

# Biotechnology Advances

## Overview of Flavin-containing Monooxygenase: History, Structures, Mechanism, Biosynthesis and Disease --Manuscript Draft--

<b>Manuscript Number:</b>	JBA-D-23-00665
<b>Article Type:</b>	Review Article
<b>Keywords:</b>	Flavin-containing Monooxygenase, Synthetic Biology, Catalytic Mechanism, Indigo and indirubin biosynthesis, Bio-indigo
<b>Corresponding Author:</b>	Yunjun Yan  Wuhan, Hubei CHINA
<b>First Author:</b>	Changxin Fan
<b>Order of Authors:</b>	Changxin Fan Ziqi Xie Yijin Li Ruihan Zhang Da Zheng Jiacheng Shi Yifei Wang Mingyuan Cheng Yu Zhou Yi Zhan Yunjun Yan
<b>Abstract:</b>	<p>The Flavin-containing Monooxygenases (FMOs) are a significant protein family that is widely distributed in bacteria, yeasts, plants, and mammals. FMOs have been demonstrated to play a crucial functional role in drug metabolism and indigo biosynthesis. The research history of FMOs was reviewed. The structural analysis of FMOs was conducted to point out the catalytic mechanism, which can aid in comprehending their functions more profoundly. The catalytic example of indole was exhibited in the docking simulation. The differences across FMOs were comprehensible among sequence alignments. The potential improvements of FMOs were provided through de novo design, directed evolution, enzyme immobilization, metabolic engineering, and fermentation engineering, which provides additional insights into indigo and indirubin biosynthesis via FMOs.</p>
<b>Suggested Reviewers:</b>	<p>Pablo Sobrado, PhD Professor, Virginia Polytechnic Institute and State University University Bookstore psobrado@vt.edu Their group studied the mechanism of mFMO and found that tryptophan-47 is important.</p> <p>Andrea Mattevi, PhD Professor, University of Pavia mattevi@ipvgen.unipv.it They solved the structure of mFMO.</p> <p>Si Wouk Kim, PhD Professor, Chosun University swkim@chosun.ac.kr The first group applied mFMO in indigo biosynthesis.</p> <p>Marco W. Fraaije, PhD Professor, University of Groningen</p>

	<p>m.w.fraaije@rug.nl They achieved the highest yield of indigo biosynthesis so far.</p>
	<p>Jin Ho Lee, PhD Professor, Kyung Sung University jhlee83@ks.ac.kr They cloned cFMO from <i>Corynebacterium glutamicum</i> to improve the yield of indigo and indirubin.</p>
	<p>Gui Hwan Han, PhD Professor, Center for Industrialization of Agricultural and Livestock Microorganisms ghhan@cialm.or.kr Great works in indigo biosynthesis</p>
	<p>Yung Hun Yang, PhD Professor, Konkuk University seokor@konkuk.ac.kr This group recently utilized some new strategies to improve the yield of bio-indigo.</p>
	<p>Byung Gee Kim, PhD Professor, Seoul National University byungkim@snu.ac.kr FMO was used to synthesize tyrian purple. A consecutive two-cell reaction system was used.</p>
	<p>Feng-Qing Wang, PhD Professor, East China University of Science and Technology fqwang@ecust.edu.cn They engineered bFMO for the efficient production of indirubin</p>

## COVER LETTER

Yunjun Yan

College of Life Science and Technology, HUST  
1037 Luoyu Road, Hongshan District, Wuhan 430074, P.R. China.

### Submission date:

September 20, 2023

Edward A. Bayer

Weizmann Institute of Science, Department of Biomolecular Sciences, Rehovot, Israel

Dear Prof. Edward A. Bayer,

We are pleased to submit our manuscript entitled “Overview of Flavin-containing Monooxygenase: History, Structures, Mechanism, Biosynthesis and Disease” for full consideration as an original article to the Journal “*Biotechnology Advances*”. This review totally investigated the research history, structures, mechanisms, and potential improvements of flavin-containing monooxygenase (FMO), specially focusing on how to improve the biosynthesis of indole and indirubin. We know the aims and scope of *Biotechnology Advances* are on current developments and future trends in biotechnology. Our review exhibits the potential application of FMOs in the industry and includes the research history and future developing trends in improving the biosynthesis of indigo and indirubin via FMOs from various aspects. We believe this article will be of special interest to the readers of *Biotechnology Advances*.

This manuscript has not been previously published and is not under consideration in the same or substantially similar form in any other peer-reviewed media.

### AUTHOR INFORMATION

#### Corresponding Authors

**Yunjun Yan** - Key Laboratory of Molecular Biophysics of the Ministry of Education,

Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [yanyunjun@hust.edu.cn](mailto:yanyunjun@hust.edu.cn)

**Yi Zhan** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [zhanyi@hust.edu.cn](mailto:zhanyi@hust.edu.cn)

**Yu Zhou** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [zhoyu@hust.edu.cn](mailto:zhoyu@hust.edu.cn)

## Authors

**Changxin Fan** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

orcid.org/0000-0002-0435-6202; Email: [cfanai@connect.ust.hk](mailto:cfanai@connect.ust.hk)

**Ziqi Xie** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

orcid.org/0000-0001-8435-1948; Email: [ziqi\\_xie@hust.edu.cn](mailto:ziqi_xie@hust.edu.cn)

**Yijin Li** - Innovation Base of Life Science and Technology, Qiming College, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [u202013531@hust.edu.cn](mailto:u202013531@hust.edu.cn)

**Ruihan Zhang** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [peterico3395@163.com](mailto:peterico3395@163.com)

**Da Zheng** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [zhengnda@outlook.com](mailto:zhengnda@outlook.com)

**Jiacheng Shi** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [shijiachengg@163.com](mailto:shijiachengg@163.com)

**Yifei Wang** - Innovation Base of Life Science and Technology, Qiming College, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [wang.ef@foxmail.com](mailto:wang.ef@foxmail.com)

**Mingyuan Cheng** - College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

Email: [chengmingyuann@163.com](mailto:chengmingyuann@163.com)

### **CRedit authorship contribution statement**

**Changxin Fan**: Conceptualization, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Ziqi Xie**: Conceptualization, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Yijin Li**: Investigation, Writing - Original Draft, Visualization. **Ruihan Zhang**: Investigation, Writing - Original Draft. **Da Zheng**: Formal analysis, Investigation, Writing - Original Draft, Visualization. **Jiacheng Shi**: Investigation, Writing - Original Draft. **Yifei Wang**: Investigation, Writing - Original Draft. **Mingyuan Cheng**: Visualization. **Yu Zhou**: Supervision. **Yi Zhan**: Supervision, Project administration, Writing - Original Draft. **Yunjun Yan**: Project administration, Writing - Review & Editing, Funding acquisition.

### **Conflict of interest**

The authors declare no financial or commercial conflict of interest.

I appreciate your consideration of this manuscript.

Sincerely,

Yunjun Yan

**Declaration of interests**

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

## Highlights

- Flavin-containing monooxygenases (FMOs) have been involved in indigo biosynthesis for over 20 years.
- The relationship between structure and catalytic mechanism of flavin-containing monooxygenases is elucidated.
- Directed evolution, *de novo* design, and genome-scale metabolic network models (GEMs) are suggested to improve indigo and indirubin biosynthesis.
- Trimethylaminuria (TMAU) is a genetic disorder caused by a defect in hFMO3.

# Overview of Flavin-containing Monooxygenase: History, Structures, Mechanism, Biosynthesis and Disease

Changxin Fan<sup>a,b,1,2</sup>, Ziqi Xie<sup>a,b,2</sup>, Yijin Li<sup>b,2</sup>, Ruihan Zhang<sup>a,b,2</sup>, Da Zheng<sup>a,b,2</sup>, Jiacheng Shi<sup>a,b</sup>, Yifei Wang<sup>b</sup>, Mingyuan Cheng<sup>a,b</sup>, Yu Zhou<sup>a,c,\*</sup>, Yi Zhan<sup>a,c,\*</sup>, Yunjun Yan<sup>a,c,\*</sup>

<sup>a</sup> College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

<sup>b</sup> Innovation Base of Life Science and Technology, Qiming College, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

<sup>c</sup> Key Laboratory of Molecular Biophysics of the Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

---

<sup>1</sup> Present address: Division of Life Science, HKUST, Clear Water Bay, Kowloon, Hong Kong, P.R.China.

<sup>2</sup> These authors contributed equally to this work.

\* Corresponding author at: College of Life Sciences and Technology, Huazhong University of Science and Technology, 1037#, Luoyu Road, Hongshan District, Wuhan 430074, P.R. China.

E-mail address: [yanyunjun@hust.edu.cn](mailto:yanyunjun@hust.edu.cn) (Y. Yan).

[zhanyi@hust.edu.cn](mailto:zhanyi@hust.edu.cn) (Y. Zhan).

[zhoyu@hust.edu.cn](mailto:zhoyu@hust.edu.cn) (Y. Zhou).



## Abstract

The Flavin-containing Monooxygenases (FMOs) are a significant protein family that is widely distributed in bacteria, yeasts, plants, and mammals. FMOs have been demonstrated to play a crucial functional role in drug metabolism and indigo biosynthesis. The research history of FMOs was reviewed. The structural analysis of FMOs was conducted to point out the catalytic mechanism, which can aid in comprehending their functions more profoundly. The catalytic example of indole was exhibited in the docking simulation. The differences across FMOs were comprehensible among sequence alignments. The potential improvements of FMOs were provided through *de novo* design, directed evolution, enzyme immobilization, metabolic engineering, and fermentation engineering, which provides additional insights into indigo and indirubin biosynthesis via FMOs.

**Keywords:** Flavin-containing Monooxygenase, Synthetic Biology, Catalytic Mechanism, Indigo and indirubin biosynthesis, Bio-indigo

29	<b>Contents</b>	
30	Abstract.....	2
31	Contents.....	3
32	Introduction .....	5
33	1. The Research history of FMOs.....	8
34	1.1 The evolution and diversity of the FMO gene family.....	8
35	1.2 From early reports to recent advances in FMO-mediated synthesis.....	13
36	2. Structures and catalytic mechanisms .....	15
37	2.1 The structure and catalytic mechanism of mammalian FMOs, for instance,	
38	AncFMO3-6.....	15
39	2.2 The structure and catalytic mechanism of mFMO.....	16
40	3. The application of FMO in biosynthesis .....	20
41	3.1 The advantages of FMO among enzymes in indigo biosynthesis.....	21
42	3.2 FMO-catalyzed biosynthesis of indigo .....	22
43	3.3 FMO-catalyzed biosynthesis of indirubin.....	26
44	3.4 Strategies for improving FMO-catalyzed biosynthesis.....	28
45	3.4.1 Insights from FPMOs .....	28
46	3.4.2 Directed evolution .....	30
47	3.4.3 Semi-rational design of FMOs .....	32
48	3.4.4 <i>De novo</i> enzyme design with high substrate specificity .....	34
49	3.4.5 The immobilization of FMOs.....	44
50	3.4.6 Regenerating NADPH in FMO catalysis .....	45
51	3.4.7 Metabolic engineering.....	46
52	3.4.8 Fermentation engineering.....	47
53	4. The Role of FMO in health and disease .....	48
54	5. Conclusions .....	50
55	Abbreviations .....	51
56	CRedit authorship contribution statement.....	53
57	Acknowledgements .....	53

58	References .....	54
59		
60		

## Introduction

In 2006, flavoprotein monooxygenases (FPMOs) were classified into six classes according to sequence and structure data (van Berkel et al., 2006). Flavin-containing Monooxygenase (FMO, EC 1.14.13.8) belongs to class B, which is encoded by a single gene and contains a tightly bound FAD cofactor. FMO retains the coenzyme NADPH/NADP<sup>+</sup> bound during catalysis and is composed of two dinucleotide binding domains (Rossmann fold) that bind FAD and NADPH, respectively. Recently, it was discovered that FMO can be classified into two subtypes: type I and type II (Jensen et al., 2012; Riebel et al., 2013). Type I FMOs depend on NADPH as the coenzyme, which are discussed mainly in this paper. While type II FMOs, such as *Stenotrophomonas maltophilia* flavin-containing monooxygenase (SMFMO), which can perform Baeyer–Villiger oxidation and accept both NADPH and NADH as coenzymes, have not been extensively studied (Ceccoli et al., 2014).

Eukaryotic FMOs are widely present in the livers of *Oryctolagus cuniculus*, *Sus scrofa*, and *Homo sapiens* and are located at the endoplasmic reticulum (ER) of hepatocytes towards the cytoplasm. Similar to cytochrome P450 (CYP450), FMOs are involved in drug metabolism and mono-oxidation N and S of xenobiotic compounds. Human FMO is classified into five categories: hFMO1, hFMO2, hFMO3, hFMO4, and hFMO5. Studies on FMO1 in pig livers, the former hotspot in the related area, have revealed the catalytic mechanism and the universality of substrates of FMOs (Table 2)

(Poulsen and Ziegler, 1979; Ball and Bruce, 1980; Beaty and Ballou, 1981a, b; Ziegler, 1993). The expression levels of FMOs vary significantly across different species (Table 1). The species specificity of FMO is likely relevant to the susceptibility and secretion efficiency of toxins and xenobiotics. Furthermore, even within humans, the expression levels of hFMOs can vary depending on factors such as tissue, age, and sex (Zane et al., 2018). hFMO1 exists mainly in fetal livers, adult kidneys, and intestines, while hFMO3 is mainly expressed in adult kidneys. hFMO3, owing to its significance in disease, is currently the most extensively studied (Phillips and Shephard, 2020a). It was found that a close link between hFMO3 mutation and trimethylaminuria (TMAU) (Treacy et al., 1998).

Unlike mammals, yeast only has one FMO isoform, yFMO, which exerts its effect on protein folding (Suh et al., 1996; Suh et al., 1999). In plants, FMO is the rate-limiting enzyme in auxin biosynthesis, catalyzing oxidative decarboxylation of indole-3-pyruvate acid (IPyA) to form indole-3-acetic acid (IAA) (Cao et al., 2019). Bacterial FMO, also known as trimethylamine monooxygenase (TMM), besides oxidizing many compounds containing nitrogen and sulfur, is also able to oxidize indole, trimethylamine (TMA), dimethylsulfate (DMS), and dimethylsulfoxide (DMSO) (Chen et al., 2011; Choi et al., 2003). Through metagenomes, TMM was found to play a pivotal but often neglected role in the global carbon and nitrogen cycles (Chen et al., 2011). The first bacterial FMO, *Methylophaga* sp. strain SK1 (mFMO) (Choi et al.,

2003), was used in engineered *Escherichia coli* to produce indigo. Further research suggested that it also catalyzes indole derivatives into indigoid dyes (Rioz-Martinez et al., 2011). It is important to note the mechanism of FMOs in the synthesis of indigo. In bacteria, tryptophan is converted to indole by tryptophanase (TRP), and FMOs subsequently oxidize indole to 2-hydroxyindole or 3-hydroxyindole, which are oxidized to isatin by oxygen. Isatins then dimerize into indigo or indirubin (Fig. 5C).

Indigo, a natural dye commonly extracted from plants, has a history of over 6,000 years of traditional extraction techniques. Indigo is widely used in the textile industry for dyeing jeans and other fabrics. However, the production of around 80,000 tonnes of indigo each year involves the use of non-renewable petrochemicals and results in the generation of toxic compounds (Linke, J.A. et al., 2023). To address these issues, bio-indigo, which involves the use of synthetic biology to synthesize indigo from tryptophan or indole, has been proposed as a sustainable and eco-friendly alternative. Although several countries, such as China, America, and France, have companies that are making efforts to commercialize bio-indigo, the process has been hindered by low yields. FMOs, which have been shown to possess the ability to synthesize bio-indigo, are regarded as promising enzymes to overcome this limitation. Indirubin, an indole alkaloid used as a drug in chronic myeloid leukemia, has shown anticancer effects (Yang et al., 2022), and traditional production methods rely on plant cell culture extraction (Han et al., 2012). Biosynthetic processes are needed to improve production

121 efficiency and conditions (Berry et al., 2002; Choi et al., 2003; Cho et al., 2011).

122 Except biosynthesis, FMOs also play crucial roles in other aspects. The superoxide  
123 anion radical generated by FMOs is considered to regulate the overall redox state of  
124 cells (Krueger and Williams, 2005). In *Caenorhabditis elegans*, *fmo-2/FMO5* is  
125 regulated by NHR-49/PPAR- $\alpha$  during infection of *Staphylococcus aureus*, revealing that  
126 FMOs are critical innate immunity effectors in animals (Wani et al., 2021). In mice, the  
127 biological clock genes regulate FMO5 expression by transcription of cis-acting  
128 elements E-box and D-box (Chen et al., 2019), suggesting that hFMOs may play a more  
129 extensive role in the human body.

130 This paper provides a comprehensive review of the research progress on FMOs,  
131 particularly on mFMO. The structure and catalytic mechanism of mFMO are  
132 highlighted, and the advances in indigo and indirubin biosynthesis are discussed,  
133 providing a prospect for scientists to improve the yields through new strategies and  
134 potentially expand the application of mFMO.

135

## 136 **1. The research history of FMOs**

### 137 **1.1 The evolution and diversity of the FMO gene family**

138 In 1972, Dr. Ziegler isolated a mixed-function enzyme from the pig liver that catalyzes  
139 N-oxidation of amine (Fig. 1) (Ziegler and Mitchell, 1972). The enzyme can oxidize a  
140 wide variety of nitrogen-, sulfur-, and phosphorus-containing xenobiotics. It was named

Flavin-containing Monooxygenases, rather than being named restrictively mixed-function, amine oxidase, or simply N-oxidase (Ziegler, 2002).

-----**Fig 1**-----

In the 1980s, Tynes et al. identified differences between hepatic and pulmonary forms of microsomal FMO in mice and rabbits (Tynes et al., 1985). Since then, over 150 types of FMOs have been successfully isolated from various species (van Berkel et al., 2006). Currently, there are 11 genes encoding hFMOs in *H. sapiens* (*FMO1-FMO5*, *FMO6P-FMO11P*), of which only *FMO1-FMO5* encode catalytic proteins. FMO 1-5 share approximately 50%-58% amino acid identity across different species (Lawton et al., 1994; Yang, 2017).

The FMO gene family is conserved across all phyla that have been examined so far. Hence, some forms of the FMO gene family can be found in all studied eukaryotes. FMO genes are defined by specific structural and functional restrictions, which have prompted the evolution of different types of FMOs to fulfill various tasks. The functional subtypes of FMOs (FMO 1-5) began to diverge before the evolution of mammals and amphibians into distinct classes. FMO5, which was identified in vertebrates, is the first functionally unique member of the FMO family and appears to have evolved earlier than other forms of FMOs. According to phylogenetic studies, the



most recent FMOs to develop into enzymes with distinct functions are FMO1 and FMO3. FMOs in invertebrates have developed polyphyletically, which means that an invertebrate evolved a phenotypically similar gene that was not passed down from a shared ancestor (Hao et al., 2009).

FMO proteins share several conserved domains are shown in Figure 2, including the FAD binding domain (GxGxxG/A) (Burnett et al., 1994), the FMO protein recognition motif (FxGxxxHxxxY/F) (Alfieri et al., 2008), and the NADP<sup>+</sup> binding domain (GxSxxG/A) (Lawton et al., 1994). Leu375 is conserved in both hFMO and AncFMO (ancient mammalian FMO), and acts as a tunnel gatekeeper (Levin, 1992). Asn78 in mFMO is conserved in all FMOs (marked green in Fig. 5), and two additional oxygen molecules are located in its side chain group, contributing to the stability of oxygen molecules through polarity and affecting enzyme activity in various FMO proteins (Alfieri et al., 2008).

It is worth noting that while hFMO is an insoluble protein, mFMO is water-soluble and its FAD is exposed to the solvent, making it easier to bind to the matrix (Phillips and Shephard, 2020b). The homology between the two proteins is only 20.7% (Fig. 2B). Furthermore, as shown in Fig. 2B, cFMO, cloned from *Corynebacterium glutamicum*, has little homology with various FMO proteins found in bacteria and mammals.

-----**Fig 2**-----

**Table 1.** The expression difference of mammalian FMOs

Species	FMO subtypes	Tissue	Developmental stage	Sex difference	References
<i>Homo sapiens</i>	FMO1	-The primary subtype in adult kidney	-High expression in fetal, silent post parturition	-No significant gender difference	Dolphin et al., 1996; Hernandez et al., 2004; Zhang and Cashman, 2006
	FMO2	-The primary subtype in <i>adult</i> lung			
	FMO3	-The primary subtype in adult liver	-Low expression in fetal, gradually increases post parturition		
	FMO4	-Liver, kidney, lung, small intestine et al.			
	FMO5	-Liver, small intestine, kidney, lung et al.			
<i>Mus musculus</i>	FMO1	-Lung, kidney, brain	-Expression in fetal liver gradually increased to reach that of female mice levels post parturition; Expression in the brain is the most in neonatal	-Gender difference is noticeable 28 days after birth. The expression in female mice remains unchanged, while that of male mice is inhibited	Cherrington et al., 1998; Janmohamed et al., 2004
	FMO2	- Low expression		-No significant gender difference	
	FMO3	-The primary subtype in adult female mouse liver	-Fetal liver expression is low and reaches adult expression level 14 days post parturition	-Gender differences demonstrated after sexual maturity	
	FMO4	-Low expression		-No obvious	

				gender difference	
	FMO5	-Liver, kidney, small intestine	-Expression is detectable 17 days after gestation and reaches the expression of adult level 2 days post parturition	-Expression level in female mice was higher than that of male mice in liver and kidney	
<i>Rattus norvegicus</i>	FMO1	-Liver		-Higher expression level in males than female	Cherrington et al., 1998; Lattard et al., 2003
	FMO2	-Kidney		-No obvious gender difference	
	FMO3	-Liver (lower expression level than that of FMO1)			
	FMO4	-Low expression in kidney and brain			
	FMO5	-Liver			
<i>Cynomolgus macaque</i>	FMO1	-Kidney			Uno et al., 2013
	FMO2	-Lung, kidney, heart, jejunum			
	FMO3	-Liver, kidney, lung, jejunum			
	FMO4	-Kidney, liver, lung, jejunum			
	FMO5	-Liver, lung, jejunum			
	FMO6	-Low expression in all tissue			
<i>Oryctolagus cuniculus</i>	FMO1	-Liver, intestinal mucosa			Larsen-Su et al., 1999;
	FMO2	-Lung	-FMO2 levels in the fetal lung are high at late gestation (except for day 28). After parturition, FMO2 levels fall dramatically,		Shehin-Johnson et al., 1995

			followed by		
			significant recovery by		
			21 days after birth		
	FMO3	-Liver (lower			
		expression level			
		than that of			
		FMO3)			
	FMO4	-Low expression			
	FMO5	-Liver			

183

## 184 1.2 From early reports to recent advances in FMO-mediated synthesis

185 In 1979, the catalytic model of mammalian FMOs was proposed based on spectrum and  
186 kinetic studies of pig FMO (pFMO) (Poulsen and Ziegler, 1979). In 1980, the catalytic  
187 C4a-(hydro)peroxide intermediate was discovered (Ball and Bruice, 1980), which was  
188 further demonstrated in 1981 (Beaty and Ballou, 1981a, 1981b). In 2008, the structure  
189 of mFMO was resolved at a 2.6 Å resolution (Fig. 4A) (Alfieri et al., 2008). In 2019, the  
190 structural movements of FMOs responding to different classes of substrates were  
191 unraveled (Fürst et al., 2019b).

192 The earliest record of indigo biosynthesis arises from the study of naphthalene  
193 dioxygenases (Ensley et al., 1983). In 1989, the pBS959 plasmid was constructed based  
194 on the naphthalene dioxygenase gene and pUC19 plasmid, which realized the synthesis  
195 of indigo in recombinant *E. coli* (Boronin et al., 1989). In 1993, a recombinant *E. coli*  
196 capable of producing indigo from glucose was also developed (Murdock et al., 1993).  
197 However, the yield was only 135 mg/L. The indigo synthesis pathways of other  
198 aromatic hydrocarbon-degrading bacteria were also identified. In 1997, two

indigo-synthesizing strains, *Pseudomonas putida* S12 and CA-3, were identified (O'Connor et al., 1997). The toluene dioxygenases from *P. putida* NCIB11767 (Stephens et al., 1989) and *P. putida* F1 (Woo et al., 2000) both demonstrated the ability to synthesize indigo. However, the ability to synthesize indigo of naphthalene dioxygenases and toluene dioxygenase is limited, and their production capacity fails to exceed 300 mg/L. Other classes of oxidases, such as toluene-4-monooxygenase from *P. mendocina* and functional enzymes similar to Multicomponent phenol hydroxylases in *Acinetobacter* sp. ST-550 (Doukyu et al., 2002; Doukyu et al., 2003) can also synthesize indigo, but their yield are inferior to mFMO, which has been widely studied for indigo biosynthesis. Choi et al. were the first to report synthesizing indigo by mFMO effectively in 2003 (Choi et al., 2003). Subsequent research on mFMO-catalyzed indigo biosynthesis was based on this study. With further optimization of the sequence and fermentation process (Han et al., 2008), Han et al. reported an indigo yield of 911 mg/L in recombinant *E. coli* (Han et al., 2011). By adjusting cysteine concentration in the tryptophan medium, the production of indigo was increased. In later studies, it was found that cysteine also had a significant impact on the catalytic selectivity of mFMO. During the oxidation catalytic process, the concentration of cysteine increased, and the reaction tended towards the indirubin synthesis (Kim et al., 2019), which is of great significance to the synthesis of indirubin by mFMO. Numerous indigo-producing bacterial FMOs were also reported (Ameria et al., 2015; Lončar et al., 2019a). However,

219 their yields were not comparable to mFMO. To overcome the limitations of indigo and  
220 indirubin biosynthesis by FMOs, various methods were employed, including enzyme  
221 engineering and fermentation engineering (Chen et al., 2021; Hsu et al., 2018; Lončar et  
222 al., 2019b; Sun et al., 2022).

223

## 224 **2. Structures and catalytic mechanisms**

### 225 **2.1 The structure and catalytic mechanism of mammalian FMOs, for instance,** 226 **AncFMO3-6**

227 The catalytic mechanism of AncFMOs is similar to that of human FMOs, with sequence  
228 identities ranging from 82% to 92%. The resolved structures of AncFMOs provide  
229 valuable insights into the catalytic mechanism and functions of mammalian FMOs  
230 (Nicoll et al., 2020).

231 Residues 510-532 of AncFMO3 form two  $\alpha$ -helices that anchor AncFMO3 tightly  
232 to the membrane in a highly hydrophobic transmembrane area (pale green in Fig. 3).  
233 AncFMO3-6 also contains two strictly conserved Rossmann fold binding areas for FAD  
234 (residues 2–154 and 331–442, pale cyan in Fig. 3) and NADP(H) (residues 155–213  
235 and 296–330, light blue in Fig. 3). An insertion constructed by three  $\alpha$ -helices (orange in  
236 Fig. 3), creates a ridged triangular fold that masks the FAD and the catalytic site (Fig.  
237 3B), providing a tunnel from the membrane to the catalytic cavity (Fig. 3C). This  
238 structure suggests that substrates enter the blocked catalytic cavity from the tunnel.

Ser62 of AncFMO3-6 can form hydrogen bonds with the N3 atom of the isoalloxazine ring and orientate the FAD towards the catalytic cavity (Fig. 3E). Additionally, Asn61, which is well-conserved in human FMOs near the C4a of the isoalloxazine ring, stabilizes the C4a-(hydro)peroxide intermediate. Mutations of Asn61 cause trimethylaminuria. The residue Leu375 protects the catalytic cavity from solvents.

-----Fig 3-----

## 2.2 The structure and catalytic mechanism of mFMO

Ongoing studies have described the complete catalytic model of FMOs (Ziegler, 1993) (Fig. 4E). The procedure involves five steps, including (1) FAD reduction to FADH<sub>2</sub> by NADPH and an additional H<sup>+</sup>, (2) interaction of FADH<sub>2</sub> with O<sub>2</sub> to form a stable C4a-(hydro)peroxide intermediate, (3) single oxidation of the closing substrates without particular binding, (4) discharge of the residuary oxygen atom as H<sub>2</sub>O, and (5) the whole cycle returns to the initial state when NADP<sup>+</sup> is separated from FMO. Since substrates do not need to provide reducing potential and specifically bind to FMO, they are extensive. Theoretically, any nucleophilic compound that could get close to FAD-OOH can become a substrate of FMO. However, molecular size and charges affect the selectivity of FMO. Substrates with one positive charge have the highest affinity to FMO (Ziegler, 2002).

mFMO has 456 amino acids and automatically forms homodimers *in vivo* (Alfieri et al., 2008) (Fig. 4A) and is not a membrane-associated protein like hFMO. There is one Rossmann fold binding area for FAD (1-169 and 281-461, pale cyan in Fig. 4), one Rossmann fold binding area for NADP<sup>+</sup> (170–280, light blue in Fig. 4), and three loops (44-80, 166-186, and 276-306, pale green in Fig. 4) linking the above two domains, lacking secondary structural elements. Both FAD and NADPH are the necessary cofactors of FMO. FMO oxidizes substrates by transferring the oxygen atom from C4a. At the same time, NADPH facilitates the reduction of FAD by transferring the hydrogen atom from C4 to N5. Additionally, NADPH's existence is essential for the stability of the C4a-(hydro)peroxide intermediate (Catucci et al., 2020). NADP<sup>+</sup> masks the catalytic cavity, providing a proper H-bonding environment that probably protects N5 of the flavine from solvents, maintains the catalytic intermediate, prolongs the intermediate half-life, and offers more chances to react. Tyr212 is a critical amino acid that assists the ribose of NADP<sup>+</sup> to protect NADPH from solvent interference (Fig. 4D). Tyr212 and residues 407-415 could facilitate the approach of substrates and catalytic intermediates. Consistent with the findings, our AutoDock (Morris et al., 2009) data indicate that Tyr212 also affects the catalytic process (Fan and Xie, 2023). Specifically, we found that Tyr212 can restrict the size of the catalytic cavity, limit the dimensions of substrates, and protect substrates from solvents, functioning as a "doorkeeper" (Fig. 5B). Moreover, Trp47 is an important residue for the rate of hydride transfer from NADPH to the flavin



(Han et al., 2013).

The AutoDock simulation revealed that Asn78 is in the catalytic cavity and proximity to indole, FAD, and NADP(H). We believe that Asn78 plays an important role in the catalytic process. Indole was found to be contained in the catalytic cavity, forming hydrogen bonds with NADP(H), and closing the isoalloxazine ring of FAD (Fig. 5A).

-----Figs 4, 5-----

**Table 2.** The  $K_{cat}$  and  $K_m$  of different FMO subtypes on various substrates

FMO subtypes	Substrates	$K_{cat}(S^{-1})$	$K_m(\mu M)$	$K_{cat}/K_m(S^{-1}M^{-1})$	References
AsFMO <sup>a</sup>	Allyl mercaptan	$0.75 \pm 0.01$	$(1 \pm 0.1) \times 10^3$	$(0.75 \pm 0.08) \times 10^3$	Valentino et al., 2020
	L-cysteine	$1.7 \pm 0.03$	$(5 \pm 0.4) \times 10^3$	$(0.34 \pm 0.03) \times 10^3$	
	N-acetyl L-cysteine	$0.57 \pm 0.004$	$(7 \pm 0.4) \times 10^3$	$(0.08 \pm 0.005) \times 10^3$	
AncFMO2	Methimazole	$0.19 \pm 0.01$	$106 \pm 22$		Nicoll et al., 2020
	Thioanisole	$0.3 \pm 0.02$	$6.9 \pm 1.6$		
	Trimethylamine	$0.16 \pm 0.008$	$445 \pm 74$		
	NADPH	$0.32 \pm 0.05$	$7.8 \pm 1.4$		
	NADPH <sub>uncoupling</sub>	$0.02 \pm 0.001$	$20 \pm 5.4$		
AncFMO3-6	Methimazole	$0.19 \pm 0.005$	$21 \pm 2.3$		
	Thioanisole	$0.1 \pm 0.008$	$128 \pm 38$		
	Trimethylamine	$0.24 \pm 0.01$	$41 \pm 6.3$		
	NADPH	$0.13 \pm 0.008$	$3.5 \pm 0.86$		
	NADPH <sub>uncoupling</sub>	$0.022 \pm 0.002$	$16 \pm 5.4$		
AncFMO5	Heptan-2-one	$0.07 \pm 0.003$	$6.36 \pm 1.2$		
	NADPH	$0.06 \pm 0.001$	$6.48 \pm 0.38$		
	NADPH <sub>uncoupling</sub>	$0.03 \pm 0.001$	$2.1 \pm 0.5$		
bFMO <sup>a</sup>	Indole	$0.22 \pm 0.01$	$(0.98 \pm 0.12) \times$	$(0.22 \pm 0.02) \times 10^3$	Sun et al.,

			$10^3$		2022
	Indole (K223R/D317S)	$0.79 \pm 0.03$	$(0.55 \pm 0.14) \times 10^3$	$(1.45 \pm 0.30) \times 10^3$	
MBP-cFMO <sup>a</sup>	Trimethylamine	2.6	575	4521.7	Jung et al., 2018
	Thiourea	0.67	380	1750	Ameria et al., 2015
	Cysteamine	1.83	6000	300	
hFMO1 <sup>a</sup>	Methimazole	1.07	7		
	Imipramine	0.85	14		Furnes and Schlenk, 2004
	Fenthion	1.55	340		
	Methyl p-tolyl sulfide	0.767	284		
hFMO2	Thiourea	0.85	27		Henderson et al., 2004
	Phorate	0.683	57		
	Ethylene thiourea	0.833	14		
	a-Naphthylthiourea	0.55	42		Krueger et al., 2002b
	1-Phenylthiourea	0.4	29		
	Disulfoton	0.333	32		Krueger et al., 2002a
hFMO3	Benzydamine	$3.12 \pm 0.20 \text{ (min}^{-1}\text{)}$	$52.0 \pm 9.0$		Bortolussi et al., 2021
	Tamoxifen	$1.13 \pm 0.70 \text{ (min}^{-1}\text{)}$	$6.40 \pm 0.70$		
	Fenthion	0.183	145		Furnes and Schlenk, 2004
	5-DPT	0.85	155		
	Trimethylamine	3.15	32		Cashman et al., 2000
	Tyramine	1.833	231		
Truncated hFMO3	Benzydamine	$4.51 \pm 0.18 \text{ (min}^{-1}\text{)}$	$53.0 \pm 6.1$		Bortolussi et al., 2021
	Tamoxifen	$1.63 \pm 0.01 \text{ (min}^{-1}\text{)}$	$5.60 \pm 0.30$		
mFMO <sup>a</sup>	Trimethylamine	$7.3 \pm 0.6$	$6.1 \pm 0.1$	$(8.4 \pm 0.7) \times 10^5$	
	Trimethylamine (E158A/E159A)	$7.1 \pm 1$	$6.4 \pm 0.3$	$(9.0 \pm 1) \times 10^5$	
	Nicotine	$130 \pm 30$	$3.0 \pm 0.2$	$(2.3 \pm 0.6) \times 10^4$	Alfieri et al., 2008
	Methimazole	$66 \pm 7$	$1.0 \pm 0.02$	$(1.5 \pm 0.2) \times 10^4$	
	N, N dimethylaniline	$232 \pm 28$	$1.8 \pm 0.4$	$(7.8 \pm 2) \times 10^3$	
	Indole	$90 \pm 14$	$0.7 \pm 0.03$	$(7.8 \pm 1) \times 10^3$	

		23 (nmol/min/mg protein)	19		
	Trimethylamine				
		29 (nmol/min/mg protein)	180		Choi et al., 2003
	Cysteamine				
		17 (nmol/min/mg protein)	390		
	Thiourea				
NiFMO <sup>a</sup>	Trimethylamine	2.01 ± 0.08	45.6 ± 9.6	44,000	
	Methimazole	1.10 ± 0.08	77 ± 10	14,000	
	Pyrrole	0.69 ± 0.03	46 ± 17	4,800	
	Indoline	0.70 ± 0.03	98 ± 14	7,100	Lončar et al., 2019a
	Indole	0.11 ± 0.01	137 ± 23	730	
	6-Bromoindole	0.09 ± 0.03	640 ± 340	140	
	Tris(hydroxymethyl) aminomethane (TRIS)	0.18 ± 0.02	21500 ± 5600	9	
pFMO1 <sup>a</sup>	Dihydrolipoic acid		1700		Taylor and Ziegler, 1987
	Lipoic acid		120		Sabourin and Hodgson, 1984
	Trimethylamine		617		
rFMO2 <sup>a</sup>	n-Heptylamine		27,000		
	N-octylamine		6700		
	N-nonylamine		1800		
	N-decylamine		400		Tynes et al., 1986
	N-undecylamine		90		
	N-dodecylamine		33		
	N-tridecylamine		13		

288 <sup>a</sup> As = *A. sativum*; b = *Methylophaga aminisulfidivorans*; c = *Corynebacterium glutamicum*; h =

289 *Homo sapiens*; m = *Methylophaga sp.* strain SK1; Ni = *Nitrincola lacisaponensis*; p = *Sus scrofa*; r =

290 *Oryctolagus cuniculus*.

291

292 **3. The application of FMO in biosynthesis**

### 3.1 The advantages of FMO among enzymes in indigo biosynthesis

In the field of indigo biosynthesis, researchers aim to achieve higher bio-yield and simpler production conditions. FMO is currently the most studied enzyme in this regard and is particularly prominent in these two aspects. While there are reports on the biosynthesis of indigo using other enzymes, most of them remain at the primary level of "observed phenomena".

For instance, Cytochrome P450 monooxygenases have been observed to synthesize indigo under specific conditions (Fiorentini et al., 2018; Kim et al., 2018), but their biosynthetic ability has not been thoroughly explored, perhaps due to the characteristics and productive ability of the enzyme. Similarly, multicomponent enzymes without heme, such as Naphthalene dioxygenases (Doukyu et al., 2002; Pathak and Madamwar, 2010; Groeneveld et al., 2016) and multicomponent phenol hydroxylases (Doukyu et al., 2003; Qu et al., 2012a), have been used as biocatalysts, but their application is limited by the fermentation conditions. The yield of indigo synthesized by these enzymes has not exceeded 300mg/L due to the need for stable conditions for each component, which makes FMOs with simple components more practical for utilization.

Furthermore, it has been shown that D flavoprotein monooxygenases, a close relative of FMO, can synthesize indigo (Dai et al., 2019), but further research is needed to understand the underlying mechanism on the biosynthetic level.

### 3.2 FMO-Catalyzed Biosynthesis of Indigo

FMOs possess a unique oxidative catalytic mechanism, and their application in biosynthesis, particularly in the synthesis of indigoid compounds, has been extensively studied (Table 3). Indigo, one of the oldest textile dyes, has traditionally been produced through plant extraction or chemical synthesis (Stasiak et al., 2014). However, to develop an environmentally friendly strategy, biosynthesis of indigo using microorganisms has been proposed (Qu et al., 2010; Honda et al., 2008; Qu et al., 2012b), and engineered *E. coli* has been validated for indigo biosynthesis (Ensley et al., 1983; Murdock et al., 1993; Berry et al., 2002; Doukyu et al., 2003). The discovery of mFMO from *Methylophaga* sp. strain SK1 has significantly improved the efficiency of heterologous biosynthesis of indigo (Choi et al., 2003).

In the biosynthetic strategy, indole is used as a substrate or intermediate in the pathway. With the help of heterologous FMO, indole is oxidized to 3-hydroxyindole (indophenol/hydroxyindole), which further forms indigo by spontaneous dimerization under aerobic conditions (Choi, 2020; Han et al., 2008). However, the productivity of the original plasmid carrying mFMO was limited, and optimization of the engineered plasmid was implemented. After removing the redundant sequences, the indigo yield of the improved plasmid reached 662 mg/L, an increase of 413% compared to the initial production of 160 mg/L (Han et al., 2008). Moreover, a yield of 920 mg/L in a 5L fermenter was achieved, and a larger scale fermentation system was developed, which

333 achieved a yield of  $911 \pm 22$  mg/L of indigo in a 3,000 L fermenter with an input of 2  
334 g/L of tryptophan (Han et al., 2011).

335 Numerous studies have been conducted to improve the properties of the enzyme  
336 and optimize biosynthetic conditions (Table 3). For instance, a consecutive two-cell  
337 reaction system was constructed using mFMO, 6-halogenase SttH, and tryptophanase  
338 TnaA to synthesize 6,6'-dibromoindigo (6BrIG) with a yield of 315.0 mg/L from  
339 tryptophan (Lee et al., 2021). The redox reaction of FAD cofactor, which tightly links  
340 with FMO, depends on the assistance of NADPH (Alfieri et al., 2008; Krueger and  
341 Williams, 2005). To recycle NADPH, an optimized strategy of fusing mFMO and  
342 phosphate dehydrogenase (PTDH) was proposed to utilize phosphate as a cheap and  
343 sacrificial substrate (Rioz-Martinez et al., 2011). This bifunctional enzyme,  
344 mFMO-PTDH, not only exhibits the ability to oxidize indole and its analogs but also  
345 shows potential for chirality selection in the biocatalytic sulfoxide oxidation of prochiral  
346 sulfides (Pereira et al., 2022; Schnepel et al., 2021; Wojaczyńska and Wojaczyński,  
347 2020). Moreover, directed evolution was applied for mFMO modification, and some  
348 mutants exhibited higher  $K_{cat}$  and lower  $K_m$ , showing higher indole affinity and superior  
349 catalytic efficiency at low substrate concentrations (Lončar et al., 2019b).

350 Since indigo itself is insoluble in water, it requires the addition of reductants such  
351 as sodium sulfate to modify it into a water-soluble dye in practical production, which  
352 increases economic and environmental burdens. A novel strategy of chemical group

353 protection was proposed to effectively solve the insolubility problem of indigo in water.  
354 A glycosyltransferase from *Polygonum tinctorium* was co-expressed with mFMO in *E.*  
355 *coli* to introduce a glucose group to protect the active hydroxyl group of hydroxyindole  
356 and form a water-soluble and stable indican to prevent indigo from further oxidation  
357 (Hsu et al., 2018). Subsequently, by adding glucosylase to remove the protective groups,  
358 the production of indigo was restored, showing excellent practical properties. To reduce  
359 the substrate cost of indigo fermentation, a system based on the co-cultivation of  
360 microorganisms was developed to convert renewable carbon substrates to indigo (Chen  
361 et al., 2021). The system can be divided into upstream and downstream parts: the  
362 upstream engineered *E. coli* is responsible for the synthesis of tryptophan from glucose  
363 or glycerol; the downstream system is responsible for the conversion of tryptophan to  
364 indole and then to indigo. Besides, the metabolism of the two strains was specially  
365 optimized, and the final indigo yield reached 104.3 mg/L, which was more than eleven  
366 folds higher than the original strain.

367 Recently, a strategy was implemented to increase the cyclopropane fatty acid (CFA)  
368 composition of phospholipid fatty acids in the cell membrane to counteract the  
369 cytotoxic effects of indole. This approach resulted in a 1.5-fold increase in indigo  
370 production compared to the control strain (Ham et al., 2023). Additionally, the enzyme  
371 TRP can convert tryptophan into indole and pyruvate (Shimada et al., 2004). As  
372 pyruvate can be used by the cell to regenerate NADPH, a bifunctional fusion enzyme of

373 mFMO-TRP was developed to produce indigo from L-tryptophan. This approach  
374 resulted in the complete conversion of 2.0 g tryptophan into 1.7 g indigo per liter of  
375 culture (Fabara and Fraaije, 2020).

376 In addition to mFMO, other FMOs isolated from various organisms have shown  
377 potential for synthesizing indigo compounds. For instance, FMO motifs encoded by  
378 *Mesorhizobia* and *Sphingomonas* isolated from wastewater sludge by metagenomic  
379 methods exhibit biocatalytic activity for indigo synthesis (Singh et al., 2010). PtFMO  
380 was cloned from *P. tinctorum* and heterologously expressed in *E. coli* BL21 (DE3),  
381 achieving a 30 mg/L indigo yield (Inoue et al., 2021). The cFMO, cloned from  
382 *Corynebacterium glutamicum*, reaches the biosynthetic indigo yield of 685 mg/L  
383 (Ameria et al., 2015) in engineered *E. coli*. Mutants F170Y, A210G, A210S, and T326S  
384 cFMOs were fused with maltose-binding protein (MBP), and the fusion protein  
385 produced 1,040 mg/L indigo and 112 mg/L indirubin with 2.5 g/L tryptophan (Jung et  
386 al., 2018). While NiFMO from *Nitrincola lacisaponensis* has a comparably lower yield  
387 of indigo, it exhibits unique thermal stability, and the optimum catalytic temperature  
388 reaches 51 °C, which is about 8 °C higher than mFMO (Lončar et al., 2019a).

389 In summary, the biosynthesis of indigo using microorganisms has been proposed as  
390 an environmentally friendly strategy, and the discovery of mFMO has significantly  
391 improved the efficiency of heterologous biosynthesis of indigo. Various studies have  
392 been carried out to optimize the biosynthetic conditions and improve the properties of



the enzyme, which contributes to the development of efficient strategies for indigo production. A discussion of potential strategies to improve yields is included in the 3.4 subsection.

### 3.3 FMO-Catalyzed Biosynthesis of Indirubin

Utilizing the improved mFMO (Han et al., 2008), the biosynthetic pathway of indirubin was further studied. Cysteine was found to increase the selectivity of FMO for 2-hydroxyindole (Han et al., 2012), leading to a yield of 223.6 mg/L under optimum conditions. In 2022, a semi-rational design was applied and combined modification to the two regions (K223R/D317S) of FMO. Based on metabolic engineering technology, *E. coli* strains were screened to obtain high indirubin production, reaching the current maximum yield of 860.7 mg/L (Sun et al., 2022).

To reduce substrate costs in fermentation, researchers have employed mFMO and *E. coli* tryptophanase (TRP) to facilitate the biological synthesis of indirubin utilizing glucose as the primary substrate. As a result, a notable yield of 0.056 g/L of indirubin has been accomplished (Du et al., 2018).

**Table 3.** The summary of production of indigo/indirubin by FMOs

Year	Strategies	Genetic sources of FMO	Chassis	Plasmid/Promoter	Production of indigo/indirubin (mg/L)	Supplements in culture	References
2003	First clone mFMO	<i>Methylophaga</i> sp. SK1	<i>E. coli</i> DH5 $\alpha$	pBlue 2.0	indigo: 160	2.00 g/L tryptophan	Choi et al., 2003
2008	Gene sequence	<i>Methylophaga</i>	<i>E. coli</i>	pBlue 1.7	indigo: 920	2.40 g/L	Han et al.,

	and culture optimization	<i>aminisulfidivorans</i> MPT	DH5 $\alpha$			tryptophan	2008
2011	Batch and continuous fermentation	<i>Methylophaga aminisulfidivorans</i> MPT	<i>E. coli</i> DH5 $\alpha$	pBlue 1.7	indigo: 911 $\pm$ 22	2.00 g/L tryptophan	Han et al., 2011
2012	Cysteine influences the regioselectivity	<i>Methylophaga aminisulfidivorans</i> MPT	<i>E. coli</i> DH5 $\alpha$	pBlue 1.7	indirubin: 223.6	2.00 g/L tryptophan, 0.36 g/L cysteine	Han et al., 2012
2015	First clone cFMO	<i>Corynebacterium glutamicum</i>	<i>E. coli</i> WCO2 1	pMCF14	indigo: 685 indirubin: 103	2.50 g/L tryptophan	Ameria et al., 2015
2018	Site-directed mutagenesis	<i>Corynebacterium</i>	<i>E. coli</i> WTS32 6	pK-T326 S	indigo: 1,040 indirubin: 112	2.50 g/L tryptophan	Jung et al., 2018
2018	Biochemical protecting group for dyeing	<i>Methylophaga aminisulfidivorans</i>	<i>E. coli</i> MG165 5	pTMH56 1	indigo: 400	3.10 g/L tryptophan	Hsu et al., 2018
2020	Plant origin	<i>Polygonum tinctorium</i>	<i>E. coli</i> BL21	pET19b	indigo: 30	0.80 mM tryptophan	Inoue et al., 2021
2020	Fusion expression	<i>Methylophaga sp.</i> SK1	<i>E. coli</i> NEB10 $\beta$	pBAD	indigo: 1700	2.00 g/L tryptophan	Fabara and Fraaije, 2020
2021	Co-culture	<i>Methylophaga aminisulfidivorans</i> MPT	<i>E. coli</i>	pCDFDue t-1	indigo: 104.3	5.00 g/L glucose or glycerol	Chen et al., 2021
2022	Structure-guided enzyme engineering	<i>Methylophaga aminisulfidivorans</i> MPT	<i>E. coli</i> BL21	pET28a	indirubin: 860.7	2.00 g/L tryptophan, 0.36 g/L cysteine	Sun et al., 2022
2023	Regulate the permeability of the cell membrane	<i>Methylophaga aminisulfidivorans</i>	<i>E. coli</i> BL21	pCDFDue t-1	indigo: 1080	10.00 mM tryptophan	Ham et al., 2023
2023	Producing isobutanol and indigo together	<i>Methylophaga aminisulfidivorans</i>	<i>E. coli</i> BL21	pETDeut-1	isobutanol: 5800 indigo: 194.1	5.00 mM tryptophan	Cho et al., 2023

### 3.4 Strategies for improving FMO-catalyzed biosynthesis

In 2023, silkworms were used to synthesize indigo, which opens a new avenue (Jia et al., 2023). Although researchers have achieved a yield of 1700 mg/L of indigo (Fabara and Fraaije, 2020), depending on tryptophan as the substrate and the lack of industrialization of indigo biosynthesis indicate the need of further improvements if we plan to use biotechnology as a substitute for the chemical method. Remodeling of FMOs, such as directed evolution and *de novo* design, holds promise for achieving higher yields. Additionally, few researchers have studied indigo biosynthesis using eukaryotes like yeasts as the chassis (Fig. 9A). To increase the yield of indigo synthesized by FMOs, several approaches can be pursued.

#### 3.4.1 Insights from FPMOs

To explore the potential of FMOs in indigo production, one effective approach is to examine studies on other enzyme families with similar mechanisms, such as flavoprotein monooxygenases (FPMOs).

Within the FPMO class B, which uses NADPH as a coenzyme, there are three subgroups, including FMOs, microbial N-hydroxylating monooxygenases (NMOs), and Baeyer-Villiger monooxygenases (BVMOs) (Paul et al., 2021; van Berkel et al., 2006). BVMOs can produce 0.031 g/L indirubin with 1 mM indole and 5 mM cysteine and convert indole to indigo only after specific mutations (Pazmiño et al., 2007; Catucci et

al., 2022). A mutant BVMO from *Acinetobacter radioresistens* (ArBVMO) with the R292A mutation can efficiently produce indigoids, with a  $k_{cat}$  of  $0.12\text{ s}^{-1}$  for indigo production. It can synthesize 0.138 g/L indirubin from 5 mM indole and 5 mM cysteine, slightly lower than that of mFMO (Catucci et al., 2022). As cysteine enhances the selectivity of FMOs for indirubin production, ArBVMO is regulated in the same way, indicating that their structure of regulation is similar.

Styrene monooxygenase (SMO), which belongs to the FPMO class E, can convert tryptophan to indigo (Cheng et al., 2016; Pan et al., 2023). The maximum yield of indigo produced by SMO was up to 787.25 mg/L after 24 hours of fermentation with 2.0 g/L tryptophan as substrates (Pan et al., 2023). Further research on indigo production by different FPMOs is recommended, as the functional conservation between these enzymes offers potential for the *de novo* design in the future.

While there is little research on the random mutagenesis of FMOs, researchers have successfully redesigned several other monooxygenases, including SMOs (Tan et al., 2019) and cyclohexanone monooxygenases (CHMOs) (Zhang et al., 2019), resulting in significantly increased catalytic activity. For example, CHMO, a BVMO named for their natural substrate cyclohexanone, was performed both local and global protein engineering using two Casting libraries surrounding  $\text{FAD}^+$  and  $\text{NADP}^+$  prosthetic groups, as well as an error-prone PCR library of the full-length CHMO. This resulted in a 50-fold increase in activity compared to the previously best-performing variant

(Zhang et al., 2019). As the structure of mFMO resembles BVMOs in their coenzyme binding mode (Alfieri et al., 2008; Paul et al., 2021), it is possible that random mutagenesis could enhance the catalytic activity of mFMO and increase indigo production.

### **3.4.2 Directed evolution**

Although FMOs (EC 1.14.13.8) are of high interest in biotechnology due to their ability to catalyze various regioselectivity and stereoselectivity monooxygenation reactions (Catucci et al., 2017; Mitsubayashi and Hashimoto, 2000), natural enzymes have inherent limitations that restrict their industrial utilization in indigo production (Ameria et al., 2015; Choi, 2020; Ma et al., 2018; Rioz-Martinez et al., 2011). Therefore, in recent years, there has been an increasing number of publications focusing on enzyme engineering to improve enzymes' catalytic activity, specificity, regioselectivity, stereoselectivity, thermostability and solvent tolerance (Choi, 2020; Fabara et al, 2020; Linke et al., 2023).

Over the past two decades, many enzyme engineering strategies have been developed, with directed evolution being the most studied (Fig. 6) (Chowdhury and Maranas, 2020; Qi et al., 2022; Xiong et al., 2021). The concept of directed evolution is a natural evolution process through iterative cycles of genetic diversification followed by high-throughput screening or selection (Xiong et al., 2021). This concept emerged

early in 20th century and continued to develop in the 1990s. Three types of methods were identified for randomly generating mutants and *in vitro* recombination (Xiong et al., 2021). The first type is error-prone PCR, based on oligonucleotide mutagenesis (Cadwell and Joyce, 1992), which was first described in 1989 (Lenug D, 1989). The second type of methods is based on homologous recombination, including DNA shuffling (Kikuchi et al., 2000), staggered extension process (StEP) (Zhao et al., 1998), random priming *in vitro* recombination (RPR) (Shao et al., 1998), and random chimeragenesis on transient templates (RACHITT) (Coco, 2003). Finally, the third type of methods includes incremental truncation for the creation of hybrid enzymes (ITCHY) (Hu et al., 1997), random multi-recombinant PCR (RM-PCR) (Tsuji et al., 2001), and sequence homology-independent protein recombination (SHIPREC) (Sieber et al., 2001).

-----**Fig 6**-----

In 2003, a mutant generation and clone screening approach was utilized with a flow cell optical rotation and a UV detector to engineer FMOs from *Oryctolagus cuniculus* (rFMO) and *Rhesus macaque* (rmFMO) for stereochemical preference control in the sulfoxidation of methyl p-tolyl sulfide. The engineered rFMO displayed up to 35% increased activity (Polyzos, 2003), meeting the demand. Furthermore, unspecific

peroxygenase (UPO, EC 1.11.2.1), a type of heme-thiolate enzyme with self-sufficient monooxygenase activity, can also catalyze the reaction from indole to indigo. The *Agrocybe aegerita* UPO1-encoding gene was subjected to directed evolution in *Saccharomyces cerevisiae*, and after five generations of evolution, nine mutations were screened, resulting in a 3250-fold activity improvement with no alteration in protein stability (Molina-Espeja et al., 2014). Additionally, after directed evolution, the half-life of a peroxidase (EC 1.11.1.16) increased from 3 to 35 minutes in the presence of 3000 equivalents of H<sub>2</sub>O<sub>2</sub> and with a 6 °C upward shift in thermostability (Gonzalez-Perez et al., 2014).

### 3.4.3 Semi-rational design of FMOs

Referring to enzyme engineering reviews (Xiong et al., 2021), semi-rational design strategy generates a mutation library on specific amino acid residues based on sequence and structural information of enzyme protein. This strategy avoids the time-consuming and labor-consuming process in directed evolution and has a convenient screening of suitable results. Semi-rational design strategy is based on two methodologies. In the sequence-based enzyme redesign method, critical conserved residues binding to substrate are identified through multiple sequence analysis of homologous protein sequences (Fig 5). In contrast, in the structure-based enzyme redesign method, the functional hot spots are pinpointed by analyzing substrate, transition state, or product

binding in the active site using docking results (Fig 4). Eventually, site-directed mutation and site-directed saturation mutation libraries are constructed and then screened for new functional enzymes.

To date, using site-directed mutagenesis, several studies have identified the role of amino acid residues associated with the structure and function of FMOs, accumulating a wealth of information on residue-function association (Wyatt et al., 1998). This information provides a valuable source for enzyme semi-rational design. For example, hFMO3 was researched through semi-rational design in different aspects. Numerous hFMO3 allelic variants were identified in patients and the 3D structural model of hFMO3 was generated by homology modeling, which was soon used to perform molecular dynamics simulations, followed by structural mapping of 12 critical polymorphic variants and molecular docking experiments with five different known substrates of hFMO3, giving insights into their binding mechanism (Gao et al., 2016).

A computational library design protocol named FRESCO is an structure-based enzyme engineering tool for stabilization (Wijma et al., 2018). In 2019, this protocol was first performed on cyclohexanone monooxygenase (EC 1.14.13.22), resulting in 128 screened point mutants. The most stable and highly active mutant displayed an increase in unfolding temperature of 13 °C and an approximately 33-fold increase in half-life at 30 °C (Fürst et al., 2019a). Then, after discovering the thermostable NiFMO (Lončar et al., 2019a), FRESCO was utilized to improve its thermostability. After



531 screening a relatively small number of enzyme mutants, the  $k_{cat}$  for indole was improved  
532 by 1.5-fold (Lončar et al., 2019b). Furthermore, in 2022, the catalytic efficiency of  
533 bFMO, cloned from *Methylophaga aminisulfidivorans*, was enhanced by 6.6-fold using  
534 FRESCO (Sun et al., 2022). Moreover, a recent study engineered mFMO for industrial  
535 application using the Protein Repair One-Stop Shop (PROSS) algorithm (Goldenzweig  
536 et al., 2016; Peleg et al., 2021; Wijma et al., 2018) and successfully redesigned a mutant  
537 with the ability to reduce TMA levels in a salmon protein hydrolysate at industrially  
538 relevant temperatures (Goris et al., 2023).

539 Further improvements could be made to FMOs in their industrial utilization for  
540 producing indigo. In 2016, a study combining structural analysis and sequence  
541 alignment successfully switched the cofactor preference of type I BVMOs from  
542 NADPH to NADH (Beier et al., 2016). As NADH cofactor is preferred for industrial  
543 purposes, this redesign could be a promising future project to apply to FMOs to acquire  
544 better industrial properties.

545

#### 546 **3.4.4 *De novo* enzyme design with high substrate specificity**

547 Recently, the limitations of directed evolution or semi-rational design strategies have  
548 become apparent due to their mutation rates, making it difficult to satisfy stability or  
549 specific activity needs for certain applications. *De novo* protein design is an emerging  
550 method for enzyme engineering that can generate a sequence based on the given

structure (Chen et al., 2022; Wang et al., 2023; Xiong et al., 2021). Computational physics-based tools for *de novo* design, such as dynamics simulation and optimizing sequences based on a given structure (Ferreira et al., 2022), can highly assemble naturally existing enzymes. Additionally, AI tools for protein design, which can generate *de novo* motifs or “hallucinate” a compatible structure based on the input structure backbone (Anishchenko et al., 2021), have become a new trend in life science (Ding et al., 2022; Kim et al., 2021; Liu and Chen, 2022; Yang et al., 2023). It is worthwhile to describe some of the most exciting tools that have been developed in recent years and their powerful potential applications in the industrialization of FMOs.

*De novo* design strategies have rapidly developed through energy computing and the development of protein modeling (Wang et al., 2023; Woolfson, 2021). There are two types of protein modeling mechanisms, which are correlated to different tools. The first type is the template-based protein modeling, which is based on multiple sequence alignment (MSA) and residue-residue interaction (RRI) data. Some successful AI-based programs in this category include AlphaFold-2, RoseTTAFold, and I-TASSER. The second type is non-template-based protein modeling, which used methods of Monte Carlo (MC) or Markov Chain Monte Carlo (MCMC) simulations to get the protein to fold into its energy-minimized state.

Introductions for bioinformatics tools developed for *de novo* enzyme design are shown in Table 4, such as RoseTTAFold (Wang et al., 2022), Message-Passing Neural

Network (MPNN) (Dauparas et al., 2022), RFdiffusion (Watson et al., 2023), and Family-wide hallucination (Yeh et al., 2023). Each tool has its own advantages and limitations.

**Table 4.** The summary of various *de novo* design tools

Protein design tools	Input	Function	Advantages	Shortcomings	Remarks	References
<i>1. Protein structure prediction</i>						
AlphaFold2	Protein sequences (MSA)	Deep learning-based protein structure prediction	High accuracy	1. High computational resource requirements; 2. Limited accuracy; 3. Cannot predict novel structures	As a baseline model for protein structure prediction, the model is modified for protein design. Introduce energy function to consider the rationality of protein structure, variant tools based on specific tasks can be flexibly applied to protein design	Jumper et al., 2021
RoseTTAFold	Protein sequences	Deep learning-based protein structure prediction		1. Limited to specific issues; 2. Energy function limits	Monomer and Multimer structure predictions were ranked at the top and fourth in CASP15	Wang et al., 2022
trRosettaX2	Protein sequences (MSA)	Deep learning-based protein structure prediction	More accurate than AlphaFold2	Multimer structure prediction		Peng et al., 2023
ESMFold	Protein sequences	Deep learning-based protein	1. Protein-specific prediction; 2. Faster than MSA-based solutions	Less accurate than AlphaFold2	A model based on information extracted from	Weissenow et al., 2022

D-I-TASSER	Protein sequences (MSA)	structure prediction			protein language models might be better for protein design than those based on MSAs	Zheng et al., 2021
		Template-based protein structure prediction		Not satisfied with multi-domain protein modeling	A model based on the threading method. The server can perform both template-based and template-free modeling	
2. Physics-based protein design						
TopoBuilder	Ideal secondary structure elements (SSEs) and epitope Coordinates, sequences, and connectivity information for a set of substructures from native protein	Backbone generation	Can construct and design any protein form description	Limits to the guidance of natively arranged SSEs	A fragment-based method works as an extension of the Fold Form Loops protocol	Harteveld et al., 2022
SEWING	Protein structures of selected protein family	Backbone generation	1. Rapidly generate large numbers of scaffolds; 2. Advantageous for functional design.		Perform structural extension with native-substructure graphs	Guffy et al., 2018
AbDesign		Backbone generation	Large-scale assembly		De novo backbone design by assembling local structures	Lipsh-Sokolik et al., 2021
HBNet	Backbone structure	Protein design	Hydrogen-bond networks design		A method that calculates cooperative hydrogen bonding networks should	Boyken et al., 2016

				be broadly useful in enzyme design	
Meta-multistate design	Backbone structure	Sequence optimization	1. Sequence design considering conformationally dynamic state; 2. Large ensemble size	A multistate design method could be used to design multisubstrate enzymes	Davey et al., 2017
Rosetta		Protein modeling and design	1. Many completely Rosetta-based protocols exist; 2. Various web servers	A powerful suite of programs for <i>de novo</i> protein design	Leman et al., 2020
IPro		Protein design	Integrate many computational protein engineering methods	Perform different functions with different modules	Pantazes et al., 2015
OSPREY	Protein structure	Protein modeling and design	With various functions related to protein design based on GPU acceleration	Use a physics-based energy function that combines both Amber and CHARMM	Hallen et al., 2018
ISAMBARD	Natural protein structures or backbones depend on the approaches	Energy prediction	User friendly	A suite of <i>de novo</i> biomolecular design tools using BUFF energy function	Wood et al., 2017
Tinker	Protein structure	Molecular mechanics (MM) and molecular dynamics (MD) simulations	1. User friendly; 2. Interoperable with other tools and access to many force fields	The main characteristic of this package is its modularity and has developed a few branches like Tinker-HP and Tinker-OpenMM	Rackers et al., 2018
OpenMM	Support for multiple input pipelines	Protein dynamics simulation	High extensibility	Based on extensibility in every layer of	Eastman et al., 2017

GROMACS	Protein structure	Protein dynamics simulation	1. Fast; 2. Various simulations; 3. Able to scale the largest machines in the world	the architecture, users can easily add new features like novel functional forms or new simulation protocols. Support simulations including leap-frog Verlet, velocity Verlet, Brownian, and stochastic dynamics, as well as calculations that do energy minimization, normal-mode analysis, and simulated annealing.	Abraham et al., 2015
		Homology modeling, energy minimization, docking, and molecular dynamic simulations	User-friendly with a built-in graphical interface	Can be used as a molecular dynamics refinement for Rosetta to broaden the substrate scope.	
MODIP	Protein sequence	Disulfide bond design	Depends on database	Used in the DSDBASE2.0 database on native and modeled disulfides to identify pairs of residues that form disulfide	Kalmankar et al., 2022

PROPKA	Protein structure	protonation states prediction			bonds A model focuses on the prediction of ligand pK <sub>a</sub> values	Søndergaard et al., 2011
MaSIF	Database of known protein structure and protein binding surface information	Protein interaction analysis	The target protein can be searched directly across the mechanism of protein-protein interaction.	1. The conformation of the designed protein is inflexible; 2. The antimutagenicity is not considered; 3. It depends on the binding surface seed bank; 4. The functional goal is single and needs artificial screening and optimization.	Techniques from the field of geometric deep learning are introduced to resolve information encoding protein interaction interfaces	Gainza et al., 2023
<i>3. Statistics-based de novo design</i>						
ProteinMPNN	Coordinates of protein (polymer)	Protein design	1. short calculation time; 2. Able to design new proteins with few restrictions	1. Computational requirements; 2. Hardware memory limitations	For the design of monomeric protein design, the Monte Carlo optimization algorithm and the hallucination strategy are combined, and the loss function is used to ensure the consistent stability of protein folding and assembly.	Dauparas et al., 2022
PoPMuSiC	Protein structure	Energy prediction	1. Rapid computation of stability changes induced by single-site mutation; 2. Good prediction performance; 3. User friendly		Use the statistical energy functions extracted from empirical data	Dehouck et al., 2011
Autodock	Protein structure	Docking	Molecular interactions		A suite of toolkits for	Goodsell et al., 2021

---

(with ligand)

various  
computational  
ligand docking  
that has  
developed over  
30 years

Use a  
surface-based  
molecular  
similarity search  
engine to rapidly  
generate a  
suitable  
conformation  
and alignment

Jain, 2003

Surflex-Dock

Protein  
structure  
(with ligand)

Docking

Superior in predicting  
the binding mode and  
binding score

CSR-SALAD

Protein  
structure

Cofactor  
specific  
switch  
analysis

User friendly

Perform the  
switch between  
NAD and NADP  
cofactors

Cahn et al.,  
2018

Combine a  
statistical energy  
function with  
van der Waals  
energy terms to  
search stable  
sequences for a  
given backbone  
structure

ABACUS

Backbone  
structure

Energy  
prediction

No boundary of  
homology between  
design targets and  
training proteins

Xiong et al.,  
2017

DenseCPD

Structure  
(atom  
distribution  
information)

Deep  
learning-based  
protein  
sequence  
design

Quick and convenient  
to generate *de novo*  
sequence

Find the most  
suitable  
sequences for the  
protein backbone

Qi and  
Zhang, 2020

RFdiffusion

Protein  
structure  
coordinates  
with Gaussian  
noise

Deep  
learning-based  
protein design

1. Better generation of  
diversity; 2. Potential to  
solve complex  
problems

The extension of  
the  
RoseTTAFold  
variant tool in  
the generative  
diffusion model

Watson et  
al., 2023

---



---

ProtGPT2	Protein sequence	Deep learning-based protein design	Discover new proteins	1. Computational requirements; 2. Hardware memory limitations	Application of GPT2 deep language model	Ferruz et al., 2022
ProteinSGM	Image-like representations of protein structures	Deep learning-based protein design		1. Computational requirements; 2. Cannot analyze inter-chain interactions	Applications of diffusion models and the continuous-time SDE framework	Lee et al., 2023

---

576

577       RoseTTAFold, developed by David Baker laboratory, is a protein structure  
578 prediction tool that can be used for designing proteins with specific functions. It is  
579 based on two schemes (Fig. 7) (Wang et al., 2022): (1) constrained hallucination  
580 (trRosetta or RoseTTAFold) to optimize known functional protein sequence and (2)  
581 inpainting ( $RF_{\text{joint}}$ ) to optimize and fill the gaps between two sequences to ensure a  
582 reasonable final synthesized protein structure.

583       Unlike energy optimization-based structure prediction tools such as RoseTTAFold  
584 and AlphaFold 2, MPNN (Dauparas et al., 2022) predicts and generates candidate  
585 protein sequences for the basic protein skeleton. This tool is useful for oligomers and  
586 protein-protein binding application, as it can ensure the formation of binding forms  
587 between protein skeletons without blindly pursuing the goal of the lowest energy.

588       Researchers have developed RFdiffusion (Watson et al., 2023), a new protein

design diffusion model that combines protein structure prediction methods with an in-depth understanding of protein structure. By using Fine-tuning RoseTTAFold as a denoising network in Denoising Diffusion Probabilistic Models (DDPMs), RFdiffusion can generate a variety of design protein skeletons. Through a multi-step denoising process, a design protein skeleton is gradually produced. The MPNN is then used to generate the protein sequence with the desired structure.

-----**Fig 7**-----

A deep-learning-based 'family-wide hallucination' approach was used to design artificial luciferases with a much higher substrate specificity than native luciferases (Yeh et al., 2023). Based on the successful examples of *de novo* design (Wicky et al., 2022), it is possible to design a new thermostable enzyme that can specifically catalyze indole and form homomultimers. One approach could be using deep network hallucination to generate a wide range of symmetric protein homo-oligomers, specifying the basic number of symmetric units and the length of the oligomer. Then, structure prediction and MPNN could be used to optimize the protein sequence to obtain a "convergent" protein sequence. One approach is Molecular Surface Interaction Fingerprinting (MaSIF) (Gainza et al., 2023), which could be used to predict the protein binding site and encode the protein binding interface to obtain the fingerprint

information of the binding interface. With the help of the binding interface fingerprint, the seeds of complementary binding motifs could be generated in the database, and the obtained seeds could be further optimized to obtain the designed protein structure. One approach could start from the natural sequence, carry out Monte Carlo search in the sequence space, and apply trRosetta structure prediction neural network to score the confidence of the predicted structure. By combining these methods, it may be possible to design a new enzyme that meets the desired specifications.

#### **3.4.5 The immobilization of FMOs**

To meet the economic requirements for enzyme usage, researchers have explored the utilization of immobilized FMOs (Fig. 8). This has been achieved by immobilizing FMOs on silica nanospheres (Biradar et al., 2010) or magnetic nanoparticles (Ramana et al., 2017). Immobilizing at the C-terminal (membrane anchor region) of hFMO3 has been shown to significantly improve its enzymatic thermostability (Gao and Zheng, 2019).

FMOs on electrodes have been studied, immobilized through dialysis or entrapment in a gel cross-linked with bovine serum albumin and glutaraldehyde (Castrignanò et al., 2010). These immobilized FMO electrodes have been used as biosensors for detecting drug and microelements in electrochemistry (Mitsubayashi and Hashimoto, 2000; Mitsubayashi and Hashimoto, 2002; Saito et al., 2008). Different methods of

immobilization have been studied, and immobilization on the gold surface has been found to be more effective in catalysis (Ferrero et al., 2008; Sadeghi et al., 2011; Sadeghi et al., 2010). The electrochemical response of the GC/DDAB/hFMO3 electrode was enhanced using graphene oxide (Castrignanò et al., 2015; Sadeghi et al., 2010).

-----Fig 8-----

#### 3.4.6 Regenerating NADPH in FMO catalysis

In the catalytic cycle of FMOs, the flavin cofactor is reduced by NADPH to initiate subsequent reactions (Eswaramoorthy et al., 2006). However, in extracellular enzyme reaction systems or recombinant cells, the rate of NADPH regeneration may limit the overall reaction rate. By fusing dehydrogenase with FMOs or constructing recombinants to express these enzymes, researchers have utilized different dehydrogenases for NADPH recycling to overcome this limitation (Doukyu et al., 2003; Rioz-Martinez et al., 2011). The choice of dehydrogenase is based on the substrates catalyzed to refill NADPH, such as phosphite, ethanol, and malate (Pan et al., 2023). The usual function of the FMOs-dehydrogenase dimers for NADPH-regeneration is to supply NADPH or NADH in a catalytic environment that lacks them, such as in organic solvents (Doukyu et al., 2003), where there are abundant organic substrates.

### 3.4.7 Metabolic engineering

In the post-genomic era, metabolic engineering has made significant advancements with the aid of genome-scale metabolic network models (GEMs) (Fig. 9B), which represent cellular metabolism mathematically and encompass known chemical reactions, metabolites, associated genes (Orth et al., 2010; Thiele and Palsson, 2010). GEMs can be refined and enhanced through the integration of multi-omics data such as transcriptomics (Colijn et al., 2009; Jensen and Papin, 2011; O'Brien et al., 2013; Ravi and Gunawan, 2021), proteomics (Chen et al., 2021; Dahal et al., 2020; Yizhak et al., 2010), and metabolomics (Filippo et al., 2022; Siddiqui et al., 2018; Yizhak et al., 2010). GEMs have been established as a standard platform for analyzing *E. coli*'s metabolism (Fang et al., 2020), and numerous studies have utilized these models for the engineering of *E. coli* to enhance the production of bioproducts.

Guided by GEMs, *E. coli* was successfully engineered to improve anaerobic functionality of the oxidative tricarboxylic acid cycle, thereby effectively generated the necessary reducing power to facilitate the biosynthesis of 1,4-butanediol (Yim et al., 2011). The flux of central and aromatic amino acid biosynthesis reactions was predicted by in silico response analysis of *E. coli* GEM (Yang et al., 2018), which showed a negative correlation between flux and the synthesis rate of D-phenyllactic acid. And knocking them out could enhance the production of aromatic polymers involving D-phenyllactic acid as a monomer.

Metabolic modifications had been used in some indigo biosynthesis research (Berry et al., 2002). However, there is still much potential for the utilization of GEMs in indigo and indirubin biosynthesis. This could lead to a more rational approach to metabolic engineering and further enhance the yield of these valuable compounds.

#### **3.4.8 Fermentation engineering**

Modern fermentation engineering integrates novel technologies such as fermentation process optimization, amplification, and precise control technology with traditional fermentation engineering, which magnificently improve the production. Optimization of composition in culture medium and coenzyme recycle is attempts to increase production of indigo aside from the endeavor in metabolic engineering.

The toxicity of indole has been shown to decrease indigo production of the producer microorganisms (Murdock et al., 1993; O'Connor et al., 1997). To address this issue, a two-phase culture system was constructed comprising organic solvent-tolerant microorganisms and non-aqueous media, which enriched indole in the hydrophobic organic solvent and reduced its concentration in the medium (Fig. 9C) (Doukyu et al., 2003). Methyl tert-butyl ether and cyclopentyl methyl ether have been found to be the best organic media for a type II flavin-containing monooxygenase (FMO-E) catalysis and a horse liver alcohol dehydrogenase (HLADH) mediated NADPH regeneration (Huang et al., 2019). Coenzyme regeneration systems have also been designed to

recycle NAD(P)H when applying FMOs in non-aqueous media (Doukyu et al., 2003; Huang et al., 2019).

Following cloning mFMO, the production of bio-indigo was optimized using a response surface methodology with a two-level central composite design to demonstrate the interactions between different pairs in the following three factors: tryptophan, yeast extract and sodium chloride, resulting in a 575% increase in production (Han et al., 2008). Large-scale fermentations were also conducted to validate the application potential of FMOs (Han et al., 2011).

-----Fig 9-----

#### **4. The role of FMO in health and disease**

TMA is a volatile tertiary amine derived from daily diet. It is primarily produced in the colon, absorbed into the bloodstream, and converted into trimethylamine oxide (TMAO) in the liver by hFMO3 (Fennema et al., 2016). FMO deficiency or excessive TMAO production is associated with various chronic diseases, such as kidney and coronary artery diseases. FMO plays a crucial role in TMA oxidation and is therefore considered a potential therapeutic target for these diseases.

Trimethylaminuria (TMAU) is a genetic disorder caused by a defect in hFMO3. Individuals affected by TMAU have a severely reduced ability to convert TMA into

TMAO, leading to the accumulated TMA that is excreted through sweat, breath, urine, and other bodily fluids. This causes an unpleasant odor resembling rotten fish (Schmidt and Leroux, 2020), which can harm individuals' psychological well-being. Gene therapy targeting hFMO3 is a promising method for treating TMAU, but limited research data is available (Donato et al., 2021). Additionally, detecting hFMO3 mutations is critical for diagnosing TMAU and expanding the corresponding gene mutation data to promote functional genomics research and explain the disease's pathogenesis. Despite numerous sequencing studies on samples of TMAU worldwide, there is limited research on the association between mutations and the loss of function of hFMO3 (Ameria et al., 2015). In 2021, modeling analysis and experimental verifications were conducted on some hFMO3 mutations considered to be polymorphic or benign. The results suggested that these mutations might damage FMO dynamics, but the causal relationship between these mutations and the pathogenic mechanism could not be confirmed.

TMAO is an atherogenic metabolite that affects platelet reactivity and thrombosis potential. Excessive TMAO production can impose a burden on the cardiovascular system (Janeiro et al., 2018; Senthong et al., 2016; Tang et al., 2019; Xu and Yang, 2021). Targeting metabolic pathways associated with TMAO may be a feasible method for treating atherosclerosis (Yang et al., 2019). TMAO has become an essential pathogenic factor and therapeutic target for treating coronary heart disease. Recent studies have shown that skin fibroblasts, vascular endothelial cells, and adipocyte



progenitors are reprogrammed into myofibroblasts through the TMA-FMO-TMAO-PERK pathway, leading to systemic sclerosis syndrome caused by intestinal dysregulation (Kim et al., 2022). hFMO3, the key enzyme in this pathway, is a potential therapeutic target for specific regulation of its association with upstream chemicals (Zhu et al., 2018).

## 5. Conclusions

The studies related to FMOs and biosynthesis have a long history, spanning over 50 years since the discovery of FMO from pig liver in 1972. The emergence of indigo biosynthesis was in 1989. And the first report of indigo biosynthesis by mFMO was published in 2003. FMOs exhibit some variations across species. In human, hFMOs are located at the ER of hepatocytes towards the cytoplasm, while in bacteria, mFMO is water-soluble. The structures and catalytic mechanisms of FMOs have been extensively studied, and FMOs are increasingly recognized for their crucial roles in biosynthesis, biosensors, diseases, and drug metabolism due to the broad substrate specificity. New biotechnologies, such as directed evolution, *de novo* design, and GEMs, can be used to further improve the yield of FMOs. However, there are still many challenges and questions that need to be addressed. This review aims to provide insights to scientists, who can use this information to further explore the potential of FMOs in various fields.

749    **Abbreviations**

750	AncFMO: ancient mammalian FMO
751	ArBVMO: BVMO cloned from <i>Acinetobacter radioresistens</i>
752	bFMO: FMO cloned from <i>Methylophaga aminisulfidivorans</i>
753	BVMOs: baeyer-villiger monooxygenases
754	CFA: cyclopropane fatty acid
755	cFMO: FMO cloned from <i>Corynebacterium glutamicum</i>
756	CHMOs: cyclohexanone monooxygenases
757	CYP450: cytochrome P450
758	DDAB: didodecylammonium bromide
759	DDPMs: denoising diffusion probabilistic models
760	DMS: dimethylsulfate
761	DMSO: dimethylsulfoxide
762	DTME: dithio-bismaleimidoethane
763	ER: endoplasmic reticulum
764	FAD: flavin adenine dinucleotide
765	FMOs: flavin-containing monooxygenases
766	FMO-E: type II flavin-containing monooxygenase cloned from <i>Rhodococcus jostii</i>
767	RHA1
768	FMO6p: flavin containing dimethylaniline monooxygenase 6, pseudogene
769	FPMOs: flavoprotein monooxygenases
770	GC: glassy carbon
771	GEMs: genome-scale metabolic network models
772	hFMO: human flavin-containing monooxygenase
773	HLADH: horse liver alcohol dehydrogenase
774	IAA: indole-3-acetic acid
775	IPyA: indole-3-pyruvate acid

776 ITCHY: incremental truncation for the creation of hybrid enzymes  
 777 MaSIF: molecular surface interaction fingerprinting  
 778 MBP: maltose-binding protein  
 779 MC: monte carlo  
 780 MCMC: markov chain monte carlo  
 781 mFMO: FMO cloned from *Methylophaga sp.* strain SK1  
 782 MPNN: message-passing neural network  
 783 MSA: multiple sequence alignment  
 784 NADPH: nicotinamide adenine dinucleotide phosphate  
 785 NiFMO: FMO cloned from *Nitrincola laccisaponensis*  
 786 NMOs: N-hydroxylating monooxygenases  
 787 pFMO: pig flavin-containing monooxygenase  
 788 PROSS: protein repair one-stop shop  
 789 PTDH: phosphate dehydrogenase  
 790 PtFMO: FMO cloned from *Polygonum tinctorium*  
 791 RACHITT: random chimeragenesis on transient templates  
 792 rFMO: rabbit FMO  
 793 rmFMO: FMO cloned from *rhesus macaque*  
 794 RM-PCR: random multi-recombinant PCR  
 795 RPR: random priming *in vitro* recombination  
 796 RRI: residue-residue interaction  
 797 SHIPREC: sequence homology-independent protein recombination  
 798 SMFMO: FMO cloned from *S. maltophilia*  
 799 SMO: styrene monooxygenase  
 800 SNPs: single nucleotide polymorphic variants  
 801 StEP: staggered extension process  
 802 TMA: trimethylamine

803 TMAO: trimethylamine oxide  
804 TMAU: trimethylaminuria  
805 TMM: trimethylamine monooxygenase  
806 TRP: tryptophanase  
807 UPO: unspecific peroxygenase  
808 6BrIG: 6,6'-dibromoindigo

809

# 810 **CRedit authorship contribution statement**

811 **Changxin Fan:** Conceptualization, Formal analysis, Investigation, Writing - Original  
812 Draft, Writing - Review & Editing, Visualization. **Ziqi Xie:** Conceptualization,  
813 Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Yijin**  
814 **Li:** Investigation, Writing - Original Draft, Visualization. **Ruihan Zhang:** Investigation,  
815 Writing - Original Draft. **Da Zheng:** Formal analysis, Investigation, Writing - Original  
816 Draft, Visualization. **Jiacheng Shi:** Investigation, Writing - Original Draft. **Yifei Wang:**  
817 Investigation, Writing - Original Draft. **Mingyuan Cheng:** Visualization. **Yu Zhou:**  
818 Supervision. **Yi Zhan:** Supervision, Project administration, Writing - Original Draft.  
819 **Yunjun Yan:** Project administration, Writing - Review & Editing, Funding acquisition.

820

# 821 **Acknowledgements**

822 We thank dockeasy.cn, biorender.com, and scidraw.io for illustrations. This work was  
823 supported by the National Funding for Undergraduate Innovation and Entrepreneurship  
824 Training of China (No.202210487002).

825   **References**

- 826   Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015.
- 827   GROMACS: High performance molecular simulations through multi-level parallelism
- 828   from laptops to supercomputers. *SoftwareX* 1-2, 19-25.
- 829   Alfieri, A., Malito, E., Orru, R., Fraaije, M.W., Mattevi, A., 2008. Revealing the
- 830   moonlighting role of NADP in the structure of a flavin-containing monooxygenase.
- 831   PNAS 105 (18), 6572-6577.
- 832   Ameria, S.P., Jung, H.S., Kim, H.S., Han, S.S., Kim, H.S., Lee, J.H., 2015.
- 833   Characterization of a flavin-containing monooxygenase from *Corynebacterium*
- 834   *glutamicum* and its application to production of indigo and indirubin. *Biotechnol. Lett.*
- 835   37 (8), 1637-1644.
- 836   Anishchenko, I., Pellock, S.J., Chidyausiku, T.M., Ramelot, T.A., Ovchinnikov, S., Hao,
- 837   J., Bafna, K., Norn, C., Kang, A., Bera, A.K., DiMaio, F., Carter, L., Chow, C.M.,
- 838   Montelione, G.T., Baker, D., 2021. *De novo* protein design by deep network
- 839   hallucination. *Nature* 600 (7889), 547-552.
- 840   Ball, S., Bruice, T.C., 1980. Oxidation of amines by a 4a-hydroperoxyflavin. *J. Am.*
- 841   *Chem. Soc.* 102 (21), 6498-6503.
- 842   Beaty, N.B., Ballou, D.P., 1981a. The oxidative half-reaction of liver microsomal
- 843   FAD-containing monooxygenase. *J. Biol. Chem.* 256 (9), 4619-4625.
- 844   Beaty, N.B., Ballou, D.P., 1981b. The reductive half-reaction of liver microsomal

845 FAD-containing monooxygenase. *J. Biol. Chem.* 256 (9), 4611-4618.

846 Beier, A., Bordewick, S., Genz, M., Schmidt, S., van den Bergh, T., Peters, C., Joosten,  
847 H.-J., Bornscheuer, U.T., 2016. Switch in cofactor specificity of a Baeyer – Villiger  
848 monooxygenase. *ChemBioChem* 17 (24), 2312-2315.

849 Berry, A., Dodge, T.C., Pepsin, M., Weyler, W., 2002. Application of metabolic  
850 engineering to improve both the production and use of biotech indigo. *J. Ind. Microbiol.*  
851 *Biotechnol.* 28 (3), 127-133.

852 Biradar, A.A., Biradar, A.V., Asefa, T., 2010. Entrapping flavin-containing  
853 monooxygenase on corrugated silica nanospheres and their recyclable biocatalytic  
854 activities. *ChemCatChem* 2 (8), 1004-1010.

855 Boronin, A., Tsoi, T., Kosheleva, I., Arinbasarov, M., Adanin, V., 1989. Cloning of  
856 *Pseudomonas putida* genes responsible for the primary stages of oxidation of  
857 naphthalene in *Escherichia coli* cells. *Genetika* 25 (2), 226-237.

858 Bortolussi, S., Catucci, G., Gilardi, G., Sadeghi, S.J., 2021. N- and S-oxygenation  
859 activity of truncated human flavin-containing monooxygenase 3 and its common  
860 polymorphic variants. *Arch. Biochem. Biophys.* 697, 108663.

861 Boyken, S.E., Chen, Z., Groves, B., Langan, R.A., Oberdorfer, G., Ford, A., Gilmore,  
862 J.M., Xu, C., DiMaio, F., Pereira, J.H., Sankaran, B., Seelig, G., Zwart, P.H., Baker, D.,  
863 2016. De novo design of protein homo-oligomers with modular hydrogen-bond  
864 network-mediated specificity. *Science* 352 (6286), 680-687.

865 Burnett, V.L., Lawton, M.P., Philpot, R.M., 1994. Cloning and sequencing of  
 866 flavin-containing monooxygenases FMO<sub>3</sub> and FMO<sub>4</sub> from rabbit and characterization of  
 867 FMO<sub>3</sub>. *J. Biol. Chem.* 269 (19), 14314-14322.

868 Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. *PCR*  
 869 *Meth. Appl.* 2 (1), 28-33.

870 Cahn, J.K.B., Brinkmann-Chen, S., Arnold, F.H., 2018. Enzyme Nicotinamide Cofactor  
 871 Specificity Reversal Guided by Automated Structural Analysis and Library Design, in:  
 872 Jensen, M.K., Keasling, J.D. (Eds.), *Synthetic Metabolic Pathways: Methods and*  
 873 *Protocols*. Springer New York, New York, pp. 15-26.

874 Cao, X., Yang, H., Shang, C., Ma, S., Liu, L., Cheng, J., 2019. The roles of auxin  
 875 biosynthesis YUCCA gene family in plants. *Int. J. Mol. Sci.* 20 (24), 6343.

876 Cashman, J.R., Akerman, B.R., Forrest, S.M., Treacy, E.P., 2000. Population-specific  
 877 polymorphisms of the human FMO<sub>3</sub> gene: Significance for detoxication. *Drug Metab.*  
 878 *Dispos.* 28 (2), 169-173.

879 Castrignanò, S., Gilardi, G., Sadeghi, S.J., 2015. Human flavin-containing  
 880 monooxygenase 3 on graphene oxide for drug metabolism screening. *Anal. Chem.* 87  
 881 (5), 2974-2980.

882 Castrignanò, S., Sadeghi, S.J., Gilardi, G., 2010. Electro-catalysis by immobilised  
 883 human flavin-containing monooxygenase isoform 3 (hFMO<sub>3</sub>). *Anal. Bioanal. Chem.*  
 884 398 (3), 1403-1409.

885 Catucci, G., Aramini, D., Sadeghi, S.J., Gilardi, G., 2020. Ligand stabilization and effect  
 886 on unfolding by polymorphism in human flavin-containing monooxygenase 3. *Int. J.*  
 887 *Biol. Macromol.* 162, 1484-1493.

888 Catucci, G., Gao, C., Sadeghi, S.J., Gilardi, G., 2017. Chemical applications of Class B  
 889 flavoprotein monooxygenases. *Rend. Lincei-Sci. Fis. Nat.* 28, 195-206.

890 Catucci, G., Turella, S., Cheropkina, H., De Angelis, M., Gilardi, G., Sadeghi, S.J., 2022.  
 891 Green production of indigo and indirubin by an engineered Baeyer–Villiger  
 892 monooxygenase. *Biocatal. Agric. Biotechnol.* 44, 102458.

893 Ceccoli, R.D., Bianchi, D.A., Rial, D.V., 2014. Flavoprotein monooxygenases for  
 894 oxidative biocatalysis: recombinant expression in microbial hosts and applications.  
 895 *Front. Microbiol.* 5, 25.

896 Chen, H., Ma, L., Dai, H., Fu, Y., Wang, H., Zhang, Y., 2022. Advances in rational  
 897 protein engineering toward functional architectures and their applications in food  
 898 science. *J. Agric. Food Chem.* 70 (15), 4522-4533.

899 Chen, M., Guan, B., Xu, H., Yu, F., Zhang, T., Wu, B., 2019. The molecular mechanism  
 900 regulating diurnal rhythm of flavin-containing monooxygenase 5 in mouse liver. *Drug*  
 901 *Metab. Dispos.* 47 (11), 1333-1342.

902 Chen, T., Wang, X., Zhuang, L., Shao, A., Lu, Y., Zhang, H., 2021. Development and  
 903 optimization of a microbial co-culture system for heterologous indigo biosynthesis.  
 904 *Microb. Cell Fact.* 20 (1), 1-11.



905 Chen, Y., Patel, N.A., Crombie, A., Scrivens, J.H., Murrell, J.C., 2011. Bacterial  
 906 flavin-containing monooxygenase is trimethylamine monooxygenase. PNAS 108 (43),  
 907 17791-17796.

908 Chen, Y., Pelt - KleinJan, E., Olst, B., Douwenga, S., Boeren, S., Bachmann, H.,  
 909 Molenaar, D., Nielsen, J., Teusink, B., 2021. Proteome constraints reveal targets for  
 910 improving microbial fitness in nutrient - rich environments. Mol. Syst. Biol. 17 (4),  
 911 e10093.

912 Cheng, L., Yin, S., Chen, M., Sun, B., Hao, S., Wang, C., 2016. Enhancing indigo  
 913 production by over-expression of the styrene monooxygenase in *Pseudomonas putida*.  
 914 Curr. Microbiol. 73 (2), 248-254.

915 Cherrington, N.J., Cao, Y., Cherrington, J.W., Rose, R.L., Hodgson, E., 1998.  
 916 Physiological factors affecting protein expression of flavin-containing monooxygenases  
 917 1, 3 and 5. Xenobiotica 28 (7), 673-682.

918 Cho, D.H., Kim, H.J., Oh, S.J., Hwang, J.H., Shin, N., Bhatia, S.K., Yoon, J.J., Jeon,  
 919 J.M., Yang, Y.H., 2023. Strategy for efficiently utilizing *Escherichia coli* cells producing  
 920 isobutanol by combining isobutanol and indigo production systems. J. Biotechnol. 367,  
 921 62-70.

922 Cho, H.J., Cho, H.Y., Kim, K.J., Kim, M.H., Kim, S.W., Kang, B.S., 2011. Structural  
 923 and functional analysis of bacterial flavin-containing monooxygenase reveals its  
 924 ping-pong-type reaction mechanism. J. Struct. Biol. 175 (1), 39-48.

925 Choi, H.S., Kim, J.K., Cho, E.H., Kim, Y.C., Kim, J.I., Kim, S.W., 2003. A novel  
 926 flavin-containing monooxygenase from *Methylophaga* sp. strain SK1 and its indigo  
 927 synthesis in *Escherichia coli*. Biochem. Biophys. Res. Commun. 306 (4), 930-936.

928 Choi, K.Y., 2020. A review of recent progress in the synthesis of bio-indigoids and their  
 929 biologically assisted end-use applications. Dyes Pigm. 181, 108570.

930 Chowdhury, R., Maranas, C.D., 2020. From directed evolution to computational enzyme  
 931 engineering—A review. AIChE J. 66 (3), e16847.

932 Coco, W.M., 2003. RACHITT: Gene family shuffling by Random Chimeragenesis on  
 933 transient templates. Methods Mol. Biol. 231, 111-127.

934 Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.-Y.,  
 935 Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting expression data with  
 936 metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production.  
 937 PLoS Comput. Biol. 5 (8), e1000489.

938 Dahal, S., Yurkovich, J.T., Xu, H., Palsson, B.O., Yang, L., 2020. Synthesizing systems  
 939 biology knowledge from omics using genome - scale models. Proteomics 20 (17-18),  
 940 1900282.

941 Dai, C., Ma, Q., Li, Y., Zhou, D., Yang, B., Qu, Y., 2019. Application of an efficient  
 942 indole oxygenase system from *Cupriavidus* sp. SHE for indigo production. Bioprocess  
 943 Biosyst. Eng. 42, 1963-1971.

944 Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky,

945 B.I.M., Courbet, A., de Haas, R.J., Bethel, N., Leung, P.J.Y., Huddy, T.F., Pellock, S.,  
 946 Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A.K.,  
 947 King, N.P., Baker, D., 2022. Robust deep learning–based protein sequence design using  
 948 ProteinMPNN. *Science* 378 (6615), 49-56.

949 Davey, J.A., Damry, A.M., Goto, N.K., Chica, R.A., 2017. Rational design of proteins  
 950 that exchange on functional timescales. *Nat. Chem. Biol.* 13 (12), 1280-1285.

951 Dehouck, Y., Kwasigroch, J.M., Gilis, D., Rooman, M., 2011. PoPMuSiC 2.1: a web  
 952 server for the estimation of protein stability changes upon mutation and sequence  
 953 optimality. *BMC Bioinform.* 12 (1), 151.

954 Ding, W., Nakai, K., Gong, H., 2022. Protein design via deep learning. *Brief. Bioinform.*  
 955 23 (3), bbac102.

956 Dolphin, C.T., Cullingford, T.E., Shephard, E.A., Smith, R.L., Phillips, I.R., 1996.  
 957 Differential developmental and tissue-specific regulation of expression of the genes  
 958 encoding three members of the flavin-containing monooxygenase family of man, FMO1,  
 959 FMO3 and FM04. *Eur. J. Biochem.* 235 (3), 683-689.

960 Donato, L., Alibrandi, S., Scimone, C., Castagnetti, A., Rao, G., Sidoti, A., D'Angelo,  
 961 R., 2021. Gut-brain axis cross-talk and limbic disorders as biological basis of secondary  
 962 TMAU. *J. Pers. Med.* 11 (2), 87.

963 Doukyu, N., Nakano, T., Okuyama, Y., Aono, R., 2002. Isolation of an *Acinetobacter* sp.  
 964 ST-550 which produces a high level of indigo in a water-organic solvent two-phase

965 system containing high levels of indole. *Appl. Microbiol. Biotechnol.* 58 (4), 543-546.

966 Doukyu, N., Toyoda, K., Aono, R., 2003. Indigo production by *Escherichia coli*

967 carrying the phenol hydroxylase gene from *Acinetobacter* sp. strain ST-550 in a

968 water-organic solvent two-phase system. *Appl. Microbiol. Biotechnol.* 60 (6), 720-725.

969 Du, J., Yang, D., Luo, Z.W., Lee, S.Y., 2018. Metabolic engineering of *Escherichia coli*

970 for the production of indirubin from glucose. *J. Biotechnol.* 267, 19-28.

971 Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y.T., Beauchamp, K.A.,

972 Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., Wiewiora, R.P., Brooks, B.R.,

973 Pande, V.S., 2017. OpenMM 7: Rapid development of high performance algorithms for

974 molecular dynamics. *PLoS Comput. Biol.* 13 (7), e1005659.

975 Ensley, B.D., Ratzkin, B.J., Osslund, T.D., Simon, M.J., Wackett, L.P., Gibson, D.T.,

976 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the

977 biosynthesis of indigo. *Science* 222 (4620), 167-169.

978 Eswaramoorthy, S., Bonanno, J.B., Burley, S.K., Swaminathan, S., 2006. Mechanism of

979 action of a flavin-containing monooxygenase. *PNAS* 103 (26), 9832-9837.

980 Fabara, A.N., Fraaije, M.W., 2020. An overview of microbial indigo-forming enzymes.

981 *Appl. Microbiol. Biotechnol.* 104 (3), 925-933.

982 Fabara, A.N., Fraaije, M.W., 2020. Production of indigo through the use of a

983 dual-function substrate and a bifunctional fusion enzyme. *Enzyme Microb. Technol.* 142,

984 109692.

985 Fang, X., Lloyd, C.J., Palsson, B.O., 2020. Reconstructing organisms in silico:  
 986 genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* 18 (12),  
 987 731-743.

988 Fennema, D., Phillips, I.R., Shephard, E.A., 2016. Trimethylamine and trimethylamine  
 989 N-oxide, a flavin-containing monooxygenase 3 (FMO3)-mediated host-microbiome  
 990 metabolic axis implicated in health and disease. *Drug Metab. Dispos.* 44 (11),  
 991 1839-1850.

992 Ferreira, P., Fernandes, P.A., Ramos, M.J., 2022. Modern computational methods for  
 993 rational enzyme engineering. *Chem Catal.* 2 (10), 2481-2498.

994 Ferrero, V.E.V., Andolfi, L., Di Nardo, G., Sadeghi, S.J., Fantuzzi, A., Cannistraro, S.,  
 995 Gilardi, G., 2008. Protein and electrode engineering for the covalent immobilization of  
 996 P450 BMP on gold. *Anal. Chem.* 80 (22), 8438-8446.

997 Ferruz, N., Schmidt, S., Höcker, B., 2022. ProtGPT2 is a deep unsupervised language  
 998 model for protein design. *Nat. Commun.* 13 (1), 4348.

999 Filippo, M.D., Pescini, D., Galuzzi, B.G., Bonanomi, M., Gaglio, D., Mangano, E.,  
 1000 Consolandi, C., Alberghina, L., Vanoni, M., Damiani, C., 2022. INTEGRATE:  
 1001 Model-based multi-omics data integration to characterize multi-level metabolic  
 1002 regulation. *PLoS Comput. Biol.* 18 (2), e1009337.

1003 Fiorentini, F., Hatzl, A.-M., Schmidt, S., Savino, S., Glieder, A., Mattevi, A., 2018. The  
 1004 extreme structural plasticity in the CYP153 subfamily of P450s directs development of

1005 designer hydroxylases. *Biochemistry* 57 (48), 6701-6714.  
 1006 Furnes, B., Schlenk, D., 2004. Evaluation of xenobiotic N-and S-oxidation by variant  
 1007 flavin-containing monooxygenase 1 (FMO1) enzymes. *Toxicol. Sci.* 78 (2), 196-203.  
 1008 Fürst, M.J.L.J., Boonstra, M., Bandstra, S., Fraaije, M.W., 2019a. Stabilization of  
 1009 cyclohexanone monooxygenase by computational and experimental library design.  
 1010 *Biotechnol. Bioeng.* 116 (9), 2167-2177.  
 1011 Fürst, M.J.L.J., Fiorentini, F., Fraaije, M.W., 2019b. Beyond active site residues: overall  
 1012 structural dynamics control catalysis in flavin-containing and heme-containing  
 1013 monooxygenases. *Curr. Opin. Struct. Biol.* 59, 29-37.  
 1014 Gainza, P., Wehrle, S., Van Hall-Beauvais, A., Marchand, A., Scheck, A., Harteveld, Z.,  
 1015 Buckley, S., Ni, D., Tan, S., Sverrisson, F., Goverde, C., Turelli, P., Raclot, C., Teslenko,  
 1016 A., Pacesa, M., Rosset, S., Georgeon, S., Marsden, J., Petruzzella, A., Liu, K., Xu, Z.,  
 1017 Chai, Y., Han, P., Gao, G.F., Oricchio, E., Fierz, B., Trono, D., Stahlberg, H., Bronstein,  
 1018 M., Correia, B.E., 2023. *De novo* design of protein interactions with learned surface  
 1019 fingerprints. *Nature* 617 (7959), 176-184.  
 1020 Gao, C., Catucci, G., Di Nardo, G., Gilardi, G., Sadeghi, S.J., 2016. Human  
 1021 flavin-containing monooxygenase 3: Structural mapping of gene polymorphisms and  
 1022 insights into molecular basis of drug binding. *Gene* 593 (1), 91-99.  
 1023 Gao, C., Zheng, T., 2019. Drug metabolite synthesis by immobilized human FMO<sub>3</sub> and  
 1024 whole cell catalysts. *Microb. Cell Fact.* 18 (1), 133.

1025 Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym,  
 1026 O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R.L., Aharoni, A., Silman, I.,  
 1027 Sussman, J.L., Tawfik, D.S., Fleishman, S.J., 2016. Automated structure- and  
 1028 sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell.*  
 1029 63 (2), 337-346.

1030 Gonzalez-Perez, D., Garcia-Ruiz, E., Ruiz-Dueñas, F.J., Martinez, A.T., Alcalde, M.,  
 1031 2014. Structural determinants of oxidative stabilization in an evolved versatile  
 1032 peroxidase. *ACS Catal.* 4 (11), 3891-3901.

1033 Goodsell, D.S., Sanner, M.F., Olson, A.J., Forli, S., 2021. The Auto Docksuite at 30.  
 1034 *Protein Sci.* 30 (1), 31-43.

1035 Goris, M., Cea-Rama, I., Puntervoll, P., Ree, R., Almendral, D., Sanz-Aparicio, J.,  
 1036 Ferrer, M., Bjerga, G.E.K., 2023. Increased thermostability of an engineered  
 1037 flavin-containing monooxygenase to remediate trimethylamine in fish protein  
 1038 hydrolysates. *Appl. Environ. Microbiol.* 89 (6), e00390-23.

1039 Groeneveld, M., Van Beek, H., Duetz, W., Fraaije, M., 2016. Identification of a novel  
 1040 oxygenase capable of regiospecific hydroxylation of D-limonene into (+)-trans-carveol.  
 1041 *Tetrahedron* 72 (46), 7263-7267.

1042 Guffy, S.L., Teets, F.D., Langlois, M.I., Kuhlman, B., 2018. Protocols for  
 1043 requirement-driven protein design in the rosetta modeling program. *J. Chem. Inf. Model.*  
 1044 58 (5), 895-901.

1045 Hallen, M.A., Martin, J.W., Ojewole, A., Jou, J.D., Lowegard, A.U., Frenkel, M.S.,  
 1046 Gainza, P., Nisonoff, H.M., Mukund, A., Wang, S.Y., Holt, G.T., Zhou, D., Dowd, E.,  
 1047 Donald, B.R., 2018. OSPREY 3.0: Open-source protein redesign for you, with powerful  
 1048 new features. *J. Comput. Chem.* 39 (30), 2494-2507.  
 1049 Ham, S., Cho, D.-H., Oh, S.J., Hwang, J.H., Kim, H.J., Shin, N., Ahn, J., Choi, K.-Y.,  
 1050 Bhatia, S.K., Yang, Y.-H., 2023. Enhanced production of bio-indigo in engineered  
 1051 *Escherichia coli*, reinforced by cyclopropane-fatty acid-acyl-phospholipid synthase  
 1052 from psychrophilic *Pseudomonas* sp. B14-6. *J. Biotechnol.* 366, 1-9.  
 1053 Han, A., Robinson, R.M., Badieyan, S., Ellerbrock, J., Sobrado, P., 2013. Tryptophan-47  
 1054 in the active site of *Methylophaga* sp. strain SK1 flavin-monooxygenase is important for  
 1055 hydride transfer. *Arch. Biochem. Biophys.* 532 (1), 46-53.  
 1056 Han, G.H., Bang, S.E., Babu, B.K., Chang, M., Shin, H.J., Kim, S.W., 2011. Bio-indigo  
 1057 production in two different fermentation systems using recombinant *Escherichia coli*  
 1058 cells harboring a flavin-containing monooxygenase gene (fmo). *Process Biochem.* 46  
 1059 (3), 788-791.  
 1060 Han, G.H., Gim, G.H., Kim, W., Seo, S.I., Kim, S.W., 2012. Enhanced indirubin  
 1061 production in recombinant *Escherichia coli* harboring a flavin-containing  
 1062 monooxygenase gene by cysteine supplementation. *J. Biotechnol.* 164 (2), 179-187.  
 1063 Han, G.H., Shin, H.J., Kim, S.W., 2008. Optimization of bio-indigo production by  
 1064 recombinant *E. coli* harboring fmo gene. *Enzyme Microb. Technol.* 42 (7), 617-623.



1065 Hao, D.C., Chen, S.L., Mu, J., Xiao, P.G., 2009. Molecular phylogeny, long-term  
 1066 evolution, and functional divergence of flavin-containing monooxygenases. *Genetica*  
 1067 137 (2), 173-187.

1068 Hartevelde, Z., Bonet, J., Rosset, S., Yang, C., Sesterhenn, F., Correia, B.E., 2022. A  
 1069 generic framework for hierarchical *de novo* protein design. *PNAS* 119 (43),  
 1070 e2206111119.

1071 Henderson, M.C., Krueger, S.K., Siddens, L.K., Stevens, J.F., Williams, D.E., 2004.  
 1072 S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton by  
 1073 human lung flavin-containing monooxygenase 2. *Biochem. Pharmacol.* 68 (5), 959-967.

1074 Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I.R., Shephard, E.A., 2004.  
 1075 Organization and evolution of the flavin-containing monooxygenase genes of human  
 1076 and mouse: identification of novel gene and pseudogene clusters. *Pharmacogenetics* 14  
 1077 (2), 117-130.

1078 Honda, K., Yamashita, S., Nakagawa, H., Sameshima, Y., Omasa, T., Kato, J., Ohtake,  
 1079 H., 2008. Stabilization of water-in-oil emulsion by *Rhodococcus opacus* B-4 and its  
 1080 application to biotransformation. *Appl. Microbiol. Biotechnol.* 78 (5), 767-773.

1081 Hsu, T.M., Welner, D.H., Russ, Z.N., Cervantes, B., Prathuri, R.L., Adams, P.D., Dueber,  
 1082 J.E., 2018. Employing a biochemical protecting group for a sustainable indigo dyeing  
 1083 strategy. *Nat. Chem. Biol.* 14 (3), 256-261.

1084 Hu, W.S., Bowman, E.H., Delviks, K.A., Pathak, V.K., 1997. Homologous

1085 recombination occurs in a distinct retroviral subpopulation and exhibits high negative  
 1086 interference. J. Virol. 71 (8), 6028-6036.

1087 Huang, L., Aalbers, F.S., Tang, W., Röllig, R., Fraaije, M.W., Kara, S., 2019.  
 1088 Convergent cascade catalyzed by monooxygenase–alcohol dehydrogenase fusion  
 1089 applied in organic media. ChemBioChem 20 (13), 1653-1658.

1090 Jung, H.S., Jung, H.B., Kim, H.S., Kim, C.G., Lee, J.H., 2018. Protein engineering of  
 1091 flavin-containing monooxygenase from *Corynebacterium glutamicum* for improved  
 1092 production of indigo and indirubin. Korean Soc. Life Sci. 28 (6), 656-662.

1093 Inoue, S., Morita, R., Minami, Y., 2021. An indigo-producing plant, *Polygonum*  
 1094 *tinctorium*, possesses a flavin containing monooxygenase capable of oxidizing indole.  
 1095 Biochem. Biophys. Res. Commun. 543, 95.

1096 Jain, A.N., 2003. Surflex: Fully Automatic Flexible molecular docking using a  
 1097 molecular similarity-based search engine. J. Med. Chem. 46 (4), 499-511.

1098 Janeiro, M.H., Ramírez, M.J., Milagro, F.I., Martínez, J.A., Solas, M., 2018. Implication  
 1099 of trimethylamine n-oxide (TMAO) in disease: Potential biomarker or new therapeutic  
 1100 target. Nutrients 10 (10), 1398.

1101 Janmohamed, A., Hernandez, D., Phillips, I.R., Shephard, E.A., 2004. Cell-, tissue-,  
 1102 sex- and developmental stage-specific expression of mouse flavin-containing  
 1103 monooxygenases (Fmos). Biochem. Pharmacol. 68 (1), 73-83.

1104 Jensen, C.N., Cartwright, J., Ward, J., Hart, S., Turkenburg, J.P., Ali, S.T., Allen, M.J.,

1105 Grogan, G., 2012. A flavoprotein monooxygenase that catalyses a Baeyer-Villiger  
 1106 reaction and thioether oxidation using NADH as the nicotinamide cofactor.  
 1107 ChemBioChem 13 (6), 872-878.

1108 Jensen, P.A., Papin, J.A., 2011. Functional integration of a metabolic network model  
 1109 and expression data without arbitrary thresholding. Bioinformatics 27 (4), 541-547.

1110 Jia, L., Lu, W., Hu, D., Feng, M., Wang, A., Wang, R., Sun, H., Wang, P., Xia, Q., Ma,  
 1111 S., 2023. Genetically engineered blue silkworm capable of synthesizing natural blue  
 1112 pigment. Int. J. Biol. Macromol. 235, 123863.

1113 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,  
 1114 Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C.,  
 1115 Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler,  
 1116 J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M.,  
 1117 Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W.,  
 1118 Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure  
 1119 prediction with AlphaFold. Nature 596 (7873), 583-589.

1120 Kalmankar, N.V., Pavalam, M., Indrakumar, S., Srinivasan, N., Sowdhamini, R., 2022.  
 1121 DSDBASE 2.0: updated version of DiSulphide dataBASE, a database on disulphide  
 1122 bonds in proteins. Database 2022, baac005.

1123 Kikuchi, M., Ohnishi, K., Harayama, S., 2000. An effective family shuffling method  
 1124 using single-stranded DNA. Gene 243 (1), 133-137.

1125 Kim, J., Lee, J., Lee, P.G., Kim, E.J., Kroutil, W., Kim, B.G., 2019. Elucidating  
 1126 cysteine-assisted synthesis of indirubin by a flavin-containing monooxygenase. ACS  
 1127 Catal. 9 (10), 9539-9544.

1128 Kim, J., Lee, P.-g., Jung, E.-o., Kim, B.-G., 2018. In vitro characterization of  
 1129 CYP102G4 from *Streptomyces cattleya*: A self-sufficient P450 naturally producing  
 1130 indigo. BBA-Proteins Proteom. 1866 (1), 60-67.

1131 Kim, J., Park, S., Min, D., Kim, W., 2021. Comprehensive survey of recent drug  
 1132 discovery using deep learning. Int. J. Mol. Sci. 22 (18).

1133 Kim, S.J., Bale, S., Verma, P., Wan, Q., Ma, F., Gudjonsson, J.E., Hazen, S.L., Harms,  
 1134 P.W., Tsou, P.S., Khanna, D., Tsoi, L.C., Gupta, N., Ho, K.J., Varga, J., 2022. Gut  
 1135 microbe-derived metabolite trimethylamine N-oxide activates PERK to drive fibrogenic  
 1136 mesenchymal differentiation. iScience 25 (7), 104669.

1137 Krueger, S.K., Martin, S.R., Yueh, M.-F., Pereira, C.B., Williams, D.E., 2002a.  
 1138 Identification of active flavin-containing monooxygenase isoform 2 in human lung and  
 1139 characterization of expressed protein. Drug Metab. Dispos. 30 (1), 34-41.

1140 Krueger, S.K., Williams, D.E., 2005. Mammalian flavin-containing monooxygenases:  
 1141 structure/function, genetic polymorphisms and role in drug metabolism. Pharmacol.  
 1142 Ther. 106 (3), 357-387.

1143 Krueger, S.K., Williams, D.E., Yueh, M.-F., Martin, S.R., Hines, R.N., Raucy, J.L.,  
 1144 Dolphin, C.T., Shephard, E.A., Phillips, I.R., 2002b. Genetic polymorphisms of

1145 flavin-containing monooxygenase (FMO). *Drug Metab. Rev.* 34 (3), 523-532.

1146 Land, H., Humble, M.S., 2018. YASARA: A Tool to Obtain Structural Guidance in  
 1147 Biocatalytic Investigations, in: Bornscheuer, U.T., Höhne, M. (Eds.), *Protein*  
 1148 *Engineering: Methods and Protocols*. Springer New York, New York, pp. 43-67.

1149 Larsen-Su, S., Krueger, S.K., Yueh, M.F., Lee, M.Y., Shehin, S.E., Hines, R.N.,  
 1150 Williams, D.E., 1999. Flavin-containing monooxygenase isoform 2: developmental  
 1151 expression in fetal and neonatal rabbit lung. *J. Biochem. Mol. Toxicol.* 13 (3-4),  
 1152 187-193.

1153 Lattard, V., Longin-Sauvageon, C., Benoit, E., 2003. Cloning, sequencing and tissue  
 1154 distribution of rat flavin-containing monooxygenase 4: two different forms are produced  
 1155 by tissue-specific alternative splicing. *Mol. Pharmacol.* 63 (1), 253-261.

1156 Lawton, M.P., Cashman, J.R., Cresteil, T., Dolphin, C.T., Elfarra, A.A., Hines, R.N.,  
 1157 Hodgson, E., Kimura, T., Ozols, J., Phillips, I.R., Philpot, R.M., Poulsen, L.L., Rettie,  
 1158 A.E., Shephard, E.A., Williams, D.E., Ziegler, D.M., 1994. A nomenclature for the  
 1159 mammalian flavin-containing monooxygenase gene family based on amino acid  
 1160 sequence identities. *Arch. Biochem. Biophys.* 308 (1), 254-257.

1161 Lee, J., Kim, J., Song, J.E., Song, W.S., Kim, E.J., Kim, Y.G., Jeong, H.J., Kim, H.R.,  
 1162 Choi, K.Y., Kim, B.G., 2021. Production of Tyrian purple indigoid dye from tryptophan  
 1163 in *Escherichia coli*. *Nat. Chem. Biol.* 17 (1), 104-112.

1164 Lee, J.S., Kim, J., Kim, P.M., 2023. Score-based generative modeling for de novo

1165 protein design. *Nat. Comput. Sci.* 3 (5), 382-392.

1166 Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F.,

1167 Aprahamian, M., Baker, D., Barlow, K.A., Barth, P., Basanta, B., Bender, B.J.,

1168 Blacklock, K., Bonet, J., Boyken, S.E., Bradley, P., Bystroff, C., Conway, P., Cooper, S.,

1169 Correia, B.E., Coventry, B., Das, R., De Jong, R.M., DiMaio, F., Dsilva, L., Dunbrack,

1170 R., Ford, A.S., Frenz, B., Fu, D.Y., Geniesse, C., Goldschmidt, L., Gowthaman, R., Gray,

1171 J.J., Gront, D., Guffy, S., Horowitz, S., Huang, P.-S., Huber, T., Jacobs, T.M., Jeliaskov,

1172 J.R., Johnson, D.K., Kappel, K., Karanicolas, J., Khakzad, H., Khar, K.R., Khare, S.D.,

1173 Khatib, F., Khramushin, A., King, I.C., Kleffner, R., Koepnick, B., Kortemme, T.,

1174 Kuenze, G., Kuhlman, B., Kuroda, D., Labonte, J.W., Lai, J.K., Lapidoth, G.,

1175 Leaver-Fay, A., Lindert, S., Linsky, T., London, N., Lubin, J.H., Lyskov, S., Maguire, J.,

1176 Malmström, L., Marcos, E., Marcu, O., Marze, N.A., Meiler, J., Moretti, R., Mulligan,

1177 V.K., Nerli, S., Norn, C., Ó'Conchúir, S., Ollikainen, N., Ovchinnikov, S., Pacella, M.S.,

1178 Pan, X., Park, H., Pavlovicz, R.E., Pethe, M., Pierce, B.G., Pilla, K.B., Raveh, B.,

1179 Renfrew, P.D., Burman, S.S.R., Rubenstein, A., Sauer, M.F., Scheck, A., Schief, W.,

1180 Schueler-Furman, O., Sedan, Y., Sevy, A.M., Sgourakis, N.G., Shi, L., Siegel, J.B., Silva,

1181 D.-A., Smith, S., Song, Y., Stein, A., Szegedy, M., Teets, F.D., Thyme, S.B., Wang,

1182 R.Y.-R., Watkins, A., Zimmerman, L., Bonneau, R., 2020. Macromolecular modeling

1183 and design in Rosetta: recent methods and frameworks. *Nat. Methods.* 17 (7), 665-680.

1184 Lenug, D.W., Che, E., Goeddel, D.V., 1989. A method for random mutagenesis of a

1185 defined DNA segment using a modified polymerase chain reaction. *Technique JMCMB*  
 1186 1, 11-15.

1187 Levin, S.A., 1992. *Mathematics and biology: The interface, challenges and*  
 1188 *opportunities.*

1189 Linke, J.A., Rayat, A., Ward, J.M., 2023. Production of indigo by recombinant bacteria.  
 1190 *Bioresour. Bioprocess.* 10 (1), 20.

1191 Linke, J.A.A., Rayat, A., Ward, J.M.M., 2023. Production of indigo by recombinant  
 1192 bacteria. *Bioresour. Bioprocess.* 10 (1), 20.

1193 Lipsh-Sokolik, R., Listov, D., Fleishman, S.J., 2021. The AbDesign computational  
 1194 pipeline for modular backbone assembly and design of binders and enzymes. *Protein*  
 1195 *Sci.* 30 (1), 151-159.

1196 Liu, H., Chen, Q., 2022. Computational protein design with data-driven approaches:  
 1197 Recent developments and perspectives. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* 13  
 1198 (3), e1646.

1199 Lončar, N., Fiorentini, F., Bailleul, G., Savino, S., Romero, E., Mattevi, A., Fraaije,  
 1200 M.W., 2019a. Characterization of a thermostable flavin-containing monooxygenase  
 1201 from *Nitrocola laciaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* 103 (4),  
 1202 1755-1764.

1203 Lončar, N., van Beek, H.L., Fraaije, M.W., 2019b. Structure-based redesign of a  
 1204 self-sufficient flavin-containing monooxygenase towards indigo production. *Int. J. Mol.*

1205 Sci. 20 (24), 6148.

1206 Ma, Q., Zhang, X., Qu, Y., 2018. Biodegradation and biotransformation of indole:  
 1207 advances and perspectives. *Front. Microbiol.* 9: 2625

1208 Mitsubayashi, K., Hashimoto, Y., 2000. Development of a gas-phase biosensor for  
 1209 trimethylamine using a flavin-containing monooxygenase 3. *Electrochemistry* 68 (11),  
 1210 901-903.

1211 Mitsubayashi, K., Hashimoto, Y., 2002. Bioelectronic nose for methyl mercaptan vapor  
 1212 using xenobiotic metabolizing enzyme: flavin-containing monooxygenase. *Sens.*  
 1213 *Actuator B-Chem.* 83 (1), 35-40.

1214 Molina-Espeja, P., Garcia-Ruiz, E., Gonzalez-Perez, D., Ullrich, R., Hofrichter, M.,  
 1215 Alcalde, M., 2014. Directed evolution of unspecific peroxygenase from *Agrocybe*  
 1216 *aegerita*. *Appl. Environ. Microbiol.* 80 (11), 3496-3507.

1217 Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S.,  
 1218 Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective  
 1219 receptor flexibility. *J. Comput. Chem.* 30 (16), 2785-2791.

1220 Murdock, D., Ensley, B.D., Serdar, C., Thalen, M., 1993. Construction of metabolic  
 1221 operons catalyzing the *de novo* biosynthesis of indigo in *Escherichia coli*. *Nat.*  
 1222 *Biotechnol.* 11 (3), 381-386.

1223 Nicoll, C.R., Bailleul, G., Fiorentini, F., Mascotti, M.L., Fraaije, M.W., Mattevi, A.,  
 1224 2020. Ancestral-sequence reconstruction unveils the structural basis of function in



1225 mammalian FMOs. Nat. Struct. Mol. Biol. 27 (1), 14-24.

1226 O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013.

1227 Genome - scale models of metabolism and gene expression extend and refine growth

1228 phenotype prediction. Mol. Syst. Biol. 9 (1), 693-693.

1229 O'Connor, K.E., Dobson, A., Hartmans, S., 1997. Indigo formation by microorganisms

1230 expressing styrene monooxygenase activity. Appl. Environ. Microbiol. 63 (11),

1231 4287-4291.

1232 Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat.

1233 Biotechnol. 28 (3), 245-248.

1234 Pan, Z., Tao, D., Ren, M., Cheng, L., 2023. A combinational optimization method for

1235 efficient production of indigo by the recombinant *Escherichia coli* with expression of

1236 monooxygenase and malate dehydrogenase. Foods 12 (3), 502.

1237 Pantazes, R.J., Grisewood, M.J., Li, T., Gifford, N.P., Maranas, C.D., 2015. The

1238 Iterative Protein Redesign and Optimization (IPRO) suite of programs. J. Comput.

1239 Chem. 36 (4), 251-263.

1240 Pathak, H., Madamwar, D., 2010. Biosynthesis of indigo dye by newly isolated

1241 naphthalene-degrading strain *Pseudomonas* sp. HOB1 and its application in dyeing

1242 cotton fabric. Appl. Biochem. Biotechnol. 160, 1616-1626.

1243 Paul, C.E., Eggerichs, D., Westphal, A.H., Tischler, D., van Berkel, W.J.H., 2021.

1244 Flavoprotein monooxygenases: Versatile biocatalysts. Biotechnol. Adv. 51, 107712.

1245 Pazmiño, D.E.T., Snajdrova, R., Rial, D.V., Mihovilovic, M.D., Fraaije, M.W., 2007.  
 1246 Altering the substrate specificity and enantioselectivity of phenylacetone  
 1247 monooxygenase by structure-inspired enzyme redesign. *Adv. Synth. Catal.* 349 (8-9),  
 1248 1361-1368.

1249 Peleg, Y., Vincentelli, R., Collins, B.M., Chen, K.-E., Livingstone, E.K., Weeratunga, S.,  
 1250 Leneva, N., Guo, Q., Remans, K., Perez, K., Bjerga, G.E.K., Larsen, O., Vanek, O.,  
 1251 Skorepa, O., Jacquemin, S., Poterszman, A., Kjaer, S., Christodoulou, E., Albeck, S.,  
 1252 Dym, O., Ainbinder, E., Unger, T., Schuetz, A., Matthes, S., Bader, M., de Marco, A.,  
 1253 Storici, P., Semrau, M.S., Stolt-Bergner, P., Aigner, C., Suppmann, S., Goldenzweig, A.,  
 1254 Fleishman, S.J., 2021. Community-wide experimental evaluation of the PROSS  
 1255 stability-design method. *J. Mol. Biol.* 433 (13), 166964.

1256 Peng, Z., Wang, W., Wei, H., Li, X., Yang, J., 2023. Improved protein structure  
 1257 prediction with trRosettaX2, AlphaFold2, and optimized MSAs in CASP15. *Proteins*  
 1258 1-8.

1259 Pereira, M.S., de Araújo, S.S., Nagem, R.A.P., Richard, J.P., Brandão, T.A.S., 2022. The  
 1260 role of remote flavin adenine dinucleotide pieces in the oxidative decarboxylation  
 1261 catalyzed by salicylate hydroxylase. *Bioorg. Chem.* 119, 105561.

1262 Phillips, I.R., Shephard, E.A., 2020a. Flavin-containing monooxygenase 3 (FMO3):  
 1263 genetic variants and their consequences for drug metabolism and disease. *Xenobiotica*  
 1264 50 (1), 19-33.

1265 Phillips, I.R., Shephard, E.A., 2020b. Flavin-containing monooxygenases: new  
 1266 structures from old proteins. *Nat. Struct. Mol. Biol.* 27 (1), 3-4.  
 1267 Polyzos, A.A., 2003. Directed evolution of a sulfoxidation biocatalyst. Dissertation,  
 1268 University of Florida.  
 1269 Poulsen, L.L., Ziegler, D.M., 1979. The liver microsomal FAD-containing  
 1270 monooxygenase. Spectral characterization and kinetic studies. *J. Biol. Chem.* 254 (14),  
 1271 6449-6455.  
 1272 Qi, Y., Zhang, J.Z.H., 2020. DenseCPD: Improving the accuracy of  
 1273 neural-network-based computational protein sequence design with DenseNet. *J. Chem*  
 1274 *Inf. Model.* 60 (3), 1245-1252.  
 1275 Qi, Y., Zhu, J., Zhang, K., Liu, T., Wang, Y., 2022. Recent development of directed  
 1276 evolution in protein engineering. *Syn. Biol.* 3 (6), 1081-1108.  
 1277 Qu, Y., Pi, W., Ma, F., Zhou, J., Zhang, X., 2010. Influence and optimization of growth  
 1278 substrates on indigo formation by a novel isolate *Acinetobacter* sp. PP-2. *Bioresour.*  
 1279 *Technol.* 101 (12), 4527-4532.  
 1280 Qu, Y., Shi, S., Zhou, H., Ma, Q., Li, X., Zhang, X., Zhou, J., 2012a. Characterization of  
 1281 a novel phenol hydroxylase in indoles biotransformation from a strain *Arthrobacter* sp.  
 1282 W1.  
 1283 Qu, Y., Zhang, X., Ma, Q., Ma, F., Zhang, Q., Li, X., Zhou, H., Zhou, J., 2012b. Indigo  
 1284 biosynthesis by *Comamonas* sp. MQ. *Biotechnol. Lett.* 34 (2), 353-357.

1285 Rackers, J.A., Wang, Z., Lu, C., Laury, M.L., Lagardere, L., Schnieders, M.J., Piquemal,  
 1286 J.P., Ren, P.Y., Ponder, J.W., 2018. Tinker 8: Software tools for molecular design. J.  
 1287 Chem. Theory Comput. 14 (10), 5273-5289.  
 1288 Ramana, P., Herman, W., Hiroux, C., Adams, E., Augustijns, P., Van Schepdael, A.,  
 1289 2017. Evaluation of immobilized hFMO<sub>3</sub> on magnetic nanoparticles by capillary zone  
 1290 electrophoresis. Bioanalysis 9 (3), 289-296.  
 1291 Ravi, S., Gunawan, R., 2021.  $\Delta$ FBA—Predicting metabolic flux alterations using  
 1292 genome-scale metabolic models and differential transcriptomic data. PLoS Comput.  
 1293 Biol. 17 (11), e1009589.  
 1294 Riebel, A., de Gonzalo, G., Fraaije, M.W., 2013. Expanding the biocatalytic toolbox of  
 1295 flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. J. Mol. Catal. B: Enzym.  
 1296 88, 20-25.  
 1297 Rioz-Martinez, A., Kopacz, M., de Gonzalo, G., Torres Pazmino, D.E., Gotor, V.,  
 1298 Fraaije, M.W., 2011. Exploring the biocatalytic scope of a bacterial flavin-containing  
 1299 monooxygenase. Org. Biomol. Chem. 9 (5), 1337-1341.  
 1300 Sabourin, P.J., Hodgson, E., 1984. Characterization of the purified microsomal  
 1301 FAD-containing monooxygenase from mouse and pig liver. Chem. Biol. Interact. 51 (2),  
 1302 125-139.  
 1303 Sadeghi, S.J., Fantuzzi, A., Gilardi, G., 2011. Breakthrough in P450 bioelectrochemistry  
 1304 and future perspectives. Biochim. Biophys. Acta 1814 (1), 237-248.

1305 Sadeghi, S.J., Meirinhos, R., Catucci, G., Dodhia, V.R., Nardo, G.D., Gilardi, G., 2010.  
 1306 Direct electrochemistry of drug metabolizing human flavin-containing monooxygenase:  
 1307 Electrochemical turnover of benzydamine and tamoxifen. *J. Am. Chem. Soc.* 132 (2),  
 1308 458-459.  
 1309 Saito, H., Shirai, T., Kudo, H., Mitsubayashi, K., 2008. Electrochemical sensor with  
 1310 flavin-containing monooxygenase for triethylamine solution. *Anal. Bioanal. Chem.* 391  
 1311 (4), 1263-1268.  
 1312 Schmidt, A.C., Leroux, J.C., 2020. Treatments of trimethylaminuria: where we are and  
 1313 where we might be heading. *Drug Discov. Today* 25 (9), 1710-1717.  
 1314 Schnepel, C., Doderio, V.I., Sewald, N., 2021. Novel arylindigoids by late-stage  
 1315 derivatization of biocatalytically synthesized dibromoindigo. *Chem.* 27 (17),  
 1316 5404-5411.  
 1317 Senthong, V., Wang, Z., Fan, Y., Wu, Y., Hazen, S.L., Tang, W.H., 2016. Trimethylamine  
 1318 N-oxide and mortality risk in patients with peripheral artery disease. *J. Am. Heart.*  
 1319 *Assoc.* 5 (10), e004237.  
 1320 Shao, Z., Zhao, H., Giver, L., Arnold, F.H., 1998. Random-priming in vitro  
 1321 recombination: An effective tool for directed evolution. *Nucleic Acids Res.* 26 (2),  
 1322 681-683.  
 1323 Shehin-Johnson, S.E., Williams, D.E., Larsen-Su, S., Stresser, D.M., Hines, R.N., 1995.  
 1324 Tissue-specific expression of flavin-containing monooxygenase (FMO) forms 1 and 2 in

1325 the rabbit. J. Pharmacol. Exp. Ther. 272 (3), 1293-1299.

1326 Shimada, A., Fujii, N., Saito, T., 2004. Chapter 26 - Tryptophanase Activity on  
 1327 D-Tryptophan, in: Pályi, G., Zucchi, C., Caglioti, L. (Eds.), Progress in Biological  
 1328 Chirality. Elsevier Science Ltd, Oxford, pp. 321-327.

1329 Siddiqui, J.K., Baskin, E., Liu, M., Cantemir-Stone, C.Z., Zhang, B., Bonneville, R.,  
 1330 McElroy, J.P., Coombes, K.R., Mathé, E.A., 2018. IntLIM: integration using linear  
 1331 models of metabolomics and gene expression data. BMC Bioinform. 19 (1), 81.

1332 Sieber, V., Martinez, C.A., Arnold, F.H., 2001. Libraries of hybrid proteins from  
 1333 distantly related sequences. Nat. Biotechnol. 19 (5), 456-460.

1334 Singh, A., Singh Chauhan, N., Thulasiram, H.V., Taneja, V., Sharma, R., 2010.  
 1335 Identification of two flavin monooxygenases from an effluent treatment plant sludge  
 1336 metagenomic library. Bioresour. Technol. 101 (21), 8481-8484.

1337 Søndergaard, C.R., Olsson, M.H.M., Rostkowski, M., Jensen, J.H., 2011. Improved  
 1338 treatment of ligands and coupling effects in empirical calculation and rationalization of  
 1339 pka values. J. Chem. Theory Comput. 7 (7), 2284-2295.

1340 Stasiak, N., Kukuła-Koch, W., Głowniak, K., 2014. Modern industrial and  
 1341 pharmacological applications of indigo dye and its derivatives--a review. Acta Pol.  
 1342 Pharm. 71 (2), 215-221.

1343 Stephens, G.M., Sidebotham, J.M., Mann, N.H., Dalton, H., 1989. Cloning and  
 1344 expression in *Escherichia coli* of the toluene dioxygenase gene from *Pseudomonas*

1345 *putida* NCIB11767. FEMS Microbiol. Lett. 57 (3), 295-300.

1346 Suh, J.-K., Poulsen, L.L., Ziegler, D.M., Robertus, J.D., 1996. Molecular cloning and  
 1347 kinetic characterization of a flavin-containing monooxygenase from *Saccharomyces*  
 1348 *cerevisiae*. Arch. Biochem. Biophys. 336 (2), 268-274.

1349 Suh, J.K., Poulsen, L.L., Ziegler, D.M., Robertus, J.D., 1999. Yeast flavin-containing  
 1350 monooxygenase generates oxidizing equivalents that control protein folding in the  
 1351 endoplasmic reticulum. PNAS 96 (6), 2687-2691.

1352 Sun, B.Y., Sui, H.L., Liu, Z.W., Tao, X.Y., Gao, B., Zhao, M., Ma, Y.S., Zhao, J., Liu,  
 1353 M., Wang, F.Q., Wei, D.Z., 2022. Structure-guided engineering of a flavin-containing  
 1354 monooxygenase for the efficient production of indirubin. Bioresour. Bioprocess. 9 (1),  
 1355 70.

1356 Tan, C., Zhang, X., Zhu, Z., Xu, M., Yang, T., Osire, T., Yang, S., Rao, Z., 2019.  
 1357 Asp305Gly mutation improved the activity and stability of the styrene monooxygenase  
 1358 for efficient epoxide production in *Pseudomonas putida* KT2440. Microb. Cell Fact. 18  
 1359 (1), 12.

1360 Tang, W.H.W., Bäckhed, F., Landmesser, U., Hazen, S.L., 2019. Intestinal microbiota in  
 1361 cardiovascular health and disease: JACC state-of-the-art review. J. Am. Coll. Cardiol.  
 1362 73 (16), 2089-2105.

1363 Taylor, K.L., Ziegler, D., 1987. Studies on substrate specificity of the hog liver  
 1364 flavin-containing monooxygenase: anionic organic sulfur compounds. Biochem.

1365 Pharmacol. 36 (1), 141-146.

1366 Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale  
1367 metabolic reconstruction. Nat. Protoc. 5 (1), 93-121.

1368 Treacy, E.P., Akerman, B.R., Chow, L.M., Youil, R., Bibeau, C., Lin, J., Bruce, A.G.,  
1369 Knight, M., Danks, D.M., Cashman, J.R., Forrest, S.M., 1998. Mutations of the  
1370 flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in  
1371 detoxication. Hum. Mol. Genet. 7 (5), 839-845.

1372 Tsuji, T., Onimaru, M., Yanagawa, H., 2001. Random multi-recombinant PCR for the  
1373 construction of combinatorial protein libraries. Nucleic Acids Res. 29 (20), e97.

1374 Tynes, R.E., Sabourin, P.J., Hodgson, E., 1985. Identification of distinct hepatic and  
1375 pulmonary forms of microsomal flavin-containing monooxygenase in the mouse and  
1376 rabbit. Biochem. Biophys. Res. Commun. 126 (3), 1069-1075.

1377 Tynes, R.E., Sabourin, P.J., Hodgson, E., Philpot, R.M., 1986. Formation of hydrogen  
1378 peroxide and N-hydroxylated amines catalyzed by pulmonary flavin-containing  
1379 monooxygenases in the presence of primary alkylamines. Arch. Biochem. Biophys. 251  
1380 (2), 654-664.

1381 Uno, Y., Shimizu, M., Yamazaki, H., 2013. Molecular and functional characterization of  
1382 flavin-containing monooxygenases in cynomolgus macaque. Biochem. Pharmacol. 85  
1383 (12), 1837-1847.

1384 Valentino, H., Campbell, A.C., Schuermann, J.P., Sultana, N., Nam, H.G., LeBlanc, S.,



1385 Tanner, J.J., Sobrado, P., 2020. Structure and function of a flavin-dependent  
 1386 S-monooxygenase from garlic (*Allium sativum*). *J. Biol. Chem.* 295 (32), 11042-11055.  
 1387 van Berkel, W.J.H., Kamerbeek, N.M., Fraaije, M.W., 2006. Flavoprotein  
 1388 monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124 (4),  
 1389 670-689.  
 1390 Wang, J., Lisanza, S., Juergens, D., Tischler, D., Watson, J.L., Castro, K.M., Ragotte, R.,  
 1391 Saragovi, A., Milles, L.F., Baek, M., Anishchenko, I., Yang, W., Hicks, D.R., Exposit,  
 1392 M., Schlichthaerle, T., Chun, J.H., Dauparas, J., Bennett, N., Wicky, B.I.M., Muenks, A.,  
 1393 DiMaio, F., Correia, B., Ovchinnikov, S., Baker, D., 2022. Scaffolding protein  
 1394 functional sites using deep learning. *Science* 377 (6604), 387-394.  
 1395 Wang, X., Xu, K., Tan, Y., Liu, S., Zhou, J., 2023. Possibilities of Using De Novo  
 1396 Design for Generating Diverse Functional Food Enzymes. *Int. J. Mol. Sci.* 24 (4), 3827.  
 1397 Wani, K.A., Goswamy, D., Taubert, S., Ratnappan, R., Ghazi, A., Irazoqui, J.E., 2021.  
 1398 NHR-49/PPAR-alpha and HLH-30/TFEB cooperate for *C. elegans* host defense via a  
 1399 flavin-containing monooxygenase. *Elife* 10, e62775.  
 1400 Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern,  
 1401 W., Borst, A.J., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Hanikel, N., Pellock, S.J.,  
 1402 Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S.V., Lauko, A.,  
 1403 De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T.S., DiMaio, F.,  
 1404 Baek, M., Baker, D., 2023. *De novo* design of protein structure and function with

1405 RFdiffusion. Nature 620 (7976), 1089-1100.

1406 Weissenow, K., Heinzinger, M., Rost, B., 2022. Protein language-model embeddings for  
 1407 fast, accurate, and alignment-free protein structure prediction. Structure 30 (8),  
 1408 1169-1177.e4.

1409 Wicky, B.I.M., Milles, L.F., Courbet, A., Ragotte, R.J., Dauparas, J., Kinfu, E., Tipps, S.,  
 1410 Kibler, R.D., Baek, M., DiMaio, F., Li, X., Carter, L., Kang, A., Nguyen, H., Bera, A.K.,  
 1411 Baker, D., 2022. Hallucinating symmetric protein assemblies. Science 378 (6615),  
 1412 56-61.

1413 Wijma, H.J., Furst, M.J.L.J., Janssen, D.B., 2018. A Computational Library Design  
 1414 Protocol for Rapid Improvement of Protein Stability: FRESCO, in: Bornscheuer, U.T.,  
 1415 Hohne, M. (Eds.), Protein Engineering: Methods and Protocols. Humana New York,  
 1416 New York, pp. 69-85.

1417 Wojaczyńska, E., Wojaczyński, J., 2020. Modern stereoselective synthesis of chiral  
 1418 sulfinyl compounds. Chem. Rev. 120 (10), 4578-4611.

1419 Woo, H.-j., Sanseverino, J., Cox, C.D., Robinson, K.G., Sayler, G.S., 2000. The  
 1420 measurement of toluene dioxygenase activity in biofilm culture of *Pseudomonas putida*  
 1421 F1. J. Microbiol. Methods. 40 (2), 181-191.

1422 Wood, C.W., Heal, J.W., Thomson, A.R., Bartlett, G.J., Ibarra, A.A., Brady, R.L.,  
 1423 Sessions, R.B., Woolfson, D.N., 2017. ISAMBARD: an open-source computational  
 1424 environment for biomolecular analysis, modelling and design. Bioinformatics 33 (19),

1425 3043-3050.

1426 Woolfson, D.N., 2021. A brief history of *de novo* protein design: Minimal, rational, and  
 1427 computational. J. Mol. Biol. 433 (20), 167160.

1428 Wyatt, M.K., Overby, L.H., Lawton, M.P., Philpot, R.M., 1998. Identification of amino  
 1429 acid residues associated with modulation of flavin-containing monooxygenase (FMO)  
 1430 activity by imipramine: Structure/function studies with FMO1 from pig and rabbit.  
 1431 Biochemistry 37 (17), 5930-5938.

1432 Xiong, P., Chen, Q., Liu, H., 2017. Computational Protein Design Under a Given  
 1433 Backbone Structure with the ABACUS Statistical Energy Function, in: Samish, I. (Ed.)  
 1434 Computational Protein Design. Springer New York, New York, pp. 217-226.

1435 Xiong, W., Liu, B., Shen, Y., Jing, K., Savage, T.R., 2021. Protein engineering design  
 1436 from directed evolution to de novo synthesis. Biochem. Eng. J. 174, 108096.

1437 Xu, J., Yang, Y., 2021. Gut microbiome and its meta-omics perspectives: profound  
 1438 implications for cardiovascular diseases. Gut Microbes 13 (1), 1936379.

1439 Yang, J.E., Park, S.J., Kim, W.J., Kim, H.J., Kim, B.J., Lee, H., Shin, J., Lee, S.Y., 2018.  
 1440 One-step fermentative production of aromatic polyesters from glucose by metabolically  
 1441 engineered Escherichia coli strains. Nat. Commun. 9 (1), 79.

1442 Yang, L., Li, X., Huang, W., Rao, X., Lai, Y., 2022. Pharmacological properties of  
 1443 indirubin and its derivatives. Biomed. Pharmacother. 151, 113112.

1444 Yang, S., Li, X., Yang, F., Zhao, R., Pan, X., Liang, J., Tian, L., Li, X., Liu, L., Xing, Y.,

1445 Wu, M., 2019. Gut microbiota-dependent marker tmao in promoting cardiovascular  
 1446 disease: Inflammation mechanism, clinical prognostic, and potential as a therapeutic  
 1447 target. *Front. Pharmacol.* 10, 1360.

1448 YANG Ya-kun, S.L., LI Yan, 2017. Role of FMOs in drug metabolism and development.  
 1449 *Acta Pharm. Sin.* 52 (10), 1485-1495.

1450 Yang, Z., Zeng, X., Zhao, Y., Chen, R., 2023. AlphaFold2 and its applications in the  
 1451 fields of biology and medicine. *Signal Transduct. Target. Ther.* 8 (1), 115.

1452 Yeh, A.H.-W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S.J., Evans, D., Ma, P., Lee,  
 1453 G.R., Zhang, J.Z., Anishchenko, I., Coventry, B., Cao, L., Dauparas, J., Halabiya, S.,  
 1454 DeWitt, M., Carter, L., Houk, K.N., Baker, D., 2023. *De novo* design of luciferases  
 1455 using deep learning. *Nature* 614 (7949), 774-780.

1456 Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J.,  
 1457 Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B.,  
 1458 Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Dien, S.V., 2011. Metabolic engineering  
 1459 of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* 7 (7),  
 1460 445-452.

1461 Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., Shlomi, T., 2010. Integrating  
 1462 quantitative proteomics and metabolomics with a genome-scale metabolic network  
 1463 model. *Bioinformatics* 26 (12), i255-i260.

1464 Zane, N.R., Chen, Y., Wang, M.Z., Thakker, D.R., 2018. Cytochrome P450 and

1465 flavin-containing monooxygenase families: age-dependent differences in expression and  
 1466 functional activity. *Pediatr. Res.* 83 (2), 527-535.

1467 Zhang, J., Cashman, J.R., 2006. Quantitative analysis of FMO gene mRNA levels in  
 1468 human tissues. *Drug Metab. Dispos.* 34 (1), 19-26.

1469 Zhang, Y., Wu, Y.-Q., Xu, N., Zhao, Q., Yu, H.-L., Xu, J.-H., 2019. Engineering of  
 1470 cyclohexanone monooxygenase for the enantioselective synthesis of (S)-omeprazole.  
 1471 *ACS Sustain. Chem. Eng.* 7 (7), 7218-7226.

1472 Zhao, H., Giver, L., Shao, Z., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution  
 1473 by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* 16 (3),  
 1474 258-261.

1475 Zheng, W., Li, Y., Zhang, C., Zhou, X., Pearce, R., Bell, E.W., Huang, X., Zhang, Y.,  
 1476 2021. Protein structure prediction using deep learning distance and hydrogen-bonding  
 1477 restraints in CASP14. *Proteins* 89 (12), 1734-1751.

1478 Zhu, W., Buffa, J.A., Wang, Z., Warriar, M., Schugar, R., Shih, D.M., Gupta, N.,  
 1479 Gregory, J.C., Org, E., Fu, X., Li, L., DiDonato, J.A., Lusi, A.J., Brown, J.M., Hazen,  
 1480 S.L., 2018. Flavin monooxygenase 3, the host hepatic enzyme in the metaorganismal  
 1481 trimethylamine N-oxide-generating pathway, modulates platelet responsiveness and  
 1482 thrombosis risk. *J. Thromb. Haemost.* 16 (9), 1857-1872.

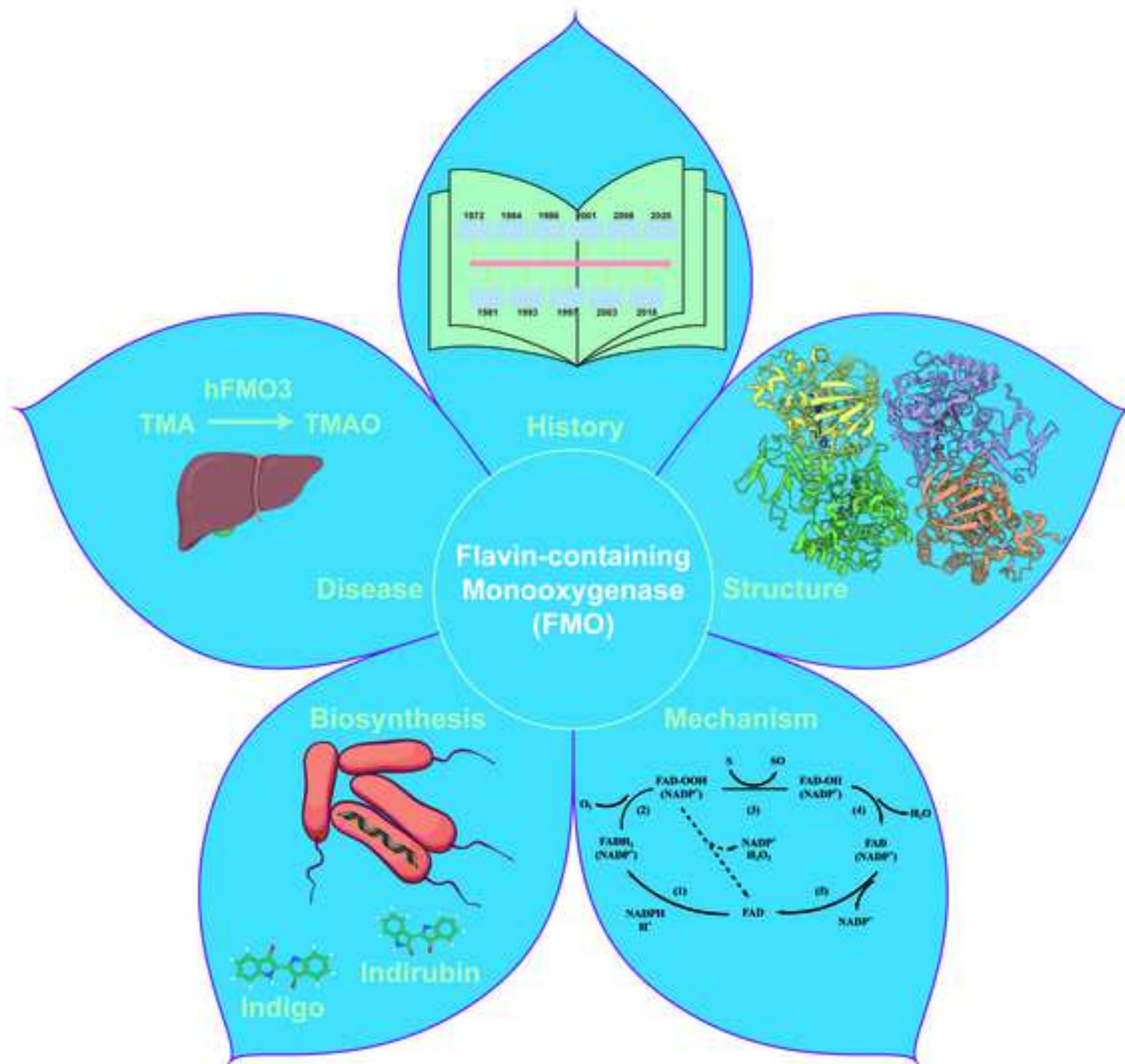
1483 Ziegler, D.M., 1993. Recent studies on the structure and function of multisubstrate  
 1484 flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* 33, 179-199.

1485 Ziegler, D.M., 2002. An overview of the mechanism, substrate specificities, and  
1486 structure of FMOs. *Drug Metab. Rev.* 34 (3), 503-511.

1487 Ziegler, D.M., Mitchell, C.H., 1972. Microsomal oxidase IV: Properties of a  
1488 mixed-function amine oxidase isolated from pig liver microsomes. *Arch. Biochem.*  
1489 *Biophys.* 150 (1), 116-125.

1490 [dataset] Fan, C., Xie, Z., 2023. AutoDock Docking Results between mFMO and Indole.  
1491 Mendeley Data, v1. <https://doi.org/10.17632/zrmhyr5y32.1>.

---



---

## Captions

**Figure 1.** The timeline of research history about FMO and its application to indigo biosynthesis (A) Crucial time points in the research history of FMO (B) Crucial time points in indigo biosynthesis with FMO.

**Figure 2.** Sequence alignments of FMOs across various species. (A) The consensus sequences of *Homo sapiens* FMO3 (hFMO), *ancient mammalian* FMO (ancFMO3-6), *Methylophaga* sp. strain SK1 FMO (mFMO), *Schizosaccharomyces pombe* FMO (spFMO), *Corynebacterium glutamicum* FMO (cFMO), and *Nitrospira lacisaponensis* FMO (NiFMO) were identified. The putative FAD, FMO-identifying motif, and NADPH pyrophosphate binding domain are colored blue. Some active sites are colored with ClustalX. The alignment was performed with ClustalW, Jalview, and ESPrpt (B) Protein sequence identity (%) between different FMOs.

**Figure 3.** The structure of AncFMO3-6. (A) The cartoon of homo-dimer AncFMO3-6 (PDB ID: 6se3.1) is exhibited from the front perspective. Pale cyan: Rossmann fold for FAD, 2-154 AA (Amino acids), 331-442 AA. Light blue: Rossmann fold for NADP(H), 155-213 AA, 296-330 AA. Orange: the ridged triangular fold, 214-295 AA. Green: 443-507 AA. Pale green: 510-532 AA. Pale yellow: FAD. Light pink: NADP(H). Red: oxygen; (B) The ridged triangular fold and the catalytic cavity are exhibited from the front perspective. The black arrow points to the catalytic cavity; (C) The tunnel is exhibited from the bottom perspective (left) and upper perspective (right). The black dotted line means the connected surface of homo-dimer AncFMO3-6. The black arrows point to the tunnel and the catalytic cavity; (D) The



---

mesh of amino acids which form hydrogen bonds with NADP(H) and FAD; (E) The amino acids (S13, E32, R33, L40, W41, R51, T60, N61, S62, V110, and T378) form hydrogen bonds with FAD; (F) The amino acids (F59, N61, L192, N194, S195, S216, E281, and Q373) form hydrogen bonds with NADP(H).

**Figure 4.** The structure and catalytic mechanism of mFMO. (A) The cartoon of homo-dimer mFMO (PDB ID: 2vq7) is exhibited from the front perspective. Pale cyan: Rossmann fold for FAD, 1-169 AA and 281-461 AA. Light blue: Rossmann fold for NADP(H), 170-280 AA. Pale yellow: FAD. Light pink: NADP(H). Pale green: three loops, 44-80 AA, 166-186 AA, and 276-306 AA; (B) The catalytic cavity is exhibited from the front perspective. The black arrow points to the catalytic cavity; (C) The three loops are exhibited linking the binding domains of NADP(H) and FAD; (D) Y212 protects the catalytic cavity from the solvents (E) The catalytic mechanism of FMOs. The sulfur atom is regarded as the substrate.

**Figure 5.** The docking result of indole and mFMO via AutoDock. (A) The cartoon of mFMO (PDB ID: 2vq7) with indole (Compound CID: 798) is exhibited from the front perspective. Pale cyan: Rossmann fold for FAD, 1-169 AA and 281-461 AA. Light blue: Rossmann fold for NADP(H), 170-280 AA. Pale yellow: FAD. Light pink: NADP(H). Pale green: three loops, 44-80 AA, 166-186 AA, and 276-306 AA. Green sticks: indole. Yellow dotted line: hydrogen bonds; (B) The surface model of mFMO with indole. Light blue dots: Tyr212 (C) The mechanism of indigo biosynthesis.

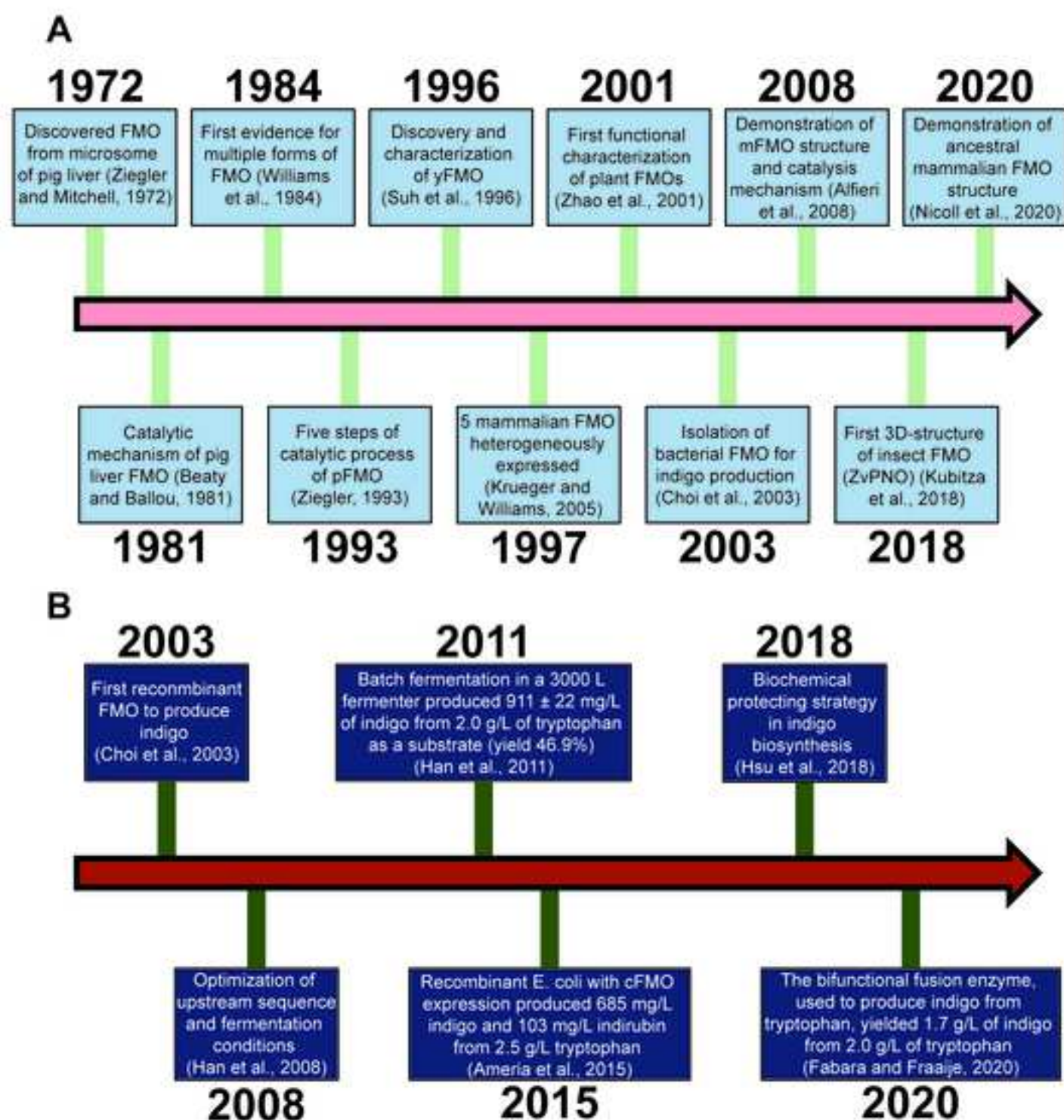
**Figure 6.** The sketch map of directed evolution to improve indigo biosynthesis via FMOs.

---

**Figure 7.** Demonstrations of some *de novo* design methods. (A) Active center design for enzymes and protein binders; (B) Sketch maps of some AI-based tools for protein scaffold design, including hallucination, inpainting, and MPNN; (C) Validation techniques for *de novo* design proteins.

**Figure 8.** Various methods of FMOs immobilization. (A) FMOs were adsorbed by support materials; (B) FMOs formed covalent bonds with support materials; (C) FMOs were cross-linked by a cross-linking agent; (D) FMOs were entrapped within a polymeric network; (E) FMOs were enclosed in a spherical semipermeable membrane; (F) FMOs were immobilized on graphene oxide (GO). Didodecyldimethylammonium bromide (DDAB) played as an interface between GO and FMOs. FMOs-DDAB-GO was on glassy carbon (GC) electrodes.

**Figure 9.** Metabolic and fermentation engineering to improve biosynthesis of FMOs. (A) A suitable chassis should be chosen for further improvements; (B) The establishment and application of genome-scale metabolic network models (GEMs); (C) Optimization in fermentation engineering to increase the production of indigo biosynthesis.

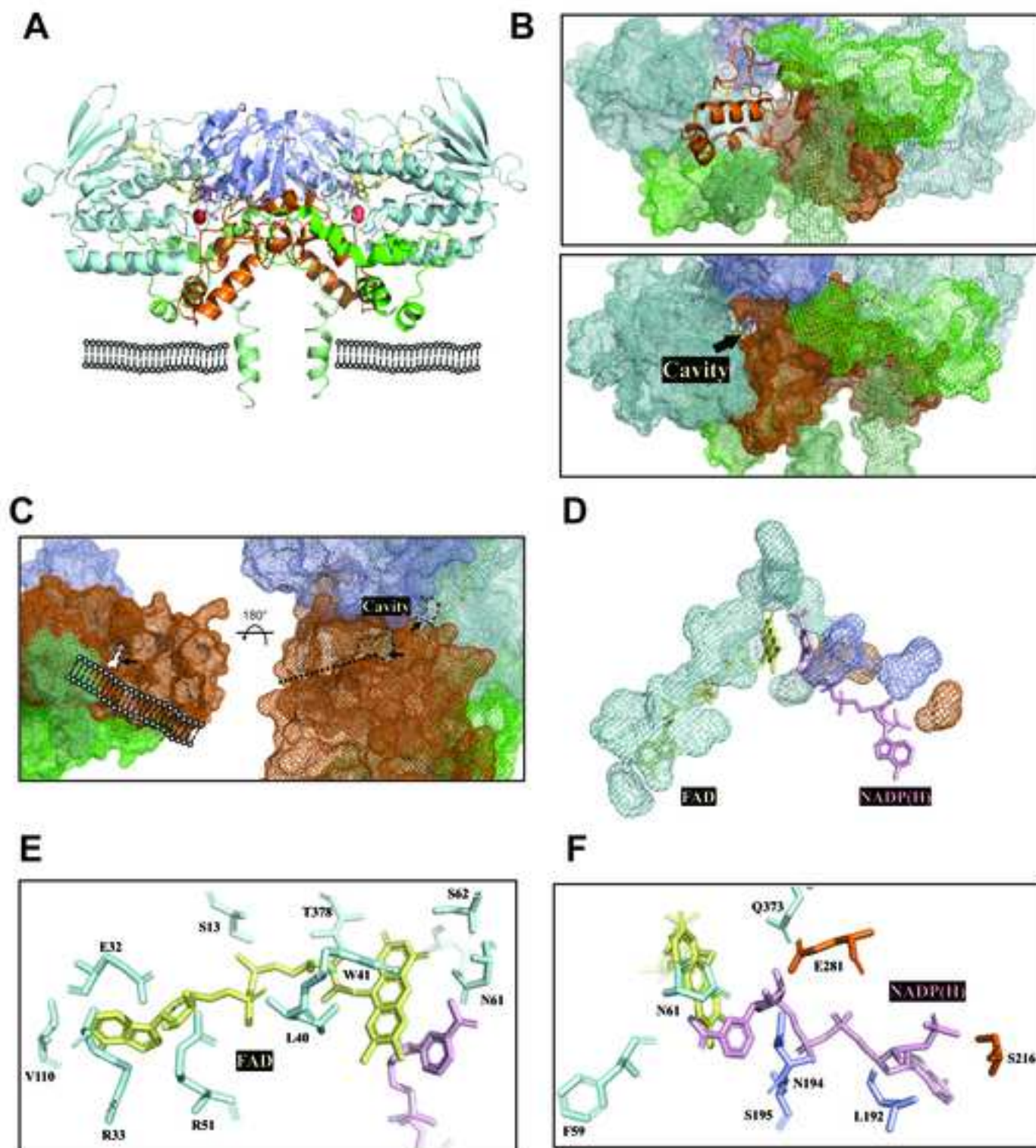


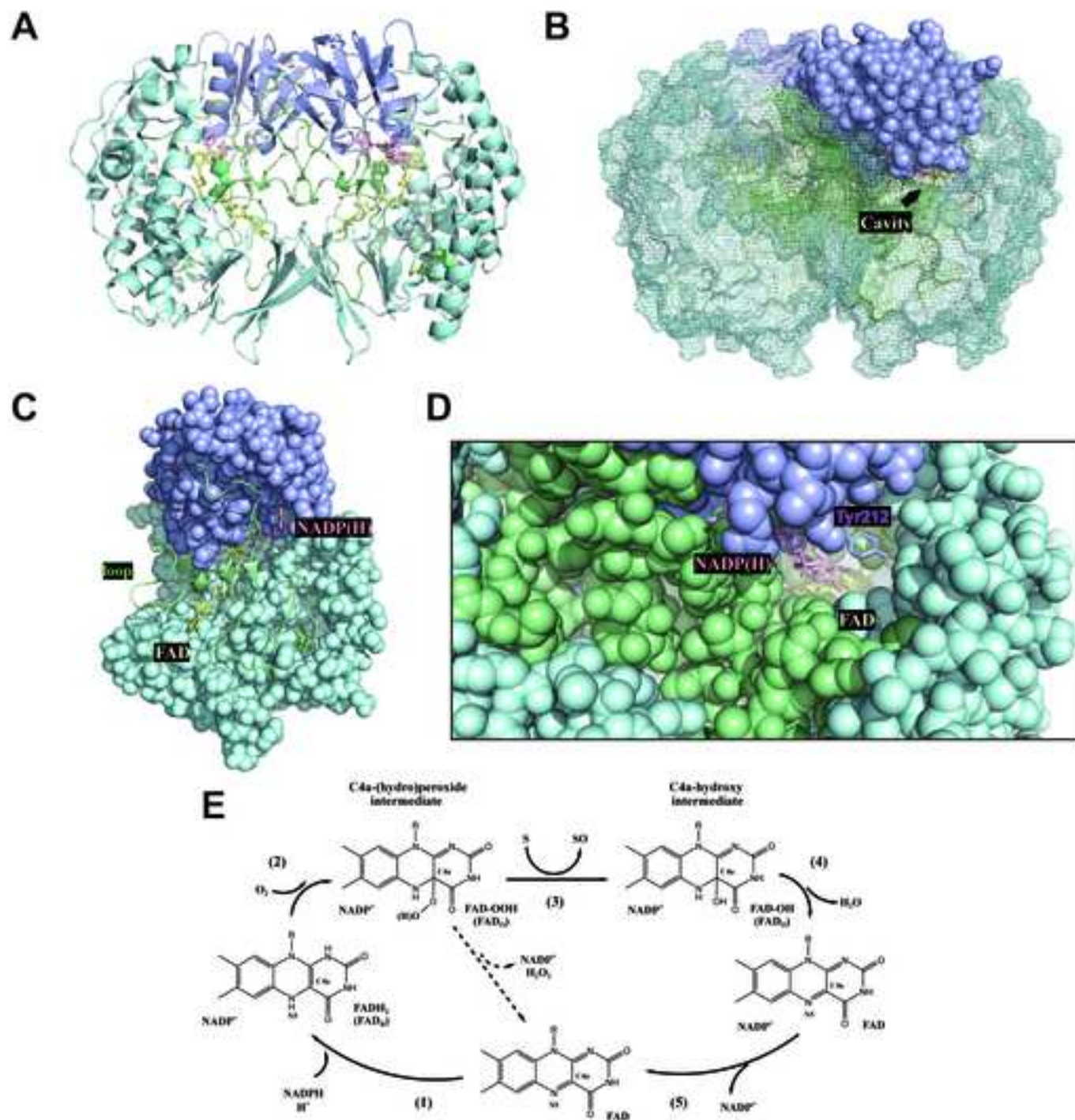
**B**

Heatmap showing the correlation of FMOs with the six FMOs. The color scale ranges from 0 (blue) to 100 (red).

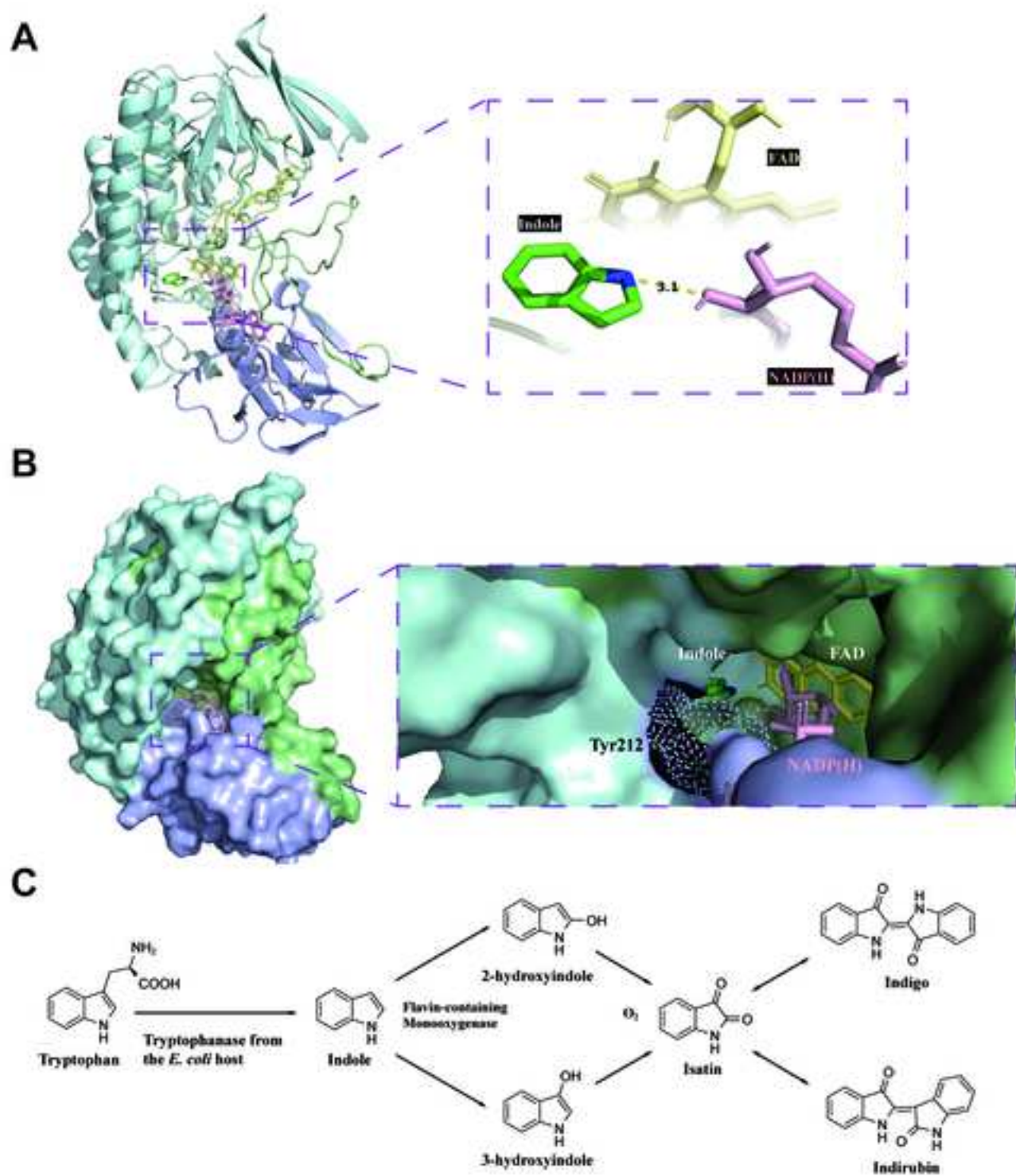
	spFMO	cFMO	hFMO3	ancFMO	mFMO	NiFMO
spFMO	100.00	9.50	18.00	18.90	20.90	10.00
cFMO	9.50	100.00	9.20	9.10	10.20	10.50
hFMO3	18.00	9.20	100.00	82.70	20.90	19.80
ancFMO	18.90	9.10	82.70	100.00	20.70	20.60
mFMO	20.90	10.20	20.90	20.70	100.00	74.10
NiFMO	10.00	10.50	19.80	20.60	74.10	100.00

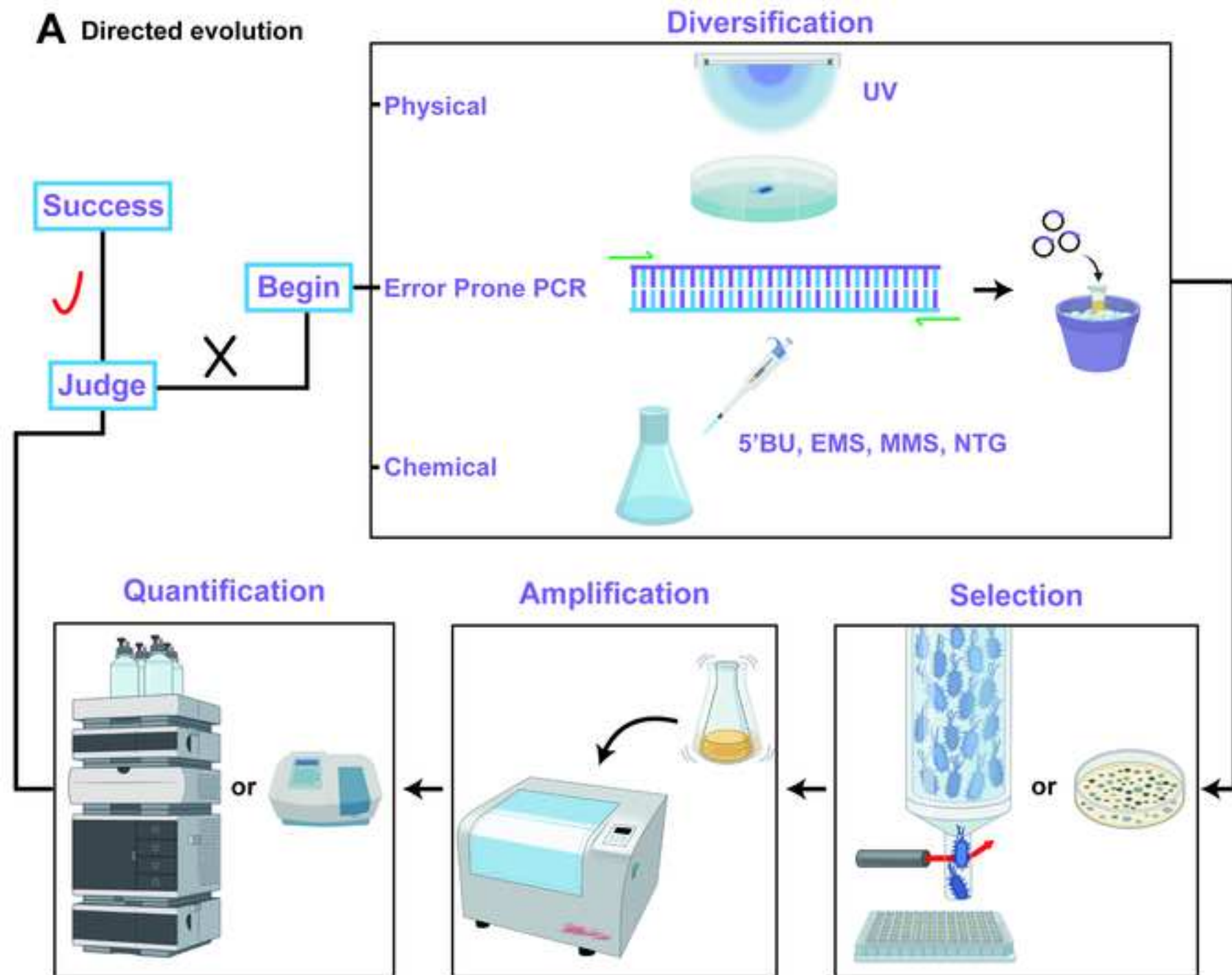




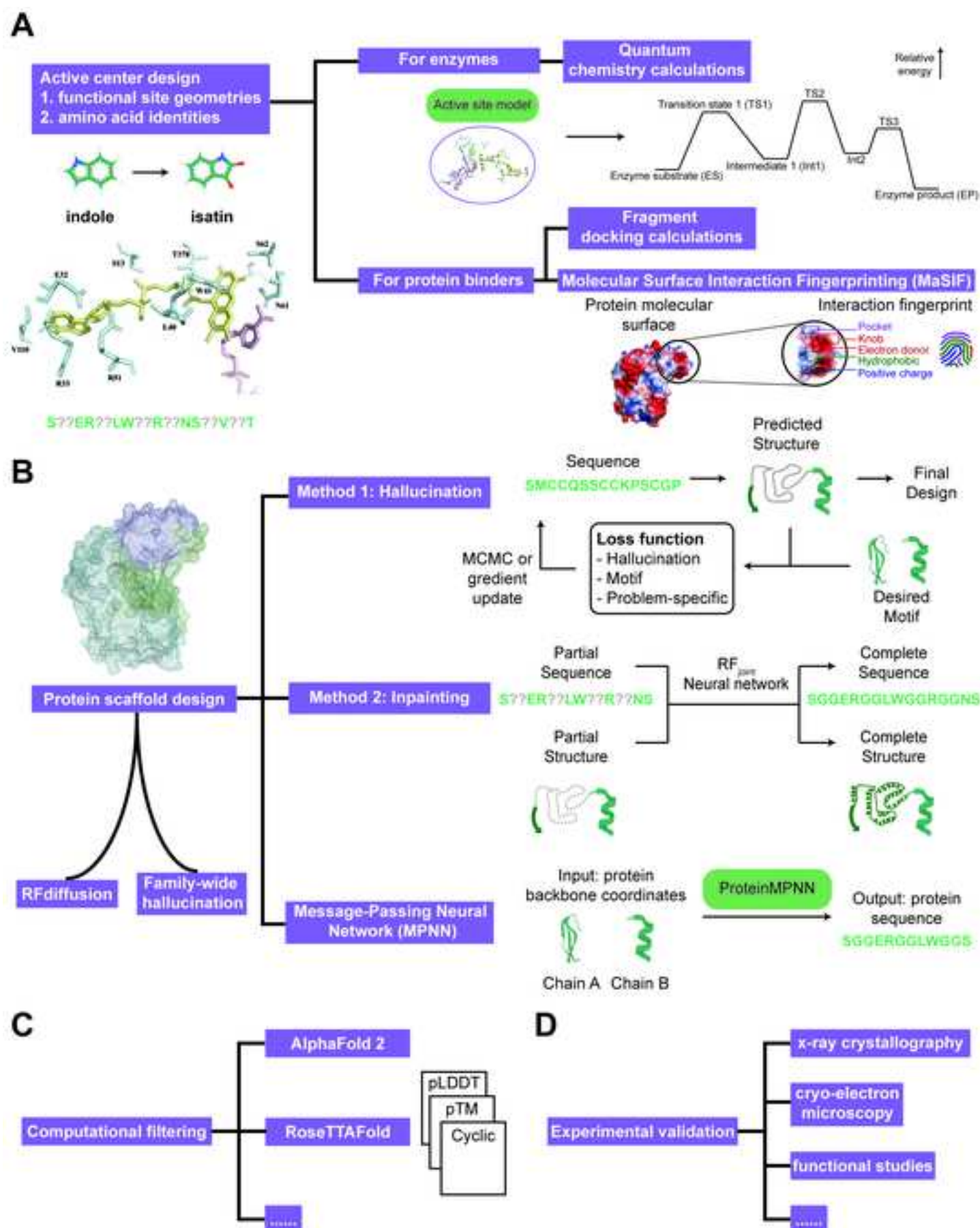


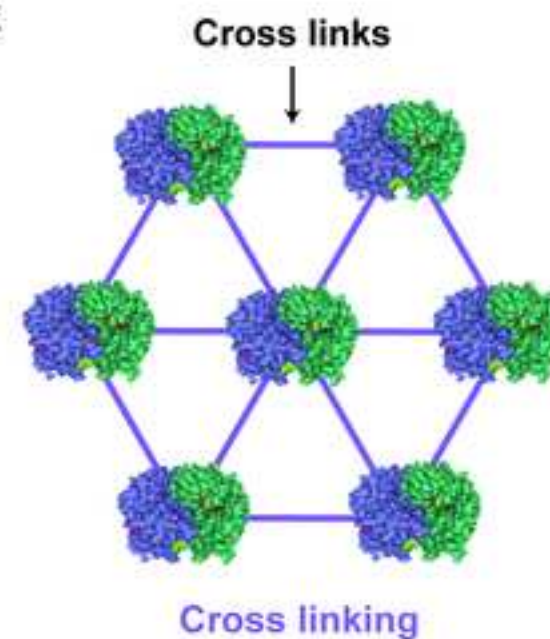
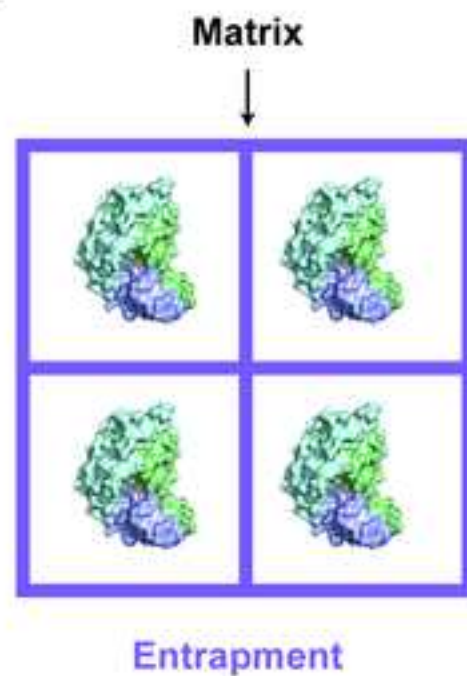
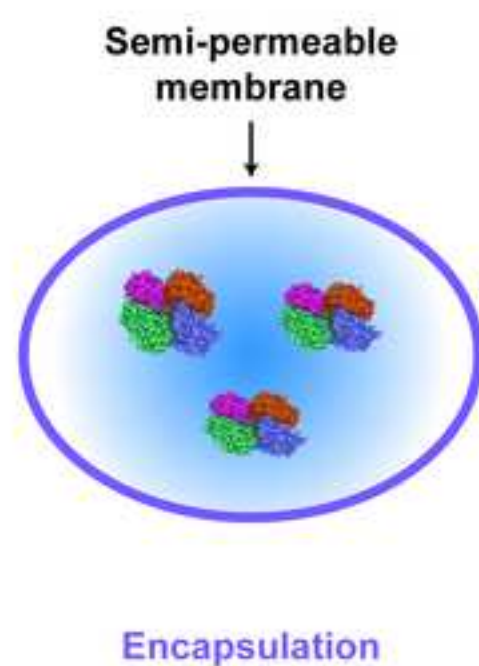


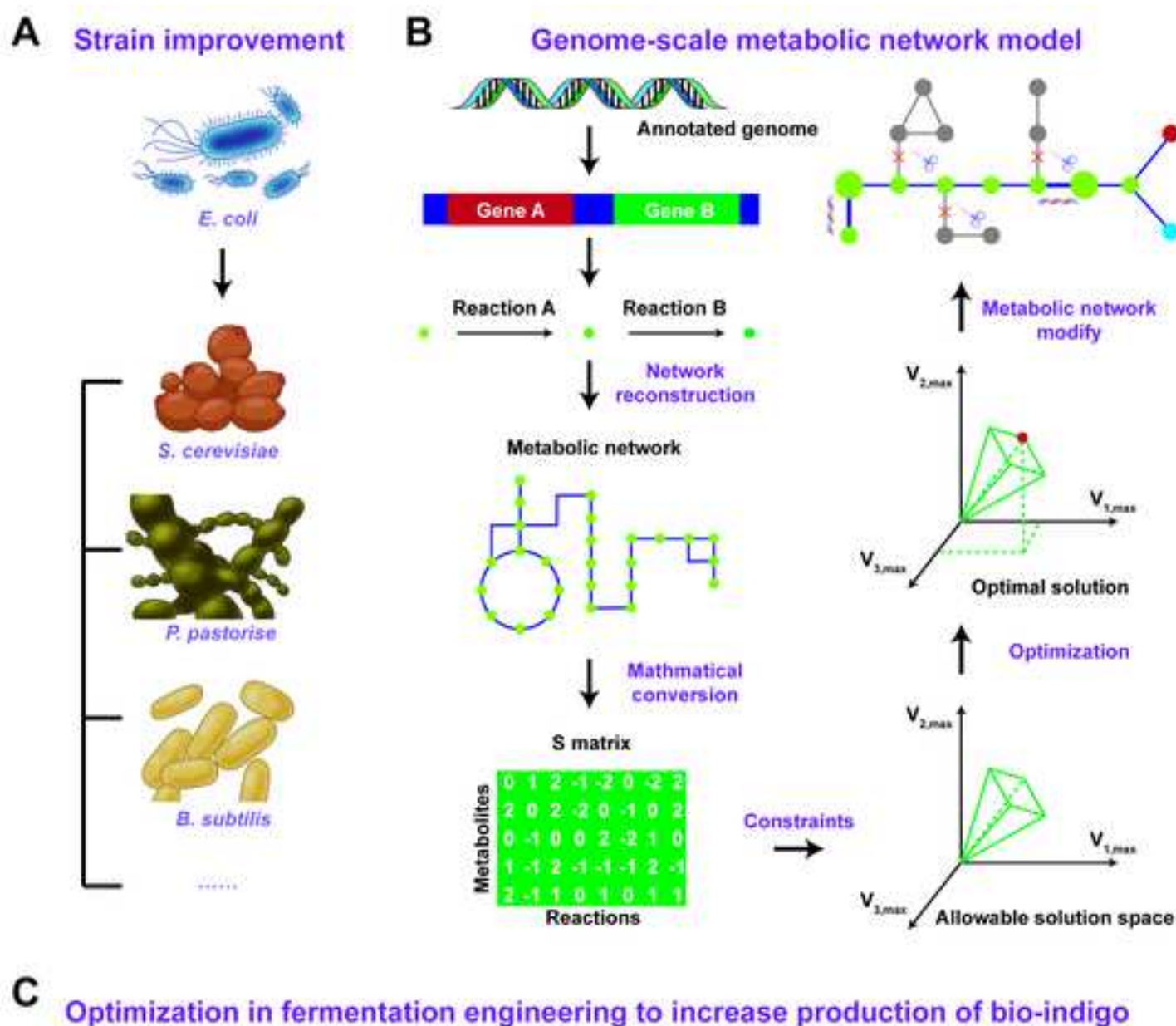








**A** Enzyme Immobilization**B****C****D****E****F**

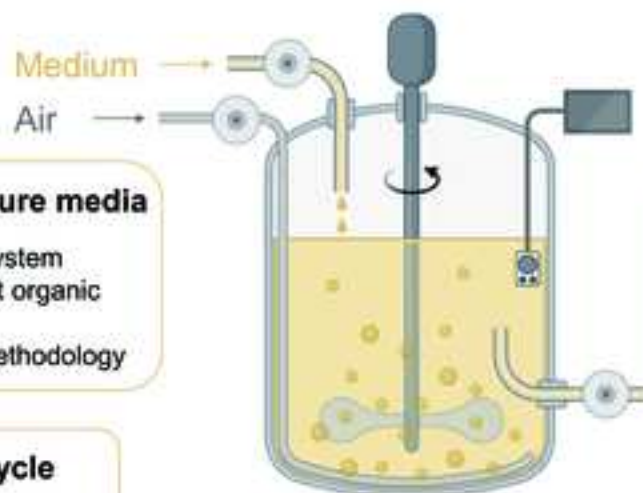


### Optimization of culture media

- two-phase culture system
- discovery of the best organic media for FMOs
- response surface methodology


### Substrate recycle

NADPH regeneration system



### Constant value control

- temperature
- pH
- agitation speed
- nutrient control
- etc.



[Click here to access/download](#)

**Data Statement**  
Data statement.docx

