2020-01-01

Structural Basis For The Stereoselective Epoxidation Reaction In Polyether Natural Product BiosynThesis

Qian Wang
University of Texas at El Paso

STRUCTURAL BASIS FOR THE STEREOSELECTIVE EPOXIDATION REACTION IN POLYETHER NATURAL PRODUCT BIOSYNTHESIS

QIAN WANG

Doctoral Program in Chemistry

APPROVED:

Chu-Young Kim, Ph.D., Chair

Ricardo A Bernal, Ph.D.

Wen-Yee Lee, Ph.D.

Suman Sirimulla, Ph.D.

Stephen Crites, Ph.D.
Dean of the Graduate School

STRUCTURAL BASIS FOR THE STEREOSELECTIVE EPOXIDATION REACTION IN POLYETHER NATURAL PRODUCT BIOSYNTHESIS

by

QIAN WANG

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry

THE UNIVERSITY OF TEXAS AT EL PASO

May 2020

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Chu-Young Kim, for accepting me into his laboratory and for his support and guidance over these past few years. I have been very lucky to have a supervisor who responded to my questions so promptly, and guided me solving problems so patiently. I deeply appreciate all the help that Dr. Kim has given to me whenever I had a difficult time.

I would also like to thank all my committee members, Dr. Ricardo A. Bernal, Dr. Wen-Yee Lee, and Dr. Suman Sirimulla for their insightful advice in my research. Particularly, I would like to thank Dr. Bernal for kindly letting me use his lab resources, and also for all the encouragement and advice in my career development.

I also want to express my gratitude to all the current and previous members of the Kim group. They have shared with me their scientific knowledge, experience, as well as friendship. I would like to thank Aparna Jain and Dr. Zilong Wang for guiding and helping me when I just joined the Kim group, Olivia Fernandez Delgado for helping me with organic chemistry problems that I encountered in my research, and Saket R. Bagde and Priyanka Gade for their constructive advice. Their endless encouragement and enthusiasm for research science has motivated me to overcome challenging times.

My research work could not be completed without the help provided by all my collaborators. First, I would like to thank Dr. Xi Chen at the Northwest University in China for her help in teaching me patiently about solving and refining the protein crystal structures. She and her master student Yaming Deng have synthesized the substrate analog for my polyether biosynthesis research. I also want to thank the beamline scientist, Dr. I. Irimpan Mathews at Stanford University for his help and advice in data collection and structure determination. I would like to thank Dr. Wen-Yee Lee and her previous graduate student Dr. Qin Gao for helping me with GC-MS analysis. Also, I would like to thank Dr. Lela Vukovic and her student Tara A. Nitka for their contribution in protein docking and dynamic simulation. Also, I would like to thank Dr. Alejandro Metta from

our X-ray facility for helping me screening the protein crystals. It was definitely a pleasure to work together with so many talented and dedicated scientists.

Last but not least, I gratefully acknowledge the funding towards my PhD from Department of Chemistry and Biochemistry teaching assistantship, and also UTEP Graduate School Scholarship.

Abstract

Polycyclic polyether natural products have received much attention due to their diverse biological activities, ranging from extreme toxicity to therapeutic properties, including antimicrobial, antifungal, and anticancer activities. They are featured by multiple cyclic ether groups that vary in the number, size, and arrangement. Ionophore polyethers contain multiple tetrahydrofuran and tetrahydropyran rings connected by either spiroketal systems or carbon-carbon single bonds. Over the past 40 years, significant progress has been made in deciphering polyether biosynthesis pathways. Remarkably, all members of the polycyclic polyether family are thought to be generated via a common biosynthetic scheme. In 1983, Cane, Celmer, and Westly proposed a biosynthetic model for producing the stereospecific ether rings of polycyclic polyethers. It postulates that linear polyene intermediates undergo a cascade of enzyme-catalyzed reactions to form the polyether rings, including stereospecific epoxidation and epoxide ring opening reactions.

In 2001, the biosynthetic gene cluster for polyether monensin was isolated from *Streptomyces cinnamonensis*. It contains the genes of a flavin-dependent epoxidase MonCI and two epoxide hydrolases MonBI and MonBII. Gene deletion and feeding experiments showed that MonCI catalyzes all three stereoselective epoxidations for the intermediate premonensin, while MonBI and MonBII are responsible for the following epoxide ring opening reactions. In 2008, the gene cluster responsible for the biosynthesis of lasalocid was identified from *Streptomyces lasaliesis*. It also includes the genes of a flavin-dependent epoxidase Lsd18, and an epoxide hydrolase Lsd19. Similar to MonCI, Lsd18 transforms two olefins in the intermediate prelasalocid into two epoxides in a stereoselective manner. Also similar to MonBI/MonBII, Lsd19 catalyzes epoxide ring opening reactions, generating the final product lasalocid with a THF-THP construction. So far, the mechanism of epoxide ring opening reactions has been extensively studied well. The Lsd19-substrate/product complex crystal structures revealed how energetically favored 5-*exo* cyclization and energetically unfavored 6-*endo* cyclization are catalyzed. Also, gene mutagenesis and the crystal structure determination of MonBI/MonBII further explained the

synergistic effects and allosteric regulation between the epoxide hydrolases. In contrast, as the first step of the polyether ring construction, the mechanism of stereoselective epoxidation is poorly understood due to the lack of epoxidases' molecular structures.

My research project is focused on determining the molecular mechanism of stereospecific epoxidation catalyzed by MonCI and Lsd18. Recombinant MonCI and Lsd18 was successfully expressed in *E.* coli BL21(DE3) and purified using standard chromatography techniques. These proteins were shown to be active by in vitro enzyme assays that included the extracted proteins MonCI/Lsd18, substrates, FAD, and reducing agent nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH). The epoxide products were detected by either GC-MS or LC-MS. We solved the MonCI crystal structure at 1.90 Å resolution and the Lsd18 ligand-free and ligand-bound complex crystal structures at 1.54 and 1.85 Å, respectively by X-ray crystallography. MonCI and Lsd18 share high similarity in both sequence (identity = 47.8%) and three-dimensional structure (RMSD = 1.63). Both MonCI and Lsd18 structures contain a fused FAD- and substrate-binding pocket with three openings. Presumably, one for substrate, one for solvent, and one for FAD reduction/NAD(P)H binding. Since they share high structural similarity to the *p*-hydroxybenzoate hydroxylase, especially the FAD-binding domain, the mechanism of FAD reduction and oxidation process is expected to be similar. The catalytic cycle consists of five steps. First step is FAD reduction. When the isoalloxazine ring of oxidized FAD moves out from the pocket ("out" conformation), it is exposed to the solvent and accepts electrons from NAD(P)H. After reduction, FAD moves back to the pocket ("in" conformation) and transforms into the intermediate C4a-(hydro)peroxyflavin by reacting with molecular oxygen. Then C4a-(hydro)peroxyflavin donates one hydroperoxyl oxygen atom to the substrate (oxidation). The final step is accomplished at the end by FAD eliminating one water molecule and turning back to the oxidized status. To investigate how epoxidases that are involved in the polyether biosynthesis achieve stereoselectivity, we conducted the docking and modeling work on MonCI and Lsd18. A molecular dynamic simulation was also performed on MonCI with its native substrate premonensin. Our findings suggest that the stereoselectivity of epoxidation is determined by the

unique preorganization of the substrate-binding pocket which permits only one face of the alkene to approach the reactive C4a-(hydro)peroxyflavin. Several key residues at the active site that play an important role in stereoselectivity have been identified and mutated to further test our hypothesis.

Overall, our crystallographic and computational studies have provided important molecular insights into how the stereoselective epoxidation is achieved by the flavin-dependent monooxygenases MonCI and Lsd18. Due to the high sequence similarity, all epoxidases that are involved the biosynthesis of polyether natural products could are likely to employ the same catalytic mechanism. Our results are expected to help bioengineering efforts for bioproduction of polyether natural product analogs for drug research.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: Background and Introduction

1.1 Natural products in life sciences

Natural products are the organic compounds produced by living organisms.(1) They can also be prepared by chemical synthesis. In fact, the fields of chemical synthesis including semi-synthesis and total synthesis, often target natural products and their derivatives.

Natural products are often divided into two major classes: primary and secondary metabolites. Primary metabolites are molecules which participate in metabolic pathways that are essential for survival of the organism, such as energy production, nutrient assimilation, and growth. Examples of primary metabolites include nucleic acids, amino acids, lipids, and carbohydrates. Secondary metabolites are molecules that are not essential for organism to survive but provide competitive advantages. Secondary metabolites act as social signaling, communication molecules, agents that solubilize and transport nutrients, and competitive weapons like repellants, venoms, and toxins. Some common examples of the secondary metabolites are antibiotics, growth factors, peptides, and nucleosides.(2) Secondary metabolites can be divided into alkaloids, phenypropanoids, polyketides, and terpenoids. Alkaloids contain basic nitrogen atoms. Phenylpropanoids are synthesized from phenylalanine and tyrosine. Polyketides are assembled from acetate and malonate. Terpenoids are constructed from the five-carbon monomer building blocks called isoprene.

Natural products exist almost everywhere in the life kingdoms including bacteria, archaea, fungi, plants, and animals. They can be extracted from cells, tissues, and secretions.(3) Diversity of their chemical structures contributes to the diversity of their biological activities. Many natural products have been extracted, identified, and processed for medical use. For example, many anti-bacterial and anti-fungal agents were found in bacteria, including tetracycline (from *Streptomyces spp*.)(4), rifamycines (from *Amycolatopsis rifamycinica*)(5), and polymyxins (from *Paenibacillus polymyxa*)(6). The glycopeptide bleomycin for cancer treatment originates from *Streptomyces verticillus*.(7) The neutrotoxin botulinum responsible for botulism is produced by *Clostridium*

botulinum.(8) Natural products not only come from bacteria but also come from fungi, plants, and animals. The first antibiotic penicillin is derived from the *Pencillium* fungus.(9) The antihypertensive drug teprotide was discovered in the venom of the Brazilian arrowhead viper *Bothrops jaraca*.(10) The antineoplastic drug trabectedin is from the sea squirt *Ecteinascidia turbinate*.(11) There are also natural products from the plants, such as the antimalarial drug artemisinin (12) and the antineoplastic drug taxol (13).

1.2 POLYKETIDE NATURAL PRODUCTS

Polyketides are one of the major natural product classes. They are assembled from acetate and malonate building blocks into complex structures, and they exhibit a wide range of structural and functional diversity. They display a variety of therapeutically important activities such as antibiotic, anticancer, antimalaria, and immunosuppressive properties.(14, 15) At the same time, many polyketides are also infamous as food-spoiling toxins or virulence factors.(16) Questions like how polyketides possess highly diverse structures and functions from very simple building blocks have triggered many researchers to investigate their biosynthetic mechanisms by chemical, genetic, and biochemical experiments.

1.2.1 Macrolides

The macrolides are a class of polyketide natural products that contain a large 14-, 15-, or 16-membered macrocyclic lactone ring to which deoxy sugars such as cladinose and desoamine are attached. Some macrolides possess antibiotic or antifungal activities and have been developed into successful pharmaceuticals. Pikromycin, the first antibiotic macrolide, was isolated in 1951.(17) Other macrolide antibiotics include erythromycin A, clarithromycin, and azithromycin. Macrolides avermectin and ivermectin are used as antihelminthics and spinosad is used as an insecticide.

1.2.2 Ansamycins

Ansamycins are macrolides with antimicrobial activities again many Gram-positive and Gram-negative bacteria. It was first discovered in 1959 from an actinomycete bacterial strain

Amycolatopsis mediterranei.(18) They are constructed uniquely with an aromatic moiety connected by an aliphatic chain. The variation among ansamycins are the aromatic moiety which can be a naphthalene ring or a napthoquinone ring as in rifamycin and the napthomycin.(19)

The benzene or benzoquinone ring system can also be a variation in geldanamycin or ansamitocin. Examples of ansamycins are rifamycins that have high potency against mycobacteria, which have been widely used in treating tuberculosis, leprosy, and mycobacterial infections.(20)

1.2.3 Acetogenins

Acetogenins are a class of polyketide natural products found in the Annonaceae family of plants. They are characterized by linear 32- and 34-carbon chains containing oxygenated functional groups including hydroxyls, ketones, epoxides, tetrahydrofurans, and tetrahydropyrans. They are often terminated with lactone or butanolide. Over 400 members of this family of compounds have been isolated from 51 different species of plants. Many acetogenins are characterized by neurotoxicity.

1.2.4 Polycyclic polyethers

The polycyclic polyethers from plants, microorganisms and marine organisms are identified by their regular occurrence of multiple C-O-C motifs. They have gathered much interest due to their fascinating structural diversity and diverse biological activities, including antibiotic, anticancer, and antimalaria to extreme toxicity.(21, 22) Based on their structural features and origin, the natural polycyclic polyethers can be divided into two groups. First group includes molecules with multiple fused cyclic ethers, which are constructed by *all-endo* cascade of epoxide openings. The second group consists of molecules characterized by their multiple rings interconnected by carbon-carbon bonds. They are produced through an *all-exo* biosynthetic cascade of epoxide openings.

1.2.4.1 Polyether ionophores

The polyether ionophores are lipophilic carboxylic acids with multiple five- and six-membered cyclic ethers as either spiroketal systems or linked bicyclic ethers (Figure 1.1). Since

the first ionophores nigericin and lasalocid were isolated in 1951, over 120 naturally occurring ionophores have been reported.(23) Most of the ionophore natural products are from the *Streptomyces* genus. These ionophores are used to control coccidiosis and as growth promoters in ruminant animals, which can specifically target the ruminal bacterial population. Commercially available ionophores include monensin **1**, **2** (Coban, Rumensin, and Coxidin), lasalocid **3** (Avatec, Bovatec), maduramicin **4** (Cygro), narasin **5** (Monteban, Maxiban), and salinomycin **7** (Bio-cox, Sacox).(24)

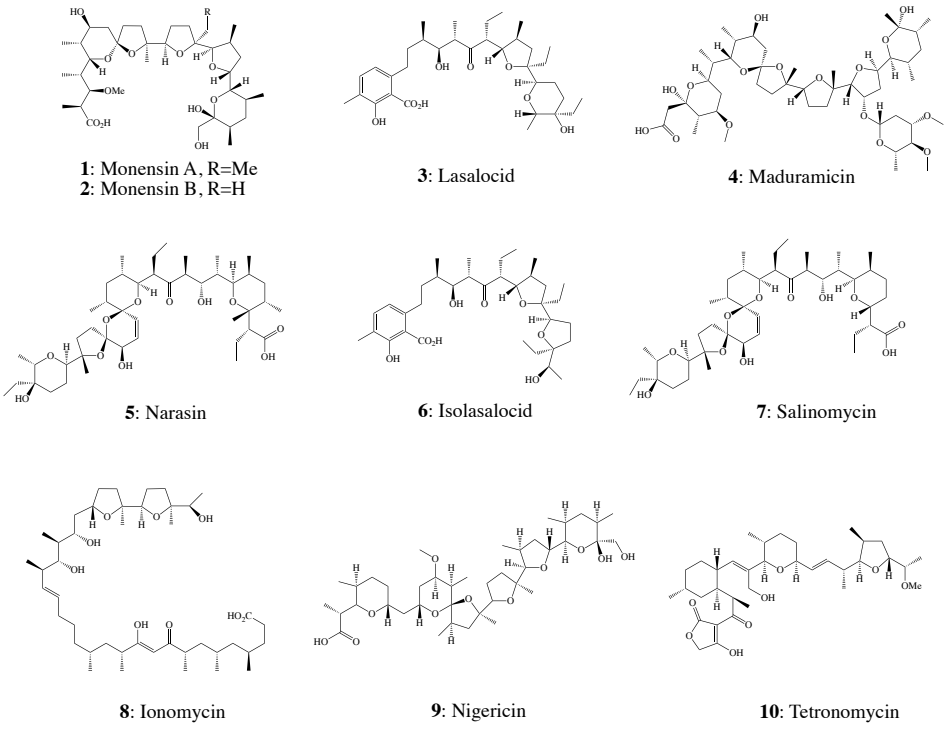


Figure 1.1: Structures of representative polyether ionophores

Biological activities of ionophore polyethers are closely related with their ability to bind metals and transport them across the bacterial cell membrane.(25) The physiological pH in ruminants is slightly acidic while the intercellular pH of most of the ruminal bacteria is near neutral. This makes an inwardly directed proton gradient. The ruminal bacteria also maintain ionic gradients across the membrane to take in nutrients efficiently. For example, it maintains a high

intracellular potassium concentration and a low sodium concentration compared with the ruminal environment. Due to their lipophilicity, polyether ionophores are able to insert into the bacterial membrane easily. Once ionophores form a complex with the metal cations like potassium or sodium, the ether oxygen of the ionophore loses its solvated water molecules and a neutral zwitterionic metal-ionophore complex is formed.(26) Then transportation though the membrane can take place. At the opposite side of membrane, the anionic ionophores are formed again after releasing the metals. It needs to bind another cation, usually a proton, to return to the initial side of the membrane.(27) The ionophores have different preferences to the metal cations. For example, monensin A **1** can bind with potassium more stable than sodium, but its transport of sodium is much faster than potassium.

1.2.4.2 Oxasqualenoids

Oxasqualenoid natural products exist in tropical plants, marine sponges and red algae.(28) They are polycyclic polyethers derived from squalene and similar to polyether ionophore structures, and contain regularly occurring tetrahydrofuran rings connected by carbon-carbon bonds (Figure 1.2). Many triterpenic polyethers have been discovered with structural and pharmacological diversities. For example, dehydrothyrsiferol **13** from the red seaweed *Laurencia viridis*, is identified as a lead compound with moderate to potent cytotoxic activities.(29) Another bromotriterpene polyether aurilol **15**, is isolated from the Indian Ocean sea hare *Dolabella auricularia*.(30) Relying on the similarity with polyether ionophores in structure and the discoveries of the steroid biosynthesis from squalene epoxide, the biogenesis of the oxasqualenoids was also proposed.(31, 32)

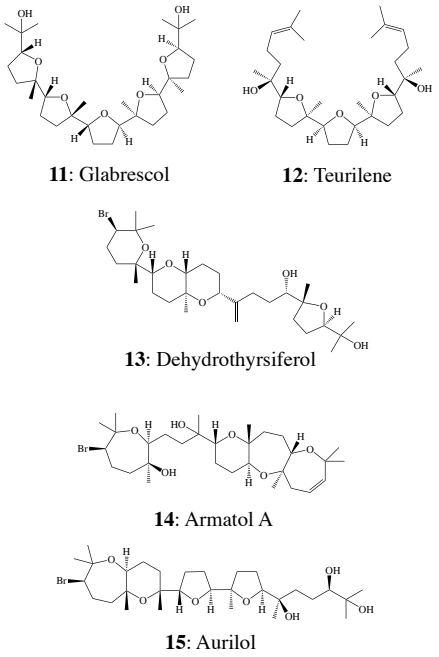


Figure 1.2: Structures of representative oxasqualenoids

1.2.4.3 Ladder polyethers

Ladder polyethers feature 4 to 32 five- to nine-membered membered cyclic ethers fused which are in a *trans-syn-trans* arrangement, thus result in a repeating C-C-O sequence that stretches throughout their polycyclic core (Figure 1.3). The first ladder polyether, brevetoxin B, was isolated in 1981 by Nakanishi and Clardy.(33) A large number of other ladder polyethers were also discovered including the smallest member, hemibrevetoxin B **19** with only four rings and the largest member, maitotoxin **16** with four polyether ladders and unequaled thirty-two rings.(34-37)

The ladder polyethers are best known for their association with harmful algal blooms, commonly referred to as red tides, which involve a rapid increase of algae concentration in coastal areas.(38) For example, *Karenia brevis*, the causative agent of algal blooms, produces the ladder polyethers that are cytotoxic and cause devastating death of fish and marine mammals. The marine species unaffected by red tide toxins can accumulate toxins, moving them upward in the food chain and causing poisoning in humans.(39, 40) Despite the effects in red tides, ladder polyethers also exhibit a variety of biological activities. Many members of this group show high binding affinity

to sodium, potassium, and calcium ion channels, which would disrupt these channel functions.(41) Some of the ladder polyethers also possess anticancer (42) and antifungal (43) properties.

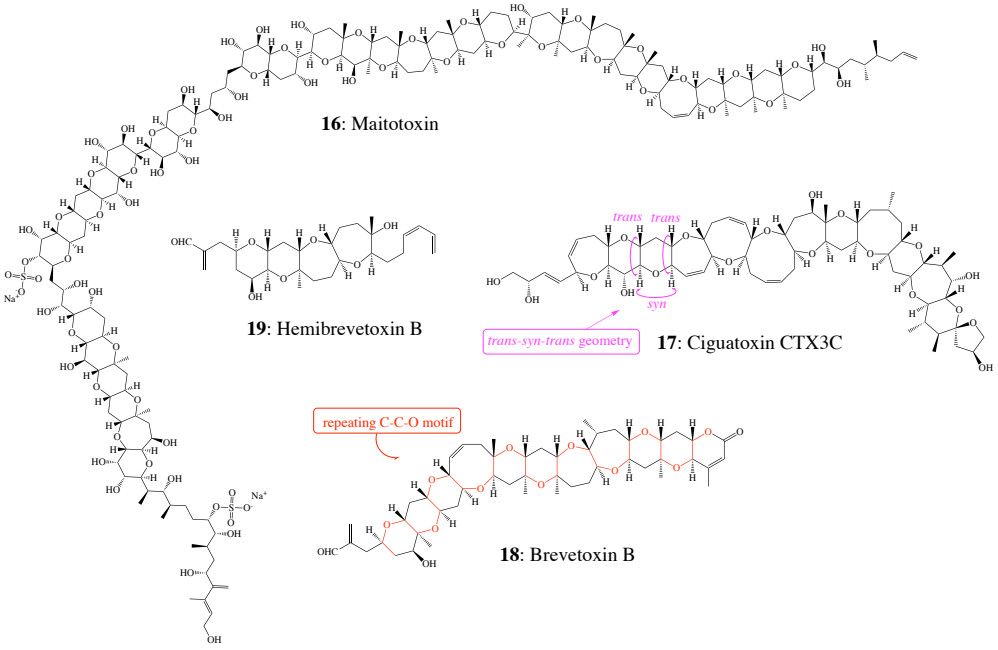


Figure 1.3: Structures of representative ladder polyethers

Based on the structure of brevetoxin B **18** and comparison to the well-studied polyether ionophores like monensin **1**, **2**, lasalocid **3**, and nigericin **9**, Nakanishi (44) and Shimizu (45) both proposed a hypothesis that the structural and stereochemical uniformity of ladder polyethers arise from epoxidation and an epoxide-opening cascade that transforms the polyepoxide into a ladder polyether.(46) The C-C-O backbone of ladder polyethers, which is shown in the figure 1.3, comes from the oxygen and carbon atoms in the epoxide rings. All ring openings take place with inversion of configuration at each epoxide that is derived from an *E* alkene. This explains how the *trans-syn-trans* topography is generated. Since all alkenes of the precursor polyene needs identical enantiofacial selectivity of epoxidation to produce either all-(*S*,*S*) or all-(*R*,*R*) polyepoxide, a single promiscuous oxidase should be sufficient for this function.(37) Feeding experiments using oxygen-18 in either molecular oxygen or acetate suggested that the ether oxygens of ladder polyethers come from molecular oxygen while the hydroxyl oxygen is derived from acetate, which

supports Nakanishi and Shimizu's hypothesis that the precursor polyene is oxidized from head to tail (or right to left) by a monooxygenase after acetate condensation.(47-49)

1.2.5 Polyketide biosynthesis

Polycyclic polyethers belong to a large family of polyketide natural products. Even though they exhibit extremely high diversity in structure and biological activity, all polyketides are constructed by successive rounds of decarboxylative Claisen condensations between a thioesterified malonate derivative and an acyl thioester. The enzymes responsible for these condensations are called polyketide synthases (PKSs) which are classified into three types I, II and III.(15, 50-53)

Type I PKSs consist of two groups: modular and iterative PKSs. The modular system contains a set of catalytic domains that are covalently linked as modules responsible, each module being responsible for the incorporation of one precursor molecule into the polyketide backbone. The typical example of PKSs is 6-deoxyerythronolide B synthase which assembles the 6-dexyerythronolide B moiety of erythromycin A.(54) A PKS module has at least three core domains: an acyl carrier protein (ACP), an acyltransferase (AT), and a ketosynthase (KS) for extending the polyketide chain by two carbon atoms. The ACP domain utilizes a phosphopantetheine arm and thioester bond to bind the growing polyketide chain and the extender unit and transfer them to other domains. The AT domain recognizes a specific extender unit and tethers it to the sulfhydryl group of the ACP prosthetic arm. The KS domain catalyzes carbon-carbon formation between the ACP-linked malonate derivative and the other ACP-linked acyl thioester through the decarboxylative Claisen condensation reaction. In addition to the KS-involved condensation, a PKS module may also contain other catalytic domains to perform different kinds of β-keto group modification. Examples of such domains include ketoreductase (KR), dehydratase (DH), and enoylreductase (ER). Once the polyketide intermediate is fully processed, the ACP domain transfers the elongated polyketide chain to the KS of the subsequent module for another round of elongation or to the thioesterase (TE) domain for release of the final

product through hydrolysis or cyclization. Iterative type I PKSs are similar to the modular PKSs in that they use the same set of catalytic domains for transferring, extending, and modifying polyketide chain. The main difference between the two PKS types is that in iterative type I PKSs, a single module acts on the substrate repetitively.(55) The biosynthetic pathway of the polyketide lovastatin is a typical iterative PKS, all 35 reactions are conducted by a single iterative dihydromonacolin L synthase.(56, 57)

Type II PKSs contain the same KS, AT, KR and DH domains like type I PKSs, but typically contains two KS domains: KS_α domain same with one in type I PKSs, and KS_β domain controlling polyketide length. Additionally, reduction of the β-keto group takes place only after the polyketide backbone is completely synthesized.(58) The anticancer drugs daunorubicin and doxorubicin are synthesized by type II PKS proteins.(59, 60)

Type III PKSs, which are distinct from type I and II PKSs, contain iterative domains as homodimers, and they are ACP independent. They act directly on acyl-Coenzyme A (CoA) thioesters and transfer the acyl group between CoA and an active site cysteine. The most well-studied type III PKS is the plant chalcone synthase, which condenses three acetate units to a 4-coumaroyl-CoA starter molecule which is cyclized to form the aromatic tetraketide product naringenin chalcone.(53)

1.3 SYNTHESIS OF POLYETHER IONOPHORES

1.3.1 Total synthesis

Over the past decades, much effort has been made to synthesize polyether ionophore antibiotics. So far, about 16 polyether ionophore natural products have been successfully prepared by total synthesis.(61, 62) The tricky part is to synthesize the polyether rings (tetrahydrofurans and tetrahydropyrans) and the spiroketal systems. The landmark total synthesis of the polyether lasalocid A was first reported by Kishi's group.(63, 64)

To synthesize polypropionates, several acyclic methods have been developed. For example, allylic 1,3 strain (A-1,3) was used to synthesize propionates and form tetrahydrofuran

through hydroboration, haloetherification, and epoxidation reactions.(65) Sharpless asymmetric epoxidation was used to produce enantiometrically pure propionate fragments.(66, 67) The Cram- and Cram-chelate-controlled addition of nucleophiles to carbonyls were also introduced by Kishi and Still to build the propionate fragments.(68) Later on, a more versatile and efficient method, chiral enolate bond construction was developed by Evans.(69) Besides, there were also several strategies developed for fragment coupling. The most popular and important reaction for fragment coupling is the aldol reaction. However, there were other reactions that have been used for the fragment couplings and the formation di- and trisubstituted carbon double bonds, including Julia olefination, Witting, and Diels-Alder reactions.

Total synthesis of lasalocid A and its isomer isolasalocid A, first polyether ionophores that were prepared by total synthesis, was achieved in 1978.(63) Lasalocid A and isolasalocid A both contain a central tetrahydrofuran B-ring and 10 chiral centers. But lasalocid A has a tetrahydropyran C-ring while isolasalocid A has a tetrahydrofuran C-ring. The total synthesis of lasalocid A involves separate synthesis of C_1-C_{11} and C_{12}-C_{24} fragments. The C_1-C_{11} fragment was prepared in two steps. The C_{12}-C_{24} fragment was prepared by two complimentary approaches using the highly regioselective and stereoselective addition of alcohols to epoxides.(70) In one approach, A-1,3 interactions was used to prepare the tetrahydrofurans by epoxidation of bishomoallylic alcohols and stereoselective reduction of ketones that is followed by epoxidation ring opening reactions. In the other alternative approach, the C_{19} alcohol was prepared by the epoxidation of the olefin followed by the reduction. The absolute stereochemistry of the intermediates was achieved by resolution, chiral pool, equilibration, Cram- and Cram-chelate-controlled addition. These two fragments were coupled together by aldol reaction.

The total synthesis of monensin A was also reported by both Kishi and Still in 1979.(71-73) Similar to lasalocid A, total synthesis of monensin A involves separate preparation of the C_1-C_7 and C_8-C_{26} fragments. A new method using hydroboration reaction was developed by Kishi for propionate synthesis. In the Kishi's total synthesis method (71, 72), the preparation of the C_1-C_7 fragment is achieved in six steps in total, which includes furan ring ozonolysis, alcohol oxidation,

and the hydroxyl protection. Synthesis of the C_8-C_{25} fragment is achieved in 15 steps, using epoxidation of bishomoallylic alcohols. The absolute stereochemistry was controlled by resolution, A-1,3 and thermodynamic equilibration. In Still's total synthesis method (74, 75), three fragments C_1-C_7, C_8-C_{15}, and C_{16}-C_{25} were prepared first. The Cram- and Cram-chelate nucleophilic addition reactions were heavily used to generate the 7 stereogenic centers of monensin A. The C_1-C_7 fragment was prepared in 10 steps through two sequential aldol reactions. The C_8-C_{15} fragment was prepared 16 steps by chelate-controlled addition, diol protection, desilylation, and bromination reactions. The fragments were coupled together at the end by aldol reactions.

Total synthesis of several other polyether natural products have been reported, including calcimycin (1979) (69), indanomycin (1981) (76), narasin and salinomycin (1981) (77), and zincophorin (1987) (78). However, there still remains some problems in synthesizing the polyether ionophores, including controlling the stereoselectivity, separating the isomers, and coupling fragments. The total yield of the final products is still very low. More efficient and predictable methods need to be developed in the future.

1.3.2 Biosynthesis

1.3.2.1 Unified model

Polyether ionophores belong to the polyketide family of natural products. Soil bacteria such as *Streptomyces* strains produce various polyether ionophores for the purpose of killing off their competitors and compositing for water, food, and other limited resources (79). So far, over hundreds of polyether natural products have been discovered in bacteria. For example, monensin A is produced by *Streptomyces cinnamonensis*, lasalocid A is produced by *Streptomyces lasaliensis*, salinomycin is produced by *Streptomyces albus*, and nigericin is produced by *Streptomyces violaceusniger*.

All polyether-producing bacteria contain a gene cluster for polyether ionophore biosynthesis. The genes in this cluster code for various enzymes responsible for biosynthesis of polyether ionophores. Biosynthesis of ionophore polyethers includes two reactions: synthesis of

linear polyene and formation of ether rings. The synthesis of linear polyene is carried out polyketide synthases. The formation of ether rings is carried out by epoxidases and hydrolyses.

A series of feeding experiments on ionophore-producing organisms revealed their polyketide biosynthetic origin.(80-82) It also indicated that their oxygen atoms come not only from the building blocks acetate, propionate, and butyrate but also from the molecular oxygen. In 1983, based on all the previous experimental results, Cane, Celmer, and Westly proposed a biosynthetic model (CCW model) that describes how polycyclic polyether skeletons with various stereocenters is constructed.(83) It suggests that a chiral polyene first undergoes sequential enantioselective epoxidation to install all the necessary stereochemistry, and then the resulting polyepoxide serves as a substrate of epoxide opening cascades to give elaborated polyethers. The first class of polyethers including monensin **1**, **2**, dianemycin, and lenoremycin, which has the identical tetrahydropyranyl rings at the termini, are produced with the building blocks as acetate, propionate, propionate, and acetate (APPA). The second class of polyethers, including lasalocid **3** and isolasalocid **6**, the building blocks vary as (PABA, BABA, and PAPA) and give either a tetrahydrofuran linked to a tetrahydropyran, or two adjacent five-membered ether rings at the termini. This was the first time that a unified stereochemical model was proposed to summarize the biosynthesis of a large number of polyethers, and it leads to the consideration that each polyene prototype is constructed by enzymes expressed by the gene cluster corresponding to the polyene chain-building sequence.

In 2001, Leadlay and co-workers identified a gene cluster for monensin biosynthesis, which consists of twelve polyketide synthase modules, MonAI to MonAX, responsible for building the twelve acyl units of monensin, one epoxidase gene MonCI, and two epoxide hydrolase genes MonBI and MonBII.(84) After a series of gene inactivation experiments in monensin biosynthesis (85, 86), the CCW model was further improved and finally established. Later on, both Oikawa's and Leadlay's group reported a gene cluster of lasalocid biosynthesis pathway containing an epoxidase Lsd18 (also called LasC) and an epoxide hydrolase Lsd19 (also called LasB), which also supports the CCW model.

1.3.2.2 Biosynthesis of monensin

Monensin A and B are the polyether ionophores that are produced by *Streptomyces cinnamonensis*. They can dissipate ionic gradient across cell membranes and are widely as food additive in poultry to prevent coccidiosis.(87) Early feeding experiments indicate that monensin A **1** is produced from five acetate, seven propionate, and one butyrate precursors, while monensin B **2** requires an extra propionate replacing the butyrate.(88) Four out of their nine oxygen atoms come from molecular oxygen while the other five are derived from the corresponding carboxylic acids.

In 2001, the gene cluster for monensin biosynthesis was identified by Leadlay's group (Figure 1.4).(84) It contains a twelve-module polyketide synthase system responsible for constructing the linear triene premonensin **20**. The loading module has an N-terminal KSQ domain that act as a malonyl-CoA decarboxylase to generate starter units *in situ*. The monensin PKSs don't have an integrated C-terminal thioesterase at the end of module 12 for the polyketide chain release. The enzymes encoded by *monCII*, *monAIX*, and *monAX* genes have been shown to function as thioesterases based on their gene sequence alignment and deletion experiments.(85, 89)

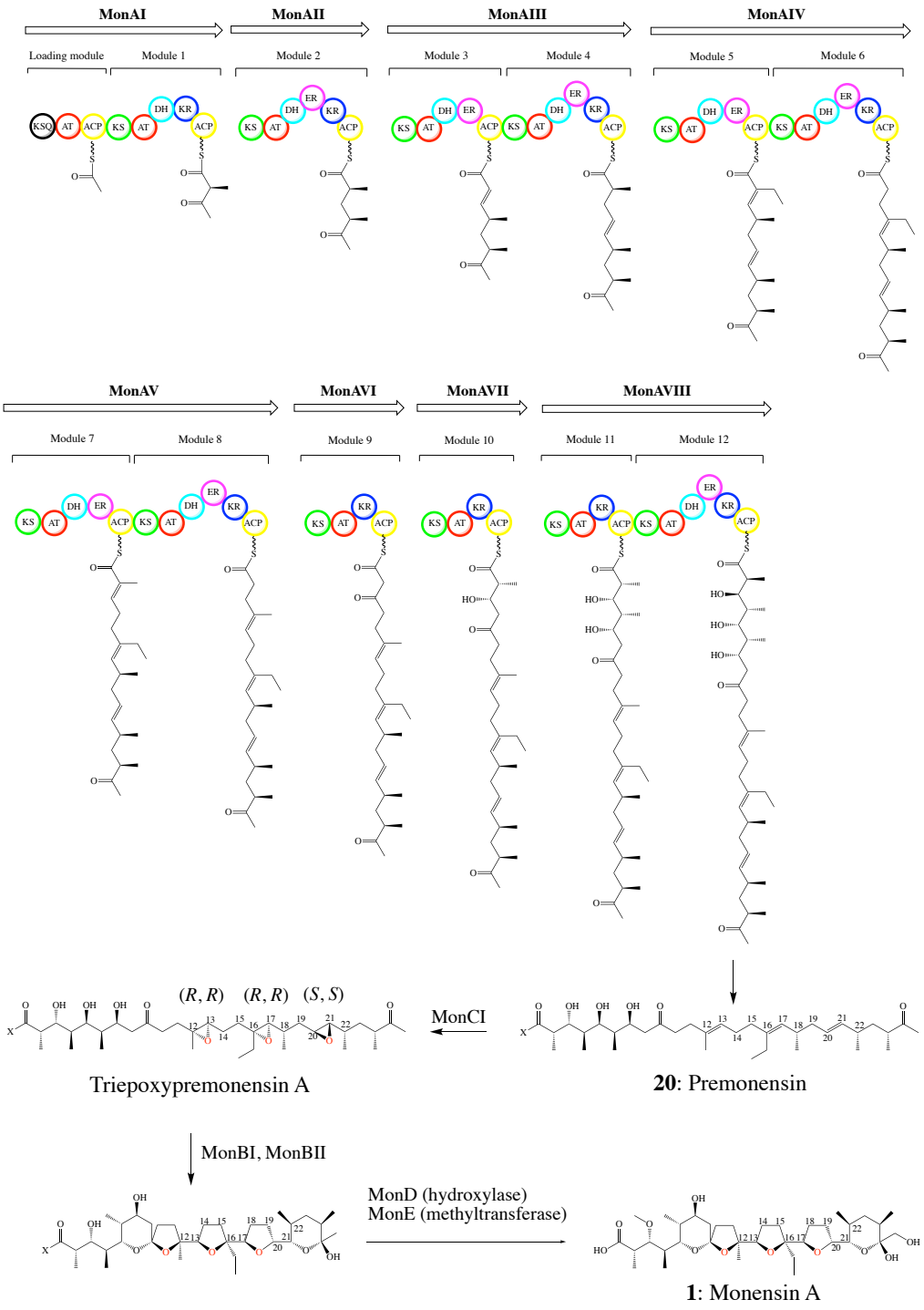


Figure 1.4: The proposed biosynthesis pathway of monensin. X= OH, SCoA, or S-protein.

After released from the PKSs, epoxidation and epoxide-opening cascades of premonensin take place. The *monCI* gene encodes a flavin-dependent monooxygenase, which is responsible for the epoxidation of all three carbon double bonds in premonensin **20**. Deletion of *monCI* gene

resulted in complete loss of the final product monensin and the accumulation of the linear *E*, *E*, *E*-triene lactones (86), which supports the CCW biosynthetic model proposed by Cane et al. The *monBI* and *monBII* genes, which are homologous to each other, encode the enzymes for epoxide ring opening and the following polyether ring formation.(90) Deletion of these genes resulted in no production of monensin. After a series of gene deletion and mutation experiments, and solving the crystal structure of MonBI by Oikawa's group, the role of MonBI and MonBII has been further understood.(91, 92) It shows that MonBI and MonBII function with synergistic effect as epoxide hydrolase and cyclase. More specifically, MonBII catalyzes two rounds of cyclization while MonBI acts as an allosteric regulator for MonBII through protein-protein interaction. Only after the conformational changes induced by MonBI, MonBII is able to catalyze multiple cyclizations.(92)

1.3.2.3 Biosynthesis of lasalocid

Lasalocid is a carboxylic ionophore produced by *Streptomyces lasalinesis*. It has been used as antibacterial and anti-coccidiosis agent, commercially available as the feed additives like Bovatec and Avatec. It can bind with divalent and monovalent cations, altering ionic transportation across lipid membranes of cells. The feeding experiments indicate that lasalocid A is produced from the three butyrate, four propionate, and five acetate building units.(63, 93, 94) The five oxygen atoms at C-1, C-3, C-11, C-13, and C-15 of lasalocid A are derived from their respective building units, while three other oxygen atoms at C-19, C-22, and C-23 come from the molecular oxygen.(95)

Around 2008, Leadlay's (96) and Oikawa's (97) group both reported a gene cluster of lasalocid A biosynthesis (Figure 1.5). It contains a seven-domain type I PKS system, named as *lsd11-lsd17* (or *lasAI-lasAVII*) including a loading module and 11 extension modules to construct the prelasalocid **21**, a linear dodecaketide from five malonyl-CoA, three ethylmalonyl-CoA, and four methylmalonyl-CoA precursors. The KSQ motif was also found in its loading module, which agrees with the feeding experiment result. It indicates that the starter unit is from malonate instead

of acetate.(98) The gene *lsd4* shows a high homology to *monAX* gene (54% identity) of monensin biosynthesis, which is a type II thioesterase for removing the aberrantly aattached acyl intermediates from the PKS.(99, 100) Unlike monensin biosynthesis which has a discrete TE domain, lasalocid biosynthesis has a fused TE domain at the Lsd17 *C*-terminal. This finding suggests that the full-length polyketide can be cleaved from the ACP domain, even before the complete formation of polyethers.

Guided by the CCW biosynthesis model, the genes of epoxidase and epoxide hydrolase were also identified from the gene cluster, which are *lsd18* (also named as *lasC*) and *lsd19* (also named as *lasB*), respectively. Lsd18 is a flavin-dependent monooxygenase with two flavin adenine dinucleotide (FAD) binding motifs, GxGxxG and GD (101, 102), sharing 52% identity with MonCI. Biotransformation of Lsd18 in an organic-solvent-resistant strain *Rhodococcus erythropolis* showed that Lsd18 catalyzes epoxidation of both carbon double bonds in a stereoselective manner. Time course analysis also showed that epoxidation starts from the terminal olefin of prelasalocid **21**.(103) According to the structural analysis and sequence alignment of its homologs like *p*-hydroxybenzoate hydroxylase (PHBH) (104), TetX (105) and RebC (106), it indicates that the isoalloxazine ring of FAD in Lsd18 might move from the "out" position to "in" position during their catalysis for the FAD oxidation and reduction.(107-109) In 2014, Oikawa's group examined six differently substituted olefins to explore Lsd18 stereoselective mechanism.(110) Based on the results, they proposed a substrate binding model for Lsd18. According to their hypothesis, the substitution patterns on the substrate olefin would affect the its transformation and enantiofacial selectivity because of the steric hindrance with the active site. However, this hypothesis hasn't been verified yet due to the lack of Lsd18 molecular structure.

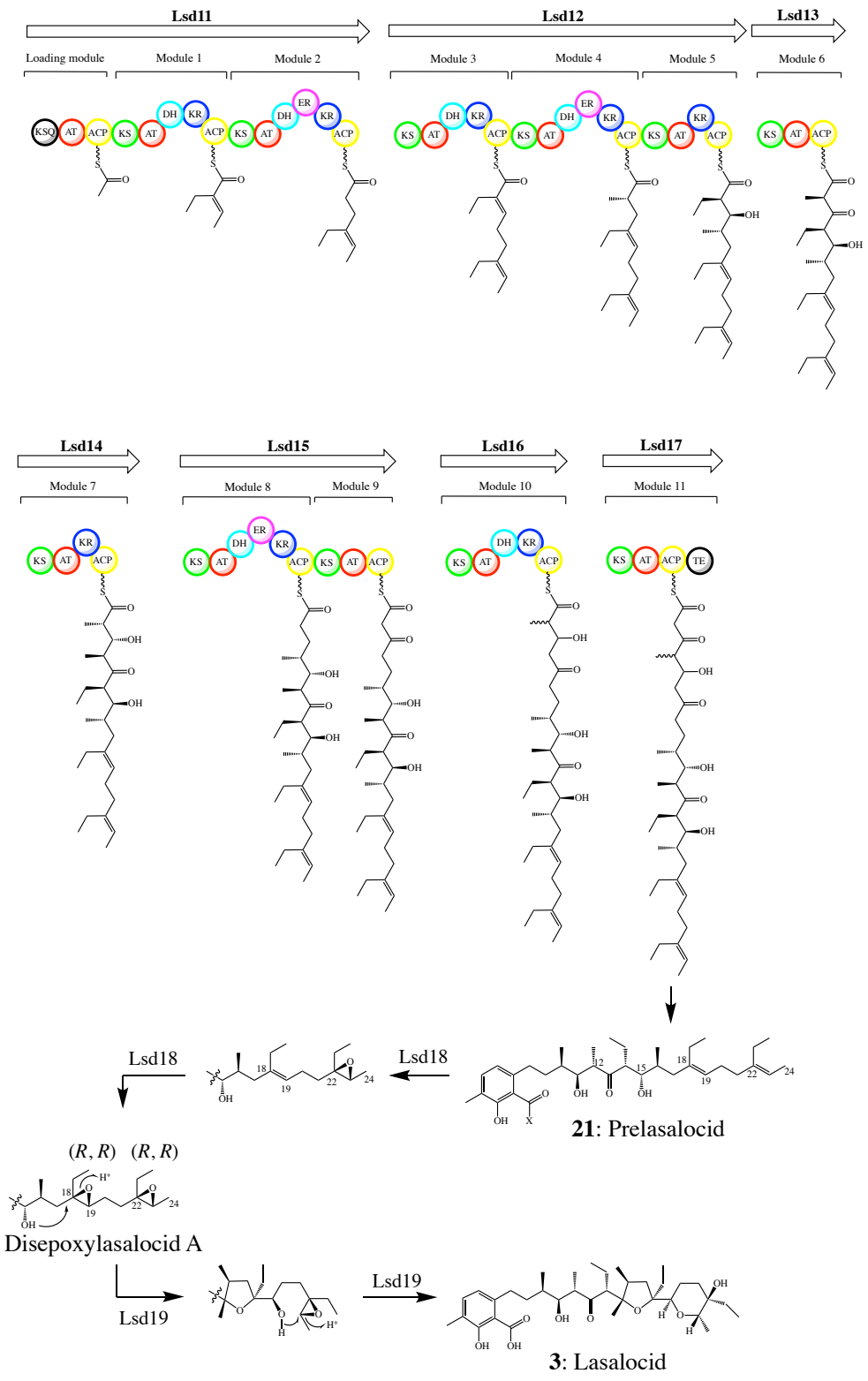


Figure 1.5: The proposed biosynthesis pathway of lasalocid. X= OH, SCoA, or S-protein.

As a homolog of epoxide hydrolases MonBI/MonBII in monensin biosynthesis, Lsd19 catalyzes two rounds of epoxide opening reactions with high regioselectivity, including an energetically favored 5-*exo* cyclization and an energetically unfavored 6-*endo* cyclization (Figure 1.5).(111, 112) The crystal structures of Lsd19 with the substrate and product analogs has been solved. It showed that Lsd19 has two fused domains, Lsd19A and Lsd19B, with two separate active sites. They are able to catalyze a single cyclization reaction independently. Structural analysis of Lsd19B by computational calculations suggests that the preorganization of the active site could be the key to initiate the second round, energetically disfavored 6-*endo* cyclization. Further studies on the Lsd19 regioselective mechanism would shed light on its bioengineering to produce various ladder polyethers.(113, 114)

1.4 FLAVIN-DEPENDENT MONOOXYGENASES

Flavin-dependent monooxygenases are the enzymes that use flavin as a cofactor and catalyze the incorporation of one oxygen atom into the substrate. Oxygen activation in the flavin-dependent monooxygenases typically involves the formation of a transiently C4a-(hydro)peroxyflavin intermediate.(115) These flavin-dependent monooxygenases are involved in various biological processes, including the epoxidation during the formation of ether rings in the biosynthesis of polyether ionophores. According to the standards of the nomenclature committee of the International Union of Biochemistry and Molecular Biology, the flavin-dependent monooxygenases are a part of the oxidoreductase subclasses 1.13 and 1.14. Based on their functional and structural properties, all the flavin-dependent monooxygenases are divided into eight groups A-H.(116)

Group A flavin-dependent monooxygenases are encoded by a single gene and use NAD(P)H as the electron donor. They contain a typical glutathionite reductase (GR-2) type Rossmann fold for cofactor flavin binding and a DG fingerprint for both flavin and NAD(P)H binding. Group A flavin-dependent monooxygenases catalyze the degradation of polyaromatic compounds and biosynthesis of some natural products. Most of the group A flavin

monooxygenases are responsible for the *ortho-* or *para*-hydroxylation od the phenolic compounds. The prototype includes PHBH that transforms 4-hydroxybenzoate into 3,4-dihydroxybenzoate.

Group B flavin-dependent monooxygenases are also encoded by a single gene. They contain two α/β Rossmann like domains for flavin and NAD(P)H binding. Unlike group A enzymes, the group B flavin monooxygenases remain bound to their pyridine nucleotide coenzymes during catalysis. This group consists of four subgroups: Baeyer-Villiger monooxygenases, *N*-hydroxylating monooxygenases, flavoprotein monooxygenases, and a more recently discovered YUCCAs.

Group C flavin-dependent monooxygenases are two-component enzyme systems that rely on a NAD(P)H-dependent flavin reductase for reducing their flavin cofactors. The prototype is luciferase in bacteria, which is the first flavoprotein with the direct nuclear magnetic resonance (NMR) proof for C4a-(hydro)peroxyflavin intermediate. The reactions that group C enzymes catalyze include light emission, Baeyer-Villiger oxidation, epoxidation, desulfurization, sulfoxidation, and hydroxylation.

Group D flavin-dependent monooxygenases are featured by an acyl-CoA dehydrogenase fold. They are also a two-component enzyme system that receive the reduced flavin from a NAD(P)H-dependent flavin reductase. They are involved in the reactions like aromatic hydroxylation and *N*-hydroxylation. The prototype enzyme is 4-hydroxyphenylacetate 3-hydroxylase.

Group E flavin-dependent monooxygenases are the epoxidases that are a two-component enzyme system and contain a PHBH (GR-2) Rossmann fold. They use a NAD(P)H-dependent flavin reductase to generate the reduced FAD. The prototype of group E enzyme is styrene monooxygenases that transform styrene derivatives into the (*S*)-styrene oxides.

Group F flavin-dependent monooxygenases are a two-component enzyme system, also generate the reduced flavin from a NAD(P)H-dependent flavin reductase. Structurally similar to group A enzymes, group F enzymes also contain a glutathione reductase (GR-2) FAD-binding fold. They catalyze halogenation reaction. The prototype is tryptophan 7-halogenase. Group G

enzymes contain a monoamine oxidase Rossmann fold and use an amino acid substrate as the electron donor. They are responsible for oxidative decarboxylation. The typical group G enzyme is tryptophan 2-monooxygenase. Group H flavin-dependent monooxygenases contain a flavin mononucleotide (FMN) as their cofactor. They are feature by a Tim-barrel (glycolate oxidase) fold. They catalyze oxidative decarboxylation and oxidative denitration reactions. The prototype of group H enzymes are lactate 2-monooxygenase and nitronate monooxygenase.

1.5 EPOXIDATION

1.5.1 Epoxidation in organic synthesis

An epoxide is a cyclic ether with a highly reactive three-atom ring. The basic structure consists of an oxygen atom bound to two carbon atoms. Epoxidation reaction is one of the most frequently used reactions in industry.

There are two main ways to prepare epoxides in organic synthesis. The most common laboratory operation includes the direct oxidation of alkenes with peroxycarboxylic acids, which can be represented by the Prilezhaev reaction (Illustration 1.1 A).(117-119) In this approach, *meta*-chloroperoxybenzoic acid (*m*-CPBA) is employed as the peroxyacid. Magnesium mono-perpthalate and peracetic acid can also be used. This reaction proceeds through a "butterfly mechanism" initially proposed by Bartlett in 1950, where the peroxy acid tends to adopt an intramolecularly hydrogen-bonded conformation in solution at the transition state (Illustration 1.1 B).(120) The peroxide is an electrophile while the alkene is a nucleophile. This high degree of peroxide polarization causes an electrophilic oxygen atom added to the alkenes. This type of epoxidation is highly stereospecific in terms of the double bond stereochemistry, which means that a *trans*-olefin only leads to a *trans*-2,3-substituted epoxide, while a *cis*-olefin would only produce the *cis*-epoxide. Hyperoxides can also be applied in the enantioselective epoxidations, such as Jacobsen epoxidation (121), Sharpless epoxidation (122), and Shi epoxidation (123, 124). All of these reactions are very useful for the enantioselective synthesis of enantiopure epoxides.

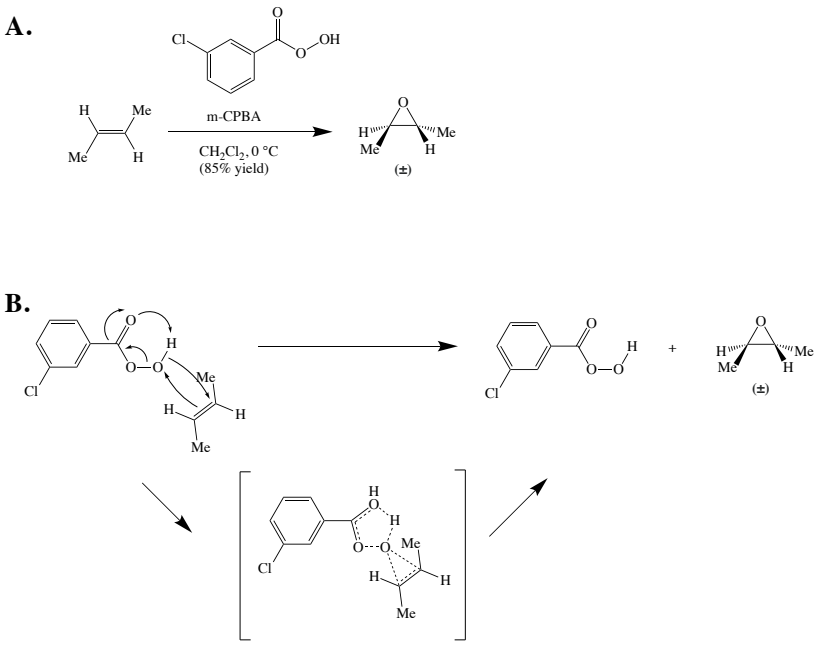


Illustration 1.1: Epoxidation reactions. (A) Prilezhaev reaction. (B) The proposed "butterfly mechanism".

The other way of synthesizing epoxides is through dehydrohalogenation, which involves the cyclization of halohydrins with base. This is a variant of the Williamson ether synthesis (125). Like the bimolecular S_N2 substitution, this reaction takes place by backside substitution of the nucleophile, the oxygen anion at the halide-bearing carbon. And this substitution requires an 'anti' relationship of the nucleophilic oxygen and the leaving halide at the transition state, which is generally achieved by a simple internal rotation in the noncyclic halohydrins (92). The representative example of this method can be found in the Johnson-Corey-Chaykovsky reaction.

1.5.2 Enzymatic epoxidation

The enzymatic epoxidation can be conducted directly by the oxygenases that incorporate one oxygen atom of either water or molecular oxygen into the substrates. They play an important role in the metabolic reactions, including the degradation of drugs, interconversion of amino acids, lipids, and hormones, and also the biosynthesis of antibiotics. So far, only few of these enzymes have been well-studied due to their flexibility and difficulty of protein purification. Among all the

identified enzymes, there are two main types: cation-dependent monooxygenases and flavin-dependent monooxygenases.

1.5.2.1 Cation-dependent epoxidases

The cation-dependent monooxygenases typically contain heme as a cofactor. They have been found in all types of living cells, including animals, insects, plants, fungi and bacteria. The cytochrome P450 monooxygenases are the important members of this type.(126) They catalyze a variety of the regio-, chemo-, and stereospecific oxidation by utilizing electrons derived from NAD(P)H.(127) For catalytic activities, cytochrome P450s need to rely on a redox partner protein to transfer the electrons from NAD(P)H to the heme center of the P450. Based on the nature of the electron transfer protein, the P450 enzymes can be further divided into several subgroups.(128) One subgroup needs both an iron-sulfur ferredoxin and a FAD-containing reductase.(129) Another subgroup only needs a FAD- or FMN-containing reductase.(130, 131)

1.5.2.2 Flavin-dependent epoxidases

The flavin-dependent monooxygenases also participate in various chemo-, regio-, and stereoselective oxygenation reactions, where one atom of the molecular oxygen is incorporated into the substrate and the other is reduced into water.(116) In nature, the cofactor flavin exists in three forms: riboflavin, flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD). They have different-length functional groups on the ribityl side chain.(116) In general, the mechanism of oxygenation by flavin-dependent monooxygenases is shown in the illustration 1.2. At first, the reduced nicotinamide NAD(P)H transfers its hydride to the isoalloxazine group of the flavin cofactor. The reduced flavin then can react with molecular oxygen, yielding a catalytically active peroxide, C4a-(hydro)peroxyflavin, which is transiently stabilized by the enzymes.(115) The oxygen-oxygen bond in peroxide is split once the peroxide reacts with the nucleophilic or electrophilic substrates. One oxygen atom is incorporated to the substrate, and the other one is reduced to water.(104, 132) Styrene monooxygenases (SMO) are one of the most-studied flavin-dependent monooxygenases, which transfer styrene into (*S*)-styrene oxide in the pathway of

styrene degradation.(133, 134) It contains two individual components: a FAD-bound styrene epoxidase (StyA) and a NADH-utilizing flavin reductase (StyB).(135, 136) During the epoxidation of styrene, molecular oxygen first reacts with the reduced FAD in StyA to produce a transiently stable flavin-C4a-peroxide adduct. Then the substrate styrene reacts with peroxide intermediate and turns into an epoxide. The crystal structure of StyA from *P. putida* S12 has been solved.(136-138) It shows that StyA has two distinct domains: one is the FAD binding site and the other is the styrene binding site.(137, 138) The kinetic studies of StyB revealed the sequential binding of NADH and flavin, and it also proved that the presence of StyA doesn't affect the NADH oxidation activity.(136)

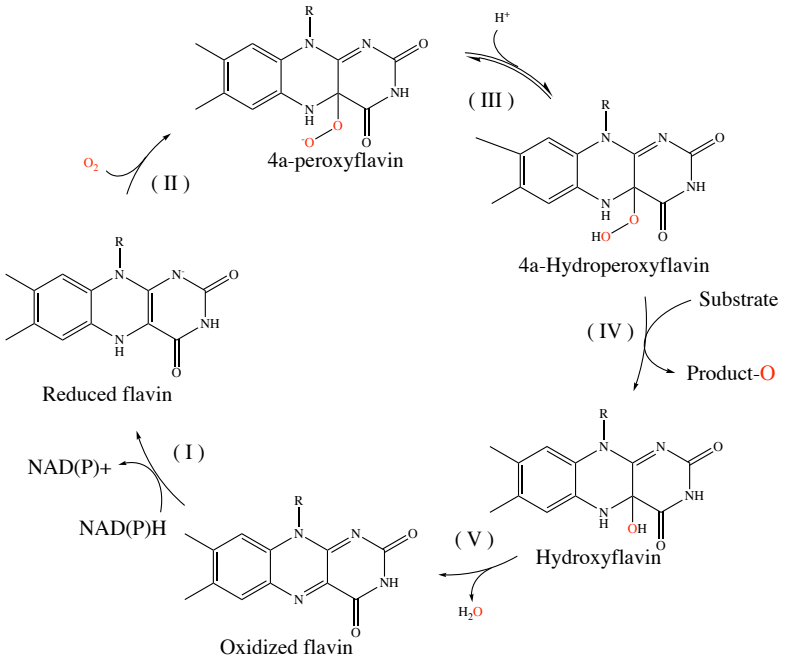


Illustration 1.2: Simplified catalytic cycle of flavin-dependent monooxygenase

1.6 EPOXIDATION INVOLVED IN POLYETHER BIOSYNTHESIS

Epoxidation is a key step to form ether rings during the biosynthesis of polyether ionophores. So far, there have been several epoxidases identified in the gene clusters responsible for polyether biosynthesis. The two most well-studied epoxidases involved in the polyether biosynthesis are MonCI and Lsd18 that are responsible for the epoxidation in the biosynthesis of monensin and lasalocid, respectively.

Since the gene cluster for the polyether monensin biosynthesis was discovered by Leadlay and his co-workers in 2001 (84), epoxidase MonCI has raised up researchers' attention. MonCI was identified as epoxidase due to the significant gene similarity to other authentic non-haem epoxidases. It was considered as a candidate to catalyze the epoxidation on all three alkenes of premonensin. For the *Streptomyces* mutant with the gene deletion of *monCI*, it only produced the linear polyene premonensin precursor instead of producing the final product monensin. This experiment indicated that MonCI plays an important role during the biosynthesis of monensin. Oliynyk et al. also demonstrated that the heterologous overexpression of protein MonCI in another strain *Streptomyces coelicolor* increased its ability to epoxide linalool, an unsaturated terpene, by a 10-20 fold, thereby further establishing MonCI as the enzyme responsible for the epoxidation step in monensin biosynthesis.(85) It has shown that MonCI is able to install (*R*, *R*)-epoxides on the two internal olefin moiety and a (*S*,*S*)-epoxide on the terminal olefin moiety.

After the discovery of MonCI, Lsd18 was also identified as an epoxidase from the gene cluster responsible for the biosynthesis of lasalocid in *Streptomyces lasaliensis*. It shares high sequence identity to MonCI. Gene deletion experiment also showed its role as epoxidase in biosynthesis of lasalocid. Recently, Oikawa's group studied Lsd18 by both *in-vivo* and *in-vitro* studies.(103) They used simple olefin and truncated diene of the Lsd18 native substrate as its substrate analogs. It showed that Lsd18 is able to conduct enantioselective epoxidation on the internal and terminal two olefin moieties one by one in a stereoselective and regioselective manner. The epoxidation starts from the terminal olefin. Lsd18 transforms both internal and terminal olefin moieties into (*R*,*R*)-epoxides. Besides, Oikawa's group also explore the mechanism of the stereoselectivity of Lsd18 by analyzing the products' stereochemistry after reaction.(110) They used different monoolefin substrate analogs with various substitution patterns such as hydrogen and methyl groups. They observed that for the substrate analogs with different substitution patterns, the enantiomeric ratio of the (*R*,*R*)-epoxide products is also different. Based on this result, Oikawa' group proposed a substrate binding model of Lsd18, which has a relatively large primary

site and a space-limited secondary site for the substrate binding. However, this binding model has be verified yet due to the lack of structural information.

Besides MonCI and Lsd18, there are also several other epoxidases identified from the gene clusters of polyether biosynthesis. For example, SalC is the epoxidase involved in biosynthesis of polyether salinomycin. It converts two carbon double bonds of the salinomycin precursor into two (*R*, *R*)-epoxides.(139) TmnC is found to be the epoxidase involved in biosynthesis of polyether tetronomycin. It transfers one terminal carbon double bond of the tetraonomycin precursor into a (*S*, *S*)-epoxide.(140) Nano and NigCI are also epoxidases responsible for the epoxidation reactions in the biosynthesis of nanchangmycin and nigericin, respectively.(141, 142) However, due to the difficulty of obtaining stable enzyme samples in vitro and lack of structural information, the overall study about the epoxidases is very limited. How these enzymes catalyze multiple epoxidation in a regio- and stereoselective manner still remains unknown.

Chapter 2: Materials and Methods

2.1 HETEROLOGOUS PROTEIN EXPRESSION AND PURIFICATION

2.1.1 Cloning

The genes *lsd18* from *Streptomyces lasalinesis* and *monCI* from *Streptomyces cinnamonensis* were separately cloned into the vector pCold I (Takara Bio) between the restriction sites NdeI and EcoRI. For the best production of protein MonCI, the *monCI*-pCold I plasmid was transformed into the BL21-AI *Escherichia coli* bacterial strain (ThermoFisher). For Lsd18, a chaperone-encoding vector pG-KJE8 (Takara Bio) was transformed together with *lsd18*-pCold I plasmid into the *Escherichia coli* BL21(DE3) bacterial strain (Sigma Aldrich). This pG-KJE8 vector encodes chaperones of dnaK, dnaJ, grpE, groES, and groEL, which can help with protein folding and increase the recovery of expressed Lsd18 in the soluble fraction. Other genes including *lsd19*, *gdh*, *fre,* and *monACPX* were all inserted the vector pET-28a(+). The genes *lsd18*, *monCI*, *lsd19*, *fre* and *monACPX* were synthesized by GenScript. The *gdh*-pET-28b plasmid was provided by courtesy of Professor Adrian Keating-Clay from University of Texas at Austin. For protein expression, the *lsd19*-pCold plasmid was transformed into *E.coli* BL21(DE3) bacterial strain together with the chaperon pG-KJE8 plasmid. The *fre*-pET28a(+) and *gdh*-pET28b plasmids were also transformed into *E.coli* BL21(DE3) bacterial strain, respectively. For the protein expression of MonACPX, the *monACPX*-pET28a(+) plasmid was transformed into the *E.coli* BAP1 bacterial strain. All the detailed gene information is listed in the table 2.1.

Table 2.1: List of all the plasmids used. Amp: ampicillin; Kan: kanamycin; Cm: chloramphenicol.

Gene	Plasmid	Expressed protein	Resistance	Origin
lsd18	pCold I	Lsd18	Amp	*Streptomyces lasalinesis*
monCI	pCold I	MonCI	Amp	*Streptomyces cinnamonensis*

lsd19	pCold I	Lsd19	Amp	*Streptomyces lasalinesis*
gdh	pET-28a(+)	Glucose-1-dehydrogenase (GDH)	Kan	*Bacillus subtilis (strain 168)*
fre	pET-28a(+)	NAD(P)H-flavin reductase (Fre)	Kan	*Escherichia coli (strain K12)*
monACPX	pET-28a(+)	MonACPX	Kan	*Streptomyces cinnamonensis*
dnaK-dnaJ-grpE-groES-groEL	pG-KJE8	DnaK, DnaJ, GrpE, GroES, GroEL	Cm	*Escherichia coli (strain K12)*

2.1.2 Site-directed mutagenesis of *lsd18* gene

The construct of Lsd18 mutants follows the protocol described before.(143) The oligonucleotides shown in the table 2.2 were used to obtain Lsd18 mutants Y218F, Y218S, H71A, H71A, H71S, I72A, I72G, Q66N, and Q66G. All the primers were synthesized by IDT company. The components for polymerase chain reaction (PCR) system are listed in the table 2.3. Detailed steps of PCR protocol are also shown in table 2.4. All the mutant constructs were verified by gene sequencing at the Genomic Analysis Core Facility of the University of Texas at El Paso.

Table 2.2: Oligonucleotide primers for constructing Lsd18 mutants.

Primers	Sequence
Lsd18 Y218F Forward Primer	5'-GTATGGTCTTTGCTACCCGTCTGTTTCGTG-3'
Lsd18 Y218F Reverse Primer	5'-CGGGTAGCAAAGACCATACCGCTATCAAC-3'
Lsd18 Y218S Forward Primer	5'-GTATGGTCTCTGCTACCCGTCTGTTTCGTG-3'
Lsd18 Y218S Reverse Primer	5'-CGGGTAGCAGAGACCATACCGCTATCAAC-3'
Lsd18 H71A Forward Primer	5'-GTCATGCGGCCATTCTGTGGTCTGGCG-3'

Lsd18 H71A Reverse Primer	5'-CACAGAATGGCCGCATGACGAGCTTGC-3'
Lsd18 H71S Forward Primer	5'-GTCATGCGTCCATTCTGTGGTCTGGCG-3'
Lsd18 H71S Reverse Primer	5'-CACAGAATGGACGCATGACGAGCTTGCG-3'
Lsd18 I72A Forward Primer	5'-CATGCGCACGCTCTGTGGTCTGGCGGTG-3'
Lsd18 I72A Reverse Primer	5'-GACCACAGAGCGTGCGCATGACGAGCTTGC-3'
Lsd18 I72G Forward Primer	5'-CATGCGCACGGTCTGTGGTCTGGCGGTG-3'
Lsd18 I72G Reverse Primer	5'-GACCACAGACCGTGCGCATGACGAGCTTGC-3'
Lsd18 Q66N Forward Primer	5'-GTGTCCCGAATGCTCGTCATGCGCACATTC-3'
Lsd18 Q66N Reverse Primer	5'-TGACGAGCATTCGGGACACCTTTGCGCAG-3'
Lsd18 Q66G Forward Primer	5'-GTGTCCCGGGAGCTCGTCATGCGCACATTC-3'
Lsd18 Q66G Reverse Primer	5'-ATGACGAGCCCTCGGGACACCTTTGCGCAG-3'

Table 2.3: PCR components for Lsd18 mutant construction. Phusion HF buffer purchased from New England Biolabs creates the optimal reaction conditions for high fidelity amplification. Phusion GC buffer purchased from New England Biolabs promotes excellent amplification of GC-rich DNA.

Components	50 µl reaction	Final concentration
Nuclease-free water	to 50 µl	/
5x Phusion HF or GC buffer	10 µl	1x
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
Template DNA (*lsd18*-pCold plasmid)	variable	120 ng
DMSO	1.5 µl	3%
Phusion DNA Polymerase	0.5 µl	1.0 units/50 µl PCR

Table 2.4: Thermocycling conditions for Lsd18 mutant PCR.

Steps	Temperature	Time
Initial Denaturation	98 ℃	30 seconds
30 cycles	98 ℃	10 seconds
	72 ℃	175 seconds
Final Extension	72 ℃	5 minutes
Hold	4 ℃	hold

2.1.3 Heterologous protein expression

For Lsd18 expression, the culture of BL21(DE3) strain carrying *lsd18*-pCold I and pG-KJE8 plasmids was grown in Luria Broth medium (Molecular Dimensions) to OD_{600nm} of 0.6. Lsd18 was induced together with the chaperones by 100 μM isopropyl-β-D-galactoside (IPTG), 0.5 mg ml^{-1} L-arabinose and 5 ng ml^{-1} tetracycline. The culture was incubated for another 20 hours at 15 ℃. After harvested by centrifugation, the cell pellets were used immediately for purification or kept at -80 ℃ for future use. For MonCI expression, the culture of BL21-AI strain carrying *monCI*-pCold I plasmid was also grown in the LB medium to OD_{600nm} of 0.6. Then it was induced by 100 μM IPTG and 2 mg ml^{-1} L-arabinose. The induction condition and cell harvesting method are the same as MonCI. Expression of Lsd19 is the same as Lsd18 while the GDH, Fre and MonACPX were expressed in BL21(DE3) *E. coli* strain with 100 μM IPTG for 20 hours at 18 ℃.

2.1.4 Recombinant protein purification

For epoxidase MonCI purification, the cell pellet from 4-liter cell culture was resuspended in the cell lysis buffer containing 50 mM sodium phosphate pH 7.4, 300 mM sodium chloride, 40 mM imidazole, 10% (v/v) glycerol and lysed by sonication. After centrifuged at 18,000 g for 40 minutes, the MonCI protein sample was first purified by immobilized metal affinity

chromatography (IMAC). The clear supernatant was loaded onto a 5 ml HisTrap column (GE Health Life Sciences) and washed with a wash buffer containing 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 70 mM imidazole, 10% (v/v) glycerol. N-terminal His-tagged MonCI proteins were eluted out with a wash buffer supplemented with 200 mM imidazole. Fractions containing MonCI were diluted by a buffer containing 20 mM Tris pH8.5, 10% (v/v) glycerol and further purified by anion exchange chromatography (IEX) using a 5 ml HiTrapQ column (GE Health Life Sciences). The MonCI was eluted out with an elution buffer 20 mM Tris pH8.5, 400 mM sodium chloride, and 10% Glycerol. After IEX, the protein samples were further purified by gel filtration using a Superdex75 10/300 GL column or Superdex200 10/300 GL column (GE Health Life Sciences) with a final protein buffer 20 mM Tris pH7.5, 150 mM sodium chloride, 10% (v/v) glycerol. The purity of purified MonCI was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE). MonCI protein sample was finally concentrated into 6.2 mg ml^{-1} for storage.

For epoxidase Lsd18 purification, the cell pellet from 2-liter cell culture was resuspended in the cell lysis buffer 50 mM sodium phosphate pH 7.4, 300 mM sodium chloride, 40 mM imidazole, 10% (v/v) glycerol and lysed by sonication. After centrifuged at 18,000 g for 40 minutes, the clear supernatant was loaded onto a 5 ml HisTrap column (GE Health Life Sciences) and washed with a wash buffer containing 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 80 mM imidazole, 10% (v/v) glycerol. N-terminal His-tagged Lsd18 proteins were eluted out with a wash buffer supplemented with 200 mM imidazole. Fractions containing Lsd18 were diluted by a buffer containing 20 mM Tris pH8.5, 10% (v/v) glycerol and further purified by IEX method using a 5 ml HiTrapQ column (GE Health Life Sciences). After IEX, the protein samples were further purified by gel filtration using a Superdex75 10/300 GL column or Superdex200 10/300 GL column (GE Health Life Sciences) with a final protein buffer 20 mM Tris pH8.5 and 100 mM sodium chloride. The purity of purified Lsd18 was checked by SDS PAGE. The protein sample was concentrated to 5.5 mg ml^{-1} for storage. For Lsd18 mutant purification, the protocol

was the same as the native Lsd18 purification except that more liters of cell culture were required due to the lower yield of Lsd18 mutant expressed in the *E.coli* bacteria.

For epoxide hydrolase Lsd19 purification, the first step, IMAC purification is the same as Lsd18. After sample loading, the column was first washed at 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 10 mM imidazole, 10% (v/v) glycerol, then Lsd19 was eluted by gradient elution. After IMAC, Lsd19 was further purified by IEX with elution buffer 20 mM Tris, pH8.5, 200 mM sodium chloride, 10% Glycerol by gradient elution. The eluted Lsd19 protein sample was concentrated to 15 mg ml^{-1} for storage. The final Lsd19 protein sample was also analyzed by SDS PAGE.

For glutamate dehydrogenase GDH purification, the cell pellet from 1-liter cell culture was first lysed by sonication with the lysis buffer 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 40 mM imidazole, 10% glycerol, and 500 μM protease inhibitor, phenylmethylsulfonyl fluoride (PMSF). After sonication, the cell lysis was centrifuged at the speed of 18,000 g for 1 hour. The supernatant was loaded onto a 5 ml Histrap column (GE Health Life Sciences). The column was washed by the buffer of 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 70 mM imidazole, and 10% (v/v) glycerol until the UV absorption at 280 nm stabilized. Then the GDH protein was eluted out by the elution buffer consisting of 50 mM sodium phosphate pH7.4, 300 mM sodium chloride, 200 mM imidazole, and 10% (v/v) glycerol. The IMAC elute was concentrated and injected onto a Superdex 75 10/300 GL column (GE Health Life Sciences) for size exclusion chromatography (SEC). After that, GDH protein was concentrated to 6.5 mg ml^{-1}. The final protein buffer was 20 mM Tris pH7.5, 150 mM sodium chloride, and 10% glycerol. The final MonCI protein sample was also analyzed by SDS PAGE.

For FAD reductase Fre purification, the protein samples were first purified by IMAC using a 5 ml HisTrap column (GE Health Life Sciences). The protein samples were washed out by gradient elution. The eluted fractions were diluted by a buffer containing 20 mM Tris pH8.5 and 10% glycerol. Then the sample was loaded onto a 5 ml HiTrapQ column (GE Health Life Sciences) and washed out by gradient elution. After IEX, the protein samples of Fre were purified by gel

filtration. The final protein buffer for Fre was 20 mM Tris pH7.5, 10% Glycerol. The final Fre protein samples were also analyzed by SDS PAGE.

For the MonACPX protein purification, the cell pellet of 2-liter cell culture was dissolved in the buffer containing 50 mM Tris pH8.0, 0.2 M sodium chloride, and 10% glycerol. After the cell pellet got lysed by sonication and span down at 18,000 g for 1 hour, the supernatant was loaded onto the 5 ml Histrap column (GE Health Life Sciences). The IMAC wash buffer was 50 mM Tris pH8.0, 0.2 M sodium chloride, and 25 mM imidazole. The MonACPX protein was eluted out by the buffer containing of 50 mM Tris pH8.0, 0.2 M sodium chloride, and 250 mM imidazole. The IMAC elute of MonACPX was about 0.21 mg ml^{-1} in 28 ml volume. It was incubated with 94 µl 1KU ml^{-1} thrombin and 3 ml 10x cleavage buffer containing 200 mM Tris pH8.5, 1.5 M sodium chloride, and 25 mM calcium chloride. The mixture was incubated at room temperature for 20 hours for his tag cleavage. After that, the mixture was injected into the gel filtration column for final purification and the separation of the cleaved his tag. The final protein buffer for MonACPX was 20 mM Tris pH7.5, 150 mM sodium chloride, and 10% Glycerol. The final MonACPX protein samples were also analyzed by SDS PAGE.

The concentration of all the protein samples including MonCI, Lsd18, Lsd19, Fre, and GDH were measure by the UV absorption. Due to the lack of aromatic amino acids in its protein sequence, the concentration of MonACPX was measured by the colorimetric dye-based method, the Bradford (Coomassie) protein assay. The first step of the Bradford (Coomassie) protein assay is the preparation of the diluted albumin (BSA) standards. The original BSA concentration is 2 mg ml^{-1}. Then it was diluted to 0.05 mg ml^{-1}, 0.033 mg ml^{-1}, 0.025 mg ml^{-1}, 0.02 mg ml^{-1}, 0.017 mg ml^{-1}, 0.0125 mg ml^{-1}, 0.010 mg ml^{-1}, and 0.0083 mg ml^{-1}. Then the Coomassie reagent was mixed with the diluted standards and the MonACPX protein sample. After incubation at the room temperature for 5 minutes, the absorption of the mixture was measured at UV 595 nm. Since there is a correlation between the protein concentration and their absorption at 595 nm after mixing with the Coomassie dye, a linear line and equation was generated by using the known protein concentrations of a series of diluted BSA standards as the x-axis, and their absorption at 595 nm

as the y-axis. The R^2 value was used to evaluate how accurate the assay was. Based on this correlationship and its absorption at 595 nm, the protein concentration of the MonACPX purified sample could also calculated. Every method to measure protein concentration has their advantages and disadvantages. For the direct measurement at UV 280nm, it is very simple and does not require the use of any assay reagents, but it is prone to have high errors with the protein mixtures or complex samples such as cell lysates. For the Bradford (Coomassie) protein assay, it is fast and easy. It can be performed at the room temperature and compatible with most solvents, salts, buffers, reducing substances, and metal-chelating agents. However, this method is not compatible with any detergents. Also, there is high protein-protein variation compared with the copper-based assays.

2.1.5 Protein molecular weight determination

After protein purification, the protein molecular weight of Lsd18 and MonCI was determined by column calibration. Size-exclusion chromatography was performed using a Superdex 10/300 GL column (GE Healthcare) and a flow rate of 0.5 ml min^{-1}. Protein standards carbonic anhydrase (29 kDa), and conalbumin (75 kDa) were purchased from GE Healthcare. The human intestinal trefoil factor monomer (6.86 kDa) and dimer (13.7 kDa) were provided by Dr. Chen Xi form the Northwest University in China. The void volume was determined by using the Blue Dextran 2,000 (GE Healthcare). The use of size-exclusion chromatography to determine protein molecular weight has been well documented. This method relies on the well-established ability of size-exclusion media to separate molecules based on their different size. The determination of the protein molecular weight was carried out by comparing the elution volume parameter Kav of the Lsd18 and MonCI to the Kav of the several protein standards carbonic anhydrase, conalbumin, and the human intestinal trefoil factor monomer and dimer. The elution volume parameter Kav of the proteins is found to have a sigmoid relationship with the logarithm values of their molecular weights. To achieve a successful molecular weight determination, all the protein standards were dissolved in the same buffer that contains 0.1 M sodium phosphate pH7.2

and 1 mM DTT. For protein stabilization, MonCI protein sample was dissolved in the buffer containing 20 mM Tris pH8.5 and 100 mM sodium chloride, while Lsd18 protein sample was dissolved in the buffer containing 20 mM Tris pH7.44, 150 mM sodium chloride, and 10% glycerol. Since the total column volume is 24 mL, 500 ul of protein samples were injected, which is less than 2% of the total volume.

2.2 ENZYME ASSAYS

2.2.1 UV absorption spectrum analysis

UV-visible absorption spectra for detecting flavin were recorded by Eppendorf biospectrometer at room temperature. Purified protein samples in buffer containing 20 mM Tris pH8.5, 200 mM sodium chloride, and 10% (v/v) glycerol were analyzed at OD_{450nm}. After that, sodium dithionite was added into the solution to reduce flavin. The spectra were recorded after 2 minutes and 90 minutes. After heated at 100 ℃ for 10 minutes and centrifugation, the flavin was separated from the protein samples. The extracted flavin supernatant was also analyzed at OD_{450nm}. The FAD purchased from Sigma Aldrich used as standards. It was dissolved in the same buffer at gradient concentrations of 5 μM, 10 μM, 50 μM, and 100 μM. To check if flavin reduction is related to the substrate-binding and catalysis or nor, MonCI and Lsd18 protein samples were added with 10x molar excess linalool and measured immediately by UV-visible absorbance. After incubated at 4 ℃ overnight, the absorbance of the protein mixture was measured again for comparison.

UV-visible spectra were also recorded by Eppendorf biospectrometer at room temperature to detect NAD(P)H. All the samples were put in a 10 mm cuvette with 100 μl solution volume and measured at 340 nm where the reduced NADH or NADPH has maximum absorption. For Fre activity assay, 3 μM Fre, 16 μM FAD, and 400 μM NADH were mixed first in the buffer solution of 20 mM Tris pH7.5 and 10% glycerol. The absorption of the mixture was measured at 340 nm every 15 seconds for 5 minutes at the room temperature. For GDH activity assay, 0.05 mg ml^{-1} GDH protein sample, 0.5 mM NADP, and 180 mM glucose were mixed together in the buffer

solution containing 50 mM HEPES pH7.5, 300 mM sodium chloride, and 10% glycerol. The absorbance of the mixture was recorded at 340 nm continuously for 10 minutes at the room temperature.

2.2.2 *In-vitro* enzyme assays

For the enzyme assays with linalool or the substrate analog **24**, each of them contained a final concentration of 80 μM of the substrate, 2 mM of NADH (or NADPH), 80 μM of FAD, and 15 μM enzyme in the buffer containing 50 mM Tris pH8.0, 300 mM sodium chloride, 5% glycerol. After incubated at 30 ℃ for 48 hours, the reaction was quenched by adding equal volume of ethyl acetate. Both the remaining substrate and the product will be extracted from the aqueous reaction mixture by vortex and centrifugation. The concentrated extracted samples were then analyzed by gas chromatography-mass spectrometry (GC-MS).

For the enzyme assay with the substrate analog **26**, it included two reactions. The reaction 1 contained a final concentration of 80 μM substrate, 1 mM NADH, 1 mM NADPH, 80 μM FAD, 15 μM Lsd18, 15 μM Fre in the buffer of 50 mM Tris pH8.0, 300 mM sodium chloride, 5% glycerol. After incubated at room temperature for 24 hours, the reaction was quenched by adding equal volume of ethyl acetate. After vapor concentration, the reaction 1 product was analyzed by liquid chromatography-mass spectrometry (LC-MS) and also used as the substrate for the second reaction. The reaction 2 contained a final concentration of 80 μM substrate, 1 mM NADH, 1 mM NADPH, 80 μM FAD, 15 μM Lsd18, 15 μM Fre and 15 μM Lsd19 in the buffer of 100 mM Bis Tris pH6.5, and 10% glycerol. After incubated at room temperature for 12 hours, the reaction mixture was extracted by the equal volume methanol. After vapor concentration, the final product was analyzed by LC-MS.

2.2.3 GC-MS analysis

The extracted samples from the enzyme assays were analyzed by GC-MS (5973 Network, Agilent) using a HP-5ms Ultra Inert GC column (30 m x 0.25 mm, 0.25 μm film thickness, J&W Scientific Agilent). An oven temperature was programmed from 40 ℃ (3 minute) to 250 ℃ at 15

℃ min^{-1} with helium (0.9 ml min^{-1}) as the carrier gas. The selected ion monitoring mode was applied, which was set to measure the specified mass with hundred-time higher sensitivity than full scan mode. The specific fragment m/z for detecting linalool and linalool oxide are 59, 71, 93, 111, and 121, selected according to the GC-MS chemical database. For substrate analog **24** and its epoxidation products, the specific m/z 55, 71, 82, 95, 112 and 135 were picked based on its standard GC-MS results. The percent yield was estimated by calculating the percentage of product peak area over the total area of both the substrate and product peaks.

2.2.4 LC-MS analysis

The extracted samples from the enzyme assay with the substrate **26** were analyzed by LC-MS (Thermo Ultimate 3000 HPLC and Thermo TSQ Endura mass spectrometry). The product mixture was separated first by HPLC and then analyzed with electrospray ionization mass spectrometry. The HPLC column we used was Agilent Porshell 300-SB-C18 (5 μM), 2.1 mM ID x 75 mM. The flow rate was 0.4 ml min^{-1}. The column temperature was 40 ℃. The pump A buffer was 100% H_2O and 0.1% formic acid. The pump B buffer was 100% methanol and 0.1% formic acid. The HPLC protocol was 0 min at 20% B; 1-6 min, linear gradient 20%-95% B; 6-10 min, 95% B; 10.2-13 min, 20% B. The detect mass range was 300-600 m/z (full scan). The base peaks 469.50-470.50+ m/z, 485.50-486.50+ m/z, and 501.50-502.50 m/z were selected specifically for detecting the substrate analog **26** and its products.

2.3 X-RAY CRYSTALLOGRAPHY

2.3.1 Protein crystallization

For initial screening, Lsd18 and MonCI samples were set up in the 3-drop 96-well sitting-drop plates automatically by Formulatrix NT8 crystallization robot. The drop contained 0.2 μl protein solution and 0.2 μl crystallization solution with 50 μl well condition. The initial screening was carried out by using the commercial crystallization screens of MSCG 1-4 (Antrance), Morpheus (Molecular Dimensions), JSCG (Molecular Dimensions), and SG1 (Molecular Dimensions). The protein concentration, buffer, salt, crystallization temperature, and other

conditions were also explored during the initial screening. For crystal optimization, the additive screen (Hampton research), silver bullets screen (Hampton research), reductive alkylation kit (Hampton research), random microseeding were all tried.

For MonCI, the crystals were obtained directly from a 1:1 mixture of protein solution (6.2 mg ml^{-1} in 20 mM Tris pH7.5, 150 mM sodium chloride, 10% (v/v) glycerol) and a reservoir solution (0.1 M Bis Tris pH6.5, 0.1 M MOPS pH6.8, 28% (w/v) PEG2000mme, 4% (v/v) glycerol) by the sitting drop diffusion method. The crystallization trays were set up by Formulatrix NT8 crystallization robot. The crystallization drop contained 0.2 µl protein solution and 0.2 µl crystallization solution with 50 µl well solution. The crystal grew up to the full size within one week. Before flash freezing, MonCI crystals were transferred to a cryoprotectant solution comprised of the crystallization buffer supplemented with 20% glycerol.

For Lsd18 crystallization, the protein sample needs to be alkylated to form crystals. After anion exchange chromatography, protein fractions were exchanged into buffer containing 50 mM HEPES pH7.5, 300 mM sodium chloride, and 10% (v/v) glycerol and then concentrated to 5 mg ml^{-1}. Each ml of samples was incubated with 20 µl 1 M dimethylamine borane complex and 40 µl 1 M acetaldehyde for 2 hours at 4 ℃. Same step was repeated again before adding 10 µl 1 M dimethylamine borane complex. After overnight incubation, 125 µl 1 M glycine was added to the mixture and incubated for 2 hours at 4 ℃. After that, the size exclusion chromatography was used for further purification and change protein buffer into 20 mM Tris pH8.5 and 100 mM sodium chloride. The hanging-drop evaporation method was applied for crystallization. The drop contained 2 µl protein solution and 2 µl crystallization solution with 800 µl well condition. The Lsd18 crystals were obtained from a 1:1 mixture of protein solution (5.5 mg ml^{-1} in 20 mM Tris pH8.5 and 100 mM sodium chloride) and a reservoir solution (0.1 M Tris pH8.5, 1 M sodium chloride, 32% (w/v) PEG8000) by the hanging-drop vapor diffusion method at 18 ℃. The crystals were harvested into a cryoprotectant containing 0.1 M Tris pH8.5, 0.85 M sodium chloride, 32% (w/v) PEG4000, and flash freezing in liquid nitrogen until X-ray data collection.

For co-crystallization, Lsd18 protein also got ethylated first. After anion exchange chromatography, protein fractions were exchanged into buffer containing 50 mM HEPES pH7.5, 300 mM sodium chloride, and 10% (v/v) glycerol and then concentrated to 5 mg ml^{-1}. Each ml of samples was incubated with 20 µl 1 M dimethylamine borane complex and 40 µl 1 M acetaldehyde for 2 hours at 4 ℃. Same step was repeated again before adding 10 µl 1 M dimethylamine borane complex. After overnight incubation, 125 µl 1 M glycine was added to the mixture and incubated for 2 hours at 4 ℃. After that, the size exclusion chromatography was used for further purification. To optimize Lsd18 co-crystals, three different protein buffers were explores. The protein buffer 1 was 20 mM Tris pH8.5 and 100 mM sodium chloride. The protein buffer 2 was 20 mM Bicine pH8.5 and 100 mM sodium chloride. The protein buffer 3 was 20 mM HEPES pH8.0 and 100 mM sodium chloride. The ethylated Lsd18 protein sample was first incubated with 100x molar excess substrate **25** in ice for over 1 hour before setting up the crystallization trays. The Lsd18-ligand co-crystals were obtained from a 1:1 mixture of protein solution (5.6 mg ml^{-1} in 20 mM Bicine pH8.5 and 100 mM sodium chloride.) and a reservoir solution (0.1 M imidazole pH8.0, 0.8 M sodium chloride, and 37% PEG4,000) by the sitting-drop vapor diffusion method at 18 ℃. The co-crystals were harvested into a cryoprotectant containing of the 80% reservoir solution and 20% glycerol, and flash freezing in liquid nitrogen until X-ray data collection.

2.3.2 X-ray diffraction, data collection and processing

For Lsd18 crystals, the diffraction data was collected at beamline 17-ID at the Advanced Photon Source, Argonne National Laboratory (Argonne, Illinois, USA). A set of 360° data was collected on a single flash-frozen crystal with a 199.990 mm distance between the crystal and the detector Dectris Pilatus 6M Pixel Array. The wavelength at 17-ID was 1.00 Å at the time of data collection. Each image was exposed for 0.248 seconds with an oscillation angle of 0.25°. Lsd18 co-crystals were diffracted by Mathews at SSRL.

MonCI crystals were diffracted at Stanford Synchrotron Radiation Lightsource (SSRL, SLAC National Accelerator Laboratory). The sodium-bromide soaked crystals were collected at

beamline BL9-2. To obtain phases by multi-wavelength anomalous diffraction (MAD) method, the native crystal was first soaked in the cryoprotectant containing 0.5 M sodium bromide for 10 seconds before diffraction. Three sets of 720° data were collected on the same single crystal at three different wavelengths in the vicinity of the absorption edge, including 0.9200 Å (peak), 0.9795 Å (high energy remote), and 0.9392 Å (inflection point). Each image was exposed for 0.2 seconds, with 15% transmission and 0.2° oscillation angle. For normal native crystals, they were diffracted at SSRL beamline 12-2 for at the wavelength of 0.9795 Å.

2.3.3 Structural determination

The diffraction data of Lsd18 crystals were indexed, integrated and scaled automatically with autoPROC (144). The initial phases were determined by molecular replacement (145) using Molrep (146). The crystal structure of MonCI was used as a search model. Refinement was performed with Phenix.refine (147) followed by manual examination and rebuilding of the refined coordinates in the program Coot (148). The cofactor FAD coordinates were first adopted from the 4-OH benzoate hydroxylase (Protein Data Bank: 1DOD), and then adjusted manually by Coot based on the electron density maps 2|Fo|-|Fc| and |Fo|-|Fc|.

The MonCI MAD diffraction data and native diffraction data were indexed, integrated, and scaled by autoXDS (149). The program SHELXC (150) was used to generate a statistical analysis of the diffraction data, determining the approximate resolution cutoff for the anomalous signal, estimating the heavy-atom structure factor F_A and phase shift α, and preparing the necessary files for the following steps. 11 bromine atoms (10 atoms with occupancy >0.2) were located using the program SHELXD (150), which helped solving the substructure. After that, the program SHELXE (150) was used to improve initial phases and electron density by density modification, and also auto-trace the poly-alanine backbone. Identification of the individual amino acids and matching the structure with its protein sequence was done by the model building program Buccaneer (151). Structural refinement was done by REFMAC (152) and Phenix_Refine (147), followed by manual examination and rebuilding of the refined coordinates in the program Coot (148).

2.4 PHYLOGENETIC ANALYSIS

The phylogenetic analysis was conducted by the online server Phylogeny.fr (153), which includes four steps: alignment (MUSCLE), curation (Gblocks), phylogeny (PhyML + aLRT), and tree rendering (TreeDyn). The gene sequential of five different epoxidases including *monCI* (*Streptomyces cinnamonensis*, GenBank: AAO65803), *nanO* (*Streptomyces nanchangesis*, GenBank: AAP42870.1), *tmnC* (*Streptomyces sp*. NRRL 11266, GenBank: BAE93732.1), *salC* (*Streptomyces albus* subsp. Albus, GenBank: CCD31908.1), and *lsd18* (*Streptomyces lasaliensis*, GenBank: CAQ64694.1) were aligned by MUSCLE (v3.8.31) configured for highest accuracy.(154, 155) After the sequential alignment, the ambiguous regions including gaps and poorly aligned were removed by Gblocks (v0.91b) by following parameters: minimum length of a block after gap cleaning is 10; no gap were allowed in the final alignment; all segments with the contiguous nonconserved positions bigger than 8 were rejected; minimum number of the sequences for a flank position is 85%.(156) After sequential alignment, the phylogenetic tree was reconstructed by using the maximum likelihood method implement in the PhyML program (v3.1/3.0 aLRT).(157, 158) The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.162) and four gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma is 97.543). Reliability of internal branch was assessed using aLRT test (SH-like). Then graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3).(159) The phylogenetic tree was generated in the online server iTOL (160, 161) with the Niewick file of the Phylogenetic analysis.

2.5 HOMOLOG MODELING

The parameters and topologies for the natural substrates were prepared and modified by several programs. The MDL Molfiles of the two-dimension chemical structures were first created by ChemDraw. The online website sever PRODRG (152) then took the Mol files as the inputs and generated a variety of topologies including the LIB files and PDB files. Alternatively, CCP4 Make

ligand-Acedrg can also generate the CIF files and PDB files from the Mol files. The chemical groups and the bonds were modified and adjusted by using Chimera (162). The substrates were initially docked into the substrate-binding pocket of either Lsd18 or MonCI crystal structures manually by Coot, Chimera, and Pymol. The resulting protein-substrate complex structures were energy-minimized with Phenix Geometry Minimization (147).

Sequence alignment of acyl carrier proteins was performed by CLUSTALW (163) and ESPript server (164), which identified the ACP domain from module 5 of MLSA1 (PDB ID: 6H0J) as the best template structure to use in homology modeling of MonACPX. The homology structure of MonCI was generated using the program MODELLER (165). Evaluation of structural models was performed by calculation of z-DOPE score with MODELLER program and the best MonACPX model was selected with comparing the score for all residues. Visualization and image processing were performed by Pymol (166) and Maestro 12.2 (Shrödinger, LLC, Portland, OR) programs.

2.6 MOLECULAR DOCKING

Structures of the MonACPX-MonCI complex were predicted using the HADDOCK2.2 webserver (167, 168). Docking was performed using the MonCl crystal structure and the homology model of *apo*-MonACPX. The HADDOCK2.2 webserver's default input settings were used, and passive residues were assigned automatically around the active residues. Active residues for MonCl were those located near the substrate-binding pocket entrance (side entrance): Asp377, Asp378, Ile381, Thr398, Asp399, Pro400, Arg401, Leu402, Ile403, Gly404, Val405, Asp406, Gln409, Arg412, Phe413, Pro440, Gln441, Ala442, Glu443, Gly445, Ser446, Asn447, Arg448, and Leu450. Two docking runs were performed using different active residues for ACP. First run was performed defining a narrower (Thr37, Gln38, Ser60, Leu61, Leu64, Glu65, Lys68, Thr69, Met80) and second run a wider (Thr37, Gln38, Ala39, Gly40, Asn59, Ser60, Leu61, Thr62, Leu64, Glu65, Thr67, Lys68, Thr69, Met80) selection of active residues surrounding Ser60 of MonACPX, the attachment site for the 4´-phosphopantetheine group. Among the docked structures that were

generated, we eliminated those in which Ser60 was not aligned with the side entrance and then chose the solution that had the best HADDOCK score (-86.2±3.6) as the final model MonACPX-MonCI.

2.7 MOLECULAR DYNAMICS SIMULATION

Each simulated system contained MonCI, FAD, and except for one system, premonensin A. The MonCI and FAD coordinates were obtained from chain A of the experimental crystal structure deposited in the PDB. 481 water molecules that are in the crystal structure and are associated with chain A were also included in simulations. In all cases, hydrogen atoms were added using the VMD psfgen plugin (169). For simulations which included premonensin A, the substrate was manually positioned within the channel with the substrate orientation varying by system. The initial substrate position was an approximate binding pose with the reactive double bond placed ~5 Å from and perpendicular to the O-H bond of the reactive -OOH group of FAD. The systems were solvated and ionized using the VMD solvate and ionize plugins (169). Final systems contained approximately 78,000 atoms, with a salt concentration of 150 mM sodium chloride to mimic physiological conditions.

Interactions for simulated molecules were defined using the CHARMM36 force-field parameters (170) and the TIP3P water model, with the CHARMM general force field (Cgenff) (171-174) used to model the substrate. MD simulations were performed in the NpT ensemble using the NAMD2.13 package (175). Temperature and pressure remained constant at 310 K and 1.01325 bar, respectively. The Langevin constant was γLang=1 ps-1. A simulation timestep of 2 fs was used, with evaluation of van der Waals interactions at every timestep and evaluation of long-range Coulomb interactions performed every 2 timesteps using the particle-mesh Ewald method (176).

All systems were initially minimized for 2,000 steps, after which water and ions were equilibrated for 1 ns while the protein-FAD-substrate complex was restrained using harmonic forces with a spring constant of 1 kcal/(mol $Å^2$). A further 1 ns of equilibration was performed with only the heavy atoms of the protein-FAD-substrate complex restrained, followed by 20 ns of

equilibration in which disordered coil protein residues were not restrained. A final 20 ns equilibration run was performed with harmonic restraints present only on the substrate atoms constituting the double bond nearest FAD. Production runs, with all restraints released, were performed for 200 ns. The analyses were performed during the last 100 ns of the production runs.

Images of simulated systems were prepared with VMD, with secondary structure information obtained using STRIDE (177). RMSDs were calculated using the VMD RMSD trajectory tool. The channel residues include MonCl residues 55, 56, 57, 58, 59, 60, 88, 92, 94, 95, 96, 97, 98, 99, 101, 112, 206, 208, 209, 210, 226, 227, 228, 229, 230, 232, 239, 240, 241, 242, 243, 244, 253, 254, 255, 256, 273, 277, 282, 329, 330, 331, 332, 333, 334, 336, 382, 385, 386, 387, 390, 414, 415, 417, 418, 421, 422, 431, 435.

2.8 CHEMICAL SYNTHESIS

2.8.1 Asymmetric epoxidation of substrate analog 24

The asymmetric epoxidation was done by using potassium peroxymonosulfate (Oxone) and a fructose-derived organocatalyst, Shi's catalyst. This reaction is believed to procced though a dioxirane intermediate generated *in situ* by oxidation of a ketone with Oxone. The added sulfate group, as a good leaving group, facilitates the ring closure of the dioxiranes. After epoxidation, ring-opening occurred under the acidic condition.

The synthetic procedure mainly followed the one as reported.(103) A solution of 38 μM Oxone in 9.1 ml Na_2(EDTA) and 9.1 ml K_2CO_3 was added drop by drop into the chilled solution containing 50 mg substrate analog **24**, 42.4 mg D-fructose-derived Shi's catalyst, 0.05 M $Na_2B_4O_7$, 7.39 mg Bu_4NHSO_4, and 5.5 ml Na_2(EDTA) during 2 hours. The mixture was then diluted with 1 N HCl solution and extracted by Et_2O. The organic layer was washed with brine, dried over anhydrous $MgSO_4$ and concentrated by air. The reaction was monitored by thin layer chromatography using phosphomolybdic acid as a developing agent.

2.8.2 Symmetric epoxidation of substrate analog 24

The symmetric epoxidation was a Prilezhaev reaction with *meta*-Chloroperoxybenzoic acid (*m*-CPBA) as a peroxy acid. This reaction proceeds by the 'butterfly mechanism', where the peracid is intramolecularly hydrogen-bonded at the transition state. The high degree of polarization induced an electrophilic oxygen atom added to the alkene.

200 mg Substrate **24** was added into the solution containing 4.2 ml methylene chloride and 101.6 mg sodium bicarbonate. The mixture was stirred and cooled in a 0 ℃ ice-water batch. A solution of 0.3 g *m*-CPBA (70%, Sigma Aldrich) in 5.94 ml methylene chloride was added slowly into the mixture over a period of 5 minutes. A heavy white precipitate formed immediately. After that, the whole mixture was stirred at 0 ℃ for 30 minutes and then transferred to room temperature. After 2 hours, equal volume of saturated sodium carbonate was added twice for washing. The separated organic layer was dried over sodium sulfate.

2.8.3 Chemical synthesis of substrate analog 25

The substrate analog **25** was chemically synthesized by Dr. Chen Xi's group in the Northwest University in China. It included 15 steps, which started by preparation of *tert-butyl*(*pent*-4-*yn*-1-yloxy)diphenylsilane. The synthetic steps are shown in the illustration 2.1. After the last synthetic step, the final product was purified by flash column chromatography (SiO_2, 30% ethyl acetate/hexane) and further purified by reverse phase high performance liquid chromatography (HPLC, UniSil 10-120 C18 Ultra column, 10 μm particle size, 21.2 mm ID x 250 mm length). The final product was monitored at 215 nm ultraviolet wavelength. The HPLC washing condition included a gradient of 20-95% methanol for 30 minutes, then 95% methanol for 30 minutes at 15 ml min^{-1} flow rate. The final product structure was confirmed by 1H NMR spectra.

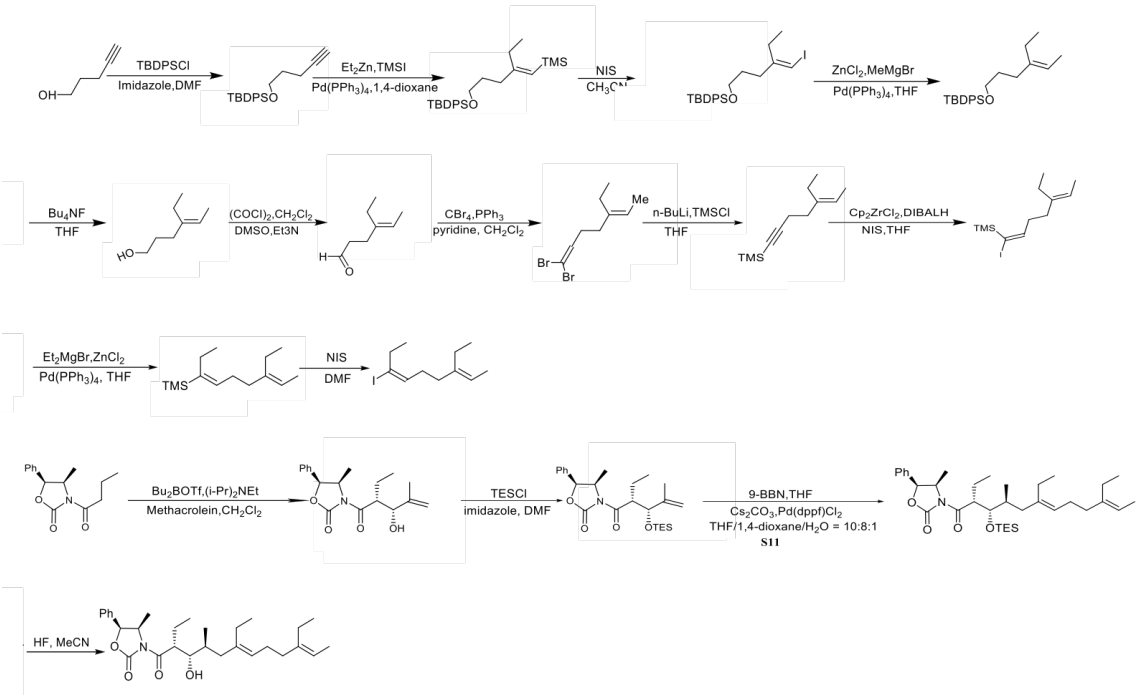


Illustration 2.1: Chemical synthesis of the substrate analog 25.

^{1}H NMR (600 MHz, $CDCl_3$) δ 7.40 (d, J = 7.6 Hz, 2H), 7.36 (dd, J = 15.4, 8.0 Hz, 1H), 7.29 (d, J = 7.4 Hz, 2H), 5.63 (d, J = 7.3 Hz, 1H), 5.16 (q, 1H), 5.08 (q, 1H), 4.80 (t, 1H), 4.15 (d, J = 3.7 Hz, 1H), 3.58 (s, 1H), 2.05-2.01 (m, J = 23.1 Hz, 8H), 1.90-1.85(m, 2H), 1.74-1.69(m, 3H) 1.62-1.55(t, J = 13.7 Hz, 5H), 0.96-0.90 (d, J = 30.9 Hz, 12H), 0.86-0.84(t, 3H).

Chapter 3: Results

3.1 PROTEIN PURIFICATION AND CHARACTERIZATION

3.1.1 Expression and purification of the recombinant proteins

Both *monCI* and *lsd18* genes were inserted into pCold I vector to facilitate protein expression and folding. The Uniprot code for the MonCI is Q846W9; the Uniprot code for the Lsd18 is B5M9L6. The GenBank code for the *monCI* gene is ANZ52466.1; the GenBank code for the *lsd18* (*lasC*) gene is BAG85033.1. The *monCI* and *lsd18* genes were first inserted into the pET-28a(+) vector. Even though the expression level of these two proteins were high, the yield of the soluble protein was very low, which means that neither MonCI nor Lsd18 protein was able to be folded very well during expression. So, we inserted both of the genes into the pCold I vector that is low-cost and efficient to fold the recombinant proteins in *E.coli*. The pCold I vector was a DNA cold-shock expression system purchased from TaKaRa. It uniquely contains the cold-shock protein A (cspA) promoter. These vectors can specifically induce the target protein expression and synthesis at a low temperature, at which condition the host protein expression is suppressed and the activity of proteases is also decreased. It would protect the target protein and result in high yield and purity of the target protein. Besides the cspA promoter, the pCold I vector also includes other elements such as the *lac* operator for controlling the protein induction, ampicillin resistance gene, his-tag sequence, and a Factor Xa cleavage site for removing the His tag from the target protein.

The plasmid *monCI*-pCold I was introduced into the *E. coli* BL21-AI. Other bacterial strains like Bl21(DE3) and Lemo21(DE3) were also tried but the expression level was lower compared with BL21-AI due to the form of inclusion bodies. The chaperone plasmids were also tried but they couldn't prevent or recover MonCI precipitation sufficiently. *E. coli* BL21-AI strain contains the gene of T7 RNA polymerase inserted into the *ara*B locus of the *ara*BAD operon. The regulation of T7 RNA polymerase is under control of the arabinose-inducible *ara*BAD promoter, which makes this strain especially useful for tight regulation and expression of protein that may

be toxic to other BL21 strains due to the leaky expression of T7 RNA polymerase. Due to the relatively low yield of MonCI, four liters of cell culture were combined together for one-time protein purification.

The recombinant 55.12-kDa MonCI protein was purified by the methods including IMAC, IEX, and SEC. 10% (v/v) glycerol added in the buffer solution plays an essential role in stabilizing MonCI proteins during purification. Without glycerol, MonCI proteins would form heavy precipitation and cannot stay in homogenous state. Fractions were collected at each purification step and analyzed by SDS PAGE. (Figure 3.1) The final purity of MonCI protein sample was higher than 95%. Purification under native conditions gave a yellow solution which is due to the presence of its flavin cofactor. The yield of MonCI was approximately 1.6 mg per liter of culture. The final purified MonCI protein samples was concentrated into 6.2 mg ml^{-1} for storage (Figure 3.1A).

The plasmid *lsd18*-pCold I was introduced together with the chaperone plasmid pG-KJE8 in *E. coli* BL21(DE3) bacteria. Recombinant Lsd18 with a N-terminal His was expressed in soluble form with the help of chaperons Dnak, DnaJ, GrpE, GroES and GroEL. These chaperones mediate the correct folding of newly synthesized proteins. DnaK and DnaJ belong to the heat shock protein 70 system, binding with the exposed hydrophobic region and stabilizing the growing peptide chain during translation. After translation, chaperone GrpE transport the unfolded polypeptides into the chamber formed by chaperones GroEL and GroES. This closed environment facilitates the correct folding of the peptide chains.

The recombinant 52.5-kDa Lsd18 was purified by several methods, including IMAC, IEX and SEC. Due to the high pI value for Lsd18, which is about 7.26, the buffer pH value we applied for anion exchange chromatography was also relatively higher. The IEX buffer we used was 20 mM Tris pH8.5. In this condition, Lsd18 was able to be purified and remained stable. We also tried to change the Lsd18 buffer to pH value lower than 7.0. However, Lsd18 protein samples would for aggregations very fast, which indicates that Lsd18 is not stable at the buffer with pH lower than 7.5. Fractions were collected after each step and analyzed in SDS PAGE (Figure 3.1).

Purification under native conditions gave a yellow solution which is due to the presence of its flavin cofactor. After the final step SEC, Lsd18 protein samples was higher than 95%, and the yield was approximately 5 mg per liter of culture. The sample was further concentrated to 5.5 mg ml^{-1} for storage (Figure 3.1B).

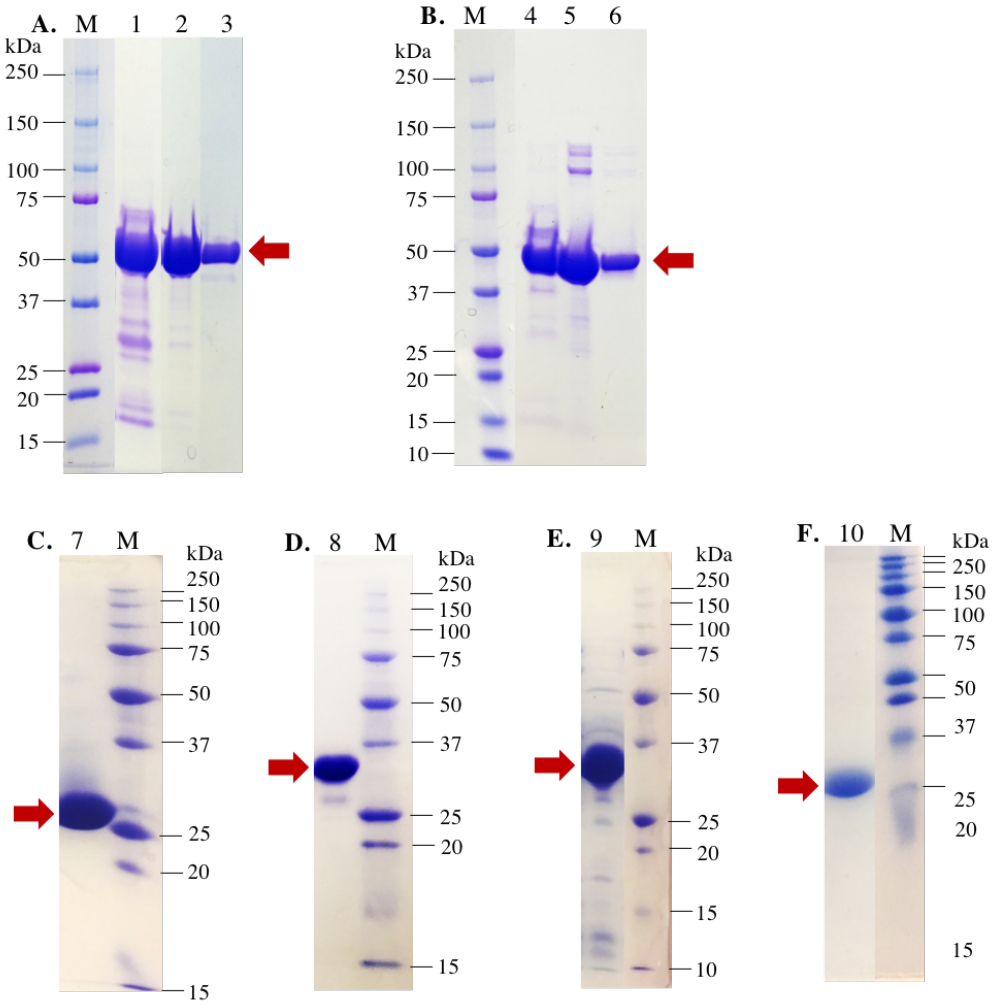


Figure 3.1: SDS-PAGE analysis of purified proteins. Lanes: M, protein marker; 1, MonCI IMAC eluate; 2, MonCI IEX eluate; 3, MonCI SEC fraction; 4, Lsd18 IMAC eluate; 5, Lsd18 IEX eluate; 6, Lsd18 SEC fraction; 7, Fre SEC fraction; 8, GDH SEC fraction; 9, Lsd19 IEX eluate; 10, MonACPX IMAC eluate. The red arrowheads indicate the recombinant proteins accordingly.

To facilitate Lsd18 crystallization, the Lsd18 needs to get through the reductive alkylation process during the protein purification. Reductive alkylation belongs to chemical modification that

is able to alter protein surface properties and crystal behavior. Reductive alkylation of the protein is a very simple method. It used the reducing agent dimethylamine-borane complex and formaldehyde (reductive methylation), acetone (reductive isopropylation), or acetaldehyde (reductive ethylation) to modify the ε-amino group of lysine or the *N*-terminal amino group that are exposed to the solvent. Reductive methylation and ethylation can prepare both mono- and di-alkyl derivatives. In the case of Lsd18, we tried all formaldehyde, acetone, and acetaldehyde. At the end, only acetaldehyde was effective to promote Lsd18 protein to form diffraction-quality crystals. The reductive ethylation mechanism involves a nucleophilic addition of an unprotonated lysine ε-amino group to acetaldehyde and formation of an *N*-ethyl moiety. Then the intermediate dehydrates and hydrogenate to produce monoethylated lysine. The equation of the reductive ethylation is shown as below:

$$R\text{-}NH_2 + CH_3CHO \rightarrow R\text{-}N{=}CHCH_3 + H2O \rightarrow R\text{-}N\text{-}CH_2CH_3$$

$R\text{-}NH_2$ is a lysine or *N*-terminal amino acid of the protein backbone. Then the partially ethylated lysine reacts with acetaldehyde to produce the final diethylated lysine due to the lower pKa of monoethyllysine than that of the lysine. The equation is shown as below:

$$R\text{-}N\text{-}CH_2CH_3 + CH3CHO \rightarrow R\text{-}N(CH_3)\text{-}CH_2CH_3 \rightarrow RN\text{-}(CH_2CH_3)_2$$

After reductive ethylation, the Lsd18 crystal transformed from 2D into 3D. According to the solved crystal structure of Lsd18, there are two lysine residues at the protein surface got ethylated.

The *fre* gene was synthesized by GenScript, then inserted into the pET-28a(+) vector and transformed in *E.coli* BL21(DE3) for protein expression. The Uniprot code for the Fre protein is P0AEN1. The GenBank code for the *fre* gene is BAE77459.1. The yield of Fre was about 6 mg per liter of cell culture. After purification, the Fre solution displayed a tint of yellow color, which was due to the bound flavin. According to the SDS PAGE, the Fre sample was higher than 90% purity which is pure enough for conducting the enzyme assays (Figure 3.1C).

The GDH plasmid inserted in the pET28b vector was a gift from Professor Keatinge-Clay from University of Texas at Austin. The Uniprot code for the GDH protein is P12310. The GenBank code of the *gdh* gene is BAA09024.1. The purification of GDH only include two parts:

IMAC and SEC. The whole purification protocol was relatively simple due to the high binding specificity between his-tagged GDH and the nickel IMAC column. Based on the SDS PAGE analysis, the final GDH protein sample had over than 95% purity, which is pure enough to be used for the enzyme assays (Figure 3.1D).

The *lsd19*-pCold I plasmid was heritage from Dr. Xi Chen who worked on Lsd19 project in out lab before. The Uniprot code for the Lsd19 protein is B6ZK72. The GenBank code for the *lsd19* (*lasB*) gene is BAG85034.1. The purification protocol of Lsd19 followed the method mentioned previously.(113) Since Lsd19 encountered the misfolding issue when expressed in *E.coli*, it was also inserted into the pCold I vector and expressed together with the chaperon plasmid pG-KJE8. To get pure samples for the enzyme assay, the *N*-terminal his-tagged Lsd19 protein sample was purified by two steps IMAC and IEX. According to the SDS PAGE, the purity of the final protein sample was higher than 90%. The final yield of Lsd19 was 2 mg per liter of the cell culture (Figure 3.1E).

The gene of *monACPX* was synthesized by GenScript. The Uniprot code for MonACPX is Q846Y3. The GenBank code for the *monACPX* gene is AAO65789.1. The MonACPX protein was expressed in the *E.coli* BAP1 bacteria strain. This special bacterial strain was heritage from Dr. Chaintan Khosla's lab in the Stanford University.(178) The genotype of the BAP1 strain is BL21(DE3) Δ*prp*RBCD::T7prom-*sfp*,T7prom-*prp*E. Through the homologous recombination, the single copy of the *sfp* gene was integrated in the *prp* opern of BL21(DE3) strain, which is also under the control of the T7 RNA polymerase promoter. The *prpE* gene which encodes the enzyme responsible for converting propionate into pronyl-CoA was also placed under the control of T7 promoter. Therefore, the Sfp and PrpE enzymes were expressed together with MonACPX after IPTG was added to the cell culture. The Sfp enzyme is a 4'-phosphopantetheinyl transferase from Bacillus subtilis (strain 168). It transfers the 4'-phosphopantetheinyl moiety of CoA to a conserved serine residue of MonACPX, which transforms MonACPX from the *apo* form to the *holo* form through thioesterification reaction. Once MonACPX is activated, it is capable to bind with the substrate, then interacts with and deliver the substrate to different unit enzymes involved in the

polyketide synthases, such as acyltransferase, ketosynthase, and epoxidase. Since MonACPX does not contain any aromatic amino acids, neither tyrosine nor tryptophan in its sequence, MonACPX protein sample did not have the absorption at UV 280nm. There was no UV absorption peak observed for MonACPX in the chromatography profile, which made it a little bit difficult to monitor and purify the protein samples. At the beginning, all the factions collected through the washing step were analyzed by running SDS PAGE to confirm where the MonACPX protein was eluted out and how pure the protein sample was. After a series of method optimization, we included the two steps, IMAC and SEC, for the MonACPX purification. After purification, the MonACPX sample was concentrated and kept at -80 ℃ for storage. The standard curve of the Bradford (Coomassie) protein assay is shown in the figure 3.2. According to the SDS PAGE, the final sample was higher than 90% purity (Figure 3.1F). The linear standard curve of the Bradford protein assay was generated by using a series of the BSA concentrations as the x-axis and the UV absorbance at 595 nm as y-axis. The equation of the linear regression was y=4.5282x-0.011. The R^2 value is 0.9925 and all the dots are very close to the regression line, which means that this is a good fit model and the result is highly accurate and reliable. Based on its UV absorbance at 595 nm, the MonACPX concentration was determined to 0.94 mg ml^{-1} after purification. The yield of MonACPX protein was 0.282 g per liter of the cell culture.

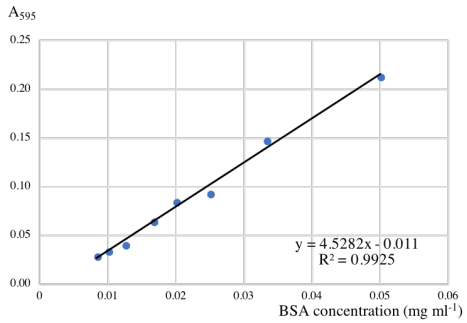


Figure 3.2: Bradford assay standard curve. The concentration of MonACPX was determined by using the equation y=4.5282x-0.011 with an R^2 value of 0.9925.

3.1.2 Column calibration

The gel filtration has been used as a way to determine the molecular weight or size by Stokes radius of proteins, which assumes that there is a sigmoid relationship between the elution volume values such as K_{av} and the logarithm of the protein molecular weight. The K_{av} value is the proportion of pores available to the molecule. In our column calibration method, the Blue Dextran 2,000, which is a high-molecular-weight glucose polymer, was injected into the size-exclusion chromatography column. Since Blue Dextran 2,000 has significantly high molecular weight, it did not go through the pores of the gel filtration column, which means that the elution volume of Blue Dextran 2,000 can be treated as the void volume of the column. Considering the theoretical molecular weight of MonCI (55.1 kDa) and Lsd18 (52.5 kDa) monomers, four different protein standards were chosen to make sure that experimental K_{av} values of the MonCI and Lsd18 can be fit in the range of the calibration standard curve. Those four standards are conalbumin (75 kDa), carbonic anhydrase (29 kDa), human intestinal trefoil factor dimer (13.7 kDa), and human intestinal trefoil factor monomer (6.86 kDa)(Figure 3.3 A). To achieve a successful molecular weight determination, the protein standards were prepared freshly in the buffer of 20 mM Tris pH8.5 and 100 mM sodium chloride. Also, the injection volume was 500 ul which is less than 2% of the total column volume. Based on the elution volume (V_e) of the protein molecules, the column void volume (V_o, elution volume of Blue Dextran 2,000), the total bed volume (V_t), the K_{av} value of each protein sample was able to be calculated as the equation shown below:

$$K_{av} = (V_e - V_o) / (V_t - V_o)$$

The linear regression was derived by using the logarithm of the protein molecular weights, LogMr values, as the x-axis, and their K_{av} values as the y-axis. The equation is calculated as y=-0.3267x + 1.665 with the 0.9984 R^2 value indicating that this equation is significantly accurate and reliable. Since MonCI and Lsd18 were injected into and eluted out from the size-exclusion column, their K_{av} values were able to be calculated. The K_{av} value for MonCI was 0.1195, while the K_{av} value for Lsd18 was 0.1482. Also, their experimental molecular weights were derived according to the equation. For MonCI protein, its experimental molecular weight was 53.8 kDa; for Lsd18, its experimental molecular weight was 43.9 kDa (Figure 3.3 B). Both of the K_{av} values were very close to their theoretical molecular weights, which indicated that both MonCI and Lsd18 exist as monomers in solution.

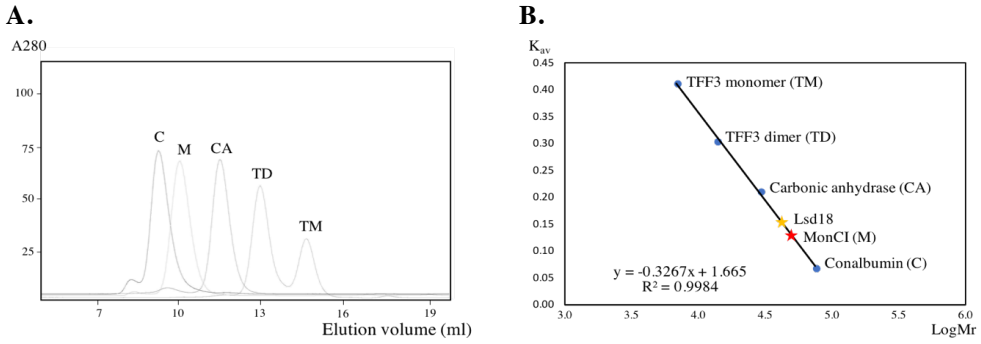


Figure 3.3: Size-exclusion chromatography analysis of MonCI and Lsd18. Standard proteins: human intestinal trefoil factor monomer (6.86 kDa), human intestinal trefoil factor dimer (13.7 kDa), carbonic anhydrase (29 kDa), conalbumin (75 kDa). Superdex 200 10/300 GL column (GE Healthcare) was used. (A) Size-exclusion chromatography profile of MonCI and standard proteins. (B) The calibration curve with LogMr as the x-axis and Kav as the y-axis. Based on the experimental Kav values, the positions of MonCI and Lsd18 are indicated by red and yellow asterisk mark on the calibration curve.

3.2 ENZYME ASSAY RESULTS

3.2.1 UV absorption spectrum

In nature, FAD is the fully oxidized form (quinone form) of flavin adenine dinucleotide. It can also exist in other redox states including flavin-N(5)-oxide, $FADH^{\bullet}$ (semiquinone form), and $FADH_2$ (hydroquinone form). Switching among these states can happen by accepting or donating

electrons. The fully oxidized FAD has an aromatic ring system, exhibiting yellow color; half reduced FADH is either blue or red depending on the solution pH value; fully reduced $FADH_2$ has high polarizability and doesn't own an aromatic ring system like FAD, so it is colorless. Catalytical cycle including flavin reduction by NAD(P)H in monooxygenases is shown in illustration 1.2. To observe the FAD reduction in MonCI, sodium dithionite was added into the MonCI solution in order to force FAD reduced. Since FAD has a typical absorbance at 350 and 450 nm while $FADH_2$ doesn't have, a distinct decrease of the UV absorption spectrum was observed once sodium dithionite was added. The absorption peak at 450 nm disappeared (Figure 3.4 A and B). Due to the aerobic condition, $FADH_2$ wasn't remained too long before turned back to its oxidized form. UV absorption increased back to the initial level (Figure 3.4 C). Color changing from yellow to clear, and back to yellow was also observed during the incubation. All these experiments showed that FAD is bound to MonCI in nature and is able to be reduced and oxidized during MonCI catalysis.

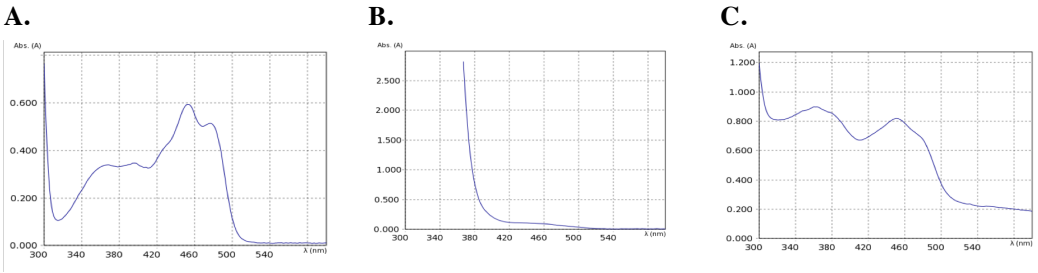


Figure 3.4: UV absorption spectra for MonCI protein samples. (A) Spectrum of the purified MonCI. (B) Spectrum after 2-minute incubation with sodium dithionite. (C) Spectrum after 90-minute incubation with sodium dithionite.

To estimate the molar ratio of our FAD and the purified enzymes MonCI and Lsd18, similar UV absorption spectrum analysis was carried out in our lab. Standard FAD compound was dissolved in the buffer containing 20 mM Tris pH8.5, 200 mM sodium chloride, and 10% glycerol with the gradient concentrations of 5 μM, 10 μM, 50 μM, and 100 μM. These FAD standard samples were scanned from 300 to 600 nm UV spectrum. Characteristic absorption peaks were observed at 362 and 451 nm. The purified protein solutions and their extracted FAD samples were also scanned from 300 nm to 600 nm UV wavelength. The cofactor FAD has a typical maximum

absorption at 450 nm. There is a proportional relation between the UV absorption and the FAD concentration in solution. The absorbance at 450 nm for 5 μM FAD standard was 0.042; for 10 μM FAD standard was 0.088; for 50 μM FAD standard was 0.449; and for 100 μM FAD was 0.836 (Figure 3.5 A). Based on these measurements, the equation was derived as y=0.0084x+0.008, where the absorbance at 450 nm was treated as the y-axis and the FAD concentration (μM) was treated as the x-axis. The R^2 value was 0.998, which indicated that this standard model is considerably accurate and reliable (Figure 3.5 B).

Since there is very limited study about MonCI, which has actually never been heterologous expressed and purified successfully, the MonCI cofactor still remains mysterious. Based on the protein sequence analysis, MonCI also has two FAD-binding motifs: GD and GxGxxG (GGGSGG in MonCI).(179, 180) UV absorption spectrum were used to analyze MonCI cofactor. All the samples were scanned from 300 to 600 nm in the spectrophotometer. The purified MonCI solution also exhibited characteristic UV absorption spectrum at 362 and 451 nm (Figure 3.5 C). The extracted FAD from the denatured MonCI also had nearly the same spectrum with the FAD standards, which indicates that MonCI naturally possesses FAD in a tightly bound form (Figure 3.5 D). The absorbance at 450 nm for the purified MonCI protein solution (4.66 mg ml^{-1}, 84.04 μM) was 0.561, which means that there was about 65.83 μM FAD in the protein solution based on the equation of FAD standards. Therefore, the molar ratio of FAD and MonCI was estimated as 0.78. Low MonCI protein stability may contribute to this result.

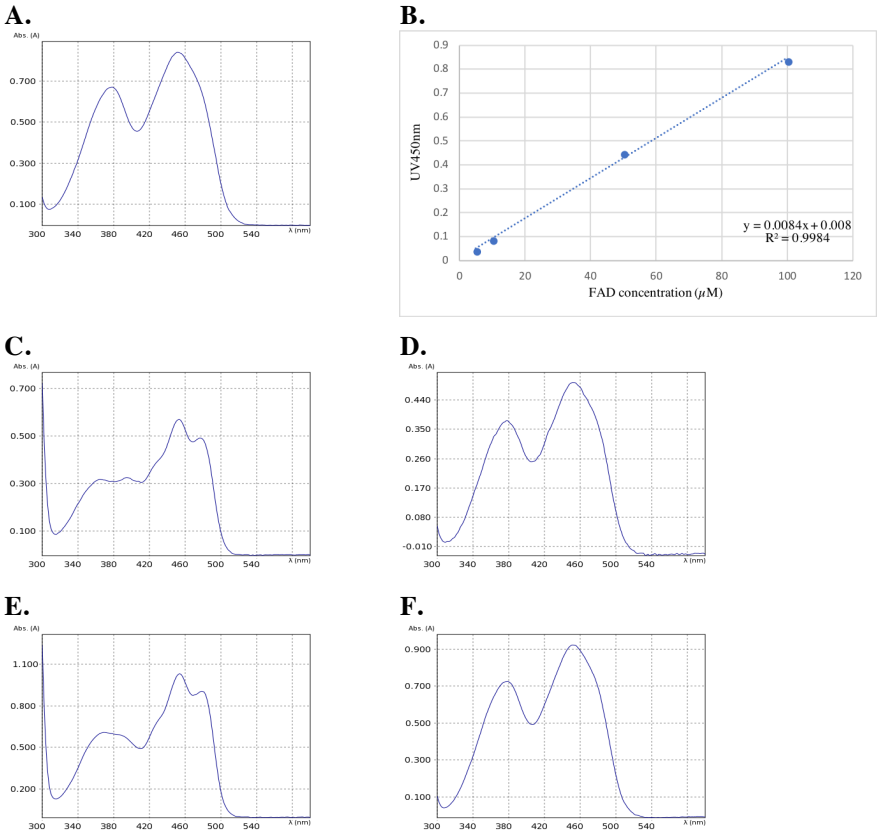


Figure 3.5: UV absorption spectrum for FAD of different samples. (A) 100 μM FAD standard. (B) Linear plot of FAD standards and their absorbance at 450 nm. (C) Purified MonCI solution. (D) Extracted FAD sample from denatured MonCI solution. (E) Purified Lsd18 solution. (F) Extracted FAD sample from denatured Lsd18 solution.

In 2012, Oikawa's group studied the cofactor requirements of Lsd18. Based on the protein sequence analysis, Lsd18 has conserved FAD-binding motifs GD and GxGxxG (GGGMAG in Lsd18).(179, 180) The purified Lsd18 solution showed typical UV absorption spectrum of flavoprotein at 362 and 451 nm (Figure 3.5 E).(103) Denatured Lsd18 samples by methanol was also ran in reverse-phase HPLC. It gave a major peak identical to the standard FAD, which means that FAD is tightly bound to Lsd18 in nature (Figure 3.5 F).(103) The absorbance at 450 nm for the purified Lsd18 protein solution (7.55 mg ml^{-1}, 143.81 μM) was 1.006, which means that there was about 118.81 μM FAD in the protein solution. According to the FAD standards, the molar ratio of FAD and Lsd18 was calculated as 0.83, consistent with Oikawa's result. UV absorption

of the extracted FAD sample was basically the same as the standards, further confirming that FAD exist as a tightly bound cofactor in Lsd18 in nature.

To investigate how substrate-binding affects the cofactor FAD and the conformation of MonCI, the UV spectra were recorded for the purified MonCI protein sample with and without adding the substrate analog linalool. Two same MonCI protein samples were prepared as 21 mg ml^{-1} in the buffer containing 20 mM Tris pH7.5, 15 mM sodium chloride, and 20% glycerol. The 3.9 μl 0.1 M linalool in methanol was added into one of the protein solution, while 3.9 μl methanol was added into the other protein solution. The blank solution is 20 mm Tris pH7.5 and 20% glycerol. As shown the figure 3.6, after adding linalool or methanol, the UV spectra were recorded immediately from 300 nm to 600 nm. There was no obvious difference between the UV spectra of these two MonCI protein samples, which either means that linalool did not interact with MonCI or means that the entrance of the substrate linalool does not affect the FAD status or the MonCI conformation. After incubating overnight at 4 ℃, there still no significant difference in the UV spectra of the two samples. It indicates that linalool entering into the substrate-binding pocket does not affect the MonCI conformation and FAD cofactor oxidation status significantly.

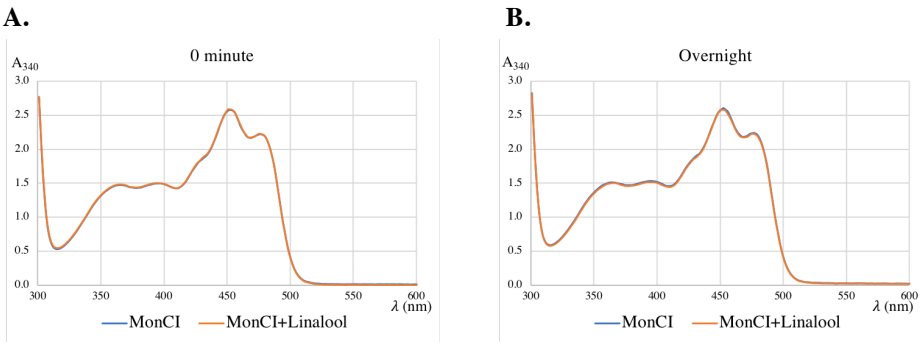


Figure 3.6: The UV absorption spectra for MonCI protein samples with and without adding the substrate analog linalool. (A) The UV spectra measured immediately after adding linalool; (B) The UV spectra measured after incubating the protein-substrate sample overnight.

To investigate how substrate-binding affects the cofactor FAD and the conformation of Lsd18, the UV spectra were recorded for the enzyme Lsd18 with and without adding the substrate

analog linalool based on time. Two same protein samples were prepared with 60 μM final Lsd18 concentration in the buffer containing 20 mM Tris pH7.5 and 100 mM sodium chloride. The 10 μl 10 mM linalool in dimethyl sulfoxide was added into one of the protein samples; the 10 μl dimethyl sulfoxide was added into the other protein sample. The blank solution contained 20 mM Tris pH7.5 and 100 mM sodium chloride. As shown in the figure 3.7, the absorbance of the samples was measured from UV 300 nm to 600 nm immediately after adding linalool or dimethyl sulfoxide. The UV spectra of the Lsd18 sample with the addition of linalool displayed a little bit different from the sample without linalool. However, the overall shape of both UV spectra still matches with the flavoprotein UV spectra. It indicates that the interaction with linalool did not change the cofactor FAD status. The differences between these two UV absorption spectra were probably caused by the conformation change of the Lsd18 protein. After incubating at 4 ℃, the absorbance at 362 nm and 451 nm of both protein samples decreased compared to the 0-minute UV spectra, which means that the FAD cofactor shifted or moved partially. However, differences between the two protein samples became less compared to the 0-minute UV spectra. It was probably because that the linalool had already entered into the substrate-binding pocket and achieved to the maximum occupancy after overnight incubation, therefore the Lsd18 conformation became more stable and turned to be more similar to the initial conformation.

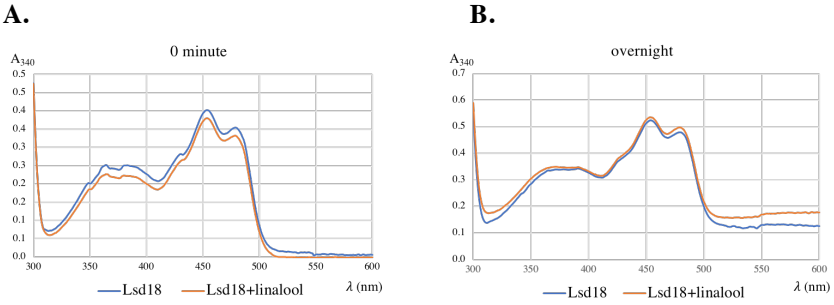


Figure 3.7: The UV absorption spectra for Lsd18 protein samples with and without adding the substrate analog linalool. (A) The UV spectra measured immediately after adding linalool; (B) The UV spectra measured after incubating the protein-substrate sample overnight.

Nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) can be reversibly reduced to NADH and NADPH by the addition of two

hydrogen ions. They play an important role as coenzymes in many reversible reactions. Especially for glycolysis, fatty acid oxidation, and tricarboxylic acid cycle, NAD is used as an acceptor of the reducing equivalents. At the same time, NADH is reoxidized by dehydrogenases or the complex I of the electron transport chain. NADP is involved in the reductive synthesis reactions like steroid synthesis. Due to an additional pair of electrons drawing as a lone pair of electrons, both of the reduced form NADH and NADPH have capability of absorbing the 340 nm-wavelength light. They have the fluorescent emission at 450 nm and excitation at 340 nm. In opposite, the oxidized form NAD and NADP do not have absorbance at 340 nm. Their physical properties allow us to evaluate how active the enzymes are by monitoring the absorbance at 340 nm.

Fre protein is an NADH-dependent flavin reductase. It reduces flavin by transferring the electrons from NADH. Therefore, as the cofactor, NADH gets oxidized into NAD. To check the enzyme activity of Fre, the absorbance of the solution samples was measured at 340 nm. The mixture included the Fre enzyme, its substrate FAD, and its cofactor NADH. The negative control was the same solution without adding the Fre enzyme. The UV absorption spectra showed that the negative control continuously had 340 nm absorbance around 2.0 during 5 minutes, while the absorbance of the Fre solution kept decreasing to nearly 0, which means that amount of NADH stayed the same in the negative control and kept decreasing in the Fre solution (Figure 3.8 A). This UV spectra indicated that the NADH was consumed continuously by Fre enzyme. It verified that the purified Fre enzyme was still active *in vivo*.

GDH protein is a glucose 1-dehydrogenase that reduce NADP to NADPH by transferring the electrons from oxidizing glucose. To check if the purified GDH protein was active, the solution of GDH, glucose, and NADP was measured for the absorbance at 340 nm. The negative control was the same condition but did not contain the GDH enzyme. As shown in the UV spectra, the negative control remained nearly 0 absorbance at 340 nm, while the GDH solution had an increasing absorbance from 0.5 to 2.5 within 1 minute and kept at 2.5 absorbance for the remaining 4 minutes, which means that the negative solution did not form any NADPH and GDH solution generated NADPH and remained at the highest amount in 5 minutes (Figure 3.8 B). This result

implicated that the our purified GDH sample was active and was able to generate the reduced NADPH.

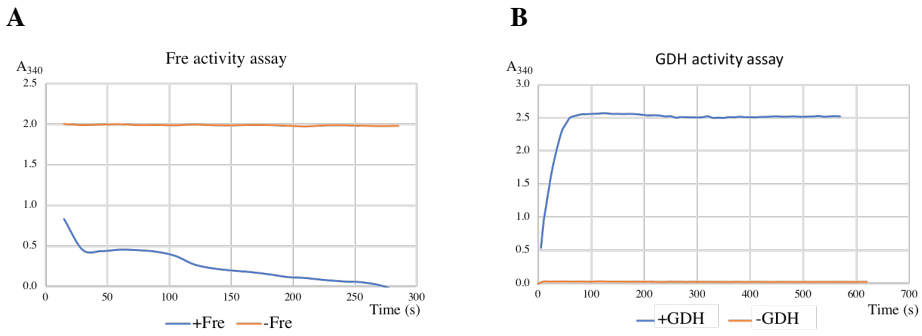


Figure 3.8: The UV absorption spectra of the Fre and GDH activity assays. (A) The UV spectra for Fre consuming NADH; (B) The UV spectra for GDH producing NADPH.

3.2.2 Epoxidation of substrate analogs

The *in vitro* enzyme activities of the purified Lsd18 and MonCI were investigated with the natural substrate analogs linalool **22** and the synthesized compound **24**. The reactions were carried out in the optimized conditions to achieve the highest yield, and the ethyl acetate-extract after incubation were analyzed by GC-MS. We also tried to run the linalool and the compound **24** by LC-MS, but there are no clear peaks shown in the liquid chromatography profile. Since we had the commercial chemical standards of linalool and linalool oxide, we were able to identify their GC-MS peaks based on the m/z of their specific fragments. Linalool is the simplest substrate analog that has been used for Lsd18 and MonCI. Both linalool **22** and linalool oxide **23** were purchased from Sigma Aldrich. Commercial linalool oxide is a mixture of isomers *cis*- and *trans*-**23**. The reactions involved in the *in vitro* enzyme assay are shown in illustration 3.1. The first step of epoxidation is catalyzed by either Lsd18 or MonCI. The second step cyclization is a nonenzymatic reaction, which can happen spontaneously after epoxidation. In GC-MS graph, linalool **22** showed up at 9.16 minutes. Linalool oxide *cis*-**23** showed up at 8.88 minute while *trans*-**23** showed up at 9.06 minute.

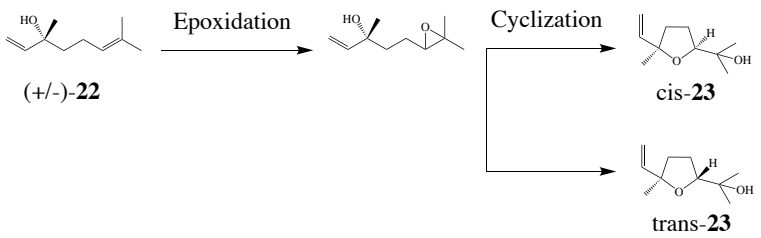


Illustration 3.1: Reactions involved in the enzyme assays. Linalool is used as the substrate analog. Enzyme Lsd18 or MonCI catalyzes epoxidation, followed by nonenzymatic cyclization.

Compound **24** was synthesized by WuXi AppTec company. It has a terminal trisubstituted olefin moiety that mimics the terminal C19-C24 part of prelasalocid **21**. Standard products **25a** and **25b** were synthesized by chemical reactions. Shi's catalyst was involved in the asymmetric epoxidation that prefers yielding (4*R*,5*R*)-epoxide instead of (4*S*,5*S*)-epoxide, while *m*-CPBA catalyst was used for symmetric epoxidation which produce (4*R*,5*R*)- and (4*S*,5*S*)-epoxides in nearly equal amounts. Same as linalool oxide, cyclization of epoxides take place spontaneously without participation of neither enzymes nor catalysts. Enzymatic catalysis of compound **24** is similar to the organic synthetic steps (illustration 3.2). In GC-MS graph, the substrate **24** showed up at 12.35 minute. The products **25** showed up at both 12.26 and 12.30 minutes. Reactions with Shi's catalyst and *m*-CPBA both have 12.26- and 12.30-minute peaks with the similar peak ratio, so whether the products **25a** and **25b** belong to these two peaks respectively hasn't been confirmed.

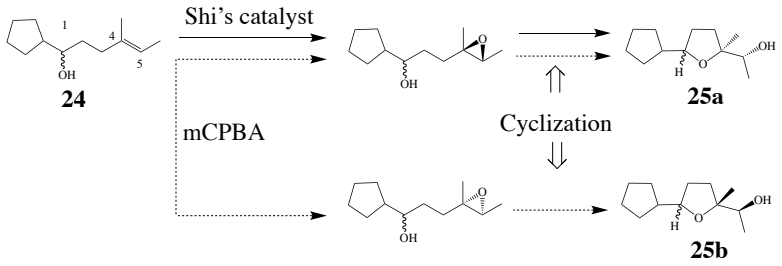


Illustration 3.2: Reactions involved in the organic synthesis and the enzyme assays. Compound **24** is used as the substrate analog. Enzyme Lsd18 or MonCI catalyzes asymmetric epoxidation, while *m*-CPBA catalyst participates symmetric epoxidation. Both is followed by nonenzymatic cyclization.

For MonCI, it is able to catalyze all three carbon double bonds of premonensin **20** with stereoselectivity according to gene deletion experiments. It installs (*R*,*R*)- and (*S*,*S*)-epoxides respectively on the internal and terminal olefins.(84, 85) So far, only *in-vivo* enzyme activity assays have been tried.(84) MonCI was heterologously over-expressed in *Streptomyces*. *Coelicolor* strain and fed with linalool. Compared with the non-transformed negative control strain, MonCI-expressed *S. coelicolor* showed 10-20-fold greater conversion. After MonCI purification, *in-vitro* enzyme activity assays were conducted with the substrates linalool and **24**. The buffer solution was optimized first to achieve the highest efficiency of MonCI. It turns out that pH valued adjusted by sulfuric acid is better that hydrochloric acid. Gradient concentrations of sodium chloride 0, 150, and 300 mM were also explored. MonCI shows the highest activity in 0 mM sodium chloride. These results indicates that the presence of chloride ions may have an inhibiting effect on MonCI activity, which has actually been observed in other flavoproteins.(181, 182) Under the optimized buffer condition, MonCI gave a 35.66% yield of linalool, and a 12.71% yield of compound **24** (Figure 3.9 and Table 3.1). Conversion of **24** was very low, which can be explained by the distinct structural difference between **24** and the MonCI natural substrate **20**. In general, the purified MonCI sample showed a lower activity than Lsd18.

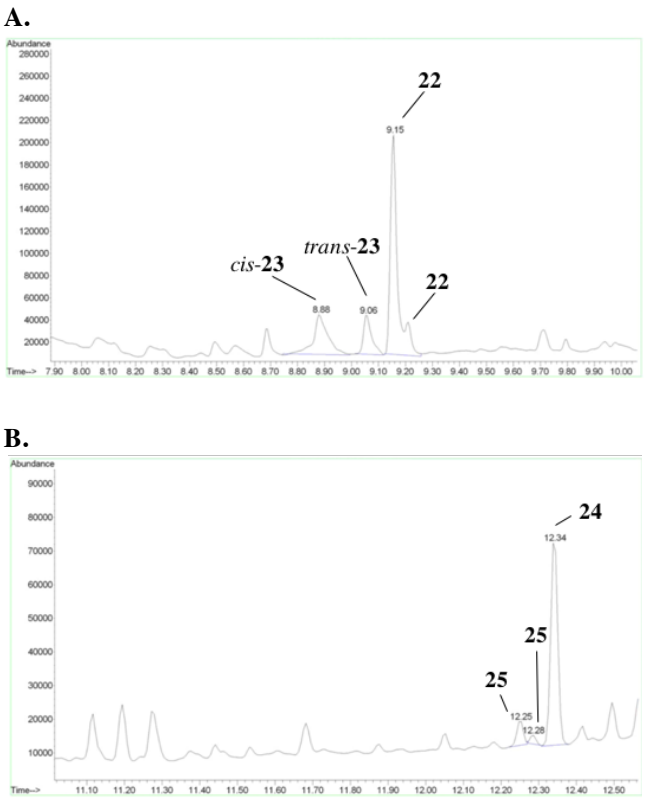


Figure 3.9: GC-MS graphs of MonCI wildtype enzyme assays. (A) Ethyl-acetate mixture after 48-hour enzyme assay with linalool as the substrate. Peaks at 8.88 min and 9.06 min are linalool oxide; peaks at 9.16 min and 9.21 min are linalool. (B) Ethyl-acetate mixture after 48-hour enzyme assay with 24 as the substrate. Peaks at 12.26 min and 12.30 min are the products 25; peaks at 12.35 min is the remaining substrate 24.

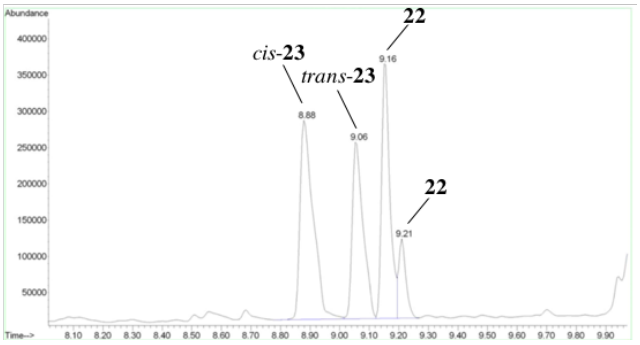
Table 3.1: Estimated percent yield of linalool oxide and **25** catalyzed by wildtype MonCI and Lsd18.

Enzyme	MonCI		Lsd18	
Substrate	Linalool	**24**	Linalool	**24**
Estimated yield %	35.66	12.71	62.92	89.70

For Lsd18, Oikawa et al. found that it catalyzes with its cofactor FAD and NAD(P)H. Both NADH and NADPH is acceptable by Lsd18, but NADH is more preferred with a higher transformation efficiency.(103) Reaction system consisting of Lsd18, FAD, NAD(P)H, and the

substrate is actually sufficient enough, which means that Lsd18 is able to act as both epoxidase and FAD reductase. However, incubating together with the Lsd19 or the flavin reductase like Fre could also help increase its efficiency. In nature, Lsd18 is able to transform both carbon double bonds of prelasalcoid into (*R*,*R*)-epoxides, starting from the terminal olefin. Compared with the linalool and linalool oxide standards, GC-MS analysis showed that Lsd18 is able to catalyze linalool without stereoselectivity, with linalool oxides in a 1/1 ratio of *syn*/*anti* (Figure 3.10 A). Compared with standards **24** and **25**, GC-MS analysis showed that Lsd18 is also able to catalyze compound **24** (Figure 3.10 B). Based on Oikawa group's investigation, Lsd18 can transform compound **24** in a stereoselective manner by esterifying products **25** with (*R*)-2-methyoxy-2-(1-naphthyl) propionate and HPLC analysis. By integrating the GC-MS peak area, the estimated product yield of linalool and **24** was also calculated, shown in table 3.1. Conversion of compound **24** is 89.70% while conversion of linalool is 62.92%, which means that Lsd18 has a higher transformation efficiency with **24** than linalool. It is reasonable because compound **24** has a higher structural similarity with the Lsd18 natural substrate.

A.



B.

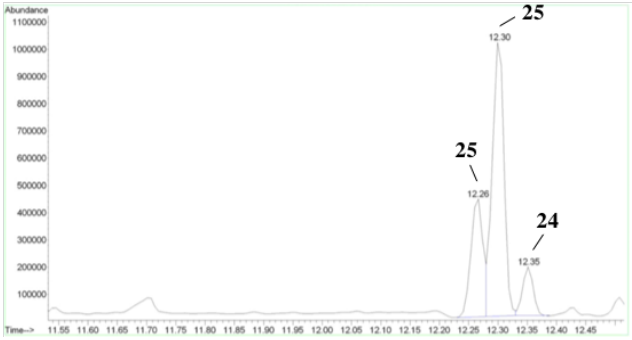


Figure 3.10: GC-MS graphs of Lsd18 wildtype enzyme assays. (A) Ethyl-acetate mixture after 48-hour enzyme assay with linalool as the substrate. Peaks at 8.88 min and 9.06 min are linalool oxide; peaks at 9.16 min and 9.21 min are linalool. (B) Ethyl-acetate mixture after 48-hour enzyme assay with 24 as the substrate. Peaks at 12.26 min and 12.30 min are the products 25; peaks at 12.35 min is the remaining substrate 24.

Besides the substrate analog linalool and **24**, we also synthesize the compound **26** as the substrate analog for Lsd18 enzyme. The compound **26** is more similar to the natural substrate prelasalocid of Lsd18 in structure. It contains the C_{13}-C_{24} part of the natural substrate and a phenyl and a five-membered ring at the other terminal. Since this substrate analog is more similar to the natural substrate, the binding affinity between the substrate and Lsd18 substrate-binding pocket is higher. Therefore, using this substrate analog can increase the transformation efficiency of the Lsd18 enzyme assays, and it is more promising to obtain the crystals of the Lsd18-substrate complex. The reactions that take place during the enzyme assays are shown in the illustration 3.3.

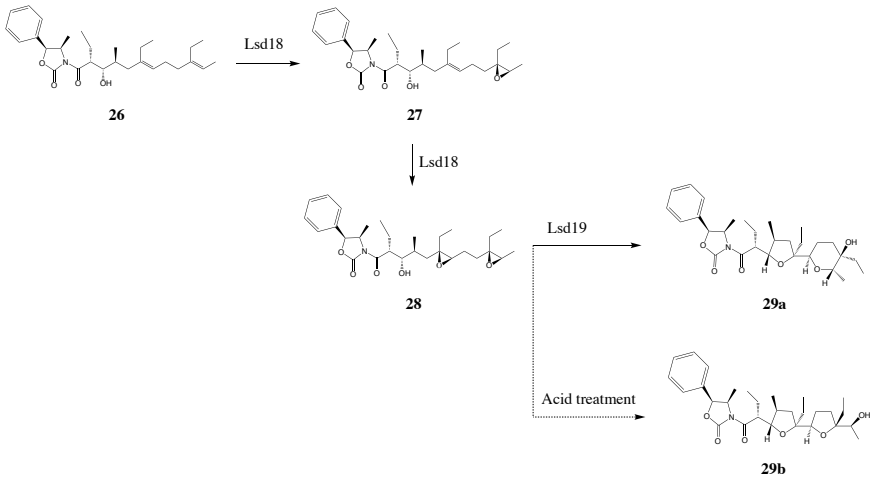


Illustration 3.3: Reactions involved in the enzyme assays. Compound 25 is used as the substrate analog. Enzyme Lsd18 catalyzes stereoselective epoxidation, followed by cyclization catalyzed either by Lsd19 with stereocontrol or by the nonenzymatic acid treatment.

The exact mass for the compound **26** is 469.32. After the terminal carbon double bond gets transformed into an epoxide by Lsd18, the intermediate product **27** has the exact mass 485.31. Furthermore, the internal carbon double bond also gets transformed into an epoxide by Lsd18. The exact mass of this product **28** that contains two epoxides is 501.31. After cyclization catalyzed either by Lsd19 or the acid treatment, the product **28** will be converted into the product **29**, the exact mass of which is also 501.31. The mass spectrometry peaks of the compounds **26**, **27**, **28** and **29** are shown in figure 3.11. The main mass spectrometry peak detected for compound **26** is 470.40 m/z; for compound **27** is 486.40 m/z; for compound **28** or **29** is 502.44 m/z. The extra mass was probably contributed by the attached protons.

Figure 3.11: Mass spectrometry profiles of the compound **26**, **27**, **28** and **29** in the reaction 1 extract.

According to Oikawa's protocol (103, 112), these two carbon double bonds get catalyzed in two reactions in vitro. The terminal carbon double bond gets epoxidized first during reaction 1 which includes Lsd18, Fre, NADH/NADPH, and the substrate **26** in the buffer containing 50 mM Tris pH8.0, 300 mM sodium chloride, and 5% glycerol. Then the intermediate product **27** gets extracted by ethyl acetate and used as the substrate in the reaction 2 that includes Lsd18, Fre, NADH/NADPH, and the substrate in the buffer containing 100 mM Bis Tris pH6.5 and 10% (v/v)

glycerol. The *in-vitro* reaction 2 should only be applied in the epoxidation reaction with two carbon double bonds, because the optimal pH condition in the epoxidation activity for the terminal and the internal carbon double bonds is different. The terminal carbon double bond prefers to getting epoxidized at basic pH like Tris pH8, while in the opposite, the internal carbon double bond prefers to getting epoxidized at acidic pH like Bis Tris pH6.5. The extract samples of the reaction 1 and 2 were analyzed by LC-MS (Figure 3.12). As indicated by the m/z values, the HPLC peaks for each compound were identified. The substrate analog **26** was eluted around 7.79 minute; the intermediate product **27** was eluted at 7.16 minute; and the diepoxide product **28** or the final product after cyclization **29** was eluted at 7.45 minute. After reaction 1, more than half of the substrate **26** was transformed into the product either **27**, **28** or **29**. Since the peak of the product **28** or **29** got detected by LC-MS for the reaction 1 extract, it indicates that Lsd18 has already proceeded to and catalyzed the internal carbon double bond. As shown in Figure 3.12(B), the reaction did not catalyzed all the intermediate product into diepoxide, which was not consist with the Oikawa's results. The difference in the terminal moiety of the substrate 26 could be the reason.

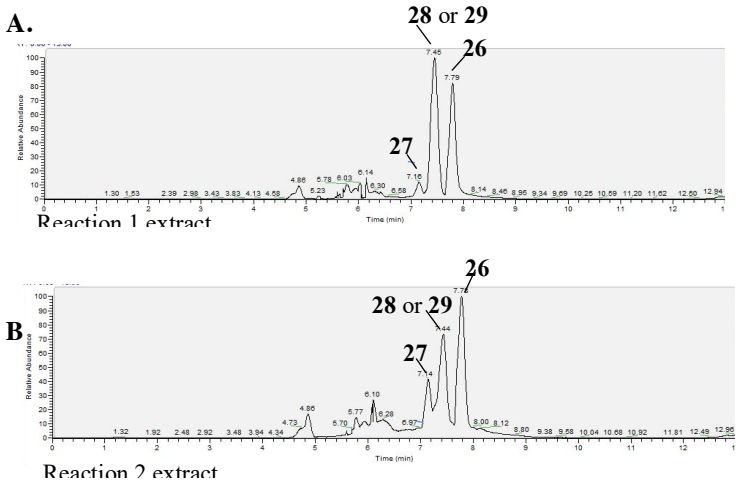


Figure 3.12: HPLC profiles of the reaction 1 and 2 extract in the Lsd18 enzyme assays. (A) The injected sample was the extract after reaction 1. (B) The injected sample was the extract after reaction 2.

Different optimization methods were also employed in the enzyme assays in the purpose of increasing the final product **29** yield. Firstly, Lsd19 was added into the protein solution. Because

Lsd19 catalyzes the cyclization reaction of the compound **28**, it can help increase Lsd18 efficiency by consuming its products. As shown in the figure 3.13, the product peaks are actually lower in the reaction with addition of Lsd19 than the reaction without Lsd19, which means that adding Lsd19 did not help increase the Lsd18 efficiency in our case. Also, we incubated the reaction mixture for a longer time to achieve the maximum yield. As shown in the figure 3.13, the product peak is the highest in the reaction mixture with 24-hour incubation time. We also notice that both the substrate and the products started degrading after 24 hours. The peaks of the 36-hour, 2-day, and 3-day reaction mixtures were much lower compared to the peaks of 12-hour and 24-hour reaction mixtures. Therefore, the optimal condition to achieve the maximum product yield is to incubate the reaction mixture for 24 hours without addition of Lsd19.

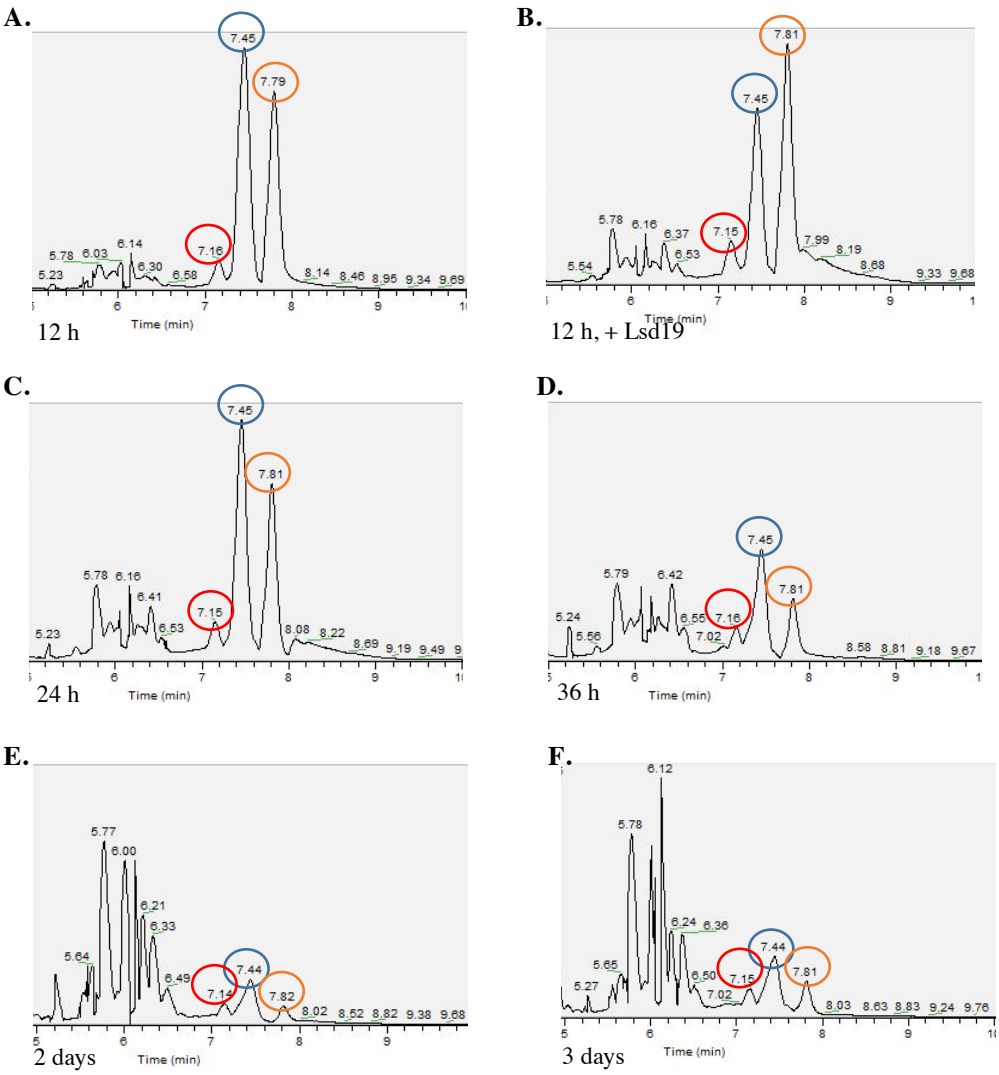


Figure 3.13: The HPLC profiles of the reaction 1 extract from the Lsd18 enzyme assays with the substrate **26**. (A) The sample mixture was incubated for 12 hours before extraction by ethyl acetate. (B) Lsd19 was added to the sample mixture. (C)-(F) Time-course HPLC analysis of the reaction 1 mixture. The sample mixture was incubated for different time periods before extracted by ethyl acetate. The peak for the substrate analog is marked by the orange circle; the peak of the product **27** is marked by the red circle; the peak for the product **28** or **29** is marked by the blue circle.

3.3 CRYSTAL STRUCTURES

To further study the mechanism of the epoxidases MonCI and Lsd18, we solved their atomic resolution structures by X-ray crystallography. Based on the SDS PAGE, the purity of both the MonCI and Lsd18 protein samples were higher than 95%, which is pure enough for

crystallization. The homogeneity of the protein sample plays an important role in obtaining the high-quality protein crystals. X-ray crystallography basically includes three steps: protein crystallization, crystal diffraction by X-ray beam (diffraction data collection), and protein structural determination (diffraction data process).

Protein crystallization is a process of the protein crystal forming.(183) During protein crystallization, it proceeds in two steps: nucleation and growth. Critical nuclei is related to supersaturation state and the surface free energy of the molecular addition.(184, 185) The crystal growth follows mainly by the mechanism of growth by two-dimensional nucleation and dislocation growth. It also follows by the mechanism of three-dimensional nucleation and normal growth.(186, 187) Both of these two steps rely on the supersaturation of the mother liquor. This supersaturated state is a non-equilibrium state when the molecular quantity is in excess of the solubility limit. Formation and development of crystals, a solid state, is a way to re-establish the equilibrium. Therefore, creating the supersaturation state is an essential step during protein crystallization. To promote supersaturation, many different methods have been developed, such as altering temperature, adding or removing salt, changing pH, and evaporation and so on. There are several common techniques to grow protein crystals, combining the methods mentioned above. These techniques include sitting-drop vapor-diffusion, hanging-drop vapor-diffusion, dialysis and microdrops under oil.(188) So far, many of the chemicals have used as the precipitants to promote protein crystallization, which includes salts, organic solvents, nonvolatile organic compounds, low-molecular-weight polymers, and long-chain polymers.

After growing to the full size, protein crystals can be diffracted by high energy level X-ray beam. The protein crystals are first soaked in the mixture of the crystallization solution and the cryoprotectant solution, then can be mounted into a glass capillary or a nylon loop that is attached to a solid rod. After that they get flash-frozen in liquid nitrogen and transferred onto a goniometer that can rotate and adjust the position of the crystals very accurately. The most common goniometer is the kappa goniometer which is able to rotate in three angles: ω angle rotating in the axis perpendicular to the beam, κ angle in an axis about 50° to the ω axis, and $\emptyset$ angle in the

capillary or loop axis. During diffraction, the protein crystals are maintained at very low temperature in order to reduce the radiation damage and the noise in the Bragg peaks caused by thermal motion. Once the crystal gets exposed to the high-energy-level X-ray beam, it scatters or deflects the X-rays into different angles. The pattern of spots or reflections can be observed and recorded on the screen behind the crystal. The spots at small angles correspond to low-resolution data, while the spots at high angles correspond to high-resolution data. The diffraction quality can be assessed such as crystal mosaicity and disorder by checking the peak width. Also, to be able to construct the protein crystal structure, one image of the reflections is not enough because it only provides the information of a small slice in the whole Fourier transform. A complete data set including many images needs to be collected by rotating and diffracting the protein crystals step-by-step at different orientations. Sometimes, multiple diffraction data sets are required to solve the phasing issues.

After collecting the sufficient diffraction images, these two-dimensional diffraction patterns get transformed to a three-dimensional electron density model by Fourier transforms. The diffraction data represents a reciprocal space of the actual crystal lattice. The diffraction spots are decided by the size and shape of the unit cell, and also the crystal symmetry. The general data processing consists of three steps: indexing, integration, scaling, and phasing. Indexing is to identify the crystal space group, the unit cell dimensions, and the image peak positions in reciprocal space. Integration is to determine the intensities by analyzing the same peak in multiple images. Scaling is to scale all images together and make them at a consistent intensity scale. Phasing is to combine all the variation and yield the whole electron density. To get the electron density map, the structure factor that is related with the amplitude and phase of a wave needs to be determined. The wave amplitude can be derived from its proportional relationship with the intensity of the diffraction spots. However, the wave phase cannot be determined directly from the experimental diffraction data, which is known as the phase issue. The phase can be determined by several different ways. One is the *ab initio* phasing, or the direct method solve the phase by using the relation between the phase of one reflection to two other phases considering the fact that

scattering density is a positive real number. This direct method is only applicable for relatively small molecules because the phase probability distribution has an inverse relation to the square root of the number of atoms. Another method to solve phase is by anomalous X-ray scattering, either multiple-wavelength anomalous diffraction (MAD) or single-wavelength anomalous diffraction (SAD) phasing. The atomic scattering factor is decided by three parts: the normal scattering factor f_0, f', and f''. The f_0 is dependent on the Bragg angle. The factor f' modifies the f_0, while f'' is 90° advanced in phase. The other two factors f' and f'' are dependent on the wavelength and represent the anomalous scattering showing up at the absorption edge. For MAD experiment, the diffraction data is collected at several different wavelengths at the absorption peak (f''), at the point of reflection on the absorption curve (f'), and also at the remote where maximizes the dispersive difference to the reflection wavelength. The MAD method normally needs to prepare protein crystals with some heavy atoms inside the protein. Compared with the MAD method, the SAD method only collects one diffraction data at a single wavelength. It provides the measurements of the anomalous that can be used to estimate the contribution of the heavy atoms to the scattering and derive the heavy-atom substructure. Thus, it solves the phase of the whole protein crystal structure. The SAD method is normally used when there are intrinsic anomalous scatters inside the protein, such as the bound ions or the sulfur atoms existing in the methionine and cysteine amino acids. The advantage of using SAD method is that it is more efficient, it needs less time in the X-ray beam, and it reduces the radiation damage to the protein crystals. Another method to solve the phase problem is molecular replacement that uses a known protein homolog structure as an initial search model to help determine the orientation and position of the protein molecules in the unit cell. This method relies on the structural similarity between the search model and the protein structure to be solved. The derived phase can help get the calculated intensities of the protein model to the observed protein structures and it can also be used to generate the electron density maps. The molecular replacement method includes two steps. The first step is to orient the model generated from the protein homologs into the unit cell by a rotational search. The second step is a translational placement of the new oriented protein model in the unit cell. The molecular

replace method has been used more and more due to its conveniences and high efficiency. It does not require extra experimental work like the MAD or SAD methods. It requires less time to determine the protein structures. It is also very easy to be understood and the model building is simplified without the needs of chain tracing.

After the generation of the electron density maps and construction of the initial protein structural model, the protein structure can be finalized by multiple cycles of model re-building and refinement. The refinement process is aimed to make the protein structure model fit better in the electron density maps obtained from the experimental X-ray diffraction data. The various refinement strategies are available in different programs such as Phenix refinement (147) and CCP4 Rafmac (189). The refinement is composed by several steps such as correction for bulk solvent, anisotropic scaling, twinning parameter refinement, ordered solvent, target weight calculation, coordinate refinement, atomic displacement parameter refinement, and occupancy refinement. After multiple cycles of refinement, the optimal protein crystal structure is obtained. The final protein structure model can be evaluated by different parameters such as R_{work}, R_{free}, B values, and root-mean-square deviations (bond lengths and angles).

3.3.1 The crystal structure of MonCI

The initial crystal hits of MonCI came from several different screening conditions, which includes 0.1 M Bis Tris pH6.5, 30% PEG4000; 0.1 M Bis Tris pH6.5, 30% PEG2000mme; 0.1 M MES pH6.5, 29% PEG3350; 0.1 M Imidazole pH6.5, 30% PEG4000; and 0.2 M sodium iodide, 20% PEG 3350. The commercial conditions that yielded the initial crystal hits were B12, C12, D11, and F1 from Anatrace MCSG1 screen, F8 and H6 from Anatrace MCSG4 screen, and C11 from Molecular Dimensions ShotGun screen. The initial crystals displayed as needle clusters in the solution, which was not large enough to be harvested or diffracted. Therefore, we applied various crystal optimization methods to improve the MonCI crystals. It included the optimization of buffer, pH, salt, precipitant, drop size, and temperature, addition of various additives, chemical modification of the MonCI protein surface, and other methods. At the end, the best crystals grew

in a combined buffer containing 0.1 M Bis Tris ph6.5, 0.1 M MOPS pH6.8, 28% PEG2000mme, and 4% glycerol by the sitting-drop vapor-diffusion method, at 18 ℃. The MonCI crystals were in rod shape and displayed in yellow color because of the tightly bound FAD in its oxidized form (Figure 3.14 A).

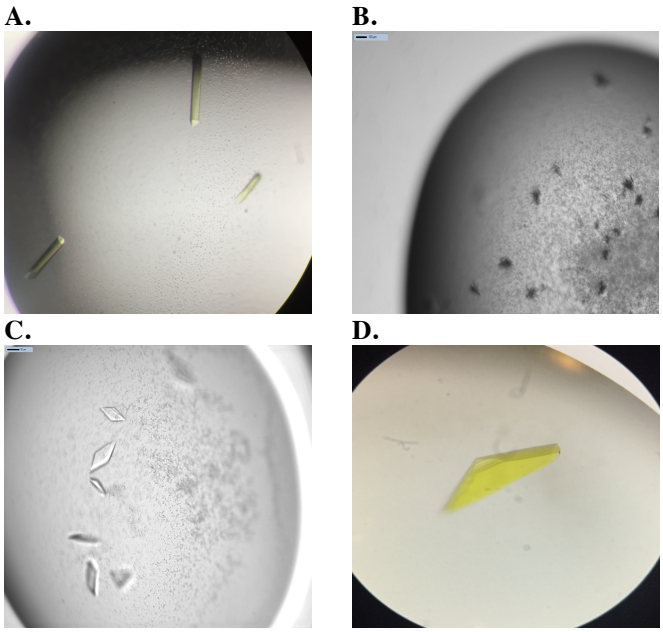


Figure 3.14: Crystal pictures of MonCI and Lsd18. (A) MonCI crystals in color. (B) Initial crystal hits of Lsd18. (C) Optimized crystals of Lsd18. (D) Single Lsd18 crystal in color.

Diffraction data collection and refinement statistics of MonCI crystal structure are shown in table 3.2, which was determined from a single crystal. The unit cell of this triclinic crystal contains two enzyme monomers. After final refinement, the Ramachandran favored is 918 (98%), and the Ramachandran outliers is 2 (0%) based on the calculation by Molprobity (190). At the end, the crystal structure of flavin-bound MonCI was determined at resolution up to 1.90 Å with the R_{work}/R_{free} as 19.20%/23.40% (Figure 3.15). The phase of the MonCI crystal was solved by MAD method using the bromide as the heavy atom inside the protein. Two molecules of MonCI are present in the crystal asymmetric unit (chain A and B). We are able to trace the entire chain A (residues 4-480) while some sections of chain B (residues 147-149 and 231) could not be modelled due to the poor electron density. One FAD cofactor and one chloride ion is found in each chain.

The final MonCI model has 7,180 protein atoms, 106 ligand atoms, 2 chloride ions, 673 water molecules.

Table 3.2: Data collection and refinement statistics for MonCI crystal structure.

	Native	Br derivative	Br derivative	Br derivative
Data collection				
Space group	*P*21	*P*21	*P*21	*P*21
Cell dimensions				
a, *b*, *c* (Å)	61.2, 52.4, 141.6			
α, β, γ (°)	90.0, 93.6, 90.0			
	Native	*Peak*	*Inflection*	*Remote*
Wavelength (Å)	0.9795	0.9200	0.9392	0.9795
Resolution (Å)	1.90 (1.95-1.90)	2.25 (2.31-2.25)	2.25 (2.31-2.25)	2.1 (2.15-2.10)
R_{sym} (%)	12.3 (131)	21.5 (161)	22.3 (166)	18.5 (141)
I/σ(I)	13.6 (1.63)	15.4 (1.31)	12.4 (1.44)	13.5 (1.62)
Completeness (%)	96.4 (95.8)	97.5 (98.3)	97.5 (98.4)	97.6 (98.0)
Redundancy	8.0 (8.1)	33.9 (16.4)	24.3 (24.6)	24.3 (24.6)
$CC_{1/2}$	0.999 (0.658)			
Refinement statistics				
Resolution limit (Å)	39.7-1.9 (1.97-1.90)			
No. of reflections	73,491 (3,632)			
R_{work} / R_{free}	0.192/0.234			
Number of atoms				
Protein	7,180			
Ligand	106			
Ion	2			
Water	673			
B-factors (Å^2)				
Protein	35.2			
Ligand	27.8			
Ion	46.3			
Water	36.8			
R.m.s deviations				
Bond lengths (Å)	0.0086			
Bond angles (°)	1.596			

* This structure was determined from one single crystal.
Values in parentheses are for the highest resolution shell.
$R_{sym} = \Sigma | I\text{avg} - I\text{i} | / \Sigma I\text{i}$, where Ii is the observed intensity and *I*avg is the average intensity of observations of symmetry-related reflections.
$R_{work} = \Sigma | F\text{p} - F\text{p(calc.)} | / \Sigma F\text{p}$, where *F*p and *F*p(calc.) are observed and calculated structure factors; R_{free} is calculated with 5% of the data.

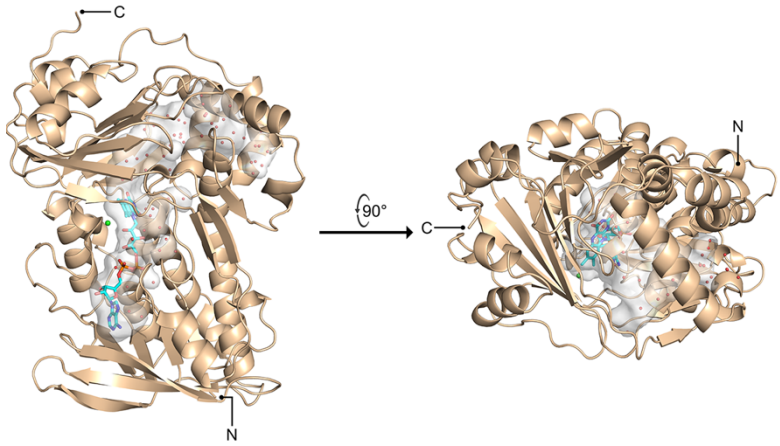


Figure 3.15: 1.9 Å resolution X-ray crystal structure of MonCI. Chloride ion and water are shown as green and red spheres, respectively. The FAD cofactor is drawn as a stick model. The fused FAD- and substrate-binding pocket is shown as a grey surface. For clarity, only those water molecules located inside the pocket are shown.

MonCI has two different entrances through which the substrate and solvent molecules can enter, a side entrance and a top entrance. The two entrances are orthogonally positioned and they both have an opening with diameter of about 6.5 Å. The top entrance is more hydrophilic while the side entrance is more hydrophobic. The side entrance is positive-charged, full of the positive residues (Figure 3.16 A). While the side entrance is connected to active site cavity by a narrow, 8.5 Å-long tunnel, the top entrance immediately leads directly to the active cavity (Figure 3.16 B). We speculate that the top entrance in MonCI functions as a solvent channel. In the MonCI crystal structure, the substrate-binding pocket is occupied by 44 water molecules. As premonensin A enters the pocket, about 24 water molecules are expected to be displaced, and the top entrance provides an exit passage for these water molecules. A similar two-entrance system was previously reported for 3-hydroxybenzoate hydroxylase from *Comamonas testosterone*. But both entrances in 3-hydroxybenzoate hydroxylase connect to the active site cavity via a short tunnel, and they lie parallel to one another (Figure 3.16 C). In agreement with our hypothesis that the side entrance of MonCI serves the substrate and the top entrance serves the water molecules, the protein surface near the side entrance is lined with hydrophobic residues, presumably enhancing with interaction

with the greasy premonensin A, while the surface near the top entrance is lined with hydrophilic residues, promoting interaction with water.

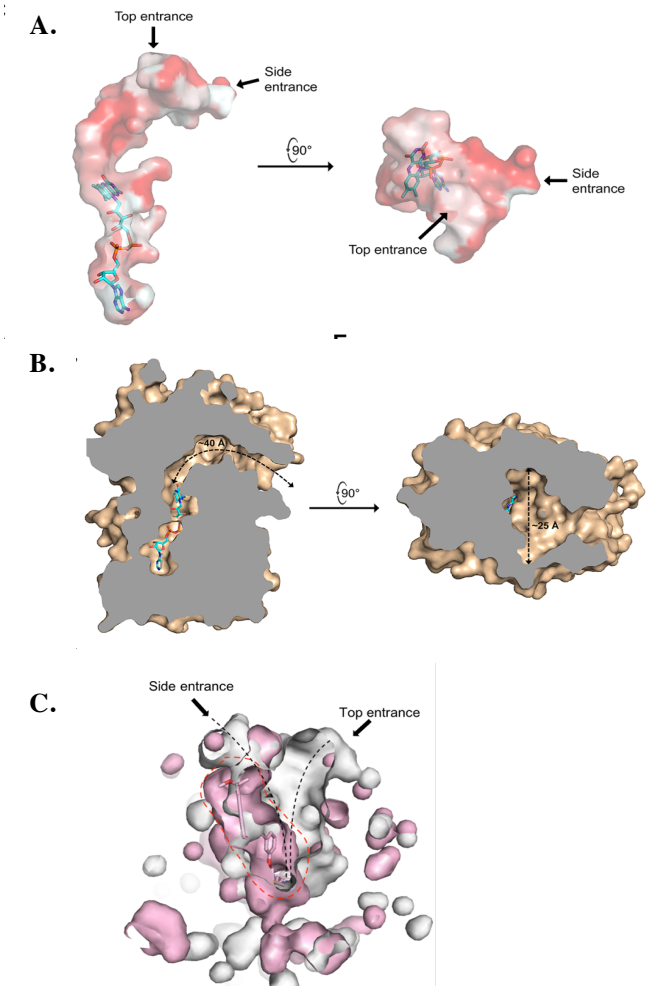


Figure 3.16: The fused FAD- and substrate-binding pocket of MonCI. (A) The top and side entrance that provide access to MonCI's active site cavity. Hydrophobic surfaces are colored in red and hydrophilic surfaces are colored in white. (B) Cross-section of MonCI showing the dimensions of the substrate-binding pocket. The indicated pocket length (~40 Å) is that of the pocket trajectory leading to the side entrance. (C) Substrate-binding pocket alignment of MonCI and human squalene epoxidase (SQLE; PDB ID: 6C6N). The pocket surface of MonCI is shown in white and the pocket surface of SQLE is shown in pink. The SQLE bound Cmpd-4 inhibitor is shown as a stick model and the binding pocket is marked with the red broken line. The two entrances in MonCI and the respective trajectory leading to the active site are marked with black broken lines.

MonCI consists of a substrate-binding domain which is located near the *C*-terminus, and a FAD-binding domain which is located near the *N*-terminus. These two domains are extensively

integrated, where each domain is composed of multiple noncontiguous segment of the protein sequence. The FAD- and the substrate-binding pocket are fused together, resulting in a long, continuous tunnel. The cofactor FAD is located at the bottom of the fused pocket and it is partially exposed to the solvent. The isoalloxazine ring of FAD adopts the so-called "in" position, which is the expected conformation of FAD during epoxidation. The substrate-binding pocket, which lies immediately on top of FAD, is about 40 Å long and about 25 Å wide (Figure 3.16 B). Interestingly, the size of the substrate-binding pocket in MonCI is significantly larger than that of the substrate molecule, premonensin A, which is approximately 30 Å long and 5 Å wide when fully extended. The significance of this size discrepancy is discussed in a later section.

The FAD cofactor makes numerous direct and water-mediated interactions with MonCI residues (Figure 3.17 A). The isoalloxazine ring of FAD makes a π-π stacking interaction with the imidazole group of His57, and its N1, N3, O2, and O4 atoms from hydrogen bonds with backbone atoms of Ala56, Leu58, Gly333, and Met336, respectively. The pyrophosphate group of FAD form hydrogen bonds with Ser15 side chain and Ala17 main chain atoms, and the adenine base of FAD forms a π-cation interaction with the guanidine group of Arg37 and hydrogen bond interaction with Ala140. The FAD-binding domain of MonCI is characteristic of flavin-dependent monooxygenase family that carries a conserved motif GxGxxG and a Rossmann dinucleotide-binding fold with central five stranded parallel β-sheet. This central β-sheet locates between a three-stranded antiparallel β-sheet and a four-helix bundle, forming a three-layer ββα substructure. The flavin phosphate makes hydrogen-binding contacts with the conserved motif GxGxxG (GGGSGG in MonCI) (Table 3.3).

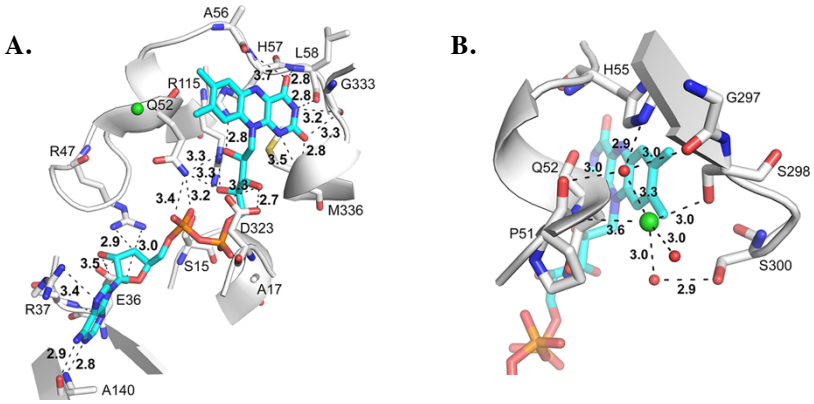


Figure 3.17: Zoomed view of cofactor FAD and chloride in MonCI. (A) Protein environment of the FAD cofactor in MonCI. Interaction distances are given in Å. (B) Protein environment of the chloride ion in MonCI. Interaction distances are given in Å.

One chloride ion was modelled next to the isoalloxazine ring of FAD. The chloride coordinates with five hydrogen bond donors (Gln52 main chain nitrogen, Ser298 side chain hydroxyl group, and three water molecules) in a distorted trigonal bipyramidal stereochemistry (Figure 3.17 B). Structural alignment of the MonCI crystal structure and the NADPH-bound 4-hydroxybenzoate hydroxylase (PDB ID: 1K0J) shows that the chloride in MonCI is located where the adenine based of NADPH is expected to bind.(108) Therefore, chloride is expected to act as a competitive inhibitor of MonCI. In agreement with this theory, chloride was previously shown to competitively inhibit the flavoenzyme *p*-hydroxybenzoate hydroxylase, a structural homolog of MonCI, although the inhibition mechanism was not determined.(181)

Three-dimensional structure alignment using the DALI server (191) identified the following proteins as the closest MonCI structural homologs: bacterial halogenase Bmp2 (PDB: 5BVA; RMSD = 2.9 Å) (192), human squalene epoxidase (PDB: 6C6N; RMSD=3.0 Å) (193), and bacterial *p*-hydroxybenzoate hydroxylase (PDB: 1PBE; RMSD=3.4 Å) (194) (Figure 3.18). These enzyme share 13 to 18% sequence identity with MonCI (Table 3.3). In general, the protein fold of the FAD-binding domains is highly conserved while that of the substrate-binding domain is more variable. The volume of the fused FAD- and the substrate-binding pocket correlates well with the size of the respective substrate molecule. MonCI has the largest pocket volume (995 $Å^3$), followed

by squalene epoxidase (629 Å^3), *p*-hydroxybenzoate (552 Å^3), and Bmp2 (294 Å^3), as calculated using CASTp 3.0.(194) The manner in which these enzymes interact with their substrate is different. MonCI and Bmp2 act on a substrate that is covalently tethered to the 4'-phosphopantetheine prosthetic group of an acyl carrier protein whereas squalene epoxidase and *p*-hydroxybenzoate hydroxylases act on the free substrate. However, there are no obvious distinguishing structural features that would explain this mechanistic difference.

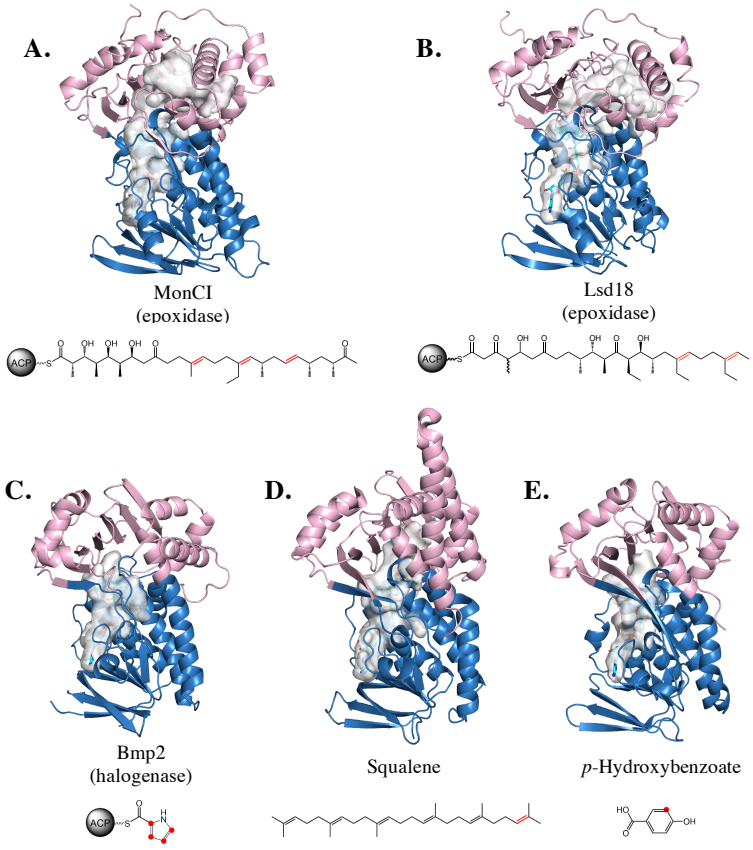


Figure 3.18: X-ray crystal structure of MonCI, Lsd18 and their structural homologs and the chemical structure of the respective substrates. The FAD-binding domain of each protein is colored in blue and the substrate-binding domain is colored in pink. The fused FAD- and substrate-binding pocket is shown in grey. Site(s) of chemical transformation in each substrate molecule is highlighted in red. ACP: Acyl carrier protein. (A) MonCI from *Streptomyces cinnamonensis*. (B) Lsd18 from *Streptomyces lasaliesis*. (C) Bmp2 from *Marinomonas mediterranea* (PDB ID: 5BVA). (D) Human squalene epoxidase (PDB ID: 6C6N). (E) *p*-Hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* (PDB ID: 1PBE).

Table 3.3: Sequence alignment of epoxidases involved in the polyether biosynthesis. MonCI (Streptomyces cinnamonensis), NaO (Streptomyces nanchangesis), TmnC (Streptomyces sp. NRRL 11266), SalC (Streptomyces albus subsp. albus), Lsd18 (Streptomyces lasaliensis), Bmp2 (Pseudoalteromonas sp. PS5), PHBH (Pseudomonas fluorescens), and SQLE (Homo sapiens). The proteins are listed in decreasing order of sequence identity to MonCI: NanO 59.8%, SalC 52.6%, TmnC 49.5%, Lsd18 47.8%, Bmp2 14.8%, PHBH 14.8%, SQLE 14.1%. The two conserved sequences (Residues 47-64 and 327-340 in MonCI), which are found only in polyether producing monooxygenases, are highlighted in red and yellow. The conserved GXGXXG FAD-binding region is highlighted in green and indicated by the green asterisk marks.

```
MonCI       ------------------------------------------------------------          0
NanO        ------------------------------------------------------------          0
SalC        ------------------------------------------------------------          0
TmnC        ------------------------------------------------------------          0
Lsd18       ------------------------------------------------------------          0
Bmp2        ------------------------------------------------------------          0
PHBH        ------------------------------------------------------------          0
SQLE        MWTFLGIATFTYFYKKFGDFITLANREVLLCVLVFLSLGLVLSYRCRHRNGGLLGRQQSG          60


MonCI       ------------------------------------------------------------MTT          3
NanO        ------------------------------------------------------------M          1
SalC        ----------------------------------------------------MPAADGRVTHLM         12
TmnC        ------------------------------------------------------MAEATRGP          8
Lsd18       ------------------------------------------------------------M          1
Bmp2        ----------------------------------------------------------GSHMNG         6
PHBH        ------------------------------------------------------------          0
SQLE        SQFALFSDILSGLPFIGFFWAKSPPESENKEQLEARRRRKGTNISETSLIGTAACTSTSS         120

                    ! ! !
MonCI       TRPAHAVVLGASMAGTLAAHVLARHVDAVTVVERDA-LPEEP--QHRKGVPQARHAHLLW          60
NanO        TTPTRAVVLGGGWAGMLTAHVLARHLESVTVVERDI-LPDGP--HHRKGQPQARHVHVLW          58
SalC        TGETHAVVLGGGLTGMLTSAVLARHLDKVTVIERDV-LPEGP--ELRKGVPQARHAHLLW          69
TmnC        NTRVHGVVLGGGLAGVLAARALRDHVDHVTVVERDT-YPDLT--EPRKGVPQGRHAHILW          65
Lsd18       TNTRSAVVLGGGMAGMLVSSMLARHVGSVTVIDRDA-FPAGP--DLRKGVPQARHAHILW          58
Bmp2        FTHYDVVIIGSGPAGSLCGIECRKKGLSVLCIEKE-QFPRFHIGESLTG----NAGQIIR          61
PHBH        -MKTQVAIIGAGPSGLLLGQLLHKAGIDNVILERQ--TPDYVLGRIRAGVLEQGMVDLLR          57
SQLE        QNDPEVIIVGAGVLGSALAAVLSRDGRKVTVIERDLKEPDRIVGEFLQP----GGYHVLK         176
                  ::*..  *   .            ::::   *                .::


MonCI       SNGARLIEEMLPGTTDRLLAAGARRL-G--------FPEDLVTLTGQGWQHRFPATQFAL         111
NanO        SSGAGIVENLLPGTAERLLAAGARRI-G--------FQSDLVTLTAWGWQYRFPATAYAM         109
SalC        SGGARAIDSVLPGTVKQLIAEGAHRL-Y--------LPRDVVWLTPHGWQHRFSGSQFMV         120
TmnC        SGGAEAIEELLPGTLDRLRAAGAHRI-G--------VKEDMVLYSAYGWQHRFPGSHYAL         116
Lsd18       SGGARIVEELLPGTTDRLLGAGAHRI-G--------IPDGQVSYTAYGWQHRFPEAQFMI         109
Bmp2        DLG----------LAEDMDAAGFPDKPGVNVIGSLSKNEFFIPILAPTW-QVR-------         103
PHBH        EAG----------VDRRMARDGL-VHEGVEIAFAGQRRRIDLK-------RLSGGK----         95
SQLE        DLG----------LGDTVEGLDAQVVNGYMIHDQESKSEVQIPYPLSENNQVQSGR----         222
            . *            :   .              :


MonCI       VASRPLLDLTVRQQALGADNITVRQRTEAVELTGSGGGSGGRVTG--VVVRDLDS--GRQ         167
NanO        MCTRPLLDWVVRDAILAGGRIEVEHGTEAVELAGD----RSRVTG--VRVRDAGG--GEP         161
SalC        TCSRALLDWVVRRQALAEPKITVRQETEVLGLLGG----AGQVTG--VRLRDR-S--GES         171
TmnC        TCSRPLLDRTVREAALDHPDTEVLTRTEAHGLLGD----RTSVTG--VRVRTS-D--GAT         167
Lsd18       ACSRALLDWTVREETLREERIALVEKTEVLALLGD----AGRVTG--VRVRDQES--GEE         161
Bmp2        ---RSDFDD-MIKRKAVEHGVEYKLGMV-TDVIKD----GEKVVG--ALYKADGV--EHQ         150
PHBH        T------------------VTVYGQTEVTRDLMEA----R-EACGATTVYQAAEVRLHDL         132
SQLE        AFHHGRFIMSLRKAAMAEPNAKFIEGVV-LQLLEE----DDVVMG--VQYKDKET--GDI         273
                                       :          . *  .  :


MonCI       EQLEADLVIDATGRGSRL---------------KQWLAALGVPALEEDVVDAGVAYATRL         212
NanO        RLLEADLVVDATGRASRL---------------GHWLAALGLPAVEQDVVDAGIGYATRM         206
SalC        TELAADLVVDAGGRASAL---------------RRWLPELGLDQVEEDVVDSGIAYATRV         216
TmnC        RELPADIVVDATGRGSRL---------------RHWLTDLGLPPAAEESVDTGLTYATRV         212
Lsd18       REVPADLVVDTTGRGSPS---------------KRLLAELGLPAPEEEFVDSGMVYATRL         206
Bmp2        --VRSKVLVDASGQNTFLS------RKGVAGKRQIEFFSQQIASFAH--------YKGVE         194
PHBH        QGERPYVTFERDGERLRLDCDYIAGCDGFHGISRQSIPAERLKVFERVY---PFGWLGLL         189
SQLE        KELHAPLTVVADG---------------LFSKFRKSLVSNKVSVSSH--------FVGFL         310
                  : .   *                        :    :    .        :


MonCI       FKAPPGATTHFPAVNIAADDRVREPGRFGVVYPIE-----------GGRWLATLSCTRGA         261
NanO        FKAPEGADGNFPAVQVAADPLTRQPGRFGVVYPQE-----------GGRWLVTLTSTRGA         255
SalC        FKAPAPVAQGFPMVNIAAAPGLGKPGQNGALVPIE-----------DGKWLVTLAGTRGG         265
TmnC        FRAPAGAPGAFPVVSVYADHRSGEPGRNGLLLPIE-----------DGRWIITLSGTRGG         261
Lsd18       FRAPEAAATNFPLVSVHADHRAGRPGCNAVLMPIE-----------DGRWIVTVSGTRGG         255
Bmp2        -RDLPPFSTN------TTILYSKQYHWS-WIIPISPDTDSLGVVIPKDLYYKECKNPDDA         246
PHBH        -ADTPPVS--------HELIYANHPRGFALCSQRS---------ATRSRYYVQVPLTEKV         231
SQLE        MKNAPQFKAN------HAELILANPSPV-LIYQISSSETRVLVDI-------RGEMPRNL         356
                                  .          .
```

```
MonCI      QLPTHEDEFIPFAE-NLNHPILADLLRDAEPLTPVFGSRSGANRRLYPERLEQWPDGLLV          320
NanO       PLPTDEDEFTGYAK-VLRHSIVSELMSVAEPISPIFQSHSGANRRMYPERMPQWPEGLLI          314
SalC       EPPTDDDSFLDFAR-GLRHPVLADLLERAEPLGPVKGSRSTVNRRLYYDRVANWPDGLLV          324
TmnC       EPTADEERFATFAR-SLRDPIIADLIEAAEPLTPVRTTRSTLNRRMHLDRLADRPEGLVA          320
Lsd18      EPPADDEGFARFARDGVRHPLVGELIAKAQPLTSVERSRSTVNRRLHYDRLATWPEGLVV          315
Bmp2       IAWGM-DHISPEL----KRRF-----KNAERQGD---SQSMADFSYRIEP--FVGDGWMC          291
PHBH       EDWSD-ERFWTELKARLPAEV-----AEKLVTGPSLEKSIAPLRSFVVEP--MQHGRLFL          283
SQLE       REYMV-EKIYPQIPDHLKEPF-----LEATDNSH---LRSMPASFLPPSS--VKKRGVLL          405
                 : :          .                              .        .

MonCI      IGDSLTAFNPIYGHGMSSAARCATTIDREFER------------SVQEGT--GSARAGTR          366
NanO       LGDSLAAFNPVYGHGMSSAARAAEALDKELAR------------D-------GFGEGGTR          355
SalC       LGDALAAFNPVYGHGMSCSALSAKALDAELGR-----------------S--GLAPGMVQ          365
TmnC       LGDCVVSLNPIHGHGMSVAARSARALEACLSR------------A----G--GLKPGLAR          362
Lsd18      LGDAVAAFNPVYGHGMSAAAHSVLALRSQLGQ-----------------R--AFQPGLAR          356
Bmp2       IGDAHRFLDPIFSYGVSFAMKEGIRAAEAIAQVVAGQDWKAPFYAYRDWSNGG-------          344
PHBH       AGDAAHIVPPTGAKGLNLAASDVS------------TLYRLLLKAYREGRGELLERY-SA          330
SQLE       LGDAYNMRHPLTGGGMTVAFKDI-------------KLWRKLLKGIPDLYDDA-------          445
            **.     *  . *:. :

MonCI      ALQKAIGAAVD-DPWILAATKDIDYVNC---------RVSATDPRLIGVDTEQRLRFAEA          416
NanO       QVQRALSEVVD-DPWIMAGLNDIQYVNC---------RNLSSDPRLTGPDVAERLKFSDF          405
SalC       AVLQKVAKVVD-DPWLLATTQDICYPGT---------KVTAQDPRIAP-RGDQEQQFADL          414
TmnC       TAQQAIAAAAD-APWLLSASQDLCYPDN---------KAAVSDPRLTT-QAAQRQGFADM          411
Lsd18      AAQRAIAVAVD-DAWVLATSHDIGYPGC---------RTQTRDPRLTR-HAGERQRVTDL          405
Bmp2       ---QQIAADLIRYFWIYP----------------IFFGYQMQNPDLRDE---V----IRL          378
PHBH       ICLRRIWKAERFSWWMTSVLHRFPD-------------TDAFSQRIQQT---E---LEYY          371
SQLE       ----AIFEAKKSFYWARKTSHSFVVNILAQALYELFSATDDSLHQLRKA---C--FLYFK          496
                :       *                              :

MonCI      ITAASIRSPKASEIVTDVMSLNAPQAELGSNRFLMAMRADERLPELTAPPFLPEELAVVG          476
NanO       LSGKSIRSPKVCEVTTSVLSLNAPQKALGDSRFLSLLRTDTSHPKLVEPPFHPEELEMVG          465
SalC       LSTAALHDPVVSAAAMQVTALAAPVSSLESPHLVAALRKGKAHEPLTAPPFKDAELAVLD          474
TmnC       VTSASLVNERVCDALTAVTTLTAPLGSLETPEFLAAMRQ-PARPPLTAAPLKDAEAAVLR          470
Lsd18      VGLTATRNQVVNRAAVALNTLSAGMASMQDPAVMAAVRRGPEVPAPTEPPLRPDEVARLV          465
Bmp2       LG-------------------------------------GCCFDCEGWKAPAI--FR          396
PHBH       LGSEAGLAT------IAENYV--GLPYEEIE-----------------------------          394
SQLE       LGGECVAGP------VGLLSVLSPNPLVLIGHFFAVAIYAV-YFCFKSEPWITKPRALLS          549
           :

MonCI      LDAATISPTPTPTPTAAVRS-----          496
NanO       LKPSGIAAKGALG------------          478
SalC       GDAAGAAASA---------------          484
TmnC       AAK----------------------          473
Lsd18      SGAGVTA------------------          472
Bmp2       NAIEEYDRKQ------------MAS          409
PHBH       -------------------------          394
SQLE       SGAVLYKACSVIFPLIYSEMKYMVH          574
```

3.3.2 The crystal structure of Lsd18

Initial crystal hits of Lsd18 came from several conditions like 0.1 M Tris pH8.5, 0.7 M sodium citrate (MCSG4_B6); 0.2 magnesium chloride, 0.1 M Tris pH8.5, 30% (w/v) PEG4000 (MCSG4_F12); 0.2 M sodium chloride, 0.1 M Tris pH8.5, 25% (w/v) PEG3350 (MCSG1_D9). These initial crystals were in small needle clusters or thin plate clusters, which were not good enough to be harvested or diffracted (Figure 3.14 B). Optimizing the crystallization conditions included varying salts, precipitants, buffers, temperature, and protein concentrations. Untagged or no-alkylated Lsd18 proteins were also tried, but none of promising hits was obtained. It turned out that salt (sodium chloride) concentration plays an important role in the crystal shape. When the concentration of sodium chloride increased from 0.1 M to 0.85 M, Lsd18 crystals switched from two-dimension plate crystals to three-dimension rhombic large crystals (Figure 3.14 C and D). It

took around 5 days for Lsd18 crystals to grow up to the full size by hanging-drop evaporation method at 18 ℃. Crystals also exhibit yellow color due to the tightly bound FAD cofactors.

Table 3.4: Data collection and refinement statistics for Lsd18 and it complex crystal structures.

	Lsd18	Lsd18-ligand complex
Data collection		
Space group	P1	P1
Cell dimensions		
a, *b*, *c* (Å)	46.37, 61.81, 75.37	47.04, 61.97, 76.29
α, β, γ (°)	74.63, 81.71, 77.11	74.80, 81.87, 76.80
Resolution (Å)	34.75 – 1.54	32.94 -1.85
R_{sym}(%)	6.005 (61.49)	11.8 (109.6)
$I / \sigma I$	10.7 (2.0)	10.18 (1.79)
$CC_{1/2}$	0.997 (0.781)	0.998 (0.704)
Completeness (%)	95.6 (94.35)	97.6 (96.3)
Redundancy	3.5 (3.6)	
Refinement		
Resolution (Å)	34.75 – 1.54	32.94-1.85
No. reflections	112052	67199
R_{work} / R_{free}	0.163/0.200	0.180/0.230
No. atoms		
Protein	7200	7083
Ligand/ion	108	177
Water	995	491
B-factors		
Protein	22.34	30.13
Ligand/ion	15.92	45.82
Water	33.05	36.48
R.m.s. deviations		
Bond lengths (Å)	0.020	0.020
Bond angles (°)	2.02	1.89

* This structure was determined from one single crystal.
Values in parentheses are for the highest resolution shell.
$R_{sym} = \Sigma | I\text{avg} - I\text{i} | / \Sigma I\text{i}$, where Ii is the observed intensity and *I*avg is the average intensity of observations of symmetry-related reflections.
$R_{work} = \Sigma | F\text{p} - F\text{p(calc.)} | / \Sigma F\text{p}$, where *F*p and *F*p(calc.) are observed and calculated structure factors; R_{free} is calculated with 5% of the data.

Diffraction data collection and refinement statistics of Lsd18 crystal structure are shown in table 3.4, which was determined from a single crystal. The unit cell of this triclinic crystal contains two enzyme monomers. After final refinement, the Ramachandran favored is 98.49%, and the Ramachandran outliers is 0% based on the calculation by Molprobity (190). At the end, the crystal structure of flavin-bound Lsd18 was determined at resolution up to 1.54 Å (Figure

3.19). The phase of the Lsd18 crystal diffraction is solved by molecular replacement, using the MonCI crystal structure as the structural model. After refinement, the crystal structure of flavin-bound Lsd18 was determined at resolution up to 1.54 Å with the R_{work}/R_{free} as 0.163/0.200. The crystal structure includes 7200 protein atoms, 106 ligand atoms, 2 chloride ions, and 995 water molecules. We are able to trace the entire chain A (residues 20-487) and chain B (residues 20-487). There are several residues exhibiting alternative conformations for their sidechains, which include Arg21, Arg91, Ile102, Ser108, Asn232, Arg315, Arg376, Thr401, Val437, and Met455 in chain A. Also based on the electron density, the residues Lys299 in both chain A and B have been ethylated. This lysine is at the surface of the Lsd18 protein, where is easier to get ethylated compared to the lysine residues inside the protein.

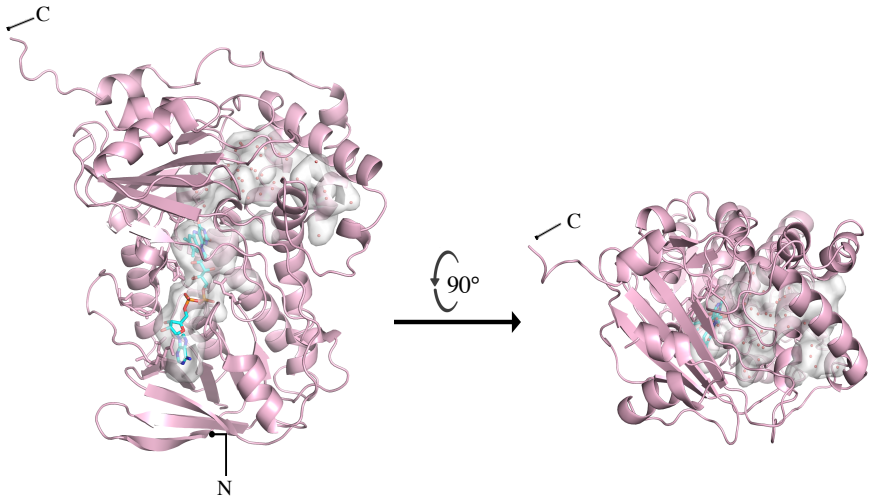


Figure 3.19: 1.46 Å resolution X-ray crystal structure of Lsd18. Chloride ion and water are shown as green and red spheres, respectively. The FAD cofactor is drawn as a stick model. The fused FAD- and substrate-binding pocket is shown as a grey surface. For clarity, only those water molecules located inside the pocket are shown.

Similar to MonCI structure, Lsd18 also has two difference entrances where its substrate and solvent molecules can enter and exist. Both of these two entrance locate at the side of the Lsd18 surface (Figure 3.20). The side entrance I is about 9.1 Å in width. The other entrance, the side entrance II is about 6.7 Å, locating at the other side of the Lsd18 surface. The side entrance II has more hydrophilic and positively charged residues, while the side entrance I is more

hydrophobic (Figure 3.20 B). Considering their hydrophobicity, we assume that the side entrance II of Lsd18 functions as a solvent channel while the side entrance I functions as a substrate-binding channel. This hypothesis agrees well with the water distribution in the substrate-binding pocket (Figure 3.20 A).

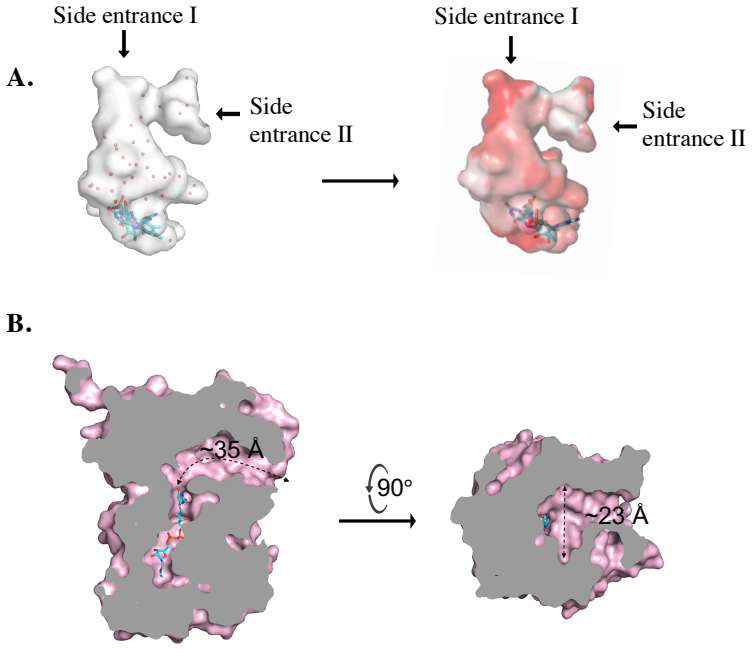


Figure 3.20: The fused FAD- and substrate-binding pocket of Lsd18. (A, B) The two side entrances that provide access to Lsd18's active site cavity. The water molecules are show in the red spheres. Hydrophobic surfaces are colored in red and hydrophilic surfaces are colored in white. (C, D) Cross-section of Lsd18 showing the dimensions of the substrate-binding pocket. The indicated pocket length (~35 Å) is that of the pocket trajectory leading to the side entrance.

The substrate-binding pocket of Lsd18 locates above the flavin isoalloxazine ring, consisting of an eight-stranded β-sheet and two separate three-helix bundles (Figure 3.18 B). It is much longer than the substrate pockets of (PDB: 1BKW) and Bmp2 halogenase (PDB: 5BVA), providing enough space for the substrate prelasalocid. Near the entrance of the substrate-binding pocket, few positive-charged residues like arginine form a surface-exposed basic path. This patch is complementary to the negative-charged surface of the acyl carrier proteins (ACPs), which indicates that the ACP-bound native substrate is presented to Lsd18 by protein-protein interactions.

According to the enzyme assays, Lsd18 was shown to function as both epoxidase and flavin reductase. Its crystal structure also reveals that Lsd18 consists of a fused pocket for both the cofactor FAD and the substrate.

The FAD-binding domain is characteristic of flavin-dependent monooxygenase family that carries a conserved motif GxGxxG and a Rossman dinucleotide-binding fold with central five stranded parallel β-sheet. This central β-sheet locates between a three-stranded antiparallel β-sheet and a three-helix bundle, forming a three-layer ββα substructure. The flavin phosphate makes hydrogen-bonding contacts with the conserved motif GxGxxG (GGGMAG in Lsd18) (Table 3.3). The adenosine diphosphate ribosyl binds at the C termini of the central sheet by hydrogen bonds. The adenine base is stacked with the guanidine group of residue Arg51 by π-cation interaction. As previously observed on PHBH (107, 108, 195, 196), the isoalloxazine ring of FAD moves between the "in" and "out" positions during its catalysis for FAD oxidation/reduction and the substrate translocation. In the Lsd18 crystal structure, FAD isoalloxazine ring is π-π stacked with the imidazole rings of His71 (Figure 3.21 A).

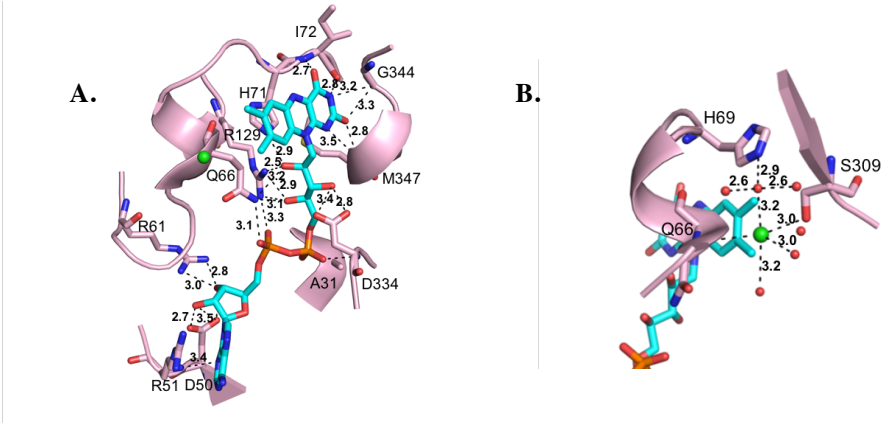


Figure 3.21: Zoomed view of cofactor FAD and chloride in Lsd18. (A) Protein environment of the FAD cofactor in Lsd18. Interaction distances are given in Å. (B) Protein environment of the chloride ion in Lsd18. Interaction distances are given in Å.

Similar to MonCI, one chloride ion was modelled next to the isoalloxazine ring of FAD in the Lsd18 crystal structure (Figure 3.21 B). The chloride coordinates with five hydrogen bond donors (Gln66 main chain nitrogen, Ser309 side chain hydroxyl group, and three water molecules) in a distorted trigonal bipyramidal stereochemistry (Figure 3.21 B). Structural alignment of the Lsd18 crystal structure and the NADPH-bound 4-hydroxybenzoate hydroxylase (PDB ID: 1K0J) shows that the chloride in Lsd18 is located where the adenine based of NADPH is expected to bind.(108) Therefore, chloride could be a competitive inhibitor for Lsd18. In agreement with this theory, chloride was previously shown to competitively inhibit the flavoenzyme *p*-hydroxybenzoate hydroxylase, a structural homolog of Lsd18, but the inhibition mechanism was not determined.(181)

In structure, the most similar structures to Lsd18 in PDB are flavin-dependent Bmp2 halogenase (PDB: 5BVA; 2.9 Å RMSD for Cα atoms; Z-score = 29.3) and *p*-hydroxybenzoate hydroxylase (PHBH; PDB: 1PBE; 3.7 Å RMSD for Cα atoms; Z-score = 29.1) (Table 3.3 and Figure 3.18). They exhibit main similarity in the FAD-binding region while the substrate-binding regions are more distinct. Among all the epoxidases available in the Protein Data Bank (PDB), nearly all of them is cation- (Zn^{+}/Fe^{2+}-) dependent.(197-200) Only styrene monooxygenase and human squalene monooxygenases are flavin-dependent. According to their functions and structures, Lsd18 and human squalene monooxygenase belong to the family of group A flavin-dependent monooxygenases that are encoded by one single gene. The oxygenase StyA and StyB are group E flavin-dependent monooxygenases that consist of a two-enzyme system. Unlike Lsd18, styrene monooxygenase has two separate components: oxygenase StyA and flavin reductase StyB, which are encoded by two individual genes.(138, 201) Structural alignment showed distinct difference between Lsd18 and styrene monooxygenases including the StyA structure (PDB: 3IHM) and StyB structures (PDB: 1USC, 1USF, 4F07). The Lsd18 and human squalene monooxygenase (PDB: 6C6N; 3.7 Å RMSD for Cα atoms; Z-score = 29.1). Both of these two enzymes are flavin-dependent epoxidases with the single-component enzyme system. They show a high structural similarity in the FAD-binding domain. However, their substrate-binding

domains are more different in structure due to the difference in their natural substrates (Figure 3.18).

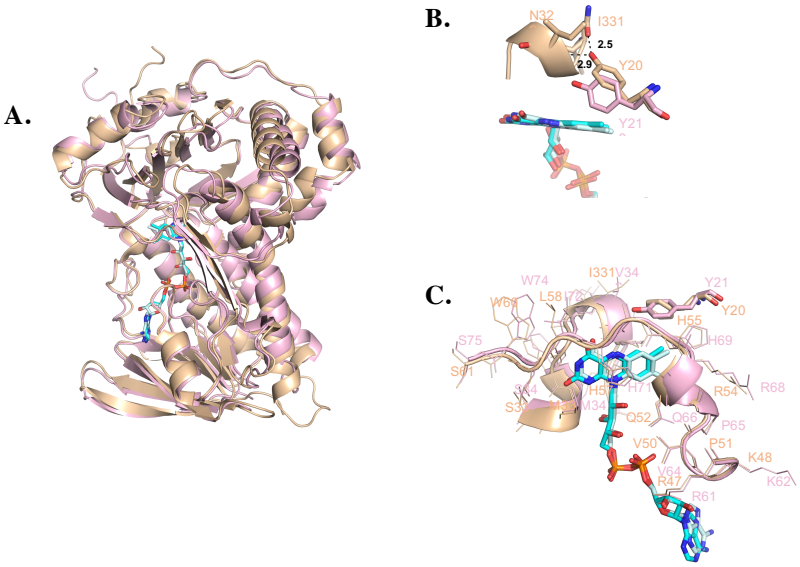


Figure 3.22: Structural comparison between MonCI and Lsd18. (A) Overall structural comparison. (B) Zoomed-view of the MonCI Tyr208 and Lsd18 Tyr218. The hydrogen-bond length of Tyr208 to Asn329 and Ile331 in MonCI is labeled. (C) The comparison of the conserved residues and the regions in MonCI and Lsd18. MonCI is colored in wheat; Lsd18 is colored in pink; the MonCI FAD is colored in cyan; the Lsd18 FAD is colored in pale cyan.

Lsd18 and MonCI share high similarity in both sequence and structure. The sequence identity is 47.8%; the structural RMSD is 1.63 (Table 3.3 and Figure 3.22 A). Their secondary structures are basically the same. Both of them have a fused pocket for FAD- and substrate-binding. The FAD-binding domain is very similar in Lsd18 and MonCI structures. Their cofactor FAD locates at the same position inside the Lsd18 and MonCI, staying in the same oxidation conformation. Both have a chloride ion near the FAD isoalloxazine ring, which could play an inhibition role during catalysis. The pocket volume of Lsd18 is 889.441 $Å^3$, a little bite smaller than the pocket volume of MonCI, 995 $Å^3$. This pocket volume difference could be related to the different sizes of their native substrates prelasalocid and premonensin. The prelasalocid is approximately 24 Å long and 4.5 Å when fully extended, while the premonensin is about 30 Å long and 5 Å wide. Their pocket entrances are also slightly different. In Lsd18, there are two side

entrances. The side entrance II (the solvent entrance) leads to the side entrance I (the substrate entrance). The side entrance I is connected directly with the active cavity. In MonCI, there are two entrances, one is at the top (solvent entrance), the other is at the side (substrate entrance). Compared to the Lsd18 pocket entrance, the top entrance of MonCI is separated further away from its side entrance, and it is connected to active site cavity by a narrow, 8.5 Å-long tunnel, while its top entrance immediately leads to the active cavity. The active cavity of Lsd18 is relatively narrower than the active cavity of MonCI. Even though their secondary structures are very similar, but the residue components are different. We have noticed that the residues at their active sites are very conservative. Both of their active sites are restrained by a tyrosine residue, a short helix region, and a loop region. The active site in Lsd18 contains Tyr218, a helix region Ala338-Ser348 (AFNPVYGHGMS) at the *re* face of FAD, and a loop region Arg61-Ser75 (RKGVPQARHAHILWS) at the *si* face of FAD. The active site in MonCI contains Tyr208, a helix region Ala327-Ser337 (AFNPIYGHGMS) at the *re* face of FAD, and a loop region Arg47-Ser61 (RKGVPQARHAHLLWS) at the *si* face of FAD (Figure 3.22 C). Even though the residues Ile72 and Val342 in Lsd18 are not exactly the same as the residues Leu58, Ile331 in MonCI, they are all non-polar residues with the side chain in the same orientations. For the tyrosine residue, the side chain of Tyr218 in Lsd18 stays closer to the FAD, making the active site relatively smaller, while the side chain of Tyr208 in MonCI is further from the FAD and forms the hydrogen bonds with the Ile331 backbone and Asn329 side-chain carbonyl group, making the active site more spacious for its native substrate with a bulky terminus (Figure 3.22 B). Generally, the active sites of Lsd18 and MonCI are significantly similar to each other. The residues surrounding the active site are also very conservative among other epoxidases involved in the polyether biosynthesis such as NanO (*Streptomyces nanchangesis*), TmnC (*Streptomyces sp*. NRRL 11266), and SalC (*Streptomyces albus* subsp. albus), which means that their epoxidation mechanism could be uniformed (Table 3.3).

3.3.3 The co-crystal structure of Lsd18-ligand complex

To achieve the Lsd18-ligand structure, various conditions and methods have been tried. We tried both soaking and co-crystallization methods for the Lsd18-ligand complex crystallization. The soaking method is to crystallize the ligand-free Lsd18 crystal first, then followed by addition of the ligands in the crystallization solution. During the soaking process, the ligand will enter inside the protein crystal and interact with the proteins. The co-crystallization method is to incubate the Lsd18 protein with the ligand first in solution, then continued by setting the crystallization trays with the Lsd18-ligand mixture. In this method, the ligand first interacts and bind with the protein in solution, then the protein-ligand complex forms crystals together. We tried both of these methods in various conditions. For soaking method, time of adding the ligand, ligand concentration, soaking time, and soaking temperature were explored. The crystal cross-linking by using glutarahyde was also carried out to stabilize and strength the Lsd18 crystals. For co-crystallization method, the protein-to-ligand molar ratio, incubation temperature, incubation time, and addition of cofactors FAD and NADH/NADPH were explored in the past. We tried all three different substrate analogs, linalool, compound **24**, and compound **26** as the ligands in both soaking and co-crystallization method. Same as the ligand-free Lsd18 crystallization, the Lsd18 protein sample needs to get reductive alkylation to achieve diffraction-quality crystals. We screened many crystals from different conditions. We got the best ligand density from a co-crystallization condition. The 5.1 mg ml^{-1} Lsd18 protein sample in the buffer containing 20 mM Bicine pH8.5 and 100 mM sodium chloride was incubated with 75x molar excess substrate analog **26** (the stock solution is 0.1 M dissolved in methanol) at 4 ℃ for one hour before setting up the crystallization trays. The crystallization solution is 0.1 M imidazole pH8.0, 0.8 M sodium chloride, and 37% PEG4000. The 0.2 μl mixture of the Lsd18 and the substrate analog **26** was mixed with the 0.2 μl crystallization solution in the sitting-drop plate with 50 μl well solution. The co-crystals were harvested around 3 months after the crystallization plate setting up. 20% (v/v) glycerol was used as the cryoprotectant. The co-crystals were diffracted at SSRL with the help of the beamline scientist Irimpan I. Mathews.

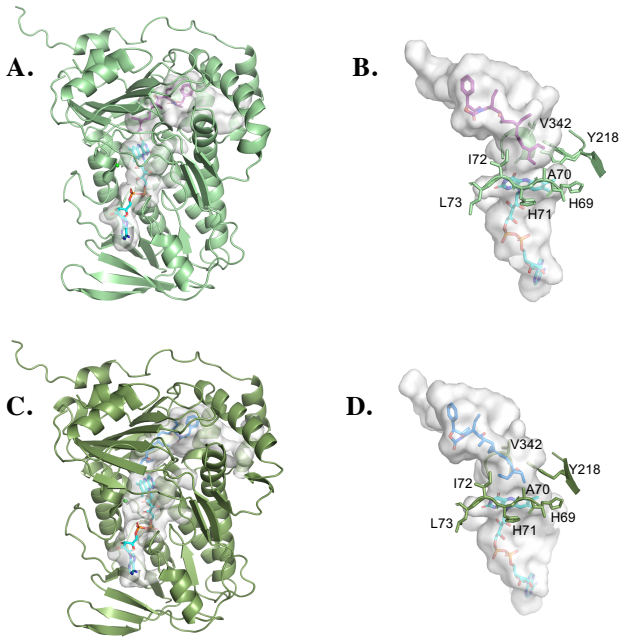


Figure 3.23: 1.80 Å resolution X-ray crystal structure of Lsd18-ligand complex structure. (A, B) The chain A structure with the substrate 26 in the pocket. The chain A is colored in pale green. The substrate 26 is shown in purple sticks. (C, D) The chain B structure with the product 27 in the pocket. The chain B is shown in smudge color. The product 27 is shown in marine color. Chloride ion is shown as the green sphere. The FAD cofactor is drawn as a stick model. The fused FAD- and substrate-binding pocket is shown as a grey surface.

Diffraction data collection and refinement statistics of Lsd18-ligand complex crystal structure are shown in table 3.4, which was determined from a single crystal. The unit cell of this triclinic crystal contains two enzyme monomers. After final refinement, the Ramachandran favored is 97.95%, the Ramachandra allowed is 1.48%, and the Ramachandran outliers is 0.22% based on the calculation by Phenix table_one program (147). At the end, the crystal structure of the Lsd18-ligand complex was determined at resolution up to 1.85 Å (Figure 3.23). The phase of the Lsd18 crystal diffraction is solved by molecular replacement, using the ligand-free Lsd18 crystal structure as the structural model. After refinement, the crystal structure of the Lsd18-ligand complex was determined at resolution up to 1.85 Å with the R_{work}/R_{free} as 0.176/0.226. The difference between R_{work} and R_{free} values seem very distinct, which means that there could be over-fitting issues with the protein refined structure. Further modeling and refinement are still in process

to achieve a better and more reliable Lsd18-ligand complex structure. The current co-crystal structure includes 7083 protein atoms, 175 ligand atoms, 2 chloride ions, and 491 water molecules. We are able to trace the entire chain A (residues 20-487) (Figure 3.23 A and B) and chain B (residues 20-487) (Figure 3.23 C and D). Based on the electron density, we also fit the substrate analog **26** in the substrate-binding pocket of chain A, and the intermediate product **27** in the substrate-binding pocket of chain B. There are several residues exhibiting alternative conformations for their sidechains, which include Tyr218 in chain A and Arg224 in chain B. The extra electron density on the terminus of the side chain also indicated that Lys299 residues in both chain A and B have been ethylated. This lysine is at the surface of the Lsd18 protein, where is easier to get ethylated compared to the lysine residues inside the protein. The overall structure of the Lsd18-ligand complex is very similar to the ligand-free Lsd18 structure. All the structures were aligned by the online server TM-align.(202)

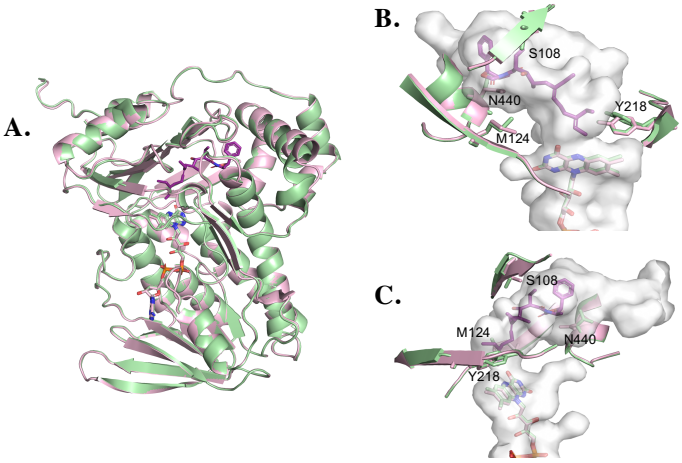


Figure 3.24: Structural comparison of the chain A structure of the native Lsd18 and the Lsd18-ligand complex crystal structures. (A) Overall structure comparison of Lsd18 chain A. (B, C) Zoomed view of the substrate-binding pocket residues that exhibit different conformations in the chain A of Lsd18 and Lsd18-ligand complex structures.

For Lsd18 chain A, the ligand-free and the ligand-bound complex structures were aligned with 467 residues (RMSD= 0.38) (Figure 3.24). The oxidized FAD cofactors are basically at the same position. Same as the chloride ion near the FAD. Their backbones are mainly the same, but some loop regions, especially the loops at the substrate-binding region, do not match very well.

For example, loop regions at the entrance of the substrate-binding pocket, residues 241-249 and residues 405-412, moved outside a little bit in the complex structure. The loop of the residues 241-249 is at the top entrance while the loop of the residues 405-411 is at the side entrance or the substrate-binding entrance. The helix region, residues 411-423 that is a part of the substrate-entrance and the substrate-binding pocket, also moved its position in the Lsd18-ligand complex structure. The backbone movement in the Lsd18-ligand complex structure expanded the space of the substrate-binding pocket, which made it easier for the substrate and enter and interact with the protein. The *N*-terminal loop in the Lsd18-ligand complex structure also moved a little bit compared to the one in the ligand-free Lsd18 structure, which is probably due to the loop flexibility at the terminus. Besides the movement of the backbone, the side chains of several residues also moved. Some of the movement was probably due to the flexibility on the protein surface, including Arg21, Lys62, Arg68, Glu139, Arg142, Arg163, Arg168, Glu173, Glu201, Arg246, Asp258, Glu278, Glu295, Ser305, Arg321, Gln367, Arg376, Arg404, Arg410, and Arg473. Unlike the residues inside the protein, the side chains on the protein surface is not restrained by any interactions, therefore they are more flexible and exhibit multiple conformations. Interestingly, some side chains of the residues along the substrate-binding pocket also moved their conformation. First is the residues near the substrate entrance changed their conformation in the Lsd18-ligand complex structure, which includes Glu414, Arg417, Leu421, Met446, Gln450, and Met455. These residues tend to change their conformations to make the substrate entrance wider in the Lsd18-ligand complex structure. Secondly, there are some residues in the substrate-binding pocket that moved their side chains. The residues Met124 and Asn440 moved their side chains from facing inside to facing outside of the substrate-binding pocket. When there is no ligand in the substrate-binding pocket, the side chains of Met124 and Asn440 are facing towards the active site, which would help guide the substrate slide into the pocket and move forward to the active site. Once the substrate enters, the side chains of Met124 and Asn440 change their conformation and face towards the entrance of the pocket, in which way it would hold the substrate inside the pocket and prevent the substrate slide out before the reaction (Figure 3.24 B and C). There is also another

residue, Ser108 in the pocket changing its conformation due to the formation of a weak hydrogen bond with the substrate hydroxy bond (Figure 3.24 B and C). This hydrogen bond also helps hold the substrate inside the pocket.

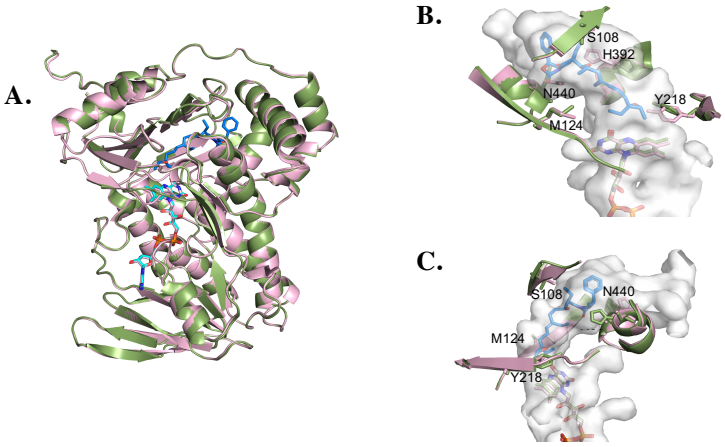


Figure 3.25: Structural comparison of the chain B structure of the native Lsd18 and the Lsd18-ligand complex crystal structures. (A) Overall structure comparison of Lsd18 chain B. (B, C) Zoomed view of the substrate-binding pocket residues that exhibit different conformations in the chain B of Lsd18 and Lsd18-ligand complex structures.

For Lsd18 chain B, the ligand-free and the ligand-bound complex structures were aligned with 467 residues (RMSD= 0.41) (Figure 3.25) The oxidized FAD cofactors are nearly at the same position. Same as the chloride ion near the FAD. The chain B backbone of the Lsd18-ligand complex structure is very similar to the one of the ligand-free Lsd18 structure. The main difference of the backbone shows at the pocket entrances. The top entrance or the solvent entrance region, including the residues 240-249 and 391-421, became smaller in the Lsd18-ligand complex structure, while the side entrance or the substrate entrance, including the residues 406-425 and 447-455, expanded wider in the Lsd18-ligand complex structure. This backbone movement is likely due to the interaction with the ligand inside the pocket. Also, there are some side chains of the residues changing their conformation in the Lsd18-ligand complex structure. Same as chain A, some residues on the protein surface changed their conformations due to their flexibility in solution. These residues include Arg21, Cys62, Trp134, Arg142, Glu150, Lys151, Glu173, Glu201, Arg285, Gln301, Ser305, Arg310, Tyr343, Gln367, Arg376, Arg400, Arg404, Arg434,

Arg460, and Glu469. There are also some residues at the substrate entrance changing their conformations, which includes Arg417, Met446, and Met455. Their side chains moved out and made the substrate entrance wider. Same for the residues Met124 and Asn440, their side chains moved towards the substrate entrance in the Lsd18-ligand complex structure in the purpose of holding the substrate in the pocket. The side chain of Ser108 also change its conformation like the one in chain A, but there are no more interactions between Ser108 and the hydroxyl group of the ligand because the ligand moved away after the reaction. However, two more residues Tyr218 and His392 in chain B attracted our attention. Tyr218 displayed in two alternative conformations in the crystal structure. One conformation is the same as the one in chain A which is more parallel to the FAD isoalloxazine ring, pointing towards the ligand. The other conformation is the side chain moving further away from the FAD and the active site. Considering that the Tyr218 is very close to the ligand we fit at the active site, it is likely that the Tyr218 plays an important role for Lsd18 functions. It forms a hydrogen bond with the epoxide we fit at the active site in chain B, in which case, the epoxide product may be leaded out of the active site by the conformation change of Tyr218 (Figure 3.25 B and C). Also, this tyrosine residue is very conservative in all other epoxidases involved in the polyether biosynthesis. To further explore its role in catalysis, we have mutated Lsd18 Tyr218 into a phenylalanine or an alanine. The Lsd18 mutant Y218A was not able to be expressed and folded very well, which means that changing Tyr218 into an alanine may affect the protein folding process. The Lsd18 mutant Y218F was expressed and purified successfully. According to the enzyme assay, the yield of the product **25** was nearly the same as the wildtype, which means that the Lsd18 catalytic efficiency was not affect by mutation. It is likely that Tyr218 may not be essential for epoxidation, but it could still play a role in the stereoselectivity of the epoxide product since the hydroxy group of the Tyr218 side chain is able to form the hydrogen bond with the epoxide group. Further studies of this tyrosine's role on the stereoselectivity of epoxidation need to be done in the future. The other residue that changed its conformation in chain B is His392 that located near the pocket entrance. When there is no ligand in the pocket, the side chain of His392 faces towards the side entrance. At this conformation, it

narrows the top entrance and promote the substrate to enter from the side entrance. When there is the ligand in the pocket, the side chain of His392 flips it conformation and widens the substrate-binding pocket (Figure 3.25 B and C). Considering that the epoxide product is the pocket of chain B, the conformation change of His392 could help the ligand exist out of the pocket after epoxidation.

Chapter 4: Discussion

4.1 PROTEIN PROPERTIES

4.1.1 FAD reductase, Fre

Fre is the NADH-dependent flavin reductase from *E.*coli (strain K12). It is a monomer of 26.2 kDa molecular weight that catalyzes the reduction of flavins by using NADH or NADPH. Its gene sequence was identified by Reichard et al in 1992.(203) Its crystal structure was solved by Eklund et al in 1999 (204), which shares a high structural similarity to the structure of the ferredoxin reductases of flavoproteins even though their sequence similarity is very low. Interestingly, Fre does not contain a binding site for the AMP moiety of FAD, therefore it only interacts with the FAD substrate by a hydrophobic isoalloxazine binding site. This binding site is consist of a serine and a threonine, and it forms hydrogen bonds with the riboflavin isoalloxazine in a substrate-binding complex. It means that the recognition of the flavin substrate mainly takes place on the isoalloxazine ring. The contribution from the binding of the flavin ribityl side chain is very limited. Fre is able to use both NADH and NADPH as the electron donor and receives various flavin analogues as the electron acceptors. According to Oikawa's results, it showed that the existence of Fre can help increase the Lsd18 efficiency. Even though Lsd18 is capable to reduce FAD by itself, the general efficiency is low. So, in our enzyme assays, Fre was included as a co-enzyme for Lsd18. We have successfully expressed and purified the Fre protein sample. The UV spectra indicated that the purified Fre sample was active and was able to keep oxidizing NADH and reducing FAD. The GC-MS analysis showed that the efficiency of Lsd18 was higher and the yield of epoxide product was also higher by addition of Fre, which was probably due to the higher FAD reduction process catalyzed by Fre.

4.1.2 Glucose 1-hydorgenase, GDH

GDH is the glucose 1-dehydrogenase from *Bacillus subtilis* (strain 168), which belongs to the family of the short-chain dehydrogenases/reductases. Its gene was first discovered by Rudikoff et al in 1986.(205) It functions as a developmentally regulated enzyme of *Bacillus subtilis*. It is

responsible for the oxidation of glucose to gluconolactone together with the concomitant reduction of NADP to NADPH. GDH is involved in the sporulation biological process and gets induced at the stage III of sporulation in *Bacillus subtilis*. It exists as a homotetramer with 28 kDa molecular weight of the monomer. The inspiration of adding GDH in the enzyme assay of Lsd18 was from the study of modular polyketide synthases ketoreductases by Komsoukaniants.(206) In their work, they employed GDH in the enzyme assays as a way to regenerate NADPH, which enabled the polyketide synthase enzymologists to perform in vitro reactions in a more affordable way. The reaction was also under the control and generated significant amount of the polyketide products. Since for the enzyme assays carried out by flavin-dependent epoxidases, it also requires NAD(P)H as the electron donor for the FAD reduction. NAD(P)H itself is not very stable in solution. It can get oxidized very fast. Therefore, GDH was purified and added into the enzyme assays in order to regenerate NADPH continuously for the epoxidases. We have successfully expressed and purified the active GDH enzymes. The UV spectra showed that the purified GDH protein sample was active and was able to produce NADPH for at least about 10 minutes. However, the GC-MS analysis showed that addition of GDH did not increase the product yield of the Lsd18 enzyme assays. It could due to the fact that Lsd18 has already reached to its highest efficiency and occupancy, and the amount of the cofactor NADPH was already enough with the help of GDH. So, for the following enzyme assays, GDH was not included in the enzyme assays since it did not help increase the product yield.

4.1.3 Lsd19

Lsd19 (or LasB) is the epoxide hydrolase involved in the biosynthesis of polyether lasalocid in the *Streptomyces lasaliensis* bacterial strain. It was identified by both Leadlay's groups in 2008 (96) and Oikawa's group in 2009 (97). It catalyzes two consecutive epoxide-opening cyclization reactions after the epoxidation reaction catalyzed by Lsd18. The crystal structures of Lsd19 contain an amino- and a carbonyl-terminal domains, Lsd19A and Lsd19B connected together by a short loop.(113) These two domains have identical backbone conformations. The

Lsd19A domain is responsible for the 5-*exo* cyclization reaction while Lsd19B is responsible for the 6-*endo* cyclization. To form the ether rings in lasalocid, Lsd18 first transforms the linear polyene intermediate, prelasalocid into a diepoxide by two rounds of epoxidation reactions. Then Lsd19 catalyzes the diepoxide into a tetrahydrofuran-tetrahydropyran (THF-THP) product which is lasalocid A. According to the enzyme assays carried out by Oikawa's group (103), the addition of Lsd19 helped increase Lsd18 efficiency. We also tried to add Lsd19 in the enzyme assays of Lsd18. Based on the GC-MS analysis, the product yield did not increase significantly by addition of Lsd19. So, in the following enzyme assays, Lsd19 was not included in the mixture. However, Lsd19 could be used to differentiate the stereochemistry of the epoxide product in the future research due to its highly regioselective manner. When the Lsd19 substrate is (22*R*,23*R*)-diepoxide, the cyclization product was THF-THP. When the Lsd19 substrate is (22*S*,23*S*)-diepoxide, the cyclization product was THF-THF (Illustration 4.1).(112) Therefore, we could use Lsd19 to different the stereochemistry of the diepoxide. It has been revealed that Lsd18 catalyzes the stereoselective epoxidation, transforming the compound into a (22*R*,23*R*)-diepoxide. So, the final product is THF-THP by adding Lsd19 in the enzyme assays. If the Lsd18 mutants exhibit the reverted stereoselectivity and product the (22*S*,23*S*)-diepoxide, the final product catalyzed by Lsd19 will be THF-THF. Therefore, once the THF-THP and THF-THF peaks are identified by LC-MS, we will be able to tell the stereoselectivity of the Lsd18 and its mutants. Furthermore, we will be able to determine which residues play an essential role in the stereoselectivity of Lsd18.

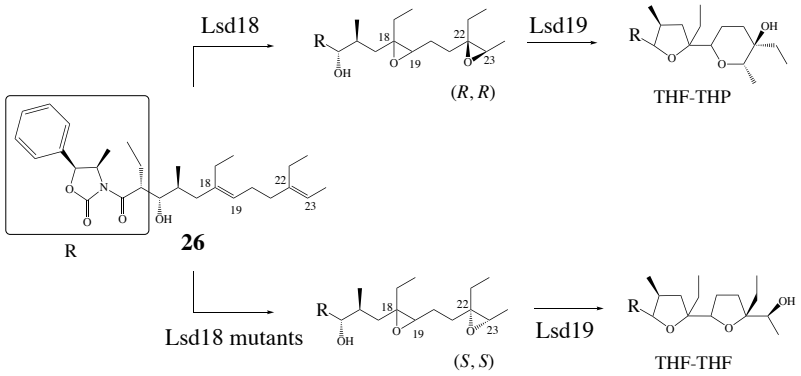


Illustration 4.1: Reactions catalyzed by Lsd18 and Lsd19 by using the compound 26 as the substrate.

4.1.4 MonACPX

MonACPX is a discrete acryl carrier protein that is involved in the biosynthesis of monensin in the *Streptomyces cinnamonensis* bacterial strain.(85) It is a relatively small protein that is about 11.26 kDa. They normally carry negative charges and contain helix-bundles with high-degree sequence and structural similarities among themselves. The previous study has revealed its role in binding the full-length linear polyketide chain and transferring it from the polyketide synthases to other individual enzymes such as MonCI, MonBI/MonBII, and MonE for oxidation cyclization and other modification reactions.(141) All the acryl carrier proteins are expressed in the inactive *apo* form first, then transformed into the active *holo* form by attaching a 4'-phosphopantetheine moiety to their conservative serine residue. For MonACPX, the conservative serine is Ser60 (Table 4.1). The attached 4'-phosphopantetheine moiety can be covalently linked with the polyether intermediates and deliver the intermediates to the active sites of different enzymes involved in the polyether biosynthesis. In our experiments, the MonACPX was expressed in the *E.coli* BAP1 bacterial strain that is genetically modified with the gene of 4'-phosphopantetheinyl transferase Sfp. Therefore, MonACPX was expressed and got immediately transformed into its *holo* form by Sfp inside bacteria. Due to the lack of aromatic amino acids, it is more difficult to purify the MonACPX samples. Also due to the aggregation issue, the final MonACPX yield was relatively low. We tried to include the MonACPX in the MonCI enzyme assays, but it did not help improve the MonCI activity. We also tried to study the binding affinity

between MonACPX and MonCI by surface plasmon resonance method, but there was no clear binding detected. We also incubated MonACPX and MonCI protein samples together and set up the crystallization trays with different commercial screens. However, none of co-crystals appeared. To study the protein-protein interactions between MonACPX and MonCI and investigate how the polyene substrate is delivered to MonCI active site by MonACPX, we conducted the modeling and docking experiments. Since all the acryl carrier proteins share significantly high similarity in their sequences and structures, we first generated the MonACPX model by using its homolog ACP domain from MLSA1 module 5 as the template. The 4-phosphopantetheine moiety was also built on Ser60 in the MonACPX model by Coot. According to the previous studies and co-crystal structures (207-212), the acryl carrier proteins interact with other enzymes primarily through the electrostatic attraction. The negative charged surface of the acryl carrier proteins can bind with the positively charged patch of other enzymes in different binding orientations. The same principles are also applied to the interactions between MonACPX and MonCI. More details are discussed in the MonCI section 4.1.5.

Table 4.1: Sequence alignment of the acyl carrier proteins. It includes the acyl carrier proteins from Streptomyces cinnamonensis (MonACPX, Uniprot ID: Q846Y3), Streptomyces erythraeus (Uniprot ID: Q03131), Escherichia coli (strain K12, Uniprot ID: P0A6A8), Bacillus subtilis (strain 168, Uniprot ID: P80643), Aquifex aeolicus (strain VF5, Uniprot: O67611), and Azospirillum brasilense (Uniprot: P94123). The conserved serine was highlighted in red.

```
S.cinnamonesis EEREKLLTDTIRTQA----GTLLNTTLSDDSNFLENGLNSLTALELTKTLMTLTGMEIAM 56
S.erythraeus   ----EALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAATGVRLPT 56
E.coli         ----STIEERVKKIIGEQLGVKQEEVTNNASFVEDLGADSLDTVELVMALEEEFDTEIPD 56
B.subtilis     ----ADTLERVTKIIVDRLGVDEADVKLEASFKEDLGADSLDVVELVMELEDEFDMEISD 56
A.acelicus     ----MSLEERVKEIIAEQLGVEKEKITPEAKFVEDLGADSLDVVELIMAFEEEFGIEIPD 56
A.brasilense   ----SDVAERVKKIVVDHLGVEESKVTENASFIDDLGADSLDTVELVMAFEEEFGCEIPD 56
                       : :        *          .   : * :** .:**   :    . .:

S.cinnamonesis VAIVENPTPAQLAHHLGQEL 76
S.erythraeus   TTVFDHPDVRTLAAHLAAEL 76
E.coli         EEAEKITTVQAAIDYINGHQ 76
B.subtilis     EDAEKIATVGDAVNYIQNQQ 76
A.acelicus     EDAEKIQTVGDVINYLKEKV 76
A.brasilense   DAAEKILTVKDAIDFIKANA 76
                   .         .:  .
```

4.1.5 MonCI

MonCI is an epoxidase that is involved in the biosynthesis of monensin from *Streptomyces cinnamonensis*. It is able to transform a terminal and two internal olefin moieties in the linear polyene, premonensin into a (*S*, *S*)- and two (*R*, *R*)-epoxides, respectively. According to the column calibration result, the MonCI protein exists as monomers in solution, which is about 55.1 kDa. MonCI is a flavin-dependent monooxygenase that relies on the cofactor FAD and NAD(P)H to catalyze the carbon double bonds. The characteristic and unique UV absorbance spectra for FAD have revealed that MonCI naturally carries the FAD cofactor in its oxidized form. The cofactor FAD can be reduced by either NAD(P)H or other reducing agents such as sodium dithionite. The oxidized FAD exhibits yellow color while the reduced FAD exhibits transparent. This is why the purified MonCI protein sample also was yellow color. The UV spectra also showed that addition of the substrate did not change the FAD status or the overall conformation of MonCI.

To study how active the MonCI sample was after purification, we also conducted the *in-vitro* enzyme assays that contained MonCI, cofactor FAD, NADH/NADPH, and the substrate linalool or the compound **24**. After incubation at the room temperature, the reaction mixture was extracted by ethyl acetate and analyzed by GC-MS. Based on their specific ionized fragments, the remining substrate and the products after reaction were able to be identified. It turned out that the product yield, either the products of linalool or the compound **24**, was not very high. It also took a very long time, at least 12 hours to reach to its maximum yield. This could possibly be caused by the low binding-affinity between the substrate and MonCI enzyme. Even though both linalool and the compound **24** contain one or more carbon double bonds, they are still not very similar to the MonCI native substrate in structure. Firstly, the native substrate of MonCI, premonensin **20**, contains a bulky carbonyl terminal before the olefin moieties. Secondly, there are different substitutions on the olefin groups among premonensin and the substrate analogs linalool and the compound **24**. Therefore, these two substrate analogs do not enter into and fit with the MonCI substrate-binding pocket very well. Thus, the product yield was relatively lower. The other possibility is that the purified MonCI sample was not very active or stable. In the process of

purification, we encountered the protein stability issue for MonCI. Firstly, MonCI was not able to be expressed and folded well inside the bacteria. We solved this problem by inserting the *monCI* gene into the pCold I vector and transforming the plasmid into the *E.coli* BL21-AI strain. Secondly, the MonCI protein tended to aggregate during purification. The MonCI proteins was eluted out in different peaks in ion exchange chromatography, which means that MonCI protein was not stable and may stay in different conformations due to misfolding. In this case, addition of 10% (v/v) glycerol did help stabilize the protein. After adding 10% (v/v) glycerol in all the buffers, the MonCI protein was more stable. However, it still precipitated and formed aggregation after overnight at room temperature. Besides, the chloride ions may play an important role in inhibiting the MonCI activity. We have compared the product yields of the MonCI enzyme assays in the different buffer systems with the chloride ion in gradient concentrations. The linalool was used as the substrate analog due to its relatively high transformation rate. For the buffer system without any chloride ion, the estimated linalool oxide yield was 57.6%; for the buffer system with 1.2 mM chloride ion, the estimated linalool oxide yield was 49.3%; for the buffer system with 150 mM chloride ion, the estimated linalool oxide yield was 47.7%; for the buffer system with 300 mM chloride ion, the estimated linalool oxide yield was 45.1%. It showed that the product yield decreased when the concentration of the chloride ion was higher. The product yield was the highest one when there was no chloride ion. This result indicates that chloride ion inhibits the MonCI enzyme activity. Based on the crystal structure analysis, the chloride ion locates near the isoalloxazine ring of FAD and at the NAD(P)H-binding site. It prevents the FAD from moving out and prevents the NAD(P)H from binding to MonCI. So, the chloride ion inhibits the MonCI enzyme activity possibly through affecting the FAD reduction. Due to all the possible reasons, the purified MonCI sample did not show a high activity on catalyzing neither linalool nor the compound **24**.

To further investigate the mechanism of the stereoselective epoxidation, we also solved the crystal structure of MonCI. The diffraction data was solved up to 1.90 Å by the MAD method. There are two identical MonCI monomers in a single unit cell. Its crystal structure reveals the FAD-binding and the substrate-binding domains are fused together as one pocket. The cofactor

FAD is tightly bound inside the protein by forming the hydrogen bonds with the residues and the water molecules. The substrate-pocket is relatively large compared to the MonCI homologs due to the larger size of the substrate. There are two entrances for its substrate-binding pocket: one is the top entrance for solvent; the other is the side entrance for the substrate. The two entrances are orthogonally positioned and they both have an opening with a diameter of about 6.5 Å. The side entrance is more hydrophobic and contains more positively charged residues for interacting with and the acryl carrier protein during the substrate delivery. We predict that premonensin A enters MonCI via the side entrance based on two observations. First, it was previously shown that premonensin A is tethered to a discrete acyl carrier protein, MonACPX, prior to oxidation by MonCI (213). Therefore, MonACPX must first dock with MonCI in order for premonensin A to enter MonCI's active site cavity. Acyl carrier proteins are known to interact with their partner proteins primarily by electrostatic interaction (207-209, 212). Our MonCI crystal structure shows that the surface surrounding the side entrance is electrostatically positive whereas the surface surrounding the top entrance is largely negative (Figure 4.1). Our homology model of MonACPX shows that Ser60, the probable attachment site of the 4´-phosphopantetheine prosthetic group, is surrounded by several acidic residues (Thr45, Thr46, Leu47, Asp49, Asp50, and Glu55) resulting in an electrostatically negatively surface (Figure 4.1). Therefore, the side entrance of MonCI is electrostatically more compatible for interaction with MonACPX than the top entrance. Second, the location of the substrate entrance in other flavin-containing monooxygenases aligns better with MonCI's side entrance than its top entrance. For example, the cavity leading to the side entrance in MonCI aligns very well with the way squalene is thought to bind to the human squalene monooxygenase (193), a close structural homolog of MonCI. These observations strongly suggest that premonensin A enters MonCI via the side entrance.

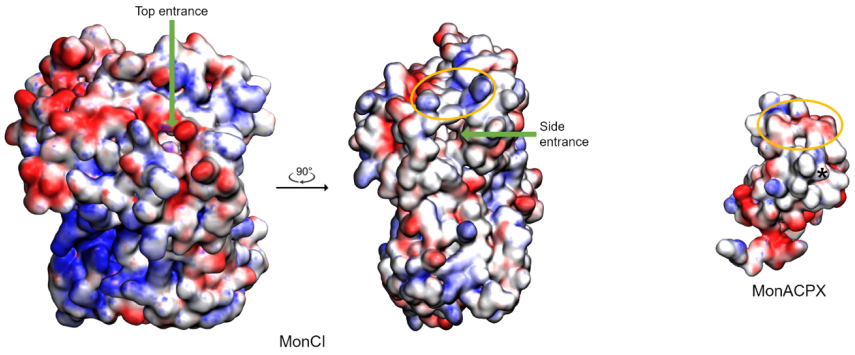


Figure 4.1: Electrostatic surface potential map of MonCI and MonACPX calculated using the VMD and APBS programs. Color scale is from -4 kbTec-1 (red) to 4 kbTec-1 (blue). Colors range from blue (positive, +0.12) to white (neutral, 0) to red (negative, -0.12). The yellow circles indicate the shared contact surface area between MonCI and MonACPX predicted by molecular docking. Asterix denotes the location of Ser60 in MonACPX.

4.1.6 Lsd18

Lsd18 is a flavin-dependent epoxidase that is involved in the biosynthesis of lasalocid A from *Streptomyces lasaliesis* bacterial strain. It is responsible to transform two olefins in the prelasalocid A into two (*R*,*R*)-epoxides. The time-course analysis showed that the initial epoxidation takes place at the terminal olefin of the prelasalocid **21**. According to the column calibration, Lsd18 exists as monomers in solution with the size of 52.5 kDa. Oikawa's group has conducted a series of *in-vitro* enzyme assays to investigate the epoxidation mechanism of Lsd18. Their results have shown that Lsd18 protein sample carried the oxidized form FAD which was able to get reduced by the reducing agent, sodium dithionite. Our UV absorption spectra also displayed the characteristic flavoprotein absorbance curve for the purified Lsd18 sample. We tried both NADH and NADPH as the reducing agents, there was no significant difference in the product yield. Therefore, an equal amount of NADH and NADPH was used in the Lsd18 enzyme assays. Based on the UV absorption spectra, addition of the Fre reductase increased the Lsd18 efficiency remarkably. Addition of GDH did not help improve the product yield. Also, addition of the substrate linalool did not change the FAD oxidation status or the protein overall conformation in a significant level.

In the enzyme assays, Lsd18, Fre reductase, cofactor FAD, cofactor NADH/NADPH, and the substrate were incubated together. After overnight incubation at the room temperature, Lsd18 was able to achieve the highest product yield. After that, more Lsd18 tended to precipitate and starting to lose the activity. The substrate and product also tended to be degraded more. There were three different substrate analogs we used for the enzyme assays. The compound linalool is the simplest one. It is commercially available, and its chemical properties has been studied very well. The compound **24** mimics the terminal olefin in the native substrate prelasalocid A. The compound **26** is the most similar substrate analog that contains both the internal and terminal olefin moieties in prelasalocid. As indicated in the GC-MS results, the transformation rate of the compound **24** is higher than the linalool transformation rate. It was possibly caused by the higher structural similarity to the Lsd18's native substrate. The LC-MS results also showed that Lsd18 was able to catalyze both of the olefins. Oikawa's group showed that the epoxidation of the internal and the terminal olefins can be catalyzed *in vitro* in a controlled manner by adjusting the pH value of the buffer systems. The terminal olefin should be catalyzed first in a Tris buffer at pH 8.0 (reaction 1). The internal olefin should only be catalyzed later in a Bis-Tris buffer with pH at 6.5 (reaction 2). However, in our experiments, the diepoxide product has already been generated after the reaction 1. It could be caused by the long incubation time, or the addition of Fre reductase. Therefore, method optimization needs to be done in the future to control the speed of epoxidation for the substrate **26**. Besides studying how active the purified Lsd18 protein is, we also want to explore the epoxidation stereoselectivity of Lsd18. The tricky part of this study is to differentiate the stereochemistry of the products. Oikawa's group differentiated the stereochemistry of the epoxide products through chemical modification. An external chemical group was attached to the epoxide products, then the stereoisomers could be separated and identified by LC-MS. We also tried the same method; however, our efforts were unsuccessful due to the lack of purification step, low transformation rate, and the limited chemicals. Therefore, the chemical modification method was not applicable in our lab. Then we came up with an alternative method that applies the regioselective property of Lsd19. As mentioned before, Lsd19 is highly regioselective. It

transforms the (22*R*,23*R*)-epoxide substrate into a THF-THP product, and transforms the (22*S*,23*S*)-epoxide substrate into a THF-THF product. Therefore, we can applied Lsd19 in the Lsd18 enzyme assays. Once the olefin is epoxidized by Lsd18, Lsd19 can continue and catalyze the cyclization reaction. Based on the final products either THF-THF or THF-THP by LC-MS analysis, the stereochemistry of the epoxide product would be deduced. This method would help us explore the key residues in Lsd18 that play an important role in the stereoselectivity.

To explore the epoxidation mechanism, we solved the crystal structures of the Lsd18 with and without ligand. The Lsd18 ligand-free crystal structure was solved with the resolution up to 1.54 Å, and the Lsd18-ligand complex crystal structure was solved up to 1.85 Å. The Lsd18 structure is very similar to the MonCI structure (RMSD= 1.63) that contains a fused FAD- and substrate-binding pocket. Each Lsd18 monomer has a tightly bound FAD. Similar to MonCI, there are also two entrances to the substrate-binding pocket. However, these two entrances both locate at the side of the protein surface. The side entrance I aligns more closely with the MonCI side entrance. It is likely to be the substrate entrance that is more hydrophobic and contains more positively charged residues. The side entrance II locates near the top entrance of MonCI. It is likely to be the solvent entrance since it is more hydrophilic. The side entrance is for water molecules to enter and exit from the pocket during the catalysis. The crystal structure of the Lsd18-ligand complex is highly similar to the ligand-free Lsd18 structure (chain A RMSD= 0.38; chain B RMSD= 0.41), except that several residues exhibit different conformations such as Ser108, ,Met124, and Asn440. The conformation changes of the residue side chains may be related to the substrate entrance and product exit. These residues could also play an important role during the catalysis. Therefore, in the future, mutation analysis could be conducted on these residues and confirm how these residues affect the Lsd18 function.

4.1.7 Epoxidases involved in the polyether biosynthesis

All ionophore polyether biosynthetic gene clusters contain a gene that encodes for an epoxidase. For example, nanchangmycin gene cluster contains *nanO* (142), salinomycin gene

cluster contains *salC* (139), tetronomycin gene cluster contains *tmnC* (214). All these genes encode for a flavin-dependent monooxygenase that, similar to MonCI and Lsd18, has the capacity to perform multiple epoxidations on the same substrate molecule, setting them apart from all other flavoenzymes. The phylogenetic tree was generated to show the similarity among all these epoxidases involved in the polyether biosynthesis (Illustration 4.2). Polyether-producing epoxidases share a high sequence identity (49 to 60%). Sequence alignment analysis of MonCI, NanO, SalC, TmnC, and Lsd18 has revealed two highly conserved regions; RKGX^{1}PQX^{2}RHX^{3}HX^{4}LWS (residues 47-64 in MonCI; X^1 = V/Q, X^2 = A/G, X^3 = A/V, X^4 = L/V/I) and AFNPX^{1}X^{2}GHGMS (residues 327-340 in MonCI; X^1 = I/V, X^2 = Y/H). The former sequence is part of an extended loop located on the *si* side of the FAD isoalloxazine ring, while the latter sequence is part of an α-helix located on the *re* side. Interestingly, the active site and the NAD(P)H-binding site we have identified in MonCI and Lsd18 are located in these conserved regions. Therefore, it appears that polyether producing epoxidases in general utilize a similar steric-driven stereocontrol mechanism. It should be noted that the two conserved sequences we have identified are present only in flavin-dependent monooxygenases involved in polyether natural product biosynthesis.

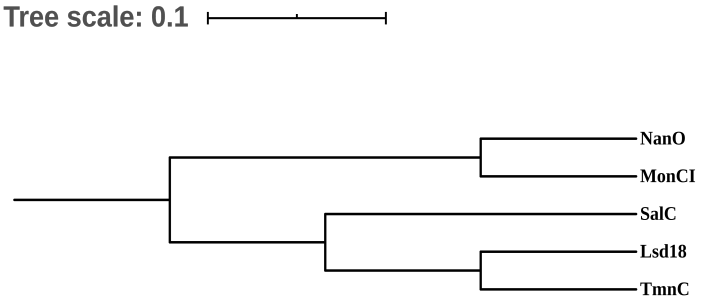


Illustration 4.2 Sequence-based phylogeny of the epoxidases involved in the polyether biosynthesis.

4.2 FAD REDUCTION MECHANISM OF MONCI AND LSD18

With the crystal structures of both MonCI and Lsd18, we are able to explore the mechanism of the stereoselective epoxidation involved in the biosynthesis of natural polyether products. Both

MonCI and Lsd18 belong to the group A flavin-dependent monooxygenases since they are both the two-enzyme system and use the NAD(P)H as the electron donor. The catalytic mechanism of the flavin-dependent monooxygenases includes an oxidative and a reductive half reactions. The FAD exists in the oxidized form in nature. The catalytic cycle starts from the NAD(P)H-mediated FAD reduction. Then the reduced FAD interacts with the molecule oxygen and form a reliable C4a-(hydro)peroxyflavin intermediate that acts as the actual oxygen transferring agent. After C4a-(hydro)peroxyflavin eliminates water and oxygen transferring to the substrate, the FAD returns to its oxidized form.

Based on the structural comparison, both MonCI and Lsd18 share relatively high structural similarity to the typical group A flavin-dependent monooxygenase, *p*-hydroxybenzoate hydroxylase (PHBH) (MonCI-to-PHBH RMSD= 3.4; Lsd18-to-PHBH RMSD= 3.7). The PHBH enzyme is also called as 4-hydroxybenzoate 3-monooxygenase that catalyzes the regioselective hydroxylation of 4-hydroxybenzoate and yield a 3,4-dihydroxybenzoate product. Like other flavin-dependent monooxygenase, PHBH functions in several steps: flavin reduction, formation of the C4a-(hydro)peroxyflavin intermediate; binding and activation of the substrate; product formation and release; also the flavin oxidation.(107) As previously observed on PHBH (222-224), the isoalloxazine ring of FAD is able to move between the "in" and "out" positions during its catalysis for FAD oxidation/reduction and the substrate translocation. Structural alignment of the MonCI and Lsd18 with PHBH "in" and "out" conformations (PDB: 1DOB) indicates that the cofactor FAD currently sits at the "in" position in the crystal structures of MonCI and Lsd18. Both of their flavin isoalloxazine rings are stabilized by the π-π stacking interaction with the imidazole ring of a conserved histidine. To force FAD move to the "out" conformation, we prepared two Lsd18 His71 mutants, His71A and His71S. The enzyme assays did show a slightly lower transformation rate of the substrate **24** for these two mutants. The estimated product yield for H71S mutant was 53.28%. The estimated product yield for H71A was 55.33%. It indicates that the His71 does play role in stabilize FAD, but it is not the key residue during catalysis. We also tried to crystallize the Lsd18 mutants, but it was unsuccessful due to the poor diffraction. In the PHBH

crystal structure (PDB: 1DOB), FAD stayed in the "out" conformation. After structural alignment (RMSD between PHBH and MonCI structures is 3.67) and fitting the PHBH "out"-position FAD into MonCI, there is only one residue, Gln52, causing steric clash (Figure 4.2 A and B). In crystal structure, Gln52 forms a hydrophobic pocket for FAD isoalloxazine ring together with Pro330 and His57. It also forms hydrogen bonds with the FAD pyrophosphate group and ribityl chain. It has the same case for Lsd18. After the structural alignment (RMSD between PHBH and Lsd18 structures is 3.68) and fitting the PHBH "out"-conformation FAD into Lsd18, only one residue, Gln66, causes steric clash (Figure 4.2 C and D). This residue Gln66 interacts with the pyrophosphate and ribityl of the "in"-position FAD. Thus, Gln52 in MonCI and Gln66 in Lsd18 could act as a gate for the FAD-binding pocket. When the FAD moves to the "out" position, this glutamine breaks the hydrogen bond with FAD and moves away from the FAD. When the reduced FAD moved back to the "in" conformation, the glutamine moves closer to FAD and isolates the FAD from the solvent, which helps stabilize the intermediate C4a-(hydro)peroxyflavin. Therefore, rupture of these interactions and conformational change of the residue glutamine could be necessary for FAD to move to the "out" position. To confirm its role in FAD reduction, we prepared two Lsd18 mutants Q66G and Q66N that have shorter side chains than the glutamine. As what was expected, the estimated product yield for Q66N was 18.05%. The estimated product yield for Q66G was 9.6%. Both of these mutants displayed much lower conversion of substrate **24**. It indicated that Gln66 does play an essential role during Lsd18 catalysis. Without the regulation and protection of the glutamine, the cofactor FAD may not be able to move smoothly between "in" and "out" conformations. The intermediate C4a-(hydro)peroxyflavin might not be stable neither.

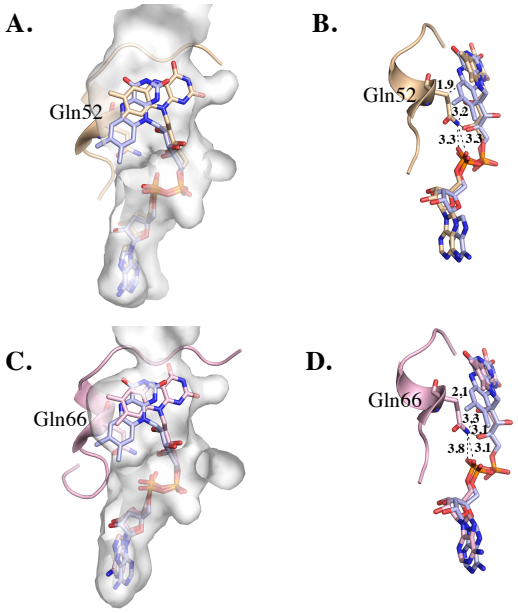


Figure 4.2: Modeling of "out"-conformation FAD in MonCI and Lsd18. (A) Modeled "out'-conformation FAD of PHBH (PDB: 1DOB) in the MonCI pocket. The MonCI FAD is shown in wheat color while the PHBH FAD is shown in light blue color. (B) Side view of the modeled FAD in MonCI. The distance between Gln52 and "in"- and "out"-conformation FAD is labeled in Angstrom. (C)) Modeled "out'-conformation FAD of PHBH (PDB: 1DOB) in the Lsd18 pocket. The Lsd18 FAD is shown in pink color while the PHBH FAD is shown in light blue color. (B) Side view of the modeled FAD in Lsd18. The distance between Gln66 and "in"- and "out"-conformation FAD is labeled in Angstrom.

Besides the FAD conformation changes, we also explored how NAD(P)H interacts with the epoxidases and transfers the electrons to FAD. Since NAD(P)H was not observed in Lsd18 crystal structure, it may indicate that NAD(P)H doesn't remain bound to the protein before and/or after FAD reduction. Based on the sequence analysis, a loop region that consists of the sequence GDSLTA in MonCI and the sequence GDAVAA in Lsd18 is likely to interact with NAD(P)H. Amine and hydroxyl groups in this loop may interact with the diphosphate group of NAD(P)H. Compared with NADPH-bound PHBH structure (PDB: 1K0J), MonCI Arg54 and Lsd18 Arg68 could also bind to the adenosyl moiety of NAD(P)H. A salt bridge may form between the guanidinium group of arginine and the ribose phosphate of NAD(P)H.

Overall, FAD reduction mechanism of the epoxidases represented by MonCI and Lsd18 could be very similar to PHBH. NAD(P)H binds to the conserved motif and some residues like

arginine in the groove in the neighbor of the FAD-binding pocket. Once FAD switches to its "out" position, the hydride transfer occurs from the pro-R side of NAD(P)H (at C4) to the re side of the flavin (at N5). (108, 215, 216) Once transferring is completed, FAD swings back to its "in" position, and NAD(P)+ would be released from the enzyme.

4.3 Stereoselective mechanism of MonCI and Lsd18

Since the crystal structures of MonCI and Lsd18 are highly similar, the MonCI crystal structure was used as the model to investigate the stereoselective mechanism by docking and dynamic simulation. Previous feeding studies and gene deletion experiments have established premonensin A, an (*E*,*E*,*E*)-triene, as the intermediate which MonCI acts on during monensin biosynthesis (217, 218). Sequential epoxidation of all three C=C groups in premonensin A can theoretically produce eight distinct triepoxide enantiomers (Figure 4.3). Remarkably, MonCI produces only the (12*R*,13*R*,16*R*,17*R*,20*S*,21*S*)-triepoxypremonensin A product. To elucidate the molecular basis for this exquisite stereocontrol, we first conducted a modeling study based on the crystal structures of MonCI.

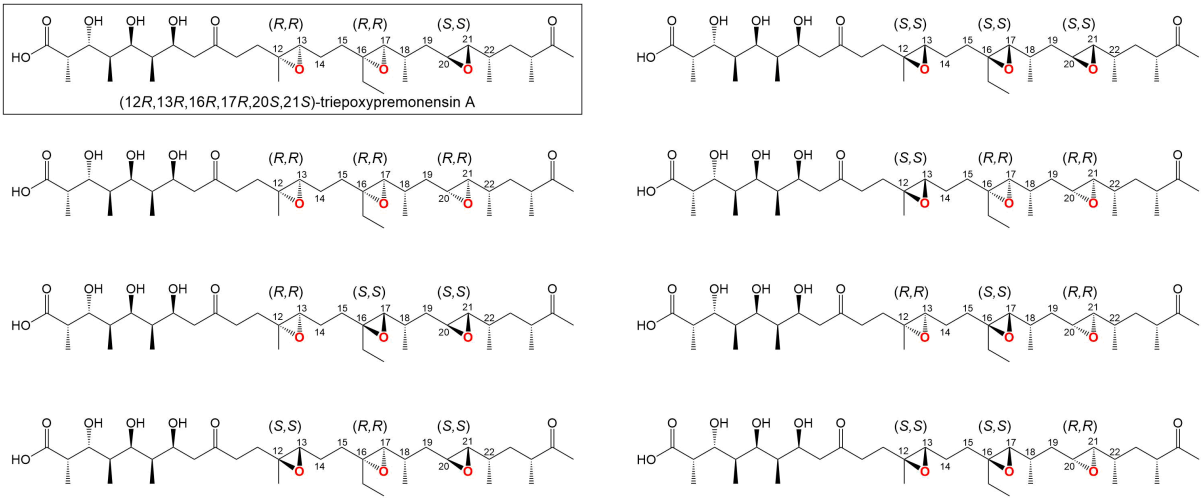


Figure 4.3: Eight possible triepoxypremonensin A enantiomers.

We analyzed the structure of other flavin-dependent monooxygenases for which both the unliganded and the substrate bound crystal structures are available. We found 18 such pairs in the

Protein Data Bank, and they all showed minimal structural differences (RMSD = 0.16 to 0.60 Å) (Table 4.2). The structural changes between the ligand-free and the ligand-bound Lsd18 crystal structures are also very minimal. This indicates that flavin-dependent monooxygenases in general do not undergo significant conformational change upon substrate binding. Therefore, we used the MonCI crystal structure coordinates directly in our modeling study.

Table 4.2: Flavin-dependent monooxygenases for which both the ligand-free and substrate-bound protein crystal structures are available.

Protein name	Unliganded protein structure (PDB accession code)	Ligand-bound protein structure (PDB accession code)	RMSD (Å)
EncM	4XLO	3W8W	0.44
Rifampicin monooxygenase	5KOW	5KOX	0.54
Rifampicin monooxygenase	5VQB	6BRD	0.51
TetX	2XYO	2Y6R	0.41
TetX	4GUV	4A99	0.38
2-hydroxybiphenyl 3-monooxygenase	4Z2U	5BRT	0.25
p-Hydroxybenzoate hydroxylase	1KOL	1KOI	0.33
3-hydroxybenzoate 6-hydroxylase	4BK2	4BK1	0.22
Geranylgeranyl reductase	4OPL	4OPT	0.17
Tetracycline destructase Tet (50)	5TUE	5TUF	0.15
Halogenase PltM	6BZI	6BZA	0.60
3-hydroxybenzoate hydroxylase	2DKI	2DKH	0.18
MTMOIV baeyer-villiger	4K5R	4K5S	0.45
TropB	6NEV	6NET	0.21
PyrH Tryptophan 5-halogenase	2WES	2WET	0.43
Thio glycine oxidase	1NG4	1NG3	0.16
Cmis2 glycine oxidase homologue	6J38	6J39	0.33
Human squalene epoxidase	6C6R	6C6N	0.30

First, we modified the FAD in the crystal structure to the reactive C4a-(hydro)peroxyflavin intermediate. Because the *si* face of the isoalloxazine ring is closely packed against the protein loop (residues His55 to Leu58), it is highly unlikely for oxygen to approach the isoalloxazine ring from this side. In contrast, the *re* face is fully exposed to the water-filled cavity, indicating that molecular oxygen will approach the flavin's C4a atom from the *re* side and produce the reactive C4a-(hydro)peroxyflavin intermediate in which the peroxide group protrudes out from the *re* side.

Next, we constructed six different models of premonensin A bound to MonCI, where each model was created by placing each of the three C=C groups of the substrate next to C4a-(hydro)peroxyflavin in two alternative conformations, one which is conducive to (*R*,*R*)-epoxide formation and the other which is conducive to (*S*,*S*)-epoxide formation (Figure 4.4). The two remaining double bonds that do not participate in the epoxidation reaction were left unmodified. Therefore, these models reflect the substrate conformation immediately prior to the first epoxidation event.

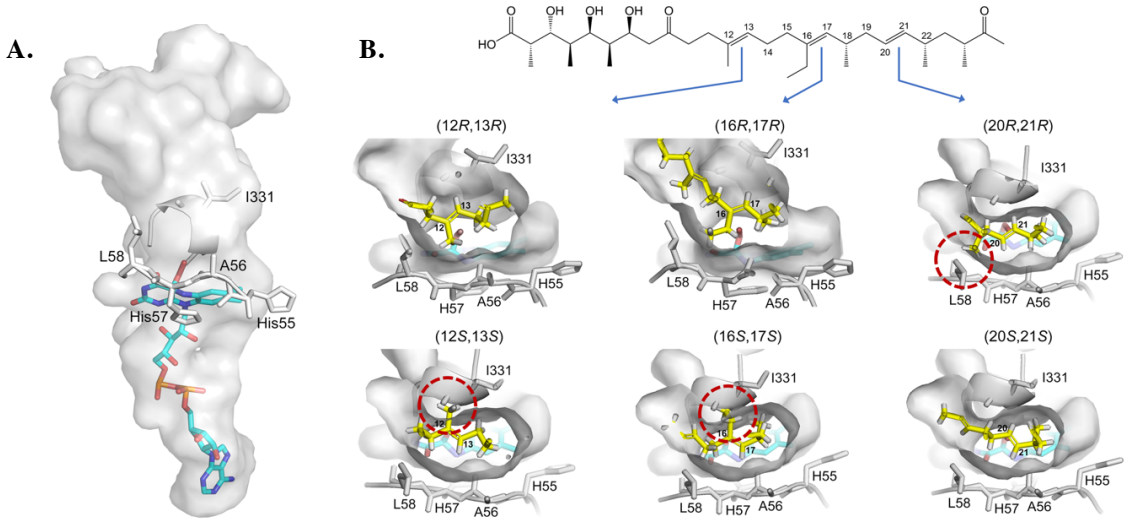


Figure 4.4: Premonensin A docking in MonCI crystal structure. (A) The fused substrate- and FAD-binding pocket of MonCI and the predicted structure of the bound C4a-(hydro)peroxyflavin adenine dinucleotide. (B) Models of the MonCI-premonensin A complex. Premonensin A was built inside the MonCI substrate-binding pocket in the conformation required to produce the six different enantiomeric products. The broken red circle indicates regions of steric clash between the protein and the substrate.

First, we generated a model of the *holo*-MonACPX-MonCI complex (Figure 4.5) by creating a homology model of *apo*-MonACPX using the program MODELLER (219) and then docking this model to the MonCI crystal structure using the program HADDOCK (220). Next, we attached the 4´-phosphopantetheine prosthetic group to the conserved Ser60 of MonACPX in the docked structure. Because polyketide substrates are tethered to the 4´-phosphopantetheine group of acyl carrier proteins via a thioester linkage, our model of the *holo*-MonACPX-MonCI complex

provides a good approximation of the location of the premonensin A C1 atom in the actual enzyme-substrate complex. Then we placed premonensin A inside the MonCI active site cavity in three different conformations, each representing epoxidation of a different double bond in premonensin A and in the orientation that would yield the correct stereoisomer. We imposed three constraints while modeling the different premonensin A conformations: 1) MonCI is completely static, 2) C1 atom of premonensin A is attached to the thiol group of the 4´-phosphopantetheine group, and 3) the C=C bond in premonensin A that is undergoing epoxidation must lie within 6 Å of the hydroperoxyl group of the FAD. Our results show that the active site cavity of MonCI can accommodate premonensin A in all three conformations necessary for epoxidation of C12=C13, C16=C17, and C20=C21 (Figure 4.5 B) The extra-large active site cavity of MonCI allows premonensin A to transition from one required conformation to another even though its C1 atom is fixed in position due to its attachment to the 4´-phosphopantetheine group of MonACPX.

For C12=C13, the substrate conformation that is conducive to (12*R*,13*R*)-epoxide formation could be modelled without any conflict. In contrast, the substrate conformation that is expected to form the (12*S*,13*S*)-epoxide creates a steric clash with the main chain atom of Ile331. For C16=C17, the substrate conformation conducive to (16*R*,17*R*)-epoxide formation could be constructed inside the pocket but not the (16*S*,17*S*)-epoxide conducive conformation, again due to a steric clash with the main chain atom of Ile331. For C20=C21, the (20*S*,21*S*)-epoxide conducive conformation could be modelled but not the (20*R*,21*R*)-epoxide conducive conformation, this time due to a steric clash with the side chain of Leu58. Taken together, our models predict that MonCI exclusively produces the (12*R*,13*R*,16*R*,17*R*,20*S*,21*S*)-triepoxypremonensin A. It is important to recognize that the stereochemical outcome is not determined by the MonCI structure alone but that it is also directed by the structure of the substrate molecule. The steric clashes that we have identified above involve the methyl group attached to C12, the ethyl group attached to C16, and the methyl group attached to C18 of premonensin A. Therefore, modification of the substrate structure at these positions can potentially alter the stereochemical outcome of the epoxidation reaction. Since premonensin A structure is determined by the nature of the modular polyketide

synthases that appear earlier in the monensin biosynthesis pathway, MonCI appears to have coevolved with the polyether producing polyketide synthases.

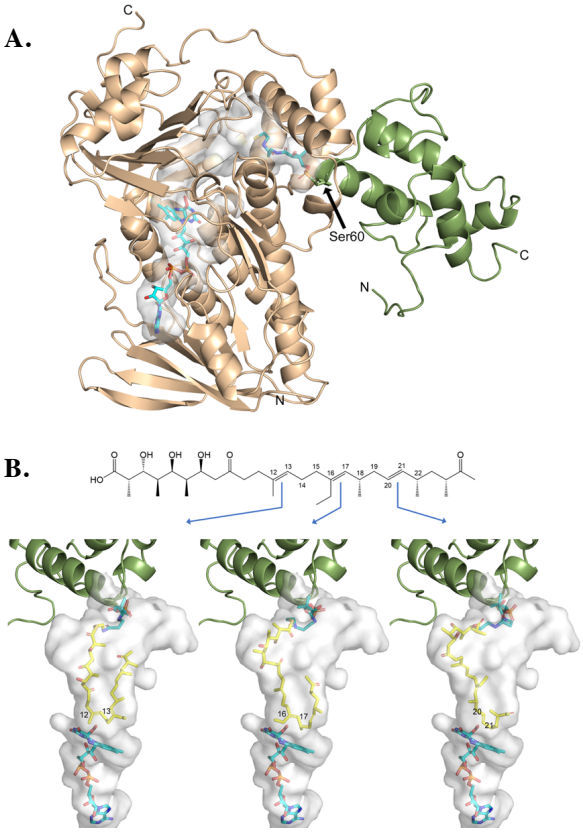


Figure 4.5: MonACPX docking and substrate delivery to MonCI. (A) Predicted structure of the *holo*-MonACPX- MonCI complex. MonACPX and MonCI are colored in green and gold, respectively. (B) Three models of the MonCI-premonensin A-MonACPX complex. Premonensin A was modelled inside the MonCI substrate-binding pocket by positioning each of the three double bonds next to the C4a atom of the isoalloxazine ring. The carbon atoms of 4´-phosphopantetheine, premonensin A, and FAD are colored in green, yellow, and light blue, respectively.

MonCI converts each of the three C=C groups present in premonensin A into an epoxide (86). Since MonCI contains only one FAD, it can only catalyze one epoxidation reaction at a time. Therefore, in order to transform all three C=C groups, either MonCI or the substrate molecule must undergo substantial structural reorganization during the reaction cycle. As previously discussed, flavin-dependent monooxygenases are, in general, rigid enzymes and therefore we predict that premonensin A, rather than MonCI, undergoes conformation change in order to successively

position each of the three C=C groups next to the reactive C4a-(hydro)peroxyflavin moiety. However, premonensin A has limited freedom of movement because it is covalently tethered to an acyl carrier protein which is fixed in position (213). To understand how premonensin A may reorganize itself inside the MonCI active site cavity under such a restrictive condition, we conducted a modeling study based on the MonCI crystal structure.

Atomistic molecular dynamics simulations were performed to investigate the dynamics of the ligand-free MonCI and premonensin A bound MonCI. Premonensin A within MonCI can be short and wide as it folds into a U-shape or a longer and narrower J-shape, depending on which C=C group is positioned at the active site. To examine if premonensin A in different shapes and orientations can fit stably within MonCI's active site cavity, RMSDs of substrate-binding channel residues were evaluated (Figure 4.6 A). In all of the systems in which premonensin A was present, RMSDs of channel residues are smaller than 1.5 Å, and similar to the RMSD for the ligand-free MonCI system. Therefore, the MonCI cavity design is such that it remains stable and robust while accommodating the substrate in all different shapes and orientations examined. Interestingly, average RMSDs of channel residues for native systems, with substrate in (12*R*,13*R*), (16*R*,17*R*) and (20*S*,21*S*) configurations, are always lower than RMSDs of non-native systems, although the differences are within standard deviations.

Epoxidation reactions of premonensin A in MonCI could occur in a predetermined order. The probability of a reaction taking place has been shown to depend on the stability of the substrate within the enzyme's active site pocket (221-224). The stability of premonensin A in different initial conformations within the MonCI active site cavity was tracked during 200 ns of MD simulations (Figure 4.6 B). In the case of the systems where C12=C13 is initially near FAD (both the *R*,*R* and *S*,*S* configurations), the substrate as a whole remains very close to its initial conformation and the binding pose. Furthermore, the C12=C13 group remains in a pose in which a chemical reaction is favored. In the systems where C16=C17 and C20=C21 are initially near FAD, substrates change shapes, and the C=C moves away from the flavin group. Our results suggest that C12=C13 of

premonensin A may be the first C=C group that is epoxidated by MonCI, since only that double bond remains stably positioned at the immediate reaction site.

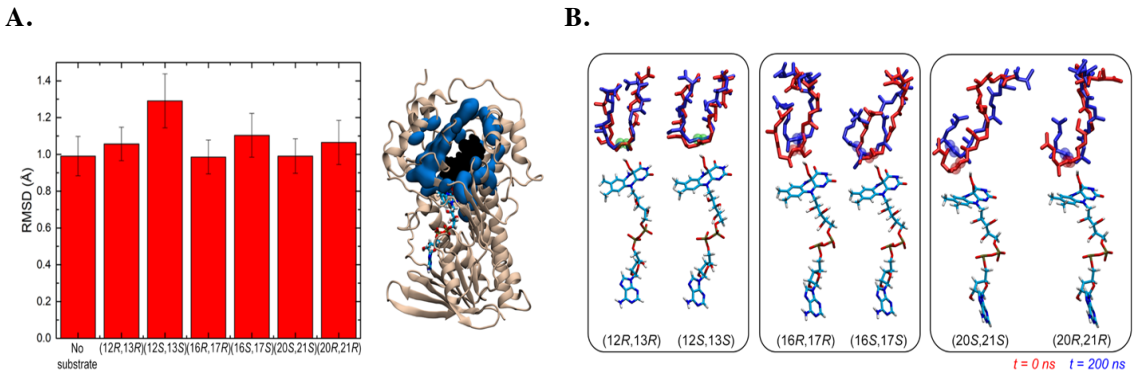


Figure 4.6: Molecular dynamic simulation of the MonCI-premonensin A complex. (A) RMSD of channel residues for each simulated system. Backbone atoms of channel residues are shown as blue surface, the substrate is shown as black surface, and FAD is shown in stick representation. (B) Conformation of premonensin A at the start (red) and after 200 ns (blue) of simulation. Carbon atoms of the double bond nearest to the flavin group are shown as transparent spheres. In the (12R,13R) and (12S,13S) systems, these carbon atoms are not displaced after 200 ns. In all other systems, these carbon atoms become displaced from their initial positions.

For Lsd18, it acts on the prelasalocid A, an (*E*,*E*)-diene, as the intermediate during the biosynthesis of lasalocid A. It transforms the diene intermediate stereoselectively into the (18*R*,19*R*,22*R*,23*R*)-disepoxylasalocid A product (Figure 4.7 C). The internal and terminal carbon double bonds were also modeled at the Lsd18 active site, which showed the same results as we observed in the MonCI modeling (Figure 4.7 B). The alkene groups were only be able to enter the active site in a certain conformation due to the space restrain. When the substrate was modeled in the unfavored conformation at the active site, the space clash between the ethyl group attached to the alkenes and the protein residues could not be avoided. The backbone of residue Val342 and the side chain of residue Ile72 play an essential role in restrain the active site space. As shown in the Lsd18-ligand complex structure, the substrate **26** stayed at the active site at the conformation to product an (*R*,*R*)-epoxide. The stereoselectivity of Lsd18 is determined by the group modification on the alkene substrate and the space restrain by residues Ile72 and Val342 (Figure

4.7 A and B). To further test our hypothesis, we mutated the residue Ile72 into an alanine or a glycine. These two mutants I72A and I72G provide more space at the Lsd18 active site due to the shorter side chains of alanine and glycine compared to isoleucine. Based on the GC-MS analysis, both of these mutants gave slightly lower product yield for the substrate **24**. However, due to the limitation on differentiating the product stereochemistry, we were not able to determine how the stereoselectivity of the mutants was. As mentioned before, the compound **26** and Lsd19 could be applied in the Lsd18 enzyme assays to identify the stereochemistry of the epoxide products.

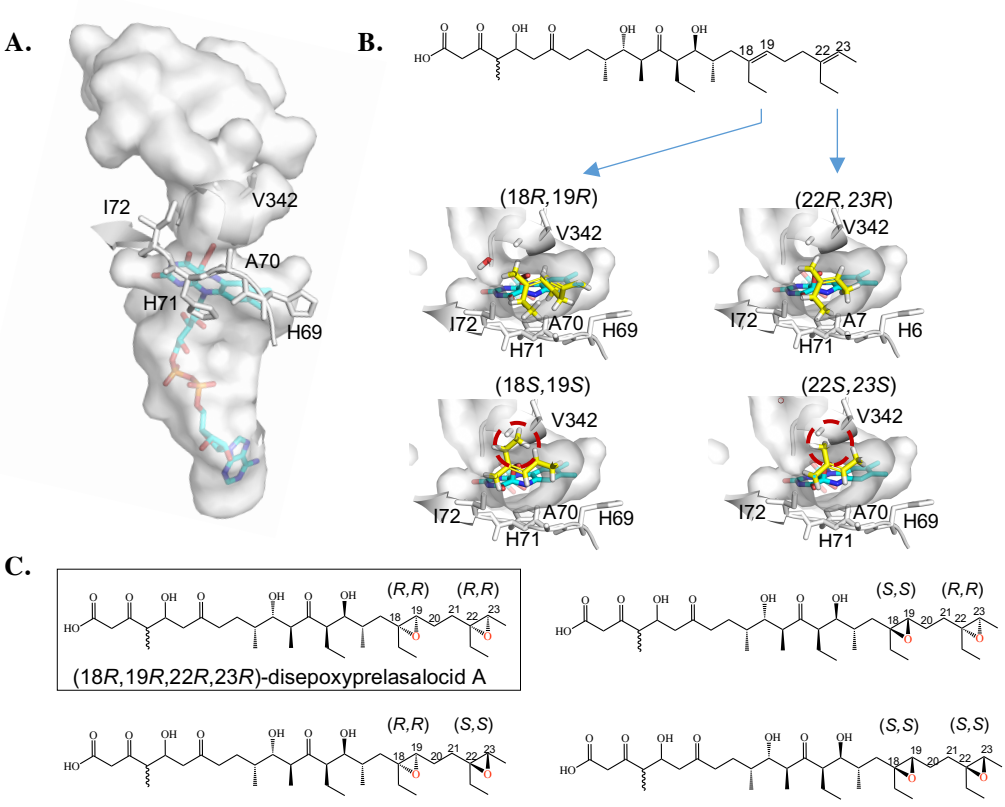


Figure 4.7: Prelasalocid A docking in Lsd18. (A) The fused substrate- and FAD-binding pocket of Lsd18 and the predicted structure of the bound C4a-(hydro)peroxyflavin adenine dinucleotide. (B) Models of the Lsd18-prelasalocid A complex. Prelasalocid A was built inside the Lsd18 substrate-binding pocket in the conformation required to produce the four different enantiomeric products. The broken red circle indicates regions of steric clash between the protein and the substrate. (C) Four possible disepoxyprelasalocid A enantiomers.

Overall, our crystallographic and computational studies of MonCI and Lsd18 have provided important molecular insights into how the epoxidases involved in the polyether

biosynthesis achieve the stereoselective epoxidation during the biosynthesis of the natural polyether products. We proposed that the epoxidases that are involved in the polyether biosynthesis exert the stereoselectivity by dictating, via pre-formed pocket shape, which face of the substrate alkene is directed at C4a-(hydro)peroxyflavin. Residues such as Leu58 and Ile331 in MonCI, Ile72 and Val342 in Lsd18, are the key residues that restrain the space of the active site; thus they may play an important role in the stereoselectivity of epoxidation.

Chapter 5: Conclusions and Future directions

5.1 MAJOR CONCLUSIONS

A combination of X-ray crystallographic and computational methods has been applied to reveal the structural basis for the stereoselective epoxidation involved in the biosynthesis of polyether natural products. The flavin-dependent monooxygenases MonCI and Lsd18 are the epoxidases responsible for the biosynthesis of monensin and lasalocid, respectively. Both of their crystal structures contain a fused FAD- and substrate-binding pocket where FAD reduction and epoxidation can be related together. Since their structures are all similar to the ones of p-hydroxybenzoate hydroxylases, the mechanism of FAD reduction could be very similar, which involved the conformation change of the FAD before and after the reduction. By comparing the structures of Lsd18 ligand-free and ligand-bound complex structures, we notice that the side chains of some residues shifted when the substrate or product exist in the pocket. The conformational change of these residues may play an important role in promoting the substrate entrance and the product exit. Through docking and molecular dynamic simulation work, we propose that the MonCI and Lsd18 control the epoxide stereochemistry by the unique shape at their active sites, which allows only one certain face of the substrate alkene to be directed at the reactive C4a-(hydro)peroxyflavin. More interestingly, MonCI and Lsd18 share high structural and sequence similarity at their active sites, where those residues are significantly conserved among all the epoxidases involved in the polyether biosynthesis. It is likely that the mechanism of epoxidation is uniformed during the polyether biosynthesis. Besides, MonCI's and Lsd18's capacity to epoxide multiple alkenes on their native substrates is made possible by their relatively large pocket volume. It enables the substantial movement and refolding of the substrate molecule when it remains bound to the enzyme.

5.2 FUTURE DIRECTION

Even though the hypothesis of the stereoselective epoxidation mechanism has been proposed, it has not been fully proved yet. We solved the complex structure of Lsd18 with its

substrate/product, the R_{work}/R_{free} ratio was not good enough. Further modeling and refinement are required to be done to achieve a more reliable Lsd18 complex structure. In this way, we would be able to study interactions between the ligand (substrate/product) and the protein residues more accurately. Based on the structural information of MonCI and Lsd18, there are several residues found to play an essential role in restraining the shape and space at the active site. Gene mutagenesis has been conducted on these residues in order to alter or remove the stereoselectivity for Lsd18. All of these Lsd18 mutants have already been constructed and successfully purified. However, more enzyme assays need to be performed with the substrate analog **26** to confirm the stereoselectivity of the mutants. Due to its regioselectivity towards the epoxides, Lsd19 enzyme will be included in the enzyme assays to help differentiate the stereochemistry of the Lsd18 epoxidation products.

Regarding to the long-term future plans, our studies could be applied to drug research by bioengineering the epoxidases to produce various polyether analogs. In recent years, much progress has been made in engineering polyketide biosynthesis pathways. For example, researchers have successfully produced modified polyketides using engineered polyketide synthases. Although there are many successful cases, it is widely known that polyketide engineering frequently results in drastically reduced product yields, generation of side products, and random failures. These issues are likely due to disrupted protein-protein interactions in the engineered enzymes and incompatibility of the modified intermediate with downstream enzymes. Access to high-resolution structural information of the biosynthetic enzymes will help avoid these problems. For example, identification of key residues involved in protein-protein recognition will allow construction of engineered enzymes that preserve this essential function rather than disrupting it. And knowledge of the enzyme active site architecture will allow structure-based remodeling of the active site pocket to accommodate nonnative substrates. Different polyether analogs could be synthesized by reshaping the active site of the epoxidases. And this reshaping can be achieved by mutating the key residues at the active site. For example, if the Lsd18 active site is modified, the (22*S*,23*S*)-epoxide could be generated instead of its native (22*R*, 23*R*)-epoxide.

After Lsd19 cyclization, a final THF-THF product will be produced instead of the native THF-THP lasalocid. However, our investigation of MonCI and Lsd18 has also uncovered an additional issue for polyketide engineering, which is that the stereochemical outcome of epoxidation may be unintentionally altered if the polyketide backbone is modified. If we wish to modify the ethyl group attached to C16 in monensin to a methyl group, we can achieve this by engineering the upstream polyketide synthase protein by substituting its ethylmalonyl-CoA specific acyltransferase domain with a methylmalonyl-CoA specific counterpart. However, based on our MonCI crystal structure and our modeling analysis, epoxidation at C16=C17 on the resulting modified polyketide chain is likely to produce an alternate stereoisomer, or a mixture of enantiomers, because there will be no steric clash with the key residue Ile331. Therefore, to preserve the stereochemistry of the natural product, it may be necessary to engineer MonCI alongside the polyketide synthase. This example illustrates that we must take into consideration all of downstream biosynthetic enzymes, and not just the polyketide synthase proteins, when engineering a polyketide biosynthesis pathway. Although this greatly complicates polyketide engineering efforts, access to high-resolution structures of biosynthetic enzyme should help navigate through this challenge.

References

1. Anonymous, All natural. *Nat Chem Biol* **3**, 351 (2007).
2. A. L. Demain, A. Fang, The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* **69**, 1-39 (2000).
3. G. Strobel, B. Daisy, Bioprospecting for microbial endophytes and their natural products. *Microbiol Mol Biol Rev* **67**, 491-502 (2003).
4. R. E. Procópio, I. R. Silva, M. K. Martins, J. L. Azevedo, J. M. Araújo, Antibiotics produced by Streptomyces. *Braz J Infect Dis* **16**, 466-471 (2012).
5. A. Saxena, R. Kumari, U. Mukherjee, P. Singh, R. Lal, Draft Genome Sequence of the Rifamycin Producer Amycolatopsis rifamycinica DSM 46095. *Genome Announc* **2** (2014).
6. S. A. Cochrane, J. C. Vederas, Lipopeptides from Bacillus and Paenibacillus spp.: A Gold Mine of Antibiotic Candidates. *Med Res Rev* **36**, 4-31 (2016).
7. R. H. Blum, S. K. Carter, K. Agre, A clinical review of bleomycin--a new antineoplastic agent. *Cancer* **31**, 903-914 (1973).
8. C. Montecucco, J. Molgó, Botulinal neurotoxins: revival of an old killer. *Curr Opin Pharmacol* **5**, 274-279 (2005).
9. E. P. Abraham, Discovery and development of penicillin. *Dent Rec (London)* **66**, 197-200 (1946).
10. S. H. Ferreira, A bradykinin-potentiating factor (BPF) present in the venom of bothrops jararca. *British journal of pharmacology and chemotherapy* **24**, 163-169 (1965).
11. R. L. Randall, M. G. Cable, Trabectedin tactics: from sea squirts to sarcomas. *Lancet Oncol* **16**, 243-244 (2015).
12. P. G. Kremsner, S. Krishna, Antimalarial combinations. *Lancet* **364**, 285-294 (2004).
13. L. Mu, S. S. Feng, A novel controlled release formulation for the anticancer drug paclitaxel (Taxol): PLGA nanoparticles containing vitamin E TPGS. *J Control Release* **86**, 33-48 (2003).
14. C. Hertweck, The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **48**, 4688-4716 (2009).
15. J. Staunton, K. J. Weissman, Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**, 380-416 (2001).
16. J. Huffman, R. Gerber, L. Du, Recent advancements in the biosynthetic mechanisms for polyketide-derived mycotoxins. *Biopolymers* **93**, 764-776 (2010).
17. H. Brockmann, W. Henkel, Pikromycin, ein bitter schmeckendes Antibioticum aus Actinomyceten (Antibiotica aus Actinomyceten, VI. Mitteil. *Chemische Berichte* **84**, 284-288 (1951).
18. P. Sensi, P. Margalith, M. T. Timbal, Rifomycin, a new antibiotic; preliminary report. *Farmaco Sci* **14**, 146-147 (1959).
19. M. Balerna, W. Keller-Schierlein, C. Martius, H. Wolf, H. Zähner, [Metabolic products of microorganisms. 72. Naphthomycin, an antimetabolite of vitamin K]. *Arch Mikrobiol* **65**, 303-317 (1969).
20. H. G. Floss, T. W. Yu, Lessons from the rifamycin biosynthetic gene cluster. *Curr Opin Chem Biol* **3**, 592-597 (1999).
21. A. Huczyński, Polyether ionophores-promising bioactive molecules for cancer therapy. *Bioorg Med Chem Lett* **22**, 7002-7010 (2012).

22. M. Nakamura *et al*., Inhibitory effects of polyethers on human immunodeficiency virus replication. *Antimicrob Agents Chemother* **36**, 492-494 (1992).
23. C. J. Dutton, B. J. Banks, C. B. Cooper, Polyether ionophores. *Nat Prod Rep* **12**, 165-181 (1995).
24. T. R. Callaway *et al*., Ionophores: their use as ruminant growth promotants and impact on food safety. *Curr Issues Intest Microbiol* **4**, 43-51 (2003).
25. D. A. Kevin Ii, D. A. Meujo, M. T. Hamann, Polyether ionophores: broad-spectrum and promising biologically active molecules for the control of drug-resistant bacteria and parasites. *Expert opinion on drug discovery* **4**, 109-146 (2009).
26. J. Rutkowski, B. Brzezinski, Structures and properties of naturally occurring polyether antibiotics. *Biomed Res Int* **2013**, 162513 (2013).
27. N. Civjan (2012) Natural products in chemical biology. (Wiley,, Hoboken, N.J.), p 1 online resource.
28. J. J. Fernández, M. L. Souto, M. Norte, Marine polyether triterpenes. *Nat Prod Rep* **17**, 235-246 (2000).
29. F. C. Pacheco *et al*., New polyether triterpenoids from Laurencia viridis and their biological evaluation. *Mar Drugs* **9**, 2220-2235 (2011).
30. K. Suenaga, T. Shibata, N. Takada, H. Kigoshi, K. Yamada, Aurilol, a Cytotoxic Bromotriterpene Isolated from the Sea Hare Dolabella auricularia. *Journal of Natural Products* **61**, 515-518 (1998).
31. N. S. Sheikh, Synthetic endeavours towards oxasqualenoid natural products containing 2,5-disubstituted tetrahydrofurans--eurylene and teurilene. *Nat Prod Rep* **31**, 1088-1100 (2014).
32. I. Vilotijevic, T. F. Jamison, Synthesis of marine polycyclic polyethers via endo-selective epoxide-opening cascades. *Mar Drugs* **8**, 763-809 (2010).
33. Y.-Y. Lin *et al*., Isolation and structure of brevetoxin B from the "red tide" dinoflagellate Ptychodiscus brevis (Gymnodinium breve). *Journal of the American Chemical Society* **103**, 6773-6775 (1981).
34. M. Murata *et al*., Partial structures and binding studies of maitotoxin, the most potent marine toxin. *Bull Soc Pathol Exot* **85**, 470-473 (1992).
35. M. Murata, H. Naoki, S. Matsunaga, M. Satake, T. Yasumoto, Structure and Partial Stereochemical Assignments for Maitotoxin, the Most Toxic and Largest Natural Non-Biopolymer. *Journal of the American Chemical Society* **116**, 7098-7107 (1994).
36. M. Murata *et al*., Structure of maitotoxin. *Journal of the American Chemical Society* **115**, 2060-2062 (1993).
37. A. R. Gallimore, J. B. Spencer, Stereochemical uniformity in marine polyether ladders--implications for the biosynthesis and structure of maitotoxin. *Angew Chem Int Ed Engl* **45**, 4406-4413 (2006).
38. K. G. Sellner, G. J. Doucette, G. J. Kirkpatrick, Harmful algal blooms: causes, impacts and detection. *J Ind Microbiol Biotechnol* **30**, 383-406 (2003).
39. J. H. Landsberg, L. J. Flewelling, J. Naar, Karenia brevis red tides, brevetoxins in the food web, and impacts on natural resources: Decadal advancements. *Harmful Algae* **8**, 598-607 (2009).
40. D. R. Hardison, W. G. Sunda, D. Shea, R. W. Litaker, Increased toxicity of Karenia brevis during phosphate limited growth: ecological and evolutionary implications. *PLoS One* **8**, e58545 (2013).

41. M. A. Poli, T. J. Mende, D. G. Baden, Brevetoxins, unique activators of voltage-sensitive sodium channels, bind to specific sites in rat brain synaptosomes. *Mol Pharmacol* **30**, 129-135 (1986).
42. H. Nagai, M. Murata, K. Torigoe, M. Satake, T. Yasumoto, Gambieric acids, new potent antifungal substances with unprecedented polyether structures from a marine dinoflagellate Gambierdiscus toxicus. *The Journal of Organic Chemistry* **57**, 5448-5453 (1992).
43. A. J. Bourdelais *et al*., Brevenal is a natural inhibitor of brevetoxin action in sodium channel receptor binding assays. *Cell Mol Neurobiol* **24**, 553-563 (2004).
44. K. Nakanishi, The chemistry of brevetoxins: A review. *Toxicon* **23**, 473-479 (1985).
45. Y. Shimizu, Biosynthesis and biotransformation of marine invertebrate toxins. *Natural Toxins. Harris JB (ed) Clarendon Press, Oxford UK*, 115-125 (1986).
46. M. S. Lee, G. Qin, K. Nakanishi, M. G. Zagorski, Biosynthetic studies of brevetoxins, potent neurotoxins produced by the dinoflagellate Gymnodinium breve. *Journal of the American Chemical Society* **111**, 6234-6241 (1989).
47. M. Izumikawa, M. Murata, K. Tachibana, T. Fujita, H. Naoki, 18O-Labelling pattern of okadaic acid from H218O in dinoflagellate Prorocentrum lima elucidated by tandem mass spectrometry. *European Journal of Biochemistry* **267**, 5179-5183 (2000).
48. M. Murata, M. Izumikawa, K. Tachibana, T. Fujita, H. Naoki, Labeling Pattern of Okadaic Acid from 18O2 and [18O2]Acetate Elucidated by Collision-Induced Dissociation Tandem Mass Spectrometry. *Journal of the American Chemical Society* **120**, 147-151 (1998).
49. M. Yamazaki *et al*., Origins of Oxygen Atoms in a Marine Ladder-Frame Polyether: Evidence of Monooxygenation by 18O-Labeling and Using Tandem Mass Spectrometry. *The Journal of Organic Chemistry* **77**, 4902-4906 (2012).
50. B. Shen, Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**, 285-295 (2003).
51. S. Smith, S. C. Tsai, The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-1072 (2007).
52. C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**, 162-190 (2007).
53. M. B. Austin, J. P. Noel, The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20**, 79-110 (2003).
54. B. J. Rawlings, Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis). *Nat Prod Rep* **18**, 190-227 (2001).
55. S. J. Moss, C. J. Martin, B. Wilkinson, Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* **21**, 575-593 (2004).
56. J. Kennedy *et al*., Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368-1372 (1999).
57. S. M. Ma, Y. Tang, Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB. *FEBS J* **274**, 2854-2864 (2007).
58. Z. Zhang, H. X. Pan, G. L. Tang, New insights into bacterial type II polyketide biosynthesis. *F1000Res* **6**, 172 (2017).
59. C. R. Hutchinson, Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem Rev* **97**, 2525-2536 (1997).

60. A. Grimm, K. Madduri, A. Ali, C. R. Hutchinson, Characterization of the Streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151**, 1-10 (1994).
61. C. J. Dutton, B. J. Banks, C. B. Cooper, Polyether ionophores. *Natural Product Reports* **12**, 165-181 (1995).
62. J. W. Westley, Antibiotic Structure and Biosynthesis. *Journal of Natural Products* **49**, 35-47 (1986).
63. T. Nakata *et al*., A total synthesis of lasalocid A. *Journal of the American Chemical Society* **100**, 2933-2935 (1978).
64. R. E. Ireland, S. Thaisrivongs, C. S. Wilcox, Total synthesis of lasalocid A (X537A). *Journal of the American Chemical Society* **102**, 1155-1157 (1980).
65. R. W. Hoffmann, Allylic 1,3-strain as a controlling factor in stereoselective transformations. *Chemical Reviews* **89**, 1841-1860 (1989).
66. S. Yamaguchi, "7 - Nuclear Magnetic Resonance Analysis Using Chiral Derivatives" in Asymmetric Synthesis, J. D. Morrison, Ed. (Academic Press, 1983), pp. 125-152.
67. G. R. Weisman, "8 - Nuclear Magnetic Resonance Analysis Using Chiral Solvating Agents" in Asymmetric Synthesis, J. D. Morrison, Ed. (Academic Press, 1983), pp. 153-171.
68. M. T. Reetz, Structural, mechanistic, and theoretical aspects of chelation-controlled carbonyl addition reactions. *Accounts of Chemical Research* **26**, 462-468 (1993).
69. M. Prudhomme, G. Dauphin, J. Guyot, G. Jeminet, Semisynthesis of A23187 (calcimycin) analogs. II. Introduction of a methyl group on the benzoxazole ring. *J Antibiot (Tokyo)* **37**, 627-634 (1984).
70. T. Nakata, Y. Kishi, Synthetic studies on polyether antibiotics. III. A stereocontrolled synthesis of isolasalocid ketone from acyclic precursors. *Tetrahedron Letters* **19**, 2745-2748 (1978).
71. G. Schmid, T. Fukuyama, K. Akasaka, Y. Kishi, Synthetic studies on polyether antibiotics. 4. Total synthesis of monensin. 1. Stereocontrolled synthesis of the left half of monensin. *Journal of the American Chemical Society* **101**, 259-260 (1979).
72. T. Fukuyama, C. L. J. Wang, Y. Kishi, Synthetic studies on polyether antibiotics. 5. Total synthesis of monensin. 2. Stereocontrolled synthesis of the right half of monensin. *Journal of the American Chemical Society* **101**, 260-262 (1979).
73. T. Fukuyama *et al*., Synthetic studies on polyether antibiotics. 6. Total synthesis of monensin. 3. Stereocontrolled total synthesis of monensin. *Journal of the American Chemical Society* **101**, 262-263 (1979).
74. W. Clark Still, J. H. McDonald, Chelation-controlled nucleophilic additions. 1. A highly effective system for asymmetric induction in the reaction of organometallics with α-alkoxyketones. *Tetrahedron Letters* **21**, 1031-1034 (1980).
75. W. Clark Still, J. A. Schneider, Chelation-controlled nucleophilic additions. 2. A highly effective system for asymmetric induction in the reaction of organometallics with β-alkoxyaldehydes. *Tetrahedron Letters* **21**, 1035-1038 (1980).
76. M. P. Edwards, S. V. Ley, S. G. Lister, B. D. Palmer, D. J. Williams, Total synthesis of the ionophore antibiotic X-14547A (Indanomycin). *The Journal of Organic Chemistry* **49**, 3503-3516 (1984).
77. Y. Kishi, S. Hatakeyama, M. D. Lewis., Total synthesis of polyether antibiotics narasin and salinomycin. *Frontiers of Chemistry* (Pergamon, 1982), pp. 287-304.

78. S. J. Danishefsky, H. G. Selnick, M. P. DeNinno, R. E. Zelle, The total synthesis of zincophorin. *Journal of the American Chemical Society* **109**, 1572-1574 (1987).
79. O. Tyc, C. Song, J. S. Dickschat, M. Vos, P. Garbeva, The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol* **25**, 280-292 (2017).
80. L. E. Day *et al.*, Biosynthesis of monensin. *Antimicrob Agents Chemother* **4**, 410-414 (1973).
81. D. E. Cane, T. C. Liang, H. Hasler, Polyether biosynthesis. 2. Origin of the oxygen atoms of monensin A. *Journal of the American Chemical Society* **104**, 7274-7281 (1982).
82. D. E. Cane, T.-C. Liang, H. Hasler, Polyether biosynthesis. Origin of the oxygen atoms of monensin A. *Journal of the American Chemical Society* **103**, 5962-5965 (1981).
83. D. E. Cane, W. D. Celmer, J. W. Westley, Unified stereochemical model of polyether antibiotic structure and biogenesis. *Journal of the American Chemical Society* **105**, 3594-3600 (1983).
84. P. F. Leadlay *et al.*, Engineering of complex polyketide biosynthesis — insights from sequencing of the monensin biosynthetic gene cluster. *Journal of Industrial Microbiology and Biotechnology* **27**, 360-367 (2001).
85. M. Oliynyk *et al.*, Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization. *Molecular Microbiology* **49**, 1179-1190 (2003).
86. A. Bhatt *et al.*, Accumulation of an E,E,E-triene by the monensin-producing polyketide synthase when oxidative cyclization is blocked. *Angew Chem Int Ed Engl* **44**, 7075-7078 (2005).
87. R. D. Goodrich *et al.*, Influence of monensin on the performance of cattle. *J Anim Sci* **58**, 1484-1498 (1984).
88. D. B. Collum, J. H. McDonald, W. C. Still, Synthesis of the polyether antibiotic monensin. 2. Preparation of intermediates. *Journal of the American Chemical Society* **102**, 2118-2120 (1980).
89. B. M. Harvey *et al.*, Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *Chembiochem* **7**, 1435-1442 (2006).
90. A. R. Gallimore *et al.*, Evidence for the role of the monB genes in polyether ring formation during monensin biosynthesis. *Chem Biol* **13**, 453-460 (2006).
91. K. Sato, A. Minami, T. Ose, H. Oguri, H. Oikawa, Remarkable synergistic effect between MonBI and MonBII on epoxide opening reaction in ionophore polyether monensin biosynthesis. *Tetrahedron Letters* **52**, 5277-5280 (2011).
92. A. Minami *et al.*, Allosteric regulation of epoxide opening cascades by a pair of epoxide hydrolases in monensin biosynthesis. *ACS Chem Biol* **9**, 562-569 (2014).
93. J. W. Westley, R. H. Evans, G. Harvey, R. G. Pitcher, D. L. Pruess, Biosynthesis of lasalocid. I. Incorporation of 13C and 14C labelled substrates into lasalocid A. *J Antibiot (Tokyo)* **27**, 288-297 (1974).
94. J. W. Westley, J. F. Blount, R. H. Evans, A. Stempel, J. Berger, Biosynthesis of lasalocid. II. X-ray analysis of a naturally occurring isomer of lasalocid A. *J Antibiot (Tokyo)* **27**, 597-604 (1974).
95. M. M. Sherman, C. R. Hutchinson, Biosynthesis of lasalocid A: biochemical mechanism for assembly of the carbon framework. *Biochemistry* **26**, 438-445 (1987).

96. L. Smith, H. Hong, J. B. Spencer, P. F. Leadlay, Analysis of specific mutants in the lasalocid gene cluster: evidence for enzymatic catalysis of a disfavoured polyether ring closure. *Chembiochem* **9**, 2967-2975 (2008).
97. A. Migita *et al*., Identification of a gene cluster of polyether antibiotic lasalocid from Streptomyces lasaliensis. *Biosci Biotechnol Biochem* **73**, 169-176 (2009).
98. C. R. Hutchinson, M. M. Sherman, A. G. McInnes, J. A. Walter, J. C. Vederas, Biosynthesis of macrolides. 6. Mechanism of stereocontrol during the formation of lasalocid A. *Journal of the American Chemical Society* **103**, 5956-5959 (1981).
99. M. L. Heathcote, J. Staunton, P. F. Leadlay, Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem Biol* **8**, 207-220 (2001).
100. Z. Hu *et al*., A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases. *Microbiology* **149**, 2213-2225 (2003).
101. M. H. Eppink, H. A. Schreuder, W. J. Van Berkel, Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci* **6**, 2454-2458 (1997).
102. O. Dym, D. Eisenberg, Sequence-structure analysis of FAD-containing proteins. *Protein Sci* **10**, 1712-1728 (2001).
103. A. Minami *et al*., Sequential enzymatic epoxidation involved in polyether lasalocid biosynthesis. *J Am Chem Soc* **134**, 7246-7249 (2012).
104. B. Entsch, W. J. van Berkel, Structure and mechanism of para-hydroxybenzoate hydroxylase. *FASEB J* **9**, 476-483 (1995).
105. W. Yang *et al*., TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *J Biol Chem* **279**, 52346-52352 (2004).
106. K. Groom, A. Bhattacharya, D. L. Zechel, Rebeccamycin and staurosporine biosynthesis: insight into the mechanisms of the flavin-dependent monooxygenases RebC and StaC. *Chembiochem* **12**, 396-400 (2011).
107. D. L. Gatti *et al*., The mobile flavin of 4-OH benzoate hydroxylase. *Science* **266**, 110-114 (1994).
108. J. Wang *et al*., Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. *Proc Natl Acad Sci U S A* **99**, 608-613 (2002).
109. C. Enroth, H. Neujahr, G. Schneider, Y. Lindqvist, The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure* **6**, 605-617 (1998).
110. G. Suzuki *et al*., Analysis of Enantiofacial Selective Epoxidation Catalyzed by Flavin-containing Monooxygenase Lsd18 Involved in Ionophore Polyether Lasalocid Biosynthesis. *Chemistry Letters* **43**, 1779-1781 (2014).
111. A. Minami *et al*., Enzymatic epoxide-opening cascades catalyzed by a pair of epoxide hydrolases in the ionophore polyether biosynthesis. *Org Lett* **13**, 1638-1641 (2011).
112. Y. Matsuura *et al*., Intriguing substrate tolerance of epoxide hydrolase Lsd19 involved in biosynthesis of the ionophore antibiotic lasalocid A. *Org Lett* **12**, 2226-2229 (2010).
113. K. Hotta *et al*., Enzymatic catalysis of anti-Baldwin ring closure in polyether biosynthesis. *Nature* **483**, 355-358 (2012).
114. F. T. Wong *et al*., Epoxide hydrolase-lasalocid a structure provides mechanistic insight into polyether natural product biosynthesis. *J Am Chem Soc* **137**, 86-89 (2015).

115. V. Massey, Activation of molecular oxygen by flavins and flavoproteins. *J Biol Chem* **269**, 22459-22462 (1994).
116. M. M. E. Huijbers, S. Montersino, A. H. Westphal, D. Tischler, W. J. H. van Berkel, Flavin dependent monooxygenases. *Archives of Biochemistry and Biophysics* **544**, 2-17 (2014).
117. N. Prileschajew, Oxydation ungesättigter Verbindungen mittels organischer Superoxyde. *Berichte der deutschen chemischen Gesellschaft* **42**, 4811-4815 (1909).
118. D. I. Metelitsa, Reaction Mechanisms of the Direct Epoxidation of Alkenes in the Liquid Phase. *Russian Chemical Reviews* **41**, 807-821 (1972).
119. R. W. Holman, Strategic Applications of Named Reactions in Organic Synthesis: Background and Detailed Mechanisms (Kürti, László; Czakó, Barbara). *Journal of Chemical Education* **82**, 1780 (2005).
120. D. A. Singleton, S. R. Merrigan, J. Liu, K. N. Houk, Experimental Geometry of the Epoxidation Transition State. *Journal of the American Chemical Society* **119**, 3385-3386 (1997).
121. E. N. Jacobsen, W. Zhang, A. R. Muci, J. R. Ecker, L. Deng, Highly enantioselective epoxidation catalysts derived from 1,2-diaminocyclohexane. *Journal of the American Chemical Society* **113**, 7063-7064 (1991).
122. T. Katsuki, K. B. Sharpless, The first practical method for asymmetric epoxidation. *Journal of the American Chemical Society* **102**, 5974-5976 (1980).
123. Z.-X. Wang, Y. Tu, M. Frohn, J.-R. Zhang, Y. Shi, An Efficient Catalytic Asymmetric Epoxidation Method. *Journal of the American Chemical Society* **119**, 11224-11235 (1997).
124. K. Matsumoto, K. Tomioka, Chiral ketone-catalyzed asymmetric epoxidation of olefins with Oxone®. *Tetrahedron Letters* **43**, 631-633 (2002).
125. A. Williamson, XLV. Theory of ætherification. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **37**, 350-356 (1850).
126. V. B. Urlacher, M. Girhard, Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol* **30**, 26-36 (2012).
127. S. G. Sligar, D. L. Cinti, G. G. Gibson, J. B. Schenkman, Spin state control of the hepatic cytochrome P450 redox potential. *Biochem Biophys Res Commun* **90**, 925-932 (1979).
128. F. Hannemann, A. Bichet, K. M. Ewen, R. Bernhardt, Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta* **1770**, 330-344 (2007).
129. G. A. Roberts, G. Grogan, A. Greter, S. L. Flitsch, N. J. Turner, Identification of a new class of cytochrome P450 from a Rhodococcus sp. *J Bacteriol* **184**, 3898-3908 (2002).
130. M. Wang *et al*., Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci U S A* **94**, 8411-8416 (1997).
131. D. Werck-Reichhart, R. Feyereisen, Cytochromes P450: a success story. *Genome Biology* **1**, reviews3003.3001 (2000).
132. D. Tischler, M. Schlömann, W. J. H. van Berkel, G. T. Gassner, FAD C(4a)-hydroxide stabilized in a naturally fused styrene monooxygenase. *FEBS Letters* **587**, 3848-3852 (2013).
133. Anonymous, Characterization of styrene catabolic pathway in Pseudomonas fluorescens ST. *International Biodeterioration & Biodegradation* **54**, 183 - 187 (2004).

134. A. Mooney, P. G. Ward, K. E. O'Connor, Microbial degradation of styrene: biochemistry, molecular genetics, and perspectives for biotechnological applications. *Appl Microbiol Biotechnol* **72**, 1 (2006).
135. A. Kantz, F. Chin, N. Nallamothu, T. Nguyen, G. T. Gassner, Mechanism of flavin transfer and oxygen activation by the two-component flavoenzyme styrene monooxygenase. *Arch Biochem Biophys* **442**, 102-116 (2005).
136. K. Otto, K. Hofstetter, M. Röthlisberger, B. Witholt, A. Schmid, Biochemical characterization of StyAB from Pseudomonas sp. strain VLB120 as a two-component flavin-diffusible monooxygenase. *J Bacteriol* **186**, 5292-5302 (2004).
137. A. Kantz, G. T. Gassner, Nature of the reaction intermediates in the flavin adenine dinucleotide-dependent epoxidation mechanism of styrene monooxygenase. *Biochemistry* **50**, 523-532 (2011).
138. U. E. Ukaegbu, A. Kantz, M. Beaton, G. T. Gassner, A. C. Rosenzweig, Structure and ligand binding properties of the epoxidase component of styrene monooxygenase. *Biochemistry* **49**, 1678-1688 (2010).
139. M. E. Yurkovich *et al*., A late-stage intermediate in salinomycin biosynthesis is revealed by specific mutation in the biosynthetic gene cluster. *Chembiochem* **13**, 66-71 (2012).
140. Y. Demydchuk *et al*., Analysis of the tetronomycin gene cluster: insights into the biosynthesis of a polyether tetronate antibiotic. *Chembiochem* **9**, 1136-1145 (2008).
141. B. M. Harvey *et al*., Insights into Polyether Biosynthesis from Analysis of the Nigericin Biosynthetic Gene Cluster in Streptomyces sp. DSM4137. *Chemistry & Biology* **14**, 703-714 (2007).
142. Y. Sun *et al*., A complete gene cluster from Streptomyces nanchangensis NS3226 encoding biosynthesis of the polyether ionophore nanchangmycin. *Chem Biol* **10**, 431-441 (2003).
143. L. Zheng, U. Baumann, J. L. Reymond, An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115 (2004).
144. C. Vonrhein *et al*., Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302 (2011).
145. P. Evans, A. McCoy, An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64**, 1-10 (2008).
146. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-25 (2010).
147. P. D. Adams *et al*., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
148. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
149. W. Kabsch, XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
150. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485 (2010).
151. K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011 (2006).
152. A. W. Schüttelkopf, D. M. van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-1363 (2004).

153. A. Dereeper *et al*., Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-469 (2008).
154. A. Dereeper, S. Audic, J. M. Claverie, G. Blanc, BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* **10**, 8 (2010).
155. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
156. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-552 (2000).
157. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
158. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**, 539-552 (2006).
159. F. Chevenet, C. Brun, A. L. Bañuls, B. Jacq, R. Christen, TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
160. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242-245 (2016).
161. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256-W259 (2019).
162. E. F. Pettersen *et al*., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
163. J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680 (1994).
164. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).
165. A. Fiser, A. Sali, Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**, 461-491 (2003).
166. D. W. . (2002) The PyMOL molecular graphics system .
167. C. Dominguez, R. Boelens, A. M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-1737 (2003).
168. G. C. P. van Zundert *et al*., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725 (2016).
169. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38 (1996).
170. R. B. Best *et al*., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2012).
171. K. Vanommeslaeghe *et al*., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671-690 (2010).
172. K. Vanommeslaeghe, A. D. MacKerell, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling* **52**, 3144-3154 (2012).

173. W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell Jr., Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* **33**, 2451-2468 (2012).
174. I. Soteras Gutiérrez *et al*., Parametrization of halogen bonds in the CHARMM general force field: Improved treatment of ligand–protein interactions. *Bioorganic & medicinal chemistry* **24**, 4812-4825 (2016).
175. J. C. Phillips *et al*., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802 (2005).
176. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993).
177. D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics* **23**, 566-579 (1995).
178. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291**, 1790-1792 (2001).
179. R. K. Wierenga, P. Terpstra, W. G. Hol, Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol* **187**, 101-107 (1986).
180. G. Eggink, H. Engel, G. Vriend, P. Terpstra, B. Witholt, Rubredoxin reductase of Pseudomonas oleovorans: Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *Journal of Molecular Biology* **212**, 135-142 (1990).
181. H. Shoun, K. Arima, T. Beppu, Inhibition of p-hydroxybenzoate hydroxylase by anions: possible existence of two anion-binding sites in the site for reduced nicotinamide adenine dinucleotide phosphate. *J Biochem* **93**, 169-176 (1983).
182. W. J. Van Berkel, W. J. Van Den Tweel, Purification and characterisation of 3-hydroxyphenylacetate 6-hydroxylase: a novel FAD-dependent monooxygenase from a Flavobacterium species. *Eur J Biochem* **201**, 585-592 (1991).
183. A. McPherson, J. A. Gavira, Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun* **70**, 2-20 (2014).
184. C. Haas, J. Drenth, Understanding protein crystallization on the basis of the phase diagram. *Journal of Crystal Growth* **196**, 388-394 (1999).
185. R. Piazza, Interactions in protein solutions near crystallisation: a colloid physics approach1This work is dedicated to Franz Rosenberger, who has been to me a valuable guide in the field of protein crystallisation, on the occasion of his (hopefully just official) retirement.1. *Journal of Crystal Growth* **196**, 415-423 (1999).
186. A. J. Malkin, Kuznetsov YuG, T. A. Land, J. J. DeYoreo, A. McPherson, Mechanisms of growth for protein and virus crystals. *Nat Struct Biol* **2**, 956-959 (1995).
187. Anonymous, \it In situ X-ray crystallography. *Journal of Applied Crystallography* **33**, 397--400 (2000).
188. Anonymous, A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *Journal of Applied Crystallography* **30**, 198--202 (1997).
189. A. A. Vagin *et al*., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195 (2004).
190. V. B. Chen *et al*., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

191. L. Holm, Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326-5327 (2019).
192. A. El Gamal *et al*., Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair. *Proc Natl Acad Sci U S A* **113**, 3797-3802 (2016).
193. A. K. Padyana *et al*., Structure and inhibition mechanism of the catalytic domain of human squalene epoxidase. *Nat Commun* **10**, 97 (2019).
194. H. A. Schreuder *et al*., Crystal structure of the p-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 A resolution. Analysis of the enzyme-substrate and enzyme-product complexes. *J Mol Biol* **208**, 679-696 (1989).
195. B. A. Palfey, D. P. Ballou, V. Massey, Flavin conformational changes in the catalytic cycle of p-hydroxybenzoate hydroxylase substituted with 6-azido- and 6-aminoflavin adenine dinucleotide. *Biochemistry* **36**, 15713-15723 (1997).
196. J. W. Setser, J. R. Heemstra, C. T. Walsh, C. L. Drennan, Crystallographic evidence of drastic conformational changes in the active site of a flavin-dependent N-hydroxylase. *Biochemistry* **53**, 6063-6077 (2014).
197. K. McLuskey, S. Cameron, F. Hammerschmidt, W. N. Hunter, Structure and reactivity of hydroxypropylphosphonic acid epoxidase in fosfomycin biosynthesis by a cation- and flavin-dependent mechanism. *Proc Natl Acad Sci U S A* **102**, 14221-14226 (2005).
198. L. C. Montemiglio *et al*., Functional analysis and crystallographic structure of clotrimazole bound OleP, a cytochrome P450 epoxidase from Streptomyces antibioticus involved in oleandomycin biosynthesis. *Biochim Biophys Acta* **1860**, 465-475 (2016).
199. P. Olivares, E. C. Ulrich, J. R. Chekan, W. A. van der Donk, S. K. Nair, Characterization of Two Late-Stage Enzymes Involved in Fosfomycin Biosynthesis in Pseudomonads. *ACS Chem Biol* **12**, 456-463 (2017).
200. L. J. Higgins, F. Yan, P. Liu, H. W. Liu, C. L. Drennan, Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature* **437**, 838-844 (2005).
201. E. Morrison, A. Kantz, G. T. Gassner, M. H. Sazinsky, Structure and mechanism of styrene monooxygenase reductase: new insight into the FAD-transfer reaction. *Biochemistry* **52**, 6063-6075 (2013).
202. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302-2309 (2005).
203. G. Spyrou *et al*., Characterization of the flavin reductase gene (fre) of Escherichia coli and construction of a plasmid for overproduction of the enzyme. *J Bacteriol* **173**, 3673-3679 (1991).
204. M. Ingelman, S. Ramaswamy, V. Nivière, M. Fontecave, H. Eklund, Crystal structure of NAD(P)H:flavin oxidoreductase from Escherichia coli. *Biochemistry* **38**, 7040-7049 (1999).
205. K. A. Lampel, B. Uratani, G. R. Chaudhry, R. F. Ramaley, S. Rudikoff, Characterization of the developmentally regulated Bacillus subtilis glucose dehydrogenase gene. *J Bacteriol* **166**, 238-243 (1986).
206. S. K. Piasecki *et al*., Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem Biol* **18**, 1331-1340 (2011).
207. C. Nguyen *et al*., Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427-431 (2014).

208. V. Agarwal, S. Lin, T. Lukk, S. K. Nair, J. E. Cronan, Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17406-17411 (2012).
209. J. E. Guy *et al*., Remote control of regioselectivity in acyl-acyl carrier protein-desaturases. *Proc Natl Acad Sci U S A* **108**, 16594-16599 (2011).
210. M. Babu *et al*., Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism. *Structure* **18**, 1450-1462 (2010).
211. K. D. Parris *et al*., Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883-895 (2000).
212. M. Su, X. Zhu, W. Zhang, Probing the Acyl Carrier Protein-Enzyme Interactions within Terminal Alkyne Biosynthetic Machinery. *AIChE J* **64**, 4255-4262 (2018).
213. B. M. Harvey *et al*., Insights into polyether biosynthesis from analysis of the nigericin biosynthetic gene cluster in Streptomyces sp. DSM4137. *Chem Biol* **14**, 703-714 (2007).
214. Y. Demydchuk *et al*., Analysis of the Tetronomycin Gene Cluster: Insights into the Biosynthesis of a Polyether Tetronate Antibiotic. *ChemBioChem* **9**, 1136-1145 (2008).
215. S. Eswaramoorthy, J. B. Bonanno, S. K. Burley, S. Swaminathan, Mechanism of action of a flavin-containing monooxygenase. *Proceedings of the National Academy of Sciences* **103**, 9832 (2006).
216. P. Chaiyen, M. W. Fraaije, A. Mattevi, The enigmatic reaction of flavins with oxygen. *Trends Biochem Sci* **37**, 373-380 (2012).
217. Z. A. Hughes-Thomas, C. B. W. Stark, I. U. Böhm, J. Staunton, P. F. Leadlay, Intermediates Released from a Polyether-Producing Polyketide Synthase Provide Insight into the Mechanism of Oxidative Cyclization. *Angewandte Chemie International Edition* **42**, 4475-4478 (2003).
218. A. Bhatt *et al*., Accumulation of an E,E,E-Triene by the Monensin-Producing Polyketide Synthase when Oxidative Cyclization is Blocked. *Angewandte Chemie International Edition* **44**, 7075-7078 (2005).
219. N. Eswar *et al*., Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics* **Chapter 5**, Unit-5.6 (2006).
220. G. C. P. van Zundert *et al*., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of molecular biology* **428**, 720-725 (2016).
221. B. K. Shoichet, W. A. Baase, R. Kuroki, B. W. Matthews, A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences* **92**, 452 (1995).
222. A. Seifert, S. Tatzel, R. D. Schmid, J. Pleiss, Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins: Structure, Function, and Bioinformatics* **64**, 147-155 (2006).
223. G. Cui, X. Li, K. M. Merz, Understanding the Substrate Selectivity and the Product Regioselectivity of Orf2-Catalyzed Aromatic Prenylations. *Biochemistry* **46**, 1303-1311 (2007).

224. M. Kujawa *et al*., Structural basis for substrate binding and regioselective oxidation of monosaccharides at C3 by pyranose 2-oxidase. *The Journal of biological chemistry* **281**, 35104-35115 (2006).

Abbreviation and Notations

PKS	polyketide synthase
ACP	acyl carrier protein
AT	acyltransferase
KS	ketosynthase
KR	ketoreductase
DH	dehydratase
ER	enoylreductase
TE	thioesterase
CoA	coenzyme A
A-1,3	allylic 1,3 strain
CCW model	biosynthetic model proposed by Cane, Celmer, and Westly in 1983
APPA	acetate, propionate, propionate, acetate
PABA	propionate, acetate, butyrate, acetate
BABA	butyrate, acetate, butyrate, acetate
PAPA	propionate, acetate, propionate, acetate
FAD	flavin adenine dinucleotide
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
FMN	flavin mononucleotide
m-CPBA	*meta*-chloroperoxybenzoic
E.coli	*Escherichia coli*
PCR	polymerase chain reaction
dNTP	deoxynucleoside triphosphate
DNA	deoxyribonucleic acid

DMSO	dimethylsulfoxide
IPTG	isopropyl-β-D-galactoside
LB medium	lysogeny broth medium
OD	optical density
IMAC	immobilized metal affinity chromatography
IEX	ion exchange chromatography
SEC	size exclusion chromatography
SDS PAG	sodium dodecyl sulfate polyacrylamide gel electrophoresis
PMSF	phenylmethylsulfonyl fluoride
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
DTT	(2*S*,3*S*)-1,4-bis(sulfanyl)butane-2,3-diol
GC-MS	gas chromatography mass spectrometry
LC-MS	liquid chromatography-mass spectrometry
Bis Tris	2[bis(2-hydroxyethyl)amino]2-2(hydroxymethyl)propane-1,3-diol
m/z	mass-to-charge ratio
HPLC	high performance liquid chromatography
MOPS	3-morpholinopropane-1-sulfonic acid
PEG	polyethylene glycol
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
Bicine	2-(bis(2-hydroxyethyl)acetic acid
MAD	multi-wavelength anomalous diffraction
SAD	single-wavelength anomalous diffraction
PDB	protein data bank
EDTA	2,2′,2″,2‴-(ethane-1,2-diyldinitrilo)tetraacetic acid
pI value	isoelectric point
pKa value	negative log of the acid dissociation constant
2D	two dimension

3D	three dimension
V_o	column void volume
V_e	elution volume of the protein molecules
V_t	total bed volume of columns
R_{work}	the measure of the quality of the atomic model obtained from the crystallographic data.
R_{free}	calculated with 5% of the data
RMSD	root-mean-square deviation of atomic positions

Appendix

Appendix 1. Codon-optimized gene sequence of *monCI*

5'-
ATGACCACCACCCGTCCGGCGCACGCGGTTGTTCTGGGCGCGAGCATGGCGGGCAC
CCTGGCGGCGCACGTTCTGGCGCGTCACGTTGACGCGGTTACCGTGGTTGAGCGTGA
TGCGCTGCCGGAGGAACCGCAGCACCGTAAGGGTGTGCCGCAAGCGCGTCATGCGC
ACCTGCTGTGGAGCAACGGTGCGCGTCTGATCGAGGAAATGCTGCCGGGTACCACC
GACCGTCTGCTGGCGGCGGGTGCGCGTCGTCTGGGTTTCCCGGAAGATCTGGTTACC
CTGACCGGCCAGGGTTGGCAACACCGTTTCCCGGCGACCCAGTTTGCGCTGGTTGCG
AGCCGTCCGCTGCTGGATCTGACCGTGCGTCAGCAAGCGCTGGGTGCGGATAACAT
CACCGTTCGTCAACGTACCGAAGCGGTGGAACTGACCGGTAGCGGTGGCGGTAGCG
GCGGTCGTGTTACCGGCGTGGTTGTGCGTGACCTGGATAGCGGTCGTCAGGAGCAAC
TGGAAGCGGACCTGGTGATTGATGCGACCGGCCGTGGTAGCCGTCTGAAGCAGTGG
CTGGCGGCGCTGGGCGTTCCGGCGCTGGAGGAAGACGTTGTGGATGCGGGTGTGGC
GTACGCGACCCGTCTGTTCAAAGCTCCGCCGGGTGCGACCACCCACTTTCCGGCGGT
TAACATTGCGGCGGATGATCGTGTGCGTGAGCCGGGCCGTTTCGGTGTTGTGTATCC
GATTGAAGGCGGTCGTTGGCTGGCGACCCTGAGCTGCACCCGTGGTGCGCAGCTGC
CGACCCATGAGGATGAGTTCATCCCGTTTGCGGAGAACCTGAACCACCCGATTCTGG
CGGACCTGCTGCGTGATGCGGAACCGCTGACCCCGGTTTTTGGCAGCCGTAGCGGTG
CGAACCGTCGTCTGTACCCGGAGCGTCTGGAACAATGGCCGGACGGCCTGCTGGTG
ATCGGTGATAGCCTGACCGCGTTCAACCCGATTTATGGTCATGGTATGAGCAGCGCG
GCGCGTTGCGCGACCACCATCGACCGTGAGTTTGAACGTAGCGTTCAGGAAGGTAC
CGGTAGCGCGCGTGCGGGTACCCGTGCGCTGCAAAAGGCGATTGGTGCGGCGGTGG
ATGATCCGTGGATCCTGGCGGCGACCAAAGACATTGATTACGTTAACTGCCGTGTGA
GCGCGACCGACCCGCGTCTGATTGGTGTGGATACCGAGCAGCGTCTGCGTTTTGCGG
AAGCGATCACCGCGGCGAGCATTCGTAGCCCGAAAGCGAGCGAGATTGTTACCGAC
GTGATGAGCCTGAACGCGCCGCAAGCGGAACTGGGTAGCAACCGTTTCCTGATGGC
GATGCGTGCGGATGAACGTCTGCCGGAACTGACCGCGCCGCCGTTTCTGCCGGAGG
AACTGGCGGTTGTGGGTCTGGATGCGGCGACCATCAGCCCGACCCCGACCCCGACC
CCGACCGCGGCGGTGCGTAGCTAAGAATTC-3'

Appendix 2. Codon-optimized gene sequence of *lsd18*

5'-
CATATGACGAACACGCGCTCGGCGGTTGTCCTGGGTGGTGGTATGGCTGGTATGCTG
GTCTCATCAATGCTGGCTCGTCACGTTGGCTCAGTGACCGTTATTGATCGTGACGCA
TTTCCGGCAGGTCCGGATCTGCGCAAAGGTGTCCCGCAAGCTCGTCATGCGCACATT
CTGTGGTCTGGCGGTGCGCGTATCGTTGAAGAACTGCTGCCGGGTACCA
CGGATCGTCTGCTGGGTGCAGGTGCTCATCGTATTGGCATCCCGGACGGTCAGGTGT
CTTATACCGCTTACGGTTGGCAGCACCGCTTTCCGGAAGCACAATTCATGATTGCCT
GCTCCCGCGCACTGCTGGATTGGACCGTTCGTGAAGAAACGCTGCGTGAAGAACGC
ATCGCCCTGGTCGAAAAAACCGAAGTGCTGGCTCTGCTGGGTGATGCAGGTCGTGTG
ACGGGCGTCCGTGTGCGTGACCAGGAAAGCGGTGAAGAACGTGAAGTTCCGGCGGA
TCTGGTGGTTGACACCACGGGTCGCGGTAGCCCGTCTAAACGTCTGCTGGCAGAACT
GGGTCTGCCGGCACCGGAAGAAGAATTTGTTGATAGCGGTATGGTCTATGCTACCCG

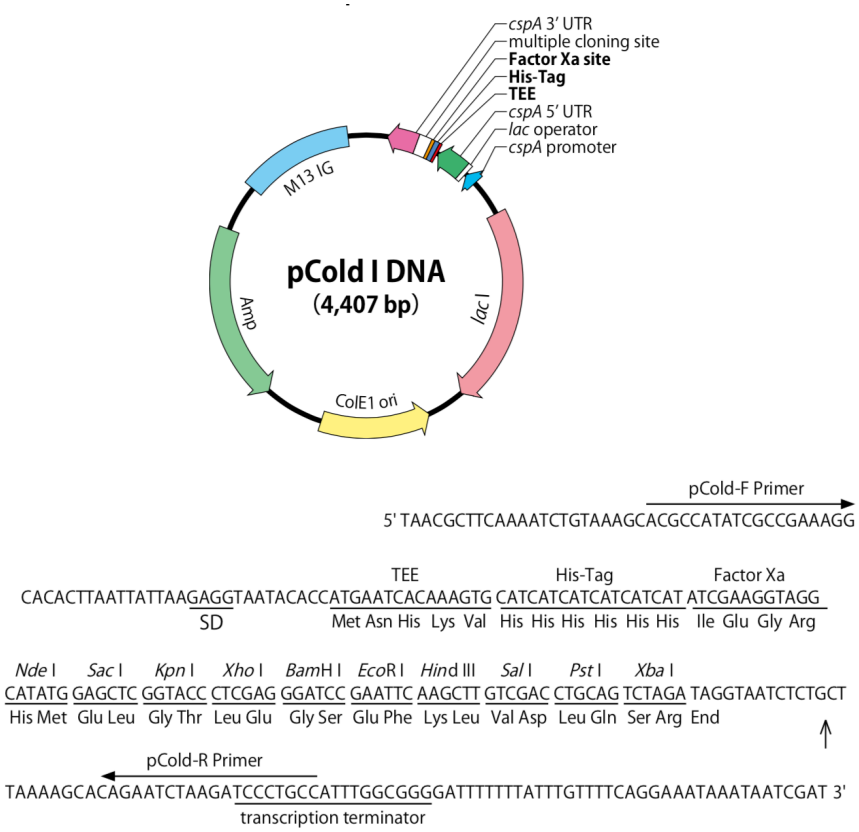
TCTGTTTCGTGCACCGGAAGCAGCAGCAACGAACTTCCCGCTGGTTTCGGTCCATGC
TGATCACCGTGCAGGTCGTCCGGGTTGTAATGCAGTGCTGATGCCGATTGAAGACGG
CCGCTGGATCGTGACCGTTAGTGGTACGCGTGGCGGTGAACCGCCGGCAGATGACG
AAGGTTTTGCCCGTTTCGCACGTGATGGTGTGCGTCATCCGCTGGTTGGTGAACTGA
TCGCGAAAGCCCAGCCGCTGACCAGCGTTGAACGTAGTCGCTCCACGGTCAACCGT
CGCCTGCACTATGATCGTCTGGCAACCTGGCCGGAAGGCCTGGTCGTGCTGGGTGAC
GCAGTCGCTGCGTTTAATCCGGTGTACGGCCATGGCATGTCAGCAGCAGCTCACTCG
GTGCTGGCACTGCGTAGCCAGCTGGGTCAACGTGCATTCCAGCCGGGTCTGGCACGT
GCAGCACAACGTGCAATTGCTGTCGCGGTGGATGACGCCTGGGTGCTGGCAACCTCT
CATGATATCGGTTACCCGGGTTGCCGTACCCAGACGCGTGACCCGCGCCTGACCCGT
CACGCAGGCGAACGTCAACGCGTTACGGATCTGGTCGGTCTGACCGCAACGCGCAA
CCAGGTTGTCAATC
GTGCAGCTGTGGCGCTGAACACCCTGAGCGCTGGCATGGCGTCTATGCAGGATCCG
GCAGTGATGGCAGCAGTTCGTCGCGGTCCGGAAGTTCCGGCACCGACCGAACCGCC
GCTGCGCCCGGATGAAGTCGCTCGTCTGGTCTCTGGTGCTGGTGTTACCGCTTAAGA
ATTC-3’

Appendix 3. Codon-optimized gene sequence of *fre*

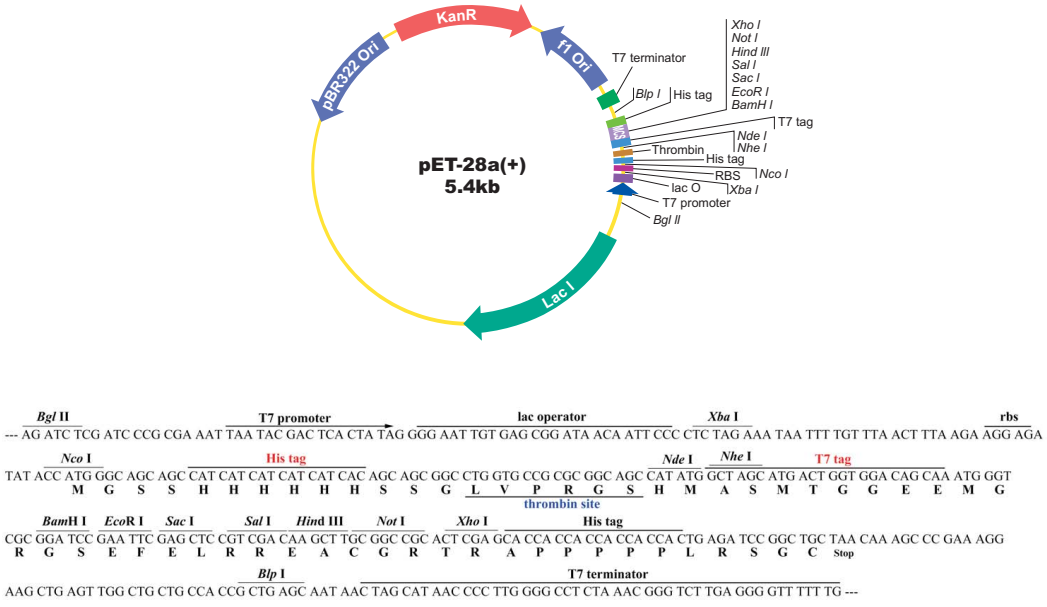
5’-
CATATGACAACCTTAAGCTGTAAAGTGACCTCGGTAGAAGCTATCACGGATACCGTA
TATCGTGTCCGCATCGTGCCAGACGCGGCCTTTTCTTTTCGTGCTGGTCAGTATTTGA
TGGTAGTGATGGATGAGCGCGACAAACGTCCGTTCTCAATGGCTTCGACGCCGGATG
AAAAAGGGTTTATCGAGCTGCATATTGGCGCTTCTGAAATCAACCTTTACGCGAAAG
CAGTCATGGACCGCATCCTCAAAGATCATCAAATCGTGGTCGACATTCCCCACGGAG
AAGCGTGGCTGCGCGATGATGAAGAGCGTCCGATGATTTTGATTGCGGGCGGCACC
GGGTTCTCTTATGCCCGCTCGATTTTGCTGACAGCGTTGGCGCGTAACCCAAACCGT
GATATCACCATTTACTGGGGCGGGCGTGAAGAGCAGCATCTGTATGATCTCTGCGAG
CTTGAGGCGCTTTCGTTGAAGCATCCTGGTCTGCAAGTGGTGCCGGTGGTTGAACAA
CCGGAAGCGGGCTGGCGTGGGCGTACTGGCACCGTGTTAACGGCGGTATTGCAGGA
TCACGGTACGCTGGCAGAGCATGATATCTATATTGCCGGACGTTTTGAGATGGCGAA
AATTGCCCGCGATCTGTTTTGCAGTGAGCGTAATGCGCGGGAAGATCGCCTGTTTGG
CGATGCGTTTGCATTTATCTGAGAATTC-3’

Appendix 3. Codon-optimized gene sequence of *monACPX*.

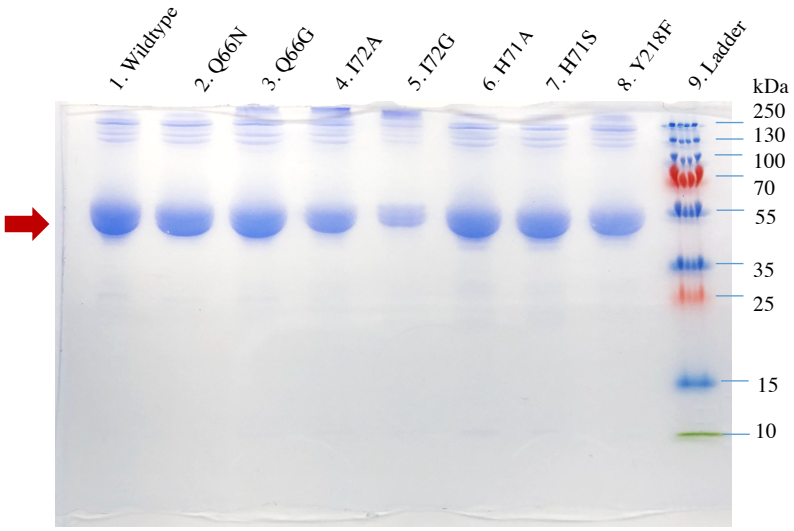
5’-
CATATGACCAGCACCGACCACACCAGCGGTCAGGATGCGACCGAGCTGGAAAAACA
GCTGGCGGCGGCGACCCCGGAGGAACGTGAGAAACTGCTGACCGACACCATCCGTA
CCCAGGCGGGTACCCTGCTGAACACCACCCTGAGCGACGATAGCAACTTCCTGGAG
AACGGCCTGAACAGCCTGACCGCGCTGGAACTGACCAAGACCCTGATGACCCTGAC
CGGTATGGAGATCGCGATGGTGGCGATTGTTGAAAACCCGACCCCGGCGCAGCTGG
CGCACCACCTGGGTCAAGAACTGGCGCACACCACCGCGTAAGAATTCGATATC-3’



Appendix 4. The vector map of pColdI that was purchased from TakaRa. The genes of *monCI*, *lsd18*, and *lsd19* were inserted at the NdeI and EcoRI cleavage sites.



Appendix 5. The vector map of pET28a(+). The genes of *fre* and *monACPX* were inserted into the cleavage sites NdeI and EcoRI in the pET28(+).



Appendix 6. SDS PAGE analysis of Lsd18 mutants. From lane 1 to lane 9: Lsd18 wildtype, Lsd18 mutants Q66N, Q66G, I72A, I72G, H71A, H71S, Y218F, and protein ladder. The lsd18 and its mutants are marked by the red arrow.

Mutant	Y218F	H71A	H71S	Q66N	Q66G	I72A	I72G
Estimated yield %	89.53	53.28	55.33	18.05	9.6	52.8	66.0

Appendix 7: Estimated percent yield of **25** catalyzed by Lsd18 mutants

Vita

Qian Wang was born in Yangzhou, China in 1992. She graduated from Xiamen University in July of 2014 with a Bachelor of Science degree in Biology science. In August of 2014, she enrolled in the National University of Singapore as a Ph.D. student in the department of Biology. In January of 2016, she transferred her Ph.D. degree to the Chemistry and Biochemistry department in the University of Texas at El Paso to continue her research with Dr. Chu-Young Kim. While pursuing the Ph.D. degree in Biochemistry, Qian was also a Teaching Assistant during the years of 2015-2019. She has assisted with the general chemistry lectures and taught organic chemistry and biochemistry laboratories.

Contact information: qwang@miners.utep.edu

This dissertation was typed by Qian Wang.