Engineering Biocatalysts for Oxidative Cross-Coupling Reactions

by

Lara E. Zetzsche

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2022

Doctoral Committee:

Associate Professor Alison R. H. Narayan, Chair
Assistant Professor Jennifer Bridwell-Rabb
Professor Robert Kennedy
Assistant Professor Kristin Koutmou
Associate Professor Pavel Nagorny

Lara E. Zetzsche
zetzsche@umich.edu
ORCID iD: 0000-0002-1644-7769

© Lara E. Zetzsche 2022

To Josti, who inspires me to pursue a career driven by my curiosities

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research mentors who have all helped shape me into the scientist that I am today. Professor Alison Narayan has been a steadfast mentor over these last five years. The endless motivation, trust, and advocacy has been pivotal in shaping me into an independent researcher. Before her, Professor Alison Fisher was the first person that allowed me to apply my scientific curiosity in a laboratory setting, inspiring me to pursue a PhD. Thank you to Dr. Margie Borra-Garske, who adopted me into her team at Codexis for a summer and taught me the ways of directed evolution. Lastly, thank you to my dissertation committee for offering their valuable time and feedback throughout my PhD.

Next, I would like to thank all of the amazing scientists that I have had the opportunity to work with over the last five years. I owe a big shoutout to the entire Narayan lab—you all made working in the lab fun and motivated me to keep pushing even when the science wasn't working. I couldn't have done this work without the P450 team, especially Suman Chakrabarty and Casey Roos, who synthesized countless substrates and product standards, as well as Meagan Hinze and Jessica Yazarians for their foundational work investigating the substrate scope of KtnC. Lastly, thank you to the computational and analytical chemists who have I have been grateful to collaborate with and have made important contributions to this project, including Leo Joyce, Spencer McMinn, Eugene Kwan, and the Kennedy lab.

Finally, to my friends, thank you for making Ann Arbor feel like home. To Graham, thank you for being my companion along this ride and always encouraging me to stand up for myself. And to my parents, thank you for always being my biggest cheerleaders.

TABLE OF CONTENTS

Dedication ii
Acknowledgements iii
List of Figures viii
List of Tables xi
List of Schemes xii
List of Abbreviations xiii
Abstract xv

CHAPTER I: Introduction: Biocatalytic C–C Bond Formation

Summary 1
Introduction 2
Alkylation and acylation reactions 3
S-adenosyl-L-methionine (SAM)-dependent enzymes 3
Prenyltransferases 5
Pyridoxal 5-phosphate (PLP)-dependent enzymes 5
Halogenated electrophiles 6
Intramolecular Friedel–Crafts alkylations 6
Friedel–Crafts acylations 8
Cyclization reactions 9
Terpene cyclases 9
Diels–Alderases 11
Ene-reductases 12

Carbene transfer reactions 13
Engineering carbene transferases 13
Olefin cyclopropanation 15
Carbene insertion into C–H bonds 15
Other reactions catalyzed by carbene transferases 16
Oxidative C–C coupling reactions 16
Intramolecular C–C coupling 17
Intermolecular C–C coupling 17
Dissertation overview 19
References 21

CHAPTER II: Development of a Biocatalytic Platform for Biaryl Coupling in Yeast
Summary 35
Introduction 36
Results and discussion 38
Experimental procedures 43
Design of expression constructs 43
Heterologous expression and biotransformations in S. cerevisiae 45
Heterologous expression and biotransformations in P. pastoris 45
Western blot analysis of protein expression 46
Heterologous expression and whole-cell reactions in E. coli 46
LC–MS analysis of reactions 47
Compound toxicity screen 47
Preparative scale biotransformations 48
Assignment of absolute configuration of biocatalytic product 58
Generation of sequence similarity networks 50
Supplementary information 54
DNA and protein sequences 54
Chemical synthesis of substrates and authentic product standards 74
LC–MS traces of biocatalytic reactions 76
NMR spectra of synthesized compounds 80

References 82

CHAPTER III: Directed Evolution of KtnC for Unnatural Cross-Coupling Reactions

Summary 87
Introduction 88
Results and discussion 89
Experimental procedures 96
Analytical scale biotransformations with KtnC 96
LC–MS analysis of analytical scale reactions 96
Design and generation of libraries using site-saturation mutagenesis 97
Generation of combinatorial libraries using multiple site-directed mutagenesis 98
High-throughput biotransformations 102
RapidFire–MS screen for activity (tier 1) 102
LC–MS screen for selectivity (tier 2) 103
Hit validation and sequencing 104
Preparative scale biotransformation with engineered LxC5 variant 104
Assignment of absolute configuration of biocatalytic product 105
Supplementary information 107
DNA and protein sequences 107
Chemical synthesis of substrates and authentic product standards 113
Standard curves and LC–MS traces of biocatalytic reactions 134
NMR spectra of synthesized compounds 160
References 196

CHAPTER IV: Engineering a Bacterial P450 for Atroposelective Biaryl Coupling Reactions

Summary 203
Introduction 204
Results and discussion 205
Experimental procedures 213
Expression and purification of CYP158A2 213
Cloning CYP158A2 into A2-RhFRed fusion construct 214
Expression and purification of P450-RhFRed 215

Reactions with purified P450-RhFRed 215
Expression of P450-RhFRed libraries in 96-well plates 215
High-throughput reactions with P450-RhFRed lysate 215
High-throughput RapidFire–MS screen for activity 216
LC–MS analysis of biocatalytic reactions 216
Design and generation of libraries using site-saturation mutagenesis 217
Generation of combinatorial libraries using multiple site-directed mutagenesis 218
High-throughput screening and analysis of libraries 218
Supplementary information 222
DNA and protein sequences 222
Chemical synthesis of substrates and authentic product standards 239
Standard curves and LC–MS traces of biocatalytic reactions 249
NMR spectra of synthesized compounds 262
References 274

CHAPTER V: Conclusions and Outlook

Summary 276
Dissertation summary 277
Future directions 280
Broader outlook: Biocatalysis in convergent synthesis 282
References 284

LIST OF FIGURES

Figure 1.1. Biocatalytic alkylation reaction using electrophilic alkyl donors..... 4
Figure 1.2. Biosynthetic and biocatalytic ring-closing reactions..... 7
Figure 1.3. Site-selective acylation of resorcinol derivatives..... 9
Figure 1.4. Cyclization reactions catalyzed by natural and engineered enzymes 10
Figure 1.5. Engineering non-natural carbene transferases..... 14
Figure 1.6. Intermolecular oxidative C–C coupling reactions in Nature..... 18
Figure 1.7. Dissertation overview..... 19
Figure 2.1. Significance and biosynthesis of biaryl molecules..... 36
Figure 2.2. Optimization of biocatalytic platform for fungal P450 biocatalysis in yeast 38
Figure 2.3. Engineering self-sufficient fusion enzyme..... 40
Figure 2.4. Application of biocatalytic platform for biaryl synthesis 41
Figure 2.5. Sequence similarity network of fungal P450s that catalyze biaryl couplings 42
Figure 2.6. Strategy for cloning KtnC-CPR self-cleaving and fusion constructs..... 44
Figure 2.7. Western blot analysis of recombinant KtnC production in yeast 46
Figure 2.8. Effect of compound and cosolvent concentration on cell survival..... 47
Figure 2.9. Assignment of absolute configuration of biocatalytic product..... 49
Figure 2.10. Sequence similarity network of full P450 family..... 53
Figure 2.11. Site-selectivity of coumarin dimerization in yeast biotransformation 76
Figure 2.12. Effect of substrate loading on reaction yield in yeast biotransformation..... 77
Figure 2.13. Effect of cell density on reaction yield in yeast biotransformation 78
Figure 2.14. Effect of chimeric P450-CPR fusion construct on yeast biotransformation..... 79
Figure 3.1. Biaryl bond formation through direct oxidative cross-coupling 88

Figure 3.2. Scope of oxidative cross-coupling reactions catalyzed by wild-type KtnC..... 90
Figure 3.3. Development of a high-throughput engineering strategy 92
Figure 3.4. Engineering P450 biocatalyst for improved activity and site-selectivity..... 93
Figure 3.5. Engineering P450 biocatalyst for improved atroposelectivity 94
Figure 3.6. Profiling the reactivity of engineered LxC variants 95
Figure 3.7. Summary of engineering data from each round of evolution..... 103
Figure 3.8. Assignment of absolute configuration of cross-coupled product 106
Figure 3.9. Standard curves for oxidative cross-coupling of **3.4** and **3.8–3.16** by KtnC..... 134
Figure 3.10. Standard curves for oxidative cross-coupling of **3.4** and **3.17**–**3.21** by KtnC..... 135
Figure 3.11. Standard curves for oxidative cross-coupling of **3.4** and **3.22**–**3.26** by KtnC..... 136
Figure 3.12. LC–MS traces for oxidative cross-coupling of **3.4** and **3.8** by KtnC..... 137
Figure 3.13. LC–MS traces for oxidative cross-coupling of **3.4** and **3.9** by KtnC..... 138
Figure 3.14. LC–MS traces for oxidative cross-coupling of **3.4** and **3.10** by KtnC..... 139
Figure 3.15. LC–MS traces for oxidative cross-coupling of **3.4** and **3.11** by KtnC..... 140
Figure 3.16. LC–MS traces for oxidative cross-coupling of **3.4** and **3.12** by KtnC..... 141
Figure 3.17. LC–MS traces for oxidative cross-coupling of **3.4** and **3.13** by KtnC..... 142
Figure 3.18. LC–MS traces for oxidative cross-coupling of **3.4** and **3.14** by KtnC..... 143
Figure 3.19. LC–MS traces for oxidative cross-coupling of **3.4** and **3.15** by KtnC..... 144
Figure 3.20. LC–MS traces for oxidative cross-coupling of **3.4** and **3.16** by KtnC..... 145
Figure 3.21. LC–MS traces for oxidative cross-coupling of **3.4** and **3.17** by KtnC..... 146
Figure 3.22. LC–MS traces for oxidative cross-coupling of **3.4** and **3.18** by KtnC..... 147
Figure 3.23. LC–MS traces for oxidative cross-coupling of **3.4** and **3.19** by KtnC..... 148
Figure 3.24. LC–MS traces for oxidative cross-coupling of **3.4** and **3.20** by KtnC..... 149
Figure 3.25. LC–MS traces for oxidative cross-coupling of **3.4** and **3.21** by KtnC..... 150
Figure 3.26. LC–MS traces for oxidative cross-coupling of **3.4** and **3.22** by KtnC..... 151
Figure 3.27. LC–MS traces for oxidative cross-coupling of **3.4** and **3.23** by KtnC..... 152
Figure 3.28. LC–MS traces for oxidative cross-coupling of **3.4** and **3.24** by KtnC..... 153
Figure 3.29. LC–MS traces for oxidative cross-coupling of **3.4** and **3.25** by KtnC..... 154
Figure 3.30. LC–MS traces for oxidative cross-coupling of **3.4** and **3.26** by KtnC..... 155
Figure 3.31. Site-selectivity in oxidative cross-coupling of **3.9** and **3.26** by KtnC..... 156
Figure 3.32. Reactivity in oxidative cross-coupling over evolution progress..... 157

Figure 3.33. Chemoselectivity in oxidative cross-coupling over evolution progress..... 158
Figure 3.34. Site-selectivity in oxidative cross-coupling over evolution progress..... 159
Figure 4.1. Significance and synthesis of chiral biaryl ligands 204
Figure 4.2. Bioinformatic analysis of P450 sequence space 206
Figure 4.3. Validation of P450-RhFRed engineering strategy for bacterial P450s 207
Figure 4.4. Exploring sequence space for P450s with complementary reactivity..... 208
Figure 4.5. Development of a high-throughput engineering strategy 209
Figure 4.6. Engineering P450 biocatalyst for improved activity and atroposelectivity 210
Figure 4.7. Trends in atroposelectivity with different cross-coupling partners 212
Figure 4.8. UPLC traces for oxidative coupling of flaviolin by CYP158A2..... 249
Figure 4.9. Quantification of percent yields using product standard curves..... 250
Figure 4.10. LC–MS traces for biocatalytic oxidative cross-coupling to form **4.21**..... 251
Figure 4.11. LC–MS traces for biocatalytic oxidative cross-coupling to form **4.22**..... 252
Figure 4.12. LC–MS traces for biocatalytic oxidative cross-coupling to form **4.23**..... 253
Figure 4.13. LC–MS traces for biocatalytic oxidative cross-coupling to form **4.24**..... 254
Figure 4.14. LC–MS traces for biocatalytic oxidative cross-coupling to form **4.25**..... 255
Figure 4.15. Site-selectivity of oxidative cross-couplings catalyzed by A2-RhFRed..... 256
Figure 4.16. Site-selectivity of oxidative cross-couplings catalyzed by P450-RhFRed..... 257
Figure 4.17. Atroposelectivity of oxidative cross-coupling to form **4.23**..... 258
Figure 4.18. Atroposelectivity of oxidative cross-coupling to form **4.1** and **4.27**..... 259
Figure 4.19. Atroposelectivity of oxidative cross-coupling to form **4.26** and **4.30**..... 260
Figure 4.20. Atroposelectivity of oxidative cross-coupling to form **4.28**..... 261
Figure 4.21. Atroposelectivity of oxidative cross-coupling to form **4.29**..... 261
Figure 5.1. Dissertation summary..... 279
Figure 5.2. Unexplored bacterial P450 sequence space..... 280
Figure 5.3. Proposed mechanism for biaryl coupling in P450 active site 281
Figure 5.4. Biocatalysis in the convergent synthesis of pharmaceutical agents..... 283

LIST OF TABLES

Table 2.1. Primers used for cloning 43
Table 2.2. Fungal P450 sequences used in sequence similarity network 50
Table 3.1. Primers used for site-saturation mutagenesis of KtnC active site residues..... 98
Table 3.2. Primers used for multiple site-directed mutagenesis combinatorial libraries 101
Table 4.1. Panel of 23 enzymes identified in SSN and screened in lysate reactions..... 215
Table 4.2. Primers used for site-saturation mutagenesis of CYP158A2 active site residues 219
Table 4.3. Primers used for multiple site-directed mutagenesis combinatorial libraries 221
Table 4.4. Screening summary for engineering campaign 221

LIST OF SCHEMES

Scheme 2.1. Preparative scale biotransformation with KtnC to form **2.5**..... 48
Scheme 3.1. Preparative scale biotransformation with LxC5 to form **3.27** 104
Scheme 3.2. Synthesis of benzofuran **3.22**..... 122
Scheme 3.3. Synthesis of 8,1′ cross-coupled product **3.27**..... 125
Scheme 3.4. Synthesis of 6,1′ cross-coupled product **3.28**..... 128
Scheme 3.5. Synthesis of 3,1′ cross-coupled product **3.29**..... 131

LIST OF ABBREVIATIONS

BBE	berberine bridge enzyme
BINOL	1,1′-bi-2-naphthol
BMG	buffered minimal glycerol medium
BMM	buffered minimal methanol medium
C–C	carbon–carbon
CPR	NAPDH-dependent cytochrome P450 reductase
DFT	discrete Fourier transform
DMSO	dimethylsulfoxide
dNTP	deoxynucleoside triphosphate
DTT	diothiothreitol
EDTA	ethylenediaminetetraacetic acid
EFI–EST	Enzyme Function Initiative–Enzyme Similarity Tool
EIC	extracted ion chromatogram
ECD	electronic circular dichroism
ESI	electrospray ionization
Fldx	flavodoxin
FldR	flavodoxin reductase
G6P	glucose-6-phosphate
G6PDH	glucose-6-phosphate dehydrogenase
HPLC	high performance liquid chromatography
IDT	Integrated DNA Technologies
IPTG	isopropyl-β-D-1-thiogalactopyranoside

IS	internal standard
LB	Luria-Bertani
LC–MS	liquid chromatography–mass spectrometry
MAFFT	multiple alignment using fast Fourier transform
MS	mass spectrometry
NADPH	nicotinamide adenine dinucleotide phosphate
NCS	norcoclaurine synthase
NEB	New England Biolabs
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
P450	cytochrome P450 monooxygenase
PCR	polymerase chain reaction
PDA	photodiode-array detection
PDB	Protein Database
PLP	pyridoxal 5-phosphate
QTOF	quadrupole time-of-flight
RhFRed	$P450_{RhF}$ reductase domain
RMS	root mean square
SAM	*S*-adenosyl-L-methionine
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SFC	supercritical fluid chromatography
SHC	squalene-hopene cyclase
S_NAc	nucleophilic acyl substitution
SSN	sequence similarity network
TDDFT	time-dependent density-functional theory
ThDP	thiamine diphosphate
TLC	thin layer chromatography
TOF	time-of-flight
UPLC	ultra performance liquid chromatography
UV	ultraviolet
YNB	yeast nitrogen base

ABSTRACT

The precision and efficiency offered by enzyme catalysis has propelled the incorporation of biocatalytic strategies into synthetic campaigns in academic and industrial spheres. Nevertheless, biocatalytic methods remain prohibitively underdeveloped compared to the vast repertoire of their small molecule-based counterparts, including for the formation of carbon–carbon (C–C) bonds. Oxidative coupling reactions offer a direct and complexity-generating approach to C–C bond formation; however, controlling the selectivity of these reactions remains an outstanding challenge in organic synthesis. In contrast, Nature has evolved a wealth of cytochrome P450 enzymes that catalyze selective oxidative C–C coupling reactions in the biosynthesis of biaryl natural products. Inspired by this catalyst-controlled selectivity, the work presented in this dissertation describes a platform for biocatalytic biaryl bond formation using P450 catalysts. Using a yeast-based platform for whole-cell biocatalysis, we demonstrate the ability of a fungal P450 to catalyze unnatural cross-coupling reactions on a panel of phenolic substrates. Moreover, over five rounds of directed evolution, we engineer the fungal P450 to possess the desired reactivity and site-selectivity by improving a low-yielding, unselective reaction 92-fold for the target cross-coupled product. We further expanded the biocatalytic platform by profiling the cross-coupling activity of a library of natural P450s using a bioinformatics-guided approach, leading to the identification of a bacterial P450 capable of cross-coupling substituted naphthols. Despite demonstrating promiscuous activity, the P450 offered little to no stereocontrol in cross-coupling reactions. To overcome this limitation, the P450 was engineered for improved atroposelectivity, ultimately providing a biocatalyst capable of cross-couplings with up to 95:5 er in the formation of privileged chiral biaryl ligands. Altogether, this streamlined method for constructing sterically hindered biaryl bonds provides a programmable platform for assembling biaryl molecules with catalyst-controlled reactivity and selectivity.

CHAPTER I

Introduction: Biocatalytic C–C Bond Formation

Enzymes exercise impeccable control over chemoselectivity, site-selectivity and stereoselectivity in reactions they mediate, such that we have witnessed a surge in the development of new biocatalytic methods. Although carbon–carbon (C–C) bonds are the central framework of organic molecules, biocatalytic methods for their formation have largely been limited to a select few lyase enzymes. Thus, despite several decades of research, there are not many biocatalytic C–C-bond-forming transformations at our disposal. This chapter describes the suite of enzymes available for highly selective biocatalytic C–C bond formation. Each class of enzyme is discussed in terms of native activity, alteration of this activity through protein or substrate engineering, and its utility in abiotic synthesis. Examples are categorized based on their application for (1) alkylation and acylation reactions, (2) cyclization reactions, (3) carbene transfer reactions, and (4) oxidative C–C coupling reactions. Finally, an overview of this thesis is provided within the context of expanding the biocatalytic toolbox for oxidative C–C coupling reactions with the formation of biaryl bonds.

This chapter was adapted with permission from Zetzsche, L. E. & Narayan, A. R. H. Broadening the Scope of Biocatalytic C–C Bond Formation. *Nat. Rev. Chem.* **2020**, *4*, 334–346. © Springer Nature Limited 2020.

INTRODUCTION

Enzymes catalyze a myriad of highly selective transformations in the biosynthesis of primary and secondary metabolites in Nature.[1-3] Major advances in enzyme discovery and engineering over the last several decades have brought with them a surge in the development and implementation of biocatalytic reactions in organic syntheses executed in both academic and industrial laboratories.[4-6] Enzymes are attractive catalysts in organic synthesis due to (1) the control they exhibit over the chemo-, site-, and stereoselectivity in the reactions they catalyze, (2) the new retrosynthetic disconnections they can offer over small molecule-mediated transformations, and (3) their sustainable footprint, especially in industrial-scale reactions.[6-9]

The major barrier in the use of biocatalysts in organic synthesis is the identification and development of a biocatalyst for a given reaction.[10] This can be challenging because, while the three-dimensional control of a protein active site lends itself toward high levels of selectivity in biocatalytic reactions, this same tight regulation can limit the versatility of a biocatalyst through a restricted substrate scope.[11] To overcome this limitation, enzymes can be engineered for the desired chemistry through rational design or random mutagenesis.[12] The power of iterative rounds of random mutagenesis and screening was recently recognized with the 2018 Nobel Prize in Chemistry, awarded for the directed evolution of enzymes. In this approach, evolutionary pressure inherent to the development of enzymes in Nature is expedited and applied artificially in a laboratory setting to evolve an enzyme for a target transformation or process.[13-15] Additionally, advances in bioinformatic tools and machine learning are progressively merging the use of random and rational protein design for increasingly rapid protein engineering.[16-20]

The utility of using enzymes to address synthetic challenges has been demonstrated over the last several decades with the increased implementation of biocatalytic processes in the synthesis of pharmaceutical compounds.[5, 6] However, a major limitation to the broader use of biocatalysts in both academic and industrial labs is the relatively narrow set of biocatalysts that can be reliably designed into a synthesis.

Carbon–carbon (C–C) bonds provide the central framework for all organic molecules and thus there are countless enzymes involved in their formation in primary and secondary metabolite pathways. Yet, the commonly employed biocatalysts for C–C bond formation are primarily restricted to a small subset of lyases, thereby limiting the scope of biocatalytic C–C bond formation

to aldol reactions, acyloin condensations, and cyanohydrin formation in organic synthesis.[21] Aldolases are by far the most commonly employed lyases for biocatalytic C–C bond formation and have been extensively developed into a versatile class of biocatalysts for site- and stereoselective aldol reactions.[22-24] Thiamine diphosphate (ThDP)-dependent enzymes perform acyloin condensation reactions and have been used as valuable biocatalysts in the synthesis of enantioenriched building blocks.[25-27] Lastly, biocatalytic cyanohydrin formation is achievable through several different mechanisms: (1) the addition of a cyanide ion to a ketone or aldehyde catalyzed by a hydroxynitrile lyase[28-30] or (2) the opening of an epoxide catalyzed halohydrin dehalogenases.[31] The breadth of research on these biocatalysts for C–C bond formation is captured in previous reviews on this topic.[24, 32, 33]

Discovery and application of enzymes that carry out new reactions are essential for the broader implementation of biocatalysts in organic synthesis, including new catalysts for C–C bond formation. Over the last several years, there have been numerous advances in biocatalytic C–C bond formation, using both wild-type enzymes and enzymes engineered for non-natural functions. This chapter provides a literature review on recent advances in biocatalytic C–C bond formation reactions including alkylation, acylation, oxidative C–C coupling, cyclization, and carbene transfer reactions.

ALKYLATION AND ACYLATION REACTIONS

Alkylation reactions are ubiquitous in Nature and are one of the key C–C bond-forming events in natural product biosynthesis. These reactions can occur through various mechanisms, however one of the most common is an S_N2-type reaction utilizing an activated electrophilic cofactor and a nucleophilic substrate (Figure 1.1).

***S*-adenosyl-L-methionine (SAM)-dependent enzymes.** SAM (**1.1**) is the major methyl donor in biological systems and the second most common enzymatic cofactor after ATP.[34, 35] The most common reaction catalyzed by enzymes harboring a SAM cofactor is the transfer of the electrophilic methyl group bound to the sulfonium center of SAM through an S_N2 reaction (Figure 1.1a). However, SAM-dependent enzymes can also perform radical reactions for the transfer of methyl or adenosyl groups,[36, 37] including those for C–C bond formation.[38] Methyltransferases utilizing SAM cofactors are incredibly important in biological systems for applications including the biosynthesis of small molecules, protein repair, gene silencing, and chromatin remodelling.[39]

Unsurprisingly, given the crucial biological ramifications of these methylation events, the ability of these methyltransferases to selectively methylate their substrates is critical. This selective alkylation of such a diverse range of biomolecules has motivated the use of SAM-dependent methyltransferases as tools for biorthogonal labelling in biotechnology.[40-44] In these systems, synthetic analogues of SAM are used to label nucleic acids or proteins by replacing the sulfonium center with an aziridine ring or larger alkyl or allyl group.[45] Upon nucleophilic attack of the target molecule on the SAM analogues, the resulting product will either be trapped by the aziridine analogue or alkylated with the alternative sulfonium substituent.[45]

Although methyltransferases have been used extensively for the modification and labelling of nucleic acids and proteins, methyltransferases using SAM analogues have only sparingly been explored for the biocatalytic C–C alkylation of small molecules in abiotic synthesis. Currently, the only example of biocatalytic C–C alkylation using SAM analogues is for a site-selective Friedel–Crafts alkylation at the unactivated C–H bond of a coumarin substrate to form **1.2** in *Streptomyces*

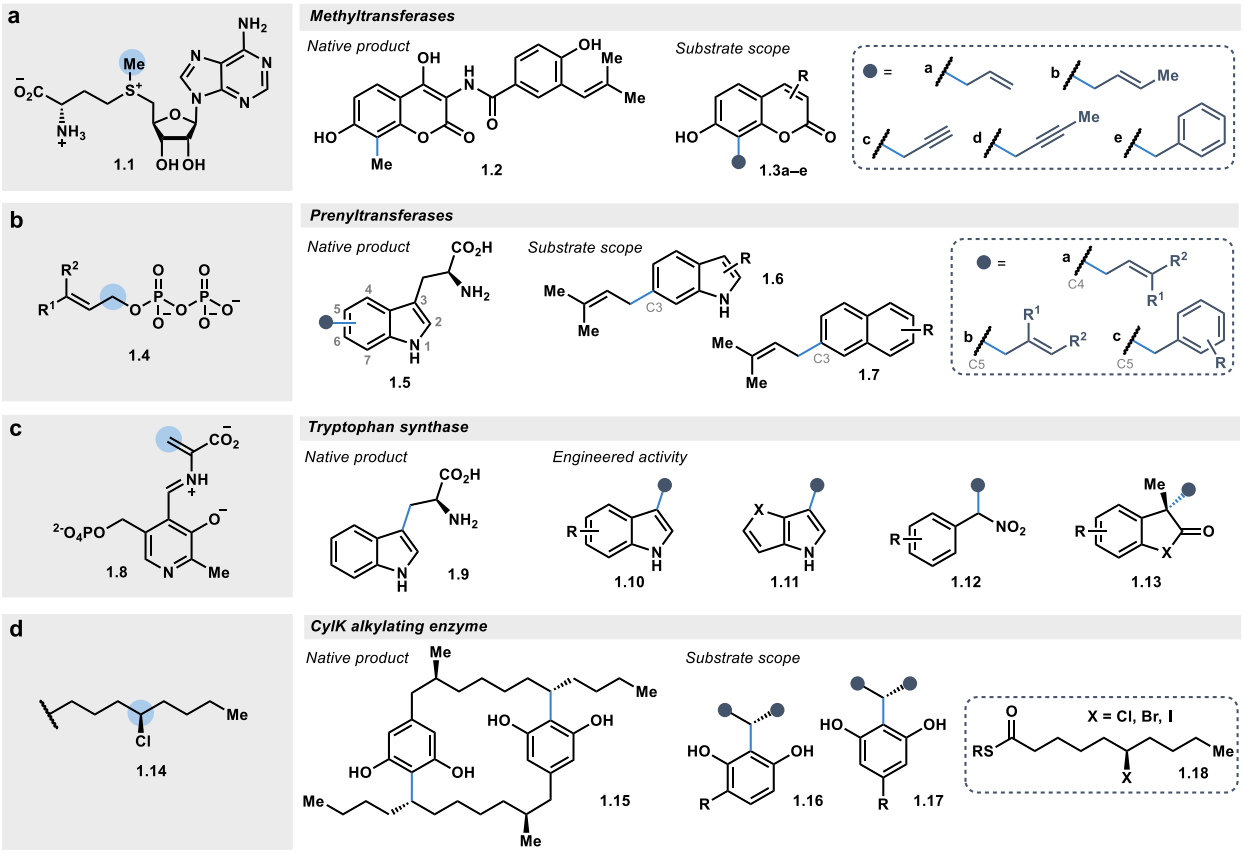


Figure 1.1. Biocatalytic alkylation reactions using electrophilic alkyl donors. (a) Methylation reactions catalyzed by SAM-dependent enzymes can be diversified with the use of synthetic SAM analogues for broader C–C alkylation reactions. (b) The diphosphate prenyl donors and acceptors in prenyltransferase-catalyzed reactions can be modified for catalyst- and substrate-controlled site-selective alkylation reactions. (c) Tryptophan synthase, a PLP-dependent enzyme, has been engineered to catalyze alkylation reactions on substituted indoles, nitro alkanes, and oxindoles. (d) CylK utilizes an alkyl halide in the Friedel–Crafts alkylation of resorcinol derivatives.

secondary metabolite pathways (Figure 1.1a).[46] In these reactions, SAM analogues synthesized with bulkier substituents at the sulfonium center were used as cofactors for the methyltransferase-catalyzed Friedel–Crafts alkylation of several different coumarin derivatives to form coumarin **1.3** modified with allyl (**a**–**b**), propargyl (**c**–**d**), and benzyl (**e**) groups with excellent site-selectivity.[46]

Prenyltransferases. Using a similar strategy to SAM-dependent methyltransferases, prenyltransferases take advantage of the excellent leaving group capability of the diphosphate group on allyl diphosphate substrates (**1.4**) for the transfer of an allylic group to a second more electron-rich substrate.[47, 48] The catalyst- and substrate-controlled site-selectivity exhibited by these enzymes have made them attractive biocatalysts for late-stage functionalization of natural products, especially due to the diverse biological effects the addition of a hydrophobic prenyl group can provide to a molecule.[49]

The investigation of the substrate scopes of several tryptophan prenyltransferases have shown that these enzymes are often promiscuous in their ability to catalyze the site-selective carbon-prenylation of a broad scope of aromatic tryptophan derivatives.[48] Like SAM-dependent methyltransferases, both the alkyl donating and accepting substrates can be modified to diversify these reactions (Figure 1.1b). A recent report demonstrated that a single wild-type prenyltransferase (FgaPT2) could donate alkyl groups from 34 different diphosphate derivatives, each forming a distinct C–C bond to give an alkylated product (**1.5a**–**c**).[50] Interestingly, these FgaPT2-catalyzed reactions exhibited substrate-dependent site-selectivity, with changes to the electronics and bulkiness of the alkyl group being donated conferring a switch in selectivity from the native C4 site of alkylation to a non-natural C5 alkylation. The divergent site-selectivity demonstrated by these reactions adds to the rapidly growing toolbox of site-selective alkylations, such as the C3 alkylations performed by the bacterial prenyltransferase PriB.[51] This recently discovered enzyme is highly promiscuous, possessing the ability to alkylate a broad range of indole (**1.6**) and naphthalene (**1.7**) derivatives.[51]

Overall, the relative promiscuity of these enzymes enables rapid substrate engineering through perturbation of the alkyl group being transferred and the substrate being alkylated, making these enzymes an attractive class of potentially robust biocatalysts.

Pyridoxal 5-phosphate (PLP)-dependent enzymes. PLP is a versatile cofactor present in enzymes that catalyze many transformations including transamination, epimerization, and decarboxylation. Despite this diversity, the shared driving force in all PLP-dependent mechanisms is

the "electron sink" character of the PLP cofactor as it uses its π-system to stabilize the negative charge that develops at its Cα position.[52] Several PLP-dependent enzymes have been identified to catalyze the formation of C–C bond forming reactions, including (1) aldol reactions catalyzed by threonine aldolase,[53] (2) the formation of alpha-amino ketones by a polyketide-like synthase,[54,55] and (3) the formation of tryptophan and tryptophan derivatives by tryptophan synthase.[56, 57]

In the catalytic mechanism of tryptophan synthase, a serine amino acid condenses onto the PLP cofactor and, upon dehydration, forms the highly electrophilic aminoacrylate intermediate species (**1.8**). Activation of this electrophilic cofactor allows for the nucleophilic attack by an indole to forge a new C–C bond, ultimately forming tryptophan (**1.9**). Tryptophan synthase is composed of an α subunit (TrpA) that acts as a regulatory subunit to the catalytic β subunit (TrpB), which binds PLP and catalyzes the C–C bond formation event.[58] Through directed evolution of the TrpB subunit, it was identified that the allosteric regulation provided by the TrpA subunit could be circumvented through a few key mutations in TrpB, increasing the biocatalytic utility of this enzyme.[59] Since this first discovery, TrpB has been engineered extensively for broader C–C bond forming reactions on substituted indoles (**1.10**), pyrroles (**1.11**), nitro alkanes (**1.12**), and oxindoles (**1.13**) to form quaternary bonds (Figure 1.1c).[59-64]

Halogenated electrophiles. Unlike the highly activated electrophilic cofactors described so far, an alkylating enzyme (CylK) was recently discovered that solely relies on a halogenated alkyl chain (**1.14**) as the electrophilic alkyl group in an intermolecular Friedel–Crafts alkylation.[65] The native reaction catalyzed by this alkylating enzyme is a dimerization reaction between two halogenated resorcinol derivatives to form **1.15**. In addition to this native reaction, CylK can catalyze the alkylation of a panel of resorcinol derivates (such as **1.16** and **1.17**) using truncated alkyl halides (**1.18**; Figure 1.1d).[66]

Intramolecular Friedel–Crafts alkylations. Intermolecular Friedel–Crafts alkylations occur routinely in primary and secondary metabolite pathways and can be catalyzed by some of the enzymes described so far; however, C–C bonds can also be formed in intramolecular Friedel–Crafts alkylations to forge a new ring, such as in the biosynthesis of many isoquinoline alkaloid and lignan natural products (Figure 1.2a,c).[67,68]

One critical enzyme in the biosynthesis of isoquinoline alkaloid natural products is norcoclaurine synthase (NCS; Figure 1.2a). NCS is a Pictet-Spenglerase that catalyzes the condensation of an aldehyde **1.20** onto an amine **1.19** to form the iminium intermediate **1.21**. This electrophilic

intermediate (**1.21**) is then poised for an intramolecular Friedel–Crafts alkylation by nucleophilic attack to forge a C–C bond with the phenolic ring A to form (*S*)-norcoclaurine (**1.22**).[69] Through the exploration NCSs for their broader biocatalytic utility for Pictet-Spengler reactions, the substrate scopes of these enzymes have been found to be relatively broad with the ability to accept a wide range of aldehyde derivates to form various tetrahydroisoquinoline products (**1.26**).[70] As such, these enzymes have been applied to the chemoenzymatic syntheses of a range of alkaloid natural products and natural product derivatives.[70] Additionally, a recent report demonstrated the ability of these enzymes to perform this ring closure with unactivated ketone coupling partners, thereby overcoming the limitation of coupling only to aldehydes.[71] Using ketone substrates, quaternary centers can be formed in chiral disubstituted- and spiro-tetrahydroisoquinolines (see **1.27** and **1.28**, respectively; Figure 1.2b).[71]

Modification of the NCS product, (*S*)-norcoclaurine (**1.22**), by several methyltransferases and a cytochrome P450 yields (*S*)-reticuline (**1.23**), a branchpoint in the biosynthesis of many isoquinoline alkaloids.[67, 72] Isoquinoline alkaloids encompass many pharmacologically valuable natural products such as morphine and berberine. As such, extensive research has focused on the metabolic engineering of microorganisms for the scalable production of these molecules.[73-77] The

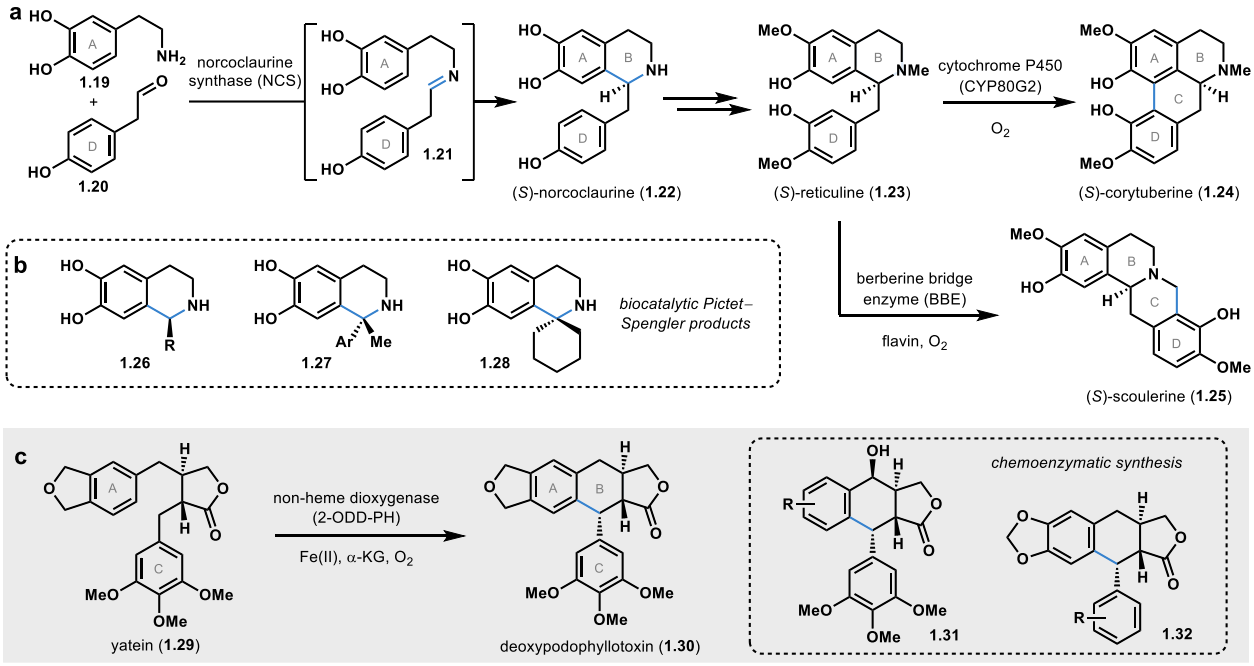


Figure 1.2. Biosynthetic and biocatalytic ring-closing reactions. (a) Several different enzymes catalyze intramolecular alkylation or oxidative coupling reactions to form a new ring in the biosynthesis of isoquinoline alkaloids. (b) Noroclaurine synthases (NCSs) have been utilized as biocatalytic Pictet-Spenglerases for the formation of a panel of products. (c) A non-heme dioxygenase catalyzes an intramolecular alkylation reaction to form a new ring in the biosynthesis of podophyllotoxin lignan products. This biocatalytic reaction has been applied to the chemoenzymatic of several natural product analogues.

in vivo production of valuable molecules can be beneficial, especially when the chemical syntheses are not amenable to large-scale synthesis or the enzymes involved in the pathway are challenging to work with *in vitro*.[78]

As a branchpoint in the biosynthesis of isoquinoline alkaloids, several different C–C bond-forming reactions can be performed on (*S*)-reticuline (**1.23**) for further diversification. In particular, a fourth ring can be formed either through an oxidative coupling reaction (described in more detail in the next section of this Review) to form (*S*)-corytuberine (**1.24**) or through another intramolecular Friedel–Crafts alkylation to form (*S*)-scoulerine (**1.25**; Figure 1.2a). The enzyme responsible for the Friedel–Crafts alkylation is the flavin-dependent berberine bridge enzyme (BBE). In all the examples described thus far, the first step before an alkylation event is the activation of an electrophilic carbon source; however, in the case of the BBE-catalyzed alkylation, the unactivated methyl group is directly attacked by the phenol in **1.23** (ring D).[79] This is made possible by the presence of a flavin cofactor in the active site of BBE which accepts a hydride from the methyl group being attacked in a single concerted step.[79] This enzyme has not been explored for use as a biocatalyst as extensively as NCS; however, it has been identified that the substituents on ring D in **1.23** can be modified with successful C–C bond formation. Additionally, the site-selectivity of the bond formation can be biased in cases where a C–F bond is present at a potential alkylation site.[80]

Beyond alkaloids, another intramolecular Friedel–Crafts alkylation was recently identified in the biosynthesis of podophyllotoxin lignans (Figure 1.2c).[81] In particular, a non-heme dioxygenase was identified to catalyze the formation of the final ring in this biosynthetic pathway to form deoxypodophyllotoxin (**1.30**).[81,82] Enzymes containing a non-heme iron are most commonly involved in hydroxylation events; however, this class of enzymes can also perform a myriad of other oxidative transformations.[83] Evaluation of the substrate scope of the C–C coupling non-heme dioxygenase, 2-ODD-PH, demonstrated its ability to perform oxidative C–C alkylations on a variety of derivatives of **1.29**, particularly with modifications to the substituents around rings A and C (Figure 1.2c).[84] This chemistry was used to perform late-stage C–C coupling reactions for the chemoenzymatic synthesis of products such as **1.31** and **1.32**.[85]

Friedel–Crafts acylations. Typically, acyl transfers in Nature are carrier protein-dependent processes that proceed through S_NAc-type reactions.[86] However, similarly to the Friedel–Crafts alkylations described in this section so far, one unique bacterial acyltransferase has been identified

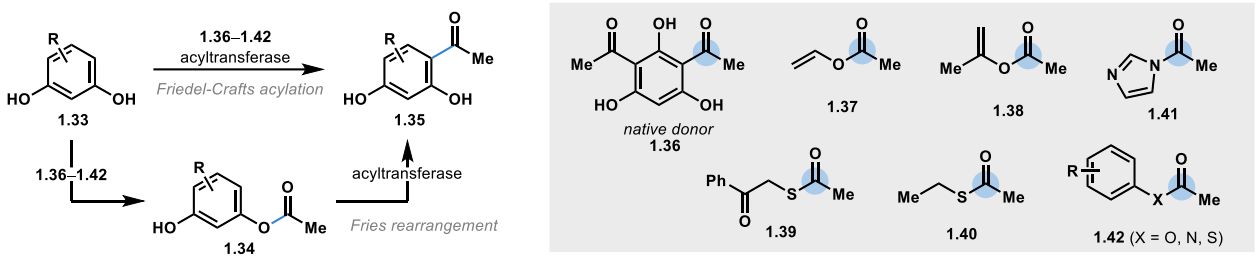


Figure 1.3. Site-selective acylation of resorcinol derivatives. In addition to the native acyl donor (**1.36**), various acyl donors (**1.37**–**1.42**) were successfully employed in a biocatalytic Friedel–Crafts acylation or Fries rearrangement-like reaction to acetylate resorcinol derivatives (**1.33**).

to catalyze a Friedel–Crafts acylation on resorcinol substrates (Figure 1.3).[87] This acyltransferase natively transfers an acetyl group from **1.36** to the resorcinol substrate (**1.33**), either by directly forming the C–C bond in the acetylated product (**1.35**), or through an ester intermediate (**1.34**).[88] In the latter mechanism, the acyltransferase can catalyze the biocatalytic equivalent of a Fries-rearrangement with an intramolecular acetyl transfer to form **1.35**.

Like the previously described intermolecular Friedel–Crafts reactions, both the acyl donor and acceptor could be modified to increase the versatility of these biocatalytic reactions. Specifically, a broad range of commercially available acetyl donors were productively utilized, including activated esters (**1.37**–**1.38**), thioesters (**1.39**–**1.40**), and amides (**1.41**–**1.42**) to the selectively acetylate various resorcinol derivates (**1.xx**).[88-90]

CYCLIZATION REACTIONS

Nature has mastered the construction of diverse scaffolds from common precursors by evolving catalysts capable of controlling the outcome of cyclization reactions. These reactions vary in complexity with some catalyzing the formation of a single C–C bond, whereas others can catalyze the stereoselective formation of a series of new C–C bonds in cascade cyclization reactions.

Terpene cyclases. Terpenes constitute the largest natural product family with over 80,000 members.[91] Despite the magnitude and structural diversity of this family, all these natural products stem from a similar five-carbon precursor unit. Most of the structural complexity of the products are installed in cascade cyclization reactions wherein multiple fused rings and stereocenters are installed by a single enzyme (such as in Figure 1.4a). The enzymes that catalyze these reactions are referred to as terpene cyclases (or terpene synthases). These enzymatic reactions are considered the most complex in Nature, with a typical cyclization reaction breaking or making new bonds at the majority of the carbon atoms comprising the backbone of the terpene precursor substrate.[91] The

complexity-generating nature of terpene cyclase-catalyzed reactions have influenced the biomimetic syntheses of terpenoid natural products, such as the use of Lewis acid-mediated opening of vinyl epoxides for the cascade polycyclizations of terpenoids, however the stereocontrol of the cyclization is still lacking in chemical methods.[92, 93]

Triterpene cyclases, such as the squalene-hopene cyclase (SHC), have been studied extensively for their structure and mechanism.[94] In the native SHC-catalyzed cascade cyclization of squalene (**1.43**) to form hopene (**1.44**), five new C–C bonds are forged to form five fused rings and ten new stereocenters (Figure 1.4a). It has been found that SHCs can catalyze a much wider variety of cyclization reactions;[95] examples of these include (1) stereoselective cascade cyclizations with smaller linear terpenes to form molecules with two or three fused rings,[96, 97] (2) Friedel–Crafts alkylations,[98] (3) Prins-type reactions,[99-101] and (4) isomerization reactions.[95] Additionally, these enzymes have demonstrated their amenability to evolution campaigns, often with a single amino acid change in the binding pocket of the enzyme resulting in a dramatic increase in activity

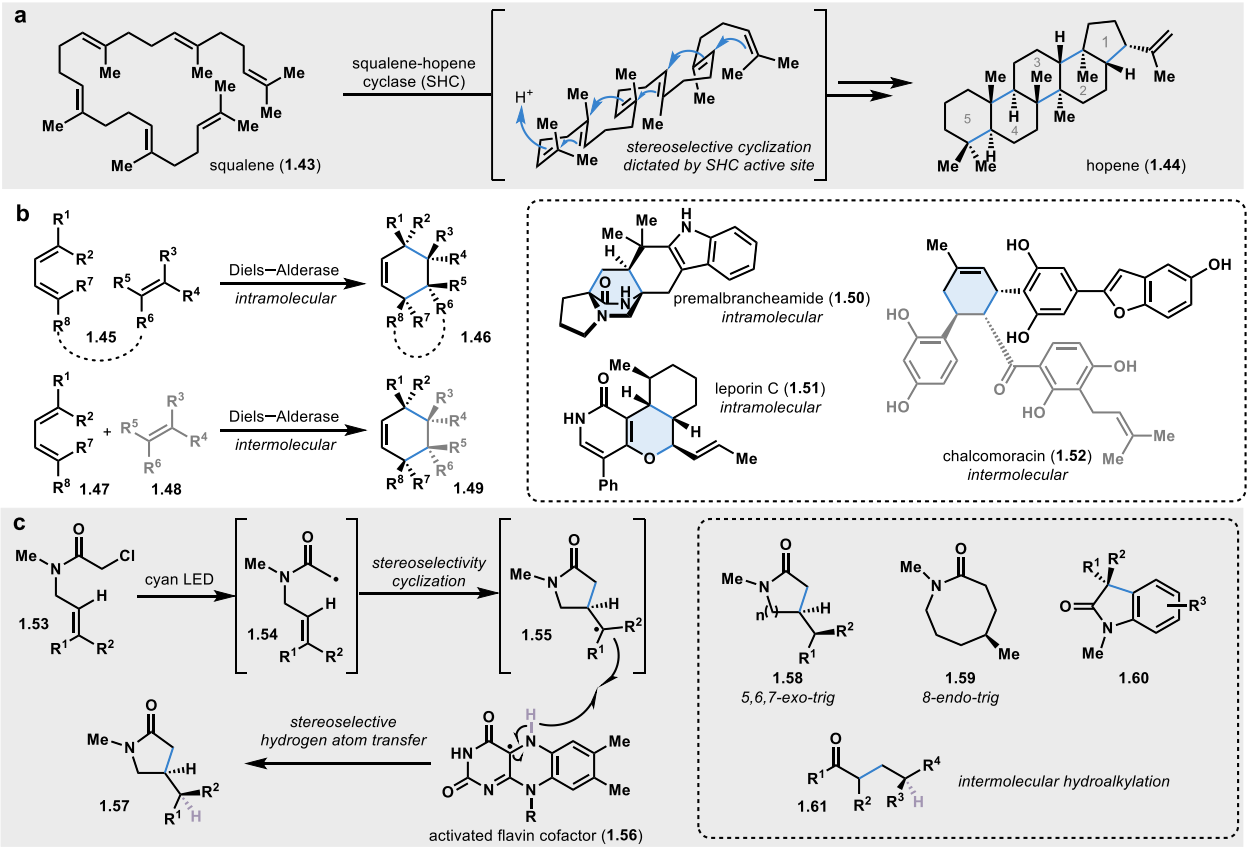


Figure 1.4. Cyclization reactions catalyzed by natural and engineered enzymes. (a) Terpene cyclases catalyze stereoselective cascade cyclization reactions for the generation of polycyclic products. (b) Diels–Alderases have been identified to catalyze intramolecular Diels–Alder reactions in secondary metabolite pathways and are hypothesized to be involved in intermolecular cycloaddition reactions. (c) Engineered flavin-dependent ene-reductases can catalyze non-natural photobiocatalytic radical cyclization reactions. Additionally, this photoenzymatic strategy for C–C bond formation has been extended to intermolecular hydroalkylation reactions.

with a non-native reaction.[95] However, this promiscuity can also lead to the formation of multiple product isomers from a single linear substrate due to the ability of the active site to accommodate multiple conformations of the carbocation skeleton during cyclization.[102] Despite this potential limitation of some terpene cyclases, engineered SHC variants have successfully been employed in chemoenzymatic syntheses, such as in the synthesis of (+)-ambrein, a high-commodity natural product that is difficult to isolate from natural sources.[103]

Diels–Alderases. In the list of classic, complexity-generating C–C bond forming reactions, the Diels–Alder reaction comes close to the top. This transformation has been central to the synthesis of six-membered rings in complex natural products and simple structures alike. This bread-and-butter reaction is embedded in the retrosynthetic logic chemists use to develop strategies toward target molecules, which has driven the development of variants of the Diels–Alder reaction that allow this cycloaddition to be applied intra- or intermolecularly and on substrates with a broad range of electronic properties (see **1.45** for intramolecular cycloaddition and **1.47**–**1.48** for intermolecular cycloaddition in Figure 1.4b). Despite the common occurrence of the cyclohexene motif in natural product cores, the first enzymatic Diels–Alder reactions were carried out with proteins engineered to this non-native reactivity by leveraging the rate enhancement possible through templating of a given diene and dienophile and the acid-base type activation possible in an active site.[104]

While a number of artificial Diels–Alderases have been developed, natural Diels–Alderases have proven to be a much more elusive target and have garnered a great deal of attention over the last several decades.[105] Elements of this controversy arise from the dual action of many "Diels–Alderases" as many enzymes given this label generate one of the Diels–Alder partners in addition to potentially catalyzing the cyclization event. This controversy has also been fueled by debate over what mechanistically constitutes a Diels–Alder cycloaddition, with strict definitions calling for concerted formation of two new C–C bonds and more liberal labelling of Diels–Alderases catalyzing these bond formations in a stepwise fashion.

Recently, crystal structures of several bifunctional enzymes were reported that catalyze intramolecular Diels–Alder reactions: (1) short-chain dehydrogenases MalC and PhqE in (+)-premalbrancheamide (**1.50**) biosynthesis[106] and (2) SAM-dependent *O*-methyltransferase LepI in leporin C (**1.51**) biosynthesis.[107, 108] Although these enzymes are homologous to functionally diverse enzyme classes, it appears that enzymes have evolved away from their canonical functions

to act as a scaffold for stereoselective Diels–Alder reactions. This possesses a challenge in assigning cyclase function based on sequence. In addition to Diels–Alderases catalyzing intramolecular cycloadditions, several intermolecular [4+2] cycloadditions have been proposed to be catalyzed by a Diels–Alderase, although few have been experimentally confirmed.[109] Only recently was the first stand-alone Diels–Alderase catalyzing an intermolecular reaction was identified in the biosynthesis of the plant natural product chalcomoracin (**1.52**).[110] This Diels–Alderase, MaDA, is a flavin-dependent enzyme performs a [4+2] cycloaddition through a concerted, asynchronous mechanism. Additionally, substrate scope analysis of the enzyme demonstrated promiscuous activity, with the ability to accept different combinations of unnatural dienes and dienophiles, ultimately allowing for the chemoenzymatic synthesis of various natural product derivatives.

Ene-reductases. Ene-reductases are a class of enzymes that natively catalyze the asymmetric reduction of activated alkenes. These flavin-dependent enzymes deliver a hydride to one face of the alkene substrate. Subsequent protonation of the reduced C–C bond by a conserved tyrosine residue reduces the alkene in a stereo-controlled fashion.[111] Ene-reductases are robust enzymes, typically with broad substrate scopes, high yields, and excellent selectivity, garnering them a great deal of attention as industrial biocatalysts.[112, 113]

Beyond their widely studied and utilized canonical function, several ene-reductases were recently employed in the reduction of other functional groups. Ene-reductase variants lacking the conserved tyrosine residue responsible for the protonation event can transfer a hydride to α,β-unsaturated aldehydes and ketones triggering a reductive cyclization event to form cyclopropanes.[114] Although these reactions only exhibit moderate stereoselectivity, these experiments represent an important proof-of-concept that the reactivity of these transformations could be controlled through protein and substrate engineering beyond the native reactivity observed for ene-reductases.

As well as the non-natural cyclopropanation reactions, several recent reports demonstrated the ability for ene-reductases to catalyze the radical cyclization of five- to eight-membered rings in photobiocatalytic reactions (Figure 1.4c).[115, 116] In contrast to the typical mechanism of ene-reductases, where hydride transfer from the flavin cofactor to the substrate initiates the reaction, the delivery of a hydrogen atom quenches a radical species such as **1.55** to form a cyclized product (**1.57**; Figure 1.4c).[117] In the first report of this photobiocatalytic reaction, a panel of chloroacetam-

ide derivatives (**1.53**) were exposed to cyan LEDs, causing the formation of a radical and subsequent radical cyclization and hydride transfer to form products such as **1.58** and **1.59**.[115] A second report demonstrated the effectiveness of this reaction on a panel of chloroacetamide substrates to form diverse oxindole products (**1.60**).[116] Additionally, beyond cyclization reactions, this photoenzymatic approach to stereoselective C(sp^3)–C(sp^3) bond formation has been applied to the hydroalkylation of alkenes in intermolecular radical coupling reactions, allowing for the versatile C–C coupling of various α-halo carbonyls and alkenes to from γ-stereogenic ketones, amides, and esters (see **1.61**; Figure 1.4c).[118-120]

In addition to engineering ene-reductases for a completely new type of reaction when combined with photocatalysis, engineered hemoproteins for carbene transfer reactions serve as an excellent case study in engineering enzymes for completely new-to-nature types of transformations.

CARBENE TRANSFER REACTIONS

The diversity of transformations achievable through carbene transfer including cyclopropanations, heteroatom-hydrogen (X–H) bond insertions, and rearrangements make carbene transformations valuable in synthetic chemistry.[121-126] Although carbene transfer is not known to occur in Nature, oxene transfer is commonly performed by cytochromes P450.[127] These enzymes harbor a heme cofactor (**1.62**) in their active sites, which upon reduction binds molecular oxygen to form a highly reactive iron-oxene (**1.63**) intermediate species.[127] This reactive iron-oxene species (**1.63**) typically performs a hydrogen atom abstraction, after which it can perform a myriad of oxidative transformations such as a rebound hydroxylation or insertion of oxygen into a double bond to form an epoxide (**1.65**; Figure 1.5a).[128, 129]

In 2013, Arnold and co-workers published the first of a series of reports describing the utilization of the bacterial P450 BM3 for carbene transfer.[130] Arnold discovered that replacement of molecular oxygen with a diazoacetate reagent (**1.66**) allowed for the formation of an iron-carbene species (**1.68**) in the P450 heme cofactor. In the presence of an olefin (**1.69**), carbene transfer could occur to form a cyclopropane (**1.70**) with the stereo-control provided by an enzyme active site, marking the first instance of a non-natural "carbene transferase" (Figure 1.5b).

Engineering carbene transferases. Following Arnold's initial report, several research groups have led the way in engineering P450 BM3 and other hemoproteins into carbene transferases for cyclopropanation of olefins, alkylation through insertion into C–H bonds, and C–C bond

formation through rearrangement reactions.[131] Engineering efforts have focused on mutagenesis and alterations to the heme cofactor and coordinating metal ion to modulate oxidation potential and overcome obstacles including (1) limited substrate scope, (2) access to a single isomer of the desired product, and (3) competition with the native iron-oxene formation.[131] The pursuit of this new-to-nature biocatalyst serves as an excellent example for how enzymes can be engineered into refined biocatalysts for a given reaction.

In 2013, Arnold's group identified that mutagenesis of the cysteine residue, which serves as the axial ligand to the P450 heme cofactor, to a serine increased the reduction potential of the resting state enzyme thereby abolishing monooxygenase activity and giving a preference to the formation of a carbene over an oxene.[132] This point mutation was then applied to this highly conserved residue in other P450s, forming a suite of enzymes dubbed "P411s" due to the change in their absorbance from 450 nm to 411 nm.[133] Additionally, nucleophilic residues in the active site

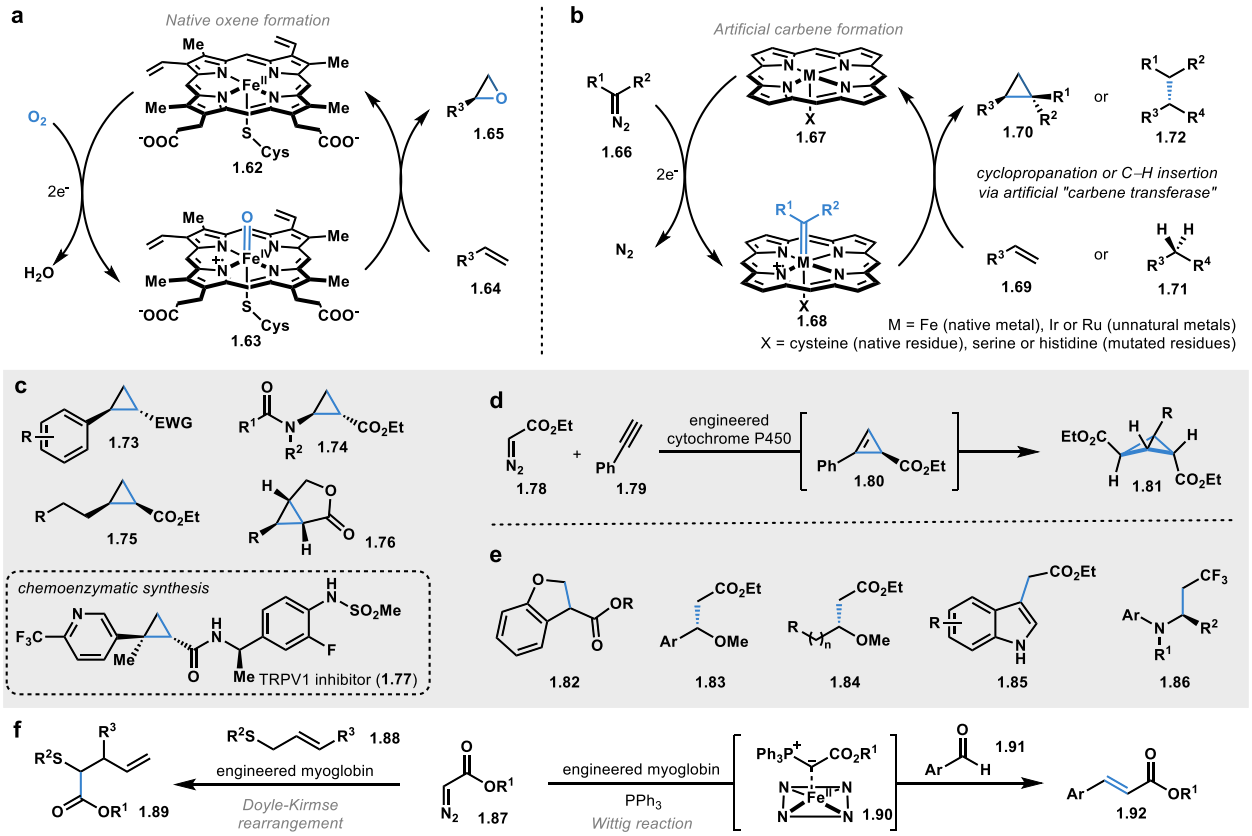


Figure 1.5. Engineering non-natural carbene transferases. (a) Natural iron-oxene formation in the heme cofactor of hemoproteins. (b) Engineering artificial carbene-transferases through iron-carbene formation. (c) Scope of cyclopropanation reactions catalyzed by carbene transferases and representative use in chemoenzymatic synthesis. (d) Carbene transfer into alkyne forms highly strained cyclopropene and bicyclobutane products. (e) Carbene insertion into C–H bonds forms alkylated products. (f) Carbene transferases have catalyzed C–C bond formation through rearrangement reactions and Wittig reactions.

causing mechanism-based inhibition were identified and removed to further promote carbene transfer to a desired substrate.[134] The Fasan group ventured beyond P450s to other hemoproteins and successfully engineered the first myoglobin capable of carbene transfer.[135] For the myoglobin-derived carbene transferases, modification of the heme cofactor through installation of alternative substitution patterns increased the efficiency of iron-carbene formation and enabled reactions to be performed aerobically without inhibition of carbene transfer by oxygen.[136]

Olefin cyclopropanation. Over the last five years, hemoproteins have been engineered for modified selectivity and substrate scope of olefin cyclopropanations performed by carbene transferases through directed evolution and rational mutagenesis of the active site.[131] These engineering efforts have provided access to stereo-divergent cyclopropane products with control over both the diastereo- and enantioselectivity in carbene transfer reactions.

The substrate scope of these reactions has moved beyond phenyl-substituted olefins (that afford **1.73** as a general product) to form cyclopropanes substituted with heteroatoms (**1.74**) or aliphatic chains (**1.75**) to access more diverse cyclopropanes (Figure 1.5c).[137-140] Furthermore, the electron withdrawing group on the diazo reagent could be modified from an ester moiety to a nitrile or trifluoromethyl substituent to form the corresponding cyclopropanes (**1.73**).[141, 142] Additionally, these cyclopropanations can occur intramolecularly to form products such as **1.76**.[143] Recently, this cyclopropanation chemistry was extended beyond alkenes to alkynes (**1.79**), in which a strained cyclopropene (**1.80**) was generated.[144] Cyclopropene products (**1.80**) are primed to undergo a second carbene transfer to form a bicyclobutane product (**1.81**) or be derivatized with chemical methods (Figure 1.5d).[144]

These reactions are generally performed in whole *E. coli* cells, requiring no purification of the active catalysts. Additionally, these reactions can typically be performed on gram-scale and have been applied to the syntheses of several cyclopropane-containing natural products and pharmacologically relevant molecules, such as the TRPV1 inhibitor (**1.77**).[145, 146]

Carbene insertion into C–H bonds. Beyond cyclopropanation, these engineered hemoproteins can perform insertion into C–H bonds to form various types of products (Figure 1.5e). In 2016, Hartwig and co-workers found that replacing the iron in the P450 or myoglobin heme with iridium promoted carbene insertion into C–H bonds to form products such as **1.82**.[147, 148] This demonstrated the ability to alter the reactivity of an enzyme by altering the catalytic metal ion without the need for extensive evolution. Shortly after these initial findings, the Fasan group

demonstrated that several first-row transitions metals such as cobalt and manganese could inflict similar changes in reactivity to enable C–H insertions.[149]

More recently, several groups have demonstrated the ability to perform sp^3 C–H alkylations on unactivated substrates using evolved P411 or myoglobin variants containing the native iron-heme cofactor.[150] These C–H insertion reactions can be performed on a wide range of substrates (such as **1.83** and **1.84**). Further evolution broadened the substrate scope even more to alkylate unprotected indoles (**1.85**)[151, 152] and *N*-aryl pyrrolidine substrates (**1.86**).[153] For the latter substrates, trifluoroethyl groups were installed through insertion into α-amino C–H bonds.[153] These reactions can either be performed with purified enzyme or in whole *E. coli* cells expressing the active catalyst as described for the cyclopropanation reactions. Additionally, the scalability and relevance of hemoprotein-catalyzed C–H insertions were demonstrated in several chemoenzymatic syntheses, such as in the synthesis of the anti-inflammatory drug indomethacin[151] and enantio-complementary alkaloids.[150]

Other reactions catalyzed by carbene transferases. Although cyclopropanations and C–H insertions have been the most well-established C–C bond-forming reactions catalyzed by engineered carbene transferases, extensive substrate and protein engineering have resulted in the discovery of other reactivity (Figure 1.5f). For example, exposure of the diazoester reagent (**1.87**) and an allylic sulfide (**1.88**) to an engineered myoglobin led to the formation of a sulfur ylide intermediate that then underwent a [2,3]-sigmatropic rearrangement to form **1.89**.[154] Alternatively, exposure of the diazoester reagent (**1.87**) and an engineered myoglobin to triphenylphosphine promoted the formation of the proposed intermediate **1.90**, which, when in the presence of an aldehyde (**1.91**) could undergo a Wittig reaction to form olefin **1.92**.[155, 156]

OXIDATIVE C–C COUPLING REACTIONS

Oxidative coupling reactions perhaps offer the most efficient approach to C–C bond formation as two building blocks are directly stitched together.[157] However, a major limitation in current synthetic methods for oxidative coupling lies in the control over the chemo-, site-, and stereochemical outcome of the reaction, particularly in the case of intermolecular coupling reactions.[158, 159] Nature has evolved enzymes that can control the oxidative dimerization of phenolic substrates with precise control of the site- and stereoselectivity of bond formation. These enzymes include cyto-

chromes P450, laccases, and peroxidases. These three classes of enzymes share the ability to initiate radical mechanisms on phenolic substrates through the abstraction of a hydrogen atom from the phenol. However, the mechanism of hydrogen abstraction and extent of control exerted over the bond formation event in intermolecular and intramolecular reactions vary among these enzymes.

Intramolecular C–C coupling. Cytochromes P450 catalyze a wide range of oxidative transformations leveraging their iron-heme cofactor.[128] Often, these oxidative transformations involve a hydrogen atom abstraction from the substrate, followed by oxygenation through a rebound hydroxylation.[127] However, the mechanism of these oxidative transformations can diverge following initial hydrogen atom abstraction, including C–C bond formation.[128] The C–C bond formed through an oxidative coupling event in a number of natural products have been attributed to P450-catalyzed reactions.[160]

Many of these P450-catalyzed C–C coupling reactions occur intramolecularly, such as in the biosynthesis of (*S*)-corytuberine (**1.24**).[161] In this reaction, cytochrome P450 80G2 initiates the oxidative mechanism on the substrate **1.23** with a hydrogen atom abstraction from one of the phenols. Instead of the canonical rebound hydroxylation typically performed by these enzymes, a C–C bond is formed either through a radical-radical coupling or by radical addition to a second phenolic group to form **1.24** (Figure 1.6a). Another key intramolecular C–C coupling catalyzed by a P450, is in the biosynthesis of the glycopeptide antibiotic vancomycin.[162] Upon identification of the P450 OxyC responsible for catalyzing this reaction, it was applied for the *in vitro* chemoenzymatic synthesis of vancomycin derivatives.[163]

Intermolecular C–C coupling. In addition to intramolecular C–C coupling, oxidative coupling reactions can occur in Nature between two phenolic or indole-containing substrates in an intermolecular fashion. These reactions are key steps in the biosynthesis of many dimeric biaryl secondary metabolites and can be catalyzed by P450s, laccases, or peroxidases (Figure 1.6).

One approach to oxidative dimerization in secondary metabolism involves enzymatic oxidation followed by rapid binding of the oxidized substrates to an auxiliary protein, wherein the templating of the two substrates in a specific orientation allows for a stereoselective radical–radical coupling reaction.[164] These oxidations are typically performed by laccases, a class of multicopper oxidases that readily oxidize diverse phenolic substrates.[165] Laccase-catalyzed oxidations are ubiquitous in fungal and plant systems, such as in the biosynthesis of lignin, and have been harnessed

for a variety of biotechnological applications.[166] However, the application of laccases in asymmetric catalysis has been limited due to the scarcity of examples in which laccases exert control over the bond formation event following the initial oxidation.[167] To overcome this limitation, auxiliary dirigent proteins can be harnessed to regain control over the selectivity of the bond formation event.[164, 168, 169] For example, in the convergent biosynthesis of viriditoxin (**1.93**), a late-stage oxidative dimerization catalyzed by a laccase and dirigent protein forms (*P*)-**1.93** with excellent atroposelectivity (>95:5 er).[169] In contrast, removal of the dirigent protein from the oxidative dimerization results in a complete loss in atroposelectivity, with the formation of a mixture of (*P*)-**1.93** and (*M*)-**1.93** in a 1:2 ratio.[169]

In addition to laccases, cytochrome P450 enzymes (P450s) have been identified as convergent catalysts in the biosynthesis of numerous dimeric natural products.[170] However, unlike most laccases, these P450s exert control over both the oxidation and selectivity of the bond formation event.[171] This catalyst-controlled selectivity has been demonstrated in the biosynthesis of an ever-expanding catalog of biaryl products from fungi and bacteria, with P450s catalyzing the selective dimerization of highly oxygenated tricyclic, coumarin, and naphthalene substrates (such as **1.94**–**1.97**).[172-175] In addition to the oxidative coupling of phenolic substrates, P450s have also been implicated in the oxidative coupling of substrates harboring indoles in the biosynthesis of alkaloids (such as **1.98**–**1.100**).[176-178]

Unlike many of Nature's biosynthetic strategies for biaryl bond formation, traditional synthetic strategies to construct biaryl bonds face fundamental challenges in controlling the selectivity

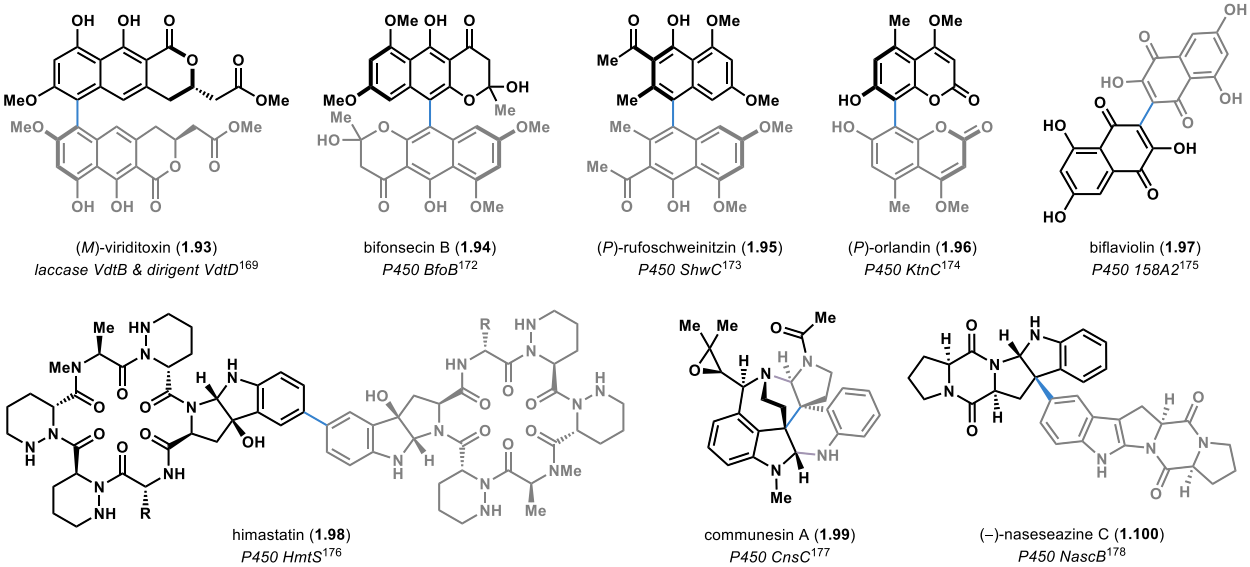


Figure 1.6. Intermolecular oxidative C–C coupling reactions in Nature.

of bond formation. For example, metal-catalyzed cross-couplings such as Suzuki and Negishi reactions require the prefunctionalization of both fragments to dictate the site of bond formation, thereby adding steps to a given synthetic campaign.[179] Alternatively, the direct oxidative coupling of two C–H bonds drastically increases the efficiency of a convergent synthetic sequence, albeit often at the expense of selectivity in the bond-forming event.[157] The powerful advantage of catalyst control offered by enzymes in these transformations positions P450s as a promising, albeit currently underexplored, class of enzymes for biocatalytic oxidative C–C bond formation.

THESIS OVERVIEW

Enzymes offer an attractive alternative to many small molecule-mediated transformations due to the high levels of chemo-, site-, and stereoselectivity they impart in their reactions. However, a major obstacle to the routine incorporation of biocatalysts in organic synthesis currently lies in the relative lack of diversity and availability of biocatalysts for a given transformation. Despite the surge in the development of new biocatalytic methods over the last several decades, additional biocatalysts need to be developed to increase the library of biocatalysts, particularly for C–C bond formation. The following chapters present the development of biocatalytic tools for biaryl C–C bond formation through cytochrome P450-catalyzed oxidative coupling reactions (Figure 1.7).

In Chapter II, the development of a biocatalytic platform for fungal P450s in yeast is outlined, motivated by the growing number of fungal P450s that have been identified to catalyze intermolecular oxidative dimerization reactions to forge biaryl bonds with catalyst-controlled site-selectivity. To harness this powerful chemistry, I developed an *in vivo* biocatalytic platform for oxidative coupling reactions catalyzed by KtnC, a fungal P450 that catalyzes a highly selective oxidative dimerization reaction in the biosynthesis of a bicoumarin in *Aspergillus niger*. Moreover,

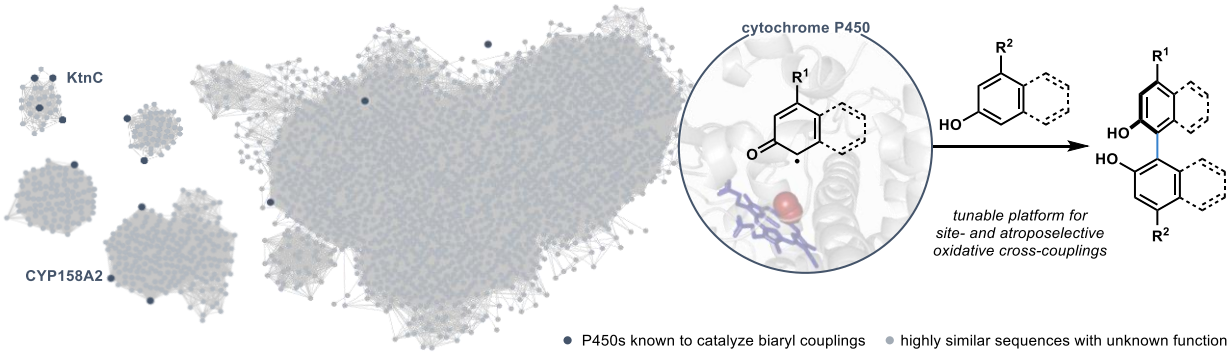


Figure 1.7. Dissertation overview.

engineering a redox self-sufficient fusion enzyme further improves the efficiency of the system, allowing for both high-throughput screening and preparative scale reactions.

In Chapter III, the *in vivo* biocatalytic platform was applied to engineer KtnC into a versatile catalyst for cross-coupling reactions. Initial substrate scope analysis performed by Dr. Jessica Yazarians demonstrated that wild-type KtnC could catalyze a range of oxidative cross-coupling reactions outside of its native dimerization reaction, however reaction yields and selectivity significantly decreased as substrate structures diverged further from the native coumarin scaffold. To overcome these limitations, I developed a platform for the directed evolution of P450s in yeast employing scanning site-saturation mutagenesis of active site residues and high-throughput mass spectrometry. I applied this platform to improve the activity, site-selectivity, and atroposelectivity in an unnatural cross-coupling reaction, demonstrating the tunability of these catalysts for a target transformation. In addition to the target cross-coupling transformation, the engineered catalysts displayed broader cross-coupling capabilities than the initial wild-type enzyme, ultimately providing a general tool for coumarin–naphthol cross-coupling reactions.

In Chapter IV, to expand beyond biocatalytic cross-coupling reactions with coumarin, I began investigating the application of bacterial P450s in naphthol–naphthol cross-coupling reactions. Using bioinformatic analysis of bacterial P450 sequence space, I generated a library of natural P450 sequences engineered as P450-RhFRed fusions with potential for cross-coupling chemistry. By profiling various cross-coupling reactions with P450s, I identified that several enzymes could catalyzed the cross-couplings of a diverse set of phenolic substrates with substituted naphthols. Not only did this demonstrate a complementary reactivity to KtnC, but it also provided a promising starting point for engineering biocatalysts for atroposelective cross-coupling reactions to provide access to privileged biaryl ligands for asymmetric catalysis. Using the platform, I already developed for directed evolution of P450s, I engineered CYP158A2, a bacterial P450 from *Streptomyces coelicolor*, for improved yield and atroposelectivity in the model cross-coupling between 2-naphthol and 3-bromo-2-naphthol.

Finally, in Chapter V, I contextualize this work in the greater field of organic chemistry and biocatalysis, offering insights into the advantages biocatalysis may offer to convergent synthesis.

REFERENCES

1. Drew, S. W.; Demain, A. L. Effect of Primary Metabolites on Secondary Metabolism. *Annu. Rev. Microbiol.* **1977**, *31*, 343–356.

2. Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why Are Secondary Metabolites (Natural Products) Biosynthesized? *J. Nat. Prod.* **1989**, *52*, 1189–1208.

3. Maplestone, R. A.; Stone, M. J.; Williams, D. H. The Evolutionary Role of Secondary Metabolites: A Review. *Gene* **1992**, *115*, 151–157.

4. Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J. Extending the Application of Biocatalysis to Meet the Challenges of Drug Development. *Nat. Rev. Chem.* **2018**, *2*, 409–421.

5. Campos, K. R.; Coleman, P. J.; Alvarez, J. C.; Dreher, S. D.; Garbaccio, R. M.; Terrett, N. K.; Tillyer, R. D.; Truppo, M. D.; Parmee, E. R. The Importance of Synthetic Chemistry in the Pharmaceutical Industry. *Science* **2019**, *363*, eaat0805.

6. 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*, 185–194.

7. Turner, N. J.; O'Reilly, E. Biocatalytic Retrosynthesis. *Nat. Chem. Biol.* **2013**, *9*, 285–288.

8. Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118*, 801–838.

9. Hönig, M.; Sondermann, P.; Turner, N. J.; Carreira, E. M. Enantioselective Chemo- and Biocatalysis: Partners in Retrosynthesis. *Angew. Chem. Int. Ed.* **2017**, *56*, 8942–8973.

10. Schoemaker, H. E.; Mink, D.; Wubbolts, M. G. Dispelling the Myths: Biocatalysis in Industrial Synthesis. *Science* **2003**, *299*, 1694–1697.

11. Hult, K.; Berglund, P. Enzyme Promiscuity: Mechanism and Applications. *Trends Biotechnol.* **2007**, *25*, 231–238.

12. Brustad, E. M.; Arnold, F. H. Optimizing Non-Natural Protein Function with Directed Evolution. *Curr. Opin. Chem. Biol.* **2011**, *15*, 201–210.

13. Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16*, 379–394.

14. Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem. Int. Ed.* **2015**, *54*, 3351–3367.

15. Turner, N. J. Directed Evolution Drives the Next Generation of Biocatalysts. *Nat. Chem. Biol.* **2009**, *5*, 567–573.

16. Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W. Improving Catalytic Function by Prosar-Driven Enzyme Evolution. *Nat. Biotechnol.* **2007**, *25*, 338–344.

17. Fox, R. J.; Huisman, G. W. Enzyme Optimization: Moving from Blind Evolution to Statistical Exploration of Sequence-Function Space. *Trends Biotechnol.* **2008**, *26*, 132–138.

18. Lutz, S. Beyond Directed Evolution: Semi-Rational Protein Engineering and Design. *Curr. Opin. Biotechnol.* **2010**, *21*, 734–743.

19. Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. Synthetic Biology for the Directed Evolution of Protein Biocatalysts: Navigating Sequence Space Intelligently. *Chem. Soc. Rev.* **2015**, *44*, 1172–1239.

20. Yang, K. K.; Wu, Z.; Arnold, F. H. Machine-Learning-Guided Directed Evolution for Protein Engineering. *Nat. Methods* **2019**, *16*, 687–694.

21. Brovetto, M.; Gamenara, D.; Saenz Méndez, P.; Seoane, G. A. C−C Bond-Forming Lyases in Organic Synthesis. *Chem. Rev.* **2011**, *111*, 4346–4403.

22. Fessner, W.-D. Enzyme Mediated C–C Bond Formation. *Curr. Opin. Chem. Biol.* **1998**, *2*, 85–97.

23. Machajewski, T. D.; Wong, C.-H. The Catalytic Asymmetric Aldol Reaction. *Angew. Chem. Int. Ed.* **2000**, *39*, 1352–1375.

24. Schmidt, N. G.; Eger, E.; Kroutil, W. Building Bridges: Biocatalytic C–C-Bond Formation toward Multifunctional Products. *ACS Catal.* **2016**, *6*, 4286–4311.

25. Pohl, M.; Sprenger, G. A.; Müller, M. A New Perspective on Thiamine Catalysis. *Curr. Opin. Biotechnol.* **2004**, *15*, 335–342.

26. Müller, M.; Gocke, D.; Pohl, M. Thiamin Diphosphate in Biological Chemistry: Exploitation of Diverse Thiamin Diphosphate-Dependent Enzymes for Asymmetric Chemoenzymatic Synthesis. *FEBS J.* **2009**, *276*, 2894–2904.

27. Pohl, M.; Lingen, B.; Müller, M. Thiamin Diphosphate-Dependent Enzymes: New Aspects of Asymmetric C-C Bond Formation. *Chem. Eur. J.* **2002**, *8*, 5288–5295.

28. Effenberger, F.; Förster, S.; Wajant, H. Hydroxynitrile Lyases in Stereoselective Catalysis. *Curr. Opin. Biotechnol.* **2000**, *11*, 532–539.

29. Griengl, H.; Schwab, H.; Fechter, M. The Synthesis of Chiral Cyanohydrins by Oxynitrilases. *Trends Biotechnol.* **2000**, *18*, 252–256.

30. Sukumaran, J.; Hanefeld, U. Enantioselective C–C Bond Synthesis Catalysed by Enzymes. *Chem. Soc. Rev.* **2005**, *34*, 530–542.

31. Schallmey, A.; Schallmey, M. Recent Advances on Halohydrin Dehalogenases - From Enzyme Identification to Novel Biocatalytic Applications. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 7827–7839.

32. Miao, Y.; Rahimi, M.; Geertsema, E. M.; Poelarends, G. J. Recent Developments in Enzyme Promiscuity for Carbon–Carbon Bond-Forming Reactions. *Curr. Opin. Chem. Biol.* **2015**, *25*, 115–123.

33. Resch, V.; Schrittwieser, J. H.; Siirola, E.; Kroutil, W. Novel Carbon-Carbon Bond Formations for Biocatalysis. *Curr. Opin. Biotechnol.* **2011**, *22*, 793–799.

34. Cantoni, G. L. Biological Methylation: Selected Aspects. *Annu. Rev. Biochem.* **1975**, *44*, 435–451.

35. Chiang, P. K.; Gordon, R. K.; Tal, J.; Zeng, G. C.; Doctor, B. P.; Pardhasaradhi, K.; McCann, P. P. *S*-Adenosylmethionine and Methylation. *FASEB J.* **1996**, *10*, 471–480.

36. Fontecave, M.; Atta, M.; Mulliez, E. *S*-Adenosylmethionine: Nothing Goes to Waste. *Trends Biochem. Sci.* **2004**, *29*, 243–249.

37. Lin, H. *S*-Adenosylmethionine-Dependent Alkylation Reactions: When Are Radical Reactions Used? *Bioorg. Chem.* **2011**, *39*, 161–170.

38. Yokoyama, K.; Lilla, E. A. C–C Bond Forming Radical SAM Enzymes Involved in the Construction of Carbon Skeletons of Cofactors and Natural Products. *Nat. Prod. Rep.* **2018**, *35*, 660–694.

39. Schubert, H. L.; Blumenthal, R. M.; Cheng, X. Many Paths to Methyltransfer: A Chronicle of Convergence. *Trends Biochem. Sci.* **2003**, *28*, 329–335.

40. Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *S*-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **2012**, *13*, 2642–2655.

41. Klimasauskas, S.; Weinhold, E. A New Tool for Biotechnology: AdoMet-Dependent Methyltransferases. *Trends Biotechnol.* **2007**, *25*, 99–104.

42. Dalhoff, C.; Lukinavičius, G.; Klimašauskas, S.; Weinhold, E. Direct Transfer of Extended Groups from Synthetic Cofactors by DNA Methyltransferases. *Nat. Chem. Biol.* **2006**, *2*, 31–32.

43. Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M. Expanding the Chemical Scope of RNA: Methyltransferases to Site-Specific Alkynylation of RNA for Click Labeling. *Nucleic Acids Res.* **2010**, *39*, 1943–1952.

44. Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T.; Duncan, K. E.; Litchfield, D. W.; Lüscher, B.; Weinhold, E. Enzymatic Site-Specific Functionalization of Protein Methyltransferase Substrates with Alkynes for Click Labeling. *Angew. Chem. Int. Ed.* **2010**, *49*, 5170–5173.

45. Deen, J.; Vranken, C.; Leen, V.; Neely, R. K.; Janssen, K. P. F.; Hofkens, J. Methyltransferase-Directed Labeling of Biomolecules and Its Applications. *Angew. Chem. Int. Ed.* **2017**, *56*, 5182–5200.

46. 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.* **2009**, *48*, 9546–9548.

47. Liang, P. H.; Ko, T. P.; Wang, A. H. Structure, Mechanism and Function of Prenyltransferases. *Eur. J. Biochem.* **2002**, *269*, 3339–3354.

48. Tanner, M. E. Mechanistic Studies on the Indole Prenyltransferases. *Nat. Prod. Rep.* **2015**, *32*, 88–101.

49. Epifano, F.; Genovese, S.; Menghini, L.; Curini, M. Chemistry and Pharmacology of Oxyprenylated Secondary Plant Metabolites. *Phytochem.* **2007**, *68*, 939–953.

50. 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. *MedChemComm* **2019**, *10*, 1465–1475.

51. Elshahawi, S. I.; Cao, H.; Shaaban, K. A.; Ponomareva, L. V.; Subramanian, T.; Farman, M. L.; Spielmann, H. P.; Phillips, G. N., Jr.; Thorson, J. S.; Singh, S. Structure and Specificity of a Permissive Bacterial C-Prenyltransferase. *Nat. Chem. Biol.* **2017**, *13*, 366–368.

52. Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **2004**, *73*, 383–415.

53. Fesko, K. Threonine Aldolases: Perspectives in Engineering and Screening the Enzymes with Enhanced Substrate and Stereo Specificities. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 2579–2590.

54. Chun, S. W.; Narayan, A. R. H. Biocatalytic Synthesis of α-Amino Ketones. *Synlett* **2019**, *30*, 1269–1274.

55. Chun, S. W.; Hinze, M. E.; Skiba, M. A.; Narayan, A. R. H. Chemistry of a Unique Polyketide-Like Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 2430–2433.

56. Miles, E. W. Tryptophan Synthase: A Multienzyme Complex with an Intramolecular Tunnel. *Chem. Rec.* **2001**, *1*, 140–151.

57. Francis, D.; Winn, M.; Latham, J.; Greaney, M. F.; Micklefield, J. An Engineered Tryptophan Synthase Opens New Enzymatic Pathways to β-Methyltryptophan and Derivatives. *ChemBioChem* **2017**, *18*, 382–386.

58. Dunn, M. F. Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienzyme Complex. *Arch. Biochem. Biophys.* **2012**, *519*, 154–166.

59. Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. Directed Evolution of the Tryptophan Synthase β-Subunit for Stand-Alone Function Recapitulates Allosteric Activation. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 14599–14604.

60. Boville, C. E.; Scheele, R. A.; Koch, P.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. Engineered Biosynthesis of β-Alkyl Tryptophan Analogues. *Angew. Chem. Int. Ed.* **2018**, *57*, 14764–14768.

61. Boville, C. E.; Romney, D. K.; Almhjell, P. J.; Sieben, M.; Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga Maritima*. *J. Org. Chem.* **2018**, *83*, 7447–7452.

62. Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations That Mimic Allosteric Activation. *Angew. Chem. Int. Ed.* **2016**, *55*, 11577–11581.

63. Dick, M.; Sarai, N. S.; Martynowycz, M. W.; Gonen, T.; Arnold, F. H. Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. *J. Am. Chem. Soc.* **2019**, 19817–19822.

64. Romney, D. K.; Murciano-Calles, J.; Wehrmüller, J. E.; Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **2017**, *139*, 10769–10776.

65. Nakamura, H.; Schultz, E. E.; Balskus, E. P. A New Strategy for Aromatic Ring Alkylation in Cylindrocyclophane Biosynthesis. *Nat. Chem. Biol.* **2017**, *13*, 916–921.

66. 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.* **2019**, *58*, 3151–3155.

67. Facchini, P. J. Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2001**, *52*, 29–66.

68. Suzuki, S.; Umezawa, T. Biosynthesis of Lignans and Norlignans. *J. Wood Sci.* **2007**, *53*, 273–284.

69. Sheng, X.; Himo, F. Enzymatic Pictet–Spengler Reaction: Computational Study of the Mechanism and Enantioselectivity of Norcoclaurine Synthase. *J. Am. Chem. Soc.* **2019**, *141*, 11230–11238.

70. Patil, M. D.; Grogan, G.; Yun, H. Biocatalyzed C−C Bond Formation for the Production of Alkaloids. *ChemCatChem* **2018**, *10*, 4783–4804.

71. Lichman, B. R.; Zhao, J.; Hailes, H. C.; Ward, J. M. Enzyme Catalysed Pictet–Spengler Formation of Chiral 1,1′-Disubstituted- and Spiro-Tetrahydroisoquinolines. *Nat. Commun.* **2017**, *8*, 14883.

72. Sato, F.; Inai, K.; Hashimoto, T., Metabolic Engineering in Alkaloid Biosynthesis: Case Studies in Tyrosine- and Putrescine-Derived Alkaloids. In *Applications of Plant Metabolic Engineering*, Verpoorte, R.; Alfermann, A. W.; Johnson, T. S., Eds. Springer Netherlands: Dordrecht, 2007; pp 145–173.

73. Leonard, E.; Runguphan, W.; O'Connor, S.; Prather, K. J. Opportunities in Metabolic Engineering to Facilitate Scalable Alkaloid Production. *Nat. Chem. Biol.* **2009**, *5*, 292–300.

74. Galanie, S.; Thodey, K.; Trenchard, I. J.; Filsinger Interrante, M.; Smolke, C. D. Complete Biosynthesis of Opioids in Yeast. *Science* **2015**, *349*, 1095–1103.

75. Diamond, A.; Desgagne-Penix, I. Metabolic Engineering for the Production of Plant Isoquinoline Alkaloids. *Plant Biotechnol. J.* **2016**, *14*, 1319–1328.

76. Thodey, K.; Galanie, S.; Smolke, C. D. A Microbial Biomanufacturing Platform for Natural and Semisynthetic Opioids. *Nat. Chem. Biol.* **2014**, *10*, 837–844.

77. Marienhagen, J.; Bott, M. Metabolic Engineering of Microorganisms for the Synthesis of Plant Natural Products. *J. Biotechnol.* **2013**, *163*, 166–178.

78. Du, J.; Shao, Z.; Zhao, H. Engineering Microbial Factories for Synthesis of Value-Added Products. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 873–890.

79. Winkler, A.; Łyskowski, A.; Riedl, S.; Puhl, M.; Kutchan, T. M.; Macheroux, P.; Gruber, K. A Concerted Mechanism for Berberine Bridge Enzyme. *Nat. Chem. Biol.* **2008**, *4*, 739–741.

80. Resch, V.; Lechner, H.; Schrittwieser, J. H.; Wallner, S.; Gruber, K.; Macheroux, P.; Kroutil, W. Inverting the Regioselectivity of the Berberine Bridge Enzyme by Employing Customized Fluorine-Containing Substrates. *Chemistry* **2012**, *18*, 13173–13179.

81. Lau, W.; Sattely, E. S. Six Enzymes from Mayapple That Complete the Biosynthetic Pathway to the Etoposide Aglycone. *Science* **2015**, *349*, 1224–1228.

82. Chang, W.-C.; Yang, Z.-J.; Tu, Y.-H.; Chien, T.-C. Reaction Mechanism of a Non-Heme Iron Enzyme Catalyzed Oxidative Cyclization via C–C Bond Formation. *Org. Lett.* **2019**, *21*, 228–232.

83. Martinez, S.; Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-Dependent Oxygenases. *J. Biol. Chem.* **2015**, *290*, 20702–20711.

84. Lazzarotto, M.; Hammerer, L.; Hetmann, M.; Borg, A.; Schmermund, L.; Steiner, L.; Hartmann, P.; Belaj, F.; Kroutil, W.; Gruber, K.; Fuchs, M. Chemoenzymatic Total Synthesis of Deoxy-, Epi-, and Podophyllotoxin and a Biocatalytic Kinetic Resolution of Dibenzylbutyrolactones. *Angew. Chem. Int. Ed.* **2019**, *58*, 8226–8230.

85. Li, J.; Zhang, X.; Renata, H. Asymmetric Chemoenzymatic Synthesis of (−)-Podophyllotoxin and Related Aryltetralin Lignans. *Angew. Chem. Int. Ed.* **2019**, *58*, 11657–11660.

86. Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat. Prod. Rep.* **2003**, *20*, 79–110.

87. Hayashi, A.; Saitou, H.; Mori, T.; Matano, I.; Sugisaki, H.; Maruyama, K. Molecular and Catalytic Properties of Monoacetylphloroglucinol Acetyltransferase from *Pseudomonas* sp. YGJ3. *Biosci. Biotechnol. Biochem.* **2012**, *76*, 559–566.

88. Schmidt, N. G.; Pavkov-Keller, T.; Richter, N.; Wiltschi, B.; Gruber, K.; Kroutil, W. Biocatalytic Friedel–Crafts Acylation and Fries Reaction. *Angew. Chem. Int. Ed.* **2017**, *56*, 7615–7619.

89. Schmidt, N. G.; Kroutil, W. Acyl Donors and Additives for the Biocatalytic Friedel–Crafts Acylation. *Eur. J. Org. Chem.* **2017**, *2017*, 5865–5871.

90. Żądło-Dobrowolska, A.; Schmidt, N. G.; Kroutil, W. Thioesters as Acyl Donors in Biocatalytic Friedel-Crafts-Type Acylation Catalyzed by Acyltransferase from *Pseudomonas Protegens*. *ChemCatChem* **2019**, *11*, 1064–1068.

91. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **2017**, *117*, 11570–11648.

92. Baunach, M.; Franke, J.; Hertweck, C. Terpenoid Biosynthesis Off the Beaten Track: Unconventional Cyclases and Their Impact on Biomimetic Synthesis. *Angew. Chem. Int. Ed.* **2015**, *54*, 2604–2626.

93. Pronin, S. V.; Shenvi, R. A. Synthesis of Highly Strained Terpenes by Non-Stop Tail-to-Head Polycyclization. *Nat. Chem.* **2012**, *4*, 915–920.

94. Siedenburg, G.; Jendrossek, D. Squalene-Hopene Cyclases. *Appl. Environ. Microbiol.* **2011**, *77*, 3905–3915.

95. Syren, P. O.; Henche, S.; Eichler, A.; Nestl, B. M.; Hauer, B. Squalene-Hopene Cyclases–Evolution, Dynamics and Catalytic Scope. *Curr. Opin. Struct. Biol.* **2016**, *41*, 73–82.

96. Hoshino, T.; Kumai, Y.; Kudo, I.; Nakano, S.; Ohashi, S. Enzymatic Cyclization Reactions of Geraniol, Farnesol and Geranylgeraniol, and Those of Truncated Squalene Analogs Having C20 and C25 by Recombinant Squalene Cyclase. *Org. Biomol. Chem.* **2004**, *2*, 2650–2657.

97. Seitz, M.; Syren, P. O.; Steiner, L.; Klebensberger, J.; Nestl, B. M.; Hauer, B. Synthesis of Heterocyclic Terpenoids by Promiscuous Squalene-Hopene Cyclases. *ChemBioChem* **2013**, *14*, 436–439.

98. 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*, 7624–7629.

99. Seitz, M.; Klebensberger, J.; Siebenhaller, S.; Breuer, M.; Siedenburg, G.; Jendrossek, D.; Hauer, B. Substrate Specificity of a Novel Squalene-Hopene Cyclase from *Zymomonas Mobilis*. *J. Mol. Catal. B Enzym.* **2012**, *84*, 72–77.

100. Siedenburg, G.; Jendrossek, D.; Breuer, M.; Juhl, B.; Pleiss, J.; Seitz, M.; Klebensberger, J.; Hauer, B. Activation-Independent Cyclization of Monoterpenoids. *Appl. Environ. Microbiol.* **2012**, *78*, 1055–1062.

101. Hammer, S. C.; Marjanovic, A.; Dominicus, J. M.; Nestl, B. M.; Hauer, B. Squalene Hopene Cyclases Are Protonases for Stereoselective Brønsted Acid Catalysis. *Nat. Chem. Biol.* **2015**, *11*, 121–126.

102. Degenhardt, J.; Kollner, T. G.; Gershenzon, J. Monoterpene and Sesquiterpene Synthases and the Origin of Terpene Skeletal Diversity in Plants. *Phytochem.* **2009**, *70*, 1621–1637.

103. Ueda, D.; Hoshino, T.; Sato, T. Cyclization of Squalene from Both Termini: Identification of an Onoceroid Synthase and Enzymatic Synthesis of Ambrein. *J. Am. Chem. Soc.* **2013**, *135*, 18335–18338.

104. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels–Alder Reaction. *Science* **2010**, *329*, 309–313.

105. Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M. Natural Diels–Alderases: Elusive and Irresistable. *J. Org. Chem.* **2015**, *80*, 11672–11685.

106. Dan, Q.; Newmister, S. A.; Klas, K. R.; Fraley, A. E.; McAfoos, T. J.; Somoza, A. D.; Sunderhaus, J. D.; Ye, Y.; Shende, V. V.; Yu, F.; Sanders, J. N.; Brown, W. C.; Zhao, L.; Paton, R. S.; Houk, K. N.; Smith, J. L.; Sherman, D. H.; Williams, R. M. Fungal Indole Alkaloid Biogenesis through Evolution of a Bifunctional Reductase/Diels–Alderase. *Nat. Chem.* **2019**, *11*, 972–980.

107. Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-c.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-Dependent Enzyme-Catalysed Pericyclic Reactions in Natural Product Biosynthesis. *Nature* **2017**, *549*, 502–506.

108. Cai, Y.; Hai, Y.; Ohashi, M.; Jamieson, C. S.; Garcia-Borras, M.; Houk, K. N.; Zhou, J.; Tang, Y. Structural Basis for Stereoselective Dehydration and Hydrogen-Bonding Catalysis by the SAM-Dependent Pericyclase Lepi. *Nat. Chem.* **2019**, *11*, 812–820.

109. Minami, A.; Oikawa, H. Recent Advances of Diels–Alderases Involved in Natural Product Biosynthesis. *J. Antibiot.* **2016**, *69*, 500–506.

110. Gao, L.; Su, C.; Du, X.; Wang, R.; Chen, S.; Zhou, Y.; Liu, C.; Liu, X.; Tian, R.; Zhang, L. FAD-Dependent Enzyme-Catalysed Intermolecular [4+2] Cycloaddition in Natural Product Biosynthesis. *Nat. Chem.* **2020**, *12*, 620–628.

111. Kohli, R. M.; Massey, V. The Oxidative Half-Reaction of Old Yellow Enzyme. The Role of Tyrosine 196. *J. Biol. Chem.* **1998**, *273*, 32763–32770.

112. Winkler, C. K.; Tasnádi, G.; Clay, D.; Hall, M.; Faber, K. Asymmetric Bioreduction of Activated Alkenes to Industrially Relevant Optically Active Compounds. *J. Biotechnol.* **2012**, *162*, 381–389.

113. Toogood, H. S.; Scrutton, N. S. Discovery, Characterisation, Engineering and Applications of Ene Reductases for Industrial Biocatalysis. *ACS Catal.* **2019**, *8*, 3532–3549.

114. Heckenbichler, K.; Schweiger, A.; Brandner, L. A.; Binter, A.; Toplak, M.; Macheroux, P.; Gruber, K.; Breinbauer, R. Asymmetric Reductive Carbocyclization Using Engineered Ene Reductases. *Angew. Chem. Int. Ed.* **2018**, *57*, 7240–7244.

115. Biegasiewicz, K. F.; Cooper, S. J.; Gao, X.; Oblinsky, D. G.; Kim, J. H.; Garfinkle, S. E.; Joyce, L. A.; Sandoval, B. A.; Scholes, G. D.; Hyster, T. K. Photoexcitation of Flavoenzymes Enables a Stereoselective Radical Cyclization. *Science* **2019**, *364*, 1166–1169.

116. Black, M. J.; Biegasiewicz, K. F.; Meichan, A. J.; Oblinsky, D. G.; Kudisch, B.; Scholes, G. D.; Hyster, T. K. Asymmetric Redox-Neutral Radical Cyclization Catalysed by Flavin-Dependent 'Ene'-Reductases. *Nat. Chem.* **2019**, 71–75.

117. Sandoval, B. A.; Meichan, A. J.; Hyster, T. K. Enantioselective Hydrogen Atom Transfer: Discovery of Catalytic Promiscuity in Flavin-Dependent 'Ene'-Reductases. *J. Am. Chem. Soc.* **2017**, *139*, 11313–11316.

118. Huang, X.; Wang, B.; Wang, Y.; Jiang, G.; Feng, J.; Zhao, H. Photoenzymatic Enantioselective Intermolecular Radical Hydroalkylation. *Nature* **2020**, *584*, 69–74.

119. Page, C. G.; Cooper, S. J.; DeHovitz, J. S.; Oblinsky, D. G.; Biegasiewicz, K. F.; Antropow, A. H.; Armbrust, K. W.; Ellis, J. M.; Hamann, L. G.; Horn, E. J.; Oberg, K. M.; Scholes, G. D.; Hyster, T. K. Quaternary Charge-Transfer Complex Enables Photoenzymatic Intermolecular Hydroalkylation of Olefins. *J. Am. Chem. Soc.* **2020**, *143*, 97–102.

120. Fu, H.; Lam, H.; Emmanuel, M. A.; Kim, J. H.; Sandoval, B. A.; Hyster, T. K. Ground-State Electron Transfer as an Initiation Mechanism for Biocatalytic C–C Bond Forming Reactions. *J. Am. Chem. Soc.* **2021**, 9622–9629.

121. Ye, T.; McKervey, M. A. Organic Synthesis with α-Diazo Carbonyl Compounds. *Chem. Rev.* **1994**, *94*, 1091–1160.

122. Doyle, M. P.; Forbes, D. C. Recent Advances in Asymmetric Catalytic Metal Carbene Transformations. *Chem. Rev.* **1998**, *98*, 911–936.

123. Davies, H. M.; Manning, J. R. Catalytic C–H Functionalization by Metal Carbenoid and Nitrenoid Insertion. *Nature* **2008**, *451*, 417–424.

124. Zhang, Y.; Wang, J. Recent Development of Reactions with α-Diazocarbonyl Compounds as Nucleophiles. *Chem. Comm.* **2009**, 5350–5361.

125. Ford, A.; Miel, H.; Ring, A.; Slattery, C. N.; Maguire, A. R.; McKervey, M. A. Modern Organic Synthesis with α-Diazocarbonyl Compounds. *Chem. Rev.* **2015**, *115*, 9981–10080.

126. Franssen, N. M. G.; Walters, A. J. C.; Reek, J. N. H.; de Bruin, B. Carbene Insertion into Transition Metal–Carbon Bonds: A New Tool for Catalytic C–C Bond Formation. *Catal. Sci. Technol.* **2011**, *1*, 153–165.

127. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, *105*, 2253–2278.

128. Isin, E. M.; Guengerich, F. P. Complex Reactions Catalyzed by Cytochrome P450 Enzymes. *Biochim. Biophys. Acta.* **2007**, *1770*, 314–329.

129. Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. Epoxidation of Olefins by Cytochrome P450: Evidence from Site-Specific Mutagenesis for Hydroperoxo-Iron as an Electrophilic Oxidant. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3555–3560.

130. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310.

131. Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111.

132. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A Serine-Substituted P450 Catalyzes Highly Efficient Carbene Transfer to Olefins *In Vivo*. *Nat. Chem. Biol.* **2013**, *9*, 485–487.

133. Heel, T.; McIntosh, J. A.; Dodani, S. C.; Meyerowitz, J. T.; Arnold, F. H. Non-Natural Olefin Cyclopropanation Catalyzed by Diverse Cytochrome P450s and Other Hemoproteins. *ChemBioChem* **2014**, *15*, 2556–2562.

134. Renata, H.; Lewis, R. D.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Wang, Z. J.; Arnold, F. H. Identification of Mechanism-Based Inactivation in P450-Catalyzed Cyclopropanation Facilitates Engineering of Improved Enzymes. *J. Am. Chem. Soc.* **2016**, *138*, 12527–12533.

135. Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem. Int. Ed.* **2015**, *54*, 1744–1748.

136. Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Stereoselective Olefin Cyclopropanation under Aerobic Conditions with an Artificial Enzyme Incorporating an Iron-Chlorin E6 Cofactor. *ACS Catal.* **2017**, *7*, 7629–7633.

137. Key, H. M.; Dydio, P.; Liu, Z.; Rha, J. Y. E.; Nazarenko, A.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. Beyond Iron: Iridium-Containing P450 Enzymes for Selective Cyclopropanations of Structurally Diverse Alkenes. *ACS Cent. Sci.* **2017**, *3*, 302–308.

138. Brandenberg, O. F.; Prier, C. K.; Chen, K.; Knight, A. M.; Wu, Z.; Arnold, F. H. Stereoselective Enzymatic Synthesis of Heteroatom-Substituted Cyclopropanes. *ACS Catal.* **2018**, *8*, 2629–2634.

139. Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenberg, O. F.; Chen, K.; Arnold, F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, 372–377.

140. Carminati, D. M.; Fasan, R. Stereoselective Cyclopropanation of Electron-Deficient Olefins with a Cofactor Redesigned Carbene Transferase Featuring Radical Reactivity. *ACS Catal.* **2019**, *9*, 9683–9697.

141. Tinoco, A.; Steck, V.; Tyagi, V.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Trifluoromethyl-Substituted Cyclopropanes via Myoglobin-Catalyzed Transfer of Trifluoromethylcarbene. *J. Am. Chem. Soc.* **2017**, *139*, 5293–5296.

142. Chandgude, A. L.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Nitrile-Substituted Cyclopropanes by Myoglobin-Mediated Carbene Transfer Catalysis. *Angew. Chem. Int. Ed.* **2018**, *57*, 15852–15856.

143. Chandgude, A. L.; Ren, X.; Fasan, R. Stereodivergent Intramolecular Cyclopropanation Enabled by Engineered Carbene Transferases. *J. Am. Chem. Soc.* **2019**, *141*, 9145–9150.

144. Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. Enzymatic Construction of Highly Strained Carbocycles. *Science* **2018**, *360*, 71–75.

145. Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem. Int. Ed.* **2016**, *55*, 16110–16114.

146. Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem. Int. Ed.* **2014**, *53*, 6810–6813.

147. Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological Catalysis by Artificial Haem Proteins Containing Noble Metals in Place of Iron. *Nature* **2016**, *534*, 534–537.

148. Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science* **2016**, *354*, 102–106.

149. Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089.

150. Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed sp^3 C–H Functionalization. *Nature* **2019**, *565*, 67–72.

151. Vargas, D. A.; Tinoco, A.; Tyagi, V.; Fasan, R. Myoglobin-Catalyzed C−H Functionalization of Unprotected Indoles. *Angew. Chem. Int. Ed.* **2018**, *57*, 9911–9915.

152. Brandenberg, O. F.; Chen, K.; Arnold, F. H. Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995.

153. Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H. Enantiodivergent α-Amino C–H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, 9798–9802.

154. Tyagi, V.; Sreenilayam, G.; Bajaj, P.; Tinoco, A.; Fasan, R. Biocatalytic Synthesis of Allylic and Allenyl Sulfides through a Myoglobin-Catalyzed Doyle–Kirmse Reaction. *Angew. Chem. Int. Ed.* **2016**, *55*, 13562–13566.

155. Tyagi, V.; Fasan, R. Myoglobin-Catalyzed Olefination of Aldehydes. *Angew. Chem. Int. Ed.* **2016**, *55*, 2512–2516.

156. Weissenborn, M. J.; Löw, S. A.; Borlinghaus, N.; Kuhn, M.; Kummer, S.; Rami, F.; Plietker, B.; Hauer, B. Enzyme-Catalyzed Carbonyl Olefination by the *E. Coli* Protein YfeX in the Absence of Phosphines. *ChemCatChem* **2016**, *8*, 1636–1640.

157. Kozlowski, M. C. Oxidative Coupling in Complexity Building Transforms. *Acc. Chem. Res.* **2017**, *50*, 638–643.

158. Kozlowski, M. C.; Morgan, B. J.; Linton, E. C. Total Synthesis of Chiral Biaryl Natural Products by Asymmetric Biaryl Coupling. *Chem. Soc. Rev.* **2009**, *38*, 3193–3207.

159. Bringmann, G.; Gulder, T.; Gulder, T. A. M.; Breuning, M. Atroposelective Total Synthesis of Axially Chiral Biaryl Natural Products. *Chem. Rev.* **2011**, *111*, 563–639.

160. Guengerich, F. P.; Yoshimoto, F. K. Formation and Cleavage of C–C Bonds by Enzymatic Oxidation–Reduction Reactions. *Chem. Rev.* **2018**, *118*, 6573–6655.

161. Ikezawa, N.; Iwasa, K.; Sato, F. Molecular Cloning and Characterization of CYP80G2, a Cytochrome P450 That Catalyzes an Intramolecular C–C Phenol Coupling of (*S*)-Reticuline in Magnoflorine Biosynthesis, from Cultured *Coptis Japonica* Cells. *J. Biol. Chem.* **2008**, *283*, 8810–8821.

162. Pylypenko, O.; Vitali, F.; Zerbe, K.; Robinson, J. A.; Schlichting, I. Crystal Structure of OxyC, a Cytochrome P450 Implicated in an Oxidative C–C Coupling Reaction During Vancomycin Biosynthesis. *J. Biol. Chem.* **2003**, *278*, 46727–46733.

163. Forneris, C. C.; Seyedsayamdost, M. R. *In Vitro* Reconstitution of OxyC Activity Enables Total Chemoenzymatic Syntheses of Vancomycin Aglycone Variants. *Angew. Chem. Int. Ed.* **2018**, *57*, 8048–8052.

164. Pickel, B.; Schaller, A. Dirigent Proteins: Molecular Characteristics and Potential Biotechnological Applications. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 8427–8438.

165. Jeon, J.-R.; Baldrian, P.; Murugesan, K.; Chang, Y.-S. Laccase-Catalysed Oxidations of Naturally Occurring Phenols: From *In Vivo* Biosynthetic Pathways to Green Synthetic Applications. *Microb. Biotechnol.* **2012**, *5*, 318–332.

166. Mate, D. M.; Alcalde, M. Laccase: A Multi-Purpose Biocatalyst at the Forefront of Biotechnology. *Microb. Biotechnol.* **2017**, *10*, 1457–1467.

167. Sagui, F.; Chirivì, C.; Fontana, G.; Nicotra, S.; Passarella, D.; Riva, S.; Danieli, B. Laccase-Catalyzed Coupling of Catharanthine and Vindoline: An Efficient Approach to the Bisindole Alkaloid Anhydrovinblastine. *Tetrahedron* **2009**, *65*, 312–317.

168. Kim, S. S.; Sattely, E. S. Dirigent Proteins Guide Asymmetric Heterocoupling for the Synthesis of Complex Natural Product Analogues. *J. Am. Chem. Soc.* **2021**, *143*, 5011–5021.

169. Hu, J.; Li, H.; Chooi, Y.-H. Fungal Dirigent Protein Controls the Stereoselectivity of Multicopper Oxidase-Catalyzed Phenol Coupling in Viriditoxin Biosynthesis. *J. Am. Chem. Soc.* **2019**, *141*, 8068–8072.

170. Liu, J.; Liu, A.; Hu, Y. Enzymatic Dimerization in the Biosynthetic Pathway of Microbial Natural Products. *Nat. Prod. Rep.* **2021**, *38*, 1469–1505.

171. Hüttel, W.; Müller, M. Regio- and Stereoselective Intermolecular Phenol Coupling Enzymes in Secondary Metabolite Biosynthesis. *Nat. Prod. Rep.* **2021**, *38*, 1011–1043.

172. Obermaier, S.; Thiele, W.; Fürtges, L.; Müller, M. Enantioselective Phenol Coupling by Laccases in the Biosynthesis of Fungal Dimeric Naphthopyrones. *Angew. Chem. Int. Ed.* **2019**, *58*, 9125–9128.

173. Thiele, W.; Froede, R.; Steglich, W.; Müller, M. Enzymatic Formation of Rufoschweinitzin, a Binaphthalene from the Basidiomycete *Cortinarius Rufoolivaceus*. *ChemBioChem* **2020**, *21*, 1423–1427.

174. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

175. Zhao, B.; Guengerich, F. P.; Bellamine, A.; Lamb, D. C.; Izumikawa, M.; Lei, L.; Podust, L. M.; Sundaramoorthy, M.; Kalaitzis, J. A.; Reddy, L. M.; Kelly, S. L.; Moore, B. S.; Stec, D.; Voehler, M.; Falck, J. R.; Shimada, T.; Waterman, M. R. Binding of Two Flaviolin Substrate Molecules, Oxidative Coupling, and Crystal Structure of *Streptomyces Coelicolor* A3(2) Cytochrome P450 158A2. *J Biol Chem* **2005**, *280*, 11599–11607.

176. Guo, Z.; Li, P.; Chen, G.; Li, C.; Cao, Z.; Zhang, Y.; Ren, J.; Xiang, H.; Lin, S.; Ju, J.; Chen, Y. Design and Biosynthesis of Dimeric Alboflavusins with Biaryl Linkages via Regiospecific C–C Bond Coupling. *J. Am. Chem. Soc.* **2018**, *140*, 18009–18015.

177. Tian, W.; Sun, C.; Zheng, M.; Harmer, J. R.; Yu, M.; Zhang, Y.; Peng, H.; Zhu, D.; Deng, Z.; Chen, S.-L. Efficient Biosynthesis of Heterodimeric C3-Aryl Pyrroloindoline Alkaloids. *Nat. Commun.* **2018**, *9*, 4428.

178. Shende, V. V.; Khatri, Y.; Newmister, S. A.; Sanders, J. N.; Lindovska, P.; Yu, F.; Doyon, T. J.; Kim, J.; Houk, K.; Movassaghi, M. Structure and Function of NzeB, a Versatile C–C and C–N Bond-Forming Diketopiperazine Dimerase. *J. Am. Chem. Soc.* **2020**, *142*, 17413–17424.

179. Yin, L.; Liebscher, J. Carbon−Carbon Coupling Reactions Catalyzed by Heterogeneous Palladium Catalysts. *Chem. Rev.* **2007**, *107*, 133–173.

CHAPTER II

Development of a Biocatalytic Platform for Biaryl Coupling in Yeast

Despite the diverse and potent bioactivities displayed by biaryl natural products, their application in drug discovery is limited by restricted access to these complex molecular scaffolds. In particular, fundamental challenges remain in controlling the site- and atroposelectivity in biaryl coupling reactions. In contrast, Nature has a wealth of biosynthetic enzymes that catalyze biaryl coupling reactions with catalyst-controlled selectivity unrivaled by current synthetic methods. In particular, a growing subset of fungal P450s have been identified to catalyze site- and atroposelective biaryl couplings. Herein, we report a whole-cell biocatalytic platform in *Pichia pastoris* to synthesize biaryl molecules through the recombinant production of the fungal P450 KtnC. Moreover, engineering a redox self-sufficient fusion enzyme further improves the efficiency of the system and sheds light on the redox compatibility of different yeast CPRs with KtnC. Altogether, this work provides a platform for biaryl coupling in yeast that can be applied to engineering a currently underexplored pool of fungal P450s into selective biocatalysts for the synthesis of complex biaryl compounds.

This chapter was adapted from Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. A Biocatalytic Platform for the Synthesis of Biaryl Molecules in Yeast. *Submitted (April 4, 2022)*.

INTRODUCTION

Compounds containing biaryl motifs have widely been acknowledged as privileged substructures in modern drug discovery.[1] Characterized by two aromatic systems connected by a single bond, the biaryl scaffold is present in many natural products with diverse biological activities and has influenced plant-based naturopathic remedies for centuries (Figure 2.1a).[2] For example, root extracts of *Ancistrocladus tectorius* are used in Chinese and Malaysian traditional medicines to treat dysentery, parasitic infections, and malaria.[3] The medicinal properties of *A. tectorius* have since been attributed to over 60 unique biaryl naphthylisoquinoline alkaloids produced by the plant, including **2.1**.[4] Despite the tremendous therapeutic potential exhibited by this class of natural products, the full exploration of their medicinal properties is limited by the ability to obtain significant quantities through isolation from their natural source or chemical synthesis.[5]

Several key structural features of biaryl compounds contribute to their synthetic challenge. First, the site of bond formation when coupling two aryl groups to from the biaryl bond can be challenging to control. The most common solution to this is to prefunctionalized the two aryl fragments at the desired site of bond formation in metal-catalyzed cross-coupling reactions.[6,7] Alternatively, a direct oxidative coupling strategy can be employed on two aryl fragments with unactivated C–H bonds; however, the reactivity and selectivity of these reactions are primarily dictated by the innate stereoelectronic properties of the substrates, limiting the versatility of the reaction.[8,9] Second, tetra-*ortho*-substituted biaryl compounds exhibit axial chirality due to hindered rotation around the biaryl C–C bond, enabling isolation of individual atropisomers of these molecules.[10] In the case with many biaryl natural products, only one of the atropisomers is bioactive as compared to the other. For example, the (–)-atropisomer of gossypol (**2.2**) displays desirable bioactivity as

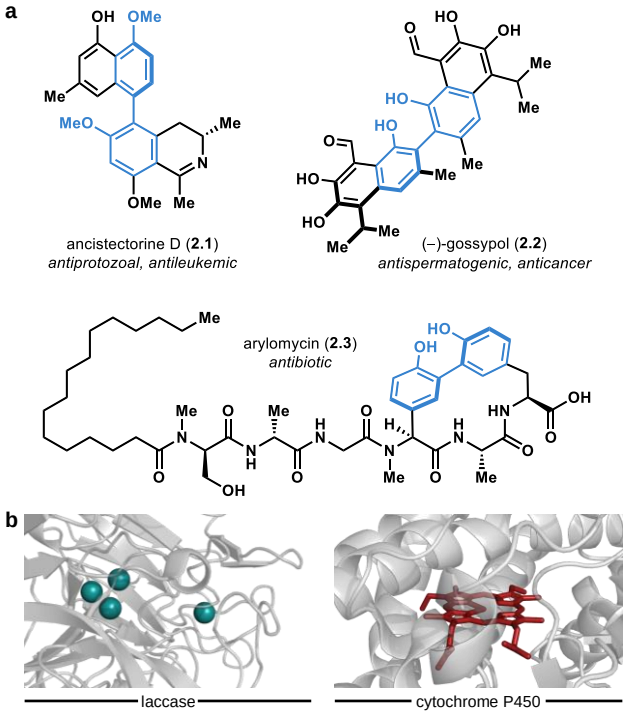


Figure 2.1. Significance and biosynthesis of biaryl molecules. (a) The biaryl scaffold is a privileged substructure in drug discovery, with a widespread presence in bioactive natural products. (b) Nature's strategies for biaryl bond formation.

anticancer and contraceptive agents.[11] Altogether, controlling the site- and atroposelectivity of biaryl coupling reactions using traditional synthetic chemistry approaches remains a fundamental challenge.

In contrast, Nature has several strategies for the site- and atroposelective formation of biaryl bonds in the biosynthesis of secondary metabolites.[12] Laccases and cytochromes P450 (P450s) are two different classes of metal-dependent enzymes that can catalyze the abstraction of a hydrogen atom from a phenolic substrate, thereby initiating an oxidative biaryl coupling reaction (Figure 2.1b).[13] However, the mechanism of control over the selectivity of the bond formation differs depending on the enzyme performing this initial oxidation. Typically, laccases are highly promiscuous multicopper oxidases that have been utilized for a variety of biotechnological applications;[14] however, the use of laccases in asymmetric catalysis has been limited due to the scarcity of reactions in which these enzymes exert control over the selectivity of the bond formation event.[15] In some instances, this lack of selectivity can be overcome by pairing laccase oxidation with an auxiliary dirigent protein, which can rapidly bind two intermediate phenoxy radical species in distinct binding pockets to guide the selectivity of the radical coupling event.[16] Alternatively, P450s exert control over the bond formation step following the initial oxidation, providing a platform for site- and atroposelective biaryl coupling reactions.[17] Although a number of these P450s are carrier protein-dependent[18] or limited to intramolecular biaryl couplings (such as in the formation of **2.3**),[19] a growing number of P450s have been identified to perform intermolecular oxidative biaryl coupling reactions.[13]

Despite the prevalence of biaryl natural products in all kingdoms of life, to date, the only P450s that have been discovered to catalyze intermolecular biaryl coupling reactions are from microbial sources.[20] Among these, a subset of fungal P450s have shown diverse reactivity with impressive catalyst-controlled selectivity in biaryl coupling reactions that is unmatched with current synthetic methods.[21] However, the membrane-bound nature and catalytic-dependence on a separate redox partner has limited the use of fungal P450s as biocatalysts in organic synthesis.[22] In this work, we disclose a platform for site- and atroposelective biaryl coupling in yeast using a fungal P450 catalyst. This *in vivo* platform is optimized by recombinant production of the P450 in *Pichia pastoris* and engineering of a self-sufficient P450-CPR fusion enzyme. We anticipate that

development of this whole-cell yeast platform will allow for the use of fungal P450s for biocatalytic biaryl coupling reactions to enable additional strategies for the synthesis of this valuable class of molecules.

RESULTS AND DISCUSSION

To develop an *in vivo* platform for biaryl coupling reactions, we first needed to select a fungal P450 known to catalyze an oxidative dimerization reaction in a natural product biosynthetic pathway. We were initially drawn to a group of *Aspergillus* P450s that catalyze the atroposelective dimerization of coumarin **2.4** to form bicoumarin isomers with complementary sites of biaryl bond formation, including the formation bicoumarin **2.5** mediated by KtnC (Figure 2.2a).[21,23] Given the demonstrated catalyst-controlled site- and atroposelectivity of these enzymes, we hypothesized

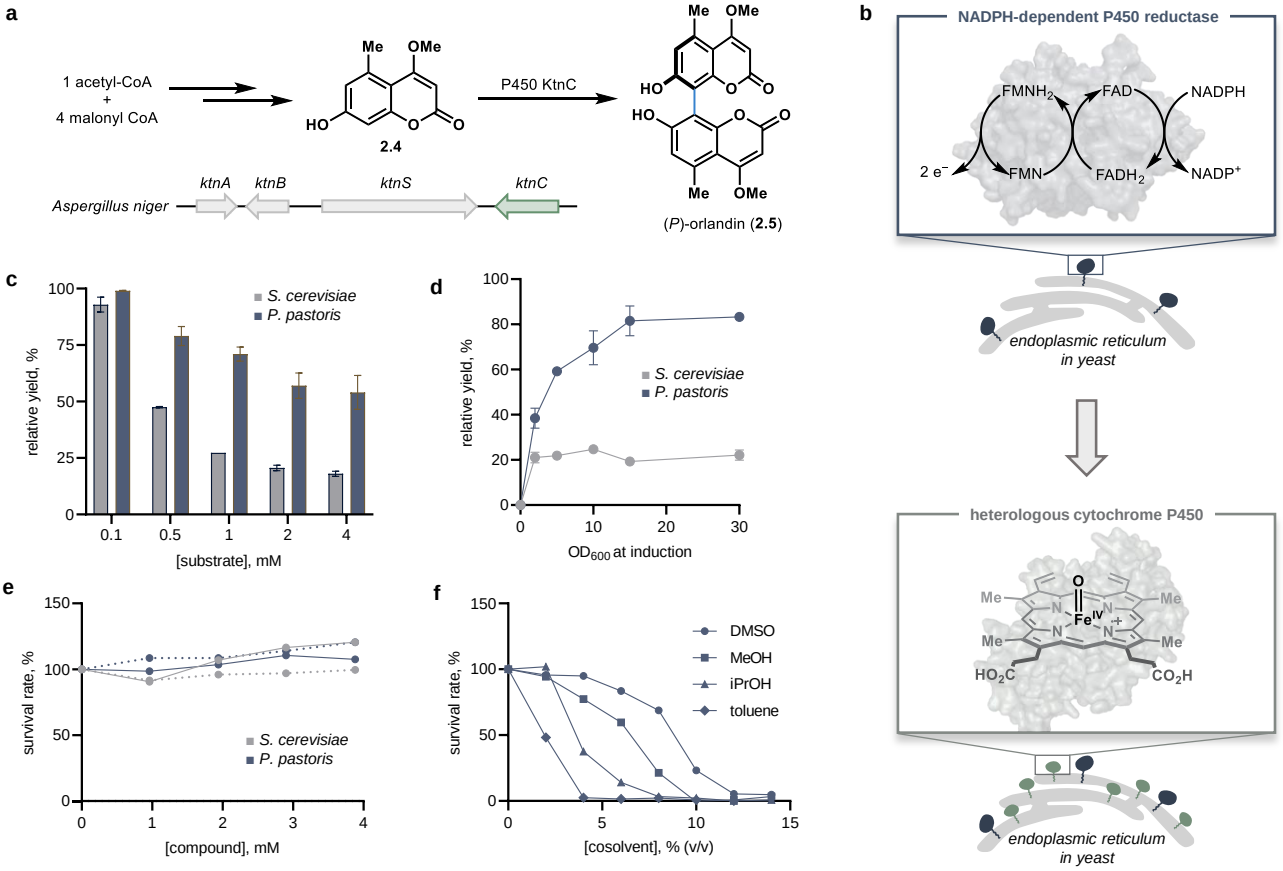


Figure 2.2. Optimization of biocatalytic platform for fungal P450 biocatalysis in yeast. (a) KtnC is a P450 that catalyzes the biaryl coupling in the biosynthesis of (*P*)-orlandin (**2.5**) in Aspergillus niger. (b) Yeast endogenously express CPR, which provides the necessary electrons for P450 catalysis through co-localization of the membrane-bound enzymes. (c) Biotransformations in *S. cerevisiae* suffered from limitations in substrate loading, however *P. pastoris* was able to accommodate higher substrate concentrations. (d) The higher cell density of *P. pastoris* expression cultures was the leading factor in higher yields. (e) No cytotoxicity effect was observed with increased concentration of substrate **2.4** (solid lines) and product **5** (dashed lines). (f) Limited amounts of co-solvent were permitted in *P. pastoris* cultures.

that a platform for preparative scale reactions with KtnC could provide a valuable tool for organic synthesis.[24]

Considering the ideal system for heterologous expression and biotransformations, we anticipated that yeast would provide an ideal platform due to the electron-transport machinery provided by endogenous levels of NADPH-cytochrome P450 reductase (CPR) enzymes in the endoplasmic reticulum (Figure 2.2b).[25] Recombinant production of KtnC in *Saccharomyces cerevisiae* reconstituted the activity of KtnC, as was previously reported by Müller and coworkers.[21] However, using *S. cerevisiae* as the host organism presented significant challenges in substrate loading, with only low concentrations of substrate **2.4** productively dimerizing to form bicoumarin **2.5** (Figure 2.2c). Moreover, the reliance of *S. cerevisiae* on simple sugars for cell growth and heterologous protein expression created concerns for downstream product isolation from the whole-cell biocatalyst.

To overcome these issues, *P. pastoris* was selected as an alternative host due to the high cell density achievable with this yeast strain and the reliance on methanol rather than galactose for recombinant protein production.[26] Recombinant production of KtnC in *P. pastoris* delivered significantly higher substrate tolerance in biotransformations, with over a 2-fold improvement in product formation at the higher substrate concentration compared to the *S. cerevisiae* system (Figure 2.2c). A comparison of the impact of the cell density at the point of induction for the two respective cultures showed that the *P. pastoris* biotransformations benefited from much higher cell densities. In contrast, increased cell concentrations did not affect the productivity of *S. cerevisiae* biotransformations (Figure 2.2d). The difference observed between the yeast strains was not due to a cytotoxicity effect in the presence of coumarin **2.4** or bicoumarin **2.5** (Figure 2.2e), however, using the live culture as the biocatalyst presented limitations in the amount of cosolvent permitted in the reactions (Figure 2.2f).

Using the optimized platform for biaryl coupling, we sought to test the capabilities of the *in vivo* platform in both preparative scale and high-throughput reactions (Figure 2.3). In a large-scale reaction, 104 mg substrate **2.4** was added to a liter of methanol-spiked minimal media containing concentrated KM71 *P. pastoris* cells producing recombinant KtnC. After three days of incubation at 30 °C, the reaction was extracted with toluene and ethyl acetate, yielding 60 mg bicoumarin **2.5** (58% yield).

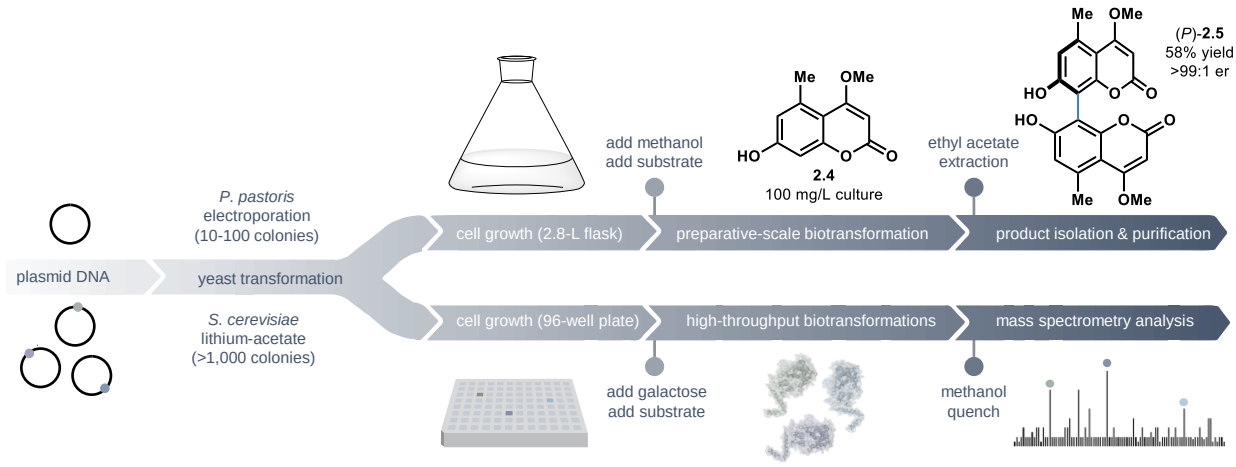


Figure 2.3. Application of biocatalytic platform for preparative scale and high-throughput biaryl synthesis.

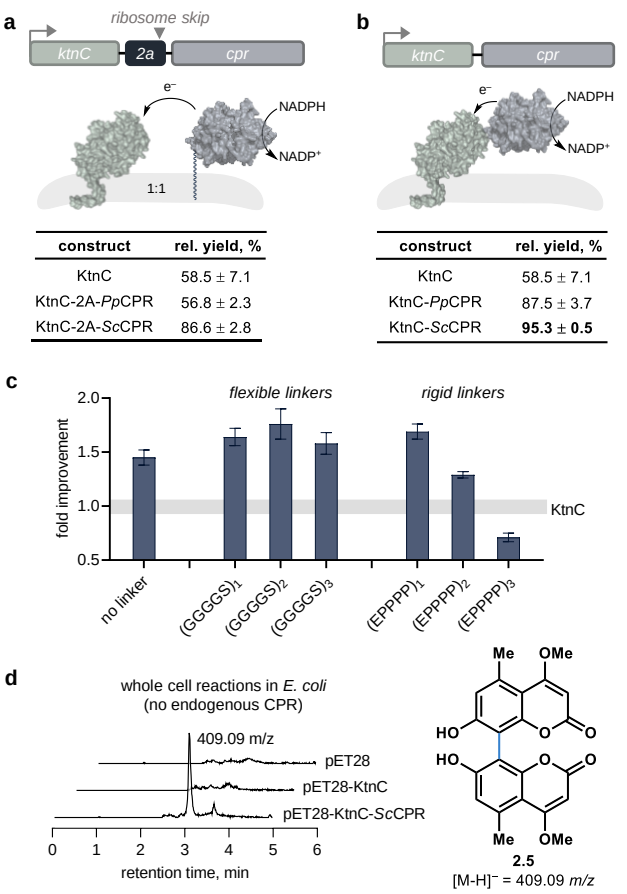
In addition to preparative scale reactions, the yeast-based platform was successfully applied to high-throughput screening in 96-well plates, allowing for screening of large substrate libraries and directed evolution.[27] Notably, while the *P. pastoris* strain was superior for preparative scale reactions, *S. cerevisiae* provided a more robust option for 96-well plate biotransformations due to the decreased concern of evaporation of the carbon source at culture volumes in the microliter range and the minimal effect of cell density on reaction outcome, alleviating the need for normalization of cell density across all wells of a given plate. Moreover, the higher transformation efficiency of DNA libraries achievable in *S. cerevisiae* provided a superior platform for protein engineering applications.

To further probe the impact of different yeast CPRs in the two yeast strains, several fusion constructs were engineered by inserting the genes encoding *Pp*CPR and *Sc*CPR downstream of the gene encoding KtnC. To achieve co-expression of the two membrane-bound enzymes, a sequence encoding a 2A peptide was inserted between *ktnC* and *cpr* to promote a ribosome skip during protein translation, thereby ensuring equimolar amounts of the redox partners (Figure 2.4a).[28] In *P. pastoris* biotransformations, only the co-expression of *Sc*CPR improved the reaction outcome, with increased amounts of *Pp*CPR providing no effect, suggesting that the CPR from *S. cerevisiae* is a more efficient redox partner for KtnC.

As an alternative approach to co-expressing the membrane-bound enzymes, fusion enzymes were engineered by insertion of a truncated form of the *cpr* directly at the 3′ end of *ktnC*, thereby providing a single polypeptide as a self-sufficient catalyst (Figure 2.4b).[29] Both of the fusion enzymes provided improved reaction outcomes in *P. pastoris* biotransformations compared

to *P. pastoris* heterologously expressing KtnC alone or with equimolar amounts of CPR, suggesting that the fusion enzyme offered more efficient electron transport. Additionally, as was previously observed, the fusion enzyme with *Sc*CPR performed better than with *Pp*CPR, providing additional support for the hypothesis that KtnC is more compatible with the CPR from *S. cerevisiae*.

To further optimize the efficiency of the KtnC-*Sc*CPR fusion enzyme, different peptide linkers separating the P450 and CPR were tested, including several flexible glycine-rich peptides and rigid proline-rich peptides (Figure 2.4c).[30] The flexible linkers offered similar to slightly improved yields in *P. pastoris* biotransformations compared to the fusion with no linker, however increasing the rigidity of the linker with proline-repeats was detrimental to catalysis. Finally, in order to confirm that the KtnC-*Sc*CPR fusion enzyme was actually acting as a self-sufficient catalyst, productive catalysis needed to be observed in an environment devoid of background redox partners. To do this, the gene encoding KtnC-*Sc*CPR was codon-optimized and transferred to a pET28 vector for heterologous expression in *E. coli*. In whole-cell reactions containing the heterologously expressed P450, only the fusion enzyme resulting the formation of dimer **2.5**, whereas the P450 alone gave no product (Figure 2.4d). These reactions verified that the fusion construct was, in fact, a self-sufficient enzyme capable of providing the electrons necessary for catalysis to the P450. Additionally, this important proof-of-concept offers future opportunities for the isolation of the self-sufficient fusion enzyme from yeast cells for *in vitro* biocatalysis.



construct	rel. yield, %
KtnC	58.5 ± 7.1
KtnC-2A-*Pp*CPR	56.8 ± 2.3
KtnC-2A-*Sc*CPR	86.6 ± 2.8

construct	rel. yield, %
KtnC	58.5 ± 7.1
KtnC-*Pp*CPR	87.5 ± 3.7
KtnC-*Sc*CPR	**95.3 ± 0.5**

Figure 2.4. Engineering self-sufficient fusion enzyme. (a) Heterologous expression of *Sc*CPR via a self-cleaving 2A resulted in increased yields in *P. pastoris* biotransformations. (b) Fusion constructs with truncated CPRs gave improved reaction yields. (c) Flexible peptides fusing KtnC and *Sc*CPR slightly improved the activity of the fusion, whereas longer, rigid linkers decreased the activity significantly. (d) The self-sufficiency of the KtnC-CPR fusion enzyme was confirmed by heterologous expression and whole-cell reactions in *E. coli* without the presence of endogenous yeast CPRs.

Beyond KtnC, a growing number of fungal P450s have been characterized to perform biaryl coupling reactions in the biosynthesis of secondary metabolites.[13] Moreover, these enzymes likely exist within a much larger pool of sequences that have yet to be associated with a secondary metabolite pathways or otherwise functionally characterized. To visualize this sequence space, a sequence similarity network (SSN) was constructed with the entire P450 protein family and then truncated to edges possessing an alignment score greater than 103 with fungal P450s known to perform oxidative biaryl coupling reactions (Figure 2.5).[31-33] This bioinformatic analysis of the fungal P450 sequence space not only shows a relationship between fungal P450 sequences that catalyze biaryl coupling, but also demonstrates that these catalysts exist within a much larger pool of P450 sequences with untapped potential for complementary biaryl coupling reactions.

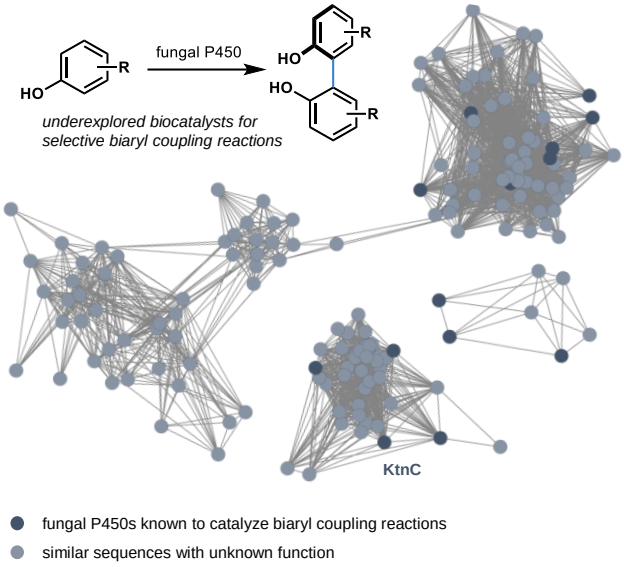


Figure 2.5. Sequence similarity network of fungal P450s that catalyze biaryl couplings. E-value: 103; minimum sequence length: 300 residues; alignment score: 88.

In conclusion, this work provides a platform for *in vivo* biocatalysis using fungal P450s for the site- and atroposelective synthesis of biaryl compounds. The platform was developed using KtnC, a fungal P450 that performs a late-stage biaryl coupling reaction in the biosynthesis of the (*P*)-orlandin (**2.5**), as a model system. Recombinant production of KtnC in *P. pastoris* in the presence of coumarin **2.4** substantially improved the production of the biaryl product (**2.5**) compared to the previously established *S. cerevisiae* platform due to the high cell density in *P. pastoris* cultures. This provided a robust platform for the preparative scale synthesis and isolation of biaryl **2.5**. Moreover, engineering KtnC as a fusion enzyme with reductases from *S. cerevisiae* and *P. pastoris* further improved the catalytic efficiency of the *in vivo* biocatalyst and provided evidence that KtnC is more compatible with reductase native to *S. cerevisiae*. Given the successful application of this *in vivo* platform for both preparative scale biaryl synthesis and high-throughput screening, we anticipate that the powerful reactivity and selectivity of further fungal P450s will be unlocked, providing additional tools for the synthesis of high-value biaryl compounds.

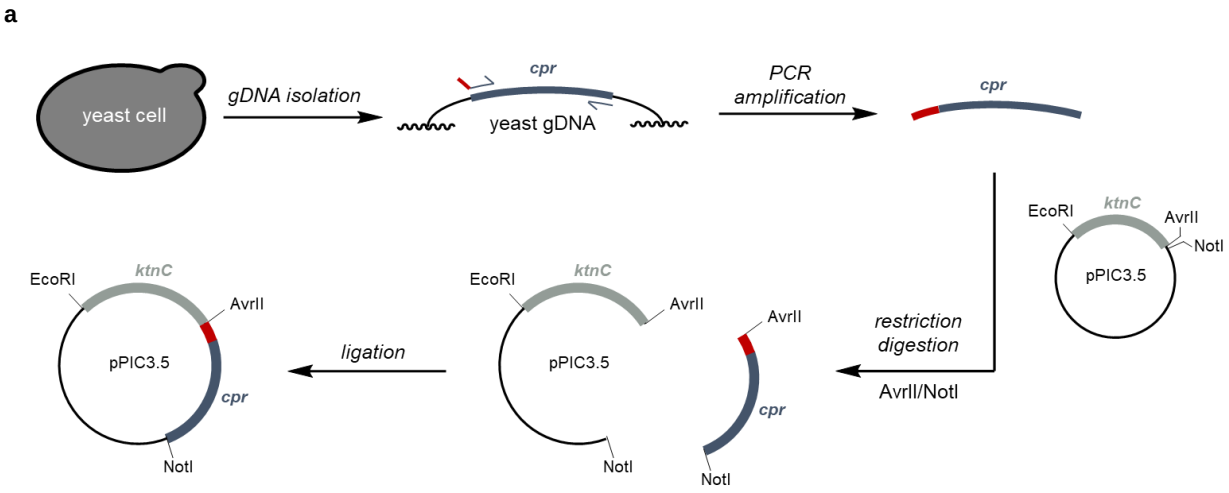
EXPERIMENTAL PROCEDURES

Design of expression constructs. The open reading frame for *ktnC* were amplified using Phusion DNA polymerase (NEB; Table 2.1). The reaction conditions were programmed as follows: 95 °C denaturation for 2 min; 30 cycles of 95 °C for 30 s, 61 °C for 1 min, 72 °C for 3 min; and a final 72 °C extension for 5 min. PCR products were purified from a 0.8% agarose gel and digested with 2 µL each of EcoRI and AvrII restriction enzymes in 50 µL Cutsmart buffer (NEB). Reactions were incubated at 37 °C for 2.5 h before quenching and purification of the DNA. Ligation reactions containing 100 ng of digested pPIC3.5 plasmid, 100 ng of digested insert, and 1 unit of T4 DNA ligase (NEB) were incubated in 10 µL ligase buffer for 2 h at room temperature, followed by a 65 °C heat inactivation for 10 min. DH5α *E. coli* cells were transformed with the ligation reaction and transformants were confirmed for gene integration by sequencing.

Table 2.1. Primers used for cloning.

Primer	Sequence (5' to 3')
Cloning KtnC into *P. pastoris* expression vector	
EcoRI_Kozak_FLAG_KtnC_fwd	CATGATGAATTCAATAATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTC
AvrII_KtnC_rvs	CATGTACCTAGGCTTCCCGTCACTTCTTGG
Cloning KtnC-2A-CPR constructs	
AvrII_1D-2A_*Pp*CPR_fwd-1	AGGGCCCGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGATACCTTGGACTTTTCT
AvrII_1D-2A_*Pp*CPR_fwd-2	CAAACTTTGAACTTTGATTTGTTGAAATTGGCTGGTGATGTTGAATCTAATCCAGGGCCCGACTACAAA
AvrII_1D-2A_*Pp*CPR_fwd-3	GCTAGAGCTCCTAGGGAAGCTAGACATAAACAAAAGATTGTTGCTCCAGTTAAACAAACTTTGAACTTTGATTTGT
NotI_*Pp*CPR_rvs	CATGTAGCGGCCGCTTACCAAACGTCTTCCTGAT
AvrII_1D-2A_*Sc*CPR_fwd-1	TCCAGGGCCCGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGCCGTTTGGAATAGAC
AvrII_1D-2A_*Sc*CPR_fwd-2	AACAAACTTTGAACTTTGATTTGTTGAAATTGGCTGGTGATGTTGAATCTAATCCAGGGCCCGA
AvrII_1D-2A_*Sc*CPR_fwd-3	GCTAGAGCTCCTAGGGAAGCTAGACATAAACAAAAGATTGTTGCTCCAGTTAAACAAACTTTGAACTTTGATTT
NotI_*Sc*CPR_rvs	CATGTAGCGGCCGCTTACCAGACATCTTCTTGGTATC
Cloning KtnC-CPR constructs	
AvrII Δ35*Pp*CPR_fwd	GCTAGAGCTCCTAGGGGAGTAGCTGGAGGATTTCA
NotI_*Pp*CPR_rvs	CATGTAGCGGCCGCTTACCAAACGTCTTCCTGAT
AvrII_Δ41*Sc*CPR_fwd	GCTAGAGCTCCTAGGGCTGTCAGCTCGGGC
NotI_*Sc*CPR_rvs	CATGTAGCGGCCGCTTACCAGACATCTTCTTGGTATC
Cloning KtnC-[linker]-*Sc*CPR constructs	
AvrII_$[GGGGS]_1$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGGTGGTGGTGGTTCTGCTGTCAGCTCGGGC
AvrII_$[GGGGS]_2$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGCTGTCAGCTCGGGC
AvrII_$[GGGGS]_3$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGCTGTCAGCTCGGGC
AvrII_$[EPPPP]_1$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGAACCTCCTCCTCCTGCTGTCAGCTCGGGC
AvrII_$[EPPPP]_2$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGAACCTCCTCCTCCTGAACCTCCTCCTCCTGCTGTCAGCTCGGGC
AvrII_$[EPPPP]_3$_Δ41*Sc*CPR_fwd	AGAGCTCCTAGGGAACCTCCTCCTCCTGAACCTCCTCCTCCTGAACCTCCTCCTCCTGCTGTCAGCTCGGGC
NotI_*Sc*CPR_linker_rvs	CATGTAGCGGCCGCTTAC

For KtnC-CPR fusion constructs, *cpr* genes were amplified from yeast genomic DNA isolated from BY4742 *S. cerevisiae* and KM71 *P. pastoris* cells using Phusion DNA polymerase



b

```
ScCPR     5    IDNTDFTVLAGLVLAVLLYVKRNSI--KELLMSDDGDITAVSSG--NRDIAQVVTENNKN   60
PpCPR     1    MDTLDLSVLIAIALALIAYFSKGSLWGKE-----DDNSVHGVAGGFQTRDLVEILNSTNKK   56

ScCPR    61    YLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVP------VIVSIF   114
PpCPR    57    ALVLYGSQTGTSEDYAHKYARELQSKFSIPTLCGDLSEFDFDNLNDIPEQVEGFTFITFF   116

ScCPR   115    ISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAA   174
PpCPR   117    MATYGEGEPTDNAVEFIEFLKN-DAEDLSNLKYTVFGLGNSTYEFYNQMGKTTNKRFSEL   175

ScCPR   175    GAIRLGKLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQY---TVLNE   231
PpCPR   176    GAQLVGTFGEGDDGQATMDEDFLAWKDSLFDTIKKDLHLEEHEVVYQPGLKVKENTALTT   235

ScCPR   232    ITDSMSLGEPSAHYLPSHQLNRNADGIQLGPFDLSQPYIAPIVKSRELFS-SNDRNCIHS   290
PpCPR   236    SSPNVSVGEPNKAYV----LREDENLLQYGPFDHTHPYIAPISSSRELCSETSERNCIHL   291

ScCPR   291    EFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIFNLDPE--TIFDLKPLDPTVKVPFPT   348
PpCPR   292    EFDLSNTNLRYSTGDHLAVWPSNANEHVESFLKVFNLTDKRSSVFDIEFLDPTVTVHFPF   351

ScCPR   349    PTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKDQFAVEITSKYFNI   408
PpCPR   352    PTTYEAVVRHHLEISGPISRQTLKQFIPYAPDQSTKQEVIRLSESKDVFHNEVTAKYYNL   411

ScCPR   409    ADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIVENFPNPEL   468
PpCPR   412    ADLLFKVSKETPWN-VPFNFLIETMPNLQHRYYSISSSSLSEKQTIHITAMIEAF-TPTG   469

ScCPR   469    PDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRL   528
PpCPR   470    SDHI-VTGVTTNLLWNIQLNQDKSTV---KAPVSYDLNGPRNLFSPYKLPVHIRRSTFKL   525

ScCPR   529    PSNPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTD-DFL   587
PpCPR   526    PSNPALPVIMIGPGTGVAPFRGFIRERC-----QQVDNGTPNIGQSILYYGCRNSEQDFL   580

ScCPR   588    YQDEWPEYAKKLDGSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDA   647
PpCPR   581    YRDEWPTYSKKLGDKFKMYTAFSR-ENSHKVYVQHRLLENSREFIELMDQGAFIYVCGDA   639

ScCPR   648    KGMAKGVSTALVGILSRGKSITTDEATELIKMLKTSGRYQEDVW   691    50% identity
PpCPR   640    GKMAKDVNKAIVEILIKEKGLSEEDATESIREFKTSNRYQEDVW   683    67% similarity
```

Figure 2.6. Strategy for cloning KtnC-CPR self-cleaving and fusion constructs. (a) The *cpr* genes were amplified from *S. cerevisiae* and *P. pastoris* genomic DNA and cloned into a pPIC3.5 vector harboring *ktnC*. (b) Sequence alignment of yeast CPRs. For P450-CPR fusion constructs, truncated CPRs were used with no N-terminal membrane-bound region (shown in underlined region).

(NEB; Table 1). The reaction conditions were programmed as follows: 95 °C denaturation for 2 min; 25 cycles of 95 °C for 30 s, 61 °C for 1 min, 72 °C for 3 min; and a final 72 °C extension for 15 min. PCR products were purified from a 0.8% agarose gel and cloned into pPIC3.5-*ktnC* at the 3′ end of the *p450* gene for expression as a fusion protein in *P. pastoris*. The vector and insert were digested using AvrII and NotI and ligated as previously described to give the fusion construct pPIC3.5-*ktnC-cpr*. The linker region between KtnC and CPR in the fusion construct was further modified to encode for (1) a self-cleaving 2A peptide, (2) flexible peptide linkers with up to three repeats of Gly-Gly-Gly-Gly-Ser (GGGGS), or (3) rigid peptide linkers with up to three repeats of Glu-Pro-Pro-Pro-Pro (EPPPP).[30] Similar cloning strategies were employed as described above (Figure 2.6).

Heterologous expression and biotransformations in *S. cerevisiae*. BY4742 *S. cerevisiae* cells were transformed with a pESC-HIS expression vector harboring *ktnC* using standard protocols[34] and incubated on histidine dropout agar containing 4% glucose for 2 days at 30 °C. Resulting colonies were grown in histidine dropout minimal medium (0.67% (w/v) yeast nitrogen base, 2 g/L drop-out mix, 450 mg/L adenine, 0.5 g/L leucine, 1.9 g/L lysine, 1.9 g/L tryptophan, 1.9 g/L uracil) containing 4% glucose overnight at 30 °C with shaking at 235 rpm. Protein expression was induced by exchange into histidine dropout minimal media containing 6% galactose to an OD_{600} of up to 4. For biotransformations, cultures were spiked with substrate **2.4** to a final concentration of 0.1-4 mM at the point of induction. Cultures were incubated in 24-well culture plates (up to 2 mL per well) for 1-4 days at 30 °C with shaking at 235 rpm.

Heterologous expression and biotransformations in *P. pastoris*. The expression plasmid harboring *ktnC* was linearized with SalI-HF restriction enzyme using the same digestion procedures as previously described. KM71 *P. pastoris* cells were transformed with up to 5 μg of linearized DNA by electroporation[35] and incubated on minimal agar with 2% glucose for 3 days at 30 °C. Resulting colonies were grown in minimal media (100 mM potassium phosphate pH 6.0, 1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin) with 1% (v/v) glycerol overnight at 30 °C with shaking. Cultures were induced for expression by resuspension in minimal media with 0.5% (v/v) methanol to an optical density of 10-30 at 600 nm. For biotransformations, cultures were spiked with substrate **2.4** to a final concentration of 0.1-4 mM at the point of induction. Cultures were incubated in 24-well culture plates (up to 2 mL per well) at 30 °C with shaking for

1-4 days and supplemented with methanol daily to maintain a concentration of 0.5% (v/v) methanol in the cultures.

Western blot analysis of protein expression. KtnC with an N-terminal FLAG-tag was expressed in yeast for 3 days at 30 °C. Equal amounts of cells from each culture based on OD_{600} were harvested. Cells were resuspended in 10 µL pellet lysis buffer (33 mM Tris acetate pH 7.9, 150 mM potassium acetate, 0.66 mM magnesium acetate, 20% (v/v) glycerol, 0.2% (v/v) Tween-20), 10 µL 1 M DTT, and 10 µL NuPAGE LDS sample buffer. Cells were lysed by bead beating, and boiled at 100 °C for 10 min. The soluble and insoluble proteins in the lysed samples were separated by centrifugations, and the proteins in the insoluble fraction were resuspended in pellet lysis buffer/DTT/LDS and resolved by SDS-PAGE. Proteins were transferred to an activated PVDF membrane using Western blot transfer buffer containing 20% ethanol. The resulting membrane was blocked overnight in a 5% (v/v) solution of evaporated milk in TBST at 4 °C followed by a 1 h incubation with monoclonal antibody (anti-FLAG M2-peroxidase, 1:1000) at room temperature. After washing the membrane with TBST, the membrane was exposed to ECL Select Western Blotting Detection Reagent for 1 min followed by visualization of heterologously expressed FLAG-tagged proteins using chemiluminescence (Figure 2.7).

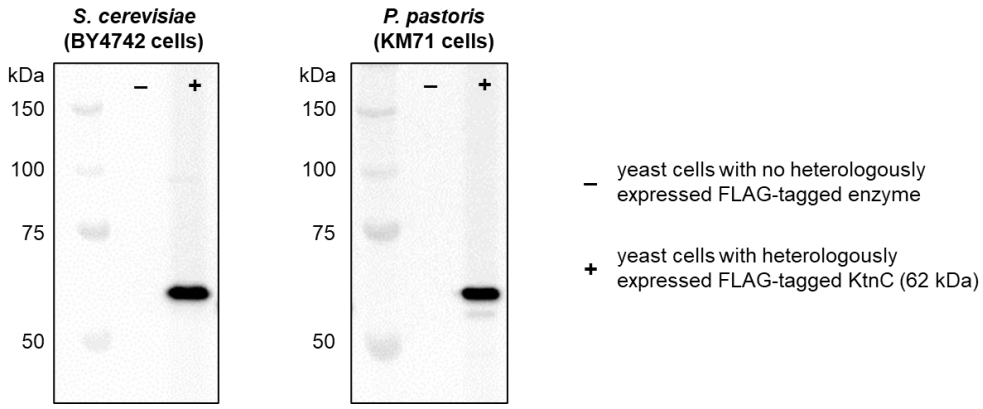


Figure 2.7. Western blot analysis of recombinant KtnC production in yeast.

Heterologous expression and whole-cell reactions in *E. coli*. BL21 (DE3) *E. coli* cells were co-transformed with a pG-Tf2 chaperone and a pET28 vector harboring the genes encoding Δ2-20-KtnC or Δ2-20-KtnC-*Sc*CPR (codon-optimized for heterologous expression in *E. coli*). Transformed cells were grown at 37 °C, 200 rpm overnight, then used to inoculate 200 mL TB containing 50 µg/mL kanamycin, 20 µg/mL chloramphenicol, and 5 ng/mL tetracycline. Cultures

were grown at 37 °C, 200 rpm until OD_{600} reached 0.7, then expression was induced with 100 μM IPTG and 1 mM 5-aminolevulinic acid. Expression cultures were grown at 16 °C, 200 rpm for 24 h. Cells were harvested and resuspended in 50 mM tricine pH 8 to a concentration of 100 mg wet cells/mL buffer. Reactions with 1 mM substrate **2.4**, 2 mM glucose-6-phosphate, 0.4 mM $NADP^+$, 0.4 units/mL glucose-6-phosphate dehydrogenase, and 10% (v/v) toluene were incubated at 30 °C, 500 rpm for 18 h.

LC–MS analysis of reactions. Reactions were quenched by addition of 3 equivalents of methanol, then centrifugation at 17,000 x *g* for 10 min. The resulting supernatant was subjected to liquid chromatography PDA spectrometry (UPLC) analysis performed on an Agilent 6230 time of flight mass spectrometer with a Dual AJS ESI source and an Agilent 1290 Infinity Series II diode array detector, autosampler, and binary pump (phase A = 95:5 water/acetonitrile (v/v); phase B = 95:5 acetonitrile/water (v/v); 0.7 mL/min). Samples were loaded onto a Waters Acquity UPLC HSST3 column (1.8 μm C18, 2.1 x 50 mm) and resolved with the following gradient: hold at 30% B for 0.5 min, increase to 45% B over 1 min, hold at 100% B for 0.5 min. Ion chromatograms were generated in negative mode and extraction ion chromatogram (EIC) peak areas were were used to calculate the relative percent yield for a given reaction.

Compound toxicity screen. BY4742 *S. cerevisiae* and KM71 *P. pastoris* cells were grown in YPD at 30 °C to saturation. Cells were distributed to each well of a 96-well culture plate containing various concentrations of coumarin **2.4** (up to 4 mM), bicoumarin **2.5** (up to 4 mM), or cosolvent (up to 14%, v/v) so that the starting OD_{600} was equal to 3 in all wells. The cultures were

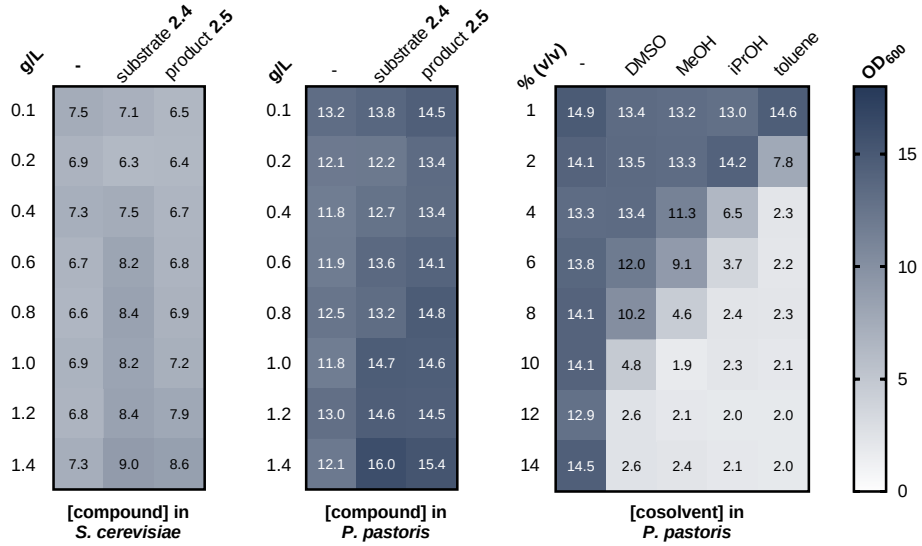
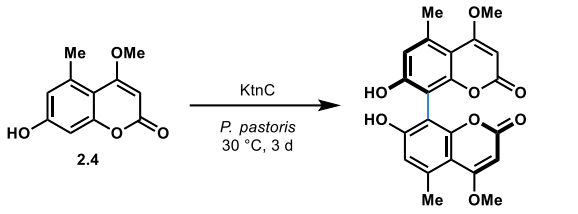


Figure 2.8. Effect of compound and cosolvent concentration on cell survival.

incubated at 30 °C with shaking and the OD_{600} of each of the wells was measured after 4 d (Figure 2.8). The survival rate of each culture was quantified relative to the OD_{600} of the cultures containing no additional compounds.

Scheme 2.1. Preparative scale biotransformation with KtnC to form 2.5.



Preparative scale biotransformation and product isolation. KM71 *P. pastoris* cells harboring the gene encoding KtnC were grown in 10 mL minimal media with 1 % (v/v) glycerol overnight at 30 °C with shaking at 235 rpm. The 10-mL culture was used to inoculate 6 L of minimal media with 1 % (v/v) glycerol, separated across 6 x 2.8-L Erlenmeyer flasks. Cells were grown overnight at 30 °C with shaking at 200 rpm until saturation. Cultures were centrifuged at 1,000 x *g* for 10 min in sterile bottles and resuspended in 1 L of minimal media with 0.5 % (v/v) methanol to an OD_{600} of 30. Substrate **4** was added to the expression culture (104 mg, 0.50 mmol, 10 mL from 50 mM stock in DMSO) and the biotransformation was incubated at 30 °C with shaking at 235 rpm. Methanol was added daily to a concentration of 0.5% (v/v) in the cultures. After 3 d, the completed biotransformation (95% conversion) was centrifuged to separate the cellular matter and the supernatant. The supernatant was extracted with a mixture of toluene and ethyl acetate (1:4 (v/v); 400 mL x 3). The cellular matter was vigorously agitated with a mixture of toluene and ethyl acetate (1:4 (v/v); 400 mL x 3). The combined organics were concentrated and purified over silica gel chromatography (5:4:1 (v/v) mixture of ethyl acetate/toluene/formic acid) to afford the dimeric product **5** as an off-white solid (60 mg, 58% isolated yield). $\boldsymbol{R_f}$ = 0.5 (5:4:1 (v/v) mixture of ethyl acetate/toluene/formic acid). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.71 (s, 2H), 5.57 (s, 2H), 4.94 (s, 6H), 2.59 (s, 6H); **^{1}H NMR** (400 MHz, CD3OD) δ 6.75 (s, 2H), 5.59 (s, 2H), 4.00 (s, 6H), 2.67 (s, 6H); **HRMS (ESI)** calculated for $C_{22}H_{19}O_8^+$ $[M+H]^+$ = 411.1074, found = 411.1077 *m/z*. All spectra obtained were in agreement with literature values.[21]

Assignment of absolute configuration of biocatalytic product. The general approach for absolute configuration assignment using ECD, including the detailed computational workflow, has been published elsewhere.[36-38] A subset of the details of the computational methodology is provided here. Conformers of each test structure were geometry optimized at the B3LYP/6-31G**

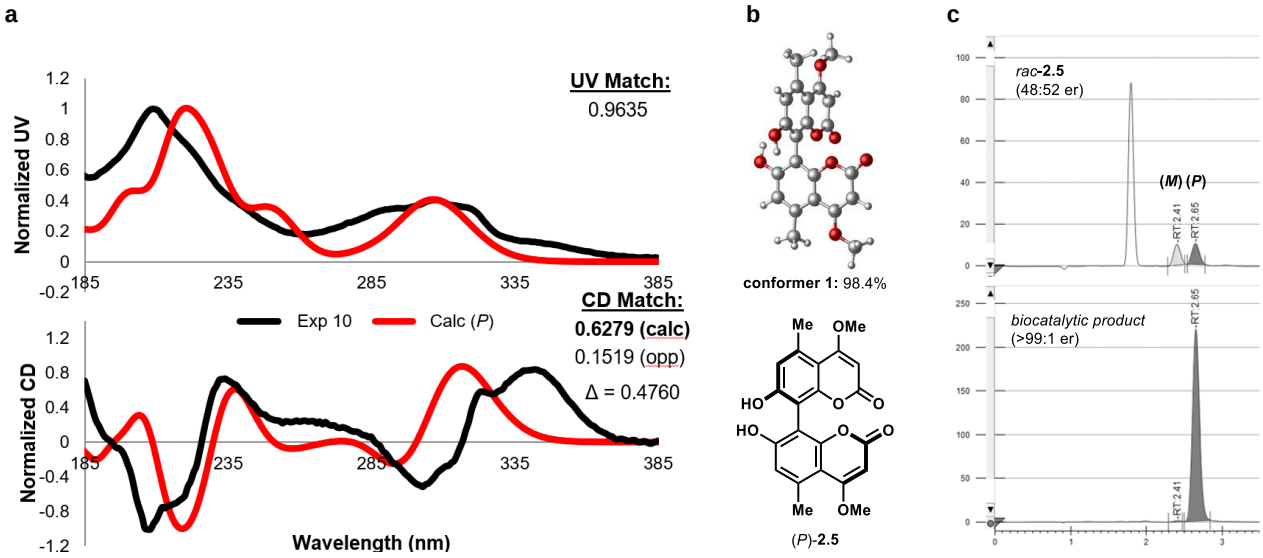


Figure 2.9. Assignment of absolute configurations of biocatalytic product. (a) Comparison between compound **2.5** experimental (black) and calculated (red) UV (top) and ECD (bottom) spectra. The calculated (*P*)-enantiomer is the better match to the experimental spectrum, with a large difference in fit (Δ=0.4760) suggesting a confident assignment. The calculated spectrum has been shifted 25 nm, and a band broadening of 0.25 eV has been applied. (b) One conformer of the (*P*)-enantiomer of compound **2.5** that contribute >2% to the Boltzmann distribution. Assigned absolute configurations of compound **2.5** based on ECD analysis. The percentage shown below is based on *in vacuo* electronic energies. (c) Chiral resolution of *rac*-**2.5** (top) and the biocatalytic product (bottom). Atropisomers were resolved using supercritical fluid chromatography with a CHIRALPAK OJ-H column (20% MeOH, CO_2, 3.5 mL/min).

level and stationary points were confirmed by performing frequency calculations.[39-47] All calculations were performed using Gaussian 09.[48] Output conformers were ranked according to DFT energy and a clustering was performed in order to remove duplicates. Initial duplicate identification was performed solely on an electronic energy basis where two compounds were considered identical if the difference in Hartrees was less than 0.01. Rounding the differences led to inconsistencies in identification of duplicates. It became better to cluster the DFT minima by energy and then re-cluster each energy bucket by structure using an all-atom RMS of 0.6 Å. This process removed just identical compounds. Two Boltzmann distributions were calculated based on the free energy (G) and the electronic energy (E). To calculate UV and ECD spectra, B3LYP geometries were used as input. The spectra were then calculated using either the B3LYP or CAM-B3LYP[49] functionals, along with the 6-31++G** basis set *in vacuo*.[50, 51] Only conformers which contributed more than 5.0% to the total in vacuo conformer distribution were selected for UV and ECD calculation. Time-dependent Density Functional Theory (TDDFT)[52] methodology was employed using the following keywords: TD=full,singlet, Nstates=100, and integral=ultrafinegrid. Spectral dis-

play, Boltzmann weighting, and curve fitting were carried out using SpecDis.[53, 54] and were displayed with a wavelength shift and band broadening sigma values in order to best match the calculated and experimental UV spectra. This shift and band broadening were then applied to the ECD spectra, and the area under the curve fit was determined by SpecDis. Conformers were visualized using CYLview Version 1.0.[55] Calculations of the ECD and UV spectra (CAM-B3LYP/6-31++G**) involved modeling the (*P*)-enantiomer of the compound. Hence, the (*P*)-enantiomer had a positive dihedral angle about the atropisomeric bond. Since the atropisomeric bond represents the only stereochemical element in the molecule, the (*M*)-enantiomer is assumed to have a spectrum that will be equal and opposite at all wavelengths.

Generation of sequence similarity networks. Sequences of 14 fungal P450s known to perform oxidative coupling reactions in the biosynthesis of biaryl secondary metabolites were uploaded to EFI-EST (Table 2.2), along with the UniRef50 sequences for the entire P450 family (PF00067).[31-33] The resulting SSN was visualized with Cytoscape software with the following parameters: E-value of 100, minimum sequence length of 300 residues, and an alignment score of 88 (Figure 2.10). The dataset was truncated to fungal sequences clustering with P450s known to perform oxidative coupling reactions. The resulting sequences represented in the truncated dataset were uploaded to EFI-EST to create a new SSN with the following parameters: E-value of 103, minimum sequence length of 300 residues, and an alignment score of 88 (see Figure 2.5).

Table 2.2. Fungal P450 sequences used in sequence similarity network.

P450	Protein sequence
KtnC (*Aspergillus niger*)[21, 23]	MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPK LRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEF ADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTIL SLCSTVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKH LSLVRERLVPEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLA DDLIFYHFELCKPTAFNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESF TKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFY DSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFT SGPVYMPNPSVMMRIRPRSDGK
DesC (*Aspergillus desertorum*)[21, 23]	MGALSEIFTAGYVALAGITGFFLVFLGFLVVISDYIDGWRSRRALGDIPIVDEGSNLSPILRWN SQPYDAEAEFTRAYYKYSKNGKPFAARIQHGGYAIVLPPSACRKVRSIGHEQLSFLDALAEFAD LSLHMDVTSRRVIEATHACNNETTIKNFQERLALECGKHLAPVFDPPQEQESTELKTLKSVFAA ISAVATALILGPDCPEALVSEVTAGATAYNEVMIQCRILRAQYPKILKPLVWRFSRTARELRAI LSRLKARLVPEIKRRIAYLEAHSTSENQAESAGSFSLLDILIQTSFKNRHLPSTPVENDKWADL LCQQALLYHFKLSRAPGTSVTFMLYRVMNHPEYATMLRDEMIAALKPSGGNWTADILQRAPKLE SFNKETFRMHDISNFVGLRVAMRPVDLRSVSPPLHLKPGTMIMTPSRTVHYDAEHYVDPLTFNG LRFYDATSNTCTPRVFTTSPTYLPFSHGTGSCPARNFATQIARMLFIRLLMGYEFELANEEMPA YGLMDGTAYFPNPEVRMRVRVRGK
BicC (*Aspergillus alliaceus*)[23]	MTVMLIFYTSQLLGWGVTAVPVTFAVVLVFCLTVVTTDYIDGWRKRRAFGDIPIVDEGSNMSVR LRWNKPEFVAEREYAKAYKQYSKTGKPYVTRVQSNDYGIVLPLNSAKEWRALPHDQLSFGHGLA KMADMCMHLNVTDRTPVQALHLCNNARSLGRFNDLVASATDRVLPLTFGPATEQKFKELNVYKV

	ILSLCSAVAMSILLGPDFCMDAPLIQTTQMYHDSIMPSCYERTGYPQILRPFVWRFSPICRALR SRLSALKMKLIPEVKHRIELVRLGRGGVERDSTQSLLDALIETAFEKGSLSRSEQSTDDGAQVG LIVEQVIMFHFELCQPISFFLCSMLYVVMDHKEYVVPLREEASQTLRTSGGHWSLDCLKNAPKL ESFARETFRLYDISALTSFRRVMKPVHLKSISMFLKPGTLVLSPCRDVHLDPDHYDNPTIFDGY RFYDAIRESCNPHMTTTSLTFLSFSHGSTACPARVIASQMSRTLFMKILLKYDMELTHEKMPAY GFTHGPVYVPNLSIMMRVKPRSNDA
AfvE (*Aspergillus flavus*)[23]	MSACLRWWSRKFDCEKEYAEAYKQYSKTGKPYATRLKNNDHGIVLPLNSTKEWRTLPHDQLSFL HALSEFADLYMHINMTDRTPLQAVHYCNNTKTLSRFNRLMVDATDRALPLIVGKDTESEWKRAN AFHTILSLCSTVAMSVLLGPEFSMDTSLIQTIMMYNTAIMPSCAKRTSYPRILRPFVWRLSPLC RAMKSDLTKTKIKLTPEIKHRIDIARSKKGWLEEGPMSLLDGLIETAFEKGCLSRSSDRGDDDQ QVALLAEEIIFYHFELSTPVAFFIIFAVYVIMNNKEYSTPLREEISEALKLSGGSFTLDTLNHA PKLASFVKETCRFFRRVMKPIHLESINLSLKPGTIIMAPGRDVHLDPDYYDNPTTFNGYRFYDA SRGTCTPHISTTSPTFLTFSHGISACPARVLATQITRTIFIMFLLKFDVELAHEEMPAYGFANG PAYLPNPSVMMRVRPCQKDVLGV
AunB (*Aspergillus niger*)[56]	MVGWLIYRLHLDPLSRIPGPLLAKFVPICNIRMVLTGRMMFTFKELHDTYGPVVRIGPSELSFA THSAFDTIYGPYGDKNFSLYGSRKGLLGSLDKENRRKLRPLVATSLNELLAAKGEEYCHLAMDE QLASHHVGQTEATPVSLSTLNNRYLWQFASMAANGNCGDETDRGTSFTNIFMGLSFMALTQVIA ICFIFRQPVHPGEVNYDRDESPFPDNLHSRLQQAASRTSLHVPEYLLQINCFVLRFSIYGTADN MLNALFYFLLRNPQCLKRLEEEVSCVGATVNELSDDRLAKLPYLNACINETFRIAPAFNGGILQ RVSCGATVDGVYVPPGVAVSVDHYTLGHDPQYWVKPDVFNPERWIDPDCKDNFKASRPFLIGAR QCPGRQMAYQMFRVCVAKLVYLYTFELLNKDFDIERDTFSSYHWTGVKLDVTMKPRTPGVLGY
BfoB (*Aspergillus brasiliensis*)[56]	MLALYLIAGLLVGLLVYRLHLDPLSHIPGPFLAKFIPICNIRMLHTGRIVFTFRELHDTYGPVV RIGPSELSFATVTAFDSIYGFEGEKKFTIYGSRRGVVSSASGTDESLGNATSKESRRKLRPLIT STLNELMASSAEEYCHLALTEQLAAHRVGQDGSTPISLSTLNYRYLWQLANMVAFGNRGQEAHR ETFNPHIRWPSPFVSFIDLLFIFCSRATIQQHARTAYKAWQAIRFVCRQTTQGPQVDFIADDST FPDNLHRRLRQAAEKAGLDDVSDFTLLVNSMILRFSVYGTSDHMINAVFYYLLRHPQCLKRLEK EVLNAGTSVEELSDNRLAKLPYLNACINETFRISPAFNGGILQRVSCGATVDGVYVPPGVAVTV DNYTLGRSKQYWENPDAFCPERWLESSDKNVFKASRPFLIGSRQCPGRQMAYQMFRILVAKLVY LYSMELVNKDFDIERDTFCGLHWADLEVDAILKPRTDVLGY
ShwC (*Xylaria schweinitzii*)[57]	MLVLYGIILIVMGWLLYRIFLDPLSRIPGPFMSKFLPLYYIRMTWNGDHVFALKALHDTYGPVV RLGPNELSFATTTAFDSIYGFAADKTFVTGGSRRSLLVPERMGEVMLAQSGNKEVRNWLRPVII STLNELVGSCAERFLNIALATGLEQHQVGQERSISLSGLNDDWMWDLGSLLARGRHATRSTRIQ FDDLINDTEYFMCLLELCACVIHRFTLQRYQEHLHRLWEIVSKLFGHQILRDACIDNVMVESSG CPIAAEDSHSARIQEAIKKAGMEDMAGTYQDFNSLATQFNIYGSMHSALNTTFSYLLKHPRCLR KLEEELLHAFQSYDEVTDSRLAKLPYLNACITESFRIMPPFNAGILERVSMGSTVDGIYVPAGT GVQVDQFTLARSKEHWDAPEEFLPERWLDSSSRKNEKAFRPFLIGSRRCPGKNMAYQIFRLAIG KMFYLYEMELVNKDYHVNRHGESRLGLSNVKLIVEMNPRVPGVLGYSAS
NsrP (*Aspergillus novofumigatus* IBT 16806)[58]	MDFQIISRFTDGGDFQWTKFGTAAFLAVLLSALAFLSYTPRVHQKSPAFTSHKLPFIGSLGFTT EQWKARNWWKRATAESKTGNFSFWLGQRHVVGVTGEAARKMFFTHEALDFVSGALIRPINIHFW PPIHDIFRPDSKSRSKNTYFLRRLYELQSTEQLNHYLPQLLKDARVGMAGLSRVTKPSVSCWET VFTQDVRLLCTDEIVADSKLLATFGRQVETLLFTFSHYNVCFPWLPSPSYYKRRQARYALYNLM EDIVNKRLKNGARGPNDPVQILLDYNDKVDHIIEFFISVLFIAPANSRIIGGQMLNIMSIYRDW QEKVYADIKAAAAAHSPDKNAPLVDQLAFIPLHAWENSFPSIDLCLQETIRMWTSFSMARLNLS PNPIPIPGSDEVIPGNTFVCYNSTEVNFSDSLYPDPKKFDPARFLDGREEFRNEAYGFLGWGRG RHPCPGMRWAKLQQNIIIAYAVAMYDWSSCDETGKPTPQAVHVKELNAARGTVLPSAYCKLVPR EKV
ClaM (*Passalora fulva*)[59]	MNSTAANVKPEPGLYGPDAAASPAIARYLIQGPSPLSIGLVVLIGAISSFLLQQFLKPAVNPLS PKFTSNTWPWLGSLGMMIGHWTFWKSCIKESKTGQFSFWLGRTHVVGLSGPDARKMFFEHPALD LHKATDLTPLGVNFWPIHAIFTPRDEKFQKSTYWLRTLTHLMKTPNMERCLPGVIQDARTGLEA LKRDTPSGIINTPKVWPVVFKQTARVFMADDVTDDPKLWATTLGAIDTILHTFSPFNTLLPWIP EPSMIRRRMARRALLRVSRGIVRDRHASPKSASKNDAVQSLINLGDPDDNISEFMINAAFVGVV NAHVIFPQLLNALAVHLDWQDKVYKEVTEVADQHCKEQCLPLVDKLFHIPLSAWENSFPNAALM MEEISRIWTCFPTARFNTLYESIPIPGTSEVIPARTHAIYNSTEIHFNPKLYDRPASFRPDRFA PESQKTWDKEPHGLNTWGTGQRLCSGMRWAKLQQNILMAVALSLYRLEKCDKDGKPDPSAYEKQ QAMDGDLDEEVAFVLPECFVKLIPRE
Oxa09230 (*Penicillium oxalicum*)[60]	MDSTYADRFHDLDLPIKVAATGLLATVTTFLAYIVFLSYTPHVDKNSPQFTSNTVPLVGSWSFF TQKWAFWKDCVAKSKTGHFSFWLGKYHVVGVSGAAARKLFLDNPQFDFVRGATLVGHGPDLVPP LHEIFHVNFQNGRTYFQRRLLDVQKSEQLEKRVPGVTKDARMTFEALRQSASSIMNPTDTCYRL VVQQASRVVCSDEIADAPAVLAKTQSVLSLLMHTSSIHAVALPYLPWLSKLKRRWGRYGLSKIV TPIVQRRRQKDAPRKDDMVQYMLDRGDSPEWMVNFFISTLFISSANAGYLSGAMLNILAHHPDW

	QSRIYREIATVAETHASNRHAPLVDQLDSIPLEAWETHFPSIDLCFKEAIRMWVAFPMIRLNMA SHAVPIPGTNEVVPAGSFAAYNTTEVHFNPELYPDPYKYDPNRFSEGREEFKKEAYGFVGWGQG RHPCLGMRWARVQLNIILAYALAMYEWSGCDENGQPSSHFDRKTDLNAPGPSLPVGLYCKYVPR EKV
Sec05419 (*Claviceps purpurea*)[60]	MTDFKNDTIPDMSWTTGSNSGLLSSSPTTLALATLATGLAVFLVYLAYTPSVDALAPEFTSDTT PLIGSWGFYSRRWSFWRDSVARSKTGQFSFWLGKNHVVGVSGAAARKMFLDHQSLDRIKAAPLH GIGPEIVPPIHFVHQPNFSKGHSYFQRRVLDLMKTEHLASRLHAATREARTVFRGFPTKQTTLK GTRSGIVSPVDSCYRLVLSQSVRMMFCDELVDTHELFEKYVAFTHELQHLNSGHTCAVPWLPSL AHTKRRYYRHRLHCLFTPLVEKRLQSQQGRAGSRANDALQILVDNGDELDNIVTFLISVLFISI ANAGKLAGVLLNILCQNPIWQDRVLSEINAATPKFWPADKPATLVEKLDGMPLEVWEASFPFVE LIIREAIRMHVAFPMTRLNVSSRAIPIPGTGQVVPSGSFAAYNTNDAHLNEALYPDPLRFDPER FETPREGAKTETYGFLGWGNGRHRCVGQRWAKLQLSINLVYAVAMYKWNSCDADGNALPPVDHS FDRDRHGNQLAQGVFCKYECRE
DmxR5 (*Cryptosporiopsis* sp. 8999)[60]	MNSTPSAPPLKMEALDLDIFQPSLTLGLGAILVVLMSFLAFLSYTPSFDKKVPAFTTHTHPFIG AADFMWRKNHFWRASMEESKTGTFSFWVGKKHVVGLSGEAARKTFMDSPGLDFVGGARIRGVSL LPNPPTPEIFRPGFHHGRSFFLRRMIDMQKTEMLKRCLPRFTSDSRGVFDELAKDPSGITNPAV ACWHVVFAHDVILFCSEEIVDDPKAFPNLLQMIDILQGLSSFTKVFLPWLPTMAGMTRARRYKY MKSIFEPLVDRRMKRGAIRREDSLQLMIDNGDRRDLIIDFVIAAVIIAPTNSRILTGQMLNVMA AYPSWQEKAYAEIKALAKSYAKDPEAPLADQLDSMPLEAWEAGFPSIDLCFKELVRMWVAISMM RRNISSRTDEFIPAGSFAIYNTTETHFSEELYPDPHTFKPERWLEGNANVQKQAYGFVGWGEGR HPCPGKRWAKLQLNITIAYALAKFKWTSVDKTEHKKSTQESGHGVPLPSKLSKFSPREEN
AacuE (*Aspergillus aculeatus*)[60]	MKAAITGLVATVTTFLAYIVFLSYTPRVDKKSPQFTPNTVPLVGSWSFFTQKWAFWQDCVAKSQ TGHFSFWLGKYHVVGVSGAAARKVFLDNPNFDFVRGATLVGHGPDFVPPLHAIFHGNFQNGRSY FQRRLVDLQKSEQLGKRLPGVTRDARGAFEALRQHASGVMNPTDACYRLVVQQACRVVCSDEIA DDPAVLARTQSVLSLLMHTSSIHAVALPYLPSLAKLKRRWGRYGLSRIVTPIVQRRLQKKKDAP RHDDVVQYMLDRGDSPEWMVAFFISTLFIASANAGYLSGAMLNILAYHPDWQARIYREIKTVAA AHAPNPHAPLVDQLDTIPLDAWESFFPSIDLCFKEAIRMWVAFPMIRLNMAPHAVPIPGTDEVV PAGTFASYNSTEVHFNPELYPDPYRYDPERFREGREEFKKEVYGFVGWGQGRHPCLGMRWAKIQ LNIILAYALAMYEWSGCDEKGQPSTHFDRKTDLNAPGPSLPVGLFCKYVPRKEV
RugG (*Talaromyces* sp.)[61]	MDLNSQDFQSPVAIGLASFVATLISLLAYMSYTPRVGKGSPAFTPDTVPFIGSWRFFTQKLPFW RNSMRLSKTGNFSFWLGKNHVVGVSGEASRKMYLDSRGLHLIKGITLIGHGPDFIDGRSTVIHD IWKAAYTNDRTYATRRLLDLQKSEQLTKRLPRVTRDARAAFEGLAASPTGVMNPAKACYRLVVL QGSRLICNNEIADDPQLLNRLVAYVSTLQSTSSLHLLAIPWLSYLSVSYWKRRWGRNGITQIVK PIVNRRMTKGAPRVDDSLQFLIDSGDSKDYITNFLISMLFIVAANAGVLSGAMLNIVAHHPKWQ EKIYGEIKAVATAHSKNKNAPLVDQLDSIPLGAWETSFPSIDLCYKEAIRMWVAFPMGRFNDTN NAIPIPGTDEVIPPGSFACYNTLDVHYNEELYPDPMRWDPERFLDGREEYKKEAYGYMGWGAGR HPCTGMRWAKLQQNIILAYALALYKWSGCDENGDPNPHFVQPTTALNELAPSLPQGLWCKYVPR EKA

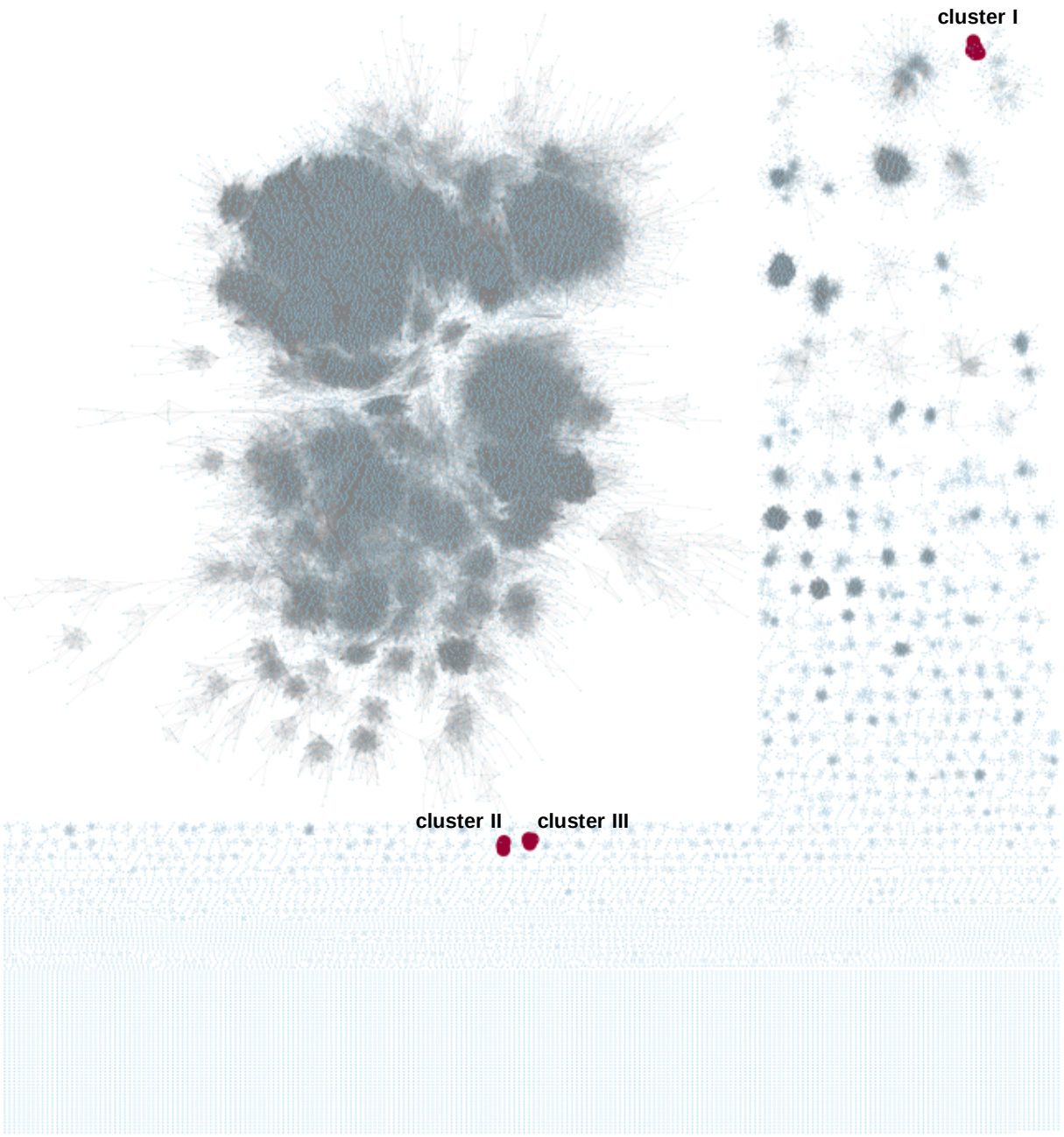


Figure 2.10. Sequence similarity network of full P450 family Nodes representing fungal P450s known to catalyze oxidative biaryl coupling reactions are highlighted with red spheres. Cluster I contains NsrP, ClaM, Oxa09230, Sec05419, DmxR5, Aa cuE, and RugG; Cluster II contains KtnC, DesC, BicC, and AfvE; Cluster III contains AunB, BfoB, and ShwC. SSN parameters: E-value = 100; minimum sequence length = 300; alignment score = 88.

SUPPLEMENTARY INFORMATION

The supplementary information for this chapter includes (I) DNA and protein sequences, (II) chemical synthesis of substrates and authentic product standards, (III) LC–MS traces of biocatalytic reactions, and (IV) NMR spectra of synthesized compounds.

DNA and protein sequences. The gene encoding KtnC in a pJET1.2 cloning plasmid and a pESC-HIS expression plasmid were generous gifts from Professor Michael Müller's lab at the Albert-Ludwigs-Universität Freiburg. The gene encoding KtnC was amplified and cloned into a pPIC3.5 plasmid (Invitrogen) for functional expression in *P. pastoris*.

KtnC sequence

```
ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG
```

```
MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGK
```

CPR sequence amplified from *P. pastoris* gDNA (*Pp*CPR)

ATGGATACCTTGGACCTTTCTGTTTTGATTGCTATTGCCTTGGCTTTGATAGCTTACTTTAGCAAAGGAT
CTCTTTGGGGCAAAGAGGATGATAACTCCGTCCACGGAGTAGCTGGAGGATTTCAGACCAGAGACCTGGT
TGAGATTTTGAACTCTACCAACAAGAAGGCACTCGTCCTTTACGGATCACAAACAGGTACATCTGAGGAT
TATGCTCATAAGTACGCTCGTGAACTGCAATCCAAGTTCAGCATTCCCACCCTCTGTGGAGACTTGTCAG
AGTTTGATTTTGACAATCTGAACGATATCCCAGAGCAGGTCGAAGGGTTCACTTTCATCACTTTCTTCAT
GGCTACGTACGGCGAAGGCGAGCCTACTGACAACGCCGTAGAGTTTATCGAGTTTCTTAAGAACGACGCC
GAGGATCTGTCTAATCTGAAGTATACTGTGTTTGGGTTGGGTAACTCTACTTACGAGTTCTATAATCAAA
TGGGTAAGACCACTAACAAGCGTTTCTCAGAACTTGGTGCTCAATTGGTTGGAACCTTTGGCGAGGGCGA
TGATGGTCAGGCAACTATGGACGAGGACTTCCTAGCATGGAAAGATTCCCTGTTTGATACAATCAAGAAG
GATTTGCATTTGGAGGAACATGAAGTTGTCTACCAGCCAGGTCTAAAAGTAAAGGAGAACACAGCGCTTA
CAACATCTTCTCCTAATGTCTCTGTTGGTGAACCTAATAAGGCTTATGTTCTTCGTGAAGATGAGAATCT
GCTGCAATACGGTCCCTTCGACCACACTCATCCTTACATTGCGCCCATCTCCAGCTCCAGAGAACTCTGT
TCTGAAACTTCGGAGAGAAACTGTATCCATCTAGAATTTGATCTCAGCAACACCAACTTGCGTTATTCCA
CTGGCGATCATTTGGCCGTGTGGCCAAGTAATGCTAATGAGCATGTCGAGAGCTTTTTGAAAGTGTTCAA
CCTCACAGACAAAAGATCTTCTGTGTTTGATATTGAGTTCCTGGACCCTACTGTCACAGTTCACTTTCCG
TTTCCAACCACATACGAAGCAGTTGTTCGTCATCACCTGGAAATTTCTGGACCAATCTCTAGACAGACAT
TGAAGCAATTCATTCCTTATGCTCCCGATCAGTCTACTAAACAGGAGGTAATCCGATTATCGGAGAGCAA
AGATGTGTTCCACAACGAAGTCACAGCAAAATACTATAACCTTGCTGACCTCTTGTTCAAAGTCTCAAAA
GAAACTCCATGGAATGTTCCATTCAACTTCCTGATCGAGACCATGCCTAATCTGCAGCATCGTTACTATT
CTATCTCATCATCTTCATTAAGTGAAAAGCAAACTATTCATATCACCGCAATGATTGAGGCCTTCACGCC
TACTGGAAGTGACCATATAGTGACAGGAGTCACCACTAATTTGCTATGGAACATTCAGCTGAATCAGGAC
AAATCTACTGTTAAGGCGCCTGTCTCCTATGATTTGAACGGCCCAAGAAATTTGTTTTCTCCATATAAAC
TTCCAGTTCATATCAGAAGGTCCACTTTCAAACTTCCAAGCAACCCGGCCCTGCCAGTGATCATGATTGG
ACCTGGTACTGGGGTAGCTCCATTCAGAGGCTTTATCAGAGAAAGATGCCAGCAAGTTGACAATGGAACT
CCTAATATTGGTCAATCTATCCTCTATTATGGCTGTAGAAATTCTGAGCAGGATTTCCTTTACCGTGACG
AATGGCCAACATATTCGAAGAAGTTAGGAGATAAGTTCAAGATGTACACTGCGTTCAGCAGAGAAAATTC
ACACAAAGTTTATGTTCAGCATAGGTTATTGGAAAACTCAAGGGAATTTATCGAACTGATGGACCAAGGT
GCTTTTATCTACGTATGTGGAGATGCTGGTAAAATGGCTAAGGATGTAAATAAGGCGATTGTTGAAATCC
TCATTAAAGAGAAAGGATTGTCAGAGGAAGATGCTACAGAAAGTATTAGAGAGTTCAAAACTTCCAACAG
ATATCAGGAAGACGTTTGGTAA

MDTLDLSVLIAIALALIAYFSKGSLWGKEDDNSVHGVAGGFQTRDLVEILNSTNKKALVLYGSQTGTSED
YAHKYARELQSKFSIPTLCGDLSEFDFDNLNDIPEQVEGFTFITFFMATYGEGEPTDNAVEFIEFLKNDA
EDLSNLKYTVFGLGNSTYEFYNQMGKTTNKRFSELGAQLVGTFGEGDDGQATMDEDFLAWKDSLFDTIKK
DLHLEEHEVVYQPGLKVKENTALTTSSPNVSVGEPNKAYVLREDENLLQYGPFDHTHPYIAPISSSRELC
SETSERNCIHLEFDLSNTNLRYSTGDHLAVWPSNANEHVESFLKVFNLTDKRSSVFDIEFLDPTVTVHFP
FPTTYEAVVRHHLEISGPISRQTLKQFIPYAPDQSTKQEVIRLSESKDVFHNEVTAKYYNLADLLFKVSK
ETPWNVPFNFLIETMPNLQHRYYSISSSSLSEKQTIHITAMIEAFTPTGSDHIVTGVTTNLLWNIQLNQD
KSTVKAPVSYDLNGPRNLFSPYKLPVHIRRSTFKLPSNPALPVIMIGPGTGVAPFRGFIRERCQQVDNGT
PNIGQSILYYGCRNSEQDFLYRDEWPTYSKKLGDKFKMYTAFSRENSHKVYVQHRLLENSREFIELMDQG
AFIYVCGDAGKMAKDVNKAIVEILIKEKGLSEEDATESIREFKTSNRYQEDVW

CPR sequence amplified from *S. cerevisiae* gDNA (*Sc*CPR)

ATGCCGTTTGGAATAGACAACACCGACTTCACTGTCCTGGCGGGGCTAGTGCTTGCCGTGCTACTGTACG
TAAAGAGAAACTCCATCAAGGAACTGCTGATGTCCGATGACGGAGATATCACAGCTGTCAGCTCGGGCAA
CAGAGACATTGCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGTTGTATGCGTCGCAGACTGGG
ACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTTCAACCTAAACGTGATGTGCG

CAGATGTTGAGAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATAGTCTCGATTTTTATCTCTAC
ATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTATTTGTAATGCGGAAGCGGGT
GCACTATCGAACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTATGAATTCTTTAATGGTGCCG
CCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGCAAGCTCGGTGAAGCTGATGA
TGGTGCAGGAACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCCTGGAGGTTTTGAAAGACGAA
CTGCATTTGGACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACACTGTGTTGAACGAAATCACTG
ACTCCATGTCGCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAGTTGAACCGCAACGCAGACGG
CATCCAATTGGGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCGTGAAATCTCGCGAACTGTTC
TCTTCCAATGACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTCTAACATCAAGTACTCCACTG
GTGACCATCTTGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAGTTCTTATCCATATTCAACCT
GGACCCTGAAACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAGTGCCCTTCCCAACGCCAACT
ACTATTGGCGCTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTCCAGACAATTGTTTTCATCTT
TGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTTTCGAAAGACAAGGACCAATT
CGCCGTCGAGATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAATATTTGTCTGATGGCGCCAAA
TGGGACACCGTACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGACTCCTCGTTACTACTCTATCT
CTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTGGAAAACTTTCCTAACCCAGA
ATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAAACATTCAATTGGCTCAAAAC
AATGTTAACATTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGGCCCACGTAAACTTTTCGCCA
ATTACAAATTGCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCCAACCCTTCCACCCCAGTTAT
CATGATCGGTCCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAGAGCGTGTCGCGTTCCTCGAA
TCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTTTTATGGATCCCGTAACACTG
ATGATTTCTTGTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGATGGTTCGTTCGAAATGGTCGT
GGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAATTAAAGGATTACGAAGACCAA
GTATTTGAAATGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGCAAAGGGTATGGCCAAGGGTG
TGTCAACCGCATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACTGATGAAGCAACAGAGCTAAT
CAAGATGCTCAAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MPFGIDNTDFTVLAGLVLAVLLYVKRNSIKELLMSDDGDITAVSSGNRDIAQVVTENNKNYLVLYASQTG
TAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVIVSIFISTYGEGDFPDGAVNFEDFICNAEAG
ALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLGKLGEADDGAGTTDEDYMAWKDSILEVLKDE
LHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQLNRNADGIQLGPFDLSQPYIAPIVKSRELF
SSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIFNLDPETIFDLKPLDPTVKVPFPTPT
TIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKDQFAVEITSKYFNIADALKYLSDGAK
WDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIVENFPNPELPDAPPVVGVTTNLLRNIQLAQN
NVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPSNPSTPVIMIGPGTGVAPFRGFIRERVAFLE
SQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLDGSFEMVVAHSRLPNTKKVYVQDKLKDYEDQ
VFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITTDEATELIKMLKTSGRYQEDVW

KtnC-2A-*Pp*CPR sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT

CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGAAGCTAGACATAAACAAAAGATTGTT
GCTCCAGTTAAACAAACTTTGAACTTTGATTTGTTGAAATTGGCTGGTGATGTTGAATCTAATCCAGGGC
CCGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGATACCTTGGACCTTTCTGTTTTGAT
TGCTATTGCCTTGGCTTTGATAGCTTACTTTAGCAAAGGATCTCTTTGGGGCAAAGAGGATGATAACTCC
GTCCACGGAGTAGCTGGAGGATTTCAGACCAGAGACCTGGTTGAGATTTTGAACTCTACCAACAAGAAGG
CACTCGTCCTTTACGGATCACAAACAGGTACATCTGAGGATTATGCTCATAAGTACGCTCGTGAACTGCA
ATCCAAGTTCAGCATTCCCACCCTCTGTGGAGACTTGTCAGAGTTTGATTTTGACAATCTGAACGATATC
CCAGAGCAGGTCGAAGGGTTCACTTTCATCACTTTCTTCATGGCTACGTACGGCGAAGGCGAGCCTACTG
ACAACGCCGTAGAGTTTATCGAGTTTCTTAAGAACGACGCCGAGGATCTGTCTAATCTGAAGTATACTGT
GTTTGGGTTGGGTAACTCTACTTACGAGTTCTATAATCAAATGGGTAAGACCACTAACAAGCGTTTCTCA
GAACTTGGTGCTCAATTGGTTGGAACCTTTGGCGAGGGCGATGATGGTCAGGCAACTATGGACGAGGACT
TCCTAGCATGGAAAGATTCCCTGTTTGATACAATCAAGAAGGATTTGCATTTGGAGGAACATGAAGTTGT
CTACCAGCCAGGTCTAAAAGTAAAGGAGAACACAGCGCTTACAACATCTTCTCCTAATGTCTCTGTTGGT
GAACCTAATAAGGCTTATGTTCTTCGTGAAGATGAGAATCTGCTGCAATACGGTCCCTTCGACCACACTC
ATCCTTACATTGCGCCCATCTCCAGCTCCAGAGAACTCTGTTCTGAAACTTCGGAGAGAAACTGTATCCA
TCTAGAATTTGATCTCAGCAACACCAACTTGCGTTATTCCACTGGCGATCATTTGGCCGTGTGGCCAAGT
AATGCTAATGAGCATGTCGAGAGCTTTTTGAAAGTGTTCAACCTCACAGACAAAAGATCTTCTGTGTTTG
ATATTGAGTTCCTGGACCCTACTGTCACAGTTCACTTTCCGTTTCCAACCACATACGAAGCAGTTGTTCG
TCATCACCTGGAAATTTCTGGACCAATCTCTAGACAGACATTGAAGCAATTCATTCCTTATGCTCCCGAT
CAGTCTACTAAACAGGAGGTAATCCGATTATCGGAGAGCAAAGATGTGTTCCACAACGAAGTCACAGCAA
AATACTATAACCTTGCTGACCTCTTGTTCAAAGTCTCAAAAGAAACTCCATGGAATGTTCCATTCAACTT
CCTGATCGAGACCATGCCTAATCTGCAGCATCGTTACTATTCTATCTCATCATCTTCATTAAGTGAAAAG
CAAACTATTCATATCACCGCAATGATTGAGGCCTTCACGCCTACTGGAAGTGACCATATAGTGACAGGAG
TCACCACTAATTTGCTATGGAACATTCAGCTGAATCAGGACAAATCTACTGTTAAGGCGCCTGTCTCCTA
TGATTTGAACGGCCCAAGAAATTTGTTTTCTCCATATAAACTTCCAGTTCATATCAGAAGGTCCACTTTC
AAACTTCCAAGCAACCCGGCCCTGCCAGTGATCATGATTGGACCTGGTACTGGGGTAGCTCCATTCAGAG
GCTTTATCAGAGAAAGATGCCAGCAAGTTGACAATGGAACTCCTAATATTGGTCAATCTATCCTCTATTA
TGGCTGTAGAAATTCTGAGCAGGATTTCCTTTACCGTGACGAATGGCCAACATATTCGAAGAAGTTAGGA
GATAAGTTCAAGATGTACACTGCGTTCAGCAGAGAAAATTCACACAAAGTTTATGTTCAGCATAGGTTAT
TGGAAAACTCAAGGGAATTTATCGAACTGATGGACCAAGGTGCTTTTATCTACGTATGTGGAGATGCTGG
TAAAATGGCTAAGGATGTAAATAAGGCGATTGTTGAAATCCTCATTAAAGAGAAAGGATTGTCAGAGGAA
GATGCTACAGAAAGTATTAGAGAGTTCAAAACTTCCAACAGATATCAGGAAGACGTTTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK

PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPREARHKQKIV
APVKQTLNFDLLKLAGDVESNPG *(ribosome skip)*
PDYKDDDDKIDPFTMDTLDLSVLIAIALALIAYFSKGSLWGKEDDNSVHGVAGGFQTRDLVEILNSTNKK
ALVLYGSQTGTSEDYAHKYARELQSKFSIPTLCGDLSEFDFDNLNDIPEQVEGFTFITFFMATYGEGEPT
DNAVEFIEFLKNDAEDLSNLKYTVFGLGNSTYEFYNQMGKTTNKRFSELGAQLVGTFGEGDDGQATMDED
FLAWKDSLFDTIKKDLHLEEHEVVYQPGLKVKENTALTTSSPNVSVGEPNKAYVLREDENLLQYGPFDHT
HPYIAPISSSRELCSETSERNCIHLEFDLSNTNLRYSTGDHLAVWPSNANEHVESFLKVFNLTDKRSSVF
DIEFLDPTVTVHFPFPTTYEAVVRHHLEISGPISRQTLKQFIPYAPDQSTKQEVIRLSESKDVFHNEVTA
KYYNLADLLFKVSKETPWNVPFNFLIETMPNLQHRYYSISSSSLSEKQTIHITAMIEAFTPTGSDHIVTG
VTTNLLWNIQLNQDKSTVKAPVSYDLNGPRNLFSPYKLPVHIRRSTFKLPSNPALPVIMIGPGTGVAPFR
GFIRERCQQVDNGTPNIGQSILYYGCRNSEQDFLYRDEWPTYSKKLGDKFKMYTAFSRENSHKVYVQHRL
LENSREFIELMDQGAFIYVCGDAGKMAKDVNKAIVEILIKEKGLSEEDATESIREFKTSNRYQEDVW

KtnC-2A-*Sc*CPR sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGAAGCTAGACATAAACAAAAGATTGTT
GCTCCAGTTAAACAAACTTTGAACTTTGATTTGTTGAAATTGGCTGGTGATGTTGAATCTAATCCAGGGC
CCGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGCCGTTTGGAATAGACAACACCGACTT
CACTGTCCTGGCGGGGCTAGTGCTTGCCGTGCTACTGTACGTAAAGAGAAACTCCATCAAGGAACTGCTG
ATGTCCGATGACGGAGATATCACAGCTGTCAGCTCGGGCAACAGAGACATTGCTCAGGTGGTGACCGAAA
ACAACAAGAACTACTTGGTGTTGTATGCGTCGCAGACTGGGACTGCCGAGGATTACGCCAAAAAGTTTTC
CAAGGAGCTGGTGGCCAAGTTCAACCTAAACGTGATGTGCGCAGATGTTGAGAACTACGACTTTGAGTCG
CTAAACGATGTGCCCGTCATAGTCTCGATTTTTATCTCTACATATGGTGAAGGAGACTTCCCCGACGGGG
CGGTCAACTTTGAAGACTTTATTTGTAATGCGGAAGCGGGTGCACTATCGAACCTGAGGTATAATATGTT
TGGTCTGGGAAATTCTACTTATGAATTCTTTAATGGTGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCC
GCGGGCGCTATCAGACTAGGCAAGCTCGGTGAAGCTGATGATGGTGCAGGAACTACAGACGAAGATTACA
TGGCCTGGAAGGACTCCATCCTGGAGGTTTTGAAAGACGAACTGCATTTGGACGAACAGGAAGCCAAGTT

CACCTCTCAATTCCAGTACACTGTGTTGAACGAAATCACTGACTCCATGTCGCTTGGTGAACCCTCTGCT
CCTATTTGCCCTCGCATCAGTTGAACCGCAACGCAGACGGCATCCAATTGGGTCCCTTCGATTTGTCTCA
ACCGTATATTGCACCCATCGTGAAATCTCGCGAACTGTTCTCTTCCAATGACCGTAATTGCATCCACTCT
GAATTTGACTTGTCCGGCTCTAACATCAAGTACTCCACTGGTGACCATCTTGCTGTTTGGCCTTCCAACC
CATTGGAAAAGGTCGAACAGTTCTTATCCATATTCAACCTGGACCCTGAAACCATTTTTGACTTGAAGCC
CCTGGATCCCACCGTCAAAGTGCCCTTCCCAACGCCAACTACTATTGGCGCTGCTATTAAACACTATTTG
GAAATTACAGGACCTGTCTCCAGACAATTGTTTTCATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCA
AGGAAAAATTGACTCTGCTTTCGAAAGACAAGGACCAATTCGCCGTCGAGATAACCTCCAAATATTTCAA
CATCGCAGATGCTCTGAAATATTTGTCTGATGGCGCCAAATGGGACACCGTACCCATGCAATTCTTGGTC
GAATCAGTTCCCCAAATGACTCCTCGTTACTACTCTATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCG
TCCATGTCACCTCCATTGTGGAAAACTTTCCTAACCCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGT
TACGACTAACTTGTTAAGAAACATTCAATTGGCTCAAAACAATGTTAACATTGCCGAAACTAACCTACCT
GTTCACTACGATTTAAATGGCCCACGTAAACTTTTCGCCAATTACAAATTGCCCGTCCACGTTCGTCGTT
CTAACTTCAGATTGCCTTCCAACCCTTCCACCCCAGTTATCATGATCGGTCCAGGTACCGGTGTTGCCCC
ATTCCGTGGGTTTATCAGAGAGCGTGTCGCGTTCCTCGAATCACAAAAGAAGGGCGGTAACAACGTTTCG
CTAGGTAAGCATATACTGTTTTATGGATCCCGTAACACTGATGATTTCTTGTACCAGGACGAATGGCCAG
AATACGCCAAAAAATTGGATGGTTCGTTCGAAATGGTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAA
AGTTTATGTTCAAGATAAATTAAAGGATTACGAAGACCAAGTATTTGAAATGATTAACAACGGTGCATTT
ATCTACGTCTGTGGTGATGCAAAGGGTATGGCCAAGGGTGTGTCAACCGCATTGGTTGGCATCTTATCCC
GTGGTAAATCCATTACCACTGATGAAGCAACAGAGCTAATCAAGATGCTCAAGACTTCAGGTAGATACCA
AGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPREARHKQKIV
APVKQTLNFDLLKLAGDVESNPG (ribosome skip)

PDYKDDDDKIDPFTMPFGIDNTDFTVLAGLVLAVLLYVKRNSIKELLMSDDGDITAVSSGNRDIAQVVTE
NNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVIVSIFISTYGEGDFPDG
AVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLGKLGEADDGAGTTDEDY
MAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQLNRNADGIQLGPFDLS
QPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIFNLDPETIFDLK
PLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKDQFAVEITSKYF
NIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIVENFPNPELPDAPPVVG
VTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPSNPSTPVIMIGPGTGVA
PFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLDGSFEMVVAHSRLPNTK
KVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITTDEATELIKMLKTSGRY
QEDVW

KtnC-*Pp*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT

GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGGAGTAGCTGGAGGATTTCAGACCAGA
GACCTGGTTGAGATTTTGAACTCTACCAACAAGAAGGCACTCGTCCTTTACGGATCACAAACAGGTACAT
CTGAGGATTATGCTCATAAGTACGCTCGTGAACTGCAATCCAAGTTCAGCATTCCCACCCTCTGTGGAGA
CTTGTCAGAGTTTGATTTTGACAATCTGAACGATATCCCAGAGCAGGTCGAAGGGTTCACTTTCATCACT
TTCTTCATGGCTACGTACGGCGAAGGCGAGCCTACTGACAACGCCGTAGAGTTTATCGAGTTTCTTAAGA
ACGACGCCGAGGATCTGTCTAATCTGAAGTATACTGTGTTTGGGTTGGGTAACTCTACTTACGAGTTCTA
TAATCAAATGGGTAAGACCACTAACAAGCGTTTCTCAGAACTTGGTGCTCAATTGGTTGGAACCTTTGGC
GAGGGCGATGATGGTCAGGCAACTATGGACGAGGACTTCCTAGCATGGAAAGATTCCCTGTTTGATACAA
TCAAGAAGGATTTGCATTTGGAGGAACATGAAGTTGTCTACCAGCCAGGTCTAAAAGTAAAGGAGAACAC
AGCGCTTACAACATCTTCTCCTAATGTCTCTGTTGGTGAACCTAATAAGGCTTATGTTCTTCGTGAAGAT
GAGAATCTGCTGCAATACGGTCCCTTCGACCACACTCATCCTTACATTGCGCCCATCTCCAGCTCCAGAG
AACTCTGTTCTGAAACTTCGGAGAGAAACTGTATCCATCTAGAATTTGATCTCAGCAACACCAACTTGCG
TTATTCCACTGGCGATCATTTGGCCGTGTGGCCAAGTAATGCTAATGAGCATGTCGAGAGCTTTTTGAAA
GTGTTCAACCTCACAGACAAAAGATCTTCTGTGTTTGATATTGAGTTCCTGGACCCTACTGTCACAGTTC
ACTTTCCGTTTCCAACCACATACGAAGCAGTTGTTCGTCATCACCTGGAAATTTCTGGACCAATCTCTAG
ACAGACATTGAAGCAATTCATTCCTTATGCTCCCGATCAGTCTACTAAACAGGAGGTAATCCGATTATCG
GAGAGCAAAGATGTGTTCCACAACGAAGTCACAGCAAAATACTATAACCTTGCTGACCTCTTGTTCAAAG
TCTCAAAAGAAACTCCATGGAATGTTCCATTCAACTTCCTGATCGAGACCATGCCTAATCTGCAGCATCG
TTACTATTCTATCTCATCATCTTCATTAAGTGAAAAGCAAACTATTCATATCACCGCAATGATTGAGGCC
TTCACGCCTACTGGAAGTGACCATATAGTGACAGGAGTCACCACTAATTTGCTATGGAACATTCAGCTGA
ATCAGGACAAATCTACTGTTAAGGCGCCTGTCTCCTATGATTTGAACGGCCCAAGAAATTTGTTTTCTCC
ATATAAACTTCCAGTTCATATCAGAAGGTCCACTTTCAAACTTCCAAGCAACCCGGCCCTGCCAGTGATC
ATGATTGGACCTGGTACTGGGGTAGCTCCATTCAGAGGCTTTATCAGAGAAAGATGCCAGCAAGTTGACA
ATGGAACTCCTAATATTGGTCAATCTATCCTCTATTATGGCTGTAGAAATTCTGAGCAGGATTTCCTTTA
CCGTGACGAATGGCCAACATATTCGAAGAAGTTAGGAGATAAGTTCAAGATGTACACTGCGTTCAGCAGA
GAAAATTCACACAAAGTTTATGTTCAGCATAGGTTATTGGAAAACTCAAGGGAATTTATCGAACTGATGG
ACCAAGGTGCTTTTATCTACGTATGTGGAGATGCTGGTAAAATGGCTAAGGATGTAAATAAGGCGATTGT
TGAAATCCTCATTAAAGAGAAAGGATTGTCAGAGGAAGATGCTACAGAAAGTATTAGAGAGTTCAAAACT
TCCAACAGATATCAGGAAGACGTTTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC

STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPRGVAGGFQTR
DLVEILNSTNKKALVLYGSQTGTSEDYAHKYARELQSKFSIPTLCGDLSEFDFDNLNDIPEQVEGFTFIT
FFMATYGEGEPTDNAVEFIEFLKNDAEDLSNLKYTVFGLGNSTYEFYNQMGKTTNKRFSELGAQLVGTFG
EGDDGQATMDEDFLAWKDSLFDTIKKDLHLEEHEVVYQPGLKVKENTALTTSSPNVSVGEPNKAYVLRED
ENLLQYGPFDHTHPYIAPISSSRELCSETSERNCIHLEFDLSNTNLRYSTGDHLAVWPSNANEHVESFLK
VFNLTDKRSSVFDIEFLDPTVTVHFPFPTTYEAVVRHHLEISGPISRQTLKQFIPYAPDQSTKQEVIRLS
ESKDVFHNEVTAKYYNLADLLFKVSKETPWNVPFNFLIETMPNLQHRYYSISSSSLSEKQTIHITAMIEA
FTPTGSDHIVTGVTTNLLWNIQLNQDKSTVKAPVSYDLNGPRNLFSPYKLPVHIRRSTFKLPSNPALPVI
MIGPGTGVAPFRGFIRERCQQVDNGTPNIGQSILYYGCRNSEQDFLYRDEWPTYSKKLGDKFKMYTAFSR
ENSHKVYVQHRLLENSREFIELMDQGAFIYVCGDAGKMAKDVNKAIVEILIKEKGLSEEDATESIREFKT
SNRYQEDVW

KtnC-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGCTGTCAGCTCGGGCAACAGAGACATT
GCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGTTGTATGCGTCGCAGACTGGGACTGCCGAGG
ATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTTCAACCTAAACGTGATGTGCGCAGATGTTGA
GAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATAGTCTCGATTTTTATCTCTACATATGGTGAA
GGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTATTTGTAATGCGGAAGCGGGTGCACTATCGA
ACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTATGAATTCTTTAATGGTGCCGCCAAGAAGGC
CGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGCAAGCTCGGTGAAGCTGATGATGGTGCAGGA
ACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCCTGGAGGTTTTGAAAGACGAACTGCATTTGG
ACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACACTGTGTTGAACGAAATCACTGACTCCATGTC
GCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAGTTGAACCGCAACGCAGACGGCATCCAATTG

GGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCGTGAAATCTCGCGAACTGTTCTCTTCCAATG
ACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTCTAACATCAAGTACTCCACTGGTGACCATCT
TGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAGTTCTTATCCATATTCAACCTGGACCCTGAA
ACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAGTGCCCTTCCCAACGCCAACTACTATTGGCG
CTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTCCAGACAATTGTTTTCATCTTTGATTCAGTT
CGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTTTCGAAAGACAAGGACCAATTCGCCGTCGAG
ATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAATATTTGTCTGATGGCGCCAAATGGGACACCG
TACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGACTCCTCGTTACTACTCTATCTCTTCCTCTTC
TCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTGGAAAACTTTCCTAACCCAGAATTGCCTGAT
GCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAAACATTCAATTGGCTCAAAACAATGTTAACA
TTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGGCCCACGTAAACTTTTCGCCAATTACAAATT
GCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCCAACCCTTCCACCCCAGTTATCATGATCGGT
CCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAGAGCGTGTCGCGTTCCTCGAATCACAAAAGA
AGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTTTTATGGATCCCGTAACACTGATGATTTCTT
GTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGATGGTTCGTTCGAAATGGTCGTGGCCCATTCC
AGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAATTAAAGGATTACGAAGACCAAGTATTTGAAA
TGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGCAAAGGGTATGGCCAAGGGTGTGTCAACCGC
ATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACTGATGAAGCAACAGAGCTAATCAAGATGCTC
AAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPRAVSSGNRDI
AQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVIVSIFISTYGE
GDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLGKLGEADDGAG
TTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQLNRNADGIQL
GPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIFNLDPE
TIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKDQFAVE
ITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIENFPNPELPDA
PPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPSNPSTPVIMIGP
GTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLDGSFEMVVAHSR
LPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITTDEATELIKMLK
TSGRYQEDVW

KtnC-[GGGGS]$_1$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA

CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGGTGGTGGTGGTTCTGCTGTCAGCTCG
GGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGTTGTATGCGTCGCAGA
CTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTTCAACCTAAACGTGAT
GTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATAGTCTCGATTTTTATC
TCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTATTTGTAATGCGGAAG
CGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTATGAATTCTTTAATGG
TGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGCAAGCTCGGTGAAGCT
GATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCCTGGAGGTTTTGAAAG
ACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACACTGTGTTGAACGAAAT
CACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAGTTGAACCGCAACGCA
GACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCGTGAAATCTCGCGAAC
TGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTCTAACATCAAGTACTC
CACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAGTTCTTATCCATATTC
AACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAGTGCCCTTCCCAACGC
CAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTCCAGACAATTGTTTTC
ATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTTTCGAAAGACAAGGAC
CAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAATATTTGTCTGATGGCG
CCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGACTCCTCGTTACTACTC
TATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTGGAAAACTTTCCTAAC
CCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAAACATTCAATTGGCTC
AAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGGCCCACGTAAACTTTT
CGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCCAACCCTTCCACCCCA
GTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAGAGCGTGTCGCGTTCC
TCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTTTTATGGATCCCGTAA
CACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGATGGTTCGTTCGAAATG
GTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAATTAAAGGATTACGAAG
ACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGCAAAGGGTATGGCCAA
GGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACTGATGAAGCAACAGAG
CTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPRGGGGSAVSS

GNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVIVSIFI
STYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLGKLGEA
DDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQLNRNA
DGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIF
NLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKD
QFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIVENFPN
PELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPSNPSTP
VIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLDGSFEM
VVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITTDEATE
LIKMLKTSGRYQEDVW

KtnC-[GGGGS]$_2$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGGTGGTGGTGGTTCTGGTGGTGGTGGT
TCTGCTGTCAGCTCGGGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGT
TGTATGCGTCGCAGACTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTT
CAACCTAAACGTGATGTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATA
GTCTCGATTTTTATCTCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTA
TTTGTAATGCGGAAGCGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTA
TGAATTCTTTAATGGTGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGC
AAGCTCGGTGAAGCTGATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCC
TGGAGGTTTTGAAAGACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACAC
TGTGTTGAACGAAATCACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAG
TTGAACCGCAACGCAGACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCG
TGAAATCTCGCGAACTGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTC
TAACATCAAGTACTCCACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAG
TTCTTATCCATATTCAACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAG
TGCCCTTCCCAACGCCAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTC

CAGACAATTGTTTTCATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTT
TCGAAAGACAAGGACCAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAAT
ATTTGTCTGATGGCGCCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGAC
TCCTCGTTACTACTCTATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTG
GAAAACTTTCCTAACCCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAA
ACATTCAATTGGCTCAAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGG
CCCACGTAAACTTTTCGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCC
AACCCTTCCACCCCAGTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAG
AGCGTGTCGCGTTCCTCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTT
TTATGGATCCCGTAACACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGAT
GGTTCGTTCGAAATGGTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAAT
TAAAGGATTACGAAGACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGC
AAAGGGTATGGCCAAGGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACT
GATGAAGCAACAGAGCTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPRGGGGSGGGG
SAVSSGNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVI
VSIFISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLG
KLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQ
LNRNADGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQ
FLSIFNLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLL
SKDKDQFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIV
ENFPNPELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPS
NPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLD
GSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITT
DEATELIKMLKTSGRYQEDVW

KtnC-[GGGGS]$_3$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA

TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGGTGGTGGTGGTTCTGGTGGTGGTGGT
TCTGGTGGTGGTGGTTCTGCTGTCAGCTCGGGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACA
AGAACTACTTGGTGTTGTATGCGTCGCAGACTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGA
GCTGGTGGCCAAGTTCAACCTAAACGTGATGTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAAC
GATGTGCCCGTCATAGTCTCGATTTTTATCTCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCA
ACTTTGAAGACTTTATTTGTAATGCGGAAGCGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCT
GGGAAATTCTACTTATGAATTCTTTAATGGTGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGC
GCTATCAGACTAGGCAAGCTCGGTGAAGCTGATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCT
GGAAGGACTCCATCCTGGAGGTTTTGAAAGACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTC
TCAATTCCAGTACACTGTGTTGAACGAAATCACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTAT
TTGCCCTCGCATCAGTTGAACCGCAACGCAGACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGT
ATATTGCACCCATCGTGAAATCTCGCGAACTGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATT
TGACTTGTCCGGCTCTAACATCAAGTACTCCACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTG
GAAAAGGTCGAACAGTTCTTATCCATATTCAACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGG
ATCCCACCGTCAAAGTGCCCTTCCCAACGCCAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAAT
TACAGGACCTGTCTCCAGACAATTGTTTTCATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAA
AAATTGACTCTGCTTTCGAAAGACAAGGACCAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCG
CAGATGCTCTGAAATATTTGTCTGATGGCGCCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATC
AGTTCCCCAAATGACTCCTCGTTACTACTCTATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCAT
GTCACCTCCATTGTGGAAAACTTTCCTAACCCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGA
CTAACTTGTTAAGAAACATTCAATTGGCTCAAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCA
CTACGATTTAAATGGCCCACGTAAACTTTTCGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAAC
TTCAGATTGCCTTCCAACCCTTCCACCCCAGTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCC
GTGGGTTTATCAGAGAGCGTGTCGCGTTCCTCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGG
TAAGCATATACTGTTTTATGGATCCCGTAACACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATAC
GCCAAAAAATTGGATGGTTCGTTCGAAATGGTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTT
ATGTTCAAGATAAATTAAAGGATTACGAAGACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTA
CGTCTGTGGTGATGCAAAGGGTATGGCCAAGGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGT
AAATCCATTACCACTGATGAAGCAACAGAGCTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAG
ATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPRGGGGSGGGG
SGGGGSAVSSGNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLN
DVPVIVSIFISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAG
AIRLGKLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHY
LPSHQLNRNADGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPL

EKVEQFLSIFNLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKE
KLTLLSKDKDQFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVH
VTSIVENFPNPELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSN
FRLPSNPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEY
AKKLDGSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRG
KSITTDEATELIKMLKTSGRYQEDVW

KtnC-[EPPPP]$_1$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGAACCTCCTCCTCCTGCTGTCAGCTCG
GGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGTTGTATGCGTCGCAGA
CTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTTCAACCTAAACGTGAT
GTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATAGTCTCGATTTTTATC
TCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTATTTGTAATGCGGAAG
CGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTATGAATTCTTTAATGG
TGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGCAAGCTCGGTGAAGCT
GATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCCTGGAGGTTTTGAAAG
ACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACACTGTGTTGAACGAAAT
CACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAGTTGAACCGCAACGCA
GACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCGTGAAATCTCGCGAAC
TGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTCTAACATCAAGTACTC
CACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAGTTCTTATCCATATTC
AACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAGTGCCCTTCCCAACGC
CAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTCCAGACAATTGTTTTC
ATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTTTCGAAAGACAAGGAC
CAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAATATTTGTCTGATGGCG
CCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGACTCCTCGTTACTACTC
TATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTGGAAAACTTTCCTAAC

CCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAAACATTCAATTGGCTC
AAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGGCCCACGTAAACTTTT
CGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCCAACCCTTCCACCCCA
GTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAGAGCGTGTCGCGTTCC
TCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTTTTATGGATCCCGTAA
CACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGATGGTTCGTTCGAAATG
GTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAATTAAAGGATTACGAAG
ACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGCAAAGGGTATGGCCAA
GGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACTGATGAAGCAACAGAG
CTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPREPPPPAVSS
GNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVIVSIFI
STYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLGKLGEA
DDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQLNRNA
DGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQFLSIF
NLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLLSKDKD
QFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIVENFPN
PELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPSNPSTP
VIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLDGSFEM
VVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITTDEATE
LIKMLKTSGRYQEDVW

KtnC-[EPPPP]$_2$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG

TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGAACCTCCTCCTCCTGAACCTCCTCCT
CCTGCTGTCAGCTCGGGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACAAGAACTACTTGGTGT
TGTATGCGTCGCAGACTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGAGCTGGTGGCCAAGTT
CAACCTAAACGTGATGTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAACGATGTGCCCGTCATA
GTCTCGATTTTTATCTCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCAACTTTGAAGACTTTA
TTTGTAATGCGGAAGCGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCTGGGAAATTCTACTTA
TGAATTCTTTAATGGTGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGCGCTATCAGACTAGGC
AAGCTCGGTGAAGCTGATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCTGGAAGGACTCCATCC
TGGAGGTTTTGAAAGACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTCTCAATTCCAGTACAC
TGTGTTGAACGAAATCACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTATTTGCCCTCGCATCAG
TTGAACCGCAACGCAGACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGTATATTGCACCCATCG
TGAAATCTCGCGAACTGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATTTGACTTGTCCGGCTC
TAACATCAAGTACTCCACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTGGAAAAGGTCGAACAG
TTCTTATCCATATTCAACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGGATCCCACCGTCAAAG
TGCCCTTCCCAACGCCAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAATTACAGGACCTGTCTC
CAGACAATTGTTTTCATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAAAAATTGACTCTGCTT
TCGAAAGACAAGGACCAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCGCAGATGCTCTGAAAT
ATTTGTCTGATGGCGCCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATCAGTTCCCCAAATGAC
TCCTCGTTACTACTCTATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCATGTCACCTCCATTGTG
GAAAACTTTCCTAACCCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGACTAACTTGTTAAGAA
ACATTCAATTGGCTCAAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCACTACGATTTAAATGG
CCCACGTAAACTTTTCGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAACTTCAGATTGCCTTCC
AACCCTTCCACCCCAGTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCCGTGGGTTTATCAGAG
AGCGTGTCGCGTTCCTCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGGTAAGCATATACTGTT
TTATGGATCCCGTAACACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATACGCCAAAAAATTGGAT
GGTTCGTTCGAAATGGTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTTATGTTCAAGATAAAT
TAAAGGATTACGAAGACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTACGTCTGTGGTGATGC
AAAGGGTATGGCCAAGGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGTAAATCCATTACCACT
GATGAAGCAACAGAGCTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAGATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPREPPPPEPPP
PAVSSGNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLNDVPVI
VSIFISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAGAIRLG
KLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHYLPSHQ
LNRNADGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPLEKVEQ
FLSIFNLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKEKLTLL
SKDKDQFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSEKQTVHVTSIV
ENFPNPELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSNFRLPS
NPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEYAKKLD
GSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRGKSITT
DEATELIKMLKTSGRYQEDVW

KtnC-[EPPPP]$_3$-*Sc*CPR fusion sequence

ATGTCTGACTACAAAGACGATGACGACAAGATTGATCCCTTCACCATGGCAGTTCATGTACCATTCAGCA
TCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCATTGTCCTCGTTCTCCTATTCGGGTT
GGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAGAGCCTTGAGAGGCATTCCGATTGTA
GACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTCGATGCGGAGAAAGAGTATGCGAGAG
CATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGATGCAGAATGATAATTATGGCATTGT
GCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGACCAGCTGAGCTTTCTTCAAGCCTTG
GCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGGACACCCATTGAAGCTGTTCATAGTT
GCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCCGGGAAACCGATACAGCCTTATCTCA
GATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACACACTGCAGACAATCCTCTCCTTATGC
TCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCTGACCCGGTGCTCCACCACCATTCGA
CGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGGGATATCCGCGCATCCTGCGGCCATT
CGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTTATCTCTCGTGAGAGAGAGACTTGTT
CCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAAACCAAGGACGTACGCCCATCTTCAT
TGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAAGCCCAGATGACCAAGGCAGAAATGA
TGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCATTTCGAACTCTGTAAACCGACCGCA
TTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAGTACAAGGCTCCTCTCCGAGAGGAGG
CACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAACTCTGAAGCACGCTCCCAAGTTAGA
AAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTTTGTCAGCTTTCGCCGTGTCATGAAA
CCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACAATCTTATTGTCTCCATGTCGGAATG
TCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATGGCTACCGCTTCTACGACTCCAGCCG
CGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTTGACATTTTCTCATGGAGCGGGCAGC
TGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTCATCAAGTTCCTGTTGCAGTACGACG
TAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTGGTCCGGTCTATATGCCTAATCCATC
AGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAGCCTAGGGAACCTCCTCCTCCTGAACCTCCTCCT
CCTGAACCTCCTCCTCCTGCTGTCAGCTCGGGCAACAGAGACATTGCTCAGGTGGTGACCGAAAACAACA
AGAACTACTTGGTGTTGTATGCGTCGCAGACTGGGACTGCCGAGGATTACGCCAAAAAGTTTTCCAAGGA
GCTGGTGGCCAAGTTCAACCTAAACGTGATGTGCGCAGATGTTGAGAACTACGACTTTGAGTCGCTAAAC
GATGTGCCCGTCATAGTCTCGATTTTTATCTCTACATATGGTGAAGGAGACTTCCCCGACGGGGCGGTCA
ACTTTGAAGACTTTATTTGTAATGCGGAAGCGGGTGCACTATCGAACCTGAGGTATAATATGTTTGGTCT
GGGAAATTCTACTTATGAATTCTTTAATGGTGCCGCCAAGAAGGCCGAGAAGCATCTCTCCGCCGCGGGC
GCTATCAGACTAGGCAAGCTCGGTGAAGCTGATGATGGTGCAGGAACTACAGACGAAGATTACATGGCCT
GGAAGGACTCCATCCTGGAGGTTTTGAAAGACGAACTGCATTTGGACGAACAGGAAGCCAAGTTCACCTC
TCAATTCCAGTACACTGTGTTGAACGAAATCACTGACTCCATGTCGCTTGGTGAACCCTCTGCTCACTAT
TTGCCCTCGCATCAGTTGAACCGCAACGCAGACGGCATCCAATTGGGTCCCTTCGATTTGTCTCAACCGT
ATATTGCACCCATCGTGAAATCTCGCGAACTGTTCTCTTCCAATGACCGTAATTGCATCCACTCTGAATT
TGACTTGTCCGGCTCTAACATCAAGTACTCCACTGGTGACCATCTTGCTGTTTGGCCTTCCAACCCATTG
GAAAAGGTCGAACAGTTCTTATCCATATTCAACCTGGACCCTGAAACCATTTTTGACTTGAAGCCCCTGG
ATCCCACCGTCAAAGTGCCCTTCCCAACGCCAACTACTATTGGCGCTGCTATTAAACACTATTTGGAAAT
TACAGGACCTGTCTCCAGACAATTGTTTTCATCTTTGATTCAGTTCGCCCCCAACGCTGACGTCAAGGAA
AAATTGACTCTGCTTTCGAAAGACAAGGACCAATTCGCCGTCGAGATAACCTCCAAATATTTCAACATCG
CAGATGCTCTGAAATATTTGTCTGATGGCGCCAAATGGGACACCGTACCCATGCAATTCTTGGTCGAATC
AGTTCCCCAAATGACTCCTCGTTACTACTCTATCTCTTCCTCTTCTCTGTCTGAAAAGCAAACCGTCCAT
GTCACCTCCATTGTGGAAAACTTTCCTAACCCAGAATTGCCTGATGCTCCTCCAGTTGTTGGTGTTACGA
CTAACTTGTTAAGAAACATTCAATTGGCTCAAAACAATGTTAACATTGCCGAAACTAACCTACCTGTTCA
CTACGATTTAAATGGCCCACGTAAACTTTTCGCCAATTACAAATTGCCCGTCCACGTTCGTCGTTCTAAC
TTCAGATTGCCTTCCAACCCTTCCACCCCAGTTATCATGATCGGTCCAGGTACCGGTGTTGCCCCATTCC
GTGGGTTTATCAGAGAGCGTGTCGCGTTCCTCGAATCACAAAAGAAGGGCGGTAACAACGTTTCGCTAGG
TAAGCATATACTGTTTTATGGATCCCGTAACACTGATGATTTCTTGTACCAGGACGAATGGCCAGAATAC

GCCAAAAAATTGGATGGTTCGTTCGAAATGGTCGTGGCCCATTCCAGGTTGCCAAACACCAAAAAAGTTT
ATGTTCAAGATAAATTAAAGGATTACGAAGACCAAGTATTTGAAATGATTAACAACGGTGCATTTATCTA
CGTCTGTGGTGATGCAAAGGGTATGGCCAAGGGTGTGTCAACCGCATTGGTTGGCATCTTATCCCGTGGT
AAATCCATTACCACTGATGAAGCAACAGAGCTAATCAAGATGCTCAAGACTTCAGGTAGATACCAAGAAG
ATGTCTGGTAA

MSDYKDDDDKIDPFTMAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIV
DEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQAL
AEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLC
STVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLV
PEVARRVAAARAADKTKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTA
FNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMK
PLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGS
CPARVLATQICRTIFIKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGKPREPPPPEPPP
PEPPPPAVSSGNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVAKFNLNVMCADVENYDFESLN
DVPVIVSIFISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNSTYEFFNGAAKKAEKHLSAAG
AIRLGKLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQYTVLNEITDSMSLGEPSAHY
LPSHQLNRNADGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLSGSNIKYSTGDHLAVWPSNPL
EKVEQFLSIFNLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGPVSRQLFSSLIQFAPNADVKE
KLTLLSKDKDQFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQMTPRYYSISSSSLSLSEKQTVH
VTSIVENFPNPELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDLNGPRKLFANYKLPVHVRRSN
FRLPSNPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHILFYGSRNTDDFLYQDEWPEY
AKKLDGSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCGDAKGMAKGVSTALVGILSRG
KSITTDEATELIKMLKTSGRYQEDVW

Δ2-20-KtnC sequence (optimized for expression in *E. coli*)

ATGGTTGTTATTGTTCTGGTGCTGCTGTTTGGTCTGGCAGTTGTTGGTAGCGATAGCCTGGATGGTTGGT
GGCAGAAACGTGCACTGCGTGGTATTCCGATTGTTGATGAAGGTAGCTATATGCGTCCGAAACTGCGTTG
GAAACGTTTTGATGCAGAAAAAGAATATGCACGTGCCTATCAGCAGTATACCAAAGCAGGTAAACCGTAT
GCAATTCGCATGCAGAATGATAACTATGGTATTGTGCTGCCGCTGAATAGCGCAAAAGAATGGCGTAGCC
TGCCGCATGATCAGCTGAGCTTTCTGCAGGCACTGGCAGAATTTGCAGATATGAATATGTATTGCGACGT
TACCGATCGTACCCCGATTGAAGCAGTTCATAGCTGTAATAATGCCGAAAGCCTGAACATTCTGAATAAA
CTGCTGGCACGTGAAACCGATACCGCACTGAGCCAGATTTTTGAACAGCCGACCGGTAAAGATTGGAAAG
AACTGAATACCCTGCAGACCATTCTGAGCCTGTGTAGCACCGTTACCATGGCACTGCTGCTGGGTCCTGA
TACAGCACCGGATCCTGTTCTGCACCATCATAGCACCAGCTTTGGTGAAGCAATTATGAGCAGCTGTTAT
CGTCGTACCGGTTATCCGCGTATTCTGCGTCCGTTTGTTTGGCGTTTTAGCAGCGAATGTCGTAATCTGC
GTAAACATCTGAGTCTGGTTCGTGAACGTCTGGTTCCGGAAGTTGCACGTCGTGTTGCAGCAGCACGTGC
AGCAGATAAAACCAAAGATGTTCGTCCGAGCAGCCTGCTGGATGCACTGATTGCAGCCGCATTTGATAAT
GGTAGCCTGAGTCCGGATGATCAGGGTCGTAATGATGCAGCACAGGTTCAGCTGCTGGCAGATGATCTGA
TCTTTTATCATTTTGAACTGTGTAAACCGACCGCCTTCAACATTATCTTTCAGCTGTATGCCATTATGGA
TCACCCGGAATATAAAGCACCGCTGCGTGAAGAAGCACTGCAGGCCCTGAAACTGACCAATGGTGATTGG
ACCGTTGAAACCCTGAAACATGCACCGAAACTGGAAAGCTTTACCAAAGAAACCTTTCGCCTGTATGACA
TTAGCGGTTTTGTTAGCTTTCGTCGTGTTATGAAACCGCTGACACTGAATAGCATTGGTCTGTCACTGCG
TCCGGGTACAATTCTGCTGAGCCCGTGTCGTAATGTTCATCTGGATCCGGAAATTTATGAAGATCCGACC
ACCTTTAATGGCTACCGTTTTTATGATAGCAGCCGTGAAGTTTGTAGTCCGCGTGTTGCAACCACCAGTC
TGACCTTTCTGACATTTAGCCATGGTGCAGGTAGCTGTCCGGCACGTGTTCTGGCAACCCAGATTTGTCG
TACCATCTTTATCAAATTTCTGCTGCAGTATGATGTTGAGCCGGTGCAGAAAGAAATCCTGCCGTATGGT
TTTACCAGCGGTCCGGTTTATATGCCGAATCCGAGCGTTATGATGCGTATTCGTCCGCGTAGTGATGGTA
AA

MVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPY
AIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNK
LLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCY
RRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADKTKDVRPSSLLDALIAAAFDN
GSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTAFNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDW
TVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPT
TFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIFIKFLLQYDVEPVQKEILPYG
FTSGPVYMPNPSVMMRIRPRSDGK

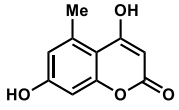
Δ2-20-KtnC-*Sc*CPR fusion sequence (optimized for expression in *E. coli*)

ATGGTTGTTATTGTTCTGGTGCTGCTGTTTGGTCTGGCAGTTGTTGGTAGCGATAGCCTGGATGGTTGGT
GGCAGAAACGTGCACTGCGTGGTATTCCGATTGTTGATGAAGGTAGCTATATGCGTCCGAAACTGCGTTG
GAAACGTTTTGATGCAGAAAAAGAATATGCACGTGCCTATCAGCAGTATACCAAAGCAGGTAAACCGTAT
GCAATTCGCATGCAGAATGATAACTATGGTATTGTGCTGCCGCTGAATAGCGCAAAAGAATGGCGTAGCC
TGCCGCATGATCAGCTGAGCTTTCTGCAGGCACTGGCAGAATTTGCAGATATGAATATGTATTGCGACGT
TACCGATCGTACCCCGATTGAAGCAGTTCATAGCTGTAATAATGCCGAAAGCCTGAACATTCTGAATAAA
CTGCTGGCACGTGAAACCGATACCGCACTGAGCCAGATTTTTGAACAGCCGACCGGTAAAGATTGGAAAG
AACTGAATACCCTGCAGACCATTCTGAGCCTGTGTAGCACCGTTACCATGGCACTGCTGCTGGGTCCTGA
TACAGCACCGGATCCTGTTCTGCACCATCATAGCACCAGCTTTGGTGAAGCAATTATGAGCAGCTGTTAT
CGTCGTACCGGTTATCCGCGTATTCTGCGTCCGTTTGTTTGGCGTTTTAGCAGCGAATGTCGTAATCTGC
GTAAACATCTGAGTCTGGTTCGTGAACGTCTGGTTCCGGAAGTTGCACGTCGTGTTGCAGCAGCACGTGC
AGCAGATAAAACCAAAGATGTTCGTCCGAGCAGCCTGCTGGATGCACTGATTGCAGCCGCATTTGATAAT
GGTAGCCTGAGTCCGGATGATCAGGGTCGTAATGATGCAGCACAGGTTCAGCTGCTGGCAGATGATCTGA
TCTTTTATCATTTTGAACTGTGTAAACCGACCGCCTTCAACATTATCTTTCAGCTGTATGCCATTATGGA
TCACCCGGAATATAAAGCACCGCTGCGTGAAGAAGCACTGCAGGCCCTGAAACTGACCAATGGTGATTGG
ACCGTTGAAACCCTGAAACATGCACCGAAACTGGAAAGCTTTACCAAAGAAACCTTTCGCCTGTATGACA
TTAGCGGTTTTGTTAGCTTTCGTCGTGTTATGAAACCGCTGACACTGAATAGCATTGGTCTGTCACTGCG
TCCGGGTACAATTCTGCTGAGCCCGTGTCGTAATGTTCATCTGGATCCGGAAATTTATGAAGATCCGACC
ACCTTTAATGGCTACCGTTTTTATGATAGCAGCCGTGAAGTTTGTAGTCCGCGTGTTGCAACCACCAGTC
TGACCTTTCTGACATTTAGCCATGGTGCAGGTAGCTGTCCGGCACGTGTTCTGGCAACCCAGATTTGTCG
TACCATCTTTATCAAATTTCTGCTGCAGTATGATGTTGAGCCGGTGCAGAAAGAAATCCTGCCGTATGGT
TTTACCAGCGGTCCGGTTTATATGCCGAATCCGAGCGTTATGATGCGTATTCGTCCGCGTAGTGATGGTA
AACCGCGTGCAGTTAGCAGCGGTAATCGTGATATTGCACAGGTTGTTACCGAGAACAACAAAAACTATCT
GGTTCTGTATGCAAGCCAGACCGGTACAGCCGAAGATTATGCCAAAAAATTCAGCAAAGAACTGGTGGCC
AAATTCAATCTGAATGTTATGTGTGCCGATGTCGAGAACTATGATTTTGAAAGTCTGAATGATGTGCCGG
TGATTGTGAGCATTTTTATCAGCACCTATGGTGAAGGTGATTTTCCGGATGGTGCCGTTAATTTTGAAGA
TTTCATTTGTAATGCGGAAGCGGGTGCACTGAGCAATCTGCGCTATAATATGTTTGGCCTGGGTAATAGT
ACCTACGAATTTTTCAATGGCGCAGCCAAAAAAGCCGAAAAACACCTGAGCGCAGCCGGTGCCATTCGTC
TGGGTAAACTGGGTGAAGCAGATGATGGCGCAGGCACCACCGATGAAGATTATATGGCATGGAAAGATAG
CATTCTGGAAGTGCTGAAAGACGAACTGCATCTGGATGAACAAGAAGCAAAATTTACCAGCCAGTTTCAG
TATACCGTGCTGAATGAAATCACCGATAGCATGAGCCTGGGTGAACCGAGCGCACATTATCTGCCGAGCC
ATCAGCTGAATCGTAATGCAGATGGTATTCAGCTGGGTCCGTTTGATCTGAGCCAGCCGTATATTGCACC
GATTGTGAAAAGCCGTGAACTGTTTAGCTCAAATGATCGCAATTGCATCCACAGCGAATTTGATCTGTCA
GGCAGCAATATCAAATATAGCACCGGTGATCATCTGGCCGTTTGGCCGAGCAATCCGCTGGAAAAAGTTG
AACAGTTTCTGAGCATCTTTAACCTGGATCCTGAAACCATTTTCGATCTGAAACCACTGGATCCGACCGT
TAAAGTTCCGTTTCCGACACCGACCACCATTGGCGCAGCAATTAAACACTATCTGGAAATCACAGGTCCG
GTTAGCCGTCAGCTGTTTAGTAGCCTGATTCAGTTTGCACCGAATGCCGATGTTAAAGAAAAACTGACCC
TGCTGAGCAAAGATAAAGATCAGTTTGCCGTTGAGATCACCAGCAAGTATTTTAACATTGCAGATGCCCT
GAAATATCTGAGTGATGGTGCCAAATGGGATACCGTTCCGATGCAGTTTCTGGTTGAAAGCGTTCCGCAG
ATGACACCGCGTTATTATAGCATTAGCAGCAGCCTGAGCGAAAAACAGACCGTTCATGTTACCAGCA

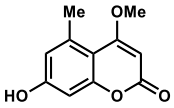
TTGTTGAAAACTTTCCGAATCCGGAACTGCCGGATGCACCGCCTGTTGTTGGTGTTACCACCAATCTGCT
GCGCAATATCCAGCTGGCACAGAATAATGTTAATATCGCAGAAACGAATCTGCCGGTGCATTATGATCTG
AATGGTCCGCGTAAACTGTTCGCAAACTATAAACTGCCGGTTCATGTTCGTCGTAGCAATTTTCGTCTGC
CGAGTAATCCGAGCACTCCGGTTATTATGATTGGTCCTGGCACCGGTGTTGCACCGTTTCGTGGTTTTAT
TCGCGAACGTGTTGCATTTCTGGAAAGCCAGAAAAAAGGTGGTAATAATGTGAGTCTGGGCAAACACATT
CTGTTTTATGGTAGCCGTAACACCGATGATTTTCTGTATCAAGATGAATGGCCTGAGTACGCCAAAAAAC
TGGATGGTAGCTTTGAAATGGTTGTTGCACATAGTCGCCTGCCGAATACCAAAAAAGTTTATGTTCAGGA
CAAGCTGAAGGACTATGAGGATCAGGTTTTTGAGATGATTAATAACGGTGCCTTCATCTATGTTTGCGGT
GATGCAAAAGGTATGGCAAAAGGTGTTAGCACCGCACTGGTTGGTATTCTGAGTCGTGGTAAAAGCATTA
CCACAGATGAAGCAACCGAACTGATTAAGATGCTGAAAACCAGTGGTCGCTATCAAGAAGATGTTTGG

MVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRFDAEKEYARAYQQYTKAGKPY
AIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDRTPIEAVHSCNNAESLNILNK
LLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAPDPVLHHHSTSFGEAIMSSCY
RRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADKTKDVRPSSLLDALIAAAFDN
GSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTAFNIIFQLYAIMDHPEYKAPLREEALQALKLTNGDW
TVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGTILLSPCRNVHLDPEIYEDPT
TFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIFIKFLLQYDVEPVQKEILPYG
FTSGPVYMPNPSVMMRIRPRSDGKPRAVSSGNRDIAQVVTENNKNYLVLYASQTGTAEDYAKKFSKELVA
KFNLNVMCADVENYDFESLNDVPVIVSIFISTYGEGDFPDGAVNFEDFICNAEAGALSNLRYNMFGLGNS
TYEFFNGAAKKAEKHLSAAGAIRLGKLGEADDGAGTTDEDYMAWKDSILEVLKDELHLDEQEAKFTSQFQ
YTVLNEITDSMSLGEPSAHYLPSHQLNRNADGIQLGPFDLSQPYIAPIVKSRELFSSNDRNCIHSEFDLS
GSNIKYSTGDHLAVWPSNPLEKVEQFLSIFNLDPETIFDLKPLDPTVKVPFPTPTTIGAAIKHYLEITGP
VSRQLFSSLIQFAPNADVKEKLTLLSKDKDQFAVEITSKYFNIADALKYLSDGAKWDTVPMQFLVESVPQ
MTPRYYSISSSSLSEKQTVHVTSIVENFPNPELPDAPPVVGVTTNLLRNIQLAQNNVNIAETNLPVHYDL
NGPRKLFANYKLPVHVRRSNFRLPSNPSTPVIMIGPGTGVAPFRGFIRERVAFLESQKKGGNNVSLGKHI
LFYGSRNTDDFLYQDEWPEYAKKLDGSFEMVVAHSRLPNTKKVYVQDKLKDYEDQVFEMINNGAFIYVCG
DAKGMAKGVSTALVGILSRGKSITTDEATELIKMLKTSGRYQEDVW

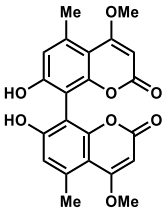
Chemical synthesis. Suman Chakrabarty synthesized and characterized the substrates and product standards reported in this section. All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenk techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, Inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Machery-Nagel 60 F_{254} precoated silica TLC plates (0.25 mm) or Merck Silica Gel 60 F_{254} precoated silica TLC plates (0.25 mm) which were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol green stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230–400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated. ^{1}H and ^{13}C NMR spectra were obtained in $CDCl_3$ at room temperature (25 °C), unless otherwise noted, on Varian 400 MHz or Varian 600 MHz spectrometers. Chemical shifts of ^{1}H NMR spectra were recorded in parts per million (ppm) on the δ scale. High resolution electrospray mass spectra were obtained on an Agilent UPLC–QTOF at the University of Michigan Life Sciences Institute or Agilent UPLC–TOF at the University of Michigan Life Sciences Institute. IR spectra were recorded on a Perkin-Elmer Spectrum BX FT-IR spectrometer.



4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (2.6). To a mixture of 5-methylresorcinol (3.000 g, 24.16 mmol, 1.000 equiv), malonic acid (2.514 g, 24.16 mmol, 1.000 equiv) and $ZnCl_2$ flame-dried under vacuum (10.21 g, 74.91 mmol, 3.100 equiv) was added $POCl_3$ (60.00 mL). The reaction was heated at 60 °C and stirred for 14.5 h. The reaction was quenched by pouring into ice water and induced the precipitation of a solid, which was isolated by vacuum filtration. The crude solid was purified by flash column chromatography (5:6:1 toluene/ethyl acetate/formic acid (v/v)) to afford 3.380 g of the title compound (73% yield) as a yellow solid with minor impurities. $\mathbf{R_f}$ = 0.37 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.06 (s, 1H), 10.39 (s, 1H), 6.54 (s,1H), 6.50 (s, 1H), 5.33 (s, 1H), 2.57 (s, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 169.1, 161.9, 160.4, 157.0, 138.7, 115.6, 106.4, 100.4, 88.0, 22.8; **HRMS** (ESI) *m/z* calculated for $C_{10}H_9O_4^+$ $[M+H]^+$ 193.0495, found 193.0505. All spectra obtained were constant with literature values.[62]



7-hydroxy-4-methoxy-5-methyl-*2H*-chromen-2-one (2.4). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**2.6**; 1.450 g, 7.546 mmol, 1.000 equiv) in MeOH (84.0 mL) was added H_2SO_4 (8.4 mL). The reaction was heated at 75 °C for 6.2 h. Incubation at 0 °C induced the precipitation of a white solid which was isolated by vacuum filtration to afford 1.070 g of the title compound (69% yield). $\mathbf{R_f}$ = 0.43 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.46 (s, 1H), 6.57 (d, J = 2.4, 1H), 6.54 (d, J = 2.4, 1H), 5.59 (s, 1H), 3.92 (s, 3H), 2.53 (s, 3H); **HRMS** (ESI) *m/z* calculated for $C_{11}H_{11}O_4^+$ $[M+H]^+$ 207.0652, found 207.0657. All spectra obtained were constant with literature values.[63]



7,7′-dihydroxy-4,4′-dimethoxy-5,5′dimethyl-*2H*,*2H*′-[8,8′-bichromene]-2,2′-dione (2.5). To a solution of 7-hydroxy-5-methyl-4-propoxy-*2H*-chromen-2-one (**2.4**; 51.5 mg, 0.250 mmol, 1.00 equiv) in CH_2Cl_2 (1.0 mL) and trifluoroacetic acid (0.25 mL) was added VOF_3 (124 mg, 0.375 mmol, 1.50 equiv) at 0 °C. The reaction was stirred for 4.0 h. The reaction was quenched with H_2O (2 mL). The aqueous phase was extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude solid was purified by flash column chromatography (8:3:1 toluene/ethyl acetate/formic acid (v/v)) then reverse-phase HPLC (Phenomenex Kinetex 5 μm C18, 150 x 21.2 mm column) under the following conditions: mobile phase A = 0.1% formic acid in water and B = 0.1% formic acid in acetonitrile; method = 75% A for 5.0 min, 75% A to 70% A over 2.0 min, 70% A for 7.0 min, 70% A to 65% A over 2.0 min, 65% A for 5.0 min, 65% A to 50% A over 3.0 min, 50% A for 1.0 min, 254 and 308 nm UV detection and 14 mL/min flow rate. A retention time of 12.65–14.77 min afforded 7.0 mg of 8,8′-product **7** (14% yield). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.69 (s, 2H), 5.55 (s, 2H), 3.94 (s, 6H), 2.59 (s, 6H); **HRMS** (ESI) *m/z* calculated for $C_{22}H_{19}O_8^+$ $[M+H]^+$ 411.1074, found 411.1070. All spectra obtained were agreement with literature values.[64]

LC–MS traces for biocatalytic reactions. High resolution electrospray mass spectra were obtained on an Agilent UPLC–QTOF at the University of Michigan Life Sciences Institute or Agilent UPLC–TOF at the University of Michigan Life Sciences Institute.

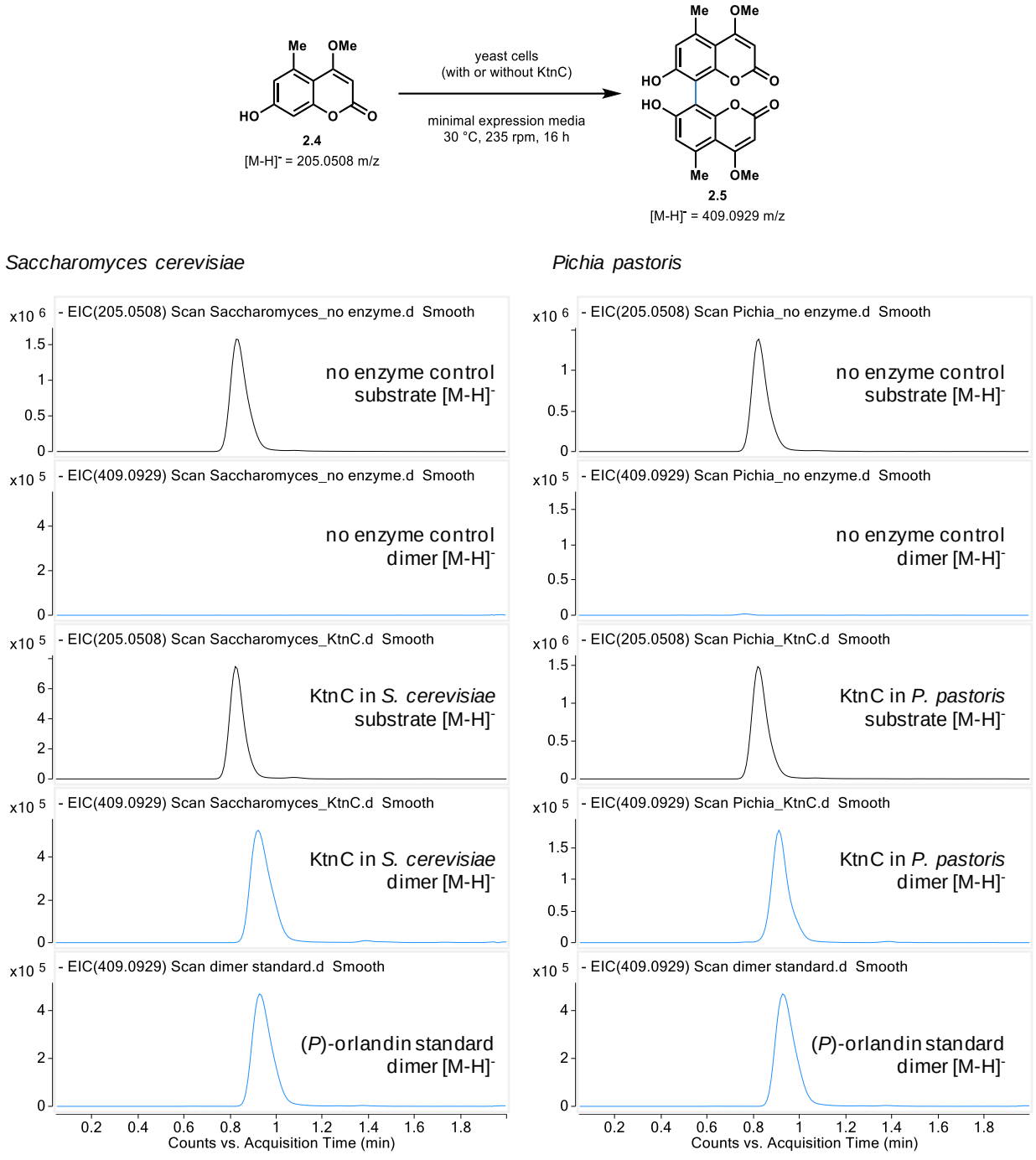


Figure 2.11. Site-selectivity of coumarin dimerization in yeast biotransformations. Extracted ion chromatograms show selective formation of the 8,8'-bicoumarin (*P*)-orlandin (**5**). Additionally, no background reactivity was observed in the no enzyme control reactions, consisting of substrate in *S. cerevisiae* or *P. pastoris* with no recombinant production of KtnC.

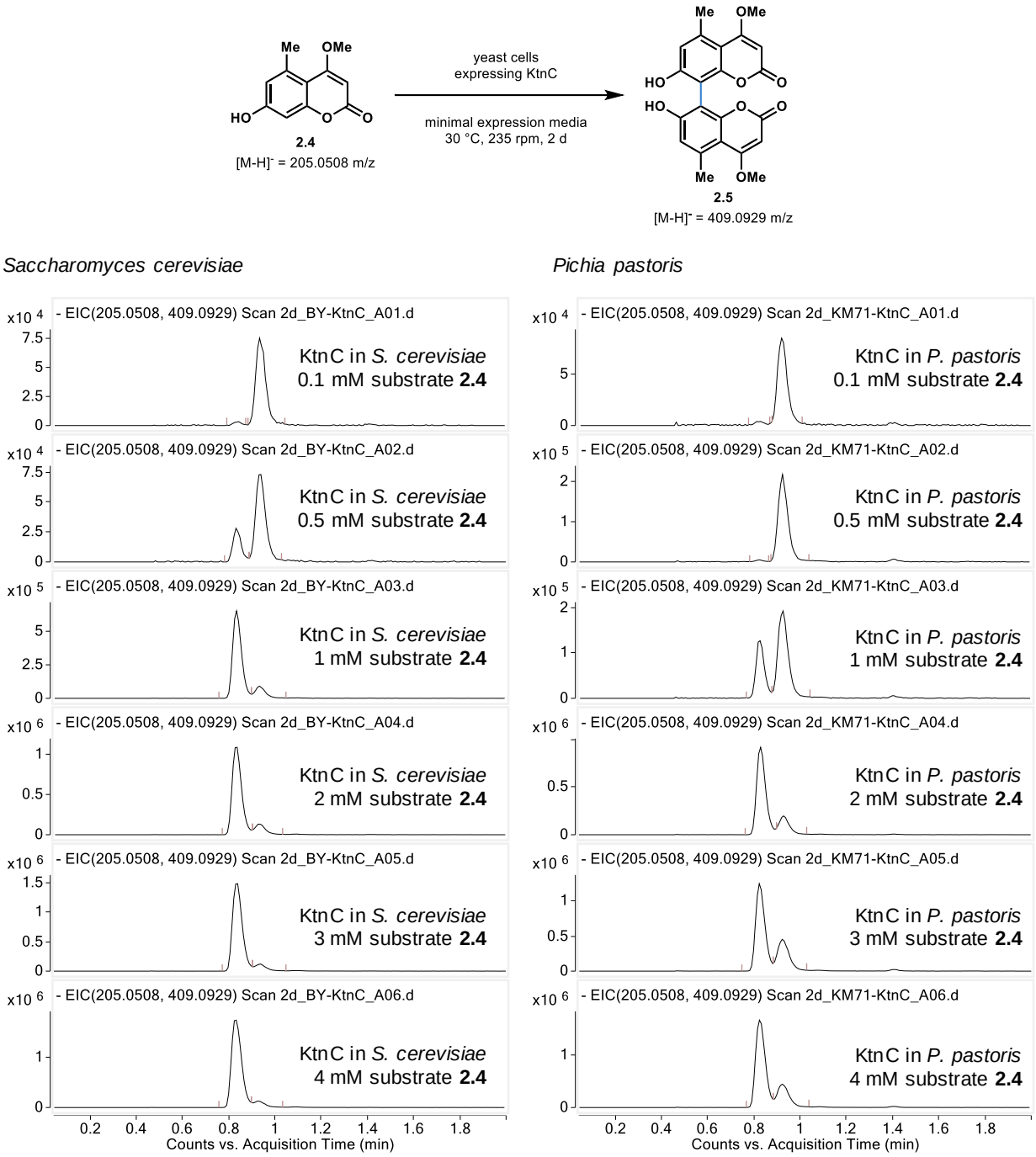


Figure 2.12. Effect of substrate loading on reaction yield in yeast biotransformations.

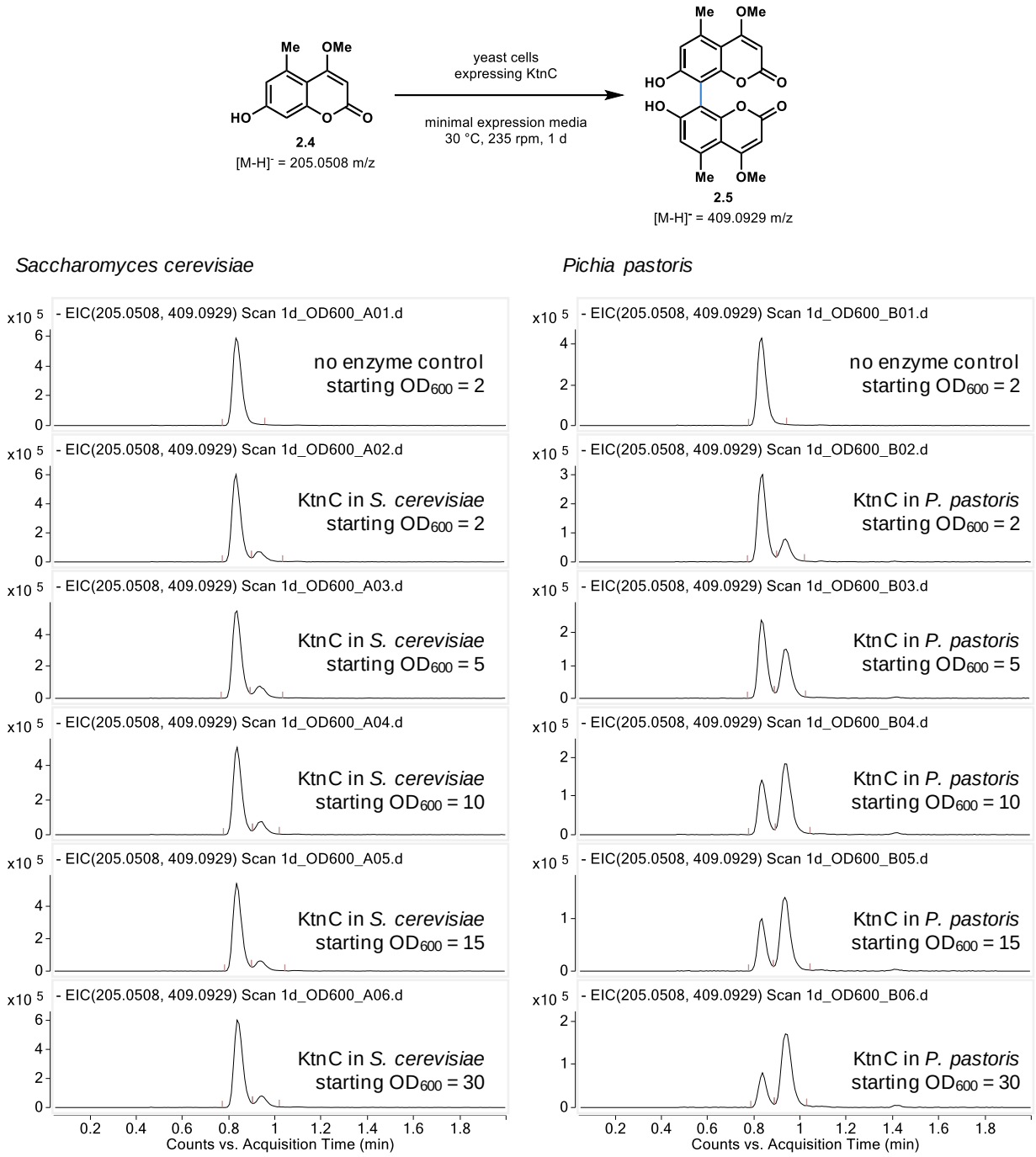


Figure 2.13. Effect of cell density on reaction yield in yeast biotransformations.

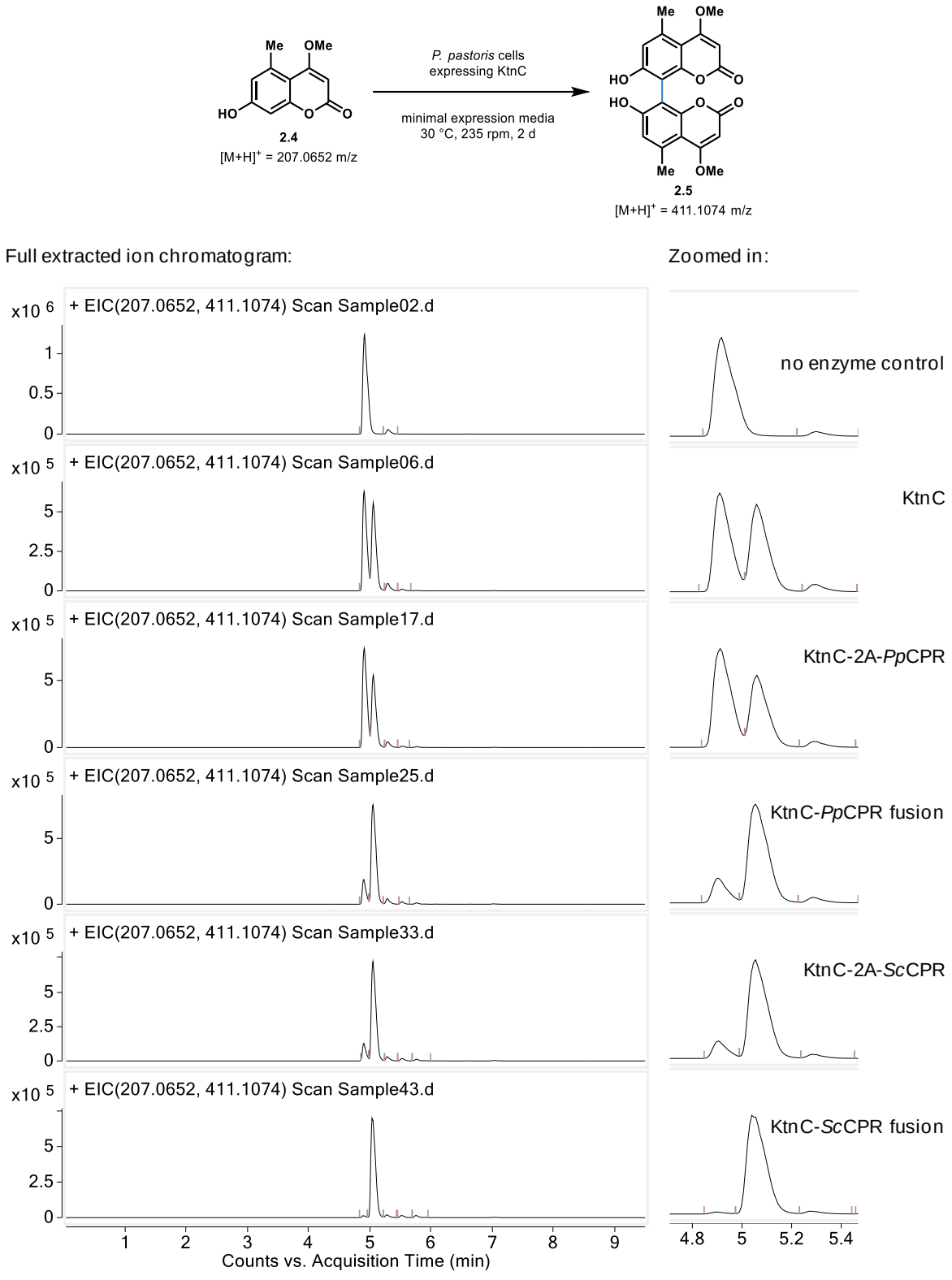
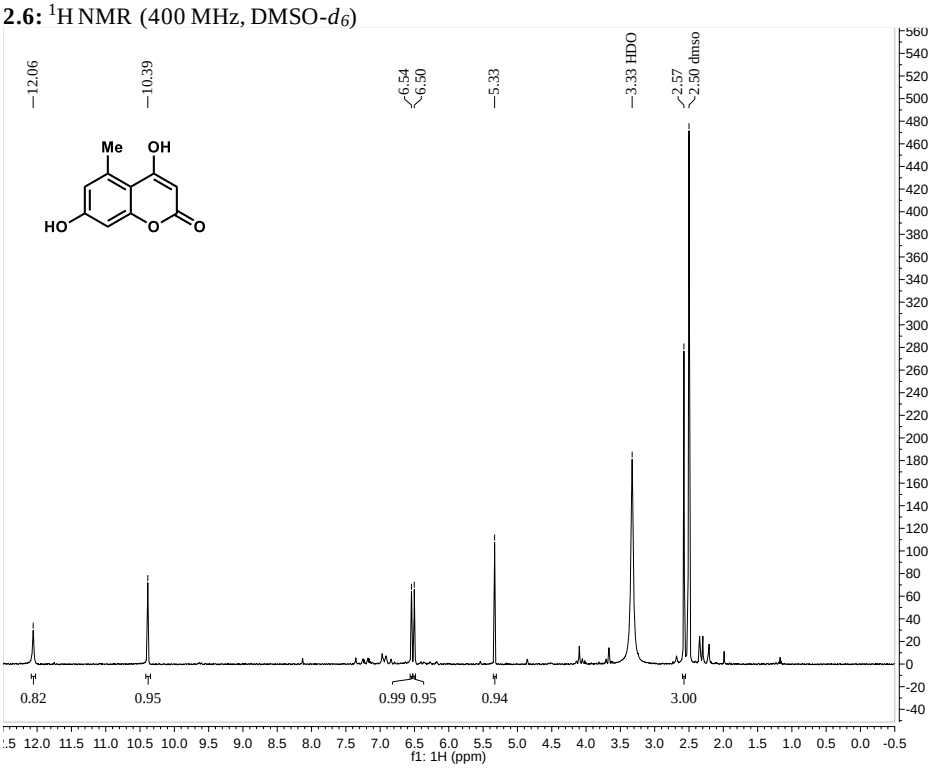


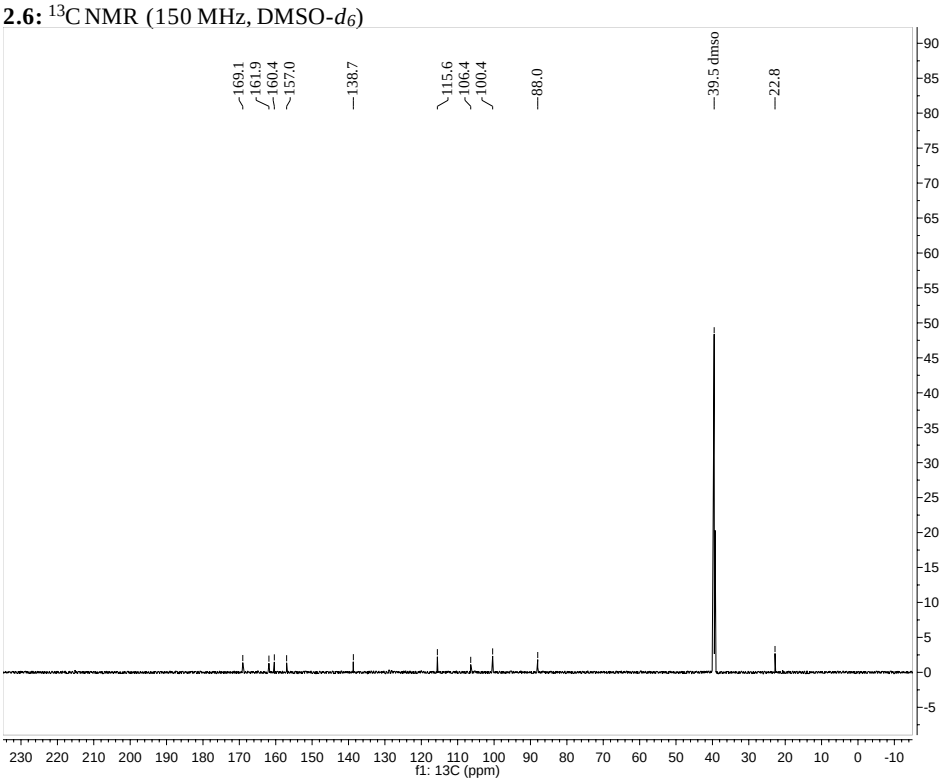
Figure 2.14. Effect of chimeric P450-CPR fusion construct on yeast biotransformation.

^{1}H NMR and ^{13}C NMR spectra of compounds. ^{1}H and ^{13}C NMR spectra were obtained at room temperature (25 °C) on Varian 400 MHz or Varian 600 MHz spectrometers.

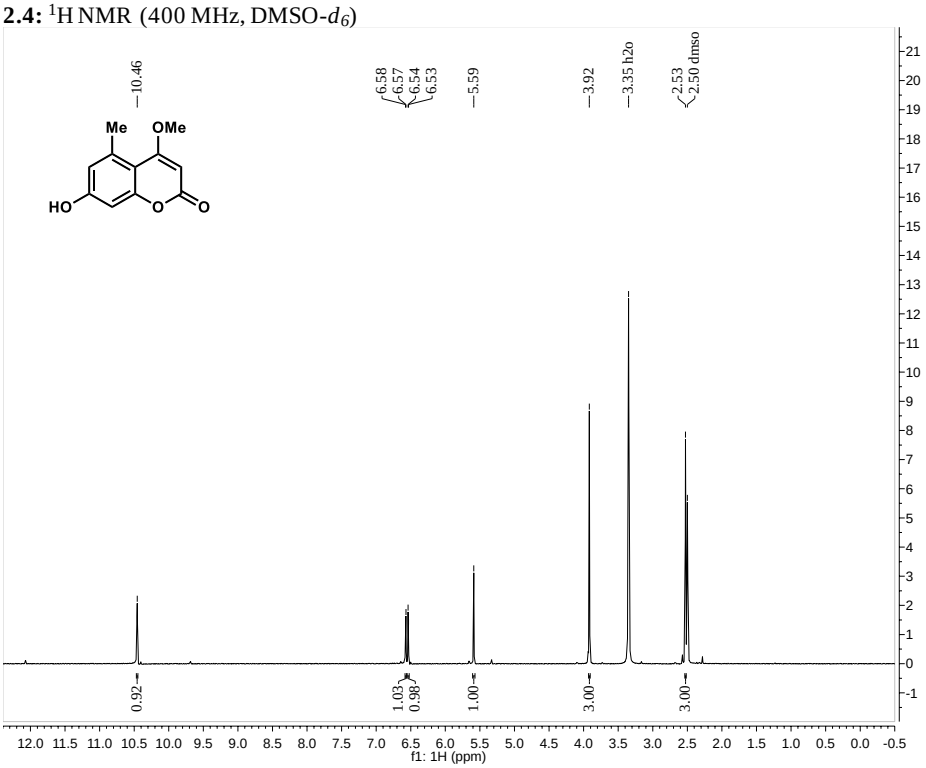
2.6: ^{1}H NMR (400 MHz, DMSO-d_6)



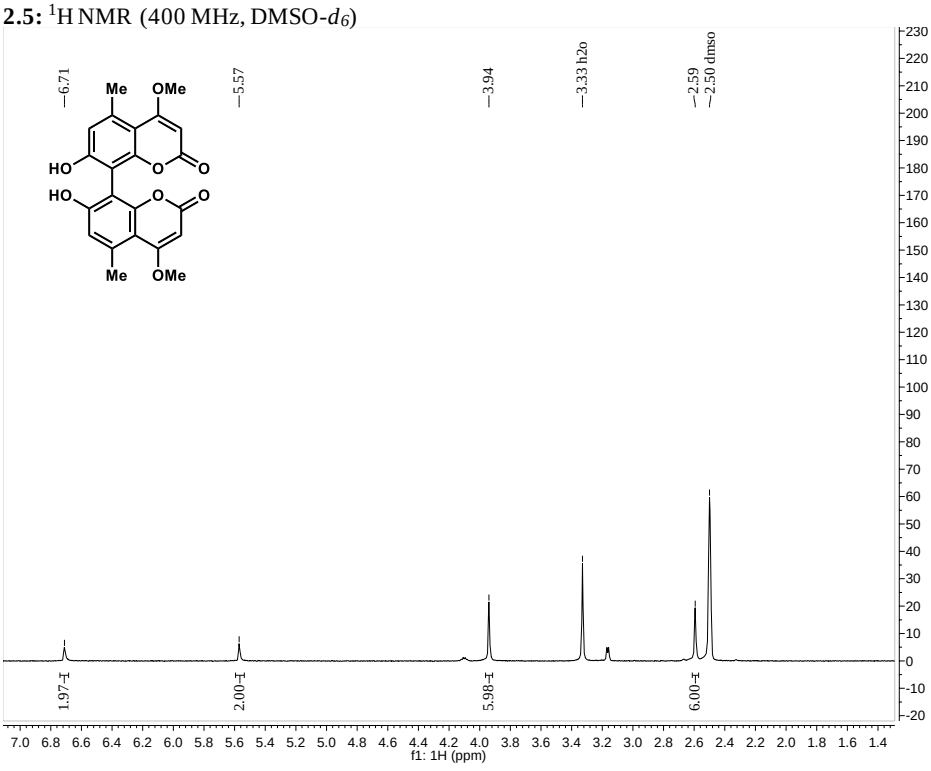
2.6: ^{13}C NMR (150 MHz, DMSO-d_6)



2.4: ^{1}H NMR (400 MHz, DMSO-d_6)



2.5: ^{1}H NMR (400 MHz, DMSO-d_6)



REFERENCES

1. Yet, L., Biaryls. In *Privileged Structures in Drug Discovery: Medicinal Chemistry and Synthesis*, First ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2018; pp 83–154.

2. Bringmann, G.; Günther, C.; Ochse, M.; Schupp, O.; Tasler, S., Biaryls in Nature: A Multi-Facetted Class of Stereochemically, Biosynthetically, and Pharmacologically Intriguing Secondary Metabolites. In *Progress in the Chemistry of Organic Natural Products*, Bringmann, G.; Günther, C.; Ochse, M.; Schupp, O.; Tasler, S.; Herz, W.; Falk, H.; Kirby, G. W.; Moore, R. E., Eds. Springer Vienna: Vienna, 2001; pp 1–249.

3. Bringmann, G.; Seupel, R.; Feineis, D.; Zhang, G.; Xu, M.; Wu, J.; Kaiser, M.; Brun, R.; Seo, E.-J.; Efferth, T. Ancistectorine D, a Naphthylisoquinoline Alkaloid with Antiprotozoal and Antileukemic Activities, and Further 5, 8'-and 7, 1'-Linked Metabolites from the Chinese Liana *Ancistrocladus Tectorius*. *Fitoterapia* **2016**, *115*, 1–8.

4. Francois, G.; Timperman, G.; Eling, W.; Assi, L. A.; Holenz, J.; Bringmann, G. Naphthylisoquinoline Alkaloids against Malaria: Evaluation of the Curative Potentials of Dioncophylline C and Dioncopeltine A against *Plasmodium Berghei In Vivo*. *Antimicrob. Agent Chemother.* **1997**, *41*, 2533–2539.

5. Bringmann, G.; Gulder, T.; Gulder, T. A. M.; Breuning, M. Atroposelective Total Synthesis of Axially Chiral Biaryl Natural Products. *Chem. Rev.* **2011**, *111*, 563–639.

6. Yin; Liebscher, J. Carbon−Carbon Coupling Reactions Catalyzed by Heterogeneous Palladium Catalysts. *Chem. Rev.* **2007**, *107*, 133–173.

7. Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug Discov.* **2018**, *17*, 709–727.

8. Yang, Y.; Lan, J.; You, J. Oxidative C–H/C–H Coupling Reactions between Two (Hetero) Arenes. *Chem. Rev.* **2017**, *117*, 8787–8863.

9. Kozlowski, M. C. Oxidative Coupling in Complexity Building Transforms. *Acc. Chem. Res.* **2017**, *50*, 638–643.

10. Smyth, J. E.; Butler, N. M.; Keller, P. A. A Twist of Nature – The Significance of Atropisomers in Biological Systems. *Nat. Prod. Rep.* **2015**, *32*, 1562–1583.

11. Keshmiri-Neghab, H.; Goliaei, B. Therapeutic Potential of Gossypol: An Overview. *Pharm. Biol.* **2014**, *52*, 124–128.

12. Aldemir, H.; Richarz, R.; Gulder, T. A. M. The Biocatalytic Repertoire of Natural Biaryl Formation. *Angew. Chem. Int. Ed.* **2014**, *53*, 8286–8293.

13. Hüttel, W.; Müller, M. Regio- and Stereoselective Intermolecular Phenol Coupling Enzymes in Secondary Metabolite Biosynthesis. *Nat. Prod. Rep.* **2021**, *38*, 1011–1043.

14. Jeon, J.-R.; Baldrian, P.; Murugesan, K.; Chang, Y.-S. Laccase-Catalysed Oxidations of Naturally Occurring Phenols: From *In Vivo* Biosynthetic Pathways to Green Synthetic Applications. *Microb. Biotechnol.* **2012**, *5*, 318–332.

15. Zetzsche, L. E.; Narayan, A. R. H. Broadening the Scope of Biocatalytic C–C Bond Formation. *Nat. Rev. Chem.* **2020**, *4*, 334–346.

16. Pickel, B.; Schaller, A. Dirigent Proteins: Molecular Characteristics and Potential Biotechnological Applications. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 8427–8438.

17. Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. The Transformative Power of Biocatalysis in Convergent Synthesis. *J. Am. Chem. Soc.* **2022**, *144*, 5214–5225.

18. Zerbe, K.; Woithe, K.; Li, D. B.; Vitali, F.; Bigler, L.; Robinson, J. A. An Oxidative Phenol Coupling Reaction Catalyzed by OxyB, a Cytochrome P450 from the Vancomycin-Producing Microorganism. *Angew. Chem. Int. Ed.* **2004**, *43*, 6709–6713.

19. Aldemir, H.; Shu, S.; Schaefers, F.; Hong, H.; Richarz, R.; Harteis, S.; Einsiedler, M.; Milzarek, T. M.; Schneider, S.; Gulder, T. A. M. Carrier Protein-Free Enzymatic Biaryl Coupling in Arylomycin A2 Assembly and Structure of the Cytochrome P450 AryC. *Chem. Eur. J.* **2022**, *28*, e20210339.

20. Liu, J.; Liu, A.; Hu, Y. Enzymatic Dimerization in the Biosynthetic Pathway of Microbial Natural Products. *Nat. Prod. Rep.* **2021**, *38*, 1469–1505.

21. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

22. Durairaj, P.; Hur, J.-S.; Yun, H. Versatile Biocatalysis of Fungal Cytochrome P450 Monooxygenases. *Microb. Cell Factories* **2016**, *15*, 125.

23. Hüttel, W.; Müller, M. Regio- and Stereoselective Intermolecular Oxidative Phenol Coupling in Kotanin Biosynthesis by *Aspergillus Niger*. *ChemBioChem* **2007**, *8*, 521–529.

24. Chakrabarty, S.; Wang, Y.; Perkins, J. C.; Narayan, A. R. H. Scalable Biocatalytic C–H Oxyfunctionalization Reactions. *Chem. Soc. Rev.* **2020**, *49*, 8137–8155.

25. Lah, L.; Kraševec, N.; Trontelj, P.; Komel, R. High Diversity and Complex Evolution of Fungal Cytochrome P450 Reductase: Cytochrome P450 Systems. *Fungal Genet. Biol.* **2008**, *45*, 446–458.

26. Karbalaei, M.; Rezaee, S. A.; Farsiani, H. *Pichia Pastoris*: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins. *J. Cell. Physiol.* **2020**, *235*, 5867–5881.

27. Zetzsche, L.; Yazarians, J.; Chakrabarty, S.; Hinze, M.; Lukowski, A.; Joyce, L.; Narayan, A. Biocatalytic Oxidative Cross-Coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603*, 79–85.

28. Kim, J. H.; Lee, S.-R.; Li, L.-H.; Park, H.-J.; Park, J.-H.; Lee, K. Y.; Kim, M.-K.; Shin, B. A.; Choi, S.-Y. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PloS One* **2011**, *6*, e18556.

29. Sadeghi, S. J.; Gilardi, G. Chimeric P450 Enzymes: Activity of Artificial Redox Fusions Driven by Different Reductases for Biotechnological Applications. *Biotechnol. Appl. Biochem.* **2013**, *60*, 102–110.

30. Bakkes, P. J.; Riehm, J. L.; Sagadin, T.; Rühlmann, A.; Schubert, P.; Biemann, S.; Girhard, M.; Hutter, M. C.; Bernhardt, R.; Urlacher, V. B. Engineering of Versatile Redox Partner Fusions That Support Monooxygenase Activity of Functionally Diverse Cytochrome P450s. *Sci. Rep.* **2017**, *7*, 9570.

31. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58*, 4169–4182.

32. Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' Genomic Enzymology Web Tools for Functional Assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85.

33. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019–1037.

34. Gietz, D.; St Jean, A.; Woods, R. A.; Schiestl, R. H. Improved Method for High Efficiency Transformation of Intact Yeast Cells. *Nucleic Acids Res.* **1992**, *20*, 1425.

35. Madden, K.; Tolstorukov, I.; Cregg, J., Electroporation of *Pichia Pastoris*. In *Genetic Transformation Systems in Fungi, Volume 1*, Springer: 2015; pp 87–91.

36. Sherer, E. C.; Lee, C. H.; Shpungin, J.; Cuff, J. F.; Da, C.; Ball, R.; Bach, R.; Crespo, A.; Gong, X.; Welch, C. J. Systematic Approach to Conformational Sampling for Assigning Absolute Configuration Using Vibrational Circular Dichroism. *J. Med. Chem.* **2014**, *57*, 477–494.

37. Liu, Z.; Shultz, C. S.; Sherwood, C. A.; Krska, S.; Dormer, P. G.; Desmond, R.; Lee, C.; Sherer, E. C.; Shpungin, J.; Cuff, J. Highly Enantioselective Synthesis of Anti Aryl β-Hydroxy α-Amino Esters via DKR Transfer Hydrogenation. *Tet. Lett.* **2011**, *52*, 1685–1688.

38. Joyce, L. A.; Nawrat, C. C.; Sherer, E. C.; Biba, M.; Brunskill, A.; Martin, G. E.; Cohen, R. D.; Davies, I. W. Beyond Optical Rotation: What's Left Is Not Always Right in Total Synthesis. *Chem. Sci.* **2018**, *9*, 415–424.

39. Petersson, a.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A Complete Basis Set Model Chemistry. I. The Total Energies of Closed-Shell Atoms and Hydrides of the First-Row Elements. *J. Chem. Phys.* **1988**, *89*, 2193–2218.

40. Petersson, G.; Al-Laham, M. A. A Complete Basis Set Model Chemistry. II. Open-Shell Systems and the Total Energies of the First-Row Atoms. *J. Chem. Phys.* **1991**, *94*, 6081–6090.

41. Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* Basis Set for Atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223–1229.

42. Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31 G* Basis Set for Third-Row Atoms. *J. Comp. Chem.* **2001**, *22*, 976–984.

43. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

44. Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

45. Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

46. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

47. Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

48. Frisch, M. J. e. a. *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford, CT, 2009.

49. Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

50. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538–4543.

51. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li–F. *J. Comp. Chem.* **1983**, *4*, 294–301.

52. Bauernschmitt, R.; Ahlrichs, R. Treatment of Electronic Excitations within the Adiabatic Approximation of Time Dependent Density Functional Theory. *Chem. Phys. Lett.* **1996**, *256*, 454–464.

53. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. Specdis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra. *Chirality* **2013**, *25*, 243–249.

54. Mazzeo, G.; Santoro, E.; Andolfi, A.; Cimmino, A.; Troselj, P.; Petrovic, A. G.; Superchi, S.; Evidente, A.; Berova, N. Absolute Configurations of Fungal and Plant Metabolites by Chiroptical Methods. ORD, ECD, and VCD Studies on Phyllostin, Scytolide, and Oxysporone. *J. Nat. Prod.* **2013**, *76*, 588–599.

55. Legault, C. Y. *Cylview 1.0b*, Université de Sherbrooke: http://www.cylview.org, 2009.

56. Obermaier, S.; Müller, M. Biaryl-Forming Enzymes from *Aspergilli* Exhibit Substrate-Dependent Stereoselectivity. *Biochemistry* **2019**, *58*, 2589–2593.

57. Thiele, W.; Froede, R.; Steglich, W.; Müller, M. Enzymatic Formation of Rufoschweinitzin, a Binaphthalene from the Basidiomycete *Cortinarius Rufoolivaceus*. *ChemBioChem* **2020**, *21*, 1423–1427.

58. Matsuda, Y.; Gotfredsen, C. H.; Larsen, T. O. Genetic Characterization of Neosartorin Biosynthesis Provides Insight into Heterodimeric Natural Product Generation. *Org. Lett.* **2018**, *20*, 7197–7200.

59. Griffiths, S.; Mesarich, C. H.; Saccomanno, B.; Vaisberg, A.; De Wit, P. J.; Cox, R.; Collemare, J. Elucidation of Cladofulvin Biosynthesis Reveals a Cytochrome P450 Monooxygenase Required for Anthraquinone Dimerization. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 6851–6856.

60. Greco, C.; de Mattos-Shipley, K.; Bailey, Andrew M.; Mulholland, N. P.; Vincent, J. L.; Willis, C. L.; Cox, R. J.; Simpson, T. J. Structure Revision of Cryptosporioptides and Determination of the Genetic Basis for Dimeric Xanthone Biosynthesis in Fungi. *Chem. Sci.* **2019**, *10*, 2930–2939.

61. Han, Y. B.; Bai, W.; Ding, C. X.; Liang, J.; Wu, S.-H.; Tan, R. X. Intertwined Biosynthesis of Skyrin and Rugulosin a Underlies the Formation of Cage-Structured Bisanthraquinones. *J. Am. Chem. Soc.* **2021**, *143*, 14218–14226.

62. Hubbard, J. S.; Harris, T. M. Condensations at the 6 Position of the Methyl Ester and the Dimethylamide of 3, 5-Dioxohexanoic Acid via 2, 4, 6-Trianions. *J. Org. Chem.* **1981**, *46*, 2566–2570.

63. Gil Girol, C.; Fisch, K. M.; Heinekamp, T.; Günther, S.; Hüttel, W.; Piel, J.; Brakhage, A. A.; Müller, M. Regio- and Stereoselective Oxidative Phenol Coupling in *Aspergillus Niger*. *Angew. Chem. Int. Ed.* **2012**, *51*, 9788–9791.

64. Cutler, H. G.; Crumley, F. G.; Cox, R. H.; Hernandez, O.; Cole, R. J.; Dorner, J. W. Orlandin: A Nontoxic Fungal Metabolite with Plant Growth Inhibiting Properties. *J. Agric. Food Chem.* **1979**, *27*, 592–595.

CHAPTER III

Directed Evolution of KtnC for Unnatural Cross-Coupling Reactions

Biaryl compounds, with two connected aromatic rings, are found across medicine, materials science, and asymmetric catalysis. The necessity of joining arene building blocks to access these valuable compounds has inspired multiple approaches for biaryl bond formation and challenged chemists to develop increasingly concise and robust methods for this task. Oxidative coupling of two C–H bonds offers an efficient strategy for the formation of a biaryl C–C bond; however, fundamental challenges remain in controlling the reactivity and selectivity for uniting a given pair of substrates. Biocatalytic oxidative cross-coupling reactions have the potential to overcome limitations inherent to small molecule-mediated methods by providing a paradigm with catalyst-controlled selectivity. In this chapter, we disclose a strategy for biocatalytic cross-coupling through oxidative C–C bond formation using cytochrome P450 enzymes. We demonstrate the ability to catalyze cross-coupling reactions on coumarin substrates using a natural P450 catalyst. Moreover, we engineer a P450 to possess the desired reactivity, site-selectivity, and atroposelectivity by transforming a low-yielding, unselective reaction into a highly efficient and selective process. The evolved reactivity of the engineered P450 catalysts is profiled across a diverse panel of phenolic substrates, demonstrating the generality of the engineered enzymes. Altogether, this streamlined method for constructing sterically hindered biaryl bonds provides a programmable platform for assembling molecules with catalyst-controlled reactivity and selectivity.

This chapter was adapted with permission from Zetzsche, L. E.; Yazarians, J. A.; Chakrabarty, S.; Hinze, M. E.; Murray, L. A. M.; Lukowski, A. L.; Joyce, L. A.; Narayan, A. R. H. Biocatalytic Oxidative Cross-coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603*, 79–85. © Springer Nature Limited 2022.

INTRODUCTION

Biaryl bonds are ubiquitous in drugs, materials, and ligands for asymmetric catalysis (Figure 3.1a).[1, 2] The most commonly employed strategy to construct these molecules relies on metal-catalyzed cross-coupling of prefunctionalized aryl fragments, thereby offering preprogrammed site-selectivity, albeit at the expense of adding extra steps in a synthetic campaign (Figure 3.1b).[3-5] These reactions can require extensive screening to identify suitable conditions for each new pair of coupling partners and are still restricted in the connectivities which are readily accessible,[6-10] with the formation of tetra-*ortho*-substituted biaryl bonds remaining an outstanding challenge.[11, 12] Oxidative coupling provides an alternative strategy for biaryl bond formation through the net transformation of two C–H bonds into a C–C bond (Figure 3.1b).[13, 14] However, these advantages come at the expense of the efficiency and selectivity of this transformation, which is typically dictated by the intrinsic stereoelectronic properties of each substrate, thereby limiting the versatility of this transformation (Figure 3.1c).[14-17] Numerous metal catalysts and metal-free oxidants have

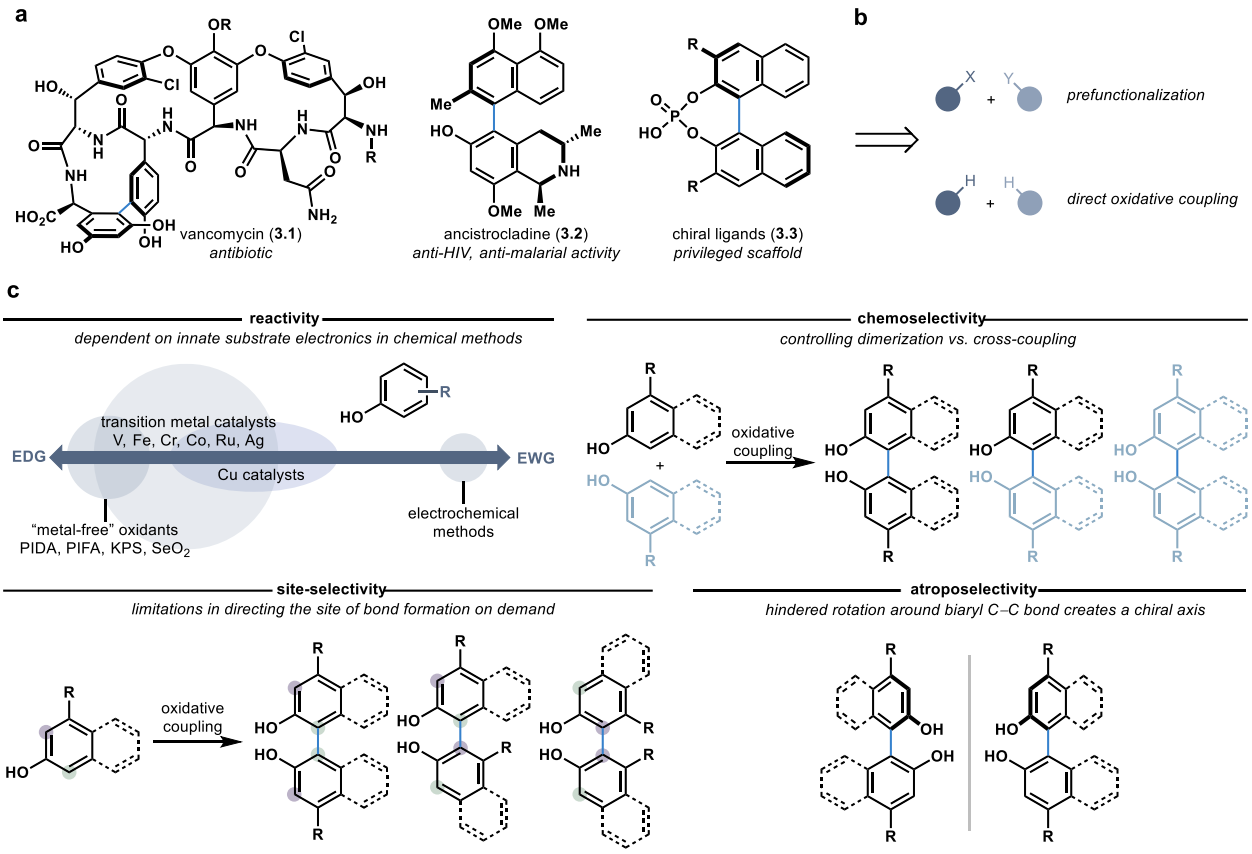


Figure 3.1. Biaryl bond formation through direct oxidative cross-coupling. (a) The biaryl scaffold is a privileged substructure in drug discovery and asymmetric catalysis. (b) Biaryl bonds can be access through metal-catalyzed cross-coupling of prefunctionalized starting materials or direct oxidative coupling. (c) Oxidative biaryl bond formation faces fundamental challenges in reactivity and selectivity.

been reported for the dimerization and cross-coupling of phenolic substrates; however, the application of these methods is commonly restricted to electron-rich phenols while oxidative coupling with electron deficient phenols remains more broadly challenging.[15-26] When general reactivity can be achieved, controlling the chemo-, site-, and atroposelectivity of the oxidative coupling presents a considerable hurdle, particularly in cases where the intrinsic steric and electronic properties of the reacting phenols prohibit access to the desired product isomer.[16, 17, 27-31]

Biocatalytic oxidative cross-coupling reactions have the potential to overcome the limitations inherent to small molecule-mediated methods for direct oxidative coupling by providing a paradigm with catalyst-controlled selectivity. Nature has evolved oxidative enzymes, including laccases and P450s, that mediate the dimerization of phenolic compounds in the biosynthesis of biaryl natural products.[32] The application of laccases in selective catalysis has been limited by the scarcity of examples in which these enzymes exert control over the selectivity of the bond formation following the initial oxidation.[33-35] In contrast, an ever-expanding repertoire of P450s mediate highly selective oxidation reactions,[36] among which exist a small subset of enzymes that catalyze site- and atroposelective dimerizations.[37] This subset of wild-type P450s that perform oxidative coupling chemistry are yet to be explored as biocatalysts for unnatural cross-coupling reactions. We hypothesized that this class of enzymes could be developed into tunable biocatalysts for convergent oxidative cross-coupling reactions with catalyst-controlled reactivity, site-selectivity, and atroposelectivity.

RESULTS AND DISCUSSION

With the complementary reactivity and selectivity of a growing set of wild-type P450s providing a robust starting point, we envisioned developing a biocatalytic method for the site- and atroposelective oxidative cross-coupling to provide a solution to this holy grail reaction. Among the P450s previously characterized to perform oxidative dimerization reactions in the biosynthesis of biaryl natural products, we were initially drawn to a small cluster of *Aspergillus* P450s that each imparts a unique site-selectivity in the oxidative dimerization of coumarin 3.**4**, as demonstrated by Müller and coworkers.[38, 39] This catalyst-controlled site-selectivity provided us a tantalizing starting point for developing an enzymatic platform for oxidative coupling reactions.

We hypothesized that the co-localization of two non-equivalent phenolic substrates in the enzyme active site could provide a route toward cross-coupled products in which selectivity is

dependent on substrate localization and positioning in the active site, rather than the inherent electronic properties of each substrate. As a starting point, fungal P450 KtnC was selected due to its ability to catalyze the site- and atroposelective dimerization of coumarin **3.4** in the biosynthesis (*P*)-orlandin (**3.6**). Using the platform for whole-cell biocatalysis in yeast developed in chapter 2, the ability for KtnC to catalyze cross-coupling reactions with native coumarin substrate **3.4** and a non-native phenolic substrate **3.5** was probed (Figure 3.2a).

In an initial proof-of-concept experiment, coumarins **3.4** and **3.15** were both added to a culture producing KtnC, leading to the formation of unnatural cross-coupled products. Further experimentation demonstrated that tuning the stoichiometry of the cross-coupling partners could maximize cross-coupled product formation over dimerization (Figure 3.2b). Using this strategy

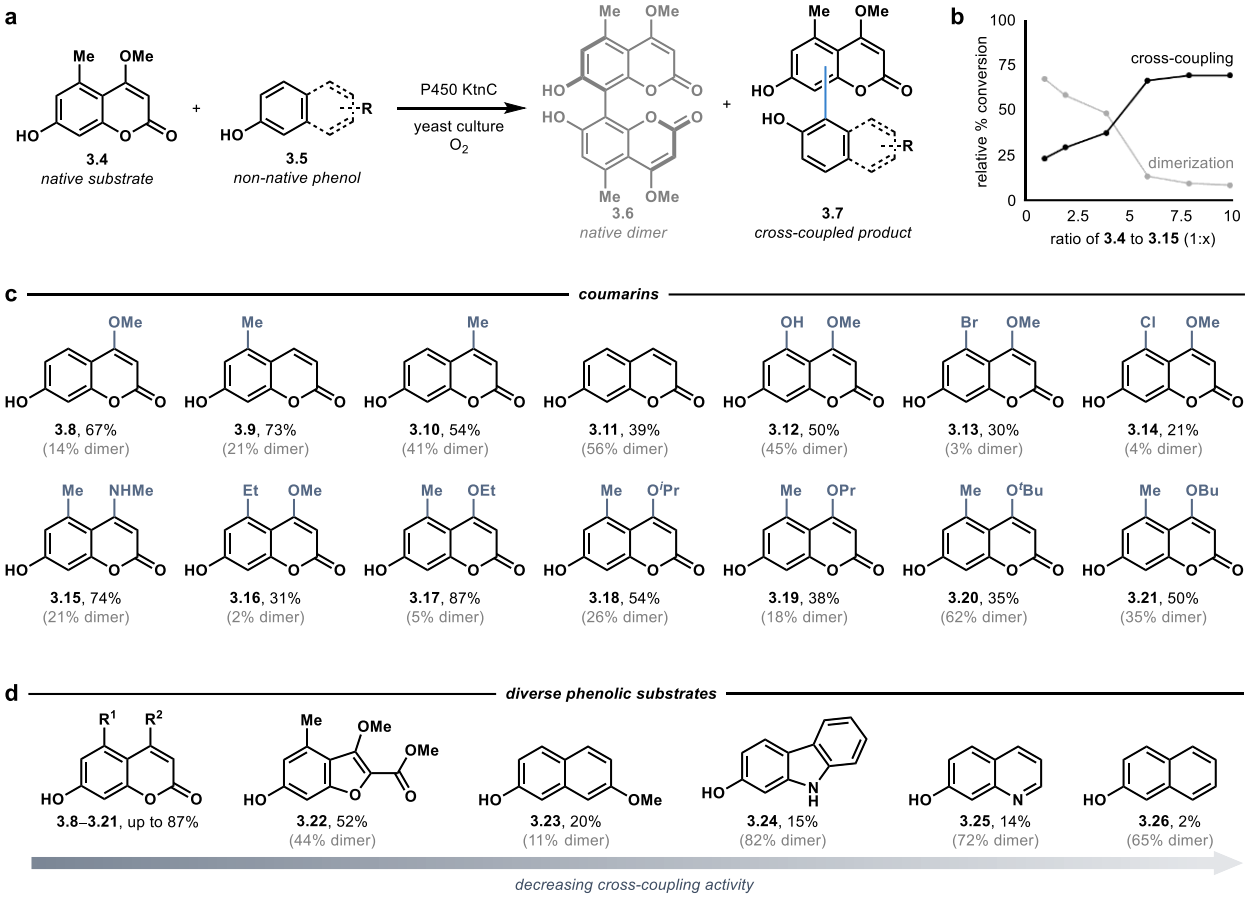


Figure 3.2. Scope of oxidative cross-coupling reactions catalyzed by wild-type KtnC. (a) Cross-coupling reactions catalyzed by KtnC were screened using yeast biotransformations. (b) Adding excess of the coupling partner resulted in more cross-coupling relative to competing dimerization. (c) KtnC displayed promiscuity in the cross-coupling of native coumarin **3.4** with a range of non-equivalent coumarins with varying steric and electronic properties. (d) KtnC catalyzed the cross-coupling with a diverse set of phenolic substrates, albeit with decreased reactivity and selectivity as the substrate structures deviated further from the native coumarin scaffold. Reaction conditions: *P. pastoris* cells expressing KtnC in the presence of **3.4** and **3.5** in up to a 1:10 molar ratio with 25–100 μM **3.4** at 30 °C with shaking at 235 rpm for 48–72 h; cross-coupling reactions reported as percent conversions; competing dimerization of substrate **3.4** reported as percent conversions in parentheses.

for KtnC-catalyzed cross-coupling reactions, substrate promiscuity was observed with a panel of coumarin substrates (Figure 3.2c). Initial cross-coupling reactions between **3.4** and **3.8**–**3.11** demonstrated that maintaining similar substitution patterns to **3.4** at the C4 and C5 positions is beneficial but not required for activity. Additionally, KtnC tolerated a range of coumarins bearing electron-rich and electron-deficient substituents (**3.12**–**2.15**), diverging from the stereoelectronic restrictions that typically govern small molecule-catalyzed oxidative cross-coupling reactions.[16, 17] Added steric bulk at the C4 and C5 positions of coumarins **3.16**–**3.21** was also readily accepted.

In addition to forming a range of unsymmetrical bicoumarin products, a panel of more diverse phenolic substrates were tested as cross-coupling partners for coumarin **3.4** (Figure 3.2d). In general, we observed superior reactivity with phenolic substrates that more closely resembled the native substrate (**3.4**), such as benzofuran **3.22** which afforded a 52% conversion to cross-coupled product. In contrast, as the substrate structures diverged further from the native coumarin scaffold, significant decreases in the cross-coupling activity and site-selectivity occurred. We hypothesized that the limitations in substrate scope and chemo-, site-, and atroposelectivity could be overcome with protein engineering.

We anticipated that protein engineering could address critical limitations in reactivity and selectivity displayed in unnatural cross-coupling reactions with the wild-type enzymes. To overcome these challenges, we developed a high-throughput directed evolution strategy that would identify P450 variants (LxCs) with both improved activity and selectivity for the target reaction. We hypothesized a semi-rational mutagenesis approach focused on targeting residues near the active site would provide the biggest impact, particularly on the selectivity of the reaction.[40] Starting with a homology model of KtnC (Figure 3.3a, residues within 12 Å of the predicted substrate binding region were selected, excluding any residues with >95% conservation based on a MAFFT alignment with up to 5,000 other sequences. The radius of the active site sphere was reduced incrementally until a total of 96 residues with ≤95% conservation were selected within an 11.25 Å radius of the active site (Figure 3.3a).

These residues were randomly mutated using site-saturation mutagenesis, generating thousands of variants harboring a single mutation within several spheres of the active site. Beneficial mutations identified through this site-scanning saturation mutagenesis approach were then folded together in combinatorial libraries made by multiple site-directed mutagenesis.[41] Variants were screened in high-throughput yeast biotransformations and analyzed for total cross-coupled product

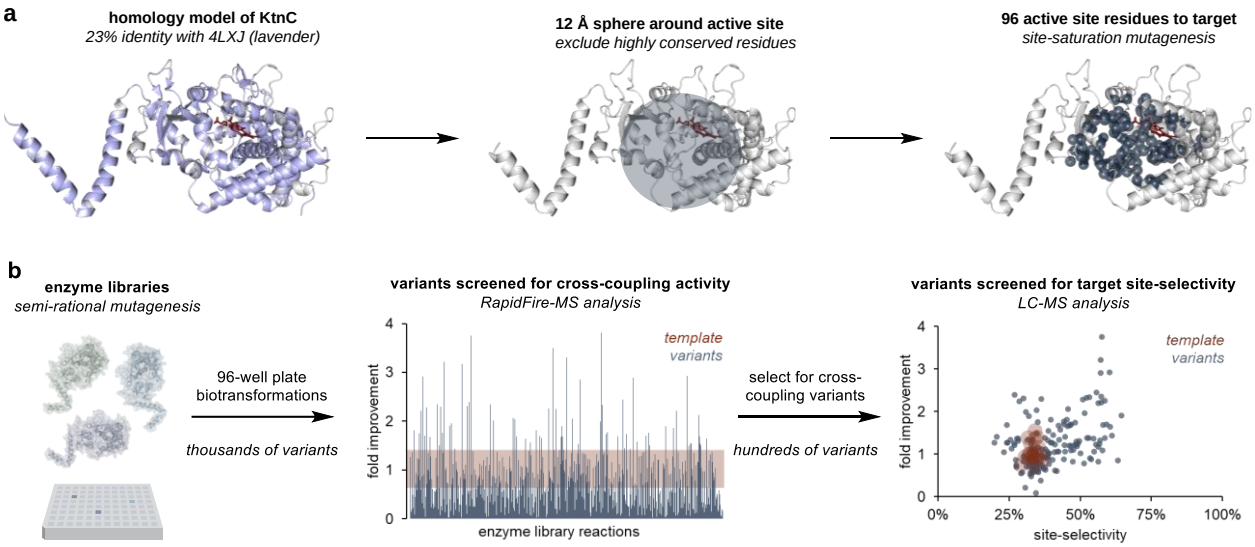


Figure 3.3. Development of a high-throughput engineering strategy. (a) Using a homology model of KtnC (gray) and based on 23% identity with lanosterol 14-alpha demethylase from *S. cerevisiae* (PDB 4LXJ; lavendar), 96 active site residues were targeted within 12 Å of active site. (b) Strategy for the directed evolution of P450s for improved activity and site-selectivity employing semi-rational mutagenesis and a two-tier mass spectrometry screen, as represented by data from round three of evolution above.

formation by RapidFire–mass spectrometry (MS). Reaction mixtures containing higher levels of the cross-coupled product were then analyzed for cross-coupling site-selectivity in a second-tier screen using LC–MS (Figure 3.3c). Ultimately, this two-tier MS screen enabled rapid identification of variants with improved activity and site-selectivity for the targeted cross-coupling reaction.

Using this platform, we targeted the cross-coupling catalyzed by wild-type KtnC between coumarin **3.9** and 2-naphthol (**3.26**) as a model system due to its low levels of cross-coupling activity (<3% conversion) and lack of site-selectivity, resulting in a complex mixture of cross-coupled and dimeric products (Figure 3.4a). We envisioned that the successful engineering of a catalyst that could transform this low-yielding, unselective biocatalytic reaction to a synthetically viable method for the preparative scale synthesis of cross-coupled product **3.27** would clearly demonstrate the potential of this platform for target-oriented synthesis. Additionally, this reaction created an opportunity to test the limits on engineering these enzymes for both the desired activity and multiple forms of selectivity.

Through iterative rounds of directed evolution, the total activity and site-selectivity in the desired cross-coupling reaction was progressively improved (Figure 3.4). Initially, improvements in total activity were targeted, resulting in a 19-fold increase in total cross-coupled product formation over the first two rounds of engineering, while keeping competing levels of coumarin dimerization low. However, the site-selectivity of the reaction was undesirable, with the C–O cross-

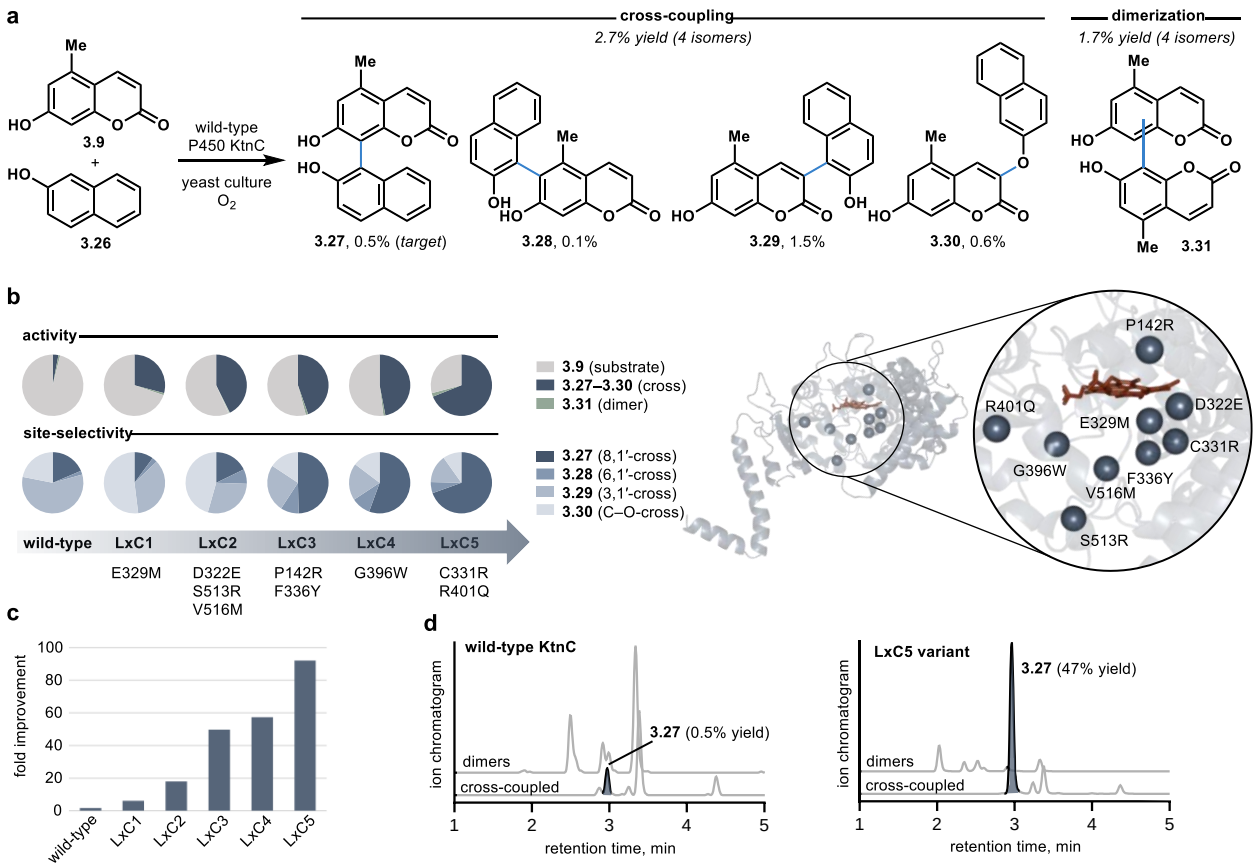


Figure 3.4. Engineering P450 biocatalyst for improved activity and site-selectivity. (a) Target reaction selected for directed evolution campaign, with wild-type KtnC catalyzing the cross-coupling at a 2.7% yield to four different cross-coupled isomers, in addition to dimeric products. (c,d,e) Over five rounds of evolution, a total of nine active site mutations were incorporated to provide LxC5 with 92-fold improvement in activity and site-selectivity toward the target cross-coupled product **3.27**.

coupled product (**3.30**) representing nearly half of all the cross-coupled products formed by LxC2. Therefore, variants with improved site-selectivity were targeted in subsequent engineering rounds, leading to greater site-selectivity for the desired 8,1′-product (**3.27**). Ultimately, a biocatalyst with a 92-fold improvement in activity was engineered through five rounds of evolution to form cross-coupled product **3.27**. Preparative scale biotransformations with LxC5 resulted in the isolation of cross-coupled product **3.27** in a 50% yield.

As variants were engineered selecting for increased cross-coupling yield and site-selectivity, a decrease in the atroposelectivity of the reaction was observed from a 20:80 er with the wild-type enzyme to a 48:52 er with LxC5 (Figure 3.5a). To regain catalyst control over atroposelectivity, two additional rounds of engineering were performed using a combination of RapidFire and chiral LC–MS. A variant (LxC7) that catalyzed the cross-coupling to **3.27** in a 23:77 er was identified, albeit with a reduction in yield (Figure 3.5b). These results demonstrate that this biocatalytic

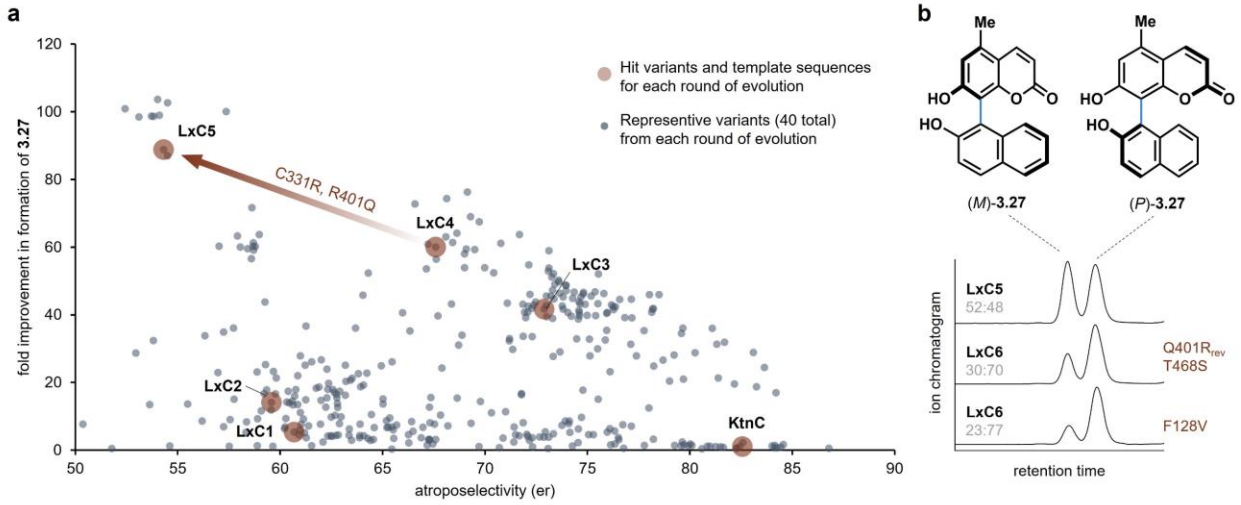


Figure 3.5. Engineering P450 biocatalyst for improved atroposelectivity. (a) A decrease in the atroposelectivity in the cross-coupling reaction was observed when up to 40 variant reactions from each round of evolution were analyzed by LC–MS using a chiral column. (b) Over two additional rounds of engineering, the atroposelectivity of the cross-coupling reaction was improved, giving the (*P*)-atropisomer as the major product.

platform is highly engineerable for the desired activity, site-selectivity, and atroposelectivity, thereby providing a programmable platform for catalyst-controlled oxidative cross-coupling reactions.

At this point, we had engineered a catalyst for a single cross-coupling reaction between coumarin **3.9** and naphthol **3.26**. However, we were curious to know whether the catalyst we had engineered was specialized for that specific reaction, or more of a general catalyst for cross-coupling reactions.[42] To probe this, a panel of diverse phenolic substrate was compiled, including coumarins, naphthols, tricycles, and phenols, and screened in cross-coupling reactions against coumarin **3.9** or naphthol **3.26** (Figure 3.6). No improvement was observed in coumarin–coumarin cross-coupling reactions, demonstrating a divergence in the evolution from the natural coumarin dimerization activity of KtnC. In contrast, significant improvements were observed in coumarin–naphthol cross-coupling reactions, suggesting that we engineered a general catalyst for coumarin–naphthol cross-coupling reactions rather than just a specialist for the cross-coupling between coumarin **3.9** and naphthol **3.26**. Beyond coumarin–naphthol cross-couplings we were surprised to see an increase in general cross-coupling reactivity with various tricyclic substrates and phenols, substrate scaffolds that had rarely shown any activity with wild-type KtnC. Altogether, these results suggest that the LxC variants not only offer a general suite of catalysts for coumarin–naphthol cross-coupling reactions, but also unlocked reactivity with various naphthol, tricyclic, and phenolic substrates.

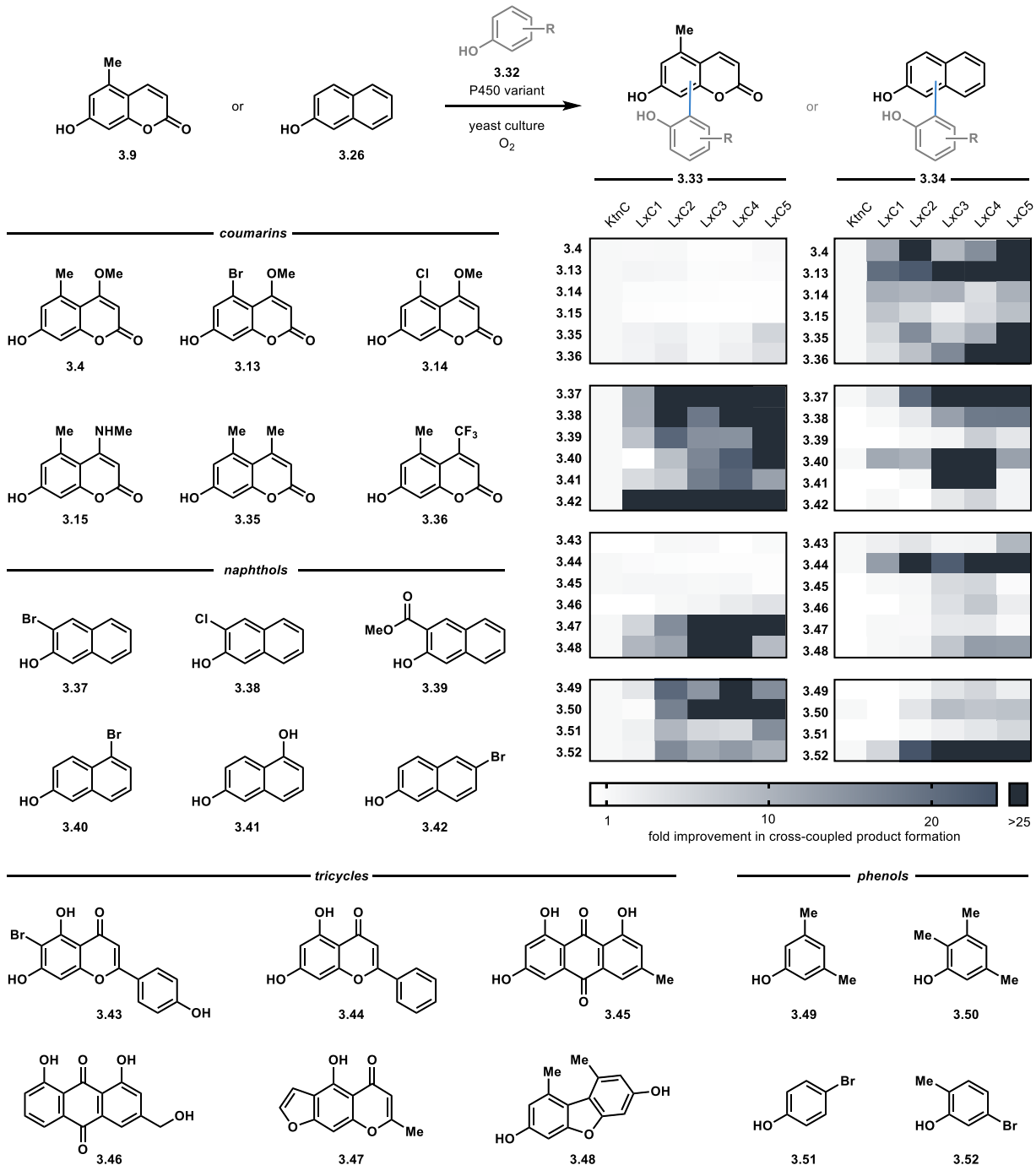


Figure 3.6. Profiling the reactivity of engineered LxC variants. Reaction conditions: *S. cerevisiae* cells expressing KtnC or LxC in the presence of 100 μM **3.9** or **3.26** and 100 μM **3.32** at 30 °C with shaking at 350 rpm for 4 d. The total cross-coupled product formed in each reaction was calculated as a fold improvement compared to the cross-coupled product formed with KtnC (set to a fold improvement of 1.0).

The challenge of overriding innate reactivity and selectivity in oxidative coupling reactions has limited the reliable use of this method for convergent synthetic routes. The presented biocata-

lytic method overcomes reactivity hurdles by surpassing steric and electronic limitations and redefining substrates that can be efficiently cross-coupled. Furthermore, using a large molecule ligand provides the opportunity to achieve catalyst-controlled site- and atroposelectivity in biaryl coupling reactions. The wealth of natural catalysts available, along with the demonstrated ability to tune enzymes for the desired substrate scope, reactivity, and selectivity, positions this biocatalytic approach to be a powerful tool in the convergent assembly of biaryl molecules.

EXPERIMENTAL PROCEDURES

Analytical scale biotransformations with KtnC. A colony of *P. pastoris* KM71 containing KtnC was inoculated in 6 mL BMG medium (100 mM potassium phosphate pH 6.0, 1.34% YNB, 4 x 10^{-5}% biotin, 1% glycerol) and grown overnight at 30 °C with shaking at 235 rpm. The culture was added to BMG medium (125 mL) and grown at 30 °C with shaking at 235 rpm until the optical density at 600 nm was in the range of 10–16. Cultures were induced for expression by resuspension in BMM medium (100 mM potassium phosphate pH 6, 1.34% YNB, 4 x 10^{-5}% biotin, 0.5% methanol) to a final optical density of 30. Substrates in a 1:10 ratio (25 μM of **3.4** and 250 μM of coupling partner or 100 μM of **3.4** and 1000 μM coupling partner) were added to 24-well plates from a 25- or 50-mM stock in DMSO, followed by 1 mL aliquots of expression cultures. The 24-well plates were sealed with an adhesive cover and grown at 30 °C with shaking at 235 rpm for 48 h and supplemented with 100 μL of 10X methanol (5% MeOH, 95% water) after 24 h. Biotransformation and standard curve cultures were subjected to chemical lysis by the addition of 200 μL CelLytic Y Cell Lysis Reagent (Sigma) and incubation at 30 °C with shaking at 235 rpm for 30 min. An aliquot (50 μL) of the samples were removed and diluted with 150 μL methanol containing an internal standard, either 1,3,5-trimethoxybenzene (final concentration of 300 μM) or 4,4′-dihydroxybenzophenone (final concentration of 40 μM). Precipitated yeast cells were pelleted by centrifugation (17,000 x *g* for 10 min) or filtered through Pall AcroPrep Advance 96-well filter plates by centrifugation (2,000 rpm for 3 min).

LC–MS analysis of analytical scale reactions. Percent conversion was calculated based on consumption of coumarin starting material **3.4** and formation of bicoumarin **3.6** compared to a standard curve with an internal standard. The concentration of the desired cross-coupled AB products was assumed to be comprised of the remaining:

(1) $$[A]_{start} = [A]_{remaining} + 2 \times [AA]_{formed} + [AB]_{formed}$$

(2) where $[AB]_{formed} = [A]_{start} - [A]_{remaining} - 2 \times [AA]_{formed}$

(3) $\% AB = [AB]_{formed} / [A]_{start} \times 100\%$

(4) $\% AA = [AA]_{formed} / [AA]_{theoretical\ max} \times 100\%$

To generate the standard curves, coumarin **3.4** (0–150 μM) and bicoumarin **3.6** (0–75 μM) were added to expression cultures of *P. pastoris* KM71 (with no KtnC present) following the same procedures as outlined for analytical scale KtnC biotransformations. The samples were subjected to liquid chromatography PDA spectrometry (UPLC) analysis with a Waters Acquity UPLC HSS T3 column (1.8 μm C18, 2.1x50 mm) on an Agilent 6230 time of flight mass spectrometer with a Dual AJS ESI source and an Agilent 1290 Infinity Series II diode array detector, autosampler, and binary pump. For positive mode analysis, an aqueous (A) mobile phase of 2 mM ammonium formate pH 3.5 in water and an organic (B) mobile phase of 95:5 acetonitrile/water (v/v) with 2 mM ammonium formate pH 3.5 was used at 0.7 mL/min flow rate (80% A held for 0.5 min, to 45% A over 2.75 min, 100% A for 1.0 min, 80% A for 1 min). For negative mode analysis, an aqueous (A) mobile phase of 95:5 water/acetonitrile (v/v) and an organic (B) phase of 95:5 acetonitrile/water (v/v) was used at 0.7 mL/min flow rate (80% A held for 0.5 min, to 45% A over 2.75 min, 100% A for 1.0 min, 80% A for 1 min).

Design and generation of libraries for site-saturation mutagenesis. All residues within ~12 Å of predicted substrate binding region in the KtnC homology model (Phyre2) active site were selected in PyMOL (command: "select bca. KtnC within 12 of substrate"). Any residues with >95% conservation based on a MAFFT alignment with up to 5,000 other sequences were removed from the pool of selected residues. The radius of the active site sphere was reduced incrementally until a total of 96 residues with ≤95% conservation were selected within an 11.25 Å radius (see Figure 3.3). Degenerate forward and reverse primers were designed for each of the 96 active site residues being targeted and purchased from IDT (Table 3.1). Blunt-end whole-plasmid PCRs were set up in a 96-well PCR plate, with each 25-uL reaction containing 1X HF Phusion buffer, 4% DMSO, 400 μM dNTPs, 0.1 ng/μL template plasmid, 0.4 units Phusion DNA polymerase, 0.8 μM degenerate forward primer, and 0.8 μM reverse primer. The reaction conditions were programmed as follows: 98 °C denaturation for 2 min; 25 cycles of 98 °C for 20 sec, 56 °C for 20 sec, 72 °C for 4.5 min; and a final 72 °C extension for 10 min. After amplification, 10 μL of each PCR product was pooled in a single tube. The crude PCR product mixture was purified by gel extraction from a 0.8% agarose gel using Qiagen spin columns. Any residual template DNA was digested in 1X

Cutsmart buffer with 2 μL DpnI (NEB) overnight at 37 °C and removed with a PCR clean-up using Qiagen spin columns. The linear DNA fragments were phosphorylated in 1X T4 DNA ligase (NEB) buffer with 2 μL PNK enzyme (NEB) at 37 °C for 30 min. T4 ligase enzyme (1 μL, NEB) was added directly to the phosphorylation reaction and incubated at room temperature for 2 h. Salts were removed from the ligation reaction with a PCR clean-up using Qiagen spin columns. High efficiency electrocompetent 5-alpha *E. coli* cells (NEB) were transformed with 2 μL of the clean DNA using standard electroporation protocols. Resulting colonies were grown up and miniprepped to give a plasmid DNA library. *S. cerevisiae* strain BY4742 cells were prepared for transformation with the plasmid DNA library through a standard protocol for lithium acetate transformations. Transformed cells were plated on histidine dropout plates containing 2% glucose and incubated at 25 °C for 3 d. Resulting colonies were used directly in library screening.

Generation of combinatorial libraries using multiple site-directed mutagenesis. Forward primers were designed to flank desired mutation with >15 complementary nucleotides on either side of mutation and purchased from IDT (Table 3.2). Equimolar amounts of each of the forward primers were combined and diluted down to 5 μM. PCRs were performed with the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent) according to their recommended protocols. After amplification, 1 μL of the supplied DpnI was added and the digestion reaction was incubated at 37 °C for 1 h. High efficiency electrocompetent 5-alpha *E. coli* cells (NEB) were transformed with 2 μL of the digested reaction using standard electroporation protocols. Resulting colonies were grown up and miniprepped to give a plasmid DNA library. *S. cerevisiae* strain BY4742 cells were prepared for transformation with the plasmid DNA library through a standard protocol for lithium acetate transformations. Transformed cells were plated on histidine dropout plates containing 2% glucose and incubated at 25 °C for 3 d. Resulting colonies were used directly in library screening.

Table 3.1. Primers used for site-saturation mutagenesis of KtnC active site residues. Primers from round 1 (Rd1) were reused in subsequent rounds unless the sequence lost complementarity to the template DNA due to new mutations being carried forward. Codons that were edited to accommodate mutations added to the template DNA are highlighted in red.

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
R62X	Rd1-fwd	NNKCCAAAGCTACGATGGAAGCG	Rd1-rvs	CATGTAGCTACCCTCGTCTACAATC
K64X	Rd1-fwd	NNKCTACGATGGAAGCGCTTCG	Rd1-rvs	TGGCCGCATGTAGCTACCC
L65X	Rd1-fwd	NNKCGATGGAAGCGCTTCGA	Rd1-rvs	CTTTGGCCGCATGTAGCTAC
R66X	Rd1-fwd	NNKTGGAAGCGCTTCGATGC	Rd1-rvs	TAGCTTTGGCCGCATGTAGC
W67X	Rd1-fwd	NNKAAGCGCTTCGATGCGG	Rd1-rvs	TCGTAGCTTTGGCCGCA
K68X	Rd1-fwd	NNKCGCTTCGATGCGGAGAA	Rd1-rvs	CCATCGTAGCTTTGGCCG
R69X	Rd1-fwd	NNKTTCGATGCGGAGAAAGAGTATG	Rd1-rvs	CTTCCATCGTAGCTTTGGCC

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
F70X	Rd1-fwd	NNKGATGCGGAGAAAGAGTATGCG	Rd1-rvs	GCGCTTCCATCGTAGCTTTG
D71X	Rd1-fwd	NNKGCGGAGAAAGAGTATGCGAGA	Rd1-rvs	GAAGCGCTTCCATCGTAGC
A72X	Rd1-fwd	NNKGAGAAAGAGTATGCGAGAGCATATC	Rd1-rvs	ATCGAAGCGCTTCCATCG
E73X	Rd1-fwd	NNKAAAGAGTATGCGAGAGCATATCAGC	Rd1-rvs	CGCATCGAAGCGCTTCC
Y76X	Rd1-fwd	NNKGCGAGAGCATATCAGCAGTACA	Rd1-rvs	CTCTTTCTCCGCATCGAAGC
R93X	Rd1-fwd	NNKATGCAGAATGATAATTATGGCATTG	Rd1-rvs	GATCGCATAAGGTTTCCCTGC
M94X	Rd1-fwd	NNKCAGAATGATAATTATGGCATTGTGC	Rd1-rvs	CCGGATCGCATAAGGTTTCC
Q95X	Rd1-fwd	NNKAATGATAATTATGGCATTGTGCTTC	Rd1-rvs	CATCCGGATCGCATAAGGTT
D97X	Rd1-fwd	NNKAATTATGGCATTGTGCTTCCC	Rd1-rvs	ATTCTGCATCCGGATCGC
Y99X	Rd1-fwd	NNKGGCATTGTGCTTCCCTTAAAC	Rd1-rvs	ATTATCATTCTGCATCCGGATCG
I101X	Rd1-fwd	NNKGTGCTTCCCTTAAACTCAGCAA	Rd1-rvs	GCCATAATTATCATTCTGCATCCG
W111X	Rd1-fwd	NNKAGGTCTCTACCACACGACCAGC	Rd1-rvs	TTCCTTTGCTGAGTTTAAGGGA
R112X	Rd1-fwd	NNKTCTCTACCACACGACCAGCTG	Rd1-rvs	CCATTCCTTTGCTGAGTTTAAGG
S120X	Rd1-fwd	NNKTTTCTTCAAGCCTTGGCAGA	Rd1-rvs	CAGCTGGTCGTGTGGTAGAGA
F121X	Rd1-fwd	NNKCTTCAAGCCTTGGCAGAGTTT	Rd1-rvs	GCTCAGCTGGTCGTGTGGT
L122X	Rd1-fwd	NNKCAAGCCTTGGCAGAGTTTGC	Rd1-rvs	AAAGCTCAGCTGGTCGTGTG
Q123X	Rd1-fwd	NNKGCCTTGGCAGAGTTTGCG	Rd1-rvs	AAGAAAGCTCAGCTGGTCGTG
A124X	Rd1-fwd	NNKTTGGCAGAGTTTGCGGATAT	Rd1-rvs	TTGAAGAAAGCTCAGCTGGTCG
L125X	Rd1-fwd	NNKGCAGAGTTTGCGGATATGAACA	Rd1-rvs	GGCTTGAAGAAAGCTCAGCTG
A126X	Rd1-fwd	NNKGAGTTTGCGGATATGAACATGTAC	Rd1-rvs	CAAGGCTTGAAGAAAGCTCAGC
E127X	Rd1-fwd	NNKTTTGCGGATATGAACATGTACTGC	Rd1-rvs	TGCCAAGGCTTGAAGAAAGC
F128X	Rd1-fwd	NNKGCGGATATGAACATGTACTGCG	Rd1-rvs	CTCTGCCAAGGCTTGAAGAAA
A129X	Rd1-fwd	NNKGATATGAACATGTACTGCGACGTC	Rd1-rvs	AAACTCTGCCAAGGCTTGAAG
D130X	Rd1-fwd	NNKATGAACATGTACTGCGACGTCA	Rd1-rvs	CGCAAACTCTGCCAAGGC
M131X	Rd1-fwd	NNKAACATGTACTGCGACGTCACG	Rd1-rvs	ATCCGCAAACTCTGCCAAG
N132X	Rd1-fwd	NNKATGTACTGCGACGTCACGGA	Rd1-rvs	CATATCCGCAAACTCTGCCA
M133X	Rd1-fwd	NNKTACTGCGACGTCACGGACA	Rd1-rvs	GTTCATATCCGCAAACTCTGCC
Y134X	Rd1-fwd	NNKTGCGACGTCACGGACAGG	Rd1-rvs	CATGTTCATATCCGCAAACTCTG
D139X	Rd1-fwd	NNKAGGACACCCATTGAAGCTGTT	Rd1-rvs	CGTGACGTCGCAGTACATGTT
	Rd4-fwd	NNKAGGACACGCATTGAAGCTGTT		
P142X	Rd1-fwd	NNKATTGAAGCTGTTCATAGTTGCAACA	Rd1-rvs	TGTCCTGTCCGTGACGTCG
I143X	Rd1-fwd	NNKGAAGCTGTTCATAGTTGCAACAAC	Rd1-rvs	GGGTGTCCTGTCCGTGACG
			Rd4-rvs	GCGTGTCCTGTCCGTGACG
V146X	Rd1-fwd	NNKCATAGTTGCAACAACGCTGAATC	Rd1-rvs	AGCTTCAATGGGTGTCCTGTC
			Rd4-rvs	AGCTTCAATGCGTGTCCTGTC
L203X	Rd1-fwd	NNKCTCGGACCAGATACAGCCCC	Rd1-rvs	CAAGGCCATAGTAACCGTTGAG
R232X	Rd1-fwd	NNKACGGGATATCCGCGCA	Rd1-rvs	TCGATAACAACTGCTCATAATTGC
G234X	Rd1-fwd	NNKTATCCGCGCATCCTGCG	Rd1-rvs	CGTTCGTCGATAACAACTGCTC
Y235X	Rd1-fwd	NNKCCGCGCATCCTGCG	Rd1-rvs	TCCCGTTCGTCGATAACAAC
P236X	Rd1-fwd	NNKCGCATCCTGCGGCCA	Rd1-rvs	ATATCCCGTTCGTCGATAACAA
R237X	Rd1-fwd	NNKATCCTGCGGCCATTCGT	Rd1-rvs	CGGATATCCCGTTCGTCG
I238X	Rd1-fwd	NNKCTGCGGCCATTCGTGTG	Rd1-rvs	GCGCGGATATCCCGTTC
L239X	Rd1-fwd	NNKCGGCCATTCGTGTGGC	Rd1-rvs	GATGCGCGGATATCCCG
R240X	Rd1-fwd	NNKCCATTCGTGTGGCGCTT	Rd1-rvs	CAGGATGCGCGGATATCC
P241X	Rd1-fwd	NNKTTCGTGTGGCGCTTTTCG	Rd1-rvs	CCGCAGGATGCGCGG
F242X	Rd1-fwd	NNKGTGTGGCGCTTTTCGTCC	Rd1-rvs	TGGCCGCAGGATGCG
L253X	Rd1-fwd	NNKAGGAAACACTTATCTCTCGTGAGAGA	Rd1-rvs	ATTTCGGCATTCGGACGA
D321X	Rd1-fwd	NNKGATCTCATTTTCTACCATTTCGAAC	Rd1-rvs	TGCCAGCAACTGTACCTGTGC
	Rd2-fwd	NNKGATCTCATTTTCTACCATTTCATGC		
	Rd3-fwd	NNKGAGCTCATTTTCTACCATTTCATGC		
D322X	Rd1-fwd	NNKCTCATTTTCTACCATTTCGAACTCTG	Rd1-rvs	ATCTGCCAGCAACTGTACCTGT
	Rd2-fwd	NNKCTCATTTTCTACCATTTCATGCTCTG		
	Rd7-fwd	NNKCTCATTTTCTACCATTTCGAACTCCG		
L323X	Rd1-fwd	NNKATTTTCTACCATTTCGAACTCTGTAAA	Rd1-rvs	ATCATCTGCCAGCAACTGTACC
	Rd2-fwd	NNKATTTTCTACCATTTCATGCTCTGTAAA	Rd3-rvs	CTCATCTGCCAGCAACTGTACC
	Rd7-fwd	NNKATTTTCTACCATTTCGAACTCCGTAAA		
I324X	Rd1-fwd	NNKTTCTACCATTTCGAACTCTGTAAACC	Rd1-rvs	GAGATCATCTGCCAGCAACTGT
	Rd2-fwd	NNKTTCTACCATTTCATGCTCTGTAAACC	Rd3-rvs	GAGCTCATCTGCCAGCAACTGT
	Rd7-fwd	NNKTTCTACCATTTCGAACTCCGTAAACC		
F325X	Rd1-fwd	NNKTACCATTTCGAACTCTGTAAACCG	Rd1-rvs	AATGAGATCATCTGCCAGCAAC
	Rd2-fwd	NNKTACCATTTCATGCTCTGTAAACCG	Rd3-rvs	AATGAGCTCATCTGCCAGCAAC

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
	Rd7-fwd	NNKTACCATTTCGAACTCCGTAAACCG		
Y326X	Rd1-fwd	NNKCATTTCGAACTCTGTAAACCGAC	Rd1-rvs	GAAAATGAGATCATCTGCCAGC
	Rd2-fwd	NNKCATTTCATGCTCTGTAAACCGAC	Rd3-rvs	GAAAATGAGCTCATCTGCCAGC
	Rd7-fwd	NNKCATTTCGAACTCCGTAAACCGAC		
H327X	Rd1-fwd	NNKTTCGAACTCTGTAAACCGACCG	Rd1-rvs	GTAGAAAATGAGATCATCTGCCAGC
	Rd2-fwd	NNKTTCATGCTCTGTAAACCGACCG	Rd3-rvs	GTAGAAAATGAGCTCATCTGCCAGC
	Rd7-fwd	NNKTTCGAACTCCGTAAACCGACCG		
F328X	Rd1-fwd	NNKGAACTCTGTAAACCGACCGCA	Rd1-rvs	ATGGTAGAAAATGAGATCATCTGCC
	Rd2-fwd	NNKATGCTCTGTAAACCGACCGCA	Rd3-rvs	ATGGTAGAAAATGAGCTCATCTGCC
	Rd7-fwd	NNKGAACTCCGTAAACCGACCGCA		
E329X	Rd1-fwd	NNKCTCTGTAAACCGACCGCATTC	Rd1-rvs	GAAATGGTAGAAAATGAGATCATCTG
	Rd4-fwd	NNKCTCTGTAAACCGACCGCATAC	Rd3-rvs	GAAATGGTAGAAAATGAGCTCATCTG
	Rd7-fwd	NNKCTCCGTAAACCGACCGCATTC		
L330X	Rd1-fwd	NNKTGTAAACCGACCGCATTCAA	Rd1-rvs	TTCGAAATGGTAGAAAATGAGATCA
	Rd4-fwd	NNKTGTAAACCGACCGCATACAA	Rd2-rvs	CATGAAATGGTAGAAAATGAGATCA
	Rd7-fwd	NNKCGTAAACCGACCGCATTCAA	Rd3-rvs	CATGAAATGGTAGAAAATGAGCTCA
C331X	Rd1-fwd	NNKAAACCGACCGCATTCAACA	Rd1-rvs	GAGTTCGAAATGGTAGAAAATGAGA
	Rd4-fwd	NNKAAACCGACCGCATACAACA	Rd2-rvs	GAGCATGAAATGGTAGAAAATGAGA
			Rd3-rvs	GAGCATGAAATGGTAGAAAATGAGC
K332X	Rd1-fwd	NNKCCGACCGCATTCAACATCA	Rd1-rvs	ACAGAGTTCGAAATGGTAGAAAATG
	Rd4-fwd	NNKCCGACCGCATACAACATCA	Rd2-rvs	ACAGAGCATGAAATGGTAGAAAATG
			Rd7-rvs	ACGGAGTTCGAAATGGTAGAAAATG
P333X	Rd1-fwd	NNKACCGCATTCAACATCATCTTCC	Rd1-rvs	TTTACAGAGTTCGAAATGGTAGAAAA
	Rd4-fwd	NNKACCGCATACAACATCATCTTCC	Rd2-rvs	TTTACAGAGCATGAAATGGTAGAAAA
			Rd7-rvs	TTTACGGAGTTCGAAATGGTAGAAAA
T334X	Rd1-fwd	NNKGCATTCAACATCATCTTCCAGC	Rd1-rvs	CGGTTTACAGAGTTCGAAATGG
	Rd4-fwd	NNKGCATACAACATCATCTTCCAGC	Rd2-rvs	CGGTTTACAGAGCATGAAATGG
			Rd7-rvs	CGGTTTACGGAGTTCGAAATGG
F336X	Rd1-fwd	NNKAACATCATCTTCCAGCTGTATGC	Rd1-rvs	TGCGGTCGGTTTACAGAGTTC
			Rd2-rvs	TGCGGTCGGTTTACAGAGCAT
			Rd7-rvs	TGCGGTCGGTTTACGGAGTTC
D393X	Rd1-fwd	NNKATCTCCGGTTTTGTCAGCTTTC	Rd1-rvs	GTACAGTCGAAAGGTCTCCTTTGT
I394X	Rd1-fwd	NNKTCCGGTTTTGTCAGCTTTCG	Rd1-rvs	ATCGTACAGTCGAAAGGTCTCCT
S395X	Rd1-fwd	NNKGGTTTTGTCAGCTTTCGCCG	Rd1-rvs	GATATCGTACAGTCGAAAGGTCTCC
	Rd7-fwd	NNKGGTTTTGTCAGCTTTCGGCG		
G396X	Rd1-fwd	NNKTTTGTCAGCTTTCGCCGTG	Rd1-rvs	GGAGATATCGTACAGTCGAAAGGT
	Rd7-fwd	NNKTTTGTCAGCTTTCGGCGTG		
F397X	Rd1-fwd	NNKGTCAGCTTTCGCCGTGTCA	Rd1-rvs	ACCGGAGATATCGTACAGTCGA
	Rd7-fwd	NNKGTCAGCTTTCGGCGTGTCA		
V398X	Rd1-fwd	NNKAGCTTTCGCCGTGTCATGA	Rd1-rvs	AAAACCGGAGATATCGTACAGTCG
	Rd7-fwd	NNKAGCTTTCGGCGTGTCATGA		
S399X	Rd1-fwd	NNKTTTCGCCGTGTCATGAAACC	Rd1-rvs	GACAAAACCGGAGATATCGTACA
	Rd7-fwd	NNKTTTCGGCGTGTCATGAAACC		
F400X	Rd1-fwd	NNKCGCCGTGTCATGAAACCTC	Rd1-rvs	GCTGACAAAACCGGAGATATCG
	Rd7-fwd	NNKCGGCGTGTCATGAAACCTC		
R401X	Rd1-fwd	NNKCGTGTCATGAAACCTCTTACTCTAAA	Rd1-rvs	AAAGCTGACAAAACCGGAGATA
I421X	Rd1-fwd	NNKTTATTGTCTCCATGTCGGAATGT	Rd1-rvs	TGTACCGGGACGAAGAGACA
L422X	Rd1-fwd	NNKTTGTCTCCATGTCGGAATGTCC	Rd1-rvs	GATTGTACCGGGACGAAGAGA
L423X	Rd1-fwd	NNKTCTCCATGTCGGAATGTCCA	Rd1-rvs	TAAGATTGTACCGGGACGAAGA
S424X	Rd1-fwd	NNKCCATGTCGGAATGTCCATTTG	Rd1-rvs	CAATAAGATTGTACCGGGACGA
P425X	Rd1-fwd	NNKTGTCGGAATGTCCATTTGGA	Rd1-rvs	AGACAATAAGATTGTACCGGGACG
C426X	Rd1-fwd	NNKCGGAATGTCCATTTGGATCC	Rd1-rvs	TGGAGACAATAAGATTGTACCGG
T468X	Rd1-fwd	NNKTTTTCTCATGGAGCGGGC	Rd1-rvs	CAAGAACGTGAGCGAGGTGG
F469X	Rd1-fwd	NNKTCTCATGGAGCGGGCAG	Rd1-rvs	TGTCAAGAACGTGAGCGAGG
			Rd7-rvs	TGACAAGAACGTGAGCGAGG
S470X	Rd1-fwd	NNKCATGGAGCGGGCAGCT	Rd1-rvs	AAATGTCAAGAACGTGAGCGA
			Rd7-rvs	AAATGACAAGAACGTGAGCGA
A473X	Rd1-fwd	NNKGGCAGCTGTCCCGCG	Rd1-rvs	TCCATGAGAAAATGTCAAGAACG
			Rd7-rvs	TCCATGAGAAAATGACAAGAACG
G474X	Rd1-fwd	NNKAGCTGTCCCGCGCGG	Rd1-rvs	CGCTCCATGAGAAAATGTCAA
			Rd7-rvs	CGCTCCATGAGAAAATGACAA

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
S475X	Rd1-fwd	NNKTGTCCCGCGCGGG	Rd1-rvs	GCCCGCTCCATGAGAAAA
P477X	Rd1-fwd	NNKGCGCGGGTTCTCGCT	Rd1-rvs	ACAGCTGCCCGCTCCA
F511X	Rd1-fwd	NNKACCAGTGGTCCGGTCTATATGC	Rd1-rvs	TCCATAGGGTAGTATCTCCTTCTGC
	Rd3-fwd	NNKACCAGGGGTCCGATGTATATGC		
T512X	Rd1-fwd	NNKAGTGGTCCGGTCTATATGCCTAA	Rd1-rvs	GAATCCATAGGGTAGTATCTCCTTCTG
	Rd3-fwd	NNKAGGGGTCCGATGTATATGCCTAA		
S513X	Rd1-fwd	NNKGGTCCGGTCTATATGCCTAATCC	Rd1-rvs	GGTGAATCCATAGGGTAGTATCTCC
	Rd3-fwd	NNKGGTCCGATGTATATGCCTAATCC		
G514X	Rd1-fwd	NNKCCGGTCTATATGCCTAATCCATC	Rd1-rvs	ACTGGTGAATCCATAGGGTAGTATCT
	Rd3-fwd	NNKCCGATGTATATGCCTAATCCATC	Rd3-rvs	CCTGGTGAATCCATAGGGTAGTATCT
P515X	Rd1-fwd	NNKGTCTATATGCCTAATCCATCAGTGATG	Rd1-rvs	ACCACTGGTGAATCCATAGGG
	Rd3-fwd	NNKATGTATATGCCTAATCCATCAGTGATG	Rd3-rvs	ACCCCTGGTGAATCCATAGGG
V516X	Rd1-fwd	NNKTATATGCCTAATCCATCAGTGATGA	Rd1-rvs	CGGACCACTGGTGAATCCA
			Rd3-rvs	CGGACCCCTGGTGAATCCA
Y517X	Rd1-fwd	NNKATGCCTAATCCATCAGTGATGATG	Rd1-rvs	GACCGGACCACTGGTGAATC
			Rd3-rvs	CATCGGACCCCTGGTGAATC
M518X	Rd1-fwd	NNKCCTAATCCATCAGTGATGATGAGA	Rd1-rvs	ATAGACCGGACCACTGGTGAA
			Rd3-rvs	ATACATCGGACCCCTGGTGAA

Table 3.2. Primers for multiple site-directed mutagenesis combinatorial libraries. Mutated codons highlighted in red.

Mutation(s)	Primer (5′ to 3′)	
L122X, A124S	Rd2-combi-1	CGACCAGCTGAGCTTTNNKCAATCCTTGGCAGAGTTTGC
I143R	Rd2-combi-2	CGGACAGGACACCCAGGGAAGCTGTTCATAGTTGCAAC
L253G	Rd2-combi-3	CGTCCGAATGCCGAAATGGGAGGAAACACTTATCTCTCGTG
D322E	Rd2-combi-4	CAGTTGCTGGCAGATGAGCTCATTTTCTACCATTTCATGC
S513R, V516M	Rd2-combi-5	CTACCCTATGGATTCACCAGGGGTCCGATGTATATGCCTAATCCATCAGTG
P142R	Rd3-combi-1	GACGTCACGGACAGGACACGCATTGAAGCTGTTCATAG
P142G	Rd3-combi-2	GACGTCACGGACAGGACAGGCATTGAAGCTGTTCATAG
V146A	Rd3-combi-3	GACACCCATTGAAGCTGCTCATAGTTGCAACAACG
F336Y	Rd3-combi-4	CTGTAAACCGACCGCATACAACATCATCTTCCAGC
M516V	Rd3-combi-5	CCTATGGATTCACCAGGGGTCCGGTGTATATGCCTAATCC
R513M	Rd3-combi-6	CCTATGGATTCACCAGTGGTCCGATGTATATGCCTAATCC
Q95T	Rd4-combi-1	GAAACCTTATGCGATCCGGATGACGAATGATAATTATGGCATTG
V146A	Rd4-combi-2	GACACGCATTGAAGCTGCTCATAGTTGCAACAACG
I324A	Rd4-combi-3	GCTGGCAGATGAGCTCGCTTTCTACCATTTCATGCTC
F325L	Rd4-combi-4	GGCAGATGAGCTCATTTTGTACCATTTCATGCTCTGTAAAC
G396W	Rd4-combi-5	CGACTGTACGATATCTCCTGGTTTGTCAGCTTTCGCCG
T468S	Rd4-combi-6	CACCTCGCTCACGTTCTTGTCATTTTCTCATGGAGC
M516V	Rd4-combi-7	GATTCACCAGGGGTCCGGTGTATATGCCTAATC
Q95T	Rd5-combi-1	GAAACCTTATGCGATCCGGATGACGAATGATAATTATGGCATTG
W111M	Rd5-combi-2	CTTAAACTCAGCAAAGGAAATGAGGTCTCTACCACACGACC
W111T	Rd5-combi-3	CTTAAACTCAGCAAAGGAAACGAGGTCTCTACCACACGACC
V146A	Rd5-combi-4	GACACGCATTGAAGCTGCTCATAGTTGCAACAACG
L253V	Rd5-combi-5	CTTTTCGTCCGAATGCCGAAATGTTAGGAAACACTTATCTCTCGTG
C331R	Rd5-combi-6	CATTTTCTACCATTTCATGCTCCGTAAACCGACCGCATACAAC
R401Q	Rd5-combi-7	CCTGGTTTGTCAGCTTTCAGCGTGTCATGAAACCTCTTAC
Q95T	Rd6-combi-1	CCTTATGCGATCCGGATGACGAATGATAATTATGGCATTGTG
I143R	Rd6-combi-2	CGGACAGGACACGCCGTGAAGCTGTTCATAGTTG
I143R, V146A	Rd6-combi-3	CGGACAGGACACGCCGTGAAGCTGCTCATAGTTGCAAC
V146A	Rd6-combi-4	GACACGCATTGAAGCTGCTCATAGTTGCAACAACG
L253V	Rd6-combi-5	GTCCGAATGCCGAAATGTGAGGAAACACTTATCTC
I324A	Rd6-combi-6	CAGTTGCTGGCAGATGAGCTCGCTTCTACCATTTC
M329E	Rd6-combi-7	TTTCTACCATTTCGAGCTCCGTAAACCGACCGCATACAAC
M329E, R331C	Rd6-combi-8	TTTCTACCATTTCGAGCTCTGTAAACCGACCGCATACAACATCATC
R331C	Rd6-combi-9	CTACCATTTCATGCTCTGTAAACCGACCGCATAC
W396G	Rd6-combi-10	GACTGTACGATATCTCCGGGTTTGTCAGCTTTC
W396V	Rd6-combi-11	CCTTTCGACTGTACGATATCTCCGTGTTTGTCAGCTTTC
Q401R	Rd6-combi-12	GTTTGTCAGCTTTCGGCGTGTCATGAAACCTC

T468S	Rd6-combi-13	CTCGCTCACGTTCTTGTCATTTTCTCATGGAGC
R69A	Rd7-combi-1	CAAAGCTACGATGGAAGGCCTTCGATGCGGAGAAAG
R69K	Rd7-combi-2	GCCAAAGCTACGATGGAAGAAGTTCGATGCGGAGAAAGAGTATG
L122G	Rd7-combi-3	GACCAGCTGAGCTTTGGTCAAGCCTTGGCAGAGTTTG
L122G, A126G	Rd7-combi-4	GACCAGCTGAGCTTTGGTCAAGCCTTGGGAGAGTTTGCGG
A126G	Rd7-combi-5	CTTTCTTCAAGCCTTGGGAGAGTTTGCGGATATG
I143R	Rd7-combi-6	CGGACAGGACACGCCGTGAAGCTGTTCATAGTTG
T334R	Rd7-combi-7	CATGCTCCGTAAACCGCGCGCATACAACATCATCTTC
W396G	Rd7-combi-8	GACTGTACGATATCTCCGGGTTTGTCAGCTTTCG
L422G	Rd7-combi-9	CGTCCCGGTACAATCGGATTGTCTCCATGTCGG
A473K	Rd7-combi-10	GTCATTTTCTCATGGAAAGGGCAGCTGTCCCGCG
Y517S	Rd7-combi-11	CCAGGGGTCCGATGAGTATGCCTAATCCATCAGTG

High-throughput biotransformations. Histidine dropout minimal media containing 4% glucose was added to each well of up to 20 sterile 96-well culture plates (VWR) to a volume of 500 µL. *S. cerevisiae* library colonies were inoculated into each well of the plate except for row E, which was reserved for *S. cerevisiae* KtnC template control colonies. Cultures were grown overnight at 30 °C with shaking at 300 rpm. Culture plates were centrifuged at 1,000 x *g* for 10 min to pellet cells and the growth media was removed. Cells were resuspended with 250 µL of histidine dropout minimal media containing 6% galactose and 100–750 µM each substrate. Biotransformations were incubated at 30 °C with shaking at 300 rpm for 2–3 days. Reactions were quenched with the addition of 2 equivalents of methanol and centrifugation at 1,000 x *g* for 10 min. The resulting supernatant was used for mass spectrometry analysis.

RapidFire–MS screen for activity (tier 1). Library reactions were analyzed for total cross-coupling activity in a high-throughput tier 1 screening using an Agilent RapidFire 365 High-throughput Mass Spectrometry System linked to an Agilent 6545 LC–Q-TOF. Library samples were prepared by diluting the library samples 2:1 in water containing 600 nM 2,6′-dihydroxyacetophenone (internal standard) in shallow 96-well plates (Axygen) and heat-sealed with pierceable aluminum covers (Agilent). Each RapidFire injection was programmed as follows: 0.6 s to aspirate the sample, 3 s to load the sample onto the C4 cartridge with water (0.5 mL/min), 8 s to elute the sample from the cartridge using 90:10 acetonitrile/water (v/v) buffered with 10 mM ammonium formate pH 8 (0.8 mL/min), and 0.5 s to reequilibrate. A blank injection was included after each sample injection to reduce carry-over between library reactions.

The peak areas for the extracted ion chromatograms for internal standard, substrate, and cross-coupled product were collected for each sample using the MassHunter Qualitative Analysis 10.0 software. The relative percent conversions to cross-coupled products were calculated and the average percent conversion resulting from the template reactions for each plate (n=10) was set to

1.0. The conversions for the variant reactions were normalized to the template reactions, giving fold improvement scores for each variant.

LC–MS screen for selectivity (tier 2). Reactions were subjected to liquid chromatography PDA spectrometry (UPLC) analysis performed on an Agilent 6230 time of flight mass spectrometer with a Dual AJS ESI source and an Agilent 1290 Infinity Series II diode array detector, autosampler, and binary pump. To determine site-selectivity of reaction, samples were analyzed by

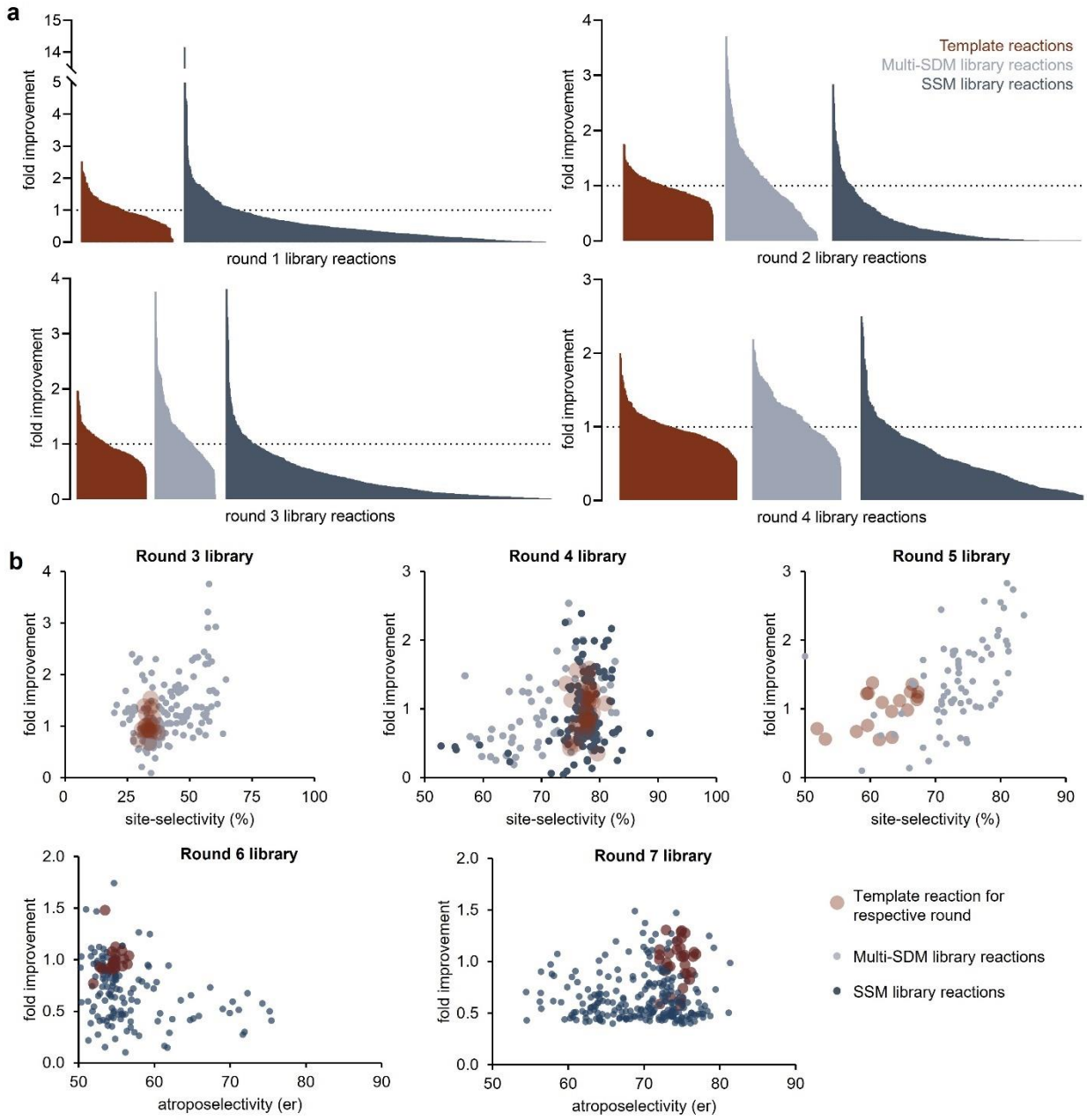
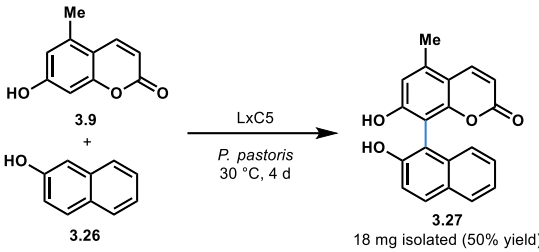


Figure 3.7. Summary of engineering data from each round of evolution. (a) Tier 1 RapidFire–MS analysis. (b) Tier 2 LC–MS analysis for selectivity.

negative mode on a Waters Acquity Premier HSS T3 1.8 μm C18, 2.1 x 50 mm column. Phase A = 95:5 water/acetonitrile (v/v); phase B = 95:5 acetonitrile/water (v/v). The pump was programmed as follows: 20% B held for 0.5 min, to 60% B over 1.0 min, hold at 90% B for 0.6 min. The flow rate was set to 0.7 mL/min flow rate, the column was heated to 45 °C, and each injection was followed by equilibration at 20% B for 0.5 min. To determine atroposelectivity of reaction, samples were analyzed by negative mode on a Daicel Chiralpak IC-3, 4.6 x 150 mm, 3 μm column. Phase A = 95:5 water/acetonitrile (v/v); phase B = 95:5 acetonitrile/water (v/v). The pump was programmed as follows: 50% B held for 2.25 min (1.5 mL/min), 100% B held over 0.5 min (2.5 mL/min), to 50% B over 0.5 min (2.5 mL/min). The MassHunter software was used to extract ion chromatograms at the target masses to characterize the percent conversion and site-selectivity of the reactions. Variants that gave the desired activity and site-selectivity were carried forward to the hit validation and sequencing step.

Scheme 3.1. Preparative scale biotransformation with LxC5 to form 3.27.



Hit validation and sequencing. Hit variants selected validated by growing up glycerol stocks of the respective cells harboring the variants, repeating biotransformations (n=3), and analyzing reactions by LC–MS using previously described procedures. Variants that showed improvement over the template after validation were sequenced to identify the active site mutations. Plasmid DNA was isolated from *S. cerevisiae* using the Zymoprep Yeast Plasmid Miniprep II kit. The resulting plasmid DNA was used to transform DH5α *E. coli* cells. Clean plasmid DNA was isolated from DH5α cells using QIAGEN miniprep kit and sent to GENEWIZ for Sanger sequencing. The sequence for the top variant from each round was defined as LxC (Lara's cross-couplase) and used as the template for the following round.

Preparative scale biotransformation with engineered LxC5 variant. A colony of *P. pastoris* KM71 containing LxC5 was inoculated in 10 mL BMG medium and grown overnight at 30 °C with shaking at 235 rpm. The starter culture was used to inoculate 6 L of BMG and grown at 30 °C with shaking at 235 rpm until an OD_{600} of 6 was reached. Cultures were induced for expression by centrifugation and resuspension in BMM medium to an OD_{600} of 25. Substrate (0.113 mmol coumarin **3.9** and 0.226 mmol naphthol **3.26**; concentration of limiting reactant = 0.1

mM) was added from a 50 mM stock in DMSO to a total of 1.135 L BMM separated across two 2.8-L flasks. Cultures were grown at 30 °C with shaking at 200 rpm for 4 d. To isolate the reaction products, the cultures were combined in a 2-L separatory funnel. This mixture was extracted with 1 L of a solvent mixture of 8:1:1 ethyl acetate/toluene/isopropanol (v/v). The extraction was repeated a total of three times. The organic extracts were combined, washed with brine (600 mL) and concentrated in a rotary evaporator. The concentrate was purified over silica gel (gradient of 20–80% ethyl acetate in hexanes) to afford the cross-coupling product **3.27** (18 mg, 50% yield) as a light brown solid. **^{1}H NMR** (600 MHz, CD_3OD) δ 8.17 (d, J = 9.6 Hz, 1H), 7.80 (t, J = 9.1 Hz, 2H), 7.26 (ddd, J = 6.9, 4.5, 1.8 Hz, 2H), 7.22 (d, J = 8.8 Hz, 1H), 7.21–7.16 (m, 1H), 6.88 (s, 1H), 6.18 (d, J = 9.6 Hz, 1H), 2.59 (s, 3H); **^{13}C NMR** (150 MHz, CD_3OD) δ 162.9, 159.4, 154.6, 152.8, 142.2, 13735, 134.0, 129.3, 128.8, 127.6, 125.8, 123.7, 122.3, 117.7, 114.2, 111.3, 111.2, 109.9, 108.9, 17.1; **HRMS** (ESI) *m/z* calculated for $C_{20}H_{15}O_4^+$ $[M+H]^+$ 319.0965, found 319.0970.

Assignment of absolute configuration of biocatalytic product 3.27. The general approach for absolute configuration assignment using ECD, including the detailed computational workflow, has been published elsewhere.[43-45] A subset of the details of the computational methodology is provided here. Conformers of each test structure were geometry optimized at the B3LYP/6-31G** level and stationary points were confirmed by performing frequency calculations.[46-54] All calculations were performed using Gaussian 09.[55] Output conformers were ranked according to DFT energy and a clustering was performed in order to remove duplicates. Initial duplicate identification was performed solely on an electronic energy basis where two compounds were considered identical if the difference in Hartrees was less than 0.01. Rounding the differences led to inconsistencies in identification of duplicates. It became better to cluster the DFT minima by energy and then re-cluster each energy bucket by structure using an all atom RMS of 0.6 Å. This process removed just identical compounds. Two Boltzmann distributions were calculated based on the free energy (G) and the electronic energy (E). To calculate UV and ECD spectra, B3LYP geometries were used as input. The spectra were then calculated using either the B3LYP or CAM-B3LYP[56] functionals, along with the 6-31++G** basis set[57, 58] in vacuo. Only conformers which contributed more than 5.0% to the total in vacuo conformer distribution were selected for UV and ECD calculation. Time-dependent Density Functional Theory (TDDFT)[59] methodology was employed using the following keywords: TD=full,singlet, Nstates=100, and integral=ultrafinegrid. Spectral display, Boltzmann weighting, and curve fitting were carried out using SpecDis.[60,

[61] and were displayed with a wavelength shift and band broadening sigma values in order to best match the calculated and experimental UV spectra. This shift and band broadening were then applied to the ECD spectra, and the area under the curve fit was determined by SpecDis. Conformers were visualized using CYLview Version 1.0.[62] Calculations of the ECD and UV spectra (CAM-B3LYP/6-31++G**) involved modeling the (*P*)-enantiomer of the compound. Hence, the (*P*)-enantiomer had a positive dihedral angle about the atropisomeric bond. Since the atropisomeric bond represents the only stereochemical element in the molecule, the (*M*)-enantiomer is assumed to have a spectrum that will be equal and opposite at all wavelengths.

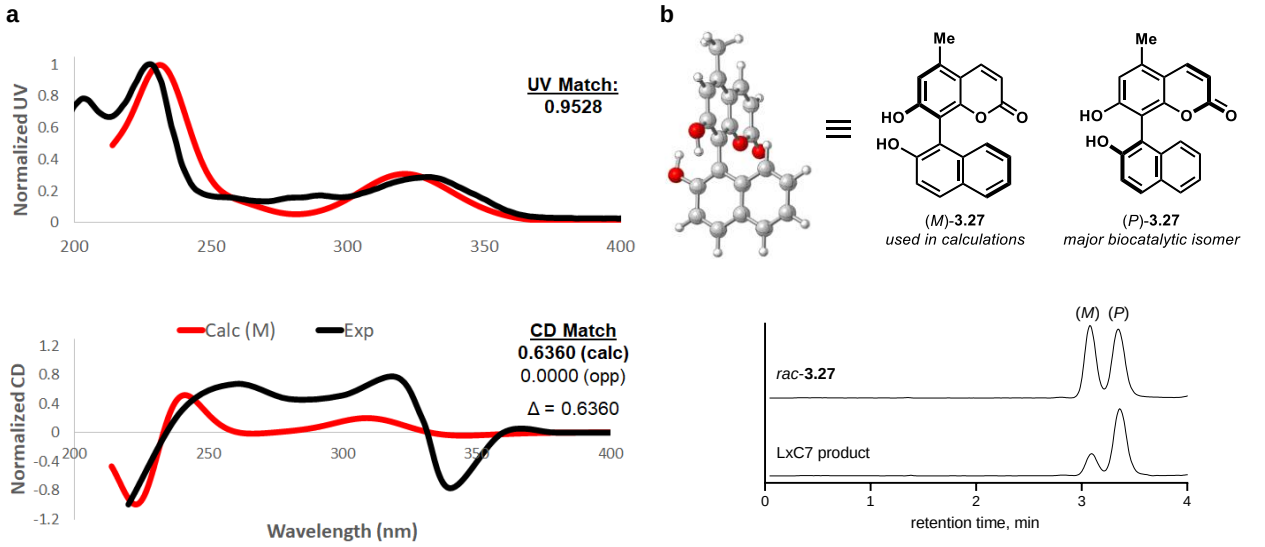


Figure 3.8. Assignment of absolute configurations of cross-coupled product. (a) Comparison between compound **3.27** (racemic, peak 1) experimental (black) and calculated (red) UV (top) and ECD (bottom) spectra. The calculated (*M*)-enantiomer is the better match to the experimental spectrum, with a large difference in fit (Δ=0.6360) suggesting a confident assignment. The calculated spectrum has been shifted 14 nm, and a band broadening of 0.30 eV has been applied. (b) One conformer of the (*M*)-enantiomer of compound **3.27** that contribute >2% to the Boltzmann distribution. Assigned absolute configurations of compound **3.27** based on ECD analysis. The first peak (rt = 3.1 min) for compound **3.27** in chiral HPLC (CHIRALPAK IC-3) is the (*M*)-isomer. The second peak (rt = 3.4 min) is the major atropisomer obtained using LxC7 and is therefore the (*P*)-isomer.

SUPPLEMENTARY INFORMATION

The supplementary information for this chapter includes (I) DNA and protein sequences, (II) chemical synthesis of substrates and authentic product standards, (III) standard curves and LC–MS traces of biocatalytic reactions, and (IV) NMR spectra of synthesized compounds.

DNA and protein sequences. The gene encoding KtnC in a pJET1.2 cloning plasmid and a pESC-HIS expression plasmid were generous gifts from Professor Michael Müller's lab at the Albert-Ludwigs-Universität Freiburg. The gene encoding KtnC was amplified and cloned into a pPIC3.5 plasmid (Invitrogen) for functional expression in *P. pastoris*.

KtnC

```
ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACCCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCA
TTTCGAACTCTGTAAACCGACCGCATTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTT
TGTCAGCTTTCGCCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTG
GTCCGGTCTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG
```

```
MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFELCKPTAFNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGK
```

LxC1 (KtnC_E329M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACCCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGATCTCATTTTCTACCA
TTTCATGCTCTGTAAACCGACCGCATTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTT
TGTCAGCTTTCGCCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGTG
GTCCGGTCTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADDLIFYHFMLCKPTAFNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTSGPVYMPNPSVMMRIRPRSDGK

LxC2 (KtnC_D322E, E329M, S513R, V516M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACCCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA

TTTCATGCTCTGTAAACCGACCGCATTCAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTT
TGTCAGCTTTCGCCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TPIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLCKPTAFNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

LxC3 (KtnC_P142R, D322E, E329M, F336Y, S513R, V516M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACGCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA
TTTCATGCTCTGTAAACCGACCGCATACAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCGGTTT
TGTCAGCTTTCGCCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TRIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLCKPTAFNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISGFVSFRRVMKPLTLNSIGLSLRPGT

ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

LxC4 (KtnC_P142R, D322E, E329M, F336Y, G396W, S513R, V516M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACGCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA
TTTCATGCTCTGTAAACCGACCGCATACAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCTGGTT
TGTCAGCTTTCGCCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TRIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLCKPTAYNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISWFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

LxC5 (KtnC_P142R, D322E, E329M, C331R, F336Y, G396W, R401Q, S513R, V516M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACGCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT

ATCTCTCGTGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA
TTTCATGCTCCGTAAACCGACCGCATACAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCTGGTT
TGTCAGCTTTCAGCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GACATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TRIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLRKPTAYNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISWFVSFQRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLTFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

LxC6 (KtnC_P142R, D322E, E329M, C331R, F336Y, G396W, T468S, S513R, V516M)

ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGTTTGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACGCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA
TTTCATGCTCCGTAAACCGACCGCATACAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCTGGTT
TGTCAGCTTTCGGCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GTCATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEFADMNMYCDVTDR
TRIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP

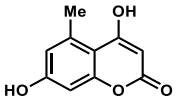
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLRKPTAYNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISWFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLSFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

LxC7 (KtnC_F128V, P142R, D322E, E329M, C331R, F336Y, G396W, T468S, S513R, V516M)

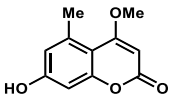
ATGGCAGTTCATGTACCATTCAGCATCCACCACGTCGTGGATATCGGCATTTCCACCGGCCCGGTAGTCA
TTGTCCTCGTTCTCCTATTCGGGTTGGCTGTCGTGGGATCCGACAGCCTTGATGGATGGTGGCAGAAAAG
AGCCTTGAGAGGCATTCCGATTGTAGACGAGGGTAGCTACATGCGGCCAAAGCTACGATGGAAGCGCTTC
GATGCGGAGAAAGAGTATGCGAGAGCATATCAGCAGTACACAAAAGCAGGGAAACCTTATGCGATCCGGA
TGCAGAATGATAATTATGGCATTGTGCTTCCCTTAAACTCAGCAAAGGAATGGAGGTCTCTACCACACGA
CCAGCTGAGCTTTCTTCAAGCCTTGGCAGAGGTGGCGGATATGAACATGTACTGCGACGTCACGGACAGG
ACACGCATTGAAGCTGTTCATAGTTGCAACAACGCTGAATCATTAAACATCCTCAATAAGCTTCTCGCCC
GGGAAACCGATACAGCCTTATCTCAGATTTTCGAGCAGCCTACAGGAAAAGACTGGAAGGAATTGAACAC
ACTGCAGACAATCCTCTCCTTATGCTCAACGGTTACTATGGCCTTGCTCCTCGGACCAGATACAGCCCCT
GACCCGGTGCTCCACCACCATTCGACGTCTTTTGGCGAAGCAATTATGAGCAGTTGTTATCGACGAACGG
GATATCCGCGCATCCTGCGGCCATTCGTGTGGCGCTTTTCGTCCGAATGCCGAAATCTGAGGAAACACTT
ATCTCTCGTGAGAGAGAGACTTGTTCCTGAGGTTGCGCGCCGTGTCGCAGCTGCACGAGCGGCGGATAAA
ACCAAGGACGTACGCCCATCTTCATTGTTGGACGCACTGATCGCGGCGGCCTTCGACAACGGCAGCTTAA
GCCCAGATGACCAAGGCAGAAATGATGCAGCACAGGTACAGTTGCTGGCAGATGAGCTCATTTTCTACCA
TTTCATGCTCCGTAAACCGACCGCATACAACATCATCTTCCAGCTGTATGCCATCATGGACCACCCAGAG
TACAAGGCTCCTCTCCGAGAGGAGGCACTCCAAGCCTTGAAGCTCACCAATGGTGACTGGACCGTTGAAA
CTCTGAAGCACGCTCCCAAGTTAGAAAGCTTTACAAAGGAGACCTTTCGACTGTACGATATCTCCTGGTT
TGTCAGCTTTCGGCGTGTCATGAAACCTCTTACTCTAAACTCCATCGGCCTGTCTCTTCGTCCCGGTACA
ATCTTATTGTCTCCATGTCGGAATGTCCATTTGGATCCCGAGATCTATGAAGACCCGACAACTTTCAATG
GCTACCGCTTCTACGACTCCAGCCGCGAGGTCTGCTCTCCACGCGTGGCAACCACCTCGCTCACGTTCTT
GTCATTTTCTCATGGAGCGGGCAGCTGTCCCGCGCGGGTTCTCGCTACTCAAATTTGTCGGACGATCTTC
ATCAAGTTCCTGTTGCAGTACGACGTAGAACCTGTGCAGAAGGAGATACTACCCTATGGATTCACCAGGG
GTCCGATGTATATGCCTAATCCATCAGTGATGATGAGAATCCGGCCAAGAAGTGACGGGAAG

MAVHVPFSIHHVVDIGISTGPVVIVLVLLFGLAVVGSDSLDGWWQKRALRGIPIVDEGSYMRPKLRWKRF
DAEKEYARAYQQYTKAGKPYAIRMQNDNYGIVLPLNSAKEWRSLPHDQLSFLQALAEVADMNMYCDVTDR
TRIEAVHSCNNAESLNILNKLLARETDTALSQIFEQPTGKDWKELNTLQTILSLCSTVTMALLLGPDTAP
DPVLHHHSTSFGEAIMSSCYRRTGYPRILRPFVWRFSSECRNLRKHLSLVRERLVPEVARRVAAARAADK
TKDVRPSSLLDALIAAAFDNGSLSPDDQGRNDAAQVQLLADELIFYHFMLRKPTAYNIIFQLYAIMDHPE
YKAPLREEALQALKLTNGDWTVETLKHAPKLESFTKETFRLYDISWFVSFRRVMKPLTLNSIGLSLRPGT
ILLSPCRNVHLDPEIYEDPTTFNGYRFYDSSREVCSPRVATTSLTFLSFSHGAGSCPARVLATQICRTIF
IKFLLQYDVEPVQKEILPYGFTRGPMYMPNPSVMMRIRPRSDGK

Chemical synthesis. Meagan Hinze, Jessica Yazarians, and Suman Chakrabarty synthesized and characterized the substrates and product standards reported in this section. All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenk techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, Inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Machery-Nagel 60 F_{254} precoated silica TLC plates (0.25 mm) or Merck Silica Gel 60 F_{254} precoated silica TLC plates (0.25 mm) which were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol green stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230–400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated.

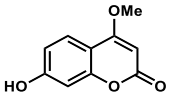


4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (3.53). To a mixture of 5-methylresorcinol (3.000 g, 24.16 mmol, 1.000 equiv), malonic acid (2.514 g, 24.16 mmol, 1.000 equiv) and $ZnCl_2$ flame-dried under vacuum (10.21 g, 74.91 mmol, 3.100 equiv) was added $POCl_3$ (60.00 mL). The reaction was heated at 60 °C and stirred for 14.5 h. The reaction was quenched by pouring into ice water and induced the precipitation of a solid, which was isolated by vacuum filtration. The crude solid was purified by flash column chromatography (5:6:1 toluene/ethyl acetate/formic acid (v/v)) to afford 3.380 g of the title compound (73% yield) as a yellow solid with minor impurities. $\mathbf{R_f}$ = 0.37 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.06 (s, 1H), 10.39 (s, 1H), 6.54 (s,1H), 6.50 (s, 1H), 5.33 (s, 1H), 2.57 (s, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 169.1, 161.9, 160.4, 157.0, 138.7, 115.6, 106.4, 100.4, 88.0, 22.8; **HRMS** (ESI) *m/z* calculated for $C_{10}H_9O_4^+$ $[M+H]^+$ 193.0495, found 193.0505. All spectra obtained were constant with literature values.[63]

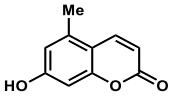


7-hydroxy-4-methoxy-5-methyl-*2H*-chromen-2-one (3.4). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**; 1.450 g, 7.546 mmol, 1.000 equiv) in MeOH (84.0 mL) was

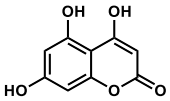
added H_2SO_4 (8.4 mL). The reaction was heated at 75 °C for 6.2 h. Incubation at 0 °C induced the precipitation of a white solid which was isolated by vacuum filtration to afford 1.070 g of the title compound (69% yield). $\mathbf{R_f}$ = 0.43 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.46 (s, 1H), 6.57 (d, J = 2.4, 1H), 6.54 (d, J = 2.4, 1H), 5.59 (s, 1H), 3.92 (s, 3H), 2.53 (s, 3H); **HRMS** (ESI) *m/z* calculated for $C_{11}H_{11}O_4^+$ $[M+H]^+$ 207.0652, found 207.0657. All spectra obtained were constant with literature values.[64]



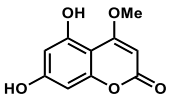
7-hydroxy-4-methoxy-*2H*-chromen-2-one (3.8). To a solution of 4,7-dihydroxycoumarin (250.0 mg, 1.403 mmol, 1.000 equiv) in MeOH (15.6 mL) was added H_2SO_4 (1.6 mL). The reaction was heated at 75 °C for 2.8 h. Incubation at 0 °C induced the precipitation of a white solid which was isolated by vacuum filtration to afford 226.2 mg of the title compound (84% yield). **^{1}H NMR** (600 MHz, DMSO-d_6) δ 7.61 (d, J =8.7, 1H), 6.78 (dd, J = 8.7, 2.3, 1H), 6.69 (d, J = 2.3, 1H), 5.67 (s, 1H), 3.96 (s, 3H); **HRMS** (ESI) *m/z* calculated for $C_{10}H_9O_4^+$ $[M+H]^+$ 193.0495, found 193.0503. All spectra obtained were constant with literature values.[65]



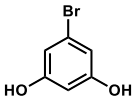
7-hydroxy-5-methyl-*2H*-chromen-2-one (3.9). Ethyl(3,3)-diethoxypropanoate (0.36 mL, 1.8 mmol, 1.0 equiv) was added neat to 5-methylresorcinol (250 mg, 2.0 mmol, 1.1 equiv). The reaction was heated at 120 °C and stirred until dissolved for 1 h. Toluene (5.0 mL) and toluene sulfonic acid monohydrate (77 mg, 0.40 mmol, 0.200 equiv) were added to the solution. The reaction was heated at 110 °C and stirred for 16 h. The reaction was quenched by pouring into ice water and further induced the precipitation of a solid, which was isolated by vacuum filtration to afford 320.0 mg of the title compound (99% yield).**^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.45 (s, 1H), 8.05 (d, J = 9.7, 1H0, 6.64 (d, J = 2.0, 1H), 6.55 (d, J = 2.0, 1H), 6.19 (d, J = 9.7, 1H), 2.42 (s, 3H); **HRMS** (ESI) *m/z* calculated for $C_{10}H_9O_3^+$ $[M+H]^+$ 177.0546, found 177.0556. All spectra obtained were constant with literature values.[66]



4,5,7-trihydroxy-*2H*-chromen-2-one (3.54). 4,5,7-trihydroxy-*2H*-chromen-2-one (**3.54**) was prepared from phloroglucinol and cyanoacetic acid according to the procedure described by Pandey et al.[67] **MP** = >310 °C; $\mathbf{R_f}$ = 0.38 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 10.40 (s, 1H), 6.17 (s, 2H), 5.19 (s, 1H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 168.3, 162.2, 161.9, 156.9, 156.3, 98.6, 96.6, 94.7, 86.6; **IR** (thin film, cm^{-1}) 3097, 2707, 2600, 1634, 1569, 1471, 1415; **HRMS** (ESI) *m/z* calculated for $C_9H_7O_5^+$ $[M+H]^+$ 195.0288, found 195.0295.

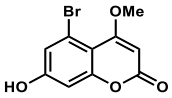


5,7-dihydroxy-4-methoxy-*2H*-chromen-2-one (3.12). To a solution of 4,5,7-trihydroxy-*2H*-chromen-2-one (**3.54**, 67.6 mg, 0.35 mmol, 1.00 equiv) in MeOH (3.0 mL) was added H_2SO_4 (0.35 mL). The reaction was heated at 80 °C for 1.5 h. Incubation at 0 °C induced the precipitation of a tan solid which was isolated by vacuum filtration to afford 19.0 mg of the title compound (26% yield). **MP** = 224.2–226.5 °C; $\mathbf{R_f}$ = 0.42 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 6.20 (d, J = 2.1, 1H), 6.17 (d, J = 2.5, 1H), 5.49 (s, 1H), 3.90 (s, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 168.8, 161.8, 161.6, 156.7, 156.0, 99.5, 96.5, 94.6, 85.2, 56.7; **IR** (thin film, cm^{-1}) 3402, 3234, 3089, 1700, 1642, 1609, 1579, 1521, 1464, 1443, 1405; **HRMS** (ESI) *m/z* calculated for $C_{10}H_9O_5^+$ $[M+H]^+$ 209.0444, found 209.0452.

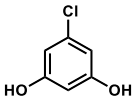


5-bromoresorcinol (3.55). To a solution of 1-bromo-3,5-dimethoxybenzene (2.00 g, 8.33 mmol, 1.00 equiv) in CH_2Cl_2 (8.3 mL) was added 1 M BBr_3 solution in CH_2Cl_2 (25.0 mL, 25.0 mmol, 3.00 equiv) at –78 °C. The reaction was warmed to room temperature and stirred for 16 h. The reaction was quenched by slow addition of water (10 mL) at 0 °C and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford an oil. The crude product was purified by flash column

chromatography (9:1 to 7:3 hexanes/ethyl acetate (v/v)) to afford 1.846 g of the title compound (85% yield) as an off-white solid. **^{1}H NMR** (400 MHz, CD_3OD) δ 6.42 (d, *J* = 2.1, 2H), 6.19 (t, *J* = 2.1, 1H). All spectra obtained were constant with literature values.[68]

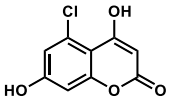


5-bromo-7-hydroxy-4-methoxy-*2H*-chromen-2-one (3.13). To a mixture of 5-bromoresorcinol (**3.55**, 422.0 mg, 2.232 mmol, 1.000 equiv), malonic acid (232.3 mg, 2.232 mmol, 1.000 equiv) and $ZnCl_2$ flame-dried under vacuum (943.0 g, 6.919 mmol, 3.100 equiv) was added $POCl_3$ (5.6 mL). The reaction was heated at 60 °C and stirred for 15.1 h. The reaction was quenched by pouring into ice water and induced the precipitation of a solid, which was isolated by vacuum filtration. The crude material was used without further purification due to decomposition under flash column chromatography conditions. To a solution of crude 5-bromo-4,7-dihydroxy-*2H*-chromen-2-one (459.0 mg, 1.785 mmol, 1.000 equiv) in MeOH (20.0 mL) was added H_2SO_4 (2.0 mL). The reaction was heated at 75 °C for 7.1 h. Partial concentration of the solution and incubation at 0 °C induced the precipitation of a white solid which was isolated by vacuum filtration to afford 34.3 mg of the title compound (13% yield). **MP** = 207.4–212.0 °C; $\mathbf{R_f}$ = 0.41 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 7.04 (d, *J* = 2.4, 1H), 6.71 (d, *J* = 2.4, 1H), 5.71 (s, 1H), 3.92 (s, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 166.4, 160.9, 160.7, 156.2, 119.7, 117.6, 106.2, 102.7, 87.9, 56.6; **IR** (thin film, cm^{-1}) 3406, 3147, 1694, 1611, 1589, 1550, 1433; **HRMS** (ESI) *m/z* calculated for $C_{10}H_8BrO_4^+$ $[M+H]^+$ 270.9600, found 270.9607.

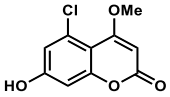


5-chlororesorcinol (3.56). To a solution of 1-chloro-3,5-dimethoxybenzene (1.00 g, 5.79 mmol, 1.00 equiv) dissolved in CH_2Cl_2 (6.0 mL) was added a 1-M solution of BBr_3 in CH_2Cl_2 (12.2 mL, 12.1 mmol, 2.10 equiv) at –78 °C. The reaction was allowed to warm to room temperature and stirred for 14 h. The reaction was quenched with the addition of water (4.0 mL). The aqueous phase was extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford an oil. The crude residue was purified by flash column chromatography (4:1 hexanes/ethyl acetate (v/v)) to afford 661 mg

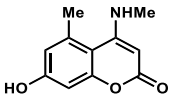
of the title compound (79% yield) as a pink solid. **^{1}H NMR** (400 MHz, acetone-d_6) δ 8.63 (s, 2H), 6.37 (d, J = 2.1, 2H), 6.30 (t, J = 2.1, 1H). All spectra obtained were constant with literature values.[69]



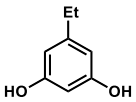
5-chloro-4,7-dihydroxy-*2H*-chromen-2-one (3.57). To a mixture of 5-chlororesorcinol (**3.56**, 300.0 mg, 2.075 mmol, 1.000 equiv), malonic acid (215.9 mg, 2.075 mmol, 1.000 equiv) and $ZnCl_2$ flame-dried under vacuum (876.7 mg, 6.432 mmol, 3.100 equiv) was added $POCl_3$ (5.2 mL). The reaction was heated at 60 °C and stirred for 14.5 h. The reaction was quenched by pouring into ice water and induced the precipitation of a solid, which was isolated by vacuum filtration. The crude solid was purified by flash column chromatography (8:3:1 toluene/ethyl acetate/formic acid (v/v)) to afford a 108.3 mg of the title compound (24% yield) as a tan solid with minor impurities. **MP** = 250.1–252.9 °C; **R_f** = 0.45 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, acetone-d_6) δ 6.88 (d, J = 2.4, 1H), 6.70 (d, J = 2.4, 1H), 5.51 (s, 1H); **^{13}C NMR** (150 MHz, acetone-d_6) δ 166.9, 161.6, 161.2, 158.5, 132.8, 116.5, 107.0, 103.2, 91.0; **IR** (thin film, cm^{-1}) 3254, 3080, 2940, 1617, 1547, 1436; **HRMS** (ESI) *m/z* calculated for $C_9H_6ClO_4^+$ $[M+H]^+$ 212.9949, found 212.9953.



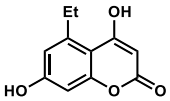
5-chloro-7-hydroxy-4-methoxy-*2H*-chromen-2-one (3.14). To a solution of 5-chloro-4,7-dihydroxy-*2H*-chromen-2-one (**3.57**, 88.0 mg, 0.41 mmol, 1.00 equiv) in MeOH (4.6 mL) was added H_2SO_4 (0.46 mL). The reaction was heated at 75 °C for 3.7 h. Incubation at 0 °C induced the precipitation of a white solid which was isolated by vacuum filtration to afford 56.9 mg of the title compound (61% yield). **MP** = >310 °C; **R_f** = 0.49 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 10.98 (br s, 1H), 6.82 (d, J = 2.3, 1H), 6.68 (d, J = 2.3, 1H), 5.71 (s, 1H), 3.93 (s, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 166.8, 160.9, 160.5, 156.2, 130.6, 115.9, 105.1, 102.2, 88.0, 56.8; **IR** (thin film, cm^{-1}) 3139, 2728, 2161, 2033, 1698, 1613, 1593,1522, 1449, 1433; **HRMS** (ESI) *m/z* calculated for $C_{10}H_8ClO_4^+$ $[M+H]^+$ 227.0106, found 227.0112.



7-hydroxy-5-methyl-4-(methylamino)-*2H*-chromen-2-one (3.15). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**, 423 mg, 2.2 mmol, 1.0 equiv) in MeCN (8.8 mL) was added benzyl triethylammonium chloride (2.0 g, 8.9 mmol, 4.0 equiv). The solution was heated at 40 °C for 10 min then $POCl_3$ (910 μL, 9.7 mmol, 4.4 equiv) was added. The reaction was heated at 90 °C for 1.0 h. The reaction was quenched by pouring into ice water and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford an oil, which was carried forward without further purification and dissolved in EtOH (12.7 mL) in a pressure vessel. To the crude solution was added triethylamine (4.20 mL) and methylamine hydrochloride (1.49 g, 22.1 mmol, 10.0 equiv). The reaction was heated at 80 °C for 14.5 h. The reaction was quenched by pouring into 1 M HCl and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford 88.7 mg of the title compound (20% yield) as an orange solid. **MP** 279.3-283.6 °C; $\mathbf{R}_f$ = 0.17 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, CD_3OD) δ 8.46 (s, 1H), 6.59 (d, *J* = 2.5, 1H), 6.54 (d, *J* = 2.6, 1H), 5.08 (s, 1H), 2.95 (s, 3H), 2.71 (s, 3H); **^{13}C NMR** (150 MHz, CD_3OD) δ 166.6, 161.5, 160.2, 157.9, 137.9, 117.8, 107.5, 102.5, 80.5, 30.5, 23.4; **HRMS** (ESI) *m/z* calculated for $C_{11}H_{12}NO_3^+$ $[M+H]^+$ 206.0812, found 206.0820.

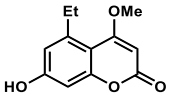


5-ethylbenzene-1,3-diol (3.58). The title compound was prepared from 1-(3,5-dihydroxyphenyl)ethan-1-one according to the procedure described by Linusson et al.[70] **^{1}H NMR** (400 MHz, acetone-d_6) δ 8.01 (s, 2H), 6.19 (d, *J* = 1.9, 2H), 6.17 (d, *J* = 2.1, 1H), 2.46 (q, *J* = 7.6, 2H), 1.14 (t, *J* = 7.6, 3H). All spectra obtained were consistent with literature values.[70]

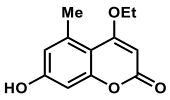


5-ethyl-4,7-dihydroxy-*2H*-chromen-2-one (3.59). To a mixture of 5-ethylbenzene-1,3-diol (**3.58**, 800.2 mg, 5.791 mmol, 1.000 equiv), malonic acid (602.6 mg, 5.791 mmol, 1.000 equiv)

and $ZnCl_2$ flame-dried under vacuum (2.446 g, 17.95 mmol, 3.100 equiv) was added $POCl_3$ (14.5 mL). The reaction was heated at 60 °C and stirred for 14.5 h. The reaction was quenched by pouring into ice water and induced the precipitation of a solid, which was isolated by vacuum filtration. The crude solid was purified by flash column chromatography (8:3:1 to 6:5:1 toluene/ethyl acetate/formic acid (v/v)) to afford a 183.5 mg of the title compound (15% yield) as a tan solid with minor impurities. **MP** = 234.1–236.8 °C; $\mathbf{R_f}$ = 0.43 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 12.11 (s, 1H), 10.41 (s, 1H), 6.57 (d, *J* = 2.4, 1H), 6.51 (d, *J* = 2.4, 1H), 5.36 (s, 1H), 2.99 (q, *J* = 7.4, 2H), 1.15 (t, *J* = 7.4, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 168.7, 161.8, 160.6, 157.2, 145.2, 114.4, 105.7, 100.6, 88.3, 28.2, 16.5; **IR** (thin film, cm^{-1}) 3086, 2971, 2599, 1646, 1598, 1555, 1510, 1437; **HRMS** (ESI) *m/z* calculated for $C_{11}H_{11}O_4^+$ $[M+H]^+$ 207.0652, found 207.0651.

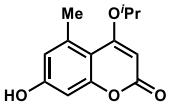


5-ethyl-7-hydroxy-4-methoxy-*2H*-chromen-2-one (3.16). To a solution of 5-ethyl-4,7-dihydroxy-*2H*-chromen-2-one (**3.59**, 141.7 mg, 0.691 mmol, 1.000 equiv) in MeOH (7.7 mL) was added H_2SO_4 (0.77 mL). The reaction was heated at 75 °C for 4.4 h. Incubation at 0 °C induced the precipitation of an orange solid which was isolated by vacuum filtration to afford 78.8 mg of the title compound (52% yield) as a tan solid with minor impurities. **MP** = 267.4–269.5 °C; $\mathbf{R_f}$ = 0.48 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 10.45 (br s, 1H), 6.58 (d, *J* = 2.3, 1H), 6.54 (d, *J* = 2.3, 1H), 5.60 (s, 1H), 3.93 (s, 3H), 2.91 (q, *J* = 7.3, 2H), 1.12 (t, *J* = 7.4, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 169.0, 161.7, 160.5, 156.2, 144.5, 114.7, 105.3, 100.6, 86.9, 56.5, 28.6, 16.1; **IR** (thin film, cm^{-1}) 3286, 1689, 1621, 1606, 1560, 1421; **HRMS** (ESI) *m/z* calculated for $C_{12}H_{13}O_4^+$ $[M+H]^+$ 221.0808, found 221.0818.

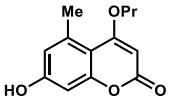


4-ethoxy-7-hydroxy-5-methyl-*2H*-chromen-2-one (3.17). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**, 510.0 mg, 2.650 mmol, 1.000 equiv) in EtOH (29.6 mL) was added H_2SO_4 (2.9 mL). The reaction was heated at 75 °C for 9.5 h. Incubation at 0 °C induced the precipitation of a pale-yellow solid which was isolated by vacuum filtration to afford 237.6 mg of the

title compound (41% yield). **MP** 276.5-278.2 °C; $\mathbf{R_f}$ = 0.58 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, DMSO-d_6) δ 10.40 (br s, 1H), 6.54 (d, J = 2.4, 1H), 6.51 (d, J = 2.4, 1H), 5.50 (s, 1H), 4.12 (q, J = 7.0, 2H), 2.51 (s, 3H), 1.40 (t, J = 7.0, 3H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 168.3, 161.8, 160.2, 156.1, 138.2, 116.0, 106.0, 100.5, 86.8, 65.2, 23.1, 13.9; **IR** (thin film, cm^{-1}) 3189, 1697, 1612, 1558, 1459, 1403; **HRMS** (ESI) *m/z* calculated for $C_{12}H_{13}O_4^+$ $[M+H]^+$ 221.0808, found 221.0813.

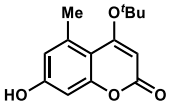


7-hydroxy-4-isopropoxy-5-methyl-*2H*-chromen-2-one (3.18). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**, 550.0 mg, 2.862 mmol, 1.000 equiv) in *i*-PrOH (32.0 mL) was added H_2SO_4 (3.2 mL). The reaction was heated at 75 °C for 4.7 h. The solution was partially concentrated under reduced pressure, diluted with H_2O and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford a solid. The crude solid was purified by flash column chromatography (9:2:1 to 8:3:1 toluene/ethyl acetate/formic acid (v/v)) to afford 51.8 mg of the title compound (8% yield) as a tan solid. **MP** = 215.1–217.1 °C; $\mathbf{R_f}$ = 0.51 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, acetone-d_6) δ 6.61 (d, J = 2.2, 1H), 6.56 (d, J = 2.5, 1H), 5.48 (s, 3H), 4.84 (heptd, J = 6.0, 2.7, 1H), 2.61 (s, 3H), 1.46 (d, J = 6.0, 6H); **^{13}C NMR** (150 MHz, acetone-d_6) δ 168.2, 162.7, 160.9, 157.8, 139.8, 116.7, 108.2, 101.5, 88.2, 73.0, 24.0, 21.8; **IR** (thin film, cm^{-1}) 3207, 2980, 2933, 1683, 1599, 1556, 1452; **HRMS** (ESI) *m/z* calculated for $C_{13}H_{15}O_4^+$ $[M+H]^+$ 235.0965, found 235.0973.

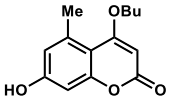


7-hydroxy-5-methyl-4-propoxy-*2H*-chromen-2-one (3.19). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**, 500.0 mg, 2.602 mmol, 1.000 equiv) in *i*-PrOH (29.0 mL) was added H_2SO_4 (2.9 mL). The reaction was heated at 75 °C for 8.6 h. The solution was partially concentrated under reduced pressure, diluted with H_2O, and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated

under reduced pressure to afford a solid. The crude solid was purified by flash column chromatography (8:3:1 toluene/ethyl acetate/formic acid (v/v)) and subjected to trituration (3x, 8:3:1 toluene/ethyl acetate/formic acid (v/v)) to afford 167.1 mg of the title compound (27% yield) as a tan solid. **MP** = 198.2–199.5 °C; $\mathbf{R_f}$ = 0.50 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, acetone-d_6) δ 9.26 (s, 1H), 6.63 (dd, J = 2.6, 1.0, 1H), 6.57 (d, J = 2.5, 1H), 5.48 (s, 1H), 4.14 (t, J = 6.3, 2H), 2.63 (s, 3H), 1.98-1.87 (m, 2H), 1.10 (t, J = 7.4, 3H); **^{13}C NMR** (150 MHz, acetone-d_6) δ 169.7, 162.6, 161.1, 157.8, 139.7, 116.8, 108.0, 101.7, 88.2, 72.1, 23.9, 22.8, 11.1; **IR** (thin film, cm^{-1}) 3148, 3090, 2974, 2935, 1709, 1697, 1689, 1625, 1451; **HRMS** (ESI) *m/z* calculated for $C_{13}H_{15}O_4^+$ $[M+H]^+$ 235.0965, found 235.0960.

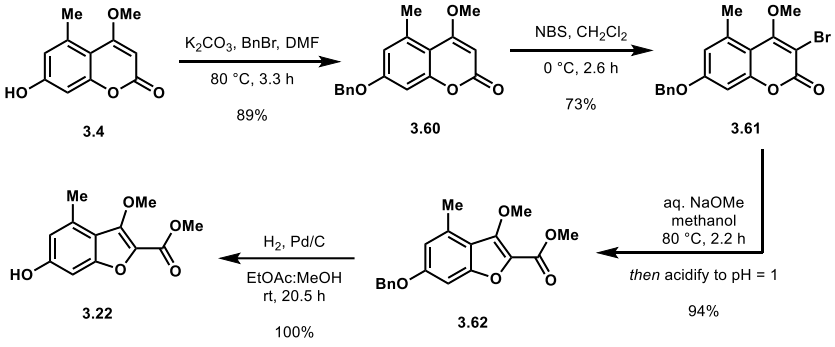


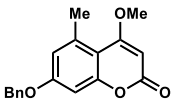
4-(*tert*-butoxy)-7-hydroxy-5-methyl-2*H*-chromen-2-one (3.20). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen- 2-one (**3.53**, 25.0 mg, 0.121 mmol) in *t*-butanol (1.20 mL, 0.10 M) was added H_2SO_4 (0.120 mL, 1.0 M) in a 10 mL cuvette. The reaction was heated at 120 °C for 1 h in a microwave reactor, rapidly forming a biphasic mixture. The reaction was quenched by the addition of water and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, and evaporated under reduced pressure to afford a yellow solid. The crude solid was purified over silica gel by flash chromatography (2–20% ethyl acetate in dichloromethane), to afford 8.9 mg (28% yield) of the title compound as a glassy yellow film. **^{1}H NMR** (600 MHz, acetone-d_6) δ 6.60 (d, J = 2.5 Hz, 1H), 6.48 (d, J = 2.4 Hz, 1H), 6.16 (s, 1H), 2.53 (s, 3H), 1.01 (s, 9H). **^{13}C NMR** (150 MHz, acetone-d_6) δ 191.5, 167.4, 159.6, 159.3, 143.8, 116.1, 108.3, 102.0, 100.6, 67.0, 28.7, 21.9. **HRMS** (ESI) *m/z* calculated for $C_{14}H_{17}O_4^+$ $[M+H]^+$ 249.1121, found 249.1123.



4-butoxy-7-hydroxy-5-methyl-*2H*-chromen-2-one (3.21). To a solution of 4,7-dihydroxy-5-methyl-*2H*-chromen-2-one (**3.53**, 400.0 mg, 2.081 mmol, 1.000 equiv) in *n*-BuOH (23.3 mL) was added H_2SO_4 (2.3 mL). The reaction was heated at 75 °C for 5.1 h. The solution was partially concentrated under reduced pressure, diluted with H_2O, and extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford a solid. The crude solid was purified by flash column chromatography (9:2:1 to 8:3:1 toluene/ethyl acetate/formic acid (v/v)) and subjected to trituration (3x, 8:3:1 toluene/ethyl acetate/formic acid (v/v)) to afford 105.2 mg of the title compound (21% yield) as a white solid. **MP** = 196.7–197.8 °C; $\mathbf{R_f}$ = 0.54 (8:3:1 toluene/ethyl acetate/formic acid (v/v)); **^{1}H NMR** (600 MHz, CD_3OD) δ 6.58 (d, *J* = 2.2, 1H), 6.54 (d, *J* = 2.2, 1H), 5.55 (s, 1H), 4.15 (t, *J* = 6.3, 2H), 2.61 (s, 3H), 1.92-1.84 (m, 2H), 1.57 (q, *J* = 7.5, 2H), 1.02 (t, *J* = 7.4); **^{13}C NMR** (150 MHz, CD_3OD) δ 171.5, 166.1, 162.2, 158.0. 140.2, 117.5, 108.0, 101.7, 87.5, 70.9, 31.8, 23.9, 20.5, 14.1; **IR** (thin film, cm^{-1}) 3279, 3122, 2957, 2872, 1675, 1610, 1555, 1448, 1409; **HRMS** (ESI) *m/z* calculated for $C_{14}H_{17}O_4^+$ $[M+H]^+$ 249.1121, found 249.1124.

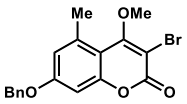
Scheme 3.2. Synthesis of benzofuran 3.22.



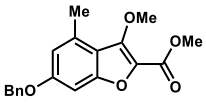


7-(benzyloxy)-4-methoxy-5-methyl-*2H*-chromen-2-one (3.60). To a solution of 7-hydroxy-4-methoxy-5-methyl-*2H*-chromen-2-one (**3.4**; 1.07 g, 5.19 mmol, 1.00 equiv) in DMF (52.0 mL) was added K_2CO_3 (1.07 g, 7.78 mmol, 1.50 equiv) and BnBr (740 µL, 6.23 mmol, 1.20 equiv).

The reaction was heated at 80 °C and stirred for 3.3 h. The reaction was quenched with MeOH (1 mL). Water was added and the aqueous phase was extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure afford 1.37 g of the title compound (89% yield) as a white solid. **MP** = 137.5–138.7 °C; $\mathbf{R_f}$ = 0.17 (2:1 hexanes/ethyl acetate (v/v)); **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.41-7.34 (m, 4H), 7.33-7.29 (m, 1H), 6.65 (d, *J* = 2.6, 1H), 6.63 (d, *J* = 2.5, 1H), 5.47 (s, 1H), 5.03 (s, 2H), 3.86 (s, 3H), 2.55 (s, 3H); **^{13}C NMR** (150 MHz, $CDCl_3$) δ 169.6, 163.0, 160.9, 156.5, 138.5, 135.9, 128.7, 128.2, 127.5, 116.1, 108.0, 99.6, 87.5, 70.1, 55.9, 23.4; **IR** (thin film, cm^{-1}) 3089, 3063, 3032, 2975, 2935, 1711, 1601, 1555, 1497, 1455, 1423; **HRMS** (ESI) *m/z* calculated for $C_{18}H_{17}O_4^+$ $[M+H]^+$ 297.1121, found 297.1122.

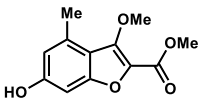


7-(benzyloxy)-3-bromo-4-methoxy-5-methyl-*2H*-chromen-2-one (3.61). To a solution of 7-(benzyloxy)-4-methoxy-5-methyl-*2H*-chromen-2-one (**3.60**, 600.0 mg, 2.02 mmol, 1.00 equiv) in CH_2Cl_2 (25.0 mL) was added recrystallized *N*-bromosuccinimide (378 mg, 2.13 mmol, 1.05 equiv) at 0 °C. The reaction was stirred for 2.6 h. The reaction was quenched with saturated $Na_2S_2O_3$ (5.0 mL). The aqueous phase was extracted with CH_2Cl_2 (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure afford 555 mg of the title compound (73% yield) as a white solid. **MP** 118.3–121.5 °C; $\mathbf{R_f}$ = 0.53 (2:1 hexanes/ethyl acetate (v/v)); **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.42-7.33 (m, 5H), 6.76 (d, 2H), 5.11 (s, 2H), 4.09 (s, 3H), 2.65 (s, 3H); **^{13}C NMR** (150 MHz, $CDCl_3$) δ 168.5, 161.4, 159.6, 155.4, 137.9, 135.7, 128.9, 128.5, 127.6, 117.2, 110.2, 99.9, 97.4, 70.5, 61.0, 22.6; **IR** (thin film, cm^{-1}) 3089, 3063, 3032, 2972, 2939, 1715, 1612, 1588, 1538, 1497, 1453; **HRMS** (ESI) *m/z* calculated for $C_{18}H_{16}BrO_4^+$ $[M+H]^+$ 375.0226, found 375.0218.

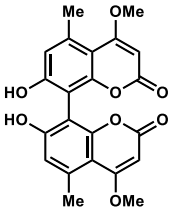


Methyl 6-(benzyloxy)-3-methoxy-4-methylbenzofuran-2-carboxylate (3.62). To a suspension of 7-(benzyloxy)-3-bromo-4-methoxy-5-methyl-*2H*-chromen-2-one (**3.61**; 418 mg, 1.11 mmol, 1.00 equiv) in MeOH (44.6 mL) was added freshly prepared 1M solution of NaOMe in MeOH

(11.1 mL, 11.1 mmol, 10.0 equiv). The reaction was heated at 80 °C and stirred for 2.2 h. The reaction was poured into 122 mL of 0.1 M HCl. The acidifed aqueous phase was extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure to afford 342 mg of the title compound (94% yield) as an off-white solid. **MP** = 57.9–58.8 °C; $\mathbf{R_f}$ = 0.62 (2:1 hexanes/ethyl acetate (v/v)); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.45-7.30 (m, 5H), 6.81 (d, *J* = 2.1, 1H), 6.74 (d, *J* = 2.1, 1H), 5.08 (s, 2H), 4.12 (s, 3H), 3.96 (s, 3H), 2.59 (s, 3H); **^{13}C NMR** (100 MHz, $CDCl_3$) δ 160.3, 159.6, 154.8, 152.4, 136.5, 134.5, 131.2, 128.8, 128.2, 127.5, 115.3, 94.4, 70.4, 63.1, 52.0, 18.2; **IR** (thin film, cm^{-1}) 3033, 2950, 2851, 1708, 1621, 1598, 1573, 1498, 1446; **HRMS** (ESI) *m/z* calculated for $C_{19}H_9O_5^+$ $[M+H]^+$ 327.1227, found 327.1231.

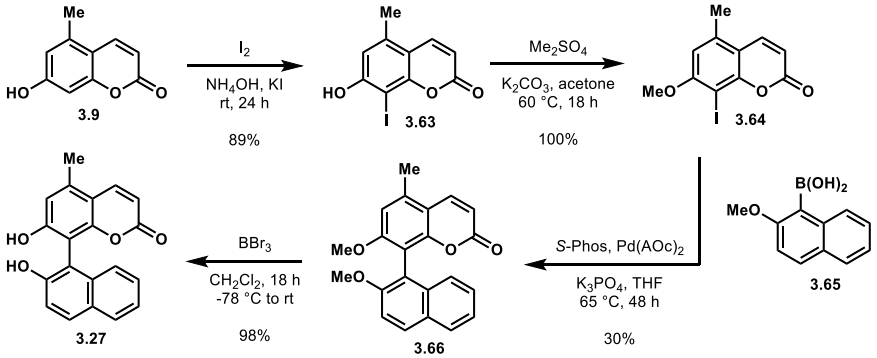


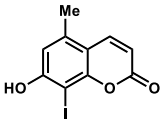
Methyl 6-hydroxy-3-methoxy-4-methylbenzofuran-2-carboxylate (3.22). A solution of methyl 6-(benzyloxy)-3-methoxy-4-methylbenzofuran-2-carboxylate (**3.62**; 502 mg, 0.923 mmol, 1.00 equiv) in MeOH (30.0 mL) was sparged with N_2 for 15 min then 10% Pd/C (30 mg, 10% (w/w)) as a slurry in ethyl acetate (1.0 mL). The reaction was sparged with N_2 for 10 min then sparged with H_2 for 10 min and stirred for 20.5 h under H_2 balloon atmosphere. The reaction was filtered through a plug of celite, rinsed with ethyl acetate, and the organic solution evaporated under reduced pressure. The crude solid was purified through a silica plug with flash chromatography (ethyl acetate) to afford 220 mg of the title compound (quantitative yield) as a light pink solid. **MP** 137.6–140.2 °C; $\mathbf{R_f}$ = 0.45 (2:1 hexanes/ethyl acetate (v/v)); **^{1}H NMR** (600 MHz, CD_3OD) δ 6.58 (d, *J* = 1.0, 1H), 6.54 (d, *J* = 1.0, 1H), 4.04 (s, 3H), 3.89 (s, 3H), 2.51 (s, 3H); **^{13}C NMR** (150 MHz, CD_3OD) δ 161.2, 160.6, 156.5, 153.6, 135.4, 131.8, 115.7, 115.1, 96.0, 63.5, 52.1, 18.1; **IR** (thin film, cm^{-1}) 3365, 2952, 2852, 1682, 1619, 1600, 1575, 1448; **HRMS** (ESI) *m/z* calculated for $C_{12}H_{13}O_5^+$ $[M+H]^+$ 237.0757, found 237.0762.



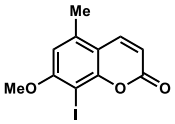
7,7′-dihydroxy-4,4′-dimethoxy-5,5′dimethyl-*2H*,*2H*′-[8,8′-bichromene]-2,2′-dione (3.6). To a solution of 7-hydroxy-5-methyl-4-propoxy-*2H*-chromen-2-one (**3.4**; 51.5 mg, 0.250 mmol, 1.00 equiv) in CH_2Cl_2 (1.0 mL) and trifluoroacetic acid (0.25 mL) was added VOF_3 (124 mg, 0.375 mmol, 1.50 equiv) at 0 °C. The reaction was stirred for 4.0 h. The reaction was quenched with H_2O (2 mL). The aqueous phase was extracted with ethyl acetate (3x). The organic layers were combined, washed with brine, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude solid was purified by flash column chromatography (8:3:1 toluene/ethyl acetate/formic acid (v/v)) then reverse-phase HPLC (Phenomenex Kinetex 5 μm C18, 150 x 21.2 mm column) under the following conditions: mobile phase A = 0.1% formic acid in water and B = 0.1% formic acid in acetonitrile; method = 75% A for 5.0 min, 75% A to 70% A over 2.0 min, 70% A for 7.0 min, 70% A to 65% A over 2.0 min, 65% A for 5.0 min, 65% A to 50% A over 3.0 min, 50% A for 1.0 min, 254 and 308 nm UV detection and 14 mL/min flow rate. A retention time of 12.65–14.77 min afforded 7.0 mg of 8,8′-product **3.6** (14% yield). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.69 (s, 2H), 5.55 (s, 2H), 3.94 (s, 6H), 2.59 (s, 6H); **HRMS** (ESI) *m/z* calculated for $C_{22}H_{19}O_8^+$ $[M+H]^+$ 411.1074, found 411.1070. All spectra obtained were agreement with literature values.[71]

Scheme 3.3. Synthesis of 8,1′ cross-coupled product 3.27.

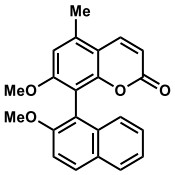




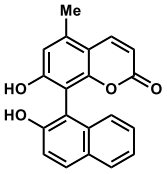
7-hydroxy-8-iodo-5-methyl-2*H*-chromen-2-one (3.63). Coumarin (**3.9**; 750 mg, 4.26 mmol, 1.00 equiv) was dissolved in 20% aqueous solution of NH_4OH (17.0 mL, 0.250 M) at room temperature in a round bottom flask open to air. I_2 (1.08 g, 4.26 mmol, 1.000 equiv) was dissolved in an 5.0% KI solution (8.50 mL, 0.125 M), and added dropwise to the reaction flask. The reaction was stirred at room temperature for 48 h, until complete by TLC. The reaction was acidified to a pH of approximately 2.0 with 6M H_2SO_4, where the product precipitated as an off-white solid. The solid was collected by vacuum filtration and dried under vacuum and carried forward without further purification, 1.15 g, 89% yield. **^{1}H NMR** (600 MHz, DMSO-d_6) δ 11.29 (s, 1H), 8.02 (d, J = 9.7 Hz, 1H), 6.75 (s, 1H), 6.23 (d, J = 9.6 Hz, 1H), 2.41 (s, 3H); **^{13}C NMR** (151 MHz, DMSO-d_6) δ 160.6, 160.2, 155.4, 141.8, 137.9, 113.2, 111.2, 111.1, 71.3, 18.0; **HRMS** (ESI) *m/z* calculated for $C_{10}H_8IO_3^+$ $[M+H]^+$ 302.9513, found 302.9513.[72]



8-iodo-7-methoxy-5-methyl-2*H*-chromen-2-one (3.64). Iodocoumarin (**3.63**; 100 mg, 0.33 mmol, 1.0 equiv) and K_2CO_3 (183 mg, 1.32 mmol, 4.00 equiv) were added to a dry 1 dram vial. Dry acetone (0.7 mL, 0.5 M) was added at room temperature followed by Me_2SO_4 (79 µL, 0.83 mmol, 2.5 equiv). The reaction was heated to reflux and allowed to stir overnight 16 h. The reaction was cooled to rt, and quenched by stirring with dilute HCl (1 mL) for 10 min at rt. The reaction was then extracted with ethyl acetate (20 mL x 3), the extracts were combined, washed with brine, dried over Na_2SO_4, and concentrated to an off-white solid (104 mg, quantitative yield). The product was carried forward to the next step without further purification. **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.76 (d, J = 9.7 Hz, 1H), 6.65 (s, 1H), 6.24 (d, J = 9.7 Hz, 1H), 3.97 (s, 3H), 2.51 (s, 3H); **^{13}C NMR** (150 MHz, $CDCl_3$) δ 161.2, 160.7, 155.4, 140.3, 137.9, 113.3, 112.8, 109.3, 73.0, 57.0, 19.0; **HRMS** (ESI) *m/z* calculated for $C_{11}H_{10}IO_3^+$ $[M+H]^+$ 316.9669, found 316.9671.



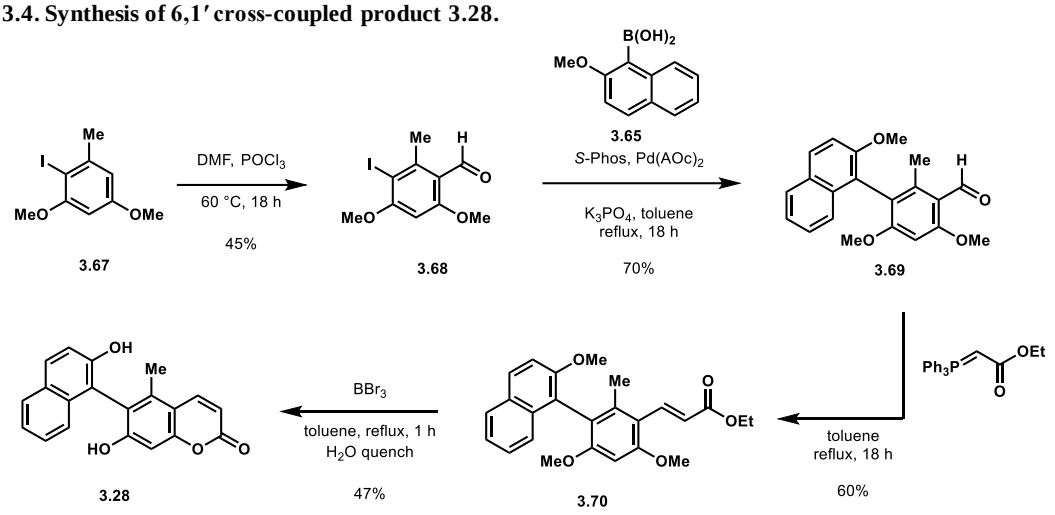
7-methoxy-8-(2-methoxynaphthalen-1-yl)-5-methyl-2*H*-chromen-2-one (3.66). Coumarin (**3.64**, 50.0 mg, 0.166 mmol, 1.00 equiv), commercial 2-methoxy-naphthalene-1-boronic acid (**3.65**; 66.9 mg, 0.331 mmol, 2.00 equiv), K_3PO_4 (105.4 mg, 0.497 mmol, 3.00 equiv), and approximately 5 mg activated 4 Å mol sieves were added to a dry round bottom flask. A stock solution of $Pd(OAc)_2$ (1.9 mg, 8.3 µmol, 5 mol %), and *S*-Phos (6.8 mg, 17 mmol, 10 mol %) in dry toluene (2.0 mL, 0.083 M) was prepared and stirred for 20 min at room temperature before transferring to the reaction vessel and heating to 110 °C. The reaction was monitored by TLC over 72 h. The reaction was cooled to rt and passed through a silica plug with 1:1 ethyl acetate: hexanes as eluent, then concentrated to afford a golden solid. The crude material was purified by column chromatography using 0–30% ethyl acetate in hexanes. The desired product was recovered as 17 mg of a white glassy solid in a 30% yield. **^{1}H NMR** (800 MHz, acetone-d_6) δ 8.15 (d, *J* = 9.7 Hz, 1H), 8.00 (d, *J* = 9.0 Hz, 1H), 7.88 (d, *J* = 8.6 Hz, 0H), 7.53 (d, *J* = 9.0 Hz, 1H), 7.31 (qt, *J* = 6.6, 3.3 Hz, 2H), 7.23 (d, *J* = 8.3 Hz, 0H), 7.12 (s, 1H), 6.17 (d, *J* = 9.8 Hz, 1H), 3.83 (s, 3H), 3.75 (s, 3H), 2.68 (s, 3H); **^{13}C NMR** (200 MHz, acetone-d_6) δ 160.4, 160.0, 155.2, 153.9, 141.0, 137.6, 133.6, 129.7, 129.3, 128.0, 126.2, 124.5, 123.3, 115.5, 114.0, 112.0, 111.9, 110.8, 109.6, 56.0, 55.6, 18.0; **HRMS** (ESI) *m/z* calculated for $C_{22}H_{19}O_4^+$ $[M+H]^+$ 347.1278, found 347.1261.

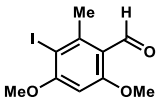


7-hydroxy-8-(2-hydroxynaphthalen-1-yl)-5-methyl-2*H*-chromen-2-one (3.27). Biaryl (**3.66**, 17 mg, 49 µmol, 1.0 equiv) was dissolved in dry CH_2Cl_2 (0.5 mL, 0.1 M) in a dry round bottom flask and cooled to –78 °C. A solution of 1 M BBr_3 in CH_2Cl_2 (200 µL, 200 µmol, 4.0 equiv) was added slowly, and the reaction was allowed to warm to room temperature and stir overnight. The reaction was quenched by the addition of water and pH adjusted to approximately 4 with aqueous saturated sodium bicarbonate, then extracted with ethyl acetate (10 mL x 3). The extracts were combined, washed with brine, dried over Na_2SO_4, and concentrated to an off-white solid as 15.3

mg of the target compound was isolated without further purification in a 98% yield. **^{1}H NMR** (600 MHz, CD_3OD) δ 8.17 (d, *J* = 9.6 Hz, 1H), 7.80 (t, *J* = 9.1 Hz, 2H), 7.26 (ddd, *J* = 6.9, 4.5, 1.8 Hz, 2H), 7.22 (d, *J* = 8.8 Hz, 1H), 7.21–7.16 (m, 1H), 6.88 (s, 1H), 6.18 (d, *J* = 9.6 Hz, 1H), 2.59 (s, 3H); **^{13}C NMR** (150 MHz, CD_3OD) δ 162.9, 159.4, 154.6, 152.8, 142.2, 13735, 134.0, 129.3, 128.8, 127.6, 125.8, 123.7, 122.3, 117.7, 114.2, 111.3, 111.2, 109.9, 108.9, 17.1; **HRMS** (ESI) *m/z* calculated for $C_{20}H_{15}O_4^+$ $[M+H]^+$ 319.0965, found 319.0970.

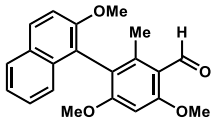
Scheme 3.4. Synthesis of 6,1′ cross-coupled product 3.28.



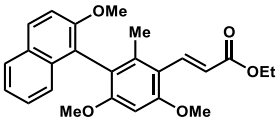


3-iodo-4,6-dimethoxy-2-methylbenzaldehyde (3.68). To an oven-dried 50 mL round-bottomed flask charged with a stir bar was added dry dimethylformamide (5 mL) under dry nitrogen atmosphere. The reaction mixture was cooled to 0 °C using an ice bath and $POCl_3$ (0.94 mL, 10.1 mmol, 2.00 equiv) was added dropwise. The ice bath was removed, and the resultant mixture was stirred at room temperature for 1 h. Subsequently, a solution of 2-iodo-1,3-dimethoxy-5-methylbenzene (**3.67**, synthesized via iodination of 3,5-dimethoxytoluene using *N*-iodosuccinimide; 1.40 g, 5.03 mmol, 1.00 equiv) in dry dimethylformamide (5 mL) was added to the reaction flask and the resultant mixture was stirred at 60 °C for 18 h. Afterwards, the reaction mixture was quenched with the addition of ice, basified to pH 13 using 10 M KOH and stirred at room temperature for 30 min. The reaction mixture was subsequently acidified to pH 2 using 5 M HCl and extracted with ethyl acetate (2 x 20 mL). The combined organic fractions were washed with water (20 mL), brine (20

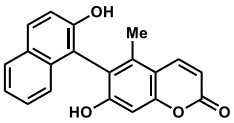
mL), dried over anhydrous Na_2SO_4 and concentrated in vacuum. Flash chromatography over silica gel (gradient of 5–60% ethyl acetate in hexanes) to afford the title compound (692 mg, 45% yield) as a yellow solid. TLC analysis (30% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, $CDCl_3$) δ 10.34 (1H, s), 6.32 (1H, s), 3.95 (3H, s), 3.92 (3H, s), 2.78 (3H, s) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 190.2, 165.3, 162.4, 146.3, 118.4, 92.6, 87.5, 56.6, 55.9, 26.2 ppm; **HRMS (ESI)** $(M+H)^+$ calculated for $C_9H_{12}IO_2^+$ = 306.9826, found = 306.9795 *m/z*.



4,6-dimethoxy-3-(2-methoxynaphthalen-1-yl)-2-methylbenzaldehyde (3.69). To an oven-dried 50 mL round bottomed charged with a stir bar was added aryl iodide **3.68** (200 mg, 0.65 mmol, 1.0 equiv), 2-methoxy-naphthalene-1-boronic acid (**3.65**; 198 mg, 0.980 mmol, 1.50 equiv), K_3PO_4 (827 mg, 3.90 mmol, 6.00 equiv), $Pd(OAc)_2$ (29 mg, 0.13 mmol, 0.20 equiv) and *S*-Phos (107 mg, 0.260 mmol, 0.400 equiv). A reflux condenser and a Teflon septum were added and the resultant mixture was subjected to vacuum and then refilled with nitrogen. The degas cycle was carried out for a total of 5 times, following which dry toluene (13 mL) was added and the resultant mixture was stirred at room temperature for 30 min. The mixture was subsequently heated to reflux and was allowed to stir at 110 °C for a total of 18 h. Afterwards, the reaction mixture was cooled to room temperature and was quenched by the addition of saturated aqueous NH_4Cl (13 mL) and was extracted with ethyl acetate (2 x 20 mL). The combined organic layers were washed with brine (10 mL) and were concentrated under vacuum. Flash chromatography over silica gel (gradient of 5–70% ethyl acetate in hexanes) afforded the title compound (152 mg, 70% yield) as a waxy solid. TLC (35% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, $CDCl_3$) δ 10.66 (s, 1H), 7.90 (d, *J* = 9.0 Hz, 1H), 7.86–7.81 (m, 1H), 7.38 (d, *J* = 9.1 Hz, 1H), 7.32 (ddt, *J* = 9.7, 6.7, 3.3 Hz, 2H), 7.25–7.19 (m, 1H), 6.51 (s, 1H), 4.00 (s, 3H), 3.83 (s, 3H), 3.69 (s, 3H), 2.20 (s, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 191.1, 165.1, 162.9, 154.5, 143.4, 133.6, 129.3, 129.1, 128.1, 126.4, 124.5, 123.5, 119.5, 118.8, 117.3, 113.7, 92.5, 56.7, 55.8, 55.8, 17.7 ppm; **HRMS (ESI)** $(M+H)^+$ calculated for $C_{21}H_{21}O_4^+$ = 337.1434, found = 337.1436 m/z.



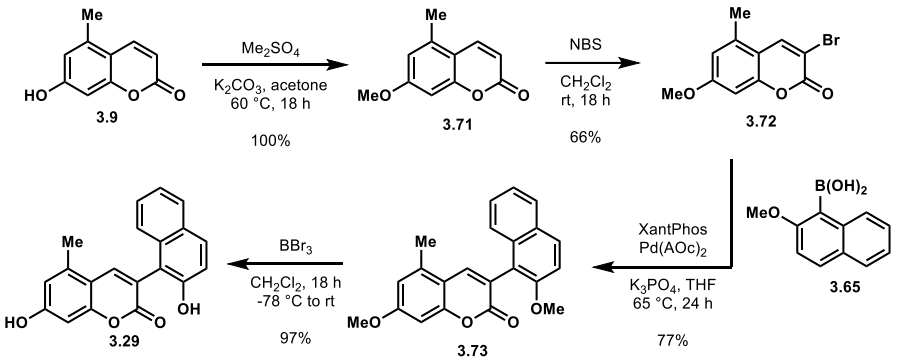
Ethyl (*E*)-3-(4,6-dimethoxy-3-(2-methoxynaphthalen-1-yl)-2-methylphenyl)acrylate (3.70). A mixture of aldehyde intermediate **3.69** (152 mg, 0.450 mmol) and Wittig reagent ethyl (triphenylphosphoranylidene)acetate (199 mg, 0.57 mmol, 1.20 equiv) were dissolved in toluene (4.5 mL) and the resultant mixture was refluxed under nitrogen for 18 h. Afterwards, the reaction mixture was concentrated under reduced pressure. Flash chromatography over silica gel (gradient of 5–50% ethyl acetate in hexanes) afforded the title compound (110 mg, 47% yield) as a white solid. TLC (25% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.00 (d, *J* = 16.1 Hz, 1H), 7.89 (d, *J* = 9.0 Hz, 1H), 7.85–7.80 (m, 1H), 7.38 (d, *J* = 8.9 Hz, 1H), 7.31 (m, 3H), 7.22 (dd, *J* = 7.8, 1.8 Hz, 1H), 6.67 (d, *J* = 16.1 Hz, 1H), 6.53 (s, 1H), 4.26 (q, J = 7.1 Hz, 2H), 3.99 (s, 3H), 3.84 (s, 3H), 3.66 (s, 3H), 2.00 (s, 3H), 1.33 (t, *J* = 7.1 Hz, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 168.6, 160.4, 159.5, 154.4, 140.8, 139.3, 133.7, 129.1, 127.9, 126.3, 124.7, 123.4, 120.5, 119.7, 117.9, 115.4, 113.8, 93.1, 60.1, 56.7, 55.8, 55.4, 17.6, 14.4 ppm; **HRMS (ESI)** $(M+H)^+$ calculated for $C_{25}H_{27}O_5^+$ = 407.1853, found = 407.1853 m/z.

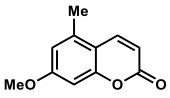


7-hydroxy-6-(2-hydroxynaphthalen-1-yl)-5-methyl-2H-chromen-2-one (3.28). This transformation was carried out with slight modifications of the original procedure[73] reported by Lavielle and coworkers as follows. To a solution of the intermediate **3.70** (110 mg, 0.270 mmol, 1.00 equiv) in dry toluene (2.7 mL) under an inert nitrogen atmosphere was added BBr_3 (0.20 mL, 2.2 mmol, 8.0 equiv) dropwise, and the resultant mixture was heated to reflux at 110 °C for 1 h. Afterwards, the reaction mixture was cooled to room temperature and subsequently to 0 °C using an ice bath and was quenched with the slow addition of water (5 mL; caution: significant exotherm noted). The quenched reaction mixture was stirred at room temperature for 30 min and was extracted with ethyl acetate (2 x 20 mL). The combined organic layers were washed with brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. Flash chromatography on silica gel (gradient of 5-80% ethyl acetate in hexanes) affords the title compound (40 mg, 47%) as a crystalline white solid: TLC (60% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.93 (m, 2H), 7.90–

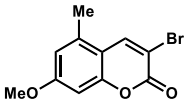
7.85 (m, 1H), 7.40 (m, 2H), 7.34 (d, *J* = 8.9 Hz, 1H), 7.19–7.13 (m, 1H), 6.95 (s, 1H), 6.28 (d, *J* = 9.7 Hz, 1H), 5.53 (s, 1H), 5.32 (s, 1H), 2.15 (s, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 161.2, 157.8, 156.6, 152.6, 140.9, 138.6, 132.85, 131.8, 129.4, 128.6, 127.9, 124.3, 123.3, 118.0, 117.1, 112.8, 112.5, 110.7, 101.9, 15.8 ppm; **HRMS (ESI)** $(M+H)^+$ calculated for $C_{20}H_{15}O_4^+$ = 319.0965, found = 319.0969 *m/z*.

Scheme 3.5. Synthesis of 3,1′ cross-coupled product 3.29.

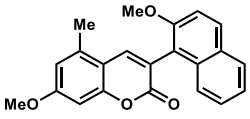




7-methoxy-5-methyl-2*H*-chromen-2-one (3.71). Coumarin (**3.9**; 100 mg, 0.57 mmol, 1.00 equiv) and K_2CO_3 (314 mg, 2.27 mmol, 4.00 equiv) were added to a round bottom flask. Dry acetone (2 mL, 0.5 M) was added at room temperature followed by Me_2SO_4 (135 µL, 1.42 mmol, 2.50 equiv). The reaction was heated to 50 °C and allowed to stir for 16 h. The reaction was cooled to room temperature and quenched by stirring with 1 mL 1.0 M NaOH for 10 min at rt. The reaction was then extracted with ethyl acetate (10 mL x 3). The extracts were combined, washed with brine, dried over Na_2SO_4, and concentrated to a white solid (110 mg, quantitative yield). The material was carried forward without further purification. **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.82 (d, J = 9.7 Hz, 1H), 6.70–6.61 (m, 2H), 6.24 (d, J = 9.7 Hz, 1H), 3.84 (s, 3H), 2.47 (s, 3H); **^{13}C NMR** (150 MHz, $CDCl_3$) δ 162.5, 161.5, 156.7, 140.6, 137.4, 114.1, 112.4, 111.7, 98.8, 55.8, 18.6; **HRMS** (ESI) *m/z* calculated for $C_{11}H_{11}O_3^+$ $[M+H]^+$ 191.0703, found 191.0705.[74]

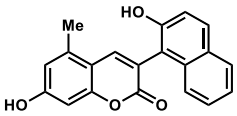


3-bromo-7-methoxy-5-methyl-2*H*-chromen-2-one (3.72). Coumarin (**3.71**, 92 mg, 0.48 mmol, 1.0 equiv) was added to a round bottom flask and dissolved in dry CH_2Cl_2 (1.4 mL, 0.35 M). NBS (130 mg, 0.73 mmol, 1.5 equiv) added at room temperature under a stream of nitrogen and allowed to stir at room temperature overnight. The reaction was quenched with water and extracted with ethyl acetate (10 mL x 3). The extracts were combined, washed with brine, dried over Na_2SO_4, and concentrated to provide an off-white solid. The crude material was purified by flash chromatography over silica gel with a gradient of 0–40% ethyl acetate in hexanes. 86 mg of the desired compound was recovered as white solid in a 66% yield. **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.18 (s, 1H), 6.71 (s, 1H), 6.67 (s, 1H), 3.85 (s, 3H), 2.47 (s, 3H); **^{13}C NMR** (151 MHz, $CDCl_3$) δ 162.8, 157.7, 155.9, 142.0, 136.9, 114.7, 110.2, 107.3, 98.7, 55.9, 18.7; **HRMS** (ESI) *m/z* calculated for $C_{11}H_{10}BrO_3^+$ $[M+H]^+$ 268.9808, found 268.9811.



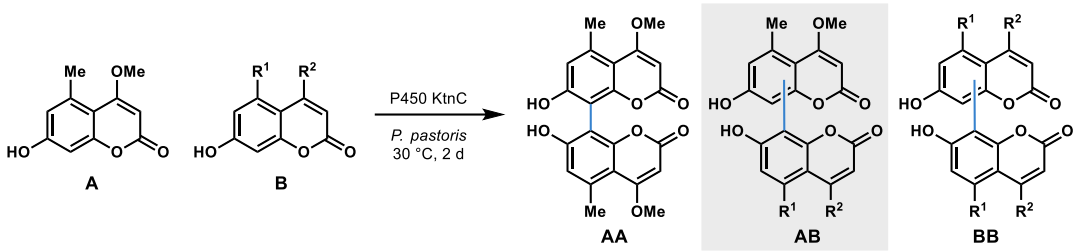
7-methoxy-3-(2-methoxynaphthalen-1-yl)-5-methyl-2*H*-chromen-2-one (3.73). Coumarin **3.72** (15 mg, 56 µmol, 1.0 equiv), boronic acid (**3.65**; 17 mg, 0.84 mmol, 1.5 equiv), K_3PO_4 (36 mg, 0.17, 3.0 equiv), $Pd(OAc)_2$ (2.5 mg, 11 µmol, 0.2 equiv), and XantPhos (13 mg, 22 µmol, 0.40 equiv) were added to a dry round bottom flask. Dry THF (2.2 mL, 0.025 M) was added, and the reaction was heated to 65 °C for 18 h. The reaction was cooled, quenched with 1 M NaOH, and extracted with CH_2Cl_2 (10 mL x 3). The combined extracts were washed with brine, dried over Na_2SO_4, and concentrated to afford a reddish-brown solid. The crude material was purified by flash chromatography over silica gel with a gradient of 5–30% ethyl acetate in hexanes, resulting in isolation of 14.8 mg of the target compound as a white solid in a 77% yield. **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.93 (d, J = 9.0 Hz, 1H), 7.84 (d, J = 7.5 Hz, 2H), 7.67 (d, J = 8.5 Hz, 1H), 7.42 (ddd, J = 8.4, 6.7, 1.4 Hz, 1H), 7.39–7.34 (m, 2H), 6.79 (d, J = 2.4 Hz, 1H), 6.73 (d, J = 2.6 Hz, 1H), 3.93 (s, 3H), 3.89 (s, 3H), 2.45 (s, 3H); **^{13}C NMR** (151 MHz, $CDCl_3$) δ 162.3, 160.9, 156.5, 155.1, 141.3, 137.3, 133.4, 130.6, 129.3, 128.4, 127.1, 124.1, 123.8, 120.35, 118.4, 114.0, 113.6,

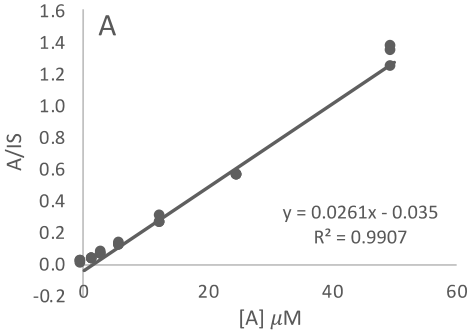
112.4, 98.6, 56.9, 55.8, 18.8; **HRMS** (ESI) *m/z* calculated for $C_{22}H_{19}O_4^+$ $[M+H]^+$ 347.1278, found 347.1284.

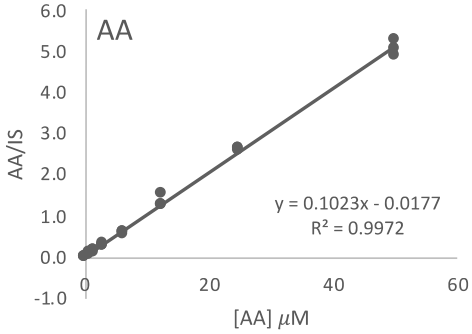


7-hydroxy-3-(2-hydroxynaphthalen-1-yl)-5-methyl-2*H*-chromen-2-one (3.29). Biaryl (**3.73**; 28 mg, 81 µmol, 1.0 equiv) was dissolved in dry CH_2Cl_2 (1.6 mL, 0.05 M) in a dry round bottom flask and cooled to –78 °C. BBr_3 (31 µL, 0.32 mmol, 4.0 equiv) was added dropwise, and the reaction mixture was allowed to warm to room temperature and stirred for 18 h. The reaction was quenched by cooling to 0 °C, then adding water and stirring for 10 minutes. The pH was adjusted to approximately 4 by addition of aqueous saturated sodium bicarbonate and extracted with ethyl acetate (10 mL x 3). The extracts were combined, washed with brine, dried over Na_2SO_4, and concentrated to a brown solid. The crude material was purified flash chromatography over silica gel 7–60% ethyl acetate in hexanes to recover 25 mg of the title compound as a white solid in a 97% yield. **^{1}H NMR** (600 MHz, acetone-d_6) δ 8.91 (s, 2H), 8.00 (s, 1H), 7.87–7.82 (m, 2H), 7.60 (dd, J = 8.5, 1.1 Hz, 1H), 7.38 (ddd, J = 8.3, 6.7, 1.3 Hz, 1H), 7.31 (ddd, J = 8.0, 6.7, 1.2 Hz, 1H), 7.27 (d, J = 8.9 Hz, 1H), 6.77 (dd, J = 2.3, 1.0 Hz, 1H), 6.70 (d, J = 2.3 Hz, 1H), 2.50 (s, 3H). **^{13}C NMR** (150 MHz, acetone-d_6) δ 161.5, 160.8, 157.6, 153.8, 143.0, 139.2, 134.9, 130.7, 129.6, 128.9, 127.3, 125.0, 123.8, 120.0, 119.3, 116.7, 115.1, 112.5, 101.2, 18.6; **HRMS** (ESI) *m/z* calculated for $C_{20}H_{13}O_4^-$ $[M-H]^-$ 317.0819, found 317.0824.

Standard curves and LC–MS traces for biocatalytic reactions. High resolution electrospray mass spectra were obtained on an Agilent UPLC–QTOF at the University of Michigan Life Sciences Institute or Agilent UPLC–TOF at the University of Michigan Life Sciences Institute.

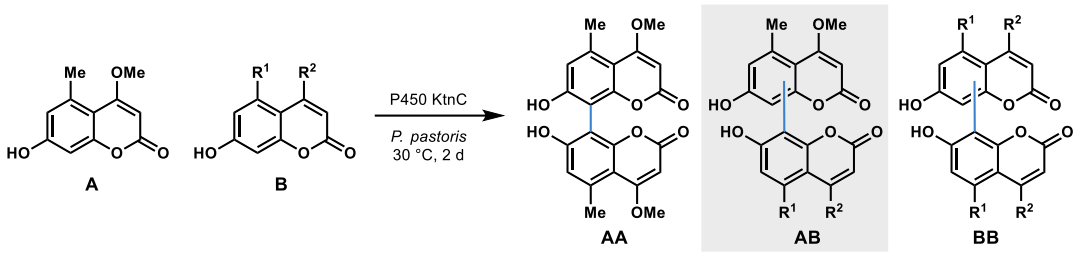


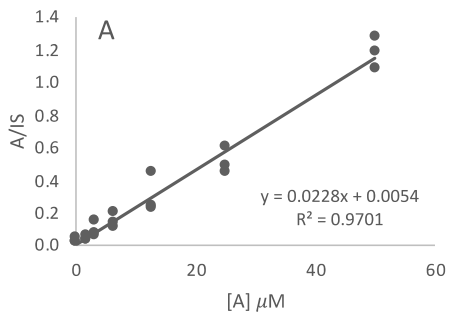


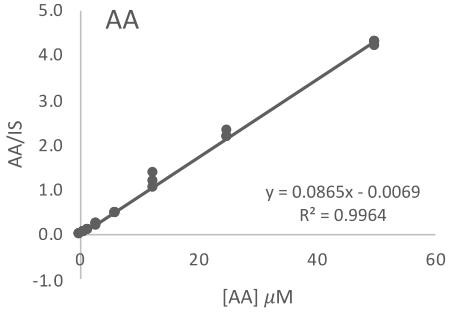


R1	R2	entry #	A/IS	AA/IS	$[A]_{remaining}$	$[AA]_{present}$	AB	% conv AB	average	% conv AA	average
H	OMe	A5	0.156	0.141	7.31570	1.54898	14.586	58.3	67.3	12.4	14.1
		A6	0.016	0.185	1.97117	1.98662	19.056	76.2		15.9	
Et	OMe	A1	0.371	0.009	15.55702	0.25933	8.924	35.7	31.0	2.1	2.2
		A2	0.431	0.013	17.84073	0.29776	6.564	26.3		2.4	
OH	OMe	C1	0.00	0.542	1.33926	5.46788	12.725	50.9	50.2	43.7	44.5
		C2	0.00	0.560	1.33926	5.64550	12.370	49.5		45.2	
Me	NHMe	A3	0.000	0.247	1.33926	2.58407	18.493	74.0	74.0	20.7	20.7
		A4	0.000	0.246	1.33926	2.57929	18.502	74.0		20.6	
Cl	OMe	C5	0.515	0.027	21.08830	0.43675	3.038	12.2	21.2	3.5	4.0
		C6	0.391	0.041	16.31739	0.57053	7.542	30.2		4.6	
Br	OMe	C3	0.396	0.016	16.50020	0.32768	7.844	31.4	29.9	2.6	2.7
		C4	0.414	0.017	17.21717	0.34356	7.096	28.4		2.7	
Me	H	B1	0.000	0.265	1.33926	2.76393	18.133	72.5	73.4	22.1	21.2
		B2	0.000	0.243	1.33926	2.54569	18.569	74.3		20.4	
H	Me	B3	0.000	0.494	1.33926	5.00279	13.655	54.6	54.0	40.0	40.6
		B4	0.000	0.510	1.33926	5.15688	13.347	53.4		41.3	
H	H	C1	0.000	0.706	1.33926	7.07236	9.516	38.1	39.0	56.6	55.7
		C2	0.000	0.683	1.33926	6.85020	9.960	39.8		54.8	

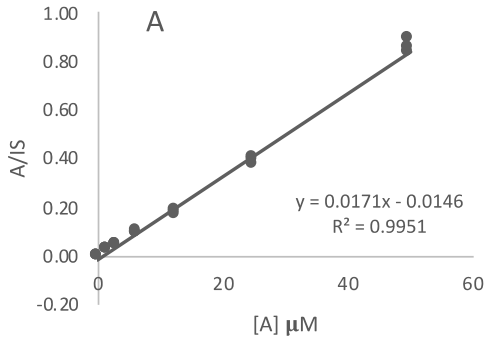
Figure 3.9. Standard curves for oxidative cross-coupling of 3.4 and 3.8–3.16 by KtnC.

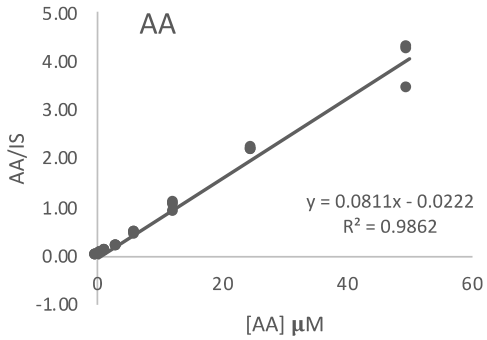






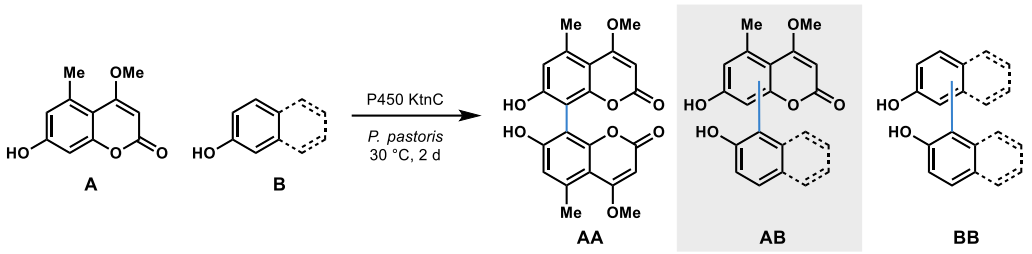
R1	R2	entry #	A/IS	AA/IS	$[A]_{remaining}$	$[AA]_{present}$	AB	% conv AB	average	% conv AA	average
Me	OEt	A1	0.01	0.05	1.67	0.64	22.05	88.19	87.3	5.12	4.7
		A2	0.03	0.04	2.32	0.53	21.62	86.49		4.24	
Me	OPr	A3	0.24	0.21	11.00	2.47	9.06	36.22	37.7	19.79	17.9
		A4	0.24	0.17	11.17	2.01	9.80	39.22		16.10	
Me	OiPr	A5	0.11	0.26	5.74	3.05	13.15	52.60	54.0	24.42	26.2
		A6	0.07	0.29	4.17	3.48	13.86	55.44		27.88	
Me	OBu	B1	0.07	0.34	3.94	3.96	13.14	52.56	49.7	31.70	34.5
		B2	0.07	0.40	3.97	4.66	11.70	46.80		37.30	

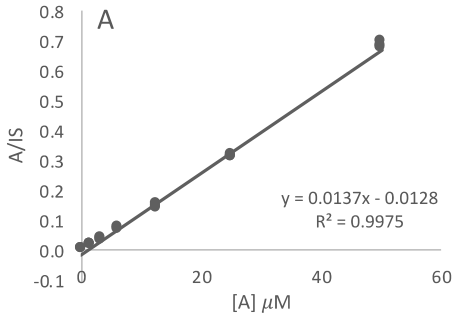


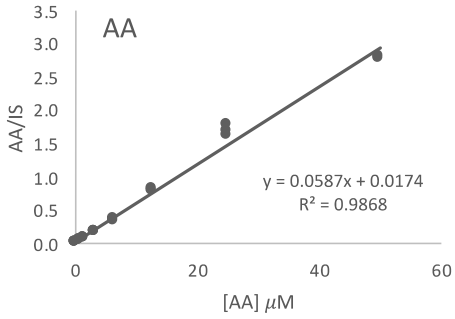


R1	R2	entry #	A/IS	AA/IS	$A_{remaining}$	$AA_{present}$	AB	% conv AB	average	% AA	average
Me	OtBu	B3	0.000	0.568	0.85302	7.27818	9.591	38.4	34.9	58.2	61.7
		B4	0.000	0.639	0.85302	8.15289	7.841	31.4		65.2	

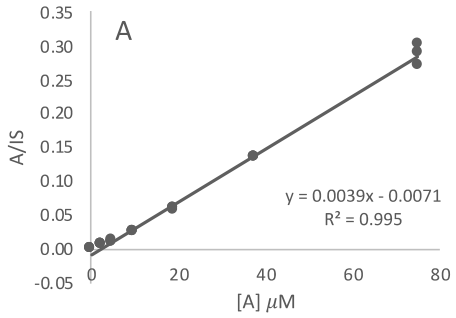
Figure 3.10. Standard curves for oxidative cross-coupling of 3.4 and 3.17–3.21 by KtnC.

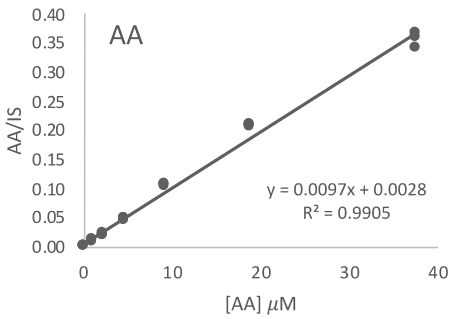






B partner	entry #	A/IS	AA/IS	$[A]_{remaining}$	$[AA]_{present}$	AB	% conv AB	average	% conv AA	average
benzofuran	B3	0.000	0.333	0.93431	5.38362	13.298	53.2	51.9	43.1	44.4
	B4	0.000	0.353	0.93431	5.70944	12.647	50.6		45.7	





Phenol	entry #	A/IS	AA/IS	$[A]_{remaining}$	$[AA]_{present}$	AB	% conv AB	average	% conv AA	average
quinoline	C3	0.051	0.340	15.02189	34.84561	15.287	15.3	14.2	69.7	72.1
	C4	0.041	0.363	12.31828	37.27717	13.127	13.1		74.6	
2-naphthol	D1	0.121	0.322	33.02458	33.02720	0.921	0.9	2.0	66.1	65.1
	D2	0.119	0.314	32.72324	32.12063	3.035	3.0		64.2	
7-methoxy-2-naphthol	D3	0.250	0.051	66.52782	4.93853	23.595	23.6	20.3	9.9	10.5
	D4	0.272	0.056	72.05196	5.51253	16.923	16.9		11.0	
carbazole	D5	0.001	0.388	2.15328	39.77408	18.299	18.3	15.4	79.5	82.0
	D6	0.005	0.412	3.05648	42.24427	12.455	12.5		84.5	

Figure 3.11. Standard curves for oxidative cross-coupling of 3.4 and 3.22–3.26 by KtnC.

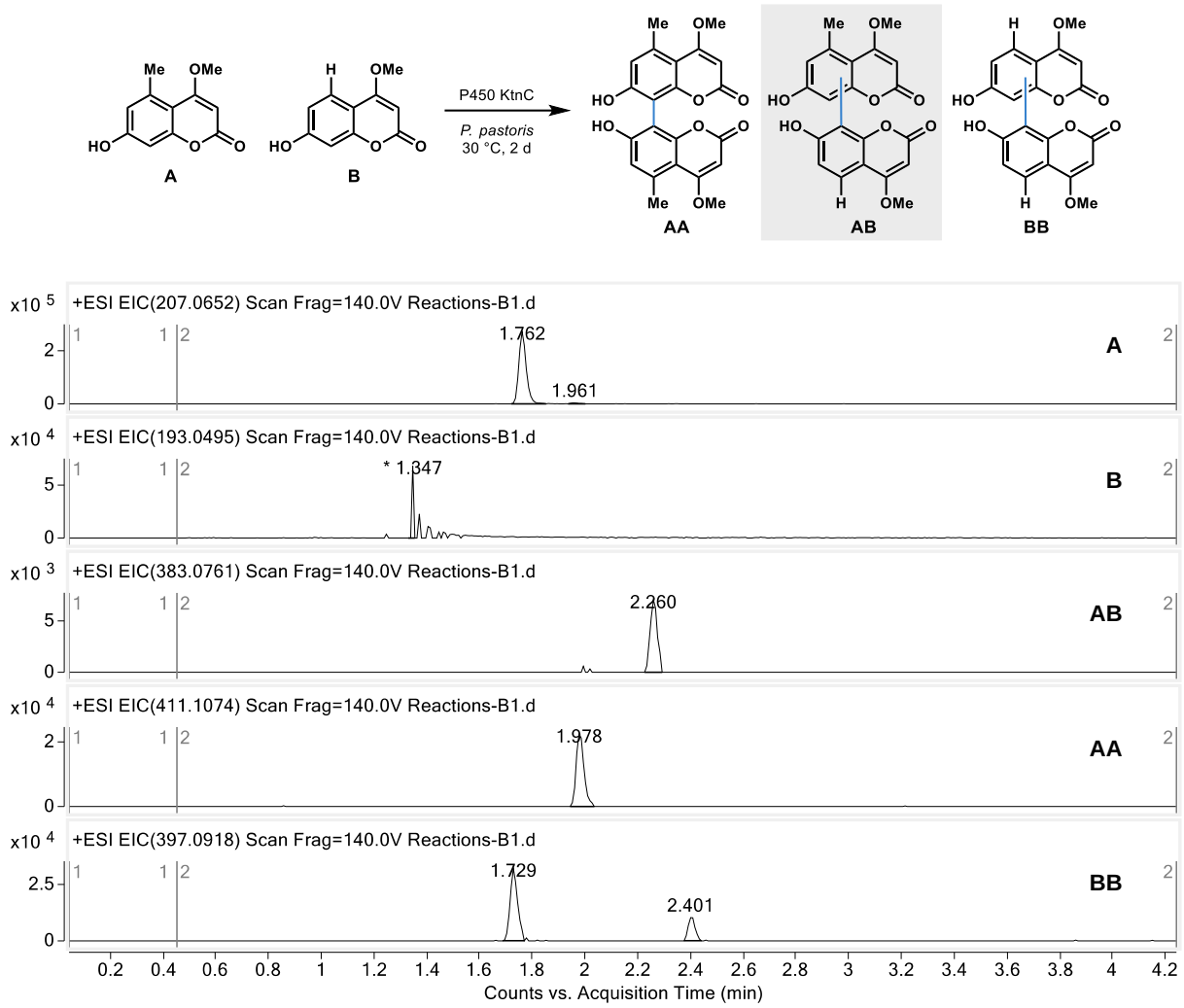


Figure 3.12. LC–MS traces for oxidative cross-coupling of 3.4 and 3.8 by KtnC.

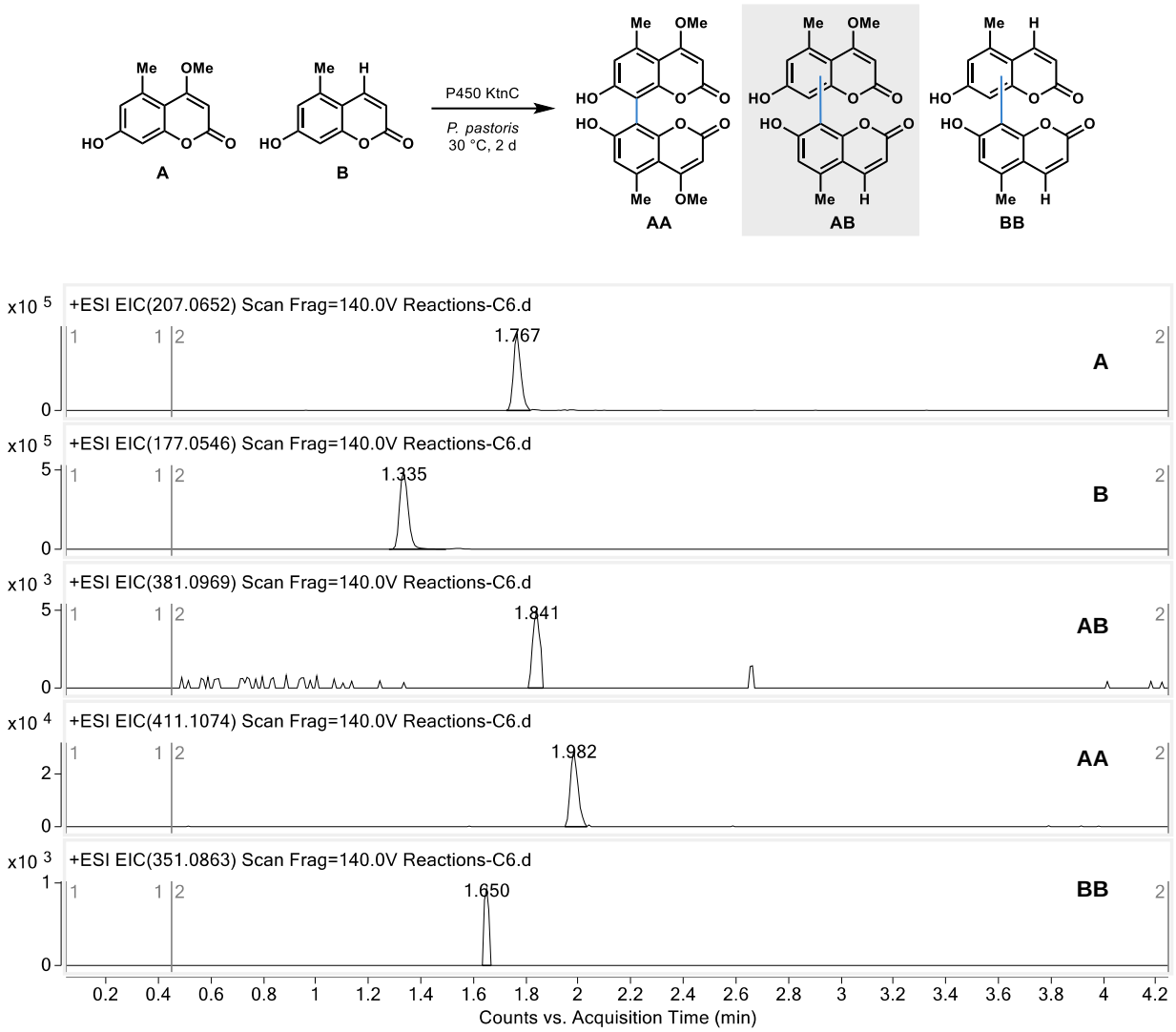


Figure 3.13. LC–MS traces for oxidative cross-coupling of 3.4 and 3.9 by KtnC.

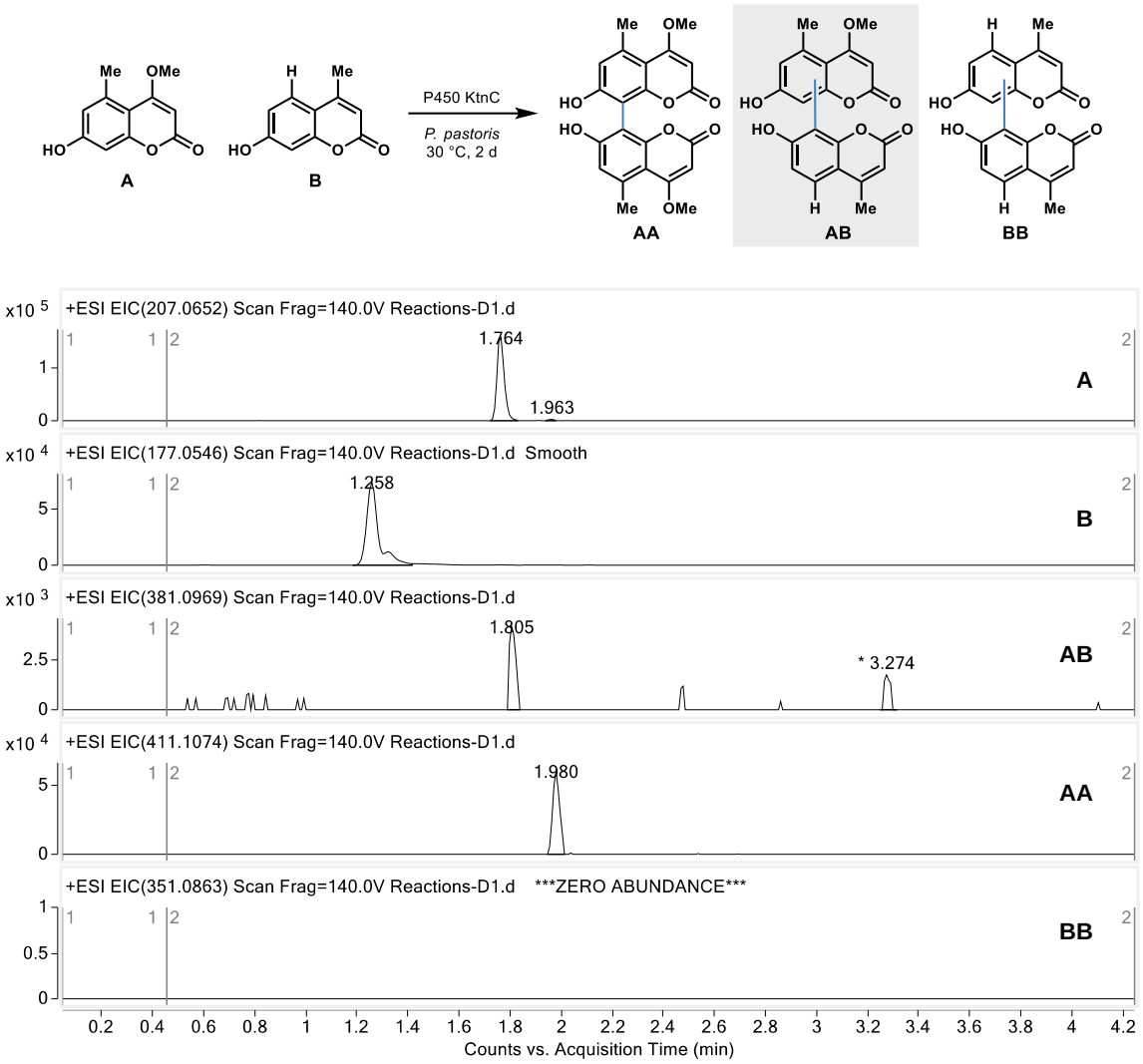


Figure 3.14. LC–MS traces for oxidative cross-coupling of 3.4 and 3.10 by KtnC.

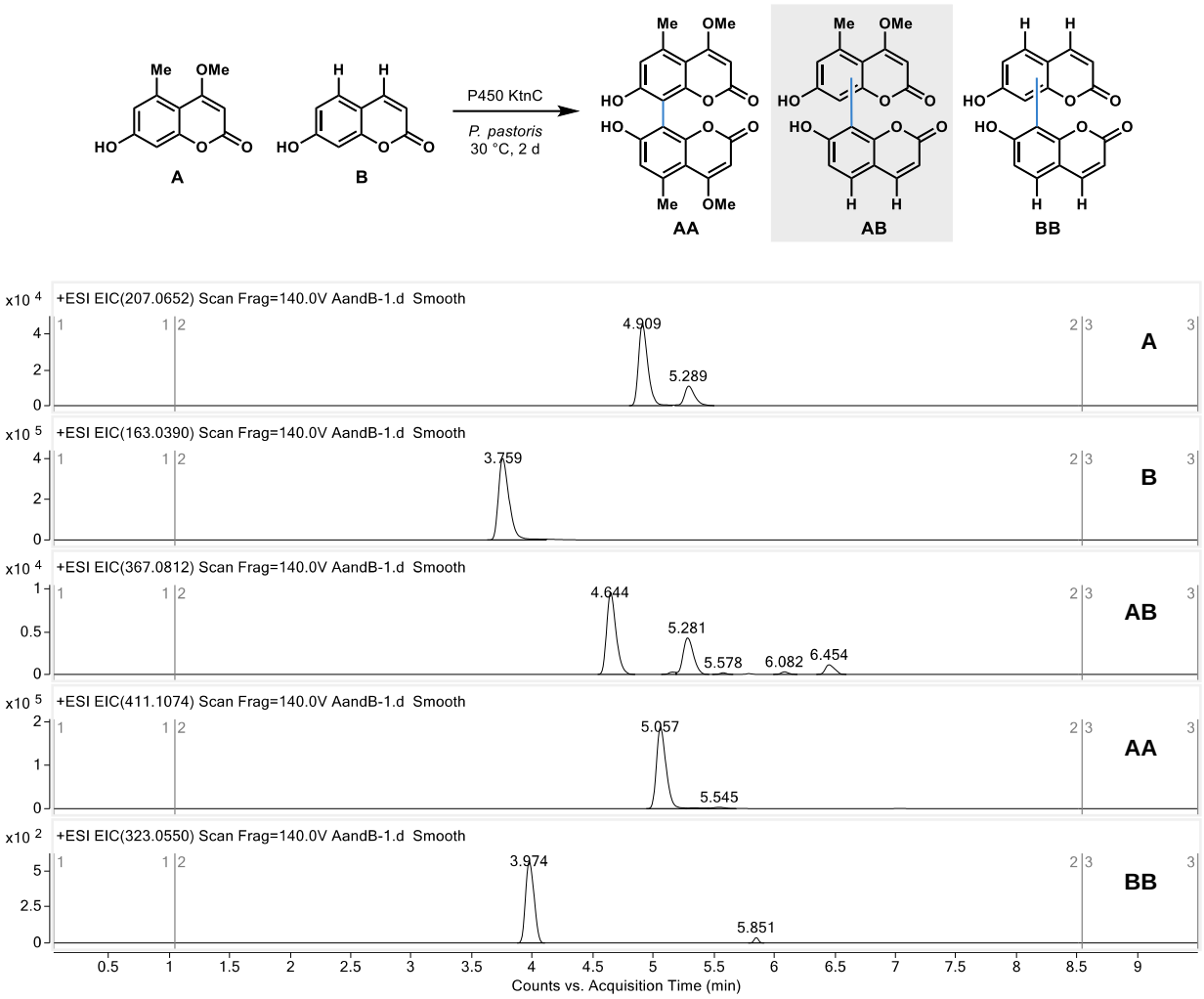


Figure 3.15. LC–MS traces for oxidative cross-coupling of 3.4 and 3.11 by KtnC.

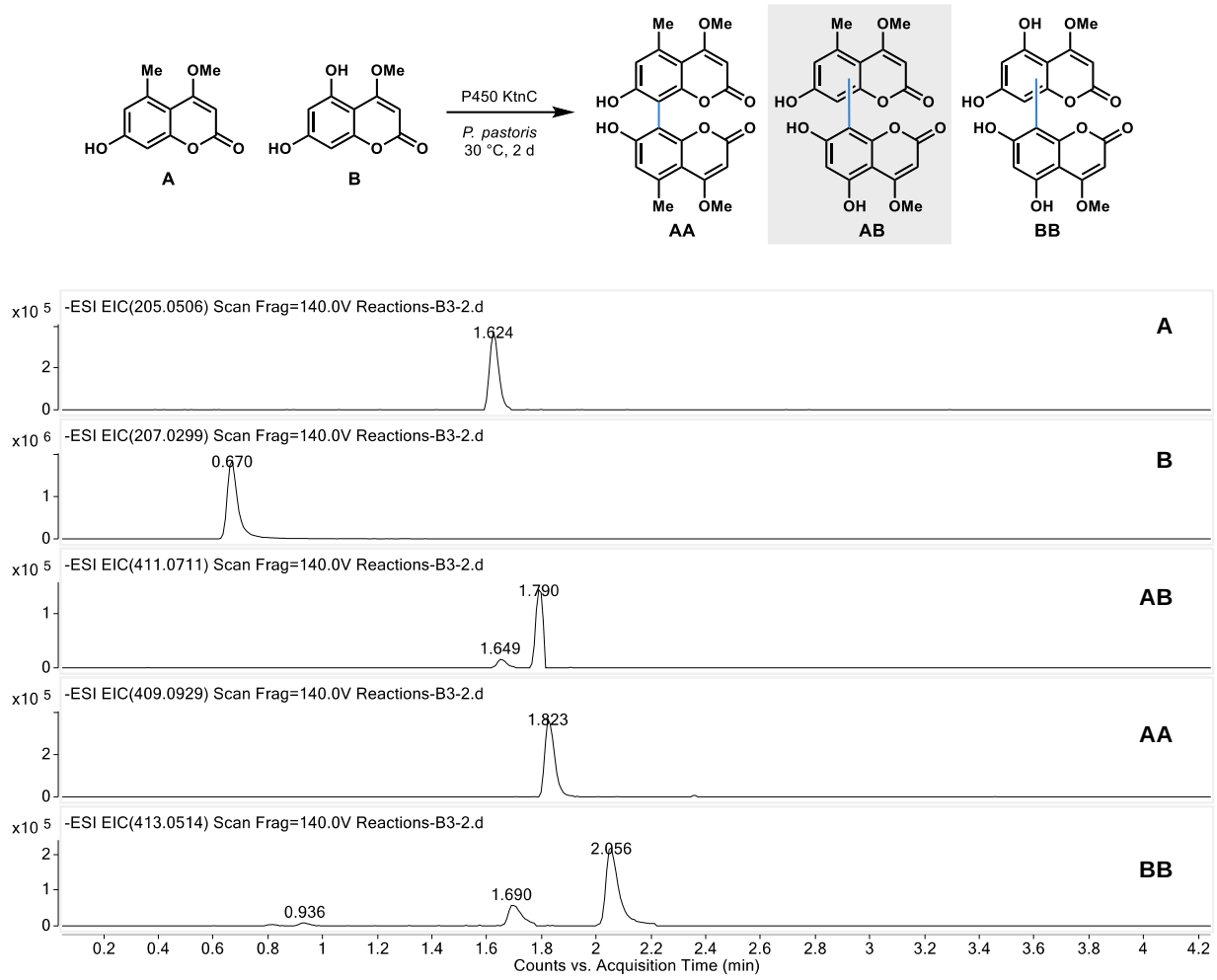


Figure 3.16. LC–MS traces for oxidative cross-coupling of 3.4 and 3.12 by KtnC.

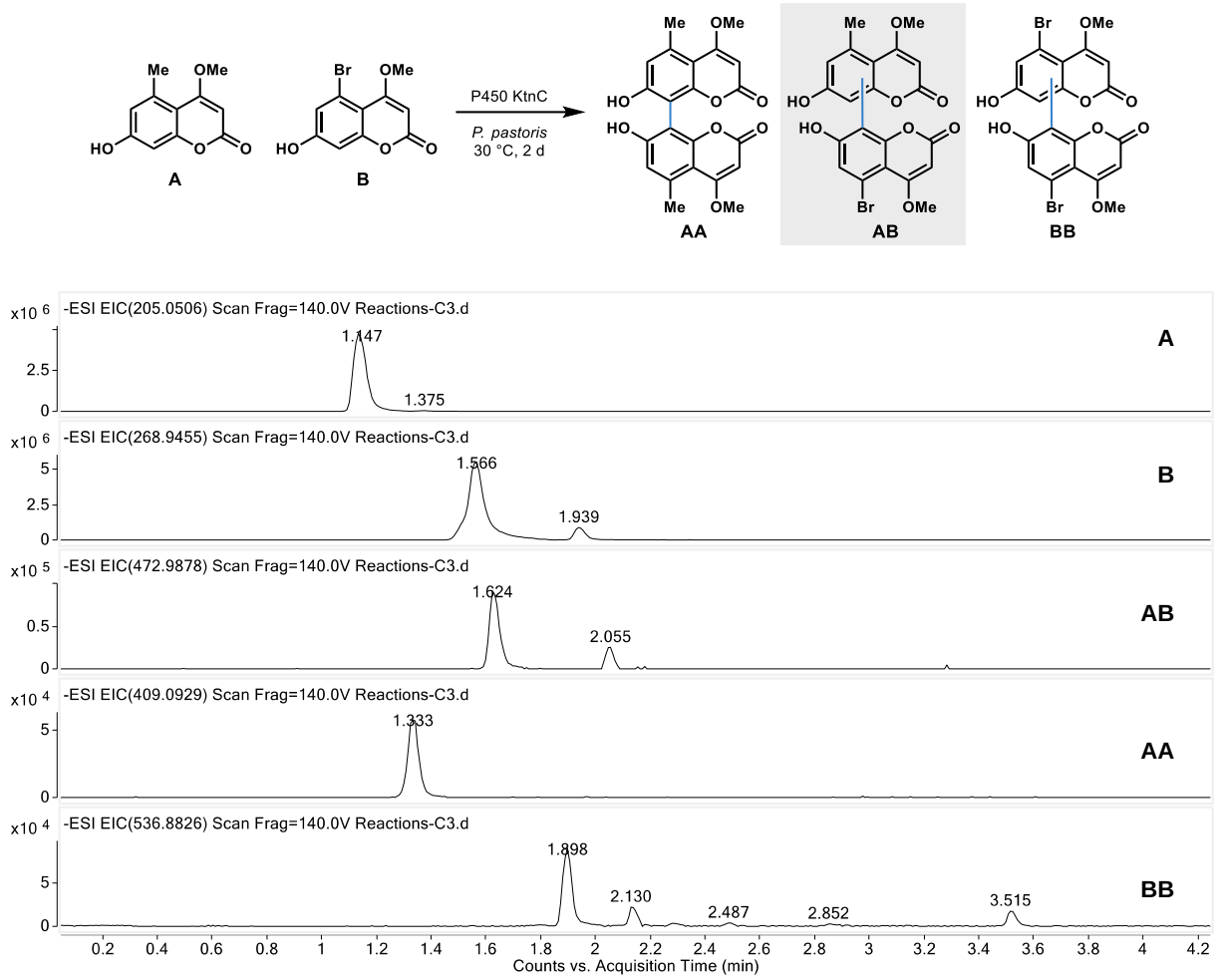


Figure 3.17. LC–MS traces for oxidative cross-coupling of 3.4 and 3.13 by KtnC.

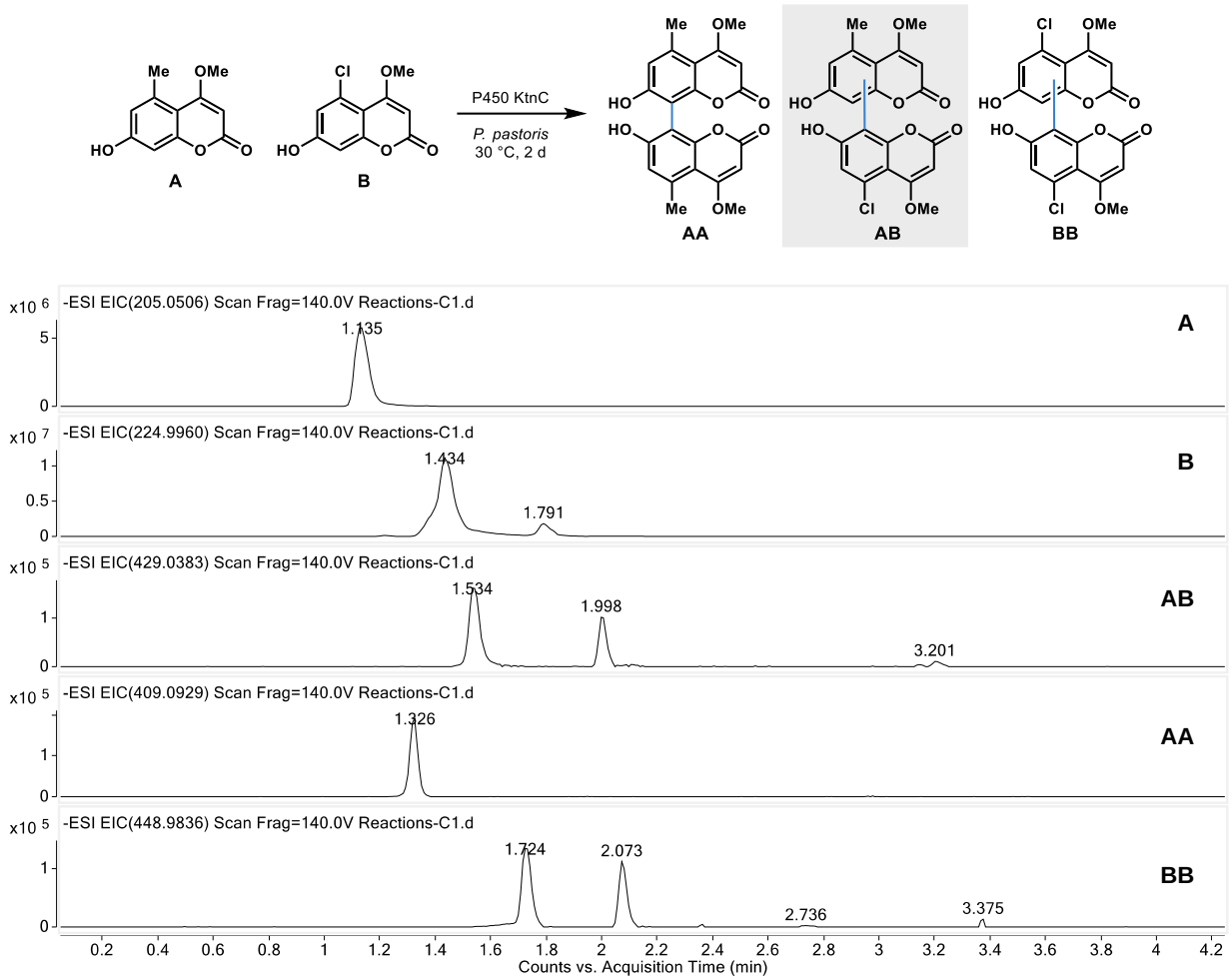


Figure 3.18. LC–MS traces for oxidative cross-coupling of 3.4 and 3.14 by KtnC.

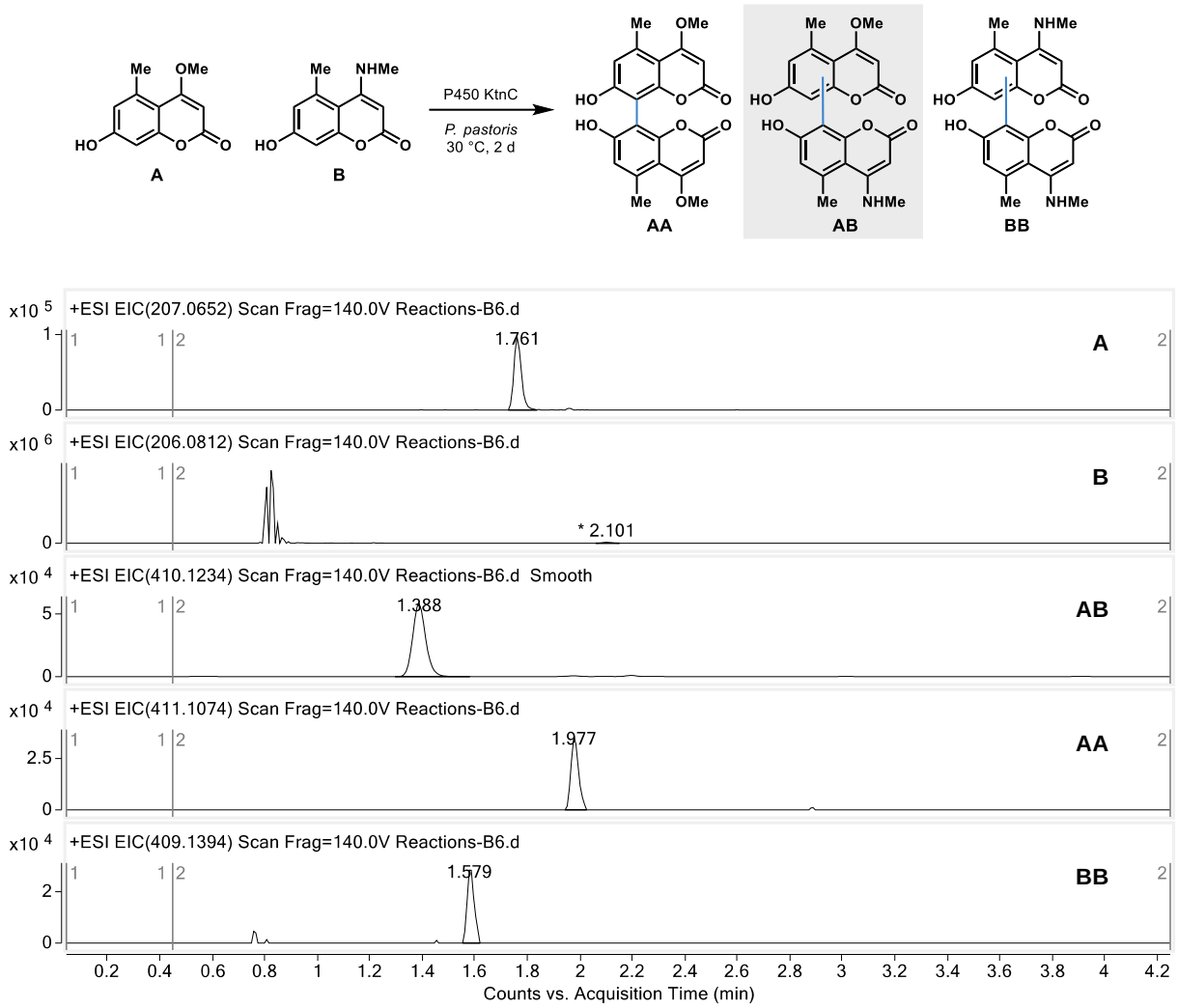


Figure 3.19. LC–MS traces for oxidative cross-coupling 3.4 and 3.15 by KtnC.

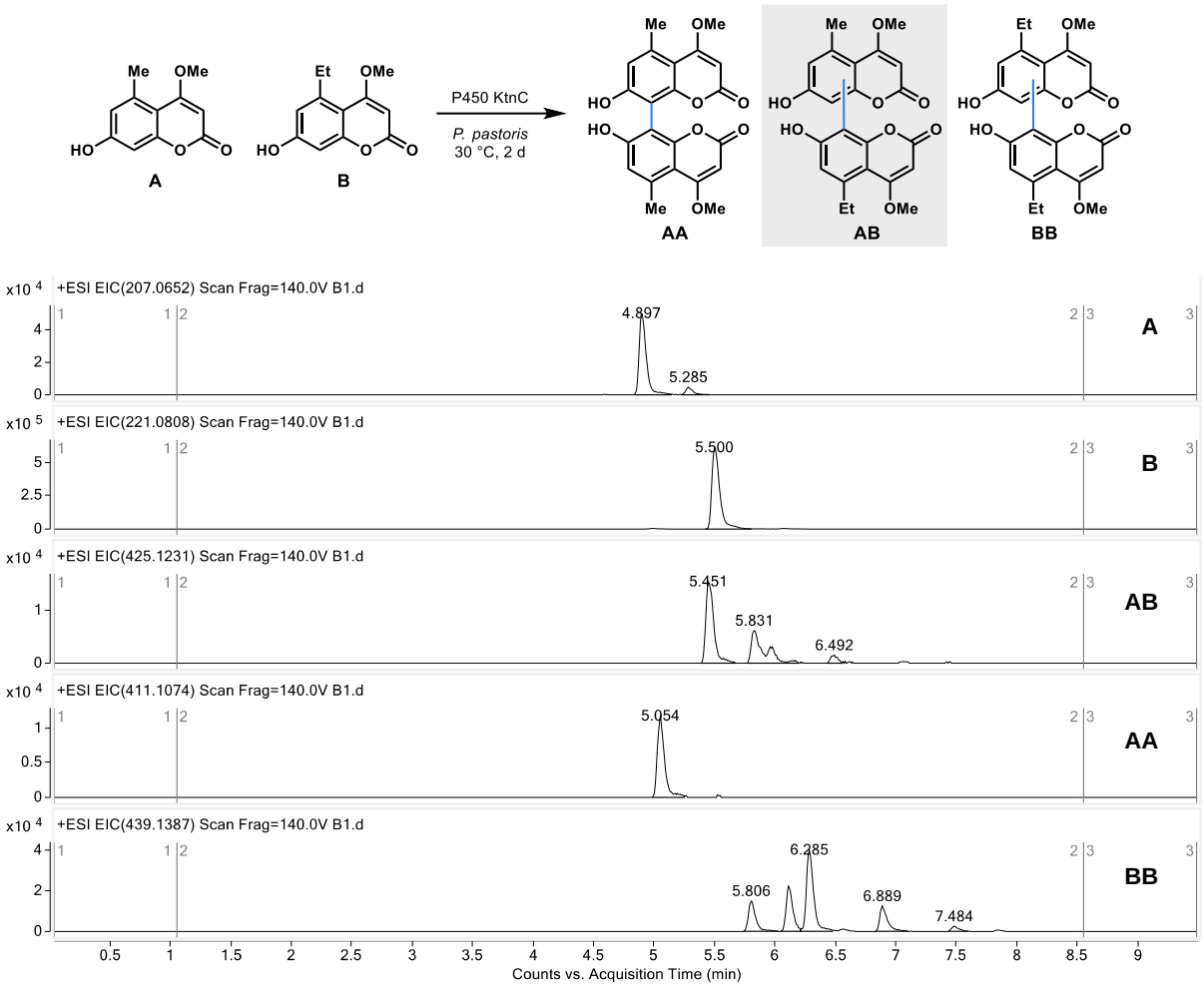


Figure 3.20. LC–MS traces for oxidative cross-coupling of 3.4 and 3.16 by KtnC.

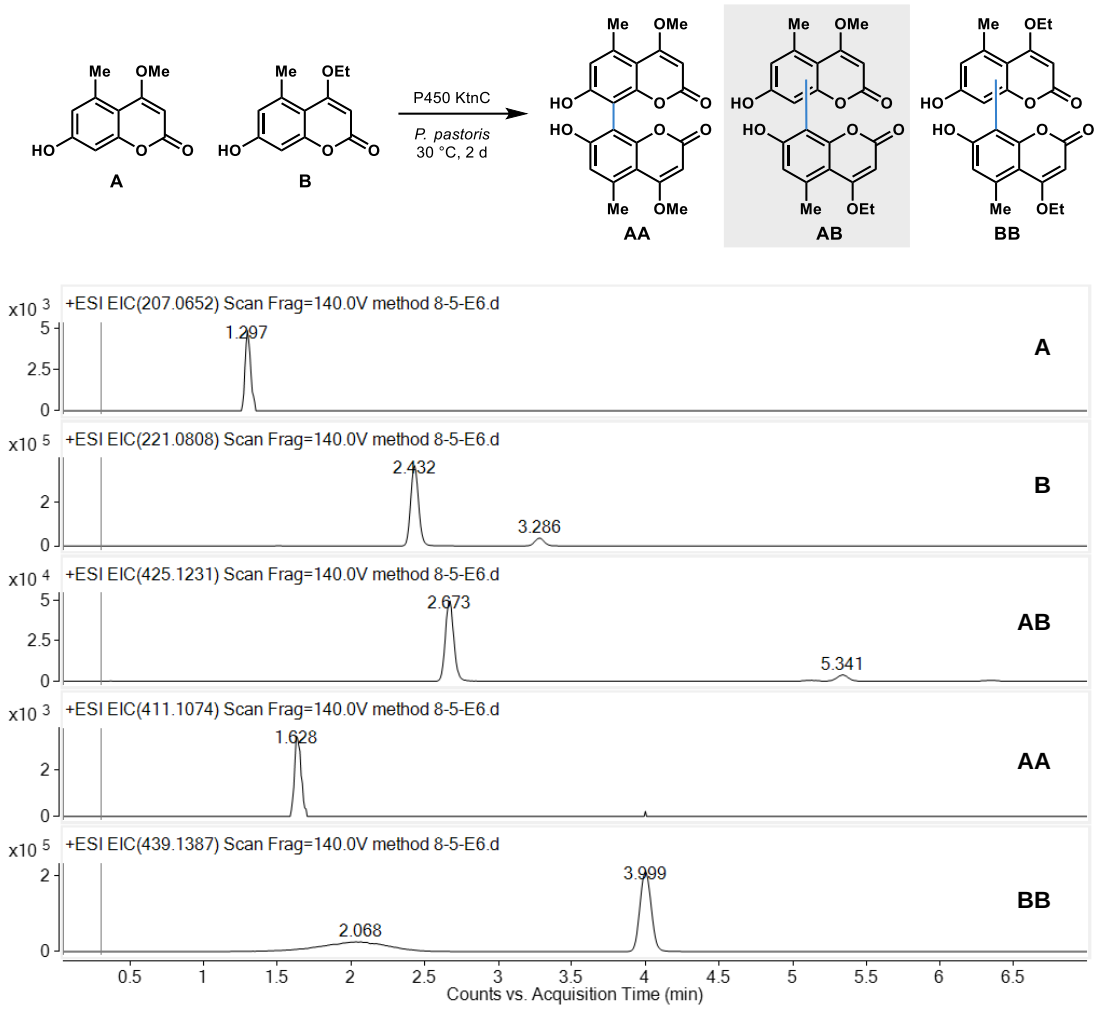


Figure 3.21. LC–MS traces for oxidative cross-coupling of 3.4 and 3.17 by KtnC.

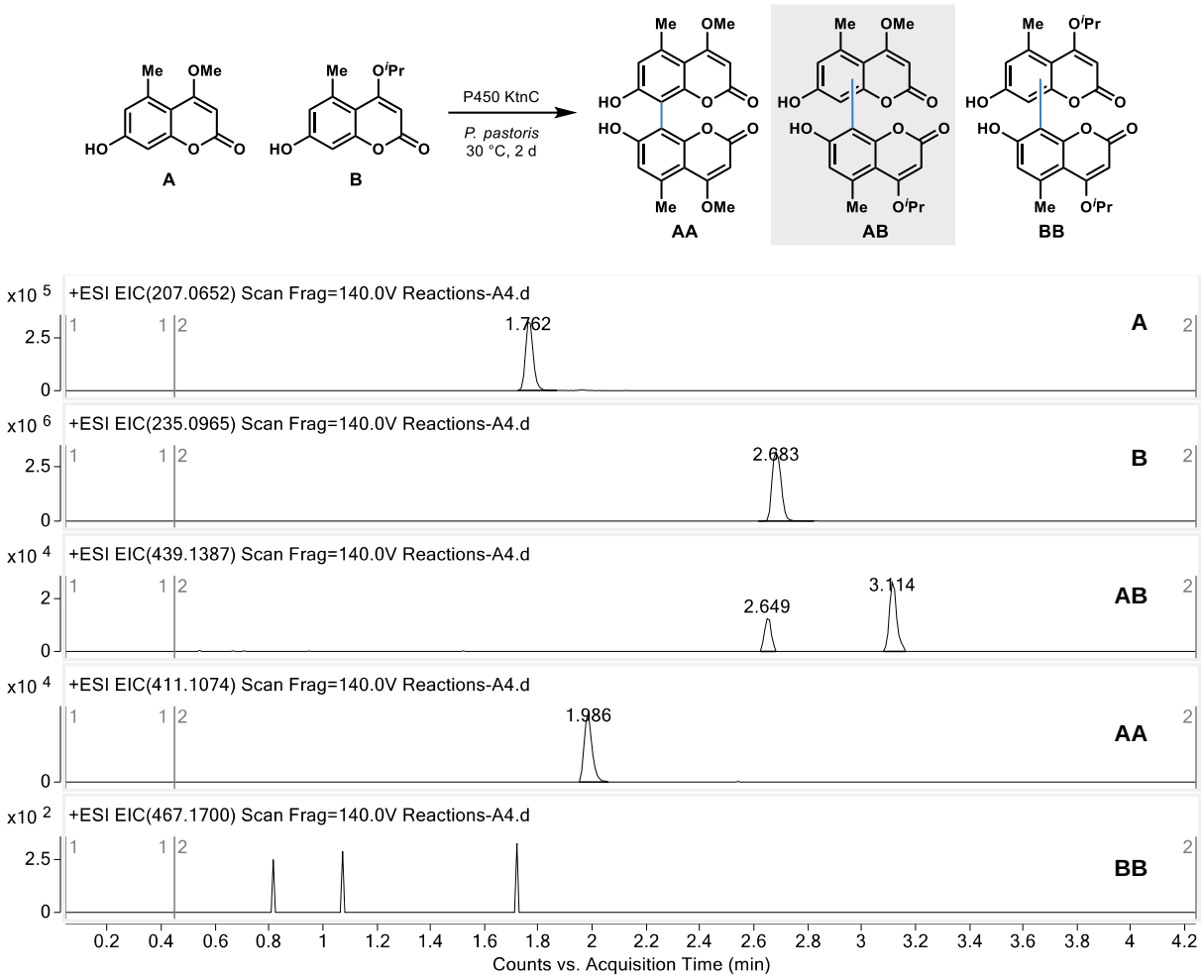


Figure 3.22. LC–MS traces for oxidative cross-coupling of 3.4 and 3.18 by KtnC.

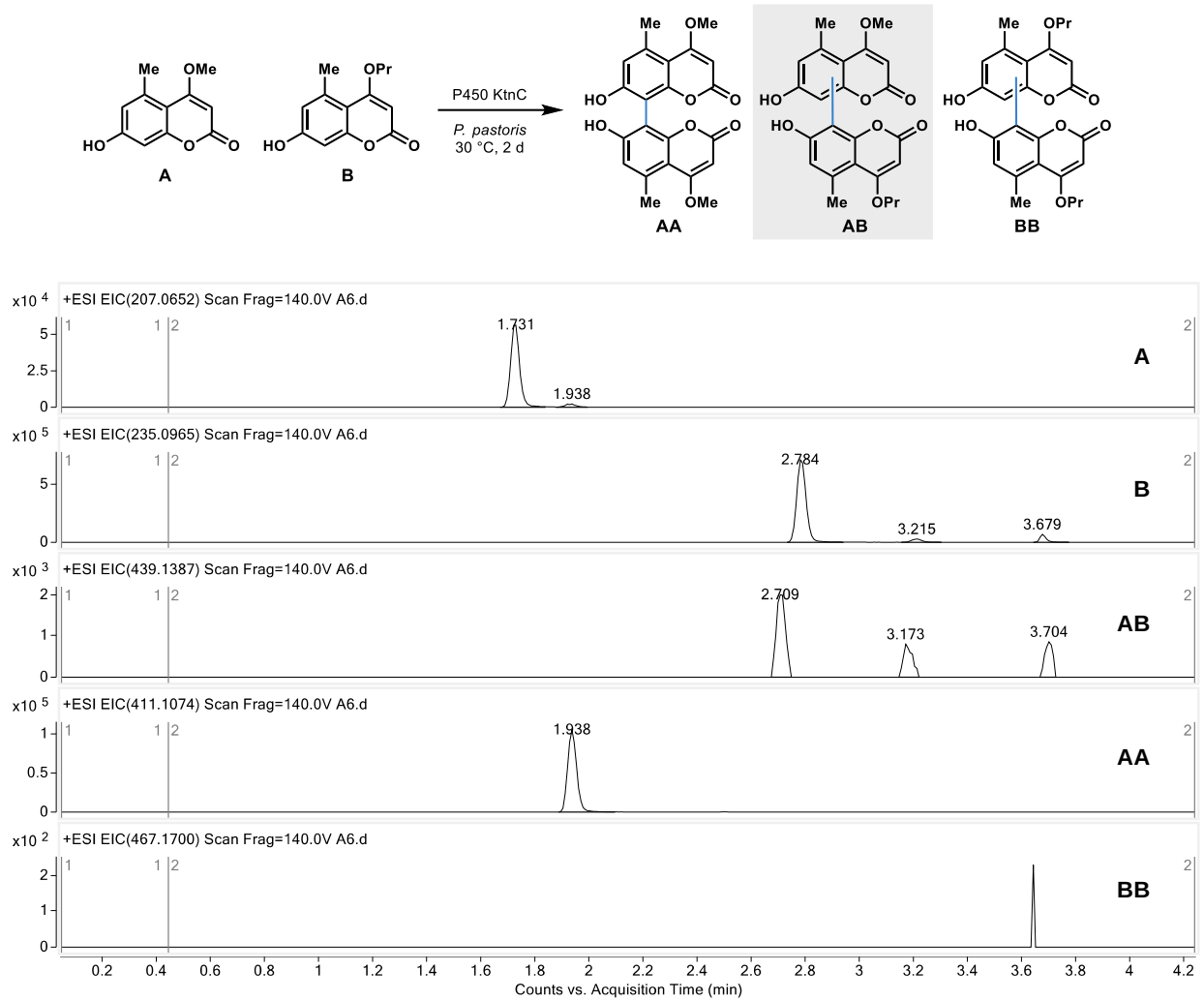


Figure 3.23. LC–MS traces for oxidative cross-coupling of 3.4 and 3.19 by KtnC.

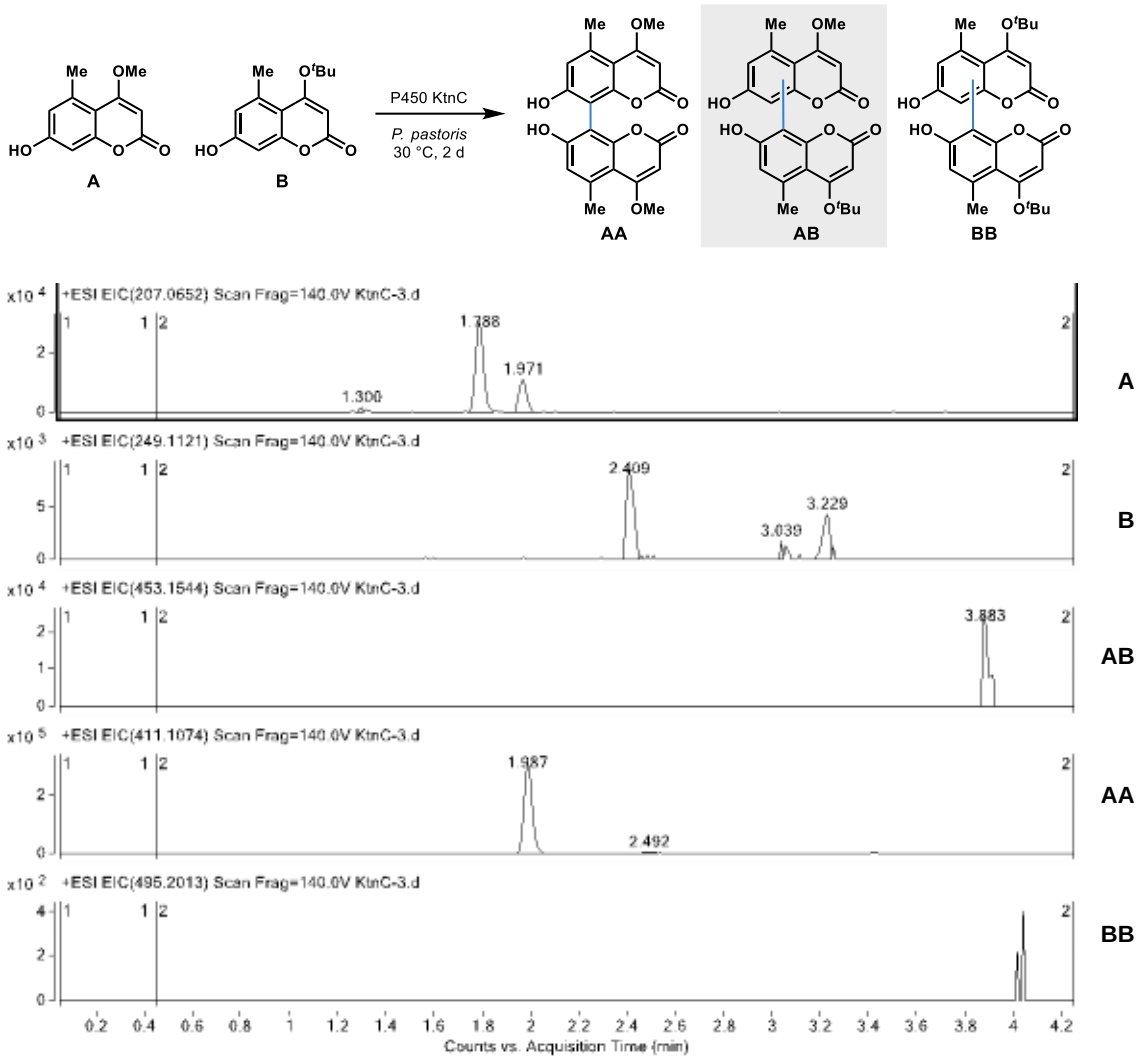


Figure 3.24. LC–MS traces for oxidative cross-coupling of 3.4 and 3.20 by KtnC.

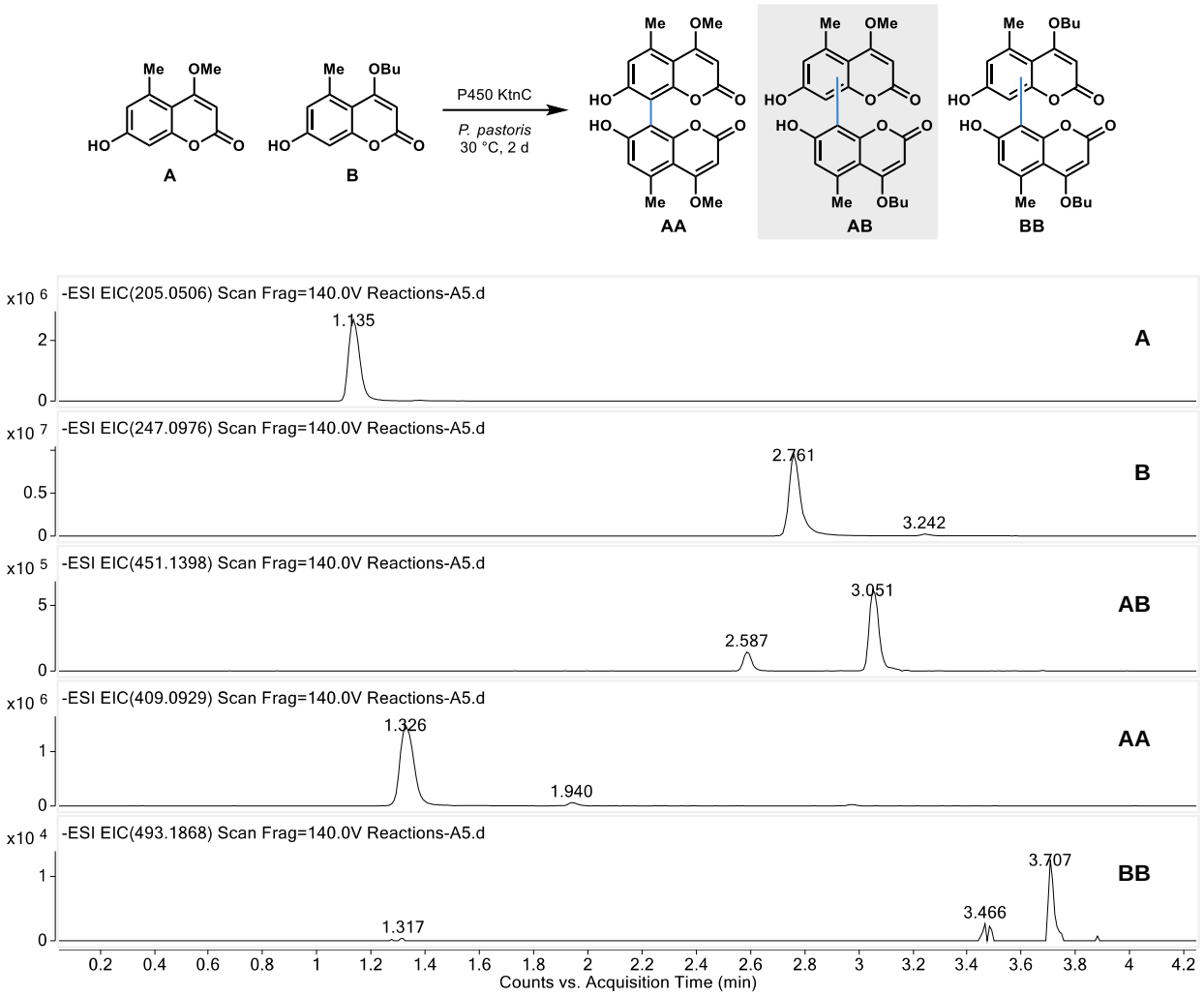


Figure 3.25. LC–MS traces for oxidative cross-coupling of 3.4 and 3.21 by KtnC.

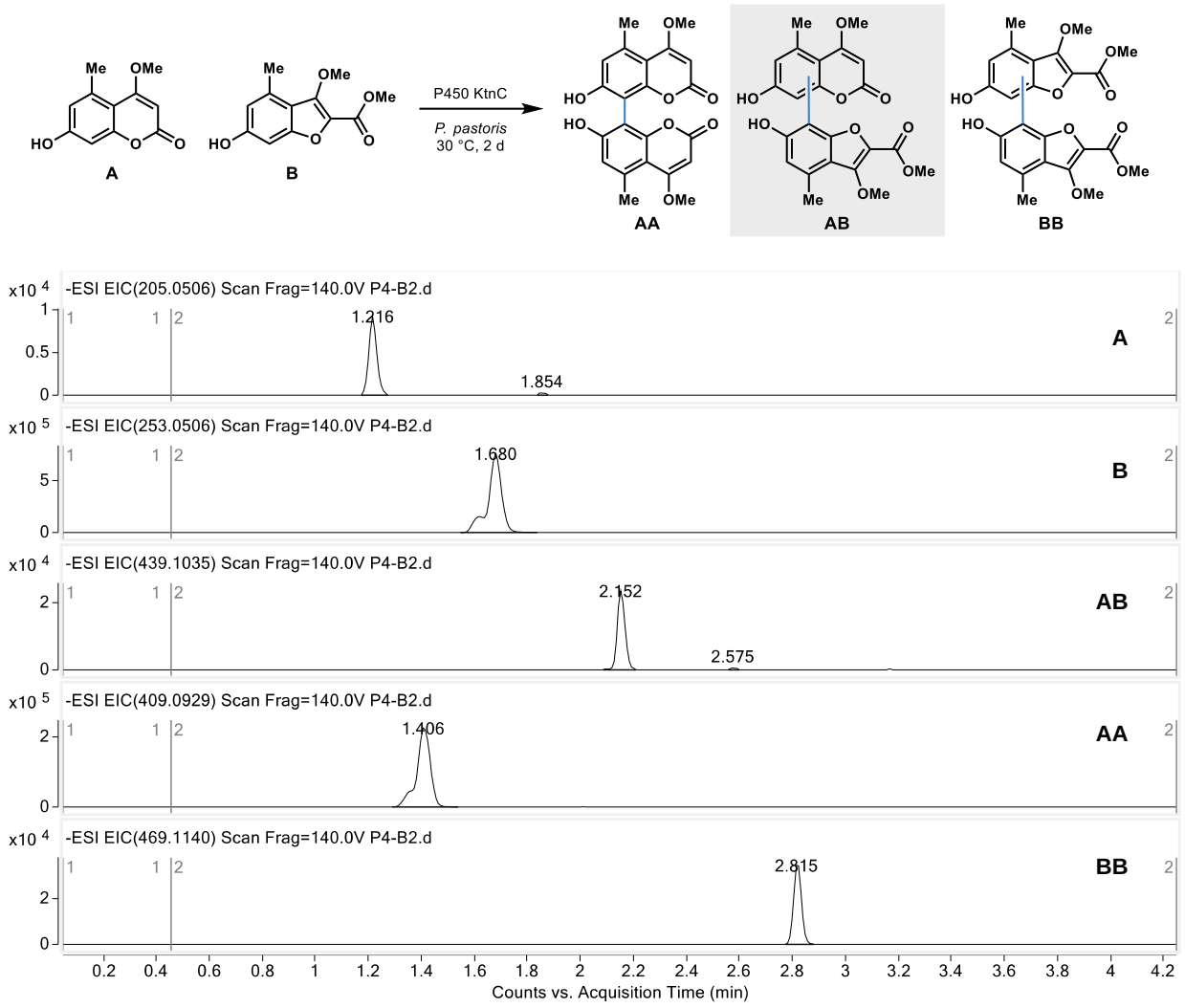


Figure 3.26. LC–MS traces for oxidative cross-coupling of 3.4 and 3.22 by KtnC.

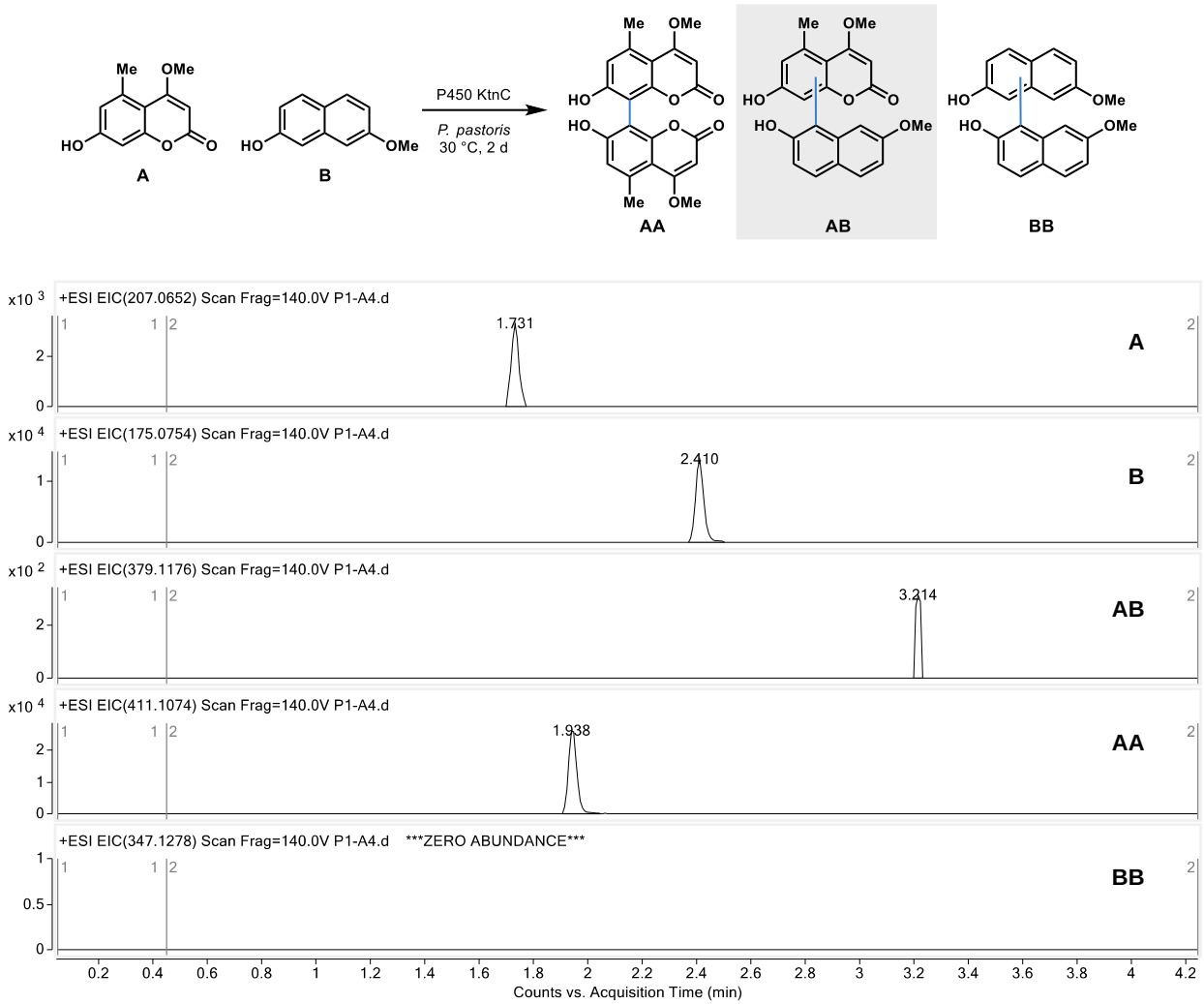


Figure 3.27. LC–MS traces for oxidative cross-coupling of 3.4 and 3.23 by KtnC.

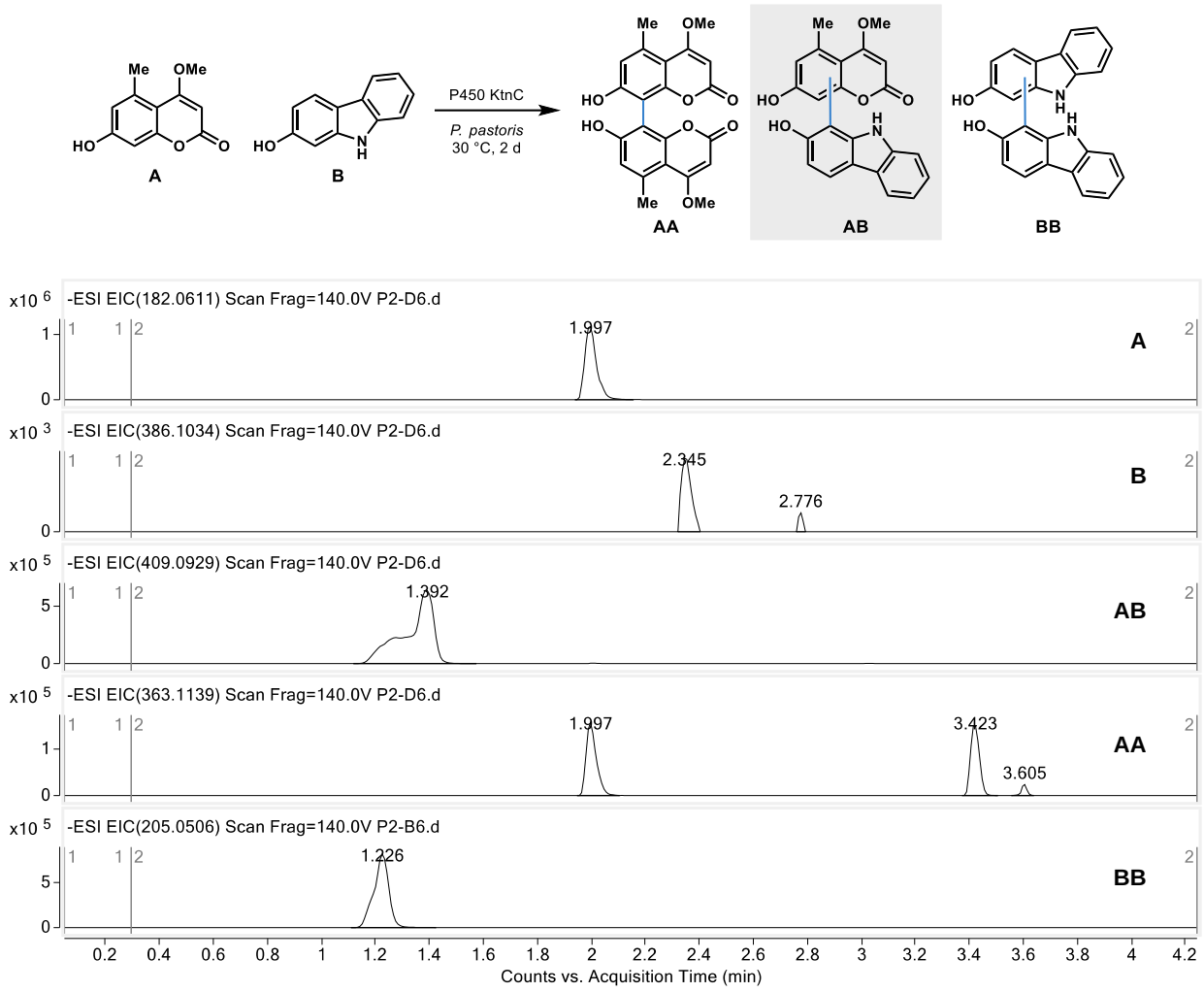


Figure 3.28. LC–MS traces for oxidative cross-coupling of 3.4 and 3.24 by KtnC.

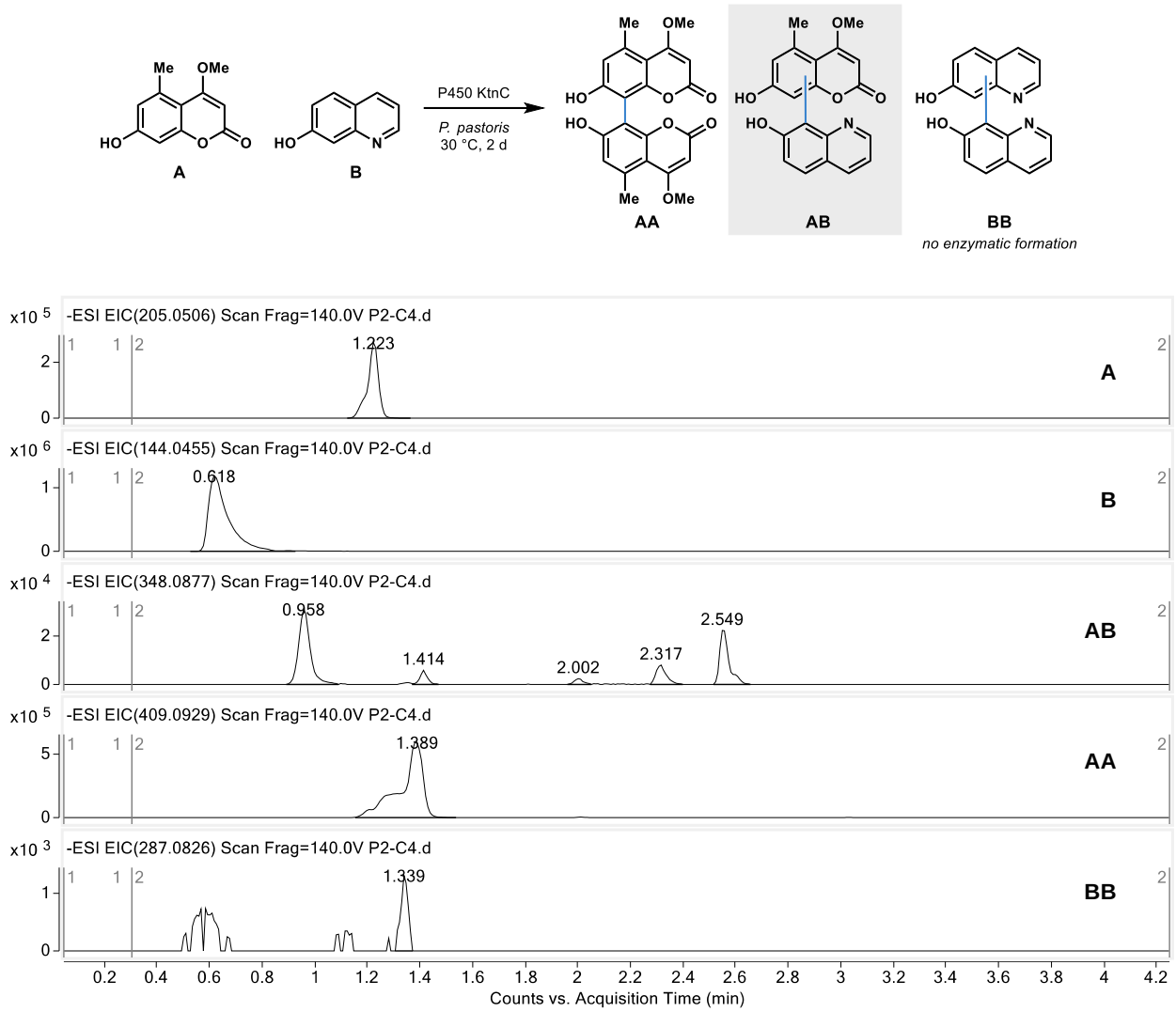


Figure 3.29. LC–MS traces for oxidative cross-coupling of 3.4 and 3.25 by KtnC.

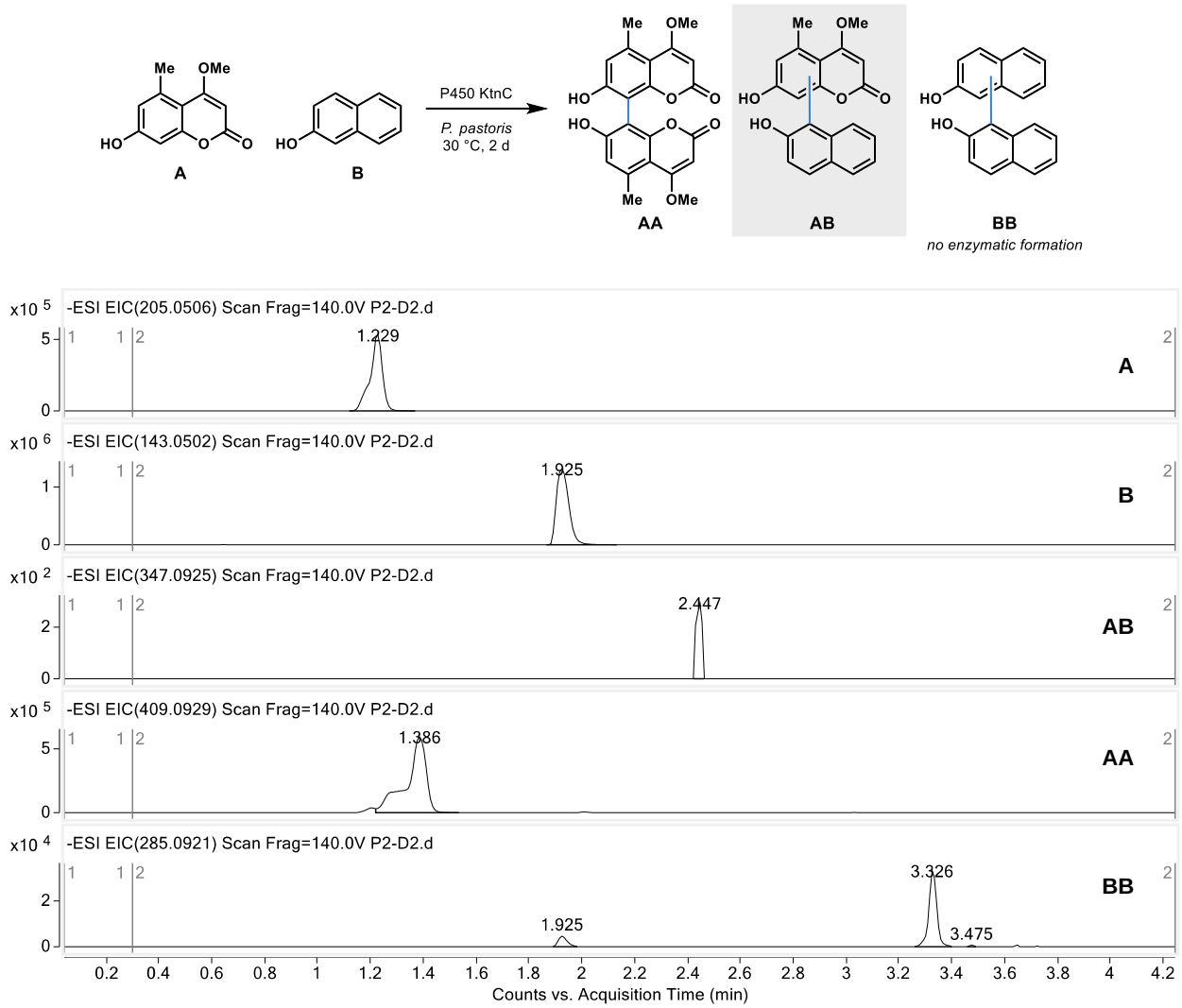
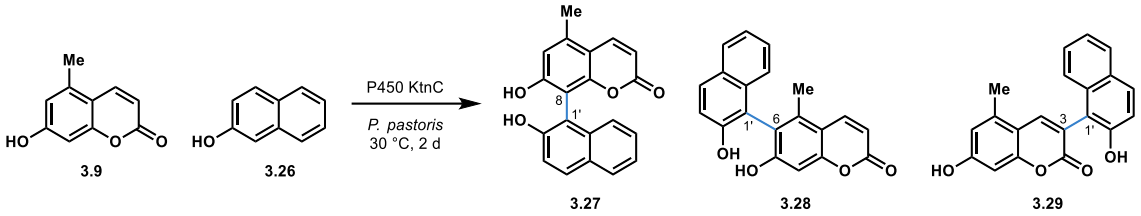


Figure 3.30. LC–MS traces for oxidative cross-coupling of 3.4 and 3.26 by KtnC.



Mixture of cross-coupled product isomers formed in biocatalytic reaction:



Biocatalytic reaction spiked with authentic product standards of three major isomers:

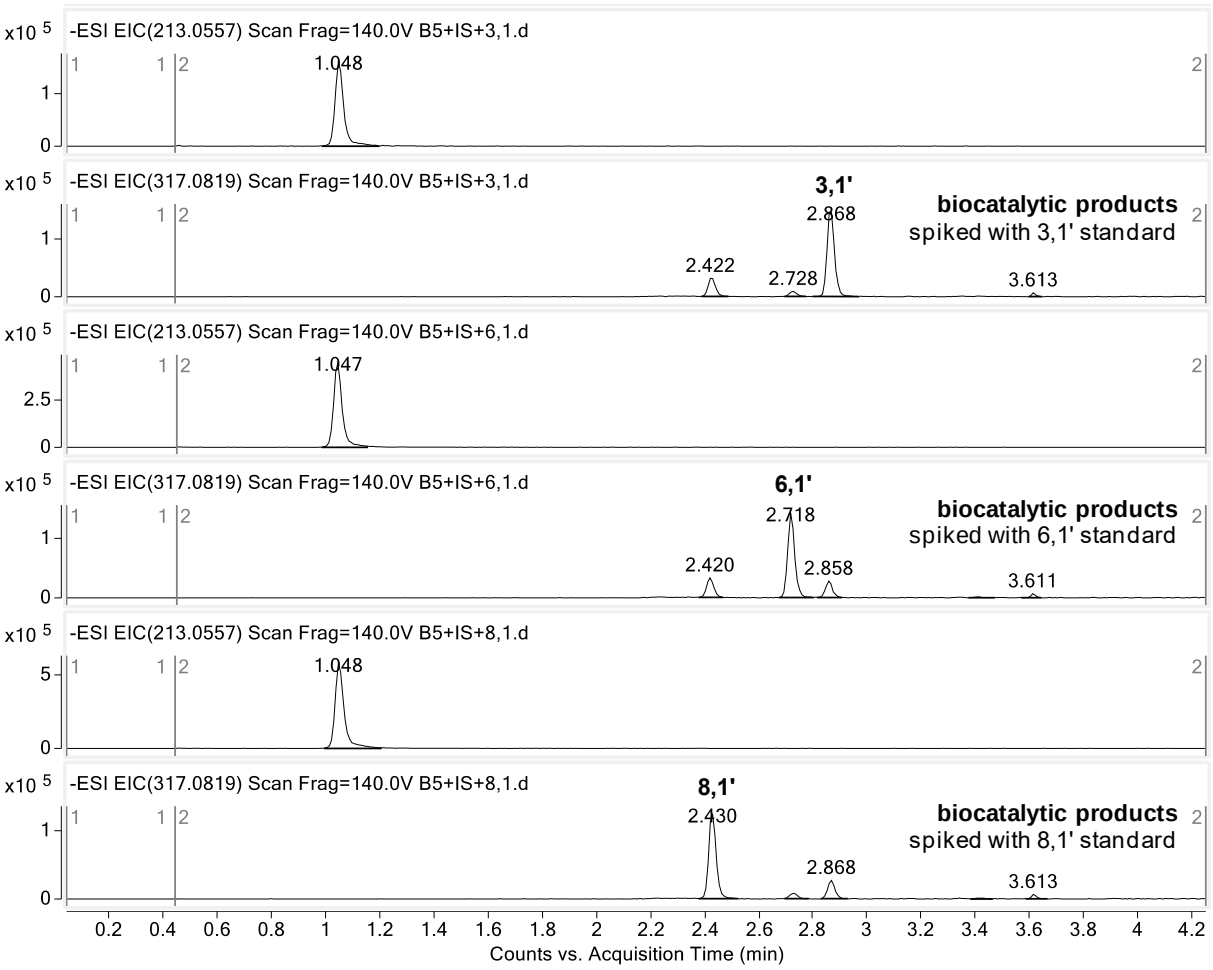


Figure 3.31. Site-selectivity in oxidative cross-coupling of 3.9 and 3.26 by KtnC. Extracted ion chromatograms include cross-coupled products (with $[M-H]^-$ of 317.0819) and internal standard (with $[M-H]^-$ of 213.0559).

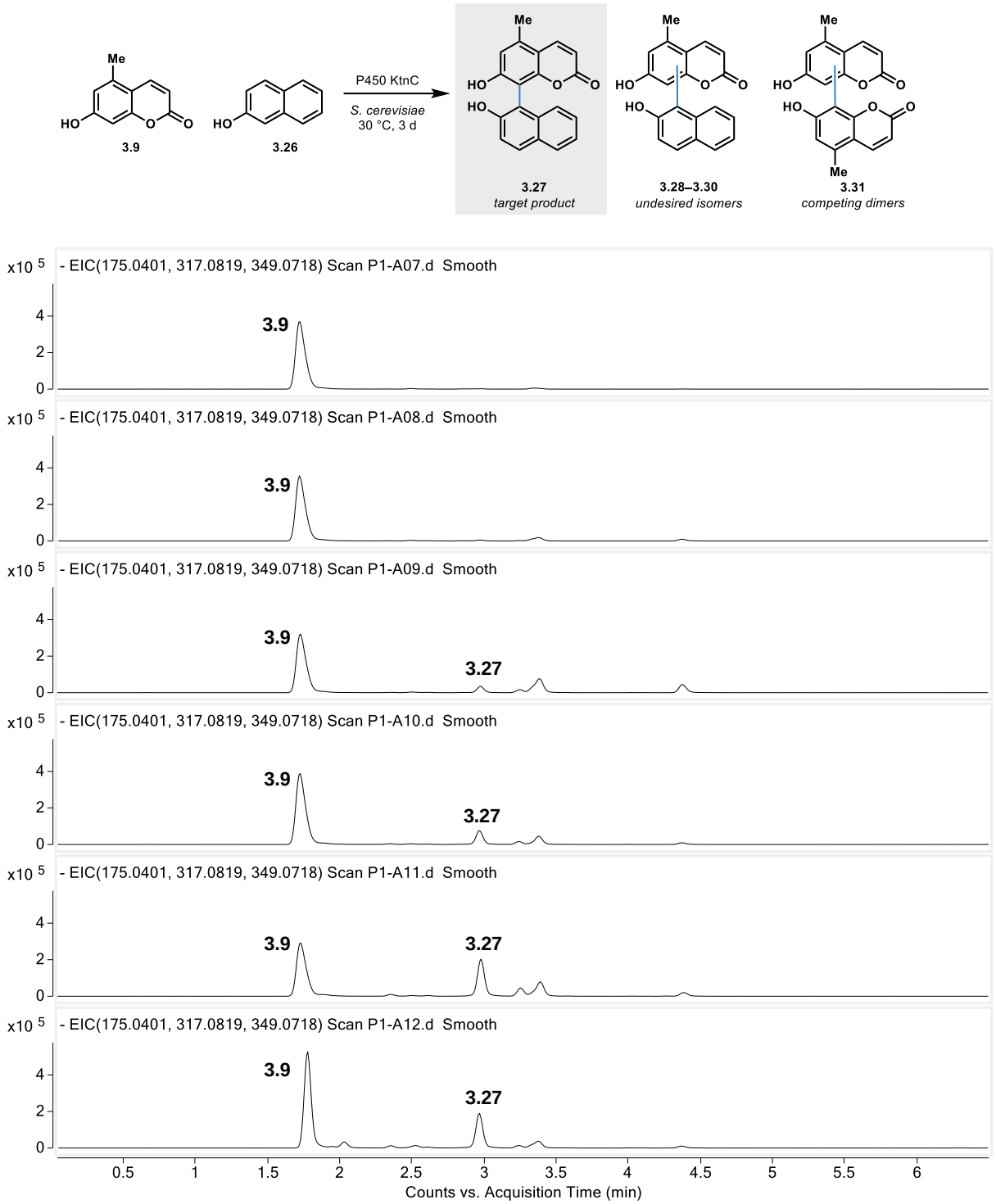


Figure 3.32. Reactivity in cross-coupling over evolution progress. Extracted ion chromatograms include coumarin substrate **3.9** (with $[M-H]^-$ of 175.0401), cross-coupled products **3.27**–**3.30** (with $[M-H]^-$ of 317.0819), and coumarin dimers **3.31** (with $[M-H]^-$ of 349.0718).

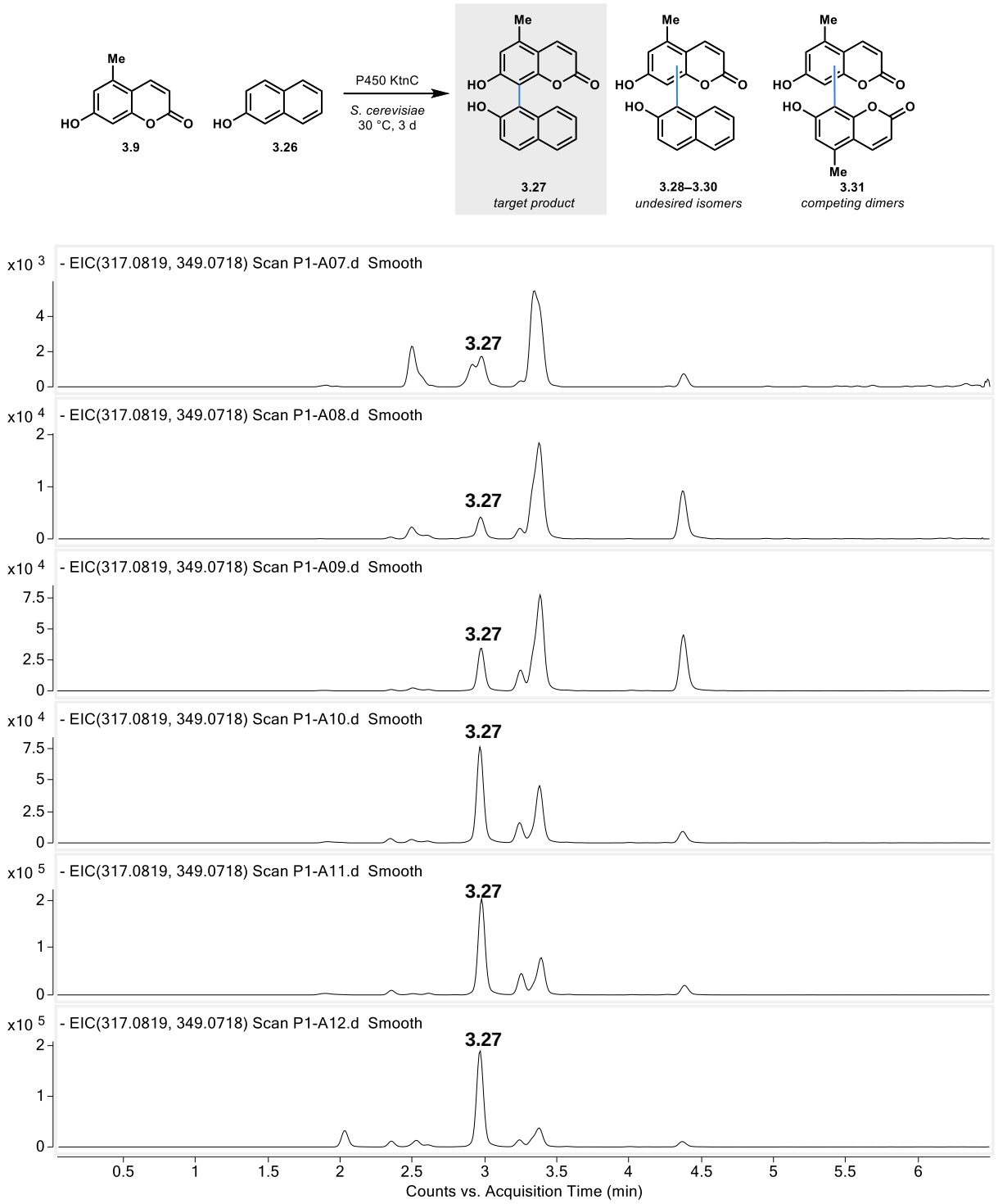


Figure 3.33. Chemoselectivity of cross-coupling over evolution progress. Extracted ion chromatograms include cross-coupled products **3.27**–**3.30** (with $[M-H]^-$ of 317.0819) and coumarin dimers **3.31** (with $[M-H]^-$ of 349.0718). The major coumarin dimer (of four total isomers) is highlighted in each chromatogram.

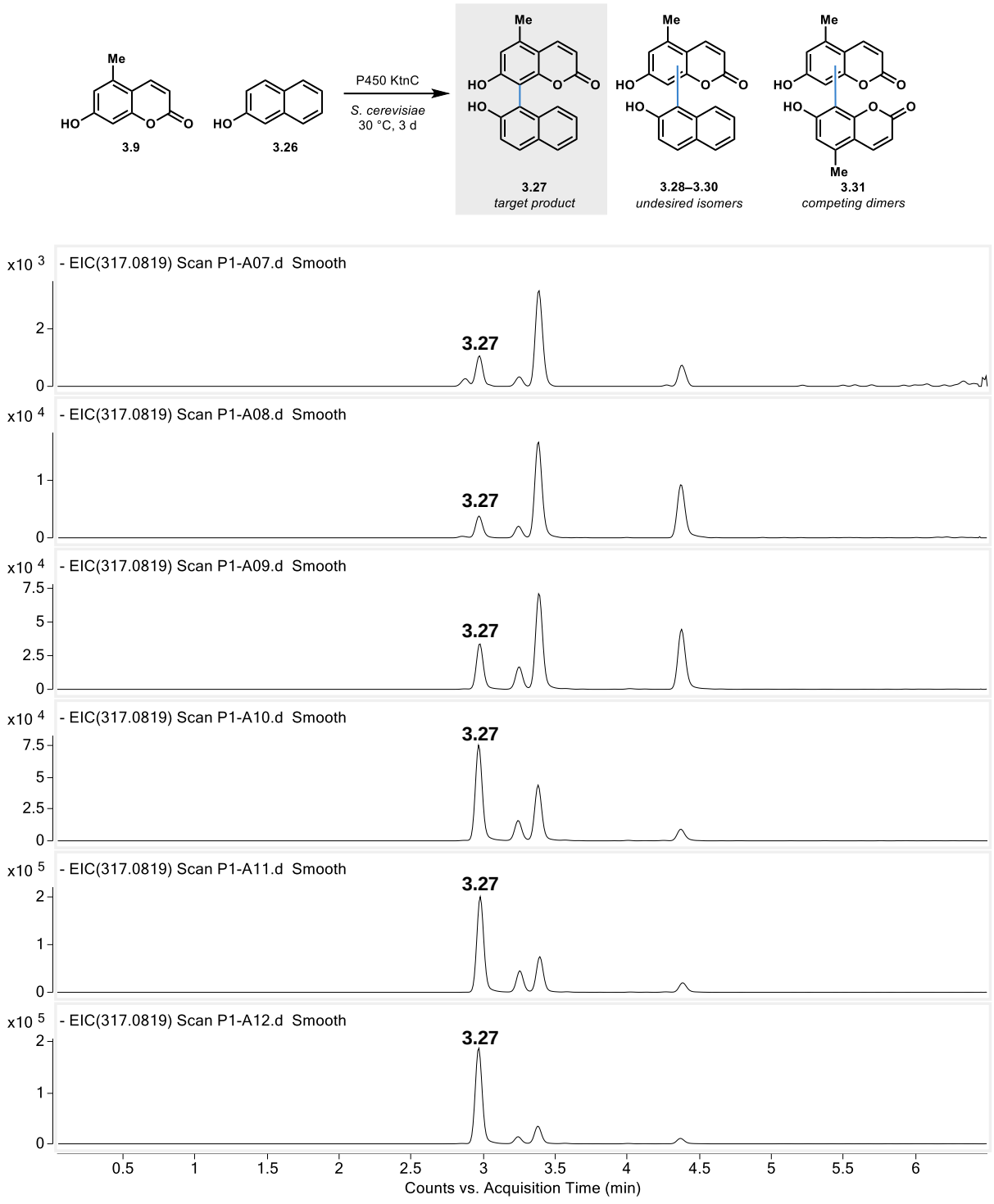
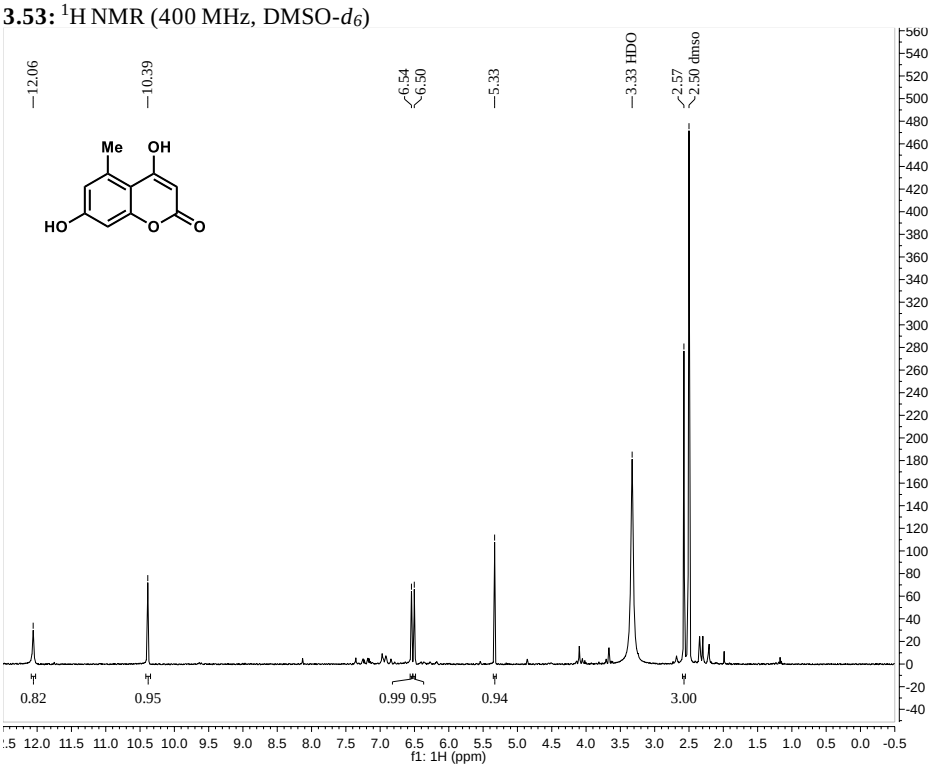


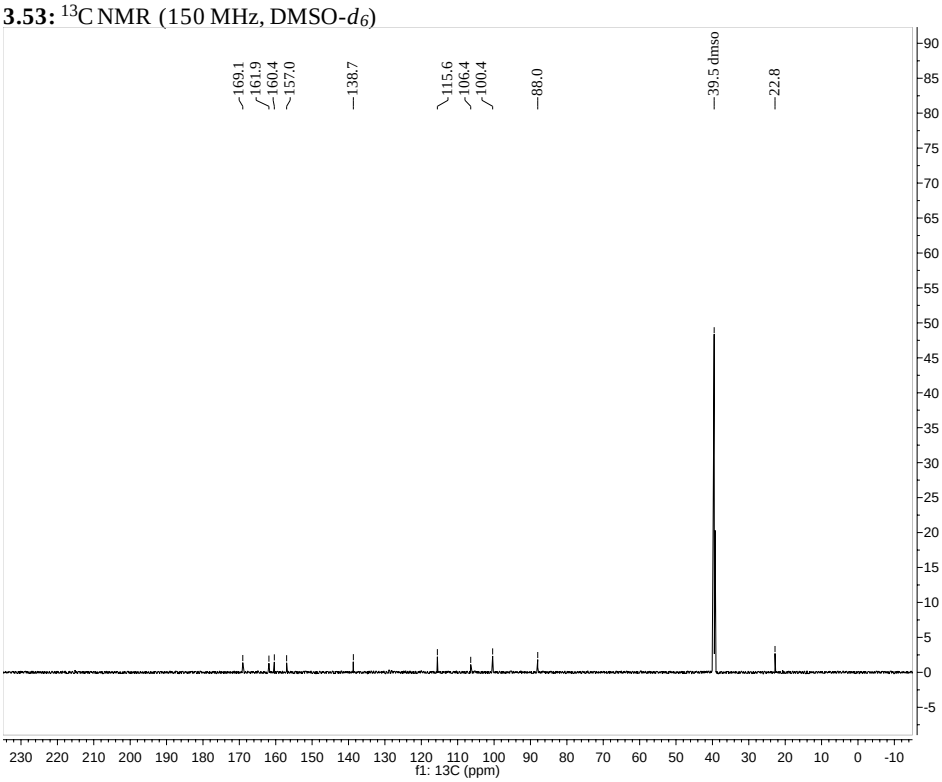
Figure 3.34. Site-selectivity of cross-coupling catalyzed by P450 variants over evolution progress. Extracted ion chromatograms include cross-coupled products **3.27**–**3.30** (with $[M-H]^-$ of 317.0819).

^{1}H NMR and ^{13}C NMR spectra of compounds. ^{1}H and ^{13}C NMR spectra were obtained at room temperature (25 °C) on Varian 400 MHz or Varian 600 MHz spectrometers.

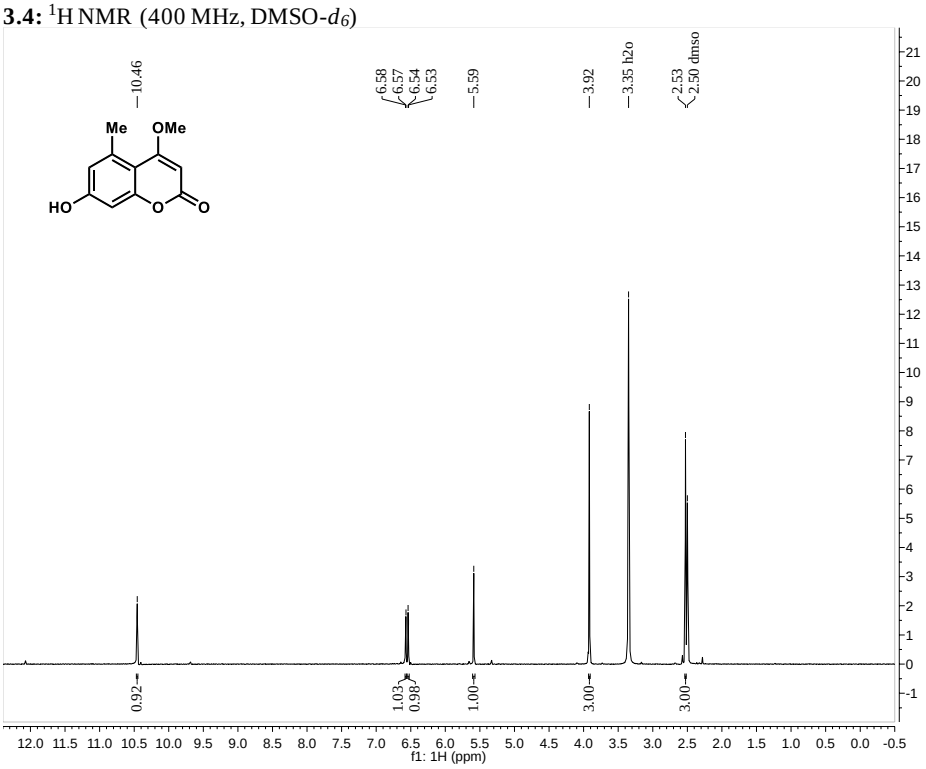
3.53: ^{1}H NMR (400 MHz, DMSO-d_6)



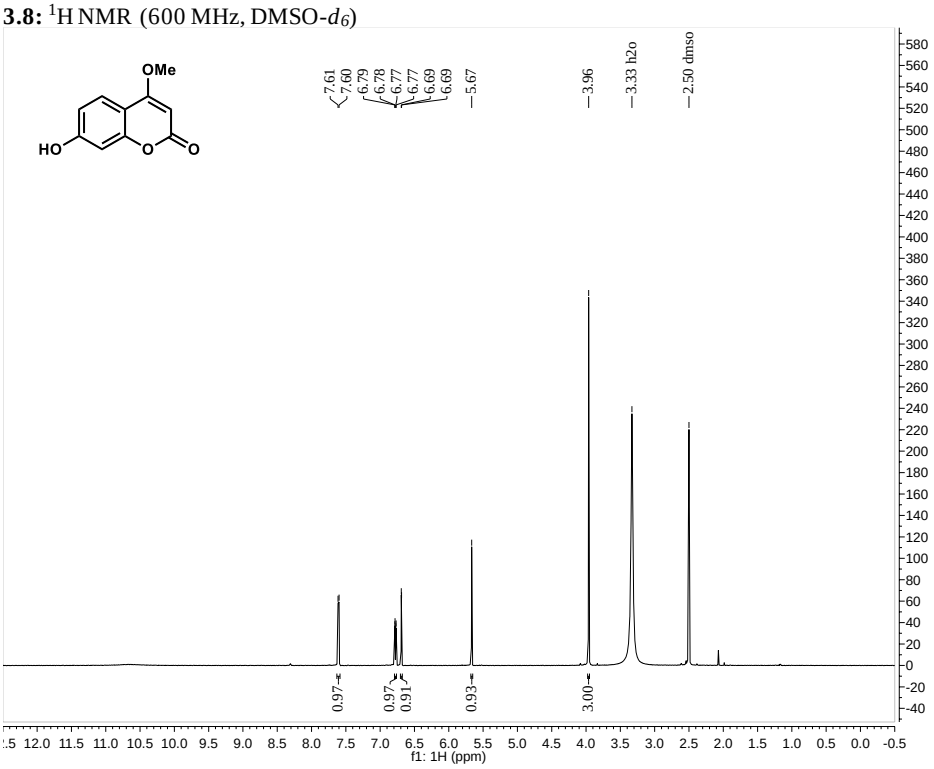
3.53: ^{13}C NMR (150 MHz, DMSO-d_6)



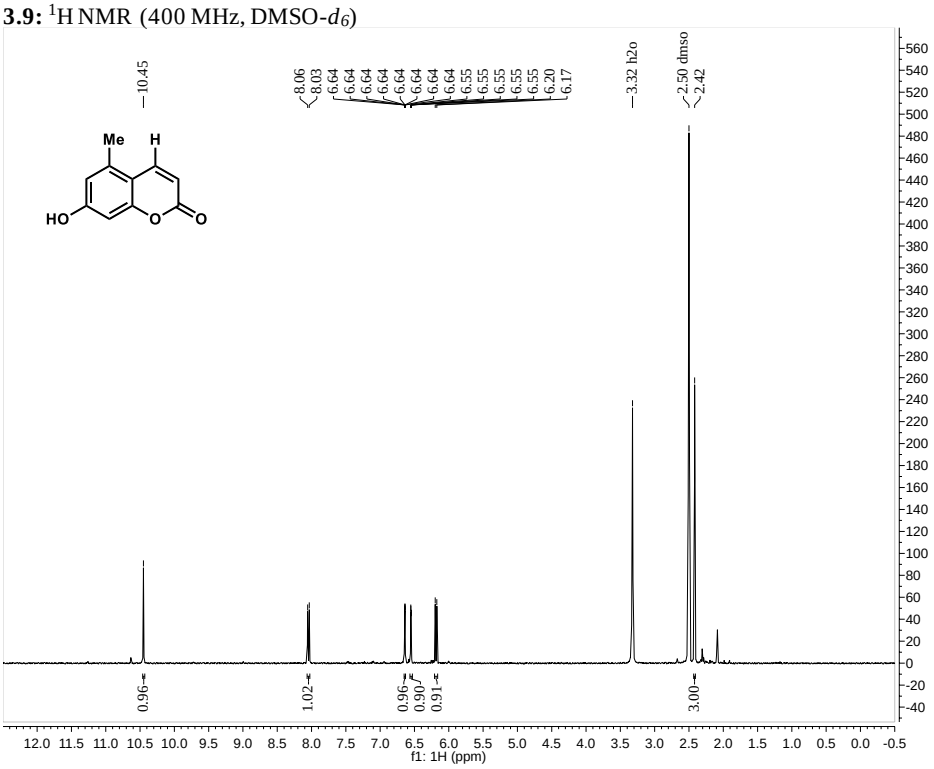
3.4: ^{1}H NMR (400 MHz, DMSO-d_6)



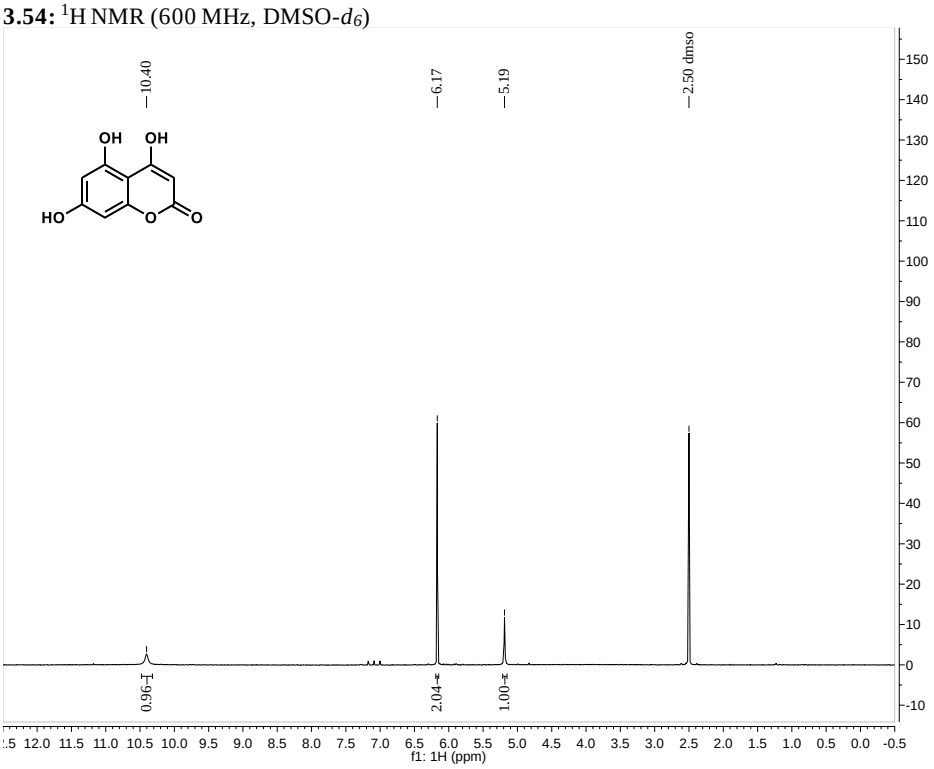
3.8: ^{1}H NMR (600 MHz, DMSO-d_6)



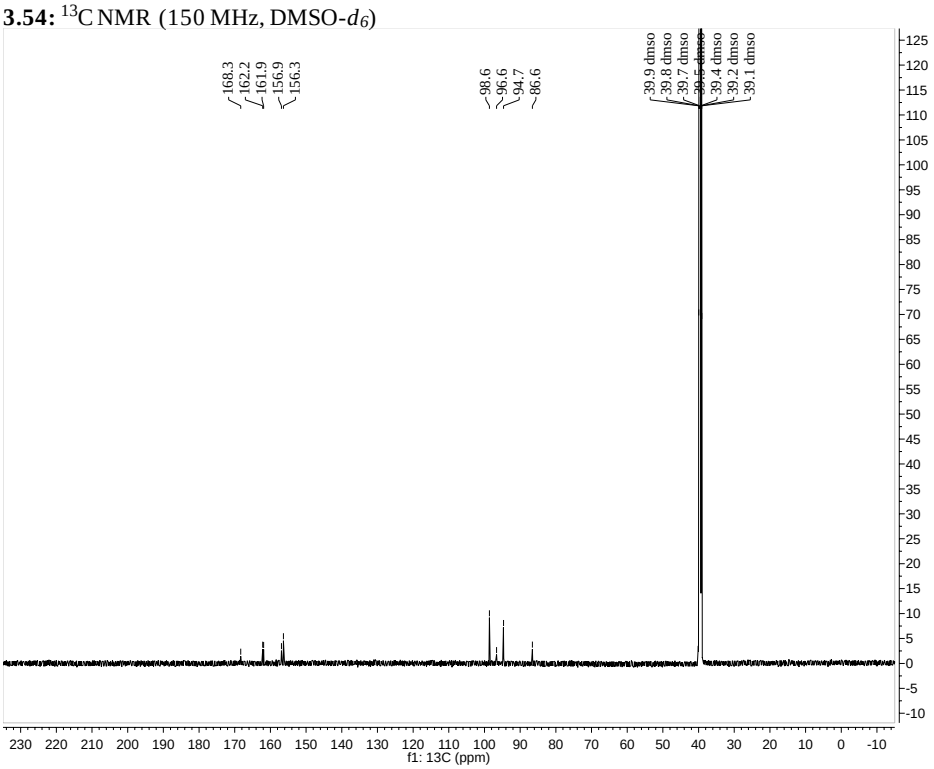
3.9: ^{1}H NMR (400 MHz, DMSO-d_6)



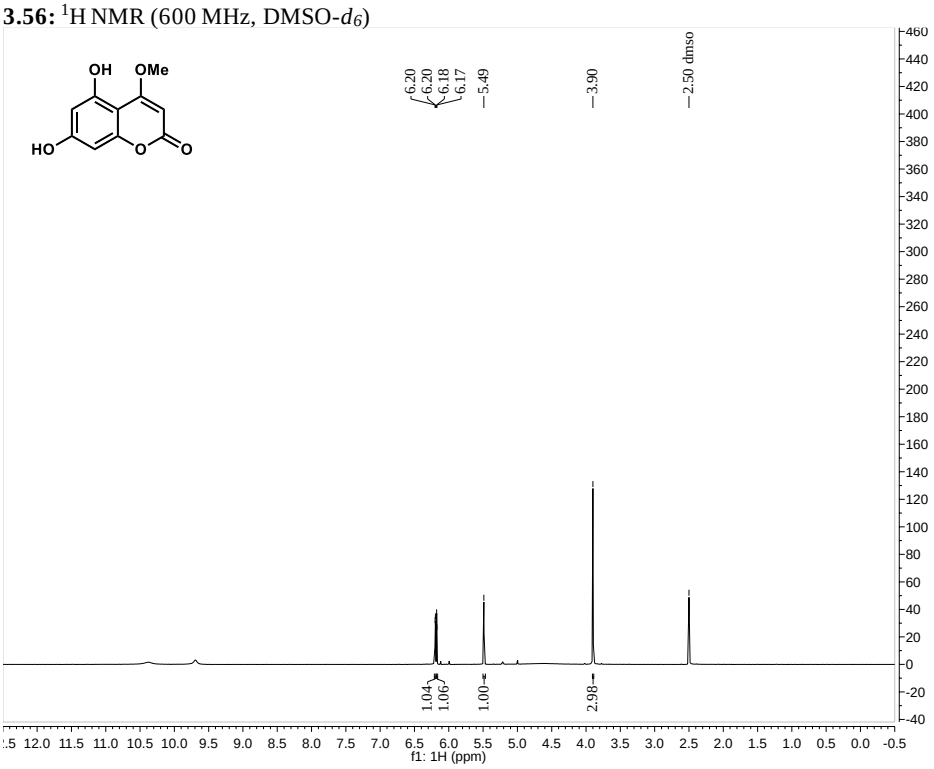
3.54: ^{1}H NMR (600 MHz, DMSO-d_6)



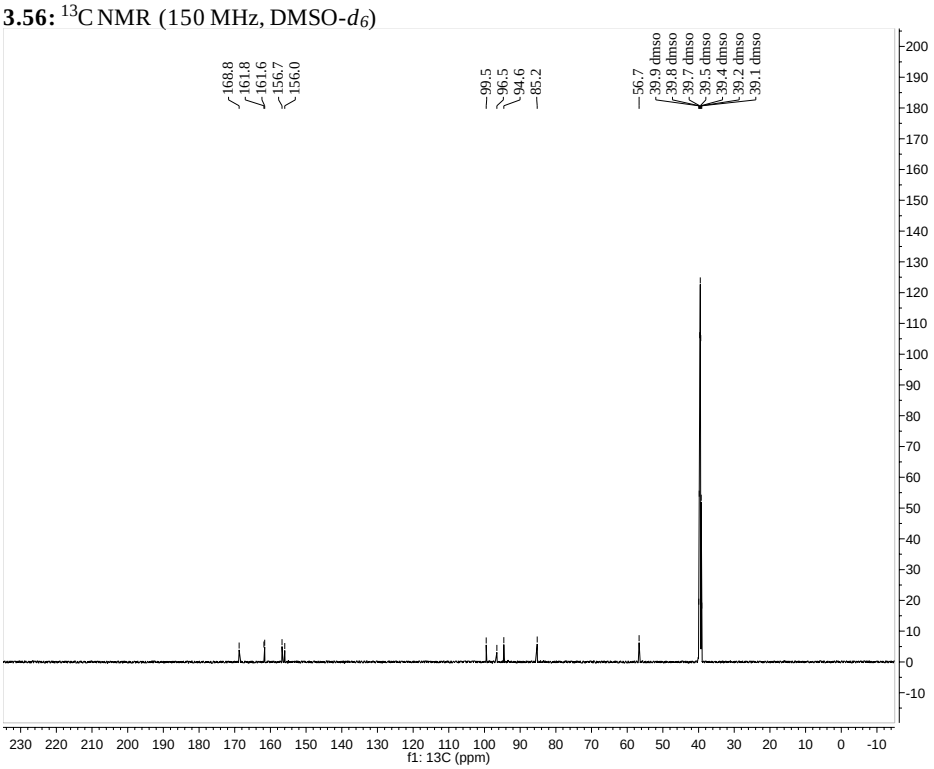
3.54: ^{13}C NMR (150 MHz, DMSO-d_6)



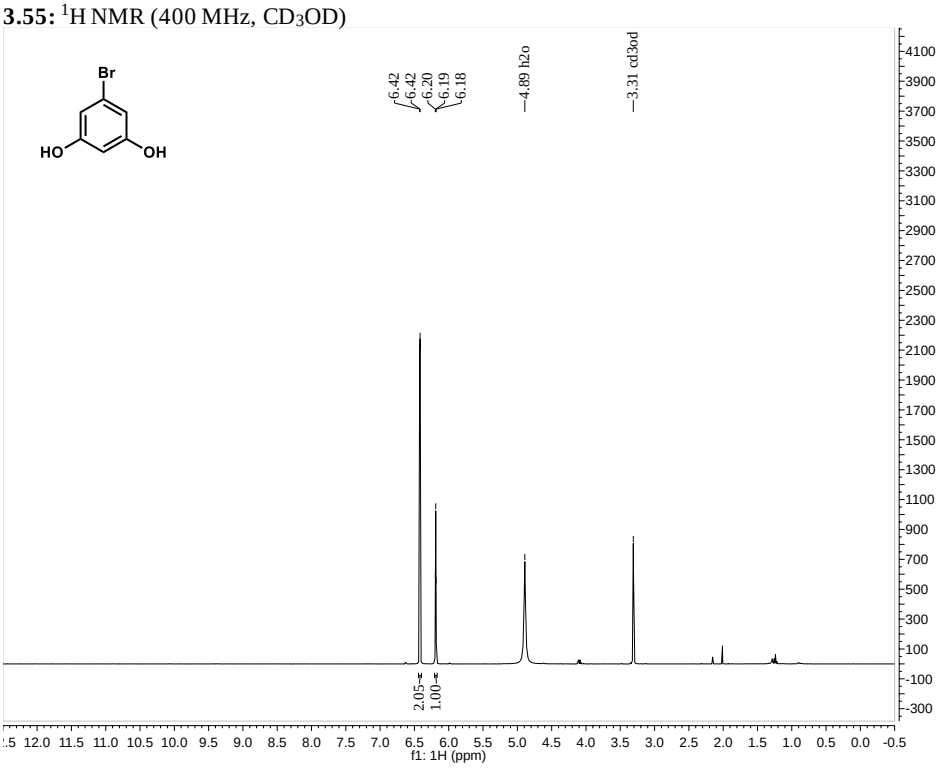
3.56: ^{1}H NMR (600 MHz, DMSO-d_6)



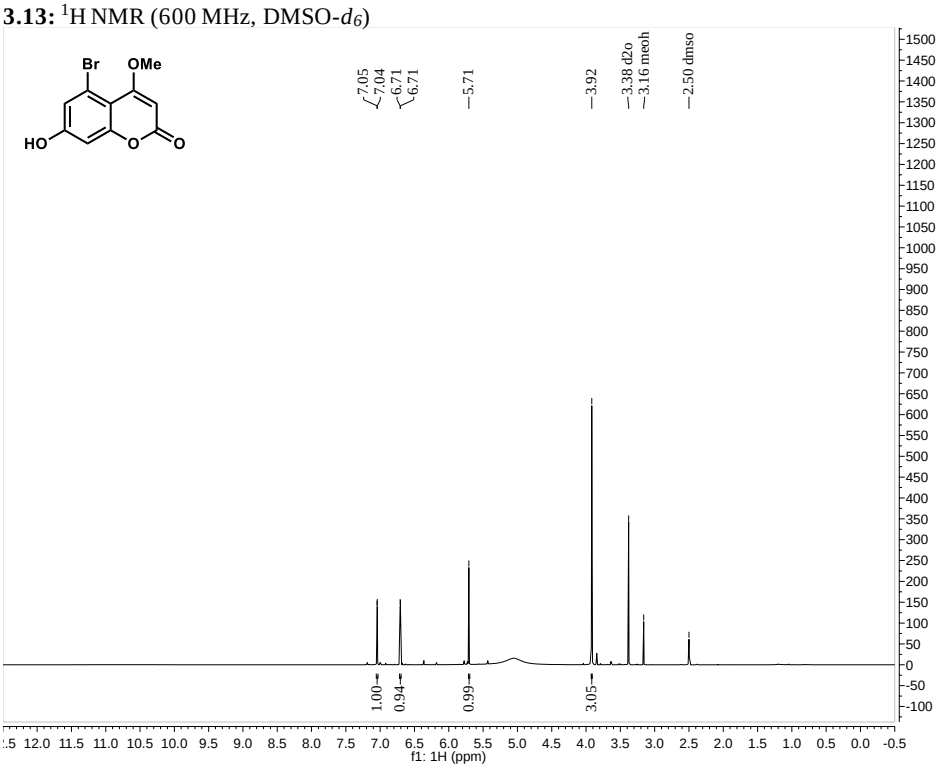
3.56: ^{13}C NMR (150 MHz, DMSO-d_6)



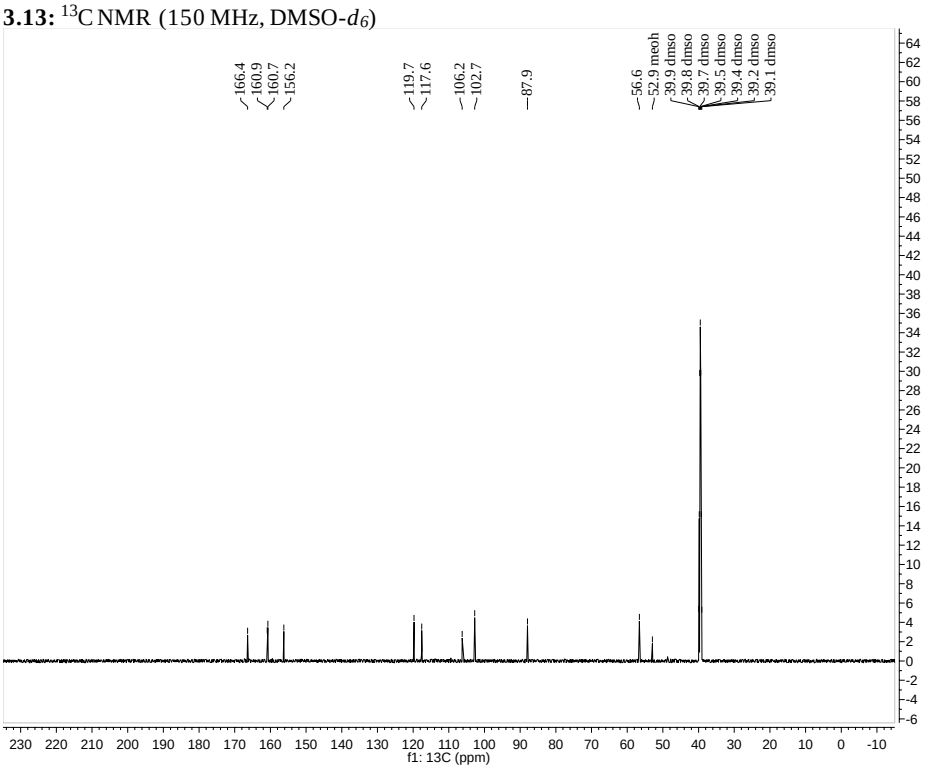
3.55: ^{1}H NMR (400 MHz, CD_3OD)



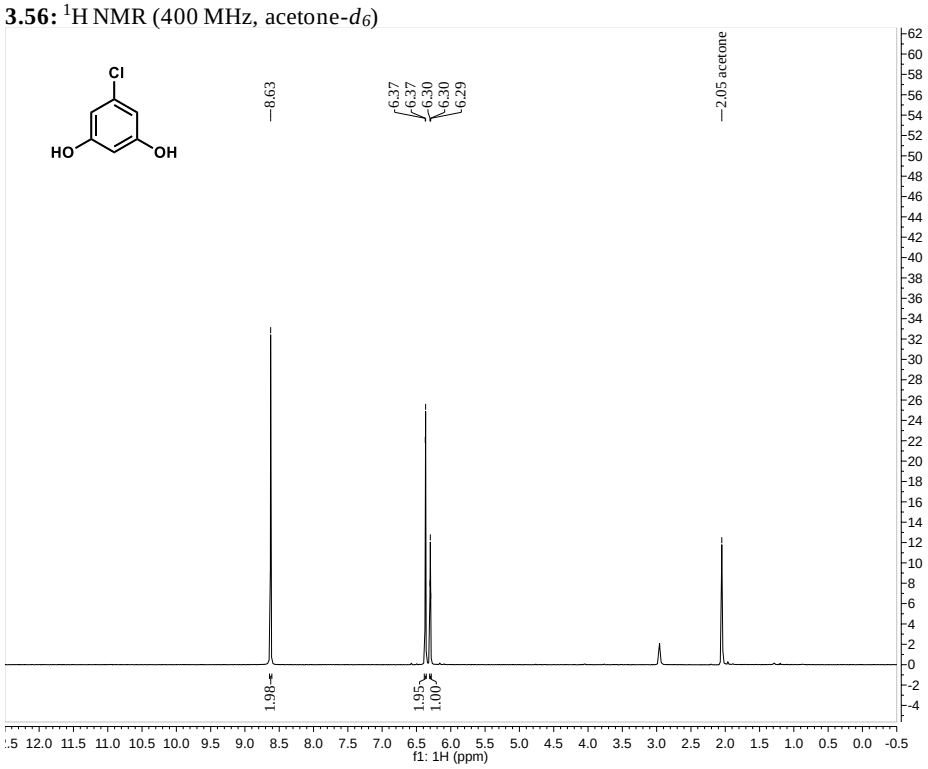
3.13: ^{1}H NMR (600 MHz, DMSO-d_6)



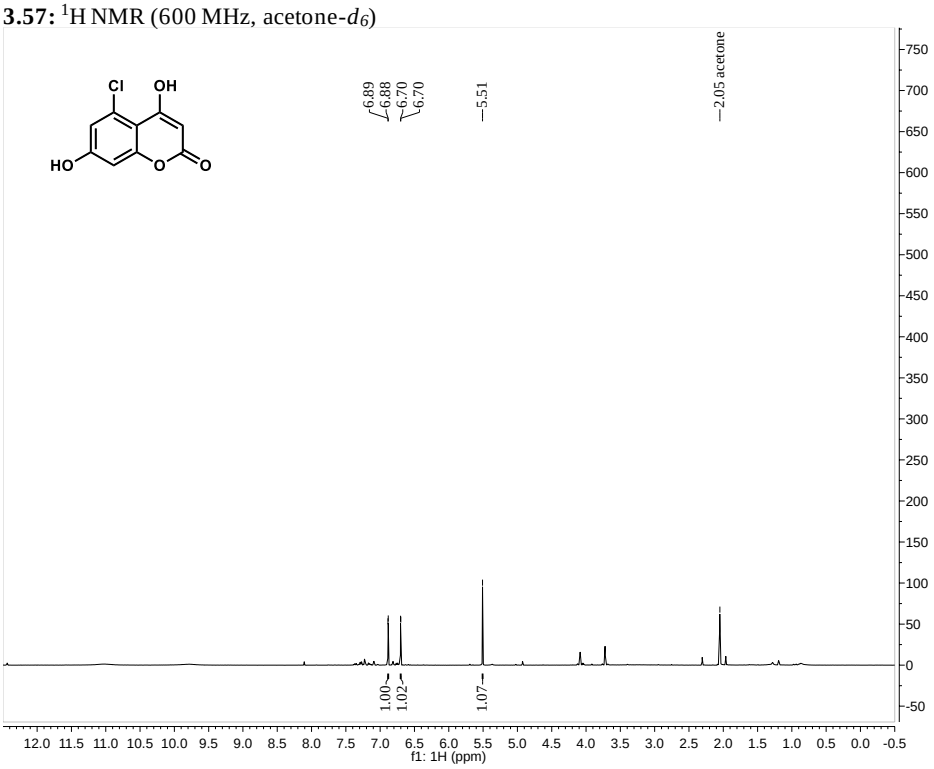
3.13: ^{13}C NMR (150 MHz, DMSO-d_6)



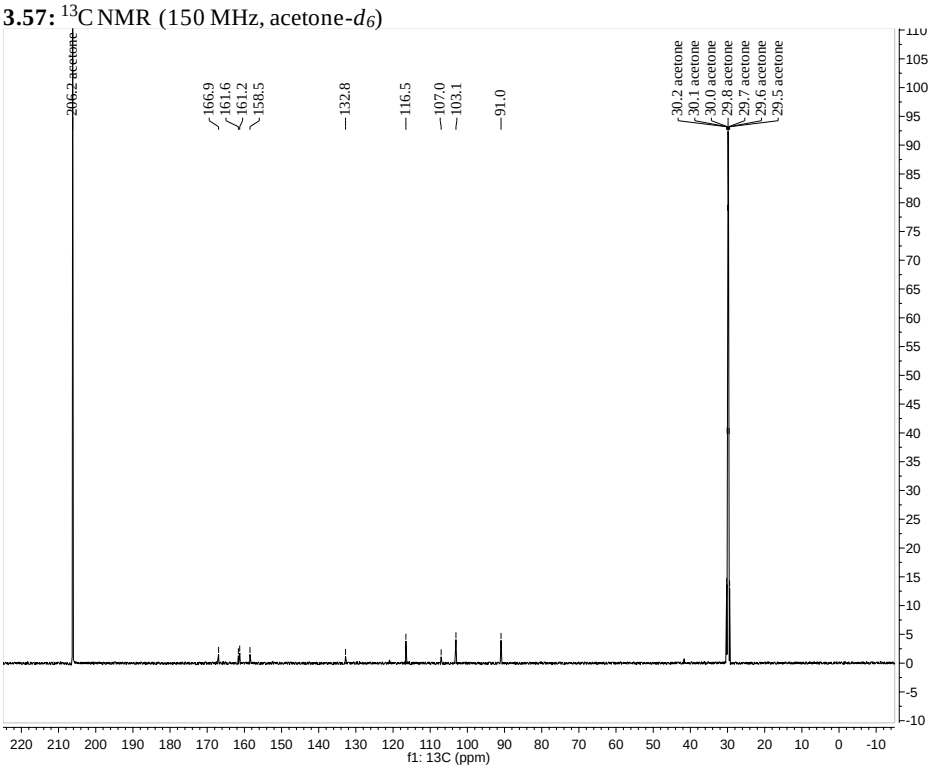
3.56: ^{1}H NMR (400 MHz, acetone-d_6)



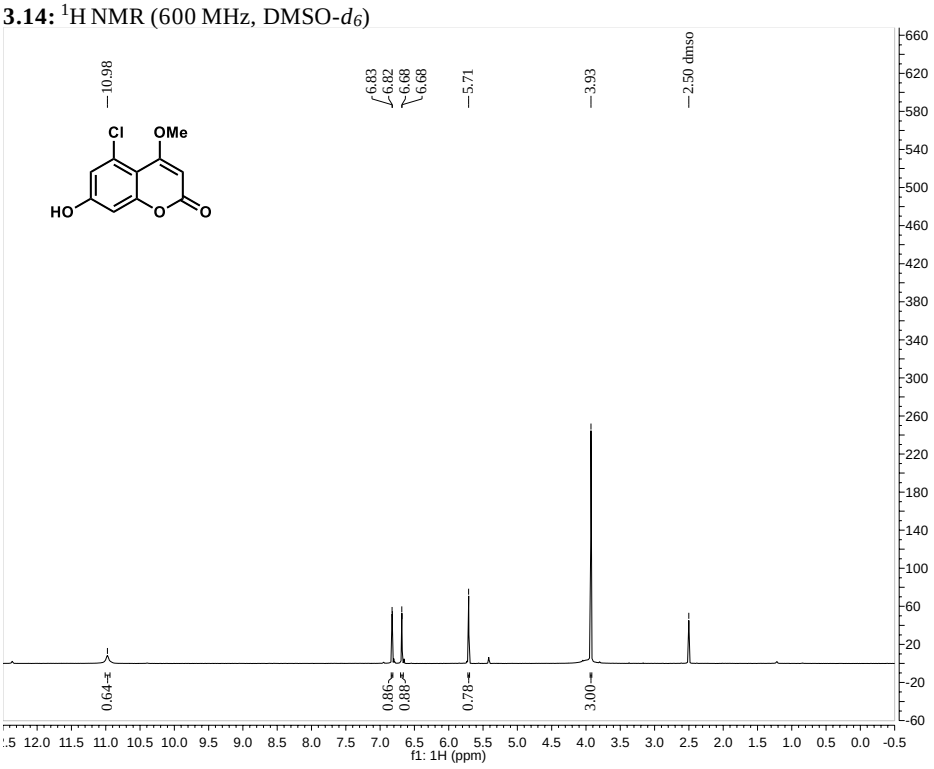
3.57: ^{1}H NMR (600 MHz, acetone-d_6)



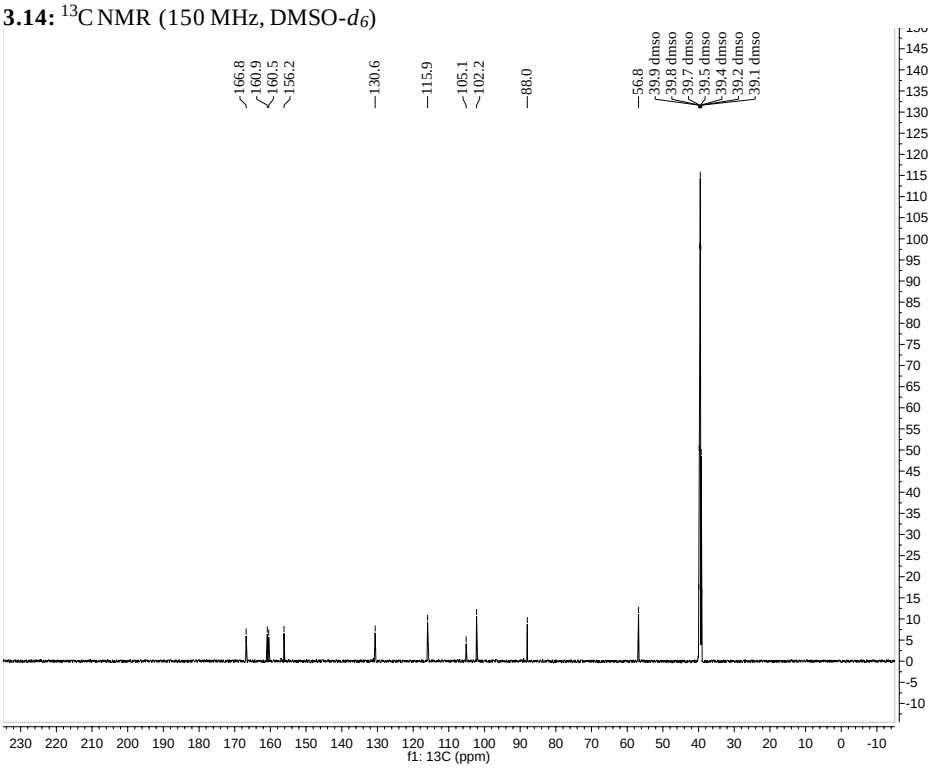
3.57: ^{13}C NMR (150 MHz, acetone-d_6)



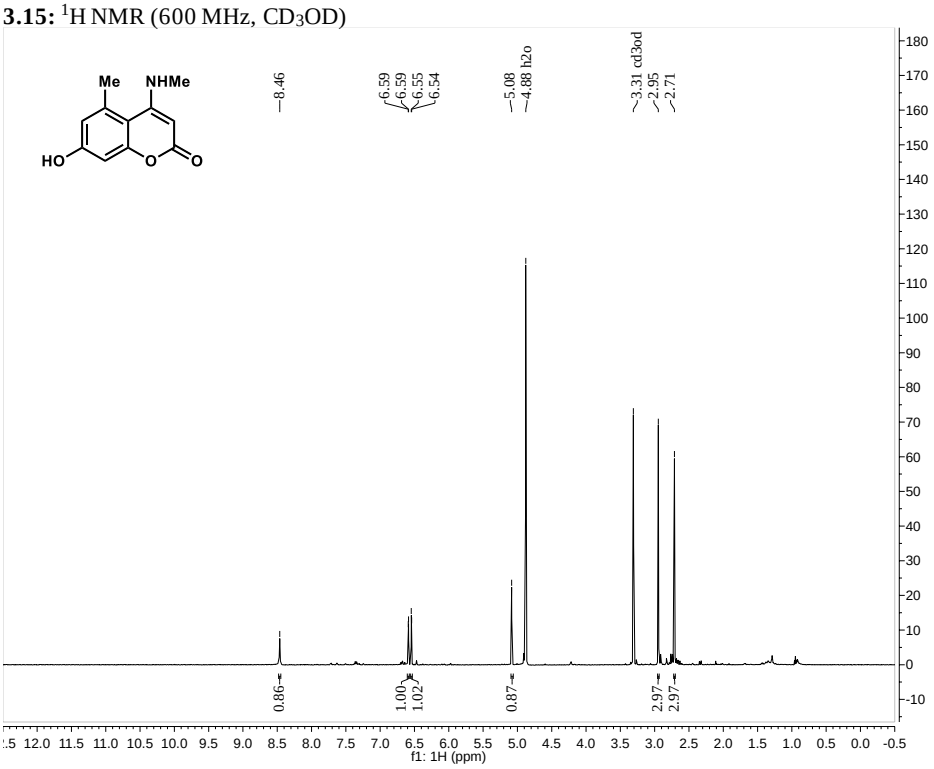
3.14: ^{1}H NMR (600 MHz, DMSO-d_6)



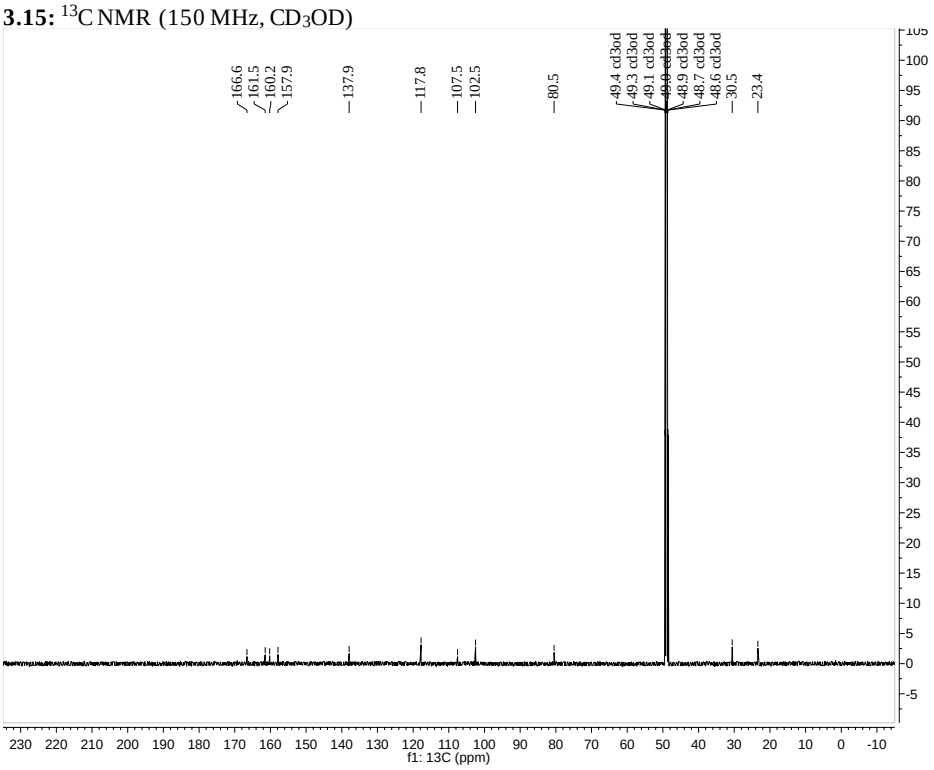
3.14: ^{13}C NMR (150 MHz, DMSO-d_6)



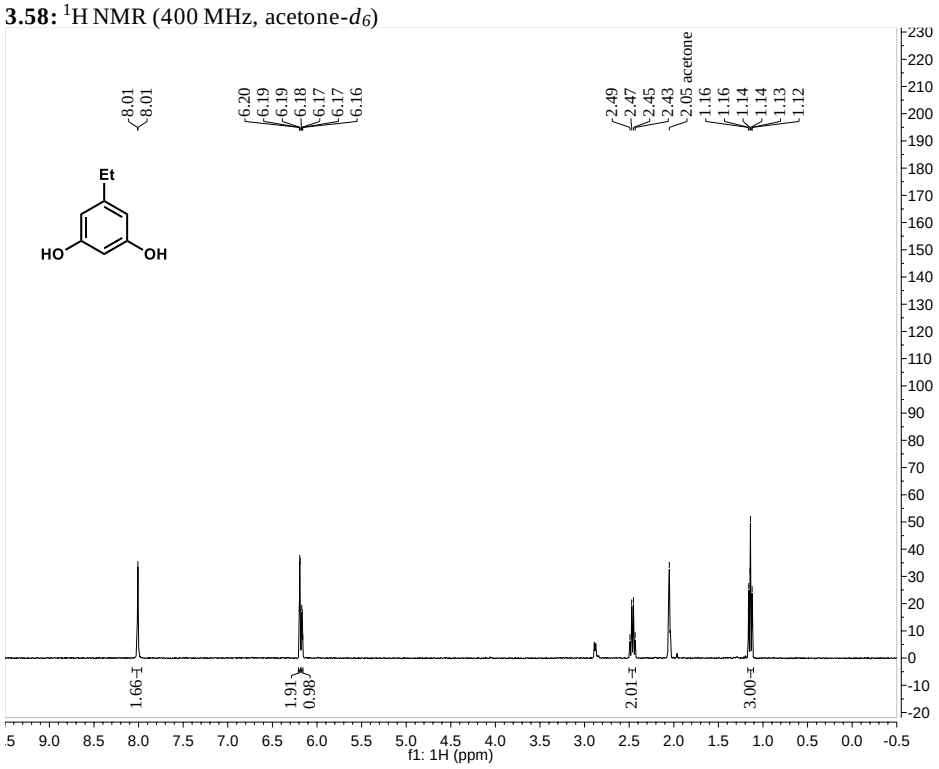
3.15: ^{1}H NMR (600 MHz, CD_3OD)



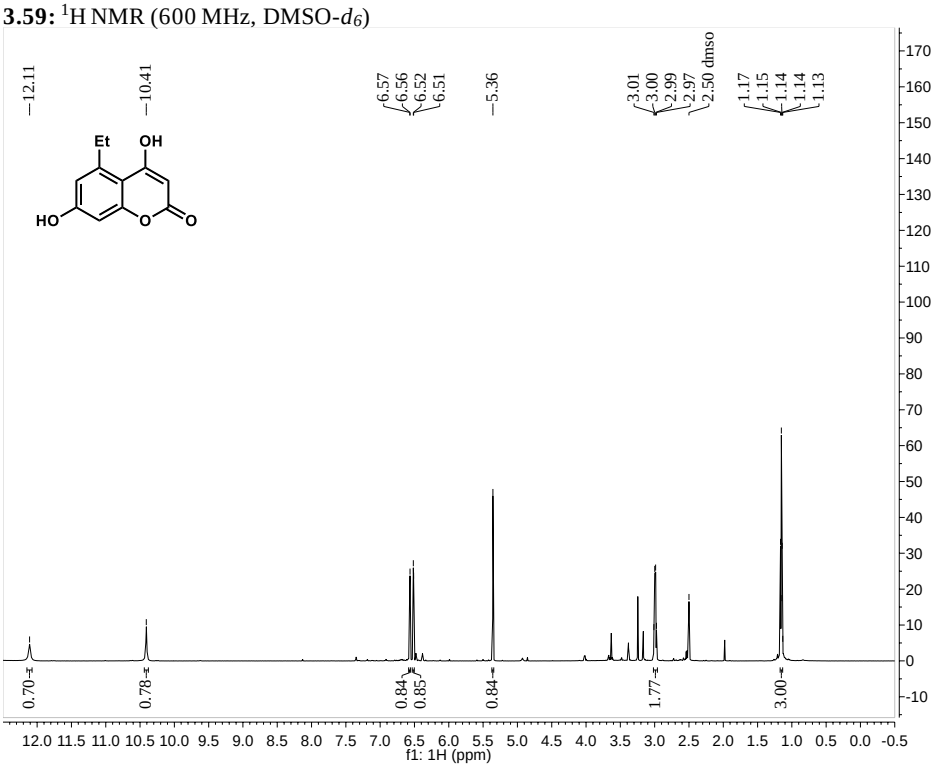
3.15: ^{13}C NMR (150 MHz, CD_3OD)



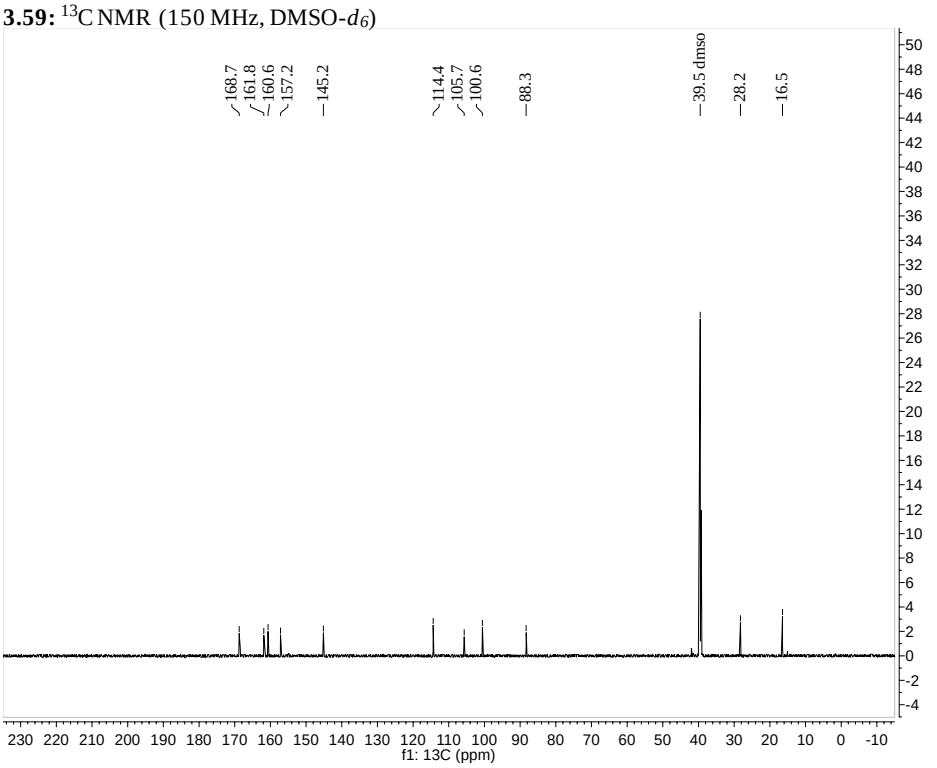
3.58: ^{1}H NMR (400 MHz, acetone-d_6)



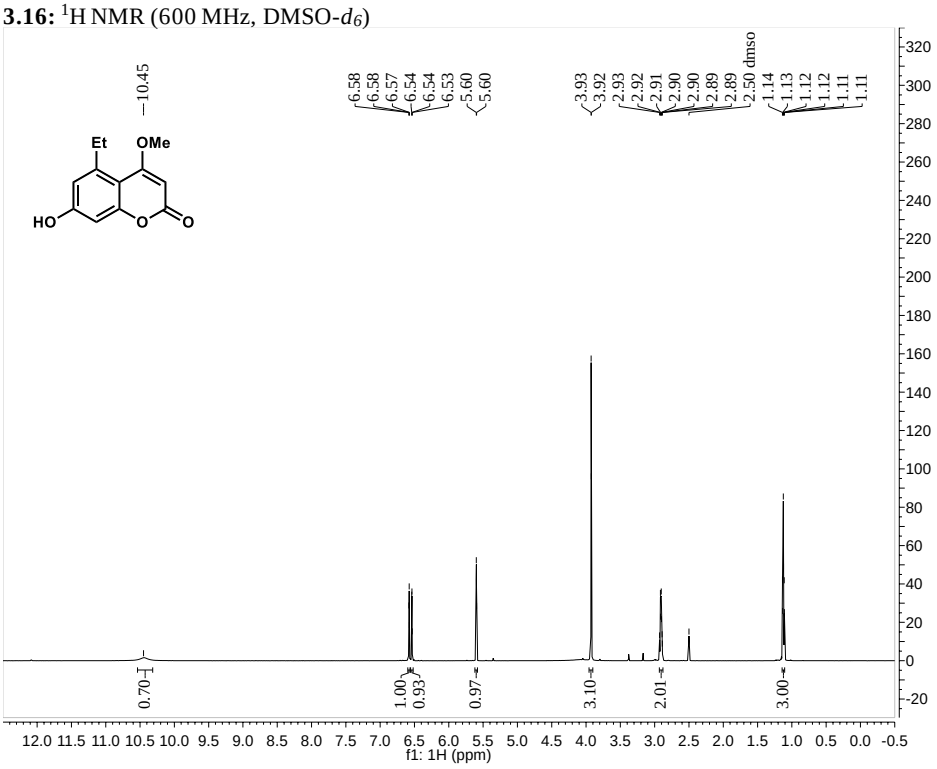
3.59: ^{1}H NMR (600 MHz, DMSO-d_6)



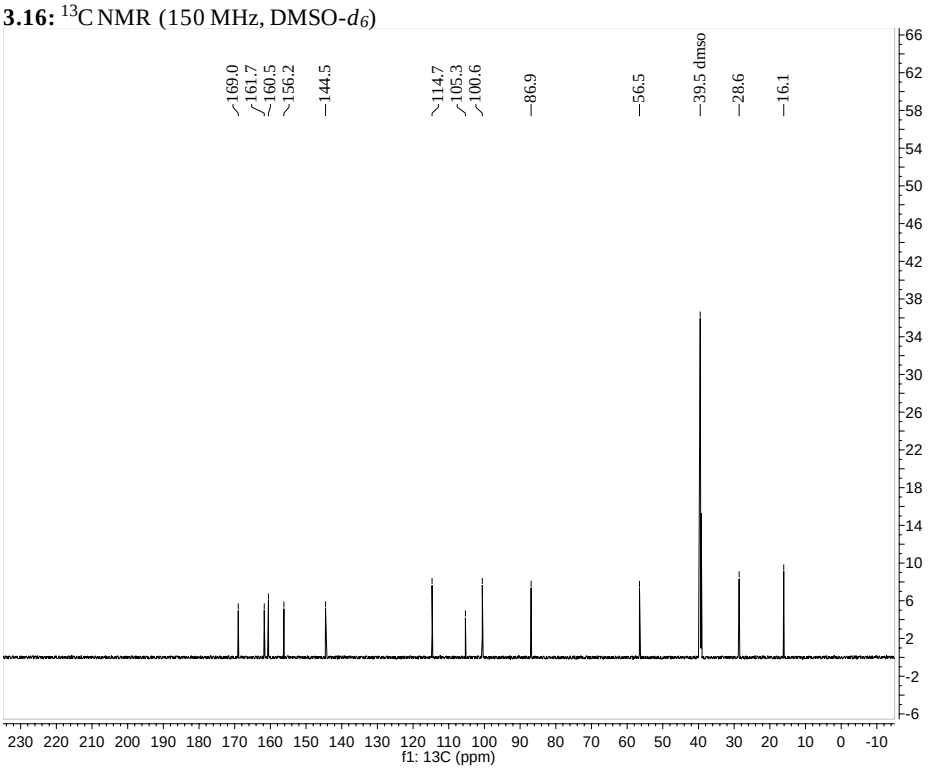
3.59: ^{13}C NMR (150 MHz, DMSO-d_6)



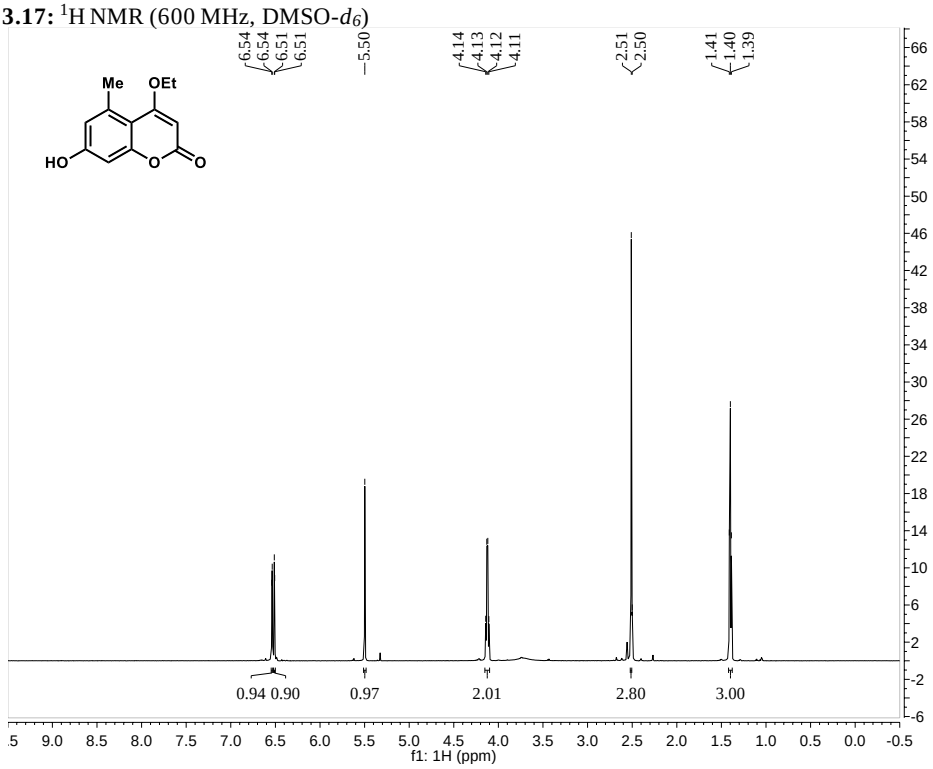
3.16: ^{1}H NMR (600 MHz, DMSO-d_6)



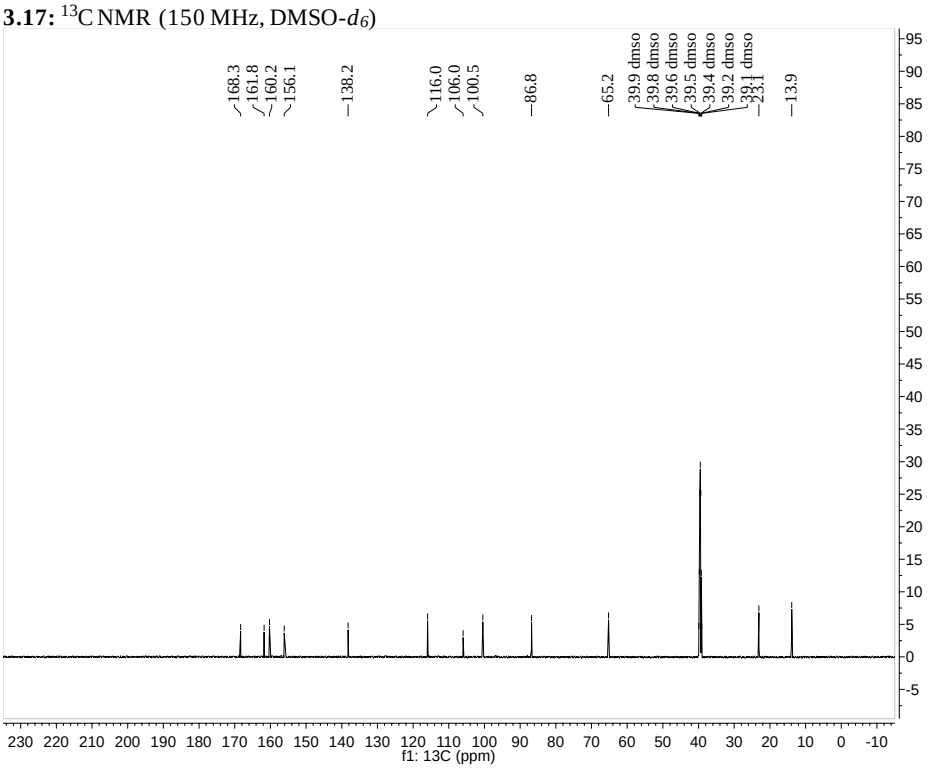
3.16: ^{13}C NMR (150 MHz, DMSO-d_6)



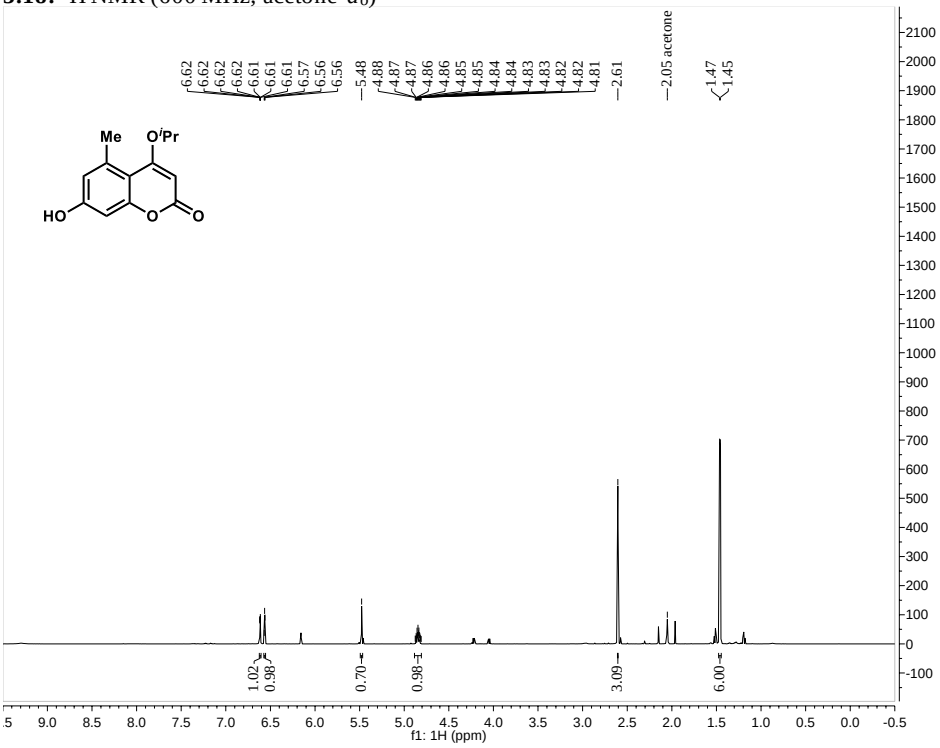
3.17: ^{1}H NMR (600 MHz, DMSO-d_6)



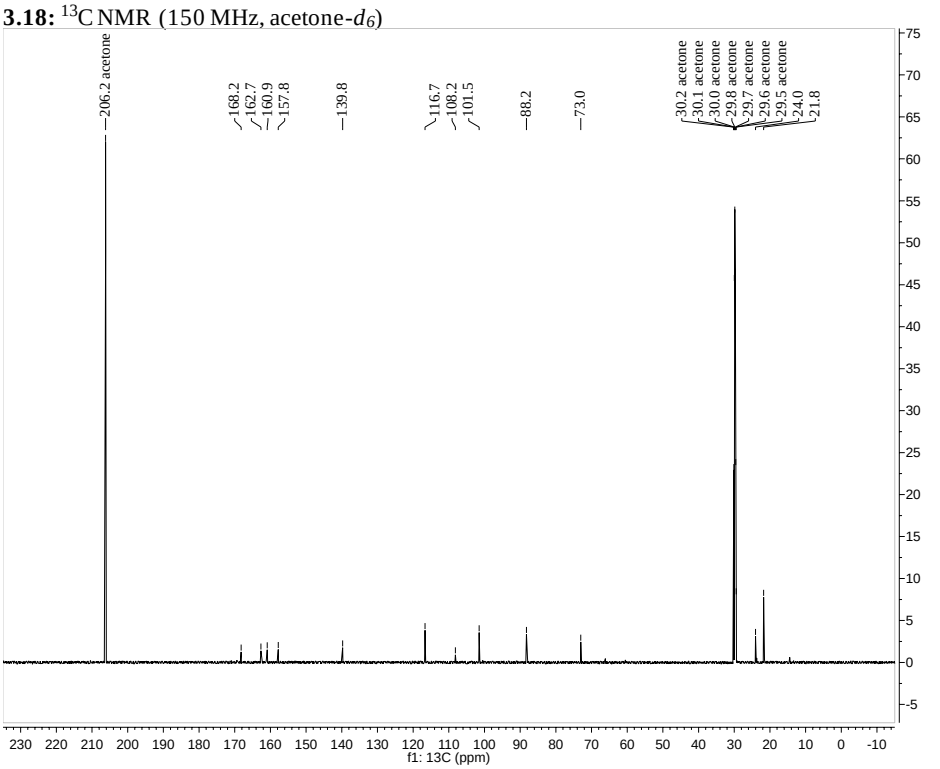
3.17: ^{13}C NMR (150 MHz, DMSO-d_6)



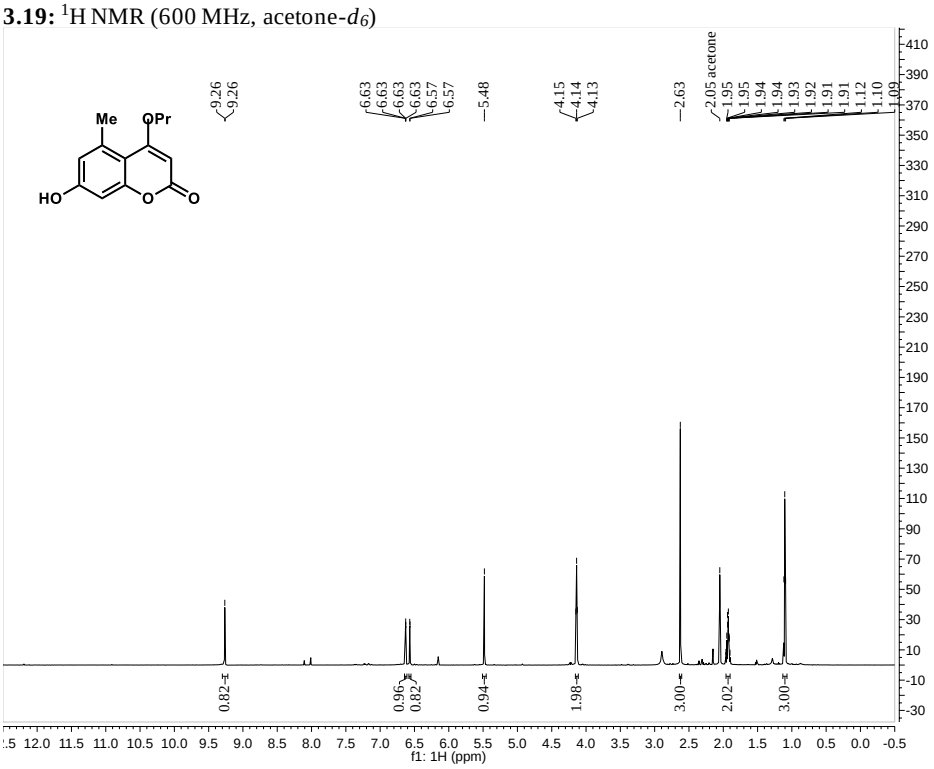
3.18: ^{1}H NMR (600 MHz, acetone-d_6)



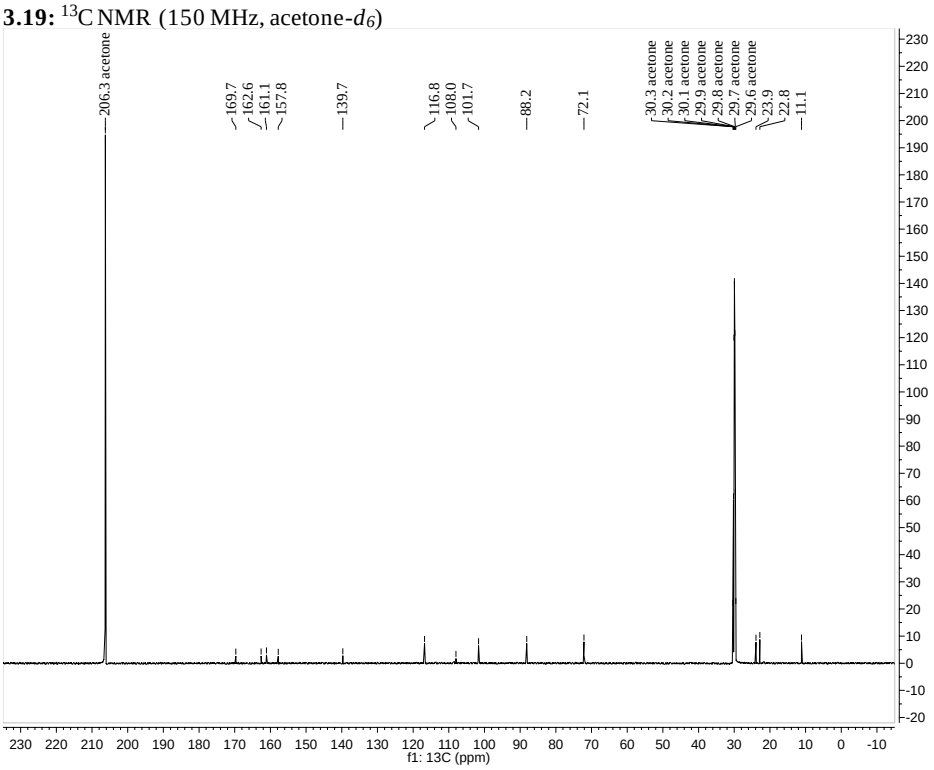
3.18: ^{13}C NMR (150 MHz, acetone-d_6)



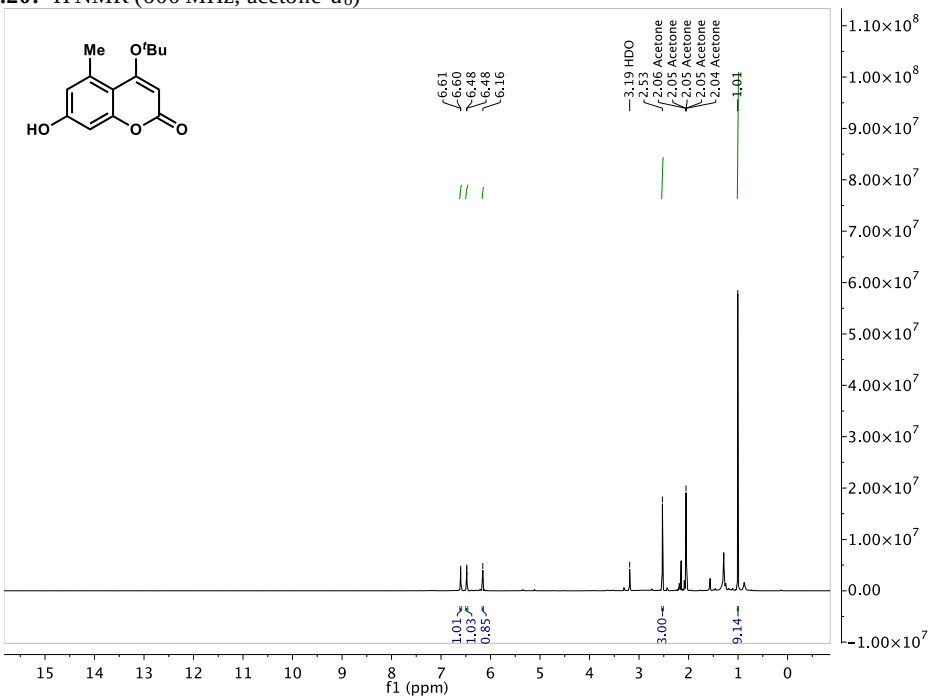
3.19: ^{1}H NMR (600 MHz, acetone-d_6)



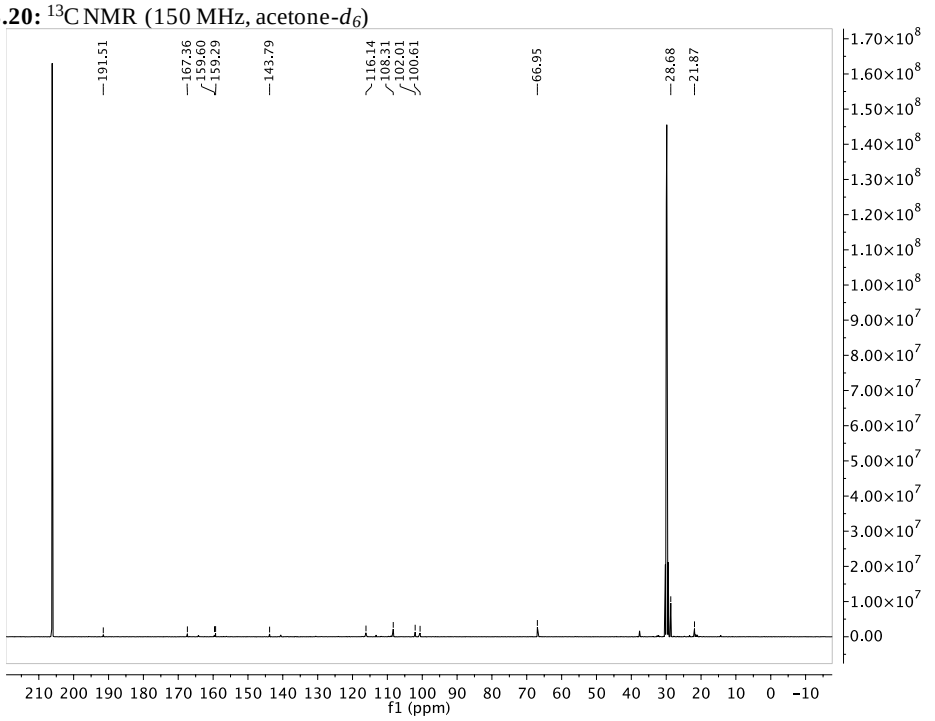
3.19: ^{13}C NMR (150 MHz, acetone-d_6)



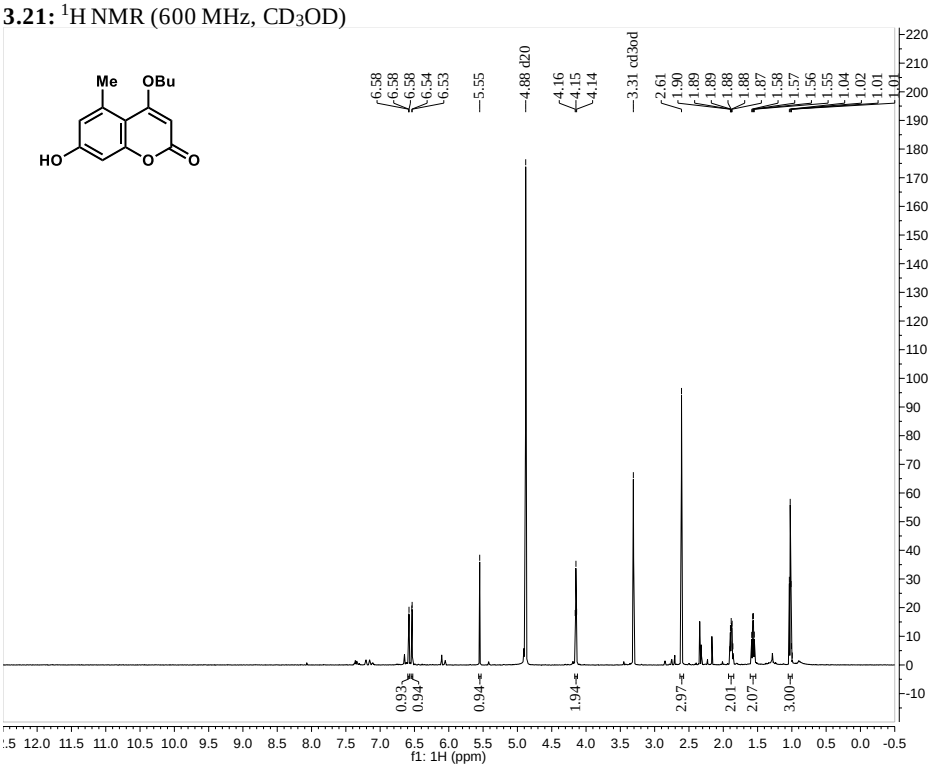
3.20: ^{1}H NMR (600 MHz, acetone-d_6)



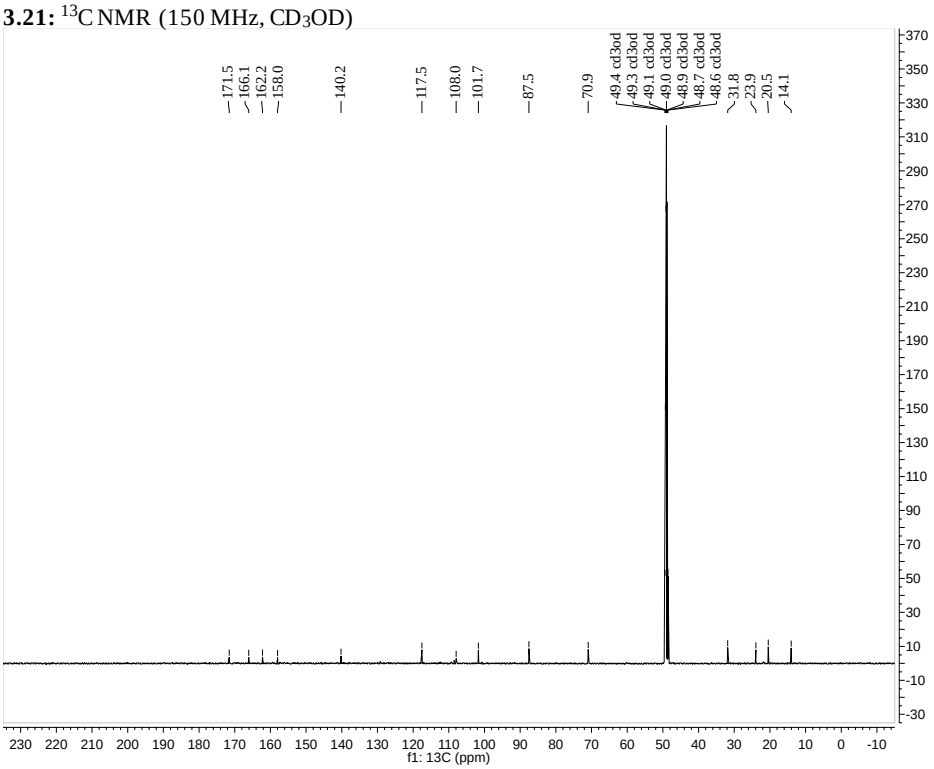
3.20: ^{13}C NMR (150 MHz, acetone-d_6)



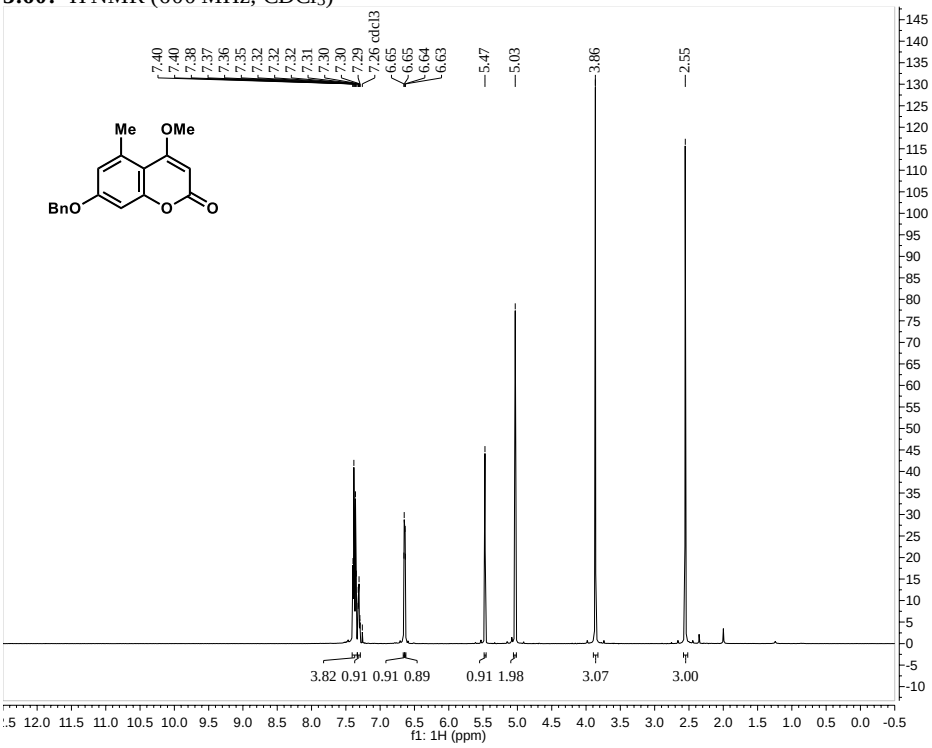
3.21: ^{1}H NMR (600 MHz, CD_3OD)



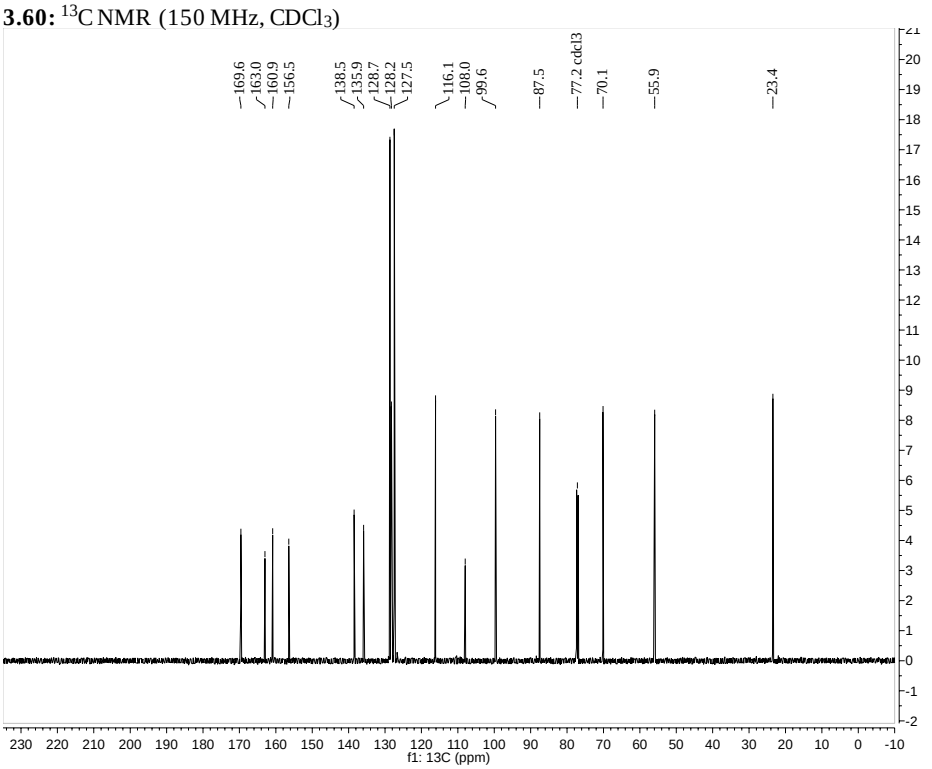
3.21: ^{13}C NMR (150 MHz, CD_3OD)



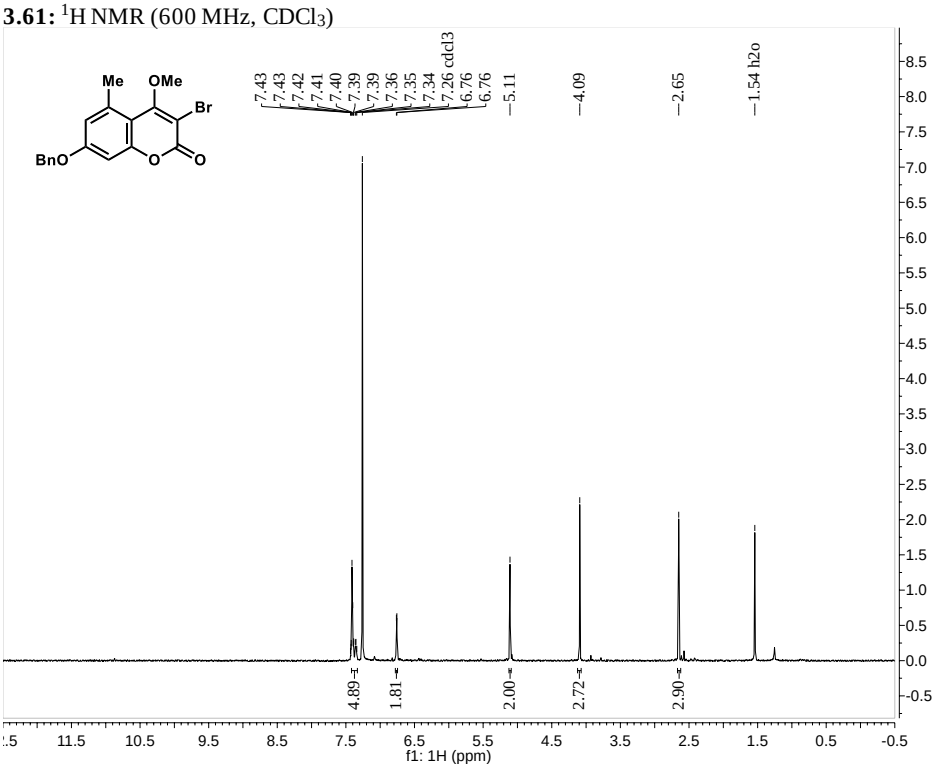
3.60: ^{1}H NMR (600 MHz, $CDCl_3$)



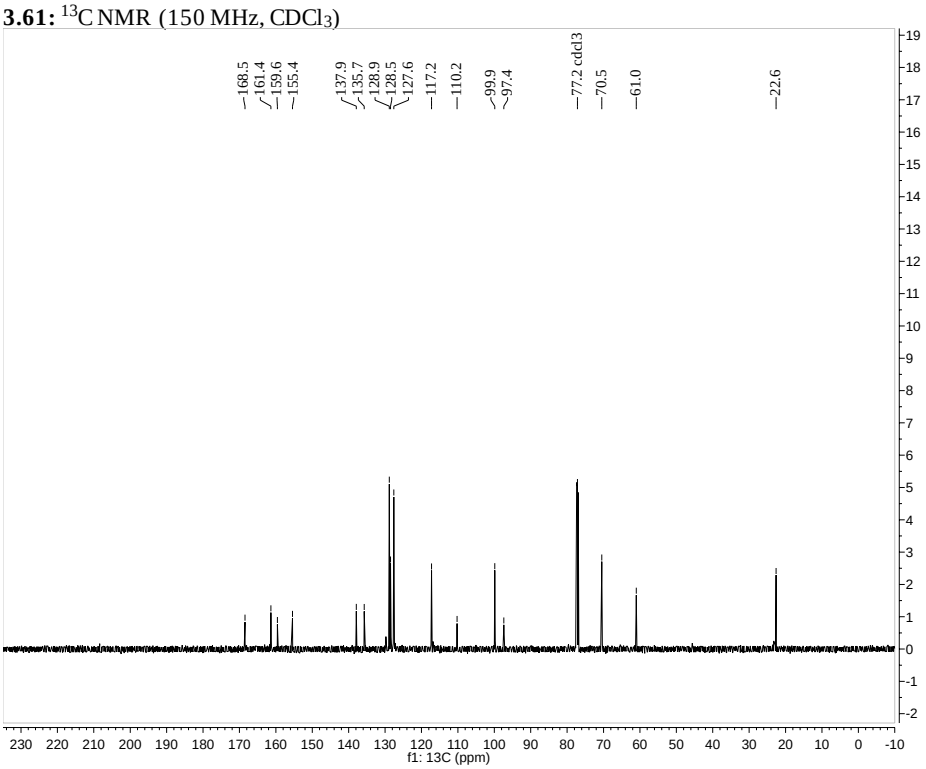
3.60: ^{13}C NMR (150 MHz, $CDCl_3$)



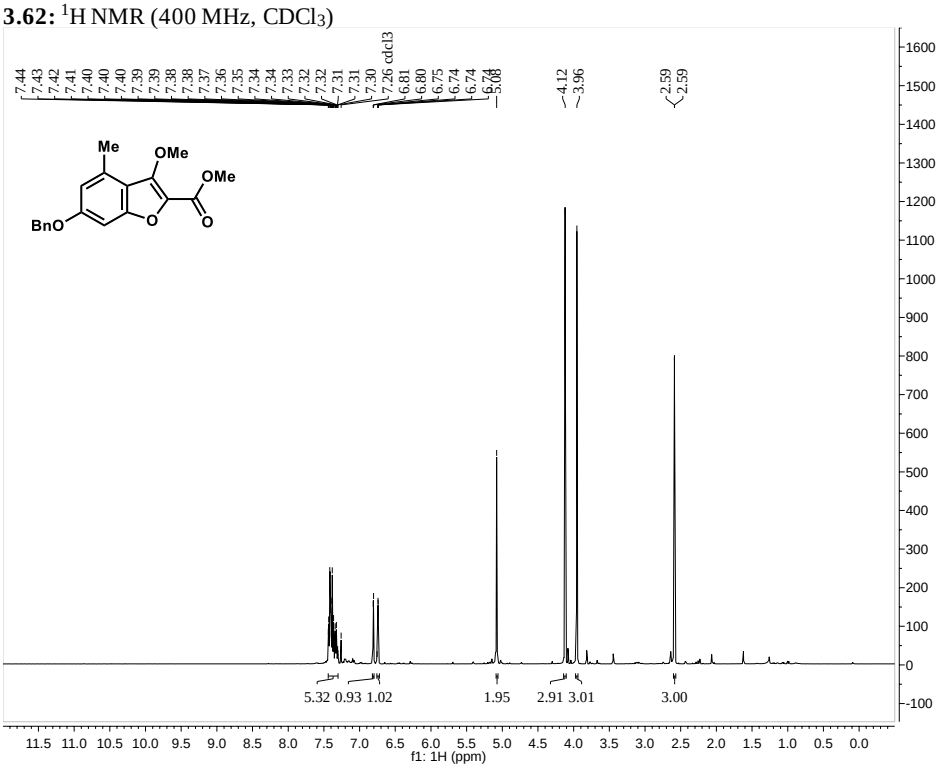
3.61: ^{1}H NMR (600 MHz, $CDCl_3$)



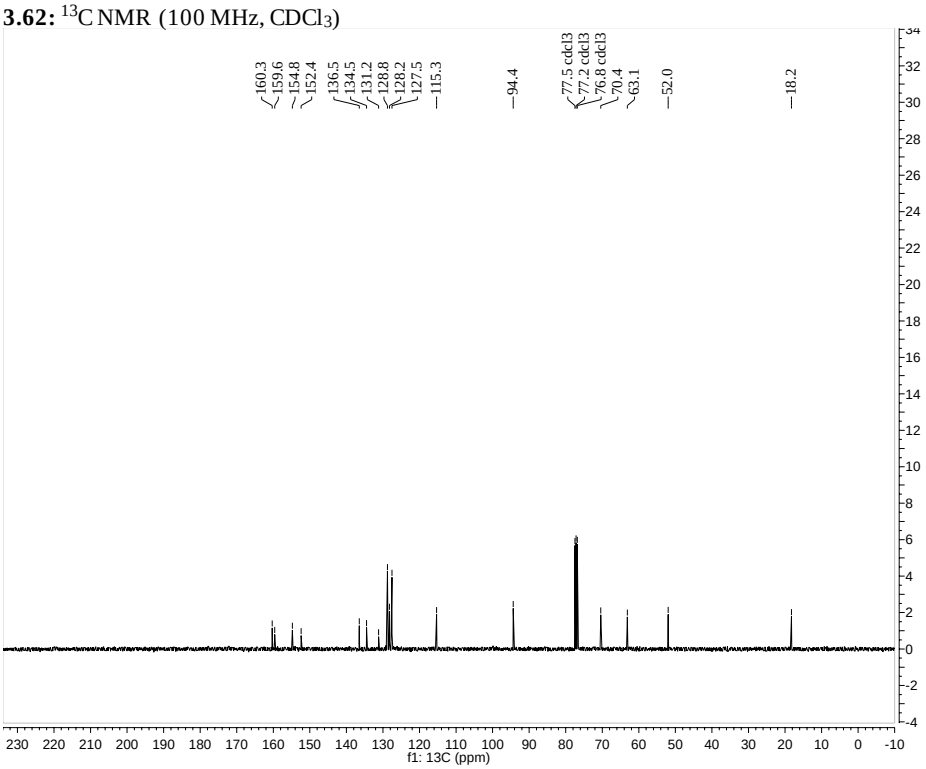
3.61: ^{13}C NMR (150 MHz, $CDCl_3$)



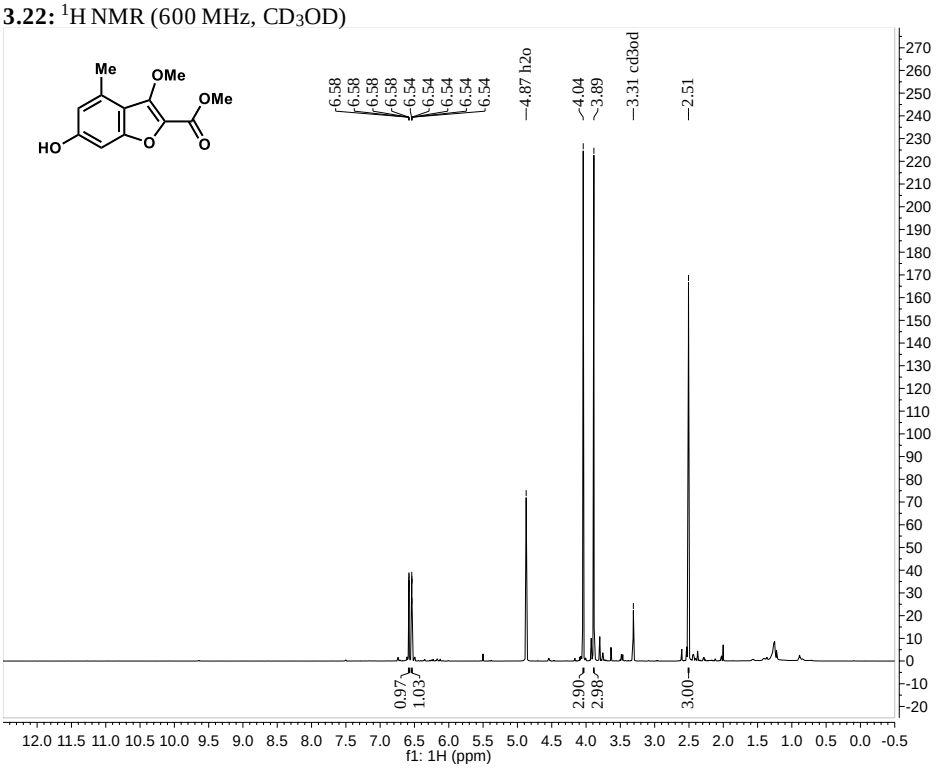
3.62: ^{1}H NMR (400 MHz, $CDCl_3$)



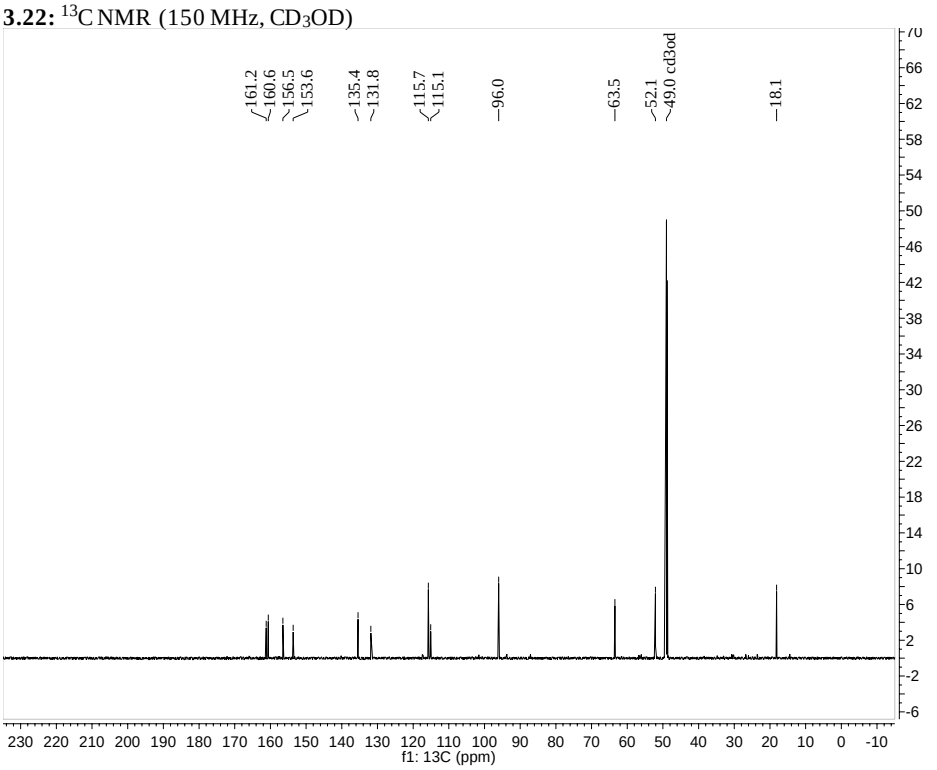
3.62: ^{13}C NMR (100 MHz, $CDCl_3$)



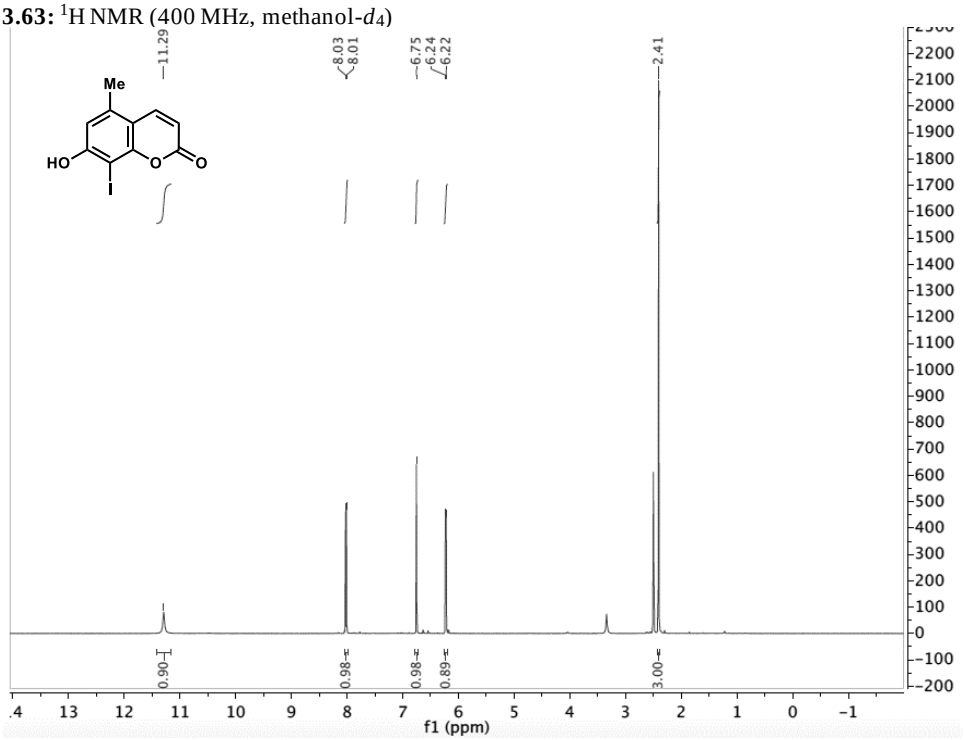
3.22: ^{1}H NMR (600 MHz, CD_3OD)



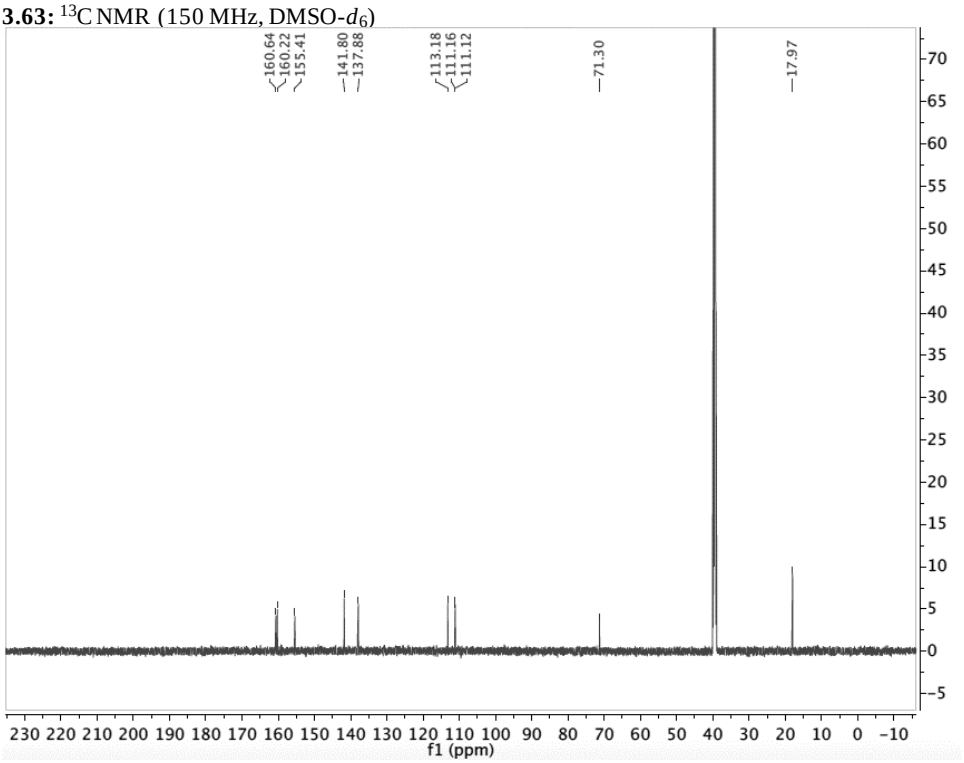
3.22: ^{13}C NMR (150 MHz, CD_3OD)



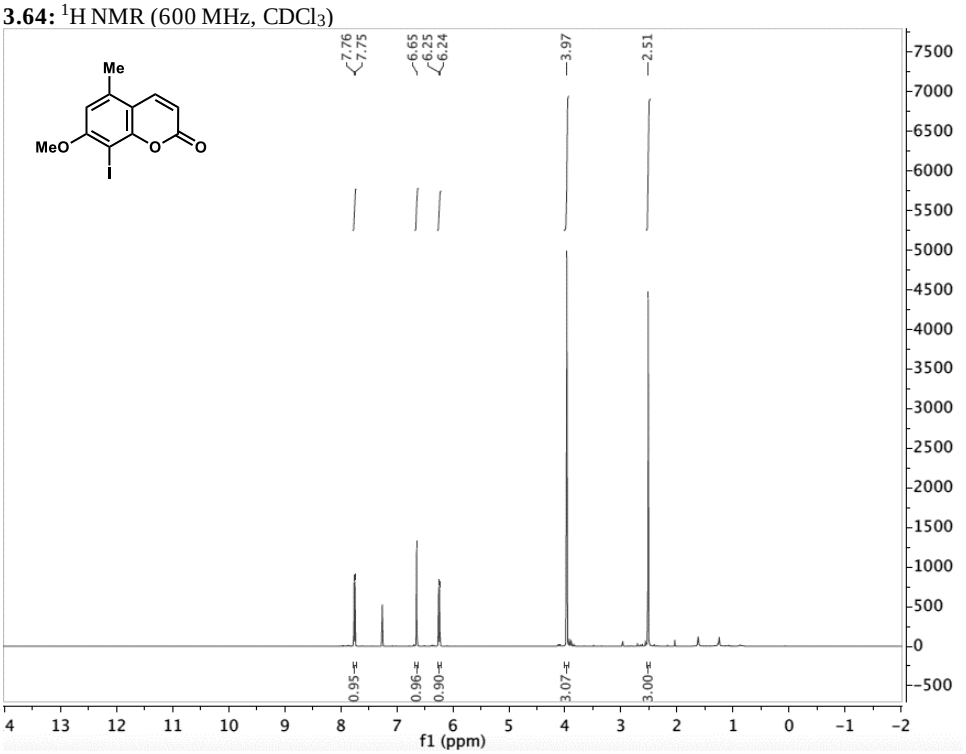
3.63: ^{1}H NMR (400 MHz, methanol-d_4)



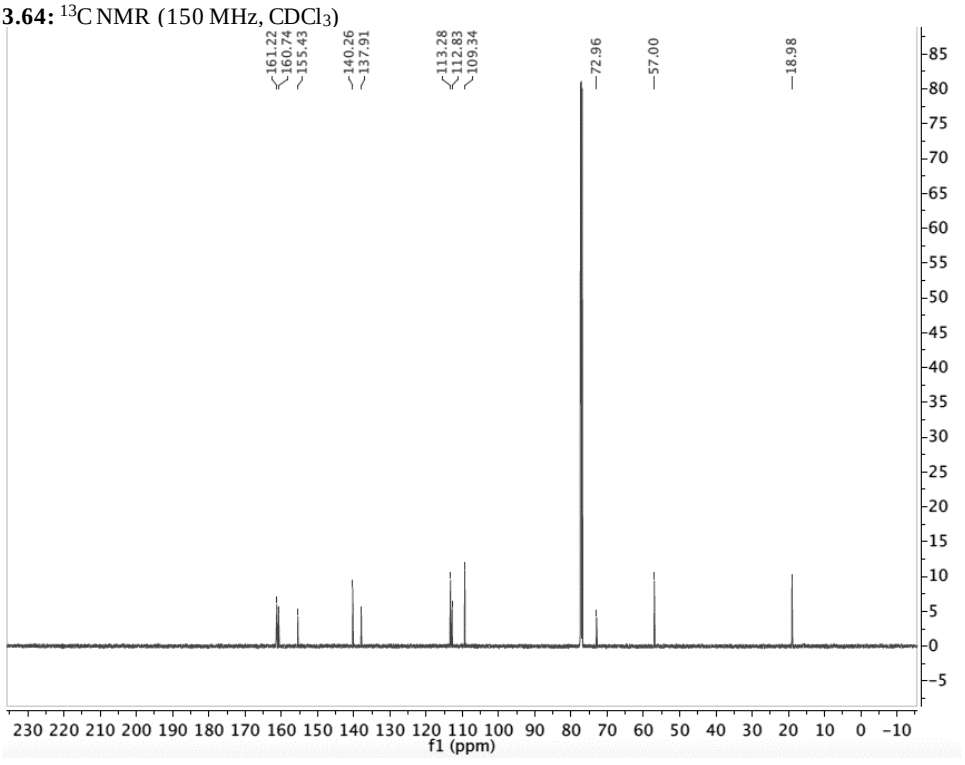
3.63: ^{13}C NMR (150 MHz, DMSO-d_6)



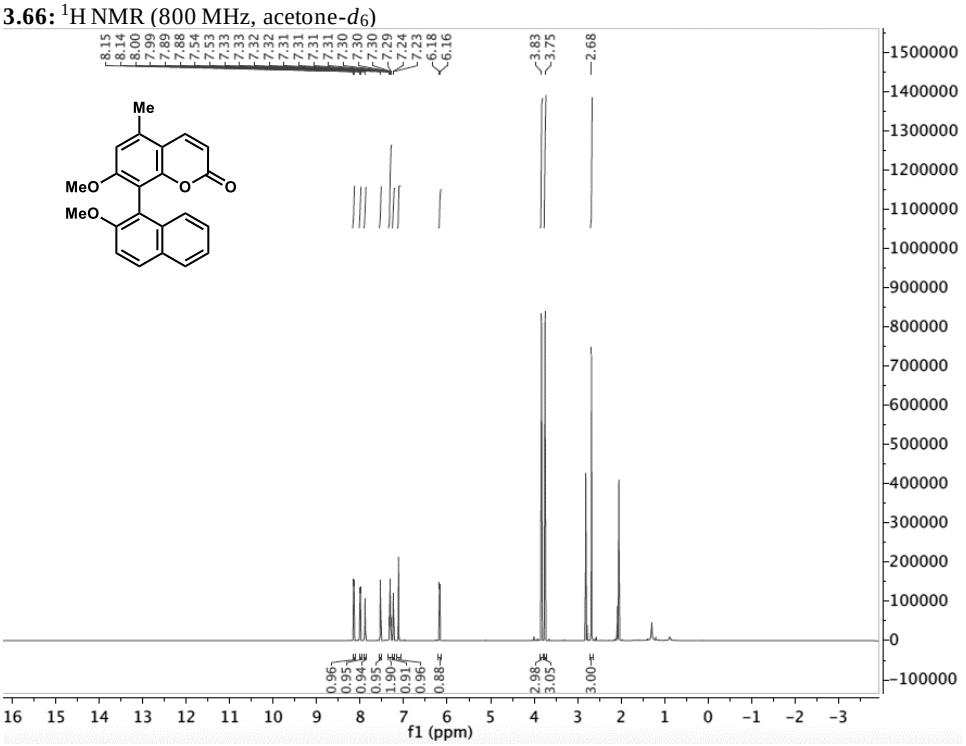
3.64: ^{1}H NMR (600 MHz, $CDCl_3$)



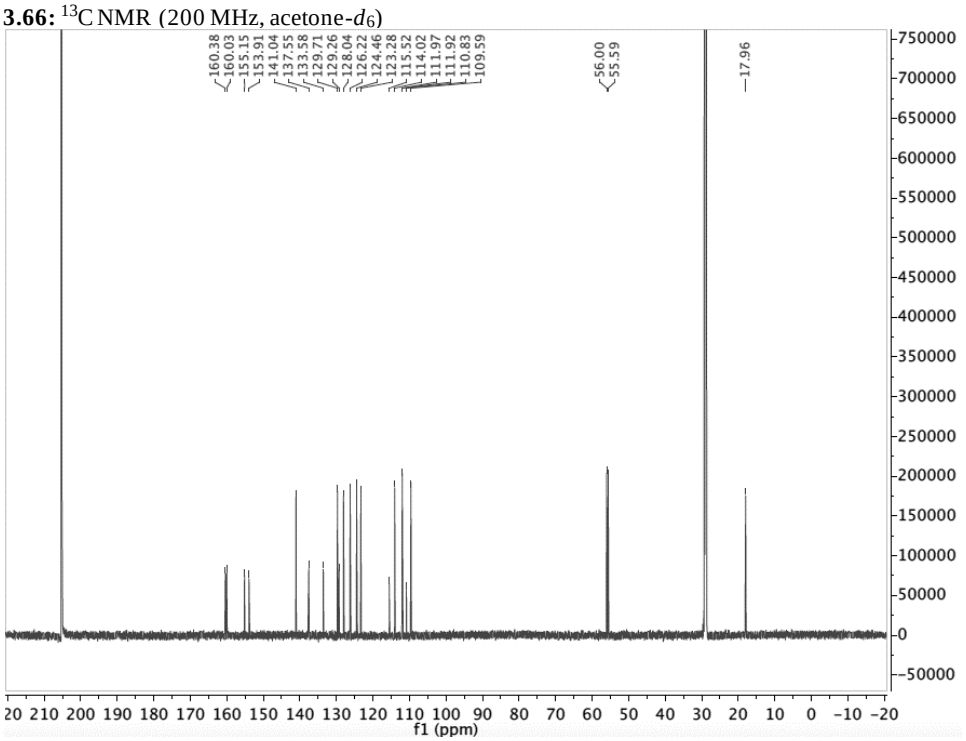
3.64: ^{13}C NMR (150 MHz, $CDCl_3$)



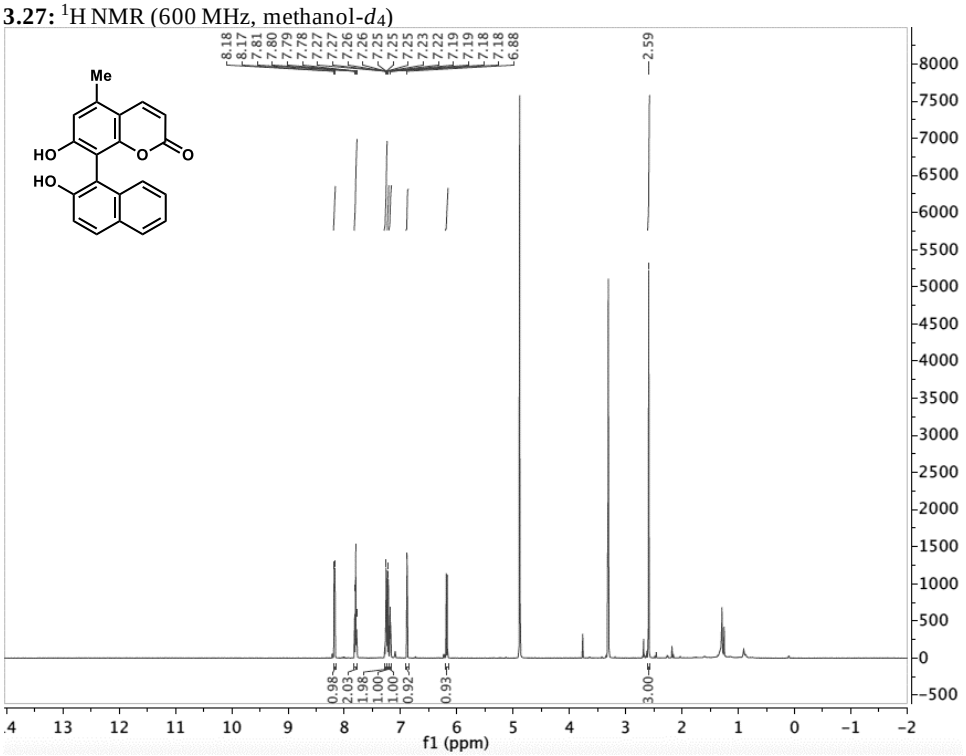
3.66: ^{1}H NMR (800 MHz, acetone-d_6)



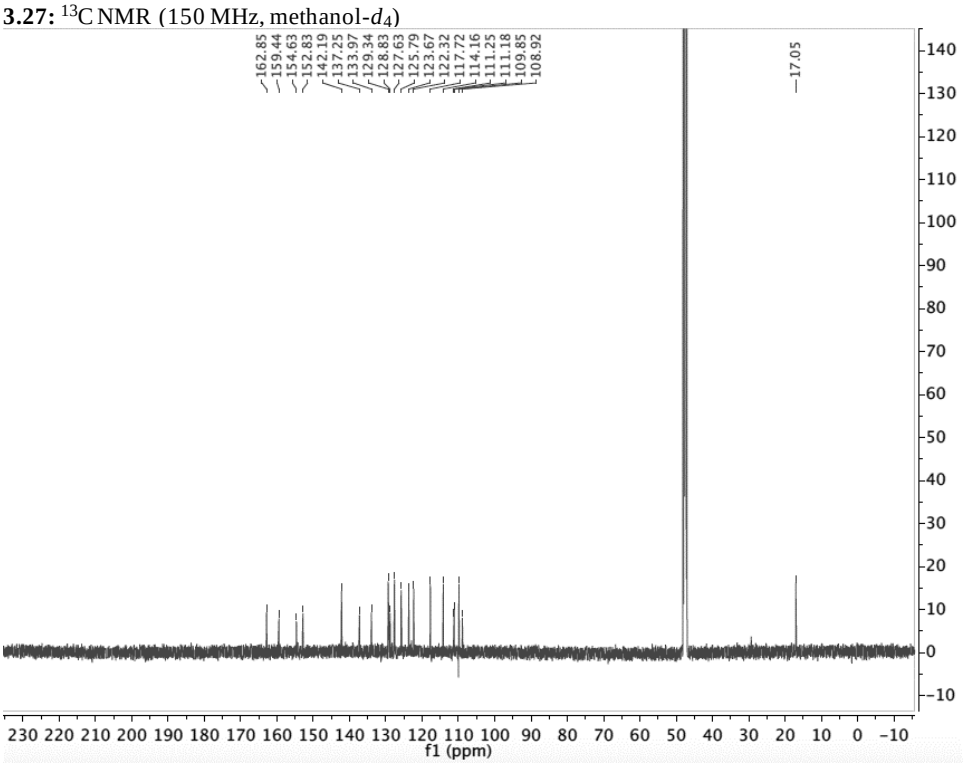
3.66: ^{13}C NMR (200 MHz, acetone-d_6)



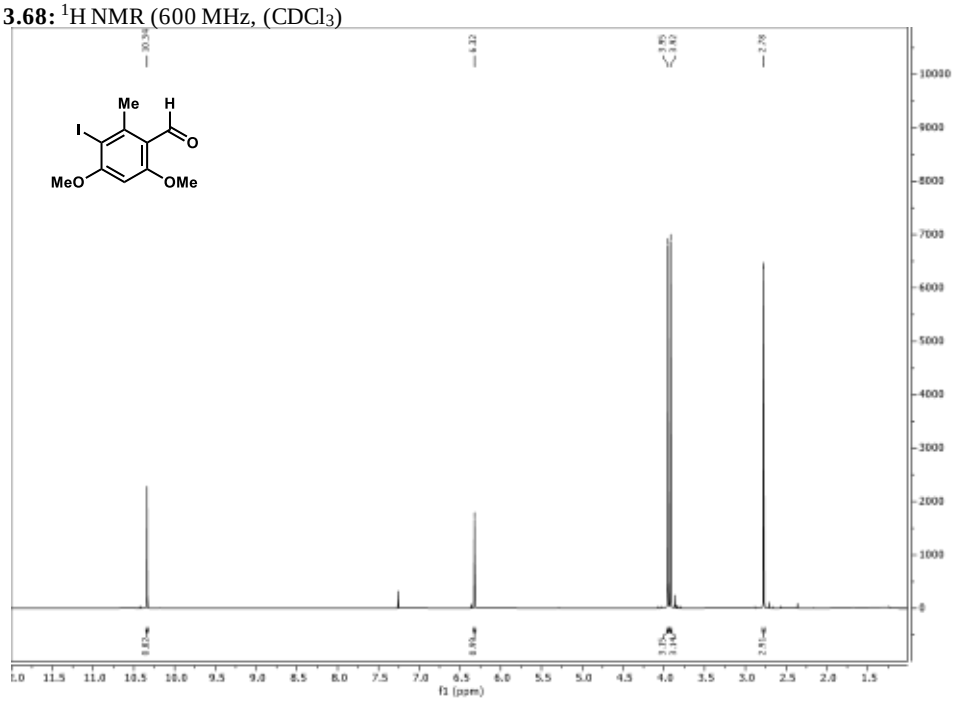
3.27: ^{1}H NMR (600 MHz, methanol-d_4)



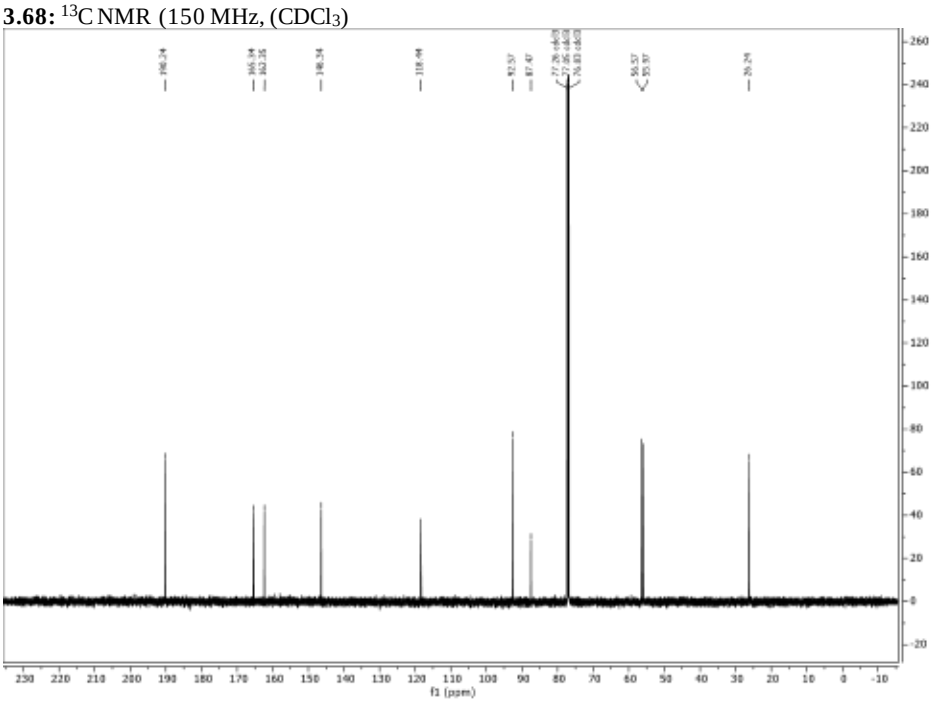
3.27: ^{13}C NMR (150 MHz, methanol-d_4)



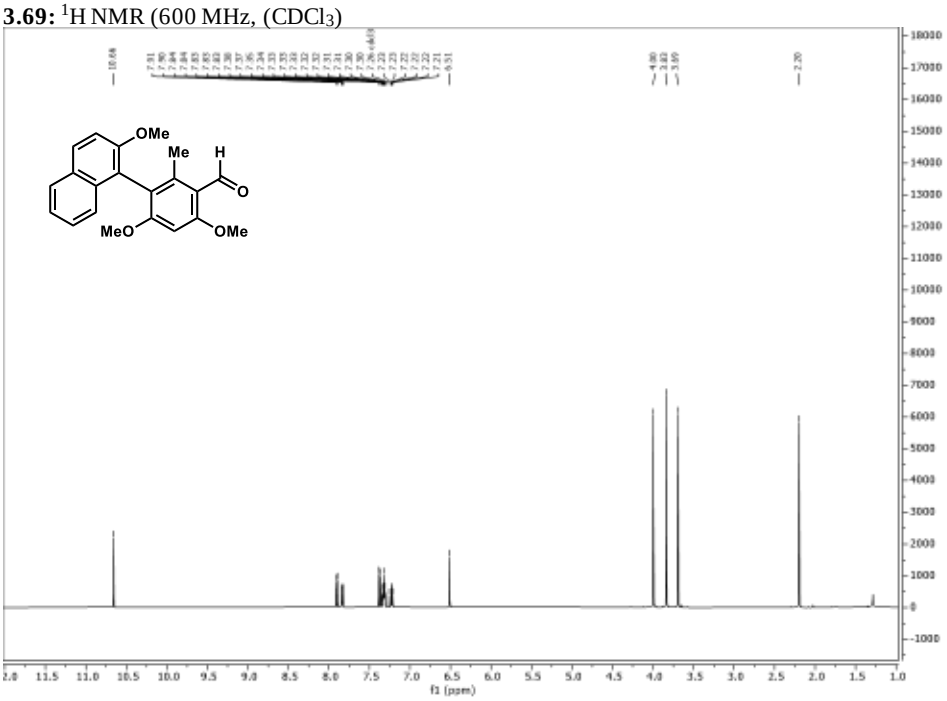
3.68: ^{1}H NMR (600 MHz, ($CDCl_3$)



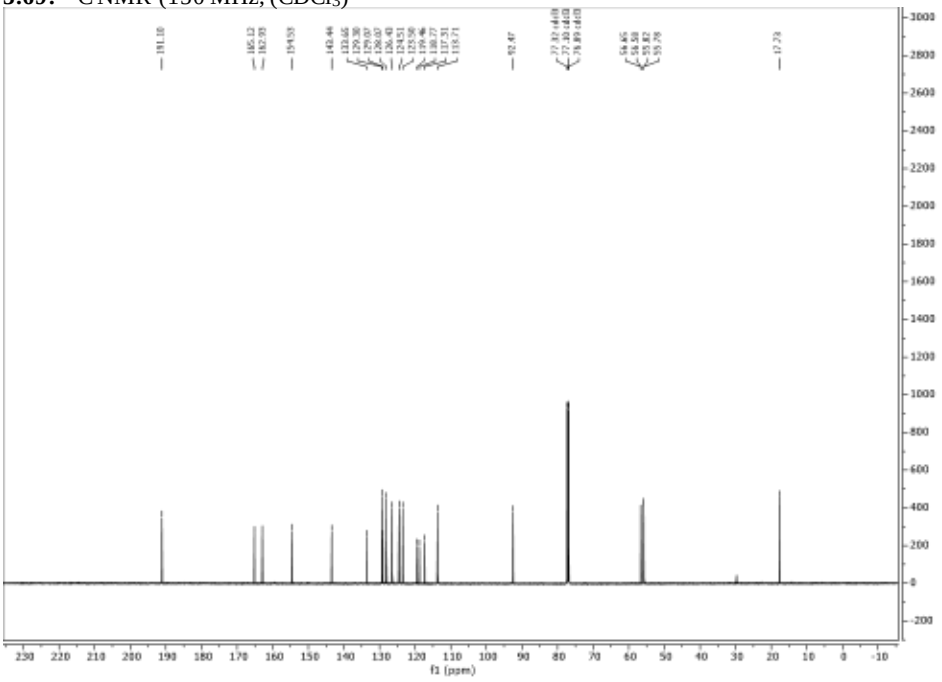
3.68: ^{13}C NMR (150 MHz, ($CDCl_3$)



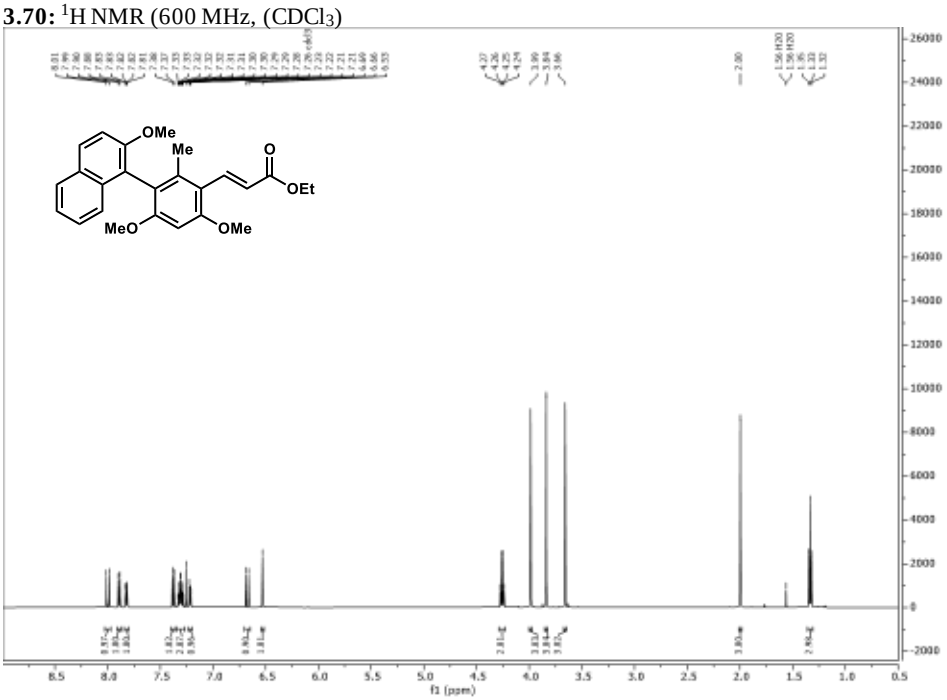
3.69: ^{1}H NMR (600 MHz, ($CDCl_3$)



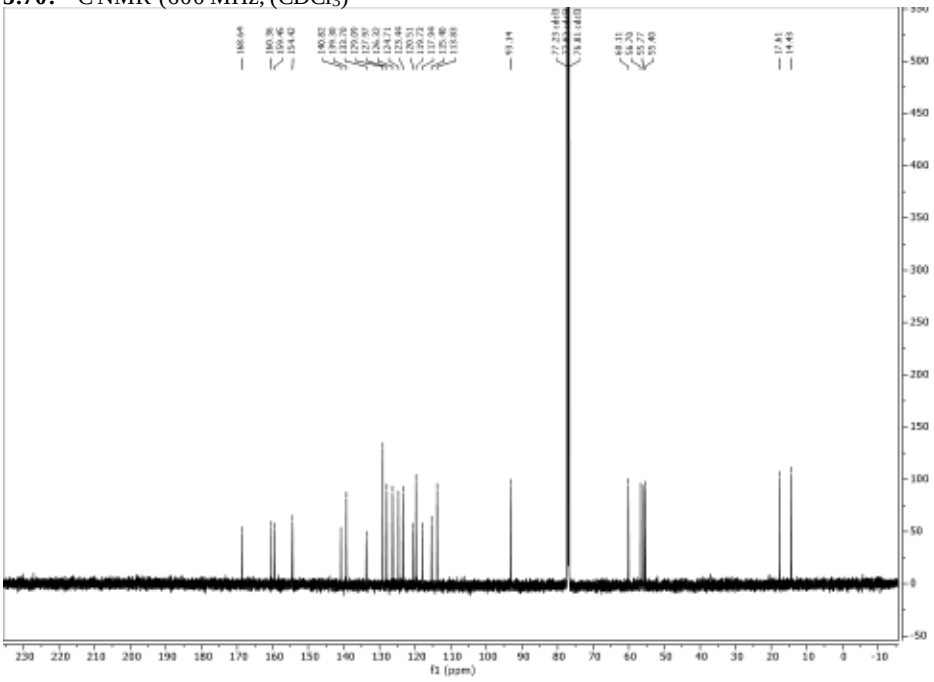
3.69: ^{13}C NMR (150 MHz, ($CDCl_3$)



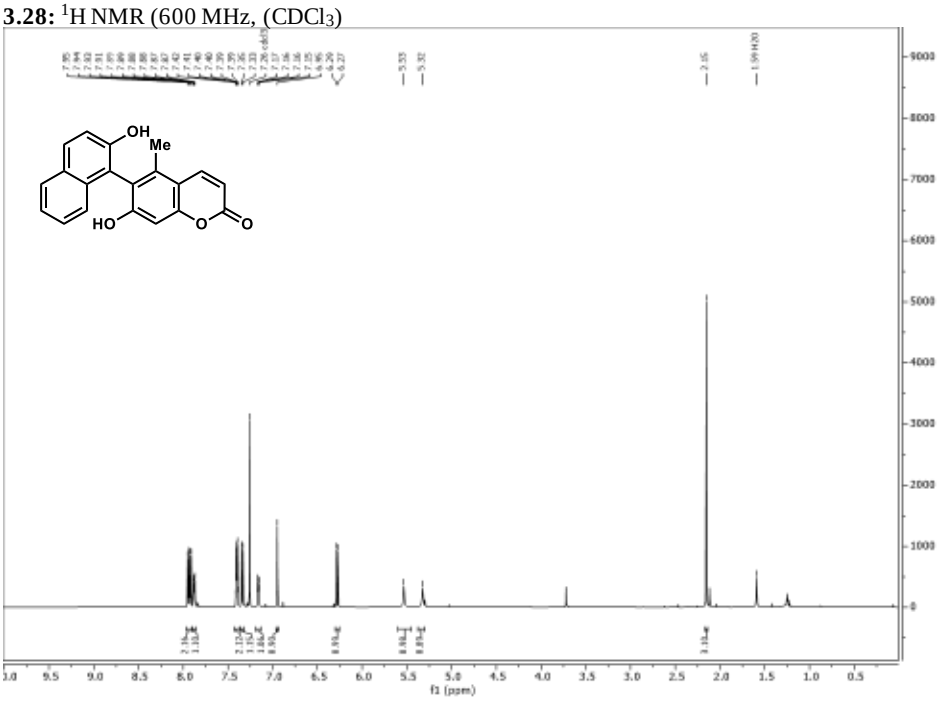
3.70: ^{1}H NMR (600 MHz, ($CDCl_3$)



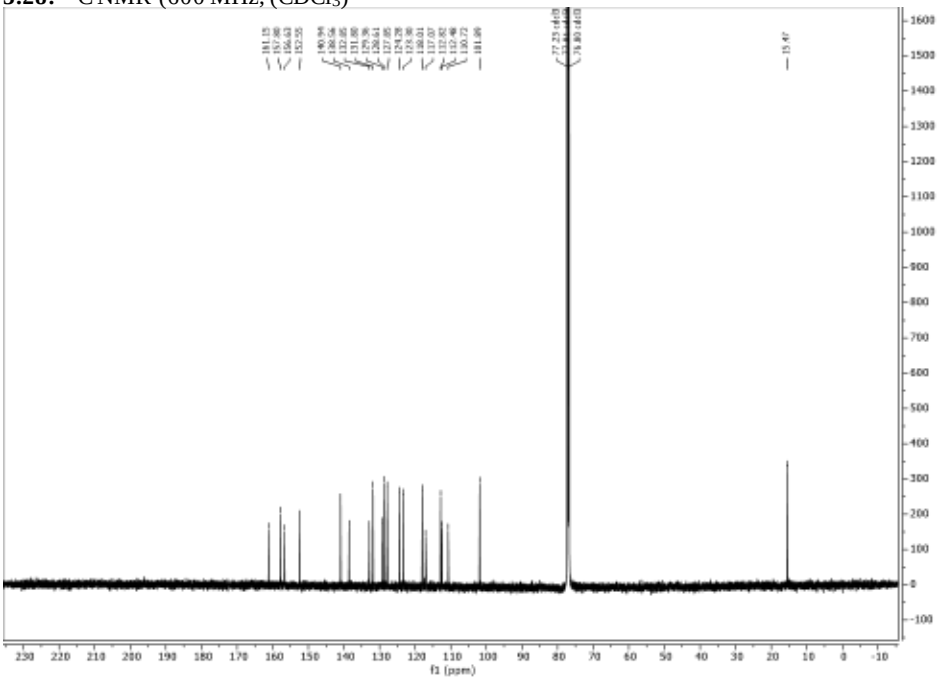
3.70: ^{13}C NMR (600 MHz, ($CDCl_3$)



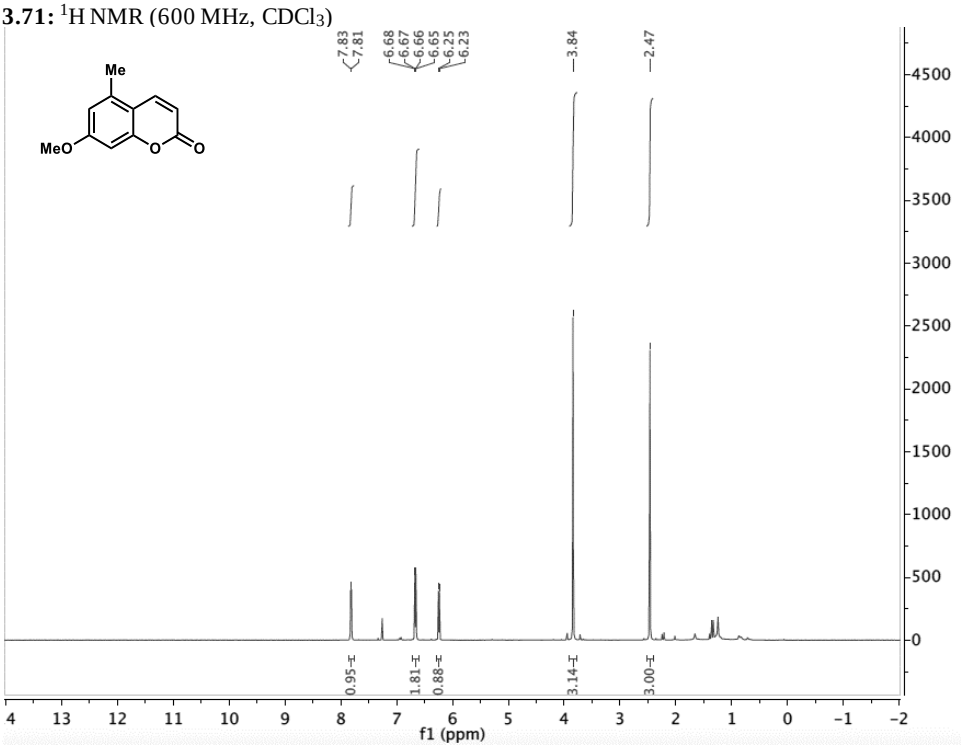
3.28: ^{1}H NMR (600 MHz, ($CDCl_3$)



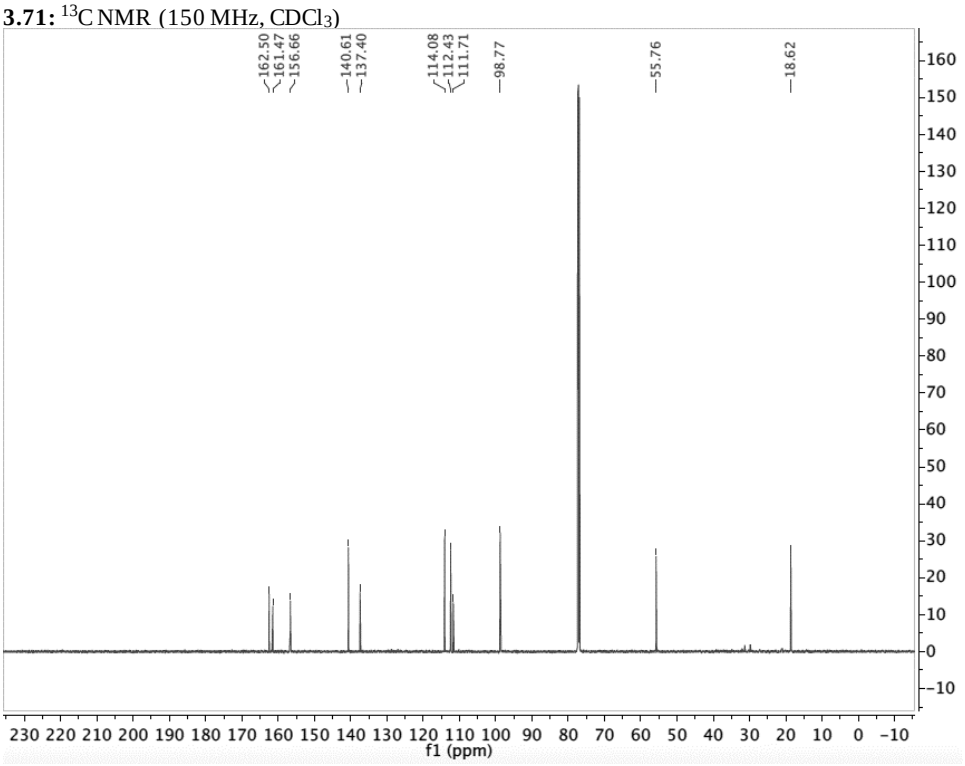
3.28: ^{13}C NMR (600 MHz, ($CDCl_3$)



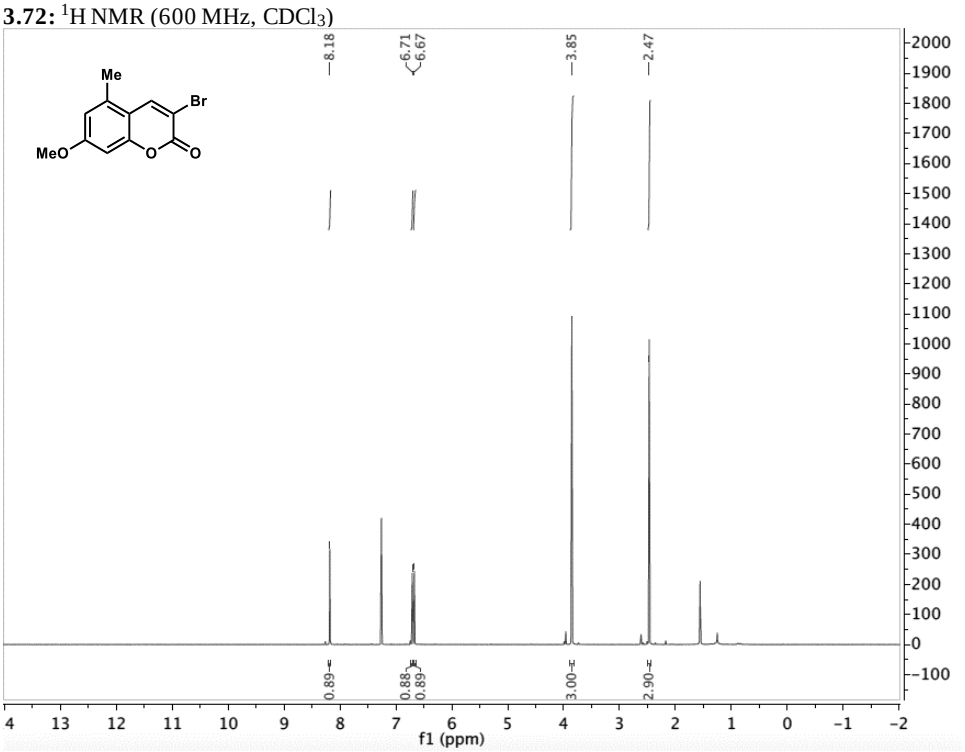
3.71: ^{1}H NMR (600 MHz, $CDCl_3$)



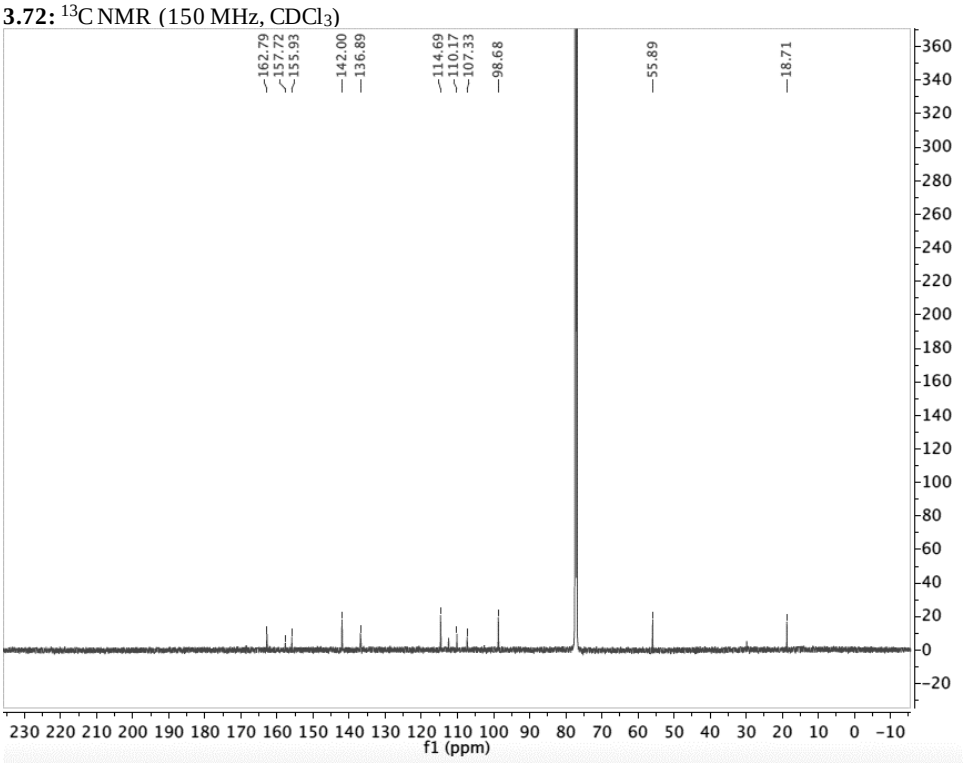
3.71: ^{13}C NMR (150 MHz, $CDCl_3$)



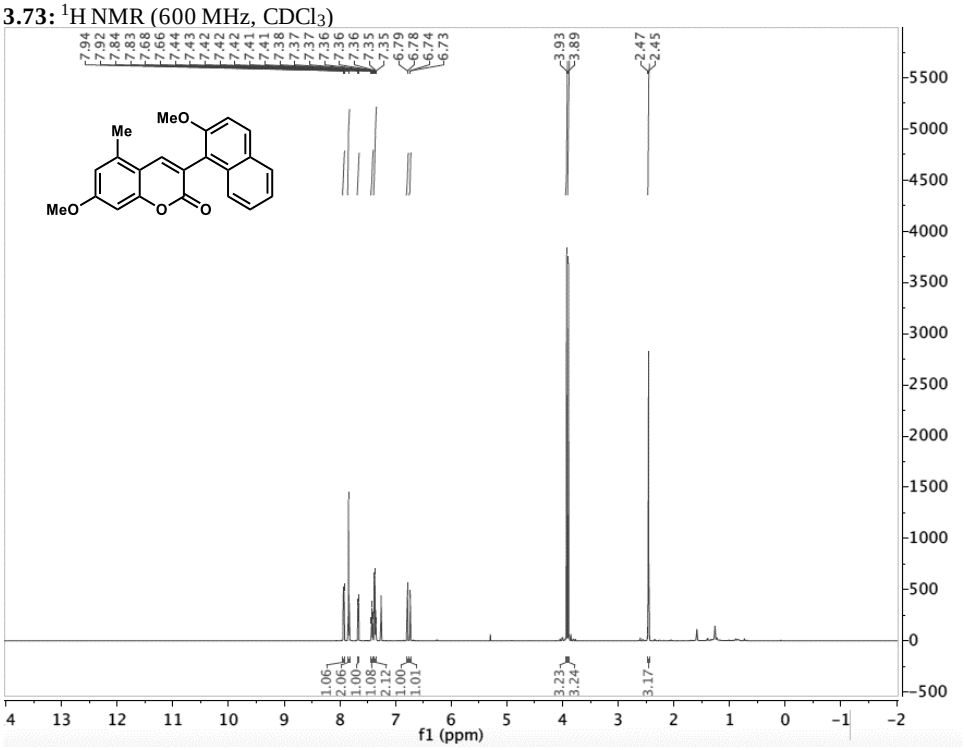
3.72: ^{1}H NMR (600 MHz, $CDCl_3$)



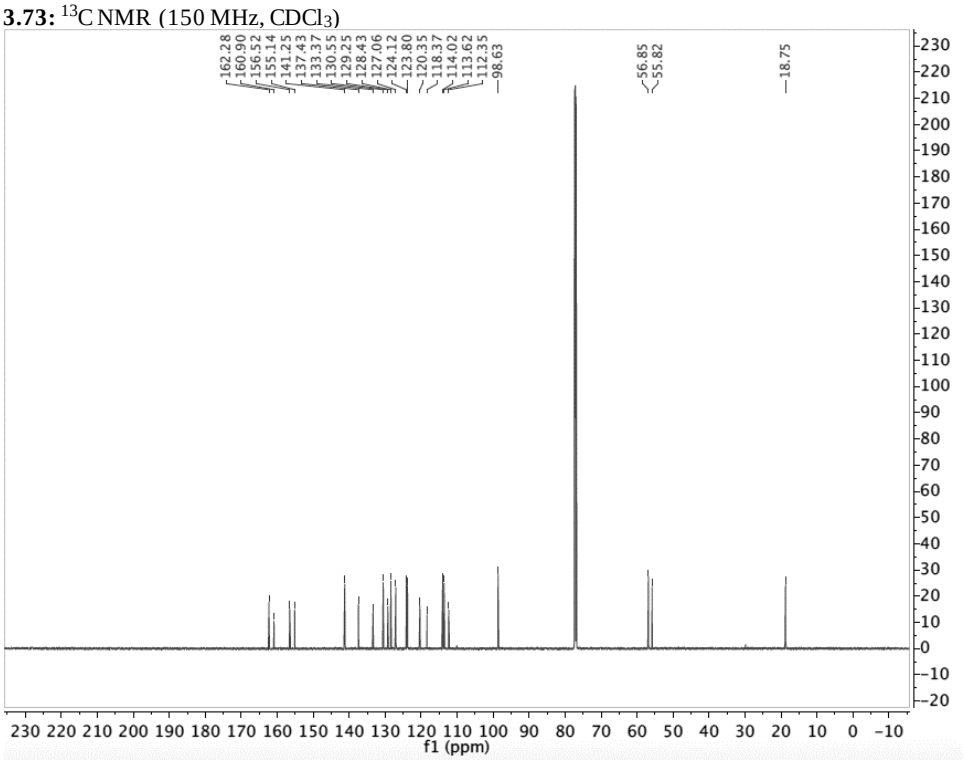
3.72: ^{13}C NMR (150 MHz, $CDCl_3$)



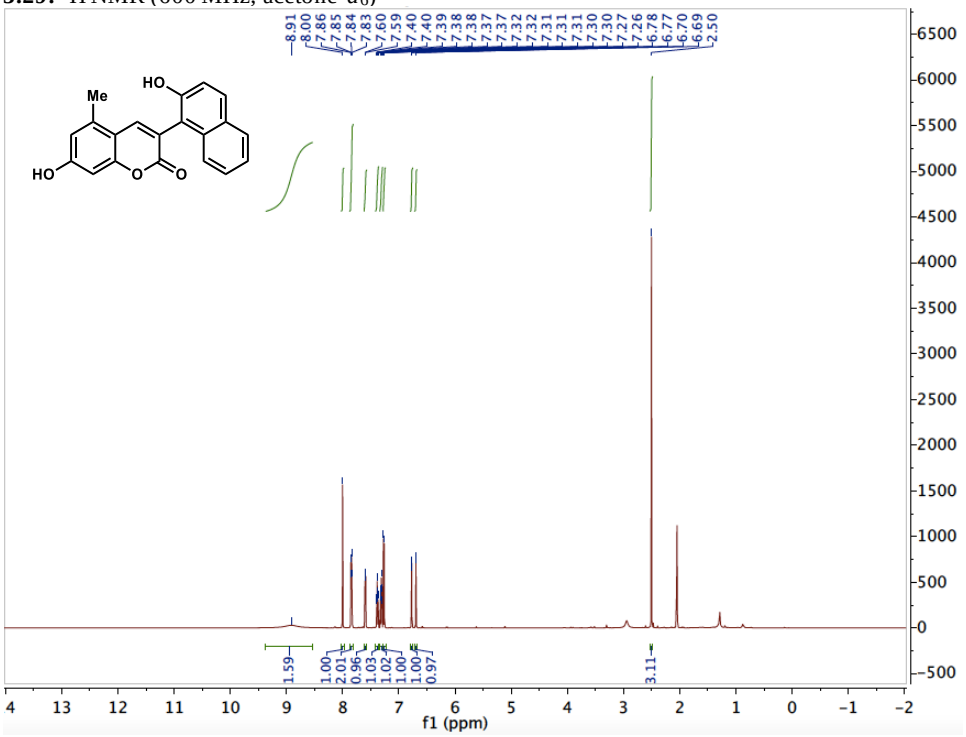
3.73: ^{1}H NMR (600 MHz, $CDCl_3$)



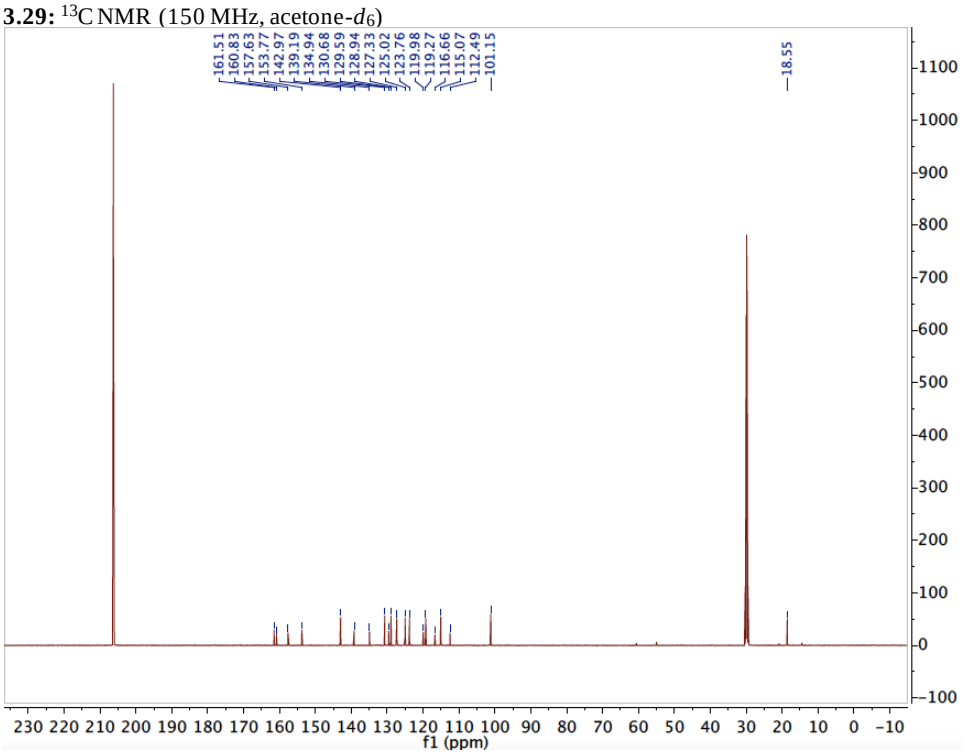
3.73: ^{13}C NMR (150 MHz, $CDCl_3$)



3.29: ^{1}H NMR (600 MHz, acetone-d_6)



3.29: ^{13}C NMR (150 MHz, acetone-d_6)



REFERENCES

1. Yet, L. *Privileged Structures in Drug Discovery: Medicinal Chemistry and Synthesis*. First ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2018.

2. Yoon, T. P.; Jacobsen, E. N. Privileged Chiral Catalysts. *Science* **2003**, *299*, 1691–1693.

3. Yin; Liebscher, J. Carbon−Carbon Coupling Reactions Catalyzed by Heterogeneous Palladium Catalysts. *Chem. Rev.* **2007**, *107*, 133–173.

4. Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug Discov.* **2018**, *17*, 709–727.

5. Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, *54*, 3451–3479.

6. Yin, J.; Rainka, M. P.; Zhang, X.-X.; Buchwald, S. L. A Highly Active Suzuki Catalyst for the Synthesis of Sterically Hindered Biaryls: Novel Ligand Coordination. *J. Am. Chem. Soc.* **2002**, *124*, 1162–1163.

7. Cammidge, A. N.; Crépy, K. V. L. Synthesis of Chiral Binaphthalenes Using the Asymmetric Suzuki Reaction. *Tetrahedron* **2004**, *60*, 4377–4386.

8. Martin, R.; Buchwald, S. L. Palladium-Catalyzed Suzuki−Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461–1473.

9. Valente, C.; Çalimsiz, S.; Hoi, K. H.; Mallik, D.; Sayah, M.; Organ, M. G. The Development of Bulky Palladium NHC Complexes for the Most-Challenging Cross-Coupling Reactions. *Angew. Chem. Int. Ed.* **2012**, *51*, 3314–3332.

10. Patel, N. D.; Sieber, J. D.; Tcyrulnikov, S.; Simmons, B. J.; Rivalti, D.; Duvvuri, K.; Zhang, Y.; Gao, D. A.; Fandrick, K. R.; Haddad, N.; Lao, K. S.; Mangunuru, H. P. R.; Biswas, S.; Qu, B.; Grinberg, N.; Pennino, S.; Lee, H.; Song, J. J.; Gupton, B. F.; Garg, N. K.; Kozlowski, M. C.; Senanayake, C. H. Computationally Assisted Mechanistic Investigation and Development of Pd-Catalyzed Asymmetric Suzuki–Miyaura and Negishi Cross-Coupling Reactions for Tetra-*Ortho*-Substituted Biaryl Synthesis. *ACS Catal.* **2018**, *8*, 10190–10209.

11. Ackermann, L.; Potukuchi, H. K.; Althammer, A.; Born, R.; Mayer, P. Tetra-*Ortho*-Substituted Biaryls through Palladium-Catalyzed Suzuki−Miyaura Couplings with a Diaminochlorophosphine Ligand. *Org. Lett.* **2010**, *12*, 1004–1007.

12. Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59*, 4443–4458.

13. Yang, Y.; Lan, J.; You, J. Oxidative C–H/C–H Coupling Reactions between Two (Hetero) Arenes. *Chem. Rev.* **2017**, *117*, 8787–8863.

14. Kozlowski, M. C. Oxidative Coupling in Complexity Building Transforms. *Acc. Chem. Res.* **2017**, *50*, 638–643.

15. Lee, Y. E.; Cao, T.; Torruellas, C.; Kozlowski, M. C. Selective Oxidative Homo- and Cross-Coupling of Phenols with Aerobic Catalysts. *J. Am. Chem. Soc.* **2014**, *136*, 6782–6785.

16. Nieves-Quinones, Y.; Paniak, T. J.; Lee, Y. E.; Kim, S. M.; Tcyrulnikov, S.; Kozlowski, M. C. Chromium-Salen Catalyzed Cross-Coupling of Phenols: Mechanism and Origin of the Selectivity. *J. Am. Chem. Soc.* **2019**, *141*, 10016–10032.

17. Shalit, H.; Dyadyuk, A.; Pappo, D. Selective Oxidative Phenol Coupling by Iron Catalysis. *J. Org. Chem.* **2019**, *84*, 1677–1686.

18. Reiss, H.; Shalit, H.; Vershinin, V.; More, N. Y.; Forckosh, H.; Pappo, D. Cobalt(II)[Salen]-Catalyzed Selective Aerobic Oxidative Cross-Coupling between Electron-Rich Phenols and 2-Naphthols. *J. Org. Chem.* **2019**, *84*, 7950–7960.

19. Röckl, J. L.; Schollmeyer, D.; Franke, R.; Waldvogel, S. R. Dehydrogenative Anodic C−C Coupling of Phenols Bearing Electron-Withdrawing Groups. *Angew. Chem. Int. Ed.* **2020**, *59*, 315–319.

20. Kang, H.; Herling, M. R.; Niederer, K. A.; Lee, Y. E.; Vasu Govardhana Reddy, P.; Dey, S.; Allen, S. E.; Sung, P.; Hewitt, K.; Torruellas, C.; Kim, G. J.; Kozlowski, M. C. Enantioselective Vanadium-Catalyzed Oxidative Coupling: Development and Mechanistic Insights. *J. Org. Chem.* **2018**, *83*, 14362–14384.

21. Libman, A.; Shalit, H.; Vainer, Y.; Narute, S.; Kozuch, S.; Pappo, D. Synthetic and Predictive Approach to Unsymmetrical Biphenols by Iron-Catalyzed Chelated Radical–Anion Oxidative Coupling. *J. Am. Chem. Soc.* **2015**, *137*, 11453–11460.

22. Morimoto, K.; Sakamoto, K.; Ohshika, T.; Dohi, T.; Kita, Y. Organo-Iodine(III)-Catalyzed Oxidative Phenol-Arene and Phenol-Phenol Cross-Coupling Reaction. *Angew. Chem. Int. Ed.* **2016**, *55*, 3652–3656.

23. More, N. Y.; Jeganmohan, M. Oxidative Cross-Coupling of Two Different Phenols: An Efficient Route to Unsymmetrical Biphenols. *Org. Lett.* **2015**, *17*, 3042–3045.

24. Egami, H.; Katsuki, T. Iron-Catalyzed Asymmetric Aerobic Oxidation: Oxidative Coupling of 2-Naphthols. *J. Am. Chem. Soc.* **2009**, *131*, 6082–6083.

25. Hovorka, M.; Günterova, J.; Zavada, J. Highly Selective Oxidative Cross-Coupling of Substituted 2-Naphthols: A Convenient Approach to Unsymmetrical 1,1′-Binaphthalene-2,2′-Diols. *Tet. Lett.* **1990**, *31*, 413‑416.

26. Li, X.; Hewgley, J. B.; Mulrooney, C. A.; Yang, J.; Kozlowski, M. C. Enantioselective Oxidative Biaryl Coupling Reactions Catalyzed by 1,5-Diazadecalin Metal Complexes: Efficient Formation of Chiral Functionalized BINOL Derivatives. *J. Org. Chem.* **2003**, *68*, 5500–5511.

27. Tian, J.-M.; Wang, A.-F.; Yang, J.-S.; Zhao, X.-J.; Tu, Y.-Q.; Zhang, S.-Y.; Chen, Z.-M. Copper-Complex-Catalyzed Asymmetric Aerobic Oxidative Cross-Coupling of 2-Naphthols: Enantioselective Synthesis of 3,3′-Substituted C1-Symmetric BINOLs. *Angew. Chem. Int. Ed.* **2019**, *58*, 11023–11027.

28. Bringmann, G.; Price Mortimer, A. J.; Keller, P. A.; Gresser, M. J.; Garner, J.; Breuning, M. Atroposelective Synthesis of Axially Chiral Biaryl Compounds. *Angew. Chem. Int. Ed.* **2005**, *44*, 5384–5427.

29. Kočovský, P.; Vyskočil, Š.; Smrčina, M. Non-Symmetrically Substituted 1,1′-Binaphthyls in Enantioselective Catalysis. *Chem. Rev.* **2003**, *103*, 3213–3246.

30. Kozlowski, M. C.; Morgan, B. J.; Linton, E. C. Total Synthesis of Chiral Biaryl Natural Products by Asymmetric Biaryl Coupling. *Chem. Soc. Rev.* **2009**, *38*, 3193–3207.

31. Bringmann, G.; Gulder, T.; Gulder, T. A. M.; Breuning, M. Atroposelective Total Synthesis of Axially Chiral Biaryl Natural Products. *Chem. Rev.* **2011**, *111*, 563–639.

32. Aldemir, H.; Richarz, R.; Gulder, T. A. The Biocatalytic Repertoire of Natural Biaryl Formation. *Angew. Chem. Int. Ed.* **2014**, *53*, 8286–8293.

33. Mate, D. M.; Alcalde, M. Laccase: A Multi-Purpose Biocatalyst at the Forefront of Biotechnology. *Microb. Biotechnol.* **2017**, *10*, 1457–1467.

34. Sagui, F.; Chirivì, C.; Fontana, G.; Nicotra, S.; Passarella, D.; Riva, S.; Danieli, B. Laccase-Catalyzed Coupling of Catharanthine and Vindoline: An Efficient Approach to the Bisindole Alkaloid Anhydrovinblastine. *Tetrahedron* **2009**, *65*, 312–317.

35. Obermaier, S.; Thiele, W.; Fürtges, L.; Müller, M. Enantioselective Phenol Coupling by Laccases in the Biosynthesis of Fungal Dimeric Naphthopyrones. *Angew. Chem. Int. Ed.* **2019**, *58*, 9125–9128.

36. Fasan, R. Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal.* **2012**, *2*, 647–666.

37. Liu, J.; Liu, A.; Hu, Y. Enzymatic Dimerization in the Biosynthetic Pathway of Microbial Natural Products. *Nat. Prod. Rep.* **2021**, *38*, 1469-1505.

38. Gil Girol, C.; Fisch, K. M.; Heinekamp, T.; Günther, S.; Hüttel, W.; Piel, J.; Brakhage, A. A.; Müller, M. Regio- and Stereoselective Oxidative Phenol Coupling in *Aspergillus Niger*. *Angew. Chem. Int. Ed.* **2012**, *51*, 9788–9791.

39. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

40. Morley, K. L.; Kazlauskas, R. J. Improving Enzyme Properties: When Are Closer Mutations Better? *Trends Biotechnol.* **2005**, *23*, 231–237.

41. Chen, M. M. Y.; Snow, C. D.; Vizcarra, C. L.; Mayo, S. L.; Arnold, F. H. Comparison of Random Mutagenesis and Semi-Rational Designed Libraries for Improved Cytochrome P450 BM3-Catalyzed Hydroxylation of Small Alkanes. *Protein Eng. Des. Sel.* **2012**, *25*, 171–178.

42. Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr. Opin. Chem. Biol.* **2006**, *10*, 498−508.

43. Sherer, E. C.; Lee, C. H.; Shpungin, J.; Cuff, J. F.; Da, C.; Ball, R.; Bach, R.; Crespo, A.; Gong, X.; Welch, C. J. Systematic Approach to Conformational Sampling for Assigning Absolute Configuration Using Vibrational Circular Dichroism. *J. Med. Chem.* **2014**, *57*, 477–494.

44. Liu, Z.; Shultz, C. S.; Sherwood, C. A.; Krska, S.; Dormer, P. G.; Desmond, R.; Lee, C.; Sherer, E. C.; Shpungin, J.; Cuff, J. Highly Enantioselective Synthesis of Anti Aryl β-Hydroxy α-Amino Esters via DKR Transfer Hydrogenation. *Tet. Lett.* **2011**, *52*, 1685–1688.

45. Joyce, L. A.; Nawrat, C. C.; Sherer, E. C.; Biba, M.; Brunskill, A.; Martin, G. E.; Cohen, R. D.; Davies, I. W. Beyond Optical Rotation: What's Left Is Not Always Right in Total Synthesis. *Chem. Sci.* **2018**, *9*, 415–424.

46. Petersson, a.; Bennett, A.; Tensfeldt, T. G.; Al‐Laham, M. A.; Shirley, W. A.; Mantzaris, J. A Complete Basis Set Model Chemistry. I. The Total Energies of Closed-Shell Atoms and Hydrides of the First-Row Elements. *J. Chem. Phys.* **1988**, *89*, 2193‐2218.

47. Petersson, G.; Al-Laham, M. A. A Complete Basis Set Model Chemistry. II. Open-Shell Systems and the Total Energies of the First-Row Atoms. *J. Chem. Phys.* **1991**, *94*, 6081‐6090.

48. Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* Basis Set for Atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223–1229.

49. Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-Row Atoms. *J. Comp. Chem.* **2001**, *22*, 976‐984.

50. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654‐3665.

51. Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

52. Becke, A. D. A New Mixing of Hartree‐Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372‐1377.

53. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

54. Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

55. Frisch, M. J. e. a. *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford, CT, 2009.

56. Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

57. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538–4543.

58. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li–F. *J. Comp. Chem.* **1983**, *4*, 294–301.

59. Bauernschmitt, R.; Ahlrichs, R. Treatment of Electronic Excitations within the Adiabatic Approximation of Time Dependent Density Functional Theory. *Chem. Phys. Lett.* **1996**, *256*, 454–464.

60. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. Specdis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra. *Chirality* **2013**, *25*, 243–249.

61. Mazzeo, G.; Santoro, E.; Andolfi, A.; Cimmino, A.; Troselj, P.; Petrovic, A. G.; Superchi, S.; Evidente, A.; Berova, N. Absolute Configurations of Fungal and Plant Metabolites by Chiroptical Methods. ORD, ECD, and VCD Studies on Phyllostin, Scytolide, and Oxysporone. *J. Nat. Prod.* **2013**, *76*, 588–599.

62. Legault, C. Y. *Cylview 1.0b*, Université de Sherbrooke: http://www.cylview.org, 2009.

63. Hubbard, J. S.; Harris, T. M. Condensations at the 6 Position of the Methyl Ester and the Dimethylamide of 3,5-Dioxohexanoic Acid via 2,4,6-Trianions. *J. Org. Chem.* **1981**, *46*, 2566–2570.

64. Gil Girol, C.; Fisch, K. M.; Heinekamp, T.; Günther, S.; Hüttel, W.; Piel, J.; Brakhage, A. A.; Müller, M. Regio- and Stereoselective Oxidative Phenol Coupling in *Aspergillus Niger*. *Angew. Chem. Int. Ed.* **2012**, *51*, 9788–9791.

65. Xie, S.-S.; Wang, X.; Jiang, N.; Yu, W.; Wang, K. D.; Lan, J.-S.; Li, Z.-R.; Kong, L.-Y. Multi-Target Tacrine-Coumarin Hybrids: Cholinesterase and Monoamine Oxidase B Inhibition Properties against Alzheimer's Disease. *Eur. J. Med. Chem.* **2015**, *95*, 153–165.

66. Cao, J.-L.; Shen, S.-L.; Yang, P.; Qu, J. A Catalyst-Free One-Pot Construction of Skeletons of 5-Methoxyseselin and Alloxanthoxyletin in Water. *Org. Lett.* **2013**, *15*, 3856–3859.

67. Pandey, G.; Muralikrishna, C.; Bhalerao, U. Mushroom Tyrosinase Catalysed Synthesis of Coumestans, Bebzofuranderivatives and Related Heterocyclic Compounds. *Tetrahedron* **1989**, *45*, 6867–6874.

68. Chaudhuri, D.; Sigmund, E.; Meyer, A.; Röck, L.; Klemm, P.; Lautenschlager, S.; Schmid, A.; Yost, S. R.; Van Voorhis, T.; Bange, S. Metal-Free Oled Triplet Emitters by Side-Stepping Kasha's Rule. *Angew. Chem. Int. Ed.* **2013**, *52*, 13449–13452.

69. Conradt, D.; Hermann, B.; Gerhardt, S.; Einsle, O.; Müller, M. Biocatalytic Properties and Structural Analysis of Phloroglucinol Reductases. *Angew. Chem. Int. Ed.* **2016**, *55*, 15531–15534.

70. Linusson, A.; Gottfries, J.; Olsson, T.; Örnskov, E.; Folestad, S.; Nordén, B.; Wold, S. Statistical Molecular Design, Parallel Synthesis, and Biological Evaluation of a Library of Thrombin Inhibitors. *J. Med. Chem.* **2001**, *44*, 3424–3439.

71. Cutler, H. G.; Crumley, F. G.; Cox, R. H.; Hernandez, O.; Cole, R. J.; Dorner, J. W. Orlandin: A Nontoxic Fungal Metabolite with Plant Growth Inhibiting Properties. *J. Agric. Food Chem.* **1979**, *27*, 592–595.

72. Devji, T.; Reddy, C.; Woo, C.; Awale, S.; Kadota, S.; Carrico-Moniz, D. Pancreatic Anticancer Activity of a Novel Geranylgeranylated Coumarin Derivative. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5770–5773.

73. Dubuffet, T.; Loutz, A.; Lavielle, G. An Efficient Large Scale Synthesis of Coumarins by a Dealkylative Boron-Mediated Ring Closure of 3-(Ortho-Methoxyaryl) Propenoic Esters. *Syn. Commun.* **1999**, *29*, 929–936.

74. Leonetti, F.; Favia, A.; Rao, A.; Aliano, R.; Paluszcak, A.; Hartmann, R. W.; Carotti, A. Design, Synthesis, and 3D QSAR of Novel Potent and Selective Aromatase Inhibitors. *J. Med. Chem.* **2004**, *47*, 6792–6803.

CHAPTER IV

Engineering a Bacterial P450 for Atroposelective Biaryl Coupling Reactions

Axially chiral biaryl ligands such as BINOL are uniquely capable of mediating an array of mechanistically diverse enantioselective transformations due to hindered rotation around the biaryl axis, giving rise to their status as privileged ligands in asymmetric catalysis. Despite continuous innovations in the application of BINOL derivatives in asymmetric synthesis, the synthesis of the vast majority of BINOL derivatives have remained rather stagnant over the last several decades. These routes rely on the sequential protection and functionalization of enantiopure BINOL over multiple steps to access C_2-symmetric ligands. Despite the promise that non-C_2-symmetric chiral ligands have shown in asymmetric catalysis, their application has been hindered by the synthetic challenge they pose. In this chapter, we disclose a strategy for biocatalytic cross-coupling of naphthol substrates to access non-C_2-symmetric BINOL ligands. Using a bioinformatics-guided approach, we identified several bacterial P450s capable of catalyzing naphthol cross-couplings and engineered one of these P450s to possess the desired reactivity and atroposelectivity in a model reaction. Initial substrate scope analysis of the P450 variants suggests that we have engineered a promiscuous biocatalyst capable of atroposelective cross-coupling reactions with a variety of substituted naphthols. Altogether, this work provides the initial framework for a biocatalytic platform for the streamlined synthesis of chiral non-C_2-symmetric BINOL derivatives.

Portions of this chapter were adapted with permission from Zetzsche, L. E.; Yazarians, J. A.; Chakrabarty, S.; Hinze, M. E.; Murray, L. A. M.; Lukowski, A. L.; Joyce, L. A.; Narayan, A. R. H. Biocatalytic Oxidative Cross-Coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603*, 79–85. © Springer Nature Limited 2022.

INTRODUCTION

Chiral catalysts are powerful tools for conferring stereocontrol in asymmetric synthesis. In particular, axially chiral biaryl ligands are uniquely capable of mediating an array of mechanistically diverse enantioselective transformations due to hindered rotation around the biaryl axis, giving rise to their status as privileged ligands in asymmetric catalysis.[1] Among biaryl ligands, 1,1′-bi-2-naphthol (BINOL, **4.1**) has endured as an extremely versatile ligand following its first application in 1979 in the asymmetric reduction of carbonyls.[2] This landmark discovery spurred the utilization of BINOL in a range of metal-catalyzed reactions, as well as the development of synthetic strategies for derivatization of the BINOL framework to access new chemistries (Figure 4.1).[3, 4] The most popular derivatives of BINOL involve modification of the free phenols access classic biaryl ligands such as BINAP (**4.2**)[5, 6] and chiral phosphoric acids (**4.3**).[7] Additionally, functionalization of the carbon backbone of BINOL can provide beneficial steric and electronic effects in metal-catalyzed reactions, such as in the asymmetric allylation of aldehydes with TRIP (**4.3**)[8] and hetero Diels–Alder cycloaddition with **4.4**.[9] The selection of a chiral ligand such as these for a given reaction is typically an empirical process, requiring the synthesis and screening of large libraries of biaryl ligands.[10] Thus, increasingly streamlined syntheses of this privileged class of ligands are needed, especially for *ortho*-substituted BINOLs.

The most commonly used strategy for accessing substituted BINOLs relies on enantiopure BINOL (**4.1**) as the starting point.[7, 11] Following protection of the free phenolic groups, functional

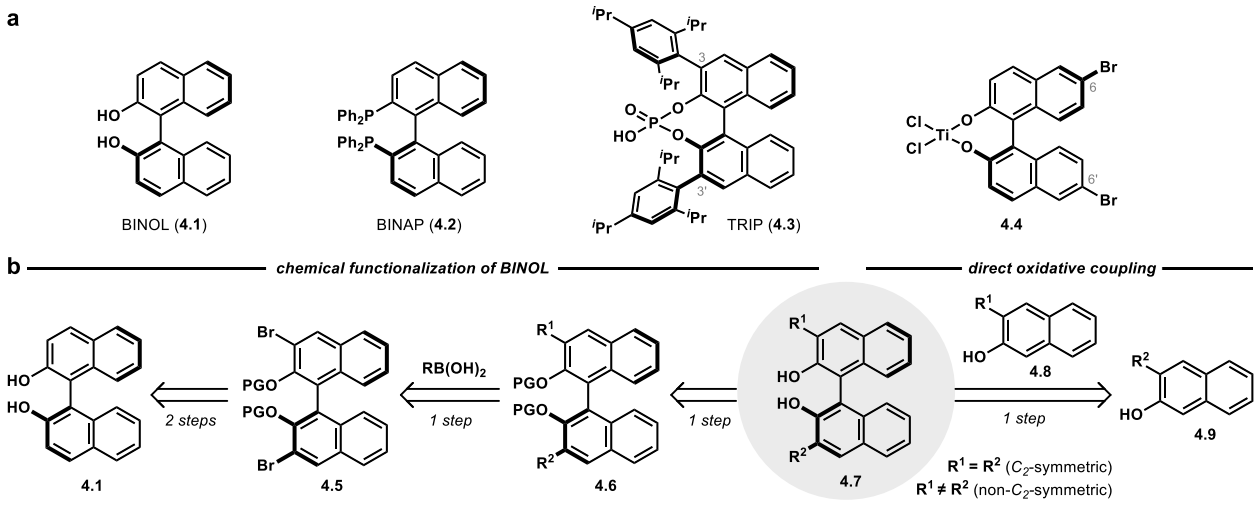


Figure 4.1. Significance and synthesis of chiral biaryl ligands. (a) Biaryl ligands are privileged structures in asymmetric catalysis. (b) The synthesis of *ortho* substituted chiral BINOLs typically relies on protection and functionalization of enantiopure BINOL over multiple steps. In contrast, the direct oxidative coupling of two naphthol substrates can allow for the formation of substituted BINOLs in a single step, albeit with fundamental challenges in controlling the selectivity of the reaction.

handles can be installed *ortho* to the protecting group through lithium-halogenation to provide the dibrominated BINOL (**4.5**). This building block can then be used in Suzuki cross-coupling reactions, often with the installation of aryl groups to increase the steric bulk of the ligand. Finally, deprotection of **4.6** yields BINOL derivatives, such as **4.7**, over a total of four steps from enantiopure material. Alternatively, a more direct and atom-economical route to substituted BINOLs can be accomplished through the direct oxidative coupling of unprotected naphthols (**4.8** and **4.9**) to form BINOL **4.7** in a single step.[12-16] However, the reactivity and selectivity of these reactions is dictated by the inherent stereoelectronic properties of the substrates, thereby restricting the *ortho* substitutions permitted in these transformations.[12, 17, 18] Atroposelectivity can be achieved in the presence of chiral catalysts, however the chemoselectivity of these transformations adds an additional layer of complexity when cross-coupling two non-equivalent naphthol partners, with homocoupling typically outcompeting the desired cross-coupling.[12, 18]

Given the current strategies accessible for the synthesis of BINOL derivatives, the vast majority of chiral biaryl ligands exhibit C_2 symmetry ($R^1 = R^2$), with only rare examples of non-symmetrically substituted ligands ($R^1 \neq R^2$) being applied in asymmetric catalysis.[3, 4, 7] Despite the promise that non-C_2-symmetric chiral ligands have shown in asymmetric catalysis,[19-22] their application has been hindered by the synthetic challenge they pose. We hypothesized that a biocatalytic strategy for the synthesis of chiral BINOL derivatives could overcome some of the limitations inherent to small molecule-mediated routes including protecting group-dependence and low levels of selectivity.[23] Previous work from our lab has shown that cytochrome P450 enzymes can be engineered into highly selective catalysts for oxidative cross-coupling reactions in the formation of chiral biaryl molecules.[24] We anticipated that this platform could be expanded to the cross-coupling of naphthol derivatives to directly access BINOL derivatives. Not only would this streamline the synthesis of BINOL derivatives, but it would also add to the underrepresented class of non-C_2-symmetric biaryl ligands for asymmetric catalysis.

RESULTS AND DISCUSSION

We anticipated that bioinformatic analysis of the natural P450 sequence space would enable the discovery of cytochrome P450 enzymes capable of oxidatively cross-coupling naphthol substrates. As a first step, a sequences similarity network (SSN) was constructed with the entire P450 protein family and then truncated to edges possessing an alignment score greater than 110 with P450s

known to dimerize phenolic substrates (Figure 4.2a).[25-28] Characterized P450s that clustered together (indicating high levels of sequence similarity) dimerized similar natural substrates with complementary site-selectivity. In contrast, characterized enzymes that cluster separately typically

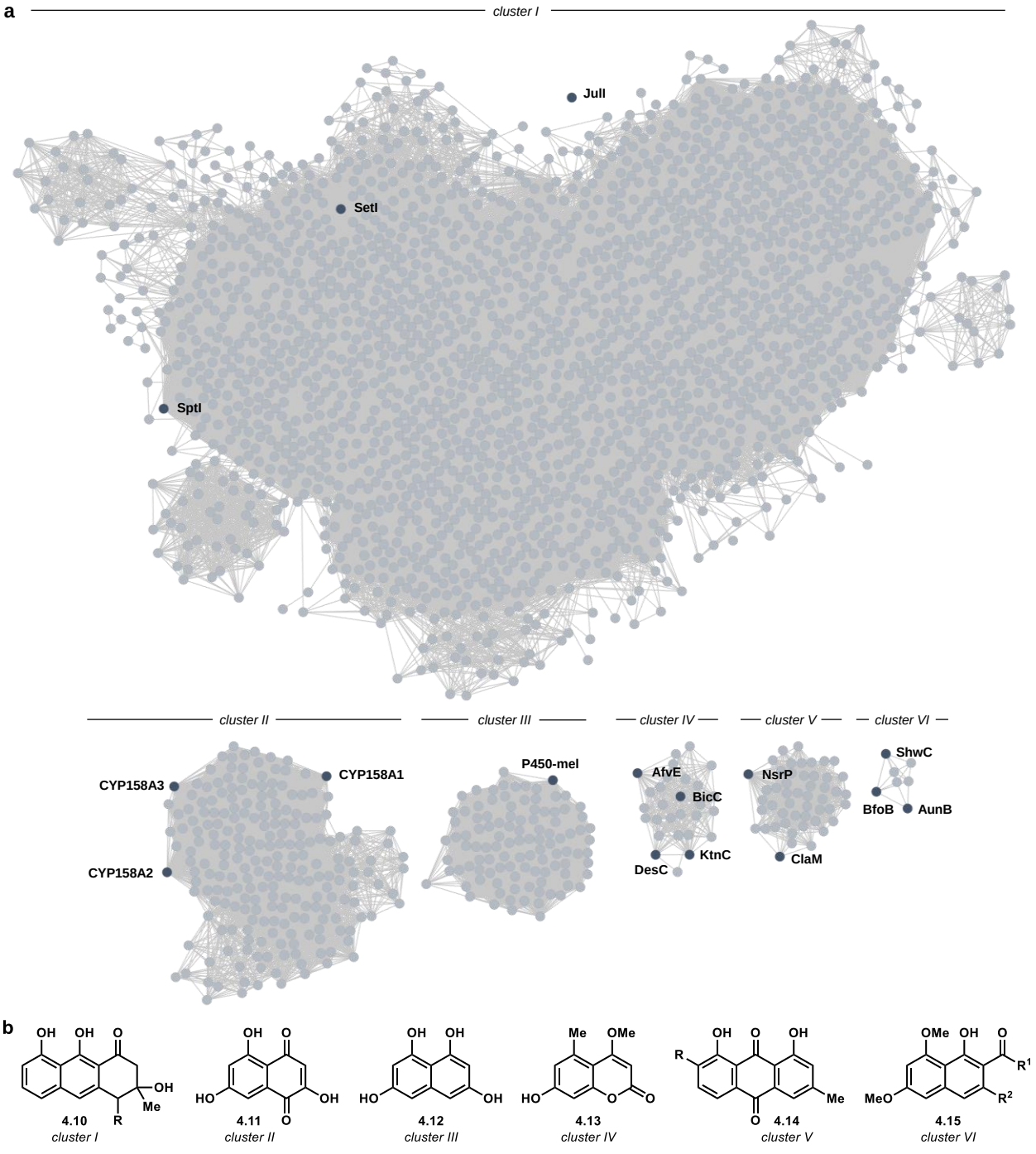


Figure 4.2. Bioinformatic analysis of P450 sequence space. (a) SSN visualizing P450 sequences that cluster with P450s known to catalyze intermolecular oxidative coupling reactions (highlighted with dark blue nodes). SSN parameters: E-value of 106, minimum sequence length of 300 residues, and an alignment score of 95. (b) Shared natural substrate structures of the known P450s in each cluster.

transformed different substrate scaffolds (Figure 4.2b). We hypothesized that we could take advantage of these trends to identify enzymes that function on scaffolds differing from the coumarin **4.13** scaffold natively dimerized by the fungal P450 KtnC.[29] Moreover, the vast majority of the nodes in clusters I–III represented bacterial sequences, offering an opportunity to break away from the *in vivo* biocatalytic platform required for fungal P450s (see Chapters II–III)..

To establish a platform for bacterial P450-catalyzed oxidative coupling reactions, CYP158A2 were selected as a model enzyme given its previous biochemical characterization and known participation in the dimerization and trimerization of flaviolin (**4.11**) in *Streptomyces coelicolor* secondary metabolism.[30] Initial reactions with the purified enzyme in the presence of flavodoxin (Fldx) and flavodoxin reductase (FldR) showed the formation of two dimers and a trimer, as was previously reported by Zhao and coworkers[30] (see Figure 4.8 in Supplemental Information). To further simplify the system and avoid the requirement for redox partners, an artificial fusion enzyme was engineered by tethering the $P450_{RhF}$ reductase domain (RhFRed) to the C-terminus of CYP158A2, thereby reconstituting the electron transport required for catalytic activity (Figure 4.3).[31]

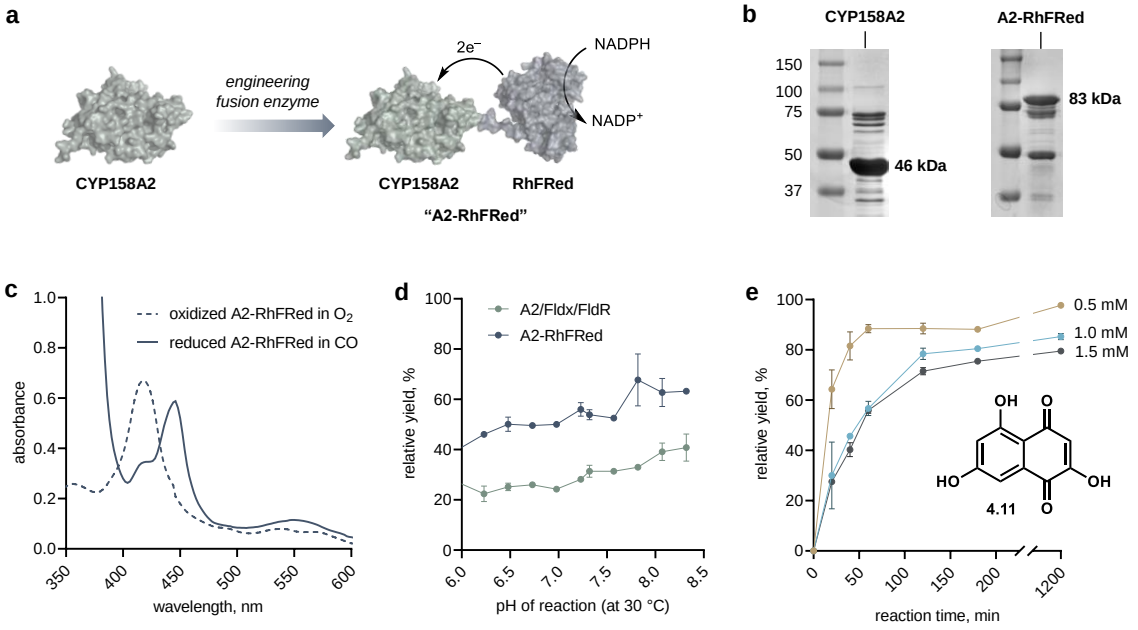


Figure 4.3. Validation of P450-RhFRed engineering strategy for bacterial P450s. (a) Strategy for P450-RhFRed engineering using CYP158A2 as a model system. (b) SDS-PAGE analysis of purified CYP158A2 and CYP158A2-RhFRed (A2-RhFRed) show the expected molecular weights. (c) UV-Vis analysis of purified A2-RhFRed show the characteristic peak at 450 nm for CO-bound heme. (d) Relative dimerization activity of CYP158A2 using a reconstituted Fldx/FldR redox system and A2-RhFRed with native flaviolin (**4.2**) substrate show improved activity with A2-RhFRed. (e) A2-RhFRed-catalyzed oxidative coupling of flaviolin (**4.11**) in the optimized reaction conditions (2.5 μM A2-RhFRed in 50 mM tricine pH 8.3 at 30 °C).

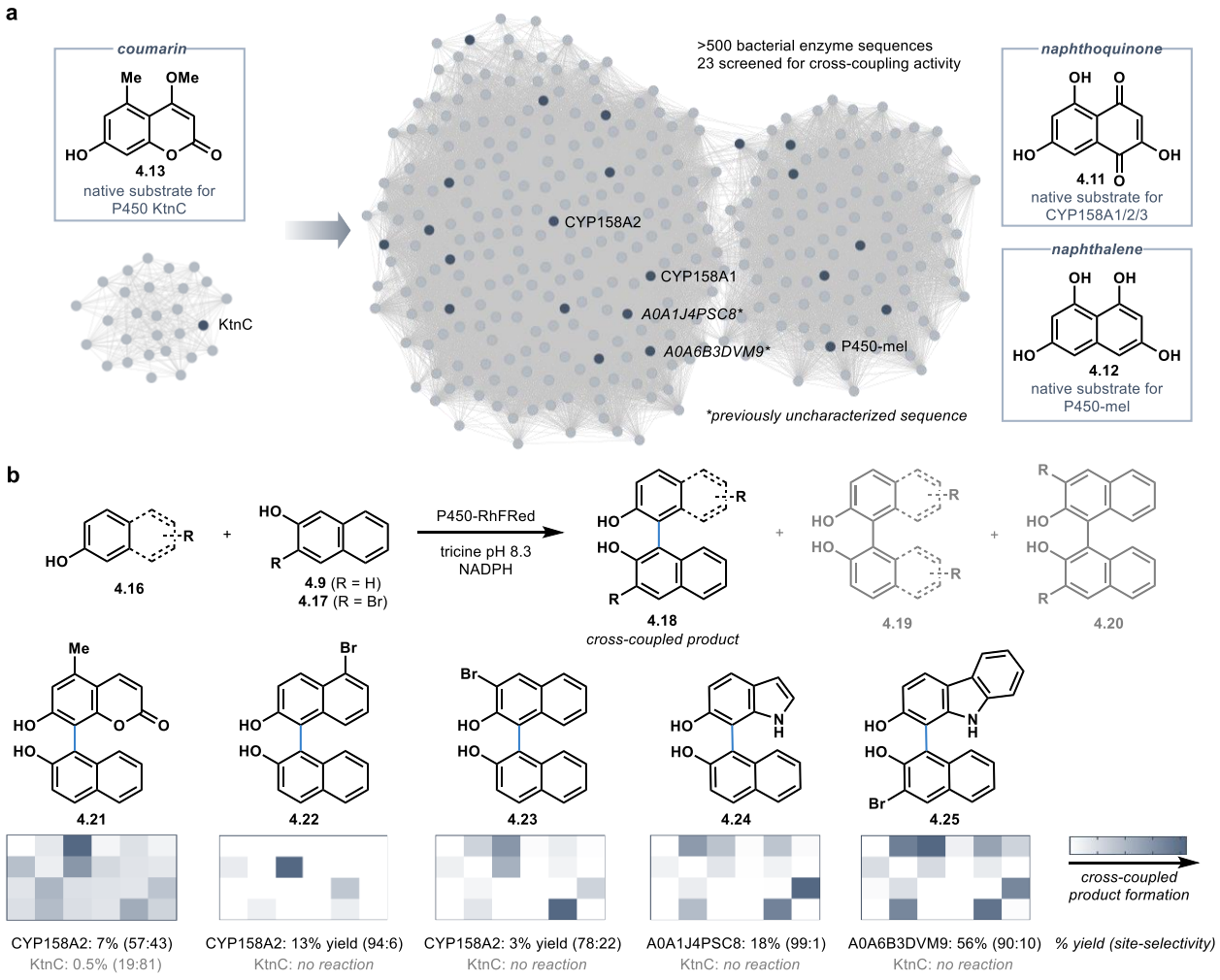


Figure 4.4. Exploring sequence space for P450s with complementary reactivity. (a) Strategy for selection of alternative catalysts for cross-coupling reactions using SSNs. (b) Bacterial fusion enzymes (P450-RhFRed) catalyze cross-coupling reactions with naphthols, demonstrating complementary reactivity to KtnC. The top hits from an initial screen of 23 enzymes were purified for further quantification of reaction yield and selectivity. Reaction conditions: P450-RhFRed in the presence of 5 mM glucose-6-phosphate, 1 mM $NADP^+$, 1 U/mL glucose-6-phosphate dehydrogenase, and substrates in a 1:1 molar ratio in lysate reactions (1 mM **4.7** and 1 mM **4.8** or **4.9**) or 1:3 molar ratio in reactions with 10 μM purified enzyme (100 μM **4.7** and 300 μM **4.8** or **4.9**) in tricine buffer pH 8.3 at 30 °C for 16 h; cross-coupling reactions reported as percent yields to major product with cross-coupling site-selectivity in parentheses.

The fusion constructs validated with CYP158A2 were applied to a panel of 22 other bacterial sequences, which included 20 uncharacterized enzymes as well as three characterized P450s known to dimerize substrates with naphthoquinone and naphthalene cores (see **4.11** and **4.12**) that differed structurally from coumarin **4.13** (Figure 4.4a).[30, 32] We anticipated that these enzymes would productively catalyze oxidative cross-coupling reactions with naphthols to provide a suite of complementary catalysts to KtnC. In a series of *E. coli* lysate screens with the P450-RhFRed library, cross-coupled products formed across a diverse set of phenolic substrates to a naphthol partner with surprising promiscuity (Figure 4.4b). The sequences catalyzing the formation of the

highest levels of cross-coupled product were further characterized using purified enzyme, including two novel sequences with previously uncharacterized function (A0A1J4PSC8 and A0A6B3DVM9). Although the cross-coupling reaction yields varied across the diverse combination of substrates, most of the transformations favored the formation of a major cross-coupled product (see **4.22**–**4.25**), demonstrating promising site-selectivity. In contrast, the atroposelectivity of the biocatalytic cross-coupling reactions were low, with only the cross-coupling to form **4.23** providing any enantioenriched product (65:35 er).

We hypothesized that the low yield and atroposelectivity of the naphthol cross-coupling reactions could be improved with directed evolution. As a starting point, we selected A2-RhFRed due to its ability to catalyze the cross-coupling between naphthols to form non-C_2-symmetric BINOLs (see **4.22** and **4.23**). Additionally, we anticipated that the previously solved crystal structure of CYP158A2 (PDB: 1T93)[30] would provide a valuable resource in the protein engineering process. In a similar strategy as employed in the directed evolution of KtnC (see Chapter III), 96 residues within 12.2 Å of the enzyme active site were selected for site-saturation mutagenesis to generate thousands of variants, each harboring a single active site mutation (Figure 4.5a). Variants generated were screened in a high-throughput manner using *E. coli* cell lysate and

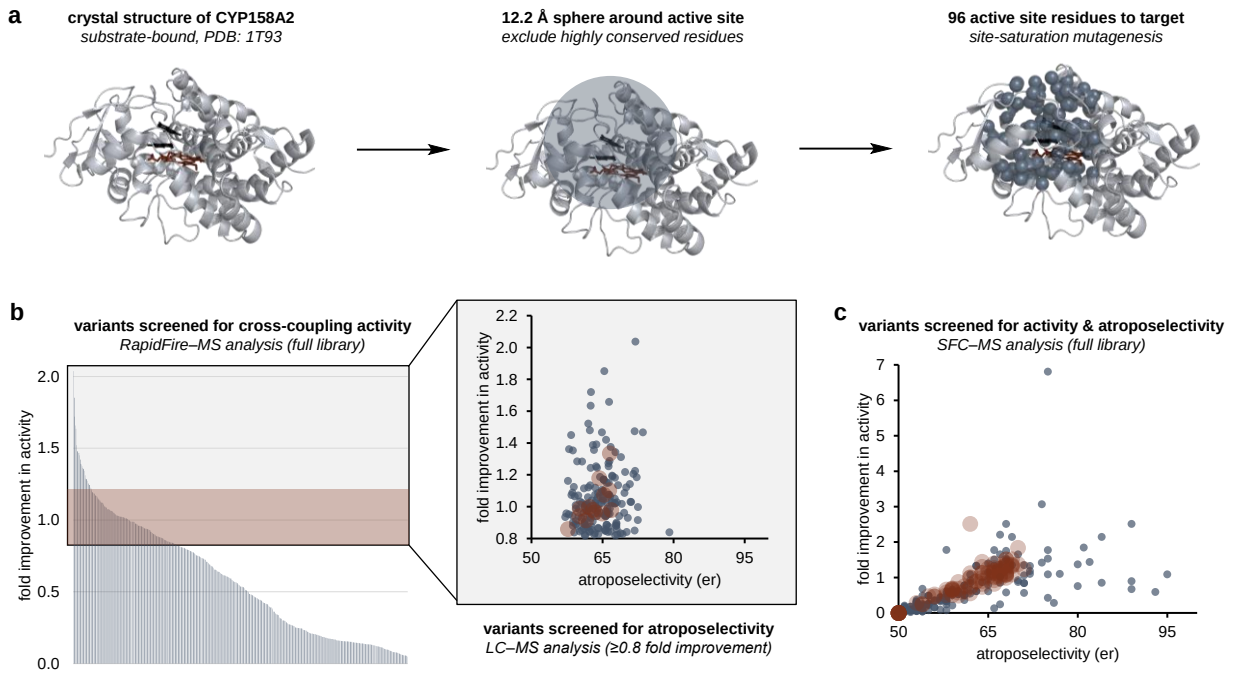


Figure 4.5. Development of a high-throughput engineering strategy. (a) Mutagenesis strategy. (b) In a two-tier screen, RapidFire–MS provided data on the activity of all variants generated, followed by LC–MS analysis on active variants. (c) In a more data-rich approach, an SFC–MS screen was developed by Merck to generate activity and selectivity data for the full library.

each reaction was analyzed for total cross-coupled product formation by RapidFire–mass spectrometry (MS). Reaction mixtures containing higher levels of the cross-coupled product were then analyzed for atroposelectivity in a second-tier screen using a chiral column on LC–MS (Figure 4.5b). Recognizing that a key limitation in the two-tier screen was the exclusion of any variants with low activity but improved atroposelectivity, an alternative single-tier MS-based screen was developed through collaboration with Spencer McMinn and Eugene Kwan at Merck. In this data-rich screen, every reactions from the entire library was subjected to SFC–MS analysis, thereby providing comprehensive data on both reaction yield and atroposelectivity (Figure 4.5c).

Using a combination of these two MS-based screening platforms, we embarked on an engineering campaign targeting the model cross-coupling reaction between 2-naphthol (**4.9**) and 3-bromo-2-naphthol (**4.17**; Figure 4.6). In the first round of evolution, several variants were identified with moderate improvements in activity and atroposelectivity, including a variant with a four-fold improvement in activity that harbored a subsitution (A245Q) on the I-helix. The I-helix is a commonly conserved motif in P450 structures that lines the substrate-binding region of the active site above the heme.[33] Given its proximity to the substrate-binding region, the I-helix is central to the catalytic activity of the enzyme and has been the focus of previous engineering efforts for P450s.[34] This result mirrors the results from the evolution campaign with KtnC, in which mutations along the I-helix were the most impactful for influencing the overall activity of the

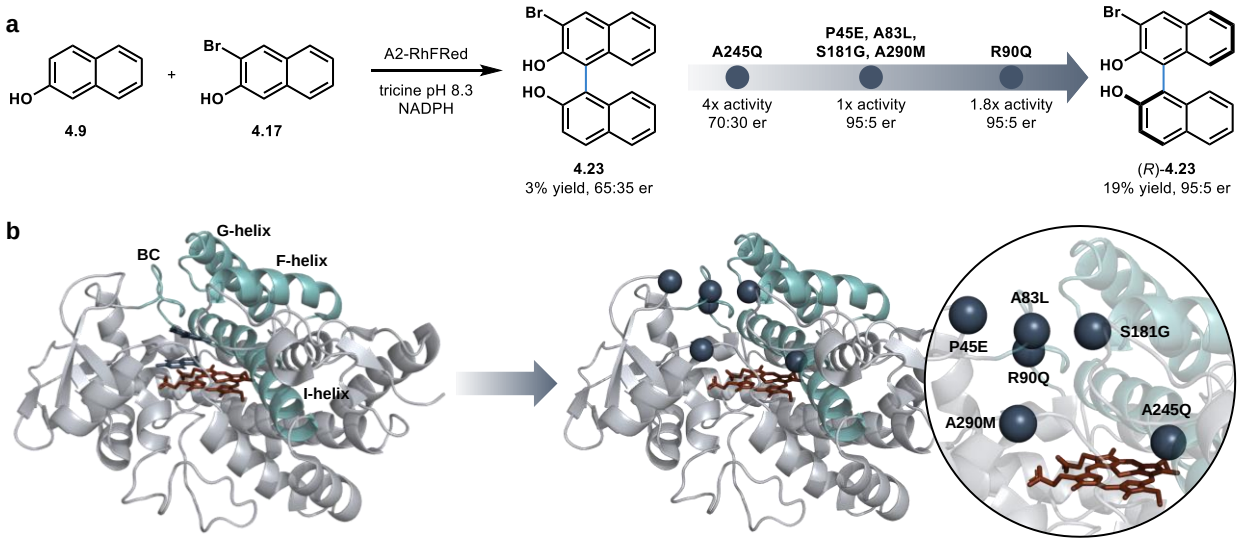


Figure 4.6. Engineering P450 biocatalyst for activity and atroposelectivity. (a) Target cross-coupling reaction for evolution campaign. (b) Flaviolin (**4.11**)-bound crystal structure of wild-type CYP158A2 (PDB: 1T93), highlighting key helices and residues previously demonstrated as important for P450 activity and selectivity. The majority of the mutations incorporated throughout evolution campaign were localized along the flexible active site opening or along the I-helix.

engineered variants.[24]

In a second round of evolution, a library consisting of variants generated with site-scanning saturation mutagenesis on the new template sequence, as well as multiple site-directed mutagenesis, in which beneficial mutations identified in the first round of evolution were folded together. The latter combinatorial strategy proved highly effective at improving the atroposelectivity of the cross-coupling, with the incoporation of four mutations (P45E, A83L, S181G, A290M) drastically increasing the atroposelectivity of the reaction from 70:30 to 95:5 er. Three of these substitutions are located on highly dynamic loop regions near the substrate access that are typically responsible for substrate recognition and binding in P450s.[34] This was phenomenon was also observed in the crystal sturcture of CYP158A2, wherein a significant conformational change occurred primarily in the BC and FG loops (up to 16 Å displacement), with residues in the BC loop dipping into the active site to contact the second substrate.[30]

In a final round of evolution, a site-scanning saturation mutagensis library was generated on the mutated template, targeting the same 96 residues within a 12.2-Å sphere of the active site. Screening this library resulted in the identification of a variant harboring R90Q, which improved the activity 1.8-fold, while maintaining high levels of atroposelectivity. Following the same trend as the previous mutations, this substitution was located on the flexible BC loop involved in substrate recognition and binding. Altogether, these results provide strong evidence that the conformational change that occurs upon substrate binding in CYP158A2 is central to the substrate orientation and therefore selectivity of the cross-coupling reaction.

At this point, we had engineered a variant with moderate activity (19% yield) and excellent atroposelectivity (95:5 er) for the oxidative cross-coupling to form (*R*)-**4.23**. We questioned whether that this engineered biocatalyst could perform additional cross-coupling reactions of naphthol substrates to form additional BINOL derivatives (Figure 4.7). Specifically, we were interested to see whether the atroposelectivity that we had engineered for would be maintained with different naphthol substrates. To investigate this, a panel of *ortho*-substituted naphthols were synthesized, specifically focusing on bulky arenes that are commonly seen BINOL derviatives, especially among chiral phosphoric acids.[7] In a series of cross-coupling reactions with the engineered biocatalyst, a number of dimeric and cross-coupled products were observed, demonstrating promiscuity. Significantly, many of these cross-coupling reactions were atroposelective, especially when one of the two partners was 2-naphthol (**4.9**). Surprisingly, cross-

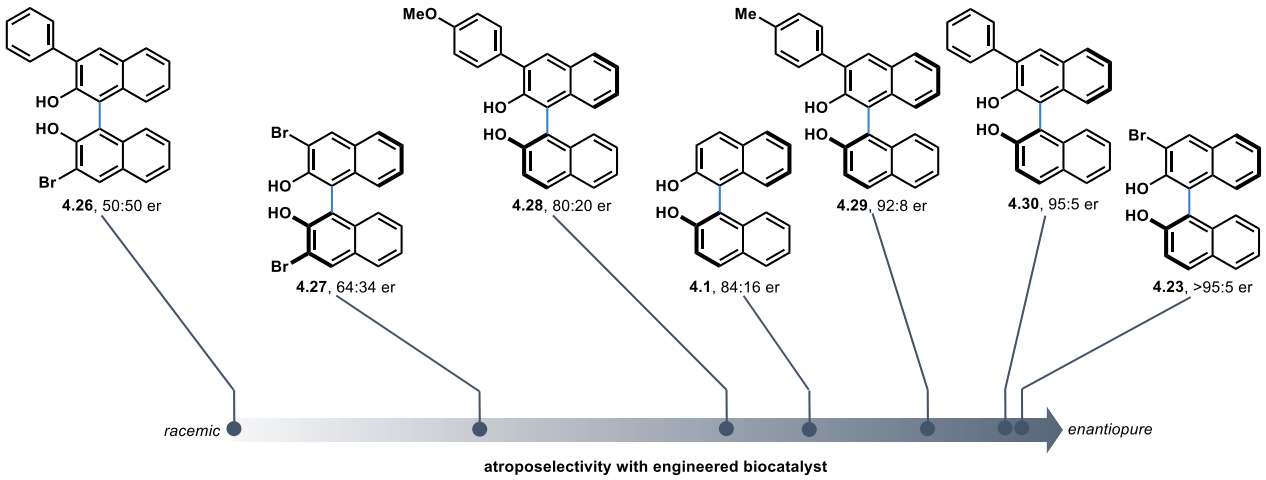


Figure 4.7. Trends in atroposelectivity with different cross-coupling partners. Biocatalytic cross-coupling reactions involving 2-naphthol were atroposelective, providing access to various non-C_2-symmetric (*R*)-BINOLs. In contrast, cross-coupling reactions involving two *ortho*-substituted naphthols resulted in racemic product.

coupling reactions in which both of the coupling partners contained an *ortho* substituent reduced the atroposelectivity of the reaction (see **4.26** and **4.27**).

Based on the activity and selectivity trends observed both with wild-type CYP158A2 and throughout the evolution campaign, a potential model for how (and whether) this enzyme exerts control over the selectivity of the cross-coupling reaction can be speculated. In its substrate-free form, CYP158A2 adopts a more "open" conformation compared to many other P450s that allow the substrates to access the active site.[30] At some point following the binding of the second substrate, a conformational change occurs that both closes the entrance to the active site and orients the second substrate for the desired coupling event. However, a lack of recognition of the second substrate could lead to the enzyme remaining in the open conformation, thereby exerting very little control over the orientation of the second substrate and ultimately the selectivity of the reaction. Throughout the evolution progress, the mutations along this dynamic substrate access channel may have tuned CYP158A2 to specifically recognize and orient 2-naphthol (**4.9**) in the second binding position, thereby enabling an atroposelective cross-coupling. In contrast, lack of recognition of an *ortho*-substituted naphthol in the second binding position may hold the enzyme in the open conformation and therefore confer no atroposelectivity. This theory would also explain the promiscuity observed with wild-type CYP158A2, wherein a diverse set of unnatural substrates could productively be cross-coupled, albeit with minimal to no atroposelectivity.

In conclusion, the work presented in this chapter provides several proof-of-concepts advances in the development of a biocatalytic platform for cross-coupling reactions with naphthol

substrates. First, bioinformatic analysis of natural P450 sequence space through SSNs offers a useful tool for identifying catalysts with diverse cross-coupling reactivity. Second, those natural catalysts can be engineered for a target atroposelectivity through directed evolution, providing access to privileged chiral ligands. In contrast to the most commonly employed strategy to access BINOL derivatives, which involves the protecting group-dependent functionalization of BINOL over a minimum of four steps,[7, 11] this biocatalytic strategy offers access to diverse BINOL derivatives in a single complexity-generating step. Efforts are currently underway to further improve these biocatalysts for higher yields, chemoselectivity, and site-selectivity in atroposelective cross-coupling of substituted naphthols to provide efficient and scalable access to chiral, non-C_2-symmetric BINOL ligands.

EXPERIMENTAL PROCEDURES

Expression and purification of CYP158A2. Chemically competent Lemo21(DE3) *E. coli* cells were transformed with a pET28a(+) plasmid encoding CYP158A2 using standard heat shock procedures and grown on LB agar plates containing 50 µg/mL kanamycin overnight at 37 °C. A single colony was used to inoculate 25 mL of LB media containing 50 µg/mL kanamycin. The culture was incubated overnight at 37 °C, 200 rpm. The overnight culture was used to inoculate 6 x 2.8-L flasks, each containing 1 L of TB media containing 4% glycerol (v/v) and 50 µg/mL kanamycin. The 1-L cultures were incubated at 37 °C, 200 rpm until reaching an optical density at 600 nm (OD600) of ~0.5-0.8. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to 0.5 mM to induce and 5-aminolevulinic acid (5-ALA) was added to 1 mM for heme synthesis. Cultures were incubated at 20 °C, 200 rpm for 20 h and harvested by centrifugation at 4,000 x *g* at 4 °C for 20 min. Harvested cells were stored at –80 °C before lysis and purification.

Harvested cells were thawed and resuspended in ~200 mL of lysis buffer (50 mM tricine pH 8, 500 mM NaCl, 10% glycerol, 0.5 mM EDTA). The cells were lysed by sonication for 5 min total in cycles of 5 seconds on and 10 seconds off. The lysed cell mixture was then centrifuged at 40,000 x *g* for 40 min at 4 °C. The clarified lysate was combined with equilibrated Ni-NTA resin (3-mL bed volume) and incubated on a rocker at 4 °C for 2 h. The resin was collected in a gravity-flow column and washed with 20 mL of lysis buffer containing 10 mM imidazole and 30 mM imidazole each. Enriched His-tagged CYP158A2 was eluted from the resin with up to 20 mL lysis buffer containing 250 mM imidazole. CYP158A2 eluted as a dark red protein. The eluted

CYP158A2 was concentrated and exchanged by dialysis into storage buffer (50 mM tricine pH 8, 150 mM NaCl, 10% glycerol, 0.5 mM EDTA). Total P450 concentration was quantified by carbon monoxide (CO) binding assays[35] and purity was assessed by SDS-PAGE analysis.

Cloning CYP158A2 into A2-RhFRed fusion construct. The open reading frame for *cyp158a2* was amplified with primers that incorporated a 5' NdeI restriction site and a 3' EcoRI restriction site. PCRs were performed with NEB reagents in 50 µL reaction volumes containing 1X HF Phusion buffer, 0.2 mM dNTPs, 0.5 µM each for forward and reverse primers (5'-GAGTTCCATATGATGACTGAAGAGACAATATCACAAGC-3' and 5'-GATAGAATTCCCAGGTGACCGGCAG-3', respectively), 20 units of Phusion DNA polymerase, and 1 µL template plasmid DNA. The reaction conditions were programmed as follows: 98 °C denaturation for 30 sec; 35 cycles of 98 °C for 10 sec, 61 °C for 30 sec, 75 °C for 1.5 min; and a final 75 °C extension for 10 min. PCR products were extracted from a 0.8% agarose gel and cloned into a pET28b vector harboring a gene encoding the RhFRed reductase domain. Digestion reactions contained 3 µg of DNA in 1X Cutsmart buffer and 2 µL each of NdeI and EcoRI restriction enzymes (NEB) in a 50 µL reaction volume. Reactions were incubated at 37 °C for 2.5 h before quenching and purification of the DNA with a PCR clean-up or gel extraction. Ligation reactions contained 100 ng of plasmid DNA and 100 ng of insert were incubated with 1X T4 DNA ligase buffer and 1 unit of T4 DNA ligase enzyme (NEB) in 10 µL reaction volumes for 2 h at room temperature, followed by a 65 °C heat inactivation for 10 min. Ligations were transformed in DH5α *E. coli* cells and transformants were confirmed for gene integration by Sanger sequencing.

Expression and purification of P450-RhFRed. Chemically competent Lemo21(DE3) cells were transformed with a pET28b plasmid encoding P450-RhFRed using standard heat shock procedures and grown on LB agar plates containing 50 µg/mL kanamycin overnight at 37 °C. A single colony was used to inoculate 25 mL of LB media containing 50 µg/mL kanamycin. The culture was incubated overnight at 37 °C, 200 rpm. The overnight culture was used to inoculate 6 x 2.8-L flasks, each containing 1 L of TB media containing 4% glycerol (v/v) and 50 µg/mL kanamycin. The 1-L cultures were incubated at 37 °C, 200 rpm until reaching an optical density at 600 nm (OD600) of ~0.5-0.8. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to 0.5 mM to induce and 5-aminolevulinic acid (5-ALA) was added to 1 mM for heme synthesis. Cultures were incubated at 20 °C, 200 rpm for 20 h and harvested by centrifugation at 4,000 x *g* at 4 °C for

20 min. Harvested cells were stored at –80 °C before lysis and purification. The same purification procedure was used as for CYP158A2.

Reactions with purified P450-RhFRed. 50 µL reactions containing 10 µM P450-RhFRed, 5 mM glucose-6-phosphate, 1 mM $NADP^+$, 1 U/mL glucose-6-phosphate dehydrogenase, and up to 1.5 mM substrate(s) in 50 mM tricine or TES pH 8.3 were prepared. Reactions were incubated at 30 °C for 3 h or 18 h and then quenched with 150 µL methanol containing 1 µM 2,6-dihydroxyacetophenone (internal standard). All reactions were performed in duplicate.

Expression of P450-RhFRed libraries in 96-well plates. Chemically competent Lemo21(DE3) cells were transformed with a pET28b plasmid encoding P450-RhFRed using standard heat shock procedures. The transformations were used to directly inoculate 400 µL LB containing 50 µg/mL kanamycin in a 96-well plate and grown overnight at 37 °C, 350 rpm. The overnight cultures (10 µL each) were used to inoculate 400 µL TB containing 4% glycerol (v/v) and 50 µg/mL kanamycin. The cultures were incubated at 37 °C, 350 rpm until reaching an optical density at 600 nm (OD600) of ~0.8–1.0. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to 0.5 mM to induce expression and 5-aminolevulinic acid (5-ALA) was added to 1 mM for heme synthesis. Cultures were incubated at 20 °C, 350 rpm for 20 h and harvested by centrifugation at 1,000 x *g* at 4 °C for 15 min. Plates containing harvested cells were stored at –80 °C before lysate reactions.

Table 4.1. Panel of 23 enzymes identified in SSN and screened in lysate reactions. Protein sequences are included in the Supplementary Information.

	1	2	3	4	5	6
A	*negative*	CYP158A1	**CYP158A2**	P450-mel	A0A4R7HLN1	A0A6H0CVT5
B	A0A221NTD0	A0A345T2R4	A0A4Y3VFB2	A0A6I6FHD3	A0A0B5DS97	A0A1C5CYG0
C	A0A0C2AVS8	A0A1G8ZS81	A0A3N6F9N6	A0A2M9JRU5	A0A0N9I1K4	**A0A1J4PSC8**
D	A0A4D4MK74	**A0A6B3DVM9**	A0A2P8Q811	A0A6G3S9J1	A0A1G9MX20	A0A2G7EEB4

High-throughput reactions with P450-RhFRed lysate. 96-well plates containing harvested cells were thawed at room temperature and resuspended in 250–300 µL of plate lysis buffer (50 mM tricine or TES pH 8.3, 150 mM NaCl, 10% glycerol, 0.5 mM EDTA, 2 mg/mL lysozyme, 100 µM PMSF). The cells were enzymatically lysed by incubation at 25 °C, 350 rpm for 2 h. The lysed plates were then centrifuged at 1,000 x *g* for 30 min at 4 °C. The clarified lysate (50 µL) was

transferred to a new plate with each well containing 50 µL reaction mix (5 mM glucose-6-phosphate, 1 mM $NADP^+$, 1 U/mL glucose-6-phosphate dehydrogenase, 1 mM substrate A, and 1 mM substrate B in 50 mM tricine pH 8.3). Reactions were incubated at 30 °C for ~18 h and then quenched with 400 µL methanol containing 1 µM 2,6-dihydroxyacetophenone (internal standard).

High-throughput RapidFire–MS screen for activity. Library reactions were analyzed for total cross-coupling activity in a high-throughput tier 1 screening using an Agilent RapidFire 365 High-throughput Mass Spectrometry System linked to an Agilent 6545 LC–Q-TOF. Library samples were prepared by diluting the library samples 2:1 in water containing 600 nM 2,6′-dihydroxyacetophenone (internal standard) in shallow 96-well plates (Axygen) and heat-sealed with pierceable aluminum covers (Agilent). Each RapidFire injection was programmed as follows: 0.6 s to aspirate the sample, 3 s to load the sample onto the C4 cartridge with water (0.5 mL/min), 8 s to elute the sample from the cartridge using 90:10 acetonitrile/water (v/v) buffered with 10 mM ammonium formate pH 8 (0.8 mL/min), and 0.5 s to reequilibrate. A blank injection was included after each sample injection to reduce carry-over between library reactions.

The peak areas for the extracted ion chromatograms for internal standard, substrate, and cross-coupled product were collected for each sample using the MassHunter Qualitative Analysis 10.0 software. The relative percent conversions to cross-coupled products were calculated and the average percent conversion resulting from the template reactions for each plate (n=10) was set to 1.0. The conversions for the variant reactions were normalized to the template reactions, giving fold improvement scores for each variant.

LC–MS analysis of biocatalytic reactions. Quenched reactions were filtered through Pall AcroPrep Advance 350 µL 0.2 um GHP Short Tip Natural PP 96-well filter plates by centrifugation (1,000 x *g* for 5 minutes). The samples were subjected to liquid chromatography PDA spectrometry (UPLC) analysis performed on an Agilent 6230 time of flight mass spectrometer with a Dual AJS ESI source and an Agilent 1290 Infinity Series II diode array detector, autosampler, and binary pump. Reactions were subjected to liquid chromatography PDA spectrometry (UPLC) analysis per-formed on an Agilent 6230 time of flight mass spectrometer with a Dual AJS ESI source and an Agilent 1290 Infinity Series II diode array detector, autosampler, and binary pump. The MassHunter software was used to extract ion chromatograms at the target masses to characterize the percent yields and selectivity of the reactions.

To determine site-selectivity of reaction, samples were analyzed by negative mode on a Waters Acquity Premier HSS T3 1.8 μm C18, 2.1 x 50 mm column. Phase A = 95:5 water/acetonitrile (v/v); phase B = 95:5 acetonitrile/water (v/v). The pump was programmed as follows: 10% B held for 0.5 min, to 40% B over 2.0 min, to 55% B over 1.75 min, to 75% B over 0.75 min, to 100% B over 0.5 min, hold at 100% B for 1.0 min. The flow rate was set to 0.7 mL/min flow rate, the column was heated to 45 °C, and each injection was followed by equilibration at 20% B for 0.5 min.

To determine atroposelectivity of reaction, samples were analyzed by negative mode on a Daicel Chiralpak IC-3, 4.6 x 150 mm, 3 μm column. Phase A = 95:5 water/acetonitrile (v/v); phase B = 95:5 acetonitrile/water (v/v). The pump was programmed as one of the two following methods: (1) 80% B held for 2.5 min, to 100% B over 0.1 min, 100% B held for 0.9 min or (2) 20% to 100% B over 10.0 min, 100% B held for 2 min. For both methods, the flow rate was set to 1.0 mL/min flow rate and each injection was followed by equilibration for 0.5 min.

Design and generation of libraries for site-saturation mutagenesis. All residues within 12.2 Å of the flaviolin molecules bound in the active site of the CYP158A2 crystal structure (PDB 1T93) were selected in PyMOL (command: "select bca. 1T93 within 12.2 of flaviolin"). Any residues with >95% conservation based on a MAFFT alignment with up to 5,000 other sequences were removed from the pool of selected residues, giving a total of 96 residues with ≤95% conservation (see Figure 4.5). Degenerate forward and reverse primers were designed for each of the 96 active site residues being targeted and purchased from IDT (Table 4.2). Blunt-end whole-plasmid PCRs were set up in a 96-well PCR plate, with each 25-uL reaction containing 1X HF Phusion buffer, 4% DMSO, 400 μM dNTPs, 0.1 ng/μL template plasmid, 0.4 units Phusion DNA polymerase, 0.8 μM degenerate forward primer, and 0.8 μM reverse primer. The reaction conditions were programmed as follows: 98 °C denaturation for 2 min; 25 cycles of 98 °C for 20 sec, 58 °C for 20 sec, 72 °C for 4.5 min; and a final 72 °C extension for 10 min. After amplification, 10 μL of each PCR product was pooled in a single tube. The crude PCR product mixture was purified by gel extraction from a 0.8% agarose gel using Qiagen spin columns. Any residual template DNA was digested in 1X Cutsmart buffer with 5 μL DpnI (NEB) overnight at 37 °C and removed with a PCR clean-up using Qiagen spin columns. The linear DNA fragments were phosphorylated in 1X Quick ligase (NEB) buffer with 2 μL PNK enzyme (NEB) at 37 °C for 30 min. Quick ligase en-

zyme (1 μL, NEB) was added directly to the phosphorylation reaction and incubated at room temperature for 15 min. Salts were removed from the ligation reaction with a PCR clean-up using Qiagen spin columns. High efficiency electrocompetent 10-beta electrocompetent *E. coli* cells (NEB) were transformed with 2 μL of the clean DNA using standard electroporation protocols. Resulting colonies were collected and miniprepped to give a plasmid DNA library. Chemically competent Lemo21(DE3) cells were transformed with the plasmid DNA library and resulting colonies were used directly in library screening.

Generation of combinatorial libraries using multiple site-directed mutagenesis. Forward primers were designed to flank desired mutation with >15 complementary nucleotides on either side of mutation and purchased from IDT (Table 4.3). Equimolar amounts of each of the forward primers were combined and diluted down to 10 μM. PCRs were performed with the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent) according to their recommended protocols. After amplification, 2 μL of the supplied DpnI was added and the digestion reaction was incubated at 37 °C for 1 h. High efficiency electrocompetent 10-beta electrocompetent *E. coli* cells (NEB) were transformed with 2 μL of the clean DNA using standard electroporation protocols. Resulting colonies were collected and miniprepped to give a plasmid DNA library. Chemically competent Lemo21(DE3) cells were transformed with the plasmid DNA library and resulting colonies were used directly in library screening.

High-throughput screening and analysis of libraries. Libraries were screened as lysate reactions in 96-well plates (as previously described; see Table 4.4 for summary of each round). Reaction products were subjected to mass spectrometry analysis, either through a combination of RapidFire–MS and LC–MS (as previously described) or SFC–MS analysis (performed by Spencer McMinn and analyzed by Eugene Kwan at Merck). Hit variants were validated by growing up glycerol stocks of the respective cells harboring the variants, repeating lysate reactions, and analyzing reactions by LC–MS. Variants that showed improvement over the template after validation were sequenced to identify the active site mutations. The sequence for the top variant from each round was used as the template for the following round.

Table 4.2. Primers used for site-saturation mutagenesis of CYP158A2 active site residues. Primers from round 1 (Rd1) were reused in subsequent rounds unless the sequence lost complementarity to the template DNA due to new mutations being carried forward. Codons that were edited to accommodate mutations added to the template DNA are highlighted in red.

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
A18X	Rd1-fwd	NNKGTGGACCTGCCGGGTTC	Rd1-rvs	CGGCCAGTCGCGAAC
L44X	Rd1-fwd	NNKCCGAATGGTGAAGGTTGGG	Rd1-rvs	ACTGATACGAGTCACTGGACC
	Rd4-fwd	NNKGAGAATGGTGAAGGTTGGG	Rd1-rvs	
P45X	Rd1-fwd	NNKAATGGTGAAGGTTGGGCTTG	Rd1-rvs	CAGACTGATACGAGTCACTGGAC
N46X	Rd1-fwd	NNKGGTGAAGGTTGGGCTTGG	Rd1-rvs	CGGCAGACTGATACGAGTCAC
			Rd4-rvs	CTCCAGACTGATACGAGTCAC
L53X	Rd1-fwd	NNKGTGACTCGTCACGATGATGTTC	Rd1-rvs	CCAAGCCCAACCTTCACC
V54X	Rd1-fwd	NNKACTCGTCACGATGATGTTCG	Rd1-rvs	CAACCAAGCCCAACCTTCAC
T55X	Rd1-fwd	NNKCGTCACGATGATGTTCGCC	Rd1-rvs	CACCAACCAAGCCCAACC
V60X	Rd1-fwd	NNKCGCCTAGTGACCAATGATCC	Rd1-rvs	ATCATCGTGACGAGTCACCAAC
V63X	Rd1-fwd	NNKACCAATGATCCGCGTTTTGG	Rd1-rvs	TAGGCGAACATCATCGTGACG
T64X	Rd1-fwd	NNKAATGATCCGCGTTTTGGCC	Rd1-rvs	CACTAGGCGAACATCATCGTG
G70X	Rd1-fwd	NNKCGCGAGGCGGTTATGG	Rd1-rvs	AAAACGCGGATCATTGGTCAC
R71X	Rd1-fwd	NNKGAGGCGGTTATGGACCGTC	Rd1-rvs	GCCAAAACGCGGATCATTG
V74X	Rd1-fwd	NNKATGGACCGTCAAGTTACGC	Rd1-rvs	CGCCTCGCGGCCAAAAC
M75X	Rd1-fwd	NNKGACCGTCAAGTTACGCGTTTG	Rd1-rvs	AACCGCCTCGCGGCC
V79X	Rd1-fwd	NNKACGCGTTTGGCGCCTC	Rd1-rvs	TTGACGGTCCATAACCGCC
	Rd4-fwd	NNKACGCGTTTGCTGCCTC		
T80X	Rd1-fwd	NNKCGTTTGGCGCCTCACTTTATC	Rd1-rvs	AACTTGACGGTCCATAACCG
	Rd4-fwd	NNKCGTTTGCTGCCTCACTTTATC		
R81X	Rd1-fwd	NNKTTGGCGCCTCACTTTATCC	Rd1-rvs	CGTAACTTGACGGTCCATAACC
	Rd4-fwd	NNKTTGCTGCCTCACTTTATCC		
L82X	Rd1-fwd	NNKGCGCCTCACTTTATCCCG	Rd1-rvs	ACGCGTAACTTGACGGTCC
	Rd4-fwd	NNKCTGCCTCACTTTATCCCG		
A83X	Rd1-fwd	NNKCCTCACTTTATCCCGGCGC	Rd1-rvs	CAAACGCGTAACTTGACGGTC
P84X	Rd1-fwd	NNKCACTTTATCCCGGCGCG	Rd1-rvs	CGCCAAACGCGTAACTTGAC
	Rd5-fwd	NNKCACTTTATCCCGGCGCA	Rd4-rvs	CAGCAAACGCGTAACTTGAC
H85X	Rd1-fwd	NNKTTTATCCCGGCGCGTGG	Rd1-rvs	AGGCGCCAAACGCGTAAC
	Rd5-fwd	NNKTTTATCCCGGCGCAGGG	Rd4-rvs	AGGCAGCAAACGCGTAAC
F86X	Rd1-fwd	NNKATCCCGGCGCGTGGC	Rd1-rvs	GTGAGGCGCCAAACGC
	Rd5-fwd	NNKATCCCGGCGCAGGGC	Rd4-rvs	GTGAGGCAGCAAACGC
I87X	Rd1-fwd	NNKCCGGCGCGTGGCGCG	Rd1-rvs	AAAGTGAGGCGCCAAACGCGTAAC
			Rd4-rvs	AAAGTGAGGCAGCAAACGCGTAAC
P88X	Rd1-fwd	NNKGCGCGTGGCGCGGTAG	Rd1-rvs	GATAAAGTGAGGCGCCAAACG
	Rd5-fwd	NNKGCGCAGGGCGCGGTAG	Rd4-rvs	GATAAAGTGAGGCAGCAAACG
A89X	Rd1-fwd	NNKCGTGGCGCGGTAGGC	Rd1-rvs	CGGGATAAAGTGAGGCGC
	Rd5-fwd	NNKCAGGGCGCGGTAGGC	Rd4-rvs	CGGGATAAAGTGAGGCAG
R90X	Rd1-fwd	NNKGGCGCGGTAGGCTTCC	Rd1-rvs	CGCCGGGATAAAGTGAGGC
G91X	Rd1-fwd	NNKGCGGTAGGCTTCCTGGAC	Rd1-rvs	ACGCGCCGGGATAAAGTG
A92X	Rd1-fwd	NNKGTAGGCTTCCTGGACCCG	Rd1-rvs	GCCACGCGCCGGGATAAAG
V93X	Rd1-fwd	NNKGGCTTCCTGGACCCGC	Rd1-rvs	CGCGCCACGCGCCGG
G94X	Rd1-fwd	NNKTTCCTGGACCCGCCG	Rd1-rvs	TACCGCGCCACGCGC
F95X	Rd1-fwd	NNKCTGGACCCGCCGGATC	Rd1-rvs	GCCTACCGCGCCACG
L96X	Rd1-fwd	NNKGACCCGCCGGATCACAC	Rd1-rvs	GAAGCCTACCGCGCCAC
W174X	Rd1-fwd	NNKACCCAGCTGATTCTGTCTAGC	Rd1-rvs	GGTATGCATGCTGTGACGATC
	Rd4-fwd	NNKACCCAGCTGATTCTGTCTGGC		
T175X	Rd1-fwd	NNKCAGCTGATTCTGTCTAGCAGC	Rd1-rvs	CCAGGTATGCATGCTGTGAC
	Rd4-fwd	NNKCAGCTGATTCTGTCTGGCAGC		
Q176X	Rd1-fwd	NNKCTGATTCTGTCTAGCAGCCATG	Rd1-rvs	GGTCCAGGTATGCATGCTG
	Rd4-fwd	NNKCTGATTCTGTCTGGCAGCCATG		
L177X	Rd1-fwd	NNKATTCTGTCTAGCAGCCATGGC	Rd1-rvs	CTGGGTCCAGGTATGCATG
	Rd4-fwd	NNKATTCTGTCTGGCAGCCATGGC		
I178X	Rd1-fwd	NNKCTGTCTAGCAGCCATGGC	Rd1-rvs	CAGCTGGGTCCAGGTATGC
	Rd4-fwd	NNKCTGTCTGGCAGCCATGGC		
L179X	Rd1-fwd	NNKTCTAGCAGCCATGGCGC	Rd1-rvs	AATCAGCTGGGTCCAGGTATG
	Rd4-fwd	NNKTCTGGCAGCCATGGCGC		
S180X	Rd1-fwd	NNKAGCAGCCATGGCGCG	Rd1-rvs	CAGAATCAGCTGGGTCCAG

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
	Rd4-fwd	NNKGGCAGCCATGGCGCG		
S181X	Rd1-fwd	NNKAGCCATGGCGCGGAG	Rd1-rvs	AGACAGAATCAGCTGGGTCC
S182X	Rd1-fwd	NNKCATGGCGCGGAGGTG	Rd1-rvs	GCTAGACAGAATCAGCTGGG
			Rd4-rvs	GCCAGACAGAATCAGCTGGG
H183X	Rd1-fwd	NNKGGCGCGGAGGTGAGC	Rd1-rvs	GCTGCTAGACAGAATCAGCTG
			Rd4-rvs	GCTGCCAGACAGAATCAGCTG
G234X	Rd1-fwd	NNKTTGGCGGTTCTGTTGCAAATTG	Rd1-rvs	GACAGCCTCAGACAGGGTG
L235X	Rd1-fwd	NNKGCGGTTCTGTTGCAAATTGG	Rd1-rvs	ACCGACAGCCTCAGACAG
A236X	Rd1-fwd	NNKGTTCTGTTGCAAATTGGCGG	Rd1-rvs	CAAACCGACAGCCTCAGAC
V237X	Rd1-fwd	NNKCTGTTGCAAATTGGCGGTG	Rd1-rvs	CGCCAAACCGACAGCC
L238X	Rd1-fwd	NNKTTGCAAATTGGCGGTGAGG	Rd1-rvs	AACCGCCAAACCGACAG
	Rd2-fwd	NNKTTGCAAATTGGCGGTGAGC		
L239X	Rd1-fwd	NNKCAAATTGGCGGTGAGGCC	Rd1-rvs	CAGAACCGCCAAACCGAC
	Rd2-fwd	NNKCAAATTGGCGGTGAGCAG		
Q240X	Rd1-fwd	NNKATTGGCGGTGAGGCCG	Rd1-rvs	CAACAGAACCGCCAAACCG
	Rd2-fwd	NNKATTGGCGGTGAGCAGG		
I241X	Rd1-fwd	NNKGGCGGTGAGGCCGTG	Rd1-rvs	TTGCAACAGAACCGCCAAAC
	Rd2-fwd	NNKGGCGGTGAGCAGGTG		
G242X	Rd1-fwd	NNKGGTGAGGCCGTGACGAAC	Rd1-rvs	AATTTGCAACAGAACCGCCAAAC
	Rd2-fwd	NNKGGTGAGCAGGTGACGAAC		
G243X	Rd1-fwd	NNKGAGGCCGTGACGAACAAC	Rd1-rvs	GCCAATTTGCAACAGAACCG
	Rd2-fwd	NNKGAGCAGGTGACGAACAAC		
E244X	Rd1-fwd	NNKGCCGTGACGAACAACAGC	Rd1-rvs	ACCGCCAATTTGCAACAGAAC
	Rd2-fwd	NNKCAGGTGACGAACAACAGC		
A245X	Rd1-fwd	NNKGTGACGAACAACAGCGGTC	Rd1-rvs	CTCACCGCCAATTTGCAACAG
V246X	Rd1-fwd	NNKACGAACAACAGCGGTCAG	Rd1-rvs	GGCCTCACCGCCAATTTG
			Rd2-rvs	CTGCTCACCGCCAATTTG
N248X	Rd1-fwd	NNKAACAGCGGTCAGATGTTTCATC	Rd1-rvs	CGTCACGGCCTCACCG
			Rd2-rvs	CGTCACCTGCTCACCG
N249X	Rd1-fwd	NNKAGCGGTCAGATGTTTCATCTG	Rd1-rvs	GTTCGTCACGGCCTCAC
			Rd2-rvs	GTTCGTCACCTGCTCAC
I285X	Rd1-fwd	NNKCCGCACCGTAATGCTGTC	Rd1-rvs	CCAACGCAGTAACTCATCGATC
	Rd4-fwd	NNKCCGCACCGTAATATGGTC		
P286X	Rd1-fwd	NNKCACCGTAATGCTGTCGGAC	Rd1-rvs	AATCCAACGCAGTAACTCATCG
	Rd4-fwd	NNKCACCGTAATATGGTCGGAC		
H287X	Rd1-fwd	NNKCGTAATGCTGTCGGACTGAG	Rd1-rvs	CGGAATCCAACGCAGTAACTC
	Rd4-fwd	NNKCGTAATATGGTCGGACTGAG		
R288X	Rd1-fwd	NNKAATGCTGTCGGACTGAGCC	Rd1-rvs	GTGCGGAATCCAACGCAG
	Rd4-fwd	NNKAATATGGTCGGACTGAGCC		
N289X	Rd1-fwd	NNKGCTGTCGGACTGAGCCG	Rd1-rvs	ACGGTGCGGAATCCAAC
	Rd4-fwd	NNKATGGTCGGACTGAGCCG		
A290X	Rd1-fwd	NNKGTCGGACTGAGCCGTATCG	Rd1-rvs	ATTACGGTGCGGAATCCAAC
V291X	Rd1-fwd	NNKGGACTGAGCCGTATCGCG	Rd1-rvs	AGCATTACGGTGCGGAATC
			Rd4-rvs	CATATTACGGTGCGGAATC
G292X	Rd1-fwd	NNKCTGAGCCGTATCGCGTTG	Rd1-rvs	GACAGCATTACGGTGCGG
			Rd4-rvs	GACCATATTACGGTGCGG
L293X	Rd1-fwd	NNKAGCCGTATCGCGTTGGAAG	Rd1-rvs	TCCGACAGCATTACGGTGC
			Rd4-rvs	TCCGACCATATTACGGTGC
S294X	Rd1-fwd	NNKCGTATCGCGTTGGAAGATGTG	Rd1-rvs	CAGTCCGACAGCATTACGG
			Rd4-rvs	CAGTCCGACCATATTACGG
R295X	Rd1-fwd	NNKATCGCGTTGGAAGATGTGG	Rd1-rvs	GCTCAGTCCGACAGCATTAC
			Rd4-rvs	GCTCAGTCCGACCATATTAC
A313X	Rd1-fwd	NNKGTGTACGTGAGCTATCTGGC	Rd1-rvs	ATCACCCGCCCGGATAC
V314X	Rd1-fwd	NNKTACGTGAGCTATCTGGCTGC	Rd1-rvs	TGCATCACCCGCCCG
Y315X	Rd1-fwd	NNKGTGAGCTATCTGGCTGCG	Rd1-rvs	GTATGCATCACCCGCCC
V316X	Rd1-fwd	NNKAGCTATCTGGCTGCGAACC	Rd1-rvs	CACGTATGCATCACCCGC
S317X	Rd1-fwd	NNKTATCTGGCTGCGAACCGTG	Rd1-rvs	GCTCACGTATGCATCACCC
Y318X	Rd1-fwd	NNKCTGGCTGCGAACCGTG	Rd1-rvs	ATAGCTCACGTATGCATCACC
L319X	Rd1-fwd	NNKGCTGCGAACCGTGACC	Rd1-rvs	GACATAGCTCACGTATGCATCAC
A320X	Rd1-fwd	NNKGCGAACCGTGACCCGG	Rd1-rvs	AGCGACATAGCTCACGTATGC
A321X	Rd1-fwd	NNKAACCGTGACCCGGAGG	Rd1-rvs	CGCAGCGACATAGCTCAC

Mutation	Forward primer (5′ to 3′)		Reverse primer (5′ to 3′)	
V344X	Rd1-fwd	NNKTCTTTCGGTTTCGGCCCG	Rd1-rvs	ATGCGGGTTCGGTGATCTTTC
S345X	Rd1-fwd	NNKTTCGGTTTCGGCCCGC	Rd1-rvs	AACATGCGGGTTCGGTG
F346X	Rd1-fwd	NNKGGTTTCGGCCCGCAC	Rd1-rvs	GAAAACATGCGGGTTCGGTG
G347X	Rd1-fwd	NNKTTCGGCCCGCACTATTGC	Rd1-rvs	ACCGAAAACATGCGGGTTC
F348X	Rd1-fwd	NNKGGCCCGCACTATTGCC	Rd1-rvs	GAAACCGAAAACATGCGGG
G349X	Rd1-fwd	NNKCCGCACTATTGCCCGG	Rd1-rvs	GCCGAAACCGAAAACATGC
P350X	Rd1-fwd	NNKCACTATTGCCCGGGCG	Rd1-rvs	CGGGCCGAAACCGAAAAC
H351X	Rd1-fwd	NNKTATTGCCCGGGCGGC	Rd1-rvs	GTGCGGGCCGAAACC
Y352X	Rd1-fwd	NNKTGCCCGGGCGGCATG	Rd1-rvs	ATAGTGCGGGCCGAAACC
C353X	Rd1-fwd	NNKCCGGGCGGCATGCTC	Rd1-rvs	GCAATAGTGCGGGCCG
P354X	Rd1-fwd	NNKGGCGGCATGCTCGCC	Rd1-rvs	CGGGCAATAGTGCGGG
G355X	Rd1-fwd	NNKGGCATGCTCGCCCGTC	Rd1-rvs	GCCCGGGCAATAGTGC
G356X	Rd1-fwd	NNKATGCTCGCCCGTCTCG	Rd1-rvs	GCCGCCCGGGCAATAG
A358X	Rd1-fwd	NNKCGTCTCGAAAGCGAATTGTTG	Rd1-rvs	GAGCATGCCGCCCGG
G391X	Rd1-fwd	NNKGCGCTGATCCGTGGTC	Rd1-rvs	CTTTTTGAACGGAACATCTTCCG
A392X	Rd1-fwd	NNKCTGATCCGTGGTCCGGAAG	Rd1-rvs	CGCCTTTTTGAACGGAACATC
L393X	Rd1-fwd	NNKATCCGTGGTCCGGAAGC	Rd1-rvs	CAGCGCCTTTTTGAACGGAAC
I394X	Rd1-fwd	NNKCGTGGTCCGGAAGCGC	Rd1-rvs	GATCAGCGCCTTTTTGAACG
R395X	Rd1-fwd	NNKGGTCCGGAAGCGCTGC	Rd1-rvs	ACGGATCAGCGCCTTTTTG

Table 4.3. Primers for multiple site-directed mutagenesis combinatorial libraries. Mutated codons highlighted in red.

Mutation(s)	Primer (5′ to 3′)	
P45E	A2Rd2-combi-1	GTGACTCGTATCAGTCTGGAGAATGGTGAAGGTTGGG
V74L	A2Rd2-combi-2	GTTTTGGCCGCGAGGCGCTTATGGACCGTCAAG
V74L, M75F	A2Rd2-combi-3	GTTTTGGCCGCGAGGCGCTTTTCGACCGTCAAGTTAC
M75F	A2Rd2-combi-4	GCCGCGAGGCGGTTTTCGACCGTCAAGTTAC
A83L	A2Rd2-combi-5	GTCAAGTTACGCGTTTGCTGCCTCACTTTATCCCGG
S181G	A2Rd2-combi-6	GCTGATTCTGTCTGGCAGCCATGGCGC
S181G, H183K	A2Rd2-combi-7	GCTGATTCTGTCTGGCAGCAAGGGCGCGGAGGTG
H183K	A2Rd2-combi-8	GATTCTGTCTAGCAGCAAGGGCGCGGAGGTG
I241L	A2Rd2-combi-9	CGGTTCTGTTGCAACTTGGCGGTGAGCAG
A290M	A2Rd2-combi-10	GGATTCCGCACCGTAATATGGTCGGACTGAGCCGTATC
V74L	A2Rd3-combi-1	GTTTTGGCCGCGAGGCGCTTATGGACCGTCAAG
V74L, M75F	A2Rd3-combi-2	GTTTTGGCCGCGAGGCGCTTTTCGACCGTCAAGTTAC
M75F	A2Rd3-combi-3	GCCGCGAGGCGGTTTTCGACCGTCAAGTTAC
A89L	A2Rd3-combi-4	CACTTTATCCCGCTGCGTGGCGCGGTAGGC
H183K	A2Rd3-combi-5	GATTCTGTCTAGCAGCAAGGGCGCGGAGGTG
I241L	A2Rd3-combi-6	CGGTTCTGTTGCAACTTGGCGGTGAGCAG

Table 4.4. Screening summary for engineering campaign.

Library	Plates	Analysis	Top hit (new mutations)	Improvement
A2Rd1	16 SSM	RapidFire–MS (tier 1) Chiral LC–MS (tier 2)	A2Rd1_08_G03 (A245Q)	4x activity, 70:30 er
A2Rd2	2 combi 18 SSM	SFC–MS (Merck)	A2Rd2_01_D09 (P45E, A83L, S181G, A290M)	1x activity, 95:5 er
A2Rd3	4 combi	Chiral LC–MS	-	No significant improvement
A2Rd4	20 SSM	RapidFire–MS (tier 1) Chiral LC–MS (tier 2)	A2Rd4_16_A11 (R90Q)	1.8x activity, 95:5 er

SUPPLEMENTARY INFORMATION

The supplementary information for this chapter includes (I) DNA and protein sequences, (II) chemical synthesis of substrates and authentic product standards, (III) standard curves and LC–MS traces of biocatalytic reactions, and (IV) NMR spectra of synthesized compounds.

DNA and protein sequences. The gene encoding CYP158A2 in a pET28a(+) vector was codon-optimized for expression in *E. coli* and purchased from Twist Bioscience. The gene encoding the $P450_{RhF}$ reductase domain (RhFRed) in pET28b vectors were generous gifts from Professor David Sherman's lab at the University of Michigan. The codon-optimized genes encoding all SSN-derived proteins in a pET28b vector harboring the RhFRed domain were purchased from Twist Bioscience.

CYP158A2

```
ATGACTGAAGAGACAATATCACAAGCTGTACCGCCTGTTCGCGACTGGCCGGCTGTGGACCTGCCGGGTT
CCGACTTCGACCCGGTTTTGACCGAGCTGATGCGTGAGGGTCCAGTGACTCGTATCAGTCTGCCGAATGG
TGAAGGTTGGGCTTGGTTGGTGACTCGTCACGATGATGTTCGCCTAGTGACCAATGATCCGCGTTTTGGC
CGCGAGGCGGTTATGGACCGTCAAGTTACGCGTTTGGCGCCTCACTTTATCCCGGCGCGTGGCGCGGTAG
GCTTCCTGGACCCGCCGGATCACACCCGTTTGCGGCGTAGCGTTGCGGCGGCGTTTACCGCACGTGGCGT
GGAGCGTGTTCGTGAGCGCTCCCGCGGTATGCTGGACGAACTGGTCGACGCCATGCTACGCGCGGGTCCG
CCAGCCGATCTGACCGAAGCCGTGCTGAGCCCGTTCCCGATTGCAGTTATTTGTGAACTTATGGGTGTTC
CGGCAACCGATCGTCACAGCATGCATACCTGGACCCAGCTGATTCTGTCTAGCAGCCATGGCGCGGAGGT
GAGCGAGCGCGCGAAAAACGAAATGAACGCCTACTTCTCCGACTTGATTGGCTTAAGATCGGACTCTGCA
GGCGAGGACGTGACGAGCCTGCTGGGCGCCGCTGTGGGTCGTGACGAGATCACCCTGTCTGAGGCTGTCG
GTTTGGCGGTTCTGTTGCAAATTGGCGGTGAGGCCGTGACGAACAACAGCGGTCAGATGTTTCATCTGCT
GCTTTCCAGGCCCGAACTGGCGGAAAGACTGCGCTCGGAACCGGAGATCCGTCCGCGTGCGATCGATGAG
TTACTGCGTTGGATTCCGCACCGTAATGCTGTCGGACTGAGCCGTATCGCGTTGGAAGATGTGGAGATTA
AGGGCGTACGTATCCGGGCGGGTGATGCAGTGTACGTGAGCTATCTGGCTGCGAACCGTGACCCGGAGGT
TTTTCCGGACCCGGACCGCATCGATTTTGAAAGATCACCGAACCCGCATGTTTCTTTCGGTTTCGGCCCG
CACTATTGCCCGGGCGGCATGCTCGCCCGTCTCGAAAGCGAATTGTTGGTTGACGCGGTACTGGACCGCG
TGCCGGGTCTTAAGCTGGCAGTTGCTCCGGAAGATGTTCCGTTCAAAAAGGGTGCGCTGATCCGTGGTCC
GGAAGCGCTGCCGGTCACCTGG
```

```
MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLPNGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLAPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSSSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEAVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNAVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTW
```

RhFRed ($P450_{RhF}$ reductase domain)

GAATTCGTGCTGCACCGGCATCAACCGGTCACCATCGGAGAACCCGCCGCCCGGGCGGTGTCCCGCACCG
TCACCGTCGAGCGCCTGGACCGGATCGCCGACGACGTGCTGCGCCTCGTCCTGCGCGACGCCGGCGGAAA
GACATTACCCACGTGGACTCCCGGCGCCCATATCGACCTCGACCTCGGCGCGCTGTCGCGCCAGTACTCC
CTGTGCGGCGCGCCCGATGCGCCGAGCTACGAGATTGCCGTGCACCTGGATCCCGAGAGCCGCGGCGGTT
CGCGCTACATCCACGAACAGCTCGAGGTGGGAAGCCCGCTCCGGATGCGCGGCCCTCGGAACCATTTCGC
GCTCGACCCCGGCGCCGAGCACTACGTGTTCGTCGCCGGCGGCATCGGCATCACCCCAGTCCTGGCCATG
GCCGACCACGCCCGCGCCCGGGGGTGGAGCTACGAACTGCACTACTGCGGCCGAAACCGTTCCGGCATGG
CCTATCTCGAGCGTGTCGCCGGGCACGGTGACCGGGCCGCCCTGCACGTGTCCGAGGAAGGCACCCGGAT
CGACCTCGCCGCCCTCCTCGCCGAGCCCGCCCCCGGCGTCCAGATCTACGCGTGCGGGCCCGGGCGGCTG
CTCGCCGGACTCGAGGACGCGAGCCGGAACTGGCCCGACGGGGCGCTGCACGTCGAGCACTTCACCTCGT
CCCTCGCGGCGCTCGATCCGGACGTCGAGCACGCCTTCGACCTCGAACTGCGTGACTCGGGGCTGACCGT
GCGGGTCGAACCCACCCAGACCGTCCTCGACGCGTTGCGCGCCAACAACATCGACGTGCCCAGCGACTGC
GAGGAAGGCCTCTGCGGCTCGTGCGAGGTCGCCGTCCTCGACGGCGAGGTCGACCATCGCGACACGGTGC
TGACCAAGGCCGAGCGGGCGGCGAACCGGCAGATGATGACCTGCTGCTCGCGTGCCTGTGGCGACCGGCT
GGCCCTGCGCCTCTGA

EFVLHRHQPVTIGEPAARAVSRTVTVERLDRIADDVLRLVLRDAGGKTLPTWTPGAHIDLDLGALSRQYS
LCGAPDAPSYEIAVHLDPESRGGSRYIHEQLEVGSPLRMRGPRNHFALDPGAEHYVFVAGGIGITPVLAM
ADHARARGWSYELHYCGRNRSGMAYLERVAGHGDRAALHVSEEGTRIDLAALLAEPAPGVQIYACGPGRL
LAGLEDASRNWPDGALHVEHFTSSLAALDPDVEHAFDLELRDSGLTVRVEPTQTVLDALRANNIDVPSDC
EEGLCGSCEVAVLDGEVDHRDTVLTKAERAANRQMMTCCSRACGDRLALRL

A2-RhFRed (CYP158A2-RhFRed fusion)

ATGACTGAAGAGACAATATCACAAGCTGTACCGCCTGTTCGCGACTGGCCGGCTGTGGACCTGCCGGGTT
CCGACTTCGACCCGGTTTTGACCGAGCTGATGCGTGAGGGTCCAGTGACTCGTATCAGTCTGCCGAATGG
TGAAGGTTGGGCTTGGTTGGTGACTCGTCACGATGATGTTCGCCTAGTGACCAATGATCCGCGTTTTGGC
CGCGAGGCGGTTATGGACCGTCAAGTTACGCGTTTGGCGCCTCACTTTATCCCGGCGCGTGGCGCGGTAG
GCTTCCTGGACCCGCCGGATCACACCCGTTTGCGGCGTAGCGTTGCGGCGGCGTTTACCGCACGTGGCGT
GGAGCGTGTTCGTGAGCGCTCCCGCGGTATGCTGGACGAACTGGTCGACGCCATGCTACGCGCGGGTCCG
CCAGCCGATCTGACCGAAGCCGTGCTGAGCCCGTTCCCGATTGCAGTTATTTGTGAACTTATGGGTGTTC
CGGCAACCGATCGTCACAGCATGCATACCTGGACCCAGCTGATTCTGTCTAGCAGCCATGGCGCGGAGGT
GAGCGAGCGCGCGAAAAACGAAATGAACGCCTACTTCTCCGACTTGATTGGCTTAAGATCGGACTCTGCA
GGCGAGGACGTGACGAGCCTGCTGGGCGCCGCTGTGGGTCGTGACGAGATCACCCTGTCTGAGGCTGTCG
GTTTGGCGGTTCTGTTGCAAATTGGCGGTGAGGCCGTGACGAACAACAGCGGTCAGATGTTTCATCTGCT
GCTTTCCAGGCCCGAACTGGCGGAAAGACTGCGCTCGGAACCGGAGATCCGTCCGCGTGCGATCGATGAG
TTACTGCGTTGGATTCCGCACCGTAATGCTGTCGGACTGAGCCGTATCGCGTTGGAAGATGTGGAGATTA
AGGGCGTACGTATCCGGGCGGGTGATGCAGTGTACGTGAGCTATCTGGCTGCGAACCGTGACCCGGAGGT
TTTTCCGGACCCGGACCGCATCGATTTTGAAAGATCACCGAACCCGCATGTTTCTTTCGGTTTCGGCCCG
CACTATTGCCCGGGCGGCATGCTCGCCCGTCTCGAAAGCGAATTGTTGGTTGACGCGGTACTGGACCGCG
TGCCGGGTCTTAAGCTGGCAGTTGCTCCGGAAGATGTTCCGTTCAAAAAGGGTGCGCTGATCCGTGGTCC
GGAAGCGCTGCCGGTCACCTGGGAATTCGTGCTGCACCGGCATCAACCGGTCACCATCGGAGAACCCGCC
GCCCGGGCGGTGTCCCGCACCGTCACCGTCGAGCGCCTGGACCGGATCGCCGACGACGTGCTGCGCCTCG
TCCTGCGCGACGCCGGCGGAAAGACATTACCCACGTGGACTCCCGGCGCCCATATCGACCTCGACCTCGG
CGCGCTGTCGCGCCAGTACTCCCTGTGCGGCGCGCCCGATGCGCCGAGCTACGAGATTGCCGTGCACCTG
GATCCCGAGAGCCGCGGCGGTTCGCGCTACATCCACGAACAGCTCGAGGTGGGAAGCCCGCTCCGGATGC
GCGGCCCTCGGAACCATTTCGCGCTCGACCCCGGCGCCGAGCACTACGTGTTCGTCGCCGGCGGCATCGG
CATCACCCCAGTCCTGGCCATGGCCGACCACGCCCGCGCCCGGGGGTGGAGCTACGAACTGCACTACTGC
GGCCGAAACCGTTCCGGCATGGCCTATCTCGAGCGTGTCGCCGGGCACGGTGACCGGGCCGCCCTGCACG

TGTCCGAGGAAGGCACCCGGATCGACCTCGCCGCCCTCCTCGCCGAGCCCGCCCCCGGCGTCCAGATCTA
CGCGTGCGGGCCCGGGCGGCTGCTCGCCGGACTCGAGGACGCGAGCCGGAACTGGCCCGACGGGGCGCTG
CACGTCGAGCACTTCACCTCGTCCCTCGCGGCGCTCGATCCGGACGTCGAGCACGCCTTCGACCTCGAAC
TGCGTGACTCGGGGCTGACCGTGCGGGTCGAACCCACCCAGACCGTCCTCGACGCGTTGCGCGCCAACAA
CATCGACGTGCCCAGCGACTGCGAGGAAGGCCTCTGCGGCTCGTGCGAGGTCGCCGTCCTCGACGGCGAG
GTCGACCATCGCGACACGGTGCTGACCAAGGCCGAGCGGGCGGCGAACCGGCAGATGATGACCTGCTGCT
CGCGTGCCTGTGGCGACCGGCTGGCCCTGCGCCTCTGA

MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLPNGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLAPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSSSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEAVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNAVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTWEFVLHRHQPVTIGEPA
ARAVSRTVTVERLDRIADDVLRLVLRDAGGKTLPTWTPGAHIDLDLGALSRQYSLCGAPDAPSYEIAVHL
DPESRGGSRYIHEQLEVGSPLRMRGPRNHFALDPGAEHYVFVAGGIGITPVLAMADHARARGWSYELHYC
GRNRSGMAYLERVAGHGDRAALHVSEEGTRIDLAALLAEPAPGVQIYACGPGRLLAGLEDASRNWPDGAL
HVEHFTSSLAALDPDVEHAFDLELRDSGLTVRVEPTQTVLDALRANNIDVPSDCEEGLCGSCEVAVLDGE
VDHRDTVLTKAERAANRQMMTCCSRACGDRLALRL

CYP158A1 (SSN_A02)

ATGACCCAAGAAACGACTACACTAACAGGACAAAGCCCGCCACCGGTGCGCGATTGGCCGGCGCTGGACC
TGGACGGTCCGGAATTTGACCCGGTACTCGCAGAGCTTATGCGTGAAGGTCCGCTCACCCGTGTTCGTTT
ACCTCACGGCGAGGGTTGGGCATGGCTGGCGACCCGTTATGATGATGTTAAAGCGATCACGAACGATCCG
CGTTTTGGTCGTGCCGAAGTAACCCAGCGTCAAATCACCCGCCTGGCCCCTCACTTTAAGCCGCGTCCGG
GTTCCCTGGCGTTCGCCGACCAGCCGGACCACAACCGTTTGAGACGCGCGGTTGCGGGTGCCTTCACCGT
TGGCGCAACGAAACGTCTGCGCCCGAGAGCGCAAGAGATCCTGGACGGGCTGGTCGATGGCATTCTGGCT
GAGGGCCCGCCAGCCGATCTGGTGGAACGCGTTTTGGAACCGTTTCCGATTGCGGTCGTGTCGGAAGTGA
TGGGTGTTCCGGCTGCTGACCGCGAGCGCGTGCACAGCTGGACCCGTCAAATCATCTCTACGAGCGGTGG
CGCGGAGGCAGCCGAGCGTGCGAAACGCGGTCTCTACGGCTGGATTACTGAGACTGTGCGTGCTCGCGCG
GGCTCCGAAGGCGGTGATGTGTACAGCATGCTGGGCGCGGCTGTGGGCCGTGGTGAGGTCGGCGAGACAG
AGGCGGTCGGTCTGGCTGGCCCACTGCAGATCGGCGGTGAGGCGGTTACCCATAACGTTGGCCAGATGCT
GTACCTGTTGCTCACCCGTCGTGAACTGATGGCACGGATGCGTGAACGTCCGGGTGCGCGAGGTACGGCA
CTGGACGAACTGTTGCGTTGGATTTCCCATCGTACCAGCGTTGGTTTAGCGCGTATCGCATTGGAAGATG
TTGAGGTTCACGGCACCCGCATTGCAGCGGGTGAGCCGGTTTACGTTAGCTATCTGGCGGCCAATCGCGA
CCCGGACGTGTTCCCGGACCCGGACAGAATCGATCTGGATCGTGACCCCAACCCGCATCTGGCGTATGGT
AATGGTCATCACTTCTGCACTGGTGCGGTGCTGGCGCGCATGCAGACCGAGCTGTTAGTGGATACCCTGC
TGGAAAGACTTCCGGGTTTACGTTTGGCGGTGCCGGCTGAACAGGTTGCGTGGCGTCGTAAGACCATGAT
TCGTGGCCCGCGTACCTTGCCGTGTACCTGG

MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLPNGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLAPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSSSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEAVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNAVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTW

P450-mel (SSN_A04)

ATGGAAAATACATCAGTACAAAACAAAGAGACGGTTCGCAACTGTCCGTTCGATTATGCACATGAATTAG
AGTTCGACCCGCAACTGCGTCAGTTATTGACGGAAGAGCCGGTCAGCCGTATTCGTATGGCGTACGGCGA
GGGCGAAGCGTGGCTCGTCACCCGTTACGAAGACGTGCGCACCGTTACCACGGATCGCAGGTTCTCTCGT
TCGGCGGTGTTGGGCCGTGATTTCCCGCGCATGACCCCGGAGCCGATTGTTCAGGCAGAATCTATCAACC
TGATGGATCCGCCTGCTAGCAGCCGTCTGAGAGGCCTGGTTGCGAAGAGCTTCACCCCGCGTCGCGTTGA
ACAAATGCGAGGTGGCACCCAGCGTGTAGTTGATCGCTTGCTGGATGAAATGGAGGAGGAGGGTAGCCCG
GCAGACTTCGTGGCCCGTGTGTCAGCTCCGCTGCCGCTTATTACCATCTGCGAAGCGCTGGACATCCCGG
AAGCGGACCGTCCGTGGTTAAGAGCGCACGCAATGACGATGATGAATGTTGGTGCGGCAGGCAAACAGGA
CGCGGTTCGGGCGAAGGCTGAACTGCGTGGTTACTTCCAAGAATTGACCGCCGACCGTCGTCGTAGCCCA
GGTGAAGACCTGATCAGCACCCTCGCGACCGCGCGTGATGGTGACGAGCTGTTGGACGATGATGAGCTGG
CCGTTATGGCAATGGTTCTGCTGATCACTGGCCAAGATACTACCACCTATCAGTTGGGAAACATCGCTTA
CACCCTGTTGACCCGTCCGGATCTGTTGCGCTCGCTGAGAGCTGAGCCGCAGCGTCTGCCGCGTACTCTG
GAGGAGCTGCTGCGCCATATCCCGTTTCGTAAAGGTGTGGGTATTCCGCGCATCGCATTGGAGGACGTCG
AACTGTCCGGTGTTCTGATCAAGGCCGGTGATGTGGTTCACGTGTCTTATCTTACAGCTAATCGTGACAG
CGCCAAGTTTGACCGCCCAGACGAGCTCGACCCGGATCGCCCGACCATTCCGCACATGACGTTTGGTTGG
GGCGCGCACCACTGCCTGGGCGCGCCACTGGCGACCATGGAACTTGAGGTGGCGTTTTCCACCCTGCTTA
CGCGTTTCCCGGCGCTGCGCCTGGATGTGCCGCCTGAGGACGTGTCCTGGAACACCACGAGCATTTGGCG
TTACCCGCTGGCGCTGCCAGTGACCTGG

MENTSVQNKETVRNCPFDYAHELEFDPQLRQLLTEEPVSRIRMAYGEGEAWLVTRYEDVRTVTTDRRFSR
SAVLGRDFPRMTPEPIVQAESINLMDPPASSRLRGLVAKSFTPRRVEQMRGGTQRVVDRLLDEMEEEGSP
ADFVARVSAPLPLITICEALDIPEADRPWLRAHAMTMMNVGAAGKQDAVRAKAELRGYFQELTADRRRSP
GEDLISTLATARDGDELLDDDELAVMAMVLLITGQDTTTYQLGNIAYTLLTRPDLLRSLRAEPQRLPRTL
EELLRHIPFRKGVGIPRIALEDVELSGVLIKAGDVVHVSYLTANRDSAKFDRPDELDPDRPTIPHMTFGW
GAHHCLGAPLATMELEVAFSTLLTRFPALRLDVPPEDVSWNTTSIWRYPLALPVTW

A0A0B5DS97 (SSN_A05)

ATGACAGAAGAGTCAACGGCTCTAACTGGACGCGAACCGCCACCGGTGAGAGACTGGCTGGCGCATGAAT
TGGACGGCACGGTGTTCGATCCGGTCCTGAGCGAGCTGATGAGCGAGGGTCCGCTGACACGCATCCGCTT
ACCGCACGGTGAAGGCTGGGCGTGGCTGGCTACCCGTTACGAGGATGTACGCACCATTACTAACGATCCG
CGTTTTTCTCGTGCCGAGGTGGTGGGCCGTCAAGTGACGCGTTTGGCTCCGCATTTTAAACCGCGTCCGG
GTAGCCTGGCTTTCGCGGATCAGCCGGATCACAACCGTCTCAGAAGAGCGGTCGCAGGTGCTTTCACCGT
TTCTGCCATGAAACGTCTGCGTCCGCGTGCGCAGCGTATTCTGGACCCGCTGGTAGAGGAGCTGGTTCGC
CAAGGTCCGCCAGCAGATCTGATTGAACGTGTTTTGGAGCCGTTTCCGCTACGCGTGGTTTCCGAGGTTA
TGGGTGTTCCGGCCGACGATCGTGAACAGGTTCACGGTTGGACCCGCCAGATCATCTCCACCAGCGGTGG
TGCCCAAGCAGCGGAACGAGCGAAGGCTGGCATGTACGGCTGGATCACGGAGGCCATCCGTGCGCGCCGT
GGCAGCGCAGGCGAGGATGTGCTCTCGCTCTTAGGTGCGGCGGTGGGCCGTGGTGAAATTAGCGAAGAGG
AGGCGGTGGGCCTGGCGGGCCCACTGCAAATCGGCGGAGAAGCGGTCACCCATAACTGCGGTCAGATGCT
GTTCCTGCTCTTGACCCGTCCGGGTCTGTTCCGCCTGCTGCAGAGTAGCCGCGGTACTCCGGAATCCTGG
GATACCGCGCTGGACGAACTGCTGCGTTGGATCCCGCACCGTACCGCGGTTGGTTTGGCCCGTGTCGCGC
TCCAAGACGTTGAGCTGCACGGCCTGTGCATTCGTGCGGGTGATGCAGTTCATGTTAGCTATCTGGCCGC
GAATCGCGACCCGGAGGTGTTCCCGGGCCCAGACCTGATCGACGTCGGCCGCGACCCCAACCCGCACCTG
TCTTTTGGTAATGGTCCGCATCATTGTACCGGCGCTGTTTTGGCACGTTTGCAACTGGAACTGCTGCTCC
ACACCCTGTTGGACCGTCTGCCAGGCCTTCGGTTGGACGTGCCGGCTGAACGTGTTCCGTGGCGTCGCCA
GACCATGATTCGTGGTCCGCAGGCACTTCCGGTTACGTGGGGT

MTEESTALTGREPPPVRDWLAHELDGTVFDPVLSELMSEGPLTRIRLPHGEGWAWLATRYEDVRTITNDP
RFSRAEVVGRQVTRLAPHFKPRPGSLAFADQPDHNRLRRAVAGAFTVSAMKRLRPRAQRILDPLVEELVR

QGPPADLIERVLEPFPLRVVSEVMGVPADDREQVHGWTRQIISTSGGAQAAERAKAGMYGWITEAIRARR
GSAGEDVLSLLGAAVGRGEISEEEAVGLAGPLQIGGEAVTHNCGQMLFLLLTRPGLFRLLQSSRGTPESW
DTALDELLRWIPHRTAVGLARVALQDVELHGLCIRAGDAVHVSYLAANRDPEVFPGPDLIDVGRDPNPHL
SFGNGPHHCTGAVLARLQLELLLHTLLDRLPGLRLDVPAERVPWRRQTMIRGPQALPVTWG

A0A0C2AVS8 (SSN_A06)

ATGAGTCTAGAACCCTCTGTAAGCTCAGCTGCGGACGTGTGGCATTGCCCGTTCGATCATGCAAATGGTC
TGGAGTTCGACCCGCTGTTGAAACGTCTTCTGACTGAGGCGCCGATCGCGCGTATTCGTATGCCATACGG
CGATGGCGAAGCGTGGCTGGTCACCCGTTATGATGACGTTAGATTCGTGACCAGCGATCGCCGCTTCAGC
CGTGCTGCGGGTATTGGTCGTGATTTACCGCGTATGACCCCGGCTCCGATTGCTCAGGCAGAAGCGATTA
ACCTGATGGATCCTCCGGAGCACAACCGTCTGAGACGCCTGGTGGCTGCCGGTTTTACCAACCGCCAAGT
CGAGCGTTTGCGCCCTTCGACCCAACGTACTGTTAATGCACTGTTAGATGACATGGCGGAGAGCGGTGCC
CCGGCGGATCTCGTCCAGCACCTGAGTGCACGACTGCCGCTGACCACCATCTGCGAATTGCTGGACATCC
CGGCTGACGACCGTGCGGACCTGCGCCGCTGGGCAATGGCAATGATGGCTGTGGGCGTGGCCGACCGCGC
CGCGGCGGATGAGGCGAAGAGCGAATTGCGCAGCTATTTTGCGGACTTGACGGCCCGTCGCCGTCGCGAT
CCGGGTGAGGACCTGTTGTCCAGCCTGGCTACCGCTCGTGACGGCGCAGAGCTGCTTGATGAGAAAGAGT
TGGCAGTGATGGGTATGGTCCTGCTCGTAACCGGTCATGACACCACTACGTACCAGCTGTCTAACATTAC
CCACACCCTGTTAACGAACCCGGAGCAACTGGCGAAGCTGAGGGCGGAGCCGGAGTTGCTCCCGACCGCG
ATCGAAGAATTGTTGCGCTTCATCCCGTTTCGTCAGGGTGTGGGAATCGCGCGTGTTGCGACCGAAGATG
TTACGATCGACGGCGTTACCGTGCGTGCAGGTGAAGCGGTTCATGTTAGCTACCTGGCGGCGAACCGTGA
CCCGGCGAAATATGAACGTCCGGATGAATTGGACCTGACGCGTGGTACGCCACATCACATGGCGTTTGGT
TCCGGCACACACCACTGCCTGGGCTCCCATCTGGCTCGCATGGAACTGTGTGTTGCCATCGGCACTCTGC
TTGAGCGTTTCCCGCGTCTCCGTCTGGCCGTGCCGGAACACGAAATTGATTGGAATACCGGCTCTATTTG
GCGTTACCCGCAGACCCTGCCGGTTACCTGGGAATTC

MSLEPSVSSAADVWHCPFDHANGLEFDPLLKRLLTEAPIARIRMPYGDGEAWLVTRYDDVRFVTSDRRFS
RAAGIGRDLPRMTPAPIAQAEAINLMDPPEHNRLRRLVAAGFTNRQVERLRPSTQRTVNALLDDMAESGA
PADLVQHLSARLPLTTICELLDIPADDRADLRRWAMAMMAVGVADRAAADEAKSELRSYFADLTARRRRD
PGEDLLSSLATARDGAELLDEKELAVMGMVLLVTGHDTTTYQLSNITHTLLTNPEQLAKLRAEPELLPTA
IEELLRFIPFRQGVGIARVATEDVTIDGVTVRAGEAVHVSYLAANRDPAKYERPDELDLTRGTPHHMAFG
SGTHHCLGSHLARMELCVAIGTLLERFPRLRLAVPEHEIDWNTGSIWRYPQTLPVTW

A0A0N9I1K4 (SSN_B01)

ATGACCACAGAAAGTACTTCAACGACACACCCAGTCCGCTTCTGGGATGTCCCAGATTTGCGTGCGACCG
ATTTTGACCCGCTCCTGGCGGAGTTGCTTAGAGACGAACCGGTGTCCCGCGTGCGTCTTCCGAACGGTGA
TGGTTACGCCTGGCTGGTTACGCGTTATGCTGACGTTCGTTTCGTGACTTCAGATCCGCGTTTTAGCCGT
CAGGCGGTGCTGGGCCGTAGCGTCACCAGACTGGCGCCTCACTTCATCCCGCTGGACGACGCGGTTGGCT
TCGCCGACCCGCCTGAACATACCCGTCTGAGACGTGCGGTGGCGAAAGCATTCAGCGCGAAGTCCGTGCA
CGCCCTCCGCGCACGTACGCAGCAACGCATGCACGGCCTGGTTGACGCAATGCAGCGTCATGGTAGCCCG
GCAGACCTGATGGAGCACGTGCATAATCCGTTGCCGTTGGCCGTCGTGAGCGACCTGATGGGTGTTCCGG
AGCAGGATCGTCCGGCAATGGCAGAGTGGTCCGGTGTTCTGCTGAGCGCTGGCCACGGCCGTGAGCGCTC
TGAACGCGCTAAGGCGGAGGTCGGCGCGTATTTTACCACCCGTCTGAGCCAACCGGGTGCTGGTAGCGAA
GATGAGTTATTAGGCCTGCTGTCGGCGGCCATCGAGCGTGAAGAGTTGACCCTGAAAGAAGCGGTTGGCC
TGGCGGTGCTGATTCAGGTTAGCGGTGCCCATGCGGTGCGTAACAACAGCGGTAATATGATGTATGCGCT
GCTCACCCATCCGGGCCACTTGGCTCGTCTGCGTACCGAACCGGGCCTGTTACCGCAAGCAATCGAAGAA
CTTCTGCGCTGGATTCCGCACCGCTCTGGCGTGGGCCTGTCTCGTATTGCAACCGAAGATGTAGCTGTAG
GCGATGTGGTGATCCGTGCGGGTGAAGCTATCTACGCGAGCTACCTGACTGCCAATCGTGATCCAGATGT
CTTTGACGACCCGGGTCACCTGGACTTCGACCGCGCGTTCAACCCGCACGTTTCTTTTGGTCATGGTCCG
CATCACTGCGTTGCATCCGCCCTGGCGCGCATGGAGTCGCGCATTATGGTTGGGACCCTGTTAGACCGCT

TGCCGGAGTTGCGACTGGCTGTGCCGGTTGAAGAGGTTCGTTGGCAACGTGCGGCACTCATCCGCGGTCC
AGAGACGCTGCCGGTTGCGTGG

MTTESTSTTHPVRFWDVPDLRATDFDPLLAELLRDEPVSRVRLPNGDGYAWLVTRYADVRFVTSDPRFSR
QAVLGRSVTRLAPHFIPLDDAVGFADPPEHTRLRRAVAKAFSAKSVHALRARTQQRMHGLVDAMQRHGSP
ADLMEHVHNPLPLAVVSDLMGVPEQDRPAMAEWSGVLLSAGHGRERSERAKAEVGAYFTTRLSQPGAGSE
DELLGLLSAAIEREELTLKEAVGLAVLIQVSGAHAVRNNSGNMMYALLTHPGHLARLRTEPGLLPQAIEE
LLRWIPHRSGVGLSRIATEDVAVGDVVIRAGEAIYASYLTANRDPDVFDDPGHLDFDRAFNPHVSFGHGP
HHCVASALARMESRIMVGTLLDRLPELRLAVPVEEVRWQRAALIRGPETLPVAW

A0A1C5CYG0 (SSN_B02)

ATGACGGAAATAGGAGAGACAACTACATGTCCTTTCGATTTTTCGGAAGGTCTCGAGTTCGACCCGCTGC
TGGCTGAGCTGATGAGCCGTAACCCGGTTACTCGCATTCGCCTTCCGTACGGCGACGCCGATGCGTGGCT
GGTCACGGGTTTTGAAGGCGTACGCCAGGTGACGACCGACCACCGTTTTAGCCGTGCAGCCATCACGGGC
CGTAACTATCCGCGTCTCACCCCAGAACCGATCGTGTCTCCGGAAAGCATTAACGTGATTGACCCGCCTC
GCAGCACCCGTTTGCGTAAGCTGGCGTCGCAAGCCTTTACCAAACGTCATGTTGATCGCATGCGTCCAGC
TGTGGAGGGCATCGCTGCGGGTCTGCTAGACGAGATGGCAGAGCACGGCAGCCCGGCCGATCTGGTAAGC
CATCTGAGCGATCGTCTGCCCCAGCACACCATTTGTGAATTGCTGGGTATAGCGCCGGAGGATAGACCGA
GTATGCAAGAGTACGCACATCAGCTGCTGTCCACCGGCCCGGAAACCCAGCAGGCGGCGGCGAACGCAAA
GAGCGCCCTGCGTGAATACTTCGGCAAATTGGTGACCGAACGTCGTCGTAGCCCGGGTGACGACTTAATT
TCTACTTTGGCGGCTGCTAAAGAAGGCGACGATGCGTTTGACGACCAAGATTTGGCGGTTATGGCGCTAA
CCCTGATGTTGTCCGGTCACGATACCGCGACGTGCGAGATCTCCAACATTGCGTATATCCTGCTGAATCG
TCCGGACCTGATGGAGCAGGTGCGTTCTCGCCCAGAGGAACTTACACAAGTGCTGGATGAGTTACTTCGT
TTCATCCCGTTTCGCAAGGGGGTCGGTATTCCGCGCATCGCTACCGAGGACGTCGAGATCGAAGGCGTTA
CGATCCGCGCAGGTGACTTCGTGCACGTGTCCTATCTGACTGCGAACAGAGACCCGAAACGTTTCACCGA
CCCAGATATTATCGATCCGAATCGTACCCCGACCCCGCACATGACCTTTGGTTGGGGTGGTCACCGCTGC
ATTGCCGCTCCGTTGGCTATGGTTGAGCTGGAAGCGGCGATCGGTGGCTTGGTTAATCGGTTCCCGCGTT
TGCGATTGGCGGTTCCGGTTGATGAGGTTCGTTGGGATACGGAAACCATCCGTCGCTTCCCGATTACCCT
GCCGGTGGCATGG

MTEIGETTTCPFDFSEGLEFDPLLAELMSRNPVTRIRLPYGDADAWLVTGFEGVRQVTTDHRFSRAAITG
RNYPRLTPEPIVSPESINVIDPPRSTRLRKLASQAFTKRHVDRMRPAVEGIAAGLLDEMAEHGSPADLVS
HLSDRLPQHTICELLGIAPEDRPSMQEYAHQLLSTGPETQQAAANAKSALREYFGKLVTERRRSPGDDLI
STLAAAKEGDDAFDDQDLAVMALTLMLSGHDTATCEISNIAYILLNRPDLMEQVRSRPEELTQVLDELLR
FIPFRKGVGIPRIATEDVEIEGVTIRAGDFVHVSYLTANRDPKRFTDPDIIDPNRTPTPHMTFGWGGHRC
IAAPLAMVELEAAIGGLVNRFPRLRLAVPVDEVRWDTETIRRFPITLPVAW

A0A1G8ZS81 (SSN_B03)

ATGACAACTGTAGCAGGACTACCCGCTTCATCTGGTGCGGAAGCGGCGGCTACCGGAGACGGCCTGCCGC
CTGTGCGTGACTGGCCGGCACCGGAGCTGAAGGGCACCGAGTTCGACCCGGTCCTGGCGGAGTTAATGCA
AGAAGGTCCAGTCACTCGCATCCGCTTGCCGCACGGCGAGGGCTGGGCTTGGCTGTTGACGCGTCATGAT
GATGTGAAGGCATTCACCAATGATCCACGTTTTAGCCGTGCTGCGGTGATGGAACGCGACGTCACTCGCC
TTGCTCCACACTTTATCCCGAGCAAAGAGGCGGTGGGTGTTCAGGATCCGCCGCATCACACCACCCTGAG
GCGTGCGGTCGCGCCGGCCTTCACGAGCCGTGGCGTTGAACGTTTACGTGAACGAGCGCAAGCAATGCTG
GATGAGCTGGTTGATCGCTTGGTTGCGCAAGGTCCGCCGGCGGACCTGACCGAGCAGGTTCTGGCTCCGT
TTCCACTGGCGGTTGTTTGTGAAGTGATGGGTGTGCCGGAGGCTGATCGTCCGCGCATGCATGCATGGAC
CGTTCGTATCCTGTCCAGCGCGGAGGGCGCGGAGCGCAGCGAGGAAGCCAAACGTGCTATGTGCGGTTAT
TTCGCCGAGCGTTTGCGTACCCAGCCGGAGGGCGAGGACGTGCTGGGTTTGCTTGCCGCGGCCGTGGCGC
GTGGTGAAGTCTCCGCTACCGAAGCGGTTGGCTTGGCGCTGCTGATTCAGATCGGTGGCGAAGCGGTTAC

GAACAACACCGGCAACATGCTGTATCTGTTGCTCACCCGTCCGGAATTGGCGGACCGCCTACGTACAGAT
AGAGACTTACGCCCGGCAGCGATTGACGAGCTCCTGAGGTACATTCCGCACCGTAACAGCGTAGGTTTGT
CTCGTATTGCAATGGAAGACATGGAGATCCGTGGCGTGCGTATCGCCGCGGGAGATGCAGTGTATGTTTC
CTACCTGGCGGCAAATCGTGACCCCGAAGTTTTCCCGGATCCGGACACGATTGATTTTACCCGCTCGCCG
AACCCGCACGTGGCATTTGGTCATGGTCCTCACTACTGCGTTGGCGGTATGTTGGCTCGTCTGGAAAGCG
AACTGCTCGTGGGCACCTTACTGGACCGCCTGCCGGATCTGAGACTTGCCGTTCCGGAGGACGAACTGCG
CTGGCGCCCGGGTGCACTGATCCGTGGTCCGGAGGGTCTGCCGGTGACCTGG

MTTVAGLPASSGAEAAATGDGLPPVRDWPAPELKGTEFDPVLAELMQEGPVTRIRLPHGEGWAWLLTRHD
DVKAFTNDPRFSRAAVMERDVTRLAPHFIPSKEAVGVQDPPHHTTLRRAVAPAFTSRGVERLRERAQAML
DELVDRLVAQGPPADLTEQVLAPFPLAVVCEVMGVPEADRPRMHAWTVRILSSAEGAERSEEAKRAMCGY
FAERLRTQPEGEDVLGLLAAAVARGEVSATEAVGLALLIQIGGEAVTNNTGNMLYLLLTRPELADRLRTD
RDLRPAAIDELLRYIPHRNSVGLSRIAMEDMEIRGVRIAAGDAVYVSYLAANRDPEVFPDPDTIDFTRSP
NPHVAFGHGPHYCVGGMLARLESELLVGTLLDRLPDLRLAVPEDELRWRPGALIRGPEGLPVTW

A0A1G9MX20 (SSN_B04)

ATGAGGGATGAAGCTAGTACAACTACGACAACCACCACGCCGGCGACCACCACGCCGACAACTGCGAAAG
CGTCCACCACCACGGCGGCCACGGGTGCCACCACTGCGGCGGCGGCGAGCGCACCGCTGCCTCCGGTTCG
CTTCTGGCCGGTTCGTGATCTAACCGGTGTCGAATTTGATCCGGACCTGGCGGCGCTGATGCGTGAAGGC
CCGGTTACTCGCATCCGTCTCCCCAACGGTGAGGGCTGGGCATGGCTGGTTTCTCGTCATGAAGACGTGC
GTACCGTTACCAACGATCCGCGTTTCAGCCGTCGCGCTGTTGTTGGTAGCGATGTGACCCGCTTGGCTCC
GCATTTTATCCCGGTGGACGGCGCTGTCGGTTTTGAAGACCCACCGGACCACACCCGTTTGCGGCGTGCG
GTTGCACCGGCCTTCACCACGCGTGGTGTAGAGCGTCTGCGCTGGCGTGCACAGGCAATGCTGGACGAAT
TGGTCGACGGCGTCCTGAGAGACGGACCGCCAGCGGATCTGGTGGCTCGTGTTCTGGCACCGTTCCCGCT
GGCCGTGGTTTGTGAACTGATGGGTGTACCGGCGGCGGACCGCCCGGCAATGCATGAATGGACCAGCCTG
ATTCTGAGCAGCGCTGAGGGTGCCGAACGCAGCGAGCGCGCACGTAACGCAATGTGCGATTACTTCGCGG
GTCAGCTGGCGGAGAGGAGAGGCACCGGTGGTGAGGATGTCGTTGCTCTGCTCTCGGCGGCCGTCGCGGG
TGGTGAGATCACCGAAGCGGAGGCCGTGGGCTTGGCGTTGCTTATCCAAATTGGTGGCGAGGCAGTGACG
AACAACACCGGTAATATGGCATACATTCTGCTGACCCGTCCGGACCTGATGGATCGTTTATGCCGTGATC
CTGGTGTTCGTGCGACGGCTGTGGACGAGCTGCTGCGTTGGATTCCGCACCGCAACGGCGTGGGCCTGTG
TCGTATTGCAACCGAGGACGTGACCGTGGGTGGCCGTGAAATCCGCCGTGGCGAGGCGGTGTATGTGTCA
TATCTGGCTGCTAATCGTGATCCTGACGTTTTTGCTGATCCGGACGGCATCGTGCTGGATCGTTCCCCGA
ATCCGCACGTGGCGTTCGGTTATGGTCCGCACTACTGCGTTGGCGCGCTCTTGGCGAGACTTGAGACAGA
ATTGTTATTTTCTGCCATGGTTGACCGATTCCCGGGCATGCGTCTGGCGGTCCCGGCTGAGGAATTGCGT
TGGCGCCCAGGCGCTCTGATCCGCGGTCCGGAAGGCCTGCCGGTGGTGTGG

MRDEASTTTTTTTPATTTPTTAKASTTTAATGATTAAAASAPLPPVRFWPVRDLTGVEFDPDLAALMREG
PVTRIRLPNGEGWAWLVSRHEDVRTVTNDPRFSRRAVVGSDVTRLAPHFIPVDGAVGFEDPPDHTRLRRA
VAPAFTTRGVERLRWRAQAMLDELVDGVLRDGPPADLVARVLAPFPLAVVCELMGVPAADRPAMHEWTSL
ILSSAEGAERSERARNAMCDYFAGQLAERRGTGGEDVVALLSAAVAGGEITEAEAVGLALLIQIGGEAVT
NNTGNMAYILLTRPDLMDRLCRDPGVRATAVDELLRWIPHRNGVGLCRIATEDVTVGGREIRRGEAVYVS
YLAANRDPDVFADPDGIVLDRSPNPHVAFGYGPHYCVGALLARLETELLFSAMVDRFPGMRLAVPAEELR
WRPGALIRGPEGLPVVW

A0A1J4PSC8 (SSN_B05)

ATGACAACACTAGAAACTTCAGCTCCCACGGGCCAAGAACCGCCACCGGTGCGTGACTGGCCTGCTCTGG
ACCTGGATGGCACCGAATTTGATCCGGTGCTGGCGGGTCTGGTTCGTGAAGGCCCGCTGACCCGCATCCG
CTTACCGTTCGGCGAGGGGTGGGCCTGGCTGGCCACCCGTTATGAGCATGTGAAGCTGATTACGAACGAT
CCGCGTTTTAGCCGTGCGGCAGTGGCGCGCGGTCAAGTCACGCGCTTGGCGCCAAACTTTGCACCGCGTC

CGGGCAGCCTGGCGTGGGCAGATCAGCCGGACCACAACCGTCTGCGCAAGCCGGTTGCGGGCGCGTTCAC
CGTTGGTGCTATGAAACGTCTGCGTCCGCGTGCCCAGGAGATCCTGGATGAGTTGGTTGACGCTCTGGTG
CACGAAGGTCCACCGGCTGATCTGGTAGAGCGTGTTTTGGAGCCGTTTCCGTTGTACGTTGTCAGCGAAG
TTATGGGTGTGCCGGACACTGACCGCGACCAGGTTCACTCCTGGACCCGTCAAATCATCAGCCTTTGCGG
TGGTGCGGAAGCTGCCGACCGTGCGAAAAACGGCCTGTACGGTTGGATCACCGATATTGTTCGTGCGCGT
GCACGGTCGGCCGGCGAGGACGTGTACTCCATGCTGGCGAGAGCGGTGGCTCGCGGCGAGATTGGTGAGG
AAGAGGCGGTGGGTTTGGCGGGTCCGCTGCAGATCGGCGGTGAGGCGGTGACGCACAACTGCGGTCAGAT
GCTGTACTTAATGTTGACGCGTCCGGAGTTGATGGCACGCATGCGTTCTCGTCCGGACGAAAGAGGCCCA
GTTTTCGATGAACTATTGCGCTGGATCCCGCACCGTAATACCGTTGGTCTCGCGCGTATCGCGCTGGAAG
ACGTCGAGATCGCGGGTCATCGTATTGCTGCGGGCGAGGCCGTGTACGTGAGCTATCTGTCTGCAAATCG
TGATCCGGAAGTGTTCCCGGACCCGGACGGCATTGCTCCGGATCGCGATCCTAATCCGCACGTGGCCTTC
GGAAACGGTCATCATTATTGTACCGGTGCTGTACTGGCACGTCTCCAACTGGAGTTGTTGGTTGACACCT
TGCTGGACCGTCTGCCGGGCCTTCGCCTGGCGGTCCCAGCAGAAGATGTTGAATGGAGACGTCAGACCAT
GATTAGAGGTCCGCGTACCCTGCCGGTTACTTGG

MTTLETSAPTGQEPPPVRDWPALDLDGTEFDPVLAGLVREGPLTRIRLPFGEGWAWLATRYEHVKLITND
PRFSRAAVARGQVTRLAPNFAPRPGSLAWADQPDHNRLRKPVAGAFTVGAMKRLRPRAQEILDELVDALV
HEGPPADLVERVLEPFPLYVVSEVMGVPDTDRDQVHSWTRQIISLCGGAEAADRAKNGLYGWITDIVRAR
ARSAGEDVYSMLARAVARGEIGEEEAVGLAGPLQIGGEAVTHNCGQMLYLMLTRPELMARMRSRPDERGP
VFDELLRWIPHRNTVGLARIALEDVEIAGHRIAAGEAVYVSYLSANRDPEVFPDPDGIAPDRDPNPHVAF
GNGHHYCTGAVLARLQLELLVDTLLDRLPGLRLAVPAEDVEWRRQTMIRGPRTLPVTW

A0A221NTD0 (SSN_B06)

ATGACGGAAGAGACATTAACACAAACTCTACCGCCCATCCGCCATTGGCCTGCGCTGGACCTGACGGGCA
CCGATTTCGATCCCGTTCTGACGGAACTGATGCACGAAGGTCCGGTTACTCGCATCAAGCTGCCGAACGG
CGAGGGCTGGGCTTGGCTGGTGACCCGTTACGATGATGTTCGTATGGTTACGAACGATCCGCGTTTTTCG
AGAGAAGCCGTTATGGATCGTCCGGTGACCCGTTTGGCTCCGCATTTTATTCCGGACCGCGGTGCCGTGG
GTTTCCTGGACCCGCCGGACCACACCCGCCTGCGTCGTGCGGTAGCGGCGGCGTTTACCGCTCGTGGTGT
TGAGCGTGTTCGAGACCGTGCCCGCGATATGCTGGACGAGCTTGTTGACGAGTTGCTGCAAGATGGACCG
CCAGCGGATCTGACCTCTGCAGTGCTGTCGCCGTTCCCGGTAGCAGTGATTTGTGAATTAATGGGTGTCC
CGGCTGGCGACCGCCACAGCATGCATAACTGGACCCAACTCATTCTGTCCAGCAGTAATGGTAAAGAAGT
TAGCGAAAAAGCAAAACGTGAAATGGCAGCGTATTTCAGCGATCTGATCGGCCTGCGCGAGAACTCTACA
GCCGAGGATGTGACCTCCCTGCTCGGTGCGGCGGTGGGCCGTGGGGAGATCACCCTGGAAGAGGCGGTGG
GCCTGGCGGGTCTGTTACAGATTGGCGGTGAAGCTGTGACCAACAACAGCGGTCAGCTGTTTTACCTGCT
GCTGACCCGTCCGGACCTGGCGGAGCGCTTGCGCGCCGAGCCGGAAATTCGTCCGCGTGCGATTGATGAG
TTGTTGCGTTATATCCCGCACCGCAACGCAGTCGGCCTGTCCCGCATCGCCTTGGAGGACGTCGACATCC
GTGGTTTCCGTATCCGTGCAGGCGACCCAGTGTATGTGAGCTACTTGGCGGCTAATCGTGATCCGGACGT
GTTTCCGGACCCGGATGCAGTTGACTTCAGCAGAAGCCCGAATCCGCATGTTTCTTTTGGCTTCGGCCCT
CACTTCTGCCCGGGTAATATGCTGGCTCGTCTTGAGTCCGAACTCTTGGTTGAGGCGTTATTGGATCGCA
TCCCGGCGCTCCGGCTGGCGGTCCCGCCAGACCAGGTGCCGTTTAAGAAGGGTGCACTGATTCGTGGTCC
GGAAGCGTTACCGGTTACTTGG

MTEETLTQTLPPIRHWPALDLTGTDFDPVLTELMHEGPVTRIKLPNGEGWAWLVTRYDDVRMVTNDPRFS
REAVMDRPVTRLAPHFIPDRGAVGFLDPPDHTRLRRAVAAAFTARGVERVRDRARDMLDELVDELLQDGP
PADLTSAVLSPFPVAVICELMGVPAGDRHSMHNWTQLILSSSNGKEVSEKAKREMAAYFSDLIGLRENST
AEDVTSLLGAAVGRGEITLEEAVGLAGLLQIGGEAVTNNSGQLFYLLLTRPDLAERLRAEPEIRPRAIDE
LLRYIPHRNAVGLSRIALEDVDIRGFRIRAGDPVYVSYLAANRDPDVFPDPDAVDFSRSPNPHVSFGFGP
HFCPGNMLARLESELLVEALLDRIPALRLAVPPDQVPFKKGALIRGPEALPVTW

A0A2G7EEB4 (SSN_C01)

ATGTCAGAATGTCCACTATATGAGGTACCCGACAGCCCGCCGGAACTGACGTTTGATCCGTTTCTCCGCA
ACGCATTGGACGCGCCCGTGAGCCGTATCCGACTTCCGTATGGTGAAGGTGATTGTTGGCTGGCTACCCG
TTACGACGACGTTCGCTTCGTGACCAGCGATCCGCGCTTCACCCGTGATATTACCGGTCGTCCGACCCCG
CGGATGACTCGTCATCTGATCCCGCTGGACCGTGCGGTGTCTTTTGTTGACCCGCCGGACCACGCGCGTG
TTCGCTCCGTGGTGGCGCCAGCTTTCACGCGTAGAGCGGTCGAACGTCTGCGTCCGAGAGCTCAAACCAT
TCTGGATGGTTTGGTTGACGATATGCTGGCTGCTGGCCCTCGGGCGGACGTGGTACGTAGCGTGGTTTCC
CCGTTCCCGCTGGCGGTCGTGGGCGAGGTCCTTGGCACCCCGCCGGAGGATCGTCCACAGCTGCGCAACT
GGGCGGAGGTTGTGTTGACACGTGCTACGGATCAGGCAGGCCTGGAACGTGCGCAACAAGCGAAGGCGGC
GGCTAGCGGCTACTTCCGTGCGCTGGCGGCCCAACGTCGTAACCGTCCGGCAGACGACCTGATGACCAGA
ATGGTTGAGGCGGTCGAGGCGGGCCGCATCGATGAAGAGGAACTGTTGGCGCTGGCCACCCTGATCGGTT
TGAATGGCTGGCATGCCGTGCGTAACAACGTGTCAAATATGGTGTACCTGTTGCTGACTCGCGCGGAGTT
GCGTGAACGCTTACTGTCTGAACCGGAGATGGTTCCGCAGGCGGTCGAGGAGCTGCTGCGTTGGATTCCG
CACAAAAACGGTGTTGGTCAGCCGCGTATCGCCACCGAAGATGTCGAGGTTTCCGGCGTTCTGATCCGTC
GTGGTGAGGTGGTGCAGGTTAGCTATGTTGCAGCCTGCCGTGATCCAGGTCATTTTCCGGACCCGGACGC
TATTGATCTGGACCGCCAGGGTCCACCGCACCTGGCGTTCGGTAATGGCCCGCACCACTGCGTTGCACCG
CTGCTGGCTCGTATGGAAGCGGAGGTGCTGTTGTCGACCTTACTGCGTAGATTGCCGGGTTTGCGCCTTG
CCGTACCTGCCCAAGAAATTGCATGGCAGCGTAGCGTTTTGATCCGCGGTCCGGTTGGCCTGCCGGTTAC
CTGG

MSECPLYEVPDSPPELTFDPFLRNALDAPVSRIRLPYGEGDCWLATRYDDVRFVTSDPRFTRDITGRPTP
RMTRHLIPLDRAVSFVDPPDHARVRSVVAPAFTRRAVERLRPRAQTILDGLVDDMLAAGPRADVVRSVVS
PFPLAVVGEVLGTPPEDRPQLRNWAEVVLTRATDQAGLERAQQAKAAASGYFRALAAQRRNRPADDLMTR
MVEAVEAGRIDEEELLALATLIGLNGWHAVRNNVSNMVYLLLTRAELRERLLSEPEMVPQAVEELLRWIP
HKNGVGQPRIATEDVEVSGVLIRRGEVVQVSYVAACRDPGHFPDPDAIDLDRQGPPHLAFGNGPHHCVAP
LLARMEAEVLLSTLLRRLPGLRLAVPAQEIAWQRSVLIRGPVGLPVTW

A0A2M9JRU5 (SSN_C02)

ATGACACTAGGATCAGATGCTAGGCAATGTCCGTTTGACTTTGCTGAAGCGCTGGAATTTGACCCGGTGC
TCCGTCACCTGATGAATGAGGGCCCAGTGTCTCGTATTCGTTTACCGTTCGGCGAAGGTACGGCCTGGCT
GGTGACCGGTTATGATGACGTGAGAACGGTGACCACCGATCGTCGTTTCAGCAGACACGCGGTCGCAGGC
CGTGACTTCCCGCGCATGACCCCTGAACCGATTGTTCAGAGCGAGGCGATTAACCTGATGGATCCGCCAG
CGGGCTCTCGTTTGCGCTCGCTGGTTAGCAAAGGTTTTGCACCGGCACACATTGAACGGATGCGTTTGCG
CACCCAAAACGTTGTTGATGAGCTGCTGGATCAGATGGCTGGCCAAGGTACACCGGCAGACATCATGGAA
CACCTGGCGGGTCCGCTGCCGCTGATTACCATTTGCGAAGTGTTGGAGATCCCCGAAGCGGATCATGCGC
AGCTGCGTGCGCACGCGCGTACCATGATGGATGTTGGTGTCGGAAACCGTGATGCAGCGGTTCGAGCGAA
AGCAGAATTGCGCGCGTACTTCGCGACCCTGACCGGCCAACGTCGTAAGAACCCGGGCGACGACCTGATC
AGCGTTTTGGCTACCGCTCGTGATGGTGACGAGATTCTGAATGATGAGGAACTCGCGGTGATGGCAATGG
TGCTGCTGATTACCGGTCAAGATACCACTACGTACGAAATCGGCAACTTAACCTACCTGTTGCTGACGCG
TCCGGACCTTTTAGCGACTGTCCGTGAACGTCCAGAGCTTCTTCCGGGTGCCATCAACGAGCTGTTACGC
TTCATCCCGTTCCGCAAGGGCGTGGGTATCCCGCGCGTGGCGGTTGAAGATGTGGTACTGAGCGGTGTCG
CCATCCAGGCTGGCGACATCGTGCATGTTTCCTATTTGACGGCTAATCGTGACCCGCGCAAGTTCGACCG
TCCGGACGAGGTTGACTTGGAGCGCTCCGGCCCACCGCACATGACCTTTGGTTGGGGTGCGCACCATTGT
CTGGGCGCTCCGCTGGCTGCTGCGGAGATCGAGATCGCACTGGGTAGCCTGTTTGAACGCTTCCCGCGTC
TGGCGCTGGCCAAACCGGCTGAGGAGCTGCAGTGGAATGCAACCAGCATTTGGCGTTATCCGTTGGCGCT
GCCGGTTACCTGG

MTLGSDARQCPFDFAEALEFDPVLRHLMNEGPVSRIRLPFGEGTAWLVTGYDDVRTVTTDRRFSRHAVAG
RDFPRMTPEPIVQSEAINLMDPPAGSRLRSLVSKGFAPAHIERMRLRTQNVVDELLDQMAGQGTPADIME

HLAGPLPLITICEVLEIPEADHAQLRAHARTMMDVGVGNRDAAVRAKAELRAYFATLTGQRRKNPGDDLI
SVLATARDGDEILNDEELAVMAMVLLITGQDTTTYEIGNLTYLLLTRPDLLATVRERPELLPGAINELLR
FIPFRKGVGIPRVAVEDVVLSGVAIQAGDIVHVSYLTANRDPRKFDRPDEVDLERSGPPHMTFGWGAHHC
LGAPLAAAEIEIALGSLFERFPRLALAKPAEELQWNATSIWRYPLALPVTW

A0A2P8Q811 (SSN_C03)

ATGACTACAGCTGACACAGATATAGGAACGGCGTGCCCGTTTCGTCACGATGACGCCTTGGAGCCGGATC
CGTTTATGACCCGTATGCGTACGGAAGCTCCGGTTACTCGCGTGCAAATGCCATACGGCACTGGCACGTG
CTGGCTGGTGACCCGTTACGAGGATGTTCAGACCGTCACATCTGATCGCCGCTTCAGCCGTGCGGCGCTC
GTCGGTCGTGACTTCCCGCGTATTACCCCGGTGCCGATTGCCCAGTCGGAAGCGATTAATCTGATGGACC
CTCCGGCGTTGAACCGCGTGCGTCGTTTGGTGATCCACGCTTTCACCACTCCGCAAGTTGAGGCGCTACG
TCCGTGGACCCAGCGTACCACGAACGCCTTGCTGGACGCTATGGCAGAACATGGTGCGCCAGCGGATATC
GCGGCACATCTGGCCAACGAACTGCCGTTGATGACCATCTCCCAGCTGATGGACGTTCCGGAAGAGGATC
GTCCTCAACTGCGTCGTTGGGCGACCGCAATGATGAGTATGAGCGCTGCTGATCGGACGTCTGCCGTTGA
GGCTAAAAGCGCATTAAGAGATTATTTCACCGCGCTCACCACCGCTCGTAGACACGCCCCGGGTGACGAC
CTGATCTCTGCAATGGCTGCGGCGCGTATCGGTGACGAGATGCTGGAGGAGGGTGAACTTGCGGTTCTGG
CAATGCTGCTGGTGGTTACGGGCCATGACACCACTACGTACGACATCAGCAATATCGTGTACACCCTGTT
GACCCATCCGGCGCACCTGGCCCAGCTGCGCACGCGCCCAGAGCTGTTACCGCAAGCACTTGAGGAACTC
TTGCGCCACATTCCGTTCCGACAAGGTGTTGGCATTCCGCGTGTTGCGCTGGAGGACGTGGAACTGGGCG
GTCAGTTGATTCGTGCTGGCGATACCGTCCAAGTGAGCTATCTGACCGCGAACCGCGACAGCCTGGTGTA
TGAACGTCCGGATGAACTCGATTTTGAACGCCAGGCACCGGCCGGTCACATGGCGTTTGGTTATGGCAGC
CACCATTGTCTGGGTTCCCACCTGGCGCGCATGGTTCTGCAGGTATCCATCGGCACCCTGCTGGCGCGTT
TTCCGACCTTGCGCCTGGCGGTTCCGGAGGACGAGGTTCGTTGGAATACCGTCAGCATCTGGCGTTACCC
GCTGGCGCTGCCGGTCACCTGG

MTTADTDIGTACPFRHDDALEPDPFMTRMRTEAPVTRVQMPYGTGTCWLVTRYEDVQTVTSDRRFSRAAL
VGRDFPRITPVPIAQSEAINLMDPPALNRVRRLVIHAFTTPQVEALRPWTQRTTNALLDAMAEHGAPADI
AAHLANELPLMTISQLMDVPEEDRPQLRRWATAMMSMSAADRTSAVEAKSALRDYFTALTTARRHAPGDD
LISAMAAARIGDEMLEEGELAVLAMLLVVTGHDTTTYDISNIVYTLLTHPAHLAQLRTRPELLPQALEEL
LRHIPFRQGVGIPRVALEDVELGGQLIRAGDTVQVSYLTANRDSLVYERPDELDFERQAPAGHMAFGYGS
HHCLGSHLARMVLQVSIGTLLARFPTLRLAVPEDEVRWNTVSIWRYPLALPVTW

A0A345T2R4 (SSN_C04)

ATGGCTACAGAAACTAGGGTATGGGATTGTCCATTCAACCACAACCAGGCGTTGGAATTTGATCCGCTGC
TTAAGCAGTTGTTGACCGAAGCGCCGATCGCGCGCATCTCCATGCCGCATGGCGATGGGCACGCATGGCT
GGTAACTCGTTACGAAGATGTTTCCACCGTTGTTACGGATCAACGTTTCAGCCGTGCGGCAGGCATCGGC
CGTGACCTTCCGCGTATGACTCCGGAGCCGATTGCCCAAAGCGAAGCTATCAACCTGATGGATCCTCCGG
CGCACACCCGCCTGCGTAAACTGGTGGCCAAGGGTTTTACCCCGGCGCAGGTCGAGCAAATGGAACCGAA
AGTTCACGCCATCGTGGACCGTCTGTTGGACGAAATGGCGGCTCAACCGGCTCCGGTTGACTTCACCCGT
CATTTTGCCAGCCGCCTGCCGATCACCACCATTTGTGAAGTGCTGGCGATCCCGGAGGCGGATCAGCCGC
GTATCCGTCAACTGGCCATGGCCATGATGAGCACTTCTAGCGGTCGTCAAGCGACCAGAGAAGCGAAAGC
AGAGCTGCGCGCATATTTCCTGGACTTATCGGCGCGCCGTCGCGAGCAGCCGGGTGAAGACCTGATTAGC
GCTTTGGCGATGGCGCGTGACGGCGACGAGCTCTTGGACCCGACCGAGCTCGCGGTGATGGCGATGGTTC
TGACGGTCACCGGTCACGATACCTCGACGTACCAGCTGTCCAATATTCTGTATACCCTGTTGACGCATCC
GGAGCAACTTGCGCTGCTGCGCTCTCGTCCAGAATTATTGAATCAGGCTGTAGAGGAACTGTTACGCTTC
ATCCCCTTCCGCCACGGCGTTGGTATTGCTCGTGTTGCTACCGAGGACGTTGAACTGGGTGGTGTGACCA
TCCGTGCAGGCGAAGTGGTGCATGTCAGCTATCTGACCGCGAATCACGATCCGGAGGTGTTTGAACGCCC
AGATGAGCTGGACATCGAACGTCCGCCTGCTATGCACATGACCTTTGGTCACGGTACACACTACTGCATT
GGCGCAAACCTGGCCCGCGCAGAACTCCGTATTGCCTTTGCGTCTCTGCTGCAGCGTTTCCCGGAGCTGC

ATCTGGCGGTGCCGGCAGAGCAGGTTCCGTGGCATATTGGCAGCATTTGGCGTTACCCGGAGGAACTCCC
GATCGCGTGG

MATETRVWDCPFNHNQALEFDPLLKQLLTEAPIARISMPHGDGHAWLVTRYEDVSTVVTDQRFSRAAGIG
RDLPRMTPEPIAQSEAINLMDPPAHTRLRKLVAKGFTPAQVEQMEPKVHAIVDRLLDEMAAQPAPVDFTR
HFASRLPITTICEVLAIPEADQPRIRQLAMAMMSTSSGRQATREAKAELRAYFLDLSARRREQPGEDLIS
ALAMARDGDELLDPTELAVMAMVLTVTGHDTSTYQLSNILYTLLTHPEQLALLRSRPELLNQAVEELLRF
IPFRHGVGIARVATEDVELGGVTIRAGEVVHVSYLTANHDPEVFERPDELDIERPPAMHMTFGHGTHYCI
GANLARAELRIAFASLLQRFPELHLAVPAEQVPWHIGSIWRYPEELPIAW

A0A3N6F9N6 (SSN_C05)

ATGGAAAATACTTCAGTAGAGAACACAGAAACGGTTCGTTCCTGCCCGTTCGACTTCGCCGAGCGCTTGG
AGTTCGACCCGCAGCTGAAACAAATGCTGACCGAAGCACCGGTCAGCCGTATTCGTATGCCGTATGGTGA
AGGTGAGGCGTGGCTGGTGACCTCGTACGAAGATGTGCGTACGGTGACGACCGACAGAAGATTCAGCCGT
AGCGCGGTCCTGGGTCGTGATTTCCCGCGTATGACTCCGGAGCCGATTGTGCAGGCAGAGAGCATCAACC
TGATGGATCCGCCTGCGTCCAGCCGTCTGCGCTCGCTGGTGGCCAAGTGCTTTACCCCGCGTCGCGTCGA
TCGTATGCGTCGCCGTACACAAGGTATTGTGGATCGCCTGTTGAATGATATGGAAGACGCTGGTCCGCCA
GCTGATTTTGTTGCCCGTCTGTCCGCTCCGCTGCCGTTGGTTACGATTTGTGAGGTTCTGGACATCCCGG
AAGCGGACCGTCCGTGGCTGCGTGCGCACGCCCTGACCATGATGAATGTTGGTGCAGCGTCTAAGGCGGA
CGCGGTGCGCGCGAAAGGCGAGTTGAGAGCGTATTTTTCTGAACTGACCGCTGAGCGCCGTCGTTCTCCG
GGTGAGGACTTGATCAGCACCTTGGCGACCGCGAGAGACGGCGACGAGCTGCTGGACGACGATGAGCTTG
CGGTTATGGCAATGACGCTGCTTATCGCCGGCCAAGATACCACCACGTACCTGCTGGGCAACATGGGTTA
TACCTTGTTGACCCGTCCAGGCCTTAAGACCGCGTTACAGGCTGAACCGGAGCGCTTCCCGCGAACCATC
GAAGAACTACTGCGCCACATTCCGTTTCGTAAAGGCGTGGGCATCCCGCGTATTGCTACCGAAGACGTTG
AACTCGGCGGTGTTCCGATCAAAGCAGGCGATGTAGTTCACGTGAGCTACCTGACTGCGAACCGTGACGC
TGCGAAGTTTGAACGGCCAGATGAGCTGGATCCGGATCGCCCGAGCATTCCGCACATGACCTTTGGTTGG
GGTGCGCATCATTGCCTGGGGGCCCCTCTCGCTACCATGGAGTTGGAGGTGGCATGTAGTACGTTGTTGG
CCCGTTTCCCAGCACTGCGCCTGGCGGTGCCGCCCGAGGACGTCCGTTGGAACACCACGAGCATCTGGCG
TCACCCGCTGGCTCTGCCGGTTACCTGG

MENTSVENTETVRSCPFDFAERLEFDPQLKQMLTEAPVSRIRMPYGGEGEAWLVTSYEDVRTVTTDRRFSR
SAVLGRDFPRMTPEPIVQAESINLMDPPASSRLRSLVAKCFTPRRVDRMRRRTQGIVDRLLNDMEDAGPP
ADFVARLSAPLPLVTICEVLDIPEADRPWLRAHALTMMNVGAASKADAVRAKGELRAYFSELTAERRRSP
GEDLISTLATARDGDELLDDDELAVMAMTLLIAGQDTTTYLLGNMGYTLLTRPGLKTALQAEPERFPRTI
EELLRHIPFRKGVGIPRIATEDVELGGVPIKAGDVVHVSYLTANRDAAKFERPDELDPDRPSIPHMTFGW
GAHHCLGAPLATMELEVACSTLLARFPALRLAVPPEDVRWNTTSIWRHPLALPVTW

A0A4D4MK74 (SSN_C06)

ATGGCTTCAACAGCGGCAGGAAGGCCACCCGACCCGACCGGTCGCCGTACCGCCATGACTTCGCAACCGC
GTACCGAACGTCCGCCGGTGCACGTGTGGGAGGTTGAGGATCTTCCGGCTTTGGAATTTGATCCGATGAT
GACCACCCTGTTACGTGATGAGCCGGTTAGCCGTATTCGTCTGCCGTTCGCAGGCGACATGGATGCCTGG
CTGGTTACCCGTTACGAAGACGTTAAAGCGGTGGCATCCGACCCACGTTTTAGCCGGGAGGCGCTGCGTG
ACATCAAGGTGACTGCGATTAGCGGCCACCGTGTTGCCGGTGCTGCTGCGTTGAACTATACCGACCCGCC
GTATCATACGAAGCTGCGCAAGATCGTTAATAAGGCGTTCACCGGTCGTCACATGAAAGGTCTGCGTCCG
ATGACCCAGCAAACCGCGGATGAGCTGCTGGACGCGATGGAAGAGCAAGGTCCACCCGCCGATCTGATGA
CGCATCTGCATGGTCCGCTGCCTCTGGCCGTTGTTTCAGATCTGCTGGGAGTTCCGAGAGAAGAGCGCGA
CAAATTTACCGTGTGGCCTGATCAGATTTTGAACGCCGGCATCGGCGCTGAGGCGTCTATGGCAGCGAAA
GCGGAAGTTACGGCGTATGTTGTCGAGTTGCTGCGTAGCCGCTTCGGCCGGGCGGCGGAAGGCACAGACG
ATCTCGCGGGGGTTCTGGCGCAGGCGTGGGAAGCCGGCGAAATTGAACAGGACGAGGCGGTGTCTTTGGC

TACCGCGATTGTAATCTCGGGTGCGCACGCAGTTCGTTACAATAGCGCAAACATGGTGTACATGCTGCTG
ACCCATCCGGAACTGATGGATCGCTTGAGAGACGACTCCGGTTTGGTGCCGCAGGCAGTCGATGAGCTGA
TTCGTCACATCCCGCATCGTAATGCTGTGGGCATCCCGCGCATTGCAATGGAAGACGTCGAGGTGGGTGG
TCACGTTATCCCGGCGGGCGACGCAGTGTACGTGAGCTACCTCGCGGCGAACCGTGATCCGGCTGTGTTC
GAGAACCCGGACGCGGTCGATTTTGACCGTCAGGGTGTCTCCCATCTGAGTTTCGGTCACGGCGTCCACC
ACTGCATGGGTGCTATGTTAGCGCGTATGGAGAGCGAAGTGATGGTTTCTAGCTTACTCGACCGCTTCCC
GAAGTTGCGCCTGGCCGGCCCACCGCAAGATACCGTTTTTCAGTCCAAAGCATTCATCCGCGGTCCGCAA
ACTCTGGTGGTGACGTGG

MASTAAGRPPDPTGRRTAMTSQPRTERPPVHVWEVEDLPALEFDPMMTTLLRDEPVSRIRLPFAGDMDAW
LVTRYEDVKAVASDPRFSREALRDIKVTAISGHRVAGAAALNYTDPPYHTKLRKIVNKAFTGRHMKGLRP
MTQQTADELLDAMEEQGPPADLMTHLHGPLPLAVVSDLLGVPREERDKFTVWPDQILNAGIGAEASMAAK
AEVTAYVVELLRSRFGRAAEGTDDLAGVLAQAWEAGEIEQDEAVSLATAIVISGAHAVRYNSANMVYMLL
THPELMDRLRDDSGLVPQAVDELIRHIPHRNAVGIPRIAMEDVEVGGHVIPAGDAVYVSYLAANRDPAVF
ENPDAVDFDRQGVSHLSFGHGVHHCMGAMLARMESEVMVSSLLDRFPKLRLAGPPQDTVFQSKAFIRGPQ
TLVVTW

A0A4R7HLN1 (SSN_D01)

ATGAGTGAAGATACTTCAACACTAACGGGACAAGCACCATCCCCGGTTAGAGTGGCAGGCCCGGCCGAGG
AGATCGTTTTGCCGGGTCACACACCGCCACCGGTTCGCGATTGGCCGCCGCCGGACCTGGATGGTACTGA
GTTCGATCCGGTGCTGGCGGGCCTGATGCGTGAAGGTCCACTGACTCGTATCCGCTTACCGTTCGGCGAG
GGCTGGGCATGGCTGGCGACCCGCTACGAAGACGTGCGCTTAATTACCAACGATCCGCGTTTTAGCCGTA
CCGAGGTGACCCGTCGTCAGGTTACCCGCCTGGCGCCGAATTTTGCACCGAGACCCGGCAGCCTGGCGTG
GGCAGATCAGCCGGAACACAACCGTCTGCGTAAACCGGTGGCGGGCGCTTTCACCGTTTCTGCCATGAAG
CGCCTCCGCCCACGCGCGCAAGAAATCCTGGACGGCCTCGTCGATGCGGTAGTTGCGGATGGTCCGCCTG
CGGATCTGGTCGAGCGAGTGCTTGAGCCTTTTCCGTTGACGGTTGTCAGCGAAGTTATGGGTGTTCCGCC
GGCTGATCGCGCTCGTGTGCATGGTTGGACCCGTCAGATTATCTCCACGTCTGGTGGCGCCGAAGCGGCG
GGCCGTGCGAAAGCTGGCCTGTACGAGTGGATTGGCACCATGGTTCGCGACCGTGCGCACAGCACGGGTG
AGGATGTGTATAGCTTGCTGGGTGCAGCCGTTACCCGTGGTGAGATCTCCGAAGAGGAAGCCGTTGGTTT
GGCCGGTCCGTTGCAAATCGGCGGTGAGGCCGTGACGCACAATTGTGGTCAGATGCTGTTCCTGCTGCTC
ACCCATCCGGATCTGATGGAACGCATGCGTGACCGTCCACTGGAACGTGGTCCGGTATTAGACGAACTGC
TGCGCCATATTCCGCACAGGTCGACCGTTGGTCTGGCGCGTATCGCACTCGAGGATGTCGAGCTGGCGGG
TCACCGTATTGCGGCGGGCGAGCCGGTCTATGTTAGCTACCTGGCCGCTAACCGGGACCCGGACGTCTTT
CCGGAACCCGACCGTATCGACCCGGACCGTGCTGCTTGCCCGCACCTGGCGTTCGGAAACGGCCCACATT
ATTGCACCGGTGCGGTGCTGGCGCGTTTACAGACCGAATTGCTGGTGGACGCAGTGCTAGACCGCCTGCC
GCGTTTACGTTTGGCGCTGCCGGCAGAGGACGTACGTTGGCGTCATAAGACCATGATTCGTGGCCCATTG
ACCCTGCCGGTGACGTGG

MSEDTSTLTGQAPSPVRVAGPAEEIVLPGHTPPPVRDWPPPDLDGTEFDPVLAGLMREGPLTRIRLPFGE
GWAWLATRYEDVRLITNDPRFSRTEVTRRQVTRLAPNFAPRPGSLAWADQPEHNRLRKPVAGAFTVSAMK
RLRPRAQEILDGLVDAVVADGPPADLVERVLEPFPLTVVSEVMGVPPADRARVHGWTRQIISTSGGAEAA
GRAKAGLYEWIGTMVRDRAHSTGEDVYSLLGAAVTRGEISEEEAVGLAGPLQIGGEAVTHNCGQMLFLLL
THPDLMERMRDRPLERGPVLDELLRHIPHRSTVGLARIALEDVELAGHRIAAGEPVYVSYLAANRDPDVF
PEPDRIDPDRAACPHLAFGNGPHYCTGAVLARLQTELLVDAVLDRLPRLRLALPAEDVRWRHKTMIRGPL
TLPVTW

A0A4Y3VFB2 (SSN_D02)

ATGTCAGAAGATACAATTATAGCAACTGCTCCGCCACATGTTGGGTGCGGTATGTCTGACGCGATGACCA
CCGGAACCGCGCCGCCACCGGTTCGTCAATGGCCTTCGCTGGATTTGACCGGTGTGGACTTTGATCCGGT

GCTGCGTGATCTGATGGACGACGGCCCGGTTACCCGCATTCAACTGCCGAATGGTGAGGGTTGGGCATGG
CTGGTTACGCGGTACGATGATGTACGTATGGTTGCGAATGATCCGCGCTTTAGCCGTGAGGCCGTGATGG
ACCGTCCGGTTACCCGCCTGGCTCCGCACTTTATTCCGGCTCGTGGCGCGGTGGGTTTCCTGGACCCGCC
GGACCACACCCGCTTGAGACGTAGCGTTGCGGCTGCGTTCACCGCCCGTGGTGTCGAACGTGTTCGTGAG
CGCAGCCGCCAAATGCTCGACGAGCTGGTCGATGAGTTGCTCCAGGATGGTCCGCCGGCGGACTTGACTG
CGACGGTGTTGTCCCCGTTCCCGATTGCAGTGATCTGCGAACTGATGGGTGTACCGGCGGCGGACCGTCA
CAACATGCACGTGTGGACCCAGCTGATCCTGTCGAGCTCCCAAGGTGCAGATGTGAGCGAAAGGGCGAAG
AAAGAAATGTCCGCATACTTCAGCGATCTGATCGGCCTGCGCGAGAACAGCACCGCGGAAGATGTCACTT
CACTTTTAGGTGCCGCGGTCGGCCGCGGCGAGATCACCCTGGAAGAGGCTGTGGGCCTGGCAGTGCTGAT
CCAGATTGGTGGCGAAGCCGTGACCAACAACAGCGGTCAGATGTTCTATCTCTTGCTTACGCGTCCGGAC
TTAGCTGAAAGACTGCGTGCAGACCCAGCGATCCGTCCGCAGGCAATTGACGAATTGTTACGTTGGATTC
CGCACCGTAATGCAGTGGGCCTGTCTCGTATTGCGACCGAAGACGTGCATATCAAAGGCCATCTGATCCG
TGCGGGTGATGCTATCTACGTGAGCTATCTGGCCGCCAACCGCGATCCGGACGTTTTTCCGGATCCGGAG
ACAATTGATTTCTCTCGTAATCCGAACCCGCATGCAGCGTTTGGCTTCGGTCCTCACTATTGTGTTGGCG
GTATGCTGGCAAAGCTGGAGAGCGAGTTGTTGGTTGACGCGCTGTTGGACCAGGTTCCGGGTCTGCGCCT
GGCTGTCCCGCCGGAGGAAGTTCCGTTTCGTAAAGGTGCGCTGATCCGCGGCCCAGAGGCTCTGCCGGTT
ACGTGG

MSEDTIIATAPPHVGCGMSDAMTTGTAPPPVRQWPSLDLTGVDFDPVLRDLMDDGPVTRIQLPNGEGWAW
LVTRYDDVRMVANDPRFSREAVMDRPVTRLAPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRE
RSRQMLDELVDELLQDGPPADLTATVLSPFPIAVICELMGVPAADRHNMHVWTQLILSSSQGADVSERAK
KEMSAYFSDLIGLRENSTAEDVTSLLGAAVGRGEITLEEAVGLAVLIQIGGEAVTNNSGQMFYLLLTRPD
LAERLRADPAIRPQAIDELLRWIPHRNAVGLSRIATEDVHIKGHLIRAGDAIYVSYLAANRDPDVFPDPE
TIDFSRNPNPHAAFGFGPHYCVGGMLAKLESELLVDALLDQVPGLRLAVPPEEVPFRKGALIRGPEALPV
TW

A0A6B3DVM9 (SSN_D03)

ATGGCTCCACCTCCCGAGGAACCGGCACCCCCACCGATCCGCGATTGGGCCGCTCTGGATTTGGCGGGCA
CCGAATTTGATCCGGTGCTGCGTGAACTGATGCGTGACGGCCCTATCACCCGTATTCGATTACCGCACGG
CGAGGGTTGGGCGTGGCTGGCCACCCGCTACGATGATGTGAAACTGATTACCAATGATCCGCGCTTCAGC
CGTGCAGAAGTTACGAGACGCCAAGTGACGCGCCTCGCTCCGCACTTCAAGCCGCGTCCGGGCAGCCTGG
CGTGGGCGGATCAGCCGGACCACAACCGTTTGCGTAAGCCGGTGTCTGCAGCCTTCACCGTTGGTGCAAT
GAAACGTCTTCGTCCCCGTGCTCAAGAACTATTAGACGAATTGGTTGACGCGCTGGTCCAGGCAGGCCCG
CCAGCGGACCTGGTTGAACGCGTTCTGGAGCCGTTTCCGATTGCGGTGGTGTCCGAGGTTATGGGTGTTC
CGCCGGCGGACCGTGAGCAGGTGCATGCATGGACCCGTCAAATCATCAGCATGTGTGGCGGTGCGGAGGC
GGCAGACCGTGCTAAGCAGGGTATGTATGGTTGGATTCGTCGTACCGTGAGAGAGCGCCGTGACAGCAGC
GCGGAGGATGTCTTGTCACTGCTTGGTGCAGCGGTCGGTCGTGGCGAGATCTCGGAAGAAGAGGCGGTGG
GCCTGGCGGGCCCACTGCAGATTGGTGGTGAAGCGGTGACGCATAACTGCGGTCAAATGCTGTACCTGCT
GTTAACTCGCCCAGAACTGATGGCACGCATGCGTGAGCGTCCGGGCGAGCGCGGCCCGGTAGTTGATGAA
TTGCTGCGCCATATTCCGCATCGTAGCACCGTTGGTTTGGCGCGTATCGCCCTCGAAGATGTAGAGCTGC
ACGGCGTTCGTATTGCCGCTGGCGAGCCGGTCTATGTGTCTTATCTGGCTGCCAACCGCGACCCGGACGT
GTTCCCGGACCCGGACCGTATCGACCCTGATAGAGAACCGAGGCCGCACCTGTCCTTTGGTAATGGTCCG
CACTACTGCACCGGTGCAGTGCTCGCGCGTTTGCAGACCGAGTTGCTGATCGATACCCTGCTGGACCGCC
TGCCGGGTCTGCGCTTGGCCGTATCCCCGGGTCAGGTTGCGTGGCGTCGTCAAACCATGATCCGTGGTCC
GCGTACGCTGCCGGTTACTTGG

MAPPPEEPAPPPIRDWAALDLAGTEFDPVLRELMRDGPITRIRLPHGEGWAWLATRYDDVKLITNDPRFS
RAEVTRRQVTRLAPHFKPRPGSLAWADQPDHNRLRKPVSAAFTVGAMKRLRPRAQELLDELVDALVQAGP
PADLVERVLEPFPIAVVSEVMGVPPADREQVHAWTRQIISMCGGAEAADRAKQGMYGWIRRTVRERRDSS
AEDVLSLLGAAVGRGEISEEEAVGLAGPLQIGGEAVTHNCGQMLYLLLTRPELMARMRERPGERGPVVDE

LLRHIPHRSTVGLARIALEDVELHGVRIAAGEPVYVSYLAANRDPDVFPDPDRIDPDREPRPHLSFGNGP
HYCTGAVLARLQTELLIDTLLDRLPGLRLAVSPGQVAWRRQTMIRGPRTLPVTW

A0A6G3S9J1 (SSN_D04)

ATGACATTTGAACCAACGACAACTTCACCCCCGGAAGTACGCTTCTGGCCTGTGGATGACTTGCCGGGCC
TGCAACCTGACCCGCTGTTTGATGAAGTCCTGCGTGAGGAGCCCGTGACCAGAATCAAACTGCCGCATGG
TTCTGGCGATGCGTGGTTGGTTACGCGTTACGAGGACGTGCGCTTCGTGACTTCAGATCCGCGTTTTAGC
CGTGAGCAAGTTGTTGGGTGCCCGGTTACGACCATGACGCCTACTCCGGTCGCGTCGCAGACCGCTGGTC
TTCAATATATCGATGCTCCGCGTCATACGCGGCTGCGCCGCGTGGTGGCGCGTGCTTTCACCGCGAGATC
GATGCAGCGTTTGCGTCCGATGGCGGAACGTACCGCGCATGCGTTGTTGGACGGCATGGAACAAAGCGGT
TCCCCGGCGGACCTGATGGAGCACCTGCATACGCCGTTTCCGATTGCGGTAGTCTGTTATTTCCTGGGTG
CAGATGAAGATGATTGGCATGCGTGGGCAGACAACAGCCGTGTTATGCTGAGCAAAACCGACGATAAAGA
CGCCGTGCAGGCGGCTCGTGCAGACACCAGAAGCCGCGTTGTCGCACTCTTACGTCGTAGGCGCACCGAC
CCGGGCGACGACCTGGCGGGTGTTCTCGCGCAGGCGGCTGCGTCCGGTGAAATCACCGATGACGAGGCAT
TGTCTCTGGCCATGGCGGTTTACGTTTCTGGCGGTCACGCAGTCGCTAACAACTCCGGCAGCATGATGTA
CGCCTTGTTGACCCATCCGGACCACCTGGACCGTCTGCGCCGTGATCCGGAAGTTGTGCCAGCTGCTGTC
GAAGAGCTGTATCGCTACGTGCCGCACCGTAACGGTGTTGGTATTCCGCGTATCGCCACCGAGGATGTGC
CGGTTGGCGGTCACGTGATCCCGAAGGGTGATGTTGTGTACAATGCGTATGTTGCCGCTAATCGTGACCC
AGAAGTGTTTCCGGATCCGGATCTGCTGGACTTTGATCGTCAGGGACCGGGCCACCTTGCGTTCGGCCAC
GGCCCACACTACTGCCTGGCTAGCCTGATGGCGCGTATGGAAGCAGAGGTGATGATTACCGTTCTGCTGG
AGCGCTTCCCGCGTCTCAGACTTGCCGTTCCGAAGGACGAGGTTGAGTTCCAGCGTGAGGGTCTGATTCG
TGGTCCGAGCACCCTGCCGGTGGCGTGG

MTFEPTTTSPPEVRFWPVDDLPGLQPDPLFDEVLREEPVTRIKLPHGSGDAWLVTRYEDVRFVTSDPRFS
REQVVGCPVTTMTPTPVASQTAGLQYIDAPRHTRLRRVVARAFTARSMQRLRPMAERTAHALLDGMEQSG
SPADLMEHLHTPFPIAVVCYFLGADEDDWHAWADNSRVMLSKTDDKDAVQAARADTRSRVVALLRRRRTD
PGDDLAGVLAQAAASGEITDDEALSLAMAVYVSGGHAVANNSGSMMYALLTHPDHLDRLRRDPEVVPAAV
EELYRYVPHRNGVGIPRIATEDVPVGGHVIPKGDVVYNAYVAANRDPEVFPDPDLLDFDRQGPGHLAFGH
GPHYCLASLMARMEAEVMITVLLERFPRLRLAVPKDEVEFQREGLIRGPSTLPVAW

A0A6H0CVT5 (SSN_D05)

ATGTCATTTCAAGAACCCACAGTACTAGAGTGCCCGTTCGACTTCGCTGAGGCGCTGGAGTTCGACCCGG
AATTGGACGCGCTCGCGCGCCGCGGTCCGGTAAGCCGTATCCGCTTACCGTATGGTGAAGCGGAGGCATG
GCTGGTGACGAGCTTTGCAGGGGTGCGCCAAGTGACCTGTGATCCACGATTCAGCCGTGCGGCGATTGTT
GGCCGTGACTACCCGCGTATGACCCCGGAGCCCATCGTGTCGCCGGAGAGCATTAACGTTACCGATCCAC
CGCACGCAACCCGTCTGCGTCATGTTGCTGCGCAGGCATTTACGCGTGACCGCGTGGCGGCAATGCGTCC
GGCTGTGGATCGTGTTGTTGCGGGTCTGCTGACGGAAATGGCAGAAACCGGTCCTCCGGCAGATTTGGTC
ACACACCTGTCTGTTCCGCTGCCTCACTTGACCATCTGCGAATTGCTGGCGGTGCCGGAGGCCGATCGCG
ACGAGCTGCGCGACCGGACCATGCGTCTGTTGGCTACTTCAGCGGATGCCCGTCAAACCGCAGCCGAAGC
GAAGGCAAGCCTGAGAGCCTATTTTGCTGGCCTGGTGCCGGAACGTCGTCGCGCCCCAGGTGAGGATCTG
TTGTCCACCCTGGCGACCGCTCCGGTCCCGGAGGGTGAGGAACCGCTGAGCGATGATGAACTCGCGGTGC
TCGCGGTCACTTTGATTCTGTCTGGCAACGACACTGCGACGTGCCAGATTTCCGACATCGCGTACGTTCT
GCTGACACATCCGGAGCAGCAAGCTGCGCTGGCCAGAGACCCGGGCCGTTTCCCTGGCGCACTGGAAGAA
CTGTTACGTTTCATCCCGTTTCGTAAAGGTGTTGGTATTCCGCGTATTGCACTTGAGGACGCGGAAGTTG
AAGGCACCCTGATCAGCGCAGGAGATTACGTTCATGTTAGCTATCTGGCGGCGAATCGTGACCCGGCGGT
CTTTCCGGATCCGCACCGTCTGGACCTGGAGCGTCCGGTGCGTCCGCACATGACCTTTGGTTGGGGTGGT
CATCACTGTCTGGCGGCTCCGCTTGCTCGCGCGGAATTGGACAGTGCTGTGTCCGGCTTGTTGACGCGCT
TCCCAACCCTGAGACTTGCAGTCGCCCCGGACCAGATCGAGTGGGATACCGGCACCATCCGTCGTTTCCC
GGTTCGCCTACCGGTGACCTGG

MSFQEPTVLECPFDFAEALEFDPELDALARRGPVSRIRLPYGEAEAWLVTSFAGVRQVTCDPRFSRAAIV
GRDYPRMTPEPIVSPESINVTDPPHATRLRHVAAQAFTRDRVAAMRPAVDRVVAGLLTEMAETGPPADLV
THLSVPLPHLTICELLAVPEADRDELRDRTMRLLATSADARQTAAEAKASLRAYFAGLVPERRRAPGEDL
LSTLATAPVPEGEEPLSDDELAVLAVTLILSGNDTATCQISDIAYVLLTHPEQQAALARDPGRFPGALEE
LLRFIPFRKGVGIPRIALEDAEVEGTLISAGDYVHVSYLAANRDPAVFPDPHRLDLERPVRPHMTFGWGG
HHCLAAPLARAELDSAVSGLLTRFPTLRLAVAPDQIEWDTGTIRRFPVRLPVTW

A0A6I6FHD3 (SSN_D06)

ATGTCATTTCAAGAACCCACAGTACTAGAGTGCCCGTTCGACTTCGCTGAGGCGCTGGAGTTCGACCCGG
AATTGGACGCGCTCGCGCGCCGCGGTCCGGTAAGCCGTATCCGCTTACCGTATGGTGAAGCGGAGGCATG
GCTGGTGACGAGCTTTGCAGGGGTGCGCCAAGTGACCTGTGATCCACGATTCAGCCGTGCGGCGATTGTT
GGCCGTGACTACCCGCGTATGACCCCGGAGCCCATCGTGTCGCCGGAGAGCATTAACGTTACCGATCCAC
CGCACGCAACCCGTCTGCGTCATGTTGCTGCGCAGGCATTTACGCGTGACCGCGTGGCGGCAATGCGTCC
GGCTGTGGATCGTGTTGTTGCGGGTCTGCTGACGGAAATGGCAGAAACCGGTCCTCCGGCAGATTTGGTC
ACACACCTGTCTGTTCCGCTGCCTCACTTGACCATCTGCGAATTGCTGGCGGTGCCGGAGGCCGATCGCG
ACGAGCTGCGCGACCGGACCATGCGTCTGTTGGCTACTTCAGCGGATGCCCGTCAAACCGCAGCCGAAGC
GAAGGCAAGCCTGAGAGCCTATTTTGCTGGCCTGGTGCCGGAACGTCGTCGCGCCCCAGGTGAGGATCTG
TTGTCCACCCTGGCGACCGCTCCGGTCCCGGAGGGTGAGGAACCGCTGAGCGATGATGAACTCGCGGTGC
TCGCGGTCACTTTGATTCTGTCTGGCAACGACACTGCGACGTGCCAGATTTCCGACATCGCGTACGTTCT
GCTGACACATCCGGAGCAGCAAGCTGCGCTGGCCAGAGACCCGGGCCGTTTCCCTGGCGCACTGGAAGAA
CTGTTACGTTTCATCCCGTTTCGTAAAGGTGTTGGTATTCCGCGTATTGCACTTGAGGACGCGGAAGTTG
AAGGCACCCTGATCAGCGCAGGAGATTACGTTCATGTTAGCTATCTGGCGGCGAATCGTGACCCGGCGGT
CTTTCCGGATCCGCACCGTCTGGACCTGGAGCGTCCGGTGCGTCCGCACATGACCTTTGGTTGGGGTGGT
CATCACTGTCTGGCGGCTCCGCTTGCTCGCGCGGAATTGGACAGTGCTGTGTCCGGCTTGTTGACGCGCT
TCCCAACCCTGAGACTTGCAGTCGCCCCGGACCAGATCGAGTGGGATACCGGCACCATCCGTCGTTTCCC
GGTTCGCCTACCGGTGACCTGG

MIEDTATRPVLGPPPVRHWPALDISGTDFDPVLAELMREGPVSRIQLPHGEGWAWLVTRYEDVRMVTNDP
RFTRAPVVDGKVTRLAPHFKPKKGSLAFADQPDHNRLRAAVAPAFTLRGVERLRSHAQDTLDGMVDALLD
AGPPADLVARVLEPFPIAVVCDVMGVPAADRERMHAWTRQIVSSSGGAEASARAKDGMYGWFTETIRSRY
GSTDEDVISLLGAAVGRGDITEEEAIGLVGPLQIGGEAVTANTGQMLYLLLTRPALLERLRTEPAARPRA
LDELLRYIPHRSSVGLARIAQEDVELHGVRISAGDAVYVSYLAANRDPDVFPDPERIDLDRAPNPHVAFG
HGPHYCMGAMLARMQTELLVDTLLDRLPATLGLAVPPQEVPWRRSTLIRGPEGLPVTW

CYP158A2_A245Q (A2Rd1_08_G03)

ATGACTGAAGAGACAATATCACAAGCTGTACCGCCTGTTCGCGACTGGCCGGCTGTGGACCTGCCGGGTT
CCGACTTCGACCCGGTTTTGACCGAGCTGATGCGTGAGGGTCCAGTGACTCGTATCAGTCTGCCGAATGG
TGAAGGTTGGGCTTGGTTGGTGACTCGTCACGATGATGTTCGCCTAGTGACCAATGATCCGCGTTTTGGC
CGCGAGGCGGTTATGGACCGTCAAGTTACGCGTTTGGCGCCTCACTTTATCCCGGCGCGTGGCGCGGTAG
GCTTCCTGGACCCGCCGGATCACACCCGTTTGCGGCGTAGCGTTGCGGCGGCGTTTACCGCACGTGGCGT
GGAGCGTGTTCGTGAGCGCTCCCGCGGTATGCTGGACGAACTGGTCGACGCCATGCTACGCGCGGGTCCG
CCAGCCGATCTGACCGAAGCCGTGCTGAGCCCGTTCCCGATTGCAGTTATTTGTGAACTTATGGGTGTTC
CGGCAACCGATCGTCACAGCATGCATACCTGGACCCAGCTGATTCTGTCTAGCAGCCATGGCGCGGAGGT
GAGCGAGCGCGCGAAAAACGAAATGAACGCCTACTTCTCCGACTTGATTGGCTTAAGATCGGACTCTGCA
GGCGAGGACGTGACGAGCCTGCTGGGCGCCGCTGTGGGTCGTGACGAGATCACCCTGTCTGAGGCTGTCG
GTTTGGCGGTTCTGTTGCAAATTGGCGGTGAGCAGGTGACGAACAACAGCGGTCAGATGTTTCATCTGCT
GCTTTCCAGGCCCGAACTGGCGGAAAGACTGCGCTCGGAACCGGAGATCCGTCCGCGTGCGATCGATGAG
TTACTGCGTTGGATTCCGCACCGTAATGCTGTCGGACTGAGCCGTATCGCGTTGGAAGATGTGGAGATTA

AGGGCGTACGTATCCGGGCGGGTGATGCAGTGTACGTGAGCTATCTGGCTGCGAACCGTGACCCGGAGGT
TTTTCCGGACCCGGACCGCATCGATTTTGAAAGATCACCGAACCCGCATGTTTCTTTCGGTTTCGGCCCG
CACTATTGCCCGGGCGGCATGCTCGCCCGTCTCGAAAGCGAATTGTTGGTTGACGCGGTACTGGACCGCG
TGCCGGGTCTTAAGCTGGCAGTTGCTCCGGAAGATGTTCCGTTCAAAAAGGGTGCGCTGATCCGTGGTCC
GGAAGCGCTGCCGGTCACCTGG

MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLPNGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLAPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSSSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEQVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNAVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTW

CYP158A2_P45E, A83L, S181G, A245Q, A290M (A2Rd2_01_D09)

ATGACTGAAGAGACAATATCACAAGCTGTACCGCCTGTTCGCGACTGGCCGGCTGTGGACCTGCCGGGTT
CCGACTTCGACCCGGTTTTGACCGAGCTGATGCGTGAGGGTCCAGTGACTCGTATCAGTCTGGAGAATGG
TGAAGGTTGGGCTTGGTTGGTGACTCGTCACGATGATGTTCGCCTAGTGACCAATGATCCGCGTTTTGGC
CGCGAGGCGGTTATGGACCGTCAAGTTACGCGTTTGCTGCCTCACTTTATCCCGGCGCGTGGCGCGGTAG
GCTTCCTGGACCCGCCGGATCACACCCGTTTGCGGCGTAGCGTTGCGGCGGCGTTTACCGCACGTGGCGT
GGAGCGTGTTCGTGAGCGCTCCCGCGGTATGCTGGACGAACTGGTCGACGCCATGCTACGCGCGGGTCCG
CCAGCCGATCTGACCGAAGCCGTGCTGAGCCCGTTCCCGATTGCAGTTATTTGTGAACTTATGGGTGTTC
CGGCAACCGATCGTCACAGCATGCATACCTGGACCCAGCTGATTCTGTCTGGCAGCCATGGCGCGGAGGT
GAGCGAGCGAGCGAAAAACGAAATGAACGCCTACTTCTCCGACTTGATTGGCTTAAGATCGGACTCTGCA
GGCGAGGACGTGACGAGCCTGCTGGGCGCCGCTGTGGGTCGTGACGAGATCACCCTGTCTGAGGCTGTCG
GTTTGGCGGTTCTGTTGCAAATTGGCGGTGAGCAGGTGACGAACAACAGCGGTCAGATGTTTCATCTGCT
GCTTTCCAGGCCCGAACTGGCGGAAAGACTGCGCTCGGAACCGGAGATCCGTCCGCGTGCGATCGATGAG
TTACTGCGTTGGATTCCGCACCGTAATATGGTCGGACTGAGCCGTATCGCGTTGGAAGATGTGGAGATTA
AGGGCGTACGTATCCGGGCGGGTGATGCAGTGTACGTGAGCTATCTGGCTGCGAACCGTGACCCGGAGGT
TTTTCCGGACCCGGACCGCATCGATTTTGAAAGATCACCGAACCCGCATGTTTCTTTCGGTTTCGGCCCG
CACTATTGCCCGGGCGGCATGCTCGCCCGTCTCGAAAGCGAATTGTTGGTTGACGCGGTACTGGACCGCG
TGCCGGGTCTTAAGCTGGCAGTTGCTCCGGAAGATGTTCCGTTCAAAAAGGGTGCGCTGATCCGTGGTCC
GGAAGCGCTGCCGGTCACCTGG

MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLENGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLLPHFIPARGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSGSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEQVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNMVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTW

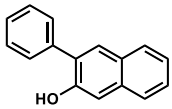
CYP158A2_P45E, A83L, R90Q, S181G, A245Q, A290M (A2Rd4_16_A11)

ATGACTGAAGAGACAATATCACAAGCTGTACCGCCTGTTCGCGACTGGCCGGCTGTGGACCTGCCGGGTT
CCGACTTCGACCCGGTTTTGACCGAGCTGATGCGTGAGGGTCCAGTGACTCGTATCAGTCTGGAGAATGG
TGAAGGTTGGGCTTGGTTGGTGACTCGTCACGATGATGTTCGCCTAGTGACCAATGATCCGCGTTTTGGC
CGCGAGGCGGTTATGGACCGTCAAGTTACGCGTTTGCTGCCTCACTTTATCCCGGCGCAGGGCGCGGTAG
GCTTCCTGGACCCGCCGGATCACACCCGTTTGCGGCGTAGCGTTGCGGCGGCGTTTACCGCACGTGGCGT
GGAGCGTGTTCGTGAGCGCTCCCGCGGTATGCTGGACGAACTGGTCGACGCCATGCTACGCGCGGGTCCG
CCAGCCGATCTGACCGAAGCCGTGCTGAGCCCGTTCCCGATTGCAGTTATTTGTGAACTTATGGGTGTTC
CGGCAACCGATCGTCACAGCATGCATACCTGGACCCAGCTGATTCTGTCTGGCAGCCATGGCGCGGAGGT

GAGCGAGCGAGCGAAAAACGAAATGAACGCCTACTTCTCCGACTTGATTGGCTTAAGATCGGACTCTGCA
GGCGAGGACGTGACGAGCCTGCTGGGCGCCGCTGTGGGTCGTGACGAGATCACCCTGTCTGAGGCTGTCG
GTTTGGCGGTTCTGTTGCAAATTGGCGGTGAGCAGGTGACGAACAACAGCGGTCAGATGTTTCATCTGCT
GCTTTCCAGGCCCGAACTGGCGGAAAGACTGCGCTCGGAACCGGAGATCCGTCCGCGTGCGATCGATGAG
TTACTGCGTTGGATTCCGCACCGTAATATGGTCGGACTGAGCCGTATCGCGTTGGAAGATGTGGAGATTA
AGGGCGTACGTATCCGGGCGGGTGATGCAGTGTACGTGAGCTATCTGGCTGCGAACCGTGACCCGGAGGT
TTTTCCGGACCCGGACCGCATCGATTTTGAAAGATCACCGAACCCGCATGTTTCTTTCGGTTTCGGCCCG
CACTATTGCCCGGGCGGCATGCTCGCCCGTCTCGAAAGCGAATTGTTGGTTGACGCGGTACTGGACCGCG
TGCCGGGTCTTAAGCTGGCAGTTGCTCCGGAAGATGTTCCGTTCAAAAAGGGTGCGCTGATCCGTGGTCC
GGAAGCGCTGCCGGTCACCTGG

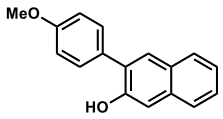
MTEETISQAVPPVRDWPAVDLPGSDFDPVLTELMREGPVTRISLENGEGWAWLVTRHDDVRLVTNDPRFG
REAVMDRQVTRLLPHFIPAQGAVGFLDPPDHTRLRRSVAAAFTARGVERVRERSRGMLDELVDAMLRAGP
PADLTEAVLSPFPIAVICELMGVPATDRHSMHTWTQLILSGSHGAEVSERAKNEMNAYFSDLIGLRSDSA
GEDVTSLLGAAVGRDEITLSEAVGLAVLLQIGGEQVTNNSGQMFHLLLSRPELAERLRSEPEIRPRAIDE
LLRWIPHRNMVGLSRIALEDVEIKGVRIRAGDAVYVSYLAANRDPEVFPDPDRIDFERSPNPHVSFGFGP
HYCPGGMLARLESELLVDAVLDRVPGLKLAVAPEDVPFKKGALIRGPEALPVTW

Chemical synthesis. Casey Roos and Suman Chakrabarty synthesized and characterized the substrates and product standards reported in this section. All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenk techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, Inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Machery-Nagel 60 F_{254} precoated silica TLC plates (0.25 mm) or Merck Silica Gel 60 F_{254} precoated silica TLC plates (0.25 mm) which were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol green stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230–400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated.

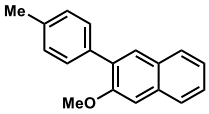


3-phenylnaphthalen-2-ol (4.31). The synthesis of this compound was carried out using Suzuki cross-coupling of phenylboronic acid with 2-bromo-3-methoxynaphthalene, followed by methyl-ether deprotection using BBr_3 as follows. A round bottom flask equipped with a stir bar was charged with K_2CO_3 (829 mg, 6.00 mmol, 3.00 equiv.), 2-bromo-3-methoxynaphthalene (472 mg, 2.00 mmol, 1.00 equiv.), phenylboronic acid (317 mg, 2.60 mmol, 1.30 equiv.) and $Pd(PPh_3)_4$ (46 mg, 0.4 mmol, 20 mol%), then evacuated and refilled with nitrogen (3x). To this reaction flask, water (3 mL), which had been degassed by sparging with nitrogen for 30 min, and 1,2-dimethoxyethane (14 mL) were added and the reaction was heated at reflux (100–110 °C) overnight. Then, the reaction flask was cooled to room temperature and extracted with ethyl acetate (20 mL x 3). The combined organic layers were dried over magnesium sulfate, filtered, and concentrated under vacuum. The crude reaction residue (560 mg) was diluted in $CHCl_3$ (6 mL) and cooled to –78 °C and to this mixture was added neat BBr_3 (0.38 mL, 4.00 mmol, 2.00 equiv.). The reaction was allowed to warm to room temperature overnight and was then quenched by the slow addition of methanol (20 mL). The mixture was concentrated under reduced pressure and more methanol was added and the mixture was concentrated (20 mL x 3). This is to ensure complete removal of boron as $B(OMe)_3$. The residue was purified by flash chromatography over silica gel (10% ethyl acetate in hexanes) to afford the pure title compound as a yellow solid (360 mg, 82%

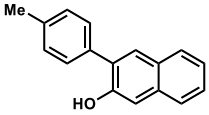
yield). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.07; **^{1}H NMR** (600 MHz, $CDCl_3$): δ 7.81 (d, *J* = 8.2 Hz, 1H), 7.80 (s, 1H), 7.73 (d, *J* = 8.3 Hz, 1H), 7.63 (d, *J* = 8.0 Hz, 2H), 7.46 (t, *J* = 7.6 Hz, 2H), 7.43 (t, *J* = 7.6 Hz, 1H), 7.39 (t, *J* = 7.4 Hz, 1H), 7.33 (t, *J* = 7.5 Hz, 1H), 7.28 (s, 1H), 7.24 (s, 1H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 153.24, 139.21, 134.64, 132.13, 130.74, 130.48, 129.72, 129.10, 128.65, 128.22, 127.28, 126.65, 124.49, 110.24 ppm; **HRMS** (ESI) calculated for $C_{16}H_{12}OH^+$ = 219.0815, found = 219.0810 *m/z*.



3-(4-methoxyphenyl)naphthalen-2-ol (4.32). A round bottom flask equipped with a stir bar was charged with tripotassium phosphate (952 mg, 2 equiv, 4.48 mmol), 3-bromo-2-naphthol (500 mg, 1.00 equiv, 2.24 mmol), (4-methoxyphenyl)boronic acid (511 mg, 1.5 equiv, 3.36 mmol), and XPhos Pd G3 (37.9 mg, 0.02 equiv, 44.8 µmol), then evacuated and refilled with nitrogen (3x). To this reaction flask, water (10 mL), which had been degassed by sparging with nitrogen for 30 minutes, and THF (5 mL) were added and the reaction was heated to 40 °C overnight. Then, the reaction flask was cooled to room temperature, the contents diluted with diethyl ether and brine, and transferred to a separatory funnel, where the aqueous layer was extracted with diethyl ether (3x). The combined organic layers were dried over magnesium sulfate, filtered, and concentrated under vacuum. The crude reaction residue was purified by silica flash chromatography (20% ethyl acetate in hexanes) to afford the title compound as an off-white solid in (174 mg, 28% yield). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.2; **^{1}H NMR** (600 MHz, CD_3CN): δ 7.79 (d, *J* = 8.2 Hz, 0H), 7.77 (s, 1H), 7.71 (d, *J* = 7.5 Hz, 0H), 7.59–7.54 (m, 2H), 7.41 (ddd, *J* = 8.2, 6.8, 1.3 Hz, 1H), 7.32 (ddd, *J* = 8.1, 6.8, 1.2 Hz, 1H), 7.26 (s, 1H), 7.19 (s, 1H), 7.04–6.99 (m, 2H), 3.84 (s, 3H) ppm; **^{13}C NMR** (150 MHz, CD_3CN): δ 160.09, 153.36, 134.90, 131.76, 131.64, 131.37, 130.33, 129.80, 128.53, 127.06, 126.61, 124.42, 114.52, 110.56, 55.89 ppm; **HRMS** (ESI) calculated for $C_{17}H_{13}O_2^-$ = 249.0921, found = 249.0916 *m/z*.

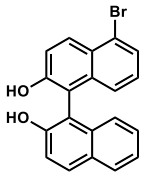


2-methoxy-3-(*p*-tolyl)naphthalene (4.33). A round bottom flask equipped with a stir bar was charged with tripotassium phosphate (895 mg, 2 equiv, 4.22 mmol), 2-bromo-3-methoxynaphthalene (500 mg, 1 equiv, 2.11 mmol), *p*-tolylboronic acid (386 mg, 1.5 equiv, 3.16 mmol) and XPhos Pd G3 (53.6 mg, 0.03 equiv, 63.3 µmol), then evacuated and refilled with nitrogen (3x). To this reaction flask, water (10 mL), which had been degassed by sparging with nitrogen for 30 min, and THF (5 mL) were added and the reaction was heated to 40 °C overnight. Then, the reaction flask was cooled to room temperature, the contents diluted with diethyl ether and brine, and transferred to a separatory funnel, where the aqueous layer was extracted with diethyl ether (3x). The combined organic layers were dried over magnesium sulfate, filtered, and concentrated under vacuum. The crude reaction residue was purified by silica flash chromatography (10% CH_2Cl_2 in hexanes) to afford the title compound as a white solid (350 mg, 66% yield). TLC (20% CH_2Cl_2 in hexanes) $\mathbf{R_f}$ = 0.34; **^{1}H NMR** (600 MHz, $CDCl_3$): δ 7.78 (d, *J* = 9.2 Hz, 1H), 7.77 (d, *J* = 8.2 Hz, 1H), 7.74 (s, 1H), 7.51 (d, *J* = 8.0 Hz, 2H), 7.44 (ddd, *J* = 8.1, 6.8, 1.3 Hz, 1H), 7.35 (ddd, *J* = 8.1, 6.8, 1.2 Hz, 1H), 7.26 (d, *J* = 8.1 Hz, 2H), 7.22 (s, 1H), 3.93 (s, 3H), 2.42 (s, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$): δ 155.51, 137.12, 135.55, 134.04, 132.53, 129.94, 129.76, 129.00, 128.92, 127.79, 126.48, 126.36, 124.02, 105.76, 55.68, 21.40 ppm.

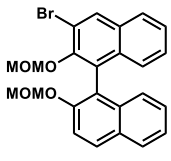


3-(*p*-tolyl)naphthalen-2-ol (4.34). A flame dried round bottom flask equipped with a stir bar was charged with 2-methoxy-3-(*p*-tolyl)naphthalene (**4.33**, 350 mg, 1 equiv, 1.41 mmol) in dry CH_2Cl_2 (5.69 mL). The solution was cooled to –78 °C and a solution of tribromoborane in CH_2Cl_2 (1.06 g, 4.23 mL of 1 M solution, 3 equiv, 4.23 mmol) was added dropwise. The reaction was allowed to warm to room temperature and monitored for full conversion to the product by TLC. Then the reaction was cooled to 0 °C and quenched by careful addition of water. The reaction was then diluted with DCM and water and transferred to a separatory funnel, where the aqueous layer was extracted with 20 mL CH_2Cl_2 (3x). The combined organic layers were dried with magnesium sulfate, filtered and concentrated under vacuum. The crude residue was purified by silica column

chromatography (20% CH_2Cl_2 in hexanes) to afford the title compound as a white solid (225 mg, 68% yield). TLC (20% CH_2Cl_2 in hexanes) $\mathbf{R_f}$ = 0.07; **^{1}H NMR** (600 MHz, CD_3CN): δ 7.80 (d, *J* = 8.2 Hz, 1H), 7.77 (s, 1H), 7.71 (d, *J* = 8.3 Hz, 1H), 7.51 (dd, *J* = 8.1, 2.0 Hz, 2H), 7.42 (ddt, *J* = 8.5, 6.8, 1.6 Hz, 1H), 7.32 (ddt, *J* = 8.5, 6.8, 1.6 Hz, 1H), 7.29–7.24 (m, 3H), 7.17 (s, 1H), 2.39 (s, 3H) ppm; **^{13}C NMR** (150 MHz, CD_3CN): δ 153.32, 138.08, 136.19, 135.02, 132.10, 130.53, 130.36, 129.75, 128.58, 127.16, 126.63, 124.44, 110.59, 21.17 ppm; **HRMS** (ESI) calculated for $C_{17}H_{13}O^-$ = 233.0972, found = 233.0972 *m/z*.

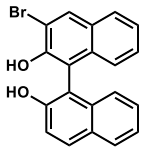


5-bromo-[1,1'-binaphthalene]-2,2'-diol (4.22). To a solution of 5-bromo-2-naphthol (100 mg, 0.45 mmol) and 2-naphthol (**4.9**, 194 mg, 1.35 mmol) in 45 mL of CH_2Cl_2 was added Cu(OH)Cl-TMEDA (10.4 mg, 5.00 mol %). The resultant mixture was stirred at room temperature in open air for 72 h. Afterwards, the reaction mixture was analyzed by TLC and LC–MS. The mixture was concentrated and purified by flash chromatography over silica gel (gradient of 5–50% ethyl acetate in hexanes) to afford the partially pure title compound (85 mg, 52% yield). The partially pure compound was resubjected to purification by flash chromatography over silica gel (gradient of 5–50% ethyl acetate in hexanes) to afford the pure title compound (45 mg, 28% yield). TLC (25% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, DMSO-d_6) δ 9.42 (s, 1H), 9.29 (s, 1H), 8.12 (d, J = 2.2 Hz, 1H), 7.86 (td, J = 8.3, 7.9, 1.1 Hz, 3H), 7.37 (d, J = 8.9 Hz, 1H), 7.34–7.29 (m, 2H), 7.25 (ddd, J = 8.1, 6.7, 1.3 Hz, 1H), 7.19 (ddd, J = 8.2, 6.7, 1.4 Hz, 1H), 6.93–6.90 (m, 1H), 6.88 (dt, J = 9.0, 0.6 Hz, 1H); **^{13}C NMR** (150 MHz, DMSO-d_6) δ 154.0, 153.5, 134.4, 133.2, 130.0, 129.8, 129.4, 129.2, 128.6, 128.4, 128.4, 127.2, 126.5, 124.7, 122.8, 120.20, 119.0, 116.2, 115.6, 115.2; **HRMS** (ESI) calculated for $C_{20}H_{12}BrO_2^-$ = 363.0026, found = 363.0065 *m/z*.



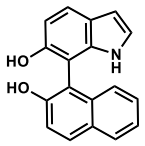
3-bromo-2,2'-bis(methoxymethoxy)-1,1'-binaphthalene (4.35). A flame dried round bottom flask equipped with a stir bar was charged with rac-2,2'-bis(methoxymethoxy)-1,1'-binaphthalene

(500 mg, 1 equiv, 1.34 mmol) and the flask was evacuated and refilled with nitrogen (3x). Then dry THF (29 mL) was added and the solution was cooled to –78 °C, followed by slow addition of n-BuLi (94.1 mg, 588 µL of 2.5 M solution in hexanes, 1.1 equiv, 1.47 mmol). The reaction was allowed to stir for 10 minutes, then warmed to 0 °C for an hour. Then the reaction was cooled back to –78 °C, and a solution of THF (3 mL) containing 1,2-dibromo-1,1,2,2-tetrachloro-ethane (652 mg, 1.5 equiv, 2.00 mmol) was added. The reaction was allowed to warm to room temperature overnight and was then quenched by addition of saturated aqueous ammonium chloride. The reaction was then diluted with diethyl ether and transferred to a separatory funnel where the aqueous layer was extracted with diethyl ether (3x). The combined organic layers were then washed with brine, dried over magnesium sulfate, filtered and concentrated. The crude product residue was then purified by silica column chromatography (10% ethyl acetate in hexanes) to afford the title compound as a sticky white solid (230 mg, 38% yield. TLC (10% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.24; **^{1}H NMR** (600 MHz, $CDCl_3$): δ 8.25 (s, 1H), 7.97 (d, *J* = 9.1 Hz, 1H), 7.86 (d, *J* = 8.2 Hz, 1H), 7.81 (d, *J* = 8.2 Hz, 1H), 7.59 (d, *J* = 9.1 Hz, 1H), 7.41 (t, *J* = 7.5 Hz, 1H), 7.36 (t, *J* = 7.4 Hz, 1H), 7.30–7.22 (m, 2H), 7.18 (d, *J* = 8.5 Hz, 1H), 7.15 (d, *J* = 8.5 Hz, 1H), 5.14 (d, *J* = 7.0 Hz, 1H), 5.04 (d, *J* = 6.9 Hz, 1H), 4.78 (d, *J* = 5.5 Hz, 1H), 4.74 (d, *J* = 5.5 Hz, 1H), 3.19 (s, 3H), 2.70 (s, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$): δ 153.06, 149.75, 134.01, 133.25, 132.54, 131.85, 130.23, 129.68, 127.99, 127.92, 127.12, 126.92, 126.69, 126.31, 125.97, 125.61, 124.33, 120.17, 117.85, 116.45, 99.12, 95.01, 56.70, 56.11 ppm.

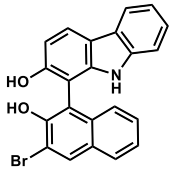


(*rac*)-3-bromo-[1,1'-binaphthalene]-2,2'-diol (4.23). A 100-mL round bottom flask was charged with 3-bromo-2,2'-bis(methoxymethoxy)-1,1'-binaphthalene (**4.35**, 50.0 mg, 1 equiv, 110 µmol), dioxane (1 mL), and 37% aqueous HCl (402 mg, 270 µL, 100 equiv, 11.0 mmol), and the reaction mixture was stirred at room temperature overnight. The mixture was partitioned between diethyl ether (100 mL) and a saturated aqueous solution of $NaHCO_3$. The organic phase was washed with brine, dried over magnesium sulfate, filtered, and concentrated in vacuo. Purification on silica gel column chromatography (10% ethyl acetate in hexanes) afforded the title compound as a foamy beige solid (40 mg, 99% yield). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.27; **^{1}H NMR** (600

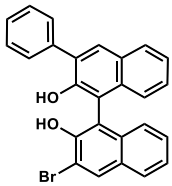
MHz, $CDCl_3$): δ 8.28 (s, 1H), 7.98 (d, *J* = 8.9 Hz, 1H), 7.90 (d, *J* = 8.1 Hz, 1H), 7.83 (d, *J* = 8.2 Hz, 1H), 7.44–7.35 (m, 3H), 7.34–7.28 (m, 2H), 7.15 (d, *J* = 8.5 Hz, 1H), 7.10 (d, *J* = 8.4 Hz, 1H), 5.56 (s, 1H), 4.91 (s, 1H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$): δ 152.34, 149.03, 133.49, 133.28, 133.03, 131.53, 130.07, 129.52, 128.57, 127.98, 127.61, 127.56, 125.21, 124.77, 124.30, 124.15, 117.90, 113.35, 112.24, 111.90; **HRMS** (ESI) calculated for $C_{20}H_{12}BrO_2^-$ = 363.0026, found = 363.0031 *m/z*.



7-(2-hydroxynaphthalen-1-yl)-1H-indol-6-ol (4.24). To a solution of 6-hydroxyindole (133 mg, 1.00 mmol, 1 equiv) and 2-naphthol (**4.9**, 144 mg, 1.00 mmol, 1.00 equiv) in 50 mL of Ch_2Cl_2 was added Cu(OH)Cl-TMEDA (23 mg, 5.0 mol %). The resultant mixture was stirred at room temperature in open air for 72 h. Afterwards, the reaction mixture was analyzed by TLC and LCMS. The reaction mixture was concentrated and purified by flash chromatography over silica gel (gradient of 10–60% ethyl acetate in hexanes) to afford the partially pure title compound (12 mg, 4.4% yield). The compound was redissolved in methanol, filtered through a 0.22 µm filter and purified by preparative HPLC using a Phenomenex Kinetex 5 µm C18, 150 x 21.2 mm column using the following conditions. Mobile phase A = deionized water + 0.1% formic acid. Mobile Phase B = acetonitrile + 0.1% formic acid. Method: 5% to 100% B over 15 minutes, 100% B over 1 minute; Flow rate = 10 mL/min. The desired product eluted at 12 min. Preparative HPLC purification afforded clean title compound as a light brown solid (2.3 mg, 0.83% yield overall). TLC (30% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, DMSO-d_6) δ 9.83 (s, 1H), 9.19 (s, 1H), 8.42 (s, 1H), 7.82 (dd, *J* = 8.0, 5.2 Hz, 2H), 7.37 (d, *J* = 8.4 Hz, 1H), 7.30 (d, *J* = 8.8 Hz, 1H), 7.26–7.19 (m, 2H), 7.12 (dd, *J* = 7.9, 1.8 Hz, 1H), 6.90 (t, *J* = 2.8 Hz, 1H), 6.75 (d, *J* = 8.4 Hz, 1H), 6.32 (t, *J* = 2.5 Hz, 1H). **^{13}C NMR** (150 MHz, DMSO-d_6) δ 153.6 150.8, 137.1, 134.4, 128.9, 128.7, 128.2, 126.0, 125.3, 123.8, 122.6, 121.5, 119.8, 119.2, 114.8, 110.2, 105.6, 101.3. **HRMS** (ESI) calculated for $C_{18}H_{12}NO_2^-$ = 274.0874, found = 274.0865 *m/z*.

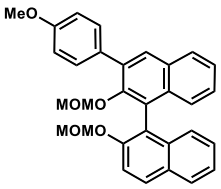


1-(3-bromo-2-hydroxynaphthalen-1-yl)-9H-carbazol-2-ol (4.25). To a solution of 2-hydroxycarbazole (366 mg, 2.00 mmol) and 3-bromo-2-naphthol (**4.17**, 446 mg, 2.00 mmol) in 100 mL of CH_2Cl_2 was added Cu(OH)Cl-TMEDA (46 mg, 5.0 mol %). The resultant mixture was stirred at room temperature in open air for 72 h. Afterwards, the reaction mixture was analyzed TLC and LC–MS. The mixture was concentrated and purified by flash chromatography over silica gel (gradient of 5–40% ethyl acetate in hexanes) to afford the partially pure title compound (20 mg, 2.5% yield). The partially pure compound was resubjected to purification by flash chromatography over silica gel (gradient of 5–25% ethyl acetate in hexanes) to afford the pure title compound (8 mg, 1%). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.5; **^{1}H NMR** (600 MHz, DMSO-d_6) δ 10.21 (s, 1H), 9.22 (s, 1H), 8.86 (s, 1H), 8.32 (s, 1H), 7.99 (d, J = 8.2 Hz, 2H), 7.89 (dd, J = 8.3, 1.3 Hz, 1H), 7.32 (ddd, J = 8.1, 6.7, 1.3 Hz, 1H), 7.28–7.25 (m, 1H), 7.25–7.23 (m, 1H), 7.18 (ddd, J = 8.1, 7.0, 1.2 Hz, 1H), 7.10 – 7.04 (m, 2H), 6.90 (d, J = 8.4 Hz, 1H); **^{13}C NMR** (150 MHz, CDCl3) δ 152.7, 148.8, 139.6, 139.4, 133.5, 132.2, 130.0, 128.0, 127.6, 125.2, 124.9, 124.6, 123.7, 122.4, 119.9, 119.6, 117.6, 112.3, 111.9, 110.6, 109.2, 101.6; **HRMS** (ESI) calculated for $C_{22}H_{13}BrNO_2^-$ = 402.0135, found = 402.0129 *m/z*.

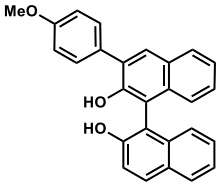


3-bromo-3'-phenyl-[1,1'-binaphthalene]-2,2'-diol (4.26). To a solution of 3-bromo-2-naphthol (**4.17**, 50.0 mg, 0.225 mmol) and 3-phenylnaphthalen-2-ol (**4.31**, 50.0 mg, 0.225 mmol) in 45 mL of CH_2Cl_2 was added Cu(OH)Cl-TMEDA (5.2 mg, 5.0 mol %). The resultant mixture was stirred at room temperature in open air for 72 h. Afterwards, the reaction mixture was analyzed TLC and LC–MS. The mixture was sent to Spencer McMinn at Merck for purification to afford the title compound (30 mg, 30%). **^{1}H NMR** (599 MHz, DMSO-d_6) δ 9.05 (s, 1H), 8.36 (s, 1H), 7.96 (s, 1H), 7.94 (d, *J* = 8.2 Hz, 1H), 7.90 (d, *J* = 8.0 Hz, 1H), 7.71–7.69 (m, 2H), 7.47 (t, *J* = 7.7 Hz, 2H), 7.38 (t, *J* = 7.4 Hz, 1H), 7.33–7.28 (m, 2H), 7.25 (ddd, *J* = 8.2, 6.8, 1.2 Hz, 1H), 7.21 (ddd, *J* =

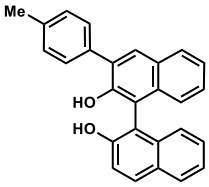
8.1, 6.8, 1.1 Hz, 1H), 6.93 (d, *J* = 8.5 Hz, 1H), 6.81 (d, *J* = 8.4 Hz, 1H); **^{13}C NMR** (151 MHz, DMSO-d_6) δ 151.16, 150.09, 138.77, 133.33, 133.19, 132.03, 131.71, 130.12, 129.58, 129.20, 128.52, 128.18, 127.99, 127.22, 127.01, 126.59, 126.28, 124.29, 123.72, 123.61, 123.02, 116.52, 114.78, 113.91; **HRMS** (ESI) calculated for $C_{26}H_{16}BrO_2^-$ = 439.0339, found = 439.0354 *m/z*.



2,2'-bis(methoxymethoxy)-3-(4-methoxyphenyl)-1,1'-binaphthalene (4.36). A round bottom flask equipped with a stir bar was charged with potassium phosphate (46.8 mg, 2 equiv, 221 µmol), 3-bromo-2,2'-bis(methoxymethoxy)-1,1'-binaphthalene (**4.35**, 50.0 mg, 1 equiv, 110 µmol), (4-methoxyphenyl)boronic acid (33.5 mg, 2 equiv, 221 µmol), and XPhos Pd G3 (1.87 mg, 0.02 equiv, 2.21 µmol) then evacuated and refilled with nitrogen (3x). To this reaction flask, THF (263 µL) and water (525 µL) were added, and the reaction was heated to 40 °C for 2 h. Then, the reaction flask was cooled to room temperature, the contents diluted with diethyl ether and brine, and transferred to a separatory funnel, where the aqueous layer was extracted with diethyl ether (3x). The combined organic layers were dried over magnesium sulfate, filtered, and concentrated under vacuum. The crude reaction residue was purified by silica flash chromatography (5:20:75 ethyl acetate/CH_2Cl_2/hexanes) to afford the title compound as an off-white solid (42 mg, 79% yield). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.4; **^{1}H NMR** (600 MHz, $CDCl_3$): δ 7.95 (d, *J* = 9.1 Hz, 1H), 7.92 (s, 1H), 7.89 (d, *J* = 8.2 Hz, 1H), 7.86 (d, *J* = 8.2 Hz, 1H), 7.66 (d, *J* = 8.6 Hz, 2H), 7.60 (d, *J* = 9.0 Hz, 1H), 7.39 (ddd, *J* = 7.9, 6.3, 1.3 Hz, 1H), 7.36 (dt, *J* = 8.1, 3.8 Hz, 1H), 7.28 (d, *J* = 3.6 Hz, 3H), 7.24 – 7.18 (m, 2H), 6.99 (d, *J* = 8.7 Hz, 2H), 5.17 (d, *J* = 6.9 Hz, 1H), 5.07 (d, *J* = 6.9 Hz, 1H), 4.35 (d, *J* = 5.8 Hz, 1H), 4.31 (d, *J* = 5.8 Hz, 1H), 3.86 (s, 3H), 3.22 (s, 3H), 2.30 (s, 3H) ppm; **^{13}C NMR** (151 MHz, $CDCl_3$): δ 159.10, 153.06, 151.25, 135.27, 134.28, 133.25, 131.60, 131.20, 130.89, 130.13, 129.89, 129.75, 128.02, 127.92, 126.61, 126.39, 126.14, 126.04, 126.02, 125.23, 124.24, 121.53, 117.00, 113.89, 98.67, 95.28, 56.10, 56.07, 55.45 ppm.

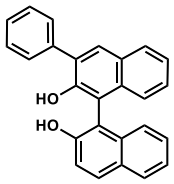


3-(4-methoxyphenyl)-[1,1'-binaphthalene]-2,2'-diol (4.28). A 100-mL round bottom flask was charged with 2,2'-bis(methoxymethoxy)-3-(4-methoxyphenyl)-1,1'-binaphthalene (**4.36**, 42 mg, 1 equiv, 87 µmol), dioxane (1 mL), and 37% aqueous HCl (0.32 g, 0.21 mL, 100 equiv, 8.7 mmol), and the reaction mixture was stirred at room temperature overnight. The mixture was partitioned between diethyl ether (100 mL) and a saturated aqueous solution of $NaHCO_3$. The organic phase was washed with brine, dried over magnesium sulfate, filtered, and concentrated in vacuo. Purification on silica gel column chromatography (20% ethyl acetate in hexanes) to afford the title compound as a foamy beige solid (31 mg, 90% yield). TLC (20% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.2; **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.02–7.95 (m, 2H), 7.90 (t, *J* = 7.1 Hz, 2H), 7.67 (d, *J* = 8.6 Hz, 2H), 7.43–7.35 (m, 3H), 7.34 – 7.27 (m, 2H), 7.23 (d, *J* = 8.4 Hz, 1H), 7.14 (d, *J* = 8.4 Hz, 1H), 7.03 (d, *J* = 8.7 Hz, 2H), 5.29 (s, 0H), 5.11 (s, 1H), 3.87 (s, 3H) ppm; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 159.53, 152.78, 150.50, 133.57, 132.87, 131.46, 131.23, 130.90, 130.47, 129.80, 129.67, 129.61, 128.55, 128.48, 127.54, 127.36, 124.49, 124.31, 124.12, 117.88, 114.15, 111.77, 111.72, 55.52 ppm; **HRMS** (ESI) calculated for $C_{27}H_{19}O_3^-$ = 391.1340, found = 391.1331 *m/z*.



3-(*p*-tolyl)-[1,1'-binaphthalene]-2,2'-diol (4.29). A round bottom flask was charged with 2-naphthol (**4.9**, 50 mg, 1 equiv, 0.35 mmol) and 3-(*p*-tolyl)naphthalen-2-ol (**4.34**, 81.3 mg, 1 equiv, 347 µmol) and di-mu-hydroxo-bis[(N,N,N'.N'-tetramethylethylenediamine)copper(II)] chloride (5.7 mg, 0.02 equiv, 6.9 µmol). CH_2Cl_2 (2.5 mL) was then added and the reaction was allowed to stir overnight at room temperature. The reaction was then filtered through a silica plug, eluting with ethyl acetate. The resulting solution was concentrated under vacuum and the crude residue was purified by silica gel column chromatography (gradient of 0–10% ethyl acetate in hexanes) to afford the title compound as a yellow foamy solid (51 mg, 78% yield). TLC (10% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.13; **^{1}H NMR** (599 MHz, $CDCl_3$) δ 8.00 (s, 1H), 7.98 (d, *J* = 9.0 Hz, 1H), 7.91 (t,

J = 8.7 Hz, 2H), 7.62 (d, *J* = 7.9 Hz, 2H), 7.42–7.35 (m, 3H), 7.33–7.28 (m, 4H), 7.23 (d, *J* = 8.4 Hz, 1H), 7.15 (d, *J* = 8.5 Hz, 1H), 5.29 (s, 1H), 5.11 (s, 1H), 2.43 (s, 3H) ppm; **^{13}C NMR** (151 MHz, $CDCl_3$): δ 152.76, 150.48, 137.84, 134.51, 133.57, 133.00, 131.43, 131.41, 130.81, 129.63, 129.60, 129.58, 129.43, 128.54, 127.52, 127.44, 124.50, 124.49, 124.33, 124.10, 117.88, 111.79, 111.77, 21.40 ppm; **HRMS** (ESI) calculated for $C_{27}H_{19}O_2^-$ = 375.1391, found = 375.1394 *m/z*.



3-phenyl-[1,1'-binaphthalene]-2,2'-diol (4.30). A round bottom flask was charged with 2-naphthol (**4.31**, 50 mg, 1 equiv, 0.35 mmol) and 3-phenylnaphthalen-2-ol (76 mg, 1 equiv, 0.35 mmol) and di-mu-hydroxo-bis[(N,N,N'.N'-tetramethylethylenediamine)copper(II)] chloride (5.7 mg, 0.02 equiv, 6.9 µmol). CH_2Cl_2 (2.5 mL) was then added and the reaction was allowed to stir overnight at room temperature. The reaction was then filtered through a silica plug, eluting with ethyl acetate. The resulting solution was concentrated under vacuum and the crude residue was purified by silica gel column chromatography (gradient of 0–10% ethyl acetate in hexanes) to afford the title compound as a yellow foamy solid (50 mg, 40% yield). TLC (10% ethyl acetate in hexanes) $\mathbf{R_f}$ = 0.13; **^{1}H NMR** (600 MHz, $CDCl_3$): δ 8.03 (s, 1H), 7.99 (d, *J* = 8.9 Hz, 1H), 7.93 (d, *J* = 8.2 Hz, 1H), 7.90 (d, *J* = 8.1 Hz, 1H), 7.75–7.72 (m, 2H), 7.50 (t, *J* = 7.7 Hz, 2H), 7.44–7.35 (m, 4H), 7.32 (dddd, *J* = 8.2, 6.6, 4.8, 1.2 Hz, 2H), 7.24 (d, *J* = 8.4 Hz, 1H), 7.16 (d, *J* = 8.4 Hz, 1H), 5.29 (s, 1H), 5.12 (s, 1H); **^{13}C NMR** (150 MHz, $CDCl_3$): δ 152.81, 150.40, 137.54, 133.55, 133.08, 131.65, 131.51, 130.83, 129.74, 129.61, 128.65, 128.61, 128.56, 127.95, 127.57, 124.55, 124.45, 124.33, 124.14, 117.90, 111.89, 111.59 ppm; **HRMS** (ESI) calculated for $C_{26}H_{17}O_2^-$ = 361.1234, found = 361.1225 *m/z*.

Standard curves and LC–MS traces for biocatalytic reactions. High resolution electrospray mass spectra were obtained on an Agilent UPLC–QTOF at the University of Michigan Life Sciences Institute or Agilent UPLC–TOF at the University of Michigan Life Sciences Institute.

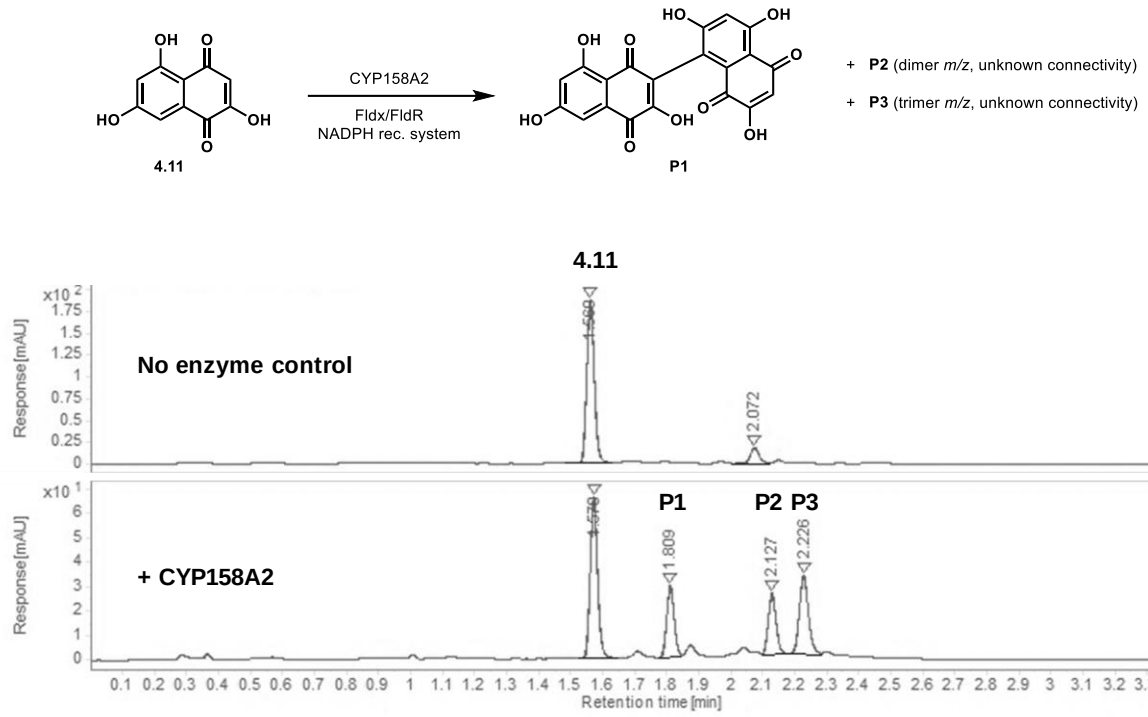
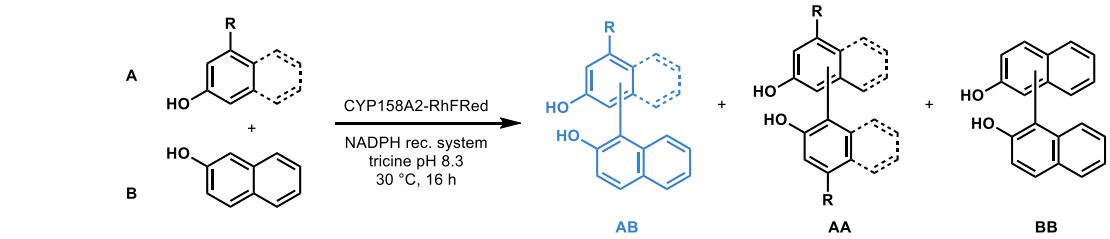
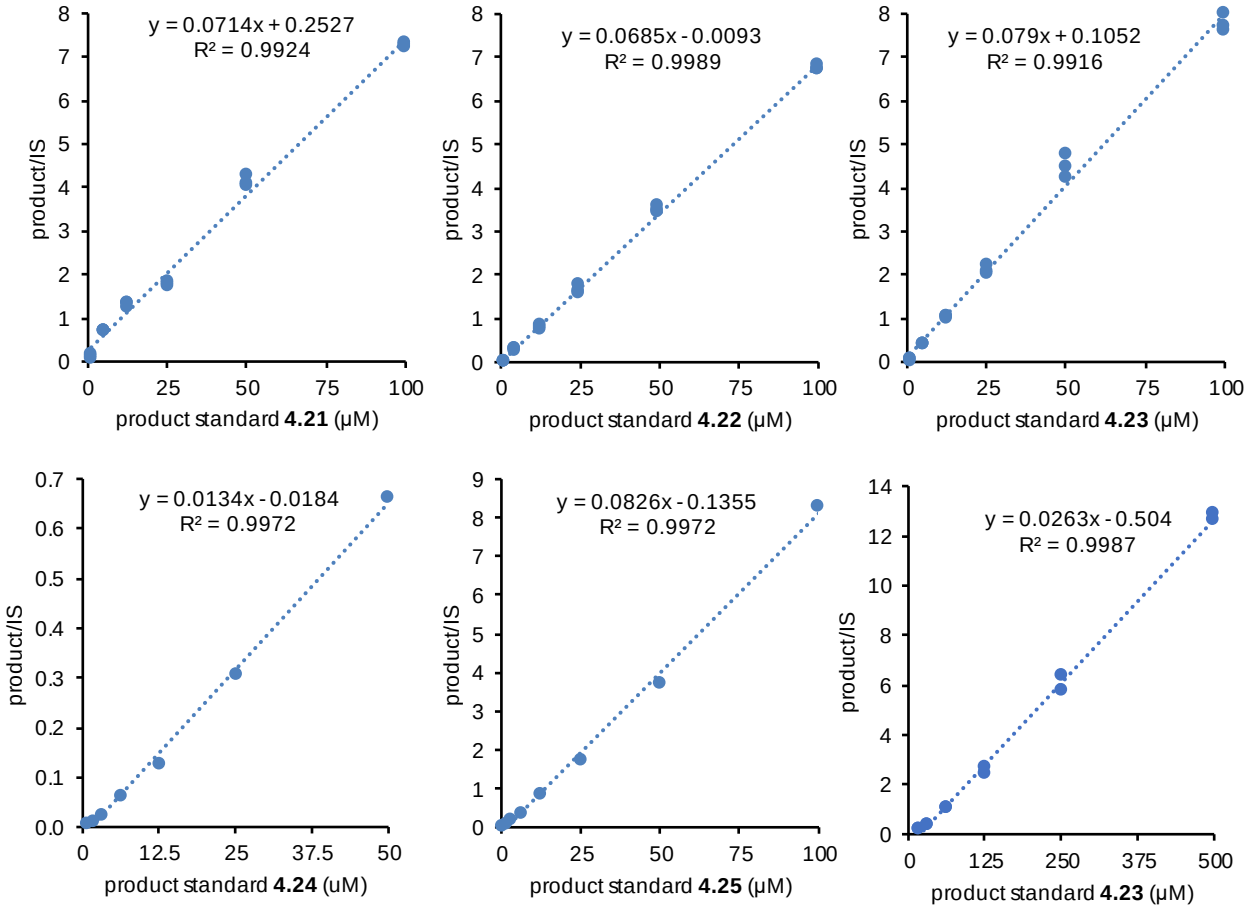


Figure 4.8. UPLC traces for oxidative coupling of flaviolin by CYP158A2. Absorbance was collected at 311 nm. Analysis of the same reaction by LC–MS shows that the first two product peaks (P1 and P2) give *m/z* corresponding to dimers, whereas the third peak (P3) gives a *m/z* corresponding to the trimer.





pdt	substrate A	substrate B	enzyme	pdt/IS	[pdt], μM	yield, %	avg yield, %
4.21	coumarin (100 μM)	2-naphthol (300 μM)	CYP158A2	0.61 0.71	6.6 7.9	6.6 7.9	7.3 ± 0.9
4.22	5-bromo-2-naphthol (100 μM)	2-naphthol (300 μM)	CYP158A2	0.85 0.87	12.6 12.9	12.6 12.9	12.8 ± 0.2
4.23	3-bromo-2-naphthol (100 μM)	2-naphthol (300 μM)	CYP158A2	0.35 0.34	3.1 2.9	3.1 2.9	3.0 ± 0.1
4.24	indole (100 μM)	2-naphthol (300 μM)	A0A1J4PSC8	0.23 0.22	18.1 18.3	18.1 18.3	18.2 ± 0.1
4.25	carbazole (100 μM)	3-bromo-2-naphthol (300 μM)	A0A6B3DVM9	4.92 4.03	61.2 50.5	61.2 50.5	55.9 ± 7.6
4.23	3-bromo-2-naphthol (500 μM)	2-naphthol (500 μM)	A2 variant	2.31 1.72	106.8 84.5	21.4 16.9	19.1 ± 3.1

Figure 4.9. Quantification of percent yields using product standard curves.

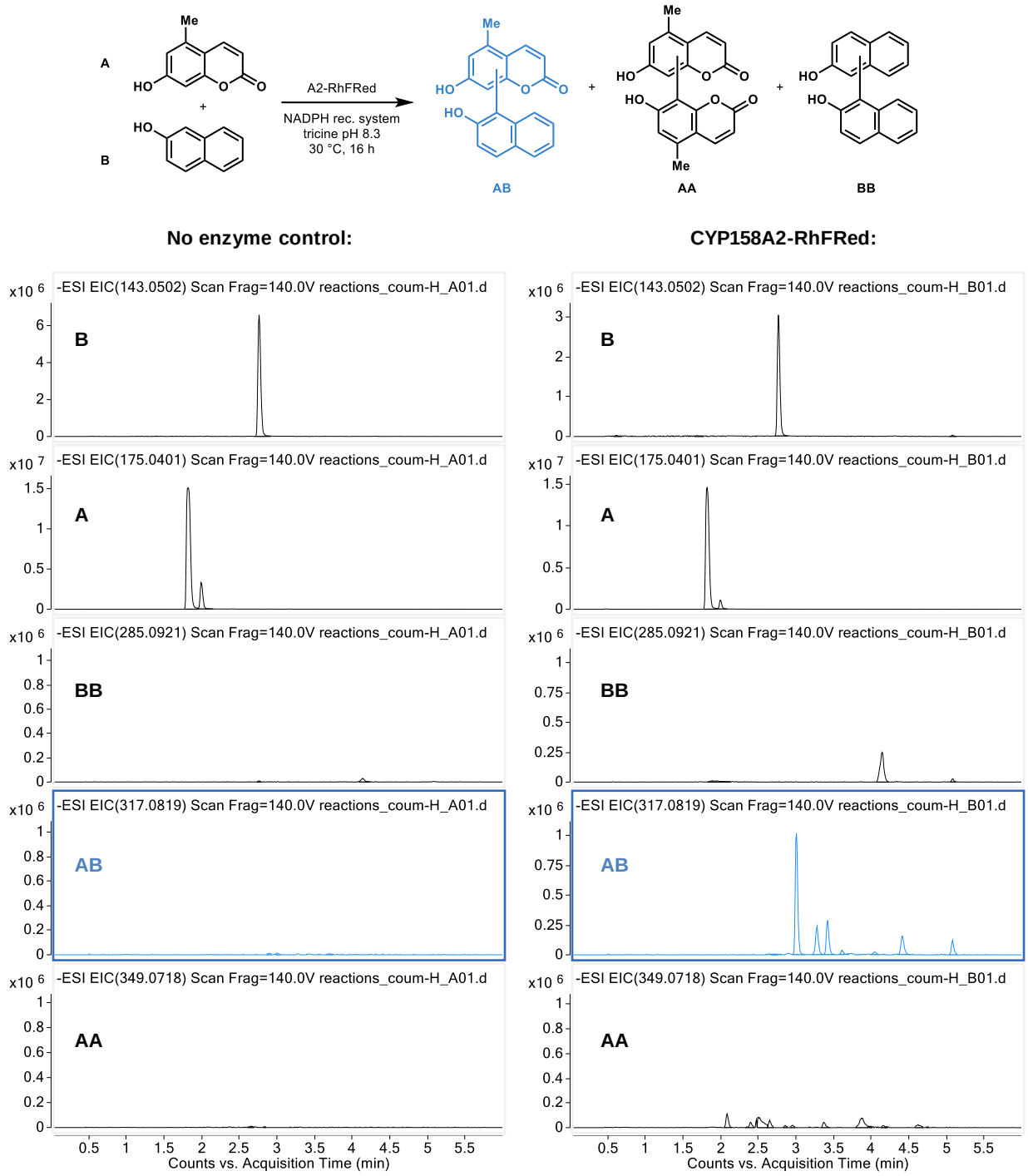


Figure 4.10. LC–MS traces for biocatalytic oxidative cross-coupling to form 4.21.

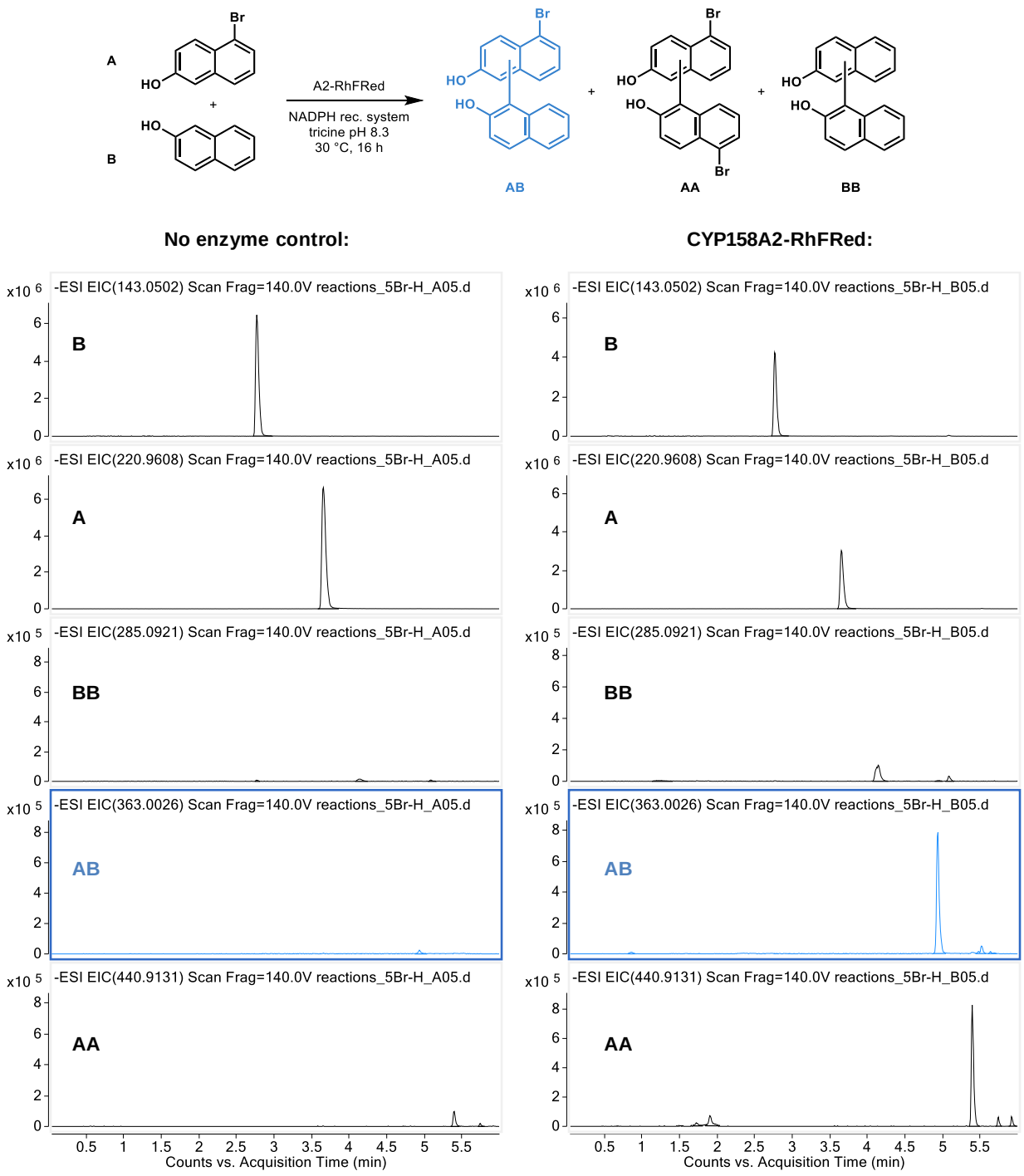


Figure 4.11. LC–MS traces for biocatalytic oxidative cross-coupling to form 4.22.

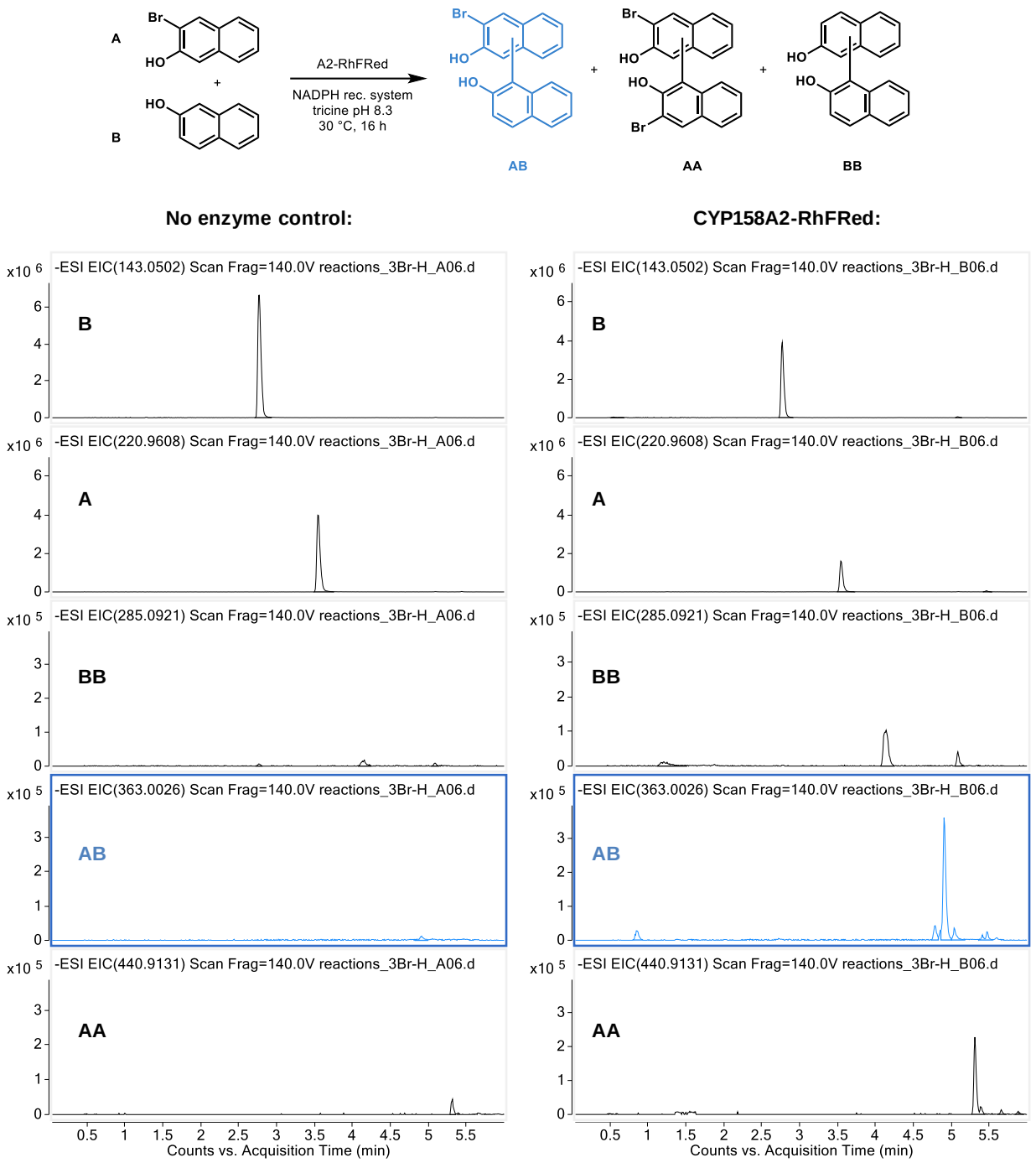


Figure 4.12. LC–MS traces for biocatalytic oxidative cross-coupling to form 4.23.

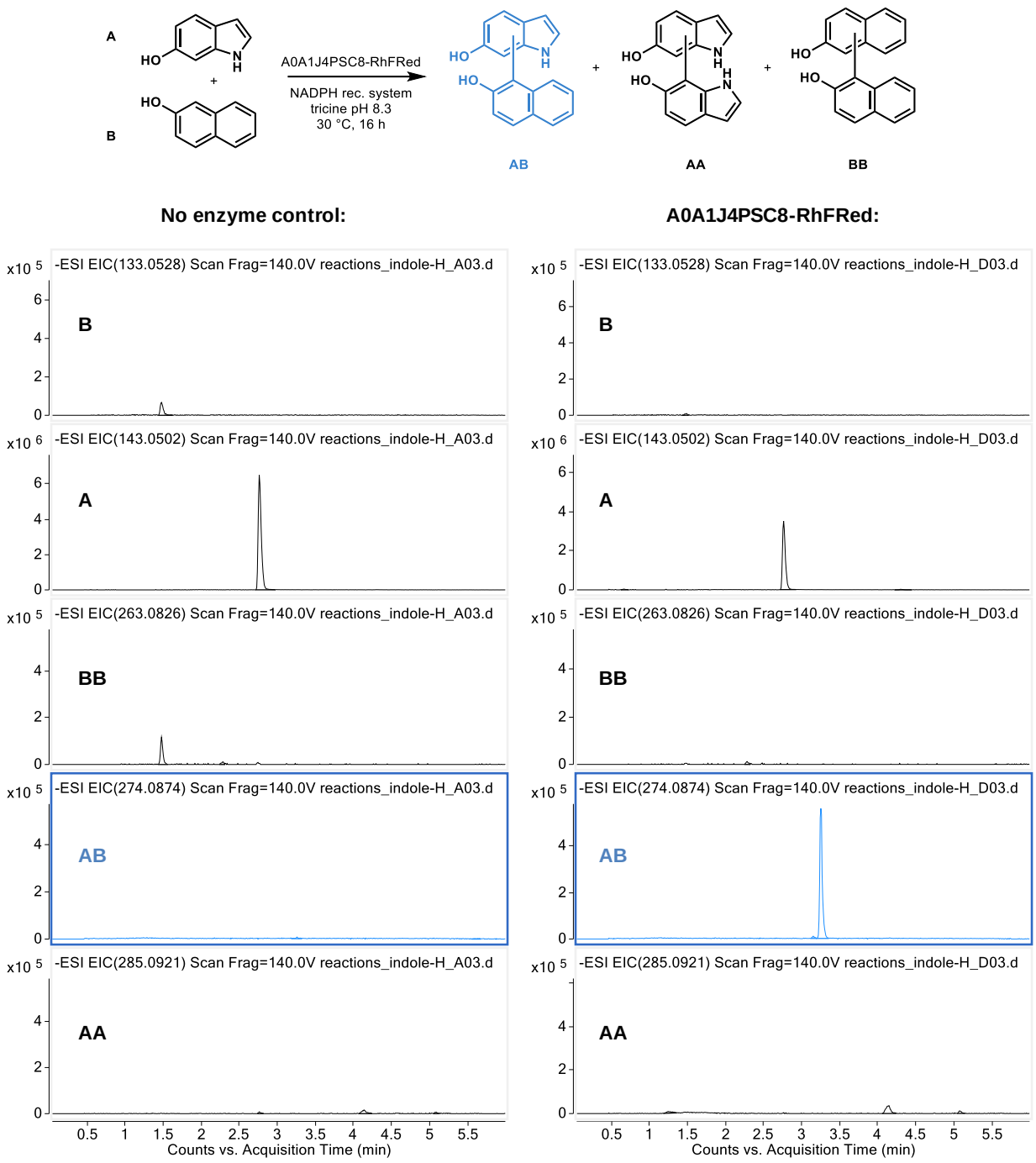


Figure 4.13. LC–MS traces for biocatalytic oxidative cross-coupling to form 4.24.

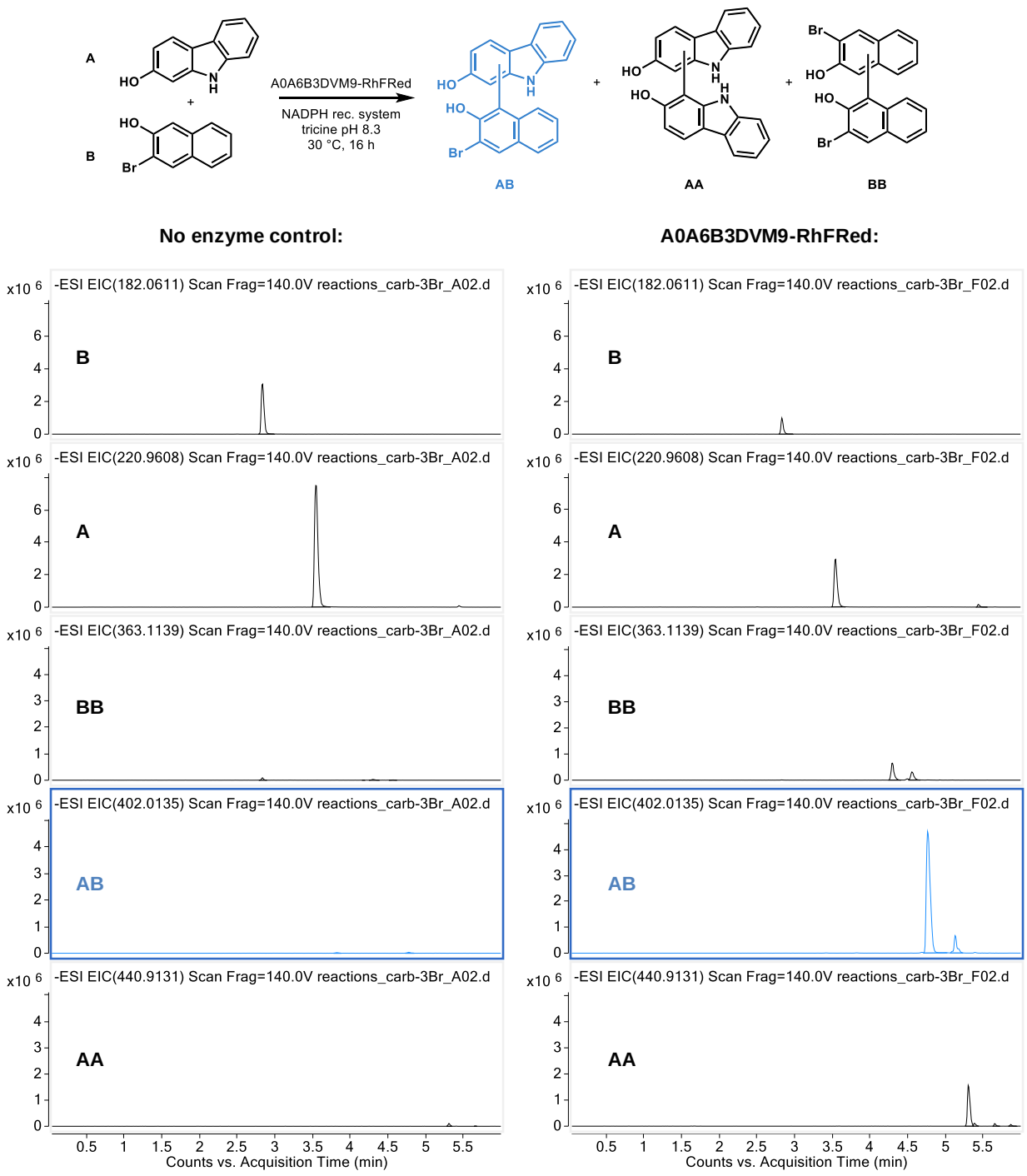


Figure 4.14. LC–MS traces for biocatalytic oxidative cross-coupling to form 4.25.

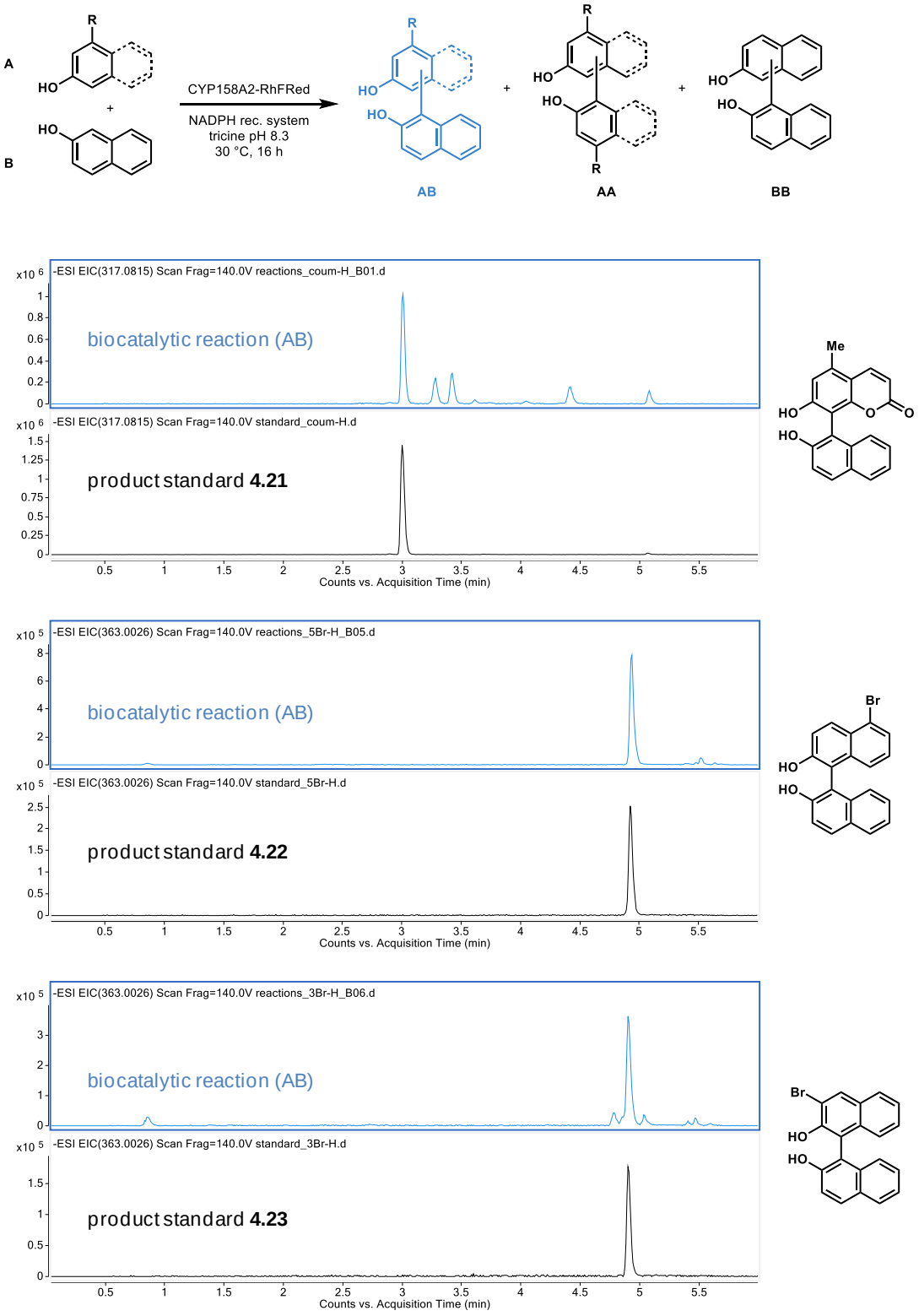


Figure 4.15. Site-selectivity of oxidative cross-couplings catalyzed by A2-RhFRed.

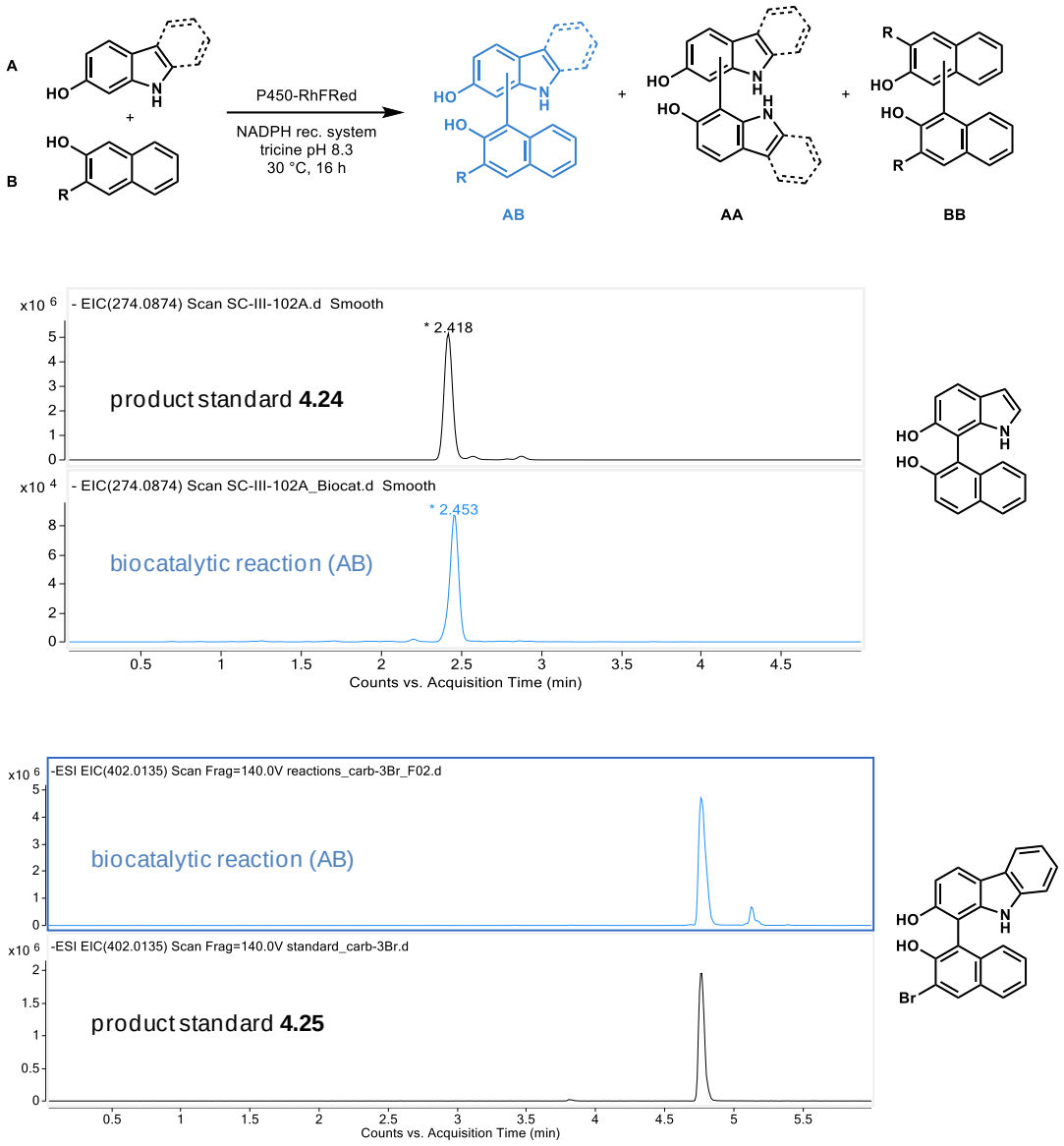


Figure 4.16. Site-selectivity of oxidative cross-coupling catalyzed by P450-RhFRed.

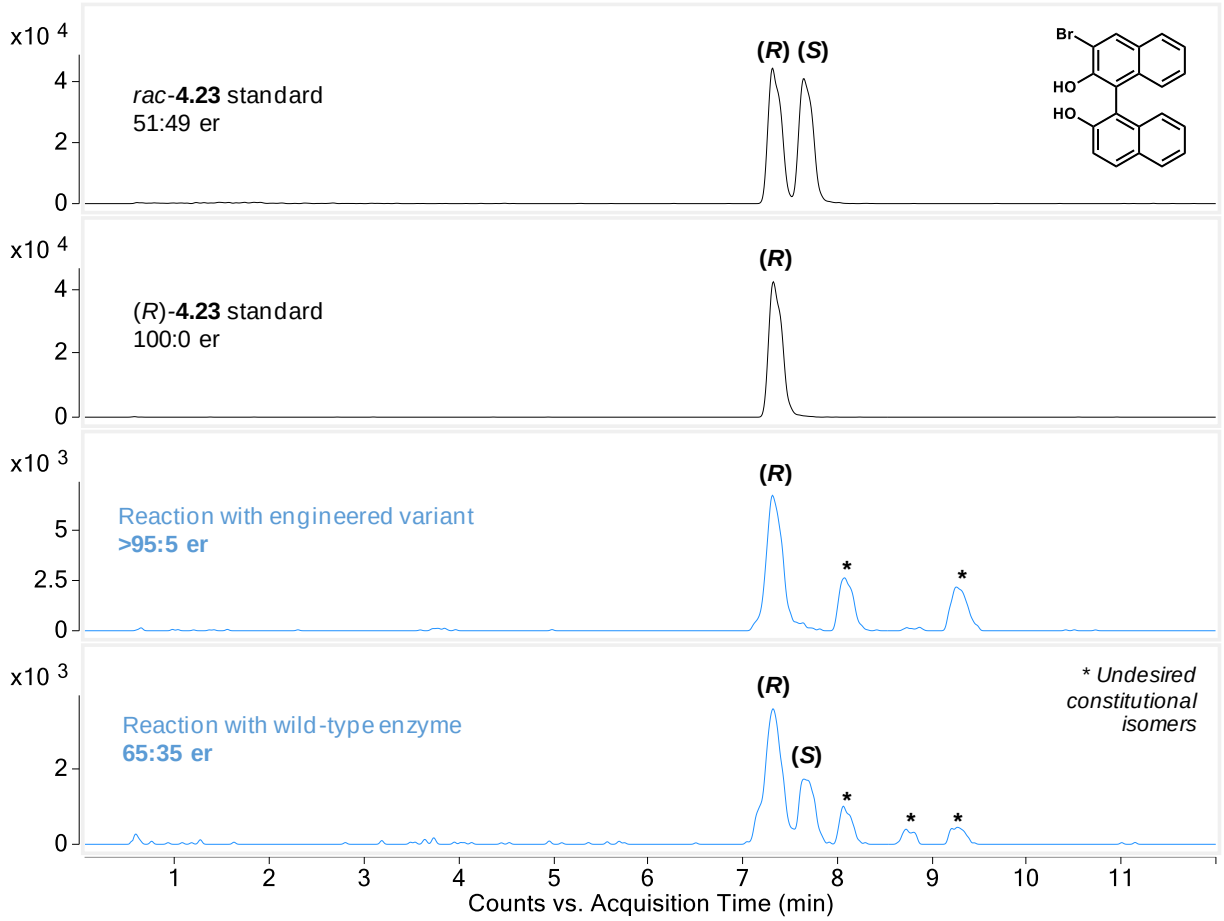


Figure 4.17. Atroposelectivity of oxidative cross-coupling to form 4.23.

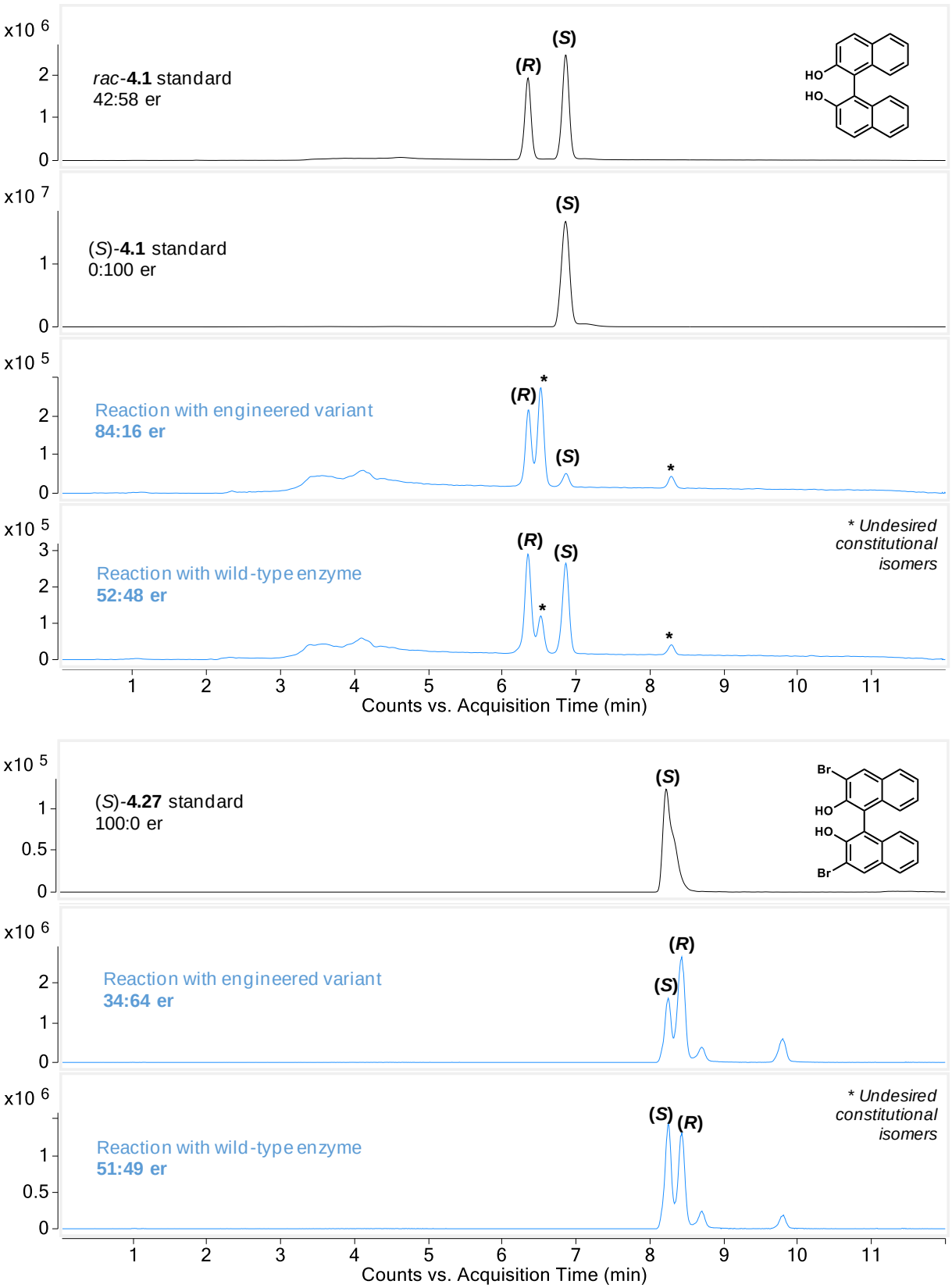


Figure 4.18. Atroposelectivity of oxidative cross-coupling to form 4.1 and 4.27.

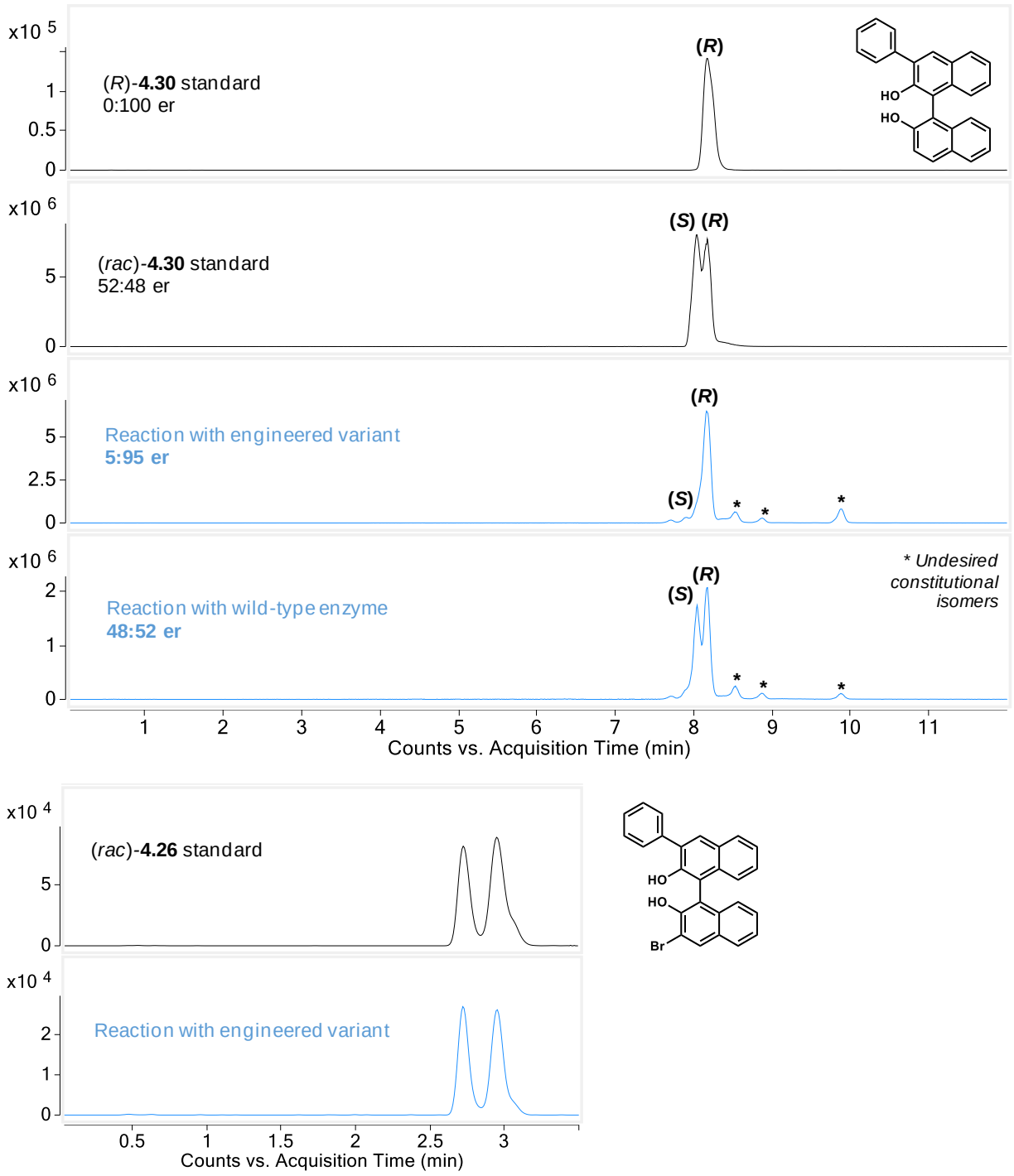


Figure 4.19. Atroposelectivity of oxidative cross-coupling to form 4.26 and 4.30.

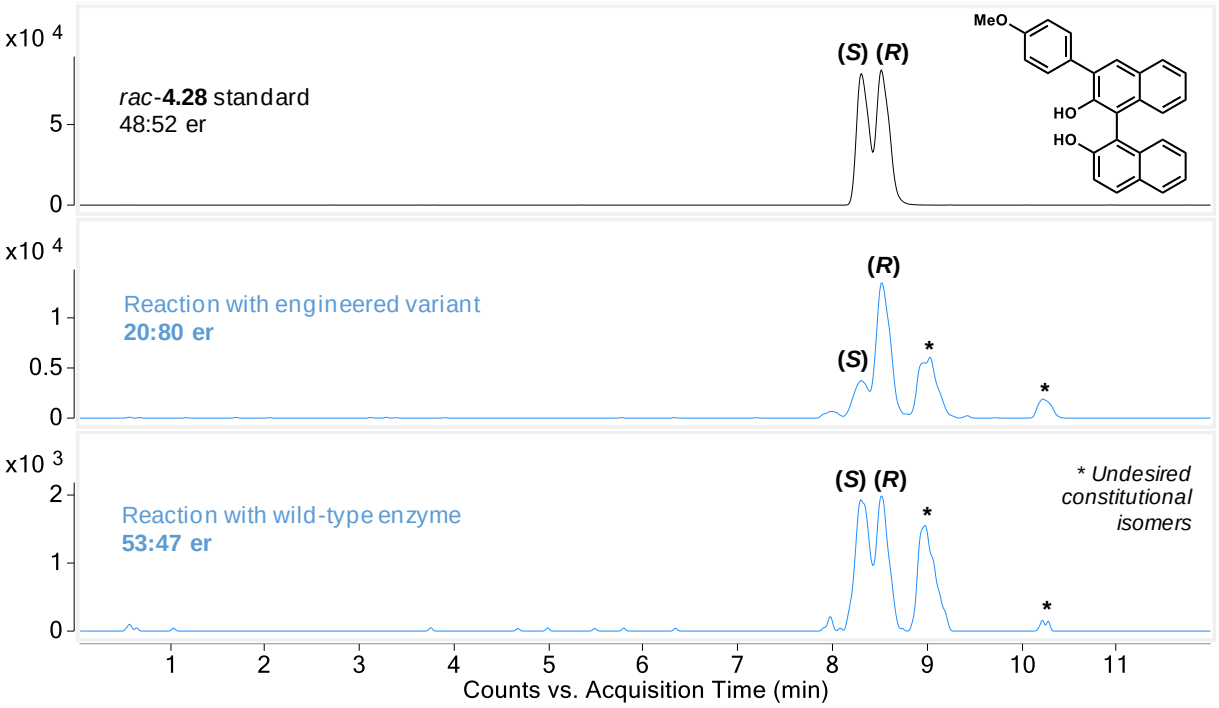


Figure 4.20. Atroposelectivity of oxidative cross-coupling to form 4.28. Absolute configuration assigned by analogy to (*R*)-**4.30**.

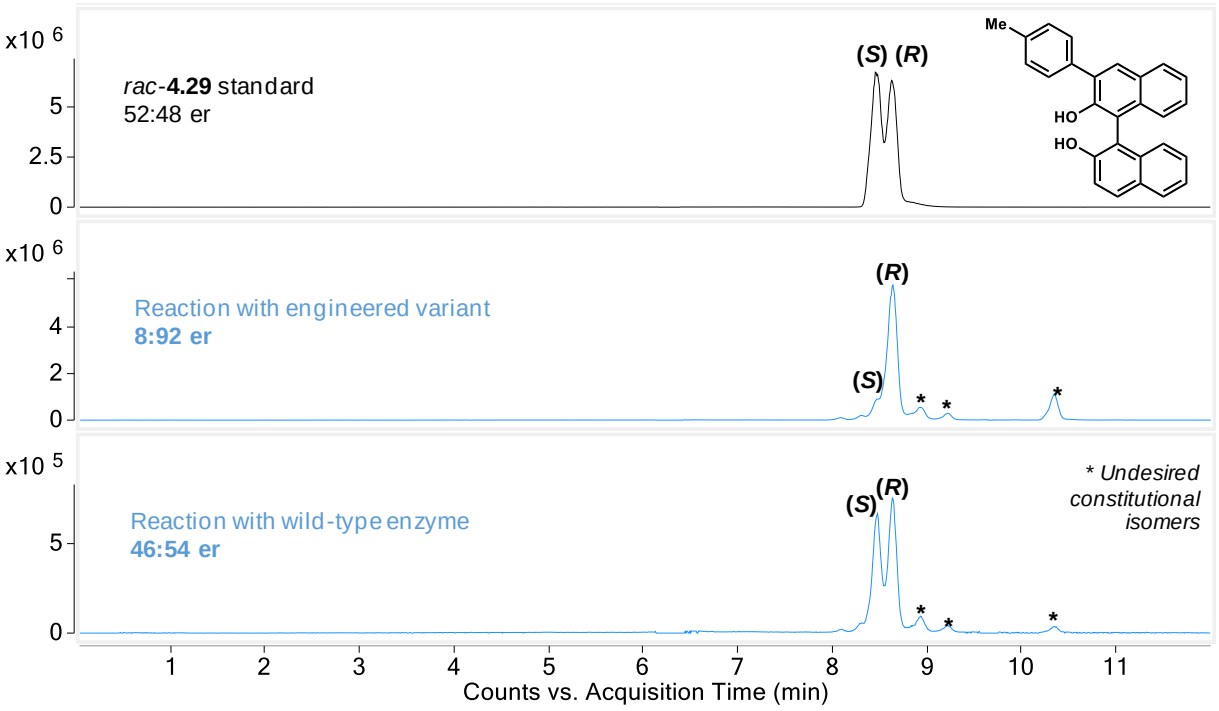
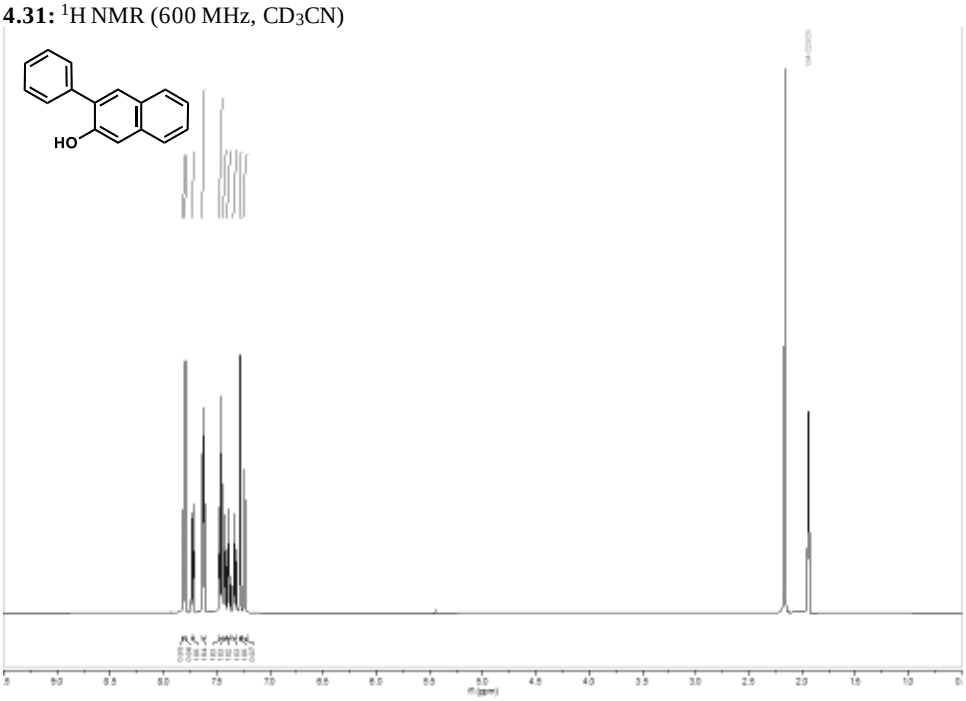


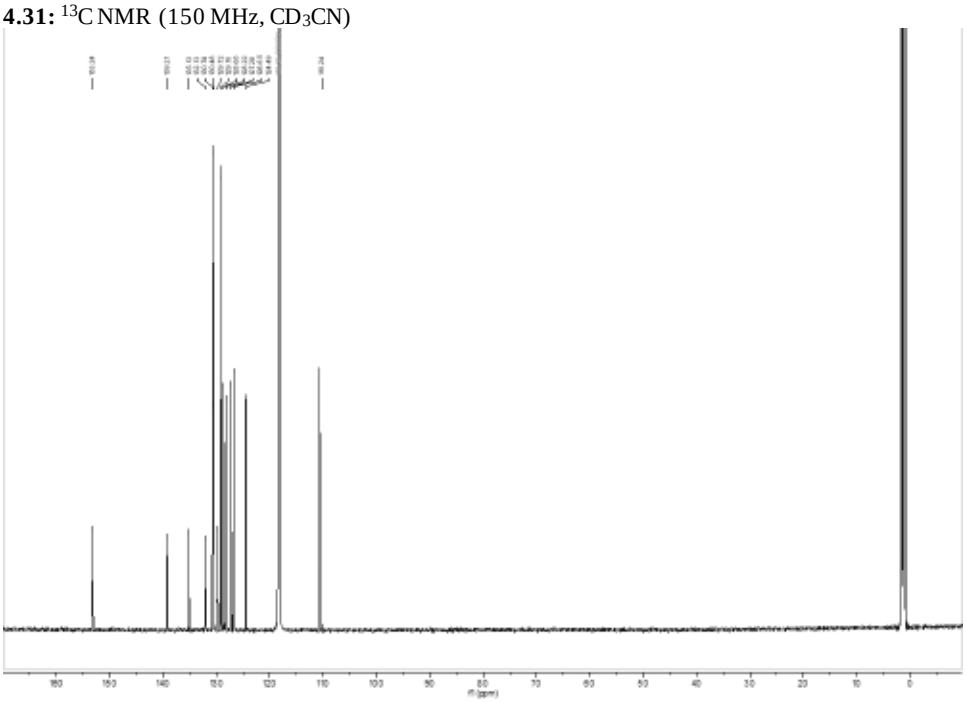
Figure 4.21. Atroposelectivity of oxidative cross-coupling to form 4.29. Absolute configuration assigned by analogy to (*R*)-**4.30**.

^{1}H NMR and ^{13}C NMR spectra of compounds. ^{1}H and ^{13}C NMR spectra were obtained at room temperature (25 °C) on Varian 400 MHz or Varian 600 MHz spectrometers.

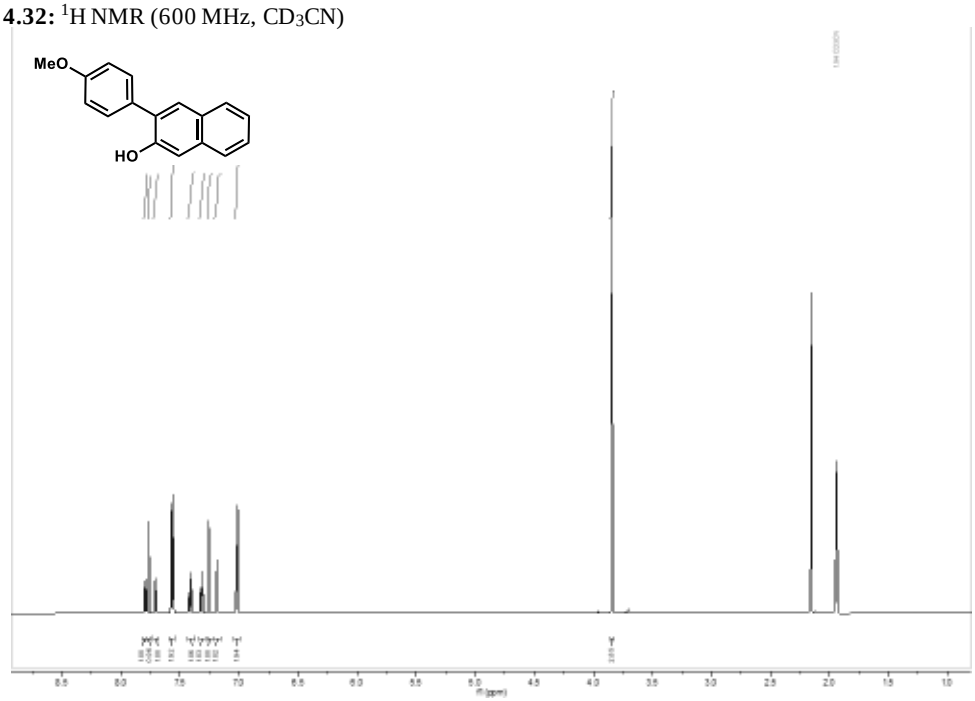
4.31: ^{1}H NMR (600 MHz, CD_3CN)



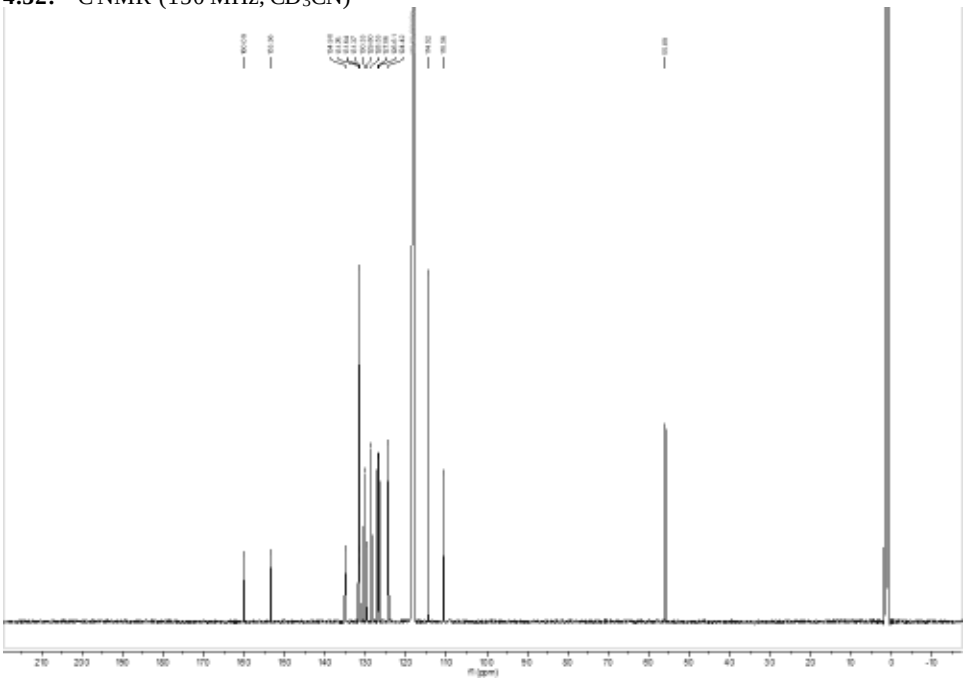
4.31: ^{13}C NMR (150 MHz, CD_3CN)



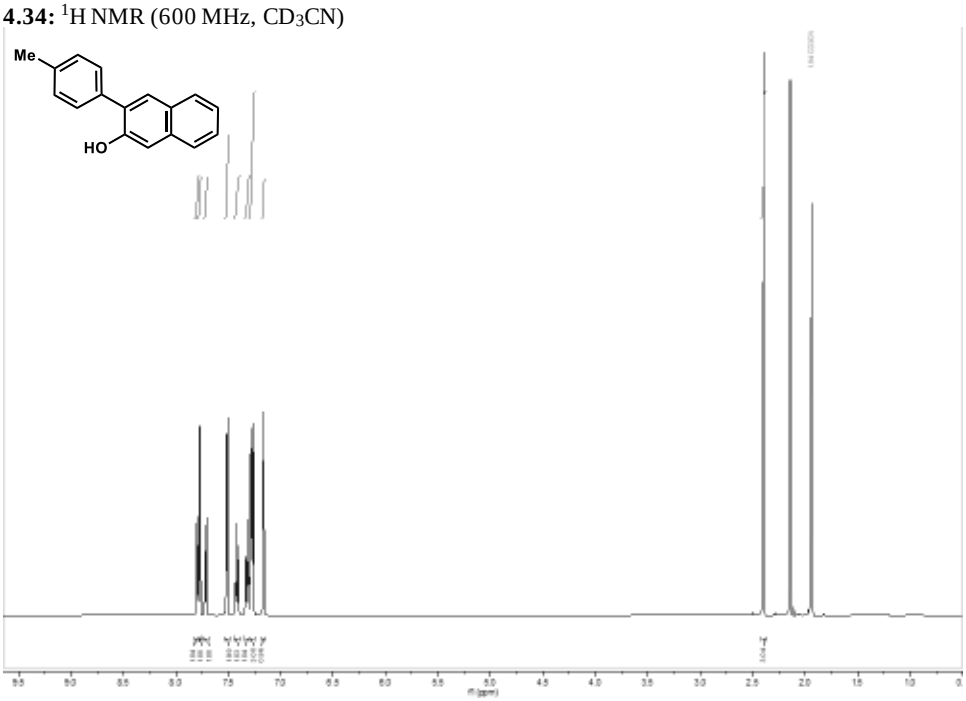
4.32: ^{1}H NMR (600 MHz, CD_3CN)



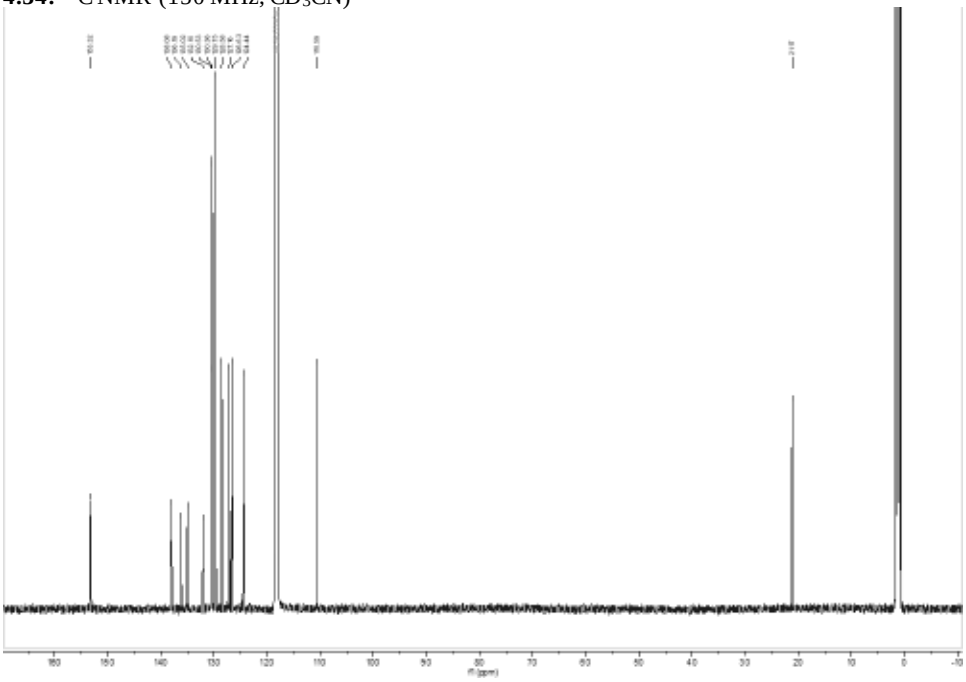
4.32: ^{13}C NMR (150 MHz, CD_3CN)



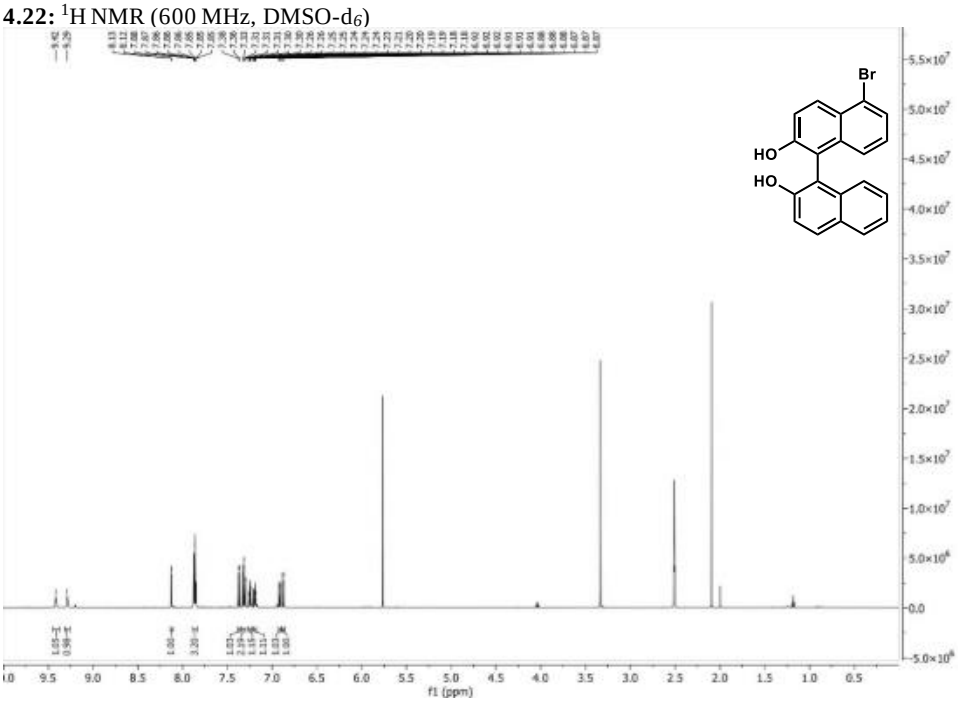
4.34: ^{1}H NMR (600 MHz, CD_3CN)



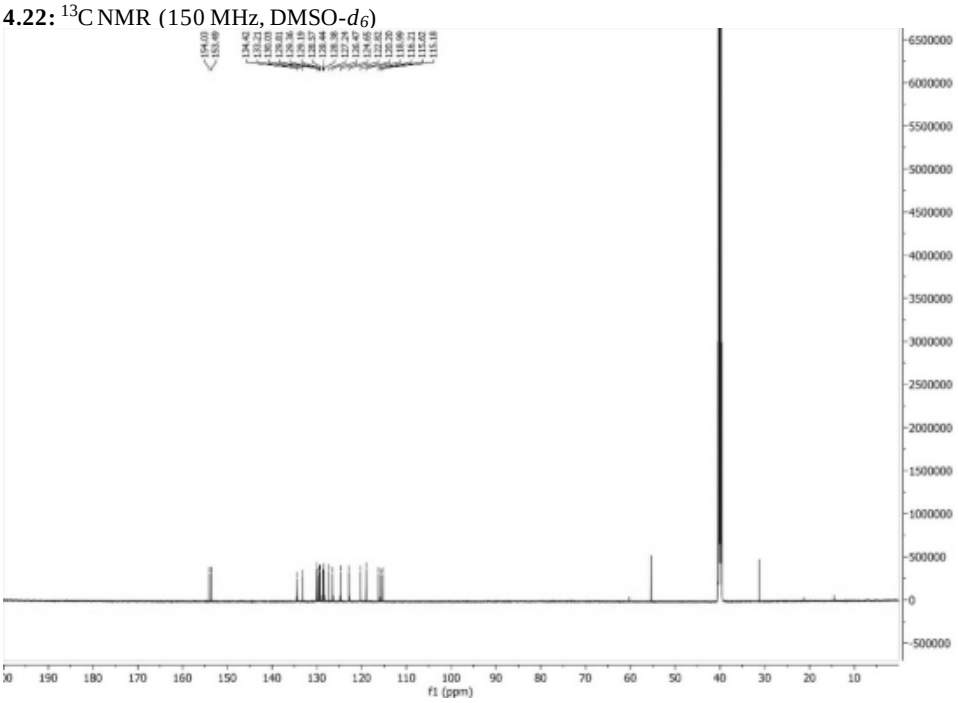
4.34: ^{13}C NMR (150 MHz, CD_3CN)



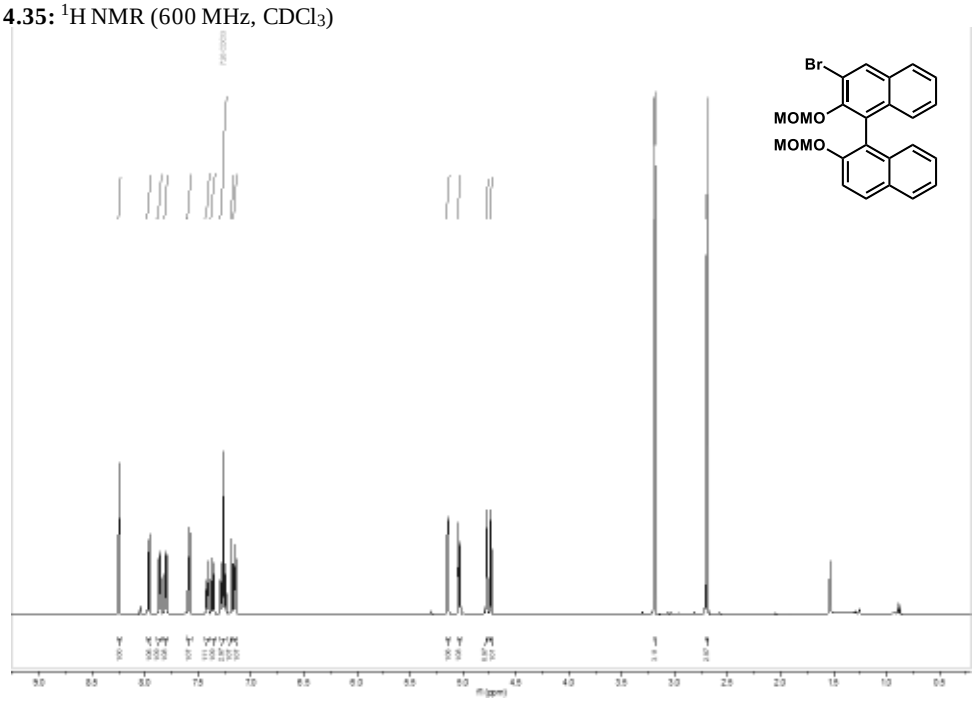
4.22: ^{1}H NMR (600 MHz, DMSO-d_6)



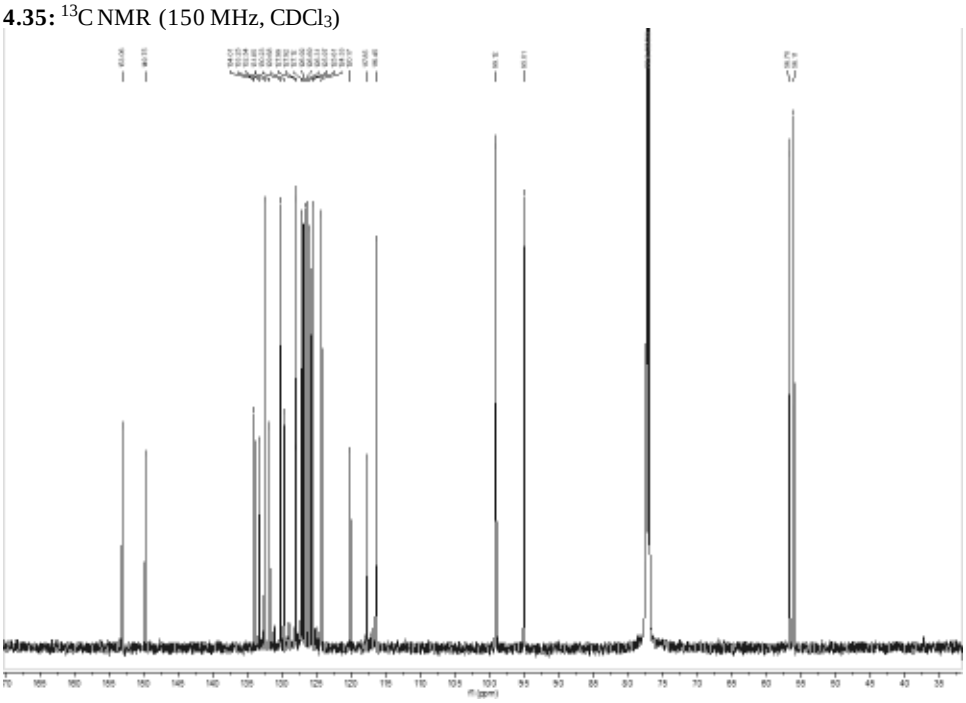
4.22: ^{13}C NMR (150 MHz, DMSO-d_6)



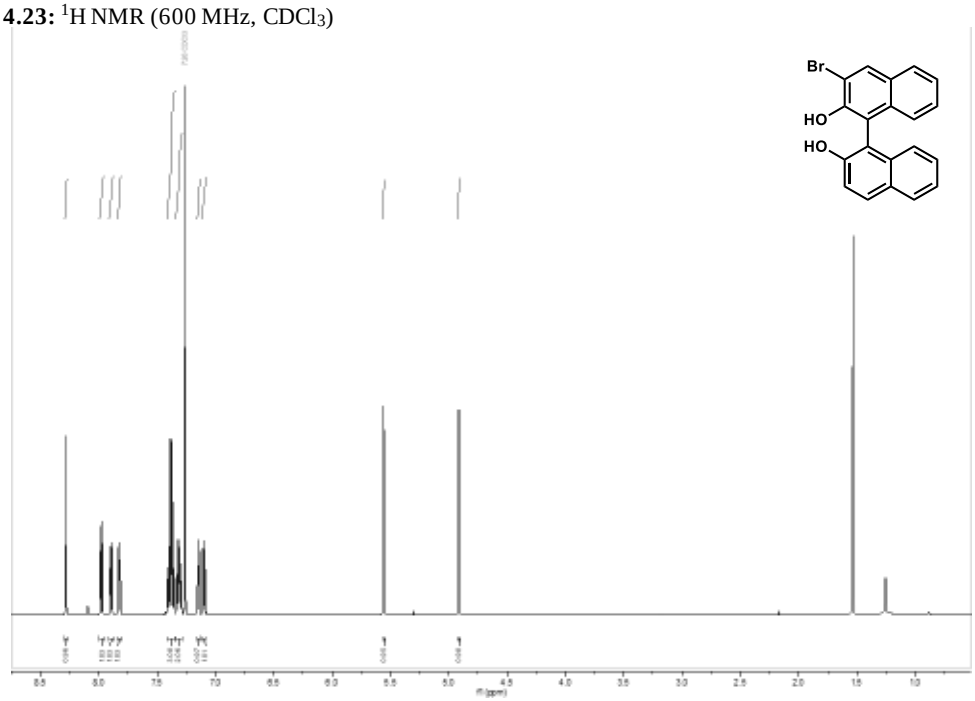
4.35: ^{1}H NMR (600 MHz, $CDCl_3$)



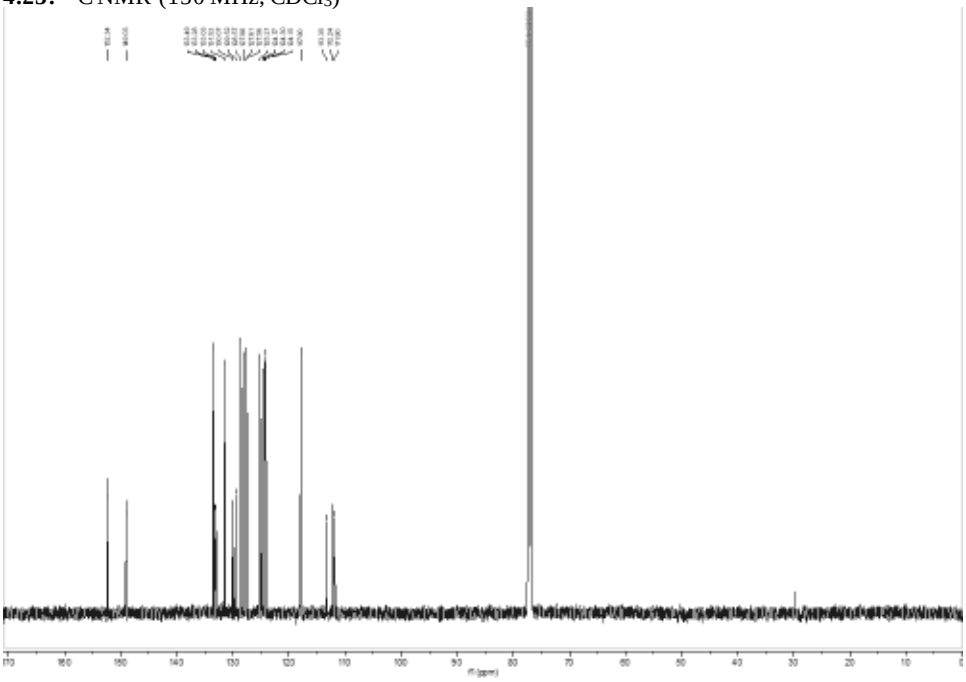
4.35: ^{13}C NMR (150 MHz, $CDCl_3$)



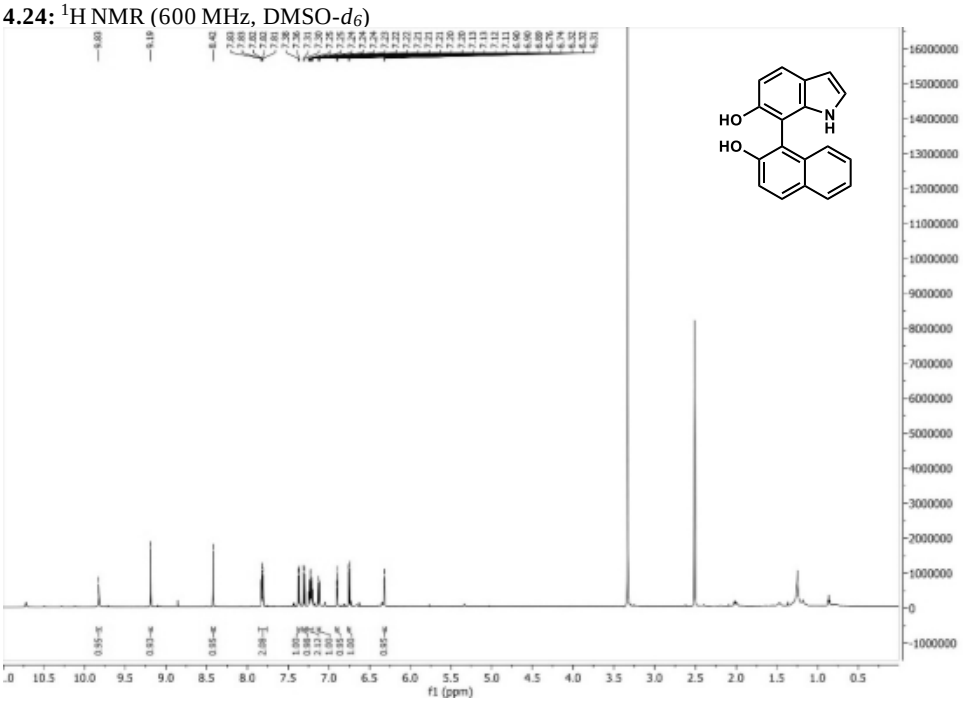
4.23: ^{1}H NMR (600 MHz, $CDCl_3$)



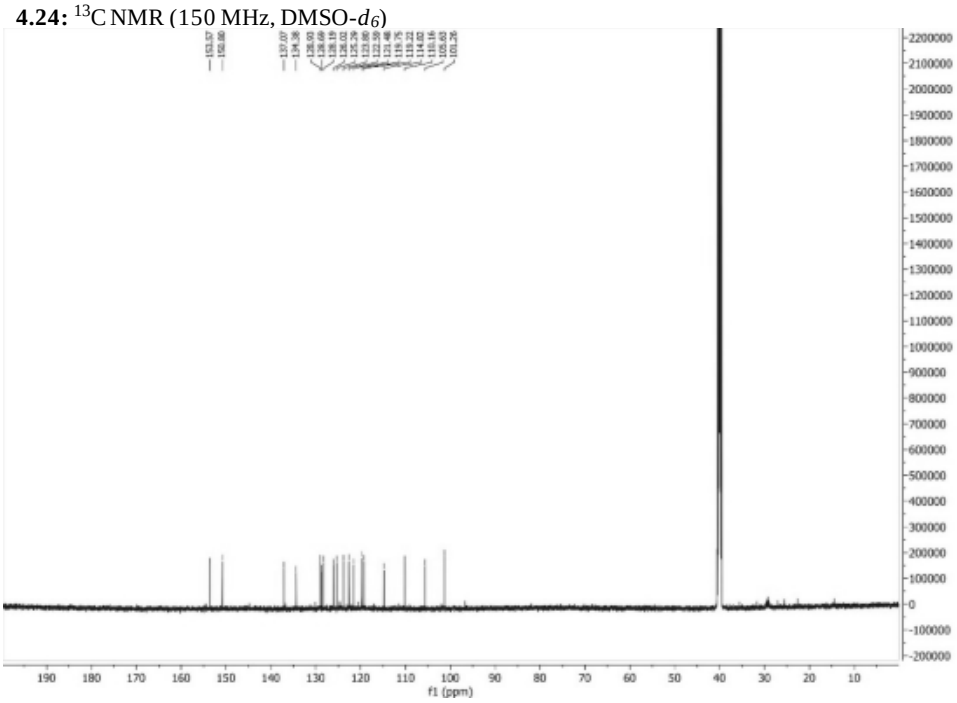
4.23: ^{13}C NMR (150 MHz, $CDCl_3$)



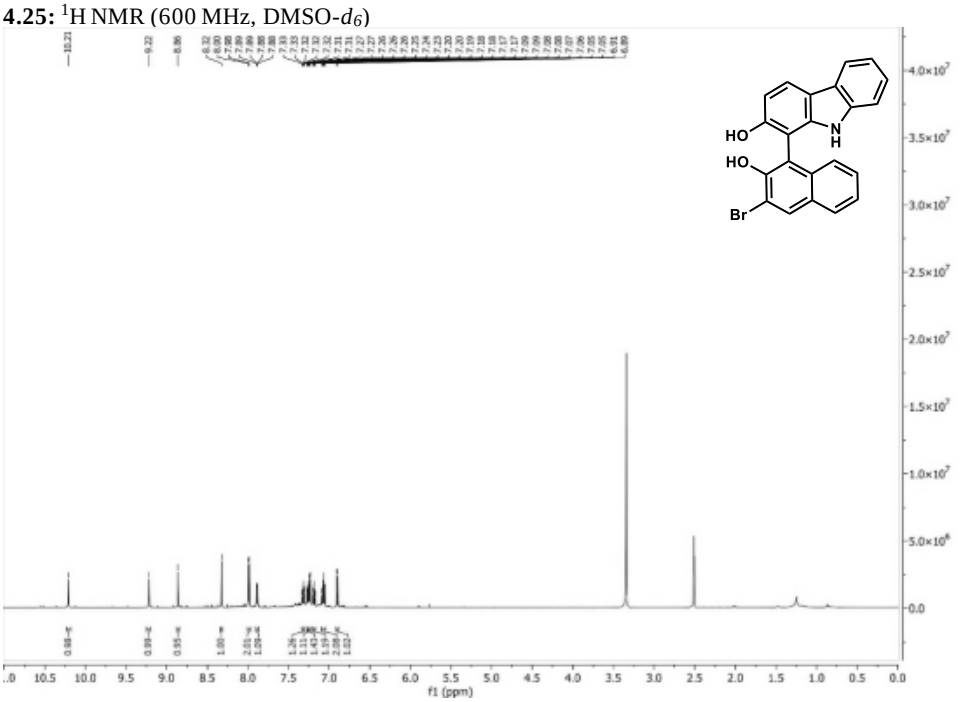
4.24: ^{1}H NMR (600 MHz, DMSO-d_6)



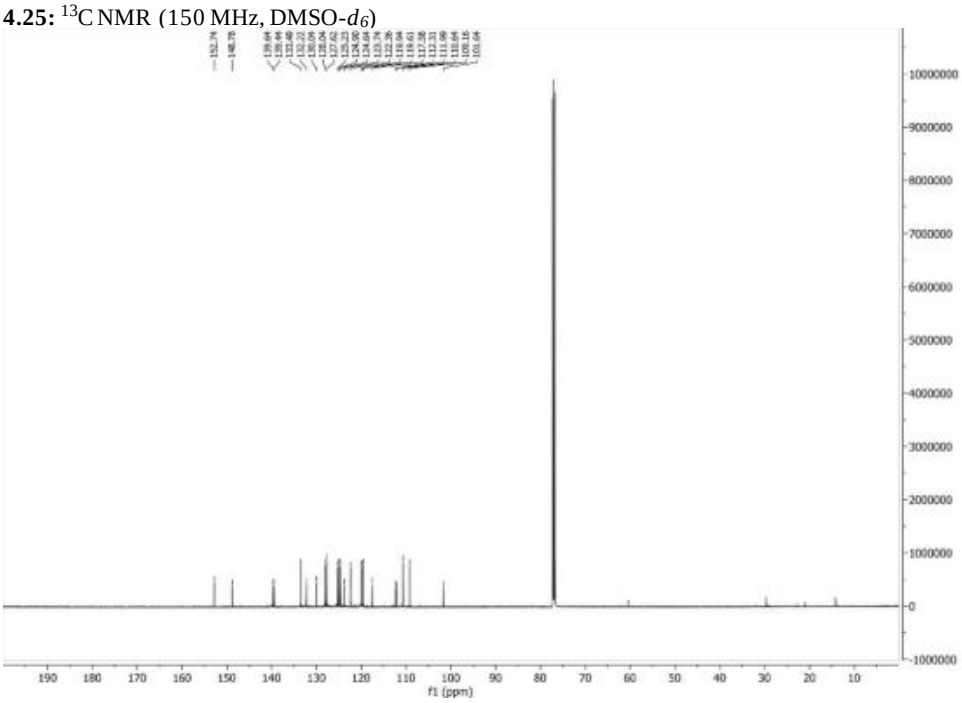
4.24: ^{13}C NMR (150 MHz, DMSO-d_6)



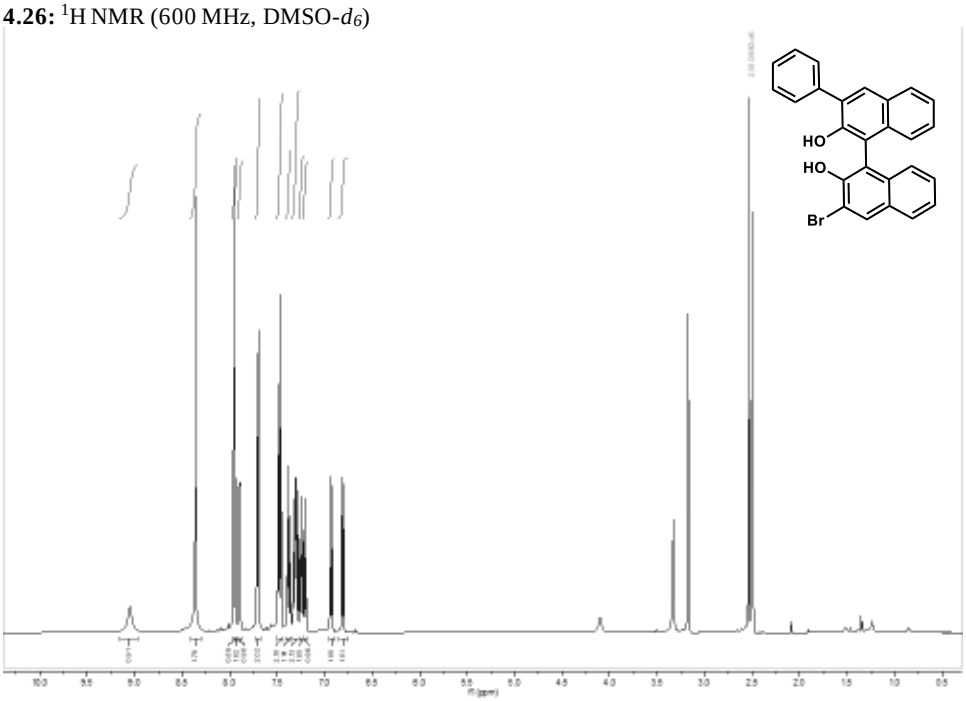
4.25: ^{1}H NMR (600 MHz, DMSO-d_6)



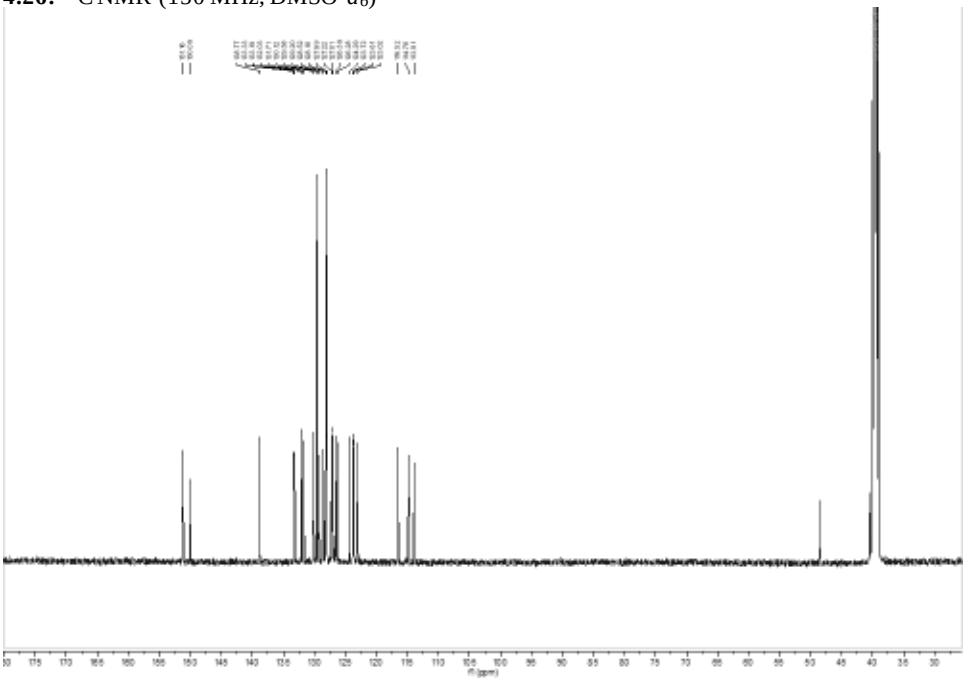
4.25: ^{13}C NMR (150 MHz, DMSO-d_6)



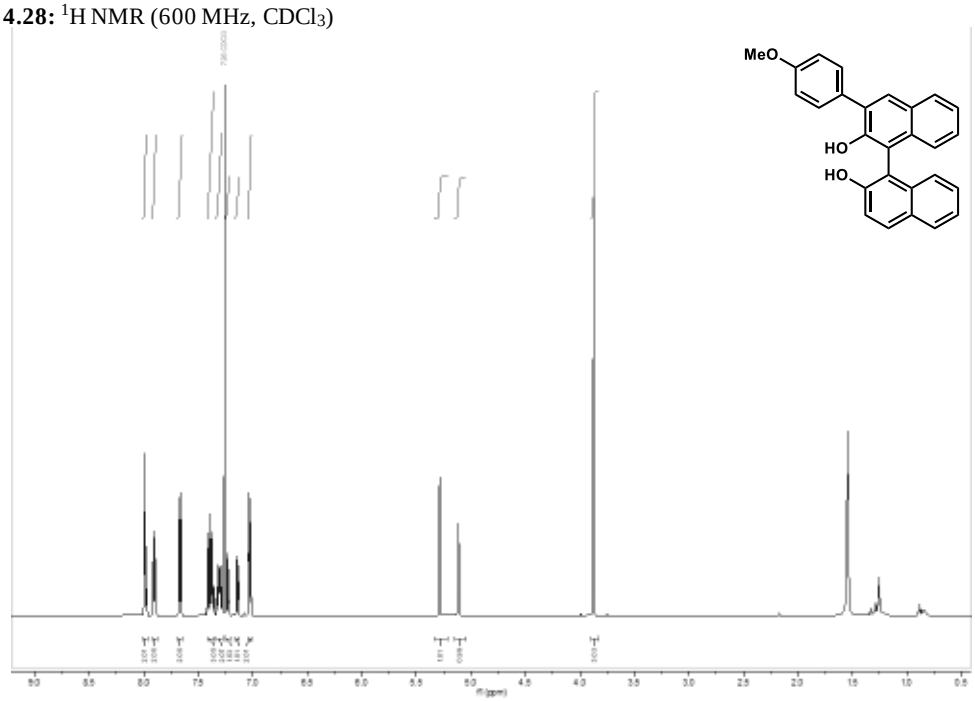
4.26: ^{1}H NMR (600 MHz, DMSO-d_6)



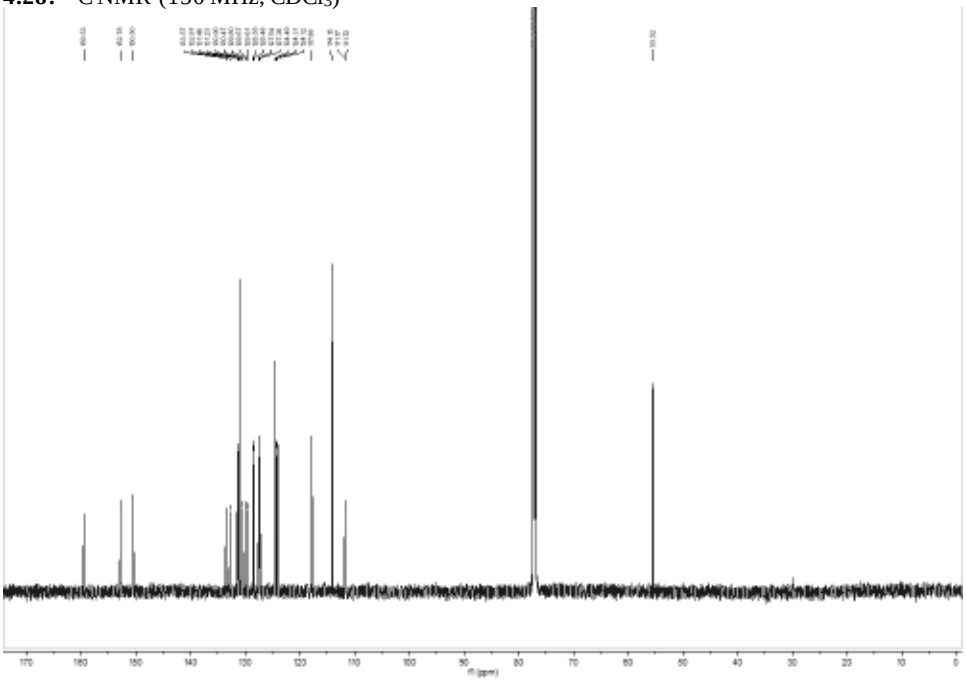
4.26: ^{13}C NMR (150 MHz, DMSO-d_6)



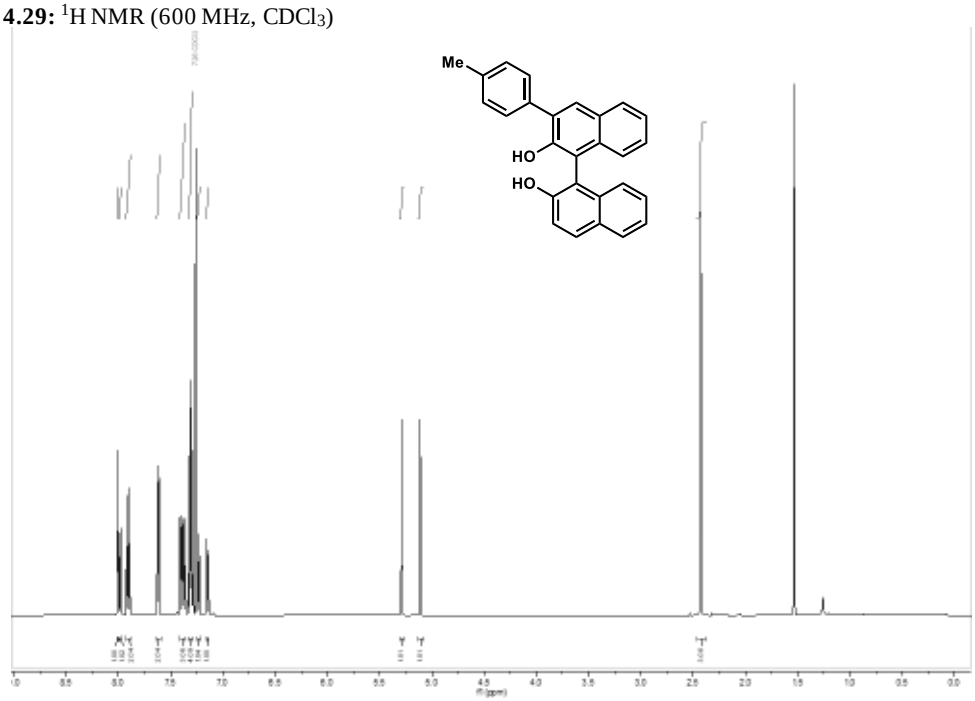
4.28: ^{1}H NMR (600 MHz, $CDCl_3$)



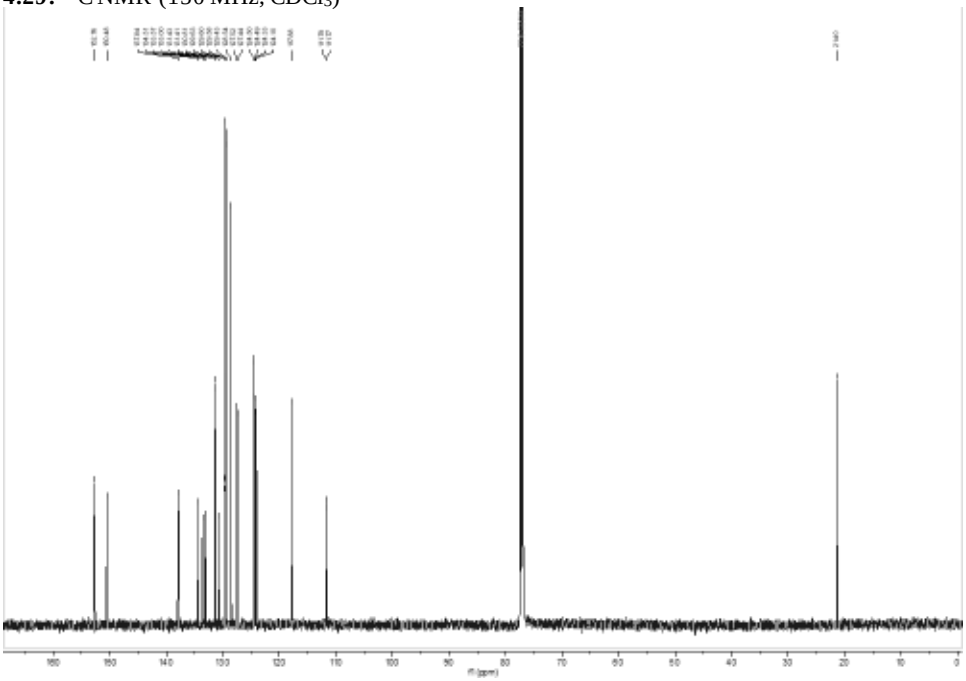
4.28: ^{13}C NMR (150 MHz, $CDCl_3$)



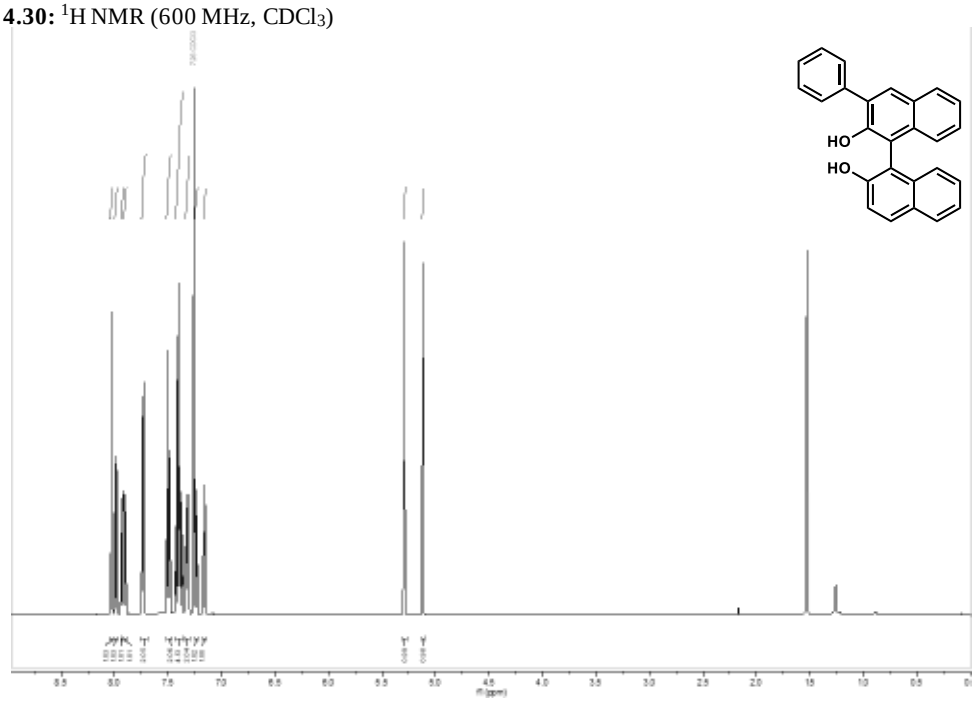
4.29: ^{1}H NMR (600 MHz, $CDCl_3$)



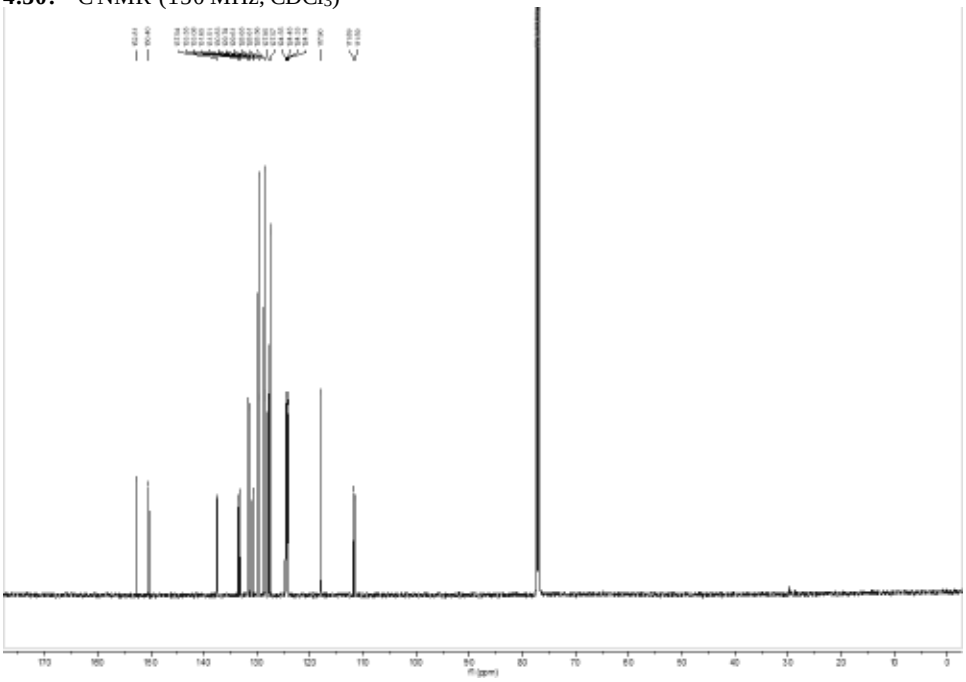
4.29: ^{13}C NMR (150 MHz, $CDCl_3$)



4.30: ^{1}H NMR (600 MHz, $CDCl_3$)



4.30: ^{13}C NMR (150 MHz, $CDCl_3$)



REFERENCES

1. Yoon, T. P.; Jacobsen, E. N. Privileged Chiral Catalysts. *Science* **2003**, *299*, 1691–1693.

2. Noyori, R.; Tomino, I.; Tanimoto, Y. Virtually Complete Enantioface Differentiation in Carbonyl Group Reduction by a Complex Aluminum Hydride Reagent. *J. Am. Chem. Soc.* **1979**, *101*, 3129–3131.

3. Brunel, J. M. BINOL: A Versatile Chiral Reagent. *Chem. Rev.* **2005**, *105*, 857–898.

4. Chen, Y.; Yekta, S.; Yudin, A. K. Modified BINOL Ligands in Asymmetric Catalysis. *Chem. Rev.* **2003**, *103*, 3155–3212.

5. Berthod, M.; Mignani, G.; Woodward, G.; Lemaire, M. Modified BINAP: The How and the Why. *Chem. Rev.* **2005**, *105*, 1801–1836.

6. Noyori, R.; Takaya, H. BINAP: An Efficient Chiral Element for Asymmetric Catalysis. *Acc. Chem. Res.* **1990**, *23*, 345–350.

7. Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M. Complete Field Guide to Asymmetric BINOL-Phosphate Derived Brønsted Acid and Metal Catalysis: History and Classification by Mode of Activation; Brønsted Acidity, Hydrogen Bonding, Ion Pairing, and Metal Phosphates. *Chem. Rev.* **2014**, *114*, 9047–9153.

8. Mukherjee, S.; List, B. Chiral Counteranions in Asymmetric Transition-Metal Catalysis: Highly Enantioselective Pd/Brønsted Acid-Catalyzed Direct α-Allylation of Aldehydes. *J. Am. Chem. Soc.* **2007**, *129*, 11336–11337.

9. Motoyama, Y.; Terada, M.; Mikami, K. Asymmetric Catalysis of (Hetero) Diels-Alder Cycloadditions by a Modified Binaphthol-Derived Titanium Complex. *Synlett* **1995**, *1995*, 967–968.

10. Patel, N. D.; Sieber, J. D.; Tcyrulnikov, S.; Simmons, B. J.; Rivalti, D.; Duvvuri, K.; Zhang, Y.; Gao, D. A.; Fandrick, K. R.; Haddad, N.; Lao, K. S.; Mangunuru, H. P. R.; Biswas, S.; Qu, B.; Grinberg, N.; Pennino, S.; Lee, H.; Song, J. J.; Gupton, B. F.; Garg, N. K.; Kozlowski, M. C.; Senanayake, C. H. Computationally Assisted Mechanistic Investigation and Development of Pd-Catalyzed Asymmetric Suzuki–Miyaura and Negishi Cross-Coupling Reactions for Tetra-*Ortho*-Substituted Biaryl Synthesis. *ACS Catal.* **2018**, *8*, 10190–10209.

11. Simonsen, K. B.; Gothelf, K. V.; Jørgensen, K. A. A Simple Synthetic Approach to 3,3′-Diaryl BINOLs. *J. Org. Chem.* **1998**, *63*, 7536–7538.

12. Tian, J.-M.; Wang, A.-F.; Yang, J.-S.; Zhao, X.-J.; Tu, Y.-Q.; Zhang, S.-Y.; Chen, Z.-M. Copper-Complex-Catalyzed Asymmetric Aerobic Oxidative Cross-Coupling of 2-Naphthols: Enantioselective Synthesis of 3,3′-Substituted C1-Symmetric BINOLs. *Angew. Chem. Int. Ed.* **2019**, *58*, 11023–11027.

13. Hovorka, M.; Günterova, J.; Zavada, J. Highly Selective Oxidative Cross-Coupling of Substituted 2-Naphthols: A Convenient Approach to Unsymmetrical 1,1′-Binaphthalene-2,2′-Diols. *Tet. Lett.* **1990**, *31*, 413–416.

14. Egami, H.; Katsuki, T. Iron-Catalyzed Asymmetric Aerobic Oxidation: Oxidative Coupling of 2-Naphthols. *J. Am. Chem. Soc.* **2009**, *131*, 6082–6083.

15. Yang, Y.; Lan, J.; You, J. Oxidative C–H/C–H Coupling Reactions between Two (Hetero) Arenes. *Chem. Rev.* **2017**, *117*, 8787–8863.

16. Kozlowski, M. C. Oxidative Coupling in Complexity Building Transforms. *Acc. Chem. Res.* **2017**, *50*, 638–643.

17. Shalit, H.; Dyadyuk, A.; Pappo, D. Selective Oxidative Phenol Coupling by Iron Catalysis. *J. Org. Chem.* **2019**, *84*, 1677–1686.

18. Egami, H.; Matsumoto, K.; Oguma, T.; Kunisu, T.; Katsuki, T. Enantioenriched Synthesis of C1-Symmetric BINOLs: Iron-Catalyzed Cross-Coupling of 2-Naphthols and Some Mechanistic Insight. *J. Am. Chem. Soc.* **2010**, *132*, 13633–13635.

19. Pfaltz, A.; Drury, W. J. Design of Chiral Ligands for Asymmetric Catalysis: From C2-Symmetric P, P-and N, N-Ligands to Sterically and Electronically Nonsymmetrical P, N-Ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5723–5726.

20. Endo, K.; Yakeishi, S.; Hamada, D.; Shibata, T. Functionalized BINOL-mono-PHOS for Multinuclear Cu-Catalysts in Asymmetric Conjugate Addition of Organozinc Reagents. *Chem. Lett.* **2013**, *42*, 547–549.

21. Kočovský, P.; Vyskočil, Š.; Smrčina, M. Non-Symmetrically Substituted 1,1′-Binaphthyls in Enantioselective Catalysis. *Chem. Rev.* **2003**, *103*, 3213–3246.

22. Trost, B. M.; Donckele, E. J.; Thaisrivongs, D. A.; Osipov, M.; Masters, J. T. A New Class of Non-C2-Symmetric Ligands for Oxidative and Redox-Neutral Palladium-Catalyzed Asymmetric Allylic Alkylations of 1,3-Diketones. *J. Am. Chem. Soc.* **2015**, *137*, 2776–2784.

23. Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. The Transformative Power of Biocatalysis in Convergent Synthesis. *J. Am. Chem. Soc.* **2022**, *144*, 5214–5225.

24. Zetzsche, L.; Yazarians, J.; Chakrabarty, S.; Hinze, M.; Lukowski, A.; Joyce, L.; Narayan, A. Biocatalytic Oxidative Cross-Coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603*, 79–85.

25. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504.

26. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta*. **2015**, *1854*, 1019–1037.

27. Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' Genomic Enzymology Web Tools for Functional Assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85.

28. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58*, 4169–4182.

29. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

30. Zhao, B.; Guengerich, F. P.; Bellamine, A.; Lamb, D. C.; Izumikawa, M.; Lei, L.; Podust, L. M.; Sundaramoorthy, M.; Kalaitzis, J. A.; Reddy, L. M.; Kelly, S. L.; Moore, B. S.; Stec, D.; Voehler, M.; Falck, J. R.; Shimada, T.; Waterman, M. R. Binding of Two Flaviolin Substrate Molecules, Oxidative Coupling, and Crystal Structure of *Streptomyces Coelicolor* A3(2) Cytochrome P450 158A2. *J Biol Chem* **2005**, *280*, 11599–11607.

31. Li, S.; Podust, L. M.; Sherman, D. H. Engineering and Analysis of a Self-Sufficient Biosynthetic Cytochrome P450 PikC Fused to the RhFRed Reductase Domain. *J. Am. Chem. Soc.* **2007**, *129*, 12940–12941.

32. Funa, N.; Funabashi, M.; Ohnishi, Y.; Horinouchi, S. Biosynthesis of Hexahydroxyperylenequinone Melanin via Oxidative Aryl Coupling by Cytochrome P-450 in *Streptomyces Griseus*. *J. Bacteriol.* **2005**, *187*, 8149–8155.

33. Werck-Reichhart, D.; Feyereisen, R. Cytochromes P450: A Success Story. *Genome Biol.* **2000**, *1*, 1–9.

34. Xu, L.-H.; Du, Y.-L. Rational and Semi-Rational Engineering of Cytochrome P450s for Biotechnological Applications. *Synth. Syst. Biotechnol.* **2018**, *3*, 283–290.

35. Guengerich, F. P.; Martin, M. V.; Sohl, C. D.; Cheng, Q. Measurement of Cytochrome P450 and NADPH–Cytochrome P450 Reductase. *Nat. Prot.* **2009**, *4*, 1245–1251.

CHAPTER V

Conclusions and Outlook

The work presented in this dissertation expands the toolbox for biocatalytic C–C bond formation through oxidative biaryl coupling reactions and provides clear evidence that biosynthetic P450s are highly tunable through protein engineering for a targeted cross-coupling activity and selectivity. This chapter provides a brief summary of this work, in addition to posing future directions for this type of transformation and more broadly for the development of convergent biocatalytic reactions, with the overarching goal of making providing reliable tools for biocatalytic fragment coupling reactions in a synthetic campaign. Achieving convergent synthetic strategies has long been a gold standard in constructing complex molecular skeletons, allowing for the rapid generation of complexity in comparatively streamlined synthetic routes. Traditionally, biocatalysis has not played a prominent role in convergent laboratory synthesis, with the application of biocatalysts in convergent strategies primarily limited to the synthesis of chiral fragments. Although the use of enzymes to enable convergent synthetic approaches is relatively new and emerging, combining the efficiency of convergent transformations with the selectivity achievable through biocatalysis creates new opportunities for efficient synthetic strategies.

Portions of this chapter were adapted with permission from Zetzsche, L. E.; Chakrabarty, S.; Narayan, A. R. H. The Transformative Power of Biocatalysis in Convergent Synthesis. *J. Am. Chem. Soc.* **2022**, *144*, 5214–5225. © American Chemical Society 2022.

DISSERTATION SUMMARY

The precision and efficiency offered by enzyme catalysis has propelled the incorporation of biocatalytic strategies into synthetic campaigns in academic and industrial spheres.[1-3] By providing a paradigm with catalyst-controlled selectivity, biocatalytic reactions have the potential to overcome limitations inherent to small molecule-mediated methods. Nevertheless, biocatalytic methods remain prohibitively underdeveloped compared to the vast repertoire of their small molecule-based counterparts, including for the formation of C–C bonds.[4] Although C–C bonds are the central framework of organic molecules, biocatalytic methods for their formation have historically been limited to a select few lyase enzymes.[5] Thus, more biocatalytic strategies for C–C bond formation need to be developed.

The work presented in this dissertation expands the toolbox for biocatalytic C–C bond formation through oxidative biaryl coupling reactions. Biaryl compounds, with two connected aromatic rings, are found across medicine, materials science, and asymmetric catalysis.[6,7] The necessity of joining arene building blocks to access these valuable compounds has inspired multiple approaches for biaryl bond formation and challenged chemists to develop increasingly concise and robust methods for this task. Oxidative coupling of two C–H bonds offers an efficient strategy for the formation of a biaryl C–C bond; however, fundamental challenges remain in controlling the reactivity and selectivity for a given pair of substrates.[8-12] In contrast, Nature has a wealth of enzymes that catalyze oxidative coupling reactions in the biosynthesis of biaryl natural products.[13] In particular, a growing subset of cytochrome P450 enzymes that perform site- and atroposelective oxidative biaryl coupling reactions have been discovered over the last decade,[14,15] offering a promising starting point in the development of selective biocatalytic strategies for this challenging reaction.

In Chapter II, the development of a platform for site- and atroposelective biaryl coupling in yeast using a fungal P450 catalyst is described.[16] This *in vivo* platform overcame challenges associated with fungal P450s, including their membrane-bound nature and catalytic-dependence on a separate redox partner, that has traditionally limited the use of these enzymes as biocatalysts in organic synthesis.[17,18] In particular, recombinant production of the fungal P450 KtnC in *Pichia pastoris*, allows for preparative scale biaryl coupling reactions on 100-mg scale.[16] Moreover, engineering a redox self-sufficient fusion enzyme further improves the efficiency of the system and sheds light on the redox compatibility of different yeast CPRs with KtnC.[16] Altogether, this work

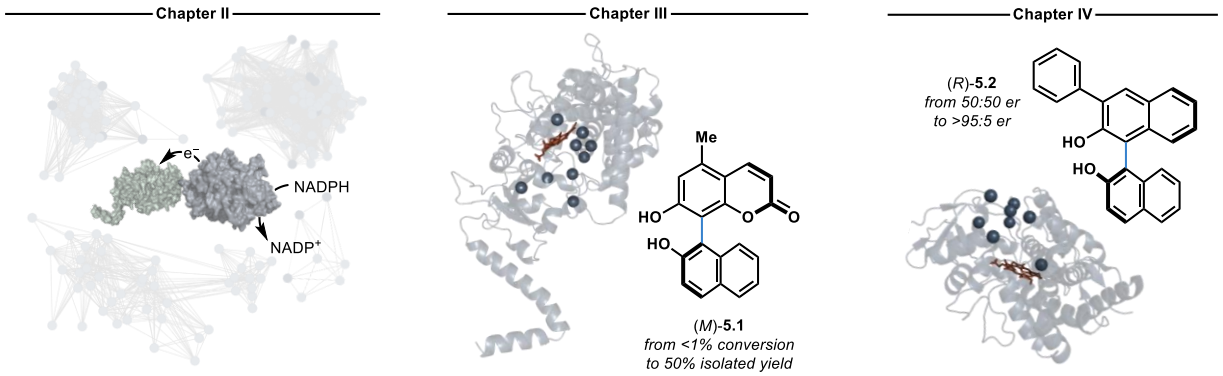


Figure 5.1. Dissertation summary. Fungal and bacterial biosynthetic P450s were engineered into self-sufficient fusion enzymes and evolved to catalyze highly selective oxidative cross-coupling reactions for biaryl bond formation.

provides a platform for biaryl coupling in yeast that can be applied to engineering the currently underexplored pool of fungal P450s into selective biocatalysts for the synthesis of complex biaryl compounds.

Next, in Chapter III, the whole-cell platform is applied to explore the biocatalytic potential of KtnC for unnatural cross-coupling reactions.[19] The ability for KtnC to catalyze a broad scope cross-coupling reactions on coumarin substrates was demonstrated, however the yield and selectivity of cross-coupling reactions with non-coumarin substrates was observed. To overcome this, KtnC was engineered through directed evolution to possess the desired reactivity, site-selectivity, and atroposelectivity in a cross-coupling between a coumarin and naphthol substrate by transforming a low-yielding, unselective reaction into a highly efficient and selective process. Moreover, the engineered biocatalysts showed improved cross-coupling activity across a diverse panel of phenolic substrates, demonstrating the generality of the engineered enzymes. Altogether, this streamlined method for constructing sterically hindered biaryl bonds provides a programmable platform for assembling molecules with catalyst-controlled reactivity and selectivity.

Finally, in Chapter IV, to expand the scope of biocatalytic cross-coupling reactions achievable, a bacterial P450 was engineered for improved activity and atroposelectivity in naphthol cross-coupling reactions to form privileged biaryl ligands. Profiling the cross-coupling substrate scope a small library of bacterial P450s engineered as P450-RhFRed fusion enzymes showed that previously uncharacterized P450s could productively catalyze naphthol cross-coupling reactions, albeit at low yields and atroposelectivity.[19] To overcome these limitations, CYP158A2 was engineered to catalyze the atroposelective cross-coupling between substituted naphthols with up to 95:5 er, providing efficient access to privileged chiral BINOL ligands.[6, 20]

FUTURE DIRECTIONS

The work in this dissertation provides clear evidence that biosynthetic P450s can be harnessed for oxidative cross-coupling chemistry beyond their native reactivity and are highly tunable with protein engineering for a target cross-coupling activity and selectivity. However, in order to reliably add these biocatalytic strategies to the synthetic chemist's toolbox for oxidative cross-coupling reactions, several key areas should be further investigated and developed.

First, the breadth of enzymes at our disposal should be expanded to increase the likelihood of identifying a hit enzyme for a given cross-coupling reaction. The work in this dissertation focused primarily on profiling and engineering two natural P450 catalysts: KtnC[21] and CYP158A2.[22]

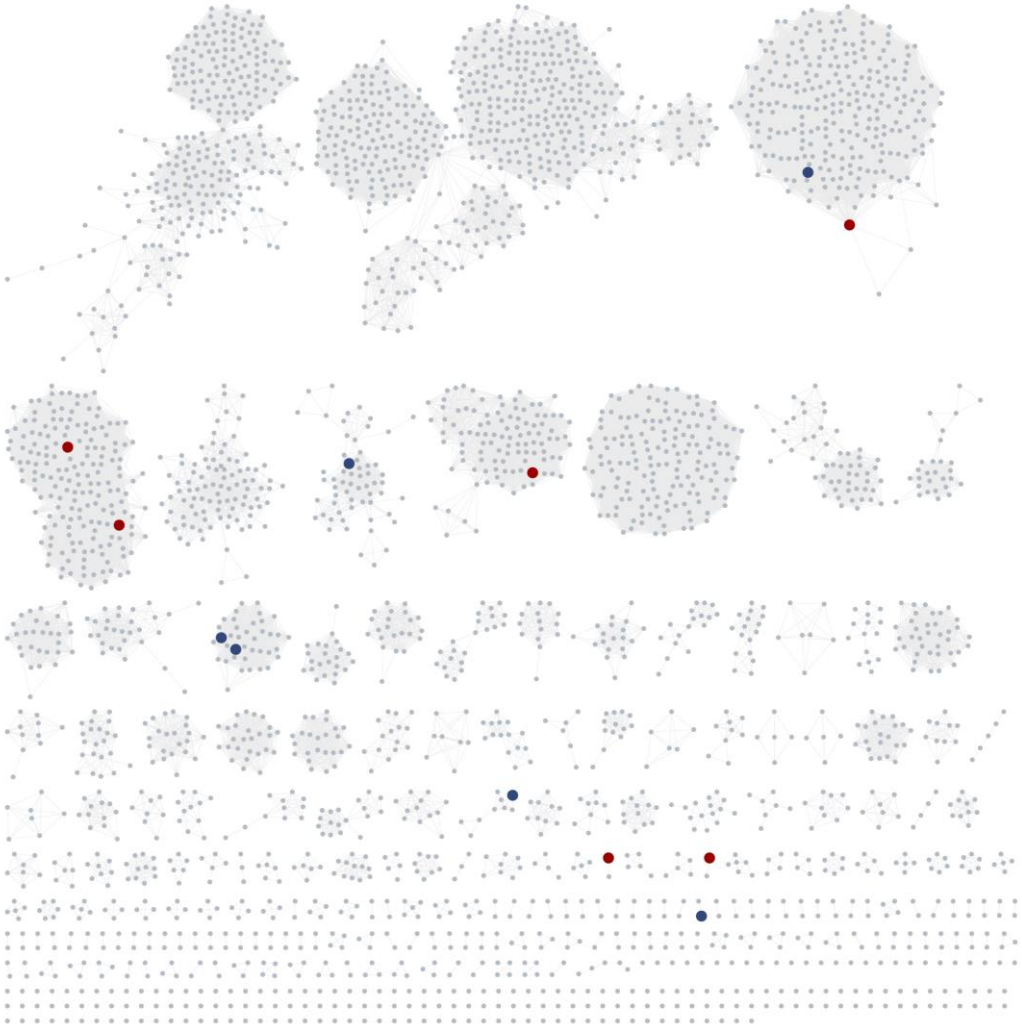


Figure 5.2. Unexplored bacterial P450 sequence space. Sequence similarity network highlighting P450 sequences with known biaryl coupling activity (red nodes), known hydroxylation activity (blue nodes), and related P450 sequences with unknown function (gray nodes). SSN parameters: E-value = 150; minimum sequence length = 300.

However, both of these enzymes exist within a much larger pool of P450 sequences with unknown function that have the potential to offer complementary cross-coupling activity and selectivity. In particular, sequence similarity network (SSN)-analysis[23-25] of bacterial P450s including CYP158A2[22] that are known to perform biaryl coupling reactions shows thousands of related sequences that have yet to be explored as biocatalysts for oxidative cross-coupling reactions (Figure 5.2).[19] Using the platform for high-throughput screening of P450-RhFRed fusion enzymes developed in Chapter 4, this natural sequence space would serve as a powerful starting point for the identification of a biocatalyst for a target reaction in a given synthetic campaign.

Second, a deeper understanding of the mechanism of oxidative coupling and factors that control substrate binding and orientation should be pursued. Although a substrate-bound crystal structure of CYP158A2 has been solved,[22] the exact mechanism of the C–C bond formation following the abstraction of the hydrogen atom from the first phenolic substrate is unknown (Figure 5.3). In one proposed mechanism, another hydrogen atom is abstracted from the second substrate, triggering a radical C–C coupling of the two oxidized substrates.[22] Alternatively, the C–C bond can form through a radical addition, in which the radical on the first substrate can add into the second, closed-shell substrate to form the C–C bond. A more rigorous analysis of this mechanism could provide insight into the types of substrates that could productively be cross-coupled. Additionally, this would provide insight into the effect different active site residues (and mutations) have on the substrate binding and orientation. With this in hand, hit enzymes identified from protein libraries of natural sequence space or an evolution campaign could more rapidly and rationally be engineered for a target activity or selectivity.[26, 27]

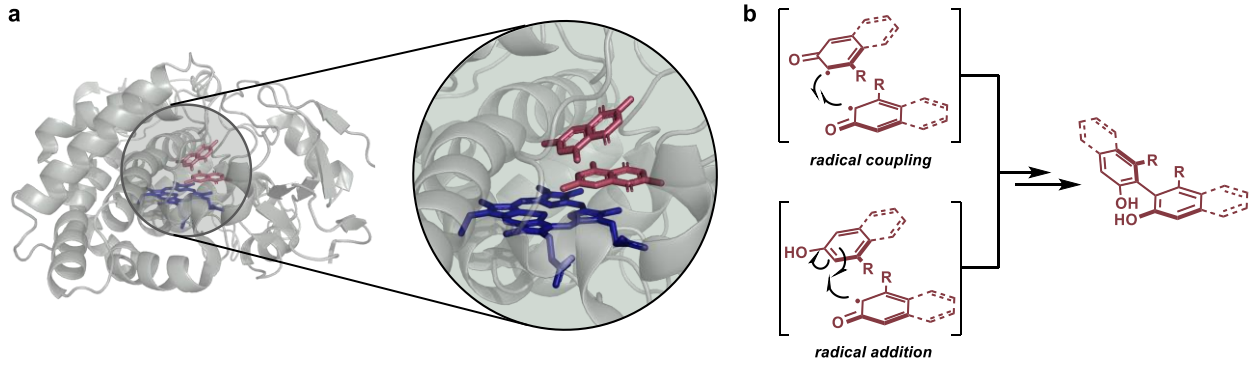


Figure 5.3. Proposed mechanism for biaryl coupling in P450 active site. (a) The crystal structure of CYP158A2 (PDB 2DKK) shows two flaviolin molecules bound in the P450 active site within 5.3 and 8.9 Å of the heme cofactor, respectively. (b) However, the exact mechanism of C–C bond formation is still unknown.

Lastly, the scalability of the biocatalytic reactions needs to be improved in order to reliably plan these enzymes into synthetic routes.[28] In order to accomplish this, both the substrate loading and enzyme stability could be improved upon. A directed evolution strategy could be employed to improve upon both of these factors, either through selection for the variants with the highest substrate turnover at increased substrate concentrations, or highest levels of activity after prolonged incubations at a target temperature and/or with increased amounts of cosolvents.[29] Additionally, alternative platforms for enzymatic catalysis could be explored including reactions with whole cells[30] or immobilized enzyme.[31] Ultimately, optimization of the biocatalytic platform to accommodate gram-scale reactions would drastically increase the applicability of these enzymes in the convergent synthesis of complex biaryl molecules.

BROADER OUTLOOK: BIOCATALYSIS IN CONVERGENT SYNTHESIS

Merging the tunable selectivity offered by biocatalysis with the efficiency of convergent transformations holds the potential for increasingly streamlined strategies for complex molecule synthesis. Although the extensive development of biocatalysts for functional group interconversions has increased the footprint of biocatalysis in the retrosynthetic playbook, biocatalysts are less commonly considered for constructing the core skeleton of molecules. Most biocatalytic reactions employed in multistep synthesis have centered on individual functional group manipulations to access chiral building blocks or late-stage functionalization reactions.[38-40] This trend can be observed in the role biocatalysis has historically played in convergent synthesis, with the utilization of enzymes for the synthesis of key fragments present in active pharmaceutical ingredients (Figure 5.4a). Enzymatic kinetic resolutions dominated the field of biocatalysis in the 1990s and early 2000s, resulting in the development of robust strategies for the preparation of chiral building blocks,[41-43] including the chiral amines present in amoxicillin (**5.3**)[44] and draflazine (**5.4**).[45,46] The development of directed evolution-based strategies to tailor biocatalysts for a targeted transformation drastically expanded the application of biocatalysis in asymmetric catalysis.[47] This new wave of biocatalysis brought the synthesis of chiral alcohols to the forefront of the field, with a wealth of catalysts capable of the stereoselective reduction of ketones, as demonstrated in the convergent synthesis of atorvastatin (**5.5**)[48] and atazanavir (**5.6**).[49,50] Additionally, biocatalysis has played a significant role in oxidative desymmetrization reactions, enabling the synthesis of the key chiral fragments present in telaprevir (**5.7**)[51] and islatravir (**5.8**).[52]

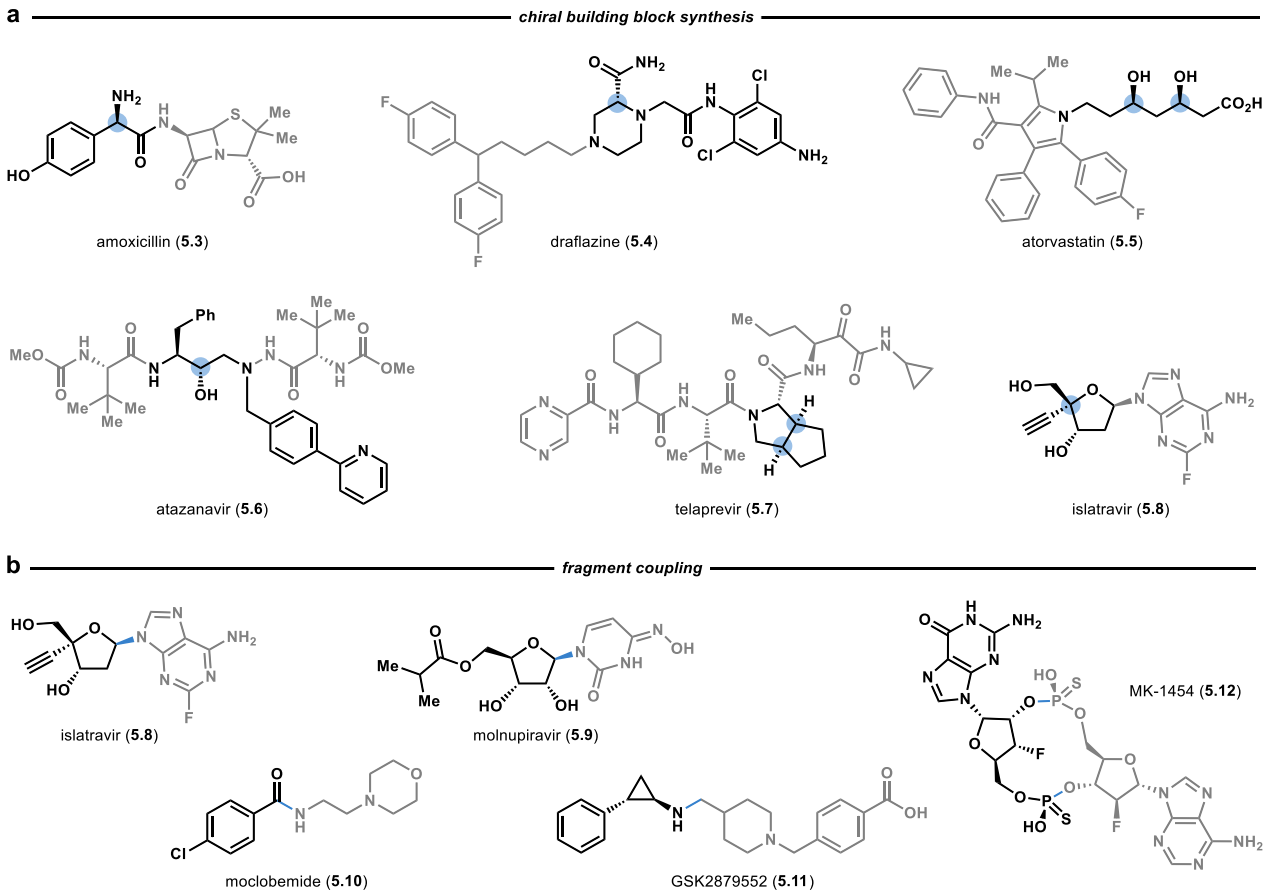


Figure 5.4. Biocatalysis in the convergent synthesis of pharmaceutical agents. (a) Traditionally, biocatalysts have been more commonly employed in the synthesis of chiral building blocks. (b) However, more recently, biocatalysts have been harnessed for fragment coupling steps.

Until recently, the area of convergent biocatalysis remained largely untapped, preventing the full realization of the impact enzymes can offer to complex molecule synthesis.[40] Encouragingly, the last decade has seen an increase in the development of convergent biocatalytic strategies, especially for C–N fragment coupling reactions (Figure 5.4b). Nucleoside phosphorylases from nucleotide salvage pathways can be engineered as convergent catalysts in the synthesis of unnatural nucleosides, as has been demonstrated in the biocatalytic synthesis of antivirals islatravir (**5.8**)[52] and molnupiravir (**5.9**).[53] Additionally, progress has recently been made in the development of convergent biocatalytic strategies for bread-and-butter reactions in medicinal chemistry such as amide bond formation and reductive aminations,[54, 55] such as in the synthesis antidepressant moclobemide (**5.10**)[56] and anticancer agent GSK2879552 (**5.11**),[29] respectively. Finally, breaking from the trend of C–N fragment coupling, a recent report from Merck disclosed the biocatalytic

cross-coupling and cyclization of two unnatural nucleotide monothiophosphates to form MK-1454 (**5.12**).[57]

Despite the recent progress in the development of convergent biocatalytic strategies, these methods remain prohibitively underdeveloped compared to the vast repertoire of small molecule catalysts and reagents that mediate fragment couplings reactions. The further development of this category of biocatalysts can potentially change the retrosynthetic framework for fragment coupling reactions by circumventing traditionally accepted requirements for protecting groups and functional handles into synthetic planning. To realize this goal, it is imperative that significant efforts are spent in the discovery and engineering of biosynthetic enzymes from secondary metabolite pathways that perform fragment coupling reactions, particularly ones that can forge C–C bonds. The work presented in this dissertation represents one such effort to add to the toolbox of biocatalysts for convergent C–C bond formation by engineering P450s for selective oxidative cross-coupling reactions.

REFERENCES

1. Winkler, C. K.; Schrittwieser, J. H.; Kroutil, W. Power of Biocatalysis for Organic Synthesis. *ACS Cent. Sci.* **2021**, *7*, 55–71.

2. Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. Chemoenzymatic Total Synthesis of Natural Products. *Acc. Chem. Res.* **2021**, *54*, 1374–1384.

3. Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem. Int. Ed.* **2021**, *60*, 88–119.

4. Zetzsche, L. E.; Narayan, A. R. H. Broadening the Scope of Biocatalytic C–C Bond Formation. *Nat. Rev. Chem.* **2020**, *4*, 334–346.

5. Schmidt, N. G.; Eger, E.; Kroutil, W. Building Bridges: Biocatalytic C-C-Bond Formation toward Multifunctional Products. *ACS Catal.* **2016**, *6*, 4286–4311.

6. Yoon, T. P.; Jacobsen, E. N. Privileged Chiral Catalysts. *Science* **2003**, *299*, 1691–1693.

7. Yet, L., Biaryls. In *Privileged Structures in Drug Discovery: Medicinal Chemistry and Synthesis*, First ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2018; pp 83–154.

8. Lee, Y. E.; Cao, T.; Torruellas, C.; Kozlowski, M. C. Selective Oxidative Homo- and Cross-Coupling of Phenols with Aerobic Catalysts. *J. Am. Chem. Soc.* **2014**, *136*, 6782–6785.

9. Yang, Y.; Lan, J.; You, J. Oxidative C–H/C–H Coupling Reactions between Two (Hetero) Arenes. *Chem. Rev.* **2017**, *117*, 8787–8863.

10. Kozlowski, M. C. Oxidative Coupling in Complexity Building Transforms. *Acc. Chem. Res.* **2017**, *50*, 638–643.

11. Nieves-Quinones, Y.; Paniak, T. J.; Lee, Y. E.; Kim, S. M.; Tcyrulnikov, S.; Kozlowski, M. C. Chromium-Salen Catalyzed Cross-Coupling of Phenols: Mechanism and Origin of the Selectivity. *J. Am. Chem. Soc.* **2019**, *141*, 10016–10032.

12. Shalit, H.; Dyadyuk, A.; Pappo, D. Selective Oxidative Phenol Coupling by Iron Catalysis. *J. Org. Chem.* **2019**, *84*, 1677–1686.

13. Bringmann, G.; Günther, C.; Ochse, M.; Schupp, O.; Tasler, S., Biaryls in Nature: A Multi-Facetted Class of Stereochemically, Biosynthetically, and Pharmacologically Intriguing Secondary Metabolites. In *Progress in the Chemistry of Organic Natural Products*, Bringmann, G.; Günther, C.; Ochse, M.; Schupp, O.; Tasler, S.; Herz, W.; Falk, H.; Kirby, G. W.; Moore, R. E., Eds. Springer Vienna: Vienna, 2001; pp 1–249.

14. Aldemir, H.; Richarz, R.; Gulder, T. A. The Biocatalytic Repertoire of Natural Biaryl Formation. *Angew. Chem. Int. Ed.* **2014**, *53*, 8286–8293.

15. Hüttel, W.; Müller, M. Regio- and Stereoselective Intermolecular Phenol Coupling Enzymes in Secondary Metabolite Biosynthesis. *Nat. Prod. Rep.* **2021**, *38*, 1011–1043.

16. Zetzsche, L. E.; Chakrabarty, S. C.; Narayan, A. R. H. A Biocatalytic Platform for the Synthesis of Biaryl Molecules in Yeast. *Submitted (April 4, 2022)*.

17. Durairaj, P.; Hur, J.-S.; Yun, H. Versatile Biocatalysis of Fungal Cytochrome P450 Monooxygenases. *Microb. Cell Factories* **2016**, *15*, 125.

18. Lah, L.; Kraševec, N.; Trontelj, P.; Komel, R. High Diversity and Complex Evolution of Fungal Cytochrome P450 Reductase: Cytochrome P450 Systems. *Fungal Genet. Biol.* **2008**, *45*, 446–458.

19. Zetzsche, L.; Yazarians, J.; Chakrabarty, S.; Hinze, M.; Lukowski, A.; Joyce, L.; Narayan, A. Biocatalytic Oxidative Cross-Coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603*, 79–85.

20. Chen, Y.; Yekta, S.; Yudin, A. K. Modified BINOL Ligands in Asymmetric Catalysis. *Chem. Rev.* **2003**, *103*, 3155–3212.

21. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

22. Zhao, B.; Guengerich, F. P.; Bellamine, A.; Lamb, D. C.; Izumikawa, M.; Lei, L.; Podust, L. M.; Sundaramoorthy, M.; Kalaitzis, J. A.; Reddy, L. M.; Kelly, S. L.; Moore, B. S.; Stec, D.; Voehler, M.; Falck, J. R.; Shimada, T.; Waterman, M. R. Binding of Two Flaviolin Substrate Molecules, Oxidative Coupling, and Crystal Structure of *Streptomyces Coelicolor* A3(2) Cytochrome P450 158A2. *J Biol Chem* **2005**, *280*, 11599–11607.

23. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019–1037.

24. Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58*, 4169–4182.

25. Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' Genomic Enzymology Web Tools for Functional Assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85.

26. Fasan, R. Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal.* **2012**, *2*, 647–666.

27. Xu, L.-H.; Du, Y.-L. Rational and Semi-Rational Engineering of Cytochrome P450s for Biotechnological Applications. *Synth. Syst. Biotechnol.* **2018**, *3*, 283–290.

28. Chakrabarty, S.; Wang, Y.; Perkins, J. C.; Narayan, A. R. H. Scalable Biocatalytic C–H Oxyfunctionalization Reactions. *Chem. Soc. Rev.* **2020**, *49*, 8137–8155.

29. Schober, M.; MacDermaid, C.; Ollis, A. A.; Chang, S.; Khan, D.; Hosford, J.; Latham, J.; Ihnken, L. A. F.; Brown, M. J. B.; Fuerst, D.; Sanganee, M. J.; Roiban, G.-D. Chiral Synthesis of Lsd1 Inhibitor Gsk2879552 Enabled by Directed Evolution of an Imine Reductase. *Nat. Catal.* **2019**, *2*, 909–915.

30. Klenk, J. M.; Nebel, B. A.; Porter, J. L.; Kulig, J. K.; Hussain, S. A.; Richter, S. M.; Tavanti, M.; Turner, N. J.; Hayes, M. A.; Hauer, B.; Flitsch, S. L. The Self-Sufficient P450 RhF Expressed in a Whole Cell System Selectively Catalyses the 5-Hydroxylation of Diclofenac. *Biotechnol. J.* **2017**, *12*, 1600520.

31. Sheldon, R. A.; van Pelt, S. Enzyme Immobilisation in Biocatalysis: Why, What and How. *Chem. Soc. Rev.* **2013**, *42*, 6223–6235.

32. McCowen, S. V.; Doering, N. A.; Sarpong, R. Retrosynthetic Strategies and Their Impact on Synthesis of Arcutane Natural Products. *Chem. Sci.* **2020**, *11*, 7538–7552.

33. Hendrickson, J. B. Systematic Synthesis Design. IV. Numerical Codification of Construction Reactions. *J. Am. Chem. Soc.* **1975**, *97*, 5784–5800.

34. Urabe, D.; Asaba, T.; Inoue, M. Convergent Strategies in Total Syntheses of Complex Terpenoids. *Chem. Rev.* **2015**, *115*, 9207–9231.

35. Smith, M. B., Synthetic Strategies. In *Organic Synthesis (Third Edition)*, Smith, M. B., Ed. Academic Press: Oxford, 2010; pp 897–997.

36. Pompeo, M. M.; Cheah, J. H.; Movassaghi, M. Total Synthesis and Anti-Cancer Activity of All Known Communesin Alkaloids and Related Derivatives. *J. Am. Chem. Soc.* **2019**, *141*, 14411–14420.

37. Lin, H.-C.; McMahon, T. C.; Patel, A.; Corsello, M.; Simon, A.; Xu, W.; Zhao, M.; Houk, K. N.; Garg, N. K.; Tang, Y. P450-Mediated Coupling of Indole Fragments to Forge Communesin and Unnatural Isomers. *J. Am. Chem. Soc.* **2016**, *138*, 4002–4005.

38. 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*, 8003–8049.

39. Hönig, M.; Sondermann, P.; Turner, N. J.; Carreira, E. M. Enantioselective Chemo - and Biocatalysis: Partners in Retrosynthesis. *Angew. Chem. Int. Ed.* **2017**, *56*, 8942–8973.

40. Turner, N. J.; O'Reilly, E. Biocatalytic Retrosynthesis. *Nat. Chem. Biol.* **2013**, *9*, 285–288.

41. Breuer, M.; Ditrich, K.; Habicher, T.; Hauer, B.; Keßeler, M.; Stürmer, R.; Zelinski, T. Industrial Methods for the Production of Optically Active Intermediates. *Angew. Chem. Int. Ed.* **2004**, *43*, 788–824.

42. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409*, 258–268.

43. Pàmies, O.; Bäckvall, J.-E. Combined Metal Catalysis and Biocatalysis for an Efficient Deracemization Process. *Curr. Opin. Biotechnol.* **2003**, *14*, 407–413.

44. 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. Process Res. Dev.* **2002**, *6*, 452–457.

45. Bruce, M.; St. Laurent, D.; Poindexter, G.; Monkovic, I.; Huang, S.; Balasubramanian, N. Kinetic Resolution of Piperazine-2-Carboxamide by Leucine Aminopeptidase. An Application in the Synthesis of the Nucleoside Transport Blocker (–)-Draflazine. *Synth. Commun.* **1995**, *25*, 2673–2684.

46. Eichhorn, E.; Roduit, J.-P.; Shaw, N.; Heinzmann, K.; Kiener, A. Preparation of (*S*)-Piperazine-2-Carboxylic Acid, (*R*)-Piperazine-2-Carboxylic Acid, and (*S*)-Piperidine-2-Carboxylic Acid by Kinetic Resolution of the Corresponding Racemic Carboxamides with Stereoselective Amidases in Whole Bacterial Cells. *Tetrahedron Asymmetry* **1997**, *8*, 2533–2536.

47. 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*, 185–194.

48. Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chem.* **2010**, *12*, 81–86.

49. de Miranda, A. S.; Simon, R. C.; Grischek, B.; de Paula, G. C.; Horta, B. A.; de Miranda, L. S.; Kroutil, W.; Kappe, C. O.; de Souza, R. O. Chiral Chlorohydrins from the Biocatalyzed Reduction of Chloroketones: Chiral Building Blocks for Antiretroviral Drugs. *ChemCatChem* **2015**, *7*, 984–992.

50. Wu, K.; Zheng, K.; Xiong, L.; Yang, Z.; Jiang, Z.; Meng, X.; Shao, L. Efficient Synthesis of an Antiviral Drug Intermediate Using an Enhanced Short-Chain Dehydrogenase in an Aqueous-Organic Solvent System. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 4417–4427.

51. Köhler, V.; Bailey, K. R.; Znabet, A.; Raftery, J.; Helliwell, M.; Turner, N. J. Enantioselective Biocatalytic Oxidative Desymmetrization of Substituted Pyrrolidines. *Angew. Chem.* **2010**, *122*, 2228–2230.

52. Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T. Design of an *In Vitro* Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366*, 1255–1259.

53. McIntosh, J. A.; Benkovics, T.; Silverman, S. M.; Huffman, M. A.; Kong, J.; Maligres, P. E.; Itoh, T.; Yang, H.; Verma, D.; Pan, W.; Ho, H.-I.; Vroom, J.; Knight, A. M.; Hurtak, J. A.; Klapars, A.; Fryszkowska, A.; Morris, W. J.; Strotman, N. A.; Murphy, G. S.; Maloney, K. M.; Fier, P. S. Engineered Ribosyl-1-Kinase Enables Concise Synthesis of Molnupiravir, an Antiviral for COVID-19. *ACS Cent. Sci.* **2021**, *7*, 1980–1985.

54. Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, *54*, 3451-3479.

55. Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug. Discov.* **2018**, *17*, 709-727.

56. Petchey, M. R.; Rowlinson, B.; Lloyd, R. C.; Fairlamb, I. J.; Grogan, G. Biocatalytic Synthesis of Moclobemide Using the Amide Bond Synthetase McbA Coupled with an ATP Recycling System. *ACS Catal.* **2020**, *10*, 4659–4663.

57. McIntosh, J. A.; Liu, Z.; Andresen, B. M.; Marzijarani, N. S.; Moore, J. C.; Marshall, N. M.; Borra-Garske, M.; Obligacion, J. V.; Fier, P. S.; Peng, F.; Forstater, J. H.; Winston, M. S.; An, C.; Chang, W.; Lim, J.; Huffman, M. A.; Miller, S. P.; Tsay, F.-R.; Altman, M. D.; Lesburg, C. A.; Steinhuebel, D.; Trotter, B. W.; Cumming, J. N.; Northrup, A.; Bu, X.; Mann, B. F.; Biba, M.; Hiraga, K.; Murphy, G. S.; Kolev, J. N.; Makarewicz, A.; Pan, W.; Farasat, I.; Bade, R. S.; Stone, K.; Duan, D.; Alvizo, O.; Adpressa, D.; Guetschow, E.; Hoyt, E.; Regalado, E. L.; Castro, S.; Rivera, N.; Smith, J. P.; Wang, F.; Crespo, A.; Verma, D.; Axnanda, S.; Dance, Z. E. X.; Devine, P. N.; Tschaen, D.; Canada, K. A.; Bulger, P. G.; Sherry, B. D.; Truppo, M. D.; Ruck, R. T.; Campeau, L.-C.; Bennett, D. J.; Humphrey, G. R.; Campos, K. R.; Maddess, M. L. A Kinase-cGAS Cascade to Synthesize a Therapeutic Sting Activator. *Nature* **2022**, *603*, 439–444.