Rieske Non-heme Iron Oxygenases in Natural Product Biosynthesis

by

Jianxin Liu

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

Doctoral Committee:

Assistant Professor Jennifer Bridwell-Rabb, Chair
Professor Nicolai Lehnert
Associate Professor Alison Narayan
Professor Stephen Ragsdale

Jianxin Liu

jianxinl@umich.edu

ORCID iD: 0000-0001-5713-434X

Dedication

To my dear motherland and to the lovely people who have supported me throughout my education and my life. Thanks for making me see this adventure through to the end and set sail for the dream.

Acknowledgements

My deepest gratitude goes first and foremost to my advisor, Prof. Jennifer Bridwell-Rabb, for her constant support and guidance through my graduate training. I was inspired by her enthusiasm and attainment and was then led into the amazing world of structural biology and biochemistry. Without her encouragement and unparalleled supervision, I would not have achieved and finished this dissertation as I literally didn't even know what SDS-PAGE gel was when I entered the graduate school.

I also want to express thanks to Prof. Alison Narayan, Prof. Nicolai Lehnert, and Prof. Stephen Ragsdale for serving on my dissertation committee. I greatly appreciate their time and dedication. I am especially thankful to Prof. Narayan for accepting me as an overseas training undergraduate student in her laboratory and giving me a great impression of Michigan Chemistry.

I had great pleasure of working with an amazing group of colleagues and collaborators during my graduate studies. Dr. Wendy Feng, Dr. Bin Li, Dr. April Lukowski, Jiayi Tian, Minshik Jo, David Boggs and Madison Knapp have been incredible people to both learn from and learn with. I also appreciate all my lab mates I have worked with.

I am also grateful to all my dear friends who accompanied me through thick and thin on this journey. Jiayi, thank you for driving me home every day and sharing your ideas not only for how to conduct better research but also for what to have for dinner. Tianyi and Tao, thank you for all the unforgettable wonderful memories that we shared on the basketball court and in the barbeque park. Most importantly, thank you Yulei! Thank you for helping me settle down and integrate into local life here in Ann Arbor; thank you for getting me out of frustration and stress during graduate school. And thanks for your accompanying through all the laughter and tears in the past four years!

Last but not least, I would like to thank my parents. Mom and Dad, although we haven't seen each other for four years, the daily FaceTime really encourages me and remind me that no matter what happens, you are always there standing with me and being my strongest support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

*At*CAO – *Arabidopsis thaliana* Chlorophyll(ide) *a* Oxygenase

Bchl – Bacteriochlorophyll

Bchlide – Bacteriochlorophyllide

CAO – Chlorophyll(ide) *a* Oxygenase

CBR – Chlorophyll(ide) *b* Reductase

Chl – Chlorophyll

Chlase – Chlorophyllase

Chlide – Chlorophyllide

*Cr*CAO – *Chlamydomonas reinhardtii* Chlorophyll(ide) *a* Oxygenase

CV – Column Volume

Da – Dalton

dcSTX – decarbamoylsaxitoxin

ddSTX – dideoxysaxitoxin

DMSO – Dimethyl sulfoxide

DTT – dithiothreitol

EIC – Extracted Ion Chromatogram

E. coli – *Escherichia coli*

FAD – Flavin Adenine Dinucleotide

Fdx – Ferredoxin

Flv – Flavodoxin

Flx – Flavodoxin reductase

FMN – Flavin mononucleotide

FNR – Ferredoxin NAD(P)$^+$ Reductase

FPLC – Fast Protein Liquid Chromatography

HCAR – 7-Hydroxylmethyl Chlorophyll(ide) *a* Reductase

HILIC – Hydrophobic Interaction Iiquid Chromatography

HPLC – High Performance Liquid Chromatography

IPTG – isopropyl-β-D-1-thiogalactopyranoside

LB – Luria Broth

LC-MS – Liquid Chromatography-Mass spectrometry

MBP – Maltose Binding Protein

*Mp*CAO – *Micromonas pusilla* Chlorophyll(ide) *a* Oxygenase

MS – Mass spectrometry

MS/MS – Tandem mass spectrometry

NAD(P) – Nicotinamide Adenine Dinucleotide (Phosphate)

NAD(P)H – Nicotinamide Adenine Dinucleotide (Phosphate) Hydride

neoSTX – neosaxitoxin

PCR – Polymerase Chain Reaction

Pchlide – Protochlorophyllide

PDB – Protein Data Bank

*Ph*CAO – *Prochlorothrix hollandica* Chlorophyll(ide) *a* Oxygenase

PST – Paralytic Shellfish Toxin

Q-ToF – Quadrupole-Time of Flight

RO – Rieske Oxygenase

SDS-PAGE – Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

STX – saxitoxin

TB – Terrific Broth

TEV – tobacco etch virus

ToF – Time of Flight

UV-Vis – Ultraviolet-Visible spectroscopy

VGSC – Voltage Gated Sodium Channel

α/β-STOH – α/β-saxitoxinol

βME – 2-Mercaptoethanol

7-OH-Chl(ide) *a* – 7-hydroxymethyl Chlorophyll(ide) *a*

Abstract

Nature has evolved a striking array of enzymes that perform precise C-H bond functionalization reactions. Rieske oxygenases represent one of the most prevalent and underexploited enzyme classes that falls in this category. Rieske oxygenases are enzymes that couple a 2-His/2-Cys ligated [2Fe-2S] cluster, or Rieske cluster, with a mononuclear iron site. The Rieske cluster accepts electrons and shuttles them to a mononuclear iron site to catalyze the desired reaction. Some enzymes in this class function as dioxygenases and incorporate both oxygen atoms of O_2 into a product, whereas others function as monooxygenases and insert only one oxygen atom into a product. Remarkably less well studied, there is even one enzyme in this class that is proposed to convert a methyl group into a formyl group via sequential mono-oxygenation reactions. Despite the ubiquity of these enzymes in Nature and their promise for producing pharmaceuticals, commodity chemicals, and facilitating bioremediation efforts, there is still a critical gap in knowledge with regards to the structure-function relationships in this class of enzymes, leaving many open questions about how these enzymes catalyze reactions selectively or catalyze reactions with different outcomes.

Saxitoxin and Chlorophyll are two natural products that are biosynthesized, in part, by Rieske oxygenases. Saxitoxin is a well-known paralytic shellfish toxin and Chlorophyll is one of the most significant pigments in photosynthesis. Two Rieske non-heme iron oxygenases, SxtT and GxtA, involved in saxitoxin biosynthesis are discussed in chapters 2 and 3. These two Rieske oxygenases share 88-pecent sequence identity with one another but catalyze hydroxylation reactions on the saxitoxin scaffold at different positions. Using X-ray crystallography, rigorous

structural analysis, and biochemical studies, we revealed the design principle for site-selective hydroxylation by SxtT and GxtA. Through analysis of the structurally available enzymes, we further showed that these design principles may be universally conserved in the Rieske oxygenase class.

Another interesting Rieske oxygenase studied here is involved in Chlorophyll biosynthesis and is discussed in chapter 4. This enzyme, Chlorophyll(ide) *a* oxygenase (CAO), is a key player in the Chlorophyll cycle and is proposed to transform Chlorophyll(ide) *a* into Chlorophyll(ide) *b* by catalyzing two sequential monooxygenation reactions that convert a C7-methyl group into a C7-formyl group. In this work, using a bottom-up approach, we overproduced and reconstituted the activity of CAO from four different organisms from different kingdoms of life. Importantly, this work demonstrates that CAO converts a Chlorophyll precursor, Chlorophyllide *a*, into Chlorophyllide *b in vitro*. We established the existence of a monooxygenated intermediate, supporting the proposal that CAO is a Rieske oxygenase which catalyzes two iterative monooxygenation reactions. We revealed key details about the stereoselectivity of this reaction and the substrate scope of CAO. Finally, through the enzymatic synthesis of an unnatural Chlorophyll pigment, we demonstrated the potential for using CAO as an enzymatic tool for synthesizing custom-tuned chlorophyll pigments. Taken together, this research adds to our fundamental understanding of how saxitoxin and Chlorophyll are biosynthesized. In addition, these studies reveal predictive power for thinking about how to control selectivity and redesign reactivity of Rieske oxygenases.

Chapter 1 Introduction

1.1 Overview of Rieske non-heme iron oxygenase enzyme superfamily

Biocatalysts that perform precise C-H bond functionalization have received attention for their potential use in building and degrading complex natural products[1-3]. The class of Rieske non-heme iron oxygenase enzymes or Rieske oxygenases, represents one of the most pervasive and underexploited enzyme classes in this context. Like other iron dependent oxygenases such as P450 and α-KG-dependent iron oxygenases, Rieske oxygenases widely exist in nature and utilize iron ion as the cofactor to catalyze a diverse set of reactions in both anabolism and catabolism. However, unlike these other iron-dependent enzymes, Rieske oxygenases are defined by the presence of a Rieske cluster and a mononuclear iron site. The Rieske cluster is a [2Fe-2S] cluster that is ligated by two Cys and two His residues that are found in a Cys-X-His-X_{17}-Cys-X_2-His sequence motif. This type of Fe-S cluster was first described in complex III of the electron transport chain by John Rieske in 1964 and was subsequently named[4]. The mononuclear iron site is ligated by two His and one Asp or Glu residue, or the so-called facial triad motif. Based on the presence of these characteristic sequence signatures, current estimates suggest that there are more than 80,000 annotated sequences that belong in this enzyme class. However, despite this large number, only a small number of these enzymes have been mechanistically studied and even less have been structurally characterized, as represented by the lack of deposited structures in the Protein Data Bank (PDB)[5]. To date, there are only 17 available structures of different Rieske oxygenases in PDB, nine of which are monooxygenases (SxtT, GxtA, OMO, DdmC, NdmA, NdmB, CntA, Stc2 and KshA) and eight of which are dioxygnenases (BPDO, NDO, PDO, CARDO, NBDO, TDO, NTDO and CumDO).

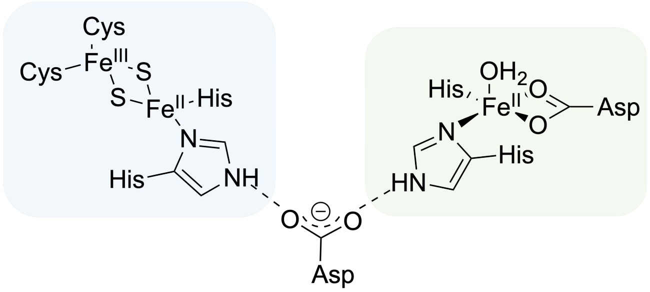


Figure 1.1. The configuration of the metalloclusters in Rieske oxygenases.
The [2Fe-2S] Rieske cluster is ligated by two-Cys and two-His residues whereas the non-heme iron is ligated by two-His and one-Asp residues. Additionally, a conserved Asp or Glu residue connects these two metallocenters as an electron bridge.

1.1.1 The metalloclusters found in Rieske oxygenases

The presence of two characteristic metallocenters, the Rieske cluster and the non-heme iron site, are defining features of a Rieske oxygenase. The Rieske cluster, or the [2Fe-2S] cluster is ligated by two Cys and two His residues. It displays a redox potential in a wide range of -150 mV to +400 mV, which is elevated relative to the four-Cys ligated [2Fe-2S] clusters[6]. The Rieske cluster also has distinct electron paramagnetic resonance (EPR) spectroscopy and ultraviolet-visible (UV-Vis) spectroscopy signal under different oxidation states. The oxidized $[2Fe\text{-}2S]^{2+}$ state is EPR silent but displays a absorption peak around 425 nm on UV-Vis spectroscopy, whereas the reduced form $[2Fe\text{-}2S]^{+}$ is UV-Vis silent but displays characteristic g values at 2.01, 1.91 and 1.45 on EPR spectroscopy[7-8]. As for the mononuclear non-heme iron site, since it lacks a chromophore and is EPR silent in the reduced state, the typical methods for investigating its geometric structure and electronic properties for mechanism study include using magnetic circular dichroism (MCD), electron nuclear double resonance (ENDOR), or using the O_2 mimic nitric oxide to convert the non-heme iron site into an EPR-active form[9].

1.1.2 The electron transfer pathway and proposed mechanisms for Rieske oxygenases

The mechanism of Rieske oxygenase catalyzed dihydroxylation and monohydroxylation reaction on aromatic substrates has been well studied[8, 10-11]. The source of the needed two electrons for Rieske oxygenase chemistry is known to originate from reduced nicotinamide adenine dinucleotide (NADH) and be transmitted to a reductase partner carrying a flavin cofactor that can mediate the conversion of two electrons into one electron events. The single electron is delivered to a [2Fe-2S] cluster and finally transferred to the Rieske cluster of the Rieske oxygenase. This electron is ultimately shuttled from the Rieske cluster through the conserved bridging Asp or Glu residue to the mononuclear iron site to facilitate the catalytic cycle (Figure 1.2.a).

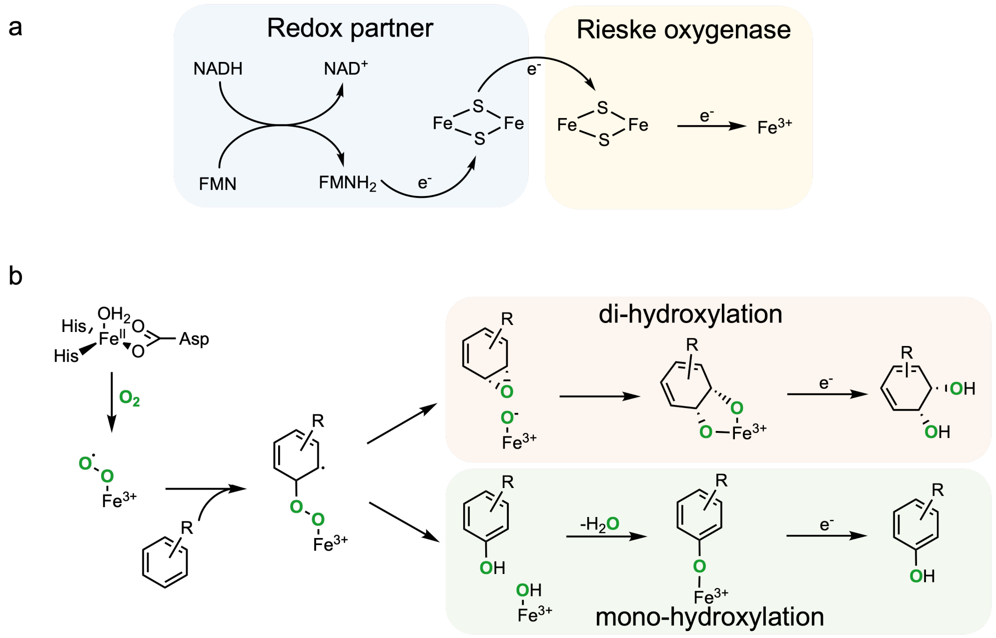


Figure 1.2. Rieske oxygenases use a common strategy to transfer electrons and may use different iron-based oxidants to facilitate their diverse catalytic functions.

(a) The needed electrons for the reaction are supplied by NADH through a flavin/iron-sulfur cluster dependent reductase partner to the Rieske cluster and ultimately to the non-heme iron site. (b) Rieske oxygenases use a common set of metallocenters to catalyze dihydroxylation and monohydroxylation. The Fe(III)-superoxo species attacks the aromatic substrate to form a C-O bond a substrate radical. It then either incorporate the second oxygen atom into the substrate to produce the di-hydroxylated product or undergo O-O cleavage to produce the mono-hydroxylated product.

The binding of O_2 to mononuclear Fe(II) was formerly considered to result in a quick transfer of one electron from the Rieske cluster, resulting in formation of either an Fe(III) hydroperoxide or an electrically equivalent species (Fe(III)-peroxo or Fe(V)-oxo-hydroxo) that may react with the substrate[10]. However, recent studies have demonstrated that the initial reaction of an activated O_2 species with the substrate arise before electron transfer from the Rieske cluster, indicating that an Fe(III)-superoxo moiety is a more plausible description for the active iron-oxygen specie[8]. In general, the active iron specie attacks the carbon atom directly, generating an oxygen-carbon bond as well as a radical on the substrate. With an electron donated from the Rieske cluster, the iron-oxygen specie either incorporates a second oxygen atom onto the substrate, forming a di-hydroxylated product, or undergoes O-O bond cleavage, resulting a mono hydroxylated product[8] (figure 1.2.b).

1.1.3 The quaternary architectures of Rieske oxygenases

Aside from sequence fingerprints, Rieske oxygenases, to date, have shown similar quaternary architectures: these enzymes are known to exist with either a homo-trimeric (α_3) or a hetero-hexameric ($\alpha_3\beta_3$) structure (Figure 1.1). In both cases, the Rieske cluster is located at the N-terminal domain of the catalytic α subunit whereas the mononuclear iron site is located nearly 45 Å away in the C-terminal domain[9, 12-13]. More importantly, in the trimeric α_3 or $\alpha_3\beta_3$ architecture, these two metallocenters end up located across a subunit-subunit interface approximately 12 Å apart[9, 12-13]. This arrangement is significant as it permits electrons to be transferred between the Rieske cluster and the mononuclear iron site, using a proposed bridging Asp or Glu residue[12-13].

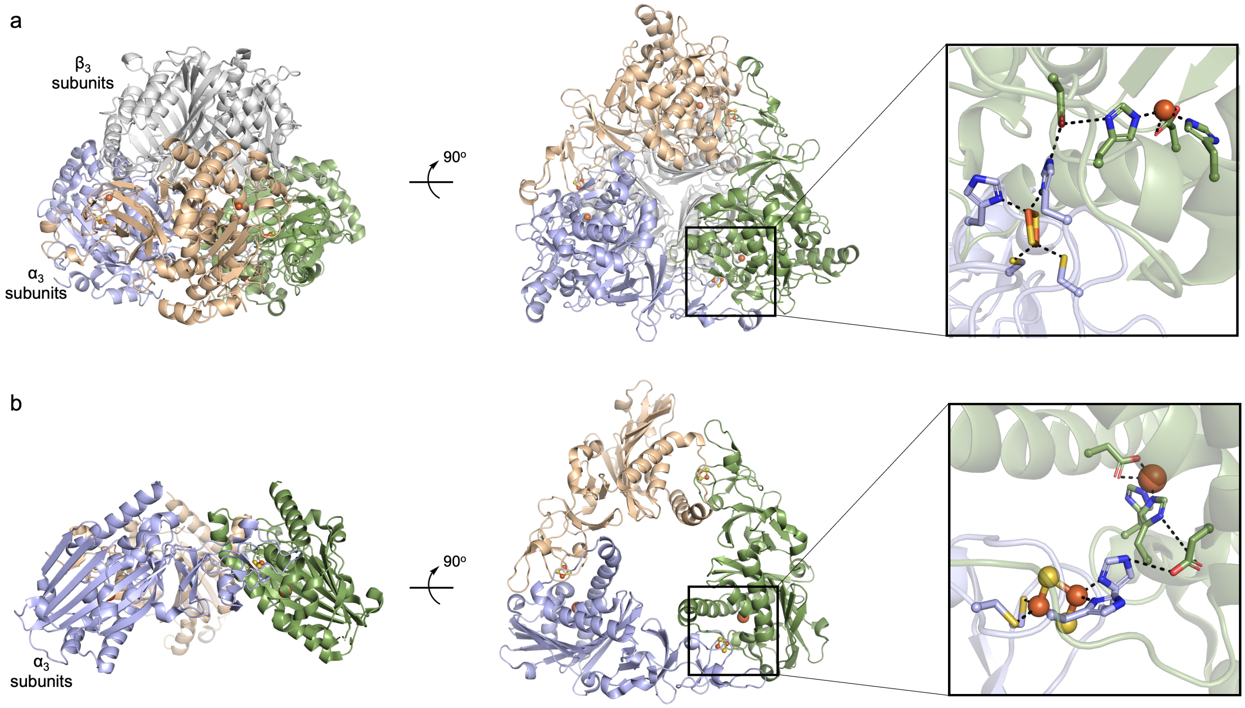


Figure 1.3. Structural features of Rieske non-heme iron oxygenases.
(a) The first structurally characterized Rieske non-heme iron oxygenase, Naphthalene 1,2-dioxygenase (NDO), showcases a hetero-hexameric ($\alpha_3\beta_3$) quaternary structure. The α_3 subunits are labeled in color whereas the β_3 subunits are gray. In each α subunit, there is a [2Fe-2S] Rieske cluster ligated by 2 Cys and 2 His residues in the C-terminal domain as well as a non-heme iron bound by 2 His and 1 Asp residues in the N-terminal domain. These is also a bridging Asp residue that connects these two metallocentres. (PDB: 1NDO). (b) Dicamba monooxygenase (DdmC), on the other hand, displays a homo-trimeric (α_3) architecture, but still contains the characteristic binding motifs for the Rieske cluster and the non heme iron site, as well as the bridging Asp residue. (PDB: 3GKE).

Despite the putative significance of the canonical trimeric structure to the catalytic activity of a Rieske oxygenase, recent studies have suggested that alternative quaternary architectures might exist as well. For instance, a Rieske oxygenase from *Micromonas pusilla*, is encoded by two genes *Mp*CAO1 and *Mp*CAO2. These genes are predicted to encode two proteins, one of which individually binds a Rieske cluster and the second of which binds the non-heme iron site[14]. These proteins are predicted to form a heterodimer that is involved in Chlorophyll *b* biosynthesis and are discussed in chapter 4. Another example of a Rieske oxygenase that is proposed to deviate from the traditional trimeric architecture comes from the Zhao group who has shown that the Rieske oxygenase PrnD, involved in aminopyrrolnitrin biosynthesis displays a homodimeric architecture[15]. Interestingly, a separate study from Brooks' laboratory suggested that two caffeine demethylating Rieske oxygenases, NdmA and NdmB, instead adopt a heterohexameric architecture in solution[16]. Finally, phthalate dioxygenase was first proposed to adopt a tetrameric (α_4) architecture and was later revealed to function as a homohexamer[17-18]. These alternative architectures suggest an extra level of structural diversity exists in this enzyme superfamily and may have important impacts on the activity of these enzymes.

1.2 Rieske oxygenases in catabolism

Rieske oxygenases are relatively well-known for the roles that they play in the degradation of small aromatic pollutants. These degradative enzymes are predominantly predicted to function using a dearomatization reaction that requires the insertion of two oxygen atoms from molecular oxygen into the substrate[11,19-23]. One such protein is Naphthalene 1,2-dioxygenase (NDO) from *Pseudomonas sp*. NCIB 9816-4. This protein was the first structurally characterized Rieske non-heme iron oxygenase[11,20]. NDO catalyzes the incorporation of two oxygen atoms from molecular oxygen into naphthalene to form (1*R*,2*S*)-dihydroxylnaphthalene (Figure 1.2.a). Several other Rieske oxygenase catalyzed reactions proceed similarly. Nitrobenzene dioxygenase (NBDO) from *Comamonas sp*. strain JS739065 catabolizes nitrobenzene to 1,2-dihydroxybenzene (catechol)[20], 3-nitro-toluene dioxygenase (NTDO) from *Diaphorobacter* generates 1,2-dihydroxytoluene (methyl-catechol)[21], and toluene dioxygenase (TDO) from *Pseudomonas putida* produces a similar dearomatization product[22] (Figure 1.2.a). Cumene dioxygenase (CumDO) from *Pseudomonas fluorescens* IP01, on the other hand, catalyzes cis-dihydroxylation in isopropylbenzyne degradation[23] (Figure 1.2.a).

In addition to dioxygenation reactions, Rieske oxygenases also catalyze monooxygenation reactions. They form C-C bonds and facilitate the cleavage of *O*-methyl and *N*-methyl bonds[8,16,24-28]. 2-oxoquinoline-8-monooxygenase (OMO), for instance, is involved in the second step of the catabolic pathway that degrades the *N*-heterocyclic quinolone by installing a hydroxyl group at the benzene ring of 2-oxoquinoline[24] (Figure 1.2.b). A similar reaction is catalyzed by Salicylate-5-hydroxylase (S5H), which converts salicylate into gentisate by adding a hydroxyl group at the C5 position of the aromatic ring[8] (Figure 1.2.b). 4-toluene sulfonate methyl monooxygenase (TsaM), on the other hand, catalyzes exocyclic monooxygenation through hydroxylation of a primary benzyl carbon atom[25] (Figure 1.2.b). Additionally, in the oxidative catabolism of caffeine and dicamba, Rieske oxygenases catalyzed monooxygenation reactions have also been reported with subsequent *N*-methyl or *O*-methyl bond cleavage[16,26] (Figure 1.2.c). Meanwhile, Pheophorbide *a* oxygenase (PAO), a Rieske oxygenase involved in Chlorophyll catabolism, is proposed to open the chlorin macrocycle by cleaving a C-C bond[27]. Another annotated Rieske oxygenase, TIC55, is proposed to then install a hydroxyl group on the C8-ethyl group of the cleavage product by PAO[28] (Figure 1.2.c)

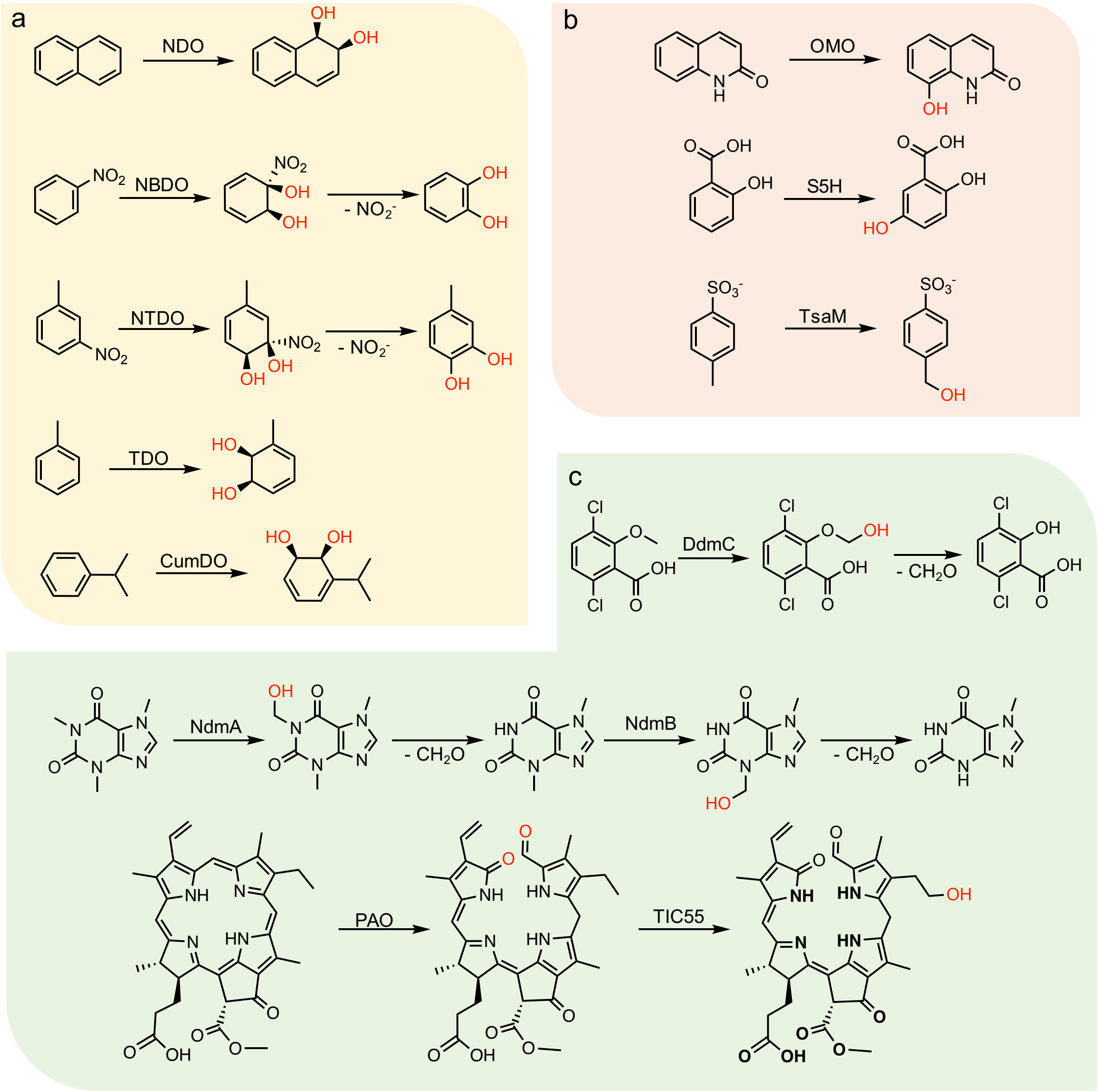


Figure 1.4. Reaction schemes for Rieske oxygenases in catabolic pathways.

(a) Rieske oxygenases catalyze dihydroxylation reactions. Shown in this panel are the reactions catalyzed by naphthalene 1,2-dioxygenase (NDO), nitrobenzene dioxygenase (NBDO), 3-nitrotoluene dioxygenase, toluene dioxygenase (TDO) and cumene dioxygenase (CumDO). NDO catalyzes the dihydroxylation to convert naphthalene to (1*R*,2*S*)-dihydroxylnaphthalene. NBDO and NTDO catalyze dihydroxylation on nitrobenzene and nitrotoluene to convert them to catechol and methyl-catechol, respectively. TDO and CumDO catalyze dihydroxylation on toluene and cumene to transform them to 3-methylcyclohexa-3,5-diene-1,2-diol and 3-isopropylcyclohexa-3,5-diene-1,2-diol, respectively. (b) Rieske oxygenases catalyze monohydroxylation reactions. Shown in this panel are the reactions catalyzed by 2-oxoquinoline-8-monooxygenase (OMO), salicylate-5-hydroxylase (S5H) and toluene sulfonate methyl monooxygenase (TsaM). OMO catalyzes the monohydroxylation at C8 position of 2-oxoquinoline. S5H catalyzes the monohydroxylation at C5 position of salicylate. TsaM catalyzes monohydroxylation at benzyl position of toluene sulfonate. (c) Rieske oxygenases catalyze *N*-methyl, *O*-methyl or C-C bond cleavage reaction. Shown in this panel are the reactions catalyzed by dicamba monooxygenase (DdmC), caffeine *N*-demethylase (NdmA and NdmB) and Pheophorbide *a* oxygenase (PAO). DdmC catalyzes *O*-demethylation of dicamba whereas NdmA and NdmB catalyze *N*-demethylation of caffeine. PAO is proposed to catalyze the ring opening reaction of pheophorbide *a* through a C-C bond cleavage reaction. Another Rieske oxygenase TIC55 is proposed to catalyze a following up hydroxylation reaction at C8 position on the ring opening compound.

1.3 Rieske oxygenases in anabolism

The function of Rieske oxygenases is not limited to catabolic processes. In fact, Rieske oxygenases are also implicated in essential biosynthetic pathways that produce a variety of natural products and secondary metabolites. For example, recent studies have shown that Rieske oxygenases catalyze C-C bond desaturation, C-C bond carbocyclization and *N*-oxidation reactions[29]. This dissertation focusses on three such anabolic Rieske non-heme iron oxygenases that catalyze C-H bond hydroxylation reactions in natural product biosynthesis. Specifically, in chapters 2 and 3, I will highlight two Rieske oxygenases, SxtT and GxtA, that are found in the biosynthetic pathway of paralytic shellfish toxins in late-stage hydroxylation reactions on tricyclic guanidinium scaffolds[30]. Another intriguing Rieske oxygenase covered in chapter 4 of this dissertation is Chlorophyll(ide) *a* Oxygenase (CAO), which is involved in the biosynthesis of Chlorophyll(ide) *b*. This Rieske oxygenase, is proposed to perform an exocyclic oxidation reaction to generate a product that is two oxidation states higher than the starting substrate, namely the oxidation of a methyl group into a formyl group[31]. This transformation, unlike for the Rieske oxygenases described above, is proposed to involve two sequential monooxygenation reactions thus transforming the C7-methyl group of Chlorophyll(ide) *a* into a hydroxymethyl group, a dihydroxymethyl functionality, and finally, a formyl group.

1.3.1 Rieske oxygenase catalyzed monohydroxylation in paralytic shellfish toxins (PST) biosynthesis

Saxitoxin and their sulfate-containing derivatives, gonyautoxins, belong to a large class of molecules collectively referred to as paralytic shellfish toxins (PST). The term saxitoxin is often used to refer to the entire suite of more than 50 structurally related neurotoxins[32-33]. These molecules each share a common 5-5-6 tricyclic bisguanidinium ion-containing core with multiple hydrogen bond donors and acceptors on the scaffold. Due to these structural features, saxitoxin has been found to bind voltage gated sodium channels (VGSC) and inhibit the signal transduction.

PSTs were first identified in the Alaskan butter clam *Saxidomus giganteus*, as a bioaccumulated product from cyanobacteria and dinoflagellates[34]. Several proposed trajectories for biosynthesizing PSTs have been reported[35]. One such pathway starts with arginine ethyl ketone. An oxidative cyclization reaction is proposed to transform this molecule to contain a tricyclic scaffold. A subsequent exocyclic carbamylation on the side chain leads to formation of the ultimate product, saxitoxin, after a late-stage hydroxylation reaction (Figure. 1.3).

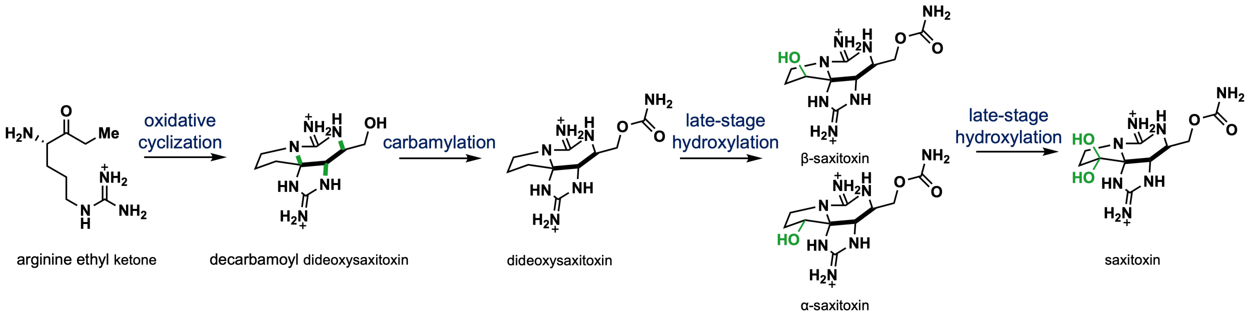
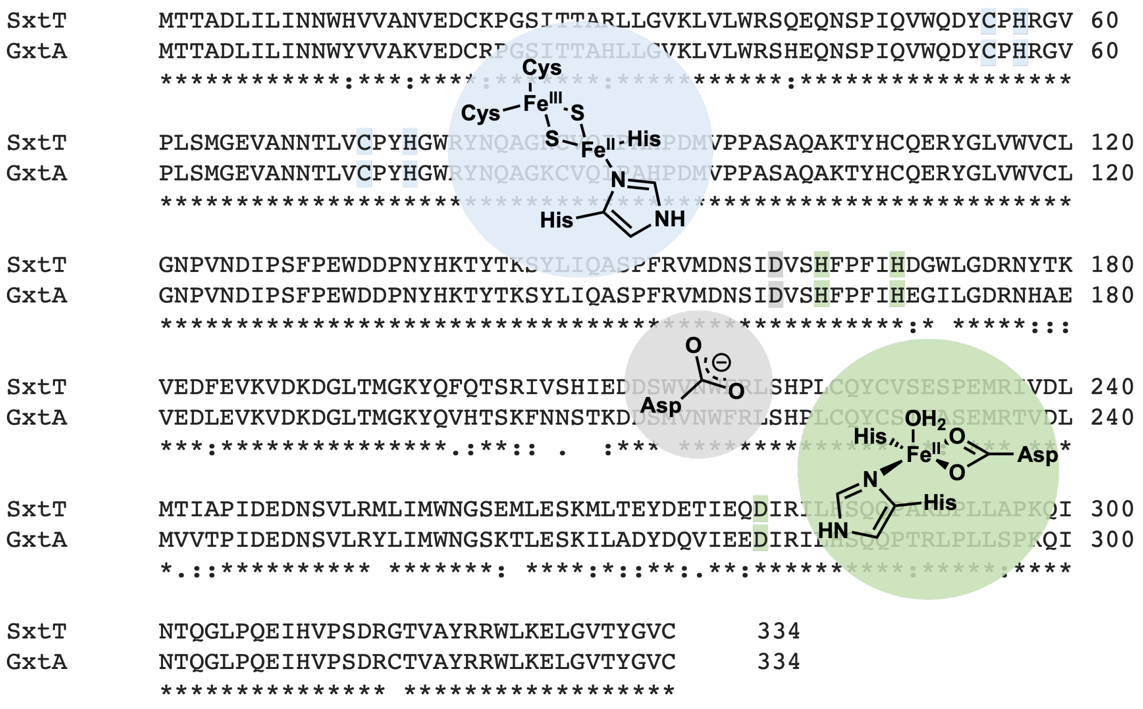


Figure 1.5. Proposed biosynthetic pathway for saxitoxin.
The proposed pathway starts with arginine ethyl ketone, which gets converted to decarbamoyl dideoxysaxitoxin via oxidative cyclization for form the tricyclic scaffold. Then, decarbamoyl dideoxysaxitoxin is converted to dideoxysaxitoxin via the exocyclic installation of a carbamoyl group. Lastly, dideoxysaxitoxin is transformed to saxitoxin by two late-stage hydroxylation steps at C12 position.

Previous studies by the Narayan group, also from the University of Michigan, resulted in the identification of three Rieske oxygenases, SxtT, GxtA and SxtH, that are all involved in saxitoxin biosynthesis[30]. Each of these enzymes is proposed to catalyze a monohydroxylation reaction in the biosynthetic pathway. Specifically, SxtT and GxtA have been shown to perform late-stage monohydroxylation reactions, whereas the substrate and activity of SxtH remains enigmatic. Impressively, a sequence alignment of SxtT and GxtA, reveals that these two Rieske oxygenases share 88-percent sequence identity and 92-percent sequence similarity with one another (Figure 1.4). Enzyme activity assays elucidated that SxtT selectively catalyzed the monohydroxylation at C12 position whereas GxtA prefers to hydroxylate the C11 position of the tricyclic scaffold (Figure 1.5). To investigate what structural aspects of these enzymes govern the selectivity of SxtT and GxtA reactions, in chapter 2, we first solved the crystal structure of these enzymes with and without substrate analogs bound. This work allowed identification of two critical residues in the active site for selectivity. However, work performed in collaboration with the Narayan laboratory showed that the substrate specificity of SxtT and GxtA active site variants



```
SxtT      MTTADLILINNWHVVANVEDCKPGSITTARLLGVKLVLWRSQEQNSPIQVWQDYCPHRGV  60
GxtA      MTTADLILINNWYVVAKVEDCRPGSITTAHLLGVKLVLWRSHEQNSPIQVWQDYCPHRGV  60
          ************:***:****:*******:***********:******************

SxtT      PLSMGEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCL  120
GxtA      PLSMGEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCL  120
          ************************************************************

SxtT      GNPVNDIPSFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHDGWLGDRNYTK  180
GxtA      GNPVNDIPSFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHEGILGDRNHAE  180
          ************************************************:* *****:::

SxtT      VEDFEVKVDKDGLTMGKYQFQTSRIVSHIEDDSWVNWFRLSHPLCQYCVSESPEMRIVDL  240
GxtA      VEDLEVKVDKDGLTMGKYQVHTSKFNNSTKDDSMVNWFRLSHPLCQYCSVNASEMRTVDL  240
          ***:***************.:**:: .  :*** *****************: *** ***

SxtT      MTIAPIDEDNSVLRMLIMWNGSEMLESKMLTEYDETIEQDIRILHSQQPARLPLLAPKQI  300
GxtA      MVVTPIDEDNSVLRYLIMWNGSKTLESKILADYDQVIEEDIRILHSQQPTRLPLLSPKQI  300
          *.::********** *******: ****:*::**:.**:**********:*****.****

SxtT      NTQGLPQEIHVPSDRGTVAYRRWLKELGVTYGVC      334
GxtA      NTQGLPQEIHVPSDRCTVAYRRWLKELGVTYGVC      334
          *************** ******************
```

Figure 1.6. Sequence alignment of the Rieske Oxygenases SxtT and GxtA shows they share 88-percent sequence identity with each other.

Both SxtT and GxtA contain the characteristic binding motif for Rieske oxygenase, including the 2Cyc- 2His sequence signature for the [2Fe-2S] Rieske cluster, the 2His-1Asp for non-heme iron site as well as the bridging Asp, which are highlighted in blue, green and gray, respectively were incompletely switched. Further structural analysis suggested that a substrate tunnel and a dynamic loop gating the entrance to the active site could be the additional structural features contributing to selectivity. Therefore, in chapter 3, we confirmed the existence of a different substrate channel in SxtT and GxtA by using both computational calculations and Xe-pressurization experiments. We also observed the different conformations of the dynamic loop that gate the substrate channel in the native substrate bound crystal structures of these two enzymes. Rational mutations made in the dynamic loop, active site and substrate tunnel, were shown to successfully invert the site selectivity of SxtT into GxtA and therefore reveal the design principle of these two Rieske oxygenases.

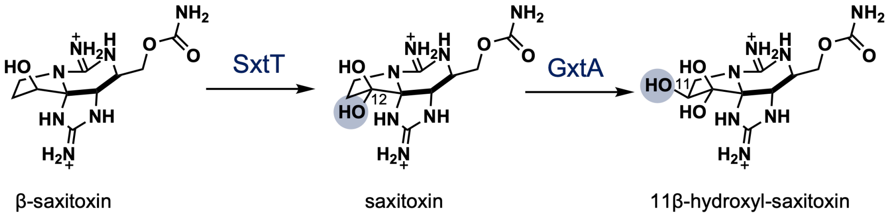
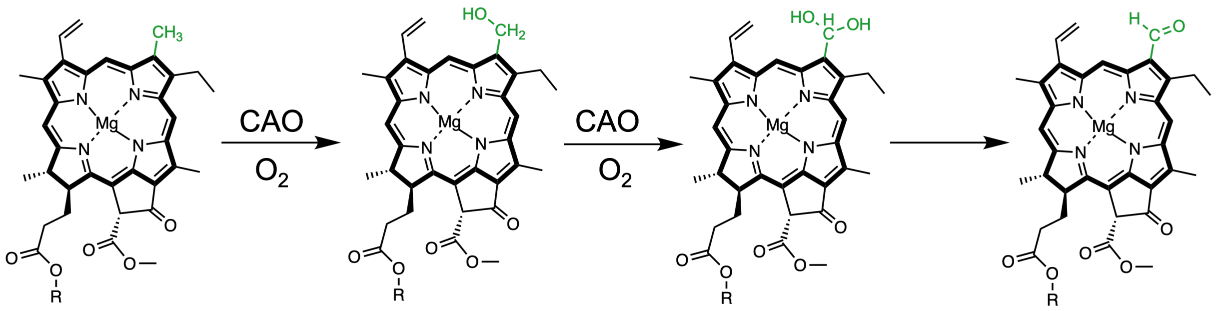


Figure 1.7. Reaction schemes of SxtT and GxtA.
SxtT selectively hydroxylates the C12 position of β-saxitoxinol whereas GxtA prefers to catalyze hydroxylation at C11 position of saxitoxin.

1.3.2 Rieske oxygenase catalyzed sequential hydroxylation in Chlorophyll biosynthesis

Photosynthesis is one of the most essential biological processes on the planet Earth. A variety of pigments are involved in this process to absorb light and capture light energy to facilitate various biochemical processes. Incredibly, photosynthesis has been found to occur under different light and oxygen conditions, presumably because it has been fine-tuned to do so through evolution. To attain this adaptability, photosynthetic organisms modify their pigment scaffolds to absorb available ambient light[36-40].

Among all the pigments involved in photosynthesis, Chlorophyll (Chl) is one of the most important compounds. Chlorophyll widely exists on our green Earth, and is found in plants, green algae and cyanobacteria. In these organisms, Chlorophyll serves to absorb light energy and help facilitate its conversion into chemical energy[36-39]. Notably, among the known Chlorophyll derivatives, Chl *a* and Chl *b* are the two most abundant photosynthetic pigments found in Nature. Chlorophyll *a* has a methyl group in the C7 position, whereas Chlorophyll *b* has a formyl group in the same position (Figure 1.6). Chlorophyll *a* and Chlorophyll *b* have distinct absorption spectra in the blue and red regions, which enable this combination of pigments to absorb wider ranges of light spectra[38]. Recent studies have demonstrated that the light harvesting complex (LHC) of plants and green algae utilized both Chl *a* and Chl *b* to absorb one photon and generate an excited electron[41]. However, the excited electron generated by Chl *a* can be directly passed to downstream cofactors and proteins through electron transport chain leading to the ultimate reduction of NADP to NADPH, but the excited electron generated by Chl *b* has to be passed to Chl *a* and then to the electron transport chain. As a result, photosynthetic organisms have developed a strategy for surviving under different light condition by optimal utilization of available light via tunning the ratio of Chl *a* and Chl *b*.



R=H, Chlorophyll *a*
R=$C_{20}H_{39}$, Chlorophyllide *a*

R=H, 7-OH-Chlorophyll *a*
R=$C_{20}H_{39}$, 7-OH-Chlorophyllide *a*

R=H, Chlorophyll *b*
R=$C_{20}H_{39}$, Chlorophyllide *b*

Figure 1.8. Proposed reaction for the CAO catalyzed transformation of Chlorophyll(ide) *a* to Chlorophyll(ide) *b* with molecular oxygen.

Formation of the C7 formyl group in Chlorophyll(ide) *b* is proposed to proceed via two sequential reactions that transform the C7-methyl group of Chlorophyll(ide) *a* into the formyl group of Chlorophyll(ide) *b* through a C7-hydroxymethyl intermediate (7-OH- Chlorophyll(ide) *a*) and a potential C7-dihydroxymethyl intermediate.

The biosynthetic route between Chl *a* and Chl *b* is referred to as the Chlorophyll cycle[42]. In this cycle, Chlorophyll(ide) *a* oxygenase (CAO), an annotated Rieske non-heme iron oxygenase, is a key player that is proposed to catalyze two sequential hydroxylation reactions to convert Chlorophyll(ide) *a* into Chlorophyll(ide) *b* (Figure 1.6). However, despite the identification of CAO more than 20 years ago[31], it has never been studied *in vitro* with purified protein, leaving many open questions about whether CAO can facilitate both oxygenation reactions in the absence of an additional cofactor or protein. In chapter 4, we demonstrate that CAO converts a Chlorophyll precursor, Chlorophyllide *a*, into Chlorophyllide *b* using high quality *in vitro* LC-MS and MS/MS data. Further analysis also highlighted the stereospecificity of the reaction and revealed the potential of CAO as an enzymatic tool for custom-tuned natural and unnatural Chlorophyll pigment synthesis.

1.4 References

1 Peters, C. & Buller, R. M. Industrial application of 2-oxoglutarate-dependent oxygenases. *Catalysts* 9, 20 (**2019**).

2 King-Smith, E., Zwick, C. R. & Renata, H. Applications of oxygenases in the chemoenzymatic total synthesis of complex natural products. *Biochemistry* 57, 403-412 (**2018**).

3 Roiban, G. D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* 51, 2208-2224 (**2015**).

4 Rieske, J.S., Maclennan, D.H., Coleman, R. Isolation and properties of iron-protein from (reduced coenzyme Q)-cytochrome c reductase complex of respiratory chain. *Biochem. Biophys. Res. Commun*. 15, 338 (**1964**)

5 Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res*. 28, 235-242 (**2000**).

6 Brown, E. N. et al. Determining Rieske cluster reduction potentials. *Journal of Biological Inorganic Chemistry*. 13 1301 (**2008**)

7 Liu, J., Chakraborty, S., Hosseinzadeh, P., et al. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev*. 114, 4366–4469. (**2014**)

8 Rogers, M.S., Lipscomb, J.D. Salicylate 5-Hydroxylase: Intermediates in Aromatic Hydroxylation by a Rieske Monooxygenase. *Biochemistry* 58, 5305-5319 (**2019**)

9 Barry, S.M., Challis, G.L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal*. 3, 2362–2370 (**2013**)

10 Wolfe, M. D., Parales, J. V., Gibson, D. T., and Lipscomb, J. D. Single turnover chemistry and regulation of O_2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J. Biol. Chem*. 276, 1945−1953. (**2001**)

11 Karlsson, A., Parales, J. V., Parales, R. E., Gibson, D. T., Eklund, H., and Ramaswamy, S. Crystal structure of naphthalene dioxygenase: Side-on binding of dioxygen to iron. *Science* 299, 1039−1042. (**2003**)

12 Capyk, J.K., D'angelo, I., Strynadka, N.C., Eltis, L.D. Characterization of 3-ketosteroid 9{alpha}-hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of Mycobacterium tuberculosis. *J. Biol. Chem*. 284, 9937–9946 (**2009**)

13 Ferraro, D.J., Gakhar, L., Ramaswamy, S. Rieske business: Structure-function of Rieske non-heme oxygenases. *Biochem. Biophys. Res. Commun*. 338, 175–190 (**2005**)

14 Kunugi, M., Takabayashi, A., Tanaka, A. Evolutionary changes in chlorophyllide a oxygenase (CAO) structure contribute to the acquisition of a new light-harvesting complex in micromonas. *J. Biol. Chem*. 288, 19330–19341 (**2013**)

15 Lee, J., Zhao, H. Mechanistic studies on the conversion of arylamines into arylnitro compounds by aminopyrrolnitrin oxygenase: Identification of intermediates and kinetic studies. *Angew. Chem. Int. Ed. Engl*. 45, 622–625 (**2006**)

16 Kim, J.H., Kim, B.H., Brooks, S., et al. Structural and mechanistic insights into caffeine degradation by the bacterial N-demethylase complex. *J. Mol. Biol*. 431, 3647–3661 (**2019**)

17 Batie, C.J., Lahaie, E., Ballou, D.P. Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from Pseudomonas cepacia. *J. Biol.Chem*. 262, 1510–1518 (**1987**)

18 Tarasev, M., Pinto, A., Kim, D., Elliott, S.J., Ballou, D.P. The "bridging" aspartate 178 in phthalate dioxygenase facilitates interactions between the Rieske center and the iron(II)–mononuclear center. *Biochemistry* 45, 10208–10216 (**2006**)

19 Kauppi, B., Lee, K., Carredano, E., et al. Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Structure* 6, 571–586 (**1998**)

20 Friemann, R., Ivkovic-Jensen, M.M., Lessner, D.J., et al. Structural insight into the dioxygenation of nitroarene compounds: The crystal structure of nitrobenzene dioxygenase. *J. Mol. Biol.* 348, 1139–1151 (**2005**)

21 Kumari, A., Singh, D., Ramaswamy, S., Ramanathan, G. Structural and functional studies of ferredoxin and oxygenase components of 3-nitrotoluene dioxygenase from Diaphorobacter sp. strain DS2. *PLoS One* 12, e0176398 (**2017**)

22 Friemann, R., Lee, K., Brown, E.N., et al. Structures of the multicomponent Rieske non-heme iron toluene 2,3-dioxygenase enzyme system. *Acta Crystallogr. D Biol.Crystallogr*. 65, 24–33 (**2009**)

23 Dong, X., Fushinobu, S., Fukuda, E., et al. Crystal structure of the terminal oxygenase component of cumene dioxygenase from Pseudomonas fluorescens IP01. *J. Bacteriol*. 187, 2483–2490 (**2005**)

24 Martins, B.M., Svetlitchnaia, T., Dobbek, H. 2-Oxoquinoline 8-monooxygenase oxygenase component: Active site modulation by Rieske-[2Fe-2S] center oxidation/reduction. *Structure* 13, 817–824 (**2005**)

25 Locher, H.H., Leisinger, T., Cook, A.M. 4-Toluene sulfonate methyl-monooxygenase from Comamonas testosteroni T-2: Purification and some properties of the oxygenase component. *J. Bacteriol.* 173, 3741–3748 (**1991**)

26 Dumitru, R., Jiang, W.Z., Weeks, D.P., Wilson, M.A. Crystal structure of dicamba monooxygenase: A Rieske nonheme oxygenase that catalyzes oxidative demethylation. *J. Mol. Biol.* 392, 498–510 (**2009**)

27 Pruzinska, A., Tanner, G., Anders, I., Roca, M., Hortensteiner, S. Chlorophyll breakdown: Pheophorbide a oxygenase is a Rieske-type iron-sulfur protein, encoded by the accelerated cell death 1 gene. *Proc. Natl. Acad. Sci. USA* 100, 15259–15264 (**2003**)

28 Hauenstein, M., Christ, B., Das, A., Aubry, S., Hortensteiner, S. A role for TIC55 as a hydroxylase of phyllobilins, the products of chlorophyll breakdown during plant senescence. *Plant Cell* 28, 2510–2527 (**2016**)

29 Perry, C., De Los Santos, E.L.C., Alkhalaf, L.M., Challis, G.L. Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat. Prod. Rep.* 35, 622–632 (**2018**)

30 Lukowski, A. L.; Ellinwood, D. C.; Hinze, M. E.; DeLuca, R. J.; Du Bois, J.; Hall, S.; Narayan, A. R. H., C-H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J Am Chem Soc* 140 (37), 11863-11869 (**2018**)

31 Tanaka, A.; Ito, H.; Tanaka, R.; Tanaka, N. K.; Yoshida, K.; Okada, K., Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proc Natl Acad Sci USA* 95 (21), 12719-23 (**1998**)

32 Llewellyn, L. E., Saxitoxin, a toxic marine natural product that targets a multitude of receptors. *Nat. Prod. Rep* 23, 200-222 (**2006**)

33 Thottumkara, A. P.; Parsons, W. H.; Du Bois, J., Saxitoxin. *Angew. Chem. Int. Ed* 53, 5760-5784 (**2014**)

34 Schantz, E.J.; Magnusson, H.W. observations on origin of paralytic poison in Alaska butter clam. *J. Protozool* 11, 239-242 (**1964**)

35 Wang, D. Z.; Zhang, S. F.; Zhang, Y.; Lin, L. Paralytic shellfish toxin biosynthesis in cyanobacteria and dinoflagellates: A molecular overview. *J. Proteomics* 135, 132-140 (**2016**)

36 Bryant, D. A.; Hunter, C. N.; Warren, M. J., Biosynthesis of the modified tetrapyrroles-the pigments of life. *J Biol Chem* 295 (20), 6888-6925 (**2020**)

37 Chen, M., Chlorophyll modifications and their spectral extension in oxygenic photosynthesis. *Annu Rev Biochem* 83, 317-40 (**2014**)

38 Chen, M.; Blankenship, R. E., Expanding the solar spectrum used by photosynthesis. *Trends Plant Sci* 16 (8), 427-31 (**2011**)

39 Hu, X.; Gu, T.; Khan, I.; Zada, A.; Jia, T., Research Progress in the Interconversion, Turnover and Degradation of Chlorophyll. *Cells* 10 (11) (**2021**)

40 Schliep, M.; Cavigliasso, G.; Quinnell, R. G.; Stranger, R.; Larkum, A. W., Formyl group modification of chlorophyll a: a major evolutionary mechanism in oxygenic photosynthesis. *Plant Cell Environ* 36 (3), 521-7 (**2013**)

41 Croce, R.; Amerongen, H.V. Light harvesting in oxygenic photosynthesis: Structural biology meets spectroscopy. *Science* 369 933 (**2020**)

42 Tanaka, R.; Tanaka, A., Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. *Biochim Biophys Acta* 1807 (8), 968-76 (**2011**)

Chapter 2 Structure Basis for Divergent C-H Hydroxylation Selectivity in Two Rieske Oxygenases

Adapted with permission from Lukowski, A. L.*; Liu, J.*; Bridwell-Rabb, J.; Narayan, A. R. H. Structural basis for divergent C–H hydroxylation selectivity in two Rieske oxygenases. *Nature communications* **2020**, (11) 2991. *Theses authors contributed equally to this work. Copyright © 2020, Springer Nature

Author contribution: A.L.L., J.L., J.B-R., and A.R.H.N. designed experiments and wrote the manuscript. A.L.L. expressed and purified proteins, performed mutagenesis, conducted all biochemical assays, obtained and analyzed LC-MS data. J.L. designed, conducted, and solved the four X-ray crystallographic structures determined in this work. J.L. conducted all structural analysis, identified residues important for the selectivity of the catalyzed reactions, and performed calculations needed to identify of the active site tunnels in SxtT and GxtA.

2.1 Abstract

Biocatalysts that perform C–H hydroxylation can exhibit exceptional substrate specificity and site-selectivity, often through the use of high-valent oxidants to activate these inert bonds. Rieske oxygenases are examples of enzymes with the ability to perform precise mono- or dioxygenation reactions on a variety of substrates. Understanding the structural features of Rieske oxygenases responsible for control over selectivity is essential to enable the development of this class of enzymes for biocatalytic applications. Decades of research has illuminated the critical features common to Rieske oxygenases; however, structural information for enzymes that functionalize diverse scaffolds is limited. Here, we report the structures of two Rieske monooxygenases involved in the biosynthesis of paralytic shellfish toxins (PSTs), SxtT and GxtA, adding to the short list of structurally characterized Rieske oxygenases. Based on these structures,

substrate-bound structures, and mutagenesis experiments, we implicate specific residues in substrate positioning and the divergent reaction selectivity observed in these two enzymes.

2.2 Introduction

Enzymes that perform C–H hydroxylation reactions often demonstrate impressive control over site- and stereoselectivity on complex scaffolds[1-3]. This precision is emulated in numerous natural product biosynthetic pathways by several classes of metalloenzymes including cytochrome P450 monooxygenases[4], non-heme α-ketoglutarate-dependent oxygenases[5], and Rieske oxygenases[6]. To accomplish selective modification of inert C–H bonds, these enzyme classes employ molecular oxygen to generate high-valent iron intermediates, which activate a substrate for hydroxylation through hydrogen atom abstraction[7-11]. Examples include the Fe(IV)-oxo and Fe(III)-hydroperoxo intermediates employed by heme-containing systems[12,13] and the Fe(III)-superoxo[14], Fe(IV)-oxo, or proposed Fe(V)-oxo species used by non-heme systems[7]. In the Rieske oxygenase class, work to date has been conducted on a small sample of enzymes, specifically those that perform dioxygenation of aromatic rings. However, enzymes in this class catalyze a plethora of reactions, including monohydroxylation[15,16], dihydroxylation[17], dealkylation of heteroatoms[18], desaturation[19], and cyclization reactions[20]. The sequence divergence within the class presents a challenge in assigning one of these functions to a given Rieske oxygenase, let alone predict the substrate, site of reactivity, or type of reaction (Figure. 2.1.a).

Structural studies on enzymes with divergent selectivities have the potential to illuminate critical features that are responsible for selectivity and provide valuable insight on how reactivity can be controlled in biocatalytic reactions. Rieske oxygenases in particular are underexplored in this regard, with only twelve available structures of different enzymes in the Protein Data Bank[21]. Five of these enzymes, including biphenyl dioxygenase[22-25], 3-nitrotoluene dioxygenase[26], naphthalene 1,2-dioxygenase[27-30], cumene dioxygenase[31], and nitrobenzene dioxygenase[32], are clustered together in the sequence similarity network (SSN) presented in Figure 2.1.a (light blue)

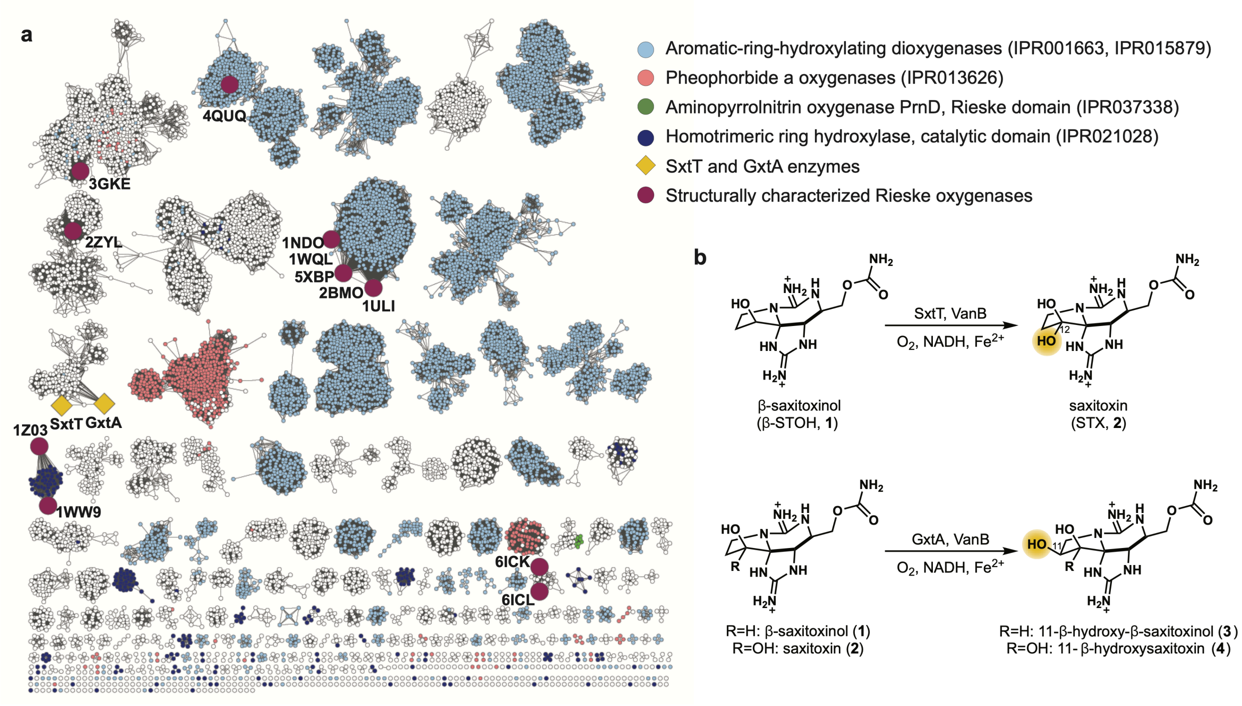


Figure 2.1. Extent of characterization of the Rieske oxygenase enzyme superfamily. (a) Sequence similarity network of Rieske oxygenases (Pfam PF00355) generated by the EFI-EST[59] server at an E-value of 5 and alignment score of 65, visualized in Cytoscape v3.7.1[60]. The network shown was trimmed to include only Rieske proteins that contain the non-heme iron catalytic domain (IPR017941, IPR005805, and IPR domains listed in the figure). PDB codes represent the first structure determined for each enzyme. 1ULI, biphenyl dioxygenase; 5XBP, nitroluene dioxygenase; 1NDO, naphthalene dioxygenase; 1WQL, cumene dioxygenase; 2BMO, nitrobenzene dioxygenase; 1WW9, carbazole 1,9α-dioxygenase; 1Z03, 2-oxoquinoline 8-monooxygenase; 3GKE, dicamba monooxygenase; 4QUQ, stachydrine demethylase; 2ZYL, 3-ketosteroid-9-α-hydroxylase; 6ICK, NdmA; 6ICL, NdmB. (b) Rieske monohydroxylases SxtT and GxtA catalyze highly selective C–H functionalization reactions in paralytic shellfish toxin biosynthesis.

and catalyze the oxidative degradation of aromatic compounds (IPR001663, IPR015879). The remaining seven structurally elucidated Rieske oxygenases are also involved in degradative pathways: carbazole 1,9α-dioxygenase[33], 2-oxoquinoline 8-monooxygenase[34], dicamba monooxygenase[35], stachydrine demethylase[36], and NdmA/NdmB[37]. These enzymes each degrade aromatic heterocycles and/or phenols, whereas 3-ketosteroid-9-α-hydroxylase (KshA[15]) catabolizes the cycloalkane rings of ketosteroids. The disparity between the large number of known Rieske oxygenases and the number that have been biochemically or structurally investigated highlights the need for further studies on the vast number of uncharacterized Rieske oxygenases.

In this work, we focus on two cyanobacterial Rieske oxygenases, SxtT and GxtA, that are involved in the biosynthesis of paralytic shellfish toxins (PSTs, Figure. 2.1.b)[16]. PST natural products exhibit high affinities for neurotoxin receptor site 1 of human voltage-gated sodium channels (VGSCs). The affinity for VGSCs varies based on the pattern of functional groups present on the PST tricyclic core.[38,39] Recently, we demonstrated that SxtT and GxtA are responsible for complementary site- and stereoselective C–H hydroxylation reactions on the PST scaffold (Figure. 2.1.b).[16] We discovered that SxtT operates on a number of tricyclic substrates, preferring β-saxitoxinol (β-STOH, **1**), to install the C12 α-hydroxyl group, generating the potent natural product, saxitoxin (STX, **2**, Figure. 2.1.b)16. GxtA, which shares 88-percent sequence identity with SxtT (Figure. 2.9), displays divergent selectivity. GxtA hydroxylates the C11 β-position of β-STOH (**1**) and STX (**2**) to afford 11-β-hydroxy-β-saxitoxinol (**3**) and 11-β-hydroxysaxitoxin (11-β-hydroxySTX, **4**), respectively (Figure. 2.1.b). Herein, we report crystal structures of SxtT and GxtA. Using the determined SxtT, GxtA, and STX analog-bound structures coupled with biochemical assays, we find that the impeccable site-selectivity of the SxtT catalyzed reaction is in part dictated by the formation of a hydrogen bond in the active site between the guanidinium moiety of the substrate and a polar residue at position 276. In GxtA, we instead find that the guanidinium moiety of the substrate is positioned by Gln226 and Asp239 residues. We identify a flexible loop that may play a role in gating access of substrate to an active site tunnel and speculate that changes to this region of the proteins also play important roles in reaction selectivity. Taken together, this work provides the structural basis for site selectivity of two Rieske monooxygenases and facilitates targeted engineering efforts to tune selectivity for the generation of non-natural or synthetically intractable PST analogs for VGSC studies.

2.3 Results

2.3.1 SxtT and GxtA showcase structurally similar trimeric architectures

The crystal structure of SxtT was determined to 1.86-Å resolution using a modified model of dicamba monooxygenase for molecular replacement (SxtT and dicamba monooxygenase share 31-percent sequence identity with one another). The structure of SxtT was subsequently used to solve the 2.10-Å resolution structure of SxtT with a STX analog bound and the 2.20-Å resolution structure of GxtA (Table 2.1). The latter was then used to solve the 2.18-Å resolution structure of substrate-analog bound GxtA (Table 2.1). As predicted based on the high sequence identity of SxtT and GxtA, the proteins are structurally similar with a root-mean-square deviation (r.m.s.d) determined by PyMOL of 0.971 Å over 950 Cα atoms. In accordance with other characterized Rieske oxygenases, SxtT and GxtA are composed of two domains that make up the catalytic α-subunit: an N-terminal Rieske domain, and a C-terminal catalytic iron-containing domain (Figure. 2.2.a). In a single monomer, the Rieske cluster and the mononuclear iron

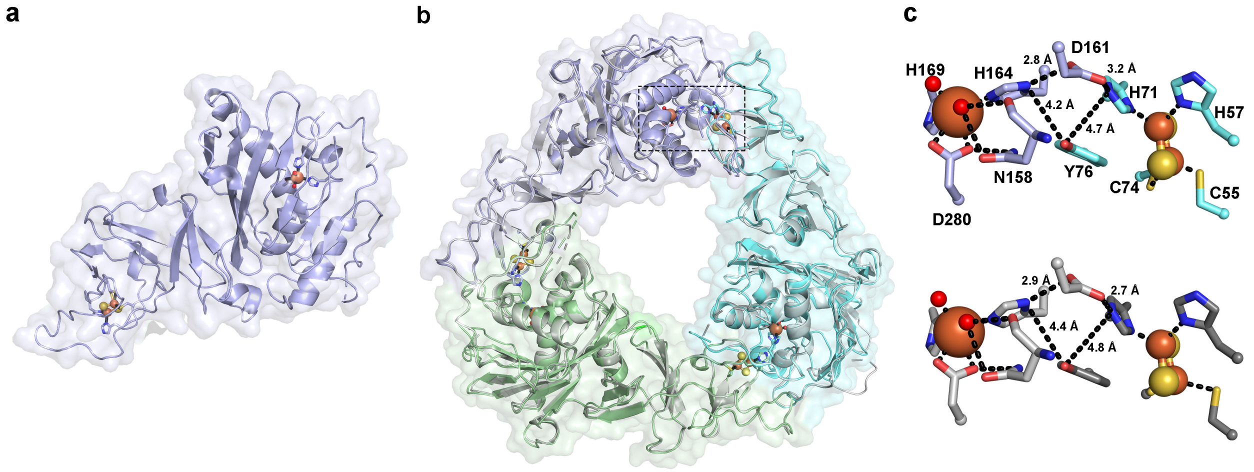


Figure 2.2. SxtT and GxtA showcase structurally similar trimeric architectures and metallocenter arrangements.
(a) The monomeric unit of SxtT reveals that the metallocenters are located approximately 45-Å apart. (b) The trimeric quaternary structures of SxtT (colors) and GxtA (gray) are overlaid with one another. The trimeric arrangement of subunits allows for the close positioning of the

metallocenters. (c) Two residues, Asp161 and Tyr76, which are found at the subunit-subunit interface of both SxtT (top panel) and GxtA (bottom panel, same residue numbering as SxtT) bridge the Rieske cluster to the non-heme iron center. SxtT is shown in purple and cyan and GxtA is gray.

site are located nearly 45 Å apart, a distance that is too far to permit intra-subunit electron transfer (Figure. 2.2.a). As observed in other Rieske oxygenases, the monomeric subunits of SxtT and GxtA arrange into an $\alpha3$ trimeric architecture to facilitate inter-subunit electron transfer from the Rieske cluster to the non-heme iron center (Figure. 2.2.b,c)[10,40,41].

According to the DALI server[42], SxtT and GxtA have the highest structural homology to dicamba monooxygenase (3GOB[35]) and KshA (4QDC[15]). Like these structures, SxtT and GxtA house the characteristic Rieske cluster motif and the 2-His, 1-Asp facial triad for coordinating the non-heme iron site[10] (His164, His169, and Asp280). The octahedral iron coordination environment in SxtT and GxtA is completed by two water molecules, one of which is held in place through interactions with the backbone and sidechain of Asn158 (Figure. 2.3.a). Although this Asn residue is conserved in many other structurally characterized α-only oxygenases[15,33-35,43], its functional significance is unclear.

Notably, comparison of the SxtT and GxtA active site pockets reveals two differences, namely the inversion of a hydrophobic residue and a polar residue at positions 255 and 276. In SxtT, these positions correspond to Met and Thr residues, whereas in GxtA, they correlate with Tyr and Val residues, respectively (Figure. 2.3.a). Importantly, these positions bracket the mononuclear iron site and are anticipated to play a role in positioning the substrate, dictating the selectivity of the catalyzed reactions (Figure. 2.3.a).

Table 2.1. Data collection and refinement statistics.

	SxtT	GxtA	SxtT + ddSTX	GxtA + ddSTX
Data collection				
Space group	$C222$	$P12_11$	$C222$	$P12_11$
Cell dimensions				
a, b, c (Å)	152.3, 159.6, 116.0	74.73, 96.93, 80.81	152.3, 159.6, 115.9	74.58, 96.98, 80.41
α, β, γ (°)	90, 90, 90	90, 107.0, 90	90, 90, 90	90, 107.2, 90
Resolution (Å)	30-1.86	40-2.20	30-2.10	40-2.18
R_{merge}	0.074 (0.907)	0.186 (0.477)	0.087 (0.436)	0.090 (0.755)
$I / \sigma I$	25.64 (2.36)	10.56 (1.56)	15.8 (2.07)	15.3 (1.84)
Completeness (%)	100 (100)	96.3 (90.9)	98.4 (89.0)	100 (100)
Redundancy	7.6 (7.5)	4.3 (2.2)	4.7 (3.4)	3.8 (3.8)
CC1/2	0996 (0.853)	0.981 (0.730)	0.984 (0.890)	0.995 (0.607)
Refinement				
Resolution (Å)	1.86	2.20	2.10	2.18
No. reflections	117619	54048	81773	56607
$R_{\mathrm{work}} / R_{\mathrm{free}}$	0.170, 0.198	0.192, 0.219	0.168, 0.207	0.171, 0.227
No. atoms				
Protein	8114	7759	8056	7721
[2Fe-2S]	12	12	12	12
Non-heme Fe	3	3	3	3
Glycerol	36	6	42	54
Sulfate	65	-	65	-
Water	885	359	443	298
ddSTX	-	-	57	57
B-factors				
Overall	40.65	43.81	45.23	39.16
Protein	39.55	43.77	44.72	38.87
ddSTX	-	-	65.34	50.75
[2Fe-2S]	28.54	30.80	32.46	28.13
Non-heme Fe	30.26	33.87	32.07	26.41
Glycerol	59.32	64.57	61.80	55.06
Sulfate	88.64	-	89.05	-
Water	46.54	49.31	46.28	40.56
R.m.s. deviations				

Bond lengths (Å)	0.008	0.012	0.008	0.008
Bond angles (°)	0.981	1.574	0.969	0.970

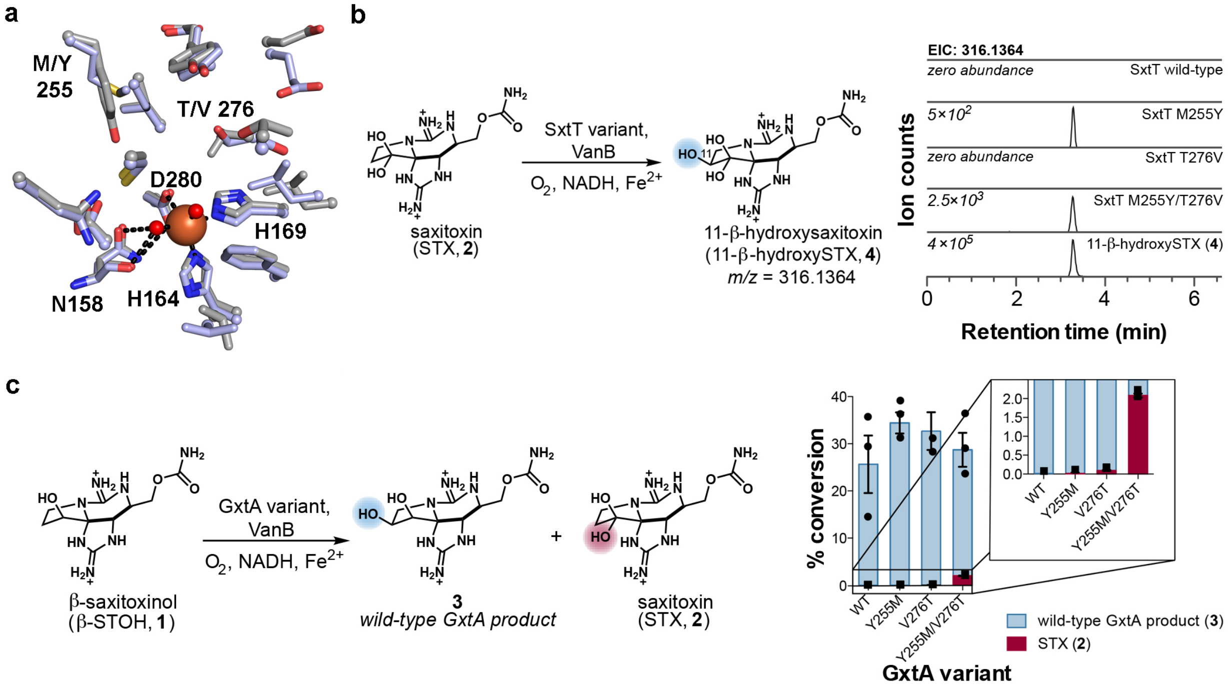


Figure 2.3. Biochemical analysis of SxtT and GxtA variants reveals changes in reaction selectivity.

(a) The active sites of SxtT (purple) and GxtA (gray) are relatively similar to one another except for the replacement of SxtT residues Met255 and Thr276 with GxtA residues Tyr255 and Val276. (b) Reactions of SxtT wild-type and variants with STX (2). (c) Reactions of GxtA wild-type and variants with β-STOH (1) and the mixtures of products generated in these reactions. Data are presented as mean values +/- SD. Reaction conditions: 5 μM SxtT/GxtA, 5 μM VanB, 200 μM β-STOH (1), 500 μM NADH, incubated at 30 °C for 2 h. The enzymatic reaction data are generated by Dr. April Lukowski.

2.3.2 Site-directed mutagenesis of SxtT and GxtA results in added functionality.

The roles of residues at the 255 and 276 positions in SxtT and GxtA were probed using site-directed mutagenesis to generate single and double variants of SxtT (M255Y, T276V, and

M255Y/T276V) and GxtA (Y255M, V276T, and Y255M/V276T, Tables 2.2 and 2.3). Each variant was expressed, purified, and analyzed by quantitative LC-MS in reactions using VanB as the electron donor (Figure. 2.3.b-c and Figure. 2.10)[16]. Each enzyme variant was assessed with β-STOH (**1**) and STX (**2**) to interrogate changes in the site-selectivity of the reactions.

Dr. April Lukowski then performed the site-directed mutagenesis for SxtT and GxtA following with activity assays. Variants possessing a single substitution at the 276 position are biased toward wild-type site-selectivity on native substrates (i.e., SxtT with β-STOH (**1**) and GxtA with STX (**2**), Figures. 2.11-13). The efficiency of the SxtT T276V variant was especially low due to precipitation upon the addition of substrate, however, a small amount of exclusively the native product STX (**2**) was formed. The GxtA V276T variant proved to be more stable under our reaction conditions, and we observed a small amount of dihydroxylated product formed in reaction with β-STOH (**1**), suggesting the variant is capable of hydroxylation at both the C11 and C12 positions (Figure. 2.14). This ability in the GxtA V276T variant, which has two H-bond-donating residues in the active site, is mirrored by the corresponding SxtT M255Y variant. Reactions with the SxtT M255Y variant revealed a small amount of non-native hydroxylation activity on STX (**2**), a substrate which is not accepted by wild-type SxtT, generating 11-β-hydroxySTX (**4**, Figure. 2.11). The corresponding GxtA variant, GxtA Y255M, was stable to standard reaction conditions with β-STOH (**1**), but did not demonstrate a change in selectivity, generating only **3**. Interestingly, both the SxtT and GxtA double variants displayed altered site-selectivity relative to the wild-type enzymes. In addition, SxtT M255Y/T276V demonstrated the ability to hydroxylate STX (2) more efficiently than SxtT M255Y. Analysis by LC-MS and MS/MS confirmed that SxtT M255Y and SxtT M255Y/T276V were capable of hydroxylating STX (**2**) at the C11 β-position, generating the native GxtA product, 11-β-hydroxySTX (**4**, Figure. 2.11). The catalytic efficiencies of these reactions were evaluated using steady-state kinetics in comparison to the wild-type GxtA reaction with STX (**2**, Figures. 2.15-17)[44]. The reaction of GxtA with STX (**2**) demonstrated an apparent catalytic efficiency (k_{cat}/K_M) of 6,600 M^{-1} s^{-1}. By contrast, the apparent k_{cat}/K_M of SxtT M255Y with STX (**2**) was ~440 times lower at 15 M^{-1} s^{-1}. However, introducing a second GxtA-like residue to generate SxtT M255Y/T276V revealed improvement in catalysis with STX (**2**) with an apparent k_{cat}/K_M of 41 M^{-1} s^{-1}.

Dr. April Lukowski also noticed that analysis of reactions between GxtA and its non-native substrate β-STOH (**1**) was complicated by co-elution of **3** and STX (**2**), the two potential products (Figure. 2.12.b). To differentiate **3** from STX (**2**) in reactions with GxtA variants, we leveraged the equilibrium exhibited by the hydrated ketone moiety of STX (**2**)[45]. At neutral pH, ethanol (EtOH) can be incorporated into STX (**2**) to form a hemiketal, enabling the unambiguous distinction between hydroxylation at C12 (EtOH incorporation) or C11 (no EtOH incorporation, Figure. 2.18-19) by LC-MS. Upon addition of EtOH to a reaction of GxtA Y255M/V276T with β-STOH (**1**), both an EtOH-incorporated product and a non-EtOH incorporated product were observed. This result suggests both STX (**2**) and **3** were generated (Figure. 2.3.c and Figure. 2.19), indicating that the GxtA double variant can perform reactions with both wild-type SxtT and GxtA activities. Comparing conversions of GxtA, GxtA Y255M, GxtA V276T, and GxtA Y255M/V276T with β-STOH (**1**) after 2 h revealed approximately equivalent total percent conversions for each enzyme (Figure. 2.3.c, Figure. 2.18-19, and Table 2.4), suggesting that the overall level of catalysis is not impaired. Small quantities of STX (**2**) were generated as substitutions were introduced, making the GxtA active site more similar to SxtT, with the most C12 hydroxylation observed in GxtA Y255M/V276T (Figure. 2.3.c). Together, these results indicate the two residues identified structurally alter C–H hydroxylation selectivity, resulting in gain of function variants.

2.3.3 Substrate analog bound structures reveal an essential hydrogen bond

To further probe the selectivity differences in SxtT and GxtA, the structures of each protein bound to dideoxysaxitoxin (ddSTX) were determined (Table 2.1). ddSTX has previously been shown to serve as a substrate for both SxtT and GxtA and exhibits reduced toxicity relative to STX and β-STOH[16]. This lowered toxicity, coupled with the availability of the molecule through synthesis, rendered ddSTX a prime target for the pursuit of substrate-bound structures. The ddSTX-bound structures are relatively similar to those of the native enzymes (r.m.s.d determined by PyMol of 0.338 Å and 0.491 Å over 1002 and 952 Cα atoms for SxtT and GxtA, respectively) except for the presence of one ddSTX molecule in the active site of each monomer (Figure. 2.4-6). The binding of ddSTX in the active site of SxtT involves hydrogen bonds between the carbamate sidechain and Ser230, Tyr273, Asp239, and Arg204, as well as interactions between the guanidium group of the six-membered ring with Asp239 and Gln226 (Figure. 2.7.a). The ddSTX molecule is also held in the active site by two interactions with the five-membered cyclic guanidinium group of ddSTX, which participates in hydrogen bonds with Thr276 and a water molecule that bridges ddSTX to the non-heme iron ligand Asp280 (Figure. 2.7.a, Figure. 2.5.a, and Figure. 2.6.a, c, d). Along with Thr276, each formed interaction anchors ddSTX such that C12 is the closest atom of the molecule to the non-heme iron site and positions

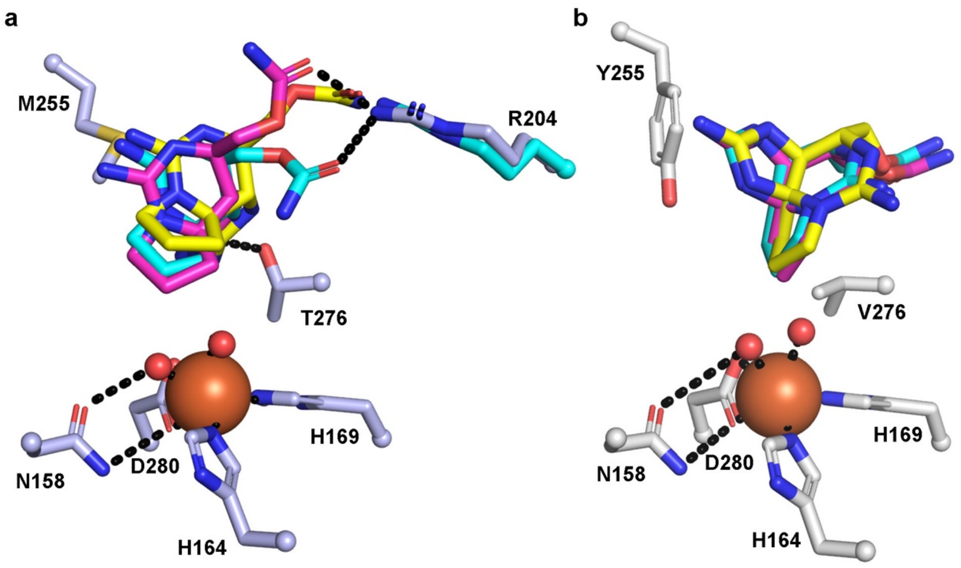


Figure 2.4. ddSTX is bound to SxtT and GxtA in each monomeric unit of the trimer.
(a) In SxtT, these orientations slightly differ in the position of the carbamate sidechains. The molecule of ddSTX found in chains A, B, and C are colored dark pink, cyan, and yellow, respectively. In chains A and B, the loop that spans the range 195-215 is ordered and Arg204 interacts with the substrate. In chain C, this loop is too disordered to be modeled. (b) In GxtA, the observed orientation of ddSTX is more consistent between chains. The molecule of ddSTX found in chains A, B, and C are colored dark pink, cyan, and yellow, respectively. The loop, despite being ordered in chain C, is found in the "open" conformation and does not interact with ddSTX.

the α-C–H bond for hydrogen atom abstraction (Figure. 2.7.a and Figure. 2.5.a). Met255 appears to be important for spatially directing the binding of ddSTX in the SxtT active site as it packs against the six-membered guanidinium ion ring. Importantly, as residues Tyr273, Asp239, and Gln226 are conserved in GxtA and Arg204 and Ser230 are substituted with Lys and Thr residues (Figure. 2.9), correct placement of the C12 α-C–H bond for hydroxylation appears to be largely dictated by the bulk and hydrophobicity of Met255 and the hydrogen bond with Thr276. Indeed, the replacement of these two residues in the GxtA active site with Tyr and Val residues results in significant movement of ddSTX such that it interacts only with conserved residues, Asp239 and Gln226 (Figure. 2.7.b, Figure. 2.5.b, and Figure. 2.6.b,e,f).

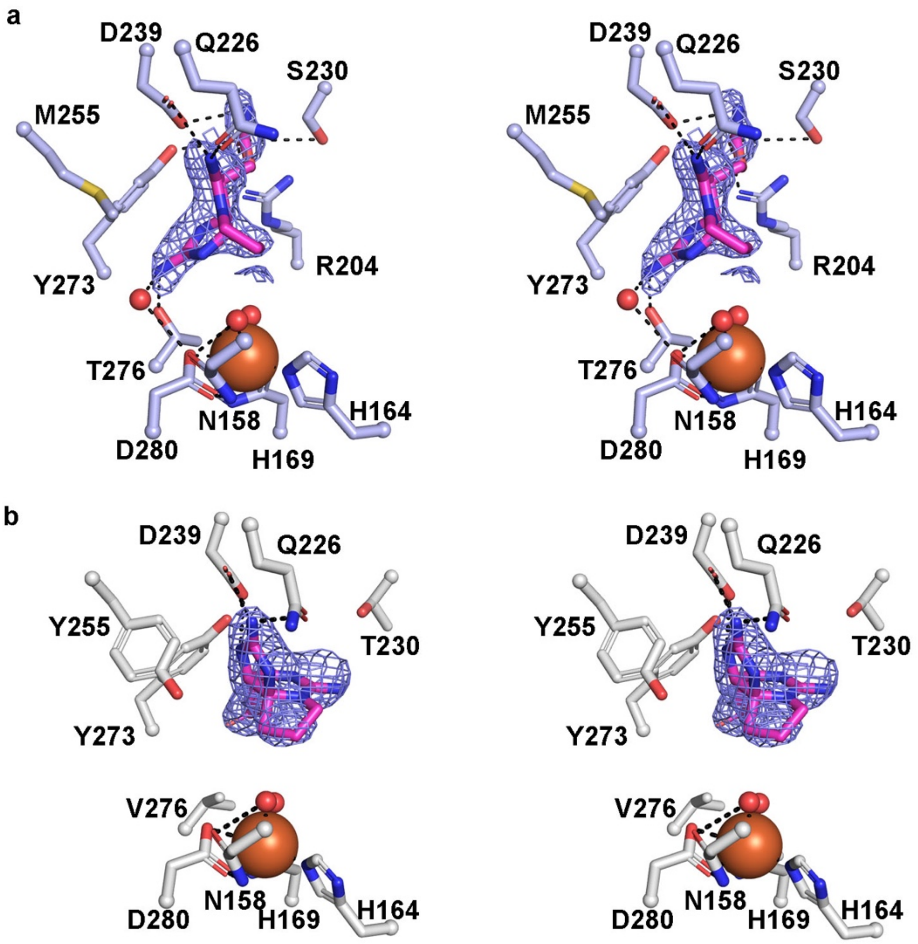


Figure 2.5. ddSTX is bound in the active sites in SxtT and GxtA.

(a) Stereo Figure of ddSTX bound to SxtT. Here, ddSTX interacts with Ser230, Tyr273, Asp239, Gln226, and Arg204. (b) Stereo figure of ddSTX bound to GxtA. In this structure, ddSTX forms interactions with Asp 239 and Gln226. Both panels are shown with 2Fo-Fc simulated annealing composite omit electron density maps contoured at 0.9σ.

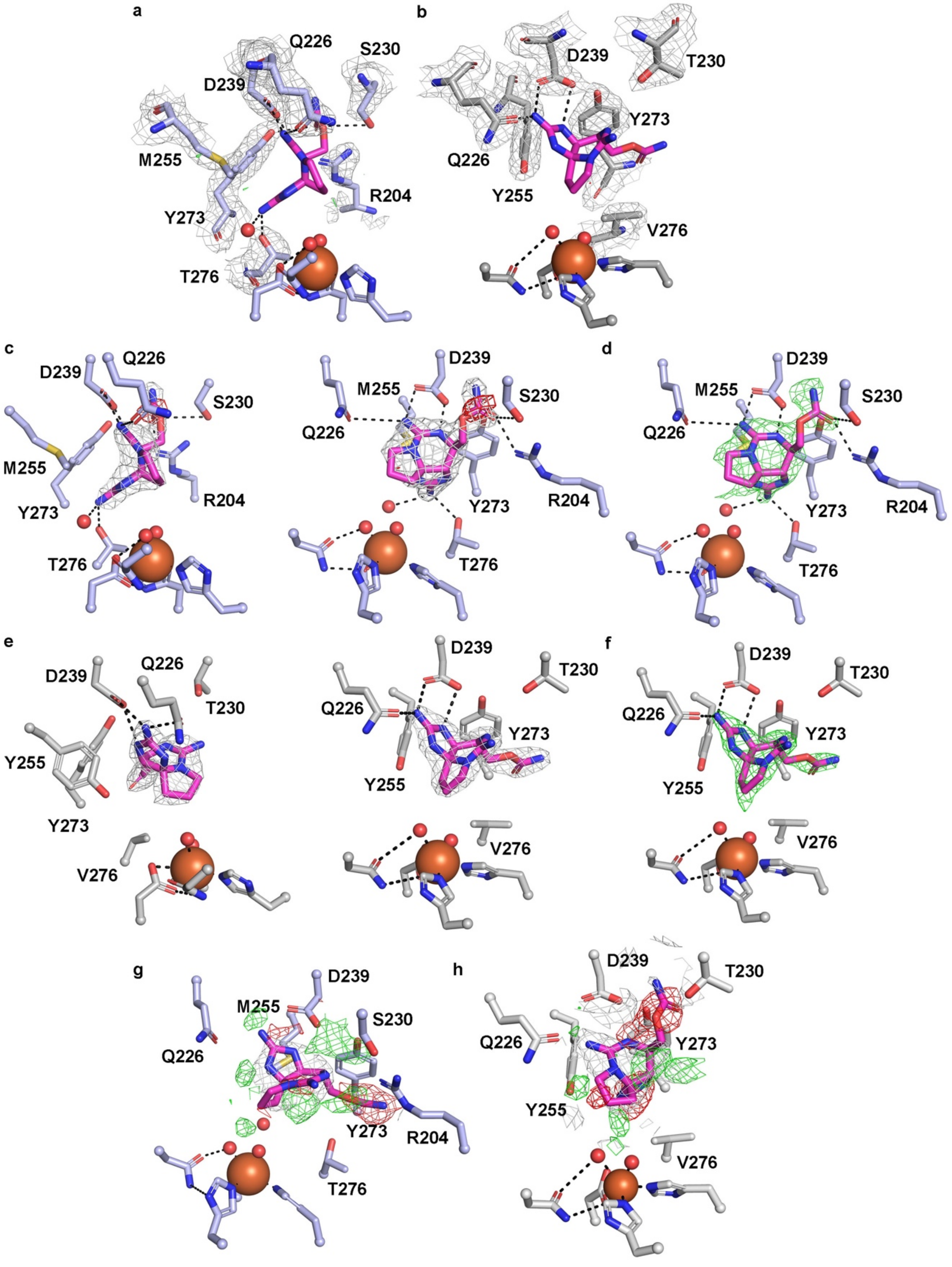


Figure 2.6. ddSTX makes specific interactions in the active sites of SxtT and GxtA.
(a) 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively are shown for the sidechains of SxtT implicated in forming the ddSTX binding site. (b) Similar to panel a, this panel shows 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ,

respectively around the sidechains of GxtA implicated in forming the ddSTX binding site. (c) 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively around ddSTX in two orientations from chain A of SxtT. (d) An Fo-Fc omit electron density map calculated after ddSTX was omitted from the refined structure of SxtT. This map is contoured at 3.0σ around ddSTX. (e) 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively around ddSTX in two orientations from chain A of GxtA. (f) An Fo-Fc omit electron density map calculated after ddSTX was omitted from the refined structure of GxtA. This map is contoured at 3.0 σ around ddSTX. (g) In this panel the orientation of ddSTX in GxtA has been refined in the active site of SxtT. This incorrect ddSTX orientation is shown with 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively. (h) Similar to panel g, in this panel, the orientation of ddSTX from SxtT has been refined in the active site of GxtA. This incorrect orientation of ddSTX is shown with 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively. In this figure, all 2Fo-Fc maps are shown in gray, whereas positive Fo-Fc difference density maps are shown in green and negative Fo-Fc difference density maps are shown in red.

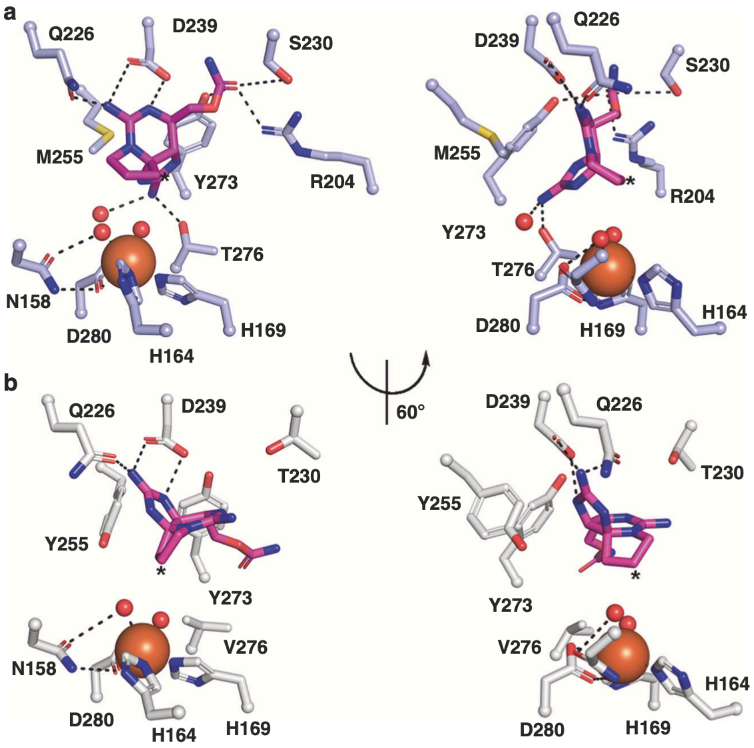


Figure 2.7. Structures of SxtT and GxtA with ddSTX bound reveal the protein interactions that are important for correctly positioning substrate for activation.

(a) ddSTX (dark pink sticks) is anchored in the active site of SxtT via interaction with several different residues that are highlighted here. A rotated view of the active site reveals that C12 (asterisk) is the closest position of ddSTX to the non-heme iron site. (b) The positioning of ddSTX in the GxtA active site places C11 (asterisk) closest to the iron center. Relative to the placement of ddSTX in SxtT, the guanidinium group that interacts with Thr276 rotates nearly 120° to istead interact with Gln226 and Asp239 in the GxtA active site. Both of these residues are conserved in SxtT.

2.3.4 Identification of a putative active site tunnel in SxtT and GxtA

The double variant enzymes unexpectedly have the ability to perform additional reactions with the selectivity of the opposite enzyme, but remain biased toward the site-selectivity of the native reaction, as evidenced by the GxtA reaction with β-STOH (**1**). Thus, we decided to investigate differences in residues located outside of the active sites. We found that each of the structures determined in this work possess subunits that are missing some residues in a loop that spans the range 195-215. These residues are too disordered to be confidently built into the refined structures. However, in one subunit of both the ligand-free SxtT and GxtA structures, the loop is ordered and can be visualized. Comparison of these ordered subunits shows that this loop assumes remarkably different conformations. In SxtT, this loop folds in and reaches toward the active site pocket, whereas in GxtA, this loop adopts an extended conformation on the surface of the protein (Figure. 2.8.a). These loop conformations were further investigated using the MOLEonline server[46,47] to calculate tunnels present in our structures. Through this analysis, we were able to visualize a tunnel in GxtA that leads from the surface of the protein to the active site (Figure. 2.8.b). An equivalent tunnel in SxtT does not exist due to the orientation of the loop, which closes off the active site cavity (Figure. 2.8.c). We likely benefitted from the crystal contacts in capturing two states of the loop orientation, but based on the location of the active site and the flexibility of this loop in all of the other subunits, we hypothesize that this loop samples a multitude of orientations that either restrict or permit access to the active site. Consistent with this loop serving as a gate to the active site, in our ddSTX-bound SxtT structure, this loop partially orders in each subunit and interacts with the bound ddSTX molecule, which is found at the base of the calculated tunnels (Figure. 2.8.d). Closer inspection of the tunnel also revealed differences in the tunnel-lining residues, which may be important for guiding the substrate into its position in the active site (Figure. 2.8.e).

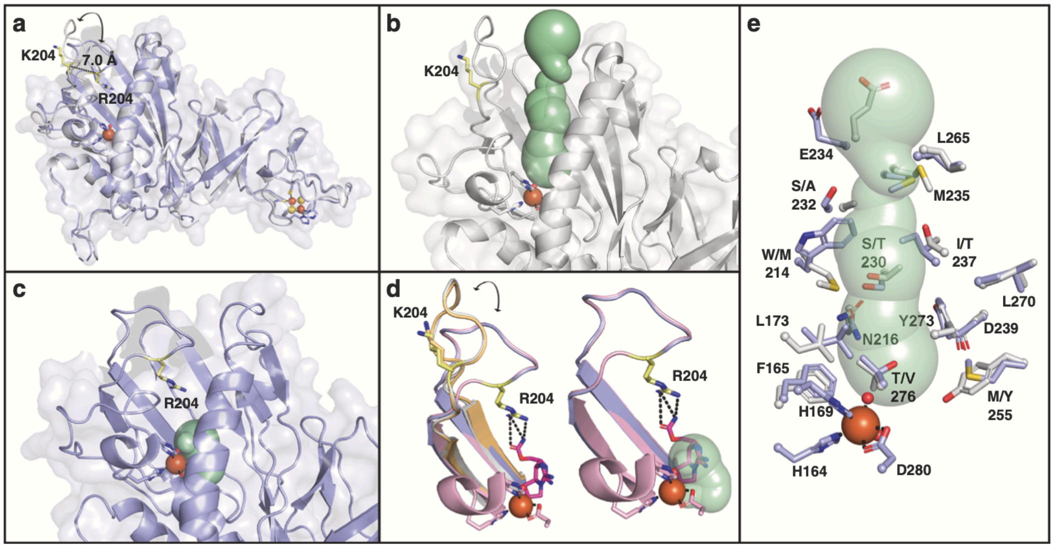


Figure 2.8. SxtT and GxtA exhibit different conformations of a loop that appears to gate access to the active site.

(a) An overlay of the SxtT and GxtA monomers reveals a 7.0 Å difference in the location of the residue at the 204 position. This difference is due to changes in the position of a loop that spans the range 195–215. (b) The open conformation of the loop in GxtA allows visualization of a tunnel that leads directly from the protein surface to the active site. (c) The closed orientation of the loop in SxtT restricts calculation of an equivalent tunnel, but instead allows visualization of the active site cavity. (d) An overlay of all four structures determined in this work reveals that Arg204 in the co-crystal structure assumes a similar position to that seen in the ligand free structure of SxtT. In this orientation, Arg204 interacts with ddSTX. (e) An overlay of SxtT and GxtA shows differences in the residues that line the putative tunnel to the active site. In all panels, structures are shown for SxtT, GxtA, ddSTX-SxtT, and ddSTX-GxtA in purple, gray, pink, and orange respectively. Tunnels are colored mint, ddSTX is dark pink, and the Arg/Lys204 residue is yellow.

2.4 Discussion

Biocatalysts for C–H hydroxylation are underexploited in chemical synthesis due to our limited understanding of structure-function relationships. This is particularly evident in the Rieske oxygenase family of enzymes, which are structurally underexplored. Here, we examined two biosynthetic Rieske monooxygenases, SxtT and GxtA, that catalyze site- and stereoselective reactions in PST biosynthesis. By solving structures of SxtT and GxtA, we identified two divergent residues in the active sites of these enzymes that bracket the non-heme iron site. We tested the importance of these residues in dictating the site-selectivity of the catalyzed reactions by replacing one or two residues in each protein and gauging the ability of these enzymes to catalyze hydroxylation at either C11 or C12 (Figure. 2.3). We found that the removal of a hydrogen bond donor from the active site to generate SxtT T276V resulted in precipitation under our reaction conditions with every substrate. The removal of a hydrogen bond donor to generate GxtA Y255M resulted in trace amounts of dihydroxylation activity. In contrast, the addition of a hydrogen bond donor to the active site of both enzymes (SxtT M255Y and GxtA V276T) permits a small amount of monohydroxylation to occur at the non-native site. For each double variant we saw that the amount of hydroxylation at these non-native positions increased.

We further revealed, by solving ddSTX-bound structures of SxtT and GxtA, that these residues are important for positioning substrate in the active site. In SxtT, Thr276 hydrogen bonds to the five-membered ring containing a guanidinium ion and Met255 sterically orients ddSTX for catalysis at C12 (Figure. 2.7.a). In GxtA, rotation about the carbamate group instead positions ddSTX for hydrogen-atom abstraction at C11 (Figure. 2.7.b). Intriguingly, each of the observed interactions between ddSTX and GxtA are with residues that are conserved in SxtT, highlighting the importance of Thr276 for correctly positioning C12 for hydroxylation. However, it has not escaped our notice that the addition of a hydroxyl group to the C12 α-position of ddSTX (identical to that found in the native GxtA substrate) would provide an additional interaction with Tyr255 in the GxtA active site, further supporting our data and the proposed importance of this residue for selectivity (Figure. 2.7.b, Figure. 2.5.b, and Figure. 2.6.b,e,f).

Surprisingly, the double variant enzymes remain biased toward native reaction site- and stereoselectivity. This incomplete switch in selectivity could be due to distinct conformational changes upon substrate binding as well as sequence differences in the composition of the flexible loop or substrate tunnel identified in these enzymes (Figure. 2.8). Indeed, although we find that these enzymes share 88-percent sequence identity with one another, the 195-215 loop is only 52-percent identical in sequence and only 29-percent identical over the portion (201-214) that plugs the active site tunnel in SxtT. Different orientations of the equivalent loop have been observed in dicamba monooxygenase[48] and NdmA/NdmB[37] and, similar to what we propose for SxtT and GxtA, have been suggested to either permit or restrict access of substrate to the active site and/or promote product release[48]. Consistent with this hypothesis, GxtA (which showcases the "open" conformation) exhibits a broader substrate scope than SxtT[16]. We also find that the residues lining the putative active site tunnel vary between SxtT and GxtA; several residues in SxtT (Trp314, Ile237, and Ser232) that are larger than their counterparts in GxtA (Met214, Thr237, and Ala232) are found near the entrance to the active site. These differences may bias substrate positioning, serve a role in substrate recognition, or gate substrate access to the active site. In agreement with this hypothesis, recent computational studies on naphthalene dioxygenase predict that residues in an equivalent region may also play roles in controlling substrate access to the active site[49]. Finally, it is also a possibility that the protein forms interactions with hydroxyl groups present on the native substrates of each enzyme that are not observed with ddSTX. Further studies to determine the importance of non-conserved residues in the tunnel and loop region and obtain crystal structures with native substrates bound in the active site of each enzyme will be pursued to further define the mechanisms of selectivity in SxtT and GxtA.

Understanding the basis for divergent selectivity has implications for engineering these enzymes as biocatalysts for the generation of nonnative PST analogs for pharmacological studies. Although there is a still much to learn about the Rieske oxygenase superfamily, the structural characterization of SxtT and GxtA provides information on this understudied class of non-heme monooxygenases and reveals several architectural components that are important for the selectivity of these enzymes.

2.5 Data availability

Protein coordinates and structure factors have been submitted to the Protein Data Bank under accession codes 6WN3 (SxtT) [http://www.rcsb.org/structure/6WN3], 6WNC (GxtA) [http://www.rcsb.org/structure/6WNC], 6WNB (ddSTX-bound SxtT) [http://www.rcsb.org/structure/6WNB], and 6WND (ddSTX-bound GxtA) [http://www.rcsb.org/structure/6WND].

2.6 Supplementary information

This section was prepared by Dr. April Lukowski

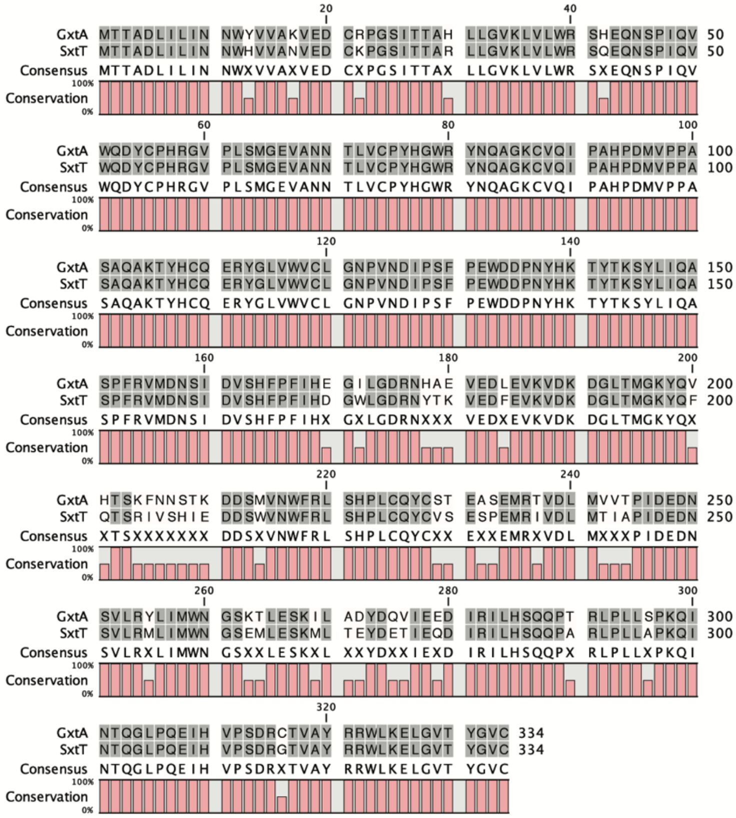


Figure 2.9. Alignment of SxtT and GxtA protein sequences reveals that these proteins share 88-percent sequence identity with each other.

Conserved residues are shown in grey and divergent residues are not highlighted. Alignment generated using ClustalW and visualized by CLC Sequence Viewer (v7.6.1).

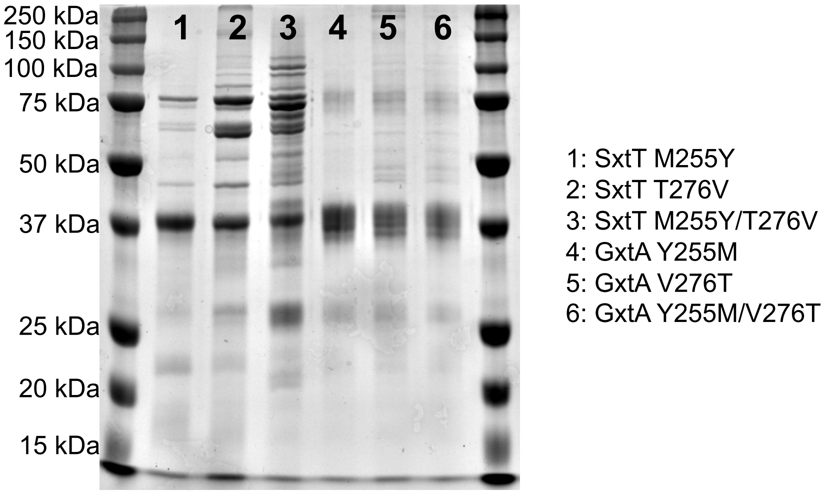


Figure 2.10. SxtT and GxtA variants were expressed, purified by Ni-affinity chromatography, and used in enzymatic assays.

Shown is an SDS-PAGE gel of each variant following the final purification step. The anticipated molecular weight of each variant is ~38 kDa.

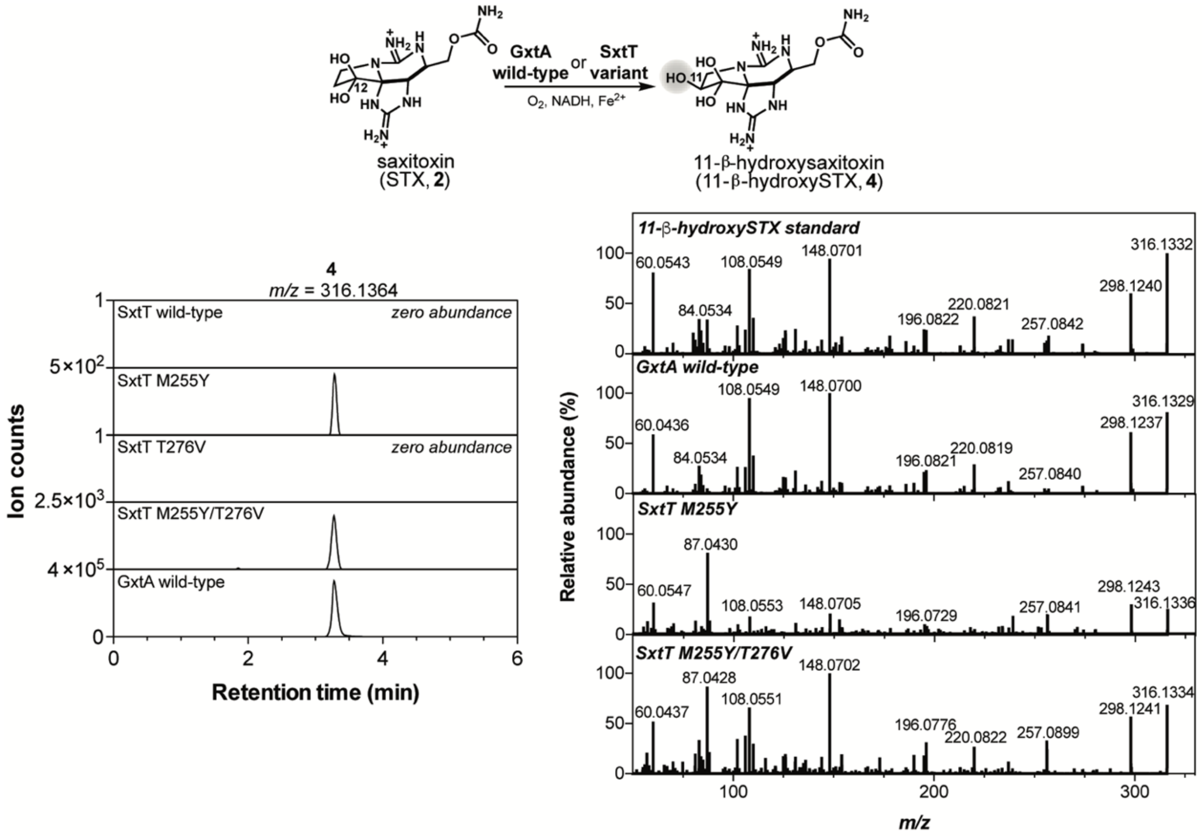


Figure 2.11. MS and MS/MS analysis of SxtT and GxtA variants reactions.
As demonstrated using HILIC-MS and targeted MS/MS, similar to the native GxtA reaction, SxtT M255Y and SxtT double variant hydroxylate STX (**2**) at the C11 position to form 11-β-hydroxySTX (**4**).

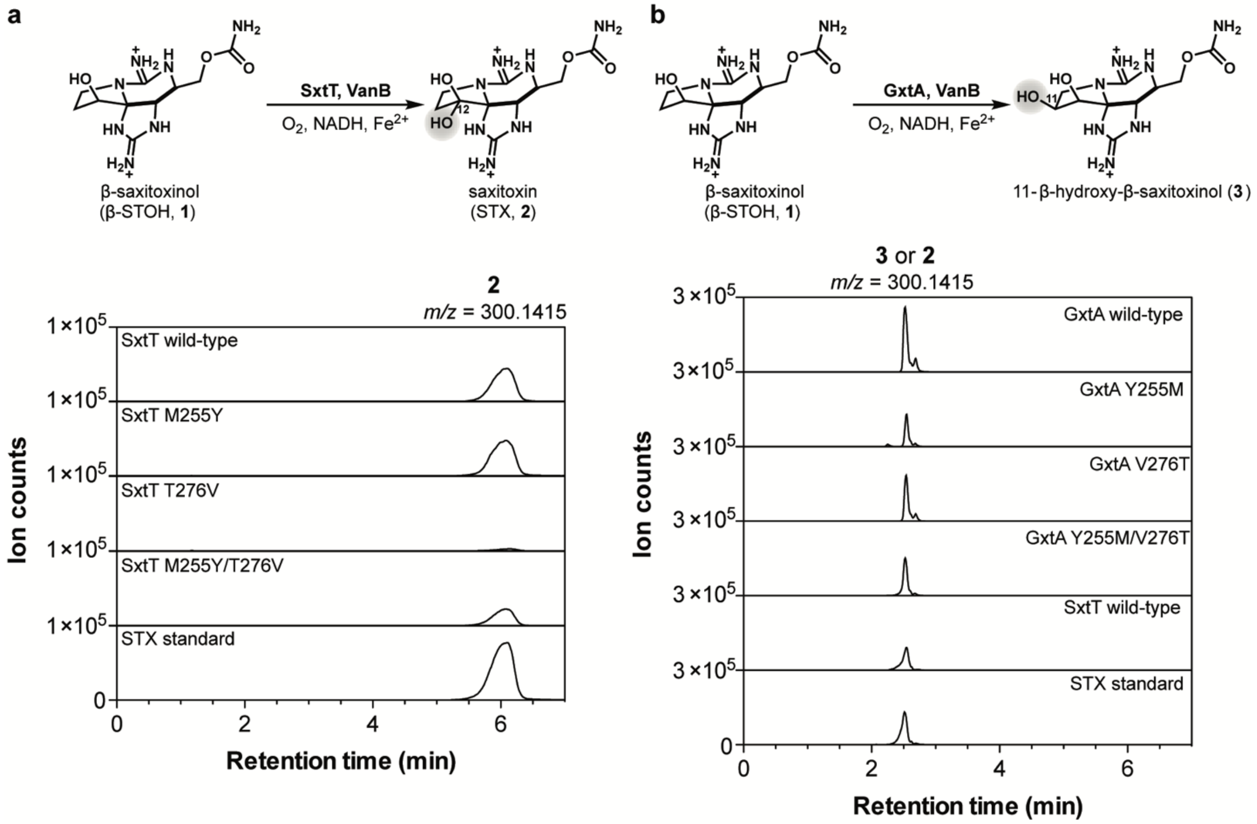


Figure 2.12. Single active site variants of SxtT and GxtA maintain wild-type monohydroxlyation activity as demonstrated using HILIC-MS extracted ion chromatograms of the SxtT and GxtA reaction product(s) with β-STOH (**1**).

m/z = 300.1415 represents the exact mass of both STX (**2**) and **3**. (a) SxtT wild-type and variants, analyzed using 15% A isocratic HILIC method. Scheme shown depicts wild-type reaction. (b) GxtA wild-type and variants, analyzed using 18% A isocratic HILIC method. Scheme shown depicts wild-type reaction.

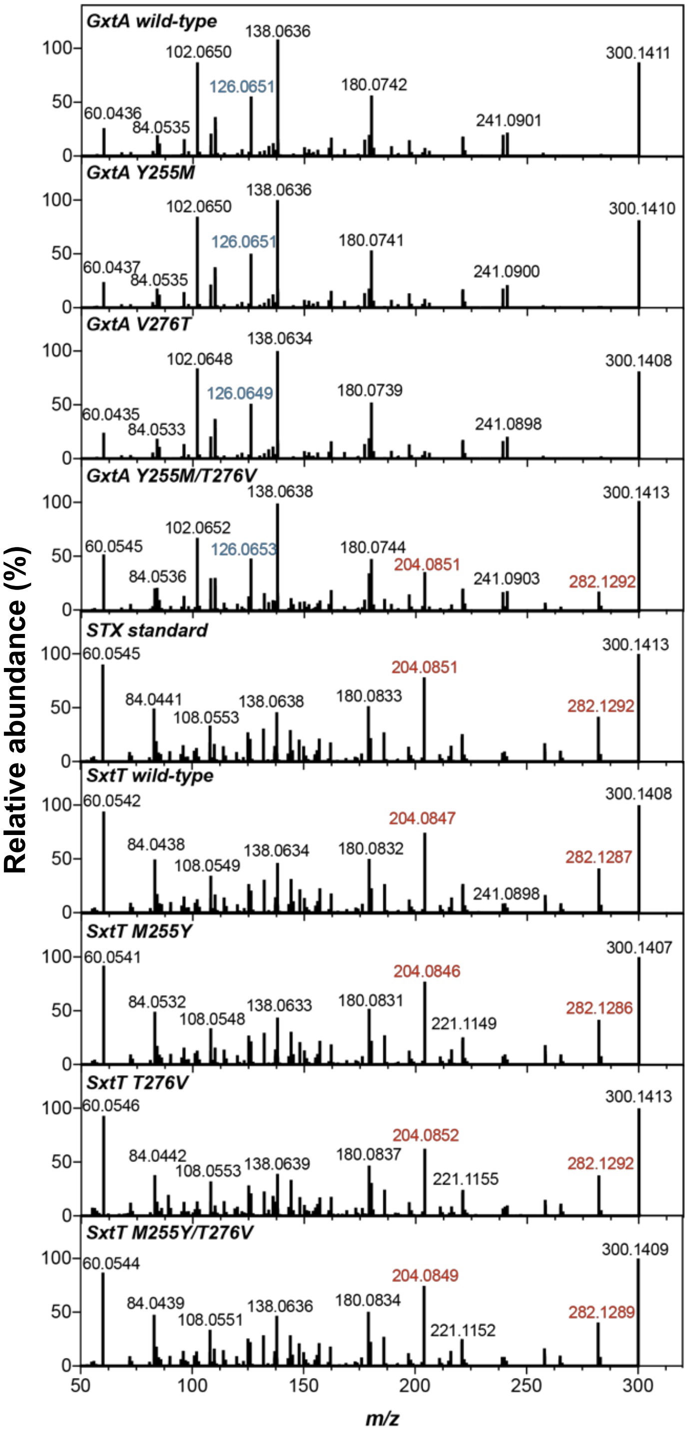


Figure 2.13. Targeted MS/MS spectra of *m/z* = 300.1415 products from reactions with SxtT and GxtA variants with β-STOH (**1**).

Numbers in red correspond to those that appear only when STX (**2**) is formed. Numbers in blue correspond to those that are prevalent only when **3** is formed. Reactions were analyzed by HILIC-MS. The exact mass of 11-β-hydroxySTX (**4**) is [M+H] = 316.1336.

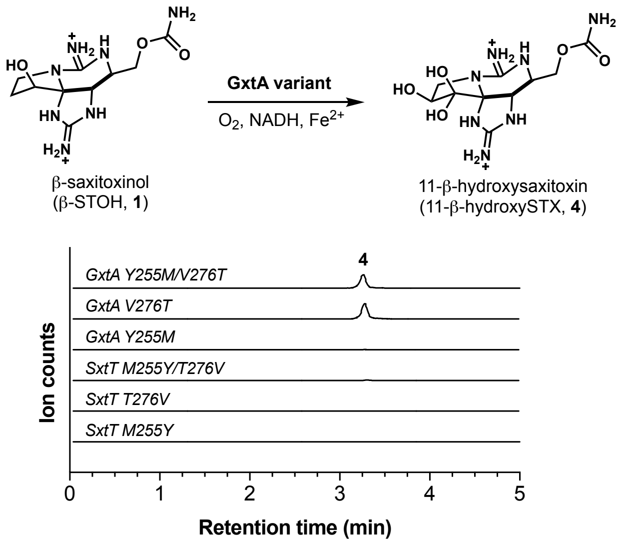


Figure 2.14. The GxtA variant V276T is able to catalyze two hydroxylation reactions when provided β-STOH (**1**) as a substrate.

Production of the dihydroxylated product, 11-β-hydroxySTX (**4**) is shown using HILIC-MS. None of the other protein variants tested in this work deviate from catalyzing a monohydroxylation reaction.

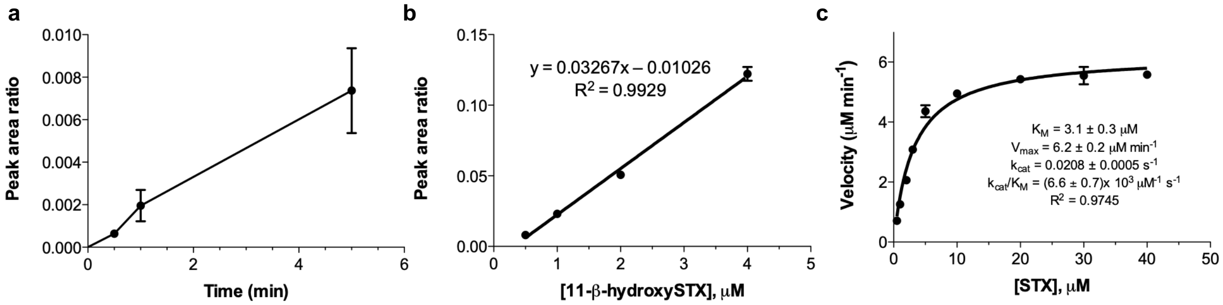


Figure 2.15. GxtA steady-state kinetic analysis with STX (**2**).

(a) Time course of 5 μM GxtA with 5 μM STX (**2**). (b) 11-β-hydroxySTX (**4**) standard curve used in the steady-state kinetic analysis. (c) Michaelis-Menten plot of GxtA reaction with STX. Data are presented as mean values +/- SD.

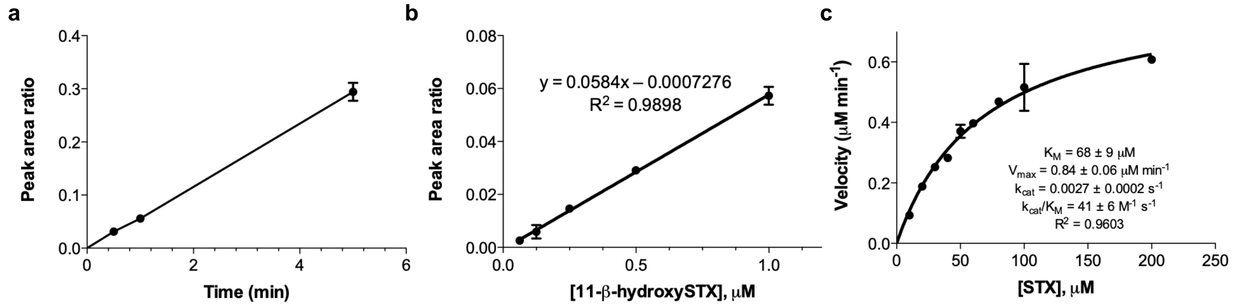


Figure 2.16. SxtT M255Y/T276V steady-state kinetic analysis with STX (**2**).
(a) Time course of 5 μM SxtT M255Y/T276V with 50 μM STX (**2**). (b) 11-β-hydroxySTX (**4**) standard curve used in the steady-state kinetic analysis. (c) Michaelis-Menten plot of SxtT M255Y/T276V reaction with STX. Data are presented as mean values +/- SD.

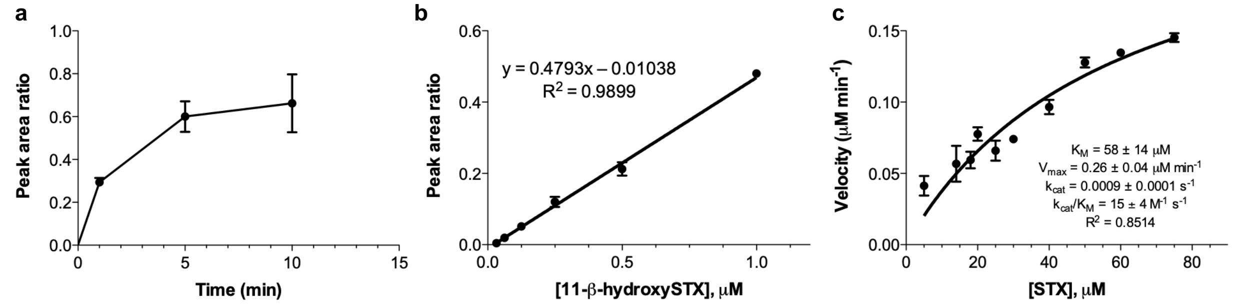
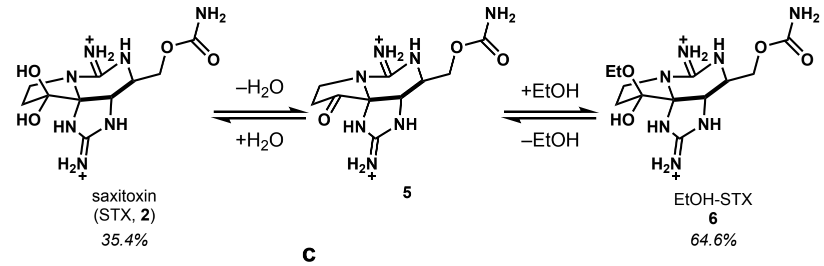
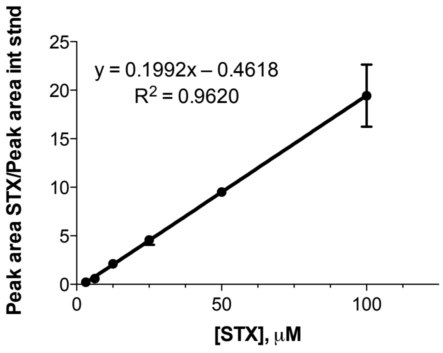


Figure 2.17. SxtT M255Y steady-state kinetic analysis with STX (**2**).
(a) Time course of 5 μM SxtT M255Y with 50 μM STX (**2**) in duplicate. (b) 11-β-hydroxySTX (**4**) standard curve used in the steady-state kinetic analysis. (c) Michaelis-Menten plot of SxtT M255Y reaction with STX. Data are presented as mean values +/- SD.

a



b



c

Trial	Peak area STX remaining	Peak area internal standard	Quantity STX remaining (μM)
1	2539449.3	176156.11	74.7
2	2704099.03	220412.48	63.9
3	2829411.92	198325.46	73.9
Average [STX] remaining			*70.8 ± 6.0 μM*
Average % STX in 200 μM			*35.4%*
Average % EtOH-STX in 200 μM			*64.6%*

Figure 2.18. Extent of ethanol incorporation into STX (**2**) under our reaction conditions. (a) Scheme of STX (**2**) equilibrium in the presence of ethanol. (b) Standard curve of STX (**2**) used to quantify remaining STX (**2**) after incubation with ethanol. Data are presented as mean values +/- SD. (c) Raw data obtained from triplicate STX (**2**) incubations with ethanol and extrapolated percent EtOH-STX (**6**) present.

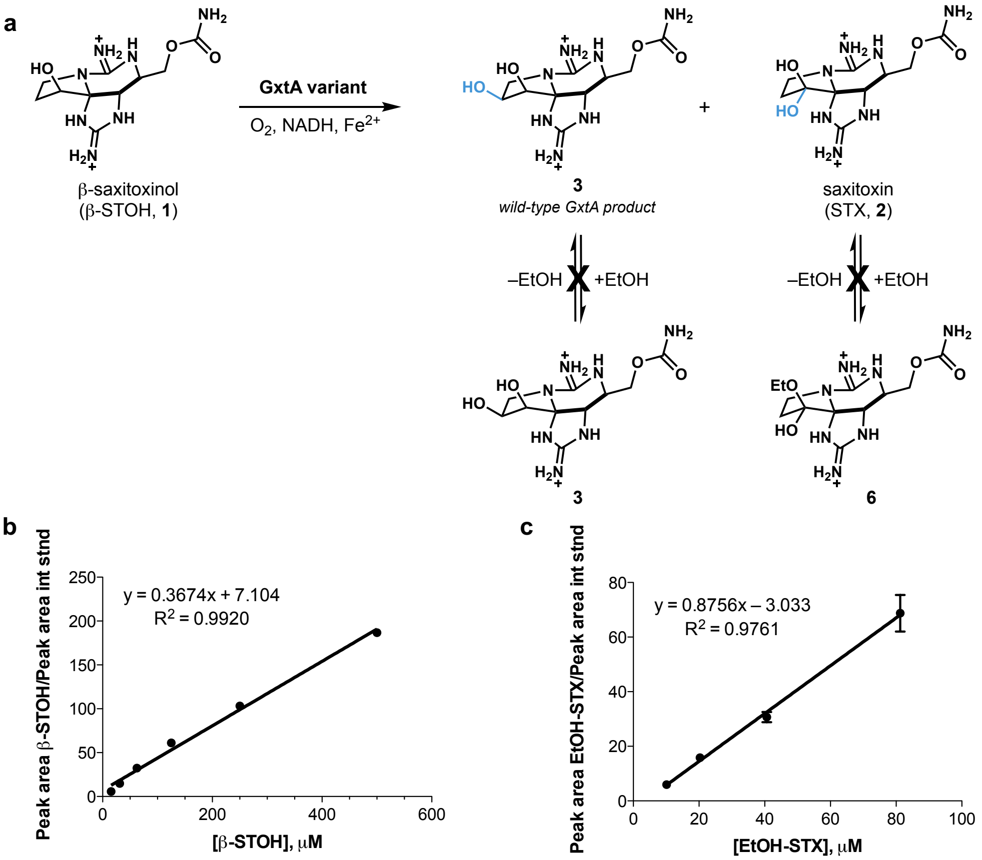


Figure 2.19. GxtA Y255M/V276T reaction scheme and standard curves.
(a) GxtA variant reaction scheme with β-STOH (**1**) and potential for incorporation of ethanol. (b) β-STOH (**1**) standard curve. (c) EtOH-STX (**6**) standard curve. Data are presented as mean values +/- SD.

Table 2.2. Protein sequence accession numbers (GenBank), DNA sources, and vectors used in this study.

Protein	Organism of origin	Accession	DNA source	Vectors
SxtT	*Microseira wollei*	ACG63840.1	GeneArt, subcloned	pET151, pMCSG9
GxtA	*Microseira wollei*	ACG63835.1	IDT, fragment	pMCSG7, pMCSG9
VanB	*Pseudomonas aeruginosa*	NP_253592.1	IDT, fragment	pMCSG7

Table 2.3. Site-directed mutagenesis primers used in this study.

Primer name	Primer DNA sequence
SxtT_M255Y	5' – GCGTTCTGCGTTATCTGATTATGTGG – 3'
SxtT_T276V	5' – CGAATATGATGAAGTGATCGAACAGGATATTC – 3'
GxtA_Y255M	5' – AGCGTTCTGCGTATGCTGATTATGTGGAATG – 3'
GxtA_V276T	5' – CAGATTATGATCAGACCATCGAAGAAGATATTC – 3'

Table 2.4. Summary of GxtA EtOH-incorporation results with β-STOH.

GxtA Variant	Trial	Quantity β-STOH remaining (μM)	Quantity EtOH-STX (μM)	Quantity 3 (μM)	Total % conversion	% conversion to 3	% EtOH-STX
Wild-type	1	430.0	0	70.0	14.0	14.0	0.0
Wild-type	2	327.1	0	172.9	34.6	34.6	0.0
Wild-type	3	357.6	0	142.4	28.5	28.5	0.0
Y255M	1	310.5	0.171	189.4	37.9	37.9	0.0
Y255M	2	324.2	0.171	175.6	35.2	35.1	0.1
Y255M	3	348.6	0.197	151.3	30.3	30.3	0.0
V276T	1	349.0	0.350	150.6	30.2	30.1	0.1
V276T	2	363.1	0.475	136.5	27.4	27.3	0.1
V276T	3	297.3	0.484	202.2	40.5	40.4	0.1
Y255M/V276T	1	385.5	9.110	105.3	22.9	21.1	1.8
Y255M/V276T	2	323.6	9.490	166.9	35.3	33.4	1.9
Y255M/V276T	3	359.5	9.933	130.6	28.1	26.1	2.0

2.7 Acknowledgements

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817)

2.8 Methods

2.8.1 Purification protocol for SxtT and GxtA for crystallography

Cell pellet from 6 L culture of MBP-SxtT or MBP-GxtA expressions (~30 g) was resuspended in 120 mL lysis buffer (20 mM Tris-HCl pH 7.0, 1 M NaCl, 1 mM DTT). Cells were lysed by sonication, 5 min total "on" time, 10 s on, 20 s off. Lysed cells were centrifuged at 40,000 x g for 30 min and the supernatant was filtered using 0.45 µM syringe filters. The sample was loaded onto an ÄKTA Pure FPLC system fitted with a 5 mL MBPTrap column, where Buffer A was the lysis buffer and Buffer B was the lysis buffer with the addition of 10 mM maltose. Lysate was loaded onto the column at 2.5 mL/min, washed with 10 CV of lysis buffer at 2.5 mL/min, and eluted in a 5 CV gradient to 100% Buffer B at 1 mL/min. Fractions containing desired protein were pooled and diluted to 30 mL. 2 mg of TEV protease was added to the diluted protein and the mixture was dialyzed overnight in 20 mM Tris-HCl pH 7.4, 50 mM NaCl, 1 mM DTT, 10% glycerol buffer. The next day, the contents of the dialysis bag were combined with 1 mL Ni-NTA resin that had been washed with water and incubated for 1 h with rocking. The mixture was poured over a 12 mL column and allowed to drain completely. The flowthrough containing proteins cleaved from MBP were collected and concentrated to 2 mL using a 30 kDa cutoff centrifugal filter. The concentrated protein was diluted to 50 mL in anion exchange binding buffer (20 mM Tris-HCl pH 7.0, 50 mM NaCl). The diluted mixture was loaded onto a 5 mL HiTrap Q HP column at 2.5 mL/min and eluted in an 8 CV gradient to 100% anion exchange elution buffer (20 mM Tris-HCl pH 7.0, 1 M NaCl). The fractions containing desired protein were pooled, concentrated to 2 mL, and injected onto a Sephacryl S-200 HR gel filtration column equilibrated with 20 mM Tris-HCl pH 7.4, 200 mM NaCl. Desired protein was pooled and exchanged into 50 mM HEPES pH

8.0, 10% glycerol buffer using a PD-10 desalting column following the manufacturer's protocol. Desalted protein was concentrated to a minimum of 10 mg/mL.

2.8.2 Crystallization of SxtT, ddSTX-bound SxtT, GxtA, and ddSTX-bound GxtA

The conditions for crystallizing SxtT were identified anaerobically using a Mosquito pipetting robot (TTP LabTech) that is housed in a Coy chamber (Coy Lab Products) maintained using a 95-percent nitrogen and 5-percent hydrogen mixture at room temperature. Initial crystals of SxtT grew within one week of mixing 0.3 µL of 10 mg/mL SxtT (50 mL HEPES pH 8.0, 10% v/v glycerol) with 0.3 µL of crystallization solution (2.0 M (NH4)2SO4, 0.1 M Bis-Tris pH 6.5). The crystals were subsequently optimized by mixing SxtT in 1:1 proportion with a well solution of (2.0 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH 6.5, 10% v/v glycerol) in a hanging drop vapor-diffusion experiment. Brown crystals appeared in three days at 20°C and achieved maximum size over the course of two weeks. Crystals of ddSTX-bound SxtT were grown similarly except the protein concentration was adjusted to 8 mg/mL and the protein buffer was altered to include 20 mM ddSTX. GxtA crystal conditions were identified similarly and grew within three days of mixing 0.3 µL of 10 mg/mL GxtA (50 mM HEPES, 10% v/v glycerol pH 8.0) with 0.3 µL of crystallization solution (0.2 M $MgCl_2$, 0.1 M Bis-Tris pH 5.5, 25% v/v PEG3350). GxtA crystals were manually optimized using hanging drop vapor diffusion by changing the concentration of the protein (24 mg/mL) and changing the well solution (0.3 M $MgCl_2$, 0.1 M Bis-Tris pH 5.5, 25% v/v PEG3350, 15% v/v glycerol). Crystals of ddSTX-bound GxtA were grown similarly except the protein concentration was adjusted to 19.2 mg/mL and the protein buffer was altered to include 20 mM ddSTX. Brown crystals appeared in one day following incubation at 20°C and achieved maximum size within one week. All crystals were harvested and cryocooled in the coy chamber with no additional cryoprotectant added.

2.8.3 Data processing and structure solution

The datasets for SxtT, GxtA, ddSTX-bound SxtT, and ddSTX-bound GxtA were collected at the Life Sciences Collaborative Access Team beamlines 21-ID-G (SxtT, Rayonix MX 300), 21-1D-F (ddSTX-bound GxtA, Rayonix MX 300), and 21-ID-D (GxtA and ddSTX-bound SxtT, Dectris Eiger 9M) at the Advanced Photon Source, Argonne National Laboratory. All datasets were collected at a temperature of 100K and wavelengths of 0.97856 (SxtT), 0.97872 (ddSTX-bound GxtA) and 1.1272 Å (GxtA and ddSTX-bound SxtT). Indexing, integration, and scaling of the data were performed in HKL2000[51]. SxtT and ddSTX-bound SxtT indexed as *C*222, whereas GxtA and ddSTX-bound GxtA indexed as $P12_11$. Both structures have three molecules in the asymmetric unit. The structure of SxtT was solved using an edited model of dicamba monooxygenase (31% identity, 48% similarity)[35] for molecular replacement. This modified model lacked the Rieske cluster, non-heme iron site, water molecules, and had the sidechains pruned to the most common atom using Phenix Sculptor[52]. With this modified model, a molecular replacement solution (LLG score of 69 and Z score of 10.8) was determined using Phenix AutoMR[53]. The resulting structure of SxtT, in the absence of water molecules and cofactors was used to subsequently determine the structures of ddSTX-bound SxtT by isomorphous replacement and GxtA by molecular replacement (LLG = 8968, Z score = 85.4). Once complete, the structure of GxtA was used to determine the structure of ddSTX-bound GxtA. For all structures described in this work, model building, model adjustment, the building of metal sites, and the addition of water molecules were performed in COOT[54]. Iterative rounds of structure and B-factor refinement were performed in Phenix[53]. An initial 5 cycles of simulated annealing were used to reduce model bias. Identical Rfree test sets composed of five percent of the original data were used for native and ddSTX-bound structures.

Small molecule parameter files were generated using the chemical SMILES string and the electronic Ligand Builder and Optimization workbench, eLBOW[55], in Phenix for ddSTX. In parallel, small molecule parameter files were generated using the Grade Web Server Global Phasing, Cambridge, UK) for comparison. Although both methods resulted in similar placement of the ligand in the electron density, the latter method resulted in more accurate refined geometry. Thus, the latter was used in refinements for both ddSTX-bound structures determined in this work. For SxtT, ddSTX is modeled at full occupancy in chains A and C, whereas it is modeled at 85-percent occupancy in chain B. For GxtA, ddSTX is modeled at full occupancy in all three chains. The final structures were analyzed using simulated annealing composite omit electron density maps. Further scrutiny of the final SxtT structure using the MolProbity program[56], which is missing residues in each monomeric subunit (A:1, 205-208, and 297-303, B: 298-303, and C: 201-212, and 298-303) showed that 96.3-, 3.7-, and 0-percent of residues are in the favored, allowed, and disallowed regions of the Ramachandran plot, respectively. Similarly, for ddSTX-bound SxtT, which is also missing residues in each monomer (A:1, 297-303, B: 1, 297-303, and C: 1, 297-303, and 204-213), there are also 96-, 4-, and 0-percent of residues are in the favored, allowed, and disallowed regions of the Ramachandran plot, respectively. The GxtA and ddSTX-bound GxtA structures are also missing residues in each subunit. For GxtA, these resides are: A: 1, 177-178, 297-303, and 201-213, B: 1, 204-208, 298-304, and C: 1 and 296-303. This structure has 96.5-, 3.5-, and 0-percent of residues in the favored, allowed, and disallowed regions of the Ramachandran plot, respectively. For the ddSTX-bound GxtA structure, which has 96.7-, 3.3-, and

0-percent of residues in the favored, allowed, and disallowed regions of the Ramachandran plot, respectively, the missing residues are A: 1, 296-305, and 202-213, B: 1, 204-209, 298-304, and C: 1 and 296-303. Data statistics are summarized in Table 2.1. All structure figures were made using PyMol and the crystallography software packages were compiled by SBGrid[57].

2.8.4 Calculation of active site tunnels in SxtT and GxtA

The existence of an active site tunnel in SxtT and GxtA was analyzed using the MOLEonline server[46,47]. In brief, the monomeric subunits which contained an ordered 195-215 loop was loaded into the server. The channel parameter settings were adjusted to tunnels that originate from the same starting point and detect pores in all cavities was turned on and the non-heme iron site and ligands were manually input as a starting point for the analysis. Using these parameters, the tunnel from the surface of GxtA to the active site and the active site cavity in SxtT were found. The tunnels and tunnel-defining residues (entrance, exit, and bottleneck) were visualized using PyMol.

2.8.5 Enzymatic assays experiment of SxtT, GxtA and their variants

This section was done by Dr. April Lukowski.

Site-directed mutagenesis

Touchdown PCR[50] was used to introduce mutations and amplify sequences. 25 μL reactions composed of 1x Phusion buffer (HF), 200 μM dNTPs, 2 ng/μL plasmid template (His-SxtT-pET151 and His-GxtA-pMCSG7), 1 μM primer (Table 2.3), 1 U Phusion DNA polymerase and 1% DMSO. Reactions were run using the following program: 98 °C 30 s, (98 °C 10 s, 69-

63 °C 30 s decreasing in 0.5 °C increments each cycle) and 72 °C 7 min) for 12 cycles followed by (98 °C 10 s, 63 °C 30 s, 72 °C 7 min) for 22 cycles, and 72 °C for 15 min. Digestion with DpnI was performed by combining 8 µL PCR reaction with 1 µL 10x CutSmart buffer and 1 µL DpnI (20,000 U/mL stock). The DpnI digest reaction was incubated at 37 °C for 3 h and the 10 µL reaction volume was transformed into DH5α chemically competent cells. Mutations were verified by Sanger sequencing (University of Michigan).

Expression protocol for SxtT and GxtA wild-type and variants

For protein to be prepared for crystallography, pMCSG9 plasmids containing *sxtT* or *gxtA* were transformed by standard heat-shock protocols into chemically competent C41(DE3) *E. coli* cells to generate MBP-SxtT and MBP-GxtA. For protein to be prepared for reactions, a pET151 vector containing *sxtT* and a pMCSG7 vector containing *gxtA* were used to generate His-SxtT and His-GxtA. The remaining expression procedures are identical for both sets of constructs. A single colony was used to inoculate 10 mL of LB containing 100 µg/mL ampicillin and incubated overnight at 37 °C, 200 rpm overnight. The overnight culture was used to inoculate a 1 L LB culture in a 2.8 L flask containing 100 µg/mL ampicillin. Cultures were grown at 37 °C and 200 rpm until the OD_{600} reached 0.6-0.8. Flasks were briefly cooled to 20 °C before addition of 0.1 mM IPTG, 0.2 mg/mL ferric ammonium citrate, and 0.4 mg/mL ferrous sulfate heptahydrate. Cultures were incubated at 20 °C and 200 rpm for approximately 18 h before harvesting. The average pellet for a 1 L culture was 5 g wet cell mass.

Expression protocol for VanB

pMCSG7 plasmid containing vanB was transformed by heat-shock into chemically competent BL21(DE3) *E. coli* cells.[16] A single colony was used to inoculate 10 mL LB containing 100 µg/mL ampicillin and incubated at 37 °C, 200 rpm overnight. The overnight culture was used to inoculate a 1 L LB culture in a 2.8 L flask containing 100 µg/mL ampicillin. Cultures were incubated at 37 °C and 200 rpm until an OD_{600} of 0.6-0.8 was achieved. Flasks were chilled at 20 °C for 1 h before induction by addition of 0.1 mM IPTG. Cultures were incubated for approximately 18 h before harvesting. The typical wet mass of a pellet from a 1 L culture was 6 g.

Purification protocol for SxtT and GxtA for reactions

Cell pellet from 6 L culture of His-SxtT or His-GxtA expressions (~30 g) was resuspended in 120 mL lysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 10% glycerol, 10 mM imidazole). Cells were lysed by sonication, 5 min total "on" time, 10 s on, 20 s off. Lysed cells were centrifuged at 40,000 x g for 30 min and the supernatant was filtered using 0.45 µM syringe filters (CellTreat). The sample was loaded onto an ÄKTA Pure FPLC system fitted with a 5 mL HisTrap column, where Buffer A was the lysis buffer and Buffer B was the lysis buffer with the addition of 400 mM imidazole. Lysate was loaded onto the column at 2.5 mL/min, washed with 6 CV of 40 mM imidazole buffer at 2.5 mL/min, and eluted in a 15 CV gradient to 100% Buffer B at 1 mL/min. 2 mL fractions were collected during elution. The fractions containing desired protein were pooled and concentrated to 2 mL using a 30 kDa cutoff centrifugal filter and diluted to 50 mL with anion exchange binding buffer (20 mM Tris-HCl pH 7.0, 50 mM NaCl). The diluted mixture was loaded onto a 5 mL HiTrap Q HP column at 2.5 mL/min and eluted in an 8 CV gradient to 100% anion exchange elution buffer (20 mM Tris-HCl pH 7.0, 1 M NaCl). The fractions containing desired protein were pooled and diluted to 30 mL with TEV dialysis buffer (20 mM Tris-HCl pH 7.4, 50 mM NaCl, 1 mM DTT, 10% glycerol). 2 mg of TEV protease was added to the diluted protein and the mixture was dialyzed overnight. The next day, the contents of the dialysis bag were combined with 1 mL Ni-NTA resin and incubated for 1 h with rocking. The mixture was poured onto a 12 mL column and allowed to drain completely. The flowthrough containing His-cleaved protein was collected and concentrated to at least 100 μM and flash frozen in liquid nitrogen for long-term storage at -80 °C.

Purification protocol for VanB

Cell pellet was resuspended in 4 mL of lysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 10 mM imidazole, 10% glycerol) per gram of pellet.16 The mixture was sonicated for 5 min total "on" time, 10 s on, 20 s off. Lysed cells were centrifuged at 40,000 x g for 30 min and the clarified supernatant was combined with 4 mL Ni-NTA resin. The mixture was incubated for 2 h and poured over a 35 mL column. The packed resin was washed with 10 mL lysis buffer followed by 10 mL each of buffer containing increasing amounts of imidazole: 25 mM, 30 mM, and 35 mM imidazole. Proteins were eluted with 20 mL elution buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Fractions containing protein were run on an SDS-PAGE gel and those containing VanB were pooled and exchanged into storage buffer (50 mM Tris-HCl pH 7.4, 10% glycerol) using a PD-10 desalting column. Proteins were concentrated to 206 μM and flash frozen in liquid nitrogen for long-term storage at -80 °C.

LC-MS and MS/MS analysis

Liquid chromatography-mass spectrometry (LC-MS) analysis was performed on an Agilent G6545A quadrupole-time of flight (QTOF) or Agilent 6230 time of flight (TOF) mass spectrometer equipped with a dual AJS ESI source and an Agilent 1290 Infinity series diode array detector, autosampler, and binary pump. Solvent A = water with 0.1% formic acid. Solvent B = 95% acetonitrile, 5% water and 0.1% formic acid. An Acquity UPLE BEH Amide 1.7 µm, 2.1 x 100 mm hydrophobic interaction liquid chromatography (HILIC) column from Waters was used for all separations. The chromatographic method was typically 18% A 0-5 min at 0.4 mL/min or 15% A 0-7 min at 0.3 mL/min. 0.3-10 µL injections were made for each sample. Targeted MS/MS was performed using the QTOF using an 18% A isocratic method, 0-5 min, to obtain fragmentation patterns. Methods were augmented to target each mass. Collision energies were set to 10, 20 and 30 eV and the resulting chromatograms were averages of the three collision energies.

SxtT and GxtA wild-type and variants reactions

Stock solutions of all substrates and product standards were prepared to final concentrations of 20 mM in dimethyl sulfoxide (DMSO, analytical grade). Enzyme aliquots were discarded after one freeze-thaw cycle. Stocks of 1 mM $Fe(NH_4)_2(SO_4)_2•6H_2O$ were prepared fresh in MilliQ water before each use. 10 mM stock solutions of nicotinamide adenine dinucleotide hydride (NADH) were prepared in water and stored at -20 °C, undergoing no more than 10 freeze-thaw cycles. All reactions were conducted in 1.5 mL plastic tubes and 96-well plates. 50 µL reactions consisting of 5-20 µM SxtT or GxtA variant, 5 µM VanB, 200 µM β-STOH (**1**) or STX (**2**), 500 µM NADH, 100 µM $Fe(NH_4)_2(SO_4)_2•6H_2O$, and 50 mM Tris-HCl pH 7.0 buffer were combined and incubated at 30 °C for 2 h and quenched by the addition of 150 µL acetonitrile. Reactions were centrifuged at 12,000 x g for 20 min to pellet precipitated material and 100 µL of the supernatant was diluted with sterile filtered acetonitrile containing 1% formic acid and 0.05% 15N-arginine as an internal standard for mass spectrometry analysis.

Ethanol incorporation into reaction products

Reactions were performed as described above. After quenching with 150 μL acetonitrile and subsequent centrifugation, reactions were diluted 1:1 with sterile filtered acetonitrile containing 1% formic acid and 0.05% 15N-arginine as an internal standard. 50 μL 100% ethanol was added to 50 μL of the reaction mixture containing internal standard and incubated at room temperature for at least 6 hours prior to analysis by LC-MS. Quantification of EtOH-STX (**6**) in samples was based on the percent EtOH-STX (**6**) present in samples of STX (**2**) incubated with ethanol prepared on the same day, under the same conditions as the reactions to be analyzed. A standard curve of STX (**2**) was used to quantify the amount of STX (**2**) remaining in the samples treated with ethanol (Figure. 2.17). The dilution factor was adjusted in the standard curve by adding additional acetonitrile in place of ethanol. The difference in ionization in a sample containing acetonitrile and a sample containing a mixture of acetonitrile and ethanol is anticipated to be negligible.

Steady-state kinetics

To determine the steady-state kinetic parameters of GxtA, SxtT M255Y, and SxtT M255Y/T276V with STX (**2**), reactions were conducted on 50 μL scale in duplicate with substrate ranging 1 μM – 200 μM in duplicate with 5 μM enzyme, 5 μM VanB, 100 μM $Fe(NH_4)_2(SO_4)_2$, and 50 mM Tris HCl pH 7 buffer in a 96-well plate. Reactions were initiated by the addition of 500 μM NADH (10 μL distributed by multichannel pipette). For GxtA and SxtT M255Y/T276V, reactions were quenched after 30 s by the addition of 10 μL of 10% formic acid in acetonitrile. 140 μL dilution mix containing internal standard was then added. 96-well plates were centrifuged at

2000 x g for 2 min, then 100 µL of the centrifuged mixture was added to a clean 96-well 0.22 µm filter plate. The plate was centrifuged over a clean 96-well plate at 2000 x g for 2 min. Product standards were prepared in the same manner with enzyme storage buffer (20 mM Tris HCl pH 7.4, 200 mM NaCl) replacing enzymes. 5 µL each sample was injected on the TOF LC-MS. The resulting Michaelis-Menten curves for each substrate are shown in Figure. 2.14-2.16 alongside standard curves used to quantify 11-β-hydroxySTX (**4**) product.

For SxtT M255Y, protocol adjustments were necessary to observe the low quantities of product being generated. The reactions were performed in quadruplicate and the most consistent three data points were used to generate the curve in Figure. 2.16.c. Reactions were quenched after 1 min by the addition of 10 µL of 10% formic acid in acetonitrile followed by the addition of 50 µL acetonitrile to precipitate protein. 96-well plates were centrifuged at 2000 x g for 2 min, then 50 µL each sample was combined with 10 µL dilution mix containing internal standard and added to a clean 96-well 0.22 µm filter plate. The plate was centrifuged over a clean 96-well plate at 2000 x g for 2 min. 10 µL each sample was injected on the QTOF LC-MS for analysis. Data was collected in Microsoft Excel and analyzed using Graphpad Prism.

Quantification of GxtA reaction products with β-STOH as a substrate

Triplicate reactions consisting of 5 µM GxtA, 5 µM VanB, 500 µM β-STOH (**1**), 500 µM NADH, 100 µM $Fe(NH_4)_2(SO_4)_2$ and 50 mM Tris HCl pH 7.0 buffer were incubated at 30 °C for 2 h. Reactions were quenched and incubated with ethanol as described above. The quantity of β-STOH (**1**) remaining in the reactions and the quantity EtOH-STX (**6**) present were determined using standard curves of β-STOH (**1**) (Figure. 2.18.b) and EtOH-STX (**6**) (Figure. 2.18.c). The quantity **3** generated in each reaction was determined by subtracting the determined quantities of β-STOH (**1**) and EtOH-STX (**6**) from 500 μM (the amount of β-STOH (**1**) added to the reaction).

Synthesis and standard preparation

Product standards of ddSTX, β-STOH (**1**), STX (2), and 11-β-hydroxySTX (**4**) were prepared as previously described[16]. Briefly, STX (**2**) was extracted from Alexandrium cultures and purified. β-STOH (**1**) was prepared from STX (**2**) by reduction with sodium borohydride. 11-β-hydroxySTX (**4**) was prepared through hydrolysis of the sulfated PST analog gonyautoxin (**3**) that had been extracted from Alexandrium cultures. ddSTX was synthesized according to procedures described by Mulcahy and coworkers.[58]

2.9 DNA and protein sequences

Microseira wollei **SxtT:**

ATGACCACCGCAGATCTGATTCTGATTAATAACTGGCATGTTGTGGCCAATGTGGAAGATTGTA
AACCGGGTAGCATTACCACCGCACGTCTGCTGGGTGTTAAACTGGTTCTGTGGCGTAGCCAAGA
ACAGAATAGCCCGATTCAGGTTTGGCAGGATTATTGTCCGCATCGTGGTGTTCCGCTGAGCATG
GGTGAAGTTGCAAATAATACCCTGGTTTGTCCGTATCATGGTTGGCGTTATAATCAGGCAGGTA
AATGTGTTCAGATTCCGGCACATCCGGATATGGTTCCGCCTGCAAGCGCACAGGCAAAAACCTA
TCATTGTCAAGAACGTTATGGTCTGGTTTGGGTTTGTCTGGGTAATCCGGTTAATGATATTCCG
AGCTTTCCGGAATGGGATGATCCGAATTATCATAAAACCTATACCAAAAGCTATCTGATTCAGG
CAAGCCCGTTTCGTGTTATGGATAATAGCATTGATGTTAGCCACTTTCCGTTTATTCATGATGG
TTGGCTGGGTGATCGTAATTATACCAAAGTCGAAGATTTCGAGGTGAAAGTGGATAAAGATGGT
CTGACCATGGGCAAATATCAGTTTCAGACCAGCCGTATTGTTAGCCATATTGAAGATGATAGCT
GGGTGAATTGGTTTCGTCTGAGCCATCCGCTGTGTCAGTATTGTGTTAGCGAAAGTCCGGAAAT
GCGTATTGTTGATCTGATGACCATTGCACCGATTGATGAAGATAATAGCGTTCTGCGTATGCTG
ATTATGTGGAATGGTAGCGAAATGCTGGAAAGCAAAATGCTGACCGAATATGATGAAACCATCG
AACAGGATATTCGCATTCTGCATAGCCAGCAGCCTGCCCGTCTGCCGCTGCTGGCACCGAAACA
AATCAATACCCAGGGTCTGCCGCAAGAAATTCATGTTCCGAGCGATCGTGGCACCGTTGCATAT
CGTCGTTGGCTGAAAGAACTGGGTGTGACCTATGGTGTTTGT

MTTADLILINNWHVVANVEDCKPGSITTARLLGVKLVLWRSQEQNSPIQVWQDYCPHRGVPLSM
GEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCLGNPVNDIP
SFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHDGWLGDRNYTKVEDFEVKVDKDG
LTMGKYQFQTSRIVSHIEDDSWVNWFRLSHPLCQYCVSESPEMRIVDLMTIAPIDEDNSVLRML
IMWNGSEMLESKMLTEYDETIEQDIRILHSQQPARLPLLAPKQINTQGLPQEIHVPSDRGTVAY
RRWLKELGVTYGVC

Microseira wollei **GxtA:**

ATGACCACCGCAGATCTGATTCTGATTAATAACTGGTATGTTGTGGCCAAGGTTGAAGATTGTC
GTCCGGGTAGCATTACCACAGCACATCTGCTGGGTGTTAAACTGGTTCTGTGGCGTAGCCATGA
ACAGAATAGCCCGATTCAGGTTTGGCAGGATTATTGTCCGCATCGTGGTGTTCCGCTGAGCATG
GGTGAAGTTGCAAATAATACCCTGGTTTGTCCGTATCATGGTTGGCGTTATAATCAGGCAGGTA
AATGTGTTCAGATTCCGGCACATCCGGATATGGTTCCGCCTGCAAGCGCACAGGCAAAAACCTA
TCATTGTCAAGAACGTTATGGTCTGGTTTGGGTTTGTCTGGGTAATCCGGTTAATGATATTCCG
AGCTTTCCGGAATGGGATGATCCGAATTATCATAAAACCTACACCAAGAGCTATCTGATTCAGG
CAAGCCCGTTTCGTGTTATGGATAATAGCATTGATGTTTCGCACTTTCCGTTTATCCATGAAGG
TATTCTGGGTGATCGTAATCATGCAGAAGTTGAAGATCTGGAAGTGAAAGTGGATAAAGATGGT
CTGACCATGGGTAAATATCAGGTTCATACCAGCAAATTCAACAACAGCACCAAAGATGATAGCA
TGGTGAATTGGTTTCGTCTGAGCCATCCGCTGTGTCAGTATTGTAGCACCGAAGCAAGCGAAAT
GCGTACCGTTGATCTGATGGTTGTTACCCCGATTGATGAAGATAATAGCGTTCTGCGTTATCTG
ATTATGTGGAATGGTAGCAAAACCCTGGAAAGCAAAATTCTGGCAGATTATGATCAGGTGATCG
AAGAAGATATTCGCATTCTGCATAGCCAGCAGCCGACACGTCTGCCGCTGCTGAGCCCGAAGCA
GATTAATACCCAGGGTCTGCCGCAAGAAATTCATGTTCCGAGCGATCGTTGTACCGTTGCATAT
CGTCGTTGGCTGAAAGAACTGGGTGTGACCTATGGTGTTTGTTAA

```
MTTADLILINNWYVVAKVEDCRPGSITTAHLLGVKLVLWRSHEQNSPIQVWQDYCPHRGVPLSM
GEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCLGNPVNDIP
SFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHEGILGDRNHAEVEDLEVKVDKDG
LTMGKYQVHTSKFNNSTKDDSMVNWFRLSHPLCQYCSTEASEMRTVDLMVVTPIDEDNSVLRYL
IMWNGSKTLESKILADYDQVIEEDIRILHSQQPTRLPLLSPKQINTQGLPQEIHVPSDRCTVAY
RRWLKELGVTYGVC
```

2.10 References

1 Peters, C. & Buller, R. M. Industrial application of 2-oxoglutarate-dependent oxygenases. *Catalysts* 9, 20 (**2019**).

2 King-Smith, E., Zwick, C. R. & Renata, H. Applications of oxygenases in the chemoenzymatic total synthesis of complex natural products. *Biochemistry* 57, 403-412 (**2018**).

3 Roiban, G. D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* 51, 2208-2224 (**2015**).

4 Greule, A., Stok, J. E., De Voss, J. J. & Cryle, M. J. Unrivalled diversity: The many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat. Prod. Rep*. 35, 757-791 (**2018**).

5 Nakamura, H., Matsuda, Y. & Abe, I. Unique chemistry of non-heme iron enzymes in fungal biosynthetic pathways. *Nat. Prod. Rep*. 35, 633-645 (**2018**).

6 Perry, C., de los Santos, E. L. C., Alkhalaf, L. M. & Challis, G. L. Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat. Prod. Rep*. 35, 622-632 (**2018**).

7 Krebs, C., Galonic Fujimori, D., Walsh, C. T. & Bollinger, J. M., Jr. Non-heme Fe(IV)-oxo intermediates. *Acc. Chem. Res*. 40, 484-492 (**2007**).

8 Solomon, E. I. et al. Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev*. 100, 235-350 (**2000**).

9 Wallar, B. J. & Lipscomb, J. D. Dioxygen activation by enzymes containing binuclear non-heme iron clusters. *Chem. Rev*. 96, 2625-2658 (**1996**).

10 Barry, S. M. & Challis, G. L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal*. 3, 2362-2370 (**2013**).

11 Solomon, E. I., Goudarzi, S. & Sutherlin, K. D. O2 activation by non-heme iron enzymes. *Biochemistry* 55, 6363-6374 (**2016**).

12 Poulos, T. L. Heme enzyme structure and function. *Chem. Rev*. 114, 3919-3962 (**2014**).

13 Matsui, T., Unno, M. & Ikeda-Saito, M. Heme oxygenase reveals its strategy for catalyzing three successive oxygenation reactions. *Acc. Chem. Res*. 43, 240-247 (**2010**).

14 Rogers, M. S. & Lipscomb, J. D. Salicylate 5-hydroxylase: Intermediates in aromatic hydroxylation by a Rieske monooxygenase. *Biochemistry* 55, 6363-6374 (**2019**).

15 Capyk, J. K., D'Angelo, I., Strynadka, N. C. & Eltis, L. D. Characterization of 3-ketosteroid 9a-hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of Mycobacterium tuberculosis. *J. Biol. Chem*. 284, 9937-9946 (**2009**).

16 Lukowski, A. L. et al. C–H hydroxylation in paralytic shellfish toxin biosynthesis. *J. Am. Chem. Soc.* 140, 11863-11869 (**2018**).

17 Gilbson, D. T. Microbial degradation of aromatic compounds. *Science* 161, 1093-1097 (**1968**).

18 Higgins, T. P., Demarco, P. & Murrell, J. C. Purification and molecular characterization of the electron transfer protein of methanesulfonic acid monooxygenase. *J. Bacteriol*. 179, 1974-1979 (**1997**).

19 Schuster, J. et al. Bacterial degradation of tert-amyl alcohol proceeds via hemiterpene 2-methyl-3-buten-2-ol by employing the tertiary alcohol desaturase function of the Rieske nonheme mononuclear iron oxygenase MdpJ. *J. Bacteriol.* 194, 972-981 (**2012**).

20 Withall, D. M., Haynes, S. W. & Challis, G. L. Stereochemistry and mechanism of undecylprodigiosin oxidative carbocyclization to streptorubin B by the Rieske oxygenase RedG. *J. Am. Chem. Soc*. 137, 7889-7897 (**2015**).

21 Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res*. 28, 235-242 (**2000**).

22 Furusawa, Y. et al. Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from Rhodococcus sp. strain RHA1. *J. Mol. Biol*. 342, 1041-1052 (**2004**).

23 Kumar, P. et al. Structural insight into the expanded PCB-degrading abilities of a biphenyl dioxygenase obtained by directed evolution. *J. Mol. Biol*. 405, 531-547 (**2011**).

24 Mohammadi, M. et al. Retuning Rieske-type oxygenases to expand substrate range. *J. Biol. Chem.* 286, 27612-27621 (**2011**).

25 Kumar, P. et al. Structural basis of the enhanced pollutant-degrading capabilities of an engineered biphenyl dioxygenase. *J. Bacteriol.* 198 (**2016**).

26 Kumari, A., Singh, D., Ramaswamy, S. & Ramanathan, G. Structural and functional studies of ferredoxin and oxygenase components of 3-nitrotoluene dioxygenase from Diaphorobacter sp. strain DS2. *PloS One* 12, e0176398 (**2017**).

27 Kauppi, B. et al. Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Structure* 6, 571-586 (**1998**).

28 Karlsson, A. et al. Crystal structure of naphthalene dioxygenase: Side-on binding of dioxygen to iron. *Science* 299, 1039-1042 (**2003**).

29 Ferraro, D. J., Okerlund, A. L., Mowers, J. C. & Ramaswamy, S. Structural basis for regioselectivity and stereoselectivity of product formation by naphthalene 1,2-dioxygenase. *J. Bacteriol.* 188, 6986-6994 (**2006**).

30 Gakhar, L. et al. Structure and increased thermostability of Rhodococcus sp. naphthalene 1,2-dioxygenase. *J. Bacteriol.* 187, 7222-7231 (**2005**).

31 Dong, X. et al. Crystal structure of the terminal oxygenase component of cumene dioxygenase from Pseudomonas fluorescens IP01. *J. Bacteriol*. 187, 2483-2490 (**2005**).

32 Friemann, R. et al. Structural insight into the dioxygenation of nitroarene compounds: The crystal structure of nitrobenzene dioxygenase. *J. Mol. Biol.* 348, 1139-1151 (**2005**).

33 Nojiri, H. et al. Structure of the terminal oxygenase component of angular dioxygenase, carbazole 1,9a-dioxygenase. *J. Mol. Biol*. 351, 355-370 (**2005**).

34 Martins, B. M., Svetlitchnaia, T. & Dobbek, H. 2-Oxoquinoline 8-monooxygenase oxygenase component: Active site modulation by Rieske-[2Fe-2S] center oxidation/reduction. *Structure* 13, 817-824 (**2005**).

35 Dumitru, R., Jiang, W. Z., Weeks, D. P. & Wilson, M. A. Crystal structure of dicamba monooxygenase: A Rieske nonheme oxygenase that catalyzes oxidative demethylation. *J. Mol. Biol.* 392, 498-510 (**2009**).

36 Agarwal, R. et al. Tracking photoelectron induced in-crystallo enzyme catalysis.

37 Kim, J. H. et al. Structural and mechanistic insights into caffeine degradation by the bacterial N-demethylase complex. *J. Mol. Biol*. 431, 3647-3661 (**2019**).

38 Thottumkara, A. P., Parsons, W. H. & Du Bois, J. Saxitoxin. *Angew. Chem. Int. Ed*. 53, 5760-5784 (**2014**).

39 Llewellyn, L. E. Saxitoxin, a toxic marine natural product that targets a multitude of receptors. *Nat. Prod. Rep*. 23, 200-222 (**2006**).

40 Capyk, J. K. & Eltis, L. D. Phylogenetic analysis reveals the surprising diversity of an oxygenase class. *J. Biol. Inorg. Chem*. 17, 425-436 (**2012**).

41 Ferraro, D. J., Gakhar, L. & Ramaswamy, S. Rieske business: Structure-function of Rieske non-heme oxygenases. Biochem. Biophys. *Res. Commun*. 338, 175-190 (**2005**).

42 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38, W545-549 (**2010**).

43 Inoue, K. et al. Specific Interactions between the ferredoxin and terminal oxygenase components of a class IIB Rieske nonheme iron oxygenase, carbazole 1,9a-dioxygenase. *J. Mol. Biol.* 392, 436-451 (**2009**).

44 Tarasev, M., Rhames, F. & Ballou, D. P. Rates of the phthalate dioxygenase reaction with oxygen are dramatically increased by interactions with phthalate and phthalate oxygenase reductase. *Biochemistry* 43, 12799-12808 (**2004**).

45 Bordner, J., Thiessen, W. E., Bates, H. A. & Rapoport, H. Structure of a crystalline derivative of saxitoxin – Structure of saxitoxin. *J. Am. Chem. Soc*. 97, 6008-6012 (**1975**).

46 Pravda, L. et al. MOLEonline: A web-based tool for analyzing channels, tunnels and pores. *Nucleic Acids Res*. 46, W368-W373 (**2018**).

47 Berka, K. et al. MOLEonline 2.0: Interactive web-based analysis of biomacromolecular channels. *Nuclear Acids Res*. 40, W222-W227 (**2012**).

48 D'Ordine, R. L. et al. Dicamba monooxygenase: Structural insights into a dynamic Rieske oxygenase that catalyzes an exocyclic monooxygenation. *J. Mol. Biol.* 392, 481-497 (**2009**).

49 Escalante, D. E., Aukema, K. G., Wackett, L. P. & Aksan, A. Simulation of the bottleneck controlling access into a Rieske active site: Predicting substrates of naphthalene 1,2-dioxygenase. *J. Chem. Inf. Model* 57, 550-561 (**2017**).

50 Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*. 19, 4008-4008 (**1991**).

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

52 Bunkoczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D Biol. Crystallogr.* 67, 303-312 (**2011**).

53 Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D-Struct. Biol*. 66, 213-221 (**2010**).

54 Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D-Struct. Biol.* 60, 2126-2132 (**2004**).

55 Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr.* Sect. D-Biol. Crystallogr. 65, 1074-1080 (**2009**).

56 Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D-Struct. Biol.* 66, 12-21 (**2010**).

57 Morin, A. et al. Collaboration gets the most out of software. *eLife* 2, 6 (**2013**).

58 Mulcahy, J. V., Walker, J. R., Merit, J. E., Whitehead, A. & Du Bois, J. Synthesis of the paralytic shellfish poisons (+)-gonyautoxin 2, (+)-gonyautoxin 3, and (+)-11,11-dihydroxysaxitoxin. *J. Am. Chem. Soc.* 138, 5994-6001 (**2016**).

59 Gerlt, J. A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta.* 1854, 1019-1037 (**2015**).

60 Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504 (**2003**).

Chapter 3 Design Principles for Site-selective Hydroxylation by A Rieske Oxygenase

Reproduced with permission from Liu, J.; Tian, J.; Perry, C.; Lukowski, A. L.; Doukov, T. I.; Narayan, A. R. H.; Bridwell-Rabb, J. Design principles for site-selective hydroxylation by a Rieske oxygenase. *Nature communications*. **2022**, (13) 255. Copyright © 2022, Springer Nature

Author contributions: All authors (J.L., J.T., C.P., A.L.L., T.D., A.R.H.N. and J.B.R.) contributed to the design of the experiments and wrote the paper. J.L., J.T., C.P. and A.L.L. performed mutagenesis experiments. J.L. and J.T. expressed and purified all proteins and variants used in this work. J.T. performed the CD experiments. J.T. and C.P. conducted all biochemical assays and obtained LC–MS data. J.L., J.T. and C.P. analyzed LC–MS data. J.L. designed, conducted, and solved the four X-ray crystallographic structures determined in this work. J.L. and J.T. conducted all structural analyses. T.D. performed the Xenon-pressurization experiments.

3.1 Abstract

Rieske oxygenases exploit the reactivity of iron to perform chemically challenging C–H bond functionalization reactions. Thus far, only a handful of Rieske oxygenases have been structurally characterized and remarkably little information exists regarding how these enzymes use a common architecture and set of metallocenters to facilitate a diverse range of reactions. Herein, we detail how two Rieske oxygenases SxtT and GxtA use different protein regions to influence the site-selectivity of their catalyzed monohydroxylation reactions. We present high resolution crystal structures of SxtT and GxtA with the native β-saxitoxinol and saxitoxin substrates bound in addition to a Xenon-pressurized structure of GxtA that reveals the location of a substrate access tunnel to the active site. Ultimately, this structural information allowed for the identification of six residues distributed between three regions of SxtT that together control the selectivity of the C–H hydroxylation event. Substitution of these residues produces a SxtT variant that is fully adapted to exhibit the non-native site-selectivity and substrate scope of GxtA. Importantly, we also found that these selectivity regions are conserved in other structurally

characterized Rieske oxygenases, providing a framework for predictively repurposing and manipulating Rieske oxygenases as biocatalysts.

3.2 Introduction

Rieske non-heme iron oxygenases, or Rieske oxygenases, represent one of Nature's solutions for performing precise site-selective C–H bond functionalization reactions. This class of enzymes, which consists of more than 80,000 annotated sequences, perform powerful oxidative chemistry with catalyst-controlled selectivity using a combination of a Rieske-type [2Fe-2S] cluster and a mononuclear non-heme iron site[1-5]. The latter metallocenter binds and activates molecular oxygen (O_2) to facilitate a myriad of reactions, including the monooxygenation, dioxygenation, and oxidative dealkylation reactions for which the Rieske oxygenases are recognized[1-6]. Currently, two main architectures have been identified within this enzyme class, a hetero-hexamer that contains three β-subunits of unknown function (α3β3) and a trimer of the catalytic metallocenter containing α-subunits (α3, Figure 3.1 a-b) [7-23]. In both cases, the active sites are buried deep within the protein interior, a feature shared by many other types of metalloenzymes and more than 60-percent of proteins [24-27]. These deeply buried active sites serve to protect the reactive metal- and substrate-based intermediates and mitigate the risk of propagating radical chemistry outside of the active site[28,29]. A buried active site in a protein also necessitates a substrate entrance tunnel[25,26,30,31]. Such tunnels have been suggested to facilitate the correct order of multi-substrate reactions[24]. In Rieske oxygenases, this feature may also help ensure that O_2 binding is coupled to substrate binding and oxidation, which is mainly governed by the oxidation states of the metallocenters and changes to the mononuclear iron site upon substrate binding[3,4,13,24,32-35].

Additionally, in line with the keyhole-lock-key model of enzyme catalysis, the substrate entrance tunnel provides extra protein structure between bulk solvent and the active site and has been proposed to play a key role in determining the substrate scope of the Rieske oxygenase naphthalene dioxygenase (Figure. 3.1 a)[24,36]. The site-selectivity of a few Rieske oxygenase catalyzed reactions has also been shown to rely on a flexible loop region located in the C-terminal non-heme iron-binding domain of the catalytic Rieske oxygenase α-subunit[8,37-39]. Notably, unlike what is suggested by the traditional lock-key model of enzyme catalysis, which emphasizes only the interactions formed between the substrate and the active site, the so-called "keyhole", or auxiliary tunnel and loop regions that interact with a substrate, are emerging as enzyme engineering hotspots, mainly recognized for their potential influence on protein stability, substrate scope, and reaction selectivity (Figure. 3.1 c)[30,31,36,40,41]. Unfortunately, an overall lack of structure-function information coupled with low sequence identity between annotated Rieske oxygenases has left many open questions regarding how such auxiliary regions contribute to catalysis.

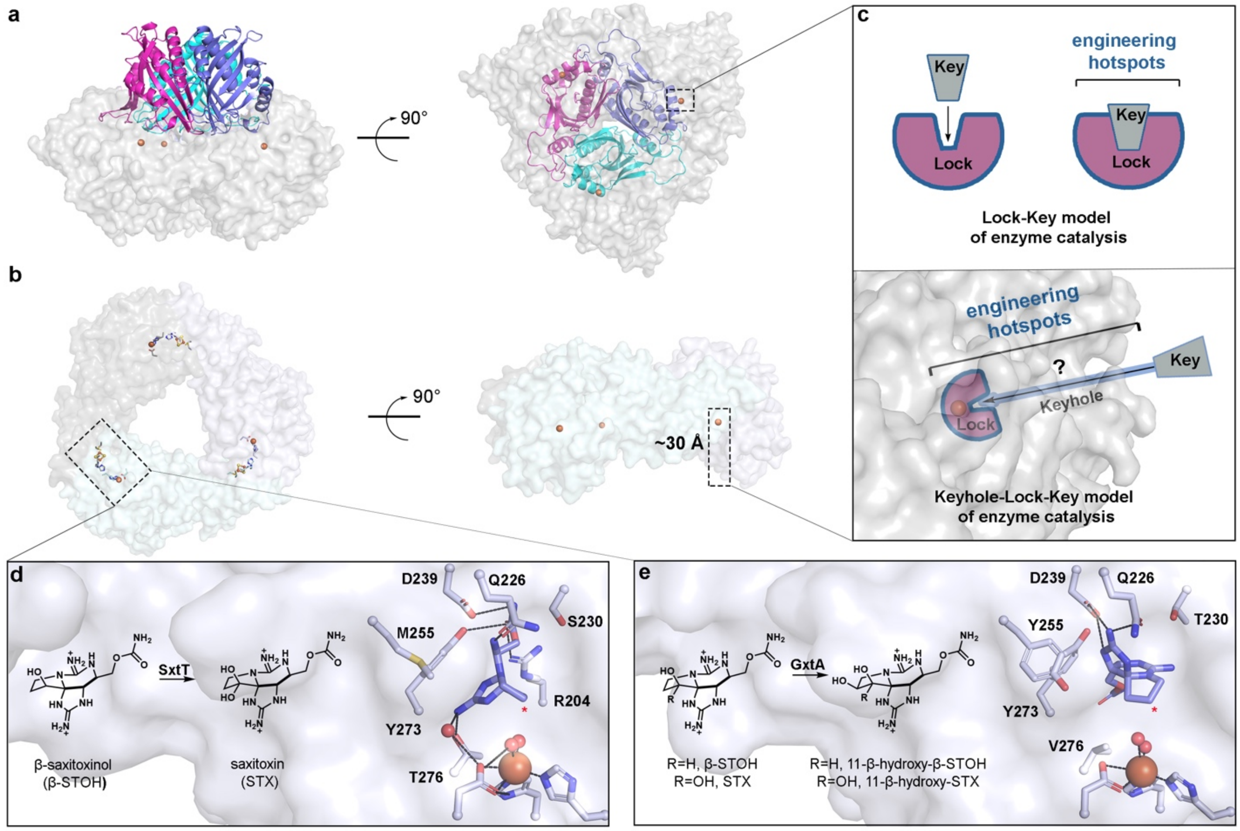
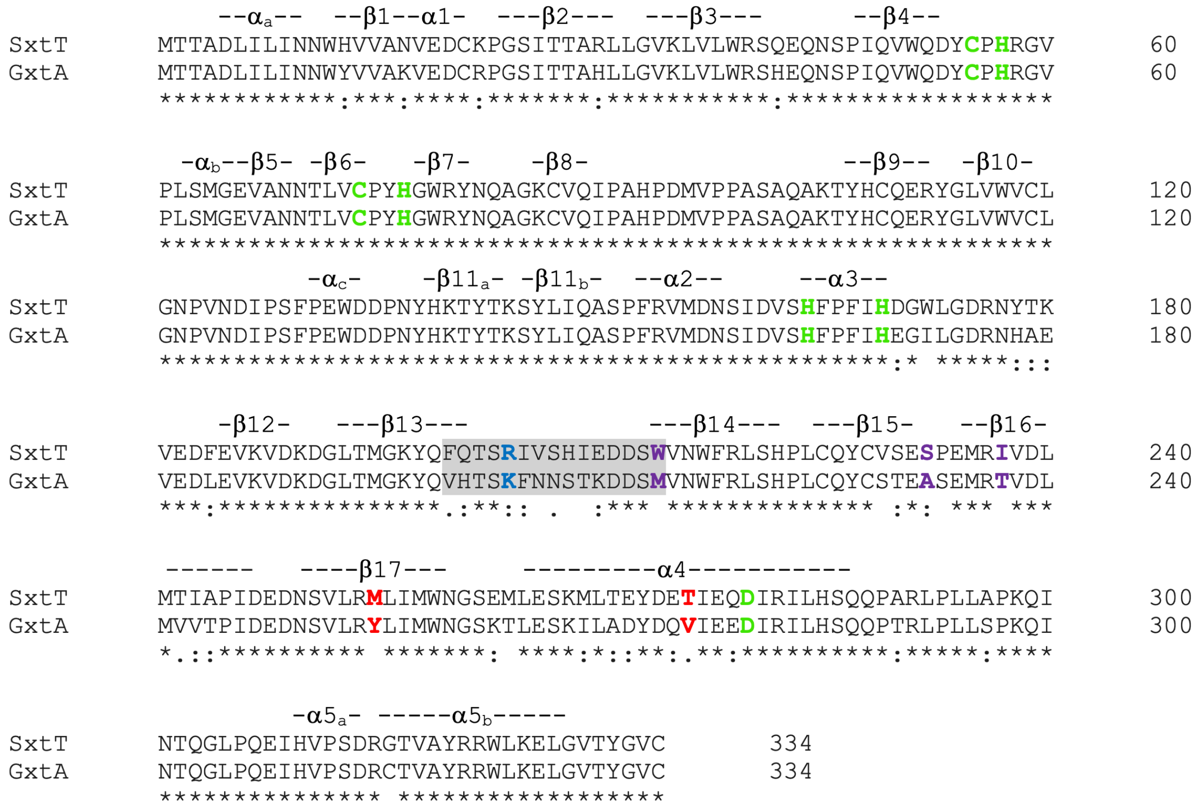


Figure 3.1. Rieske oxygenases adopt a trimeric architecture that sequesters the metal-based active site and reactive pathway intermediates in the interior of the protein.
(a) The quaternary architecture of an α3β3 Rieske oxygenase that is observed primarily in Rieske oxygenases that catalyze cis-dihydroxylation reactions[15-23]. This panel is shown in two orientations

that differ by a 90° rotation (PDB: 1NDO[22]). Here, the three catalytic α-subunits are shown in gray and the β-subunits are shown in pink, cyan, and purple. (b) The quaternary architecture of a homotrimeric α3 Rieske oxygenase is observed in 10 of the 18 available non-redundant Rieske oxygenase crystal structures [7-14] (PDB: 6WN3[7]). As shown in panel b, the active site in the α3 Rieske oxygenases is also buried deep within the protein core. (c) The architecture of α3β3 and α3 Rieske oxygenases require the substrate to travel to reach the active site before catalysis can occur. The extra architectural region that the substrate must traverse is known as the "keyhole" and unlike the traditional Lock-Key model of enzyme catalysis, requires the substrate to interact with protein regions outside of the active site[36]. The identity and importance of these auxiliary "keyhole" regions and their relationship to the selectivity of a Rieske oxygenase catalyzed reaction is yet to be determined. (d) SxtT catalyzes the conversion β-saxitoxinol (β-STOH) to saxitoxin (STX) and (e) GxtA catalyzes the conversion of β-STOH and STX to 11-β-hydroxy-β-saxitoxinol and 11-β-hydroxysaxitoxin, respectively[42]. The selectivity exhibited by SxtT and GxtA can be partially attributed to two active site residues that position the substrate correctly for hydroxylation at the respective positions (marked with a red asterisk)[7].

To investigate the architectural parameters that dictate the substrate specificity and site-selectivity of a Rieske oxygenase catalyzed reaction, we focused on two Rieske oxygenases, SxtT and GxtA, which are involved in the biosynthesis of paralytic shellfish toxins. These enzymes share 88-percent sequence identity with one another and selectively install a hydroxyl group on adjacent C12 and C11 positions of a tricyclic saxitoxin scaffold, respectively (Figure. 3.1 d-e and Figure. 3.2)[42]. We previously found that the differences between the SxtT and GxtA mononuclear iron-containing active sites are limited to just two residues that play a minor role in controlling C11 versus C12 hydroxylation[7]. In particular, in SxtT, these residues correspond to Met255 and Thr276, which sterically orient and hydrogen bond with the substrate analog dideoxysaxitoxin (ddSTX, Figure. 3.1 d)[7]. In GxtA, these positions instead correspond to Tyr255 and Val276, neither of which interact with ddSTX (Figure. 3.1 e)[7]. Nevertheless, in GxtA, ddSTX undergoes a dramatic rotation in the active site that places C11 as the closest atom to the non-heme iron center[7] (Figure. 3.1 e). Despite the noted importance of these residues for orienting ddSTX in the active sites of SxtT and GxtA, the activity of these enzymes is not inverted by creating double active site variants. Instead, each double variant remains biased towards its native site-selectivity and substrate[7]. Thus, in this work, we undertook studies to determine the design principles outside of the active site that control site-selectivity of the SxtT and GxtA catalyzed reactions.



```
              --αa--  --β1--α1-  ---β2---  ---β3---     --β4--
SxtT      MTTADLILINNWHVVANVEDCKPGSITTARLLGVKLVLWRSQEQNSPIQVWQDYCPHRGV      60
GxtA      MTTADLILINNWYVVAKVEDCRPGSITTAHLLGVKLVLWRSHEQNSPIQVWQDYCPHRGV      60
          ************:***:****:*******:***********:******************

            -αb--β5- -β6-  -β7-    -β8-                 --β9--  -β10-
SxtT      PLSMGEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCL      120
GxtA      PLSMGEVANNTLVCPYHGWRYNQAGKCVQIPAHPDMVPPASAQAKTYHCQERYGLVWVCL      120
          ************************************************************
                    -αc-    -β11a- -β11b- --α2--     --α3--
SxtT      GNPVNDIPSFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHDGWLGDRNYTK      180
GxtA      GNPVNDIPSFPEWDDPNYHKTYTKSYLIQASPFRVMDNSIDVSHFPFIHEGILGDRNHAE      180
          *************************************************:* *****:::

              -β12-   ---β13---            ---β14---  ---β15-   --β16-
SxtT      VEDFEVKVDKDGLTMGKYQFQTSRIVSHIEDDSWVNWFRLSHPLCQYCVSESPEMRIVDL      240
GxtA      VEDLEVKVDKDGLTMGKYQVHTSKFNNSTKDDSMVNWFRLSHPLCQYCSTEASEMRTVDL      240
          ***:***************.:**:: .  :*** ************** :*: *** ***

           ------   ----β17---    ---------α4-----------
SxtT      MTIAPIDEDNSVLRMLIMWNGSEMLESKMLTEYDETIEQDIRILHSQQPARLPLLAPKQI      300
GxtA      MVVTPIDEDNSVLRYLIMWNGSKTLESKILADYDQVIEEDIRILHSQQPTRLPLLSPKQI      300
          *.::********** *******: ****:*::**:.**:**********:*****:****

                   -α5a- -----α5b-----
SxtT      NTQGLPQEIHVPSDRGTVAYRRWLKELGVTYGVC       334
GxtA      NTQGLPQEIHVPSDRCTVAYRRWLKELGVTYGVC       334
          *************** ******************
```

Figure 3.2. Sequence similarity between SxtT and GxtA.

SxtT and GxtA share high levels of sequence identity with one another. The secondary structure observed in the crystal structures of SxtT and GxtA are marked above the alignment. The α-helices marked a, b, and c are not universally conserved among structurally characterized Rieske oxygenases[1]. The metal-binding residues are marked in green. The flexible loop region is highlighted in grey. Arg204 from the loop region is shown in blue, whereas tunnel lining residues identified in this work as being important for selectivity are shown in purple. Active site residues that are important for selectivity are shown in red.

Towards this end, we used X-ray crystallography and Xenon (Xe)-pressurization experiments to identify global differences in the SxtT and GxtA architectures and establish the route of substrate access to the active site. Coupled with site-directed mutagenesis and activity assays, these experiments demonstrated that SxtT and GxtA dictate the site-selectivity of their reactions through a coordinated interplay of three protein regions located inside and outside of the active site. Through the change of only six residues, SxtT, like GxtA, can be made to primarily catalyze hydroxylation at the C11, rather than the C12 position on the tricyclic scaffold and demonstrate a preference for a saxitoxin (STX) substrate rather than a native β-saxitoxinol (β-STOH) substrate. Importantly, we also determined that the implicated protein regions involved in dictating selectivity are conserved in other structurally characterized Rieske oxygenases, suggesting their control over selectivity may be widespread in this enzyme class. Thus, this work outlines the design principles for engineering Rieske oxygenases to have improved activity, broader substrate scopes, or altered reaction specificity.

3.3 Results

3.3.1 Native substrate-bound structures reveal the role of active site residue Tyr255 in selectivity

To investigate how the native hydroxylated substrates bind to SxtT and GxtA and to identify global architectural differences between these proteins, we solved the structures of SxtT and GxtA with β-STOH bound, as well as that of GxtA with STX bound to 1.79-, 1.79-, and 1.74-Å resolution, respectively (Figure. 3.3, Figure. 3.4, Figure. 3.5 and Table 3.1, Table 3.2). Each structure showcases an α3 trimeric architecture, and each protomer of the trimer contains a Rieske cluster in the N-terminal domain and a non-heme iron center in the C-terminal domain (Figure. 3.3.a, Figure. 3.6). The trimeric quaternary architecture places these metallocenters within electron transfer distance of one another across the subunit-subunit interfaces (Figure. 3.3.a, Figure. 3.6). In SxtT, the presumed native substrate β-STOH, shows slight deviations in its position in the active site relative to that observed with ddSTX, but it maintains interactions with Arg204, Cys228, Gln226, Ser230, Asp239, Tyr273, and Thr276 (Figure. 3.1.d and Figure. 3.3.b). The differentiating feature between β-STOH and ddSTX, the C12 β-hydroxyl group, forms only one additional interaction in the SxtT active site with Tyr198 (Figure. 3.3.b). This interaction is mediated by a water molecule and Tyr198, like Cys228, Asp239, and Tyr273, is conserved in GxtA. In GxtA, STX also forms extra active-site interactions (Figure. 3.3.c). Here, the addition of a hydroxyl group at the C12 α-position allows for formation of hydrogen bonds with Ser159 and Tyr255, explaining the previously noted structural importance of Tyr255 for selectivity[7]. The remaining GxtA interactions are formed between β-STOH and STX with Asp239 and Gln226 (Figure. 3.3.c-d).

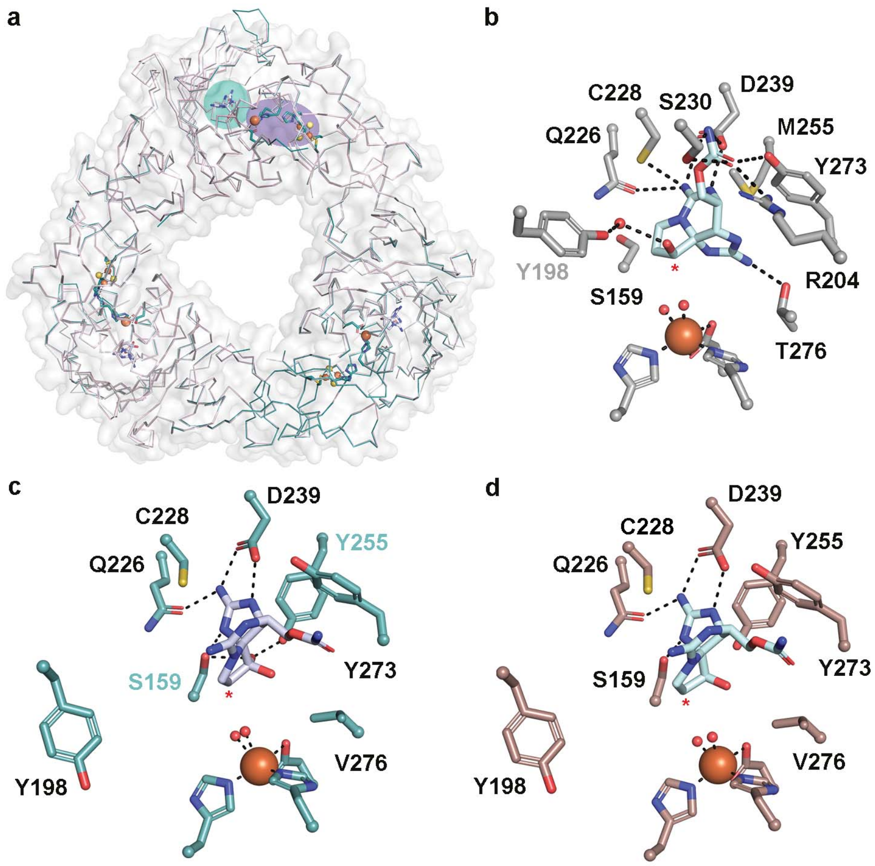
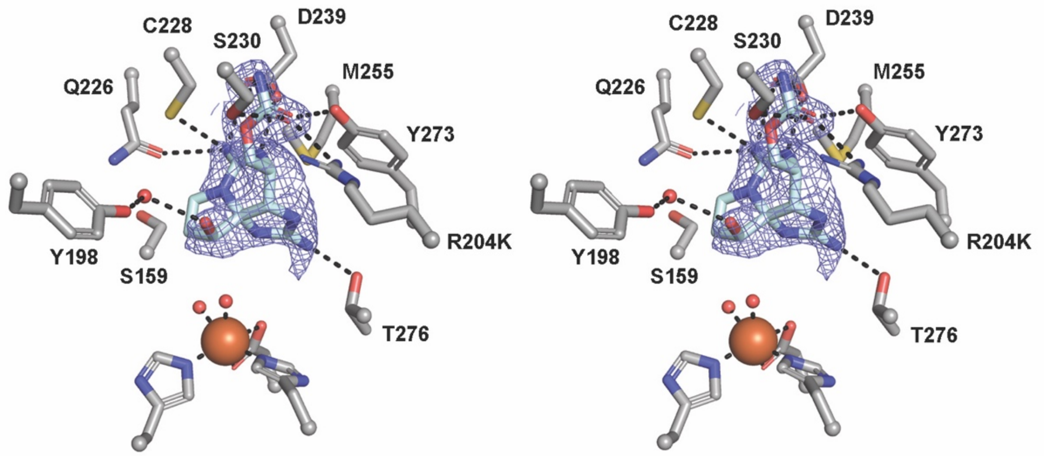
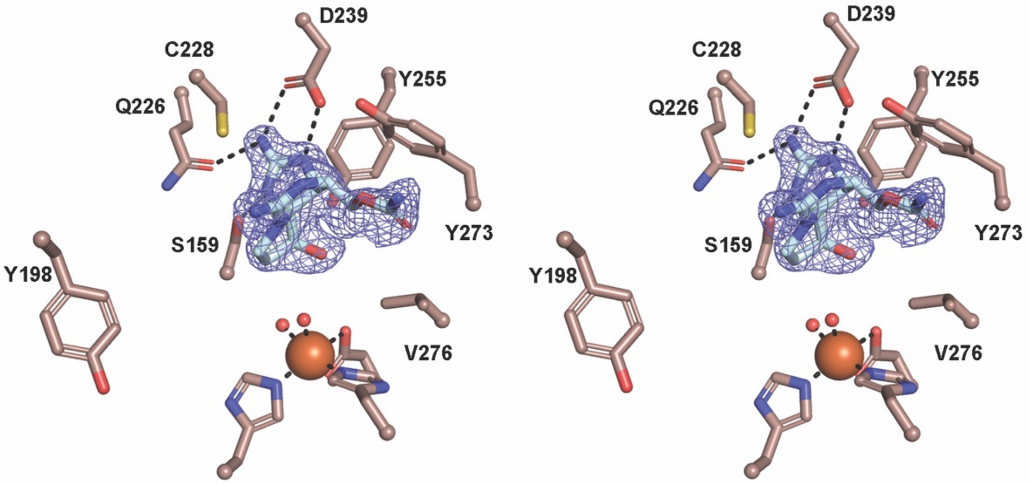


Figure 3.3. The structures of SxtT and GxtA with their native hydroxylated substrates bound reveal the essential active site interactions.

(a) The quaternary trimeric architectures of each determined structure are shown as ribbon diagrams and overlaid with each other (SxtT with β-STOH bound is shown in gray, GxtA with β-STOH bound is shown in pink, and GxtA with STX bound is shown in teal). The metallocenters that bridge two subunits are highlighted in purple and the substrate binding site of one subunit is highlighted in teal. (b) The structure of SxtT with β-STOH bound reveals the myriad of interactions that this molecule forms in the active site. Tyr198 (gray font) interacts with the β-hydroxyl group of β-STOH via a water molecule. (c) STX binds in the active site of GxtA. The α-hydroxyl group interacts with active site residues Ser159 and Tyr255 (cyan font). (d) Despite the similar makeup of the GxtA active site relative to SxtT, β-STOH binds in a remarkably different orientation and interacts with conserved residues Gln226 and Asp239. In all panels the position of hydroxylation is marked with a red asterisk.

The most significant structural difference between these substrate-bound structures of SxtT and GxtA is the orientation of a loop that connects the β13 and β14 strands of the C-terminal iron binding domain. This loop, which is composed of residues 200-214 shows low sequence identity between the proteins (4 out of 15 residues are conserved) and is flexible, as evidenced by the higher-than-average B-factors in this region of the protein (Figure. 3.2 and Figure. 3.7). The flexibility of this loop region means that for several subunits of both SxtT and GxtA, this loop region was not able to be built. However, in one subunit of SxtT and in two subunits of GxtA, this loop is ordered. In SxtT, this loop extends towards the active site and interacts with β-STOH as well as with non-conserved residues Ser230 and Glu272 (Figure. 3.8.a). In contrast, the two loop orientations in GxtA extend away from the active site. In one of these orientations, the loop is stabilized by crystal packing (Figure. 3.8.b, Figure. 3.9), whereas in the second orientation, loop residues Asn206, Ser208, and Thr209 interact with the backbone and sidechain of Ser233 and the backbone of Glu234 on the surface of the protein (Figure. 3.8.c). Two of the corresponding loop residues in SxtT (Val206 and Ile209) are hydrophobic and unlikely to reside on the protein surface or form similar interactions with the corresponding surface residues Pro233 and Glu234. Accordingly, whereas both orientations of the GxtA loops reveal an open path from the surface of the protein to the substrate binding site, the SxtT loop, which is tucked into the protein interior, blocks the analogous path in the crystal structure (Figure. 3.8.d-e).


C228
S230
D239
M255
Q226
Y273
Y198
S159
R204K
T276
C228
S230
D239
M255
Q226
Y273
Y198
S159
R204K
T276


D239
C228
Y255
Q226
S159
Y273
Y198
V276
D239
C228
Y255
Q226
S159
Y273
Y198
V276

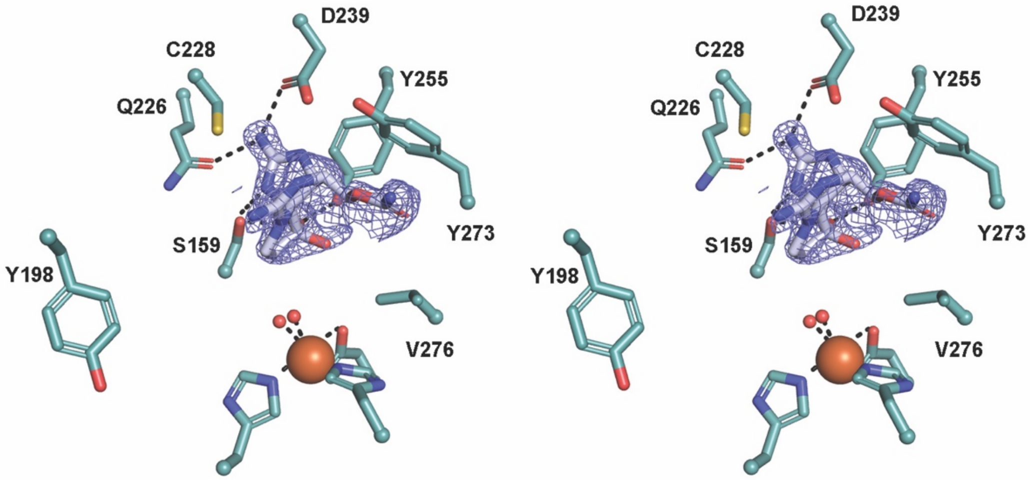

D239
C228
Y255
Q226
S159
Y273
Y198
V276
D239
C228
Y255
Q226
S159
Y273
Y198
V276

Figure 3.4. Stereo figures of β-STOH bound to SxtT (top), GxtA (middle), and STX bound to GxtA (bottom).
All panels are shown with 2Fo-Fc simulated annealing composite omit electron density maps. These maps are contoured at 0.9σ for SxtT and 1.0σ for GxtA.

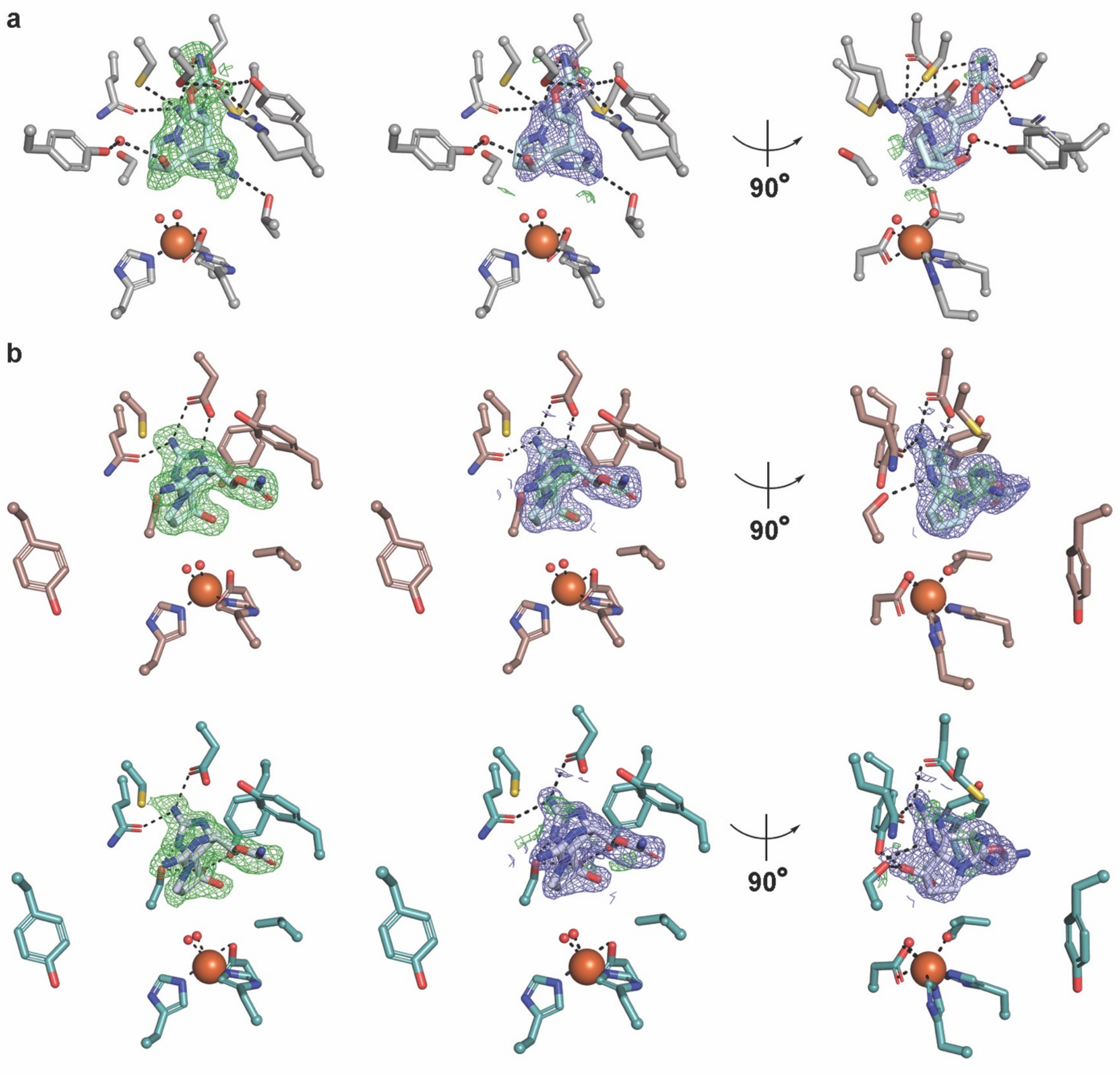


Figure 3.5. β-STOH and STX bind in remarkably different orientations in SxtT and GxtA.
(a) In the left panel, a Fo-Fc omit electron density map contoured at ±3.0σ was calculated after β-STOH was omitted from the refined structure of β-STOH-bound SxtT. In the right two panels, two orientations of β-STOH bound to SxtT are shown with 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively. (b) Similar to panel a, a Fo-Fc omit electron density map contoured at ±3.0σ around β-STOH bound to GxtA as well as refined 2Fo-Fc and Fo-Fc

electron density maps contoured at 1.0σ and ±3.0σ, respectively, are shown. (c) STX makes extra interactions in the GxtA active site. STX is shown in the left panel with an omit electron density map contoured at 3.0σ after STX was omitted from the STX-bound structure. The right panels show 2Fo-Fc and Fo-Fc electron density maps contoured at 1.0σ and ±3.0σ, respectively.

Table 3.1. Data collection and refinement statistics.

	β-STOH-SxtT 7SZH	β-STOH-GxtA 7SZF	STX-GxtA 7SZE	Xe-GxtA§ 7SZG
Data collection				
Space group	*C*222	$P12_11$	$P12_11$	$P12_11$
Cell dimensions				
a, *b*, *c* (Å)	153.3, 158.5, 115.8	74.73, 96.93, 80.81	74.98, 96.93, 80.67	75.20, 96.64, 80.72
α,β,γ (°)	90, 90, 90	90, 107.0, 90	90, 106.91, 90	90, 107.1, 90
Resolution (Å)	50-1.79	50-1.79	50-1.74	50-2.69
$R_{\mathrm{meas}}^{\ddagger}$	0.068 (1.702)	0.062 (0.723)	0.076 (1.035)	0.141 (1.393)
$I/\sigma I^{\ddagger}$	22.05 (1.86)	14.77 (1.93)	16.75 (2.05)	19.63 (2.42)
Completeness (%)‡	98.5 (97.8)	98.2 (96.8)	96.9 (96.1)	99.7 (100)
Redundancy‡	13.7 (13.4)	3.5 (3.3)	7.0 (6.8)	18.5 (17.5)
CC1/2‡	1.000 (0.815)	0.999 (0.737)	0.999 (0.756)	0.999 (0.793)
Refinement				
Resolution (Å)	1.79	1.79	1.74	2.69
No. reflections	129304	102236	109917	30537
R_{work} / R_{free}	0.170, 0.198	0.158, 0.197	0.161, 0.197	0.199, 0.235
No. atoms				
Protein	8073	7900	8165	7721
[2Fe-2S]	12	12	12	12
Non-heme Fe	3	3	3	3
Glycerol	108	84	72	36
Sulfate	30	-	-	-
Water	834	926	939	254
STX molecule	60	60	63	-
Cl	2	7	8	3
Xe	-	-	-	3
DTT	-	-	-	32
B-factors				
Overall	44.57	31.34	31.40	50.62
Protein	43.27	30.22	30.12	50.53
STX molecule	51.40	30.59	38.16	-
[2Fe-2S]	30.34	22.14	20.86	40.48

Non-heme Fe	29.75	20.87	20.51	46.13
Glycerol	65.40	49.63	53.31	66.05
Sulfate	90.54	-	-	-
Water	51.74	39.33	40.45	61.77
Cl	76.98	40.26	47.40	61.47
Xe	-	-	-	82.60
DTT	-	-	-	64.47
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.008	0.011
Bond angles (°)	0.918	0.897	0.950	1.586

‡Values in parentheses are for highest-resolution shell.

§Bijvoet pairs were not merged during data processing

Table 3.2. Refined structure geometry statistics.

Structure	Missing residues	Ramachandran favored	Ramachandran allowed	Ramachandran disallowed	Rotamer outliers
β-STOH-SxtT	A: 1-2, 207-208, 297-303; B:297-303; C:203-211, 298-303	96.96%	3.04%	0%	0.99%
β-STOH-GxtA	A: 1,204-211, 297-303; B: 1, 203-210, 298-304; C: 1, 297-303	97.14%	2.86%	0%	0.33%
STX-GxtA	A: 201-211, 297-303; B: 300-304; C: 297-304	96.98%	3.02%	0	0.33%
Xe-GxtA	A: 1-2, 203-212, 297-303; B: 1, 203-211, 297-304; C: 298-304	96.71%	3.29%	0	0.92%

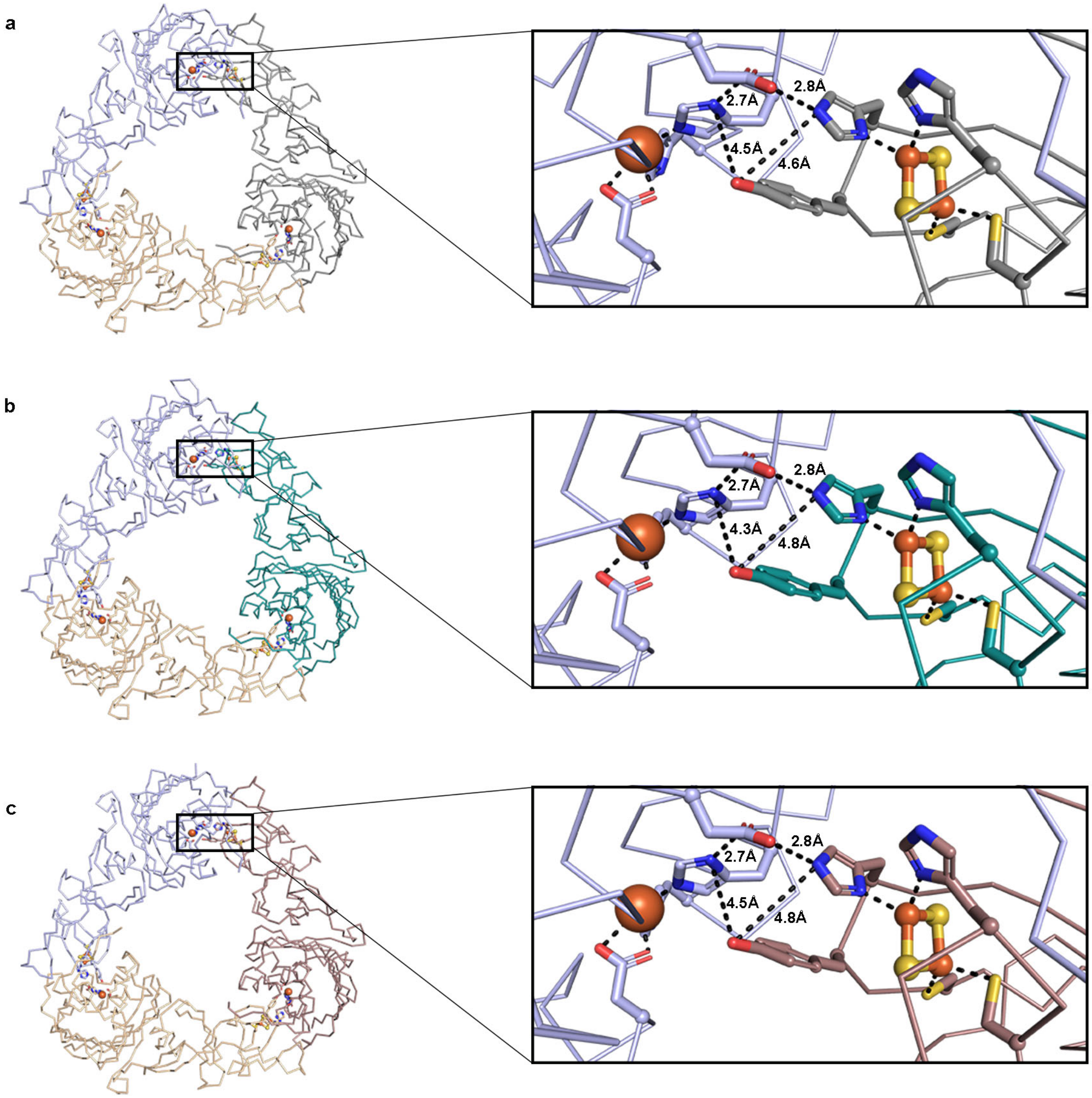


Figure 3.6. The structures of SxtT and GxtA determined in this work each has a trimeric architecture.

Each monomeric unit of the contains a [2Fe-2S] Rieske cluster in the N-terminal domain and a mononuclear iron center in the C-terminal domain. As observed in other Rieske oxygenase structures, in all of the substrate bound structures determined here, the metallocenters of two different subunits are placed within electron transfer distance of one another across the subunit-subunit interfaces. The path between the Rieske cluster from one subunit and the mononuclear iron from a second subunit are shown in enlarged panels for the three substrate bound structures determined here. (a) The structure of SxtT with β-STOH bound. (b) The structure of GxtA with STX bound. (c) The structure of GxtA with β-STOH bound.

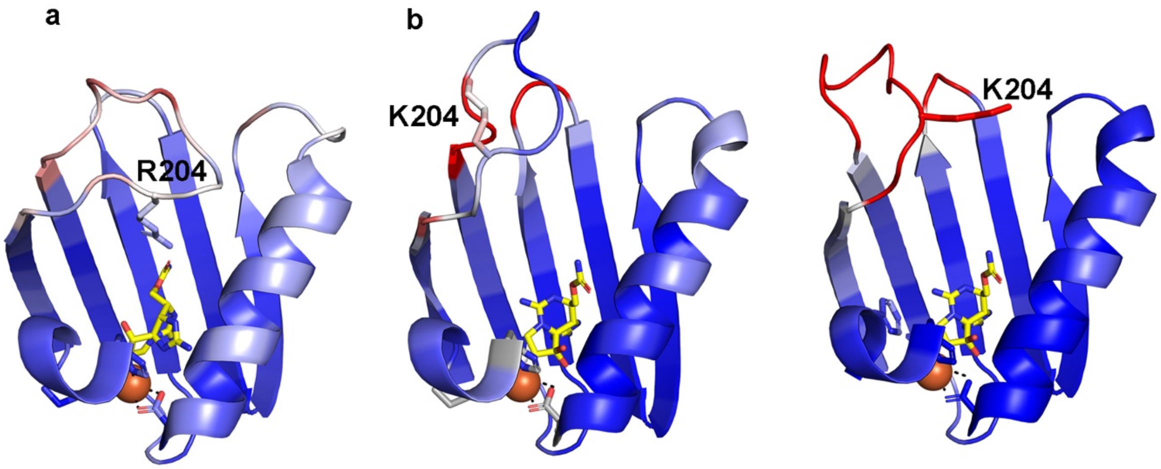


Figure 3.7. Identification of a flexible loop that is important for selectivity.
The loop that connects the β13 and β14 strands of the C-terminal iron binding domain in SxtT and GxtA is highly flexible as evidenced by the B-factors. (a) The flexible loop in SxtT reaches toward the active site and Arg204 interacts with β-STOH. (b) The two orientations of the loop in STX-bound GxtA extend away from the active site in both the β-STOH and STX-bound structures. In all panels, the B-factors are plotted on a scale of 20–60 Å2 with increasing values colored in blue, white, and red, respectively.

3.3.2 Identification of a flexible loop residue that is involved in selectivity

To determine whether the noted differences in sequence and orientation of the flexible loop contribute to the selectivity of the catalyzed reactions, site-directed mutagenesis was used to create a full SxtT chimeric loop variant (F200V/Q201H/R204K/I205F/V206N/S207N/H208S/I209T/E210K/W214M), a series of shorter chimeric loop variants (F200V/Q201H/R204K, F200V/Q201H/R204K/I205F/V206N/S207N, and F200V/Q201H/R204K/I205F/V206N/S207N/H208S/I209T/E210K), and a single SxtT loop variant at the 204 position (R204K) (Figure. 3.10 and Table 3.3). This single variant was chosen based on the observed interaction between this residue

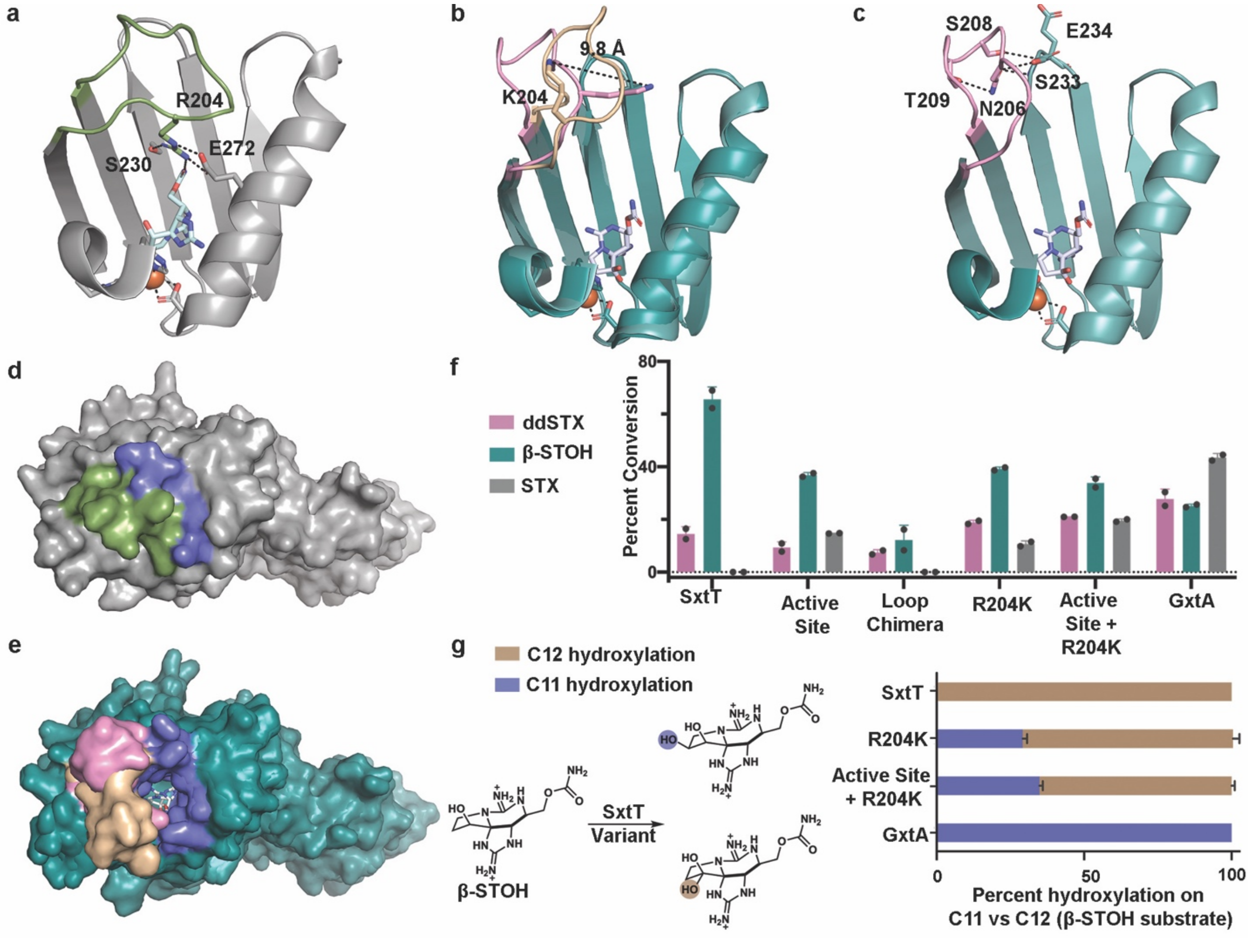


Figure 3.8. Identification of a second structural region in SxtT and GxtA that is involved in selectivity.

Different conformations of a flexible loop that connects the β13 and β14 strands of (a) SxtT and (b) GxtA are observed in each substrate bound structure. In SxtT this loop (green) interacts with the substrate as well as Ser230 and Glu272. In GxtA, this flexible loop is found in two different orientations (pink and wheat). The difference between the position of Lys204 on these loops is 9.8 Å. (c) The pink loop orientation shown in panel b interacts with residues Ser233 and Glu234 on the surface of the protein. (d) A surface rendering of SxtT with β-STOH bound shows that the loop (green) closes the entry (blue) into the active site. (e) A similar surface rendering of GxtA reveals that the different observed loop orientations (pink and wheat) extend away from the active site and leave the entrance (blue) wide open. (f) A chimeric variant in the loop region of SxtT (F200V, Q201H, R204K, I205F, V206N, S207N, H208S, I209T, E210K, W214M) shows low levels of activity on ddSTX and β-STOH, whereas the single R204K loop variant is able to hydroxylate ddSTX, β-STOH, and STX. Similarly, creation of a triple variant that combines residue changes at the R204 position with changes in the active site region (M255Y, T276V) amplify the ability to hydroxylate STX. (g) As evidenced by incorporation of ethanol into the reaction product, a mixture of hydroxylation at the C11 (teal) and C12 (purple) positions is observed with the R204K and M255Y/T276V/R204K variants when β-STOH is used as a substrate. In panels f and g, data was measured in duplicate, and the data is presented as mean values ± SD. In all panels, ddSTX corresponds to dideoxysaxitoxin, β-STOH corresponds to β-saxitoxinol, and STX corresponds to saxitoxin.

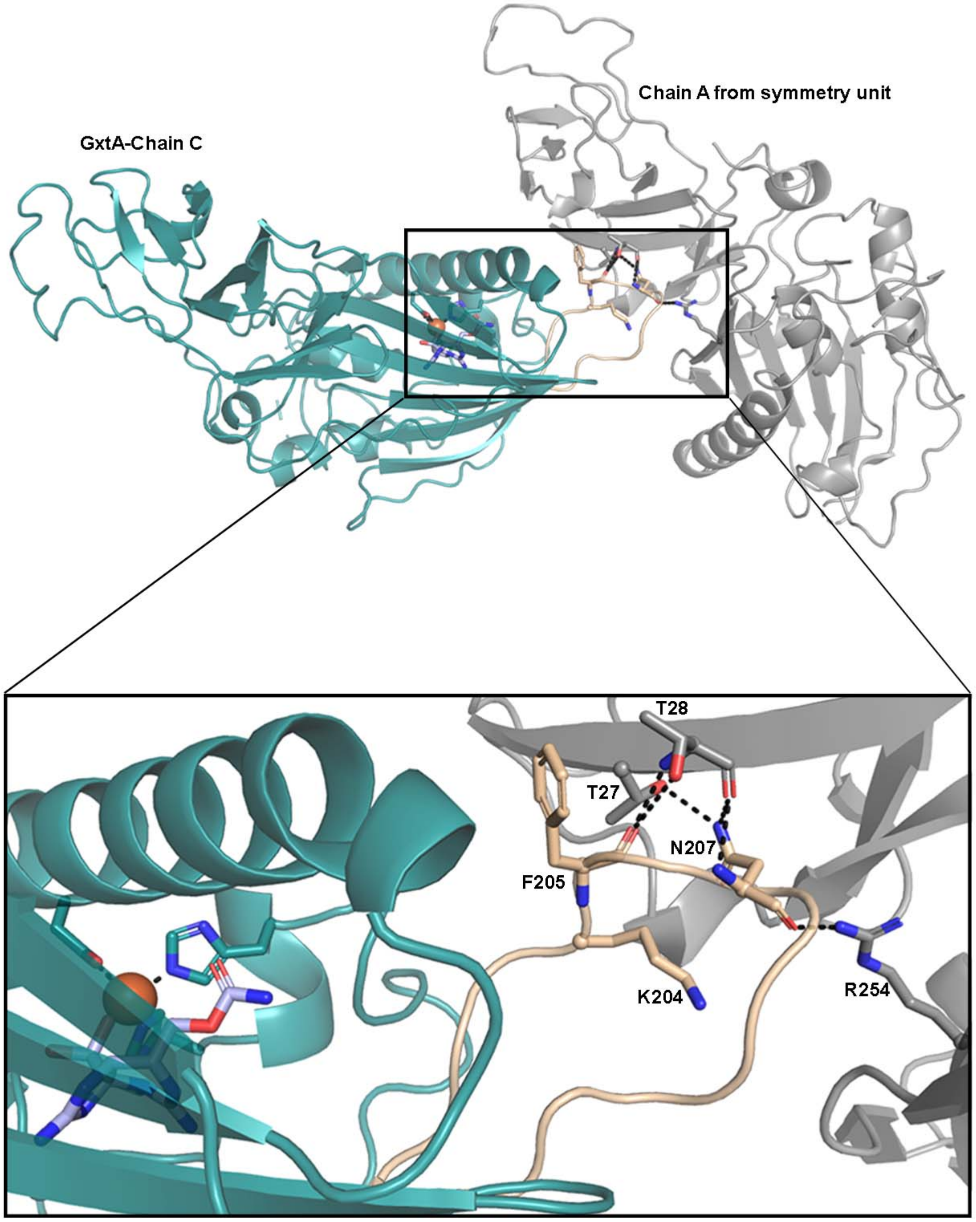
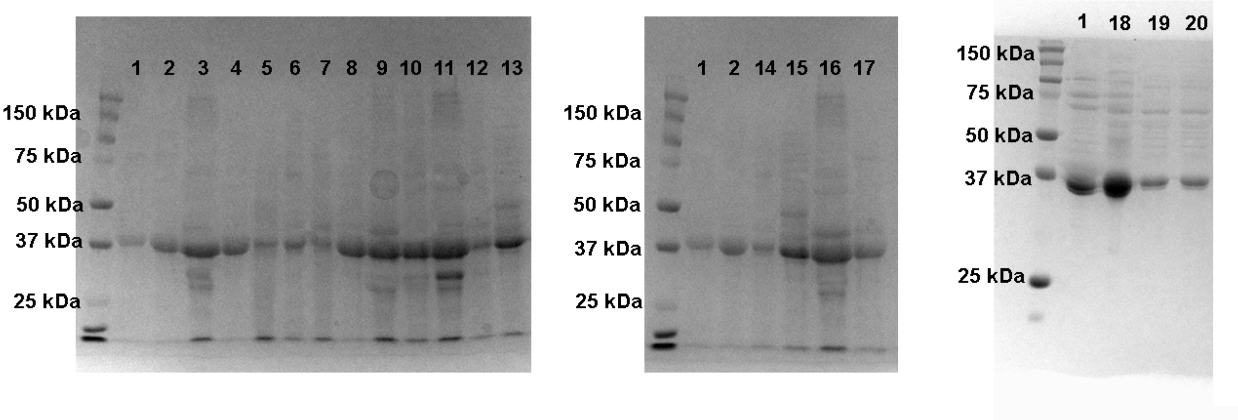


Figure 3.9. The structure of GxtA shows two orientations of the flexible loop.
These orientations can be seen in Figure 3b. One of the loop orientations is extended (wheat) and stabilized by interactions with a symmetry related molecule in the crystal packing. This extended loop is shown here in wheat as part of the dark teal monomer. The symmetry related molecule that is interacting with this loop is shown in gray. A zoomed in view of the interactions made between the loop and the symmetry molecule are shown in the bottom panel.

and the carbamate sidechain in the structures of ddSTX and β-STOH bound-SxtT (Figure. 3.1.d and 3.3.b). The activity of these variants was tested using VanB as the reductase and ddSTX, β-STOH, and STX as substrates. To our surprise, LC-MS experiments showed that the full chimeric loop variant of SxtT has only a low level of activity on ddSTX and β-STOH and no activity on STX (Figure. 3.8.f and Figure. 3.11). The observed lack of hydroxylation on STX suggests the selectivity is unchanged in the chimeric loop SxtT variant. However, to ensure that this lack of activity is due to changes in the primary structure of the loop region, rather than changes to the folding of the variant relative to the wild-type SxtT, a circular dichroism (CD) experiment was used to demonstrate that the full chimeric loop variant is similar in secondary structure to wild-type SxtT (Figure. 3.12.a).



1. SxtT
2. GxtA
3. M255Y/T276V/R204K
4. W214M/S232A/I237T (tunnel chimera)
5. M255Y/T276V/R204K/S232A
6. M255Y/T276V/R204K/I237T
7. M255Y/T276V/R204K/W214M
8. M255Y/T276V/R204K/W214M/S232A/I237T
9. M255Y/T276V/W214M/S232A/I237T
10. S232A
11. I237T
12. S232A/I237T
13.W214M/I237T
14. F200V/Q201H/R204K/I205F/V206N/S207N/ H208S/I209T/E210K/W214M (loop chimera)
15. M255Y/T276V/R204K
16. M255Y/T276V
17.R204K
18.F200V/Q201H/R204K
19.F200V/Q201H/R204K/I205F/V206N/S207N
20.F200V/Q201H/R204K/I205F/V206N/S207N/ H208S/I209T/E210K

Figure 3.10. Gel of SxtT variants
The wild type SxtT and GxtA proteins as well as the SxtT variants used in this work were expressed and purified using affinity and gel filtration chromatography. The purity of each protein was

assessed using SDS-PAGE gels. As predicted, each wild-type and variant protein has a molecular weight of approximately 38 kDa (judged based on comparison to a protein standard). Aliquots of these purified proteins were used once and discarded to ensure that consistent activity was observed across the batch.

Table 3.3. Sequences of primers used in this study.

Primer name	Primer DNA sequence
SxtT_M255Y	5' – gatgaagataatagcgttctgcgttatctgattatgtggaatggtagcga – 3'
SxtT_T276V	5' – aaatgctgaccgaatatgatgaagtcatcgaacaggatattcgcattc – 3'
SxtT_W214M	5' – agccatattgaagatgatagcatggtgaattggtttcgtctgag – 3'
SxtT_S230T	5' – cgctgtgtcagtattgtgttaccgaaagtccgg – 3'
SxtT_S232A	5' – gctgtgtcagtattgtgttagcgaagctccggaaatgcgta – 3'
SxtT_I237T	5' – gaaagtccggaaatgcgtactgttgatctgatgacc – 3'
SxtT_R204K	5' – tgggcaaatatcagtttcagaccagcaagattgttagccatattgaagatgatag – 3'
SxtT_F200V/Q201H	5' – gaccatgggcaaatatcaggttcataccagcaagattgttagc – 3'
SxtT_I205F/V206N/S207N	5' – gcaaatatcaggttcataccagcaagtttaataaccatattgaagatgatagctgggtga – 3'
SxtT_H208S/I209T/E210K	5' – aggttcataccagcaagtttaataacagtactaaagatgatagctgggtgaattggtttc – 3'

In contrast to the chimeric loop variant, the R204K-SxtT variant shows activity on each of the three provided substrates (Figure. 3.8.f and Figure. 3.11). To identify the specific residue that leads to the drastic difference in activity between the single and full loop variants, the ability of the shorter chimeric loop variants to hydroxylate ddSTX, β-STOH, and STX, was probed. Here, it was determined that each tested variant showed decreased activity on ddSTX and β-STOH relative to wild-type SxtT (Figure. 3.13). Only the R204K- and F200V/Q201H/R204K-SxtT loop variants show any activity on STX. As the R204K-SxtT variants shows the highest level of activity on STX, a study to identify the position of hydroxylation when provided a β-STOH substrate was undertaken. Here, we capitalized on a previously established method to distinguish between hydroxylation at the C11 and C12 positions[7]. This method, which relies on the reaction of one molecule of ethanol with STX to form the hemiketal, reveals when hydroxylation has occurred at the C12 position (Figure. 3.14.a). This experiment demonstrated that the R204K-SxtT variant, like the M255Y/T276V double active site variant of SxtT, produces both an ethanol-incorporated product and a non-ethanol incorporated product, suggesting that this variant displays some degree of altered site-selectivity relative to the wild-type enzyme (Figure. 3.8.g and Figure. 3.14). This R204K SxtT variant shows less C12 hydroxylation than the double

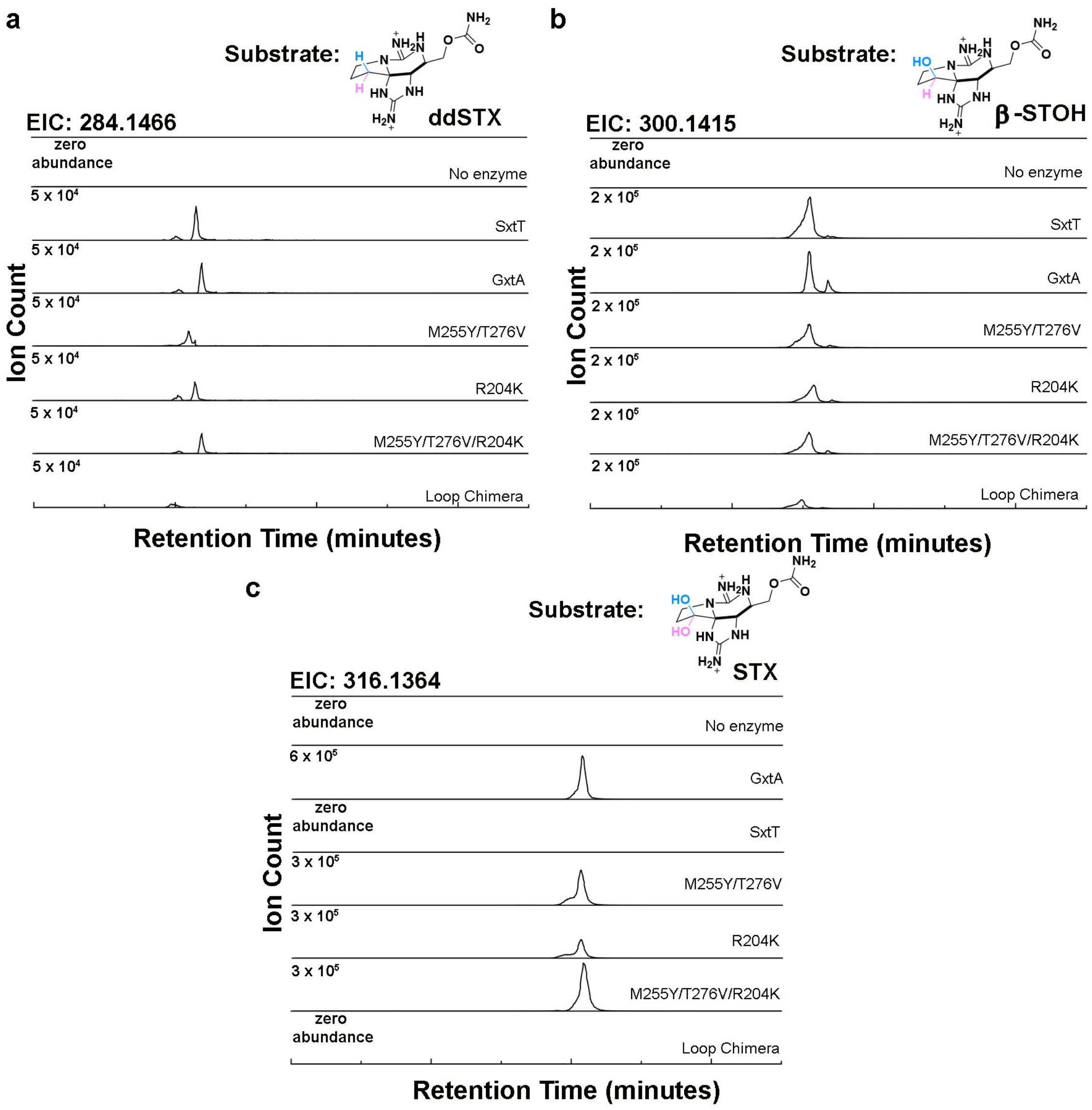


Figure 3.11. A combination of changes to the active site and flexible loop of SxtT leads to an increased ability to hydroxylate STX.

(a) The extracted ion chromatograms for SxtT variant reaction products with a ddSTX substrate. The *m/z* = 284.1466 is the mass of a hydroxylated ddSTX product. (b) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a β-STOH substrate (*m/z* = 300.1415). (c) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a STX substrate (*m/z* = 316.1364).

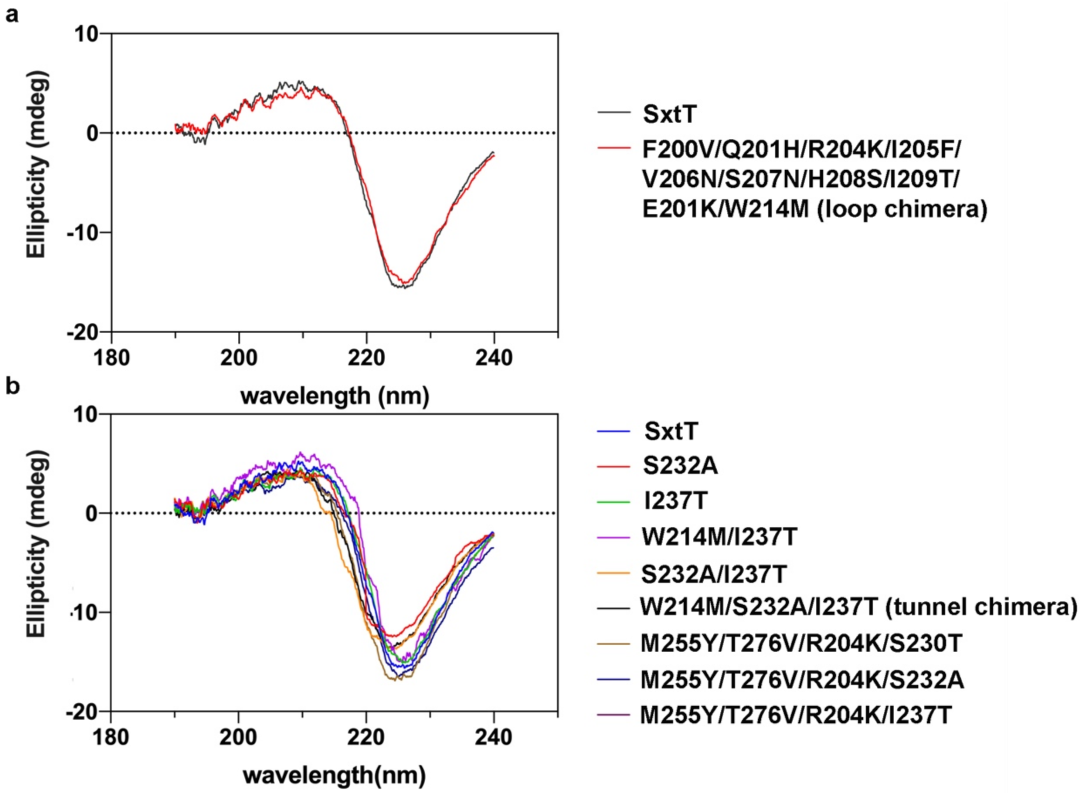
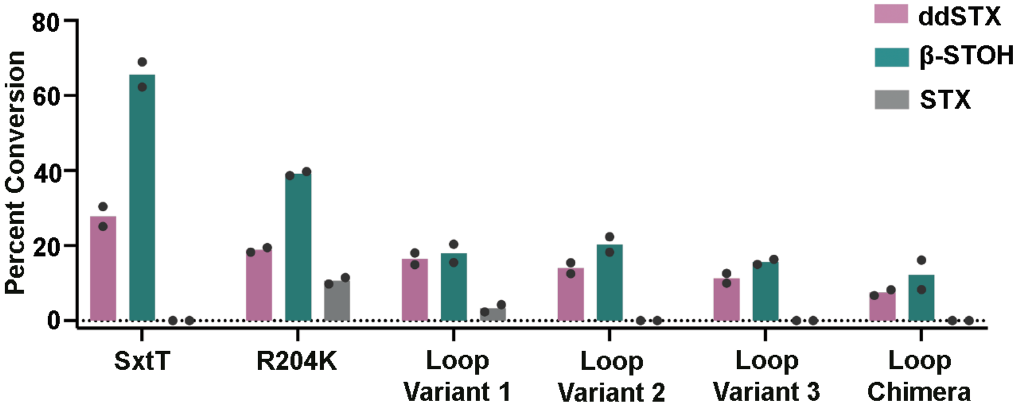


Figure 3.12. Circular dichroism spectra of SxtT variants.
This figure was prepared by Jiayi Tian. Despite low levels of measured activity, variants in the loop and tunnel region of SxtT are folded. (a) Circular dichroism (CD) data shows that the SxtT loop chimera variant (F200V, Q201H, R204K, I205F, V206N, S207N, H208S, I209T, E210K, W214M) resembles wild-type SxtT. (b) Similarly, CD data on single, double and triple variants in the tunnel region resemble wild-type SxtT. This experiment, along with gel filtration chromatography, iron analysis, and SDS-PAGE, was performed with the samples assayed in this work to ensure that the proteins used were pure, homogenous, reconstituted, and folded.

active site variant, but the selectivity remains biased toward the native site-selectivity (Figure. 3.8.g and Figure. 3.14). This finding prompted an investigation into whether a triple variant (M255Y/T276V/R204K) that combines changes in the active site and the loop would show increased affinity for hydroxylating at C11. Indeed, the combination of the active site substitutions with the R204K loop substitution shows an increase in the ability to hydroxylate STX and an approximate nine-percent increase in the amount of hydroxylation at C11 relative to the R204K variant, suggesting that the flexible loop region and active site integrate to confer selectivity in SxtT and GxtA (Figure. 3.8.f-g, Figure. 3.11 and 3.14). However, these results also show that these two structural regions are not enough to fully account for the selectivity differences between SxtT and GxtA.



Loop Variant 1: F200V/Q201H/R204K
Loop Variant 2: F200V/Q201H/R204K/I205F/V206N/S207N
Loop Variant 3: F200V/Q201H/R204K/I205F/V206N/S207N/H208S/I209T/E210K
Loop Chimera: F200V/Q201H/R204K/I205F/V206N/S207N/H208S/I209T/E210K/W214M

Figure 3.13. Changes made to the flexible loop of SxtT so that it resembles GxtA leads to a decreased ability to hydroxylate ddSTX and β-STOH.

As shown in Figure. 3.8, an R204K SxtT variant shows activity on each of the three provided substrates whereas a full loop chimera shows reduced activity on ddSTX and β-STOH and no activity on STX. Only one of the shorter loop chimera variants, F200V/Q201H/R204K, shows any activity on STX. Of note, this variant has reduced activity on all three provided substrates relative to wild type SxtT and the R204K variant. Data was measured data was measured using n=2 independent experiments.

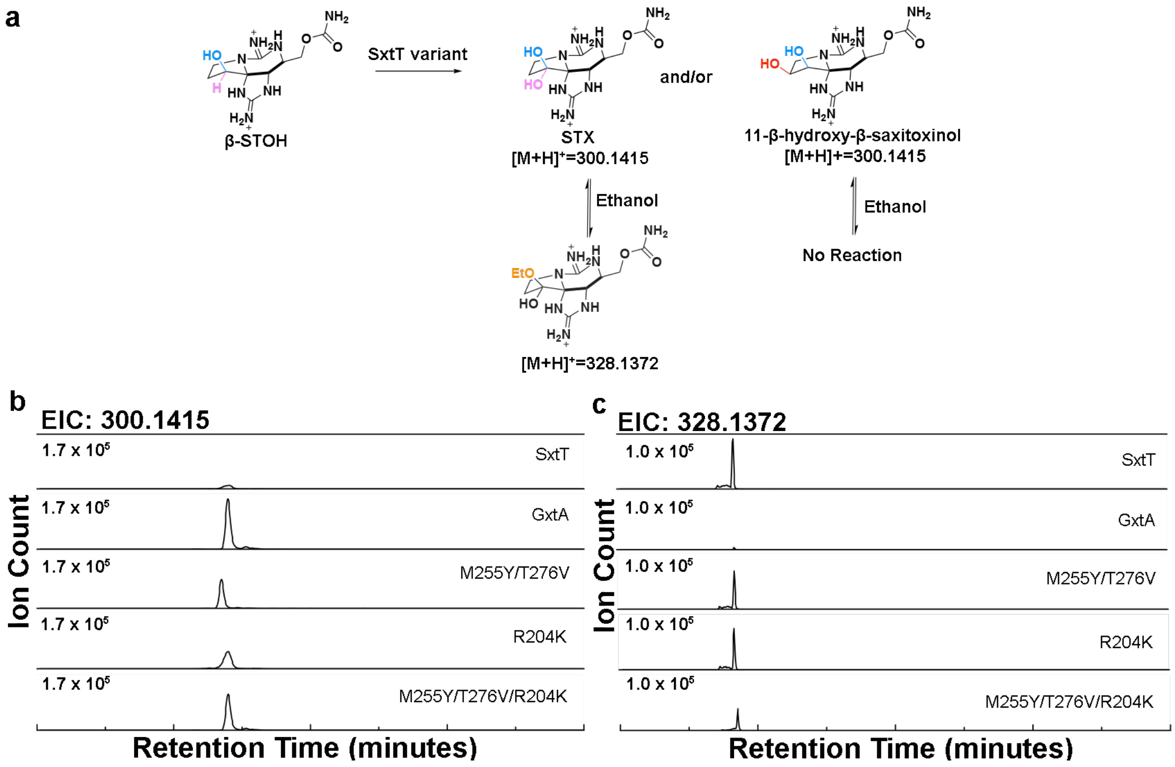


Figure 3.14. Hydroxylation on β-STOH happens at the C11 and C12 positions in SxtT M255Y/T276V, R204K, and M255Y/T276V/R204K variants.

(a) If the product of β-STOH hydroxylation by a SxtT variant is STX (hydroxylation at the C12 position), it can be monitored by looking for incorporation of ethanol into the product. If the SxtT variant instead catalyzes hydroxylation at the C11 position, ethanol will not be incorporated. (b) Extracted ion chromatograms of the SxtT, GxtA, and M255Y/T276V, R204K, and M255Y/T276V/R204K variant reaction product with of β-STOH. The *m/z* = 300.1415 value represents the mass of 11-β-hydroxy-β-STOH or STX. (c) Extracted ion chromatograms of the SxtT, GxtA, and M255Y/T276V, R204K, and M255Y/T276V/R204K variant reaction product with β-STOH. The *m/z* = 328.1372 value represents the mass of an ethanol incorporated product, which is only formed when hydroxylation happens at the C12 position.

3.3.3 Identification of a substrate entrance tunnel that is important for selectivity

Based upon the success of identifying a residue in the flexible loop that is involved in site selectivity, we experimentally probed the location of a substrate entrance tunnel to the active site. Crystals of SxtT and GxtA were pressurized with Xe, cryo-cooled, and placed in the X-ray beam. A resultant 2.69-Å resolution Xe-pressurized crystal of GxtA revealed the presence of Xe in each

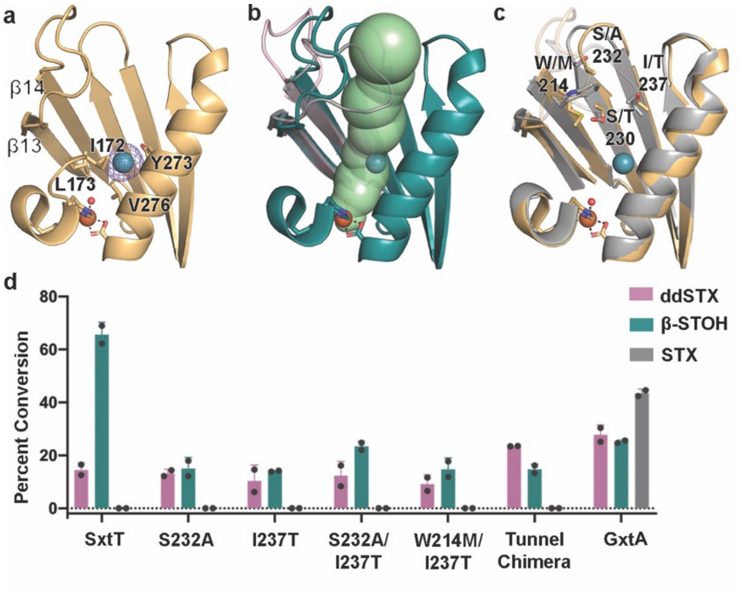


Figure 3.15. The route of substrate entrance into the active site was visualized using crystals of GxtA that were pressurized with Xenon (Xe).

(a) A Xe-pressurized structure of GxtA shows the presence of Xe near the mononuclear iron-containing active site. The location of Xe was verified by calculation of anomalous maps, which are shown in purple mesh and contoured around Xe at 3.0σ. (b) An approximate 33-Å long tunnel connects the surface of GxtA to the non-heme iron containing active site. This calculated tunnel overlaps with the position of Xe. The loops of GxtA (pink and teal) leave this tunnel open, whereas the loop of SxtT (gray) closes off this path. (c) The SxtT and GxtA regions around the tunnel differ at several positions. The flexible loops are shown at low transparency to show the differences more clearly in tunnel lining residues. (d) Only low levels of activity are observed when ddSTX and β-STOH are combined with single (S232A, I237T), double (S232A/I237T, W214M/I237T), and chimeric (W214M/S232A/I237T) SxtT tunnel variants. Of note, there is no activity on STX using SxtT or any tunnel variant tested. In panel d, data was measured in duplicate, and the data is presented as mean values ± SD. In all panels, ddSTX corresponds to dideoxysaxitoxin, β-STOH corresponds to β-saxitoxinol, and STX corresponds to saxitoxin.

protomer of the trimeric architecture (Figure. 3.15.a). The Xe binding sites were confirmed by calculation of anomalous maps (Figure. 3.15.a). Each Xe is located at the base of a computationally predicted tunnel for transporting substrate into the active site of GxtA[7]. This tunnel is impassable by the loop orientation in SxtT but is wide-open in GxtA, as reported previously[7] (Figure. 3.15.b). The Xe binding site is formed mainly by hydrophobic residues Val276, Leu173, Ile172, as well as

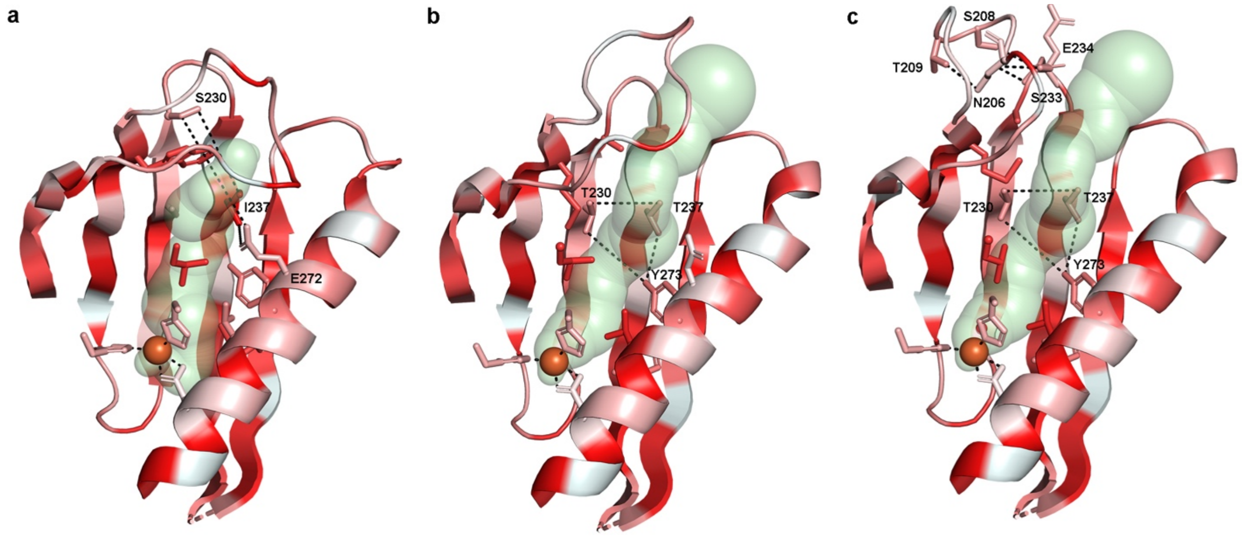


Figure 3.16. Comparison of the tunnels and hydropathy index of SxtT and GxtA.
(a) The residues that define the bottleneck of SxtT are labeled and dashed lines are drawn between them. The bottleneck of the tunnel appears to be at the entrance. Of note, due to the closed nature of the loop, the tunnel was calculated without residue Arg204 being present. (b-c) An equivalent portion of GxtA to that detailed for SxtT in panel a is shown with the residues that define the bottleneck labeled and connected by dashed lines. Unlike SxtT, the bottleneck of the GxtA tunnel is centrally located. In addition, the tunnel in GxtA is more hydrophobic than the one in SxtT and the flexible loop is more polar. All panels are colored based on the Eisenberg hydrophobicity scale[2].

Tyr273 (Figure. 3.15.a). Confirmation of the substrate entrance tunnel prompted calculation of an analogous tunnel in SxtT using the MOLEonline server43,44. This task required removal of Arg204 from the structure prior to calculation. In both proteins, the tunnels span a similar distance of approximately 33 Å. The major differences between the corresponding tunnels are the residues located at the 214 (bridges the loop and tunnel), 230, 232, and 237 positions, the overall tunnel hydrophobicity, and the width of the tunnels (Figure. 3.15.c and Figure. 3.16). In particular, the GxtA tunnel showcases a higher hydropathy index than that of the tunnel found in SxtT and a larger bottleneck radius of 2.7 Å (Figure. 3.16). This tunnel width is mainly limited by residues in the center of the tunnel (Thr230, Thr237, and Tyr273), whereas the smaller 1.7 Å tunnel radius observed in SxtT is limited by residues located near the entrance (Ser230, Ile237, and Glu272, Figure. 3.16).

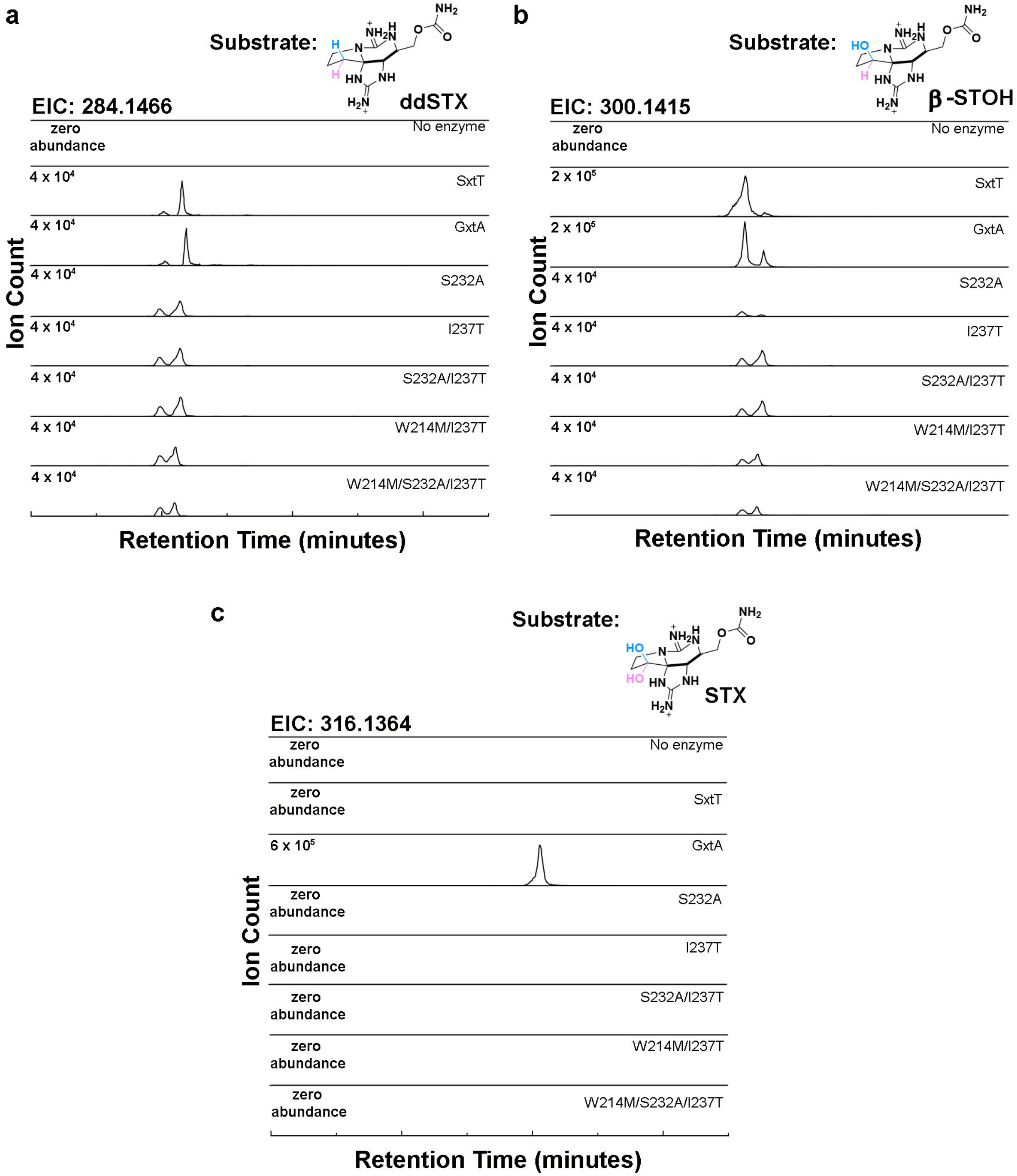


Figure 3.17. Variants of SxtT in the tunnel region show low levels of activity on ddSTX, β-STOH, and no activity on STX.

(a) The extracted ion chromatograms for SxtT tunnel variant reaction products with a ddSTX substrate. The *m/z* = 284.1466 is the mass of a hydroxylated ddSTX product. (b) The extracted ion chromatograms of the hydroxylated product formed for SxtT tunnel variants with a β-STOH substrate (*m/z* = 300.1415). (c) The extracted ion chromatograms of the hydroxylated product formed for SxtT tunnel variants with a STX substrate (*m/z* = 316.1364). As shown here, none of the single, double, or tunnel chimera variant have activity on STX.

Using the experimentally confirmed substrate access tunnel as a guide, we tested the importance of tunnel lining and bottleneck residues in dictating the site selectivity of the SxtT catalyzed reaction using site-directed mutagenesis experiments (Figure. 3.10, Table 3.3, and Figure. 3.15.d). Targeted residues in SxtT (Trp214, Ser232, and Ile237) included those that are larger, or differ in polarity from the corresponding residues in GxtA (Met214, Ala232, and Thr237). Again, as described for the SxtT loop variants, to determine whether these tunnel variants showed changes in site selectivity, we tested their activity using VanB as an electron donor and ddSTX, β-STOH, and STX as substrates. Here, LC-MS experiments showed that single (I237T, S232A), double (S232A/I237T and W214M/I237T), and chimeric (W214M/S232A/I237T) tunnel variants, despite showing similar CD spectra to wild-type SxtT, show no activity on STX and lower amounts of activity on ddSTX and β-STOH than GxtA (Figure. 3.15.d, Figure. 3.12b and 3.17).

3.3.4 Coordinated interplay of protein regions dictates substrate-specificity and site-selectivity

As additively varying the sequences in the active site and loop regions led to an increased ability to hydroxylate at C11, the effect of combining changes in the tunnel, active site, and loop regions was subsequently tested. First, individual tunnel residue mutations were introduced into the construct that contains two mutations in the active site and one on the loop to produce

M255Y/T276V/R204K/S230T, M255Y/T276V/R204K/S232A, and M255Y/T276V/R204K/I237T variants (Figure. 3.10 and Table 3.3). These new variants, which were confirmed to be properly folded by CD, still lacked the propensity to hydroxylate STX

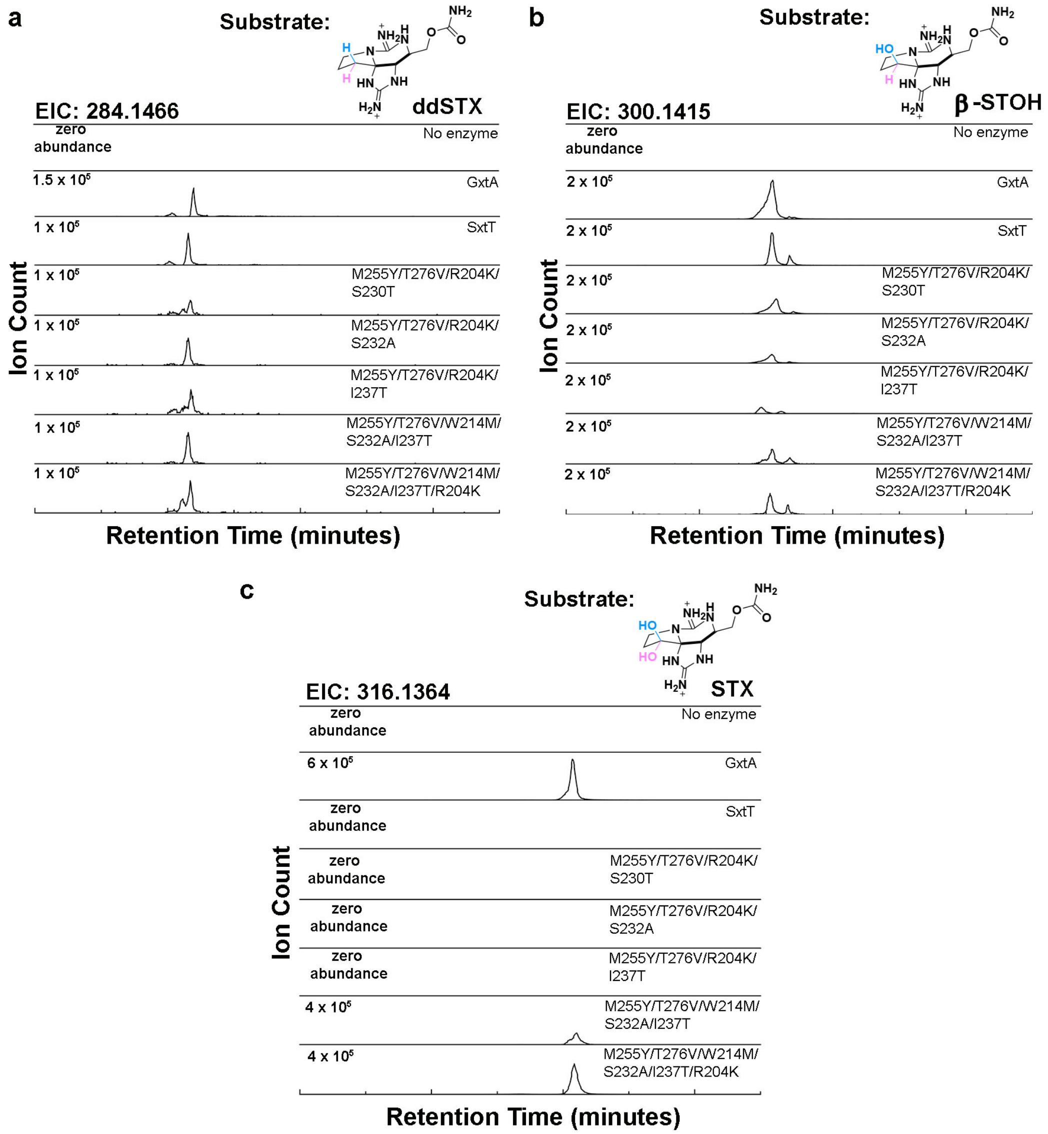


Figure 3.18. Variants of SxtT that combine changes in the tunnel chimera with changes in the active site and loop region show differing levels of activity on ddSTX, β -STOH, and STX. (a) The extracted ion chromatograms for SxtT variant reaction products with a ddSTX substrate. The *m/z* = 284.1466 is the mass of a hydroxylated ddSTX product. (b) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a β-STOH substrate (*m/z* = 300.1415). (c) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a STX substrate (*m/z* = 316.1364). As shown here, the M255Y/T276V/R204K,

M255Y/T276V/W214M/S232A/I237T, and M255Y/T276V/R204K/W214M/S232A/I237T variants can each hydroxylate STX, suggesting they can now hydroxylate at C11.

(Figure. 3.24.a, Figure. 3.12.b and Figure. 3.18). However, when we cumulatively changed three tunnel lining residues and two active site residues (M255Y/T276V/W214M/S232A/I237T), activity was detected on ddSTX, β-STOH, and STX (Figure. 3.24.a and Figure. 3.18). Introduction of the R204K loop mutation into this construct (M255Y/T276V/W214M/S232A/I237T/R204K) showed, for the first time, the highest level of activity with the STX substrate, suggesting that the selectivity of SxtT had been inverted to resemble GxtA (Figure. 3.24.a and Figure. 3.18). The ability to catalyze hydroxylation at C11 in this latter variant was confirmed by looking for ethanol incorporation in the reaction that used β-STOH as a substrate (Figure. 3.24.b and Figure. 3.19). In this case, the essential lack of identification of an ethanol-incorporated product confirmed that 99-percent of hydroxylation occurred at the C11, rather than the C12 position on the substrate (Figure. 3.24.b). These results suggest that the flexible loop region, the tunnel, and the active site integrate to confer selectivity and specificity in SxtT and GxtA.

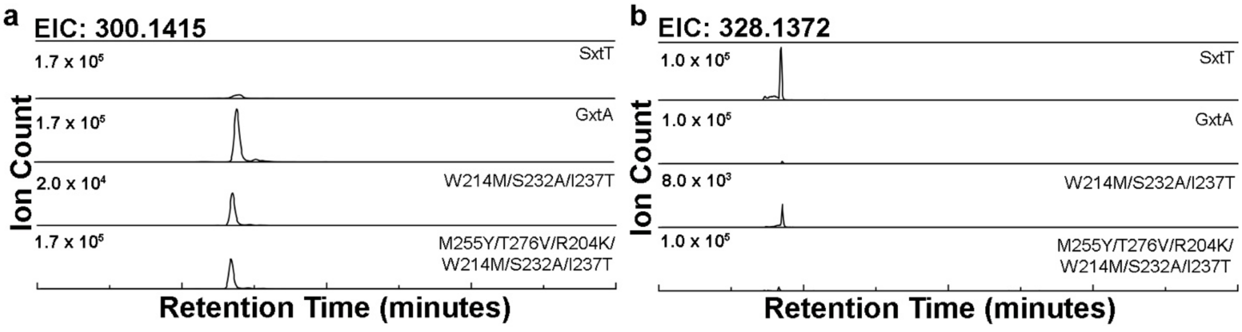


Figure 3.19. A SxtT variant that combines changes in the active site, loop, and tunnel region, mirrors the selectivity of GxtA and catalyzes hydroxylation primarily at C11.
(a) Extracted ion chromatograms of the SxtT, GxtA, W214M/S232A/I237T variant, and M255Y/T276V/R204K/W214M/S232A/I237T variant reaction product with of β-STOH. The *m/z* = 300.1415 value represents the mass of 11-β-hydroxy-β-STOH. (b) Extracted ion chromatograms of the SxtT, GxtA, W214M/S232A/I237T variant, and

M255Y/T276V/R204K/W214M/S232A/I237T variant reaction product with β-STOH. The *m/z* = 328.1372 value represents the mass of an ethanol incorporated product. The reaction scheme for this figure can be seen in Figure 3.14.a.

Intriguingly, our structures did not reveal any interactions that were formed between tunnel lining residues and substrate in SxtT or GxtA. Thus, we hypothesize that the tunnel is involved in transporting and handing-off the substrate to the active site in the correct orientation for hydroxylation. Commensurate with this hypothesis, 90-percent of the hydroxylated product made by the SxtT tunnel chimera variant (W214M/S232A/I237T), when β-STOH was provided as a substrate, was hydroxylated at the C11 position (Figure. 3.24.b and Figure. 3.19).

3.3.5 Identification of the architectural parameters that dictate substrate scope

The altered selectivity of the SxtT tunnel chimera variant and the ability of the active site-tunnel-loop variant (M255Y/T276V/W214M/S232A/I237T/R204K) to hydroxylate STX inspired an investigation into whether these variants, like GxtA42, would also hydroxylate other STX-derived compounds, including α-STOH, decarbamoyl (dc)-α-STOH, dc-STX, ddSTX, and dc-ddSTX (Figure. 3.24.c). The three α-hydroxyl group-containing STX analogs (α-STOH, dc-α-STOH, dc-STX) are not accepted by wild-type SxtT. However, as described for STX, when VanB was used as a reductase and LC-MS was used to assess hydroxylation, the M255Y/T276V/R204K, W214M/S232A/I237T, and M255Y/T276V/R204K/W214M/S232A/I237T variants show an increasing ability to hydroxylate these compounds (Figure. 3.24c and Figure. 3.20). In particular, as previously described for wild-type GxtA, the tunnel chimera variant (W214M/S232A/I237T) catalyzes hydroxylation on α-STOH at the C11 position (Figure. 3.24.b and Figure. 3.21). Again, this data suggests that the tunnel is important for transporting and placing the substrate in the correct orientation for hydroxylation in the active site. Intriguingly, despite the noted ability of the tunnel chimera variant to accept substrates with an α-hydroxyl group, it does not hydroxylate STX.

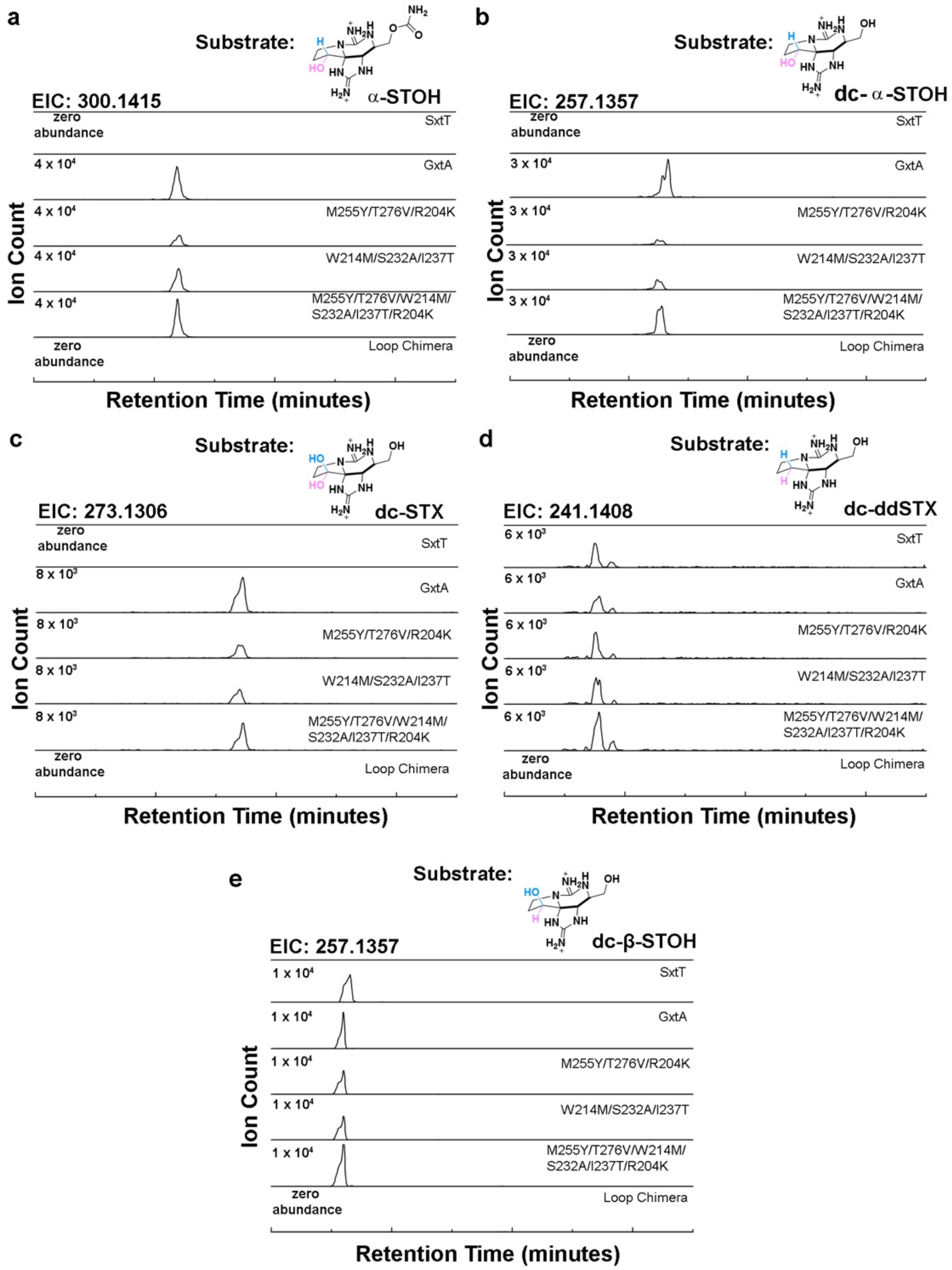


Figure 3.20. Variants of SxtT that combine changes in the active site, tunnel, and loop region show activity on α-STOH, dc-α-STOH, dc-STX, dc-ddSTX and dc-β-STOH.
(a) The extracted ion chromatograms for SxtT variant reaction products with an α-STX substrate. The *m/z* = 300.1415 is the mass of a hydroxylated α-STOH product. (b) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a dc-α-STOH substrate

(*m/z* = 257.1357) (c) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a dc-STX substrate (*m/z* = 273.1306) (d) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a dc-ddSTX substrate (*m/z* = 241.1408) (e) The extracted ion chromatograms of the hydroxylated product formed for SxtT variants with a dc-β-STOH substrate (*m/z* = 257.1357). In these panels, ddSTX corresponds to dideoxysaxitoxin, β/α-STOH corresponds to β/α-saxitoxinol, and STX corresponds to saxitoxin. The abbreviation dc-represents dicarbamoyl.

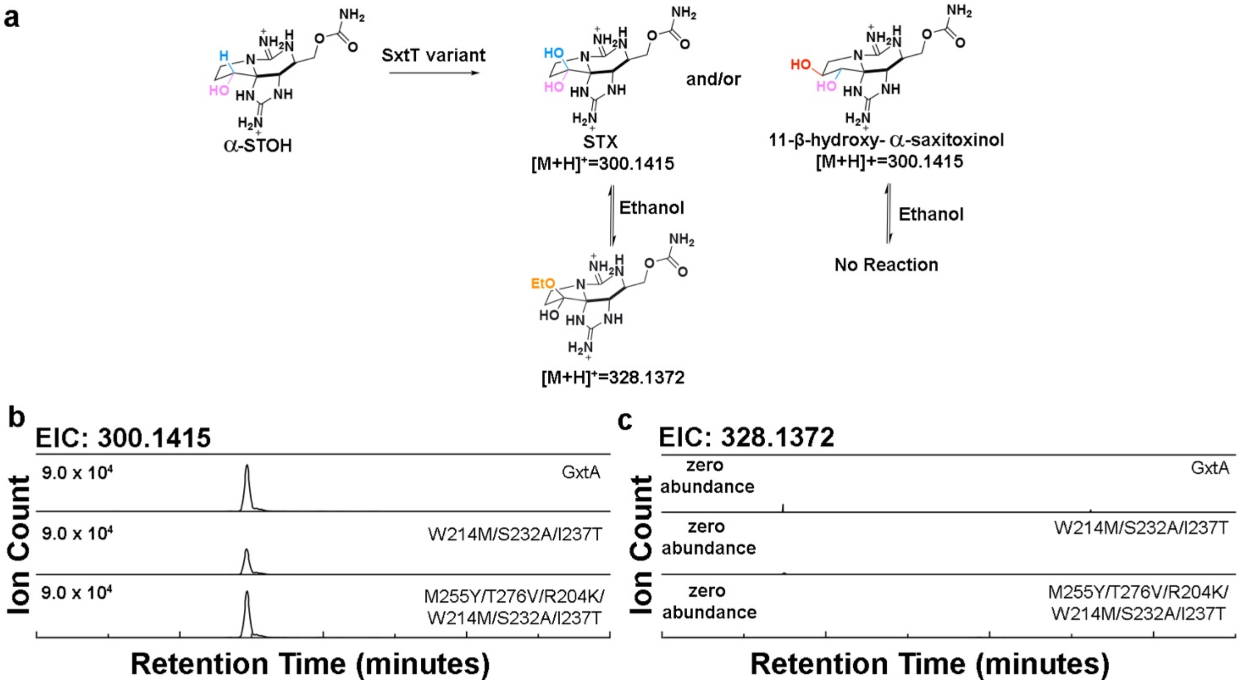


Figure 3.21. The SxtT tunnel chimera variant and variant that combines changes in the active site, loop, and tunnel region catalyzes hydroxylation at the C11 position of α-STOH.
(a) Extracted ion chromatograms of the GxtA, W214M/S232A/I237T variant, and M255Y/T276V/R204K/W214M/S232A/I237T variant reaction product with α-STOH. The *m/z* = 300.1415 value represents the mass of 11-β-hydroxy-α-STOH. (b) Extracted ion chromatograms of the GxtA, W214M/S232A/I237T variant, and M255Y/T276V/R204K/W214M/S232A/I237T variant reaction product with α-STOH. The *m/z* = 328.1372 value represents the mass of an ethanol incorporated product, which would only be observed if hydroxylation occurred at C12.

To accomplish this feat, the variant that also showcases changes in the active site and loop region (M255Y/T276V/R204K/W214M/S232A/I237T) is required (Figure. 3.24.b and Figure. 3.19).

Like wild-type GxtA, this variant also hydroxylates α-STOH at the C11 position (Figure. 3.24.b and Figure. 3.21).

As expected for wild-type SxtT, the highest amount of activity is observed on a β-STOH substrate (Figure. 3.24.c). Correspondingly, activity on this native substrate decreases as SxtT is mutated to resemble GxtA. For ddSTX, dc-ddSTX, and dc-β-STOH, each of which have been previously shown to be hydroxylated by both wild-type enzymes[42], the highest level of activity is observed with wild-type GxtA (Figure. 3.24.c). For ddSTX, dc-ddSTX, and dc-β-STOH, consistent low levels of activity are encountered in the W214M/S232A/I237T variant, whereas activity is amplified to resemble wild-type GxtA by also changing the active site and loop regions (Figure. 3.24.c, Figure. 3.17, 3.18 and 3.20). Collectively, these results confirm that the substrate specificity and selectivity of SxtT can be changed to mirror that of GxtA by changing just six residues that span the active site, tunnel, and loop regions of the protein.

3.3.6 Identification of conserved loop and tunnel regions in additional Rieske oxygenases

The parallels that can be drawn between SxtT, GxtA, and other Rieske oxygenases make it tempting to consider whether these design principles are conserved in other enzymes of this class. Indeed, our results on the M255Y/T276V/R204K variant, which shows a pronounced ability to hydroxylate STX, are reminiscent of what is observed in the caffeine demethylating Rieske oxygenases NdmA and NdmB8. In this system, replacement of NdmA loop with the corresponding loop residues from NdmB, coupled with changes in the active site, results in altered *N*-demethylase activity8. Structural alignment of these enzymes with SxtT and GxtA reveals that the implicated selectivity loop in NdmA, NdmB, SxtT, and GxtA, similarly occurs between the β13 and β14 strands (Figure. 3.22.a-b). To determine whether the flexible loop and tunnel are more common among Rieske oxygenases, we analyzed the B-factors of the additional six other structurally characterized α3-oxygenases (Figure. 3.22). These structurally characterized enzymes, include those that are involved in *N*-demethylation of *N*,*N*- dimethylproline (Stc2)[12], as well as those that are responsible for the degradation of cholesterol (KshA)[10], carnitine (CntA)[11], 2-oxoquinoline

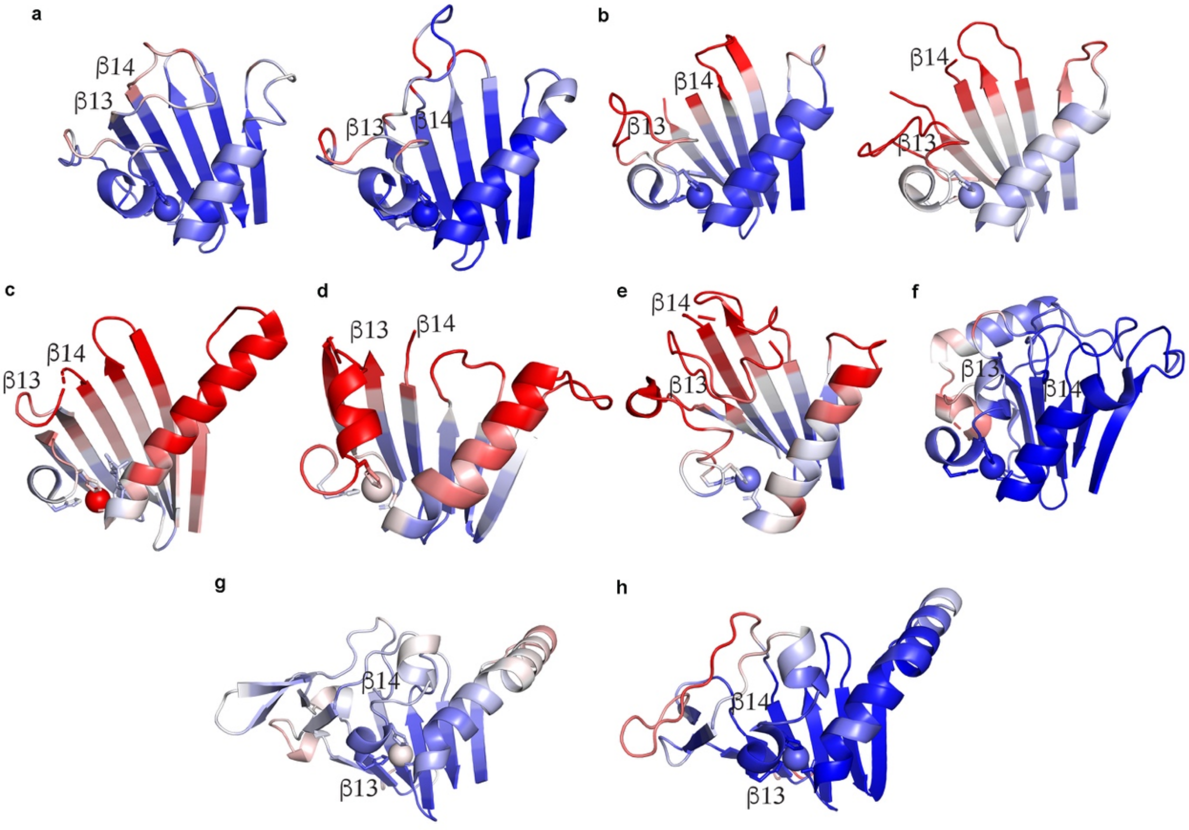


Figure 3.22. The loop that connects the β13 and β14 strands of the C-terminal iron binding domain in SxtT and GxtA, as well as in most other structurally characterized α3 Rieske oxygenases, is flexible.

(a) For reference the structures of SxtT (left) and GxtA (right) are provided for comparison. (b) The loop that connects β13 and β14 in NdmA (left) and NdmB (right) is missing from the crystal structures due to disorder (PDB: 6ICK, 6ICL)[3]. (c) A flexible loop that is partly missing connects β13 and β14 in KshA (PDB: 4QCK)[4]. (d) CntA (PDB: 6Y9D)[5], like NdmA and NdmB, has a β13 and β14 connecting loop that is mostly missing. (e) DdmC (PDB: 3GKE)[6] also has a slight disordered and unmodeled loop that connects β13 and β14 in the C-terminal domain. (f) There is some amount of flexibility in the Stc2 (PDB: 3VCP)[7] loop, but it is longer than the SxtT and GxtA loops. (g) Finally, CARDO (PDB: 2DE7)[8,9] and (h) OMO (1Z03)[10] both showcase very long loops that have extra secondary structure. The loop of OMO is more flexible than that of CARDO. In all panels, the B-factors are plotted on a scale of 20–60 Å^2 with increasing values colored in blue, white, and red, respectively.

(OMO)[13], dicamba (DdmC)[9], and carbazole (CARDO)[14,45]. Through analysis of the structures' B-factors, a loop of high flexibility that exists between the β13 and β14 strands of the C-terminal domain, analogous to that observed in SxtT and GxtA, can be identified in the structures of NdmA, NdmB, KshA, CntA, and DdmC (Figure. 3.22 and Figure. 3.25.a). Consistently, the loop implicated in NdmA and NdmB selectivity is so flexible that it was not modeled into the crystal structures[8] (Figure. 3.22.b). Similarly, the flexible loop region in KshA is partially missing and has been shown to play a role in dictating the substrate preference of KshA homologs[10,38] (Figure. 3.22.c). A disordered and partially missing loop is also observed in the apo- and substrate-bound structures of CntA11 and in DdmC[9] (Figure. 3.22.d-e). Importantly, in these five Rieske oxygenases a tunnel that leads from this loop region to the mononuclear iron site could be calculated using the MOLEonline server[43,44] (Figure. 3.23). In all cases, this tunnel traverses across the same secondary structure seen in SxtT and GxtA and is oriented similarly with respect to the flexible loop (Figure. 3.23 and Figure. 3.25.a). In Stc2, there is flexibility in the analogous loop, but based on its secondary structure which folds in towards the substrate bound in the active site, an active site tunnel is not obvious, nor can it be calculated (Figure. 3.23.f). In CARDO, a tunnel was also not observed, presumably due to the region that connects β13 and β14 being much longer and containing an extra β-strand and β-hairpin relative to its counterparts in SxtT and GxtA. Nevertheless, residues in this region have been suggested to be dynamic and to gate the active site upon substrate binding[45-48] (Figure. 3.22.g). Finally, in OMO, an ordered, mainly hydrophobic loop that resembles the long loop in CARDO similarly precludes identification of an active site tunnel (Figure. 3.22.h). Importantly, however, an adjacent region in the tertiary architecture of OMO has been implicated in substrate channeling to the active site[13]. Thus, these results suggest the possibility that like SxtT and GxtA, these other structurally characterized Rieske oxygenases will employ similar architectural features, including a loop, tunnel, and active site to recognize and orient a substrate for site-selective chemistry.

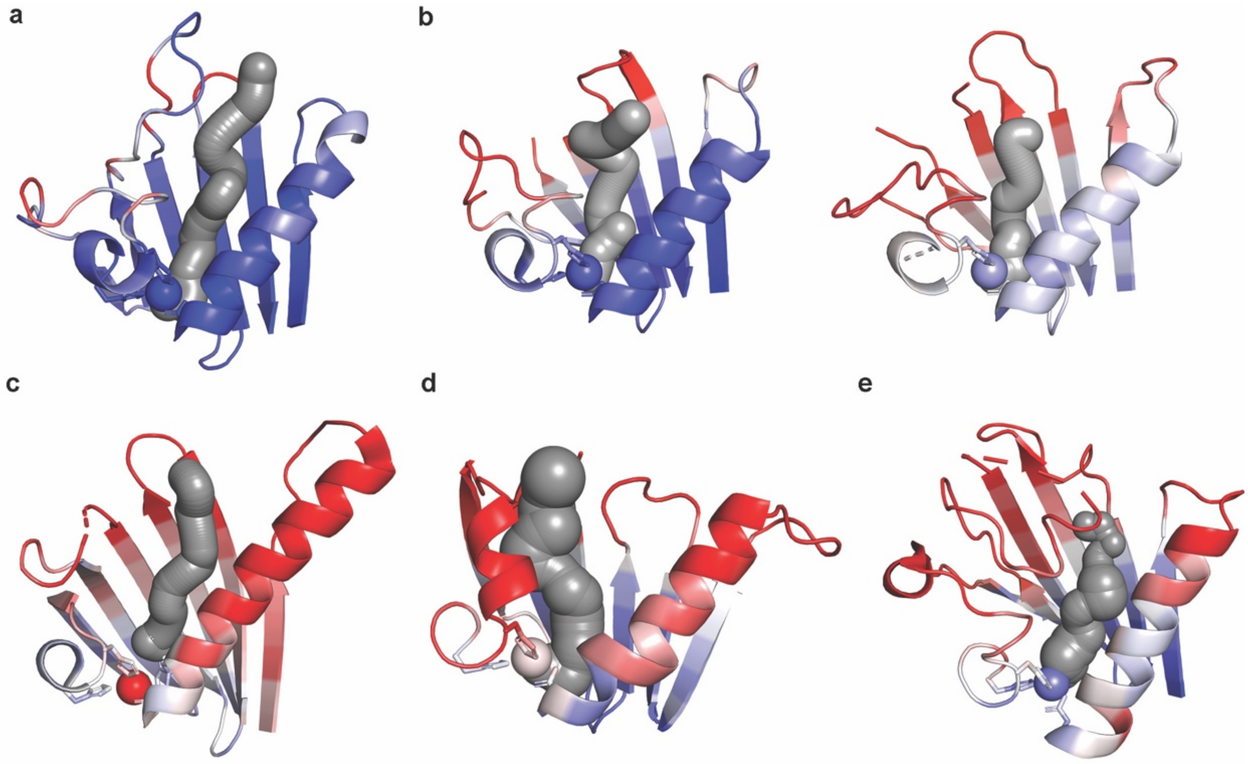
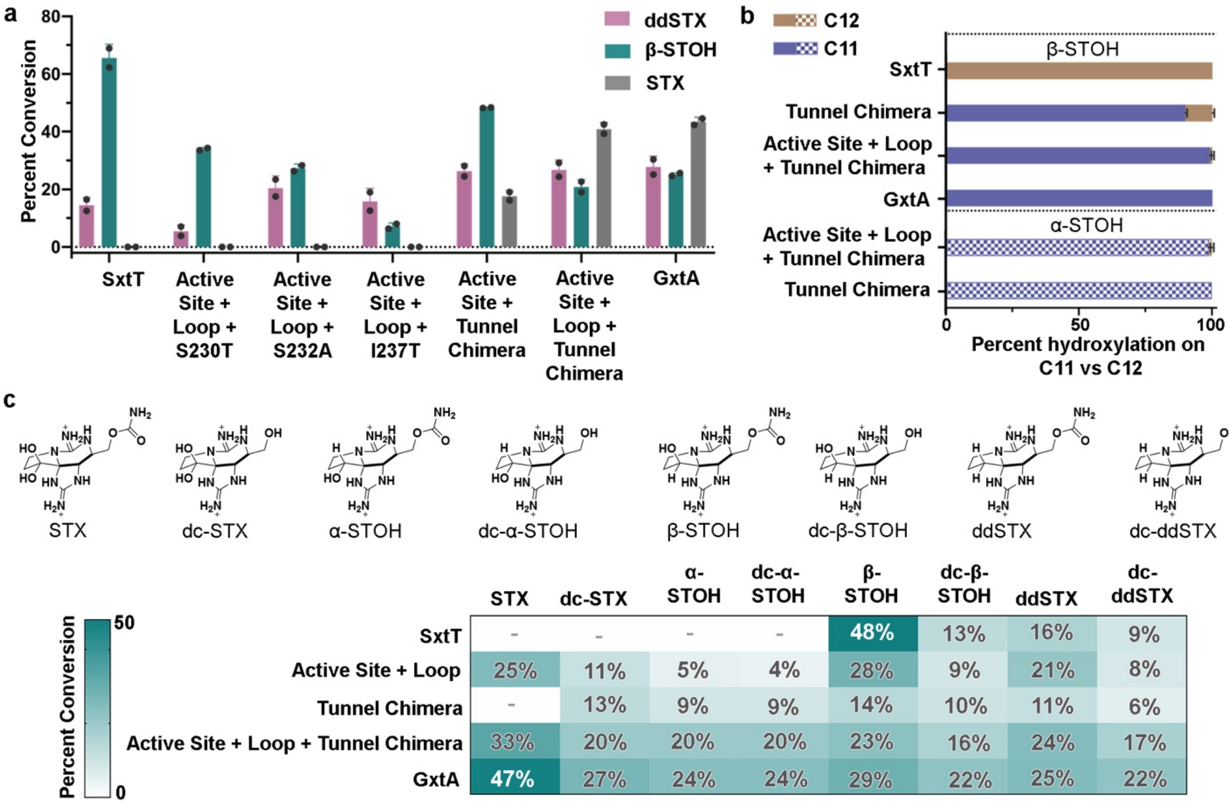


Figure 3.23. A tunnel that spans a similar region to that observed in GxtA was also observed in NdmA, NdmB, KshA, CntA, and DdmC.

(a) The structure of GxtA with its calculated tunnel determined in this work is provided for comparison. (b) Both NdmA (left) and NdmB (right) have tunnels that lead from the surface of the protein to the non-heme iron center (PDB: 6ICK, 6ICL)[3]. (c) A tunnel can also be calculated in KshA that is oriented similarly to that observed in GxtA (PDB: 4QCK)[4]. (d) The CntA (PDB: 6Y9D)[5] tunnel occludes visualization of the disordered loop that is shown in Figure 3.19.d. (e) DdmC (PDB: 3GKE)[6] also showcases a tunnel for substrate access that appears to be gated by the flexible loop that connects β13 and β14.

3.4 Discussion

Here, we elaborated on the design principles that dictate how the Rieske oxygenases SxtT and GxtA catalyze site-selective monohydroxylation reactions. In particular, the specificity and selectivity of these enzymes requires a coordinated interplay of three protein regions: the active



	STX	dc-STX	α-STOH	dc-α-STOH	β-STOH	dc-β-STOH	ddSTX	dc-ddSTX
SxtT	-	-	-	-	48%	13%	16%	9%
Active Site + Loop	25%	11%	5%	4%	28%	9%	21%	8%
Tunnel Chimera	-	13%	9%	9%	14%	10%	11%	6%
Active Site + Loop + Tunnel Chimera	33%	20%	20%	20%	23%	16%	24%	17%
GxtA	47%	27%	24%	24%	29%	22%	25%	22%

Figure 3.24. A combination of residues from the active site, loop, and tunnel region are responsible for the selectivity and substrate scope of the SxtT catalyzed reaction.
(a) The effect of combining residue variants from the active site (M255Y, T276V), loop (R204K), and tunnel region (W214M/S232A/I237T) was tested using ddSTX, β-STOH, and STX as substrates. (b) The position of hydroxylation on a β-STOH substrate was confirmed using an ethanol-incorporation experiment. The essential lack of an ethanol-derivatized product in the active site + loop + tunnel chimera variant (99-percent of hydroxylation at C11) suggested that selectivity had been changed to resemble GxtA (solid colors). The position of hydroxylation when α-STOH was used as a substrate was tested in the tunnel chimera and the active site + loop + tunnel chimera variants (patterned colors). Both variants, like GxtA, showed a preference for hydroxylating at the C11 position. In the tunnel chimera, using an α-STOH substrate, 98-percent of the product is hydroxylated at the C11 position, whereas in the active site + loop + tunnel chimera variant, 100-percent of the product showcases a hydroxyl group at the C11 position. (c)

The ability to hydroxylate naturally occurring saxitoxin analogs (top panel) was evaluated for the active site + loop, tunnel chimera, and the active site + loop + tunnel chimera variants. The activity is shown as a heat map with lowest activity in white and the highest activity in dark teal. The percent conversion of the native substrates of SxtT and GxtA are shown in white. In panels a and b, data was measured in duplicate, and the data is presented as mean values ± SD. In panel c, the data was also measured in duplicate and is presented as the mean value of these measurements. In panels a and c, ddSTX corresponds to dideoxysaxitoxin, β/α-STOH corresponds to β/α-saxitoxinol, and STX corresponds to saxitoxin. The abbreviation dc-represents decarbamoyl.

site, a flexible loop, and the substrate entrance channel (Figure. 3.24 and Figure. 3.25a). Structurally it was shown that the native hydroxylated substrates of SxtT and GxtA bind in the active site similarly to ddSTX, albeit with some important key differences. Namely, β-STOH in SxtT makes a through-water interaction with active site residue Tyr198 and STX in GxtA interacts with Tyr155 (Figure. 3.3). A comparison of these structures and those previously determined[7] also revealed important orientations of a flexible loop that can serve as a gate to the active site by assuming open (GxtA) or closed states (SxtT, Figure 3.8.a-e). The importance of this loop is demonstrated by the chimeric loop variants described in this work, which show that changing the SxtT loop to resemble the GxtA counterpart results in lower levels of activity ddSTX and β-STOH (Figure. 3.8.f and Figure. 3.13). Outside of considering the overall loop architecture, in SxtT, the loop is important as it interacts with a bound substrate using residue Arg204 (Figure. 3.1.d and 3.3.b). This Arg204 interaction, coupled with additional interactions that are formed between active site residues (Thr276, Gln226, Ser230, Asp239, Tyr273) and the guanidinium groups of the five- and six-membered rings of β-STOH, ensures that three regions of ddSTX and β-STOH are recognized by SxtT (Figure. 3.25.b). The lower activity of SxtT on ddSTX shown here, and previously noted[42], is likely correlated to the ability of β-STOH to interact with Tyr198 via a water molecule (Figure. 3.3.b). In GxtA, there are only two recognition points for the native STX substrate. Gln226 and Asp239 interact with the guanidinium of the five-membered ring of ddSTX, β-STOH, and STX, and Tyr255 interacts with the α-hydroxyl group of STX, which is installed by SxtT (Figure. 3.1.e, 3.3.c-d, and Figure. 3.25.c). The conservation of Gln226 and Asp239 between SxtT and GxtA suggests that any conformational freedom afforded in the SxtT active site could allow rotation of the substrate and formation of the interactions that permit hydroxylation at C11 (Figure. 3.25.b-c).

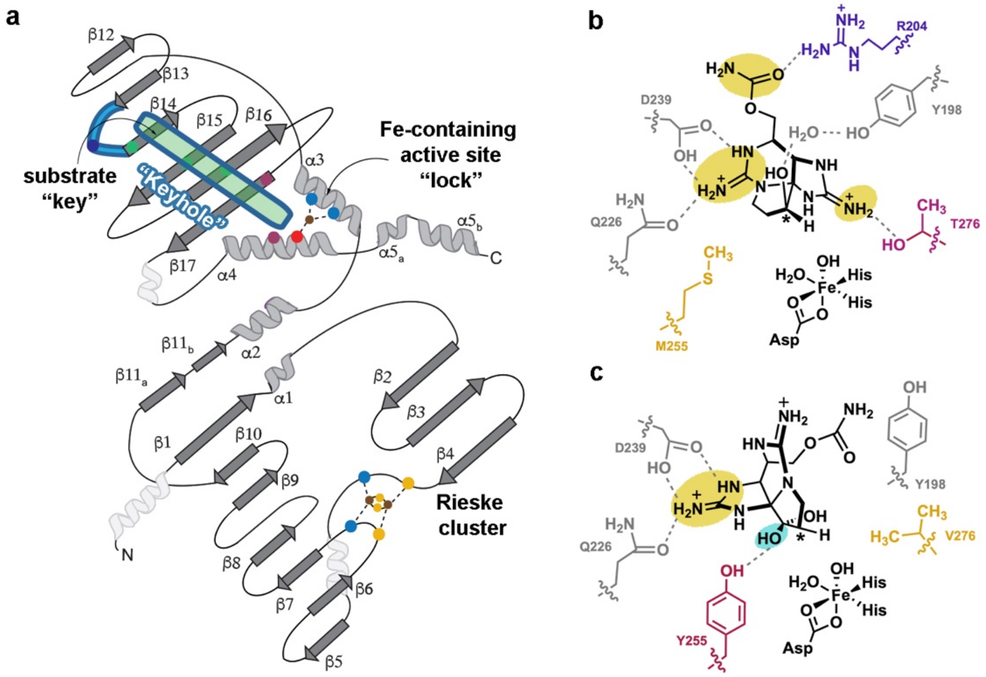


Figure 3.25. A coordinated interplay of three protein regions in SxtT are important for dictating selectivity.

(a) A topology diagram of SxtT/GxtA. The active site (dark pink circles represent residues 255 and 276), loop region (light blue, dark blue circle represents residue 204) and tunnel (green, green circles represent residues at the 214, 232, and 237 positions) each have residues that are involved in dictating selectivity. The so-called "keyhole" (loop and tunnel region) of these proteins is outlined in blue. The secondary structure that is conserved among the structurally characterized α3 Rieske oxygenases is shown in dark grey and light grey[5]. The three helices in white are not conserved. (b) Three regions of β-STOH that are also found in STX are recognized by SxtT (yellow): the carbamate is recognized by Arg204 from the flexible loop and the guanidium ions are recognized by active site residues Thr276, Gln226, and Asp239. Of these residues, only Thr276 and Arg204 are unique to SxtT. (c) Two regions of STX are recognized by active site residues in GxtA: the guanidinium moiety of the five-membered ring interacts with Gln226 and Asp239 and the α-hydroxyl interacts with Tyr255. Of these recognition points, only the α-hydroxyl is unique (teal) to STX and only Tyr255 is unique to GxtA.

Indeed, in the R204K variant, the SxtT substrate loses one of its two unique interacting residues, and a low level of non-native hydroxylation is observed (Figure. 3.8.f-g). A similar phenomenon is observed in the single and double active site variants of SxtT; loss of Thr276, which interacts with the guanidinium of the five-membered ring in ddSTX and β-STOH, allows SxtT to catalyze non-native hydroxylation at C11[7]. Likewise, loss of interactions with the SxtT active site by removal of the carbamate sidechain from the native substrates would account for the lower levels of activity observed when SxtT is given dc-ddSTX or dc-β-STOH as a substrate (Figure. 3.24.c). Finally, we now know that Tyr255 in GxtA interacts with the α-hydroxyl group of STX (Figure. 3.3.c). Introduction of this residue in place of Met255 in SxtT in isolation or combined with a T276V and R204K mutation can account for the ability of previously investigated SxtT M255Y and M255Y/T276V variants[7] as well as the M255Y/T276V/R204K variant to catalyze low levels of hydroxylation on STX and at the C11 position of β-STOH (Figure. 3.8.f-g). Similarly, the presence of Tyr255 is likely the reason that the M255Y/T276V/R204K SxtT variant can accommodate additional substrates that like STX, contain a C12 α-hydroxyl group (α-STOH, dc-α-STOH, and dcSTX) (Figure. 3.24.c and Figure. 3.25.b-c).

As described for the double active site SxtT variant[7], changing just the active site and loop residue Arg204 to resemble GxtA adds functionality to SxtT as it can now hydroxylate a substrate at either the C11 or C12 position and accommodate a broader range of substrates (Figure. 3.8.f). However, the M255Y/T276V/R204K variant still shows a preference for a β-STOH substrate and hydroxylation at C12 (Figure. 3.8.g). To bestow SxtT both the substrate-specificity and site-selectivity of GxtA requires consideration of a third architectural parameter, a tunnel that leads from the surface of the protein into the active site. Indeed, in this work, we demonstrated that changing just three residues in the tunnel of SxtT to mirror GxtA (W214M/S232A/I237T) allows SxtT to preferentially hydroxylate β-STOH at the C11 position (Figure. 3.24.a-b). The W214M/S232A/I237T variant, like the M255Y/T276V/R204K variant, can also hydroxylate α-STOH, dc-α-STOH, and dc-STX substrates, albeit to a greater extent (Figure. 3.24.c). The latter α-STOH substrate, like β-STOH, is preferentially hydroxylated at the C11 position (Figure. 3.24.b). Yet, despite the observed ability of this tunnel variant to catalyze hydroxylation at C11, like wild-type SxtT, it does not accept STX as a substrate. However, when these tunnel residues changes are accompanied by the addition of GxtA active site residues (M255Y/T276V/W214M/S232A/I237T), STX is hydroxylated. Finally, when these tunnel residue changes are accompanied by both changes to the active site and to the loop (M255Y/T276V/R204K/W214M/S232A/I237T), STX conclusively becomes the preferred substrate (Figure. 3.24.a-b). These results indicate that SxtT, can showcase the substrate-specificity and site-selectivity of GxtA by (i) the removal of both unique residues that interact with the native substrate of SxtT (Thr276 and Arg204), (ii) the addition of a residue that interacts with the α-hydroxyl of STX in GxtA (Tyr255), and (iii) the replacement of three SxtT tunnel lining residues with the GxtA counterparts to change the primary site of hydroxylation from C12 to C11. As described in the keyhole-lock-key model of enzyme catalysis[36], the substrates of these enzymes must be recognized and positioned by the "keyhole", or the loop and tunnel regions, to be correctly fit into the active site for hydroxylation (Figure. 3.25.a).

To investigate which architectural features of SxtT and GxtA are more prevalent in the Rieske oxygenase family, which contains more than 80,000 annotated Rieske oxygenase sequences and only 18 unique crystal structures[5,7], we undertook an investigation into whether α3 enzymes would showcase similar loop and tunnel regions. We found that in almost all cases, a loop of high flexibility exists in an analogous region to that observed in SxtT and GxtA (Figure. 3.22). Furthermore, for most of the proteins that possess a flexible loop, we can calculate a tunnel to the active site that is comparable to what is seen in SxtT and GxtA (Figure. 3.23). These results suggest the possibility that like SxtT and GxtA, some of these other structurally characterized Rieske oxygenases will employ three architectural regions to orient and position substrate, and to dictate substrate scope. It is also intriguing to consider whether these architectural parameters will be employed in the α3β3 Rieske oxygenases. Based on recent computational studies performed on Naphthalene 1,2 dioxygenase, that showed the value of using an entrance tunnel to predict substrate scope, we hypothesize the active site tunnels have thus far been underappreciated in studying Rieske oxygenase selectivity[24]. Additionally, based on recent work that shows changes to the flexible loop region in cumene dioxygenase gives rise to altered reactivity, substrate scope and reaction selectivity, and may alter the substrate access tunnel[37], we hypothesize that the cooperativity between the three protein regions in SxtT and GxtA may also be used by the α3β3 enzymes. Thus, we are excited to explore the possibility that these design principles will be conserved across the Rieske oxygenase class of enzymes and to decipher whether they can be exploited to predictively repurpose and manipulate the diverse catalytic repertoire of Rieske oxygenases.

3.5 Data availability

Protein coordinates and structure factors have been submitted to the Protein Data Bank under accession codes 7SZH [http://www.rcsb.org/structure/7SZH] (SxtT with β-STOH bound), 7SZF [http://www.rcsb.org/structure/7SZF] (GxtA with β-STOH bound), 7SZE [http://www.rcsb.org/structure/7SZE] (GxtA with STX bound), and 7SZG [http://www.rcsb.org/structure/7SZG] (Xenon-pressurized GxtA).

3.6 Acknowledgements

Use of the Stanford Synchrotron Radiation Light Source, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894).

3.7 Methods

3.7.1 Site-directed mutagenesis

The Agilent QuickChange Lightning Site-Directed Mutagenesis Kit and Bio-Rad C1000 Thermal Cycler were used to introduce the desired mutations into SxtT. In brief, each 50 µL reaction mixture contained 5 µL of 10× reaction buffer, 1 µL of 2 ng/µL pMCSG-9-SxtT-containing construct, 0.5 µL or 125 ng of each oligonucleotide primer, 1 µL of dNTP mix, 1.5 µL of Agilent QuickSolution reagent, and 1 µL of Agilent QuickChange Lightning Enzyme. Each mutagenesis reaction was subsequently run for 18 cycles (95 °C 20 s, 60 °C 10 s, 68 °C 5 min). Following amplification, the reaction mixtures were digested at 37 °C for 10 min following the addition of 2 µL of DpnI restriction enzyme. Finally, 5 µL of the DpnI digest reaction mixture was transformed into XL10-Gold ultracompetent cells, and mutations were verified by Sanger sequencing (Genewiz). Primer sequences can be found in Table 3.3.

3.7.2 Protein expression conditions for SxtT, GxtA, and variants

To produce wild-type and variant protein samples for use in X-ray crystallography and activity assays, we capitalized on the expression and purification methods that were used in our previous work[7]. In short, single colonies of C41(DE3) Escherichia coli cells containing the sxtT or gxtA pMCSG9 plasmids were used to inoculate 5 mL of LB containing 100 µg/mL ampicillin. These starter cultures were incubated overnight at 37 °C and subsequently used to inoculate 1 L of LB media containing 100 µg/mL ampicillin. These larger cultures were grown in similar conditions to the starter culture (37 °C and 200 rpm). At the time when the OD600 reached 0.6–0.8, the temperature was decreased to 20 °C and 0.2 mM IPTG, 0.2 mg/mL ferric ammonium citrate, and 0.4 mg/mL ferrous sulfate heptahydrate were added. Each of these large-scale cultures was incubated for an additional 18 h prior to harvesting. The average pellet for a 1 L wild-type SxtT culture was 5 g of wet cell mass. For the SxtT tunnel chimera, SxtT loop chimera, and SxtT M255Y/T276V/R204K/S230T, M255Y/T276V/R204K/S232A, and I237T variants, the average pellet mass was 2–2.5 g for 1 L of cell culture. The wet cell mass for all other SxtT variants except for the 5 variants mentioned above was 3.5–4.5 g for 1 L cell culture.

3.7.3 Purification protocol for SxtT and GxtA for reactions and crystallography

To obtain homogeneous SxtT and GxtA protein for X-ray crystallography and activity assay experiments, a similar protocol to that previously described was followed[7]. Briefly, the cell pellet produced from a 2 L culture that contained the over-expressed MBP-tagged SxtT and GxtA was resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% glycerol). Cells were lysed by sonication using a program that proceeds through repetitive cycles of 4 s sonication and 15 s rest until reaching a total on-time of 5 min. Lysed cells were clarified in an Eppendorf centrifuge 5810R at 12,000 × g for 40 min. The supernatant, which contained the MBP-tagged protein SxtT or MBP-tagged GxtA was loaded onto a 5 mL MBP-Trap column (Cytiva). This column was washed with 10 column volumes of Buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% glycerol) and eluted with a five-column volume gradient of Buffer B (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% glycerol, 10 mM maltose). Fractions containing the MBP-SxtT or MBP-GxtA were pooled and dialyzed into Buffer C (20 mM Tris-HCl pH 7.4, 50 mM NaCl, 1 mM DTT, 10% glycerol). Following buffer exchange, the MBP-tagged SxtT and GxtA proteins were cleaved using 5 mg of TEV protease. This tag-free protein was subsequently loaded onto a 5 mL His-Trap (Cytiva) for further purification and removal of both the TEV protease and the cleaved His-MBP tag. The flow-through from the column, which contained the desired tag-free proteins was concentrated and loaded onto a HiPrep 16/60 Sephacryl S200-HR (Cytiva) gel filtration column that was pre-equilibrated in Buffer D (50 mM HEPES, pH 8.0, 200 mM NaCl, 10% glycerol). The trimeric protein fractions were pooled and concentrated to 10 mg/mL. This concentrated protein was aliquoted into 50 µL fractions and flash-frozen using liquid nitrogen for storage in the −80 °C freezer. For crystallography, the trimeric protein fractions were pooled, and buffer exchanged into Buffer E (50 mM HEPES, pH = 8.0, 10% glycerol) using PD-10 desalting columns (BioRad). The cell pellet from a 1 L culture of wild-type SxtT typically results in a yield of ~5 mg of homogenous protein. For the SxtT tunnel chimera, SxtT loop chimera, and SxtT M255Y/T276V/R204K/S230T, M255Y/T276V/R204K/S232A, and I237T variants, the average protein yield was 1–1.5 mg from 1 L of cell pellet. In contrast, all other variants yielded 2–3 mg protein per 1 L of cell pellet.

3.7.4 Protein expression conditions for the reductase protein VanB

The methods for purifying VanB were previously described and were followed here[7]. First, a single colony of BL21(DE3) *E. coli* that contained the pMCSG7-vanB plasmid was used to inoculate a 5 mL starter culture of LB containing 50 µg/mL ampicillin. Second, following overnight incubation at 37 °C, the starter culture was used to inoculate 1 L of LB containing 100 µg/mL ampicillin. These larger cultures were incubated at 37 °C until the OD600 reached 0.6–0.8. At this point, flasks were cooled at 20 °C for 1 h prior to the addition of 0.2 mM IPTG. Following an 18 h period of incubation at 20 °C, the cells were harvested, and the protein was purified as described below.

3.7.5 Purification protocol for the reductase protein VanB

The cell pellet from a 2 L culture of VanB was resuspended in 50 mL of VanB Buffer A (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole, 10% glycerol). The mixtures were lysed by sonication using the program described above for SxtT and GxtA. Lysed cells were centrifuged in an Eppendorf centrifuge 5810 R at 12,000 × g for 40 min. The supernatant was loaded onto a Bio-Rad FPLC system fitted with a 5 mL HisTrap HP column (Cytiva). The column was equilibrated using where VanB Buffer A (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole, 10% glycerol) and the protein was eluted using VanB Buffer B (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 200 mM imidazole, 10% glycerol). The VanB-containing fractions were pooled and exchanged into storage VanB Buffer C (50 mM HEPES pH 8.0, 100 mM NaCl, 10% glycerol) using a PD-10 desalting column (BioRad). Finally, VanB was concentrated to approximately 200 μM and flash-frozen by liquid nitrogen for long-term storage at −80 °C.

3.7.6 Circular Dichroism (CD) experiments

This experiment was carried out by Jiayi Tian. Stock solutions of wild-type SxtT and SxtT variants stored at concentrations of 250 μM were diluted to 5 μM using a buffer containing 50 mM HEPES pH=8.0 and 150 mM NaCl. For each CD measurement, 350 μL of the diluted wild-type SxtT or SxtT variant solution was transferred into a 10 mm quartz cuvette (Hellma). The CD spectra were recorded using a Jasco J-1500 CD spectrometer with 0.1 nm data pitch and 20 nm/min scan speeds. The baseline was measured with the same buffer used to do dilution (50 mM HEPES pH=8.0 and 150 mM NaCl). Each sample spectrum is an average of 5 cumulative spectra.

3.7.7 Crystallization of substrate bound SxtT and GxtA

Conditions for crystallizing wild-type and substrate-analog bound SxtT and GxtA were previously identified by our laboratory[7]. To produce crystals of SxtT with its native substrate β-STOH bound, 1 µL of 10 mg/mL SxtT (50 mM HEPES, pH 8.0, 10% v/v glycerol) was combined with 1 µL of well solution of (2.0 M (NH4)2SO4, 0.1 M Bis–Tris pH 6.5, 10% v/v glycerol and 20 mM of β-STOH). In a hanging drop setup, brown crystals appeared within two days of being stored in a 20 °C incubator housed within a Coy chamber (Coy Lab Products) that is maintained using a 95% nitrogen and 5% hydrogen mixture. Likewise, crystals of β-STOH-bound GxtA and STX-bound GxtA were grown by mixing 1 µL of 15 mg/mL GxtA (50 mM HEPES, pH 8.0, 10% v/v glycerol) with 1 µL of well solution (0.3 M MgCl2, 0.1 M Bis–Tris pH 5.5, 25% v/v PEG3350, 15% v/v glycerol, and either 20 mM β-STOH or 20 mM STX). Brown crystals appeared within one day following anaerobic incubation at 20 °C. All crystals were harvested and cryocooled in liquid nitrogen without the addition of cryoprotectant in the Coy chamber.

3.7.8 Xe pressurization experiment with GxtA

The conditions for crystallizing wild-type and substrate-analog bound SxtT and GxtA were previously identified by our laboratory and are described above[7]. However, for this Xe-pressurized structure, we deviated from the conditions described above. To grow crystals for Xe-pressurization experiments, the reservoir solutions did not contain STX or β-STOH and the sitting drop plates used were MiTeGen In-Situ-1 Crystallization Plates. Once brown crystals were identifiable for SxtT and GxtA, the MiTeGen In-Situ-1 Crystallization Plates were shipped to Stanford Synchrotron Radiation Lightsource (SSRL) for Xe incorporation and data collection by Dr. Tzanko I Doukov. Once at SSRL, the crystallization drops were harvested under Paratone N oil using Mitegen loops. Any water-based liquid around any crystals and any extra oil were carefully removed during the harvesting process. These minimal oil-covered crystals were placed inside the SSRL pressurizing apparatus49. The GxtA and SxtT crystals were pressurized to 400 psi of Xe gas for 3 min. A subsequent depressurizing period of 15 s was used to return to atmospheric pressure. This step was followed by plunging the pin that contained the looped crystal into a liquid nitrogen container. The frozen Xe-derivatized crystal was transferred and diffraction was measured on an EIGER 16 M detector at SSRL beamline BL14-1. The beam size was adjusted to 100 by 80 microns to match the crystal shape. A total of 4320 images with 0.5 degrees oscillation were collected with inverse beam protocol (every 30 degrees of data were collected close in time to the 180 + 30 degrees of data) to maximize the anomalous signal. For GxtA, three datasets were collected at 6000 eV (2.06633 Å), where the Xe anomalous signal is the strongest. This method resulted in a crystal structure of GxtA that has one atom of Xe in each protomer. These Xe atoms are modeled at an occupancy of 0.8 and their correct placement was confirmed by the calculation of anomalous maps. In contrast to GxtA, the Xe-pressurized SxtT crystals lost diffraction with increased Xe pressures and were not amenable to this experiment.

3.7.9 Data processing and structure solution

The datasets for β-STOH-bound SxtT, β-STOH-bound GxtA, and STX-bound GxtA were collected at SSRL beamline 9-2 (detector: Dectris Pilatus 6 M). The Xe-GxtA dataset was collected at SSRL beamline 14-1 (described above). For each of the former datasets, the temperature of data collection was 100 K and the wavelength used for data collection was 0.97946 Å (SxtT + β-STOH, GxtA + β-STOH, and GxtA + STX). All datasets were indexed, integrated, and scaled in XDS[50,51]. Similar to that observed for the previously determined structures[7] of SxtT and GxtA, SxtT indexed as *C222* and GxtA indexed as *P12₁1*. Edited versions of these previously determined structures, which lacked water molecules, metal centers, and small molecules, were used to determine the new β-STOH-SxtT, β-STOH-GxtA, STX-GxtA, and Xe-pressurized structures. Once solved, each model was subjected to five cycles of simulated annealing in Phenix[52]. Subsequent rounds of manual building and adjusting in COOT[53] coupled with cycles of positional and individual B-factor refinement in Phenix[52] were used until each model was complete. In general, for each of the structures, protein sidechains were adjusted first, metal centers and water molecules were added second, and ligands were added in the final refinement stages. For the structures of β-STOH-SxtT, β-STOH-GxtA, and STX-GxtA, small molecule parameter files previously[7] generated using the Grade Web Server Global Phasing (Cambridge, UK) for STX analogs, were also included in refinement. Whereas, for the β-STOH bound GxtA structure, the occupancy of the ligand is set at 1 in each chain of the structure, for β-STOH bound-SxtT and STX-bound GxtA, we see lower occupancy in some of the chains (β-STOH bound-SxtT occupancy = 0.8, 0.6, 0.8 and STX-bound GxtA occupancy = 1, 0.8, 1 for chains A, B, and C, respectively). As the disordered loop in chain A of SxtT with STX bound packs against the same disordered loop of a symmetry mate leaving substantial positive difference electron density at the interface, but no clear way of building the loops without intersection, we re-evaluated the space groups using Pointless54 and XTriage in Phenix[52]. We also evaluated other space groups for SxtT, including *C121*, *C222₁*, *P12*, and *P12₁*. However, none of these space groups, other than *C121* allowed us to build a model of SxtT. The resultant lower symmetry model, similar to that observed in the *C222* space group also placed the disordered loops from two adjacent protomers close together, confirming our original choice of *C222*. The R_{free} test sets composed of five percent of the original datasets[7] were used to evaluate the progression of refinement.

After the final refinement stages were completed, the structures were inspected using simulated annealing composite omit electron density maps calculated in Phenix[52] and the MolProbity program[55]. Ramachandran statistics and omitted residues for each structure determined in this work can be found in Table 3.2. The data collection and refinement statistics for each of the structures are summarized in Table 3.1. Each of the structure figures presented in this work was made using PyMol. The hydropathy plots of SxtT and GxtA were produced in PyMol using the Eisenberg hydrophobicity scale[56]. The crystallography software packages used to determine these structures were compiled by SBGrid[57].

3.7.10 Tunnel calculations for proteins studied in this work

The substrate-channeling tunnels in SxtT and GxtA were analyzed as previously described[7] using the MOLEonline server[43,44]. For tunnel calculation, each of the Rieske oxygenases was loaded into Pymol and exported as monomeric files that contained only the carbon-chain and non-heme iron sites. Additional cofactors if present were removed. Each monomeric subunit was then loaded into the MOLEonline server[43,44]. This analysis was performed for SxtT, GxtA, NdmA (6ICK)[8], NdmB (6ICL)[8], Stc2 (3VCP)[12], KshA (4QCK)[10], CntA (6Y9D)[11], OMO (1Z03)[13], DdmC (3GKE)[9], and CARDO (2DE7)[14,45].

3.7.11 SxtT and GxtA wild type and variants reactions

Enzymatic reactions were prepared and undertaken following the literature procedure[7]. Wild-type SxtT, SxtT variants, and GxtA were purified as described above and stored in final concentrations of 10 mg/mL. For each enzymatic reaction, a fresh protein aliquot was removed from the freezer. The reactions consisted of either 50 μM wild-type SxtT, SxtT variant, or GxtA. In addition, these reactions contained 36 μM VanB, 200 μM substrate, 500 μM NADH, 100 μM $Fe(NH_4)_2(SO_4)_2$, and 50 mM Tris-HCl pH 7.0. Once combined, the reactions were mixed and incubated at 30 °C for 2 h. Reactions were then quenched by the addition of 150 μL acetonitrile and centrifuged. Following centrifugation, 100 μL of the supernatant was diluted with 100 μL of acetonitrile solution that also contained 1% of our mass spectrometry internal standard 15N-arginine. The activity of the wild-type SxtT and GxtA proteins, as well as the SxtT variant proteins, was determined by analyzing the amount of substrate consumed as previously described[42]. In particular, the decrease of the substrate peak relative to the control was converted into a percentage of substrate conversion. All reactions presented in this work were run in duplicate.

3.7.12 Ethanol incorporation into reaction products

To evaluate whether hydroxylation occurred at the C11 or C12 position, ethanol incorporation into the reaction product was probed. First, reactions were conducted as described above. However, following the quenching step and the subsequent addition of the internal standard mixture, 50 μL of 100% ethanol was added to a 50 μL aliquot of each reaction mixture. The resulting mixture was incubated at room temperature for 12 h before analysis by LCMS. To determine the extent of ethanol incorporation into the product, we used a previously documented quantification method, which involves evaluating the amount of incorporated-STX and STX products using standard curves[7]. Again, each reaction presented was run in duplicate. For all assays, the hydroxylation percentage was calculated using Excel and the figures were then made using GraphPad Prism.

3.7.13 LC-MS and MS/MS analysis

In general, methods described in our previous work, which used an Agilent G6545A quadrupole-time of flight (TOF) or an Agilent 6230 TOF mass spectrometer equipped with a dual AJS ESI source and an Agilent 1290 Infinity series diode array detector, autosampler, and binary pump, were followed here7. Solvent A = water with 0.1% formic acid. Solvent B = 95% acetonitrile, 5% water and 0.1% formic acid. An Acquity UPLE BEH Amide 1.7 µm, 2.1 × 100 mm hydrophobic interaction liquid chromatography column was used for all separations (Waters). The chromatographic method was typically 18% A 0–5 min at 0.4 mL/min or 15% A 0–7 min at 0.3 mL/min. 1.0 µL injections were made for each sample.

3.8 References

1 Perry, C., de Los Santos, E. L. C., Alkhalaf, L. M. & Challis, G. L. Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat. Prod. Rep*. 35, 622-632 (**2018**).

2 Kovaleva, E. G. & Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nat. Chem. Biol*. 4, 186-193, doi:10.1038/nchembio.71 (**2008**).

3 Barry, S. M. & Challis, G. L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal.* 3, 2362-2370 (**2013**).

4 Ferraro, D. J., Gakhar, L. & Ramaswamy, S. Rieske business: structure-function of Rieske non-heme oxygenases. Biochem. Biophys. *Res. Commun.* 338, 175-190, (**2005**).

5 Knapp, M., Mendoza, J. & Bridwell-Rabb, J. *In Encyclopedia of Biological Chemistry*, 3rd Edition (**2021**).

6 Li, B. & Bridwell-Rabb, J. Aerobic Enzymes and Their Radical SAM Enzyme Counterparts in Tetrapyrrole Pathways. *Biochemistry* 58, 85-93, (**2019**).

7 Lukowski, A. L., Liu, J., Bridwell-Rabb, J. & Narayan, A. R. H. Structural basis for divergent C-H hydroxylation selectivity in two Rieske oxygenases. *Nat. Commun*. 11, 2991, (**2020**).

8 Kim, J. H. et al. Structural and Mechanistic Insights into Caffeine Degradation by the Bacterial N-Demethylase Complex. *J. Mol. Biol*. 431, 3647-3661, (**2019**).

9 Dumitru, R., Jiang, W. Z., Weeks, D. P. & Wilson, M. A. Crystal structure of dicamba monooxygenase: a Rieske nonheme oxygenase that catalyzes oxidative demethylation. *J. Mol. Biol.* 392, 498-510 (**2009**).

10 Capyk, J. K., D'Angelo, I., Strynadka, N. C. & Eltis, L. D. Characterization of 3-ketosteroid 9a-hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of Mycobacterium tuberculosis. *J. Biol. Chem*. 284, 9937-9946 (**2009**).

11 Quareshy, M. et al. Structural basis of carnitine monooxygenase CntA substrate specificity, inhibition and inter-subunit electron transfer. *J. Biol. Chem.*, (**2020**).

12 Daughtry, K. D. et al. Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J. Am. Chem. Soc*. 134, 2823-2834 (**2012**).

13 Martins, B. M., Svetlitchnaia, T. & Dobbek, H. 2-Oxoquinoline 8-monooxygenase oxygenase component: active site modulation by Rieske-[2Fe-2S] center oxidation/reduction. *Structure* 13, 817-824 (**2005**).

14 Nojiri, H. et al. Structure of the terminal oxygenase component of angular dioxygenase, carbazole 1,9a-dioxygenase. *J. Mol. Biol*. 351, 355-370 (**2005**).

15 Friemann, R. et al. Structural insight into the dioxygenation of nitroarene compounds: the crystal structure of nitrobenzene dioxygenase. *J. Mol. Biol.* 348, 1139-1151, (**2005**).

16 Friemann, R. et al. Structures of the multicomponent Rieske non-heme iron toluene 2,3-dioxygenase enzyme system. *Acta Crystallogr. D. Biol. Crystallogr.* 65, 24-33, (**2009**).

17 Kumari, A., Singh, D., Ramaswamy, S. & Ramanathan, G. Structural and functional studies of ferredoxin and oxygenase components of 3-nitrotoluene dioxygenase from Diaphorobacter sp. strain DS2. *PLoS One* 12, e0176398 (**2017**).

18 Furusawa, Y. et al. Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from Rhodococcus sp. strain RHA1. *J. Mol. Biol.* 342, 1041-1052 (**2004**).

19 Karlsson, A. et al. Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science* 299, 1039-1042 (**2003**).

20 Dong, X. et al. Crystal structure of the terminal oxygenase component of cumene dioxygenase from Pseudomonas fluorescens IP01. *J. Bacteriol.* 187, 2483-2490 (**2005**).

21 Jakoncic, J., Jouanneau, Y., Meyer, C. & Stojanoff, V. The crystal structure of the ring-hydroxylating dioxygenase from Sphingomonas CHY-1. *FEBS J.* 274, 2470-2481 (**2007**).

22 Kauppi, B. et al. Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Structure* 6, 571-586 (**1998**).

23 Hou, Y. J., Guo, Y., Li, D. F. & Zhou, N. Y. Structural and Biochemical Analysis Reveals a Distinct Catalytic Site of Salicylate 5-Monooxygenase NagGH from Rieske Dioxygenases. *Appl. Environ. Microbiol.* 87 (**2021**).

24 Escalante, D. E., Aukema, K. G., Wackett, L. P. & Aksan, A. Simulation of the Bottleneck Controlling Access into a Rieske Active Site: Predicting Substrates of Naphthalene 1,2-Dioxygenase. *J. Chem. Inf. Model* 57, 550-561 (**2017**).

25 Pravda, L. et al. Anatomy of enzyme channels. *BMC Bioinformatics* 15, 379, (**2014**).

26 Urban, P., Lautier, T., Pompon, D. & Truan, G. Ligand Access Channels in Cytochrome P450 Enzymes: A Review. *Int. J. Mol. Sci.* 19 (**2018**).

27 Yu, X., Cojocaru, V. & Wade, R. C. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnol. Appl. Biochem.* 60, 134-145 (**2013**).

28 Bridwell-Rabb, J., Grell, T. A. J. & Drennan, C. L. A Rich Man, Poor Man Story of S-Adenosylmethionine and Cobalamin Revisited. *Annu. Rev. Biochem.* 87, 555-584, (**2018**).

29 Dowling, D. P., Croft, A. K. & Drennan, C. L. Radical use of Rossmann and TIM barrel architectures for controlling coenzyme B12 chemistry. *Annu. Rev. Biophys.* 41, 403-427, (**2012**).

30 Kress, N., Halder, J. M., Rapp, L. R. & Hauer, B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* 47, 109-116, (**2018**).

31 Banerjee, R. & Lipscomb, J. D. Small-Molecule Tunnels in Metalloenzymes Viewed as Extensions of the Active Site. *Acc. Chem. Res*. 54, 2185-2195, (**2021**).

32 Yang, T. C. et al. Modulation of substrate binding to naphthalene 1,2-dioxygenase by rieske cluster reduction/oxidation. *J. Am. Chem. Soc.* 125, 2034-2035, (**2003**).

33 Yang, T. C. et al. Substrate binding to NO-ferro-naphthalene 1,2-dioxygenase studied by high-resolution Q-band pulsed 2H-ENDOR spectroscopy. *J. Am. Chem. Soc.* 125, 7056-7066, (**2003**).

34 Wolfe, M. D., Parales, J. V., Gibson, D. T. & Lipscomb, J. D. Single turnover chemistry and regulation of O2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J. Biol. Chem.* 276, 1945-1953, (**2001**).

35 Ohta, T., Chakrabarty, S., Lipscomb, J. D. & Solomon, E. I. Near-IR MCD of the nonheme ferrous active site in naphthalene 1,2-dioxygenase: correlation to crystallography and structural insight into the mechanism of Rieske dioxygenases. *J. Am. Chem. Soc.* 130, (**2008**).

36 Kokkonen, P., Bednar, D., Pinto, G., Prokop, Z. & Damborsky, J. Engineering enzyme access tunnels. *Biotechnol. Adv*. 37, 107386, (**2019**).

37 Heinemann, P. M., Armbruster, D. & Hauer, B. Active-site loop variations adjust activity and selectivity of the cumene dioxygenase. *Nat. Commun.* 12, 1095, (**2021**).

38 Petrusma, M., Dijkhuizen, L. & van der Geize, R. Structural features in the KshA terminal oxygenase protein that determine substrate preference of 3-ketosteroid 9alpha-hydroxylase enzymes. *J. Bacteriol.* 194, 115-121 (**2012**).

39 Ferraro, D. J. et al. Structural investigations of the ferredoxin and terminal oxygenase components of the biphenyl 2,3-dioxygenase from Sphingobium yanoikuyae B1. *BMC Struct. Biol.* 7, 10, (**2007**).

40 Nestl, B. M. & Hauer, B. Engineering of Flexible Loops in Enzymes. *ACS Catal.* 4, 3201-3211, (**2014**).

41 Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Ges. Dtsch. Chem* 27, 2985 (**1984**).

42 Lukowski, A. L. et al. C-H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J. Am. Chem. Soc.* 140, 11863-11869, (**2018**).

43 Pravda, L. et al. MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.* 46, W368-W373 (**2018**).

44 Berka, K. et al. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.* 40, W222-227 (**2012**).

45 Ashikawa, Y. et al. Electron transfer complex formation between oxygenase and ferredoxin components in Rieske nonheme iron oxygenase system. *Structure* 14, 1779-1789 (**2006**).

46 Ashikawa, Y. et al. Structural insight into the substrate- and dioxygen-binding manner in the catalytic cycle of rieske nonheme iron oxygenase system, carbazole 1,9a-dioxygenase. *BMC Struct. Biol.* 12, 15 (**2012**).

47 Inoue, K. et al. Structural basis of the divergent oxygenation reactions catalyzed by the rieske nonheme iron oxygenase carbazole 1,9a-dioxygenase. *Appl. Environ. Microbiol.* 80, 2821-2832 (**2014**).

48 Ashikawa, Y. et al. Crystallization and preliminary X-ray diffraction studies of the ferredoxin reductase component in the Rieske nonhaem iron oxygenase system carbazole 1,9a-dioxygenase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 499-502 (**2007**).

49 Soltis, S. M., Stowell, M. H. B., Wiener, M. C., Phillips, G. N. & Rees, D. C. Successful flash-cooling of xenon-derivatized myoglobin crystals. *J Appl Crystallogr* 30, 190-194, (**1997**).

50 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66, 133-144 (2010).

51 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* 66, 125-132 (**2010**).

52 Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221 (**2010**).

53 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132 (**2004**).

54 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D* 62, 72-82 (**2006**).

55 Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* 66, 12-21 (**2010**).

56 Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 179, 125-142, (**1984**).

57 Morin, A. et al. Collaboration gets the most out of software. Elife 2, e01456, (**2013**).

Chapter 4 Rieske Oxygenase Catalyzed C-H Bond Functionalization Reaction in Chlorophyll *b* Biosynthesis

Reproduced with permission from Liu, J.; Knapp, M.; Jo, M.; Dill, Z; Bridwell-Rabb, J. Rieske oxygenase catalyzed C-H bond functionalization reaction in Chlorophyll *b* biosynthesis. *Submitted*

Author Contributions: J.L., M.K., M.J., and J.B.R. contributed to the design of the experiments and wrote the manuscript. J.L. and M.K. performed all biochemical assays, obtained all LC-MS data, and performed the synthetic experiments to produce the needed substrates and products. J.L., M.K., and Z.D. purified all proteins used in this work. M.J. designed and optimized the chlorophyllase construct, as well as the expression, purification, and reaction conditions.

4.1 Abstract

Rieske oxygenases perform precise C–H bond functionalization reactions at a variety of traditionally inert substrate centers. These functionalization reactions are integral to both anabolic and catabolic pathways and generally involve the insertion of one (monooxygenation) or two (dioxygenation) oxygen atoms into a substrate. Other more divergent reactions are also catalyzed by Rieske oxygenases. For example, Chlorophyll(ide) *a* oxygenase (CAO) is proposed to catalyze sequential monooxygenation reactions to transform a methyl group into a formyl group that is characteristic of Chlorophyll *b*. This formyl group, similar to the formyl groups found in Chlorophyll pigments *d* and *f*, tunes the absorption spectra of Chlorophyll *b* and plays a key role in the ability of several photosynthetic organisms to adapt to available environmental light. Despite the importance of this reaction and the overall lack of information regarding how photosynthetic organisms append these important modifications to the Chlorophyll scaffold, CAO has never been studied *in vitro* with purified protein, leaving many open questions regarding whether CAO can facilitate both oxygenation reactions using just the Rieske oxygenase machinery. In this study, we

overproduced and reconstituted the activity of CAO from four different organisms. We demonstrated that CAO, in partner with a non-native Rieske reductase, converts a Chlorophyll *a* precursor, Chlorophyllide *a*, into Chlorophyllide *b in vitro*. Further, analysis of this reaction confirmed the existence of the proposed key intermediate, highlighted the stereospecificity of the reaction, and revealed the potential of CAO as an enzymatic tool for synthesizing custom-tuned natural and unnatural Chlorophyll pigments. This work thus adds to our fundamental understanding of Chlorophyll biosynthesis and extends the catalytic repertoire of the Rieske oxygenase superfamily.

4.2 Introduction

Chlorophylls (Chls) are a class of naturally occurring pigments that play key roles in photosynthetic organisms. These pigments capture solar energy and facilitate its transformation into chemical energy[1-4]. The diversity of this class of molecules, in some cases, arises from C–H bond functionalization reactions that decorate the macrocyclic scaffold and adjust its spectroscopic properties[1-4]. One such scaffold modification that is found in Chl derivatives *b*, *d*, and *f*, is a formyl group[5] (Figure 4.1.a). These modifications have received attention for their ability to customize and extend the range of usable light in photosynthetic organisms[5]. For example, Chl *d* and Chl *f*, allow photosynthetic microorganisms to absorb red and far-red light, which is typically found in soils, mats, and other shaded environments[6]. Chl *b*, on the other hand, is an integral accessory pigment that has enhanced absorption features for blue and green light[3]. For Chl f, formylation is catalyzed by a membrane-bound photooxidoreductase, Chl *f* synthase, while the enzyme involved in the biosynthesis of Chl *d* has yet to be identified (Figure 4.1.a)[7-9]. Lastly, the C7-formyl group of Chl *b* is installed by Chlorophyll(ide) *a* oxygenase (CAO), an annotated member of the large Rieske non-heme iron oxygenase, or Rieske oxygenase superfamily of enzymes (Figure 4.1.a)[10,11].

Rieske oxygenases are classified by the presence of motifs for binding an N-terminal Rieske [2Fe-2S] cluster and a C-terminal mononuclear non-heme iron center. The Rieske cluster accepts external electrons from a dedicated reductase protein and shuttles them to the mononuclear iron site for activation of molecular oxygen (O_2)[12-16]. The formed oxidant is then used to facilitate subsequent chemistry. In most cases, characterized Rieske oxygenases have been shown to use their metallocenters to facilitate single monooxygenation or dioxygenation events[12-16]. Using the Rieske machinery, CAO is proposed to catalyze sequential monooxygenation reactions to form

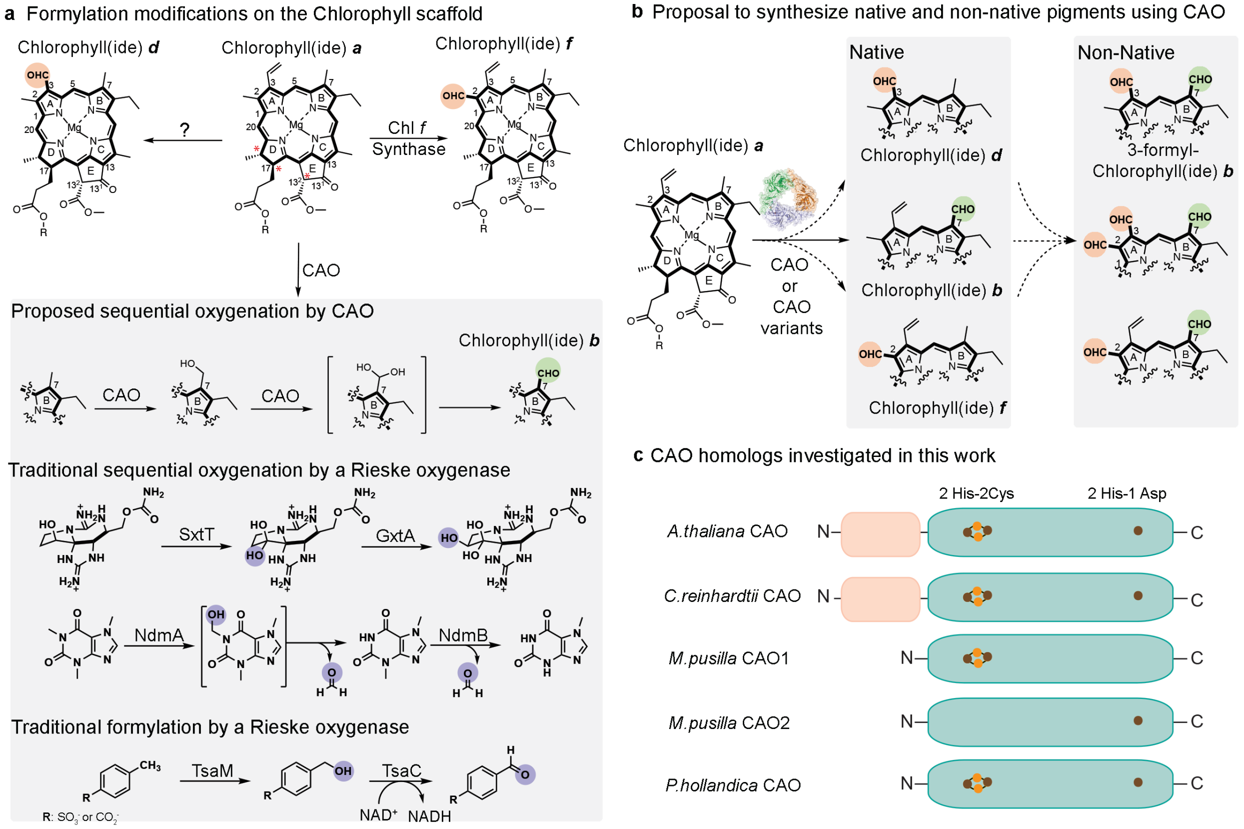


Figure 4.1. Formyl groups are abundant modifications made on the Chlorophyll (Chl) scaffold. (a) The pigments Chl *b*, *d*, and *f*, each bear formyl groups on their macrocyclic chlorin scaffolds at various positions. These modifications contribute to their characteristic absorbance patterns. The modifications found in Chl *d* and *f* are formed using an as-yet unidentified enzyme and the photooxidoreductase Chl f synthase[7-9], respectively. Formation of the C7 formyl group in Chlorophyll(ide) *b* is instead proposed to be installed via two sequential Chlorophyll(ide) *a* oxygenase (CAO)-catalyzed reactions that transform the C7-methyl group of Chlorophyll(ide) *a* into the formyl group of Chlorophyll(ide) *b* through C7-hydroxymethyl and C7-dihydroxymethyl intermediates[10]. All three chiral centers in Chl *a*, *b* and intermediates are labeled with a red asterisk. This proposal for CAO is unlike that for other transformations that proceed through more than one monooxygenation reaction and require two Rieske oxygenases to be completed[9-11]. This proposal is also unlike that needed to convert a methyl group into a formyl group in the catabolism of 4-toluene sulfonate, which requires both a Rieske oxygenase and dehydrogenase[13]. (b), CAO has the propensity to be used as a tool for formylating the Chl pigment scaffold to produce custom-tuned native and non-native pigments. (c), The CAO homologs studied in this work have different domain architectures. All homologs of CAO are predicted to have a Rieske [2Fe-2S] cluster and a mononuclear non-heme iron site in their catalytic domains (blue). These metallocenters are coordinated by two His and two Cys ligands and a facial triad of residues, respectively. The *M. pusilla* CAO homolog is found in two polypeptide chains and the *A. thaliana* and *C. reinhardtii*

homologs have N-terminal regulatory domains (peach)[25]. As *Ph*CAO appears to contain the simplest architecture of all four homologs, it was the first CAO homolog characterized in this work.

both a hydroxymethyl and di-hydroxymethyl intermediate[10]. The latter species is then proposed to spontaneously lose water to become the formyl group found at the C7-position of Chl *b*[10] (Figure 4.1.a). In support of this proposal, early *in vivo* biochemical characterization of CAO in *Chorella vulgaris* determined that the formyl group oxygen atom originates from O_2[17]. Other seminal studies on the CAO homolog from *Arabidopsis thaliana* (*At*CAO) demonstrated that lysed cells, from which *At*CAO was recombinantly expressed, could facilitate the conversion of a small amount of the Chl precursor, Chlorophyllide (Chlide) *a*, into Chlide *b* when combined with a reducing system[10]. This result, along with data that shows a Chl *b* deficiency can be traced to a single gene in multiple photosynthetic organisms suggests that CAO, in the absence of a dedicated partner protein, is sufficient to synthesize the formyl group[10, 18-20].

Therefore, CAO represents the seemingly simplest enzyme to formylate a pigment and has the potential to serve as a tool for catalyzing late-stage formylation (or use oxidation) reactions at different positions on the pigment scaffold (Figure 4.1.a). Remarkably, however, whether the substrate of this enzyme is Chl *a* or Chlide *a* is still in debate, no purification and *in vitro* reconstitution of a CAO homolog has been demonstrated, and the activity of CAO has never been demonstrated in isolation[4]. As this sequential oxygenation reaction represents a divergence from traditional Rieske oxygenase chemistry, many questions remain regarding whether the proposed sequential oxygenation reaction can occur in the absence of an additional component. For example, in other biosynthetic pathways where sequential oxygenation reactions are required, multiple proteins work together (Figure 4.1.a). The two required oxygenations on the saxitoxin scaffold are made by the Rieske oxygenases SxtT and GxtA (Figure 4.1.a)[21-22]. Likewise, two Rieske oxygenases, NdmA and NdmB, catalyze sequential oxidative demethylation reactions in caffeine degradation (Figure 4.1.a)[23]. More interestingly, however, is the Rieske oxygenase 4-toluenesulfonate methyl monooxygenase (TsaM), which also catalyzes a methyl- to formyl-group transformation[24]. This enzyme is involved in converting 4-tolunesulfonate into 4-sulfobenzoate, but interestingly requires the help of a nicotinamide adenine dinucleotide (NAD^+)-dependent dehydrogenase to transform the monohydroxy intermediate into the product (Figure 4.1.a)[24].

Thus, in this work, we designed methods to study four homologs of CAO that showcase different predicted architectures[25] *in vitro* with purified protein (Figure 4.1.c). We sought to determine whether these homologs could facilitate a sequential oxygenation reaction in the absence of an additional Rieske oxygenase, dehydrogenase, or cofactor. Through these studies, a non-native Rieske reductase was identified that could work in partnership with each CAO homolog to convert a Chl *a* precursor, Chlide *a*, into Chlide *b in vitro*. Analysis of this reaction using mass spectrometry (MS) confirmed the identity of the proposed monooxygenated intermediate and showed that it too can be synthesized and accepted by CAO as a substrate. Finally, this work revealed intriguing details regarding the stereoselectivity of this reaction, the substrate scope of CAO, and the potential for using CAO as an enzymatic tool for synthesizing custom Chl pigments, those of which are utilized broadly in the cosmetic, food, agricultural, and pharmaceutical industries. This enzymatic reaction also serves as a convenient and green method to synthesize late-stage oxidative chlorophyll species that otherwise would be synthetically complicated to obtain.

4.3 Results

4.3.1 Purification of four different CAO homologs

To address the ability of CAO to form chlorophyll(ide) *b in vitro*, genes encoding the CAO homologs from the prokaryotic organism *Prochlorothrix hollandica* (*Ph*CAO), the model plant *Arabidopsis thaliana* (*At*CAO), and two green algae *Chlamydomonas reinhardtii* (*Cr*CAO) and *Micromonas pusilla* (*Mp*CAO), were synthesized and codon-optimized for expression in *E. coli.* Due to the simpler annotated domain structure[25], the CAO homolog from *P. hollandica* (*Ph*CAO) was investigated first (Figure 4.1.c). Here, it was determined that a construct of *Ph*CAO that contained an N-terminal His-maltose binding protein (MBP) tag could be overexpressed in *E. coli* C41(DE3) cells and purified by exploiting the affinity of MBP for amylose. Tag cleavage using Tobacco Etch Virus protease and tag removal using a second nickel affinity chromatography step resulted in isolation of a pure sample of *Ph*CAO. Subsequent polishing via gel filtration

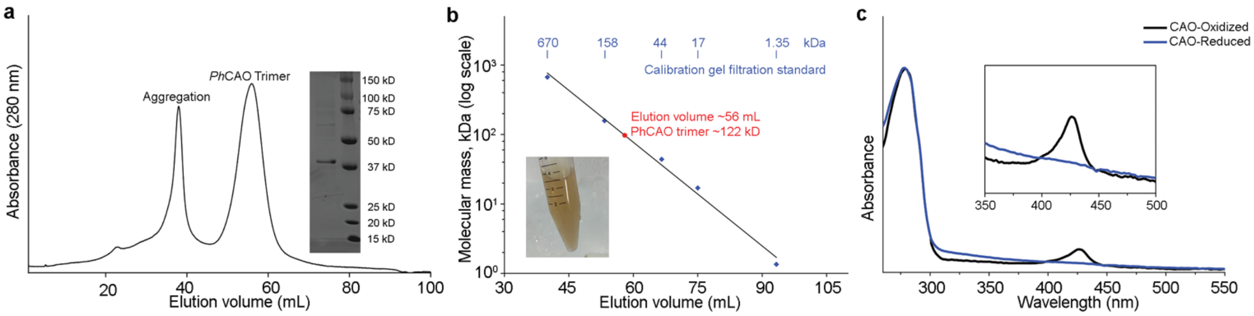
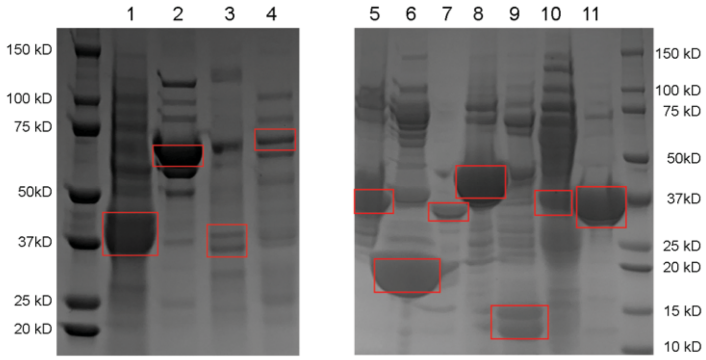


Figure 4.2. *P. hollandica* CAO can be purified and used for *in vitro* biochemical studies.
(a), A gel filtration profile of PhCAO reveals that CAO purifies with the expected trimeric quaternary structure of a Rieske oxygenase. An inset of an SDS-PAGE gel reveals that the isolated trimeric PhCAO is pure (expected monomeric molecular weight of 41 kDa). (b), The molecular weight of native PhCAO from the gel filtration chromatography step was estimated using a calibration curve (logarithm of molecular mass versus elution volume) generated from a Bio-Rad gel filtration standard (blue squares). (c), The UV-Vis absorption spectra of the oxidized [2Fe-2S] Rieske cluster of as isolated CAO (black trace) was measured as was the absorption spectra in the presence of the chemical reductant sodium dithionite (blue trace).

chromatography revealed that, as expected for a Rieske oxygenase[15-16], *Ph*CAO has a trimeric quaternary architecture (Figure 4.2.a-b and Figure 4.3). In addition, it was determined that following metallocenter reconstitution, *Ph*CAO contained the requisite three iron ions per monomer and displayed the distinctive Fe-S cluster absorption peak at approximately 430 nm (Figure 4.2.c). This peak, which is characteristic of the oxidized [2Fe-2S] cluster, could be quenched by the addition of sodium dithionite (Figure 4.2.c). By capitalizing on the lessons learned in the purification of *Ph*CAO, His-tagged-*At*CAO and *Cr*CAO were subsequently isolated (Figure 4.3). For *Mp*CAO, both polypeptide chains (*Mp*CAO1 and *Mp*CAO2) were co-expressed in a pRSFDuet-1 plasmid and the complex was purified similar to that described for the other homologs (Figure 4.3).



Lane #	Protein	Predicted Size (KDa)
1	*P. hollandica* CAO	40.8
2	*A. thaliana* CAO	60.3
3	*M. pusilla* CAO 1&2	31.7&42.4
4	*C. reinhardtii* CAO	72.1
5	VanB	35.9
6	*E. coli* flavodoxin reductase	19.7
7	*E. coli* flavodoxin	27.7
8	DdmA	44.6
9	DdmB	12.4
10	TsaB	34.2
11	Chlorophyllase	35.7

Figure 4.3. An SDS-PAGE gel of the proteins purified in this study.
Each purified protein band is outlined in a red box. The predicted protein sizes were determined by the ProtParam tool on the Expasy server and are indicated in the table on the right.

4.3.2 CAO converts Chlorophyllide *a* into Chlorophyllide *b*

Once purified, the ability of CAO to convert both Chl *a* and Chlide *a* into Chl *b* and Chlide *b*, respectively, was investigated (Figure 4.4). As these experiments require both substrate and product standards, Chl *a* and Chl *b* were purchased and methods for producing Chlide *a* and Chlide *b* were designed. To obtain these molecules, a protocol for recombinantly expressing and purifying the *Triticum aestivum* homolog of the Chl hydrolase or chlorophyllase (chlase), which catalyzes the first step of Chl catabolism, the conversion of Chl into Chlide, was developed (Figure 4.5 and Figure 4.4.a). The ability of the purified chlorophyllase protein to catalyze production of the needed Chlide *a* and Chlide *b* pigments was subsequently probed using LC-MS (Figure 4.4.a, b). These experiments revealed that an approximate 1-hour incubation of chlorophyllase with Chl *a* or Chl *b* resulted in near complete conversion into the respective Chlide *a* or Chlide *b* pigments (Figure 4.4.b, and Figures 4.6-7). To test the ability of CAO to convert these potential substrates into their respective product molecules, an electron donor for the reaction also needed to be identified. As the partner reductase protein for CAO is unknown, activity was first tested by combining Chl *a* or the enzymatically produced Chlide *a* with *Ph*CAO and the chemical reductants: ascorbate, sodium dithionite, titanium (III) chloride ($TiCl_3$), and dithiothreitol (DTT) in an aerobic environment. Here, it was unexpectedly determined that despite the ability of sodium dithionite to reduce the Rieske cluster, that neither combination of it, nor the other tested reductants with *Ph*CAO and the possible substrates resulted in production of Chl *b* or Chlide *b* (Figure 4.2.c and Figure 4.8.a, b).

Similarly, the ability of CAO to form product, was probed using the so-called peroxide shunt reaction[26]. This reaction uses hydrogen peroxide (H_2O_2) and directly forms an activated oxygen species in the active site[26]. However, even under these conditions, no production of Chlide *b* or Chl *b* was detected (Figure 4.8.c and 4.8.d). These experiments suggested that a needed component for the reaction was missing from the experimental setup. To identify this missing component, assays were performed by combining Chl *a*, *Ph*CAO, and a small amount of *A. thaliana* cell lysate (Figure 4.4.c). This combination did not produce any Chl *b* product above background levels (Fig 4.4.d). Replacement of *Ph*CAO with *At*CAO, *Cr*CAO, or *Mp*CAO, in an

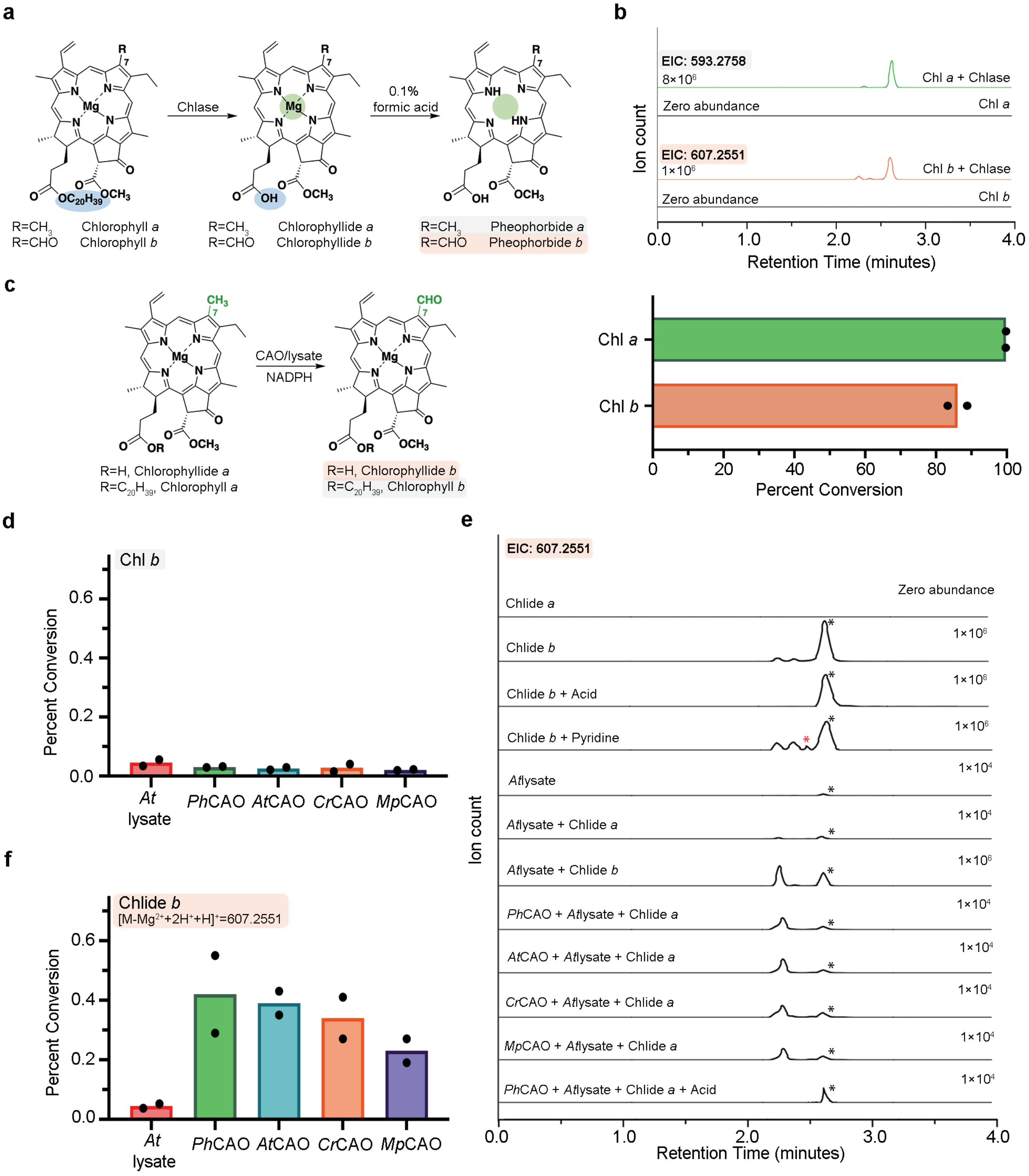


Figure 4.4. CAO homologs transform Chlide *a*, not Chl *a*, into Chlide *b* in cell lysate.

(a), A scheme that involves combination of a Chl pigment with recombinantly expressed and purified *T. aestivum* chlorophyllase (chlase) enzyme permits formation of the desired Chlide a and Chlide *b* pigments. (b), Activity of chlorophyllase on Chl *a* and Chl *b*. Extracted ion chromatograms of chlorophyllase activity assays product with Chl *a* (top trace) and Chl *b* (bottom trace). The LC-MS method designed and employed for pigment separation in this work relies on an acidic running solvent and causes loss of the central Mg^{2+} ion and the addition of two protons to the pigments under study. Therefore, the *m/z*=593.2758 and *m/z*=607.2551 represents the $[M+H]^+$ of Chlide a and Chlide b minus Mg^{2+} plus $2H^+$, respectively (left panel). Chlorophyllase converts nearly 100- and 85-percent of Chl a and Chl b into their Chlide counterparts, respectively (right panel). (c), Proposed reaction scheme catalyzed by CAO in cell lysate. (d), None of the CAO homologs can transform Chl *a* into Chl *b* in *A. thaliana* cell lysate. (e), The extracted ion chromatograms for the CAO homolog reaction products when combined with a Chlide *a* substrate and *A. thaliana* cell lysate reveal that all four CAO homologs can convert Chlide *a* into Chlide *b*. Of note, the cell lysate, pyridine, and acid can each shift the diastereomer equilibrium of the Chlide *b* standard (traces 3-4 and 7). The black asterisk indicates the major peak of the standard which is also observed in the assays. (f), PhCAO shows the highest percent conversion among all the four homologs in the presence of *A. thaliana* cell lysate and a Chlide substrate. All data shown in the bar graphs was performed in duplicate and data are presented as mean values.

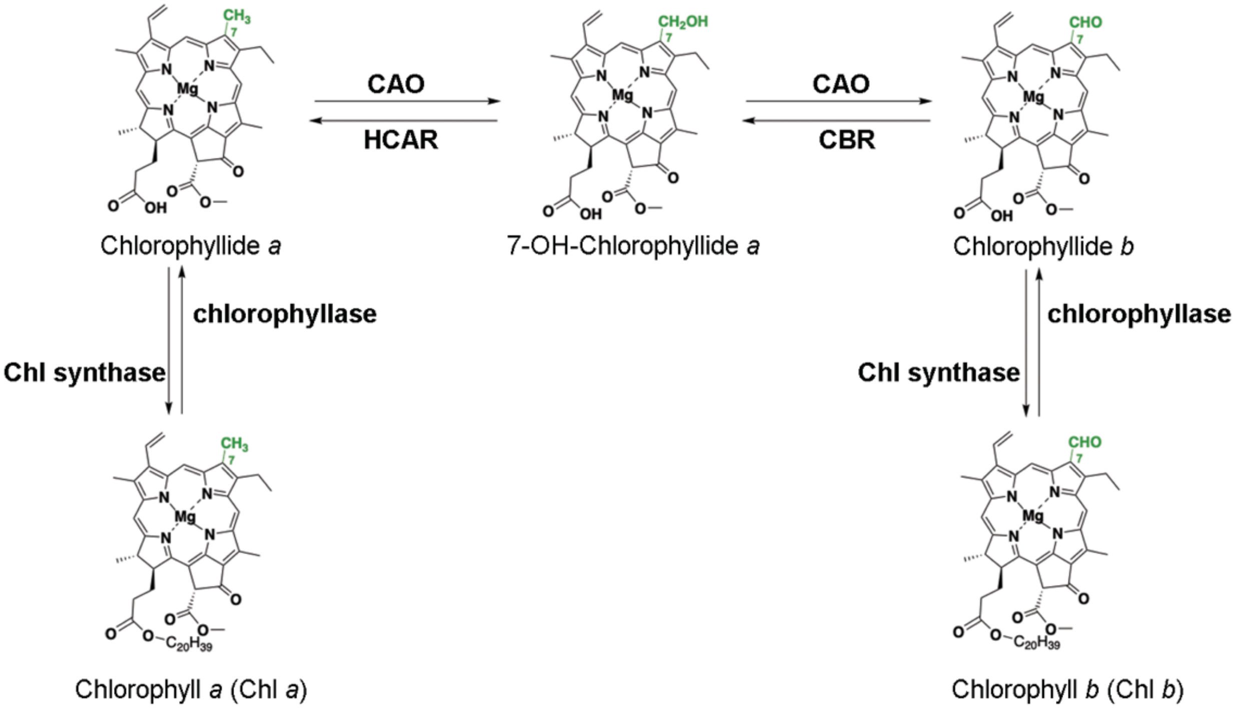


Figure 4.5. Proposed Chlorophyll metabolic pathway.

Chlorophyllase can hydrolyze Chlorophyll into Chlorophyllide by removing the phytol group and Chlorophyll synthase catalyze the reverse reaction. CAO, HCAR and CBR work together to fulfill the so-called Chl-cycle to transform Chlorophyllide a into Chlorophyllide b via 7-hydroxyl-Chlorophyllide a or the reverse reaction.

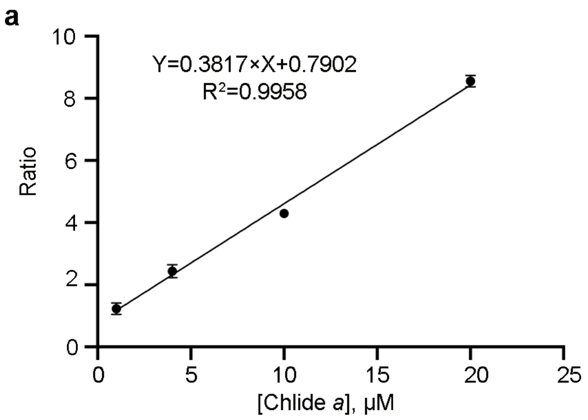


Figure 4.6. Standard curve for Chlide *a* and Chlide *b*.

A standard curve was used to convert for Q-TOF LC-MS peak area into the amount of pigments that were produced and consumed in the reactions. (a), A standard curve for Chlorophyllide *a*. (b), A standard curve for Chlorophyllide *b*. In both panels, the x-axis is the concentration of each Chlorophyllide compound whereas the y-axis is the ratio of the peak area integration of the Chlorophyllide compound to the 1,3,5-trimethoxybenzene (0.5 mM) internal standard. All data points were measured in duplicate.

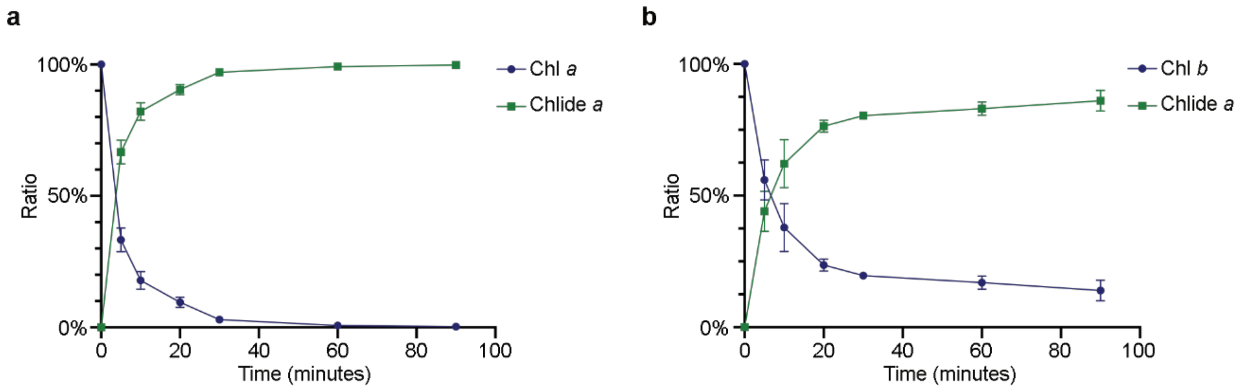


Figure 4.7. Chlorophyllase can be used as a tool to produce the needed Chlide *a* and Childe *b* standards.

(a), A time dependent measurement of Chl *a* (1 mM, blue) consumption and Chlide *a* production using chlorophyllase (1.5 µM, green). The chlorophyllase reaction reaches the maximum amount of substrate conversion at 1 hour, which corresponds to conversion of nearly 100-percent Chl *a* into Chlide *a*. (b), A time dependent measurement of Chl b (1 mM, blue) consumption and Chlide *b* production using chlorophyllase (1.5 µM, green). The chlorophyllase reaction reaches the maximum amount of substrate conversion at 1.5 hours. At this point, nearly 85-percent of the Chl *b* substrate has been converted into Chlide *b*. For both panels, the data points were measured in duplicate.

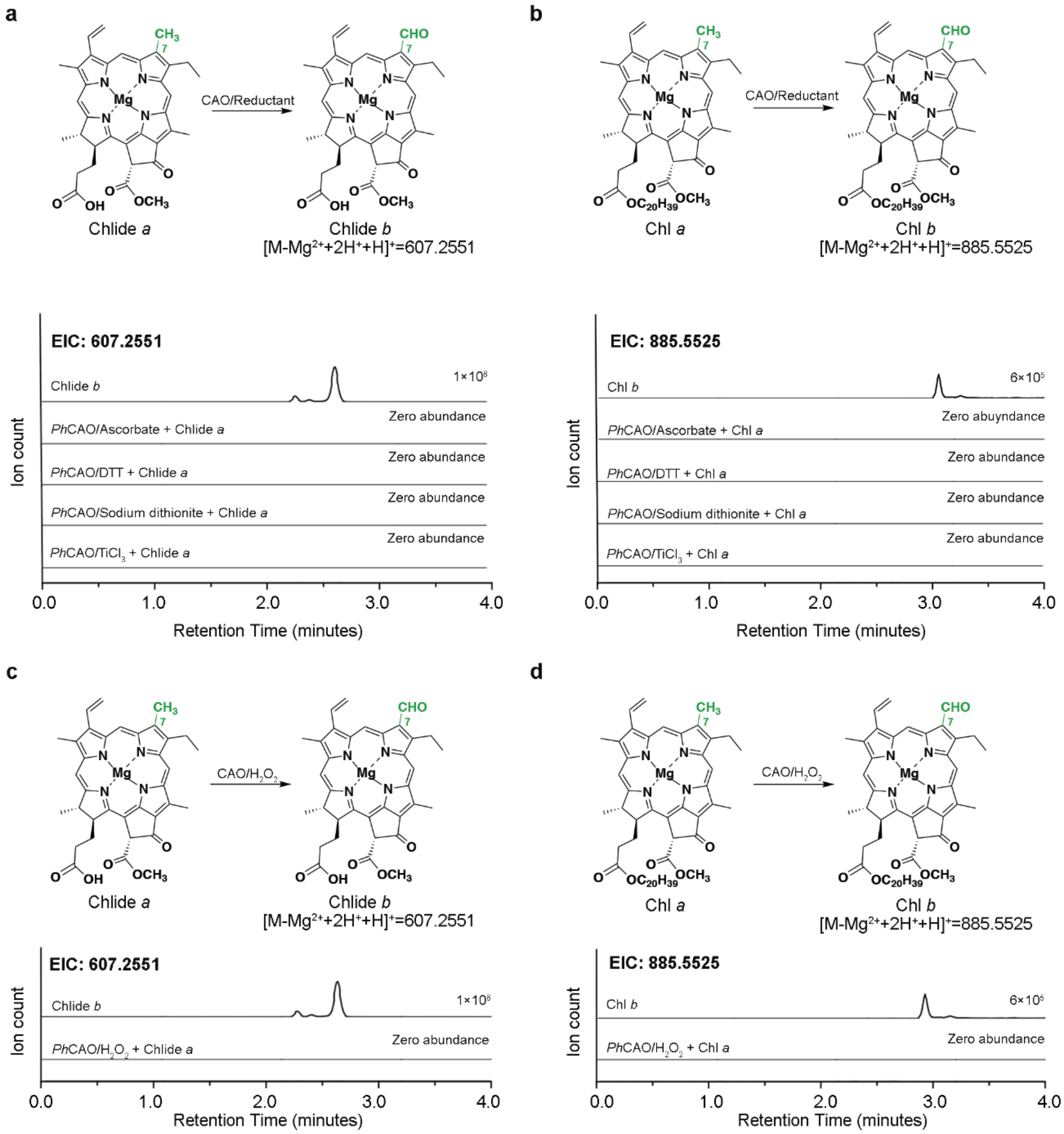


Figure 4.8. Several chemical reductants and the peroxide (H_2O_2) shunt reaction were tested in the CAO reactions.

(a), CAO is unable to convert Chlide *a* into Chlide *b* using ascorbate, dithiothreitol (DTT), sodium dithionite, or titanium (III) chloride ($TiCl_3$) as a reductant. (b), Similarly, CAO does not catalyze the conversion of Chl *a* into Chl *b* using a chemical reductant. (c), CAO is unable to transform Chlide *a* into Chlide *b* using H_2O_2 to bypass the need for an electron donor. (d), Likewise, CAO does not show any activity on Chl *a* with H_2O_2. In the extracted ion chromatograms (bottom panels), the *m/z* of 607.2551 and 885.5525 represent the mass of $[M+H]^+$ of Chlorophyllide *b*, Chlorophyll *b* minus Mg^{2+} plus $2H^+$, respectively. All data was performed in triplicate.

equivalent assay likewise showed no production of Chl *b* (Figure 4.4.d). However, when the same reaction was performed with Chl *a*, chlorophyllase, *Ph*CAO, and *A. thaliana* cell lysate, two new peaks were formed that matched the exact mass of the expected Chlide *b* product (Figure 4.4.e and Figure 4.9). The retention time of one of these peaks matched the major peak of the Chlide *b* standard (Figure 4.4.e, black asterisk). To investigate the nature of the second peak, samples of the Chlide *b* standard were treated with acid and pyridine. The latter of these molecules has been previously shown to change the stereochemistry at the C-13^2 position of Chl pigments, presumably via removal of the acidic proton at this position and enolization[27-29]. Consistently, in both cases, the LC-MS peak distribution of the standard changed (Figure 4.4.e). Similarly, incubation of the Chlide *b* standard in *A. thaliana* cell lysate showed interconversion of peaks, suggesting that there may be an enzyme in the lysate, which like acid and base, interchanges the stereochemistry of Chlide *b* (Figure 4.4.e). From these results, it was hypothesized that the different observed LC-MS peaks from the reaction represented diastereomers of Chlide *b*, which contains three chiral centers at C13^2, C17 and C18 (Figure 4.1.a). This assignment is further supported by MS/MS data, which yielded identical fragmentation patterns between the Chlide *b* standard and the pyridine-treated sample (Figures 4.10 and 4.11). Substitution of *Ph*CAO in the assays that contained Chl *a*, chlorophyllase, and *A. thaliana* cell lysate with *At*CAO, *Cr*CAO, or *Mp*CAO resulted in similar formation of product diastereomers (Figure 4.4.f and Figure 4.9).

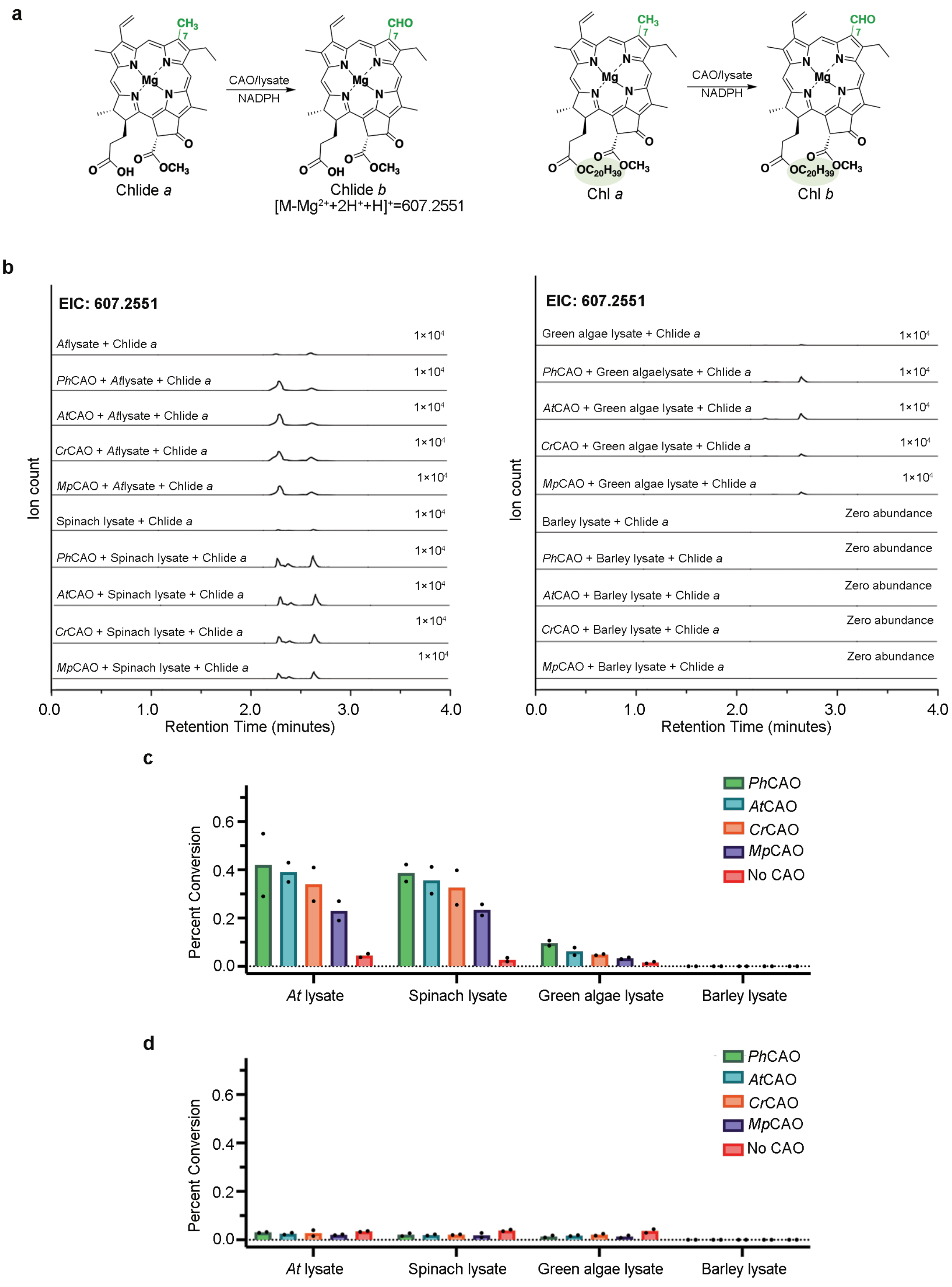


Figure 4.9. CAO activity with different cell lysate.

The ability of cell lysate from four different photosynthetic organisms was tested for its ability to support Chlide *b* production. (a), Reaction scheme for the CAO catalyzed conversion of Chlide *a* into Chlide *b* or Chl *a* into Chl *b* in different cell lysates. (b), The extracted ion chromatograms of the CAO-catalyzed reactions performed in cell lysate with a Chlide *a* substrate. (c), A bar chart to compare the activity of CAO homologs in different cell lysates using Chlide *a* as the substrate. Of note, *Arabidopsis thaliana* cell lysate allows for the highest production of Chlide *b* for each of the different CAO homologs. Spinach lysate shows similar level activity compared with *Arabidopsis thaliana* lysate. Green algae lysate, on the other hand, shows much lower activity. Finally, barley lysate failed to support the ability of any CAO homolog to produce Chlide *b*. (d), A bar chart to compare the activity of CAO homologs with different cell lysates using Chl *a* as the substrate. None of the tested cell lysates shows the ability to support any of the CAO homologs to transform Chl *a* into Chl *b*. All data in the bar graph was measured in duplicate and are shown as mean values ± SD.

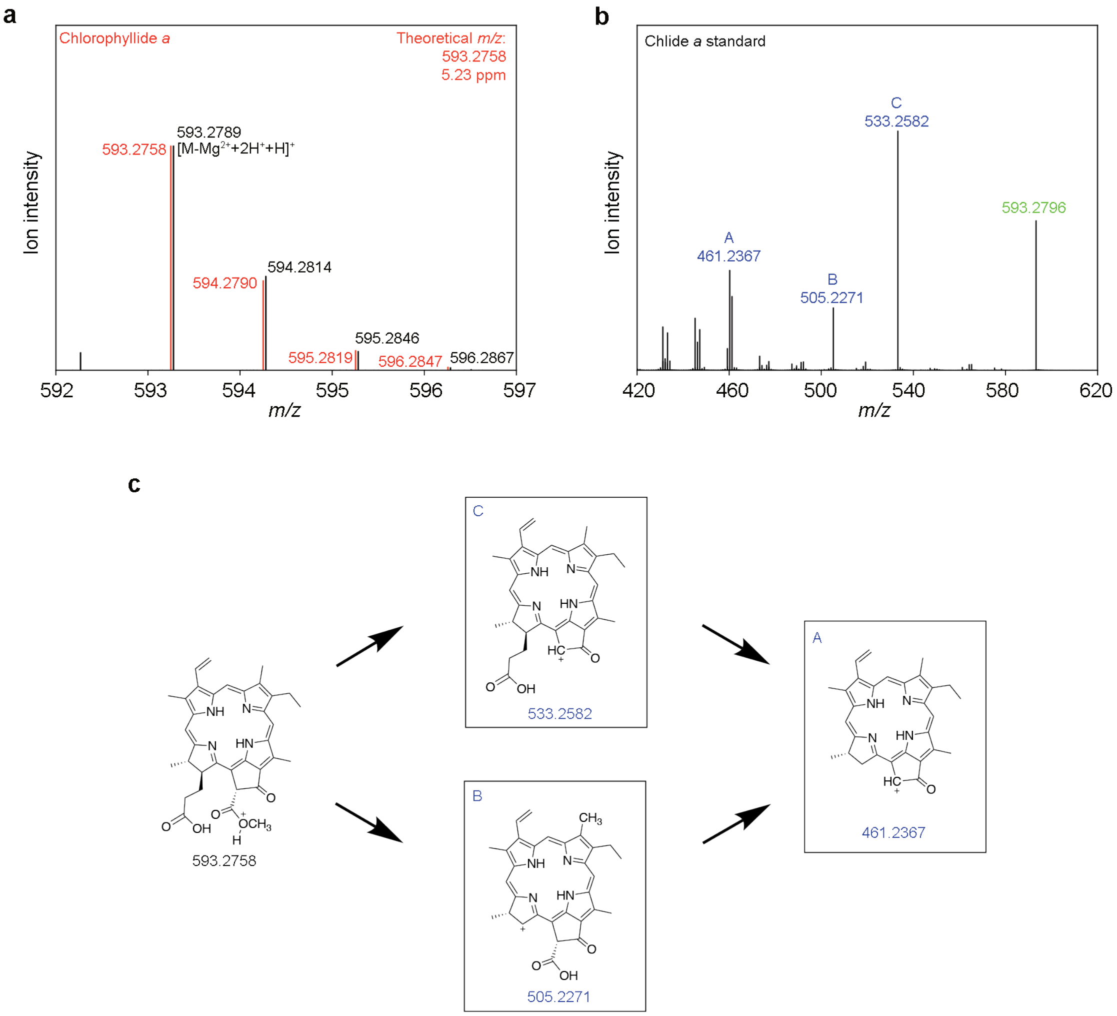
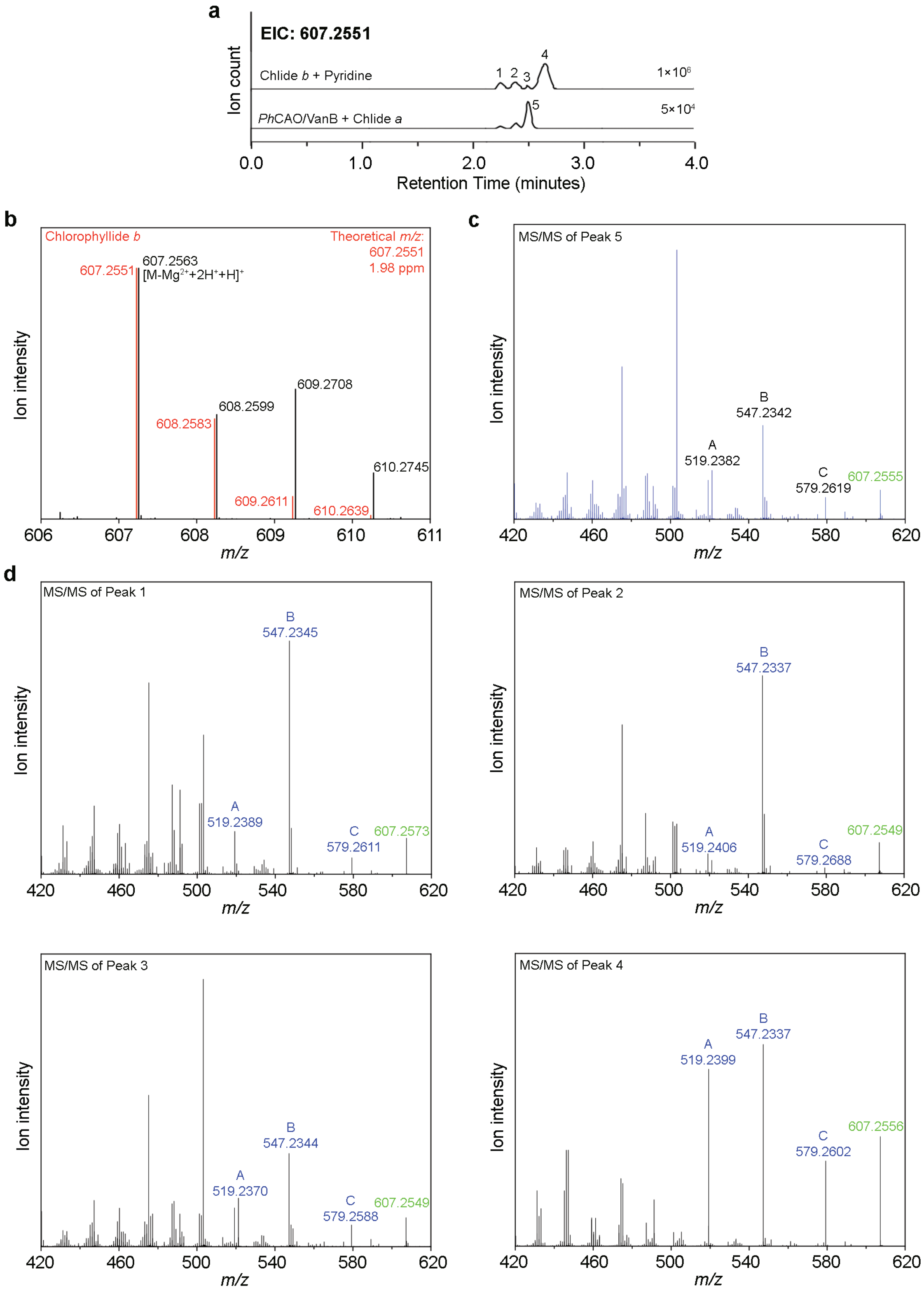


Figure 4.10. MS and MS/MS analysis of the enzymatically produced Chlide *a* standard.

(a), The predicted (red) and the observed (from the chlorophyllase reaction, black) isotope distribution for Chlide *a* show good agreement. (b), MS/MS spectrum of the Chlide *a* standard shows three major fragments of interest, labelled as A, B and C. (c), The proposed fragmentation pathway for Chlide *a*. Of note, the major fragments (C and A) have a mass loss of 60.0176 Da ($C_2H_4O_2$) and 132.0391 ($C_5H_8O_4$), suggesting that the side chains on C13^2 and C17 are relatively easy to dissociate.

Additional studies were performed to evaluate whether cell lysate from spinach, green algae, or barley would support higher conversion of the other purified CAO homologs (Figure 4.9.a-c). These experiments revealed that substitution of *A. thaliana* cell lysate with spinach cell lysate in the assays, results in a comparable amount of Chlide *b* formation in all four homologs (Figure 4.9.b-c). The substitution of green algae lysate, however, resulted in a much lower yield relative to *A. thaliana*. Barley lysate didn't facilitate the CAO-mediated production of Chlide *b* at all (Figure 4.9.b-c). Of note, these additional cell lysate options also failed to support the ability of CAO to convert Chl *a* into Chl *b* (Figure 4.9.d). Finally, to evaluate whether an additional component that is found exclusively in photosynthetic organisms is required to support the activity of the different CAO homologs, the assays were also performed in *E. coli* cell lysate (Figure 4.12.a). Unlike that previously described[10], without the addition of an external reducing system, these assays showed the production of a small amount of Chlide *b* and suggested the possibility that the *in vitro* activity of these proteins may be supported using a non-native reductase (Figure 4.12.a).


a
EIC: 607.2551
Ion count
Chlide *b* + Pyridine
1 2 3 4
1×10⁶
*Ph*CAO/VanB + Chlide *a*
5
5×10⁴
0.0 1.0 2.0 3.0 4.0
Retention Time (minutes)
b
Chlorophyllide *b*
Theoretical *m/z*: 607.2551 1.98 ppm
607.2563 [M-Mg²⁺+2H⁺+H]⁺
607.2551
608.2599
608.2583
609.2708
609.2611
610.2745
610.2639
Ion intensity
606 607 608 609 610 611
m/z
c
MS/MS of Peak 5
B 547.2342
A 519.2382
C 579.2619
607.2555
Ion intensity
420 460 500 540 580 620
m/z
d
MS/MS of Peak 1
B 547.2345
A 519.2389
C 579.2611
607.2573
MS/MS of Peak 2
B 547.2337
A 519.2406
C 579.2688
607.2549
MS/MS of Peak 3
B 547.2344
A 519.2370
C 579.2588
607.2549
MS/MS of Peak 4
B 547.2337
A 519.2399
C 579.2602
607.2556
Ion intensity
420 460 500 540 580 620
m/z

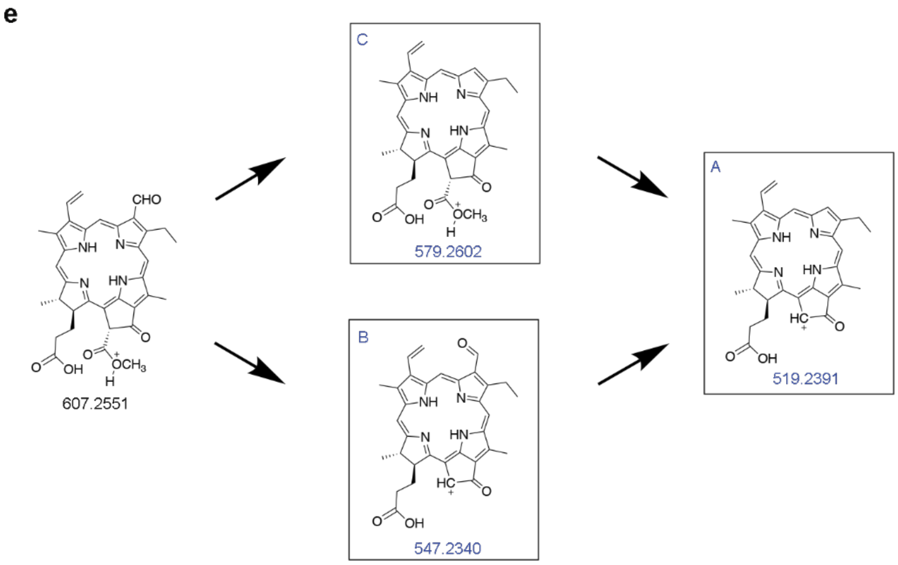


Figure 4.11. Analysis of the enzymatically (chlorophyllase) produced Chlide *b* standard and the product of the CAO reaction.

(a), Extracted ion chromatograms for Chlide *b* standard and the product of the *Ph*CAO reaction. There are four peaks shown in Chlide *b* standard (labeled peaks 1-4) and the product peak of the *Ph*CAO reaction (peak 5) has the same retention time as peak 3 from the standard. (b), The predicted (red) and observed (from the *Ph*CAO reaction, black) isotope distribution for Chlide *b* show good agreement. (c), The MS/MS spectrum of peak 5 from the *Ph*CAO reaction. (d), The MS/MS spectrum of peak 1-4 from Chlide *b* standard. Each of the peaks 1-4 show different fragmentation patterns consistent with them being diastereomers of Chlide *b*. Comparison of the retention time (panel a) and MS/MS spectrum of the product peak of CAO reaction (panel c) with the Chlide *b* standard (panel d), suggests that the major product of CAO reaction shows the same diastereomer configuration as peak 3 in the Chlide *b* standard. (e), A proposed fragmentation pathway for Chlide *b* that gives rise to the fragments that are labeled A, B, and C, and are observed in all five MS/MS spectra. Fragment C, which demonstrates a mass loss of 27.9949 Da corresponds to the loss of a formyl group.

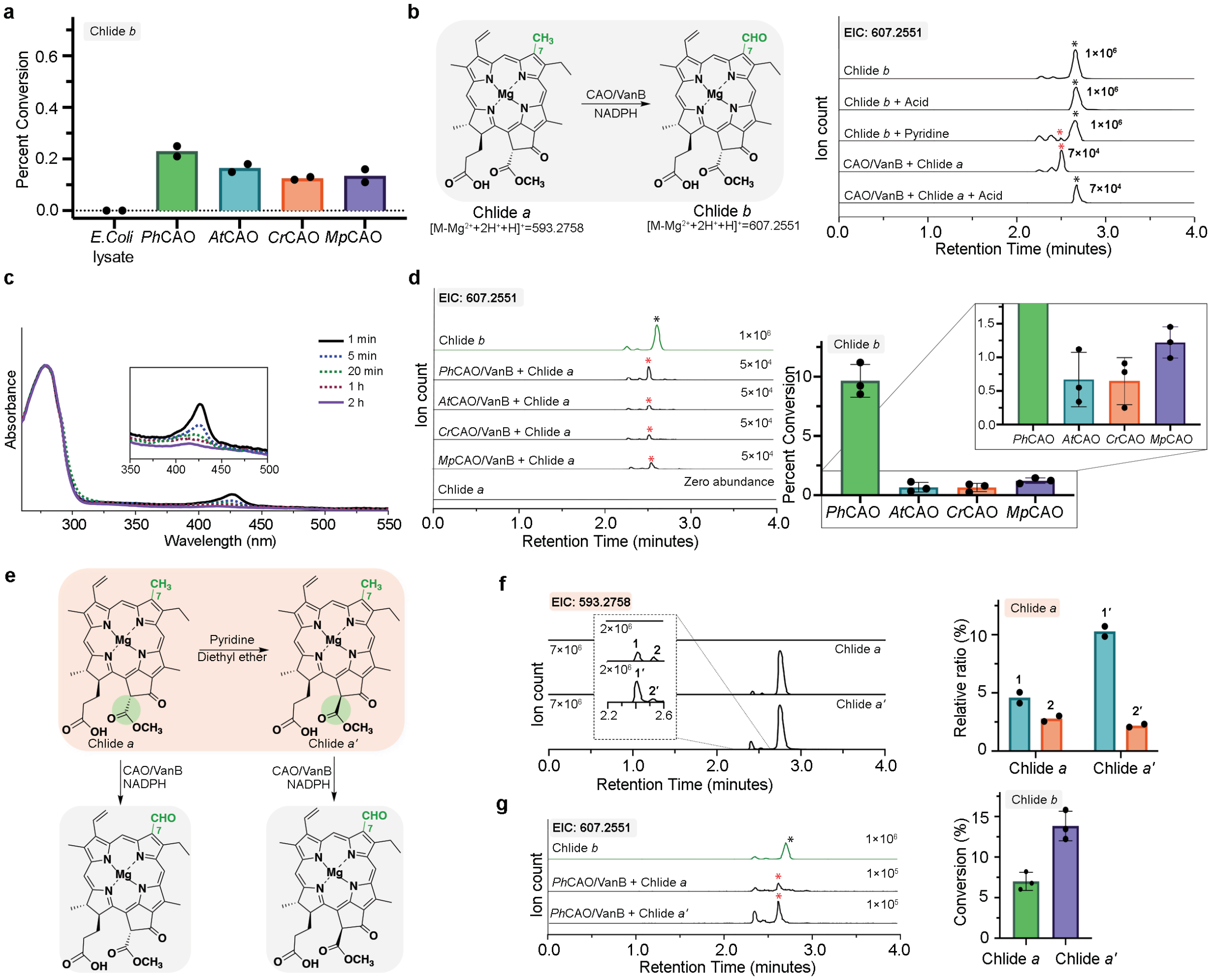


Figure 4.12. CAO homologs transform Chlorophyllide *a* into Chlorophyllide *b* in the presence of the non-native reductase VanB.

(a), All four CAO homologs show the ability to convert Chlide *a* into Chlide *b* in *E. coli* cell lysate, suggesting that a non-native reductase may work with CAO to facilitate the reaction. *Ph*CAO shows the highest percent conversion among all four homologs. (b), Reaction scheme of the CAO/VanB catalyzed conversion of Chlide *a* into Chlide *b* (left panel). Treatment of the Chlide *b* standard with acid or pyridine shows new diastereomer peaks. Similarly, the reaction product from the CAO homolog reactions can be converted into the main diastereomer of the Chlide *b* standard under acidic conditions (right panel). (c), UV-Vis absorption spectra of the Rieske [2Fe-2S] cluster in *Ph*CAO shows that it can be reduced by VanB over time. (d), The extracted ion chromatograms for the product formed when the CAO homologs are combined with Chlide *a*, VanB, and NADPH. This data shows a diastereomer of Chlide *b* is formed in these reactions. Again, *Ph*CAO shows the highest percent conversion among all four homologs. (e), Reaction scheme for converting Chlide *a* into Chlide *a′* with pyridine. (f), An extracted ion chromatogram of the Chlide *a* peak suggests the Chlide *a′* sample shows a different diastereomer distribution than Chlide *a*. (g), PhCAO shows a higher percent conversion on Chlide *a′* than Chlide *a*. In panels d and g, reactions were performed in triplicate and in a and f, they were performed in duplicate. In all bar graphs presented, data are shown as mean values.

4.3.3 The combination of CAO with a non-native reductase allows for production of Chlorophyllide *b*

To evaluate whether a non-native reductase system, rather than cell lysate, could facilitate the CAO reaction, a flavodoxin–flavodoxin reductase system from *E. coli* and several different known Rieske reductase systems were purified. These include the ferredoxin-NAD$^+$ reductases, VanB and TsaB, and the two-component [2Fe-2S] cluster containing ferredoxin (DdmB) and ferredoxin–NAD$^+$ reductase system (DdmA). These proteins, as well as a spinach ferredoxin–ferredoxin reductase system, were subsequently tested as candidate reductases for the CAO homologs (Figure 4.13). Here, the combination of a CAO homolog with Chl *a*, chlorophyllase, and either *E. coli* flavodoxin–flavodoxin reductase or spinach ferredoxin–ferredoxin reductase did not result in any detectable production of Chlide *b* (Figure 4.13). In contrast, the use of VanB, DdmB/DdmA, or TsaB and NADPH in the assays resulted in production of a compound, which has the same molecular weight as Chlide *b* but is shifted in retention time relative to the enzymatically produced Chlide *b* standard (Figure 4.12.b red asterisk and Figure 4.13). The highest amount of this compound was produced in the assays that contained VanB (Figure 4.13, Figure 4.12.b). Although produced to a much greater extent, like that described above, the retention time of this molecule could be shifted to match the Chlide *b* standard by the addition of *A. thaliana* cell lysate or by treatment with acid (Figure 4.4.e and 4.12.b). Similarly, formation of a matching peak could be made in the Chlide b standard via the addition of pyridine (Figure. 4.12.b, Figure 4.14). Last, as it appeared that VanB could support the CAO-catalyzed production of Chlide *b*, it was investigated whether it was providing the needed electrons for the reaction. Indeed, the Rieske

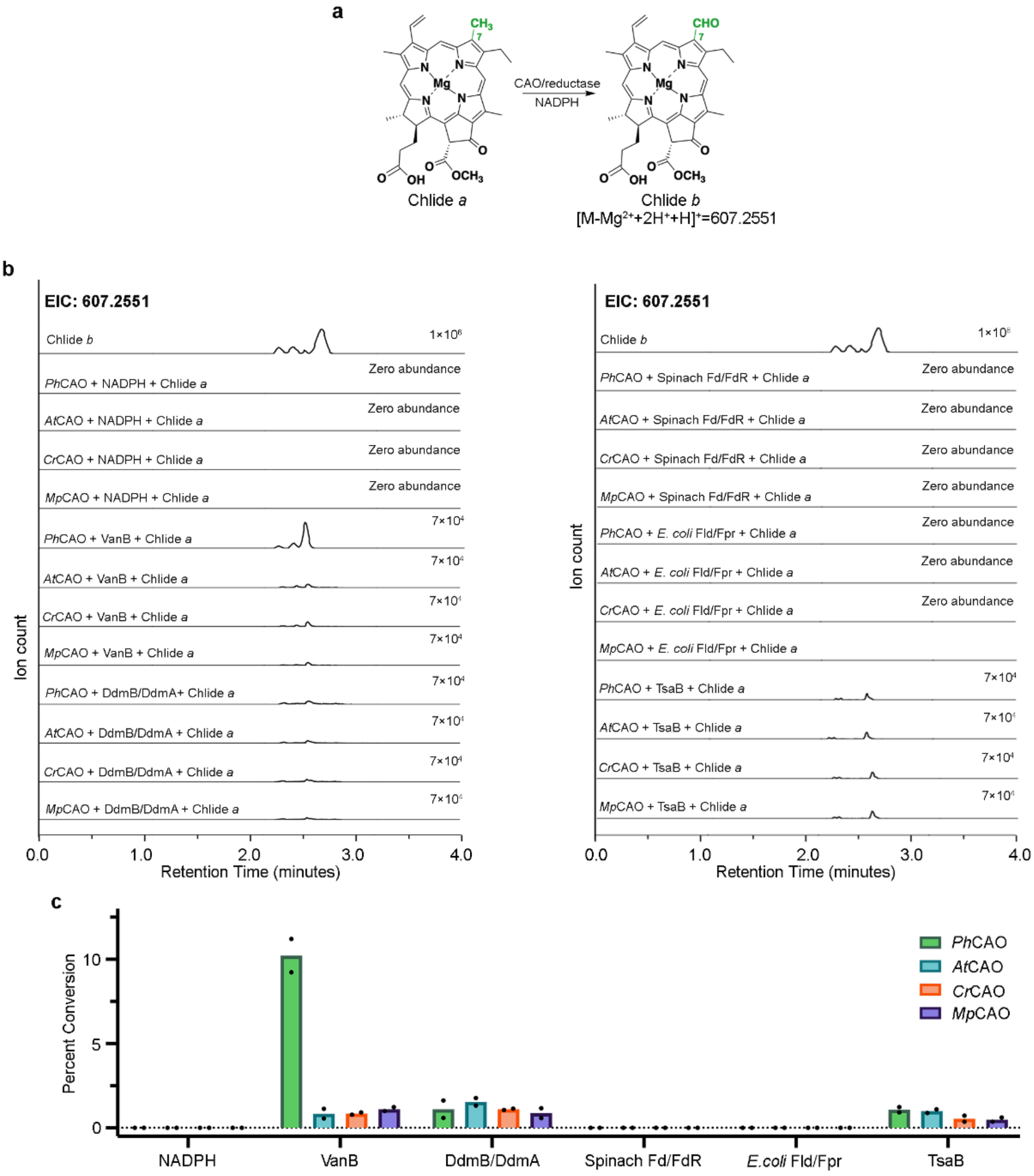


Figure 4.13. CAO activity with different non-native reductase systems.

The ability of five different non-native reductase systems were tested in their ability to support Chlide *b* production. (a), Reaction scheme of the CAO catalyzed conversion of Chlide *a* into Chlide *b* with NADPH and with NADPH and different reductase systems. (b), The extracted ion chromatograms of the products from the CAO reactions that were performed using NADPH and five different reductases (VanB, DdmB/DdmA, Spinach ferredoxin/ferredoxin reductase, E. coli flavodoxin/flavodoxin reductase, and TsaB). (c), A bar chart to compare the activity of the CAO homologs with NADPH and the different tested reductases. VanB, the reductase partner of a

Rieske oxygenase, vanillate *O*-demethylase, shows the highest activity among all tested reductase systems. All data in the bar graph was performed in duplicate and are shown as mean values ± SD.

[2Fe-2S] cluster in *Ph*CAO was shown to be reduced by addition of NADPH and VanB over time using UV-Vis (Figure 4.12.c). The *At*CAO, *Cr*CAO, and *Mp*CAO homologs were then tested in equivalent assays that contained VanB, NADPH, Chl *a* and chlorophyllase, and it was demonstrated that each homolog yielded the same diastereomer of Chlide *b*, albeit to a lower extent than *Ph*CAO (Figure 4.12.d, red asterisk). Finally, it was also tested whether NADPH alone was sufficient to aid in CAO chemistry. However, in this experiment it was determined that an electron mediator between NADPH and CAO is required to support activity (Figure 4.13).

4.3.4 CAO shows a preference for Chlide *a′* diastereomer

The intriguing observation that CAO produces a different major diastereomer of Chlide *b* than the standard, prompted an investigation into the preferred Chlide a substrate diastereomer. As described above, previous work has demonstrated that the addition of pyridine to a Chl pigment can change the absolute configuration at the C-13^2 position[27-29]. A similar phenomenon has also been shown to occur following addition of triethylamine to Chl[27]. Therefore, to investigate the preferred diastereomer of the CAO-catalyzed reaction, both methods were tested for their ability to produce chlorophyll *a′* (Chl *a′*) from Chl *a*. Chl *a′* has an (*S*)-13^2-carbomethoxy group rather than the (*R*)-13^2-carbomethoxy group found in Chl a. In these experiments, only subtle changes were noticed in the LC-MS analysis after addition of triethylamine (Figure 4.14). In contrast, the product of the pyridine-treated reaction, which is a mixture of Chl *a′* and Chl *a*[29], shows a near two-fold increase in the size of one diastereomer peak following the chlorophyllase catalyzed hydrolysis reaction (Figure 4.15 and Figure 4.12.f). Despite these differences, the chlorophyllase-cleaved samples of Chl *a*′ and Chl *a* behave similarly in cell lysate, again suggesting the presence of an enzyme that can change the stereochemistry of Chlide molecules (Figure 4.16). More interestingly, however, combination of the produced Chlide *a*′ mixture with *Ph*CAO, VanB, and NADPH resulted in almost twice as much Chlide *b* product (Figure 4.12.g).

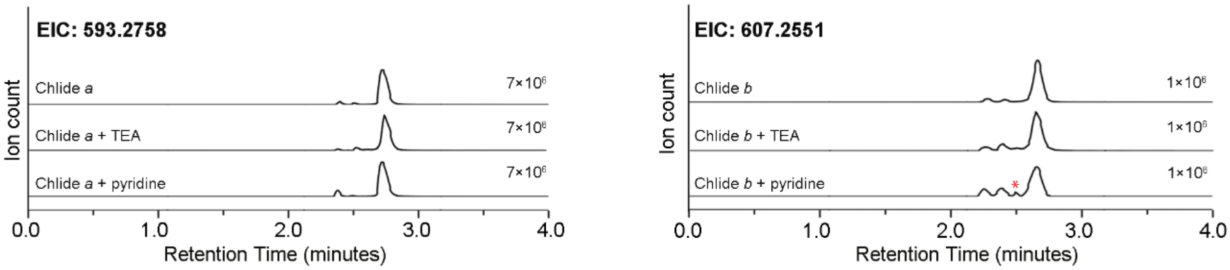


Figure 4.14. The diastereomer distribution of Chlide *a* (left) and Chlide *b* (right) can be shifted by triethylamine (TEA) or pyridine.
The peak in Chlide *b* with pyridine that matches the product peak in CAO reaction is labelled with a red asterisk. The *m/z*=593.2758 and *m/z*=607.2551 represents the $[M+H]^+$ of Chlide *a* and Chlide *b* minus Mg^{2+} plus $2H^+$, respectively.

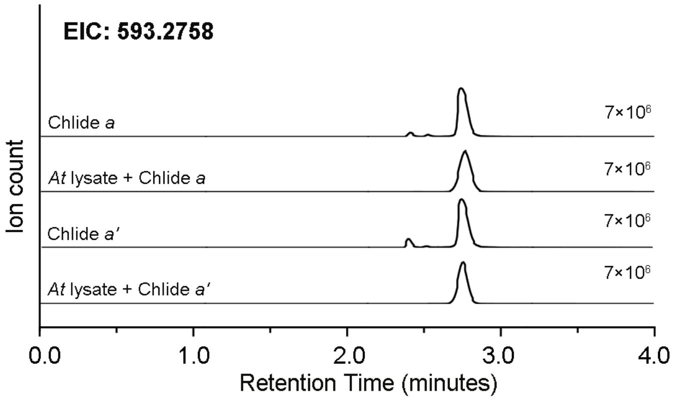


Figure 4.15. The diastereomer equilibrium of the Chlide *a* and Chlide *a*′ can be shifted by *A. thaliana* cell lysate.

Related experiments on extracted chlorophyllase from *Melia azedarach* and *Tetragonia ezpansa* have also revealed that chlorophyllase is stereospecific and prefers a substrate that has an

(*R*)-13^2-carbomethoxy group. More specifically, chlorophyllase shows a preference for hydrolyzing Chl *a* into Chlide a rather than Chl *a′* into Chlide *a′*[30]. Our data on chlorophyllase is consistent with this previous work: the purchased Chl *a* is a mixture of diastereomers, as reported from supplier, that appear as different peaks in the LC-MS experiment. Incubation of this mixture for an extended amount of time with chlorophyllase results in complete hydrolysis but the major diastereomer of the purchased Chl *a* is consumed at a faster rate (Figures 4.15). Treatment of Chl *a* with pyridine and subsequent hydrolysis with chlorophyllase also leads to complete conversion into Chlide, but as described above, the product mixture has a different diastereomer distribution

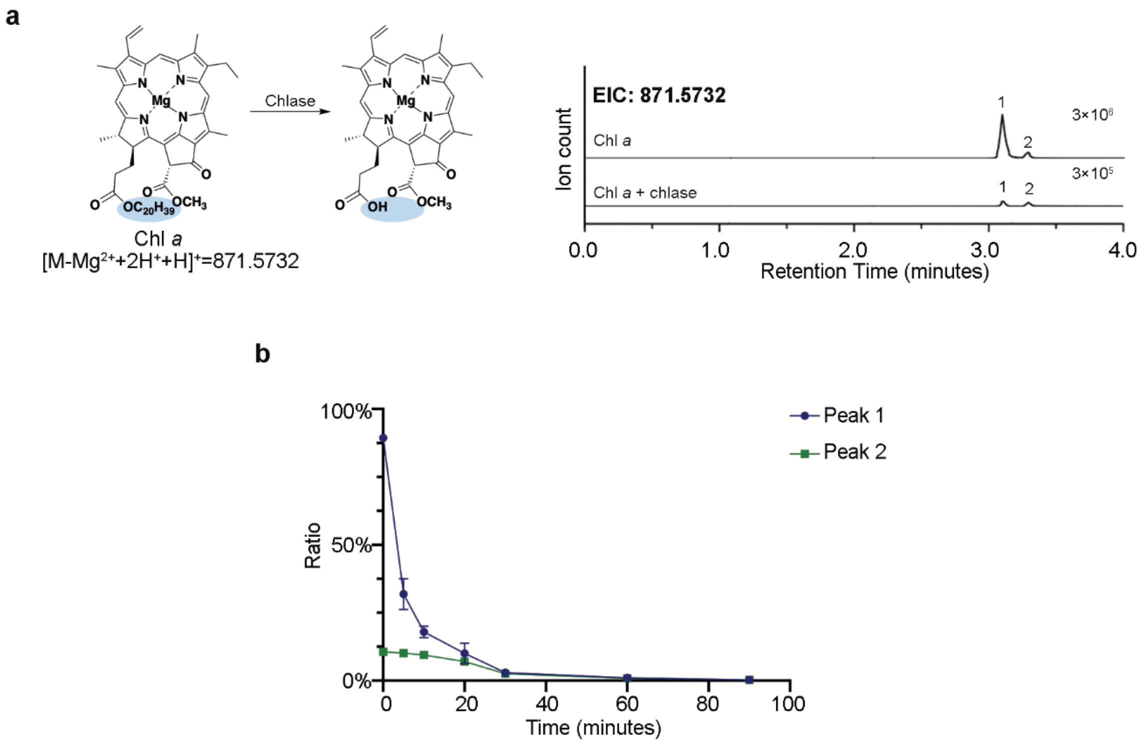


Figure 4.16. A certain diastereomer of Chl *a* is preferred by chlorophyllase**.**
(a), Reaction scheme and the extracted ion chromatograms of the substrate (Chl *a*) from the hydrolysis reaction by chlorophyllase (abbreviated chlase). (b), A time dependent measurement of the consumption of the two diastereomers present in the Chl *a* standard. The diastereomer of peak 1 is preferred by chlorophyllase although both diastereomers get are mostly consumed after the 1-hour incubation.

than the non-pyridine treated Chlide product (Figure 4.12.f). Combined with the data that shows incubation of the Chl *b* product standard with pyridine results in the production of a peak that matches the major peak produced in the *in vitro* CAO reaction, this data suggests that CAO shows a clear preference for the pyridine treated substrate and produces a major product that is the equivalent to the sample of pyridine treated Chlide *b* (Figure 4.12.e and 4.12.g). As observed

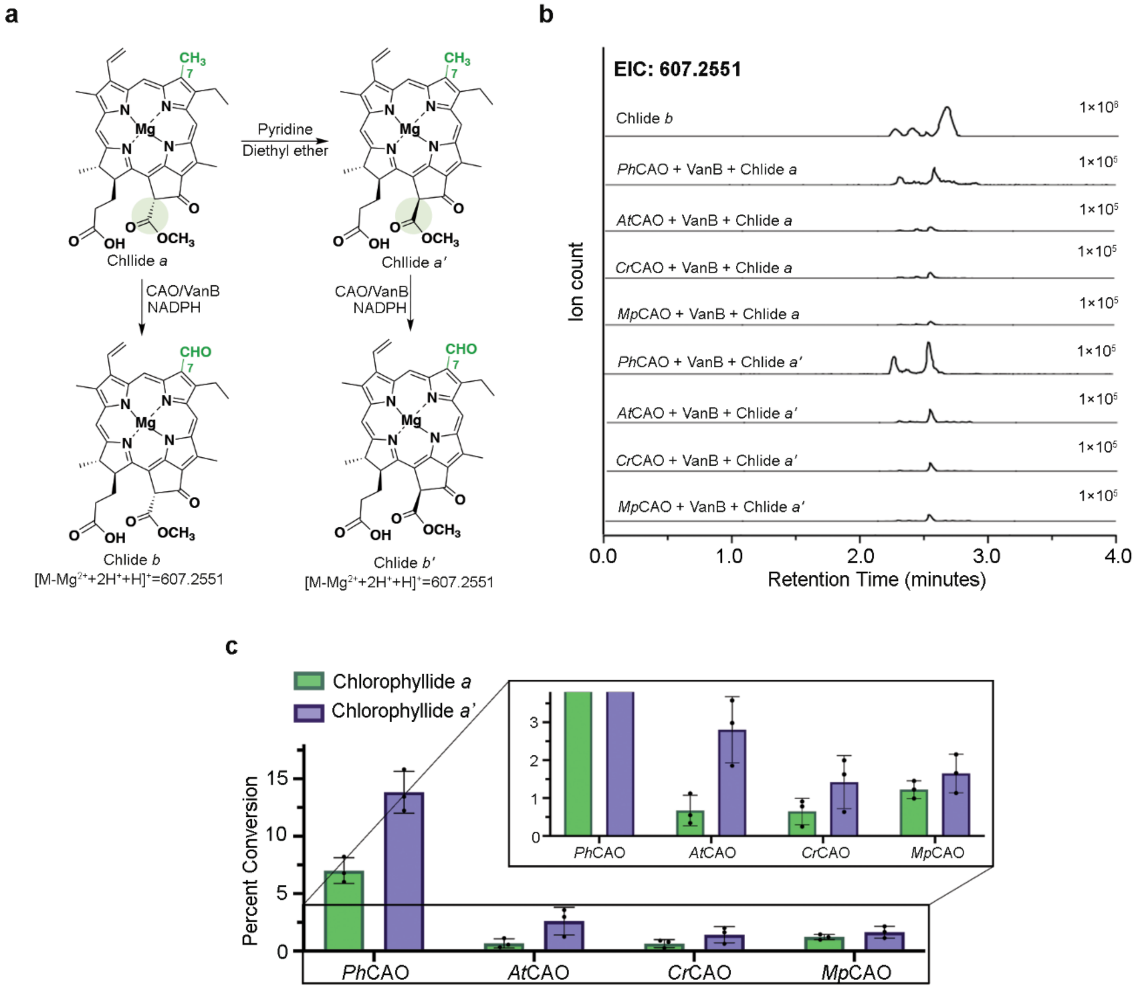


Figure 4.17. The Chlide *a′* diasteromer is preferred by all tested homologs of CAO. (a), Reaction scheme that shows the pyridine-mediated transformation of Chlide a and Chlide *a′*. (b), The extracted ion chromatograms of the CAO reaction product using either Chlide *a* or Chlide *a′* as a substrate. (c), A bar chart to compare the percent yield of CAO homologs with Chlide *a*

and Chlide *a′*. This data revealed that all four homologs show a preference for Chlide *a′* rather than Chlide *a*. All data points were measured in triplicate and are shown as mean values ± SD.

for *Ph*CAO, the other homologs are also able to use Chl *a′* as a substrate, albeit to different extents (Figure 4.17).

4.3.5 CAO is a Rieske oxygenase that catalyzes sequential monooxygenation reactions

The proposal in Figure 4.1.a suggests that CAO works by catalyzing sequential monooxygenation reactions that transform the C7-methyl group into a formyl group via a 7-hydroxymethyl Chl(ide) *a* intermediate (7-OH-Chl(ide) *a*). To probe whether 7-OH-Chlide *a* is a true intermediate of the CAO reaction *in vitro*, 7-OH-Chlide *a* was produced in two steps. First, the formyl group of Chl *b* was reduced with sodium borohydride to produce 7-OH-Chl *a*[31-32] and second, the phytol tail of 7-OH-Chl *a* was cleaved using purified *T. aestivum* chlorophyllase (Figure 4.18.a, Figures 4.19 and 4.20). The successful production of the 7-OH-Chlide *a* molecule was then evaluated using LC-MS and UV-Vis spectroscopy (Figure 4.18.b-c). LC-MS demonstrated production of a single peak and the Q bands of the UV-Vis absorption spectrum of this pigment, as previously described[31-32], are blue shifted relative to Chlide *a* (Figure 4.18.b-c). The produced 7-OH-Chlide *a* compound was then used as a standard to look for its production in the assays that contained *Ph*CAO, VanB, NADPH, Chl *a*, and chlorophyllase (Figure 4.18.d). This experiment was complicated by the existence of two peaks in the LC-MS trace that showcase different retention times, have the expected mass of the 7-OH-Chlide *a* intermediate, and are present in the control reaction that lacks CAO (Figure 4.18.e). However, in the LC-MS trace from the reaction that contained CAO, one peak at relatively low intensity was present that did not appear in the control (Figure 4.18.e). This peak displayed a different retention time compared to the 7-OH-Chlide *a* standard, but like what was observed in the CAO-Chlide *a* reaction, the retention time of this peak shifted when treated with acid (Figure 5e). Again, this peak was hypothesized to be a diastereomer of 7-OH-Chlide *a* and MS/MS analysis revealed that this peak had a similar fragmentation pattern to the produced 7-OH-Chlide *a* standard (Figure 4.21).

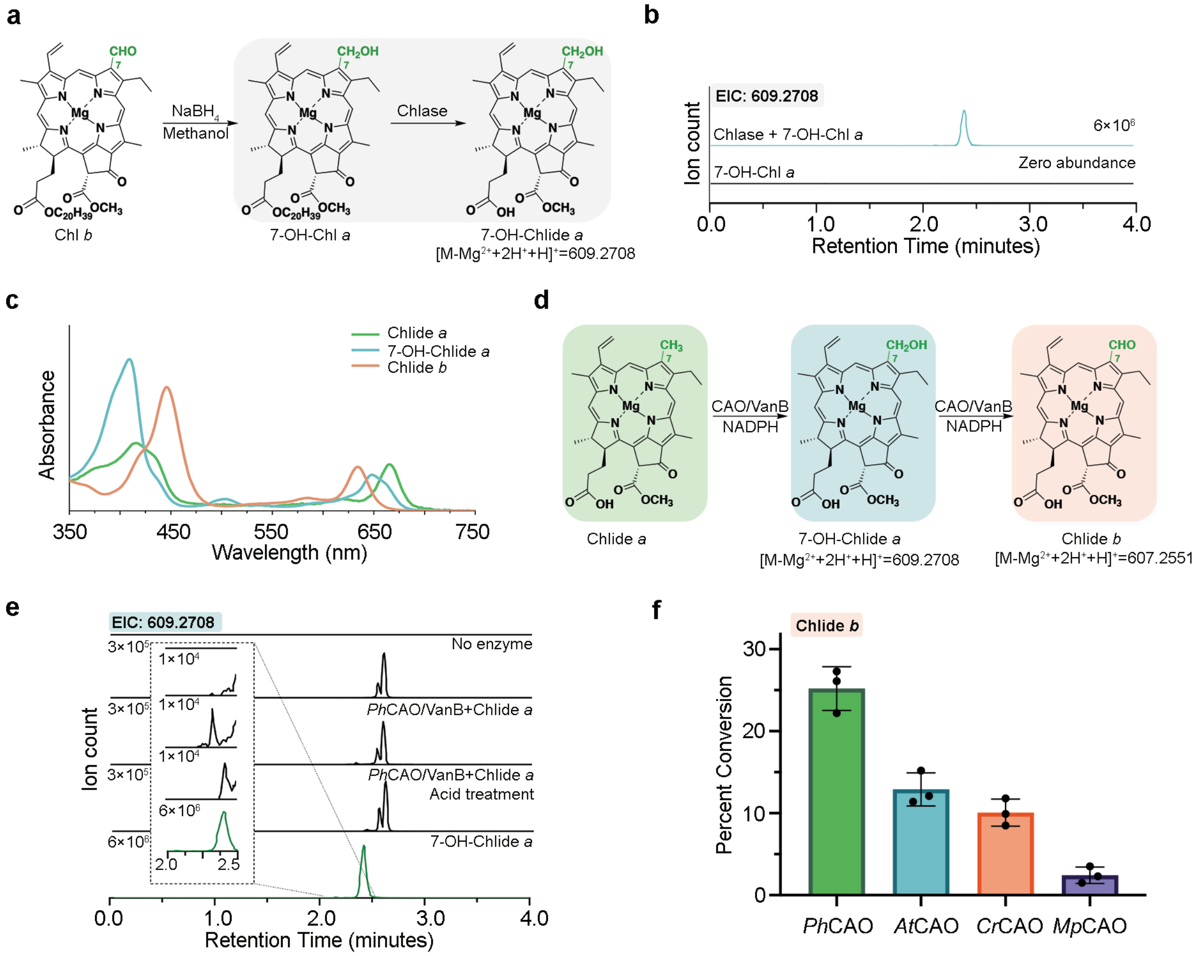


Figure 4.18. 7-OH-Chlide *a* can be transformed into Chlide *b* by combination of the CAO homologs with VanB and NADPH.

(a), Reaction scheme for synthetic conversion of Chl *b* into 7-OH-Chlide *a*. (b), Extracted ion chromatograms of chlorophyllase catalyzed hydrolysis of 7-OH-Chl *a*. The *m/z*=609.2708 represents the $[M+H]^+$ of 7-OH-Chlide *a* minus Mg^{2+} plus $2H^+$ (see Figure 3a). (c), Absorbance spectra of the Chlide *a*, Chlide *b* and 7-OH Chlide *a* that were enzymatically and synthetically produced in this work. (d), Reaction scheme of sequential hydroxylation reactions that is catalyzed by CAO. (e), Extracted ion chromatograms for the *Ph*CAO reaction products reveals the formation of a 7-OH-Chlide *a* intermediate. (f), All four CAO homologs show the ability to transform the intermediate (7-OH-Chlide *a*) into the final product (Chlide *b*). Again, *Ph*CAO shows the highest percent conversion among all four homologs.

To further investigate the sequential hydroxylation proposal, 7-OH-Chlide *a*, the proposed intermediate, was added into an assay that contained CAO, VanB, and NADPH as a substrate. In this experiment, it was observed using LC-MS, that all four CAO homologs could produce Chlide *b*. In these reactions, the product displayed the exact same retention time as the main peak of the Chlide *b* standard, which is presumably due to it being synthesized from the same Chl *b* precursor (Figure 4.22). Again, when provided with the 7-OH-Chlide *a* substrate, *Ph*CAO showed the highest percent conversion among all four CAO homologs (Figure 4.18.f). Therefore, the steady-state kinetic behavior was tested for *Ph*CAO using a 7-OH Chlide *a* substrate. Here, it was determined that the Michaelis constant is 7.8 μM and the k_{cat}= 0.12 min^{-1} (Figure 4.23). Collectively, these results show that CAO is a Rieske oxygenase that performs sequential oxygenation reactions *in vitro* when combined with the annotated ferredoxin-NAD$^+$ Rieske reductase VanB.

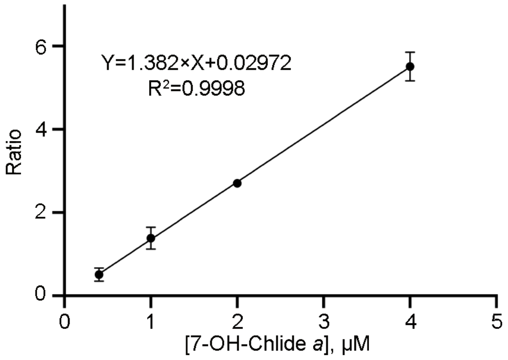


Figure 4.19. A standard curve for Q-TOF LC-MS was used to determine the amount of intermediate (7-OH-Chlorophyllide *a*) in the reactions.

The x-axis is the concentration of 7-OH-Chlorophyllide *a* whereas the y-axis is the ratio of the peak area integration of the 7-OH-Chlorophyllide *a* to the 1,3,5-trimethoxybenzene (0.5 mM) internal standard. All data points were measured in duplicate.

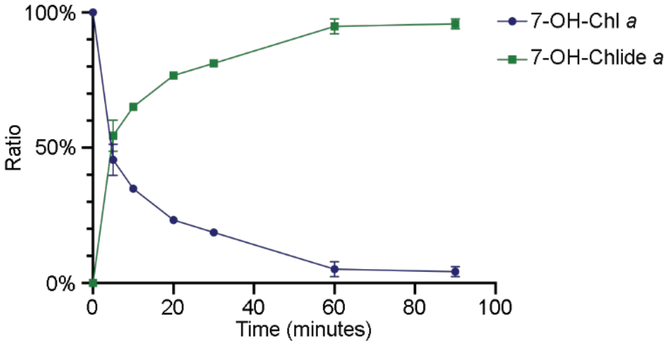


Figure 4.20. Chlorophyllase can be used as a tool to produce a 7-OH-Chlide *a* standard.
A time dependent measurement of 7-OH-Chl *a* (1 mM, blue) consumption and 7-OH-Chlide *a* production using chlorophyllase (1.5 µM, green) shows that the chlorophyllase reaction reaches the maximum amount of substrate conversion at 1 hour. At this time point, approximately 95-percent of the 7-OH-Chl *a* has been converted into 7-OH-Chlide *a*. All data points were measured in duplicate.

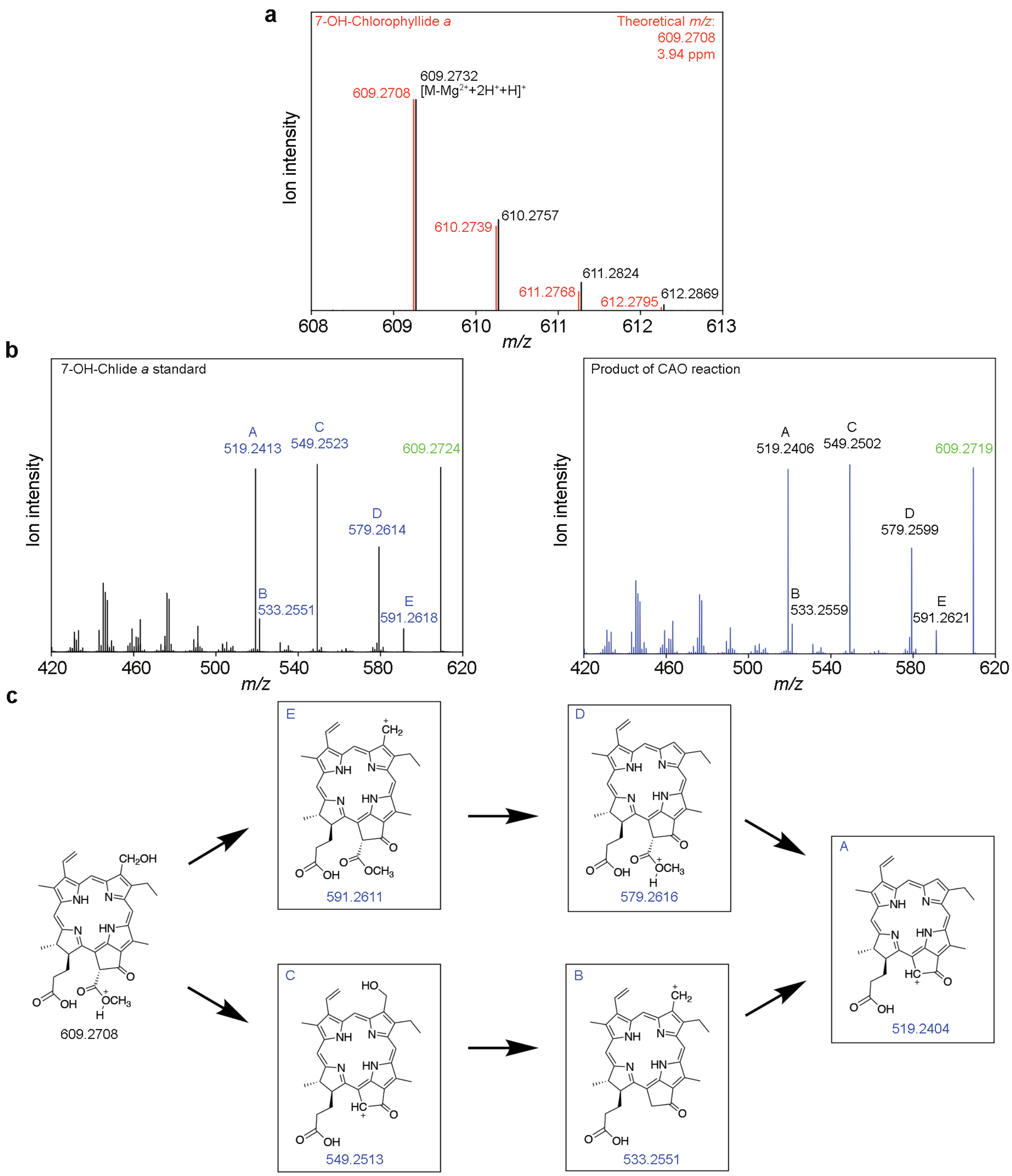


Figure 4.21. Analysis of the chemically produced 7-OH-Chlide *a* standard and the intermediate produced in the CAO reaction.

(a), The predicted (red) and observed (from the *Ph*CAO reaction, black) isotope distribution for 7-OH-Chlide *a* show good agreement. (b), The MS/MS spectrum of produced 7-OH-Chlide *a* standard (left panel) is compared to the MS/MS spectrum of the proposed 7-OH-Chlide *a* intermediate from the *Ph*CAO reaction (right panel). (c), Proposed fragmentation pathway for 7-OH-Chlide *a*. The existence of fragments D and E are consistent with the presence of a hydroxymethyl group in the compound being tested due to the mass loss of 18.0097 Da (H_2O) and 30.0092 Da (CH_2O).

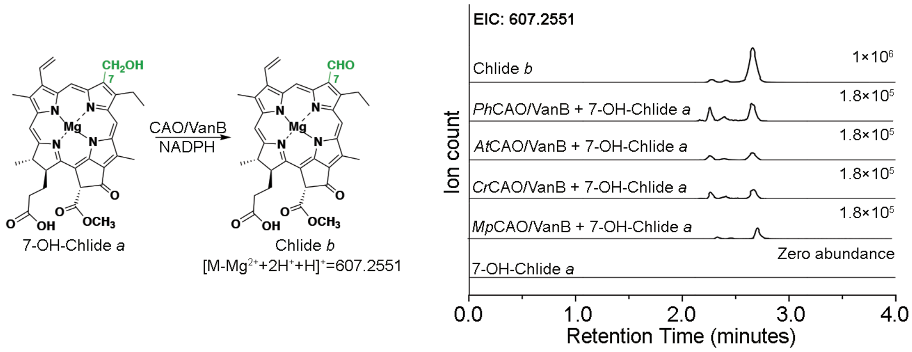


Figure 4.22. Extracted ion chromatograms for the CAO homolog reaction products when combined with VanB, NADPH, and a 7-OH-Chlide *a* substrate.
These traces reveal that 7-OH-Chlide *a* can be converted into Chlide *b*.

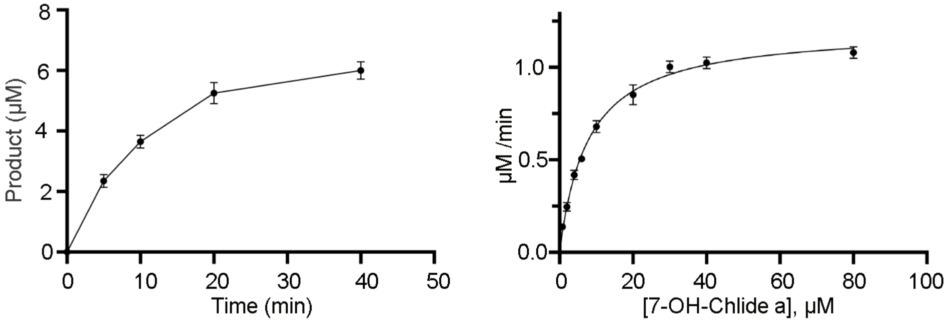


Figure 4.23. Steady-state kinetics study for CAO with 7-OH-Chlide *a*.
A time course of *Ph*CAO with a 7-OH-Chlide *a* substrate was measured to find the linear range of product formation (left panel). This data demonstrates that 5 min is an optimal time to capture the initial reaction rate. The steady-state kinetics of *Ph*CAO with 7 OH-Chlide *a* substrate was fitted to the Michaelis-Menten equation with K_M= 7.8 ± 0.9 μM and V_{max}=1.2 ± 0.1 μM min^{-1}(right panel). All data points in this figure were measured in duplicate.

4.3.6 Factors that dictate the substrate scope of CAO

As described above, none of the CAO homologs purified in this work are able to oxidize the C7-methyl group of Chl *a* into the formyl group of Chl *b* either in the presence or absence of cell lysate (Figure 4.4.d and Figure 4.24.a). Rather, each of the tested homologs shows

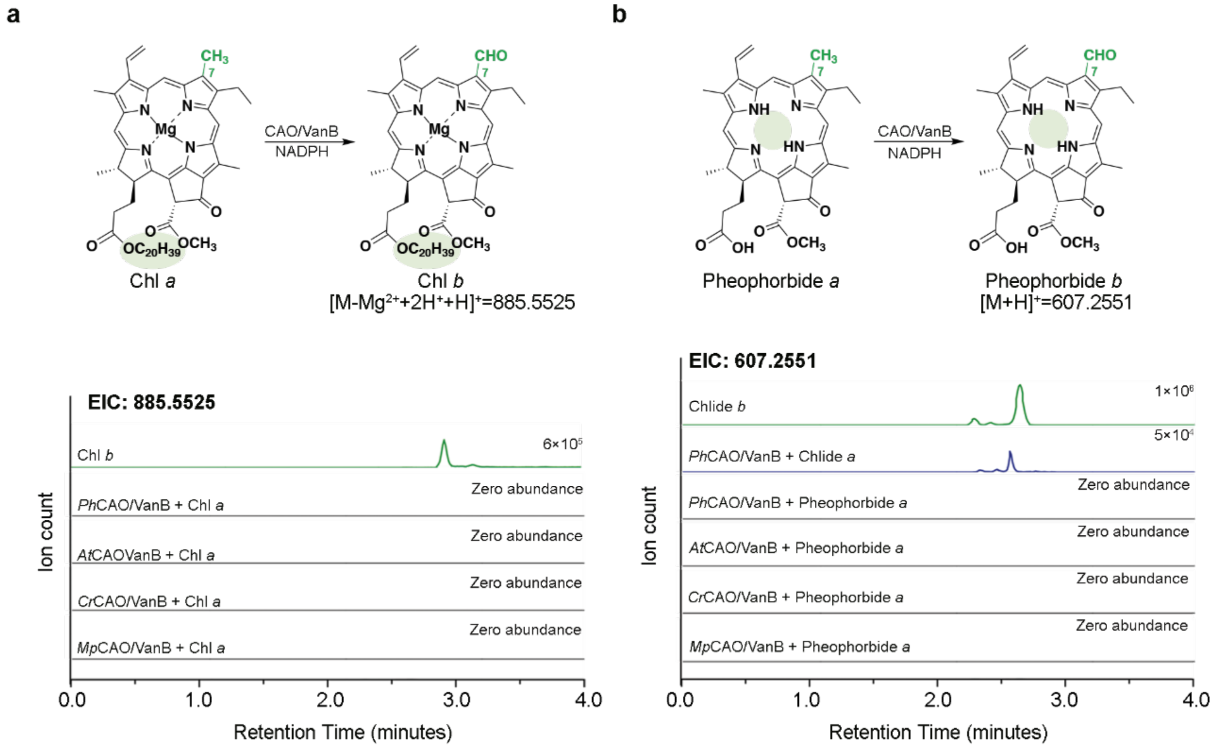


Figure 4.24. Chlorophyll *a* and Pheophorbide *a* are not substrates of the CAO homologs.
(a), An extracted ion chromatogram of the expected product of a CAO-VanB-catalyzed reaction with a Chlorophyll *a* substrate does not reveal production of a hydroxylated product, suggesting the long phytol group of Chlorophyll *a*, relative to Chlide *a*, impairs the activity of CAO. (b), Similarly, none of the CAO homologs show production of a hydroxylated product when provided with a Pheophorbide *a* substrate, suggesting that the central metal ion (Mg^{2+}) is also important for CAO activity.

a preference for a Chlide *a* or Chlide *a′* substrate. To further explore the scope of substrates accepted by CAO, several Chlide *a* analogs, including Pheophorbide *a*, bacteriochlorophyll (Bchl) *a*, bacteriochlorophyllide *a* (Bchlide *a*), and Chlide *d* were tested for their propensity to be oxidized by *Ph*CAO. Here, it was determined that Pheophorbide *a*, which lacks central Mg^{2+} ion of Chlide *a*, is not a substrate of *Ph*CAO (Figure 4.24.b). Similarly, Bchl *a*, which resembles Chl *a* but instead contains a bacteriochlorin scaffold and a C3-acetyl group, rather than a chlorin scaffold and a C3-vinyl group, is also not a substrate of *Ph*CAO (Figure 4.25). Bchlide *a*, produced via combination of commercially purchased Bchl *a* with chlorophyllase, was also not formylated by

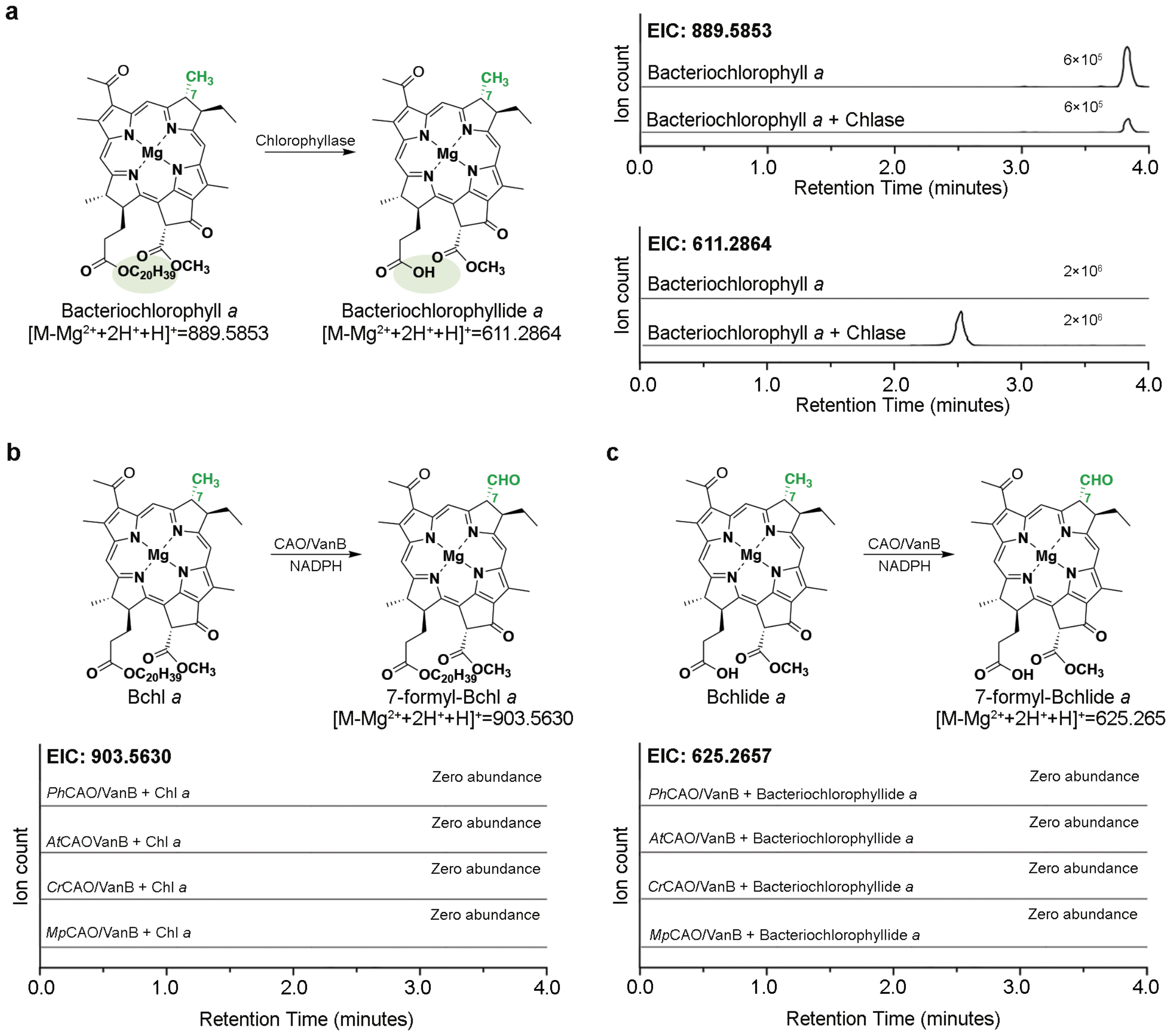


Figure 4.25. The activity of the CAO homologs was tested on Bacteriochlorophyll *a* and Bacteriochlorophyllide *a* substrates.

(a), Chlorophyllase was demonstrated to hydrolyze the phytol tail of Bacteriochlorophyll a to produce Bacteriochlorophyllide *a*. (b), The extracted ion chromatograms of the CAO-VanB reaction products when incubated with a Bacteriochlorophyll *a* substrate reveals that none of the CAO homologs accept Bacteriochlorophyll *a* as a substrate. (c), Similarly, the CAO homologs do not show activity on Bacteriochlorophyllide *a* using VanB as a reductase. Collectively, these data suggests that substrates that contain a bacteriochlorin scaffold cannot be oxidized by CAO.

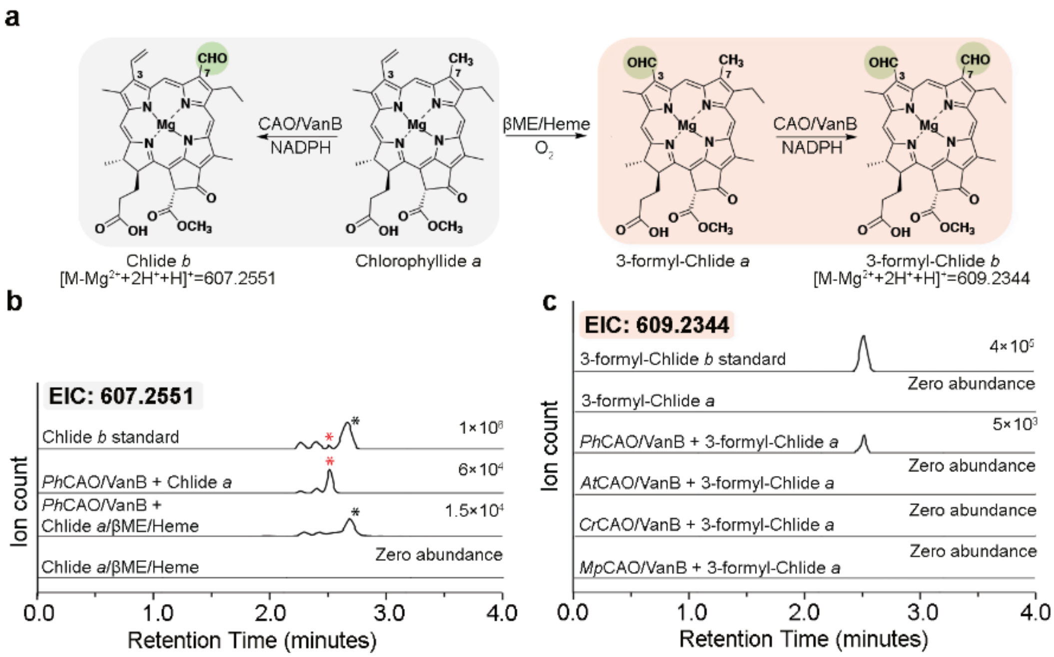


Figure 4.26. 3-formyl-Chlide *a* (Chlide *d*) can be transformed into 3-formyl-Chlide *b* by combination of the *Ph*CAO with VanB and NADPH.
(a), Reaction scheme to synthesize 3-formyl-Chlide *a* and the proposed route to C7-oxygenation by CAO. (b), The extracted ion chromatograms for the product formed when Chlide *a* or Chlide *a* with βME and heme was provided to *Ph*CAO as a substrate. (c), The extracted ion chromatograms show that *Ph*CAO can transform 3-formyl-Chlide *a* (Chlide *d*) into 3-formyl-Chlide *b*.

*Ph*CAO (Figures 4.25). Last, a report which showed that a doubly formylated Chl species, 7-formyl Chl *d*, could be produced by transformation of *Ph*CAO into the Chl *d* producer, *Acarychloris marina*, was investigated[33]. Specifically, as this *in vivo* experiment could not provide details about the order of the formylation reactions, it was explored whether CAO could formylate Chlide *d in vitro*. For this experiment, Chlide *a* was produced using chlorophyllase and then a recently described protocol[34] to produce Chl *d in vitro* was used to instead yield the desired Chlide *d* pigment as a mixture (Figure 4.26.a and Figure 4.27). When this reaction mixture was given to *Ph*CAO as a substrate, formation of both Chlide *b* and a second product was observed (Figure 4.26.b-c). This second product is present at relatively low yields but has a mass that is consistent with it being 3-formyl-Chlide *b* (Figure 4.26.c). To investigate the nature of this molecule,

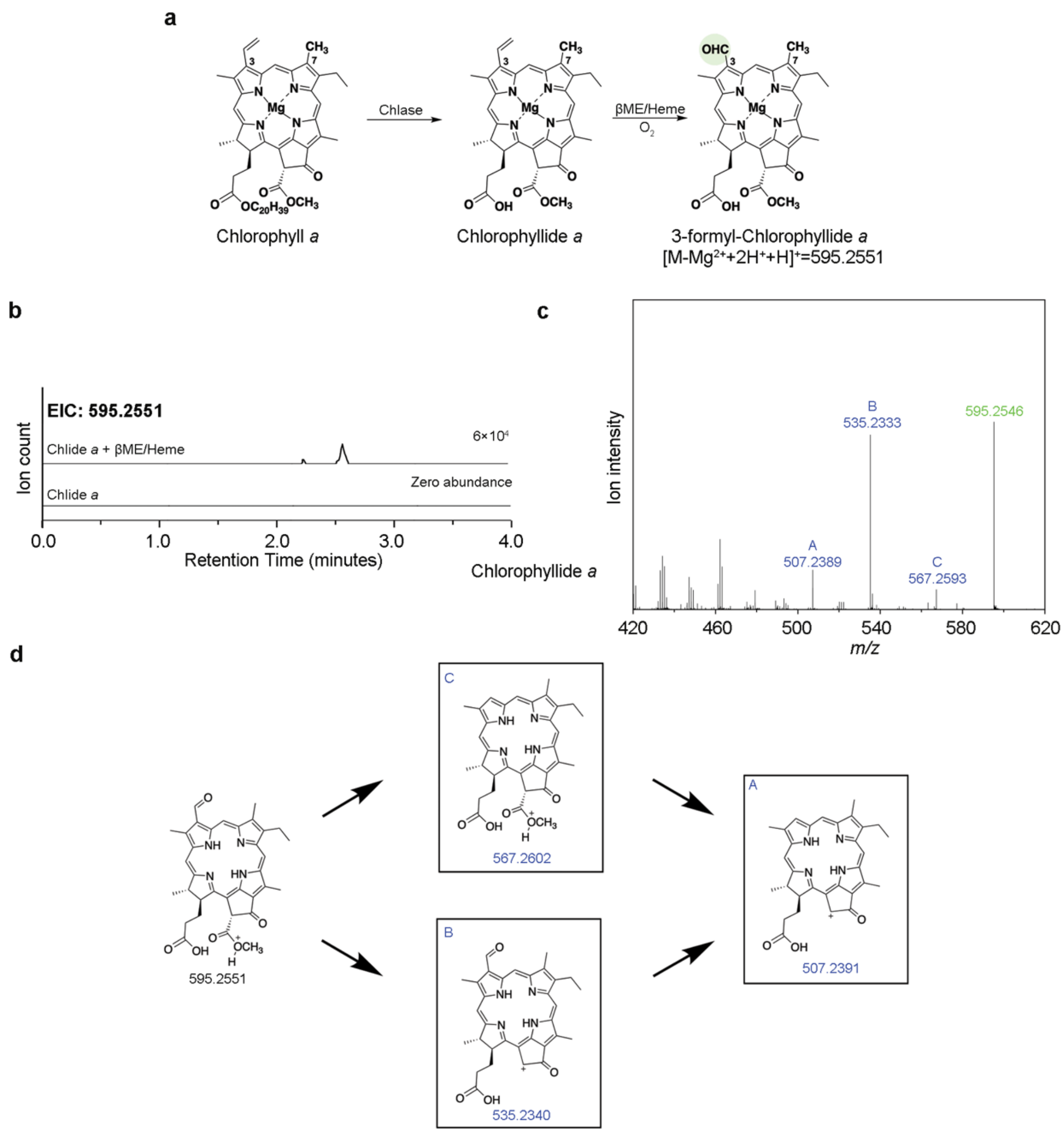


Figure 4.27. Analysis for the chemically produced 3-formyl-Chlide *a*.
(a), Reaction scheme of enzymatically and chemically synthesis for 3-formyl-Chlide *a*. (b), Extracted ion chromatogram of the product produced by following the published methods for using β-mercaptoethanol (βME) and heme to convert Chlide *a* into 3-formyl-Chlide *a*. (c), MS/MS spectrum 3-formyl-Chlide *a* shows three major fragments of interest, labelled as A, B and C. (d), A proposed fragmentation pathway for 3-formyl-Chlide *a*. Fragment C, which demonstrates a mass loss of 27.9949 Da corresponds to the loss of a formyl group.

a standard of 3-formyl-Chlide *b* was produced similarly to Chlide *d*, except that Chl *b* was used as a starting material (Figure 4.26.c). This standard showed the same retention time as the product made by *Ph*CAO and had a similar MS/MS fragmentation pattern, indicating that *Ph*CAO can accept a pigment that is formylated at the C3 position to produce a non-natural 3-formyl-Chlide *b* pigment (Figure 4.28).

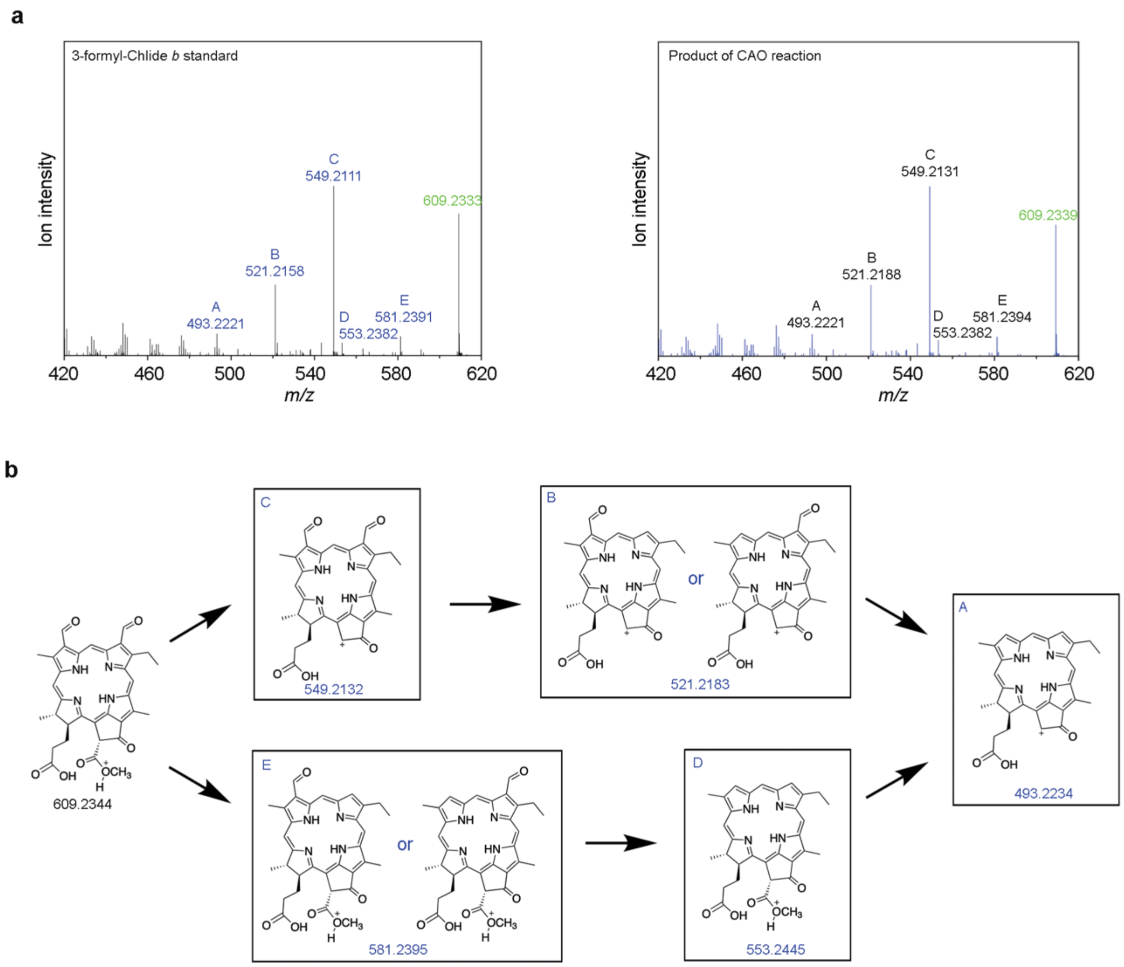


Figure 4.28. Analysis for the 3-formyl-Chlide *b* produced in the CAO reaction.
(a), The MS/MS spectrum of produced 3-formyl-Chlide *b* standard (left panel) is compared to the MS/MS spectrum of the 3-formyl-Chlide *b* from the *Ph*CAO reaction (right panel). (b), Proposed fragmentation pathway for 3-formyl-Chlide *b*. The existence of fragments D and E are consistent with the presence of two formyl groups in the compound being tested due to the mass loss of 27.9949 Da (CO) and 55.9899 Da (C_2O_2).

4.4 Discussion

Despite the identification of Chlorophyll(ide) *a* Oxygenase (CAO) over 20 years ago, the mechanism of the required methyl-to-formyl group transformation in Chlorophyll *b* biosynthesis has never been studied *in vitro* with purified protein. Traditionally, these studies have been hindered by the lack of published protocols for the recombinant expression and purification of a CAO homolog, the reported insolubility and reactive nature of the proposed substrates, and the lack of an annotated reductase for the reaction. To overcome these obstacles, in this work, a homogeneous and reconstituted "simple" CAO that lacks the accessory domains found in *A. thaliana* and *C. reinhardtii* homologs and carries both the Rieske cluster and non-heme iron site on a single polypeptide chain was recombinantly expressed and purified (Figure 4.2 and Figure 4.3). The successful methods for isolating *Ph*CAO were then extended to isolate other CAO homologs (Figure 4.3). Each of these purified CAO homologs were then shown to convert Chlide *a* into Chlide *b* rather than Chl *a* into Chl *b* in cell lysate and in the presence of the non-native Rieske reductase VanB (Figures 4.4 and 4.12). Thus, this work suggests that despite previous proposals, Chl *a* is not formylated by CAO even in the presence of other photosynthetic proteins that could function as carriers of Chl in the native organism[4, 35-36]. Likewise, the ability of chlorophyllase to cleave Chl into Chlide in our assays negates previous proposals[35-36] that the insolubility of Chl prohibits its ability to serve as an enzyme substrate.

This work also showed that all four CAO homologs, which based on sequence, are predicted to have different architectures[25], function similarly and can use a combination of NADPH and the same non-native reductase as a source of electrons. The ability of *Mp*CAO to accept electrons from VanB and produce Chlide *b* is particularly surprising as it has been previously suggested to exist as a heterodimer[25], rather than as a prototypical Rieske oxygenase trimer. This different quaternary architecture may explain the lower turnover rate of *Mp*CAO relative to the three other homologs studied in this work (Figure 4.2.a). Similarly, the presence of extra N-terminal regulatory domains in the *A. thaliana* and *C. reinhardtii* homologs, which have been suggested to sense Chl *b* accumulation and/or destabilize CAO[37], may be partially responsible for their lower observed *in vitro* activity relative to *Ph*CAO (Figure 4.2.a). Most notably, in this work, it was determined that these CAO-catalyzed reactions form a different diastereomer of Chlide *b* than what is found in the Chlide *b* standard (Figures 4.4 and 4.12). As the product formed can be converted into the standard by the addition of pyridine, we propose that the CAO product is natively Chlide *b′* (Figures 4.4 and 4.12). Curiously, it is also known that the last step of Chl synthesis requires the Chl synthetase-catalyzed attachment of the hydrophobic tail[38]. This protein, like chlorophyllase is known to show a preference for both Chlide *a* and Chlide *b* substrates, rather than Chlide *a*′ and Chlide b′[38]. The observation that cell lysate shows the remarkable ability to interchange diastereomer peaks may suggest that there are epimerase enzymes present that are involved in regulating which Chl and Chlide diastereomers are present in the cell (Figure 4.4.e).

This work also used MS experiments to reveal that 7-OH-Chlide *a*, is a true intermediate of the reaction and shows that CAO is a Rieske oxygenase that can facilitate the sequential oxygenation reaction in lieu of an additional Rieske protein, dehydrogenase, or cofactor (Figure 5). Using 7-OH-Chlide *a* as a substrate, it was determined that *Ph*CAO displays a K_M of 7.8 ± 0.9 μM and a k_{cat} of 0.12 min^{-1} (Figure 5h). This K_M is consistent with that determined for other Chl biosynthetic enzymes, including AcsF (7.0 μM)[39] and light-dependent protochlorophyllide oxidoreductase (8.6 μM)[40]. The k_{cat} for the second oxygenation reaction catalyzed by CAO, on the other hand, is lower than that recently described for AcsF (0.9 min^{-1})[39] and may reflect the complexity of the reaction, which requires chlorophyllase to cleave the phytol tail of the substrate and a non-physiological reductase to deliver electrons (Figure 4.223). In addition, it may correlate with providing the non-preferred substrate diastereomer in the assay or may suggest that as previously described for other Chl biosynthetic enzymes[41], the *in vivo* rate is amplified by protein-protein interactions with other pathway enzymes.

Last but not least, along with Chl *a*, several Chlide *a* analogs were used to test the substrate requirements of the CAO reaction. Chl *a*, Pheophorbide *a*, Bchl *a*, and Bchlide *a*, were each shown to not be formylated by CAO (Figure 4.24 and 4.25). These results suggest that the presence of a phytol side chain at the C17 position, the lack of a central metal ion, and the bacteriochlorin scaffold each effect the reactivity of CAO. We propose that the phytol chain and metal ion are important dictators of substrate binding in CAO. In addition, these data suggest that the electronics of the substrate may be significant for the CAO reaction. In the case of both Bchl *a* and Bchlide *a*, the B ring is oxidized by two electrons relative to Chl *a* and Chlide *a* (Figure 4.25). This structural feature means that for CAO to accept these molecules as a substrate, that a C7-substrate radical must be produced adjacent to an sp^3-, rather than an sp^2-hyridized carbon atom. We hypothesize that due to this difference, CAO is unable to abstract the needed hydrogen atom to initiate the formylation reaction. Protochlorophyllide (Pchlide) *a* also resembles Chlide *a*, but it has a 22-electron-containing scaffold. Conflicting previous experiments performed in cell lysate suggested that CAO either could[42] or could not[10] convert a small amount of this molecule into Pchlide *b*. Intriguingly, however, a second Rieske oxygenase, protochlorophyllide *a* oxygenase (PTC52), is instead credited for this transformation[42]. Most interestingly, it was determined here that 3-formyl-Chlorophyllide *a*, or Chlorophyllide *d*, can be accepted by *Ph*CAO and a second formyl group can be installed at the C7 position to produce 3-formyl-Chlide *b* (Figure 4.26). Collectively, these results highlight the remarkable flexibility of the active site for binding pigment substrates, that lack a phytol tail and contain a central metal ion and a chlorin scaffold.

Therefore, this work adds to our knowledge of chlorophyll biosynthesis, extends the known reactivity of the Rieske oxygenase class, and provides a framework for developing the CAO-VanB system as a tool to produce non-native Chl pigments (Figure 4.1.b). The procedures established in this work for expressing, purifying, and reconstituting four CAO homologs work can now be used to facilitate future structural and biochemical studies on CAO. Likewise, the procedures developed for production of the desired Chlide molecules and bottom-up approach for measuring CAO activity have revealed intriguing details regarding the stereoselectivity and substrate scope of the reaction. Future work will be aimed at pinpointing key amino acids involved in the transformation of Chlide *a* into Chlide *b*, optimizing the partner reductase used in the reaction, identifying the component(s) in cell lysate that change the stereochemistry of Chlide, and increasing the yields of native and non-native pigments. This works further elucidates the sequential oxidation mechanism performed by Rieske oxygenases as well as demonstrates the ability to produce non-native pigments that have further implications in industrial applications, such as cosmetic pigments or food dyes. From a biosynthetic standpoint, it suggests that these enzymes can be further engineered to allow for greater reactivity/selectivity that would play a greater role in the overall understanding of photosynthesis in various organisms.

4.5 Methods

4.5.1 Protein production and purification

Protein expression protocols for MBP-tagged *Ph*CAO

Small-scale expression tests on an N-terminal 6X-His-tagged construct of *Ph*CAO in different *E. coli* strains (BL21(DE3), C41(DE3) and *ΔiscR* Bl21-(DE3)) revealed no significant overexpression under different levels of Isopropyl β-D-1-thiogalactopyranoside or at different temperatures. A C-terminal 6X-His-tagged construct was subsequently tested and shown to yield better overexpression of the target *Ph*CAO protein in the *ΔiscR* Bl21-(DE3) cell line. However, analysis by gel filtration chromatography suggested that this purified protein existed predominantly as an aggregate. Thus, a new codon optimized construct was cloned into the pMCSG9 vector (Genscript). This construct contains an N- terminal His-MBP tag and was tested for expression in the *E. coli* strains listed above. The pMCSG9 plasmid containing *Ph*CAO was transformed by standard heat-shock protocols into C41(DE3) chemically competent E. coli cells, which showed the highest level of over-expression. A 5 mL starter was grown from one colony in Luria Broth (LB) containing 50 µg/mL ampicillin in a constant temperature incubator at 37°C and 200 rpm. After an overnight incubation, the starter culture was used to inoculate larger 1 L Terrific Broth (TB) cultures containing 50 µg/mL ampicillin. These 1 L cultures were grown at 37°C and shaken at 200 rpm until the optical density at 600 nm (OD_{600}) reached a value of approximately 0.8-1.0. At this point, flasks were incubated at 20°C, left for 2 hours, and induced by the addition of 0.2 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), 0.2 mg/mL ferric ammonium citrate, and 0.4 mg/mL ferrous sulfate heptahydrate. The temperature of the incubator was then decreased to 18°C and left for an additional 16 h before harvesting.

Protein expression protocols for His-tagged *At*CAO and His-tagged *Cr*CAO

As described for *Ph*CAO, both *At*CAO and *Cr*CAO were codon optimized, synthesized, and cloned into pET-41b(+) plasmids by Genscript. These *At*CAO and *Cr*CAO plasmids were transformed by standard heat shock protocols into Δ*iscR* Bl21-(DE3) chemically competent cells and plated on LB plates containing 50 μg/mL kanamycin. 5 mL starter cultures were grown and used to inoculate 1 L TB cultures containing 50 μg/mL kanamycin. Expression of both *At*CAO and *Cr*CAO was induced as described for *Ph*CAO.

Protein Expression Protocols for *Mp*CAO1 and *Mp*CAO2

A pRSFDuet-1 plasmid containing codon optimized *Mp*CAO1 and *Mp*CAO2 genes was transformed by standard heat-shock protocols into C41(DE3) chemically competent cells. Starters were grown as described for *Ph*CAO and subsequently used to inoculate 1 L cultures of LB containing 50 μg/mL kanamycin. Cultures were grown at 37°C, 200 rpm until the OD_{600} reached 0.6-0.8. Following the addition of 0.2 mM isopropyl-β-D-1- thiogalactopyranoside (IPTG), Flasks were then incubated at 20°C for 16 hours before harvesting.

Protein Expression Conditions for the reductase proteins VanB, DdmB, DdmA, TsaB, and *E. coli* flavodoxin (Flv) and *E. coli* flavodoxin (Flx) reductase.

The methods for purifying VanB, DdmB, DdmA, and *E. coli* Flv and Flx were previously described and were followed here[22,43]. In brief, these methods require a pMCSG7-*vanB*, pMCSG7-*ddmB*, pMCSG7-*ddmA*, pET28a-*flv*, and pET28a-*flx* plasmids, as well as BL21(DE3) (VanB, Flv, and Flx) and C41(DE3) (DdmB and DdmA) *E. coli* cells. TsaB, on the other hand, was codon optimized, synthesized, and cloned into a pet-28a(+)-TEV plasmid by Genscript. This pET28a(+)-

TEV-tsaB plasmid was transformed into BL21(DE3) *E. coli* cells. TsaB was then grown and expressed in cultures of TB containing 50 µg/mL kanamycin until the OD_{600} reached 0.6-0.8. At this point, expression was induced with 0.1 mM IPTG. These cultures were then left at 20°C for 16 hours.

Protein Expression Protocols for *T. aestivum* Chlorophyllase

A pET21d plasmid containing an N-terminal 6X-His-tagged *T. aestivum* chlorophyllase was synthesized and codon-optimized for *E. coli* expression by Genscript. However, this initial plasmid did not result in overexpression of the chlorophyllase protein. Thus, the codon-optimized chlorophyllase gene was subcloned into a pET21d plasmid that contained a C-terminal 6X-His-tag. This new construct was transformed into BL21(DE3) *E. coli* competent cells. To produce chlorophyllase protein, a single colony of the transformant was grown in LB medium that contained 50 µg/mL of ampicillin at 37°C. When the OD_{600} of this culture reached 0.8, protein expression was induced with 0.5 mM IPTG for 16 hours at 16°C.

Protein Purification Protocols for His-MBP-tagged *Ph*CAO

The cell pellet produced from 1 L of culture that contained the over-expressed His-MBP-tagged *Ph*CAO was resuspended in 50 mL of lysis buffer (50 mM Tris-HCl (pH 7.5), 200 mM NaCl, and 5% glycerol). Resuspended cells were lysed by sonication and clarified by centrifugation in an Eppendorf centrifuge 5810R at 12,000 x g for 40 min. The supernatant, which contained the MBP-tagged protein *Ph*CAO, was loaded onto 5 mL MBP-Trap column (Cytiva). This column was washed with 10 column volumes of Buffer A (50 mM Tris-HCl (pH 7.5), 200 mM NaCl, and 5% glycerol) and eluted with a five-column volume gradient of Buffer B (50 mM

Tris-HCl (pH 7.5), 200 mM NaCl, 5% glycerol, and 10 mM maltose). Fractions that bound to column that contained the MBP-*Ph*CAO were pooled and dialyzed into Buffer C (20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, and 10% glycerol). Following buffer exchange, the His-MBP tag from the PhCAO protein was cleaved using 5 mg of Tobacco Etch Virus (TEV) protease. Following a 20 h incubation at 4°C, the tag cleavage reaction was complete. The tag-free protein was loaded onto a 5mL His-Trap (Cytiva) for further purification and removal of TEV protease and the cleaved His-MBP tag. The flow-through from the column, which contained the desired tag-free protein was concentrated and loaded onto a HiPrep 16/60 Sephacryl S200-HR (Cytiva) gel filtration column that was pre-equilibrated in Buffer D (50 mM HEPES (pH 8.0), 200 mM NaCl, and 10% glycerol).

Protein Purification Protocols for *At*CAO and *Cr*CAO

The cell pellet produced from 1 L of culture that contained the over-expressed 6X-His-tagged *At*CAO or *Cr*CAO was resuspended in Buffer A (50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 10 mM imidazole, and 5% glycerol). Cells were lysed by sonication, centrifuged, and loaded onto 5 mL His-Trap column. This column was washed with 10 column volumes of Buffer A and eluted with a five-column volume gradient of Buffer B (50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 200 mM imidazole, and 5% glycerol). Fractions containing the *At*CAO or *Cr*CAO were pooled and exchanged into storage buffer (50 mM HEPES (pH 8.0), 200 mM NaCl, and 10% glycerol) using a PD-10 desalting column (BioRad) to remove imidazole.

Protein Purification Protocols for *Mp*CAO1 and *Mp*CAO2

The cell pellet produced from a 2 L culture of *Mp*CAO1 and *Mp*CAO2 was resuspended in 50 mL lysis buffer (50 mM HEPES (pH 7.5), 250 mM NaCl, and 5% glycerol). Cells were lysed by sonication and centrifuged. The supernatant, which contained the His-tagged proteins *Mp*CAO1 and *Mp*CAO2 was loaded onto a 5 mL His- Trap column. This column was performed, and the protein was stored as described for *At*CAO and *Cr*CAO.

Purification protocol for the reductase proteins VanB, DdmB, DdmA, TsaB, and *E. coli* Flv and Flx

The methods for purifying VanB, DdmB, DdmA, were previously described and were followed here[22]. In short, the cells from 2 L cultures were resuspended in 50 mL of Buffer A (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole, and 10% glycerol). These proteins were then lysed, centrifuged, and purified using nickel affinity and gel filtration chromatography. *E. coli* Flv and Flx were purified using a similar protocol with small deviations. Specifically, the Flv and Flx buffers, as previously described[43], contained 20 mM Tris–HCl (pH 8.5), 1 M NaCl, 10 μM FMN (Flv) or 10 μM FAD (Flx), and either 20 mM or 200 mM imidazole, Purified proteins were concentrated to approximately 200 μM and flash frozen by liquid nitrogen and stored at -80°C.

Protein Purification Protocol for *T. aestivum* Chlorophyllase

The cell pellet harvested after overexpressing chlorophyllase was resuspended in Buffer A (50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 5 mM imidazole). The cells were lysed by sonication, centrifuged, and loaded onto 5 mL Ni-NTA column. Once the supernatant was all loaded, Buffer A was used to wash the column for 10 column volumes. C-terminal His-tagged chlorophyllase was eluted using the buffer B containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 200 mM imidazole. The purified Chlorophyllase was concentrated to 5 ml and was loaded onto Superdex 200 size exclusion column that was pre-equilibrated with buffer C containing 50 mM HEPES pH 7.5 and 100 mM NaCl. Chlorophyllase was eluted as a dimer in the size exclusion column with Buffer C. Chlorophyllase was concentrated to around 60 μM and flash frozen using liquid nitrogen for long-term storage in -80°C freezer.

4.5.2 Preparation of substrate and product standards

Chlorophyllase reaction to produce Chlorophyllide species

Purified His-tagged chlorophyllase (chlase) was stored in the -80°C freezer at a stock concentration of 30 μM. To perform a reaction, 2 μM Chlase was combined with 160 μM of the desired chlorophyll molecule in DMSO (Chl *a*, Chl *b*, 7-OH-Chl *a*, and Bchl *a*), and 50 mM Tris-HCl (pH 7.2). Reactions were initiated by addition of a Chl molecule and incubated at 30°C for 2 h in the dark. Reactions were quenched by the addition of 150 μL acetonitrile (HPLC-grade). Quenched reactions were centrifuged at 17,000 x g for 10 min. 50 μL of the supernatant was diluted with 150 μL acetonitrile containing 0.5 mM 1,3,5- trimethoxy-benzene as internal standard.

Preparation of 7-OH-Chlorophyllide *a*

The preparation of 7-OH-Chlorophyllide *a* was accomplished in two steps. The first step, production of 7-OH-Chlorophyll *a*, was performed using previously described methods[31]. In short, 1 mg of Chlorophyll *b* from spinach (Sigma Aldrich) was dissolved in 5 mL of methanol containing 1.0 mg of $NaBH_4$. The reaction was stirred for 2 min at room temperature and then quenched by the addition of 5 ml solution of saturated NaCl. The reaction products were then transferred into dichloromethane (DCM) and dried by anhydrous sodium sulfate. A rotary evaporator was then used to remove extra solvent. The second step was to convert the produced 7-OH-Chlorophyll *a* into 7-OH- Chlorophyllide *a*. This step was accomplished using purified chlorophyllase and the methods described above.

Preparation of Chlorophyllide *a′*

To prepare Chlorophyllide *a′*, we capitalized on a previous method[27] and made small modifications. In short, Chlorophyllide *a′* was prepared by dissolving 1 mg of Chlorophyll *a* (Frontier Scientific) in 5 mL of diethyl ether. To this solution 50 µL of pyridine was added to a final volume of 1%. The reaction was stirred at room temperature for 1.5 hours before removal of the solvent using a rotary evaporator. The resulting liquid containing both diastereomers of Chlorophyllide *a* was then dissolved in 200 µL of DMSO for storage. A second step that employed chlorophyllase and the methods described above was followed to produce Chlorophyllide *a′*.

Preparation of Chlorophyllide *b′*

The preparation of Chlorophyllide b¢ was accomplished using 1 mg of Chlorophyll *b* (Sigma Aldrich). The protocol followed was like that described previously[27], and for Chlorophyllide *a′*. First, Chlorophyll *b* was dissolved in 1 mL of diethyl ether. Then 200 µL of this solution was added into an additional 1.6 mL of diethyl ether and 15% pyridine. The reaction was stirred at room temperature for 14 h before removal of the solvent via using a rotary evaporator. The resulting liquid containing both diastereomers was then dissolved in 200 µL of DMSO for storage. Finally, chlorophyllase was used to complete production of Chlorophyllide *b′*.

Preparation of 3-formyl-Chlorophyllide *a* and 3-formyl- Chlorophyllide *b*

To prepare 3-formyl-Chlorophyllide *a* and 3-formyl-Chlorophyllide *b*, we capitalized on a previous method (DOI: 10.1038/srep06069) and made small modifications. Briefly, 1 mM Chlorophyll *a* or Chlorophyll *b* was mixed with 10 μM chlorophyllase and incubated at 30°C for 1 h to produce Chlorophyllide a. 5 μL of β-mercaptoethanol (βME) and 200 μM heme chloride were then added into this reaction mixture for another 4-hour incubation at 30°C. The resulting liquid containing a mixture of Chlorophyllide *a* and 3-formyl-Chlorophyllide *a* or a mixture of Chlorophyllide *b* and 3-formyl-Chlorophyllide *b* was then fast-frozen by liquid nitrogen for storage for future assays.

4.5.3 Enzymatic reactions

CAO reactions in cell lysate

Stocks of each CAO homologs were prepared from purified (and frozen) stocks of *Ph*CAO, or His-tagged *At*CAO, *Cr*CAO and *Mp*CAO at a concentration of 100 μM. *A. thaliana*, spinach, green algae, barely and *E. coli* cell lysate (GenLysate, GBiosciences) were reconstituted by adding 100 μL MQ water, respectively. Reactions consisting of 20 μM CAO homolog, 10 μL lysate solution, 1 μM Chlase, 100 μM Chl a, 500 μM NADPH, 200 μM $Fe(NH_4)_2(SO_4)_2$ and 50 mM Tris-HCl pH 7.2 were mixed and incubated at 30°C for 3 hours in the dark. These reactions were then quenched by the addition of 150 μL acetonitrile. Quenched reactions were centrifuged at 17,000 x g for 10 min and 50 μL of the supernatant was diluted with 150 μL acetonitrile containing an internal standard (0.5 mM 1,3,5-trimethoxy-benzene as internal standard). Equivalent reactions were performed with Spinach, Barley, Green Algae, and *E. coli* lysate (GenLysate, GBiosciences).

CAO reactions with non-native reductases TsaB, DdmA/B, VanB, Spinach ferredoxin/ferredoxin reductase and *E. coli* flavodoxin/flavodoxin reductase

Stocks of each CAO homologs were prepared from purified (and frozen) stocks of *Ph*CAO, or His-tagged *At*CAO, *Cr*CAO and *Mp*CAO at a concentration of 100 μM. The ability of VanB to reduce the Fe-S cluster of PhCAO was tested by titrating 1.5 equivalents of VanB into a sample that contains 45 μM of *Ph*CAO with 1 mM NADPH at room temperature. The UV-Vis spectrum was measured by NanoDrop (ThermoScientific). Reactions consisting of 20 μM CAO, 40 μM VanB, 1 μM Chlase, 100 μM Chl *a* or 80 μM 7-OH-Chlorophyllide *a*, 500 μM NADPH, 200 μM $Fe(NH_4)_2(SO_4)_2$, and 50 mM Tris-HCl pH 7.2 were mixed and incubated at 30°C in the dark. After 3 h, these reactions were quenched by the addition of 150 μL acetonitrile. Quenched reactions were centrifuged at 17,000 x g for 10 min and 50 μL of the supernatant was diluted with 150 μL acetonitrile containing 0.5 mM 1,3,5-trimethoxy-benzene as internal standard. Equivalent reactions were performed with 40 μM of TsaB, 32 μM DdmA/DdmB, 20 μM *E. coli* flavodoxin/60 μM flavodoxin reductase, or 20 μM spinach ferredoxin/60 μM ferredoxin reductase (Sigma Aldrich). However, it was determined that VanB provided the highest level of activity for the assays. In addition, equivalent reactions were also performed with Chl *a*, Bchl *a* Bchlide *a*, 3-formyl-Chlide *a* (Chlide *d*), and Pheophorbide *a*.

LC-MS and MS/MS analysis

Liquid chromatography-mass spectrometry (LC-MS) analysis was performed on an Agilent G6545A quadrupole-time of flight (Q-TOF), or an Agilent 6230 time of flight (TOF) mass spectrometer equipped with a dual AJS ESI source and an Agilent 1290 Infinity series diode array detector, autosampler, and binary pump. Solvent A contained water with 0.1% formic acid. Solvent

B contained 95% acetonitrile, 5% water and 0.1% formic acid. For all separations, a ZORBAX RRHT StableBond Aq (2.1 x 50 mm, 1.8 µm, 80Å) column from Agilent was used. The employed chromatographic method used was made up of three steps: (step i) 40% Solvent B 0-1 min, (step ii) a linear gradient to 100% Solvent B over 1 min, and (step iii) 100% Solvent B for 2 min. 2.0 µL injections were made for each sample and the column was run at 0.4 mL/min.

Steady-state kinetics of *Ph*CAO with 7-OH-Chlorophllide *a*

To determine the linear range for kinetic analysis, duplicate 50 µL reactions with 10 µM *Ph*CAO and 10 µM 7-OH-Chlorophyllide *a* were performed and analyzed compared to an internal standard (0.5 mM 1,3,5-trimethoxy-benzene) by LC/MS. A 5 min time point was demonstrated to be the optimal time to capture the initial reaction rate.

To determine the steady-state kinetic parameters of *Ph*CAO with a 7-OH- Chlorophyllide *a* substrate, reactions were conducted on a 50 µL scale with substrate ranging 0.4 µM – 80 µM in duplicate with 10 µM *Ph*CAO, 15 µM VanB, 500 µM NADPH, 200 µM $Fe(NH_4)_2(SO_4)_2$ and 50 mM Tris-HCl (pH 7.2) buffer. Prior to reaction initiation, 1 mM 7-OH-Chlorophyll *a* was incubated with 5 µM chlorophyllase at 30°C for 2 h to generate the needed substrate, 7-OH-Chlorophyllide *a*. CAO reactions were then initiated by the addition of freshly made 7-OH-Chlorophyllide *a* and quenched after 5 min by the addition of 150 uL acetonitrile. Quenched reactions were centrifuged at 17,000 x g for 10 min and 50 µL of the supernatant was diluted with 150 µL acetonitrile containing 0.5 mM 1,3,5-trimethoxy-benzene as an internal standard. 2 uL each sample was injected on the Q-TOF LC/MS. The result data points were then plotted and fitted into Michaelis-Menten curve using GraphPad Prism.

4.6 DNA and protein sequences

4.6.1 Sequences of CAO homologs

Prochlorothrix hollandica Chlorophyllide _a_ oxygenase (_Ph_CAO):

ATGAATAATAGCCTCAACGTATCCGCAACCCTAGATTTAGCGAACGGATTACGCAACTTTTGGT
ATCCCGTTGAGTTTTCTAAGAATTTGGGGATGGCGGATCCCCTGGGCTTCGAGTTATTTGATCA
ATGTTGGGTTTTGTTTCGCGATGATCAGGGAACTGCCGCCTGTATTCTGGATGAATGTGCCCAC
CGCGCTTGTCCCCTATCCCTGGGCAAGGTGATCCAAGGGCGGATCCAGTGCCCCTACCATGGTT
GGGAGTACGATCGCCAGGGGGAGTGTGTCCATATGCCCTCCTGCCAAGCCATCAGCAATCCAAT
TTTGACCTTGCCAGTGATGGAGCAAGGGGGCATGATTTGGGTTTGGCCCGGAACCGACGAGCCA
GGTGCCTTGCCATCCTTAGCCCCCACCCTACCGGACAATTTCACCCTCCAGGCTGAATTGGTGA
TGGATTTGGAGGTGGAGCATGGCTTAATGCTGGAAAACCTCTTAGATCTCGCCCATGCCCCCTT
TACCCACACCGGCACCTTTGCCAAGGGCTGGCCCGTCCCCCCTTTTGTCCGGTTTGCCAATGCT
GCCACCACACCCTGGACCGGGCACTGGGATCCCTATCCCATCCACATGACCTTTGAACCCCCTT
GCTTTGTCATTAGCACCATCGGCTTGCGGGGCAAAGACTGTGGCCGCCATCTGCATCAGGTTCA
CGCCTGTCTCCCTCGGGGCCAGGGGCGCACTCGCTTGCTCTACCGCTTAGCCTTAGATTTTGGC
CACTGGTTACGCTGGGTTCCGGGGACCCATTGTCTCTGGCAACATTTGGCCAACCGAGTCATCC
AGGAAGACTTACGCTTAGTCCAGGGCCAACAGGAGCGTCTCAAGGGGGGGGCTAATGTGTGGAA
CCAACCCGTGGGCTATGACAAGTTGGGGGTGGCCTATCGCCACTGGCGCAATCAAGTGGAACGC
CACGGTTCTGACTGGCCTGAATCTCCCGCTGATGAGGGTAGGGAGCCTGCTCTCAATGCGATCG
TCACGGGGTCAGATGCTCCGATCACGGGTTCTGTGGTTTCCCTGCCGCCGTCCCAGGCTCCTCC
CACGGGGCACTGA

MNNSLNVSATLDLANGLRNFWYPVEFSKNLGMADPLGFELFDQCWVLFRDDQGTAACILDECAH
RACPLSLGKVIQGRIQCPYHGWEYDRQGECVHMPSCQAISNPILTLPVMEQGGMIWVWPGTDEP
GALPSLAPTLPDNFTLQAELVMDLEVEHGLMLENLLDLAHAPFTHTGTFAKGWPVPPFVRFANA
ATTPWTGHWDPYPIHMTFEPPCFVISTIGLRGKDCGRHLHQVHACLPRGQGRTRLLYRLALDFG
HWLRWVPGTHCLWQHLANRVIQEDLRLVQGQQERLKGGANVWNQPVGYDKLGVAYRHWRNQVER
HGSDWPESPADEGREPALNAIVTGSDAPITGSVVSLPPSQAPPTGH

Chlamydomonas reinhardtii Chlorophyllide _a_ oxygenase (_Cr_CAO):

ATGCTGCCGGCGAGCCTGCAGCGTAAAGCGGCGGCGGTTGGTGGCCGTGGTCCGACCAACCAGA
GCCGTGTGGCGGTTCGTGTGAGCGCGCAACCGAAAGAGGCGCCGCCGGCGAGCACCCCGATCGT
GGAGGATCCGGAAAGCAAGTTCCGTCGTTACGGTAAACACTTTGGTGGCATTCACAAGCTGAGC
ATGGACTGGCTGGATAGCGTTCCGCGTGTTCGTGTGCGTACCAAAGACAGCCGTCAGCTGGACG
ATATGCTGGAGCTGGCGGTGCTGAACGAACGTCTGGCGGGTCGTCTGGAGCCGTGGCAGGCGCG
TCAAAAGCTGGAATACCTGCGTAAGCGTCGTAAAAACTGGGAGCGTATCTTCGAATATGTTACC
CGTCAAGATGCGGCGGCGACCCTGGCGATGATTGAGGAAGCGAACCGTAAGGTGGAGGAAAGCC
TGAGCGAGGAAGCGCGTGAGAAAACCGCGGTTGGTGACCTGCGTGATCAGCTGGAAAGCCTGCG

TGCGCAAGTGGCGCAGGCGCAAGAGCGTCTGGCGATGACCCAGAGCCGTGTGGAACAGAACCTG
CAACGTGTTAACGAGCTGAAGGCGGAAGCGACCACCCTGGAGCGTATGCGTAAGGCGAGCGACC
TGGATATCAAAGAGCGTGAACGTATCGCGATTAGCACCGTTGCGGCGAAAGGTCCGGCGAGCAG
CAGCAGCAGCGCGGCGGCGGTGAGCGCGCCGGCGACCAGCGCGACCCTGACCGTTGAGCGTCCG
GCGGCGACCACCGTGACCCAAGAAGTTCCGAGCACCAGCTACGGTACCCCGGTTGATCGTGCGC
CGCGTCGTAGCAAAGCGGCGATCCGTCGTAGCCGTGGCCTGGAGAGCAGCATGGAAATTGAGGA
AGGTCTGCGTAACTTCTGGTATCCGGCGGAATTTAGCGCGCGTCTGCCGAAGGACACCCTGGTG
CCGTTCGAGCTGTTTGGCGAACCGTGGGTTATGTTCCGTGATGAGAAGGGTCAGCCGAGCTGCA
TTCGTGATGAATGCGCGCACCGTGGCTGCCCGCTGAGCCTGGGTAAAGTGGTTGAGGGCCAAGT
TATGTGCCCGTACCACGGTTGGGAATTCAACGGTGATGGTGCGTGCACCAAGATGCCGAGCACC
CCGTTTTGCCGTAACGTTGGCGTGGCGGCGCTGCCGTGCGCGGAGAAAGACGGTTTTATCTGGG
TGTGGCCGGGTGATGGCCTGCCGGCGGAAACCCTGCCGGACTTTGCGCAGCCGCCGGAGGGCTT
TCTGATCCACGCGGAAATTATGGTTGATGTGCCGGTTGAGCACGGTCTGCTGATTGAAAACCTG
CTGGATCTGGCGCATGCGCCGTTCACCCACACCAGCACCTTTGCGCGTGGCTGGCCGGTGCCGG
ACTTCGTTAAGTTTCACGCGAACAAAGCGCTGAGCGGTTTCTGGGACCCGTACCCGATCGATAT
GGCGTTTCAGCCGCCGTGCATGACCCTGAGCACCATCGGTCTGGCGCAACCGGGCAAGATTATG
CGTGGTGTTACCGCGAGCCAGTGCAAAAACCACCTGCACCAACTGCACGTGTGCATGCCGAGCA
AGAAAGGTCACACCCGTCTGCTGTATCGTATGAGCCTGGATTTCCTGCCGTGGATGCGTCACGT
GCCGTTTATCGACCGTATTTGGAAGCAGGTGGCGGCGCAAGTTCTGGGCGAGGATCTGGTGCTG
GTTCTGGGTCAGCAAGACCGTATGCTGCGTGGTGGCAGCAACTGGAGCAACCCGGCGCCGTACG
ACAAACTGGCGGTGCGTTATCGTCGTTGGCGTAACGGCGTTAACGCGGAAGTGGCGCGTGTTCG
TGCGGGTGAACCGCCGAGCAACCCGGTGGCGATGAGCGCGGGCGAGATGTTCAGCGTTGACGAA
GACGATATGGATAACAGCAGCGGTGAAAACCTGTATTTTCAGGGTAGCGGCAGCAGC

MLPASLQRKAAAVGGRGPTNQSRVAVRVSAQPKEAPPASTPIVEDPESKFRRYGKHFGGIHKLS
MDWLDSVPRVRVRTKDSRQLDDMLELAVLNERLAGRLEPWQARQKLEYLRKRRKNWERIFEYVT
RQDAAATLAMIEEANRKVEESLSEEAREKTAVGDLRDQLESLRAQVAQAQERLAMTQSRVEQNL
QRVNELKAEATTLERMRKASDLDIKERERIAISTVAAKGPASSSSSAAAVSAPATSATLTVERP
AATTVTQEVPSTSYGTPVDRAPRRSKAAIRRSRGLESSMEIEEGLRNFWYPAEFSARLPKDTLV
PFELFGEPWVMFRDEKGQPSCIRDECAHRGCPLSLGKVVEGQVMCPYHGWEFNGDGACTKMPST
PFCRNVGVAALPCAEKDGFIWVWPGDGLPAETLPDFAQPPEGFLIHAEIMVDVPVEHGLLIENL
LDLAHAPFTHTSTFARGWPVPDFVKFHANKALSGFWDPYPIDMAFQPPCMTLSTIGLAQPGKIM
RGVTASQCKNHLHQLHVCMPSKKGHTRLLYRMSLDFLPWMRHVPFIDRIWKQVAAQVLGEDLVL
VLGQQDRMLRGGSNWSNPAPYDKLAVRYRRWRNGVNAEVARVRAGEPPSNPVAMSAGEMFSVDE
DDMDN

Arabidopsis thaliana **Chlorophyllide *a* oxygenase (*At*CAO):**

ATGAACGCGGCGGTGTTTAGCCCGAGCGCGCTGAGCCTGCCGATCAGCTTTAGCAAAACCCGTA
GCAGCTTCCTGAGCCGTAAGAAAGGTGTTAAGGGCGAGTTCCGTGTGTTTGCGGTTTTCGGCGA
CGAGAGCGGTCTGGTGGAAAAGAAAAGCCAGTGGCGTCCGCTGTTCGACGTTGAGGATCCGCGT
AGCAAAGCTCCGCCGTACAAGGGTAAATTTCTGGACGTGAACCAAGCGATCGAAGTTGCGCGTT

TCGACATTCAGTATCTGGATTGGCGTGCGCGTCAAGATCTGCTGACCATCATGATTCTGCACGA
CAAAGTGGTTGATGTGCTGAACCCGCTGGCGCGTGAGTACAAGAGCATCGGCACCGTTAAGAAA
GAACTGGCGGGTCTGCAGGAAGAGCTGAGCAAAGCGCACCAGCAAGTGCACATTAGCGAGGCGC
GTGTTAGCACCGCGCTGGACAAGCTGGCGCACATGGAGGAACTGGTGAACGATCGTCTGCTGCC
GGGTCGTGTGGTTACCGAACTGGACAAACCGAGCAGCAGCACCACCGCGAGCGCGGTTGAGCTG
GATCGTGAAAAAACCAACACCGGCGCGAAGAGCCTGAACGTGAGCGGTCCGGTTCCGCCGTACA
GCCCGCACCTGAAAAACTTTTGGTATCCGGTGGCGTTCACCGCGGACCTGAAGCACGATACCAT
GGTGCCGATCGAGTGCTTTGAACAGCCGTGGGTTATTTTCCGTGGTGAGGACGGCAAACCGGGT
TGCGTTCGTAACACCTGCGCGCACCGTGCGTGCCCGCTGGATCTGGGCACCGTTAACGAAGGTC
GTATCCAATGCCCGTACCACGGCTGGGAGTATAGCACCGACGGTGAATGCAAGAAAATGCCGAG
CACCAAGCTGCTGAAGGTTAAAATTAAGAGCCTGCCGTGCCTGGAGCAGGAAGGCATGATCTGG
ATTTGGCCGGGTGATGAGCCGCCGGCGCCGATCCTGCCGAGCCTGCAACCGCCGAGCGGCTTTC
TGATTCACGCGGAGCTGGTGATGGATCTGCCGGTTGAACACGGTCTGCTGCTGGACAACCTGCT
GGATCTGGCGCACGCGCCGTTTACCCACACCAGCACCTTCGCGAAAGGCTGGAGCGTGCCGAGC
CTGGTTAAGTTTCTGACCCCGACCAGCGGCCTGCAGGGTTACTGGGACCCGTATCCGATCGATA
TGGAGTTCAAACCGCCGTGCATTGTGCTGAGCACCATCGGTATTAGCAAACCGGGCAAGCTGGA
AGGCAAGAGCACCCAGCAATGCGCGACCCACCTGCACCAGCTGCACGTTTGCCTGCCGAGCAGC
AAAAACAAGACCCGTCTGCTGTACCGTATGAGCCTGGACTTCGCGCCGATCCTGAAAAACCTGC
CGTTTATGGAGCACCTGTGGCGTCACTTCGCGGAGCAAGTGCTGAACGAAGATCTGCGTCTGGT
TCTGGGCCAGCAAGAACGTATGCTGAACGGTGCGAACATTTGGAACCTGCCGGTGGCGTACGAC
AAACTGGGCGTTCGTTATCGTCTGTGGCGTAACGCGGTGGATCGTGGTGACGATAAGCTGCCGT
TTAGCGGCAGCAGCGGTGAAAACCTGTATTTCCAAGGTAGCGGCAGCAGC

MNAAVFSPSALSLPISFSKTRSSFLSRKKGVKGEFRVFAVFGDESGLVEKKSQWRPLFDVEDPR
SKAPPYKGKFLDVNQAIEVARFDIQYLDWRARQDLLTIMILHDKVVDVLNPLAREYKSIGTVKK
ELAGLQEELSKAHQQVHISEARVSTALDKLAHMEELVNDRLLPGRVVTELDKPSSSTTASAVEL
DREKTNTGAKSLNVSGPVPPYSPHLKNFWYPVAFTADLKHDTMVPIECFEQPWVIFRGEDGKPG
CVRNTCAHRACPLDLGTVNEGRIQCPYHGWEYSTDGECKKMPSTKLLKVKIKSLPCLEQEGMIW
IWPGDEPPAPILPSLQPPSGFLIHAELVMDLPVEHGLLLDNLLDLAHAPFTHTSTFAKGWSVPS
LVKFLTPTSGLQGYWDPYPIDMEFKPPCIVLSTIGISKPGKLEGKSTQQCATHLHQLHVCLPSS
KNKTRLLYRMSLDFAPILKNLPFMEHLWRHFAEQVLNEDLRLVLGQQERMLNGANIWNLPVAYD
KLGVRYRLWRNAVDRGDDKLPFSG

Micromonas pusilla **Chlorophyllide *a* oxygenase1 (*Mp*CAO1):**

ATGATCCCGTTCGATCTGTTCAACGTCCCGTGGGTCGCGTTCCGCGACCAGGACGGCATGGCGG
GGTGCATCAAGGACGAGTGCGCGCACAGGGCGTGCCCGATATCCCTCGGGAAAGTCGTCGAAGG
ACGCGTGCAGTGCCCGTACCACGGGTGGGAGTACACCTCGGGAGGGGAGTGCAAGAAGATGCCG
TCGATAAAAAACCTCCTCCCGAACGTCTACGTCGACGCCGCGCCGATCGTGGAGCGCGACGGGC
TGTTATACGTGTGGGCGGGGGTGTGGGAGCCGGAGAGGGCGGAGGAGATATTGAGCGAGCTCCC
GCCGAGCGCGGCGACCGCGCCGCCGAGCGGGTTCGCCGCGATGGCTGAGGTCACCGTCGACGTC
CCGCTCGACGCGCCCGCGATTTTGTCGCGACTGATGGACGAGAACAAGGTGCCGTTCACGCGCG

TGGACACGACGACGCTGAGCGACGACGTCTTCCCGAAGGTGATCGCGAAGGTGTTGCGGGGGTT
CGGTAAGCCCGCGCCGAAACGCGTGGAGTTCACCCCGGCGTGCATACTCGACAGCACGATCGGT
CTGGACGGCGTGGGCGGACAAGACTGGAACGTGCATCAGACGCACGTCGTGTTACCCTCGCGGC
CCGGGAAGGCGCGGGTGTTGTACCGCCTCTCCGTGGACTTTGTCGTCGGCGCGGAGATCGCGCG
GACCGTGGGCGGGCAGGTGTGGCAGAACCTCGCGGAGATGATCCTCCAAGAGCAGCTCGAGGGG
ATCCGGGGGGGGAGGTTCGAGGACGACAGCGTCGGGGAGCAGGCCGCGGACGTGTCGCAGAGCT
ACGACGAGTGGATGGAAGAGATCCAGGCGCCGCGGTGA

MIPFDLFNVPWVAFRDQDGMAGCIKDECAHRACPISLGKVVEGRVQCPYHGWEYTSGGECKKMP
SIKNLLPNVYVDAAPIVERDGLLYVWAGVWEPERAEEILSELPPSAATAPPSGFAAMAEVTVDV
PLDAPAILSRLMDENKVPFTRVDTTTLSDDVFPKVIAKVLRGFGKPAPKRVEFTPACILDSTIG
LDGVGGQDWNVHQTHVVLPSRPGKARVLYRLSVDFVVGAEIARTVGGQVWQNLAEMILQEQLEG
IRGGRFEDDSVGEQAA DVSQSYDEWMEEIQAPR

***Micromonas pusilla* Chlorophyllide *a* oxygenase2 (*Mp*CAO2):**

ATGGCGCCCGAGGTGTCCTCCTCCCCCGCGCCCGTGGACGCCTCGAGGGGGGGCGCTGAGAAAG
CCGCGCGCAGGCTCGGCATGGGACGAAGGTTCGCGGACGAGGTCTCCCTGGACAACCCCTCCGA
CGTCGGCGCGATCCGCAACTATTGGTACCCGATCCACTTCATCTCGAAGCTGAACAAGGGCGAC
GCGGCCACCTCCTTCGTGCTCTTCGGCGAGCGGTGGGAGCTCGTCGCCGACGACGACGCCGCCG
TCGCCGCCGCGAAGACCGCCGTCGGCGTCTTCGGCCCGGAGTACGCGGAGACGCAGGCGCACCT
GGTGGACGGCGCCGCACAGAGGTGGACGTGCCGCTCGCGCGACGACGCGACGCGCTTCCTCCCG
ATCGGGCTGCAGGACGGCCTCGTCATGCCCGACGTCGCGTTGCCGACGACGTTTACACCCCCGG
CGGGGTACACGACGCACGCGGAGCTCATCATCGAGGACGTCCCCGTGGAGCACGGGCTGCTGAT
GGAGAACCTGCTGGACCTCGCGCACGCGCCGTTCACGCACACCGGGACGTTCGCGAAGGGCTGG
GGCGTCCCCACGTTCGTCGAGTTCGTCACGTCGAAGCTCCGGAGGGAAGGGGACGGGTGGCAGG
ACATGGCGCGAGGGCTCACGCGCGAGGGGATCGGGCTGGGGTCGCAGCAGGGGTCGTGGAACCC
GTACCCGATCGACATGAAGTTCGTCACGCCGTGCATGGTGGACTCGCACATCGGGATGTCGCAG
GCTGGCGCCGCGGGGAAGGGGGCGCAGTTTGAAGAGGGCGTCCAGTGCGCGGAGTGCAGCAACC
ACCTCCACCAACTCCACGTGTGCGTGCCGTCCGAGCCCGGGCGGACGCGCCTGCTGTATCGGAT
GGCGCTCGATTTCGCCGGGTGGGCGAAGTACGTCCCGGGGATCGAGCTCGTGTGGACGGAGATG
GCGAACCAGGTCTTGGGCGAGGACTTGAGGCTGGTGACCGGGCAGCAGGATCGGATGCGGAGAG
GCGGACGGGTGTGGGCGCATCCGGTGGCGTACGATAAGCTCGGGCTGGTGTACCGCAGGTGGAG
GAACTTCAGCGTTGGCGAGGCGTGCGACGTGAGCGCGGGTATCGGCGCGGAGGGCGCGGGGGAG
ACGGCGCAGCGGTCGCGGTGA

MAPEVSSSPAPVDASRGGAEKAARRLGMGRRFADEVSLDNPSDVGAIRNYWYPIHFISKLNKGD
AATSFVLFGERWELVADDDAAVAAAKTAVGVFGPEYAETQAHLVDGAAQRWTCRSRDDATRFLP
IGLQDGLVMPDVALPTTFTPPAGYTTHAELIIEDVPVEHGLLMENLLDLAHAPFTHTGTFAKGW
GVPTFVEFVTSKLRREGDGWQDMARGLTREGIGLGSQQGSWNPYPIDMKFVTPCMVDSHIGMSQ
AGAAGKGAQFEEGVQCAECSNHLHQLHVCVPSEPGRTRLLYRMALDFAGWAKYVPGIELVWTEM

ANQVLGEDLRLVTGQQDRMRRGGRVWAHPVAYDKLGLVYRRWRNFSVGEACDVSAGIGAEGAGE
TAQRSR

4.6.2 Sequences of reductases

***Pseudomonas* VanB:**

ATGTACTTCCAATCCAATGCAATGATCGAGGTTTTGGTGGCAGCGCGCCACCACGAGGCGCTGG
ATATTTGCGGCTACGAACTGACTGCTGTTGACGGTCAGCCACTGCCGGCATTTACGGCGGGTGC
ACACATTGATGTCCACCTTCCTACTGGGATGATTCGCCAATATTCTTTATGCAATCACCCTGAG
GAACGCCATCGCTACCTTATTGGTGTTCTGCGTGACCCCGCATCCCGTGGTGGGTCGCGCGCGA
TGCACGAATTGATCGAGCCTGGTACACGTTTGCAAATCTCAGAGCCGCGCAATCTTTTTCCTTT
GGCCCCTGAGGCTCAACGTTCACTGCTTTTTGCGGGAGGGATCGGGATCACCCCGATCCTTTCA
ATGGCCGAGTTTCTGGCCCAAGACGGTGCCGCTTTCGAGCTGCACTACTGTGCGCGTTCCCGTG
AGCGCGCAGCTTTCGTAGAACGTTTGCGTTTGTCACCTTACGCGGACCGTGTCTTCCTTCATTT
CGATGAAGAGCCGGACACCTTGCTGGACGCCGCTGCAGTCCTGTGCGCGCCAACTGATGATCTT
CATTTATATGTCTGCGGACCCGGCGGATTCATGCAATATATTTTAGATACGGCGCGCCATCATG
GGTGGGAAGAGACCCGCCTTCACCGTGAGTATTTTTCCGCAGCGCCAGTTGATACGCGTGCGGA
TGGTTCTTTCAGTGTAAAGCTGGCGCGCTCGGGGCAGGTTTTCGACATCCCGGCAGATCGCTCT
GTAGTCCAAGTGTTGGAATCACACGGGATCGAAATTCCAATTAGTTGTGAGCAGGGCATCTGTG
GAACCTGCTTGACCCGCGTCTTGGAAGGGGTCCCAGAACATCGCGATATGTTTTTAACTGAAGC
AGAGCAGGCGTGTAATGATCAATTCACGCCCTGTTGCTCACGCTCAAAGACGCCTCTTCTGGTC
CTTGACCTGTAGCATTGGAAGTGGATAA

MIEVIVGAIRLEAQDIHSFELFRADGAALPSFEPGAHIDLHLPNGLVRQYSLCGPAERPRHYRI
AVLRCRDSRGGSATLHAELRVGQRLHIGEPRNLFPLSPEPGPHLLFAGGIGITPLLAMAERLAR
DGADFQLHYCAHSGERAAFVDYLGRCAFADRVHCHFDHGESSRRADLRALLATSPRDAQLYLCG
PAGFMQWIEESARELGWEASRLHREHFAAAPRDASADGTFEVQLASNGALIRVAAGQTVLAALR
EAGVDLPASCEQGICG TCLTRVLDGEPEHRDLYLSEEEQAANDCFTPCCSRSRSPRLVLDL

***Stenotrophomonas maltophilia* Dicamba monooxygenase reductase, ferredoxin component (DdmB):**

ATGCCGCAGATTACCGTCGTCAACCAGTCGGGTGAAGAATCCAGCGTCGAGGCGAGTGAAGGCC
GCACCCTGATGGAAGTCATCCGCGACAGCGGTTTTGACGAACTCCTGGCGCTTTGCGGCGGCTG
CTGCTCGTGCGCGACCTGCCACGTCCACATCGACCCGGCCTTCATGGACAAGCTGCCGGAGATG
AGCGAAGACGAGAACGACCTGCTCGACAGCTCGGACCACCGCAACGAGTACTCGCGTCTCTCGT
GCCAGATTCCGGTCACCGGCGCCCTCGAAGGCATCAAGGTGACGATCGCGCAGGAAGACTGA

MPQITVVNQSGEESSVEASEGRTLMEVIRDSGFDELLALCGGCCSCATCHVHIDPAFMDKLPEM
SEDENDLLDSSDHRNEYSRLSCQIPVTGALEGIKVTIAQED

Stenotrophomonas maltophilia Dicamba monooxygenase reductase (DdmA):

ATGAGCAAGGCAGACGTCGTAATCGTGGGAGCCGGGCATGGCGGCGCACAGTGCGCGATCGCCC
TTCGCCAGAACGGCTTCGAAGGAACCATCACCGTCATCGGTCGTGAGCCGGAATATCCCTATGA
GCGTCCGCCGCTCTCGAAGGAATATTTCGCGCGCGAGAAGACCTTCGACCGCCTCTACATCCGT
CCGCCGACGTTCTGGGCCGAGAAGAACATCGAGTTCAAGCTTGGCACCGAAGTCACCAAGGTCG
ATCCCAAGGCGCACGAACTGACGCTCTCCAACGGCGAGAGCTACGGTTATGGCAAGCTCGTCTG
GGCCACCGGCGGCGATCCGCGTCGCCTTTCTTGCCAGGGGGCCGACCTCACCGGCATCCACGCC
GTGCGCACCCGCGAGGACTGCGACACGCTGATGGCCGAAGTCGATGCGGGCACGAAGAACATCG
TCGTCATCGGCGGCGGCTACATCGGTCTGGAAGCCGCTGCGGTGCTGTCCAAGATGGGCCTCAA
GGTCACCCTGCTCGAAGCGCTTCCGCGCGTGCTGGCGCGCGTTGCGGGTGAAGACCTCTCGACC
TTCTACCAGAAGGAACATGTCGATCACGGCGTCGACCTGCGCACCGAAGTCATGGTCGACAGCC
TCGTCGGCGAAAACGGCAAGGTCACCGGCGTGCAGCTTGCCGGCGGCGAAGTGATCCCGGCCGA
AGGCGTCATCGTCGGCATCGGCATCGTGCCTGCCGTCGGTCCGCTGATCGCGGCCGGCGCGGCC
GGTGCCAACGGCGTCGACGTGGACGAGTACTGCCGCACCTCGCTGCCCGACATCTATGCGATCG
GCGACTGTGCGGCTTTCGCCTGCGACTACGCCGGCGGCAACGTGATGCGCGTGGAATCGGTCCA
GAACGCCAACGACATGGGCACCTGCGTGGCCAAGGCGATCTGCGGCGACGAGAAGCCCTACAAG
GCGTTCCCGTGGTTCTGGTCCAACCAGTACGACCTCAAGCTGCAGACCGCCGGCATCAACCTGG
GCTTCGACAAGACCGTGATCCGCGGCAATCCGGAGGAGCGCAGCTTCTCGGTCGTCTATCTCAA
GGACGGCCGCGTGGTCGCGCTGGACTGCGTGAACATGGTCAAGGATTACGTGCAGGGCCGCAAG
CTGGTCGAAGCCGGGGCCACCCCCGACCTCGAAGCGCTGGCCGATGCCGGCAAGCCGCTCAAGG
AACTGCTCTAG

MSKADVVIVGAGHGGAQCAIALRQNGFEGTITVIGREPEYPYERPPLSKEYFAREKTFDRLYIR
PPTFWAEKNIEFKLGTEVTKVDPKAHELTLSNGESYGYGKLVWATGGDPRRLSCQGADLTGIHA
VRTREDCDTLMAEVDAGTKNIVVIGGGYIGLEAAAVLSKMGLKVTLLEALPRVLARVAGEDLST
FYQKEHVDHGVDLRTEVMVDSLVGENGKVTGVQLAGGEVIPAEGVIVGIGIVPAVGPLIAAGAA
GANGVDVDEYCRTSLPDIYAIGDCAAFACDYAGGNVMRVESVQNANDMGTCVAKAICGDEKPYK
AFPWFWSNQYDLKLQTAGINLGFDKTVIRGNPEERSFSVVYLKDGRVVALDCVNMVKDYVQGRK
LVEAGATPDLEALADAGKPLKELL

Comamonas testosteroni TsaB:

ATGAGCGCCGATGTGCCCGTGACCGTGGCCGCCGTGCGGGCGGTGGCCCGCGACGTGCTGGCGC
TGGAACTGCGCCACGCCAACGGCCAGCCCCTGCCCGGCGCCAGCGCCGGTGCCCACATTGACCT
CGCCCTGCCCAATGGCCTGGTGCGCCAGTACTCGCTGGTGAACGCCACCGGCCAGGCCACCATG
GACTGCTACCAGGTGGCCGTGGGCTGGGACGCCAACAGCCGTGGCGGCTCGGTGTGGATTCACG
AAAAGCTCAAGGTGGGCCAGGCCCTGCGCGTGACGCACCGCGCAACCTGTTCCGAGATGGCGCC

CGAGCACCGCCGCGTGCTGCTGCTGGCCGGTGGCATTGGCGTCACGCCCATCTACGCCATGGCC
CAGGCTTGCGCGCAGCAAGGCGTGGACGTTGAGCTGTGGGCCAGCGCCCGCTCGGCCCCGCGCC
TGGCCTACTTGGAAGAACTCAAGGCACTGCTGGGCCAGCGCCTGCACCTGCATGCCGACGACGA
GCAGGGCGGCCCCATGAACCTCACCGAGCGCCTGGCCACCCAGCGCTGGGACGCGGTTTACGCC
TGCGGCCCCGCGCCCATGCTGGACGCGCTCACCGCCGCCACCGCCCACTGGGCGCCGGGCTCGG
TGCGCATGGAACGCTTCAAGGGCGCCGAGCAGCCCGCCAGCGAGCGCCAGCCCTTTGAGCTGGT
GCTGCAGCGCGCCGGCCTGAGCACCACGGTGGACGCGCACGAGAGCGTGCTCGACGCCATGGAG
CGCGTGGGCGTGGACTTCCCTTGGTCGTGCCGCGAAGGCATTTGCGGCACCTGCGAAGCGCCGG
TGCTCGAAGGCGAGGTGCAGCACCTCGATTACGTGCTCTCGCCCGAAGAACGCGCCGAACAGCG
GCGAATGATGGTCTGCGTGTCGCGTTGTGGCGGCGGCCGGCTGGTCCTGGACATCTGA

MSADVPVTVAAVRAVARDVLALELRHANGQPLPGASAGAHIDLALPNGLVRQYSLVNATGQATM
DCYQVAVGWDANSRGGSVWIHEKLKVGQALRVTHRATCSEMAPEHRRVLLLAGGIGVTPIYAMA
QACAQQGVDVELWASARSAPRLAYLEELKALLGQRLHLHADDEQGGPMNLTERLATQRWDAVYA
CGPAPMLDALTAATAHWAPGSVRMERFKGAEQPASERQPFELVLQRAGLSTTVDAHESVLDAME
RVGVDFPWSCREGICGTCEAPVLEGEVQHLDYVLSPEERAEQRRMMVCVSRCGGGRLVLDI

Escherichia coli **Flavodoxin reductase:**

ATGGCTGATTGGGTAACAGGCAAAGTCACTAAAGTGCAGAACTGGACCGACGCCCTGTTTAGTC
TCACCGTTCACGCCCCCGTGCTTCCGTTTACCGCCGGGCAATTTACCAAGCTTGGCCTTGAAAT
CGACGGCGAACGCGTCCAGCGCGCCTACTCCTATGTAAACTCGCCCGATAATCCCGATCTGGAG
TTTTACCTGGTCACCGTCCCCGATGGCAAATTAAGCCCACGACTGGCGGCACTGAAACCAGGCG
ATGAAGTGCAGGTGGTTAGCGAAGCGGCAGGATTCTTTGTGCTCGATGAAGTGCCGCACTGCGA
AACGCTATGGATGCTGGCAACCGGTACAGCGATTGGCCCTTATTTATCGATTCTGCAACTAGGT
AAAGATTTAGATCGCTTCAAAAATCTGGTCCTGGTGCACGCCGCACGTTATGCCGCCGACTTAA
GCTATTTGCCACTGATGCAGGAACTGGAAAAACGCTACGAAGGAAAACTGCGCATTCAGACGGT
GGTCAGTCGGGAAACGGCAGCGGGGTCGCTCACCGGACGGATACCGGCATTAATTGAAAGTGGG
GAACTGGAAAGCACGATTGGCCTGCCGATGAATAAAGAAACCAGCCATGTGATGCTGTGCGGCA
ATCCACAGATGGTGCGCGATACACAACAGTTGCTGAAAGAGACCCGGCAGATGACGAAACATTT
ACGTCGCCGACCGGGCCATATGACAGCGGAGCATTACTGGTAA

MADWVTGKVTKVQNWTDALFSLTVHAPVLPFTAGQFTKLGLEIDGERVQRAYSYVNSPDNPDLE
FYLVTVPDGKLSPRLAALKPGDEVQVVSEAAGFFVLDEVPHCETLWMLATGTAIGPYLSILQLG
KDLDRFKNLVLVHAARYAADLSYLPLMQELEKRYEGKLRIQTVVSRETAAGSLTGRIPALIESG
ELESTIGLPMNKETSHVMLCGNPQMVRDTQQLLKETRQMTKHLRRRPGHMTAEHYW

Escherichia coli **Flavodoxin:**

ATGGCTATCACTGGCATCTTTTTCGGCAGCGACACCGGTAATACCGAAAATATCGCAAAAATGA
TTCAAAAACAGCTTGGTAAAGACGTTGCCGATGTCCATGACATTGCAAAAAGCAGCAAAGAAGA

TCTGGAAGCTTATGACATTCTGCTGCTGGGCATCCCAACCTGGTATTACGGCGAAGCGCAGTGT
GACTGGGATGACTTCTTCCCGACTCTCGAAGAGATTGATTTCAATGGCAAACTGGTTGCGCTGT
TTGGTTGTGGTGACCAGGAAGATTACGCCGAATATTTCTGCGACGCATTGGGCACCATCCGCGA
CATCATTGAACCGCGCGGTGCAACCATCGTTGGTCACTGGCCAACTGCGGGCTATCATTTCGAA
GCATCAAAAGGTCTGGCAGATGACGACCACTTTGTCGGTCTGGCTATCGACGAAGACCGTCAGC
CGGAACTGACCGCTGAACGTGTAGAAAAATGGGTTAAACAGATTTCTGAAGAGTTGCATCTCGA
CGAAATTCTCAATGCC

MAITGIFFGSDTGNTENIAKMIQKQLGKDVADVHDIAKSSKEDLEAYDILLLGIPTWYYGEAQC
DWDDFFPTLEEIDFNGKLVALFGCGDQEDYAEYFCDALGTIRDIIEPRGATIVGHWPTAGYHFE
ASKGLADDDHFVGLAIDEDRQPELTAERVEKWVKQISEELHLDEILNA

4.6.3 Sequence of chlorophyllase

***Triticum aestivum* Chlorophyllase:**

ATGGCTAGCGCAGCAGCAGCACCCGCAGAGACGATGAACAAGTCCGCCGCCGGCGCCGAGGTTC
CCGAGGCGTTCACATCGGTGTTCCAGCCGGGGAAGCTTGCGGTCGAGGCGATTCAGGTGGATGA
GAATGCGGCGCCGACACCACCGATCCCGGTGCTGATCGTCGCACCCAAGGATGCAGGAACCTAC
CCCGTGGCCATGCTCTTGCACGGCTTCTTCCTCCATAACCACTTCTACGAACACCTTCTCCGGC
ACGTCGCATCCCACGGCTTCATCATTGTCGCGCCCCAGTTCAGCATCAGTATCATACCTTCGGG
TGACGCAGAGGACATCGCCGCGGCAGCCAAGGTGGCAGACTGGCTCCCCGACGGCCTCCCGTCC
GTGCTGCCCAAAGGCGTCGAGCCGGAGCTCTCGAAGCTCGCCTTGGCCGGCCACAGCCGAGGAG
GCCACACGGCTTTCTCCCTGGCCTTGGGGCACGCCAAGACCCAGCTAACCTTCTCCGCGCTCAT
CGGACTCGACCCCGTCGCCGGCACGGGGAAGTCCTCCCAGCTCCAGCCCAAGATCCTCACCTAC
GAGCCGTCCTCCTTCGGCATGGCGATGCCGGTGCTGGTCATCGGCACCGGGCTCGGCGAGGAGA
AGAAGAACATATTCTTCCCTCCCTGCGCACCCAAGGACGTGAACCACGCCGAGTTCTACCGCGA
GTGCAGGCCGCCCTGCTACTACCTTGTGACCAAGGACTACGGGCATCTGGACATGCTGGACGAC
GACGCCCCCAAGTTCATCACCTGCGTCTGCAAGGATGGGAACGGGTGCAAGGGCAAGATACGGA
GGTGCGTTGCTGGGATCATGGTGGCATTTCTTAATGCTGCCTTGGGTGAGAAAGATGCAGATCT
TGAGGCCATACTGAGAGACCCGGCGGTTGCACCCACCACGCTTGATCCGGTTGAGCACCGCGTG
GCGGCGGCCGCACTCGAG

MASAAAAPAETMNKSAAGAEVPEAFTSVFQPGKLAVEAIQVDENAAPTPPIPVLIVAPKDAGTY
PVAMLLHGFFLHNHFYEHLLRHVASHGFIIVAPQFSISIIPSGDAEDIAAAAKVADWLPDGLPS
VLPKGVEPELSKLALAGHSRGGHTAFSLALGHAKTQLTFSALIGLDPVAGTGKSSQLQPKILTY
EPSSFGMAMPVLVIGTGLGEEKKNIFFPPCAPKDVNHAEFYRECRPPCYYLVTKDYGHLDMLDD
DAPKFITCVCKDGNGCKGKIRRCVAGIMVAFLNAALGEKDADLEAILRDPAVAPTTLDPVEHRV
AAAALE

4.7 References

1. Bryant, D. A.; Hunter, C. N.; Warren, M. J., Biosynthesis of the modified tetrapyrroles-the pigments of life. *J Biol Chem* **2020**, 295 (20), 6888-6925.
2. Chen, M., Chlorophyll modifications and their spectral extension in oxygenic photosynthesis. *Annu Rev Biochem* **2014**, 83, 317-40.
3. Chen, M.; Blankenship, R. E., Expanding the solar spectrum used by photosynthesis. *Trends Plant Sci* **2011**, 16 (8), 427-31.
4. Hu, X.; Gu, T.; Khan, I.; Zada, A.; Jia, T., Research Progress in the Interconversion, Turnover and Degradation of Chlorophyll. *Cells* **2021**, 10 (11).
5. Schliep, M.; Cavigliasso, G.; Quinnell, R. G.; Stranger, R.; Larkum, A. W., Formyl group modification of chlorophyll a: a major evolutionary mechanism in oxygenic photosynthesis. *Plant Cell Environ* **2013**, 36 (3), 521-7.
6. Tros, M.; Bersanini, L.; Shen, G.; Ho, M. Y.; van Stokkum, I. H. M.; Bryant, D. A.; Croce, R., Harvesting far-red light: Functional integration of chlorophyll f into Photosystem I complexes of Synechococcus sp. PCC 7002. *Biochim Biophys Acta Bioenerg* **2020**, 1861 (8), 148206.
7. Ho, M. Y.; Shen, G.; Canniffe, D. P.; Zhao, C.; Bryant, D. A., Light-dependent chlorophyll f synthase is a highly divergent paralog of PsbA of photosystem II. *Science* **2016**, 353 (6302).
8. Murray, J. W., Sequence variation at the oxygen-evolving centre of photosystem II: a new class of 'rogue' cyanobacterial D1 proteins. *Photosynth Res* **2012**, 110 (3), 177-84.

9. Shen, G.; Canniffe, D. P.; Ho, M. Y.; Kurashov, V.; van der Est, A.; Golbeck, J. H.; Bryant, D. A., Characterization of chlorophyll f synthase heterologously produced in Synechococcus sp. PCC 7002. *Photosynth Res* **2019**, 140 (1), 77-92.

10. Oster, U.; Tanaka, R.; Tanaka, A.; Rudiger, W., Cloning and functional expression of the gene encoding the key enzyme for chlorophyll b biosynthesis (CAO) from Arabidopsis thaliana. *Plant J* **2000**, 21 (3), 305-10.

11. Oster, U.; Bauer, C. E.; Rudiger, W., Characterization of chlorophyll a and bacteriochlorophyll a synthases by heterologous expression in Escherichia coli. *J Biol Chem* **1997**, 272 (15), 9671-6.

12. Perry, C.; de Los Santos, E. L. C.; Alkhalaf, L. M.; Challis, G. L., Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat. Prod. Rep*. **2018**, 35 (7), 622-632.

13. Kovaleva, E. G.; Lipscomb, J. D., Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nat. Chem. Biol*. **2008**, 4 (3), 186-93.

14. Barry, S. M.; Challis, G. L., Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal*. **2013**, 3, 2362-2370.

15. Ferraro, D. J.; Gakhar, L.; Ramaswamy, S., Rieske business: structure-function of Rieske non-heme oxygenases. *Biochem Biophys Res Commun* **2005**, 338 (1), 175-90.

16. Knapp, M.; Mendoza, J.; Bridwell-Rabb, J., *In Encyclopedia of Biological Chemistry*, 3rd Edition, **2021**.

17. Schneegurt, M. A.; Beale, S. I., Origin of the chlorophyll b formyl oxygen in Chlorella vulgaris. *Biochemistry* **1992**, 31 (47), 11677-83.

18. Tanaka, A.; Ito, H.; Tanaka, R.; Tanaka, N. K.; Yoshida, K.; Okada, K., Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proc Natl Acad Sci* U S A **1998**, 95 (21), 12719-23.

19. Chunaev, A. S.; Mirnaya, O. N.; Maslov, V. G.; Boschetti, A., Chlorophyll B-Deficient and Loroxanthin-Deficient Mutants of Chlamydomonas-Reinhardii. *Photosynthetica* **1991**, 25 (2), 291-301.

20. Simpson, D. J.; Machold, O.; Hoyerhansen, G.; Vonwettstein, D., Chlorina Mutants of Barley (Hordeum-Vulgare L). *Carlsberg Res Commun* **1985**, 50 (4), 223-238.

21. Lukowski, A. L.; Ellinwood, D. C.; Hinze, M. E.; DeLuca, R. J.; Du Bois, J.; Hall, S.; Narayan, A. R. H., C-H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J Am Chem Soc* **2018**, 140 (37), 11863-11869.

22. Lukowski, A. L.; Liu, J.; Bridwell-Rabb, J.; Narayan, A. R. H., Structural basis for divergent C-H hydroxylation selectivity in two Rieske oxygenases. *Nat Commun* **2020**, 11 (1), 2991.

23. Kim, J. H.; Kim, B. H.; Brooks, S.; Kang, S. Y.; Summers, R. M.; Song, H. K., Structural and Mechanistic Insights into Caffeine Degradation by the Bacterial N-Demethylase Complex. *J Mol Biol* **2019**, 431 (19), 3647-3661.

24. Locher, H. H.; Leisinger, T.; Cook, A. M., 4-Toluene sulfonate methyl-monooxygenase from Comamonas testosteroni T-2: purification and some properties of the oxygenase component. *J Bacteriol* **1991**, 173 (12), 3741-8.

25. Kunugi, M.; Takabayashi, A.; Tanaka, A., Evolutionary changes in chlorophyllide a oxygenase (CAO) structure contribute to the acquisition of a new light-harvesting complex in micromonas. *J Biol Chem* **2013**, 288 (27), 19330-41.

26. Wolfe, M. D.; Lipscomb, J. D., Hydrogen peroxide-coupled cis-diol formation catalyzed by naphthalene 1,2-dioxygenase. *J Biol Chem* **2003**, 278 (2), 829-35.

27. Helfrich, M.; Bommer, B.; Oster, U.; Klement, H.; Mayer, K.; Larkum, A. W.; Rudiger, W., Chlorophylls of the c family: absolute configuration and inhibition of NADPH: protochlorophyllide oxidoreductase. *Biochim Biophys Acta* **2003**, 1605 (1-3), 97-103.

28. Katz, J. J.; Norman, G. D.; Svec, W. A.; Strain, H. H., Chlorophyll Diastereoisomers. Nature of Chlorophylls a' and B' and Evidence for Bacteriochlorophyll Epimers from Proton Magnetic Resonance Studies. *Journal of the American Chemical Society* **1968**, 90 (24), 6841-&.

29. Watanabe, T.; Nakazato, M.; Honda, K., Epimerization of Chlorophyll Derivatives .2. Kinetic and Thermodynamic Parameters for the Pheophytin a/a' Epimerization in Organic-Solvents. *Chem Lett* **1986**, (2), 253-256.

30. Fiedor, L.; Rosenbach-Belkin, V.; Scherz, A., The stereospecific interaction between chlorophylls and chlorophyllase. Possible implication for chlorophyll biosynthesis and degradation. *J Biol Chem* **1992**, 267 (31), 22043-7.

31. Wang, X.; Liu, L., Crystal Structure and Catalytic Mechanism of 7-Hydroxymethyl Chlorophyll a Reductase. *J Biol Chem* **2016**, 291 (25), 13349-59.

32. Ito, H.; Ohtsuka, T.; Tanaka, A., Conversion of chlorophyll b to chlorophyll a via 7-hydroxymethyl chlorophyll. *J Biol Chem* **1996**, 271 (3), 1475-9.

33. Tsuchiya, T.; Mizoguchi, T.; Akimoto, S.; Tomo, T.; Tamiaki, H.; Mimuro, M., Metabolic engineering of the Chl d-dominated cyanobacterium Acaryochloris marina: production of a novel Chl species by the introduction of the chlorophyllide a oxygenase gene. *Plant Cell Physiol* **2012**, 53 (3), 518-27.

34. Loughlin, P. C.; Willows, R. D.; Chen, M., *In vitro* conversion of vinyl to formyl groups in naturally occurring chlorophylls. *Sci Rep* **2014**, 4, 6069.

35. Xu, H.; Vavilin, D.; Vermaas, W., Chlorophyll b can serve as the major pigment in functional photosystem II complexes of cyanobacteria. *Proc Natl Acad Sci* U S A **2001**, 98 (24), 14168-73.

36. Tanaka, R.; Tanaka, A., Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. *Biochim Biophys Acta* **2011**, 1807 (8), 968-76.

37. Yamasato, A.; Nagata, N.; Tanaka, R.; Tanaka, A., The N-terminal domain of chlorophyllide a oxygenase confers protein instability in response to chlorophyll B accumulation in Arabidopsis. *Plant Cell* **2005**, 17 (5), 1585-97.

38. Helfrich, M.; Schoch, S.; Lempert, U.; Cmiel, E.; Rudiger, W., Chlorophyll synthetase cannot synthesize chlorophyll a'. *Eur J Biochem* **1994**, 219 (1-2), 267-75.

39. Chen, G. E.; Adams, N. B. P.; Jackson, P. J.; Dickman, M. J.; Hunter, C. N., How the O2-dependent Mg-protoporphyrin monomethyl ester cyclase forms the fifth ring of chlorophylls. *Nat Plants* **2021**, 7 (3), 365-375.

40. Heyes, D. J.; Martin, G. E.; Reid, R. J.; Hunter, C. N.; Wilks, H. M., NADPH: protochlorophyllide oxidoreductase from Synechocystis: overexpression, purification and preliminary characterisation. *FEBS Lett* **2000**, 483 (1), 47-51.

41. Shepherd, M.; McLean, S.; Hunter, C. N., Kinetic basis for linking the first two enzymes of chlorophyll biosynthesis. *FEBS J* **2005**, 272 (17), 4532-9.

42. Reinbothe, S.; Bartsch, S.; Rossig, C.; Davis, M. Y.; Yuan, S.; Reinbothe, C.; Gray, J., A Protochlorophyllide (Pchlide) a Oxygenase for Plant Viability. *Front Plant Sci* **2019**, 10, 593.

43. Lanz, N. D.; Grove, T. L.; Gogonea, C. B.; Lee, K. H.; Krebs, C.; Booker, S. J., RlmN and AtsB as models for the overproduction and characterization of radical SAM proteins. *Methods Enzymol* **2012**, 516, 125-52.

Chapter 5 Conclusion and Future Directions

5.1 Summary

Tens of millions of enzymes have evolved in Nature to facilitate various chemical transformations in living organisms. From photosynthesis to nitrogen fixation, starch hydrolysis to DNA synthesis, enzymes exhibit astonishing functions that facilitate almost all metabolic processes in the cell and sustain life[1]. Due to the exquisite selectivity, unparalleled reactivity, and ubiquity of enzymes, they have been increasingly recognized and utilized as biocatalysts. Indeed, the first example of using an enzyme to facilitate catalysis was in 1858 when Louis Pasteur demonstrated that (-)-tartaric acid, rather than a racemic mixture, could be exclusively produced through fermentation with a variety of microorganisms[2-3]. Since this time, the rising demands for selective, safe, and sustainable industrial processes has only increased the number of studies aimed at repurposing and customizing both the selectivity and reactivity of enzymes for use in industrial manufacturing and pharmaceutical production[3-4]. Meanwhile, discovery of new enzymes, and the number of methods for engineering and diversifying enzymes has exponentially increased. Such methods, including directed evolution, exploitation of cofactors for new reactivities, and use of synthetic reagents and non-natural conditions are the most powerful and prevalent strategies for the discovery of new enzyme functions[5].

Among these strategies, the massive efforts put into engineering proteins has mainly focused on tuning the catalytic activity of an enzyme by merely changing one or a few amino acids. These changes have also historically focused on making mutations in the active site. However, this simplistic view of enzyme activity, in line with the traditional lock-key model, suggests that the enzyme-substrate interaction is limited to the active site. A recent paradigm shift, known as the keyhole-lock-key model has expanded this vision[8]. This model expands the focus beyond the interactions formed between the substrate and the active site and adds in the so-called "keyhole". This region, which seems to be especially important for metalloenzymes includes the regions between the buried metallocenter-containing active site and the surface of the protein, namely the substrate entrance tunnel[9]. Along with dynamic loop regions, such secondary sphere regions are emerging as enzyme engineering hotspots and have been recognized for their potential influence on protein stability, substrate scope, and reaction selectivity[6-10]. Remarkably, we showed, in chapter 3, that just focusing on the tunnel resulted in an approximate 90-percent change in the selectivity of the SxtT catalyzed reaction. This result thus shows that this region not only allows substrate to enter the active site but plays a key role in positioning it for chemistry.

In chapters 2 and 3, we demonstrated the potential of using rational engineering guided by structural analysis to engineer the selectivity of a Rieske oxygenase. We first solved the crystal structures of two Rieske oxygenases, SxtT and GxtA with and without substrate analog bound. Two critical residues in the active site were then identified that played a crucial role in determining the site selectivity between these two enzymes. However, site-directed mutagenesis and activity assays revealed that residues beyond just these two residues are required for site selectivity. This result suggests that the traditional lock-key model is insufficient to describe the selectivity and guide further protein engineering in this case. With the inspiration of the modified "keyhole-lock-key" paradigm, we then identified the so-called "keyhole" component, or a dynamic loop and a substrate tunnel that also contribute to its selectivity. Indeed, site-directed mutagenesis results indicated that SxtT can showcase the substrate-specificity and site-selectivity of GxtA by (i) the removal of the two unique residues that interact with the native substrate of SxtT (Thr276 and

Arg204), (ii) the addition of a residue that interacts with the α-hydroxyl of STX in GxtA (Tyr255), and (iii) the replacement of three SxtT tunnel lining residues with the GxtA counterparts. As shown in chapter 3, this latter tunnel region was the most important contributing factor to changing the primary site of hydroxylation from C12 to C11. In short, the substrate of these enzymes must be recognized and positioned by the "keyhole", or the loop and tunnel regions, to be correctly fit into the active site for selective hydroxylation.

More interestingly, as the flexible loop and substrate tunnel region are prevalent in all structurally characterized α3 Rieske oxygenases, we suggest that this modified "keyhole-lock-key" may play a more universal role in selectivity. Analysis of the other structurally characterized α3 Rieske oxygenases revealed that a loop of high flexibility exists between the β13 and β14 strands, an equivalent region to that observed in SxtT and GxtA. Meanwhile, a comparable substrate tunnel that leads from the surface of the protein towards the active site can be visualized computationally for most of the Rieske oxygenases with the flexible loop. These results demonstrated the possibility that other Rieske oxygenases also employ these three architectural regions to control selectivity and specificity. Interestingly, recent studies also showed that the reactivity and selectivity of metalloenzymes from different families can be altered by engineering a loop or a tunnel region other than active site. For example, a recent study from Dodani et al. describes a single mutation in the dynamic loop of the nitrating cytochrome P450 enzyme, TxtE, completely shifts the regioselectivity from the C4 to the C5 position of L-tryptophan[11]. Another study from Garcia-Borras et al. highlighted that activity and chemoselectivity can be modulated by engineering the dynamic loop that covers the substrate access of *Rhodothermus marinus* cytochrome c heme protein (*Rma* cyt c), of which the native function is electron transport[12]. An interesting study by Banerjee et al. revealed the relationship between the rigorous substrate selectivity of diminutive substrates and the small molecule tunnels of metalloenzymes[7]. With each of these examples in mind, investigating the cumulative effects of the roles that the active site and the regions beyond the active site should be top of mind for future rational protein engineering.

In chapter 4, we showed, for the first time, the *in vitro* activity of purified recombinant CAO protein with high quality mass spectrometry data. We further demonstrated the potential of CAO as an enzymatic tool for synthesizing custom-tuned natural and unnatural Chlorophyll pigments. This work is significant because, Chlorophyll(ide) an oxygenase (CAO) was identified over 20 years ago, but the mechanism of the proposed methyl- to formyl- group transformation by two sequential hydroxylation, has never been studied *in vitro* with purified protein. A few challenges include but are not limited to the lack of published protocols for the recombinant expression and purification of a CAO homolog, the light sensitivity of the proposed Chl substrate, the lack of an annotated reductase partner, and the deficiency of a feasible and accurate detection method for enzymatic assays. To overcome these obstacles, we first developed a viable method to overexpress and purify *Ph*CAO which lacks the accessory domain found in *At*CAO and *Cr*CAO. Second, using a bottom-up approach that relied on photosynthetic organism cell lysate, *E. coli* cell lysate, and extensive screening of non-native reductases revealed that CAO can convert Chlide *a* into Chlide *b* rather than Chl *a* into Chl *b*, monitored by a newly developed LC-MS method. These approaches were then extended to other CAO homologs and each of these isolated CAO homologs exhibited activity on Chlide *a* rather than Chl *a.* Interestingly, we also noticed that all CAO homologs purified in this work displayed a preference on a specific diastereomer, Chlide *a′*, indicating that there might be an epimerase enzyme present that is involved in regulating Chl cycle by controlling the ratio of different diastereomers. Meanwhile, MS data also suggested that 7-OH-Chlide *a* is a true intermediate of the reaction and revealed that CAO is a Rieske oxygenase that can facilitate the sequential oxygenation reaction in lieu of an additional Rieske protein, dehydrogenase, or cofactor. Finally, substrate scope screening with several Chl derivatives revealed that the lack of a phytol tail, the central metal ion and the 20-electron containing chlorin scaffold are nonnegligible for CAO activity. But most interestingly, it was determined in this work that 3-formyl-Chlide *a* (Chlide *d*) can be accepted by *Ph*CAO and a second formyl group can be installed at C7 position to produce the non-natural pigment 3-formyl-Chlide *b*. This result demonstrated the potential for using CAO as an enzymatic tool for synthesizing custom Chl pigments, those of which are utilized broadly in the cosmetic, food, agricultural, and pharmaceutical industries.

5.2 Future directions

Looking forward, our group will continue working on Rieske non-heme iron oxygenases and here are a few potential directions. First, it has previously been highlighted that a bridging Asp or Glu residue is significant for connecting the Rieske cluster and the non-heme iron site for electron transport during a catalytic cycle. Intriguingly, in chapter 2, we noticed that a Tyr residue at the subunit-subunit interface could be another candidate that bridges the two metallocenters of SxtT and GxtA to facilitate electron transfer. More surprisingly, this Tyr is conserved in all structurally characterized Rieske oxygenases in the sequence of Cys-X-His-X_{17}-Cys-X-Tyr-His, with the exception of KshA, which has a Phe residue instead. The role of this potential extra electron transferring residue needs further investigation to elucidate whether it is also involved in the mechanism of electron transfer, either serving a compensatory role and serving to regulate the flux of electrons or a complementary role to the bridging Asp/Glu residue. On the other hand, the [2Fe-2S] Rieske clusters possess a wide range electron reduction potential from -150 to +400 mV, which makes Rieske oxygenases exceptionally adaptable in catalyzing a diverse range of chemistry. Further study will focus on engineering the chemical environment of the Rieske cluster and the electron transport chain to manipulate the activity of Rieske oxygenases. Site-directed mutagenesis, along with Electron Paramagnetic Resonance Spectroscopy (EPR) and Neutron scattering would be powerful tools to answer these questions.

In addition, as mentioned in chapter 3, we revealed the design principles for selectivity in the Rieske oxygenase, SxtT. We successfully showed that the site-selectivity and substrate scope of SxtT was inverted and enlarged, respectively by introducing only six mutations into the SxtT encoding gene. These six mutations come from three architectural regions, the dynamic loop, active site, and substrate entrance channel. Further structural analysis revealed that these structural regions were conserved among all structurally characterized α_3 Rieske oxygenases, suggesting that this model could be applied to all these ROs for rational protein engineering. For example, in chapter 4, since we have demonstrated the ability of a Rieske oxygenase, *Ph*CAO, to produce the non-natural Chlorophyll pigment, it would be extremely exciting to apply the aforementioned design principles to engineer CAO to install an aldehyde group at a desired position on the Chlorophyll scaffold. Also, several substrates screened in chapter 4 showed that CAO had a relatively narrow substrate scope. It would be sensational if the engineered CAO could accept a broader substrate scope to synthesize different natural and non-natural pigments, those of which are utilized broadly in the cosmetic, food, agricultural, and pharmaceutical industries. This enzymatic reaction also serves as a convenient and green method to synthesize late-stage oxidative chlorophyll species that otherwise would be synthetically complicated to obtain.

As for Chlorophyll, Chlorophyll *a* has been demonstrated to be involved in harvesting light energy and in driving electron transfer in the reaction center whereas Chlorophyll *b* is only engaged in light harvesting. It is also known that the ratio of Chl *a* to Chl *b* is higher under high light growth conditions than in low light growth conditions[13-14]. Therefore, the interconversion of Chl(ide) *a* and Chl(ide) *b*, or the so-called Chl cycle, plays a crucial role in various developmental stages in green plants by tunning the ratio of Chl *a* to Chl *b*. Previous studies suggested that CAO, CBR (Chlorophyll(ide) *b* reductase) and HCAR (7-Hydroxymethyl-Chlorophyll(ide) *a* reductase) are involved in the Chl cycle, but there is currently a lack of compelling *in vitro* data to show their specific roles. In chapter 4, we demonstrated that CAO catalyzes two sequential monooxygenation reactions on a Chlorophyll *a* precursor, Chlorophyllide *a*, to transform it into Chlorophyllide *b*. Further studies should focus on elucidating how CBR and HCAR convert Chlorophyll(ide) *b* back to Chlorophyll(ide) *a* and how Chlorophyll(ide) *b* levels are sensed to regulate CAO stability and activity. Likewise, as there is not a homolog of the non-native reductase used in our work, future studies need to address the identity of the physiological reductase for the formation of chlorophyll *b*.

Finally, it is astonishing how Rieske oxygenases are able to catalyze so many different types of chemistry, including C-C bond cleavage, oxidative decarboxylation and carbocyclization etc. Recent work from Lipscomb group showed that Rieske aromatic monooxygenase and dioxygenase used the same reactive species, an Fe(III)-superoxo or its electronic equivalent form, to initiate the catalysis[15]. Therefore, further mechanistic studies on these enzymes to identify the possible activated oxygen intermediates and catalytic cycles that give rise to the different options, particularly for those that oxidize non-aromatic compounds, would definitely be one of the most important future directions in this field. Notably, as for hydroxylation oxygenation reactions, in particular, Rieske oxygenases can catalyze mono-, di-, or even sequential oxygenation reactions in the absence of an auxiliary cofactor. It would also be fascinating to see what structure features, especially for the non-heme iron site, of these Rieske oxygenases are responsible for the difference in activity.

5.3 References

1 Stryer, L.; Berg, J.M.; Tymoczko, J.L. (**2002**) *Biochemistry* (5th ed.)

2 Pasteur, L. C.R. *Seances Acad. Sci* (**1958**) 46, 615.

3 Hughes, G.J; Lewis, J.C. Introduction: Biocatalysis in Industry *ACS Chem. Rev.* (**2018**) 118, 1-3

4 Truppo, M.D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed *ACS Med. Chem. Lett.* (**2017**) 8, 476-480

5 Chen, K.; Arnold, F.H. Engineering new catalytic activities in enzymes *Nature Catalysis* (**2020**) 3, 203-213

6 Kress, N., Halder, J. M., Rapp, L. R. & Hauer, B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* (**2018**) 47, 109-116.

7 Banerjee, R. & Lipscomb, J. D. Small-Molecule Tunnels in Metalloenzymes Viewed as Extensions of the Active Site. *Acc. Chem. Res*. (**2021**) 54, 2185-2195.

8 Kokkonen, P., Bednar, D., Pinto, G., Prokop, Z. & Damborsky, J. Engineering enzyme access tunnels. *Biotechnol. Adv*. (**2019**) 37, 107386.

9 Escalante, D. E., Aukema, K. G., Wackett, L. P. & Aksan, A. Simulation of the bottleneck controlling access into a Rieske active site: Predicting substrates of naphthalene 1,2-dioxygenase. *J. Chem. Inf. Model* 57, 550-561 (**2017**).

10 Nestl, B. M. & Hauer, B. Engineering of Flexible Loops in Enzymes. *ACS Catal.* (**2014**) 4, 3201-3211.

11 Dodani, S.C.; Kiss. G.; et al. Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nature Chemistry* (**2016**) 8, 2474

12 Garcia-Borras, M.; et al. Origin and Control of Chemoselectivity in Cytochrome c Catalyzed Carbene Transfer into Si–H and N–H bonds *J. Am. Chem. Soc.* (**2021**) 143, 18, 7114-7123

13 Kunugi, M., Takabayashi, A., Tanaka, A. Evolutionary changes in chlorophyllide a oxygenase (CAO) structure contribute to the acquisition of a new light-harvesting complex in micromonas. *J. Biol. Chem*. (**2013**) 288, 19330–19341

14 Tanaka, A.; Ito, H.; Tanaka, R.; Tanaka, N. K.; Yoshida, K.; Okada, K., Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proc Natl Acad Sci USA* (**1998**) 95 (21), 12719-23

15 Rogers, M.S., Lipscomb, J.D. Salicylate 5-Hydroxylase: Intermediates in Aromatic Hydroxylation by a Rieske Monooxygenase. *Biochemistry* 58, 5305-5319 (**2019**)