

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Characterization and Biocatalytic Potential of an Enzymatic Friedel–Crafts Alkylation from Cylindrocyclophane Biosynthesis

Citation

Braffman, Nathaniel. 2022. Characterization and Biocatalytic Potential of an Enzymatic Friedel–Crafts Alkylation from Cylindrocyclophane Biosynthesis. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37372291

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Chemistry & Chemical Biology

have examined a dissertation entitled:

Characterization and Biocatalytic Potential of an Enzymatic Friedel–Crafts Alkylation from Cylindrocyclophane Biosynthesis

presented by: Nathaniel R. Braffman

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____
Typed name: Professor Emily Balskus

Signature _____
Typed name: Professor Eric Jacobsen

Signature _____
Typed name: Professor Christina Woo

Date: 6 May 2022

Characterization and Biocatalytic Potential of an Enzymatic Friedel–Crafts Alkylation from Cylindrocyclophane Biosynthesis

A dissertation presented

by

Nathaniel R. Braffman

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts, USA

May 2022

© 2022 Nathaniel R. Braffman
All rights reserved.

Dissertation Advisor: Professor Emily P. Balskus Nathaniel R. Braffman

Characterization and Biocatalytic Potential of an Enzymatic Friedel–Crafts Alkylation from Cylindrocyclophane Biosynthesis

Abstract

The cyanobacterial enzyme CylK assembles the cylindrocylophane natural products by performing two sequential Friedel–Crafts alkylation reactions, forming new alkyl–aryl carbon–carbon bonds between resorcinol nucleophiles and secondary alkyl chloride electrophiles. Other enzymatic transformations analogous to the Friedel–Crafts alkylation have been reported, however they require highly activated electrophilic substrates. Remarkably, the substitution reaction catalyzed by CylK occurs in a selective fashion with inversion of configuration at the alkyl chloride stereocenter. This transformation is one of only a few known biological examples of 'cryptic halogenation', a biosynthetic logic that produces a halogenated intermediate and leverages its increased reactivity to form a new bond, releasing the corresponding halide anion as a side product. Although multiple classes of halogenating enzymes are mechanistically characterized, the cognate halide-utilizing enzymes involved in cryptic halogenation are underexplored. Moreover, the reaction catalyzed by CylK is the only known biological example of aromatic ring alkylation with an alkyl halide electrophile, and broadly, structures and biochemical mechanisms for enzymatic activation of alkyl halide substrates as biosynthetic intermediates are unknown. Here, we describe our efforts to characterize CylK in order to evaluate its potential as a biocatalyst and to enhance our fundamental knowledge of this enzyme family.

In Chapter 2 we describe our work to improve CylK enzyme preparations and to elucidate the substrate scope and selectivity of this reaction. These efforts were motivated by the biocatalytic potential of CylK and the desire to develop an understanding of mechanistic and substrate binding requirements that might influence substrate tolerance and reactivity. Our results reveal that CylK can mediate Friedel–Crafts alkylations with an array of non-biological substrate pairs while maintaining the exquisite regio- and stereoselectivity of the native biological reaction; however, CylK has a strict requirement for nucleophilic

resorcinol substrates with free phenolic substituents. Subsequently, collaborators discovered that an enzyme homologous to CylK instead catalyzes carbon–oxygen bond formation between carboxylate nucleophiles and secondary alkyl chloride electrophiles, suggesting that while the broader enzyme family likely shares an alkyl halide activating strategy, CylK may have a distinct active site architecture that enables resorcinol engagement.

In Chapter 3, in collaboration with the Boal Lab at Penn State University, we determine the structural basis for the CylK-catalyzed reaction in order to understand how it orients and activates resorcinol nucleophiles and alkyl halide electrophiles for stereospecific nucleophilic substitution. Our X-ray crystal structures reveal that CylK is a distinctive fusion of Ca^{2+}-binding and β-propeller domains, and site-directed mutagenesis and anomalous diffraction with bound bromide ions locate the active site to the domain interface. This implicates hydrogen bonding interactions with Arg105 and Tyr473 in CylK's alkyl halide activation strategy, and additional density functional theory and molecular dynamics modeling suggests that the nucleophilicity of resorcinol substrates is enhanced via deprotonation by Asp440. Notably, free halide anions do not inhibit the reaction catalyzed by CylK, suggesting that this enzyme precisely calibrates the concerted activation of both reaction partners. Bioinformatic analysis of the enzyme family establishes that these three key catalytic residues are conserved in homologous enzymes from other cyanobacteria that likely use additional nucleophilic substrates beyond resorcinols and carboxylates.

Finally, in Chapter 4 we apply our knowledge of CylK's structure and substrate scope to begin preliminary protein engineering efforts. Motivated by the therapeutic potential of alkylated stilbenes accessed in trace quantities with wild-type CylK, we develop screening methodology for enzyme evolution that leverages the putative biological activity of the CylK reaction products. This strategy, dubbed bioactivity-guided directed evolution, seeks to enable biocatalyst development in parallel with optimization of biological activity. Additionally, we construct libraries of CylK mutants and identify two enhanced variants that may serve as starting points for additional rounds of mutagenesis and screening. Future efforts are focused on applying insights from our structural, mechanistic, and bioinformatic work to guide protein engineering, natural product discovery, and further enzyme characterization.

Acknowledgements

First, I need to thank my dissertation advisor, mentor, and supervisor, Prof. Emily Balskus. Your lab has been an extraordinary place to work and learn. It is a rigorous and supportive environment where we are pushed to think outside of our immediate discipline, and to strive for scientific excellence. It has been my privilege to watch, and to be a part of, the research team and program as it has grown and evolved. Thank you for the freedom to explore my scientific and professional interests, and for the countless insights and gentle nudges in the 'right' direction that you have brought to my work and academic development.

To all my research advisors, Jennifer Gordinier, Dr. Barbara Grosz, Prof. Scott Snyder, Prof. Seth Darst, Prof. Elizabeth Campbell, and Dr. Jason King; thank you for the knowledge and mentorship at different stages of my education and career. This work builds from the foundation developed under your guidance.

Prof. Eric Jacobsen and Prof. Christina Woo, thank you for serving on my thesis and advising committees over the past five years. This work reflects your engagement and challenging questions posed as we struggled to work with and understand this unique enzyme.

To Myles and Erica, now Prof. Smith and Prof. Schultz! Thank you for the excellent hands-on training and mentorship as a young scientist. I am eager to watch your research and academic programs develop, and I know that you both will be exceptional professors.

To my collaborators, Prof. Erica Schultz, Dr. Michael Luescher, Harry Hager, Prof. Amie Boal, Prof. C. Denise Okafor, Terry Ruskoski, Prof. Katherine Davis, Dr. Nate Glasser, Cassidy Johnson, Prof. Pedro Leão, João Reis, Michelle Wang; this work would not be possible without your contributions. Thank you for your hard work and ideas.

I would also like to acknowledge the invaluable technical assistance that I have received from Prof. Paul Boudreau, Prof. Monica McCallum, Dr. Andrew Bendelsmith, Martin Reiterer, Dr. Grace Kenney, Beverly Fu, Dominick Jenkins, Dr. Li Zha, Dr. Doug Kenny, and Dr. Alexis Caulier. Thank you for the generosity of your time and expertise.

To Drs. Jack Nicoludis, Matt Wilson, Matt Volpe, Vayu Maini Rekdal, Stephania Irwin, Chip Li, Yolanda Huang, Abe Waldman, Spencer Peck, and Nitzan Koppel; thank you for excellent advice and wisdom, both solicited and unsolicited, over the years. It has been my privilege to learn from and alongside you all.

To the members of the Balskus Lab, and our superb administrator Rhonda Pautler, thank you bringing your varied interests, passions, and efforts to the workplace. All of you individually contribute to this dynamic learning environment that has shaped my outlook on science, and the work presented below. I hope you enjoy the new lab space!

Finally, to our union, HGSU-UAW Local 5118, thank you for the solidarity and support over what has been several challenging years. Playing a small role in building our union has been the most meaningful experience of my graduate career. Thank you for holding Harvard accountable, advocating for better conditions for all workers on campus, and demonstrating how local communities can support each other.

Table of Contents

Title page i
Copyright ii
Abstract iii
Acknowledgements v
Table of contents vii
List of figures x
List of schemes xii
List of tables xiii
List of abbreviations xiv

1 Introduction to cryptic halogenation and cylindrocyclophane biosynthesis 1
1.1 Natural products and biosynthetic enzymes are a source of inspiration 1
1.2 Cryptic halogenation strategies in natural product biosynthesis 5
1.3 Cylindrocyclophane biosynthesis 9
1.4 Enzymatic Friedel–Crafts alkylations 18
1.5 Synthetic strategies to form alkyl–aryl C–C bonds 22
1.6 Chapter preview 28
2 Elucidating the substrate scope of CylK and homologous enzymes 30
2.1 Introduction 30
2.2 Results and Discussion 36
2.2.1 Efforts to optimize enzyme overexpression and purification methods 36
2.2.2 Biocatalytic application and optimization of non-native CylK reaction 42
2.2.3 Characterization and enantiomeric excess of a non-native CylK catalyzed reaction 45
2.2.4 Alkyl halide electrophile substrate scope of CylK mediated Friedel–Crafts alkylation 50
2.2.5 Resorcinol nucleophile substrate scope of CylK mediated Friedel–Crafts alkylation 53
2.3 Conclusions 58
2.4 Materials and Methods 62
2.4.1 General procedures and materials 62
2.4.2 Cloning of N-terminal fusions and plasmids used in this study 64
2.4.3 Overexpression and purification of CylI 66
2.4.4 Soluble overexpression and purification of CylK 66
2.4.5 Pilot attempts to refold CylK on a small scale 67
2.4.6 Insoluble overexpression, refolding, and purification of CylK 68
2.4.7 Optimized refolding and purification of CylK 69
2.4.8 Overexpression and purification of ULP protease 70
2.4.9 Measuring native CylK activity in lysate 71

2.4.10 Experimental Data for Enzymatic Reactions 73
2.4.11 Experimental Data for Synthetic Substrates and Standards 82
2.4.12 High Resolution Mass Spectra 101
2.4.13 NMR Characterization Data 113
3 Structural and mechanistic basis for the CylK-catalyzed Friedel–Crafts alkylation 131
3.1 Introduction 131
3.2 Results and Discussion 137
3.2.1 Density functional theory modeling and mechanistic speculation 137
3.2.2 Structural determination effort 140
3.2.3 Locating the CylK active site 148
3.2.4 Identifying key catalytic residues 156
3.2.5 Modelling substrate orientation in the active site 158
3.2.6 Comparison to other CylK crystal structures 161
3.2.7 Targeted mutagenesis 167
3.2.8 Halide inhibition 172
3.2.9 Mechanistic proposal 178
3.2.10 Bioinformatic support for proposed mechanism and nucleophile diversity 182
3.3 Conclusions 190
3.4 Materials and Methods 195
3.4.1 DFT Calculations 195
3.4.2 Refolding and purification of apoCylK-Strep 195
3.4.3 Competitive lanthanide substitution 196
3.4.4 Refolding and purification of CylK-Strep for crystallography 196
3.4.5 Refolding and purification of chloride-free CylK-Strep 197
3.4.6 Crystallization and structure determination 200
3.4.7 Mutant enzyme activity screening with non-biological substrates 205
3.4.8 Mutant enzyme activity screening with biological substrates 206
3.4.9 Molecular dynamics simulations 209
3.4.10 Kinetic experiments with halide ions as potential inhibitors 210
3.4.11 Bioinformatic analysis 210
4 Enzyme engineering and method development 212
4.1 Introduction 212
4.2 Results and Discussion 215
4.2.1 Bioactivity-guided directed evolution method development 215
4.2.2 Luciferase reporter validation and AhR activity of novel benvitimod derivatives 223
4.2.3 Generation of CylK mutant libraries 225
4.2.4 Pilot of bioactivity-guided direction evolution 228
4.2.5 UPLC-MS guided protein evolution 231

4.2.6 Validation and characterization of enhanced CylK variants 237
4.3 Conclusions and Future Directions 243
4.4 Materials and Methods 248
4.4.1 LC-MS conditions 248
4.4.2 Cell culture and qRT-PCR detection of AhR activity 249
4.4.3 Luciferase AhR reporter assays 249
4.4.4 Error prone PCR library construction 250
4.4.5 Site saturation mutagenesis libraries construction 252
4.4.6 Cloning CylK-Strep-GFP constructs 254
4.4.7 Cloning CylK-Strep-GFP-6xHis constructs 255
4.4.8 Plate-based CylK growth, expression and activity assay 260
5 References 262

Figures

Figure 1.1. Select natural products with known physiological roles..... 2
Figure 1.2. Image of *Cylindrospermum stagnale* and notable cyanobacterial natural products 4
Figure 1.3. Cryptically halogenated natural products and halogenated precursors 6
Figure 1.4. Structures of the cylindrocyclophane natural products..... 9
Figure 1.5. Structures of related cyanobacterial cyclophane natural products..... 11
Figure 1.6. Feeding studies and β-methyl installation identify biosynthetic gene cluster 12
Figure 1.7. *in vitro* biochemistry and feeding study highlights dimerization biosynthetic logic..... 13
Figure 2.1. The four possible diastereomers of CylK's native monomeric substrate..... 34
Figure 2.2. CylK constructs designed for protein preparation development 37
Figure 2.3. Expression and activity of CylK N-terminal fusions 40
Figure 2.4. HPLC and chiral LC-MS analysis of CylK reactions..... 49
Figure 2.5. Electrophilic substrates not accepted by CylK 52
Figure 2.6. Nucleophilic substrates not accepted by CylK 55
Figure 2.7. AhR agonists and synthetic precursor stilbene nucleophiles..... 55
Figure 2.8. Extended alkyl portions of electrophilic substrates..... 59
Figure 2.9. Necessary 1,3-diol relationship of nucleophilic substrates 59
Figure 2.10. Protein SDS-PAGE and size exclusion chromatogram 72
Figure 2.11. LC-MS analysis of alkyl fluoride and chloride electrophile substrates..... 98
Figure 3.1. Dichloroethane catabolism and mechanism of haloalkane dehalogenase..... 136
Figure 3.2. Electronic considerations of resorcinol nucleophile activation 138
Figure 3.3. The CylK X-ray crystal structure..... 142
Figure 3.4. N-terminal domain architecture of CylK..... 143
Figure 3.5. Structural comparison of N-terminal Ca^{2+} binding domain of CylK to PAP protease 144
Figure 3.6. C-terminal domain architecture of CylK..... 146
Figure 3.7. Comparison of CylK's C-terminal β-propeller domain with structurally related enzymes..... 147
Figure 3.8. Potential active site cavities of CylK..... 149
Figure 3.9. Mutant screen to identify active site at the domain interface 151
Figure 3.10. Anomalous diffraction of bound bromide in active site and mechanistic implications 153
Figure 3.11. Mutant screen of CylK active site suggests 'Br1' is key catalytic center 156
Figure 3.12. Mutagenesis of key CylK residues reveals a single catalytic site 158
Figure 3.13. Molecular dynamics simulations of the intermolecular alkylation catalyzed by CylK 159
Figure 3.14. Molecular dynamics simulations of second alkylation 160
Figure 3.15. Comparison of CylK crystal structures from different research groups 162
Figure 3.16. Substrate analogs and active site comparison between CylK crystal structures 164

Figure 3.17. Comparison of active site residues and distances between CylK crystal structures..... 165
Figure 3.18. Comparison of substrate orientations between CylK crystal structure and modeling 166
Figure 3.19. Point and combinatorial mutagenesis of key substrate activating residues..... 169
Figure 3.20. Chemdraw representations of enzyme engagement of alkyl chlorides 171
Figure 3.21. The effect of millimolar concentrations of chloride on the CylK-catalyzed reaction 173
Figure 3.22. The effect of micromolar concentrations of halides on the CylK-catalyzed reaction..... 174
Figure 3.23. Resorcinol–alkyl chloride substitution reaction transition states and electrostatics 177
Figure 3.24. Proposed CylK catalytic cycle 179
Figure 3.25. Structural comparison of alkyl halides bound in enzyme active sites..... 181
Figure 3.26. Annotated phylogenetic tree of likely alkyl halide-utilizing CylK homologs..... 184
Figure 3.27. Conservation mapping analysis of likely CylK homologs..... 186
Figure 3.28. Chlorinated glycolipids from *N. punctiforme* ATCC 29133 187
Figure 3.29. CylK homolog from *N. punctiforme* oligomerizes glycolipid with hydroxyl nucleophile 189
Figure 3.30. Comparison of 7-bladed β-propeller enzymes 191
Figure 3.31. Size exclusion chromatography and SDS-PAGE of chloride-free CylK 199
Figure 3.32. Western blotting of CylK mutants 208
Figure 4.1. AhR ligands and uncharacterized analogs 218
Figure 4.2. AhR activity of stilbene nucleophiles 220
Figure 4.3. AhR activity of pseudo reaction mixtures of 3,5-dihydroxystilbene and benvitimod 222
Figure 4.4. Luciferase reporter validation and bioactivity of CylK-accessed stilbene products..... 224
Figure 4.5. Sites selected for single saturation mutagenesis 227
Figure 4.6. Timing of luciferase reporter response 229
Figure 4.7. Relative bioactivity of enzymatic reaction products..... 230
Figure 4.8. Histogram of mutant screen for enhanced CylK activity..... 234
Figure 4.9. Putative N-terminally truncated CylK variant 236
Figure 4.10. Activity validation of enhanced CylK variants..... 238
Figure 4.11. Activity and fluorescence of CylK-GFP..... 239
Figure 4.12. Activity and fluorescence of CylK-GFP variants..... 240
Figure 4.13. Western blotting of evolved CylK variants 242

Schemes

Scheme 1.1. Cryptic halogenation in coronatine biosynthesis 7
Scheme 1.2. Proposed biosynthesis of cylindrocyclophane F 15
Scheme 1.3. Synthetic Friedel–Crafts alkylation reactions and mechanisms 19
Scheme 1.4. Enzymatic electrophilic aromatic substitution reactions 20
Scheme 1.5. Retrosyntheses of cylindrocyclophane macrocycle and alkyl–aryl bond 23
Scheme 1.6. Total synthesis of cylindrocyclophane F with asymmetric aryl allylation 26
Scheme 1.7. Alkyl–aryl coupling reactions with secondary alkyl and allyl chlorides 27
Scheme 2.1. The role of CylK in cylindrocyclophane biosynthesis and non-biological coupling 33
Scheme 2.2. Stereoselectivity and stereoinversion of the CylK-catalyzed reaction 35
Scheme 2.3. Synthesis of racemic CylK product standard 47
Scheme 2.4. CylK homolog BrtB catalyzes esterification of alkyl chloride substrates 60
Scheme 3.1. CylK and related enzymes use alkyl chloride substrates as biosynthetic intermediates 133
Scheme 3.2. Putative mechanistic similarities between CylK and BrtB 139
Scheme 4.1. Convergent biocatalytic strategies 213

Tables

Table 2.1. Optimization of growth and expression conditions enhancing CylK activity in lysate 38
Table 2.2. Activity of denatured and refolded polyhistidine-tagged CylK constructs 41
Table 2.3. Optimization of non-native CylK reaction with model substrate pair 44
Table 2.4. Further optimization of 1 mL scale reactions 45
Table 2.5. ^{1}H NMR comparison of product 8 prepared enzymatically or via chemical synthesis 48
Table 2.6. ^{13}C NMR comparison of product 8 prepared enzymatically or via chemical synthesis 48
Table 2.7. Alkyl halide electrophile substrate scope 51
Table 2.8. Resorcinol nucleophile substrate scope 54
Table 2.9. Alkylated stilbenes accessed enzymatically with CylK 57
Table 2.10. Primers used to clone N-terminal fusions of CylK-Strep 65
Table 2.11. Alkyl halide electrophile product masses. 76
Table 2.12. Resorcinol nucleophile product masses. 78
Table 3.1. ICP-OES analysis of lanthanide reconstituted CylK 141
Table 3.2. Conservation of CylK residues selected for mutant screen to locate active site 152
Table 3.3. Conservation of CylK active site residues selected for mutant screen 155
Table 3.4. Data collection and refinement statistics for Tb-soaked CylK structures 202
Table 3.5. Structural homologs of the N-terminal Ca^{2+} binding domain of CylK 203
Table 3.6. Structural homologs of the C-terminal β-propeller domain of CylK 203
Table 3.7. Data collection and refinement statistics for Br-soaked CylK structures 204
Table 3.8. Point mutants of CylK-Strep 207
Table 4.1. Random mutagenesis library statistics 226
Table 4.2. Select CylK reactions for activity optimization 232
Table 4.3. Sequencing results of most active CylK variants 235
Table 4.4. Tandem mass spectrometry parameters 248
Table 4.5. Primers used to quantify AhR activity via CYP1A1 expression 249
Table 4.6. Primers used to clone epPCR library of CylK-Strep-GFP11 251
Table 4.7. Primers used to clone SSM libraries of CylK-Strep-GFP11 253
Table 4.8. Primers used to clone CylK-Strep-GFP 255
Table 4.9. Primers used to clone CylK-Strep-GFP-6xHis 255
Table 4.10. DNA sequences of unpublished CylK constructs 256

List of Abbreviations

°C	degree Celsius
d	deuterium
DFT	density-functional theory
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
FT-IR	Fourier-transform infrared spectroscopy
g	gram
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid
HPLC	high performance liquid chromatography
Hz	hertz
ICP-OES	inductively coupled plasma-optical emission spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	coupling constant
kDa	kilodalton
L	liter
LC-MS	liquid chromatography-mass spectrometry
M	molar
MHz	megahertz
min	minute
mL	milliliter
mol	mole
MS/MS	tandem mass spectrometry
MWCO	molecular weight cut-off
NAD	nicotinamide adenine dinucleotide

Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NRPS	nonribosomal peptide synthetase
OD	optical density
PCR	polymerase chain reaction
PKS	polyketide synthase
PQQ	pyrroloquinoline quinone
qRT-PCR	real-time quantitative reverse transcription polymerase chain reaction
RiPP	ribosomally synthesized and post-translationally modified peptide
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TLC	thin layer chromatography
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
µg	microgram
µL	microliter
µm	micrometer
UPLC-MS	ultra performance liquid chromatography-tandem mass spectrometry
UV	ultraviolet
V	volt
Vis	visible
x *g*	times force of gravity

1 Introduction to cryptic halogenation and cylindrocyclophane biosynthesis

1.1 Natural products and biosynthetic enzymes are a source of inspiration

Over the course of evolution, across all domains of life on Earth, Nature has developed incredible molecular complexity that fascinates scientists and can be harnessed for the benefit of humankind. For instance, small molecule natural products produced by living organisms are widely employed as life-saving therapeutics,[1,2] while enzymes and even entire cells can be engaged and engineered to manufacture pharmaceutical compounds[3,4] and other anthropogenic materials.[5] The focus of this thesis will be on a family of cyanobacterial natural products and the powerful enzymatic strategies Nature has deployed to construct these small molecule metabolites.

Articulated by Firn and Jones in 1991, a unifying evolutionary theory of natural products is the 'screening hypothesis',[6] or the diversity-based model,[7] which postulates that organisms produce and in effect 'screen' large numbers of small molecule metabolites such that one or more of these molecular entities may confer a potential benefit to the host organism or ecosystem (Figure 1.1). This hypothesis seeks to explain an apparent quirk of natural product producers: the observation that biosynthetic pathways tend to generate many small molecule products,[8] some without potent biological activity or a clear physiological role. Notably, this screening strategy neatly parallels the practice of contemporary medicinal chemistry. Upon identifying a bioactive small molecule scaffold, medicinal chemists then make systematic changes to its structure in order to determine which molecular features contribute to that desired activity, and what alterations can be made that might improve potency, bioavailability, and reduce potential toxicities – among other properties.[9] Moreover, many small molecule therapeutics are natural products themselves (~5%), or their structures are inspired by natural products (~22%) but contain synthetic modifications to enhance therapeutic performance.[2] In biological systems, however, the primary determinant of natural product structure and diversity are genetically encoded biosynthetic enzymes that catalyze the chemical reactions necessary to assemble these small molecule products.

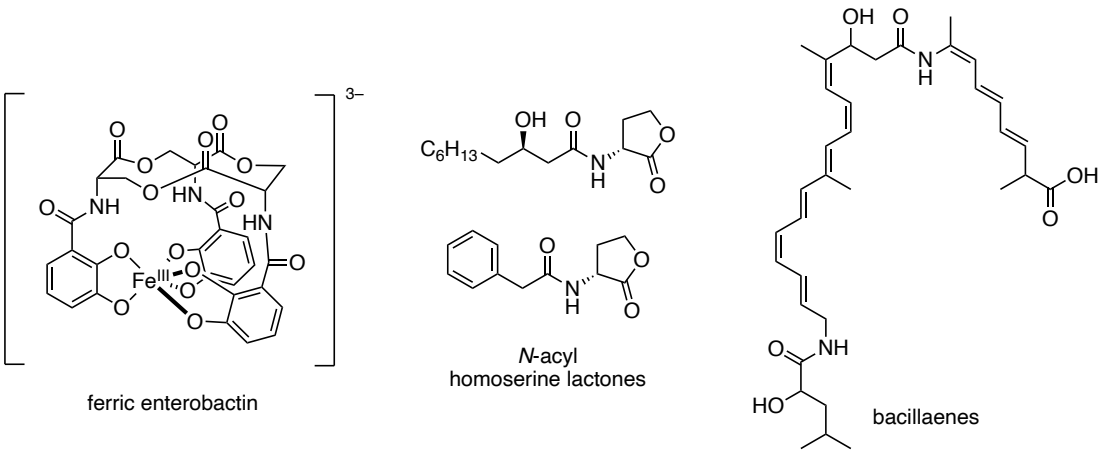


Figure 1.1. Select natural products with known physiological roles

Enterobactin is a siderophore that chelates extracellular iron(III) for use by the producing organism, the *N*-acyl homoserine lactones are implicated in community signaling and regulation, while the bacillaenes are antibiotic compounds that aid the producer in competing for its ecological niche.

From this genetic vantage point, natural product-producing organisms might maintain the genes encoding biosynthetic enzymes if they provide a selective advantage. For example, enzymes that construct small molecule metabolites to scavenge the surrounding environment for limiting nutrients provide a growth advantage to the encoding organism.[10] Other interesting examples include the production of metabolites that play a role in communication and community structure,[11,12] or antibiotics that inhibit bacterial competitors and help the producing organism establish colonization.[13,14] Alternatively, biosynthetic genes can also persist in organisms if their presence and maintenance do not also confer competitive disadvantages, such as self-directed toxicity[15] or growth limitations.[16] In this evolutionary niche somewhere between 'essential' and 'not harmful', biosynthetic enzymes seem to have proliferated and acquired remarkable, and sometimes promiscuous, chemical activities. Fischbach and Clardy suggest that evolutionary competition has rewarded biosynthetic enzymes with flexible substrate scopes because the enzymes themselves might be fewer genetic steps away from gaining new and beneficial reactivity.[7] In support of this, it has been widely observed that distinct chemical reactions can be catalyzed by related enzymes within the same superfamily, suggesting that they arose from a common ancestor.[17] This promiscuity and potential ease of genetic reprogramming also comes as an advantage to the fields of

biocatalysis and enzyme engineering which seek to apply enzymes in synthetic contexts.[18] Although entirely synthetic enzyme scaffolds have been designed,[19] the majority of biocatalysts are derived from native biological enzymes. In this way, both natural products and biosynthetic enzymes responsible for their production serve as a source of inspiration for downstream applications.

Our work will focus on a remarkable family of cyanobacterial enzymes and their associated polyketide natural products. Cyanobacteria (Figure 1.2A) and other microorganisms are prolific producers of secondary metabolites.[20,21] Other notable natural product producers include fungi and plants; extracts of these organisms have been used as medicines as early as 2600 B.C.E.[22] Natural products and their respective biosynthetic enzymes can be classified into five major classes: polyketides, peptide based natural products, saccharides, alkaloids, and isoprenoids. More than 90 genera of cyanobacteria are known natural product producers, encompassing structural diversity across all of these natural product classes (Figure 1.2B) with varied bioactivities that have found clinical relevance.[20]

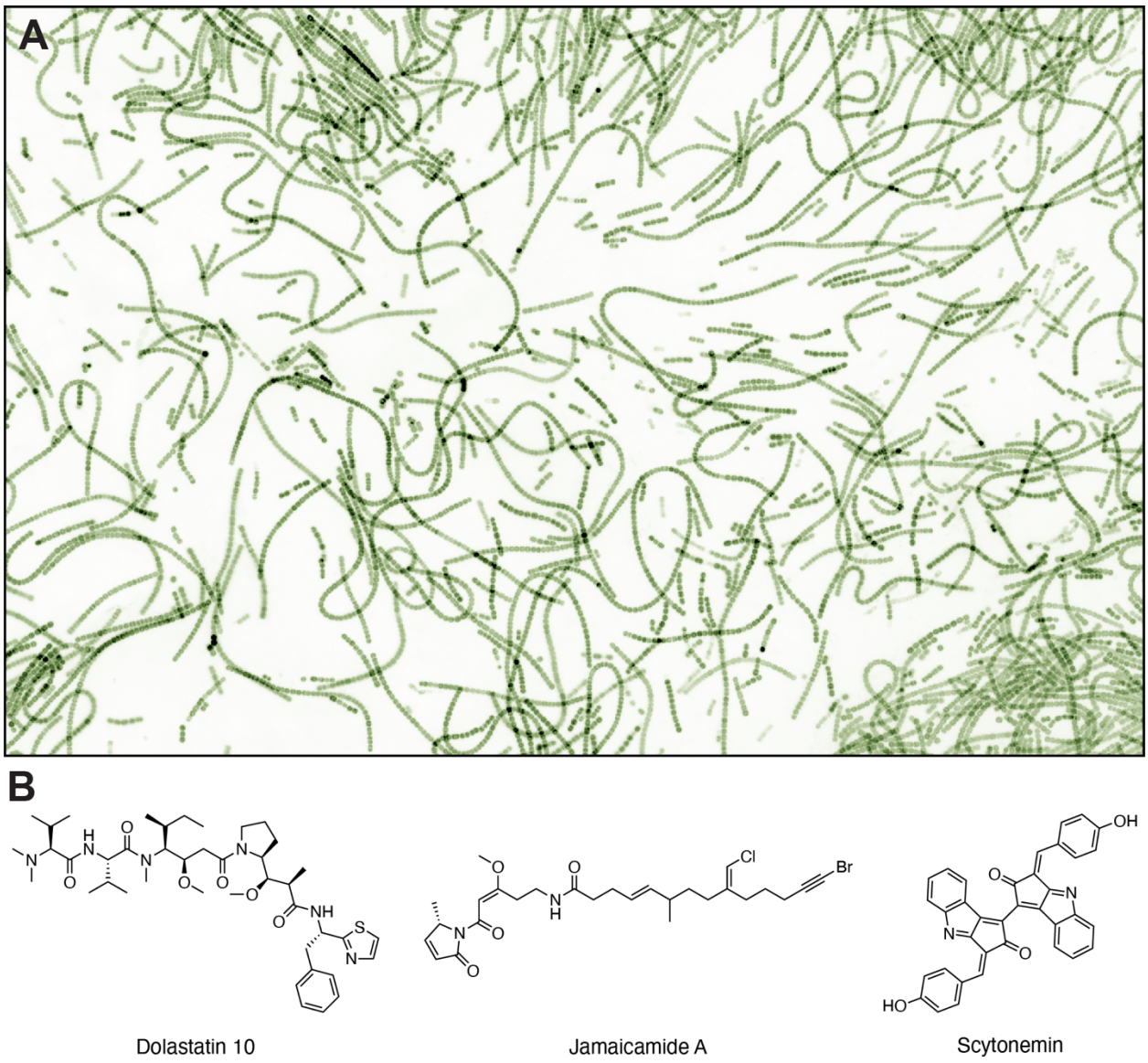


Figure 1.2. Image of *Cylindrospermum stagnale* and notable cyanobacterial natural products

(*A*) Colorized optimal microscopy image of cylindrocyclophane natural product producer, *C. stagnale* ATCC 29204, captured with Dr. John M. Nicoludis and Dr. Nitzan Koppel. Filaments of vegetative *C. stagnale* cells are visible. The ends of *Cylindrospermum* filaments are generally terminated with specialized heterocyst cells, believed to be the site of nitrogen fixation. (*B*) Representative cyanobacterial natural products from peptide based, hybrid polyketide/nonribosomal peptide, and alkaloid classes.[20,23,24]

Polyketides are iteratively assembled by modular and multifunctional enzymes known as polyketide synthases (PKSs).[25] This biosynthetic logic parallels that of fatty acid biosynthesis, although PKSs generate metabolites with greater structural diversity.[26] Facilitating the study of biosynthesis, microorganisms tend to

physically cluster the genes associated with producing natural products.[27] We refer to these groups of genes as biosynthetic gene clusters (BGCs). Clustered in this way, microorganisms can co-regulate these genes, and they have developed sophisticated strategies to horizontally transfer DNA[28,29] and along with it, the capacity to produce particular natural products.[30] The advent of cheap DNA sequencing technologies has greatly facilitated the study of natural product biosynthesis and biosynthetic enzymes. Now, not only can gene and enzyme sequences be associated with specific reactivity and their role in biosynthesis, but all sequenced genomes and metagenomes can be searched for relatives of known biosynthetic genes in order to discover new enzymes and natural products. This genome mining work has revealed that bacteria harbor a large number of silent or 'cryptic' BGCs that are not yet linked to the production natural products, and these gene clusters may contain uncharacterized enzymes of unknown function which can make natural product structure prediction elusive without further interrogation.[31] DNA sequencing technology has ushered in an unprecedented era in which the sequences of genes and enzymes are the starting point for hypothesis generation, and ultimately, the discovery of new chemical activities and metabolites that impact multiple scientific disciplines and commercial industries.

1.2 Cryptic halogenation strategies in natural product biosynthesis

Chemists utilize alkyl and aryl halides (molecules with sp^2 and sp^3 C–X bonds, X = F, Cl, Br, or I) as key synthetic reagents because of their accessibility and ease of activation;[32,33] however, examples of their use as intermediates in biological systems are rare. A few characterized natural product biosynthetic pathways generate transiently halogenated intermediates that facilitate downstream biosynthetic reactions in a strategy known as 'cryptic halogenation'.[34] Diverse features of natural products constructed via cryptic halogenation include carbon–nitrogen (C–N) biaryl couplings,[35] terminal alkyne formation,[36] and cyclopropanation[37] (Figure 1.3). The biosynthetic logic of cryptic halogenation requires two partner enzymes: a halogenase, which introduces the halide substituent, and a second enzyme that utilizes the halogenated intermediate as a substrate, leveraging its increased reactivity and releasing the corresponding halide anion as a side product.

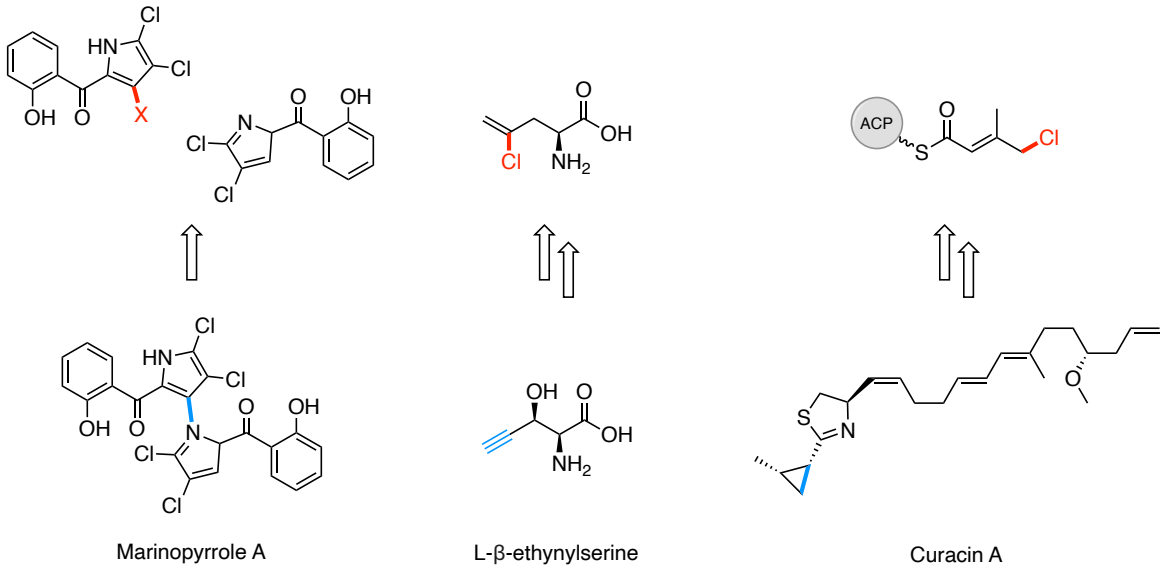
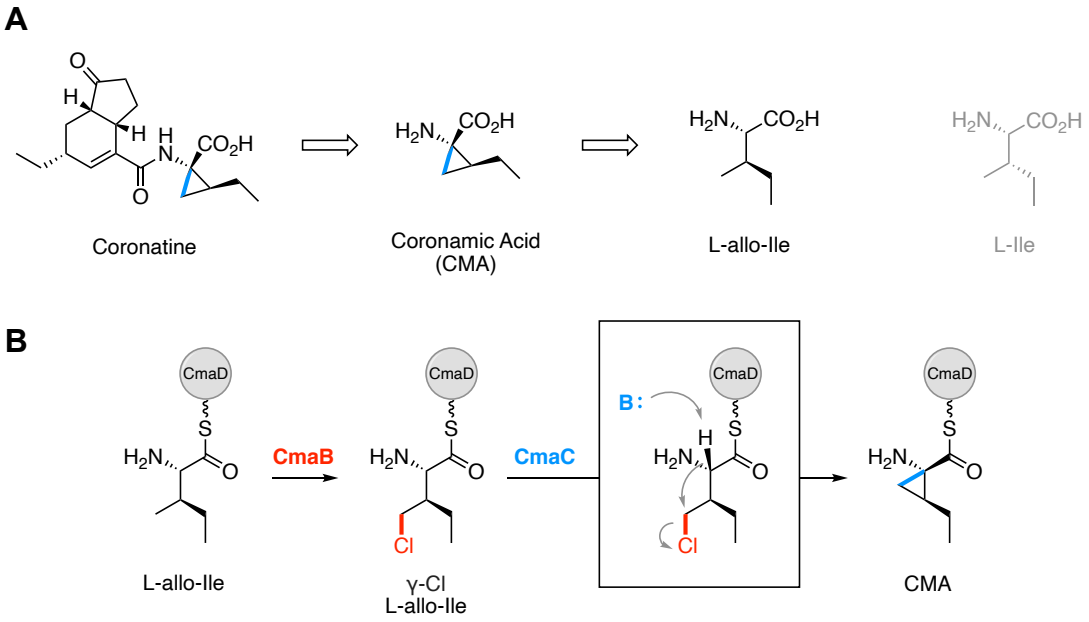


Figure 1.3. Cryptically halogenated natural products and halogenated precursors

Three representative natural products assembled via cryptic halogenation. The bonds displayed in blue were formed by engaging halogenated intermediates and releasing the halide anion as side product. Halogenated intermediates are displayed above retrobiosynthetic arrows, with halide substituents represented in red. Note, the halogenated intermediate precursor to marinopyrrole A is proposed and has not been isolated.[35]

The first reported example of cryptic halogenation in natural product biosynthesis was the cyclopropyl-containing metabolite coronatine (Scheme 1.1A).[38] Biochemical work revealed that coronamic acid, bearing the cyclopropyl moiety, is a key building block incorporated into the final natural product by nonribosomal peptide synthetase (NRPS) modules. It was further demonstrated that L-allo-Ile, a diastereomer of L-isoleucine, is the biosynthetic precursor of coronamic acid. Intriguingly, the key carbon–carbon (C–C) bond of the cyclopropyl ring in coronamic acid was shown to be formed in two steps (Scheme 1.1B). First, the iron(II) 2-(oxo)-glutarate (Fe^{II}/2OG)-dependent halogenase CmaB selectively chlorinates the unactivated γ carbon of L-allo-Ile tethered via a thioester linkage to the phosphopantetheine (ppant) arm of the freestanding thiolation (T) domain CmaD. Then, a Zn^{2+}-dependent enzyme, CmaC, activates this chlorinated intermediate to form the cyclopropyl ring.[39] It is proposed that CmaC deprotonates the alpha carbon of L-allo-Ile, and the resulting thioester enolate undergoes intramolecular nucleophilic addition with the alkyl chloride to form the key C–C bond of the cyclopropyl ring. More work is necessary to determine the structure of CmaC, its active site residues, and to shed light on the specifics of its mechanism and the

potential role of zinc in catalysis. Other biosynthetic intramolecular nucleophilic substitution strategies to form cyclopropyl rings invoke *S*-adenosylmethionine (SAM), utilizing its highly reactive sulfonium as a leaving group to facilitate intramolecular C–C bond formation.[40-42] However, unlike SAM, activating the relatively inert carbon–chloride (C–Cl) bond of γ-Cl L-allo-Ile for substitution is chemically challenging.



Scheme 1.1. Cryptic halogenation in coronatine biosynthesis

(*A*) Retrobiosynthesis of coronatine from L-allo-Ile, an isomer of L-Ile. (*B*) The Fe^{II}/2OG-dependent halogenase CmaB generates chlorinated L-allo-Ile for Zn^{2+}-dependent CmaC, which catalyzes cyclopropyl formation by intramolecular nucleophilic substitution of chloride. Deprotonation and substitution are shown as a single step but a thioester enolate intermediate is possible. The halogenase and alkyl halide bond are displayed in red, while the halide-utilizing enzyme and corresponding C–C bond formed are represented in blue.

An analogous cryptic chlorination and cyclopropanation strategy occurs during the biosynthesis of kutzneride, however, the nucleophilic displacement to form the cyclopropyl ring is catalyzed by flavin-containing acyl-CoA dehydrogenase KtzA in a redox neutral fasion.[43] Distinctly, the cyclopropyl group in curacin A is formed from an allyl chloride-containing intermediate (see Figure 1.3), and enoyl reductase CurF catalyzes cyclopropanation and reduction of this intermediate.[44] Other scaffolds constructed via

cryptic halogenation utilize additional classes of halide-utilizing enzymes. Terminal alkyne formation in L-β-ethynylserine is catalyzed by PLP-dependent enzyme BesB from a vinyl chloride intermediate.[36] Genetic deletion experiments revealed that three enzymes are required to construct the C–N biaryl bond of marinopyrrole A, two flavin-dependent halogenases paired with a flavin reductase, and although this suggestive of cryptic halogenation, a halogenated intermediate precursor to marinopyrrole A has not been isolated and the roles of these enzymes are unclear.[35] In general, enzymatic strategies to activate alkyl and aryl halide substituents are of fundamental significance, especially alkyl chlorides.

Halogenating enzymes in their own right have generated considerable interest as biocatalysts,[45] and multiple classes of halogenases have been structurally and mechanistically characterized.[46] Halogenases can be classified into three distinct categories based on their mechanisms of halogenation: (1) haloperoxidases[47] and flavin-dependent halogenases[48] generate electrophilic halogen species that react with alkenes and aromatic rings, (2) nucleophilic halogenases utilize halide anions (X^-) to attack electrophilic carbon centers such as SAM,[49] and (3) iron(II) 2-(oxo)-glutarate (FeII/2OG)-dependent halogenases can selectively halogenate unactivated sp^3 carbon centers with a radical based mechanism.[50] However, the cognate halide-utilizing enzymes in pathways that employ cryptic halogenation remain underexplored. In particular, the structures and biochemical mechanisms for enzymatically engaging and activating alkyl halide substrates as intermediates in natural product biosynthesis are unknown. Interrogating cylindrocyclophane biosynthesis presents an exciting opportunity to study a notable example of cryptic halogenation and alkyl halide utilization in biological systems.

1.3 Cylindrocyclophane biosynthesis

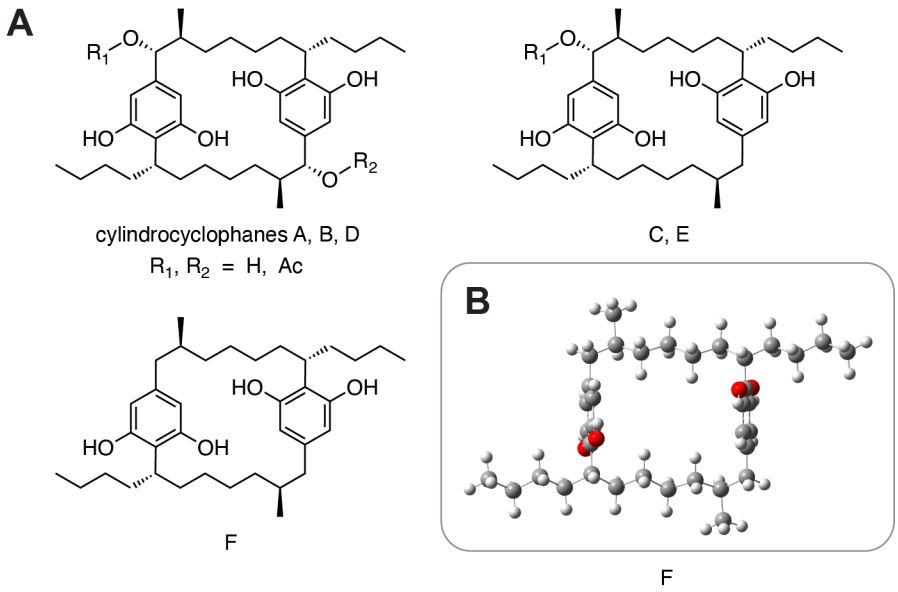


Figure 1.4. Structures of the cylindrocyclophane natural products

(*A*) The cylindrocyclophanes vary in hydroxylation and acetylation patterns. The simplest member – and the primary focus of this work is cylindrocyclophane F. (*B*) DFT optimized ground state of cylindrocyclophane F (B3LYP, MIDIX, D3BJ, PCM=Et_2O) illustrates predicted orientation of the resorcinol rings relative to the alkyl backbone, consistent with crystal structures of related paracyclophanes.[51,52] Oxygen atoms shown as red spheres, carbons in grey, and hydrogens in white.

The cylindrocyclophanes are a family of natural products produced by the fresh water cyanobacterium *Cylindrospermum licheniforme*[i] (Figure 1.4).[51,53] The genus *Cylindrospermum* has been found in damp and flooded soils, and in true aquatic habitats.[54] *C. licheniforme* is reported to natively grow in wet ecosystems throughout Europe and in Brazil, although some of this field work predates contemporary genomic identification.[55] Following detection of antitumor activity against KB and LoVo tumor cell lines from extracts of this organism, the first cylindrocyclophane natural product was isolated and structurally characterized from *C. licheniforme* in 1990 by the Moore and Clardy labs.[51] Notably, this was the first example of biological

[i] Several isolates of *C. licheniforme* produce the cylindrocyclophanes, including ATCC 29412 and 29204 (a.k.a. PCC 7417 and *C. stagnale*). Partial genome sequence of ATCC 29412 is available on IMG-JGI, including most of the *cyl* gene cluster (*cylA–L*). NCBI contains all protein sequences from the *cyl* gene cluster (CylA–P) from *C. licheniforme* 'UTEX B 2014'.

[n.n]paracyclophanes to be discovered. Previously, non-biological paracyclophanes had been chemically synthesized to study the physical properties of these rigid systems, including altered absorption spectra attributed to aromatic ring distortions and through-space π–π interactions, as well as host–guest complex formation in aqueous solution.[56-58] Chiral paracyclophanes are also of interest as binders of asymmetric guest molecules such as amino acids and other non-ionic species.[59]

Notably, several structurally related paracyclophanes are produced by other *Nostocaceae* cyanobacterium, including the nostocyclophanes (A-D),[51,60] halogenated cylindrocyclophanes (A_1-A_4, A_{B4}, C_1-C_4, F_4),[61] carbamidocyclophanes (A-L),[62-64] merocyclophanes (A-D),[52,65] and ribocyclophanes (A-E).[66] (Figure 1.5). All of these natural products share the common [7.7]paracyclophane carbon scaffold but differ in functional groups appended to the core structure. The producing organisms have been isolated from locations as distant as Joliet, Illinois (USA) and Vietnam, suggesting that cylindrocyclophane production is not geographically constrained to one ecosystem. Many of these natural products have demonstrated bioactivities including mammalian cell cytotoxicity and antibiotic activity, however, their native physiologic role(s) remain unknown. Interestingly, both (–)-cylindrocyclophane A and its unnatural enantiomer (+)-cylindrocyclophane A demonstrated comparable growth inhibition of human colorectal carcinoma HCT-116 cells, indicating that chirality does not play a role in cylindrocyclophane cytotoxicity.[67] Despite this, the cylindrocyclophane biosynthetic pathway produces only one enantiomer and it is unknown if this chirality is important for their native function in the producing organisms and associated ecosystems. The methods developed to isolate the cylindrocyclophanes and our work reproducing these isolation techniques suggests that these relatively non-polar metabolites are predominantly associated with the cyanobacterial cell mass. This is in contrast to natural products found in the culture media, as observed with actively exported metabolites such as siderophores and antibiotics.[10,68]

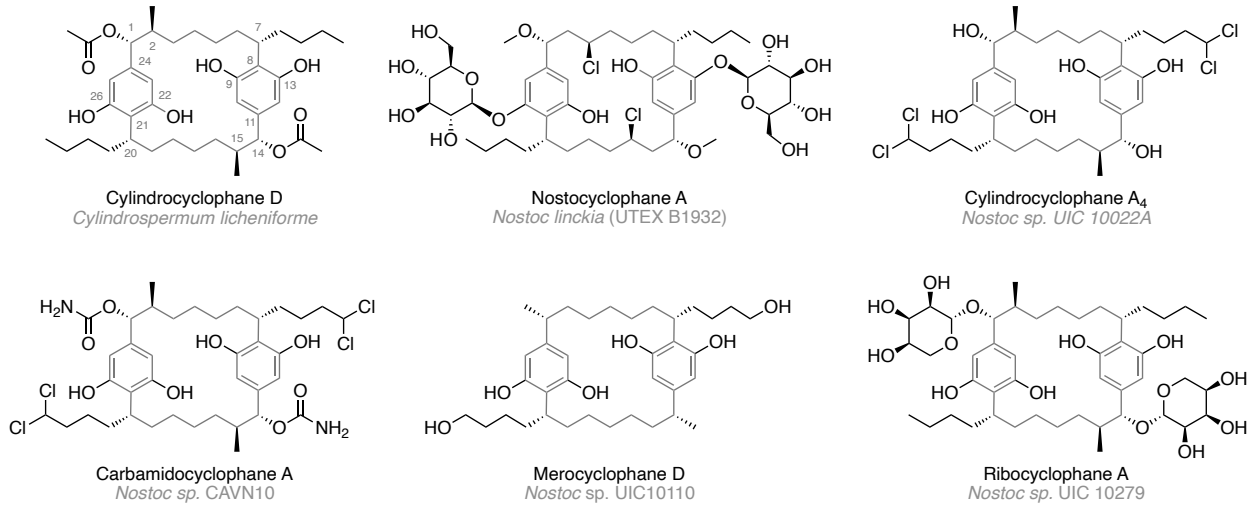


Figure 1.5. Structures of related cyanobacterial cyclophane natural products

The carbon scaffolds of [7.7]paracyclophane natural products vary in halogenation, methylation, and hydroxylation patterns; while their hydroxyl and phenolic groups can be differentially methylated, acetylated, carbamoylated, and glycosylated. Within each family of paracyclophanes, the most elaborately decorated member is displayed. The shared [7.7]paracyclophane scaffold is rendered in grey for clarity.

Elucidating the key steps of cylindrocyclophane biosynthesis occurred nearly three decades after the initial isolation work; this effort revealed novel enzymology involving cryptic halogenation and alkyl halide utilization shared among these related cyclophane natural product families.[69] Initial feeding studies by Bobzin and Moore had established that the cylindrocyclophanes and nostocyclophanes were of polyketide origin (Figure 1.6A), and the symmetric labelling pattern from incorporation of heavy sodium acetate ([1,2-$^{13}C_2$] or [2-^{13}C,d_3]) suggested that the [7.7]paracyclophane core structure arose from a head-to-tail dimerization of alkyl resorcinol monomers with disconnections at carbons 7 and 8, and 20 and 21;[70] however, it was unknown how this cyclization event took place and what precisely the monomeric structures were. Nineteen years later, Nakamura et al. identified the cylindrocyclophane biosynthetic gene cluster in *C. licheniforme* ATCC 29412 by isolating its genomic DNA and utilizing degenerate PCR to amplify putative homologs of hydroxymethylglutaryl-CoA (HMG-CoA) synthase involved in acetate-derived β-methyl installation (Figure 1.6B).[23,37,71,72] These β-methyl groups are present in the cylindrocyclophanes, carbamidocyclophanes, and ribocyclophanes, at carbons 2 and 15 but are absent in the nostocyclophanes

and merocyclophanes. Importantly, the HMG-CoA synthase identified from *C. licheniforme* (*cylF*) was located in a putative BGC with a number of polyketide synthases (PKSs) and potential tailoring enzymes (Figure 1.6C).

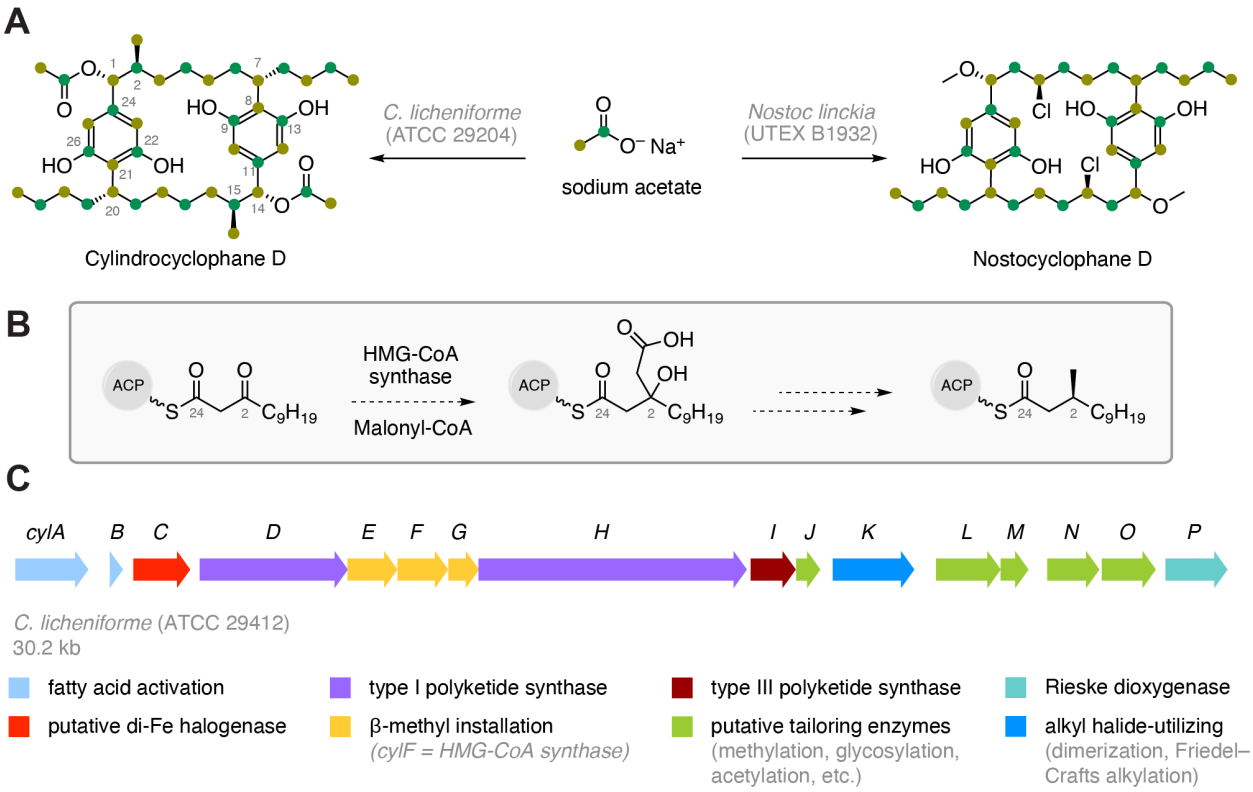


Figure 1.6. Feeding studies and β-methyl installation identify biosynthetic gene cluster

(*A*) *C. licheniforme* and *N. linckia* cultures supplemented with sodium acetate labelled at carbon 2 [2-^{13}C,d_3] or both carbons 1 and 2 [1,2-$^{13}C_2$] incorporated these heavy elements into cylindrocyclophane and nostocyclophane D with a distinctive labelling pattern suggestive of PKS biosynthetic logic and dimerization of alkyl resorcinol monomers to afford the full [7.7]paracyclophane. (*B*) Acetate-derived β-methyl group of the cylindrocyclophanes can be introduced into polyketides with an enzyme homologous to HMG-CoA synthase. (*C*) Annotated *cyl* gene cluster from *C. licheniforme* responsible for cylindrocyclophane biosynthesis containing PKSs and a homolog of HMG-CoA synthase, *cylF*.

Then, using a combination of *in vitro* biochemistry and a feeding study with d_{19}-decanoic acid, Nakamura et al. definitively established the link between this gene cluster and cylindrocyclophane biosynthesis.[71] Notably, they demonstrated that CylI catalyzed the formation of an alkyl resorcinol monomer (**1.2**) from a decanoic acid derived β-ketothioester (**1.1**) *in vitro* (Figure 1.7A); and monomer **1.2** was consistent with the disconnection implied by the earlier Bobzin and Moore feeding studies. Furthermore, the incorporation of d_{19}-decanoic acid into the final natural product as d_{36}-cylindrocyclophane F established that carbons 7 and 20 that undergo C–C bond formation enter the biosynthetic pathway as unactivated methylene groups (Figure 1.7B). Excitingly, this suggested that construction of the [7.7]paracyclophane core structure likely requires enzymatic C–H functionalization events.

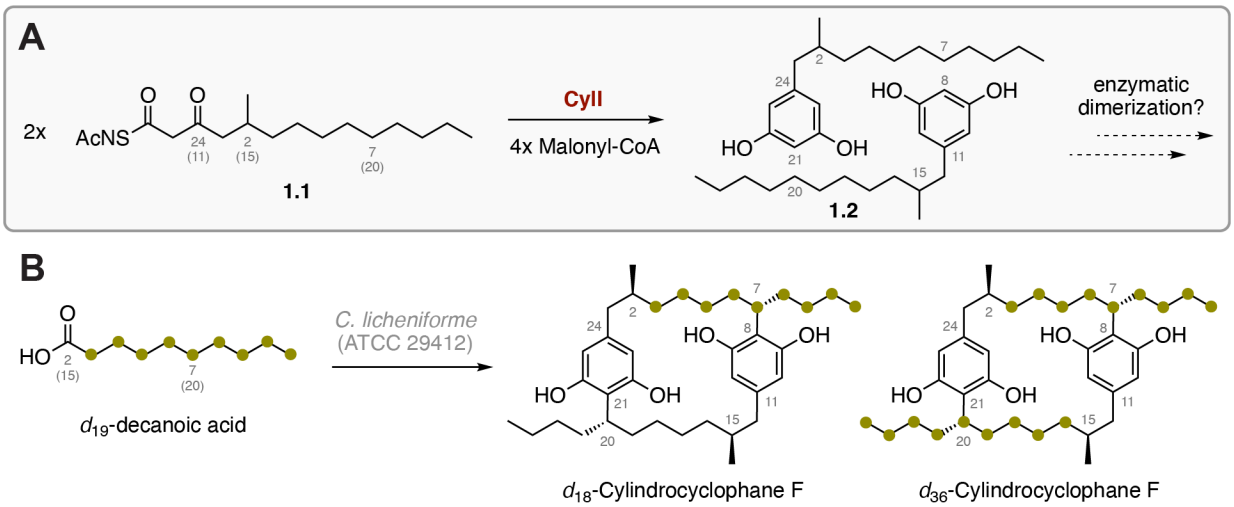
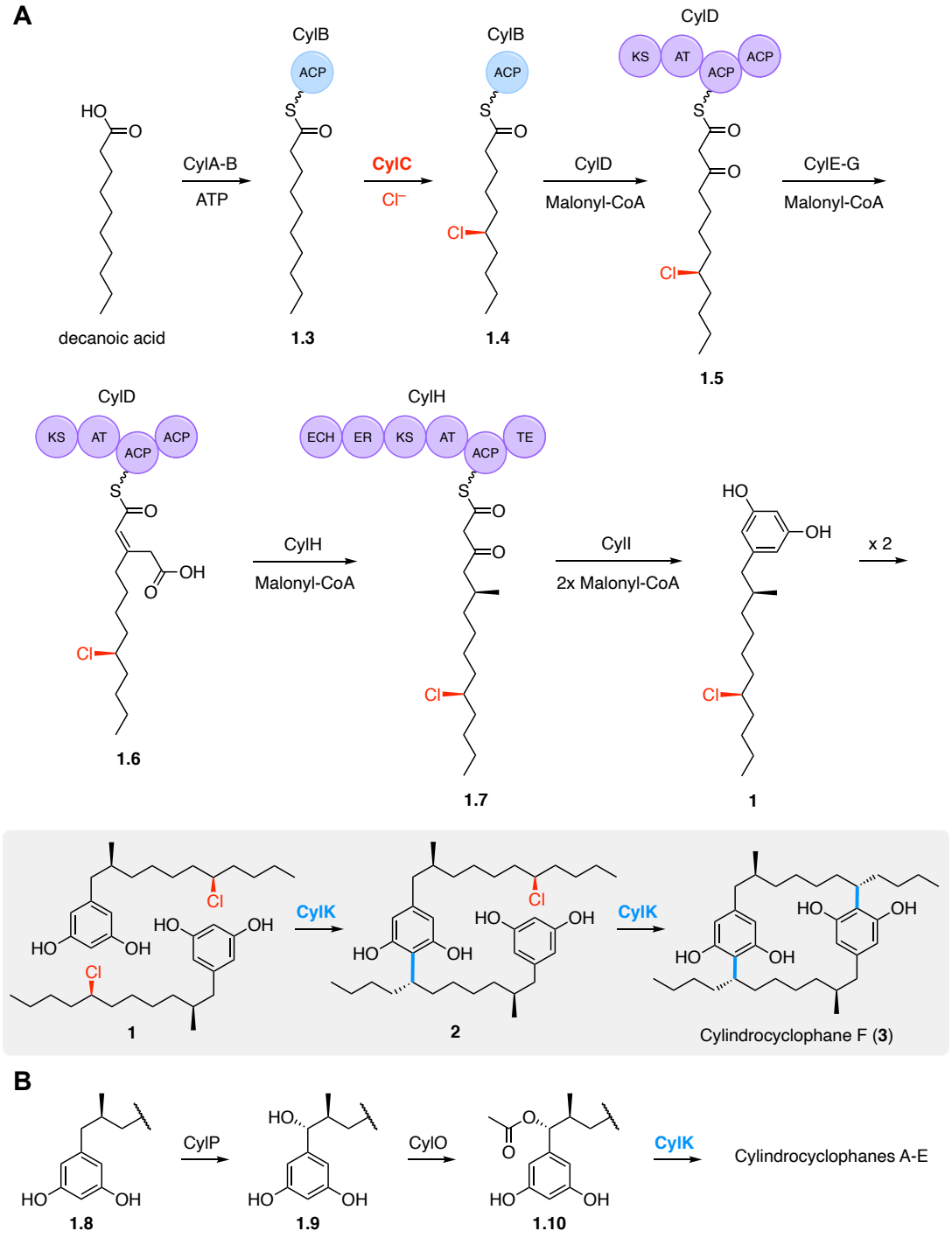


Figure 1.7. *in vitro* biochemistry and feeding study highlights dimerization biosynthetic logic

(*A*) Purified CylI enzyme with SNAc-tethered β-ketothioester intermediate **1.1** and malonyl-CoA produces an alkyl resorcinol monomer **1.2** with disconnections at carbons 7 and 8, and 20 and 21. (SNAc = *N*-Acetylcysteamine) (*B*) *C. licheniforme* cultures supplemented with deuterated decanoic acid incorporated one or two equivalents into cylindrocyclophane F.

Having identified the cylindrocyclophane BGC, the remaining enzymes were examined for their potential roles in dimerization. One enzyme in particular, CylC, is putative diiron-carboxylate enzyme with predicted secondary sequence homology to *N*-oxygenase AurF, an enzyme that catalyzes sequential oxidation of anilines to nitroarenes.[73,74] More broadly, members of this enzyme family can hydroxylate or desaturate unactivated sp^3 carbon centers.[75-77] Notably, more than 30 uncharacterized enzymes with significant homology to CylC are encoded in cyanobacterial genomes (40% amino acid identity, 56% similarity),[69] several of which reside in BGCs associated with chlorinated natural products such as the microginins,[78] columbamides,[79,80] and bartolosides.[81] None of these gene clusters encode known classes of halogenating enzymes, suggesting that these CylC homologs might have chlorinating activity. Upon examination *in vitro*, Nakamura et al. discovered that the enzyme CylC chlorinates a methylene carbon of decanoic acid, tethered via thioester linkage to the ppant arm of the acyl carrier protein (ACP) CylB (**1.3**), suggesting that the dimerization occurs via cryptic halogenation (Scheme 1.2A).[69] Excitingly, the halogenating activity of CylC is a novel function for this enzyme family and represents a new class of halogenase that appears to be operative in other, unrelated cyanobacterial BGCs.



Scheme 1.2. Proposed biosynthesis of cylindrocyclophane F

(*A*) Cylindrocyclophane biosynthesis involves cryptic halogenation, in particular, the halogenase CylC generates an alkyl chloride substrate for CyK, which catalyzes two stepwise Friedel–Crafts alkylations to construct the [7.7]paracyclophane macrocycle of cylindrocyclophane F. CylC and the chloride substituent are depicted in red, CylK and the C–C bond it generates are displayed in blue. (*B*) CylP hydroxylates alkyl resorcinols at the benzylic position, and CylO likely acetylates this group. CylK might then homo or heterodimerize these intermediates to give rise to cylindrocyclophanes A-E, although this order of events has not been demonstrated.

This alkyl chloride intermediate (**1.4**) is then further elaborated by CylD-I, with four equivalents of malonate. Next, CylI terminates and aromatizes the growing polyketide to produce resorcinol-containing alkyl chloride **1**.[82] This intermediate is now primed to be dimerized and form the final alkyl–aryl bonds at the alkyl chloride carbons. However, there was not an obvious enzyme candidate to facilitate the C–C bond formation via activation of the alkyl chloride moiety. A few enzymes could be ruled out as they were not found in related BGCs encoded by other cyclophane producers, namely CylL-O; while CylJ was an unlikely candidate because it has homology to carboxyl and phenol methyltransferases. Ultimately, it was discovered that the novel enzyme CylK dimerizes this intermediate to produce cylindrocyclophane F (**3**) (Scheme 1.2A).[69] The CylK-catalyzed dimerization and cyclization proceeds via two sequential alkylation reactions that displace chloride. CylK first catalyzes an intermolecular alkylation reaction between two equivalents of **1** to generate mono-alkylated intermediate **2**, followed by an intramolecular alkylation and cyclization to afford cylindrocyclophane F (**3**). In this way, cylindrocyclophane biosynthesis is enabled by a cryptic halogenation strategy. The proposed biosynthetic logic is supported by the isolation of analogous monomeric and mono-alkylated natural products from *C. stagnale* PCC 7417.[83]

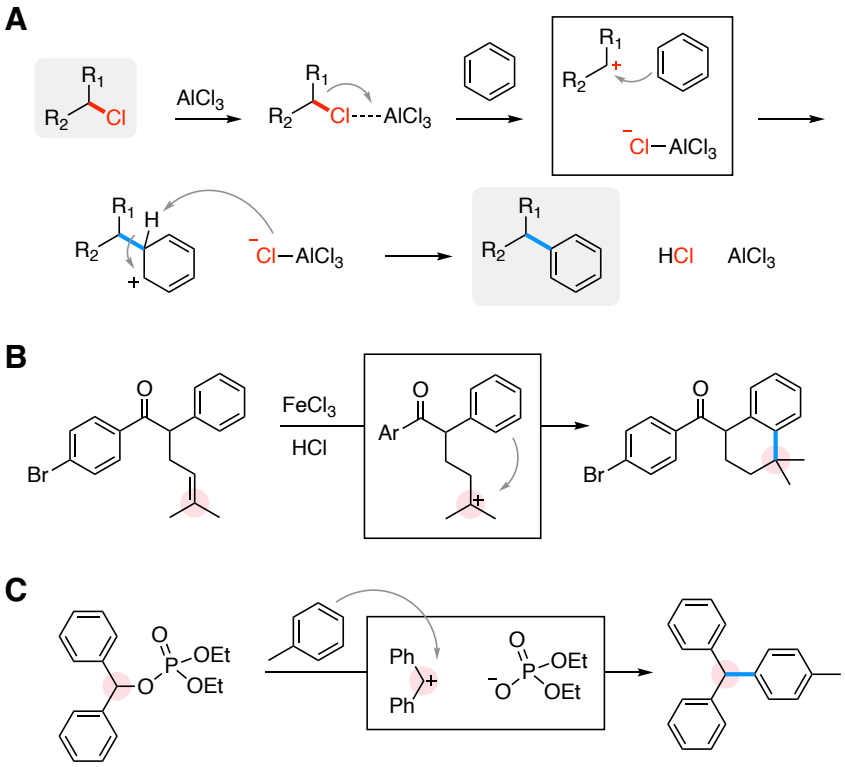
Moreover, it was revealed by chemoenzymatic synthesis and *in vitro* enzymatic assays that CylK only accepts diastereomers of **1** with the *R* configuration the alkyl chloride carbon, demonstrating that the CylK-catalyzed transformation must proceed with inversion of configuration at this alkyl chloride stereocenter (see Scheme 2.2 for more discussion). Furthermore, CylK is the only enzyme known to catalyze arene alkylation with an alkyl halide electrophile, a challenging reaction that mirrors a classical non-enzymatic reaction known as the Friedel–Crafts alkylation. The traditional chemical Friedel–Crafts reaction typically suffers from a lack of stereo- and regiocontrol, unwanted overalkylation events, and the generation of carbocationic intermediates that can undergo unproductive rearrangements (discussed further in Chapter 1.4).[84,85] With this in mind, the selectivity and activity of the CylK-catalyzed reaction is remarkable, and warrants further investigation from both a fundamental and applied perspective. At the time of its discovery, CylK was annotated as a fusion of a Ca^{2+} binding domain and a β-propeller fold, but the roles of these predicted protein domains in catalysis were unknown; and uncharacterized enzymes with homology to CylK

are found in numerous cyanobacteria and we speculated that they might catalyze analogous transformations involved in natural product biosynthesis.

The remaining cylindrocyclophanes (A-E) contain all possible combinations of 1 or 2 hydroxyl groups and/or O-acetyl groups at the benzylic positions of their resorcinol rings. Over the course of examining uncharacterized Cyl enzymes for dimerization activity, Nakamura et al. discovered that *E. coli* cultures heterologously expressing putative Rieske dioxyganse CylP could hydroxylate an alkyl resorcinol substrate (**1.8**) at the benzylic position (Scheme 1.2B).[69] Although they did not perform this reaction with chlorinated alkyl resorcinol **1**, it is reasonable to assume that CylP can similarly hydroxylate this substrate. Of note, CylP hydroxylation activity was not observed on the macrocycle, cylindrocyclophane F (**3**).[86] Additionally, the cylindrocyclophane biosynthetic gene cluster contains a putative acetyl transferase, CylO, that may be responsible for acetylating this benzylic hydroxyl group before or after dimerization and macrocycle formation. Cumulatively, this suggests that CylK can also accept alkyl resorcinol substrates hydroxylated and potentially *O*-acetylated at the benzylic position, although this order of events has not been conclusively demonstrated. If CylK has some tolerance for substrate promiscuity, it could present a useful starting point for its use as a biocatalyst and future directed evolution efforts. Furthermore, improving our understanding of these alkyl halide-utilizing enzymes may inform the design of non-enzymatic catalysts, and enable genome mining to uncover new natural products constructed by related biosynthetic strategies.

1.4 Enzymatic Friedel–Crafts alkylations

The Friedel–Crafts alkylation is widely used to construct diverse alkyl–aryl linkages such as those formed by the CylK-catalyzed steps of cylindrocyclophane biosynthesis.[84,87] Prototypically, this chemical reaction generates C–C bonds by harnessing Lewis acid catalysis to activate alkyl halides or alkenes for electrophilic aromatic substitution by arene nucleophiles.[88] The critical step of this transformation is formation of the carbocationic intermediate from the electrophile coupling partner (Scheme 1.3A). Generation of this charged and highly reactive species then promotes the subsequent addition of the arene nucleophile to the carbocationic electrophile, followed by deprotonation of the arene at the site of alkylation, restoring aromaticity and generating an equivalent of hydrochloric acid (HCl) as a byproduct. While carbocation intermediates are generated from alkyl halides via halide removal, similar carbocations can be accessed from alkenes via protonation (Scheme 1.3B). These carbocationic intermediates can rearrange, sometimes unproductively, via hydride and alkyl shifts. Moreover, the stereo- and regiochemical outcomes are difficult to control, considering the nature of these planar and reactive carbocationic intermediates. Over time, the definition of the Friedel–Crafts alkylation has expanded to accommodate new methodology that takes advantage of pseudo-alkyl halide electrophiles. In one interesting example, benzylic phosphates are harnessed as labile leaving groups with triflic acid at room temperature to generate a stabilized carbocationic intermediate for subsequent arene alkylation (Scheme 1.3C).[89] This use of phosphorylated intermediates neatly parallels a widespread biosynthetic strategy employed by isoprenoid synthases to generate carbocations for coupling and cyclization reactions.[90]

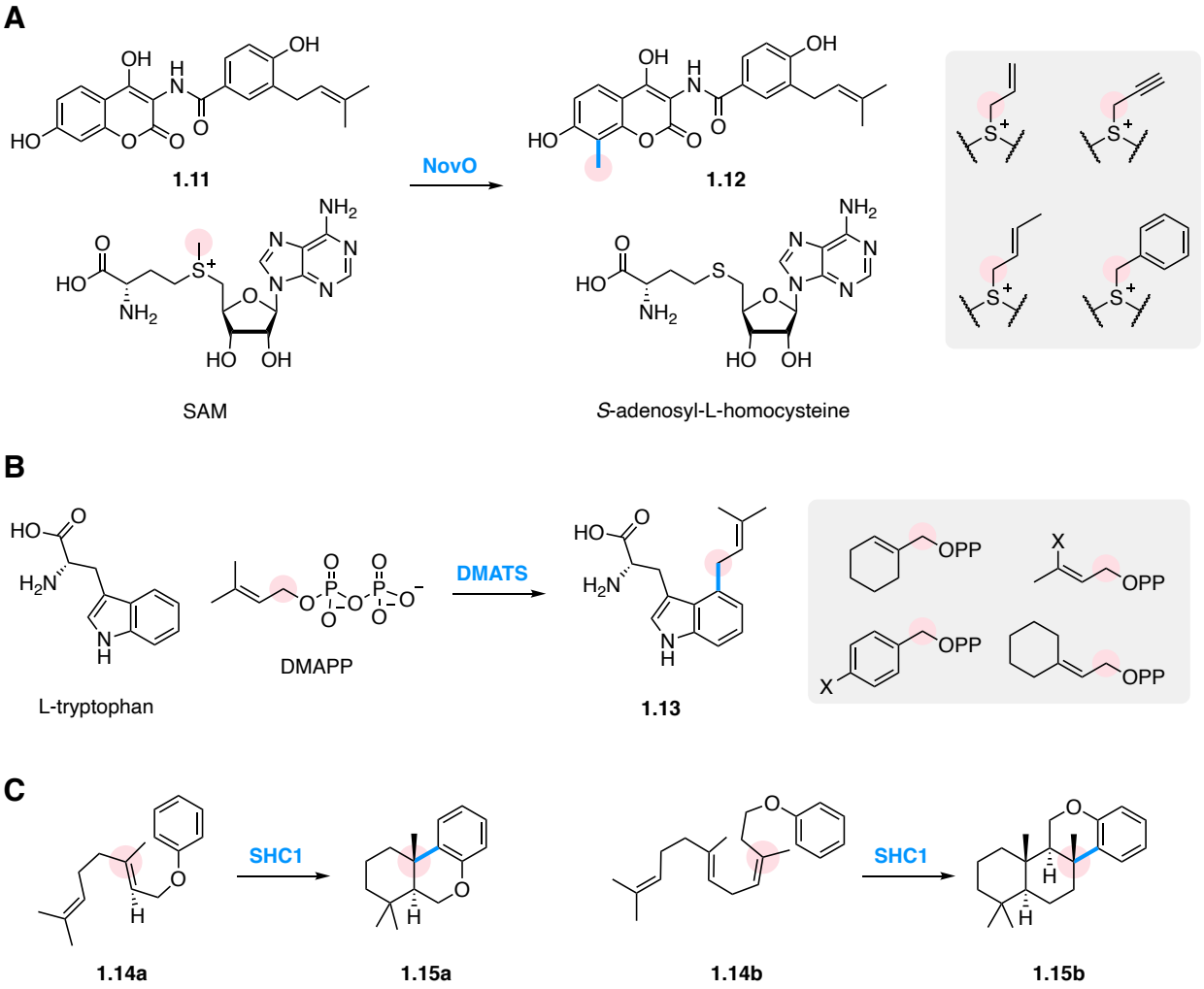


Scheme 1.3. Synthetic Friedel–Crafts alkylation reactions and mechanisms

(*A*) The mechanism of the Friedel–Crafts alkylation reaction with an alkyl chloride electrophile and benzene nucleophile, as originally pioneered by Friedel and Crafts. The alkyl halide is displayed in red, while the C–C bond generated is represented in blue. (*B*) An intramolecular Friedel–Crafts alkylation with alkene electrophile.[91] The electrophilic carbon is highlighted with a light red circle. (*C*) An example of a benzylic phosphate electrophile used to generate the reactive carbocationic intermediate for the Friedel–Crafts alkylation.

Native biological reactions analogous to the Friedel–Crafts alkylation have been observed, and some of the enzymes responsible for this chemistry can be employed as C–C bond-forming biocatalysts.[92] In novobiocin biosynthesis by *Streptomyces spheroides*, SAM is harnessed as a methyl donor by the enzyme NovO in order to alkylate coumarin **1.11**, to afford methylated coumarin **1.12** (Scheme 1.4A).[93] This reaction takes advantage of SAM's highly reactive sulfonium as a leaving group to facilitate the substitution by the coumarin nucleophile. Deprotonation of the phenol and an S_N2-like substitution reaction are proposed as the key mechanistic steps of this enzyme, although a concerted reaction is plausible.[94] A similar methylation strategy is also present in coumermycin biosynthesis catalyzed by the enzyme CouO from *S. rishiriensis*. A number of non-biological SAM analogs with different alkyl groups have been chemically synthesized

(from alkyl halides) and can be used as electrophilic substrates by these *C*-methyl transferases (Scheme 1.4A).[95,96]



Scheme 1.4. Enzymatic electrophilic aromatic substitution reactions

(*A*) SAM-dependent methyltransferase NovO catalyzes the methylation of coumarin **1.11**. Non-native SAM electrophiles are accepted by this enzyme. The electrophilic carbon is highlighted with a light red circle, while the C–C bond generated is represented in blue. (*B*) Prenyltransferases use native and non-biological allyl diphosphates as electrophilic substrates to selectively alkylate tryptophan. (*C*) Wild-type squalene cyclase (SHC1) catalyzes non-biological cyclization and electrophilic aromatic substitution of polyprenyl phenyl ethers **1.14a-b**.

The arene cores of numerous indole alkaloid natural products are prenylated, and a family of *C*-prenyltransferase enzymes natively performs these biochemical Friedel–Crafts alkylation reactions.[97] The first committed step of ergot alkaloid biosynthesis by *Claviceps purpurea* is the C4 prenylation of tryptophan by the enzyme DMATS (4-dimethylallyltryptophan synthase, a.k.a. FgaPT2) with dimethylallyl diphosphate (DMAPP) as the electrophilic substrate to afford prenylated tryptophan **1.13** (Scheme 1.4B).[98] Based on biochemical work and a structure of this enzyme with substrate analogs, it is proposed that DMATS leverages the reactivity of the electrophile's diphosphate group (OPP) in order to generate a carbocationic intermediate that is then subjected to electrophilic aromatic substitution by *N*-deprotonated tryptophan.[99,100] However, a more concerted or S_N2-like substitution mechanism without a discrete carbocation is also plausible. Notably, additional prenyltransferases have been characterized that can selectively alkylate at the other six positions of the indole ring; several tryptophan nucleophile derivatives and allyl diphosphate electrophiles are accepted by these enzymes (Scheme 1.4B).[101] Related prenyltransferases can perform analogous alkylation reactions with distinct phloroglucinol nucleophiles.[102,103]

In an interesting example, the squalene hopene cyclase from *Zymomonas mobilis* (SHC1) can be used as a biocatalyst to promote non-native Friedel–Crafts alkylation reactions with polyprenyl phenyl ethers **1.14a-b** (Scheme 1.4C).[104] This affords cyclized products **1.15a-b** in low yield, in addition to side products where aromatic alkylation did not occur. It is proposed these reactions proceed in a manner similar to squalene cyclization,[105] in which the reaction is initiated by protonating a double bond in order to generate a reactive carbocation electrophile, followed by a cascade of electrophilic additions followed by electrophilic aromatic substitution. In the native squalene cyclization, the reaction is terminated by elimination or hydration of the final carbocation to yield an olefin or an alcohol moiety, respectively.

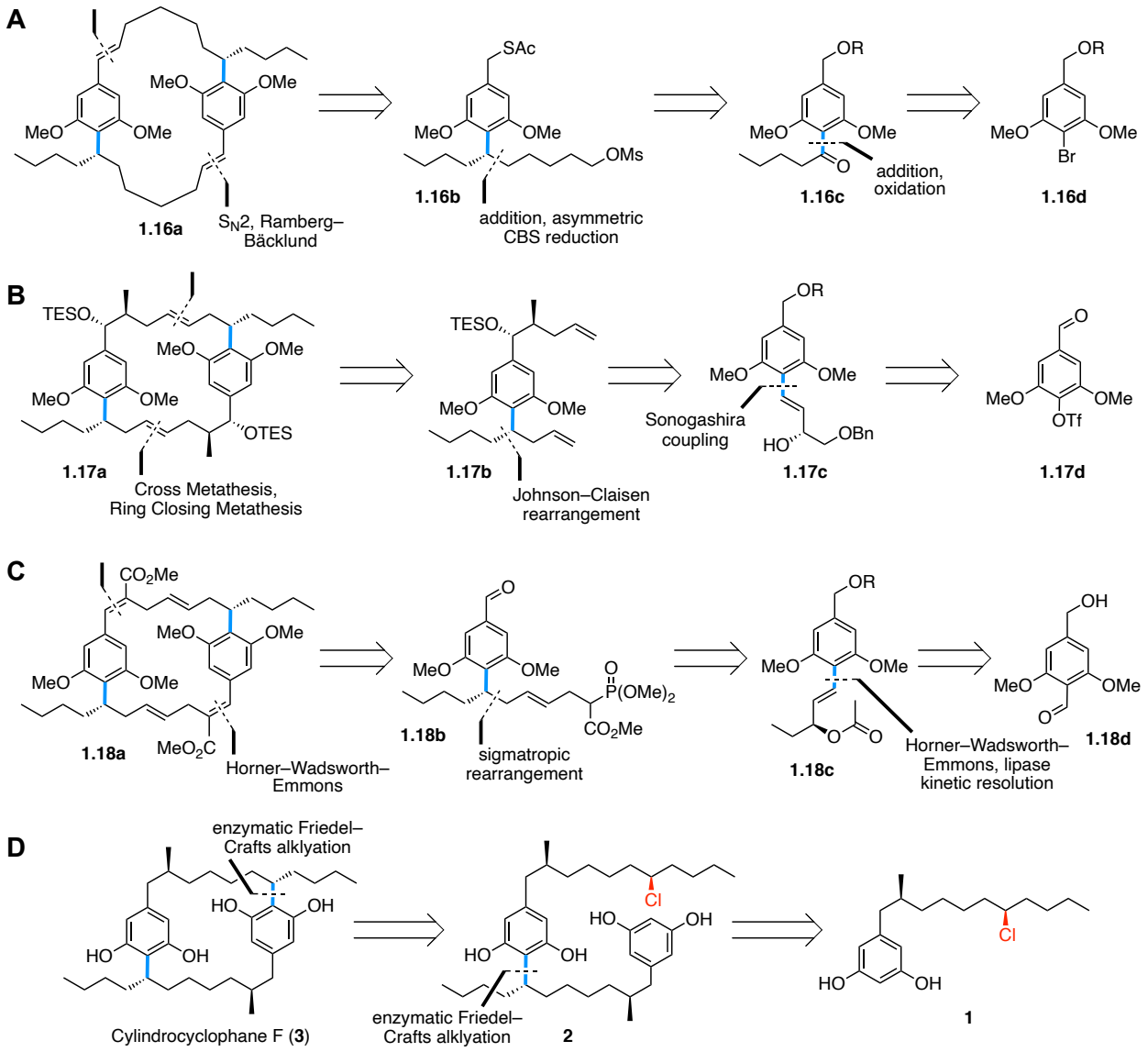
These biological examples of enzymatic electrophilic aromatic substitution are representative of cation- and electrophile-driven reaction manifolds. Although the alkyl–aryl bonds formed are analogous to the C–C bonds found in the cylindrocylophanes, these enzymes require pre-activated electrophilic substrates such as SAM or allyl diphosphates to initiate substitution and/or carbocation generation. These electrophiles are

relatively labile, and the kinetic barrier to reaction is minimized. Furthermore, from a synthetic perspective, these substrates are less accessible than alkyl halides. In contrast, the CylK-catalyzed alkylation reaction enzymatically activates relatively inert alkyl chlorides for electrophilic aromatic substitution. CylK is the only enzyme known to do this, and structures and mechanisms of enzymes that engage secondary alkyl halides are unknown.[ii]

1.5 Synthetic strategies to form alkyl–aryl C–C bonds

Chemists have developed powerful strategies to assemble C–C bonds, and these methodologies enable academic research, pharmaceutical discovery, and chemical manufacturing.[106,107] However, despite innovations, constructing asymmetric C(sp^2)–C(sp^3) bonds with modern cross coupling methods remains a challenge.[108,109] In particular, the chiral alkyl–aryl linkages found in the cylindrocyclophanes are difficult to access due to steric congestion on the arene and the need for stereocontrol. With this challenge in mind, a number of research groups have targeted these natural products for total synthesis. In one instance, Nicolaou and coworkers assembled cylindrocyclophanes A and F through a head-to-tail S_N2 cyclodimerization and Ramberg–Bäcklund reaction of monomer **1.16b** (Scheme 1.5A).[110] This dimerization event forms two C–C bonds that are subsequently reduced to an sp^3–sp^3 linkage in the final [7.7]paracyclophane. Proceeding this dimerization, the key secondary alkyl (2° alkyl)–aryl linkage is formed in three steps, first the addition of aryl bromide **1.16d** to pentanal, followed by oxidation to ketone **1.16c** and a subsequent addition step. The chirality at this carbon center is set with an asymmetric Corey–Bakshi–Shibata (CBS) reduction to ultimately afford a single enantiomer of monomer **1.16b**.

[ii] See Chapter 3.1 for a discussion of structurally characterized enzymes that engage primary alkyl halide substrates.



Scheme 1.5. Retrosyntheses of cylindrocyclophane macrocycle and alkyl–aryl bond

Disconnections affording final macrocycle or synthetic precursor and key 2° alkyl–aryl linkage from cylindrocyclophane total syntheses by (*A*) K. C. Nicolaou et al., (*B*) Y. Iwabuchi, (*C*) T. R. Hoye, and (*D*) the biosynthesis by *C. licheniforme.* The alkyl halide is displayed in red, while the 2° alkyl–aryl C–C bond is represented in blue.

With a distinct head-to-tail disconnection, the Iwabuchi and Smith groups independently targeted the cylindrocyclophane natural products with a cross metathesis and ring-closing metathesis strategy (Scheme 1.5B).[67,111,112] Notably, an earlier route by Smith and coworkers was the first reported total synthesis of any cylindrocyclophanes, employing a ring-closing metathesis step following a pseudo-dimerization via a Myers

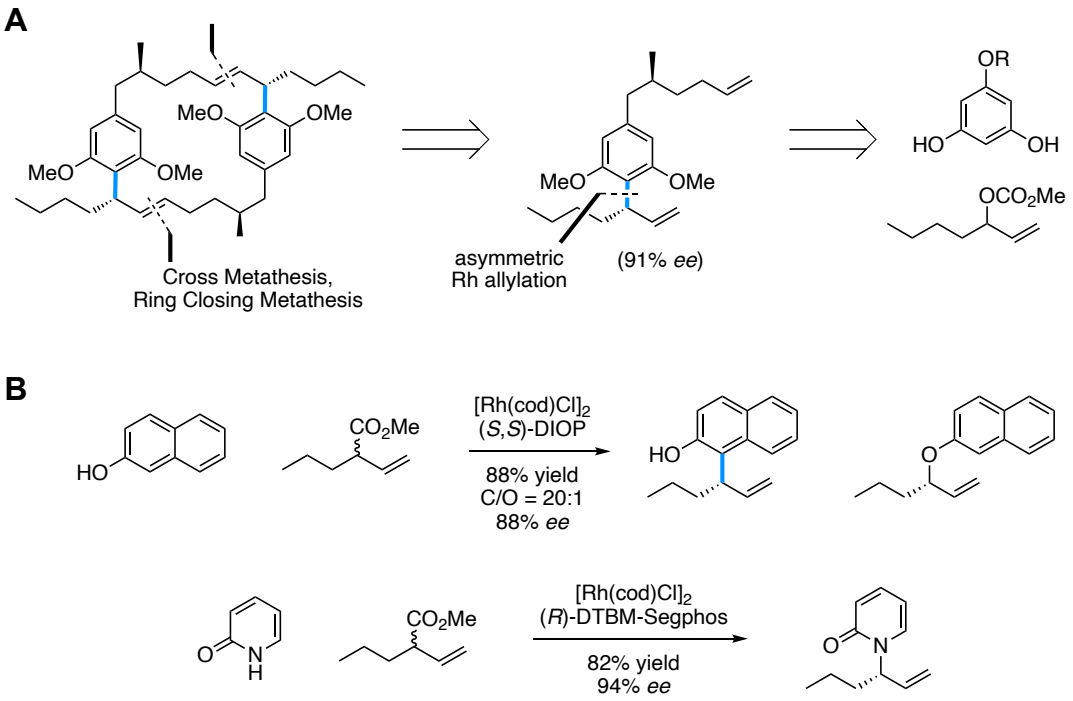
reductive coupling.[113] Building from this initial route, Smith et al. then developed the cross metathesis and ring-closing metathesis strategy.[111,112] Years later, Iwabuchi et al. similarly assembled cylindrocyclophane A from diolefin monomer **1.17b**. This metathesis dimerization forms two alkenes that are eventually reduced to a $C(sp^3)$–$C(sp^3)$ bond of the cylindrocyclophane alkyl backbone. Proceeding this, the intermediate is assembled in several steps. First, a chiral acetylene is coupled with aldehyde **1.17d** via Sonogashira reaction and reduced to afford **1.17c**. The second arm of the alkyl group is then installed with a Johnson–Claisen rearrangement followed by additional manipulations to access diolefin monomer **1.17b**.

In another total synthesis, the Hoye group developed a route to cylindrocyclophane A that exploited a double Horner–Wadsworth–Emmons head-to-tail dimerization of bifunctional monomer **1.18a** (Scheme 1.5C).[114] This dimerization constructs two alkenes that are eventually reduced to a $C(sp^3)$–$C(sp^3)$ bond in the final product. The positioning of these alkenes on the alkyl backbone enables the installation of the benzylic hydroxyl groups found in cylindrocyclophane A. Before dimerization, the secondary alkyl group is assembled on the arene by coupling alcohol **1.18d** with a ketone-containing phosphonate via a Horner–Wadsworth–Emmons reaction. Of note, chirality is introduced by reducing the ketone, acetylating the resultant hydroxyl group, and then performing a lipase catalyzed kinetic resolution to afford a single enantiomer of **1.18c**. This primary alkyl group is transformed into a secondary alkyl group by promoting a sigmatropic rearrangement. This rearrangement efficiently transfers chirality from the O-acetyl group, and additional manipulations afford the bifunctional monomer **1.18b**.

Although each synthetic route to the cylindrocyclophanes is distinct, they all employ strategies that require pre-installation of the $C(sp^2)$–$C(sp^3)$ chiral center prior to a dimerization step. This is in contrast to the biosynthetic route of *C. licheniforme* and other *Nostocaceae* cyanobacterium, in which two resorcinol-containing alkyl chloride monomers (**1**) are dimerized by the enzyme CylK to sequentially form the two alkyl–aryl linkages found in the natural products (**3**) (Scheme 1.5D). In this biosynthesis, the asymmetry at

the alkyl chloride stereocenter is introduced via enzymatic halogenation,[iii] and the stereochemical configuration of the alkyl chloride is inverted during the CylK-catalyzed Friedel–Crafts alkylation. This transformation would be challenging to perform with existing synthetic methodology due to the limitations of the chemocatalytic Friedel–Crafts alkylation, and the difficulties in using alkyl chlorides for other cross coupling reactions. Furthermore, the formation of the chiral $C(sp^2)$–$C(sp^3)$ linkage requires lengthy manipulations, as reported in earlier studies. Notably, in a more recent cylindrocyclophane total synthesis by Berthold and Breit, this 2° alkyl–aryl linkage is established in one step with high enantiomeric excess (ee) of 91% (Scheme 1.6A).[115] Their route takes advantage of the cross metathesis and ring closing metathesis strategy originally employed by the Iwabuchi and Smith groups. Enabling monomer synthesis, they developed a rhodium (Rh)-catalyzed asymmetric allylation of phenols and 2-hydroxypyridines with esters and carbamates (Scheme 1.6B).[116] This work represents a notable advance in asymmetric alkyl–aryl coupling chemistry.

[iii] The stereochemical outcome of the reaction catalyzed by putative diiron-carboxylate halogenase CylC has not been definitely determined but we can deduce it from the stereoselectivity of CylK. Furthermore, the absolute configuration of the internal alkyl chloride of the columbamides is known, and this carbon center is likely halogenated by an enzyme homologous to CylC.

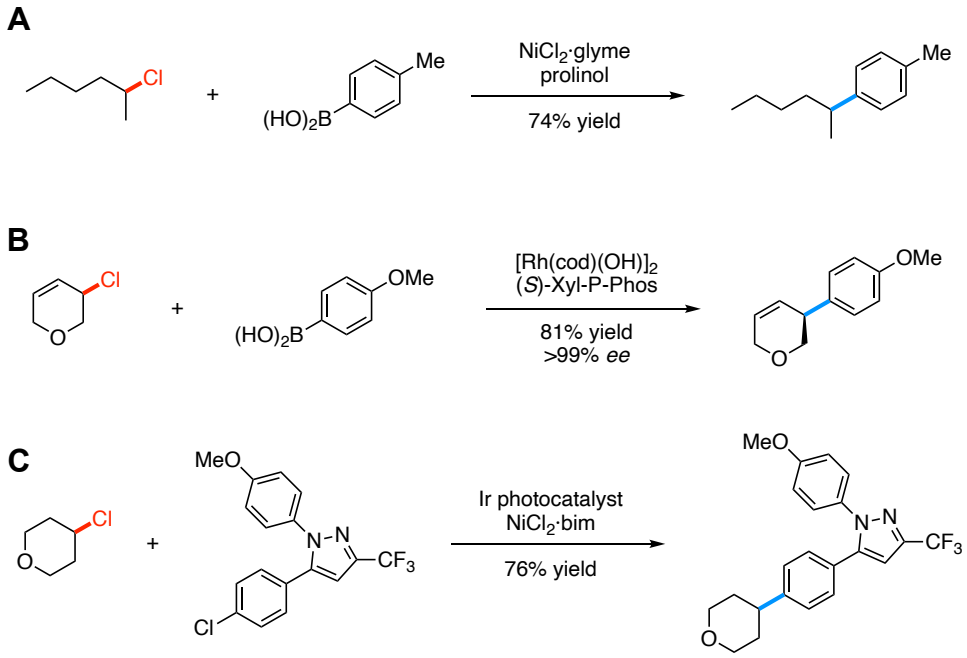


Scheme 1.6. Total synthesis of cylindrocyclophane F with asymmetric aryl allylation

(*A*) Retrosynthesis of synthetic precursor to final natural product emphasizing head-to-tail dimerization and key 2° alkyl–aryl linkage from B. Breit et al. (*B*) Rhodium-catalyzed asymmetric aryl allylation developed by the same group. The 2° alkyl–aryl C–C bond is represented in blue.

The use of alkyl chlorides as substrates for coupling reactions has not been as widely explored compared to their alkyl bromide and iodide counterparts.[117,118] Fu and coworkers have pioneered the use of alkyl chlorides to construct C(sp^2)–C(sp^3) linkages via the Suzuki cross-coupling manifold. In one example, they found that nickel catalysts with amino alcohol ligands can activate primary and secondary alkyl halides for Suzuki reactions with a variety of aryl boronic acids (Scheme 1.7A).[119] Fu and others have developed analogous conditions to access asymmetric alkyl–alkyl coupling products from unactivated alkyl chloride electrophiles.[120] In another instance, Fletcher and coworkers have developed Rh-catalyzed methodology to asymmetrically couple racemic allyl chlorides to aryl boronic acids (Scheme 1.7B).[121] Of note, both enantiomers of allyl chloride are converted to a single enantiomer of product. Although allyl chlorides are inherently more reactive than alkyl chlorides, this method is a useful strategy to access asymmetric alkyl–aryl linkages found in natural products and pharmaceuticals.[122] More recently, MacMillan and coworkers

have harnessed unactivated alkyl chlorides in photocatalytic cross-electrophile coupling reactions and used this methodology to access medicinally relevant compounds (Scheme 1.7C).[123] Although this specific method affords racemic products, this represents a novel coupling manifold that may be rendered asymmetric in the future.



Scheme 1.7. Alkyl–aryl coupling reactions with secondary alkyl and allyl chlorides

Representative reactions from (*A*) G. C. Fu, (*B*) S. P. Fletcher, and (*C*) D. W. C. MacMillan and coworkers highlighting methodology developed to engage secondary alkyl and allyl chlorides as coupling partners. The alkyl halide is displayed in red, while the C–C bond generated is represented in blue.

Alkyl chlorides offer several advantages over other electrophilic substrates, including their availability from commercial and biological sources, reduced toxicities, and increased chemical stability over the course of total chemical synthesis.[123-125] However, this stability presents a challenge in activating alkyl chlorides as substrates, considering that C–Br and C–I bonds are inherently weaker and better electrophiles than C–Cl bonds.[126,127] This highlights why alkyl bromides and iodides are more widely used as electrophilic coupling partners than alkyl chlorides. With this in mind, it is remarkable that the enzyme CylK can activate relatively

inert C–Cl bonds at room temperature for a stereospecific aryl substitution. This further motivates our work to determine the biocatalytic potential, structure, and mechanism of this powerful enzyme.

1.6 Chapter preview

The overarching goal of this thesis is to characterize the cyanobacterial enzyme CylK that plays a key role in cylindrocyclophane biosynthesis. The reaction catalyzed by CylK is unprecedented in the biochemical literature; it is the only enzyme known to catalyze aromatic ring alkylation with an alkyl halide electrophile, mirroring the classical Friedel–Crafts alkylation. We evaluate its substrate scope and mechanistic requirements, determine its structure via X-ray crystallography (in collaboration with the Boal Lab at Penn State University), and access CylK mutants with enhanced activity for further directed evolution.

In **Chapter 2**, we develop and optimize protein preparation methods to access CylK, and we probe the non-biological substrate scope of this enzyme in order to assess its utility as a biocatalyst. Through this work, we begin to understand CylK's binding and recognition elements, and its electronic requirements for substrate engagement. Furthermore, we determine the regio and stereoselectivity of a model non-native reaction, and use CylK to access derivatives of a therapeutic compound. CylK's tolerance for non-native substrates presents a useful starting point for protein engineering.

Chapter 3 focuses on solving the structure of CylK and identifying its active site and key catalytic residues responsible for resorcinol nucleophile and alkyl chloride electrophile activation. Based on this information, we propose a mechanism for the CylK-catalyzed alkylation and provide *in silico* modelling and bioinformatic support. This work further sheds light on the reactivity of homologous alkyl halide-utilizing enzymes found in numerous other cyanobacteria. We also identify a new reaction catalyzed by a CylK homolog. This work enhances our understanding of how secondary alkyl halides can be activated and harnessed by enzymes as coupling partners.

Finally, in **Chapter 4**, we develop methods to screen enzyme libraries for enhanced catalytic properties and the production of bioactive products, construct random and site saturation mutagenesis libraries of CylK, and identify two enzyme variants with superior activity in plate-based and cell lysate reactions. This work will enable future directed evolution of CylK and homologs. Finally, we propose future directions for this project that focus on applying insights from our structural, mechanistic, and bioinformatic work to guide protein engineering, natural product discovery, and further enzyme characterization.

2 Elucidating the substrate scope of CylK and homologous enzymes

Components of this chapter were adapted from a previous publication.[128]

Portions of the work reported were conducted by other lab members and collaborators including Dr. Erica Schultz, Dr. Michael U. Luescher, Harry H. Hager, João P. A. Reis, and Dr. Hitomi Nakamura. Their contributions will be recognized as footnotes and citations throughout the chapter.

2.1 Introduction

We set out to characterize the unique and powerful carbon–carbon (C–C) bond forming reaction catalyzed by the enzyme CylK. This cyanobacterial enzyme assembles the cylindrocyclophane natural products by performing two sequential Friedel–Crafts alkylation reactions, dimerizing alkyl resorcinol substrates by forming new C–C bonds between aromatic rings and secondary alkyl chlorides. (Scheme 2.1A).[69] This transformation is unprecedented in biology, and the enzyme itself might find use as a biocatalyst. Other biochemical reactions analogous to the Friedel–Crafts alkylation have been reported, however, they require highly activated electrophiles that are not easily amenable to synthesis.[iv] Remarkably, CylK has evolved to activate relatively inert and synthetically accessible alkyl chlorides in a highly selective fashion. Notably, CylK alkylates at the 2-position of the resorcinol ring, and the substitution reaction occurs with inversion of configuration at the alkyl chloride electrophile stereocenter. Despite the intriguing reactivity and potential applications of this enzyme, the substrate requirements and mechanistic details of this biochemical reaction remain unknown. Furthermore, accessing the enzyme via traditional heterologous expression methods is challenging.

In this Chapter, we first highlight our efforts to improve the enzyme preparation. With reliable access to purified enzyme, we then set out to elucidate the non-biological substrate scope and selectivity of this enzymatic reaction in order to evaluate its potential application as a biocatalyst, and to begin developing an understanding of its putative mechanistic and substrate binding requirements that may influence the

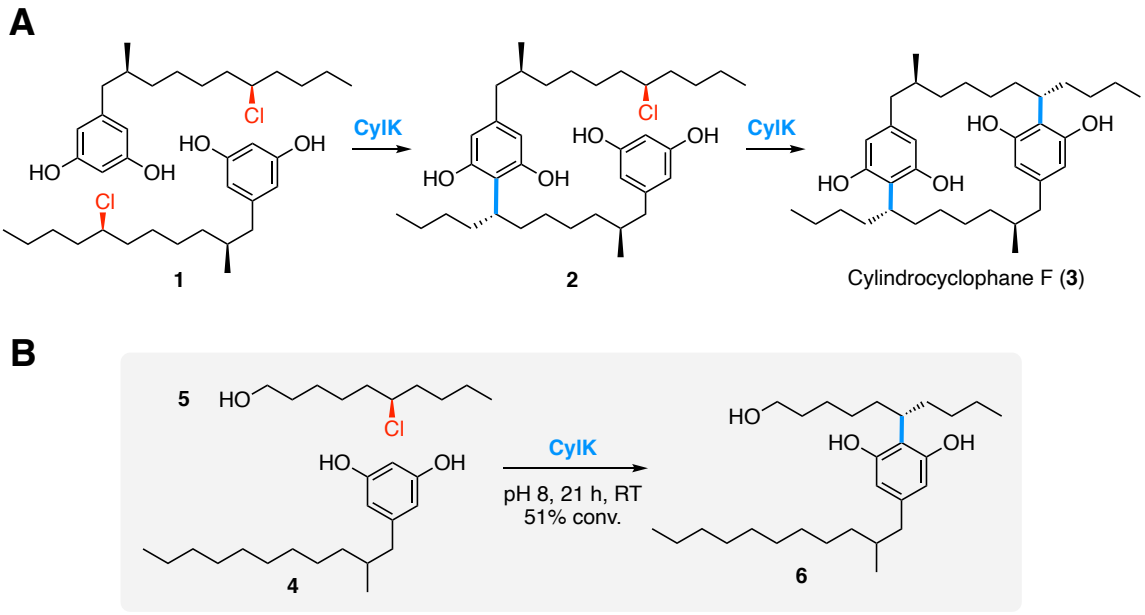
[iv] See Chapter 1.4 for a discussion of other enzymatic Friedel–Crafts alkylations.

native enzyme's substrate tolerance and reactivity. This work has enhanced our fundamental understanding of CylK and how it may catalytically engage resorcinol and alkyl halide substrates.

Contemporary synthetic methods to efficiently and selectively form bonds between carbon atoms have revolutionized small molecule synthesis, empowering academic research, pharmaceutical discovery, and chemical manufacturing.[106,107] Although these advances have been celebrated with Nobel Prizes, challenges remain in asymmetrically constructing C(sp^2)–C(sp^3) bonds with modern cross coupling strategies.[108,109] Asymmetric alkyl–aryl linkages, such as those of the cylindrocyclophanes, are particularly challenging to access. This is evidenced by lengthy total syntheses of these natural products that employ macrocycle disconnections distinct from their biosynthesis in *C. licheniforme* and other cyanobacteria.[v] First discovered in 1877,[84] the Friedel–Crafts alkylation has been widely used to construct such alkyl–aryl linkages,[87] forming C–C bonds between nucleophilic aryl rings and an electrophilic alkyl halides or alkenes. This reaction is promoted by strong Lewis acid catalysts such as aluminum chloride.[88] Although vastly improved since its initial discovery, this reaction suffers from several limitations including a lack of stereo and regiocontrol, the fact that multiple alkylation events are more likely to occur than single ones, and the requirement for strong Lewis acids to activate alkyl halide electrophiles, which generates carbocation intermediates that can undergo unproductive rearrangements. Enzyme-based catalysts such as CylK provide a potential solution to this challenge. Biocatalysis can enable efficient and highly selective transformations on multiple reaction scales and it is growing in popularity due to increased access to powerful enzymes and robust protein-engineering platforms.[129] Biocatalysts are additionally attractive as alternatives to classic organic and organometallic catalysts due to their minimal toxicity and the use of mild reaction conditions.[130] With this in mind, we set out to determine if CylK could be potentially used as a biocatalyst by characterizing its non-biological substrate scope and determining if these non-native reactions maintain the exquisite selectivity of the native reaction.

[v] See Chapter 1.5 for further discussion of synthetic strategies to assemble cylindrocyclophane natural products.

In their initial CylK discovery paper, Nakamura et al. preliminarily established that a pair of discrete resorcinol nucleophile and alkyl chloride electrophile substrates are accepted by the enzyme with modest conversions (~50%, Scheme 2.1B).[69] They challenged the native enzyme with resorcinol nucleophile **4**, which lacks the alkyl chloride substituent, and alkyl chloride electrophile **5**, which has a hydroxyl group in place of the resorcinol ring. These substrates resemble the native bifunctional substrate (**1**) but each individually cannot be homodimerized by CylK. The progress of this non-biological reaction was measured by HPLC with UV absorbance detection at 210 nm, analogous to how the native reaction was monitored *in vitro*, but without a product standard. Gratifyingly, a new peak appeared only when the two substrates and CylK enzyme were included. The retention time of this new peak (~20 min) suggested it was more hydrophobic than resorcinol **4** (~18.5 min), as expected for the more hydrophobic product (**6**). LC-MS analysis confirmed that the parent mass of this new peak matched product **6**; $[M\text{-}H^+]$ = 433.3687. These results preliminarily suggested that product **6** was formed, and that the enzyme was capable of coupling a non-biological substrate pair, however, it was not determined if this model reaction involved alkylation at the 2-position of the resorcinol ring, and if the substitution reaction occurred with inversion of configuration at the alkyl chloride electrophile stereocenter, as is observed with the native substrate.



Scheme 2.1. The role of CylK in cylindrocyclophane biosynthesis and non-biological coupling

(*A*) CylK catalyzes two sequential head-to-tail Friedel–Crafts alkylations to construct the paracyclophane macrocycle of the cylindrocyclophanes. This involves an intermolecular reaction between two resorcinol-containing alkyl chloride substrates (**1**) to generate intermediate **2**, followed by an intramolecular alkylation to afford cylindrocyclophane F (**3**). (*B*) CylK reactivity toward discrete resorcinol and alkyl chloride substrates suggests that the enzyme might have promise as a biocatalyst.

Nakamura et al. definitively established that the native CylK reaction proceeds with inversion of stereochemical configuration at the alkyl chloride stereocenter by chemoenzymatically synthesizing all four possible diastereomers of resorcinol-containing alkyl chloride substrate **1** (Figure 2.1). The key stereocenter in question is the carbon bearing the alkyl chloride electrophile, although they also examined the β-methyl group to determine if its orientation influences the outcome of the CylK-catalyzed reaction. The absolute configuration of the natural product had been clearly established and confirmed by the original isolation and characterization work and several total syntheses.[51,53,67] Nakamura et al. demonstrated that after exposing the mixture of all possible diastereomers of substrate **1** to CylK, only two isomers were appreciably consumed, both with the (*R*)-configuration at the chlorinated carbon: (*S*,*R*)-**1** and (*R*,*R*)-**1**.

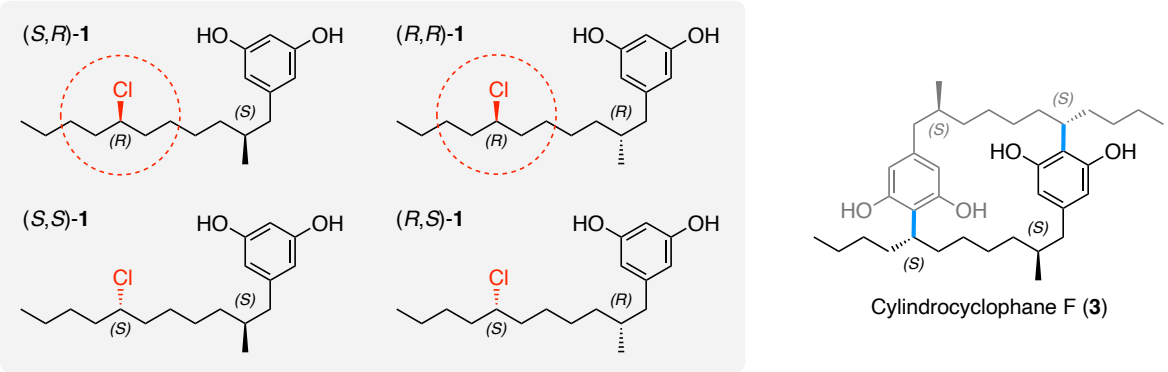
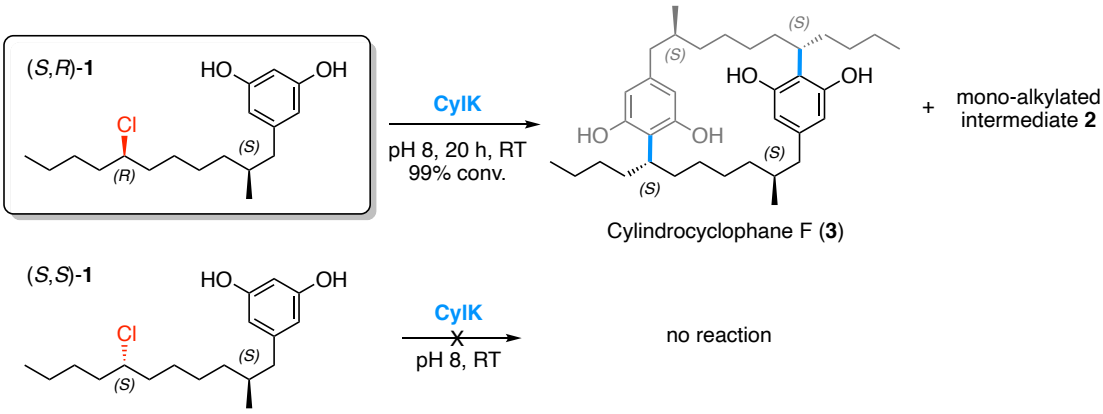


Figure 2.1. The four possible diastereomers of CylK's native monomeric substrate

CylK natively dimerizes (*S*,*R*)-**1** to produce mono-alkylated intermediate **2** and cylindrocyclophane F (**3**), however it can promiscuously utilize (*R*,*R*)-**1** as a bifunctional substrate because it has the same (*R*)-configuration at the chlorinated carbon. These two diastereomers with (*R*)-chlorides are highlighted with red dotted circles. All alkyl chloride bonds are displayed in red, while the C–C bonds formed by CylK are represented in blue.

Then, they individually challenged the enzyme with (*S,R*)-**1** or (*S*,*S*)-**1**, with both configurations at the chlorinated carbon, and only (*S,R*)-**1** produced mono-alkylated intermediate **2** and cylindrocyclophane F (**3**) (Scheme 2.2). Considering that this is the opposite configuration found in the final natural product, Nakamura et al. concluded that CylK must invert the stereocenter over the course of its reaction. This might suggest that CylK catalyzes a substitution reaction with a mechanism akin to the canonical S_N2 reaction, or similar in principle to the inverting glycoside hydrolases.[131] Alternate mechanisms evoking a stabilized carbocation or covalent enzyme intermediates are also plausible. The other isomer consumed, (*R*,*R*)-**1**, corresponds to the (*R*)-configuration at the alkyl chloride carbon but the opposite stereochemical orientation at the β-methyl group compared to the natural product. This suggests that CylK has some promiscuity for non-native substrates, or at the very least, the enzyme tolerates substrate modifications distal to the reaction center. Furthermore, based on the accumulation of unreactive mono-alkylated intermediate **2** when CylK is challenged with the diastereomeric mixture, we speculate that all four diastereomers are competent nucleophilic substrates while only the two with (*S*)-alkyl chlorides cannot be engaged by CylK as electrophiles. Cumulatively, this work suggests that CylK is stereoselective for the (*R*)-alkyl chloride of its native substrate.



Scheme 2.2. Stereoselectivity and stereoinversion of the CylK-catalyzed reaction

After 20 h, CylK dimerizes (*S,R*)-**1** to produce cylindrocyclophane F (**3**), and trace amounts of mono-alkylated intermediate **2**. No reaction is observed with the diastereomer of substrate **1** with an (*S*)-chloride, (*S,S*)-**1**, suggesting that CylK is stereoselective for the (*R*)-configuration at the chlorinated carbon.

Building from this foundational work, we sought to improve methods to access CylK, and to determine the substrate scope and selectivity of this enzymatic transformation. As enzymes with tolerance for non-native substrates present useful starting points for future protein engineering efforts,[132] we believe this effort will set the stage to use CylK as a biocatalyst and to develop a fundamental mechanistic understanding of this novel enzyme class.

2.2 Results and Discussion

2.2.1 Efforts to optimize enzyme overexpression and purification methods

Following the initial discovery of CylK, further efforts to study its native reactivity and substrate scope were hampered by unreliable and inefficient heterologous expression in *E. coli*. These challenges might be explained by the differences in evolutionary history, cellular environment, and/or codon usage between cyanobacteria (such as the native CylK-encoding organism, *C. licheniforme*) and Proteobacteria (in particular, common laboratory strains of *E. coli* used for recombinant protein expression).[133] After screening a number of expression conditions including N- and C-terminal affinity tags, *E. coli* host strains, expression temperature/time, and $CaCl_2$ supplementation, the best preparation of CylK to date had been from a plasmid-encoded C-terminal *Strep*-tag II construct (CylK-Strep) (Figure 2.2A) expressed in *E. coli* BL21 Gold CodonPlus® (DE3) RIL (Agilent).[69,86,134] This strain of *E. coli* is protease deficient and contains an additional plasmid encoding tRNAs for rare codons AGG/AGA (Arg), AUA (Ile), CUA (Leu). Of note, all four of these rare codons are found in the native *cylK* gene, as well as in our expression constructs. The instances of soluble expression of CylK-Strep in this strain might be explained by a number of factors including codon availability, growth/expression rate, and affinity tag properties, however, yields of soluble CylK remained low and variable.

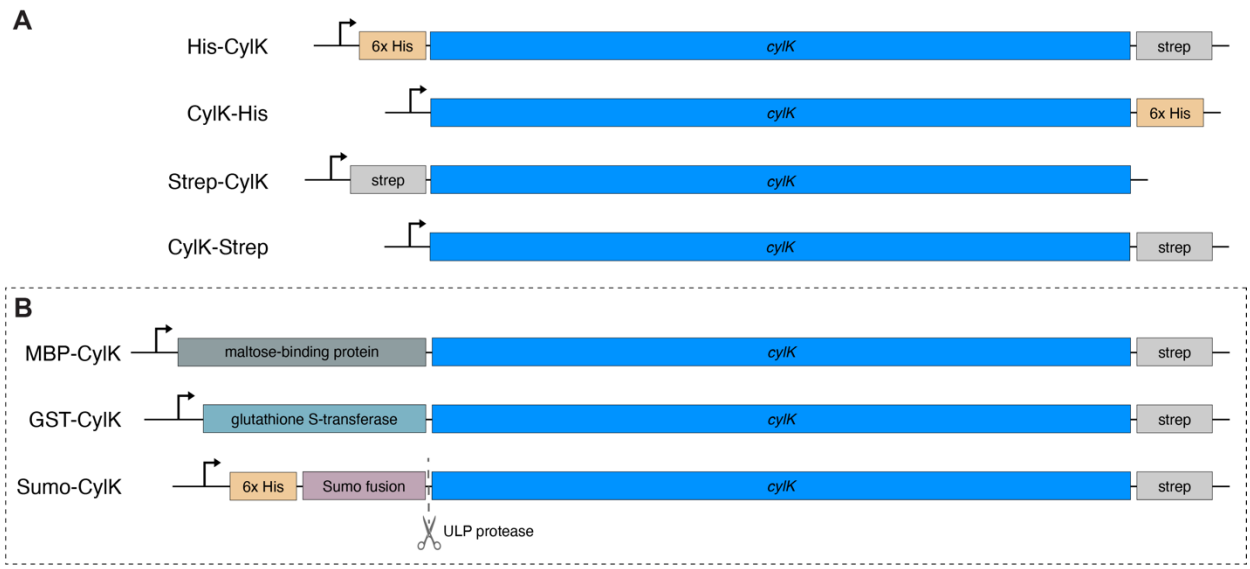


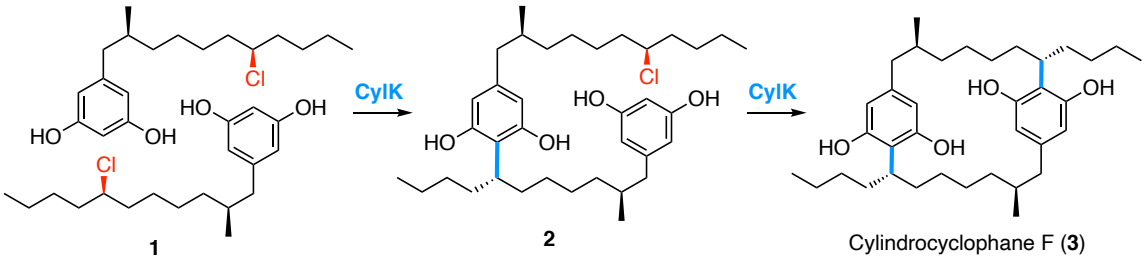
Figure 2.2. CylK constructs designed for protein preparation development

(*A*) Affinity tagged CylK constructs prepared by Nakamura et al. Of these, CylK-Strep had detectable solubility and activity.[69] (*B*) In this work, we constructed additional N-terminal fusions to CylK-Strep to improve protein preparations. Constructs and components are not drawn to scale.

In order to assess and potentially improve the reproducibility of this expression system, we reasoned that the 12–18 h overnight induction time (following IPTG addition) might be a significant source of variability affecting protein expression and solubility. Supporting this hypothesis, it has been shown that after induction, heterologous protein yields in *E. coli* can peak at a specific time point, and can then decrease beyond that window.[135] Accordingly, we designed a new expression protocol that included a shorter and consistently timed induction window of 4 h, before harvesting cells for subsequent lysis. We also began each expression attempt by freshly transforming BL21 Gold CodonPlus® (DE3) RIL with the plasmid encoding CylK-Strep, rather than using a frozen glycerol stock of this strain. We found that these changes resulted in a significant increase in CylK activity, solubility, and batch-to-batch consistency. In one key experiment, we found that reducing the induction time from 12–18 h down to 4 h increased activity of CylK lysate more than 4-fold (Table 2.1), suggesting that the enzyme's solubility, expression, and/or protein folding are also improved. We also observed a modest increase in CylK activity from cultures that had been supplemented with 5.0 mM $CaCl_2$ just preceding induction. Of note, moderately increasing the induction

time from 4 h to 6 or 7 h did not significantly alter CylK solubility or activity (data not shown), suggesting that the soluble protein yield may not drop off until closer to 12–18 h of induction. With these new preparation conditions, we were able to consistently obtain protein yields of approximately 120 µg/L of culture. Although this marked a significant improvement in CylK preparation reproducibility, yields remained frustratingly low, limiting access to sufficient quantities of enzyme for subsequent biochemical and structural studies.

Table 2.1. Optimization of growth and expression conditions enhancing CylK activity in lysate



1 **2** Cylindrocyclophane F (**3**)

Entry	Induction Time (h)	$CaCl_2$ (mM)	Conversion to **2** (%)	Conversion to **3** (%)
1	16	–	6	1
2	16	5.0	9	2
3	4	–	33	2
4	4	5.0	44	3

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are the mean of three biological replicates. All clarified lysate reactions were run for 1 h at room temperature.

To further improve soluble protein yield, we next cloned three N-terminal fusion constructs of CylK-Strep (Figure 2.2B). Three fusion partners – maltose binding protein (MBP),[136] glutathione S-transferase (GST),[137] and small ubiquitin-related modifier (Sumo)[138] – were chosen because they are readily available and have been demonstrated to increase the solubility of challenging proteins of interest.[139] With the three N-terminal fusion constructs in hand, small scale expression (50 mL cultures) and lysate activity assays were conducted (Figure 2.3A). Unfortunately, none of the CylK constructs appeared to have significant soluble protein expression, as visualized by SDS-PAGE gel stained with Coomassie. However, high levels of insoluble protein were apparent for CylK-Strep (74.8 kDa) and Sumo-CylK-Strep (87.8 kDa), while the MBP-CylK-Strep fusion (117.8 kDa) expressed insolubly at a molecular weight similar to CylK-Strep. This might be explained by proteolysis or the initiation of protein translation at the native ATG start site of CylK resulting in a truncated construct. The GST fusion protein (107.0 kDa) did not visibly express by SDS-PAGE stained with Coomassie. Next, all soluble fractions were examined for enzymatic activity on the native CylK substrate **1** (Figure 2.3B). All constructs appeared to have comparable activity to CylK-Strep except for the GST fusion protein which had no measurable activity.

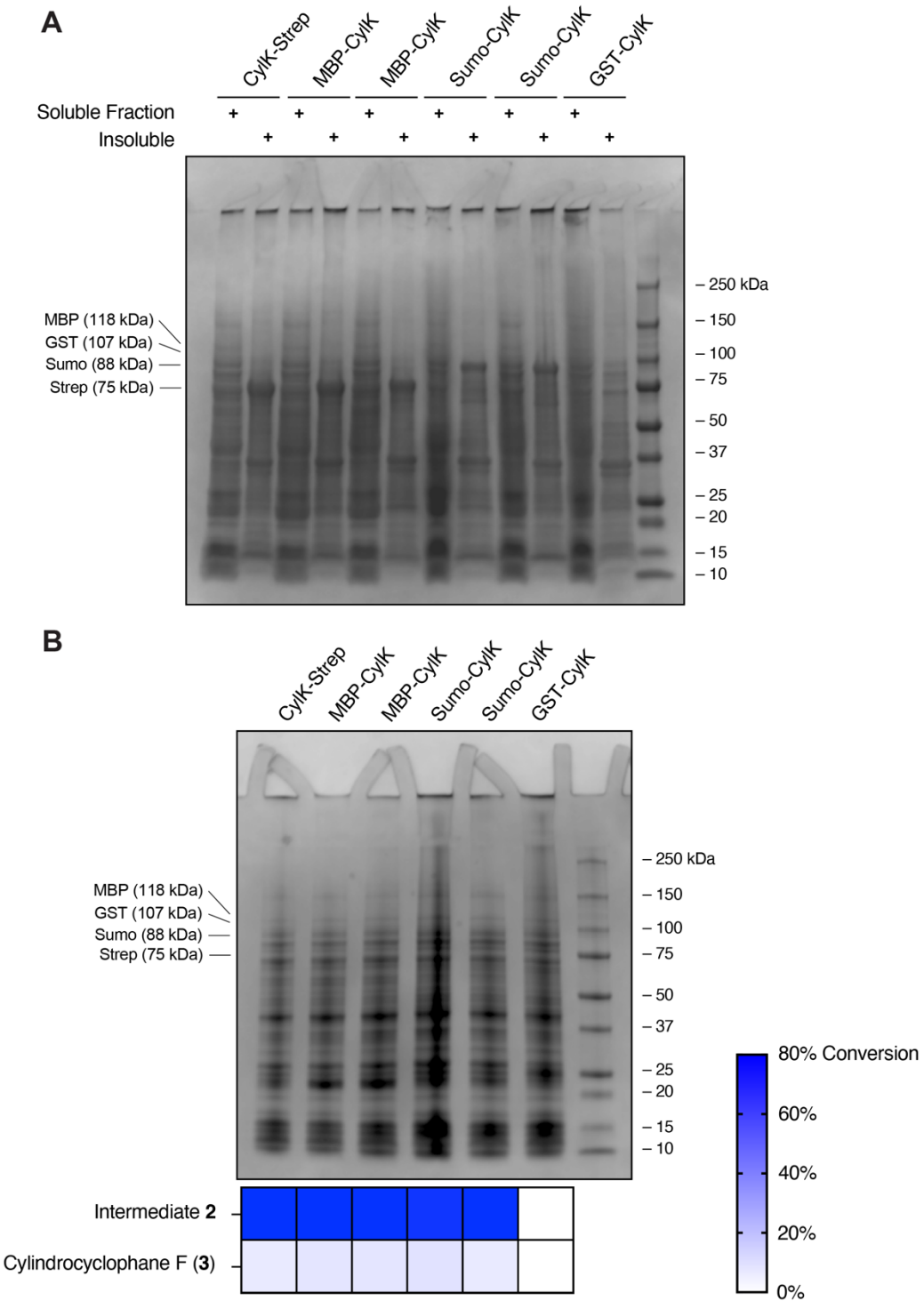


Figure 2.3. Expression and activity of CylK N-terminal fusions

(*A*) SDS-PAGE of soluble and insoluble fractions from 50 mL scale expression cultures. (*B*) SDS-PAGE of soluble lysate (from *A*) and heat map of corresponding enzymatic activity. All Blue Precision Plus Protein Prestained Standard (Bio-Rad) was used to visualize molecular weight standards. Clarified lysate reactions were run for 50 min at room temperature. Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are from single replicates.

Considering the high levels of insoluble protein expression, we next attempted to denature and refold CylK from whole cell lysate in order to increase enzyme yields. Three polyhistidine-tagged CylK constructs were chosen to use in screens of refolding conditions because the Strep-Tactin resin (used to purify Strep-tagged CylK) would be denatured under these conditions. These CylK constructs were expressed on a small scale (50 mL) in *E. coli* BL21 (DE3) in order to maximize insoluble protein yields, and then were resolubilized in lysis buffer containing 6 M urea. After batch nickel-nitrilotriacetic acid (Ni-NTA) affinity purification in the presence of 6 M urea, the denatured proteins were refolded by removing urea via three dialysis steps. The resulting solutions were tested for enzymatic activity on native substrate **1** (Table 2.2). Gratifyingly, all were active after urea removal, suggesting that all three CylK constructs had been successfully refolded. The His-Sumo-CylK-Strep construct was selected for further optimization because it demonstrated the highest enzymatic activity after 17 h, and the Sumo fusion can be selectively removed by incubating with Ulp1 protease from *Saccharomyces cerevisiae*,[140] allowing for an additional subtractive Ni-NTA step to further purify CylK.

Table 2.2. Activity of denatured and refolded polyhistidine-tagged CylK constructs

Entry	Construct	Reaction Time (h)	Conversion to **2** (%)	Conversion to **3** (%)
1	His-CylK	1	86	9
2		17	84	15
3	CylK-His	1	87	9
4		17	85	15
5	His-Sumo-CylK-Strep	1	92	5
6		17	71	29

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are from single replicates. All enzymatic reactions were run at room temperature.

The workflow described above was then translated to large scale growth cultures (2 L) of His-Sumo-CylK-Strep in *E. coli* BL21 (DE3) and carried out using analogous refolding conditions. Additional protein purification steps were added; including ultracentrifugation to remove soluble aggregates, Ulp1 proteolysis

to access CylK-Strep and enable subtractive Ni-NTA, and size exclusion chromatography (Superdex 200) as a final polishing step and to confirm homogeneity. This procedure routinely afforded 2.0 mg of CylK-Strep per L of culture (see Materials and Methods 2.4.6). This represents a significant increase in protein yield, approximately 20-times more than soluble expression of CylK-Strep in BL21 Gold CodonPlus® (DE3) RIL. Importantly, this refolding protocol gave enzyme with comparable activity and an identical oligomeric state to solubly expressed CylK-Strep. Subsequently, we have further optimized this refolding protocol by initially lysing in buffer without urea in order to isolate the protein in the insoluble fraction and crudely purify it from the contaminating *E. coli* components of the soluble fraction. Additionally, minor buffer changes were made to standardize buffer component concentrations across refolding steps. This modified procedure routinely afforded 4.5 mg of CylK-Strep per L of culture (see Materials and Methods 2.4.7).

Of note, Wang et al. reported a preparation of CylK that involves refolding and includes several modifications to our method.[141] Their enzyme construct appears to be untagged, and can be sufficiently purified as inclusion bodies from *E. coli* cell debris by washing cell pellets and lysate with a buffer containing 0.1% Triton X-100. Subsequently, CylK is solubilized in a 6.7 M urea buffer and the pH is rapidly increased to 10.5 (our pH is held constant at 8), before refolding by removing urea via two dialysis steps. The pH returns to approximately pH 8 during refolding dialysis steps. The key differences between the two procedures seem to be the pH change and the presence of 10 mM glycine and 5 mM β-cyclodextrin over the course of refolding. This CylK preparation is reported to similarly afford a monomeric species, with comparable activity on the native substrate.

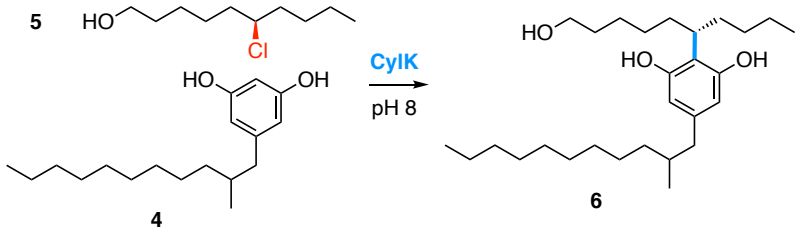
2.2.2 Biocatalytic application and optimization of non-native CylK reaction

With a reliable enzyme preparation in hand, we next explored the substrate scope of CylK in order to assess its potential use as a biocatalyst. In the initial CylK discovery paper, Nakamura et al. established that discrete resorcinol nucleophile **4** and alkyl chloride electrophile **5** are accepted with modest conversions (~50%),[69] suggesting that CylK might catalyze a Friedel–Crafts alkylation with an array of discrete substrate pairs. While resorcinol **4** and alkyl chloride **5** resemble the native bifunctional substrate (**1**), they lack either

an alkyl chloride or resorcinol substituent, respectively, rendering them unable to homodimerize. We initially began this investigation of CylK's substrate scope by optimizing the reaction conditions for these model non-native substrates to afford adduct **6** with high conversions and low enzyme loading.[128,vi] At room temperature, a 1:2 ratio of **4** + **5** gave 51% conversion to alkylated resorcinol **6** after 24 h (Table 2.3, entry 1). We reasoned that the rate of reaction and overall conversions could be enhanced if we increased the reaction temperature to 37 °C; hypothesizing that while the enzyme might denature and lose activity over the course of 24 h at an elevated temperature, the additional heat may accelerate the reaction rate more than the rate of enzyme inactivation. We also used a 1:1 ratio of **4** + **5** in order to determine if it is necessary to have the alkyl chloride electrophile (**5**) in excess. With these modifications, the reaction now afforded 97% conversion to alkylated resorcinol **6** (Table 2.3, entry 2). We then sequentially decreased enzyme loading from 1 μM to 0.25 μM and found no detectable loss in overall conversion (Table 2.3, entries 2–4). Of note, substrate loadings could also be increased from 200 μM to 500 μM with some loss in conversion (Table 2.3, entries 5–7), although we suspect that this is due to enzyme inhibition by dimethyl sulfoxide (DMSO) introduced from substrate stock solutions (see Table 2.4 discussed in more detail below). Finally, the reaction volume could be scaled up ten-times from 0.1 mL to 1.0 mL with near quantitative conversion at 0.5 μM enzyme loading (Table 2.3, entry 9).

[vi] This *in vitro* biochemical work was primarily conducted by Dr. Erica Schultz and adapted from Schultz, Braffman, et al. 2019.

Table 2.3. Optimization of non-native CylK reaction with model substrate pair

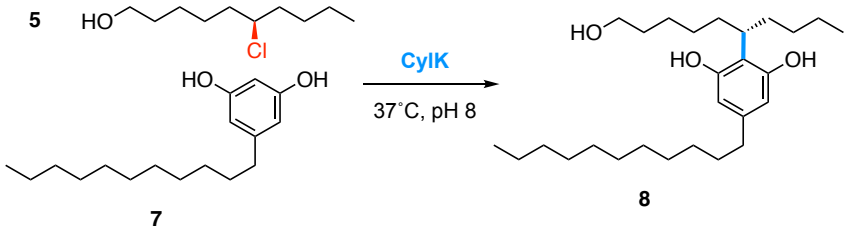


Entry	[CylK] (μM)	[**4**] (μM)	[**5**] (μM)	Temp. (°C)	Time (h)	Volume (μL)	Conversion to **6** (%)	TON
1	1.0	100	200	23	24	100	51	51
2	1.0	200	200	37	24	100	97	194
3	0.5	200	200	37	24	100	97	388
4	0.25	200	200	37	24	100	98	784
5	0.25	300	300	37	22	100	90	1080
6	0.25	400	400	37	22	100	72	1152
7	0.25	500	500	37	22	100	63	1260
8	0.25	200	200	37	17	1000	20	160
9	0.5	200	200	37	17	1000	98	392

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are from single replicates. Turnover number (TON) defined by [product]/[catalyst].

We then tested a desmethyl analog of **4**, undecylresorcinol **7**, as a nucleophilic substrate to minimize the number of stereocenters in the resulting product (**8**), facilitating the determination of enantiomeric excess (ee) of the CylK catalyzed reaction. Of note, we observed a significant increase in conversion when running the reaction in glass vials (amber glass, Agilent 5182-0716) over plastic vials (polypropylene, VWR 87003-294) (Table 2.4, entries 1–2). This might be due to enhanced affinity of the substrate(s) for the non-polar plastic surface over the hydrophilic reaction buffer, rendering the effective substrate concentration lower. DMSO concentrations in the range of 2–3% (v/v) gave near quantitative conversion (Table 2.4, entries 2–3) while 4% (v/v) DMSO appeared to be inhibitory (Table 2.4, entry 4). We suspect that low concentrations of DMSO facilitate the reaction by solubilizing the non-polar substrates in aqueous buffer, while high concentrations of DMSO become inhibitory as they may alter substrate-enzyme binding affinity and/or promote enzyme inactivation via unfolding.

Table 2.4. Further optimization of 1 mL scale reactions



Entry	[CylK] (μM)	[**7**] (μM)	[**5**] (μM)	Vial	DMSO (%)	Volume (mL)	Conversion to **8** (%)	TON
1	1.0	200	204	Plastic	2.0	1.0	64	128
2	1.0	200	204	Glass	2.0	1.0	98	196
3	1.0	200	204	Glass	3.0	1.0	99	198
4	1.0	200	204	Glass	4.0	1.0	84	168

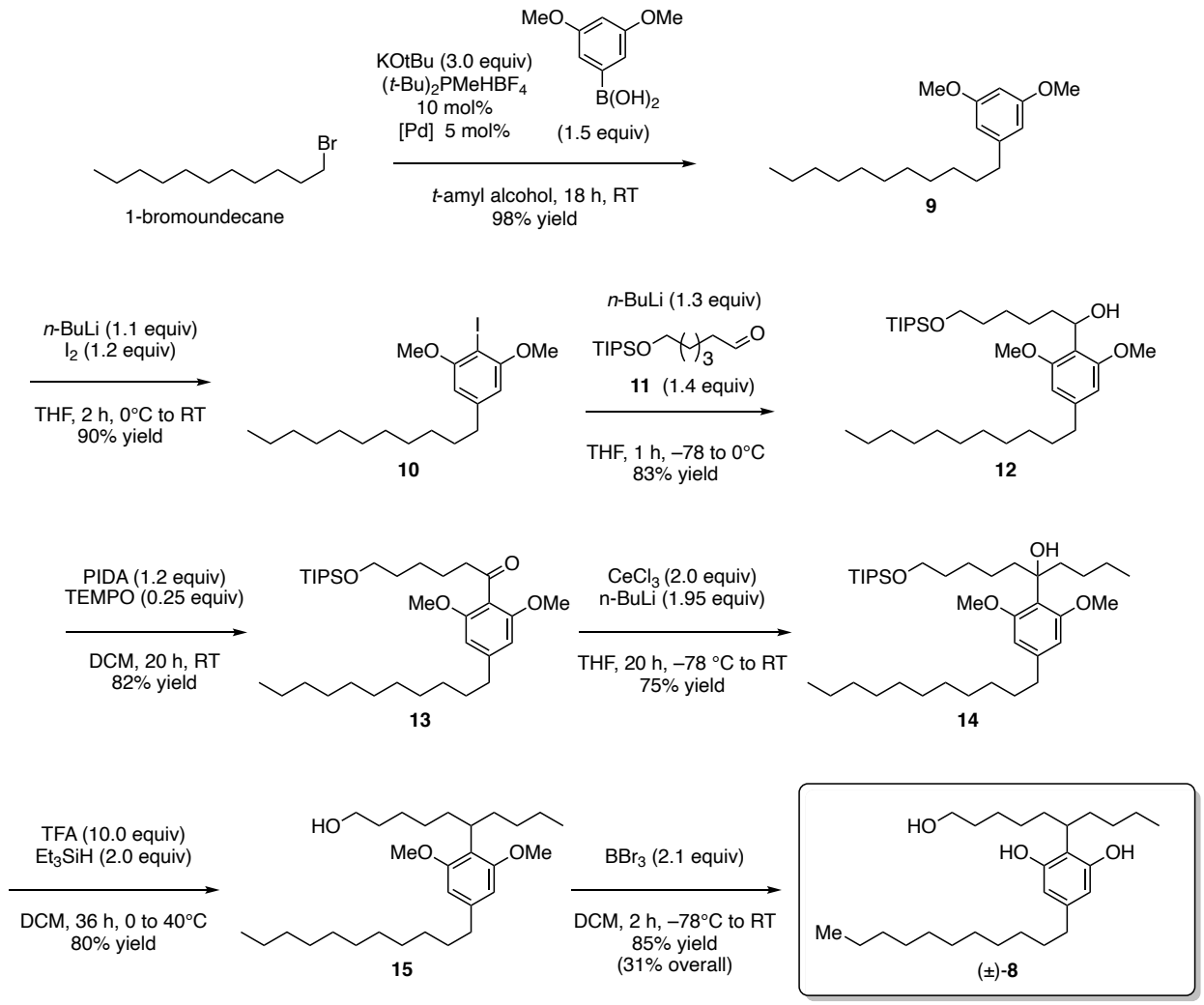
Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are from single replicates. All enzymatic reactions were run for at 37 °C for 24 h, with 150 rpm shaking. Turnover number (TON) defined by [product]/[catalyst].

2.2.3 Characterization and enantiomeric excess of a non-native CylK catalyzed reaction

With optimized biocatalytic reaction conditions identified, we set out to determine if a non-native CylK catalyzed reaction maintains the regioselectivity and stereospecificity of the native CylK reaction. Although Nakamura et al. firmly established that CylK natively alkylates at the 2-position of the resorcinol ring in biological substrate **1**, and the substitution reaction occurs with inversion of configuration at the alkyl chloride electrophile stereocenter, it remained unknown to what extent this held true for discrete and non-biological substrates.[69] To test this, we sought to chemically synthesize a racemic standard of product **8** in order to compare it to the product generated in the enzyme catalyzed reaction using NMR spectroscopy and chiral LC-MS. We developed a chemosynthetic route to (±)-**8** (Scheme 2.3)[vii] with inspiration from an elegant asymmetric total synthesis of the cylindrocyclophanes from the K. C. Nicolaou group.[110] Unlike the biological reaction, the key alkyl–aryl C–C bond was formed by lithiating iodoresorcinol **10** followed by addition into aldehyde **11**. Of note, it is important to store the aldehyde as a solution in DCM at –20 °C and

[vii] This chemosynthetic work was initially developed and conducted by Nathaniel R. Braffman, with invaluable input from Dr. Michael U. Luescher, Dr. Monica E. McCallum, and characterization assistance from Dr. Paul D. Boudreau. Dr. Michael U. Luescher optimized key synthetic steps for publication.

use within a few days for best results. The resulting alcohol (**12**) was then oxidized to ketone **13**, treated with *n*-butyllithium, dehydrated and reduced, and finally deprotected to access racemic product **8**. Gratifyingly, the NMR spectra of enzymatically accessed product **8** were consistent with those of the synthetic chemical standard (Table 2.5 & Table 2.6). This demonstrates that CylK couples discrete resorcinol nucleophiles and alkyl halide electrophiles with the same regioselectivity and overall connectivity as the native biological reaction. Next, we determined that the enzymatic reaction yields product in greater than 99% ee (Figure 2.4F), suggesting that the reaction catalyzed by CylK is stereospecific. From this result, we infer that this reaction proceeds with inversion of configuration at the alkyl chloride stereocenter to afford (*S*)-**8**, paralleling CylK's native biological activity. Finally, when challenged with two equivalents racemic electrophile **5**, the ee of the product was 88% (Figure 2.4G) with 79% conversion, relative to one equivalent of resorcinol nucleophile **7**. The results of this kinetic resolution suggests that CylK is stereoselective for one enantiomer of electrophile **5**, and it slowly reacts with the other enantiomer to give a trace amount of (*R*)-**8**.



Scheme 2.3. Synthesis of racemic CylK product standard

Seven step linear synthesis of product (±)-**8** to afford a racemic chemical standard for chiral LC-MS and NMR spectroscopic characterization. Note, aldehyde **11** should be stored at –20 °C as a solution in anhydrous dichloromethane and used within a few days for best results.

Table 2.5. ^{1}H NMR comparison of product 8 prepared enzymatically or via chemical synthesis

Enzymatic (500 MHz)	**Synthetic** (500 MHz)	Δ δH (ppm)	Comments
6.15 (s, 2 H)	6.15 (s, 2 H)	0.00	-
4.60 (s, 2 H, phenolic OH)	4.53 (br s, phenolic OH)	- 0.07	OH groups
3.58 (q, *J* = 6.2 Hz, 2 H)	3.59 (t, *J* = 6.6 Hz, 2 H)	- 0.01	-
3.14–2.99 (m, 1 H)	3.14–2.95 (m, 1 H)	- 0.02	-
2.47–2.38 (m, 2 H)	2.49–2.37 (m, 2 H)	+ 0.01	-
1.95–1.76 (m, 2 H)	1.93–1.78 (m, 2 H)	0.00	-
1.68–1.58 (m, 2 H)	1.68–1.58 (m, 6 H)	0.00	-
1.58—1.42 (m, 4 H)	1.58–1.44 (m, 4 H)	+ 0.01	Difficult to integrate in the enzymatic sample due to the presence of water
1.43–1.10 (m, 24 H + OH)	1.40–1.15 (m, 24 H + OH)	+ 0.01	-
0.88 (t, *J* = 6.8 Hz, 3 H)	0.88 (t, *J* = 6.9 Hz, 3 H)	0.00	-
0.84 (t, *J* = 7.1 Hz, 3 H)	0.84 (t, *J* = 7.1 Hz, 3 H)	0.00	-

Table 2.6. ^{13}C NMR comparison of product 8 prepared enzymatically or via chemical synthesis

Enzymatic (125 MHz)	**Synthetic** (125 MHz)	Δ δ13C (ppm)	Comments
155.17 (weak, broad signal)	155.32 (broad signal)	+ 0.15	Broad signal in both samples
142.24	142.12	- 0.12	-
114.92	115.03	+ 0.11	-
108.91 (weak, broad signal)	108.61 (broad signal)	- 0.30	Broad signal in both samples
63.15	63.07	- 0.08	-
35.59	35.53	- 0.06	-
35.50	35.44	- 0.06	-
33.88	33.90	+ 0.02	-
33.77	33.65	- 0.12	-
32.70	32.52	- 0.18	-
32.07	32.07	0.00	-
31.07	31.08	+ 0.01	-
30.75	30.75	0.00	-
29.83	29.84	+ 0.01	-
29.80	29.80	0.00	-
29.75	29.76	+ 0.01	-
29.68	29.69	+ 0.01	-
29.55	29.59	+ 0.04	-
29.50	29.50	0.00	-
27.95	27.78	- 0.17	-
25.68	25.53	- 0.15	-
23.02	23.03	+ 0.01	-
22.84	22.83	- 0.01	-
14.27	14.26	- 0.01	-

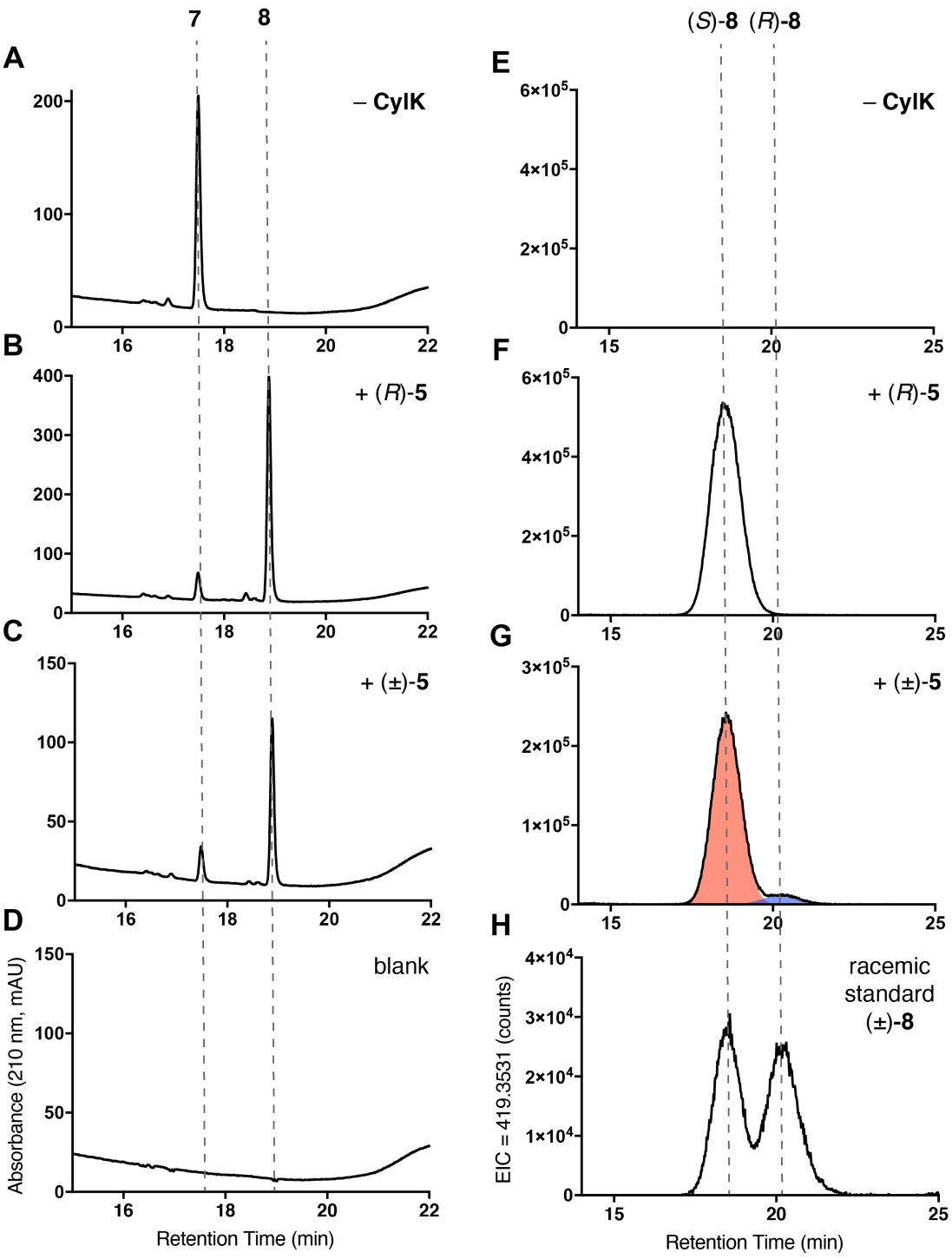


Figure 2.4. HPLC and chiral LC-MS analysis of CylK reactions

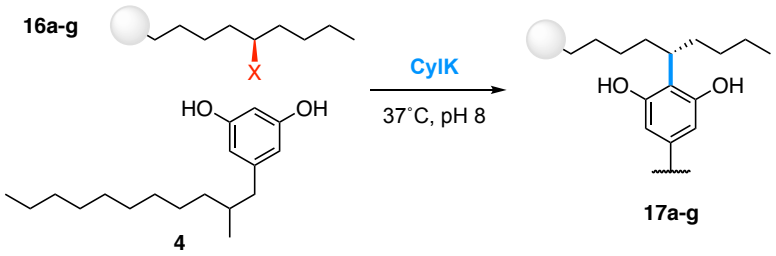
All enzymatic reactions were run for at 37 °C for 24 h. (Left) HPLC chromatograms of reactions (*A*) without CylK, (B) with enantioenriched racemic electrophile (*R*)-**5**, (*C*) racemic electrophile (±)-**5**, or (*D*) blank injection. (Right) Extracted ion chromatograms (EICs) of chiral LC-MS analysis on reactions (*E*) without CylK, (*F*) with enantioenriched racemic electrophile (*R*)-**5**, (*G*) racemic electrophile (±)-**5**, or (*H*) synthetic racemic standard (±)-**8**. Area of overlapping peaks was determined by modeling Gaussian distributions (shown in red and blue).

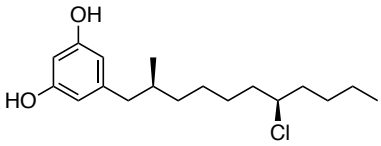
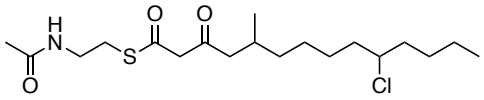
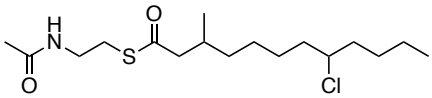
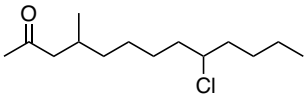
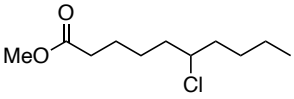
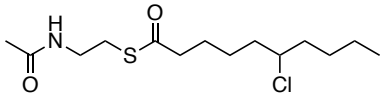
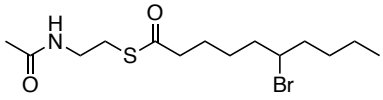
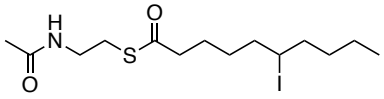
2.2.4 Alkyl halide electrophile substrate scope of CylK mediated Friedel–Crafts alkylation

We further examined CylK's tolerance for various alkyl halide electrophile substrates to inform our understanding of the enzyme's binding/recognition and electronic requirements, and to further ascertain the potential utility of the native enzyme as a biocatalyst (Table 2.7).[128,viii] In addition to hydroxyl group-containing substrate **5**, we found that CylK accepts a variety of functional groups in place of the native resorcinol ring (**1**), including amides and thioesters (**16a**,**b**,**e-g**), ketones (**16c**), and esters (**16d**) located four or more methylenes away from the alkyl halide substituent. Interestingly, CylK also accepts alkyl bromides (**16f**) and alkyl iodides (**16g**) as electrophilic substrates, in addition to alkyl chlorides. Chloride **16e** and bromide **16f** yielded product in 88% and 76% conversion, respectively, while alkyl iodide **16g** afforded product in 29% conversion. Notably, this observed trend is the inverse of their inherent reactivity, as the C–I bond is weaker and a better electrophile than C–Br and C–Cl bonds.[126,127] However, the size of the alkyl halide substituent increases from chloride, bromide, to iodide,[142] which might suggest that CylK's active site has evolved to preferentially bind and activate alkyl chloride electrophiles based on their size. This could imply that, while the C–I bond is most labile, the enzyme does not readily activate it for resorcinol substitution. Although we do not yet know if CylK's partner chlorinase, CylC, has promiscuous brominating activity, a number of natively chlorinating halogenases can also brominate,[143] and brominated cylindrocyclophane and carbamidocyclophane natural products are produced from *Nostoc* sp. UIC 10022A and CAVN2, respectively, in bromide-enriched growth media.[61,144] *Nostoc* sp. CAVN2 contains enzymes homologous to CylK and CylC in a putative carbamidocyclophane biosynthetic gene cluster. This information, in combination with CylK's ability to interface with brominated electrophiles, suggests that brominated metabolites might be produced as cryptic intermediates in cylindrocyclophane biosynthesis under certain nutrient contexts.

[viii] This *in vitro* biochemical work was primarily conducted by Dr. Erica Schultz, with contributions from Harry H. Hager and Dr. Michael U. Luescher, and adapted from Schultz, Braffman, et al. 2019.

Table 2.7. Alkyl halide electrophile substrate scope



	Substrate	Conversion (%)	TON
native substrate **1**		–	–
16a		100	200
16b		100	200
16c		97	194
16d		22	44
16e		88	176
16f		76	152
16g		29	58

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are the mean of two biological replicates. All enzymatic reactions were run for at 37 °C for 24 h, with 200 µM of racemic alkyl halide **16** and 100 µM of resorcinol **4**. Turnover number (TON) defined by [product]/[catalyst].

CylK demonstrated no activity towards an alkyl fluoride electrophile (**16h**) (Figure 2.5). This is in agreement with the fact that the C–F bond in fluoride **16h** likely has a higher bond dissociation energy (~108 kcal/mol) than new the C–C bond in product **17** (~93 kcal/mol, based on toluene),[126,127] suggesting that the defluorination reaction is thermodynamically unfavorable, in contrast to the dissociation energies of C–Cl, C–Br, and C–I bonds (83.5–56 kcal/mol) which are lower than the alkyl–aryl C–C bond. Further, CylK does not accept alkyl chlorides containing a free carboxylic acid (**16n**), secondary alkyl chlorides without extended alkyl chains and polar head groups (**16q-v**), primary chlorides (**16j,k**), tertiary chlorides (**16i**), alkenes (**16l**), and hydroxyl (**16m**), tosyl (**16o**), and mesyl (**16p**) groups as electrophiles (Figure 2.5). This might be due to factors including sterics, active site substrate affinity, and mechanistic considerations.

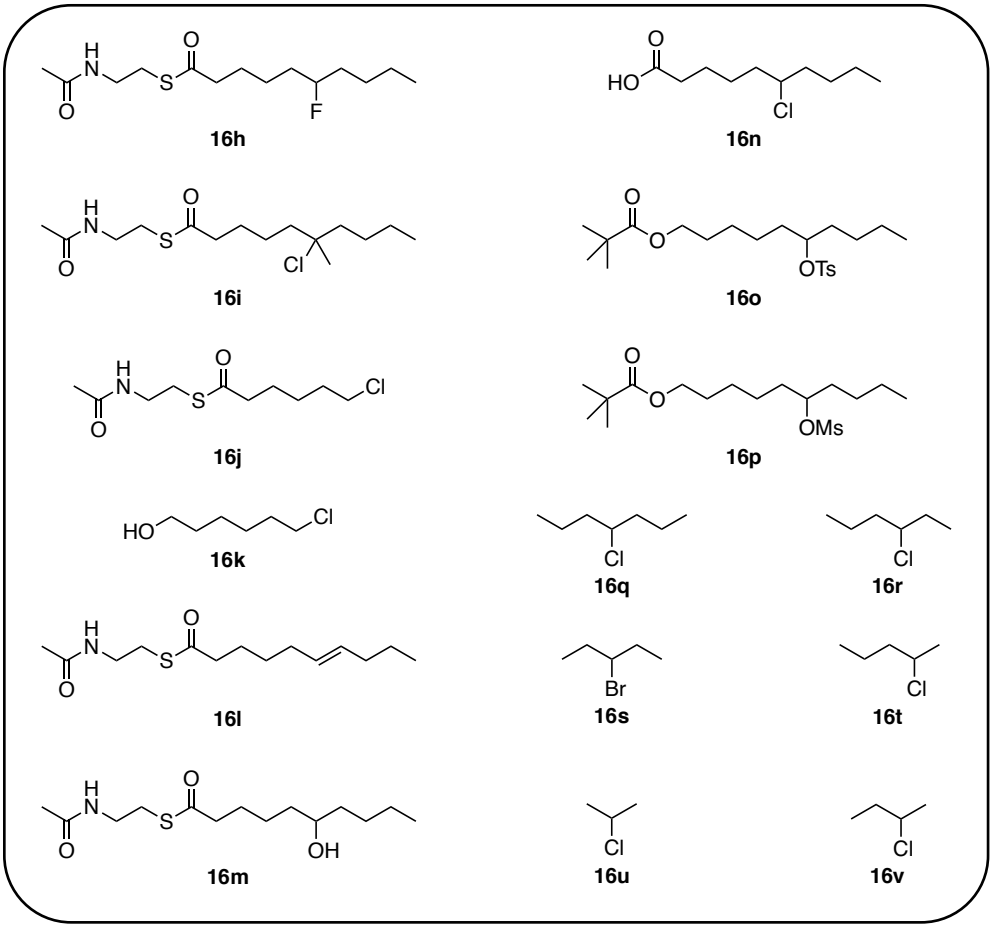


Figure 2.5. Electrophilic substrates not accepted by CylK

2.2.5 Resorcinol nucleophile substrate scope of CylK mediated Friedel–Crafts alkylation

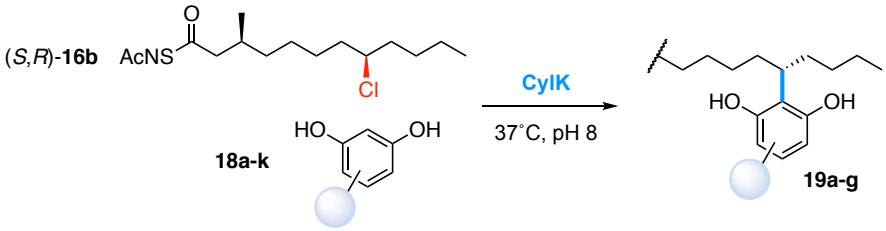
We next examined CylK's ability to interface with alternate arene nucleophiles to further elaborate the substrate scope of the native enzyme.[ix] This work demonstrated that CylK can accept resorcinol substrates with a variety of substituents at the 4- and 5-positions of the arene ring (Table 2.8). Substitution at the 5-position was more readily tolerated than the 4-position, which parallels the connectivity of the alkyl tail and resorcinol ring of native substrate **1**. Electron rich substrates (**18a-b,g-h**) generally gave higher conversions than electron poor substrates (**18c-d,h-i**), supporting a role for the arene ring as the nucleophilic partner, however the inherent the nucleophilicity of these substrates does not fully capture all factors influencing conversion. Binding affinities, steric factors, and potential product inhibition likely influence reaction rates and end point conversions. Further, resorcinol (**18f**), a free hydroxyl group (**18e**), fused aryl ring (**18k**), and aryl bromide groups (**18d,j**) were tolerated. The brominated nucleophilic substrates are notable because they can serve as synthetic handles for downstream coupling reactions.[145]

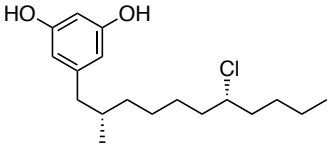
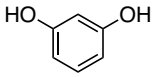
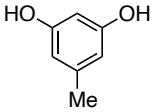
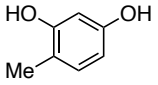
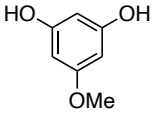
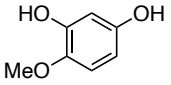
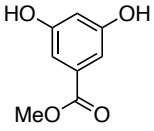
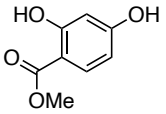
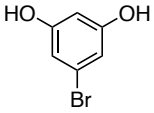
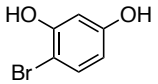
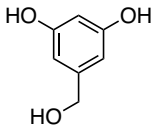
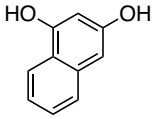
We found that the nucleophile substrate scope of CylK was strictly limited to resorcinol moieties with two free phenolic groups at the 1- and 3-positions of the arene ring (Figure 2.6). Substrates with a single phenol were not tolerated (**18m,q**), nor were substrates containing a mono-protected resorcinol (**18n**) with only one free phenol. Similarly, 3-aminophenol (**18p**) was not accepted despite the 1,3 relationship of the amino and phenolic groups. Interestingly, catechol **18l**, with a 1,4 di-phenolic relationship, was not a competent nucleophile, nor was di-protected resorcinol **18o**. These results strongly suggest that the nucleophile substrate scope is limited to resorcinols with two free phenolic groups, only substituted at the 4- and/or 5-positions. Unsurprisingly, di-amine (**18r**) and indole (**18s**) were not accepted by CylK. Of note, nitrile-substituted resorcinol (**18t**) failed to show even trace conversion to product. This might be due to the electron deficiency of this nucleophile, however, additional nitrogen-containing resorcinols (**18o,v**) were also not competent nucleophiles. This might suggest that nitriles, pyridines, and anilines are incompetent

[ix] This *in vitro* biochemical work was conducted by Dr. Erica Schultz, Nathaniel R. Braffman, and Harry H. Hager, and adapted from Schultz, Braffman, et al. 2019.

nucleophilic substrates due to a combination of intersecting factors including nucleophilicity, charge, and enzyme binding considerations.

Table 2.8. Resorcinol nucleophile substrate scope



	Substrate	Conversion (%)	TON		Substrate	Conversion (%)	TON
native substrate **1**				**18f**		22	44
18a		56	112	**18g**		17	34
18b		51	102	**18h**		9	18
18c		34	68	**18i**		< 1	< 2
18d		27	54	**18j**		7	14
18e		37	74	**18k**		25	50

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are the mean of two biological replicates. All enzymatic reactions were run for at 37 °C for 24 h, with 200 µM of alkyl chloride (*S*,*R*)-**16b** and 200 µM of resorcinol **18**. Turnover number (TON) defined by [product]/[catalyst].

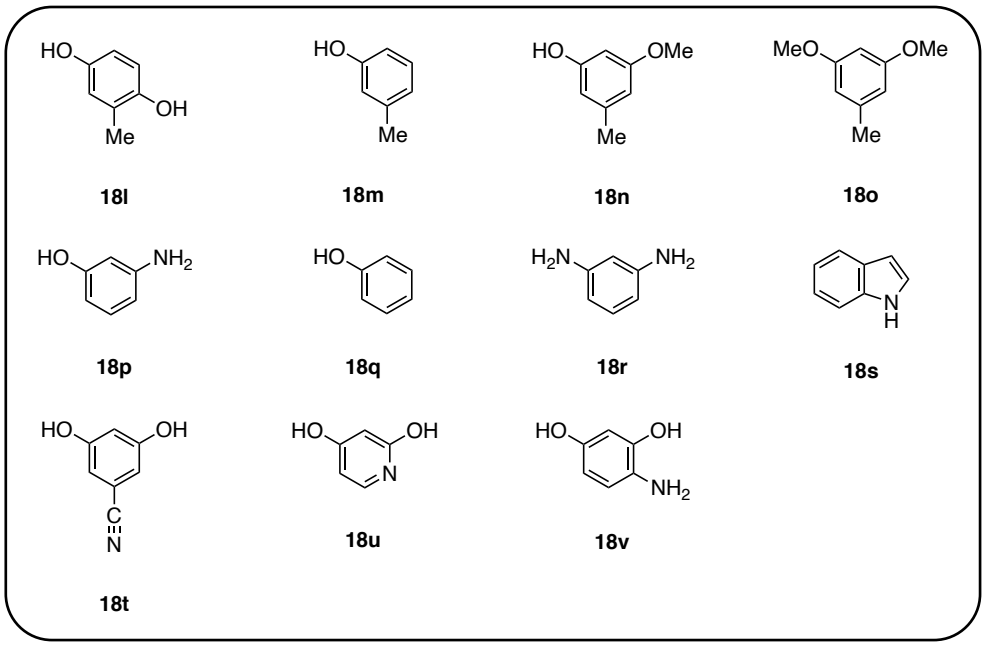


Figure 2.6. Nucleophilic substrates not accepted by CylK

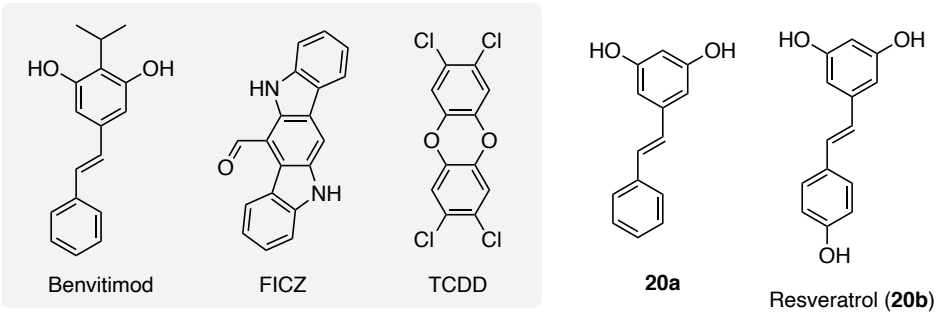
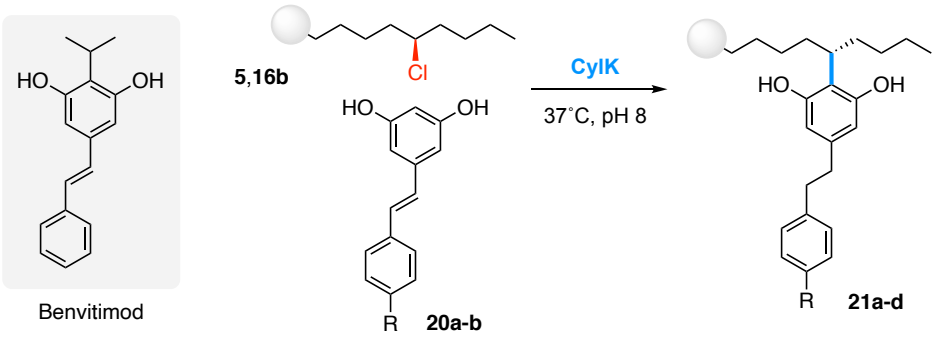


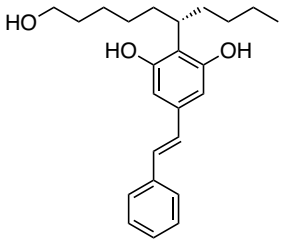
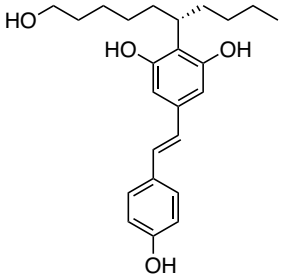
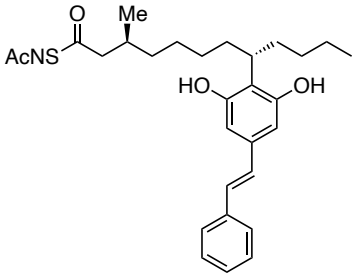
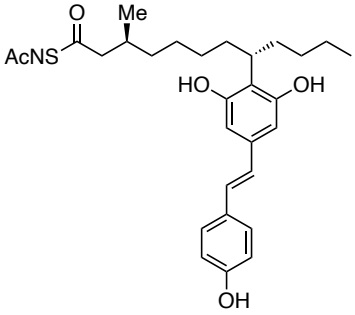
Figure 2.7. AhR agonists and synthetic precursor stilbene nucleophiles

Three known aryl hydrocarbon receptor (AhR) ligand agonists are highlighted in the grey box. FICZ (6-formylindolo(3,2-b)carbazole) is believed to be an endogenous human AhR ligand produced from tryptophan.[146] TCDD (2,3,7,8-Tetrachlorodibenzo-*p*-dioxin) is an anthropogenic product produced as an industrial side product.[147] 3,5-dihydroxystilbene (**20a**) and resveratrol (**20b**) are CylK nucleophiles that can afford access to alkylated benvitimod-like products.

Upon examining the distinctive dialkylresorcinol scaffold that can now be readily accessed with CylK, we recognized that this feature was found in at least one notable bioactive molecule, referred to as benvitimod or tapinarof interchangeably (Figure 2.7). This alkylated stilbene natural product was first isolated from *Xenorhabdus luminescens*,[148] and it is in clinical trials as a topical treatment for mild to moderate plaque psoriasis.[149] There is mounting evidence that suggests psoriasis is related to a dysregulated immune system, in particular, a dependence on tumor necrosis factor alpha (TNFα) and over activation of helper T cells 1 (T_H1) and 17 (T_H17).[150-152] It has been demonstrated that benvitimod may resolve the symptoms of psoriasis by reducing interleukin-17A (IL-17A) transcript levels in T_H17 cells by activating the aryl hydrocarbon receptor (AhR).[149] Of note, benvitimod resembles other AhR agonists, many of which contain similar planar and highly conjugated structures (Figure 2.7).[146] Current synthetic methods to access this alkylated stilbene require toxic reagents or rely on a harsh deprotection conditions to remove the methoxy protecting groups on the resorcinol moiety.[153,154] Notably, one process route to benvitimod involves a Friedel–Crafts alkylation reaction between a dimethoxy-protected resorcinol compound and 2-bromopropane, catalyzed by $AlCl_3$ and at reflux in CS_2 for 7 days.[155] Use of this transformation inherently limits the alkyl substituent to the isopropyl group, because this alkyl group is symmetric and lacks any methylene carbons, preventing carbocationic rearrangements. Larger and asymmetric secondary alkyl groups cannot be selectively accessed with this reaction. In contrast to this chemocatalytic Friedel–Crafts alkylation, we wondered if CylK could directly mediate the alkylation of 3,5-dihydroxystilbene **20a** with secondary alkyl halide substrates in a single step. Excitingly, we observed alkylation activity of both **20a** and a related stilbene scaffold, resveratrol (**20b**), with two of our discrete electrophilic substrates. In total, we accessed four novel derivatives of benvitimod (**21a-d**) using CylK (Table 2.9). Although these alkylations proceed with relatively low conversions (1–11%), and smaller alkyl halides are not accepted by CylK as electrophiles (see Figure 2.5 for electrophile limitations), this result potentially provides an exciting starting point to engineer CylK to more efficiently access these and related bioactive scaffolds.

Table 2.9. Alkylated stilbenes accessed enzymatically with CylK



	Product	Conv. (%)		Product	Conv. (%)
21a		4	**21b**		1
21c		11	**21d**		3

Product formation (percent conversion) was quantified by HPLC, as described in Materials and Methods, values reported are the mean of two biological replicates. All enzymatic reactions were run for at 37 °C for 24 h, with 200-800 µM of alkyl chloride and 200 µM of stilbene. Turnover number (TON) defined by [product]/[catalyst].

2.3 Conclusions

In this Chapter, we highlighted our efforts to improve access to CylK via both soluble expression and enzyme refolding. We also demonstrated that this enzyme can mediate a Friedel–Crafts alkylation with an array of non-biological substrate pairs. This transformation has important features not associated with the classical organic reaction, including regioselectivity for alkylation at the 2-position of the resorcinol ring and stereospecific inversion of configuration at the alkyl chloride electrophile stereocenter. To conclusively demonstrate this in a non-native substrate context, we chemically synthesized a racemic standard of CylK product and scaled up the corresponding enzymatic reaction. NMR spectroscopy confirmed that the regioselectivity and overall connectivity of these products are identical, while chiral LC-MS analysis demonstrated that the CylK-catalyzed reaction exclusively affords a single enantiomer of product. These selectivities, coupled with CylK's use of synthetically accessible secondary alkyl halides, mild reaction conditions, and avoidance of protecting groups, all represent significant advantages of the enzymatic reaction in comparison to its chemocatalytic counterpart.

CylK can tolerate a number of different electrophilic alkyl halide substrates, including alkyl bromides and iodides. Of note, the enzyme does not accept short chain secondary alkyl halides without polar head groups. This might be due to challenges in binding such substrates, considering both hydrophobic effects and the potential necessity for a polar head group to anchor electrophilic substrates in the active site. All of the competent alkyl halide substrates contained extended alkyl chains with at least four unfunctionalized sp^3 carbons on both sides of the halogenated carbon, resembling the native substrate (Figure 2.8). We did not systematically vary the length of these alkyl groups or location of the halide substituent relative to the alkyl backbone, however, this work could reveal additional CylK promiscuity or highlight challenges and opportunities for future enzyme engineering efforts. Furthermore, the lack of CylK reactivity towards alkyl substrates with alternate leavings groups including hydroxyl, alkene, and mesylate moieties, suggests that CylK specifically engages and activates alkyl halide electrophiles for this substitution reaction.

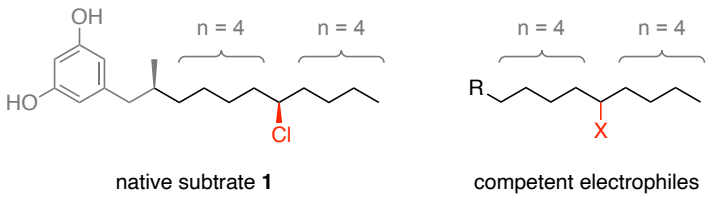


Figure 2.8. Extended alkyl portions of electrophilic substrates

On the other end, CylK can accept nucleophilic resorcinol substrates with a variety of substituents at the 4- and 5-positions of the arene ring, and we were able to use this transformation to access novel analogs of the potent AhR agonist, benvitimod. Interestingly, the CylK catalyzed reaction has a strict requirement for the two free phenolic groups of the nucleophilic resorcinol moiety, as found in the native substrate (Figure 2.9). This key feature suggests that the reactivity and/or binding mode of resorcinol substrates is essential for catalysis. Changes to the resorcinol moiety might disfavor the kinetics of the enzyme catalyzed reaction pathway or the 1,3-diol may be necessary to promote substrate binding. The 1,3-diol-containing aryl substrates that were not accepted, including nitrile, pyridine, and aniline moieties, might not be accepted due to their altered nucleophilicity, potential charge, and other factors that may influence substrate binding. Despite these limitations, however, the promiscuity of the native enzyme suggests that future enzyme engineering efforts to expand the substrate scope and efficiency of CylK might be fruitful.

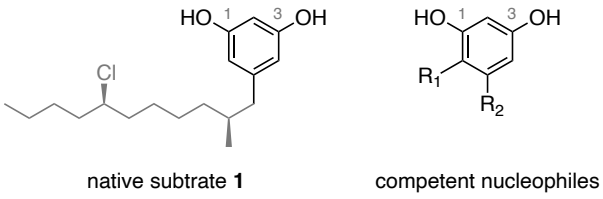
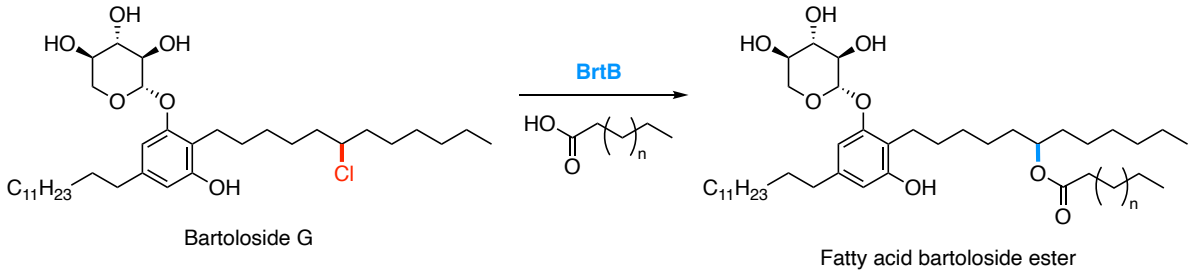


Figure 2.9. Necessary 1,3-diol relationship of nucleophilic substrates

Following this work, our collaborators, Reis et al. reported the activity of a CylK homolog, BrtB (30% amino acid identity, 46% similarity to CylK) encoded in the *brt* gene cluster that produces alkyl chloride-containing dialkylresorcinol natural products known as the bartolosides.[81,156] Interestingly, this CylK homolog catalyzes a carbon–oxygen (C–O) bond formation between the carboxylate groups of fatty acids and the alkyl chlorides of the bartolosides (Scheme 2.4). This distinct reactivity was discovered by chance, and these fatty acid derivatives had originally been overlooked in the producing cyanobacteria (*Synechocystis salina* LEGE 06099) because they do not readily elute from non-polar reversed-phase chromatography columns. Furthermore, we had assumed that BrtB was potentially an inactive CylK homolog because the resorcinols in the bartolosides bear primary alkyl groups at the 2-position and there is no other obvious nucleophilic partner produced by this biosynthetic pathway. This unexpected discovery suggests that the mechanism by which CylK binds and activates alkyl chlorides for substitution reactions might be shared with BrtB and other homologs. Furthermore, this may suggest that the family of CylK- and BrtB-like enzymes may natively use diverse nucleophilic substrates, broadening their potential applications in biocatalysis as well as the potential diversity of associated natural products. It is not yet known if there is cross nucleophile reactivity between CylK and BrtB, and our preliminary efforts have failed to demonstrate CylK-catalyzed C–O bond formation between fatty acids and CylK electrophiles.



Scheme 2.4. CylK homolog BrtB catalyzes esterification of alkyl chloride substrates

BrtB catalyzes an analogous but chemically distinct C–O bond-forming event between chlorinated bartolosides and fatty acid nucleophiles. By LC-MS, the most abundance bartoloside G ester is C16:0 (n=16), followed by C18:1 (n=18 and with one desaturation). Fatty acid bartoloside A esters are also found with similar abundance of fatty acid chain length adducts. The stereochemical configuration of bartoloside alkyl halides and ester adducts are unknown. The content of this figure is reproduced from work performed by Reis et al.[156]

In summary, CylK performs a challenging chemical reaction without precedent in biology. This enzymatic alkylation might find application in constructing diverse alkyl–aryl C–C bonds outside of natural pathways, and CylK and its homologs may also be promising starting points for enzyme and natural product discovery, enzyme engineering, and they might inspire the design of biomimetic chemocatalysts that engage secondary alkyl halides in substitution reactions. By improving access to the enzyme and characterizing its substrate scope in this section, we set the stage to begin unraveling the molecular details of its mechanism in Chapter 3. A fundamental understanding of CylK's mechanism is unknown, as are structures of enzymes that engage alkyl halides as intermediates in natural product biosynthesis.

2.4 Materials and Methods

2.4.1 General procedures and materials

All non-aqueous reactions were performed in dried glassware under an atmosphere of dry nitrogen gas. Reaction mixtures were stirred magnetically and monitored by thin layer chromatography (TLC) on Merck precoated glass-backed silica gel 60 F-254 0.25 mm plates with visualization by fluorescence quenching at 254 nm. TLC plates were stained using potassium permanganate or p-anisaldehyde TLC solutions. Concentration of reaction product solutions and chromatography fractions under reduced pressure was performed by rotary evaporation at 35–40 °C at the appropriate pressure and then at room temperature (RT), ca. 0.1 mmHg (vacuum pump) unless otherwise indicated. Yields given refer to chromatographically purified and spectroscopically pure compounds unless otherwise stated.

All chemical reagents were obtained from commercial suppliers (Sigma-Aldrich, VWR, or TCI) and used as received, unless otherwise indicated. Water was obtained by passing distilled water through a Milli-Q (Millipore Sigma) filtration system. Anhydrous tetrahydrofuran (THF), toluene, dimethylformamide (DMF), diethyl ether (Et_2O), and dichloromethane (CH_2Cl_2) were obtained by passing commercially available pre-dried, oxygen-free formulations through activated alumina columns. Lennox Luria-Bertani (LB) medium was purchased from EMD Millipore or Alfa Aesar. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Teknova. All NMR solvents were purchased from Cambridge Isotope Laboratories. Nickel-nitrilotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen and Thermo Scientific, and Strep-Tactin resin was purchased from IBA. SDS-PAGE gels were purchased from BioRad and Thermo Fischer.

Analytical high performance liquid chromatography (HPLC) was carried out with a Hypersil Gold aQ C18 column (3.0 x 150 mm, Thermo Scientific) on a Dionex UltiMate 3000 instrument (Thermo Scientific) in order to determine percent conversions by comparing the relative size of product and resorcinol starting material absorbance peaks at 210 nm. The HPLC method was as follows: 10 µL injection volume, 0.5 mL/min solvent flow rate, 25 °C column temperature, 20% solvent B in solvent A for 2 min, a gradient increasing to 100% solvent B over 16 min, 100% solvent B for 5 min, a gradient decreasing to 20% solvent

B in solvent A for 1 min, 20% solvent B in solvent A for 10 min (solvent A = water, solvent B = acetonitrile), unless otherwise specified. Preparative scale purifications were carried out with forced flow eluent at 0.3–0.5 bar on Silicycle Silica Flash F60 (230–400 Mesh) silica gel.

High-resolution mass spectral (HRMS) data were obtained on an Accurate-Mass 6530 Q-TOF LC-MS (Agilent) using dual electrospray ionization (ESI). Liquid chromatography (LC) employed a Hypersil Gold aQ C18 column (3.0 x 150 mm, Thermo Scientific) and the following method: 1 µL injection volume, 0.5 mL/min solvent flow rate, 25 °C column temperature, 20% solvent B in solvent A for 2 min, a gradient increasing to 99% solvent B over 16 min, 99% solvent B for 5 min, a gradient decreasing to 20% solvent B in solvent A for 2 min, 20% solvent B in solvent A for 5 min (solvent A = water, solvent B = acetonitrile), unless otherwise specified. The first 6 min of the LC method were diverted to waste, the subsequent eluate was analyzed by mass spectrometry (MS) performed in positive or negative ion mode, 190 °C gas temperature, 8 L/min drying gas, 22 psi nebulizer, 4500 V voltage capillary, over a 100–1700 m/z mass window in low mass mode, unless otherwise specified. Chiral separation of racemic CylK product standard (±)-**8** is achieved on a Lux 5 µM Amylose-1 column (100 x 4.6 mm, Phenomenex), with a mobile phase of 35% acetonitrile, 32.5% methanol, 32.5% water, 0.04% formic acid (20 µL injection volume, 0.6 mL/min solvent flow rate, 25 °C column temperature). All samples were collected with dual spray of a calibration reference mix.

Proton nuclear magnetic resonance (^{1}H NMR) and carbon nuclear magnetic resonance (^{13}C NMR) spectra were obtained on Varian Inova-500 (500 MHz) spectrometers at the Laukien-Purcell Instrumentation Center in the Department of Chemistry and Chemical Biology at Harvard University. Data for ^{1}H NMR chemical shifts are internally referenced to residual proton signals in $CDCl_3$ (7.26 ppm) and reported as follows: chemical shift (δ ppm), splitting patterns (br = broad, s = singlet, d = doublet, t = triplet, q = quartet, quin = quintuplet, m = multiplet), coupling constant (Hz), and integration. Chemical shifts are reported in parts per million (ppm) downfield from tetramethylsilane using the solvent resonance as the internal standard for ^{1}H

($CDCl_3$, H = 7.26 ppm, acetone-d_6 = 2.05) and ^{13}C ($CDCl_3$, C = 77.16 ppm). Furthermore, all ^{13}C spectra were measured with complete proton decoupling.

Fourier-transform infrared spectroscopy (FT-IR) data were recorded on a Bruker ALPHA FT-IR instrument and reported as wavenumber (cm^{-1}) of the absorption maxima for the range between 4000 cm^{-1} and 750 cm^{-1} with only major peaks reported.

2.4.2 Cloning of N-terminal fusions and plasmids used in this study

Plasmids pPR-IBA1-CylK,[69] pET-41 a(+), pMAL-c2X, pET His6 Sumo TEV LIC cloning vector 2S-T (gift from Scott Gradia, Addgene plasmid #29711), and pHYRS52 (gift from Hideo Iwai, Addgene plasmid #31122),[157] were used as cloning and expression vectors for the enzymes described in this study.

The MBP-CylK-Strep and GST-CylK-Strep plasmids were constructed via restriction enzyme digestion and ligation. The CylK-Strep insert was prepared by PCR amplification from pPR-IBA1-CylK using Primers 1 and 2 (Table 2.10). PCR reactions contained 1 µL of template DNA (63 ng), 0.5 µM of each primer, 50 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 100 µL. PCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 56 °C for 30 s, and elongation at 72 °C for 60 s. The amplified insert was then PCR purified (Cytiva, 28903470). Next, insert and backbone vectors were digested with BamHI and SalI (NEB) for 2 h at 37 °C. Digests contained 30 µL of DNA (1–2 µg), 4 µL 10X CutSmart Buffer (NEB), 1 µL BamHI-HF, 1 µL SalI-HF, and 4 µL of water. Following digestion, backbone vector reactions were supplemented with 1 µL of Quick CIP (NEB) and incubated at 37 °C for an additional h. All reactions were PCR purified (Cytiva, 28903470) before performing the ligation reaction. Ligations were incubated at 25 °C for 2 h and contained 50–150 fmol digested insert, 5–15 fmol digested backbone, 1 µL T4 DNA ligase (NEB), 2 µL 10X T4 DNA Ligase Buffer (NEB), and water to 20 µL. 10 µL of each ligation mixture was transformed into chemically competent *E. coli* Top10 cells. Construct identities were verified by Sanger sequencing (Eton Bioscience Inc. and Genewiz/Azenta).

The Sumo-CylK-Strep plasmid (pET-His6-Sumo-CylK-Strep) was constructed via Gibson assembly. The CylK-Strep insert was prepared by PCR amplification from pPR-IBA1-CylK using Primers 3 and 4 (Table 2.10). PCR reactions contained 1 µL template DNA (63 ng), 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 100 µL. PCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 51 °C for 30 s, and elongation at 72 °C for 60 s. The Sumo backbone was prepared by PCR amplification from pET His6 Sumo TEV LIC cloning vector 2S-T using Primers 5 and 6 (Table 2.10). PCR reactions contained 0.5 µL template DNA (36 ng), 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. PCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing and elongation at 72 °C for 200 s. All amplicons were purified before performing the Gibson assembly reaction. The reactions were incubated at 37 °C for 1 h and contained 50 ng insert, 50 ng backbone, 5 µL Gibson Assembly Master Mix (NEB), and water to 10 µL. The entire 10 µL reaction was transformed into chemically competent *E. coli* Top10 cells. Construct identities were verified by Sanger sequencing (Eton Bioscience Inc. and Genewiz/Azenta).

Table 2.10. Primers used to clone N-terminal fusions of CylK-Strep

Primer	Nucleotide Sequence (5′ to 3′)	Amplicon
1	AAGGATCCATGAAAAAAAACAAAAAAACAACAAAATCC	BamHI-CylK-Strep-SalI
2	AAGTCGACTTATTTTTCGAACTGCGGGTGG	
3	AGAGAACAGATTGGTGGGATGAAAAAAAACAAAAAAACAACA	CylK-Strep (Sumo Homology)
4	AGCAGCCGGATCAGCGCGTTATTTTTCGAACTGCGGGTGG	
5	CGCGCTGATCCGGCTGCTAAC	Sumo Backbone
6	CCCACCAATCTGTTCTCTGTGAGCC	

Restriction or homology sites underlined.

2.4.3 Overexpression and purification of CylI

CylI was expressed and purified according to a published method,[71] reproduced here: A 50 mL starter culture of LB medium supplemented with 50 µg/mL kanamycin was prepared from frozen glycerol stock of pET-28a-CylI in *E. coli* BL21 (DE3). The starter culture was incubated overnight at 37 °C with orbital shaking. The overnight culture was diluted 1:50 into 2 L of LB medium containing kanamycin. The cultures were incubated at 37 °C with shaking. At OD600 = 0.5, the cultures were cooled to 15 °C, protein expression was induced with 500 µM IPTG, and incubated for approximately 16 h. Cells were pelleted by centrifugation (6,770 x *g* for 10 min). All subsequent steps were carried out at 4 °C. Cells were resuspended in 40 mL of lysis buffer (20 mM Tris, 500 mM NaCl, 10 mM $MgCl_2$, pH 8) and lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3), and the lysate was clarified by centrifugation (20,000 x *g* for 30 min). The supernatant was incubated with 2 mL of Ni-NTA resin and 5 mM imidazole for 2 h. Following this, the supernatant was decanted and the Ni-NTA resin was resuspended in 10 mL of wash buffer (20 mM Tris, 500 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, pH 8), loaded onto a fritted glass column, and washed with 10 mL of wash buffer. Protein was washed and eluted with a stepwise imidazole gradient in wash buffer (25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 200 mM imidazole) and collecting 2 mL fractions. Fractions containing CylI were combined and dialyzed twice against 2 L of storage buffer (20 mM Tris, 50 mM NaCl, 10% v/v glycerol, pH 8), concentrated, and flash frozen in liquid nitrogen. This procedure afforded 20 mg of CylI per L of culture.[x]

2.4.4 Soluble overexpression and purification of CylK

CylK-Strep was expressed and purified according to a modified literature method.[69,128] pPR-IBA1-CylK was freshly transformed into electrocompetent *E. coli* BL21 Gold CodonPlus® (DE3) RIL cells (Agilent) before each expression, and an individual colony was used directly to prepare a 50 mL starter culture in LB medium supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. The starter culture was incubated overnight at 37 °C and 190 rpm shaking. Overnight cultures were diluted 1:100 into 2 L of LB medium

[x] The protein purification of CylI was conducted by Dr. Hitomi Nakamura.

containing ampicillin and chloramphenicol. The cultures were incubated at 37 °C with 190 rpm shaking. At OD600 = 0.4, the cultures were cooled to 15 °C and further incubated with 175 rpm shaking. When the cultures reached OD600 = 0.6-0.7, 5 mM $CaCl_2$ was added to each culture, and protein expression was induced with 250 µM IPTG. The cultures were incubated for 4 h at 15 °C with 175 rpm shaking. Cells from the 2 L culture were pelleted by centrifugation (6,770 x *g* for 15 min) and resuspended in 50 mL lysis buffer (20 mM HEPES, 500 mM NaCl, 20 mM $CaCl_2$, 10 mM $MgCl_2$, pH 8) supplemented with EDTA-free Pierce Protease Inhibitor Tablets (Thermo Fisher). All subsequent steps were carried out at 4 °C. The cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3), and the lysate was clarified by centrifugation (20,000 x *g* for 30 min). The supernatant was applied to 4 mL of *Strep*-Tactin resin preequilibrated with lysis buffer. The column was then washed with 12 mL of lysis buffer. Protein was eluted from the column in elution buffer (lysis buffer containing 2.5 mM desthiobiotin) in multiple 4 mL fractions. Fractions containing the desired protein, as determined by SDS-PAGE, were combined and dialyzed twice against 2 L of storage buffer (20 mM HEPES, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, pH 8.0, 10% glycerol). The enzyme was further purified by applying to a Superdex 200 30/100 gel filtration column (GE Healthcare) equilibrated in storage buffer and eluted at 0.5 mL/min. The fractions containing enzyme were pooled and concentrated to 37 mg/mL (500 µM), flash frozen in liquid nitrogen, and stored at –80 °C. This procedure afforded 0.1 mg of CylK-Strep per L of culture.

2.4.5 Pilot attempts to refold CylK on a small scale

CylK constructs were transformed into chemically competent *E. coli* BL21 (DE3) cells, and 5–10 colonies were used to directly inoculate 50 mL cultures in LB medium supplemented with 100 µg/mL ampicillin/carbenicillin or 50 µg/mL kanamycin. The liquid cultures were incubated at 37 °C and 200 rpm shaking. At OD400 = 0.5, the cultures were cooled to 15 °C, maintaining shaking. Following an additional 30 min, cultures were supplemented with 1 mM IPTG. The cultures were then incubated for 20 h at 15 °C with shaking. Following expression, cell pellets were collected by centrifugation (3,220 x *g* for 10 min) and resuspended in 800 µL of lysis buffer (15 mM Tris pH 8.0, 375 mM NaCl, 7.5 mM $MgCl_2$, 7.5 mM $CaCl_2$, 6 M urea). Samples were lysed by sonication (Branson, 25% amplitude, 1 min) on ice, and the soluble fraction

was collected by centrifugation (16,100 x *g* for 20 min at 4 °C). The following steps were carried out at 4 °C. The supernatant was combined with 500 µL of Ni-NTA and allowed to incubate for 30 min. The resin was washed sequentially with a stepwise gradient of imidazole in lysis buffer (5 mM, 30 mM, 50 mM) and then eluted with 250 mM imidazole. Elution fractions were diluted to approximately 1.0 µM, and 1 mL of each CylK construct was sequentially dialyzed against 250 mL of buffer 1 (17.5 mM Tris pH 8, 440 mM NaCl, 9 mM $MgCl_2$, 9 mM $CaCl_2$, 3 M urea) for approximately 14 h, 250 mL of buffer 2 (20 mM Tris pH 8, 500 mM NaCl, 10 mM $MgCl_2$,10 mM $CaCl_2$) for 3 h, and 500 mL of buffer 3 (20 mM Tris pH 8, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) for 3 h. Following dialysis, the refolded constructs were assayed for activity. Reactions were initiated by combining 44.5 µL assay buffer (20 mM HEPES pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$), 0.5 µL $CaCl_2$ (1 M), 2 µL malonyl-coenzyme A (10 mM, MilliporeSigma), 2 µL substrate (*S*,*R*)-**22** in DMSO (5 mM, see Materials and Methods 2.4.10 & 2.4.11), 1 µL CylI (219 µM, see Materials and Methods 2.4.3), with 50 µL of refolded CylK constructs. This affords access to native substrate **1** *in vitro*, that can then be dimerized/macrocyclized by CylK. Reactions were sealed and incubated at RT for 1 h or 17 h. Reactions were quenched with the addition of 1:1 methanol/acetonitrile (200 µL) and centrifuged (6,000 x *g* for 15 min). The supernatant was then subjected to analytical HPLC to determine percent conversion.

2.4.6 Insoluble overexpression, refolding, and purification of CylK

pET-His6-Sumo-CylK-Strep was freshly transformed into chemically competent *E. coli* BL21 (DE3) cells. Transformants were selected in the presence of ampicillin (100 µg/mL) and an individual colony was used directly to prepare a 100 mL starter culture of LB medium supplemented with ampicillin (LB-Amp). Following overnight incubation at 37 °C and 180 rpm shaking, 40 mL of the starter culture was used to inoculate four 2 L cultures of LB-Amp which were further incubated at 37 °C and 180 rpm shaking. At OD600 = 0.2, the cultures were cooled to 15 °C maintaining 180 rpm shaking. After 30 min, protein expression was induced with 1 mM IPTG and allowed to incubate at 15 °C for an additional 20 h. Following expression, cells were harvested by centrifugation (6,770 x *g* for 15 min), flash frozen in liquid nitrogen, and stored as pellets at –70 °C. Cell pellets were resuspended in 200 mL lysis buffer (15 mM Tris pH 8, 375 mM NaCl, 7.5 mM

$MgCl_2$, 7.5 mM $CaCl_2$, 6 M urea) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet (Thermo Fisher) and 5 mM imidazole. The cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The cell lysate was clarified twice by centrifugation (20,000 x *g* for 15 min) and the supernatant was applied to 20 mL of Ni NTA resin equilibrated in lysis buffer + 5 mM imidazole. The column was sequentially washed with 2 column volumes (CVs) of lysis buffer + 5 mM imidazole, and 2 CVs of modified lysis buffer (3 M urea) + 30 mM imidazole. Bound species were eluted with lysis buffer + 250 mM imidazole. Elution fractions were pooled and diluted to 200 mL with lysis buffer for a final protein concentration of 2-4 μM, as measured by absorbance at 280 nm. All subsequent steps were carried out at 4 °C. The enzyme was refolded by sequentially dialyzing this solution against 2 L of buffer 1 (17.5 mM Tris pH 8, 440 mM NaCl, 9 mM $MgCl_2$, 9 mM $CaCl_2$, 3 M urea) for 10 h, 2 L of buffer 2 (20 mM Tris pH 8, 500 mM NaCl, 10 mM $MgCl_2$,10 mM $CaCl_2$) for 3 h, and 2 L of buffer 3 (20 mM Tris pH 8, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 5% glycerol) for 3 h. Following dialysis, the refolded enzyme solution was centrifuged (100,000 x *g* for 30 min) to pellet soluble aggregates and the supernatant was incubated overnight with ULP protease (2 mg, 1:25 w/w) in order to cleave off the Sumo-tag, providing CylK-Strep. After confirming proteolysis by SDS-PAGE, a subtractive Ni NTA step (10 mL resin equilibrated in buffer 3 + 5 mM imidazole) removed trace uncleaved enzyme and free Sumo-tag. The enzyme was further purified by applying to a Superdex 200 26/600 gel filtration column (GE Healthcare) equilibrated in storage buffer (20 mM HEPES pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol) and eluted at 0.5 mL/min. The fractions containing enzyme were pooled and concentrated to 37 mg/mL (500 μM), flash frozen in liquid nitrogen, and stored at –70 °C. This procedure afforded 2.0 mg of CylK-Strep per L of culture.

2.4.7 Optimized refolding and purification of CylK

CylK was expressed and purified as described previously (see Materials and Methods 2.4.6) with several modifications detailed below. Following overexpression in *Escherichia coli* BL21(DE3) as a Sumo fusion construct, cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet. Cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The pellet was collected by centrifugation (20,000

× *g* for 30 min) and resuspended in lysis buffer supplemented with 6 M urea and 5 mM imidazole. The denatured enzyme solution was clarified twice by centrifugation (20,000 × *g* for 15 min), and the supernatant was applied to 20 mL of Ni NTA resin. The column was washed with lysis buffer containing 6 M urea and 5 mM imidazole. Bound species were eluted with lysis buffer containing 6 M urea and 250 mM imidazole. All subsequent steps were carried out at 4 °C. The enzyme was refolded at a concentration of 1-10 μM by sequentially dialyzing this solution against buffer 1 (lysis buffer supplemented with 3 M urea) overnight, buffer 2 (lysis buffer lacking urea) for 4 h, and buffer 3 (lysis buffer supplemented with 5% glycerol) for 4 h. Following dialysis, the refolded enzyme solution was centrifuged (100,000 × *g* for 30 min), and the supernatant was collected and incubated overnight with ULP protease (1:25 w/w). Then, following a subtractive Ni NTA step, the enzyme was applied to a Superdex 200 26/600 gel filtration column (GE Healthcare) equilibrated in protein storage buffer (20 mM HEPES pH 7.8, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol). The fractions containing pure CylK were pooled, concentrated (500–700 μM, determined by extinction coefficient and absorbance at 280 nm), and flash frozen in liquid nitrogen. This procedure afforded 4.5 mg of CylK-Strep per L of culture. Representative SDS-PAGE and size exclusion chromatogram can be found below (Figure 2.10).

2.4.8 Overexpression and purification of ULP protease

ULP protease was expressed and purified according to a modified literature method.[158] pHYRS52 (gift from Hideo Iwai, Addgene plasmid #31122) was freshly transformed into chemically competent *E. coli* BL21 (DE3) cells. Transformants were selected in the presence of ampicillin (100 μg/mL) and two plates of transformants were used to directly inoculate two 2 L cultures of LB medium supplemented with ampicillin (LB-Amp). These cultures were incubated at 37 °C and 180 rpm shaking. At OD600 = 0.4, the cultures were cooled to 15 °C with 200 rpm shaking. After 30 min, protein expression was induced with 1 mM IPTG and allowed to incubate at 15 °C for an additional 6 h. Following expression, cells were harvested by centrifugation (6,770 x *g* for 15 min), flash frozen in liquid nitrogen, and stored as pellets at –70 °C. Cell pellets were resuspended in 50 mL lysis buffer (50 mM Tris pH 8, 200 mM NaCl, 2 mM 2-mercaptoethanol) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet (Thermo Fisher) and 5 mM imidazole.

The cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The cell lysate was clarified twice by centrifugation (20,000 x *g* for 15 min) and the supernatant was applied to 5 mL of Ni NTA resin equilibrated in lysis buffer + 5 mM imidazole. The column was sequentially washed with 10 column volumes (CVs) of lysis buffer + 5 mM imidazole, lysis buffer + 30 mM imidazole, and lysis buffer + 50 mM imidazole. Bound species were eluted with lysis buffer + 250 mM imidazole. Elution fractions were pooled and concentrated to 6 mg/mL (220 μM), flash frozen in liquid nitrogen, and stored at –70 °C. This procedure afforded 6.0 mg of ULP protease per L of culture.

2.4.9 Measuring native CylK activity in lysate

CylK constructs were freshly transformed into chemically or electrocompetent *E. coli* BL21 Gold CodonPlus® (DE3) RIL cells (Agilent), and 5–10 colonies were used to directly inoculate 50 mL cultures in LB medium supplemented with 100 μg/mL ampicillin or carbenicillin and 34 μg/mL chloramphenicol. The liquid cultures were incubated at 37 °C and 190 rpm shaking. At OD400 = 0.4, the cultures were cooled to 15 °C, maintaining shaking. Following an additional 1 h, cultures were supplemented with 5 mM $CaCl_2$ and protein expression was induced with 250 μM IPTG (unless noted). The cultures were then incubated for 4 h at 15 °C with 190 rpm shaking (unless noted). Following expression, cell pellets were collected by centrifugation (3,220 x *g* for 10 min) and resuspended in 800 μL of lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$). Samples were lysed by sonication (Branson, 25% amplitude, 1 min) on ice, and the soluble fraction was collected by centrifugation (16,100 x *g* for 20 min at 4 °C) and analyzed by SDS-PAGE and/or used to assay *in vitro* activity. Reactions were initiated by combining 43.2 μL assay buffer (20 mM HEPES pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$), 0.5 μL $CaCl_2$ (1 M), 2 μL malonyl-coenzyme A (10 mM, MilliporeSigma), 2 μL substrate (*S,R*)-**22** in DMSO (5 mM, see Materials and Methods 2.4.10 & 2.4.11), 2.3 μL CylI (219 μM, see Materials and Methods 2.4.3), with 50 μL of clarified CylK lysate. This affords access to native substrate **1** *in vitro*, that can then be dimerized/macrocyclized by CylK. Reactions were sealed and incubated at RT for 50 min (unless noted). Reactions were quenched with the addition of 1:1 methanol/acetonitrile (200 μL) and centrifuged (6,000 x *g* for 15 min). The supernatant was then subjected to analytical HPLC to determine percent conversion.

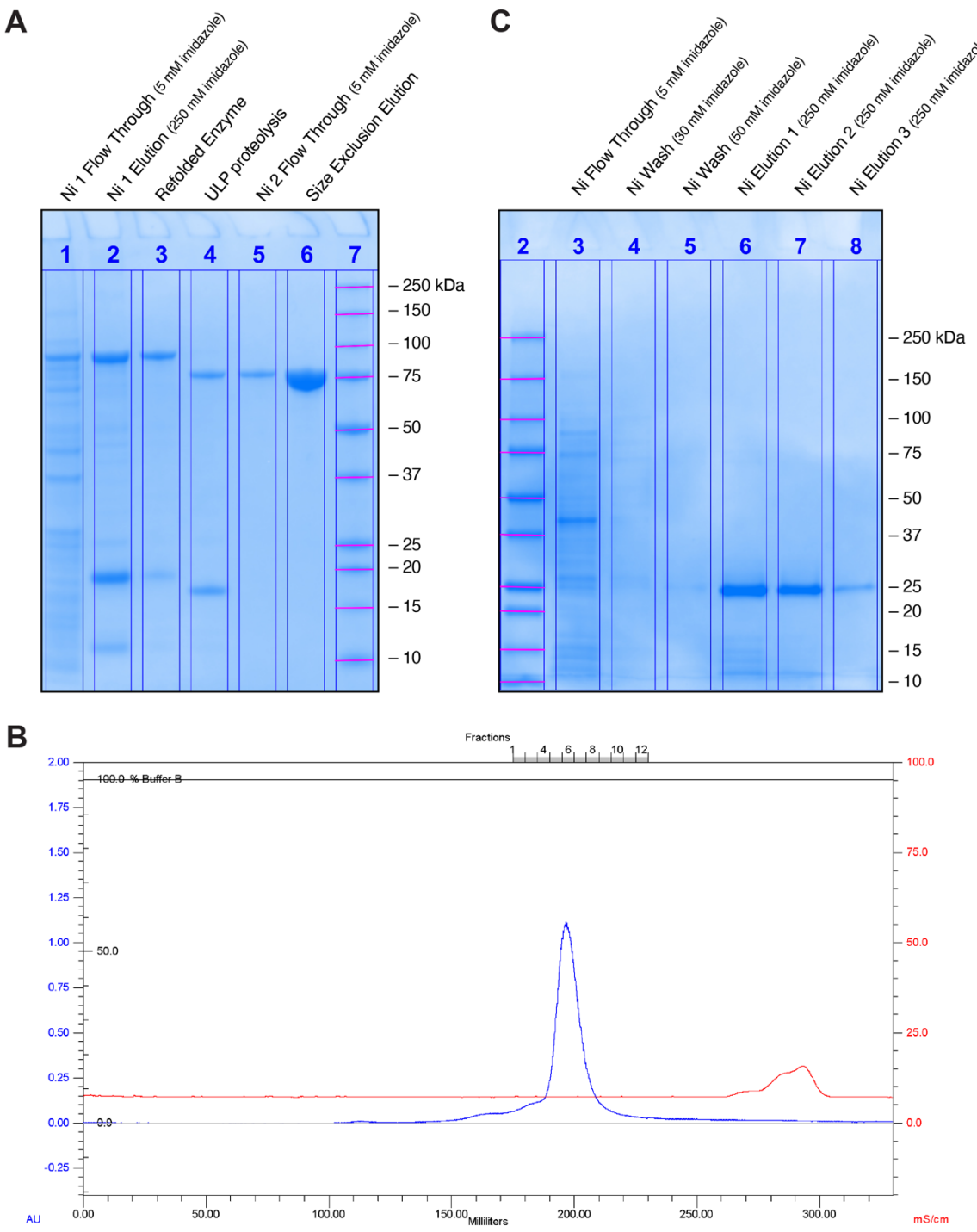
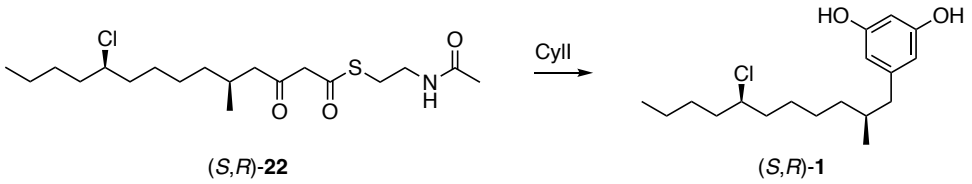


Figure 2.10. Protein SDS-PAGE and size exclusion chromatogram

(*A*) SDS-PAGE and (*B*) size exclusion chromatogram of refolded CylK. (*C*) SDS-PAGE of ULP protease purification. All Blue Precision Plus Protein Prestained Standard (Bio-Rad) was used to visualize molecular weight standards. Size exclusion chromatography run on Superdex 200 26/600 gel filtration column (GE Healthcare). Absorbance measured at 280 nm (blue line).

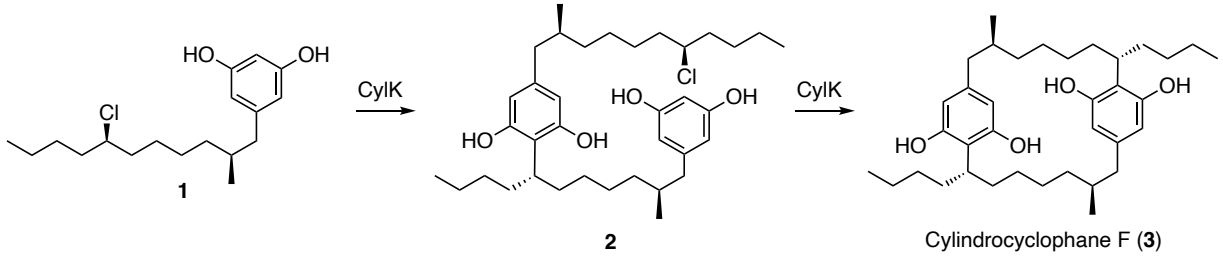
2.4.10 Experimental Data for Enzymatic Reactions



(*S*,*R*)-**22** (*S*,*R*)-**1**

5-((2*S*,7*R*)-7-chloro-2-methylundecyl)benzene-1,3-diol ((*S*,*R*)-**1**)

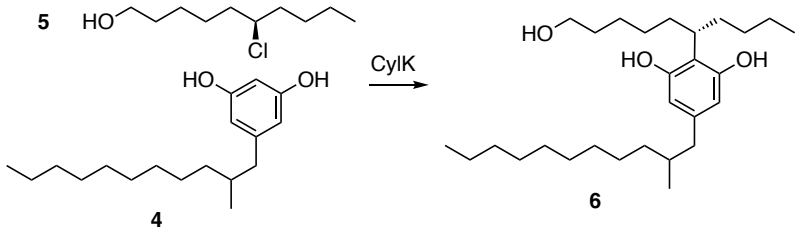
(*S*,*R*)-**1**. Prepared according to a previously reported literature procedure[69] using enantiomerically pure (*S*,*R*)-**22** and malonyl-coenzyme A (Malonyl-CoA) as starting material, with heterologously expressed and purified CylI enzyme (see Materials and Methods 2.4.3) to catalyze resorcinol formation *in vitro*. HPLC and LC-MS analyses matched retention time (16.5 min) and m/z ($C_{18}H_{28}ClO_2$ [M - H^+] 311.1783) reported in the literature.[69] A maximum UV/Vis absorbance at 205 nm is diagnostic of **1**, while **22** does not have significant absorbance measured near 200-280 nm.



1 **2** Cylindrocyclophane F (**3**)

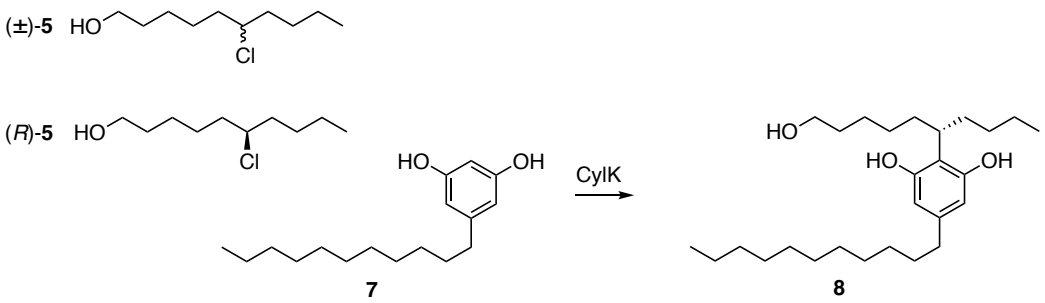
mono-alkylated intermediate (**2**) & cylindrocyclophane F (**3**)

2 & **3**. Prepared according to a previously reported literature procedure[69] using enantiomerically pure (*S*,*R*)-**1** accessed *in situ* by combining (*S*,*R*)-**22**, Malonyl-CoA, and purified CylI enzyme. Subsequently, CylK is added as lysate, clarified lysate, or as purified enzyme to access **2** and **3**. HPLC and LC-MS analyses matched retention times (19 min for **2**, 18 min for **3**) and m/z ($C_{36}H_{56}ClO_4$ [M - H^+] 587.3873 for **2**, $C_{36}H_{55}O_4$ [M - H^+] 551.4106 for **3**) reported in the literature.[69] A maximum UV/Vis absorbance of 207 nm is diagnostic of **2**, and 209 nm is diagnostic of **3**, while 205 is diagnostic of **1**, if measured near 200-280 nm. Note, each alkylation event increases the max absorbance of these resorcinol species by ~2 nm from starting material (**1**) to intermediate (**2**) to cylindrocyclophane F final product (**3**).



2-((*S*)-10-Hydroxydecan-5-yl)-5-(2-methylundecyl)benzene-1,3-diol (**6**)

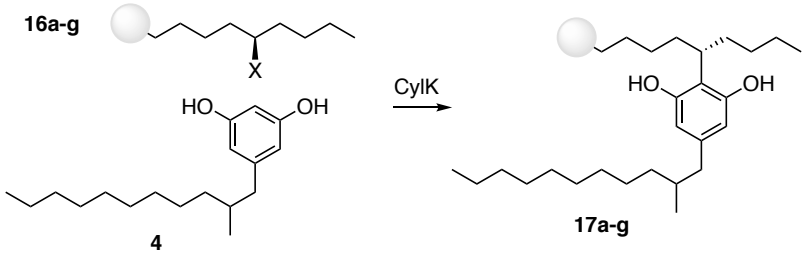
(*S*)-**6.** To an Eppendorf tube containing reaction buffer (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$) were added resorcinol **4** (from 10 mM solution in DMSO), alkyl chloride (*R*)-**5** (from 10 mM solution in DMSO), and freshly thawed CylK-Strep (from 500 μM solution in storage buffer [20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol]) to a final volume of 100 μL or 1.0 mL. The reaction mixture was gently mixed and allowed to incubate at 23 or 37 °C for 17–24 h. Upon completion, the reaction was quenched with the addition of 0 °C 1:1 methanol:acetonitrile (200 μL or 2.0 mL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatant was then subjected to analytical HPLC and HRMS to measure percent conversion and ion abundance, respectively. As a control, reactions without enzyme were performed. HRMS (ESI) calculated for $C_{28}H_{49}O_3$ [M - H^+] 433.3687, found 433.3698 only in the reactions incubated with CylK-Strep.



(*S*)-2-(10-Hydroxydecan-5-yl)-5-undecylbenzene-1,3-diol ((*S*)-**8**)

(*S*)-**8.** To an Eppendorf tube or VWR amber glass vial containing reaction buffer (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$) were added resorcinol **7** (from 40 mM solution in DMSO), alkyl chloride (*R*)-**5** or (±)-**5** to (from 40 mM solution in DMSO), and freshly thawed CylK-Strep to 1.0 μM (from 500 μM solution in storage buffer [20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol]) to a final volume of 30 μL or 1.0 mL. The reaction mixture was allowed to incubate at 37 °C for

24 h, with 150 rpm shaking. Upon completion, the reaction was quenched (or 30 μL of 1 mL reactions) with the addition of 1:1 methanol/acetonitrile (60 μL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatant was then subjected to analytical HPLC and HRMS to measure percent conversion and ion abundance, respectively. Enantiomeric excess was determined to be >99% or 88% by comparing (*S*)-**8** to a synthetic racemic standard **8** via chiral LC-MS. As a control, reactions without enzyme were performed. (*S*)-**8** was isolated by combining fifty 1 mL-scale reactions. The resultant mixture was purified by flash column chromatography (C18 silica gel, 1:4 to 1:0 acetonitrile/H_2O) to afford (*S*)-**8** as a white solid (3.1 mg, 73% yield). R_f = 0.41 (silica gel, 1:1 hexanes:EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.15 (s, 2 H), 4.60 (s, 2 H), 3.58 (q, *J* = 6.2 Hz, 2 H), 3.14–2.99 (m, 1 H), 2.47–2.38 (m, 2 H), 1.95–1.76 (m, 2 H), 1.68–1.58 (m, 2 H), 1.58—1.42 (m, 4 H), 1.43–1.10 (m, 24 H + OH), 0.88 (t, *J* = 6.8 Hz, 3 H), 0.84 (t, *J* = 7.1 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 155.2, 142.2, 114.9, 108.9, 63.1, 35.6, 35.5, 33.9, 33.8, 32.7, 32.1, 31.1, 30.7, 29.8, 29.8, 29.7, 29.7, 29.6, 29.5, 27.9, 25.7, 23.0, 22.8, 14.3; HRMS (ESI) calculated for $C_{27}H_{47}O_3$ [M - H^+] 419.3531, found 419.3531 only in the reactions incubated with CylK-Strep.



General procedure for CylK reactions with alkyl chloride electrophiles

17a-g. To Eppendorf tubes containing reaction buffer (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$) were added resorcinol **4** to 100 μM (from 10 mM solution in DMSO), electrophiles **16a-g** to 200 μM (from 10 mM solution in DMSO), and freshly thawed CylK-Strep to 1.0 μM (from 500 μM solution in storage buffer [20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol]) to a final volume of 100 μL. The reaction mixtures were gently mixed and allowed to incubate at 37 °C for 24 h. Upon completion, the reactions were quenched with the addition of 0 °C 1:1 methanol:acetonitrile (200 μL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatants were then subjected to analytical

HPLC and HRMS to measure percent conversion and ion abundance, respectively. As a control, reactions without enzyme were performed. Product masses were found in only the reactions incubated with CylK-Strep.

Table 2.11. Alkyl halide electrophile product masses.

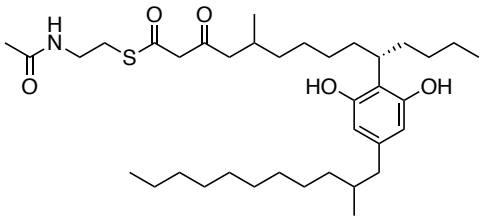
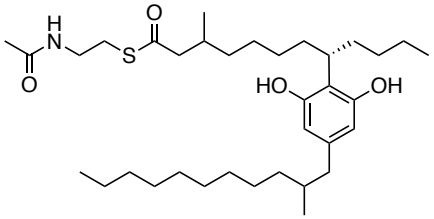
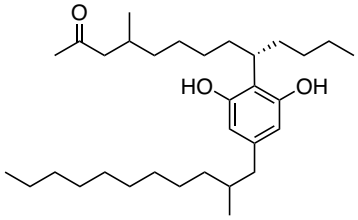
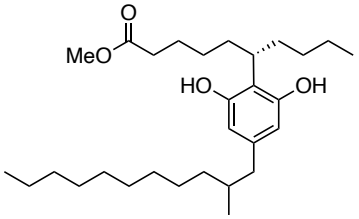
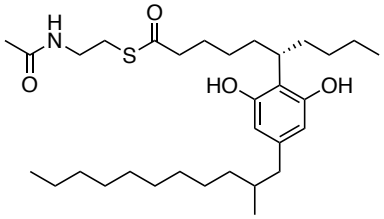
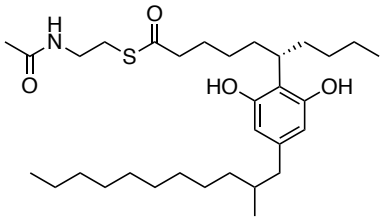
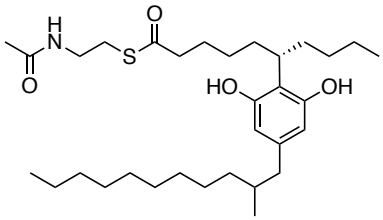
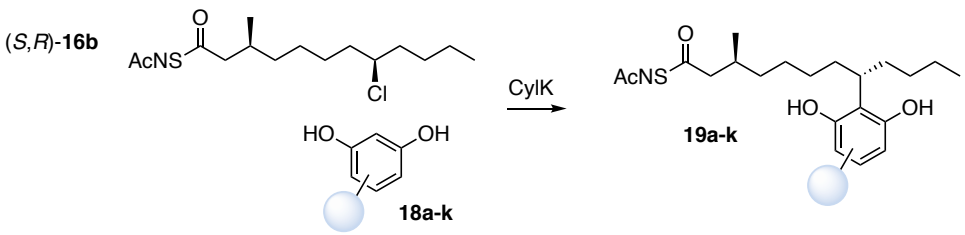
	Product	Exact Mass	HRMS(ESI) Found
17a		$[M+H^+]$ $C_{37}H_{64}NO_5S$ 634.4500	634.4507
17b		$[M+H^+]$ $C_{35}H_{62}NO_4S$ 592.4394	592.4381
17c		$[M-H^+]$ $C_{32}H_{55}O_3$ 487.4157	487.4157
17d		$[M+H^+]$ $C_{29}H_{49}O_4$ 463.3782	463.3776
17e		$[M+H^+]$ $C_{32}H_{56}NO_4S$ 550.3925	550.3933

Table 2.11. (Continued)

17f		[M+H⁺] $C_{32}H_{56}NO_4S$ 550.3925	550.3924
17g		[M+H⁺] $C_{32}H_{56}NO_4S$ 550.3925	550.3910



General procedure for CylK reactions with resorcinol nucleophiles

19a-k. To Eppendorf tubes containing reaction buffer (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$) were added electrophile (*S*,*R*)-**16b** to 200 μM (from 10 mM solution in DMSO), various commercially available resorcinol nucleophiles **18a-k** to 200 μM (from 10 mM solution in DMSO), and freshly thawed CylK-Strep to 1.0 μM (from 500 μM solution in storage buffer [20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol]) to a final volume of 100 μL. The reaction mixtures were gently mixed and allowed to incubate 37 °C for 24 h. Upon completion, the reactions were quenched with the addition of 0 °C 1:1 methanol:acetonitrile (200 μL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatants were then subjected to analytical HPLC and HRMS to measure percent conversion and ion abundance, respectively. The HPLC method was modified as follows: 0% solvent B in solvent A for 2 min, a gradient increasing to 100% solvent B over 16 min, 100% solvent B for 5 min, a gradient decreasing to 0% solvent B in solvent A for 1 min, 0% solvent B in solvent A for 10 min (solvent A = H_2O, solvent B =

acetonitrile). LC-MS method was modified as follows: 1% solvent B in solvent A for 2 min, a gradient increasing to 99% solvent B over 16 min, 99% solvent B for 5 min, a gradient decreasing to 1% solvent B in solvent A for 2 min, 1% solvent B in solvent A for 5 min equilibration (solvent A = H_2O, solvent B = acetonitrile). The first 3 min of the LC method were diverted to waste, and the subsequent eluate was analyzed by mass spectrometry (MS) performed in positive ion mode. As a control, reactions without enzyme were performed. Product masses were found in only the reactions incubated with CylK-Strep.

Table 2.12. Resorcinol nucleophile product masses.

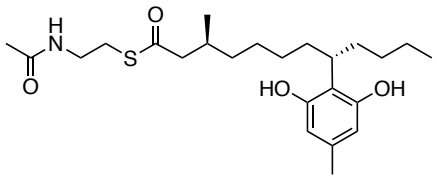
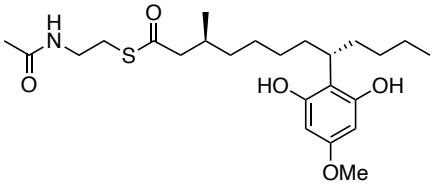
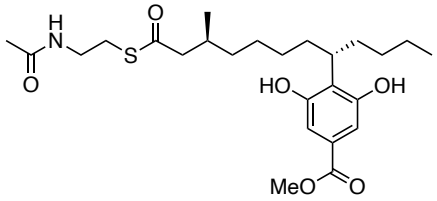
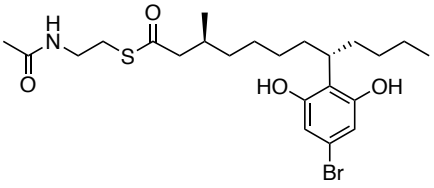
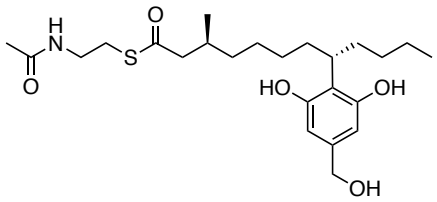
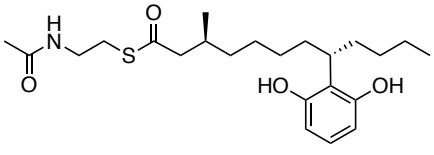
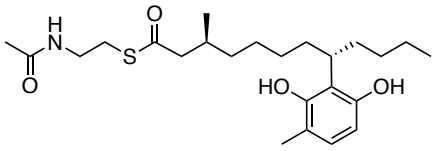
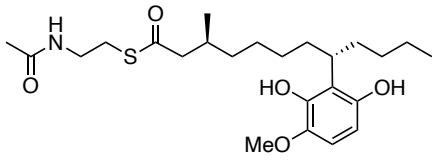
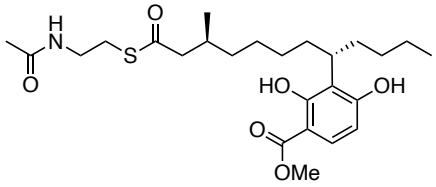
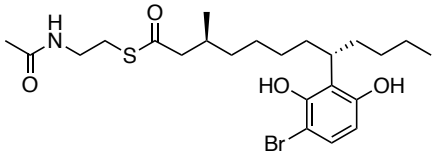
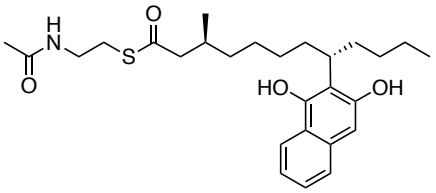
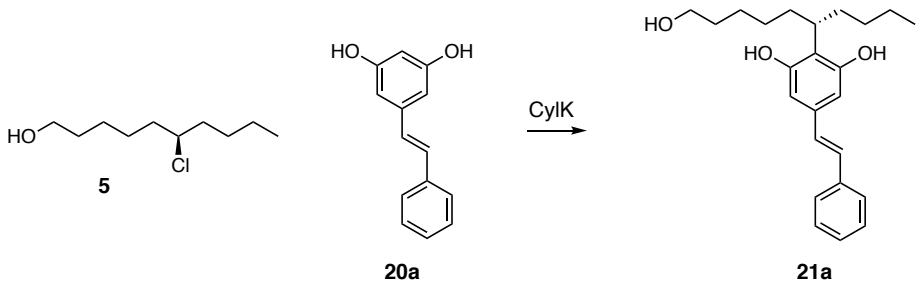
	Product	Exact Mass	HRMS(ESI) Found
19a		[M+H+] $C_{24}H_{40}NO_4S$ 438.2673	438.2672
19b		[M+H+] $C_{24}H_{40}NO_5S$ 454.2622	454.2649
19c		[M+H+] $C_{25}H_{40}NO_6S$ 482.2571	482.2600
19d		[M+H+] $C_{23}H_{37}BrNO_4S$ 502.1621, 504.1601	502.1652, 504.1633
19e		[M+H+] $C_{24}H_{40}NO_5S$ 454.2622	454.2643
19f		[M+H+] $C_{23}H_{38}NO_4S$ 424.2516	424.2540

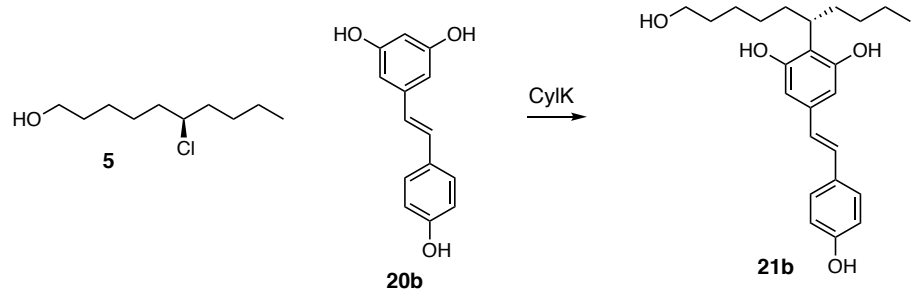
Table 2.12. (Continued)

19g		[M+H+] $C_{24}H_{40}NO_4S$ 438.2673	438.2693
19h		[M+H+] $C_{24}H_{40}NO_5S$ 454.2622	454.2631
19i		[M+H+] $C_{25}H_{40}NO_6S$ 482.2571	482.2571
19j		[M+H+] $C_{23}H_{37}BrNO_4S$ 502.1621, 504.1601	502.1637, 504.1620
19k		[M+H+] $C_{27}H_{40}NO_4S$ 474.2673	474.2661



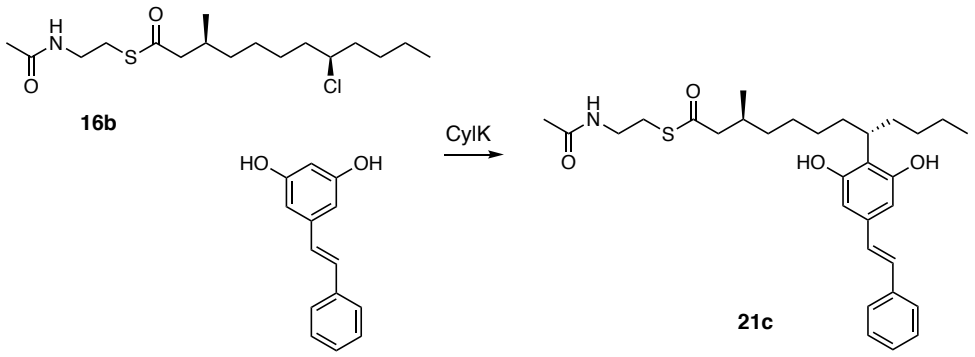
(*S*,*E*)-2-(10-Hydroxydecan-5-yl)-5-styrylbenzene-1,3-diol (**21a**)

21a. To an Eppendorf tube containing reaction buffer (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$) were added stilbene **20a** to 200 μM (0.15 μL, 40 mM stock in DMSO), alkyl chloride (*R*)-**5** to 800 μM (0.60 μL, 40 mM stock in DMSO), and freshly thawed CylK-Strep to 0.5 μM (0.4 μL, 37 μM in storage buffer [20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol]) to a final volume of 30 μL. The reaction mixture was gently mixed and allowed to incubate at 37 °C for 24 h. Upon completion, the reaction was quenched with the addition of 1:1 methanol:acetonitrile (60 μL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatant was then subjected to analytical HPLC and HRMS to measure percent conversion (4%) and ion abundance, respectively. As a control, reactions without enzyme were performed. HRMS (ESI) calculated for $C_{24}H_{32}O_3$ [M - H^+] 367.2279, found 367.2281 only in the reaction incubated with CylK-Strep.



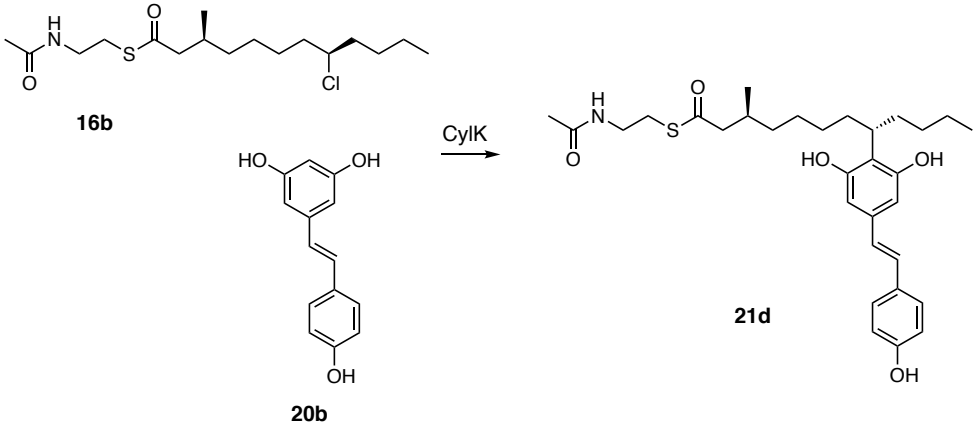
(*S*,*E*)-2-(10-Hydroxydecan-5-yl)-5-(4-hydroxystyryl)benzene-1,3-diol (**21b**)

21b. Prepared according to the procedure for **21a**, substituting commercially available resveratrol (**20b**) as the nucleophile, 200 μM (0.15 μL, 40 mM stock in DMSO). HPLC: 1% Conversion. HRMS (ESI) calculated for $C_{24}H_{31}O_4$ [M - H^+] 383.2228, found 383.2227 only in the reaction incubated with CylK-Strep.



S-(2-Acetamidoethyl) (3*S*,8*S*)-8-(2,6-dihydroxy-4-((*E*)-styryl)phenyl)-3-methyldodecanethioate (**21c**)

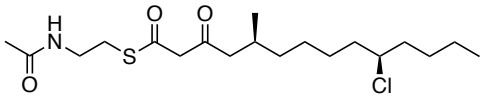
21c. Prepared according to the procedure for **21a**, substituting (*S,R*)-**16b** as the electrophile, 200 μM (0.60 μL, 10 mM stock in DMSO). HPLC: 11% Conversion. HRMS (ESI) calculated for $C_{31}H_{44}NO_4S$ [M + H^+] 526.2986, found 528.2988 only in the reaction incubated with CylK-Strep.



S-(2-Acetamidoethyl) (3*S*,8*S*)-8-(2,6-dihydroxy-4-((*E*)-4-hydroxystyryl)phenyl)-3-methyldodecanethioate (**21d**)

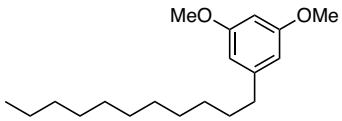
21d. Prepared according to the procedure for **21a**, substituting (*S,R*)-**16b** as the electrophile, 800 μM (0.60 μL, 10 mM stock in DMSO) and commercially available resveratrol (**20b**) as the nucleophile, 200 μM (0.15 μL, 40 mM stock in DMSO). HPLC: 3% Conversion. HRMS (ESI) calculated for $C_{31}H_{44}NO_5S$ [M + H^+] 542.2935, found 542.2916 only in the reaction incubated with CylK-Strep.

2.4.11 Experimental Data for Synthetic Substrates and Standards



S-(2-acetamidoethyl) (5*S*,10*R*)-10-chloro-5-methyl-3-oxotetradecanethioate ((*S*,*R*)-**22**)

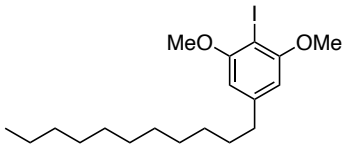
(*S*,*R*)-**22.** Prepared according to known literature procedures,[69] using enantiomerically pure (3*S*,8*R*)-8-chloro-3-methyldodecanoic acid as starting material, as prepared in the literature.[69] All spectra matched data previously reported in the literature.[69] ^{1}H NMR (500 MHz, $CDCl_3$): Keto form: δ 5.92 (br s, 1 H, NH), 3.94–3.82 (m, 1 H), 3.72 (s, 2 H), 3.50–3.43 (m, 2 H), 3.12–3.07 (m, 2 H), 2.51 (dd, *J* = 16.6, 5.7 Hz, 1 H), 2.34 (dd, *J* = 16.6, 7.8 Hz, 1 H), 1.98 (s, 3 H), 1.98–1.88 (m, 1 H), 1.75–1.64 (m, 4 H) 1.56–1.15 (m, 10 H), 0.95–0.88 (m, 6 H); Enol form: δ 12.60 (s, 1 H), 5.92 (br s, 1 H, NH), 5.45 (s, 1 H), 3.92–3.83 (m, 1 H), 3.50–3.43 (m, 2 H), 3.12–3.07 (m, 2 H), 2.18 (dd, *J* = 7.8, 2.3 Hz, 1 H), 2.08–19.5 (m, 1 H), 1.98 (s, 3 H,), 1.97–1.88 (m, 1 H), 1.75–1.64 (m, 4 H) 1.56–1.15 (m, 10 H), 0.96–0.88 (m, 6 H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 202.2, 194.4, 192.5, 177.0, 170.6, 170.5, 100.6, 64.5, 57.9, 51.0, 51.0, 42.8, 42.7, 40.1, 39.5, 38.7, 38.5, 36.8, 31.3, 29.9, 29.5, 29.1, 28.9, 28.1, 26.8, 26.7, 23.5, 23.4, 22.5, 19.9, 19.8, 14.2.



1,3-Dimethoxy-5-undecylbenzene (**9**)

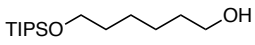
9. 1-Bromoundecane (2.85 mL, 3.00g, 12.8 mmol, 1.00 equiv) was added dropwise to a degassed mixture of $Pd(OAc)_2$ (143 mg, 0.638 mmol, 5 mol%), di-*t*-butylmethylphosphonium tetrafluoroborate (315 mg, 1.28 mmol, 10 mol%), KO*t*Bu (4.30 g, 38.2 mmol, 3.00 equiv), and (3,5-dimethoxyphenyl)boronic acid (3.48 g, 19.1 mmol, 1.50 equiv) in anhydrous *t*-amyl alcohol (60 mL) at RT. The resulting suspension was stirred at RT for 18 h before being diluted with Et_2O (50 mL) and filtered over celite® (Et_2O rinse). The filtrate was concentrated under reduced pressure and the resulting residue was purified by flash column chromatography (hexanes:EtOAc 100:0 to 50:1) affording the desired product **9** (3.60 g, 98%) as a colorless oil. R_f = 0.35 (silica gel, hexanes:EtOAc 50:1); IR (film) ν_{max} 2997, 2922, 2852, 1594, 1460, 1427, 1347,

1323, 1292, 1204, 1148, 1060, 993, 925, 827, 722, 695, 540 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.35 (s, 2 H), 6.30 (s, 1 H), 3.78 (s, 6 H), 2.54 (t, *J* = 15.0 Hz, 2 H), 1.63–1.57 (m, 2 H), 1.42–1.18 (m, 16 H), 0.88 (t, *J* = 15.0 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 160.8, 145.6, 106.6, 97.7, 55.4, 36.5, 32.1, 31.4, 29.8, 29.8, 29.7, 29.5, 29.5, 29.5, 22.8, 14.3; HRMS (ESI): calculated for $C_{19}H_{33}O_2^+$ [M + H^+] 293.2475, found 293.2476.



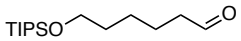
2-Iodo-1,3-dimethoxy-5-undecylbenzene (**10**)

10. *n*-BuLi (2.5 M in hexanes; 5.42 mL, 13.5 mmol, 1.10 equiv) was added dropwise to a solution of dimethoxybenzene **9** (3.60 g, 12.3 mmol, 1.00 equiv) in THF (28 mL) at 0 °C. Stirring at 0 °C was continued for 1.5 h before a solution of iodine (3.75 g, 14.8 mmol, 1.20 equiv) in THF (28 mL) was added dropwise. The resulting mixture was allowed to warm to RT and stirring was continued for 2 h. Upon completion, the mixture was cooled to 0 °C and quenched with the dropwise addition of H_2O (15 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with saturated aqueous $Na_2S_2O_3$ (2 x 15 mL), H_2O (15 mL) and brine (2 x 15 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure to afford an orange oil. Purification by flash column chromatography (hexanes:EtOAc 100:0 to 50:1) afforded the desired aryl iodide **10** (4.66 g, 90%) as pale yellow oil. R_f = 0.18 (silica gel, hexanes:EtOAc 50:1); IR (film) ν_{max} 3367, 3001, 2921, 2851, 1578, 1453, 1409, 1377, 1345, 1309, 1234, 1205, 1149, 1122, 1049, 1019, 820, 721, 669, 655, 614, 514 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.34 (s, 2 H), 3.88 (s, 6 H), 2.58 (t, *J* = 15.0 Hz, 2 H), 1.64–1.58 (m, 2 H), 1.38–1.22 (m, 16 H), 0.88 (t, *J* = 10.0 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 159.4, 145.7, 104.7, 74.0, 56.6, 36.6, 32.1, 31.6, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 22.8, 14.3; HRMS (ESI): calculated for $C_{19}H_{32}IO_2^+$ [M + H^+] 419.1441, found 419.1439.



6-((Triisopropylsilyl)oxy)hexan-1-ol (**23**)

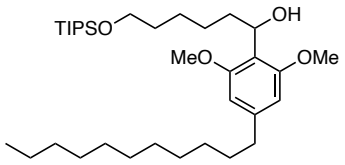
23. A procedure from the literature[159] was modified as follows: Imidazole (5.30 g, 78.0 mmol, 1.50 equiv) was added in one portion to a solution of hexane-1,6-diol (18.4 g, 156 mmol, 3.00 equiv) in CH_2Cl_2 (258 mL) at RT. The resulting mixture was stirred until a homogeneous suspension was obtained and then cooled to 10 °C before chlorotriisopropylsilane (11.1 mL, 10.0 g, 51.9 mmol, 1.00 equiv) was added dropwise over 10 min. Stirring was continued at RT for 18 h. The resulting mixture was washed with saturated aqueous NH_4Cl (2 x 30 mL), H_2O (30 mL), and brine (30 mL), dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 7:3) afforded the mono-protected diol **23** (14.1g, 99%) as colorless oil. R_f = 0.38 (silica gel, hexanes:EtOAc 7:3); ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.68 (t, *J* = 6.5 Hz, 2 H), 3.65 (t, *J* = 6.6 Hz, 2 H), 1.70–1.42 (m, 4 H + OH), 1.42–1.33 (m, 4 H), 1.14–0.97 (m, 21 H). These spectral characteristics are identical to those previously reported.[159]



6-((Triisopropylsilyl)oxy)hexanal (**11**)

11. A procedure from the literature[160] was repeated as reported: DMSO (1.98 mL, 2.18 g, 32.1 mmol, 2.20 equiv) was added to a solution of the alcohol **23** (4.00 g, 14.6 mmol, 1.00 equiv) in CH_2Cl_2 (92 mL) at –78 °C. The reaction mixture was stirred for 5 min before $(COCl)_2$ (1.38 mL, 2.03 g, 16.0 mmol, 1.10 equiv) was added dropwise over 30 min. Stirring was continued for 10 min at –78 °C followed by the dropwise addition of Et_3N (10.2 mL, 7.37 g, 72.9 mmol, 5.00 equiv) over 5 min. The reaction flask was kept at –78 °C for 30 min, then warmed to RT and stirred for 2 h. The brown reaction mixture was diluted with H_2O (50 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (3 x 30 mL) and the combined organic extracts were washed with saturated aqueous NH_4Cl (2 x 30 mL) and brine (30 mL) before being dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 15:1) afforded the desired aldehyde **11** (3.45 g, 88%) as colorless oil. R_f

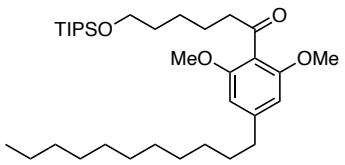
= 0.38 (hexanes:EtOAc 10:1); ^{1}H NMR (400 MHz, $CDCl_3$): δ 9.77–9.76 (m, 1 H), 3.68 (t, *J* = 6.1 Hz, 2 H), 2.46–2.41 (m, 2 H), 1.79–1.50 (m, 4 H), 1.45–1.38 (m, 2 H), 1.10–0.98 (m, 21 H). These spectral characteristics were identical to those previously reported.[160] Note: Aldehyde **11** should be stored at –20 °C as a solution in $CHCl_3$ and used within a few days for best results.



1-(2,6-Dimethoxy-4-undecylphenyl)-6-((triisopropylsilyl)oxy)hexan-1-ol (**12**)

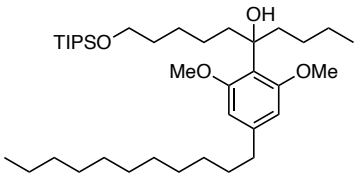
12. *n*-BuLi (2.5 M in hexanes; 3.10 mL, 7.77 mmol, 1.30 equiv) was added dropwise over 15 min to a solution of iodobenzene **10** (2.50 g, 5.98 mmol, 1.00 equiv) in THF (40 mL) at –78 °C. Stirring was continued for 0.5 h before the mixture was allowed to warm to –30 °C over a period of 10 min. Stirring at –30 ± 3 °C (internal temp.) was continued for 0.5 h before the solution was then cooled to –78 °C at which time aldehyde **11** (2.28 g, 8.38 mmol, 1.40 equiv) in THF (4 mL) was added dropwise over 15 min. The cooling bath was removed and stirring was continued until the internal temperature reached 0 °C at which time stirring was continued for 1 h. Saturated aqueous NH_4Cl (15 mL) was added dropwise, the layers were separated and the aqueous layer was extracted with EtOAc (3 x 20 mL). The combined organic layers were washed with H_2O (20 mL) and brine (2 x 20 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure to afford an orange oil. Purification by flash column chromatography (hexanes:EtOAc 30:1 to 20:1) afforded benzyl alcohol **12** (2.80 g, 83%) as a colorless oil. R_f = 0.35 (silica gel, hexanes:EtOAc 10:1); IR (film) ν_{max} 3568, 2924, 2862, 1610, 1583, 1461, 1420, 1225, 1120, 1104, 882, 678 cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.37 (s, 2 H), 5.06 (br d, *J* = 7.2 Hz, OH), 3.81 (s, 6 H), 3.68–3.62 (m, 3 H), 2.59–2.50 (m, 2 H), 1.93–1.81 (m, 1 H), 1.72–1.64 (m, 1 H), 1.59–1.50 (m, 4 H + H_2O), 1.40–1.21 (m, 20 H), 1.06–1.04 (m, 21 H), 0.88 (t, *J* = 6.8 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 157.6, 143.6, 117.7, 104.6, 68.0, 63.7, 55.8, 37.9, 36.7, 33.2, 32.1, 31.6, 29.8, 29.8, 29.8, 29.7, 29.5, 29.5, 26.2, 26.0, 22.8, 18.2, 14.3, 12.2; HRMS (ESI): calculated for $C_{34}H_{64}O_4SiNa^+$ [M + Na^+] 587.4466, found 587.4470. Note: It

is important to measure and follow the indicated temperature during the lithiation process to obtain best results and minimize the amount of protodeiodinated side product formation.



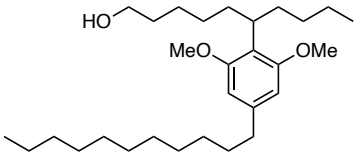
1-(2,6-Dimethoxy-4-undecylphenyl)-6-((triisopropylsilyl)oxy)hexan-1-one (**13**)

13. TEMPO (110 mg, 0.708 mmol, 0.250 equiv) and (diacetoxyiodo)benzene (1.09 g, 3.40 mmol, 1.20 equiv) were added in one portion to a stirring solution of alcohol **12** (1.60 g, 2.83 mmol, 1.00 equiv) in CH_2Cl_2 (22 mL) at RT. The resulting mixture was allowed to stir in the dark for 12 h before additional TEMPO (110 mg, 0.708 mmol, 0.250 equiv) and (diacetoxyiodo)benzene (734 mg, 2.28 mmol, 0.800 equiv) were added. Stirring was continued for 8 h before the reaction was quenched with a mixture of saturated aqueous $NaHCO_3$:NaS_2O_3:H_2O (1:1:1, 30 mL); stirring was continued for 10 min. The layers were separated and the aqueous layer was extracted with CH_2Cl_2 (3 x 15 mL). The combined organic layers were washed with H_2O (20 mL) and brine (2 x 20 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure affording an orange solid. Purification by flash column chromatography (hexanes:EtOAc 40:1 to 20:1) afforded the desired ketone **13** (1.30 g, 82%) as a colorless oil. R_f = 0.40 (silica gel, hexanes:EtOAc 10:1); IR (film) ν_{max} 2925, 2863, 1706, 1607, 1581, 1461, 1415, 1230, 1130, 1107, 882, 680 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.35 (s, 2 H), 3.76 (s, 6 H), 3.67 (t, *J* = 6.6 Hz, 2 H), 2.74 (t, *J* = 7.4 Hz, 2 H), 2.60–2.53 (m, 2 H), 1.73–1.55 (m, 6 H), 1.44–1.23 (m, 18 H), 1.07–1.04 (m, 21 H), 0.88 (t, *J* = 6.7 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 205.7, 156.8, 146.3, 118.4, 104.3, 63.5, 55.9, 45.0, 36.9, 33.0, 32.1, 31.5, 29.8, 29.8, 29.8, 29.7, 29.5, 29.5, 25.6, 23.6, 22.8, 18.2, 14.3, 12.2; HRMS (ESI): calculated for $C_{34}H_{63}O_4Si^+$ [M + H^+] 563.4490, found 563.4495.



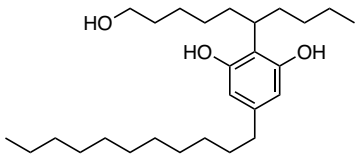
5-(2,6-Dimethoxy-4-undecylphenyl)-10-((triisopropylsilyl)oxy)decan-5-ol (**14**)

14. A flask was charged with $CeCl_3$ (anhydrous, beads, 1 mesh; 1.22 g, 4.97 mmol, 2.00 equiv). THF (10 mL) was added in one portion followed by immediate cooling to –78 °C. *n*-BuLi (2.5 M in hexanes; 1.94 mL, 4.85 mmol, 1.95 equiv) was added dropwise over 10 min and stirring at the indicated temperature was continued for 0.5 h before ketone **13** (1.40 g, 2.49 mmol, 1.00 equiv) in THF (2.5 mL) was added dropwise over 5 min. Stirring at –78 °C was continued for 2 h before the cooling bath was removed and stirring at RT was continued for another 18 h. The reaction was quenched with saturated aqueous $NaHCO_3$ (5 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with H_2O (20 mL) and brine (2 x 20 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:MTBE:$CHCl_3$ 25:1:1 + 0.1% Et_3N) afforded the desired alcohol **14** (1.15 g, 75%) as a colorless oil. R_f = 0.20 (silica gel, hexanes:MTBE:$CHCl_3$ 25:1:1); IR (film) ν_{max} 2923, 2855, 1606, 1574, 1455, 1414, 1232, 1127, 1110, 882, 680 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.35 (s, 2 H), 3.76 (s, 6 H), 3.67 (t, *J* = 6.6 Hz, 2 H), 2.79–2.66 (m, 2 H), 2.61–2.53 (m, 2 H), 1.72–1.64 (m, 2 H), 1.62–1.52 (m, 6 H), 1.44–1.23 (m, 20 H + OH), 1.06–1.03 (m, 26 H), 0.91–0.85 (m, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 158.2, 142.7, 119.0, 106.4, 80.2, 63.7, 56.4, 42.5, 42.2, 36.2, 33.3, 32.1, 31.3, 29.8, 29.8, 29.8, 29.7, 29.6, 29.5, 26.6, 26.6, 24.3, 23.4, 22.8, 18.2, 14.3, 14.3, 12.2; HRMS (ESI): calculated for $C_{38}H_{72}O_4SiNa^+$ [M + Na^+] 643.5092, found 643.5088. Note: It was observed that the more powdered the $CeCl_3$ the less conversion was obtained. Therefore, $CeCl_3$ beads gave the best results. Also, the desired product, alcohol **14**, is not stable under acidic conditions.



6-(2,6-Dimethoxy-4-undecylphenyl)decan-1-ol (**15**)

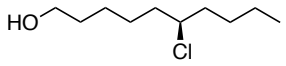
15. A flask was charged with alcohol **14** (400 mg, 0.644 mmol, 1.00 equiv) in CH_2Cl_2 (0.322 mL). The solution was cooled to 0 °C before Et_3SiH (205 μL, 150 mg, 1.29 mmol, 2.00 equiv) and TFA (424 μL, 631 mg, 6.44 mmol, 10.0 equiv) were added dropwise over 5 min. The flask was heated to 40 °C and held at this temperature for 36 h. The reaction mixture was allowed to cool to RT then diluted with EtOAc (50 mL) and washed with saturated aqueous $NaHCO_3$ (2 x 10 mL), H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 10:1) afforded the desired hydrodeoxygenated **15** (232 mg, 80%) as a colorless oil. R_f = 0.08 (silica gel, hexanes:EtOAc 10:1); IR (film) ν_{max} 3323, 2922, 2853, 1607, 1579, 1454, 1419, 1231, 1204, 1127, 1093, 825 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.34 (s, 2 H), 3.75 (s, 6 H), 3.57 (t, *J* = 6.6 Hz, 2 H), 3.27–3.17 (m, 1 H), 2.59–2.51 (m, 2 H), 1.86–1.70 (m, 2 H), 1.68–1.44 (m, 6 H) , 1.37–1.02 (m, 24 H + OH), 0.88 (t, *J* = 6.9 Hz, 3 H), 0.82 (t, *J* = 7.3 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 159.0, 141.7, 119.2, 104.8, 63.1, 56.1, 55.5, 36.6, 34.9, 33.7, 33.6, 32.8, 32.0, 31.5, 30.7, 29.8, 29.8, 29.8, 29.7, 29.5, 28.0, 25.8, 23.0, 22.8, 14.2, 14.2; HRMS (ESI): calculated for $C_{29}H_{53}O_3^+$ [M + H^+] 449.3989, found 449.3972.



2-(10-Hydroxydecan-5-yl)-5-undecylbenzene-1,3-diol ((±)-**8**)

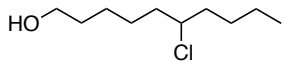
(±)-**8.** A flask was charged with dimethoxy benzene **15** (100 mg, 0.223 mmol, 1.00 equiv) in CH_2Cl_2 (1 mL). The solution was cooled to –78 °C before BBr_3 (1.0 M in CH_2Cl_2; 468 μL, 0.468 mmol, 2.10 equiv) was added dropwise over 10 min. The resulting mixture was stirred at –78 °C for 1 h before being allowed to warm to RT over 1 h. The resulting orange solution was cooled to 0 °C and quenched with saturated

aqueous $NaHCO_3$ (2 mL). The layers were separated and the aqueous layer was extracted with CH_2Cl_2 (2 x 5 mL). The combined organic layers were washed with H_2O (5 mL) and brine (5 mL), dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 3:1) afforded the desired resorcinol (±)-**8** (77.8 mg, 83%) as a colorless oil. R_f = 0.18 (silica gel, hexanes:EtOAc 3:1); IR (film) ν_{max} 3388, 2923, 2854, 1654, 1636, 1610, 1465, 1394, 1377, 1314, 1202, 1049, 1024, 897 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.15 (s, 2 H), 4.53 (br s, phenolic OH's), 3.59 (t, *J* = 6.6 Hz, 2 H), 3.14–2.95 (m, 1 H), 2.49–2.37 (m, 2 H), 1.93–1.78 (m, 2 H), 1.68–1.58 (m, 2 H), 1.58–1.44 (m, 4 H), 1.40–1.15 (m, 24 H + OH), 0.88 (t, *J* = 6.9 Hz, 3 H), 0.84 (t, *J* = 7.1 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 155.3, 142.1, 115.0, 108.6, 63.1, 35.5, 35.4, 33.9, 33.7, 32.5, 32.1, 31.1, 30.7, 29.8, 29.8, 29.8, 29.7, 29.6, 29.5, 27.8, 25.5, 23.0, 22.8, 14.3; HRMS (ESI): calculated for $C_{27}H_{49}O_3^+$ [M + H^+] 421.3676, found 421.3670.



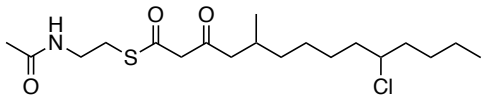
(*R*)-6-Chlorodecanol ((*R*)-**5**)

(*R*)-**5.** Prepared according to a known literature procedure[69] using enantiomerically pure (*R*)-5-(oxiran-2-yl)pentan-1-ol to set the stereocenter. All spectra matched data reported in the literature.[69] ^1H NMR (500 MHz, $CDCl_3$): δ 3.93–3.75 (m, 1 H), 3.64 (t, *J* = 6.6 Hz, 2 H), 1.81–1.42 (m, 14 H), 0.90 (t, *J* = 7.2 Hz, 3 H); ^{13}C NMR (125 MHz; $CDCl_3$): δ 64.2, 62.8, 38.4, 38.3, 32.6, 28.7, 26.3, 25.3, 22.3, 14.0.



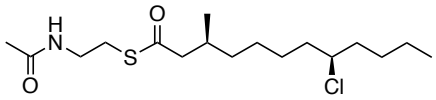
(±)-6-Chlorodecanol ((±)-**5**)

(±)-**5.** Prepared according to a known literature procedure using racemic starting materials.[69] All spectra matched data reported in the literature.[69] ^1H NMR (500 MHz, $CDCl_3$): δ 3.95–3.78 (m, 1 H), 3.64 (t, *J* = 6.6 Hz, 2 H), 1.81–1.42 (m, 14 H), 0.90 (t, *J* = 7.2 Hz, 3 H); ^{13}C NMR (125 MHz; $CDCl_3$): δ 64.2, 62.8, 38.4, 38.3, 32.6, 28.7, 26.3, 25.3, 22.3, 14.0.



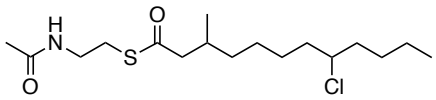
S-(2-Acetamidoethyl) 10-chloro-5-methyl-3-oxotetradecanethioate (**16a**)

16a. Prepared according to a known literature procedure.[69] All spectra matched data reported in the literature.[69] ^{1}H NMR (500 MHz, $CDCl_3$): Keto form: δ 5.92 (br s, 1 H, NH), 3.94–3.82 (m, 1 H), 3.72 (s, 2 H), 3.50–3.43 (m, 2 H), 3.12–3.07 (m, 2 H), 2.51 (dd, *J* = 16.6, 5.7 Hz, 1 H), 2.34 (dd, *J* = 16.6, 7.8 Hz, 1 H), 1.98 (s, 3 H), 1.98–1.88 (m, 1 H), 1.75–1.64 (m, 4 H) 1.56–1.15 (m, 10 H), 0.95–0.88 (m, 6 H); Enol form: δ 12.60 (s, 1 H), 5.92 (br s, 1 H, NH), 5.45 (s, 1 H), 3.92–3.83 (m, 1 H), 3.50–3.43 (m, 2 H), 3.12–3.07 (m, 2 H), 2.18 (dd, *J* = 7.8, 2.3 Hz, 1 H), 2.08–19.5 (m, 1 H), 1.98 (s, 3 H,), 1.97–1.88 (m, 1 H), 1.75–1.64 (m, 4 H) 1.56–1.15 (m, 10 H), 0.96–0.88 (m, 6 H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 202.2, 194.4, 192.5, 177.0, 170.6, 170.5, 100.6, 64.5, 57.9, 51.0, 51.0, 42.8, 42.7, 40.1, 39.5, 38.7, 38.5, 36.8, 31.3, 29.9, 29.5, 29.1, 28.9, 28.1, 26.8, 26.7, 23.5, 23.4, 22.5, 19.9, 19.8, 14.2.



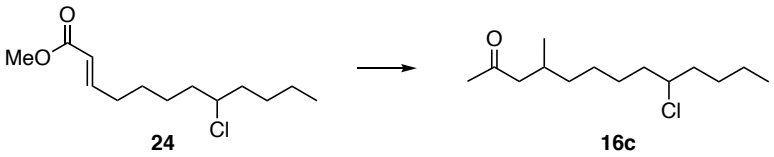
S-(2-Acetamidoethyl) (3*S*,8*R*)-8-chloro-3-methyldodecanethioate ((*S,R*)-**16b**))

(*S*,*R*)-**16b**. Prepared according to a known literature procedure[69] using enantiomerically pure (3*S*,8*R*)-8-chloro-3-methyldodecanoic acid as starting material. All spectra matched data reported in the literature.[69] ^{1}H NMR (500 MHz, $CDCl_3$): δ 5.90 (br s, 1 H, NH), 3.92–3.82 (m, 1 H), 3.42 (q, *J* = 6.2 Hz, 2 H), 3.02 (t, *J* = 6.4 Hz, 2 H), 2.55 (dd, *J* = 14.6, 6.0 Hz, 1 H), 2.38 (dd, *J* = 14.6, 8.1 Hz, 1 H), 2.08–1.92 (m, 1 H), 1.97 (s, 3 H), 1.74–1.64 (m, 4 H), 1.53–1.44 (m, 2 H), 1.42–1.17 (m, 8 H), 0.98–0.78 (m, 6 H); ^{13}C NMR (125 MHz; $CDCl_3$): δ 199.7, 170.2, 64.2, 51.3, 39.8, 38.4, 38.2, 36.4, 31.0, 28.6, 28.5, 26.5, 26.4, 23.2, 22.3, 19.5, 14.0.



S-(2-Acetamidoethyl) (±)-8-chloro-3-methyldodecanethioate ((±)-**16b**)

(±)-**16b.** Prepared according to a known literature procedure using racemic starting materials.[69] All spectra matched data reported in the literature.[69] ^{1}H NMR (500 MHz, $CDCl_3$): δ 5.90 (br s, 1 H, NH), 3.94–3.80 (m, 1 H), 3.42 (q, *J* = 6.2 Hz, 2 H), 3.02 (t, *J* = 6.4 Hz, 2 H), 2.55 (dd, *J* = 14.6, 6.0 Hz, 1 H), 2.38 (dd, *J* = 14.6, 8.1 Hz, 1 H), 2.06–1.90 (m, 1 H), 1.97 (s, 3 H), 1.74–1.64 (m, 4 H), 1.53–1.44 (m, 2 H), 1.42–1.17 (m, 8 H), 0.99–0.78 (m, 6 H); ^{13}C NMR (125 MHz; $CDCl_3$): δ 199.7, 170.2, 64.2, 51.3, 39.8, 38.4, 38.2, 36.4, 31.0, 28.6, 28.5, 26.5, 26.4, 23.2, 22.3, 19.5, 14.0.

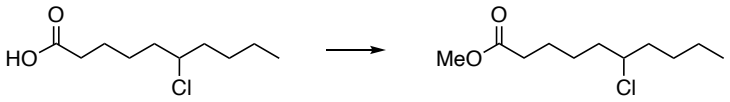


24 **16c**

9-Chloro-4-methyltridecan-2-one (**16c**)

16c. A flask containing a suspension of copper iodide (120 mg, 0.630 mmol, 1.00 equiv) in Et_2O (5.0 mL) was cooled to –20 ºC before MeMgBr (3.0 M in Et_2O; 1.05 mL, 3.15 mmol, 5.00 equiv) was added dropwise over 5 min. Stirring at –20 °C was continued for 1.5 h. Ester **24**, prepared as described previously,[69] (154 mg, 0.630 mmol, 1.00 equiv) in Et_2O (3.0 mL) was added dropwise over 5 min and the reaction mixture was allowed to stir at –20 °C for 2 h. The reaction was quenched with the slow addition of MeOH (1 mL) followed by saturated aqueous $NaHCO_3$ (30 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 20 mL). The combined organic layers were washed with H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure affording an orange oil. Purification by flash column chromatography (hexanes:Et_2O 15:1 to 10:1) afforded ketone **16c** (54 mg, 35%). R_f = 0.70 (silica gel, hexanes:EtOAc 9:1); ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.95–3.78 (m, 1 H), 2.39 (dd, *J* = 16.0, 5.7 Hz, 1 H), 2.23 (dd, *J* = 16.0, 7.9 Hz, 1 H), 2.12 (s, 3 H), 2.04–1.90 (m, 1 H), 1.76–1.62 (m, 4 H), 1.57–1.44 (m, 2 H), 1.44–1.11 (m, 8 H), 0.95–0.79 (m, 6 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 209.1, 64.4, 51.4, 38.6, 38.4, 36.9, 30.6, 29.3, 28.8, 26.7, 26.7, 22.4, 19.9, 14.1; HRMS (ESI): calculated for

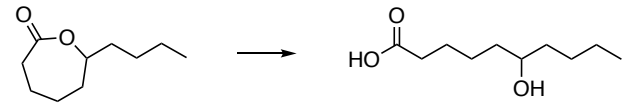
$C_{14}H_{27}ClNaO^+$ [M + Na^+] 269.1643, found 269.1633; calculated for $C_{14}H_{27}ClKO^+$ [M + K^+] 285.1382, found 285.1404.



25 **16d**

Methyl 6-chlorodecanoate (**16d**)

16d. (Trimethylsilyl)diazomethane (2.0 M in Et_2O; 0.120 mL, 0.240 mmol, 1.50 equiv) was added dropwise to a solution of 6-chlorodecanoic acid **25**. (33.0 mg, 0.160 mmol, 1.00 equiv) in a mixture of CH_2Cl_2:MeOH (2.0 mL; 1.2:0.8) and stirring at RT was continued for 0.5 h. The reaction was quenched with the addition of acetic acid (0.5 mL) and diluted with EtOAc (10 mL). The layers were separated and the organic layer was washed with saturated aqueous $NaHCO_3$ (2 x 10 mL) and brine (10 mL), dried over $NaSO_4$, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 20:1) afforded ester **16d** (28.0 mg, 79%). R_f = 0.63 (silica gel, hexanes:EtOAc 4:1); ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.93–3.80 (m, 1 H), 3.67 (s, 3 H), 2.32 (t, *J* = 7.3 Hz, 2 H), 1.78–1.23 (m, 12 H), 0.90 (t, *J* = 7.3 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 174.1, 64.0, 51.7, 38.4, 38.3, 34.1, 28.8, 26.2, 24.6, 22.4, 14.1; HRMS (ESI): calculated for $C_{11}H_{22}ClO^+$ [M + H^+], 221.1303; found 221.1323.

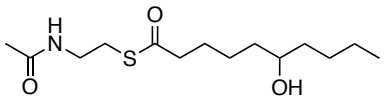


7-butyloxepan-2-one **26**

6-Hydroxydecanoic acid (**26**)

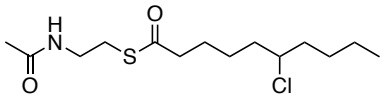
26. KOH (1.30 g, 23.5 mmol, 2.00 equiv) was added portion-wise to a solution of commercially available 7-butyloxepan-2-one (2.00 g, 11.8 mmol, 1.00 equiv) in H_2O (15 mL) at RT. The resulting mixture was stirred for 12 h before being quenched with the dropwise addition of aqueous HCl (1.0 M; 30 mL) over 10 min. The layers were separated and the aqueous layer was extracted with EtOAc (3 x 40 mL). The combined organic layers were washed with brine (3 x 40 mL), dried over Na_2SO_4, filtered and concentrated under reduced

pressure affording commercially available carboxylic acid **26** (2.20 g, quant.) as a white solid that was used in the next reaction without further purification.



S-(2-Acetamidoethyl) 6-hydroxydecanethioate (**27**)

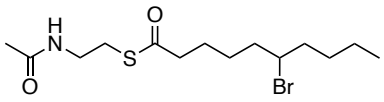
27. EDC•HCl (613 mg, 3.20 mmol, 2.00 equiv) was added portionwise to a solution of acid **26** (300 mg, 1.60 mmol, 1.00 equiv) in CH_2Cl_2 (30 mL) at 0 °C. Stirring at 0 °C was continued for 20 min, followed by the addition of *N*-acetylcysteamine (190 μL, 213 mg, 1.76 mmol, 1.10 equiv) and DMAP (226 mg, 1.86 mmol, 1.15 equiv). The resulting mixture was allowed to warm to RT and stirring was continued for 3 h. Upon completion, the reaction was quenched with the addition of H_2O (30 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 40 mL). The combined organic layers were washed with brine (3 x 40 mL), dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (EtOAc) afforded thioester **27** (340 mg, 74%) as a white solid. R_f = 0.18 (silica gel, EtOAc); IR (film) ν_{max} 3294, 3085, 2930, 2860, 1690, 1656, 1555, 1438, 1375, 1292, 1239, 1198, 1099, 1002, 735, 699, 602 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 5.85 (br s, 1 H, NH), 3.59 (s, 1 H), 3.44 (q, *J* = 15.0 Hz, 2 H), 3.02 (t, *J* = 15.0 Hz, 2 H), 2.59 (t, *J* = 15.0 Hz, 2 H), 1.96 (s, 3 H), 1.78–1.54 (m, 2 H), 1.65–1.22 (m, 11 H), 0.91 (t, *J* = 15.0 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 200.2, 170.4, 71.8, 44.1, 39.8, 37.5, 37.0, 28.7, 28.0, 25.8, 25.1, 23.4, 22.9, 14.2; HRMS (ESI): calculated for $C_{14}H_{28}NO_3S^+$ [M + H^+] 290.1784, found 290.1785.



S-(2-Acetamidoethyl) 6-chlorodecanethioate (**16e**)

16e. A flask was charged with PPh_3 (118 mg, 0.450 mmol, 1.30 equiv) in CH_2Cl_2 (3 mL). The resulting solution was cooled to 0 °C before *N*-chlorosuccinimide (60.0 mg, 0.450 mmol, 1.30 equiv) was added in portions. Stirring was continued for 15 min before alcohol **27** (100 mg, 0.350 mmol, 1.00 equiv) was added.

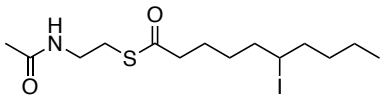
The resulting mixture was allowed to warm to RT and stirring was continued for 18 h. The reaction mixture was cooled to 0 °C and quenched with saturated aqueous $NaHCO_3$ (3 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 5 mL). The combined organic layers were washed with H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 1:1 to 1:2) afforded the desired chloride **16e** (80.8 mg, 75%) as a white solid. TLC: R_f = 0.34 (silica gel, hexanes:EtOAc 1:2); IR (film) ν_{max} 3279, 3078, 2934, 2862, 1689, 1653, 1550, 1457, 1434, 1373, 1360, 1288, 1237, 1185, 1095, 997, 752, 734, 698, 661, 631, 601, 496, 472, 419, 409 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 5.87 (br s, 1 H, NH), 3.90–3.78 (m, 1 H), 3.42 (q, *J* = 20.0 Hz, 2 H), 3.02 (t, *J* = 10.0 Hz, 2 H), 2.59 (t, *J* = 15.0 Hz, 2 H), 1.96 (s, 3 H), 1.78–1.28 (m, 12 H), 0.90 (t, *J* = 15.0 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 199.9, 170.4, 63.9, 44.0, 39.8, 38.4, 38.2, 28.8, 28.7, 25.9, 25.2, 23.3, 22.4, 14.1; HRMS (ESI): calculated for $C_{14}H_{27}ClNO_2S^+$ [M + H^+] 308.1446, found 308.1443.



S-(2-Acetamidoethyl) 6-bromodecanethioate (**16f**)

16f. A flask was charged with alcohol **27** (50.0 mg, 0.173 mmol, 1.00 equiv) and Et_3N (33.7 μL, 24.5 mg, 0.242 mmol, 1.40 equiv) in CH_2Cl_2 (1 mL). The resulting solution was cooled to –10 °C in a salt and ice bath before MsCl (14.5 μL, 21.4 mg, 0.187 mmol, 1.08 equiv) was added dropwise over 3 min. Stirring was continued for 1 h at –10 °C and 1 h at RT before the reaction was quenched with saturated aqueous NH_4Cl (2 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 5 mL). The combined organic layers were washed with H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure affording the desired mesylate (60 mg, 95%) as a clear colorless oil. ^1H NMR (500 MHz, $CDCl_3$) δ 5.89 (br s, 1 H, NH), 4.78–4.65 (m, 1 H), 3.48–3.37 (m, 2 H), 3.03 (t, *J* = 6.4 Hz, 2 H), 3.00 (s, 3 H), 2.64–2.56 (m, 2 H), 1.97 (d, *J* = 1.0 Hz, 3 H), 1.79–1.28 (m, 12 H), 0.92 (t, *J* = 4.7 Hz, 3 H). The crude material was immediately used in the next transformation without further purification.

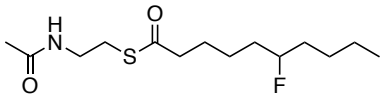
A flask was charged with the previously obtained mesylate (60.0 mg, 0.163 mmol, 1.00 equiv) in THF (3.5 mL). LiBr (70.8 mg, 0.815 mmol, 5.00 equiv) was added in one portion and the resulting mixture was heated to 60 °C for 12 h. The reaction was allowed to cool to RT and quenched with H_2O (5 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 5 mL). The combined organic layers were washed with H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 1:2) afforded the desired bromide **16f** (43 mg,75%) as a white solid. TLC: R_f = 0.30 (silica gel, hexanes:EtOAc 1:2); IR (film) ν_{max} 3280, 3076, 2954, 2931, 2860, 1690, 1654, 1549, 1457, 1433, 1374, 1361, 1290, 1236, 1190, 997; ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.78 (br s, 1 H, NH), 4.07–3.92 (m, 1 H), 3.44 (q, *J* = 6.2 Hz, 2 H), 3.03 (t, *J* = 6.4 Hz, 2 H), 2.60 (t, *J* = 7.4 Hz, 2 H), 1.97 (s, 3 H), 1.82–1.33 (m, 12 H), 0.91 (t, *J* = 7.2 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 200.0, 170.4, 58.3, 44.0, 39.9, 39.1, 38.8, 29.9, 28.7, 27.1, 25.2, 23.4, 22.3, 14.1; HRMS (ESI): calculated for $C_{14}H_{27}BrNO_2S^+$ [M + H^+] 352.0940, 354.0920; found 352.0937, 354.0918.



S-(2-Acetamidoethyl) 6-iododecanethioate (**16g**)

16g. A flask was charged with PPh_3 (63.4 mg, 0.242 mmol, 1.40 equiv) and imidazole (16.5 mg, 0.242 mmol, 1.40 equiv) in CH_2Cl_2 (3.0 mL). The resulting solution was cooled to 0 °C before iodine (61.4 mg, 0.242 mmol, 1.40 equiv) was added in portions. Stirring was continued for 15 min before the alcohol **27** (50.0 mg, 0.173 mmol, 1.00 equiv) in CH_2Cl_2 (1.0 mL) was added dropwise over 5 min. The resulting mixture was allowed to warm to RT and stirring was continued for 18h. The reaction mixture was cooled to 0 °C and quenched with saturated aqueous $NaHCO_3$ (2 mL). The layers were separated and the aqueous layer was extracted with EtOAc (3 x 5 mL). The combined organic layers were washed with H_2O (10 mL) and brine (2 x 10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (EtOAc) afforded the desired iodide **16g** (45 mg, 65%) as a white solid. TLC: R_f = 0.30 (silica gel, EtOAc); IR (film) ν_{max} 3272, 3077, 2962, 2928, 2858, 1677, 1654, 1548, 1458, 1372, 1286, 1180, 1119, 997, 721, 694, 540; ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.83 (br s, 1 H, NH),

4.14–4.03 (m, 1 H), 3.43 (q, J = 6.3 Hz, 2 H), 3.03 (t, J = 6.3 Hz, 2 H), 2.60 (t, J = 7.4 Hz, 2 H), 1.96 (s, 3 H), 1.91–1.80 (m, 2 H), 1.75–1.25 (m, 10 H), 0.91 (t, J = 7.2 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 199.9, 170.3, 44.0, 40.6, 40.3, 39.8, 39.7, 31.8, 29.0, 28.7, 24.9, 23.4, 22.1, 14.1; HRMS (ESI): calculated for $C_{14}H_{26}INO_2SNa^+$ [M + Na^+], 422.0621; found 422.0627.



S-(2-Acetamidoethyl) 6-fluorodecanethioate (**16h**)

16h. Diethylaminosulfur trifluoride (DAST) (45.6 μL, 55.7 mg, 0.346 mmol, 2.00 equiv) was added dropwise to a solution of alcohol **27** (50.0 mg, 0.173 mmol, 1.00 equiv) in CH_2Cl_2 (1.1 mL) at –78 °C. The resulting mixture was stirred for 15 min before the cooling bath was removed and stirring was continued for 1 h. The reaction mixture was diluted with EtOAc (20 mL), cooled to 0 °C and quenched with saturated aqueous $NaHCO_3$ (5mL). The layers were separated and the aqueous layer was extracted with EtOAc (2 x 5 mL). The combined organic layers were washed with H_2O (2 x 5 mL), dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography (EtOAc) afforded fluoride **16h** (21 mg, 42%) as a white solid. TLC: R_f = 0.40 (EtOAc); IR (film) ν_{max} 3293, 3079, 2952, 2929, 2870, 2860, 1693, 1641, 1552, 1465, 1410, 1373, 1359, 1294, 1138, 948, 844, 730, 678, 607; ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.79 (br s, 1 H, NH), 4.56–4.34 (m, 1 H), 3.44 (q, J = 6.0 Hz, 2 H), 3.11–2.96 (m, 2 H), 2.59 (t, J = 7.5 Hz, 2 H), 1.96 (s, 3 H), 1.74–1.25 (m, 12 H), 0.91 (t, J = 7.1 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 200.0, 170.4, 94.3 (d, $^1J_{CF}$ = 167.3 Hz), 44.0, 39.8, 35.0 (d, $^2J_{CF}$ = 20.9 Hz), 34.9 (d, $^2J_{CF}$ = 21.1 Hz), 28.6, 27.4 (d, $^3J_{CF}$ = 4.6 Hz), 25.6, 24.6 (d, $^3J_{CF}$ = 4.3 Hz), 23.3, 22.7, 14.1; HRMS (ESI): calculated for $C_{14}H_{27}FNO_2S^+$ [M + H^+], 292.1741; found 292.1717.

HRMS data fluoride **16h** purified using flash column chromatography showed traces of chloride species **16e** of which the origin is unknown (Figure 2.11). Using this fluoride sample in the biocatalytic CylK mediated Friedel–Crafts alkylation, some conversion was observed. To exclude the participation of the contaminating chloride species in the reaction, HPLC purification was pursued. Preparative HPLC

purification to remove all traces of chloride species **16e** was carried out with a Hypersil Gold aQ C18 column (20 x 250 mm, Thermo Scientific) on a Dionex UltiMate 3000 instrument (Thermo Scientific). The HPLC method was as follows: 500 μL injection volume (5.0 mg **16h** in 500 μL H_2O:acetonitrile 1:1), 8.0 mL x min^{-1} solvent flow rate, 25 °C column temperature, 20% solvent B in solvent A for 5 min, a gradient increasing to 100% solvent B over 20 min, 100% solvent B for 5 min with solvent A = H_2O + 0.1% formic acid and solvent B = acetonitrile + 0.1% formic acid. The desired product eluted with a retention time of 26.5 min affording fluoride **16h** free of chloride **16e** (Figure 2.11). Using the HPLC purified fluoride **16h**, no conversion in the CylK mediated alkylation was observed, indicating this compound is not accepted by the enzyme.

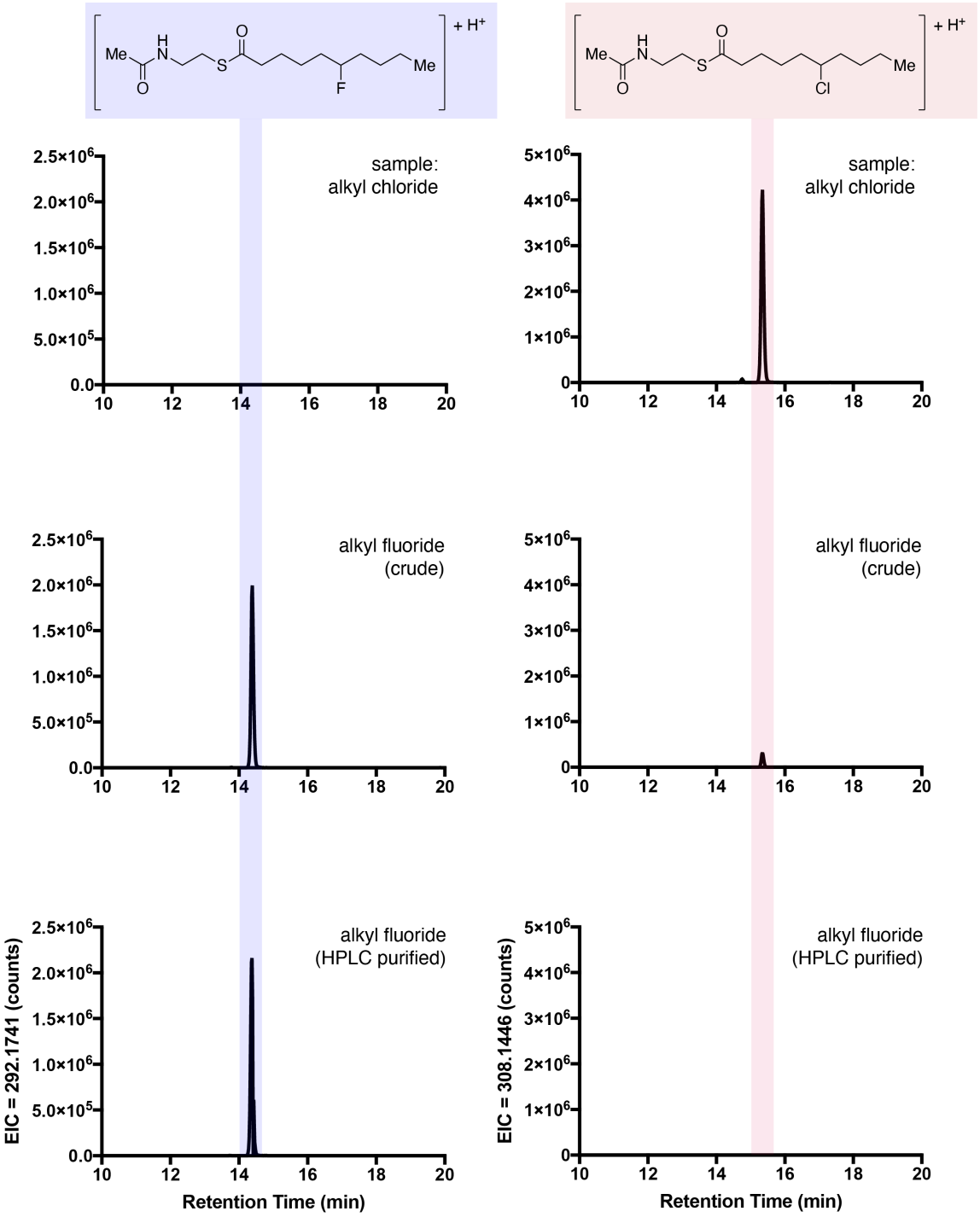
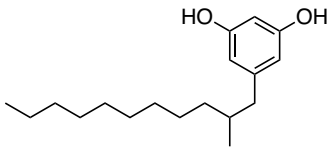


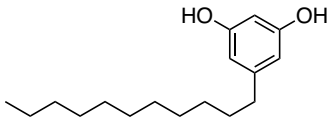
Figure 2.11. LC-MS analysis of alkyl fluoride and chloride electrophile substrates

Extracted ion chromatograms (EICs) of substrate stock solutions alkyl chloride **16e**, crude alkyl fluoride **16h** (purified by flash column chromatography), and HPLC purified alkyl fluoride **16h**. The origins of trace chloride species **16e** in crude alkyl fluoride **16h** are unknown. Pure alkyl fluoride **16h** was obtained by preparatory HPLC purification.



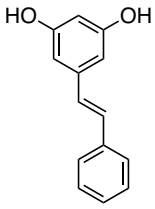
5-(2-Methylundecyl)benzene-1,3-diol (**4**)

4. Prepared according to a known literature procedure.[71] The spectral characteristics are identical to those reported. ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.27 (d, *J* = 2.3 Hz, 2 H), 6.22 (t, *J* = 2.1 Hz, 1 H), 6.08 (br s, 2 H, OH), 2.52 (dd, *J* = 13.3, 5.5 Hz, 1 H), 2.18 (dd, *J* = 13.3, 8.7 Hz, 1 H), 1.72–1.60 (m, 1 H), 1.55–1.21 (m, 15 H), 1.20–1.10 (m, 1 H), 0.91 (t, *J* = 6.9 Hz, 3 H), 0.81 (d, *J* = 6.4 Hz, 3 H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 156.2, 145.5, 109.4, 100.6, 43.8, 37.2, 34.9, 32.2, 30.2, 30.0, 29.9, 29.6, 27.4, 22.9, 19.5, 14.4.



5-Undecylbenzene-1,3-diol (**7**)

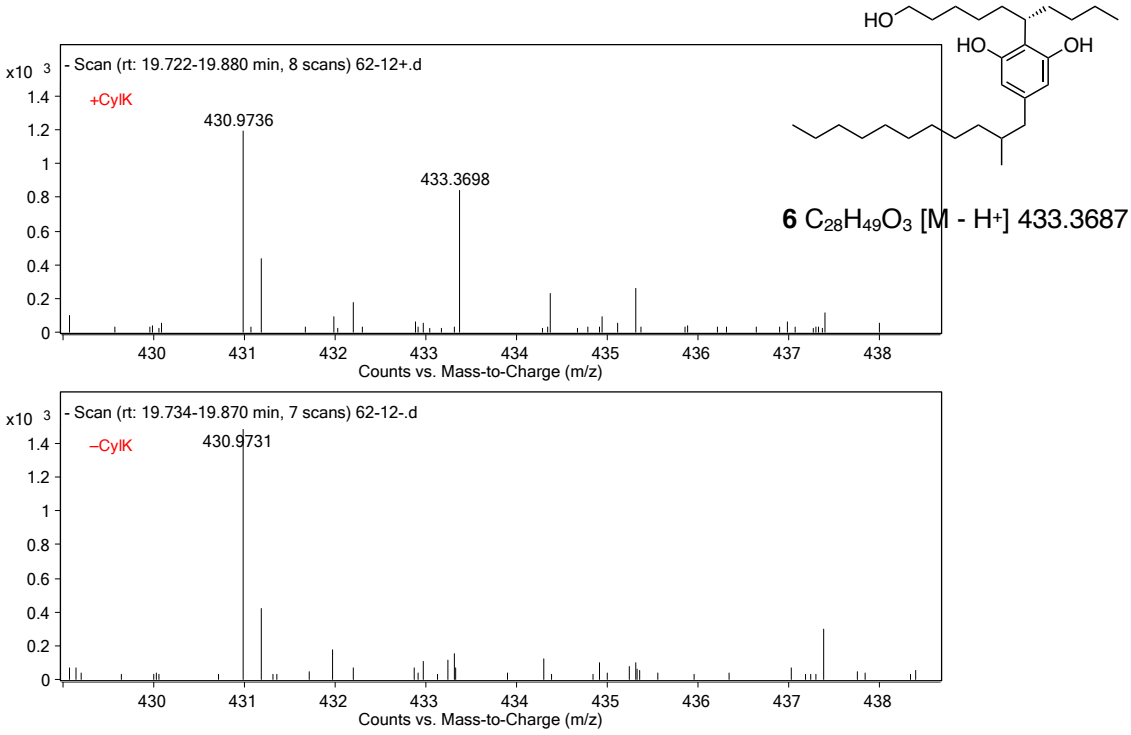
7. Boron tribromide (1.0 M in CH_2Cl_2; 4.72 mL, 4.72 mmol, 3.00 equiv) was added dropwise over 10 min to a solution of dimethoxybenzene **9** (460 mg, 1.57 mmol, 1.00 equiv) in CH_2Cl_2 (10 mL) at –78 °C. The resulting mixture was allowed to warm to RT after which stirring was continued for 14 h. Upon completion, the reaction was quenched with the slow addition of saturated aqueous $NaHCO_3$ (20 mL). The layers were separated, and the aqueous layer was extracted with Et_2O (3 x 20 mL). The combined organic layers were washed with brine (2 x 20 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. Purification by flash column chromatography (hexanes:EtOAc 2:1) afforded diol **7** (363 mg, 87% yield) as a white solid. R_f = 0.41 (silica gel, hexanes:EtOAc 1:1); IR (film) ν_{max} 3291, 2961, 2916, 2849, 1629, 1598, 1508, 1481, 1474, 1463, 1451, 1380, 1344, 1332, 1306, 1286, 1213, 1159, 1147, 998, 978, 836, 817, 728, 718, 696, 653, 626, 585, 523 cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.24 (s, 2 H), 6.17 (s, 1 H), 4.63–4.61 (m, 2 H), 2.48 (t, *J* = 15.0 Hz, 2 H), 1.60–1.54 (m, 2 H), 1.30–1.23 (m, 16 H), 0.88 (t, *J* = 10.0 Hz, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 156.7, 146.3, 108.2, 100.3, 36.0, 32.1, 31.2, 29.8, 29.8, 29.7, 29.7, 29.5, 29.4, 22.8, 14.3; HRMS (ESI) calculated for $C_{17}H_{27}O_2$ [M - H$^+$] 263.2017, found 263.2022.

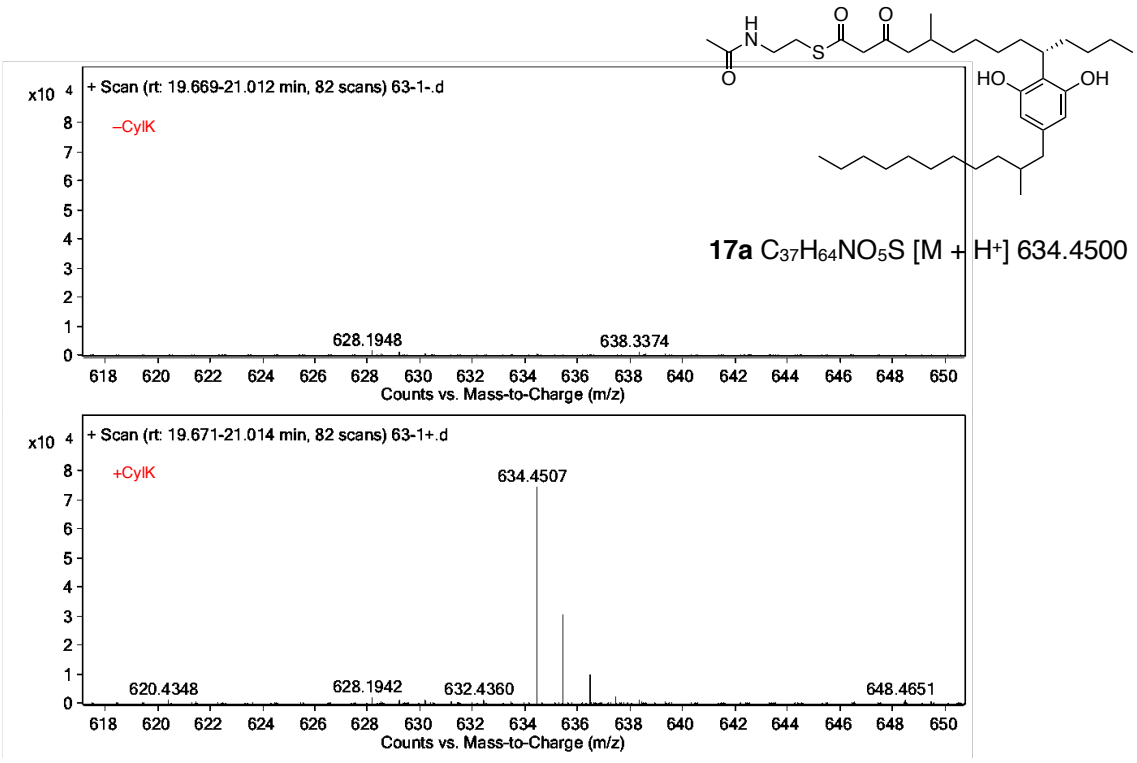


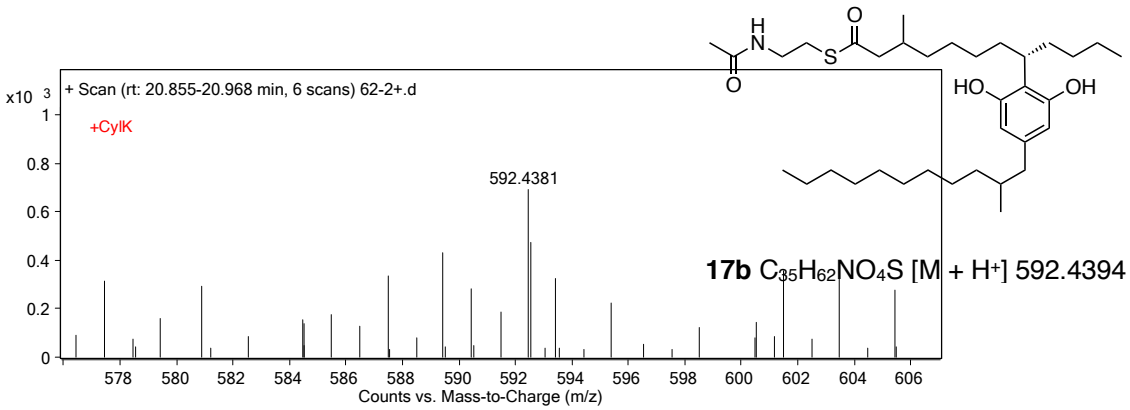
(*E*)-5-Styrylbenzene-1,3-diol (**20a**)

20a. Prepared according to a known literature procedure.[161] The spectral characteristics are identical to those reported.[161] ^{1}H NMR (500 MHz, acetone-d_6) δ 8.21 (s, 2 H), 7.54 (d, *J* = 7.8 Hz, 2 H), 7.32 (t, *J* = 7.5 Hz, 2 H), 7.23–7.20 (m, 1 H), 7.08 (d, *J* = 16.8 Hz, 1 H), 7.04 (d, *J* = 16.8 Hz, 1H), 6.56 (d, *J* = 2.1 Hz, 2 H), 6.28 (t, *J* = 2.2 Hz, 1 H).

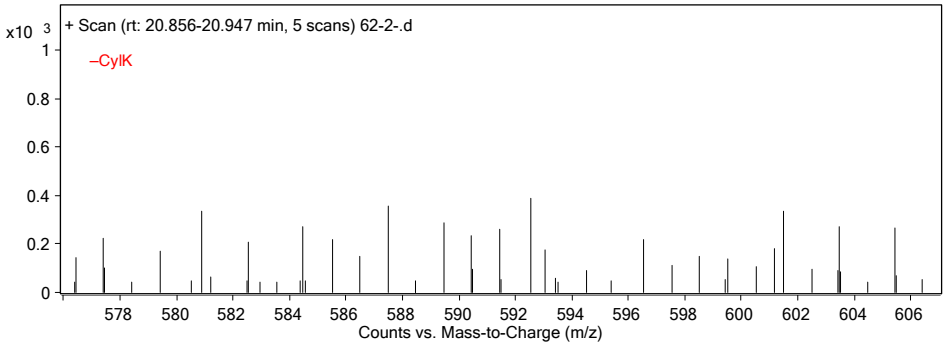
2.4.12 High Resolution Mass Spectra

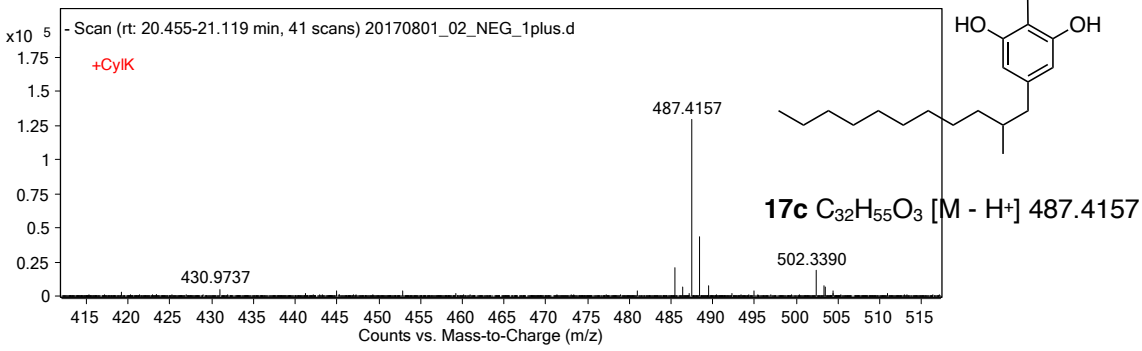




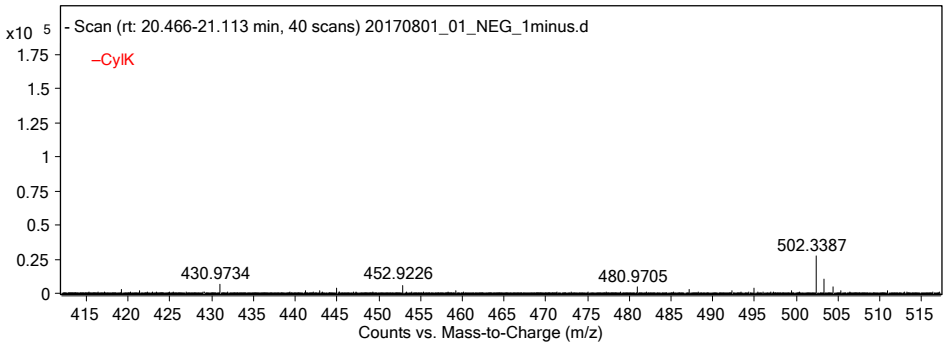


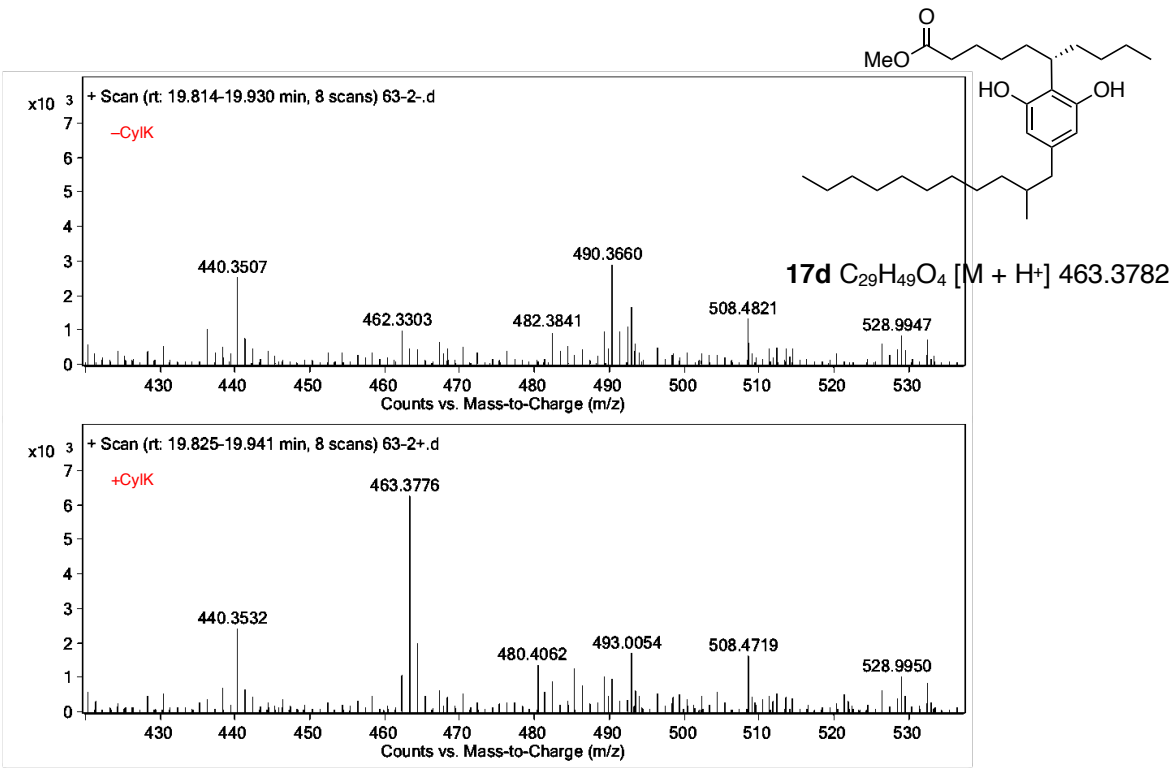
17b $C_{35}H_{62}NO_4S$ [M + H$^+$] 592.4394



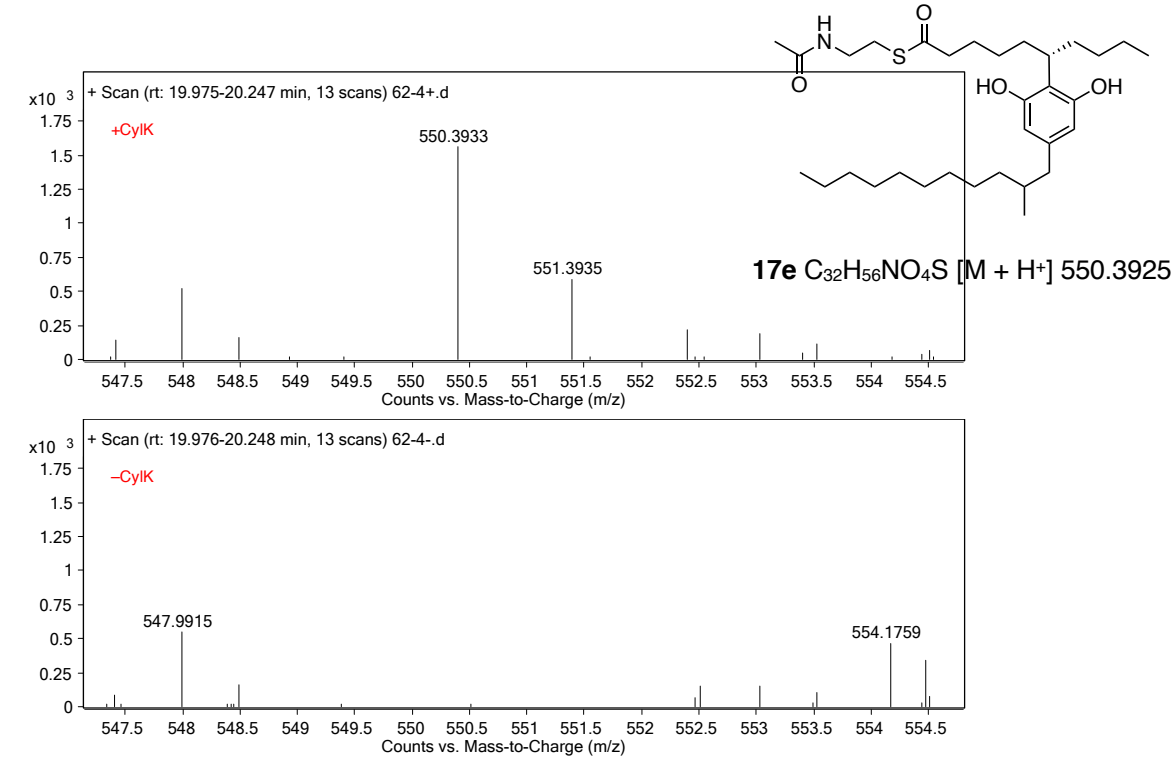


17c $C_{32}H_{55}O_3$ [M - H$^+$] 487.4157

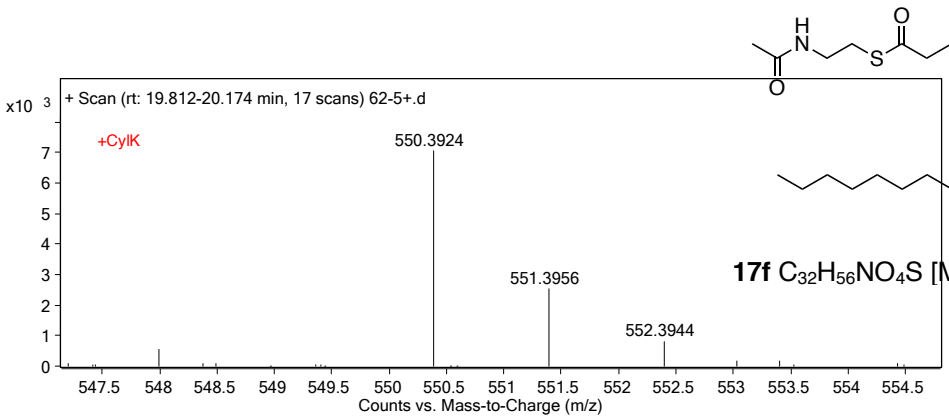


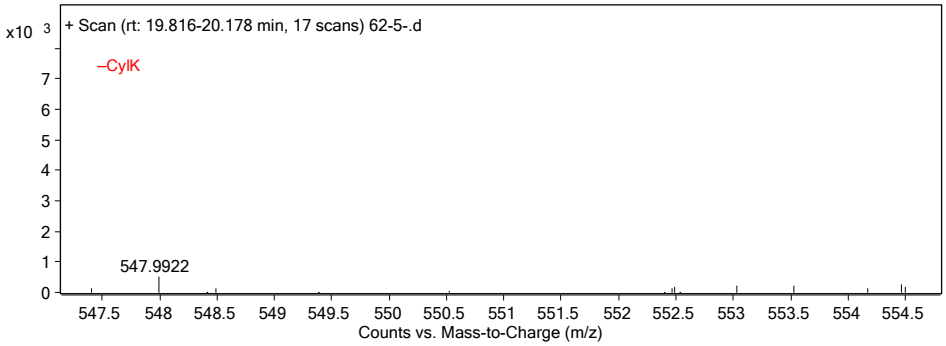


17d $C_{29}H_{49}O_4$ [M + H+] 463.3782

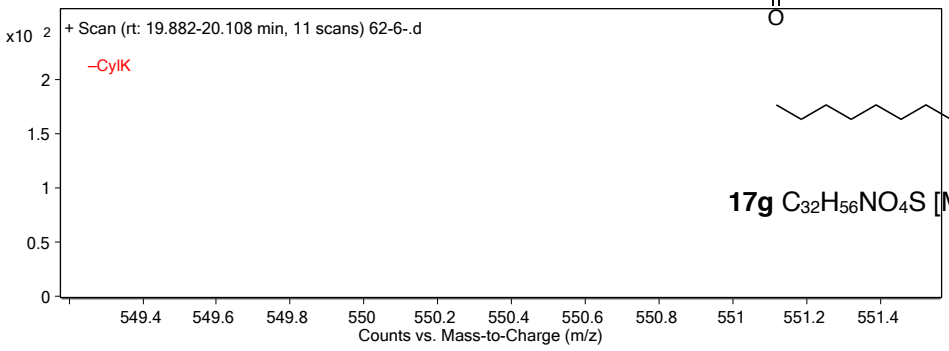


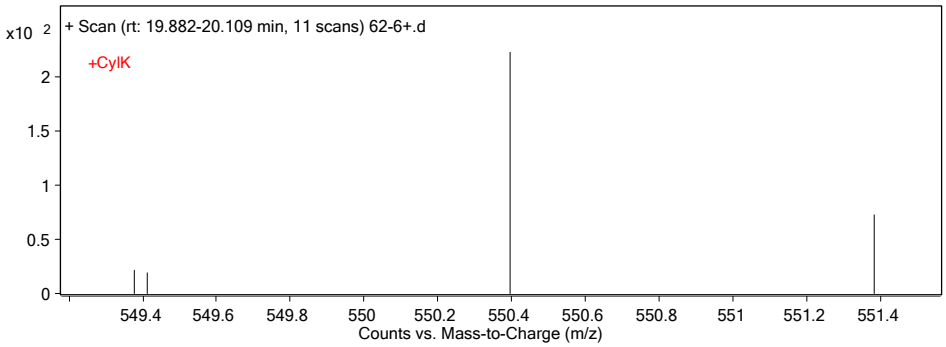
17e $C_{32}H_{56}NO_4S$ [M + H+] 550.3925



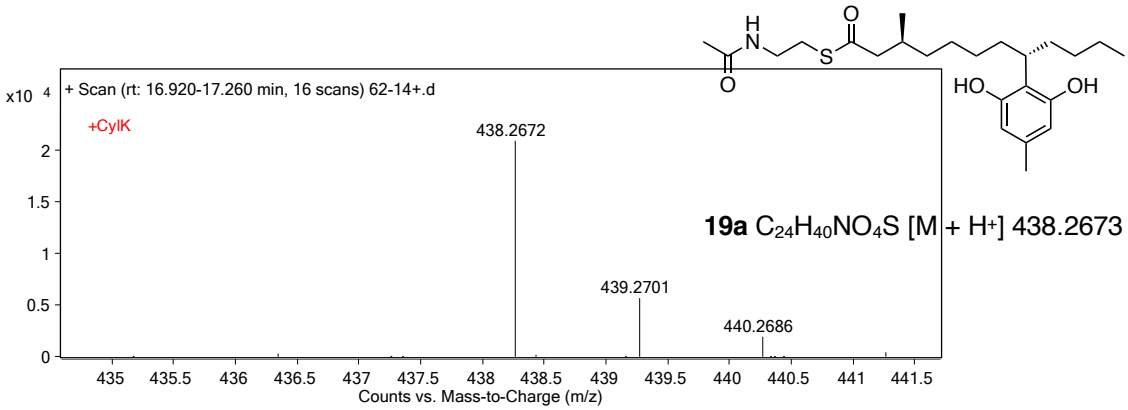


17f $C_{32}H_{56}NO_4S$ [M + H+] 550.3925

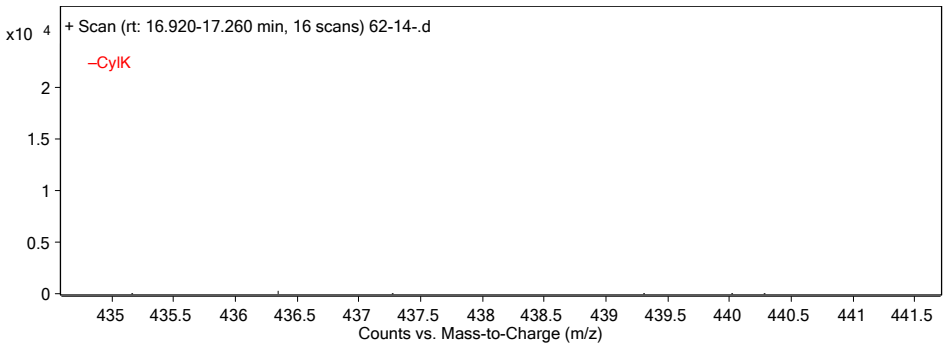


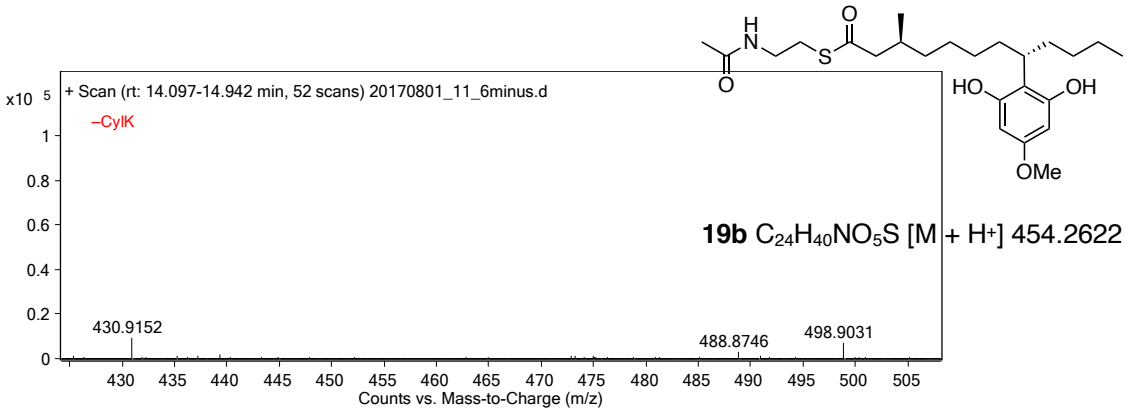


17g $C_{32}H_{56}NO_4S$ [M + H+] 550.3925

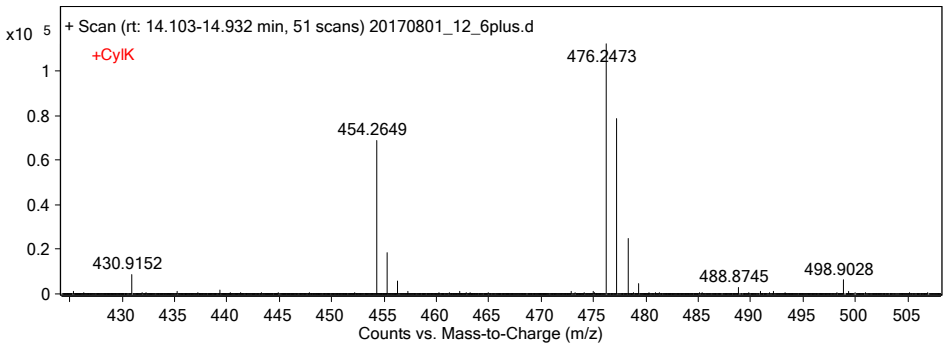


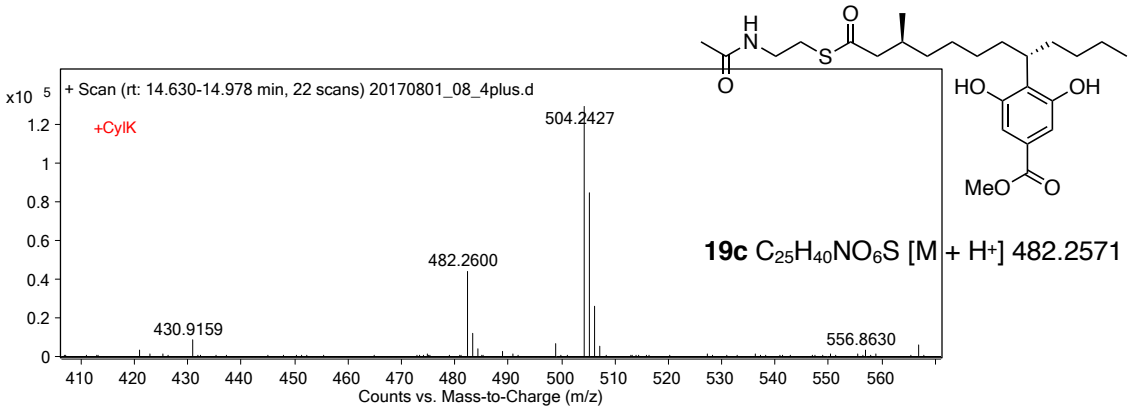
19a $C_{24}H_{40}NO_4S$ [M + H$^+$] 438.2673



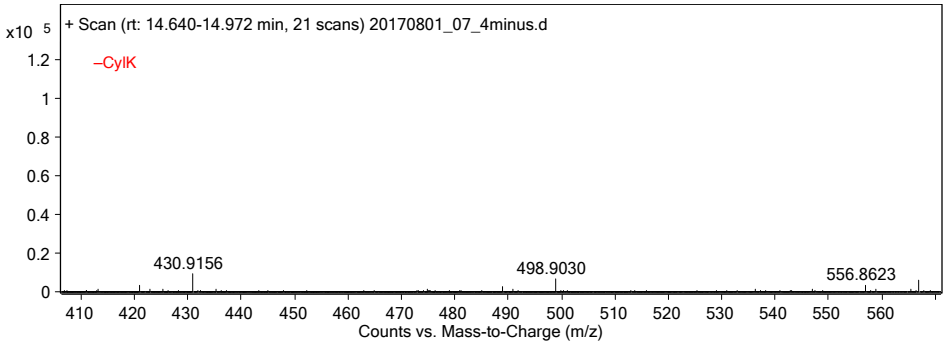


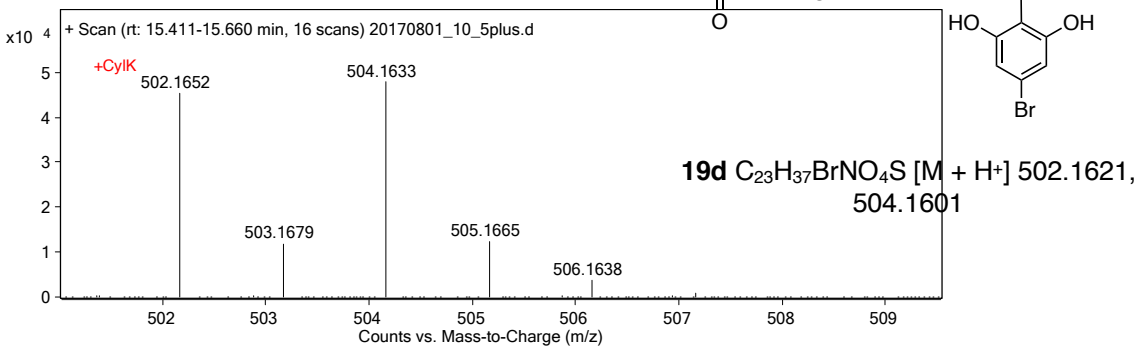
19b $C_{24}H_{40}NO_5S$ [M + H$^+$] 454.2622



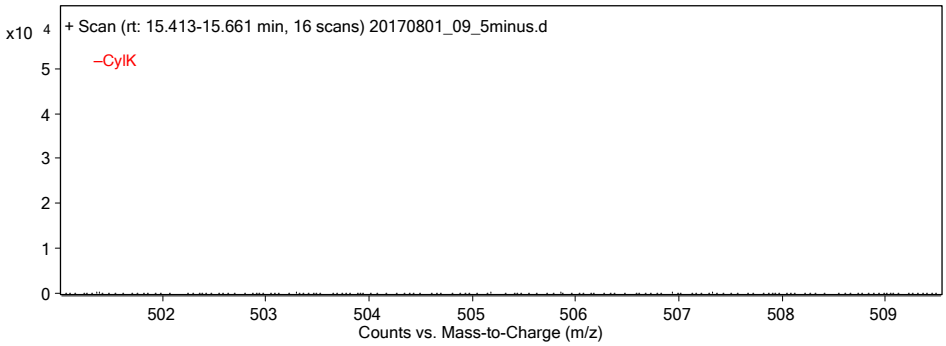


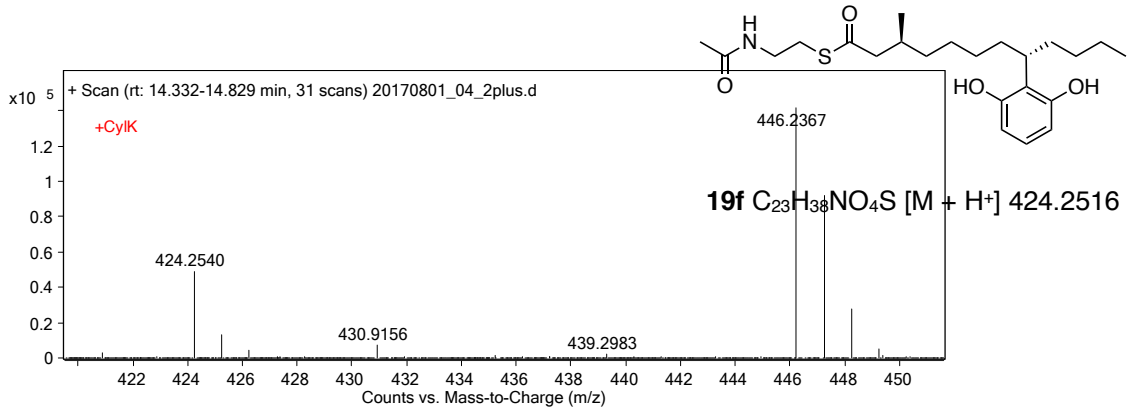
19c $C_{25}H_{40}NO_6S$ [M + H+] 482.2571



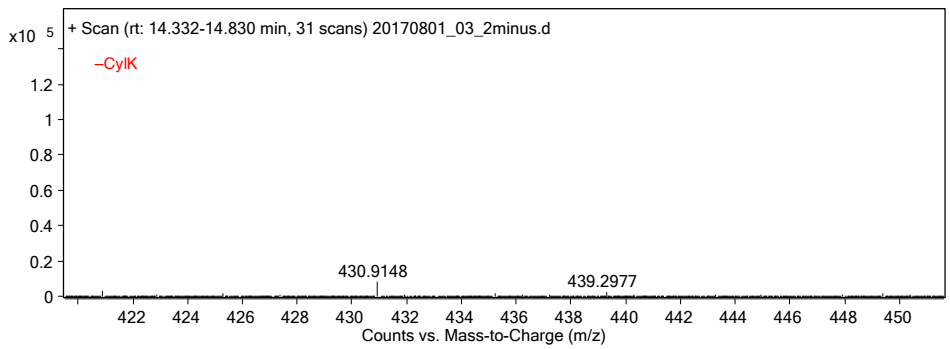


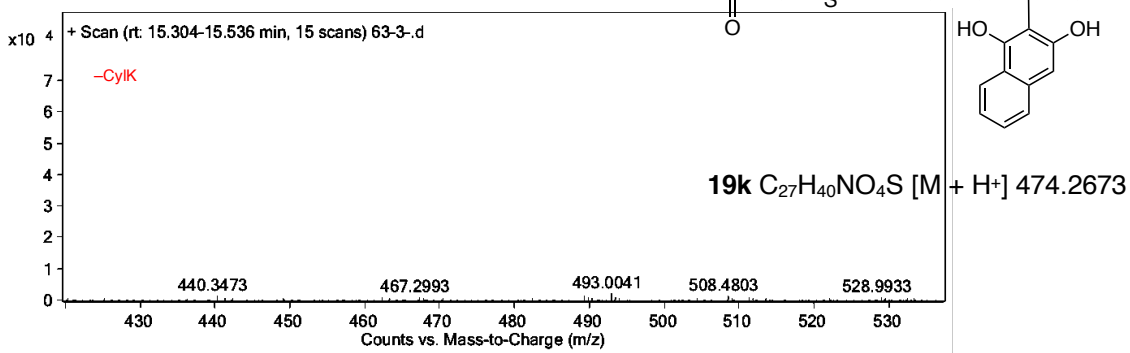
19d $C_{23}H_{37}BrNO_4S$ [M + H+] 502.1621, 504.1601



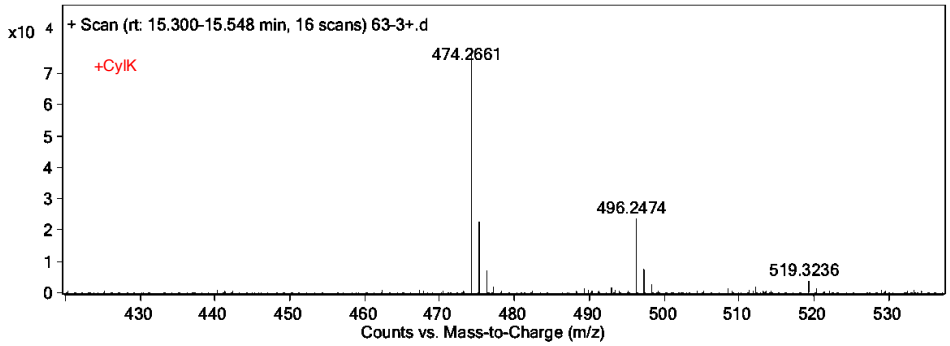


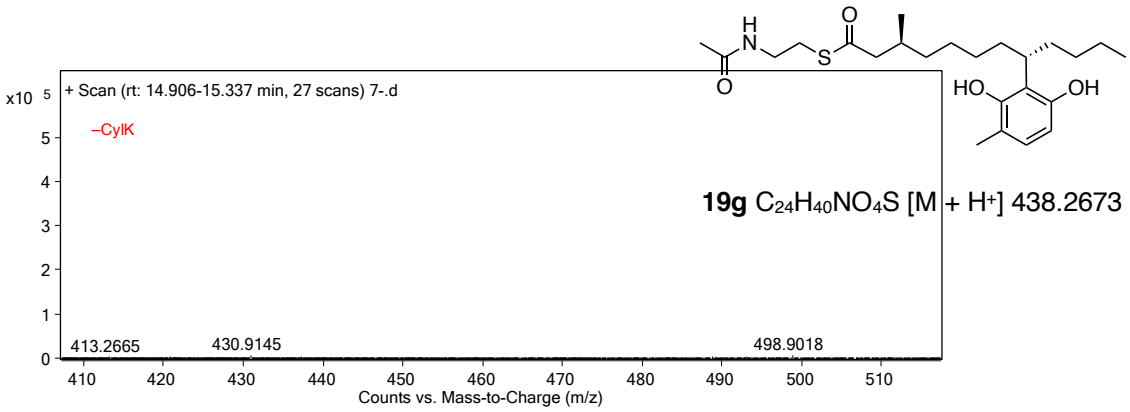
19f $C_{23}H_{38}NO_4S$ [M + H+] 424.2516



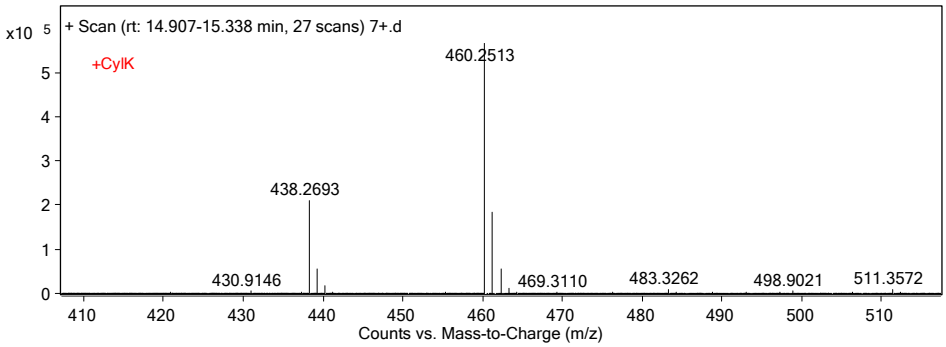


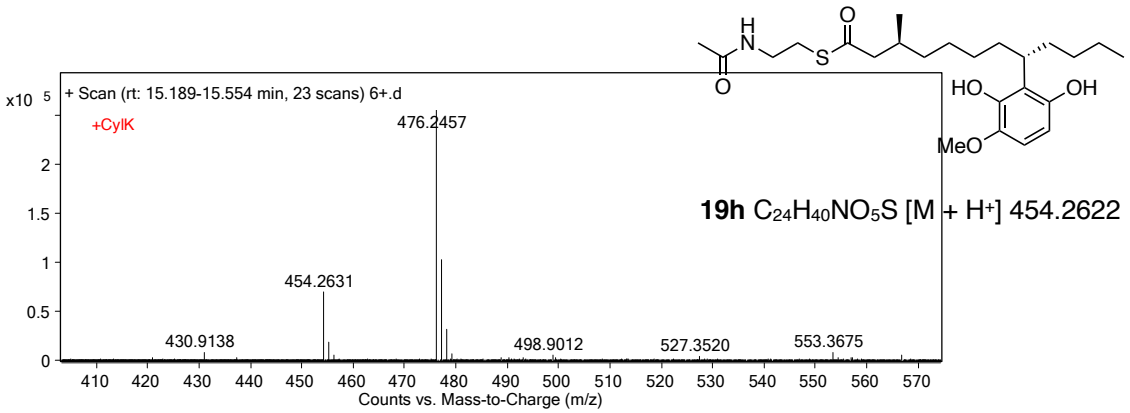
19k $C_{27}H_{40}NO_4S$ [M + H+] 474.2673



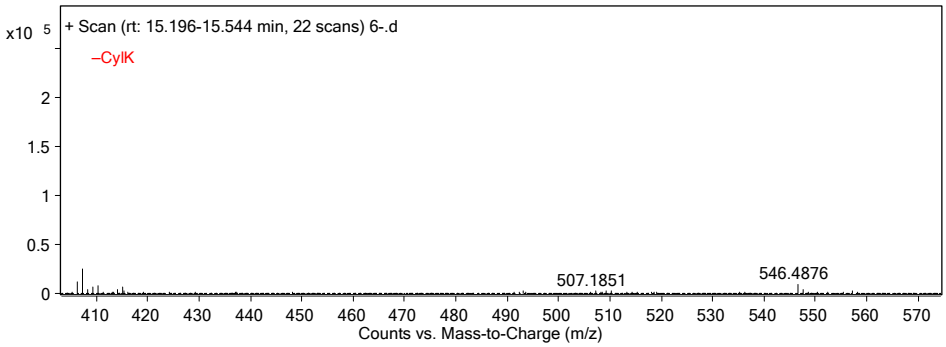


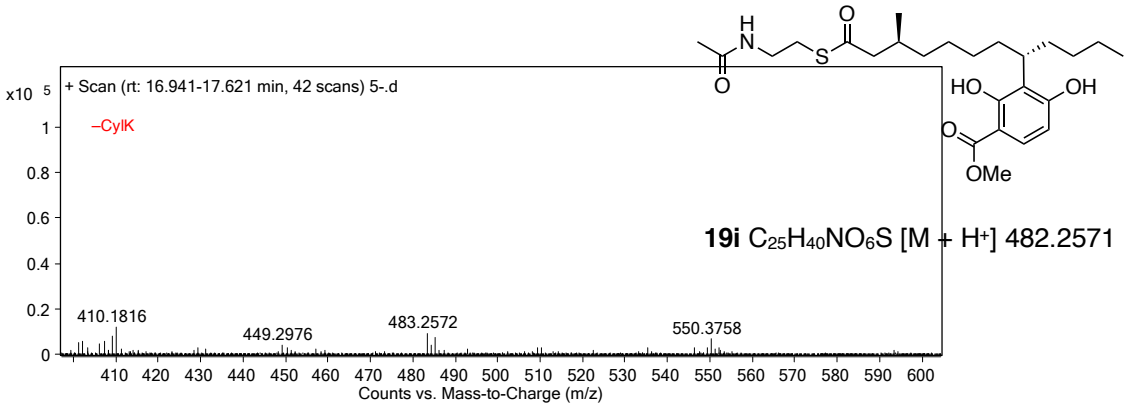
19g $C_{24}H_{40}NO_4S$ [M + H⁺] 438.2673

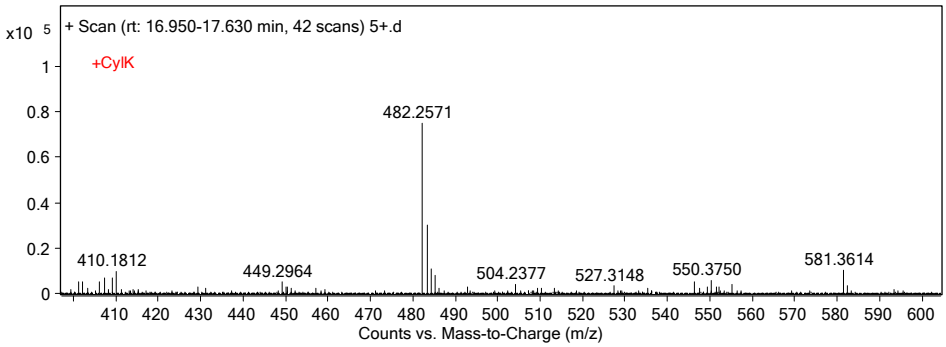




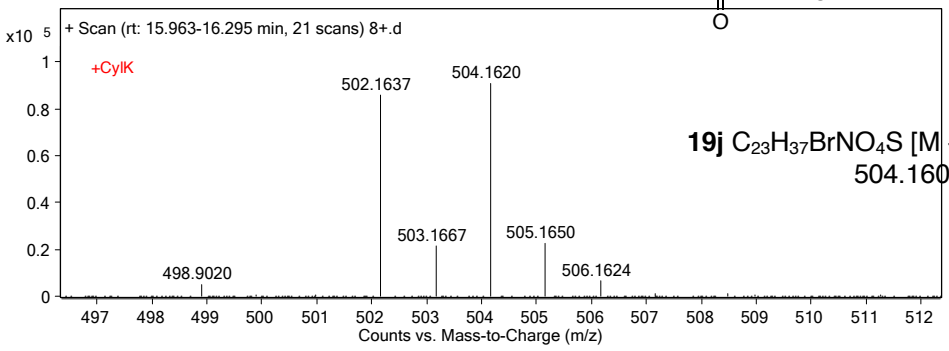
19h $C_{24}H_{40}NO_5S$ [M + H⁺] 454.2622

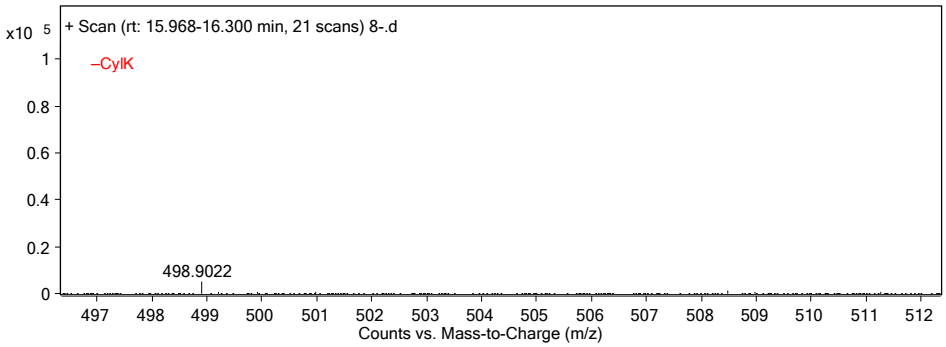




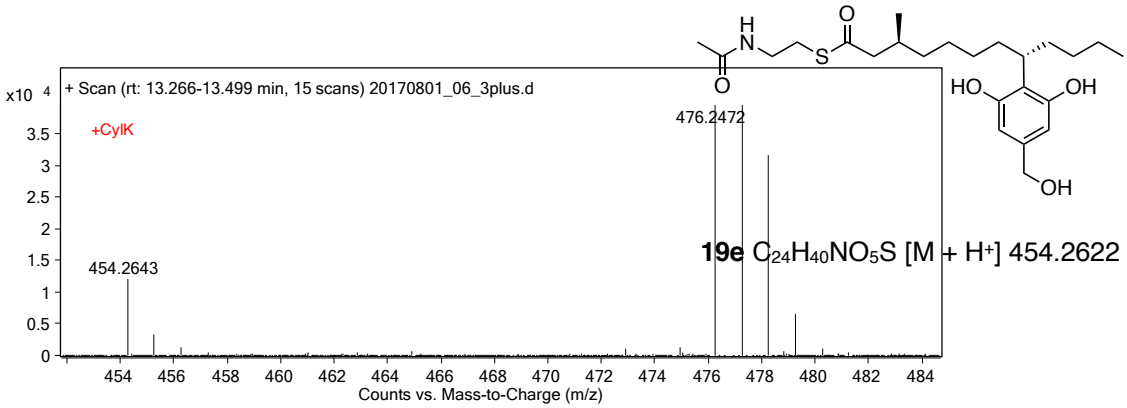


19i $C_{25}H_{40}NO_6S$ [M + H⁺] 482.2571

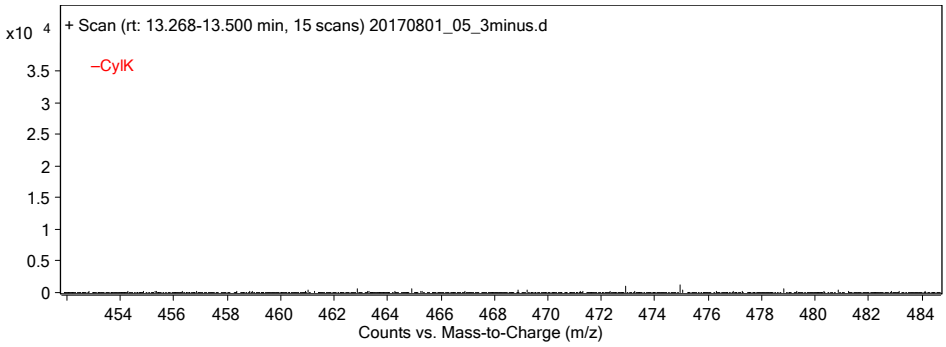


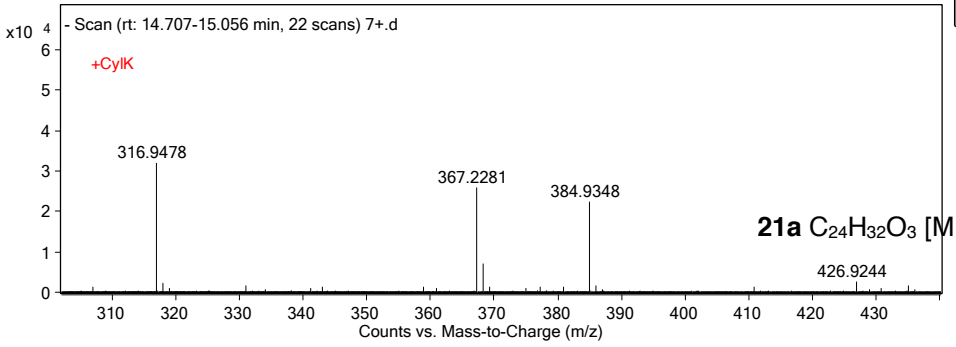


19j $C_{23}H_{37}BrNO_4S$ [M + H⁺] 502.1621, 504.1601

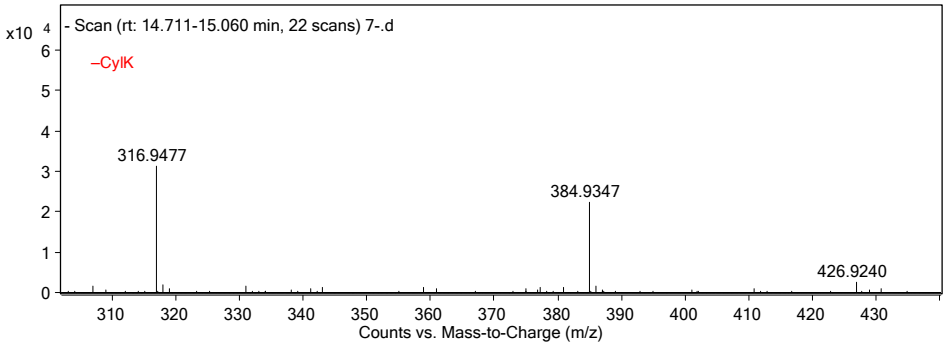


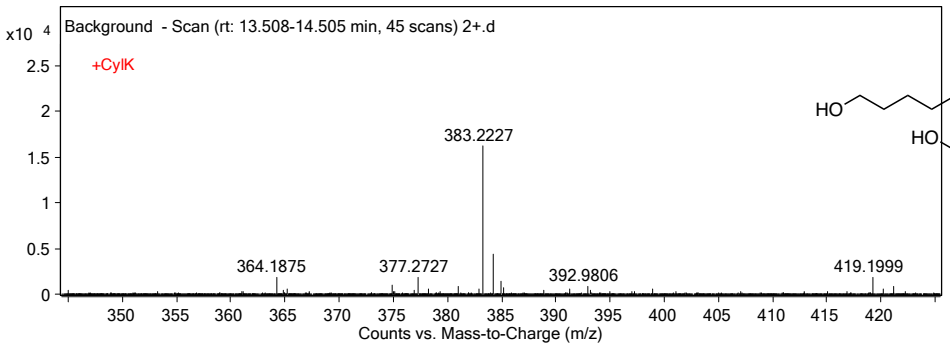
19e $C_{24}H_{40}NO_5S$ [M + H⁺] 454.2622

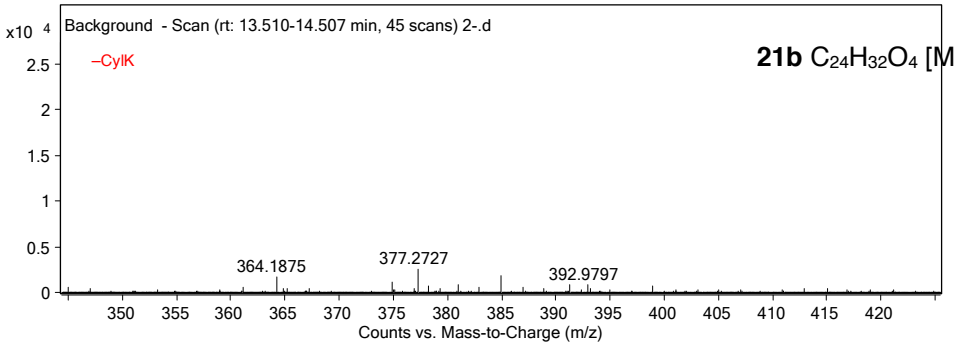




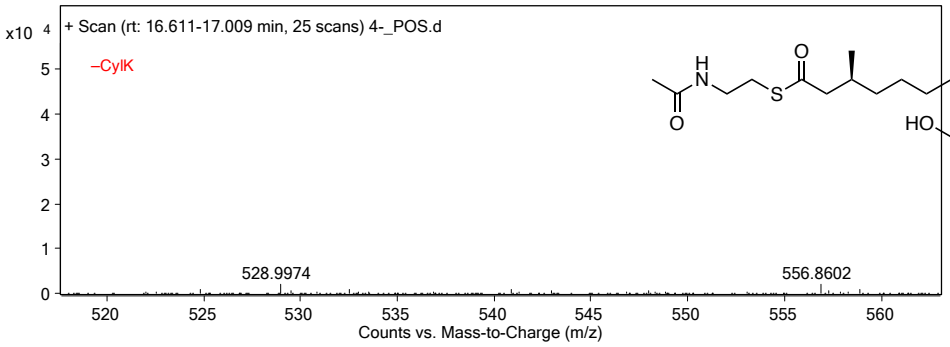
21a $C_{24}H_{32}O_3$ [M - H⁺] 367.2279

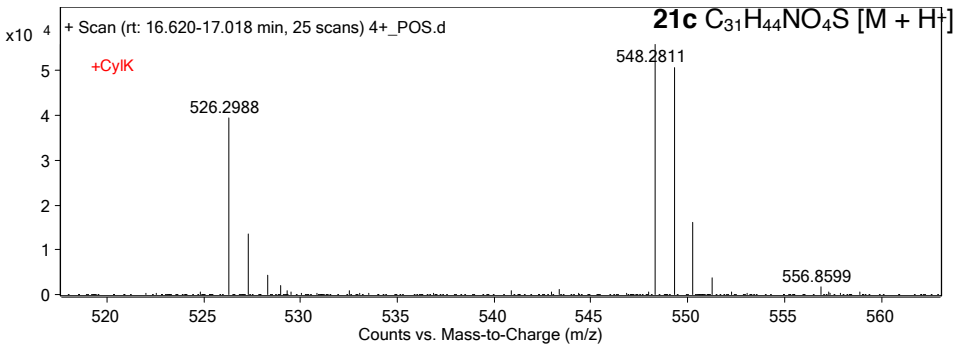




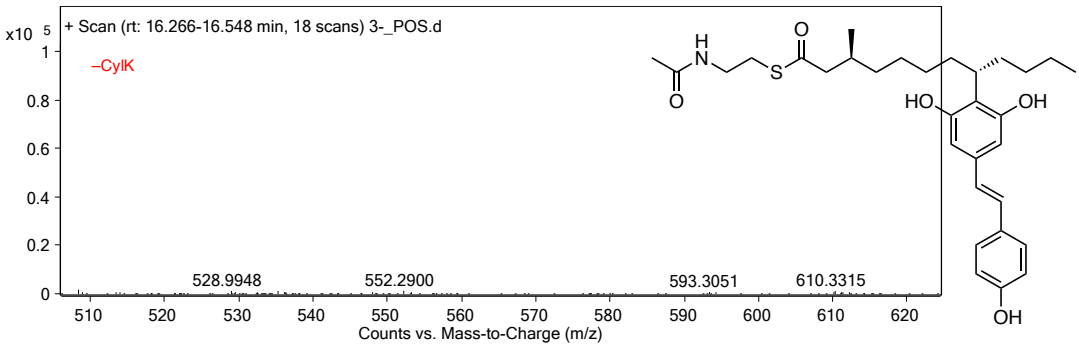


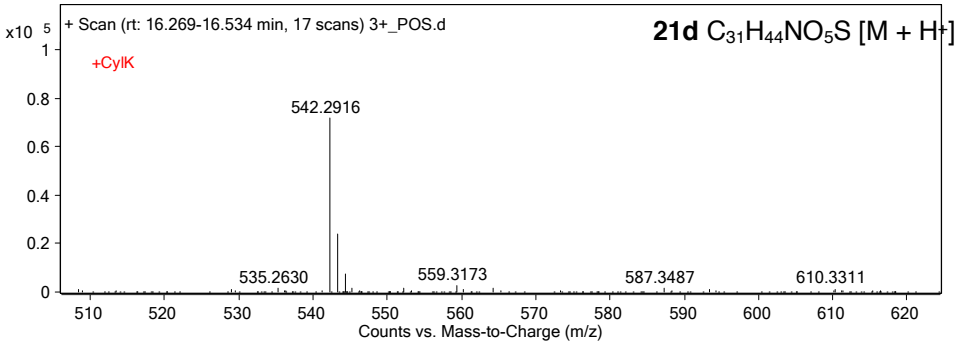
21b $C_{24}H_{32}O_4$ [M - H$^+$] 383.2228





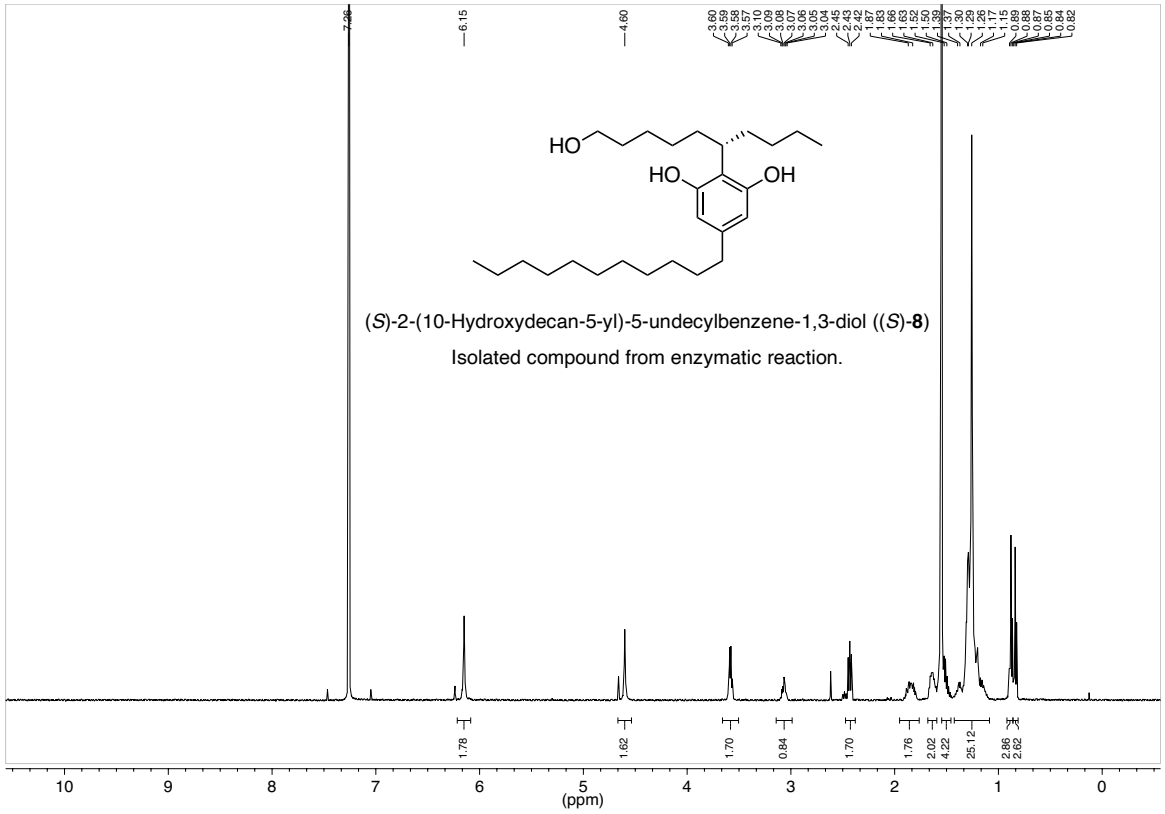
21c $C_{31}H_{44}NO_4S$ [M + H$^+$] 526.2986

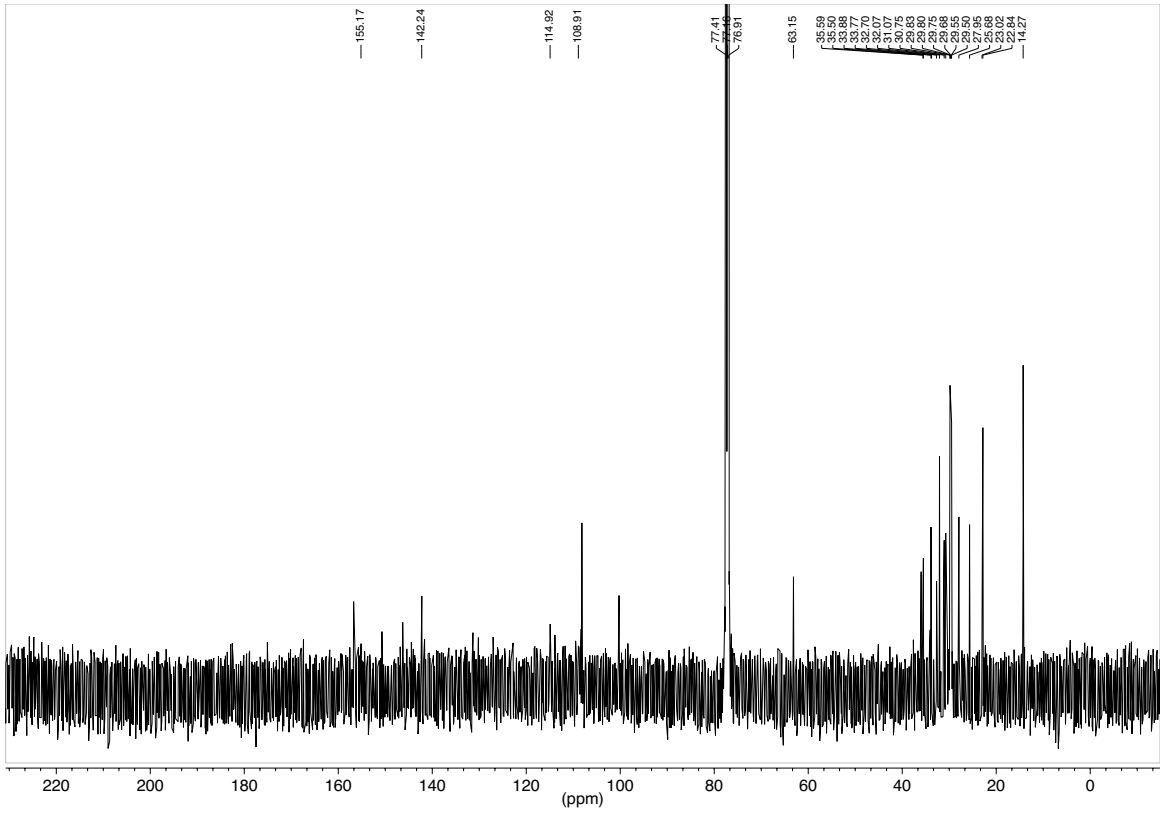


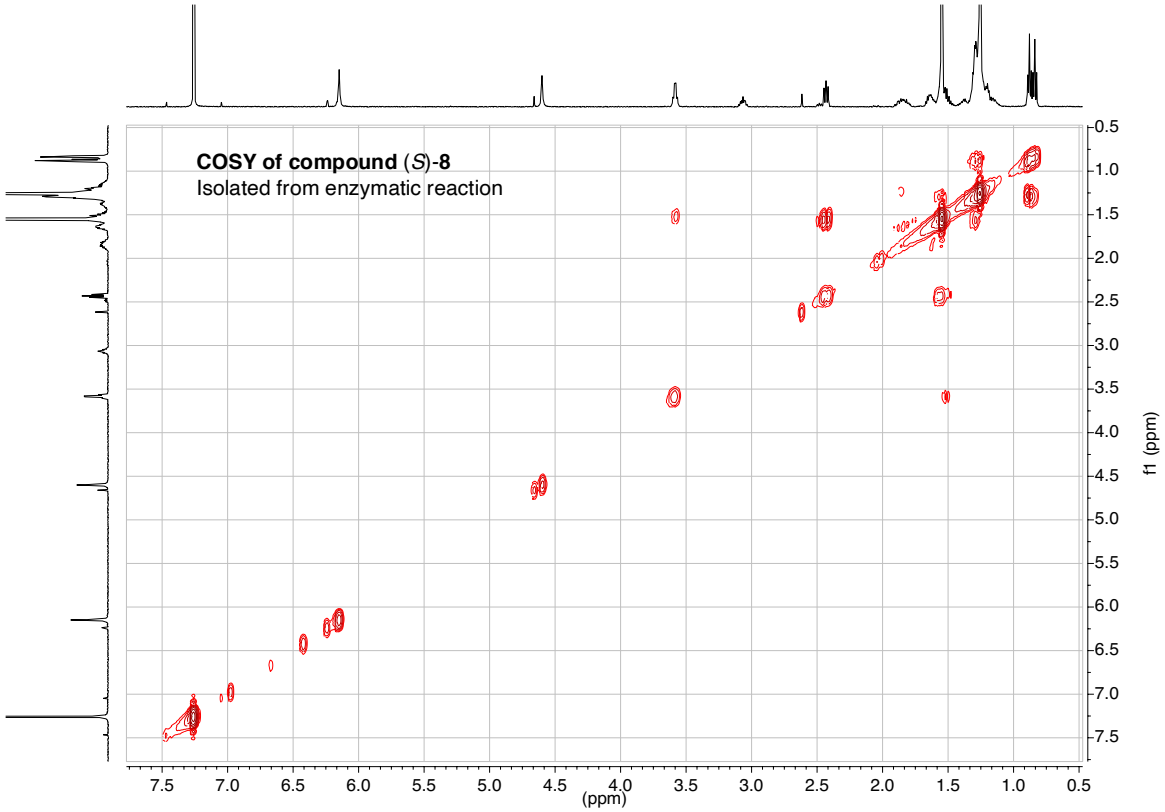


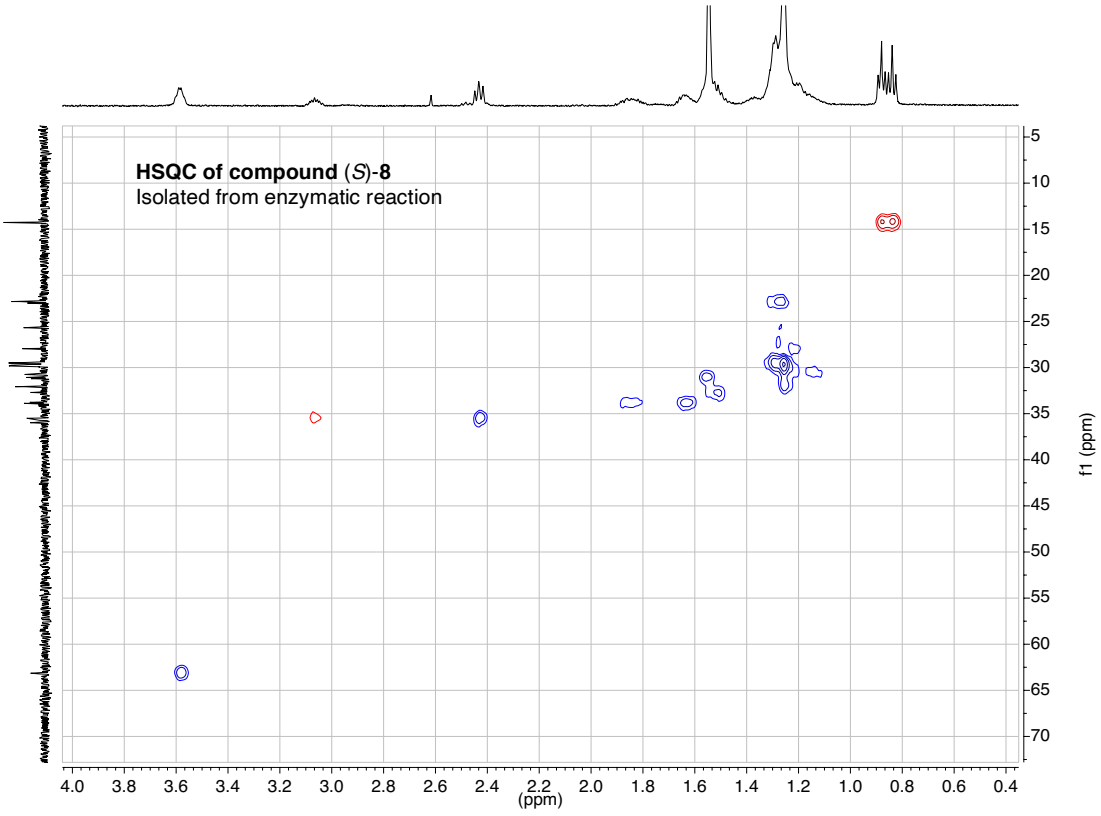
21d $C_{31}H_{44}NO_5S$ [M + H$^+$] 542.2935

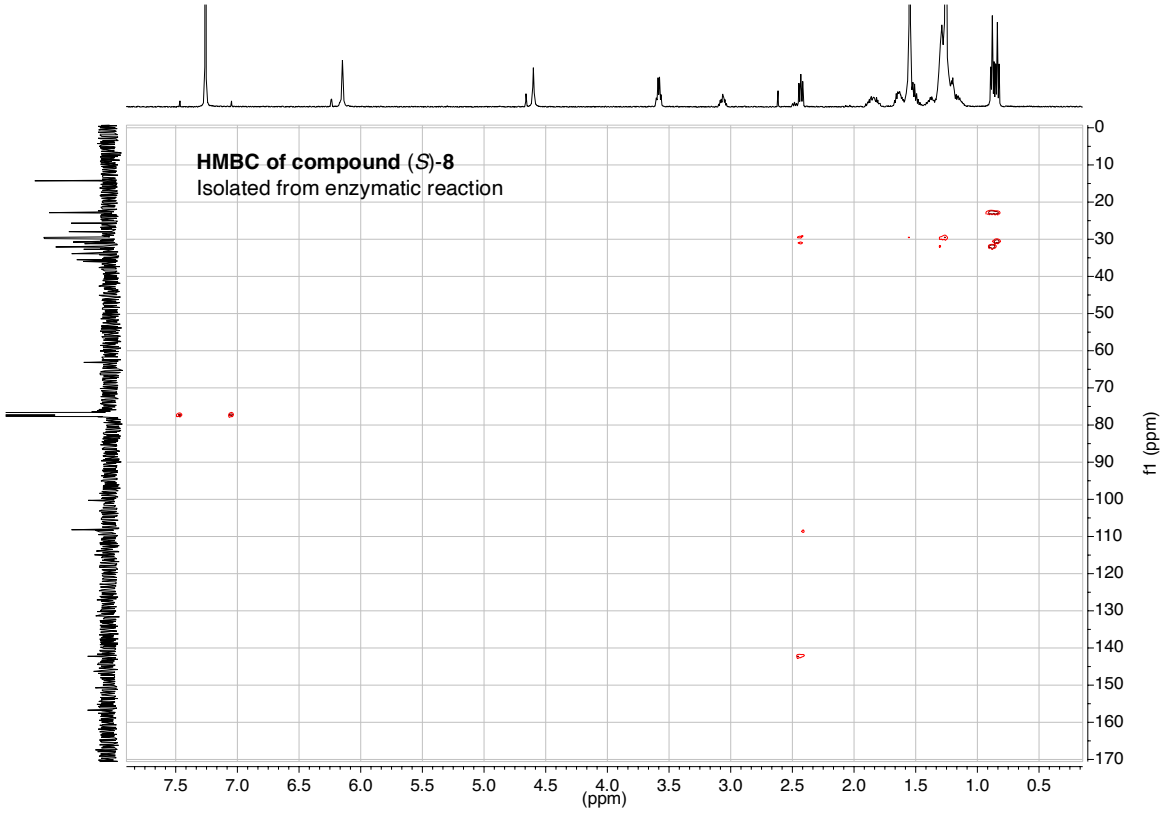
2.4.13 NMR Characterization Data

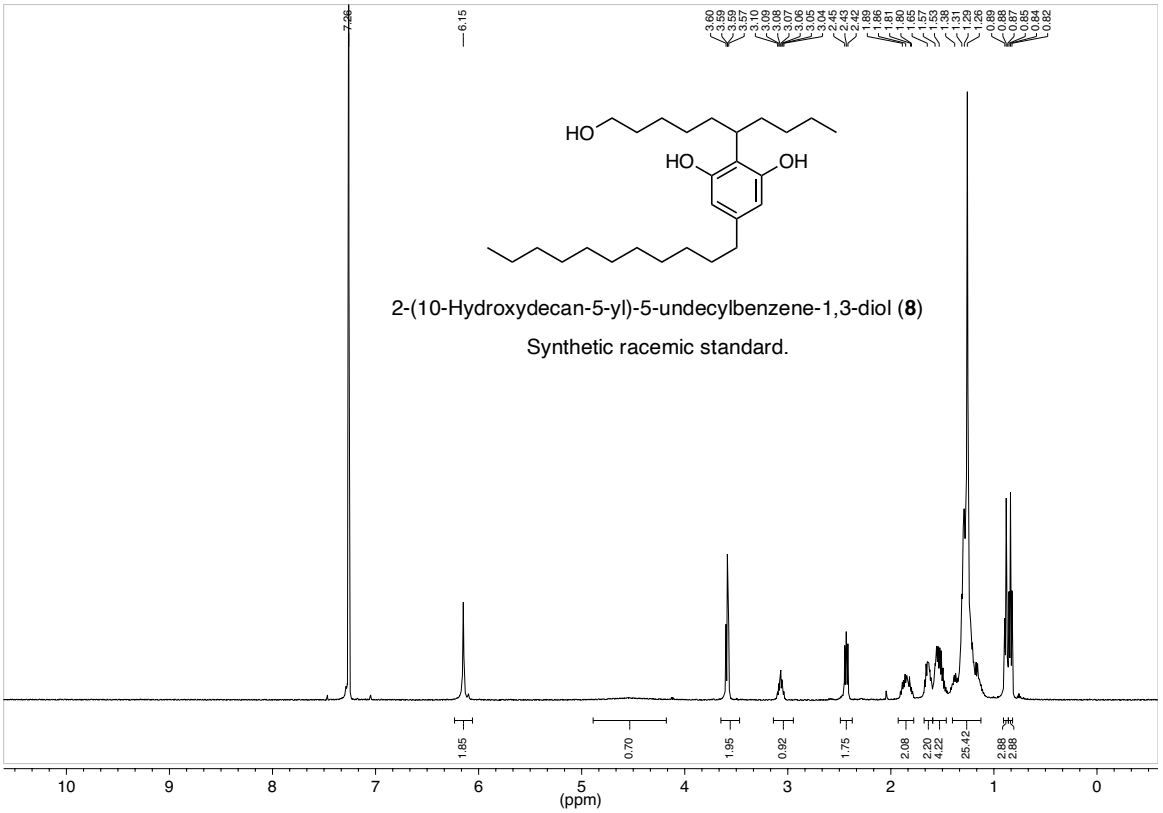





COSY of compound (S)-8
Isolated from enzymatic reaction
(ppm)
f1 (ppm)

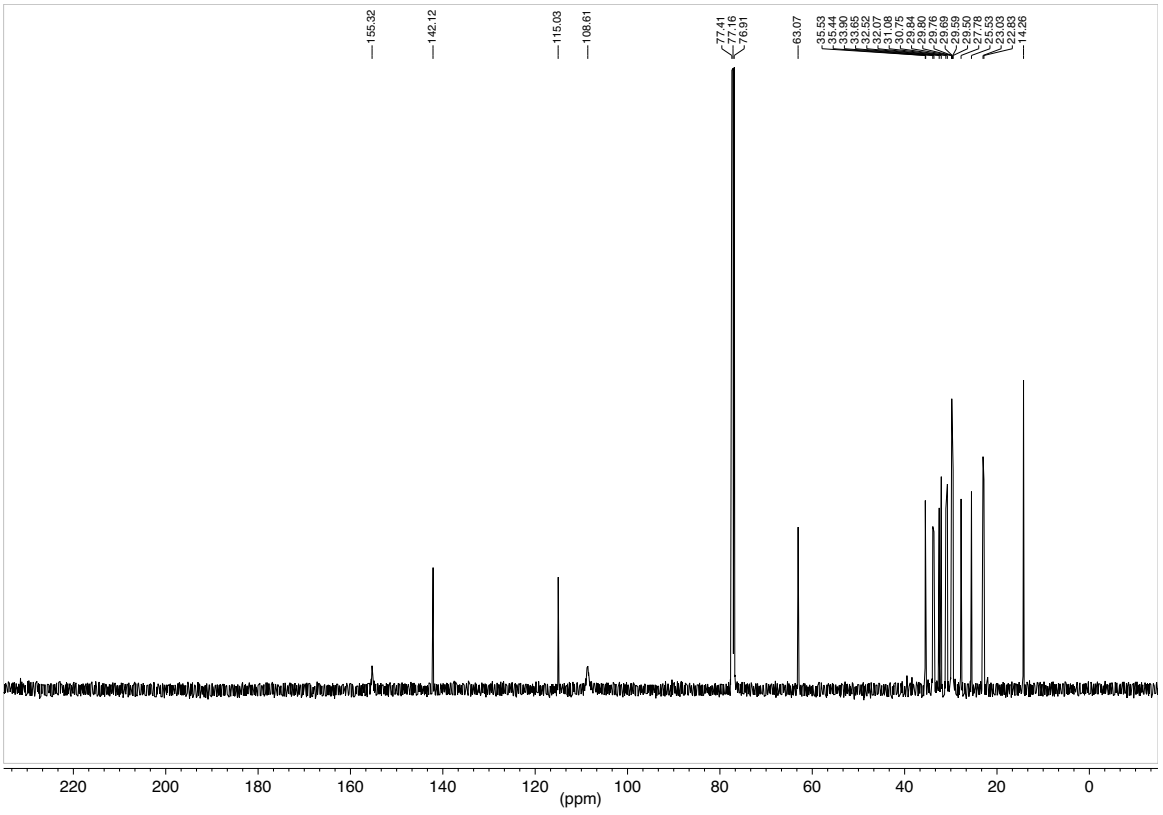

HSQC of compound (S)-8
Isolated from enzymatic reaction
(ppm)
f1 (ppm)

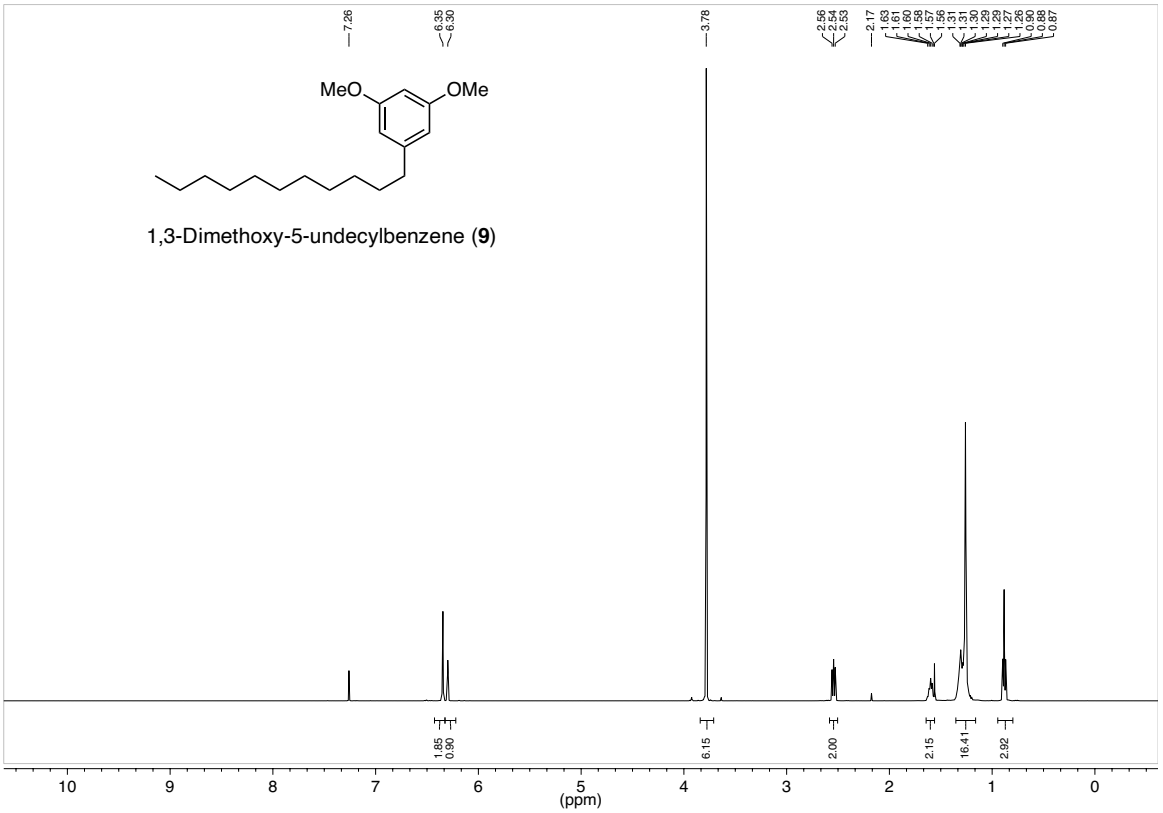

HMBC of compound (S)-8
Isolated from enzymatic reaction
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
(ppm)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
f1 (ppm)

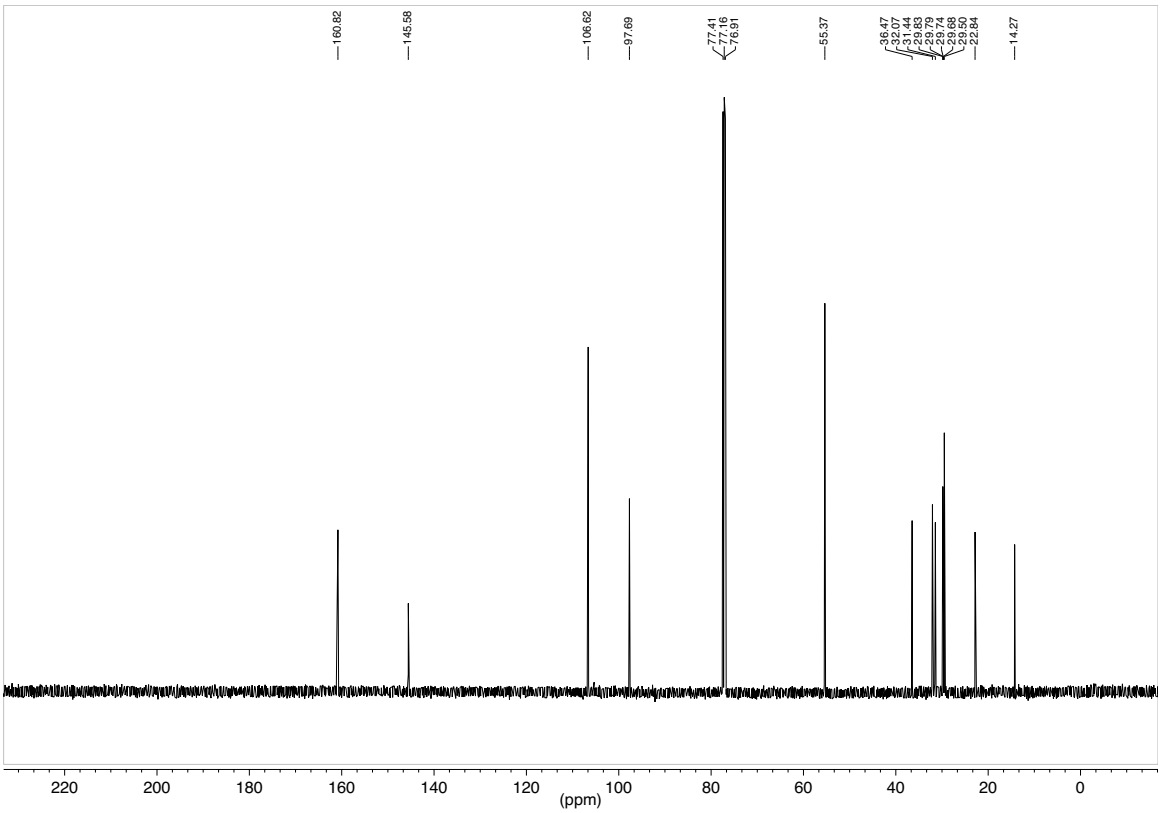


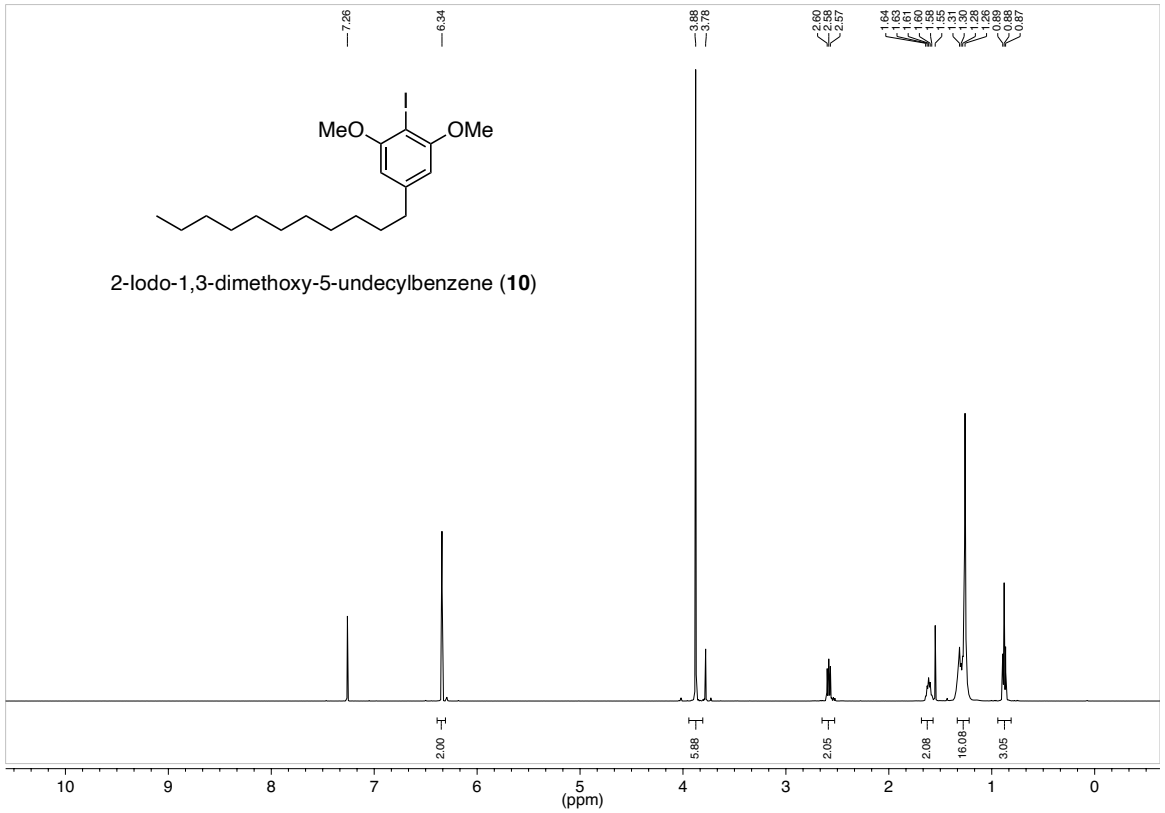
2-(10-Hydroxydecan-5-yl)-5-undecylbenzene-1,3-diol (**8**)

Synthetic racemic standard.

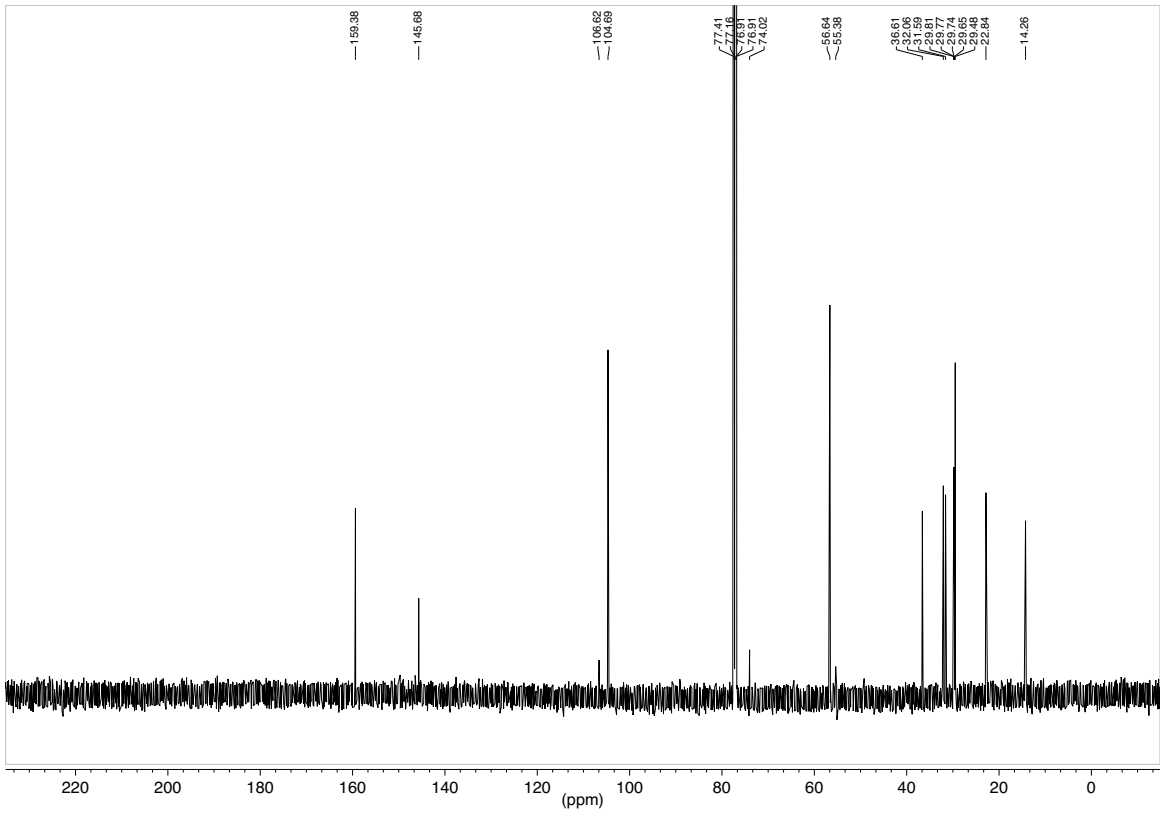


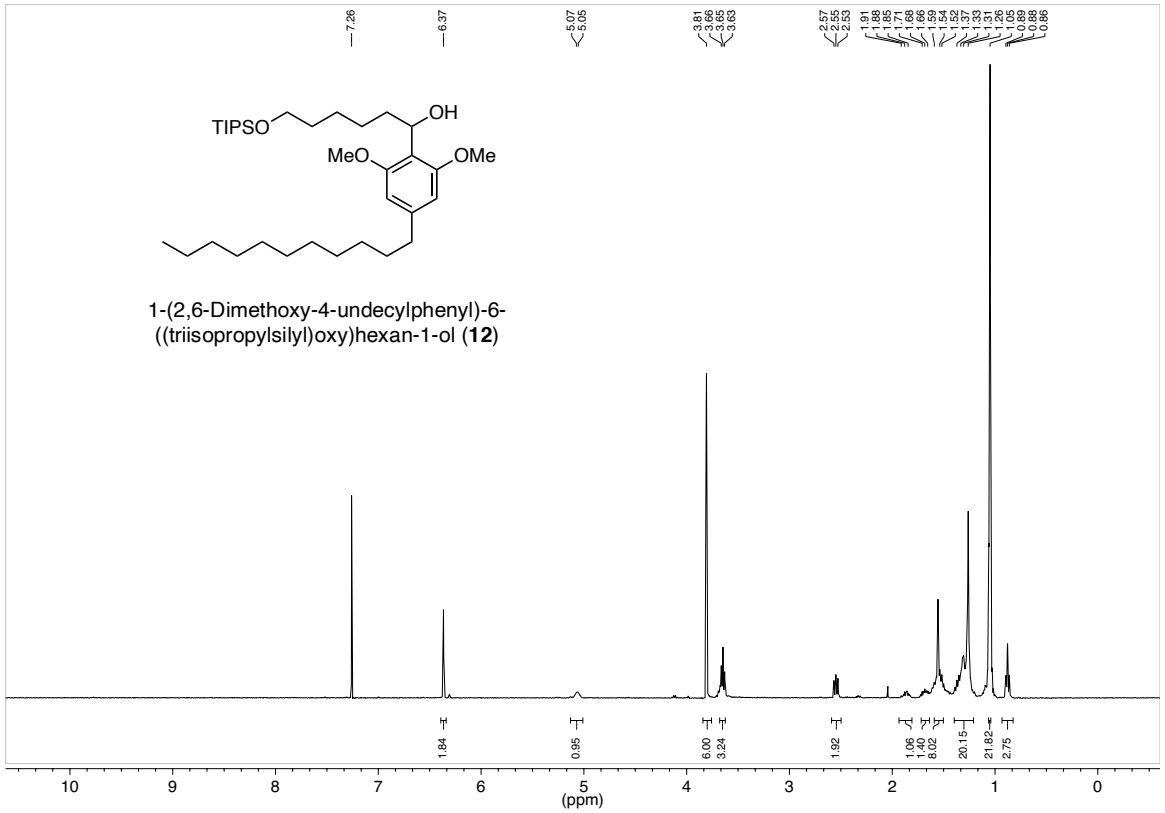

7.26
6.35
6.30
3.78
2.56
2.54
2.53
2.17
1.63
1.61
1.60
1.58
1.57
1.56
1.31
1.30
1.29
1.29
1.27
1.26
0.90
0.88
0.87
MeO
OMe
1,3-Dimethoxy-5-undecylbenzene (9)
1.85
0.90
6.15
2.00
2.15
16.41
2.92
10
9
8
7
6
5
4
3
2
1
0
(ppm)


160.82
145.58
106.62
97.69
77.41
77.16
76.91
55.37
36.47
34.07
31.44
29.83
29.79
29.74
29.68
29.50
22.84
14.27
220
200
180
160
140
120
100
80
60
40
20
0
(ppm)

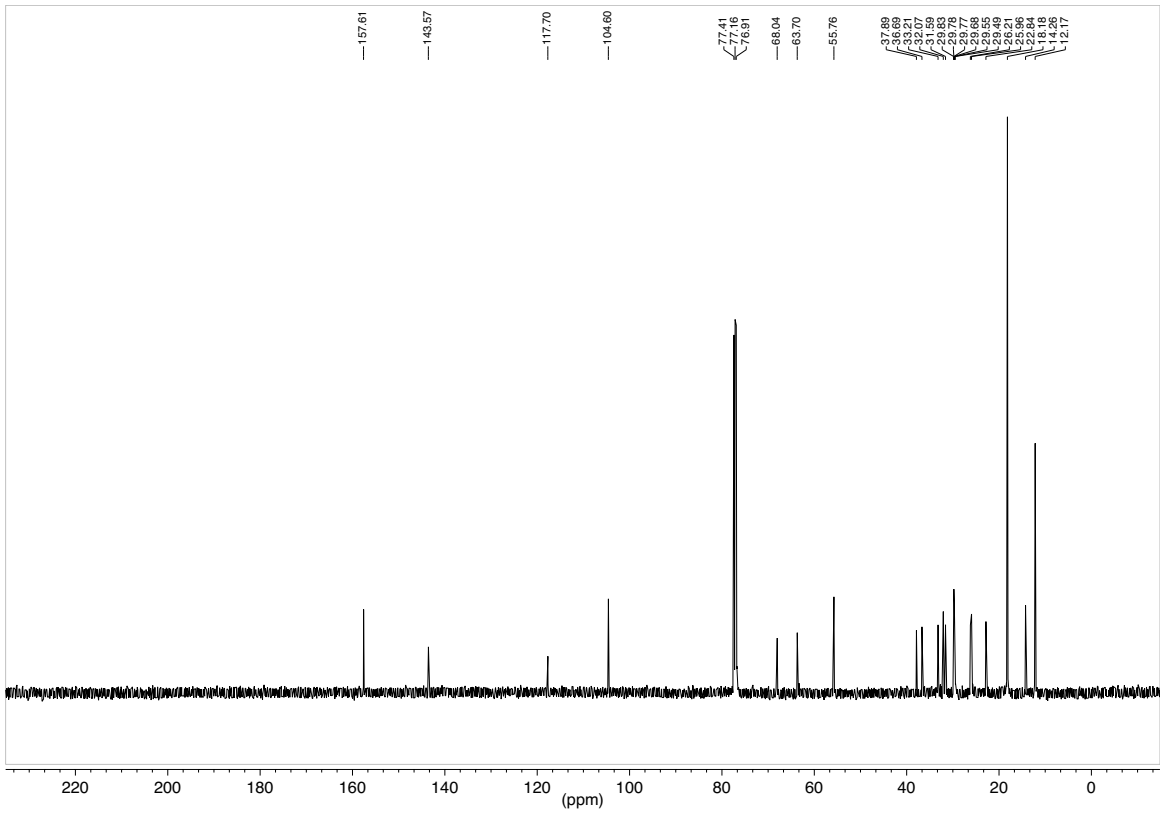


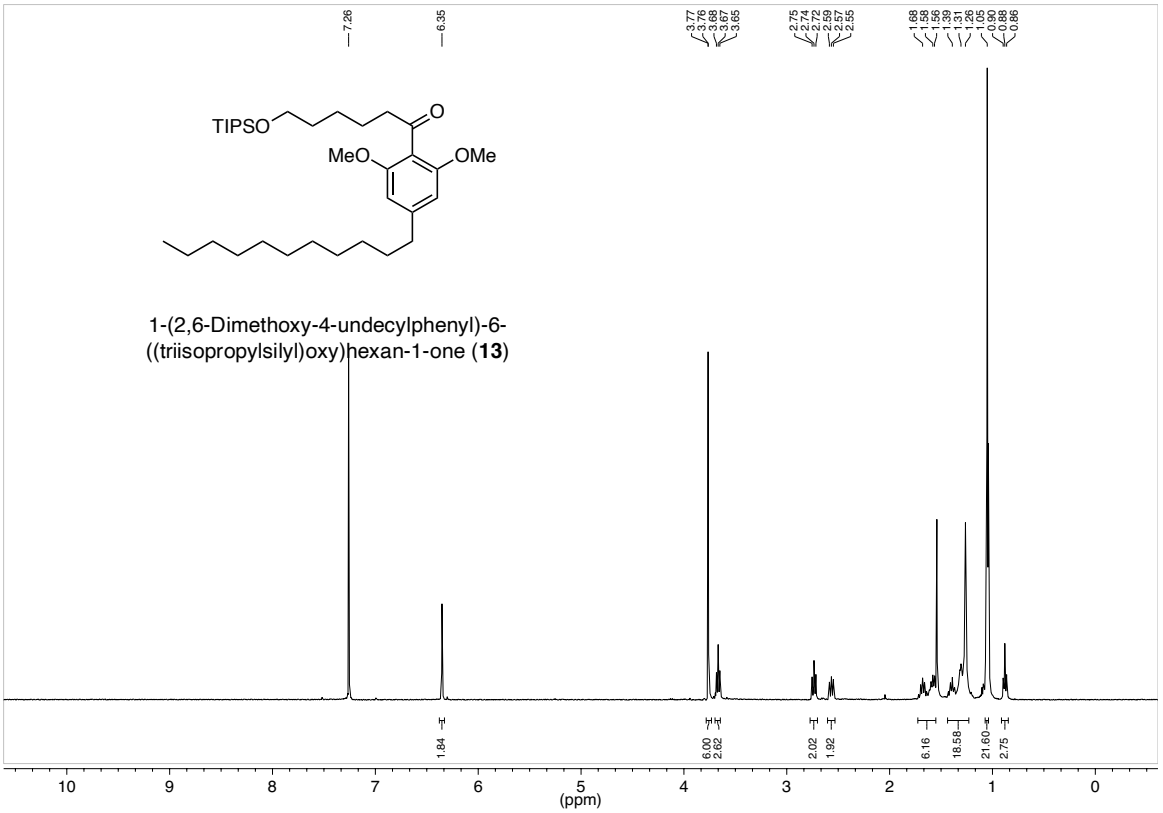
2-Iodo-1,3-dimethoxy-5-undecylbenzene (**10**)

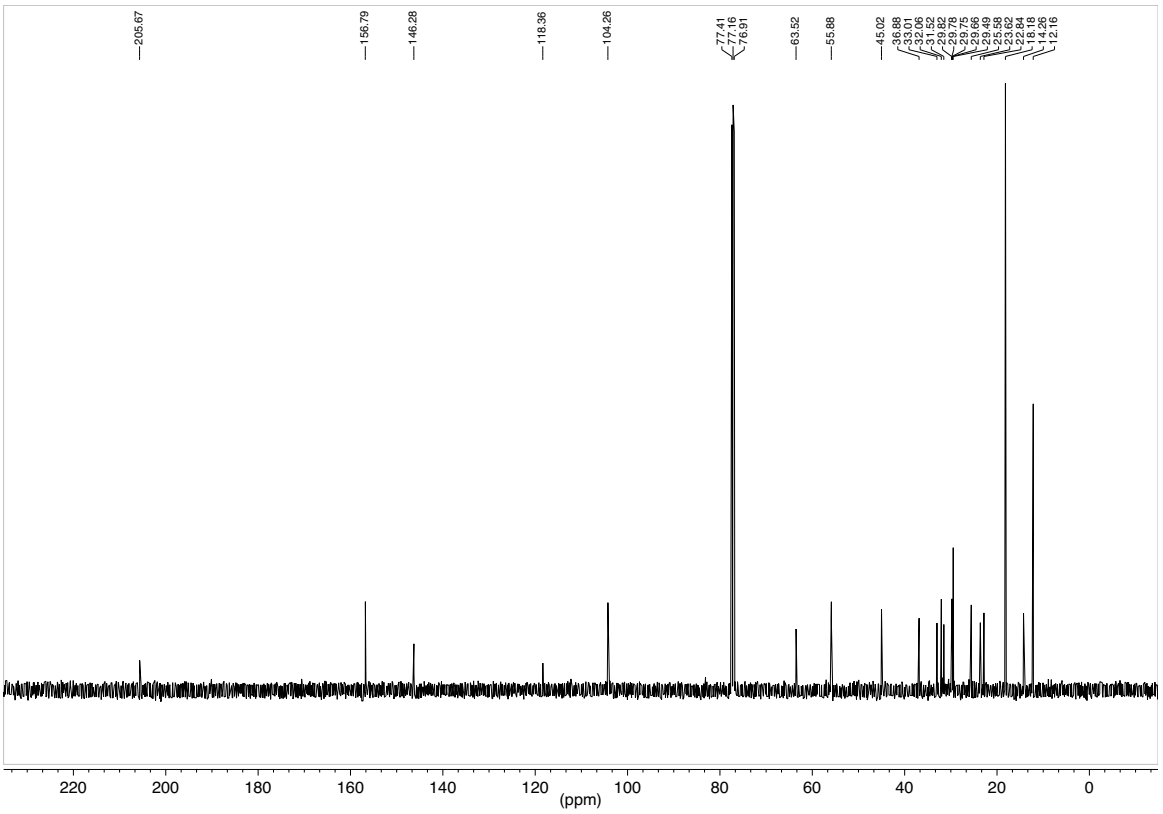


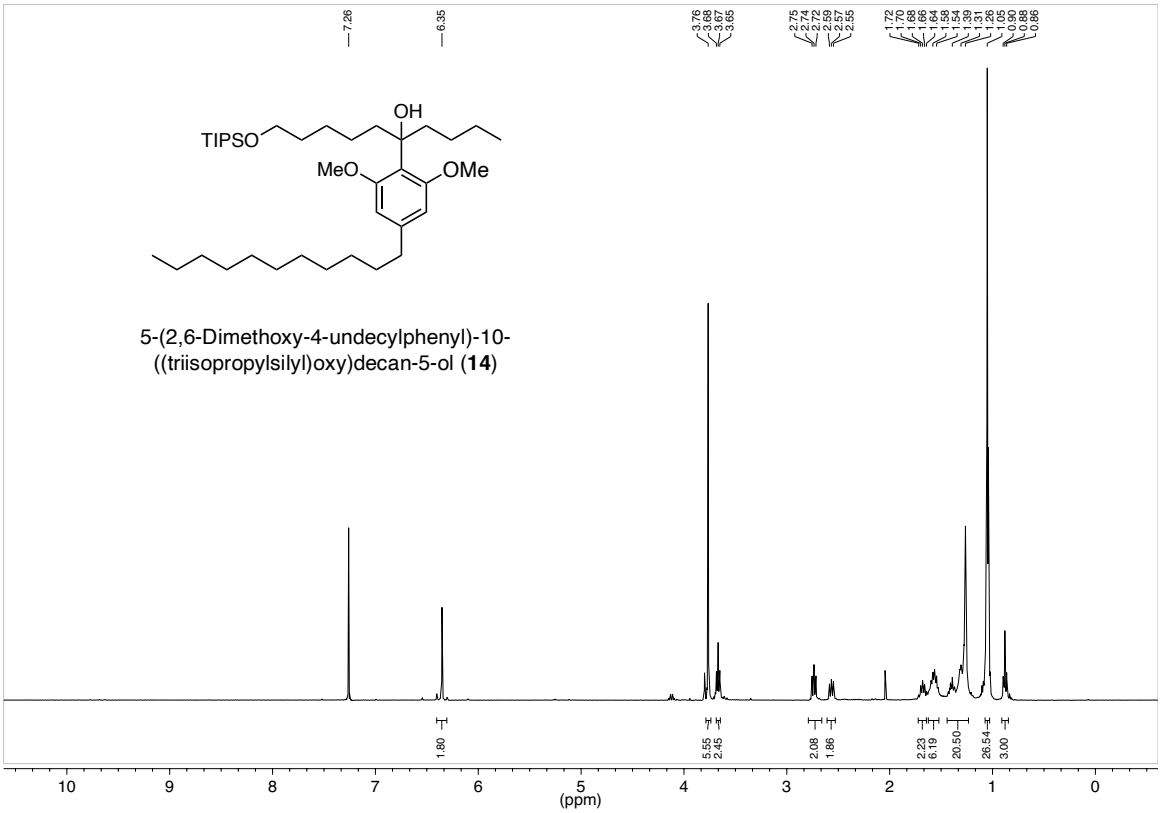


1-(2,6-Dimethoxy-4-undecylphenyl)-6-((triisopropylsilyl)oxy)hexan-1-ol (**12**)

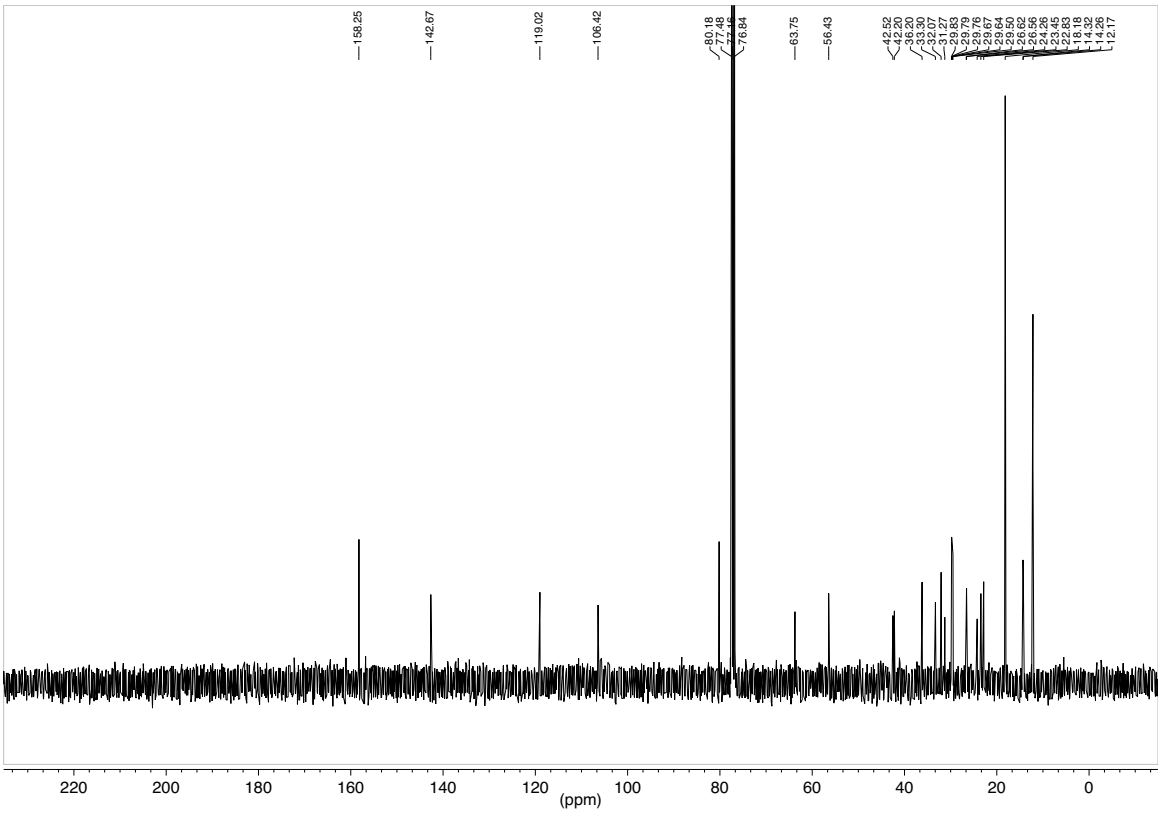


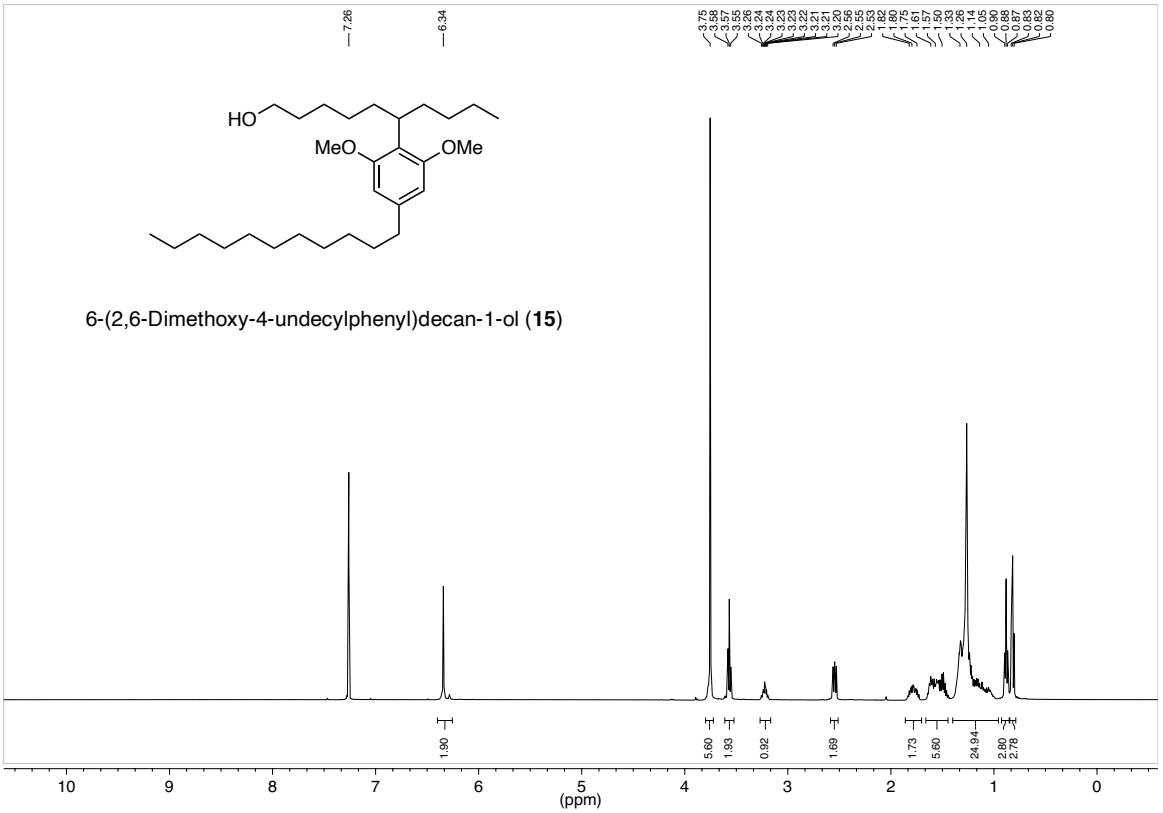

7.26
6.35
3.77
3.76
3.68
3.67
3.65
2.75
2.74
2.72
2.59
2.57
2.55
1.68
1.58
1.56
1.39
1.31
1.26
1.05
0.90
0.88
0.86
TIPSO
O
MeO
OMe
1-(2,6-Dimethoxy-4-undecylphenyl)-6-
((triisopropylsilyl)oxy)hexan-1-one (**13**)
1.84
6.00
2.62
2.02
1.92
6.16
18.58
21.60
2.75
10
9
8
7
6
5
4
3
2
1
0
(ppm)


205.67
156.79
146.28
118.36
104.26
77.41
77.16
76.91
63.52
55.88
45.02
36.88
33.91
32.06
31.52
29.82
29.78
29.75
29.66
29.49
25.58
23.82
22.84
18.18
14.26
12.16
220
200
180
160
140
120
100
80
60
40
20
0
(ppm)

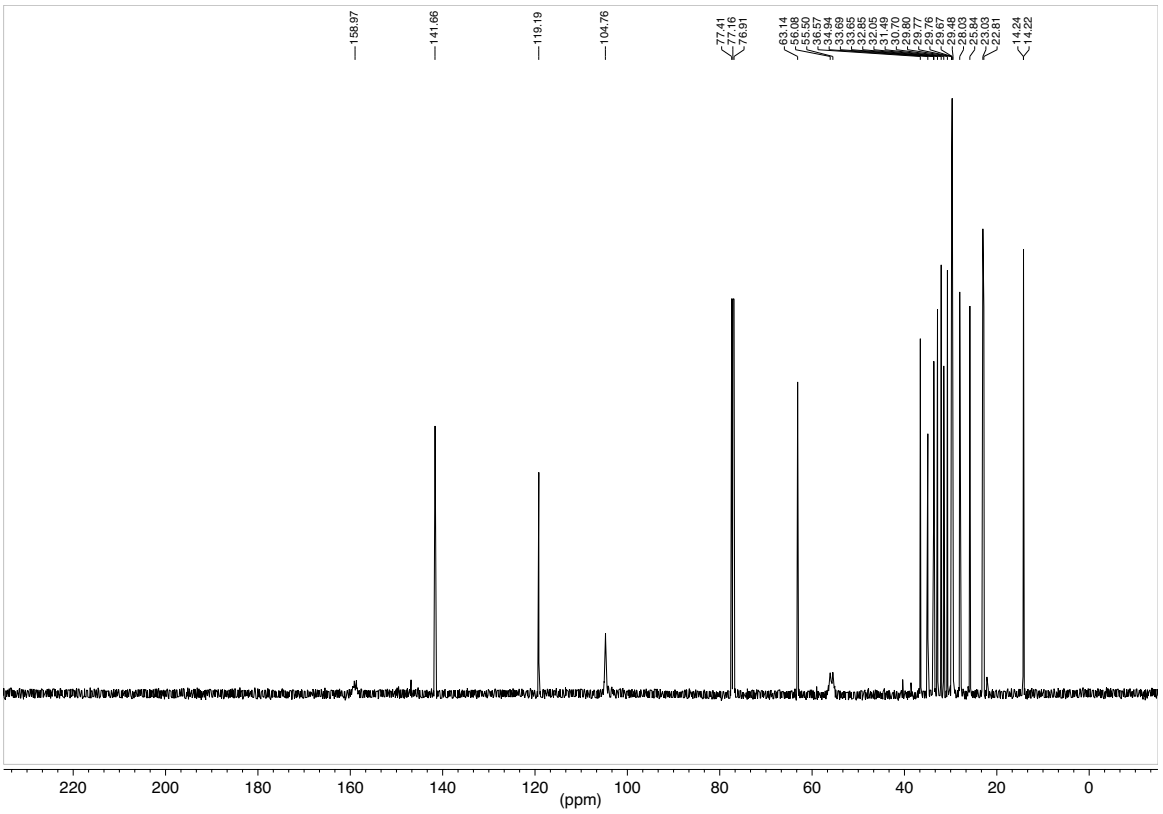


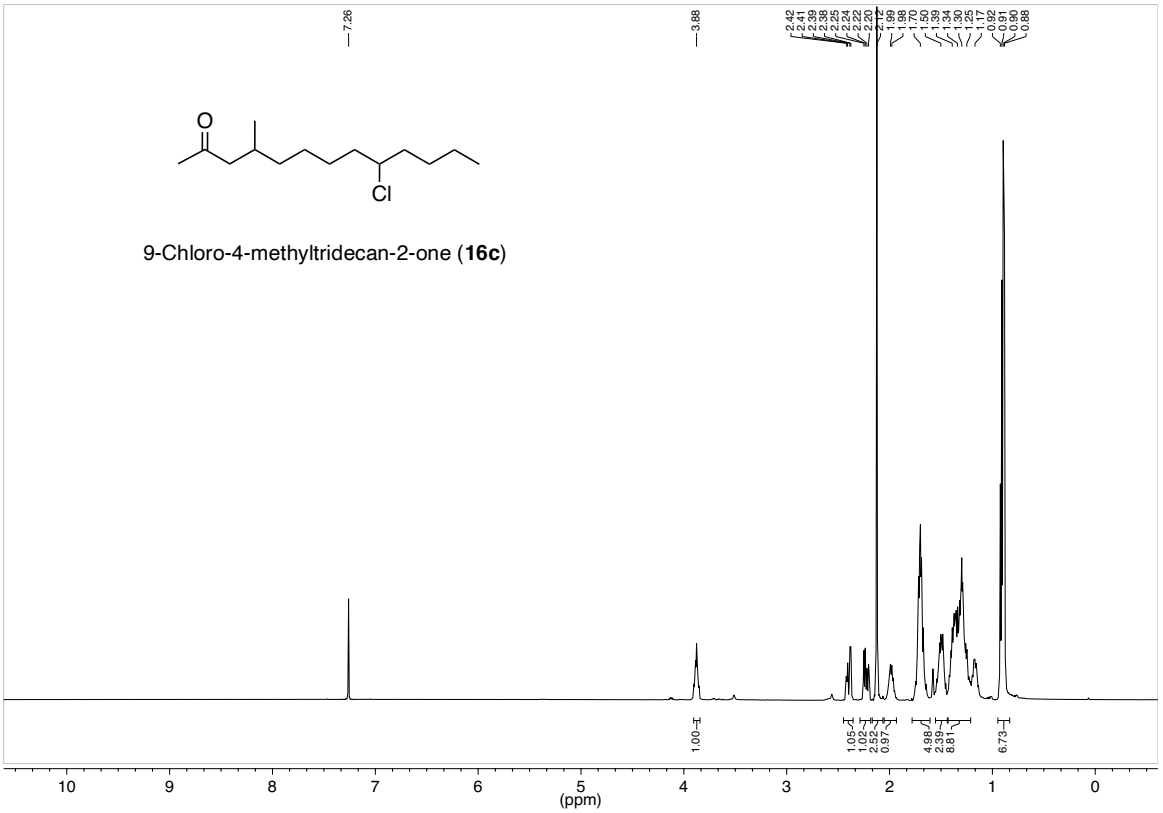
5-(2,6-Dimethoxy-4-undecylphenyl)-10-((triisopropylsilyl)oxy)decan-5-ol (**14**)



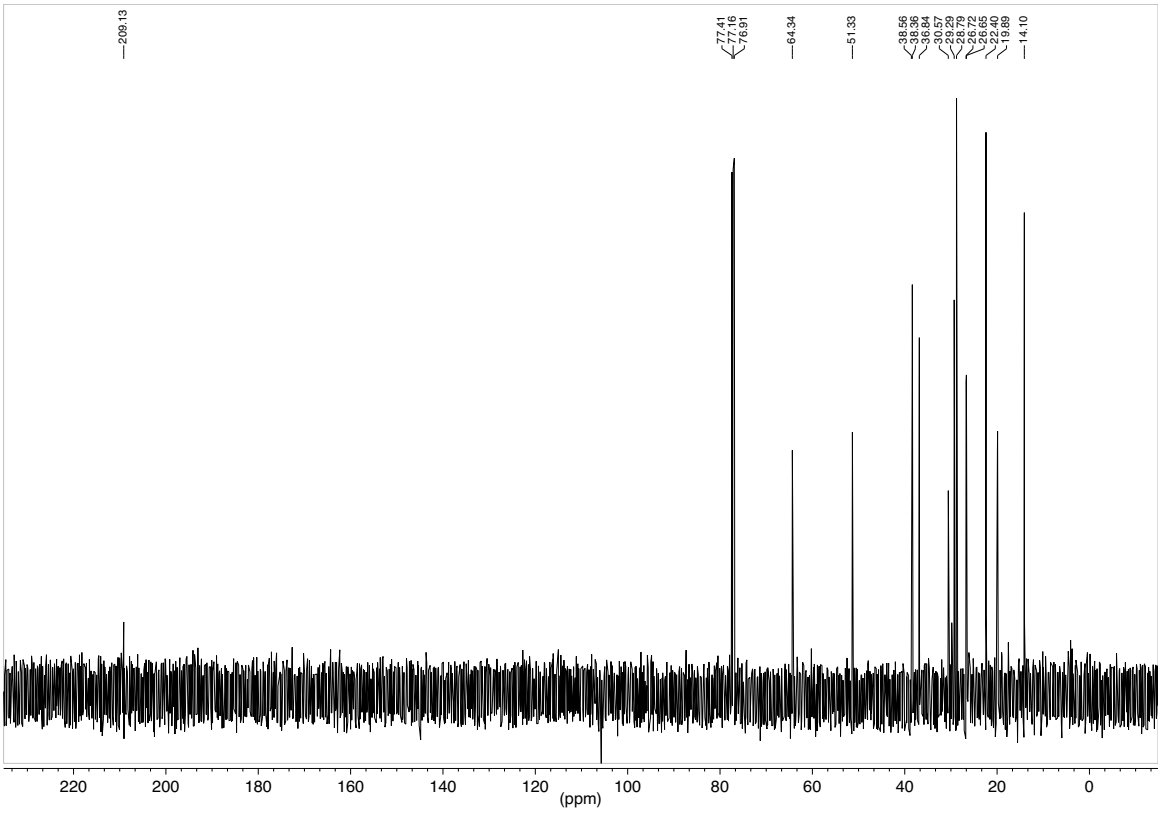


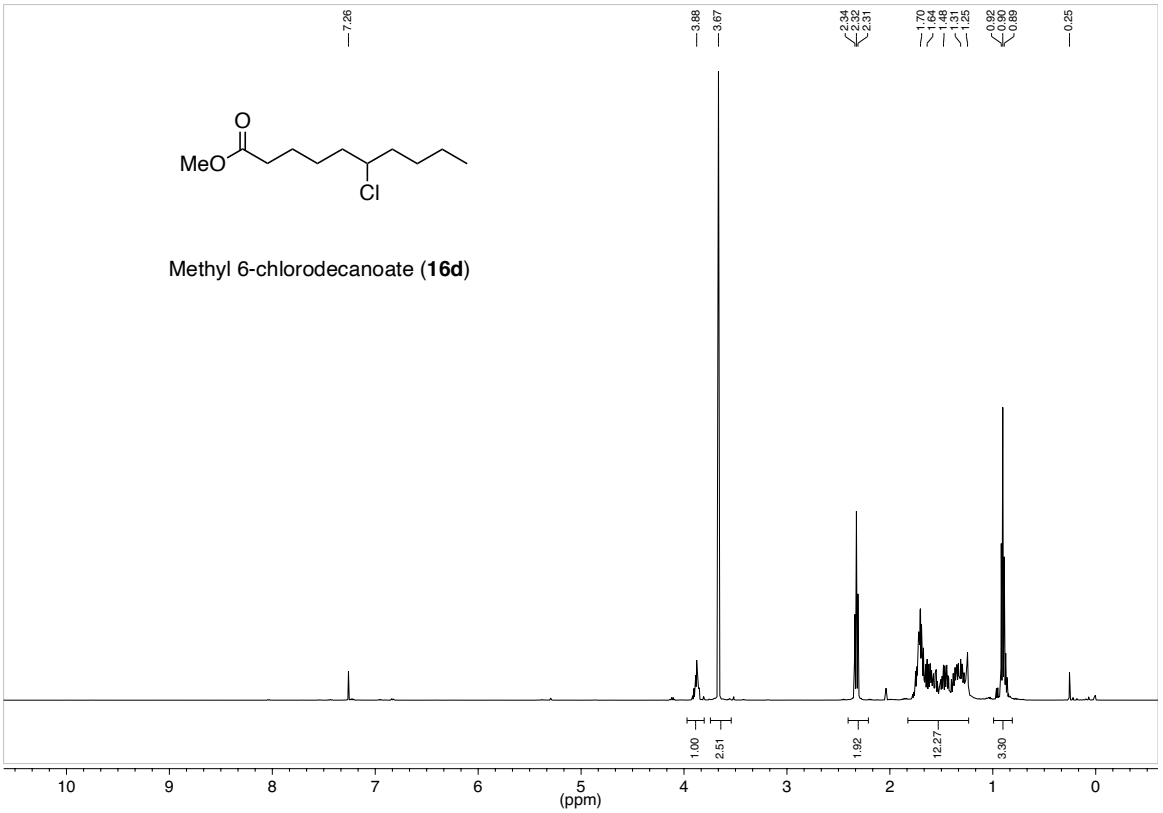
6-(2,6-Dimethoxy-4-undecylphenyl)decan-1-ol (**15**)



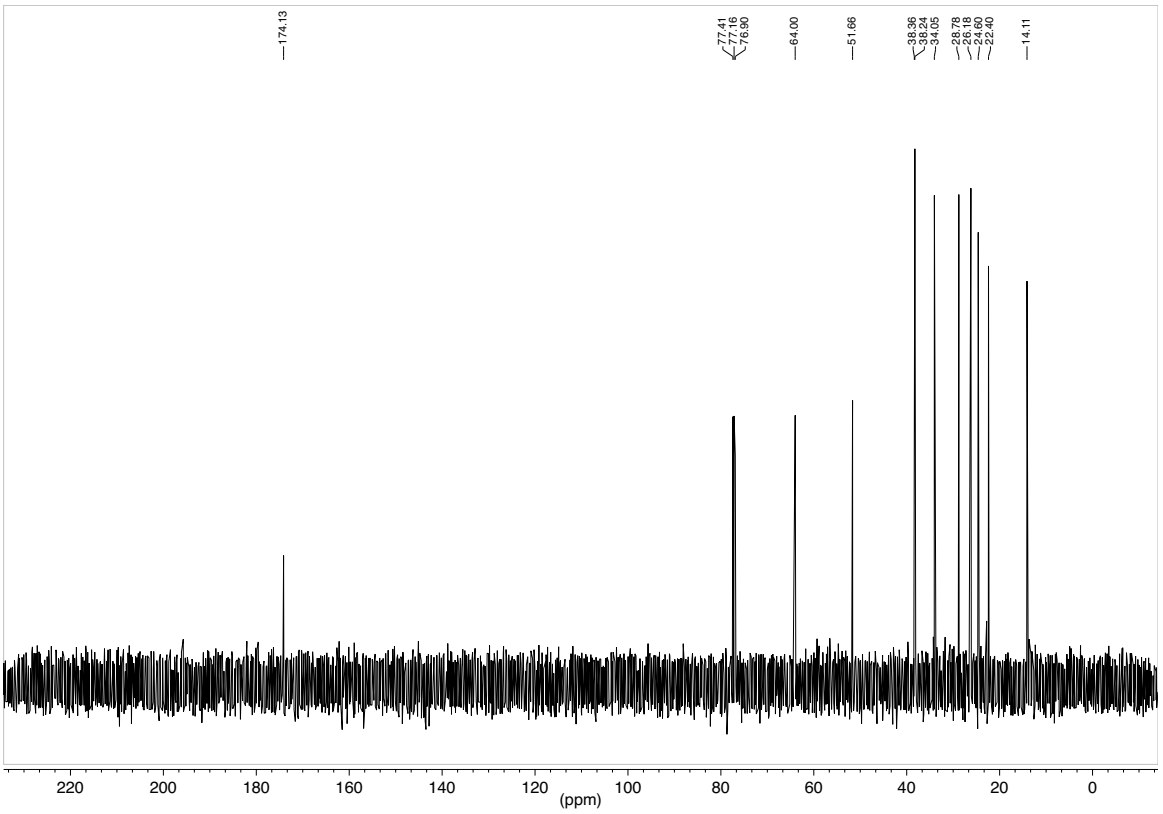


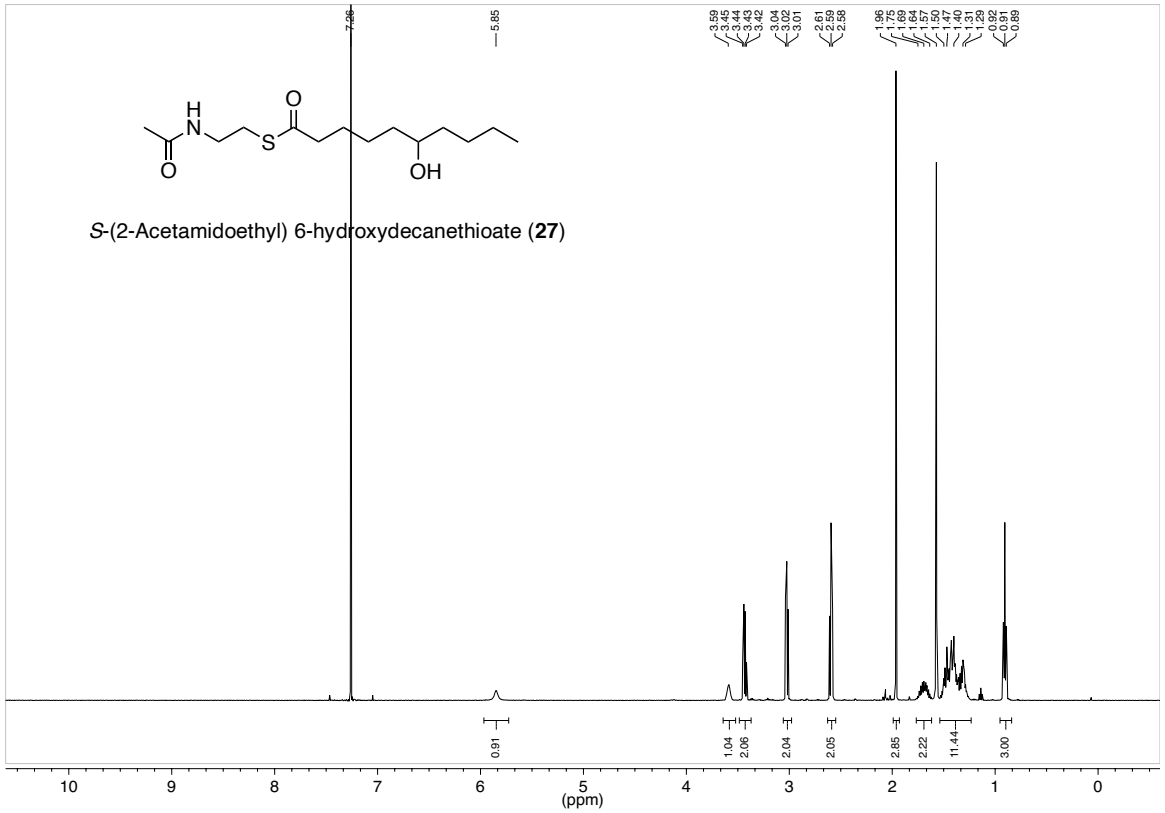
9-Chloro-4-methyltridecan-2-one (**16c**)



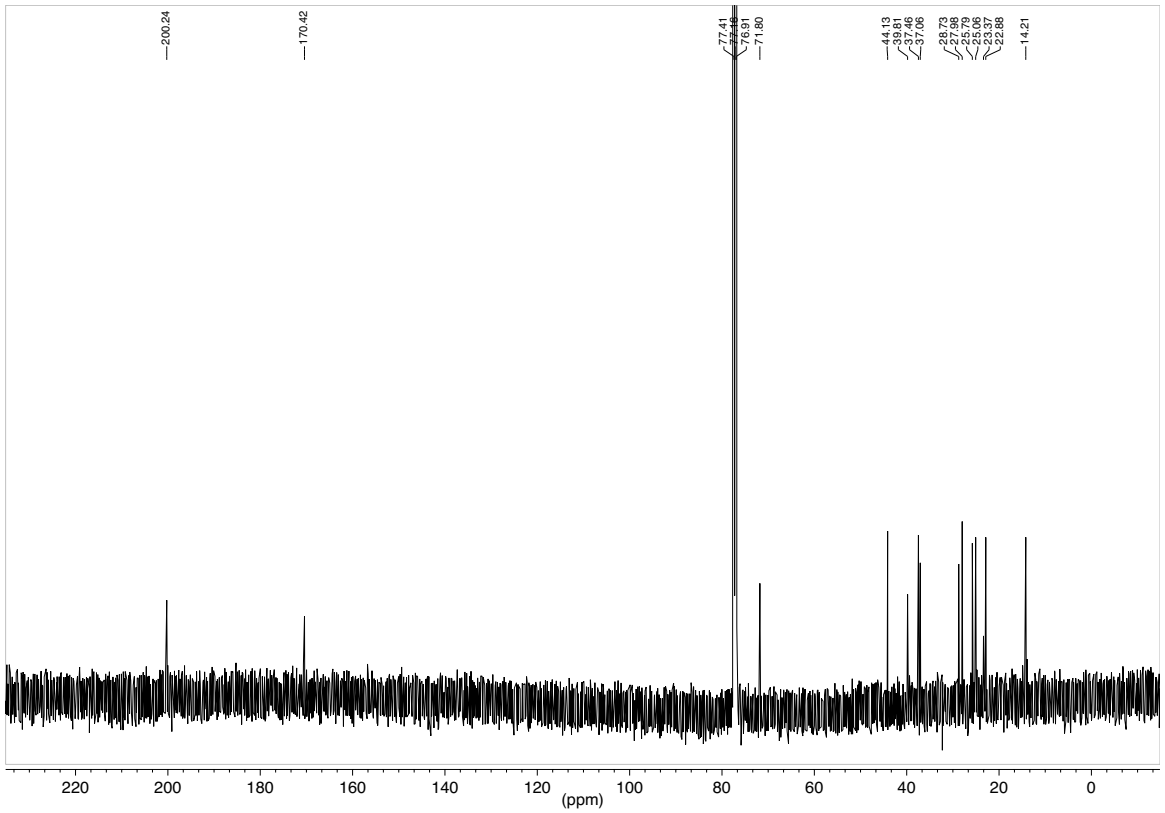


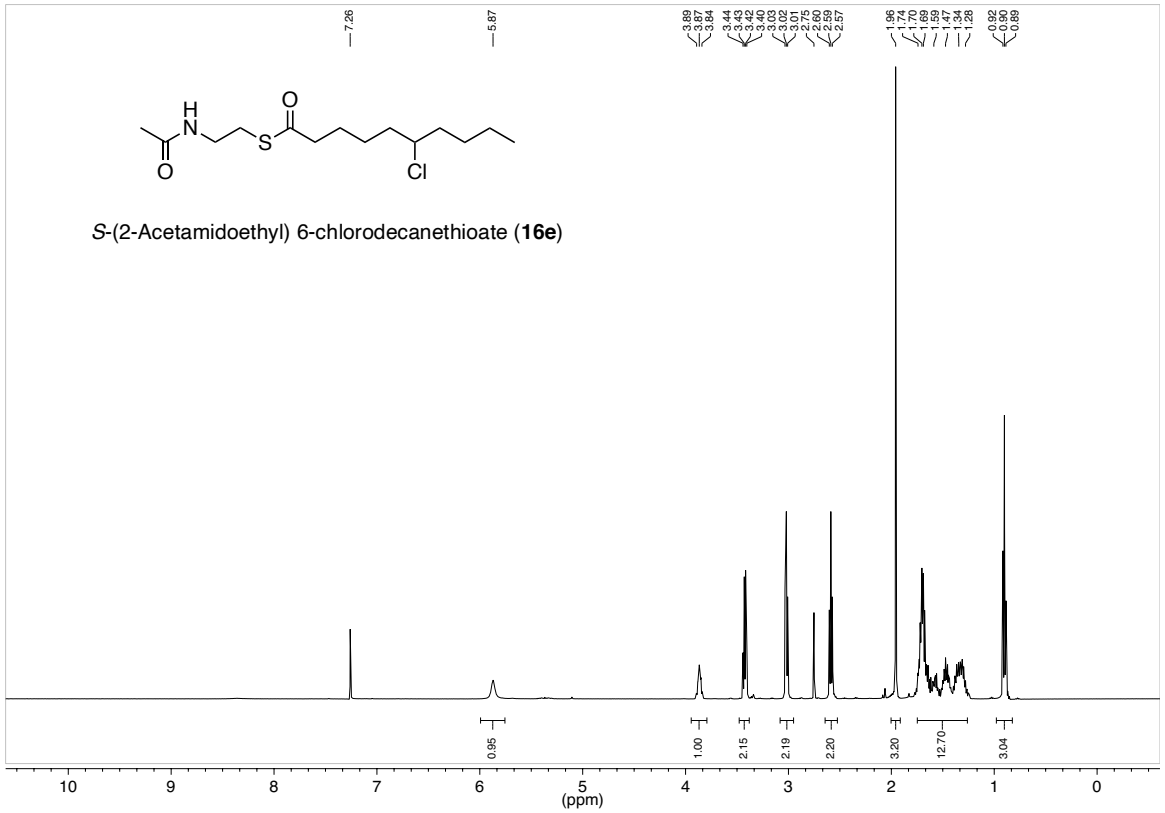
Methyl 6-chlorodecanoate (**16d**)



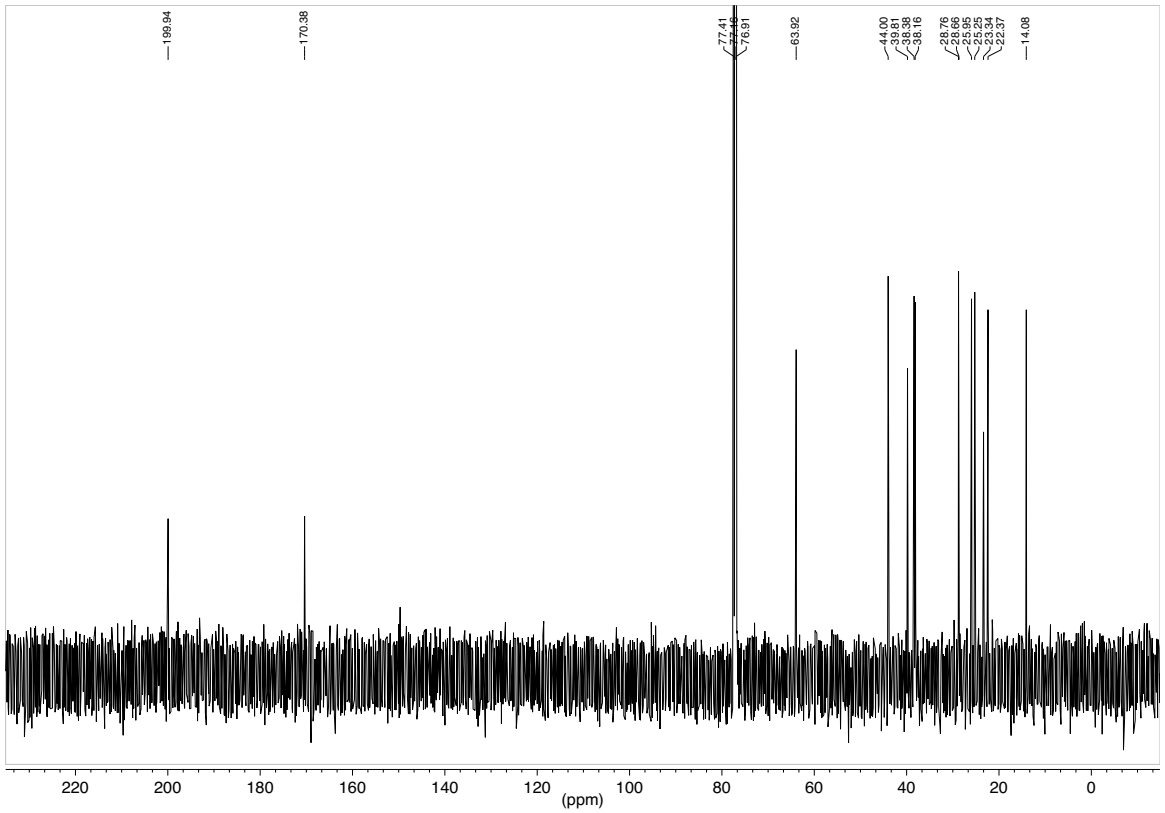


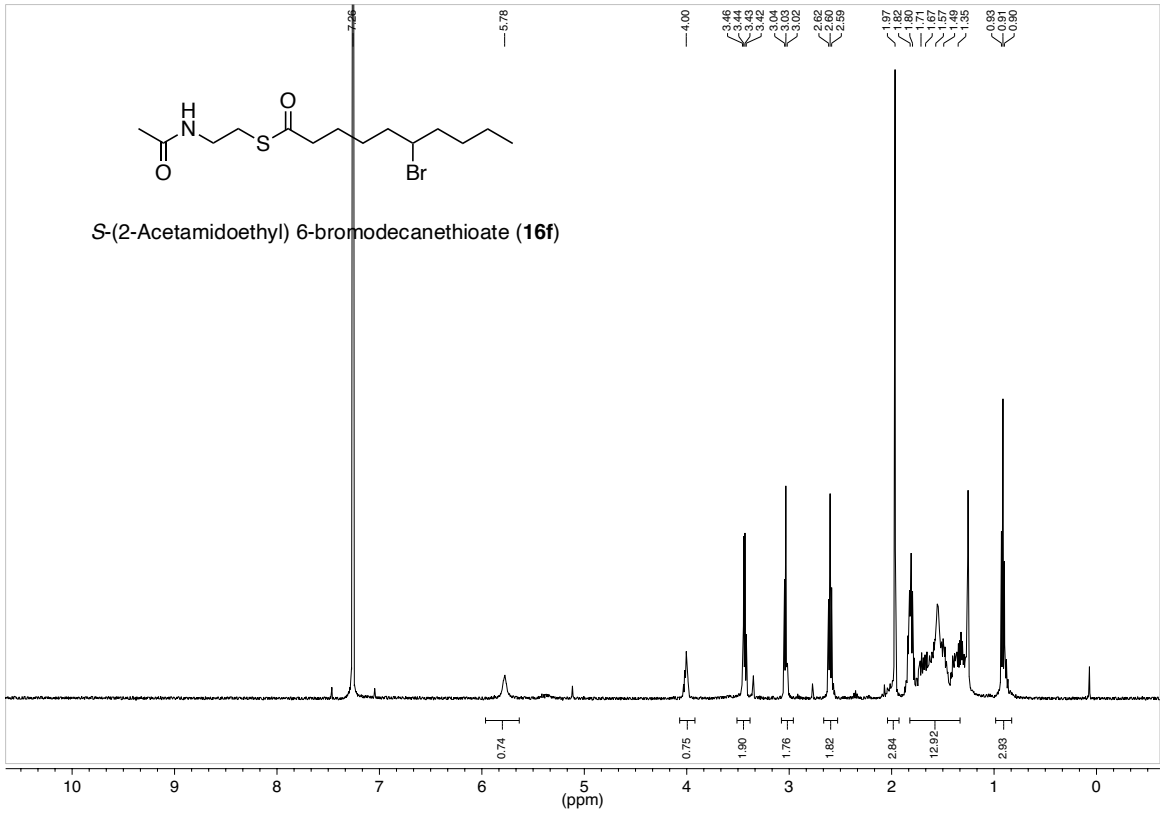
S-(2-Acetamidoethyl) 6-hydroxydecanethioate (**27**)

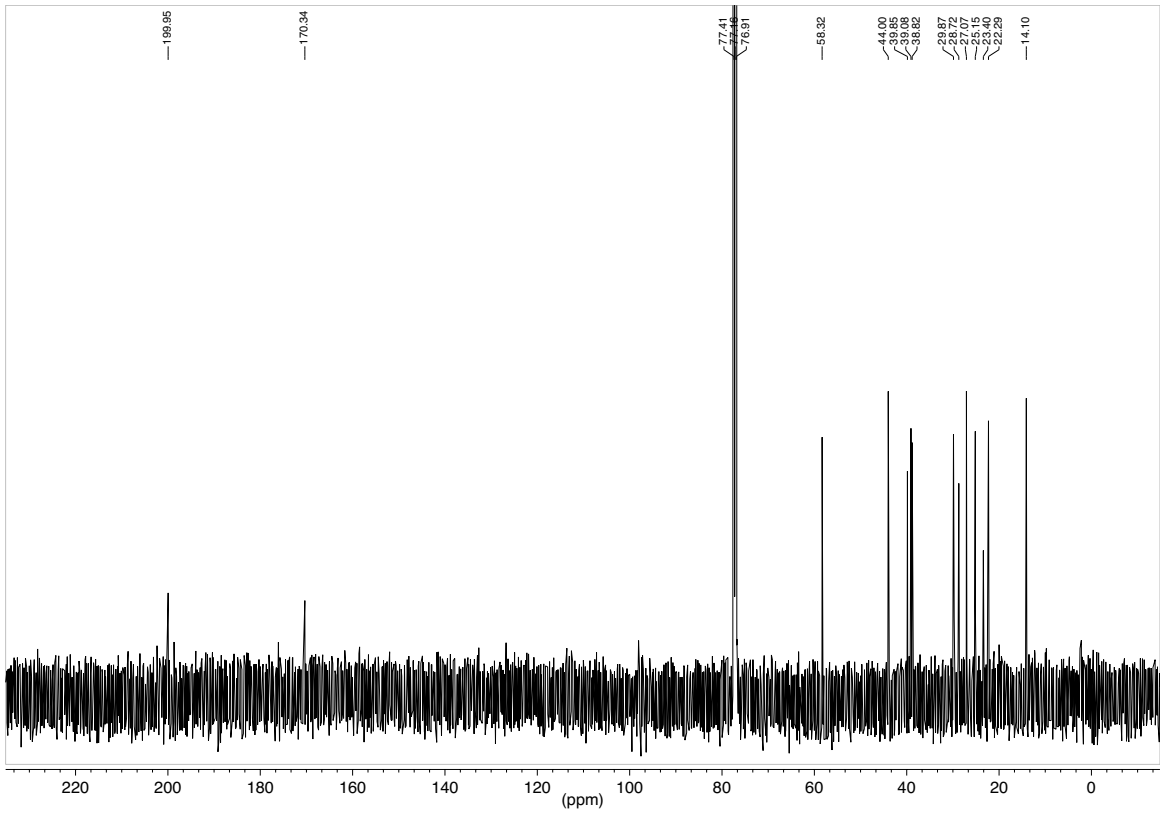


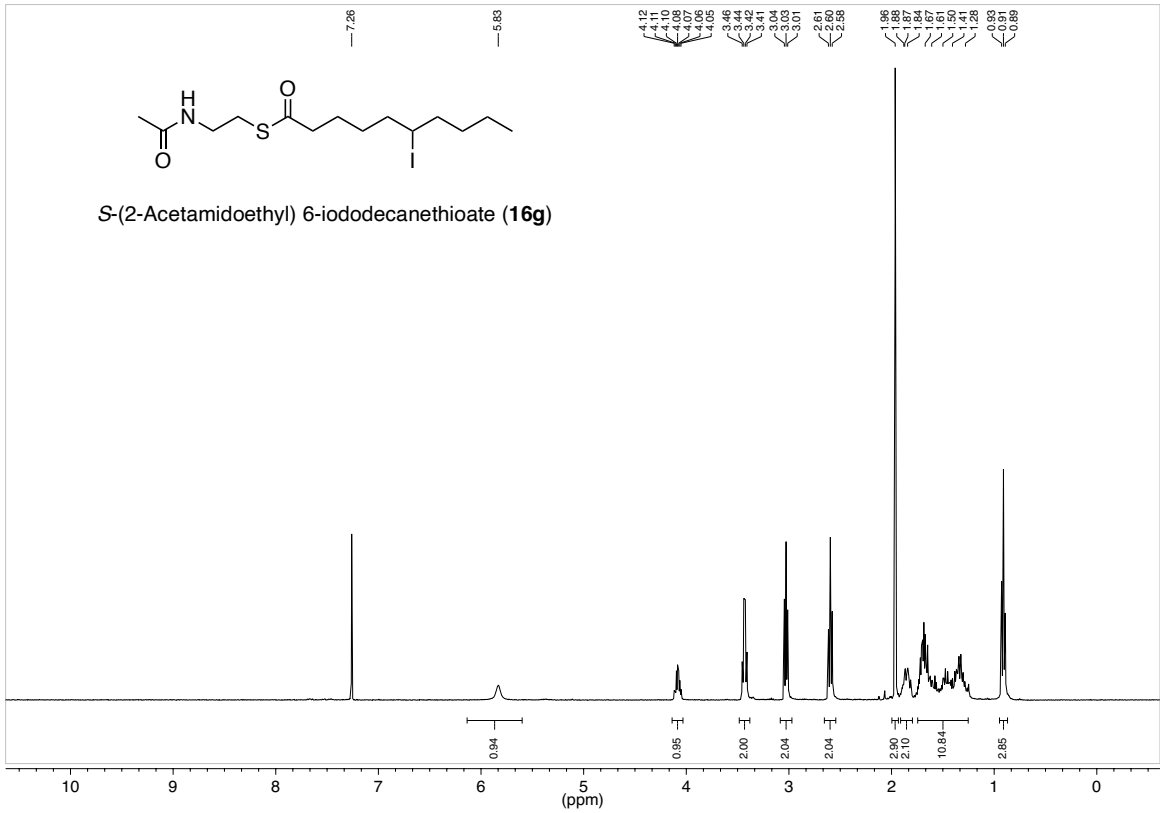


S-(2-Acetamidoethyl) 6-chlorodecanethioate (**16e**)

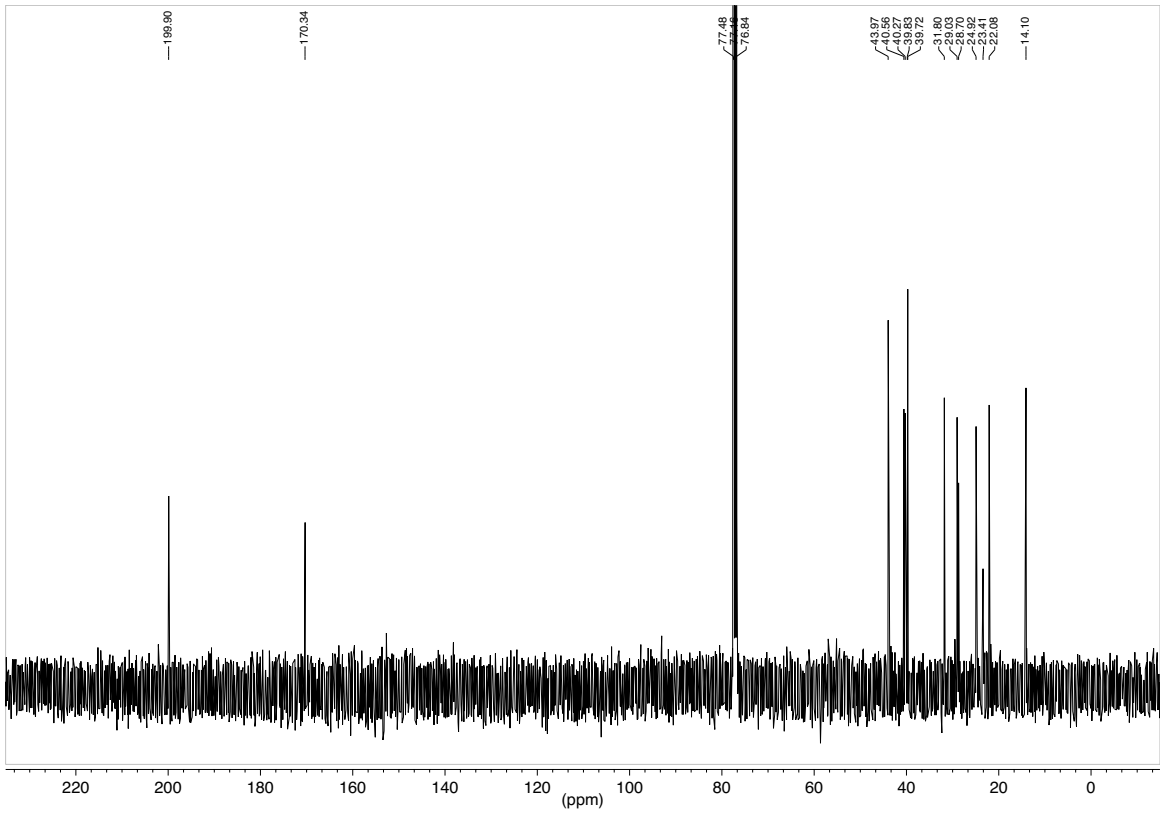


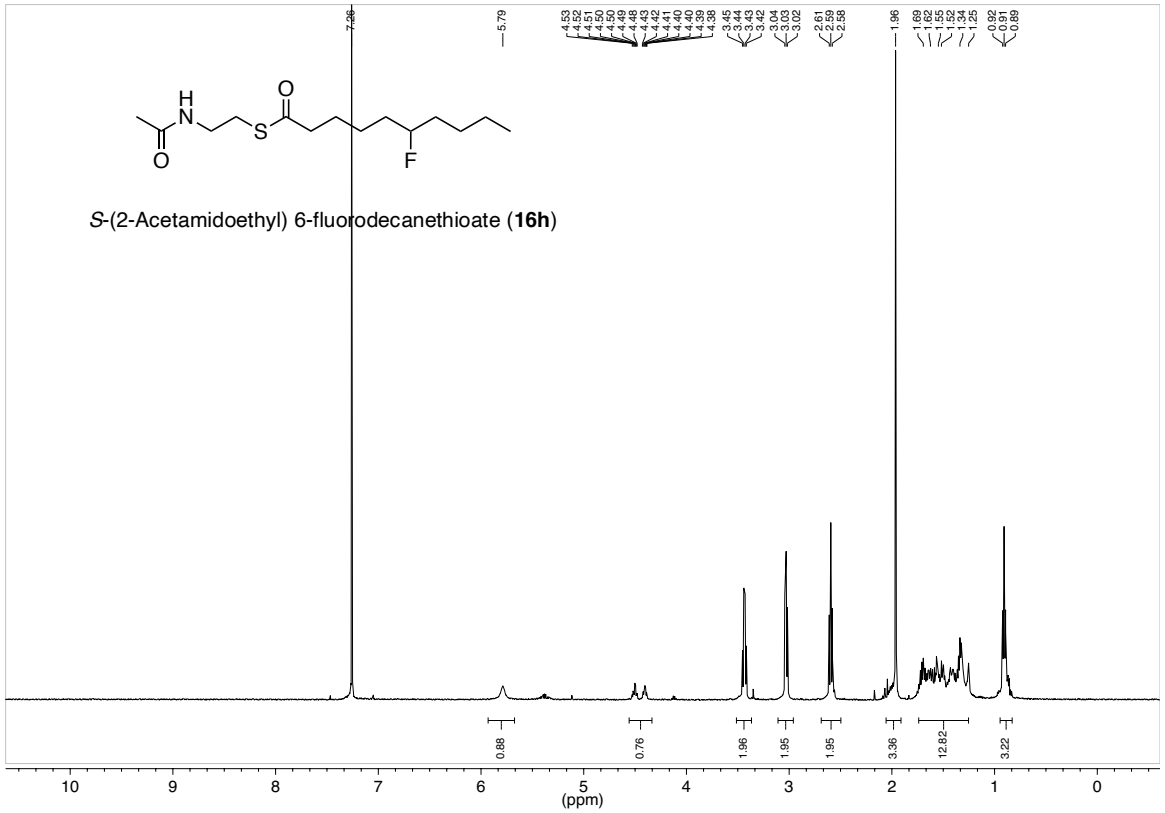

7.26
5.78
4.00
3.46
3.44
3.43
3.42
3.04
3.03
3.02
2.62
2.60
2.59
1.97
1.82
1.80
1.71
1.67
1.57
1.49
1.35
0.93
0.91
0.90
0.74
0.75
1.90
1.76
1.82
2.84
12.92
2.93
(ppm)


199.95
170.34
77.41
77.16
76.91
58.32
44.00
39.85
39.08
38.82
29.87
29.72
27.07
25.15
23.40
22.29
14.10
(ppm)

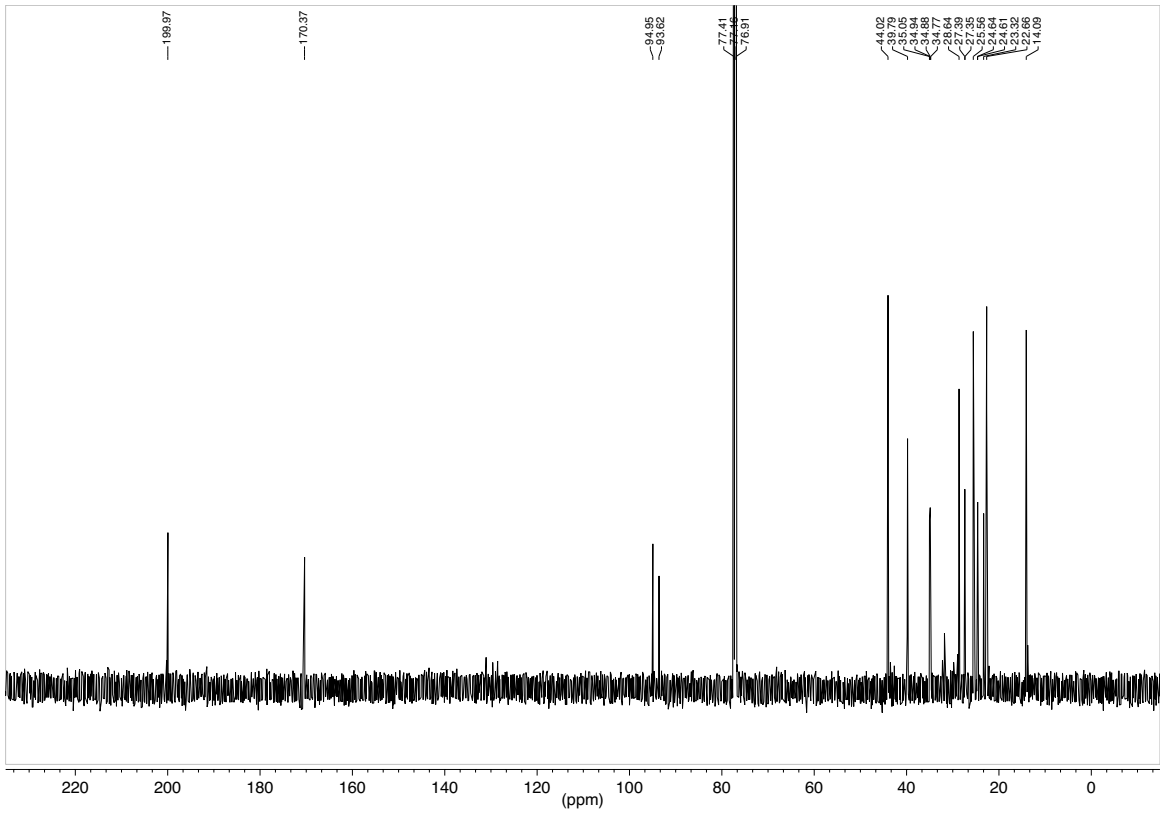


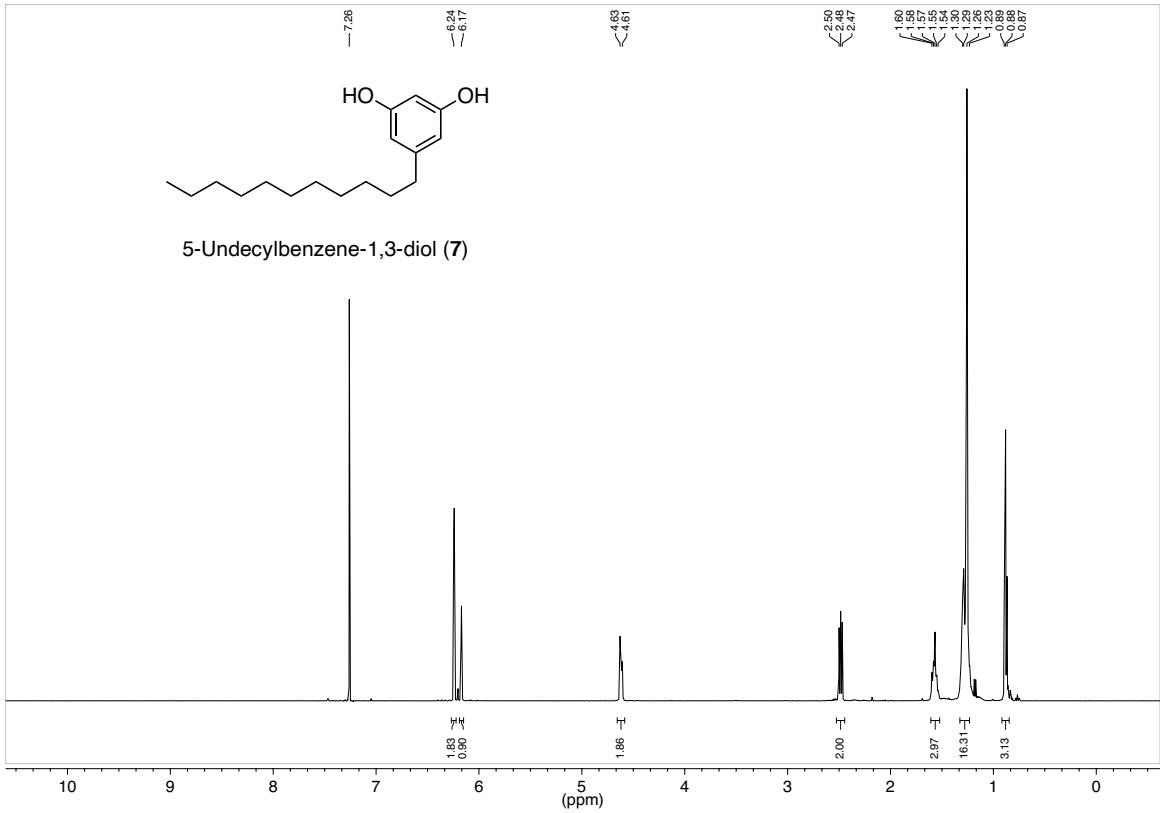
S-(2-Acetamidoethyl) 6-iododecanethioate (**16g**)



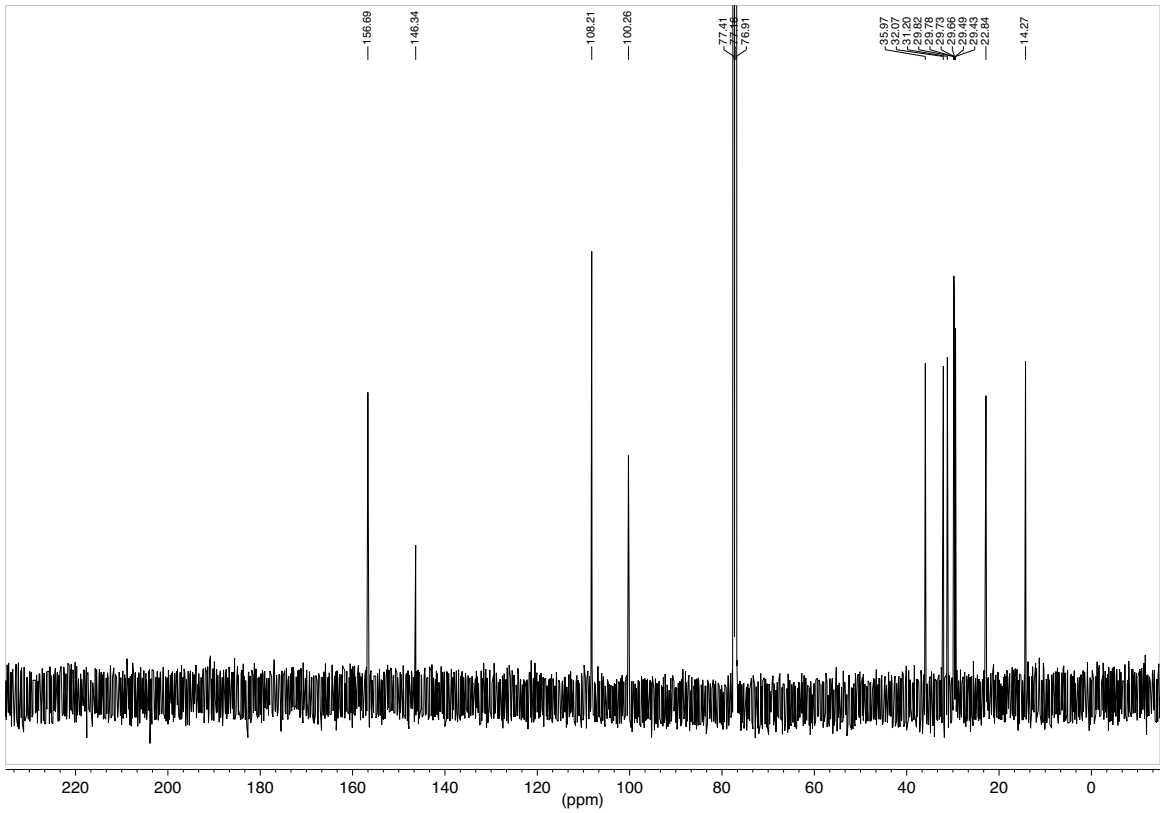


S-(2-Acetamidoethyl) 6-fluorodecanethioate (**16h**)





5-Undecylbenzene-1,3-diol (**7**)



3 Structural and mechanistic basis for the CylK-catalyzed Friedel–Crafts alkylation

Components of this chapter were adapted from previous publications.[128,162] The eLife article is distributed under the terms of a Creative Commons Attribution License that permits unrestricted use and redistribution provided that the original author and source are credited.

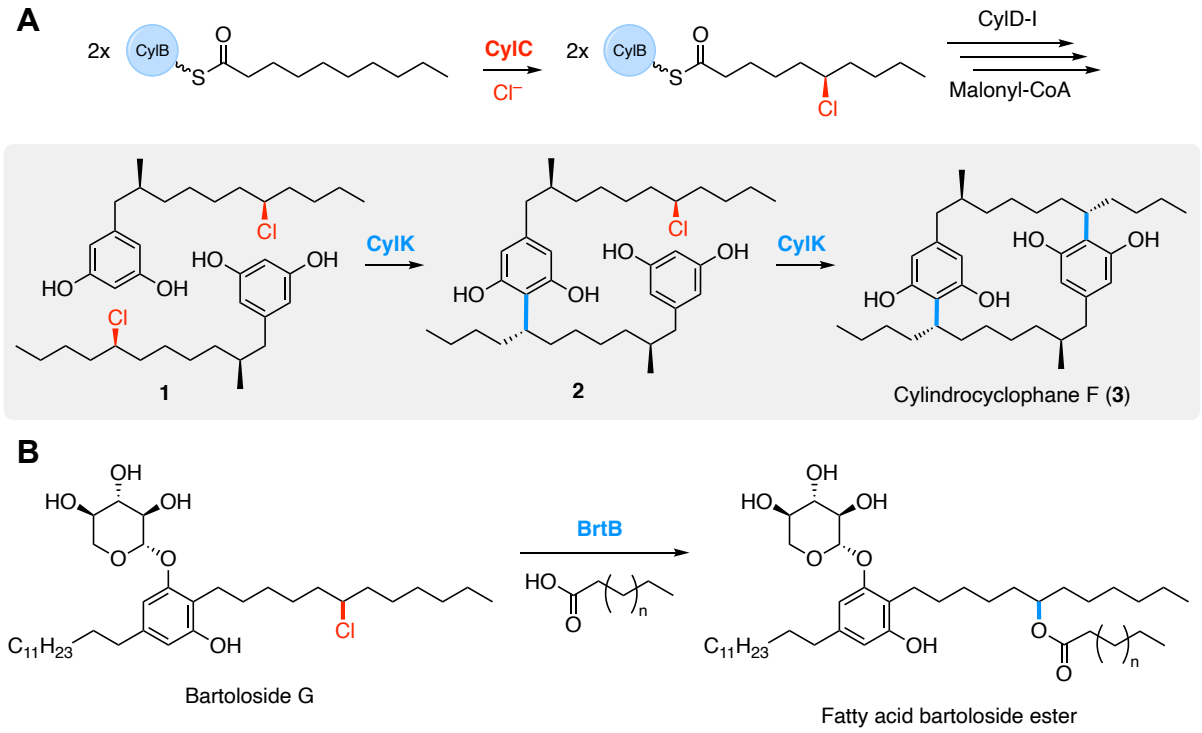
Portions of the work reported were conducted by other lab members and collaborators including Terry B. Ruskoski, Dr. Katherine M. Davis, Dr. Nathaniel R. Glasser, Dr. C. Denise Okafor, and Cassidy Johnson. Their contributions will be recognized as footnotes and citations throughout the chapter.

3.1 Introduction

The cyanobacterial enzyme CylK assembles the cylindrocyclophane natural products by performing two unusual alkylation reactions, forming new carbon–carbon (C–C) bonds between resorcinol aromatic rings and secondary alkyl chloride electrophiles. This transformation is unprecedented in biology and the structure and mechanism of CylK were unknown at the outset of this thesis research. This is one of a few known examples of 'cryptic halogenation', a biosynthetic logic that requires two partner enzymes: a halogenase, which introduces the alkyl halide substituent, and a second enzyme that utilizes the halogenated intermediate as a substrate, leveraging its increased reactivity for bond formation and releasing the corresponding halide anion as a side product (see Chapter 1.2 for more discussion).[34] Multiple classes of halogenating enzymes have been structurally and mechanistically characterized,[46] but the cognate halide-utilizing enzymes in pathways that employ cryptic halogenation remain underexplored. In particular, the structures and biochemical mechanisms of enzymes that engage and activate alkyl halide substrates as intermediates in natural product biosynthesis are unknown. In gaining a molecular understanding of the unusual biosynthetic transformation catalyzed by CylK, this work fills a gap in our understanding of how alkyl halides are activated and used by enzymes as biosynthetic intermediates, informing enzyme engineering, catalyst design, and natural product discovery.

Interrogating cylindrocyclophane biosynthesis by the cyanobacterium *Cylindrospermum licheniforme* presents an exciting opportunity to study alkyl halide utilization in biological systems (see Chapter 1.3 for more discussion). In this organism, the putative diiron carboxylate halogenase CylC generates an alkyl chloride intermediate which is further elaborated to produce resorcinol-containing alkyl chloride **1**. Two

molecules of **1** are then dimerized by an alkyl chloride-utilizing enzyme, CylK, via the formation of two new C–C bonds to construct a [7.7]paracyclophane ring system (Scheme 3.1A).[69] CylK is annotated as a fusion of a Ca^{2+} binding domain and a β-propeller fold, but the roles of these predicted protein domains in catalysis were unknown at the outset of this work. Initial biochemical studies revealed that this transformation involves two stereoselective alkylation events that occur in a stepwise fashion, with inversion of configuration at the alkyl chloride stereocenter.[69] CylK is the only enzyme known to catalyze aromatic ring alkylation with an alkyl halide electrophile, a reaction that mirrors a classical non-enzymatic reaction known as the Friedel–Crafts alkylation, which is an important transformation in organic synthesis. The traditional non-enzymatic Friedel–Crafts reaction suffers from a lack of stereo- and regiocontrol, unwanted overalkylation events, and the generation of carbocationic intermediates that can undergo unproductive rearrangements.[84,85] In contrast, the CylK-catalyzed Friedel–Crafts alkylation overcomes these limitations, likely by tightly controlling reactivity within the enzyme active site. This enzymatic transformation holds promise for biocatalytic applications, and we have demonstrated CylK's ability to accept multiple discrete resorcinol nucleophiles and alkyl halide electrophiles with varying substitution and electronic character (see Chapter 2 and Schultz, Braffman et al. for this work).[128] Moreover, uncharacterized enzymes with homology to CylK are found in numerous cyanobacteria, one of which (BrtB) was recently revealed to catalyze carbon–oxygen (C–O) bond formation between the carboxylate groups of fatty acids and the bartolosides, alkyl chloride-containing natural products (Scheme 3.1B).[156] This suggests that the mechanism by which CylK binds and activates alkyl chlorides might be shared with BrtB and other CylK homologs that may use different nucleophilic substrates.



Scheme 3.1. CylK and related enzymes use alkyl chloride substrates as biosynthetic intermediates

(*A*) The halogenase CylC generates an alkyl chloride substrate for CylK, which catalyzes two stepwise Friedel–Crafts alkylations to construct a paracyclophane macrocycle in cylindrocyclophane biosynthesis. This involves an intermolecular reaction between two resorcinol-containing alkyl chloride substrates (**1**) to generate intermediate **2**, followed by an intramolecular alkylation to afford cylindrocyclophane F (**3**). (*B*) The related enzyme BrtB (30% amino acid identity, 46% similarity) catalyzes an analogous but chemically distinct C–O bond-forming event between chlorinated bartolosides and fatty acid nucleophiles.

CylK and its homologs are not closely related to any enzymes that have been structurally or mechanistically characterized, and therefore key details about this enzyme family such as active site location, essential catalytic residues, cofactor or metal dependence, and roles for both N-terminal Ca^{2+} binding and C-terminal β-propeller domains were unknown at the outset of this thesis research. At the time of initial enzyme discovery, Nakamura et al. demonstrated that CylK's activity was dependent on calcium.[69] They first noticed that upon treatment with EDTA (ethylenediaminetetraacetic acid), CylK was rendered inactive. Furthermore, size exclusion chromatography demonstrated a shift in CylK's oligomeric state from monomeric to dimeric. EDTA is a common water-soluble chelating agent that is widely used to sequester alkaline-earth and other metal ions in solution.[163] In this way, ions such as Ca^{2+} and Mg^{2+} that were

bioinformatically predicted to be bound by CylK could be removed from this enzyme. The loss of CylK activity and oligomeric state change upon EDTA treatment suggested that metal ions had been removed. Notably, CylK activity could be restored upon additional supplementation with $CaCl_2$ while other metals including iron(II), manganese, zinc, and magnesium failed to rescue CylK's activity. In fact, Zn^{2+} and to a lesser extent Fe^{2+}, in combination with Ca^{2+} was found to moderately inhibit CylK activity. This might be due to competitive binding in the enzyme's metal ion binding sites. Although this work clearly demonstrated the importance of Ca^{2+}, it was unclear if this ion played a direct role in catalysis and/or provided structural integrity to the overall enzyme. Furthermore, it was unknown if EDTA removed all metal equivalents because while the N-terminal domain of CylK is predicted to bind calcium, the C-terminal domain of CylK has homology to monodomain β-propeller enzyme virginiamycin B lyase (Vgb) that binds an essential magnesium ion.[164,165] Vgb confers resistance to type B streptogramin antibiotics by catalyzing an elimination reaction that results in antibiotic macrocycle linearization. It is proposed that the initial deprotonation of this elimination is facilitated by the reduced pK_a of the Mg^{2+}-substrate complex. Structures demonstrate that Mg^{2+} interacts with Glu residues of Vgb, and a phenol and amide carbonyl of the substrate. With this in mind, we wondered if CylK could similarly use calcium or magnesium ions to coordinate resorcinol nucleophiles or stabilize chloride leaving groups.

In addition to the lack of structural and mechanistic information about CylK itself, there are few well-characterized examples of enzymes that engage and activate alkyl halide substrates. In one notable case, the bacteria *Xanthobacter autotrophicus* can grow on media containing 1,2-dichloroethane as a sole carbon source.[166] In order to catabolize this anthropogenic compound, *X. autotrophicus* has evolved an interesting enzymatic strategy to break two C–Cl bonds (Figure 3.1A). The first step is catalyzed by haloalkane dehalogenase (DhlA) which hydrolyzes dichloroethane to produce 2-chloroethanol.[167] This is followed by oxidation to chloroacetaldehye by a periplasmic PQQ-dependent alcohol dehydrogenase, further oxidation to chloroacetate by an NAD-dependent aldehyde dehydrogenase, and a final dehalogenation reaction catalyzed by haloacetate dehalogenase (DhlB) to afford 2-hydroxyacetic acid (glycolic acid). The first step of this pathway, catalyzed by haloalkane dehalogenase, highlights a strategy by which primary alkyl

chlorides can be activated for substitution (Figure 3.1B).[167] First, dichloroethane is bound in the active site by two Phe and two Trp residues with apparent alkyl halide–aryl and hydrogen bonding interactions. Then, the alkyl chloride is displaced by the carboxylate side chain of Asp124 in a S_N2 substitution reaction. It is proposed that hydrogen bond contacts between the labile alkyl chloride and two indole N–H bonds activate the C–Cl bond for substitution by stabilizing the partial negative charge that develops on the chloride in the transition state. Structures of this enzyme have been solved with both bound starting material (PDB 2DHC) and the covalent enzyme–alkyl intermediate (PDB 2DHD), and it is worth noting that Phe172 is positioned to potentially assist in binding and stabilizing the labile chloride with an alkyl halide–aryl interaction (see Figure 3.25 for an illustration of this interaction). Finally, the covalent intermediate is hydrolyzed by a water molecule activated by His289, to afford the final product 2-chloroethanol. Interestingly, mutating His289 arrests the catalytic cycle at the enzyme–alkyl intermediate[168] and this mutant serves as the basis of the HaloTag chemical tool used for protein and cell labeling.[169]

Although the reaction catalyzed by CylK is distinct from that of haloalkane dehalogenase in a number of ways, including its use of a resorcinol nucleophile and secondary alkyl halide electrophile substrates, we wondered its mode of alkyl halide activation might be analogous to that of haloalkane dehalogenase. Notably, CylK's inversion of configuration at the alkyl chloride stereocenter suggests than an enzyme–substrate intermediate is unlikely. However, the carboxylate nucleophiles of BrtB do resemble haloalkane dehalogenase's Asp124 nucleophile, and similar strategies of alkyl chloride activation may be operative in both BrtB and CylK. As we began our work to structurally and mechanistically characterize CylK, we hypothesized that similar alkyl chloride–enzyme hydrogen bonds may be important for catalysis, recognizing that secondary alkyl halides are less reactive than primary and may require a higher degree of activation.

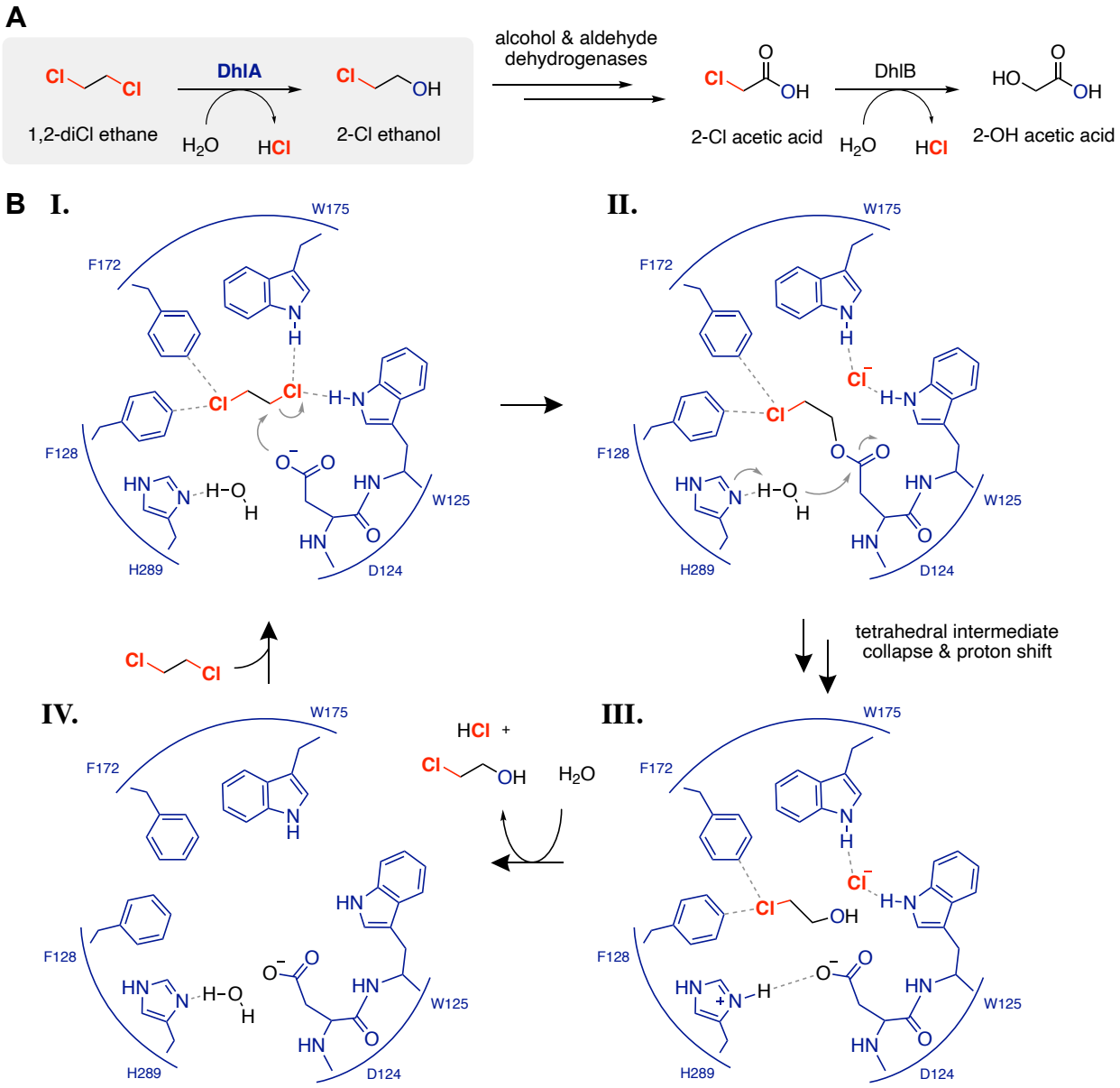


Figure 3.1. Dichloroethane catabolism and mechanism of haloalkane dehalogenase

(*A*) Dehalogenation steps of 1,2-dichloroethane by *X. autotrophicus* to produce 2-hydroxyacetic acid. (*B*) Proposed mechanism of haloalkane dehalogenase (DhlA) invoking concerted nucleophilic attack and primary alkyl chloride activation (complex I), via the carboxylate of Asp124 and adjacent Trp indole N–H–Cl hydrogen bonds, respectively, followed by hydrolysis (complex II) to afford chloroethanol for subsequent oxidation and dehalogenation by DhlB.

Despite the intriguing reactivity and potential applications of CylK and related alkyl halide activating enzymes, our knowledge of its structure and mechanism was limited at the outset of this dissertation research. In this Chapter, we present a crystal structure of CylK, identify critical active site residues, provide experimental and bioinformatic support for their roles in catalysis, and propose a mechanism by which alkyl chloride substrates are activated for stereospecific nucleophilic substitution. In determining how CylK performs this Friedel–Crafts alkylation, we have enhanced our fundamental understanding of how enzymes engage alkyl halide substrates. Our structural information and mechanistic model will guide future enzyme engineering efforts, inform the design of non-enzymatic catalysts, and enable genome mining to uncover new natural products constructed by related biosynthetic strategies.

3.2 Results and Discussion

3.2.1 Density functional theory modeling and mechanistic speculation

Based on our evaluation of CylK's substrate scope in Chapter 2, we recognized that this enzyme tolerates a variety of functional groups appended to the resorcinol nucleophile substrate, but it strictly requires two free phenols with the 1,3 relationship that defines resorcinol rings. Substrates in which one or both of these phenols were converted to methoxy groups were not accepted as competent nucleophiles. With this enzyme specificity in mind, we employed density functional theory (DFT) calculations to investigate the inherent nucleophilicity of the resorcinol ring in order to gain insight about what, if any, enzymatic activation by CylK might be required.

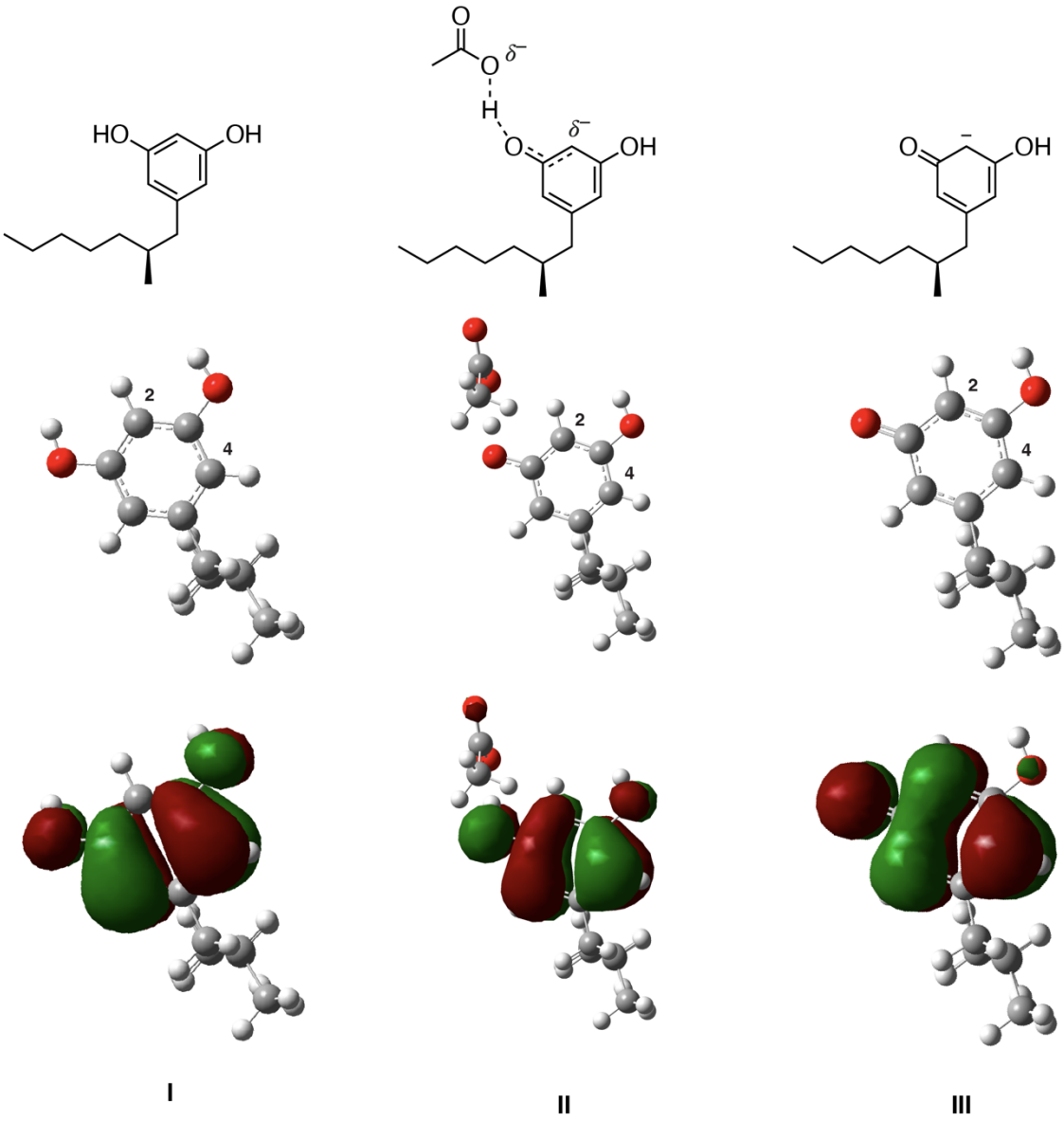
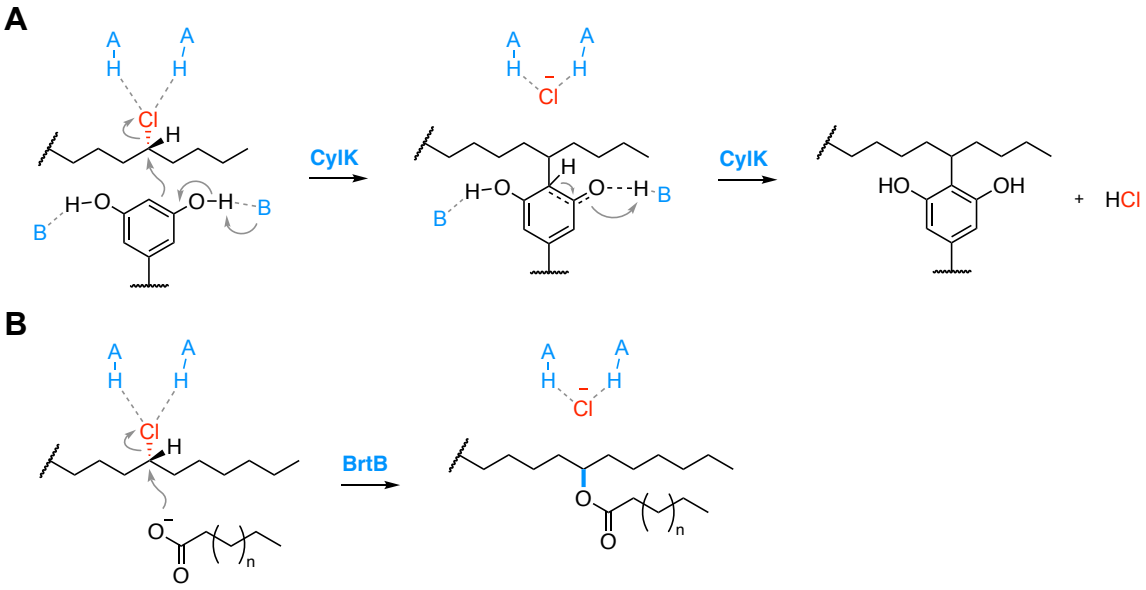


Figure 3.2. Electronic considerations of resorcinol nucleophile activation

Drawn, ball and stick, and highest occupied molecular orbital (HOMO) representations of DFT calculated ground states of 5-(heptan-2-yl)-resorcinol (I), hydrogen bonding to acetate (II), and deprotonated (III). HOMO electron density is present at the 2-position only when hydrogen bonded to acetate (II) or deprotonated (III). DFT parameters were B3LYP functional, 6-31G(d) basis set, D3BJ4 correction, and polarizable continuum model for Et_2O

We found that a truncated alkyl resorcinol (5-(heptan-2-yl)-resorcinol) (I) resembling the native substrate contained a node at the 2-position of the ring (Figure 3.2). This suggests that this carbon is not nucleophilic, and that resorcinol substrates must be activated in some manner to place electron density at this carbon. We then found that upon hydrogen bonding (II) to acetate, or full deprotonation (III) of a phenol substituent,

the highest occupied molecular orbital (HOMO) of this alkyl resorcinol shifts to place electron density at the 2-position. This result suggests that CylK might alter the nucleophilicity of its resorcinol substrate by partial or full deprotonation of one of its phenolic groups. Furthermore, considering the relative stability of the alkyl chloride electrophile, we began to envision a generic catalytic mechanism in which CylK activates both substrates with a network of hydrogen bonds for a concerted (S_N2-like) or stepwise (S_N1-like) substitution reaction (Scheme 3.2A).



Scheme 3.2. Putative mechanistic similarities between CylK and BrtB

(*A*) Generic concerted mechanism of CylK catalyzed alkylation with hypothetical network of hydrogen bond donors (A) and acceptors (B). (*B*) Analogous proposal for the BrtB catalyzed O-alkylation. The alkyl chloride electrophile is displayed in red, while the halide-utilizing enzyme and corresponding bond formed are represented in blue.

This may also suggest that the mechanism by which CylK binds and activates alkyl chlorides is shared with BrtB. This homolog of CylK has a similar di-domain structure and 30% amino acid sequence identity and 46% similarity to CylK. However, BrtB catalyzes a distinct reaction between fatty acid nucleophiles and alkyl chloride bartoloside electrophiles (see Scheme 2.4 for more discussion). At physiological pH, carboxylate

substrates will be predominately deprotonated, indicating that BrtB might not need to activate these nucleophiles for catalysis. In considering these potential differences, we were further motivated to solve the structure of CylK and to determine the location of its active site and key catalytic residues. If shared, the enzymatic strategy by which CylK and BrtB activate alkyl chloride electrophiles might be operative in other family members that may utilize additional nucleophilic substrates.

3.2.2 Structural determination effort

With our optimized CylK preparation and several substrate analogs in hand, we initiated a collaboration with the Amie K. Boal Group at Penn State University in order to solve a crystal structure of CylK.[xi] Early work demonstrated that CylK crystallized in the presence of undecylresorcinol (**7**), and after subsequent optimization and X-ray data collection it was revealed that this crystal form was in the $C222_1$ space group and there was a single monomer of CylK in the asymmetric unit. Suitable molecular replacement (MR) models for phasing could not be identified because CylK is not closely related to structurally characterized enzymes. Instead, we refolded CylK without the Ca^{2+} and Mg^{2+} salts (apoCylK) customarily present in our buffers, and then we attempted to substitute potential metal ion binding sites with ytterbium (Yb), thulium (Tm), and terbium (Tb) chloride. A competitive reconstitution experiment with all three lanthanides revealed that CylK had a slight preference for Tb (Table 3.1). Of note, this work and follow up activity assays confirmed that apoCylK was dimeric and inactive, and that its activity could be rescued by the addition of 5 mM $CaCl_2$, as first observed by Nakamura et al.[69] Magnesium could not be excluded from our CylK preparations, likely due to buffer contamination.

[xi] Structural work was performed by Terry B. Ruskoski, with contributions from Dr. Katherine M. Davis.

Table 3.1. ICP-OES analysis of lanthanide reconstituted CylK

Sample	Ca (ppb / **equiv.**)	Mg	Yb	Tm	Tb
buffer blank	-0.2 / **0.0**	15.2 / **2.7**	0.0 / **0.0**	0.3 / **0.0**	0.7 / **0.0**
apoCylK	3.7 / **0.2**	22.5 / **1.7**	0.1 / **0.0**	0.2 / **0.0**	0.8 / **0.0**
apoCylK, $YbCl_3$, $TmCl_3$, $TbCl_3$	1.5 / **0.2**	21.0 / **3.7**	22.3 / **0.6**	24.4 / **0.7**	28.3 / **0.8**

Metal abundance was quantified by ICP-OES at the Quantitative Bio-element Imaging Center at Northwestern University. Following treatment with a mixture of lanthanides, apoCylK was buffer exchanged with iterative centrifugal concentrations to remove unbound metals.

The Boal Group then obtained phase information by soaking CylK crystals in a solution containing 100 mM $TbCl_3$, followed by collection of X-ray diffraction datasets at the Tb X-ray absorption peak energy. The structure of CylK was solved to 1.68 Å resolution via a combined MR and single-wavelength anomalous diffraction (SAD) approach, with a poly-alanine 7-bladed β-propeller MR search model.[xii] This structure reveals an unprecedented fusion of two protein folds, an N-terminal Ca^{2+} binding domain and a C-terminal β-propeller domain that have not been found fused together in other proteins (Figure 3.3).

The N-terminal Ca^{2+} binding domain of CylK contains a short helical component that packs against a β-roll core with at least 4 calcium ions bound in hexacoordinate sites comprised of four backbone carbonyls and two Asp carboxylates. (Figure 3.4). This core fold is similar to repeat-in-toxin (RTX) motifs, found in bacterial extracellular proteins.[170] Structural homologs of this domain include the C-terminal RTX domains of secreted toxins[171] and bacterial surface-layer proteins (see Table 3.5 in Materials and Methods).[172] In these systems, this domain facilitates Ca^{2+}-dependent folding or assembly in the calcium-rich environment outside of the cell. Enzyme with structural homology to the N-terminal Ca^{2+} binding domain of CylK include extracellular proteases, hydrolases, and epimerases from gram-negative bacteria.[173-175] The RTX motif in these enzymes can be fused both N- and C-terminally to their catalytic domains. It is believed that the RTX domain facilitates extracellular secretion and Ca^{2+}-dependent folding. In one representative instance, this RTX motif is fused C-terminally to the catalytic domain of zinc-dependent protease PAP and a structural

[xii] Alanine β-propeller model was generated by Dr. Nathaniel R. Glasser.

comparison to CylK reveals how the RTX motif can be functionally modular with no direct connection to the active site, as well as the differences in the Ca^{2+} binding motif itself. (Figure 3.5).[176] The N-terminal domain of CylK differs from typical RTX motifs in that it is more asymmetric and only contains one consensus Ca^{2+} binding motif. Notably, the section of this β-roll with Ca^{2+} sites packs closely into the center of the adjacent C-terminal domain of CylK, unlike PAP in which this motif is on the protein surface.

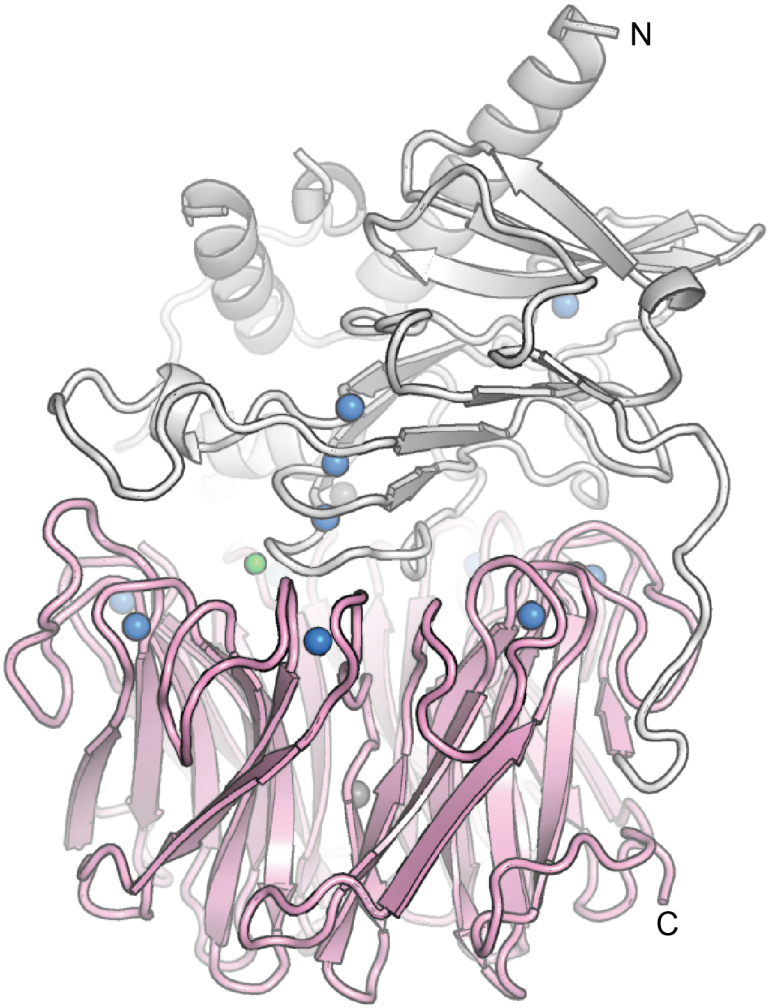


Figure 3.3. The CylK X-ray crystal structure

An overall view of the structure of CylK (PDB 7RON). In this image, calcium ions are shown as blue spheres, magnesium ions as dark grey spheres, and chloride ions as green spheres. The N-terminal Ca^{2+} binding domain is depicted as a light grey ribbon diagram and the C-terminal β-propeller domain is shown as a pink ribbon diagram. This figure was adapted from images generated by Terry B. Ruskoski.

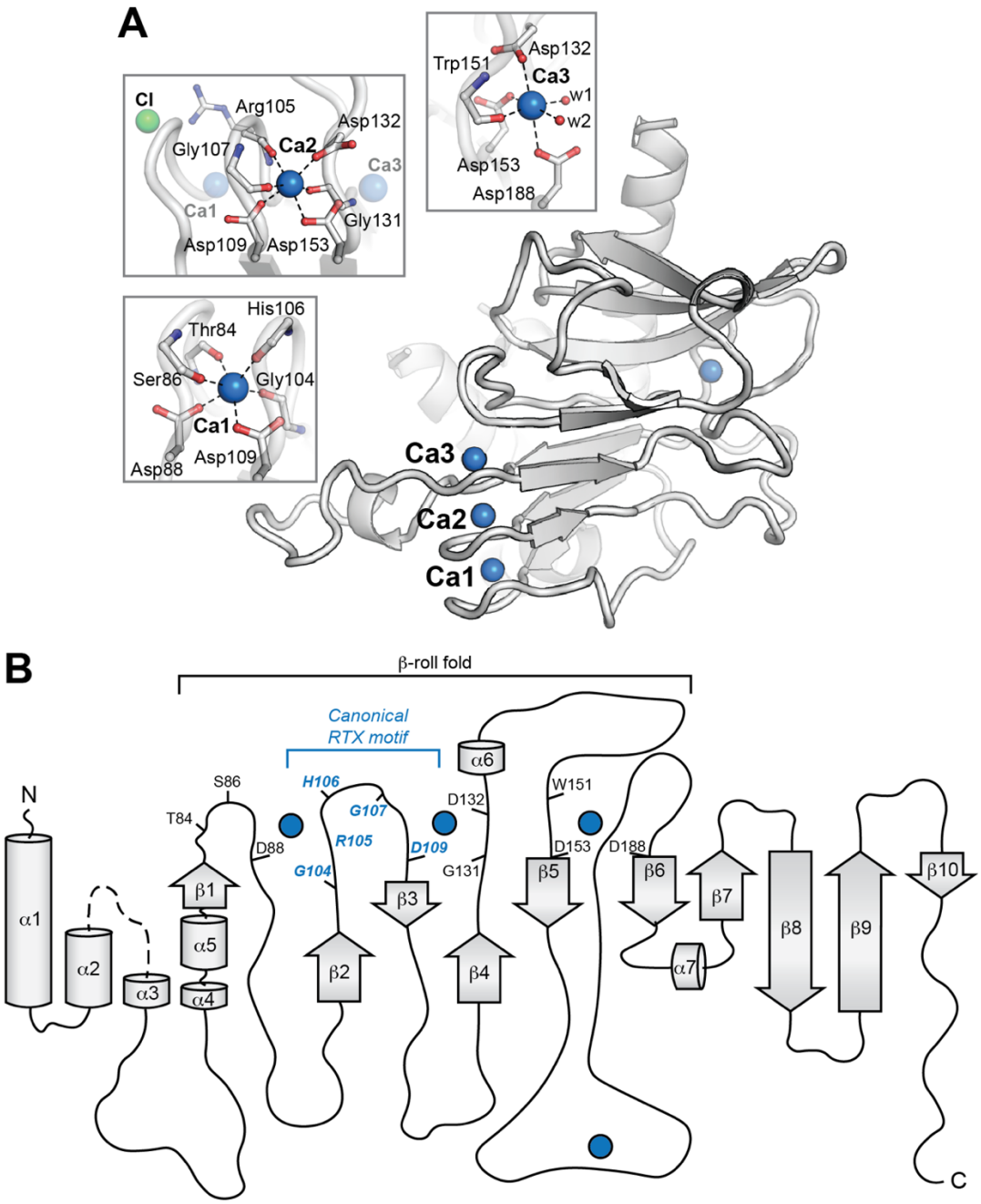


Figure 3.4. N-terminal domain architecture of CylK

(*A*) CylK's N-terminal domain contains a right-handed parallel β-roll stabilized by three Ca^{2+} ions capped with additional secondary structures (residues 7-251). (*B*) Topology diagram of N-terminal domain. Calcium ions are shown as blue spheres/circles, chloride ions as green spheres, and Ca^{2+}-coordinating residues that follow canonical repeat-in-toxin (RTX) motifs are shown in blue italics. This figure was adapted from images generated by Terry B. Ruskoski.

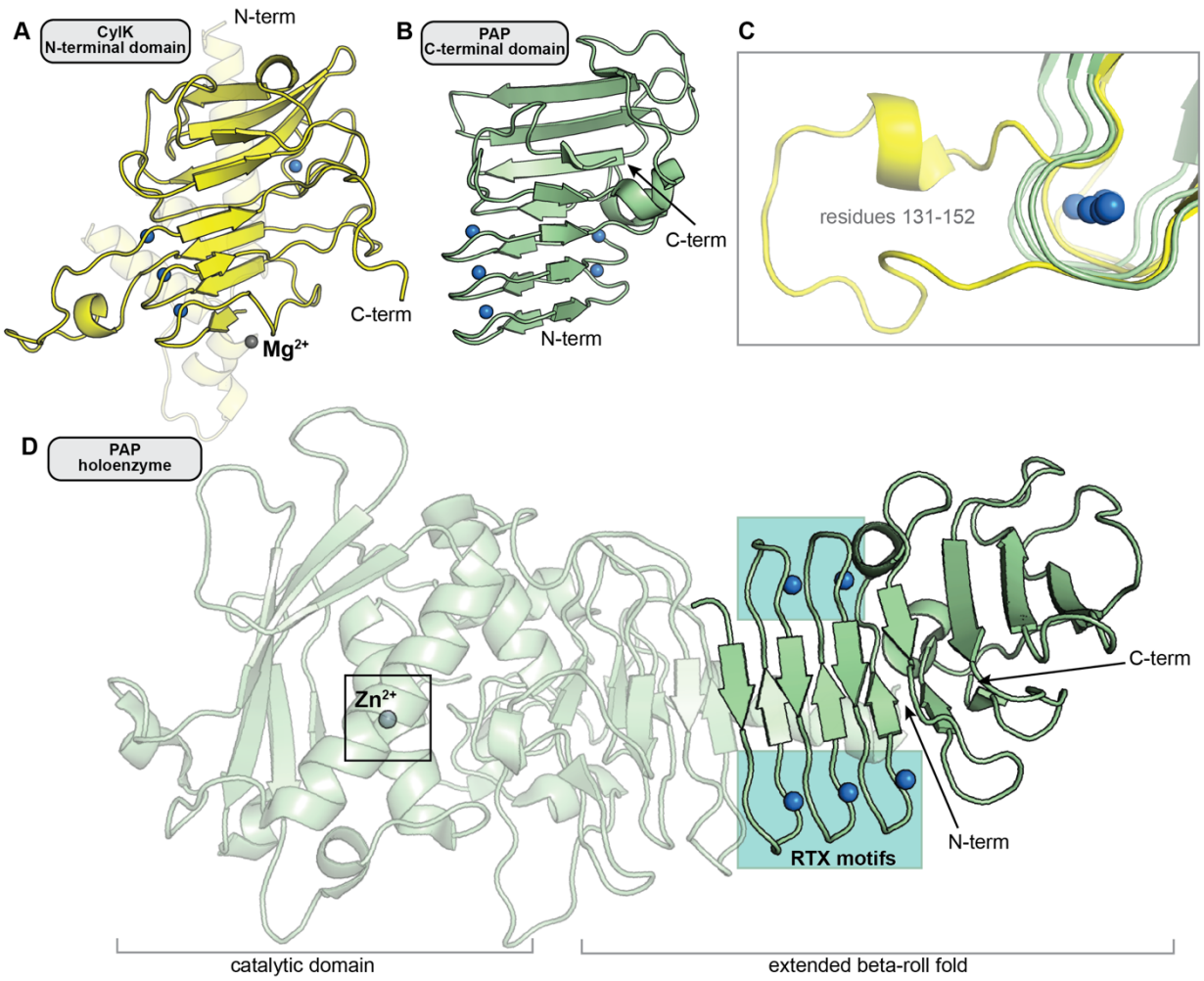


Figure 3.5. Structural comparison of N-terminal Ca^{2+} binding domain of CylK to PAP protease

(A) The N-terminal domain of CylK depicted in yellow (residues 7-251). *(B)* The C-terminal RTX motif of zinc-dependent protease PAP depicted in green (residues 326-461, PDB 1OMJ) that structurally aligns with the N-terminal domain of CylK (PDB 1OMJ). *(C)* CylK Ca^{2+}-coordinating loops (yellow) extend beyond the tight RTX motif turn of PAP (green). *(D)* Domain arrangement of PAP suggests RTX motif is not directly involved in catalysis, zinc ion of active site is highlighted (box). Calcium ions are shown as blue spheres, magnesium and zinc ions as dark grey spheres. This figure was generated by Terry B. Ruskoski.

The C-terminal domain of CylK adopts a 7-bladed β-propeller fold (Figure 3.6). The loops that connect subsequent blades contain unique blade-bridging Ca^{2+} binding sites comprised of backbone carbonyls and carboxylate side chains, suggesting a shared role for this divalent metal in structural stabilization of both domains of CylK. The β-propeller is a widespread structural motif is found in both eukaryotic and prokaryotic protein structures,[177] and proteins with homology to this domain are overrepresented in BLAST searches of CylK and BrtB. In eukaryotes, the β-propeller fold facilitates protein-protein interactions[178] via its 'top face', the same interface that interacts with the N-terminal domain of CylK. In prokaryotes and some eukaryotes (plants), β-propeller folds serve as catalytic domains (see examples in Figure 3.7).[177] CylK is structurally similar to a class of monodomain bacterial β-propeller enzymes implicated in streptogramin antibiotic resistance (see Table 3.6 in Materials and Methods), including virginiamycin B lyase (Vgb).[164,165] However, the chemistry performed by these two enzymes differs. While CylK forms two C–C bonds to generate a cyclic product, the streptogramin resistance enzymes linearize cyclic peptide substrates by cleaving a C–O bond via an elimination reaction. X-ray crystal structures of an inactive variant of Vgb bound to a substrate analog provide insight into the location of the active site and mechanism of macrocycle opening (Figure 3.7B).[165] In Vgb, the substrate analog and a Mg^{2+} ion necessary for catalysis bind near conserved polar side chains in blades 6 and 7 of the propeller fold. In CylK, these sites are substituted for hydrophobic side chains, suggesting that, despite the structural similarities to streptogramin lyases, CylK engage substrates in a different location.

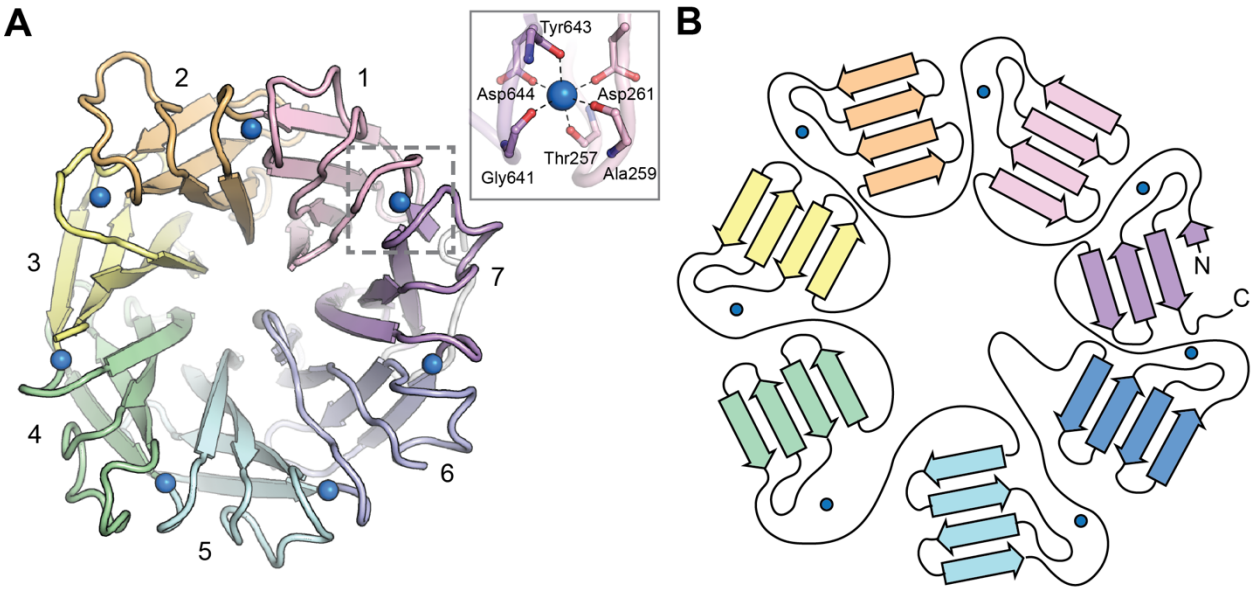


Figure 3.6. C-terminal domain architecture of CylK

(*A*) A top-down view of the C-terminal 7-bladed β-propeller domain of CylK (residues 252-662). The propeller blades are numbered from the N-terminal end of the domain and colored by blade. The inset shows a representative view of the Ca^{2+} coordination environment. A similar binding site exists at each blade junction. (*B*) Topology diagram of the C-terminal domain, blades are colored as in (A). Calcium ions are shown as blue spheres/circles, magnesium ions as dark grey spheres. This figure was adapted from images generated by Terry B. Ruskoski.

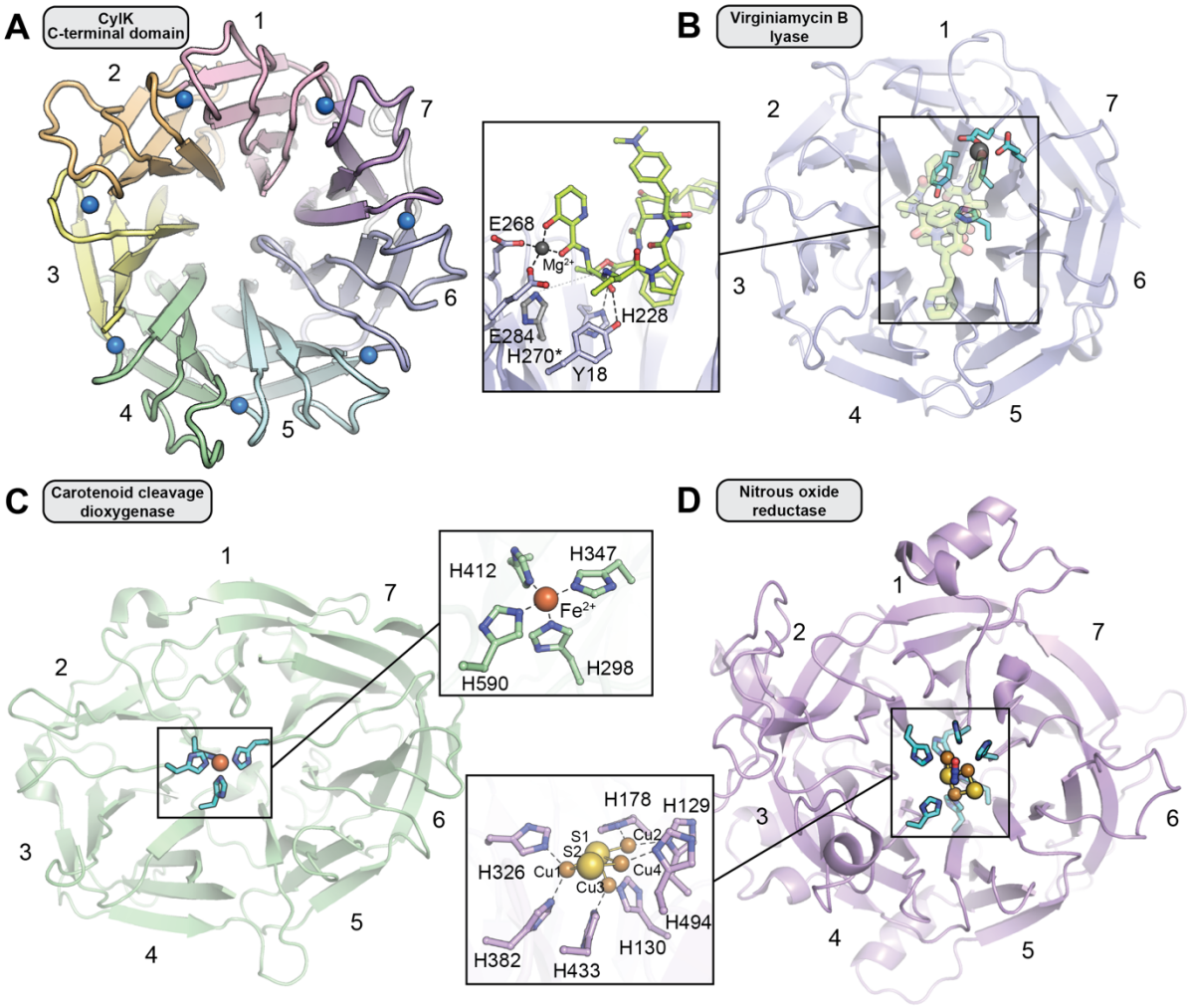


Figure 3.7. Comparison of CylK's C-terminal β-propeller domain with structurally related enzymes

Panels show a top-down ribbon diagram of β-propeller domains with residues implicated in catalysis colored cyan. *(A)* The C-terminal domain of CylK. Calcium ions are shown as blue spheres, magnesium ions as dark grey spheres. *(B)* Virginiamycin B lyase (VrgB) complexed with substrate analog quinupristin (lime green) (PDB 2Z2P, H270 from PDB 2Z2O). *(C)* Carotenoid cleavage dioxygenase (CCD) (PDB 3NPE). Ferrous ions are shown as orange spheres. *(D)* Nitrous oxide reductase (NOR) complexed with N_2O (blue and red) (PDB 3SBR) and the active site Cu-S cluster. Copper ions are shown orange spheres, and sulfur atoms are shown as yellow spheres. This figure was adapted from images generated by Terry B. Ruskoski.

3.2.3 Locating the CylK active site

Although CylK had been cocrystalized in the presence of undecylresorcinol (**7**), we could not accurately place one or two equivalents of this substrate analog in any of the remaining unmodeled and unaccounted for electron density (F_o–F_c). This led us to conclude that water, inorganic ions, and/or other buffer components might be responsible for this electron density. In an effort to locate the active site, we focused on two solvent accessible cavities (Figure 3.8A). Cavity 1 is lined by the top face of the C-terminal β-propeller domain and by the N-terminal Ca^{2+} binding domain, forming a tunnel that is ~18 Å deep and 15 Å wide at its largest point. Notably, the walls of this cavity contain both polar amino acid residues as well as hydrophobic patches (Figure 3.8B-C), ideal for interaction with—and activation of—amphipathic resorcinol-containing alkyl chloride substrates. In particular, we expected hydrophobic residues to assist in binding the substrates' alkyl substituents, while polar or charged residues might bind and activate the alkyl chloride and the resorcinol's phenolic groups. We also considered cavity 2, the central channel located exclusively within the C-terminal domain and opening to the opposite (bottom face) side of the propeller motif. Cavity 2 as the potential active site location would be analogous to other β-propeller-containing enzymes mentioned earlier (see examples in Figure 3.7).

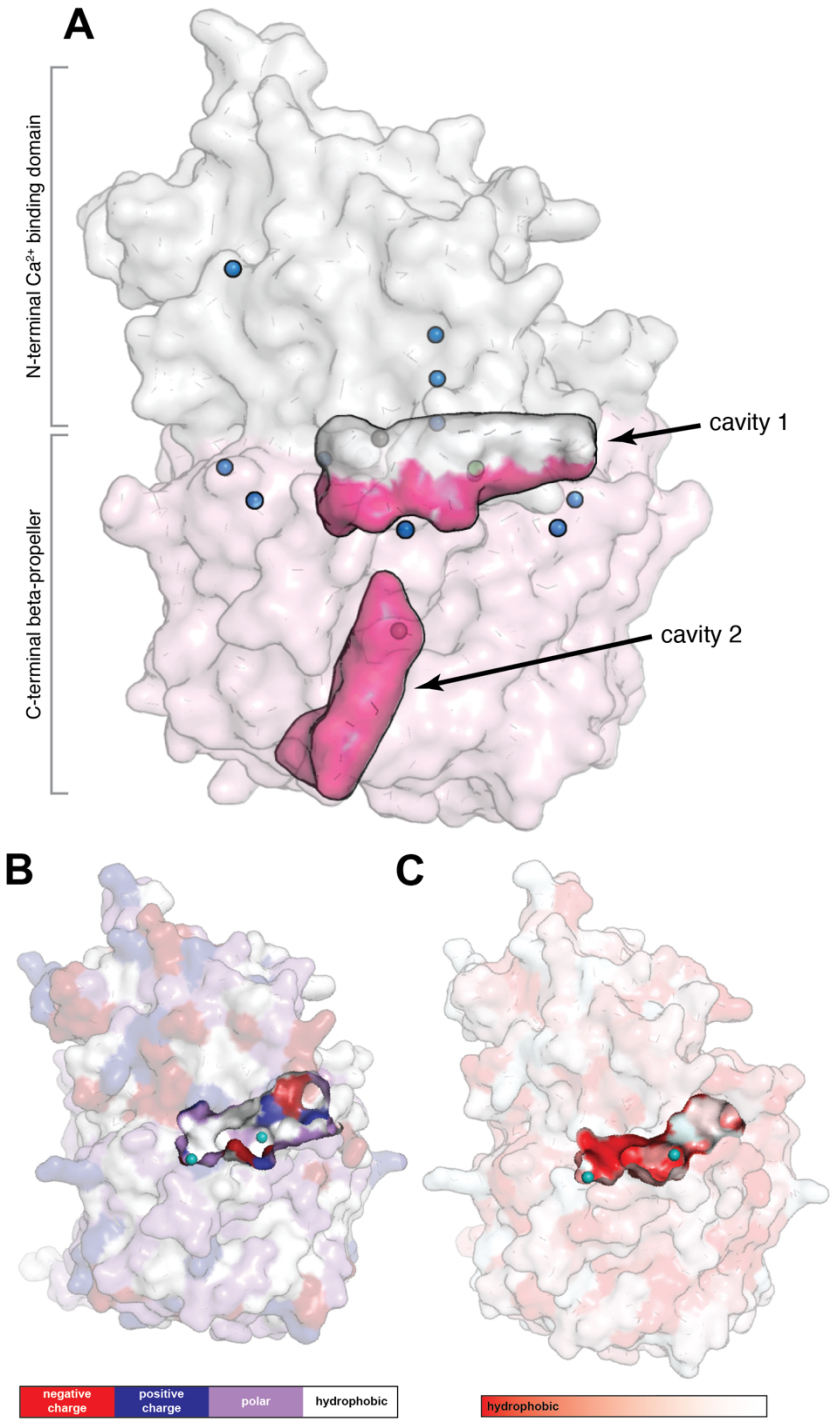


Figure 3.8. Potential active site cavities of CylK

(*A*) Cavity mapping analysis[179] of CylK reveals two solvent accessible cavities that may be large enough to accommodate substrates (PDB 7RON). (*B*) Side chain properties and (*C*) hydrophobicity of cavity 1, located at the junction between the N-terminal and C-terminal domains of CylK (PDB 7ROO). Calcium ions are shown as blue spheres, magnesium ions as dark grey spheres, and halide ions as green spheres. This figure was adapted from images generated by Terry B. Ruskoski.

To identify the active site, we individually mutagenized seventeen polar residues spanning both cavities 1 and 2 that are conserved in putative CylK homologs from genetically sequenced cyanobacteria known to produce structurally related natural products (Figure 3.9A-B & Table 3.2).[64,65,81,83] Not all of these residues are conserved in BrtB as we expected differences between their modes of catalysis, particularly their modes of nucleophile activation. To rapidly test variant proteins for activity, we designed a plate-based lysate activity assay using non-native substrate pair **7** + **16e** (Figure 3.9C). This alkyl resorcinol nucleophile and alkyl chloride electrophile are accepted by CylK (see Chapter 2).[128] All variants associated with the central/bottom channel of the C-terminal β-propeller retained full alkylation activity (Figure 3.9D). Strikingly, substitution of two residues located at the domain interface between the top face of the β-propeller and the N-terminal domain, Arg105 and Tyr473, completely abolished all activity on substrate pair **7** + **16e** (Figure 3.9D). Notably, although located in close proximity, Arg105 is supplied by the N-terminal domain, while Tyr473 is from the C-terminal β-propeller, suggesting that both domains play an essential role in catalysis. These data preliminarily supported the location of the active site as the cleft formed between the two domains of CylK (cavity 1).

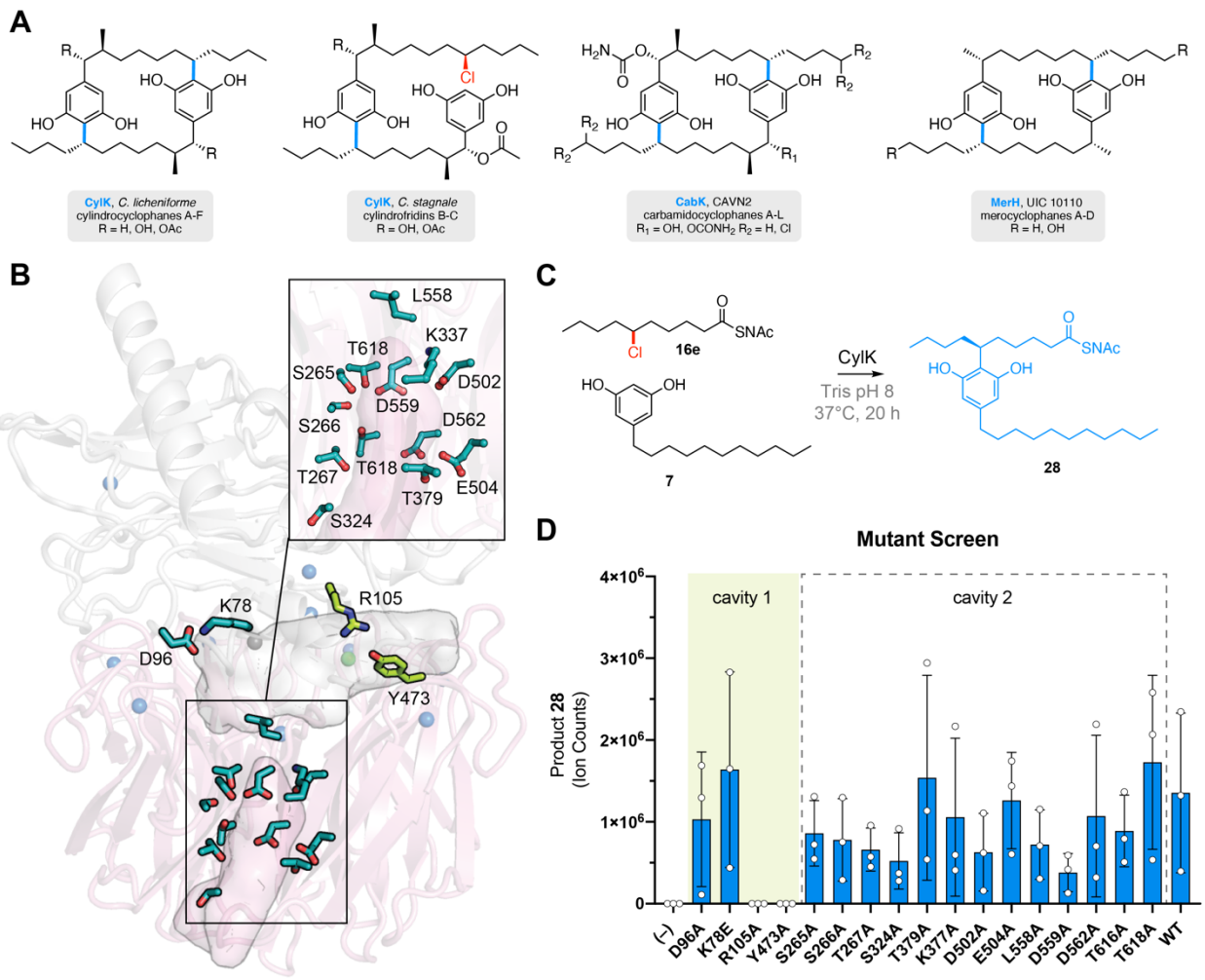


Figure 3.9. Mutant screen to identify active site at the domain interface

(*A*) Structurally related natural products produced by cyanobacteria with related biosynthetic gene clusters and CylK homologs. The bonds highlighted in blue are likely constructed by their respective CylK enzyme. Note that *C. stagnale* produces cylindrocyclophane natural products in addition to the cylindrofridins.[83] (*B*) Selected amino acids in CylK near cavities 1 and 2 subjected to mutant scanning are shown in stick format. Side chains shown in yellow near cavity 1 correspond to sites found to be essential for activity. (*C*) Non-native substrate pair used in CylK activity assays and product of the alkylation reaction (SNAc = *N*-Acetylcysteamine). (*D*) Screen of mutant activity. Product formation was measured by LC-MS, error bars represent the standard deviation from the mean of three biological replicates. Panel B was adapted from images generated by Terry B. Ruskoski.

Table 3.2. Conservation of CylK residues selected for mutant screen to locate active site

Position	**Residue** (CylK, *C. licheniforme*)	Residue (CylK, *C. stagnale*)	Residue (CabK)	Residue (MerH)	Residue (BrtB)	**Mutant**
78	Lys	Lys	Lys	Lys	Lys	Glu
96	Asp	Asp	Asp	Asp	Asp	Ala
105	Arg	Arg	Arg	Arg	Arg	Ala
473	Tyr	Tyr	Tyr	Tyr	Tyr	Ala
265	Ser	Ser	Ser	Ala	Ala	Ala
266	Ser	Ser	Ser	Ser	Thr	Ala
267	Thr	Thr	Thr	Thr	Ala	Ala
324	Ser	Ser	Ser	Ser	Ser	Ala
377	Lys	Lys	Arg	Arg	Gly	Ala
379	Thr	Thr	Thr	Thr	Ala	Ala
502	Asp	Asp	Asp	Asp	Asp	Ala
504	Glu	Glu	Glu	Glu	Ala	Ala
558	Leu	Leu	Leu	Leu	Met	Ala
559	Asp	Asp	Asp	Asp	Gly	Ala
562	Asp	Asp	His	Glu	Val	Ala
616	Thr	Thr	Thr	Thr	Gly	Ala
618	Thr	Thr	Thr	Ser	Ser	Ala

Residue numbering is relative to CylK from *Cylindrospermum licheniforme* ATCC 29412.

In parallel, we hypothesized that the alkyl chloride functional group may be an important substrate-binding determinant based on our inability to obtain a full occupancy complex of alkyl resorcinol **7** in CylK crystals. Resorcinol **7** resembles the native substrate (**1**) but lacks the β-methyl and alkyl chloride substituents. Although **7** was required for CylK crystallization, we could not model this alkyl resorcinol substrate in the resulting structure. Instead, to visualize locations within CylK capable of binding an alkyl halide or free halide, we soaked CylK crystals with NaBr solutions. The Br^- ions are surrogates for the native Cl^- side product and are generated in reactions with alkyl bromide electrophiles, which CylK accepts with similar efficiency to alkyl chlorides (see Table 2.7).[128] We chose to use Br^- because bromine X-ray absorption energies can be accessed at a conventional synchrotron X-ray source for anomalous diffraction experiments, allowing us to pinpoint the location of the halide in the structure. Excitingly, initial refinements identified two strong positive peaks located ~15 Å apart in F_o–F_c electron density maps of the interdomain cavity (cavity 1). Anomalous diffraction datasets collected at the absorption edge of bromine confirmed assignment of these peaks as Br^- ions (Figure 3.10A). Importantly, anomalous peaks were not found in the

central channel of the β-propeller domain (cavity 2). These data further support our identification of the interdomain cleft (cavity 1) as the active site of CylK.

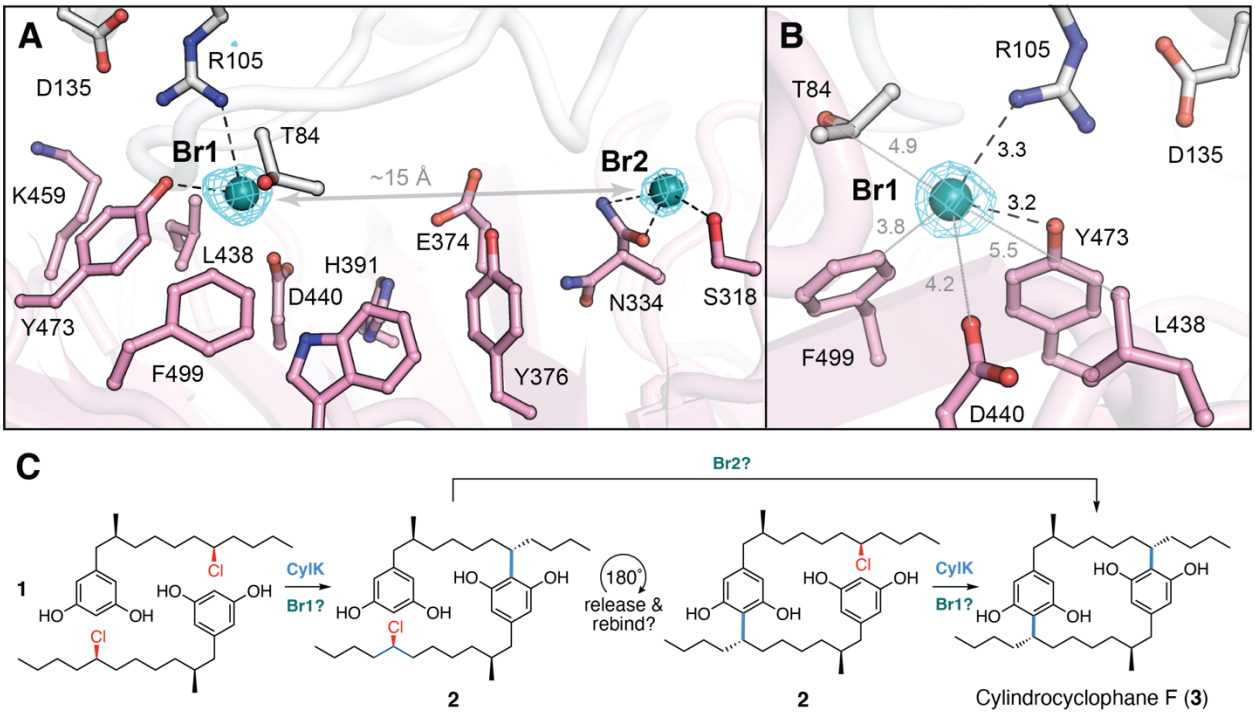


Figure 3.10. Anomalous diffraction of bound bromide in active site and mechanistic implications

(*A*) A soak of NaBr into CylK crystals reveals two strong peaks in the anomalous difference electron density map (cyan mesh, contoured at 5.0 σ) within cavity 1. Bromide ion 1 (Br1) and 2 (Br2) are shown as teal spheres. The N-terminal domain is colored pink and the C-terminal domain is colored white. (*B*) Alternate view of Br1, polar contacts shown as black dashed lines, other distances shown as grey lines. All distances measured in angstroms. (PDB 7ROO) (*C*) Alkylation scheme with potential roles for two bromide binding sites in paracyclophane formation, or invoking release and rebinding of mono-alkylated intermediate **2** for alkylation at a single bromide site. Panels A-B were adapted from images generated by Terry B. Ruskoski.

Br1, the bromide peak of highest intensity (~33 σ) in the anomalous difference electron density map, is located just inside the opening to the interdomain cavity 1 (Figure 3.10B). This bromide is modeled at 85% occupancy and it appears to be bound by the side chains of Arg105 from the N-terminal domain and Tyr473 from the C-terminal domain, the two residues identified in our mutagenesis scan. We hypothesize that these

interactions are Arg N–H–Br and Tyr O–H–Br hydrogen bonds, respectively, based on their orientations and bond lengths.[180] Of note, Asp440 and Leu438, undergo a conformational change upon bromide binding, however, the distance between Br1 and Asp440 (4.2 Å) suggests that this does not result in hydrogen bond formation. Water coordination changes, electrostatics, or subtle changes to the active site structure upon bromide binding might explain these movements. Notably, the bromide-free CylK structure also contained a positive F_o–F_c electron density peak at the Br1 site. Given the presence of chloride salts in the protein storage buffer, we modeled this site as a chloride ion in the bromide-free structure (see Figure 3.3). Interestingly, Br1 resides within 8 Å of the calcium-binding sites in the N-terminal domain, and the backbone carbonyl of Arg105 coordinates the central Ca^{2+} ion (Ca2) in the β-roll motif of the N-terminal domain (see Figure 3.4A). This observation suggests that Ca^{2+} might play a role in maintaining active site architecture, in addition to overall structural stabilization of both domains. Br1 is located within 5-6 Å of other side chains that form a portion of the solvent accessible channel nearest to the protein surface, including a number of polar residues that may be implicated in activation of the resorcinol nucleophile and/or the alkyl chloride electrophile.

A second bromide peak, Br2, is located deeper within the interdomain cavity. This ion is modeled at 48% occupancy and makes hydrogen bonding contact with Ser318, Asn334, and Trp320, all contributed by the C-terminal domain. While residues surrounding both bromides are generally conserved in closely related CylK homologs, those associated with Br2 are not conserved in the C–O bond forming family member BrtB (Table 3.3). This observation suggests that while Br2 residues might be relevant for paracyclophane forming enzymes, they may not be necessary for all family members.

With this in mind, we wondered if the two discrete alkylation reactions natively catalyzed by CylK occurred at one or both bromide binding sites. If only a single bromide site is involved in catalysis, we speculate that the mono-alkylated intermediate **2** is released from the active site, rotates 180°, and must rebind in the active site in a different orientation for the second alkylation event to occur (Figure 3.10C). To test this proposal, we individually mutagenized several residues closely associated with or appearing to directly bind

either bromide, and examined their ability to couple non-native substrate pair **7** + **16e** as before. This revealed that alanine mutants of Asn334 and Ser318 maintained near wild-type catalytic activity, while all other mutations had low or undetectable activity (Figure 3.11). In addition to Arg105 and Tyr473, Glu374, Leu438, Asp440, and Phe499 stood out as potentially important residues due to the exceedingly low activities of their corresponding mutants. Although the limitations of this assay include the inability to control for enzyme solubility/concentration and the use of a non-native substrate pair than cannot undergo a second intramolecular alkylation (macrocyclization), these results suggest that cavity 1 is the location of the CylK active site. Furthermore, this helped us prioritize key active site residues for further work.

Table 3.3. Conservation of CylK active site residues selected for mutant screen

Position	**Residue** (CylK, *C. licheniforme*)	Residue (CylK, *C. stagnale*)	Residue (CabK)	Residue (MerH)	Residue (BrtB)	**Mutant**
105	Arg	Arg	Arg	Arg	Arg	Ala, Lys
473	Tyr	Tyr	Tyr	Tyr	Tyr	Ala, Phe
440	Asp	Asp	Asp	Asp	Glu	Ala, Asn
442	Trp	Trp	Trp	Trp	Trp	Phe
438	Leu	Leu	Leu	Leu	Leu	Ala
459	Lys	Lys	Lys	Arg	Lys	Ala
135	Asp	Asp	Asp	Asp	Val	Ala
391	His	His	Asn	His	His	Ala
376	Tyr	Tyr	Phe	Phe	Met	Ala
374	Glu	Glu	Glu	Glu	Arg	Ala
84	Thr	Thr	Thr	Thr	Thr	Val
499	Phe	Phe	Phe	Phe	Tyr	Ala
334	Asn	Asn	Asn	Asn	Trp	Ala
318	Ser	Ser	Ser	Thr	Ile	Ala

Residue numbering is relative to CylK from *Cylindrospermum licheniforme* ATCC 29412.

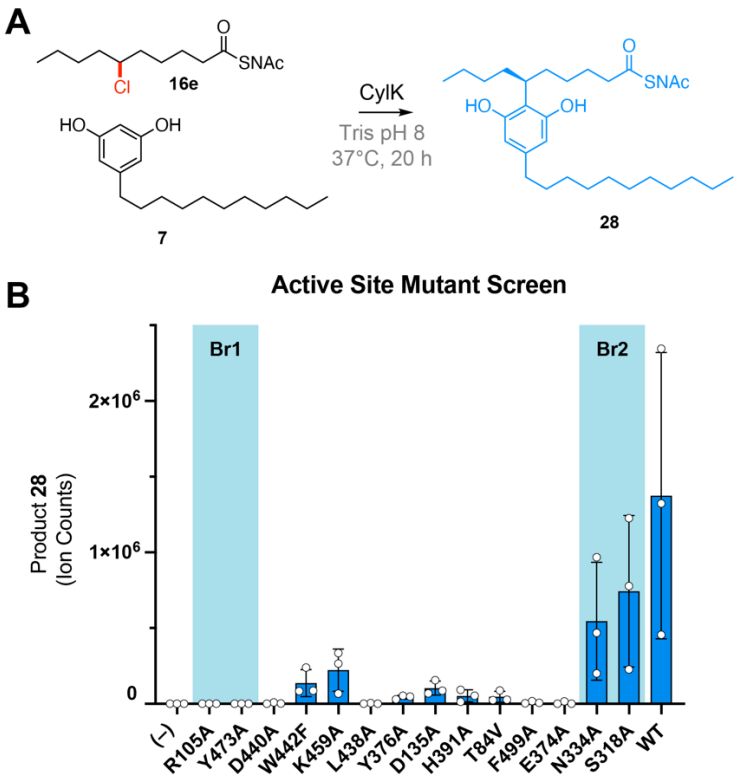


Figure 3.11. Mutant screen of CylK active site suggests 'Br1' is key catalytic center

(*A*) Non-native substrate pair used in CylK activity assays and product of the alkylation reaction (SNAc = *N*-Acetylcysteamine). (*B*) Screen of mutant activity. Product formation was measured by LC-MS, error bars represent the standard deviation from the mean of three biological replicates.

3.2.4 Identifying key catalytic residues

Based on our bromide anomalous data, the CylK active site cleft (cavity 1) may have one or two alkyl halide binding/activating regions (Br1 and Br2). Further, the stereospecificity and potentially concerted nature of the CylK-catalyzed reaction suggests that resorcinol activating residues may be located near to the site(s) of alkyl halide activation. Screening data with the non-native substrate pair (**7** + **16e**) clearly demonstrated the importance of Br1 for catalysis, however this reaction could not assess a potential role for Br2 in the native CylK dimerization and macrocyclization reactions. To further explore this possibility, we selected several mutants closely associated with either bromide, as well as nearby residues that had the lowest activity in our plate based assay, and examined their ability to perform the first and second alkylation reactions with native substrates **1** and **2**.

Considering that the first alkylation is effectively faster than the second,[69] likely due to substrate and/or product inhibition, we monitored the first alkylation by interrogating the 1 h time point of the native reaction. At this time point, *in vitro* reactions with wild-type enzyme at 37 °C have fully transformed both equivalents of **1** to intermediate **2**. Notably, we found that substitution of the residues that comprise the Br1 binding site (Arg105, Tyr473) abolished CylK activity on substrate **1** at 1 h, while mutations that disrupt Br2 (Ser318, Asn334) had no apparent effect, retaining wild-type levels of activity (Figure 3.12B). Variants targeting three additional residues located between the two bromide sites (Glu374, Leu438, Asp440) had very low or no activity and stood out as potential candidates for substrate binding or nucleophile activation due to their side chain functionalities and vicinity to Br1 on the cleft surface. Based on our earlier DFT work, we speculated that carboxylate side chains could be involved in resorcinol activation (see Figure 3.2). Additionally, mutating Phe499 to Ala did not significantly reduce conversion.

We then examined the second alkylation reaction using a mixture of intermediate **2** containing a small amount of cylindrocyclophane F (~18%) as the substrate. Consistent with the results obtained for the first alkylation reaction, variants that disrupted Br1 did not have appreciable activity on intermediate **2**, while substitution of residues comprising Br2 maintained near wild-type activity (Figure 3.12C). Mutating Glu374, Leu438, and Asp440 also resulted in no significant activity on **2** after 22 h, while the Phe499 to Ala mutant retained appreciable activity as before. Based on these results, we propose that Br1 is the site of alkyl chloride binding and both aromatic ring alkylation events. Additional work is required to determine if the residues that comprise Br2 assist in substrate binding, but it is clear from these data that they are not essential for catalysis. Furthermore, the site of catalysis at Br1 is located near the outermost portion of the solvent exposed interdomain cavity; this is in agreement with our earlier work that demonstrated CylK's ability to accept non-native substrates with rigid substitutions on the resorcinol ring nucleophile (see Table 2.8) which would necessitate a flexible or open active site.[128]

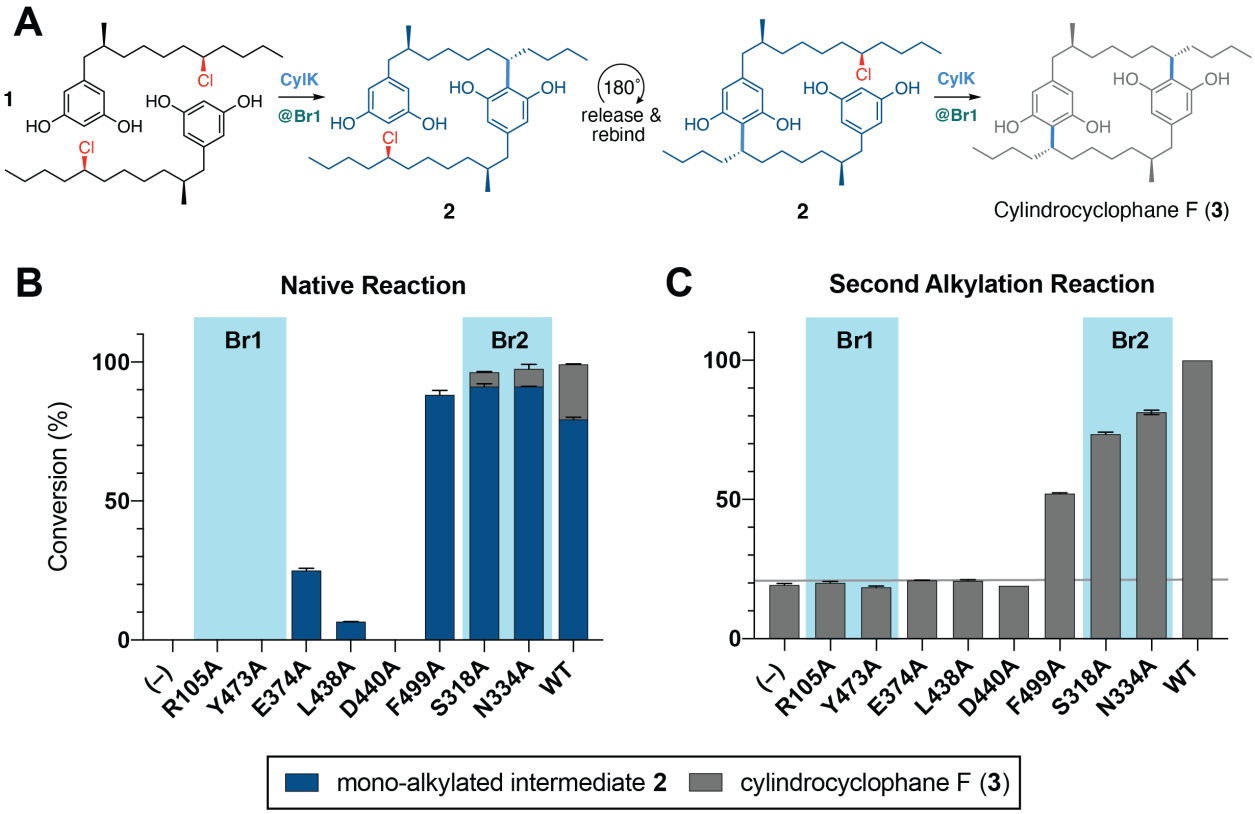


Figure 3.12. Mutagenesis of key CylK residues reveals a single catalytic site

(*A*) Operative alkylation scheme invoking release and rebinding of mono-alkylated intermediate **2** for catalysis at a single site. (*B*) End-point activity at 1 h of select mutants performing the native reaction with substrate **1**, highlighting the residues associated with Br1 or Br2. (*E*) End-point activity at 22 h of select mutants performing the second alkylation reaction with intermediate **2**. Product formation was quantified by HPLC, error bars represent the standard deviation from the mean of two biological replicates.

3.2.5 Modelling substrate orientation in the active site

After identifying the site of alkylation and potential key residues, we initiated a collaboration with the C. Denise Okafor lab at Penn State University in order to determine the specific mode of substrate binding and to assess the relative contributions of interactions with active site amino acids to resorcinol nucleophile and alkyl chloride electrophile activation.[xiii] All molecular dynamics (MD) simulations were performed with Amber20.[181] We began by computationally docking two equivalents of substrate **1** into the active site of

[xiii] This work was conducted by C. Denise Okafor with contributions from Nathaniel R. Braffman, Terry B. Ruskoski, and Amie K. Boal.

CylK (Figure 3.13). The initial complexes were generated manually using the Br1 site to anchor the electrophilic alkyl chloride substituent of one equivalent of **1**. Potential interactions with important carboxylate residues Asp440 and Glu374 guided the placement of the nucleophilic resorcinol of the other equivalent of substrate **1**. Our earlier work suggested that a resorcinol–carboxylate hydrogen bond (O–H–O) would enhance nucleophilicity of the aromatic ring (see Figure 3.2).[128] Additionally, we oriented the nucleophilic and electrophilic carbon centers at an appropriate distance and arrangement to accommodate the known stereochemical inversion of the substitution reaction.

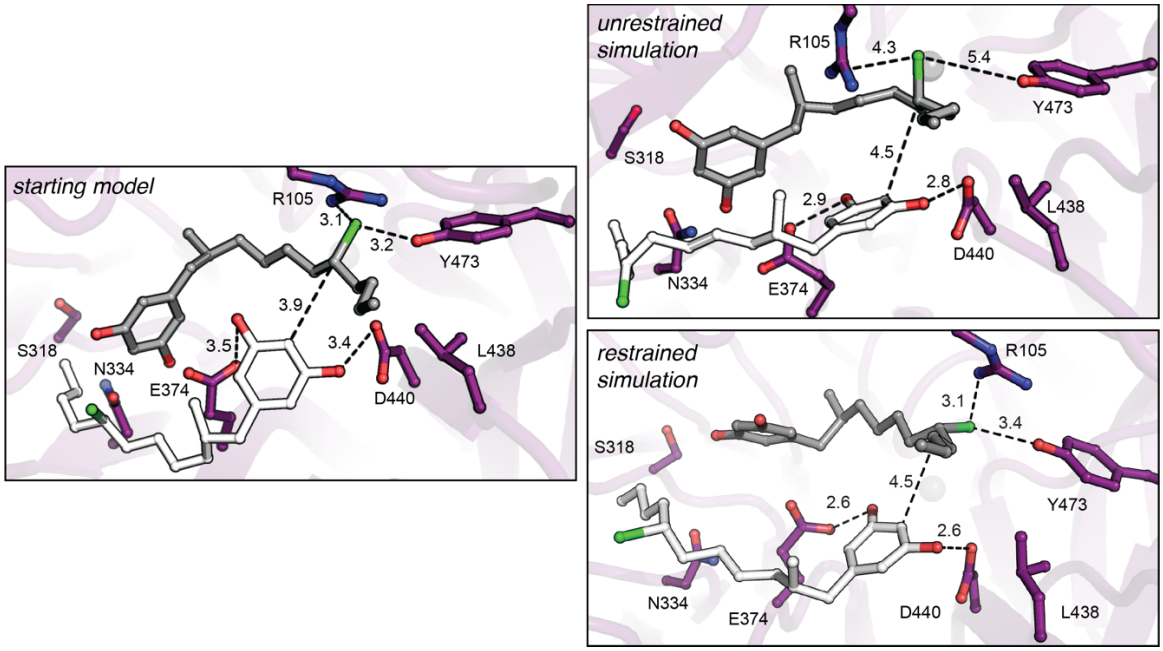


Figure 3.13. Molecular dynamics simulations of the intermolecular alkylation catalyzed by CylK

An energy minimized docking model of CylK in complex with two chlorinated alkyl resorcinol substrates in cavity 1 before (starting model) and after unrestrained or restrained MD simulations. Select residues shown as dark purple sticks, substrate molecules as white or grey sticks, oxygens in red, chloride in green, and distances measured in angstroms. This figure was adapted from images generated by Terry B. Ruskoski.

This starting complex was subjected to restrained and unrestrained simulations. Restrained simulations included 1 kcal/mol·Å^2 restraints on all protein atoms to allow substrates to search the active site. These calculations revealed that the active site cleft can accommodate two equivalents of substrate **1**, and

furthermore, Asp440 and Glu374 maintained hydrogen bonds with the resorcinol nucleophile (Figure 3.13). For the majority of the simulation time, the halide binding pocket formed by Arg105 and Tyr473 (Br1) maintained contact with the alkyl chloride electrophile. This suggests that Asp440 and Glu374 are resorcinol nucleophile binding and activating residues, respectively, while Arg105 and Tyr473 bind and activate the alkyl chloride electrophile. This further emphasizes that the source of CylK's regio- and stereoselectivity may be the orientation in which the resorcinol nucleophile is held in close proximity to the backside of the C–Cl bond for a concerted substitution reaction. Simulations of intermediate **2**, positioned for the second alkylation reaction, maintained similar alkyl chloride electrophile interactions, however, the resorcinol nucleophile did not remain at an appropriate distance or orientation for the reaction (Figure 3.14). More work is required to accurately model the second alkylation step, however, it is clear from our mutagenic work that the same residues are essential for both alkylation events.

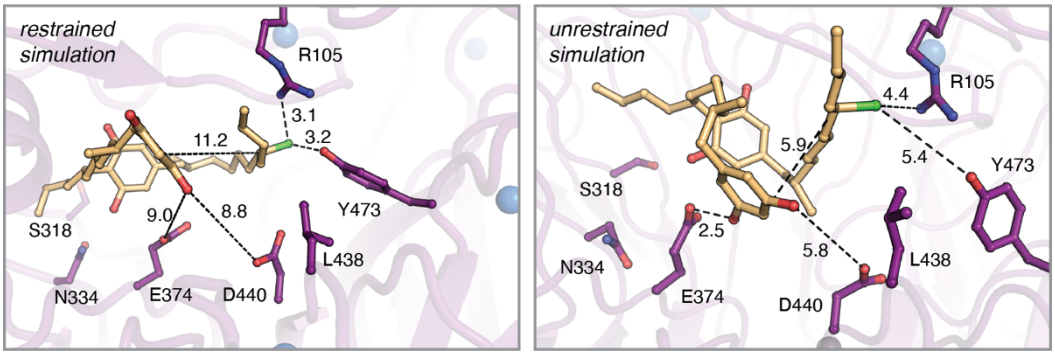


Figure 3.14. Molecular dynamics simulations of second alkylation

Restrained and unrestrained MD simulations of intermediate **2** in the CylK active site poised for the second alkylation reaction. Select residues shown as dark purple sticks, substrate molecule as yellow sticks, oxygens in red, chloride in green, and distances measured in angstroms. Starting complex not shown. This figure was adapted from images generated by Terry B. Ruskoski.

3.2.6 Comparison to other CylK crystal structures

Concurrently with our work, Wang et al. published three crystal structures of CylK and reached similar conclusions regarding the active site location and key catalytic residues.[141] Their structures are of: (1) wild-type CylK (truncating eight residues of the N-terminus) PDB 7FH6, (2) Y37F CylK with three bound fluorinated derivatives of substrate **1** (F-**1**) PDB 7FH7, and (3) H391A CylK with a fluorinated derivative of intermediate **2** (F-**2**) PDB 7FH8. Resolutions ranged from 1.32–1.55 Å, and R-work and R-free values are appropriate for this resolution. Ligand structure quality assessments are low, likely due to low occupancy and/or goodness-of-fit, although flexible alkyl halide substrates such as these are expected to be challenging to model. Their crystal form is similarly orthorhombic but in the $P2_12_12_1$ space group while ours is $C222_1$. Overall, their protein structures are strikingly similar to our CylK structures (Figure 3.15). The key difference appears to be a 'downward' shift of the alpha helices of the N-terminal domain towards the RTX motif Ca^{2+}-binding region. Although one of these alpha helices contains the Y37F mutation (Figure 3.16A), this shift is also present in the wild-type and H391A structures. This suggests that this relative domain positioning is inherent to the crystal form. However, such dynamic protein movement may be relevant to catalysis and the behavior of this enzyme in solution or in the context of the cyanobacterial cell, although more work would be required interrogate this proposal.

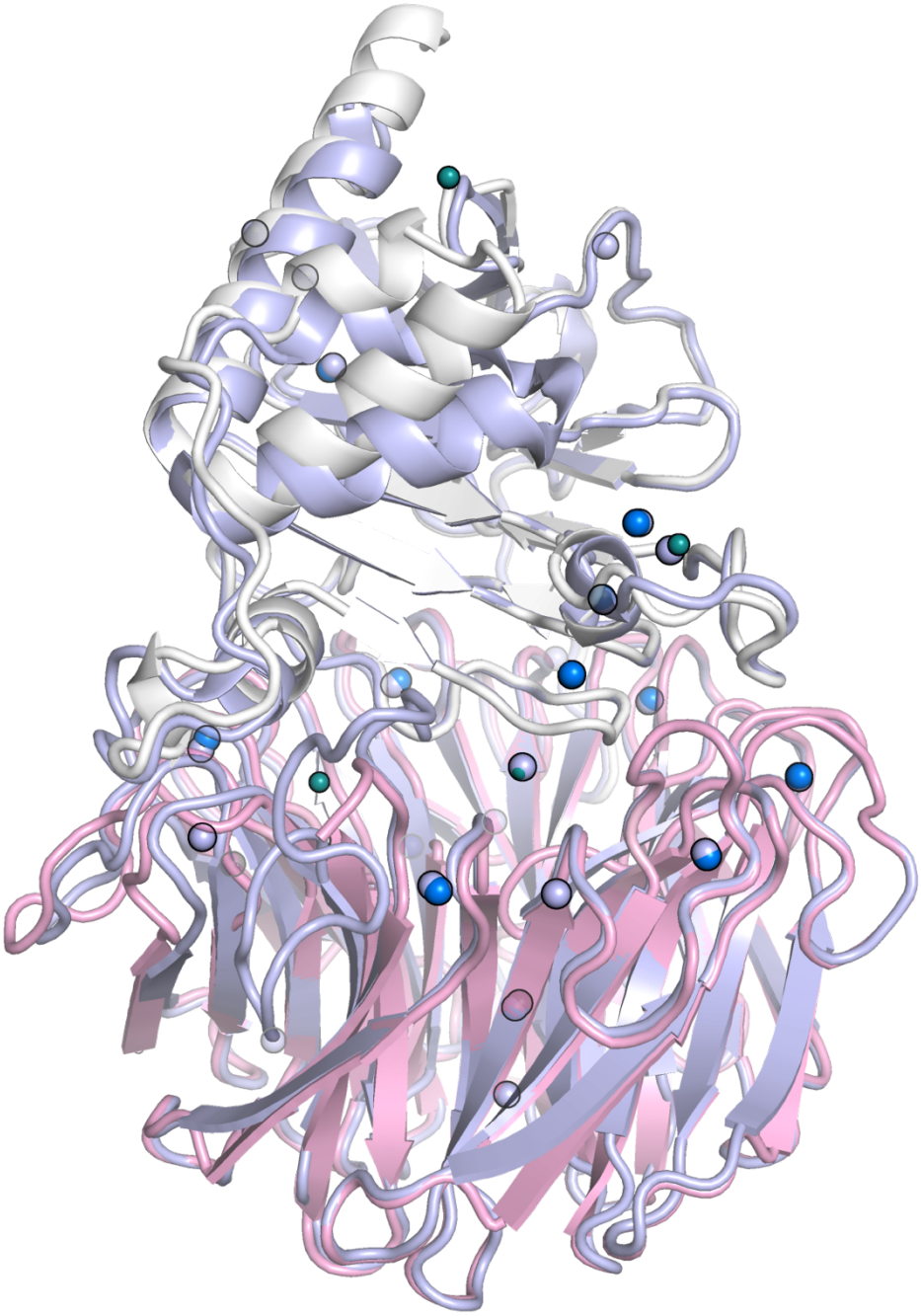


Figure 3.15. Comparison of CylK crystal structures from different research groups

Superimposed structures of our Br-soaked CylK (PDB 7ROO) and Y37F CylK with fluorinated **1** (F-**1**) (PDB 7FH7) from Wang et al. In our Br-soaked CylK, the N-terminal domain is colored light grey, the C-terminal domain is pink, calcium ions are blue, magnesium ions are dark grey, and bromide ions are teal. The Y37F CylK structure and its associated ions are rendered in light purple.

The most informative structure is Y37F CylK with three bound derivatives of F-**1** (PDB 7FH7) because it depicts two of these substrates in a pseudo competent conformation for an S_N2-like reaction. The third equivalent of substrate appears to sit above the active site, its alkyl tail interacting with the side chain of the Phe37 mutation, while a phenol of its resorcinol ring interacts with the neighboring resorcinol ring of electrophile F-**1** (Figure 3.16A). It is unclear if this bound species is biologically significant although it is interesting to speculate about a potential allosteric or regulatory role at this site. Notably, the two catalytically relevant F-**1** species are positioned in the active site alongside a bound chloride ion (Figure 3.16B). This chloride aligns with our anomalous difference electron density for bromide (Br1), suggesting this assignment as chloride is accurate. We speculate that this halide is bound as if it were the coproduct of the reaction. Importantly, all active site residues sit in a nearly identical conformation to our bromide-bound structure (except two rotamers of Glu374 are modelled). As in our structure, their halide ion is coordinated by Arg105 and Tyr473 with distances suggestive of N–H–X and O–H–X hydrogen bonds.[180] This halide also makes non-polar interactions with Thr84 and Phe499 (see Figure 3.17 for distances). Analogous to our modelling work, the resorcinol nucleophile interacts with Asp440 and Glu374 with resorcinol–carboxylate hydrogen bonds (O–H–O). Their alkyl fluoride electrophile sits between these species in what we propose is an inactive conformation with its lowest unoccupied molecular orbital (LUMO), σ^*_{C-Cl}, pointing away from the resorcinol nucleophile. In particular, the trajectory of attack from the resorcinol's nucleophilic 2-carbon to the line of the C–F bond is 64 º (C–C–F), in contrast to 180 º expected for a concerted substitution reaction. This suggests that the alkyl halide electrophile might reorient itself into a catalytically competent conformation when the bound halide species is not present. We speculate that chloride is bound in this structure because of the low occupancy of the F-**1** substrates in addition to the high concentration of chloride during crystallization (100 mM NaCl, 200-300 mM $MgCl_2$ or $BaCl_2$).[xiv]

[xiv] Notably, we do not detect competitive halide inhibition, see Chapter 3.2.8 for discussion.

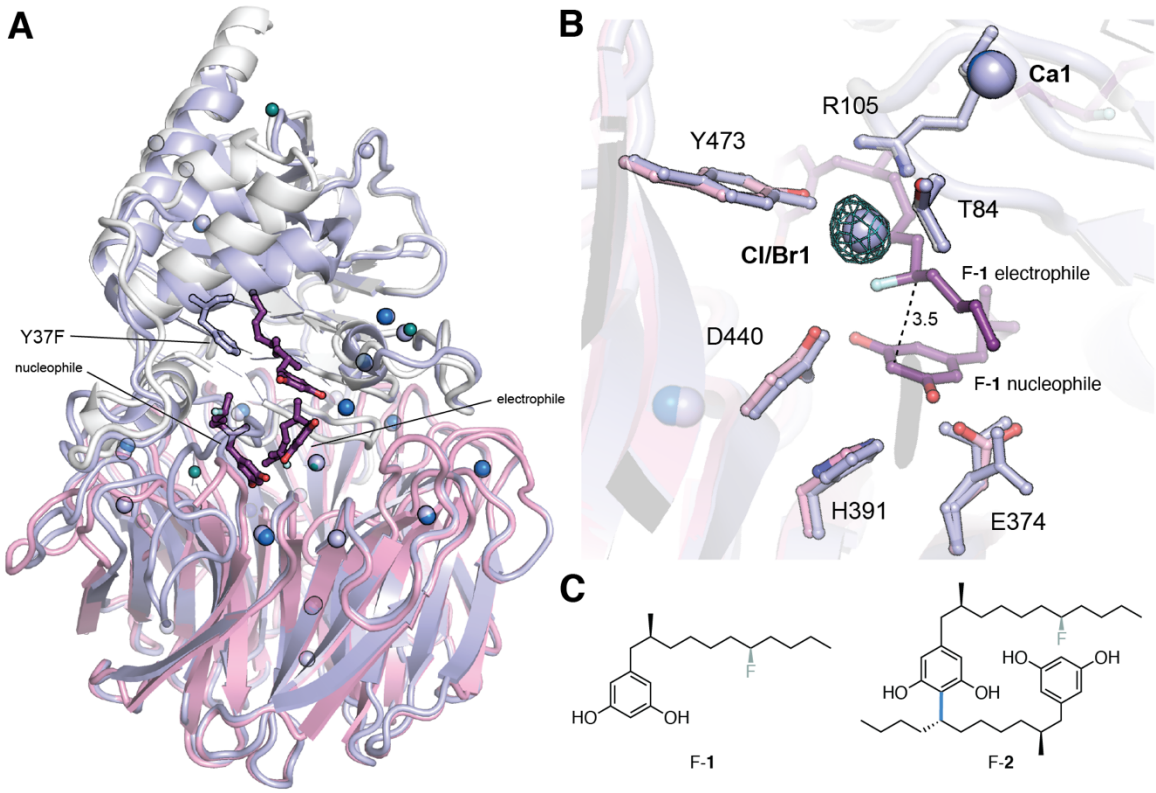


Figure 3.16. Substrate analogs and active site comparison between CylK crystal structures

Superimposed structures of our Br-soaked CylK (PDB 7ROO) and Y37F CylK with F-**1** (PDB 7FH7) from Wang et al. as in Figure 3.15, including substrates (F-**1**) rendered in dark purple. (*A*) Overall view and (*B*) active site perspective highlighting key catalytic residues. Some residues and distances are omitted for clarity, see Figure 3.17. Note, between panels A and B, the perspective is rotated ~180° around the resorcinol C–C alkyl halide reaction coordinate. In our Br-soaked CylK, the N-terminal domain is colored light grey, the C-terminal domain is pink, calcium ions are blue, magnesium ions are dark grey, bromide ions are teal, and bromine anomalous difference electron density map is cyan mesh. The Y37F CylK structure and its associated ions are rendered in light purple, and F-**1** substrates are in dark purple. Distances measured in angstroms. (*C*) Inactive fluorinated derivatives of substrates **1** and **2**.

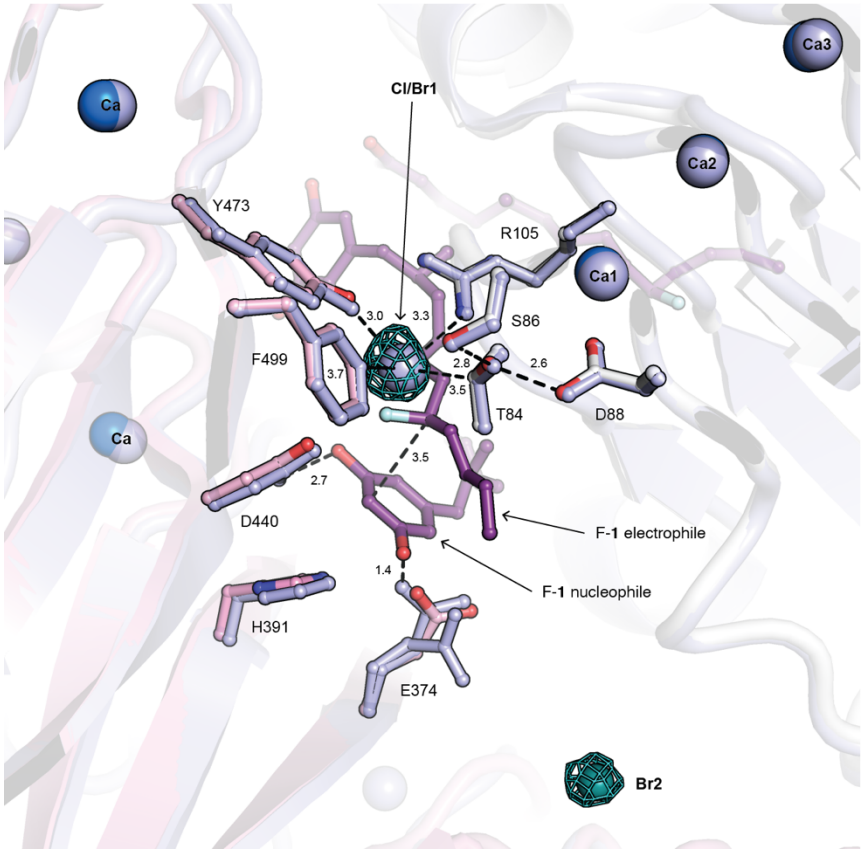


Figure 3.17. Comparison of active site residues and distances between CylK crystal structures

Superimposed active sites of our Br-soaked CylK (PDB 7ROO) and Y37F CylK with F-**1** (PDB 7FH7) from Wang et al. as in Figure 3.16B, including additional distances and residues F499, S86, D88. In our Br-soaked CylK, the N-terminal domain is colored light grey, the C-terminal domain is pink, calcium ions are blue, magnesium ions are dark grey, bromide ions are teal, and bromine anomalous difference electron density map is cyan mesh. The Y37F CylK structure and its associated ions are rendered in light purple, and F-**1** substrates are in dark purple. Distances measured in angstroms.

Although our structural and modelling work demonstrate the same key catalytic interactions as observed in the Y37F CylK structure from Wang et al., their F-**1** electrophilic substrate is rotated approximately 180° around the resorcinol C–C alkyl halide reaction coordinate relative to our substrate modeling work, and while the resorcinol nucleophiles are similarly oriented, the alkyl halide tail of their nucleophilic substrate (F-**1**) is rotated to face in the opposite direction relative to our substrate model (Figure 3.18). Importantly, both substrate orientations engage the same resorcinol- and halide-interacting residues, while the key differences emerge when considering the full mode of substrate binding. Further mutagenesis experiments

paired with binding affinity determinations might shed light on substrate interactions distal to the catalytic center. Regardless, based on our work, and in agreement with the structures from Wang et al., we maintain our assignments of Asp440 and Glu374 as resorcinol nucleophile binding and activating residues, and Arg105 and Tyr473 as alkyl halide activating residues. Furthermore, we hypothesize that Thr84 and Phe499 contribute nonpolar interactions to bind the alkyl halide and stabilize the transition state.

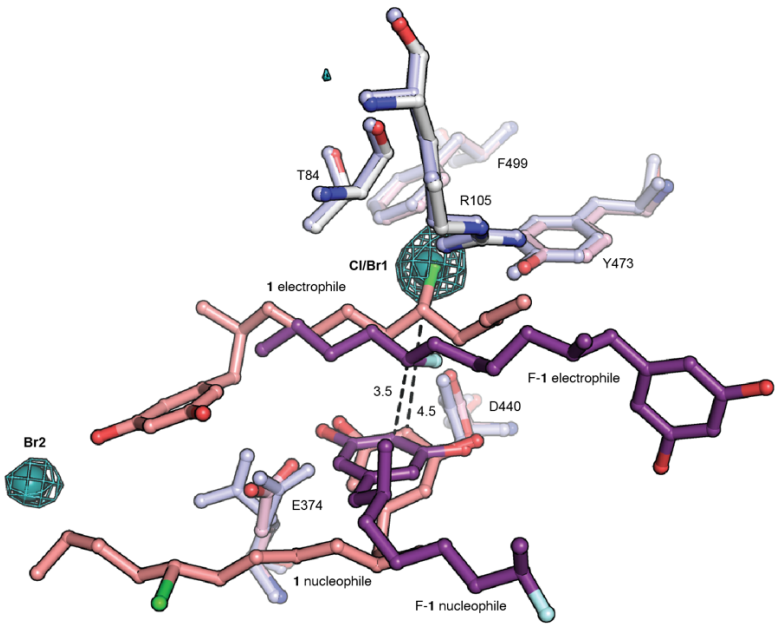


Figure 3.18. Comparison of substrate orientations between CylK crystal structure and modeling

Superimposed active site residues from our Br-soaked CylK (PDB 7ROO) and Y37F CylK with F-**1** (PDB 7FH7) from Wang et al. F-**1** substrates are rendered in dark purple, and MD substrates (**1**) from a restrained simulation are overlaid and colored salmon. From our Br-soaked CylK, N-terminal domain residues are colored light grey, C-terminal domain residues are pink, bromide ions are teal, and bromine anomalous difference electron density map is cyan mesh. Residues from the Y37F CylK structure are rendered in light purple.

In contrast, Wang et al. propose a role for the hydroxyl group of Thr84 as an alkyl chloride activating hydrogen bond donor based on quantum mechanics/molecular mechanics modeling of the substitution reaction.[141] We note that all structures of CylK feature this hydroxyl group facing away from the bound halide ion (as represented in Figure 3.17). Considering that both an alkyl halide and C–X bond-breaking

transition state are less polar and charged than a free halide ion, we believe that non-polar interactions between Thr84 and the alkyl halide are likely maintained over the course of reaction, contributing to the architecture of an alkyl halide biding pocket of both hydrogen bond donors (Arg105, Tyr473) and non-polar interactions (Thr84, Phe499). Although it is conceivable that this side chain can access alternate rotamers, its positioning in all crystal structures is stabilized by hydrogen bonds to the side chains of Ser86 and Asp88 (depicted in Figure 3.17). However, more work is required to fully understand the role of this residue in the reaction catalyzed by CylK.

The cocrystal structure of H391A CylK with F-**2**, a non-reactive intermediate **2** analog, features this molecule bound as the product of the first alkylation.[141] As discussed previously in Chapter 3.2.4, we speculate that this intermediate must exit the active site and reorient 180° before rebinding in a reactive conformation at the same site of the first alkylation reaction. This inactive orientation of F-**2** in the CylK crystal structure suggests that intermediate **2** might competitively inhibit the second alkylation reaction. This is potentially in agreement with the observation that the first alkylation reaction occurs much more rapidly than the second.

3.2.7 Targeted mutagenesis

Our structural and mutagenesis experiments established that Arg105 and Tyr473 are key catalytic residues likely involved in alkyl chloride electrophile activation via hydrogen bonding, while our mutagenesis and MD simulations demonstrated that Asp440 is an essential catalytic residue likely involved in resorcinol nucleophile activation. Structural and studies from Wang et al. and our computational modeling both implicate Glu374 as a resorcinol binding residue, however, we found that the alanine mutant of Glu374 retained measurable activity for the first alkylation reaction (see Figure 3.12) which led us to conclude that Asp440 is the key resorcinol activating residue (likely via deprotonation by its carboxylate side chain) while Glu374 participates in resorcinol binding as an hydrogen bond donor. Computational modeling and mutagenesis work from Wang et al. is in agreement with our analysis of Glu374. Considering that the alanine mutant of Glu374 did not measurably perform the second alkylation reaction, we suggest that this

residue might be necessary to position intermediate **2** in the reactive conformation for catalysis, specifically orienting the resorcinol nucleophile for both alkylation reactions, instead of acting as a base.

Based on our structural and modeling work, we propose hydrogen bond interactions between the side chains of Arg105 and the alkyl halide (N–H–X), Tyr473 and the alkyl halide (O–H–X), and Asp440 and the resorcinol phenol (O–H–O). Notably, these interactions are consistent with the structures reported by Wang et al. and observed hydrogen bond lengths (2.7–3.3 Å).[180,182] In order to further test the importance of these interactions for catalysis, we made additional mutations to these three catalytic residues that maintained similar properties such as charge, steric bulk, and/or hydrogen bonding capacity to the original side chains, but altered key features that might contribute to their interactions with substrates. Namely, we mutated Arg105 to Lys (charge, hydrogen bond donor), Tyr473 to Phe (steric bulk, aromaticity), and Asp440 to Asn (steric bulk, hydrogen bond donor/acceptor). Mutating Arg105 to Lys, and Asp440 to Asn resulted in no activity for either alkylation reaction, suggesting that the size, hydrogen bonding capacity, and/or pK_a of these residues might be important for catalysis (Figure 3.19B,D). However, mutating Tyr473 to Phe only moderately reduced activity and might suggest that the aromatic nature and/or size of this residue is critical for alkylation rather than the proposed hydrogen bonding interaction with the phenol group of Tyr that is absent in Phe.

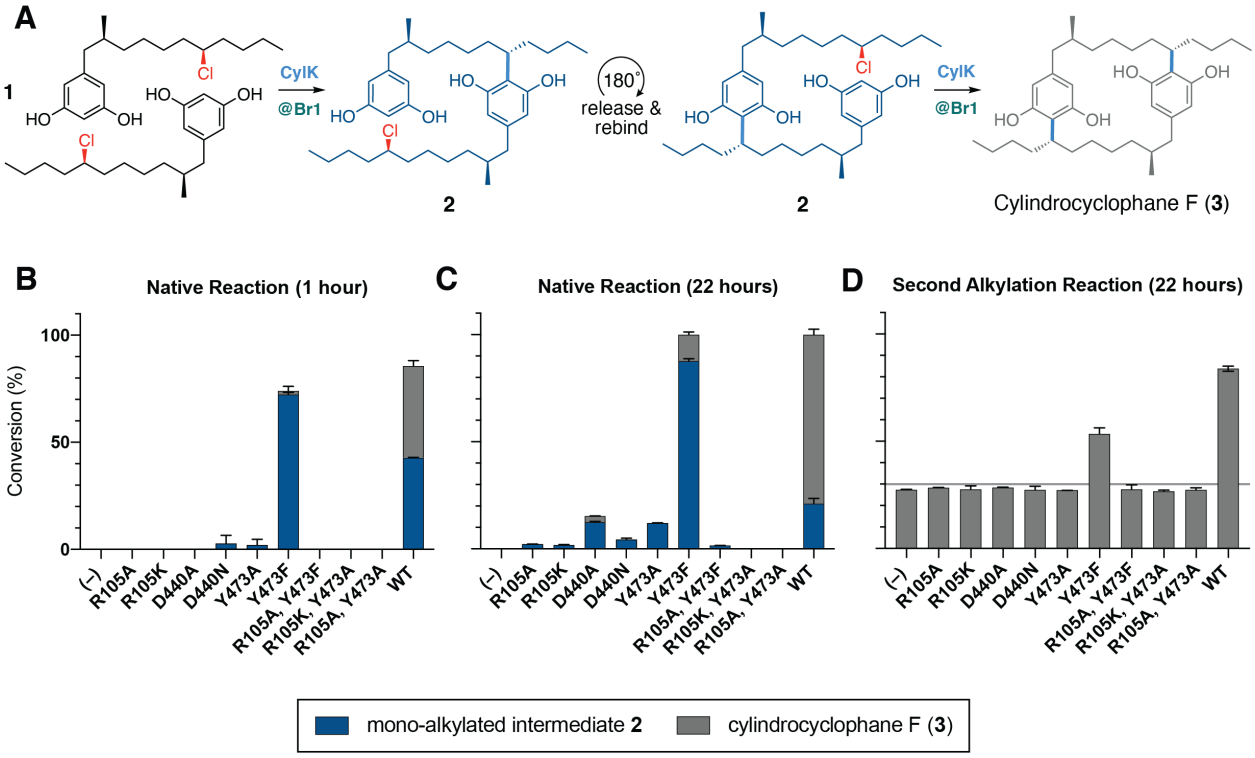


Figure 3.19. Point and combinatorial mutagenesis of key substrate activating residues

(*A*) Operative alkylation scheme invoking release and rebinding of mono-alkylated intermediate **2** for catalysis at a single site. End-point activity at (*B*) 1 h or (*C*) 22 h of select mutants performing the native reaction with substrate **1**. (*D*) End-point activity at 22 h of select mutants performing the second alkylation reaction with intermediate **2**. Product formation was quantified by HPLC, error bars represent the standard deviation from the mean of two biological replicates.

The retention of activity of the Tyr473 to Phe mutant, which is consistent for both the first and second alkylation reactions, might imply that the alkyl chloride–Tyr473 interface resembles an anion–π interaction,[183,184] in contrast to our proposed tyrosine–alkyl chloride hydrogen bond (O–H–Cl). Alternatively, if the phenol of Tyr473 does provide a key hydrogen bond with its phenol group, the active site of the Phe mutant may be able to rearrange and partially compensate for this loss with new interactions. In support of this type of rearrangement, we observed that if the native reaction is run for 22 h, the CylK active site can somewhat compensate for the loss of any of the key catalytic residues (Figure 3.19C). At this timepoint, detectable amounts of mono-alkylated intermediate **2** are produced by the alanine mutants of Arg105, Tyr473, and Asp440 (2-12% conversion). This suggests that although these singe point mutants are poor

catalysts, they are not entirely inactivated and can partially stabilize the resorcinol–alkyl halide transition state (C–C–Cl). Of note, double point mutants of Arg105 and Tyr473 were inactive at both 1 and 22 h, except for trace activity of the Ala & Phe double mutant, respectively (Figure 3.19B-D). With this in mind, it appears that Phe mutants of Tyr473 are not as disruptive as other substitutions, and the CylK active site can more readily compensate for the loss of the Tyr473 hydrogen bond donor with an alkyl halide–aryl interaction and other potential rearrangements (Figure 3.20B). There is precedent for such alkyl halide–aryl interactions in proteins,[185] although they are less thoroughly studied than anion–π interactions. Notable enzymatic examples include haloalkane dehalogenase and the SAM-dependent chlorinase, SalL (Figure 3.20C-D).[167,186] The active sites of both of these examples engage primary alkyl chloride substrates or products with alkyl halide–aryl edge interactions contributed by aromatic amino acid side chains including Phe and Tyr. In haloalkane dehalogenase, two Phe residues coordinate one alkyl chloride of its dichloroethane substrate; note that this alkyl chloride is not the reactive bond that is hydrolyzed. Crystal structures of these enzymes and further discussion can be found below (see Chapter 3.2.9) as examples of alkyl halide bond-forming/breaking enzymology.

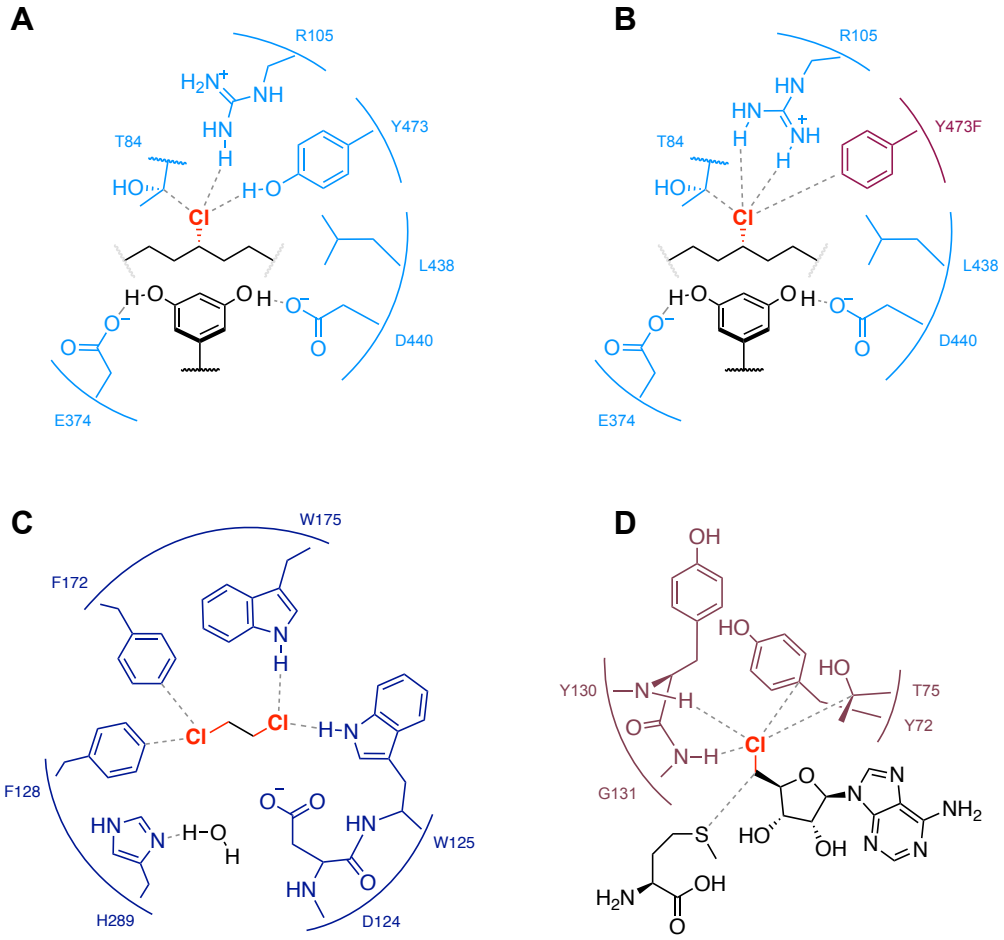


Figure 3.20. Chemdraw representations of enzyme engagement of alkyl chlorides

(*A*) Proposed interactions between CylK catalytic residues and alkyl chloride electrophile and resorcinol nucleophile substrates. For clarity, Phe499 is not represented. Phe499 sits behind the plane of the page and likely provides an alkyl chloride–aryl edge interaction, as seen with free halides (see Figure 3.17). (*B*) Hypothetical interactions and rearrangements of the Phe mutant of Tyr473 to compensate with an alkyl chloride–aryl edge interaction. Representations of the active sites of (*C*) haloalkane dehalogenase and (*D*) SAM-dependent chlorinase, highlighting analogous alkyl chloride–aryl edge interactions, based on PDBs 2DHC and 2Q6I, respectively. See Figure 3.25 for crystal structures and further discussion.

3.2.8 Halide inhibition

Considering that our CylK structures and those of Wang et al. featured chloride or bromide ions bound by proposed alkyl halide activating residues, we sought to determine if free halides inhibit the CylK-catalyzed reaction. We hypothesized that these anions were likely not inhibitory because all of our purification and reaction buffers generally contain 50-500 mM NaCl in addition to 5-10 mM of $MgCl_2$ and $CaCl_2$, with enzyme present at concentrations of 0.25-2.0 μM. However, understanding the extent to which halide inhibition occurs, and how CylK might have overcome this potential problem could shed light on its mechanism and inform the design of analogous chemocatalysts that engage alkyl halide substrates in substitution reactions.

To test for halide inhibition, we diluted CylK into a reaction buffer that excluded sources of chloride, substituting 10 μM $CaSO_4$ for our standard calcium supplementation source ($CaCl_2$). The lowest concentration of chloride that could be achieved was approximately 500 μM due to contaminating NaCl, $MgCl_2$, and $CaCl_2$ from the standard CylK preparation (see Materials and Methods 3.4.4). Regardless, we analyzed the rate of this CylK reaction with increasing concentrations of supplemented NaCl above the background level, up to 200 mM. Interestingly, increasing the concentration of NaCl had no apparent effect on the reaction rate in this context (Figure 3.21).

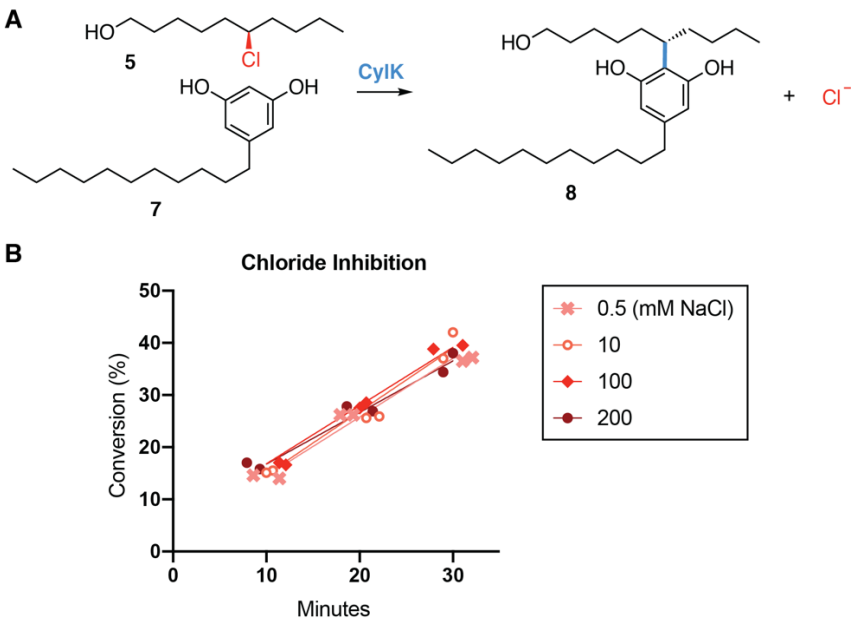


Figure 3.21. The effect of millimolar concentrations of chloride on the CylK-catalyzed reaction

(*A*) Reaction used in kinetic assays. (*B*) Plot of conversion over time with added NaCl, quantified by HPLC with two biological replicates for each condition. Lines represent linear regression of each NaCl set. All reactions were run for at 37 °C for 10–30 min, with 200 μM of alkyl chloride (*R*)-**5**, 200 μM of resorcinol **7**, and 2 μM of CylK-Strep.

Although promising, our substrate concentrations (at 200 μM) were less than our lowest NaCl condition (500 μM), suggesting that any additional effect of halide inhibition might be suppressed under these conditions. To further explore the possibility of halide inhibition, we modified our CylK preparation protocol to eliminate all sources of chloride (see Materials and Methods 3.4.5). In short, after standard refolding and ULP proteolysis in the presence of NaCl, $MgCl_2$, and $CaCl_2$, CylK was dialyzed into a buffer containing 100 mM $NaNO_3$ in lieu of NaCl. Ca^{2+} and Mg^{2+} sources were excluded. Size exclusion chromatography was then conducted in the same buffer system. This analysis revealed that this CylK preparation was a 1:1 mixture of monomeric and dimeric species, likely due to the exclusion of Ca^{2+} (see Materials and Methods 3.4.5 and Figure 3.31). Dimeric species are generally negligible in our standard preparations. Then, as before, we analyzed the rate of this reaction supplemented with a range of concentrations of NaCl; and we also included NaBr and NaI. Excitingly, assays using this enzyme preparation revealed that 10–500 μM of

NaCl and NaBr, or 100 μM NaI had no measurable effect on CylK's reaction rate (Figure 3.22). Overall, the slower reaction rate of this enzyme preparation is likely due to inactive dimeric species.

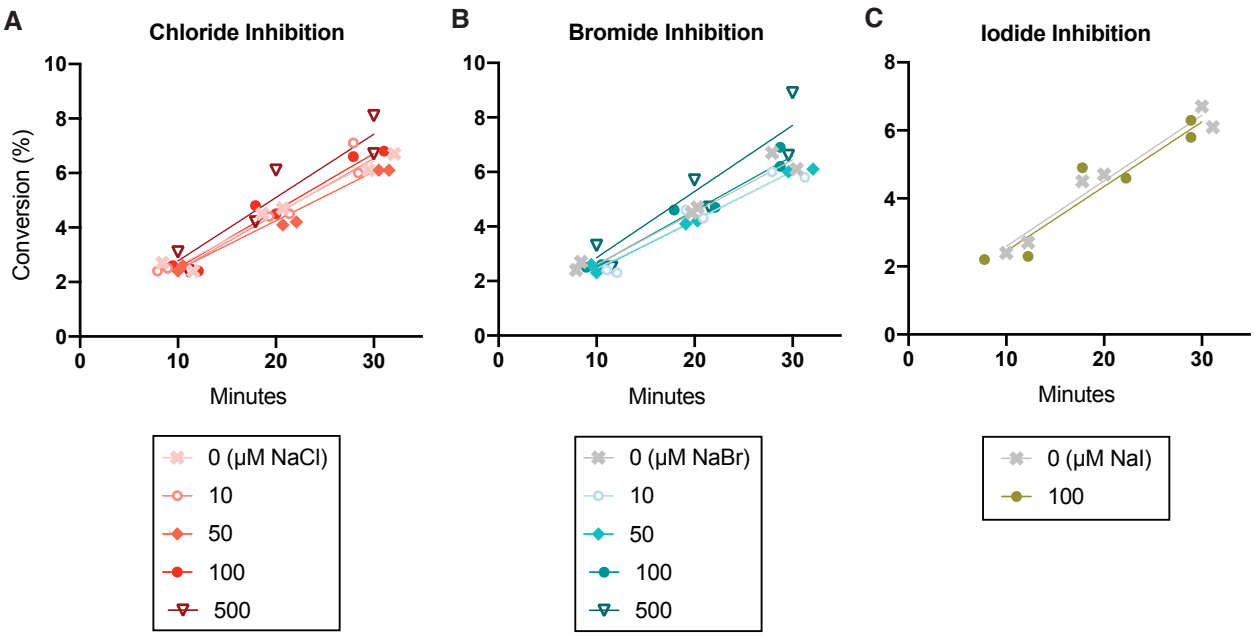


Figure 3.22. The effect of micromolar concentrations of halides on the CylK-catalyzed reaction

Plot of conversion over time with added (*A*) NaCl, (*B*) NaBr, (*C*) NaI, quantified by HPLC with two biological replicates for each condition. Lines represent linear regression of each NaX set. All reactions were run for at 37 °C for 10–30 min, with 200 μM of alkyl chloride (*R*)-**5**, 200 μM of resorcinol **7**, and 2 μM of chloride-free CylK-Strep.

This remarkable feature of CylK suggests that the enzyme can preferentially bind the resorcinol–alkyl halide transition state ($C^{\delta-}$---C---$X^{\delta-}$) over free halide leaving group. Our structure has bound chloride (or bromide) at this site likely due to its high concentration in protein storage buffer and crystallization conditions (25 mM NaCl, or 500 mM NaBr). The structural work from Wang et al. also contains bound chloride and accordingly reports to use high levels of chloride from multiple sources during crystallization (100 mM NaCl, 200-300 mM $MgCl_2$ or $BaCl_2$) relative to 1-2 mM of substrate analogs F-**1** and F-**2** at low occupancy in these protein crystals. This might explain why all CylK structures contain bound halides. However, the ability of CylK to catalyze the 'forward' substitution reaction using an alkyl halide electrophile in the presence of high

concentrations of free halides, further speaks to why understanding its mechanism is of fundamental interest. Avoiding this mode of inhibition is a challenge for analogous small molecule catalytic methods such as thiourea-based anion-abstraction catalysis, which can be suppressed by exogenous halide.[187] In these systems, the thiourea catalysts likely have higher affinity for the halide leaving group in comparison to the alkyl halide bond-breaking transition state, potentially due to the increased Lewis basicity of free halides.[188] Conversely, CylK appears to construct its alkyl halide binding pocket from a suite of polar hydrogen bond donors (Arg105, Tyr473) as well as non-polar residues (Phe499 and the alkyl portion of Thr84) (see Figure 3.17 and Figure 3.27 for illustrations of this pocket), and we hypothesize these hydrophobic side chains might diminish the pocket's affinity for free halides.

One potential consequence of this active site architecture is that CylK may promote nucleophile activation in concert with alkyl halide electrophile bond-weakening. In other words, the enzyme may use an S_N2-like reaction manifold consistent with the stereochemical outcome of the reaction. In this way, CylK may have evolved to minimize its affinity for free halides by not too strongly nor exclusively promoting leaving group activation as would be required for an S_N1-like reaction. However, if the CylK-catalyzed reaction did proceed by generating a discrete carbocation, the enzymatic architecture needed to stabilize a C–X bond breaking transition state ($C^{\delta+}$---$X^{\delta-}$) would likely have a higher affinity for free halides, and might therefore be more prone to product halide inhibition. We speculate that this would occur because a C–X bond breaking transition state ($C^{\delta+}$---$X^{\delta-}$) bears more localized charges than the resorcinol–alkyl halide ($C^{\delta-}$---C---$X^{\delta-}$) transition state in which the charge is more widely distributed across more atoms, including an aryl system. Compared to the latter S_N2-like transition state, the former (S_N1-like transition state with more localized charges) is more akin to a free halide ion and would likely necessitate more polar active site residues to bind and stabilize the developing anionic halide leaving group.

In order to model this, we calculated DFT transition states for protonated (Figure 3.23A) and deprotonated (Figure 3.23B) alkyl resorcinol–alkyl chloride transition state complexes. The reality of the CylK-catalyzed reaction may lie in between these two extremes, with partial deprotonation of one or both phenols by

Asp440 and Glu374. Interestingly, the protonated complex appears to have a late transition state in which the breaking alkyl chloride bond is longer than the developing resorcinol–alkyl bond. This suggests that even a concerted reaction in the absence of resorcinol nucleophile activation (via deprotonation) contains more localized negative charge on the halide ion (yellow) in its transition state, and might be susceptible to halide inhibition. Conversely, the deprotonated complex representative of resorcinol nucleophile activation (via deprotonation) contains a symmetric transition state and distributes negative charge throughout the complex (yellow & red). With all of this in mind, we speculate that CylK promotes an S_N2-like reaction with concerted nucleophile and electrophile activation, and has evolved an active site architecture that does not strongly bind free halides and has a higher affinity for the concerted substitution reaction's transition state. In this way, CylK strikes a careful balance between promoting alkyl halide bond breaking while not 'too strongly' activating alkyl halides in a manner more akin to an S_N1-like reaction, which may be prone to halide inhibition. Further evidence against a stepwise or S_N1-like reaction is demonstrated by the lack of alkyl chloride hydrolysis upon exposing a model electrophilic substrate to CylK for 24 h at 37 °C (Figure 3.23C). This might also suggest that the mechanism of BrtB-catalyzed fatty acid alkylation is similarly concerted, and that it requires close positioning of its carboxylate nucleophiles and alkyl halide electrophiles to promote substitution.

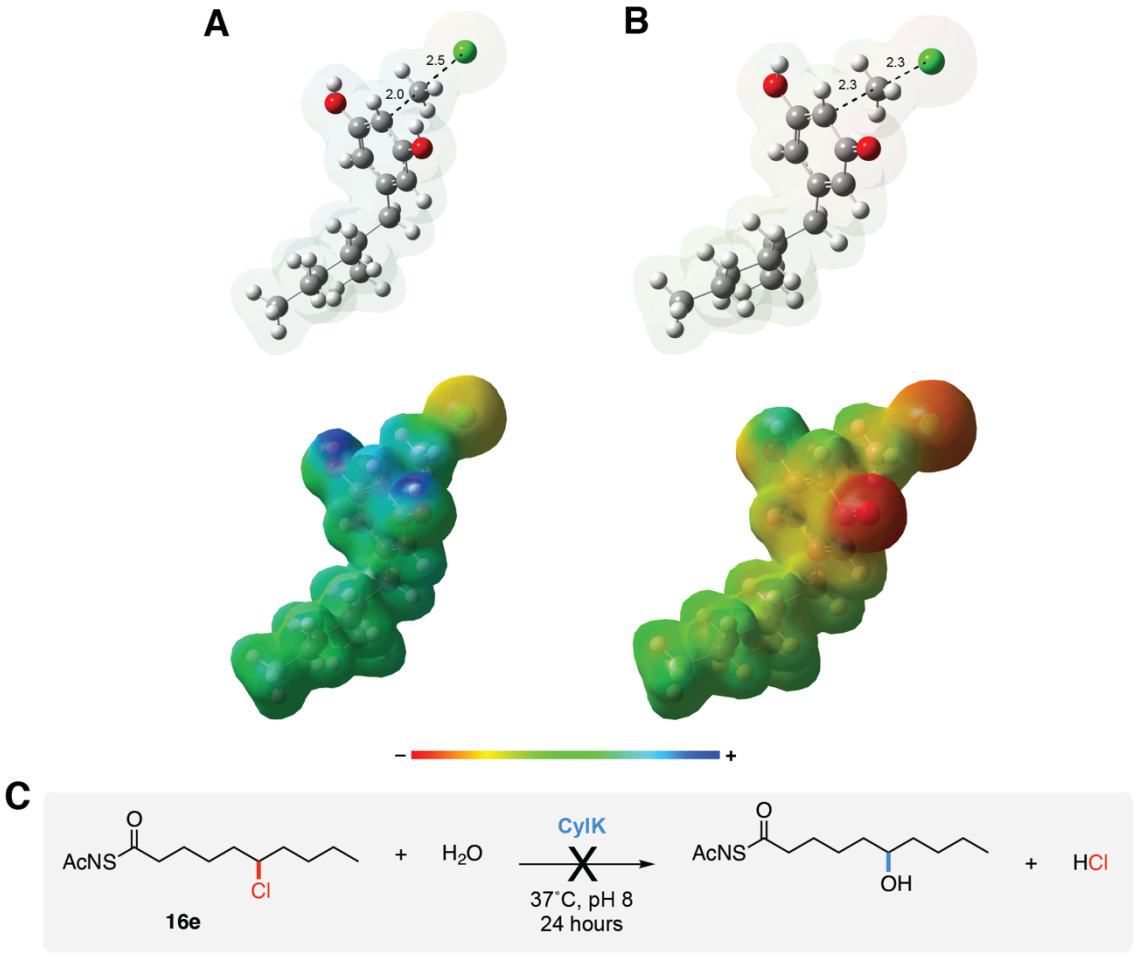


Figure 3.23. Resorcinol–alkyl chloride substitution reaction transition states and electrostatics

(*A*) DFT optimized transition state and electrostatics of a substitution reaction between 5-(heptan-2-yl)-resorcinol and chloromethane. (*B*) Deprotonated version of A. Transition state complexes were generated by scanning the bond length between a deprotonated resorcinol nucleophile and chloromethane electrophile to identify the highest energy complex and optimizing that saddle point (B3LYP, 631GD, D3BJ4), either deprotonated or protonated. Oxygen atoms shown as red spheres, carbons in grey, chlorides in green, and hydrogens in white. (*C*) CylK does not catalyze hydrolysis of alkyl chloride electrophile.

3.2.9 Mechanistic proposal

Based on our structural, mutagenetic, and modeling work – and in agreement with the structures of CylK from Wang et al. – we propose a plausible catalytic mechanism that invokes positioning the reactive partners in close proximity for a concerted S_N2-like reaction, consistent with the known inversion of configuration at the alkyl chloride stereocenter upon substitution (Figure 3.24, complex I). In particular, we suggest that the side chain carboxylates of Asp440 and Glu374 hydrogen bond with the two phenol substituents of the resorcinol nucleophile. By deprotonating one phenol, the essential residue Asp440 enhances the nucleophilicity of the reacting carbon atom. Analogous deprotonation of a resorcinol with a carboxylate side chain has been invoked as a critical step in an enzymatic Friedel–Crafts acylation reaction with acetyl arene electrophiles, lending support to our proposal for resorcinol nucleophile activation.[189] Simultaneously, we hypothesize that Arg105, Tyr473, Phe499, and the hydrophobic portion of Thr84 form an alkyl halide binding pocket as observed in the bromide soaked CylK structure. Hydrogen bonding interactions between the polar residues (Arg105 and Tyr473), and the alkyl halide weaken the C–X bond, reducing the overall activation energy barrier for substitution to occur in a concerted fashion (complex II).

Following alkylation, the halide anion product diffuses out of the active site, and the aromaticity of the resorcinol is restored upon proton transfer to water (complex III), perhaps via His391. Finally, intermediate **2** is released from the active site (complex IV), must reorient 180°, and re-bind for the second intramolecular alkylation to occur in the same fashion. This reorientation likely requires significant energy to disrupt protein-substrate interactions and might suggest active site conformational changes occur to accommodate intermediate **2** in the catalytically competent orientation, however, this warrants further investigation. Although we cannot rule out a radical mechanism, the architecture of the active site and lack of radical initiation sources suggests that one-electron processes are unlikely to be involved in catalysis by CylK.

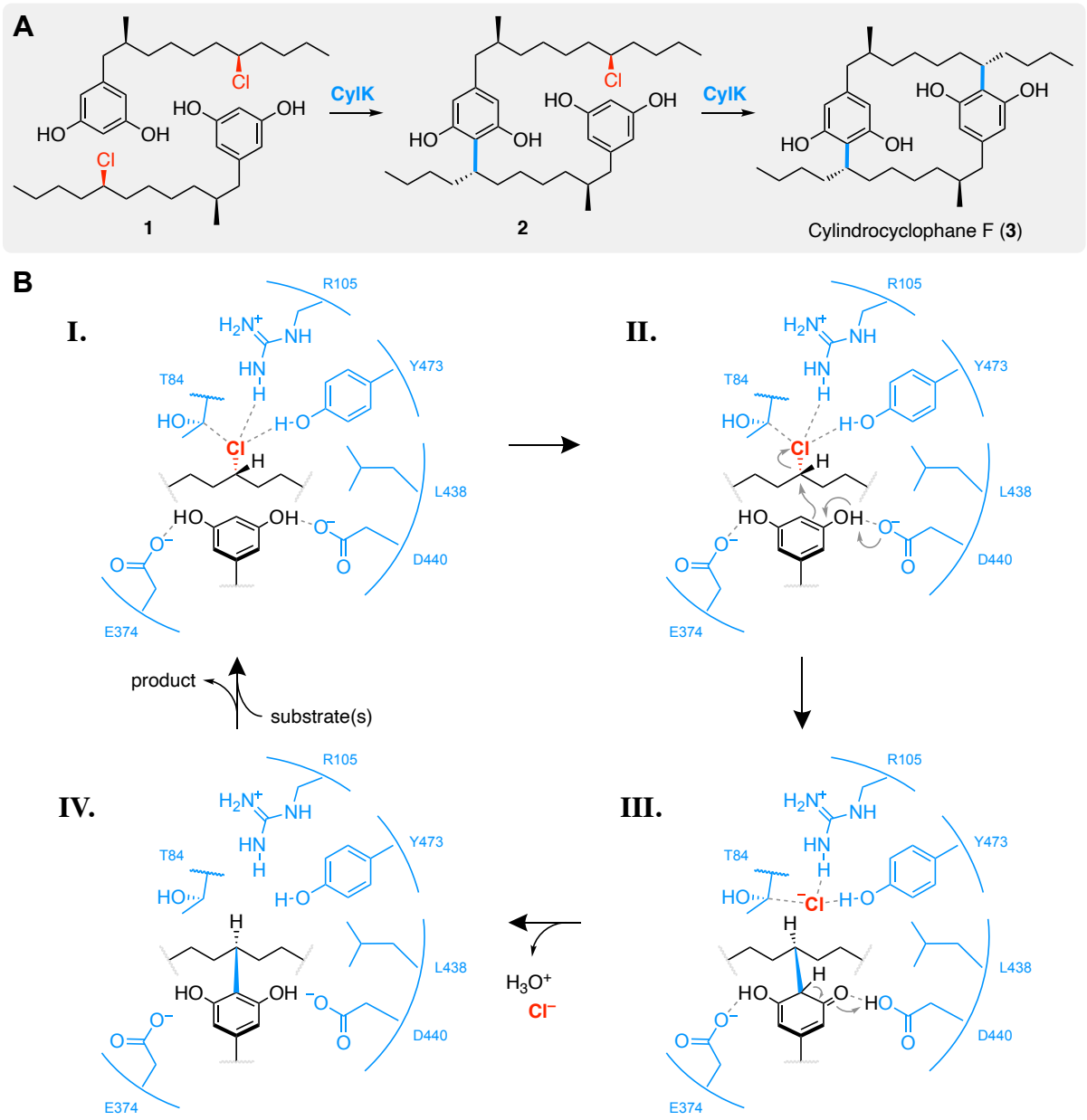


Figure 3.24. Proposed CylK catalytic cycle

(*A*) Native alkylation reactions catalyzed by CylK. This involves an intermolecular reaction between two resorcinol-containing alkyl chloride substrates (**1**) to generate intermediate **2**, followed by release and rebinding of intermediate **2** for a subsequent intramolecular alkylation at the same catalytic site. (*B*) Proposed mechanism for a single cycle of the CylK-catalyzed reaction with key residues illustrated. For clarity, Phe499 is not represented as it sits behind the plane of the page, although it likely interacts with the alkyl chloride. The alkyl chloride electrophile is displayed in red, while CylK residues and the corresponding C–C bond formed are represented in blue.

Interestingly, this alkyl halide binding pocket is reminiscent of the active sites of the SAM-dependent chlorinase[186] and fluorinase[49] enzymes. In these structures, similar ensembles of polar and aliphatic residues bind and stabilize a halide anion as it reacts with SAM in a S_N2 substitution reaction that resembles the reverse of the transformation catalyzed by CylK (Figure 3.25A-B). These crystal structures contain bound alkyl halide and methionine products. In particular, the chlorinase SalL binds alkyl chloride product (5'-chloro-5'-deoxyadenosine) with an apparent hydrophobic interaction between the alkyl chloride and Thr75 and an alkyl chloride–aryl edge interaction with Tyr72 (in addition to polar interactions with backbone amides). Interestingly, the chlorinase reaction can also occur in reverse, suggesting that this binding mode may weaken the C–Cl bond as in CylK.[186] In fact, the reverse reaction is approximately 13x faster than the forward chlorination reaction. Notably, our proposed alkyl chloride binding mode is distinct from that of the well-studied haloalkane dehalogenase enzyme, which primarily uses the indole N–H bonds of adjacent Trp residues to activate a primary alkyl chloride for alkylation by the side chain carboxylate of an active site Asp via hydrogen bonding (Figure 3.25C).[167] We speculate that the differences between these two enzymes may arise from their distinct evolutionary histories and the decreased hydrophobicity and steric factors, and increased reactivity of a primary alkyl chloride substrate.

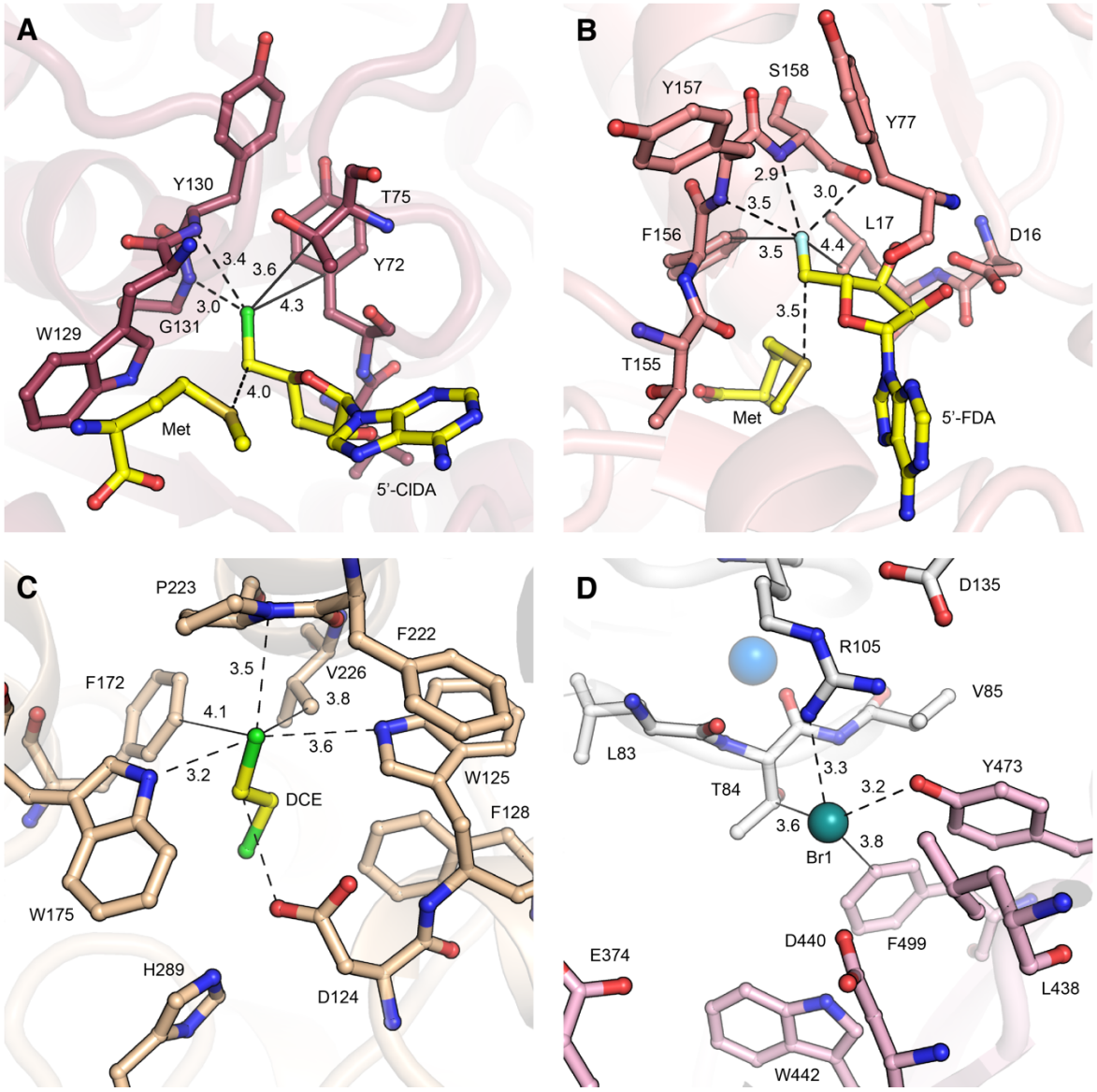


Figure 3.25. Structural comparison of alkyl halides bound in enzyme active sites

(A) 5'-chloro-5'-deoxyadenosine (5'-ClDA) and Met products bound in the active site of wild-type SalL SAM-dependent chlorinase (PDB 2Q6I). The alkyl chloride forms apparent polar interactions (dashed lines) with the backbone amides of Gly131 and Tyr130, and non-polar interactions (solid lines) with Tyr72 and Thr75. *(B)* 5'-fluoro-5'-deoxyadenosine (5'-FDA) and Met products bound in the active site of wild-type FlA SAM-dependent fluorinase (PDB 1RQR). The alkyl fluoride forms apparent polar interactions with the backbone amides of Ser158 and Tyr157, and the side chain of Ser158, and non-polar interactions with Phe156 and Leu17. *(C)* 1,2-dichloroethane (DCE) substrate bound in the active side of wild-type haloalkane dehalogenase (PDB 2DHC). The reactive alkyl chloride forms apparent polar interactions with Trp125, 175, and the backbone amide of Pro223, and non-polar interactions with Phe172 and Val226. Asp124 is the nucleophilic residue that displaces chloride. *(D)* Bromide ion product analog bound in the active site of CylK. The Br^- ion forms polar interactions with Arg105 and Tyr473, and non-polar interactions with Thr84 and Phe499. Select amino acid residues are represented in stick format, bromide ions are shown as teal spheres, calcium ions are shown as blue spheres, alkyl chloride in green, and alkyl fluoride in light aqua.

We have emphasized the distinction between the proposed resorcinol nucleophile-activating residues and those that likely activate the alkyl chloride electrophile in order to compare CylK with the divergent C–O bond forming family member, BrtB (30% amino acid identity, 46% similarity). BrtB catalyzes C–O bond formation between the carboxylate groups of fatty acids and the alkyl chloride-containing bartoloside natural products. We hypothesize that both enzymes utilize a similar strategy to enhance the reactivity of their alkyl chloride substrates, although it remains to be determined if BrtB is similarly stereoinvertive. In agreement with this proposal, BrtB contains identical residues at the Arg105 and Tyr473 positions that may play a role in alkyl chloride activation. In place of the nucleophile-activating Asp440 found in CylK, BrtB contains a functionally similar Glu, and notably, Glu374 in CylK is substituted with a positively charged Arg in BrtB. These two residues might interact with deprotonated carboxylate nucleophiles, although their precise roles remain to be determined. Based on our analysis of CylK, we speculate that BrtB performs an analogous concerted substitution reaction by positioning the reactive partners in close proximity and enhancing the nucleophilicity of a deprotonated carboxylate while simultaneously weakening the C–Cl bond of the alkyl chloride electrophile. The divergent reactivity of CylK and BrtB hints that other relatives might perform unique chemistry and may even be amenable to protein engineering efforts to alter their substrate scopes and reactivity.

3.2.10 Bioinformatic support for proposed mechanism and nucleophile diversity

We next sought to apply the functional insights gained from our structural analysis of CylK to improve our understanding of less closely related family members. Previously, bioinformatically identifying CylK-like enzymes was challenging because proteins encoding β-propeller motifs are common in protein databases, making it unclear which hits were true family members. With the knowledge that both protein domains of CylK are necessary for catalysis, we could now accurately locate CylK-like enzymes encoded in sequenced genomes. From a BLAST search, we identified over 700 proteins with >24% amino acid identity to CylK or BrtB. That group was pared down to 286 unique enzyme sequences containing both N- and C-terminal domains found in CylK, and their relationship was assessed by constructing a maximum-likelihood

phylogenetic tree (Figure 3.26). We further analyzed these data by looking for the presence or absence of a partner CylC halogenase encoded in the same genome. A subset of CylK homologs mostly clustered together on the phylogenetic tree are found in organisms that also have a putative CylC halogenase. Of these CylK and CylC enzyme pairs, ~80% are co-localized in their respective genomes, hinting that they might work together within a biosynthetic pathway. We found very few other types of halogenases co-localized with CylK homologs; those identified were of the iron(II) 2-(oxo)-glutarate (FeII/2OG)-dependent enzyme family but were also clustered with CylC halogenases.[xv] We examined each putative CylK sequence for the presence of the proposed alkyl halide activating residues Arg105 and Tyr473. Intriguingly, nearly all of the candidate CylK enzymes containing both key residues were from organisms that also encode a CylC halogenase (42 of 53 organisms), suggesting that they are likely using halogenated substrates and that Arg105 and Tyr473 are indeed important for alkyl halide activation. We also looked for the presence of Thr84 or the functionally similar Ser to interrogate the mechanism proposed by Wang et al.[141] This mechanism invokes hydrogen bonding between Thr84 and alkyl halide electrophiles, whereas all CylK structures suggest that the hydrophobic portion of Thr84 interacts with halide ions while its hydroxyl group faces away and engages in hydrogen bonds with distal residues. Notably, only 23 of 48 unique sequences with Arg105 and Tyr473 also have Thr84 (or Ser84), leading us to conclude that Thr84 likely does not participate as a key hydrogen bond donor in this putative family of alkyl halide-utilizing enzymes. Instead, other residues with alkyl side chain substituents can likely substitute for the interaction between Thr84 and the alkyl halide electrophile.

[xv] The bioinformatic work to identify putative halogenases was conducted by Cassidy Johnson.

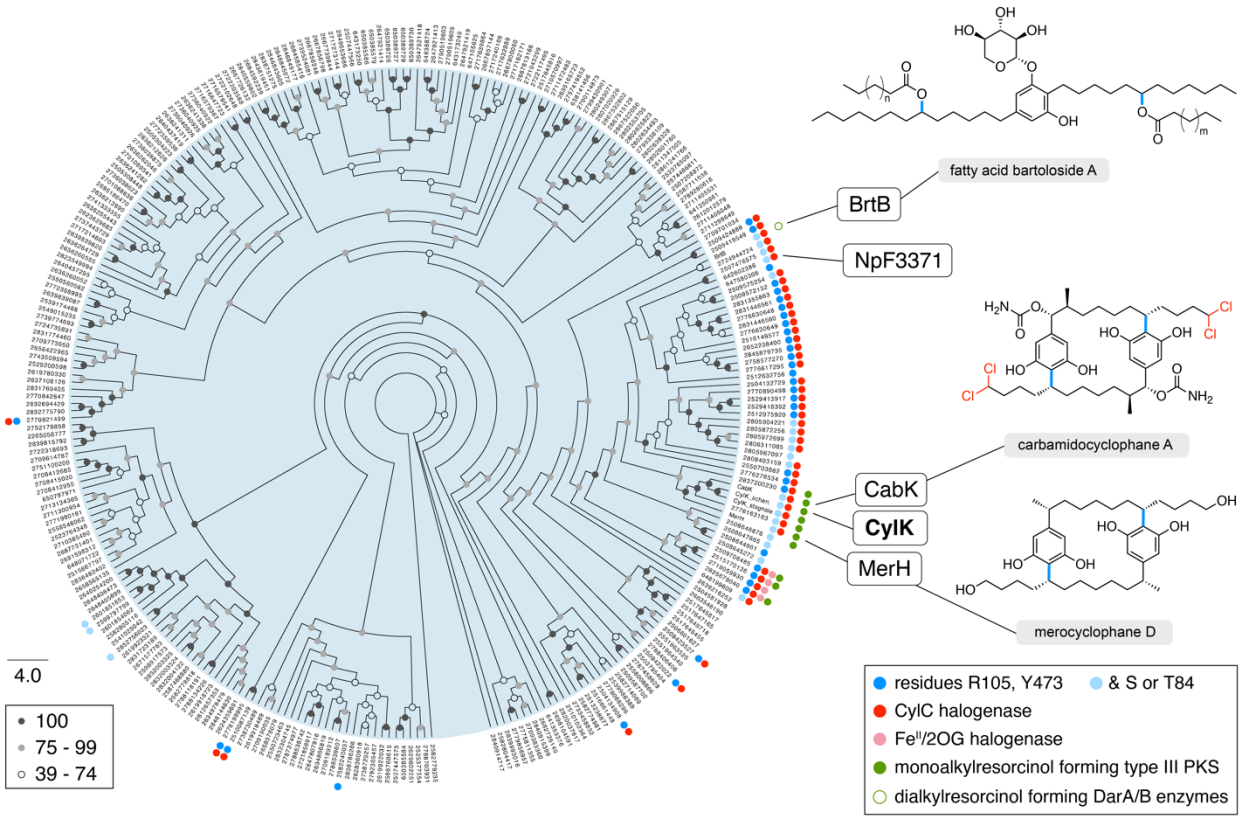


Figure 3.26. Annotated phylogenetic tree of likely alkyl halide-utilizing CylK homologs

Maximum-likelihood phylogenetic tree of protein sequences homologous to CylK and BrtB that contain both N- and C-terminal domains (>24% amino acid identity), highlighting key conserved catalytic residues, clustered halogenases, and associated monoalkylresorcinol forming enzymes. Dialkylresorcinol forming enzymes (DarA/B-like) were not found except as expected with BrtB and the bartoloside producing cluster. Bootstrap values are shaded based on the legend. Natural products associated with select CylK homologs are displayed; the bonds highlighted in blue are constructed by their respective CylK.

Next, the subset of CylK homologs with both Arg105 and Tyr473 residues was aligned and visualized which revealed that, while the resorcinol nucleophile activating residue Asp440 is also highly conserved, Glu374 is highly variable. Thr84, Leu438, Phe499 and other active site residues are not highly prevalent within this subset of homologs (Figure 3.27). This demonstrates that the correlation between Arg105, Tyr473, and the presence of a CylC halogenase is not simply due to the overall similarity of this subset of enzymes. Considering that every CylK homolog from an organism that encodes a CylC halogenase also conserves Arg105 and Tyr473, we believe our assignment of these residues as key alkyl halide activating side chains is correct. Notably, residues surrounding Br2 are highly variable. These and other differences within the active site may suggest that a diversity of substrates can be used by these candidate enzymes.

To further examine the potential for alternate substrate use, we also looked for co-localized monoalkylresorcinol (MAR) forming enzymes encoded near putative CylK homologs in order to ascertain if resorcinols are likely to serve as their native nucleophilic substrates. This search was inspired by the biosynthetic logic of the *cyl* pathway. To our knowledge, biosyntheses of MARs only involves type III PKS enzymes such as CylI.[190] Surprisingly, we found only 10 CylI homologs clustered with CylK homologs, representing just 19% of the candidate enzymes that have Arg105 and Tyr473 residues and are expected to have alkyl chloride activating activity (shown as green circles in Figure 3.26). This observation suggests that a broader diversity of nucleophilic substrates is likely engaged by these uncharacterized alkyl halide-utilizing enzymes. Of note, dialkylresorcinol forming enzymes such as those involved in bartoloside biosynthesis were not found in CylK-encoding gene clusters beyond the bartoloside gene cluster. Furthermore, the observation that the majority of CylK homologs do not have proposed alkyl chloride activating residues, partner halogenases, nor resorcinol forming enzymes, indicates that this enzyme family may catalyze other reactions not involving alkyl halides or resorcinols. This information will enable future studies by prioritizing organisms that might produce more distantly related natural products and/or CylK-like enzymes with alternate substrate scopes.

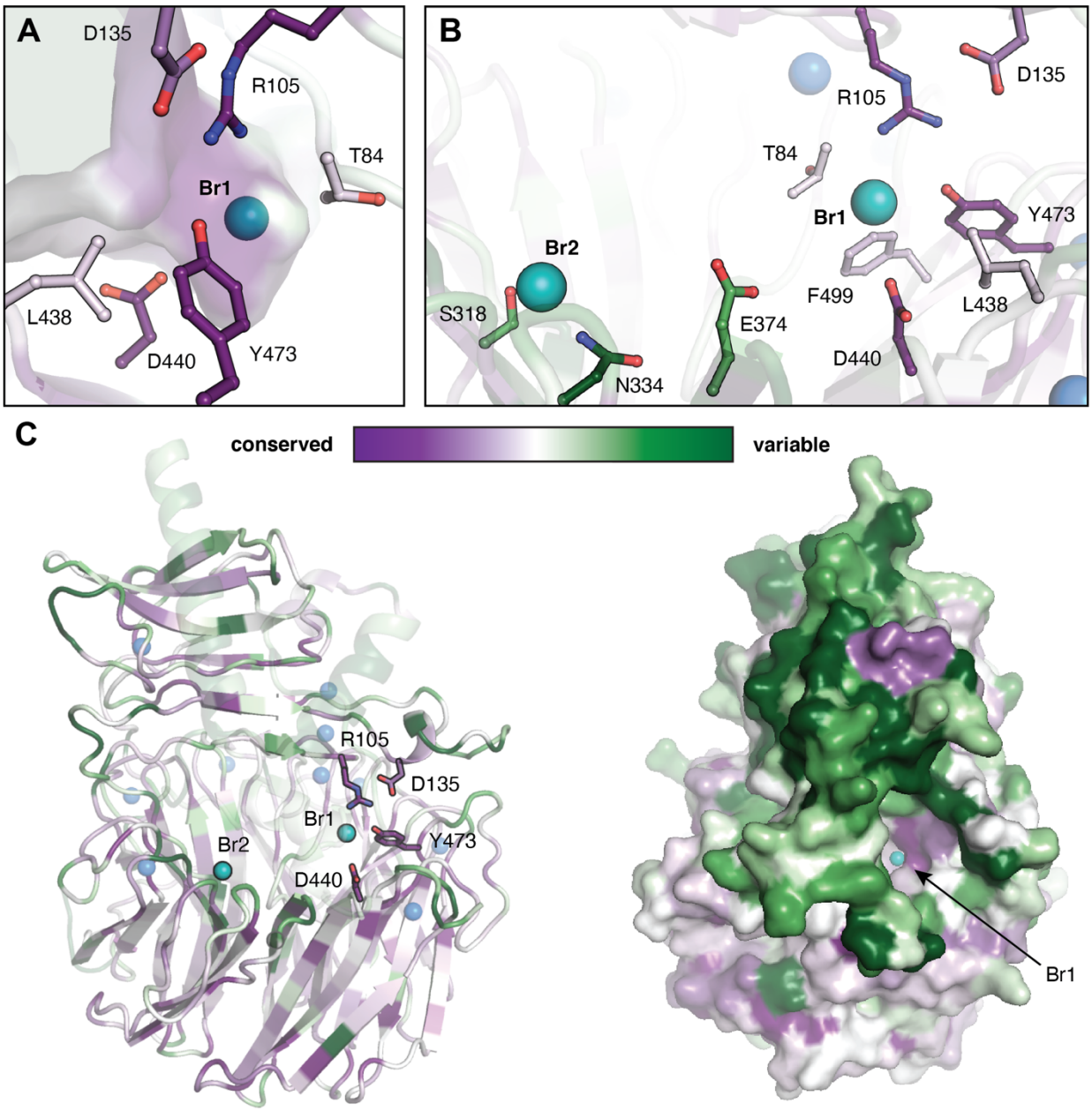


Figure 3.27. Conservation mapping analysis of likely CylK homologs

Representations of sequence conservation generated in ConSurf[191] using a subset of cyanobacterial CylK homologs that have Arg105 and Tyr473 residues (48 unique sequences). Of those homologous enzymes, 37 contain a partner CylC halogenase in the genome of the encoding organism (~80% are co-localized with CylK). *(A)* The alkyl halide binding pocket. *(B)* Residues surrounding Br1 and Br2. *(C)* Ribbon and surface diagrams of overall sequence conservation in this subset of homologs. Select amino acid residues are represented in stick format, bromide ions are shown as teal spheres, and calcium ions are shown as blue spheres. Arrow denotes Br1 and the opening of the interdomain cavity containing the active site of CylK (cavity 1). This figure was adapted from a figure generated by Terry B. Ruskoski.

In support of our speculation that the family of CylK-like alkyl halide-utilizing enzymes use diverse nucleophilic substrates, Dr. Nathaniel R. Glasser has identified a chlorinated glycolipid natural product from *Nostoc punctiforme* ATCC 29133 that is produced by a biosynthetic gene cluster containing CylK and CylC homologs, NpF3370 and NpF3358, respectively (Figure 3.28). Production of the glycolipid is eliminated if the entire cluster is genetically deleted, while selectively knocking out the CylK homolog, NpF3370, increases the levels of this metabolite. This suggests that the chlorinated lipid may be an intermediate towards a larger product, and it is likely a substrate for NpF3370. Note, this CylK-like enzyme can be found in our phylogenetic tree, and it contains key alkyl halide activating residues Arg105 and Tyr473. With this information, we speculated that NpF3370 analogously uses alkyl halide electrophiles for substitution chemistry, however it was unclear what the nucleophilic functional group might be as there is no resorcinol nor carboxylate group in this metabolite. Furthermore, although C_{16} and C_{18} fatty esterified derivatives of the glycolipid were detected and isolated, the chloride group remained intact in these molecules and this ester linkage was appended to a hydroxyl group on the lipid backbone.

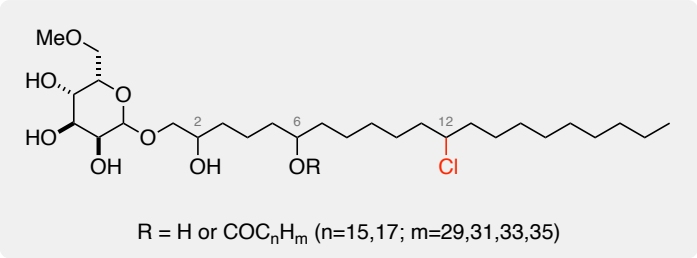


Figure 3.28. Chlorinated glycolipids from *N. punctiforme* ATCC 29133

The structure proposed are based on LC-MS and NMR characterization by Dr. Nathaniel R. Glasser. Stereochemical configurations at carbons 2, 6, and 12 are unknown.

In order to determine if NpF3370 had alkyl halide-utilizing activity, heterologous expression of this CylK homolog was attempted. Initial expression tests were unsuccessful, however, we inspected our multiple sequence alignment of NpF3370 and CylK and noticed that NpF3370 has a ~200 residue C-terminal extension beyond the terminus of CylK's 7-bladed β-propeller domain. Of note, both BrtB and MerH have analogous C-terminal extensions of ~80 and ~260 amino acids, respectively. There is some conservation between the C-terminal extensions of NpF3370 and MerH, and they are predicted to contain a repeat-in-toxin (RTX) motif similar to CylK's N-terminus (CylK residues 104-109). However, it is unknown what function(s) these C-terminal extensions may have *in vivo*. Truncating this extension from the heterologous expression construct afforded soluble NpF3370, and a crude preparation of the enzyme could oligomerize the glycolipid up to eight times (Figure 3.29). MS fragmentation data suggests that one hydroxyl group of the lipid backbone serves as the nucleophilic substrate, which displaces an alkyl chloride electrophile of a second equivalent of lipid to form an ether linkage. Furthermore, identical dimeric and trimeric species could be detected from cultures of *N. punctiforme*. A head-to-tail cyclized product could not be detected. Notably, NpF3370 has identical residues at the Thr84, Arg105, Tyr473, and Phe499 positions that likely play a role in alkyl chloride activation. In place of the nucleophile-activating Asp440 found in CylK, NpF3370 contains a functionally similar Glu (similar to BrtB), while Glu374 in CylK is substituted with Gly in NpF3370. The precise roles of these residues in NpF3370 remain to be determined. This finding suggests that CylK homologs catalyze at least three distinct substitution reactions involving alkyl halide electrophiles, and we hypothesize this family may have additional reactivity yet to be discovered.

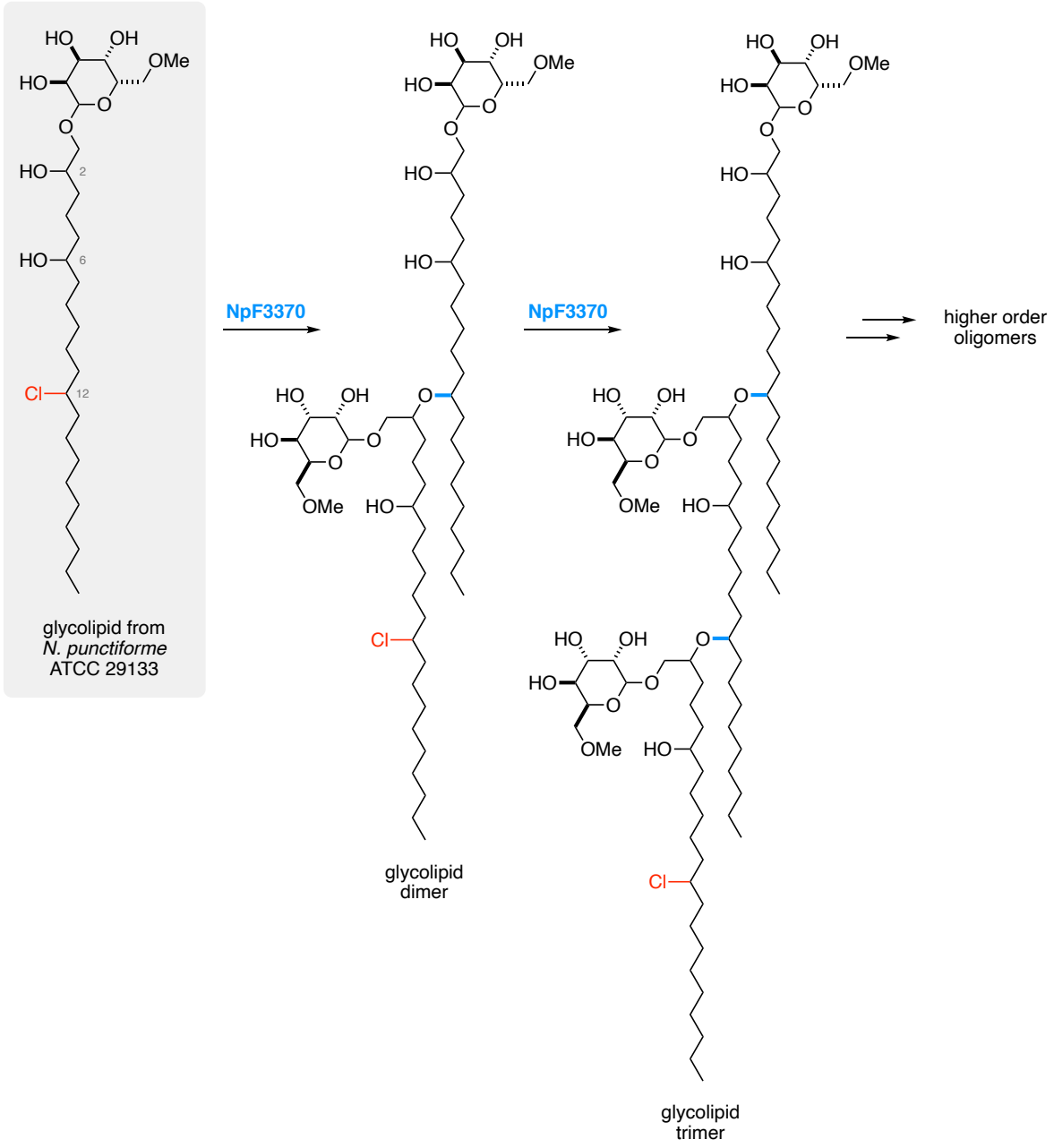


Figure 3.29. CylK homolog from *N. punctiforme* oligomerizes glycolipid with hydroxyl nucleophile

Putative oligomeric structures constructed by CylK homolog NpF3370. Hydroxyl group at carbon 2 represented as nucleophilic species, however an ester linkage at carbon 2 or 6 could not be distinguished by MS fragmentation data. Dimeric and trimeric species were most abundant, although higher order oligomers were detected *in vitro*.

3.3 Conclusions

Our results provide a structural basis for understanding the unusual enzymatic Friedel–Crafts alkylation catalyzed by CylK in cylindrocyclophane biosynthesis. This structural model, along with supporting biochemical experiments and bioinformatic analysis, has provided fundamental insight into this intriguing reaction that represents the only known example of enzymatic aromatic ring alkylation with an alkyl halide electrophile. We identified the active site of CylK at the interface of its N- and C-terminal domains. Together with previously characterized enzymes, our results show how structurally related β-propeller enzymes have divergently adapted this fold for catalytic function. For instance, in two metalloenzymes, carotenoid cleavage dioxygenase (CCD)[192] and nitrous oxide reductase (NOR),[193] an active site metallocofactor is lodged on the top face of the propeller and coordinated by His side chains from inner β-strands of the fold (Figure 3.30). Like CylK, both of these enzymes have a capping motif on the top face of the propeller. CCD contains more extensive internal loops than CylK, and these adopt structured helix-loop motifs that bury the top face of the propeller. NOR is a di-domain dimeric enzyme that employs an electron transfer domain from an adjacent monomer to shield the catalytic metallocofactor in the β-propeller motif. In both, these domain arrangements create a favorable environment for substrate binding and catalysis, as we observe in CylK.

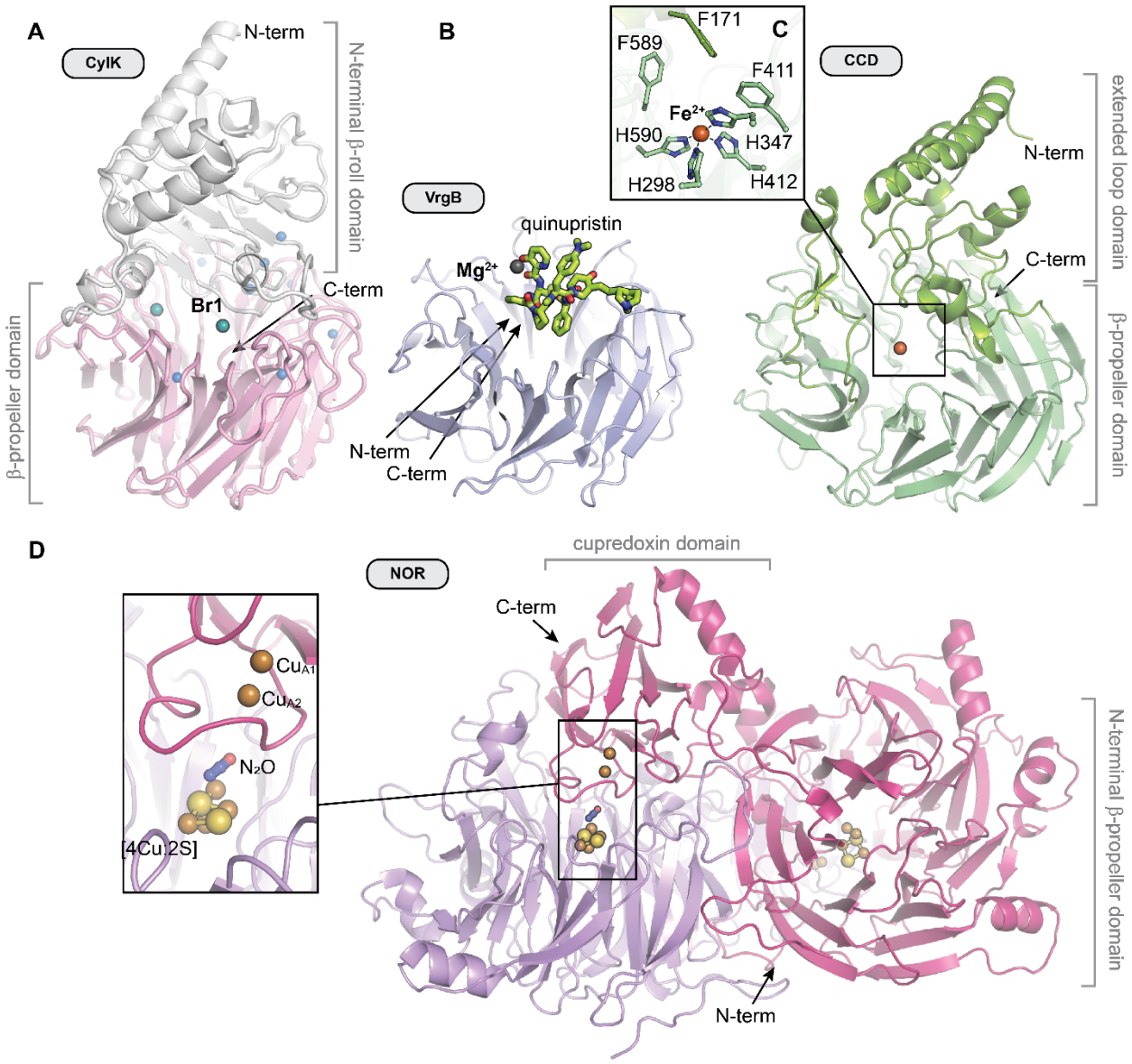


Figure 3.30. Comparison of 7-bladed β-propeller enzymes

(A) CylK is a fusion of a Ca^{2+}-binding β-roll fold and a β-propeller with the active site located at the domain interface. *(B)* Virginiamycin B lyase (VrgB) is a single domain enzyme in which the active site is located on the top face of the propeller. Quinupristin substrate analog shown as green sticks (PDB 2Z2P). *(C)* The extended loops of carotenoid cleavage dioxygenase (CCD) form a helical domain on the top face of the propeller, oriented similarly to the N-terminal domain of CylK. The active site, delineated by the location of the iron cofactor (orange sphere), is located at the interface of these two domains (PDB 3NPE). *(D)* Dimeric nitrous oxide reductase (NOR) contains an active site at the interface between a C-terminal cupredoxin electron transfer domain (Cu_A) contributed by the other monomer and the β-propeller catalytic domain (PDB 3SBR). Unless otherwise noted, ions are shown as spheres and colored as in Figure 3.7. This figure was adapted from a figure generated by Terry B. Ruskoski.

Both domains of CylK have multiple bound calcium ions, a row of which are near the proposed alkyl chloride binding site. However, the closest Ca^{2+} ion is approximately 8 Å from this pocket. Although a functional requirement for calcium has been previously demonstrated,[69] our structural work suggests that a direct role in catalysis is unlikely. Removing calcium might cause a significant conformational change that precludes access to or alters the architecture of the active site, as the treatment of CylK with EDTA was observed to inactive the enzyme and cause a shift in oligomeric state from monomer to a dimer. This might be indicative of structural changes within each unit of CylK that are incompatible with catalysis. Structural homologs of CylK's N-terminal Ca^{2+} binding fold include RTX domains that may facilitate Ca^{2+}-dependent folding (see Chapter 3.2.2 for examples and further discussion). It is unknown if calcium plays a regulatory role for CylK activity and cylindrocyclophane biosynthesis *in vivo* although it is a component of the cyanobacterial BG-11 growth medium.

The reaction catalyzed by CylK represents a unique approach for enzymatic C–C bond formation. Other biological strategies to form analogous alkyl–aryl C–C bonds involve radical intermediates that must be precisely held and oriented in their enzyme active sites in order to prevent side reactions. Examples include the radical SAM and Fe^{II}/2OG enzymes involved in streptide[194] and etoposide[195] biosynthesis, respectively. In contrast to CylK, these enzymes form bonds between arenes and unactivated carbon centers. Because the target C(sp^3) coupling sites of their substrates are inherently unactivated, these alkylating enzymes must stringently discriminate between the alkyl C–H coupling site and its chemically equivalent neighbors. In contrast, the CylK catalyzed reaction leverages the preinstalled alkyl chloride substituent in order to direct reactivity and stereochemistry. Therefore, the cylindrocyclophane biosynthetic logic can be described as dividing the task of determining reaction specificity between CylK and its partner halogenase, CylC. This novel halogenase family likely has strict requirements for substrate positioning,[196] although it remains to be seen if CylC is promiscuous. This feature of distributing selectivity and alkylation across two enzymes might rationalize why wild-type CylK can accept multiple non-native substrates, whereas active site specificity for unactivated coupling partners might be more stringent.

CylK performs a challenging chemical reaction without precedent in biology. By elucidating its structure, we have begun to unravel the molecular details underlying the use of alkyl halides as cryptic intermediates in natural product biosynthesis. This enzymatic Friedel–Crafts alkylation might also find application in constructing diverse alkyl–aryl C–C bonds outside of natural pathways. With structural information about CylK, notably the identification of active site residues, efforts to further expand its substrate specificity and improve its stability via enzyme engineering can begin in earnest, and preliminary work towards these goals is reported in Chapter 4. In addition to its unique reactivity, the large size and solvent accessibility of CylK's active site suggest it might be engineered to accept diverse and bulky substrates. Furthermore, our structure and proposed mechanism have increased our fundamental understanding of enzymatic interactions with alkyl halides and may inspire the design of biomimetic chemocatalysts that engage secondary alkyl halides in substitution reactions.[197,198] Now armed with the ability to functionally annotate CylK-like genes, we can prioritize orphan biosynthetic gene clusters and producing organisms in order to discover additional structurally unique natural products and enzymes, as demonstrated by Dr. Nathaniel R. Glasser's work discovering the chlorinated glycolipids and distinct reactivity from CylK-homolog NpF3370. These new natural products may be candidate therapeutics and/or play important physiological roles in cyanobacteria.

Exciting future directions for this work include interrogating aspects of our proposed CylK mechanism, especially as it relates to homologs such as BrtB and NpF3370 which use different nucleophilic substrates. It remains to be determined if CylK can use carboxylate or hydroxyl group nucleophiles, and if the other two enzymes have any analogous promiscuity. Mutagenetic work that elucidates the active site determinants of this differential reactivity would be revealing. Along the same lines, further characterization of CylK mutants at residues Thr84 and Phe499 might enhance our understanding of the mode of alkyl halide activation. Additional crystallography work excluding sources of chloride (or bromide) might enable access to structures with substrate analogs bound in the reactive conformation, and our preparation of CylK with $NaNO_3$ instead of NaCl may be a good starting point for such studies. Supplementation with non-chloride-containing calcium and magnesium sources may be beneficial to improve this preparation. Additionally, any

potential protein movement in solution may be interesting to explore with site specific labelling and fluorescence spectroscopy,[199] and might be relevant for substrate binding, re-binding, and catalysis. Cryo-electron microscopy might also enable structural visualization of different protein conformations of the CylK catalytic cycle.[200] The different crystal forms of CylK have distinct N-terminal domain arrangements relative to the C-terminal β-propeller domain, and this may be indicative of states accessible to the protein in solution. Finally, the role of calcium could be potentially probed further by attempting growth experiments with the native producers with limiting calcium, or by mutating calcium binding sites, as well as solving structures of apoCylK.

3.4 Materials and Methods

3.4.1 DFT Calculations

Calculations were performed at Harvard University using Gaussian 09 with the B3LYP functional, with default convergence and integration settings. For each molecule or complex of interest, the geometry of the lowest-energy conformation was optimized with the 6-31G(d) basis set, D3BJ4 empirical dispersion correction, and polarizable continuum model (PCM) for Et_2O. Transition state complexes were generated by scanning the bond length between a deprotonated resorcinol nucleophile and chloromethane electrophile to identify the highest energy complex and optimizing that saddle point (B3LYP, 631GD, D3BJ4), either deprotonated or protonated. Checkpoint files were generated by including the keywords %chk=g09.chk and pop=(chelpg,npa), and subsequently converted to formatted checkpoint files, from which .cube files containing electron density, electrostatic potential, and highest occupied molecular orbital (HOMO) data were generated. HOMOs were visualized with GaussView.

3.4.2 Refolding and purification of apoCylK-Strep

CylK without bound Ca^{2+} (apoCylK) was expressed and purified as described previously (see Materials and Methods 2.4.6) with several modifications detailed below. Following overexpression in *Escherichia coli* BL21(DE3) as a Sumo fusion construct, cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet. Cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The pellet was collected by centrifugation (20,000 × *g* for 30 min) and resuspended in lysis buffer supplemented with 6 M urea and 5 mM imidazole. The denatured enzyme solution was clarified twice by centrifugation (20,000 × *g* for 15 min), and the supernatant was applied to 20 mL of Ni NTA resin. The column was washed with lysis buffer containing 6 M urea and 5 mM imidazole. Bound species were eluted with lysis buffer containing 6 M urea and 250 mM imidazole. All subsequent steps were carried out at 4 °C. The enzyme was refolded at a concentration of 1-10 μM by sequentially dialyzing this solution against buffer 1 (20 mM Tris pH 8.0, 500 mM NaCl, 3 M urea) overnight, buffer 2 (20 mM Tris pH 8.0, 500 mM NaCl) for 4 hr, and buffer 3 (20 mM Tris pH 8.0, 500 mM NaCl, 5% v/v glycerol) for 4 hr. Following dialysis, the refolded enzyme solution

was centrifuged (100,000 × *g* for 30 min), and the supernatant was collected and incubated overnight with ULP protease (1:25 w/w, see Materials and Methods 2.4.8 for ULP preparation). Then, following a subtractive Ni NTA step, the enzyme was applied to a Superdex 200 26/600 gel filtration column (GE Healthcare) equilibrated in protein storage buffer (20 mM HEPES pH 7.8, 50 mM NaCl, 10% glycerol). The fractions containing pure CylK were pooled, concentrated (50–60 μM, determined by extinction coefficient and absorbance at 280 nm), and flash frozen in liquid nitrogen. This procedure afforded 2.4 mg of apoCylK-Strep.

3.4.3 Competitive lanthanide substitution

Aliquots of apoCylK-Strep were diluted to 0.9 μM in 5.722 mL of protein storage buffer (20 mM HEPES pH 7.8, 50 mM NaCl, 10% glycerol) and $YbCl_3$, $TmCl_3$, and $TbCl_3$ were successively added in four 25 μL portions over 30 min for a final concentration of 3.0 μM for each lanthanide. This sample was incubated overnight and at 4 °C. No precipitation was observed. Lanthanide treated apoCylK was then buffer exchanged by successive centrifugal concentration/dilution with protein storage buffer. Following buffer exchange, theoretical final concentrations of unbound lanthanides were ~0.4 nM while the enzyme was 21.6 μM. Samples and standards for ICP-OES was prepared by diluting 60 μL of these solutions into 5.760 mL water supplemented with 180 μL nitric acid, and stored at –20 °C until running ICP-OES at the Quantitative Bio-element Imaging Center at Northwestern University.

3.4.4 Refolding and purification of CylK-Strep for crystallography

CylK was expressed and purified for crystallographic characterization as described previously (see Materials and Methods 2.4.6) with several modifications detailed below. Following overexpression in *Escherichia coli* BL21(DE3) as a Sumo fusion construct, cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet. Cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The pellet was collected by centrifugation (20,000 x *g* for 30 min) and resuspended in lysis buffer supplemented with 6 M urea and 5 mM imidazole. The denatured enzyme solution was clarified twice by

centrifugation (20,000 x *g* for 15 min), and the supernatant was applied to 20 mL of Ni NTA resin. The column was washed with lysis buffer containing 6 M urea and 5 mM imidazole. Bound species were eluted with lysis buffer containing 6 M urea and 250 mM imidazole. All subsequent steps were carried out at 4 °C. The enzyme was refolded at a concentration of 1-10 μM by sequentially dialyzing this solution against buffer 1 (lysis buffer supplemented with 3 M urea) overnight, buffer 2 (lysis buffer lacking urea) for 4 h, and buffer 3 (lysis buffer supplemented with 5% glycerol) for 4 h. Following dialysis, the refolded enzyme solution was centrifuged (100,000 x *g* for 30 min) and the supernatant was collected and incubated overnight with ULP Protease (1:25 w/w, see 2.4.8 for ULP preparation). Then, following a subtractive Ni NTA step, the enzyme was applied to a Superdex 200 26/600 gel filtration column (GE Healthcare) equilibrated in protein storage buffer (20 mM HEPES pH 7.8, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol). The fractions containing pure CylK were pooled, concentrated (500-700 μM, determined by extinction coefficient and absorbance at 260 nm) and flash frozen in liquid nitrogen. This procedure afforded 4.5 mg of CylK-Strep per L of culture.

3.4.5 Refolding and purification of chloride-free CylK-Strep

CylK was expressed and purified for halide inhibition experiments as described previously (see Materials and Methods 2.4.6) with several modifications detailed below. Following overexpression in *Escherichia coli* BL21(DE3) as a Sumo fusion construct, cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) supplemented with an EDTA-free Pierce Protease Inhibitor Tablet. Cells were lysed using a continuous-flow homogenizer (Avestin EmulsiFlex-C3). The pellet was collected by centrifugation (20,000 x *g* for 30 min) and resuspended in lysis buffer supplemented with 6 M urea and 5 mM imidazole. The denatured enzyme solution was clarified twice by centrifugation (20,000 x *g* for 15 min), and the supernatant was applied to 20 mL of Ni NTA resin. The column was washed with lysis buffer containing 6 M urea and 5 mM imidazole. Bound species were eluted with lysis buffer containing 6 M urea and 250 mM imidazole. All subsequent steps were carried out at 4 °C. The enzyme was refolded at a concentration of 1-10 μM by sequentially dialyzing this solution against buffer 1 (lysis buffer supplemented with 3 M urea) overnight, buffer 2 (lysis buffer lacking urea) for 4 h, and buffer 3 (lysis buffer supplemented

with 5% glycerol) for 4 h. Following dialysis, the refolded enzyme solution was centrifuged (100,000 x *g* for 30 min) and the supernatant was collected and incubated overnight with ULP Protease (1:25 w/w, see 2.4.8 for ULP preparation). Then, following a subtractive Ni NTA step, the enzyme was dialyzed twice against chloride-free buffer (20 mM HEPES pH 7.8, 100 mM $NaNO_3$, 10% glycerol). Following dialysis, chloride concentration was calculated to be less than 0.5 μM. Then, the chloride-free enzyme was applied to a Superdex 200 16/600 gel filtration column (GE Healthcare) equilibrated in chloride-free storage buffer (20 mM HEPES pH 7.8, 100 mM $NaNO_3$, 10% glycerol). Size exclusion chromatography revealed that this chloride-free CylK-Strep preparation was a 1:1 mixture of monomeric and dimeric species (Figure 3.31). The fractions containing pure CylK were pooled, concentrated (150-200 μM, determined by extinction coefficient and absorbance at 260 nm) and flash frozen in liquid nitrogen. This procedure afforded 2.5 mg of CylK-Strep per L of culture.

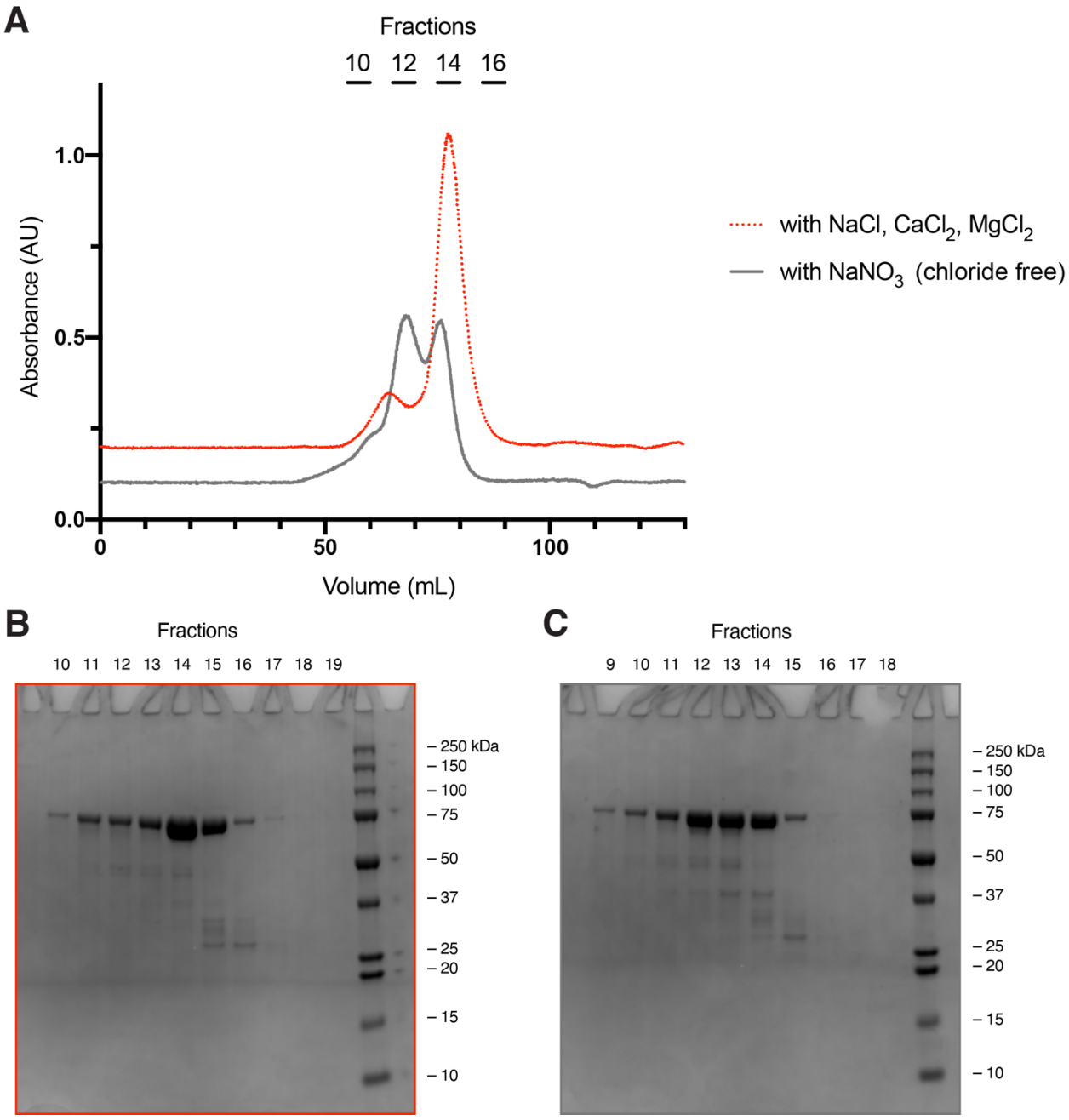


Figure 3.31. Size exclusion chromatography and SDS-PAGE of chloride-free CylK

(*A*) Size exclusion chromatograms of chloride-free CylK (grey solid line) or in standard buffer (red dotted line). Peak at ~68 mL likely corresponds to dimeric species, peak at ~75 mL is monomeric CylK. Absorbance measured at 280 nm. SDS-PAGE of (*B*) standard buffer CylK or (*C*) chloride-free CylK. All Blue Precision Plus Protein Prestained Standard (Bio-Rad) was used to visualize molecular weight standards. Size exclusion chromatography run on Superdex 200 16/600 gel filtration column (GE Healthcare).

3.4.6 Crystallization and structure determination

This work was conducted by Terry B. Ruskoski with contributions from Nathaniel R. Braffman, Katherine M. Davis, Nathaniel R. Glasser, Andrew J. Mitchell, and Jonathan A. Bergman; and supervised by Amie K. Boal. Prior to crystallization trials, purified CylK protein in storage buffer was combined with 5-undecylbenzene-1,3-diol (**7**), a substrate analog synthesized as described previously[128] (also see Materials and Methods 2.4.11) and dissolved in protein storage buffer supplemented with 22% DMSO (v/v). The final protein solution contained 10 mg/mL CylK and 1.66 mM of **7**. Crystals were obtained by using the hanging drop vapor diffusion method in 2 µL drops mixed in 1:1 ratio with a precipitant solution of 1.8 M sodium malonate, pH 7.0. To aid in obtaining phase information, the native Ca(II) sites in CylK crystals were substituted with a lanthanide via soak in a solution of 2.0 M sodium malonate, 100 mM terbium(III) chloride for 8 min at room temperature. The Tb-soaked crystals were transferred to a cryoprotectant solution containing 2.0 M sodium malonate supplemented with 20% ethylene glycol, mounted on rayon loops, and flash frozen by direct plunge into liquid nitrogen. Diffraction datasets were collected at 1.6314 Å (Tb anomalous) and 0. 97872 Å (native) on these crystals at beamlines 23-ID-B [General Medical Sciences and Cancer Institutes Collaborative Access Team (GM/CA-CAT), Advanced Photon Source (APS)] and 21-ID-F [Life Sciences Collaborative Access Team (LS-CAT), APS], respectively. Diffraction datasets were processed with the HKL2000 software package.[201] The structure was solved by using the MR-SAD method. PHASER-EP,[202,203] as implemented within the PHENIX software package,[204] was used to calculate initial phases. A poly-alanine model generated from the core β-propeller domain of *Staphylococcus cohnii* streptogramin B lyase (PDB accession code 2QC5, with all loops and secondary structure connections truncated)[164] was used as the initial search model for molecular replacement. PHENIX.autobuild generated an initial model containing 504 residues in 15 chain fragments with initial R_{work}/R_{free} values of 29.9%/37.4%. The model was iteratively improved via refinement in Refmac5 against a native dataset and model building in Coot,[205] yielding R_{work}/R_{free} values of 15.6%/18.9% (Table 3.4). The final model contains residues 7-45, 49-392, and 413-662 with 11 Ca^{2+} ions, 2 Mg^{2+} ions, 1 Cl^- ion, and 339 water molecules. Tb ions were not modeled because they were not present at high occupancy. Ramachandran analysis[204] indicates a single outlier, Val620. Although **7** was required for CylK crystallization, and weak electron density resembling the

compound could be found at several points near the surface of the protein, likely mediating crystal lattice contacts, we could not confidently model the alkyl resorcinol. No density for **7** could be identified in cavity 1, the putative active site of CylK.

To identify possible alkyl halide-binding sites within the protein, CylK crystals were soaked in 2 M sodium malonate with 500 mM NaBr for up to 1 min. The Br-soaked crystals were transferred to a cryoprotectant solution consisting of the soak solution supplemented with 20% ethylene glycol. Diffraction datasets were collected at 0.9184 Å (Br anomalous) and 0.97872 Å (native) on these crystals at beamlines 23-ID-B (GM/CA-CAT, APS) and 21-ID-F (LS-CAT, APS), respectively. Phase information was obtained by molecular replacement using the CylK model as the initial search model in Phaser MR, implemented within CCP4. The model was iteratively built and refined in Coot and Refmac5, respectively, resulting in a R_{work}/R_{free} of 16.9%/19.8% (Table 3.7). The final model contains residues 7-393 and 410-667 with 10 Ca^{2+} ions, 2 Mg^{2+} ions, 1 Na^{+} ion, 4 Br^{-} ions, and 403 water molecules. Anomalous maps were generated using CAD and FFT programs implemented within CCP4. Figures were generated in PyMOL (Schrödinger, LLC). The information in Table 3.4 and Table 3.7 was generated using HKL2000 (data collection statistics), Refmac5 (refinement statistics including resolution limits, R_{work}/R_{free}, rms deviations), the PBD validation server (no. of atoms), and PHENIX (*B*-factors, Molprobity clashscore, rotamer outliers, Ramachandran statistics) software packages.

Table 3.4. Data collection and refinement statistics for Tb-soaked CylK structures

	Tb anomalous	Native (final model)
Data collection		
Space group	*C* 2 2 2_1	*C* 2 2 2_1
Wavelength (Å)	1.6314	0.97872
Cell dimensions		
a, *b*, *c* (Å)	94.265, 139.286, 100.381	93.667, 138.613, 99.438
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50-2.38 (2.42-2.38)	50.0-1.68 (1.71-1.68)
R_{merge}	0.218 (0.670)	0.101 (1.326)
R_{pim}	0.041 (0.192)	0.026 (0.346)
$I / \sigma I$	25.2 (3.8)	29.3 (2.15)
$CC_{1/2}$	1.010 (0.906)	0.999 (0.780)
Completeness (%)	99.5 (99.9)	99.8 (99.3)
Redundancy	29.5 (11.4)	14.8 (14.3)
Refinement		
Resolution (Å)		42.41 - 1.68
No. reflections		68,476
R_{work} / R_{free}		0.1556/0.1894
No. atoms		5292
Protein		4903
Ligand/ion		50
Water		339
B-factors (Å^2)		
Protein		18.20
Ligand/ion		28.54
Water		25.82
R.m.s. deviations		
Bond lengths (Å)		0.016
Bond angles (°)		1.921
Molprobity clashscore		2.29 (99th percentile)
Rotamer outliers (%)		0.78
Ramachandran favored (%)		96.33
Ramachandran allowed (%)		3.51
Ramachandran outlier† (%)		0.16
PDB accession code		7RON

*Values in parentheses are for highest-resolution shell.

Table 3.5. Structural homologs of the N-terminal Ca^{2+} binding domain of CylK

Rank	PDB ID	Functional description	Z-score	Rmsd	No. residues aligned	Total residues
1	1OMJ-A	Psychrophilic alkaline metalloprotease (PAP); serralysin family	11.2	6.8	135	456
2	2Z8Z-A	MIS38 lipase	11.1	8.5	116	616
3	6SUS-A	RTX domain of blocks IV and V of adenylate cyclase toxin	11.0	3.1	120	258
4	2QUA-A	LipA, extracellular lipase from Serratia marcescens	10.6	2.3	110	615
5	2ML1-A	AlgE6R1, Mannuronan C5-epimerase	10.6	3.0	122	153

This table was populated by Terry B. Ruskoski with a structural comparison to other proteins in the PDB.[206]

Table 3.6. Structural homologs of the C-terminal β-propeller domain of CylK

Rank	PDB ID	Functional description	Z-score	Rmsd	No. residues aligned	Total residues
1	1SQ9-A	Antiviral protein Ski8	31.2	2.4	286	378
2	4H5J-B	Guanine nucleotide-exchange factor Sec12	31.1	2.9	300	347
3	5TF2-A	Prolactin regulatory element-binding protein	31.0	2.7	291	338
4	4J0X-B	Ribosomal RNA-processing protein 9	30.9	2.4	290	366
5	2Z2O-C	Virginiamycin B lyase	30.5	2.4	286	299

This table was populated by Terry B. Ruskoski with a structural comparison to other proteins in the PDB.[206]

Table 3.7. Data collection and refinement statistics for Br-soaked CylK structures

	Native	Br anomalous
Data collection		
Space group	*C* 2 2 2_1	*C* 2 2 2_1
Wavelength (Å)	0.97872	0.9184
Cell dimensions		
a, *b*, *c* (Å)	93.470, 138.656, 99.511	93.557, 138.726, 99.706
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0-1.52 (1.55-1.52)	50-1.54 (1.57-1.54)
R_{merge}	0.076 (0.703)	0.725 (4.755)
R_{pim}	0.029 (0.295)	0.149 (1.140)
$I / \sigma I$	23.3 (2.14)	24.0 (2.06)
$CC_{1/2}$	0.990 (0.843)	0.978 (0.124)
Completeness (%)	99.7 (98.9)	100.0 (100.0)
Redundancy	7.0 (5.5)	24.2 (16.9)
Refinement		
Resolution (Å)	46.78 - 1.52	
No. reflections	89,122	
R_{work} / R_{free}	0.1693/0.1978	
No. atoms	5,437	
Protein	4,989	
Ligand/ion	45	
Water	403	
B-factors (Å^2)		
Protein	16.68	
Ligand/ion	23.03	
Water	23.77	
R.m.s. deviations		
Bond lengths (Å)	0.013	
Bond angles (°)	1.823	
Molprobity clashscore	1.64 (99th percentile)	
Rotamer outliers (%)	0.76	
Ramachandran favored (%)	96.41	
Ramachandran allowed (%)	3.59	
Ramachandran outlier (%)	0.00	
PDB accession code	7ROO	

*Values in parentheses are for highest-resolution shell.

3.4.7 Mutant enzyme activity screening with non-biological substrates

CylK mutants were expressed and assayed for activity according to a modified method[128] (see Materials and Methods 2.4.9). Point mutants in plasmid pPR-IBA1-CylK were synthesized and sequence verified by Genewiz (South Plainfield, NJ, USA) (Table 3.8). Enzyme activity assays using substrate pair **7** + **16e** were carried out as follows. Chloro-SNAc electrophile **16e** was accessed as described previously (see Materials and Methods 2.4.11).[128] Respective mutant, wild-type, and negative control plasmids (Addgene #31122), were freshly transformed into electrocompetent *E. coli* BL21 Gold CodonPlus® (DE3) RIL cells (Agilent), and an individual colony was used to directly inoculate 1 mL cultures of LB medium supplemented with 100 µg/mL ampicillin or carbenicillin and 34 µg/mL chloramphenicol in a 96 deep well plate (2 mL capacity, VWR), sterility was maintained with standard techniques and a gas permeable rayon film (VWR). The liquid culture plate was incubated for 5 h at 37 °C and 250 rpm shaking. After five h, and confirming visible growth in each well, the liquid culture plate was cooled to 15 °C, maintaining shaking. Following an additional 30-45 min of incubation, protein expression was induced with 250 µM IPTG. The cultures were incubated for 4 h at 15 °C with 250 rpm shaking. Following expression, 10 µL of each culture was removed from each well and sub-cultured for liquid culture RCA-based DNA sequencing to confirm identity and rule out culture cross contamination (Genewiz). The remaining cultures were concentrated ~10X by centrifugation (3,220 x *g* for 10 min) and resuspended in ~200 µL of spent media supernatant. Reactions were initiated in opaque 96-well plates (Costar) by combining 100 µL of concentrated cell suspension with 2 µL of lysozyme mix (9 mg lysozyme & 1 mg EDTA-free Pierce Protease Inhibitor Tablet per 500 µL of 500 mM Tris, 50 mM EDTA, pH 8.0) and mono-alkylation substrates in DMSO for a final concentration of nucleophile **7** at 150 µM, electrophile **16e** at 300 µM, and DMSO at 3.4%. Reaction plates were sealed with aluminum seals (VWR) and incubated at 37 °C with 190 rpm shaking for 20 h. Reactions were quenched with the addition of 1:1 methanol/acetonitrile (200 µL) and centrifuged (3,220 x *g* for 15 min). The supernatant was then subjected to LC-MS to measure ion abundance of product **28** ($C_{31}H_{53}ClNO_4S$ [M + H^+] = 536.3768 ± 5 ppm) and starting material **16e** ($C_{14}H_{26}ClNO_2S$ [M + H^+] = 308.1446 ± 5 ppm) as a loading control. LC-MS conditions as described previously (see Materials and Methods 2.4.1).[128] Reactions were repeated in biological triplicate.

3.4.8 Mutant enzyme activity screening with biological substrates

Enzyme activity on native substrates **1** or **2** was assayed as follows. Substrate **1** and intermediate **2** were accessed chemoenzymatically as described previously (see Materials and Methods 2.4.10).[69] Enzymatic reactions affording intermediate **2** were lyophilized, dissolved in DMSO, and used neat. Care was taken to minimize glycerol because it was determined to inhibit CylK activity. Respective mutant and wild-type plasmids (Table 3.8) were freshly transformed into electrocompetent *E. coli* BL21 Gold CodonPlus® (DE3) RIL cells (Agilent), and 5-10 colonies were used to directly inoculate 100 mL cultures in LB medium supplemented with 100 µg/mL ampicillin or carbenicillin and 34 µg/mL chloramphenicol. The liquid cultures were incubated at 37 °C and 190 rpm shaking. At OD400 = 0.4, the cultures were cooled to 15 °C, maintaining shaking. Following an additional 1 h of growth, protein expression was induced with 250 µM IPTG. The cultures were then incubated for 4 h at 15 °C with 190 rpm shaking. Following expression, cell pellets were collected by centrifugation (3,220 x *g* for 10 min) and resuspended in ~2 mL of assay buffer (20 mM HEPES pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$). Cell concentration was normalized across mutants by measuring OD600 and diluting with an appropriate amount of assay buffer. A 600 µL aliquot of each mutant was lysed by sonication (Branson, 25% amplitude, 1 min) on ice. The soluble fraction was collected by centrifugation (16,100 x *g* for 20 min) and analyzed by anti-Strep-HRP (IBA) western blotting to confirm mutant solubility and approximate protein concentration (Figure 3.32). Reactions were initiated by combining 4.3 µL of **1** or **2** with 25.7 µL of soluble lysate for a final concentration of **1** at 400 µM and 4% DMSO, and **2** at 200 µM and 2% DMSO. Reactions were sealed and incubated at 37 °C for 1 or 20 h. Reactions were quenched with the addition of 1:1 methanol/acetonitrile (60 µL) and centrifuged (6,000 x *g* for 15 min). The supernatant was then subjected to analytical HPLC to determine percent conversion as described previously (see Materials and Methods 2.4.1). Reactions were performed in biological duplicate and repeated again on a second day to minimize any variation. Results were consistent on both days.

Table 3.8. Point mutants of CylK-Strep

mutant,	codon change,	position
K78E:	Replaced AAG with GAG	at position 232-234
D96A:	Replaced GAT with GCT	at position 286-288
R105A:	Replaced CGT with GCC	at position 313-315
Y473A:	Replaced TAT with GCT	at position 1417-1419
S265A:	Replaced AGT with GCC	at position 793-795
S266A:	Replaced TCC with GCC	at position 796-798
T267A:	Replaced ACC with GCT	at position 799-801
S324A:	Replaced AGT with GCT	at position 970-972
K377A:	Replaced AAA with GCC	at position 1129-1131
T379A:	Replaced ACC with GCT	at position 1135-1137
D502A:	Replaced GAC with GCC	at position 1504-1506
E504A:	Replaced GAA with GCC	at position 1510-1512
L558A:	Replaced TTA with GCT	at position 1672-1674
D559A:	Replaced GAT with GCT	at position 1675-1677
D562A:	Replaced GAT with GCC	at position 1684-1686
T616A:	Replaced ACG with GCT	at position 1846-1848
T618A:	Replaced ACT with GCT	at position 1852-1854
D440A:	Replaced GAC with GCC	at position 1318-1320
W442F:	Replaced TGG with TTT	at position 1324-1326
L438A:	Replaced CTT with GCT	at position 1312-1314
K459A:	Replaced AAA with GCT	at position 1375-1377
D135A:	Replaced GAT with GCT	at position 403-405
H391A:	Replaced CAC with GCC	at position 1171-1173
Y376A:	Replaced TAT with GCC	at position 1126-1128
E374A:	Replaced GAA with GCT	at position 1120-1122
T84V:	Replaced ACA with GTG	at position 250-252
F499A:	Replaced TTT with GCC	at position 1495-1497
N334A:	Replaced AAC with GCC	at position 1000-1002
S318A:	Replaced AGC with GCC	at position 952-954
R105K:	Replaced CGT with AAG	at position 313-315
Y473F:	Replaced TAT with TTT	at position 1417-1419
D440N:	Replaced GAC with AAC	at position 1318-1320
R105A/Y473F:	Replaced CGT with GCC	at position 313-315
	and TAT with TTT	at position 1417-1419
R105K/Y473A:	Replaced CGT with AAG	at position 313-315
	and TAT with GCT	at position 1417-1419
R105A/Y473A:	Replaced CGT with GCC	at position 313-315
	and TAT with GCT	at position 1417-1419

Point mutations constructed and sequence verified by Genewiz (South Plainfield, NJ, USA).

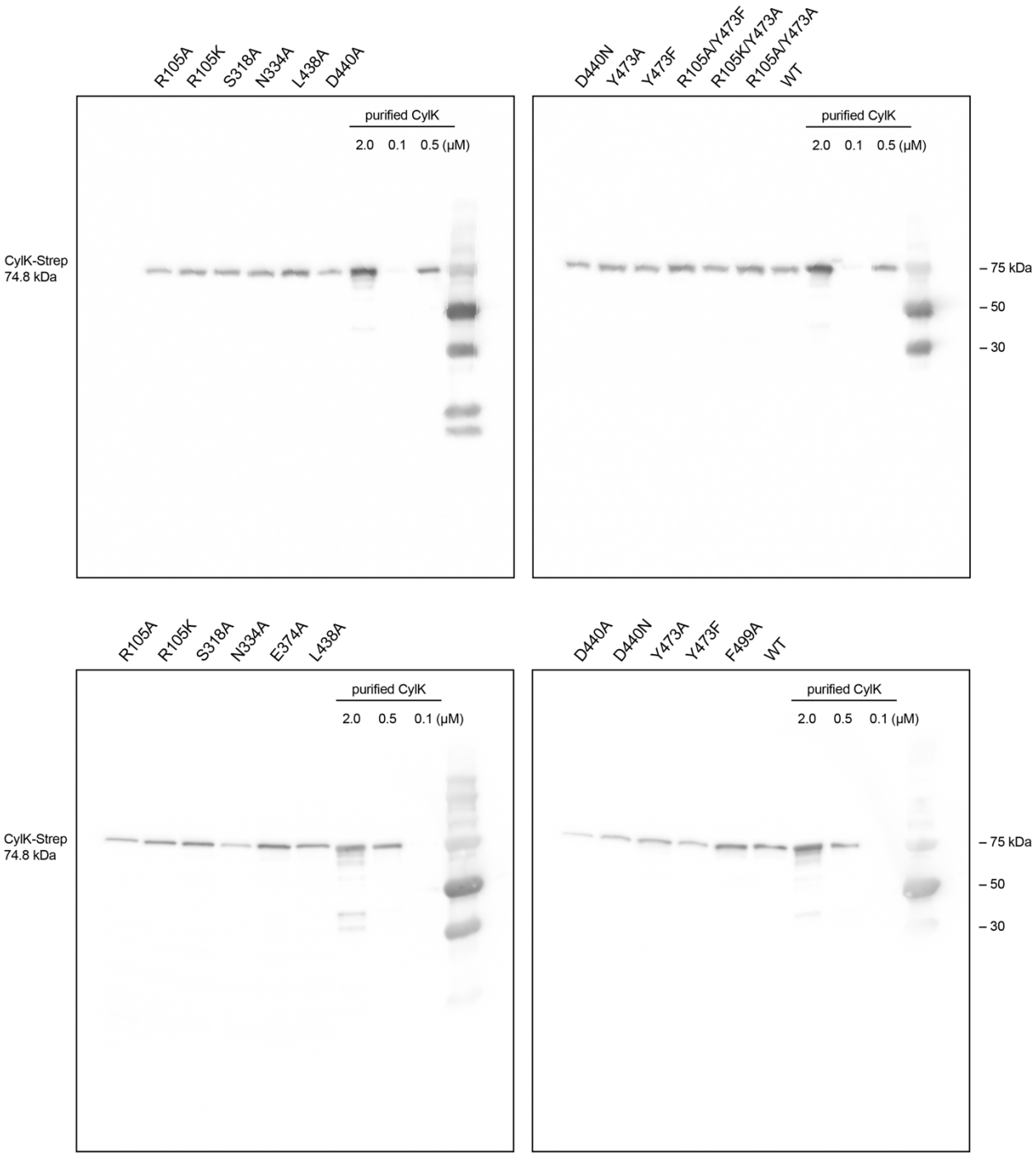


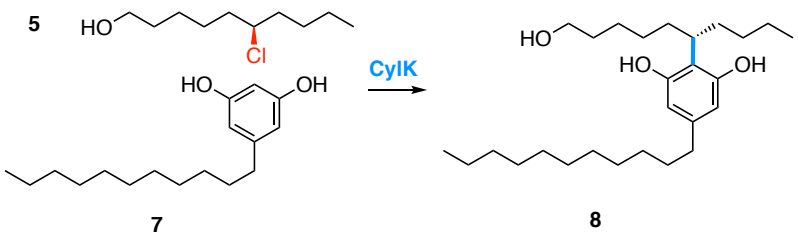
Figure 3.32. Western blotting of CylK mutants

Western blots of CylK-Strep mutants from soluble lysate with anti-Strep-HRP (horseradish peroxidase, IBA) in order to ensure mutant enzyme expression and solubility. Serial dilution of purified CylK-Strep included to visualize protein concentration range. All Blue Precision Plus Protein Prestained Standard (Bio-Rad) was used to visualize molecular weight standards.

3.4.9 Molecular dynamics simulations

This work was conducted by C. Denise Okafor with contributions from Nathaniel R. Braffman, Terry B. Ruskoski, and Amie K. Boal. Coordinates for the native substrates **1** and **2** were prepared in ACEDRG[207] from MOL file descriptions of the molecules generated in ChemDraw. Substrate molecules were docked manually into cavity 1 of CylK using Br1 to place the alkyl chloride electrophile substituent and interactions with Asp440 and E374 to guide placement of the resorcinol nucleophile substituent. The resulting CylK complexes were prepared for molecular dynamics (MD) simulations using the Xleap module of AmberTools.[181] Substrates were parameterized using Antechamber[208,209] in AmberTools, while the FF14SB forcefield[210] was used for protein atoms. Complexes were solvated in a 10 Å octahedral box of in TIP3P water.[211] Sodium and chloride ions were added to neutralize the complexes and achieve a final concentration of 150 mM NaCl. All minimizations and production runs were performed with Amber20.[181] Minimization was performed using 5000 steps of steepest descent and 5000 steps of conjugate gradient. In the first round, 500 kcal/mol·Å^2 restraints on all protein, ligand atoms as well as the metal ion cofactors. Restraints were reduced to 100 kcal/mol·Å^2 for an additional round of solvent minimization. Next, restraints were retained only ligands and metal ions for a third round of minimization. Finally, all restraints were removed for the last stage of minimization. After minimization, a 100 ps run was used to heat complexes from 0 to 300 K using constant volume periodic boundaries with 10 kcal/mol·Å^2 restraints on all solute atoms. Equilibration was performed using two 10 ns runs (10 kcal/mol·Å^2 and 1 kcal/mol·Å^2 restraints) on protein, ligand and metal ions. Following equilibration, simulations were run as either restrained or unrestrained. For restrained simulations, we obtained 100 ns trajectories with 1 kcal/mol·Å^2 restraints on protein atoms. Unrestrained simulations (100 ns) were run with all restraints removed.

3.4.10 Kinetic experiments with halide ions as potential inhibitors



7 8

Relative rate of CylK catalysis on substrate pair **5** + **7** was assayed as follows. Enzyme master mix (2x) with 4 µM (chloride-free) CylK-Strep, 40 mM HEPES pH 7.8, 2.4 mM $NaNO_3$, 20 µM $CaSO_4$, and 0.2% v/v glycerol, and substrate master mix (2x) with 400 µM electrophile **5**, 400 µM nucleophile **7**, 6% v/v DMSO, and 0–5000 µM NaX were prepared. Master mixes were preincubated to 37 °C, and then reactions were initiated by combining 15 µL of enzyme master mix with 15 µL substrate master mix to a final volume of 30 µL and final concentrations of 2 µM (chloride-free) CylK-Strep, 200 µM electrophile **5**, 200 µM nucleophile **7**, 3% v/v DMSO, 20 mM HEPES pH 7.8, 1.2 mM $NaNO_3$, 10 µM $CaSO_4$, and 0.1% v/v glycerol. Reactions were quenched after 10, 20, or 30 min with the addition of 1:1 methanol:acetonitrile (60 µL) and centrifuged (16,100 x *g* for 10 min at 4 °C) twice. The supernatant was then subjected to analytical HPLC to measure percent conversion.

3.4.11 Bioinformatic analysis

The CylK (GenBank: ARU81125.1) and BrtB (GenBank: AOH72618.1) protein sequences were independently used in a Basic Local Alignment Search Tool (BLAST) search of the Joint Genome Institute-Integrated Microbial Genomes & Microbiomes (IMG-JGI) database of all isolates, metagenome-assembled genomes (MAGs), and single-amplified genomes (SAGs). Each BLAST search resulted in 500 hits (>24% amino acid sequence identity) that were merged to form a non-redundant list of 715 enzyme candidates (available online as Supplementary file 1).[162] Parent enzyme sequences (CylK, BrtB) and likely CylK homologs (MerH, CabK, CylK from *C. stagnale*) were added to this dataset. A targeted BLAST search of the genomes from this list was performed to identify all putative halogenases (CylC-like, Fe^{II}/2OG, SAM-dependent, flavin-dependent, & haloperoxidase) and resorcinol forming enzymes (CylI, BrtC, BrtD).[162] All

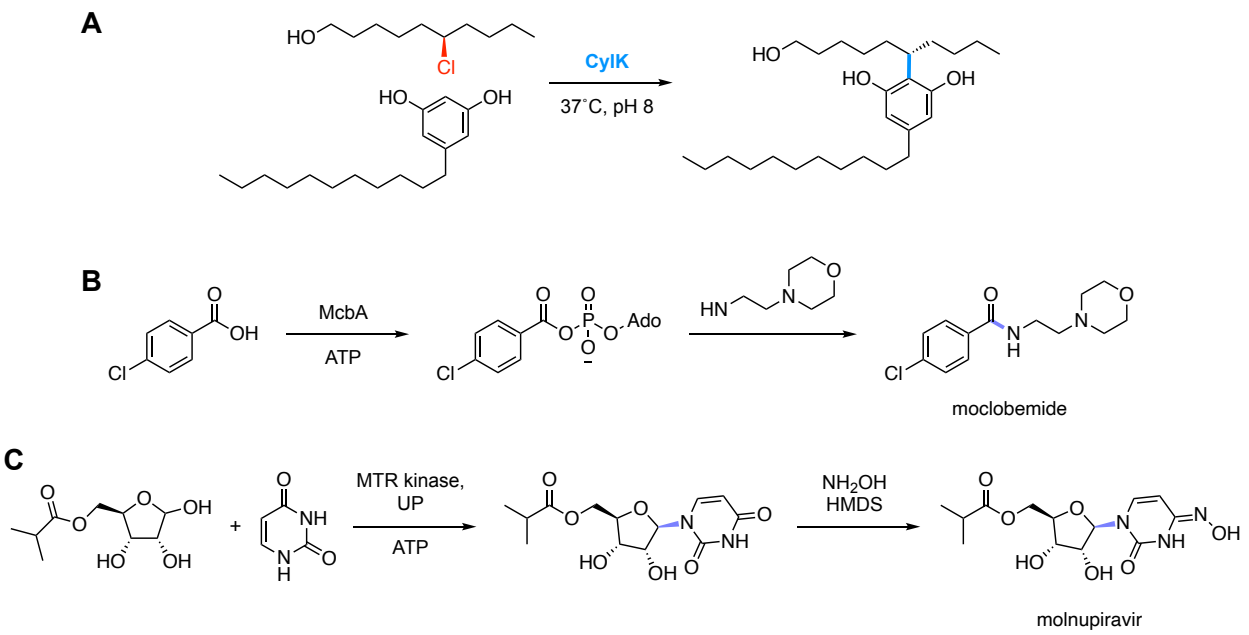
non-CylC halogenases were confirmed to contain key catalytic residues and examined for co-localization with CylK. All candidate CylC halogenases were included, although the overwhelming majority (63 of 78 enzymes) were fewer than 18 genes away from CylK. CylC's catalytic residues remain unknown. Monoalkylresorcinol forming CylI homologs were included if clustered with CylK; 11 of 54 candidate enzymes were less than 12 genes away from CylK. Next, candidate CylK enzymes were filtered by constructing an individual MUSCLE alignment[212,213] of each hit to the CylK parent sequence and identifying at least 135 amino acid residues before the C-terminal domain. Lys240 or its aligned equivalent was defined as the junction point between protein domains. This analysis resulted in 286 unique protein sequences (available online as Supplementary file 2)[162] that were aligned (MUSCLE, EMBL-EBI), sites containing gaps at ≥ 90% of sequences were removed, and a maximum-likelihood phylogenetic tree was constructed using IQ-TREE v1.6.12 [214] and visualized with FigTree v1.4.4. The best fit model (VT+F+G4) was selected by ModelFinder.[215] Bootstrap values were calculated using the UFBoot2 method.[216] These candidate enzymes were analyzed for the presence or absence of Arg105, Tyr473, and Thr84 residues by constructing an individual MUSCLE alignment of each hit to the CylK parent sequence. The results were manually mapped onto the phylogenetic tree. The subset CylK homologs with Arg105 and Tyr473 residues were MUSCLE aligned and visualized with ConSurf.[191] Code used to perform iterative MUSCLE alignments available (https://github.com/nbraffman/CylK-homologs).

4 Enzyme engineering and method development

Portions of the work reported were conducted in collaboration with other lab members including Dr. Douglas J. Kenny, Michelle Wang, and Vanessa Dippon. Their contributions will be recognized as footnotes throughout the chapter.

4.1 Introduction

Considering the challenges associated with contemporary synthetic methods for constructing asymmetric alkyl–aryl linkages (see Chapter 1.5 for examples), CylK and the broader family of alkyl halide-utilizing enzymes hold promise for synthetic applications. Furthermore, as demonstrated in Chapter 2, we can use wild-type CylK to access novel derivatives of the therapeutic compound benvitimod (see Table 2.9 for reaction conditions), and although conversions were low, this presents a potentially useful starting point for enzyme engineering efforts to improve CylK's properties including expanded scope, increased substrate loading, enhanced solubility, high expression levels, and improved catalytic rate. Engineered variants of CylK and its homologs may find use as biocatalytic coupling reagents in order to access novel scaffolds with alkyl–aryl, ester, or ether linkages. More broadly, the practical use of enzymes that enable convergent synthetic strategies, including CylK, represents a new paradigm in the field of biocatalysis. This view has been persuasively articulated by Narayan and coworkers in a recent perspective article detailing notable enzymes that catalyze fragment coupling reactions.[217] Traditionally, biocatalysts have been incorporated into syntheses to introduce chiral centers with control of stereochemistry or to enable late-stage functional group interconversions, while their use as coupling reagents to assemble core skeletons has been limited.[218,219] Notable exceptions include the syntheses of moclobemide[220] and molnupiravir[221] which both use enzymes to form transiently phosphorylated intermediates that are used in amide bond formation and glycosylation reactions (Scheme 4.1). Building from our initial protein engineering efforts described in this Chapter, we envision that CylK may find similar use in convergent biocatalytic syntheses.



Scheme 4.1. Convergent biocatalytic strategies

(*A*) Representative non-biological coupling reaction catalyzed by CylK. (*B*) Biocatalytic synthesis of monoamine oxidase A inhibitor moclobemide relies on ATP-dependent amide bond synthetase McbA to form activated adenylated intermediate followed by nucleophilic amine displacement to access amide coupled product. 'Ado' = adenosine (*C*) Biocatalytic synthesis of COVID-19 antiviral molnupiravir involves several enzymatic steps but notably, phosphorylation of a ribose by MTR kinase, followed by nucleophilic amide displacement catalyzed by UP to afford the uracil coupled product. 'MTR' = methylthioribose, 'UP' = uridine phosphorylase.

While the number of biocatalytic methods are dwarfed by their non-enzymatic counterparts, biocatalysis offers several advantages over traditional chemical methods, including high regio- and stereoselectivity, broad functional group tolerance (oftentimes obviating the need for protecting groups), generally mild reaction conditions, minimal use of toxic/dangerous reagents and solvents, and decreased costs overall.[222] These benefits often translate into more environmentally friendly processes. However, wild-type enzymes are rarely used as biocatalysts in synthetic campaigns without some degree of primary sequence modification and optimization, which can be challenging or slow to realize. The growing use of biocatalysis in both academic and industrial settings is largely due to the advent of protein engineering techniques, notably directed evolution.[129] Early uses of biocatalysis took advantage of wild-type enzymes to mediate transformations in a non-biological context, examples include supplementing laundry detergents with

proteases to enhance cleaning power at lower temperatures and harnessing penicillin G acylase to access semisynthetic β-lactam antibiotics including ampicillin and amoxicillin on industrial scales.[223,224] The nascent field of protein engineering then began to grow rapidly when structural information and contemporary molecular biology techniques could be used to semi-rationally generate enzyme mutants with properties such as enhanced stability or affinity for non-native substrates.[225] This approach is now broadly known as directed evolution, and it involves iterative cycles of genetic diversification and screening for the desired activity or property. Structural models, bioinformatic information, or *in silico* methods can tailor genetic diversification towards particular residues or important protein regions in order to minimize the sequence space necessary to screen.[226] However, a frequent bottleneck in the directed evolution workflow is the screening method itself. Optical readouts can be rapidly assayed in a plate-based format, however, these systems are largely substrate specific and are not easily generalizable to new reactions.[227] HPLC and LC-MS based methods tend to be low throughput, however, some mass spectrometry (MS)-based methods have been used successfully, particularly in the context of pharmaceutical synthesis, and some contemporary methods can be as fast as seconds per sample.[227,228] With these challenges in mind, and with the knowledge of CylK's structure and wild-type substrate scope, we set out to develop new high-throughput screening methodology to guide enzyme evolution and to begin preliminary protein engineering efforts in order to enhance CylK's overall activity, expression levels, and solubility. This work may serve as a starting point for additional rounds of directed evolution towards CylK variants with expanded substrate scopes and favorable biocatalytic properties including increased substrate loadings and catalytic rate.

4.2 Results and Discussion

4.2.1 Bioactivity-guided directed evolution method development

In industrial pharmaceutical settings, biocatalysis is typically incorporated into synthetic campaigns if it can bring cost savings or process improvements to the large-scale syntheses of lead compounds. Initiating protein engineering efforts, which can be time consuming and expensive, generally takes place after the target compound has been designed, chemically synthesized (on a small scale), and may even have already demonstrated favorable biological activity in pre-clinical investigations. Considering the short timelines of the pharmaceutical industry, it can be a challenge to then rapidly engineer and incorporate novel biocatalysts into the manufacturing process before the overall synthetic route must be finalized for regulatory filings.[129] A number of advances in technologies that support protein engineering including DNA synthesis,[229] microfluidics-based liquid handling,[230] and *in silico* protein structure prediction[231] may speed up the lag time from reaction conception to having a mature biocatalyst in hand. However, we and others suggest that it may also be beneficial to incorporate enzymatic chemistry earlier in the drug development timeline.[232] Part of this challenge will be making enzymes more user-friendly so they can be widely available as synthetic reagents during the iterative medicinal chemistry process. For example, commercially supplying robust biocatalysts as lyophilized powder or easy-to-handle liquid solutions may accelerate their adoption by users unfamiliar with molecular biology and heterologous protein expression.[xvi] Furthermore, for late-stage enzymatic transformations that might require significant development and engineering efforts, we also envision merging the processes of medicinal chemistry with protein engineering to potentially identify lead compounds and key biocatalytic steps in tandem. Our work with CylK, and in particular the discovery that we can use this enzymatic transformation to access novel derivatives of the clinical candidate benvitimod inspired the idea of this merged screening approach, which we refer to as bioactivity-guided directed evolution.

[xvi] Notably, companies including Codexis have made biocatalyst kits available to consumers for limited reaction types.

Considering the challenges associated with directed evolution screening techniques (discussed in the introduction to this Chapter), we set out to identify a new, more rapid method to screen enzyme variants for the ability to produce bioactive products with therapeutic potential. It also occurred to us that robust and high-throughput assays already exist for detecting numerous biological activities, and that they could potentially be applied to measure the enzymatic conversion of biologically 'inactive' starting materials to novel products that might elicit a desired phenotypic response in the context of a screen. In this way, both enzymatic and biological activity can potentially be evaluated simultaneously. Notably, there is significant precedent for this type of bioactivity-guided screening in other systems. A semi-analogous case of activity screening for small molecule production comes from antibiotic discovery efforts, in which extracts from putative metabolite producing organisms are fractionated and then assayed for the phenotype of growth impairment or killing of target bacterial strains.[233] The active antibiotics from these screens can then be further investigated as potential drug candidates, while the producing organism or biosynthetic pathway affords access to these compounds and may even be amenable to scale up with downstream protein and/or metabolic engineering. We reasoned that one could analogously assay enzyme mutant libraries for the production of compounds with desired biological activities (beyond antibiotic-type inhibition) by testing them in relatively high-throughput against mammalian cell lines with reporter assays for the desired bioactivities. Two general requirements for implementing this workflow are: (1) starting materials and other reaction components should not elicit the reporter response, and (2) mixtures of 'products' and 'starting materials' (representing incomplete conversion to product) should elicit a significant reporter response, as enzyme variants screened in this format hits may afford trace conversions as a starting point for subsequent rounds of directed evolution. Note, precedence for an analogous bioactivity-based workflow can be found in recent work in using antibiotic activity as a readout to screen a 6,000 member library of unnatural RiPPs accessed via combinatorial shuffling of peptide modules.[234] This workflow was miniaturized using small alginate hydrogel compartments to encapsulate producing and reporter bacterial cells, and similar methodologies exist for manipulating small populations of mammalian cells in parallel, and assaying each population with an optical readout.[235]

In the context of our work, we had used CylK to access small quantities of novel derivatives of benvitimod (Figure 4.1, and see Table 2.9 for reaction conditions), which has been shown to resolve skin inflammation associated with psoriasis by activating the aryl hydrocarbon receptor (AhR).[149] We wondered if these analogs of benvitimod had biological activity, and if mutant libraries of CylK could be subjected to bioactivity-guided directed evolution in order to improve yields and/or access new analogs with AhR bioactivity. To utilize this screening approach, we first needed to determine if our starting materials were inactive, or 'silent', in our reporter assay, as this would establish the background level of bioactivity and if our assay signal had sufficient dynamic range.[236] The precise structural determinates of AhR agonism (activation) versus antagonism (repression) have not been elucidated. Interestingly, the plant-derived natural product, and competent CylK nucleophile, resveratrol (**20b**) is an AhR antagonist,[237] although it is closely structurally related to benvitimod, an AhR agonist (Figure 4.1). The key differences between the two molecules are the presence of a third phenolic group and the absence of the isopropyl group. Furthermore, our stilbene starting material, 3,5-dihydroxylstilbene (**20a**), lacks both of these moieties and its AhR activity was unknown at the time of this work.

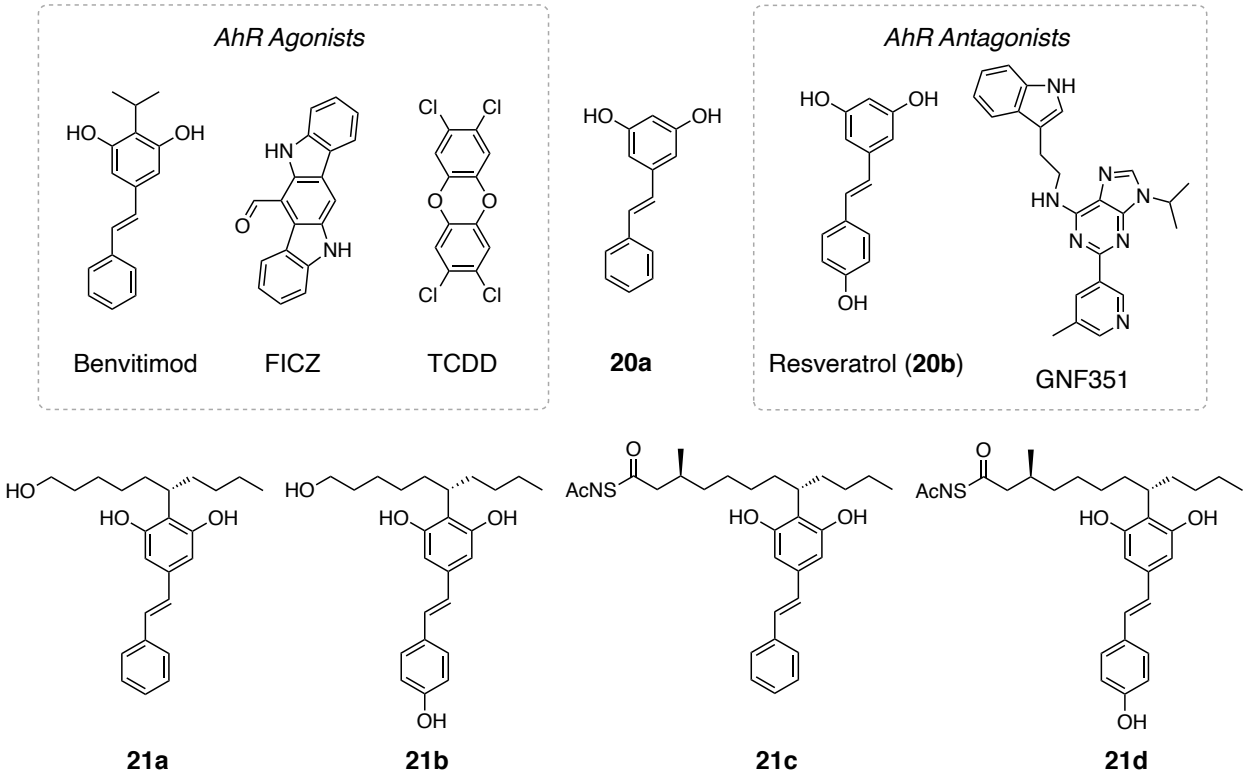


Figure 4.1. AhR ligands and uncharacterized analogs

Three known aryl hydrocarbon receptor (AhR) agonists and two AhR agonists are highlighted in grey boxes. FICZ (6-formylindolo(3,2-b)carbazole) is believed to be an endogenous human AhR ligand produced from tryptophan.[146] TCDD (2,3,7,8-Tetrachlorodibenzo-*p*-dioxin) is an anthropogenic product produced as an industrial side product that demonstrates AhR antagonism.[147] GNF351 (*N*-(2-(1*H*-indol-3-yl)ethyl)-9-isopropyl-2-(5-methylpyridin-3-yl)-9*H*-purin-6-amine) is a synthetic AhR ligand with complete antagonism.[238] 3,5-dihydroxystilbene (**20a**) and resveratrol (**20b**) are competent CylK nucleophiles that afford access to alkylated benvitimod-like products (**21a-d**). Resveratrol is an AhR antagonist while the bioactivities of **20a** and **21a-d** were unknown at the time of this work.

In order to enable bioactivity-guided direction evolution on this system, we also needed a robust and high-throughput method to detect AhR activity. Excitingly, there exists a human HT29 colon adenocarcinoma cell line engineered with a luciferase-based reporter of AhR activity (Invivogen) that is amenable to 96-well plate-based assaying and detection. However, we first utilized qRT-PCR to determine if our starting material ligand (**20a**) would likely be 'silent' in the wild-type cell-based assay before acquiring the engineered reporter cell line. To measure AhR activity, we synthesized primers for the gene encoding cytochrome P450-dependent monooxygenase 1A1 (CYP1A1), as its expression is induced by AhR activation.[239,240] Interestingly, this enzyme is implicated in the oxidation of AhR ligands themselves, which may serve as a feedback loop by modifying their potency.[241] We then tested the activity of these ligands by exposing human Caco-2 (colorectal adenocarcinoma) cells to 10 nM or μM of each ligand for 4 h.[xvii] Following RNA extraction, qRT-PCR analysis revealed that benvitimod upregulated CYP1A1 at both concentrations, while 3,5-dihydroxystilbene (**20a**) and resveratrol (**20b**) do not, and they may even slightly downregulate CYP1A1 expression (Figure 4.2). This suggests that 3,5-dihydroxystilbene (**20a**) does not activate AhR in this *in vitro* context, and that the isopropyl group of benvitimod appears to be the key determinant of its bioactivity. Excitingly, this suggests that our stilbene starting materials (**20a-b**) can potentially be used as substrates for bioactivity-guided direction evolution as they do not contribute significant background activity. Of note, CYP1A1 upregulation by benvitimod was more consistent at 10 μM (Figure 4.2B).

[xvii] This work was done in collaboration with Dr. Douglas Kenny.

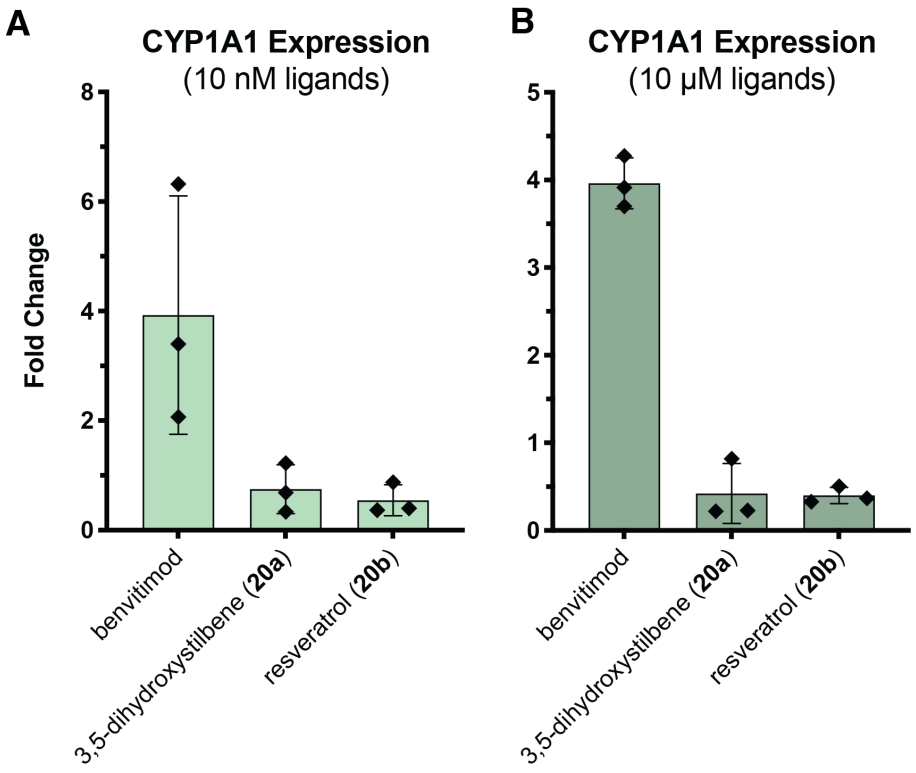


Figure 4.2. AhR activity of stilbene nucleophiles

CYP1A1 expression levels are measured as a surrogate for AhR activity in cultured Caco-2 cells exposed to stilbene ligands for 4 h at (*A*) 10 nM or (*B*) 10 μM concentrations. CYP1A1 expression was quantified by qRT-PCR, error bars represent the standard deviation from the mean of three biological replicates, and biological replicate values (diamonds) are reported as the average of three technical replicates.

Next, we set out to determine if mixtures of 'products' and 'starting materials' could be detected in this assay format. We reasoned that incomplete conversion to product would be highly likely in biocatalyst activity screens, especially if the starting enzyme construct cannot initially perform the desired transformation. Accounting for this might be additionally important in our system considering that the stilbene nucleophile substrates may moderately suppress AhR activity. Furthermore, we also wondered if relative conversion could be measured with a bioactivity readout, as we potentially wanted to apply bioactivity-guided direction evolution towards improving existing enzyme activity, in addition to identifying enzyme variants with new reactivity. To interrogate this, we exposed human HT29 colon adenocarcinoma cells to mixtures of 3,5-dihydroxystilbene (**20a**) 'starting material' and benvitimod 'product' ligands for 1 h, representing 1% and 10% conversion.[xviii] We also included lysate reaction components at analogous dilutions including LB, *E. coli*, and lysozyme. In order to enhance the bioactivity signal, we shortened the 4 h stimulation used previously, hypothesizing that RNA levels may decrease over time following initial ligand exposure. We also utilized HT29 cells instead of the Caco-2 cell line because, although similar, the engineered luciferase reporter cell line we planned to use was derived from HT29. Excitingly, following stimulation and RNA extraction, qRT-PCR analysis revealed that mixtures of 3,5-dihydroxystilbene (**20a**) and benvitimod elicited a strong CYP1A1 expression response (>10 fold change), even at a ratio equivalent to 1% conversion (9.9 nM stilbene **20a**, and 0.1 nM benvitimod) at 10 nM total ligand concentration (Figure 4.3A). The signal response at 1% conversion was further enhanced at 10 μM total ligand concentration (~50 fold change) (Figure 4.3B). Furthermore, LB, *E. coli*, and lysozyme (individually and together) did not elicit a significant CYP1A1 response; and mixtures of stilbene **20a** or benvitimod with *E. coli* lysate components had comparable activity to those ligands without lysate. Of note, a dose response that correlated to ratios of 'product' to 'starting material' (simulating percent conversion) could be observed at the lower stimulatory total ligand concentration of 10 nM (Figure 4.3A).

[xviii] This work was done in collaboration with Dr. Douglas Kenny.

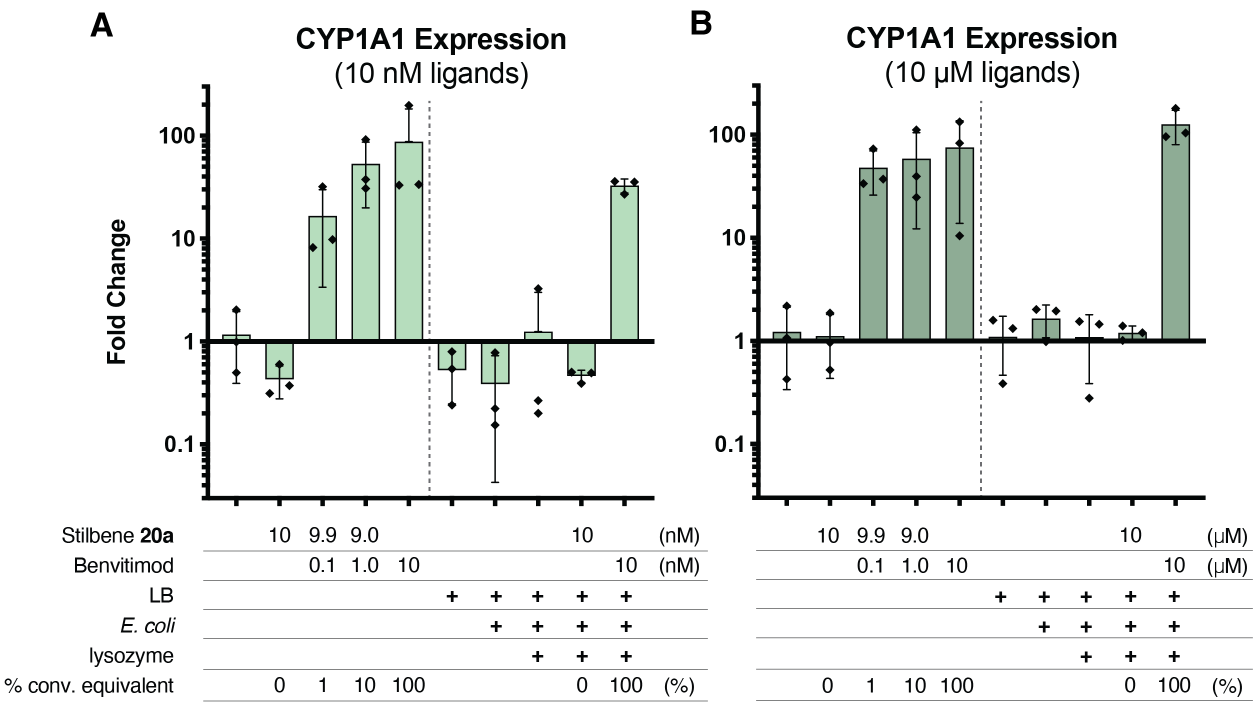


Figure 4.3. AhR activity of pseudo reaction mixtures of 3,5-dihydroxystilbene and benvitimod

CYP1A1 expression levels are measured as a surrogate for AhR activity in cultured HT29 cells exposed to mixtures of stilbene ligands and cell lysate for 1 h at (*A*) 10 nM or (*B*) 10 μM concentrations (or dilution equivalents). CYP1A1 expression was quantified by qRT-PCR, error bars represent the standard deviation from the mean of three biological replicates. Percent conversion of pseudo reaction mixtures are indicated as (% conv. equivalent).

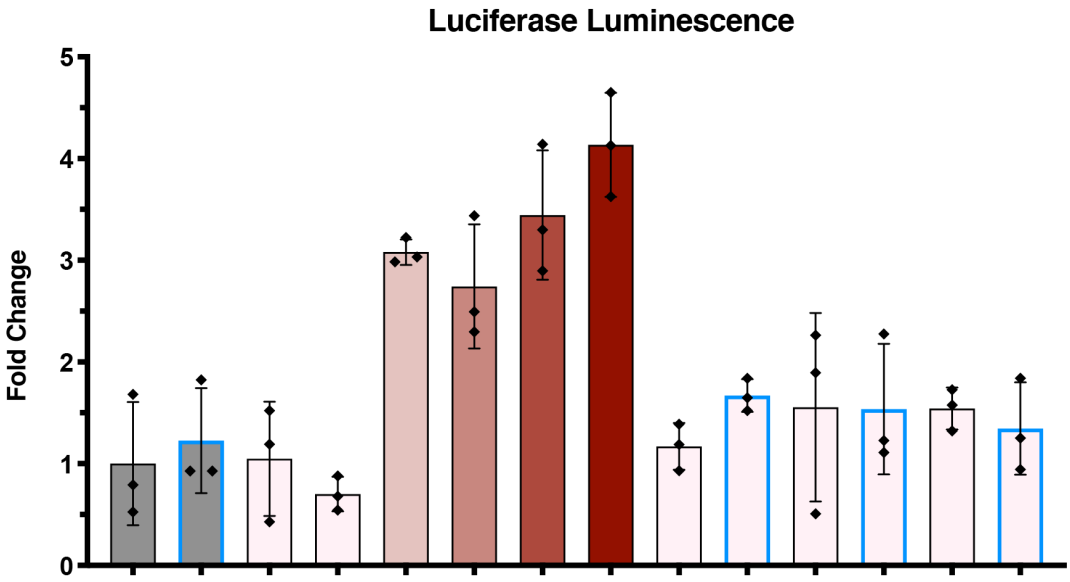
Collectively, these results suggest that an AhR-related response might enable bioactivity-guided directed evolution of CylK, provided that the luciferase reporter-based assay mirrors the CYP1A1 expression response. Interestingly, it appears that both total ligand concentrations (10 μM or 10 nM) elicited AhR activity, however, we speculate that diluting lysate reactions to 10 μM (200-fold from 200 μM standard substrate loading conditions) would be more appropriate when screening our CylK enzyme libraries for new reactivity, while diluting to 10 nM (200,000-fold from 200 μM) is best for subsequent rounds of enzyme evolution conducted to enhance reactivity already catalyzed by CylK or a variant. Note, these dilution schemes will vary by system, and target concentrations can be empirically determined during assay development.

4.2.2 Luciferase reporter validation and AhR activity of novel benvitimod derivatives

Having established the theoretical feasibility of bioactivity-guided directed evolution of CylK and stilbene ligands with an AhR-related response, we then acquired the human HT29 colon adenocarcinoma cell line engineered with a luciferase-based reporter of AhR activation (Invivogen) in order to benchmark CYP1A1 expression levels to a luciferase response more amenable to high-throughput screening. We chose to use a 1.0 µM total ligand concentration as an intermediate dosing level with the knowledge that 10 nM and 10 µM both elicit a strong response, and that 10 nM demonstrates dose response for ligand mixtures of 3,5-dihydroxystilbene (**20a**) and benvitimod. We also included extracts of *in vitro* CylK reactions producing alkylated stilbenes (**21a-d**) in order to assess their bioactivity in parallel. First, we verified that the starting materials, 3,5-dihydroxystilbene (**20a**) and resveratrol (**20b**), and the enzyme alone did not demonstrate luminescence (as a surrogate AhR activity), while 1.0 µM of benvitimod elicited a response (Figure 4.4). As before, we then assayed a gradient of 3,5-dihydroxystilbene (**20a**) and benvitimod mixtures in order to simulate reaction mixtures, although this time we included the lower equivalent of 0.1% conversion as well. Excitingly, this demonstrated that even 0.1% benvitimod (1 nM) was sufficient to elicit an AhR response in the presence of 99.9% 3,5-dihydroxystilbene (**20a**) (999 nM). This suggests that we can use this reporter cell line to assay CylK reaction mixtures for the production of trace amounts of bioactive products.

Now having validated this methodology to assess the relative bioactivity of mixtures of stilbenes, we also exposed the reporter cell line to crude CylK reaction mixtures that produced adducts (**21a-c**), along with their respective negative controls lacking enzyme. We verified by HPLC and LC-MS that adducts **21a-c** were present in reactions with CylK. Interestingly, no significant bioactivity was observed, suggesting that our alkylated stilbenes are not potent AhR agonists relative to benvitimod. Despite this, the realization that the isopropyl group is a key determinant of benvitimod's AhR activity may warrant further investigation to optimize this substituent. Engineered variants of CylK might provide access to benvitimod itself, or products derived from other secondary alkyl groups not currently accepted as electrophiles (see Figure 2.5 for examples), and/or stilbene nucleophiles with a variety of functionalization patterns. In particular, bioactivity-guided direction evolution could speed up the process of identifying new CylK variants that produce potent

AhR ligands that may be challenging to chemically synthesize. More standard MS-guided directed evolution towards these goals may also afford access to CylK variants that can produce benvitimod derivatives, however, a bioactivity-guided approach may increase screening throughput by reducing the detection time per sample, and by potentially enabling a pooled substrate approach that could be challenging to detect by MS-based methods due to ion suppression of similar species.



CylK	+								+		+		+	
Resveratrol **20b**		+										+	+	
Stilbene **20a**			1000	999	990	900	(nM)	+	+	+	+			
Benvitimod				1	10	100	1000							
Electrophile **5**								+	+			+	+	
Electrophile **16b**										+	+			
% conv. (or equivalent)			0	0.1	1	10	100		5		1		~0.5	

Figure 4.4. Luciferase reporter validation and bioactivity of CylK-accessed stilbene products

All enzymatic reactions were run for at 37 °C for 15 h, with 200 µM of alkyl chloride, 200 µM of stilbene, and 2 µM of CylK. The presence of CylK stilbene adducts was verified by HPLC and LC-MS. All substrates and mixtures were diluted 200-fold for stimulation, to afford 1000 nM of substrates and 100 nM of CylK, unless noted. Bar edges colored blue correspond to added CylK. Gradient mixture of 3,5-dihydroxystilbene (**20a**) and benvitimod are colored from light pink to dark red, corresponding to increased benvitmod concentration. AhR activity is measured as luminescence after 14 h of stimulation, error bars represent the standard deviation from the mean of three biological replicates.

4.2.3 Generation of CylK mutant libraries

With the CylK structure solved, and knowledge of its active site and key residues, we set out to construct mutant libraries of CylK for protein engineering with the goals of enhancing its overall activity, expression levels, and/or solubility. These objectives were chosen as a starting point because although CylK can be heterologously expressed in *E. coli*, yields remain low and the majority of enzyme is insoluble (see Chapter 2.2.1 for more discussion). We envisioned identifying a CylK variant with improved soluble expression and activity, and then using that mutant as a starting enzyme scaffold for further engineering, targeting properties such as an expanded substrate scope and organic solvent tolerance. With this initial goal in mind, we constructed two types of mutant CylK libraries using random mutagenesis- and site saturation mutagenesis. Random mutagenic strategies can be beneficial when it is not known which specific residues of an enzyme should be targeted for modification.[226] We chose to employ this approach because it was unclear what changes to the primary sequence of CylK might enhance its soluble expression levels. However, because random mutagenesis libraries tend to have high wild-type background levels and require larger screens to identify desirable variants, we also chose several amino acid sites for single site saturation mutagenesis (SSM) in which a single residue is mutated to all possible amino acids in a pooled library format. This type of focused mutagenesis strategy has demonstrated considerable success in numerous enzyme directed evolution campaigns,[242] and 95% probability of full sequence space coverage at a single mutagenized site can achieved with approximately 117 samples.[243] In this way, a single 96-well plate would be nearly sufficient to obtain coverage at each mutagenized site. Furthermore, the availability of structural information meant we could target residues located near or distal to CylK's catalytic center for SSM, increasing the odds of discovering beneficial mutations in key protein regions that may be important for enhanced activity, based on our understanding of this enzyme.

To construct a pooled random mutagenesis library of CylK without targeting the plasmid vector, we utilized restriction cloning and error-prone PCR (epPCR).[xix] The epPCR method relies on the low fidelity of DNA

[xix] This work was done in collaboration with Michelle Wang.

polymerases when *in vitro* PCR reactions are supplemented with manganese or non-native deoxyribonucleoside triphosphate (dNTP) analogues.[226,244,245] In short, we PCR amplified the *cylK* gene from our expression construct which already contained XbaI and NdeII restriction sites, followed by a second round of PCR supplemented with 100 μM $MnCl_2$, and final vector assembly. Sequencing 20 colonies from this library revealed 50% wild-type background rate, 10% truncated insert, and approximately 1-2 mutations/kilobase in the remaining 40% (Table 4.1). CylK is slightly greater than 2 kilobases. We were encouraged by the abundance of missense and silent mutations that might be expected to alter CylK expression and/or folding. As expected, a higher manganese level of 250 μM resulted an amplified mutational rate (~20 mutations/kilobase), while a lower manganese level of 25 μM reduced the mutational rate to <1 mutation/kilobase.

Table 4.1. Random mutagenesis library statistics

Mutation Type	Count
wild-type (no change)	10
missense (point mutation)	4
nonsense (premature stop)	1
silent (synonymous codon)	1
indel (frame shift)	3

Error prone PCR (epPCR) with 100 μM was used to construct random mutagenesis CylK library. Statistics are from Sanger sequencing runs with ~75% coverage of the full gene.

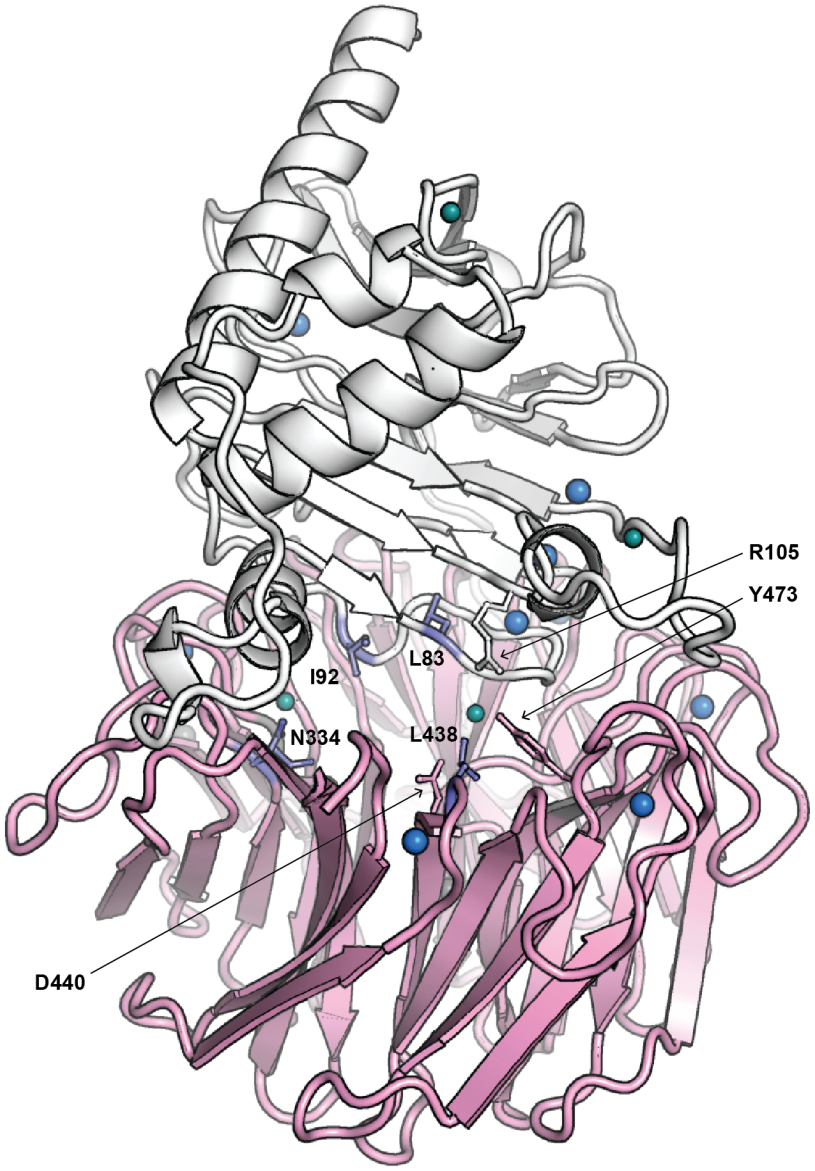


Figure 4.5. Sites selected for single saturation mutagenesis

An overall view of the bromide-soaked structure of CylK (PDB 7ROO). Select residues are displayed in stick format and labelled, residues targeted for mutagenesis are purple. Calcium ions are shown as blue spheres, bromide ions in teal, magnesium in dark grey spheres. The N-terminal domain is depicted as a light grey ribbon diagram and the C-terminal domain is shown as a pink ribbon diagram.

Next, we selected four residues for SSM, namely Leu83, Ile92, Leu438, and Asn334 (Figure 4.5).[xx] These residues span the active site cleft but are not essential for catalysis (see Figure 3.17 for a structural

[xx] This work was done in collaboration with Michelle Wang.

representation of key residues). We hypothesized that these sites might play minor roles in substrate binding and/or structural roles that may influence the enzyme's overall solubility and activity. To construct a pooled SSM library at each residue, we used a variation of golden gate cloning[246] combined with a pooled primer approach[247,248] in order to achieve a codon distribution that avoids premature stops and encodes a uniform amino acid distribution, in other words, 1/20 odds for each residue. Gratifyingly, this cloning methodology was straightforward to implement, and Sanger sequencing confirmed the desired distribution of DNA bases at these selected sites.

4.2.4 Pilot of bioactivity-guided direction evolution

Having demonstrated the theoretical feasibility of using bioactivity-guided directed evolution with stilbene ligand starting materials and an AhR bioactivity readout, we next wanted to pilot this methodology by examining the potential ability of CylK mutant libraries to produce benvitimod from 3,5-dihydroxystilbene (**20a**) and 2-chloropropane (*i*-PrCl). Although wild-type CylK can use stilbene **20a** as a nucleophile, it does not accept *i*-PrCl as an electrophile. First, we examined a time course of our reporter cell line (HT29-Lucia™ AhR, Invivogen) exposed to benvitimod because we had previously varied stimulation time from 1–14 h dependent on qRT-PCR or luciferase luminescence readouts. We expected the ideal exposure time to vary between these assay formats because a transcriptional response precedes a translational one. Furthermore, the cell line supplier suggested measuring luminescence after 24–48 h although we hoped to minimize incubation times. This demonstrated that the luminescence increased linearly from 16 to 24 and 49 h for both conditions (Figure 4.6A). The highest relative luminescence fold change was observed at 16 h, however, the 24 and 48 h time points had less variability (Figure 4.6B), and so we concluded that taking luminescence measurements after 24 h might minimize false negative results.

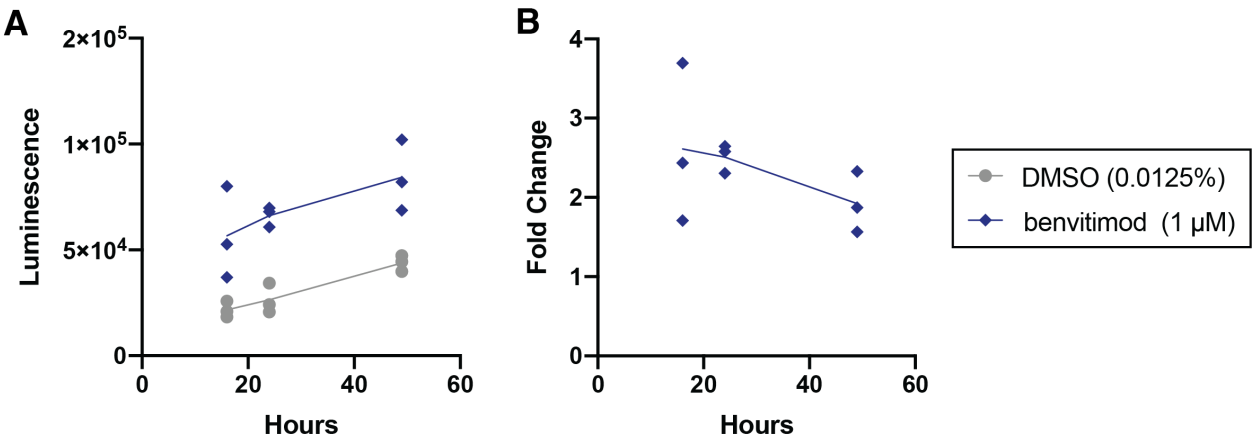


Figure 4.6. Timing of luciferase reporter response

Replicates of HT29-Lucia AhR cells (~300,000) were stimulated with 1.0 μM benvitimod or DMSO vehicle for 16, 24 or 49 h. Each replicate culture was 1.25 mL, and in 24-well tissue culture dish estimated to contain 6.25x greater surface area than the standard 96-well dishes used for larger screens. Volume and cell count were scaled analogously. Time points were taken by removing 20 μL of culture supernatant and storing at 4 °C until measuring luminescence in parallel. Lines represent the mean of three biological replicates (plotted as diamonds or circles).

Next, to pilot bioactivity-guided direction evolution of CylK towards accepting *i*-PrCl as an electrophile (in order to produce benvitimod) we selected Leu83 and Ile92 as residues to target for mutagenesis considering their proximity to the active center and putative binding site of the electrophile's alkyl tail. Furthermore, we more generally speculate that hydrophobic portions of the active site might play a role in alkyl halide binding, and structural changes to these regions could potentially accommodate smaller alkyl halide electrophiles. With our SSM plasmid libraries in hand, we used our plate-based lysate activity assay (initially developed in Chapter 3) to screen 46 members from both Leu38 and Ile92 libraries. This represents a 99% probability of discovering one of the top two performing two variants at each site,[243] presuming two mutants exist with the desired activity. Note, full sequence coverage at 95% or 99% probability would require screening 117 or 149 clones, respectively. We also included two wild-type CylK controls with competent substrates **5** + **7** to ensure proper expression and reaction conditions, two wells expressing enzyme but no added substrates as negative controls, and two wells with benvitimod as positive controls for bioactivity. Following growth and expression, enzymatic reactions were initiated with the addition of substrate-lysozyme mix and allowed to react for 21 h at 30 °C. This lower temperature (typically 37 °C) was

selected to prevent substantial *i*-PrCl evaporation near its boiling point of 36 °C. Then, our reporter cell line was stimulated with dilutions of crude reaction lysate (or 1% benvitimod/stilbene **20a**) for a final total ligand concentration of 1.0 µM (as previously in Chapter 4.2.2) for 41 h. Our results suggest that none of the CylK mutants assayed coupled stilbene **20a** with *i*-PrCl to produce benvitimod, as even the highest luciferase levels (~40,000) fall within a gaussian-like distribution below the positive control values (Figure 4.7). By LC-MS, we verified that the CylK reaction controls coupled substrates **5** + **7** to produce adduct **8**, and mutant enzyme reactions with the five highest luminescence values did not contain trace benvitimod product.

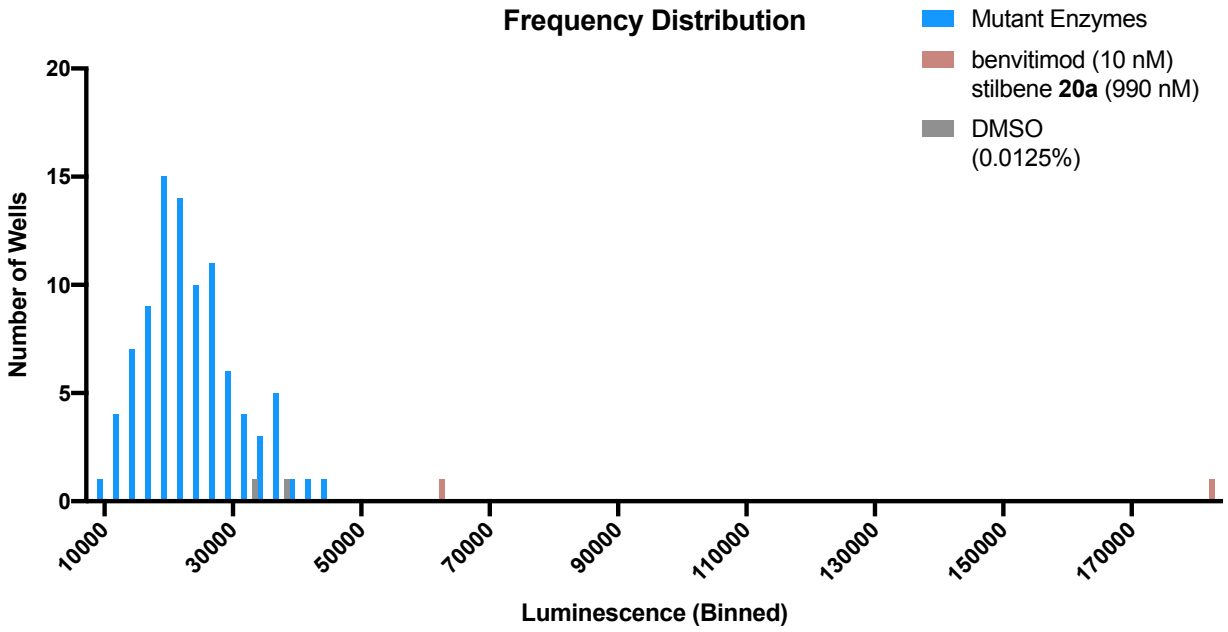


Figure 4.7. Relative bioactivity of enzymatic reaction products

HT29-Lucia AhR cells (~45,000) were stimulated with 1.0 µM total ligand for 41 h. Mutant enzyme lysate reactions with stilbene **20a** and *i*-PrCl are displayed in blue, positive controls with benvitimod and stilbene **20a** are in dark red, and negative controls with enzyme lysate and DMSO vehicle are grey. Cultures were assayed at 200 µL volume in a 96-well tissue culture dish. Luciferase reporter values represent single biological replicates, and are displayed in a histogram format to determine the frequency distribution of these data points.

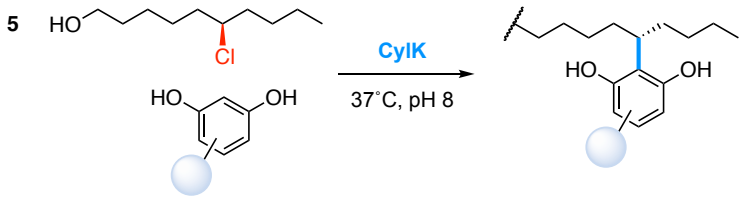
Overall, we were encouraged by the feasibility of this workflow from a technical perspective, and anticipate that liquid handling robotics could help scale up and streamline the entire process. From the results of this

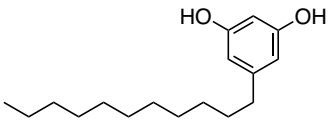
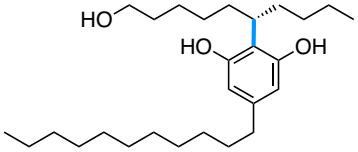
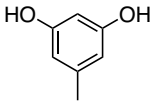
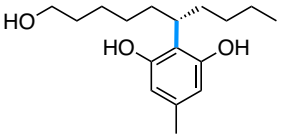
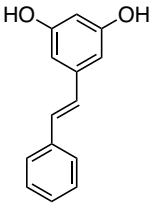
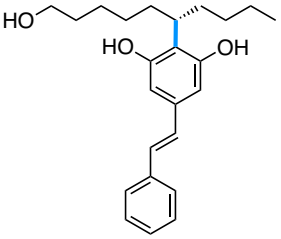
pilot run, we recommend that enzymatic reactions are repeated in biological triplicate, as variability was observed for the CylK reaction controls (with substrates **5** + **7**, measured by LC-MS) as well as the bioactivity positive controls (1% benvitimod conversion equivalent, measured by luminescence). This also highlights the challenges of using enzyme evolution to directly identify mutants that accept a new substrate in contrast to a substrate walking approach.[249]

4.2.5 UPLC-MS guided protein evolution

Based on our pilot of bioactivity-guided directed evolution, we speculated that it would first be beneficial to improve the overall activity, expression levels, and/or solubility of CylK before attempting to expand the substrate scope to include new electrophiles such as *i*-PrCl. This may be especially worthwhile considering that, although stilbene **20a** is a competent CylK substrate, conversions with wild-type enzyme are limited to ~5-10% which further reduces the odds of a productive reaction between stilbene **20a** and i-PrCl. To directly screen for enhanced activity, we developed rapid UPLC-MS screening conditions to detect adducts **8**, **29a**, **21a** that arise from the reactions between alkyl chloride electrophile **5** and undecylresorcinol (**7**), orcinol (**18a**), and 3,5-dihydroxystilbene (**20a**) (Table 4.2). With this method we can analyze a full 96-well plate in approximately 5 h. Of note, we also confirmed that these products were detectable at a concentration reflecting approximately 0.5% conversion and in the presence of *E. coli* cell lysate.

Table 4.2. Select CylK reactions for activity optimization



Resorcinol Nucleophile	Product	Conversion (%)
 **7**	 **8**	99
 **18a**	 **29a**	20
 **20a**	 **21a**	4

Product formation (percent conversion) of **8** and **21a** is reproduced from Table 2.4 and Table 2.9, minor variability is observed between enzyme preparations. Product formation of **29a** was quantified by HPLC, as described in Materials and Methods, value reported is from one biological replicate. All enzymatic reactions were run for at 37 °C for 24 h, with 200-400 µM of alkyl chloride and 200-400 µM of resorcinol nucleophile.

Next, to begin screening CylK variants for enhanced activity, expression, and/or solubility, we selected our epPCR random mutagenesis library in addition to our Asn334 and Leu438 SSM libraries for this assay.[xxi] We reasoned that epPCR could access sequence space that may improve solubility and/or expression. While Asn334 is not believed to be substrate interacting, mutants at this site might alter overall active site

[xxi] This work was done in collaboration with Vanessa Dippon and Michelle Wang.

architecture or activity. Leu438 is a nonessential substrate interacting residue, and mutants at this site may enhance the binding or reactivity of non-native substrates. In total, we screened 12 epPCR library members, 31 Asn334 SSM library members (>95% probability of encountering one of two top variants), and 47 Leu438 SSM library members (>99% probability of encountering one of two top variants).[243] We had hoped to screen approximately 90 members per library, however our numbers were limited by transformation efficiency. For future efforts, the use of electrocompetent or fresh chemically competent cells would be advisable to increase transformation efficiency. We additionally included wild-type CylK and sterile controls.

Following growth and expression, three distinct enzymatic reactions were initiated (alkyl chloride **5** + nucleophile **7**, **18a**, or **20a**) by splitting the expression culture into three aliquots and combining with substrate-lysozyme mix, and allowed to react for 22 h at 37 °C. Unfortunately, alkylated orcinol (**29a**) was not detected in this lysate reaction format, and the ion counts of alkylated stilbene (**21a**) were too low to reliably quantify. However, LC-MS analysis of dialkylated resorcinol (**8**) reactions revealed clear outliers with significantly higher activity than both wild-type enzyme and the general population (Figure 4.8). Interestingly, the majority of enzyme constructs (62 of 90) appeared to be inactive and the Leu438 SSM library accounted for 46 of these inactive variants, suggesting that substituting this residue is not beneficial in this context. Of the remaining inactive variants, 3 and 13 came from the epPCR and SSM Asn334 libraries, respectively. Of the top 10 performing constructs, 6 were from the epPCR library, and 3 and 1 came from the SSM Asn334 and Leu438 libraries, respectively. Of note, we ensured that these 'winners' reflected true hits with high product ion counts, and not inflated values from our relative activity normalization formula.

We then sequenced the top 5 hits and determined that 3 were wild-type, suggesting that our assay has a high false positive hit rate (Table 4.3). Interestingly, the two non-wild-type constructs were from the SSM Asn334 and Leu438 libraries, and contained a silent codon change and 25 residue N-terminal truncation, respectively. The N-terminal truncation resulted from an indel that caused a premature stop at residue 18 and then followed by an alternate start codon at residue 26 (Figure 4.9). This indel might originate from

trace contamination from primer synthesis or PCR error as this site overlapped with a cloning primer. Although these two CylK variants could be false positives, a codon change could alter translation kinetics and protein folding and an N-terminal truncation might result in improved solubility, expression, and/or activity; and so we carried these hits forward for validation.

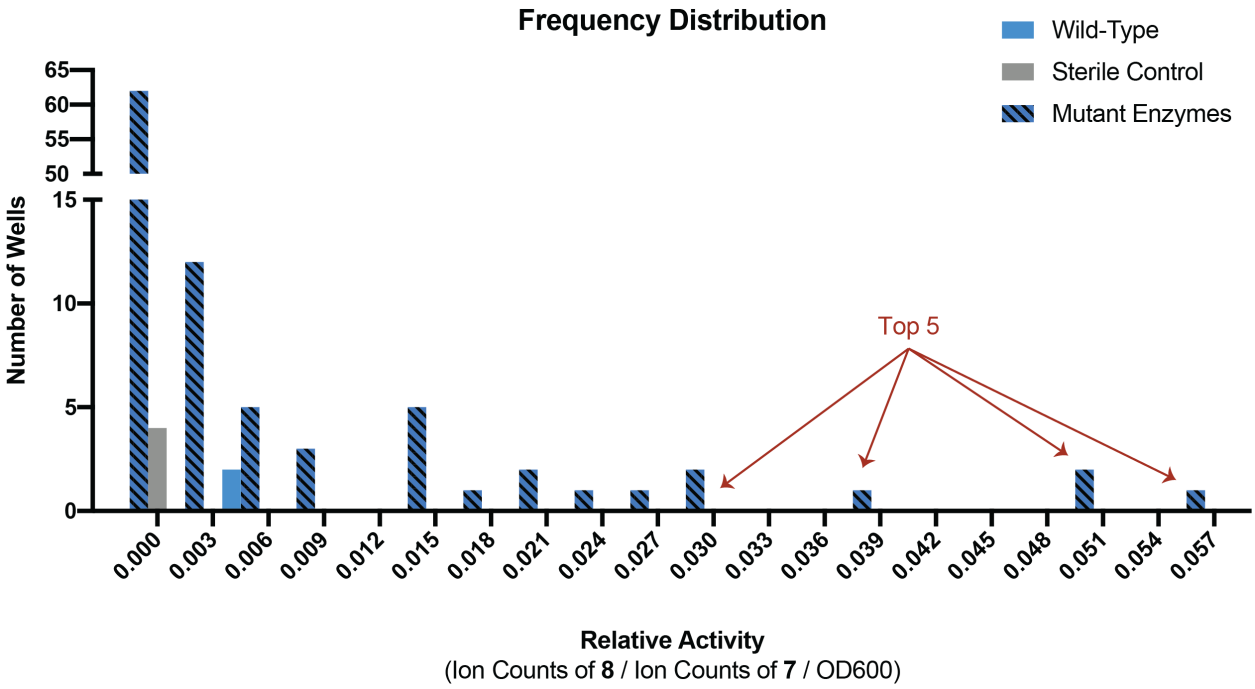


Figure 4.8. Histogram of mutant screen for enhanced CylK activity

LC-MS results of mutant enzyme lysate reactions producing adduct **8**. Wild-type enzyme values are displayed in blue, mutants in blue/black stripes, and sterile negative controls in grey. Relative activity values represent single biological replicates, and are defined the three-part quotient of product **8** ion counts, divided by starting material **7** ion counts, divided optical density at 600 nm (OD600) in order to normalize by autosampler injection and growth differences, respectively. Results are displayed in a histogram format to determine the frequency distribution of these data points.

Table 4.3. Sequencing results of most active CylK variants

Well	Library	Mutation
A3	epPCR	wild-type
A9	epPCR	wild-type
A12	epPCR	wild-type
C6	SSM N334	silent (AAC>AAT)
H9	SSM L438	N-terminal truncation

Full plasmids were analyzed by next-generation sequencing conducted at the Massachusetts General Hospital Center for Computational & Integrative Biology DNA Core.

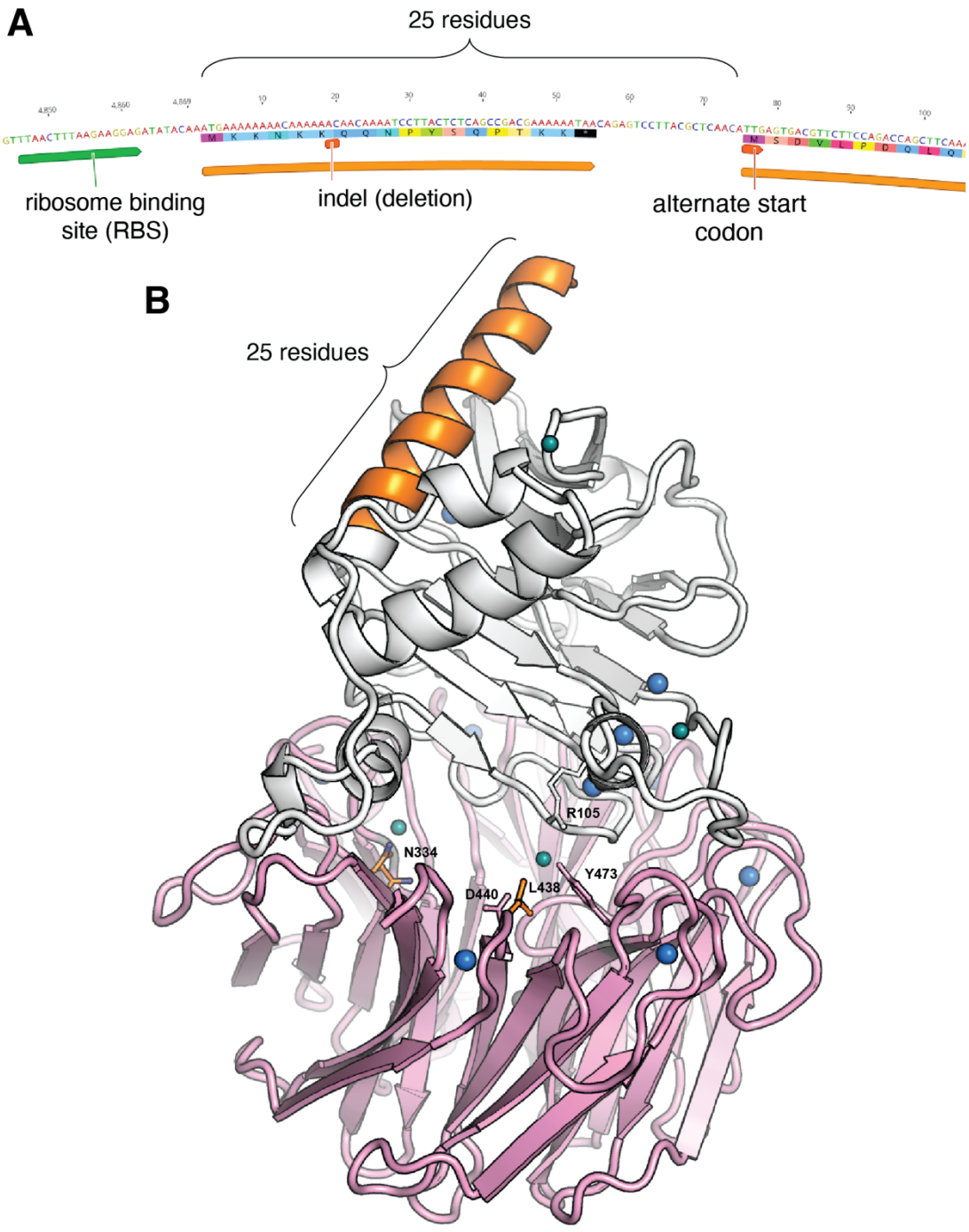


Figure 4.9. Putative N-terminally truncated CylK variant

(*A*) Primary sequence view of N-terminal truncation variant of CylK. (*B*) An overall view of bromide-soaked CylK (PDB 7ROO) highlighting 25 residue N-terminal truncation in orange. Select residues are displayed in stick format and labelled, residues targeted for mutagenesis are orange. Calcium ions are shown as blue spheres, bromide ions in teal, magnesium in dark grey spheres. The N-terminal domain is depicted as a light grey ribbon diagram and the C-terminal domain is shown as a pink ribbon diagram.

4.2.6 Validation and characterization of enhanced CylK variants

Next, we repeated these reactions in a similar assay format but included six replicates of each construct or control in order to increase the statistical significance. We measured relative ion counts of dialkylresorcinol **8** in two ways, first normalizing product ion counts by starting material ion counts in order to control for potential loading differences (Figure 4.10A), and then additionally normalizing by optical density at 600 nm (OD600) to control for growth differences (Figure 4.10B). We also checked raw product ion counts to ensure these formulas did not artificially inflate relative ion counts. Based on these data, both mutant constructs appear to have enhanced CylK activity when controlling for growth, while the N334 silent codon change (from AAC to AAT) has enhanced CylK activity when normalizing by starting material alone, although there seems to be significant variability between replicates. Furthermore, average raw product ion counts (MS/MS) for both mutants were greater than 8000, while wild-type had ~6000 representing at least a 33% activity enhancement. We speculate that the N334 silent codon change increases soluble expression and the pool of active enzyme, while the N-terminal truncation might be more inherently active and/or soluble.

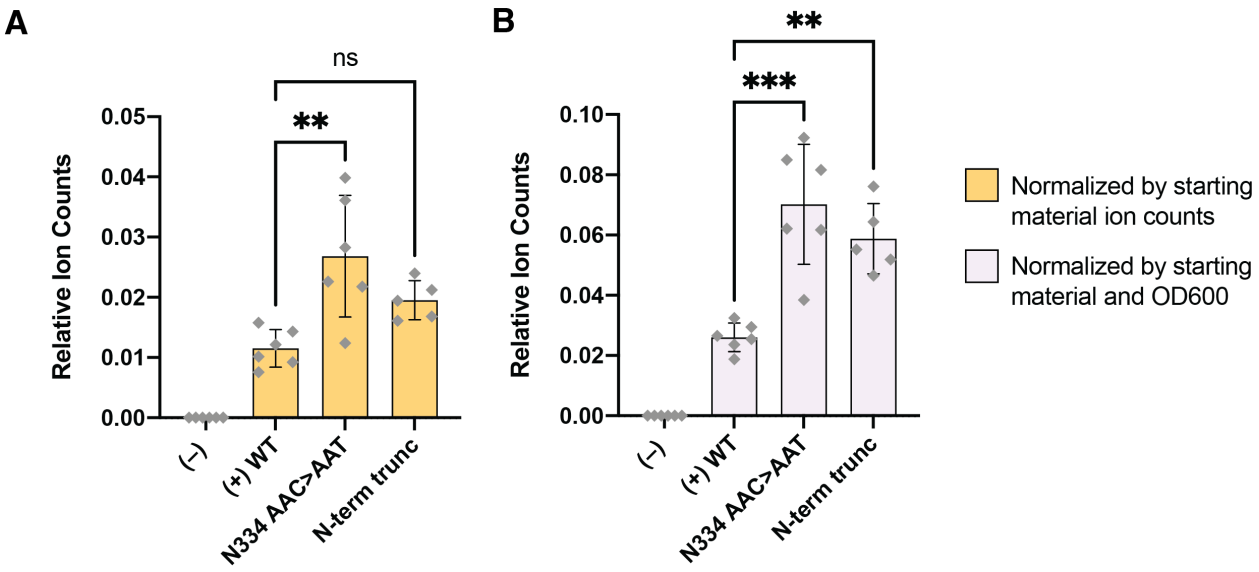


Figure 4.10. Activity validation of enhanced CylK variants

LC-MS results of CylK lysate reactions producing adduct **8**. The no enzyme sterile negative control is referred to as '(–)' and wild-type enzyme as '(+) WT'. Bar heights represent the mean of six biological replicates (plotted as grey diamonds), and error bars represent the standard deviation from the mean. Relative ion counts are defined by (*A*) product ion counts divided by starting material ion counts, or (*B*) that quotient additionally divided by OD600. *P* value was determined by ordinary one-way ANOVA; 'ns' not significant, (*) 0.0332, (**) 0.0021, (***) 0.0002.

To further characterize these CylK variants, we cloned a C-terminally tagged green fluorescent protein (GFP) fused construct of each CylK variant with the fluorescent protein appended after the Strep tag (see Materials and Methods for construct information) in order to quantify relative protein expression levels that may contribute to activity differences from wild-type. First, we validated that wild-type CylK with and without the GFP fusion had comparable activity levels (Figure 4.11A-B). Next, we measured OD600 and GFP fluorescence which revealed that although the GFP fusion construct slightly outgrew wild-type CylK, only the GFP fusion demonstrated fluorescence (Figure 4.11C).

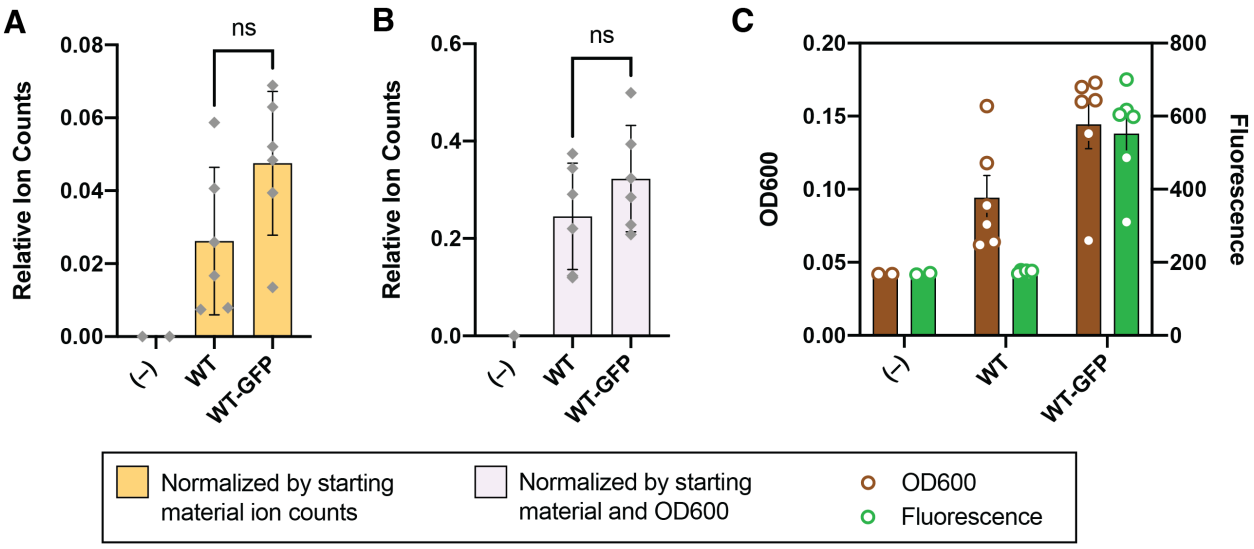


Figure 4.11. Activity and fluorescence of CylK-GFP

LC-MS, UV-Vis, and fluorescence analyses of CylK expression cultures and lysate reactions producing adduct **8**. The no enzyme sterile negative control is referred to as '(–)', wild-type enzyme as 'WT', and GFP-fused wild-type enzyme as 'WT-GFP'. (*A*) LC-MS analysis of lysate reactions. Relative ion counts are defined by product ion counts divided by starting material ion counts, or (*B*) that quotient additionally divided by OD600. *P* value was determined by ordinary one-way ANOVA; 'ns' not significant. Bar heights represent the mean of six or two biological replicates (plotted as grey diamonds), and error bars represent the standard deviation from the mean. (*C*) UV-Vis and fluorescence analyses of enzyme expression cultures. Growth is measured by OD600 and is plotted in brown on the left y-axis, while green fluorescence is plotted in green and on the right y-axis.

Then, having validated the activity and fluorescence of the CylK-GFP fusion construct, we repeated this assay on both CylK variants as GFP fusion proteins which revealed that there was no statistically significant increase in product conversion relative to wild-type, and a high degree of variability between biological replicates (Figure 4.12A-B). This suggests that if these two mutations do result in some form(s) of enzyme enhancement, the GFP fusion construct might suppress any differences from wild-type. However, it was striking to note that the N-terminally truncated CylK variant (ΔN25-CylK) had considerably lower fluorescence levels (Figure 4.12C), suggesting that this mutation results in decreased enzyme expression levels. This is particularly interesting considering the fact that ΔN25-CylK demonstrated comparable (or higher) activity to wild-type in the same assay format.

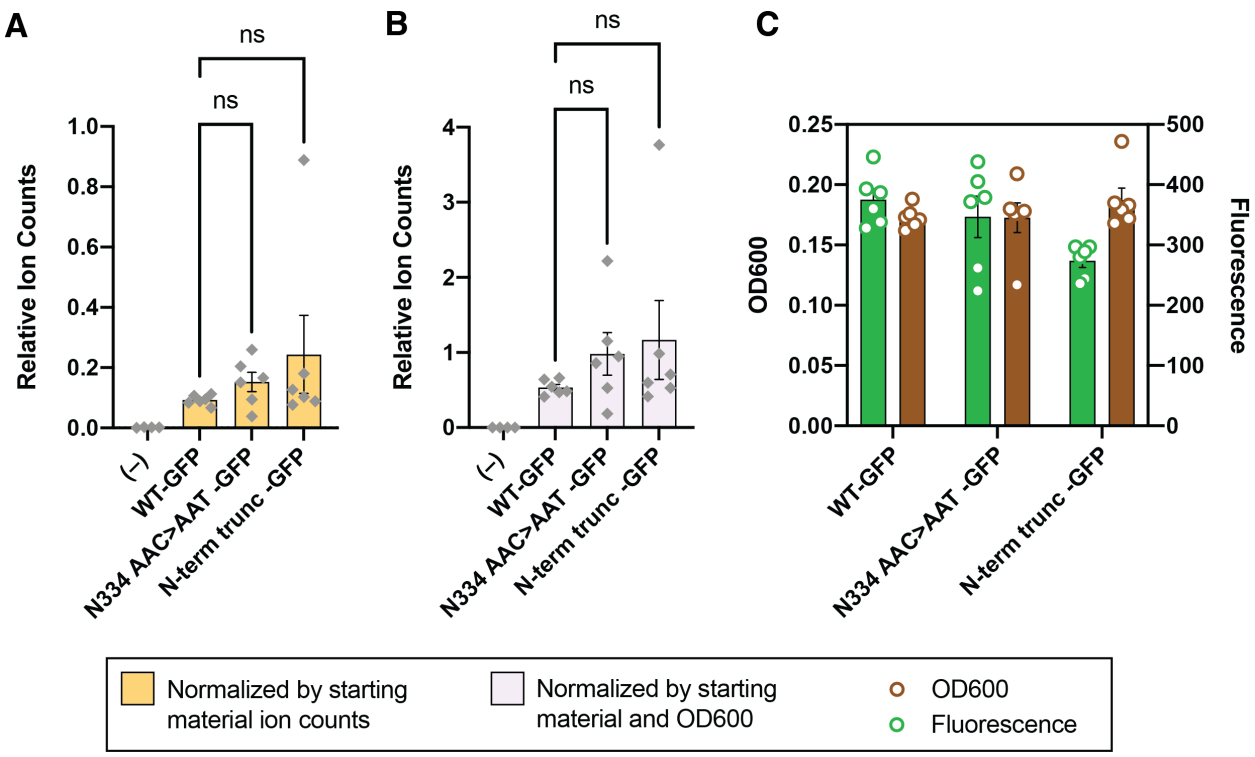


Figure 4.12. Activity and fluorescence of CylK-GFP variants

LC-MS, UV-Vis, and fluorescence analyses of CylK expression cultures and lysate reactions producing adduct **8**. The no enzyme sterile negative control is referred to as '(–)', GFP-fused wild-type as 'WT-GFP', and similarly for GFP-fused variants. (*A*) LC-MS analysis of lysate reactions. Relative ion counts are defined by product ion counts divided by starting material ion counts, or (*B*) that quotient additionally divided by OD600. *P* value was determined by ordinary one-way ANOVA; 'ns' not significant. Bar heights represent the mean of six or four biological replicates (plotted as grey diamonds), and error bars represent the standard deviation from the mean. (*C*) UV-Vis and fluorescence analyses of enzyme expression cultures. Growth is measured by OD600 and is plotted in brown on the left y-axis, while green fluorescence is plotted in green and on the right y-axis.

We speculate that lower expression of ΔN25-CylK is likely due to do the use of 'UUG' as a start codon instead of the standard start codon 'AUG'. In *E. coli*, 'AUG' is used more frequently than 'UUG' (90% versus 1%), and it has been demonstrated that β-galactosidase with a 'UUG' start had 11.2% activity relative to the same construct with a 'AUG' start.[250] Furthermore, the presence of a leader peptide in our ΔN25-CylK construct, as well as the increased distance between ribosome binding and site start codon (8 versus 82 bases) likely effects translation kinetics and fidelity. Despite potentially decreased translation, more of the total ΔN25-CylK protein may be soluble or active compared to wild-type and could explain the enhanced

activity of this mutant seen previously. Pilot expressions of all CylK-GFP constructs on a 2 L scale confirmed that ΔN25-CylK had significantly lower GFP levels by visual inspection. Notably, western blotting of soluble lysate from these expressions revealed that soluble enzyme levels were comparable across variants (Figure 4.13), suggesting that the ΔN25-CylK-GFP has increased soluble expression relative to total protein content.

Efforts to purify and quantify relative activities of wild-type and ΔN25-CylK-GFP were complicated by a proteolysis event that resulted in a new protein species between ~80–90 kDa that co-eluted with CylK-GFP and was not previously observed in soluble lysate. We speculate that including protease inhibitors in affinity purification buffers might suppress the formation of this putative truncation product, as protease inhibitors were included in lysis buffers. Future efforts to examine the CylK variants without GFP fusions could reveal if both the N334 silent codon change and ΔN25-CylK have enhanced soluble expression by western blotting; proteolysis should be minimized in the constructs without GFP. Furthermore, the GFP fusions might alter the behavior of these systems. Rigorous protein purification of ΔN25-CylK could determine if its catalytic activity is altered relative to wild-type. We suggest that it would also be revealing to alter the start codon of wild-type CylK to 'UUG' from 'AUG', remove the leader peptide from ΔN25-CylK, alter the start codon of ΔN25-CylK to 'AUG' from 'UUG', and/or combine ΔN25-CylK with the N334 silent codon change. This work may identify an enhanced CylK expression construct to be used for subsequent rounds of directed evolution.

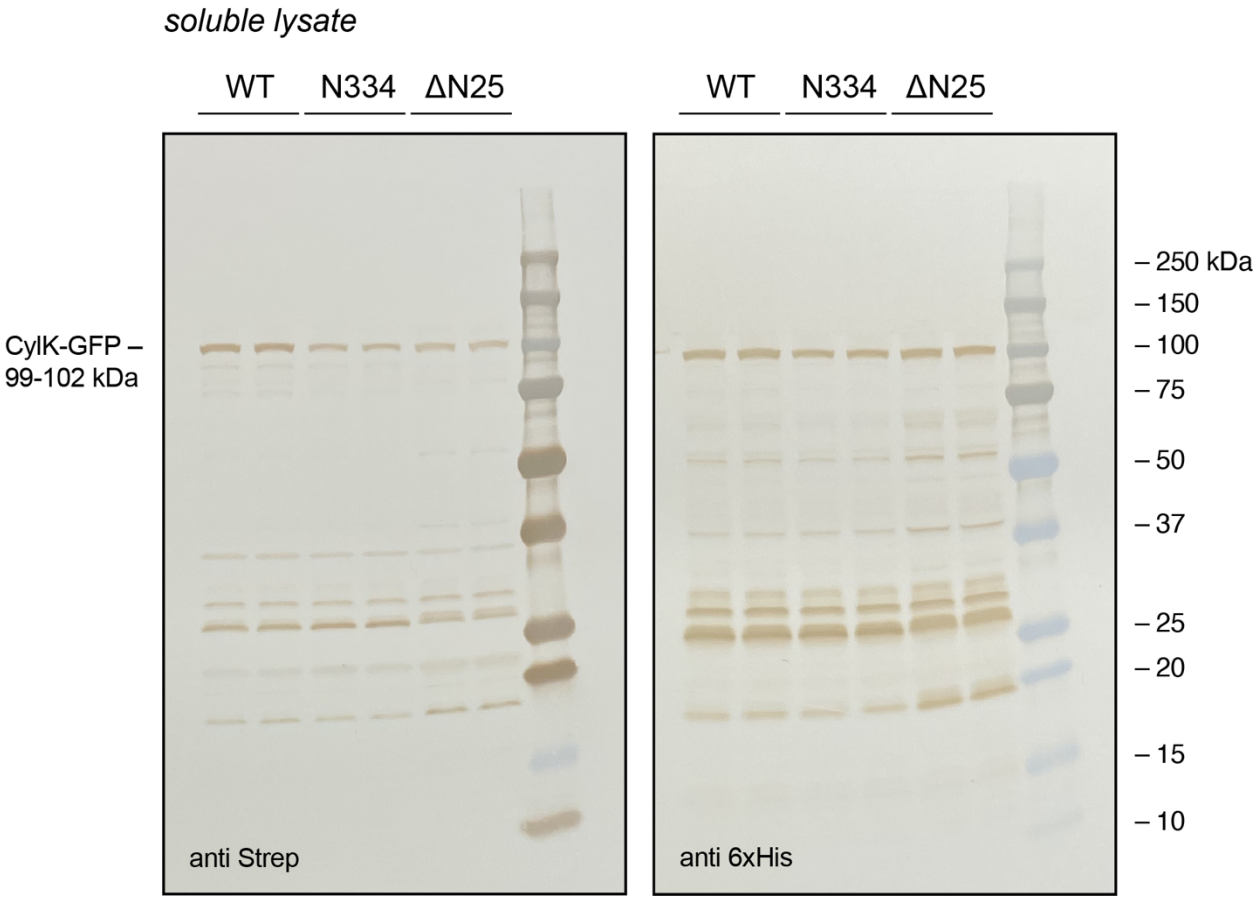


Figure 4.13. Western blotting of evolved CylK variants

Western blots of CylK-GFP variants expressed on a 2 L scale with anti-Strep-HRP (IBA) or anti-6xHis-HRP (Sigma) in order to visualize variant enzyme expression and solubility. WT = wild-type, N334 = silent codon change (from AAC to AAT), ΔN25 = N-terminal truncation. The soluble fraction of each variant is loaded twice (technical duplicates). All Blue Precision Plus Protein Prestained Standard (Bio-Rad) was used to visualize molecular weight standards.

4.3 Conclusions and Future Directions

In summary, we have characterized the enzyme CylK by elucidating its non-biological substrate scope, determining the structural basis for this transformation, and identifying promising starting points and methodologies for protein engineering. This work reveals that CylK and homologous enzymes, with their inherent substrate promiscuity, are poised to be applied and engineered as biocatalysts using our structural information and the platforms we developed for directed evolution as a guide. Furthermore, we have proposed a plausible mechanism for catalysis that invokes positioning and activating the reactive partners in close proximity for a concerted S_N2-like reaction that involves Arg105, Tyr473, and Asp440 as key catalytic residues. Cumulatively, this effort has provided fundamental insight into this intriguing reaction that represents the only known biological example of aromatic ring alkylation with an alkyl halide electrophile.

The enzyme's active site location, at the solvent exposed opening to the interdomain cleft, is in agreement with our observation that CylK can accept resorcinol derivatives with rigid substitutions on the aryl core. However, CylK's strict requirement for resorcinol nucleophiles is likely due to an active site architecture that evolved to bind and activate resorcinol rings by interacting with their phenolic groups. Future structural and biochemical work with BrtB and NpF3370 might reveal differences in nucleophile interactions that could shed light on the mechanistic differences and residues that play a role in determining nucleophile specificity, and this would also be of great interest for directed evolution. Considering the electrophilic reaction partners, CylK's inability to accept smaller secondary alkyl halides (<10 carbons) without polar head groups is most likely attributable to the specific mode of binding these substrates, as features such as polar functional groups on electrophilic substrates are not expected to play a role in catalysis. Future characterization of BrtB and NpF3370 might illuminate the binding modes of their substrates, and structures of CylK with non-biological electrophiles lacking the resorcinol moiety would also be of interest. With practical applications in mind, directed evolution of CylK towards accepting alternate alkyl halide substrates might benefit from a substrate walking approach[249] in which the alkyl halide electrophile is successively modified between rounds of evolution, rather than immediately screening for activity on a substrate considerably different than the native electrophile that might require more significant active site changes.

In this final Chapter, we highlight our preliminary efforts to engineer the enzyme CylK for enhanced solubility and activity, and pilot a new directed evolution screening methodology. This screening strategy, which we refer to as bioactivity-guided directed evolution, seeks to accelerate protein engineering campaigns by harnessing cell-based biological reporter assays to screen reaction products in high-throughput. We envision applying this methodology for late-stage enzymatic transformations that may give rise to a diversity of products. In this way, the early stages of the medicinal chemistry process, such as target compound optimization, can potentially be merged with protein engineering in order to identify lead compounds and key biocatalytic steps in tandem. In the case of CylK, we propose initially using this workflow to identify enzyme variants that can couple stilbene nucleophiles and alkyl halide electrophiles to produce novel derivatives of benvitimod with analogous or enhanced bioactivity, namely, AhR agonism. Towards this goal, we demonstrated that a model unalkylated starting material (**20a**) does not elicit an AhR-related response, while mixtures of this starting material and as little as 0.1% product (benvitimod, representing a crude reaction mixture) activates the AhR in a plate- and human cell-based luciferase assay. We envision screening workflows in which discrete substrate pairs are used for the enzymatic reaction step, or even pooled substrate panels are mixed together. In the pooled approach, after wells are assayed for bioactivity, the crude reaction mixtures from select wells can be further examined by LC-MS in lower throughput to identify the presence of putative products.

In a pilot run of bioactivity-guided directed evolution with 3,5-dihydroxystilbene (**20a**) and *i*-PrCl substrates, we did not observe the production of benvitimod, however we speculate that CylK variant(s) that can perform this transformation might be several steps away from wild-type considering that *i*-PrCl is not accepted as an electrophile and 3,5-dihydroxystilbene (**20a**) does not give significant conversion with other electrophiles. Instead, it might be beneficial to first identify a variant of CylK that can more readily accept this stilbene nucleophile paired with a competent alkyl halide electrophile; in parallel, one could also use a substrate walking approach[249] to identify a CylK variant that accepts *i*-PrCl as an electrophile when paired with a more readily accepted alkyl resorcinol nucleophile. Then, the mutations associated with these two

hypothetical CylK variants could be merged or shuffled, and one of these scaffolds could be subjected to additional rounds of genetic diversification and screening for the production of benvitimod. At this point, a bioactivity-guided approach might be especially beneficial to rapidly screen large sequence space that would be slow to assay by LC-MS. Engineering of enzymes that catalyze intermolecular reactions, such as CylK, may require larger screens and careful optimization of each substrate's unique mode of binding and reactivity. For instance, although transaminases are a prominent and widely used class of biocatalyst, the co-substrate of isopropylamine is not broadly accepted as an amine donor and both protein engineering and enzyme discovery efforts have sought to identify transaminases that can tolerate high loadings of isopropylamine with model partner substrates.[251] We envision that a similar approach might be beneficial for CylK protein engineering.

Another obstacle associated with the use of CylK as a biocatalyst is its limited soluble expression in *E.* coli. The fact that the majority of heterologously expressed CylK is insoluble and inactive also presents a challenge for protein engineering in that low enzyme loadings in lysate reaction format might suppress the detection of any enhanced activity, such the use of new substrates not accepted by wild-type CylK. Instead, in order to identify CylK variants with inherently improved activity, expression levels, and/or solubility, we used a rapid UPLC-MS method to screen CylK mutant libraries for improved conversion in a reaction already catalyzed by wild-type enzyme. We examined this activity in CylK variants derived from random and targeted mutagenesis libraries, and although assay variability likely contributed to a high false positive rate, we identified two CylK mutants that may have enhanced activity. The first is a silent codon change at Asn334, which may improve the kinetics of protein expression and folding, resulting in a greater fraction of soluble enzyme. The second is an N-terminal truncation of the first 25 residues of CylK. Interestingly, this N-terminal truncation mutant appears to have lower expression levels than wild-type, suggesting that more of the total enzyme is soluble. Further characterization of these variants via protein purification is warranted, and it may be beneficial to combine these changes into a single construct. Either of these two variants may serve as a starting point for additional rounds of mutagenesis and screening.

Looking forward, future protein engineering efforts might involve docking non-biological substrates into the active sites of CylK and its homologs in order to identify additional residues to target for mutagenesis. It would be particularly interesting to visualize the differences between the three characterized family members as well as uncharacterized members. Furthermore, the random and targeted mutagenesis libraries we created could be probed further because our sampling of this genetic diversity was limited to pilot scale work on only a few substrate pairs. However, DNA shuffling between homologous enzymes might be another fruitful and interesting source of genetic diversity.[226] Codon optimization or harmonization might also be beneficial for this system. Future screening approaches may consider assaying enzyme variants in biological triplicate in order to minimize the high rate of false positives we observed. Another strategy to potentially control for variability would be to maintain the GFP fusion within enzyme variant libraries so that enzyme loading could be accounted for. The fluorescence readout from GFP might even be used to screen for enhanced solubility, although we suspect that fluorescence correlates to CylK expression levels and not solubility.

More broadly, while our work characterizing CylK has enhanced our fundamental understanding of enzymatic interactions with alkyl halide substrates, it has also generated a number of unanswered questions and exciting future research directions. In particular, the specific mode of alkyl halide binding can be probed further by cocrystalizing CylK with alkyl halide substrates in the absence of free halides. Furthermore, structurally characterizing the Phe mutant at Tyr473 might inform our understanding of alkyl halide interactions with this residue. Additional mechanistic information might also be gleaned from quantum mechanics/molecular mechanics modeling of a transition state in the active site, with special attention paid to substrate-residue interactions and potential changes to these binding modes over the course of reaction. Further insight, in addition to practical applications, might also be attained if a CylK-inspired biomimetic catalyst could be designed to catalyze analogous reactions with secondary alkyl halide electrophiles. Despite these open mechanistic questions, we now have ability to functionally annotate CylK-like genes, and we can prioritize other biosynthetic gene clusters that encode related enzymatic machinery. Other classes of natural products or distinct metabolic processes may be associated with distantly related

CylK homologs, such as those without partner halogenases and halide activating residues. Considering that the cylindrocyclophane producers have been isolated from diverse locations across the world, many more unsequenced cyanobacteria with CylK homologs likely exist. The metabolites associated with CylKs may play interesting roles in cyanobacterial physiology or ecosystem dynamics, and might serve as starting points for the identification of new therapeutic compounds.

4.4 Materials and Methods

4.4.1 LC-MS conditions

Ultra performance liquid chromatography (UPLC) was conducted using a Waters Acquity UPLC H-Class System (Waters) equipped with an Acquity UPLC BEH C18 column (130 Å, 1.7 µm, 2.1 mm X 50 mm) with VanGuard pre-column (130 Å, 1.7 µm, 2.1 mm X 5 mm) and the following method: 10 µL injection volume, 0.8 mL/min solvent flow rate, 40 °C column temperature, 20% solvent B in solvent A at 0 min, a gradient increasing to 100% solvent B over 0.1 min, 100% solvent B for 0.9 min, a gradient decreasing to 20% solvent B in solvent A for 0.1 min, 20% solvent B in solvent A for 0.5 min (solvent A = water with 0.1% v/v formic acid, solvent B = acetonitrile with 0.1% v/v formic acid), unless otherwise specified. The first 0.5 min and last 0.3 min of the LC method were diverted to waste, the subsequent eluate was analyzed by tandem mass spectrometry (Waters Xevo TQ-S) with electron spray ionization in positive mode (ESI+), 500 °C desolvation temperature, 1000 L/h desolvation gas, 150 L/h cone gas, 7 bar nebulizer gas, 2800 V voltage capillary. MS parameters for each compound are described below (Table 4.4).

Table 4.4. Tandem mass spectrometry parameters

Compound	Parent Ion $[M+H]^+$ (m/z)	Daughter Ion $[M+H]^+$ (m/z)	Dwell (s)	Cone (V)	Collision (V)
8	421.3676	265.2162	0.025	50	18
7	265.2162	111.0441	0.025	50	18
29a	281.2111	125.0597	0.025	50	18
18a	125.0597	110.0362	0.025	50	18
21a	369.2424	213.0910	0.025	50	18
benvitimod	255.1380	213.0910	0.025	50	18
20a	213.0910	111.0441	0.025	50	18

4.4.2 Cell culture and qRT-PCR detection of AhR activity

Caco-2 or HT29 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (ThermoFischer 11995073, 4.5 g/L glucose, 2 mM glutamine, 110 mg/L pyruvate) supplemented with 10% v/v heat inactivated fetal bovine serum (FBS), 100 IU/mL penicillin and 100 µg/mL streptomycin (Pen-Strep) (Corning 30-002-CI). Growth conditions of 37 °C and 5% CO_2 were maintained. Following thawing, cells were passaged at least 4 times before assays were conducted. To prepare cells for assaying, approximately 20,000 cells (in 98 µL volume) were added to each well of a tissue culture treated 96-well plate. Ligand stimulation was initiated by adding 2 µL of each compound or mixture in DMSO and gently mixing. Then following 1 or 4 h incubation under standard growth conditions, RNA was extracted (Dynabeads mRNA DIRECT, Invitrogen), and complementary DNA (cDNA) was generated (High-Capacity cDNA Reverse Transcription) by combining 10 µL template RNA (2-4 ng/µL) with 10 µL kit master mix (without RNAse) and incubating at 25 °C for 10 min, 37 °C for 120 min, and 85 °C for 5 min. Then qRT-PCR was conducted (QuantiTect SYBR Green PCR Kit, Qiagen) by combining 5 µL of kit master mix with 0.2 µL cDNA and 0.3 µM of either primer pair (Table 4.5), and with the following parameters: initial denaturation at 95 °C for 15 m, denaturation at 95 °C for 15 s, annealing at 60 °C for 30 s, and elongation at 72 °C for 30 s.

Table 4.5. Primers used to quantify AhR activity via CYP1A1 expression

Primer	Nucleotide Sequence (5′ to 3′)	Details
L13A-F	CCTGGAGGAGAAGAGGAAAGAGA	control gene
L13A-R	GAGGACCTCTGTGTATTTGTCAA	
CYP1A1-F	TACCTCAGCCACCTCCAAGAT	target
CYP1A1-R	GAGGTCTTGAGGCCCTGATT	

4.4.3 Luciferase AhR reporter assays

HT29-Lucia™ AhR cells (InvivoGen) were cultured in Dulbecco's Modified Eagle Medium (DMEM) (ThermoFischer 11995073, 4.5 g/L glucose, 2 mM glutamine, 110 mg/L pyruvate) supplemented with 10%

v/v heat inactivated fetal bovine serum (FBS), 100 IU/mL penicillin and 100 µg/mL streptomycin (Pen-Strep) (Corning 30-002-CI), 100 µg/mL Normocin (InvivoGen), and 100 µg/mL Zeocin (InvivoGen). Growth conditions of 37 °C and 5% CO_2 were maintained. Following thawing, cells were passaged at least 4 times before assays were conducted. To prepare cells for luciferase assay, growth media was exchanged with pyruvate-free DMEM (ThermoFischer 11960069, 4.5 g/L glucose) supplemented with 10% heat inactivated FBS and Pen-Strep, and approximately 50,000 cells (in 190 µL volume) were added to each well of a tissue culture treated 96-well plate. Ligand stimulation was initiated by adding 10 µL of each compound or mixture in DMSO and gently mixing. Following 14 h under standard growth conditions, luciferase activity was measured by collecting 20 µL of cell culture supernatant, combining with 50 µL of QUANTI-Luc™ (InvivoGen), and measuring luminescence on a BioTek plate reader (auto-gain, 0.1 integration time). Each condition was repeated in biological triplicate. For each sample, the luminescence was reported as fold change relative to vehicle (DMSO) negative controls.

4.4.4 Error prone PCR library construction

Plasmid encoding CylK-Strep-GFP11 (Table 4.10), derived from pPR-IBA1-CylK, was synthesized and sequence verified by Genewiz (South Plainfield, NJ, USA). From this plasmid, an error prone PCR (epPCR) library was constructed via PCR and restriction enzyme digestion and ligation. The CylK insert was prepared by initial PCR amplification using Primers 7 and 8, followed by epPCR amplification with Primers 7 and 8 (Table 4.6). Initial PCR reaction contained 5 µL template DNA (50 ng), 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. Initial PCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 65 °C for 15 s, and elongation at 72 °C for 45 s. Following PCR purification (Cytiva, 28903470), epPCR was conducted on a dilution of this amplicon. The epPCR reaction contained 100 µM $MnCl_2$, 5 µL template DNA (50 ng), 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. epPCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 40 cycles of denaturation at 98 °C for 10 s, annealing at 65 °C for 15 s, and elongation at 72 °C for 45 s. Backbone was prepared by PCR using Primers 9 and 10 (Table 4.6). Backbone PCR

reaction contained 5 µL template DNA (50 ng), 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. Backbone PCR was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 66 °C for 15 s, and elongation at 72 °C for 70 s. Then, following gel purification (Cytiva, 28903470) of the epPCR insert, or PCR purification of the backbone amplicon, both were digested with DpnI (NEB) for 30 min at 30 °C, and XbaI and NdeI (NEB) for 4 h at 37 °C. Digests contained 1–2 µg DNA, 3 µL 10X CutSmart Buffer (NEB), 3 µL DpnI, and water to 30 µL. Following 30 min at 30 °C, digests were supplemented with 1 µL of XbaI and 1 µL of NdeI, incubated for 4 h at 37 °C, then 80 °C for 20 min. Following digestion, backbone vector reactions were supplemented with 2 µL of Quick CIP (NEB) and incubated at 37 °C for 2 h. Finally, all reactions were PCR purified before performing the ligation reaction. Ligations were incubated at 16 °C for 16 h and contained 156 ng digested insert, 40 ng digested backbone, 1 µL T4 DNA ligase (NEB), 1 µL 10X T4 DNA Ligase Buffer (NEB), and water to 10 µL, performed in duplicate. 5 µL of each ligation mixture was transformed into chemically competent *E. coli* Top10 cells. Two plates of ~800 colonies were resuspended in 7 mL of LB medium supplemented with 100 µg/mL ampicillin or carbenicillin, incubated at 37 °C for 1 h, before extracting plasmid DNA (Omega Bio-Tek). Mutation rate was determined by Sanger sequencing (Genewiz) 20 individual colonies from this library.

Table 4.6. Primers used to clone epPCR library of CylK-Strep-GFP11

Primer	Nucleotide Sequence (5′ to 3′)	Amplicon
7	CACAACGGTTTCCCTCTAGAAATAATTTTG	XbaI-CylK-Strep-GFP11-NdeI
8	CATGCAGCACCATATGATCGC	
9	CGCGATCATATGGTGCTGCATG	NdeI-GFP11-pPR-IBA1-XbaI
10	CAAAATTATTTCTAGAGGGAAACCGTTGTG	

Restriction sites underlined.

4.4.5 Site saturation mutagenesis libraries construction

Plasmid encoding CylK-Strep-GFP11 (Table 4.10), derived from pPR-IBA1-CylK, was synthesized and sequence verified by Genewiz (South Plainfield, NJ, USA). From this plasmid, single site saturation mutagenesis (SSM) libraries were constructed via PCR and modified golden gate assembly[246] with a pooled primer approach to ensure uniform distribution of all 20 amino acids.[247,248]

For the Leu83 SSM library, the CylK insert was prepared as two fragments by PCR amplification using Primers 11 and 12, or 13 and 14. At Ile92, with Primers 11 and 15, or 16 and 14. At Asn334, with Primers 11 and 17, or 18 and 14. At Leu438, with Primers 11 and 19, or 20 and 14. Backbone vector fragment was prepared by PCR amplification using Primers 21 and 22.

PCR reactions contained 50 ng template DNA, 0.5 µM of each primer or total pooled primers, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. PCR of CylK insert fragments was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 15 s, and elongation at 72 °C for 30–60 s. PCR of backbone vector fragment was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 15 s, and elongation at 72 °C for 90 s. Golden gate reactions contained approximately 50 ng of each CylK insert fragment, 75 ng of backbone vector fragment, 2.0 µL Golden Gate Assembly Kit BsaI-HF (NEB), 2.0 µL 10X T4 DNA Ligase Buffer (NEB) and water to 20 µL (or negative controls without assembly mix). Golden gate reactions were carried out using the following parameters: 30 cycles of digestion at 37 °C for 1 min and ligation at 16 °C for 1 min followed by final heat inactivation at 60 °C for 5 min. Following digestion/ligation, reactions were supplemented with 0.2 µL of DpnI (NEB) and incubated at 37 °C for 2 h followed by final heat inactivation at 80 °C for 20 min. 5 µL of each ligation mixture or negative control was transformed into chemically competent *E. coli* Top10 cells. Negative control plates generally had ~100-fold fewer colonies than positive ligation plates, suggesting low background. Plates of colonies (200–800) were resuspended in 7 mL of LB medium supplemented with 100 µg/mL ampicillin or carbenicillin, incubated at 37 °C for 1 h, before

extracting plasmid DNA (Omega Bio-Tek). Construction and degeneracy at codon of interest were verified by submitting an aliquot of plasmid library for Sanger sequencing (Genewiz/Azenta).

Table 4.7. Primers used to clone SSM libraries of CylK-Strep-GFP11

Primer	Nucleotide Sequence (5′ to 3′)	Amplicon
11	TTGGTCTCTACAAATGAAAAAAAACAAAAAAACAACAAAA	CylK(1-Leu83)
12	TTGGTCTCTACTGT**AHN**TACATTACTATACTTGGCAAAAATG (12) TTGGTCTCTACTGT**TKB**TACATTACTATACTTGGCAAAAATG (6) TTGGTCTCTACTGT**CAT**TACATTACTATACTTGGCAAAAATG (1) TTGGTCTCTACTGT**CCA**TACATTACTATACTTGGCAAAAATG (1)	
13	TTGGTCTCTCAGTTTCCAGCGATCCTAATTTCATAACATC	CylK(84-STOP)
14	TTGGTCTCTTTAGGTAATGCCCGCCGCG	
15	TTGGTCTCTGATGT**AHN**GAAATTAGGATCGCTGGAAACTG (12) TTGGTCTCTGATGT**TKB**GAAATTAGGATCGCTGGAAACTG (6) TTGGTCTCTGATGT**CAT**GAAATTAGGATCGCTGGAAACTG (1) TTGGTCTCTGATGT**CCA**GAAATTAGGATCGCTGGAAACTG (1)	CylK(1-Ile92) *with primer 11*
16	TTGGTCTCTCATCAGAAGATGAAGAGGTCAAAATCATTT	CylK(93-STOP) *with primer 14*
17	TTGGTCTCTGTGGT**AHN**TCCCGCAACGTAAATATTGCTCC (12) TTGGTCTCTGTGGT**TKB**TCCCGCAACGTAAATATTGCTCC (6) TTGGTCTCTGTGGT**CAT**TCCCGCAACGTAAATATTGCTCC (1) TTGGTCTCTGTGGT**CCA**TCCCGCAACGTAAATATTGCTCC (1)	CylK(1-Asn334) *with primer 11*
18	TTGGTCTCTCCACAGGCCAGCTAGAAAATAACACCGTT	CylK(335-STOP) *with primer 14*
19	TTGGTCTCTTCGTC**AHN**TGTGATCGTACCAAACTGTG (12) TTGGTCTCTTCGTC**TKB**TGTGATCGTACCAAACTGTG (6) TTGGTCTCTTCGTC**CAT**TGTGATCGTACCAAACTGTG (1) TTGGTCTCTTCGTC**CCA**TGTGATCGTACCAAACTGTG (1)	CylK(1-Leu438) *with primer 11*
20	TTGGTCTCTACGACAATTGGGGTGTCGCAGC	CylK(439-STOP) *with primer 14*
21	TTGGTCTCTTTGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC	pPR-IBA1 backbone
22	TTGGTCTCTCTAATAAGCTTGATCCGGCTGCTAAC	

Restriction sites and downstream cut sites underlined. Mutagenized codon in bold. Values in parenthesis indicate pooled primer ratios.

4.4.6 Cloning CylK-Strep-GFP constructs

Plasmid encoding CylK-Strep-GFP11, derived from pPR-IBA1-CylK, was synthesized and sequence verified by Genewiz (South Plainfield, NJ, USA). From this plasmid, or from variants derived from single site saturation mutagenesis (SSM) libraries (N334-CylK-Strep-GFP11 & ΔN25-CylK-Strep-GFP11), CylK-Strep-GFP constructs were assembled via PCR amplification and golden gate assembly. Plasmid encoding GFP1-10, pBAD-6xHis-GFP1-10, was synthesized and sequence verified by Genewiz (South Plainfield, NJ, USA). All insert DNA sequences can be found below (Table 4.10).

PCR reactions contained approximately 50 ng template DNA, 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. PCR of CylK+backbone fragment was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 67 °C for 30 s, and elongation at 72 °C for 5 min. PCR of GFP1-10 insert was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 64 °C for 30 s, and elongation at 72 °C for 30 s. 8 µL of each PCR reaction was combined with 1 µL of DpnI (NEB), and 1 µL 10X CutSmart Buffer (NEB), and incubated at 37 °C for 2 h, followed by heat inactivation at 80 °C for 20 min. Following digestion, all reactions were PCR purified (Cytiva, 28903470) before assembly. Golden gate reactions contained approximately 75 ng of each CylK+backbone fragment, 25 ng of GFP1-10 insert, 1.0 µL Golden Gate Assembly Kit BsaI-HF (NEB), 2.0 µL 10X T4 DNA Ligase Buffer (NEB) and water to 20 µL. Golden Gate reactions were carried out using the following parameters: 30 cycles of digestion at 37 °C for 1 min and ligation at 16 °C for 1 min followed by final heat inactivation at 60 °C for 5 min. 5 µL of each ligation mixture was transformed into chemically competent *E. coli* Top10 cells. Constructs were verified by Sanger (Genewiz/Azenta) and next-generation sequencing (Massachusetts General Hospital Center for Computational & Integrative Biology DNA Core).

Table 4.8. Primers used to clone CylK-Strep-GFP

Primer	Nucleotide Sequence (5′ to 3′)	Amplicon
23	TTGGTCTCTGCGTGAGCAAGGGCGAGGAG	GFP1-10
24	TTGGTCTCTCGCTTCTCGTTGGGGTCTTTGC	
25	TTGGTCTCTACGCCGCCTTTTTCGAACTGCG	GFP11-backbone-CylK-Strep
26	TTGGTCTCTAGCGCGATCATATGGTGCTGCATG	

Restriction sites and downstream cut sites underlined.

4.4.7 Cloning CylK-Strep-GFP-6xHis constructs

Plasmids encoding CylK-Strep-GFP were used as template DNA (see Materials and Methods 4.4.6), and from these, constructs containing C-terminal 6xHis fusions were assembled via PCR amplification and Gibson Assembly. DNA encoding His tag and linker was introduced via primers. PCR reactions contained approximately 50 ng template DNA, 0.5 µM of each primer, 25 µL of Phusion High-Fidelity PCR Master Mix (NEB), and water to 50 µL. PCR of CylK+backbone fragment was carried out using the following parameters: denaturation for 30 s at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 66 °C for 30 s, and elongation at 72 °C for 2.5 min. 8 µL of each PCR reaction was combined with 1 µL of DpnI (NEB), and 1 µL 10X CutSmart Buffer (NEB), and incubated at 37 °C for 3 h, followed by heat inactivation at 80 °C for 20 min. Following digestion, all reactions were PCR purified (Cytiva, 28903470) before assembly. Gibson assembly reactions were incubated at 50 °C for 45 min and contained 5 µL CylK+backbone fragment, 5 µL Gibson Assembly Master Mix (NEB). 5 µL of each reaction was transformed into chemically competent *E. coli* Top10 cells. Constructs were verified by Sanger (Azenta) and next-generation sequencing (Mass General Hospital Center for Computational & Integrative Biology DNA Core).

Table 4.9. Primers used to clone CylK-Strep-GFP-6xHis

Primer	Nucleotide Sequence (5′ to 3′)	Amplicon
27	TCTCACCATCACCATCACCATTAATAAGCTTGATCCGGCTGCTAAC	CylK-Strep-GFP11-6xHis
28	ATGGTGATGGTGAGAAGAACCGGTAATGCCCGCCGCGTTC	

Site of homology and linker/6xHis addition are underlined.

Table 4.10. DNA sequences of unpublished CylK constructs

Construct	Nucleotide Sequence
CylK-Strep-GFP11	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCCGCGATCATATGGTGCTGCATGAATATGTGAACGCGGCGGGC ATTACCTAA
N334-CylK-Strep-GFP11 (codon change)	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAATACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCCGCGATCATATGGTGCTGCATGAATATGTGAACGCGGCGGGC ATTACCTAA
ΔN25-CylK-Strep-GFP11	ATGAAAAAAAACAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTGA GTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAATC TCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAGT ATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGCC GTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTCG GTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATTC TGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGACT TTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACAG

Table 4.10. (Continued)

	CAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGTG ACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAATG ATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGAC GCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGCT CAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAGC TAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATTA AACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGTC ACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGCT ACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGGG GTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGGA CAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACTC CGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGACT TGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAAC AGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACGG GAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGCA ATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAAC TTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTAG ACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGGA GCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCCGCGATCATATGGTGCTGCATGAATATGTGAACGCGGCGGGCA TTACCTAA
CylK-Strep-GFP	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGC AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC GTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACC CTGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAAC GTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAA TATGTGAACGCGGCGGGCATTACCTAA
N334-CylK-Strep-GFP (codon change)	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT

Table 4.10. (Continued)

	GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAATACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGC AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC GTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACC CTGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAAC GTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAA TATGTGAACGCGGCGGGCATTACCTAA
ΔN25-CylK-Strep-GFP	ATGAAAAAAAACAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTGA GTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAATC TCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAGT ATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGCC GTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTCG GTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATTC TGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGACT TTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACAG CAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGTG ACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAATG ATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGAC GCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGCT CAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAGC TAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATTA AACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGTC ACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGCT ACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGGG GTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGGA CAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACTC CGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGACT TGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAAC AGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACGG GAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGCA ATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAAC TTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTAG ACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGGA GCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCC TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGCA AGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG TCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACCC TGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT TCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAACG TCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG ACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAAT ATGTGAACGCGGCGGGCATTACCTAA
6xHis-GFP1-10	ATGGGTTCTTCTCACCATCACCATCACCATGGTTCTTCTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCC ATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATT GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACC TACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC

Table 4.10. (Continued)

	TACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGAC ACCCTGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC AACTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCAC AACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG CCCGACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGGGCACCGGTGGAGGCGGTAGC TAG
CylK-Strep-GFP-6xHis	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGC AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC GTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACC CTGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAAC GTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAA TATGTGAACGCGGCGGGCATTACCGGTTCTTCTCACCATCACCATCACCATTAA
N334-CylK-Strep-GFP-6xHis (codon change)	ATGAAAAAAAACAAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTG AGTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAAT CTCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAG TATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGC CGTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTC GGTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATT CTGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGAC TTTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACA GCAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGT GACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAAT GATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGA CGCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGC TCAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAATACCACAGGCCAG CTAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATT AAACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGT CACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGC TACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGG GGTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGG ACAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACT CCGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGAC TTGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAA CAGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACG GGAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGC AATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAA

Table 4.10. (Continued)

	CTTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTA GACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGG AGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGC AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC GTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACC CTGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAAC GTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAA TATGTGAACGCGGCGGGCATTACCGGTTCTTCTCACCATCACCATCACCATTAA
ΔN25-CylK-Strep-GFP-6xHis	ATGAAAAAAAACAAAAAACAACAAAATCCTTACTCTCAGCCGACGAAAAAATAACAGAGTCCTTACGCTCAACATTGA GTGACGTTCTTCCAGACCAGCTTCAAACATATATCAGAACTGTCTTACAATTCTCTGGTAGACCGGAGGGGGCCAATC TCCTAACAGGCCCCAATACTGAGATAGAATTTTTCAGTCAAGATCCCAATAAGAATTTTCCCAACATTTTTGCCAAGT ATAGTAATGTACTGACAGTTTCCAGCGATCCTAATTTCATAACATCAGAAGATGAAGAGGTCAAAATCATTTGGGGCC GTCATGGAAGTGATAGCCTAATTGGTTTTGACCCTGGTGCTGACCTTGTTGGTAAACGCAGAATTGATATTTTCCTCG GTGATTTTATCGATGAACAATTCAATCCAATACCTGGCGCATTGAACGCAGGTAAATCGTGGTCAGACAGATTTATTC TGGGTGATTGGCAAAAACCCTACTATTTTGAGGATGACGAAACCTTGGGTTTAAACCAATCAGCTATGATTTTAGACT TTAACCCCAATGAAGATGTCATTCAGCTCCACGGCGATCGCCAAGACTATGAACTGGTAAATATTAGCTTGGGAACAG CAATATTTTGGCGGGAGAAAAAAGGTTATGACCTCATCGGCGTTTTAGGTGGAGTCTCTGATTTAAGTTTAAAGGGTG ACTACTTCGAGTTCAAGGGCAACACCGCACCAAAAACAGTCCTAAAGACGGCTGAACATATTGGGACTGCCGCGAATG ATTACATATTTAGTTCCACCGTAGATGCTAAGGGCAATTTCTATGTTGGAGGAGGAACCGGGGGATCTTTAGGAGGAC GCAACATAGGAGCTAGGGACGCTTGGCTGGCCAAATACGATAGCAATGGCAACCAGCGATGGAGTAGACAATTTGGCT CAACCGGCACTGAAAGCTTATGGGGTATGGCTAGTGACGGGAGCAATATTTACGTTGCGGGAAACACCACAGGCCAGC TAGAAAATAACACCGTTAAAGGCGGCAACGATGCTTATTTAGCTAAATACGATAGTGATGGCAACCAAGTGTGGATTA AACAGAATGGGACTTACACACTTGAGGAATCTTATAAAATTACCGTCGATTCCAGTGGCAATATTTACACTGCTGGTC ACACTTTTGGCAGCTTGGGTGGGCCAAACCAAAACTTGGAACAAGGCGAAGTATTCGAGTTACCATCGACTGATGGCT ACGTGGCCAAGTTCGACAGCAATGGTAATCAGTTGTGGGTTGCACAGTTTGGTACGATCACACTTGACGACAATTGGG GTGTCGCAGCCGACAACAATGGCAATGTCTTCGCTGGAGGAAATACAAAAGGCAGTTTCGGAGCAAAGAATACTGGGA CAGCGGGAGAATATGATGCCTGGCTAGTTAAGTTGAACAAAGATGGTCAGACTGATTGGGTTAGACAGTTTGGCACTC CGAATTATGACTTTATGTGGGACATTGAAACGGATAGTCTAGGCGATATCTATGCGACCGGATGGACTCTGGGAGACT TGGGGGGAAAAAATGCCGGCTCTTATGATGTCTGGCTGGCCAAGTACAACACTAATGGCAACCAGTTGTGGATCAAAC AGTTTGGCACTAGTGAAGATGATGCACCCTTTTTAGATGGGATAGATATAGACGCAAATGACAATATTTTCCTGACGG GAAATACAAACGGCAACTTAGGCGGAGCTAATGCTGGATCTTATGATGCTTGGGCGGCGAAGTTTGACAAGGATGGCA ATCAGTTATGGCTCAAACAGTTTGGTACCCCCGACTATGACACTGCTACCACGGTAACTGCCGTTAATTTCGGTAAAC TTTATGTTTCGGGAATTACTGAGGGAAGTTTGGGAACTACCAATGCTGGGTCTTATGACTCTTGGGCGCTCAAACTAG ACGCTGATAACGGAGAGATACAAGATTTTAATTCTTCTACTAATACTTTTGGCCAAACGGGTTTTTTGAATTTAGGGA GCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGGCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCC TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCATTGGCA AGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG TCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTGGTGAAGTTCGAGGGCGACACCC TGGTGAACCGCATCGAGCTGAAGGGCACCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT TCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCACCGTGCGCCACAACG TCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG ACAACCACTACCTGAGCACCCAGACCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTGCTGCATGAAT ATGTGAACGCGGCGGGCATTACCGGTTCTTCTCACCATCACCATCACCATTAA

In all CylK constructs, underlined sites indicate deletion (AAC>AC), alternate start codon (AAT), and residue N334 (AAC or AAT).

4.4.8 Plate-based CylK growth, expression and activity assay

CylK libraries or individual variants were expressed and assayed for activity according to a modified method (Materials and Methods 3.4.7). Respective CylK plasmids or plasmid libraries were freshly transformed into electrocompetent *E. coli* BL21 Gold CodonPlus® (DE3) RIL cells (Agilent), and an individual colony was

used to directly inoculate 1 mL cultures of LB medium supplemented with 100 µg/mL ampicillin or carbenicillin and 34 µg/mL chloramphenicol in a 96 deep well plate (2 mL capacity, VWR), sterility was maintained with standard techniques and a gas permeable rayon film (VWR). The liquid culture plate was incubated for 5 h at 37 °C and 250 rpm shaking. After five h, and confirming visible growth in each well, the liquid culture plate was cooled to 15 °C, maintaining shaking. Following an additional 30-45 min of incubation, protein expression was induced with 250 µM IPTG. The cultures were incubated for 4 h at 15 °C with 250 rpm shaking. Following expression, reactions were initiated in opaque 96-well plates (Costar) by combining 100 µL of concentrated cell suspension with 2 µL of lysozyme mix (9 mg lysozyme & 1 mg EDTA-free Pierce Protease Inhibitor Tablet per 500 µL of 500 mM Tris, 50 mM EDTA, pH 8.0) and mono-alkylation substrates in DMSO for a final concentration of nucleophile at 200 µM, electrophile 200 µM, and DMSO at 2.5%. OD_{600} and GFP measurements were taken by plate reader (BioTek) by placing 200 µL of expression cultures in opaque 96-well plates with transparent bottoms (Corning). Reaction plates were sealed with aluminum seals (VWR) and incubated at 37 °C with 190 rpm shaking for 20-22 h. For luciferase-based detection, reactions were diluted 10-fold, and then 10 µL of diluted reactions were combined with 190 µL of HT29-Lucia™ AhR cells (Materials and Methods 4.4.3). For LC-MS detection, reactions were quenched with the addition of 1:1 methanol/acetonitrile (200 µL) and centrifuged (3,220 x *g* for 15 min). The supernatant was then subjected to LC-MS to measure ion abundance of product and nucleophile starting material as a loading control. LC-MS conditions as described previously (Materials and Methods 4.4.1). Select expression cultures were subcultured into LB medium supplemented with 100 µg/mL ampicillin for plasmid preparation (Omega D6943-00S).

5 References

(1) Cragg, G. M.; Newman, D. J.; Snader, K. M. Natural Products in Drug Discovery and Development. *J Nat Prod* **1997,** *60* (1), 52.

(2) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **2020,** *83* (3), 770.

(3) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010,** *329* (5989), 305.

(4) Ro, D. K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440* (7086), 940.

(5) Lee, S. Y.; Kim, H. U.; Chae, T. U.; Cho, J. S.; Kim, J. W.; Shin, J. H.; Kim, D. I.; Ko, Y.-S.; Jang, W. D.; Jang, Y.-S. A comprehensive metabolic map for production of bio-based chemicals. *Nat Catal* **2019,** *2* (1), 18.

(6) Firn, R. D.; Jones, C. G. Natural products – a simple model to explain chemical diversity. *Nat Prod Rep* **2003,** *20* (4), 382.

(7) Fischbach, M. A.; Clardy, J. One pathway, many products. *Nat Chem Biol* **2007,** *3* (7), 353.

(8) MacMillan, J. Biosynthesis of the gibberellin plant hormones. *Nat Prod Rep* **1997,** *14* (3), 221.

(9) Lombardino, J. G.; Lowe, J. A. The role of the medicinal chemist in drug discovery — then and now. *Nat Rev Drug Discov* **2004,** *3* (10), 853.

(10) Crosa, J. H.; Walsh, C. T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **2002,** *66* (2), 223.

(11) Liao, L.; Schaefer Amy, L.; Coutinho Bruna, G.; Brown Pamela, J. B.; Greenberg, E. P. An aryl-homoserine lactone quorum-sensing signal produced by a dimorphic prosthecate bacterium. *P Natl Acad Sci USA* **2018,** *115* (29), 7587.

(12) Mukherjee, S.; Bassler, B. L. Bacterial quorum sensing in complex and dynamically changing environments. *Nat Rev Microbiol* **2019,** *17* (6), 371.

(13) Ogran, A.; Yardeni Eliane, H.; Keren-Paz, A.; Bucher, T.; Jain, R.; Gilhar, O.; Kolodkin-Gal, I.; Parales Rebecca, E. The plant host induces antibiotic production to select the most-beneficial colonizers. *Appl Environ Microbiol* **2019,** *85* (13), e00512.

(14) Butcher Rebecca, A.; Schroeder Frank, C.; Fischbach Michael, A.; Straight Paul, D.; Kolter, R.; Walsh Christopher, T.; Clardy, J. The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. *P Natl Acad Sci USA* **2007,** *104* (5), 1506.

(15) Solbiati, J. O.; Ciaccio, M.; Farías, R. N.; Salomón, R. A. Genetic analysis of plasmid determinants for microcin J25 production and immunity. *J Bacteriol* **1996,** *178* (12), 3661.

(16) Price, M. N.; Wetmore, K. M.; Deutschbauer, A. M.; Arkin, A. P. A comparison of the costs and benefits of bacterial gene expression. *PLoS One* **2016,** *11* (10), e0164314.

(17) Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M. Enzyme evolution: innovation is easy, optimization is complicated. *Curr Opin Struc Biol* **2018,** *48*, 110.

(18) Bornscheuer, U. T.; Kazlauskas, R. J. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angew Chem Int Ed Engl* **2004,** *43* (45), 6032.

(19) Crawshaw, R.; Crossley, A. E.; Johannissen, L.; Burke, A. J.; Hay, S.; Levy, C.; Baker, D.; Lovelock, S. L.; Green, A. P. Engineering an efficient and enantioselective enzyme for the Morita–Baylis–Hillman reaction. *Nature Chemistry* **2022,** *14* (3), 313.

(20) Demay, J.; Bernard, C.; Reinhardt, A.; Marie, B. Natural products from cyanobacteria: focus on beneficial activities. *Mar Drugs* **2019,** *17* (6), 320.

(21) Pham, J. V.; Yilma, M. A.; Feliz, A.; Majid, M. T.; Maffetone, N.; Walker, J. R.; Kim, E.; Cho, H. J.; Reynolds, J. M.; Song, M. C.et al. A review of the microbial production of bioactive natural products and biologics. *Front Microbiol* **2019,** *10*, 1404.

(22) Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012,** *2* (2), 303.

(23) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Cell Chem Biol* **2004,** *11* (6), 817.

(24) Balskus, E. P.; Case, R. J.; Walsh, C. T. The biosynthesis of cyanobacterial sunscreen scytonemin in intertidal microbial mat communities. *FEMS Microbiol Ecol* **2011,** *77* (2), 322.

(25) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* **2014,** *53* (18), 2875.

(26) Smith, S.; Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **2007,** *24* (5), 1041.

(27) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* **2015,** *11* (9), 625.

(28) Koonin, E. V.; Makarova, K. S.; Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **2001,** *55* (1), 709.

(29) Thomas, C. M.; Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **2005,** *3* (9), 711.

(30) Egan, S.; Wiener, P.; Kallifidas, D.; Wellington, E. M. H. Phylogeny of Streptomyces species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Anton Leeuw* **2001,** *79* (2), 127.

(31) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A natural product chemist's guide to unlocking silent biosynthetic gene clusters. *Annu Rev Biochem* **2021,** *90* (1), 763.

(32) Rudolph, A.; Lautens, M. Secondary Alkyl Halides in Transition-Metal-Catalyzed Cross-Coupling Reactions. *Angew Chem Int Ed Engl* **2009,** *48*, 2656.

(33) Miyaura, N.; Suzuki, A. Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem Rev* **1995,** *95* (7), 2457.

(34) Adak, S.; Moore, B. S. Cryptic halogenation reactions in natural product biosynthesis. *Nat Prod Rep* **2021,** *38*, 1760.

(35) Yamanaka, K.; Ryan, K. S.; Gulder, T. A. M.; Hughes, C. C.; Moore, B. S. Flavoenzyme-catalyzed atropo-selective N,C-bipyrrole homocoupling in marinopyrrole biosynthesis. *J Am Chem Soc* **2012,** *134* (30), 12434.

(36) Marchand, J. A.; Neugebauer, M. E.; Ing, M. C.; Lin, C. I.; Pelton, J. G.; Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **2019,** *567* (7748), 420.

(37) Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W. H. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* **2004,** *67* (8), 1356.

(38) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **2005,** *436*, 1191.

(39) Kelly, W. L.; Boyne, M. T.; Yeh, E.; Vosburg, D. A.; Galonić, D. P.; Kelleher, N. L.; Walsh, C. T. Characterization of the Aminocarboxycyclopropane-Forming Enzyme CmaC. *Biochemistry* **2007,** *46* (2), 359.

(40) Ma, S.; Mandalapu, D.; Wang, S.; Zhang, Q. Biosynthesis of cyclopropane in natural products. *Natural Product Reports* **2022**, DOI:10.1039/D1NP00065A 10.1039/D1NP00065A.

(41) White, M. F.; Vasquez, J.; Yang, S. F.; Kirsch, J. F. Expression of apple 1-aminocyclopropane-1-carboxylate synthase in Escherichia coli: kinetic characterization of wild-type and active-site mutant forms. *P Natl Acad Sci USA* **1994,** *91* (26), 12428.

(42) Zha, L.; Jiang, Y.; Henke, M. T.; Wilson, M. R.; Wang, J. X.; Kelleher, N. L.; Balskus, E. P. Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. *Nat Chem Biol* **2017,** *13* (10), 1063.

(43) Neumann, C. S.; Walsh, C. T. Biosynthesis of (–)-(1S,2R)-allocoronamic acyl thioester by an FeII-dependent halogenase and a cyclopropane-forming flavoprotein. *J Am Chem Soc* **2008,** *130* (43), 14022.

(44) Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L.; Sherman, D. H.; Gerwick, W. H.; Håkansson, K.; Smith, J. L. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase. *Structure* **2015,** *23* (12), 2213.

(45) Schnepel, C.; Sewald, N. Enzymatic Halogenation: A Timely Strategy for Regioselective C-H Activation. *Chem Eur J* **2017,** *23*, 12064.

(46) Neumann, C. S.; Fujimori, D. G.; Walsh, C. T. Halogenation Strategies In Natural Product Biosynthesis. *Chem Biol* **2008,** *15* (2), 99.

(47) Winter, J. M.; Moore, B. S. Exploring the chemistry and biology of vanadium-dependent haloperoxidases. *J Biol Chem* **2009,** *284* (28), 18577.

(48) Yeh, E.; Cole, L. J.; Barr, E. W.; Bollinger, J. M., Jr.; Ballou, D. P.; Walsh, C. T. Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry* **2006,** *45* (25), 7904.

(49) Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J. B.; O'Hagan, D.; Naismith, J. H. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **2004,** *427*, 561.

(50) Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006,** *440* (7082), 368.

(51) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E.; Brinen, L. S.; Kato, Y.; Clardy, J. [7.7]Paracyclophanes from blue-green algae. *J Am Chem Soc* **1990,** *112* (10), 4061.

(52) Kang, H.-S.; Santarsiero, B. D.; Kim, H.; Krunic, A.; Shen, Q.; Swanson, S. M.; Chai, H.; Kinghorn, A. D.; Orjala, J. Merocyclophanes A and B, antiproliferative cyclophanes from the cultured terrestrial Cyanobacterium Nostoc sp. *Phytochemistry* **2012,** *79*, 109.

(53) Moore, B. S.; Chen, J.-L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyphanes a-f. *Tetrahedron* **1992,** *48* (15), 3001.

(54) Johansen, J. R.; Bohunická, M.; Lukešová, A.; Hrčková, K.; Vaccarino, M. A.; Chesarino, N. M. Morphological and molecular characterization within 26 strains of the genus *Cylindrospermum* (Nostocaceae, Cyanobacteria), with descriptions of three new species. *J Phycol* **2014,** *50* (1), 187.

(55) Bornet, E. *Revision des nostocacées hétérocystées contenues dans les principaux herbiers de France, par Éd. Bornet et Ch. Flahault*: France, 1888.

(56) Cram, D. J.; Steinberg, H. Macro Rings. I. Preparation and Spectra of the Paracyclophanes. *J Am Chem Soc* **1951,** *73* (12), 5691.

(57) Dale, J. Conformational Aspects of Many-Membered Rings. *Angew Chem Int Ed Engl* **1966,** *5* (12), 1000.

(58) Odashima, K.; Itai, A.; Iitaka, Y.; Koga, K. Host-guest complex formation between a water-soluble polyparacyclophane and a hydrophobic guest molecule. *J Am Chem Soc* **1980,** *102* (7), 2504.

(59) Takahashi, I.; Odashima, K.; Koga, K. Diastereomeric host-guest complex formation by an optically active paracyclophane in water. *Tetrahedron Lett* **1984,** *25* (9), 973.

(60) Chen, J. L.; Moore, R. E.; Patterson, G. M. L. Structures of nostocyclophanes A-D. *J Org Chem* **1991,** *56* (14), 4360.

(61) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the Cyanobacterium Nostoc sp. *J Nat Prod* **2010,** *73* (9), 1529.

(62) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. Carbamidocyclophanes A–E, Chlorinated Paracyclophanes with Cytotoxic and Antibiotic Activity from the Vietnamese Cyanobacterium Nostoc sp. *J Nat Prod* **2007,** *70* (4), 499.

(63) Luo, S.; Kang, H. S.; Krunic, A.; Chlipala, G. E.; Cai, G.; Chen, W. L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. Carbamidocyclophanes F and G with anti-Mycobacterium tuberculosis activity from the cultured freshwater cyanobacterium Nostoc sp. *Tetrahedron Lett* **2014,** *55*, 686.

(64) Preisitsch, M.; Harmrolfs, K.; Pham, H. T. L.; Heiden, S. E.; Füssel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H. J.; Müller, R.et al. Anti-MRSA-acting carbamidocyclophanes H–L from the Vietnamese cyanobacterium Nostoc sp. CAVN2. *J Antibiot* **2015,** *68*, 165.

(65) May, D. S.; Chen, W. L.; Lantvit, D. D.; Zhang, X.; Krunic, A.; Burdette, J. E.; Eustaquio, A.; Orjala, J. Merocyclophanes C and D from the Cultured Freshwater Cyanobacterium Nostoc sp. (UIC 10110). *J Nat Prod* **2017,** *80*, 1073.

(66) May, D. S.; Kang, H. S.; Santarsiero, B. D.; Krunic, A.; Shen, Q.; Burdette, J. E.; Swanson, S. M.; Orjala, J. Ribocyclophanes A–E, Glycosylated Cyclophanes with Antiproliferative Activity from Two Cultured Terrestrial Cyanobacteria. *J Nat Prod* **2018,** *81*, 572.

(67) Yamakoshi, H.; Ikarashi, F.; Minami, M.; Shibuya, M.; Sugahara, T.; Kanoh, N.; Ohori, H.; Shibata, H.; Iwabuchi, Y. Syntheses of naturally occurring cytotoxic [7.7]paracyclophanes, (–)-cylindrocyclophane A and its enantiomer, and implications for biological activity. *Org Biomol Chem* **2009,** *7* (18), 3772.

(68) Peterson, E.; Kaur, P. Antibiotic resistance mechanisms in bacteria: relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens. *Front Microbiol* **2018,** *9*.

(69) Nakamura, H.; Schultz, E. E.; Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat Chem Biol* **2017,** *13* (8), 916.

(70) Bobzin, S. C.; Moore, R. E. Biosynthetic origin of [7.7]paracyclophanes from cyanobacteria. *Tetrahedron* **1993,** *49* (35), 7615.

(71) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane Biosynthesis Involves Functionalization of an Unactivated Carbon Center. *J Am Chem Soc* **2012,** *134* (45), 18518.

(72) Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat Prod Rep* **2008,** *25* (5), 845.

(73) Choi, Y. S.; Zhang, H.; Brunzelle, J. S.; Nair, S. K.; Zhao, H. In vitro reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis. *P Natl Acad Sci USA* **2008,** *105* (19), 6858.

(74) Zocher, G.; Winkler, R.; Hertweck, C.; Schulz, G. E. Structure and action of the N-oxygenase AurF from Streptomyces thioluteus. *J Mol Biol* **2007,** *373* (1), 65.

(75) Krebs, C.; Bollinger, J. M.; Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like 'di-iron-carboxylate' proteins. *Curr Opin Chem Biol* **2011,** *15* (2), 291.

(76) Fox, B. G.; Lyle, K. S.; Rogge, C. E. Reactions of the diiron enzyme stearoyl-acyl carrier protein desaturase. *Acc Chem Res* **2004,** *37* (7), 421.

(77) Merkx, M.; Kopp, D. A.; Sazinsky, M. H.; Blazyk, J. L.; Müller, J.; Lippard, S. J. Dioxygen Activation and Methane Hydroxylation by Soluble Methane Monooxygenase: A Tale of Two Irons and Three Proteins. *Angew Chem Int Ed Engl* **2001,** *40* (15), 2782.

(78) Ishida, K.; Matsuda, H.; Murakami, M. Four new microginins, linear peptides from the cyanobacterium Microcystis aeruginosa. *Tetrahedron* **1998,** *54* (44), 13475.

(79) Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.et al. Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J Nat Prod* **2015,** *78* (7), 1671.

(80) Lopez, J. A. V.; Petitbois, J. G.; Vairappan, C. S.; Umezawa, T.; Matsuda, F.; Okino, T. Columbamides D and E: chlorinated fatty acid amides from the marine cyanobacterium moorea bouillonii collected in malaysia. *Org Lett* **2017,** *19* (16), 4231.

(81) Leão, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P. Biosynthesis-assisted structural elucidation of the bartolosides, chlorinated aromatic glycolipids from cyanobacteria. *Angew Chem Int Edit* **2015,** *54* (38), 11063.

(82) Nakamura, H.; Wang, J. X.; Balskus, E. P. Assembly line termination in cylindrocyclophane biosynthesis: discovery of an editing type II thioesterase domain in a type I polyketide synthase. *Chem Sci* **2015,** *6* (7), 3816.

(83) Preisitsch, M.; Niedermeyer, T. H. J.; Heiden, S. E.; Neidhardt, I.; Kumpfmüller, J.; Wurster, M.; Harmrolfs, K.; Wiesner, C.; Enke, H.; Müller, R.et al. Cylindrofridins A–C, Linear Cylindrocyclophane-Related Alkylresorcinols from the Cyanobacterium Cylindrospermum stagnale. *J Nat Prod* **2015,** *79*, 106.

(84) Friedel, C.; Crafts, J. M. Sur une nouvelle méthode générale de synthèse d'hydrocarbures, d'acétones, etc. *Comptes rendus des séances* **1877,** *84*, 1392.

(85) Roberts, R. M.; Khalaf, A. A. *Friedel–Crafts Alkylation Chemistry: A Century of Discovery*; Marcel Dekker, Inc.: New York, 1984.

(86) Nakamura, H., Harvard University, 2016.

(87) Heravi, M. M.; Zadsirjan, V.; Saedi, P.; Momeni, T. Applications of Friedel–Crafts reactions in total synthesis of natural products. *RSC Adv* **2018,** *8* (70), 40061.

(88) Rueping, M.; Nachtsheim, B. J. A review of new developments in the Friedel–Crafts alkylation – From green chemistry to asymmetric catalysis. *Beilstein J. Org. Chem.* **2010,** *6* (6).

(89) Pallikonda, G.; Chakravarty, M. Benzylic phosphates in Friedel–Crafts reactions with activated and unactivated arenes: access to polyarylated alkanes. *J Org Chem* **2016,** *81* (5), 2135.

(90) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem Rev* **2017,** *117* (17), 11570.

(91) Watson, R. B.; Schindler, C. S. Iron-catalyzed synthesis of tetrahydronaphthalenes via 3,4-dihydro-2H-pyran intermediates. *Org Lett* **2018,** *20* (1), 68.

(92) Zetzsche, L. E.; Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nat Rev Chem* **2020,** *4* (7), 334.

(93) Pacholec, M.; Tao, J.; Walsh, C. T. CouO and NovO: C-methyltransferases for tailoring the aminocoumarin scaffold in coumermycin and novobiocin antibiotic biosynthesis. *Biochemistry* **2005,** *44* (45), 14969.

(94) Sadler, J. C.; Chung, C.-w. H.; Mosley, J. E.; Burley, G. A.; Humphreys, L. D. Structural and functional basis of C-methylation of coumarin scaffolds by NovO. *ACS Chem Biol* **2017,** *12* (2), 374.

(95) Stecher, H.; Tengg, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M. Biocatalytic Friedel-Crafts alkylation using non-natural cofactors. *Angew Chem Int Ed Engl* **2009,** *48* (50), 9546.

(96) Tengg, M.; Stecher, H.; Offner, L.; Plasch, K.; Anderl, F.; Weber, H.; Schwab, H.; Gruber-Khadjawi, M. Methyltransferases: green catalysts for Friedel-Crafts alkylations. *ChemCatChem* **2016,** *8* (7), 1354.

(97) Tanner, M. E. Mechanistic studies on the indole prenyltransferases. *Nat Prod Rep* **2015,** *32* (1), 88.

(98) Gebler, J. C.; Poulter, C. D. Purification and characterization of dimethylallyl tryptophan synthase from Claviceps purpurea. *Arch Biochem Biophys* **1992,** *296* (1), 308.

(99) Metzger, U.; Schall, C.; Zocher, G.; Unsöld, I.; Stec, E.; Li, S.-M.; Heide, L.; Stehle, T. The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *P Natl Acad Sci USA* **2009,** *106* (34), 14309.

(100) Gebler, J. C.; Woodside, A. B.; Poulter, C. D. Dimethylallyltryptophan synthase. An enzyme-catalyzed electrophilic aromatic substitution. *J Am Chem Soc* **1992,** *114* (19), 7354.

(101) Bandari, C.; Scull, E. M.; Bavineni, T.; Nimmo, S. L.; Gardner, E. D.; Bensen, R. C.; Burgett, A. W.; Singh, S. FgaPT2, a biocatalytic tool for alkyl-diversification of indole natural products. *Med Chem Commun* **2019,** *10* (8), 1465.

(102) Sasaki, K.; Mito, K.; Ohara, K.; Yamamoto, H.; Yazaki, K. Cloning and characterization of naringenin 8-prenyltransferase, a flavonoid-specific prenyltransferase of Sophora flavescens. *Plant Physiol* **2008,** *146* (3), 1075.

(103) Zhou, K.; Ludwig, L.; Li, S.-M. Friedel–Crafts alkylation of acylphloroglucinols catalyzed by a fungal indole prenyltransferase. *J Nat Prod* **2015,** *78* (4), 929.

(104) Hammer, S. C.; Dominicus, J. M.; Syrén, P.-O.; Nestl, B. M.; Hauer, B. Stereoselective Friedel–Crafts alkylation catalyzed by squalene hopene cyclases. *Tetrahedron* **2012,** *68* (37), 7624.

(105) Siedenburg, G.; Jendrossek, D. Squalene-hopene cyclases. *Appl Environ Microbiol* **2011,** *77* (12), 3905.

(106) Johansson Seechurn, C. C. C.; Kitching, M. O.; Colacot, T. J.; Snieckus, V. Palladium-Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize. *Angew Chem Int Ed Engl* **2012,** *51* (21), 5062.

(107) Suzuki, A. Cross-Coupling Reactions Of Organoboranes: An Easy Way To Construct C–C Bonds (Nobel Lecture). *Angew Chem Int Ed Engl* **2011,** *50* (30), 6722.

(108) Kambe, N.; Iwasaki, T.; Terao, J. Pd-catalyzed cross-coupling reactions of alkyl halides. *Chem Soc Rev* **2011,** *40* (10), 4937.

(109) Dombrowski, A. W.; Gesmundo, N. J.; Aguirre, A. L.; Sarris, K. A.; Young, J. M.; Bogdan, A. R.; Martin, M. C.; Gedeon, S.; Wang, Y. Expanding the Medicinal Chemist Toolbox: Comparing Seven C(sp2)–C(sp3) Cross-Coupling Methods by Library Synthesis. *ACS Med Chem Lett* **2020,** *11* (4), 597.

(110) Nicolaou, K. C.; Sun, Y. P.; Korman, H.; Sarlah, D. Asymmetric Total Synthesis of Cylindrocyclophanes A and F through Cyclodimerization and a Ramberg–Bäcklund Reaction. *Angew Chem Int Ed Engl* **2010,** *49* (34), 5875.

(111) Smith, A. B.; Kozmin, S. A.; Adams, C. M.; Paone, D. V. Assembly of (–)-cylindrocyclophanes A and F via remarkable olefin metathesis dimerizations. *J Am Chem Soc* **2000,** *122* (20), 4984.

(112) Smith, A. B.; Adams, C. M.; Kozmin, S. A.; Paone, D. V. Total Synthesis of (–)-Cylindrocyclophanes A and F Exploiting the Reversible Nature of the Olefin Cross Metathesis Reaction. *J Am Chem Soc* **2001,** *123* (25), 5925.

(113) Smith, A. B.; Kozmin, S. A.; Paone, D. V. Total Synthesis of (–)-Cylindrocyclophane F. *J Am Chem Soc* **1999,** *121* (32), 7423.

(114) Hoye, T. R.; Humpal, P. E.; Moon, B. Total Synthesis of (–)-Cylindrocyclophane A via a Double Horner-Emmons Macrocyclic Dimerization Event. *J Am Chem Soc* **2000,** *122* (20), 4982.

(115) Berthold, D.; Breit, B. Total synthesis of (–)-cylindrocyclophane F: a yardstick for probing new catalytic C–C bond-forming methodologies. *Chem Eur J* **2018,** *24* (63), 16770.

(116) Li, C.; Breit, B. Rhodium-catalyzed dynamic kinetic asymmetric allylation of phenols and 2-hydroxypyridines. *Chem Eur J* **2016,** *22* (41), 14655.

(117) Lu, Z.; Fu, G. C. Alkyl-alkyl suzuki cross-coupling of unactivated secondary alkyl chlorides. *Angew Chem Int Ed Engl* **2010,** *49* (37), 6676.

(118) Atack, T. C.; Cook, S. P. Manganese-catalyzed borylation of unactivated alkyl chlorides. *J Am Chem Soc* **2016,** *138* (19), 6139.

(119) González-Bobes, F.; Fu, G. C. Amino alcohols as ligands for nickel-catalyzed suzuki reactions of unactivated alkyl halides, including secondary alkyl chlorides, with arylboronic acids. *J Am Chem Soc* **2006,** *128* (16), 5360.

(120) Owston, N. A.; Fu, G. C. Asymmetric alkyl–alkyl cross-couplings of unactivated secondary alkyl electrophiles: stereoconvergent suzuki reactions of racemic acylated halohydrins. *J Am Chem Soc* **2010,** *132* (34), 11908.

(121) Sidera, M.; Fletcher, S. P. Rhodium-catalysed asymmetric allylic arylation of racemic halides with arylboronic acids. *Nat Chem* **2015,** *7* (11), 935.

(122) González, J.; van Dijk, L.; Goetzke, F. W.; Fletcher, S. P. Highly enantioselective rhodium-catalyzed cross-coupling of boronic acids and racemic allyl halides. *Nat Protoc* **2019,** *14* (10), 2972.

(123) Sakai, H. A.; Liu, W.; Le, C.; MacMillan, D. W. C. Cross-electrophile coupling of unactivated alkyl chlorides. *J Am Chem Soc* **2020,** *142* (27), 11691.

(124) Ertl, P.; Schuhmann, T. A systematic cheminformatics analysis of functional groups occurring in natural products. *J Nat Prod* **2019,** *82* (5), 1258.

(125) Sobol, Z.; Engel, M. E.; Rubitski, E.; Ku, W. W.; Aubrecht, J.; Schiestl, R. H. Genotoxicity profiles of common alkyl halides and esters with alkylating activity. *Mutat Res, Genet Toxicol Environ Mutagen* **2007,** *633* (2), 80.

(126) Kerr, J. A. Bond Dissociation Energies by Kinetic Methods. *Chem Rev* **1966,** *66* (5), 465.

(127) Benson, S. W. III - Bond energies. *J Chem Educ* **1965,** *42* (9), 502.

(128) Schultz, E. E.; Braffman, N. R.; Luescher, M. U.; Hager, H. H.; Balskus, E. P. Biocatalytic Friedel-Crafts alkylation using a promiscuous biosynthetic enzyme. *Angew Chem Int Ed Engl* **2019,** *58* (10), 3151.

(129) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med Chem Lett* **2017,** *8* (5), 476.

(130) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem Rev* **2018,** *118* (2), 801.

(131) Koshland Jr, D. E. Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev* **1953,** *28* (4), 416.

(132) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.et al. Biocatalysis. *Nat Rev Methods Primers* **2021,** *1* (1), 46.

(133) Zhu, Q.; Mai, U.; Pfeiffer, W.; Janssen, S.; Asnicar, F.; Sanders, J. G.; Belda-Ferre, P.; Al-Ghalith, G. A.; Kopylova, E.; McDonald, D.et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* **2019,** *10* (1), 5477.

(134) Schmidt, T. G. M.; Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2007,** *2* (6), 1528.

(135) Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci* **2009,** *18* (5), 936.

(136) di Guan, C.; Li, P.; Riggs, P. D.; Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **1988,** *67* (1), 21.

(137) Smith, D. B.; Johnson, K. S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988,** *67* (1), 31.

(138) Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expres Purif* **2005,** *43* (1), 1.

(139) Costa, S.; Almeida, A.; Castro, A.; Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front Microbiol* **2014,** *5*, 63.

(140) Mossessova, E.; Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **2000,** *5* (5), 865.

(141) Wang, H.-Q.; Mou, S.-B.; Xiao, W.; Zhou, H.; Hou, X.-D.; Wang, S.-J.; Wang, Q.; Gao, J.; Wei, Z.; Liu, L.et al. Structural basis for the Friedel–Crafts alkylation in cylindrocyclophane biosynthesis. *ACS Catal* **2022,** *12* (3), 2108.

(142) Slater, J. C. Atomic Radii in Crystals. *J Chem Phys* **1964,** *41* (10), 3199.

(143) Vaillancourt, F. H.; Vosburg, D. A.; Walsh, C. T. Dichlorination and Bromination of a Threonyl-S-Carrier Protein by the Non-heme FeII Halogenase SyrB2. *ChemBioChem* **2006,** *7* (5), 748.

(144) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H. J.; Schneefeld, M.; Herrmann, J.; Kumpfmüller, J.; Thürmer, A.; Neidhardt, I.; Wiesner, C.et al. Effects of Halide Ions on the Carbamidocyclophane Biosynthesis in Nostoc sp. CAVN2. *Mar Drugs* **2016,** *14* (1).

(145) Ackerman, L. K. G.; Lovell, M. M.; Weix, D. J. Multimetallic catalysed cross-coupling of aryl bromides with aryl triflates. *Nature* **2015,** *524* (7566), 454.

(146) Wincent, E.; Bengtsson, J.; Mohammadi Bardbori, A.; Alsberg, T.; Luecke, S.; Rannug, U.; Rannug, A. Inhibition of cytochrome P4501-dependent clearance of the endogenous agonist FICZ as a mechanism for activation of the aryl hydrocarbon receptor. *Proc Natl Acad Sci USA* **2012,** *109* (12), 4479.

(147) Baccarelli, A.; Pesatori, A. C.; Masten, S. A.; Patterson, D. G.; Needham, L. L.; Mocarelli, P.; Caporaso, N. E.; Consonni, D.; Grassman, J. A.; Bertazzi, P. A.et al. Aryl-hydrocarbon receptor-dependent pathway and toxic effects of TCDD in humans: a population-based study in Seveso, Italy. *Toxicol Lett* **2004,** *149* (1), 287.

(148) Richardson, W. H.; Schmidt, T. M.; Nealson, K. H. Identification of an anthraquinone pigment and a hydroxystilbene antibiotic from Xenorhabdus luminescens. *Appl Environ Microbiol* **1988,** *54* (6), 1602.

(149) Smith, S. H.; Jayawickreme, C.; Rickard, D. J.; Nicodeme, E.; Bui, T.; Simmons, C.; Coquery, C. M.; Neil, J.; Pryor, W. M.; Mayhew, D.et al. Tapinarof Is a Natural AhR Agonist that Resolves Skin Inflammation in Mice and Humans. *J Invest Dermatol* **2017,** *137* (10), 2110.

(150) Tsoi, L. C.; Spain, S. L.; Knight, J.; Ellinghaus, E.; Stuart, P. E.; Capon, F.; Ding, J.; Li, Y.; Tejasvi, T.; Gudjonsson, J. E.et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* **2012,** *44* (12), 1341.

(151) Nestle, F. O.; Kaplan, D. H.; Barker, J. Psoriasis. *New Engl J Med* **2009,** *361* (5), 496.

(152) Elder, J. T.; Bruce, A. T.; Gudjonsson, J. E.; Johnston, A.; Stuart, P. E.; Tejasvi, T.; Voorhees, J. J.; Abecasis, G. R.; Nair, R. P. Molecular dissection of psoriasis: integrating genetics and biology. *J Invest Dermatol* **2010,** *130* (5), 1213.

(153) Kronenwerth, M.; Dauth, C.; Kaiser, M.; Pemberton, I.; Bode, H. Facile Synthesis of Cyclohexanediones and Dialkylresorcinols — Bioactive Natural Products from Entomopathogenic Bacteria. *Eur J Org Chem* **2014,** *2014* (36), 8026.

(154) Krow, G. R.; Miles, W. H.; Smiley, P. M.; Lester, W. S.; Kim, Y. J. Synthesis of antibiotic stilbenes using organomanganese arene complexes. *J Org Chem* **1992,** *57* (14), 4040.

(155) Chen, G.; Webster, J.; Li, J.; Hu, K.; Zhu, J., 2000; Vol. US20030171429A1.

(156) Reis, J. P. A.; Figueiredo, S. A. C.; Sousa, M. L.; Leão, P. N. BrtB is an O-alkylating enzyme that generates fatty acid-bartoloside esters. *Nat Commun* **2020,** *11* (1), 1458.

(157) Muona, M.; Aranko, A. S.; Iwai, H. Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing. *Chembiochem* **2008,** *9* (18), 2958.

(158) Reverter, D.; Lima, C. D. Preparation of SUMO proteases and kinetic analysis using endogenous substrates. *Methods Mol Biol* **2009,** *497*, 225.

(159) Mills, L. R.; Barrera Arbelaez, L. M.; Rousseaux, S. A. L. Electrophilic Zinc Homoenolates: Synthesis of Cyclopropylamines from Cyclopropanols and Amines. *J Am Chem Soc* **2017,** *139* (33), 11357.

(160) Neufeind, S.; Hülsken, N.; Neudörfl, J. M.; Schlörer, N.; Schmalz, H. G. Total synthesis of cyclomumbaistatin analogues through anionic homo-Fries rearrangement. *Chem Eur J* **2011,** *17* (9), 2633.

(161) Uzura, S.; Sekine-Suzuki, E.; Nakanishi, I.; Sonoda, M.; Tanimori, S. A facile and rapid access to resveratrol derivatives and their radioprotective activity. *Bioorg Med Chem Lett* **2016,** *26* (16), 3886.

(162) Braffman, N. R.; Ruskoski, T. B.; Davis, K. M.; Glasser, N. R.; Johnson, C.; Okafor, C. D.; Boal, A. K.; Balskus, E. P. Structural basis for an unprecedented enzymatic alkylation in cylindrocyclophane biosynthesis. *eLife* **2022,** *11*, e75761.

(163) Hart, J. R. *Ethylenediaminetetraacetic acid and related chelating agents*; Wiley: Weinheim, Germany, 2011.

(164) Lipka, M.; Filipek, R.; Bochtler, M. Crystal structure and mechanism of the *Staphylococcus cohnii* virginiamycin B lyase (Vgb). *Biochemistry* **2008,** *47* (14), 4257.

(165) Korczynska, M.; Mukhtar, T. A.; Wright, G. D.; Berghuis, A. M. Structural basis for streptogramin B resistance in Staphylococcus aureus by virginiamycin B lyase. *P Natl Acad Sci USA* **2007,** *104* (25), 10388.

(166) Janssen, D. B.; van der Ploeg, J. R.; Pries, F. Genetics and biochemistry of 1,2-dichloroethane degradation. *Biodegradation* **1994,** *5* (3-4), 249.

(167) Verschueren, K. H. G.; Seljée, F.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **1993,** *363* (6431), 693.

(168) Pries, F.; Kingma, J.; Krooshof, G. H.; Jeronimus-Stratingh, C. M.; Bruins, A. P.; Janssen, D. B. Histidine 289 Is essential for hydrolysis of the alkyl-enzyme intermediate of haloalkane dehalogenase. *J Biol Chem* **1995,** *270* (18), 10405.

(169) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008,** *3* (6), 373.

(170) Linhartová, I.; Bumba, L.; Masin, J.; Basler, M.; Osicka, R.; Kamanová, J.; Procházková, K.; Adkins, I.; Hejnová-Holubová, J.; Sadílková, L.et al. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* **2010,** *34* (6), 1076.

(171) Motlova, L.; Klimova, N.; Fiser, R.; Sebo, P.; Bumba, L. Continuous assembly of beta-roll structures Is implicated in the type I-dependent secretion of large repeat-in-toxins (RTX) proteins. *J Mol Biol* **2020,** *432* (20), 5696.

(172) von Kügelgen, A.; Tang, H.; Hardy, G. G.; Kureisaite-Ciziene, D.; Brun, Y. V.; Stansfeld, P. J.; Robinson, C. V.; Bharat, T. A. M. In situ structure of an intact lipopolysaccharide-bound bacterial surface layer. *Cell* **2020,** *180* (2), 348.

(173) Tanaka, S.; Saito, K.; Chon, H.; Matsumura, H.; Koga, Y.; Takano, K.; Kanaya, S. Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic archaeon: evidence for Ca^{2+}-induced folding. *J Biol Chem* **2007,** *282* (11), 8246.

(174) Buchinger, E.; Knudsen, D. H.; Behrens, M. A.; Pedersen, J. S.; Aarstad, O. A.; Tondervik, A.; Valla, S.; Skjak-Braek, G.; Wimmer, R.; Aachmann, F. L. Structural and functional characterization of the R-modules in alginate C-5 epimerases AlgE4 and AlgE6 from Azotobacter vinelandii. *J Biol Chem* **2014,** *289* (45), 31382.

(175) Meier, R.; Drepper, T.; Svensson, V.; Jaeger, K. E.; Baumann, U. A calcium-gated lid and a large beta-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. *J Biol Chem* **2007,** *282* (43), 31477.

(176) Ravaud, S.; Gouet, P.; Haser, R.; Aghajari, N. Probing the Role of Divalent Metal Ions in a Bacterial Psychrophilic Metalloprotease: Binding Studies of an Enzyme in the Crystalline State by X-Ray Crystallography. *J Bacteriol* **2003,** *185* (14), 4195.

(177) Chen, C. K.; Chan, N. L.; Wang, A. H. The many blades of the beta-propeller proteins: conserved but versatile. *Trends Biochem Sci* **2011,** *36* (10), 553.

(178) Schapira, M.; Tyers, M.; Torrent, M.; Arrowsmith, C. H. WD40 repeat domain proteins: a novel target class? *Nat Rev Drug Discov* **2017,** *16* (11), 773.

(179) Ho, B. K.; Gruswitz, F. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Struct Biol* **2008,** *8*, 49.

(180) Steiner, T. Hydrogen-bond distances to halide ions in organic and organometallic crystal structures: up-to-date database study. *Acta Cryst B* **1998,** *54* (4), 456.

(181) Case, D.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.et al. AMBER 2020. *University of California, San Francisco* **2020**.

(182) Lin, J.; Pozharski, E.; Wilson, M. A. Short carboxylic acid-carboxylate hydrogen bonds can have fully localized protons. *Biochemistry* **2017,** *56* (2), 391.

(183) Schottel, B. L.; Chifotides, H. T.; Dunbar, K. R. Anion-p interactions. *Chem Soc Rev* **2007,** *37*, 68.

(184) Garau, C.; Quiñonero, D.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M. Anion–p interactions: must the aromatic ring be electron deficient? *New J Chem* **2003,** *27*, 211.

(185) Parisini, E.; Metrangolo, P.; Pilati, T.; Resnati, G.; Terraneo, G. Halogen bonding in halocarbon–protein complexes: a structural survey. *Chem Soc Rev* **2011,** *40*, 2267.

(186) Eustáquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat Chem Biol* **2008,** *4*, 69.

(187) Ford, D. D.; Lehnherr, D.; Kennedy, C. R.; Jacobsen, E. N. Anion-abstraction catalysis: the cooperative mechanism of α-chloroether activation by dual hydrogen-bond donors. *ACS Catal* **2016,** *6* (7), 4616.

(188) Levi Samuel, M.; Li, Q.; Rötheli Andreas, R.; Jacobsen Eric, N. Catalytic activation of glycosyl phosphates for stereoselective coupling reactions. *P Natl Acad Sci USA* **2019,** *116* (1), 35.

(189) Pavkov-Keller, T.; Schmidt, N. G.; Żądło-Dobrowolska, A.; Kroutil, W.; Gruber, K. Structure and Catalytic Mechanism of a Bacterial Friedel–Crafts Acylase. *ChemBioChem* **2018,** *20*, 88.

(190) Martins, T. P.; Rouger, C.; Glasser, N. R.; Freitas, S.; de Fraissinette, N. B.; Balskus, E. P.; Tasdemir, D.; Leão, P. N. Chemistry, bioactivity and biosynthesis of cyanobacterial alkylresorcinols. *Nat Prod Rep* **2019,** *36*, 1437.

(191) H., A.; S., A.; E., M.; O., C.; I., M.; T., P.; N., B.-T. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **2016,** *44*, W344.

(192) Messing, S. A.; Gabelli, S. B.; Echeverria, I.; Vogel, J. T.; Guan, J. C.; Tan, B. C.; Klee, H. J.; McCarty, D. R.; Amzel, L. M. Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. *Plant Cell* **2010,** *22* (9), 2970.

(193) Pomowski, A.; Zumft, W. G.; Kroneck, P. M.; Einsle, O. N2O binding at a [4Cu:2S] copper-sulphur cluster in nitrous oxide reductase. *Nature* **2011,** *477* (7363), 234.

(194) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat Chem* **2015,** *7*, 431.

(195) Lau, W.; Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015,** *349*, 1224.

(196) Matthews, M. L.; Neumann, C. S.; Miles, L. A.; Grove, T. L.; Booker, S. J.; Krebs, C.; Walsh, C. T.; Bollinger, J. M. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *P Natl Acad Sci USA* **2009,** *106*, 17723.

(197) Park, Y.; Harper, K. C.; Kuhl, N.; Kwan, E. E.; Liu, R. Y.; Jacobsen, E. N. Macrocyclic bis-thioureas catalyze stereospecific glycosylation reactions. *Science* **2017,** *13*, 162.

(198) Brak, K.; Jacobsen, E. N. Asymmetric Ion-Pairing Catalysis. *Angew Chem Int Edit* **2012,** *52* (2), 534.

(199) Orevi, T.; Lerner, E.; Rahamim, G.; Amir, D.; Haas, E. In *Fluorescence Spectroscopy and Microscopy: Methods and Protocols*; Engelborghs, Y.;Visser, A. J. W. G., Eds.; Humana Press: Totowa, NJ, 2014, DOI:10.1007/978-1-62703-649-8_7 10.1007/978-1-62703-649-8_7.

(200) Matsumoto, S.; Ishida, S.; Araki, M.; Kato, T.; Terayama, K.; Okuno, Y. Extraction of protein dynamics information from cryo-EM maps using deep learning. *Nat Mach Intell* **2021,** *3* (2), 153.

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

(202) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **2007,** *40* (Pt 4), 658.

(203) Read, R. J.; McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr D* **2011,** *67* (Pt 4), 338.

(204) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **2019,** *75* (Pt 10), 861.

(205) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **2004,** *60*, 2126.

(206) Holm, L. *Using Dali for Protein Structure Comparison*; Humana: New York, NY, 2020.

(207) Long, F.; Nicholls, R. A.; Emsley, P.; Graæulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **2017,** *73*, 112.

(208) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J Am Chem Soc* **2001,** *222*, U403.

(209) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157.

(210) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015,** *11* (8), 3696.

(211) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983,** *79* (2), 926.

(212) Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004,** *32*, 1792.

(213) Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *BMC Bioinformatics* **2004,** *5*, 113.

(214) Nguyen, L. T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* **2015,** *32*, 268.

(215) Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* **2017,** *14*, 587.

(216) Hoang, D. T.; Chernomor, O.; von Haeseler, A.; Minh, B. Q.; Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **2017,** *35*, 518.

(217) Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. The transformative power of biocatalysis in convergent synthesis. *J Am Chem Soc* **2022,** *144* (12), 5214.

(218) May, O.; Verseck, S.; Bommarius, A.; Drauz, K. Development of dynamic kinetic resolution processes for biocatalytic production of natural and nonnatural L-amino acids. *Org Proc Res Dev* **2002,** *6* (4), 452.

(219) Yi, D.; Bayer, T.; Badenhorst, C. P. S.; Wu, S.; Doerr, M.; Höhne, M.; Bornscheuer, U. T. Recent trends in biocatalysis. *Chem Soc Rev* **2021,** *50* (14), 8003.

(220) Petchey, M. R.; Rowlinson, B.; Lloyd, R. C.; Fairlamb, I. J. S.; Grogan, G. Biocatalytic synthesis of moclobemide using the amide bond synthetase McbA coupled with an ATP recycling system. *ACS Catal* **2020,** *10* (8), 4659.

(221) McIntosh, J. A.; Benkovics, T.; Silverman, S. M.; Huffman, M. A.; Kong, J.; Maligres, P. E.; Itoh, T.; Yang, H.; Verma, D.; Pan, W.et al. Engineered ribosyl-1-kinase enables concise synthesis of molnupiravir, an antiviral for COVID-19. *ACS Cent Sci* **2021,** *7* (12), 1980.

(222) Sandoval, B. A.; Hyster, T. K. Emerging strategies for expanding the toolbox of enzymes in biocatalysis. *Curr Opin Chem Biol* **2020,** *55*, 45.

(223) Leisola, M.; Jokela, J.; Pastinen, O.; Turunen, O. Industrial use of enzymes. *Physiology and Maintenance* **2011,** *2*.

(224) Bruggink, A.; Roos, E. C.; de Vroom, E. Penicillin acylase in the industrial production of β-lactam antibiotics. *Org Process Res Dev* **1998,** *2* (2), 128.

(225) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185.

(226) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **2015,** *16* (7), 379.

(227) Pluchinsky, A. J.; Wackelin, D. J.; Huang, X.; Arnold, F. H.; Mrksich, M. High throughput screening with SAMDI mass spectrometry for directed evolution. *J Am Chem Soc* **2020,** *142* (47), 19804.

(228) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B. MISER chromatography (multiple injections in a single experimental run): the chromatogram is the graph. *Tetrahedron-Asymmetr* **2010,** *21* (13), 1674.

(229) Eisenstein, M. Enzymatic DNA synthesis enters new phase. *Nat Biotechnol* **2020,** *38* (10), 1113.

(230) Weng, L.; Spoonamore, J. E. Droplet microfluidics-enabled high-throughput screening for protein engineering. *Micromachines (Basel)* **2019,** *10* (11), 734.

(231) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583.

(232) Goodwin, N. C.; Morrison, J. P.; Fuerst, D. E.; Hadi, T. Biocatalysis in medicinal chemistry: challenges to access and drivers for adoption. *ACS Med Chem Lett* **2019,** *10* (10), 1363.

(233) Imai, Y.; Meyer, K. J.; Iinishi, A.; Favre-Godal, Q.; Green, R.; Manuse, S.; Caboni, M.; Mori, M.; Niles, S.; Ghiglieri, M.et al. A new antibiotic selectively kills Gram-negative pathogens. *Nature* **2019,** *576* (7787), 459.

(234) Schmitt, S.; Montalbán-López, M.; Peterhoff, D.; Deng, J.; Wagner, R.; Held, M.; Kuipers, O. P.; Panke, S. Analysis of modular bioengineered antimicrobial lanthipeptides at nanoliter scale. *Nat Chem Biol* **2019,** *15* (5), 437.

(235) Mocciaro, A.; Roth, T. L.; Bennett, H. M.; Soumillon, M.; Shah, A.; Hiatt, J.; Chapman, K.; Marson, A.; Lavieu, G. Light-activated cell identification and sorting (LACIS) for selection of edited clones on a nanofluidic device. *Commun Biol* **2018,** *1*, 41.

(236) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* **1999,** *4* (2), 67.

(237) Casper, R. F.; Quesne, M.; Rogers, I. M.; Shirota, T.; Jolivet, A.; Milgrom, E.; Savouret, J. F. Resveratrol has antagonist activity on the aryl hydrocarbon receptor: implications for prevention of dioxin toxicity. *Mol Pharmacol* **1999,** *56* (4), 784.

(238) Smith, K. J.; Murray, I. A.; Tanos, R.; Tellew, J.; Boitano, A. E.; Bisson, W. H.; Kolluri, S. K.; Cooke, M. P.; Perdew, G. H. Identification of a high-affinity ligand that exhibits complete aryl hydrocarbon receptor antagonism. *J Pharmacol Exp Ther* **2011,** *338* (1), 318.

(239) Kobayashi, A.; Sogawa, K.; Fujii-Kuriyama, Y. Cooperative interaction between AhR.Arnt and Sp1 for the drug-inducible expression of CYP1A1 gene. *J Biol Chem* **1996,** *271* (21), 12310.

(240) Ye, W.; Chen, R.; Chen, X.; Huang, B.; Lin, R.; Xie, X.; Chen, J.; Jiang, J.; Deng, Y.; Wen, J. AhR regulates the expression of human cytochrome P450 1A1 (CYP1A1) by recruiting Sp1. *Febs J* **2019,** *286* (21), 4215.

(241) Schiering, C.; Wincent, E.; Metidji, A.; Iseppon, A.; Li, Y.; Potocnik, A. J.; Omenetti, S.; Henderson, C. J.; Wolf, C. R.; Nebert, D. W.et al. Feedback control of AHR signalling regulates intestinal immunity. *Nature* **2017,** *542* (7640), 242.

(242) Li, A.; Acevedo-Rocha, C. G.; Reetz, M. T. Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Appl Microbiol Biot* **2018,** *102* (14), 6095.

(243) Nov, Y. When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl Environ Microbiol* **2012,** *78* (1), 258.

(244) Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol* **1996,** *255* (4), 589.

(245) Lin-Goerke, J. L.; Robbins, D. J.; Burczak, J. D. PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques* **1997,** *23* (3), 409.

(246) Püllmann, P.; Ulpinnis, C.; Marillonnet, S.; Gruetzner, R.; Neumann, S.; Weissenborn, M. J. Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci Rep* **2019,** *9* (1), 10932.

(247) Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y. Economical analysis of saturation mutagenesis experiments. *Sci Rep* **2015,** *5*, 10654.

(248) Tang, L.; Gao, H.; Zhu, X.; Wang, X.; Zhou, M.; Jiang, R. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques* **2012,** *52* (3), 149.

(249) Chen, Z.; Zhao, H. Rapid creation of a novel protein function by in vitro coevolution. *J Mol Biol* **2005,** *348* (5), 1273.

(250) O'Donnell, S. M.; Janssen, G. R. The initiation codon affects ribosome binding and translational efficiency in Escherichia coli of cI mRNA with or without the 5' untranslated leader. *J Bacteriol* **2001,** *183* (4), 1277.

(251) Leipold, L.; Dobrijevic, D.; Jeffries, J. W. E.; Bawn, M.; Moody, T. S.; Ward, J. M.; Hailes, H. C. The identification and use of robust transaminases from a domestic drain metagenome. *Green Chem* **2019,** *21* (1), 75.