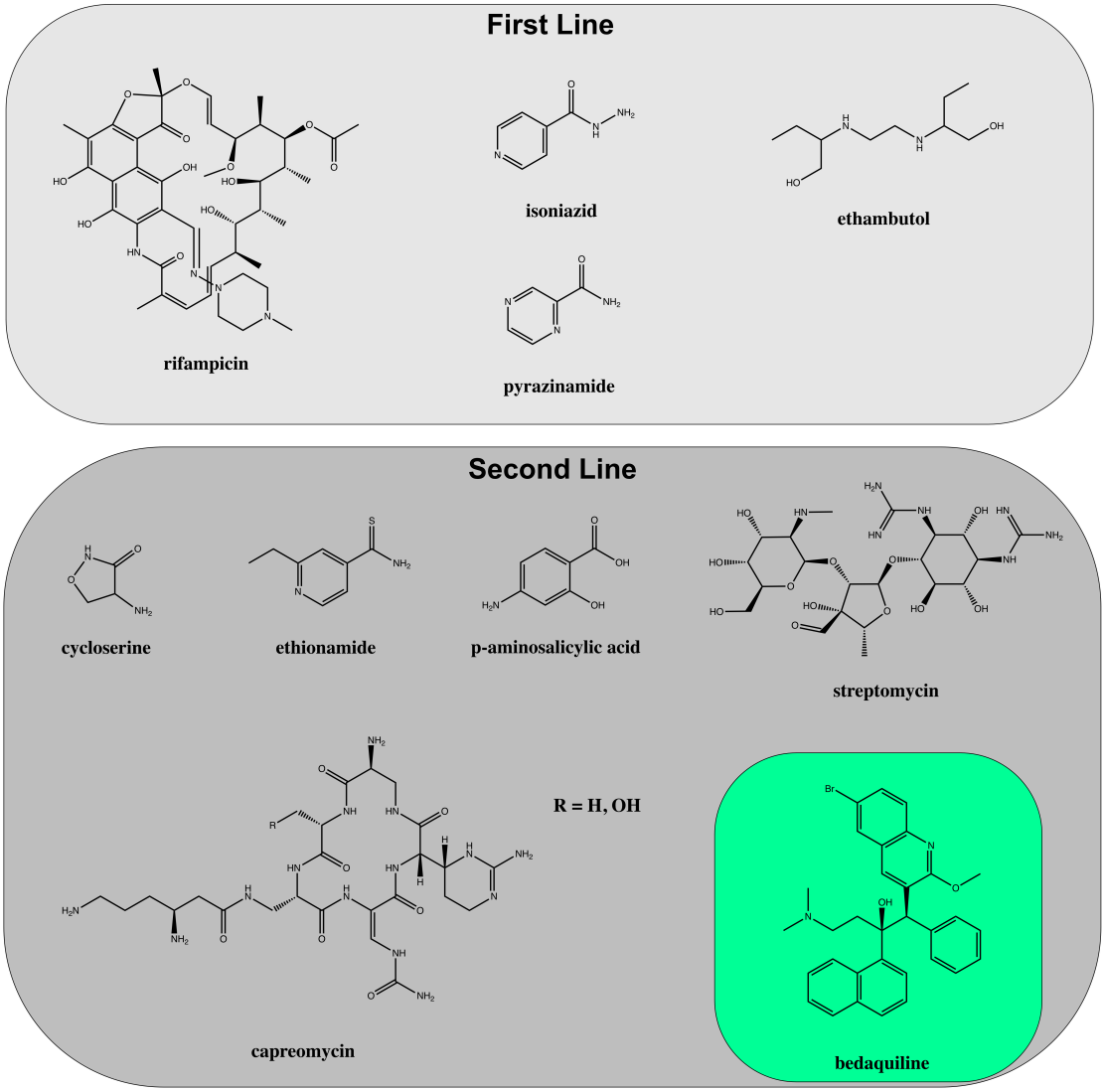


Figure 22. First and Second Line Therapeutics Approved for Treatment of TB by the US FDA

Further, the first anti-mycobacterial therapeutics, para-aminosalicylic acid (PAS) and streptomycin, were discovered in the mid-1940s and finally provided viable treatment

options for patients, particularly when co-administered.[165] Subsequent drug discovery efforts yielded isoniazid and rifampicin, two front line TB therapeutics, as well as D-cycloserine, ethambutol, pyrazinamide, and others.[157] However, unfortunately this time of opulence for antibacterial drug discovery has now stagnated, not only for TB, but for most bacterial diseases. This reduction in efforts and success in producing new antibacterial agents was highlighted in the 1980s and 1990s as there was a sharp increase in global TB cases and resistance to current frontline TB drugs.[166, 167] This increase, greatly assisted by the HIV epidemic[168] resulted in a 16% increase in TB cases in the US from 1985-1990,[169] which was coupled with identification that 26% of TB isolates were resistant to at least isoniazid, and that 19% were resistant to both isoniazid and rifampicin.[166, 167] Together, these concerns regenerated interest in developing therapeutics against this pathogen.[170] Most recently, in 2012 a new drug, bedaquiline, an ATP synthase inhibitor, was approved for treatment of multidrug-resistant TB,[171] but this therapy produces frequent and significant side effects that may limit its usage.[172] In addition, it cannot be used in conjunction with rifampicin as rifampicin treatment results in upregulation of cytochrome P450 CYP3A4[173] which degrades bedaquiline to a less potent form.[174] As rifampicin is the most active of the frontline TB drugs,[175] this metabolic issue may further limit bedaquiline's therapeutic potential as this modified bedaquiline product is rapidly cleared from the system leading to lower plasma concentration than required for therapeutic efficacy.[176] While this progress is heading in the right direction, TB still poses a global concern owing to the prevalence of its drug resistance. For example, in 2019 of the 10 million people that fell ill with TB, 0.5 million were diagnosed with TB resistant to the

frontline drug rifampicin.[146] Of these 0.5 million resistant cases, 78% of the strains were multidrug-resistant (resistant to rifampicin and isoniazid, MDR).[146] Together, since the initial infection of patients with TB thousands of years ago, we have been able to identify the etiological agent that afflicted death and destruction to hundreds of millions of people and have developed effective therapies to assist in treatment of this disease. However, research on this complex pathogen must continue in order to gain enhanced understanding of this bacterium in order to develop novel therapeutics to combat resistant strains of *Mtb*.

ii. Mtb Infection, Progression, and Key Discoveries

Owing to centuries of research conducted on *Mtb*, the scientific community has developed extensive knowledge about how this complex bacterium functions. One of the key clinical features of *Mtb* infection is the formation of a granuloma.[177, 178] A granuloma, in its simplest sense, is an organized collection of macrophages, a type of white blood cell. In the case of *Mtb* infection of the human host, initial inhalation of the aerosolized bacilli (1-10+ are required for infection) results in phagocytosis of the bacilli into alveolar macrophages and dendritic cells.[179, 180] Subsequent invasion of these macrophages into the epithelial layer results in a local inflammatory response leads to the recruitment of mononuclear cells from neighboring blood vessels.[181] These cells become infected by the rapidly growing bacilli and form an early granuloma, a diagnostic characteristic of Mycobacterial infection.

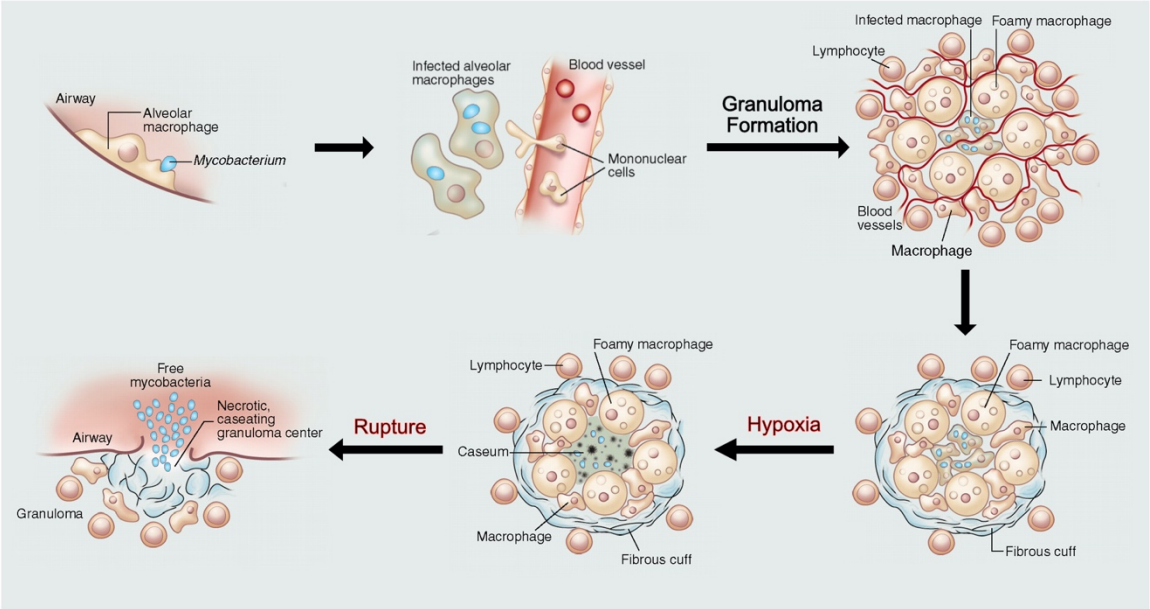


Figure 23. Infection Cycle of *Mtb* in Human Host
Adapted with modification from Russell, D. G., Cardona, P. J., Kim, M. J., Allain, S., and Altare, F. (2009) Foamy macrophages and the progression of the human tuberculosis granuloma, *Nat Immunol* 10, 943-948.

Most infectious pathogens engulfed by phagocytic cells ultimately succumb to the harsh conditions which ensue in the trafficking and maturation of the phagosome, ranging from antimicrobial peptides, high concentrations of reactive oxygen and nitrogen species, and acidic pH.[182] However, *Mtb* is able to manipulate the phagosome by varied mechanisms including pH modulation, quenching of reactive species, among others that are described in detail elsewhere.[183, 184] This bypass of traditional antimicrobial measures by host phagocytic cells thereby allows *Mtb* to use these cells as a primary reservoir within the human host granuloma. In addition, the oxygenated forms of mycolic acids that decorate the cellular surface of *Mtb* have been shown to play a role in the enrichment of foamy

macrophages in *Mtb* infected granulomas.[181,185] This class of macrophages are referred to as "foamy" owing to their high lipid content which has been proposed to possibly serve as a carbon source for the granuloma-ridden bacterium.[185] In about 90% of cases TB cases, the bacilli remains in this latent and arrested state in the granuloma, but given the right circumstances, such as immune collapse by coinfection with other pathogens like HIV,[168] or not fully understood perturbations within the granuloma that disrupt its integrity, reactivation of the disease can occur.[186]

In the shift to reactivation, initially cellular necrosis occurs at the core of the mature granuloma and results in the establishment of caseous tissue, which has been identified to contain cholesterol, cholesterol esters, triacylglycerols, and lactosylceramides.[187] This lipid-rich environment provides the bacilli with a dense nutrient source, albeit under hypoxic conditions.[187] While TB can remain in this stage for the rest of its lifecycle within the host, it can also cause liquification and cavitation of this caseous tissue (thought to be mediated in part by *Mtb* metalloproteases).[188] This ultimately results in the rupture of the granuloma and release thousands of bacilli into the airways of the host, thus completing the infection cycle.[186] On its face, this progression of TB infection may seem biphasic, but ongoing research suggest this is far from the case. Instead, it is thought that mycobacterial infection of granulomas is quite heterogeneous and activation/reactivation of infection may be driven by a mixture of global and local perturbations to this granuloma-ridden *Mtb*.[189] Thus, this heterogeneous character of infection likely plays an integral role in the reactivation of TB infection in latent patients[190] but key immunological and metabolic determinants of this reactivation are still poorly understood.[177, 189, 191]

While researchers have struggled to decipher the heterogeneity of TB infection in individuals, much success has occurred in gleaning an enhanced understanding of how the intracellular environments that *Mtb* experience may perturb its metabolism and gene/protein expression. Studies as early as the 1950s by Bloch and Segal[192] observed that *Mtb* harvested from the infected cells of mice preferentially catabolized fatty acid carbon sources. This suggested that the intracellular environments the bacilli experienced *in vivo* resulted in differential catabolism of carbon sources, likely due to altered gene expression compared to *Mtb* grown in liquid medium. Sequencing of the *Mtb* genome in 1998,[193] and putative assignment of protein functions, was a critical juncture in enhancing the accessibility to explore molecular functions and essentiality of enzymes and proteins within the bacilli. One key identification in this genomic study was that *Mtb* contained a very large number of lipid metabolism genes (comprising ~250 distinct enzymes), including enzymes that were normally only found in mammals or plant systems.[193] When compared to *E.coli,* it was seen that *Mtb* had a 5-fold increase in the number of encoded enzymes for lipid metabolism, albeit having a similar genome size (4.1 Mbp[193] vs. 4.6 Mbp[194], respectively).[195] This strongly suggested that *Mtb* lipid metabolism plays a key role in the ability of this bacterium to be such an effective pathogen. This genomic information also allowed researchers to use approaches such as saturating transposon mutagenesis to identify essential *Mtb* genes.[196, 197] This method leverages the lack of specificity of the Himar1 transposon to disrupt a gene via insertion of a transposable element (inserts at TA dinucleotide sites, thus approximately 17,000 insertions possible in *Mtb* genome).[196] Then, using deep sequencing[198] or microarray hybridization,[196] one can

identify genes that underwent transposition thereby suggesting which genes are conditionally essential and non-essential genes.[196, 197]

This analysis has proven critical in defining new drug targets and elucidating new metabolic pathways. One example of this is the important identification of the gene *mce4* being required for the survival of *Mtb* in mice.[197, 199] Subsequent biochemical characterization of this gene identified it expressed an ATP-dependent transporter for cholesterol import.[200] This finding resulted in the critical understanding that *Mtb* can efficiently use cholesterol as a carbon source, thereby linking cholesterol import to the ability of the bacterium to maintain a chronic infection.[200] The more recently, the use of metabolomics has facilitated enhanced and in depth understanding of *Mtb* metabolism.[179, 201, 202] This approach has resulted in significant advances in elucidating the precise roles enzymes play under conditions ranging from hypoxia to where *Mtb* utilizes fatty acids and cholesterol as carbon sources.[203, 204] This powerful approach has enlightened key flux points in *Mtb* metabolic networks and the critical players under these conditions, one of which is the enzyme isocitrate lyase.[179, 203-209]

iii. Isocitrate lyase: A Key Metabolic Enzyme for Mtb Persistence

The protein isocitrate lyase from *Mtb* has two isoforms that are expressed in *Mtb* cells coined ICL1 and ICL2.[207] From a functional perspective, isocitrate lyase enzymes catalyze the retro-aldol cleavage of the substrate D-isocitrate into glyoxylate and succinate as products.[210, 211] This reaction comprises the first step of the glyoxylate shunt, an anaplerotic bifurcation of the tricarboxylic acid cycle that is utilized by bacterium, plants,

and protists.[212] This pathway provides microorganisms an efficient pathway to assimilate two molecules of acetyl-CoA when using acetate, fatty acids, or ketogenic amino acids as carbon sources while eliminating the loss of CO_2 via the decarboxylation of D-isocitrate and alpha-ketoglutarate.[213] Seminal studies by Bentrup *et al*. identified that *icl1* was overexpressed in *Mtb* grown with fatty acids as a sole carbon source (acetate and palmitate) while it was expressed much less robustly under glucose conditions.[207] Further, through recombinant expression and purification of ICL1 and ICL2, it was identified that both enzymes were active, with ICL1 being more active than ICL2, and both requiring Mg^{2+} for maximal activity.[207] Of interest, the *icl2* gene was identified to contain a frameshift mutation in some strains of *Mtb* (H37Rv and the Euro-American strain 630) which produces two distinct ICL2 derived gene products (Rv1915 and Rv1916) that were postulated to yield inactive enzymes.[208] However, recent characterization of recombinant Rv1916 revealed it was an active isocitrate lyase enzyme, suggesting that the activity was not fully lost due to the frameshift mutation.[209] Further, in a more elegant study, full length ICL2 from the CDC1551 *Mtb* strain was identified to require allosteric binding of acetyl CoA to obtain full activity.[214]

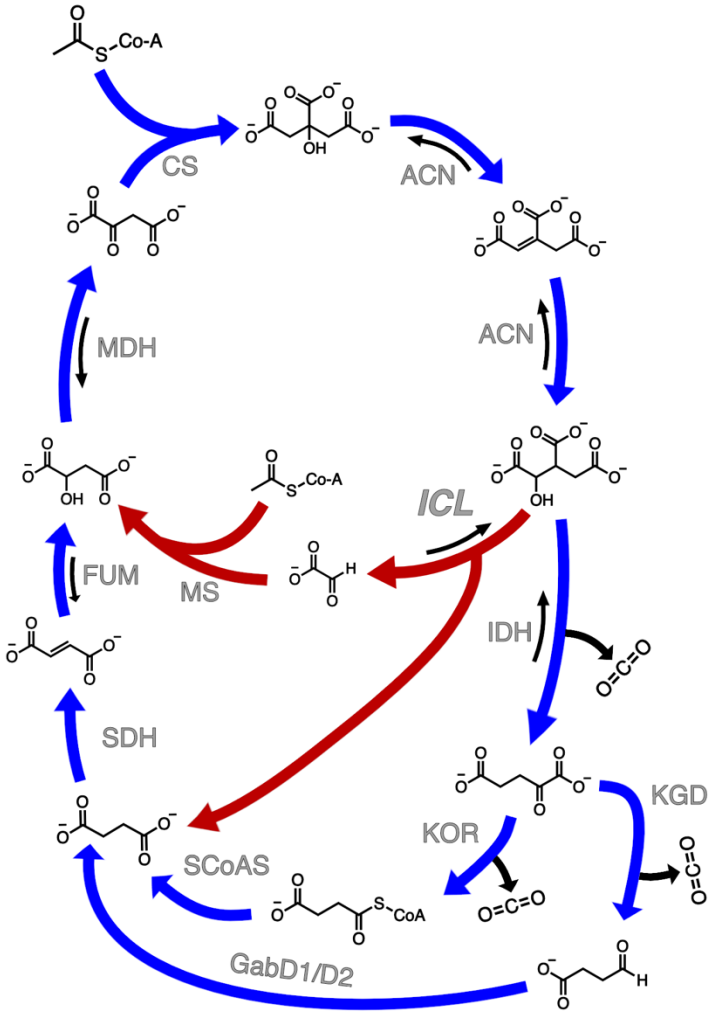


Figure 24. The TCA Cycle (blue) and the Glyoxylate Shunt (red) of *Mycobacterium tuberculosis*

The glyoxylate shunt is primarily utilized for fatty acid metabolism of even chain fatty acids. CS, citrate synthase; ACN, aconitase; ICL, isocitrate lyase; IDH, isocitrate dehydrogenase; KOR, α-ketoglutarate ferredoxin oxidoreductase; KGD, α-ketoglutarate decarboxylase; GabD1/D2, succinate-semialdehyde dehydrogenase; SCoAS, succinyl-CoA synthetase; SDH, succinate dehydrogenase; FUM, fumerase; MS, malate synthase; MDH, malate dehydrogenase.

This underscores why previous recombinantly produced ICL2 enzymes, which were often truncated to remove this additional domain (to which acetyl CoA binds) for enhanced stability and solubility, exhibited weak catalytic activity compared to the ICL1 enzyme.[215] From a cellular context, the early work of McKinney et. al identified that deletion of the *icl1/2* from *Mtb* resulted in clearance of the bacilli from the lungs of mice.[216] In addition, ICL1 is one of the most highly upregulated genes in *Mtb* during mouse infection and in *Mtb* infected macrophages by qRT-PCR analysis.[217] This upregulated production was further corroborated by producing an *Mtb* strain containing an ICL1/2-GFP (green fluorescent protein) fusion protein which upon macrophage infection resulted in heightened GFP signal (compared to non-infected cells), corresponding to enhanced ICL expression.[216] Further biochemical experiments went on to identify that *icl1* and *icl2* were jointly required for virulence and *in vivo* growth of *Mtb*, thus verifying that the expression of the ICLs comprise an essential process in the effective persistence of TB infection.[218] However, it intrigued the researchers that ICL1/2 activity was conditionally essential for *Mtb* growth under propionate medium. This intrigue stemmed from the fact that typical microorganisms breakdown propionate chiefly through the methylcitrate cycle,[219] a pathway that canonically uses three enzymes: methylcitrate synthase (MCS), methylcitrate dehydratase (MCD), and 2-methyl-isocitrate lyase (MCL)(Figure 25). Together these three enzymes convert propionyl CoA into pyruvate and succinate which can then be incorporated into the TCA cycle for energy production.

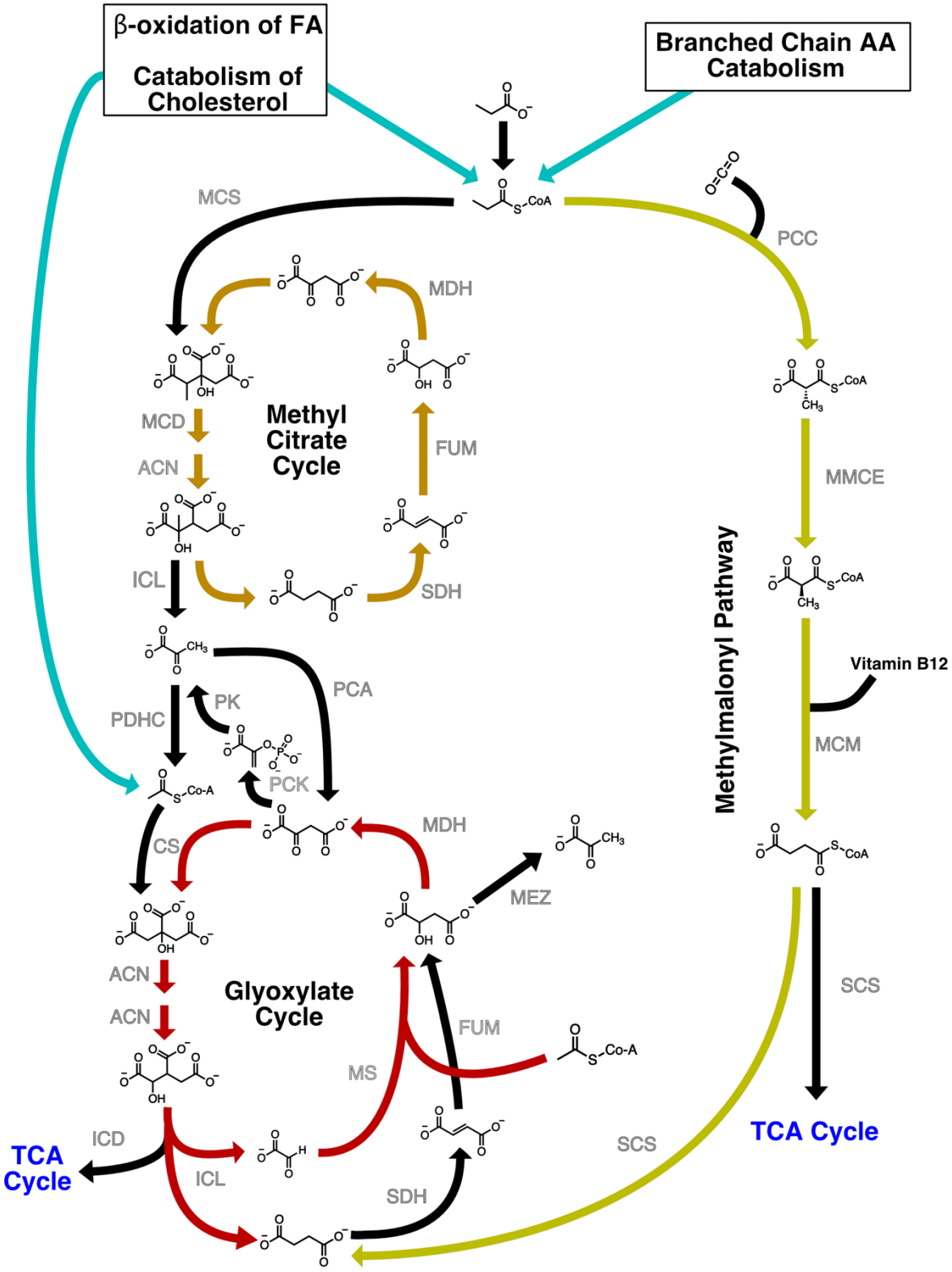

β-oxidation of FA
Catabolism of Cholesterol
Branched Chain AA Catabolism
MCS
PCC
MDH
MCD
ACN
ICL
Methyl Citrate Cycle
FUM
SDH
MMCE
Methylmalonyl Pathway
Vitamin B12
MCM
PDHC
PK
PCA
PCK
CS
MDH
MEZ
ACN
ACN
Glyoxylate Cycle
FUM
MS
SCS
ICD
TCA Cycle
ICL
SDH
SCS
TCA Cycle

Figure 25. *Mtb* Catabolism of Odd Chain Fatty Acids is Conducted via the Methycitrate Cycle and Methylmalonyl Pathways.

The catabolism of odd chain fatty acids is conducted via the methycitrate cycle, whereby the MCL activity of ICL is essential for completing this pathway. The methylmalonyl pathway is functionally independent of ICL, but requires cofactor vitamin B12. MLS, malate synthase; PCC, propionyl-CoA carboxylase; MCD, malonyl-CoA decarboxylase; ACN, aconitase; SDH, succinate dehydrogenase; FUM, fumerase; MDH, malate dehydrogenase; ICL, isocitrate lyase; MMCE, methylmalonyl-coenzyme A epimerase; PDHC, pyruvate dehydrogenase complex; PCA, pyruvate carboxylase; PK, pyruvate kinase, PCK, phosphoenolpyruvate carboxykinase; CS, citrate synthase; MCM, methylmalonyl-CoA mutase; MDH, malate dehydrogenase; MEZ, malic enzyme; ICD, isocitrate dehydrogenase; SCS, succinate synthase. Adapted and modified from Savvi *et al*. J. Bacteriol. **2008**, 190(11) 3886-389.

However, analysis of the *Mtb* genome revealed that while it encodes for MCS and MCD genes, it apparently lacks a MCL enzyme. This suggested that ICL may play a bifunctional role in *Mtb* where it catalyzed the conversion of isocitrate to glyoxylate and succinate in the glyoxylate shunt and of 2-methyl-isocitrate to pyruvate and succinate in the methylcitrate cycle. Through incubating 2-methylisocitrate with whole cell lysates of wild-type and icl mutant stains (Δ*icl1*, and Δ*icl1/2*), Muñoz-Elías *et al*. identified that *Mtb* MCL activity was specific to the presence of the genes that express ICL1 and 2 and that this activity was upregulated under propionate growth conditions.[220] Further work by Gould *et al*. showed that recombinantly produced ICL1 exhibited both ICL and MCL activity (ICL2 exhibited poor activity) and resolved a crystal structure of an ICL1-$C_{191}S$ mutant that had been treated with 2-methy-isocitrate.[221] This structure revealed the reaction products of pyruvate and succinate were bound in two of the four subunits, consistent with the ability of ICL1 to bind to and catalyze the cleavage of 2-methyl-

isocitrate.[221] More recent studies demonstrated that full-length ICL2 with acetyl CoA bound exhibited MCL activity of nearly identical efficiency to ICL1, suggesting that its activity may be regulated in the intracellular content of fatty acids using acetyl CoA as a signaling molecule to stimulate its activity.[214] Together, this information suggested that the ICLs may play a critical role in the methylcitrate cycle via the retro-aldol cleavage of 2-methylisocitrate. However, initial experiments contradicted this hypothesis, and instead postulated that the methylcitrate cycle may be dispensable in *Mtb*. Observation that deletion of two genes in the methylcitrate cycle (prpC and prpD), which are required for conversion of propionyl-CoA to 2-methylisocitrate, did not cause growth defects in *Mtb* infected mouse lungs, leading to the suggestion that the critical role of the ICLs *in vivo* may be independent of their MCL activity. Follow-up experiments proposed that the essentiality of the ICLs may instead owe to their elimination of toxic metabolites derived from propionyl-CoA metabolism. In agreement with this, Eoh and Rhee,[203] in an elegant ^{13}C metabolomics tracing study, identified that lack of the MCL activity of the ICLs was the chief metabolic defect responsible for the bactericidal effects of acetate and propionate carbon sources to Δ*icl Mtb* strains. These defects were attributed to the depletion of oxaloacetate from the TCA cycle and gluconeogenic pathways, accumulation of 2-methyl-isocitrate which inhibits fructose 1,6-bisphophatase, among other enzymes,[222] and accumulation of propionyl-CoA and methylmalonyl-CoA which results in perturbations to the NAD/NADH flux and perturbation of intracellular pH.[203] Together, this collection of work highlights the essentiality of the ICLs for *Mtb* persistence and growth under carbon source conditions (cholesterol[223] and fatty acids) similar to what it experiences in

the granuloma, thus validating the ICLs as drug targets for anti-mycobacterial development.

iii. Isocitrate lyase: Structure, Function, and Regulation

The identification of an enzyme that catalyzed the cleavage of isocitric acid into glyoxylate and succinate was first reported in 1954 and cleverly named isocitrase.[224] In the many years since, this enzyme has been identified to play a key role in the glyoxylate shunt of organisms ranging from Archea, bacteria, fungi, nematodes, plants, to protists.[212] This homotetrameric enzyme requires an active site divalent cation (Mg^{2+} or Mn^{2+}) and

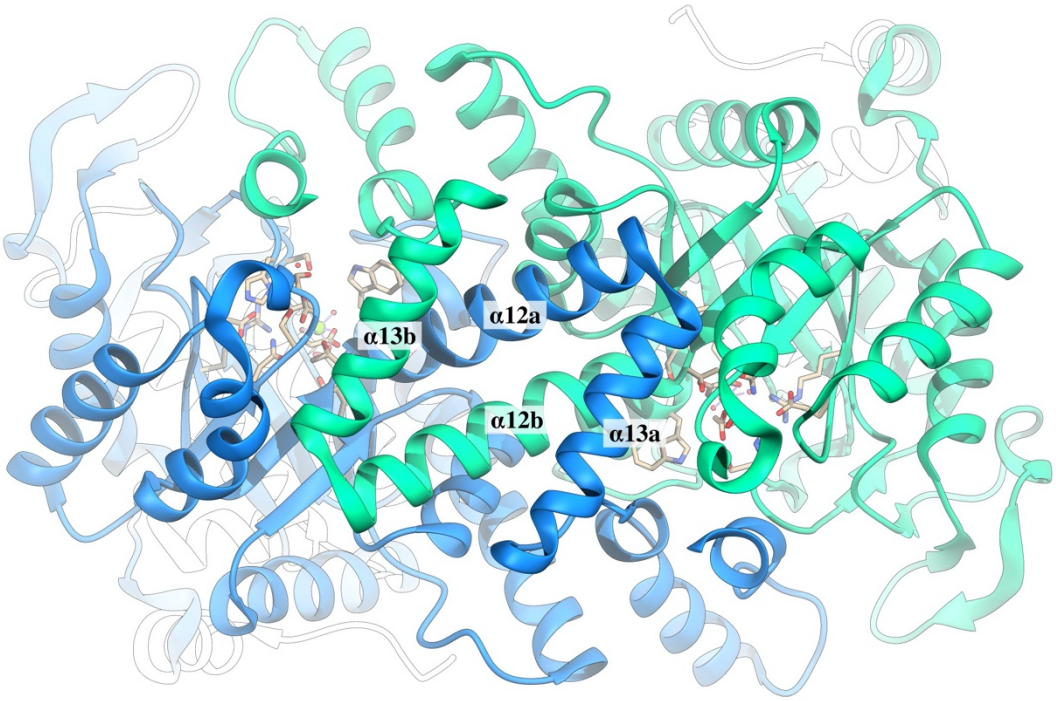


Figure 26. Helix Swapping of *Mtb* ICL1

ICL1 crystal structure (PDB: 1f8i) showing the helix swapping of the $\alpha 12$ and $\alpha 13$ helices of chain A and B of the structure.

cysteine residue that resides in the signature ICL catalytic sequence motif (KKCGH) for efficient catalysis.

In the case of *Mtb* ICL1, the formation of this homotetrameric structure, or dimer of dimers, results from helix swapping of the α12 and α 13 helix with the noncrystallographically related subunits of this unique α/β folded protein (Figure 26).[225] This helix swapping has been shown to occur in other ICLs[226] as well as the tetrameric phosphoenolpyruvate mutase[227] and is proposed to assist in the formation of a stable dimer.[228] Close inspection of the active site of the apo enzyme reveals the ICL signature sequence resides on a mobile loop that in its "open" state is exposed to solvent (Figure 27A). Visualization of the active site cavity of *Mtb* ICL1 reveals a quite small and highly electronegative space (Figure 27B). In particular, the region of residues seen to complex the Mg^{2+} divalent cation (Asp 108, Glu 155) and surrounding area (Asp 153 and Glu 182) are key contributors to this overwhelmingly electronegative active site (Figure 27A). This adds an additional challenge from the perspective of drug design as ICL enzymes by generality have a very stringent tolerance with respect to both the size and charge of the molecule it will tolerate in its active site cavity. These stringencies may explain why high-throughput screening against *Mtb* ICL1 both in the pharmaceutical industry and in academia has led to no viable lead compounds. However, it has been seen that *Mtb* ICL1 will tolerate small polar covalent inhibitors such as 3-bromopyruvate and 3-nitropropionate, which undergo alkylation with the catalytic Cys_{191} of *Mtb* ICL1 and lock the mobile active site loop (~15Å) in a closed conformation. Owing to a lack of viable lead pharmacophores, non-drug-like compounds like 3-brompyruvate and 3-

nitropropionate still serve as benchmark standards for *Mtb* ICL1 inhibitor design. While these molecules effect robust activity against the enzyme, they also exhibit toxicity [229, 230] and thus it is necessary to develop new discovery approaches to develop specific inhibitors for this sought after *Mtb* drug target.

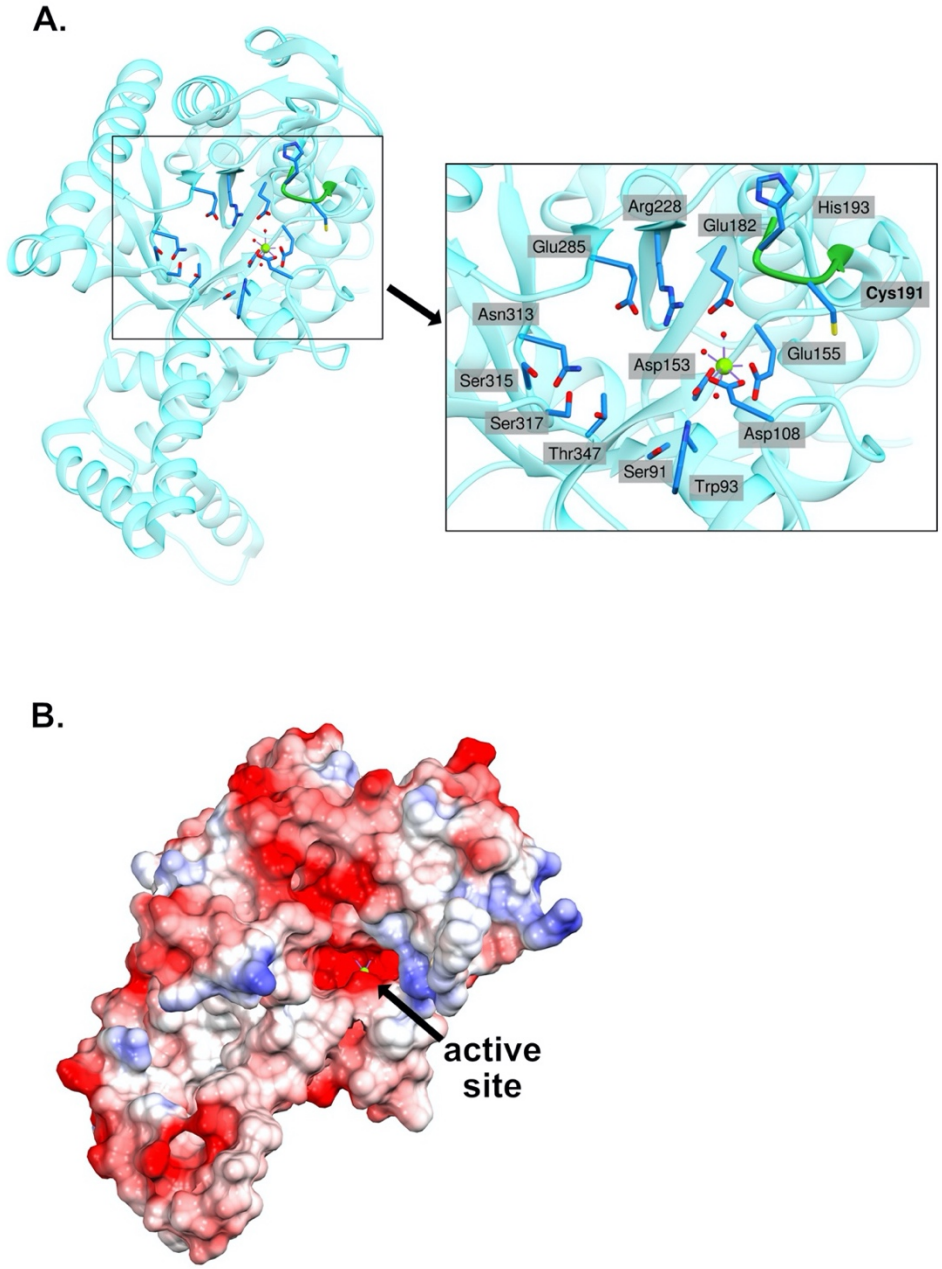


Figure 27. ICL1 Active Site Residues and Electrostatic Surface Map.

A. ICL monomer from PDB: 1f61 where key active site residues are shown in blue and the ribbon form of the active site mobile loop is shown in lime green which contains the ICL signature sequence KKCGH. **B**. Electrostatic surface potential map for ICL1 monomer from PDB: 1f61 where the red is electronegative, white is neutral, and blue is electropositive.

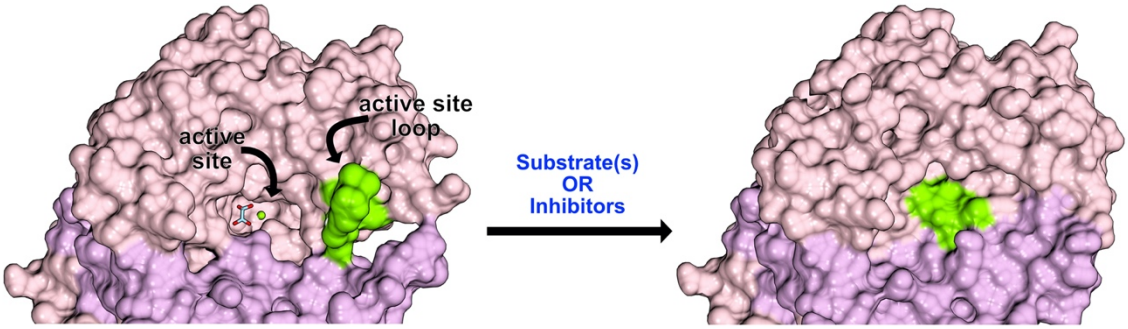


Figure 28. ICL1 Loop Closure.

Morph of ICL1 loop closure with PDB: 1f61 and 6vb9 showing the conformational shift that occurs following the binding of a substrate or inhibitor to the enzyme.

From an enzymology perspective, *Mtb* ICL1 carries out the retro-aldol cleavage of the substrate D-threo-isocitrate (D-IC) to produce glyoxylate and succinate which desorb from the enzyme in an ordered fashion (Figure 29A).[231] In the reverse direction, initial complexation of glyoxylate to the active site divalent Mg^{2+} cation is followed by ordered binding of succinate.[211] Subsequent aldol condensation of the substrates forms the product D-IC. Through combining detailed kinetic studies[211] with structural data,[225] researchers have been able to propose putative chemical mechanisms by which *Mtb* ICL1 catalyzes the cleavage of D-IC and condensation to form D-IC (Figure 29B). Briefly, in the direction of D-IC cleavage, the substrate binds to *Mtb* ICL1, likely as a Mg^{2+}-D-IC complex,[232, 233] following its positioning in the active site by unknown residues. The abstraction of a proton from the C2 carboxylate of D-IC by an unknown catalytic base forms an aci-succinate intermediate and glyoxylate. The collapse of this intermediate results in abstraction of a proton from the Cys_{191} thiol to from succinate and an active site thiolate.

Subsequent ordered desorption of the product succinate is followed by the release of glyoxylate, thereby reading the enzyme for another round of catalysis.

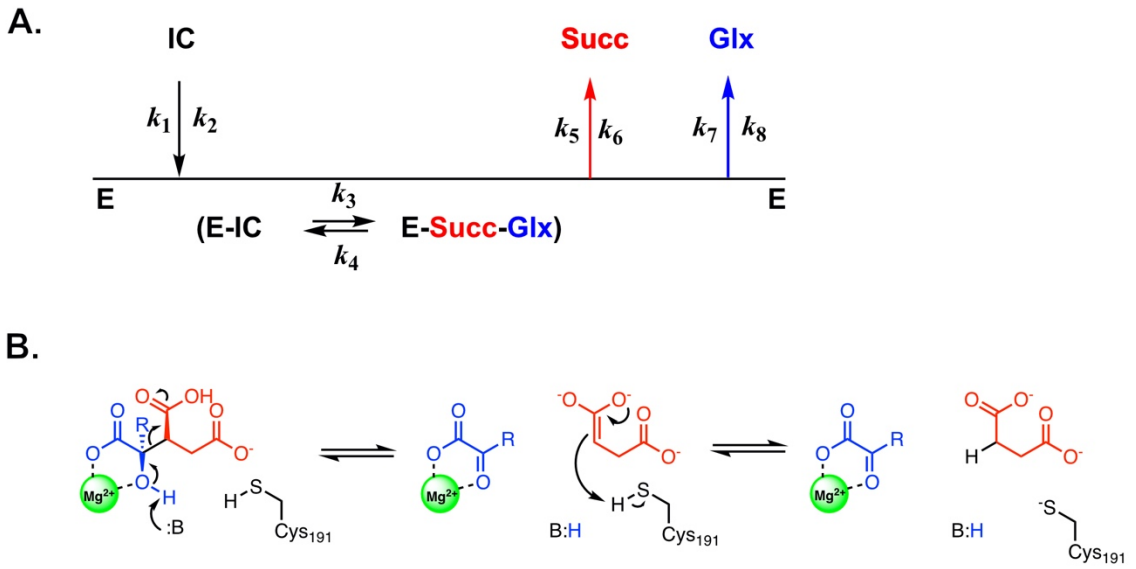


Figure 29. Proposed Kinetic and Chemical Mechanism of *Mtb* ICL1 in the direction of D-isocitrate cleavage.

Since the best-known molecules that blockade ICL enzyme activity via covalent bond formation with the active site catalytic cysteine, we envisioned that development of mechanism-based inactivators and less toxic inactivators of ICL may comprise a reasonable strategy to target this protein. For example, two well-known ICL inactivators, 3-nitropropionate[234] and 3-bromopyruvate[235] both exhibit toxicities. 3-Nitropropionate is neurotoxic, and is a mechanism-based inactivator of succinate dehydrogenase,[229] whereas 3-bromopyruvate reacts readily with thiols, as well as the active site cysteines of other enzymes such as glyceraldehyde-3-phosphate dehydrogenase.[230] In an effort to thwart this toxicity and develop more specific molecules for ICL inactivation we used the approach of developing mechanism-based inactivators (MBIs). As discussed in Chapter I, MBIs

comprise molecules that masquerade as enzyme substrates, but upon specific enzyme catalysis produce a reactive species which can undergo reaction with the enzyme.[7] This may be a viable strategy to target *Mtb* ICL1 owing to both its nucleophilic active site thiol and the active site loop closure of *Mtb* ICL1 which will, in theory, entrap the reactive species of the MBI within the active site cavity until it undergoes reaction with the enzyme. Therefore, we designed a possible mechanism based inactivator of the *Mtb* ICLs, 2-vinyl-(2*R*,3*S*)-isocitrate (2-VIC) (Figure 30). This molecule was found to be

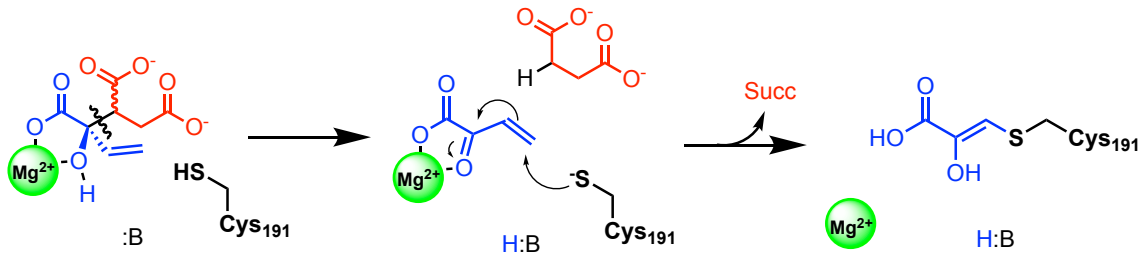


Figure 30. Proposed Mechanism of 2-VIC Inactivation of ICL1

synthetically complex to produce (26 linear steps), but indeed served as a mechanism-based inactivator of *Mtb* ICL1 and 2 as determined by using kinetics, protein mass spectrometry, and X-ray crystallography. From a mechanistic perspective, inactivation of *Mtb* ICL1 by 2-VIC is proposed to be initiated in an identical fashion to D-IC catalysis, whereby an unknown base abstracts a proton from the carbon-2 hydroxyl of 2-VIC to form an aci-succinate intermediate and the reactive metabolite 2-vinyl glyoxylate. Collapse of the aci-succinate intermediate unmasks a Michael acceptor, 2-vinyl-glyoxylate (2-VG), that efficiently forms a covalent adduct with Cys_{191}.[215] While this molecule proved to be an efficient *Mtb* ICL MBI, one challenge we encountered was that 2-VIC was liable to

free thiols and that 2-VIC mediated inactivation was readily reversible upon incubation of the enzyme with thiols. We postulate the latter to owe to the ordered release of *Mtb* ICL product release which exposes the active site of *Mtb* ICL1 during the dissociation of succinate,[225] thereby resulting in the apparent interception of 2-VG by free thiols *in vitro* in lieu of its reaction with Cys_{191}. This thiol reactivity could in part account for its lack of anti-mycobacterial activity. We therefore sought to develop more durable covalent inactivators of the *Mtb* ICLs that might arise from succinate analogs bearing an electrophilic group.

Towards this end, we analyzed a panel of di-carboxylic acids and found that cis-olefinic-di-carboxylic acids, some with substitution at carbon-2, effected uncompetitive inhibition versus D-IC while their trans-di-carboxylic acid counterparts were D-IC competitive inhibitors. Structural similarity suggested that (2R,3S)-epoxy-succinate (*cis*-EpS) could bind to the succinate sub-site of *Mtb* ICL1, and the proximity of Cys_{191} to the epoxide group may result in the formation of a covalent adduct concomitant with opening of the epoxide ring. Here we report that *cis*-EpS is a selective and potent covalent inactivator of *Mtb* ICL1, whose action is potentiated by the binding of isocitrate, glyoxylate, or glycolate.

B. Experimental Methods

i. Protein Expression and Purification

A plasmid containing *Mtb icl1* was a gift from Dr. Andrew Murkin of Buffalo University, NY.[211] This construct contains the ICL1 gene (Uniprot: P9WKK6) with an N-terminal, thrombin-cleavable His_6 epitope tag. The recombinant protein was expressed in *E.coli* BL21(DE3) cells, and was expressed and purified as previously described.[211, 215] Pure ICL1 was stored in a buffer of 50 mM Tris-HCl (pH 7.0), 150 mM NaCl, and 10% glycerol at –80 °C, or in 25 mM Tris-HCl (pH 7.5), 75 mM NaCl, 0.5 mM DTT and 50% (v/v) glycerol at –20 °C. For mass spectrometry, ICL1 was further purified via gel filtration (HiLoad® 26/600 Superdex® 200 pg, GE Healthcare), in which the column was pre-equilibrated with 50 mM HEPES (pH 7.5) and 150 mM NaCl. Fractions containing the pure tetramer were pooled, concentrated, and stored at -80°C. For crystallography, the His_6 epitope tag was removed by incubation with thrombin, in both the presence and absence of D-isocitrate to protect the tag-free ICL1 from further proteolysis. Truncated ICL2 (Uniprot: Q8VJU4) was expressed and purified as previously described.[215] *E. coli* isocitrate dehydrogenase (ICDH) gene was obtained from an Aska (-) clone containing the *icd* gene in a pCA24N plasmid.[211, 236]

ii. Enzyme Assays

Unless otherwise specified, all assays were conducted in clear 96-well plates consisting of 250-µL reaction mixtures containing 50 mM HEPES (pH 7.5), 5 mM $MgCl_2$, and 1 mM DTT at 37ºC. In the direction of isocitrate cleavage, product glyoxylate was reacted with

10 mM phenyl-hydrazine-HCl to form its phenyl-hydrazone product, which was measured spectrophotometrically at A_{324nm} (ε_{324} = 17,000 $M^{-1}cm^{-1}$). In the direction of isocitrate synthesis, product isocitrate was converted to α-ketoglutarate using a coupled-enzyme assay containing 100 nM *E. coli* isocitrate dehydrogenase and 0.25 mM NADP. Absorbance of the product NADPH was measured spectrophotometrically at $A_{340\ nm}$ (ε_{340} = 6,220 $M^{-1}cm^{-1}$). Alternately, isocitrate synthesis was measured using a two-enzyme coupled system containing 0.1 units of bovine isocitrate dehydrogenase (Sigma Aldrich), 0.3 mM $NADP^+$, 0.1 units diaphorase from *Clostridium kluyveri* (Sigma Aldrich) and 50 μM resazurin in the assay buffer described above, but without the addition of 1 mM DTT. The formation of D-isocitrate, concomitant with NADPH formation by isocitrate dehydrogenase, was followed by the conversion of NADPH to $NADP^+$, and resazurin was converted to resorufin by diaphorase catalysis, allowing the detection of resorufin at 525/598 nm (ex/em). This assay was conducted at 25°C in black clear-bottom 96-well half-volume plates (Greiner) with a final reaction volume of 150 μL. All spectrophotometric and fluorescent measurements were recorded on a Biotek® Synergy M2 plate reader equipped with temperature control.

iii. Native and Denatured Mass Spectrometry

Samples of 22.5 μM ICL1 in 50 mM HEPES, 150 mM NaCl, (pH 7.5) were incubated with *cis*-Eps (112.5 μM) alone or with *cis*-Eps (112.5 μM) and glyoxylate (1 mM) or glycolate (1 mM) overnight. The protein was then exchanged into a volatile buffer, 200 mM ammonium acetate (pH 7.4), using a Micro Bio-Spin 6 device (Bio-Rad). For analysis

of native protein, instrumental parameters were tuned to preserve non-covalent interactions, and to minimize gas-phase activation on a Thermo Exactive Plus with an extended mass range Orbitrap mass spectrometer. Gas-phase activation was achieved with increasing collision voltage in source (150 V) and HCD cell (200 V). Proteins were also denatured using 1% formic acid and introduced into the mass spectrometer.

iv. Intrinsic Protein Fluorescence

ICL1 (1 μM monomers) in 50 mM HEPES and 10 mM $MgCl_2$ (pH 7.5) was combined with either *cis*-EpS (1 μM) or *cis*-EpS (1 μM) and glyoxylate (100 μM). Reaction mixtures (200 μL) were excited at λ_{ex} = 295 nm, and the intrinsic fluorescence emission spectra (λ_{em} = 315–380 nm) was determined at multiple time points (room temperature) using a Biotek Synergy plate reader and black clear-bottom 96-well plates (Greiner). The λ_{max} of ICL1 fluorescence was determined to be 325 nm. All data were corrected by subtracting the background fluorescence observed in a sample of buffer. Upon addition of ligand the spectrum was recorded multiple time points. Each spectrum was normalized, assigning the fluorescent intensity of protein alone at 325 nm to 1. For the samples containing both *cis*-EpS and glyoxylate, we subtracted the glyoxylate spectrum in order to elucidate the relative fluorescent change resulting from *cis*-EpS binding to ICL1. Relative fluorescent change of ICL1 at 325 nm in response to ligand binding were plotted using $F_t = (F_0 - F_{min})*\exp(-k_{obs}*t) + F_{min}$ in which F_t, F_0, and F_{min} are the respective protein fluorescence measured at time t, t=0, respectively, and k_{obs} is the apparent first-order rate constant of diminution of fluorescence.

v. X-Ray Crystallography

Samples of tag-free ICL1 (10 mg/ml) in 50 mM Tris (pH 8.0), 10 mM $MgCl_2$, and 1 mM DTT were treated with 2 mM *cis*-epoxy-succinic acid for 1 h prior to set up for crystallization. Crystals were produced by the hanging-drop vapor diffusion method at 17 °C in 0.1 M Tris-HCl (pH 8.0), 0.2 M sodium acetate, and 20-30% PEG4000 mixed at 1:1 or 1:3 volume ratio with the protein solution. For the X-ray data collection, crystals were flash-frozen in liquid nitrogen after cryoprotection with 20% glycerol added to the mother liquor. High-resolution diffraction data for ICL1-EpS co-crystals were collected at beamline 19ID of the Structural Biology Center (SBC, Advanced Photon Source, Argonne National Laboratory). Detailed information about the data transformation and structural refinement is described in[14]

vi. Analysis of Kinetic Data

Time courses of enzyme inactivation were fitted to eq. 3, a modification of an equation for fitting time-dependent inhibition wherein the final steady-state rate is zero, due to apparent irreversible inactivation. [P] is the concentration of glyoxylate, v_i is the initial rate (mOD/s) or ([Glx]/s), t is the time of reaction, and k_{obs} is the apparent first-order rate constant of inactivation obtained at each concentration of inactivator. Values of k_{obs} (the maximal observed rate constant (s^{-1})) were re-plotted using eq. 4, wherein [I] is the concentration of inactivator, k_{inact} is the maximal rate constant of inactivation, and K_{inact}' is the mechanism-dependent, apparent concentration of inactivator at which the observed rate constant of inactivation is half that of k_{inact}.

$$[P] = \frac{v_i(1-e^{-k_{obs}t})}{k_{obs}} \quad \textbf{(eq. 3)}$$

$$k_{obs} = (k_{inact}[I])/(K_{inact}(1 + K_a/[A])+[I]) \quad \textbf{(eq. 4)}$$

Apparent values of k_{obs}, vs. [isocitrate] or [succinate] as shown in Fig. 1c were respectively fitted to eq 5 and 6, and K_{inact}' at variable concentrations of D-isocitrate or succinate [A], wherein K_a are the Michaelis constants of the two substrates, were fitted to eqs 7 and 8 for uncompetitive and competitive inactivators, respectively.

$$k_{obs'} = k_{obs}\left(1 + \frac{K_a}{[A]}\right) \quad \textbf{(eq. 5)}$$

$$k_{obs'} = k_{obs}\left(1 + \frac{[A]}{K_a}\right) \quad \textbf{(eq. 6)}$$

$$K_{inact'} = K_{inact}\left(1 + \frac{K_a}{[A]}\right) \quad \textbf{(eq. 7)}$$

$$k_{inact'} = k_{inact}\left(1 + \frac{K_a}{[A]}\right) \quad \textbf{(eq. 8)}$$

Data of [glyoxylate] vs. time at changing-fixed concentrations of *cis*-EpS were also fitted globally, that is, all concentrations of the inactivator simultaneously, to eq. 9, for which

K_{i1} is the inhibition constant for binding of I to E-Glx, and k_{inact} and K_{inact} are limiting values as defined below.

$$[P] = \left[\frac{\frac{k_{cat}[E_t][A]}{\left(Ka+[A]\left(1+\frac{[I]}{K_{i1}}\right)\right)}}{\left(\frac{k_{inact}[I]}{K_{inact}\left(1+\frac{Ka}{[A]}\right)+[I]}\right)}\right]\left[1 - \exp\left(\frac{-tk_{inact}[I]}{K_{inact}\left(1+\frac{Ka}{[A]}\right)+[I]}\right)\right] \quad \textbf{(eq. 9)}$$

C. Results and Discussion

i. Identification of cis-EpS as a Covalent Inactivator of ICL1

In this section, I will briefly detail initial work my co-author and former lab mate Dr. Truc Pham conducted which led to the initial identification of *cis*-EpS as an ICL1 inactivator. The kinetic mechanism of ICL1 is ordered uni-bi; desorption of succinate (Succ) from (ICL-Succ-Glx) precedes that of glyoxylate (Glx) (Figure 31. Therefore, we sought to uncover if succinate bound to the enzyme in its *synperiplanar* or *antiperiplanar* form. Thus, we analyzed the compounds maleate and fumarate, which are the fixed constrained isomers of succinate in their *cis* and *trans* forms, respectively. From this we identified that maleate exhibited uncompetitive inhibition of ICL1 versus isocitrate, consistent with its binding to the succinate site with a 2-fold improvement in affinity (maleate: K_{ii} = 300 ± 20 μM; succinate: K_{ii} = 650 ± 50 μM). In the reverse reaction of ICL1, maleate was a competitive inhibitor versus succinate (K_{is} = 190 ± 10 μM). Accordingly, maleate binds exclusively to the ICL1-Glx complex. The ratio of K_{ii}/K_{is} provides a measurement of the desorption rate of glyoxylate (k_7) from ICL1-Glx with

respect to k_{cat} of the forward reaction ($K_{ii}/K_{is} = k_7/k_{cat} = 1.6 \pm 0.2$). Consequently, the release of glyoxylate from ICL1-Glx is likely the rate-limiting step for the forward reaction, so that at saturating concentrations of isocitrate the predominant enzyme species will be ICL1-Glx, to which succinate, maleate, and their analogs bind.[210, 211] Conversely, the fixed *trans*-isomer of maleate, fumarate, was a poor competitive inhibitor versus isocitrate ($K_{is} = 1.9 \pm 0.4$ mM), suggesting that the *cis*-isomer of but-2-en-1,4-dioic acid is preferred, and that succinate binds to the ICL1-Glx complex in a synperiplanar conformation. This encouraged the evaluation of other compounds which maintained the *synperiplanar* conformations of succinate, such as (2R,3S)-epoxy-succinate (*cis*-EpS), as potential covalent inactivators of ICL. Incubation of ICL1 with *cis*-EpS in the presence of 200 μM isocitrate (~5K_M) demonstrated time-dependent loss of enzyme activity commensurate with increasing concentrations of *cis*-EpS (0.16 – 2.0 μM) as shown in Figure 1b. The time courses of glyoxylate formation at each concentration of *cis*-EpS were fitted to eq. 3, from which replotting of the apparent rate constants of inactivation (k_{obs}) versus [*cis*-EpS] using eq. 4 indicated hyperbolic binding (inset, Figure 31.). From this replot we obtained values of: $k_{inact} = (1.09 \pm 0.009) \times 10^{-2}$ s^{-1} and $K_{inact} = 0.7 \pm 0.1$ μM. These data indicated saturation of *cis*-EpS at sub-micromolar concentrations, a signature of a two-step inactivation mechanism. Global nonlinear regression analysis was conducted by fitting the time-course data in Figure 1b to eq. S11 ([isocitrate] = 200 μM). This resulted in values of $k_{inact} = (9.2 \pm 0.4) \times 10^{-3}$ s^{-1}, $K_{inact} = 0.62 \pm 0.03$ μM, and $K_{i1} = 2.0 \pm 0.2$ μM. Similarly, time-dependent inactivation of ICL2 by *cis*-EpS under the same conditions

resulted in a diminished binding affinity and a lower apparent rate of inactivation compared to ICL1 (apparent K_{inact} = 35 ± 2 μM and k_{inact} = (1.4 ± 0.1) x 10^{-3} s^{-1}).

In contrast, *trans*-2,3-epoxy-succinate (*trans*-EpS) exhibited competitive inhibition of ICL1 versus isocitrate (K_{is} = 0.8 ± 0.2 mM), but was not an inactivator. After overnight incubation with up to 100 μM *trans*-EpS, no effect was observed on the catalytic activity of ICL1. Hence, *trans*-EpS binds preferentially to free ICL1, similar to that of the competitive inhibitor (*2R)*-malate.[215] The value of K_{is} of racemic *trans*-EpS was twice that of (2*R*)-malate, suggesting that only the (*R*,*R)*-isomer of *trans*-EpS binds to ICL1. The differential binding of *trans*-EpS and *cis*-EpS to the respective free ICL1 and ICL1-glyoxylate complexes suggested that these geometric isomers structurally resemble different features of the substrate isocitrate, wherein *trans*-EpS more closely mimics the substrate isocitrate and *cis*-EpS serves as a succinate analog, reminiscent of the different sub-site binding by inhibitors glycolate and maleate.

We next evaluated the effect of isocitrate concentration (25-200 μM) on the values of k_{obs}. At 0.25 μM *cis*-EpS the k_{obs} of ICL1 inactivation reached an apparent maximal k_{obs} value as fixed concentrations of isocitrate exceeded 6K_m (Figure 31.) providing additional evidence that *cis*-EpS acts as an uncompetitive inactivator of ICL1. In contrast, increasing concentrations of succinate (0 – 4.0 mM) resulted in a hyperbolic decrease in apparent values of k_{obs} (Figure 31.), highlighting that the binding of succinate to the ICL1-Glx complex blocks the binding of *cis*-EpS, and the diminution of k_{obs} arises from apparent saturation of the ICL1-Succ-Glx complex. The mode of inactivation of *cis*-EpS in the

presence of isocitrate and succinate supports the conclusion that *cis*-EpS preferentially binds to the ICL1-Glx complex.

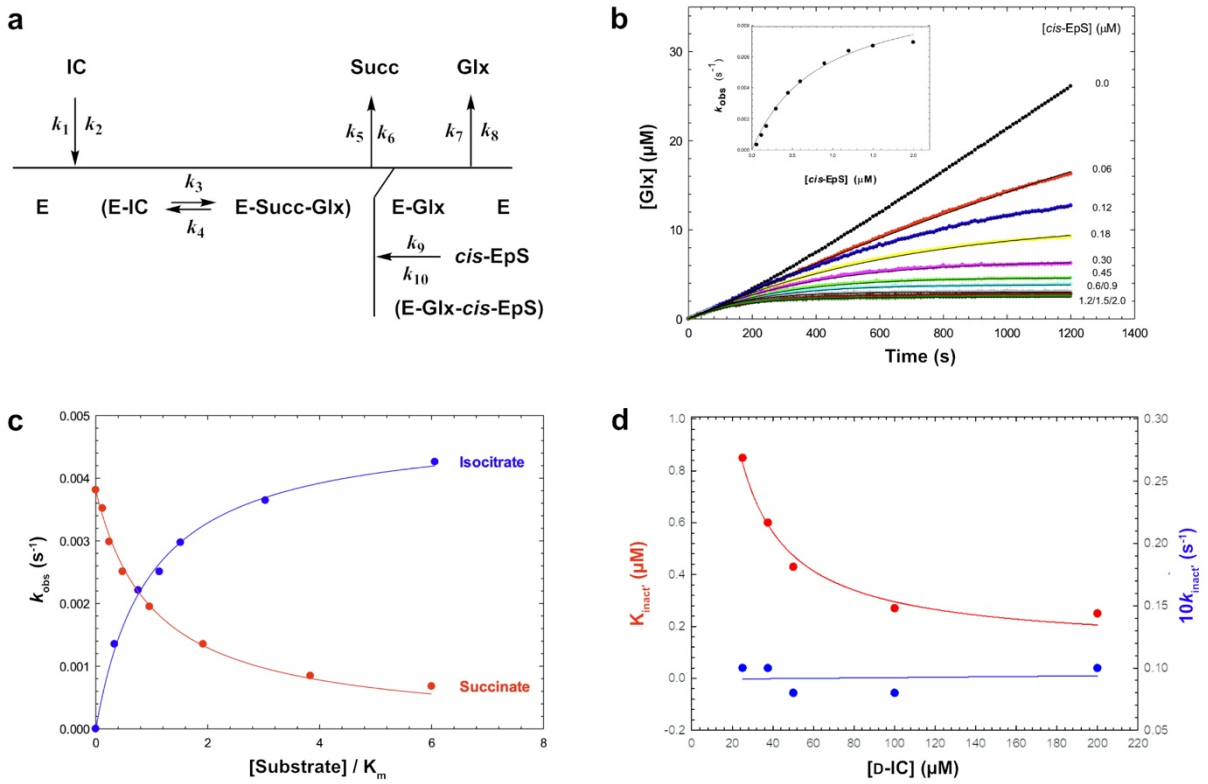


Figure 31. Proposed Kinetic Mechanism of ICL1 and Time-Dependent Inactivation of ICL1 with *cis*-EpS

a) Ordered uni-bi kinetic mechanism of ICL1. The E-Glx-*cis*-EpS complex proceeds to form a covalent adduct. **(b)** Time-dependent inactivation of ICL1 by variable concentrations of *cis*-EpS in the presence of 200 μM D-isocitrate and 2 nM ICL1. The curves drawn through the experimental data points were obtained by fitting data to eq. 3 for each concentration of *cis*-EpS. (Inset) Replot of the observed rates of inactivation (k_{obs}) were fitted to eq. 4, yielding: $k_{inact} = (1.09 \pm 0.009) \times 10^{-2}$ s^{-1} and $K_{inact} = 0.7 \pm 0.1$ μM. **(c)** Inactivation of ICL1 by *cis*-EpS in the presence of fixed concentrations of isocitrate or succinate. Values of k_{obs} obtained from time courses at 0.25 μM *cis*-EpS at changing-fixed concentrations of isocitrate (blue) and succinate (red), plotted as $[S]/K_M$, were fitted to eqs. 5 and 6, respectively. **(d)** Replots of the apparent values of k_{inact} and K_{inact} values obtained at variable concentrations of isocitrate. The curve drawn through the data points for the plot of K_{inact}, vs. [D-isocitrate] is from linear regression via fitting to eq. 7, while k_{inact}, vs. [D-isocitrate] was fitted to a straight line.

Apparent values of k_{obs} versus [isocitrate] and [succinate] in Figure 1c were fitted to eqs. 5 and 6, respectively, for which eq. 5 describes an increase in the rate of inactivation as isocitrate concentration is increased, akin to uncompetitive inhibition, while eq. 6 describes its decrease as the concentration of succinate is increased, akin to competitive inhibition. Fitting of k_{obs} versus [isocitrate] data to eq. 7 led to limiting values of k_{inact} = (9.5 ± 0.5) x 10^{-3} s^{-1}, K_{inact} = 260 ± 30 nM, and k_{inact}/K_{inact} = 36,000 ± 6,000 M^{-1} s^{-1}. As shown in Figure 31., apparent values of k_{inact}' and K_{inact}', obtained from global fitting of the time-course data with various concentrations of *cis*-EpS at 25-200 μM isocitrate demonstrated that the value of K_{inact}', but not k_{inact}', was dependent on the concentration of isocitrate. Upon extrapolation of concentrations of isocitrate to infinity, the limiting values of the kinetic parameters became K_{inact} = 250 ± 10 nM, k_{inact} = 1.0 ± 0.1 x 10^{-2} s^{-1}, and k_{inact}/K_{inact} = 40,000 ± 6,400 $M^{-1}s^{-1}$. These kinetic parameters of inactivation are the largest yet observed for an inactivator of ICL1.

ii. Inactivation of ICL1 by cis-EpS is Potentiated by Added Glyoxylate or Glycolate

To elucidate the role of glyoxylate in *cis*-EpS inactivation of ICL1, we monitored the time-dependent change in intrinsic protein fluorescence of ICL1. This fluorescence change likely arises in part from the closure of the active site loop of (Trp_{93} is proximal to the bound glyoxylate), which moves approximately 15 Å upon ligand binding,[225] allowing one to monitor substrate, inhibitor, or inactivator binding to ICL1.[215] In the presence of *cis*-EpS alone, mixed with ICL1 at a 1:1 molar ratio, we observed a slow decrease in fluorescence intensity, (half-life = 53.3 min, Figure 32a, c) indicative of active site loop

closure. Conversely, supplementation of 100 μM glyoxylate to ICL1 and *cis*-EpS at a 1:1 molar ratio resulted in a more rapid loss of signal (half-life = 5.4 min, Figure 32b, c). Analysis of ICL1 inactivation with or without glyoxylate (Figure 32c), demonstrated that the rate of inactivation upon glyoxylate addition is 10-fold faster (k_{obs} = 2.2 x 10^{-3} s^{-1} vs. k_{obs} = 2.2 x 10^{-4} s^{-1}). This was further corroborated by kinetic analysis of *cis*-EpS inactivation of ICL1 with the glyoxylate analog, glycolate (Glc), which when added at 100 μM with *cis*-EpS, exhibited a 8.5-fold faster apparent rate of inactivation when compared to its absence (not shown). These data indicate that the ternary ICL1-Glx/Glc-*cis*-EpS complex is far more favorable to afford reaction of the epoxide of *cis*-EpS with Cys_{191} than the binary ICL1- *cis*-EpS complex.

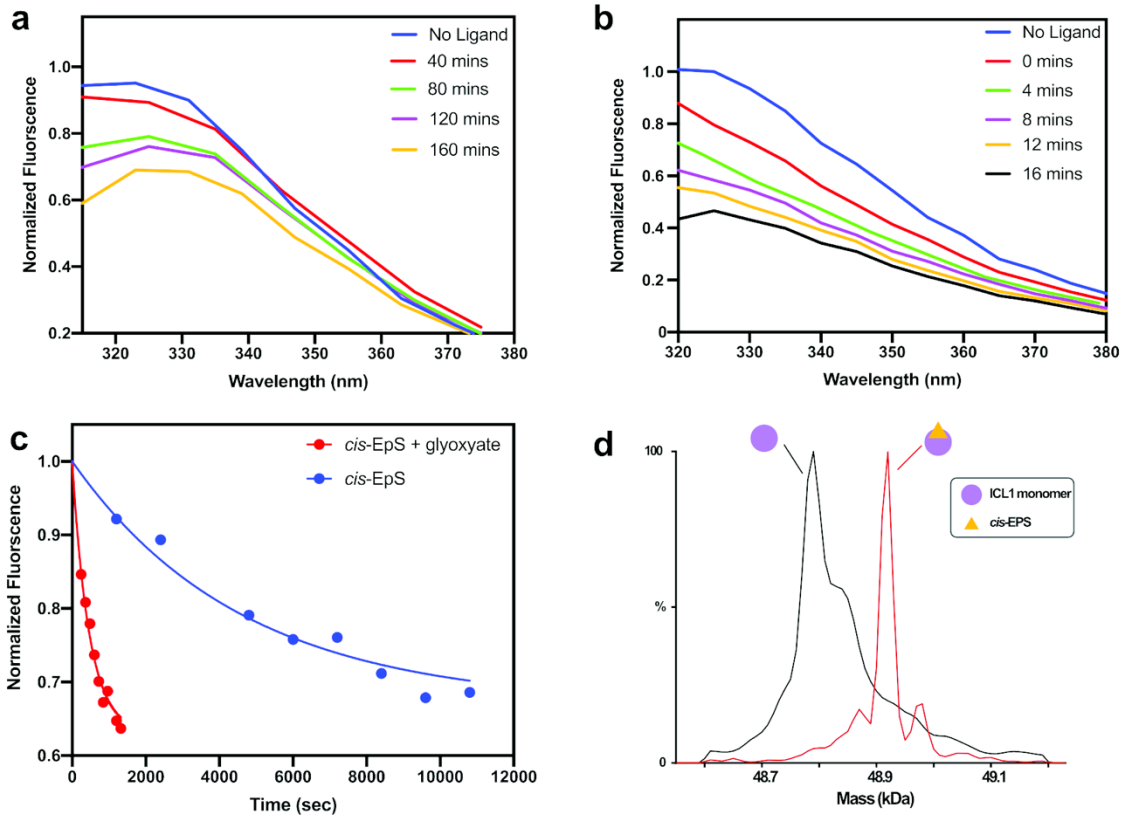


Figure 32. ICL1 Inactivation is Potentiated by Glyoxylate and is Covalent and Irreversible

(a) 1 µM ICL1 in the absence of added co-ligands (blue line) displayed fluorescence maxima at λ_{max} = 325 nm, which is slowly diminished upon binding of 1 µM *cis*-EpS alone in a time-dependent fashion. **(b)** Rapid time-dependent loss in fluorescence of 1 µM ICL1 in the presence of 1 µM *cis*-EpS and 100 µM glyoxylate. **(c)** Replots of the apparent time-dependent inactivation kinetics exhibited by *cis*-EpS from the decrease in intrinsic tryptophan fluorescence in (a) and (b). Curves drawn through the data points were obtained by fitting the data to $F_t = (F_0 - F_{min})*\exp(-k_{obs}*t) + F_{min}$ yielding apparent k_{obs} rates of 2.2 x 10^{-3} sec^{-1} in the presence of *cis*-EpS and glyoxylate, and 2.2 x 10^{-4} sec^{-1} in the presence of *cis*-EpS alone. **(d)** Deconvoluted mass spectra of WT ICL1 versus ICL1 incubated with *cis*-EpS overnight at 25°C. Theoretical and measured mass of monomer bound *cis*-EpS are 48919.07 Da and 48918.32 Da (15.26 ppm), respectively, consistent with the addition of one molecule of *cis*-EpS (132.01 Da) per monomer of ICL1 (48,787.52 Da). Mass spectrometry analysis was conducted in collaboration with Zahra Moghadamchargari in the laboratory of Dr. Arthur Laganowsky.

iii. Inactivation of ICL1 by cis-EpS is Irreversible

Pre-incubation of ICL1 (monomer concentration of 0.5 µM) with *cis*-EpS (0-2 µM) with or without added glycolate (100 µM) for 20 h revealed stoichiometric loss of ICL1 activity (molar ratio of ≥ 1:1) (data not shown). Further, repeated spin-dialysis of inactivated enzyme (4 rounds over 72 h, equivalent to a 12,500-fold dilution) to remove unbound *cis*-EpS did not lead to any appreciable recovery of enzyme activity. Therefore, the inactivation was apparently irreversible, and the molar ratio of inactivator:ICL1 monomer was 1:1 (data not shown). Co-incubation of 1 µM ICL1 and 1 µM *cis*-EpS with

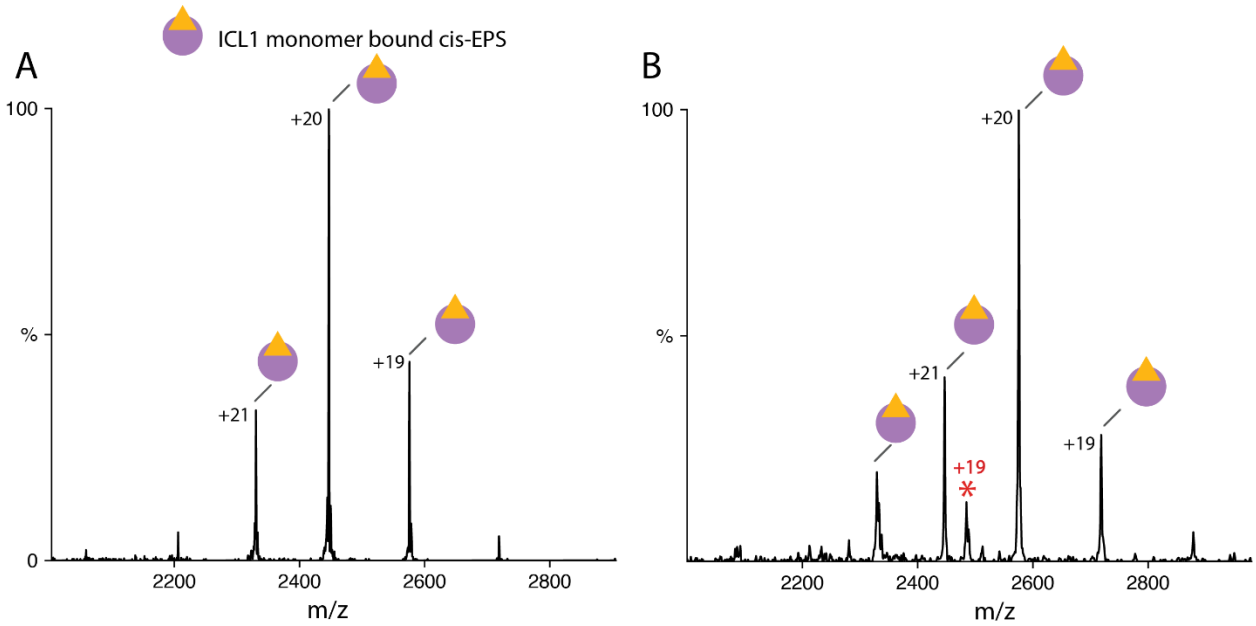


Figure 33. Mass Spectra for cis-EpS Bound ICL1 Monomer Treated with Glycolate.

The mass spectra shown for cis-EPS bound ICL1 monomer (A) treated in the presence of 1 mM glycolate, and (B) treated in the presence of 1 mM glycolate after a four-day incubation at 4 0C. The star indicates a truncated monomer with a measured mass of 47190 Da. Mass spectrometry analysis was conducted in collaboration with Zahra Moghadamchargari in the laboratory of Dr. Arthur Laganowsky.

1 mM DTT or 1 mM reduced glutathione (pH 7.5, $t \leq 20$ hrs) caused no change in the ability of *cis*-EpS to inactivate ICL1, or the reactivation of ICL1 activity, indicating stability of the ICL1 adduct to free thiols.

The mode of binding of *cis*-EpS to ICL1 was investigated using high-resolution mass spectrometry (MS). Both denatured MS and collision-induced dissociation (CID) of tetrameric ICL1 following incubation of ICL1 with *cis*-EpS showed covalent addition of *cis*-EpS to ICL1 at a 1:1 molar ratio (Figure 32d, Figure 33). Further, the modification of ICL1 by *cis*-EpS was observed in the same 1:1 ratio in presence of the competitive inhibitor glycolate, even following four days of incubation at 4°C (Figure 33). Together, these results demonstrated that *cis*-EpS inactivation of ICL1 is covalent and irreversible. The uniformity and efficiency of this modification makes *cis*-EpS a possible active site titrant not only for ICL1 from *Mtb*, but for other isocitrate lyases as well.

iv. Cis-EpS Inhibits Growth of Mycobacterium tuberculosis in Cell Culture

In collaboration with Dr. Inna Krieger in Dr. James Sacchettini's lab, we identified that treatment of M. tuberculosis (mc2 7000 strain) cell culture (grown on acetate as a sole carbon source) with cis-EpS, resulted in an MIC_{50} of 100 ± 10 μM (Figure 34). No inhibition of M. tuberculosis growth by *cis*-EpS was observed when cultures were grown on dextrose as a carbon source, indicating that the effect of cis-EpS was specific to the growth conditions where ICL activity has been shown to be essential for the survival of Mtb.[203] In addition, at a similar concentration range tested, trans-EpS had no effect on mycobacterial growth.

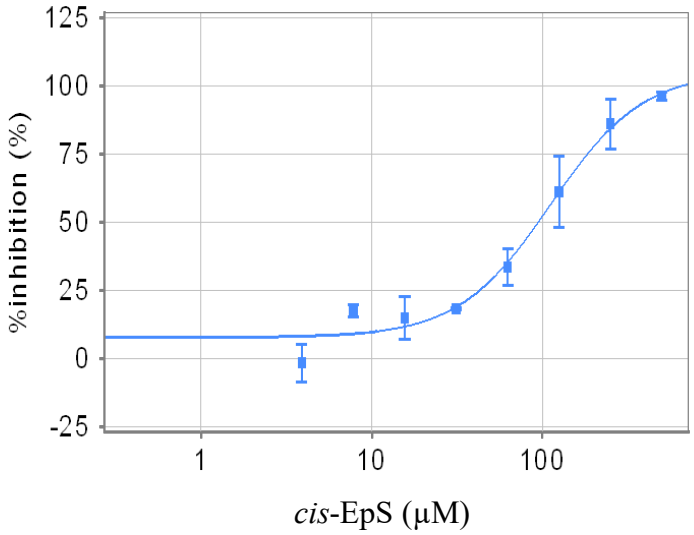


Figure 34. Inhibition of *Mtb mc² 7000* Growth by *cis*-EpS

Effects of *cis*-EpS on *Mtb mc² 7000* grown using acetate as a carbon source. The dose-response curve of c*is*-EpS-induced growth inhibition was fitted to equation $y = 100 - 100/(1+ [I]/IC_{50})$ where y is normalized inhibition (%) and [I] is the concentration of *cis*-EpS (x-axis) and IC_{50} represents the concentration of *cis*-EpS that achieves 50% inhibition (IC_{50} = 100 ± 10 µM).

v. Crystal Structures of ICL1 Treated with cis-EpS

Tag-free ICL1 was co-crystallized with *cis*-EpS, and crystals belonging to the $P2_12_12_1$ space group with the tetramer of ICL1 in the asymmetric unit were produced and diffracted to high resolution (1.88Å, Table 6). In all four subunits of the treated ICL1 tetramer, electron density corresponding to a covalent adduct attached to Cys_{191} was present. This adduct apparently resulted from the attack of the sulfur of Cys_{191} on the 2*R*-carbon of (2*R*,3*S*)-*cis*-EpS, with attending ring-opening of the epoxide to form the S-

malyl-Cys_{191} (Mal-Cys_{191})-ICL1 complex (Figure 35a). The S-malylated active sites of ICL1 also contained a second ligand, bound to the active site Mg^{2+} ion. We modeled it as glyoxylate, its presence likely resulting from the addition of isocitrate to ICL1 during the thrombin-catalyzed removal of the His6 tag. After refinement, a positive electron density

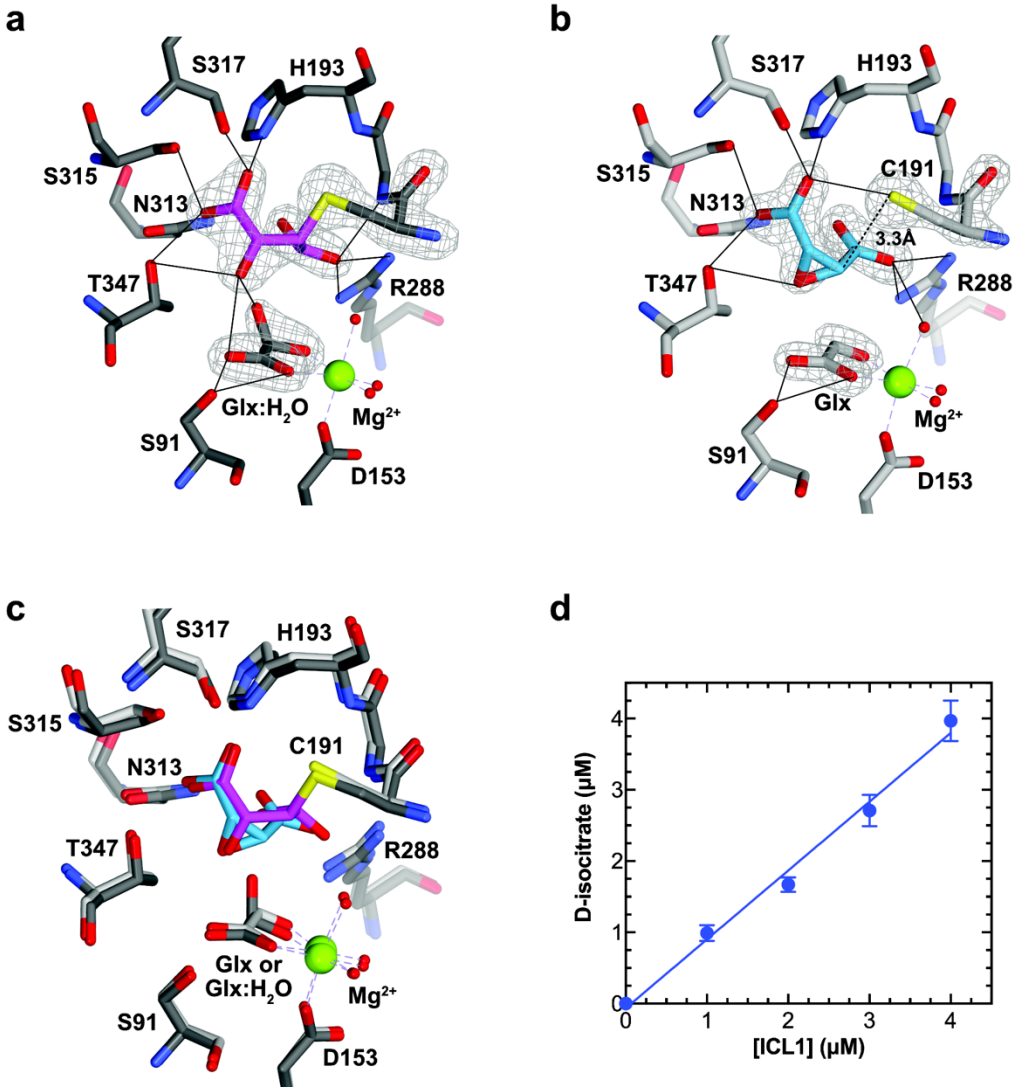


Figure 35. Crystallographic Analysis of ICL1 Inactivation by *cis*-EpS.

(a) Crystal structure of ICL1 covalently inactivated by *cis*-EpS. *Cis*-EpS-treated ICL1 resulted in a S-malatyl-Cys$_{191}$ adduct (Mal-Cys$_{191}$) shown in magenta, surrounded by proximal protein residues (dark gray carbons); oxygens are shown in red, nitrogens in blue, sulfur in yellow and Mg^{2+} in green. Hydrated glyoxylate (Glx-OH_2) is coordinated to the Mg^{2+} ion. Solid lines indicate potential hydrogen bonds and electrostatic interactions. Polder omit electron density maps of glyoxylate hydrate and the Mal-Cys$_{191}$ adduct are shown as gray mesh and contoured at 7σ. **(b)** Crystal structure of ICL1 with non-covalently bound *cis*-EpS shown in blue, surrounded by proximal protein residues (light gray). Polder omit electron density maps of glyoxylate, *cis*-EpS, and Cys$_{191}$ are shown in gray and contoured at 10, 7, and 8 σ respectively. **(c)** Overlay of the ICL1-*cis*-EpS covalent structure (contact-making residues, carbons in dark grey and ligand in magenta) with the non-covalently bound *cis*-EpS structure (contact making residues, carbons in light grey and ligand in blue). **(d)** Formation of D-isocitrate by incubating varied concentrations of ICL1 in the presence of saturating succinate.

peak appeared off the C2 carbon of refined glyoxylate. Similar electron density coordinated to Mg^{2+} was observed by Ray *et al.* in their 3-NP modified ICL1 structure (PDB: 6C4A),[234] which was modeled as pyruvate. However, based on the fact that our sample was exposed to isocitrate, which would yield glyoxylate upon cleavage, and in accordance with our hypothesized reaction mechanism (Figure 37), we modeled Mg^{2+}-bound glyoxylate as glyoxylate hydrate which refined well into the electron density (Figure 35a).

The Mal-Cys_{191} adduct occupies the succinate sub-site, and like succinate, the malyl group is positioned to form hydrogen bonds with NH1 and NH2 of Arg_{228}, OG1 of Thr_{347}, ND2 of Asn_{313}, OG of Ser_{315}, OG of Ser_{317}, ND1 of His_{193}, the backbone nitrogen of Gly_{192}, OH of glyoxylate hydrate, and Mg^{2+}-coordinated water (Figure 35a). These hydrogen-bonding interactions with other residues of the active site of ICL1 are similar to those observed between 3-nitropropionate and ICL1 (PDB: 1F8I). Since Moynihan and Murkin[211] showed that Cys_{191} is a thiolate when glyoxylate is bound to ICL1, where its apparent catalytic role is to deprotonate succinate, it is likely that the thiolate form of Cys_{191} attacks *cis*-EpS.

To investigate the interaction of *cis*-EpS with ICL1 in the absence of glyoxylate, we produced crystals of the *cis*-EpS-treated ICL1 that were prepared with no exposure to substrates or products of ICL1 at any step of purification or removal of the H_6 tag. To our surprise, we discovered that the active sites of all four subunits not only contained intact *cis*-EpS bound non-covalently to ICL1, but also exhibited electron density for an apparent two-carbon keto acid, reminiscent of glyoxylate, coordinated to the Mg^{2+} ion (Figure 35b).

Inspection of the electron density after refinement, revealed a positive peak of ~ 4σ in the F_o-F_c electron density map between the Cys_{191} and 2*R*-carbon of the *cis*-EpS, demonstrating that trace amounts of the inactivator reacted with ICL1, but the majority of protein molecules are bound to intact *cis*-EpS. Attempts to refine a covalent adduct between Cys_{191} and *cis*-EpS resulted in a negative peak of 4σ in the F_o-F_c electron density map, indicating that the intact *cis*-EpS molecule was the best-fitted ligand. This intact *cis*-EpS molecule formed nearly identical hydrogen bonds to those seen in the covalently-bound *cis*-EpS structure (Figure 35 a, b). The 2*R*-carbon of the intact *cis*-EpS is positioned 3.3 Å away from the sulfur of active site Cys_{191}, but nevertheless remains sufficiently proximal to the thiolate for attack on the epoxide (Figure 35b). In contrast to the ICL1-Mal-Cys_{191} structure, owing potentially to the absence of a hydroxyl group of Mal-Cys_{191}, no hydrogen bonds were formed between the bound two-carbon keto-acid and intact *cis*-EpS. To determine the identity of the keto-acid co-ligand coordinated to ICL1, we subjected the protein to high-resolution native mass spectrometry. Analysis of purified recombinant ICL1 showed that the homo-tetrameric complex had an additional mass of 310 Da per tetramer compared to the calculated mass of the ICL1 tetramer (195,148 Da vs. 195,461.78 Da), equating to an addition of 77 ± 17 Da (90.38 ppm) per subunit (Figure 36B). When the tetrameric complex was dissociated in the gas phase, the measured mass of one subunit, 48,787.94 Da, was nearly equal to the calculated mass of un-liganded ICL1, 48,787.52 Da, suggesting the ligand bound to the protein non-covalently. This mass shift per subunit is in agreement with the mass of glycolate (76.02 Da) and glyoxylate

(74.00 Da), both of which can coordinate to the Mg^{2+} of ICL1, and are also present in the recombinant expression host, *E. coli*.

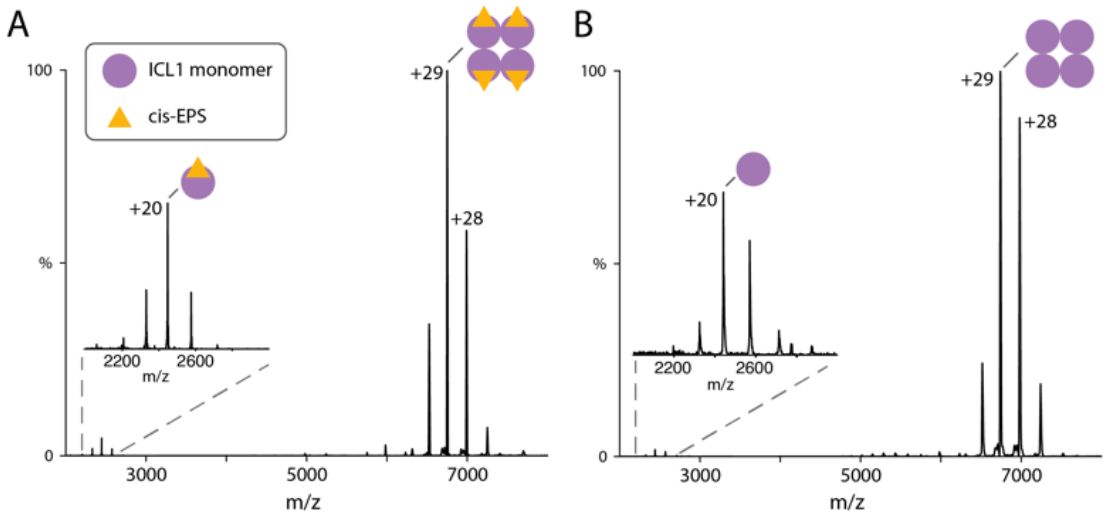


Figure 36. Nano-ESI-MS of Tetrameric ICL1 Obtained Under Activating Energy Conditions.

The collision-induced dissociation spectra of (A) ICL1 in the presence of *cis*-EpS and (B) free ICL1. The mass spectra shown in **A** demonstrates that ICL1 is bound to *cis*-EpS. ICL1 was incubated with *cis*-EpS overnight, buffer exchanged into 200 mM ammonium acetate (pH 7.4), and then introduced into a Thermo Exactive Plus system with an extended mass-range Orbitrap mass spectrometer. The inset is the expansion of peaks at lower values of m/z which show ejected monomers. The theoretical and observed masses of free and *cis*-EpS-bound ICL1 are 48919.07 and 48918.32 (15.26 ppm) Da, respectively. **B**. The tetramer of ICL1 has a calculated mass of 195.148 Da, whereas the observed mass was 195,461.78 Da, indicating an addition of 77±17 Da (90.38 ppm) per monomer. When dissociated into monomers the observed mass was 48,787.94 Da, nearly equal to the calculated mass of an unmodified ICL1 monomer, 48,787.52 Da, suggesting that a non-covalent ligand is bound to the protein. This mass spectrometry analysis was conducted in collaboration with Zahra Moghadamchargari in the laboratory of Dr. Arthur Laganowsky.

To test if this molecule was glyoxylate, we monitored the formation of D-isocitrate upon succinate addition, quantifiable by a coupled assay which reports on isocitrate formation. We observed formation of D-isocitrate that approximated a 1:1 molar ratio with

respect to the ICL1 monomer concentration, identifying the molecule as glyoxylate (Figure 35d). These biochemical data agree with our mass spectrometry and structural data; under our purification conditions, recombinant ICL1 apparently retains bound glyoxylate, which enables *cis*-EpS to bind to the ICL1-Glx complex, as per the ordered mechanism. Further, the limited prevalence of the ICL1-*cis*-EpS covalent adduct, and the ability to capture this transient intermediate *in crystallo*, may arise from the diminished rate of inactivation in the absence of exogenous glyoxylate as seen in Figure 32 a, c. This slowed rate suggests an altered protonation state of Cys_{191} and/or differing accessibility to a water molecule that is required for *cis*-EpS inactivation of ICL1. Close inspection of the F_o-F_c electron density corresponding to the C2 of glyoxylate in the non-covalent *cis*-EpS structure revealed a weak positive density peak that was not present in all chains. Conversely, the covalent structure exhibited a strong positive F_o-F_c electron density peak present near the C2 of glyoxylate, consistent with a hydrated form of glyoxylate, suggestive that the appearance of this density occurs concomitantly with alkylation of Cys_{191} by *cis*-EpS. Taken together, these two structures provide direct insight into the

inactivation of ICL1 by *cis*-EpS, and allow one to propose plausible chemical mechanisms by which this inactivation may proceed (Figure 37).

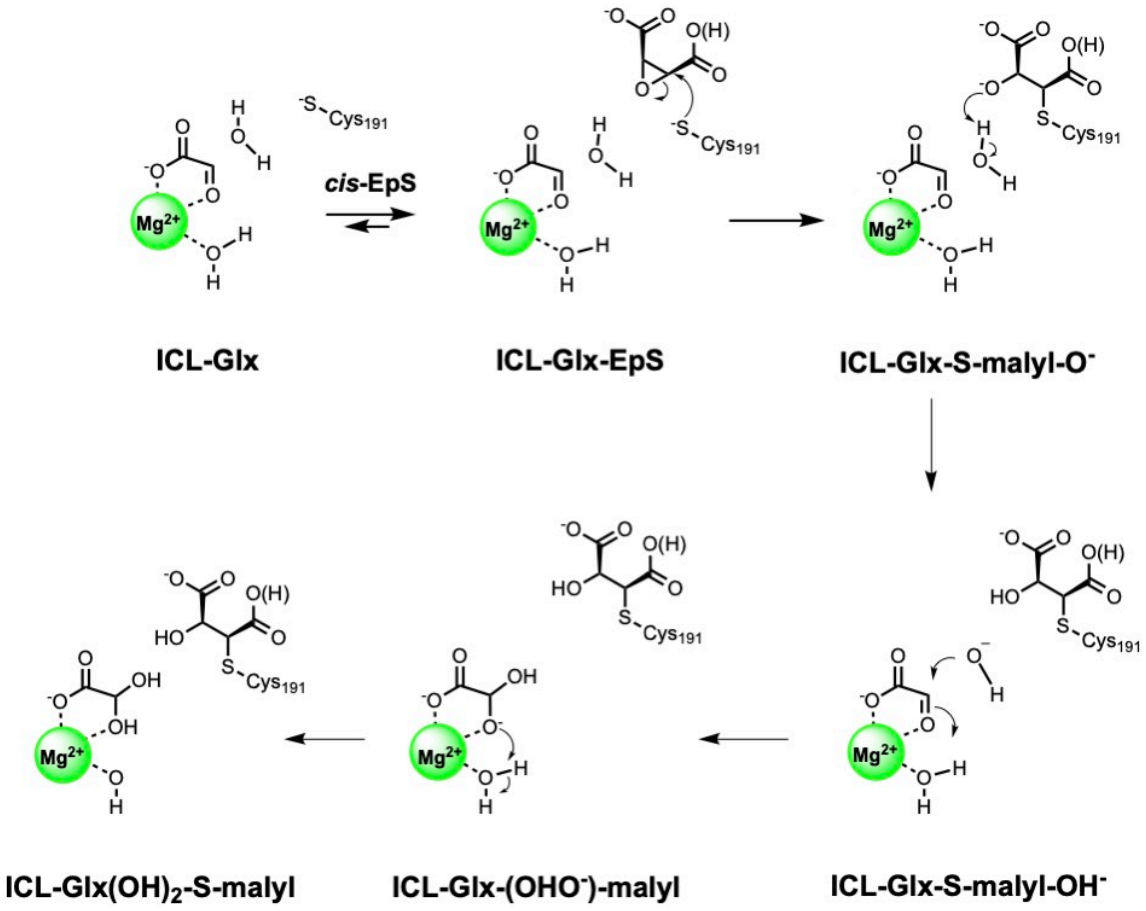


Figure 37. Proposed Chemical Mechanism of *cis*-EpS Inactivation of ICL1

vi. Proposed Chemical Mechanism of cis-EpS Inactivation of ICL1

Nucleophilic attack of epoxides generally involves one of two chemical mechanisms (a) protonation of the epoxide ring oxygen, followed by attack of the

nucleophile on the more sterically-hindered carbon, as a result of the enhanced stabilization of an incipient carbocation at the more substituted carbon, or (b) direct attack of the nucleophile on the least sterically-hindered carbon of the epoxide ring.[237]

The protonation of an epoxide oxygen prior to nucleophilic attack lowers the activation energy of the nucleophilic addition, as was shown for HIV protease to allow efficient covalent reaction to proceed both computationally[238] and experimentally with 1,2-epoxy-(4-nitrophenoxy)propane.[59] However, in lieu of ring oxygen protonation, it has been suggested by Meara and Rich[239] that hydrogen-bonding of the epoxide-ring oxygen of the *trans*-epoxysuccinylamide group of an E-64 analog is sufficient to induce reaction with the Cys_{25} thiolate of papain. Alternately, E-64 inactivation of a similar cysteine protease, cathepsin B, has been proposed to use a proton relay between His_{199} and a water molecule to protonate the alkoxide formed following ring opening of the epoxide.[240] In support of direct thiolate attack onto the epoxide, model reaction studies identified that the thiolate of methanethiol, a formidable nucleophile mimicking the thiolate of cysteine, reacts with the anionic oxirane 2-carboxylate ($k = 0.01$ min^{-1}),[241] albeit at a 60-fold slower rate than that of *cis*-EpS inactivation of ICL1 ($k_{inact} = 0.6$ min^{-1}). In addition, direct attack of the Cys_{115} thiolate of MurA onto the epoxide antibiotic fosfomycin results in highly efficient alkylation of the enzyme (k_{inact} 7.4 min^{-1}), underscoring the ability of an active site thiolate to react efficiently with an unactivated expoxide.[87]

When glyoxylate is bound to ICL1, we propose that the chemical mechanism of inactivation by *cis*-EpS proceeds by the mechanisms shown in Figure 37, similar to that proposed for the mechanism of inactivation of cathepsin B by E-64 or MurA by

fosfomycin for which there is no pre-protonation of the ring oxygen of the epoxide. Initially glyoxylate is coordinated to the Mg^{2+} of ICL1 as suggested by native-protein mass spectrometry, structural, and kinetic data.[211, 242] Glyoxylate likely binds to the protein during its heterologous expression in *E. coli*, and remained bound during the purification process. Subsequently, the binding of *cis*-EpS to the succinate site of the enzyme results in loop closure of the active site, and formation of hydrogen bonds with the carboxylate groups of *cis*-EpS in a nearly analogous fashion to that of succinate. Based on our crystallographic data, the presence of glyoxylate hydrate in the *cis*-EpS modified structure (Figure 35a) indicates that water may be involved in the inactivation of ICL1 by *cis*-EpS. Therefore, we propose that thiolate attack of Cys_{191}-S^- onto the *pro*-R carbon of *cis*-EpS results in alkylation of ICL1 and ring opening of the epoxide ring. The alkoxide produced from ring opening then abstracts a proton from an active site water molecule to form the S-malylated product and a hydroxide ion (Figure 37). Subsequent attack of the hydroxide ion onto the proximal glyoxylate aldehyde, followed by proton abstraction from a metal-bound water forms glyoxylate hydrate as seen in our crystal structures. This metal-bound water has previously been proposed to protonate the C2 hydroxyl group of D-isocitrate in the direction of D-isocitrate synthesis.[211, 242] This C2 hydroxyl group of D-isocitrate is derived from glyoxylate in this synthesis, and as such, it is likely that the deprotonated oxygen of glyoxylate produced upon hydroxide addition could be protonated by the same metal-bound water. An additional possibility is that glyoxylate becomes hydrated prior to epoxide ring opening, and one of its *gem*-hydroxyl groups could protonate the alkoxide of the ring-opened *cis*-EpS. However, this mechanism is less consistent with our

crystallographic snapshot data (Figure 35b), wherein the epoxide ring is closed, but a small subset of the density also shows alkylation of ICL1 by *cis*-EpS. The observed densities seen on the C2 of glyoxylate and between Cys_{191} and *cis*-EpS are present in similar magnitudes, suggesting that alkylation and hydration likely occur in concert as proposed in Figure 37.

Alternately, when the bound co-ligand is glycolate, we propose that the S-malylation of ICL1 proceeds under the same mechanism seen in Figure 5, except that the hydroxide ion product is unable to form an adduct with glycolate, but could potentially react with and become a ligand of the divalent magnesium ion (not shown). Otherwise, these alkylation mechanisms are identical, which is in accord with the identical values of k_{inact} (0.01 s^{-1}) for *cis*-EpS inactivation ascertained at 2-fold the values of K_m for D-isocitrate (effectively the E-glyoxylate complex) and K_{is} for glycolate, while K_{inact} for glycolate is nearly twice that observed for the E-glyoxylate complex. These results suggest that the attack of the Cys_{191} thiolate on *cis*-EpS comprises the slowest step of ICL1 inactivation regardless if glyoxylate or glycolate is bound to the enzyme.

D. Conclusions and Future Outlook

We have discovered that *cis*-EpS is a potent and selective inactivator of *Mtb* ICL1 as determined by kinetics, mass spectrometry, protein fluorescence, and crystallographic characterization. Kinetic studies demonstrated that time-dependent inactivation of ICL1 by *cis*-EpS is 35-fold and 750-fold, more efficient than 3-bromopyruvate and 2-vinyl isocitrate, respectively. Indeed, *cis*-EpS is apparently the most potent inhibitor or

inactivator yet discovered for *Mtb* ICL1 (limiting values of: K_{inact} = 250 nM, k_{inact} = 0.01 s^{-1}). We found that ICL1 co-purifies with glyoxylate, explaining why inactivation of the enzyme is observed even in the absence of exogenous glyoxylate, albeit at a slower rate. High-resolution crystal structures of ICL1 both alkylated by *cis*-EpS and bound to intact *cis*-EpS provided insight into the mechanism of the enzyme inactivation which likely requires activation of the *cis*-EpS epoxide ring oxygen to afford effective reaction with the thiolate of Cys_{191} of ICL1.

Cis-EpS also inhibited *Mtb* cell growth in M9-acetate media, albeit at relatively higher concentrations (MIC_{50} = 100 μM) (Figure 34). Even though the anionic nature of *cis*-EpS probably led to limited cellular uptake of the compound, exploration of amide and ester derivatives of *cis*-epoxy-succinate is ongoing to identify ICL1 inactivators of higher cellular permeability and improved anti-mycobacterial activity, while maintaining the apparent selectivity of *cis*-EpS.

CHAPTER IV

MECHANISM-BASED INACTIVATION OF *MYCOBACTERIUM TUBERCULOSIS* ISOCITRATE LYASE 1 BY 5-DESCARBOXY-5-NITRO-D-ISOCITRIC ACID

A. Introduction

As detailed in Chapter III, *Mycobacterium tuberculosis* (*Mtb*), the etiological agent of tuberculosis (TB), currently afflicts a quarter of the world's population, and is the leading cause of death from an infectious disease.[146] The success of this pathogen can be attributed to many factors, but one of its most impressive feats lies in its ability to persist in human macrophages, often undetected for decades, in a state of metabolic dormancy.[243] In this carbon-limited environment, *Mtb* consumes lipids, imported cholesterol, and other carbon sources and nutrients it can scavenge or import from its host.[200, 244-246] This shift in nutrients results in metabolic remodeling of *Mtb* and strong upregulation of *icl1*, the gene which encodes isocitrate lyase-1.[207, 217] This enzyme, and its ortholog ICL2, catalyze the retro-aldol cleavage of D-isocitrate (D-IC) to produce glyoxylate (Glx) and succinate (Succ), and of 2-methylisocitrate to form pyruvate and succinate (Figure 38A).[213, 221, 225] These reactions shift the flux of the TCA cycle through the anaplerotic glyoxylate shunt, and activate the methylcitrate cycle, respectively.[203, 218] Deletion of *icl1* and *icl2* has been shown to clear *Mtb* infection from the lungs of mice,[216] and also to enhance the efficacy of front-line anti-tubercular drugs, highlighting a possible role of the ICLs in multi-drug and extreme-drug resistant *Mtb*.[205, 216] Thus, the essentiality of *icl1* and *icl2* for the

persistence of *Mtb* infection and their absence in humans makes the ICLs attractive and validated drug targets.

Mechanism-based inactivators of enzymes comprise a class of molecules that structurally mimic the substrate of an enzyme, but through inadvertent catalysis are transformed into electrophilic species which inactivate the target enzyme by the formation of a covalent adduct.[7] We previously characterized 2-vinyl-D-isocitrate as a highly efficient mechanism-based inactivator of the *Mtb* ICLs.[215] Upon catalysis, the retro-aldol cleavage of the C2-C3 bond of 2-vinyl-D-isocitrate unmasks an electrophilic 2-vinylglyoxylate species that effects covalent alkylation of the active site Cys_{191} of ICL1 (Figure 38B). Unfortunately, this alkylation was readily reversible by free thiols, thereby eliminating its prospect as a therapeutic agent. As such, we sought to design mechanism-based inactivators that undergo more durable covalent inactivation, by producing a more stable adduct with the active site Cys_{191} of ICL1.

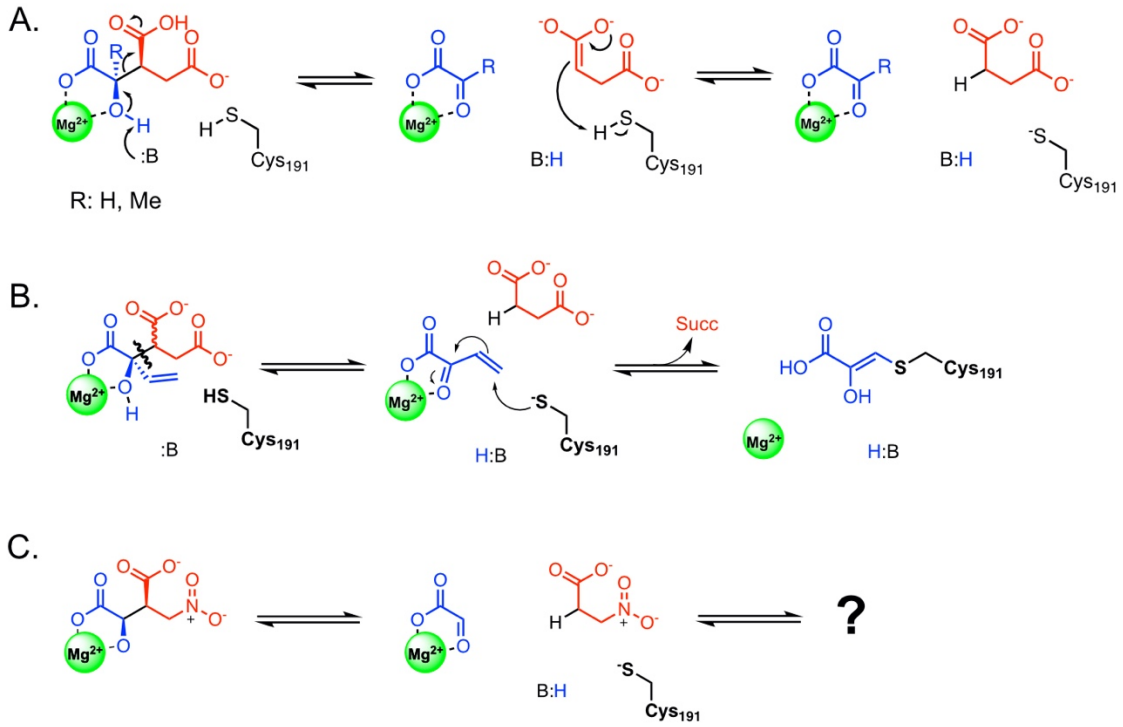


Figure 38. Catalysis of and Effects of Mechanism-Based Inactivators on Isocitrate Lyase 1

A. Proposed catalytic mechanism of ICL1 retro-aldol cleavage of D-isocitrate (R = H) or methylisocitrate (R = Me) to produce the products of glyoxylate and succinate or pyruvate and succinate, respectively.[25] **B**. Mechanism-based inactivation of ICL1 by 2-VIC.[215] **C**. Proposed mechanism of 5-NIC catalyzed-scission by ICL1 to produce glyoxylate and 3-nitropropionic acid.

Inspired by the work of Schloss and Cleland[247] and Ray *et al.*,[234] we envisioned that replacement of a carboxylic acid in the C3 or C5 position of isocitrate with a nitro group would produce a mechanism-based inactivator of ICL1. Previous work by Schloss and Cleland demonstrated that an isocitrate analog containing a nitro group in the C3 position of isocitrate (3-NIC) to replace the 3-carboxy group comprised an unstable and poor substrate of *Pseudomonas indigofera* ICL, which collapses without catalysis into

glyoxylate and 3-nitroproponic acid, particularly at elevated pH and in the presence of divalent magnesium ions.[247] Recently, 3-nitropropionic acid (3-NP) was identified to form a covalent adduct with the catalytic Cys_{191} of ICL1 which may explain why 3-NIC exhibited such low rates of catalysis as measured by glyoxylate production (0.8% when compared to D-IC catalysis).[234] Therefore, in an effort to explore a more viable nitro-analogue of isocitrate to act as a mechanism-based inactivator that upon ICL catalysis elaborates the electrophile 3-NP, we synthesized (*2R*,*3S*)-2-hydroxy-3-(nitromethyl)succinic acid (5-NIC). We hypothesized that ICL1 catalysis of 5-NIC would produce glyoxylate and the ICL1-reactive molecule 3-NP in the active site of the enzyme (Figure 38C), and subsequently lead to robust inactivation of ICL1. Herein, we describe the synthesis, kinetic, crystallographic, and mass spectrometric characterization of 5-NIC, a potent mechanism-based inactivator of ICL1.

B. Experimental Methods and Materials

i. Materials

All chemical reagents used were obtained from commercial vendors and used without purification unless otherwise specified.

ii. Protein Expression and Purification

A plasmid containing *Mtb icl*1 was a gift from Prof. Andrew Murkin of Buffalo University, NY. This construct contains the ICL1 gene (Uniprot: P9WKK6) with an N-terminal, thrombin-cleavable His6 epitope tag. The recombinant protein was expressed in

E.coli BL21(DE3) cells, and expressed protein was purified as previously described.[14, 211, 215] Pure ICL1 was stored in a buffer of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 10% glycerol at –20 °C. For mass spectrometry, ICL1 was further purified via gel filtration (HiLoad® 26/600 Superdex® 200 pg, GE Healthcare), with elution using a buffer of 50 mM HEPES (pH 7.5) and 150 mM NaCl. Fractions containing the pure tetramer were pooled, concentrated, and stored at -80°C. For crystallography, the His_6 epitope tag was removed by overnight incubation with thrombin at 4°C followed by removal of thrombin (1 mL benzamidine-Sepharose 4 fast flow, GE Healthcare), and unprocessed protein (5 mL HisTrap HP, GE Healthcare) to yield pure, tagless ICL1.

iii. Inactivation Kinetics

Unless otherwise specified, all assays were conducted in 96-well half-volume clear plates (Greiner) consisting of 100-µL reaction mixtures containing 50 mM HEPES (pH 7.5) and 10 mM $MgCl_2$ at 37ºC. In the direction of isocitrate cleavage, product glyoxylate was reacted with 10 mM phenylhydrazine to form its phenylhydrazone product, which was measured spectrophotometrically at A324 nm ($\varepsilon 324 = 17{,}000\ M^{-1}cm^{-1}$).[215] For continuous analysis of ICL1 activity, 2-4 nM of the enzyme in 50 mM HEPES (pH 7.5), 5 mg/mL BSA, 10 mM $MgCl_2$, pH 7.5 was added to a reaction mixture containing variable concentrations of 5-NIC (0-100 µM) and 200 µM D-isocitrate. For protection studies a fixed concentration of 5-NIC (10 µM) was used with the ICL1 competitive inhibitors D-malate and hydroxycitrate. Nonlinear regression analysis of the resulting data was fitted using Graphpad Prism 8 to the equations specified below.

iv. Intrinsic Protein Fluorescence

ICL1 (2 μM of ICL1 monomer) in 50 mM HEPES, 10 mM $MgCl_2$, 1 mM DTT (pH 7.5) was combined with 5-NIC (0-20 μM). Reaction mixtures (200 μL) were excited at λ_{ex} = 295 nm, and the intrinsic fluorescence emission spectra (λ_{em} = 315–380 nm) was determined at multiple time points (37°C) using a Biotek Synergy plate reader with black, black-bottom, 96-well plates (Corning). All data were corrected by subtracting the background fluorescence observed in a sample containing buffer. Upon addition of ligand, the spectrum was recorded at multiple time points. The area under the curve of each spectrum was determined, corrected for photobleaching, and normalized to a time zero control. These relative fluorescent changes of ICL1 in response to ligand binding were plotted using $F_t = (F_0 - F_{min})*\exp(-k_{obs}*t) + F_{min}$ in which F_t, F_0, and F_{min} are the respective protein fluorescence measured at time t, t=0, respectively, and k_{obs} is the apparent first-order rate constant of diminution of fluorescence.

v. ICL1 Denaturing Mass Spectrometry

ICL1 (22.5 μM) in 50 mM HEPES (pH 7.5) and 150 mM NaCl was incubated with either 500 μM 5-NIC or with 500 μM glyoxylate in the presence of 500 μM 3-NP for 1.5 h at 37 °C. The buffer in these mixtures were then exchanged into 200 mM ammonium acetate (pH 7.4) buffer using a (Micro Bio-Spin 6 Column, Bio-Rad). The samples were treated with 0.5% formic acid, and denatured mass spectra were collected using Thermo Exactive with extended mass range (EMR). Mass spectra were deconvoluted using Unidec.[248]

vi. Determination of the pK_I of 5-NIC Inactivation.

Assays were conducted in clear 96-well half-volume clear plates (Greiner) consisting of 100-µL reaction mixtures containing 25 mM MES/25 mM Tris-HCl (pL 5.5-8.6) and 10 mM $MgCl_2$, 10 mM phenylhydrazine, and 1 mM DTT and 200 µM D-isocitrate for 5-NIC inactivation at 37ºC. For steady-state analysis the substrate D-isocitrate was dissolved in water, and 5-µL aliquots of variable concentrations (5-5000 µM) were added to the assay buffer. For inactivation, 5-NIC was dissolved in water, and 5-µL aliquots of variable concentrations (2-100 µM) were added to the assay wells. Subsequently, the reaction was initiated by the addition of ICL1 (2 nM for steady-state kinetics and 4 nM for inactivation kinetics) that had been diluted into 25 mM MES/25 mM Tris-HCl, 5 mg/mL BSA, 10 mM $MgCl_2$ (pH: 5.5-8.6**)**. Nonlinear regression analysis of the resulting data was fitted using Prism 8 and the equations as specified.

vii. Determination of pK_a Values of 5-NIC and 3-NP

The fractional conversion of 5-NIC and 3-NP to their nitronate forms was analyzed in various buffers: 50 mM HEPES/citrate, 150 mM NaCl (pH: 6.22-8.19), 50 mM HEPES/CHES, 150 mM NaCl (pH: 9.22-10.03), 50 mM HEPES/CAPS, 150 mM NaCl (pH: 11.03), or 50 mM HEPES/Na-Pi, 150 mM NaCl (pH: 12.02-13). Addition of 50 µM of the respective compound to a 1-mL quartz cuvette containing the appropriate buffer allowed for monitoring of nitronate formation at 232 nm using a SpectraMax M2 Plate

reader. Normalization and replotting of the data versus pH allowed fitting of the data to eq. 16 to determine the acid dissociation constant for both compounds.

viii. Analysis of Kinetics of 5-NIC Nitronate Formation and Protonation

The rate of nitronate formation for both 5-NIC and 3-NP was determined by incubation of 50 μM of the respective compound with 50 mM HEPES/citrate, 150 mM NaCl (pH = 6.2-8.2), 50 mM HEPES/CHES, 150 mM NaCl (pH = 9.2-10.0), 50 mM HEPES/CAPS, 150 mM NaCl (pH: 11.0), or 50 mM HEPES/Na-Pi, 150 mM NaCl (pH = 12-13) in a 1-mL quartz cuvette. Nitronate formation was monitored at 232 nm using a SpectraMax M2 plate reader. The protonation of the 5-NIC and 3-NP nitronate was conducted by incubating the compounds with 0.1 N NaOH at RT for 10 min, followed by adding 50 μM (5 μL) of the molecules to 10 mM potassium phosphate buffer (pH: 7.0) (995 μL), and monitoring the change of absorbance at 232 nm. Fitting of data to equation 14 was used to determine the rate of protonation and the half-lives of 5-NIC and 3-NP nitronate species, respectively.

ix. Mass Spectrometry Analysis of Deuterium Incorporation into 5-NIC

A stock solution of 5-NIC (10 mM) or 3-NP prepared in ddH_2O, and was either added to 99.9% D_2O (Millipore Sigma) or ddH_2O to a final concentration of 50 μM. Subsequently, the molecules were incubated for the specified time, and analyzed using an ISQ EM (ThermoFisher Scientific), equipped with an injector (Harvard Apparatus 11 plus), in which the sample was injected at a flow rate of 0.250 mL/min allowing analysis using

negative mode ionization (15V CID, sheath gas; 50 psig, aux gas; 2 psig, sweep gas; 0.5 psig, vaporizer; 350°C, ion transfer tube; 300°C). Data were acquired over 5 min in the range of 185-200 m/z for 5-NIC and D-IC and 105-150 m/z for 3-NP. The ions were normalized to the highest intensity ion observed in the spectral ranges collected.

x. Crystallography Analysis of 5-NIC Treated ICL1

Tag-free ICL1 (11 mg/mL) in 50 mM Tris (pH 8.0), 10 mM $MgCl_2$, and 1 mM DTT was treated with 1.2 mM 5-NIC overnight at room temperature prior to setting crystal drops. Crystals were produced in 5-7 days by the hanging-drop vapor diffusion method at 17 °C in 0.1 M Tris-HCl (pH 8.0), 0.2 M sodium acetate, and 20-30% PEG4000 mixed at 1:1 or 1:3 volume ratio with the protein solution. For the X-ray data collection, crystals were flash-frozen in liquid nitrogen after cryoprotection with 20% glycerol added to the mother liquor. High-resolution diffraction data for ICL1-5-NIC co-crystals were collected at the 19ID beamline of the APS synchrotron. Data were indexed, integrated and scaled using HKL2000 suite.[249] The phase was solved by molecular replacement using previously solved ICL1 closed form structure as a search model. Final model was produced by iterative building in Coot and refinement using Phenix refine software.[250, 251] Data and refinement statistics are listed in Table 7. Graphical rendering was conducted using UCSF Chimera.[252]

xi. Analysis of Kinetic Data

Time courses of inactivation by 5-NIC and 3-NP the data were fitted to:

$$P = (\frac{V_o}{kobs})(1 - e^{-k_{obs}*t}) \quad \textbf{(eq. 3)}$$

Where P is the concentration of the product glyoxylate, V_o is the initial rate, k_{obs} is the observed rate of inactivation, and t is time.

The observed rate constants of inactivation (k_{obs}) were then replotted vs. inactivator concentration and fitted to:

$$k_{obs} = \frac{k_{inact}*I}{(K_I\left(1+\frac{S}{K_m}\right)+I)} \quad \textbf{(eq. 10)}$$

where k_{inact} is the intrinsic maximal rate of inactivation, I is the concentration of inactivator, K_I is the concentration of inactivator that gives 0.5 k_{inact}, S is the substrate concentration, and K_m is the Michaelis constant for the substrate. For inactivation at pH 7.5 in HEPES buffer the K_m used was (37 μM), as we previously reported.[215] For pH rate studies we employed mixed buffers, and the values of K_m were experimentally determined for each pH value used by fitting the observed kinetics rates vs. substrate concentration, and fitting the data to the Michaelis-Menten equation (data not published).

When plots of k_{obs} vs. [I] were not hyperbolic (lack of saturation with I), the second-order rate constant was approximated from fitting the data to:

$$k_{obs} = \left(\frac{k_{inact}}{K_I}\right)\left(\frac{K_m}{K_m + S}\right) I \qquad \textbf{(eq. 11)}$$

In which K_m is the experimentally determined Michaelis constant under the given conditions, S is substrate concentration in the assay, and I is inhibitor or inactivator concentration.

Determination of the fractional protection of ICL1 from hydroxycitrate and D-malate is given by the following;

$$1 - \left(\frac{k_{obs}^{no\ ligand}}{k_{obs}^{ligand}}\right) = \alpha\left(\frac{[L]}{K_d^n + [L]^n}\right) \qquad \textbf{(eq. 12)}$$

in which $k_{\text{obs no ligand}}$ is the rate of the enzyme inactivation in the absence of inhibitor, $k_{\text{obs ligand}}$ is the rate of the enzyme inactivation in the presence of added inhibitor or inactivator, α is a proportionality factor that represents the maximal degree of protection, [L] is the concentration of added inhibitor or inactivator, K_d is the apparent dissociation constant of L, and n is the Hill coefficient.

Data of partition analysis of 5-NIC-treated ICL1 were fitted to:

$$\frac{v_i}{v_0} = 1 - \left(\frac{[5NIC]}{[E_t](1+p)}\right) \quad \textbf{(eq. 13)}$$

In which [5NIC] is the concentration of 5-NIC in μM, [E_t] is the enzyme concentration in μM, and p is the partition coefficient.

For nonlinear fitting of the partition ratio we fitted the data to:

$$\frac{v_i}{v_0} = 1 - \left(\frac{([Et] + [5NIC] + K'_d) - \sqrt{([E_t + [5NIC] + K_d)^2 - 4[E_t][5NIC]}}{2(1+p)[E_t]}\right) \textbf{(eq. 14)}$$

To obtain the protonation rate of 5-NIC and 3-NP and the observed decrease in protein fluorescence, the data were fitted to the following:

$$P = (Y_{max} - Y_{min})e^{(-k_{obs}*t)} + Y_{min} \quad \textbf{(eq. 15)}$$

Where Y_{min} is the lower-bound rate value at a given time, Y_{max} is the upper-bound rate value at a given time, k_{obs} is the apparent observed rate, and t is time.

To obtain the rate of deuterium incorporation into 5-NIC by mass spectrometry, data were fitted to eq. 16:

$$P = Y_{min} + \left(\frac{Y_{max}}{Y_{min}}\right)(1 - e^{-k_{obs}*t}) \qquad \textbf{(eq. 16)}$$

In which Y_{min} is the lower-bound rate value at a given time, Y_{max} is the upper-bound rate value at a given time, k_{obs} is the apparent observed rate, and t is time.

To determine values of pK_a of 5-NIC and 3-NP, we incubated 50 μM of each molecule in mixed buffers, and monitored the change in absorbance at 232 nm until signals were stable and had plateaued. The data were then normalized to the largest value of absorbance, followed by fitting to eq. 17:

$$P = Total\left(\frac{1}{1 + (\frac{1}{10^{(pH-pKa)}})}\right) \qquad \textbf{(eq. 17)}$$

In which Total is the highest value observed and pK_a is the apparent acid dissociation constant.

For pH-rate profiles the steady-state Michaelis-Menten kinetics were determined at each pH value and the -log(K_m) was plotted versus pH followed by fitting to eq. 18:

$$P = log\left(\frac{c}{1 + (10^{(pKa1-pH)} + 10^{(pH-pKb)}}\right) \qquad \textbf{(eq. 18)}$$

in which *c* is the pH-independent limiting value, and pK_{a1} and pK_b are the apparent acid and base dissociation constants.

For inactivation with 5-NIC the rates of inactivation were fitted using eq. 3, and resulting values of k_{obs} were then replotted vs [5-NIC], followed by fitting to eq. 10 in cases in which a hyperbolic curve was observed, and to eq. 11 in which a hyperbolic curve was absent. Upon obtaining the values of k_{inact}, K_I, and k_{inact}/K_I using the experimentally determined K_m values at each pH value, the data were plotted as pK_I vs. pH and fitted to eq. 18.

C. Results and Discussion

i. 5-NIC Effects Time-Dependent Inactivation of ICL1

Reactions in assay mixtures containing variable concentrations of 5-NIC (0-100 μM) and 200 μM (~5.5K_m) D-IC were initiated by the addition of ICL1. The resulting kinetic data were consistent with time-dependent inactivation kinetics, as exemplified by the concave-downward, curvilinear time courses shown in Figure 39A. Increasing concentrations of 5-NIC resulted in more rapid conversion of initial inhibition to complete loss of ICL1 activity, and at 10 μM 5-NIC, ICL1 activity was eliminated after ~1000 s. Global fitting of the data in Figure 39A to eq. 3, followed by replotting of the individual observed rates of inactivation (k_{obs}) vs. 5-NIC concentration resulted in a hyperbolic dependence of 5-NIC concentration (Figure 39B). The observation of apparent saturation by 5-NIC is a hallmark of two-step, time-dependent inactivation (Figure 39B). Fitting of these data to eq. 10 resulted in $K_I = 6.5 \pm 0.7$ μM (the concentration of 5-NIC which gives

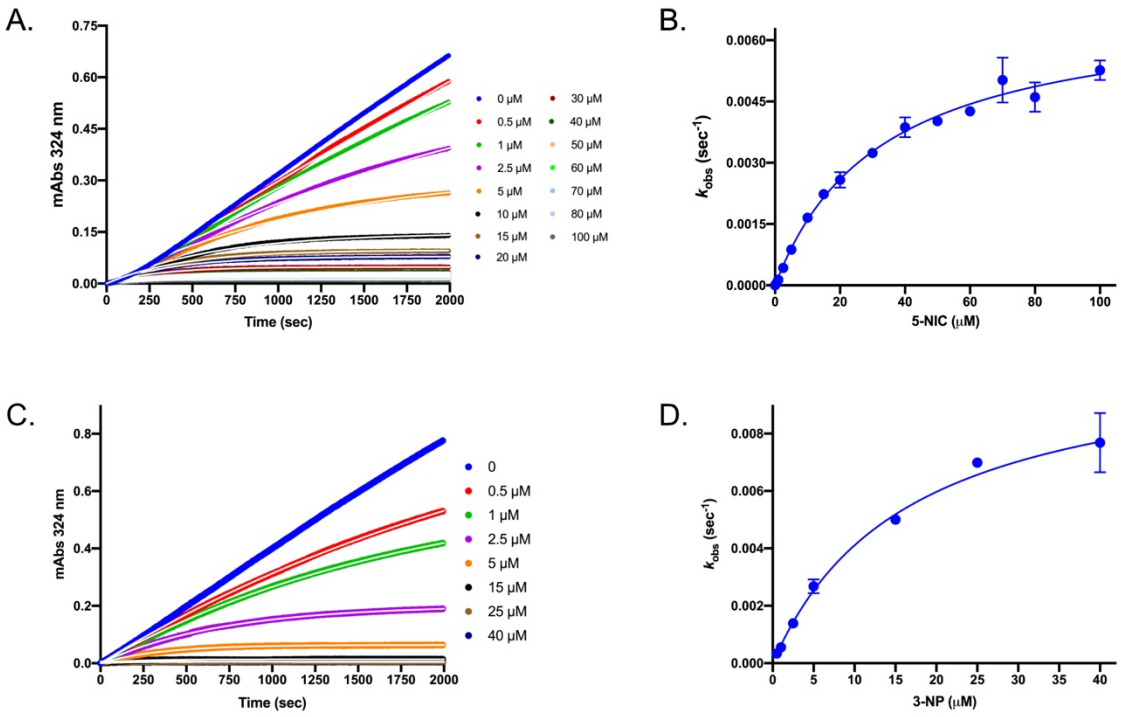


Figure 39. ICL1 Inactivation by 5-NIC and 3-NP.

A. Representative inactivation kinetics of ICL1 in the presence of variable concentrations of 5-NIC with fixed D-isocitrate. Curves drawn through the experimental data points (white) correspond to fitting of data to eq. 3**. B**. Replot of the rate of inactivation (k_{obs}) vs. [5-NIC] from data in 2A, which were fitted to eq. 10 as shown by the curve drawn through the experimental data points. **C**. **A**. Representative kinetic traces of glyoxylate formation (mAbs at 234 nm) at variable concentrations of 3-NP at fixed 200 µM D-isocitrate, under identical assay conditions used for 5-NIC inactivation studies as in Figure 2A. Data were fitted to eq. 3 to determine the k_{obs} constants of ICL1 inactivation by 3-NP. **B**. The replot of k_{obs} values vs. inactivator concentration was fitted to eq. 10, yielding limiting values of k_{inact} = 0.65 ± 0.04 min^{-1}, K_I = 3.1 ± 0.4 µM, k_{inact}/K_I = (2.5 ± 0.3) x 10^{-1} µM^{-1} min^{-1}.

half-maximal inactivation), and a maximal rate of inactivation k_{inact} = 0.49 ± 0.02 min^{-1}, yielding a value of k_{inact}/K_I of (7.5 ± 0.9) × 10^{-2} µM^{-1} min^{-1}. The value of k_{inact}/K_I (the larger the value the more efficient the inactivator), known as the specificity constant of inactivation, is analogous to the specificity constant of substrate catalysis, k_{cat}/K_m. Thus,

comparison of these two values provides the relative efficiency of the inactivator vs. steady state substrate turnover. In the case of ICL1, the k_{cat}/K_m is 27 ± 7 μM^{-1} min^{-1} is ~390-fold more efficient than that of 5-NIC inactivation ($k_{inact}/K_I = (6.9 \pm 2) \times 10^{-2}$ μM^{-1} min^{-1}). The value of K_I (which is the "Michaelis constant" of mechanism-based inactivation) is ~6-fold lower than $K_{D\text{-}IC}$ (37 μM). Accordingly, while 5-NIC binds more tightly to ICL1 than D-IC, the 390-fold difference of k_{cat}/K_m and k_{inact}/K_I arises from the slower rate of inactivation vs. substrate turnover (k_{cat}/k_{inact} = 2000), suggesting that any chemical steps which contribute to the value of k_{inact} are much slower than product formation. This is typical for mechanism-based inactivators.[7] However, when comparing 5-NIC to our previously reported ICL1 mechanism-based inactivator 2-VIC,[215] we see that 5-NIC effects 21-fold more efficient inactivation ($k_{inact}/K_I = (7.5 \pm 0.9) \times 10^{-2}$ μM^{-1} min^{-1} versus k_{inact}/K_I of $(3.5 \pm 0.7) \times 10^{-3}$ μM^{-1} min^{-1}). In addition, under identical assay conditions employed for 5-NIC inactivation, we determined that the rate of ICL1 inactivation by 3-NP was slightly (3.3-fold) more efficient than 5-NIC inactivation ($k_{inact}/K_I = (2.5 \pm 0.3) \times 10^{-1}$ μM^{-1} min^{-1}) which predominantly resulted from 3-NP binding 2.5-fold more tightly to the enzyme than 5-NIC, while the rate of 3-NP inactivation (k_{inact}) was only 1.3-fold faster (Figure 39C,D). These results lead us to speculate that the slightly diminished rate of k_{inact} for 5-NIC compared to 3-NP may result from the prerequisite catalytic step of C2-C3 bond scission of 5-NIC to produce glyoxylate and 3-NP which ultimately leads to the inactivation of ICL1.

ii. Intrinsic Tryptophan Fluorescence of ICL1 Inactivation by 5-NIC.

In previous studies of ICL1 we have shown that intrinsic tryptophan fluorescence reports on the closure of the active site loop of ICL1 which occurs upon ligand binding.[14, 215] Therefore, we analyzed the rate of decay of fluorescence intensity (k_{obs}) upon incubation of fixed concentrations of ICL1 (2 μM) with variable 5-NIC (0-20 μM) to determine the rate of 5-NIC inactivation in the absence of exogenous D-IC. We observed that the apparent rate (k_{obs}) of 5-NIC inactivation in the absence of substrate was up to 4–fold faster compared to the rate of k_{obs} in the presence of 200 μM D-IC (ICL1, 4 nM) with the same 5-NIC concentrations (Figure 40C). To ensure that the decay in fluorescence reports on the inactivation rate, we diluted ICL1-treated with 5-NIC 200-fold, and analyzed the residual activity in the presence of saturating D-IC (1 mM). Samples treated with 10 and 20 μM 5-NIC recovered less than 2% initial activity, whereas samples treated with 5 μM 5-NIC maintained ~20% initial activity, consistent with the incomplete saturation of tryptophan fluorescence seen at 5 μM 5-NIC (data not shown). Overall, these data indicate that 5-NIC binds to the active site of ICL1, and that the rate of inactivation in the absence of substrate is enhanced, consistent with it being competitive versus D-IC.

iii. 5-NIC Inactivation is Competitive Versus D-isocitrate and is Abrogated by Competitive Inhibitors of D-isocitrate.

Assessment of 5-NIC inactivation of ICL1 in the presence of fixed 5-NIC concentrations (5-20 μM) and varied concentrations of D-IC (15-300 μM) resulted in a hyperbolic decrease of the apparent k_{obs} constant with respect to increasing D-IC

concentration, consistent with 5-NIC being competitive versus D-IC (data not shown). Replotting of the $1/k_{obs}$ values versus D-IC concentration showed that the apparent rate of k_{obs} decreased as a function of increasing D-IC concentration, wherein $1/k_{obs}$ reaches its minimal value as D-IC concentration approaches zero (Figure 40D). Figure 40D is in effect a Dixon plot of inactivation wherein D-IC acts like a competitive inhibitor. To further confirm the competitive nature of binding of 5-NIC versus D-IC, ICL1 was treated with 10 μM 5-NIC and variable concentrations of the D-IC-competitive inhibitors D-malate and hydroxycitrate. With both competitive inhibitors a decrease in the curvilinear behavior of 5-NIC inactivation was observed as the concentration of the inhibitors was increased (Figure 40E). In the case of D-IC -malate (0-6 mM), nearly complete protection of ICL1 from 5-NIC inactivation was provided ($appK_d$ = 843 ± 112 μM). Although this binding constant is 2.7-fold higher than the reported inhibition constant of D-malate versus ICL1 (K_i = 310 ± 30 μM),[215] this is in agreement with the tighter binding of 5-NIC compared to the substrate D-IC. The more potent D-IC competitive inhibitor hydroxycitrate (K_i = 8 μM)[14] afforded complete protection versus inactivation by 5-NIC with an apparent dissociation constant of 52 μM (data not shown). Collectively these data suggested that 5-NIC inactivation is competitive vs. D-IC, and that 5-NIC likely binds to the active site of ICL1 in a similar fashion to D-IC.

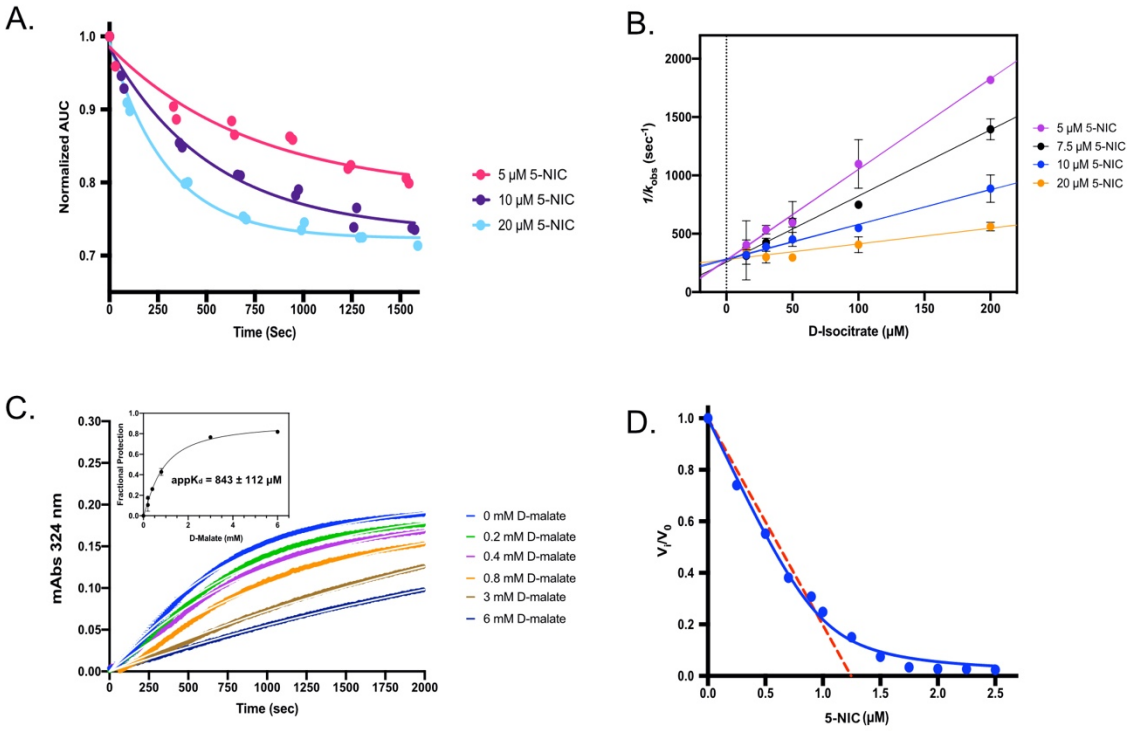


Figure 40. Analysis of ICL1 Inactivation by 5-NIC

A. Apparent time-dependent inactivation of 5-NIC versus ICL1 as determined by the diminution of the intrinsic protein fluorescence signal. Curves drawn through the data points were obtained by fitting the data to eq. 15. **B**. Replotting of $1/k_{obs}$ obtained from treating ICL1 with variable concentrations of D-IC in the presence of fixed concentrations of 5-NIC. **C**. Representative protection of ICL1 inactivation by 5-NIC with increasing concentrations of D-malate at a fixed concentration of D-IC. Curves drawn through points (white) correspond to fitting to eq. 3. (Inset) Replot showing the apparent K_d of D-malate in the presence of D-IC and 5-NIC as fitted to eq. 12. **D**. Residual activity of ICL1 (1 µM) following an overnight incubation of ICL1 in the presence of varied 5-NIC (0-2.5 µM). The dotted line represents a partition coefficient of 0.25 ± 0.02 which was obtained when fitting the data to eq. 13. Data were fitted to eq. 14.

iv. 5-NIC Inactivation of ICL1 is Covalent and Irreversible and has a Partition Coefficient of Less Than 1.

Having identified that 5-NIC likely functions as a mechanism-based inactivator of ICL1, we sought to determine if this inactivation was covalent and irreversible. Analysis of ICL1 treated with 5-NIC ($5K_I$) followed by successive rounds of dialysis (125,000-fold dilution) for over 72 h, resulted in negligible recovery of ICL1 activity (< 6%) compared to the untreated control, suggesting that 5-NIC forms a stable, effectively irreversible, covalent adduct with ICL1, although not all turnovers of 5-NIC result in covalent bond formation. In an effort to determine the efficiency of 5-NIC inactivation we measured the partition ratio. This ratio reports on the fraction of the putative electrophilic product of 5-NIC that forms a covalent adduct with ICL1 versus the fraction that dissociates. Following a 20-h pre-incubation of ICL1 (1 μM) with varied concentrations of 5-NIC (0-2.5 μM), reaction mixtures were diluted 100-fold into assay buffer containing saturating D-IC (1 mM) to evaluate the residual enzyme activity (v_i/v_o) (Figure 40D). At the highest concentration of 5-NIC tested (2.5 μM) only ~2% enzyme activity remained, but consistent with our dialysis experiment, complete loss in activity was never observed. When fitting the data to eq. 12 we determine the partition coefficient to be 0.25 ± 0.03, which demonstrates that 1.25 molecules of 5-NIC are required to inactivate 1 molecule of ICL1 monomer, underscoring the high efficiency of 5-NIC inactivation of ICL1.

v. Mass Spectrometry of ICL1 treated with 5-NIC.

To further verify the covalent addition of a catalytic product of 5-NIC to ICL1, we subjected the protein to denaturing mass spectrometry. Analysis of untreated ICL1 produced a mass of 48790.78 ± 1.4 da, consistent with the theoretical mass of the ICL1

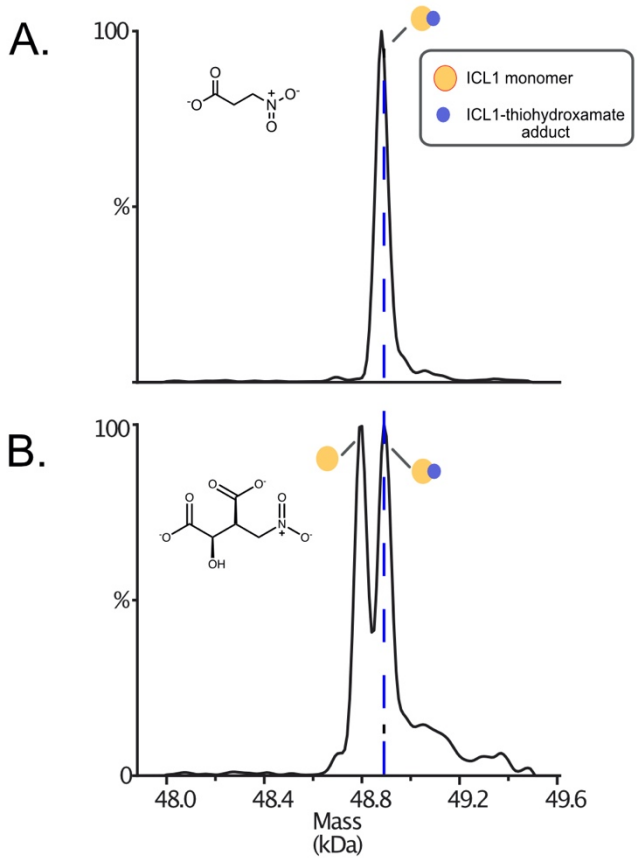


Figure 41. 5-NIC Effects Covalent Modification of ICL1 Like That of 3-NP.

A. ICL1 treated with 3-NP and subjected to denatured mass spectrometry produced an ICL1 species with an additional mass of 100.2 ± 0.4 Da per monomer, consistent with the formation of a thiohydroxamate adduct. **B**. ICL1 treated with 5-NIC produced two major species comprising apo ICL1 and an ICL1 adduct with an additional mass of 100.9 ± 1.9 Da per monomer, consistent with the formation of the thiohydroxamate adduct seen in 3-NP treated ICL1. Mass spectrometry analysis was conducted by Zahra Moghadamchargari from the laboratory of Dr. Arthur Laganowsky.

monomer (48,787.52 Da) (Figure 41). Following incubation of 5-NIC with ICL1 followed

by denaturing electrospray mass spectrometry, we observed a mass of 48891.67 ± 1.9 Da, an addition of 100.9 ± 1.9 Da per monomer (Figure 41B). This mass shift is consistent with the reaction of Cys_{191} with the nitronate of 5-NIC derived 3-NP, and the subsequent elimination of a water molecule to form a Cys_{191}-thiohydroxamate adduct, as previously seen by the Murkin group.[234] To verify this, we treated ICL1 with 3-NP, and the resulting mass of 3-NP-treated ICL1 was 48890.95 ± 0.4 Da, an addition of 100.2 ± 0.4 Da per monomer (Figure 41B). This is consistent with the 5-NIC adduct produced being analogous to that seen for 3-NP-inactivated-ICL1. Together, this suggests that 5-NIC inactivation of ICL1 likely occurs in a stepwise fashion wherein ICL1 initially cleaves the C2-C3 bond of 5-NIC to form both glyoxylate and 3-NP in the active site of the enzyme, followed by the reaction of the ICL1 Cys_{191} thiolate with the nitroic acid of 3-NP to form a thiohydroxamate adduct as the final product.

vi. The nitronate form of 5-NIC is not required for ICL1 binding.

Deprotonation of the C4 carbon of 5-NIC would produce the nitronate form of 5-NIC (N-5-NIC), which may be the preferred form for efficient binding to ICL1. If this were the case, the maximal affinity of 5-NIC would occur near pH 9.4 (the experimentally determined pK_a of 5-NIC is 9.4 ± 0.09, Figure 42A). Plotting the pK_I of 5-NIC inactivation vs. pH (Figure 42D) yielded values of pK_a = 6.1 ± 0.2 and pK_a = 7.6 ± 0.1, which do not

indicate that N-5-NIC is preferred for binding, rather, deprotonation of a group of pK_b = 7.6 ± 0.1, well below the value of 9.4 ± 0.09 for "chemical" deprotonation of 5-NIC eliminates binding and inactivation. This finding is consistent with the results of

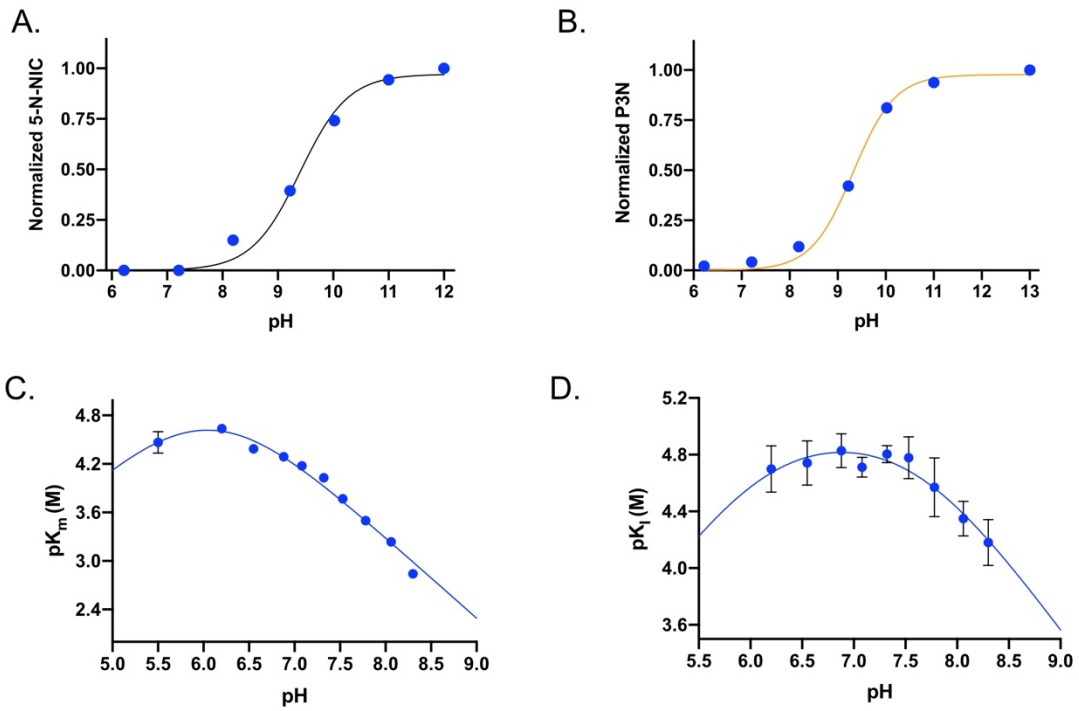


Figure 42. 5-NIC pH Dependence of UV Spectra of 5-NIC and 3-NP and pH Rate Profiles of D-IC and 5-NIC

A. The pK_a of 5-NIC was determined by monitoring the formation of its nitronate at 232 nm at pH 6-12. The change in absorbance at 232 nm, which reports on de-protonation, were normalized and plotted vs. pH, from which fitting of the data to eq. 17 yielded pK_a = 9.4 ± 0.09. **B**. The pK_a of 3-NP was determined in an analogous fashion to **A** and yielded a pK_a of 9.3 ± 0.06. **A**. Plotting of the pK_m of D-IC versus pH which was fitted to eq. 18 yielding limiting values of 5.7 ± 0.2 and 6.4 ± 0.1. **B**. Plotting of the pK_I of 5-NIC versus pH which was fitted to eq. 18 yielding the limiting values of 6.1 ± 0.2 and 7.6 ± 0.1.

Moynihan *et al.* who suggested that the nitronate form of 3-NP, P3N, was not required for ICL1 binding, but rather was generated in the active site of ICL1 upon binding.[211]

vii. Formation of the 5-NIC Nitronate Enhances ICL1 Inactivation Efficiency but Undergoes Rapid Protonation.

Previous work has shown that treatment of ICL with the nitronate form of 3-NP (P3N) (Figure 44A) results in a marked enhancement in the rate (k_{obs}) of ICL inactivation.[234, 247] Formation of P3N occurs via incubation of 3-NP with 0.1 N K/NaOH, and the subsequent dilution of this preformed P3N into 10 mM potassium phosphate (K-Pi) buffer (pH = 7.0) results in a prolonged half-life (~15 min) of the nitronate species, thus allowing for facile analysis into its effect on ICL activity. Ray *et al*. identified this P3N species exerted ~100-fold more efficient inactivation of ICL1 compared to untreated 3-NP, suggesting that the anionic P3N species is a key intermediate in ICL1 inactivation.[234] Owing to the presumed ability of 5-NIC to be converted to its nitronate form (N-5-NIC) (Figure 44A) under basic conditions, we sought to determine if the formation of N-5-NIC, followed by addition to ICL1, effects similar enhancement in the rate of inactivation. Surprisingly, incubation of ICL1 with preformed N-5-NIC in K-Pi buffer (pH 7.0) only yielded a mere 1.7-fold enhancement in inactivation efficiency when compared to 5-NIC under the same conditions ($k_{inact}/K_I = 1.7 \pm 0.3 \times 10^{-3}$ μM^{-1} min^{-1} versus 9.9×10^{-4} μM^{-1} min^{-1}, respectively) (Figure 43). While this result suggests that the nitronate form of P3N, derived from N-5-NIC, likely serves as a reaction intermediate in ICL1 inactivation, we sought to determine why such a small enhancement in inactivation efficiency was observed. We initially hypothesized a key difference may be due to the rate of N-5-NIC protonation in solution when compared to P3N. Consistent with this hypothesis, the protonation of preformed N-5-NIC proceeded ~86-fold faster than that of

preformed P3N (k = 3.8 x 10^{-1} sec^{-1} versus k = 4.4 x 10^{-3} sec^{-1}) under the same conditions (Figure 44C). Owing to this significant difference in the rate of protonation for the two molecules, we investigated the ability of deuterium to incorporate into samples of 5-NIC

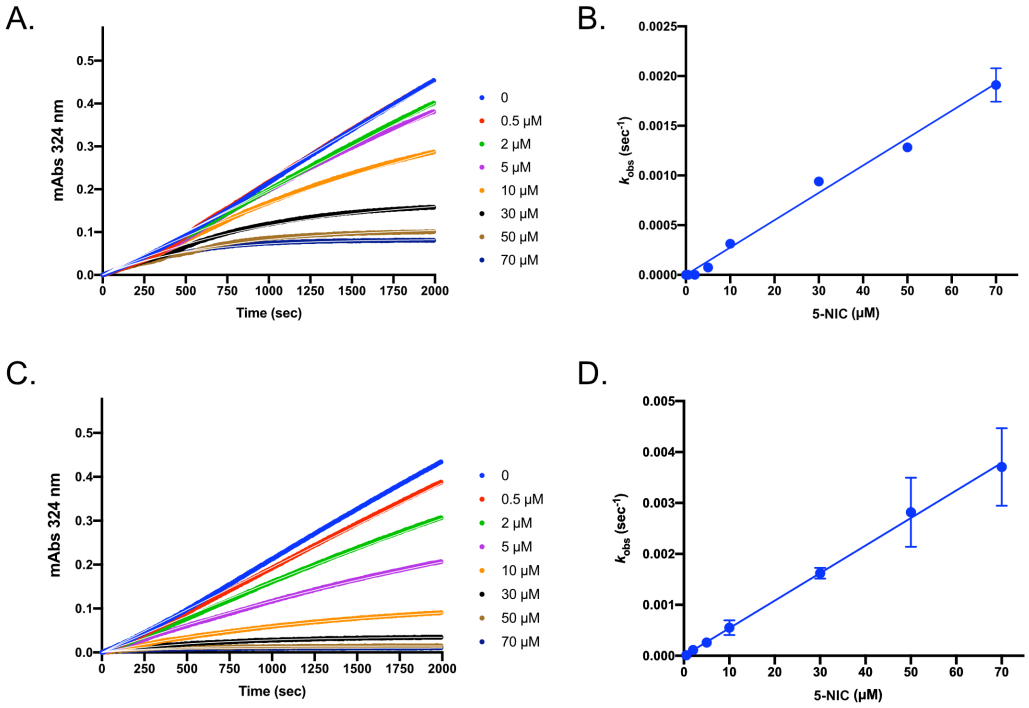


Figure 43. Time Courses of 5-NIC Inactivation in Presence and Absence of N-5-NIC in 10 mM K-Pi Buffer.

A. Representative plot of ICL1 (2 nM) inactivation in the presence of variable [5-NIC] with fixed D-IC (200 µM) in 10 mM K-Pi buffer (pH 7.0). The curves (white) drawn through the experimental data were fitted using eq. 3 to yield apparent k_{obs} values. **B**. Replotting of k_{obs} vs. [5-NIC]. The line drawn through the experimental data was from fitting to eq. 11 to give k_{inact}/K_I = (6.0 ± 0.1) x10^{-3} μM^{-1} min^{-1}. **C**. Representative plot of ICL1 (2 nM) inactivation in the presence of variable [5-NIC] pretreated with 0.1N NaOH with fixed D-IC (200 µM) in 10 mM K-Pi buffer (pH 7.0). The curves (white) drawn through the experimental data were fitted using eq. 3 to yield apparent k_{obs} values. **D**. Replotting of k_{obs} vs [5-NIC] that had been treated with 0.1N NaOH. The line drawn through the experimental data was from fitting to eq. 11 giving a value of k_{inact}/K_I of (1.1 ± 0.04) x10^{-2} μM^{-1} min^{-1}

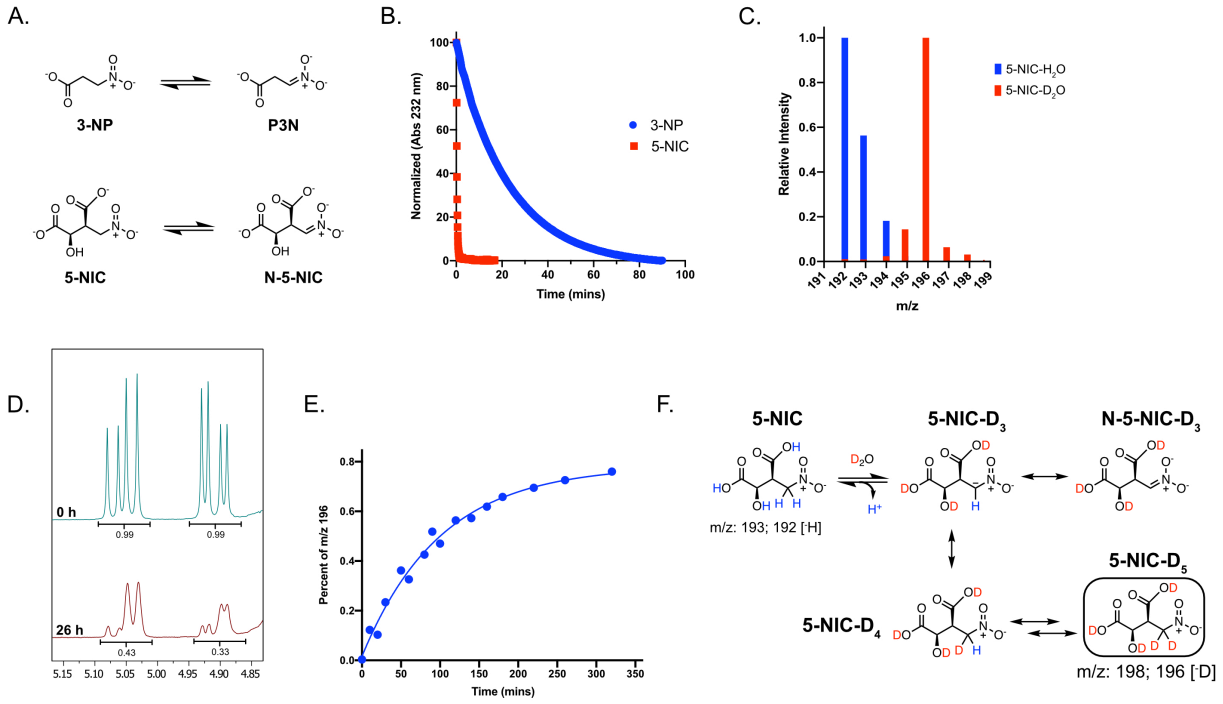


Figure 44. Carbon-4 of 5-NIC Undergoes Rapid Incorporation of Deuterium when Incubated in D_2O

A. The structures of 3-NP and 5-NIC and their respective nitronate forms (P3N and N-5-NIC). **B**. The rate of nitronate to nitro conversion at 50 µM 5-NIC and 3-NP upon preincubation in 0.1N NaOH, followed by dilution (100-fold) into 10 mM K-Pi buffer (pH 7.0). Curves were fitted to eq. 15 to determine the rate of conversion. **C**. Direct infusion mass spectrometry analysis of 50 µM 5-NIC incubated with either H_2O or D_2O for 26 h. **D**. NMR analysis of the C4 protons of 5-NIC following 26 h incubation in D_2O. **E**. Direct infusion mass spectrometry analysis of the formation of the 5-NIC m/z 196 species using negative mode ionization. The curve drawn through the experimental data points is from fitting the data to eq. 16 **F**. Proposed conversion of 5-NIC in D_2O to produce the diminished signal of the C4 protons observed in NMR and the m/z 196 species observed in mass spectrometry studies.

and 3-NP dissolved in D_2O by direct infusion mass spectrometry using negative mode ionization. Following 26 h incubation, we identified that 5-NIC underwent near total conversion to a +3 m/z [$^-$D] species compared to the mass of 5-NIC incubated in H_2O (Figure 44C).

We tentatively assign this species to **5-NIC-D_5**, as this +5 m/z species would undergo the loss of a deuterium from either the C1 or C3 carboxylic acid of 5-NIC-D_5 during negative mode ionization to form the +3 m/z [$^-$D] species (Figure 44F). Conversely, even following 72 h incubation, no substantial conversion of 3-NP to a +1 or +2 m/z [$^-$D] species was observed (not shown). Subsequent NMR analysis confirmed the incorporation of deuterium at the C4 position of 5-NIC, demonstrating that robust interconversion between the nitro and nitroic acid/nitronate forms of 5-NIC occurs in bulk D_2O solution (Figure 44D). In agreement with the MS studies, 3-NP yielded no detectible deuterium incorporation, suggesting that the rate of 5-NIC tautomerzation occurs more rapidly compared to that of 3-NP (not shown).

We next monitored the rate of formation of the +3 m/z [$^-$D] 5-NIC species over time using direct infusion mass spectrometry. Incubation of 5-NIC in D_2O, followed by injection at discreet time points, identified the initial rapid formation of a +1 m/z [$^-$D] species, consistent with the facile proton/deuteron exchange from solvent for groups on the C1, C2, and C3 of 5-NIC (C1 and C3 carboxylic acids and C2 hydroxyl. However, after 5 h, near quantitative conversion of 5-NIC to its +3 m/z [$^-$D] species occurred at a rate of 0.01 ± 0.001 min^{-1}, which reported on the incorporation of deuterium into the C4 of 5-NIC (Figure 44C,E). Close analysis of the three dimensional structures of 5-NIC and 3-NP shows that the C2 hydroxyl of 5-NIC is poised to serve as a potential intramolecular proton donor which could assist in the conversion of N-5-NIC to 5-NIC. Consistent with this, Porter and Bright reported that a hydroxyl group in the 2-position of 3-NP (3-nitro-2-hydroxy-propionate, 3-NHP) exhibited a 16-fold faster rate of protonation (in pH 7.5

buffer) when compared to 3-NP (in pH 7.0 buffer).[253] Comparison of the three dimensional structure of 3-NHP and 5-NIC revealed that the C2-hydroxyl oxygen atom resides respectively 2.4 and 3.2 Å away from the carbanion that would be formed in the tautomerization of both molecules, suggesting that the proton or deuterium of this hydroxyl, instead of solvent, may assist in the intramolecular protonation of both

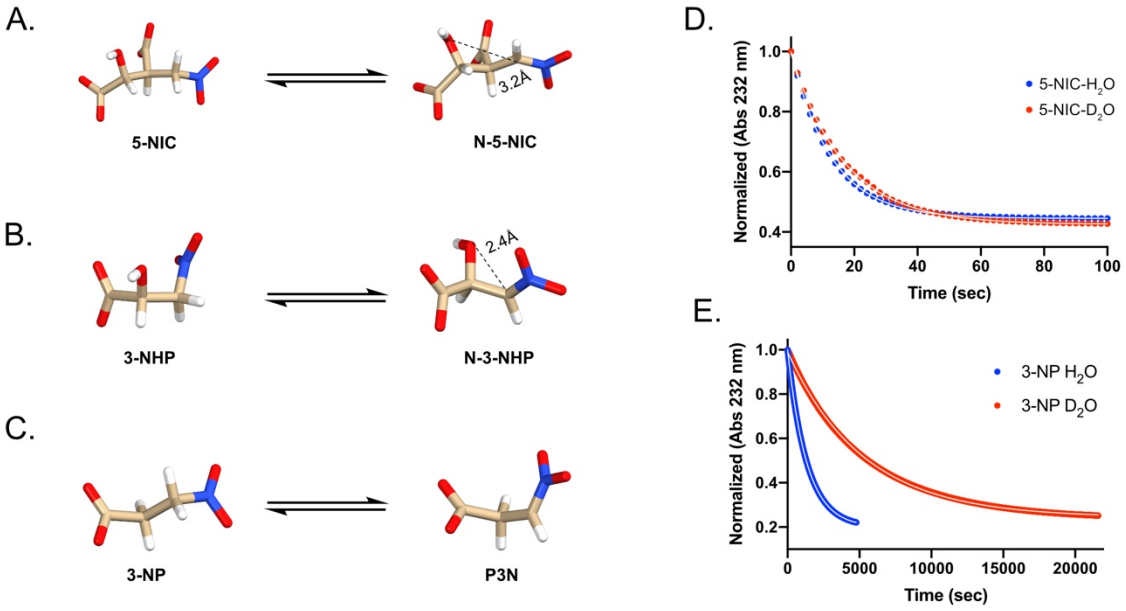


Figure 45. Three-dimensional Structures of 5-NIC, 3-NHP, and 3-NP with their Respective Nitronate Forms and sKIEs of N-5-NIC and P3N Protonation.

A. The 5-NIC and nitronate N-5-NIC form of the molecule. The distance of the C2 hydroxyl oxygen to the presumed carbanion at the C4 is shown via a dotted line. **B**. The 3-NHP and nitronate N-3-NHP form of the molecule. The distance of the C2 hydroxy oxygen to the presumed carbanion at the C3 is shown via a dotted line. **C**. The 3-NP and nitronate P3N form of the molecule. **D**. Representative plot of protonation of 5-NIC following preincubation in 0.1N NaOH (10 min) and 100-fold dilution into 10 mM K-Pi buffer (pH 7.0) as determined by its absorbance of 232 nm. The curves drawn through the experimental data were from fitting to eq. 15. The k_H/k_D determined for the reaction was 1.3 ± 0.01. **E**. Representative plot of 3-NP protonation under identical conditions as those used for 5-NIC. The curves drawn through the experimental data were from fitting to eq. 15. The k_H/k_D determined for the reaction was 4.3 ± 0.2.

molecules (Figure 45A,B)). To test this hypothesis, we pretreated 3-NP and 5-NIC with 0.1 N NaOH, diluted the solution into 10 mM K-Pi buffer prepared in both H_2O and D_2O (pL 7.0), and analyzed the rate of protonation by spectrophotometric measurement at 232 nm. We observed a solvent kinetic isotope effect (sKIE) for N-5-NIC protonation of k_H/k_D = 1.3 ± 0.01 whereas the sKIE of P3N protonation was much larger: k_H/k_D = 4.3 ± 0.2 (Figure 45). We interpret this small normal sKIE effect for N-5-NIC protonation to be consistent with a either an early or late transition state for proton transfer for 5-NIC, while a more symmetric transition state exists for proton transfer in 3-NP. Together, these data suggested that the 5-NIC C2 hydroxyl group likely accounts for the rapid protonation of N-5-NIC, allowing the facile incorporation of deuterium into the C4 position of 5-NIC, and as a consequence, limited enhancement in ICL1 inactivation was observed for the 5-NIC nitronate when compared to the 3-NP nitronate.

viii. Crystallographic Analysis of 5-NIC Inactivated ICL1.

To further verify our kinetic and mass spectrometry data that suggested 5-NIC serves as a covalent, mechanism-based inactivator of ICL1, we sought to determine the crystal structure of the ICL1 treated with 5-NIC. In approximately one-week ICL1-5-NIC co-crystals formed via hanging-drop vapor diffusion method which diffracted to high resolution (~1.8Å), and were of the $P2_12_12_1$ space group. Fo-Fc difference electron density showed positive peaks in the closed active site of ICL1. We identified electron density consistent with a molecule covalently bound to Cys_{191} of ICL1 apparently resulting from the attack of the Cys_{191} sulfur onto the C3 of 3-NP, a product of the retro-aldol cleavage

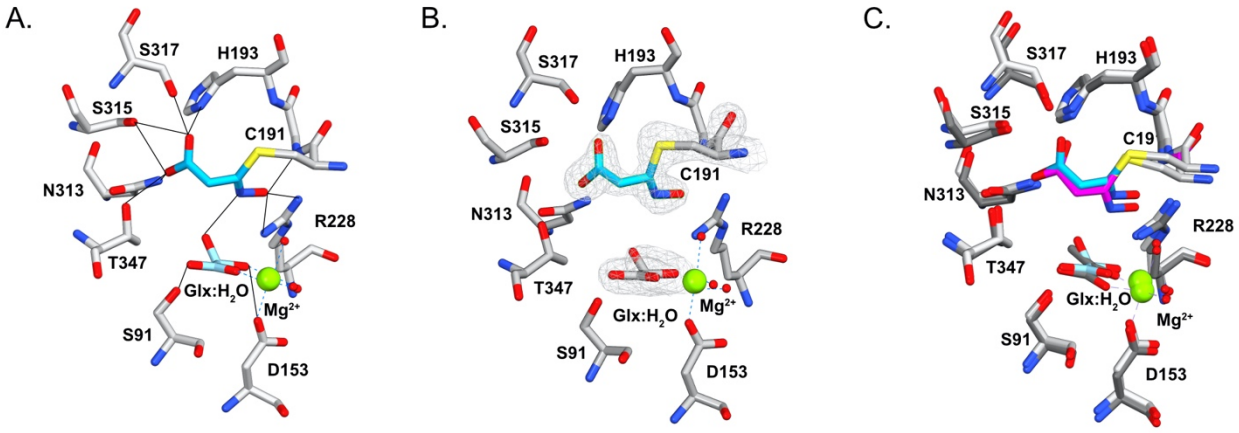


Figure 46. 5-NIC Inactivation of ICL1 Produces a Cys_{191}-thiohydroxamate Adduct and Glyoxylate Hydrate.

A. Structure of ICL1 covalently bound to the 5-NIC-derived product 3-NP, which underwent subsequent dehydration, and glyoxylate hydrate which is bound to the divalent Mg^{2+} ion. The thiohydroxamate adduct is shown with carbons in light blue and the surrounding protein residues (gray carbons) whereby red signifies oxygen atoms, nitrogen atoms shown in blue, sulfur atoms shown in yellow, and Mg^{2+} shown in green. Solid lines indicate potential hydrogen bond contacts and electrostatic interactions. **B**. Crystal structure as shown in **A**. with 2Fo-Fc omit map density contoured at 1.5 σ for C_{191}-thiohydroxamate and 1.2 σ for glyoxylate hydrate. **C**. Overlay of 5-NIC-treated ICL1 (gray) with the 3-NP-treated ICL1 structure[17] dark gray carbons and magenta C_{191}-thiohydroxamate adduct.

of 5-NIC (Figure 46A). Consistent with the mass spectrometry analysis, this adduct was lacking one oxygen atom from the presumed nitro group, as previously reported for exogenously-added 3-NP by Ray *et al.*[234] Therefore, we modeled this adduct as (*E*)-3-(hydroxyimino)propanoic acid whereby the C3 forms a covalent bound with the sulfur of Cys_{191}, in which this ligand refined favorably into the electron density (Figure 46B). This Cys_{191}-thiohydroxamate adduct observed is within hydrogen bonding distance to ND1 of His_{193}, NH1 of Arg_{228}, OG1 of Ser_{315}, OG1 of Ser_{317}, ND2 of Asn_{313}, OG1 of Thr_{347}, the backbone nitrogen of Gly_{192}, and Mg^{2+}-coordinated water. When directly comparing our

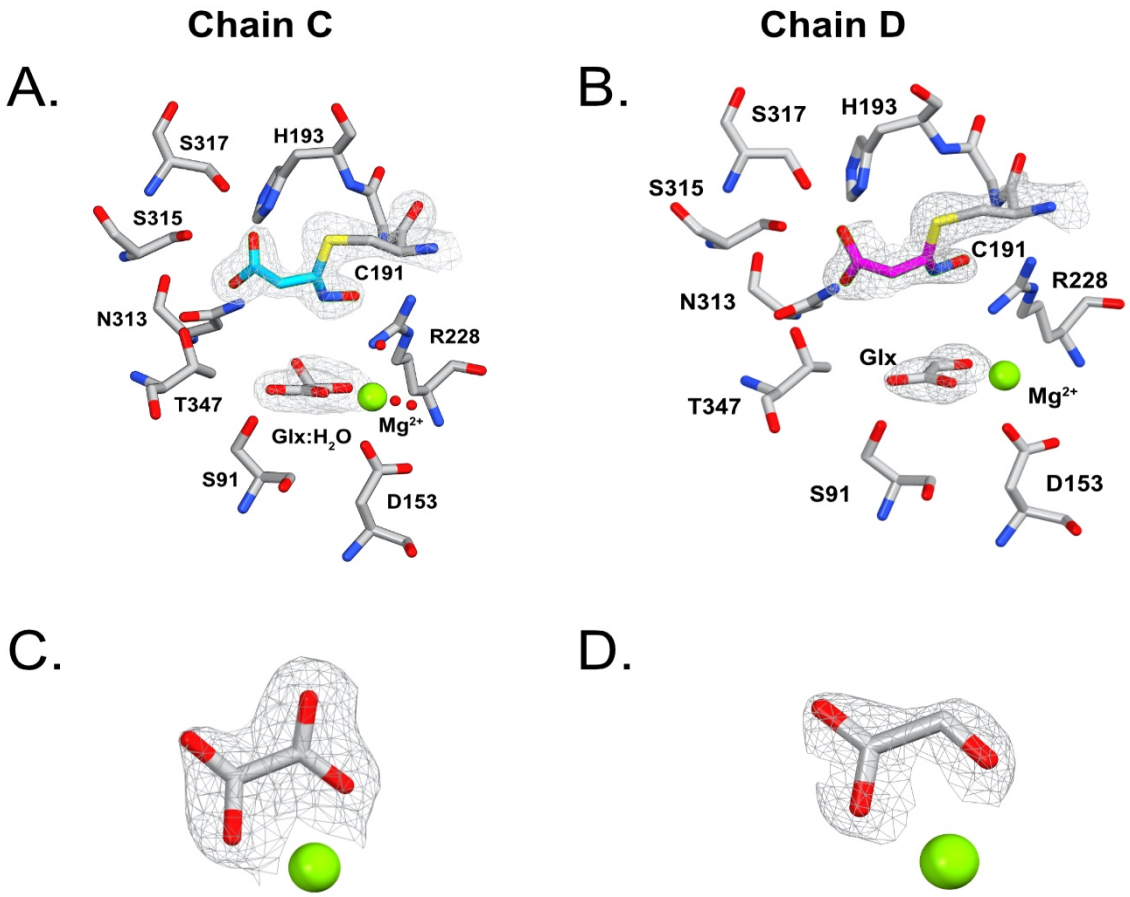


Figure 47. X-Ray Crystal Structures of ICL1 Treated with 5-NIC (chains C and D)

A. Chain C, as seen in Figure 46, exhibits strong density for the Cys_{191}-thiohydroxamate adduct (1.5 σ) and the glyoxylate hydrate (1.2 σ). **B**. Chain D exhibits weaker density for the Cys_{191}-thiohydroxamate adduct (0.9 σ) and strong density for glyoxylate in its aldehyde form (1.9 σ) demonstrating that this active site of ICL1 exhibits lower occupancy of the covalent adduct between Cys_{191} and the thiohydroximate species which in turn led to a minimization of density appearance off the C2 of glyoxylate. **C**. Blowup of 2Fo-Fc omit density map of chain C contoured at 1.2 σ and fit to glyoxylate hydrate. **D**. Blowup of 2Fo-Fc omit density map of chain D contoured at 1.9 σ and fit to glyoxylate.

structure with that of 3-NP covalently bound to ICL, the covalent ligands overlay nicely as active site adducts, and apparently form hydrogen bonds with the same key residues of ICL1 (Figure 46C). In addition, we visualized strong density for a keto acid coordinated

to the active site divalent Mg^{2+} ion in the ICL1 active site, consistent with Glx being one of the cleavage products of 5-NIC. After the refinement of glyoxylate, Fo-Fc difference electron density displayed a positive peak off of its C2 carbon. In our previous work we identified similar density on the C2 of glyoxylate upon inactivation of ICL1 by *cis*-epoxysuccinate (*cis*-EpS), that we assigned as the hydrate of glyoxylate.[14] Subsequent diminishment of this C2 density in a crystallographic snapshot of *cis*-EpS inactivation allowed us to propose a model whereby water may serve as proton donor of the alkoxide which results from ring opening of the *cis*-EpS epoxide, followed by addition of the resulting hydroxide to the aldehyde of glyoxylate. Therefore, the presence of this additional density of the C2 carbon of glyoxylate in our 5-NIC structure suggested that the water molecule abstracted from 3-NP, in order to form the Cys_{191}-thiohydroxamate adduct, could undergo addition to the aldehyde of glyoxylate, forming the *gem*-diol seen in the structure. Thus, we modeled the acid coordinated to Mg^{2+} of ICL1 as glyoxylate hydrate which fits well into the density. In addition, one active site of ICL1 underwent incomplete inactivation as visualized by the lack of strong homogeneous electron density between Cys_{191} and the active site ligand. In this active site we observed a diminished level of density off of the C2 of glyoxylate, suggesting that hydration likely occurs concomitant with the dehydration of the nitronic acid of 3-NP (Figure 47 and Figure 48). Collectively, our structural data coupled with our kinetic and mass spectrometry data allows the proposal of a putative chemical mechanism as to how 5-NIC inactivation of ICL1 precedes (Figure 48).

ix. Proposed 5-NIC Inactivation Mechanism of ICL1.

From a biochemical perspective, the nitro group has been shown to undergo various chemical transformations in the active sites of enzymes and in cells resulting in its formation into an electrophilic species. Some of the earliest reported cases of covalent modification of enzymes by nitro groups came from the laboratory of Harold Bright, who found that a variety of nitro- containing compounds underwent covalent addition to N5 of the flavin cofactor of D-amino acid oxidase.[254] Subsequent studies from Bright, and a host of others, identified many nitro compounds as substrates, inhibitors, and suicide substrates of numerous enzymes.[255] In particular, the molecule 3-NP, and derivatives thereof, has been extensively used in studies and has been identified as a substrate analog, mechanism-based inactivator, and covalent modifier of various enzymes,[255] including ICL1 and succinate dehydrogenase.[229, 234, 247] Recently, the Schreiber laboratory identified nitroisoxazole ligands as masked electrophiles in which the nitro group undergoes dehydration in the reductive intracellular environment to yield a nitrile oxide, which ultimately affords efficient covalent modification of a selenocysteine of glutathione peroxidase 4, a cancer target.[256] In addition, it has been uncovered in *Mtb* that benzothiazinones molecules, containing a nitro group off the C8 of the benzothiazinone pharmacophore, effect covalent inactivation of the enzyme decaprenylphosphoryl-*ß*-D-ribofuranose-2′-epimerase (DprE1, DprE2), a key enzyme in cell wall biosynthesis. This molecule is then proposed to be transformed from its native nitro to a strongly electrophilic

and reactive nitroso through the FAD-dependent conversion, which undergoes subsequent covalent addition to Cys_{394} of the enzyme.[257]

In the case of the mechanism-based inactivator 5-NIC, we propose that similar to the inactivation of DprE1 by the benzothiazinone, BTZ043, ICL1 catalysis unmasks the 5-NIC nitro electrophile which results in covalent addition to its active site Cys_{191}. Owing to the elegant work conducted on inactivation of ICL by 3-NP,[234, 247] and mechanistic analysis of D-IC cleavage,[242] we were able to subsume these findings into a putative

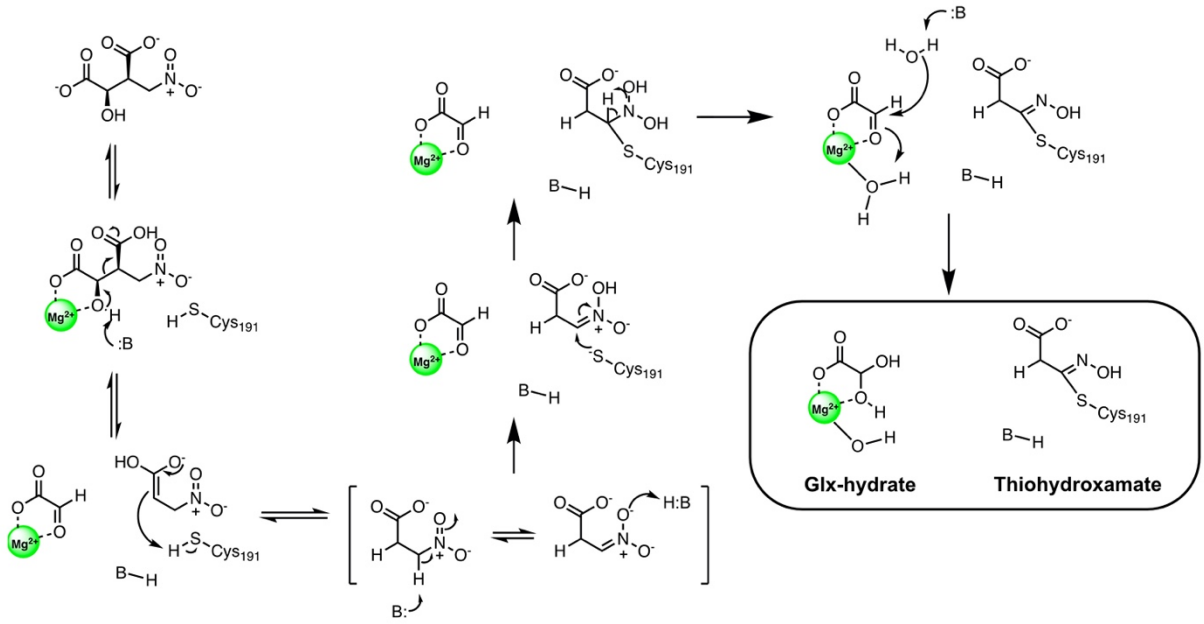


Figure 48. Proposed Inactivation Mechanism of ICL1 by 5-NIC

mechanism for the inactivation of ICL1 by 5-NIC (Figure 48). Initially, binding of 5-NIC competes with substrate D-IC for access to the active site of ICL1. Subsequently, abstraction of a proton from the C2 hydroxyl group of 5-NIC initiates C2-C3 bond cleavage resulting in the formation of glyoxylate and an *aci*-3-nitropropionate

intermediate. This intermediate then abstracts a proton from the catalytic Cys_{191}, resulting in the formation of the thiolate of Cys_{191} and 3-NP in the ICL1 active site. Tautomerization of the 5-NIC derived 3-NP to its P3N form, followed by protonation by an unknown acid, yields the nitronic acid form of P3N. Consequent attack of the nascent Cys_{191} thiolate in a Michael-like addition, forms a covalent bond with 3-NP. The elimination of water from the nitronic acid leads to the formation of the stable thiohydroxamate adduct as is observed in the crystal structure (Figure 4), and supported by our mass spectrometry studies (Figure 41) as well as prior literature.[234] We then propose that the water molecule produced by the dehydration of the P3N nitronic acid in the formation of the thiohydroxamate adduct is activated by an unknown base, and undergoes reaction with the aldehyde group of glyoxylate. Subsequent protonation of the resulting alkoxide by a water-bound Mg^{2+} results in the formation of glyoxylate hydrate, as seen previously,[14] and in Figure 46A,B. This water-bound magnesium has been proposed to protonate the C2 hydroxyl of glyoxylate in the aldol condensation between glyoxylate and succinate to synthesize D-IC.[242]

D. Conclusions and Future Outlook

The development of mechanism-based inactivators of enzymes requires a detailed understanding of an enzymatic mechanism and, often, elaborate synthetic chemistry. We prepared the compound (*2R,3R*)-2-hydroxy-3-(nitromethyl)succinic acid, or 5-NIC, which is in effect a composite of glycolate and 3-nitropropionic acid (3-NP). 5-NIC proved to be the most potent mechanism-based inactivator of mycobacterial ICL1 to date. 5-NIC

undergoes retro-aldol cleavage by ICL1 to generate 3-NP in its active site, which subsequently reacts with the Cys_{191} thiolate to covalently modify ICL1, as determined by kinetics, mass spectrometry, and crystallography. Akin to 3-NP, the nitronate form of 5-NIC (N-5-NIC) is not required for productive binding to ICL1, but does affect more robust inactivation of ICL1, likely through enrichment of the P3N species proposed to be required for ICL1 inactivation. As we observed limited enhancement in inactivation efficiency for the nitronate form of 5-NIC compared to the nitronate form of 3-NP (P3N), we used a combination of absorbance spectroscopy, mass spectrometry, NMR, and solvent kinetic isotope effects to uncover that not only was the rate of N-5-NIC protonation more rapid that P3N, but also that the rapid protonation of N-5-NIC may be due to an intramolecular proton transfer from its C2 hydroxyl to the C4 carbon of the nitronate. Together, this work highlights the development, characterization, and unique physiochemical properties of a masked electrophilic isocitrate substrate surrogate which effects robust covalent inactivation of a key mycobacterial enzyme.

Moving forward, we seek to further optimize and develop mechanism-based inactivators for key *Mtb* enzymes such as ICL1. One of the principal challenges we, and others[170] have faced is the limited ability to translate these molecules from potent *in vitro* molecules into whole cell active anti-mycobacterial agents. Thus, one key area to focus on may be the development of activatable pro-drug forms of these molecules. One approach that is often used in drug design it to take advantage of intracellular esterases. In the case of 5-NIC we noted that one key liability may be its di-acid character which prohibits its entry and thus development of ester prodrugs may be one possible approach.

At present there are 40 putatively predicted esterases, a number of which have been verified using activity based protein profiling with serine reactive fluorophosphonate probes[258] as well as using far-red fluorogenic substrates.[259] While these studies indeed identified various hydrolases in the cells that can de-esterify substrates, two key features seen with these *Mtb* esterases is that they prefer longer chain (fatty acyl) esters[259] and that there is a significant decrease in *Mtb* esterase activity under hypoxic conditions. While the former issue may not pose critical concerns for drug-design purposes, the latter may. The addition of long chain fatty acyl esters to a molecule will markedly enhance the cLogP of a molecule, which may assist in the ability of the molecule to more effectively traverse the highly hydrophobic mycolic acid decorated *Mtb* cell wall. However, even if this molecule can now traverse the cell wall more effectively, there will now be the concern of the molecule not undergoing efficient activation by esterases owing to their low expression under the hypoxic state that *Mtb* resides in most of its life cycle. Of further importance, the vast majority of TB infections are latent (~90%) whereby the bacilli reside in a quiescent state in the hypoxic environment of the granuloma. Thus, new therapeutics for TB treatment must have the ability to treat both actively replicating and bacteria that is not only an active state, but also a hypoxic and quiescent state. Therefore, more clever approaches that simple esterification of a molecule may be required to result in efficient delivery of molecules to *Mtb*, particularly bacilli that are residing within granulomas.

CHAPTER V

RV1151C FROM *MYCOBACTERIUM TUBERCULOSIS* IS A PROTEIN LYSINE DESUCCINYLASE AND LONG-CHAIN FATTY DEACYLASE

A. Introduction

The post-translational modification (PTM) of the ε-amine of lysine residues serves as a rapid chemical switch by which organisms can modulate the function of proteins.[260, 261] Generally, these modifications are affixed to lysine residues of a protein by an acyl transferase enzyme or a chemical ligation event with a reactive metabolite such as succinyl CoA.[262] The subsequent reversal of these acyl modifications, via the hydrolysis of an amide bond, is catalyzed by either a NAD^+-dependent sirtuin-like enzyme or a metal dependent histone deacetylase.[263] In addition to these classes of enzymes, recent work using activity-based protein profiling identified an orphan serine hydrolase enzyme that removed various lysine acyl modifications from peptides and proteins, suggesting there may be other uncharacterized enzymes that carry out lysine deacylation.[264]

Silent regulators of information (SIR) genes modulate of the post-translational modification of lysine residues on proteins. These enzymes, conserved from prokaryotes to eukaryotes, catalyze the hydrolysis of an amide bond of acyl modified lysine residues in an NAD^+-dependent fashion, liberating one molecule of nicotinamide and 2'-O-acyl-ADP-ribose per deacylation event (Figure 49). In humans, there are seven known sirtuin enzymes (SIRT1-7) that mediate the removal of a large array of acyl modifications.[265] These enzymes, in addition to their drastically differing substrate specificities, have been

demonstrated to play a role in diverse biological processes such as aging, gene regulation, energy metabolism, signal transduction, and clinically relevant pathologies in areas such as cancer, cardiovascular diseases, neurodegeneration, and metabolic disease.[266]

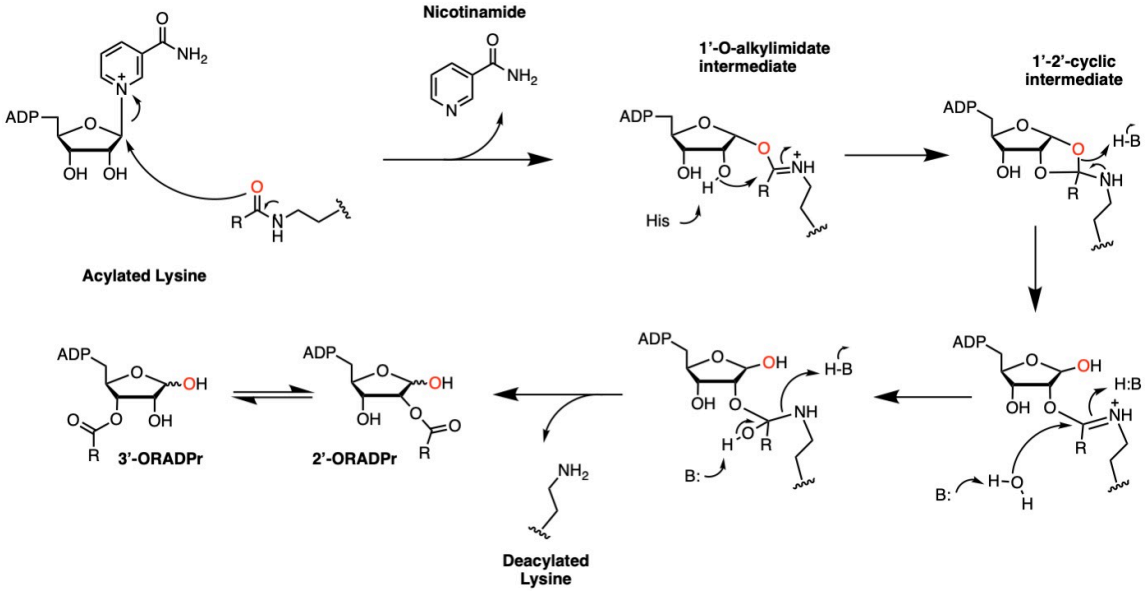


Figure 49. Proposed Chemical Mechanism of an NAD^+-dependent Sirtuin.

Of the myriad of lysine PTMs that exist, one of the most extensively studied is lysine acetylation (Kac).[267] This modification is abundant in organisms ranging from prokaryotes to eukaryotes, and has been shown to play a role in metabolic regulation through modulating the activity of enzymes, such as acetyl CoA synthetase.[268] Recent advances in mass spectrometry have elucidated dozens of new and structurally diverse types of lysine acyl modifications that contain charged groups (glutararyl, malonyl, and succinyl), variable chain lengths (propionyl, butryl, crotonyl, myristoyl, palmiytoyl, etc.) and even cofactors such as biotin and lipoamide.[269]

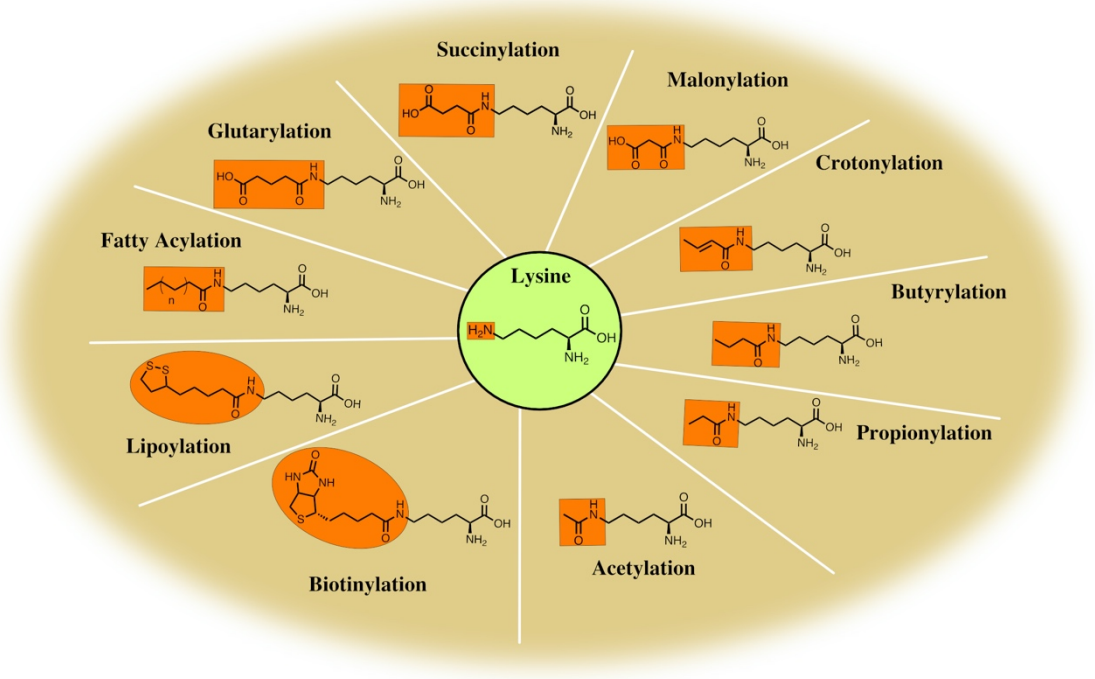


Figure 50. Chemical Structures of Select Lysine Post-Translational Modifications that Exist in Nature

While the focus of this work is on small-molecule, amide-based, lysine PTMs, we would be remiss to not mention that other important forms of lysine PTMs exist, such as methylation and ubiquitination.[270] Collectively, many of these lysine PTMs, partially in prokaryotic systems, have yet to be been functionally characterized with respect to the role they may play on key modified protein targets and the ability of these diverse functionalities to serve as substrates of deacylase enzymes.[271]

Until recent years, lysine acylations were not heavily explored in prokaryotes as researchers believed these modifications were not as prevalent or important in regulation when compared to eukaryotic systems.[272] However, proteomic data for many of gram negative and positive bacterium highlights the abundance of lysine PTMs on many

proteins that play critical roles in central carbon metabolism.[272] Of particular interest to us is post-translational modification of the highly successful bacterial pathogen *Mycobacterium tuberculosis.* This organism has been shown to undergo acetylation, glutarylation, propionlyation, and succinylation on thousands of lysine residues on hundreds of discreet proteins.[273-276] However, there still remains the possibility that *Mtb* contains additional uncharacterized lysine PTMs of varying chemical diversity, which may encompass hundreds, if not thousands, of additional modification sites. Characterized modifications have been shown to reside on a range of proteins, but are particularly enriched on enzymes that play key roles metabolic pathways within *Mtb*, suggestive that some of these modifications may be involved in the regulation of cellular metabolism.[273-276] In the case of *Mtb*, akin to many other prokaryotes, there is only one annotated NAD^+-dependent SIR gene encoded in its genome. Although there are limited comprehensive studies on the substrate specificity of bacterial Sir2 enzymes, many have demonstrated efficient hydrolysis of chemically diverse lysine acyl modifications. For example, the *E.coli* enzyme CobB has been identified to hydrolyze lysine acyl modifications ranging from N-acetyl, to N-succinyl, N-lipoamide, and most recently N-2-hydroxyisobutyrylation.[277-280] However, for many of these bacterial deacylases, such as the *Mtb* deacylase, Rv1151c (which we name *Mt*-Sirt), have yet to have their substrate specificity and kinetic mechanism characterized. Of particular interest to *Mtb* biology, is the ability of *Mt*-Sirt to remove lysine modifications on essential proteins which may play a role in the lifecycle of *Mtb* within the host. Work by Ritterhaus *et al*. demonstrated that an acetyltransferase, Mt-PAT, plays a critical role in the survival of *Mtb* under non-

replicating growth conditions, owing to its ability to remodel the metabolic state of *Mtb* to favor reductive reactions.[281] It is possible that *Mt*-Sirt plays a role in this remodeling and although it has been shown to be non-essential for *Mtb* growth under varied carbon sources,[197, 282, 283] is still important to understand the intricate role(s) that this deacylase plays within this intracellular pathogen.

Here, we present the first detailed kinetic study of *Mt*-Sirt. Analysis of synthetic substrates derived from the sequence of the known *Mt*-Sirt substate, acetyl CoA synthetase (ACS), showed that *Mt*-Sirt it is an efficient desuccinylase and fatty deacylase. Using a fatty acid alkyne probe, we identified that protein lysine fatty acylation occurs in the *Mtb* proteome, suggestive that the *in vitro* fatty deacylase activity of *Mt*-Sirt may play a yet uncharacterized role *in vivo*. Additional studies verified the inhibitory properties of nicotinamide, ADP-ribose, and NADH against *Mt*-Sirt, and synthesis of thioamide and thiourea mechanism-based inhibitors, mimicking a fatty acyl lysine, provided potent inhibitors against this enzyme.

B. Experimental Methods and Materials

i. Reagents and General Synthetic Considerations

All chemical reagents were purchased from commercial providers and used without purification unless otherwise noted. Reactions were run under an atmosphere of nitrogen or argon and at ambient temperature unless otherwise noted. Reaction progress was monitored using thin-layer chromatography and by analysis employing an HPLC−MS (UltiMate 3000 equipped with a diode array coupled to an MSQ Plus single or ISQ-EM

single quadrupole mass spectrometer, Thermo Fisher Scientific) using electrospray positive and negative ionization detectors. Reported liquid chromatography retention times (t_R) were established using the following conditions: column: Phenomenex Luna 5µm C18(2) 100 Å, 4.6 mm, 50 mm; mobile phase A: water with 0.1% formic acid (v/v); mobile phase B: MeCN with 0.1% formic acid (v/v); temperature: 25°C; gradient: 0−100% B over 6 min, then a 2 min hold at 100% B; flow: 1 mL min^{-1}; and detection: MS and UV at 254, 280, 214, and 350 nm. Semi-preparative HPLC purification of compounds was performed on a Thermo Fisher Scientific UltiMate 3000 with a single wavelength detector coupled to a fraction collector. Purifications were conducted using the following conditions: column: Phenomenex Luna 5µmC18(2) 100 Å, 21.2 mm, 250 mm; mobile phase A: water with 0.1%formic acid (v/v); mobile phase B: MeCN with 0.1% formic acid (v/v); temperature: room temperature; gradient: 0−100% B over 30 min, then a 5 min hold at 100% B; flow: 20 mL min−1; and detection: UV (350 nm).$^{1}H/^{13}C$ NMR magnetic resonance spectra were obtained in $CDCl_3$,CD_3OD, or DMSO-d6 at 400 MHz/100 MHz at 298 K on a Bruker AVANCE III Nanobay console with an Ascend magnet unless otherwise noted. The following abbreviations were utilized to describe peak patterns when appropriate: br = broad, s = singlet, d = doublet, q= quartet, t = triplet, and m = multiplet. All final compounds used for testing in assays and biological studies had purities that were determined to be >95% as evaluated by their proton NMR spectra and their HPLC/MS based on ultraviolet detection at 254 nm. Masses detected were in the range of 100−1000 Da and were detected in the positive or negative mode, depending on the ionization of the molecule.

ii. Expression and Purification of Rv1151c (Mt-Sirt) and Mutants Thereof

A pET28a vector containing the sequence for the Rv1151c gene with an N-terminal hexahistidine tag (a gift from the laboratory of Dr. John Blanchard) was transformed into the T7 Express lys^{Y}/I^{q} strain of *E. coli*. A single colony was picked and expanded overnight at 37°C in LB media with 30 ug/mL kanamycin. A 10 mL inoculum was added to 1L of LB media containing 30 ug/mL kanamycin and grown at 37°C shaking at 200 rpm until reaching midlog phase (OD_{600} ~0.6). Protein synthesis was induced by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and the cells were allowed to grow overnight at 25°C. The cells were harvested by centrifugation (5K rpm) and stored at -20°C. The cells were lysed in 50 mM sodium phosphate buffer, 300 mM NaCl, and 10 mM imidazole (pH 7.5) by sonication (50% duty cycle) using a 5 sec pulse for 20 min with 30 sec in between pulses. The resulting solution was centrifuged at 16,000g for 45 min at 4°C to remove cell debris and the supernatant was filtered with a 0.45 µm filter and loaded onto a 5 mL Nickel NTA agarose column (Thermo Fisher) equilibrated with lysis buffer. The solution was allowed to bind to the column for 1 h and then was washed three times with 50 mM sodium phosphate buffer, 300 mM NaCl, and 25 mM imidazole (pH 7.5) and the protein was eluted in 50 mM sodium phosphate buffer, 300 mM NaCl, and 250 mM imidazole (pH 7.5). Following SDS-PAGE analysis of the fractions, the fractions containing pure Rv1151c (~26 kDa) were pooled and dialyzed three times against 50 mM Tris-HCl, 150 mM NaCl, pH 7.5. The protein was stored in a 50% v/v solution of glycerol at -80 ºC. For the preparation of Rv1151c mutants we used the overlapping primer design approach and carried out mutagenesis using the Pfu Turbo DNA polymerase (Agilent

Technologies). We verified the mutations to the gene by sequencing and then transformed, expressed and purified the mutant proteins ($Y_{53}F$ and $R_{56}M$) using the same methods as detailed above.

iii. Synthesis of Cbz-Lysine-7-amino-4-methylcoumarin Substrates for Mt-Sirt

Method A. Amide formation with 7-amino-4-methylcoumarin

Cbz-Lys and 7-amino-4-methylcoumarin (1mmol/1mmol lysine) was purged with $N_{2(g)}$ yielding positive pressure. Anhydrous tetrahydrafuran (12 mL/mmol lysine) was was added followed by the addition of N,N-diisopropylethylamine (3 mmol/1 mmol lysine). Propylphosphonic anhydride (T3P) in a 50% wt. solution of ethyl acetate (2 mmol/1 mmol lysine) was added dropwise followed by allowing the reaction to proceed overnight. Solvents were removed under reduced pressure and the yellow oil was solubilized in DCM and washed with water (4x) and brine (2x). The organic layer was dried over Na_2SO_4 and removed under reduced pressure.

N^6-acetyl-N^2-((benzyloxy)carbonyl)-L-lysine Cbz-Lys(ac)-OH).[284] A flask charged with Cbz-Lys-OH (300 mg, 1.07 mmol) was purged with $N_{2(g)}$. Then anhydrous pyridine (10 mL) was added followed by the dropwise addition of acetic anhydride (101 uL, 1.07 mmol). The reaction was allowed to proceed overnight followed by acidification to pH 1 with 2 N HCl and extraction with ethyl acetate (3x 20 mL). The pooled organic layers were then washed with 5% $NaCO_3$ and this bicarbonate layer was acidified to pH 1 and extracted with ethyl acetate (3x, 20 mL). To combined organic layers were washed with

brine (2x, 20 mL), dried over Na_2SO_4, and concentrated under reduced pressure. The product was moved directly to the next step with no purification.

benzyl (S)-(6-acetamido-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxohexan-2-yl)carbamate (Cbz-Lys(ac)-AMC). Method A. Z-Lys(ac)-OH (256 mg, 0.8 mmol) and 7-amino-4-methylcoumarin (140 mg, 0.8 mmol), N,N-diisopropylethylamine (420 µL, 2.44 mmol), T3P (950 mL, 1.6 mmol). Semi-preparative HPLC of the molecule resulted in 90 mg (23.5% yield). LC-MS t_R: 4.14 min, m/z 480.07 [M+1H], $C_{16}H_{22}N_2O_5$, Calcd. 480.21 [M+1H]

benzyl (S)-(1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxo-6-propionamidohexan-2-yl)carbamate (Cbz-Lys(prop)-AMC). Z-Lys-OH (150 mg, 0.535 mmol) was dissolved in 1N NaOH (2 mL) / THF (2mL), covered with a rubber septum, purged with $N_{2(g)}$ and cooled to 0ºC on an ice bath. Propionyl chloride (50 uL, 0.56 mmol) was added dropwise and the reaction was allowed to warm to RT and was stirred overnight. The reaction was then cooled in an ice bath and washed with ice-cold Et_2O (10 mL). The aqueous phase was acidified to pH ~4 with 1N HCl and extracted with ethyl acetate (20 mL, 3x). The combined organic phases were dried over Na_2SO_4, and concentrated under reduced pressure to yield the crude product which was progressed to the next step without purification. Subsequent coupling using Method A. Cbz-Lys(prop)-OH (78 mg, 0.24 mmol), 7-amino-4-methyl coumarin (42 mg, 0.24 mmol), THF (5 mL), DIEA, (167 µL, 0.96 mmol), T3P (286 µL, 0.48 mmol). Semi-preparative HPLC of the

crude molecule resulted in pure title molecule (28.4 mg, 24% yield). LC-MS t_R: 4.37 min, m/z 492.16 [M-1H], $C_{27}H_{31}N_3O_6$, Calcd. 492.22 [M-1H]

benzyl (S)-(6-butyramido-1-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-1-oxohexan-2-yl)carbamate (Cbz-Lys(but)-AMC). Z-Lys-OH (150 mg, 0.535 mmol) was dissolved in 1N NaOH (2 mL) / THF (2mL), covered with a rubber septum, purged with $N_{2(g)}$ and cooled to 0ºC on an ice bath. Butyryl chloride (58 uL, 0.56 mmol) was added dropwise and the reaction was allowed to warm to RT and was stirred overnight. The reaction was then cooled in an ice bath and washed with ice-cold Et_2O (10 mL). The aqueous phase was acidified to pH ~4 with 1N HCl and extracted with ethyl acetate (20 mL, 3x). The combined organic phases were dried over Na_2SO_4, and concentrated under reduced pressure to yield the crude product which was progressed to the next step without purification. Subsequent coupling using Method A. Cbz-Lys(but)-OH (112 mg, 0.32 mmol), 7-amino-4-methylcoumarin (56 mg, 0.24 mmol), THF (7 mL), DIEA, (223 µL, 1.29 mmol), T3P (381 µL, 0.64 mmol). Semi-preparative HPLC of the crude molecule resulted in pure title molecule (45.6 mg, 28% yield). LC-MS t_R: 4.580 min, m/z 506.19 [M+1H], $C_{28}H_{33}N_3O_6$, Calcd. 506.24 [M+1H]

benzyl tert-butyl (6-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-6-oxohexane-1,5-diyl)(S)-dicarbamate (Cbz-Lys(boc)-AMC). Method A. Cbz-Lys(boc)-OH (400 mg, 1.05 mmol), 7-amino-4-methylcoumarin (184 mg, 1.05 mmol), THF (12 mL), DIEA (550 µL, 3.15 mmol), T3P (1.25 mL, 2.1 mmol). Semi-preparative HPLC yielded (344 mg,

61% yield). LC-MS t_R: 5.27 min, m/z 438.1 [M+1H, -Boc], $C_{29}H_{35}N_3O_7$, Calcd. 538.25 [M+1H], 438.2[M+1, -Boc]

(S)-4-((5-(((benzyloxy)carbonyl)amino)-6-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-6-oxohexyl)amino)-4-oxobutanoic acid (Cbz-Lys(succ)-AMC). A round bottom flask charged with Z-Lys(boc)-AMC (150 mg, 0.28 mmol) was purged with $N_{2(g)}$ yielding positive pressure. Anhydrous DCM (4 mL) was added followed by 33% v/v TFA (2 mL) and the reaction was allowed to proceed for 1.5 h at which time the solvents were removed under reduced pressure. The residue was co-evaporated with Et_2O (2x, 10 mL) and then moved to the next step. To the flask succinic anhydride (28 mg, 0.28 mmol) was added followed by purging the flask with $N_{2(g)}$ to yield positive pressure. Anhydrous DMF (3 mL) and then DIEA (300 uL, 1.7 mmol) was added. The reaction was allowed to proceed overnight at RT at which time the reaction mixture was diluted with H_2O (20 mL), acidified to pH 2, and extracted with ethyl acetate (3x, 20 mL). The pooled organic layer was washed with brine (2x, 15 mL) dried over Na_2SO_4, and concentrated under reduced pressure. This crude product was dissolved in a minimal amount of DMF and subjected to semi-preparative HPLC purification to yield the pure title compound (94 mg, 62% yield). LC-MS t_R: 4.03 min, m/z 538.15 [M+1H], $C_{28}H_{31}N_3O_8$, Calcd. 538.21 [M+1H]

(*S*)-3-((5-(((benzyloxy)carbonyl)amino)-6-((4-methyl-2-oxo-2*H*-chromen-7-yl)amino)-6-oxohexyl)amino)-3-oxopropanoic acid (Cbz-Lys(mal)-AMC). A round bottom flask charged with Z-Lys(boc)-AMC (150 mg, 0.28 mmol) was purged with $N_{2(g)}$

yielding positive pressure. Anhydrous DCM (4 mL) was added followed by 33% v/v TFA (2 mL) and the reaction was allowed to proceed for 1.5 h at which time the solvents were removed under reduced pressure to yield Cbz-Lys-AMC • TFA. The residue was co-evaporated with Et_2 (2x, 10 mL) and then moved to the next step. To the flask, mono-*t*-butyl malonate (71 mg, 0.46 mmol) was added followed by purging with $N_{2(g)}$ to yield positive pressure. Anhydrous DCM (5 ml) was added followed by the addition of N,N-diisopropylethylamine (265 μL, 1.52 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (452 mL, 0.76 mmol) was added dropwise and allowed to proceed overnight at RT. The reaction mixture was diluted in DCM (~30 mL) and washed with water (15 mL, 4x) and brine (15 mL, 2x). The organic layer was dried over Na_2SO_4, and solvents were removed under reduced pressure. The crude product was then solubilized in DCM (5 mL) and 40% TFA v/v (2 mL) and the solvents were removed upon completion as determined by LC-MS. The crude residue was dissolved in a minimal amount of DMF and subjected to semi-preparative HPLC purification to yield the pure title compound (90.6 mg, 62% yield). LC-MS t_R: 4.08 min, m/z 522.16 [M-1H], $C_{27}H_{29}N_3O_8$, Calcd. 524.2 [M-1H]

5-(*tert*-butoxy)-5-oxopentanoic acid. A round bottom flask charged with glutaric anhydride (2 g, 17.5 mmol) and $ZnCl_2$ (360 mg, 2.6 mmol) was purged with $N_{2(g)}$ yielding positive pressure. Tert-butyl alcohol (10 ml) was added and the reaction was heated to 60ºC and monitored by TLC (p-anisaldehyde stain) and was stopped after 18 h. The reaction mixture was allowed to cool followed by the addition of sat. $NaCO_3$ (20 mL)

followed by extraction with DCM (50 mL, 3x). The aqueous phase was then acidified with conc. H_2SO_4 to pH 1 and extracted with DCM (50 mL, 3x). The acidified organic extract was then washed with brine (2x, 40 mL), dried over Na_2SO_4 and concentrated to yield the title compound as a clear oil in high purity (1.04 g, 31.5% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.29 (s, 1H), 1.46 (s, 9H), 1.94 (q, J = 7.3 Hz, 2H), 2.32 (t, J = 7.3 Hz, 2H), 2.43 (t, J = 7.4 Hz, 2H).

(*S*)-5-((5-(((benzyloxy)carbonyl)amino)-6-((4-methyl-2-oxo-2*H*-chromen-7-yl)amino)-6-oxohexyl)amino)-5-oxopentanoic acid (Cbz-Lys(glut)-AMC). A round bottom flask charged with Z-Lys(boc)-AMC (150 mg, 0.28 mmol) was purged with $N_{2(g)}$ yielding positive pressure. Anhydrous DCM (4 mL) was added followed by 33% v/v TFA (2 mL) and the reaction was allowed to proceed for 1.5 h at which time the solvents were removed under reduced pressure to yield Cbz-Lys-AMC • TFA. The residue was co-evaporated with Et_2 (2x, 10 mL) and then moved to the next step. To the flask 5-(*tert*-butoxy)-5-oxopentanoic acid (105 mg, 0.56 mmol) was added followed by purging the flask with $N_{2(g)}$ to yield positive pressure. Anhydrous DCM (5 ml) was added followed by the addition of N,N-diisopropylethylamine (243 µL, 1.4 mmol). Propylphosphonic anhydride in a 50% wt. solution of ethyl acetate (333 mL, 0.56 mmol) was added dropwise and allowed to proceed overnight at RT. The reaction mixture was diluted in DCM (~30 mL) and washed with water (15 mL, 4x) and brine (15 mL, 2x). The organic layer was dried over Na_2SO_4, and solvents were removed under reduced pressure. The crude product was then solubilized in DCM (5 mL) and 40% TFA v/v (2 mL) and the solvents were

removed upon completion as determined by LC-MS. The crude residue was dissolved in a minimal amount of DMF and subjected to semi-preparative HPLC purification to yield the pure title compound (80.0 mg, 52% yield). LC-MS t_R: 4.06 min, m/z 550.19 [M-1H], $C_{29}H_{33}N_3O_8$, Calcd. 550.23 [M-1H]

iv. Synthesis and Analysis of Acyl ACS Peptides

Peptides corresponding to residues 614-618 of *Mycobacterium tuberculosis* acetyl CoA synthetase (Ac-RSGKI-*W*-OH) containing varied acyl groups at the K_{617} position were synthesized on a 0.100 mmol scale using Fmoc-Trp(boc)-Wang resin (Chem Impex and CEM) on a Liberty Blue automated microwave peptide synthesizer (CEM) under standard conditions. To produce the varied acyl modified lysine residues two strategies were employed: synthesis of Fmoc-Lys(R)-OH amino acids, where R is a predefined chemical group or synthesis of the peptide using an Fmoc-Lys(Mtt)-OH allowing selective deprotection of the lysine amine side chain, affording the use of anhydride chemistry or amide bond coupling to produce the peptide of interest. All peptides were acetylated on the N-terminus with a solution of 25% Ac_2O in N,N-dimethylformamide (DMF) and 1.5 equivalents of N,N-diisopropylethylamine (DIEA) followed by 45 min of agitation at room temperature and ninhydrin test to verify completion. Select peptides synthesized with an Mtt (4-methyltrityl) group underwent treatment with 3% v/v TFA in DCM/5% triisopropylsilane for 20 min (2x) to cleave the Mtt group. Subsequently the resin was washed with DCM (2x), MeOH (2x), DCM (2x), DMF containing 1% v/v DIEA (2x), and DMF (2x). The now free ε-amine of lysine was reacted with an acyl anhydride (10 eq of

anhydride in either DMF or toluene and 1.5 eq of DIEA) ranging from 1 h to overnight (2x) to give the appropriate modified lysine as monitored by ninhydrin test and test cleavage followed by HPLC-MS analysis. The peptides were cleaved (2x for 2-4 h) from the resin using 92.5% trifluoroacetic acid, 5% triisopropylsilane, and 2.5% water or 85% trifluoroacetic acid, 5% w/v phenol, 5% v/v thioanisole, and 5% v/v water. The cleavage solution was evaporated under reduced pressure or under a stream of $N_{2(g)}$ and the resulting residue was precipitated with ice cold diethyl ether, dried over $N_{2(g)}$, dissolved in DMF, and purified on C18 reverse phase semi-preparative HPLC (Luna 5 um C18(2) 100 Å, 21.2 mm, 250 mm, Phenomenex). Fractions containing peptide that was >90% pure, as determined by analytical reverse phase HPLC-MS analysis (Luna 5 um C18(2) 100 Å, 4.6 mm, 250 mm, or a Luna 5 um C18(2) 100 Å, 4.6 mm, 50 mm, Phenomenex) were pooled and lyophilized. Mass spectrometric analysis of the peptides by ESI (ThermoFisher) verified the identity of the peptide(s).

v. Synthesis of Fmoc-modified Lysine Molecules for ACS Peptide Synthesis

Method B. Synthesis of Fmoc-Lysine(x)-OH amino acids

The appropriate carboxylic acid chain and N-hydroxysuccinimide (0.96 mmol/1 mmol acid) was dissolved in anhydrous DMF (0.75 mL/mmol acid) under a positive pressure of $N_{2(g)}$. The dropwise addition of N,N' diisopropylcarbodiimide (0.96 mmol/1 mmol acid) was added to the reaction and stirred for 2 h. The reaction mixture was filtered and the filtrate was directly transferred to a flask containing Fmoc-Lys-OH (0.75 mmol/1 mmol acid) dissolved in DMF (2 mL/1 mmol acid) and N,N-diisopropylethylamine (1.9 mmol/1

mmol acid). Following 30 min, the reaction was quenched with water (1 mL/1 mmol acid), acidified with 1 M HCl (1 mL/1 mmol acid), and extracted with ethyl acetate (3x). The organic layers were washed with saturated bicarbonate (3x), brine (2x), and dried over Na_2SO_4. Silica chromatography (0-3% MeOH/DCM, 0.5% formic acid) gave the title compounds as white/off white fluffy solids.

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-propionyl-*L*-lysine (Fmoc-Lys(prop)-OH).

Method B. Propionic acid (1.2 mL, 16.3 mmol), N-hydroxysuccinimide (1.81 g, 15.7 mmol), DMF (12 mL), N,N' diisopropylcarbodiimide (2.44 mL, 15.7 mmol), Fmoc-Lys-OH (4g, 10.85 mmol), DIEA (5.48 mL, 31.5 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (2.59 g, 56% yield). LC-MS t_R: 4.48 min, m/z 425.27 [M+1H], $C_{24}H_{28}N_2O_5$, Calcd. 425.2 [M+1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-butyryl-*L*-lysine (Fmoc-Lys(but)-OH). Method B. Butanoic acid (1.87 mL, 20.4 mmol), N-hydroxysuccinimide (2.27 g, 19.7 mmol), DMF (15 mL), N,N' diisopropylcarbodiimide (3.05 mL, 19.7 mmol), Fmoc-Lys-OH (5g, 13.6 mmol), DIEA (6.86 mL, 39.4 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (4.21 g, 71% yield). LC-MS t_R: 4.69 min, m/z 439.22 [M+1H], $C_{25}H_{30}N_2O_5$, Calcd. 439.22 [M+1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-tetradecanoyl-*L*-lysine (Fmoc-Lys(myr)-OH). Method B. Crotonic acid (1.75 g, 20.4 mmol), N-hydroxysuccinimide (2.27 g, 19.7 mmol), DMF (15 mL), N,N' diisopropylcarbodiimide (3.05 mL, 19.7 mmol), Fmoc-Lys-OH (5g, 13.6 mmol), DIEA (6.86 mL, 39.4 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (3.0 g, 50% yield). LC-MS t_R: 4.72 min, m/z 437.2 [M+1H], $C_{25}H_{28}N_2O_5$, Calcd. 437.20 [M+1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-tetradecanoyl-*L*-lysine (Fmoc-Lys(myr)-OH). Method B. Myristic acid (2.3 g, 10.2 mmol), N-hydroxysuccinimide (1.3 g, 9.9 mmol), DMF (8 mL), N,N' diisopropylcarbodiimide (1.53 mL, 9.9 mmol), Fmoc-Lys-OH (2.5 g, 6.8 mmol), DIEA (3.43 mL, 19.7 mmol), and DMF (20 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (1.74 g, 44% yield). LC-MS t_R: 7.47 min, m/z 579.2[M+1H], $C_{35}H_{50}N_2O_5$, Calcd. 579.37 [M+1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-(3-(*tert*-butoxy)-3-oxopropanoyl)-*L*-lysine (Fmoc-Lys(mal-O-*t*But)-OH). Method B. Mono-*t*-butyl malonate (3.14 g, 20.4 mmol), N-hydroxysuccinimide (2.27 g, 19.7 mmol), DMF (15 mL), N,N' diisopropylcarbodiimide (3.05 mL, 19.7 mmol), Fmoc-Lys-OH (5g, 13.6 mmol), DIEA (6.86 mL, 39.4 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (5.3 g, 77% yield). LC-MS t_R: 5.06 min, m/z 511.19 [M+1H], $C_{28}H_{34}N_2O_7$, Calcd. 511.24 [M+1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-(4-(*tert*-butoxy)-4-oxobutanoyl)-*L*-lysine (Fmoc-Lys(succ-O-*t*But)-OH). Method B. Mono-*t*-butyl succinate (3.53 g, 20.4 mmol), N-hydroxysuccinimide (2.27 g, 19.7 mmol), DMF (15 mL), N,N' diisopropylcarbodiimide (3.05 mL, 19.7 mmol), Fmoc-Lys-OH (5g, 13.6 mmol), DIEA (6.86 mL, 39.4 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (4.1 g, 57% yield). LC-MS t_R: 5.15 min, m/z 523.32 [M-1H], $C_{29}H_{36}N_2O_7$, Calcd. 524.25 [M-1H]

N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-(5-(*tert*-butoxy)-5-oxopentanoyl)-*L*-lysine (Fmoc-Lys(glut-O-*t*But)-OH). Method B. Mono-*t*-butyl glutaryl (3.06 g, 16.3 mmol), N-hydroxysuccinimide (1.81 g, 15.7 mmol), DMF (12 mL), N,N' diisopropylcarbodiimide (2.44 mL, 15.7 mmol), Fmoc-Lys-OH (4g, 10.85 mmol), DIEA (5.48 mL, 31.5 mmol), and DMF (35 mL). Silica column (0-5% MeOH/DCM, 0.5% formic acid) to yield (3.9 g, 66% yield). LC-MS t_R: 5.23 min, m/z 537.32 [M+1H], $C_{30}H_{38}N_2O_7$, Calcd. 537.27 [M+1H]

vi. Characterization of ACS Acyl Peptides

Peptide	Observed m/z [M+1]	Calculated m/z [M+1]	t_R
ACSK617 formyl	816.4	816.43	2.82 min
ACSK617 acetyl	830.5	830.44	2.53 min
ACSK617 propionyl	844.6	844.46	2.62 min
ACSK617 butyryl	858.5	858.48	2.76 min
ACSK617 crotonyl	856.4	856.46	2.74 min
ACSK617 hexanoyl	886.5	886.51	3.26 min
ACSK617 octanoyl	914.5	914.54	3.89 min
ACSK617 decanoyl	942.5	942.57	4.26 min
ACSK617 dodecanoyl	970.7	970.60	4.32 min
ACSK617 myristoyl	998.8	998.63	4.25 min
ACSK617 palmitoyl	1026.8	1026.66	16.03 min[a]
ACSK617 malonyl	874.5	874.43	2.53 min
ACSK617 succinyl	888.7	888.45	2.57 min
ACSK617 glutaryl	902.6	902.47	2.52 min
ACSK617 biotinoyl	1014.4	1014.51	2.67 min
ACSK617 thiomyristol	1014.5	1014.61	4.70 min

[a] Analysis was conducted using a Luna 5 um C18(2) 100 Å, 4.6 mm, 250 mm at a flow rate of 1 mL/min and a gradient of (30-52% ACN, 0.1%FA (20 mins), 52-100% ACN, 0.1%FA (30 min)).

vii. HPLC Kinetic Analysis of Peptide Hydrolysis by Mt-Sirt

A typical assay steady state assay contained variable concentrations of ACS modified peptide, 50 mM Tris-HCl, 1 mM NAD^+, 1 mM DTT, 0-10% v/v DMSO, pH 8.0, and 0.1 to 1 uM of *Mt*-Sirt to a final reaction volume of 60 uL. This reaction incubated for 5 min at 37°C at which time it was quenched with 60 uL of 10% TFA in water or 0.5N HCl in MeOH followed by centrifugation for 10 min at 15,000 rpm to remove the precipitated protein. The supernatant (110 uL) was then loaded into a 96 well plate and analyzed by HPLC (40 uL injection) with a Kinetex PS C18 reverse phase column (Kinetex® 2.6 µm PS C18 100 Å, LC Column 150 x 3 mm, Phenomenex). For peptides with acidic lysine modifications the gradient was as follows: a 10-min gradient: 95% H_2O/5% MeCN increasing to 60% H_2O/40% MeCN where all solvents contained 0.1% formic acid. For

peptides containing acyl modifications (acetyl to butryl) we employed a 10-min gradient: 95% H_2O/5% MeCN increasing to 40% H_2O/60% MeCN and for longer chain peptides we used a 10-min gradient that started at 95% H_2O/5% MeCN and increased to 0% H_2O/100% MeCN. Product formation was measured by the increase of the product peak using absorbance at 280 nm in comparison to a no enzyme control. All peptide concentrations were verified using an HPLC standard curve of Ac-Trp-NH_2. The area of the product peak was integrated and divided by the sum of the area of the starting material peak and the product peak. This value was then multiplied by the concentration of substrate to report on the conversion in terms of μM. Division of this value by the time of the reaction then gives one a rate with respects to concentration and time (eg. μM sec^{-1}). To ensure steady state conditions, the rates of product formation were all ensured to be less than 10%. The obtained rates were then replotted versus substrate concentration and fitted to the Michaelis-Menten equation (eq. 1) using GraphPad Prism 8.

viii. Alkyne Fatty Acid Labeling of $mc^2$7000

To a 10 mL culture of $mc^2$7000 in 7H9/OADC medium at OD_{600} 0.8-1, a 50 mM solution of Alk-12 in DMSO was added to the cells to reach a final concentration of 50 uM (10 uL). In addition, a control culture with DMSO alone was also created. These cells were grown for an additional 24 h at 37ºC at which time they were spun down, washed 6x with ice cold PBS buffer and stored at -20°C. The subsequent pellet was resuspended in resuspended in 50 μL lysozyme (0.1 mg/mL) and 500 μL 200 mM Tris-HCl pH 8, 4% CHAPS, 1 M NaCl, 8 M urea, 5 mM nicotinamide, and 50 uL cOmplete EDTA-free

Protease Cocktail Inhibitor (1 tablet dissolved in 2 mL ddH_20, 25x, Roche). Cells were disrupted using a bead-beating homogenizer with silica-zirconium beads for 60 sec, followed by 3 min on ice (8x). The lysate was centrifuged at 13,000 rpm at 4 °C for 15 min and the supernatant was removed followed by determination of the protein concentration by BCA assay. Labeling of the Alk-12 treated cell lysates occurred in a stepwise fashion whereby addition of TBTA in 1:4 DMSO: n-butanol (v/v) (100 μM final) was added to $CuSO_4$ in water (1 mM final), followed by the addition of Cy7-azide (Click Chemistry Tools) in DMSO (20 μM final), 24 μg cell lysate, and finally sodium ascorbate in water (2.5 mM final). This reaction was incubated at room temperature for 1.5 h, and quenched by the addition ice-cold methanol (200 μL), chloroform (75 μL), and water (150 μL), followed by vortexing and centrifugation at 15,000 x g at 4 ºC for 15 min. The top layer of liquid was carefully removed and discarded, and the protein precipitate formed between two liquid phases was not disturbed. An additional 1 mL of ice-cold methanol was added into the tube, vortexed, and centrifuged at 15,000 x g at 4 ºC for 20 min. The supernatant was discarded and this was repeated an additional time. Following removal of the supernatant the protein pellet was allowed to air-dry at room temperature. To the dried pellet 4% SDS buffer (4% SDS (w/v), 50 mM TEA, 150 mM NaCl, pH 7.4) was added, and the pellet was sonicated to achieve a clarified solution. Then, the solution was either added to 4x loading dye (240 mM Tris-HCl, 8% SDS (w/v), 0.004 % bromophenol blue (to avoid fluorescence from the dye), DTT (5 mM final), and MilliQ water to a final volume of 24 μLor 4x loading dye (240 mM Tris-HCl, 8% SDS (w/v), 0.004 % bromophenol blue (to avoid fluorescence from the dye), DTT (5 mM final), 60 mM

hydroxylamine (pH 7.2), and MilliQ water to a final volume of 24 µL. The samples were heated to 95 ºC for 10 min, followed by a brief centrifugation at 15,000 x g for 1 min. The ladder (Bio-Rad, Precision Plus Protein™ Dual Color Standards) was diluted 10,000-fold in 4% SDS buffer and 4x loading dye to avoid strong fluorescence from the ladder during imaging. Samples were analyzed on a 4-12% Bis-Tris Nu-PAGE gel (Invitrogen). To each well, 20 µg of labeled lysate was loaded, and the gel was run at 60 V for 30 min followed by 120 V for ~90 min. The gels were destained (10% acetic acid, 40% methanol, 50% water) for 1 hr to overnight in the dark followed by imaging the gel on a ChemiDoc imager (Bio-Rad) using the Cy7 blot setting. To visualize the two fluorescent proteins of the protein standard ladder, we used the rhodamine setting. The gels were also analyzed by Coomassie Brilliant Blue R-250 staining (Bio-Rad) to ensure equal loading.

ix. Inhibition Kinetics of Mt-Sirt

Variable concentration of inhibitor in DMSO or water, fixed concentration of ACSK617W succinyl peptide (100-200 µM), 50 mM Tris-HCl, 1 mM NAD^+, 1 mM DTT, pH 8.0, was combined and then 0.05-0.2 µM of *Mt*-Sirt was added to initiate the reaction. The reaction was allowed to incubate for 5-30 min at 37°C at which time it was quenched with 60 uL of 10% TFA in water which was followed by centrifugation for 10 min at 15,000 rpm to remove the precipitated protein. The supernatant (110 uL) was loaded into a 96 well plate and analyzed by HPLC (40 uL injection) with a Kinetex PS C18 reverse phase column (Kinetex® 2.6 µm PS C18 100 Å, LC Column 150 x 3 mm, Phenomenex) and analyzed using a 10-min gradient: 95% H_2O/5% acetonitrile (MeCN) increasing to 60% H_2O/40%

MeCN where all solvents contained 0.1% formic acid. Analysis of product peaks and calculation of reaction velocity was conducted identically to that described in the steady-state kinetics section.

x. Synthesis of Inhibitors of Mt-Sirt

benzyl tert-butyl (6-oxo-6-(phenylamino)hexane-1,5-diyl)(S)-dicarbamate (Cbz-Lys(Boc)-Aniline). Cbz-Lys(Boc)-OH (500 mg, 1.31 mmol) was added to a round bottom flask and purged with Ar(g) and dissolved in anhydrous DCM (25 mL). Subsequently, DIEA (687 µL, 3.94 mmol) and aniline (120 µL, 1.31 mmol) were added to the solution followed by the dropwise addition of 50% T3P in ethyl acetate (1.56 mL, 2.63 mmol) to the flask. Following two h, or until HPLC showed depletion of starting material the reaction was quenched with H_2O (20 mL), followed by washing of the organic layer with H_2O (4x, 20 mL), 0.5 M HCl (1 x 15 mL), and brine (2 x 20 mL). The organic extract was dried over anhydrous $NaSO_4$, filtered, and concentrated under reduced pressure and the off-white solid was moved directly to the next step without purification. LC-MS t_R: 5.7 min, m/z 356.1 [M+1H, -Boc], $C_{25}H_{33}N_3O_5$, HF-18 Calcd. 456.24 [M+1H], 356.19 [M+1, -Boc]

benzyl (S)-(6-amino-1-oxo-1-(phenylamino)hexan-2-yl)carbamate (Cbz-Lys-Aniline • TFA). Cbz-Lys(Boc)-aniline (482 mg, 1.06 mmol) was added to a flask and dissolved in DCM (20 mL) under Ar(g) followed by the addition of 25% v/v trifluoroacetic acid (TFA), 6.5 mL). The reaction was stirred for 4 h, or until deemed complete by TLC and

LC-MS at which time solvent was removed under reduced pressure. This intermediate was moved to the next step without further purification. LC-MS t_R: 3.39 min, m/z 356.1 [M+1H], $C_{20}H_{25}N_3O_3$, Calcd. 356.19 [M+1H]

benzyl (S)-(1-oxo-1-(phenylamino)-6-thiocyanatohexan-2-yl)carbamate (Cbz-Lys(SCN)-aniline. Cbz-Lys-aniline • TFA (150 mg, 0.42 mmol) was added to a vial and purged with $N_{2(g)}$ followed by the addition of anhydrous THF (6 mL). To this solution, solid 1,1′-thiocarbonyl diimidazole (90 mg, 0.51 mmol) and triethylamine (TEA, 145 uL, 1.05 mmol) were added and the reaction was allowed to proceed overnight at RT. The crude reaction mixture was concentrated under reduced pressure and dissolved in ethyl acetate (40 mL) followed by washing with H_2O (15 mL, 3x), brine (15 mL, 1x). The collected organic layer was dried over Na_2SO_4 and concentrated by rotary evaporation to yield an off-white solid which was moved directly to the next step without further purification. LC-MS t_R: 5.87 min, m/z 398.1 [M+1H], $C_{21}H_{23}N_3O_3S$, Calcd. 398.15 [M+1H]

benzyl (*S*)-(6-(3-methylthioureido)-1-oxo-1-(phenylamino)hexan-2-yl)carbamate (1). Cbz-Lys(SCN)-aniline (50 mg, 0.125 mmol) was added to a vial and purged with $N_{2(g)}$ followed by the addition of anhydrous N-N-dimethylformamide (DMF) (1 mL). To this solution, methylamine (33% (w/w) in absolute EtOH, (17.5 µL, 0.18 mmol) and TEA (35 µL, 0.25 mmol) was added at 0ºC. The reaction was allowed to warm to room temperature and proceed overnight at RT at which time it was diluted in H_2O and extracted with ethyl

acetate (3x, 20 mL) and the organic layer was washed brine (15 mL, 2x), dried over dried over Na_2SO_4 and concentrated by rotary evaporation. The crude material was dissolved in a small volume of DMF and subjected to semi-preparative HPLC yielding the pure title compound as a white solid (47.2 mg, 92% yield). LC-MS t_R: 4.80 min, m/z 429.1 [M+1H], $C_{22}H_{28}N_4O_3S$, Calcd. 429.19 [M+1H]

benzyl (S)-(6-(3-dodecylthioureido)-1-oxo-1-(phenylamino)hexan-2-yl)carbamate. (2) Cbz-Lys(SCN)-aniline (50 mg, 0.125 mmol) was added to a vial and purged with $N_{2(g)}$ followed by the addition of anhydrous DMF (1 mL). To this solution, dodecylamine (35 mg, 0.19 mmol) and TEA (35 µL, 0.25 mmol) was added at 0ºC. The reaction was allowed to warm to room temperature and proceed overnight at RT at which time it was diluted in H_2O and extracted with ethyl acetate (3x, 20 mL) and the organic layer was washed brine (15 mL, 2x), dried over dried over Na_2SO_4 and concentrated by rotary evaporation. The crude material was purified silica column chromatography (30-50% ethyl acetate/hexanes) which following concentration under reduced pressure yielded 48.6 mg, 69% yield as a white solid. LC-MS t_R: 7.6 min, m/z 583.3 [M+1H], $C_{33}H_{50}N_4O_3S$, Calcd. 583.36 [M+1H]

***tert*-butyl 2-isothiocyanatoacetate.** t-butyl glycinate (500 mg, 3.8 mmol) was added to a round bottom flask, purged with Ar(g), followed by the addition of DCM (25 mL). To this solution triethylamine (2.64 mL, 19 mmol) and carbon disulfide (0.69 mL, 11.4 mmol) were added followed by stirring the solution for 1 h. The reaction mixture was then cooled to 0°C in an ice bath and 50% T3P in ethyl acetate (4.52 mL, 7.6 mmol) was added

dropwise. The reaction was allowed to warm to room temperature and stirred for an additional 2 h, or until TLC analysis showed reaction completion using a p-anisaldehyde staining. Following completion, the reaction mixture was quenched with 10 mL of water for 30 min to hydrolyze residual T3P and diluted with DCM (250 mL). The organic layer was washed with H2O (2 x 20 mL), 1 M HCL (2 x 20 mL), H2O (2 x 20 mL), saturated sodium bicarbonate (2 x 20 mL), H2O (1 x 20 mL), and brine (1 x 20 mL). The organic extract was then dried over anhydrous NaSO4, filtered, and concentrated under reduced pressure. The crude residue was purified via silica column chromatography (2% ethyl acetate/hexanes) which following concentration yielded the title product as a pale yellow liquid (460 mg, 69%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.44 (s, 3H), 4.03 (s, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 27.97, 47.06, 83.93, 138.21, 164.98.

***tert*-butyl 3-thiocyanatopropanoate. β-alanine t-butyl ester hydrochloride** (1 g, 5.5 mmol) was added to a round bottom flask, purged with Ar(g), followed by the addition of DCM (50 mL). To this solution triethylamine (3.84 mL, 27.5 mmol) and carbon disulfide (0.99 mL, 16.5 mmol) were added followed by stirring the solution for 1 h. The reaction mixture was then cooled to 0°C in an ice bath and 50% T3P in ethyl acetate (5.9 mL, 9.9 mmol) was added dropwise. The reaction was allowed to warm to room temperature and stirred for an additional 2 h, or until TLC analysis showed reaction completion using a p-anisaldehyde staining. Following completion, the reaction mixture was quenched with 10 mL of water for 30 min to hydrolyze residual T3P and diluted with DCM (500 mL). The organic layer was washed with H2O (2 x 40 mL), 1 M HCL (2 x 40 mL), H2O (2 x 40

mL), saturated sodium bicarbonate (2 x 40 mL), H2O (1 x 40 mL), and brine (1 x 40 mL). The organic extract was then dried over anhydrous NaSO4, filtered, and concentrated under reduced pressure. The crude residue was purified via silica column chromatography (2% ethyl acetate/hexanes) which following concentration yielded the title product as a pale yellow liquid (863 mg, 83%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.49 (s, 1H), 2.61 (dd, *J* = 7.2, 5.9 Hz, 1H), 3.75 (dd, *J* = 7.2, 5.9 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 28.02, 35.87, 41.00, 81.82, 132.09, 168.96.

(**S)-((5-(((benzyloxy)carbonyl)amino)-6-oxo-6-(phenylamino)hexyl)carbamothioyl)glycine (3).** Cbz-Lys-Aniline • TFA (50 mg, 0.14 mmol) was added to a round bottom flask, purged with Ar(g), followed by the addition of anhydrous DMF (1 mL) and N,N-diisopropylethylamine (DIEA) (49 μL, 0.28 mmol). Then, tert-butyl 2-isothiocyanatoacetate (27.2 mg, 0.16 mmol) was added to the flask followed by stirring overnight at RT. Subsequently the reaction was diluted with H2O, extracted with ethyl acetate (3 x 100 mL) and the combined organic layers were washed with brine (1 x 75 mL), dried over anhydrous NaSO4 and concentrated under reduced pressure. The crude product was then dissolved in anhydrous DCM (2 mL) and 15% v/v TFA (0.3 mL). Following 1 h, or upon completion as determined by LC-MS, the reaction was concentrated under reduced pressure. Purification via semi-preparative HPLC yielded (48.6 mg, 73.5% yield) All compounds were verified by NMR and HPLC-MS data. LC-MS t_R: 4.6 min, m/z 473.1 [M+1H], $C_{23}H_{28}N_4O_5S$, Calcd. 473.18 [M+1H]

(S)-3-oxo-1-phenyl-5-(phenylcarbamoyl)-11-thioxo-2-oxa-4,10,12-triazapentadecan-15-oic acid (4). Cbz-Lys-Aniline • TFA (50 mg, 0.14 mmol) was added to a round bottom flask, purged with Ar(g), followed by the addition of anhydrous DMF (1 mL) and N,N-diisopropylethylamine (DIEA) (49 µL, 0.28 mmol). Then, tert-butyl 3-thiocyanatopropanoate (29.3 mg, 0.155 mmol) was added to the flask followed by stirring overnight at RT. Subsequently the reaction was diluted with H2O, extracted with ethyl acetate (3 x 100 mL) and the combined organic layers were washed with brine (1 x 75 mL), dried over anhydrous NaSO4 and concentrated under reduced pressure. The crude product was then dissolved in anhydrous DCM (2 mL) and 15% v/v TFA (0.3 mL). Following 1 h, or upon completion as determined by LC-MS, the reaction was concentrated under reduced pressure. Purification via semi-preparative HPLC yielded (29.4 mg, 43% yield). LC-MS t_R: 4.63 min, m/z 487.2 [M+1H], $C_{24}H_{30}N_4O_5S$, Calcd. 488.19 [M+1H]

***N*^{2}-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-*N*^{6}-tetradecanethioyl-*L*-lysine (Fmoc-Lys(thiomyr)-OH)** A round bottom flask charged with Z-Lys(myr)-OH (1 g, 1.73 mmol) and Lawesson's reagent (630 mg, 1.56 mmol) was purged with $N_{2(g)}$ yielding positive pressure. Anhydrous THF (30 mL) was added and the reaction was allowed to react at RT overnight. The solvents were removed and the crude reaction mixture was purified by silica column (0.5% MeOH/DCM, 1% formic acid) to yield the title compound (250 mg, 24% yield). LC-MS t_R: 8.1 min, m/z 595.3 [M+1H], $C_{35}H_{50}N_2O_4S$, Calcd. 595.35 [M+1H]

xi. Kinetic Analysis

Replotting of the initial velocity data versus substrate concentration resulted in hyperbolic plots which were fitted to eq. 1:

$$v = \frac{V_{max}\,[S]}{K_m + [S]} \qquad (\mathbf{eq.\,1})$$

Where V_{max} is the maximal velocity achieved, [S] is the molar concentration of the substrate, and K_m is the substate concentration at which the velocity is half V_{max}. Thereby allowing determination of the rate of catalysis, or the k_{cat} using eq. 2:

$$V_{max} = k_{cat}\,[E_t] \qquad (\mathbf{eq.\,2})$$

Where, k_{cat} is the number of molecules made per enzyme per a unit of time and $[E_t]$ is the molar concentration of enzyme in the assay conditions.

Determination of the half-maximal inhibitory concentration (IC_{50}) we fit the data to the following equations:

$$Y = \frac{100}{(1 + 10^{([logI] - logIC50)})} \qquad (\mathbf{eq.\,19})$$

Where [logI] is the $\log_{10}$ of the inhibitor concentration in molar.

Or,

$$Y = \frac{y_{min} + (y_{max} - y_{min})}{(1 + 10^{(logIC50 - [logI] * h)})} \quad \textbf{(eq. 20)}$$

Where y_{min} and y_{max} represent the lower and upper bound of the percent reduction of enzyme activity and *h* is the Hill slope.

C. Results and Discussion

i. Mt-Sirt Contains and Active Site YxxR Motif

Initial sequence alignment of *Mt*-Sirt with other known sirtuins revealed that it contains an YxxR motif (Y_{53} and R_{56}), reminiscent of human sirtuin 5 (hSirt5) and bacterial sirtuins from *E.coli* and *S. typhimurium* (CobB) which also contain this conserved motif (Figure 51A). Although we made extensive efforts to determine a crystal structure of *Mt*-Sirt, we were unable to produce crystals that diffracted to high resolution. Therefore, to try and better understand how this enzyme may function, we created a homology model of the protein using Phyre2 (Figure 51B).[285] Overlay of this homology model with crystal structures of archaeal, human, and bacterial NAD^+ dependent sirtuin enzymes assisted in identifying key active site motifs required for substrate recognition. Initially, superposition with the structure of *E.coli* CobB (PDB: 6RXJ)[286] bound to an acetyl lysine peptide resulted in a moderate overall structural homology with *Mt*-Sirt (RMSD: 0.93 Å). In particular, our attention was drawn to helix bundle six, which contains the YxxR motif of CobB (Y_{92} and R_{95}) (Figure 51C). While the tyrosine residues of both proteins overlaid nicely, we noticed that the R_{95} guanidinium group of *E. coli* CobB pointed away from the active site cavity, while the R_{56} of *Mt*-Sirt pointed inward towards this cavity. The positioning of R_{95} away from the active site cavity may be due to electrostatic repulsion between the hydrophobic methyl of the bound acetyl lysine peptide and the electropositive R_{95} guanidinium. Further, this "shift" is allowed as the area proximal to R_{95} in the CobB structure contains a polar Q_{98} residue that may afford

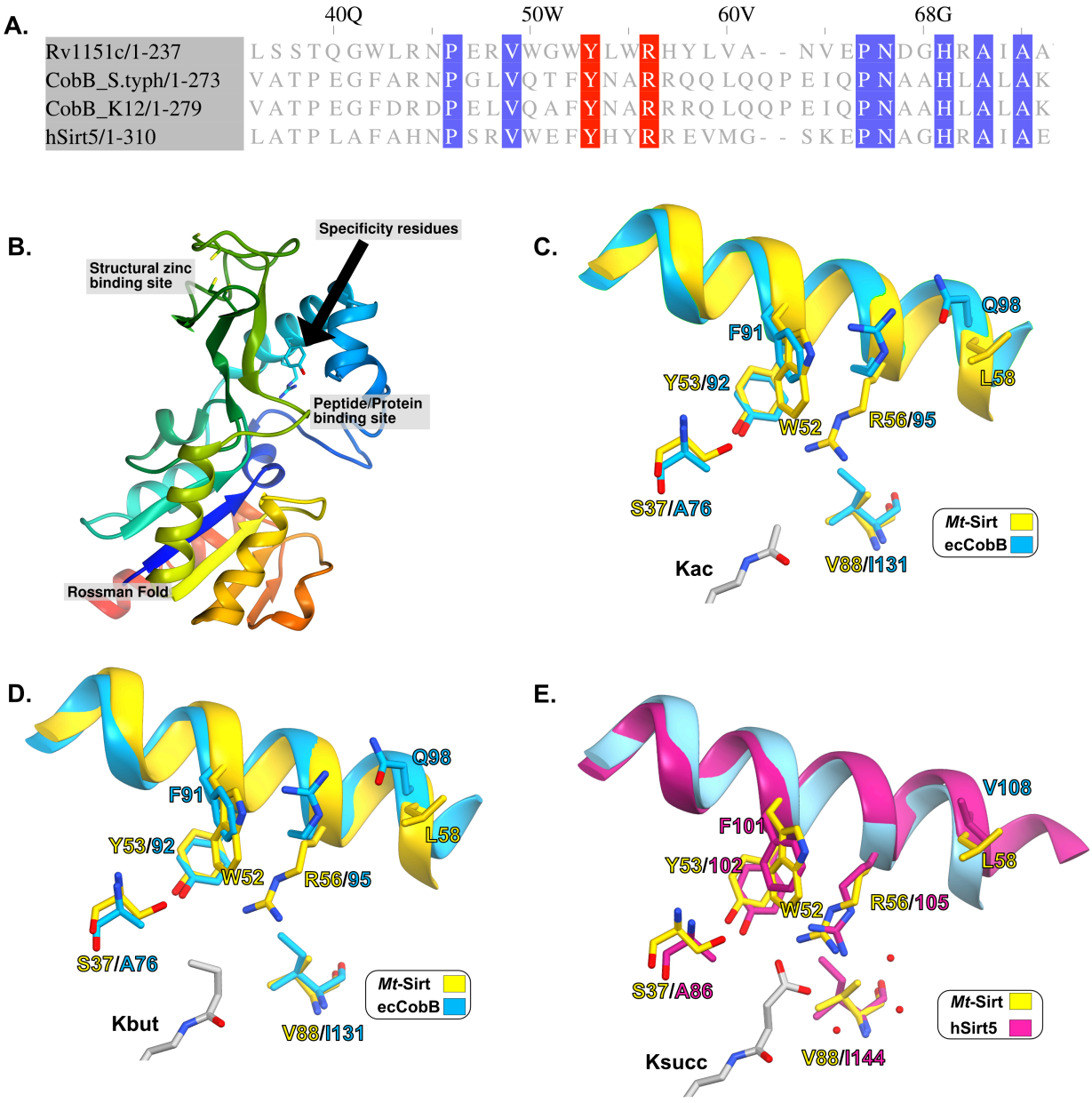


Figure 51. Sequence Homology of *Mt*-Sirt and Overlay of Homology Model

A. Sequence alignment of Rv1151c (*Mt*-Sirt) with two bacterial NAD^+ dependent deacylases and human sirtuin 5. Blue highlighted regions demarcate 80% identity and the red highlighted regions correspond to the YxxR motif of the enzymes. **B**. Homology model of *Mt*-Sirt. **C**. Overlay of *Mt*-Sirt homology model (gold) with *E.coli* CobB (PDB: 6RXJ) (blue) bound to an acetylated peptide (gray). **D.** Overlay of *Mt*-Sirt homology model (gold) with *E.coli* CobB (PDB: 6RXK) (blue) bound to an butyryl lysine peptide (gray). **E**. Overlay of *Mt*-Sirt homology model (gold) with human sirtuin 5 (PDB: 3RIY) (pink) bound to a succinyl peptide (cyan).

hydrogen bonding with R_{95} guanidinium and a nonpolar F_{91} which could afford cation-π interaction.

Conversely, this type of movement is unlikely to be observed in *Mt*-Sirt, as the Q_{98} has been replaced with a leucine (L_{59}). Beyond this difference, it has been recently shown in CobB that A_{76}, in concert with I_{131}, may play a role in the efficient catalysis of short chain fatty acyl peptides through a substrate induced-fit mechanism.[286] Interestingly, overlay of *Mt*-Sirt with CobB revealed that in lieu of an alanine at this position, *Mt*-Sirt contained a serine (S_{37}) (Figure 51C,D). Close inspection of CobB bound to a butyryl lysine peptide (PDB: 6RXK) reveals that the methyl of the butyryl lysine group resides proximal to A_{76} (3.4 Å), whereas the observed proximity of the polar S_{37} of *Mt*-Sirt to the overlaid butyryl lysine peptide may suggest *Mt*-Sirt has a reduced preference for short chain hydrophobic lysine residues (Figure 51D). Subsequent alignment of the *Mt*-Sirt model with the known desuccinylase, human sirtuin 5 (hSirt5), bound to a succinyl peptide (PDB: 3RIY),[287] resulted in clear overlay of arginine and tyrosine sidechains of both *Mt*-Sirt and hSirt5 (Figure 51E). In addition, hSirt5 and *Mt*-Sirt both contain a hydrophobic residue (V_{108} and L_{58}, respectively) where the Q_{98} residue was seen for CobB. This hydrophobic substitution would presumably disallow masking of the polar character of the arginine guanidinium group, and thus may play role in the limited tolerance of hSirt5 for short chain hydrophobic lysine PTMs[287, 288] when compared to *E.coli* CobB.[289] Together, this information suggests that *Mt*-Sirt may prefer acidic lysine modifications, such as succinyllysine, when compared to shorter chain hydrophobic PTMs such as acetyllysine.[278, 287]

ii. Mt-Sirt Shows a Preference for N-succinyllysine

It has been previously shown that *Mt*-Sirt can utilize the acetyl-lysine containing carboxybenzyl-lysine(acetyl)-7-amino-4-methylcoumarin (Cbz-Lys(acetyl)-AMC) as a substrate, albeit no kinetic parameters were reported.[290] Therefore, to test the ability of *Mt*-Sirt to hydrolyze the amide bond of varied lysine PTMs, we synthesized six different Cbz-Lys(**R**)-AMC substrates (Figure 52A). Initial analysis demonstrated that *Mt*-Sirt tolerates all of the molecules as substrates, but most robustly deacylated Cbz-Lys(succ)-AMC (Figure 52B). This is consistent with our hypothesis that *Mt*-Sirt has a preference for succinyl lysine substrates and corroborates the recent finding that *Mt*-Sirt can desuccinylate chemically succinylated EchA19, which was published while were conducting this work.[291] Subsequent kinetic analysis determined that *Mt*-Sirt bound to this succinyllysine substrate 10-fold tighter (K_m = 16.5 μM vs. 155 μM) and catalyzed the deacylation 1.6-fold faster (0.015 vs. 0.009 s^{-1}) when compared to its acetyllysine counterpart (data not shown). While we were unable to conduct complete kinetic analysis of these substrates due to poor solubility, we did note that *Mt*-Sirt exerts weak demalonase and deglutarase activity, suggestive that the N-succinyl lysine PTM may provide optimal binding interactions with the active site Y_{53} and R_{56}. In addition, the poor activity observed with short chain hydrophobic lysine PTMs may result from weak binding to the enzyme (no saturation up to 250 μM for propionyl and butyryl, data not shown) due to the polar character of R_{56} or S_{37}.

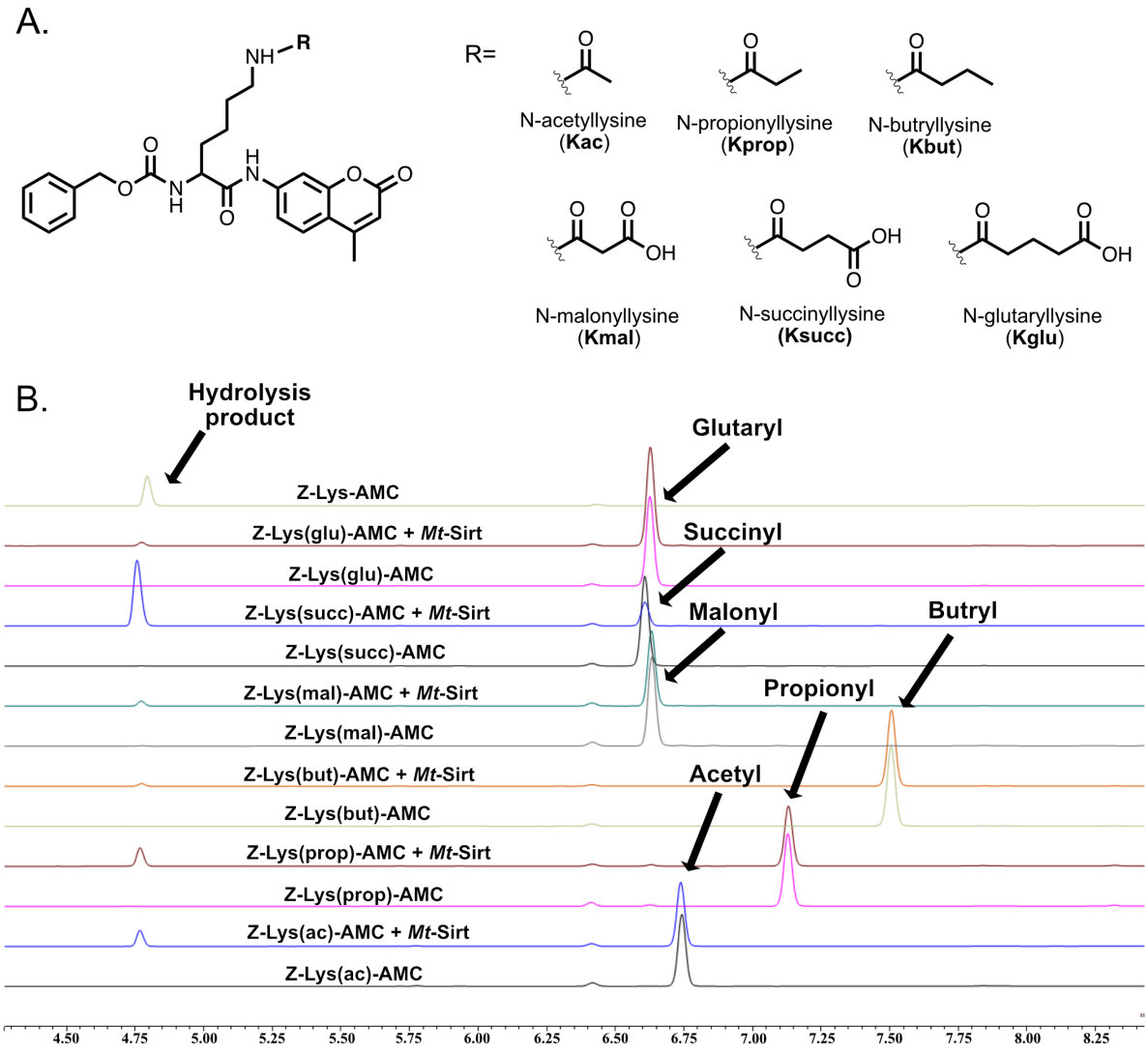


Figure 52. *Mt*-Sirt Preferentially Catalyzes the Deacylation of Succinyl Lysine *in vitro*

A. Panel of 7-amino-4-methylcoumarin containing acylated mono-lysine compounds synthesized. **B**. High performance liquid chromatography traces demonstrating the Rv1151c catalyzed hydrolysis of the different acyl sidechains of the mono-lysine-AMC substrates shown in A. Reactions were carried out with 50 uM substrate, 1 uM *Mt*-Sirt, 50 mM Tris-HCl, 1 mM NAD^+ and 1 mM DTT, pH 8.0 at 37°C for 30 min followed by quenching with 10% TFA v/v in water.

iii. Design and Synthesis of an Acetyl CoA Synthetase-derived Substrate Series

As mono-lysine substrates containing a bulky, hydrophobic, N-terminal carboxybenzyl and C-terminal 7-amino-4-methylcoumarin groups do not recapitulate a protein or peptide substrate that *Mt*-Sirt would function on in the cell, we prepared a series of peptidic substrates to gain further insight into the substrate specificity of this enzyme. These peptides are derived from the acetyl CoA synthetase (ACS), specifically the sequence encompassing lysine 617. This lysine residue has been shown to be acetylated by Ms-Pat, a cAMP dependent GCN5-related N-acetyltransferase (GNAT), which leads to specific inactivation of the acetyl-adenylate forming reaction of ACS.[292] Reversion of ACS activity is obtained by the removal of the acetyl group from K_{617} by *Mt*-Sirt, underscoring how lysine PTMs can provide *Mtb* a reversible regulatory mechanism to fine-tune the activity of metabolic enzymes.

These ACS peptides were prepared using solid phase peptide synthesis and were six amino acids in length (Ac-*RSGK*(**R**)*I*-W-OH). Five of these residues comprised the $ACSK_{617}$ sequence (shown in italics) and the sixth is a C-terminal tryptophan residue to aid in HPLC assay analysis and for purification ease. In these peptides **R** is made up a variety of lysine PTMs, some of which are verified lysine PTMs that occur in *Mtb*, while others have yet to be identified (Figure 53). Together, these substrates provide substantial chemical diversity that allows us to probe the substrate specificity of *Mt*-Sirt.

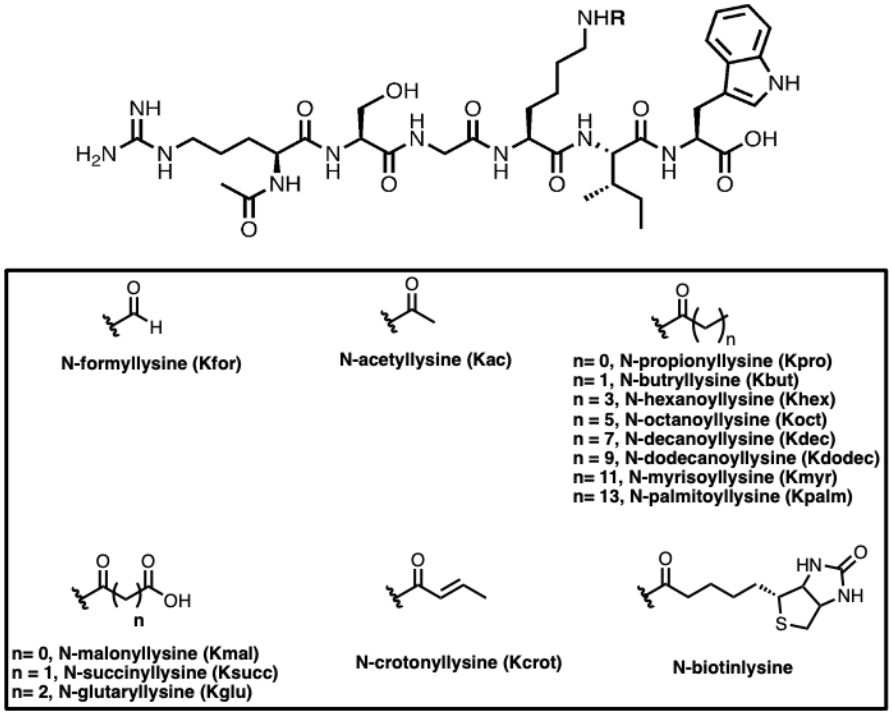


Figure 53. Structure of ACS Peptide Employed and of the Varied ε-Amine Lysine Groups Appended to this Peptide.

iv. Mt-Sirt is a Robust Desuccinylase and Fatty Deacylase

Characterization of *Mt*-Sirt deacylase activity with the synthetic acetyl CoA synthetase K_{617} derived peptides was performed using saturating NAD^+ (1 mM) and variable peptide concentrations to determine the steady state kinetic parameters k_{cat}, $K_{m,substrate}$, and $k_{cat}/K_{m,substrate}$. Initially, we employed a nicotinamide coupled assay,[293] to determine deacylation rates of *Mt*-Sirt continuously, but owing to detection limitations for poorer substrates and interference of fatty acyl peptides with the assay, we instead employed a discontinuous HPLC based assay for all substrates. Consistent with our previous analysis of the mono-lysine AMC substrates, *Mt*-Sirt displayed the highest level

of catalytic efficiency for the succinyl lysine ACS modified peptide ($k_{cat}/K_m = 3.8 \pm 0.3$ x 10^3 sec^{-1} M^{-1}) (Figure 54). The enzyme displayed quite poor tolerance for the malonyllysine peptide compared to the succinyllysine peptide (105-fold reduction in k_{cat}/K_m, respectively), which was principally derived from a 32-fold increase in the K_m of this substrate, suggesting the bound malonyl side chain may not be within hydrogen bonding distance of R_{56} of *Mt*-Sirt, or may bind in a conformation that results in a clash with hydrophobic residues proximal to R_{56}.

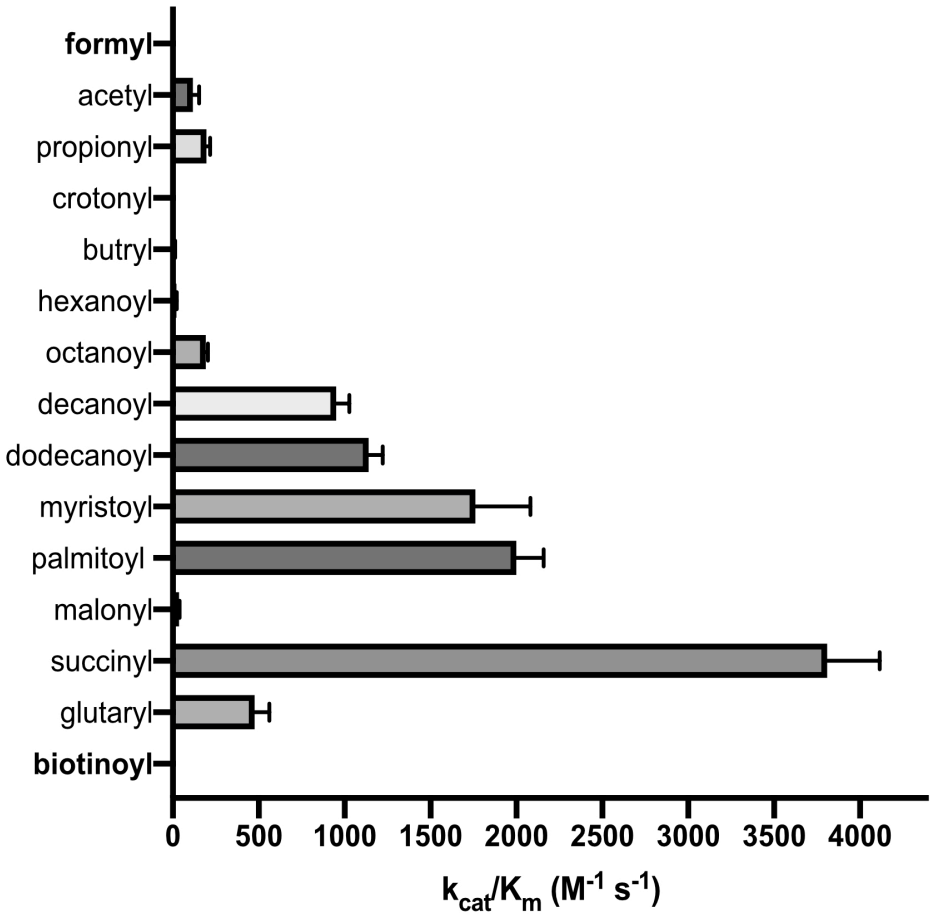


Figure 54. Catalytic Efficiency of the ACS Peptide Substrates Tested

Subsequent analysis of peptides containing short chain acyl groups, such as N-acetyl, -propionyl, and -butyryllysine, also resulted in a marked decrease in catalytic efficiency when compared to that of N-succinyllysine. This poor catalytic efficiency was also mainly derived from heightened K_m (24.5- and 16-fold increase of K_m of N-acetyllysine and N-propionyllysine versus N-succinyllysine, respectively). To probe if the increased K_m resulted from the polar character of R_{56} and Y_{53} we mutated these residues and conducted kinetic analysis using both the propionyllysine and succinyllysine peptides.

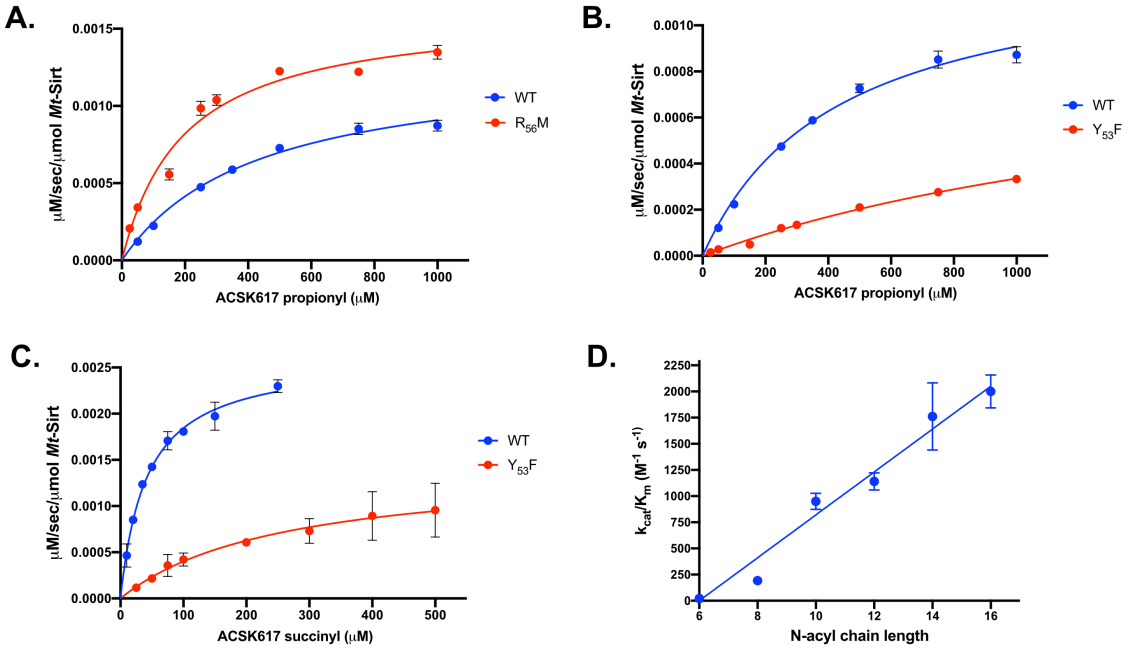


Figure 55. Analysis of Mt-Sirt Mutants with Propionyl and Succinyl Peptides

A. Steady-state rates of WT and $R_{56}M$ *Mt*-Sirt with the ACSK617 propionyl peptide fit to eq. 1. B. Steady-state rates of WT and $Y_{53}F$ *Mt*-Sirt with the ACSK617 propionyl peptide fit to eq. 1. C. Steady-state rates of WT and $Y_{53}F$ *Mt*-Sirt with the ACSK617 succinyl peptide fir to eq. 1. D. Plotting of the N-acyl carbon chain length of ACS peptides versus k_{cat}/K_m of the substrate. Line through data is determined by fitting to y = mx + b.

Table 2. Steady-State Kinetics of *Mt*-Sirt Mutants with a Propionyl and Succinyl Peptide

Substrate	Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
ACSK617 Kprop	$Y_{53}F$	0.06 ± 0.008	1885 ± 344	3.2 ± 0.7 x 10^1
	$R_{56}M$	0.1 ± 0.02	196 ± 28	5.1 ± 0.1 x 10^2
ACSK617 Ksucc	$Y_{53}F$	0.09 ± 0.01	258 ± 89	3.2 ± 1.0 x 10^2
	$R_{56}M$	N/A[a]	N/A[a]	N/A[a]

[a]No detectible product was formed in the presence of 200 μM ACSK617 Ksucc.

Steady state analysis of the propionyllysine peptide with a $R_{56}M$ mutant resulted in a 2.1-fold reduction in the K_m compared to wild-type *Mt*-Sirt (196 μM vs. 410 μM) and a 1.25-fold increase in k_{cat} (0.1 vs. 0.08 sec^{-1}), respectively (Figure 55A, Table 2). Further analysis with a *Mt*-Sirt $Y_{53}F$ mutant led to a 4.6-fold increase in K_m and a 1.3-fold decrease in k_{cat} when compared to wild-type (Figure 55B, Table 2). In contrast, evaluation of the succinyllysine peptide with the $R_{56}M$ mutant with the resulted in no detectible desuccinylase activity, which is consistent with previous findings in hSirt5 and highlights the importance of this arginine residue.[287] The incubation with the succinylysine substrate with the $Y_{53}F$ mutant resulted in a 6.2-fold increase in the observed K_m and a 1.8-fold reduction in k_{cat} when compared to wild-type (Figure 55C, Table 2).. In sum, substitution of the polar arginine 56 with more nonpolar methionine affords enhanced selectivity and slightly more robust catalysis for the short chain hydrophobic propionyllysine substrate, but mutation of Y_{53} to F_{53} results in kinetic perturbations for both substrates. This may indicate that in addition to the phenol of Y_{53} likely forming a hydrogen bond with the succinyllysine peptide that this polar character may also be important in the positioning of

other key active site residues required for efficient binding of the propionyllysine substrate.

Beyond study of peptides containing short chain acyl and acidic lysine PTMs, we also prepared longer chain fatty acyl lysine peptides. It has been observed that hSirt5, which has a similar active site to *Mt*-Sirt (Figure 51), exhibits fatty deacylase activity.[288] Therefore, we sought to determine if *Mt*-Sirt also has fatty deacylase activity. To our surprise, we noticed that as acyl chain length extended beyond butyryllysine, *Mt*-Sirt began to exert increasingly robust deacylation of peptide substrates, consistent with efficient fatty deacylase activity (Figure 54). In fact, the best fatty acyl lysine substrate (ACSK617 palmitoyl) exhibited only a 2-fold reduction in k_{cat}/K_m when compared to the succinyllysine substrate. We also noticed a linear correlation ($R^2 = 0.95$) between fatty acyl lysine chain length (hexanoyl to palmitoyl) of the peptide substrate and catalytic efficiency (Figure 55). Close inspection of the kinetic data reveals the major contributing factor to this increase in k_{cat}/K_m is derived from a reduction in K_m with respect to increasing chain length of the fatty acyl lysine peptide. At present, no crystal structures of a sirtuin-like enzyme containing a YxxR motif in complex with a fatty acyl peptide exist to our knowledge. This lack of information into key binding interactions and possible structural shifts that may occur upon fatty acyl lysine binding to hSirt5 and *Mt*-Sirt complicate our ability to understand the key molecular interactions required for this activity. Determination of such crystal structures should be the focus of future work.

v. Fatty Acylation of Proteins Occurs in the Mtb Proteome

Being as we identified that *Mt*-Sirt can remove fatty acyl lysine PTMs, we sought to determine the abundance of fatty acylated proteins in the *Mtb* proteome. Therefore, we cultured an attenuated *Mtb* strain (mc^{2}7000) in the presence of the fatty alkyne metabolic probe, Alk-12. Upon lysis of these cells, we conducted copper mediated click-chemistry to append a fluorogenic Cy7 azide to the alkyne modified target proteins. Following labeling, we treated the samples with hydroxylamine to remove any thioester linked fatty acids, thereby, in theory, leaving only proteins containing amide-linkages to the fatty alkyne probe. Resolving the samples with gel electrophoresis, followed by visualization using in-gel fluorescence, identified a reasonable abundance of Alk-12 modified proteins that exist in the *Mtb* proteome; thus, suggestive that the fatty deacylase activity of *Mt*-Sirt may play relevant and yet to be described intracellular roles.

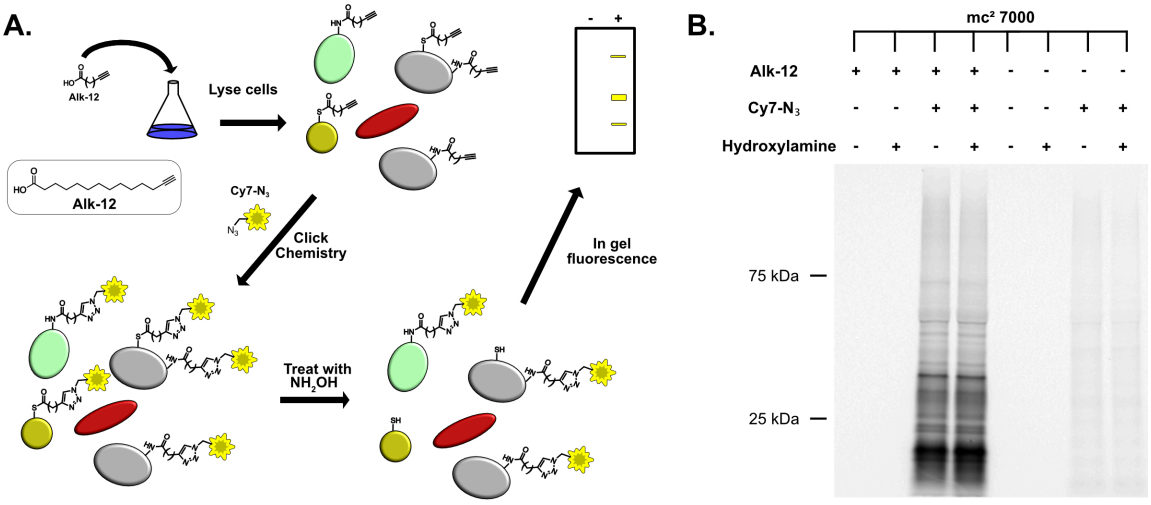


Figure 56. *Mtb* Proteins are Labeled with Alk-12

A. Method of using Alk-12 to detect fatty acylation in *Mtb*. B. Profiling of mc^{2}7000 with Alk-12.

vi. Inhibition Analysis of Mt-Sirt

We initially tested the ability of the known sirtuin inhibitors nicotinamide (NAM), adenosine 5′-diphosphoribose (ADPR), and β-nicotinamide adenine dinucleotide, reduced (NADH) to inhibit *Mt*-Sirt in the presence of saturating succinyllysine peptide (200 μM) and NAD^+ (1 mM) (Figure 58B, Table 3). We observed that NAM was a moderate inhibitor of *Mt*-Sirt, with an IC_{50} of 530 ± 200 μM, while ADPR exerted weaker activity of IC_{50} = 3.0 ± 0.7 mM and NADH resulted in a mere 20% reduction of activity at a concentration of 5 mM (Figure 58B, Table 3).

We next shifted to developing inhibitors for this enzyme that exploit its covalent catalytic mechanism. Previous work has identified that thioamide and thiourea-based inhibitors comprise specific and often potent mechanism-based inhibitors of sirtuins.[294] This inhibition occurs through stalling the formation of the 1',2'-cyclic intermediate between the 2' hydroxyl of the NAD ribose and the carbonyl oxygen of the peptide-derived amide, as has been observed *in crystallo*.[295, 296] Mechanistically, the exchange of the amide carbonyl oxygen to sulfur is proposed to result in non-ideal bond angles for the attack of the 2'-oxygen of the ribose ring into the 1'-S-alkylamidate intermediate when compared to the 1'-O-alkylamidate (Figure 57).[297] Further, the poorer electrophilicity of the 1'-S-alkylamidate also likely plays a role in this slowed transition.[297] To exploit this phenomenon, we prepared thiourea inhibitors with a fixed carboxybenzyl-lysine-aniline scaffold and varied thiourea motifs (**1**-**4**, Figure 58C) to elucidate which motifs are best tolerated and result in robust inhibition.

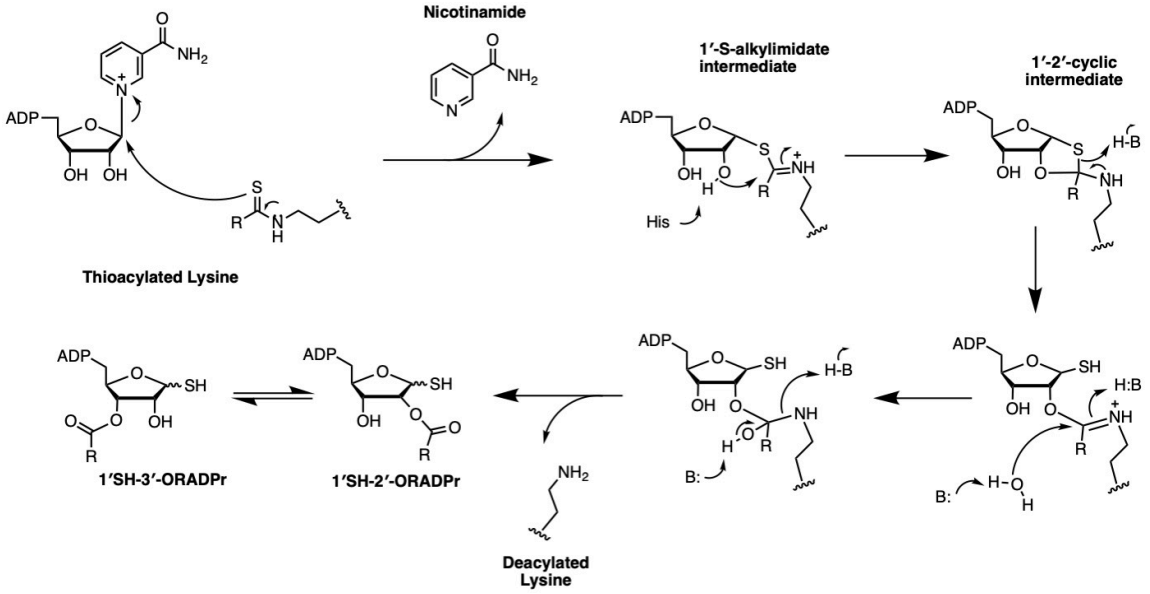


Figure 57. Proposed Mechanism of Thioamide/Thiourea Mechanism-Based Inhibition of Sirtuins

Table 3. IC_{50}s for Inhibition of *Mt*-Sirt with NAM, ADPR, NADH, and Mechanism-Based Inhibitors

Inhibitor	IC_{50} (μM)
NAM	530 ± 200
ADPR	3000 ± 715
NADH	>5000[a]
1	>100[b]
2	6.8 ± 1.2
3	>100[b]
4	184 ± 59
ACSK617 thiomyristoyl	4.6 ± 2.4

[a]IC_{50} value could not be well determined from the concentration-response curve (20% at 5 mM)

[b]We observed less than 10% reduction of activity at 100 μM inhibitor concentration.

These inhibitors were synthesized by preparing and coupling the appropriate *t*-butyl isothiocyanate,[298] with the ε-amine of Cbz-Lys-aniline, followed by acidic deprotection of the *t*-butyl group to yield the desired thiourea, or by converting the ε-amine of Cbz-Lys-aniline to an isothiocyanate, followed by reacting a primary amine of interest with the isothiocyanate to yield the desired thiourea.[299] Analysis of these molecules identified that thiourea mimics of succinyllysine and propionyllysine exhibited weak inhibition of *Mt*-Sirt (<10% reduction of activity at 100 μM). The addition of a methylene to the thiourea succinyllysine mimic, to make a thiourea mimic of glutaryl lysine, resulted in moderate inhibition with an IC_{50} of 184 ± 59 μM (Figure 58D). Consistent with our kinetic analysis showing fatty acyl lysine peptides had low K_m, the production of a thiourea mimic of myristoyllysine proved to provide a potent *Mt*-Sirt inhibitor with an IC_{50} of 7.0 ± 0.3 μM.

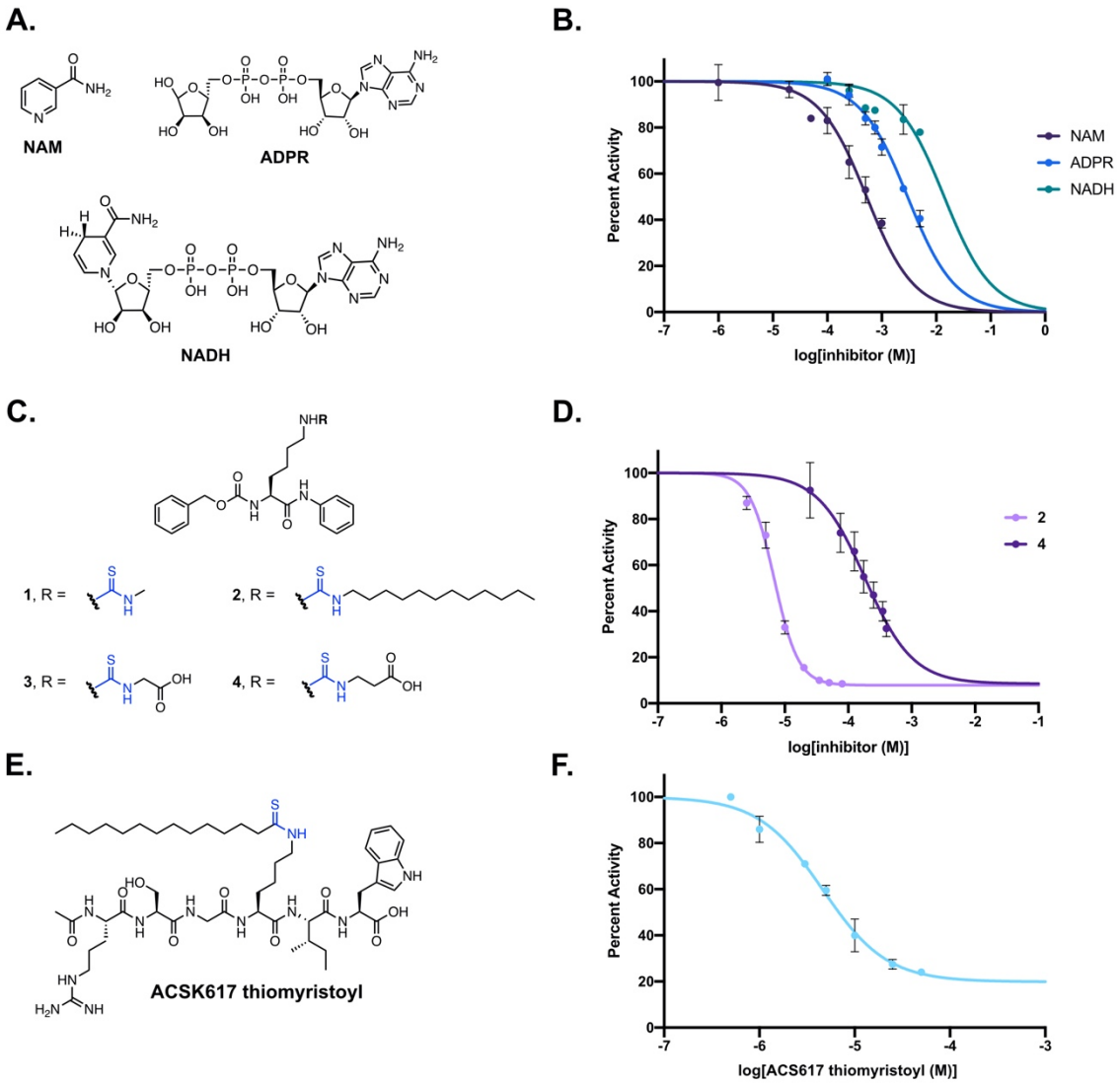


Figure 58. Inhibition of *Mt*-Sirt with NAM, ADPR, NADH, and Mechanism-Based Inhibitors.

A. Structures of NAM, ADPR, and NADH. **B**. Residual activity of *Mt*-Sirt observed in the presence of varied inhibitor with fixed NAD (1 mM) and peptide substrate (200 µM ACSK617 succinyl). **C**. Structures of thiourea containing mechanism-based inhibitors prepared. D. Residual activity of *Mt*-Sirt observed in the presence of varied inhibitor using the same conditions as **B**. **E**. Structure of thioamide containing peptide inhibitor of *Mt*-Sirt. **F**. Residual activity of *Mt*-Sirt observed in the presence of varied inhibitor using the same conditions as **B.**

Production of a hexapeptide containing thiomyristollysine yielded a nearly equipotent inhibitor of *Mt*-Sirt (IC_{50} = 4.6 ± 2.4 μM), which, assuming competitive inhibition, would yield a K_i of 1.3 ± 0.7 μM (Figure 58F). We do note that this class of inhibitors may exhibit time-dependent behavior, as has been previously displayed,[294] and thus these inhibition values only report on a "snapshot" of the inhibition at a 30-min incubation in the presence of substrate (2.4x K_m). Also, we note that the enzyme always exhibited some level of residual activity in the presence of the thioamide and thiourea containing inhibitors, even at concentration 10x higher than IC_{50}. This residual activity was greater with our thioamide containing peptide, consistent with recent findings that dethioacylation of the thioamide peptides occurs more robustly than previously thought.[300] While these findings observed that dethioacylation occurred with less efficiency as acyl chain length increased, we did indeed note thiodeacylation of our thiomyristoyl peptide when incubated with *Mt*-Sirt and NAD^+ in the absence of substrate (data not shown).

D. Conclusions and Future Outlooks

Through a combination of sequence alignment and homology modeling, we proposed that *Mt*-Sirt, the sole Sir2-like lysine deacylase of *Mtb*, may exert robust lysine desuccinyase activity. Synthesis and steady-state characterization of mono-lysine and peptide substrates containing varied N-acyl modifications with *Mt*-Sirt verified the enzyme was highly active against succinyllysine containing substrates, but also unexpectedly identified that it catalyzes efficient lysine fatty deacylation. Using a fatty acid metabolic alkyne probe, we identified that fatty acylation of *Mtb* proteins occurs in

cells, suggesting that the fatty deacylase activity of *Mt*-Sirt may play yet uncharacterized role in *Mtb*. Analysis of the known sirtuin inhibitors, nicotinamide and adenosine 5′-diphosphoribose, effected moderate inhibition of the enzyme, while reduced β-nicotinamide adenine dinucleotide was a poor inhibitor. Subsequently, we identified that mechanism-based thiourea and thioamide inhibitors mimicking myristoyllysine were potent against the enzyme (IC_{50} = 4.6-6.8 μM). Together, this comprises the first comprehensive characterization of the enzyme *Mt*-Sirt from *Mtb* and may serve to assist in better understanding the complex landscape of lysine post-translational modifications in this deadly pathogenic bacterium. Future experimental efforts will benefit from working to identify the key intracellular targets of *Mt*-Sirt and elucidating how this enzyme may regulate metabolic enzymes which undergo post-translational modification.

CHAPTER VI

A CLINICAL-STAGE PROTEASE INHIBITOR BLOCKS SARS-COV-2 INFECTION OF HUMAN AND MONKEY CELLS*

A. Introduction

The COVID-19 outbreak of 2019[301, 302] has led to a global health crisis of a magnitude not seen since the influenza pandemic of 1918. By March 2021, more than 118 million people have been infected worldwide, including more than 29 million in the US.[303, 304] Accordingly, there is an urgent need for the rapid identification of therapies that limit the pathology caused by SARS-CoV-2 for the management of COVID-19.[305, 306] The severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) outbreaks of 2003 and 2012,[307, 308] caused by betacoronaviruses highly related to SARS-CoV-2 provided important clues to the mechanism of viral infection of host cells. Two virally encoded cysteine proteases, 3CL protease (3CLpro)[55] and a papain-like protease (PLpro)[309] are essential to coronaviral maturation.

---

*This work has been published in and is reprinted in part with permission from: Mellott, D. M., Tseng, C.-T., Drelich, A., Fajtová, P., Chenna, B. C., Kostomiris, D. H., Hsu, J., Zhu, J., Taylor, Z. W., Kocurek, K. I., Tat, V., Katzfuss, A., Li, L., Giardini, M. A., Skinner, D., Hirata, K., Yoon, M. C., Beck, S., Carlin, A. F., Clark, A. E., Beretta, L., Maneval, D., Hook, V., Frueh, F., Hurst, B. L., Wang, H., Raushel, F. M., O'Donoghue, A. J., de Siqueira-Neto, J. L., Meek, T. D., and McKerrow, J. H. (2021) A Clinical-Stage Cysteine Protease Inhibitor blocks SARS-CoV-2 Infection of Human and Monkey Cells, *ACS Chemical Biology* 16(4), 642-650.

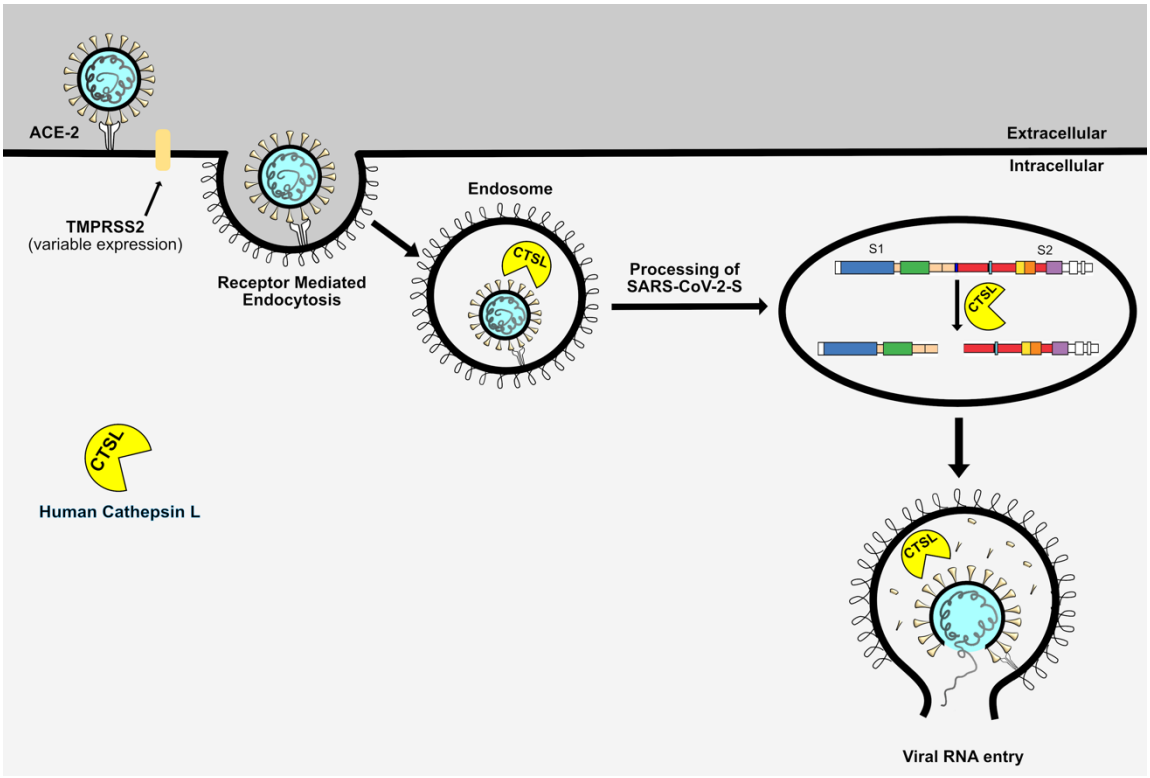


Figure 59. Cellular Entry of SARS-CoV-2.

Initial binding of a spike protein coating the viral capsid to the angiotensin converting enzyme 2 (ACE-2 receptor) results in attachment of the viral particle to the cellular surface. Subsequently the spike proteins of the capsid can undergo proteolysis via the transmembrane protease serine-2 (TMPRSS2) and furin-like proteases on the extracellular surface which can allow for efficient plasma membrane fusion into the intracellular space where the virus can replicate. Alternately, the viral particle receptor mediated endocytosis whereby the spike protein is proposed to undergo proteolytic processing by the cysteine protease cathepsin L (CTSL) at a site known as the S1/S2 site. Proteolysis at this site allows for the virus to efficiently leave the endocytic space and enter the intracellular milleu where it can replicate.

In addition, the trimeric coronaviral spike glycoprotein of SARS-CoV-2 (76% sequence identity to SARS-CoV-1 spike protein[310]), is involved in the fusion of the viral envelope with mammalian cell membranes, followed by the release of the genomic viral RNA and the nucleocapsid complex into the host cell. [92, 311] Studies of the SARS-CoV-1 spike protein reveal that it adheres to the extracellular domain of the angiotensin converting

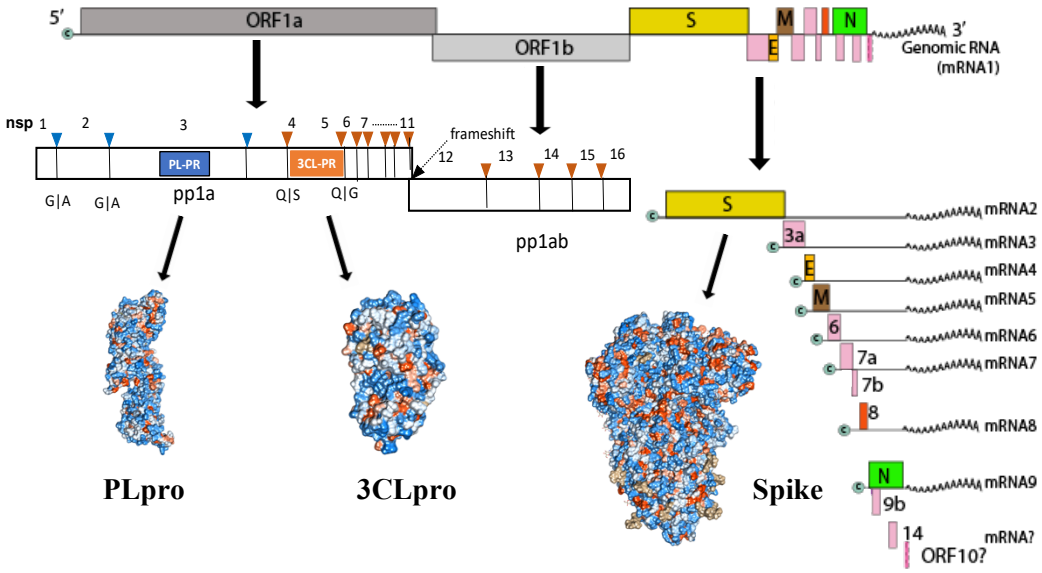


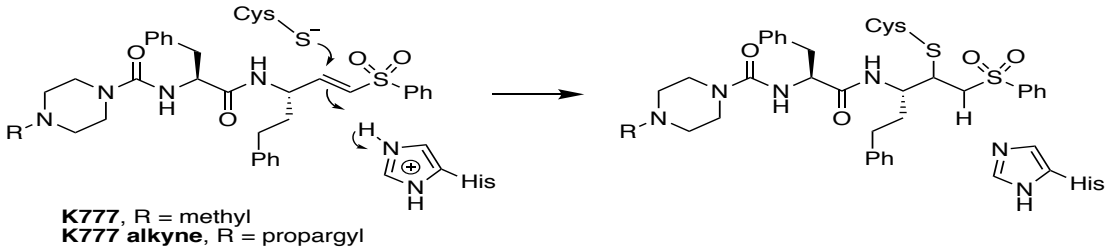
Figure 60. SARS-CoV-2 Genome and Key Proteins Produced

The SARS-CoV-2 genome is comprised of 12 canonical viral open reading frames (ORFs) which together comprise ~30,0000 nucleotides. The ORF1a and ORF1b contain the primary nucleotide content and are translated into two polyproteins (pp1a and pp1b). These polyproteins undergo excision from these polypeptide chains by the viral proteases 3CLpro and PLpro (cut sites shown as green and blue triangles, respectively) to produce 16 viral proteins. In addition, other proteins (spike, membrane, envelope, nucleocapsid, and accessory proteins) are produced independent of proteolytic processing.

enzyme-2 (ACE2) receptor in susceptible human cells.[310, 312, 313] In SARS-CoV-2, the receptor binding domain of the spike protein subunit S1 binds to the ACE2 receptor, and subunit S2 is involved in the viral cell membrane fusion to facilitate viral entry.[93, 314] Prior

to cellular entry, the spike protein requires processing by host cell proteases for activation (Figure 59).[93, 122, 313, 315] In particular, the transmembrane protease serine-2 (TMPRSS-2), the cysteine protease cathepsin L (CTSL), and furin-like proteases, have been proposed to play a role in preparing SARS-CoV-2 to enter cells by either endocytosis or traversing the cellular membrane.[93, 122, 314, 316, 317]

The proposed roles of host proteases in viral infection have stimulated studies to explore whether serine and/or cysteine protease inhibitors reduce coronaviral entry. Inhibitors of TMPRSS-2, such as camostat, reduced infectivity in selected human cell lines.[93] Broad spectrum cysteine protease inhibitors, such as E-64d, also inhibited coronaviral entry in certain cell types.[93, 316, 317] Riva *et. al* also reported in a drug repositioning screen that several cathepsin inhibitors have the ability to block SARS-CoV-2 entry into cells at sub-micromolar concentrations.[318]



K777, R = methyl
K777 alkyne, R = propargyl

Figure 61. Structure of K777 and its Alkyne Analog.

Chemical Mechanism of Irreversible Covalent Inactivation of Clan CA Cysteine Proteases

K777 (also known as, K11777, S-001, SLV213, 4-methyl-N-((S)-1-oxo-3-phenyl-1-(((S)-1-phenyl-5-(phenylsulfonyl)pentan-3-yl)amino)propan-2-yl)piperazine-1-carboxamide) is a highly potent, irreversible, covalent inhibitor of mammalian cathepsin L and other cysteine proteases of clan CA (Figure 61).[319, 320] K777 was originally developed as an

inhibitor of cathepsin S,[319] but later showed promise as an anti-parasitic agent.[320-326] In 2015, Zhou and colleagues reported that K777 blocked entry of pseudovirus forms of SARS-CoV-1 and MERS into Vero E6 or HEK293 cells, likely due to inactivation of cathepsin L (CTSL) on cell surfaces and/or within endosomes.[122] These results suggested that K777 might also be a potential therapeutic agent for the treatment of COVID-19. Here we show that K777 blocks CTSL-mediated processing of the spike protein, and blocks infection of SARS-CoV-2 in several mammalian cell lines. In addition, Selva Therapeutics has an open IND with the FDA for K777 as a possible therapeutic treatment of COVD-19 infection (communicated by Selva Therapeutics, Del Mar, CA).

B. Experimental Methods and Materials

i. Chemicals and Proteins

The synthesis of K777, its N-propargyl analog, are described elsewhere.[35, 321] Cyanine7-azide was obtained from Click Chemistry Tools. Recombinant proteases were obtained from the following vendors: human cathepsin L (Millipore Sigma, Athens Research and Technology, Inc., or R&D Systems, SARS-CoV-2 PLpro (Acro Biosystems, PAE-C5184), human cathepsin S (Millipore Sigma), bovine spleen cathepsin C (Millipore Sigma), human liver cathepsin B (Millipore Sigma), and human cathepsin K (Enzo Sciences, Inc.). Substrates were purchased from the following vendors: Z-FR-AMC (EMD Millipore), GF-AMC (MP Biomedicals), and Z-LR-AMC (Enzo Life Sciences, Inc.). Recombinant SARS-CoV-1 spike protein was obtained from SinoBiological and the SARS-CoV-2 spike protein was obtained from Genscript and Acro Biosystems.

ii. Cell Cultures, SARS-CoV-2 Infection, and Evaluation of Protease Inhibitors in a Coronaviral Infection Assay

All methods used to conduct these experiments can be found in Mellott *et al*.[138]

iii. Expression and Purification of 3CLpro

Based on the three-dimensional structure of SARS CoV-2 3CLpro (PDB 6Y84.pdb)[327], we designed expression constructs comprising nucleotides 10055-10972 of ORF1AB from GenBank (protein ID QHO62106.1, genome sequence MN988668.1, which encodes 306 amino acids.[328] This gene was *E. coli* optimized and ligated into the into the *Bam*HI and *Xho*I restriction sites of the pGEX-6p1 plasmid (Genscript, USA), resulting in the 3CLpro coding sequence being flanked by an N-terminal GST domain followed by a 3CLpro cleavage sequence (SAVLQ*SGF) and an C-terminal sequence containing a modified PreScission protease sequence (SGVTFQ*GP) preceding a His_6 sequence.[55, 329] Upon expression, auto-proteolysis from 3CLpro removed the N-terminal GST tag, yielding the authentic N-terminus (Ser-Gly-Phe). Purification of the processed protein using immobilized nickel affinity chromatography, followed by treatment with HRV 3C protease then generates 3CLpro with authentic N and C termini. This construct was transformed into BL21-DE3 *E. coli* cells for protein expression and a single colony of the transformed cells was used to inoculate a culture of LB broth containing 100 ug/mL ampicillin and was incubated at 37°C overnight. Subsequently, 1 L of LB media containing 100 ug/mL ampicillin was inoculated with the starter culture, and incubated at 37°C until reaching O.D. 600 of 0.6-0.8, at which time expression was induced by the

addition of 1 mM isopropyl β-thiogalactoside (IPTG). The cells were allowed to continue growth at 37°C for 4-5 h. and were then harvested by centrifugation (6,300 g at 4°C), and either stored at -80°C or lysed immediately for purification. Cells were suspended in 12 mM Tris-HCl, 120 mM NaCl, 0.1 mM EDTA, 2 mM DTT, pH: 7.5 (Buffer A). The cell slurry was then lysed using either a French press (25,000 psi) or by sonication. Lysates were centrifuged at 26,000g to remove cell debris, and clarified lysates were filtered with a 0.45 μm filter. The filtrates were loaded onto a HisTrap HP column (GE Healthcare), and washed with Buffer A. This was directly followed by elution using a linear gradient to 35% buffer B (12 mM Tris-HCl, 120 mM NaCl, 500 mM imidazole, 0.1 mM EDTA, 2 mM DTT, pH: 7.5) over 25 column volumes. The fractions containing pure 3CLpro-PreScision site-His_6 protein, as determined by SDS-PAGE, were pooled, and the protein concentration was determined using the presumed monomeric molecular weight of 33.9 kDa and an extinction coefficient of 32,800 M^{-1} cm^{-1}. 3CLpro was twice dialyzed against buffer A at 4°C. Proteolysis of the C-terminal H_6 tag of the auto processed translated gene product from the pGEX-6p1 plasmid (3CLpro-HRV 3C protease-His_6) was conducted by incubating 3.5 units of Pierce™ HRV 3C Protease (Thermo Fisher Scientific) per mg of 3CLpro overnight at 4°C overnight in buffer A. Subsequently, the protein mixture was successively loaded onto a 5 mL GSTrap HP column and a 5-mL HisTrap HP column (GE Healthcare), to remove the GST-fused HRV 3C protease and undigested H_6 tagged protein. The flow through was collected, analyzed by SDS-PAGE, and pure fractions of the tagless 3CLpro were pooled and concentrated (10 kDa molecular weight cutoff filter, GE Healthcare). The protein was deemed to be ≥95% pure by SDS-PAGE, and was stored

at – 80°C in 12 mM Tris-HCl, 120 mM NaCl, 0.1 mM EDTA, 2 mM DTT, (pH 7.5) with 50% glycerol (v/v). Analytical gel filtration using a Superdex 200 Increase 10/300 GL column (Buffer A at a flow rate of 0.7 mL/min), indicated that native 3CLpro was the expected homodimer.

iv. Kinetic Analysis of SARS-CoV-2 Proteases

Recombinant SARS-CoV-2 3CLpro (25-50 nM) was assayed at 25°C in 50-µL reaction mixtures containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 2 mM DTT, and 10% DMSO in 96-well plates (Greiner, flat-bottom half volume, clear black plates), using the FRET-based substrate Abz-SAVLQSGFRK(DNP)-NH2 wherein peptidolysis was measured at 320/420 nm (ex/em) (Biotek® Synergy M2) in the presence and absence of 0-50 µM K777 in duplicate. Recombinant SARS-CoV-2 PL was assayed in 50 mM HEPES pH 7.5, 5 mM DTT, 0.1 mg/mL BSA, 2% DMSO, 50 µM Z-RLRGG-AMC (Bachem, I1690), and 10 nM PLpro, where rates in the presence and absence of K777 were measured at 360/460 nm (ex/em).

v. Kinetic Analysis of K777 as an inhibitor of Clan CA Cysteine Proteases

Assays were conducted at 25°C in 50 mM sodium acetate (pH 5.5), 1 mM EDTA, 1 mM CHAPS, 10% (v/v) DMSO, and 5 mM DTT containing 28 ng/mL (~ 1 nM) cathepsin B , 1.1-1.9 ng/mL (~30-50 pM) cathepsin S, 8.3 ng/mL (~320 pM) cathepsin K, 25 ng/mL (~1 nM) cathepsin L, and 3.3 µg/mL (~132) nM in the presence of 0-0.1 mM K777 and 0.01 mM Z-FR-AMC for cathepsin S, cathepsin B, and cathepsin L, GF-AMC for

cathepsin C, and Z-LR-AMC for cathepsin K) 96-well black microplate (Corning Costar®). Rates of peptidolysis of the dipeptide-AMC substrate(s) were measured at either a SpectraMax M5 (Molecular Devices) in or a Synergy HTX (Biotek, Wisnooki, VT) microplate reader with λ_{ex} = 360 nm, λ_{em} = 460 nm in 8-60 sec intervals.

To determine the mechanism of inhibition, cathepsin L (250 ng/mL) or cathepsin S (92 ng/mL) was incubated with 10 nM K777, cathepsin B (490 ng/mL) or cathepsin K (97 ng/mL) with 200 nM K777, and cathepsin C (37.3 µg/mL) with 100 µM K777 for 1 h at room temperature, followed by 50-fold dilution into assay buffer and their residual activity was measured under nearly identical conditions to assay described above.

vi. Fitting of Kinetic Data

For cathepsins that were irreversibly inactivated by K777 the data was fit to:

$$Product = (\frac{V_o}{k})(1 - e^{-kobs*t}) \quad \textbf{(eq 3)}$$

where V_o is the initial rate, k_{obs} is the observed rate of inactivation, and t is time.

For cathepsins that were reversibly inactivated by K777 the data were fit to:

$$Product = V_s t + \left(\frac{V_o - V_s}{k}\right)(1 - e^{-kobs*t}) \quad (\textbf{eq. 21})$$

where V_0 and V_s are respectively the initial and steady-state rates, and k_{obs} is the rate constant of conversion of V_0 to V_s, and t is time.

The observed rates of inactivation were then replotted as a function of K777 concentration and fit to:

$$k_{obs} = k_{react} + \frac{k_{inact} * I}{(K_I\left(1+\frac{S}{K_M}\right)+I)} \quad \textbf{(eq 10)}$$

where k_{inact} is the intrinsic rate of inactivation, k_{react} is the rate of reactivation, I is the concentration of inhibitor: K777, K_I is the dissociation constant, S is the concentration of substrate, and K_M is the Michaelis constant for the substrate. K_M was determined for enzyme-substrate pairs by plotting observed catalytic rate versus substrate concentration and fitting the data to the Michaelis-Menten equation (data not published). For irreversible inhibitors, the rate of reactivation is essentially zero (i.e. $k_{react} = 0$).

When plots of k_{obs} vs. [I] were not hyperbolic (lack of saturation with I), the second-order rate constant was approximated from fitting the data to:

$$k_{obs} = \left(\frac{k_{inact}}{K_I}\right) I + k_{react} \quad (\textbf{eq. 11})$$

k_{react} was determined through fitting the results of the jump-dilution experiment to eq.20.

For cases in which inhibition or inactivation was readily reversible, the steady-state rate was plotted against concentration of K777 and fit to:

$$Rel. Activity = \frac{100}{1+\left(\frac{IC_{50}}{I}\right)^H} \quad \textbf{(eq. 22)}$$

where *H* is the Hill slope.

vii. In gel fluorescence of K777 alkyne treated Vero E6 cells Click Chemistry

K777 alkyne was added at concentrations of 1 μM to Vero E6 cells, with and without infection by SARS CoV-2 viral particles at an MoI of 0.1. For samples treated with both

K777 and the K777 alkyne, Vero E6 cells were first pre-treated with K777 (1 μM) for 1 h, prior to the addition of K777 alkyne (1 μM). Cell growth continued for 4 days, followed by harvest, and ~ 20 million cells were suspended in 1 mL of lysis buffer (200 mM Tris, 4% CHAPS, 1M NaCl, 8 M urea, pH 8.0) and protease inhibitor cocktail (Sigma Aldrich, #8340) at a ratio of 50:1 (v/v). Cells were lysed by sonication (ThermoFisher), protein concentration was quantified by use of the BCA reagent (Pierce). Labeling of the K777 alkyne treated cell lysates occurred in a stepwise fashion whereby addition of TBTA in 1:4 DMSO: n-butanol (v/v) (100 μM final) was added to $CuSO_4$ in water (1 mM final), followed by the addition of Cy7-azide (Click Chemistry Tools) in DMSO (20 μM final), 24 μg cell lysate, and finally sodium ascorbate in water (2.5 mM final). This reaction was incubated at room temperature for 1.5 h, and quenched by the addition ice-cold methanol (200 μL), chloroform (75 μL), and water (150 μL), followed by vortexing and centrifugation at 15,000 x g at 4 ºC for 15 min. The top layer of liquid was carefully removed and discarded, and the protein precipitate formed between two liquid phases was not disturbed. An additional 1 mL of ice-cold methanol was added into the tube, vortexed, and centrifuged at 15,000xg at 4 ºC for 20 min. The supernatant was discarded and this was repeated an additional time. Following removal of the supernatant the protein pellet was allowed to air-dry at room temperature. To the dried pellet 4% SDS buffer (4% SDS (w/v), 50 mM TEA, 150 mM NaCl, pH 7.4) was added, and the pellet was sonicated to achieve a clarified solution. Then, the solution was added to 4x loading dye (240 mM Tris-HCl, 8% SDS (w/v), 0.004 % bromophenol blue (to avoid fluorescence from the dye), DTT (5 mM final), and MilliQ water was added to a total volume of 24 μL. Samples were

heated to 95 ºC for 5 min, followed by a brief centrifugation at 5,000xg for 1 min. The ladder (Bio-Rad, Precision Plus Protein™ Dual Color Standards) was diluted 10,000 fold in 4% SDS buffer and 4x loading dye to avoid strong fluorescence from the ladder during imaging. Samples were analyzed on a 4-12% Bis-Tris Nu-PAGE gel (Invitrogen). To each well, 20 μg of labeled lysate was loaded, and the gel was run at 60 V for 30 min followed by 120 V for ~90 min. The gels were destained (10% acetic acid, 40% methanol, 50% water) for 1 h to overnight in the dark followed by imaging the gel on a ChemiDoc imager (Bio-Rad) using the Cy7 blot setting. To visualize the two fluorescent proteins of the protein standard ladder, we used the rhodamine setting. Densitometry of in-gel fluorescence was conducted using ImageJ. The gels were also analyzed by Coomassie Brilliant Blue R-250 staining (Bio-Rad).

viii. Mass Spectrometry for K777 Alkyne Enrichment Analysis

K777 alkyne[321] was added at concentrations of 0-10 μM to Vero E6 cells. Cell lysates were prepared as described above. Enrichment of alkyne-modified proteins was conducted using the Click-&-Go Protein Enrichment Kit (Click Chemistry Tools) and the protocols therein. The resin-bound proteins were reduced with DTT, alkylated with iodoacetamide, and digested on resin using 1 μg of Trypsin Gold (Promega) at 37°C overnight with agitation. The digested peptides were eluted from the column, desalted (Pierce™ C18 Tips, Thermo Fisher Scientific), concentrated to dryness, and suspended in 25 mM ammonium bicarbonate. NanoLC-MS/MS analysis was performed using an UltiMate 3000 HPLC system (Thermo Scientific) coupled to a Thermo Scientific Orbitrap Fusion™

Tribrid™ mass spectrometer. 1 µl of each sample was injected onto a PepMap100 C18 5 µm trap cartridge (0.3 × 5 mm) followed by an Acclaim PepMap C18 column (0.075 mm × 150 mm, particle size 3 µm, pore size 100 Å) at a flow rate of 30 µl/min. Peptides were eluted at a flow rate of 0.2 µl/min, using 98% water/2% acetonitrile with 0.1% formic acid (A), and 2% water/98% acetonitrile with 0.1% formic acid (B) as mobile phase. The gradient used was as follows: equilibration at 2% B for 5 min, ramping up to 45% at 37 min, 90% B at 40 to 46 min, ramping down to 2% B at 47 min, followed by re-equilibration at 2% B until the end of the run (60 min).

Following LC separation, samples were introduced into the mass spectrometer by nanoelectrospray ionization at a spray voltage of 2450 V, with the ion transfer tube temperature set to 275 °C. Data were acquired in top speed mode, with the cycle duration set to 3 sec. Full scan data were acquired in the Orbitrap in the 400-1600 m/z range at a resolution of 120,000 at m/z 200. MS/MS data were acquired in the ion trap in rapid scan mode, using an isolation window of 1.6 m/z with HCD at a fixed normalized collision energy of 28%. Dynamic exclusion was set to 60 sec.

Data were processed using MaxQuant 1.6.14. Label-free quantification was enabled at the default parameters. The green monkey (*Chlorocebus sabaeus*) proteome (UniProt ID UP000029965; 19,229 protein sequences) was used as the target database. MaxQuant result files were analyzed using Perseus 1.6.14. Reverse hits, potential contaminant hits, hits only identified by site, and hits for which only one peptide was detected were removed. Only proteins which were not detected in the control samples were considered. Intensity values were logarithmized. Finally, proteins which were

detected in at least two test samples were hierarchically clustered at default parameters (column clustering was disabled) and plotted as a heat map.

ix. Western Blotting

Cell lysates were prepared as stated above. Samples were separated via SDS-PAGE (Invitrogen) and semi-dry transferred (Bio-Rad) onto an activated PDVF membrane (GE Healthcare Life Sciences) followed by blocking with TBST/5% wt/vol non-fat milk (1 h, RT). Anti-actin antibody (Sigma Aldrich) or anti-Flag (Cell Signaling Technologies) was added at a 1:100,000 or 1:25,000 dilution respectively and incubated overnight at 4°C. The membrane was washed 5x with TBST, probed with goat-anti rabbit HRP conjugate antibody (1:3000) (Bio-Rad) in TBST/5% wt/vol non-fat milk (1 h, RT), washed (5x, TBST) followed by addition of ECL substrate (Bio-Rad) and imaged (ChemiDoc, Bio-Rad).

x. Statistical analysis.

All densitometry and gel quantification analysis was conducted using ImageJ.[330] For analysis of the statistical significance of the changes in pixel intensity we used an unpaired parametric t-test. P values ≤0.05 (*), ≤0.01 (**), ≤0.001 (***)

xi. Synthesis and Characterization of 3CLpro Peptide Substrate.

General synthetic considerations. The reagents and starting materials used were obtained from commercial vendors and used as received without any purification. Reactions were

carried in an inert atmosphere of nitrogen unless otherwise specified. Progress of the reactions were monitored using Thin Layer Chromatography (TLC) and LC-MS analysis, by employing an HPLC-MS (UltiMate 3000 equipped with a diode array coupled to a MSQ Plus Single Quadrupole Mass Spectrometer, ThermoFisher Scientific) using electrospray positive and negative ionization detectors. HPLC conditions used: column: Phenomenex Luna 5 μm C18(2) 100 Å, 4.6 mm, 50 mm, Mobile phase A: water with 0.1% formic acid (v/v). Mobile phase B: MeCN with 0.1% formic acid (v/v). Temperature: 25 °C. Gradient: 0−100% B over 6 min, then a 2 min hold at 100% B. Flow: 1 mL/min. Detection: MS and UV at 254, 280, 214, and 350 nm. $^{1}H/^{13}C$ NMR spectra were obtained in $CDCl_3$, CD_3OD, or DMSO-d_6 at 400MHz/100MHz at 298 K on a Bruker Avance III NanoBay console with an Ascend magnet. The following abbreviations were utilized to describe peak patterns when appropriate: br = broad, s = singlet, d = doublet, q = quartet, t = triplet, and m = multiplet. The final compounds used for testing in assays and biological studies had purities that were determined to be >95% as evaluated by their proton NMR spectra and/or their HPLC/MS traces based on ultraviolet detection at 254 nm or 350 nm. All reagents and starting materials were obtained from commercial suppliers, and used without further purification unless otherwise stated. Solution phase reactions were conducted under an atmosphere of nitrogen at ambient temperature unless otherwise noted. Reaction progress was monitored using thin-layer chromatography and by HPLC–MS (UltiMate 3000 equipped with a diode array coupled to an ISQ EM single quadrupole mass spectrometer, Thermo Fisher Scientific) using electrospray positive and negative ionization detectors. Reported liquid chromatography retention times (t_R) were established

using the following conditions: column: Phenomenex Luna 5 μm C18(2) 100 Å, 4.6 mm, 50 mm; mobile phase A: water with 0.1% formic acid (v/v); mobile phase B: MeCN with 0.1% formic acid (v/v); temperature: 25 °C; gradient: 0–100% B over 6 min, then a 2 min hold at 100% B; flow: 1 mL min^{-1}; and detection: MS and UV at 254, 280, 214, and 350 nm. Semi-preparative HPLC purification of compounds was performed on a Thermo Fisher Scientific UltiMate 3000 with a single wavelength detector coupled to a fraction collector. Purifications were conducted using the following conditions: column: Phenomenex Luna 5 μm C18(2) 100 Å, 21.2 mm, 250 mm; mobile phase A: water with 0.1% formic acid (v/v); mobile phase B: acetonitrile with 0.1% formic acid (v/v); temperature: room temperature.

Fmoc-Lys(DNP)-OH. A round bottom flask charged with Fmoc-Lys-OH (2 g, 5.43 mmol) was purged with N_2 gas to yield positive pressure, followed by the addition of 50 mL anhydrous dichloromethane. Subsequently, DIPEA (2.84 mL, 16.3 mmol) was added, followed by the dropwise addition of 1-fluoro-2,4-dinitrobenzene (700 μL, 5.54 mmol). The mixture was allowed to react for 3.5 h at which time the reaction mixture was washed with 1N HCl (1 time), water (3 times), brine (1 time), dried over Na_2SO_4, and concentrated under reduced pressure. The crude product was subjected to flash purification on silica (2% MeOH:DCM), which following concentration under reduced pressure yielded a bright yellow fluffy powder (1.89 g, 65% yield). LC-MS t_R: 5.99 min, m/z 535.12 [M+1H], $C_{27}H_{26}N_4O_8$ Calcd. 535.18 [M+1H].

Boc-2-Abz-OH. A round bottom flask charged with 2-aminobenzoic acid (2 g, 14.58 mmoles) was dissolved in 20 mL of water and the pH was adjusted to 8 by adding 10 N NaOH dropwise. Subsequently, di-tert-butyl dicarbonate (3.5 g, 16.04 mmol) was dissolved in 20 mL of anhydrous THF was added dropwise to the reaction which was allowed to proceed overnight. The organic layer was removed under reduced pressure, and the aqueous layer was acidified with 2N HCL and extracted with ethyl acetate 3 times. The organic layer was washed with water (2 times), and brine (2 times), and dried over anhydrous sodium sulfate, and removed under reduced pressure yielding an off-white powder. (2.76 g, 81% yield) LC-MS t_R: 5.36 min, m/z 236.08 [M-1H], $C_{12}H_{15}NO_4$ Calcd. 236.10 [M-1H].

Fmoc-SAVLQSGFRK(DNP)-NH_2 . To a syringe fitted with a frit, 200 umol of Rink Amide AM resin (Novabiochem, 0.73 mmol/g) was washed with dichloromethane (DCM), and swelled in N,N-dimethylformamide (DMF). The Fmoc resin was deprotected using 20% piperidine in DMF (v/v) for 15 min (3 times), followed by five washes with DMF. Coupling of Fmoc-Lys(DNP)-OH was conducted using 3-fold excess of the amino acid versus resin loading (0.6 mmol), (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate (COMU) (0.6 mmol), and DIPEA (1.2 mmol) under agitation for 3 h (2 times). Washing of the resin with DMF, followed by ninhydrin analysis verified the coupling was complete. The resin was then capped by 45 min of agitation with 25% acetic anhydride in DMF (v/v) and DIEA (0.3 mmol), followed by DMF washes (5 times). Subsequent synthesis of the

peptide occurred in a stepwise fashion: deprotection of the Fmoc protecting group by 20% piperidine in DMF (v/v) for 15 min (2 times), DMF wash (5 times), ninhydrin analysis, addition of Fmoc-AA-OH (0.6 mmol), COMU (0.6 mmol), and DIEA (1.2 mmol) (allowed to pre-activate for 2 min prior to addition to resin) and coupling for 1 h, washing of the resin with DMF (5 times), ninhydrin analysis, capping of unreacted amine groups with 25% acetic anhydride in DMF (v/v) and DIPEA (0.3 mmol) for 15 min, and washing of the resin with DMF (5 times). All steps were carried out under constant agitation. Test cleavage of the peptide was conducted using ~5-10 mg of resin suspended in hexafluoroisopropanol (HFIP) containing 0.5 N HCl (aq), 5% v/v water, and 2.5% v/v triisopropylsilane for 2 h. The solution was then removed under a stream of N_2 gas, and the resulting material was dissolved in acetonitrile, filtered, and analyzed by LC-MS, showing that the peptide was the major product (m/z 740.4 [M+2H]; 1479.7 [M+1H] or 740.4 [M+2H] calcd.). The peptide-laden resin was then washed with DCM (3 times), dried under vacuum, and stored in a desiccator for further synthesis.

Abz-SAVLQSGFRK(DNP)-NH_2. To a syringe fitted with a frit, half of dry resin bound Fmoc-(DNP)-Rink Amide peptide was suspended in DMF for 20 min. Treatment with 20% piperidine in DMF (v/v) for 15 min (3 times), followed by DMF washes (5 times) and a ninhydrin test confirmed removal of the Fmoc group. To the free amine, Boc-2-Abz-OH (0.3 mmol), COMU (0.3 mmol), and DIPEA (0.6 mmol) (allowed to pre-activate for 2 min) was added and allowed to couple overnight. The resin was washed with DMF (5 times), and the coupling was repeated for another 4 h to enhance the yield. The resin was

then washed with DMF (5 times), DCM (3 times), and MeOH (3 times). Subsequent cleavage of the peptide with HFIP containing 0.5 N HCl (aq), 5% v/v water, and 2.5% v/v triisopropylsilane for 4 h (2 times) was followed by removal of the solution via rotary evaporation, precipitation of the peptide in ice cold diethyl ether. The crude mixture was solubilized in a small volume of DMF, and subjected to semi-preparative HPLC (5%-52% B over 20 min, then 52-100% B to 26 min, and 100% B until 29 min at 21.2 ml/min). The fractions containing pure peptide (≥95% pure) were then lyophilized to yield a fluffy yellow powder (42.4 mg, 29% yield as a formic acid salt) LC-MS t_R: 5.36 min, m/z 688.89 [M+2H], 459.79 [M+3H], $C_{61}H_{89}N_{19}O_{18}$ Calcd 1376.66 [M+1H], 688.83 [M+2H], 459.56 [M+3H].

C. Results and Discussion

i. K777 Blocks SARS-CoV-2 Infectivity of Host Cells

Table 4 shows EC_{50} (or EC_{90} for Caco-2 cells) values for inhibition of viral invasion of host cells by K777, usually assessed by reduction of the cytopathic effect (CPE). The EC_{50} is defined as the concentration of K777 that reduces viral infectivity by 50% whereas the EC_{90} reports on a 90% reduction. For Caco-2 cells, virus yield reduction was used, as CPE was not observed. EC_{50} values ranged from 4 nM for HeLa/ACE2 cells, 7 nM for Calu-3/2B4 cells, ≥70 nM for Vero cells, <80 nM for A549/ACE2 cells, and >10 micromolar for Calu-3 (ATCC) cells. Caco-2 cells showed an EC_{90} of 4.3 μM. No host cell toxicity was seen for any cell line at inhibitor concentrations of 10-100 μM. The differences between Calu-3 and Calu-3 2B4 cells is especially notable, and likely is the result of enrichment of the ACE2 receptor in the latter cells.[331] Figure 62 shows an

example of dilution curves used to calculate EC_{50} values. For comparison, the EC_{90} of the known anti-viral remdesivir is 0.10 μM in the Caco-2 assay (data not shown)

Table 4. Anti-Coronaviral Activity of K777 in Infected Primate and Human Cells[1]

Cell Line	Cell Type	Coronavirus	Laboratory	Inhibition of Viral Infectivity EC_{50} or LC*** (μM)	Cytotoxicity to Host Cells CC_{50} (μM)
Vero E6	African Green Monkey / kidney epithelial	SARS CoV-1*	UTMB	2.5***	>10
Vero E6		MERS*	UTMB	0.62***	>10
Vero E6		SARS CoV-2*	UTMB	0.62***	>10
Vero E6		SARS CoV-2	USU	<0.1	>10
Vero E6		SARS CoV-2	UCSD	0.07	>10
Caco-2	Human colon	SARS-CoV-2	USU	4.3	>10
A549/ACE-2	Human lung carcinoma expressing ACE-2 receptor	SARS CoV-1**	UTMB	0.08***	>10
A549/ACE-2		MERS**		0.08	>10
A549/ACE-2		SARS CoV-2**		<0.08	>10
HeLa/ACE-2	Human cancer cells expressing ACE-2 receptor	SARS CoV-2	UCSD	0.004	>10
Calu-3/2B4	Human lung epithelial cells	SARS CoV-2	UTMB	1***	>10
Calu-3	Human lung epithelial cells	SARS CoV-2	UCSD	3.7***	>10

[1]Infected with 100* or 500** viral particles/sample; Values of EC_{50} were obtained from multiple concentrations of K777 at which CPE is observed or is not present. ***LC is the lowest concentration of K777 at which CPE is not observed. The laboratory sites in which these studies were conducted are: UTMB, University of Texas, Medical Branch; USU, Utah State University; UCSD, University of California, San Diego.

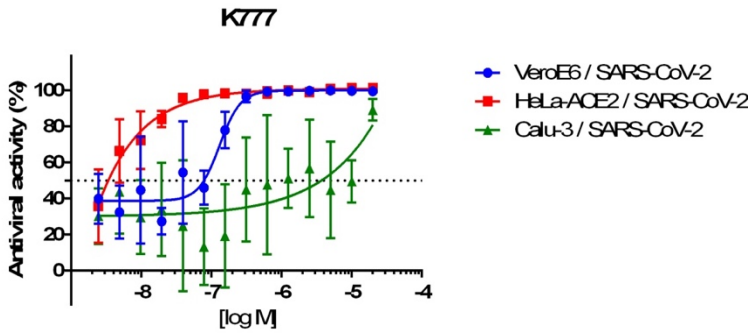


Figure 62. Dose Dilution Curves for K777 Effect on Virus Infection of Three Cell Lines

ii. Analysis of K777 versus SARS-CoV-2 and Clan CA Cysteine Proteases

In assays carried out at both UC San Diego and Texas A&M University, there was no inhibition of either the PL and 3CL proteases of SARS-CoV-2 (25-100 nM enzyme concentration) at up to 100 μM K777 (Table 5; Figure 63). Conversely, K777 exhibited potent activity against human cathepsin L, where the second-order rate constant of inactivation, k_{inact}/K_I, was found to be 50-fold and 333-fold greater than cathepsin K and cathepsin B, respectively (Table 5). K777 formed apparent irreversible covalent adducts with these enzymes, but exhibited weak, slowly reversible, inhibition of cathepsin C, and potent, but rapidly reversible inhibition of cathepsin S.

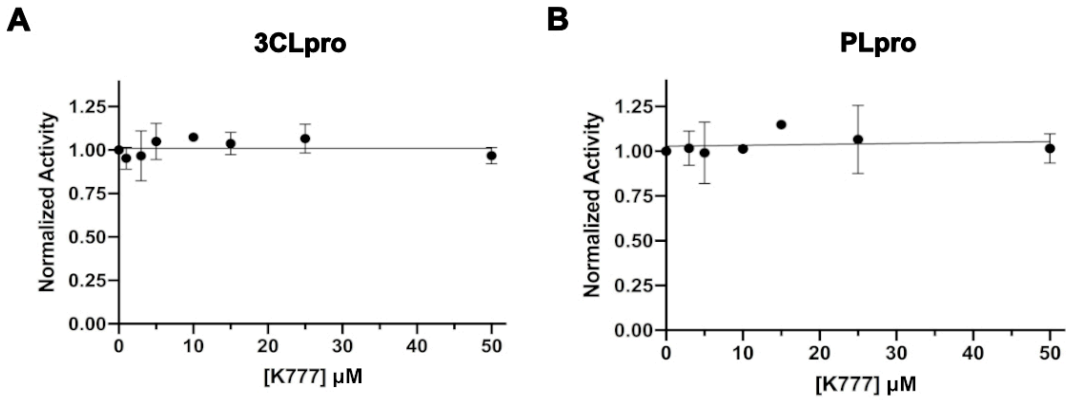


Figure 63. K777 Does Not Inhibit SARS-CoV-2 3CLpro or PLpro.

Average initial rates determined upon incubation of 25 nM 3CLpro in the presence of 50 μM of Abz-Ser-Ala-Val-Leu-Gln*Ser-Gly-Phe-Arg-Lys-(DNP)-NH_2 with varied concentrations of K777 (1-50 μM) at 25°C for 30 min. **C.** Average initial rates determined upon incubation of 10 nM PLpro in the presence of the substrate Arg-Leu-Arg-Gly-Gly-AMC with varied concentrations of K777 (1-50 μM) at 25°C for 30 min. Data was fit to a straight line.

Table 5. Inactivation or Inhibition of Mammalian and CoV-2 Cysteine Proteases

Enzyme	K_I (nM)	k_{inact} (s^{-1})	k_{inact}/K_I (M^{-1} s^{-1})
Human Cathepsin L	50 ± 20	0.013 ± 0.002	(3 ± 1) x 10^6
Human Cathepsin K	400 ± 100	0.022 ± 0.004	(6 ± 2) x10^4
Human Cathepsin B	3,000 ± 1000	0.026 ± 0.007	(9 ± 4) x10^3
Bovine Cathepsin C	>100,000	ND	(1.2 ± 0.4) x10^{1} †
Human Cathepsin S	2.0 ± 0.1‡	ND	ND
SARS-CoV-2 3CLpro	>100,000	ND	ND
SARS-CoV-2 PL	>100,000	ND	ND

[a] pH = 5.5 for cathepsins and pH 7.0 for CoV-2 proteases; ND, not determinable.

†Value approximated from slope of k_{obs} vs inhibitor.

‡Value represents overall dissociation constant.

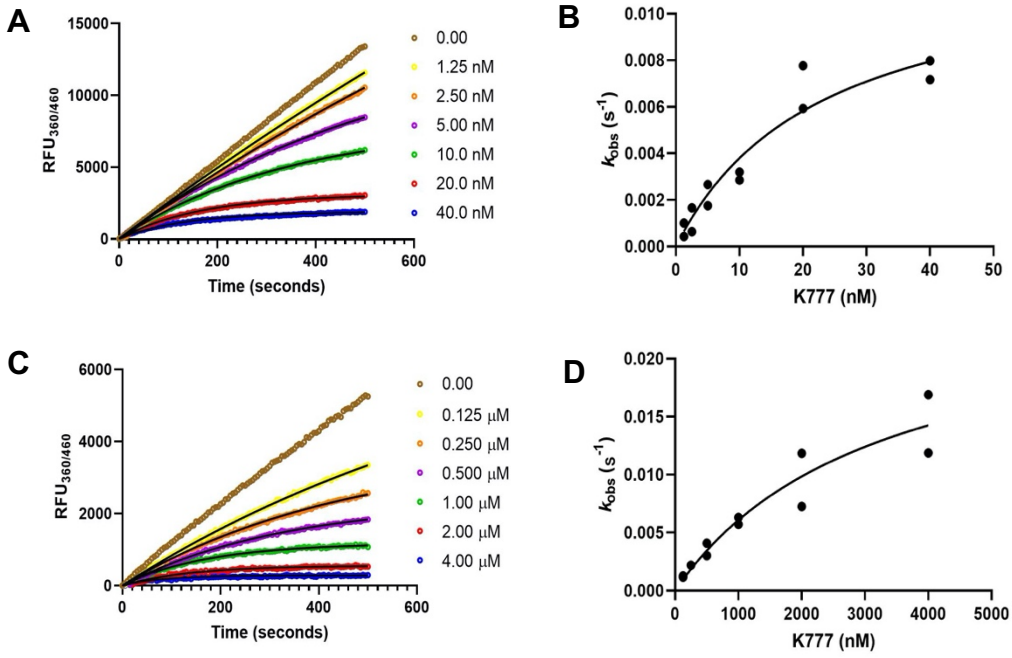


Figure 64. Cathepsin L and B Inactivation by K777 and Replot of Inactivation Rates Versus [K777].

Cathepsin activity monitored by hydrolysis of dipeptide-AMC substrates. Shown are representative plots of inactivation where data was fit to eq. 3: **A, C**. Representative plot of Cathepsin L (**A**) and B (**C**) inactivation with Z-F-R-AMC. **B,D**. Replot of k_{obs} values from A (**B**) and C (**D**) versus concentration of K777, fit to eq. 10 .

iii. Activity-Based Protein Profiling of K777 Protein Targets in Vero E6 Cells.

We prepared a K777 analog, K777 alkyne[321]; (Figure 61) to elucidate the cellular target(s) of K777. This activity-based protein probe has been used previously to identify the cathepsin-L-like and cathepsin-B-like proteases found in the parasitic protozoan, *Trypanosoma brucei.*[321] The K777 alkyne probe facilitates copper-mediated cycloaddition chemistry between an azide and alkyne to both enrich target proteins for mass spectrometry analysis, and for visualization of target proteins via in-gel fluorescence. In both virally and non-virally infected Vero E6 cells, the protein(s) bound to the K777-Cy7 conjugate had a molecular weight of approximately 25 kDa (Figure 65A), a molecular weight that could correspond to either cathepsin B (mature form: 25 kDa) or cathepsin L (mature form: 24.2 kDa). To confirm that the specificity of the K777 alkyne was analogous to that of K777, cells were also preincubated with 1 μM K777 prior to treatment with the K777 alkyne probe.

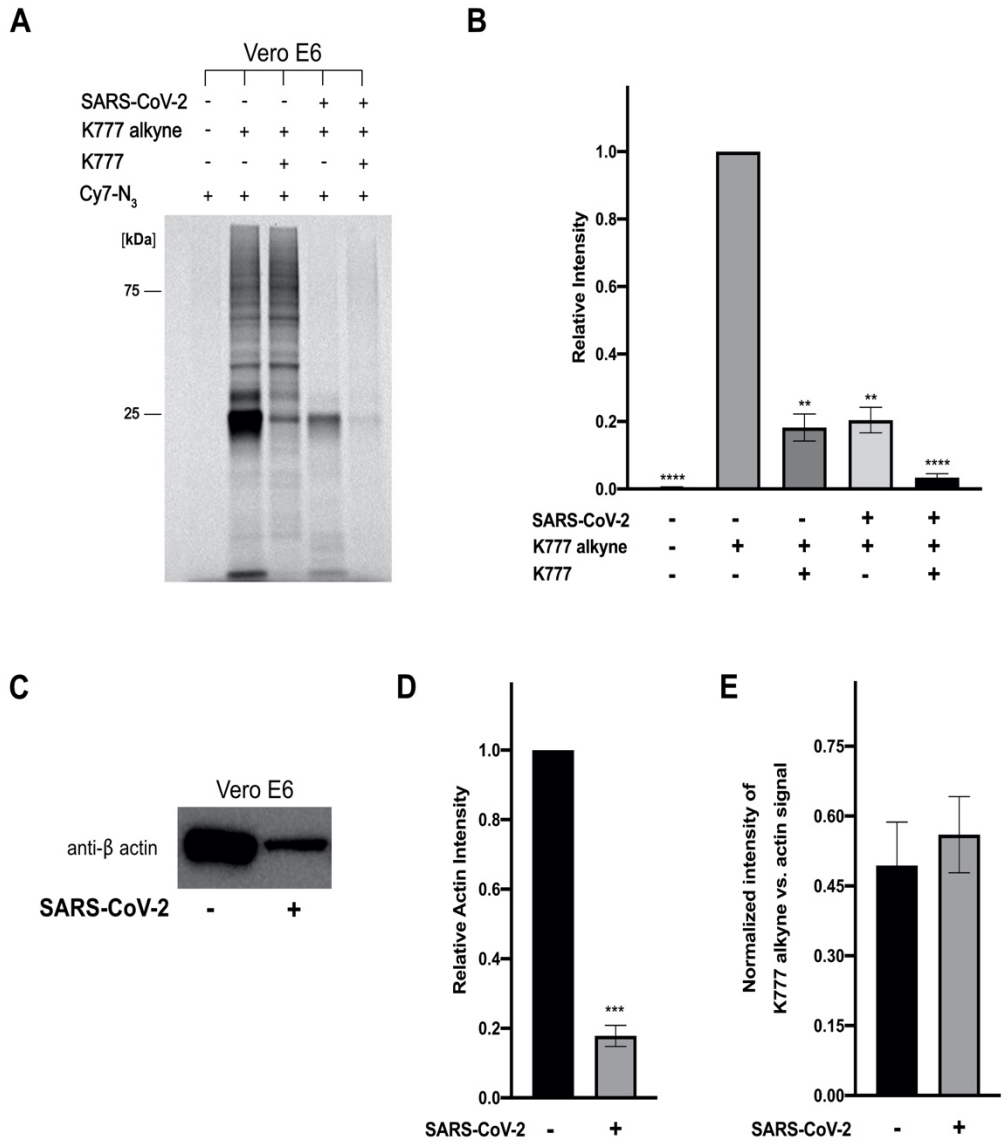


Figure 65. K777 Alkyne Specifically Targets a Non-viral Protein in SARS-CoV-2 Infected and Uninfected Vero E6 Cells

A. Cy7 azide labeled proteins (SARS-CoV-2 infected and uninfected) at 1 μM K777 alkyne is blocked by pre-treatment with 1 μM K777. **B.** Densitometry analysis of **A,** and a replicate in gel fluorescence experiment. **C**. Blotting of total β-actin levels in SARS-CoV-2 infected and non-infected cells. **D.** Densitometry analysis of relative actin levels. **E.** Comparison of the signal intensity of the 25 kDa band enriched in the presence of K777 alkyne (SARS-CoV-2 infected and uninfected) versus the β-actin signal (SARS-CoV-2 infected and uninfected).

This resulted in >80% blockage of the 25-kDa target protein signal in both virally and non-virally infected cells (Figure 65B). As the other minor protein bands visualized were unchanged upon K777 pre-treatment, it is concluded that these proteins are not targets of K777.

When comparing the intensity of the 25-kDa band in SARS-CoV-2-infected cells to that of its uninfected counterpart there was nearly a 5-fold decrease in intensity (Figure 65A-B). Reduction of the signal for this band in the virally infected samples may be due to either a reduction of total protein or differential expression of the target(s) of the K777 alkyne. To investigate this, we immunoblotted the housekeeping protein human β-actin to compare its relative amounts in both sets of samples. Upon quantifying the difference in β-actin content between the SARS-CoV-2-infected and uninfected cells, we measured a 5.7-fold decrease in total β-actin signal in the Vero E6 SARS-CoV-2 infected samples (Figure 65C-D). Comparison of the relative intensities of the β-actin band versus the 25-kDa enriched protein for both SARS-CoV-2 infected and uninfected cells approximated a 1:1 ratio (Figure 65E). These results suggested that coronaviral infection of Vero E6 causes a global reduction in protein expression, in agreement with recent reports that non-structural protein-1 of SARS-CoV-2 suppresses protein translation in infected cells.[332]

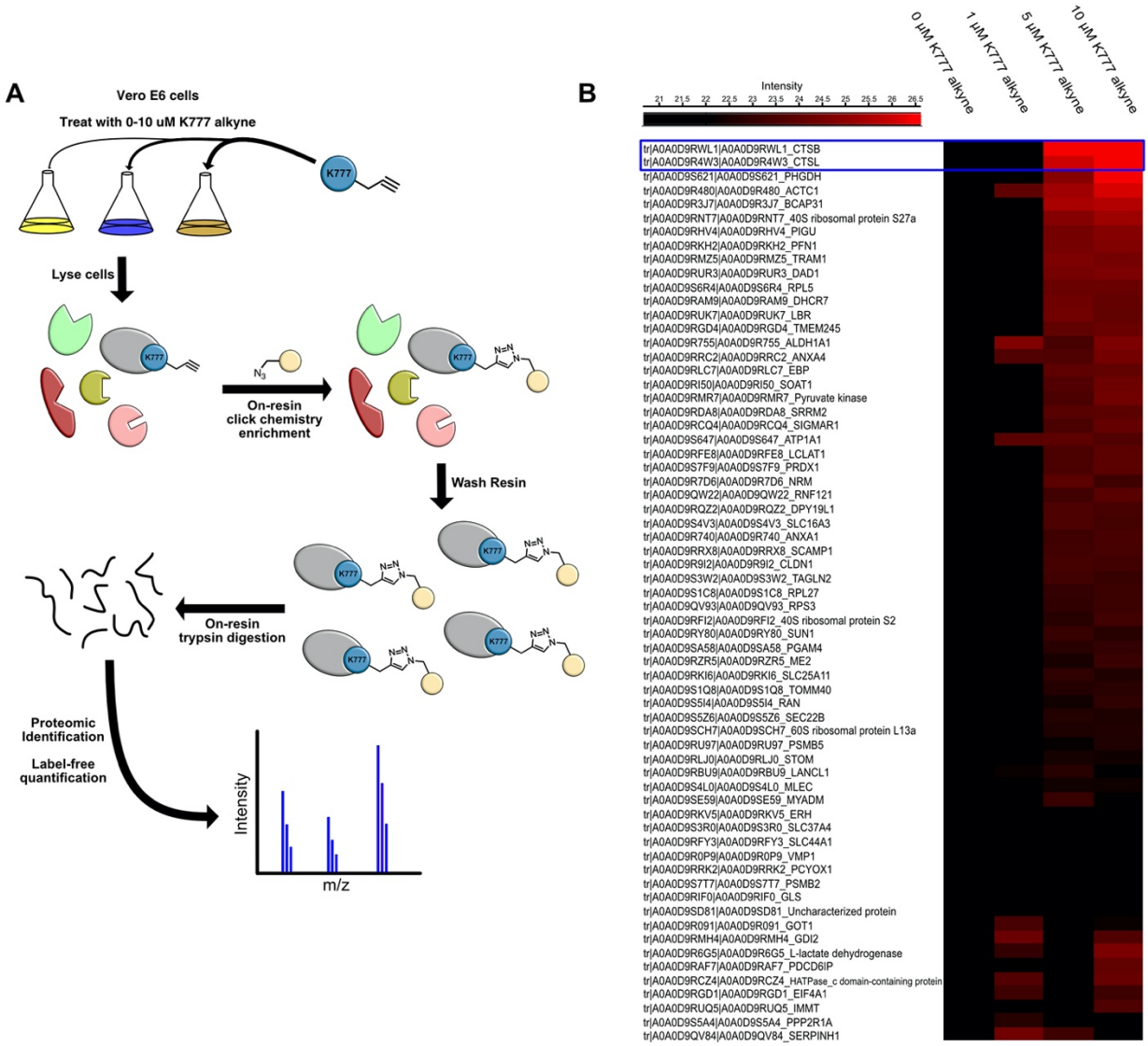


Figure 66. Cathepsins B and L are Strongly Enriched Upon K777 Alkyne Treatment

A. Schematic of enrichment strategy employed to prepare samples for proteomic processing. **B**. Heat map of proteomic data clustered from highest to lowest intensity. Intensity reports on the summation of the peptide intensities within a particular protein group.

iv. Identification of Cathepsin B and L as K777 Targets by Affinity Enrichment and Proteomic Analysis.

To identify the molecular target(s) of the K777 alkyne, affinity enrichment and proteomic analysis of treated Vero E6 cell lysates was carried out with variable concentrations of the K777 alkyne (0-10 μM) (Figure 66A). 117 proteins at a false discovery rate of 1% were identified, of which 65 were present in the K777 alkyne-treated samples, but not the untreated control (Figure 66B). Proteins of the highest abundance in the treated (5 and 10 μM samples), as determined by MaxQuant LFQ analysis,[333] were cathepsin B (CTSB) and cathepsin L (CTSL).

v. Cathepsin L but not Cathepsin B Catalyzes Cleavage of the SARS-CoV-2 Spike Protein.

Having identified CTSB and CTSL as putative targets of K777 in cells, we sought to elucidate the ability of these proteins to process the SARS-CoV-1 and SARS-CoV-2 spike (S) proteins *in vitro*. Previous studies of CTSL proteolysis of the SARS-CoV-1 S suggested that CTSL functionally cleaves the S protein proximal to the S1 cleavage site (R_{667}*S_{668}) at T_{678}*M_{679}, which produces two polypeptides of ~90 and ~70 kDa corresponding to the S1 and S2 domains, respectively.[317] This, coupled with extensive biochemical and inhibitor-based studies, indicated that CTSL likely plays a role in viral entry of SARS-CoV-1 via proteolytic processing of the S protein.[312, 316, 317] These prior studies, and recent studies in SARS-CoV-2, have failed to rule out the role of CTSB in the proteolytic processing of the CoV-1 and CoV-2 S proteins,[312, 316, 334] as there is conflicting evidence as to if the inhibitors used (CA-074 and CA-074-Me) can enter cells or exhibit

selectivity towards CTSB.[89, 335] Recently, Jaimes and colleagues analyzed the cleavage of an 11 amino-acid FRET peptide corresponding to the unique S1/S2 multibasic site of CoV-2 S protein (Mca-TNAPRRAR*SVA-Dnp), and determined that furin, trypsin, PC1, matriptase, CTSB, and CTSL can catalyze cleavage of this peptide.[336] With regard to cathepsins B and L, this is unsurprising given their proclivities for arginine residues in the P_1 sidechains of substrates.[139] In particular, this substrate should be processed by CTSB, as this protease tolerates an alanine in the P_2 position of its substrates.[337] Conversely, this substrate should be less effective as a substrate of CTSL as this protease generally recognizes substrates with a hydrophobic amino acid in its P_2 position.[139, 140] Therefore, to determine if these findings held true with SARS-CoV-2 S *in vitro*, we assessed the ability of trypsin, CTSB, and CTSL to catalyze proteolysis of both recombinant, insect-cell expressed SARS-CoV-2 S protein containing the intact S1, S2, and extracellular domains, and a stable trimer of SARS-2-S expressed in HEK-293 cells containing R683A and R685A mutations at the putative cleavage site of furin. As expected, trypsin (pH 7.5) was able to catalyze cleavage of both proteins efficiently, producing two protein products corresponding to the S1 and S2 domains, as previously reported. (Figure 67A,B). Recent viral infectivity studies have shown that a mutant of the S1/S2 site retaining a singular monobasic site (R_{682}) is still able to, albeit at lower efficiency, enter cells likely using both the trypsin-like TMPRSS-2 and cathepsins.[338] Treatment of SARS-CoV-2 S protein with trypsin and CTSL produced different protein fragments (Figure 68A), demonstrating that the two proteases recognize different substrate sites in the CoV-2 S proteins tested.

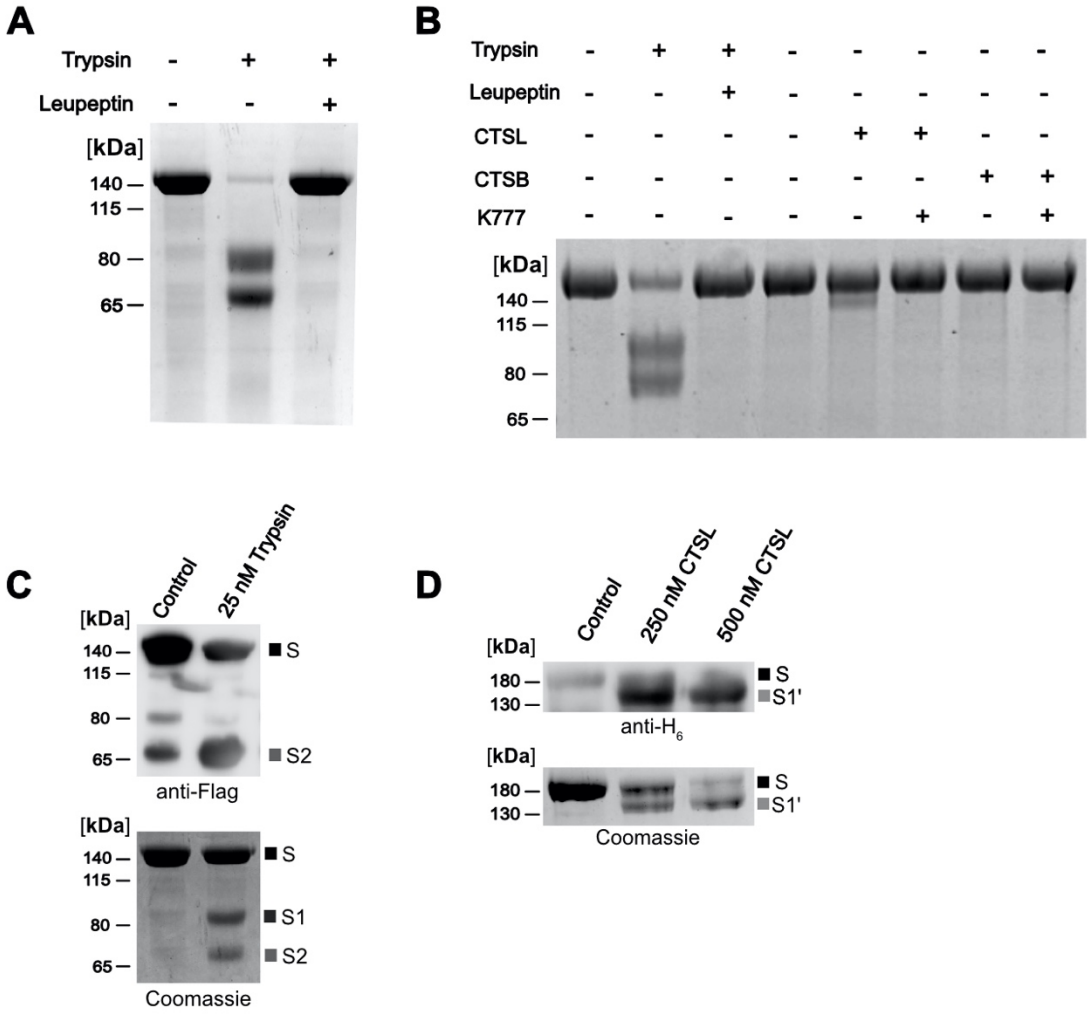


Figure 67. Processing of SARS-CoV-2 Spike Protein by Trypsin, CSTB, and CTSL.

A. Trypsin proteolysis of SARS-CoV-2 S produced ~80 kDa and ~65 kDa fragments, which can be blocked by leupeptin. **B.** Analysis of processing of the trimeric SARS-CoV-2 S protein by trypsin (25 nM), CTSB (250 nM), and CTSL (250 nM) in the presence and absence of inhibitors. **C.** Representative immunoblotting of trypsin treated SARS-CoV-2 S demonstrates cleavage at the S1/S2 site, which results in signal for the full length S protein and a ~65 kDa fragment corresponding to the cleaved S2 domain. **D.** Immunoblotting of **B** demonstrates that the S1' product produced contains a C-terminal H_6 tag, suggesting that CTSL processing occurred in the S1 domain of the protein.

In addition, no proteolysis was observed with human CTSB for either CoV-2 S protein, highlighting that the S1/S2 FRET peptides did not recapitulate the proteolysis seen with recombinant CoV-2 S protein (Figure 68A, Figure 67A,B). In the presence of CTSL (at pH 5.5, mimicking the pH of the endosome), we observed a prominent, seemingly singular, ~120 kDa band for the monomer (Figure 68B, C) and ~140 kDa for the trimer (Figure 67B); more extensive glycosylation in the latter accounts for its slightly higher apparent molecular weight. The formation of a faint ~40 kDa band and indistinct protein bands at less than ~40 kDa, suggest that more extensive proteolytic degradation of the other protein product had occurred. As processing by both CTSL and trypsin resulted in apparently identical products formed from the trimer and monomer CoV-2 S proteins, we used the monomer throughout the remainder of the study.

We next directly compared the processing of the SARS-CoV-2 S to that of CoV-1 S. Treatment of the CoV-1 and CoV-2 S proteins with CTSL produced different proteolytic fragments (Figure 68A), revealing that CTSL catalyzes cleavage at different sites in the two proteins. With CoV-1 S protein, proteolysis catalyzed by CTSL produced three unique bands, one at ~105 kDa, another at ~95 kDa, and a third poorly defined band at ~55 kDa. Interestingly, CSTB-catalyzed proteolysis produced a band seemingly analogous to that resulting from treatment with CTSL, as well as a poorly defined band at ~60 kDa. To identify which domain of the CoV-2 S protein was proteolyzed by CTSL, we immunoblotted against the CTSL proteolyzed CoV-2-S protein containing a C-terminal Flag epitope. We observed immuno-reactivity with both the intact S protein as well as the ~120 kDa protein fragment (Figure 68C), indicating that this ~120 kDa protein contains

an intact S2 domain, and that CTSL catalyzes proteolysis in the S1 domain of intact CoV-2 S at an unknown location, which we coin the S1' site). When comparing of the S1 and S2 domains of CoV-1 S versus CoV-2-S proteins revealed that there is only a 64% sequence identity between the S1 domain of these two proteins.[339]

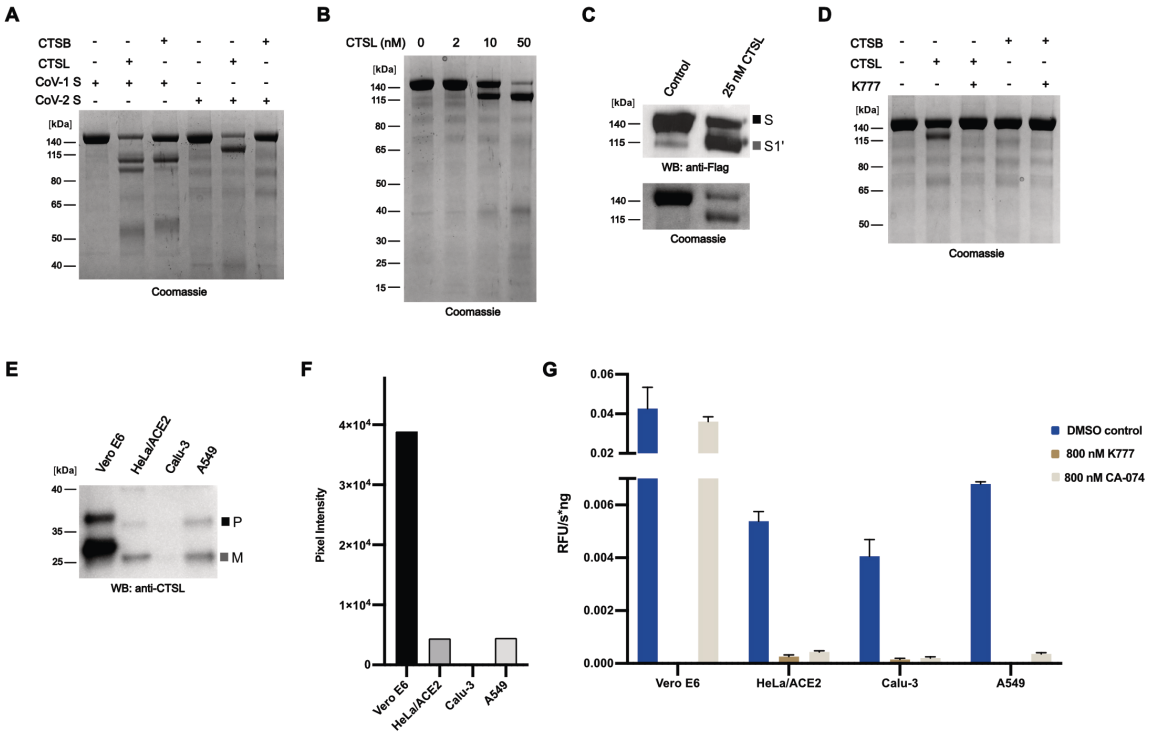


Figure 68. CTSL and CTSB Catalyze Differential Processing of the SARS-CoV-2 Spike Protein and CTSL Levels Vary in Cells.

A. Processing of SARS-CoV-1 spike protein and SARS-CoV-2 spike protein by CTSB (250 nM) and CTSL (25 nM) **B**. Concentration-dependent cleavage of the SARS-CoV-2 spike protein by CTSL. **C**. Western blotting against the C-terminal FLAG epitope of SARS-CoV-2 spike demonstrates that CTSL cleavage occurs in the S1 domain. **D.** CTSL (25 nM) proteolysis of the SARS-CoV-2 spike protein is inhibited by K777. **E**. Western blotting against CTSL shows differential expression between cell lines (P: pro form of CTSL; M: mature form of CTSL). **F.** Densitometry analysis of the mature CTSL in **E. G**. Activity of cell lysates with Z-FR-AMC in the presence and absence of inhibitors.

Close analysis of structure of the CoV-2 S protein reveals multiple unresolved loop regions, presumably of high disorder, in the S1 domain of the CoV-2-S structure which comprise different primary sequences than that of CoV-1-S protein, tempting one to speculate that one of these loop regions could be proteolyzed by CTSL.[340]

We next investigated the effects of added K777 on CTSL-catalyzed cleavage of CoV-2-S protein. We observed robust inhibition of proteolysis of CoV-2 S protein by CTSL in the presence of K777 (Figure 68D), suggesting that the inactivation of CTSL by K777 in host cells may contribute to lowered viral infectivity. Together, these *in vitro* data suggest that K777 effectively blocks processing of the CoV-2 S protein by CTSL, and that the CoV-2 S protein is processed by CTSL at a proposed novel S1' site. Further studies will be necessary to elucidate the biological significance of this altered processing, and determine if this altered processing is indeed recapitulated in virally infected SARS-CoV-2 cells or in animal models. In addition, the finding that CoV-1 S protein can be cleaved *in vitro* by CTSB and CTSL may lend insight as to why K777 and other CTSL-specific protease inhibitors have shown enhanced efficacy against SARS-CoV-2 infected cells when compared to their effects on SARS-CoV-1. In particular, for K777, the potency of inactivation of CTSL is approximately 330-fold greater than CTSB *in vitro*, which is reflected in the ability of K777 to prevent CTSL-catalyzed processing of CoV-2 S.

vi. Cathepsin L is Present in All Cell Lines Tested but Levels Vary

Although K777 targets both CTSL and CTSB, inhibition of CTSL is responsible for its potent antiviral effects. To investigate why virally infected A549/ACE2,

HeLa/ACE2, and Vero E6 cells are more sensitive to K777 treatment than Calu-3 cells, we determined the relative abundance of CTSL protein in each cell line. Human CTSL shares 96% protein sequence identity with CTSL in African green monkeys and therefore both the proprotein (upper band) and mature active enzyme (lower band) can be detected by immunoblot using an anti-human CTSL antibody (Figure 68E-F). Mature CTSL was 8.8-fold and 8.7-fold higher in Vero E6 extracts compared to HeLa/ACE2 and A549 respectively. Under these exposure conditions, CTSL was not detectable in Calu-3 cell extracts, consistent with low abundance of CTSL in Calu-3 cells as has been previously described. [341, 342] The large difference in CTSL abundance between these cell lines was further validated by quantitative proteomic analysis using a CTSL-derived peptide (NHCGIASSASYPTV) that is identical in humans and African green monkeys (data not shown). This peptide was between 32- and 57-fold more abundant in Vero E6 cells compared to the other three cell lines, therefore supporting the immunoblotting studies.

Finally, we quantified the specific activity of CTSL in all four cell lines using the fluorogenic substrate Z-FR-AMC. The cell extracts were first treated with broad-spectrum serine, aspartic acid, and metallo-protease inhibitors to inactivate non-cysteine proteases that may hydrolyze this substrate. The remaining activity in each cell line was eliminated by K777 confirming that only cysteine proteases were active in the non-K777 treated samples. As K777 inhibits both CTSL and CTSB, we treated the cell lines with CA-074, a selective CTSB inhibitor.[89] Cysteine protease activity in A549, HeLa/ACE2 and Calu-3 cell lysates was found to be mostly due to CTSB, while activity in Vero E6 cells was mostly CTSL. The specific activity of CTSL was determined to be 0.1, 0.2, 0.5 and 36.7

RFU s^{-1} μg^{-1} for Calu-3, A549, HeLa/ACE2, and Vero E6, respectively (Figure 68G) revealing that cell lines used for SARS-CoV-2 infection assay have high variability in CTSL abundance.

D. Conclusion and Future Outlook

Three host proteases, TMPRSS-2, furin, and CTSL, have been proposed as key enzymes for the processing of the coronaviral spike protein.[93, 122, 316, 317] In our analysis, we have shown that K777, a potent inhibitor of CTSL activity, is also a potent anti-SARS-CoV-2 infection agent in several host cell models. Although the maximal rates of inactivation, k_{inact}, for cathepsins L, B, and K are similar, the second-order rate constant of time-dependent inactivation (k_{inact}/K_I) varies by over two orders-of-magnitude for these enzymes and by five orders-of-magnitude compared to cathepsin C, highlighting the functional selectivity of K777 for CTSL. Using a propargyl analog of K777 as an affinity probe, we determined that CTSB and CTSL were intracellular targets of this inhibitor in Vero E6 cells. However only CTSL, cleaved the SARS-CoV-2 spike protein. This cleavage occurred in the S1 domain of the spike protein at a site different from that observed in the S protein of SARS-CoV-1. The S1 domains of the spike proteins from SARS-CoV-1 and SARS-CoV-2 share 64% sequence identity, and analysis of the SARS-CoV-2 spike protein structure revealed multiple unresolved loop regions in the S1 domain, suggesting that CTSL proteolysis may occur in one of these unstructured regions.[336, 340]

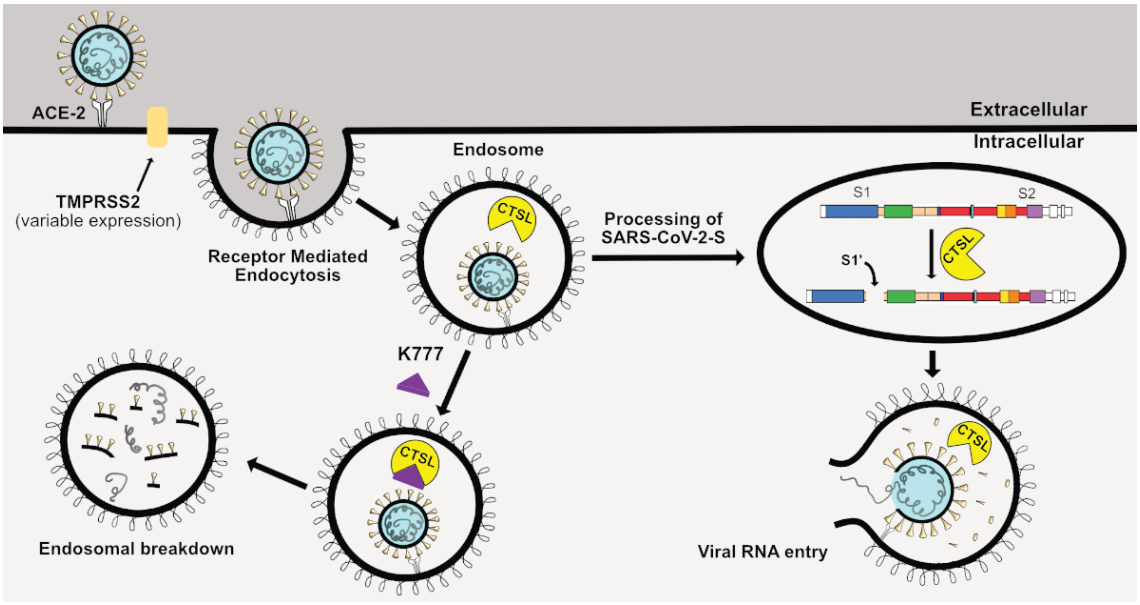


Figure 69. Proposed Mechanism of Anti-SARS-CoV-2 Activity of K777

K777 showed antiviral activity in three host cell models including cell lines derived from African green monkey renal epithelium (Vero E6), human cervical epithelium (HeLa/ACE2) and human lung epithelium (A549/ACE2) with nanomolar potency. Antiviral inhibition was lower (EC_{50} = 3.7 μM) in a second human lung epithelium cell line (Calu-3) and a human colon epithelium-derived cell line (Caco-2; EC_{50} = 4.3 μM). Why is the efficacy of K777 different in different cell lines? Quantification of differential levels of CTSL in cell lines via Western blotting and an enzyme assay confirmed that Vero E6 cells expressed the highest levels of CTSL, while HeLa/ACE2 and A549/ACE2 had lower amounts. Calu-3 cells had the lowest amount of CTSL protein, but were relatively insensitive to K777 treatment. This is consistent with previous reports that the broad cysteine protease inhibitor E64d exhibited diminished efficacy of against MERS and SARS-CoV-2 infected Calu-3 cells, which was attributed to the cells

containing low levels CTSL, but high levels of TMPRSS-2.[341, 342] Correspondingly, a recent CRISPR knockout screen that identified essential host genes for SARS-CoV-2 infection found that TMPRSS-2, but not CTSL, was essential for viral infection of Calu-3 cells.[343, 344] However, the treatment of SARS-CoV-2 infected Calu-3 cells with the TMPRSS-2 inhibitor, camostat mesylate (EC_{50} = 1 μM), reduced, but did non abrogate infection, leading Hoffman *et al.*[93] to propose that residual spike protein might require CTSL processing. In addition, SARS-CoV-2 patient data indicated heightened levels of CTSL in lung tissue and plasma correlate with fatal or severe disease outcomes, suggesting that CTSL, not TMPRSS-2, may be a key protease for efficient viral entry and inflammatory reactions that lead to severe disease.[345, 346] Therefore, if primary infection occurs in the lung epithelium, it could be argued that the diminished efficacy of K777 in Calu-3 cells may result in a reduced ability to prevent primary lung infections, but its robust inhibition of viral infection in another lung epithelial cell line, A549/ACE2, underscores the differential gene expression profiles of lung epithelial cells. (Table 1). Moreover, both Calu-3 and A549/ACE2 cells are derived from adenocarcinomas of the lung, not primary lung epithelium, and therefore may have multiple mutations that could alter infection potential. Thus, future experiments should evaluate a combination of K777 and an appropriate TMPRSS-2 or furin inhibitor.[347] This combination therapy might protect all host cells from viral infection even if some utilize both CTSL and TMPRSS-2 (or another serine protease) for processing of the SARS-CoV-2 spike protein.[345, 346] Confirmation of these results, and the potential efficacy of dual protease inhibition, will come from further tests in animal models, and validation in human clinical trials. Support

for the rapid advancement of K777 as potential COVID-19 therapy comes from the fact that K777 is orally bioavailable, and safety, tolerability and PK/PD have been evaluated in several preclinical model systems, including non-human primates.[326] Selva Therapeutics has been granted an open IND for K777 with the FDA (Selva Therapeutics, Del Mar, CA).

CHAPTER VII

CONCLUSIONS

Covalent drug discovery using mechanism-based approaches comprises a powerful and attractive methodology for enzyme-based drug discovery. This dissertation has described various studies that highlight the importance of developing a keen understanding of the kinetic and mechanistic intricacies of each enzyme target and how to employ this information to identify, create, or repurpose chemical matter to produce potent enzyme inhibitors and inactivators.

The first study in this dissertation focused on the synthesis and kinetic characterization of cruzain, a cysteine protease from the parasitic protozoa *Trypanosoma cruzi*. Synthesis and kinetic analysis of dipeptide substrates containing varied natural and nonnatural amino acids assisted in identification of cruzain's ability to tolerate polar amino acids in its P_2 subsite. From this information, inhibitors were prepared containing an electrophilic warhead, in lieu of the hydrolysable 7-amino-4-methylcoumarin moiety. Analysis of the kinetic data revealed a strong linear relationship between the K_m of the fluorogenic substrate and the tight-binding inhibition constant (K_i*), suggesting the key driver of potency for this class of cruzain inhibitors is efficient binding of the peptide scaffold to the enzyme target and not the electrophilic warhead.

In order to obtain more selective substrates and inhibitors, future studies should focus on developing targeted combinatorial libraries to screen variable di- and tri-fluorogenic peptide substrates, containing primarily non-natural amino acids, against

cruzain and structurally similar cysteine proteases. While previous studies using large combinatorial libraries have detailed the substrate specificity of various cysteine proteases, many employed strictly natural amino acids and focused on the change in relative fluorescent units at fixed substrate concentrations.[139, 140] Ideally, using a smaller and more targeted library would allow researchers to vary substrate concentrations and elucidate kinetic constants such as k_{cat}, K_m, and k_{cat}/K_m. This information could then guide further optimization of selective substrate scaffolds prior to appending an electrophilic warhead to the non-natural peptide scaffold. This approach may prove optimal for developing highly selective inhibitors against cruzain and is supported by the success of previous work which developed selective non-natural peptide scaffolds for malarial cathepsin C versus human and bovine forms[348] as well as for caspases.[349]

The second study of this dissertation identified *cis*-epoxy succinic acid (*cis*-EpS) as a potent uncompetitive inactivator of isocitrate lyase-1 (ICL1) from *Mycobacterium tuberculosis;* and, under conditions where ICL1 is essential, *cis*-EpS exerted moderate anti-mycobacterial activity. Kinetic, structural, and mass spectrometric analysis of ICL1 with *cis*-EpS identified it to be a covalent and irreversible inactivator through alkylation of the catalytic cysteine 191 of ICL1. X-ray structures of ICL1 bound to *cis*-EpS revealed a covalent thioether bond between *cis*-EpS and Cys_{191} and glyoxylate hydrate coordinated to the active site Mg^{2+} cation in the closed active site cavity. A crystallographic snapshot of *cis*-EpS inactivation suggested that the hydration of glyoxylate proceeds concurrent with *cis*-EpS covalent bond formation, thus allowing for the proposal of a putative chemical mechanism by which inactivation proceeds.

Future work is already underway in our lab to identify and produce analogs of *cis*-EpS that may exert more potent anti-mycobacterial effects. One approach would be to introduce different warheads, such as aziridines. However, since the epoxide warhead is so effective against ICL1, it may be more viable to produce analogs that retain the epoxide, but vary the modifications on C1 and C4 of the cis-EpS molecule. These modifications could range from esters (for prodrugs), to substitution of the carboxylic acid groups with bioisosteric replacements such as triazoles, tetrazoles, sulfonamides, or even squaric acids.[350] In addition, one could envision the installation of an oxirane ring within ICL1 substrate analogs, which would afford effective positioning of this weakly electrophilic warhead to undergo thiolate attack. This has been shown to be an effective strategy with isocitrate lyase from watermelon cotyledons using itaconate expoxide.[351] Together, these strategies could prove useful in developing new analogs of *cis*-EpS that still effectively target ICL1 while enhancing the anti-mycobacterial activity of the molecule.

The third study in this dissertation also focused on ICL1, but, in this case, we developed 5-descarboxy-5-nitro-D-isocitric acid, or 5-NIC, a mechanism-based inactivator of the enzyme. Using enzyme kinetics, X-ray crystallography, and mass spectrometry we identified that 5-NIC effects covalent and effectively irreversible inactivation of ICL1 via the formation glyoxylate hydrate and a thiohydroximate adduct with cysteine 191 in the enzyme active site. This data strongly suggests that 5-NIC undergoes cleavage of the carbon-2 carbon-3 bond to produce glyoxylate and 3-nitropropionate within the ICL1 active site. 5-NIC effects highly efficient mechanism-based inactivation of ICL1 with a partition ratio of 0.25, demonstrating 1.25 molecules of

5-NIC are required to inactivate 1 molecule of ICL1. pH rate profile studies identified that 5-NIC bound to ICL1 in its nitro form, suggesting the 3-nitropropionate produced by 5-NIC cleavage must undergo interconversion to its nitronate form in the enzyme active site in order to result in effective ICL1 inactivation. The nitronate form of 5-NIC was determined to undergo rapid reversion back to its nitro form in neutral pH solution. Using a mixture of mass spectrometry and NMR it was identified that 5-NIC underwent facile exchange with deuterium oxide at its carbon 4 position, suggestive that it rapidly interconverts between its nitronate and nitro forms. We propose this phenomenon may be assisted by the carbon-2 hydroxyl of 5-NIC through an intramolecular proton transfer.

When designing mechanism-based inactivators (MBIs) of enzymes, one of the critical challenges is to produce a molecule that both effectively engages with the enzyme target of interest, but also undergoes efficient inactivation of the enzyme. This is particularly true for MBIs that generate strongly electrophilic products, in the active site of the enzyme. Our lab has successfully designed two different types of MBIs for ICL1 and found that 5-NIC effects much more potent inactivation when compared to 2-VIC (discussed in Chapter IV). We posit this result, as well as our findings in Chapter III, suggests that succinate-based electrophiles effect more robust inactivation of ICL1 versus glyoxylate-based electrophiles. This is likely for two major reasons, 1) cysteine 191 resides more proximal the succinate binding site and 2) the ordered product release of succinate followed by glyoxylate (in the direction of D-isocitrate synthesis). Due to this fact, we postulate that further development of succinate-based electrophiles which are either appended to an isocitrate-like scaffold or are confined to a *synperiplanar*

conformation may provide the best ICL1 covalent inactivators and mechanism-based inactivators of ICL1.

The fourth study in this thesis characterized the substrate specificity of *Mt*-Sirt (Rv1151c), the sole annotated NAD^+-dependent lysine deacylase from *Mycobacterium tuberculosis*. We identified that this enzyme has a preference for succinylated and long chain fatty acyl modified lysine substrates. Mutation of a key arginine residue of *Mt*-Sirt ($R_{56}M$), proposed to be required for efficient binding of succinyllysine peptides, resulted in loss of desuccinylation activity, but increased the activity and specificity of the enzyme for a propionyllysine peptide. Incubation of a fatty acid alkyne probe with mc^27000, an auxotrophic *Mtb* strain, resulted in clear visualization of fatty acylation of proteins in *Mtb*, suggesting *Mt*-Sirt fatty acyl lysine deacylation may play an uncharacterized role *in vivo*. Further study characterized the inhibition of *Mt*-Sirt by nicotinamide, adenosine 5′-diphosphoribose, and reduced β-nicotinamide adenine dinucleotide, as well as, mechanism-based thiourea and thioamide inhibitors which proved to potently inhibit the enzyme.

Ongoing studies are underway in our laboratory to elucidate the detailed kinetic mechanism of *Mt*-Sirt and attempt to better understand the fatty acyl lysine deacylase activity of the enzyme. Subsequent work for Mt-Sirt should further investigate this protein from both an enzymology/*in vitro* and an *in vivo* perspective. One important missing piece to understanding this enzyme is a crystal structure or solution phase NMR structure. Unfortunately, our attempts to crystalize various constructs of this protein in over 25,000 sparce matrix conditions and produce sufficient quantities of labeled protein for NMR

studies were unsuccessful. However, this structural information is critical in understanding how long chain fatty acyl peptides bind to enzymes like *Mt*-Sirt that contain a YxxR motif, but do not bind to *E.coli* CobB which contains the same motif and similar active site pockets.[289] Therefore, future studies should focus on deciphering the structural requirements for fatty acyl chain binding of proteins containing a YxxR motif. Such studies may be amenable for human sirtuin 5 as dozens of high-resolution crystal structures exist for this protein, thus highlighting its propensity to form protein crystals.

Shifting to *in vivo* studies, a very important endeavor is identifying the intracellular targets of *Mt*-Sirt. Proteomic studies have identified nearly 1,800 sites of lysine succinylation on nearly 700 proteins within the *Mtb* proteome, but at present a limited amount of these modified proteins have been investigated.[273] One way to limit the amount of protein targets that need to be interrogated, and to identify the *bona fide* protein targets of *Mt*-Sirt, is by conducting quantitative proteomic analysis.

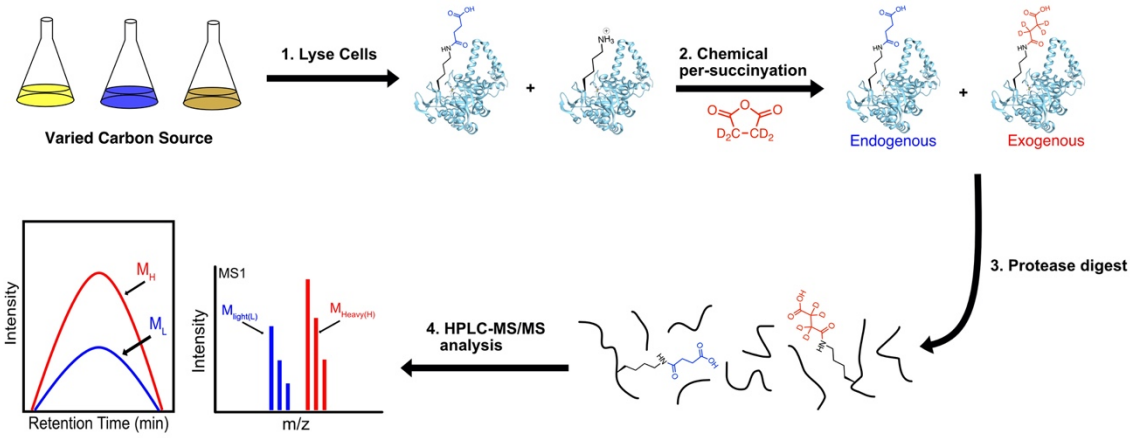


Figure 70. Lysine PTM Stoichiometry Workflow

Mtb cell lysate is treated with succinic anhydride-d4 to modify non-succinylated lysine residues on proteins. The labeled proteome is digested by GluC results in peptide fragments which are then injected into a mass spectrometer. The data obtained can be utilized to calculate the ratio of light versus heavy modification, thus allowing determination of stoichiometry

In the case of identifying putative protein substrates that Mt-Sirt can desuccinylate *in vivo*, one would culture both wild-type and Δ*mt-sirt Mtb* strains, chemically label the cell lysate with succinic anhydride-d4 to modify all free amine lysine residues in the proteome, and proteolyze the proteins with GluC. Subsequent LC/MS/MS analysis would allow quantitation of the heavy to light ratio for each peptide fragment observed and thus allow for tentative assignment of the protein substrates of *Mt*-Sirt. Beyond these studies for succinyl lysine, it would be ideal to conduct analogous proteomic studies for other lysine PTMs in *Mtb* as the *in vitro* activity of *Mt*-Sirt may not be fully recapitulated *in vivo*. Moreover, these studies should be conducted under varied carbon source conditions as the

abundance of lysine PTMs, such as acetylation and succinylation, have been observed to be highly dependent on carbon source. [273, 352]

In the fifth and final study in this dissertation, we identified a vinyl sulfone containing protease inhibitor, called K777, which exerted potent activity against SARS-CoV-2 infected human and monkey cells. Kinetic analysis identified that K777 did not inhibit the cysteine proteases 3CLpro or PLpro from SARS-CoV-2, but underwent robust covalent inactivation with human cathepsin L and, to a lesser extent, irreversibly inactivated human cathepsins B and K. Synthesis of a propargyl analog of K777 and treatment of SARS-CoV-2 infected and uninfected cells with this probe, followed by lysis and 1,3 dipolar cycloaddition of the K777 alkyne to a Cy7 azide, revealed robust modification of a primary protein target at ~25 kDa. Subsequent enrichment of the alkyne modified proteins, followed by proteomic analysis, resulted in enrichment of human cathepsins B and L. To identify which of these enzymes was responsible for spike protein processing, we incubated both cathepsins with recombinant SARS-CoV-2 spike protein. We observed the spike protein was proteolyzed by cathepsin L, but not cathepsin B, and that this processing was blocked by K777, suggestive that cathepsin L is the intracellular target of K777 and blocks cathepsin L mediated SARS-CoV-2 spike protein processing.

Beyond cathepsin L, there are other proteases that can proteolytically process the spike protein and result in efficient invasion of SARS-CoV-2 into the human cells. These include furin-like proteases and transmembrane protease serine 2.[353] While a singular protease inhibitor such as K777 may have a pronounced effect on the ability of the virus to invade cells that contain abundant cathepsin L, the effect of K777 on cells with less

cathepsin L expression may not be as well protected. Therefore, it may be wise to develop therapies which contain both K777, or a similar cathepsin L inhibitor, as well as other small molecule inhibitors which block furin-like proteases and the trypsin-like transmembrane protease serine 2.

REFERENCES

[1] Lee, C. U., and Grossmann, T. N. (2012) Reversible covalent inhibition of a protein target, *Angew Chem Int Ed Engl 51*, 8699-8700.
[2] Johansson, M. (2012) Reversible Michael Additions- Covalent Inhibitors and Prodrugs, *Mini-Reviews in Medicinal Chemistry 12*, 14.
[3] Serafimova, I. M., Pufall, M. A., Krishnan, S., Duda, K., Cohen, M. S., Maglathlin, R. L., McFarland, J. M., Miller, R. M., Frödin, M., and Taunton, J. (2012) Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles, *Nature chemical biology 8*, 471-476.
[4] Krishnan, S., Miller, R. M., Tian, B., Mullins, R. D., Jacobson, M. P., and Taunton, J. (2014) Design of reversible, cysteine-targeted Michael acceptors guided by kinetic and computational analysis, *Journal of the American Chemical Society 136*, 12624-12630.
[5] Bradshaw, J. M., McFarland, J. M., Paavilainen, V. O., Bisconte, A., Tam, D., Phan, V. T., Romanov, S., Finkle, D., Shu, J., Patel, V., Ton, T., Li, X., Loughhead, D. G., Nunn, P. A., Karr, D. E., Gerritsen, M. E., Funk, J. O., Owens, T. D., Verner, E., Brameld, K. A., Hill, R. J., Goldstein, D. M., and Taunton, J. (2015) Prolonged and tunable residence time using reversible covalent kinase inhibitors, *Nature chemical biology 11*, 525-531.
[6] Latorre, A., Schirmeister, T., Kesselring, J., Jung, S., Johé, P., Hellmich, U. A., Heilos, A., Engels, B., Krauth-Siegel, R. L., Dirdjaja, N., Bou-Iserte, L., Rodríguez, S., and González, F. V. (2016) Dipeptidyl Nitroalkenes as Potent Reversible Inhibitors of Cysteine Proteases Rhodesain and Cruzain, *ACS Medicinal Chemistry Letters*.
[7] Silverman, R. B. (1995) [10] Mechanism-based enzyme inactivators, In *Methods in Enzymology*, pp 240-283, Academic Press.
[8] Walsh, C. (1982) Suicide substrates: mechanism-based enzyme inactivators, *Tetrahedron 38*, 871-909.
[9] Walsh, C. T. (1984) SUICIDE SUBSTRATES, MECHANISM-BASED ENZYME INACTIVATORS: RECENT DEVELOPMENTS, *Annual Review of Biochemistry 53*, 493-535.
[10] Pauling, L. (1946) Molecular Architecture and Biological Reactions, *Chemical & Engineering News Archive 24*, 1375-1377.
[11] Pauling, L. (1948) Nature of Forces between Large Molecules of Biological Interest*, *Nature 161*, 707-709.
[12] Lewandowicz, A., Tyler, P. C., Evans, G. B., Furneaux, R. H., and Schramm, V. L. (2003) Achieving the Ultimate Physiological Goal in Transition State Analogue Inhibitors for Purine Nucleoside Phosphorylase*, *Journal of Biological Chemistry 278*, 31465-31468.
[13] Schramm, V. L. (2013) Transition States, analogues, and drug development, *ACS chemical biology 8*, 71-81.
[14] Pham, T. V., Mellott, D. M., Moghadamchargari, Z., Chen, K., Krieger, I., Laganowsky, A., Sacchettini, J. C., and Meek, T. D. (2021) Covalent Inactivation

of Mycobacterium tuberculosis Isocitrate Lyase by cis-2,3-Epoxy-Succinic Acid, *ACS Chemical Biology 16*, 463-470.
[15] Copeland, R. A., Pompliano, D. L., and Meek, T. D. (2006) Drug–target residence time and its implications for lead optimization, *Nature Reviews Drug Discovery 5*, 730-739.
[16] Walkup, G. K., You, Z., Ross, P. L., Allen, E. K. H., Daryaee, F., Hale, M. R., O'Donnell, J., Ehmann, D. E., Schuck, V. J. A., Buurman, E. T., Choy, A. L., Hajec, L., Murphy-Benenato, K., Marone, V., Patey, S. A., Grosser, L. A., Johnstone, M., Walker, S. G., Tonge, P. J., and Fisher, S. L. (2015) Translating slow-binding inhibition kinetics into cellular and in vivo effects, *Nature Chemical Biology 11*, 416-423.
[17] Werkheiser, W. C. (1961) Specific Binding of 4-Amino Folic Acid Analogues by Folic Acid Reductase, *Journal of Biological Chemistry 236*, 888-893.
[18] Rawat, R., Whitty, A., and Tonge, P. J. (2003) The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: adduct affinity and drug resistance, *Proceedings of the National Academy of Sciences of the United States of America 100*, 13881-13886.
[19] Moellering, R. E., and Cravatt, B. F. (2012) How chemoproteomics can enable drug discovery and development, *Chemistry & biology 19*, 11-22.
[20] Spradlin, J. N., Zhang, E., and Nomura, D. K. (2021) Reimagining Druggability Using Chemoproteomic Platforms, *Accounts of Chemical Research 54*, 1801-1813.
[21] Hinman, A., Chuang, H.-h., Bautista, D. M., and Julius, D. (2006) TRP channel activation by reversible covalent modification, *Proceedings of the National Academy of Sciences 103*, 19564.
[22] Macpherson, L. J., Dubin, A. E., Evans, M. J., Marr, F., Schultz, P. G., Cravatt, B. F., and Patapoutian, A. (2007) Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines, *Nature 445*, 541-545.
[23] Fu, H., Park, J., and Pei, D. (2002) Peptidyl Aldehydes as Reversible Covalent Inhibitors of Protein Tyrosine Phosphatases, *Biochemistry 41*, 10700-10709.
[24] Fontanillo, M., and Köhn, M. (2018) Microcystins: Synthesis and structure–activity relationship studies toward PP1 and PP2A, *Bioorganic & Medicinal Chemistry 26*, 1118-1126.
[25] Wang, J., Zhang, J., Zhang, C.-J., Wong, Y. K., Lim, T. K., Hua, Z.-C., Liu, B., Tannenbaum, S. R., Shen, H.-M., and Lin, Q. (2016) In situ Proteomic Profiling of Curcumin Targets in HCT116 Colon Cancer Cell Line, *Scientific Reports 6*, 22146.
[26] Kalesh, K. A., Clulow, J. A., and Tate, E. W. (2015) Target profiling of zerumbone using a novel cell-permeable clickable probe and quantitative chemical proteomics, *Chemical Communications 51*, 5497-5500.
[27] Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E. P., Wolff, B., Yoshida, M., and Horinouchi, S. (1999) Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central

conserved region, *Proceedings of the National Academy of Sciences of the United States of America 96*, 9112-9117.
[28] Kwok, B. H. B., Koh, B., Ndubuisi, M. I., Elofsson, M., and Crews, C. M. (2001) The anti-inflammatory natural product parthenolide from the medicinal herb Feverfew directly binds to and inhibits IκB kinase, *Chemistry & Biology 8*, 759-766.
[29] Berdan, C. A., Ho, R., Lehtola, H. S., To, M., Hu, X., Huffman, T. R., Petri, Y., Altobelli, C. R., Demeulenaere, S. G., Olzmann, J. A., Maimone, T. J., and Nomura, D. K. (2019) Parthenolide Covalently Targets and Inhibits Focal Adhesion Kinase in Breast Cancer Cells, *Cell chemical biology 26*, 1027-1035.e1022.
[30] Böttcher, T., and Sieber, S. A. (2010) Showdomycin as a Versatile Chemical Tool for the Detection of Pathogenesis-Associated Enzymes in Bacteria, *Journal of the American Chemical Society 132*, 6964-6972.
[31] Greger, H., Hofer, O., Zechner, G., Hadacek, F., and Wurz, G. (1994) Sulphones derived from methylthiopropenoic acid amides from Glycosmis angustifolia, *Phytochemistry 37*, 1305-1310.
[32] Astelbauer, F., Obwaller, A., Raninger, A., Brem, B., Greger, H., Duchêne, M., Wernsdorfer, W., and Walochnik, J. (2010) High antitrypanosomal activity of plant-derived sulphur-containing amides, *Int J Antimicrob Agents 36*, 570-572.
[33] Liu, Y., Shreder, K. R., Gai, W., Corral, S., Ferris, D. K., and Rosenblum, J. S. (2005) Wortmannin, a Widely Used Phosphoinositide 3-Kinase Inhibitor, also Potently Inhibits Mammalian Polo-like Kinase, *Chemistry & Biology 12*, 99-107.
[34] Weerapana, E., Simon, G. M., and Cravatt, B. F. (2008) Disparate proteome reactivity profiles of carbon electrophiles, *Nature Chemical Biology 4*, 405-407.
[35] Chenna, B. C., Li, L., Mellott, D. M., Zhai, X., Siqueira-Neto, J. L., Calvet Alvarez, C., Bernatchez, J. A., Desormeaux, E., Alvarez Hernandez, E., Gomez, J., McKerrow, J. H., Cruz-Reyes, J., and Meek, T. D. (2020) Peptidomimetic Vinyl Heterocyclic Inhibitors of Cruzain Effect Antitrypanosomal Activity, *Journal of Medicinal Chemistry 63*, 3298-3316.
[36] Kuter, D. J., Boccia, R. V., Lee, E.-J., Efraim, M., Tzvetkov, N., Mayer, J., Trněný, M., Kostal, M., Hajek, R., McDonald, V., Bandman, O., Burns, R., Neale, A., Thomas, D., and Cooper, N. (2019) Phase I/II, Open-Label, Adaptive Study of Oral Bruton Tyrosine Kinase Inhibitor PRN1008 in Patients with Relapsed/Refractory Primary or Secondary Immune Thrombocytopenia, *Blood 134*, 87-87.
[37] Krishnan, S., Miller, R. M., Tian, B., Mullins, R. D., Jacobson, M. P., and Taunton, J. (2014) Design of reversible, cysteine-targeted Michael acceptors guided by kinetic and computational analysis, *J Am Chem Soc 136*, 12624-12630.
[38] Chen, X., Wang, Y., Ma, N., Tian, J., Shao, Y., Zhu, B., Wong, Y. K., Liang, Z., Zou, C., and Wang, J. (2020) Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification, *Signal Transduction and Targeted Therapy 5*, 72.

[39] Niphakis, M. J., and Cravatt, B. F. (2014) Enzyme Inhibitor Discovery by Activity-Based Protein Profiling, *Annual Review of Biochemistry 83*, 341-377.
[40] van Esbroeck, A. C. M., Janssen, A. P. A., Cognetta, A. B., Ogasawara, D., Shpak, G., van der Kroeg, M., Kantae, V., Baggelaar, M. P., de Vrij, F. M. S., Deng, H., Allarà, M., Fezza, F., Lin, Z., van der Wel, T., Soethoudt, M., Mock, E. D., den Dulk, H., Baak, I. L., Florea, B. I., Hendriks, G., De Petrocellis, L., Overkleeft, H. S., Hankemeier, T., De Zeeuw, C. I., Di Marzo, V., Maccarrone, M., Cravatt, B. F., Kushner, S. A., and van der Stelt, M. (2017) Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor BIA 10-2474, *Science 356*, 1084.
[41] Wright, M. H., and Sieber, S. A. (2016) Chemical proteomics approaches for identifying the cellular targets of natural products, *Natural Product Reports 33*, 681-708.
[42] Staub, I., and Sieber, S. A. (2008) β-Lactams as Selective Chemical Probes for the in Vivo Labeling of Bacterial Enzymes Involved in Cell Wall Biosynthesis, Antibiotic Resistance, and Virulence, *Journal of the American Chemical Society 130*, 13400-13409.
[43] Böttcher, T., and Sieber, S. A. (2008) β-Lactones as Privileged Structures for the Active-Site Labeling of Versatile Bacterial Enzyme Classes, *Angewandte Chemie International Edition 47*, 4600-4603.
[44] Staub, I., and Sieber, S. A. (2009) β-Lactam Probes As Selective Chemical-Proteomic Tools for the Identification and Functional Characterization of Resistance Associated Enzymes in MRSA, *Journal of the American Chemical Society 131*, 6271-6276.
[45] Kong, K.-F., Schneper, L., and Mathee, K. (2010) Beta-lactam antibiotics: from antibiosis to resistance and bacteriology, *APMIS 118*, 1-36.
[46] Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Lechner, A., Kale, A. J., Kobayashi, Y., Reynolds, K. A., and Moore, B. S. (2009) Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adeozyl-L0 methionine, *Proceedings of the National Academy of Sciences 106*, 12295.
[47] Medicine, U. S. N. L. o. (2021) A Phase III Trial of With Marizomib in Patients With Newly Diagnosed Glioblastoma (MIRAGE), U.S National Library of Medicine
[48] Matsumoto, K., Mizoue, K., Kitamura, K., Tse, W.-C., Huber, C. P., and Ishida, T. (1999) Structural basis of inhibition of cysteine proteases by E-64 and its derivatives, *Peptide Science 51*, 99-107.
[49] Meng, L., Mohan, R., Kwok, B. H. B., Elofsson, M., Sin, N., and Crews, C. M. (1999) Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity, *Proceedings of the National Academy of Sciences 96*, 10403.
[50] Kim, K. B., and Crews, C. M. (2013) From epoxomicin to carfilzomib: chemistry, biology, and medical outcomes, *Natural Product Reports 30*, 600-604.

[51] Wilson, M. R., Jiang, Y., Villalta, P. W., Stornetta, A., Boudreau, P. D., Carrá, A., Brennan, C. A., Chun, E., Ngo, L., Samson, L. D., Engelward, B. P., Garrett, W. S., Balbo, S., and Balskus, E. P. (2019) The human gut bacterial genotoxin colibactin alkylates DNA, *Science (New York, N.Y.) 363*, eaar7785.
[52] Xue, M., Kim, C. S., Healy, A. R., Wernke, K. M., Wang, Z., Frischling, M. C., Shine, E. E., Wang, W., Herzon, S. B., and Crawford, J. M. (2019) Structure elucidation of colibactin and its DNA cross-links, *Science (New York, N.Y.) 365*, eaax2685.
[53] Dembitsky, V. M., Terent'ev, A. O., and Levitsky, D. O. (2013) Aziridine Alkaloids: Origin, Chemistry and Activity, pp 977-1006, Springer-Verlag: Berlin, Heidelberg.
[54] Nakao, Y., Fujita, M., Warabi, K., Matsunaga, S., and Fusetani, N. (2000) Miraziridine A, a Novel Cysteine Protease Inhibitor from the Marine Sponge Theonella aff. mirabilis1, *Journal of the American Chemical Society 122*, 10462-10463.
[55] Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, *Science 368*, 409.
[56] Wofsy, L., Metzger, H., and Singer, S. (1962) Affinity labeling—a general method for labeling the active sites of antibody and enzyme molecules, *Biochemistry 1*, 1031-1039.
[57] Tuley, A., and Fast, W. (2018) The Taxonomy of Covalent Inhibitors, *Biochemistry 57*, 3326-3337.
[58] Krantz, A., Copp, L. J., Coles, P. J., Smith, R. A., and Heard, S. B. (1991) Peptidyl (acyloxy)methyl ketones and the quiescent affinity label concept: the departing group as a variable structural element in the design of inactivators of cysteine proteinases, *Biochemistry 30*, 4678-4687.
[59] Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M. L., Magaard, V. W., and Debouck, C. (1989) Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease, *Proceedings of the National Academy of Sciences 86*, 1841.
[60] Grant, S. K., Moore, M. L., Fakhoury, S. A., Tomaszek, T. A., and Meek, T. D. (1992) Inactivation of HIV-1 protease by a tripeptidyl epoxide, *Bioorganic & Medicinal Chemistry Letters 2*, 1441-1445.
[61] Watts, D. C., Rabin, B. R., and Crook, E. M. (1961) The reaction of iodoacetate and iodoacetamide with proteins as determined with a silver/silver iodide electrode, *Biochimica et Biophysica Acta 48*, 380-388.
[62] Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry, *Annual Review of Biochemistry 77*, 383-414.

[63] Roberts, A. M., Ward, C. C., and Nomura, D. K. (2017) Activity-based protein profiling for mapping and pharmacologically interrogating proteome-wide ligandable hotspots, *Current opinion in biotechnology 43*, 25-33.
[64] Tooke, C. L., Hinchliffe, P., Bragginton, E. C., Colenso, C. K., Hirvonen, V. H. A., Takebayashi, Y., and Spencer, J. (2019) β-Lactamases and β-Lactamase Inhibitors in the 21st Century, *Journal of molecular biology 431*, 3472-3500.
[65] Konaklieva, M. I. (2014) Molecular Targets of β-Lactam-Based Antimicrobials: Beyond the Usual Suspects, *Antibiotics (Basel) 3*, 128-142.
[66] Böttcher, T., and Sieber, S. A. (2012) β-Lactams and β-lactones as activity-based probes in chemical biology, *MedChemComm 3*, 408-417.
[67] Laxminarayan, R., Van Boeckel, T., Frost, I., Kariuki, S., Khan, E. A., Limmathurotsakul, D., Larsson, D. G. J., Levy-Hara, G., Mendelson, M., Outterson, K., Peacock, S. J., and Zhu, Y.-G. (2020) The Lancet Infectious Diseases Commission on antimicrobial resistance: 6 years later, *The Lancet Infectious Diseases 20*, e51-e60.
[68] Abraham, E. P., and Chain, E. (1940) An enzyme from bacteria able to destroy penicillin, *Nature 146*, 837-837.
[69] Howarth, T. T., Brown, A. G., and King, T. J. (1976) Clavulanic acid, a novel β-lactam isolated from Streptomyces clavuligerus; X-ray crystal structure analysis, *Journal of the Chemical Society, Chemical Communications*, 266b-267.
[70] Knowles, J. R. (1985) Penicillin resistance: the chemistry of. beta.-lactamase inhibition, *Accounts of Chemical Research 18*, 97-104.
[71] Imtiaz, U., Billings, E., Knox, J. R., Manavathu, E. K., Lerner, S. A., and Mobashery, S. (1993) Inactivation of class A. beta.-lactamases by clavulanic acid: the role of arginine-244 in a proposed nonconcerted sequence of events, *Journal of the American Chemical Society 115*, 4435-4442.
[72] Brown, R. P., Aplin, R. T., Schofield, C. J., and Frydrych, C. H. (1997) Mass spectrometric studies on the inhibition of TEM-2β-lactamase by clavulanic acid derivatives, *The Journal of antibiotics 50*, 184-185.
[73] Brown, R. P., Aplin, R. T., and Schofield, C. J. (1996) Inhibition of TEM-2 β-lactamase from Escherichia coli by clavulanic acid: observation of intermediates by electrospray ionization mass spectrometry, *Biochemistry 35*, 12421-12432.
[74] English, A. R., Retsema, J. A., Girard, A. E., Lynch, J. E., and Barth, W. E. (1978) CP-45,899, a Beta-Lactamase Inhibitor That Extends the Antibacterial Spectrum of Beta-Lactams: Initial Bacteriological Characterization, *Antimicrobial Agents and Chemotherapy 14*, 414.
[75] Micetich, R. G., Maiti, S. N., Spevak, P., Hall, T. W., Yamabe, S., Ishida, N., Tanaka, M., Yamazaki, T., Nakai, A., and Ogawa, K. (1987) Synthesis and .beta.-lactamase inhibitory properties of 2.beta.-[(1,2,3-triazol-1-yl)methyl]-2.alpha.-methylpenam-3.alpha.-carboxylic acid 1,1-dioxide and related triazolyl derivatives, *Journal of Medicinal Chemistry 30*, 1469-1474.
[76] Stachyra, T., Péchereau, M.-C., Bruneau, J.-M., Claudon, M., Frère, J.-M., Miossec, C., Coleman, K., and Black, M. T. (2010) Mechanistic Studies of the Inactivation

of TEM-1 and P99 by NXL104, a Novel Non-β-Lactam β-Lactamase Inhibitor, *Antimicrobial Agents and Chemotherapy 54*, 5132.
[77] Coleman, K. (2011) Diazabicyclooctanes (DBOs): a potent new class of non-β-lactam β-lactamase inhibitors, *Current Opinion in Microbiology 14*, 550-555.
[78] Ehmann, D. E., Jahić, H., Ross, P. L., Gu, R.-F., Hu, J., Kern, G., Walkup, G. K., and Fisher, S. L. (2012) Avibactam is a covalent, reversible, non–β-lactam β-lactamase inhibitor, *Proceedings of the National Academy of Sciences 109*, 11663.
[79] Shirley, M. (2018) Ceftazidime-Avibactam: A Review in the Treatment of Serious Gram-Negative Bacterial Infections, *Drugs 78*, 675-692.
[80] Patel, T. S., Pogue, J. M., Mills, J. P., and Kaye, K. S. (2018) Meropenem-vaborbactam: a new weapon in the war against infections due to resistant Gram-negative bacteria, *Future Microbiol 13*, 971-983.
[81] Wong, D., and Van Duin, D. (2017) Novel beta-lactamase inhibitors: unlocking their potential in therapy, *Drugs 77*, 615-628.
[82] Naas, T., Oueslati, S., Bonnin, R. A., Dabos, M. L., Zavala, A., Dortet, L., Retailleau, P., and Iorga, B. I. (2017) Beta-lactamase database (BLDB)–structure and function, *Journal of enzyme inhibition and medicinal chemistry 32*, 917-919.
[83] Brown, E. D., Marquardt, J. L., Lee, J. P., Walsh, C. T., and Anderson, K. S. (1994) Detection and characterization of a phospholactoyl-enzyme adduct in the reaction catalyzed by UDP-N-acetylglucosamine enolpyruvoyl transferase, MurZ, *Biochemistry 33*, 10638-10645.
[84] Kim, D. H., Lees, W. J., and Walsh, C. T. (1994) Formation of two enzyme-bound reaction intermediate analogs during inactivation of UDP-GlcNAc enolpyruvyl transferase by (E)-or (Z)-3-fluorophosphoenolpyruvate, *Journal of the American Chemical Society 116*, 6478-6479.
[85] Zhu, J.-Y., Yang, Y., Han, H., Betzi, S., Olesen, S. H., Marsilio, F., and Schönbrunn, E. (2012) Functional Consequence of Covalent Reaction of Phosphoenolpyruvate with UDP-N-acetylglucosamine 1-Carboxyvinyltransferase (MurA)*, *Journal of Biological Chemistry 287*, 12657-12667.
[86] Kahan, F. M., Kahan, J. S., Cassidy, P. J., and Kropp, H. (1974) THE MECHANISM OF ACTION OF FOSFOMYCIN (PHOSPHONOMYCIN), *Annals of the New York Academy of Sciences 235*, 364-386.
[87] Marquardt, J. L., Brown, E. D., Lane, W. S., Haley, T. M., Ichikawa, Y., Wong, C.-H., and Walsh, C. T. (1994) Kinetics, Stoichiometry, and Identification of the Reactive Thiolate in the Inactivation of UDP-GlcNAc Enolpyruvoyl Transferase by the Antibiotic Fosfomycin, *Biochemistry 33*, 10646-10651.
[88] Han, H., Yang, Y., Olesen, S. H., Becker, A., Betzi, S., and Schönbrunn, E. (2010) The Fungal Product Terreic Acid Is a Covalent Inhibitor of the Bacterial Cell Wall Biosynthetic Enzyme UDP-N-Acetylglucosamine 1-Carboxyvinyltransferase (MurA), *Biochemistry 49*, 4276-4282.
[89] Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S., and Greenbaum, D. (2000) Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs, *Chemistry & Biology 7*, 27-38.

[90] Greenbaum, D., Medzihradszky, K. F., Burlingame, A., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools, *Chemistry & Biology 7*, 569-581.
[91] Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J., and Tanaka, I. (1978) Isolation and Characterization of E–64, a New Thiol Protease Inhibitor, *Agricultural and Biological Chemistry 42*, 523-528.
[92] Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., and Li, F. (2020) Cell entry mechanisms of SARS-CoV-2, *Proceedings of the National Academy of Sciences 117*, 11727.
[93] Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., and Pöhlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell 181*, 271-280.e278.
[94] Liu, S., and Hanzlik, R. P. (1992) Structure-Activity Relationships for Inhibition of Papain by Peptide Michael Acceptors, *Journal of Medicinal Chemistry 35*, 1067-1075.
[95] Craig, M., Luu, H. A., McCready, T. L., Holmes, C. F. B., Williams, D., and Andersen, R. J. (1996) Molecular mechanisms underlying the interaction of motuporin and microcystins with type-1 and type-2A protein phosphatases, *Biochemistry and Cell Biology 74*, 569-578.
[96] MacKintosh, R. W., Dalby, K. N., Campbell, D. G., Cohen, P. T. W., Cohen, P., and MacKintosh, C. (1995) The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1, *FEBS Letters 371*, 236-240.
[97] Sun, Q., Carrasco, Y. P., Hu, Y., Guo, X., Mirzaei, H., MacMillan, J., and Chook, Y. M. (2013) Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1, *Proceedings of the National Academy of Sciences 110*, 1303.
[98] Newlands, E. S., Rustin, G. J. S., and Brampton, M. H. (1996) Phase I trial of elactocin, *British Journal of Cancer 74*, 648-649.
[99] Walker, J. S., Hing, Z. A., Harrington, B., Baumhardt, J., Ozer, H. G., Lehman, A., Giacopelli, B., Beaver, L., Williams, K., Skinner, J. N., Cempre, C. B., Sun, Q., Shacham, S., Stromberg, B. R., Summers, M. K., Abruzzo, L. V., Rassenti, L., Kipps, T. J., Parikh, S., Kay, N. E., Rogers, K. A., Woyach, J. A., Coppola, V., Chook, Y. M., Oakes, C., Byrd, J. C., and Lapalombella, R. (2021) Recurrent XPO1 mutations alter pathogenesis of chronic lymphocytic leukemia, *Journal of Hematology & Oncology 14*, 17.
[100] Syed, Y. Y. (2019) Selinexor: First Global Approval, *Drugs 79*, 1485-1494.
[101] Kalakonda, N., Maerevoet, M., Cavallo, F., Follows, G., Goy, A., Vermaat, J. S. P., Casasnovas, O., Hamad, N., Zijlstra, J. M., Bakhshi, S., Bouabdallah, R., Choquet, S., Gurion, R., Hill, B., Jaeger, U., Sancho, J. M., Schuster, M., Thieblemont, C., De la Cruz, F., Egyed, M., Mishra, S., Offner, F., Vassilakopoulos, T. P., Warzocha, K., McCarthy, D., Ma, X., Corona, K., Saint-Martin, J.-R., Chang, H., Landesman, Y., Joshi, A., Wang, H., Shah, J.,

Shacham, S., Kauffman, M., Van Den Neste, E., and Canales, M. A. (2020) Selinexor in patients with relapsed or refractory diffuse large B-cell lymphoma (SADAL): a single-arm, multinational, multicentre, open-label, phase 2 trial, *The Lancet Haematology 7*, e511-e522.
[102] Sutanto, F., Konstantinidou, M., and Dömling, A. (2020) Covalent inhibitors: a rational approach to drug discovery, *RSC Medicinal Chemistry 11*, 876-884.
[103] Zhai, X., Ward, R. A., Doig, P., and Argyrou, A. (2020) Insight into the Therapeutic Selectivity of the Irreversible EGFR Tyrosine Kinase Inhibitor Osimertinib through Enzyme Kinetic Studies, *Biochemistry 59*, 1428-1441.
[104] Singh, J., Petter, R. C., Baillie, T. A., and Whitty, A. (2011) The resurgence of covalent drugs, *Nat Rev Drug Discov 10*, 307-317.
[105] Organization, P. A. H. (2017) Chagas in the Americas for Public Health Workers; 2017, Pan American Health Organization
[106] Organization, W. H. (2021) Chagas disease (American trypanosomiasis) Fact Sheet, World Health Organization
[107] Montgomery, S. P., Parise, M. E., Dotson, E. M., and Bialek, S. R. (2016) What Do We Know About Chagas Disease in the United States?, *Am J Trop Med Hyg 95*, 1225-1227.
[108] Meymandi, S., Hernandez, S., Park, S., Sanchez, D. R., and Forsyth, C. (2018) Treatment of Chagas Disease in the United States, *Curr Treat Options Infect Dis 10*, 373-388.
[109] Nunes Maria Carmo, P., Beaton, A., Acquatella, H., Bern, C., Bolger Ann, F., Echeverría Luis, E., Dutra Walderez, O., Gascon, J., Morillo Carlos, A., Oliveira-Filho, J., Ribeiro Antonio Luiz, P., Marin-Neto Jose, A., and null, n. (2018) Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association, *Circulation 138*, e169-e209.
[110] Gürtler, R. E., Kitron, U., Cecere, M. C., Segura, E. L., and Cohen, J. E. (2007) Sustainable vector control and management of Chagas disease in the Gran Chaco, Argentina, *Proceedings of the National Academy of Sciences 104*, 16194.
[111] Gaspe, M. S., Provecho, Y. M., Fernández, M. P., Vassena, C. V., Santo Orihuela, P. L., and Gürtler, R. E. (2018) Beating the odds: Sustained Chagas disease vector control in remote indigenous communities of the Argentine Chaco over a seven-year period, *PLOS Neglected Tropical Diseases 12*, e0006804.
[112] Carrilero, B., Murcia, L., Martinez-Lage, L., and Segovia, M. (2011) Side effects of benznidazole treatment in a cohort of patients with Chagas disease in non-endemic country, *Rev Esp Quimioter 24*, 123-126.
[113] Olivera, M. J., Cucunubá, Z. M., Álvarez, C. A., and Nicholls, R. S. (2015) Safety Profile of Nifurtimox and Treatment Interruption for Chronic Chagas Disease in Colombian Adults, *Am J Trop Med Hyg 93*, 1224-1230.
[114] Pérez-Molina, J. A., Pérez-Ayala, A., Moreno, S., Fernández-González, M. C., Zamora, J., and López-Velez, R. (2009) Use of benznidazole to treat chronic Chagas' disease: a systematic review with a meta-analysis, *Journal of Antimicrobial Chemotherapy 64*, 1139-1147.

[115] Jackson, Y., Alirol, E., Getaz, L., Wolff, H., Combescure, C., and Chappuis, F. (2010) Tolerance and Safety of Nifurtimox in Patients with Chronic Chagas Disease, *Clinical Infectious Diseases 51*, e69-e75.
[116] Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007) A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immunodeficient-Mouse Model of Infection, *Antimicrobial Agents and Chemotherapy 51*, 3932-3939.
[117] Tomás, A. M., and Kelly, J. M. (1996) Stage-regulated expression of cruzipain, the major cysteine protease of Trypanosoma cruzi is independent of the level of RNA, *Molecular and Biochemical Parasitology 76*, 91-103.
[118] Sajid, M., Robertson, S. A., Brinen, L. S., and McKerrow, J. H. (2011) Cruzain, In *Cysteine Proteases of Pathogenic Organisms* (Robinson, M. W., and Dalton, J. P., Eds.), pp 100-115, Springer US, Boston, MA.
[119] Culp, E., and Wright, G. D. (2017) Bacterial proteases, untapped antimicrobial drug targets, *The Journal of Antibiotics 70*, 366-377.
[120] McKerrow, J. H., Engel, J. C., and Caffrey, C. R. (1999) Cysteine protease inhibitors as chemotherapy for parasitic infections, *Bioorganic & Medicinal Chemistry 7*, 639-644.
[121] Wlodawer, A., and Vondrasek, J. (1998) INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.
[122] Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion, R., Jr., Nunneley, J. W., Barnard, D., Pöhlmann, S., McKerrow, J. H., Renslo, A. R., and Simmons, G. (2015) Protease inhibitors targeting coronavirus and filovirus entry, *Antiviral research 116*, 76-84.
[123] Mellott, D. M., Tseng, C.-T., Drelich, A., Fajtová, P., Chenna, B. C., Kostomiris, D. H., Hsu, J., Zhu, J., Taylor, Z. W., Kocurek, K. I., Tat, V., Katzfuss, A., Li, L., Giardini, M. A., Skinner, D., Hirata, K., Yoon, M. C., Beck, S., Carlin, A. F., Clark, A. E., Beretta, L., Maneval, D., Hook, V., Frueh, F., Hurst, B. L., Wang, H., Raushel, F. M., O'Donoghue, A. J., de Siqueira-Neto, J. L., Meek, T. D., and McKerrow, J. H. (2021) A Clinical-Stage Cysteine Protease Inhibitor blocks SARS-CoV-2 Infection of Human and Monkey Cells, *ACS Chemical Biology*.
[124] Zhai, X., and Meek, T. D. (2018) Catalytic Mechanism of Cruzain from Trypanosoma cruzi As Determined from Solvent Kinetic Isotope Effects of Steady-State and Pre-Steady-State Kinetics, *Biochemistry 57*, 3176-3190.
[125] Ábrányi-Balogh, P., Petri, L., Imre, T., Szijj, P., Scarpino, A., Hrast, M., Mitrović, A., Fonovič, U. P., Németh, K., Barreteau, H., Roper, D. I., Horváti, K., Ferenczy, G. G., Kos, J., Ilaš, J., Gobec, S., and Keserű, G. M. (2018) A road map for prioritizing warheads for cysteine targeting covalent inhibitors, *European Journal of Medicinal Chemistry 160*, 94-107.
[126] Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. , *Biochem Biophys Res Commun 425*, 497-502.
[127] Berger, A., and Schechter, I. (1970) Mapping the Active Site of Papain with the Aid of Peptide Substrates and Inhibitors, *Philosophical Transactions of the Royal Society B: Biological Sciences 257*, 249-264.

[128] Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012) Cysteine cathepsins: from structure, function and regulation to new frontiers, *Biochim Biophys Acta 1824*, 68-88.
[129] Turk, B. (2006) Targeting proteases: successes, failures and future prospects, *Nat Rev Drug Discov 5*, 785-799.
[130] Siklos, M., BenAissa, M., and Thatcher, G. R. J. (2015) Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors, *Acta Pharmaceutica Sinica B 5*, 506-519.
[131] Mott, B. T., Ferreira, R. S., Simeonov, A., Jadhav, A., Ang, K. K.-H., Leister, W., Shen, M., Silveira, J. T., Doyle, P. S., Arkin, M. R., McKerrow, J. H., Inglese, J., Austin, C. P., Thomas, C. J., Shoichet, B. K., and Maloney, D. J. (2010) Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB, *Journal of medicinal chemistry 53*, 52-60.
[132] Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Pandey, K. C., Caffrey, C. R., Legac, J., Hansell, E., McKerrow, J. H., Craik, C. S., Rosenthal, P. J., and Brinen, L. S. (2009) Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design, *Journal of Biological Chemistry 284*, 25697-25703.
[133] Ferreira, L. G., and Andricopulo, A. D. (2017) Targeting cysteine proteases in trypanosomatid disease drug discovery, *Pharmacol Ther 180*, 49-61.
[134] Boudreau, P. D., Miller, B. W., McCall, L.-I., Almaliti, J., Reher, R., Hirata, K., Le, T., Siqueira-Neto, J. L., Hook, V., and Gerwick, W. H. (2019) Design of Gallinamide A Analogs as Potent Inhibitors of the Cysteine Proteases Human Cathepsin L and Trypanosoma cruzi Cruzain, *Journal of Medicinal Chemistry 62*, 9026-9044.
[135] Wiggers, H. J., Rocha, J. R., Fernandes, W. B., Sesti-Costa, R., Carneiro, Z. A., Cheleski, J., da Silva, A. B. F., Juliano, L., Cezari, M. H. S., Silva, J. S., McKerrow, J. H., and Montanari, C. A. (2013) Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay, *PLOS Neglected Tropical Diseases 7*, e2370.
[136] Ferreira, R. S., Simeonov, A., Jadhav, A., Eidam, O., Mott, B. T., Keiser, M. J., McKerrow, J. H., Maloney, D. J., Irwin, J. J., and Shoichet, B. K. (2010) Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors, *Journal of medicinal chemistry 53*, 4891-4905.
[137] McKerrow, J. H. (2018) Update on drug development targeting parasite cysteine proteases, *PLOS Neglected Tropical Diseases 12*, e0005850.
[138] Mellott, D. M., Tseng, C.-T., Drelich, A., Fajtová, P., Chenna, B. C., Kostomiris, D. H., Hsu, J., Zhu, J., Taylor, Z. W., Kocurek, K. I., Tat, V., Katzfuss, A., Li, L., Giardini, M. A., Skinner, D., Hirata, K., Yoon, M. C., Beck, S., Carlin, A. F., Clark, A. E., Beretta, L., Maneval, D., Hook, V., Frueh, F., Hurst, B. L., Wang, H., Raushel, F. M., O'Donoghue, A. J., de Siqueira-Neto, J. L., Meek, T. D., and McKerrow, J. H. (2021) A Clinical-Stage Cysteine Protease Inhibitor blocks SARS-CoV-2 Infection of Human and Monkey Cells, *ACS Chemical Biology 16*, 642-650.

[139] Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proc Natl Acad Sci U S A 97*, 7754-7759.
[140] Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., Ellman, J. A., and Craik, C. S. (2006) Substrate Profiling of Cysteine Proteases Using a Combinatorial Peptide Library Identifies Functionally Unique Specificities, *Journal of Biological Chemistry 281*, 12824-12832.
[141] Polticelli, F., Zaini, G., Bolli, A., Antonini, G., Gradoni, L., and Ascenzi, P. (2005) Probing the Cruzain S2 Recognition Subsite: A Kinetic and Binding Energy Calculation Study, *Biochemistry 44*, 2781-2789.
[142] Chen, Y. T., Brinen, L. S., Kerr, I. D., Hansell, E., Doyle, P. S., McKerrow, J. H., and Roush, W. R. (2010) In vitro and in vivo studies of the trypanocidal properties of WRR-483 against Trypanosoma cruzi, *PLoS neglected tropical diseases 4*, e825.
[143] Jones, B. D., Tochowicz, A., Tang, Y., Cameron, M. D., McCall, L. I., Hirata, K., Siqueira-Neto, J. L., Reed, S. L., McKerrow, J. H., and Roush, W. R. (2016) Synthesis and Evaluation of Oxyguanidine Analogues of the Cysteine Protease Inhibitor WRR-483 against Cruzain, *ACS Med Chem Lett 7*, 77-82.
[144] Brady, K., and Abeles, R. H. (1990) Inhibition of chymotrypsin by peptidyl trifluoromethyl ketones: determinants of slow-binding kinetics, *Biochemistry 29*, 7608-7617.
[145] Rubach, J. K., Cui, G., Schneck, J. L., Taylor, A. N., Zhao, B., Smallwood, A., Nevins, N., Wisnoski, D., Thrall, S. H., and Meek, T. D. (2012) The Amino-Acid Substituents of Dipeptide Substrates of Cathepsin C Can Determine the Rate-Limiting Steps of Catalysis, *Biochemistry 51*, 7551-7568.
[146] Organization, W. H. (2020) *Global tuberculosis report 2020*.
[147] Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., and Vincent, V. (2005) Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis, *PLOS Pathogens 1*, e5.
[148] Kappelman, J., Alçiçek, M. C., Kazancı, N., Schultz, M., Özkul, M., and Şen, Ş. (2008) First Homo erectus from Turkey and implications for migrations into temperate Eurasia, *American Journal of Physical Anthropology 135*, 110-116.
[149] Bedeir, S. A. (2004) Tuberculosis in Ancient Egypt, In *Tuberculosis* (Madkour, M. M., Ed.), pp 3-13, Springer Berlin Heidelberg, Berlin, Heidelberg.
[150] Garg, R. K., and Somvanshi, D. S. (2011) Spinal tuberculosis: a review, *J Spinal Cord Med 34*, 440-454.
[151] Donoghue, H. D., Lee, O. Y. C., Minnikin, D. E., Besra, G. S., Taylor, J. H., and Spigelman, M. (2010) Tuberculosis in Dr Granville's mummy: a molecular re-examination of the earliest known Egyptian mummy to be scientifically examined and given a medical diagnosis, *Proc Biol Sci 277*, 51-56.
[152] Barberis, I., Bragazzi, N. L., Galluzzo, L., and Martini, M. (2017) The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus, *J Prev Med Hyg 58*, E9-E12.

[153] Daniel, T. M., Bates, J. H., and Downes, K. A. (1994) History of tuberculosis, *Tuberculosis: pathogenesis, protection, and control*, 13-24.
[154] Frith, J. (2014) History of tuberculosis. Part 1 - phthisis, consumption and the white plague, *Journal of Military and Veterans' Health 22*, 29-35.
[155] Sergio, S. (2004) Historical insights into tuberculosis,
Infez Med 2, 284-291.
[156] Dubos, R. J., and Dubos, J. (1987) *The white plague: tuberculosis, man, and society*, Rutgers University Press.
[157] Murray, J. F., Schraufnagel, D. E., and Hopewell, P. C. (2015) Treatment of Tuberculosis. A Historical Perspective, *Annals of the American Thoracic Society 12*, 1749-1759.
[158] Sakula, A. (1983) Robert koch: centenary of the discovery of the tubercle bacillus, 1882, *Can Vet J 24*, 127-131.
[159] Koch, R. (1884) Die aetiologie der tuberkulose.
[160] Robin, E. H., Maybelle, F. S., and John B. Bass, Jr. (1993) The Tuberculin Skin Test, *Clinical Infectious Diseases 17*, 968-975.
[161] Sakula, A. (1983) BCG: who were Calmette and Guérin?, *Thorax 38*, 806-812.
[162] Aronson, J. D., Aronson, C. F., and Taylor, H. C. (1958) A twenty-year appraisal of BCG vaccination in the control of tuberculosis, *AMA archives of internal medicine 101*, 881-893.
[163] Luca, S., and Mihaescu, T. (2013) History of BCG vaccine, *Maedica 8*, 53-58.
[164] Escobar, L. E., Molina-Cruz, A., and Barillas-Mury, C. (2020) BCG vaccine protection from severe coronavirus disease 2019 (COVID-19), *Proceedings of the National Academy of Sciences 117*, 17720.
[165] Wier, J. A. (1956) Streptomycin, Isoniazid, and Para-Aminosalicylic Acid in the Treatment of Pulmonary Tuberculosis, *American Review of Tuberculosis and Pulmonary Diseases 73*, 117-122.
[166] Frieden, T. R., Sterling, T., Pablos-Mendez, A., Kilburn, J. O., Cauthen, G. M., and Dooley, S. W. (1993) The Emergence of Drug-Resistant Tuberculosis in New York City, *New England Journal of Medicine 328*, 521-526.
[167] Nachega, J. B., and Chaisson, R. E. (2003) Tuberculosis Drug Resistance: A Global Threat, *Clinical Infectious Diseases 36*, S24-S30.
[168] Kwan, C. K., and Ernst, J. D. (2011) HIV and tuberculosis: a deadly human syndemic, *Clin Microbiol Rev 24*, 351-376.
[169] Jereb, J. A., Kelly, G. D., Dooley, S. W., Cauthen, G. M., and Snider, D. E. (1991) Tuberculosis Morbidity in the United States: Final Data, 1990, *Morbidity and Mortality Weekly Report: Surveillance Summaries 40*, 23-27.
[170] Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011) The challenge of new drug discovery for tuberculosis, *Nature 469*, 483-490.
[171] Mahajan, R. (2013) Bedaquiline: First FDA-approved tuberculosis drug in 40 years, *Int J Appl Basic Med Res 3*, 1-2.
[172] Guglielmetti, L., Jaspard, M., Le Dû, D., Lachâtre, M., Marigot-Outtandy, D., Bernard, C., Veziris, N., Robert, J., Yazdanpanah, Y., Caumes, E., and Fréchet-Jachym, M. (2017) Long-term outcome and safety of prolonged bedaquiline

treatment for multidrug-resistant tuberculosis, *European Respiratory Journal 49*, 1601799.
[173] Chen, J., and Raymond, K. (2006) Roles of rifampicin in drug-drug interactions: underlying molecular mechanisms involving the nuclear pregnane X receptor, *Ann Clin Microbiol Antimicrob 5*, 3-3.
[174] van Heeswijk, R. P. G., Dannemann, B., and Hoetelmans, R. M. W. (2014) Bedaquiline: a review of human pharmacokinetics and drug–drug interactions, *Journal of Antimicrobial Chemotherapy 69*, 2310-2318.
[175] Barry, C. E., 3rd. (2009) Unorthodox approach to the development of a new antituberculosis therapy, *N Engl J Med 360*, 2466-2467.
[176] Healan, A. M., Griffiss, J. M., Proskin, H. M., Riordan, M. A., Gray, W. A., Salata, R. A., and Blumer, J. L. (2018) Impact of Rifabutin or Rifampin on Bedaquiline Safety, Tolerability, and Pharmacokinetics Assessed in a Randomized Clinical Trial with Healthy Adult Volunteers, *Antimicrobial Agents and Chemotherapy 62*, e00855-00817.
[177] Martin, C. J., Carey, A. F., and Fortune, S. M. (2016) A bug's life in the granuloma, *Semin Immunopathol 38*, 213-220.
[178] Ramakrishnan, L. (2012) Revisiting the role of the granuloma in tuberculosis, *Nature Reviews Immunology 12*, 352-366.
[179] Ehrt, S., Schnappinger, D., and Rhee, K. Y. (2018) Metabolic principles of persistence and pathogenicity in Mycobacterium tuberculosis, *Nat Rev Microbiol 16*, 496-507.
[180] Russell, D. G., Barry, C. E., and Flynn, J. L. (2010) Tuberculosis: What We Don't Know Can, and Does, Hurt Us, *Science 328*, 852.
[181] Russell, D. G., Cardona, P. J., Kim, M. J., Allain, S., and Altare, F. (2009) Foamy macrophages and the progression of the human tuberculosis granuloma, *Nat Immunol 10*, 943-948.
[182] Weiss, G., and Schaible, U. E. (2015) Macrophage defense mechanisms against intracellular bacteria, *Immunol Rev 264*, 182-203.
[183] Upadhyay, S., Mittal, E., and Philips, J. A. (2018) Tuberculosis and the art of macrophage manipulation, *Pathogens and Disease 76*.
[184] Stallings, C. L., and Glickman, M. S. (2010) Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence, *Microbes Infect 12*, 1091-1101.
[185] Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffe, M., Emile, J. F., Marchou, B., Cardona, P. J., de Chastellier, C., and Altare, F. (2008) Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence, *PLoS Pathog 4*, e1000204.
[186] Russell, D. G. (2007) Who puts the tubercle in tuberculosis?, *Nature Reviews Microbiology 5*, 39-47.
[187] Kim, M.-J., Wainwright, H. C., Locketz, M., Bekker, L.-G., Walther, G. B., Dittrich, C., Visser, A., Wang, W., Hsu, F.-F., Wiehart, U., Tsenova, L., Kaplan, G., and Russell, D. G. (2010) Caseation of human tuberculosis granulomas

correlates with elevated host lipid metabolism, *EMBO Molecular Medicine 2*, 258-274.
[188] Elkington, P. T., D'Armiento, J. M., and Friedland, J. S. (2011) Tuberculosis immunopathology: the neglected role of extracellular matrix destruction, *Science translational medicine 3*, 71ps76-71ps76.
[189] Cadena, A. M., Fortune, S. M., and Flynn, J. L. (2017) Heterogeneity in tuberculosis, *Nature Reviews Immunology 17*, 691-702.
[190] Silva Miranda, M., Breiman, A., Allain, S., Deknuydt, F., and Altare, F. (2012) The Tuberculous Granuloma: An Unsuccessful Host Defence Mechanism Providing a Safety Shelter for the Bacteria?, *Clinical and Developmental Immunology 2012*, 139127.
[191] Russell, D. G., Huang, L., and VanderVen, B. C. (2019) Immunometabolism at the interface between macrophages and pathogens, *Nature Reviews Immunology 19*, 291-304.
[192] Bloch, H., and Segal, W. (1956) Biochemical differentiation of Mycobacterium tuberculosis grown in vivo and in vitro, *Journal of bacteriology 72*, 132-141.
[193] Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence, *Nature 393*, 537-544.
[194] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The Complete Genome Sequence of Escherichia coli K-12, *Science 277*, 1453.
[195] Riley, M. (1993) Functions of the gene products of Escherichia coli, *Microbiol Rev 57*, 862-952.
[196] Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria, *Proceedings of the National Academy of Sciences of the United States of America 98*, 12712-12717.
[197] Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis, *Molecular Microbiology 48*, 77-84.
[198] Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J., and Sassetti, C. M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism, *PLoS Pathog 7*, e1002251.
[199] Joshi, S. M., Pandey, A. K., Capite, N., Fortune, S. M., Rubin, E. J., and Sassetti, C. M. (2006) Characterization of mycobacterial virulence genes through genetic

interaction mapping, *Proceedings of the National Academy of Sciences 103*, 11760.
[200] Pandey, A. K., and Sassetti, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol, *Proceedings of the National Academy of Sciences 105*, 4376-4380.
[201] de Carvalho, L. P. S., Fischer, S. M., Marrero, J., Nathan, C., Ehrt, S., and Rhee, K. Y. (2010) Metabolomics of Mycobacterium tuberculosis Reveals Compartmentalized Co-Catabolism of Carbon Substrates, *Chemistry & Biology 17*, 1122-1131.
[202] Hartman, T. E., Wang, Z., Jansen, R. S., Gardete, S., and Rhee, K. Y. (2017) Metabolic Perspectives on Persistence, *Microbiol Spectr 5*.
[203] Eoh, H., and Rhee, K. Y. (2014) Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty acids, *Proceedings of the National Academy of Sciences 111*, 4976-4981.
[204] Eoh, H., and Rhee, K. Y. (2013) Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences 110*, 6554-6559.
[205] Nandakumar, M., Nathan, C., and Rhee, K. Y. (2014) Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis, *Nature Communications 5*, 4306.
[206] Beste, D. J. V., Bonde, B., Hawkins, N., Ward, J. L., Beale, M. H., Noack, S., Nöh, K., Kruger, N. J., Ratcliffe, R. G., and McFadden, J. (2011) 13C Metabolic Flux Analysis Identifies an Unusual Route for Pyruvate Dissimilation in Mycobacteria which Requires Isocitrate Lyase and Carbon Dioxide Fixation, *PLOS Pathogens 7*, e1002091.
[207] Höner Zu Bentrup, K., Miczak, A., Swenson, D. L., and Russell, D. G. (1999) Characterization of activity and expression of isocitrate lyase in Mycobacterium avium and Mycobacterium tuberculosis, *Journal of bacteriology 181*, 7161-7167.
[208] Carey, A. F., Rock, J. M., Krieger, I. V., Chase, M. R., Fernandez-Suarez, M., Gagneux, S., Sacchettini, J. C., Ioerger, T. R., and Fortune, S. M. (2018) TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities, *PLoS pathogens 14*, e1006939-e1006939.
[209] Antil, M., Sharma, J., Brissonnet, Y., Choudhary, M., Gouin, S., and Gupta, V. (2019) Structure-function insights into elusive Mycobacterium tuberculosis protein Rv1916, *International Journal of Biological Macromolecules 141*, 927-936.
[210] McFadden, B. A., Williams, J. O., and Roche, T. E. (1971) Mechanism of action of isocitrate lyase from Pseudomonas indigofera, *Biochemistry 10*, 1384-1390.
[211] Moynihan, M. M., and Murkin, A. S. (2014) Cysteine is the general base that serves in catalysis by isocitrate lyase and in mechanism-based inhibition by 3-nitropropionate, *Biochemistry 53*, 178-187.
[212] Lorenz, M. C., and Fink, G. R. (2002) Life and Death in a Macrophage: Role of the Glyoxylate Cycle in Virulence, *Eukaryotic Cell 1*, 657-662.

[213] Dolan, S. K., and Welch, M. (2018) The Glyoxylate Shunt, 60 Years On, *Annual Review of Microbiology 72*, 309-330.
[214] Bhusal, R. P., Jiao, W., Kwai, B. X. C., Reynisson, J., Collins, A. J., Sperry, J., Bashiri, G., and Leung, I. K. H. (2019) Acetyl-CoA-mediated activation of Mycobacterium tuberculosis isocitrate lyase 2, *Nature Communications 10*, 4639.
[215] Pham, T. V., Murkin, A. S., Moynihan, M. M., Harris, L., Tyler, P. C., Shetty, N., Sacchettini, J. C., Huang, H.-l., and Meek, T. D. (2017) Mechanism-based inactivator of isocitrate lyases 1 and 2 from Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences of the United States of America 114*, 7617-7622.
[216] McKinney, J. D., zu Bentrup, K. H., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W.-T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., and Russell, D. G. (2000) Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase, *Nature 406*, 735-738.
[217] Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., Dolganov, G., Efron, B., Butcher, P. D., Nathan, C., and Schoolnik, G. K. (2003) Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment, *J Exp Med 198*, 693-704.
[218] Munoz-Elias, E. J., and McKinney, J. D. (2005) Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence, *Nat Med 11*, 638-644.
[219] Dolan, S. K., Wijaya, A., Geddis, S. M., Spring, D. R., Silva-Rocha, R., and Welch, M. (2018) Loving the poison: the methylcitrate cycle and bacterial pathogenesis, *Microbiology 164*, 251-259.
[220] Muñoz-Elías, E. J., Upton, A. M., Cherian, J., and McKinney, J. D. (2006) Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence, *Molecular Microbiology 60*, 1109-1122.
[221] Gould, T. A., van de Langemheen, H., Munoz-Elias, E. J., McKinney, J. D., and Sacchettini, J. C. (2006) Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis, *Mol Microbiol 61*, 940-947.
[222] Beach, R. L., Aogaichi, T., and Plaut, G. W. (1977) Identification of D-threo-alpha-methylisocitrate as stereochemically specific substrate for bovine heart aconitase and inhibitor of TPN-linked isocitrate dehydrogenase, *Journal of Biological Chemistry 252*, 2702-2709.
[223] Griffin, Jennifer E., Pandey, Amit K., Gilmore, Sarah A., Mizrahi, V., McKinney, John D., Bertozzi, Carolyn R., and Sassetti, Christopher M. (2012) Cholesterol Catabolism by Mycobacterium tuberculosis Requires Transcriptional and Metabolic Adaptations, *Chemistry & Biology 19*, 218-227.
[224] Smith, R. A., and Gunsalus, I. C. (1954) ISOCITRITASE: A NEW TRICARBOXYLIC ACID CLEAVAGE SYSTEM, *Journal of the American Chemical Society 76*, 5002-5003.
[225] Sharma, V., Sharma, S., zu Bentrup, K. H., McKinney, J. D., Russell, D. G., Jacobs, W. R., and Sacchettini, J. C. (2000) Structure of isocitrate lyase, a

persistence factor of Mycobacterium tuberculosis, *Nat Struct Mol Biol 7*, 663-668.
[226] Britton, K. L., Langridge, S. J., Baker, P. J., Weeradechapon, K., Sedelnikova, S. E., De Lucas, J. R., Rice, D. W., and Turner, G. (2000) The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans, *Structure 8*, 349-362.
[227] Huang, K., Li, Z., Jia, Y., Dunaway-Mariano, D., and Herzberg, O. (1999) Helix swapping between two α/β barrels: crystal structure of phosphoenolpyruvate mutase with bound Mg2+–oxalate, *Structure 7*, 539-548.
[228] Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) 3D domain swapping: a mechanism for oligomer assembly, *Protein Sci 4*, 2455-2468.
[229] Alston, T. A., Mela, L., and Bright, H. J. (1977) 3-Nitropropionate, the toxic substance of Indigofera, is a suicide inactivator of succinate dehydrogenase, *Proceedings of the National Academy of Sciences of the United States of America 74*, 3767-3771.
[230] Ganapathy-Kanniappan, S., Geschwind, J.-F. H., Kunjithapatham, R., Buijs, M., Vossen, J. A., Tchernyshyov, I., Cole, R. N., Syed, L. H., Rao, P. P., Ota, S., and Vali, M. (2009) Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is pyruvylated during 3-bromopyruvate mediated cancer cell death, *Anticancer Res 29*, 4909-4918.
[231] Quartararo, C. E., Hadi, T., Cahill, S. M., and Blanchard, J. S. (2013) Solvent isotope-induced equilibrium perturbation for isocitrate lyase, *Biochemistry 52*, 9286-9293.
[232] Giachetti, E., Pinzauti, G., Bonaccorsi, R., and Vanni, P. (1988) Isocitrate lyase from Pinus pinea, *European Journal of Biochemistry 172*, 85-91.
[233] Ruá, J., Soler, J., Busto, F., and de Arriaga, D. (1997) Effect of pH on the role of Mg2+ and Mn2+ on Phycomyces isocitrate lyase kinetics, *Biochimie 79*, 179-186.
[234] Ray, S., Kreitler, D. F., Gulick, A. M., and Murkin, A. S. (2018) The Nitro Group as a Masked Electrophile in Covalent Enzyme Inhibition, *ACS chemical biology 13*, 1470-1473.
[235] Ko, Y. H., and McFadden, B. A. (1990) Alkylation of isocitrate lyase from Escherichia coli by 3-bromopyruvate, *Archives of Biochemistry and Biophysics 278*, 373-380.
[236] Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of Escherichia coli ASKA library (A Complete S et of E. coli K -12 ORF A rchive): Unique Resources for Biological Research, *DNA Research 12*, 291-299.
[237] Parker, R. E., and Isaacs, N. S. (1959) Mechanisms Of Epoxide Reactions, *Chemical Reviews 59*, 737-799.
[238] Kóňa, J. (2008) Theoretical study on the mechanism of a ring-opening reaction of oxirane by the active-site aspartic dyad of HIV-1 protease, *Organic & Biomolecular Chemistry 6*, 359-365.

[239] Meara, J. P., and Rich, D. H. (1996) Mechanistic Studies on the Inactivation of Papain by Epoxysuccinyl Inhibitors, *Journal of Medicinal Chemistry 39*, 3357-3366.
[240] Mladenovic, M., Schirmeister, T., Thiel, S., Thiel, W., and Engels, B. (2007) The Importance of the Active Site Histidine for the Activity of Epoxide- or Aziridine-Based Inhibitors of Cysteine Proteases, *ChemMedChem 2*, 120-128.
[241] Bihovsky, R. (1992) Reactions of .alpha.,.beta.-epoxy carbonyl compounds with methanethiolate: regioselectivity and rate, *The Journal of Organic Chemistry 57*, 1029-1031.
[242] Quartararo, C. E., Hadi, T., Cahill, S. M., and Blanchard, J. S. (2013) Solvent Isotope-Induced Equilibrium Perturbation for Isocitrate Lyase, *Biochemistry 52*, 9286-9293.
[243] Gengenbacher, M., and Kaufmann, S. H. E. (2012) Mycobacterium tuberculosis: success through dormancy, *FEMS Microbiology Reviews 36*, 514-532.
[244] O'Garra, A., Redford, P. S., McNab, F. W., Bloom, C. I., Wilkinson, R. J., and Berry, M. P. (2013) The immune response in tuberculosis, *Annu Rev Immunol 31*, 475-527.
[245] Russell, D. G., VanderVen, B. C., Lee, W., Abramovitch, R. B., Kim, M., Homolka, S., Niemann, S., and Rohde, K. H. (2010) Mycobacterium tuberculosis wears what it eats, *Cell host & microbe 8*, 68-76.
[246] Daniel, J., Maamar, H., Deb, C., Sirakova, T. D., and Kolattukudy, P. E. (2011) Mycobacterium tuberculosis Uses Host Triacylglycerol to Accumulate Lipid Droplets and Acquires a Dormancy-Like Phenotype in Lipid-Loaded Macrophages, *PLOS Pathogens 7*, e1002093.
[247] Schloss, J. V., and Cleland, W. W. (1982) Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analog, *Biochemistry 21*, 4420-4427.
[248] Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K., Benesch, J. L., and Robinson, C. V. (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles, *Anal Chem 87*, 4370-4376.
[249] Otwinowski, Z., and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode, In *Methods in Enzymology*, pp 307-326, Academic Press.
[250] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.
[251] Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.
[252] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—A visualization system

for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.
[253] Porter, D. J., and Bright, H. J. (1980) 3-Carbanionic substrate analogues bind very tightly to fumarase and aspartase, *Journal of Biological Chemistry 255*, 4772-4780.
[254] Porter, D. J. T., Voet, J. G., and Bright, H. J. (1973) Direct Evidence for Carbanions and Covalent N5-Flavin-Carbanion Adducts as Catalytic Intermediates in the Oxidation of Nitroethane by d-Amino Acid Oxidase, *Journal of Biological Chemistry 248*, 4400-4416.
[255] Alston, T. A., Porter, D. J. T., and Bright, H. J. (1983) Enzyme inhibition by nitro and nitroso compounds, *Accounts of Chemical Research 16*, 418-424.
[256] Eaton, J. K., Furst, L., Ruberto, R. A., Moosmayer, D., Hilpmann, A., Ryan, M. J., Zimmermann, K., Cai, L. L., Niehues, M., Badock, V., Kramm, A., Chen, S., Hillig, R. C., Clemons, P. A., Gradl, S., Montagnon, C., Lazarski, K. E., Christian, S., Bajrami, B., Neuhaus, R., Eheim, A. L., Viswanathan, V. S., and Schreiber, S. L. (2020) Selective covalent targeting of GPX4 using masked nitrile-oxide electrophiles, *Nature Chemical Biology 16*, 497-506.
[257] Neres, J., Pojer, F., Molteni, E., Chiarelli, L. R., Dhar, N., Boy-Röttger, S., Buroni, S., Fullam, E., Degiacomi, G., Lucarelli, A. P., Read, R. J., Zanoni, G., Edmondson, D. E., De Rossi, E., Pasca, M. R., McKinney, J. D., Dyson, P. J., Riccardi, G., Mattevi, A., Cole, S. T., and Binda, C. (2012) Structural Basis for Benzothiazinone-Mediated Killing of <em>Mycobacterium tuberculosis</em&gt, *Science Translational Medicine 4*, 150ra121.
[258] Tallman, K. R., Levine, S. R., and Beatty, K. E. (2016) Small-Molecule Probes Reveal Esterases with Persistent Activity in Dormant and Reactivating Mycobacterium tuberculosis, *ACS Infectious Diseases 2*, 936-944.
[259] Tallman, K. R., Levine, S. R., and Beatty, K. E. (2016) Profiling Esterases in Mycobacterium tuberculosis Using Far-Red Fluorogenic Substrates, *ACS Chemical Biology 11*, 1810-1815.
[260] Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y., Ning, Z. B., Zeng, R., Xiong, Y., Guan, K. L., Zhao, S., and Zhao, G. P. (2010) Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux, *Science 327*, 1004-1007.
[261] Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling, *Nat Rev Mol Cell Biol 15*, 536-550.
[262] Wagner, G. R., Bhatt, D. P., O'Connell, T. M., Thompson, J. W., Dubois, L. G., Backos, D. S., Yang, H., Mitchell, G. A., Ilkayeva, O. R., Stevens, R. D., Grimsrud, P. A., and Hirschey, M. D. (2017) A Class of Reactive Acyl-CoA Species Reveals the Non-enzymatic Origins of Protein Acylation, *Cell Metabolism 25*, 823-837.e828.
[263] Haberland, M., Montgomery, R. L., and Olson, E. N. (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy, *Nature Reviews Genetics 10*, 32-42.

[264] Rajendran, A., Vaidya, K., Mendoza, J., Bridwell-Rabb, J., and Kamat, S. S. (2019) Functional Annotation of ABHD14B, an Orphan Serine Hydrolase Enzyme, *Biochemistry*.
[265] Bheda, P., Jing, H., Wolberger, C., and Lin, H. (2016) The Substrate Specificity of Sirtuins, *Annual Review of Biochemistry 85*, 405-429.
[266] Haigis, M. C., and Sinclair, D. A. (2010) Mammalian Sirtuins: Biological Insights and Disease Relevance, *Annual Review of Pathology: Mechanisms of Disease 5*, 253-295.
[267] Ali, I., Conrad, R. J., Verdin, E., and Ott, M. (2018) Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics, *Chemical Reviews 118*, 1216-1252.
[268] Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S., and Verdin, E. (2006) Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2, *Proceedings of the National Academy of Sciences 103*, 10224.
[269] Rajabi, N., Galleano, I., Madsen, A. S., and Olsen, C. A. (2018) Chapter Two - Targeting Sirtuins: Substrate Specificity and Inhibitor Design, In *Progress in Molecular Biology and Translational Science* (Zheng, W., Ed.), pp 25-69, Academic Press.
[270] Wang, Z. A., and Cole, P. A. (2020) The Chemical Biology of Reversible Lysine Post-translational Modifications, *Cell Chemical Biology 27*, 953-969.
[271] Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019) Protein post-translational modifications in bacteria, *Nature Reviews Microbiology*.
[272] Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., and Wolfe, A. J. (2019) Post-translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions, *Front Microbiol 10*, 1604-1604.
[273] Yang, M., Wang, Y., Chen, Y., Cheng, Z., Gu, J., Deng, J., Bi, L., Chen, C., Mo, R., Wang, X., and Ge, F. (2015) Succinylome Analysis Reveals the Involvement of Lysine Succinylation in Metabolism in Pathogenic Mycobacterium tuberculosis, *Molecular & Cellular Proteomics 14*, 796-811.
[274] Zhou, M., Xie, L., Yang, Z., Zhou, J., and Xie, J. (2016) Lysine succinylation of Mycobacterium tuberculosis isocitrate lyase (ICL) fine-tunes the microbial resistance to antibiotics, *J Biomol Struct Dyn*, 1-12.
[275] Xie, L., Wang, X., Zeng, J., Zhou, M., Duan, X., Li, Q., Zhang, Z., Luo, H., Pang, L., Li, W., Liao, G., Yu, X., Li, Y., Huang, H., and Xie, J. (2015) Proteome-wide lysine acetylation profiling of the human pathogen Mycobacterium tuberculosis, *The International Journal of Biochemistry & Cell Biology 59*, 193-202.
[276] Xie, L., Wang, G., Yu, Z., Zhou, M., Li, Q., Huang, H., and Xie, J. (2016) Proteome-wide Lysine Glutarylation Profiling of the Mycobacterium tuberculosis H37Rv.

[277] Zhao, K., Chai, X., and Marmorstein, R. (2004) Structure and Substrate Binding Properties of cobB, a Sir2 Homolog Protein Deacetylase from Escherichia coli, *Journal of Molecular Biology 337*, 731-741.
[278] Colak, G., Xie, Z., Zhu, A. Y., Dai, L., Lu, Z., Zhang, Y., Wan, X., Chen, Y., Cha, Y. H., Lin, H., Zhao, Y., and Tan, M. (2013) Identification of lysine succinylation substrates and the succinylation regulatory enzyme CobB in Escherichia coli, *Molecular & cellular proteomics : MCP 12*, 3509-3520.
[279] Rowland, E. A., Greco, T. M., Snowden, C. K., McCabe, A. L., Silhavy, T. J., and Cristea, I. M. (2017) Sirtuin Lipoamidase Activity Is Conserved in Bacteria as a Regulator of Metabolic Enzyme Complexes, *mBio 8*, e01096-01017.
[280] Dong, H., Zhai, G., Chen, C., Bai, X., Tian, S., Hu, D., Fan, E., and Zhang, K. (2019) Protein lysine de-2-hydroxyisobutyrylation by CobB in prokaryotes, *Science Advances 5*, eaaw6703.
[281] Rittershaus, E. S. C., Baek, S.-H., Krieger, I. V., Nelson, S. J., Cheng, Y.-S., Nambi, S., Baker, R. E., Leszyk, J. D., Shaffer, S. A., Sacchettini, J. C., and Sassetti, C. M. (2018) A Lysine Acetyltransferase Contributes to the Metabolic Adaptation to Hypoxia in Mycobacterium tuberculosis, *Cell Chemical Biology 25*, 1495-1505.e1493.
[282] Sassetti, C. M., and Rubin, E. J. (2003) Genetic requirements for mycobacterial survival during infection, *Proceedings of the National Academy of Sciences of the United States of America 100*, 12989-12994.
[283] DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., Rubin, E. J., Schnappinger, D., Ehrt, S., Fortune, S. M., Sassetti, C. M., and Ioerger, T. R. (2017) Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis, *mBio 8*, e02133-02116.
[284] Heltweg, B., Dequiedt, F., Marshall, B. L., Brauch, C., Yoshida, M., Nishino, N., Verdin, E., and Jung, M. (2004) Subtype Selective Substrates for Histone Deacetylases, *Journal of Medicinal Chemistry 47*, 5235-5243.
[285] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nature protocols 10*, 845-858.
[286] Spinck, M., Neumann-Staubitz, P., Ecke, M., Gasper, R., and Neumann, H. (2020) Evolved, Selective Erasers of Distinct Lysine Acylations, *Angewandte Chemie International Edition 59*, 11142-11149.
[287] Du, J., Zhou, Y., Su, X., Yu, J. J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J. H., Choi, B. H., He, B., Chen, W., Zhang, S., Cerione, R. A., Auwerx, J., Hao, Q., and Lin, H. (2011) Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase, *Science (New York, N.Y.) 334*, 806-809.
[288] Feldman, J. L., Baeza, J., and Denu, J. M. (2013) Activation of the Protein Deacetylase SIRT6 by Long-chain Fatty Acids and Widespread Deacylation by Mammalian Sirtuins, *Journal of Biological Chemistry 288*, 31350-31356.

[289] Ringel, A. E., Roman, C., and Wolberger, C. (2014) Alternate deacylating specificities of the archaeal sirtuins Sir2Af1 and Sir2Af2, *Protein science : a publication of the Protein Society 23*, 1686-1697.
[290] Gu, J., Deng, J.-Y., Li, R., Wei, H., Zhang, Z., Zhou, Y., Zhang, Y., and Zhang, X.-E. (2009) Cloning and characterization of NAD-dependent protein deacetylase (Rv1151c) from Mycobacterium tuberculosis, *Biochemistry (Moscow) 74*, 743-748.
[291] Bonds, A. C., Yuan, T., Werman, J. M., Jang, J., Lu, R., Nesbitt, N. M., Garcia-Diaz, M., and Sampson, N. S. (2020) Post-translational Succinylation of Mycobacterium tuberculosis Enoyl-CoA Hydratase EchA19 Slows Catalytic Hydration of Cholesterol Catabolite 3-Oxo-chol-4,22-diene-24-oyl-CoA, *ACS Infectious Diseases 6*, 2214-2224.
[292] Xu, H., Hegde, S. S., and Blanchard, J. S. (2011) Reversible acetylation and inactivation of Mycobacterium tuberculosis acetyl-CoA synthetase is dependent on cAMP, *Biochemistry 50*, 5883-5892.
[293] Smith, B. C., Hallows, W. C., and Denu, J. M. (2009) A continuous microplate assay for sirtuins and nicotinamide-producing enzymes, *Analytical biochemistry 394*, 101-109.
[294] Rajabi, N., Auth, M., Troelsen, K. R., Pannek, M., Bhatt, D. P., Fontenas, M., Hirschey, M. D., Steegborn, C., Madsen, A. S., and Olsen, C. A. (2017) Mechanism-Based Inhibitors of the Human Sirtuin 5 Deacylase: Structure-Activity Relationship, Biostructural, and Kinetic Insight, *Angewandte Chemie (International ed. in English) 56*, 14836-14841.
[295] Zhou, Y., Zhang, H., He, B., Du, J., Lin, H., Cerione, R. A., and Hao, Q. (2012) The bicyclic intermediate structure provides insights into the desuccinylation mechanism of human sirtuin 5 (SIRT5), *The Journal of biological chemistry 287*, 28307-28314.
[296] Hawse, W. F., Hoff, K. G., Fatkins, D. G., Daines, A., Zubkova, O. V., Schramm, V. L., Zheng, W., and Wolberger, C. (2008) Structural insights into intermediate steps in the Sir2 deacetylation reaction, *Structure (London, England : 1993) 16*, 1368-1377.
[297] Smith, B. C., and Denu, J. M. (2007) Mechanism-Based Inhibition of Sir2 Deacetylases by Thioacetyl-Lysine Peptide, *Biochemistry 46*, 14478-14486.
[298] Janczewski, Ł., Gajda, A., Frankowski, S., Goszczyński, T. M., and Gajda, T. (2018) T3P® – A Benign Desulfurating Reagent in the Synthesis of Isothiocyanates, *Synthesis 50*, 1141-1151.
[299] Farooqi, A. S., Hong, J. Y., Cao, J., Lu, X., Price, I. R., Zhao, Q., Kosciuk, T., Yang, M., Bai, J. J., and Lin, H. (2019) Novel Lysine-Based Thioureas as Mechanism-Based Inhibitors of Sirtuin 2 (SIRT2) with Anticancer Activity in a Colorectal Cancer Murine Model, *Journal of Medicinal Chemistry 62*, 4131-4141.
[300] Rajabi, N., Nielsen, A. L., and Olsen, C. A. (2020) Dethioacylation by Sirtuins 1–3: Considerations for Drug Design Using Mechanism-Based Sirtuin Inhibition, *ACS Medicinal Chemistry Letters 11*, 1886-1892.

[301] Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature 579*, 270-273.
[302] Morse, J. S., Lalonde, T., Xu, S., and Liu, W. R. (2020) Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV, *ChemBioChem 21*, 730-738.
[303] Dong, E., Du, H., and Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time, *The Lancet Infectious Diseases 20*, 533-534.
[304] Times, N. Y. (2020) Coronavirus in the U.S.: Latest Map and Case Count.
[305] Pillaiyar, T., Meenakshisundaram, S., and Manickam, M. (2020) Recent discovery and development of inhibitors targeting coronaviruses, *Drug discovery today 25*, 668-688.
[306] Guy, R. K., DiPaola, R. S., Romanelli, F., and Dutch, R. E. (2020) Rapid repurposing of drugs for COVID-19, *Science 368*, 829.
[307] Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M.-h., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C. T., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Günther, S., Osterhaus, A. D. M. E., Drosten, C., Pallansch, M. A., Anderson, L. J., and Bellini, W. J. (2003) Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome, *Science 300*, 1394.
[308] Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia, *New England Journal of Medicine 367*, 1814-1820.
[309] Báez-Santos, Y. M., St John, S. E., and Mesecar, A. D. (2015) The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds, *Antiviral research 115*, 21-38.
[310] Lv, L., Li, G., Chen, J., Liang, X., and Li, Y. (2020) Comparative genomic analysis revealed specific mutation pattern between human coronavirus SARS-CoV-2 and Bat-SARSr-CoV RaTG13, *bioRxiv*, 2020.2002.2027.969006.
[311] Heald-Sargent, T., and Gallagher, T. (2012) Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence, *Viruses 4*, 557-580.
[312] Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., and Qian, Z. (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, *Nature Communications 11*, 1620.
[313] Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, *Proceedings of the National Academy of Sciences 106*, 5871.

[314] Huang, Y., Yang, C., Xu, X.-f., Xu, W., and Liu, S.-w. (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19, *Acta Pharmacologica Sinica 41*, 1141-1149.
[315] Millet, J. K., and Whittaker, G. R. (2014) Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein, *Proceedings of the National Academy of Sciences 111*, 15214.
[316] Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry, *Proceedings of the National Academy of Sciences of the United States of America 102*, 11876.
[317] Bosch, B. J., Bartelink, W., and Rottier, P. J. M. (2008) Cathepsin L Functionally Cleaves the Severe Acute Respiratory Syndrome Coronavirus Class I Fusion Protein Upstream of Rather than Adjacent to the Fusion Peptide, *Journal of Virology 82*, 8887.
[318] Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., Burgstaller-Muehlbacher, S., De Jesus, P. D., Teriete, P., Hull, M. V., Chang, M. W., Chan, J. F.-W., Cao, J., Poon, V. K.-M., Herbert, K. M., Cheng, K., Nguyen, T.-T. H., Rubanov, A., Pu, Y., Nguyen, C., Choi, A., Rathnasinghe, R., Schotsaert, M., Miorin, L., Dejosez, M., Zwaka, T. P., Sit, K.-Y., Martinez-Sobrido, L., Liu, W.-C., White, K. M., Chapman, M. E., Lendy, E. K., Glynne, R. J., Albrecht, R., Ruppin, E., Mesecar, A. D., Johnson, J. R., Benner, C., Sun, R., Schultz, P. G., Su, A. I., García-Sastre, A., Chatterjee, A. K., Yuen, K.-Y., and Chanda, S. K. (2020) Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing, *Nature*.
[319] Palmer, J. T., Rasnick, D., Klaus, J. L., and Bromme, D. (1995) Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors, *Journal of Medicinal Chemistry 38*, 3193-3196.
[320] Engel, J. C., Doyle, P. S., Hsieh, I., and McKerrow, J. H. (1998) Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection, *J Exp Med 188*, 725-734.
[321] Yang, P.-Y., Wang, M., He, C. Y., and Yao, S. Q. (2012) Proteomic profiling and potential cellular target identification of K11777, a clinical cysteine protease inhibitor, in Trypanosoma brucei, *Chemical Communications 48*, 835-837.
[322] Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007) A Cysteine Protease Inhibitor Cures Chagas Disease in an Immunodeficient-Mouse Model of Infection, *Antimicrobial Agents and Chemotherapy 51*, 3932.
[323] Barr, S. C., Warner, K. L., Kornreic, B. G., Piscitelli, J., Wolfe, A., Benet, L., and McKerrow, J. H. (2005) A cysteine protease inhibitor protects dogs from cardiac damage during infection by Trypanosoma cruzi, *Antimicrobial agents and chemotherapy 49*, 5160-5161.
[324] Abdulla, M.-H., Lim, K.-C., Sajid, M., McKerrow, J. H., and Caffrey, C. R. (2007) Schistosomiasis Mansoni: Novel Chemotherapy Using a Cysteine Protease Inhibitor, *PLOS Medicine 4*, e14.

[325] Ndao, M., Nath-Chowdhury, M., Sajid, M., Marcus, V., Mashiyama, S. T., Sakanari, J., Chow, E., Mackey, Z., Land, K. M., Jacobson, M. P., Kalyanaraman, C., McKerrow, J. H., Arrowood, M. J., and Caffrey, C. R. (2013) A Cysteine Protease Inhibitor Rescues Mice from a Lethal Cryptosporidium parvum Infection, *Antimicrobial Agents and Chemotherapy 57*, 6063.
[326] McKerrow, J. H. (2018) Update on drug development targeting parasite cysteine proteases, *PLoS neglected tropical diseases 12*, e0005850-e0005850.
[327] C. D. Owen, C. D., Lukacik, P., Strain-Damerell, C. M., Douangamath, A., Powell, A. J., Fearon, D., Brandao-Neto, J., Crawshaw, A. D., Aragao, D., Williams, M., Flaig, R., Hall, D. R., McAuley, K. E., Mazzorana, M., Stuart, D. I., Delft, F. v., and Walsh, M. A. (2020) 6Y84, SARS-CoV-2 main protease with unliganded active site (2019-nCoV, coronavirus disease 2019, COVID-19), RCSB Protein Data Bank, RCSB Protein Data Bank.
[328] Chen, L., Liu, W., Zhang, Q., Xu, K., Ye, G., Wu, W., Sun, Z., Liu, F., Wu, K., Zhong, B., Mei, Y., Zhang, W., Chen, Y., Li, Y., Shi, M., Lan, K., and Liu, Y. (2020) RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak, *Emerging Microbes & Infections 9*, 313-319.
[329] Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M., and Rao, Z. (2007) Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction, *Journal of molecular biology 366*, 965-975.
[330] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis, *Nature Methods 9*, 671-675.
[331] Yoshikawa, T., Hill, T. E., Yoshikawa, N., Popov, V. L., Galindo, C. L., Garner, H. R., Peters, C. J., and Tseng, C.-T. (2010) Dynamic Innate Immune Responses of Human Bronchial Epithelial Cells to Severe Acute Respiratory Syndrome-Associated Coronavirus Infection, *PLOS ONE 5*, e8729.
[332] Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., Straub, J. H., Stürzel, C. M., Fröhlich, T., Berninghausen, O., Becker, T., Kirchhoff, F., Sparrer, K. M. J., and Beckmann, R. (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2, *Science 369*, 1249.
[333] Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Molecular & cellular proteomics : MCP 13*, 2513-2526.
[334] Huang, I. C., Bosch, B. J., Li, F., Li, W., Lee, K. H., Ghiran, S., Vasilieva, N., Dermody, T. S., Harrison, S. C., Dormitzer, P. R., Farzan, M., Rottier, P. J. M., and Choe, H. (2006) SARS Coronavirus, but Not Human Coronavirus NL63, Utilizes Cathepsin L to Infect ACE2-expressing Cells, *Journal of Biological Chemistry 281*, 3198-3203.

[335] Poreba, M., Groborz, K., Vizovisek, M., Maruggi, M., Turk, D., Turk, B., Powis, G., Drag, M., and Salvesen, G. S. (2019) Fluorescent probes towards selective cathepsin B detection and visualization in cancer cells and patient samples, *Chemical Science 10*, 8461-8477.
[336] Jaimes, J. A., Millet, J. K., and Whittaker, G. R. (2020) Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site, *iScience 23*, 101212.
[337] Biniossek, M. L., Nägler, D. K., Becker-Pauly, C., and Schilling, O. (2011) Proteomic Identification of Protease Cleavage Sites Characterizes Prime and Non-prime Specificity of Cysteine Cathepsins B, L, and S, *Journal of Proteome Research 10*, 5363-5373.
[338] Zhu, Y., Feng, F., Hu, G., Wang, Y., Yu, Y., Zhu, Y., Xu, W., Cai, X., Sun, Z., Han, W., Ye, R., Chen, H., Ding, Q., Cai, Q., Qu, D., Xie, Y., Yuan, Z., and Zhang, R. (2020) The S1/S2 boundary of SARS-CoV-2 spike protein modulates cell entry pathways and transmission, *bioRxiv*, 2020.2008.2025.266775.
[339] Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., and Whittaker, G. R. (2020) Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop, *Journal of molecular biology 432*, 3309-3325.
[340] Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell 181*, 281-292.e286.
[341] Park, J.-E., Li, K., Barlan, A., Fehr, A. R., Perlman, S., McCray, P. B., and Gallagher, T. (2016) Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism, *Proceedings of the National Academy of Sciences 113*, 12262.
[342] Laporte, M., Stevaert, A., Raeymaekers, V., Van Berwaer, R., Martens, K., Pöhlmann, S., and Naesens, L. (2020) The SARS-CoV-2 and other human coronavirus spike proteins are fine-tuned towards temperature and proteases of the human airways, *bioRxiv*, 2020.2011.2009.374603.
[343] Wang, R., Simoneau, C. R., Kulsuptrakul, J., Bouhaddou, M., Travisano, K. A., Hayashi, J. M., Carlson-Stevermer, J., Zengel, J. R., Richards, C. M., Fozouni, P., Oki, J., Rodriguez, L., Joehnk, B., Walcott, K., Holden, K., Sil, A., Carette, J. E., Krogan, N. J., Ott, M., and Puschnik, A. S. (2020) Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses, *Cell*.
[344] Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., Strine, M. S., Zhang, S.-M., Graziano, V. R., Schmitz, C. O., Chen, J. S., Mankowski, M. C., Filler, R. B., Ravindra, N. G., Gasque, V., de Miguel, F. J., Patil, A., Chen, H., Oguntuyo, K. Y., Abriola, L., Surovtseva, Y. V., Orchard, R. C., Lee, B., Lindenbach, B. D., Politi, K., van Dijk, D., Kadoch, C., Simon, M. D., Yan, Q., Doench, J. G., and Wilen, C. B. (2020) Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection, *Cell*, S0092-8674(0020)31392-31391.

[345] Zhao, M.-M., Yang, W.-L., Yang, F.-Y., Zhang, L., Huang, W., Hou, W., Fan, C., Jin, R., Feng, Y., Wang, Y., and Yang, J.-K. (2020) Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development, *medRxiv*, 2020.2010.2025.20218990.
[346] Wu, M., Chen, Y., Xia, H., Wang, C., Tan, C. Y., Cai, X., Liu, Y., Ji, F., Xiong, P., Liu, R., Guan, Y., Duan, Y., Kuang, D., Xu, S., Cai, H., Xia, Q., Yang, D., Wang, M.-W., Chiu, I. M., Cheng, C., Ahern, P. P., Liu, L., Wang, G., Surana, N. K., Xia, T., and Kasper, D. L. (2020) Transcriptional and proteomic insights into the host response in fatal COVID-19 cases, *Proceedings of the National Academy of Sciences 117*, 28336.
[347] Liu, T., Luo, S., Libby, P., and Shi, G.-P. (2020) Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients, *Pharmacol Ther 213*, 107587.
[348] Poreba, M., Mihelic, M., Krai, P., Rajkovic, J., Krezel, A., Pawelczak, M., Klemba, M., Turk, D., Turk, B., Latajka, R., and Drag, M. (2014) Unnatural amino acids increase activity and specificity of synthetic substrates for human and malarial cathepsin C, *Amino Acids 46*, 931-943.
[349] Poreba, M., Kasperkiewicz, P., Snipas, S. J., Fasci, D., Salvesen, G. S., and Drag, M. (2014) Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates, *Cell Death & Differentiation 21*, 1482-1492.
[350] Ballatore, C., Huryn, D. M., and Smith, A. B., 3rd. (2013) Carboxylic acid (bio)isosteres in drug design, *ChemMedChem 8*, 385-395.
[351] Jameel, S., El-Gul, T., and McFadden, B. A. (1985) Modification of the active site of isocitrate lyase from watermelon cotyledons, *Archives of Biochemistry and Biophysics 236*, 72-81.
[352] Kosono, S., Tamura, M., Suzuki, S., Kawamura, Y., Yoshida, A., Nishiyama, M., and Yoshida, M. (2015) Changes in the Acetylome and Succinylome of Bacillus subtilis in Response to Carbon Source, *PLOS ONE 10*, e0131169.
[353] Seth, S., Batra, J., and Srinivasan, S. (2020) COVID-19: Targeting Proteases in Viral Invasion and Host Immune Response, *Frontiers in Molecular Biosciences 7*.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER III

Table 6. Data Collection and Refinement Statistics for Crystallographic Analysis of *cis*-EpS-treated ICL1.

	Covalent adduct of cis-2,3-epoxysuccinic acid to ICL1 (6VB9)	Intact cis-2,3-epoxysuccinic acid bound to ICL1 (6WSI)
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	75.06,129.06, 167.82	79.17, 134.21, 161.04
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	48.94 – 1.88	71.05 - 1.75
R_{sym} or R_{merge}	0.3 (2.6)	0.1 (1.008)
$I/\sigma I$	2.2 (1.2)	10.1 (1.0)
Completeness (%)	90.7 (53.1)	99.9 (98.6)
Redundancy	9 (4.9)	5.9 (3.3)
Refinement		
Resolution (Å)	48.94 – 1.88	71 – 1.75
No. reflections	119322	172629
$R_{work/}$ R_{free}	0.18 / 0.23	0.16 / 0.196
No. atoms		
Protein	13345	13322
Ligand/ion	103	170
Water	971	1176
B-factors		
Protein	25	22
Ligand/ion	31.5	26
Water	21	24
R.m.s deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	0.85	0.84

*Highest resolution shell is shown in parenthesis.

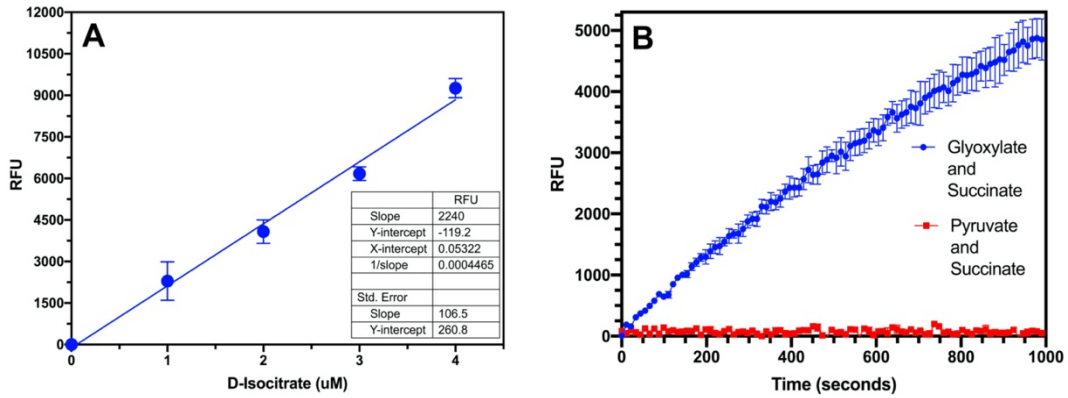


Figure 71. Detection of D-isocitrate Using an Isocitrate Dehydrogenase and Diaphorase Coupled Assay.

A. Standard curve of D-isocitrate versus relative fluorescence units (RFUs) when incubated with the coupled assay system. **B**. Incubation of 30 nM ICL1 in the presence of 2 mM glyoxylate and 2 mM succinate or 2 mM pyruvate and 2 mM succinate was analyzed to ensure the assay was specific for the formation of D-isocitrate, and would not detect methyl isocitrate produced from added pyruvate.

APPENDIX B

SUPPLEMENTAL INFORMATION FOR CHAPTER IV

Table 7. Data Collection and Refinement Statistics for ICL1-5-NIC Co-Crystal Structure

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	75.07, 140.30, 160.69
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	50 – 1.64
R_{sym} or R_{merge}	0.16 (2.975)
$I/\sigma I$	3.4 (0.98)
Completeness (%)	90.9 (46)
Redundancy	5.5 (2)
Refinement	
Resolution (Å)	42.57 – 1.7
No. reflections	168890
R_{work}/ R_{free}	0.1810 / 0.2137
No. atoms	
Protein	13284
Ligand/ion(s)	107
Water	802
B-factors	
Protein	32
Ligand/ion(s)	37
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (º)	1.429

*Highest resolution shell is shown in parenthesis.

APPENDIX C

SUPPLEMENTAL INFORMATION FOR CHAPTER V

Table 8. Steady-State Rates of Acetyl CoA Synthetase Peptide Substrates with *Mt*-Sirt

Substrate	k_{cat} (s^{-1})	K_m (μM)	K_{cat}/K_m (s^{-1} M^{-1})
ACSK617 formyl	N/A[a]	N/A[a]	N/A[a]
ACSK617 acetyl	0.12 ± 0.02	1030 ± 273	1.2 ± 0.4 x 10^2
ACSK617 propionyl	0.08 ± 0.003	410 ± 43	2.0 ± 0.2 x 10^2
ACSK617 butryl	0.02 ± 0.001	1346 ± 134	1.3 ± 0.5 x 10^1
ACSK617 crotonyl	0.007 ± 0.001	4011 ± 933	1.7 ± 0.5 ± x 10^0
ACSK617 hexanoyl	0.02 ± 0.001	975 ± 59	2.0 ± 0.2 x 10^1
ACSK617 octanoyl	0.036 ± 0.0008	182 ± 11	1.9 ± 0.1 x 10^2
ACSK617 decanoyl	0.023 ± 0.0004	24 ± 1.9	9.5 ± 0.8 x 10^2
ACSK617 dodecanoyl	0.047 ± 0.002	42 ± 2.9	1.1 ± 0.08 x 10^3
ACSK617 myristoyl	0.03 ± 0.002	18 ± 3.0	1.8 ± 0.3 x 10^3
ACSK617 palmitoyl	0.016 ± 0.004	8 ± 0.6	2.0 ± 0.2 x 10^3
ACSK617 malonyl	0.048 ± 0.003	1356 ± 130	3.6 ± 0.4 x 10^1
ACSK617 succinyl	0.16 ± 0.004	42 ± 3.2	3.8 ± 0.3 x 10^3
ACSK617 glutaryl	0.05 ± 0.004	105 ± 17	4.8 ± 0.9 x 10^2
ACSK617 biotinoyl	N/A[a]	N/A[a]	N/A[a]

[a]No product formation was detected at 1 mM substrate, 1 μM *Mt*-Sirt, and 1 mM NAD^+ incubated for 30 min at 37ºC