2024

Improvement of Isoeugenol Dioxygenases for the Sustainable Production of Lignin-derived Vanillin

Mario De Simone

Dissertation presented to obtain the

Ph.D degree in Molecular Biosciences

Speciality in Biochemistry

Oeiras, November, 2024

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA ANTÓNIO XAVIER | UNIVERSIDADE NOVA DE LISBOA

Improvement of Isoeugenol Dioxygenases for the Sustainable Production of Lignin-derived Vanillin

Mario De Simone

Dissertation presented to obtain the Ph.D. degree in Molecular Biosciences

Instituto de Tecnologia Química e Biológica António Xavier
Universidade Nova de Lisboa

Supervisor: Prof. Lígia Oliveira Martins

Oeiras, November 2024

First Edition, November 2024
Microbial and Enzyme Technology Laboratory
Instituto de Tecnologia Química e Biológica António Xavier
Universidade Nova de Lisboa
Av. da República, Oeiras

ACKNOWLEDGMENTS

Emperor Augustus would say, *"If you enjoyed the show, applaud and let your joy be known."* Farrok Bulsara from Zanzibar would sing, *"The Show must Go On"*. I agree with the latter; this is another chapter in my life that began at Carvoeiro beach when, fresh out of university and enjoying a well-deserved vacation, Professor Lígia O. Martins invited me to Oeiras for an interview, marking the beginning of my time at ITQB-NOVA.

So, with that, let’s start the closing credits.
First and foremost, my deepest gratitude goes to my supervisor, Professor Lígia O. Martins, who accepted me into her laboratory and made this work possible. It was just as I expected: I knew it would be challenging, but I also knew it would be worth it. She has been an inexhaustible source of knowledge and taught me what it truly means to work in science.

I am grateful to everyone I met through the SMARTBOX project, especially Andrea Mattevi, Marco Fraaije, Ivana Maric, Laura Alvigini, and Lur Alonso, whose contributions were crucial to the outcome of this thesis. I extend my thanks, once again, to Andrea Mattevi and João Vicente for accepting my invitation to join my PhD thesis committee and for providing valuable feedback and advice.

My thanks go to Vânia Brissos, who I bothered particularly in the early years and who was often a companion in my (mis)adventures. I am grateful to Diogo Silva, Carolina Rodrigues, and André Taborda, fellow PhD candidates and companions in laughter, beers, and therapeutic conversations. I also wish to thank Carolina Dias, Catarina, Tiago, Magalí, and Riccardo, who, though we spent less time together, contributed positively to my experience.

To Ana, without whom I would have been just an *"Englishman in New York"* — interestingly, she was the one who found the scholarship application posted by Lígia, making her, in a way, the architect of all this. My gratitude also extends to her family, who welcomed me like a son from the first moment.

I thank my parents, who raised me, provided me with everything, and, most importantly, gave me the freedom and support to pursue my many interests and dreams — a list that was anything but short.

Lastly, I extend my gratitude to Fundação para a Ciência e a Tecnologia (FCT) for funding my thesis work through a Ph.D. grant (2020.08246.BD), Bio-based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 Research and Innovation Program (GA 837890), and COST-Action COZYME (CA21162).

The struggle itself towards the heights is enough to fill a man's heart. One must imagine Sisyphus happy – The Myth of Sisyphus, Albert Camus.

ABBREVIATIONS

$t_{1/2}$ - Half-life time

$(NH_4)2Fe(SO_4)_2$ - Ferrous ammonium sulfate

4PG - 4-*n*-propylguaiacol

Abs - Absorbance

ABTS - 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)

Ado - Aromatic dioxygenase

ASR - Ancestral-sequence reconstruction

AY - Activity yield

***Bc*Pad -** Phenolic acid decarboxylase from *Bacillus coagulans*

C3H - Coumarate 3-hydroxylase

CCO - Carotenoid cleavage oxygenase

CCR - Cinnamoyl-CoA reductase

$CDCl_3$ – Deuterated chloroform

CoA - Coenzyme A

COMT - Caffeate-O-methyltransferase

Cso2 - Oxygenase 2 from *Caulobacter segnis*

Cys or **C -** Cysteine

dd - doublet of doublets

DMP - 2,6-dimethoxyphenol

DMSO - Dimethyl sulfoxide

dNTP - Deoxyribonucleotide

dq - Doublet of quartets

DyP - Dye-decolorizing peroxidase

ECH - Enoyl-CoA hydratase/aldolase

EDTA - Ethylenediaminetetraacetic acid

***ep*PCR** - *error-prone* polymerase chain reaction

ε - Molar extinction coefficient

FAD - Flavin adenine dinucleotide

FCS - Feruloyl-CoA synthetase

Fdc - Ferulic acid decarboxylase

$FeSO_4$ - Ferrous sulfate

FRESCO - Framework for Rapid Enzyme Stabilization by Computational Libraries

f_U - Unfolded protein

GdnHCl - Guanidine hydrochloride

GSH - Glutathione

GSSG - Glutathione dimer

GST - Glutathione-S-transferase

H_2O_2 - Hydrogen peroxide

HCl - Chloridric acid

HPLC - High performance liquid chromatography

***Hs*CCO** - Carotenoid cleavage oxygenase from *Herbaspirillum seropedicae*

IE - Immobilization efficiency

IEM - Isoeugenol monooxygenase

IEUG - Isoeugenol

IPTG - Isopropyl-β-D-thiogalactopyranoside

IY - Immobilization yield

k_{cat} - Turnover number

k_{cat} /K_m - Catalytic efficiency

KLD - Kinase-Ligase-DpnI mixture

k_{in} - Inactivation constant

K_m - Michaelis constant

KPi - Potassium phosphate

Lac - Laccase

LB - Luria Bertani broth

LiP - Lignin peroxidase

LME - Lignin modifying enzyme

LOX - Lipoxygenase

LsdA - Lignostilbene dioxygenase A from *Sphingomonas paucimobilis*

m - Multiplet

MCO - Multicopper oxidases

McoA - Multicopper oxidase from *Aquifex aeolicus*

MD - Molecular dynamics

$MgCl_2$ - Magnesium chloride

$MnCl_2$ - Manganese chloride

MnP - Manganese peroxidase

MSA - Multiple sequence alignment

NaCl - Sodium chloride

NaOH - Sodium hydroxide

NMR - Nuclear magnetic resonance spectroscopy

NOV1 or **LsdD -** Lignostilbene dioxygenase D from *Novosphingobium aromatocivorans*

OD - Optical density

PCR - Polymerase chain reaction

PIC - Protein Interaction Calculator

***Pn*IEM -** Isoeugenol monooxygenase from *Pseudomonas nitroreducens*

***Pp*IEM -** Isoeugenol monooxygenase from *Pseudomonas putida*

PROGO - 4PG oxidase from *Rhodococcus jostii* RHA1

PROSS - Protein Repair One Stop Shop

QM - Quantum refinement

***Rb*CCO -** Carotenoid cleavage oxygenase from *Rhodobacteraceae bacterium*

RCF - Reductive catalytic fractionation

RDF - Radial Distribution Function

SCO -Stilbene cleavage oxygenase

SDS-PAGE - Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SMARTBOX - Selective Modifications of ARomatics through Biocatalytic Oxidations

T_{agg} - Onset of aggregation temperature

TAO*sm* - *t*-anethole oxygenase from *Stenotrophomonas maltophilia*

T_m - Melting temperature

TPTZ - 2,4,6-tris(2-pyridyl)-s-triazine

Trp or **W** - Tryptophane

TTN - Total turnover number

VAO - Vanillyl alcohol oxidase

VP - Versatile peroxidase

***Vp*VAN -** Hydratase/lyase-type from *Vanilla planifolia*

ABSTRACT

Lignocellulose is the main component of all plant materials, primarily composed of three biopolymers: lignin, cellulose, and hemicellulose. While cellulose and hemicellulose are widely used as renewable sources, lignin often remains a low-value by-product despite being the most abundant source of phenolics. All lignocellulose components, including lignin, must be fully valorized to make the biorefinery concept feasible and sustainable. Lignin can be transformed into value-added products, including vanillin, the most used flavor and fragrance globally. Due to the limited cultivation of vanilla plants, most commercial vanillin is produced through chemical synthesis from crude oil. The growing consumer demand for natural and bio-based products has spurred interest in eco-friendly alternative routes for vanillin production, with lignin-derived vanillin being particularly advantageous over its crude oil-derived counterpart, having approximately a 90% lower carbon footprint.

This thesis was initially developed as part of the SMARTBOX European project (https://www.smartbox-project.eu), which aimed to provide bio-based industries access to inexpensive, robust, and efficient oxidative enzymes that can be readily engineered to meet industrial requirements. A key feature of the project was its exclusive focus on products derived from reductive catalytic fractionation (RCF), such as 4-*n*-propylguaiacol (4PG), to be converted into high-value products like vanillin. The proposed pathway consisted of two reaction steps: (1) converting 4PG into isoeugenol, and (2) the oxidative cleavage of the C=C bond to produce vanillin. While the optimization of the first step, catalyzed by an engineered vanillyl alcohol oxidase (VAO), was carried out by Prof. Marco Fraaije's group at the

University of Groningen, the optimization of the second step was undertaken by Ligia Martins' group at ITQB and developed during my thesis work.

Initially, two bacterial isoeugenol oxygenases (IEMs) were considered: *Pp*IEM from *Pseudomonas putida* and *Pn*IEM from *Pseudomonas nitroreducens*. However, due to their tendency to aggregate and low production yields, a more suitable alternative was searched, and attention shifted to NOV1, an isoeugenol dioxygenase from *Novosphingobium aromatocivorans*, which exhibited significantly higher solubility and production yields than the previous enzymes. The enzymes were kinetically and biochemically characterized; NOV1 displayed a K_m of 0.6 ± 0.1 mM and a k_{cat} of 7.5 ± 0.2 s^{-1}, with a half-life of 2 h at 25 °C. Additionally, NOV1 demonstrated a higher melting temperature (T_m) of 58 °C compared to *Pn*IEM’s 43 °C, indicating higher thermal stability, making it a promising candidate for vanillin production. Based on these findings and the availability of a crystal structure in the database, NOV1 was selected as the most promising catalyst for protein engineering experiments.

In the first stage, a rationally guided mutagenesis near the active site was performed to enhance NOV1’s catalytic efficiency. Among thirty-five variants, S283F was selected, showing a 2-fold increase in activity and a 20-fold increase in half-life compared to the wild-type enzyme. Biochemical and molecular dynamics studies revealed that the S283F mutation enhanced the stabilization of the enzyme’s iron cofactor, which was identified as a critical factor for maintaining enzyme stability. Furthermore, resting *E. coli* cells, previously overproduced NOV1-S283F, converted over 99% of 100 mM isoeugenol into vanillin within 24 h. We developed a coupled whole-cell reaction system to convert 4PG into vanillin. We used the engineered VAO to convert 4PG into isoeugenol and NOV1-S283F for the second reaction step. Following optimization, this approach achieved a 90% vanillin yield starting from commercial 4PG and a 66% yield using 4PG derived from natural lignin oil. This result demonstrated, for the first time, the feasibility of enzymatic conversion of lignin-derived 4PG to vanillin, highlighting the potential of NOV1-S283F as an industrial biocatalyst while providing essential insights into the structure-function relationships in this class of enzymes.

In the second stage, we aimed to improve NOV1 using the computational tool Zymspot to identify distal hotspots located more than 9 Å from the active site. Forty-one variants were constructed using site-directed mutagenesis, and six of the most active variants were selected for further recombination. Two variants, 12G2 (M324K_L372A) and 1D2 (I222V_L286W_L372A), exhibited a 10-fold increase in activity and up to a 40-fold improvement in operational stability compared to the wild-type enzyme. These two variants' mutations are placed in a second shell, indirectly interacting with the iron-binding histidines. The potential of these two catalysts was tested in bioconversions with immobilized enzymes. Enzyme immobilization is crucial for industrial applications, offering advantages such as increased enzyme stability and recyclability. These variants show 90 to 100% immobilization efficiency in metal affinity resins, compared to approximately 60% for the wild-type. In bioconversion tests, where 50 mM of isoeugenol was added stepwise over 24-hour cycles, the 1D2 variant produced approximately 144 mM of vanillin after six reaction cycles, corresponding to around 22 mg, indicating a 35% molar conversion yield. This output was around 2.5 times higher than that obtained using the wild-type. Our findings highlight the efficacy of distal protein engineering in enhancing enzyme functions such as activity, stability, and metal binding selectivity.

Finally, we used a novel computational tool to predict optimal tag sites for immobilization, considering the enzyme’s secondary structure, solvent accessibility, proximity to the active site, and residue conservation. State-of-the-art immobilization strategies often lead to suboptimal enzyme orientations, reducing activity and stability. This tool was validated using a heme-dependent L-tyrosine hydroxylase from *Streptomyces sclerotialus*, chosen for its potential in L-DOPA production. Two immobilization strategies were computationally designed and tested: (1) Lys-tags for covalent immobilization and (2) His-tags for affinity immobilization. A total of 20 variants were constructed, produced, and assessed for activity after immobilization. The computational predictions proved highly effective for His-tagged enzymes: our data suggests that four His residues are sufficient to ensure specific binding and that loop location is essential. Notably, the immobilized activity of our His-tagged variants was up to 5-fold higher than that of the N-terminal 6His-Wild type, which served as a benchmark, highlighting the effectiveness and potential of our strategy to enhance enzyme immobilization.

In conclusion, the work in this thesis paves the way for the enzymatic conversion of RCF-derived 4PG to vanillin. We expanded the NOV1 dioxygenase enzyme's structure-function knowledge, identifying activity determinants, substrate specificity and kinetic stability. We developed and optimized two bioconversion approaches, whole-cell and immobilized enzymes, that achieved high yields of vanillin. Additionally, we highlighted the potential of computational tools to improve enzyme features such as activity, stability, and immobilization efficiency. Future investigations may elucidate more about NOV1's biological function, expand its substrate range, and provide deeper insights into the catalytic mechanisms of this class of enzymes, enhancing their industrial potential.

RESUMO

A lignocelulose é o principal componente de todos os materiais vegetais, composta principalmente por três biopolímeros: lignina, celulose e hemicelulose. Enquanto a celulose e a hemicelulose são amplamente utilizadas como fontes renováveis, a lignina muitas vezes permanece um subproduto de baixo valor, apesar de ser a fonte mais abundante de compostos fenólicos. Para que o conceito de biorrefinaria seja realmente viável e sustentável, todos os componentes da lignocelulose, incluindo a lignina, devem ser plenamente valorizados. A lignina pode ser transformada em vários produtos de valor acrescentado, incluindo a vanilina, o aroma e fragrância mais utilizados globalmente. A maioria da vanilina comercial é produzida por síntese química a partir do petróleo bruto, devido à limitada produção de plantas de baunilha. A crescente procura dos consumidores por produtos naturais e de base biológica estimulou o interesse por rotas alternativas e ecológicas para a produção de vanilina, sendo a vanilina derivada da lignina particularmente vantajosa em comparação com a sua contraparte derivada do petróleo, por apresentar uma emissão de CO_2 aproximadamente 90% inferior.

Esta tese foi inicialmente desenvolvida no âmbito do projeto europeu SMARTBOX (https://www.smartbox-project.eu), cujo objetivo era proporcionar às bioindústrias o acesso a enzimas oxidativas económicas, robustas e eficientes, que possam ser facilmente modificadas para cumprir os requisitos industriais. Uma característica fundamental do projeto foi o foco exclusivo em produtos derivados da fracionação catalítica redutiva (RCF), como o 4-*n*-propilguaiacol (4PG), a ser convertido em produtos de alto valor, como a vanilina. O percurso proposto consistiu em duas etapas de reação: (1)

a conversão do 4PG em isoeugenol e (2) a clivagem oxidativa da ligação C=C para produzir vanilina. Enquanto a otimização da primeira etapa, catalisada por uma vanilil-álcool oxidase (VAO) modificada, foi realizada pelo grupo do Prof. Marco Fraaije na Universidade de Groningen, a otimização da segunda etapa foi desenvolvida pelo grupo da Dra. Lígia Martins no ITQB e durante o trabalho desta tese.

Inicialmente, foram consideradas duas isoeugenol oxigenases bacterianas (IEMs): *Pp*IEM de *Pseudomonas putida* e *Pn*IEM de *Pseudomonas nitroreducens*. No entanto, devido à tendência de agregação e aos baixos rendimentos de produção, procurou-se uma alternativa mais adequada. Como resultado, a atenção foi direcionada para a NOV1, uma isoeugenol dioxigenase de *Novosphingobium aromatocivorans*, que apresentou significativamente maior solubilidade e rendimentos de produção em comparação com as enzimas anteriores. NOV1 exibiu um K_m de 0.6 ± 0.1 mM e um k_{cat} de 7.5 ± 0.2 s^{-1}, com uma meia-vida de 2 horas a 25 °C. Além disso, a NOV1 demonstrou uma temperatura de fusão (T_m) superior, de 58 °C, em comparação com os 43 °C de *Pn*IEM, indicando maior estabilidade térmica, tornando-a uma candidata promissora para a produção de vanilina. Com base nestes achados e na disponibilidade de uma estrutura cristalina na base de dados, a NOV1 foi selecionada como o catalisador mais promissor para experiências de engenharia de proteínas.

Primeiro, realizámos uma mutagénese guiada racionalmente próxima ao sítio ativo para melhorar a eficiência catalítica da NOV1. Entre trinta e cinco variantes, a S283F mostrou um aumento de 2 vezes na atividade e de 20 vezes na meia-vida em comparação com a variante *wild-type*. Estudos bioquímicos e de dinâmica molecular revelaram que a mutação S283F melhorou a estabilização do cofator de ferro (II) da enzima, identificado como um fator crítico para a manutenção da estabilidade da enzima. Além disso, células *E. coli* que anteriormente produziram NOV1-S283F converteram mais de 99% de 100 mM de isoeugenol em vanilina em 24 horas. Para cumprir os objetivos do projeto SMARTBOX, desenvolvemos um sistema de reação acoplada em *whole-cell* para converter 4PG em vanilina, utilizando a VAO modificada para converter 4PG em isoeugenol e a NOV1-S283F para produzir vanilina. Após a otimização, esta abordagem atingiu um rendimento de 90% de vanilina a partir de 4PG comercial e 66% utilizando 4PG derivado de óleo de lignina natural. Demonstrámos, pela primeira vez, a viabilidade da conversão enzimática de 4PG derivado da lignina em vanilina, destacando o potencial da NOV1-S283F como um biocatalisador industrial e proporcionando importante conhecimento sobre as relações estrutura-função nesta classe de enzimas.

A seguir, visámos melhorar NOV1 usando a ferramenta computacional Zymspot para identificar *hotspots* distais localizados a mais de 9 Å do sítio ativo. Foram construídas quarenta e uma variantes por mutagénese, e seis das variantes mais ativas foram selecionadas para recombinação. Duas variantes, 12G2 (M324K_L372A) e 1D2 (I222V_L286W_L372A), exibiram um aumento de 10 vezes na atividade e até 40 vezes na estabilidade operacional em comparação com a enzima wild-type. A estabilidade cinética pode resultar da estabilização do cofator de ferro, possivelmente por um efeito conformacional entre as mutações distais e o sítio ativo, semelhante ao observado com a S283F. As mutações nestas duas variantes estão, de facto, posicionadas numa segunda camada, interagindo indiretamente com as histidinas de ligação ao ferro. O potencial destes dois catalisadores foi testado em bioconversões com enzimas imobilizadas. A imobilização enzimática é crucial para aplicações industriais, oferecendo vantagens como maior estabilidade e reciclabilidade das enzimas. Estas variantes apresentam uma eficiência de imobilização de 90 a 100% em resinas de afinidade a metais, em comparação com aproximadamente 60% da enzima *wild-type*. Nos testes de bioconversão, onde 50 mM de isoeugenol foram adicionados em etapas ao longo de ciclos de 24 horas, a variante 1D2 produziu aproximadamente 144 mM de vanilina após seis ciclos de reação, correspondendo a cerca de 22 mg, indicando um rendimento molar de conversão de 35%. Este resultado foi cerca de 2.5 vezes superior ao obtido com a enzima *wild-type*. Os nossos resultados destacam a eficácia da engenharia de proteínas distais na melhoria de funções enzimáticas, como a atividade, a estabilidade e a seletividade de ligação a metais.

As estratégias de imobilização de última geração muitas vezes resultam em orientações enzimáticas subótimas, reduzindo a atividade e a estabilidade. No capítulo 5, utilizámos uma nova ferramenta computacional para prever os locais ótimos de marcação para imobilização, considerando a estrutura secundária da enzima, acessibilidade ao solvente, proximidade ao sítio ativo e conservação dos resíduos. Esta ferramenta foi validada utilizando uma L-tirosina hidroxilase dependente de heme de *Streptomyces sclerotialus*, escolhida pelo seu potencial na produção de L-DOPA. Foram desenhadas e testadas duas estratégias de imobilização: (1) marcas de lisina para imobilização covalente e (2) marcas de histidina para imobilização por afinidade. Foram construídas, produzidas e avaliadas um total de 20 variantes em ensaios de atividade após a imobilização. As previsões computacionais mostraram-se altamente eficazes para as enzimas com marcação de histidina: os nossos dados sugerem que quatro resíduos de histidina são suficientes para garantir uma ligação específica, e que a localização na alça é essencial. Importa referir que a atividade das variantes marcadas com histidina após imobilização foi até 5 vezes superior à da enzima *wild-type* com 6 histidinas no N-terminal, que serviu como referência, destacando a eficácia e o potencial da nossa estratégia para melhorar a imobilização enzimática.

Em conclusão, esta tese abriu o caminho para a conversão enzimática de 4PG derivado de RCF em vanilina. Ampliámos o conhecimento sobre as relações estrutura-função da enzima NOV1 dioxigenase, identificando determinantes de atividade, especificidade de substrato e estabilidade cinética. Desenvolvemos e optimizámos dois processos de bioconversão, utilizando células inteiras e enzimas imobilizadas, que atingiram elevados rendimentos de vanilina. Adicionalmente, destacámos o potencial de ferramentas computacionais para melhorar características enzimáticas, como a sua atividade, estabilidade e eficiência de imobilização. Investigações futuras poderão elucidar mais sobre a função biológica da NOV1, expandir a sua gama de substratos e fornecer um conhecimento mais profundo sobre os mecanismos catalíticos nesta classe de enzimas, ampliando o seu potencial industrial.

List of Publications

Mario De Simone, Laura Alvigini, Lur Alonso-Cotchico, Vânia Brissos, Jonatan Caroli, Maria Fátima Lucas, Emanuele Monza, Eduardo Pinho Melo, Andrea Mattevi, and Lígia O. Martins (2022). Rationally Guided Improvement of NOV1 Dioxygenase for the Conversion of Lignin-Derived Isoeugenol to Vanillin. *ACS Biochemistry*.
https://doi.org/10.1021/acs.biochem.2c00168

Ivana Marić, Yiming Guo, Maximilian J.L.J. Fürst, Korneel Van Aelst, Sander Van den Bosch, **Mario De Simone**, Lígia O. Martins, Bert F. Sels, Marco W. Fraaije. A One-Pot, Whole-Cell Biocatalysis Approach for Vanillin Production using Lignin Oil". *Advanced Synthesis and Catalysis* 365 22 (2023): 3987-3995. https://doi.org/10.1002/adsc.202300868

Mario De Simone, Lur Alonso-Cotchico, Maria Fátima Lucas, Vânia Brissos, Lígia O. Martins. Distal mutations enhance efficiency of free and immobilized NOV1 dioxygenase for vanillin synthesis. *Journal of Biotechnology 391,* (2024), 92-98.
https://doi.org/10.1016/j.jbiotec.2024.06.012

Table of Contents

Acknowledgements ii
Abbreviations iv
Abstract viii
Resumo xiv
List of Publications xx
Table of Contents xxii
Chapter 1 1
- **1.1. Introduction** 3
- **1.2. Lignin Extraction and Depolymerization — General Considerations** 6
  - 1.2.1. Chemocatalytic Processes 9
  - 1.2.2. Enzymatic Depolymerization 10
- **1.3. Downstream Processing Methods** 14
  - 1.3.1. Chemocatalytic Funneling 16
  - 1.3.2. Biological Funneling 17
- **1.4. Overcoming the Enzyme Limitations for Industrial Application** 23
  - 1.4.1. Protein Engineering 23
  - 1.4.2. Whole-cell Reactions and Enzyme Immobilization 26
- **1.5. Concluding Remarks and Thesis Outline** 28
- **1.6. REFERENCES** 32

Chapter 2 43
- **2.1. ABSTRACT** 45
- **2.2. INTRODUCTION** 47
- **2.3. MATERIALS AND METHODS** 50
  - 2.3.1. Production of *Pp*IEM, *Pn*IEM and NOV1 50
  - 2.3.2. *Pn*IEM Purification from the Soluble Fraction 51
  - 2.3.3. *Pn*IEM Purification from Inclusion Bodies 51
  - 2.3.4. *In silico* stability design 52
  - 2.3.5. Biochemical Characterization 52
  - 2.3.6. Thermostability Analysis 53

2.3.7. NMR analysis of coniferyl alcohol conversion 53
2.3.8. Whole-cell Reaction 54
2.3.9. Other Methods 54
2.4. RESULTS AND DISCUSSION 55
2.4.1. Production of *Pp*IEM and *Pn*IEM 55
2.4.2. Biochemical Characterization of *Pn*IEM 58
2.4.3. Computational Engineering to Improve Solubility and Stability 60
2.4.4. Production of NOV1 from *N. aromaticivorans* 63
2.4.5. Substrate Scope of NOV1 from *N. aromatocivorans* 66
2.4.6. Isoeugenol to Vanillin Conversion by Whole-Cells 69
2.5. CONCLUDING REMARKS 69
2.6. REFERENCES 70
Chapter 3 75
3.1. ABSTRACT 77
3.2. INTRODUCTION 79
3.3. MATERIALS AND METHODS 84
3.3.1. NOV1 Mutants Design with Rosetta Coupled Moves 84
3.3.2. NOV1 Mutants Construction using Site-directed Mutagenesis 84
3.3.3. Activity Screenings in 96-well plates 85
3.3.4. NOV1 Production and Purification 86
3.3.5. Apparent Steady-state Kinetic Analysis 87
3.3.6. Thermodynamic Stability 88
3.3.7. Kinetic Stability 88
3.3.8. Protein crystallization, Data collection, and X-ray Structure Determination and Refinement 89
3.3.9. Molecular Dynamics Simulations 89
3.3.10 Inter-residue Interactions 90
3.3.11. Ensemble Protein-Ligand Docking 90
3.3.12. Bioconversions of Isoeugenol to Vanillin 91
3.3.13. Bioconversion of 4PG to Vanillin by Coupled Reactions 92
3.3.14. HPLC analysis 92
3.3.15. Isolation of Vanillin and NMR Analysis 93
3.4 RESULTS AND DISCUSSION 94

3.4.1. Active Site Design: Construction and Characterization of Variants 94
3.4.2. Biochemical and Kinetic Characterization 97
3.4.3. NOV1 stability and Iron Incorporation 99
3.4.4. Structural Characterization of the S283F variant 102
3.4.5. Molecular Dynamics Simulations and Substrate Docking 102
3.4.6. Enzymatic Vanillin Bio-production 105
3.4.7. Activity of Individual Whole-Cells toward 4PG and Isoeugenol 108
3.4.8. Optimizing the Productivity 110
3.4.9. Cascade Conversions: Simultaneous vs. Stepwise Addition of Whole-Cell Catalysts 111
3.4.10. One-Pot Conversion of Lignin Oil 114
3.5. CONCLUDING REMARKS 117
3.6. REFERENCES 118
Chapter 4 123
4.1. ABSTRACT 125
4.2. INTRODUCTION 127
4.3. MATERIALS AND METHODS 129
4.3.1. Bacterial strains, plasmids, and culture media 129
4.3.2. Identification of NOV1 Distal Hotspot Residues 129
4.3.3. Construction of *nov1* Variants using Site-directed Mutagenesis 130
4.3.4. Recombination by DNA Shuffling 130
4.3.5. High-throughput Activity Screening in 96-well plates 131
4.3.6. Enzyme Production and Purification 132
4.3.7. Apparent Steady-state Kinetic Analysis 132
4.3.8. Thermodynamic and Kinetic Stability 133
4.3.9. Immobilization on Metal-loaded Resin by Affinity Interaction 133
4.3.10. Bioconversions of Isoeugenol to Vanillin by Immobilized Enzymes 134
4.4. RESULTS AND DISCUSSION 135
4.4.1. Identification of NOV1 Distal Hotspots 135
4.4.2. Construction, Screening, and Characterization of Variants 135
4.4.3. Mapping Mutations in the NOV1 X-ray Structure 138
4.4.4. Recombination of Mutations by DNA Shuffling 139
4.4.5. Affinity Immobilization of Wild-type and Variants 142

4.4.6. Bioconversion of Isoeugenol into Vanillin using Immobilized Enzymes 144

4.5. CONCLUDING REMARKS **147**

4.6. REFERENCES **147**

Chapter 5 .... **151**

5.1. ABSTRACT **153**

5.2. INTRODUCTION **155**

5.3. MATERIALS AND METHODS **158**

5.3.1. Design of Variants 158

5.3.2. Construction of Variants 158

5.3.3. Production and Purification 159

5.3.4. Biochemical Characterization 160

5.3.5. Immobilization on Metal-loaded Resin by Affinity Interaction 160

5.4 RESULTS AND DISCUSSION **162**

5.4.1. Design and Construction of the Mutants 162

5.4.2. Biochemical Characterization 164

5.4.3. Immobilization on Epoxy Resin by Lys-tag 166

5.3.4. Immobilization by His-tag on Metal-loaded Resin by Affinity Interaction 168

5.5. CONCLUDING REMARKS **170**

5.6. REFERENCES **171**

Chapter 6 .... **173**

6.1. Conclusions and Future Perspectives **175**

6 2. REFERENCES **178**

Annexes **S1**

Annex I **S3**

Annex II **S13**

Annex III **S17**

Chapter 1

Vanillin from Lignin: an Overview of the Current Chemo- and Biocatalytic Processes

1.1. INTRODUCTION

One of the 21st-century scientific research challenges is developing new sustainable bioprocesses that utilize renewable products, materials, fuels, and energy sources. These new bioprocesses are essential to address fossil feedstocks' depletion and global environmental concerns. Over the past decades, numerous strategies have been developed to valorize biomass into high-value products. Among these, plant-based biomass derived from lignocellulose is one of the most abundant and exploited (Linger *et al.* 2014). Lignocellulose is one of the few renewable resources that is CO_2-neutral, widely available and does not compete with food production. Lignocellulose is the main component of all plant materials and is mainly constituted by three biopolymers: lignin (15-30%), cellulose (40-60%), and hemicellulose (10-40%), each of them showing a peculiar structure (Cano-Díaz *et al.* 2021; Mori *et al.* 2015) (**Fig.1.1**). A classic example of a lignocellulosic biorefinery is a pulp or paper mill, which produces high-quality pulps for the paper and cardboard industry (Mongkhonsiri *et al.* 2020). Another well-known example is the fermentative production of bioethanol from lignocellulosic carbohydrates (Limayem and Ricke 2012; Hahn-Hagerdal *et al.* 2006). In these examples, the primary products are derived from the carbohydrate fraction, while the lignin portion is often considered a low-value by-product or a cheap energy source. Therefore, the wood industry generates more than 100 million tons of lignin residue annually, but it has no edible utility and remains underutilized (Bajwa *et al.* 2019). Despite that, to achieve economic viability and environmental sustainability of a biorefinery and overcome the initial energy cost associated with processing and pretreatment, the valorization of lignin should be fully integrated (Beckham *et al.* 2016; Borchert, Henson, and Beckham 2022; Linger *et al.* 2014; Ragauskas *et al.*

2014). In this regard, the catalytic conversion of lignin is highly challenging, mainly due to the robustness and complexity of its structure (Beckham *et al.* 2016; Zakzeski *et al.* 2010). As the largest renewable source of aromatic building blocks in nature, lignin has significant potential as a starting material for the production of many bulk or functionalized aromatic compounds, offering a suitable alternative to petroleum-derived chemicals (Borchert, Henson, and Beckham 2022; Sun *et al.* 2018). To now, lignin has been successfully transformed into value-added products such as polyhydroxyalkanoates (bioplastic) (Tomizawa *et al.* 2014; Xu *et al.* 2019), lipids (often used as biofuel) (Chatterjee *et al.* 2020), animal feed additive (Doherty, Mousavioun, and Fellows 2011), pesticides (Turan *et al.* 2019), biofertilizers (Klammsteiner *et al.* 2020), muconic acid (Becker *et al.* 2018) and vanillin (Harshvardhan *et al.* 2017).

The case of vanillin is an excellent example of the potential of lignin valorization. Vanillin, widely used as a flavor and fragrance compound (Banerjee and Chattopadhyay 2019), showed antimicrobial and antioxidant properties (Tomadoni *et al*. 2016) and can also serve as an aromatic building block to produce epoxy resins, polyesters, food preservatives, and pharmaceutical drugs (Fache, Boutevin, and Caillol 2016; Walton, Mayer, and Narbad 2003). Natural vanillin is traditionally extracted from Vanilla pods of the *Vanilla* genus, primarily of the species *Vanilla planifolia*. These plants are native to tropical regions and require specific conditions such as warm temperatures, high humidity, and partial shade for optimal growth. Approximately 1000 kg of vanilla pods are needed to recover approximately 2 kg of vanillin, and its cost is highly dependent on the availability of the pods.

Additionally, vanilla cultivation is exceptionally labor-intensive, discouraging farmers from large-scale cultivation. Consequently, the production of natural vanilla extract represents less than 1% of the total vanillin production, amounting to around 50 tons per year worldwide. Most commercial vanillin is chemically synthesized from crude oil through the guaiacol route (80%) or derived from lignin (20%) (Fache, Boutevin, and Caillol 2016). Consumer demand for inexpensive and natural vanillin has led to investigating the biotechnological routes to satisfy the need for more ecological biosynthetic procedures to make natural flavors. Lignin-derived vanillin is advantageous over vanillin made from crude oil, considering its bio-based nature and presenting approximately 90% lower carbon footprint (Fache, Boutevin, and Caillol 2016). Despite this advantage, these methods still need to improve yield, selectivity, and process efficiency, which slow their implementation in the current biorefinery industry.

This chapter aims to summarize and provide perspectives on the processes involved in converting lignin to vanillin, from the separation (fractionation) from lignocellulose biomass to the final upgrading, emphasizing the research and implementation of enzymatic reactions in biorefinery. Strategies such as protein engineering, whole-cell bioconversions, and enzyme immobilization, which are employed to overcome enzymatic limitations, will be discussed.

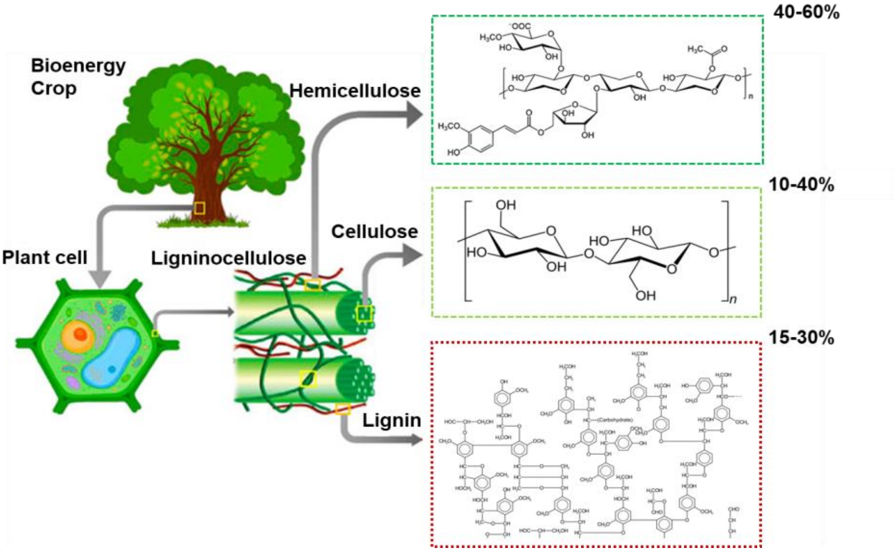
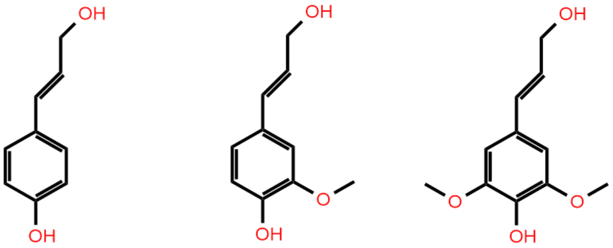


Figure 1.1. Lignocellulose biomass composition. Adapted from (Sankaran *et al.* 2021).

1.2. Lignin Extraction and Depolymerization — General Considerations

The main impediment to using lignin as a rich natural feedstock for aromatic compounds is its heterogeneous and complex structure, complicating its efficient conversion to high-value chemical products. Lignin's biological function is to provide rigidity to the plant cell wall and resistance to microbial attack, thus contributing to biomass recalcitrance (DeMartini *et al.* 2013; Ono and Erhard 2011; Wyman *et al.* 2013). Lignin is defined as an irregular polymer of phenylpropane units linked by C-O-C and C-C bonds with a variable degree of methoxylation (Vanholme *et al.*, 2012). It is constituted of three main building blocks (monolignols), coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S) units, with their ratios varying by species (**Fig. 1.2**).



Coumaryl alcohol (H) Coniferyl alcohol (G) Sinapyl alcohol (S)

Figure 1.2. Chemical structure of the three main precursors of lignin.

For example, softwood lignin mainly comprises G units, while hardwood lignin generally contains both G and S units, with a prevalence of S units (80%) (Li and Wilkins 2021). Lignin-derived monomers most often resemble their parent monolignol structure. Softwood lignin is typically considered the ideal source for vanillin production, as its depolymerization under mild conditions produces a relatively narrow range of methoxy-phenolic monomers, including vanillin and others such as eugenol, isoeugenol, coniferyl alcohol, and ferulic acid, which can be then converted to vanillin more easily (Rodrigues Pinto, Borges da Silva, and Rodrigues 2012; Wu, De Bruyn, and Barta 2023).

The conversion of lignin into value-added chemicals consists mainly of three key biorefinery procedures: lignocellulose fractionation, lignin depolymerization, and subsequent upgrading to desired chemicals (**Fig. 1.3**). One primary challenge in this process is developing industrially feasible lignin isolation methods that minimize structural degradation while maximizing lignin yield and purity (Renders *et al.* 2017).

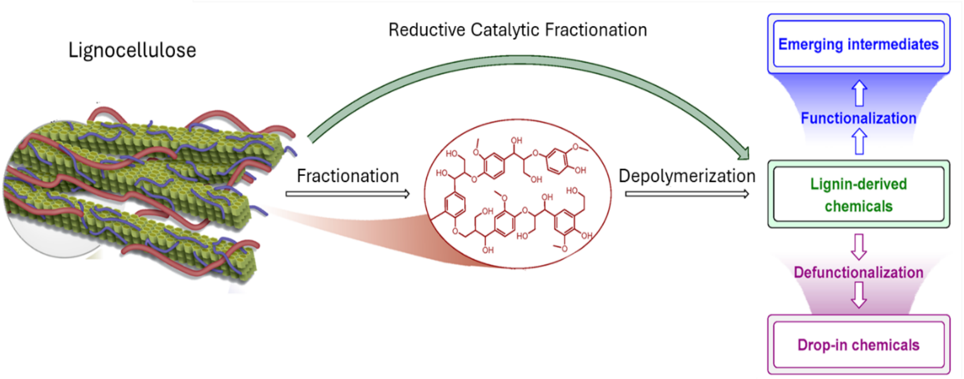


Figure 1.3. Schematic view of lignin valorization steps from the initial lignocellulose biomass to the final upgrade. Adapted from (Korányi *et al.* 2020).

The first step of lignin valorization is its extraction and separation from the other two components, cellulose and hemicellulose, of lignocellulosic biomass. Every fractionation method induces structural alterations to some extent, which can adversely affect the subsequent conversion of lignin into chemicals. Preserving the original lignin structure, particularly the reactive β-O-4 bonds, is crucial to enhance its potential for valorization routes (Schutyser *et al.* 2018). Generally, milder processing conditions are preferred to limit ether bond cleavage and effectively preserve the β-O-4 linkages (Schutyser *et al.* 2018). For example, ammonia-based strategies tend to be milder compared to NaOH-based methods. Analogously, for efficient depolymerization of lignin, the second stage of the lignin-to-chemicals valorization chain, it is crucial to avoid the formation of stable C–C bonds. Since most depolymerization methods cannot cleave C–C bonds in lignin, increasing C-C bond content at the expense of cleavable ether bonds reduces the potential for depolymerization (Schutyser *et al.* 2018).

One strategy to mitigate repolymerization and formation of C-C bonds involves performing lignin depolymerization directly on native lignin, simultaneously executing biomass fractionation and depolymerization. A notable process that accomplishes this integration is reductive catalytic fractionation (RCF). In RCF, native lignin is extracted from lignocellulosic biomass using solvents, such as methanol, and simultaneously depolymerized in the presence of a heterogeneous redox catalyst and a hydrogen source, largely avoiding repolymerization (Galkin and Samec 2016; Parsell *et al.* 2015; Renders *et al.* 2017; Van den Bosch *et al.* 2015). RCF yields a depolymerized lignin oil rich in phenolic monomers, dimers, and oligomers, making it a more suitable feedstock for enzymatic conversion into high-value products. Additionally, RCF offers tunability towards the selective formation of specific lignin monomers (Van den Bosch *et al.* 2015). The final product yield also depends strongly on the lignocellulose source. Generally, monomer yields increase in the order of softwoods, herbaceous crops, and hardwoods, correlating with their respective β-O-4 content (Laskar *et al.* 2013; Mottiar *et al.* 2016; Van den Bosch *et al*. 2015).

1.2.1. Chemocatalytic Processes

Thermochemical lignin conversion into simpler forms occurs in the presence of high temperatures, organic solvents, additives, and catalysts via hydrolysis, oxidation, pyrolysis, or gasification (Chauhan *et al.* 2022). Industrially, the aerobic oxidative depolymerization of lignin in alkaline environments is the most advanced and widely used technology for producing aromatic compounds, including vanillin (Upton and Kasko 2016). This process is based on simultaneous oxidation, using in particular molecular oxygen (O_2) and hydrogen peroxide (H_2O_2), and alkaline hydrolysis at high temperatures in the presence of a transition-metal catalyst

(**Fig. 1.4**), promoting the dissolution of lignin and generating aldehyde groups. Vanillin and other monomers are finally obtained through alkaline hydrolysis with ether bond cleavage (Gillet *et al.* 2017; Pacek *et al.* 2013). The total monomer yield depends on the lignin source and method: alkaline oxidation of kraft lignin generally yields lower than 10 wt% vanillin; oxidation of coniferous lignin in alkaline media can produce up to 28 wt% vanillin, while hardwood lignin can yield up to 51% of vanillin and syringaldehyde mixtures (Luo *et al.* 2021; Tarabanko *et al.* 2017; Xu *et al.* 2023). Although this process described above produces relatively good vanillin yields, high temperatures (120–200 °C), long reaction times, and costly infrastructure remain critical for sustainable production (Pandey and Kim 2011; Xu *et al.* 2023).

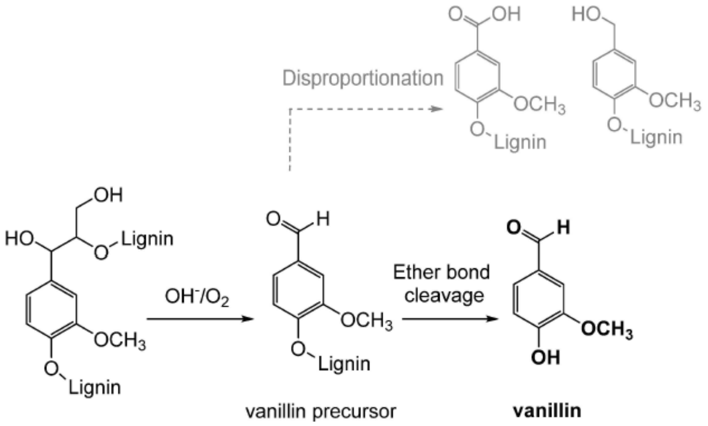


Figure 1.4. Schematic view of vanillin production through alkaline oxidative degradation of lignin (D'Arrigo *et al.* 2024).

1.2.2. Enzymatic Depolymerization

Unlike thermochemical processes, bioprocessing offers distinct advantages characterized by high specificity, low energy requirements, and cost-effectiveness (Chen and Wan 2017). The degradation of lignin by microorganisms and enzymes is a promising area of research for sustainable bioproduction processes. Many fungi and bacteria have developed specialized enzymes and metabolic pathways to break down lignin into valuable compounds like vanillin and other aromatic chemicals. The main classes of ligninolytic enzymes that are grouped under the name of lignin modifying enzymes (LMEs) can be divided into peroxidases, laccases, and β-etherases (Bugg 2024; Rosini *et al.* 2023) (**Table 1.1**).

Peroxidases are heme-containing enzymes that depolymerize lignin (Datta *et al.* 2017; Laskar *et al.* 2013; Mottiar *et al.* 2016). They act on a wide range of lignin-based phenolic and non-phenolic compounds (Houtman *et al.* 2018; Qin *et al.* 2017; Singh *et al.* 2021) with H_2O_2 consumption. This group includes lignin peroxidase (LiP), manganese peroxidase (MnP), versatile peroxidase (VP), and dye-decolorizing peroxidase (DyP). LiPs have been extensively documented to cleave the Cα-Cβ and C-O-C linkages in lignin and oxidize benzyl alcohols to aldehydes (Qin *et al.* 2017). In the case of MnP, the reaction involves the oxidation of Mn^{2+} to Mn^{3+}, which oxidizes many phenolic substrates. MnP can cleave β-O-4 linkages in phenolic structures but also Cα-Cβ and β-aryl ether bonds in non-phenolic (Datta *et al.* 2017). A versatile peroxidase (VP) exhibits both LiP and MnP activity and, indeed, presents a hybrid molecular architecture that combines many oxidation-active regions, allowing high redox catalytic capacity for a wide range of substrates (Gonzalez-Perez and Alcalde 2018; Gonzalez-Perez *et al.* 2016; Liu *et al.* 2019). DyP is a family of heme-containing peroxidases that has received significant attention due to its ability to dye decolorization, degradation of lignin, and other compounds (Silva, Rodrigues, *et al.* 2023; Sugano and Yoshida 2021). Notably, DyP enzymes can cleave C-C bonds without requiring any oxidative mediator (Linde *et al.* 2021; Xu *et al.* 2021).

DyPs can oxidize kraft lignin, syringyl- and guaiacyl-type phenolics (Linde *et al.* 2015; Silva, Sousa, *et al.* 2023). However, comparative studies have shown that only fungal DyPs can oxidize nonphenolic model dimers, which are representative of the most recalcitrant moiety of the lignin macromolecule (Liers *et al.* 2013; Linde *et al.* 2021; Linde *et al.* 2015). Although bacterial DyPs were initially thought to degrade lignin, they are ineffective at oxidizing and breaking down non-phenolic lignin model compounds (Lončar, Colpa, and Fraaije 2016). Therefore, their role in lignin degradation is controversial, suggesting that their eco-physiological role contributes to the catabolism of lignin products resulting from fungal attacks on native lignin (Linde *et al.* 2021). The mechanism of action of peroxidases is well understood; however, the formation of phenoxy radicals usually leads to (re)polymerization (Zeng *et al.* 2017). Depolymerization must occur instead to obtain low molecular weight polymers or specific monomers. The current challenge is determining the factors controlling depolymerization and repolymerization for each specific enzyme to drive the reaction toward the desired product (Cajnko *et al.* 2021).

Laccases (LACs) are multicopper-containing oxidoreductases (MCOs) that oxidize diverse phenolic, non-phenolic, and lignin model compounds utilizing O_2 as an electron acceptor, resulting in water generated as a by-product through a chemical reaction (Jones and Solomon 2015; Mehra *et al.* 2018; Pogni *et al.* 2015). In the reaction catalyzed by LACs, electrons are typically transferred from the substrate to the t_1-Cu site and then to the trinuclear center through a His-Cys-His pathway (Jones and Solomon 2015). The oxidation of phenolic substrates by LACs forms phenoxyl free radical that promotes Cα oxidation, alkyl-aryl cleavage, and Cα-Cβ cleavage, which are typical lignin bonds (Hautphenne, Penninckx, and Debaste 2016). As in the case of peroxidases, laccase activity also tends to result in repolymerization of the aromatic oligomers or monomers due to phenoxy radical coupling (Arzola *et al.* 2006; Hilgers *et al.* 2018).

β-etherases belong to the superfamily of glutathione-S-transferases (GSTs), which are involved in a wide range of processes, including protection against oxidative stress and biodegradation of lignin (Mazari *et al.* 2023; Picart, de Maria, and Schallmey 2015). β-etherases attack lignin on C=O groups by the addition of glutathione (GSH), producing a GSH-conjugate and releasing a monomer; a glutathione lyase uses a second GSH molecule to release GSH from the GSH-conjugates producing a second free monomer and a GSH dimer GSSG (Picart, de Maria, and Schallmey 2015; Picart *et al.* 2017).

In addition to the classes of enzymes mentioned above, an array of auxiliary enzymes (aryl alcohol oxidases, pyranose oxidase, galactose oxidase, vanillyl alcohol oxidase) work in synergy with LMEs, facilitating lignin deconstruction. These enzymes produce compounds that serve as substrates (e.g., H_2O_2 needed by peroxidases) and metabolites that stabilize and chelate enzyme-generated radicals (Bugg 2024).

In conclusion, significant research efforts are still required to understand the viability of this concept for biorefinery applications. Realistic biological processes generally require high titers, rates, and yields of products. Not all compounds may be fully processed through the funnel, potentially leading to lower conversion rates. It is important to recognize that chemocatalytic and biocatalytic strategies have inherent limitations, and one approach does not necessarily exclude the other.

Table 1.1. Major lignin modifying enzymes with their corresponding cofactor, co-substrate, and reaction.

Enzyme	Cofactor	Co-substrate	Reaction	Reference
LiP	Heme	H_2O_2	Cα-Cβ, Cα ox, β-O-4 aryl	Qin *et al.*, 2017
MnP	Heme	H_2O_2, Mn^{2+}	Cα-Cβ	Datta *et al.*, 2017
VP	Heme	H_2O_2	Cα-Cβ, Cα ox	Liu *et al.*, 2019
DyP	Heme	H_2O_2	Cα-Cβ, aryl–Cα	Xu *et al.*, 2021
Laccase	Trinuclear-Cu^{2+}	O_2	Cα-Cβ, Cα ox, aryl–Cα	Hautphenne *et al.*, 2016
β-etherase	-	GSH	β-O-4 aryl	Picart *et al*., 2015

1.3. Downstream Processing Methods

Some monomeric compounds obtained from the various depolymerization approaches can be directly applied in end-use applications, such as vanillin, commonly used as a vanilla fragrance and flavoring agent. However, the depolymerization of lignin often yields complex mixtures of mono-phenolic aldehydes, ketones, and acids (Hu *et al.* 2021; Mei *et al.* 2020; Walch *et al.* 2021). Oxidative depolymerization methods typically produce 5–30 wt% of lignin-derived monomers, primarily comprising vanillin, syringaldehyde, benzaldehyde, and their corresponding acids, depending on the feedstock and method used (Wu, De Bruyn, and Barta 2023). Similarly, reductive approaches like RCF result in a monomer fraction predominantly constituted by 4-propanol guaiacol and 4-propyl guaiacol (Wu, De Bruyn, and Barta 2023). Other compounds usually present in these mixtures are eugenol, isoeugenol, coniferyl and coumaryl alcohols, lactones, and aliphatic compounds (Rodrigues Pinto, Borges da Silva, and Rodrigues 2012). After depolymerization, the resulting crude mixture must first be treated by physicochemical methods like acidification and ultrafiltration to remove residual lignin and extract the phenolic fraction. One approach to enhance selectivity towards aldehydes, particularly useful when producing vanillin, is the bisulfitation of the crude lignin depolymerization mixture (Abdelaziz *et al.* 2019; Rodrigues Pinto, Borges da Silva, and Rodrigues 2012; Tarabanko *et al.* 2013) (**Fig. 1.5**). In this process, the hydrosulfite anion selectively reacts with the aldehyde moiety, while all the high molecular weight products precipitate due to an increase in pH. After removing this sediment, the aldehydes are recovered by further acidification. This method often results in the extraction of a mixture of vanillin, syringaldehyde, and *p*-hydroxybenzaldehyde, depending on the initial feedstock. These extraction methods have significant drawbacks, such as using large quantities of organic solvents like benzene or toluene and extensive acid usage for neutralization and acidification (Fache, Boutevin, and Caillol 2016). Another approach uses ultrafiltration to remove high molecular weight lignin fragments, followed by acidification to extract the phenolic fraction (Zabkova, da Silva, and Rodrigues 2007). However, even with these methods, the resulting product stream primarily consists of crude vanillin mixed with other phenolics. Structurally similar compounds complicate the separation process, leaving food-grade vanillin with significant impurities. One strategy to address this challenge is "funneling", which involves transforming diverse product mixtures into smaller central intermediates that can be more efficiently converted into desired final products (Borchert, Henson, and Beckham 2022; Linger *et al.* 2014). This strategy can involve both chemocatalytic and biological transformations.

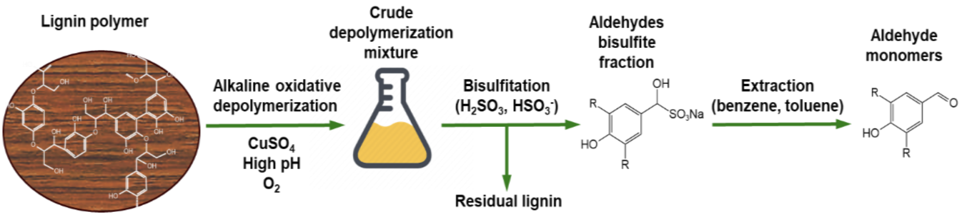


Figure 1.5. Scheme of aldehydes production from lignin by alkaline oxidative depolymerization and bisulfitation (Wu, De Bruyn, and Barta 2023).

1.3.1. Chemocatalytic Funneling

Chemocatalytic funneling transformations can be mainly distinct into transformations that affect the phenolic core and those that change the side-chain structure. Transformations that alter the phenolic core include hydroxygenation and oxidation reactions that decrease product complexity. In the case of vanillin, the phenolic core must be preserved. Therefore, only side-chain transformations are performed, such as dehydrogenation, isomerization, oxidation, and decarboxylation. Many methoxyphenols like eugenol, isoeugenol, coniferyl alcohol, and ferulic acid can be easily converted to vanillin. One well-known example is the isomerization of eugenol to isoeugenol using potassium hydroxide, followed by oxidation to produce vanillin, a method established in the late 19th century to meet the growing demand for vanilla aroma (Fache, Boutevin, and Caillol 2016; Hocking 1997). To yield vanillin, eugenol can also be oxidized with nitrobenzene in a basic dimethyl sulfoxide (DMSO) solution. More recently, an alternative strategy allowed the direct oxidation of ferulic acid to vanillin using metal oxide catalysts (Sanchez-Gonzalez *et al.* 2017; Yepez *et al.* 2015). Alkaline wet oxidation by Cu-Mn mixed oxide catalysts can convert coniferyl alcohol into vanillin using H_2O_2 as an oxidizing agent between

120–180 °C (Jeon *et al.* 2020). However, like chemical depolymerization, chemocatalytic funneling has limitations, like low selectivity with side-product production, use of heavy metals as catalysts, and high temperatures. Furthermore, in these processes, O_2 partial pressure and reaction time must be finely tuned to avoid oxidation of vanillin to vanillic acid.

1.3.2. Biological Funneling

One significant advantage of biological funneling over chemocatalytic processes is the high selectivity and tunability toward specific products (Chubukov *et al.* 2016). Critical enzymes in this process include demethylases, hydroxylases, decarboxylases, and oxygenases (Erickson *et al.* 2022; Ji *et al.* 2024). In plants, vanillin is synthesized from ferulic acid by the hydratase/lyase-type enzyme *Vp*VAN (Gallage *et al.* 2014). Despite the potential of this enzyme, its biotechnological application has not progressed significantly, primarily due to ongoing disagreements about the catalytic activity of *Vp*VAN, as attempts to express this enzyme in heterologous organisms have failed to confirm its activity (Diamond, Barnabé, and Desgagné-Penix 2023; Yang *et al.* 2017). One of the first enzymatic pathways to be proposed for vanillin production is the coenzyme-dependent non-β-oxidative pathway, consisting of two enzymes, feruloyl-CoA synthetase (FCS) and enoyl-CoA hydratase/aldolase (ECH) (**Fig. 1.6**) (Chattopadhyay, Banerjee, and Sen 2018). However, this pathway has notable limitations, such as reliance on expensive cofactors (ATP and CoA), a propensity to produce by-products, and the relatively high cost of ferulic acid (US $180/kg). An alternative was discovered by identifying Cso2 from *Caulobacter segnis.* This enzyme, which belongs to the family of carotenoid cleavage oxygenases (CCOs), can catalyze the cleavage of the C=C bond of 4-vinylguaiacol in a single, coenzyme-independent step, producing vanillin and making this route more advantageous. When combined with a decarboxylase, Fdc from *Bacillus pumilus*, a coenzyme-independent two-step pathway was successfully constructed in *E. coli* (Furuya, Miura, and Kino 2014). This two-stage production process, tested using a whole-cell reaction, resulted in a vanillin yield of 68% starting from ferulic acid in 8 h. However, more recently, ferulic acid conversion to vanillin was made even easier by engineering the active site of the CCO Ado from *Thermothelomyces thermophila*, unlocking activity directly on ferulic acid and producing vanillin in just one step without requiring more enzymes (Fujimaki *et al.* 2024). *E. coli* cells expressing this variant produced 7 mM, 1.1 g/L vanillin from 10 mM ferulic acid in 24 h (**Table 1.2**).

Other vanillin biosynthetic pathways included the oxidative catabolic and the epoxide-diol pathways that start from eugenol and isoeugenol, respectively. The oxidative catabolic pathway, which was first found in *Pseudomonas sp*. HR199 utilizes eugenol as the substrate, which is catalyzed by eugenol hydroxylase, coniferyl alcohol dehydrogenase, and coniferyl aldehyde dehydrogenase to form ferulic acid, which is then converted to vanillin (Overhage *et al.* 1999; Overhage, Priefert, and Steinbüchel 1999). Although eugenol (US $20/kg) is much cheaper than ferulic acid, the drawbacks of this pathway are the yield, which is approximately 1 g/L and the requirement of multiple enzymes (Chen *et al.* 2022). A novel and more straightforward biosynthetic pathway for converting eugenol into vanillin was proposed by introducing a cinnamoyl-CoA reductase (CCR), which catalyzes the conversion of coniferyl aldehyde to feruloyl-CoA, bypassing the need for coniferyl aldehyde dehydrogenase and feruloyl-CoA synthetase (Zhu *et al.* 2024). Introducing this pathway in *E. coli* cells 33 mM eugenol was converted to vanillin with a yield of 88.2% after 24 h reaction. A path that transforms the eugenol isomer, isoeugenol, to vanillin via oxidation of the side chains of isoeugenol has been revealed in bacteria such as *Pseudomonas* (Ryu *et al.* 2013; Yamada *et al.* 2007). The key enzyme in this pathway is isoeugenol monooxygenase (IEM) (belonging to the superfamily of CCOs), encoded by the gene *iem* to convert isoeugenol to vanillin. This pathway has been widely explored for vanillin production due to the high yield and the cheap cost of isoeugenol (US $23–26/kg). *E. coli* cells overexpressing IEM from *Pseudomonas putida* could convert 81% of 230 mM isoeugenol in vanillin after 6 h reaction (Yamada *et al.* 2007). Another dioxygenase able to oxidatively cleave isoeugenol to vanillin is the non-heme iron-containing Lipoxygenase (LOX), which is widely available in plants. However, only low yields were achieved (<40%) using crude enzyme preparations (Liu *et al.* 2020). Recently, a new class of bacterial heme-containing oxygenases that cleave C=C was discovered. The *t*-anethole oxygenase from *Stenotrophomonas maltophilia* (TAO*Sm*) was found to accept *t*-anethole, isoeugenol, and O-methyl isoeugenol as substrates, converting them into the corresponding aromatic aldehydes without the need of any cofactor regeneration (Purwani *et al.* 2024). The application of this last class of enzymes in biocatalysis still needs to be explored due to their very recent discovery.

Another route was recently proposed starting from 4-*n*-propylguaiacol (4PG), which is one of the most common products from RCF (Sun *et al.* 2018; Van den Bosch *et al.* 2015). The FAD-dependent eugenol oxidase from *Rhodococcus jostii* RHA126 was chosen as a promising candidate to catalyze the first step reaction, which consists of dehydrogenation of 4PG to produce isoeugenol (Nguyen *et al.* 2016). Optimization through computational-driven and structure-based mutagenesis increased thermostability and activity toward 4PG obtaining isoeugenol with high chemoselectivity and good yield at the gram scale (Guo *et al.* 2022). This route can then be completed using an isoeugenol oxygenase such as IEM, LOX and TAO.

It should also be noted that vanillin synthesis from coumaryl alcohol derivatives (H units) is technically feasible. For instance, *p*-coumaric acid can be converted to vanillin via ferulic acid by introducing a coumarate 3-hydroxylase (C3H) and a caffeate-O-methyltransferase (COMT) (Ni *et al.* 2015). However, this pathway is still in the proof-of-concept stage and faces limitations such as low yields and the use of expensive cofactors. For this reason, most current studies focus on funneling H units toward the microbial central metabolism to produce biofuels and bioplastics instead of vanillin (Weng, Peng, and Han 2021).

To conclude, while beyond the scope of this review, researchers have proposed enzymatic pathways for vanillin production as well from renewable sources not derived from lignin. Some examples include glucose, poly(ethylene terephthalate) (PET), stilbenoids, and capsaicinoids (Zhang *et al.* 2025).

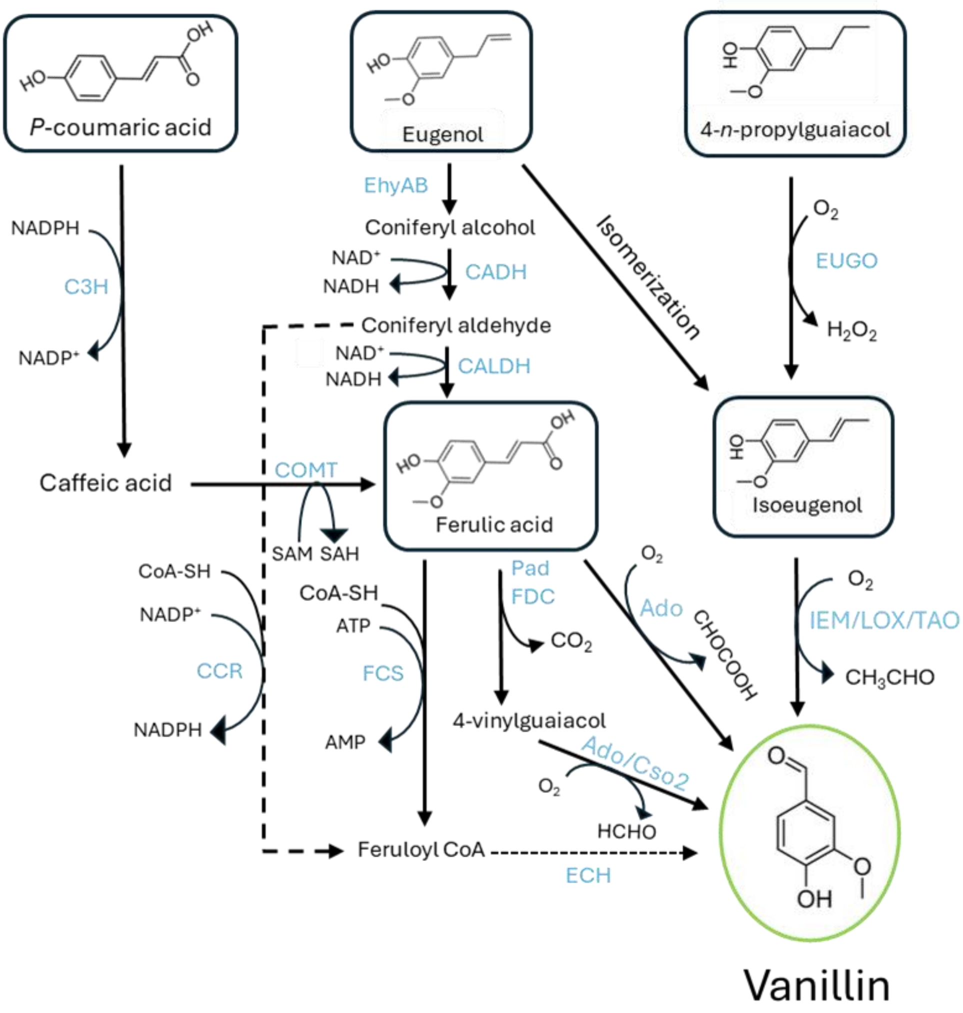


Figure 1.6. Enzymatic conversion of lignin-derived compounds to vanillin. COMT: Catechol-O-methyltransferase; 3DSD: 3-dehydroshikimate dehydratase; SAM: S-adenosylmethionine; SAH: S-adenosylhomocysteine; EhyAB: Eugenol hydroxylase; CADH: Coniferyl alcohol dehydrogenase; CALDH: Coniferyl aldehyde dehydrogenase; FCS: Feruloyl-CoA synthetase; ECH: Enoyl-CoA hydratase/aldolase; Pad: Phenolic acid decarboxylase; Ado: Aromatic dioxygenase; FDC: Ferulic acid decarboxylase; CSO2: carotenoid cleavage oxygenase; EUGO: eugenol oxidase; IEM: isoeugenol monooxygenase; LOX: lipoxygenase; TAO: *t*-anethole oxygenase.

Table 1.2. Bioconversions of isoeugenol (IEUG) and ferulic acid to vanillin using whole-cell and immobilized enzymes system.

Enzyme	Substrate	System	Result	Reference
Ado-m258 (*T. thermophila*)	Ferulic acid	Whole-cell	70% yield from 10 mM ferulic acid in 24 h	Fujimaki *et al*. 2024
Fcs (*Streptomyces sp*. V-1)/ECH (Streptomyces sp. V-1)	Ferulic acid	Whole-cell	75% yield from 20 mM ferulic acid in 6 h	Chen *et al*. 2022
Fdc (*B. pumilus) /* Cso2 (*C. segnis*)	Ferulic acid	Whole-cell	69% yield from 75 mM ferulic acid in 24 h	Furuya *et al*. 2015
Pad (*B. coagulans*)/Ado (*T. thermophila*)	Ferulic acid	Whole-cell	99% yield from 30 mM ferulic acid in 2 h	Ni *et al*. 2018
LOX	Isoeugenol	Crude enzyme	40% yield from 36 mM IEUG in 24 h	Liu *et al*. 2020
IE27 (*P. putida*)	Isoeugenol	Whole-cell	71% yield from 230 mM IEUG in 6 h	Yamada *et al*. 2007
IEM (*P. nitroreducens*)	Isoeugenol	Whole-cell	82% yield from 300 mM isoeugenol in 8 h	Wang *et al*. 2021
HsCCO (*H. seropedicae*)	Isoeugenol	Whole-cell	80% yield from 10 mM isoeugenol in 24 h	Han *et al*. 2019
RbCCO (*R. bacterium*)	Isoeugenol	Whole-cell	75% yield from 10 mM isoeugenol in 24 h	Han *et al*. 2019
Cso2 (*C. segnis*)	Isoeugenol	Immobilized enzyme	6.8 mg yield from isoeugenol in 10 reaction cycles	Furuya *et al*. 2017
Fdc (*B. pumilus) /* Cso2 (*C. segnis*)	Ferulic acid	Immobilized enzyme	2.5 mg yield from ferulic acid in 10 reaction cycles	Furuya *et al*. 2017

1.4. Overcoming the Enzyme Limitations for Industrial Applications

1.4.1. Protein Engineering

Low reaction rates, solubility issues, loss of activity, and specificity often limit the use of enzymatic reactions in biorefineries. One strategy to overcome these limitations is to enhance specific properties through protein engineering. Protein engineering approaches can be broadly categorized into three main methods: (*i*) directed evolution, (*ii*) rational design, and (*iii*) semi-rational design (**Fig. 1.7**).

Directed evolution involves generating diverse enzyme variants through random mutagenesis and screening to identify the best-performing variants. This process mimics natural evolution, applying iterative rounds of mutagenesis and screening to enhance a given property or create new protein catalysts for non-natural chemical transformations (Zeymer and Hilvert 2018). Notably, it only requires partial knowledge of the enzyme's structure or specific mutational targets. However, this method can be time-consuming, and its success largely depends on the effectiveness of a high-throughput screening method. Directed evolution is mainly applied through error-prone polymerase chain reaction (*ep*PCR). In this technique, low-fidelity polymerases, which lack 3′-5′ proofreading exonuclease activity, elevate the error rate of DNA amplification under nonstandard conditions: increased $MgCl_2$, presence of $MnCl_2$ and biased dNTP concentrations (Brissos *et al.* 2024; Buller *et al.* 2018; Herrmann *et al.* 2022; Huang *et al.* 2020; Otten *et al.* 2018). In the field of lignin valorization, a remarkably successful case was the one of McoA from *Aquifex aeolicus* that, after rounds of directed evolution and rational design, led to the identification of a hit variant with a 400-fold increased efficiency (k_{cat}/K_m) than the wild-type for 2,6-

dimethoxyphenol (DMP) (Brissos *et al.* 2024). One limitation of directed evolution is that even the most high-throughput screening can sample only a fraction of the possible sequences that can be made (Yang, Wu, and Arnold 2019). Recombination methods may allow for bigger jumps in sequence space, but these methods are restricted to combinations of previously explored mutations. While directed evolution discards information from unimproved sequences, machine-learning methods use this information to expedite evolution and achieve higher fitness levels than those possible through directed evolution alone by intelligently selecting new variants to screen (Bedbrook *et al.* 2017; Bedbrook *et al.* 2019; Wu *et al.* 2019). However, for broader machine learning applications, scientists must take advantage of unlabeled protein sequences other than those of specific interest to the protein engineer (Yang, Wu, and Arnold 2019).

Rational and semi-rational design strategies utilize knowledge of enzyme structure-function relationships to suggest a library of the most promising target residues for modification to enhance a given property. Algorithms or models often aid these methods in better predicting the success of designed sequences (Marques, Planas-Iglesias, and Damborsky 2021). However, such information is not always readily accessible for a given enzyme and is rarely proficient in predicting distal mutations that may enhance the desired activity (Calzini *et al.* 2021; Yang, Miton, and Tokuriki 2020). Typically, the most common strategy to improve enzyme activity towards a specific substrate involves introducing mutations in the active site. Other approaches have recently proven successful, such as engineering access tunnels, modifying dynamic properties, editing recognition elements like loops, or targeting allosteric sites (Marques, Planas-Iglesias, and Damborsky 2021). Several protein design algorithms have been developed for a variety of applications.

Some prominent examples include the Rosetta computer package (Alford *et al.* 2017), HotSpot Wizard algorithms (Sumbalova *et al.* 2018), and FuncLib (Khersonsky *et al.* 2018). Generally, these tools involve sequential bioinformatics and structure-based analysis (structure prediction, QM energy refinement, sequence conservation, and other techniques) to recommend residues for genetic diversification and experimental screening.

For many biotechnological applications, enzymes must perform for more extended periods or harsher conditions than native variants typically could. Enzyme stability can be enhanced through various approaches, including (*i*) energy calculations, (*ii*) phylogenetic analysis, (*iii*) machine learning, and (*iv*) a combination of these methods. Many computational design methodologies assist this approach, such as FRESCO (Framework for Rapid Enzyme Stabilization by Computational Libraries) (Wijma *et al.* 2014), PROSS (Protein Repair One Stop Shop) (Khersonsky *et al.* 2018), and Fireprot (Musil *et al.* 2024), which can be used to predict enzymes with increased stability. These tools provide in-depth analysis to predict “small but smart” mutant libraries with a high likelihood of enhancing desired enzymatic properties. Another approach, ancestral-sequence reconstruction (ASR), uses phylogenetic relationships among homologous proteins to predict likely ancestral sequences that often show improved thermal stability (Babkova *et al.* 2017; Gumulya *et al.* 2018) and extended substrate range (Clifton *et al.* 2017).

Regarding enzymatic vanillin production, a notable success was obtained with a triple mutant of IEM (K83R_K95R_L273F) from *P. nitroreducens*, showing increased 5-fold k_{cat} and 3-fold half-life at 25 °C (Lu *et al.* 2021). Instead, a semi-rational approach based on position and distance from the active site was used to engineer IEM720 from metagenome (Zhao *et al.*

2018). After screening the site-saturation libraries of five positions, the best hit displayed a 2-fold k_{cat} value compared to the wild-type. A remarkable example was the engineering of Ado from *T. thermophilus*: combining a rational and semi-rational approach, authors could generate a variant capable of producing vanillin from ferulic acid in a one-step reaction (Fujimaki *et al.* 2024).

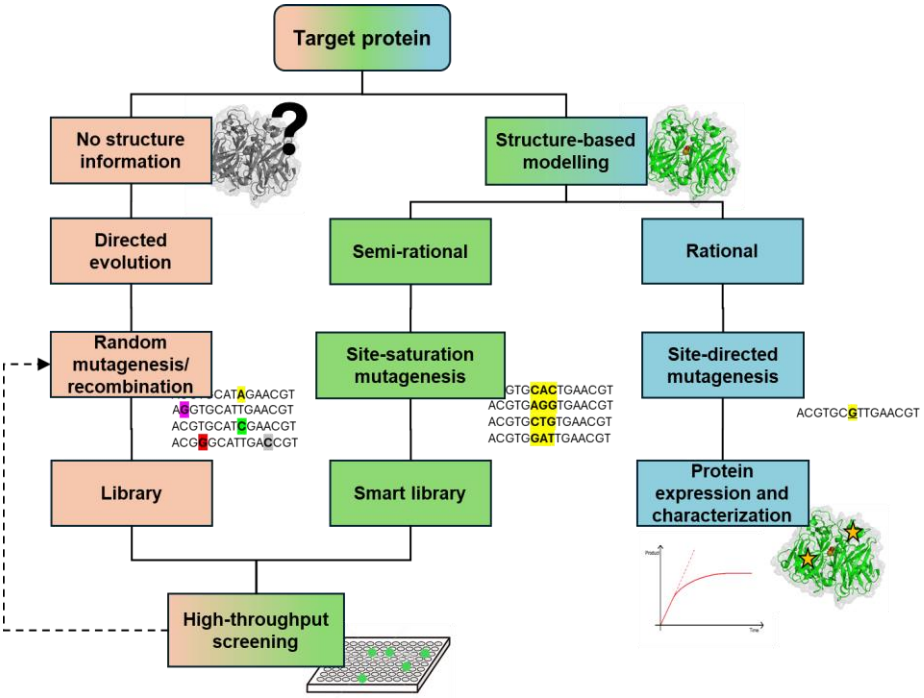


Figure 1.7. Schematic representation of the main protein engineering strategies consisting of directed evolution, semi-, and ration design.

1.4.2. Whole-cell Reactions and Enzyme Immobilization

Whole-cell biocatalysts or enzyme immobilization are essential strategies to promote the implementation of enzymes in biorefineries. These approaches offer significant advantages, including increased enzyme stability and recyclability, making industrial processes more efficient and cost-effective. Some relevant and successful cases are shown in **Table 2**.

Whole-cell biocatalysis has attracted considerable interest in industrial applications due to its cost-effectiveness by eliminating the need for cell lysis and enzyme purification (de Carvalho 2017). Additionally, whole cells can protect the expressed enzymes from potentially harmful environments, maintaining their natural enzymatic conditions for efficient conversion. To apply this strategy, it is crucial that the cell membrane is permeable to the substrate or that the enzyme is secreted to allow catalysis. Regarding vanillin production, the recombinant *E. coli* expressing IEM from *P. nitroreducens* Jin1 achieved a vanillin production of 115 g/L from isoeugenol corresponding to 82% yield (Wang *et al.* 2021). Analogously, *E. coli* cells overproducing *Hs*CCO and *Rb*CCO from *Herbaspirillum seropedicae* and *Rhodobacteraceae bacterium* displayed conversion rates of isoeugenol to vanillin of 80 and 75%, respectively (Han, Long, and Ding 2019). Setting up a decarboxylase/oxygenase cascade to convert ferulic acid to vanillin, *E. coli* cells expressing Cso2 from *C. segnis* efficiently oxidized 4-vinylguaiacol to vanillin. The concentration of vanillin reached 52 mM in 24 h, starting from 75 mM ferulic acid (Furuya *et al.* 2015). The same cascade was also tested, replacing Fdc and Cso2 with thermostable *Bc*Pad from *Bacillus coagulans* DSM1 and Ado from *Thermothelomyces thermophila,* converting 99% of 30 mM ferulic acid in 2 h (Ni *et al.* 2018).

Similarly, enzyme immobilization enhances enzyme stabilization, simplifies product separation, and enables enzyme recycling, thereby improving the efficiency and cost-effectiveness of industrial enzyme applications. Based on the type of interaction with the carrier, these methods can be divided into reversible techniques, such as adsorption, ionic interaction, and metal binding encapsulation, and irreversible techniques, including cross-linking, covalent binding, and entrapment (Maghraby *et al.* 2023). Generally, non-covalent methods offer rapid immobilization, preserve enzyme activity, and allow for catalyst recovery. However, the more unstable interactions make the enzyme more susceptible to leaking. Conversely, covalent binding can sometimes lead to enzyme inactivation, but it provides a robust and durable interaction, even in the presence of high concentrations of substrates or robust buffer solutions. Common drawbacks of enzyme immobilization include the high cost of carriers, low immobilization efficiency, and loss of enzyme activity. However, developing new supports using affinity interactions allows for higher immobilization efficiency, faster immobilization, and the potential for “one-pot” purification and immobilization. These methods offer advantages over conventional cross-linking, covalent, and ionic binding immobilization techniques. While vanillin production using whole cells has been extensively investigated, there are fewer reports on its production using immobilized enzymes. Various immobilization methods have been applied so far. To solve the instability problem of Cso2 and to simplify the process, Cso2 was immobilized on Sepabeads EC-EA anion exchange carriers, enabling the production of 6.8 mg vanillin from isoeugenol at a 1 mL scale over ten cycles of reaction (Furuya, Kuroiwa, and Kino 2017). When Fdc and Cso2 were immobilized simultaneously, the system could synthesize 2.5 mg vanillin from ferulic acid after ten reaction cycles.

1.5. Concluding Remarks and Thesis Outline

Bioconversion of lignin into various value-added compounds can significantly reduce our dependence on petroleum-based fuels and chemicals, a necessity given the limited availability of fossil resources. The degradation of lignin by microorganisms and enzymes is a promising area of research for sustainable bioproduction processes. So far, many microorganisms have evolved specialized enzymes and metabolic pathways to break down lignin into valuable compounds like vanillin and other aromatic chemicals. However, using microbes and enzymes to convert lignin-derived compounds to value-added products is still in infancy. Significant research efforts are needed to understand if this concept will be viable for biorefinery applications in the long term. For example, one limitation in using LMEs in the deconstruction of lignin is represented by repolymerization due to phenoxy radical coupling instead of depolymerization, making their application still far.

On the other hand, enzymatic funneling is an attractive alternative to chemocatalytic methods as they offer high specificity and selectivity and do not require harsh conditions for reaction. However, in this case, some challenges must be overcome before application. In particular, enzymatic stability and activity of enzymes in their wild-type form are unsuitable for prolonged industrial processes. Therefore, protein engineering and strategies like whole-cell and immobilization are often required to allow their implementation. As shown in this chapter, the case of lignin-derived vanillin has been relatively widely explored by researchers in the last two decades and led to the identification and characterization of new classes of enzymes with high biotechnological potential. The last bioconversion studies demonstrated that implementing these enzymes for biorefinery is already at more than a proof-of-concept phase. However, experiments at a larger scale to assess the potential for real socioeconomic impact still need to be included in the literature.

This thesis was developed with the objective of the SMARTBOX European project (https://www.smartbox-project.eu) (Grant Agreement 837890), which aimed to provide bio-based industries access to inexpensive, robust, and efficient oxidative enzymes that can be readily engineered to meet industrial requirements. This project utilized advanced computational engineering platforms as alternatives to traditional directed evolution methods to accelerate and reduce biocatalyst development costs. ZYMVOL-Biomodeling SA (Barcelona, Spain) contributed by integrating quantum mechanics, machine learning, and molecular dynamics experiments into engineering. Another key feature of the project was the exclusive focus on products derived from the RCF method. Bert Sels' group at KU Leuven, recognized for their extensive work in lignin deconstruction (Renders *et al.* 2017, Van Aelst *et al.* 2020; Van den Bosch *et al*. 2015), was tasked with providing the lignin oil. Their RCF method involves simultaneous solvolysis and catalytic hydrogenolysis of softwood using a Ru-on-carbon catalyst in a methanol solvent, resulting in a lignin oil enriched with phenolic monomers, primarily 4PG and 4-*n*-propylsyringol (Van Aelst *et al.* 2020). Indeed, one of the main advantages of RCF is its tunability towards the formation of specific lignin monomers. By carefully choosing the feedstock, such as softwood, where lignin contains over 95% guaiacyl (G) units, achieving a high yield of the optimal substrate 4PG for enzymatic conversion into vanillin is possible. As outlined in section 3.2, the proposed pathway consisted of two reaction steps: (1) converting 4PG into isoeugenol and (2) the oxidative cleavage of the C=C bonds to produce vanillin as the final product. The optimization of the first step, catalyzed by a vanillyl alcohol oxidase (VAO), was carried out by Prof. Marco Fraaije's group at the University of Groningen and resulted in the construction of a high specific and thermostable VAO variant (PROGO) (Guo *et al.* 2022). The second step, focused on optimizing the oxidative cleavage of the C=C bond to produce vanillin from isoeugenol, was undertaken by Lígia Martins' group at ITQB-NOVA and developed during my thesis work.

In **Chapter 2**, we tested the production in *E. coli* of three different isoeugenol oxygenases and characterized their kinetic and stability parameters. We successfully performed an initial test of isoeugenol-to-vanillin bioconversion using whole-cells overproducing these enzymes. NOV1 from *Novosphingobium aromatocivorans* was selected as the most suitable enzyme for further protein engineering and bioconversion experiments.

Chapter 3 details a rational, computer-guided design of 35 variants targeting the active site of NOV1 to enhance its activity against the isoeugenol substrate. The variant S283F was identified and selected for its improved activity and stability. Biochemical and structural, in collaboration with Prof. Andrea Mattevi of the University of Pavia and molecular dynamics experiments in partnership with Zymvol, revealed iron cofactor retention as the key factor in the enzyme's stability. The S283F variant was also tested in coupled reaction experiments with the PROGO enzyme to demonstrate the potential of a new enzymatic route for valorizing the RCF-derived intermediate 4PG to vanillin. After optimization, a one-pot reaction using whole-cells was set up, yielding high amounts of vanillin from synthetically pure 4PG and natural lignin oil containing 4PG.

Chapter 4 uses a computational tool to predict distal mutations that could improve NOV1 activity on isoeugenol in collaboration with Zymvol. This work identified six best-performing variants with improved k_{cat} compared to the wild-type. DNA shuffling technique was used to recombine these mutations, resulting in the selection of the 12G2 (M324K_L372A) and 1D2 (I222V_L286W_L372A) mutants, which exhibited up to a 7-fold increase in

k_{cat} and 40-fold more excellent stability compared to the wild-type. Finally, a fast and user-friendly immobilization by affinity protocol was developed to perform bioconversions, achieving high vanillin yields and excellent recyclability using the 1D2 triple mutant.

Chapter 5 will move beyond the scope of lignin-to-vanillin conversion and isoeugenol oxygenase. Still, it addresses a related topic in terms of techniques and objectives: the immobilization of variants of a peroxygenase, an L-tyrosine hydroxylase from *Streptomyces sclerotialus*, an enzyme of interest for L-DOPA production and other biotechnological applications to validate a computational software designed by Zymvol to predict optimal surface sites.

1.6. REFERENCES

Abdelaziz, O. Y., S. Meier, J. Prothmann, C. Turner, A. Riisager, and C. P. Hulteberg. 2019. 'Oxidative Depolymerisation of Lignosulphonate Lignin into Low-Molecular-Weight Products with Cu-Mn/δ-AlO'. *Top. Catal.* 62: 639-48.

Alford, R. F., A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme, and J. J. Gray. 2017. 'The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design'. *J. Chem. Theory Comput.* 13: 3031-48.

Arzola, K. G., O. Polvillo, M. E. Arias, F. Perestelo, A. Carnicero, F. J. González-Vila, and M. A. Falcón. 2006. 'Early attack and subsequent changes produced in an industrial lignin by a fungal laccase and a laccase-mediator system: an analytical approach'. *Appl. Microbiol. Biotechnol.* 73: 141-50.

Babkova, P., E. Sebestova, J. Brezovsky, R. Chaloupkova, and J. Damborsky. 2017. 'Ancestral Haloalkane Dehalogenases Show Robustness and Unique Substrate Specificity'. *Chembiochem* 18: 1448-56.

Bajwa, D. S., G. Pourhashem, A. H. Ullah, and S. G. Bajwa. 2019. 'A concise review of current lignin production, applications, products and their environmental impact'. *Ind. Crops Prod.* 139: 111526.

Banerjee, G., and P. Chattopadhyay. 2019. 'Vanillin biotechnology: the perspectives and future'. *J. Sci. Food Agric.* 99: 499-506.

Becker, J., M. Kuhl, M. Kohlstedt, S. Starck, and C. Wittmann. 2018. 'Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin'. *Microb. Cell Fact.* 17.

Beckham, G. T., C. W. Johnson, E. M. Karp, D. Salvachúa, and D. R. Vardon. 2016. 'Opportunities and challenges in biological lignin valorization'. *Curr. Opin. Biotechnol.* 42: 40-53.

Bedbrook, C. N., K. K. Yang, A. J. Rice, V. Gradinaru, and F. H. Arnold. 2017. 'Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization'. *PLoS Comput. Biol.* 13.

Bedbrook, C. N., K. K. Yang, J. E. Robinson, E. D. Mackey, V. Gradinaru, and F. H. Arnold. 2019. 'Machine learning-guided channelrhodopsin engineering enables minimally invasive optogenetics'. *Nat. Methods* 16: 1176-+.

Borchert, A. J., W. R. Henson, and G. T. Beckham. 2022. 'Challenges and opportunities in biological funneling of heterogeneous and toxic substrates beyond lignin'. *Curr. Opin. Biotechnol.* 73: 1-13.

Brissos, V., P. T. Borges, F. Sancho, M. F. Lucas, C. Frazao, F. Conzuelo, and L. O. Martins. 2024. 'Flexible active-site loops fine-tune substrate specificity of hyperthermophilic metallo-oxidases'. *J. Biol. Inorg. Chem.* 29: 339-51.

Brissos, V., M. Renio, M. Lejmel, R. Estevinho, M. P. Robalo, M. R. Ventura, and L. O. Martins. 2024. 'Unlocking Lignin's Potential: Engineered Bacterial Laccases to Produce Biologically Active Molecules'. *Chemsuschem* e202401386.

Bugg, T. D. H. 2024. 'The chemical logic of enzymatic lignin degradation'. *Chem. Commun.* 60: 804-14.

Buller, A. R., P. van Roye, J. K. B. Cahn, R. A. Scheele, M. Herger, and F. H. Arnold. 2018. 'Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble'. *J. Am. Chem. Soc.* 140: 7256-66.

Cajnko, M. M., J. Oblak, M. Grilc, and B. Likozar. 2021. 'Enzymatic bioconversion process of lignin: mechanisms, reactions and kinetics'. *Bioresour. Technol.* 340.

Calzini, M. A., A. A. Malico, M. M. Mitchler, and G. J. Williams. 2021. 'Protein engineering for natural product biosynthesis and synthetic biology applications'. *Protein Eng. Des. Sel.* 34.

Cano-Díaz, G. S., A. Rosas-Aburto, E. Vivaldo-Lima, L. Flores-Santos, M. A. Vega-Hernández, M. G. Hernández-Luna, and A. Martinez. 2021. 'Determination of the Composition of Lignocellulosic Biomasses from Combined Analyses of Thermal, Spectroscopic, and Wet Chemical Methods'. *Ind. Eng. Chem. Res.* 60: 3502-15.

Chatterjee, A., D. M. DeLorenzo, R. Carr, and T. S. Moon. 2020. 'Bioconversion of renewable feedstocks by'. *Curr. Opin. Biotechnol.* 64: 10-16.

Chattopadhyay, P., G. Banerjee, and S. K. Sen. 2018. 'Cleaner production of vanillin through biotransformation of ferulic acid esters from agroresidue by'. *J. Clean. Prod.* 182: 272-79.

Chauhan, Prakram Singh, Ruchi Agrawal, Alok Satlewal, Ravindra Kumar, Ravi P. Gupta, and S. S. V. Ramakumar. 2022. 'Next generation applications of lignin derived commodity products, their life cycle, techno-economics and societal analysis'. *Int. J. Biol. Macromol.* 197: 179-200.

Chen, Q. H., D. T. Xie, S. Qiang, C. Y. Hu, and Y. H. Meng. 2022. 'Developing efficient vanillin biosynthesis system by regulating feruloyl-CoA synthetase and enoyl-CoA hydratase enzymes'. *Appl. Microbiol. Biotechnol.* 106: 247-59.

Chen, Z., and C. X. Wan. 2017. 'Biological valorization strategies for converting lignin into fuels and chemicals'. *Renew. Sust. Energy Rev.* 73: 610-21.

Chubukov, V., A. Mukhopadhyay, C. J. Petzold, J. D. Keasling, and H. G. Martín. 2016. 'Synthetic and systems biology for microbial production of commodity chemicals'. *NPJ Syst. Biol. Appl.* 2.

Clifton, B. E., J. H. Whitfield, I. Sanchez-Romero, M. K. Herde, C. Henneberger, H. Janovjak, and C. J. Jackson. 2017. 'Ancestral Protein Reconstruction and Circular Permutation for Improving the Stability and Dynamic Range of FRET Sensors'. *Methods Mol. Biol.* 1596: 71-87.

D'Arrigo, P., L. A. M. Rossato, A. Strini, and S. Serra. 2024. 'From Waste to Value: Recent Insights into Producing Vanillin from Lignin'. *Molecules* 29.

Datta, R., A. Kelkar, D. Baraniya, A. Molaei, A. Moulick, R. S. Meena, and P. Formanek. 2017. 'Enzymatic Degradation of Lignin in Soil: A Review'. *Sustainability* 9.

de Carvalho, C. C. C. R. 2017. 'Whole cell biocatalysts: essential workers from Nature to the industry'. *Microb. Biotechnol.* 10: 250-63.

DeMartini, J. D., S. Pattathil, J. S. Miller, H. J. Li, M. G. Hahn, and C. E. Wyman. 2013. 'Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass'. *Energy Environ. Sci.* 6: 898-909.

Diamond, A., S. Barnabé, and Desgagné-Penix. 2023. 'Is a spice missing from the recipe? The intra-cellular localization of vanillin biosynthesis needs further investigations'. *Plant Biol.* 25: 3-7.

Doherty, W. O. S., P. Mousavioun, and C. M. Fellows. 2011. 'Value-adding to cellulosic ethanol: Lignin polymers'. *Ind. Crops Prod.* 33: 259-76.

Erickson, E., A. Bleem, E. Kuatsjah, A. Z. Werner, J. L. DuBois, J. E. McGeehan, L. D. Eltis, and G. T. Beckham. 2022. 'Critical enzyme reactions in aromatic catabolism for microbial lignin conversion'. *Nat. Catal.* 5: 86-98.

Fache, M., B. Boutevin, and S. Caillol. 2016. 'Vanillin Production from Lignin and Its Use as a Renewable Chemical'. *ACS Sustainable Chem. Eng.* 4: 35-46.

Fujimaki, S., S. Sakamoto, S. Shimada, K. Kino, and T. Furuya. 2024. 'Engineering a coenzyme-independent dioxygenase for one-step production of vanillin from ferulic acid'. *Appl. Environ. Microbiol.* 90: e0023324.

Furuya, T., M. Kuroiwa, and K. Kino. 2017. 'Biotechnological production of vanillin using immobilized enzymes'. *J. Biotechnol.* 243: 25-28.

Furuya, T., M. Miura, and K. Kino. 2014. 'A Coenzyme-Independent Decarboxylase/Oxygenase Cascade for the Efficient Synthesis of Vanillin'. *ChemBioChem* 15: 2248-54.

Furuya, T., M. Miura, M. Kuroiwa, and K. Kino. 2015. 'High-yield production of vanillin from ferulic acid by a coenzyme-independent decarboxylase/oxygenase two-stage process'. *New Biotechnol.* 32: 335-39.

Galkin, M. V., and J. S. M. Samec. 2016. 'Lignin Valorization through Catalytic Lignocellulose Fractionation: A Fundamental Platform for the Future Biorefinery'. *ChemSusChem* 9: 1544-58.

Gallage, N. J., E. H. Hansen, R. Kannangara, C. E. Olsen, M. S. Motawia, K. Jorgensen, I. Holme, K. Hebelstrup, M. Grisoni, and B. L. Moller. 2014. 'Vanillin formation from ferulic acid in Vanilla planifolia is catalysed by a single enzyme'. *Nat. Commun.* 5.

Gillet, S., M. Aguedo, L. Petitjean, A. R. C. Morais, A. M. D. Lopes, R. M. Lukasik, and P. T. Anastas. 2017. 'Lignin transformations for high value applications: towards targeted modifications using green chemistry'. *Green Chem.* 19: 4200-33.

Gonzalez-Perez, D., and M. Alcalde. 2018. 'The making of versatile peroxidase by directed evolution'. *Biocatal. Biotransform.* 36: 1-11.

Gonzalez-Perez, D., I. Mateljak, E. Garcia-Ruiz, F. J. Ruiz-Dueñas, A. T. Martinez, and M. Alcalde. 2016. 'Alkaline versatile peroxidase by directed evolution'. *Catal. Sci. Technol.* 6: 6625-36.

Gumulya, Y., J. M. Baek, S. J. Wun, R. E. S. Thomson, K. L. Harris, D. J. B. Hunter, J. B. Y. H. Behrendorff, J. Kulig, S. Zheng, X. M. Wu, B. Wu, J. E. Stok, J. J. De Voss, G. Schenk, U. Jurva, S. Andersson, E. M. Isin, M. Bodén, L. Guddat, and E. M. J. Gillam. 2018. 'Engineering highly functional thermostable proteins using ancestral sequence reconstruction'. *Nat. Catal.* 1: 878-88.

Guo, Y. M., L. Alvigini, M. Trajkovic, L. Alonso-Cotchico, E. Monza, S. Savino, A. Mattevi, and M. W. Fraaije. 2022. 'Structure- and computational-aided engineering of an oxidase to produce isoeugenol from a lignin-derived compound'. *Nat. Commun.* 13.

Hahn-Hagerdal, B., M. Galbe, M. F. Gorwa-Grauslund, G. Lidén, and G. Zacchi. 2006. 'Bio-ethanol - the fuel of tomorrow from the residues of today'. *Trends Biotechnol.* 24: 549-56.

Han, Z. C., L. K. Long, and S. J. Ding. 2019. 'Expression and Characterization of Carotenoid Cleavage Oxygenases From Herbaspirillum seropedicae and Rhodobacteraceae bacterium Capable of Biotransforming Isoeugenol and 4-Vinylguaiacol to Vanillin'. *Front. Microbiol.* 10: 1-11.

Harshvardhan, K., M. Suri, A. Goswami, and T. Goswami. 2017. 'Biological approach for the production of vanillin from lignocellulosic biomass'. *J. Clean. Prod.* 149: 485-90.

Hautphenne, C., M. Penninckx, and F. Debaste. 2016. 'Product formation from phenolic compounds removal by laccases: A review'. *Environ. Technol. Innov.* 5: 250-66.

Herrmann, K. R., Brethauer, C., Siedhoff, N. E., Hofmann, I., Eyll, J., Davari, M. D., Schwaneberg, U., and Ruff, A. J. 2022. 'Evolution of E. coli Phytase Toward Improved Hydrolysis of Inositol Tetraphosphate', *Front. Chem. Eng.*, 4.

Hilgers, R., Vincken, J. P., Gruppen, H., and Kabel, M. A. 2018. 'Laccase/Mediator Systems: Their Reactivity toward Phenolic Lignin Structures', *ACS Sustainable Chem. Eng.*, 6: 2037-46.

Hocking, M. B. 1997. 'Vanillin: Synthetic flavoring from spent sulfite liquor', *J. Chem. Educ.*, 74: 1055-59.

Houtman, C. J., Maligaspe, E., Hunt, C. G., Fernández-Fueyo, E., Martínez, A. T., and Hammel, K. E. 2018. 'Fungal lignin peroxidase does not produce the veratryl alcohol cation radical as a diffusible ligninolytic oxidant', *J. Biol. Chem.*, 293: 4702-12.

Hu, Y., Yan, L., Zhao, X., Wang, C., Li, S., Zhang, X., Ma, L., and Zhang, Q. 2021. 'Mild selective oxidative cleavage of lignin C–C bonds over a copper catalyst in water', *Green Chem.*, 23: 7030-40.

Huang, L., Ma, J. Y., Sang, J. C., Wang, N., Wang, S., Wang, C., Kang, H. W., Liu, F. F., Lu, F. P., and Liu, Y. H. 2020. 'Enhancing the thermostability of phospholipase D from *Streptomyces halstedii* by directed evolution and elucidating the mechanism of a key amino acid residue using molecular dynamics simulation', *Int. J. Biol. Macromol.*, 164: 3065-74.

Jeon, W., Choi, I. H., Park, J. Y., Lee, J. S., and Hwang, K. R. 2020. 'Alkaline wet oxidation of lignin over Cu-Mn mixed oxide catalysts for production of vanillin', *Catal. Today*, 352: 95-103.

Ji, T., Liaqat, F., Khazi, M. I., Liaqat, N., Nawaz, M. Z., and Zhu, D. 2024. 'Lignin biotransformation: Advances in enzymatic valorization and bioproduction strategies', *Ind. Crops Prod.*, 216: 118759.

Jones, S. M., and Solomon, E. I. 2015. 'Electron transfer and reaction mechanism of laccases', *Cell. Mol. Life Sci.*, 72: 869-83.

Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D. L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D. S., and Fleishman, S. J. 2018. 'Automated Design of Efficient and Functionally Diverse Enzyme Repertoires', *Mol. Cell*, 72: 178-+.

Klammsteiner, T., Turan, V., Juárez, M. F. D., Oberegger, S., and Insam, H. 2020. 'Suitability of Black Soldier Fly Frass as Soil Amendment and Implication for Organic Waste Hygienization', *Agron. Basel*, 10.

Korányi, T. I., Fridrich, B., Pineda, A., and Barta, K. 2020. 'Development of "Lignin-First" Approaches for the Valorization of Lignocellulosic Biomass', *Molecules*, 25.

Laskar, D. D., Yang, B., Wang, H. M., and Lee, J. 2013. 'Pathways for biomass-derived lignin to hydrocarbon fuels', *Biofuels Bioprod. Biorefining*, 7: 602-26.

Li, M., and Wilkins, M. 2021. 'Lignin bioconversion into valuable products: fractionation, depolymerization, aromatic compound conversion, and bioproduct formation', *Syst. Microbiol. Biomanuf.*, 1: 166-85.

Liers, C., Pecyna, M. J., Kellner, H., Worrich, A., Zorn, H., Steffen, K. T., Hofrichter, M., and Ullrich, R. 2013. 'Substrate oxidation by dye-decolorizing peroxidases (DyPs) from wood- and litter-degrading agaricomycetes compared to other fungal and plant heme-peroxidases', *Appl. Microbiol. Biotechnol.*, 97: 5839-49.

Limayem, A., and Ricke, S. C. 2012. 'Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects', *Prog. Energy Combust. Sci.*, 38: 449-67.

Linde, D., Ayuso-Fernández, I., Laloux, M., Aguiar-Cervera, J. E., de Lacey, A. L., Ruiz-Dueñas, F. J., and Martínez, A. T. 2021. 'Comparing Ligninolytic Capabilities of Bacterial and Fungal Dye-Decolorizing Peroxidases and Class-II Peroxidase-Catalases', *Int. J. Mol. Sci.*, 22.

Linde, D., Ruiz-Dueñas, F. J., Fernández-Fueyo, E., Guallar, V., Hammel, K. E., Pogni, R., and Martínez, A. T. 2015. 'Basidiomycete DyPs: Genomic diversity, structural-functional aspects, reaction mechanism and environmental significance', *Arch. Biochem. Biophys.*, 574: 66-74.

Linger, J. G., Vardon, D. R., Guarnieri, M. T., Karp, E. M., Hunsinger, G. B., Franden, M. A., Johnson, C. W., Chupka, G., Strathmann, T. J., Pienkos, P. T., and Beckham, G. T. 2014. 'Lignin valorization through integrated biological funneling and chemical catalysis', *PNAS. USA*, 111: 12013-18.

Liu, H. M., Zou, Y., Yao, C. Y., and Yang, Z. 2020. 'Enzymatic synthesis of vanillin and related catalytic mechanism', *Flavour Fragr. J.*, 35: 51-58.

Liu, J. S., Zhang, S., Shi, Q. P., Wang, L., Kong, W., Yu, H. B., and Ma, F. Y. 2019. 'Highly efficient oxidation of synthetic and natural lignin-related compounds by *Physisporinus vitreus* versatile peroxidase', *Int. Biodeterior. Biodegrad.*, 136: 41-48.

Lončar, N., Colpa, D. I., and Fraaije, M. W. 2016. 'Exploring the biocatalytic potential of a DyP-type peroxidase by profiling the substrate acceptance of *Thermobifida fusca* DyP peroxidase', *Tetrahedron*, 72: 7276-81.

Lu, X. Y., Wu, X. M., Ma, B. D., and Xu, Y. 2021. 'Enhanced Thermostability of *Pseudomonas nitroreducens* Isoeugenol Monooxygenase by the Combinatorial Strategy of Surface Residue Replacement and Consensus Mutagenesis', *Catalysts*, 11.

Luo, H., Weeda, E. P., Alherech, M., Anson, C. W., Karlen, S. D., Cui, Y., Foster, C. E., and Stahl, S. S. 2021. 'Oxidative Catalytic Fractionation of Lignocellulosic Biomass under Non-alkaline Conditions', *J. Am. Chem. Soc.*, 143: 15462-70.

Maghraby, Y. R., El-Shabasy, R. M., Ibrahim, A. H., and Azzazy, H. M. E. 2023. 'Enzyme Immobilization Technologies and Industrial Applications', *ACS Omega*.

Marques, S. M., Planas-Iglesias, J., and Damborsky, J. 2021. 'Web-based tools for computational enzyme design', *Curr. Opin. Struct. Biol.*, 69: 19-34.

Mazari, A. M. A., Zhang, L. L., Ye, Z. W., Zhang, J., Tew, K. D., and Townsend, D. M. 2023. 'The Multifaceted Role of Glutathione S-Transferases in Health and Disease', *Biomolecules*, 13.

Mehra, R., Muschiol, J., Meyer, A. S., and Kepp, K. P. 2018. 'A structural-chemical explanation of fungal laccase activity', *Sci. Rep.*, 8.

Mei, C., Hu, C., Hu, Q., Sun, C., Li, L., Liang, X., Dong, Y., and Gu, X. 2020. 'Effective Depolymerization of Sodium Lignosulfonate over SO42−/TiO2 Catalyst', *Catalysts*.

Mongkhonsiri, G., Charoensuppanimit, P., Anantpinijwatna, A., Gani, R., and Assabumrungrat, S. 2020. 'Process development of sustainable biorefinery system integrated into the existing pulping process', *J. Clean. Prod.*, 255.

Mori, T., Tsuboi, Y., Ishida, N., Nishikubo, N., Demura, T., and Kikuchi, J. 2015. 'Multidimensional High-Resolution Magic Angle Spinning and Solution-State NMR Characterization of 13C-labeled Plant Metabolites and Lignocellulose', *Sci. Rep.*, 5.

Mottiar, Y., Vanholme, R., Boerjan, W., Ralph, J., and Mansfield, S. D. 2016. 'Designer lignins: harnessing the plasticity of lignification', *Curr. Opin. Biotechnol.*, 37: 190-200.

Musil, M., Jezik, A., Horackova, J., Borko, S., Kabourek, P., Damborsky, J., and Bednar, D. 2024. 'FireProt 2.0: web-based platform for the fully automated design of thermostable proteins', *Brief. Bioinform.*, 25.

Nguyen, Q. T., de Gonzalo, G., Binda, C., Rioz-Martinez, A., Mattevi, A., and Fraaije, M. W. 2016. 'Biocatalytic Properties and Structural Analysis of Eugenol Oxidase from Rhodococcus jostii RHA1: A Versatile Oxidative Biocatalyst', *ChemBioChem*, 17: 1359-66.

Ni, J., Tao, F., Du, H. Q., and Xu, P. 2015. 'Mimicking a natural pathway for de novo biosynthesis: natural vanillin production from accessible carbon sources', *Sci. Rep.*, 5.

Ni, J., Wu, Y. T., Tao, F., Peng, Y., and Xu, P. 2018. 'A Coenzyme-Free Biocatalyst for the Value-Added Utilization of Lignin-Derived Aromatics', *J. Am. Chem. Soc.*, 140: 16001-05.

Ono, Kanji, and Anton Erhard. 2011. 'Nondestructive Testing, 3. Ultrasonics'. In Ullmann's Encyclopedia of Industrial Chemistry, (Ed.).

Otten, R., Liu, L., Kenner, L. R., Clarkson, M. W., Mavor, D., Tawfik, D. S., Kern, D., and Fraser, J. S. 2018. 'Rescue of conformational dynamics in enzyme catalysis by directed evolution', *Nat. Commun.*, 9.

Overhage, J., Priefert, H., Rabenhorst, J., and Steinbuchel, A. 1999. 'Biotransformation of eugenol to vanillin by a mutant of Pseudomonas sp. strain HR199 constructed by disruption of the vanillin dehydrogenase (vdh) gene', *Appl. Microbiol. Biotechnol.*, 52: 820-8.

Overhage, J., Priefert, H., and Steinbüchel, A. 1999. 'Biochemical and genetic analyses of ferulic acid catabolism in Pseudomonas sp. strain HR199', *Appl. Environ. Microbiol.*, 65: 4837-47.

Pacek, A. W., Ding, P., Garrett, M., Sheldrake, G., and Nienow, A. W. 2013. 'Catalytic conversion of sodium lignosulfonate to vanillin: engineering aspects. Part 1. Effects of processing conditions on vanillin yield and selectivity', *Ind. Eng. Chem. Res.*, 52: 8361-72.

Pandey, M. P., and Kim, C. S. 2011. 'Lignin depolymerization and conversion: a review of thermochemical methods', *Chem. Eng. Technol.*, 34: 29-41.

Parsell, T., Yohe, S., Degenstein, J., Jarrell, T., Klein, I., Gencer, E., Hewetson, B., Hurt, M., Kim, J. I., Choudhari, H., Saha, B., Meilan, R., Mosier, N., Ribeiro, F., Delgass, W. N., Chapple, C., Kenttämaa, H. I., Agrawal, R., and Abu-Omar, M. M. 2015. 'A synergistic biorefinery based on catalytic conversion of lignin prior to cellulose starting from lignocellulosic biomass', *Green Chem.*, 17: 1492-99.

Picart, P., de Maria, P. D., and Schallmey, A. 2015. 'From gene to biorefinery: microbial β-etherases as promising biocatalysts for lignin valorization', *Front. Microbiol.*, 6.

Picart, P., Liu, H. F., Grande, P. M., Anders, N., Zhu, L. L., Klankermayer, J., Leitner, W., de María, P. D., Schwaneberg, U., and Schallmey, A. 2017. 'Multi-step biocatalytic depolymerization of lignin', *Appl. Microbiol. Biotechnol.*, 101: 6277-87.

Pogni, R., Baratto, M. C., Sinicropi, A., and Basosi, R. 2015. 'Spectroscopic and computational characterization of laccases and their substrate radical intermediates', *Cell. Mol. Life Sci.*, 72: 885-96.

Purwani, N. N., Rozeboom, H. J., Willers, V. P., Wijma, H. J., and Fraaije, M. W. 2024. 'Discovery of a new class of bacterial heme-containing C=C cleaving oxygenases', *New Biotechnol.*, 83: 82-90.

Qin, X., Sun, X. H., Huang, H. Q., Bai, Y. G., Wang, Y., Luo, H. Y., Yao, B., Zhang, X. Y., and Su, X. Y. 2017. 'Oxidation of a non-phenolic lignin model compound by two Irpex lacteus manganese peroxidases: evidence for implication of carboxylate and radicals', *Biotechnol. Biofuels*, 10.

Ragauskas, A. J., Beckham, G. T., Biddy, M. J., Chandra, R., Chen, F., Davis, M. F., Davison, B. H., Dixon, R. A., Gilna, P., Keller, M., Langan, P., Naskar, A. K., Saddler, J. N., Tschaplinski, T. J., Tuskan, G. A., and Wyman, C. E. 2014. 'Lignin valorization: improving lignin processing in the biorefinery', *Science*, 344: 709-+.

Renders, T., Van den Bosch, S., Koelewijn, S. F., Schutyser, W., and Sels, B. F. 2017. 'Lignin-first biomass fractionation: the advent of active stabilisation strategies', *Energy Environ. Sci.*, 10: 1551-57.

Rodrigues Pinto, P. C., Borges da Silva, E. A., and Rodrigues, A. E. 2012. 'Lignin as source of fine chemicals: vanillin and syringaldehyde.' in Baskar, C., Baskar, S., and Dhillon, R. S. (eds.), *Biomass Conversion: The Interface of Biotechnology, Chemistry and Materials Science* (Springer: Berlin, Heidelberg).

Rosini, E., Molinari, F., Miani, D., and Pollegioni, L. 2023. "Lignin valorization: production of high value-added compounds by engineered microorganisms." In *Catalysts*.

Ryu, J. Y., Seo, J., Park, S., Ahn, J. H., Chong, Y., Sadowsky, M. J., and Hur, H. G. 2013. 'Characterization of an isoeugenol monooxygenase (Iem) from Pseudomonas nitroreducens Jin1 that transforms isoeugenol to vanillin', *Biosci. Biotechnol. Biochem.*, 77: 289-94.

Sanchez-Gonzalez, E., Lopez-Olvera, A., Monroy, O., Aguilar-Pliego, J., Flores, J. G., Islas-Jacome, A., Rincon-Guevara, M. A., Gonzalez-Zamora, E., Rodriguez-Molina, B., and Ibarra, I. A. 2017. 'Synthesis of vanillin via a catalytically active Cu(ii)-metal organic polyhedron', *Crystengcomm*, 19: 4142-46.

Sankaran, R., Markandan, K., Khoo, K. S., Cheng, C. K., Ashokkumar, V., Deepanraj, B., and Show, P. L. 2021. 'The expansion of lignocellulose biomass conversion into bioenergy via nanobiotechnology', *Front. Nanotechnol.*, 3.

Schutyser, W., Renders, T., Van den Bosch, S., Koelewijn, S. F., Beckham, G. T., and Sels, B. F. 2018. 'Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading', *Chem. Soc. Rev.*, 47: 852-908.

Silva, D., Rodrigues, C. F., Lorena, C., Borges, P. T., and Martins, L. O. 2023. 'Biocatalysis for biorefineries: the case of dye-decolorizing peroxidases', *Biotechnol. Adv.*, 65.

Silva, D., Sousa, A. C., Robalo, M. P., and Martins, L. O. 2023. 'A wide array of lignin-related phenolics are oxidized by an evolved bacterial dye-decolourising peroxidase', *New Biotechnol.*, 77: 176-84.

Singh, A. K., Bilal, M., Iqbal, H. M. N., and Raj, A. 2021. 'Lignin peroxidase in focus for catalytic elimination of contaminants─a critical review on recent progress and perspectives', *Int. J. Biol. Macromol.*, 177: 58-82.

Sugano, Y., and Yoshida, T. 2021. 'DyP-type peroxidases: recent advances and perspectives', *Int. J. Mol. Sci.*, 22.

Sumbalova, L., Stourac, J., Martinek, T., Bednar, D., and Damborsky, J. 2018. 'HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information', *Nucleic Acids Res.*, 46: W356-W62.

Sun, Z. H., Fridrich, B., de Santi, A., Elangovan, S., and Barta, K. 2018. 'Bright side of lignin depolymerization: toward new platform chemicals', *Chem. Rev.*, 118: 614-78.

Tarabanko, V. E., Chelbina, Y. V., Kudryashev, A. V., and Tarabanko, N. V. 2013. 'Separation of vanillin and syringaldehyde produced from lignins', *Sep. Sci. Technol.*, 48: 127-32.

Tarabanko, V. E., Kaygorodov, K. L., Skiba, E. A., Tarabanko, N., Chelbina, Y. V., Baybakova, O. V., Kuznetsov, B. N., and Djakovitch, L. 2017. 'Processing pine wood into vanillin and glucose by sequential catalytic oxidation and enzymatic hydrolysis', *J. Wood Chem. Technol.*, 37: 43-51.

Tomadoni, B., Viacava, G., Cassani, L., Moreira, M. R., and Ponce, A. 2016. 'Novel biopreservatives to enhance the safety and quality of strawberry juice', *J. Food Sci. Technol.*, 53: 281-92.

Tomizawa, S., Chuah, J. A., Matsumoto, K., Doi, Y., and Numata, K. 2014. 'Understanding the limitations in the biosynthesis of polyhydroxyalkanoate (PHA) from lignin derivatives', *ACS Sustainable Chem. Eng.*, 2: 1106-13.

Turan, V., Schröder, P., Bilen, S., Insam, H., and Juárez, M. F. D. 2019. 'Co-inoculation effect of Rhizobium and Achillea millefolium L. oil extracts on growth of common bean (Phaseolus vulgaris) and soil microbial-chemical properties', *Sci. Rep.*, 9.

Upton, B. M., and Kasko, A. M. 2016. 'Strategies for the conversion of lignin to high-value polymeric materials: review and perspective', *Chem. Rev.*, 116: 2275-306.

Van Aelst, K., Van Sinay, E., Vangeel, T., Cooreman, E., Van den Bossche, G., Renders, T., Van Aelst, J., Van den Bosch, S., and Sels, B. F. 2020. 'Reductive catalytic fractionation of pine wood: elucidating and quantifying the molecular structures in the lignin oil', *Chem. Sci.*, 11: 11498-508.

Van den Bosch, S., Schutyser, W., Vanholme, R., Driessen, T., Koelewijn, S. F., Renders, T., De Meester, B., Huijgen, W. J. J., Dehaen, W., Courtin, C. M., Lagrain, B., Boerjan, W., and Sels, B. F. 2015. 'Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps', *Energy Environ. Sci.*, 8: 1748-63.

Vanholme, R., Morreel, K., Darrah, C., Oyarce, P., Grabber, J. H., Ralph, J., and Boerjan, W. 2012. 'Metabolic engineering of novel lignin in biomass crops', *New Phytol.*, 196: 978-1000.

Walch, F., Abdelaziz, O. Y., Meier, S., Bjelić, S., Hulteberg, C. P., and Riisager, A. 2021. 'Oxidative depolymerization of Kraft lignin to high-value aromatics using a homogeneous vanadium–copper catalyst', *Catal. Sci. Technol.*, 11: 1843-53.

Walton, N. J., Mayer, M. J., and Narbad, A. 2003. 'Molecules of interest - vanillin', *Phytochemistry*, 63: 505-15.

Wang, Q., Wu, X. M., Lu, X. Y., He, Y. C., Ma, B. D., and Xu, Y. 2021. 'Efficient biosynthesis of vanillin from isoeugenol by recombinant isoeugenol monooxygenase from Pseudomonas nitroreducens Jin1', *Appl. Biochem. Biotechnol.*, 193: 1116-28.

Weng, C. H., Peng, X. W., and Han, Y. J. 2021. 'Depolymerization and conversion of lignin to value-added bioproducts by microbial and enzymatic catalysis', *Biotechnol. Biofuels*, 14.

Wijma, H. J., Floor, R. J., Jekel, P. A., Baker, D., Marrink, S. J., and Janssen, D. B. 2014. 'Computationally designed libraries for rapid enzyme stabilization', *Protein Eng. Des. Sel.*, 27: 49-58.

Wu, X. Y., De Bruyn, M., and Barta, K. 2023. 'Deriving high value products from depolymerized lignin oil, aided by (bio)catalytic funneling strategies', *Chem. Commun.*, 59: 9929-51.

Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., and Arnold, F. H. 2019. 'Machine learning-assisted directed protein evolution with combinatorial libraries', *PNAS. U.S.A.*, 116: 8852-58.

Wyman, C., Dale, B., Balan, V., Elander, R. T., Holtzapple, M., Ramirez, R. S., Ladisch, M., Mosier, N., Lee, Y. Y., Gupta, R., St. Thomas, B., Hames, B., Warner, R., and Kumar, R. 2013. *Aqueous pretreatment of plant biomass for biological and chemical conversion to fuels and chemicals*.

Xu, L. X., Sun, J. Z., Qaria, M. A., Gao, L., and Zhu, D. C. 2021. 'Dye decoloring peroxidase structure, catalytic properties and applications: current advancement and futurity', *Catalysts*, 11.

Xu, X. W., Li, P. H., Zhong, Y. D., Yu, J. D., Miao, C., and Tong, G. L. 2023. 'Review on the oxidative catalysis methods of converting lignin into vanillin', *Int. J. Biol. Macromol.*, 243.

Xu, Z. X., Lei, P., Zhai, R., Wen, Z. Q., and Jin, M. J. 2019. 'Recent advances in lignin valorization with bacterial cultures: microorganisms, metabolic pathways, and bio-products', *Biotechnol. Biofuels*, 12.

Yamada, M., Okada, Y., Yoshida, T., and Nagasawa, T. 2007. 'Purification, characterization and gene cloning of isoeugenol-degrading enzyme from Pseudomonas putida IE27', *Arch. Microbiol.*, 187: 511-17.

Yang, G., Miton, C. M., and Tokuriki, N. 2020. 'A mechanistic view of enzyme evolution', *Protein Sci.*, 29: 1724-47.

Yang, H. L., Barros-Rios, J., Kourteva, G., Rao, X. L., Chen, F., Shen, H., Liu, C. G., Podstolski, A., Belanger, F., Havkin-Frenkel, D., and Dixon, R. A. 2017. 'A re-evaluation of the final step of vanillin biosynthesis in the orchid', *Phytochemistry*, 139: 33-46.

Yang, K. K., Wu, Z., and Arnold, F. H. 2019. 'Machine-learning-guided directed evolution for protein engineering', *Nat. Methods*, 16: 687-94.

Yepez, R., García, S., Schachat, P., Sánchez-Sánchez, M., González-Estefan, J. H., González-Zamora, E., Ibarra, I. A., and Aguilar-Pliego, J. 2015. 'Catalytic activity of HKUST-1 in the oxidation of trans-ferulic acid to vanillin', *New J. Chem.*, 39: 5112-15.

Zabkova, M., da Silva, E. A. B., and Rodrigues, A. E. 2007. 'Recovery of vanillin from Kraft lignin oxidation by ion-exchange with neutralization', *Sep. Purif. Technol.*, 55: 56-68.

Zakzeski, J., Bruijnincx, P. C. A., Jongerius, A. L., and Weckhuysen, B. M. 2010. 'The catalytic valorization of lignin for the production of renewable chemicals', *Chem. Rev.*, 110: 3552-99.

Zeng, J. J., Mills, M. J. L., Simmons, B. A., Kent, M. S., and Sale, K. L. 2017. 'Understanding factors controlling depolymerization and polymerization in catalytic degradation of β-ether linked model lignin compounds by versatile peroxidase', *Green Chem.*, 19: 2145-54.

Zeymer, C., and Hilvert, D. 2018. 'Directed evolution of protein catalysts', *Annu. Rev. Biochem.*, 87: 131-57.

Zhang, Zhi-Gang, Jiang, Shi-Kai, Shen, Xi, Lin, Jia-Chun, Yi, Yan, and Ji, Xiao-Jun. 2025. 'Biocatalytic synthesis of vanillin from biomass-derived compounds: a review', *Catal. Today*, 445: 115077.

Zhao, L. Q., Xie, Y. M., Chen, L. Y., Xu, X. F., Zha, C. X., and Cheng, F. 2018. 'Efficient biotransformation of isoeugenol to vanillin in recombinant strains of Escherichia coli by using engineered isoeugenol monooxygenase and sol-gel chitosan membrane', *Process Biochem.*, 71: 76-81.

Zhu, X. C., Wu, J. Y., Li, S. Z., Xiang, L., Jin, J. M., Liang, C. N., and Tang, S. Y. 2024. 'Artificial biosynthetic pathway for efficient synthesis of vanillin, a feruloyl-CoA-derived natural product from eugenol', *J. Agric. Food Chem.*, 72: 6463-70.

Chapter 2

Identification of a Suitable Oxygenase for Vanillin Production from Isoeugenol

Author contributions:

The author of this thesis conducted production, purification and biochemical characterization of *Pp*IEM, *Pn*IEM and NOV1. Furthermore, he performed whole-cell reactions and characterized NOV1 substrate scope. NMR analysis of the product of coniferyl alcohol conversion was performed under the guidance of Cristiano Conceição at Lab of Dr. Rita Ventura’s Lab in ITQB. Purification attempts of *Pn*IEM from the insoluble fraction was performed by Francisco Nunes during his Master’s thesis work. Attempts of *Pn*IEM crystallization were carried out by Laura Alvigini at the University of Pavia, under the supervision of Prof. Andrea Mattevi.

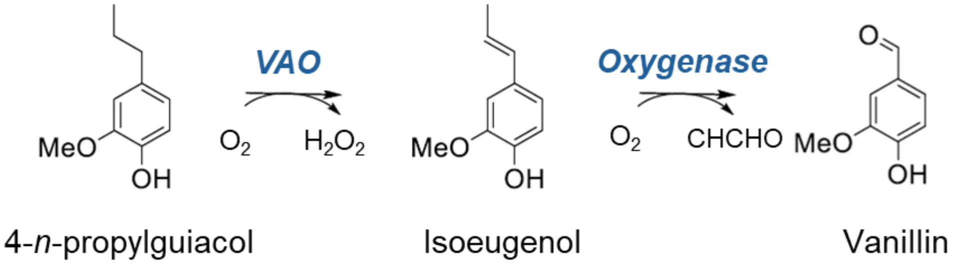
2.1. ABSTRACT

This chapter describes the work to identify the most suitable isoeugenol oxygenase for converting lignin-derived isoeugenol to vanillin. Initially, two bacterial isoeugenol oxygenases (IEMs) were considered: *Pp*IEM from *Pseudomonas putida* and *Pn*IEM from *Pseudomonas nitroreducens*. The genes were synthesized and expressed, producing protein in *E. coli*. Due to the lack of detectable activity of *Pp*IEM in *E. coli* cell extracts' supernatants, *Pn*IEM was purified and chosen for further biochemical characterization. This enzyme has a K_m of 0.25 ± 0.03 mM and a turnover number (k_{cat}) of 5.1 ± 0.2 s^{-1} for isoeugenol. However, its tendency to aggregate resulted in very low production soluble yields. Attempts to improve its solubility were unsuccessful, and efforts to determine its crystal structure compromised the success of the following protein engineering experiments. These reasons led us to search for an alternative enzyme, and NOV1 from *Novosphingobium aromaticivorans* was identified as a suitable candidate. NOV1 displayed high solubility, with ten times higher production levels than the *Pn*IEM. NOV1 revealed slightly higher catalytic efficiency than *Pn*IEM and a similar half-life at 25 °C ~2 h, but with a higher melting temperature (T_m) at 58 °C than *Pn*IEM's T_m of 43 °C. Based on these results and the availability of a crystal structure in the database, NOV1 was selected for future protein engineering experiments.

Keywords: Protein expression, Protein stability, Biocatalysis, Whole-cell reactions

2.2. INTRODUCTION

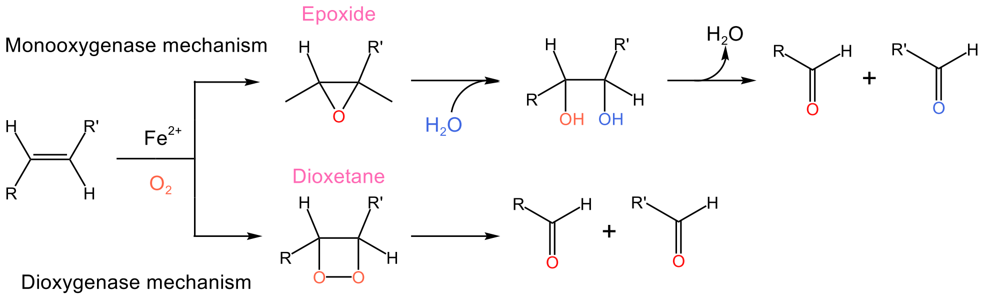
The conversion of 4-*n*-Propylguaiacol (4PG) obtained from reductive catalytic fractionation (RCF) of lignin to more valuable compounds is crucial for the economic sustainability of lignin biorefineries (Sun *et al.* 2018; Van den Bosch *et al.* 2015). In the SMARTBOX project, the proposed pathway to convert 4PG to vanillin involved first the dehydrogenation of the alkyl moiety to produce isoeugenol, which can then be converted to vanillin through oxygenation at the double bond (**Scheme 2.1**).



Scheme 2.1. Oxidation of 4-*n*-propyl guaiacol (4PG) into vanillin. 4PG is first converted to isoeugenol by a VAO-type oxidase; an isoeugenol oxygenase cleaves the double bond, producing vanillin.

The most promising candidates for catalyzing the second step were isoeugenol oxygenases (1.13.11.88) of the carotenoid cleavage oxygenases (CCOs) superfamily, as they are reported to cleave C=C bonds in one step without the use of expensive cofactors (Daruwalla and Kiser 2020). CCOs constitute a superfamily of mononuclear non-heme iron-dependent enzymes capable of oxidatively cleaving the C=C bonds in carotenoid, apocarotenoid, stilbenoid, and phenylpropanoid compounds, forming aldehydes or ketones as products (Daruwalla and Kiser 2020; Sui *et al.* 2017). CCOs are present in all life domains, and their activity is critical for different biological processes such as hormone signaling, light sensation, and production of aesthetically pleasing compounds in plants, including those responsible for the smell of roses (Huang *et al.* 2009) and the aroma of saffron and vanillin (Frusciante *et al.* 2014). As shown in section 1.3.2, *Chapter 1*, alkene-cleaving oxygenases have raised interest in bioindustry applications in the last few years. These enzymes are expressed in soil-based fungi and bacteria to detoxify plant antimicrobials and degrade biomass. Indeed, their substrates are stilbene and phenylpropanoids, which are secondary metabolites synthesized by plants generally as a defense against abiotic and biotic agents exploiting their antimicrobial and antioxidant properties or as lignin building blocks (Ortiz and Sansinenea 2023). The first reports of these enzymes go back to the early 1990s when researchers described the first identified enzymes capable of cleaving the interphenyl α,β double bond of trans-stilbenes from bacteria of the soil (Kamoda and Saburi 1993, 1995). In the 2000s-2010s, the first isoeugenol oxygenases were identified from bacterial genomes, including *Pp*IEM from *Pseudomonas putida* (Yamada *et al.* 2007) and *Pn*IEM from *Pseudomonas nitroreducens* (Ryu *et al.* 2010). These two enzymes share 81.4% identity and showed high vanillin yields from isoeugenol by whole-cell conversions (Ryu *et al.* 2013; Yamada *et al.* 2008). It is controversial whether these enzymes' incorporation of two oxygen atoms during catalysis occurs via a mono- or dioxygenase mechanism (**Scheme 2.2**). In the first case, one oxygen atom is first incorporated from O_2, forming an epoxide intermediate, and then a second atom is added from the water solvent. In the second case, both the oxygen atoms come from O_2, generating a dioxetane intermediate. Both mechanisms have been suggested for this enzyme family (Holger *et al.* 2006; Marasco and Schmidt-Dannert 2008; Ryu *et al.* 2013; Yamada *et al.* 2007). Subsequent quantum mechanics/molecular mechanics studies and oxygen-labeling experiments indicated that the dioxygenase mechanism was the most plausible for some alkene-cleaving oxygenases (Bai *et al.* 2019; Lu and Lai 2019; Sui *et al.* 2017). However, discrepancies between these studies might also indicate that both mono- and dioxygenase mechanisms could exist within the CCO family.

Due to the unsatisfactory production of *Pn*IEM in the soluble fraction and difficulties crystallizing the protein, we sought an alternative enzyme. We identified NOV1, a stilbene cleavage dioxygenase (SCD) from *Novosphingobium aromaticivorans* (Marasco and Schmidt-Dannert 2008), which shares 41% sequence identity with *Pn*IEM. Although NOV1 was described in the literature as resveratrol cleavage dioxygenase, it was reported to have activity on isoeugenol (McAndrew *et al.* 2016). However, no characterization was performed at the time to compare with *Pn*IEM. Importantly, NOV1 was described as a very soluble protein, and its structure was already available in the database, facilitating future protein engineering experiments.



Scheme 2.2. Putative mono- and dioxygenase mechanisms of carotenoid cleavage oxygenases (CCOs).

This chapter describes selecting the most suitable isoeugenol oxygenase for converting isoeugenol into vanillin. Towards this goal, *Pp*IEM, *Pn*IEM, and later NOV1 were heterologously produced *E. coli* and characterized. The three proteins were evaluated based on their production yields, kinetic and stability parameters, tendency to aggregate, and the availability of their crystal structures. Finally, the most promising catalyst was tested in the bioconversion of isoeugenol to vanillin by whole-cell reaction.

2.3. MATERIALS AND METHODS

2.3.1. Production of *Pp*IEM, *Pn*IEM and NOV1

Genes coding for both *Pp*IEM and *Pn*IEM wild-type, Fireprot and PROSS designed enzymes were synthesized, with codons optimized for *E. coli*, and cloned into pET-15b under a T7 promoter by GenScript. The gene coding for NOV1 oxygenase was provided by the University of Pavia and cloned into pET28b with a SUMO-tag sequence. *Pp*IEM, *Pn*IEM, and NOV1 production tested *in E. coli* host strains: Tuner, Bl21, Bl21Star, and Rosetta. Recombinant *E. coli* strains containing plasmids for IEMs and NOV1 were grown in 0.1-1 L of LB medium, supplemented with either 100 µg/mL ampicillin or 50 µg/mL kanamycin, respectively. Cultures were incubated in a 0.1-5 L Erlenmeyer flask with shaking at 100 rpm (Innova 44, New Brunswick Scientific) at 37 °C until reaching an optical density of 0.6 at 600 nm. At this point, 0.1 mM IPTG and 0.5 mM $FeSO_4$ were added, and the temperature was reduced to 25 °C. Incubation continued for an additional 16 h before cells were harvested by centrifugation.

*Pn*IEM production was also tested with the co-expression of two genes coding for the chaperones, GroES and GroEL. *E. coli* Tuner cells were transformed with the pGro7 plasmid, containing the GroES and GroEL genes. These chaperones enhance the enzyme folding process during recombinant enzyme expression. Growth was conducted as previously described, and the co-expression of chaperons was performed by adding 0.5 mg/mL of L-Arabinose and 0.1 mM IPTG to induce *PnI*EM expression.

2.3.2. *Pn*IEM Purification from the Soluble Fraction

Cell pellets were suspended in 20 mM sodium phosphate buffer, 0.5 M NaCl, and 20 mM imidazole, pH 7.6 (buffer A), containing DNaseI (2 µg/mL), 5 mM $MgCl_2$, and a protease inhibitor mixture (Complete^{TM} mini-EDTA-free protease inhibitor mixture tablets; Roche, Basel, Switzerland). Cells were disrupted by three cycles at French press (900 psi), followed by centrifugation (18,000g, 1 h, 4 °C). The resulting soluble extracts were loaded onto an affinity chromatography HisTrap^{TM} column (GE Healthcare) equilibrated with buffer A. Elution was performed with a linear imidazole gradient (0 to 0.5 M in 20 min) in the same buffer. All purification steps were carried out at room temperature using an ÅKTA purifier (GE Healthcare). The purity of eluted fractions was assessed by SDS-PAGE analysis. NOV1 active fractions were digested with SUMO-protease (1 unit per 5 µg of protein) overnight at 4 °C, then concentrated by ultrafiltration (30 KDa cut-off) and equilibrated to 20 mM Tris-HCl and 0.2 M NaCl, pH 7.6.

2.3.3. *Pn*IEM Purification from Inclusion Bodies

*Pn*IEM inclusion bodies were isolated by resuspending the cell pellets from a 50 mL cell culture in 10 mL of 20 mM Tris-HCl buffer, pH 7.6, supplemented with 0.5% Triton X-100. The mixture was incubated at room temperature for 5 min, then centrifuged for 5 min at 18,000 rpm, 4 °C. The cell debris was recovered and resuspended in 9 mL of 20 mM Tris-HCl buffer, pH 7.6, supplemented with 8 M urea, 2 mM reduced glutathione, and 0.2 mM oxidized glutathione. After incubating for 1 h at room temperature, the mixture was centrifuged for 5 min at 18,000 rpm at 4 °C. The supernatant, containing the urea-solubilized protein, was recovered for the renaturation step. Two different renaturation solutions were tested. The first solution consisted of 20 mM Tris-HCl buffer pH 7.6 supplemented with 2 mM reduced Glutathione, 0.2 mM oxidized Glutathione, and 100 μM $FeSO_4$. The second solution was identical to the first but supplemented with 5 mM ascorbic acid. Forty milliliters of either solution were slowly added to the fractions. The second refolding solution was added under anaerobic conditions using an anaerobic chamber. These mixtures were incubated for 2 h at room temperature and then dialyzed using an Amicon® centrifugal filter to remove urea traces and concentrate the refolded protein.

2.3.4. *In silico* Stability Design

The variant structure model of *Pn*IEM wild-type was obtained by submitting the protein sequence in Phyre2 web portal (www.sbg.bio.ic.ac.uk/phyre2). The structure was modeled using four crystal structures available in database as templates (6ojr, 5vd2d, 5j54, 5u90), which were selected based on percentage identity and alignment coverage. The resulting model was submitted to PROSS (https://pross.weizmann.ac.il/) and FireProt1.2 (https://loschmidt.chemi.muni.cz/FireProtweb/) servers to improve protein stability and solubility. Submissions were performed in standard settings. Mutations in positions close at < 4Å from Fe(II)-center were excluded from the final designs.

2.3.5. Biochemical Characterization

Enzymatic activities of *Pn*IEM, *Pp*IEM, and NOV1 were monitored using a *Synergy2* microplate reader (BioTek, Vermont, USA) at 25 °C, following vanillin absorbance peak at 340 nm ($\varepsilon_{340} = 15.97\ M^{-1}\ cm^{-1}$). The effect of pH on activity was studied in 100 mM Britton-Robinson buffer (pH 3−11). The optimal reaction temperatures were determined in the 10−50 °C range. In this case, the reaction mixtures were incubated in 1 mL glass cuvettes in a UVvisible spectrophotometer (Shimadzu UV-1603) equipped with a Peltier module system. The steady-state kinetic analysis was performed with isoeugenol (0.01−4 mM) in 100 mM Tris-HCl, pH 9. Kinetic data were fitted directly to the Michaelis−Menten equation using Origin software. All enzymatic assays were performed at least in triplicate.

2.3.6. Thermostability Analysis

Thermal inactivation was studied by incubating enzyme preparations in 20 mM Tris-HCl, pH 7.6, at 25 °C. At appropriate time points, aliquots were withdrawn and examined for activity. Inactivation constants (k_{in}) were obtained by linear regression of logarithm activity versus time. Half-life ($t_{1/2}$) was calculated using the equation $t_{1/2} = \ln 2/k_{in}$.

Protein unfolding was monitored using a Cary Eclipse spectrofluorometer with an excitation wavelength of 296 nm and by recording the fluorescence emission of tryptophyl residues at 340 nm. Enzyme preparations in 20 mM Tris-HCl and 0.2 M NaCl, pH 7.6, were placed onto a thermostatically thermal block and heated at 1 °C/min to 100 °C. Thermal unfolding was analyzed using an $N \rightleftharpoons U$ two-state model (Eftink 1998). Static light scattering using the spectrofluorometer's excitation and emission wavelengths of 500 nm followed enzyme aggregation and determined the onset of aggregation temperature (T_{agg}).

2.3.7. NMR analysis of coniferyl alcohol conversion

The product of coniferyl alcohol conversion was studied in a 50 mL Erlenmeyer flask with a reaction volume of 10 mL containing 20 mM coniferyl alcohol in 0.1 M TrisHCl pH 9 at room temperature and 150 rpm. The reaction started after adding 0.1 U/mL of pure NOV1 (1 U is defined as the amount of enzyme that produced 1 μmol min^{-1} of product). After 24 h the reaction mixture was lyophilized, and the resulting powder was resuspended in 0.7 mL of deuterated chloroform ($CDCl_3$), transferred into a clean tube and analyzed by NMR. Spectra of ^1H (400 MHz) were collected on Bruker AVANCE IT+ 400 MHz for reaction samples and samples of vanillin and coniferyl alcohol as standards. Chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane (0 ppm). Coupling constants are reported in Hertz (Hz). Multiplicity is reported with the usual abbreviations (s: singlet, d: doublet, dq: doublet of quartets, dd: doublet of doublets, m: multiplet).

2.3.8. Whole-cell Reaction

Whole-cell reactions were carried in 100 mL Erlenmeyer flasks holding 20 mL of reaction buffer (100 mM Tris-HCl, pH 9) and 2 OD/mL *E. coli* cells that had overproduced NOV1 dioxygenase. The addition of 30 mM isoeugenol started the reactions. Time-course bioconversions were carried out at room temperature and 150 rpm for 24 h.

2.3.9. Other Methods

Protein concentrations were estimated using the Bradford colorimetric assay. The concentration of purified protein preparations was assessed using the molar absorption coefficient of *Pn*IEM (ε_{280} = 85830 M^{-1} cm^{-1}) and NOV1 (ε_{280} = 81930 M^{-1} cm^{-1}) calculated from the protein sequences using the ExPASy Bioinformatics Resource Portal (http://web.expasy.org). The iron content of purified protein samples was determined by a colorimetric assay resorting to the chelator TPTZ (2,4,6-tris(2-pyridyl)-s-triazine) (Fischer and Price 1964) and by atomic emission spectroscopy at Laboratório Central de Análises, Universidade de Aveiro (Aveiro, Portugal). Conversion of phenolics used to explore NOV1 substrate scope was assessed using the molar absorption coefficients in the literature (Silva *et al.* 2023)

2.4. RESULTS AND DISCUSSION

2.4.1. Production of *Pp*IEM and *Pn*IEM

The two isoeugenol oxygenases, *Pp*IEM from *P. putida* and *Pn*IEM from *P. nitroreducens*, were identified from databases and literature (Yamada *et al.* 2007; Ryu *et al.* 2010). Their application was attractive as they do not require external cofactors, and their catalysis in whole-cell systems was already demonstrated, achieving vanillin yields above 70% (Yamada *et al.* 2008; Ryu *et al.* 2010).

Their gene expression was tested in Tuner and BL21 *E. coli* strains. The overproduction levels of recombinant enzymes were compared by SDS-PAGE analysis (**Fig. 2.1**). Protein production was higher in the Tuner strain; however, both proteins, *Pp*IEM and *Pn*IEM, were primarily found in the insoluble fraction of cell crude extracts. Moreover, no activity was detected from *Pp*IEM supernatants, while *Pn*IEM displayed 0.05 ± 0.01 U/mg specific activity on isoeugenol.

Considering the low yields of *Pn*IEM in the soluble fraction, an unfolding and refolding protocol was tested to take advantage of the high amounts of enzyme in inclusion bodies. (**Fig. 2.2A**). High levels of refolded protein were obtained (~25 mg), but no enzymatic activity could be detected after refolding, indicating that the Fe^{2+} cofactor was possibly lost during the unfolding-refolding process. Therefore, $FeSO_4$ was added to the refolding buffer to allow the putative Fe^{2+} incorporation into the catalytic center, and ascorbic acid was added to create a reducing environment to prevent Fe^{2+} oxidation. However, the results did not differ from the ones observed in previous conditions. A second approach was taken to increase the amount of enzyme in the soluble fraction through the co-expression of two chaperones in *E. coli* host strains (**Fig. 2.2B**). However, *Pn*IEM was not produced at increased soluble levels compared to enzyme productions without the chaperone co-expression.

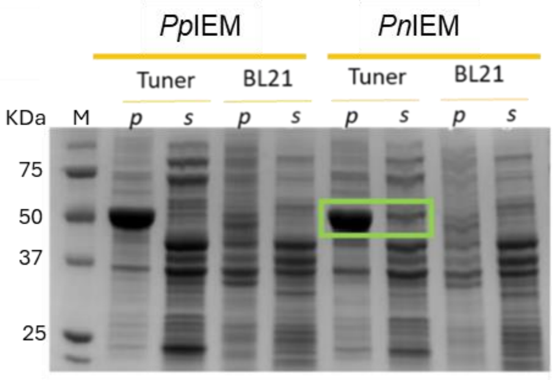


Figure 2.1. SDS-PAGE analysis of soluble (s-crude extracts) and insoluble (p-cell pellets) protein fractions from *E. coli* Tuner and BL21 strains cultures, transformed with pET15b plasmid containing the genes coding for *Pp*IEM and *Pn*IEM. The two enzymes have a molecular mass of ~54.5 kDa. In the green square, *Pn*IEM production in the Tuner strain is highlighted.

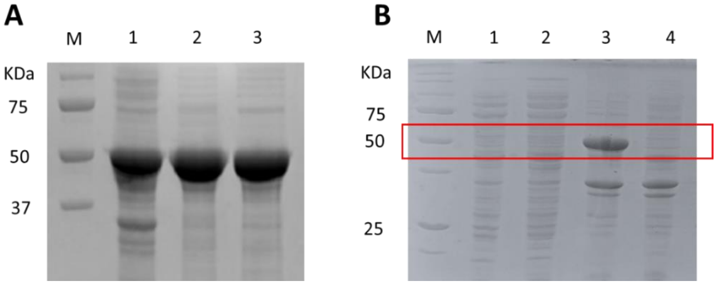


Figure 2.2. (A) *Pn*IEM inclusion body refolding procedure. M - molecular marker; 1 - cell pellet after inclusion body solubilization step; 2 - soluble fraction after unfolding step; 3 - soluble fraction after folding step. (**B)**. SDS-PAGE analysis of *Pn*IEM production in *E. coli* co-expressed with GroES and GroEL chaperones. M - molecular marker; 1 - soluble fraction of induced cells: 2 - soluble fraction of not induced cells; 3 - cell pellet of induced cells; 4 - not induced cells. *Pn*IEM has a molecular mass of approximately 54.5 kDa.

Overall, the highest levels of soluble protein were achieved with the production of *Pn*IEM in the *E. coli* Tuner strain. Therefore, this enzyme and this strain were chosen for production at a 1 L scale. An affinity chromatography method was used to purify *Pn*IEM from cell crude extracts, and the protein was eluted using an increasing imidazole gradient. However, final protein yields were about 0.5 mg/L culture, and enzyme preparations resulted in iron-depleted and very prone to aggregation. Some adjustments to the protocol were made to improve protein yields and iron loading. The volume to resuspend the cellular pellet was increased from 1 to 5 mL buffer/100 mL culture to disfavor aggregation throughout the process. The addition of $FeSO_4$ was replaced by the more stable $(NH_4)_2Fe(SO_4)_2$ salt to promote iron incorporation into the enzyme. Furthermore, 20% glycerol was added to aliquots just after elution. These adjustments resulted in an increase of 20-fold in the yield, which resulted in 10 mg protein/L culture and 60% iron loading after His-trap purification. Specific activity was also considerably higher, resulting in 2.9 ± 0.2 U/mg (**Fig. 2.3**).

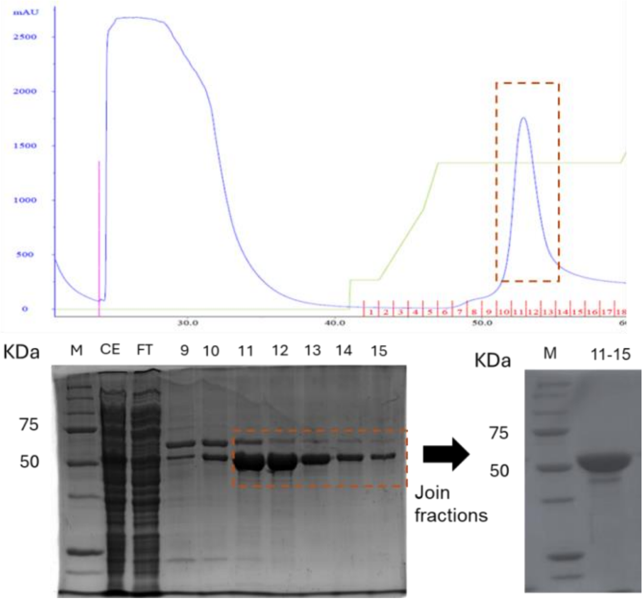


Figure 2.3. Purification chromatogram (top) of *E. coli* Tuner expressing *pn*IEM. In blue, the absorbance registered at 280 nm in the ÄKTA purifier system. In the red square, the peak corresponds to eluted *Pn*IEM fractions. SDS-PAGE analysis (bottom) resultant from the purification.

2.4.2. Biochemical Characterization of *Pn*IEM

A bell-shaped curve describes the pH profile for isoeugenol bioconversion to vanillin using *Pn*IEM (**Fig. 2.4A**) with an optimum pH of 9. Temperatures tested for isoeugenol conversion to vanillin ranged from 25 °C to 50 °C, with an optimum at around 30 °C (**Fig. 2.4B**). The kinetic parameters were calculated to evaluate *Pn*IEM efficiency in the conversion of isoeugenol to vanillin (**Fig. 2.5A**). The affinity constant (K_m) proved to be relatively low, 0.25 ± 0.03 mM, indicating high substrate affinity to the enzyme. The calculated turnover number (k_{cat}) resulted in 5.1 ± 0.2 s^{-1}, comparable to the previously reported value (Ryu *et al.* 2013). This resulted in the catalytic efficiency (k_{cat}/K_m) of (25.5 ± 2.3) x 10^3 $M^{-1}\cdot s^{-1}$. The kinetic thermostability was assessed by incubating *Pn*IEM at 25 °C (**Fig. 2.5B**). The enzyme showed a temperature-dependent activity decay, with a half-life time ($t_{1/2}$) of 1.6 ± 0.3 h. In thermal unfolding experiments (**Fig. 2.6A**), *Pn*IEM starts to denature at approximately 36 °C and displays a T_m of 45 ± 3 °C. Following at the same time the aggregation signal at 500 nm, it was possible to determine a T_{agg} at 43 °C (**Fig. 2.6B**).

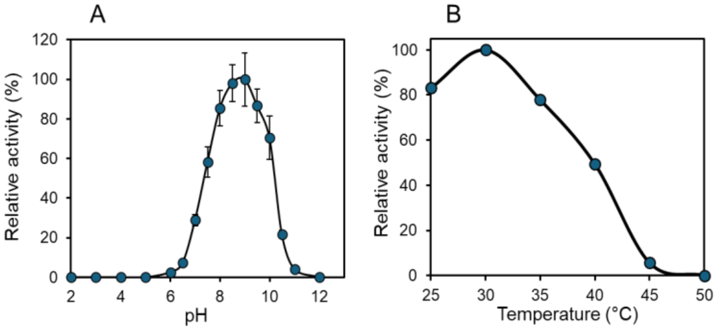


Figure 2.4. (**A**) pH profile of pure *Pn*IEM. Reactions were performed using Britton Robinson buffer (pH range from 2 to 12) in 1 mM isoeugenol at room temperature. (**B**) Temperature dependence of *Pn*IEM enzymatic activity. Reactions were performed in 100 mM Tris-HCl buffer, pH 9, in 1 mM isoeugenol.

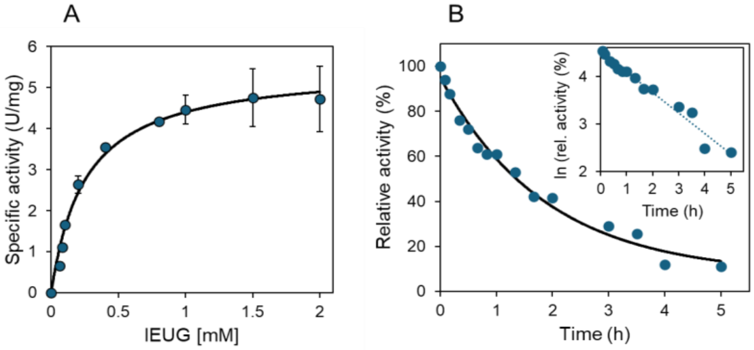


Figure 2.5. (**A**) Steady-state kinetic analysis of *Pn*IEM. Reactions were performed at room temperature using isoeugenol concentrations from 0 to 2 mM in 100 mM Tris-HCl buffer, pH 9. The activity was determined by measuring vanillin production at 340 nm (ε_{340} = 15.97 mL μmol^{-1} cm^{-1}). Kinetic data were fitted with the Michaelis-Menten equation using Origin software. (**B**) Kinetic thermostability of *Pn*IEM. An aliquot was incubated at 25 °C in 20 mM Tris-HCl buffer, pH 9, and at appropriate times, aliquots were withdrawn and tested for activity at room temperature.

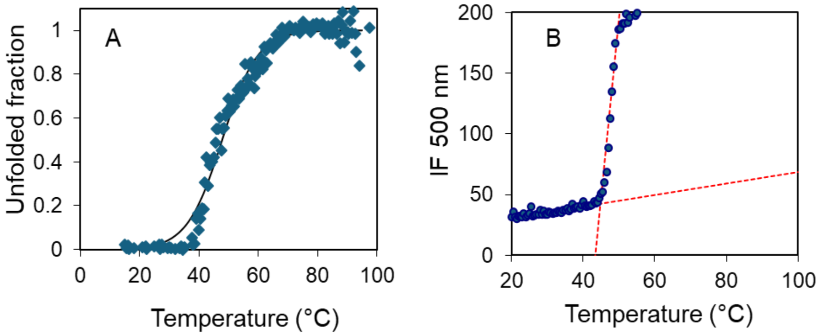


Figure 2.6. (**A**) Thermal unfolding of *Pn*IEM following fluorescence emission of tryptophan residues at 340 nm (T_m = 45 ± 3 °C) The unfolded fraction was measured from 10 to 100 °C with an increasing gradient of 1 °C/min. (**B**) Static light scattering at 500 nm (T_{agg} = 43 °C).

2.4.3. Computational Engineering to Improve Solubility and Stability

One stone of the SMARTBOX project was engineering the selected isoeugenol oxygenase to improve activity using computational tools. Therefore, it was strongly advised that a crystal structure model of the protein be available. For this reason, we extensively attempted the structural elucidation of *Pn*IEM. Although the protein could be effectively purified, the protein tended to precipitate. Therefore, it was not surprising that the enzyme resisted our extensive attempts at crystallization. As a more drastic approach, we employed FireProt1.2 and PROSS online tools to engineer *Pn*IEM to enhance its stability and solubility and reduce its tendency to aggregate. Tools such as FireProt (Bednar, Beerens *et al.* 2015, Musil, Stourac *et al.* 2017), PROSS (Goldenzweig, Goldsmith *et al.* 2016) have a multilevel framework combining consensus analysis, energy calculations and in FRESCO, *de novo design*, to analyze possibly stabilizing mutations from sequence-based and energy-based analysis, calculates possible antagonistic effects among predicted mutations and provide a list of target positions for experimental verification. These tools generated two variant designs, introducing 64 and 70 mutations, respectively (**Table 2.1**). Ten of these mutations targeted the same positions, while other nine were identical across both designs.

Table 2.1. Mutations designed by Fireport (sixty-six) and PROSS (seventy). Mutations in the same positions are underlined; identical ones are in bold. The asterisk indicates mutations at < 4 Å from the active center that were excluded from the final designs.

Fireprot		**PROSS**	
L4P	I247P	L11M	**L273F**
N5P	Y248F	L15F	K274E
T14W	E251D	L16A	L278C
D30E	G255L	S36D	W279Y*
S36E	S268G	A65V	F281G*
I37L	**L273F**	D72N	V284M
E47D	W279F*	F77Y	V289E
Q49F	N309D	**I78R**	T291N
T51P	Q311D	S79N	A316D
T57W	N312G	A85E	S334G
I59F	Q313F	T88E	I392K
S66M	A316I	L92A	W393Y
I78R	V321W	**K95R**	N395H
K83R	E328T	T109P	H396S
R86K	S334N	**S122N***	R397T
R94G	S335D	S125W	D399E
K95R	M350F	R145D	D401Q
S96A	I352R	F153Y	L402Y
S122N*	L359G	D176E	S405A
S125V	N362Y	A183E	A408G
H127A	Y364H	A188G	A410Y
Y155F	G365F	P190R	A411V
Y170I	A377W	H200N	**T422D**
D176E	A381G	**G214P**	A424P
L177M	F385W	**I222V**	Y429W
S182Y	N388A	A234L	G435S
T207I	S389A	**S239E**	L437Y
Q211A	N395F	A243K	**T449A**
G214P	**T422D**	K244R	Q453S
I222V	G435L	G255P	T459R

W226Y	**T499A**	I258F	K461R
S227V	A468W	R264H	L466V
S239E	C473N	Q266T	A468Y
		G267A	A469G
		Q269D	L470F*

The variants were expressed in *E. coli* strains; however, both attempts were unsuccessful, resulting in inactive variants and no improvement in solubility. In a final effort, we considered using a DNA shuffling technique to recombine the FireProt variant with the wild-type, aiming to remove destabilizing mutations selectively. However, after characterizing the T_m of this variant, which showed no improvement (45 °C) compared to the wild-type (**Fig. 2.7**), we opted to pursue an alternative enzyme candidate instead. This result underscores the limited reliability of rational design strategies in the absence of an X-ray crystal structure.

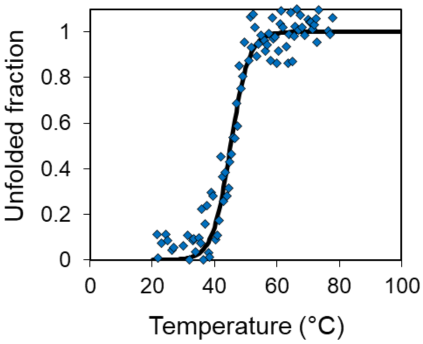


Figure 2.7. Thermal unfolding of *Pn*IEM FireProt1.2 variant following fluorescence emission of tryptophan residues at 340 nm (T_m = 45 ± 3 °C) The unfolded fraction was measured from 10 to 100 °C with an increasing gradient of 1 °C/min.

2.4.4. Production of NOV1 from *N. aromaticivorans* and comparison to *Pn*IEM

Considering these results, we searched for alternatives to *Pn*IEM. Sequence comparisons highlighted the *Lignostilbene dioxygenase* A from *Sphingomonas paucimobilis* (LsdA) and the stilbene dioxygenase (SCD) from *N. aromaticivorans* (NOV1), with 43% and 41% identity to *Pn*IEM, respectively. Past literature described that most of the recombinant LsdA could not easily be expressed in to *Pn*IEM.

The pH and the temperature profiles of NOV1 are very similar to those observed *E. coli,* whereas NOV1 was reportedly expressed as an active and soluble protein (Kuatsjah *et al.* 2019; McAndrew *et al.* 2016). The nov1 gene was chosen to clone in the plasmid pET28a with a SUMO-tag for these reasons. The overproduction of recombinant NOV1 dioxygenase was tested in several *E. coli* strains and BL21star was selected as the host since it resulted in higher yields of the soluble enzyme (data not shown). As observed by SDS-PAGE analysis (**Fig. 2.8A**), the protein is mainly produced in the soluble fraction, in contrast to what was observed with IEMs (**Fig. 2.1**). After one purification step by affinity chromatography, it was possible to obtain a protein yield of 100 mg/L culture (10-fold higher than *PnI*EM) with a high purity level, as observed in the SDS-PAGE (**Fig. 2.8B**). NOV1 presented an iron load of around 50%, similar to *Pn*IEM, with an optimal pH of 9 (**Fig. 2.9A**) and an optimal temperature of around 30 °C (**Fig. 2.9B**). The kinetic parameters were determined to evaluate NOV1 in converting isoeugenol to vanillin (**Fig. 2.10A**). The catalytic efficiency of 13×10^3 $M^{-1}{\cdot}s^{-1}$ is lower than that of *Pn*IEM ($k_{cat}/K_m = 5.1 \times 10^3$ $M^{-1}{\cdot}s^{-1}$) mainly due to a 2-fold increased K_m. However, the k_{cat} value is slightly higher, 7.5 ± 0.2 s^{-1} (**Table 2.2**). The kinetic thermostability of NOV1 was assessed at 25 °C (**Fig.**

2.10B), and the calculated half-life ($t_{1/2}$) resulted in 1.7 ± 0.4 h, the same as *Pn*IEM. Thermodynamic stability was investigated, and a T_m of 58 ± 1 °C (**Fig. 2.11A**) was calculated, indicating that NOV1 is more thermostable than *Pn*IEM (T_m = 43 ± 3 °C). Moreover, NOV1 aggregates at a slightly higher temperature than *Pn*IEM (T_{agg} = 47 °C) (**Fig. 2.11B**).

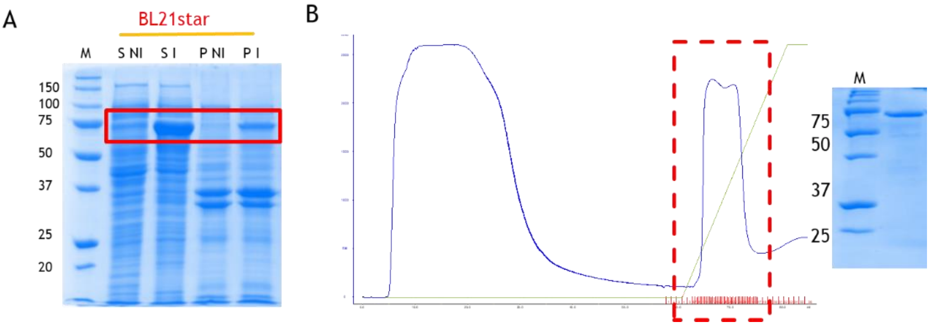


Figure 2.8. (**A**) SDS-PAGE analysis of soluble (s-crude extracts) and insoluble (p-cell pellets) protein fractions from cultures of IPTG-induced (I) and non-induced (NI) recombinant *E. coli* BL21star strain transformed with a plasmid containing the gene coding for NOV1. In the red square, the NOV1-SUMO band is highlighted (70 KDa). (**B**) Purification chromatogram of His-trap purification. In blue, the absorbance registered at 280 nm by the ÄKTA purifier system. In green is the percentage of elution buffer. In the red square are the recovered fractions and corresponding SDS-PAGE analysis.

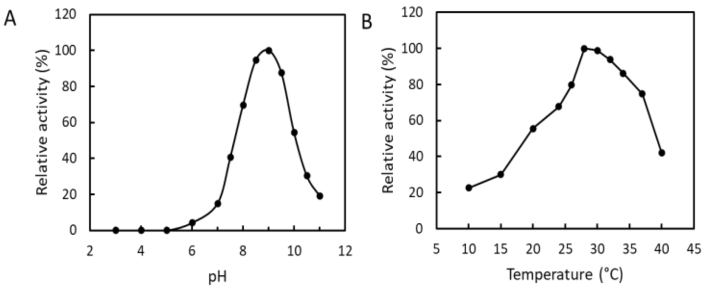


Figure 2. 9. (**A**) pH profile of pure NOV1. Reactions were performed in Britton Robinson buffer (pH range from 2 to 12). (**B**) Temperature dependence of NOV1 enzymatic activity. Reactions were performed in 100 mM Tris-HCl buffer, pH 9, and 1 mM isoeugenol.

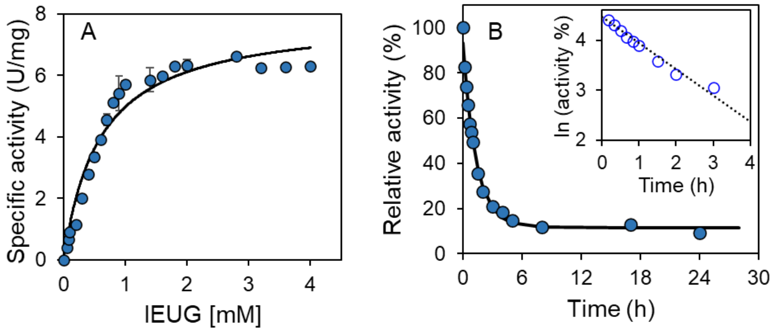


Figure 2.10. (**A**) Steady-state kinetic analysis of NOV1. Reactions were performed at room temperature using isoeugenol concentrations from 0 to 4 mM in 100 mM Tris-HCl buffer, pH 9. Kinetic data were fitted with the Michaelis-Menten equation using Origin software. (**B**) Kinetic thermostability of NOV1. An aliquot was incubated at 25 °C in 20 mM Tris-HCl buffer pH 9, and at appropriate times, aliquots were withdrawn and tested for activity at room temperature.

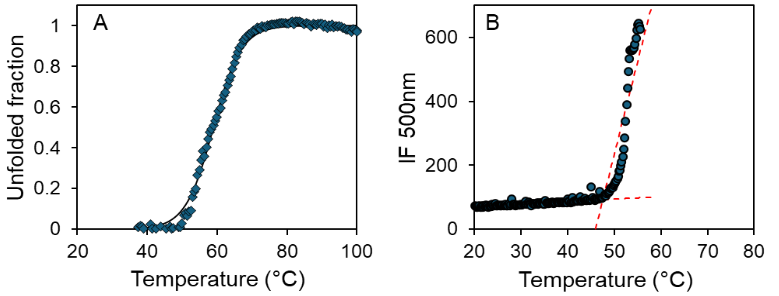


Figure 2.11. (**A**) Thermodynamic stability of NOV1 (T_m = 58 ± 1 °C). The unfolded fraction was measured from 10 to 100 °C with an increasing gradient of 1 °C/min. (**B**) Static light scattering at 500 nm (T_{agg} = 47 °C).

Table 2.2. Apparent steady-state kinetic parameters of *Pn*IEM and NOV1 for isoeugenol. Assays were performed at room temperature in 0.1 M Tris-HCl, pH 9.

	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$) × 10^3
*Pn*IEM	5.1 ± 0.2	0.25 ± 0.03	20 ± 2
NOV1	7.5 ± 0.2	0.6 ± 0.1	13 ± 2

2.4.5. Substrate Scope of NOV1 from *N. aromaticivorans*

CCOs possess relevant biological importance, constituting a family of ubiquitous enzymes involved in many processes. A recent study by Metz *et al.* (2024) identified NOV1 (LsdD) as key in the degradation of lignin-derived dimers, indicating a significant role in lignin catabolism. To characterize the substrate scope of NOV1, the lignin-related phenolic compounds were measured in the presence of molecular oxygen and hydrogen peroxide **(Table 2.3)**. We have tested H_2O_2 as a substrate since data from De Vitis *et al.* (2023) showed that Cso2 oxygenase oxidized 2,6-dimethoxyphenol (DMP) using H_2O_2 despite the lack of prior reports of peroxidase active among CCO enzymes. However, dual oxygenase/peroxidase activity is not unprecedented among iron-dependent enzymes (Hrycay and Bandiera 2015; Wang *et al*. 2013). In expanding NOV1's substrate scope, we observed peroxidase-like activity on phenolics, including DMP, guaiacol, and ABTS (**Table 2.3**). Notably, activity on DMP was optimal at a pH of 10 (**Fig. 2.12A**). However, the K_m for H_2O_2 is relatively high (1.3 ± 0.3 mM) (**Fig. 2.12B**), indicating that NOV1's primary function likely remains the cleavage of C=C bonds in stilbenoid and phenylalkyloid compounds. On the other hand, the K_m for DMP (0.3 ± 0.1 mM) and the k_{cat} ($1.0 \pm 0.1\ s^{-1}$) are within the range observed for specific peroxidases (Linde *et al.* 2021; Rodrigues *et al*. 2021; Silva *et al.* 2023). Notably, NOV1 showed no inhibition even at H_2O_2 concentrations up to 10 mM (**Fig. 2.12B**). It is unclear whether, in the NOV1 case, peroxidase activity has biological relevance. However, the alkaline optimal pH and lack of H_2O_2 inhibition at high concentrations are valuable characteristics for industrial applications. Interestingly, NOV1 in a dioxygenase activity can use coniferyl alcohol, albeit at levels five times lower than those for isoeugenol for the production of vanillin as it was confirmed by NMR (**Fig. 2.13**), highlighting NOV1's potential to catalyze the one-step conversion of another key intermediate in lignin depolymerization into vanillin.

Table 2.3. NOV1 specific activity on phenolic substrates and ABTS, using O_2 or H_2O_2 (4 mM) as co-substrate. Reactions were performed at room temperature in 0.1 M Tris-HCl buffer pH 9 or pH 7 for ABTS. Reactions were monitored at the spectrophotometer following substrate consumption or product formation.

Substrates	**λ (nm)**	**ε (M^{-1} cm^{-1})**	**Activity (μmol min^{-1} mg^{-1})**	
Syringyl-type phenolics			$+O_2$	$+H_2O_2$
2,6-DMP	468*	49600	N.D.	0.9 ± 0.2
Sinapyl alcohol	275	2503	0.03 ± 0.01	N.D.
Acetosyringone	360	11162	N.D.	0.04 ± 0.01
Methyl syringate	280	8056	N.D.	0.11 ± 0.01
Syringaldehyde	360	21812	N.D.	0.02 ± 0.01
Sinapic acid	310	14820	N.D.	N.D.
Syringic acid	260	7473	N.D.	N.D.
Guaiacyl-type phenolics				
Isoeugenol	340*	15970	3.1 ± 0.2	N.D.
Guaiacol	470*	26600	N.D.	0.6 ± 0.2
Coniferyl alcohol	340*	15970	0.6 ± 0.2	N.D
Coniferyl aldehyde	340*	15970	N.D.	0.14 ± 0.02
Ferulic acid	300	13495	N.D.	N.D
Vanillyl alcohol	280	2570	N.D.	N.D
Vanillin	340	15970	N.D.	N.D
Hydroxybenzene-type phenolics				
Caffeic acid	320	3646	N.D.	N.D.
p-Coumaric acid	300	13642	N.D.	N.D.
Gallic acid	260	7040	N.D.	N.D.
Resveratrol	306	31800	0.3 ± 0.1	N.D.
ABTS	420	36000	N.D.	0.12 ± 0.02 (pH 7)

*product absorbance peak; N.D.= not detected.

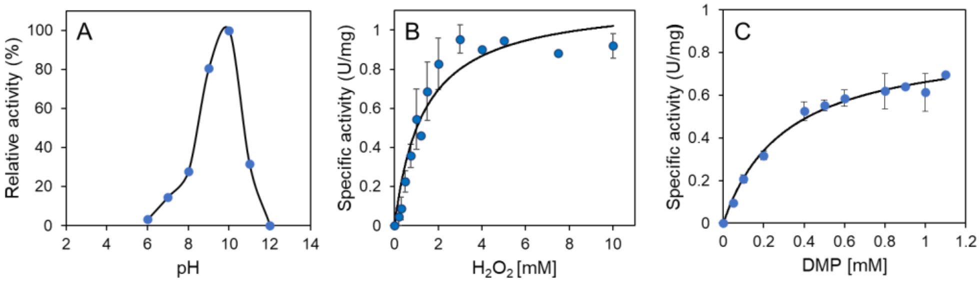


Figure 2.12. Characterization of NOV1 peroxidase activity on 2,6-DMP. (**A**) pH profile of NOV1 peroxidase activity. Reactions were performed at room temperature using Britton Robinson buffer (pH range from 4 to 12), 1 mM DMP, 4 mM H_2O_2 at room temperature. (**B**) Steady-state kinetic analysis of NOV1 for H_2O_2 in concentrations from 0 to 10 mM, in 1 mM 2,6-DMP. (**C**) Steady-state kinetic analysis for 2,6-DMP using concentrations from 0 to 1.2 mM in 4 mM H_2O_2. Reactions were performed in 100 mM glycine-NaOH buffer, pH 10. The activity was determined by measuring absorbance increase at 468 nm ($\varepsilon_{420} = 49.6 \times 10^3$ M^{-1} cm^{-1}). Kinetic data were fitted with the Michaelis-Menten equation using Origin software.

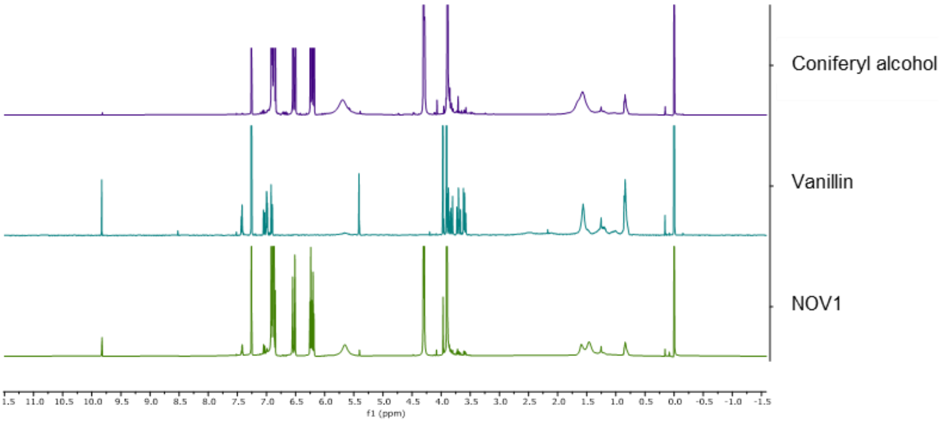


Figure 2.13. ^{1}H-NMR spectra of coniferyl alcohol, vanillin and conversion catalyzed by NOV1 wild type. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): coniferyl alcohol: 3.90 (s, 3H), 4.29(dd, J = 1.2, 6.0 Hz, 2H), 5.69 (s, 1H), 6.23 (dt, J = 6.0, 15.8 Hz, 1H), 6.53 (d, J = 15.8 Hz, 1H), 6.84-6.91 (m, 3H); vanillin: 9.82 (s, 1H), 7.44 – 7.40 (m, 2H), 7.04 (d, J = 8.5 Hz, 1H), 6.27 (s, 1H), 3.96 (s, 3H). Numbers in labels represent the value of the integral of each peak, as MestReNova gives.

2.4.6. Isoeugenol to Vanillin Conversion by Whole-Cells

For biotechnological purposes, whole-cell systems overproducing the enzymes of interest for bioconversions can be of interest as they are more cost-effective than purified enzyme preparations. Hence, we investigated whole-cell systems for isoeugenol bioconversion to vanillin. Reactions by whole-cells, that overproduced NOV1, resulted in 90% of isoeugenol bioconversion to vanillin after 16 h in the presence of 30 mM substrate (**Fig. 2.14**).

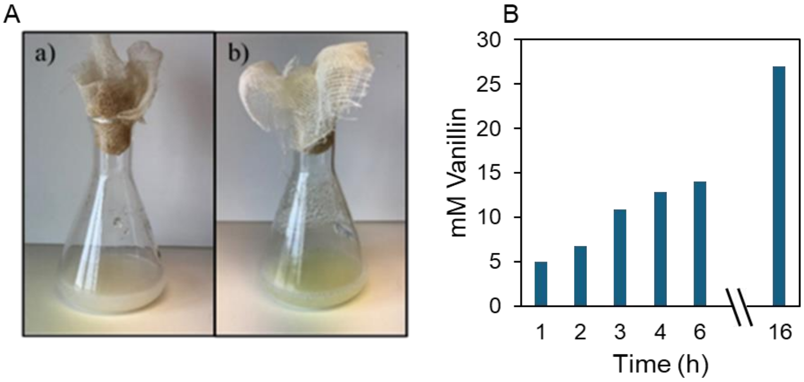


Figure 2.14. Isoeugenol bioconversion to vanillin in whole-cell reactions. (**A**) Whole-cell reaction setting-up: aspect of reaction solution a) at initial time and b) after 24 h reaction. (**B**) Vanillin yields over time using 2 OD/mL resting cells overproducing NOV1 in 0.1 M Tris-HCl, pH 9, at 30 °C, 150 rpm.

2.5. CONCLUDING REMARKS

We have expressed the genes of three different enzymes showing isoeugenol oxygenase activity (*P. putida Pp*IEM, *P. nitroreducens Pn*IEM, and *N. aromaticivorans* NOV1) in *E. coli* strains and have produced, purified, characterized, and tested the conversion of isoeugenol into vanillin using the *Pn*IEM and NOV1 enzymes. *N. aromatocivorans* NOV1, a stilbene cleavage oxygenase, was selected for further investigations considering the higher levels of soluble enzyme production (10-fold higher than *Pn*IEM), improved thermostability (~10 °C higher T_m), and solubility as compared to *Pn*IEM. Moreover, and importantly, crystal structures of this enzyme are available (PDB codes: 5J53, 5J54 and 5J55), paving the way to the computational *in silico* design of mutants with improved catalytic efficiency and enhanced thermostability required for industrial applications.

2.6. REFERENCES

Bai, J., Hou, Q. Q., Zhu, W. Y., and Liu, Y. J. 2019. 'Mechanical insights into the oxidative cleavage of resveratrol catalyzed by dioxygenase NOV1 from Novosphingobium aromaticivorans: confirmation of dioxygenase mechanism by QM/MM calculations', *Catal. Sci. Technol.*, 9: 444-55.

Bednar, D., K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, Z. Prokop, J. Brezovsky, D. Baker, and J. Damborsky. 2015. 'FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants', *PLoS Comput. Biol.*, 11

Daruwalla, A., and Kiser, P. D. 2020. 'Structural and mechanistic aspects of carotenoid cleavage dioxygenases (CCDs)', *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, 1865: 158590.

De Vitis, V., P. Cannazza, L. Mattio, D. Romano, A. Pinto, F. Molinari, T. Laurenzi, I. Eberini, and M. L. Contente. 2023. 'Caulobacter segnis Dioxygenase CsO2: A Practical Biocatalyst for Stilbenoid Ozonolysis', *ChemBioChem*, 24: e202300477.

Eftink, M. R. 1998. 'The use of fluorescence methods to monitor unfolding transitions in proteins', *Biochemistry (Mosc)*, 63: 276-84.

Fischer, D. S., and Price, D. C. 1964. 'A simple serum iron method using the new sensitive chromogen tripyridyl-s-triazine', *Clin. Chem.*, 10: 21-31.

Frusciante, S., Diretto, G., Bruno, M., Ferrante, P., Pietrella, M., Prado-Cabrero, A., Rubio-Moraga, A., Beyer, P., Gomez-Gomez, L., Al-Babili, S., and Giuliano, G. 2014. 'Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis', *Proc. Natl. Acad. Sci. USA*, 111: 12246-51.

Goldenzweig, A., M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. L. Lieberman, A. Aharoni, I. Silman, J. L. Sussman, D. S. Tawfik, and S. J. Fleishman. 2016. 'Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability', *Mol. Cell*, 63: 337-46.

Hrycay, E. G., and S. M. Bandiera. 2015. 'Monooxygenase, peroxidase and peroxygenase properties and reaction mechanisms of cytochrome P450 enzymes', *Adv Exp Med Biol*, 851: 1-61.

Holger, S., Kurtzer, R., Eisenreich, W., and Schwab, W. 2006. 'The carotenase AtCCD1 from Arabidopsis thaliana is a dioxygenase', *J. Biol. Chem.*, 281: 9845-51.

Huang, F. C., Horváth, G., Molnár, P., Turcsi, E., Deli, J., Schrader, J., Sandmann, G., Schmidt, H., and Schwab, W. 2009. 'Substrate promiscuity of RdCCD1, a carotenoid cleavage oxygenase from', *Phytochemistry*, 70: 457-64.

Kamoda, S., and Saburi, Y. 1993. 'Structural and enzymatical comparison of lignostilbene-alpha,beta-dioxygenase isozymes, I, II, and III, from Pseudomonas paucimobilis TMY1009', *Biosci. Biotechnol. Biochem.*, 57: 931-4.

Kamoda, S., and Saburi, Y. 1995. 'Cloning of a lignostilbene-alpha,beta-dioxygenase isozyme gene from Pseudomonas paucimobilis TMY1009', *Biosci. Biotechnol. Biochem.*, 59: 1866-8.

Kuatsjah, E., Verstraete, M. M., Kobylarz, M. J., Liu, A. K. N., Murphy, M. E. P., and Eltis, L. D. 2019. 'Identification of functionally important residues and structural features in a bacterial lignostilbene dioxygenase', *J. Biol. Chem.*, 294: 12911-20.

Linde, D., I. Ayuso-Fernández, M. Laloux, J. E. Aguiar-Cervera, A. L. de Lacey, F. J. Ruiz-Dueñas, and A. T. Martínez. 2021. 'Comparing Ligninolytic Capabilities of Bacterial and Fungal Dye-Decolorizing Peroxidases and Class-II Peroxidase-Catalases', *Int J Mol Sci*, 22.

Lu, J. R., and Lai, W. Z. 2019. 'Mechanistic insights into a stilbene cleavage oxygenase NOV1 from quantum mechanical/molecular mechanical calculations', *ChemistryOpen*, 8: 228-35.

Marasco, E. K., and Schmidt-Dannert, C. 2008. 'Identification of bacterial carotenoid cleavage dioxygenase homologues that cleave the interphenyl alpha,beta double bond of stilbene derivatives via a monooxygenase reaction', *Chembiochem*, 9: 1450-61.

McAndrew, R. P., Sathitsuksanoh, N., Mbughuni, M. M., Heins, R. A., Pereira, J. H., George, A., Sale, K. L., Fox, B. G., Simmons, B. A., and Adams, P. D. 2016. 'Structure and mechanism of NOV1, a resveratrol-cleaving dioxygenase', *PNAS. USA*, 113: 14324-29.

Metz, F., A. M. Olsen, F. C. Lu, K. S. Myers, M. N. Allemann, J. K. Michener, D. R. Noguera, and T. J. Donohue. 2024. 'Catabolism of β-5 linked aromatics by', *mBio*, 15.

Musil, M., J. Stourac, J. Bendl, J. Brezovsky, Z. Prokop, J. Zendulka, T. Martinek, D. Bednar, and J. Damborsky. 2017. 'FireProt: web server for automated design of thermostable proteins', *Nucleic Acids Res.*, 45: W393-W99.

Ortiz, A., and Sansinenea, E. 2023. 'Phenylpropanoid derivatives and their role in plants' health and as antimicrobials', *Current Microbiol.*, 80.

Rodrigues, C. F., P. T. Borges, M. F. Scocozza, D. Silva, A. Taborda, V. Brissos, C. Frazao, and L. O. Martins. 2021. 'Loops around the Heme Pocket Have a Critical Role in the Function and Stability of DyP from Bacillus subtilis', *Int J Mol Sci*, 22.

Ryu, J. Y., Seo, J., Park, S., Ahn, J. H., Chong, Y., Sadowsky, M. J., and Hur, H. G. 2013. 'Characterization of an isoeugenol monooxygenase (Iem) from Pseudomonas nitroreducens Jin1 that transforms isoeugenol to vanillin', *Biosci. Biotechnol. Biochem.*, 77: 289-94.

Ryu, J. Y., Seo, J., Unno, T., Ahn, J. H., Yan, T., Sadowsky, M. J., and Hur, H. G. 2010. 'Isoeugenol monooxygenase and its putative regulatory gene are located in the eugenol metabolic gene cluster in Pseudomonas nitroreducens Jin1', *Arch. Microbiol.*, 192: 201-09.

Silva, D., A. C. Sousa, M. P. Robalo, and L. O. Martins. 2023. 'A wide array of lignin-related phenolics are oxidized by an evolved bacterial dye-decolourising peroxidase', *New Biotechnol*, 77: 176-84.Sui, X. W., Weitz, A. C., Farquhar, E. R., Badiee, M., Banerjee, S., von Lintig, J., Tochtrop, G. P., Palczewski, K., Hendrich, M. P., and Kiser, P. D. 2017. 'Structure and spectroscopy of alkene-cleaving dioxygenases containing an atypically coordinated non-heme iron center', *Biochemistry*, 56: 2836-52.

Sun, Z. H., Fridrich, B., de Santi, A., Elangovan, S., and Barta, K. 2018. 'Bright side of lignin depolymerization: toward new platform chemicals', *Chem. Rev.*, 118: 614-78.

Van den Bosch, S., Schutyser, W., Vanholme, R., Driessen, T., Koelewijn, S. F., Renders, T., De Meester, B., Huijgen, W. J. J., Dehaen, W., Courtin, C. M., Lagrain, B., Boerjan, W., and Sels, B. F. 2015. 'Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps', *Energy Environ. Sci.*, 8: 1748-63.

Wang, C., W. C. Chang, Y. Guo, H. Huang, S. C. Peck, M. E. Pandelia, G. M. Lin, H. W. Liu, C. Krebs, and J. M. Bollinger, Jr. 2013. 'Evidence that the fosfomycin-producing epoxidase, HppE, is a non-heme-iron peroxidase', *Science*, 342: 991-5.

Yamada, M., Okada, Y., Yoshida, T., and Nagasawa, T. 2007. 'Purification, characterization and gene cloning of isoeugenol-degrading enzyme from Pseudomonas putida IE27', *Arch. Microbiol.*, 187: 511-17.

Yamada, M., Okada, Y., Yoshida, T., and Nagasawa, T. 2008. 'Vanillin production using Escherichia coli cells over-expressing isoeugenol monooxygenase of Pseudomonas putida', *Biotechnol. Lett.*, 30: 665-70.

Chapter 3

Engineering NOV1 Oxygenase and Its Implementation in a Catalytic Cascade for Efficient Vanillin Production

This chapter contains data published in:

Mario De Simone, Laura Alvigini, Lur Alonso-Cotchico, Vânia Brissos, Jonatan Caroli, Maria Fátima Lucas, Emanuele Monza, Eduardo Pinho Melo, Andrea Mattevi and Lígia O. Martins (2022). Rational-Guided Improvement of NOV1 Dioxygenase for the Conversion of Lignin-Derived Isoeugenol to Vanillin. *Biochemistry 2023, 62, 2, 419–428.*
https://doi.org/10.1021/acs.biochem.2c00168

Ivana Marić, Yiming Guo, Maximilian J. L. J. Fürst, Korneel Van Aelst, Sander Van den Bosch, **Mario De Simone**, Lígia O. Martins, Bert F. Sels, Marco W. Fraaije (2023). A One-Pot, Whole-Cell Biocatalysis Approach for Vanillin Production using Lignin Oil. *Advanced Synthesis & Catalysis 2023, 365, 22, 3987-3995.*
https://doi.org/10.1002/adsc.202300868

Author contributions:

The author of this thesis performed kinetic and biochemical characterization experiments, including enzyme characterization and set-up the conversions of isoeugenol to vanillin by whole-cells. Vânia Brissos supervised mutagenesis, cloning, and gene expression. Eduardo Pinho e Melo supervised the stability experiments. Laura Alvigini and Jonatan Cairoli crystallized the variant, collected data, and analyzed the structure at University of Pavia. Lur Alonso-Cotchico, Emanuele Monza and Maria Fatima Luca designed the variants with Rosetta and performed molecular dynamics and docking experiments at Zymvol Biomodeling (Barcelona). Optimization and analysis of one-pot cascade reactions were performed by Ivana Marić, Yiming Guo, and Maximilian J.L.J. Fürst at the University of Groningen. Korneel Van Aelst, Sander Van den Bosch and Bert F. Sels prepared and analyzed the lignin oil from wood sawdust at KU Leuven University.
This work was conceptualized, supervised and revised by Andrea Mattevi, Marco Fraaje and Ligia O Martins.

3.1. ABSTRACT

In this chapter, we made advancements in understanding the structure-function relationships of NOV1 through rational mutagenesis near the active site. We integrated computational, kinetic, structural, and biophysical approaches, leading to identifying the NOV1-S283F variant that exhibited a 2-fold increase in activity and a 20-fold longer half-life compared to the wild-type. Biochemical and molecular dynamics experiments elucidated that S283F replacement in the active site enhances the stabilization of the iron cofactor, revealing cofactor retention as a critical factor for enzyme stability. The potential of NOV1 was further demonstrated in whole-cell reactions, where cells producing NOV1-S283F efficiently converted >99% of 100 mM isoeugenol to vanillin within 24 h. Subsequently, we set up a coupled reaction in which a eugenol oxidase initially converts lignin-derived 4-*n*-propyl guaiacol (4PG) into isoeugenol, followed by its conversion to vanillin by NOV1-S283F. This approach led to a 90% vanillin yield starting from commercial 4PG and a substantial 66% yield using 4PG from natural lignin oil. This integrative strategy provided a new enzyme for biotechnological applications and mechanistic insights to facilitate the future design of robust and efficient biocatalysts.

Keywords: Biocatalysis, Oxygenases, Non-heme iron proteins, Cascade reactions, Lignin valorization, Biorefinery, Green Chemistry

3.2. INTRODUCTION

NOV1, also known lignostilbene dioxygenase (LsdD), is a non-heme, iron-dependent dioxygenase that catalyzes the cleavage of double bonds in stilbene compounds such as resveratrol, oxyresveratrol, and piceatannol, yielding small-modified benzaldehydes (Marasco and Schmidt-Dannert 2008; McAndrew *et al.* 2016). In its natural host, *Novosphingobium aromatocivorans,* NOV1 plays a key role in the breakdown of aromatic dimers, converting them to aromatic monomers, like vanillin and 5-formylferulate, that can be later processed through the aromatic central pathway (Metz *et al.* 2024). NOV1 has also shown activity on single-ring aromatic compounds, such as isoeugenol, where it oxidatively cleaves the C-C double bond of the propenyl functional group to produce vanillin, albeit at low efficiencies (McAndrew *et al.* 2016**).** Notably, this reaction does not require coenzymes, simplifying the bioprocess and reducing the costs (Marasco and Schmidt-Dannert 2008; Sui *et al.* 2015). Analogously to all CCO enzymes, the NOV1 structure is a seven-bladed β-propeller (**Fig. 3.1A**). A ferrous ion (Fe^{2+}), which is essential for catalysis, is placed at the center of the central axis (McAndrew et al. 2016) and is directly coordinated by four conserved histidine residues (H167, H218, H284, H476) (**Fig. 3.1B**), three of which are stabilized by three second-sphere glutamate residues (E135, E353, E418) by hydrogen bonding interactions. The geometry of this metal binding is best described as either square pyramidal or trigonal bipyramidal with iron-His bond lengths in the range of 2.1-2.2 Å. Helical and loop segments connecting the β-strands form a cap over one face of the entire. While the β-propeller core is structurally preserved across the CCO superfamily, the cap region shows significant variability and conformational flexibility, influencing substrate recognition and function (Daruwalla and Kiser 2020;

Sui *et al.* 2017). The size of the active site correlates with substrate range. For example, canonical CCOs typically feature long tunnels to accommodate carotenoid molecules (Kiser *et al.* 2009; Kloer *et al.* 2005; Messing *et al.* 2010). In contrast, enzymes of the subfamily of stilbene cleavage oxygenases (SCOs), including NOV1, possess a narrower and flatter active site pocket, consistent with the expected planar structure of the stilbenoid substrates (Kuatsjah *et al.* 2019; McAndrew *et al.* 2016; Sui *et al.* 2017). The substrate-binding cleft is predominantly covered by aromatic side chains, which may facilitate substrate binding and stabilize reaction intermediates through van der Waals forces or π-stacking interactions (Sui *et al.* 2017). Additionally, these structures revealed a pre-active site binding pocket closer to the surface of the protein where the substrate may initially interact before moving to the site of catalysis.

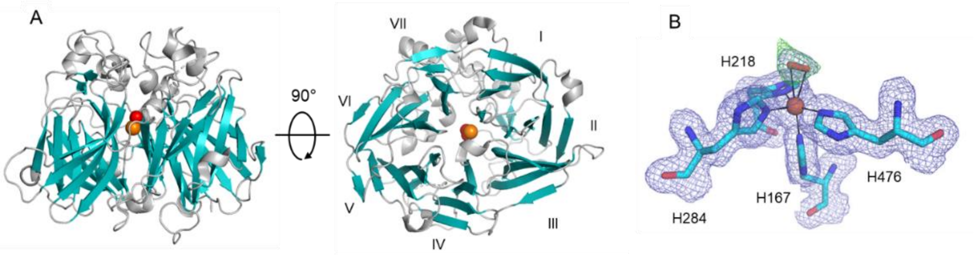
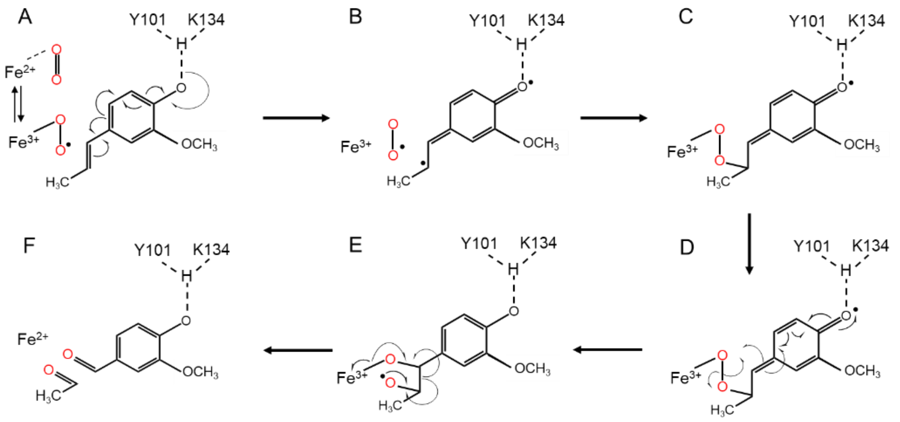


Figure 3.1. **A)** Side and under view of NOV1 (*N. aromatocivorans*) crystal structure (5J53). The seven β-sheets are highlighted in cyan, and the Fe^{2+} cofactor and O_2 are shown as orange-red spheres, respectively. **B)** Coordination of the Fe^{2+} cofactor by the four His residues. Structural images were generated using PyMOL. Figure B was adapted from (McAndrew *et al.* 2016).

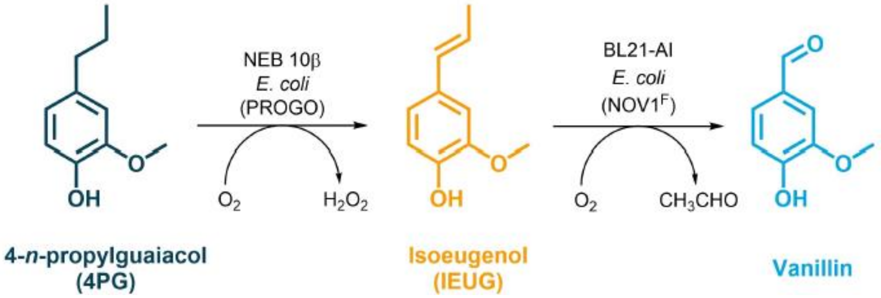
In the NOV1 substrate binding pocket, the side chains of conserved Y101 and K134 interact with the 4’-hydroxyl group of the substrate, increasing its nucleophilicity and helping the delocalization of electrons necessary for subsequent C=C cleavage. Substitution of these residues dramatically decreases catalytic activity, supporting their key functional roles (Kuatsjah *et al.* 2019). The conserved interactions involving these residues help explain the absolute requirement for a 4’-hydroxyl group in the substrates of these enzymes (Kamoda, Terada, and Saburi 2003; McAndrew *et al.* 2016). In the proposed mechanism of reaction (**Scheme 3.1**), the catalysis occurs as follows: (A) Activation of the 4'-OH group of the substrate promotes delocalized electron density toward the central double, forming the ternary complex of Fe^{3+}-superoxide-substrate; deprotonation of 4′-OH by Tyr and Lys residues contributes to the activation of the substrate. (B) Intermediate formed before the formation of the first C−O bond. (C) Intermediate formed after the formation of the first C−O bond. (D) Cleavage of the O−O bond and formation of the second C−O bond. (E) Cleavage of the C−C bond and restoration of the Fe^{2+}-enzyme. (F) Release of two aldehyde products. Other residues in the active site are responsible for enzyme specificity. For instance, in NOV1, a small serine residue (S283) is placed perpendicular to the plane of the substrate to allow both the binding of the stilbene compound and isoeugenol. In isoeugenol oxygenases from *Pseudomonas nitroreducens* and *Pseudomonas putida*, a bulky phenylalanine is placed in the same position, which makes these enzymes specific for phenylpropanoids (Ryu *et al.* 2013; Yamada *et al.* 2007).



Scheme 3.1. The proposed NOV1 reaction mechanism for isoeugenol was adapted from McAndrew et al. (2016). (**A**) Formation of the ternary complex of Fe(III)-superoxo, O_2, and isoeugenol substrate, and contributions of deprotonation of 4'-OH by Y101 and K135 in activation of the substrate. (**B**) Intermediate formed before the formation of the first C-O bond. (**C**) Intermediate formed after the formation of the first C-O bond. (**D**) Cleavage of the O-O bond and formation of the second C-O bond. (**E**) Cleavage of the C-C bond and restoration of the Fe(II) enzyme. (**F**) Vanillin and acetaldehyde are the reaction products.

In this chapter, our primary goal was to enhance the activity of NOV1 towards isoeugenol. To achieve this, we designed thirty-five variants based on a detailed structural analysis of NOV1, incorporating *in silico* docking studies, comparative structural alignments, and Rosetta computational design. These variants were constructed and tested for their activity against the isoeugenol substrate. Among them, the S283F variant emerged as the most active and stable, showing great potential as a biocatalyst for vanillin production. Additionally, a combination of kinetic studies, stability assessments, X-ray diffraction, and molecular dynamics simulations helped to identify the molecular mechanisms underlying the improved properties of the S283F variant.

Next, we developed a biocatalytic cascade to produce vanillin from lignin-derived 4-*n*-propylguaiacol (4PG). Specifically, we selected an engineered 4PG oxidase (PROGO) from *Rhodococcus jostii* RHA1 for the selective production of isoeugenol (Guo *et al.* 2022) and NOV1-S283F to produce vanillin from isoeugenol (**Scheme 3.2**). For this purpose, whole-cell biocatalysis with both engineered enzymes, PROGO and NOV1-S283F since it avoids the cost contribution of cell lysis and enzyme purification (Wang *et al.* 2020; Yao *et al.* 2020) and, at the same time, can protect the expressed enzymes from potentially harmful surroundings (Wachtmeister and Rother 2016). We employed a one-pot approach with further process scale-up and related downstream processing to reduce produced waste (Schmidt, Schallmey, and Kourist 2021). The reaction parameters were optimized by employing 4PG, and RCF lignin oil was used in bioconversion. The resulting vanillin was isolated and characterized, confirming the effectiveness of the developed process for lignin valorization.



Scheme 3.2. The developed whole-cell cascade for vanillin production from 4PG.

3.3. MATERIALS AND METHODS

3.3.1. NOV1 Mutants Design with Rosetta Coupled Moves

Isoeugenol was included in the active site of wild-type NOV1 using Autodock VINA (Trott and Olson 2010). The vanillin present in the X-ray structure (PDB code: 5J55) was used to define the center of the simulation box, comprising 3 Å from the ligand, and then it was manually removed. The best-scored binding mode of isoeugenol at the active site was used as a starting point for Rosetta Coupled Moves experiments (Ollikainen, de Jong, and Kortemme 2015). Three groups of non-catalytic active site residues close to the isoeugenol substrate were selected for mutagenesis (**Fig. 3.2C**); *i*) N120 and T121, *ii*) F281, S283, and F307, and *iii*) F354, L473, and F475. For each group, a Monte Carlo algorithm explored changes over one, two, or three positions simultaneously, resulting in a library of single, double, and triple variants. Repacking was allowed for all the residues in contact with any modified position. A Boltzmann constant of 0.6 and ligand weight of 1.0 were used. Every individual experiment consisted of 100 Coupled Moves runs of 1000 trials. All variants were filtered by protein-ligand docking, retaining the variants that attained pre-catalytic distances (the distances between the two carbons forming the double bond of isoeugenol and the oxygen molecule bound to Ir (*d1*) and the hydroxyl group of the substrates and the catalytic residues Y101 and K134 (*d2*)< 4 Å). The final list of variants was determined by the TotalScore, LigandScore (from Rosetta), and binding energy (from docking) values (**Fig. S3.1,** *Annex I*)).

3.3.2. NOV1 Mutants Construction using Site-directed Mutagenesis

Single, double, or triple amino acid substitutions in the *nov1* gene were constructed using the Quick-Change site-directed mutagenesis protocol

(Stratagene). The plasmid pET-28a (+) containing the *nov1* gene was used as a template using appropriated primers (**Table S3.1,** Annex *I*), except the triple variants F281M_283I_F307H and F281M_S283T_F307H, constructed using the DNA template of variant F281M_S283I and F281M_S283T, respectively. PCRs were performed in 50 μL reaction volumes containing 3 ng of DNA template, 2 μM of primers, 200 μM of dNTPs, NZYProof polymerase buffer, and 1.25 U of NZYProof polymerase (NZYTech). After an initial denaturation period of 5 min at 94 °C, the following steps were repeated for 20 cycles in a thermal cycler (MyCycler thermocycler, Biorad): 1 min at 94 °C, 1 min at 55-58 °C, 10 min at 72 °C, followed by a final 10 min period at 72 °C. The amplified products were purified using GFX PCR DNA and a Gel Band Purification kit (GE Healthcare). The final PCR products were digested with *Dpn*I to eliminate the wild-type template and were used to transform electrocompetent *E. coli* strain DH5α (Novagen) cells. The presence of the desired mutation in the resulting plasmid was confirmed by DNA sequencing.

3.3.3. Activity screenings in 96-well plates

Electroporation introduced the genes coding for the wild-type and variant enzymes into the host strain E. coli BL21 star (DE3, Novagen). In *E. coli* BL21 star, the genes are controlled by the T7 promoter and are induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). Single colonies were picked from a fresh agar plate to 96-well microplates filled with 200 μL of Luria-Bertani medium (LB) supplemented with 50 μg mL^{-1} kanamycin. Cultures were cultivated at 37 °C, 750 rpm for 24 h. The following day, 20 μL of these cultures were inoculated with 180 μL of LB medium in 96-well microplates and cultivated for 4 h at 37 °C. At this time, gene expression was induced with 0.1 mM IPTG; the medium was supplemented with 0.5 mM $FeSO_4$ and cultivated at room temperature at 750 rpm. After 24 h, cells were harvested by centrifugation, re-suspended in 20 mM Tris-HCl, pH 7.6, disrupted by 3 cycles of freezing in liquid nitrogen and thawing at room temperature for 5 min, followed by re-suspension in the same buffer with lysozyme (0.5 mg ml^{-1}). After cell disruption, plates were centrifuged at 4,000 rpm for 30 min at 4 °C, and supernatants (cell crude extracts) were used for enzymatic activity measurements. The activity was assessed after mixing 20 µl of crude extracts with 160 µl of 100 mM Tris-HCl, pH 9.0, containing 1 mM isoeugenol. Reactions were monitored following vanillin production at 340 nm (ε_{340}= 15,97 M^{-1} cm^{-1}).

3.3.4. NOV1 production and purification

The recombinant *E. coli* strains containing plasmids coding for wild-type and S283F were grown in 1 L LB medium supplemented with kanamycin in 5-L Erlenmeyer with 120 rpm shaking (Innova® 44, New Brunswick Scientific). The cells were grown at 37 °C until an optical density of 0.6 at 600 nm was reached, after which 0.1 mM IPTG and 0.5 mM $FeSO_4$ were added to the culture media, and the temperature was reduced to 25 °C. The incubation was continued for a further 16 h, after which the cells were harvested by centrifugation. The cell pellets were suspended in 20 mM sodium phosphate buffer, 0.5 M NaCl, 20 mM Imidazole pH 7.6 (buffer A), containing DNaseI (2 µg mL^{-1}), 5 mM $MgCl_2$, and a mixture of protease inhibitors (CompleteTM mini-EDTA free protease inhibitor mixture tablets; Roche, Basel, Switzerland). Cells were disrupted in a French press cell (at 9,000 psi), followed by centrifugation (18,000 g, 1 h, 4 °C). The resulting soluble extracts were loaded onto an affinity chromatography HisTrap^{TM} column (GE Healthcare) equilibrated with buffer A. Elution was carried out with a linear imidazole gradient (0–0.5 M in 20 min) in the same buffer. All purification steps were performed at room temperature with an Åkta Purifier (GEHealthcare). SDS-PAGE analyzed the purity of eluted fractions. The active fractions were pooled and digested by SUMO-protease (1 U for 5µg of protein), removing the SUMO and His-tag, overnight at 4 °C and then concentrated by ultrafiltration (cut-off of 30 kDa), and equilibrated to 20 mm Tris-HCl, 0.2 M NaCl pH 7.6. The concentration of purified protein preparations was estimated using the molar absorption coefficient of NOV1 (ε_{280} = 81,930 M^{-1} cm^{-1}) calculated from the protein sequence using the ExPASy Bioinformatics Resource Portal (http://web.expasy.org). Iron content of purified protein samples was determined by colorimetric assay resorting to the chelator TPTZ (2, 4, 6-Tris(2-pyridyl)-s-triazine) (Fischer and Price 1964) and by atomic emission spectroscopy at Laboratório Central de Análises, Universidade de Aveiro, Portugal.

3.3.5. Apparent Steady-state Kinetic Analysis

The effect of pH on activity was studied in 100 mM Britton-Robinson buffer (pH 3-11), and the optimal temperature was determined at the temperature range of 10-40 °C. Reactions were performed with isoeugenol (0.01-4 mM) in 100 mM Tris-HCl, pH 9, and monitored at 340 nm (see above). Reactions with resveratrol (0.01-0.32 mM) were monitored following the decrease of absorbance at 306 nm (ε_{306} = 31,800 M^{-1} cm^{-1}) in 100 mM Tris-HCl, pH 9. Enzyme activity was also measured by monitoring the oxygen consumption with an oxygen electrode (Oxygraph; Hansatech) at 25 °C. Reactions were started by adding the enzyme to a nitrogen-purged mixture containing 4 mM isoeugenol in 100 mM Tris-HCl, pH 9. Kinetic data were fitted directly to the Michaelis-Menten equation using the Origin® software. All enzymatic assays were performed at least in triplicate.

3.3.6. Thermodynamic Stability

Protein unfolding was monitored using a Cary Eclipse spectrofluorimeter with an excitation wavelength of 296 nm, and the fluorescence emission of tryptophyl residues was recorded at 340 nm. Guanidine hydrochloride (GdnHCl) concentrations in the range 0–3 M in 20 mM Tris-HCl, 0.2 M NaCl, pH 7.6, were used to induce protein unfolding after incubation at room temperature for 5 min. The fraction of unfolded protein (f_U) was plotted versus [GdnHCl] and fitted according to the equation $f_U = \exp(-\Delta G°/RT)/(1+\exp(-\Delta G°/RT))$. The chemical-induced unfolding was analyzed using the equations previously described based on a two-state process (Fernandes, Martins, and Melo 2009). Thermal unfolding was performed for the iron-depleted enzyme preparations (10 μM) with EDTA for 12 h at room temperature. Before measurement, EDTA was removed by ultrafiltration using Vivaspin^{TM}. Enzyme preparations in 20 mM Tris–HCl, 0.2 NaCl, and pH 7.6, were placed onto a thermostatically thermal block and heated at 1 °C/min up to 100 °C. Thermal unfolding was analyzed using a two-state model $N \rightleftharpoons U$ (Eftink 1998). Static light scattering using excitation and emission wavelengths at 500 nm in the spectrofluorometer was used to follow enzyme aggregation.

3.3.7. Kinetic Stability

Thermal inactivation was studied by incubating enzyme preparations in 20 mM Tris-HCl, pH 7.6 at 25 °C. $FeSO_4$ was added to the solution at a concentration of 100-fold the enzyme molarity when required. At appropriate time points, aliquots were withdrawn and examined for activity. Inactivation constants k_{in} were obtained by linear regression of logarithm activity versus time. The half-life $t_{1/2}$ was calculated using the equation: $t_{1/2} = \ln 2/k_{in}$.

3.3.8. Protein crystallization, Data collection, and X-ray Structure determination and Refinement

The S283F enzyme was crystallized according to the crystallization conditions reported in the literature for NOV1 wild-type (sitting drop vapor diffusion, 0.16 M $MgCl_2$, 0.08 M Tris-HCl pH 8.5, 18% wt/vol PEG 4000, and 20% vol/vol glycerol at 20 °C). For data collection, crystals were flash-cooled in a gaseous nitrogen stream at 100 K. X-ray diffraction data were measured at beamlines of the Swiss-light source (SLS, Villigen-CH). Data processing and scaling were performed using XDS (Kabsch 2010) and the CCP4 package (Collaborative Computational Project 1994) (**Table S3.2,** *Annex I*). The coordinates of the NOV1 bound to vanillin (PDB: 5J55) deprived of all ligands and waters were used as the initial model for structure determination. The program Coot (Emsley *et al.* 2010) was used for electron density inspection and model building, whereas crystallographic refinement was performed with the program REFMAC5 (Murshudov *et al.* 2011).

3.3.9. Molecular Dynamics Simulations

Eight hundred ns trajectories (200 ns per 4 replicas) were run for four different systems: 1) the wild-type NOV1 in its Fe(II)-O_2 substrate-free form, 2) the S283F variant in its Fe(II)-O_2 substrate-free form, 3) the wild-type NOV1 in its *apo* form (no metal and no ligand) and 4) the S283F variant in its *apo* form. The preparation of the NOV1 structures and the MD simulations were carried out using Yasara (Krieger and Vriend 2015). The protein and water solvent were described based on the AMBER14 force field (Hornak *et al.* 2006) and the TIP3P model (Mark and Nilsson 2001). The resulting conformations of the *holo* proteins were used for structural analysis (free-volume analysis and inter-residue interactions) and ensemble docking. Those *apo*proteins were used for pK_a studies on the histidines composing the catalytic center to assess the effect of the mutation S283F on their flexibility and orientation. MDpocket, part of the Fpocket suite (Le Guilloux, Schmidtke, and Tuffery 2009; Schmidtke *et al.* 2010), was used to assess the free volume available inside the core of the protein and active site along with the MD simulation timescale.

3.3.10. Inter-residue interactions

The Protein Interactions Calculator (Tina, Bhadra, and Srinivasan 2007) webserver was used to identify the hydrophobic patches on wild-type NOV1 and S283F PDB structures.

3.3.11. Ensemble Protein-Ligand Docking

The isoeugenol and resveratrol substrates were systematically docked every 40 ps along with the MD simulations of the ligand-free systems (20000 structures for each system) using a simulation box of 8 Å from the oxygen molecule. 96 docking runs and a clustering cut-off of 2 Å of RMSD were established for each MD snapshot. The protein was kept rigid while the rotors of the substrate were allowed to move. The distances between the two carbons forming the double bond of the ligands and the oxygen molecule bound to Fe (*d1*), the hydroxyl group of the substrates and the catalytic residues Y101 and K134 (*d2*) were measured. The dockings were performed with AutoDock VINA (Trott and Olson 2010) using the YAMBER force field (Krieger *et al.* 2004), and the structural analysis was done with the Yasara Python module. For all the binding modes identified as pre-catalytic structures (the distances *d1* and *d2* were <4 Å), both the $\Delta G_{binding}$ (it estimates the quality of ligand-protein interactions) and $\Delta E_{dihedral}$ (it measures how strong the distortion of the ligand is in comparison to its protein-free form) energies were assessed.

To dock *E*- and *Z*-isoeugenol into the S283F variant, we downloaded the PDB-REDO version3 of NOV1 wild-type in complex with vanillin (PDB ID 5J55). This structure was deemed more suitable for computation than pdb 7QR6 (NOV1-S283F) due to its higher resolution, the presence of the vanillin product, and the absence of electron density for F59 in 7QR6, and because the overall structure of the enzyme was essentially identical (Cα RMSD = 0.48Å). We used the YASARA molecular modeling program (Krieger, Koraimann, and Vriend 2002) to prepare the structure by fixing common issues (YASARA commands Clean and OptHydAll), followed by energy minimization with vanillin, oxygen, iron, and the iron-chelating histidine side chains frozen. Next, Ser283 was swapped for Phe (SwapRes), the residue was optimized with YASARA's OptimizeRes using the SCWALL method (Canutescu, Shelenkov, and Dunbrack 2003; Krieger *et al.* 2009) and another minimization was performed, now with vanillin kept flexible. Finally, the aldehyde oxygen of vanillin was swapped for an ethyl group to convert vanillin into isoeugenol (setting dihedral angles such that the molecule was either in trans or cis configuration) and the complex was energy minimized once more. These complex structures were used for an initial inspection to identify active-site residues involved in substrate binding. Subsequently, we removed the ligand, froze the whole structure except for the side-chains of residues F59, M216, F283, F307, F354, and L475, and defined a simulation box 2 Å around all non-frozen atoms. Using default parameters, YASARA was then used to execute 999 docking runs with VINA (Trott and Olson 2010).

3.3.12. Bioconversions of Isoeugenol to Vanillin

Conversion of isoeugenol to vanillin was studied in 50 mL-Erlenmeyer with a reaction volume of 10-mL containing 10, 25, 50, and 100 mM isoeugenol in 0.1 M glycine-NaOH, pH 9, at room temperature, 150 rpm. After adding 1 U mL^{-1} of purified wild-type and S283F variant (defined as the amount of enzyme that produced 1 mol of vanillin per min) and whole-cells of *E. coli* overproducing the variant S283F at OD_{600} = 2, which corresponds to 1.3 mg cell dry weight ml^{-1}. Reactions containing up to 3.5% ethanol were used to prepare the substrate's stock solution (1-2 M isoeugenol in 70% ethanol). The progress of reactions was monitored at 340 nm by quantifying vanillin production by absorbance (see above) and HPLC (see below).

3.3.13. Bioconversion of 4PG to Vanillin by Coupled Reactions

E. coli NEB 10β cells harboring PROGO and *E. coli* BL21AI cells harboring NOV1-S283F were used, resulting in the expression of PROGO and NOV1-S283F after induction by L-arabinose and IPTG, respectively. The incubation continued until final OD_{600} values were above 10 and 5 for PROGO and NOV1-S283F, respectively, after which the cells were harvested and resuspended in potassium phosphate (KPi) buffer (50 mM, pH 7) to yield the desired OD before being used as biocatalysts.

Unless stated otherwise, all small-scale conversions (250 μL) were performed in 20 mL glass vials with 50 mM KPi buffer, pH 7.0, at 150 rpm. Both individual conversions on targeted substrates were tested with cells with various OD values at 25 °C. Subsequently, the temperature in 3 gradients was applied to the conversion. P cells were fixed at OD 50 and mixed with N cells to initialize the one-pot conversion (250 μL). In comparison, the step-wise approach of 4PG conversion was separated by the later addition of N cells and doubled the reaction volume (500 μL) upon adding the N cells.

3.3.14. HPLC analysis

Bioconversions of isoeugenol to vanillin by pure NOV1 wild-type and S283F were analyzed in a Waters Alliance 2695 HPLC System with a Purospher STAR RP-18c column (125 × 4 mm), 5 µm particle size (Merck KGaA, Germany). The column was maintained at 40 °C, the flow rate was set to 1.0 mL min$^{-1,}$ and the volume injected was 25 µL. A linear gradient was set starting from 70% of 0.5% acetic acid (solvent A) and 30% methanol with 0.5% acetic acid (solvent B), going from 0 to 80% solvent B in 25 min and maintaining for 10 min. After this, initial conditions were resumed in 2 min and maintained for 8 min. The absorption was monitored between 200 and 500 nm by a Waters Photodiode Array Detector 2996 operated by Empower Pro, version 5, 2002 (Waters Chromatography). A time-course investigation of whole-cell biocatalysis was analyzed. In 20 mL glass vials, the samples were quenched with acetonitrile at a 20-fold dilution. After centrifuging the mixture, the supernatants were analyzed using HPLC. All chemicals were separated with a gradient control method and calibrated using standards on the HPLC. Buffer A is 12 mM KPi (pH 7), and Buffer B is pure acetonitrile. Control method: 25% to 95% gradient acetonitrile for 12 min, remaining 95% acetonitrile for 0.5 min and following 95% to 25% gradient acetonitrile decrease for 5.5 min.

3.3.15. Isolation of Vanillin and NMR analysis

After 48 h, the reaction mixture was spun down, and the supernatant was separated from the biomass by decanting. Firstly, the supernatant was extracted with ethyl acetate, followed by the extraction of the precipitate. To facilitate the collection of organic layers from the mixture with biomass, each extraction step was followed by spinning down the mixture. Combined organic layers were washed with water, saturated solution of NaCl and dried over anhydrous $MgSO_4$. The organic layer was concentrated in vacuo, and the residue was purified by column chromatography using Merck 60 A 230-

400 mesh silica gel, and solvent composition hexane/Ethyl acetate = 9/1 followed by 7/3. Combined fractions were evaporated, and NMR analyzed the residue. The yield was ~40%, and NMR analysis indicated the presence of water and impurities that probably originated from lignin oil (no isoeugenol or 4PG were detected). Water was added to recrystallize the remaining solid, and the flask was heated until the compound was completely dissolved. The water layer was separated from the oil by pipetting it into another clean flask. The remaining solution was allowed to cool down slowly at room temperature, leading to vanillin crystals formation. The crystals were isolated by filtration and dried. The isolated yield was ~10%, and NMR indicates pure vanillin with some traces of water. The identity of products was determined using NMR. NMR data were collected on a Varian MercuryPlus (^{1}H at 400 MHz) equipped with a 400 Autosw probe. Chemical shifts are reported in parts per million (ppm) relative to the residual solvent peak ($CDCl_3$: ^{1}H: 7.26 ppm). Coupling constants are reported in Hertz (Hz). Multiplicity is reported with the usual abbreviations (s: singlet, d: doublet, dq: doublet of quartets, dd: doublet of doublets, m: multiplet).

3.4. RESULTS AND DISCUSSION

3.4.1. Active Site Design: Construction and Characterization of Variants

The analysis of vanillin and resveratrol binding to NOV1 (PDBs 5J55 and 5J54) suggested that isoeugenol can be docked into the NOV1 active site with the hydroxyl group pointed towards Y101 and K134 residues, and the methoxy substituent, to the pocket created by N120, T121 and L475 (**Fig. 3.2A-C**, Region A). In this position, the isoeugenol aromatic ring plane sits parallel to the side chain of F59. Moreover, the reactive double bond of isoeugenol is located right above the iron center. In contrast, the terminal methyl group is hosted in a large and hydrated niche in the rear of the cavity close to F281, S283, F307, and F354 (**Fig. 3.2A-C**, Region B). The alignment of NOV1 with *P. nitroreducens* isoeugenol monooxygenase shows a phenylalanine residue at the position of S283 in this enzyme (Yamada *et al*. 2007).

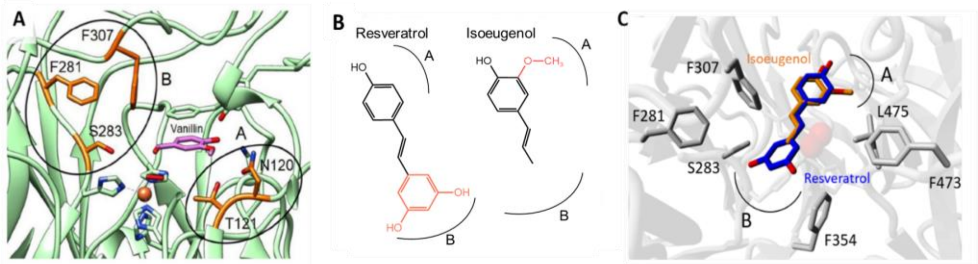


Fig. 3.2. Structural Basis for Rational Design of NOV1. (A) NOV1 active site in complex with vanillin (PDB code 5J55) with surrounding residues highlighted. (**B**) Resveratrol and isoeugenol chemical structures. The regions at which the different chemical groups of the ligands bind to the active site are indicated. (**C**) Comparison between the binding of isoeugenol and resveratrol (PDB code 5J54) to the NOV1 binding site. Regions A and B outline the amino acids binding the isoeugenol hydroxyl propenyl groups.

Furthermore, the X-ray crystal structure analysis supported that S283F replacement could improve isoeugenol catalysis by fulfilling the free space left in the active site upon binding the single-ring isoeugenol substrate. Mutations N120L, F281W, S283Q, F307H, F307W, and L475S were suggested by sequence comparisons and biochemical information from the literature (Zhao *et al*. 2018; McAndrew *et al*. 2016; Yamada *et al*. 2007; Ryu *et al*. 2013). Positions 120, 121, 281, 283, 307, 354, 473, and 475 were submitted to mutagenesis using Rosetta and filtered by protein-ligand docking; see the computational methods section for further details. Thirty-five variants based on the rational and computational approaches were constructed using site-directed mutagenesis and characterized (**Table 3.1**). The enzymatic activity of isoeugenol was tested in cell crude extracts in 96-well microplates. Results showed that mutations in positions 120, 121, 473, and 475 impaired the isoeugenol activity. Three double variants, F281M_S283T, F281M_S283V, and F281C_S283I, showed specific activities comparable to the wild-type, and the single variant S283F stood out with almost 2-fold higher enzymatic activity as compared to the wild-type.

Table 3.1. Mutant design and enzymatic activity. Reactions were performed using cell-crude extracts in 0.1 M Tris-HCl, pH 9, 1 mM isoeugenol at room temperature.

Enzyme	**Rational design**	**Initial Activity (i_A) (U/mL)**	**Relative activity to wild-type ($i_{AVariant}/i_{AWT}$)**
Wild-type		0.252 ± 0.027	-
N120L	*(Zhao, 2018)*	nd	nd
N120S	Rosetta + docking	0.019 ± 0.006	0.07 ± 0.03
T121P	Rosetta	nd	nd
T121A	Rosetta + docking	0.052 ± 0.009	0.18 ± 0.03
F281W	*(Ryu, 2013)*	0.047 ± 0.011	0.23 ± 0.05
S283F	*(Yamada, 2007)*	0.433 ± 0.082	1.50 ± 0.29
S283Q	*(Zhao, 2018)*	0.031 ± 0.005	0.15 ± 0.02
S283I	Rosetta + docking	0.044 ± 0.019	0.18 ± 0.08
S283E	Rosetta + docking	0.040 ± 0.010	0.17 ± 0.04
S283N	Rosetta + docking	nd	nd
S283V	Rosetta + docking	0.118 ± 0.027	0.61 ± 0.14
S283C	Rosetta + docking	0.012 ± 0.004	0.05 ± 0.02
S283T	Rosetta + docking	0.063 ± 0.018	0.26 ± 0.02
F307H	Visual inspection	0.017 ± 0.010	0.07 ± 0.04
F307W	(McAndrew, 2016)	0.030 ± 0.006	0.10 ± 0.02

F309W	Rosetta + docking	0.016 ± 0.003	0.05 ± 0.01
F473W	Rosetta + docking	0.020 ± 0.005	0.07 ± 0.02
F473E	Rosetta + docking	0.018 ± 0.004	0.06 ± 0.01
F473Q	Rosetta + docking	0.014 ± 0.005	0.06 ± 0.02
L475S	*(Zhao, 2018)*	0.011 ± 0.004	0.04 ± 0.02
L475G	Rosetta	0.017 ± 0.004	0.07 ± 0.02
L475T	Rosetta + docking	0.016 ± 0.003	0.05 ± 0.01
E353D	Rosetta + docking	0.013 ± 0.006	0.05 ± 0.03
F281C S283I	Rosetta + docking	0.283 ± 0.055	1.02 ± 0.18
F281C S283T	Rosetta + docking	0.015 ±0.005	0.05 ± 0.02
F281I S283V	Rosetta + docking	0.012 ± 0.004	0.05 ± 0.02
F281H S283I	Rosetta + docking	0.193 ± 0.043	0.70 ± 0.14
F281H S283V	Rosetta + docking	0.011 ± 0.003	0.04 ± 0.02
F281M S283I	Rosetta + docking	0.273 ± 0.048	0.98 ± 0.16
F281M S283T	Rosetta	0.282 ± 0.061	1.02 ± 0.20
F281M S283V	Rosetta + docking	0.252 ± 0.023	0.90 ± 0.09
F473W L475G	Rosetta + docking	0.022 ± 0.004	0.09 ± 0.02
F473W L475T	Rosetta + docking	0.017 ± 0.003	0.07 ± 0.03
F281M S283I F307H	Rosetta + docking	0.038 ± 0.015	0.20 ± 0.10
F281M S283T F307H	Rosetta + docking	0.014 ± 0.007	0.04 ± 0.03

nd – not detected

3.4.2. Biochemical and Kinetic Characterization

Wild-type and variant S283F were produced at an Erlenmeyer scale and purified. Both enzymes showed an optimal pH of 9, an optimal temperature of around 28 °C for the wild-type enzyme, and 32 °C for the S283F variant (**Fig. 3.3A, B**). Purified enzyme preparations displayed approximately 0.5 mol of iron per mol of protein. They were partially iron-depleted, similar to the observed in isoeugenol oxygenases from *P. putida* (Yamada *et al.* 2007) and *P. nitroreducens* (Ryu *et al.* 2013). The steady-state kinetic analysis revealed that the S283F variant features a 2-fold higher k_{cat} than the wild-type, and a slightly higher K_m value for isoeugenol, leading to a catalytic efficiency (k_{cat}/K_m) comparable to the wild-type (**Table 3.2**). The activity using resveratrol as substrate was also investigated; the S283F variant showed a 5-fold lower k_{cat} and 4-fold higher K_m, resulting in a sharp 20-fold decrease of the k_{cat}/K_m, compared to the wild-type. Notably, the kinetic analysis for O_2 in the presence of isoeugenol showed that the S283F displayed an enhanced binding to molecular oxygen, with a ~2-fold lower K_m and a 4-fold higher catalytic efficiency than the wild-type (**Table 3.2**). This is an important asset to overcome the usually limiting levels of soluble O_2 in large-scale industrial processes; indeed, enhancing O_2 binding and catalysis has been a critical challenge in the application of oxygenases (Garcia-Ochoa *et al.* 2010; Sigurdardóttir *et al.* 2018). The results supported the prediction that introducing the bulky and hydrophobic phenylalanine at position 283 impairs resveratrol oxidation, shifting the specificity of NOV1 towards smaller substrates such as isoeugenol and molecular oxygen.

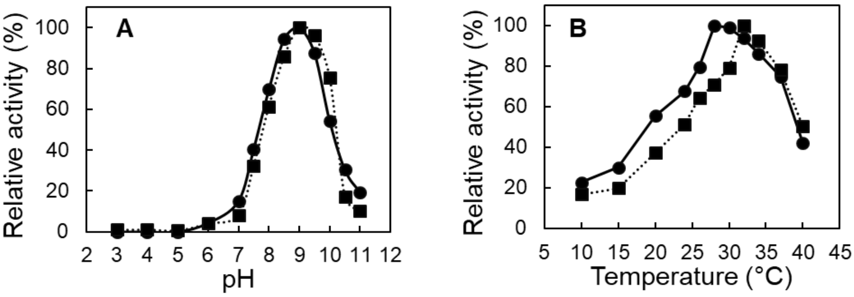


Figure 3.3. pH and temperature profiles. (A) pH activity profile of wild-type (circles) and S283F NOV1 (squares). Reactions were performed in Britton-Robinson buffer (in the pH range from 3 to 11) in 1 mM isoeugenol at room temperature. **(B)** Temperature dependence of enzymatic activity in reactions performed in 100 mM Tris-HCl buffer, pH 9.

Table 3.2. Apparent steady-state kinetic parameters of wild-type and S283F NOV1 for isoeugenol, molecular oxygen (in 4 mM isoeugenol), and resveratrol. Kinetic assays were performed at room temperature in 0.1 M Tris-HCl, pH 9.

		Isoeugenol	**O_2 (isoeugenol)**	**Resveratrol**
wild-type	k_{cat} (s^{-1})	7.3 ± 0.2	10.9 ± 0.8	0.38 ± 0.02
	K_m (mM)	0.6 ± 0.1	$(0.7 \pm 0.1) \times 10^{-2}$	0.06 ± 0.01
	k_{cat}/K_m ($M^{-1}\ s^{-1}$)	$(12.2 \pm 2) \times 10^3$	$(15.6 \pm 4) \times 10^5$	$(6.3 \pm 0.2) \times 10^3$
S283F	k_{cat} (s^{-1})	14.5 ± 0.8	17.3 ± 0.6	0.078 ± 0.004
	K_m (mM)	0.8 ± 0.1	$(0.29 \pm 0.04) \times 10^{-2}$	0.227 ± 0.002
	k_{cat}/K_m ($M^{-1}\ s^{-1}$)	$(18.1 \pm 4) \times 10^3$	$(59.7 \pm 10) \times 10^5$	$(0.34 \pm 0.02) \times 10^3$

3.4.3. NOV1 Stability and Iron Incorporation

The kinetic or operational stability is relevant to assess the biocatalyst performance under specific operating conditions, for example, at a given temperature, and to study pathways that lead to the formation of irreversibly inactivated enzyme states (Sanchez-Ruiz 2010). The thermal inactivation assays revealed that the S283F amino acid replacement drastically improves the enzyme kinetic stability. While the wild-type NOV1 displayed inferior stability with a half-life ($t_{1/2}$) of around 1 h at 25 °C, the variant S283F exhibited a half-life of 29 h (**Fig. 3.4**). Notably, incubation of wild-type with iron increased the half-life 10-fold but not of the variant, indicating that *i*) binding of the iron cofactor seems the critical determinant of thermostability of NOV1 enzyme, and *ii*) the wild-type enzyme loses iron more quickly than the S283F variant. The chemical unfolding of wild-type and S283F was assessed using fluorescence emission of tryptophan residues.

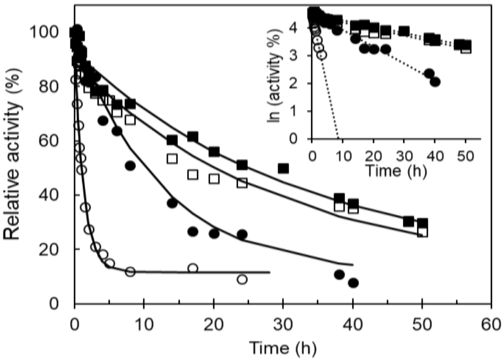


Figure 3.4. Kinetic stability. Stability of wild-type (circles) and S283F (squares) NOV1 at 25 °C in the absence (open symbols) and the presence of 100 equivalents of $FeSO_4$ (closed symbols). Inset: linear regression of logarithm activity versus time. Without iron, the half-time at 25 °C was 1.3 ± 0.2 h and 29 ± 3.4 h for wild-type and S283F variant, respectively. The addition of iron increased 10-fold the half-life time (11.4 ± 0.7 h) of the wild-type, whereas the stability of S283F remained similar (30.4 ± 1.2 h).

The folded and unfolded states were the only states that accumulated in significant amounts, and a two-stage process accurately fit the unfolding process (**Fig. 3.5A**), and both enzymes displayed similar stability; the guanidinium hydrochloride mid-point concentration at around 0.6 M (where 50% of molecules are unfolded), and the native state free energy is 3.5 kcal mol^{-1}. In thermal unfolding experiments, the fluorescence emission from tryptophan residues increased in the 30-40 °C range (**Fig. 3.6**), indicating the iron cofactor's release, a known fluorescence quencher (Dunning Hotopp *et al.* 2003). Incubation of wild-type with 2000 molar equivalents of EDTA abolished this effect (**Fig. 3.6A**), in contrast to S283F, which shows a persisting "iron quenching effect" even after incubation with EDTA, suggesting a significantly higher affinity for iron (**Fig. 3.6B**), in line with the higher half-life values of kinetic stability. The apparent melting temperatures (T_m) are very similar in both enzymes at 57-59 °C (**Fig. 3.5B**); static light scattering at 500 nm revealed a highly-aggregation tendency with an onset of aggregation (T_{agg}) at 47 °C before enzyme unfolding (**Fig. 3.5C**). These results showed that both enzymes share a relatively high thermal robustness of the enzymes' native state. The higher kinetic thermostability observed in the S283F variant compared to wild-type most likely results from an enhanced stabilization of iron cofactor inside the catalytic cavity.

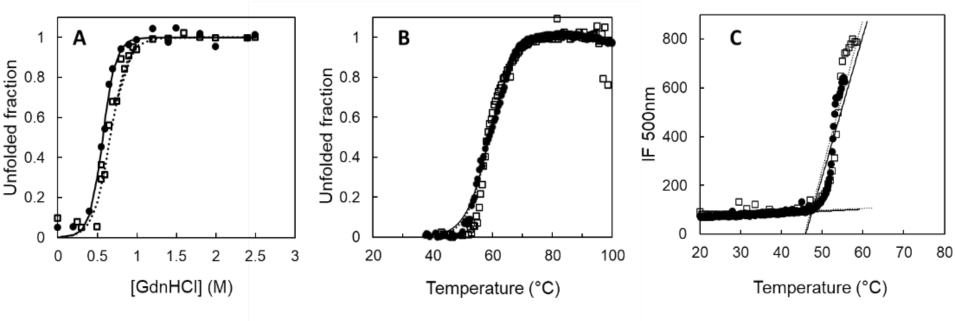


Figure 3.5. (A) NOV1 wild-type (closed circles) and S283F (open squares) fraction unfolded by Gdn-HCl as measured by fluorescence emission of tryptophyl residues at 340 nm. Measurements were performed by reading fluorescence at excitation wavelengths of 296 nm and emission wavelengths of 340 nm. The solid line is the fit according to the equation $f_U = \exp(-\Delta G^0/RT)/1+\exp(-\Delta G^0/RT)$, which assumes the equilibrium $N \rightleftharpoons U$ (Fernandes, Martins, and Melo 2009). **B)** Thermal unfolding following fluorescence emission of tryptophan residues at 340 nm (T_m = 57-59 °C) for wild-type (filled circles) and S283F variant (empty squares). **C)** Static light scattering at 500 nm (T_{agg} = 47 °C) for wild-type (filled circles) and S283F variant (empty squares).

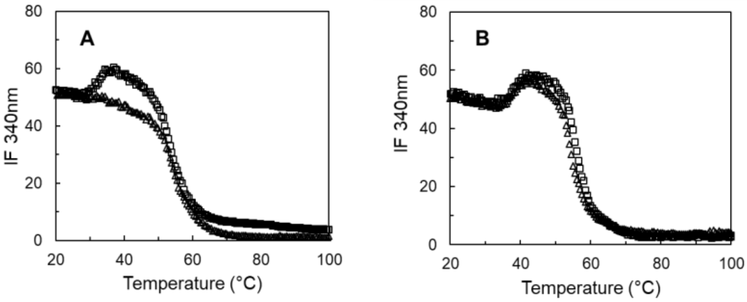


Figure 3.6. Tryptophan fluorescence emission at 340 nm at increasing temperatures of wild-type **(A)** and S283F variant **(B)** in the absence (squares) and after incubation with 2000 equivalents of EDTA (triangles).

3.4.4. Structural Characterization of the S283F Variant

The crystal structure of NOV1-S283F was solved at 2.9 Å using synchrotron radiation (**Table S3.2,** *Annex I*), revealing the recognizable electron density of a phenylalanine side chain at position 283 (**Fig. 3.7A-C**). The aromatic ring of F283 was found to occupy the active site, cluttering the catalytic cavity. The Fe(II)-O_2 complex, required for activity, is coordinated by four histidine residues: H167, H218, H284, and H476 (**Fig. 3.6B, C**). F283 is predicted to reach out and interact with isoeugenol through hydrophobic contacts (**Fig. 3.1C**), whereas K134 and Y101 create the proper hydrogen bonding environment with the 4-hydroxy group of the substrate (**Fig. 3.6C, D**).

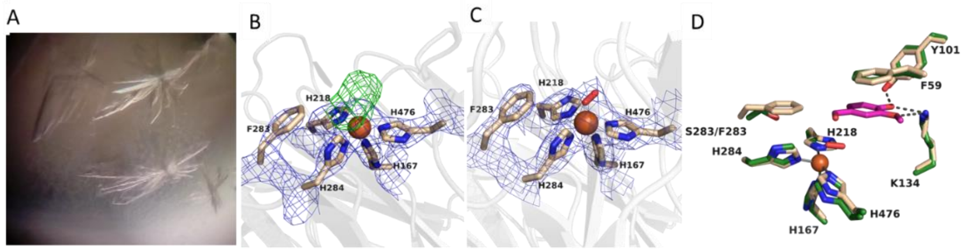


Figure 3.7. (**A**) Crystals of S283F NOV1 variant. (**B**) Weighted 2Fo-Fc electron density of the active site. The contour level is 1.2 σ. The side chain of F283 is well recognizable with its aromatic ring close to the His-coordinated iron. The difference Fourier Fo-Fc map (contoured at 3.0 σ level; green) showed a residual electron density interpreted as bound oxygen. (**C**) The oxygen-bound active-site structure of the final model. (**D**) S283F NOV1 crystal structure with the reaction product (vanillin, carbons in purple) modeled in the cavity. The model used the complex structure between the wild-type enzyme and vanillin as a reference (carbons in green, PDB code 5J55). F283 interacts with the edge of the substrate ring, whereas F59 is involved in a π–π stacking with the aromatic ring of the substrate.

3.4.5. Molecular Dynamics Simulations and Substrate Docking

Starting from the experimental wild-type (PDB code 5J55) and S283F structures (PDB code 7QR6), the computational analysis showed that a comparable number of interactions is present in both wild-type and S283F enzymes and the active site cavity along the MD trajectories (800 ns), which remained in a relatively stable conformation with a constant volume in both structures (**Fig. S3.2**). However, S283F replacement resulted in local re-arrangement of a sizeable hydrophobic core present in wild-type to create small hydrophobic patches near the active site (**Fig. S3.3**); the average cavity volume is expanded in S283F. Consequently, a higher number of water molecules was identified inside the active site of the mutant (**Fig. S3.4**). The more expanded shape of the S283F active site could make the catalytic center more accessible to the solvent and, by extension, other small molecules, such as molecular oxygen and isoeugenol. Furthermore, the presence of the bulky phenylalanine at the active site seems to distort the resveratrol conformation significantly when approaching pre-catalytic states (**Fig. 3.11A, B** and **Table S3.3,** *Annex I*). In contrast, the distortion of the isoeugenol structure when approaching pre-catalytic binding modes appeared only slightly higher in the variant than in the wild-type. These data support the experimental observation that resveratrol is a poorer substrate for the variant than wild-type (4-fold higher K_m and 5-fold lower k_{cat}). Furthermore, MD trajectories of the ligand-, oxygen- and iron-free systems of wild-type showed that the four iron-binding histidines are kept in a stable conformation that is not compatible with iron coordination, particularly for H167 and H218 (**Fig. 3.12**). This was different from S283F, where the presence of phenylalanine in the iron neighborhood increased the fluctuation of the His residues, which visited conformations compatible with iron-binding (**Fig. 3.12**), indicating a lower relative transition energy barrier between iron-free and iron-bound forms in S283F. This increased flexibility of the iron-coordinating histidines could be related to alterations in the H-bond network of the active site, where three conserved carboxylate residues (E135, E353, and E418) are responsible for stabilizing H218, H284, and H476 (McAndrew *et al.* 2016; Daruwalla and Kiser 2020). In addition, lower pK_a values for all histidine side chains, particularly H167 and H218, were found in the S283F variant (**Fig. 3.11C**), suggesting that their dissociation equilibrium favors the unprotonated state required to bind the iron metal. These electrostatic calculations support a higher affinity of iron to the S283F variant with kinetic stability and thermal unfolding/EDTA chelation, which showed that wild-type loses iron more readily than S283F (**Fig. 3.4** and **3.6**).

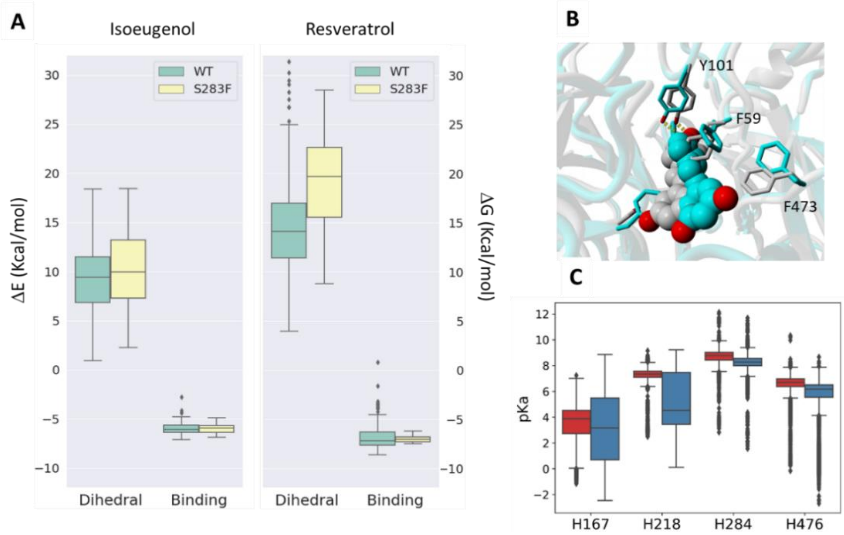


Figure 3.11. (A) Dihedral (ΔE) and binding energies (ΔG) were estimated for ensemble docking of isoeugenol and resveratrol to wild-type NOV1 and S283F variants. $\Delta E_{dihedral}$ measures how strong the ligand distortion is compared to its protein-free form, $\Delta G_{binding}$ estimates the quality of ligand-protein interactions. **(B)** Representation of resveratrol binding to wild-type NOV1 (grey) and the S283F variant (blue). The substrate is represented as balls, and the residues as sticks. **(C)** pK_a of iron-coordinating histidines 167, 218, 284, and 476 were measured every 40 ps along the 800 ns MD trajectory for the *apo* form of the wild-type (red) and the S283F mutant (blue).

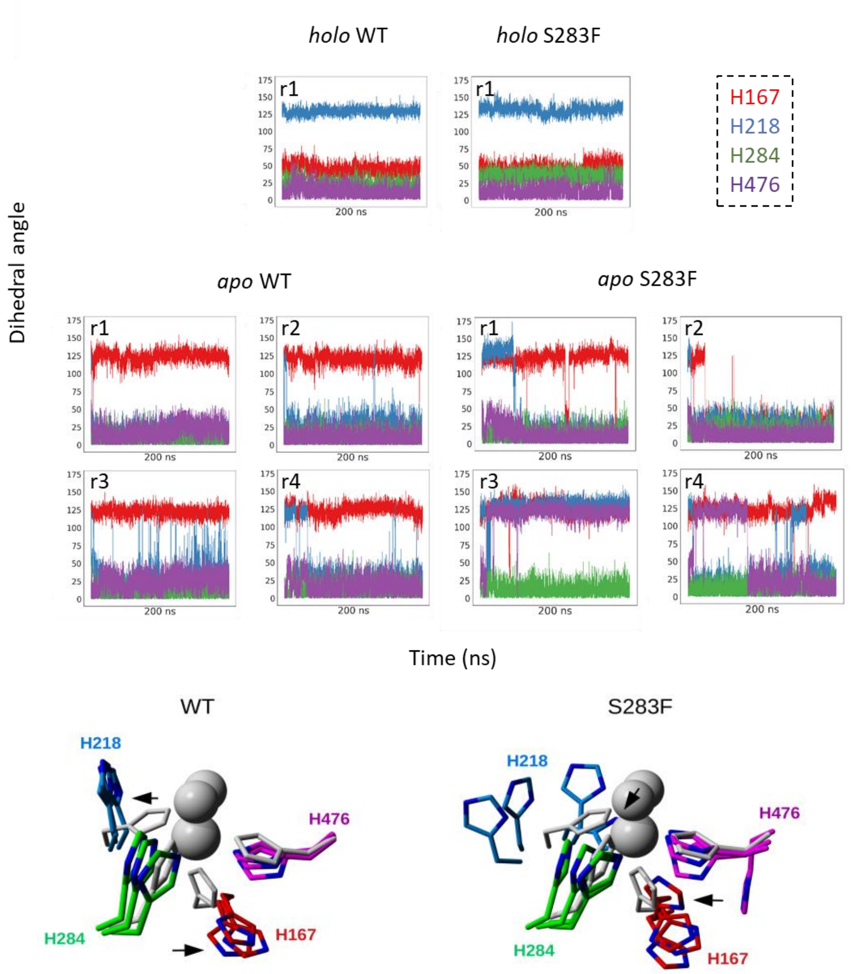


Figure 3.12. On the top, dihedral angles [C-CA-CB-CG] define the orientation of the histidine 167 (red), 218 (blue), 284 (green), and 476 (purple) side chains, measured on the four replicas of 200 ns MD simulation. The values for the wild-type and mutant in their *holo* (Fe-bound) and *apo* (Fe-free) forms are represented. For simplicity, only one replica is shown for the *holo* form (which is consistent for the four replicas). The bottom illustrates the most representative clusters concerning the histidine conformations, along with the MD simulation of the iron-free form of both wild-type and mutant. For comparison, these appear superposed with the catalytic center of the Fe(II)-O_2 bound state as it appears in the 5J55 X-ray structure.

3.4.6. Enzymatic Vanillin Bio-Production

Bioconversion assays confirmed that purified S283F converts isoeugenol to vanillin more efficiently than wild-type: >99% conversion of 10 mM isoeugenol after 1.5 h of reaction was achieved, which favorably compares with the 67% conversion obtained using the wild-type enzyme (**Fig. 3.13A**). The reaction product was confirmed to be vanillin by HPLC following the previous identification (McAndrew *et al.* 2016). Further addition of isoeugenol (10 mM) resulted in 96% and 74% conversion yields after 24 h for S283F and wild-type, respectively. The total turnover number (TTN), defined as the total moles of vanillin produced per mole of enzyme over the entire length of the reaction, is 5.1 and 2.2 for S283F and wild-type, respectively. Remarkably, the k_{cat} of the S283F variant for isoeugenol is 2 to 10 times higher than the values previously reported for isoeugenol oxygenases (Han, Long, and Ding 2019; Yamada *et al.* 2007; Ryu *et al.* 2013; Zhao *et al.* 2018) further endorsing this enzyme as a superior biocatalyst in non-limiting substrate concentrations, typical of industrial setups. We set up time-course bioconversion assays using whole cells that overproduced the S283F variant to reduce the costs associated with enzyme purification (**Fig. 3.13B, C**) (Lin and Tao 2017; Fernandes *et al.* 2021). Our results revealed that excellent molar conversion yields, >99%, were achievable within 24 h of reaction using concentrations of isoeugenol up to 100 mM, in the presence of small amounts (3.5%) of ethanol, a bio-solvent, in the whole-cell catalysis mixture. Notably, the obtained conversion yields were similar to those obtained with *Pseudomonas* isoeugenol monooxygenases in the presence of organic solvents or enzyme aggregates (Zhao *et al.* 2019; Yamada *et al.* 2008) and higher than those obtained with oxygenases from *Herbaspirillum seopedicae* and *Rhodobacteraceae bacterium* (**Table 3.3**) (Han, Long, and Ding 2019) Furthermore, the NOV1 system tolerated concentrations of substrate up to 100 mM in contrast to other enzymatic systems inhibited by isoeugenol, and force reactions to be performed at significantly lower concentrations of substrate (Han, Long, and Ding 2019; Furuya *et al.* 2015).

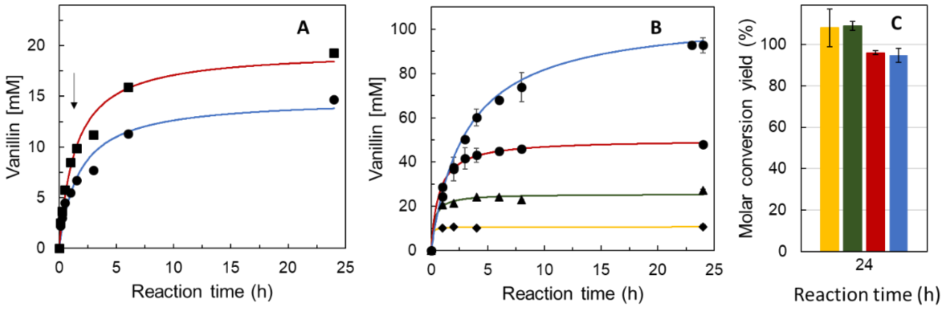


Figure 3.13. Bioconversions of isoeugenol to vanillin. **(A)** Time course of vanillin production using 1 U mL^{-1} of wild-type (blue) and S283F NOV1 (red) purified enzymes; reactions started with 10 mM of the substrate, and additional supplementation with 10 mM of isoeugenol occurred after 1.5 h of reaction (arrow). **(B)** Time course of vanillin production using recombinant *E. coli* whole-cells overproducing S283F NOV1 variant (final OD_{600} = 2) in reactions containing initial concentrations of 10 (yellow), 25 (green), 50 (red) and 100 (blue) mM isoeugenol. **(C)** Molar conversion yields after 24 h in reactions containing initial concentrations of 10 (yellow), 25 (green), 50 (red), and 100 (blue) mM isoeugenol. Reactions were performed in 0.1 M glycine-NaOH buffer, pH 9, at room temperature, 150 rpm.

Table 3.3. Comparison of performance of whole-cell reactions in the conversion of isoeugenol to vanillin.

Enzyme	**Cell dry weight (mg/mL)**	**Organic solvent % (v/v)**	**Substrate (mM)**	**Conversion (%)**	**Ref.**
NOV1-S283F (*N. aromatocivorans*)	1.3	3.5% Ethanol	100	>99	This study
IEM *(P. putida)*	15	10% DMSO	230	81	Yamada *et al.* 2008
IEM720-F281Q (metagenome)	34	10% DMSO	100	75	Zhao *et al.*, 2018
*Hs*CCO (*H. seropedicae*)	18	10% glycerol	10	80	Han *et al.*, 2019
*Rb*CCO (*R. bacterium*)	18	10% glycerol	10	75	Han *et al.*, 2019

3.4.7. Activity of Individual Whole-Cells toward 4PG and Isoeugenol

We investigated the feasibility of a one-pot conversion of 4PG to vanillin, utilizing *E. coli* NEB 10β cells harboring PROGO (P cells) and *E. coli* BL21-AI cells containing NOV1-S283F (N cells). The number of cells varied to determine the optimal amount for reaching a complete or close to complete conversion. After 24 h, a >95% conversion of 50 mM 4PG was achieved using P cells at an OD_{600} of 75 (**Fig. 3.14A**). Lower amounts of P cells resulted in less efficient yet considerable conversions, indicating that lowering the OD_{600}, alongside optimization of other reaction conditions (temperature, pH, reaction time), could lead to satisfactory yields. The N cells showed activity on commercially available *E*-isoeugenol already at OD_{600} 2. Production of vanillin yielded approximately 85% when the OD_{600} value was above 10 (**Figure 3.14B**). Despite increasing the amounts of N cells to OD_{600} 50, the vanillin yield remained at around 85%.

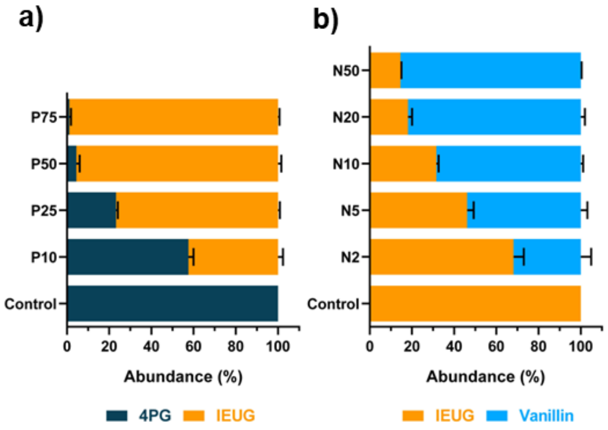


Figure 3.14. Conversions of 50 mM substrate of **(A)** 4PG by P cells at different OD_{600} values (10, 25, 50, and 75); and **(B)** isoeugenol (IEUG) by N cells at different OD_{600} values (2, 5, 10, 20, and 50). Reactions were performed in 50 mM KPi buffer, pH 7 in 10% (v/v) DMSO at 25 °C for 24 h.

Previous reports indicate that P cells catalyze the formation of a mixture of *E*- and *Z*-isoeugenol (Guo *et al.* 2022). To anticipate the actual reaction conditions of the cascade, we followed the depletion of a mixture of isoeugenol isomers (*E*/*Z*=75/25) catalyzed by N cells over time. We hypothesized that the ratio of unreacted isoeugenol, remaining after the cascade, should remain constant if the N cells are equally active on both isomers. The NMR analysis of the crude reaction mixture at different time intervals points toward the preference of NOV1-S283F for the *E*-isomer, as the ratio of *E*/*Z* changes from 3 to 0.8 after 24 h (**Fig. 3.15A**). To further support this evidence, a docking study using the crystal structure of NOV1-S283F revealed a minimal clashing between the CH_3 moiety of the alkyl chain of the *Z*-isomer and F283 (**Fig. 3.15B**). It appears that *Z*-isoeugenol can adapt within the active site of the enzyme, thus being still a possible substrate but with hampered reactivity when compared to the *E*-isomer. Thus, we expected that the conversion of both isomers would occur; however, it would be more readily on *E*-isoeugenol, which might affect the final vanillin yield.

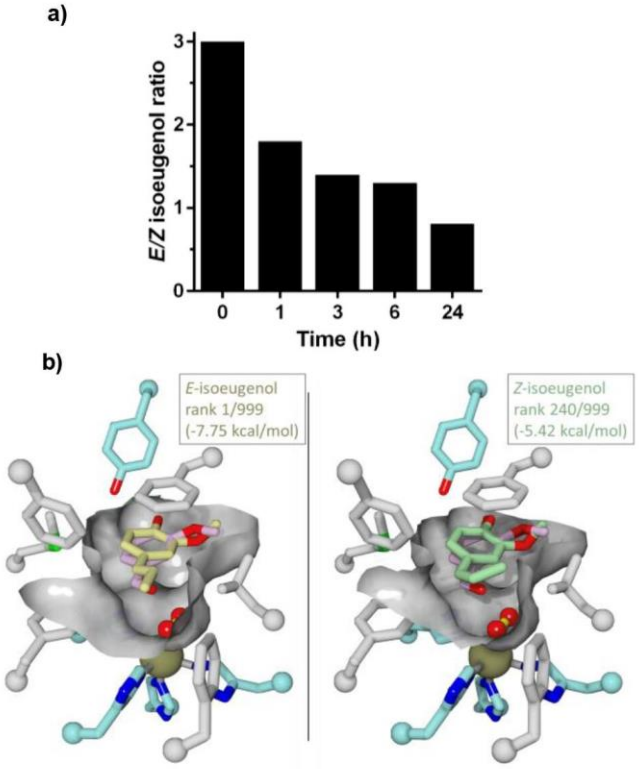


Figure 3.15. Substrate preference of NOV1-S283F. **A)** NMR analysis of the whole-cell conversion of 50 mM *E*/*Z*-isoeugenol (starting ratio *E*/*Z*=75/25). N cells (OD_{600} 10) were used in 50 mM KPi buffer, pH 7, with 10% (v/v) DMSO as a cosolvent, at 25 °C; **B)** Docking result for NOV1-S283F with the *E*- (pale yellow left) and *Z*- (pale green, right) isomers of isoeugenol. Active site residues that were flexible during docking are white, frozen residues cyan, and the catalytic iron and oxygen are in gold and red, respectively. For comparison, the active site-bound vanillin from PDB:5J55 is shown in purple. While *E*-isoeugenol docked readily into the active site and with −7.75 kcal/mol binding energy is the highest ranking docking result out of 999 VINA runs, the Z-isomer docks mostly in non-catalytic positions (where the ligand’s alcohol oxygen atom was more than 3.5 Å from the catalytic oxygen of Y101). The best (<3.5 Å O–O distance) catalytic pose of *Z*-isoeugenol binds with 5.42 kcal/mol as rank 240 – significantly less than the top-ranked (noncatalytic) pose, which binds with −7.17 kcal/mol.

3.4.8. Optimizing the Productivity

Different temperatures, buffer systems, and pH values for the two individual biocatalysts were compared to identify the most suitable reaction conditions to perform the two-step cascade process in one pot. The cellular confinement protects enzymes from the surrounding environment, stabilizing the enzyme's catalytic performance (Mueller and Levin 2020). We used these advantages to determine isoeugenol and vanillin productivity in whole-cell conversions at higher temperatures. It has been reported that isolated eugenol oxidase exhibits a relatively low conversion rate at 25 °C (Nguyen *et al.* 2016). Isoeugenol productivity was significantly improved from 25 °C to 37 °C (**Fig. S3.5A**). In contrast, the highest vanillin yields using N cells was obtained at 25 °C (**Fig. S3.5B**). Conversions of 4PG and isoeugenol were performed in Tris-HCl and in KPi buffers (**Fig. S3.5C, D)**. To avoid or minimize the cell lysis, our studies were performed at 50 KPi buffer (pH 7) at 25 °C.

3.4.9. Cascade Conversions: Simultaneous vs. Stepwise Addition of Whole-cell Catalysts

To enhance the efficiency of the multi-enzyme process, we aimed to optimize the stoichiometry of the biocatalysts. Combining the two cell types in different ratios slightly influenced the oxidation of 4PG (**Fig. 3.16A**). However, the vanillin yield significantly decreased in the cascade reaction, compared to an 85% yield in the transformation of isoeugenol by N cells alone. We speculated that the by-product, hydrogen peroxide, may inhibit the performance of the N cells. Theoretically, the biocatalytic oxidation of 50 mM 4PG affords 50 mM H_2O_2, which might exceed the quenching ability of the naturally present catalase in cells (Kumar and Imlay 2013; Seaver and Imlay 2001). Therefore, 50 mM sodium sulfite was added to the reaction to diminish the potentially damaging effect of H_2O_2. However, we observed a significant decrease in vanillin production probably related to the oxidation of sodium sulfite, which can inhibit the activity of NOV1-S283F, an iron-containing dioxygenase (Nguyen and Kumar 2022; Ranguelova, Bonini, and Mason 2010). Following this hypothesis, the biological alternative, 1 to 5 mg of catalase was introduced, but the amount of vanillin produced increased only slightly. This demonstrates that H_2O_2 bioproduct plays a minor role in limiting the conversion of isoeugenol to vanillin. Another possible explanation for the inefficiency of the one-pot approach is a partial substrate inhibition of NOV1-S283F by 4PG (**Fig. S3.6**). Isoeugenol and 4PG are very similar compounds, only differing in the propyl moiety. Therefore, 4PG may also bind in the active site of NOV1-S283F as a substrate analogue, competing with isoeugenol binding.

To avoid this interference, a step-by-step strategy was used to achieve a one-pot biocatalytic production of vanillin. In such an approach, all reaction steps of the cascade are still performed in one pot but are separated in time. Furthermore, separating the consecutive reactions in the time allowed us to apply in each step the optimal temperatures of the independent biocatalysts, thus maximizing the production of the desired product. P cells were added to a 50 mM 4PG to start the cascade. The N cells were introduced after 24 h and the reaction temperature decreased from 37 °C to 25 °C. The newly added biocatalyst reacted for another 24 h before analysis. Gratifyingly, a high yield of vanillin (80%) was obtained, while only a small portion of isoeugenol remained unreacted. (**Fig. 3.16B**). Clearly, the stepwise approach successfully boosted the conversion of 4PG to vanillin, probably by eliminating the inhibiting effect of 4PG on NOV1-S283F activity.

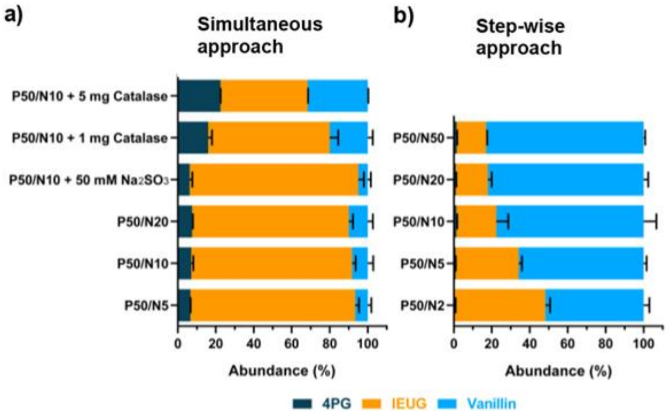


Figure 3.16. Biocatalytic cascade reactions using P and N cells in one pot. **a)** One-pot conversions of 4PG to vanillin using a whole-cell multi-enzyme cascade in a simultaneous approach. Sodium sulfite and catalase were supplied to quench the excess of H_2O_2. **b)** One-pot conversions of 4PG to vanillin using a whole-cell multi-enzyme cascade step-wise. Different amounts of N cells were added after an almost full conversion of 50 mM 4PG, which was reached utilizing P cells at OD_{600} 50. Reactions were shaken at 150 rpm at 25 °C for 24 h for a simultaneous approach and at 37 °C for 24 h, followed by the change to 25 °C for 24 h for the step-wise method.

In addition, we investigated the feasibility of our biocatalytic cascade reaction at higher substrate concentrations. Remarkably, 92% of 100 mM 4PG was converted, while with 260 mM 4PG, only 69% was converted to give a mixture of isoeugenol and vanillin (**Fig. S3.7**). Despite the difference in yields, the obtained vanillin production was 9.9 g/L (for 100 mM) and 8.7 g/L (for 260 mM), values that are similar to that obtained for 50 mM 4PG conversions (6.1 g/L). The obtained values are within the average for previously reported procedures using isoeugenol as substrate, while other biotechnological approaches and substrates remain inferior to our process (Kaur and Chakraborty 2013).

3.4.10. One-Pot Conversion of Lignin Oil

After optimizing the reaction conditions on a model system with pure 4PG, we shifted our attention to a more complex starting material: a raw depolymerized lignin oil obtained by RCF on spruce wood. Using the established procedure, 0.5 g of RCF lignin oil, containing 27.1% (wt) 4PG, resuspended in 50 mM KPi buffer pH 7, 10% (v/v) DMSO, was first reacted with P cells (OD_{600} 25) for 24 h at 37 °C, followed by a conversion by N cells (OD_{600} 10) for 24 h at 25 °C. Regardless of the cosolvent's presence, 4PG in the mixture was converted to yield 15% of isoeugenol and 66% of vanillin within 48 h. Although cosolvent has no evident effect on the conversion, we note that the use of DMSO was more practical in handling lignin-derived oil. It is worth noting that the one-pot conversion of pure 4PG using the previously optimized settings for the conversion on a small scale resulted in an approximate 90% yield of vanillin (**Fig. 3.17**). Notably, the obtained conversion yields using oil produced by RCF as a starting material were lower than those obtained using pure and commercially available 4PG. However, it should be considered that the applied lignin oil comprises about 32% phenolic monomers (27.1% being 4PG) alongside lignin-derived dimers and small oligomers (Mw <1000 g/mol) (Van Aelst *et al.* 2020). Other structurally similar monomers, such as 4-ethylguaiacol or 4-propanolguaiacol can be potential enzyme inhibitors. Yet, despite the complexity of the reaction mixture, the biocatalysts remained remarkably efficient, showing promise to valorize the lignin oil from lignocellulose into a value-added chemical with a satisfactory yield.

Finally, we aimed at simplifying the process of bioconversion. To do so, cells, after the growth phase and in their native media, were considered ready-to-use biocatalysts. 0.5 g of lignin oil in DMSO cosolvent was mixed with P cells in LB medium for 24 h at 37 °C. After 24 h, N cells, also in LB medium, were added, and conversion continued for 24 h at 25 °C. Both types of cells were used at their final OD_{600} values: OD_{600} 17 for P cells and OD_{600} 8 for N cells, which are lower than what was found to be the optimum for the highest vanillin production. Nonetheless, vanillin was produced in 60% yield, which is comparable to results obtained for KPi buffered systems (**Fig. 3.17**). The successful conversion using cells expressing the corresponding enzymes in growth media demonstrates the robustness and utility of this process. It encourages the translation of this biobased laboratory protocol into a viable industrial process.

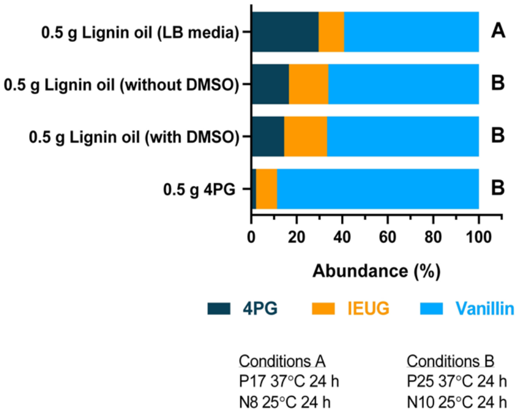


Figure 3.17. One-pot conversions of commercial 4PG and lignin-derived oils using a whole-cell, multi-enzyme cascade stepwise under different experimental conditions. One-pot conversions of 0.5 g lignin oil or 4PG in 50 mM KPi buffer, pH 7. For conversions of 0.5 g lignin oil, the effect of 10% (v/v) DMSO as cosolvent was tested. To catalyze the first step of the process, P cells at OD_{600} 25 and N cells at OD_{600} 10 were used in the second step. A conversion of 0.5 g RCF oil was performed using P and N cells at OD_{600} 17 and OD_{600} 8, respectively, resuspended in the growth media, i.e., LB media. All the reactions were shaken at 150 rpm at 37 °C for 24 h, then changed to 25 °C for 24 h.

To isolate and fully characterize vanillin produced through the biocatalytic cascade, 1 g of RCF lignin oil containing 27.1% (wt) of 4PG (279 mg, 1.7 mmol) was reacted for 48 h in one-pot system to produce vanillin. The reaction mixture was purified by column chromatography, followed by recrystallisation. The first purification step yielded 40% (99 mg, 0.7 mmol) of vanillin. At the same time the NMR analysis indicated that the isolated product contained traces of impurities (**Fig. 3.20A**). Then, the same sample was recrystallized, which significantly reduced the yield of vanillin (10%, 26 mg, 0.2 mmol) while the purity increased (**Fig. 3.20B**). Note that this protocol was not optimized and should be considered a first proof-of-concept. Further purification process development is required, building upon industrially existing best practices. At the same time, the NMR analysis of unreacted isoeugenol, obtained through column chromatography, sheds some light on the reasons behind its incomplete conversion to vanillin. The integration of corresponding CH_3 signals of isoeugenol isomers indicates that the *Z*-isoeugenol is the more dominant component of the mixture (**Fig. 3.20C**), which is in agreement with NMR and docking studies.

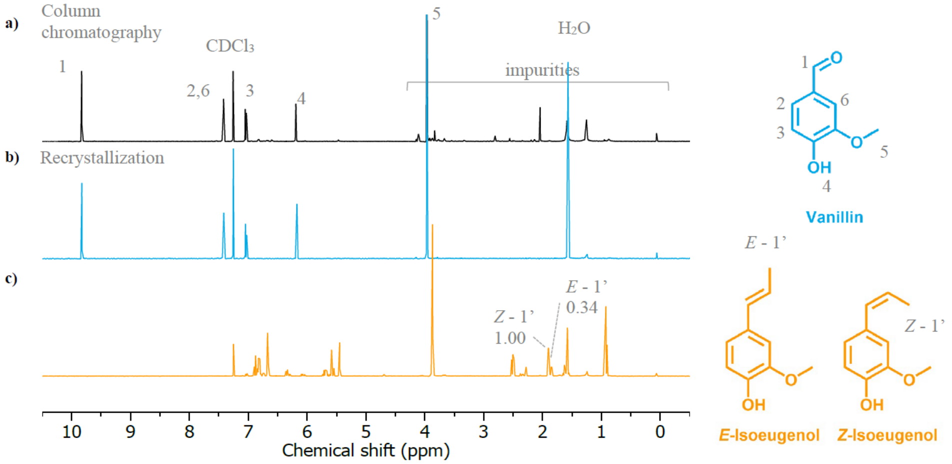


Figure 3.20. NMR spectra of **A)** vanillin isolated via column chromatography; **B)** vanillin isolated by column chromatography followed by the recrystallization. ^{1}H NMR (400 MHz, Chloroform-d) δ 9.83 (s, ^{1}H), 7.46 – 7.40 (m, ^{2}H), 7.04 (d, J = 8.5 Hz, ^{1}H), 6.18 (s, ^{1}H), 3.97 (s, ^{3}H). **C)** unreacted mixture of *E*- and *Z*-isoeugenol isolated by column chromatography. ^{1}H NMR (400 MHz, CDCl3) δ (ppm): *E*-isomer: 6.91 – 6.78 (m, ^{3}H, ArH), 6.39 – 6.28 (m, ^{1}H, (C)CH), 6.08 (dq, J = 15.7, 6.6 Hz, ^{1}H, $CHCHCH_3$), 5.55 (s, ^{1}H, OH), 3.90 (s, 1H, OCH_3), 1.86 (dd, J = 6.6, 1.7 Hz, ^{3}H, CH_3). *Z*-isomer: 6.91 – 6.78 (m, ^{3}H, ArH), 6.39 – 6.28 (m, ^{1}H, (C)CH), 5.70 (dq, J = 11.6, 7.2 Hz, ^{1}H, $CHCHCH_3$), 5.59 (s, ^{1}H, OH), 3.89 (s, ^{1}H, OCH_3), 1.90 (dd, J = 7.2, 1.9 Hz, ^{3}H, CH_3). Numbers in labels represent the value of the integral of each peak, as MestReNova gives.

3.5. CONCLUDING REMARKS

In this study, we rationally designed, constructed, and tested thirty-five variants of NOV1, ultimately selecting the S283F variant for its superior selectivity and catalytic efficiency towards isoeugenol compared to the wild-type enzyme. This variant also showed enhanced kinetic thermostability due to increased stabilization of the iron cofactor inside the catalytic cavity. Incidentally, the primary molecular determinant of NOV1 kinetic stability is suggested to be iron depletion of the active site. MD analyses supported increased iron retention in the active site of S283F and enhanced kinetic stability.

The biotransformation of isoeugenol using whole-cells expressing the improved S283F variant achieved remarkable conversion levels to vanillin, with concentrations reaching up to 100 mM at room temperature in relatively short reaction times. Enzyme deactivation and stability under harsh conditions are significant challenges in industrial biocatalysis. Thus, S283F, with its increased activity and stability, was the optimal candidate for testing in a biocatalytic cascade aimed at vanillin production from RCF lignin-derived 4PG. By combining *E. coli* cells expressing an engineered 4PG oxidase (PROGO) with NOV1-S283F, we optimized the conditions for a whole-cell biocatalytic process. This approach yielded a high vanillin production rate of 90% from synthetically pure 4PG and 66% from real lignin oil, which contains 27.1% 4PG, derived from RCF biorefining of spruce wood. The pre-purification vanillin yield was 18% from RCF lignin oil and 3% from the initial spruce feedstock, representing an estimated four-fold improvement over current industrial wood-based vanillin production processes (Breiner 2021). Notably, the relative increase for the post-purification vanillin yield is expected to be even larger due to the lack of a homogeneous metal catalyst in the vanillin product stream presented here. Furthermore, we showed that cells in a growth medium could be used as ready-to-use biocatalysts, making this one-pot conversion of lignin oil into vanillin a rather facile process. The presented lignin refining strategy paves the way for further development to sustainably produce biobased vanillin using lignin as feedstock.

3.6 REFERENCES

Breiner, M., Strugatchi, J., Waldvogel, S. R. 2021. 'Vanillin from lignin', *Wiley Anal. Sci.*

Canutescu, A. A., Shelenkov, A. A., Dunbrack, R. L. 2003. 'A graph-theory algorithm for rapid protein side-chain prediction', *Protein Sci.*, 12: 2001-14.

Collaborative Computational Project, Number. 1994. 'The CCP4 suite: programs for protein crystallography', *Acta Crystallogr. D Struct. Biol.*, 50: 760-3.

Daruwalla, A., Kiser, P. D. 2020. 'Structural and mechanistic aspects of carotenoid cleavage dioxygenases (CCDs)', *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, 1865: 158590.

Dunning Hotopp, J. C., Auchtung, T. A., Hogan, D. A., Hausinger, R. P. 2003. 'Intrinsic tryptophan fluorescence as a probe of metal and alpha-ketoglutarate binding to TfdA, a mononuclear non-heme iron dioxygenase', *J. Inorg. Biochem.*, 93: 66-70.

Eftink, M. R. 1998. 'The use of fluorescence methods to monitor unfolding transitions in proteins', *Biochemistry (Mosc)*, 63: 276-84.

Emsley, P., Lohkamp, B., Scott, W. G., Cowtan, K. 2010. 'Features and development of Coot', *Acta Crystallogr. D Struct. Biol.*, 66: 486-501.

Fernandes, A., Pinto, B., Bonardo, L., Royo, B., Robalo, M. P., Martins, L. O. 2021. 'Wasteful Azo Dyes as a Source of Biologically Active Building Blocks', *Front. Bioeng. Biotechnol.*, 9.

Fernandes, A. T., Martins, L. O., Melo, E. P. 2009. 'The hyperthermophilic nature of the metallo-oxidase from Aquifex aeolicus', *Biochim. Biophys. Acta-Proteins Proteomics*, 1794: 75-83.

Fischer, D. S., Price, D. C. 1964. 'A Simple Serum Iron Method Using the New Sensitive Chromogen Tripyridyl-s-triazine', *Clin. Chem.*, 10: 21-31.

Furuya, T., Miura, M., Kuroiwa, M., Kino, K. 2015. 'High-yield production of vanillin from ferulic acid by a coenzyme-independent decarboxylase/oxygenase two-stage process', *New Biotechnol.*, 32: 335-39.

Garcia-Ochoa, F., Gomez, E., Santos, V. E., Merchuk, J. C. 2010. 'Oxygen uptake rate in microbial processes: An overview', *Biochem. Eng. J.*, 49: 289-307.

Guo, Y. M., Alvigini, L., Trajkovic, M., Alonso-Cotchico, L., Monza, E., Savino, S., Mattevi, A., Fraaije, M. W. 2022. 'Structure- and computational-aided engineering of an oxidase to produce isoeugenol from a lignin-derived compound', *Nat. Commun.*, 13.

Han, Z. C., Long, L. K., Ding, S. J. 2019. 'Expression and Characterization of Carotenoid Cleavage Oxygenases From Herbaspirillum seropedicae and Rhodobacteraceae bacterium Capable of Biotransforming Isoeugenol and 4-Vinylguaiacol to Vanillin', *Front. Microbiol.*, 10: 1-11.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. 2006. 'Comparison of multiple amber force fields and development of improved protein backbone parameters', *Proteins*, 65: 712-25.

Kabsch, W. 2010. 'Xds', *Acta Crystallogr. D Struct. Biol.*, 66: 125-32.

Kamoda, S., Terada, T., Saburi, Y. 2003. 'A Common Structure of Substrate Shared by Lignostilbenedioxygenase Isozymes from Sphingomonas paucimobilis TMY1009', *Biosci. Biotechnol. Biochem.*, 67: 1394-96.

Kaur, B., Chakraborty, D. 2013. 'Biotechnological and Molecular Approaches for Vanillin Production: a Review', *Appl. Biochem. Biotechnol.*, 169: 1353-72.

Kiser, P. D., Golczak, M., Lodowski, D. T., Chance, M. R., Palczewski, K. 2009. 'Crystal structure of native RPE65, the retinoid isomerase of the visual cycle', *Proc. Natl. Acad. Sci. U.S.A.*, 106: 17325-30.

Kloer, D. P., Ruch, S., Al-Babili, S., Beyer, P., Schulz, G. E. 2005. 'The structure of a retinal-forming carotenoid oxygenase', *Science*, 308: 267-69.

Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A., Vriend, G. 2004. 'Making optimal use of empirical energy functions: force-field parameterization in crystal space', *Proteins*, 57: 678-83.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., Karplus, K. 2009. 'Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8', *Proteins-Struct. Funct. Bioinform.*, 77: 114-22.

Krieger, E., G. Koraimann, and G. Vriend. 2002. 'Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field', *Proteins-Struct. Funct. Bioinform.*, 47: 393-402.

Krieger, E., and G. Vriend. 2015. 'New ways to boost molecular dynamics simulations', *J. Comput. Chem.* 36: 996-1007.

Kuatsjah, E., M. M. Verstraete, M. J. Kobylarz, A. K. N. Liu, M. E. P. Murphy, and L. D. Eltis. 2019. 'Identification of functionally important residues and structural features in a bacterial lignostilbene dioxygenase', *J. Biol. Chem.* 294: 12911-20.

Kumar, S. R., and J. A. Imlay. 2013. 'How *Escherichia coli* tolerates profuse hydrogen peroxide formation by a catabolic pathway', *J. Bacteriol.* 195: 4569-79.
Le Guilloux, V., P. Schmidtke, and P. Tuffery. 2009. 'Fpocket: An open source platform for ligand pocket detection', *BMC Bioinformatics* 10.
Lin, B. X., and Y. Tao. 2017. 'Whole-cell biocatalysts by design', *Microb. Cell Fact.* 16.

Marasco, E. K., and C. Schmidt-Dannert. 2008. 'Identification of bacterial carotenoid cleavage dioxygenase homologues that cleave the interphenyl alpha,beta double bond of stilbene derivatives via a monooxygenase reaction', *ChemBioChem* 9: 1450-61.

Mark, P., and L. Nilsson. 2001. 'Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K', *J. Phys. Chem. A* 105: 9954-60.

McAndrew, R. P., N. Sathitsuksanoh, M. M. Mbughuni, R. A. Heins, J. H. Pereira, A. George, K. L. Sale, B. G. Fox, B. A. Simmons, and P. D. Adams. 2016. 'Structure and mechanism of NOV1, a resveratrol-cleaving dioxygenase', *PNAS USA* 113: 14324-29.

Messing, S. A. J., S. B. Gabelli, I. Echeverria, J. T. Vogel, J. C. Guan, B. C. Tan, H. J. Klee, D. R. McCarty, and L. M. Amzel. 2010. 'Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid', *Plant Cell* 22: 2970-80.

Metz, F., A. M. Olsen, F. C. Lu, K. S. Myers, M. N. Allemann, J. K. Michener, D. R. Noguera, and T. J. Donohue. 2024. 'Catabolism of β-5 linked aromatics by Novosphingobium aromaticivorans', *ASM Appl. Ind. Microbiol.* 15.

Mueller, E. A., and P. A. Levin. 2020. 'Bacterial cell wall quality control during environmental stress', *ASM Appl. Ind. Microbiol.* 11.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin. 2011. 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallogr. D Biol. Crystallogr.* 67: 355-67.

Nguyen, K., and P. Kumar. 2022. 'Morphological phenotypes, cell division, and gene expression of *Escherichia coli* under high concentration of sodium sulfate', *Microorganisms* 10.

Nguyen, Q. T., G. de Gonzalo, C. Binda, A. Rioz-Martínez, A. Mattevi, and M. W. Fraaije. 2016. 'Biocatalytic properties and structural analysis of eugenol oxidase from

Rhodococcus jostii RHA1: A versatile oxidative biocatalyst', *ChemBioChem* 17: 1359-66.

Ollikainen, N., R. M. de Jong, and T. Kortemme. 2015. 'Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity', *PLoS Comput. Biol.* 11: e1004335.

Ranguelova, K., M. G. Bonini, and R. P. Mason. 2010. '(Bi)sulfite oxidation by copper,zinc-superoxide dismutase: sulfite-derived, radical-initiated protein radical formation', *Environ. Health Perspect.* 118: 970-75.

Ryu, J. Y., J. Seo, S. Park, J. H. Ahn, Y. Chong, M. J. Sadowsky, and H. G. Hur. 2013. 'Characterization of an isoeugenol monooxygenase (Iem) from *Pseudomonas nitroreducens* Jin1 that transforms isoeugenol to vanillin', *Biosci. Biotechnol. Biochem.* 77: 289-94.

Sanchez-Ruiz, J. M. 2010. 'Protein kinetic stability', *Biophys. Chem.* 148: 1-15.

Schmidt, S., A. Schallmey, and R. Kourist. 2021. 'Multi-Enzymatic Cascades In Vitro.' in Selin Kara and Florian Rudroff (eds.), *Enzyme Cascade Design and Modelling* (Springer International Publishing: Cham).

Schmidtke, P., V. Le Guilloux, J. Maupetit, and P. Tufféry. 2010. 'fpocket: online tools for protein ensemble pocket detection and tracking', *Nucleic Acids Res.* 38: W582-W89.

Seaver, L. C., and J. A. Imlay. 2001. 'Hydrogen peroxide fluxes and compartmentalization inside growing *Escherichia coli*', *J. Bacteriol.* 183: 7182-89.

Sigurdardóttir, S. B., J. Lehmann, S. Ovtar, J.-C. Grivel, M. Della Negra, A. Kaiser, and M. Pinelo. 2018. 'Enzyme immobilization on inorganic surfaces for membrane reactor applications: mass transfer challenges, enzyme leakage and reuse of materials', *Adv. Synth. Catal.* 360: 2578-607.

Sui, X. W., M. Golczak, J. Y. Zhang, K. A. Kleinberg, J. von Lintig, K. Palczewski, and P. D. Kiser. 2015. 'Utilization of dioxygen by carotenoid cleavage oxygenases', *J. Biol. Chem.* 290: 30212-23.

Sui, X. W., A. C. Weitz, E. R. Farquhar, M. Badiee, S. Banerjee, J. von Lintig, G. P. Tochtrop, K. Palczewski, M. P. Hendrich, and P. D. Kiser. 2017. 'Structure and spectroscopy of alkene-cleaving dioxygenases containing an atypically coordinated non-heme iron center', *Biochemistry* 56: 2836-52.

Tina, K. G., R. Bhadra, and N. Srinivasan. 2007. 'PIC: Protein Interactions Calculator', *Nucleic Acids Res.* 35: W473-6.

Trott, O., and A. J. Olson. 2010. 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *J. Comput. Chem.* 31: 455-61.

Van Aelst, K., E. Van Sinay, T. Vangeel, E. Cooreman, G. Van den Bossche, T. Renders, J. Van Aelst, S. Van den Bosch, and B. F. Sels. 2020. 'Reductive catalytic fractionation of pine wood: elucidating and quantifying the molecular structures in the lignin oil', *Chem. Sci.* 11: 11498-508.

Wachtmeister, J., and D. Rother. 2016. 'Recent advances in whole cell biocatalysis techniques bridging from investigative to industrial scale', *Curr. Opin. Biotechnol.* 42: 169-77.

Wang, F., J. Zhao, Q. Li, J. Yang, R. J. Li, J. Min, X. J. Yu, G. W. Zheng, H. L. Yu, C. Zhai, C. G. Acevedo-Rocha, L. X. Ma, and A. T. Li. 2020. 'One-pot biocatalytic route from cycloalkanes to α,ω-dicarboxylic acids by designed *Escherichia coli* consortia', *Nat. Commun.* 11.

Yamada, M., Y. Okada, T. Yoshida, and T. Nagasawa. 2007. 'Purification, characterization and gene cloning of isoeugenol-degrading enzyme from *Pseudomonas putida* IE27', *Arch. Microbiol.* 187: 511-17.

Yamada, M., Y. Okada, T. Yoshida, and T. Nagasawa. 2008. 'Vanillin production using *Escherichia coli* cells over-expressing isoeugenol monooxygenase of *Pseudomonas putida*', *Biotechnol. Lett.* 30: 665-70.

Yao, J., Y. He, N. N. Su, S. R. Bharath, Y. Tao, J. M. Jin, W. Chen, H. Song, and S. Y. Tang. 2020. 'Developing a highly efficient hydroxytyrosol whole-cell catalyst by de-bottlenecking rate-limiting steps', *Nat. Commun.* 11.

Zhao, L. Q., Y. Z. Jiang, H. Y. Fang, H. C. Zhang, S. Cheng, M. S. R. Rajoka, and Y. G. Wu. 2019. 'Biotransformation of isoeugenol into vanillin using immobilized recombinant cells containing isoeugenol monooxygenase active aggregates', *Appl. Biochem. Biotechnol.* 189: 448-58.

Zhao, L. Q., Y. M. Xie, L. Y. Chen, X. F. Xu, C. X. Zha, and F. Cheng. 2018. 'Efficient biotransformation of isoeugenol to vanillin in recombinant strains of *Escherichia coli* by using engineered isoeugenol monooxygenase and sol-gel chitosan membrane', *Process Biochem.* 71: 76.

Chapter 4

Distal Mutations Enhance Efficiency of Free and Immobilized NOV1 Dioxygenase for Vanillin Synthesis

This chapter contains data published in:

Mario De Simone, Lur Alonso-Cotchico, Maria Fátima Lucas, Vânia Brissos, and Lígia O. Martins (2024). Distal Mutations Enhance Efficiency of Free and Immobilized NOV1 Dioxygenase for Vanillin Synthesis. *J. Biotechnol, 391, 92-98.*
https://doi.org/10.1016/j.jbiotec.2024.06.012

Author contributions:

The author of this thesis performed construction, purification and biochemical characterization of all NOV1 variants in this chapter. Furthermore, he performed the construction of DNA-shuffling library, the screening in 96-well plate and the enzyme immobilization experiments. Distal Hotspot identification and their analysis was performed by Lur Alonso-Cotchico at Zymvol Biomodeling (Barcelona) under Maria Fatima ´Lucas’s supervision. Vania Brissos: aided part of the experimental work and reviewed and edited the writing. This work was supervised and revised by Prof. Ligia O Martins.

4.1. ABSTRACT

Protein engineering is crucial to improve enzymes' efficiency and robustness for industrial biocatalysis. This study aims to enhance NOV1 activity and operational stability by identifying distal hotspots located at more than 9 Å from the active site using Zymspot. This tool predicts advantageous distant mutations, streamlining protein engineering. Fourty-one variants were constructed using site-directed mutagenesis, and the six most active enzyme variants were then recombined. Two variants, with two and three mutations, showed nearly a 10-fold increase in activity and up to 40-fold higher operational stability than the wild-type. Furthermore, these variants show 90% to 100% immobilization efficiency in metal affinity resins, compared to approximately 60% for the wild-type. In bioconversion tests, where 50 mM of isoeugenol was added stepwise over 24-hour cycles, the 1D2 variant produced approximately 144 mM of vanillin after six reaction cycles, corresponding to around 22 mg, indicating a 35% molar conversion yield. This output was around 2.5 times higher than that obtained using the wild-type. Our findings highlight the efficacy of distal protein engineering in enhancing enzyme functions like activity, stability, and metal binding selectivity, thereby fulfilling the criteria for industrial biocatalysts.

Keywords: Biocatalysis, Non-heme iron proteins, Immobilization metal affinity, Protein Engineering

4.2. INTRODUCTION

In the previous chapter, we have made progress in understanding the one-step, coenzyme-free conversion of isoeugenol to vanillin by NOV1 dioxygenase. Advancements in comprehending the structure-function relationships of NOV1 through rational mutagenesis near the active site have been made (De Simone *et al.* 2023). Furthermore, its potential has been recognized for effectively utilizing lignin-derived intermediates in whole-cell bioprocesses to produce valuable compounds

In this study, we have two objectives: 1) use computational tools to predict beneficial mutations distal to active sites, and 2) optimize the immobilization of NOV1 variant enzymes onto solid metal supports by affinity, proposed to be easier and more advantageous over conventional methods of immobilization. Protein engineering is crucial to optimizing enzymes and improving their efficiency and robustness for industrial biocatalysis (Bornscheuer *et al.* 2012). Most rational and semi-rational engineering designs target residues near the active or substrate binding sites. Still, it is becoming increasingly clear that distal sites ≥10 Å significantly contribute to the enzymatic function (Gu, Xu, and Nie 2023; Lee and Goodey 2011), although the mechanisms underlying their influence are still unclear (Motlagh *et al.* 2014; Wilding *et al.* 2019). Current strategies for identifying distal hotspots in enzymes involve assessing the enzyme's conformational landscape to understand its allosteric behavior (Osuna 2021; Wang *et al.* 2020), relying on lengthy molecular dynamics simulations. Emerging machine-learning models offer a promising approach to identifying these functional sites (Cagiada *et al.* 2023). As part of Zymvol’s proprietary software for enzyme engineering, Zymevolver (trade secret), Zymspot was validated. This algorithm integrates bioinformatics and structure-based methods to predict long-range mutations with minimal human intervention, allowing the construction of "smart libraries" that significantly accelerate protein engineering (Casilli *et al.* 2024; Zymvol-Biomodeling). Zymspot does not rely on time-consuming strategies such as molecular dynamics simulations to assess the impact of the mutations on the conformational dynamics of the protein but quickly extracts bioinformatics and structural metrics that were benchmarked against experimental data. Based on the assumption that sequence similarity implies functional similarity (Yu *et al.* 2022), Zymspot selects the most representative amino acids at each position from a multiple sequence alignment (MSA). This process helps identify positions conserved across related sequences but differ in the target protein, pinpointing potential sites for functional improvement. On the other hand, immobilization confers enzyme stabilization, simplifies the separation of products, and enables recycling, enhancing the efficiency and cost-effectiveness of the industrial application of enzymes (Basso *et al.* 2022; Sheldon and van Pelt 2013; Winand *et al.* 2023). While vanillin production using whole cells has been well investigated, few reports concerning its production using immobilized enzymes can provide advantages in specific applications (Furuya, Kuroiwa, and Kino 2017; Liu *et al.* 2023). The high costs of the carrier, low immobilization efficiency, and loss of activity are common drawbacks of enzyme immobilization (Es, Vieira, and Amaral 2015). Still, the development of new supports by affinity interaction allows higher immobilization efficiency, faster immobilization, and even the possibility to perform a "one-pot" purification and immobilization with clear advantages over conventional cross-linking, covalent and ionic binding immobilization techniques (Barbosa *et al.* 2015).

Herein, we employed Zymspot to identify advantageous single distal mutations on the NOV1 scaffold. We generated variants with enhanced catalytic rates and stability compared to the wild-type through site-directed mutagenesis and recombination. Double and triple-mutation-containing variants were selected and immobilized on Chromalite© MIDA/M particles with different metal ions, utilizing their His-tags' metal affinity. These engineered immobilized enzymes outperformed the wild-type in vanillin production, demonstrating significant improvements. Our study highlights the importance of distal interactions in improving enzymatic properties, i.e., the importance of developing strategies for distal-mutation-based protein engineering.

4.3. MATERIALS AND METHODS

4.3.1. Bacterial strains, plasmids, and culture media

E. coli strain DH5α (Novagen) was used to amplify plasmid constructs. *E. coli* BL21 star (DE3, Invitrogen) strain was used to express the *nov1* gene cloned in pET-28a(+) plasmid (Novagen) and its variants. In the BL21star strain, the expression of the *nov1* gene is controlled by the T7 promoter and induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). Luria-Bertani medium (LB) was used to grow *E. coli* strains, supplemented with kanamycin (50 μg mL^{-1}).

4.3.2. Identification of NOV1 Distal Hotspot Residues

Our strategy comprised Zymvol's in-house algorithm, Zymevolver, and its utility to depict distal hotspots, known as Zymspot (Zymvol-Biomodeling). This algorithm deploys bioinformatics and structure-based approaches to depict long-range mutations with the potential to perturb the conformational dynamics of the enzyme, ultimately influencing the population of catalytically productive conformations. It has consistently demonstrated its efficiency in more than 20 enzyme engineering campaigns to improve catalytic properties (mainly activity, selectivity, enzyme stability, operational stability, and expression) in approximately 30% of the predicted mutant libraries. In this work, it has been used to improve NOV1 activity against isoeugenol.

4.3.3. Construction of *nov1* Variants using Site-directed Mutagenesis

Single amino acid substitutions in the *nov1* gene were inserted using the Quick-Change site-directed mutagenesis protocol (Stratagene). The plasmid pET-28a(+) containing the *nov1* gene was used as a template using appropriate primers (**Table S4.1,** Annex II), and the PCRs were performed as previously described (De Simone *et al.* 2023). The amplified products were purified using GFX PCR DNA and a Gel Band Purification kit (GE Healthcare). The final PCR products were digested with *Dpn*I to eliminate the wild-type template and were used to transform electrocompetent *E. coli* strain DH5α cells. The presence of the desired mutations in the resulting plasmids and the absence of additional inadvertent mutations in other regions of the inset were confirmed by DNA sequencing.

4.3.4. Recombination by DNA Shuffling

DNA shuffling was performed after the amplification of genes coding for variants I222V, L286W, T299S, M324K, L372A and N478S using primers pET21Fwd and pET21Rev (**Table S4.1,** Annex II). A mixture containing 200 ng of each parental gene was digested with 0.04 U of DNaseI in 200 mM Tris-HCl, pH 7, with 0.2 M $MnCl_2$ for 10 min at 15 °C in a thermal cycler. Digestion was stopped by adding 50 mM EDTA. The PCR reassembly was performed in a 20 µL reaction volume containing 5 µL of DNA fragments, 200 µM of dNTPs, NZYProof polymerase buffer and 2.5 U of NZYProof polymerase. After an initial denaturation period of 3 min at 96 °C, the subsequent steps were repeated for 45 cycles in a thermal cycler: 1 min at 94 °C, 90 s at 59 °C, 90 s at 56 °C, 90 s at 53 °C, 90 s at 50 °C, 90 s at 47 °C, 90 s at 44 °C, 90 s at 41 °C and 1 min + 5 s/cycle at 72 °C followed by a final 10 min period at 72 °C. The PCR reassembly products were amplified by PCR using the primers NOV1GApET28a Fwd and NOV1GApET28a Rev (**Table S4.1,** *Annex II*) as previously described (De Simone et al. 2023). The vector pET28a was amplified by PCR using primers pET28aGAnov1 Fwd and pET28aGAnov1 Rev (**Table S4.1,** *Annex II*) with the following conditions: 50 µL reaction volume containing 2.5 ng of DNA template, 200 µM dNTPs, Q5 polymerase buffer and 2.5 U of Q5® High-Fidelity DNA polymerase (New England Biolabs). After an initial denaturation period of 1 min at 98 °C, the subsequent steps were repeated for 35 cycles in a thermal cycler: 30 s at 98 °C, 20 s at 72 °C, 2 min at 72 °C, followed by a final 10 min period at 72 °C. The amplified products were purified using GFX PCR DNA and Gel Band Purification kit (GE Healthcare, Chicago, IL, USA). The ligation between shuffling products and vector pET28a was performed using the NEBuilder® HiFi DNA Assembly (New England Biolabs). The constructed library was first introduced into E. cloni 10G ELITE electrocompetent cells (Lucigen), the plasmids were extracted and then propagated in *E. coli* BL21 star cells.

4.3.5. High-throughput Activity Screening in 96-well plates

Single colonies were picked from a fresh agar plate and transferred to 96-well microplates containing 200 µL of LB medium supplemented with 50 µg mL^{-1} kanamycin. Cell cultivation, disruption, and activity screenings in crude extracts for isoeugenol following vanillin production at 340 nm (ε_{340} = 15,97 M^{-1} cm^{-1}) were performed as previously described (De Simone *et al.* 2023).

4.3.6. Enzyme Production and Purification

The cultivation of recombinant *E. coli* BL21star strains in 100 mL LB medium supplemented with kanamycin in 500 mL Erlenmeyer's at 37 °C, 100 rpm shaking and the induction of gene expression with 0.1 mM IPTG and 0.5 mM $(NH_4)_2Fe(SO_4)_2(H_2O)_6$ was performed as previously described (De Simone *et al.* 2023). The protein purification protocol, using an affinity chromatography HisTrap™ column, and the digestion of the pure protein with SUMO-protease were performed as previously described (De Simone *et al.* 2023). The purified protein concentration was estimated using the molar absorption coefficient of NOV1 (ε_{280} = 81,930 M^{-1} cm^{-1}) calculated from the protein sequence using the ExPASy Bioinformatics Resource Portal (http://web.expasy.org). The iron content of purified protein samples was determined by colorimetric assay resorting to the chelator TPTZ (2, 4, 6-Tris(2-pyridyl)-s-triazine) (Fischer and Price 1964).

4.3.7. Apparent Steady-state Kinetic Analysis

The activity dependence on pH was measured using Britton-Robinson buffer (pH 3-11), and the optimal temperature was determined at the temperature range of 10-40 °C. Apparent steady-state kinetic parameters were measured at 340 nm by adding the enzyme to a mixture containing isoeugenol (0.01−4 mM) in 100 mM Tris-HCl, pH 9. Kinetic data were fitted directly to the Michaelis-Menten equation using the Origin® software. All enzymatic assays were performed at least in triplicate.

4.3.8. Thermodynamic and Kinetic Stability

Protein unfolding was monitored using a Cary Eclipse spectrofluorometer with an excitation wavelength of 296 nm and recording the fluorescence emission of tryptophan residues at 340 nm as previously described (De Simone *et al.* 2023). Thermal unfolding was analyzed according to a two-state model N ⇌ U. (Eftink 1998). Thermal inactivation was studied as previously described (De Simone *et al.* 2023). In brief, enzyme solutions were incubated at 25 °C in 20 mM Tris-HCl, pH 7.6, and at fixed time intervals, aliquots were withdrawn and tested for activity.

4.3.9. Immobilization on Metal-loaded Resin by Affinity Interaction

Chromalite MIDA/M is a microporous iminodiacetic acid polymethacrylic affinity resin, which can be loaded with divalent transition metals. The purified NOV1 wild-type and variant enzymes 12G2 (M324K/L372A), and 1D2 (I222V/L286W/L372A) were immobilized by affinity interaction on Chromalite™ MIDA/M resins loaded with Ni^{2+} (LS04058-186), Co^{2+} (LS04054-186), Zn^{2+} (LS04057-186), Cu^{2+} (LS04055-186) and Fe^{2+} (LS04056-186), taking advantage of the His-tag of the recombinant enzymes. Resins were first equilibrated, washing twice with 1 mL of 50 mM sodium phosphate, pH 8. The immobilization was performed by incubating 0.1 mg protein/10 mg dry resin in the same buffer at 25 °C, under rotary mixing at 15 rpm (Hybrid Mini 10 incubator) for one hour, in a total volume of 0.5 mL. After immobilization, resins were centrifuged, the supernatants were collected, and the carrier was washed with 0.5 mL of buffer. The total free protein was quantified using the Bradford method, and the amount of enzyme immobilized enzyme was calculated from a mass balance. The immobilization yield (1) was determined to evaluate the immobilization process.

$$(1)\quad Immobilization\ yield\ (\%) = 100 * \left(\frac{P1-P2}{P1}\right)$$

P1 is the initial protein concentration, and P2 is the free protein concentration after immobilization in the supernatant and washing.

The immobilized enzymes were resuspended in 0.5 mL of buffer and used in enzymatic assays to measure the specific activity and compared with that of the free enzyme to calculate the activity yield (2) and immobilization efficiency (3). Assays were performed in 1 mL volume, adding 20-100 μl of immobilized enzyme in 100 mM Tris-HCl buffer, pH 9, containing 2 mM isoeugenol in a 96-deep well plate on a shaker at 1000 rpm at room temperature. Reactions were followed, taking aliquots over time and reading absorbance at 340 nm.

$$(2)\quad Activity\ Yield\ (\%) = 100 * \left(\frac{immobilized\ activity}{initial\ activity}\right)$$

$$(3)\quad Immobilization\ Efficiency\ (\%) = 100 * \left(\frac{Activity\ Yield}{Immobilization\ Yield}\right)$$

The kinetic stability of immobilized enzymes was assessed using 0.3 mg protein on 30 mg Chromalite^{TM} MIDI/M/Ni^{2+} in 50 mM sodium phosphate, pH 8, and incubating at room temperature. Samples are withdrawn, cooled, and tested for activity at appropriate times. Inactivation constants k_{in} and half-life values are calculated as previously described (De Simone *et al.* 2023).

4.3.10. Bioconversions of Isoeugenol to Vanillin by Immobilized Enzymes

Bioconversions were performed using 1.2 mg of enzyme immobilized on 120 mg of Chromalite^{TM} MIDI/M/Ni^{2+}. Immobilized enzymes were resuspended in 1 mL of 100 mM Tris-HCl, pH 9, with 50 mM isoeugenol. Reaction mixtures were transferred into a flask at room temperature at 10,000 rpm. At specific times, aliquots were withdrawn to quantify vanillin production at 340 nm and calculate the % of substrate conversion (defined as [product]/[initial substrate]*100). After 24 h, all the reaction mixtures were transferred to an Eppendorf tube and centrifuged for 2 min, at 12,000 rpm, to recover the supernatant. The immobilized enzymes were washed with buffer, centrifuged, and resuspended with fresh reaction mixtures to start new conversion cycles.

4.4. RESULTS AND DISCUSSION

4.4.1. Identification of NOV1 Distal Hotspots

A comprehensive computational screening was performed in NOV1 dioxygenase using Zymspot to identify positions that increased NOV1 activity against isoeugenol. As a result, 62 distal positions, covering 13% of the protein, were identified to have a key role in modulating the enzyme properties. This list was shortened to the best-ranked 20 positions, representing 4% of the protein. Driven by phylogenetic and structural information, the amino acid variability was assessed at each position, resulting in a final list of 41 single-point residues.

4.4.2. Construction, Screening, and Characterization of Variants

The distal variants (41) resulting from the computational study were constructed using site-directed mutagenesis, and their activity was measured for isoeugenol (**Table 4.1**). Twenty enzyme variants had no detectable activity, ten showed lower activity than wild-type, and eleven (T299S, L372A, L286V, M324K, L286W, M374A, I222V, N478S, L286I, T299V, and I222L), around 27% of the library, showed improved activity as compared to wild-type. Notably, seven of these improved variants are placed in three positions: T299 (2), L286 (3) and I222 (2), suggesting these positions are very permissive to mutations. The eleven variants were purified and characterized. The purified enzyme preparations are partially iron-depleted, displaying approximately 0.5 mol of iron per mol of protein (**Table 4.2**), lower amounts as compared to wild-type, as previously observed in our previous studies (De Simone *et al.* 2023) and with other isoeugenol oxygenases (Ryu *et al.* 2013; Yamada *et al.* 2007). Notably, all variants displayed an increase in k_{cat} up to 4-fold compared to wild-type except I222L and M374A, although the K_m values are generally 2-fold higher (**Table 4.2**).

Table 4.1. Enzymatic activity of constructed variants. Reactions were performed using cell-crude extracts in 1 mM isoeugenol, 100 mM Tris-HCl, pH 9, at room temperature. The results are the mean ± standard error of three independent experiments. Errors for $iA_{variant}/iA_{WT}$ were obtained by propagation.

#	Enzyme	Relative activity ($iA_{Variant}/iA_{WT}$)	#	Enzyme	Relative activity ($iA_{Variant}/iA_{WT}$)
	Wild-type	-			
1	T299S	2.2 ± 0.2	22	K86G	nd
2	L372A	2.1 ± 0.4	23	K86R	nd
3	L286V	2.0 ± 0.2	24	R116K	nd
4	M324K	2.0 ± 0.4	25	N122A	nd
5	L286W	2.0 ± 0.3	26	N122S	nd
6	M374A	1.9 ± 0.2	27	F124I	nd
7	I222V	1.8 ± 0.3	28	F124L	nd
8	N478S	1.5 ± 0.2	29	F164M	nd
9	L286I	1.4 ± 0.2	30	M195F	nd
10	T299V	1.4 ± 0.2	31	M195Y	nd
11	I222L	1.2 ± 0.2	32	L286A	nd
12	L372V	0.8 ± 0.2	33	T299G	nd
13	F124V	0.7 ± 0.1	34	E301R	nd

14	T299L	0.5 ± 0.1	35	W330I	nd
15	T299A	0.5 ± 0.1	36	R356M	nd
16	M177L	0.4 ± 0.1	37	R356V	nd
17	W330F	0.4 ± 0.1	38	T362L	nd
18	M195G	0.3 ± 0.1	39	Q417G	nd
19	T362V	0.2 ± 0.1	40	Q417S	nd
20	W330Y	0.2 ± 0.1	41	N478T	nd
21	F13Y	nd			

nd – not detected.

Table 4.2. The apparent steady-state kinetic parameters of NOV1 wild-type and the eleven variants show the highest activity for isoeugenol. Activity assays were performed using 100 mM Tris-HCl buffer, pH 9, in the presence of 0.01−4 mM isoeugenol at room temperature. Activity values were normalized by mol Fe^{2+}/mol of protein (%). The results are the mean ± standard error of three independent experiments. Errors for k_{cat}/K_m were obtained by propagation.

Enzymes	**mol Fe^{2+}/mol of protein (%)**	**k_{cat} (s^{-1})**	**K_m (mM)**	**k_{cat}/K_m ($M^{-1}\ s^{-1}$) × 10^3**
Wild-type	68 ± 5	7.0 ± 0.3	0.6 ± 0.1	12 ± 2
M324K	40 ± 5	33 ± 3	1.7 ± 0.3	20 ± 4
L372A	51 ± 4	30 ± 3	1.9 ± 0.3	16 ± 3
N478S	40 ± 3	26 ± 4	2.5 ± 0.3	11 ± 2
L286W	60 ± 5	23 ± 1	1.6 ± 0.2	15 ± 2
I222V	30 ± 4	23 ± 1	1.9 ± 0.2	12 ± 1
T299S	64 ± 4	21 ± 2	1.3 ± 0.3	16 ± 4
L286V	41 ± 2	17 ± 1	1.5 ± 0.2	12 ± 2
L286I	35 ± 2	13 ± 1	1.0 ± 0.1	13 ± 2
T299V	36 ± 3	11 ± 1	1.0 ± 0.2	11 ± 2
I222L	33 ± 2	4.2 ± 0.3	1.2 ± 0.2	3.5 ± 0.6
M374A	40 ± 3	-	-	5 ± 1

4.4.3. Mapping Mutations in the NOV1 X-ray Structure

The analysis of the NOV1 crystal structure of NOV1 (5J53) shows that the six mutations that resulted in the most improved catalytic features are located at distances ≥10 Å from the Fe^{2+} cofactor (**Fig. 4.1A, Table S4.2,** *Annex II*). Except for the mutation M324K at the surface, the remaining mutations are placed in the enzyme's core at the second shell of the non-heme cofactor in proximity to one of the four iron-coordinating histidines**.** For example, T299S and L286W are relatively close to H284 (**Fig. 4.1B**), whereas N478S is near H476 (**Fig. 4.1C**) at distances around 5 Å. The influence of L372A on H476 is probably mediated through its proximity (4 Å) to E418, a conserved glutamate residue that stabilizes H476 (**Fig. 4.1C**). Similarly, I222V is only at 4 Å from K169, stabilizing H167 by H-bond (**Fig. 4.1D**). M324K appears at the walls of the active site entrance tunnel, close to the substrate entrance loops, and can impact the substrate migration pathway (**Fig. 4.1E**). These mutations, or a combination of mutations (please see below), can modulate the activity and stability of the enzyme.

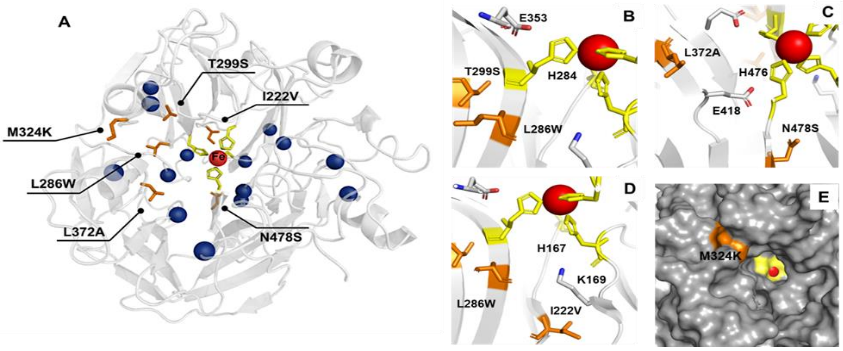


Figure 4.1. (**A**) Cartoon representation of NOV1 crystal structure (PDB code 5J53) showing the distribution of hotspots (blue spheres) and the six best hits (in orange). The four iron-coordinating histidines are in yellow, and their surroundings are highlighted for H284 in (**B**), H476 in (**C**), and H167 in (**D**). The surface visualization of NOV1 in (**E**) shows the only solvent-exposed mutation M324K located close to the access to the substrate pocket in yellow. Fe^{2+} cofactor is shown as a red sphere.

4.4.4. Recombination of Mutations by DNA Shuffling

To combine the best six mutations (I222V, L286W, T299S, M324K, L372A, and N478S) and attempt to further increase the activity towards isoeugenol, the random recombination of genes was performed using DNA shuffling (Stemmer 1994). This excellent in vitro gene recombination technique allows the randomization of gene mutations, concentrating on beneficial mutations. It also facilitates the assessment of potential synergistic conformational effects arising from side-chain orientations, which are unpredictable using single-site mutagenesis. After screening a library of 1700 clones for isoeugenol activity, the best 25 variants were selected for re-screening **(Fig. 4.2)**, and the variants showing 2-fold higher activity than the wild-type were chosen for DNA sequence analysis. The two variants with the highest activity were selected for further studies: variant 12G2, harboring mutations M324K and L372A, and variant 1D2, with mutations I222V, L286W, and L372A. Both variants show comparable optimal activity to the wild-type at pH 9 and 28-32 °C. Notably, these variants showed approximately 7-fold higher k_{cat} than that of the wild-type (**Table 4.3**) and 3-fold the k_{cat} of the best variant (S283F) identified and studied in our previous work (De Simone *et al.* 2023). The catalytic efficiency is only 2-fold higher than that of the wild-type because the K_m values are 4-fold higher; however, the turnover number (k_{cat}) is considered the most important parameter for biotechnological applications since bioprocesses usually take place at high concentrations of substrate, i.e., the enzyme's activity is not limited by substrate concentration. The other identified single mutations and mutation S283F from our previous study were inserted using site-directed mutagenesis in each variant. Still, these substitutions did not increase further the activity.

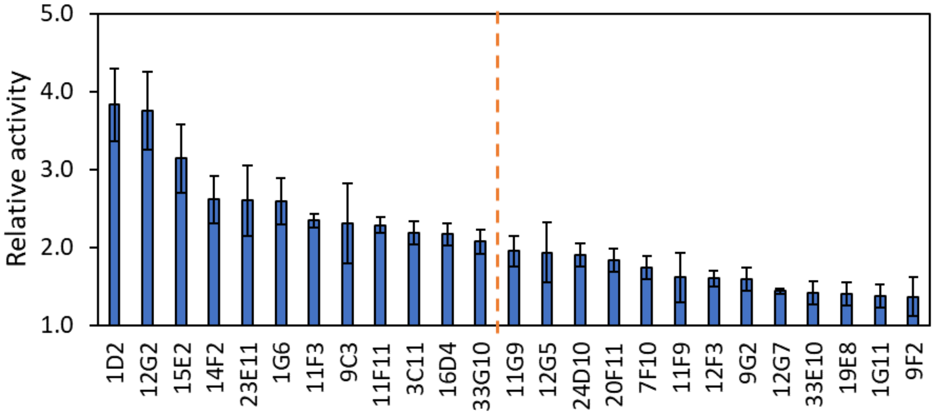


Figure 4.2. Relative activity for isoeugenol (to wild-type) of the twenty-five DNA shuffling variants selected for re-screening. Reactions were performed using 100 mM Tris-HCl buffer, pH 9, in 2 mM isoeugenol at room temperature. The best 12 variants that showed relative activity > 2 (left of red dash line) were selected for DNA sequence analysis.

Table 4.3. Apparent steady-state kinetic parameters of variants 12G2 and 1D2 that resulted from the recombination of genes using DNA shuffling, measured at room temperature in 100 mM Tris-HCl buffer, pH 9 in 0.1-4 mM isoeugenol. The results are the mean ± standard error of three independent experiments. Errors for k_{cat}/K_m were obtained by propagation.

Enzymes	**mol Fe^{2+}/mol of protein (%)**	**k_{cat} (s^{-1})**	**K_m (mM)**	**k_{cat}/K_m (M^{-1} s^{-1}) × 10^3**
Wild-type	68 ± 5	7.0 ± 0.3	0.6 ± 0.1	12 ± 2
12G2 (M324K/L372A)	43 ± 4	53 ± 3	2.4 ± 0.3	22 ± 3
1D2 (I222V/L286W/L372A)	50 ± 5	47 ± 3	2.3 ± 0.3	20 ± 3

Thermal unfolding and kinetic stability of NOV1 variants were assessed. Assuming the denaturation process as a two-state model native (N) ⇌ unfolded (U) (Eftink 1998), the apparent melting temperature (T_m) is defined as the temperature at which half of the protein population is in the unfolded state. The T_m of 12G2 and 1D2 variants are comparable to the wild-type (58

± 1 °C), suggesting the robustness of the enzyme's native structure is maintained (**Table 4.4**). The kinetic or operational stability of an enzyme is helpful to evaluate its resistance to inactivation under specific operating conditions, for example, at a given temperature. It is generally expressed as a half-life time ($t_{1/2}$) (Sanchez-Ruiz 2010). Notably, the study of the kinetic stability at 25 °C showed a half-life time of 50 ± 5 h for 12G2 and 40 ± 5 h for 1D2 (**Fig. 4.3, Table 4.4**), which were significantly higher than that of the wild-type (1.3 ± 0.3) or of any other studied NOV1 variant (variant S283F shows a half-life of 29 ± 3 h (De Simone *et al.* 2023)). Our previous results indicated that the primary molecular determinant of NOV1 kinetic stability is the depletion of iron cofactor from the active site during the enzyme incubation at 25 °C. Therefore, the enhanced kinetic stability of both 12G2 and 1D2 variants may result from stabilizing the iron cofactor, hypothetically, like what was observed with S283F, throughout a conformational effect between the distal mutations and the active site.

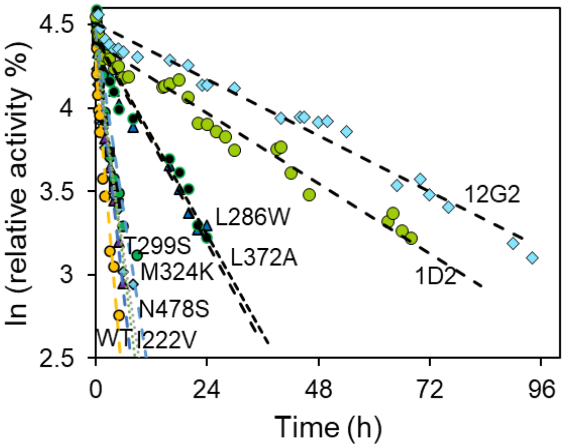


Figure 4.3. Kinetic stability of wild-type and NOV1 variants at 25 °C. Data are shown as linear regression.

Table 4.4. Thermal stability of wild-type and variant enzymes as monitored by thermal inactivation at 25 °C and by fluorescence emission at 330 nm. Half-life at 25 °C is the time after which 50% of activity is achieved. T_m is the melting temperature at which 50% of the protein molecules are unfolded. The results are the mean ± standard error of three independent experiments for kinetic stability and two for thermal unfolding.

Enzymes	**Half-life at 25 °C (h)**	**T_m (°C)**
Wild-type	1.7 ± 0.4	58 ± 1
12G2 (M324K/L372A)	50 ± 5	58 ± 2
1D2 (I222V/L286W/L372A)	40 ± 5	58 ± 1
L372A	14 ± 2	58
L286W	14 ± 2	56
M324K	4 ± 1	55
T299S	3 ± 1	53
N478S	3 ± 1	55
I222V	3 ± 1	53

4.4.5. Affinity Immobilization of Wild-type and Variants

In this study, NOV1 wild-type and variants were produced with a SUMO-tag at the N-terminal to promote higher protein solubility and facilitate purification by affinity chromatography. The His-tag allowed binding to Chromalite® MIDA/M resin loaded with Ni^{2+}, Co^{2+}, Zn^{2+}, Cu^{2+} and Fe^{2+} metals (**Fig. 4.4**). The long distance between the His-tag and the active site can provide a good orientation for catalysis, eventually resulting in good enzymatic activity and stability compared to other immobilization methods (Basso *et al.* 2022). The immobilization yield (defined as the percentage of protein bound to the resin) was above 80% for all resins except Fe^{2+}-loaded resin, resulting in the lowest (20-30%). On the other hand, although the immobilization yield on Cu^{2+} was 99%, most of the protein is inactive, resulting in the lowest immobilization efficiency (20-40%). Previous works suggested that different metal ions exhibit varying His-tag selectivity: $Cu^{2+} < Ni^{2+} < Co^{2+} < Zn^{2+} < Fe^{2+}$, influencing the amount and the way enzymes are immobilized (Basso *et al.* 2022; Winand *et al.* 2023). Accordingly, the lower selectivity of resins with Cu^{2+} putatively promotes unspecific interactions with protein surface impairing activity, while the higher selectivity of resins with Fe^{2+} most likely results in a weaker protein binding.

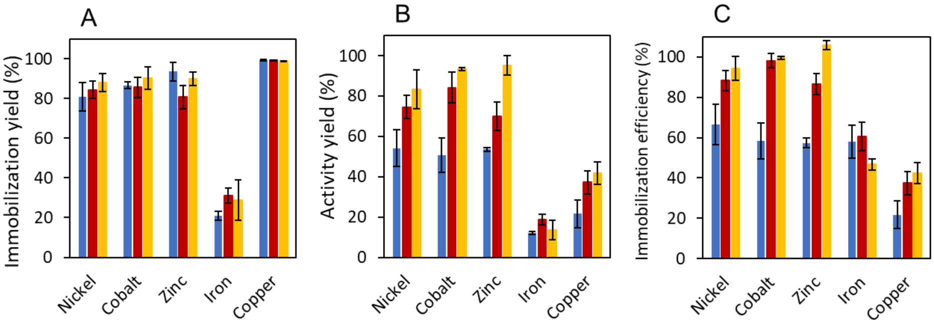


Figure 4.4. Immobilization yield (**A**), Activity yield (**B**) and Immobilization efficiency (**C**) of NOV1 wild-type (blue), 12G2 (red), and 1D2 (orange) immobilized onto Chromalite™ MIDA/M loaded with Nickel, Cobalt, Zinc, Iron, and Copper metals. The immobilization was performed by incubating 10 mg protein/g resin in 50 mM sodium phosphate buffer pH 8 for 1 h at 25 °C.

Wild-type and both variants show similar and high immobilization yields in Ni^{2+}, Co^{2+}, and Zn^{2+} loaded resins; however, a significantly higher immobilized activity yield (% of immobilized enzyme activity to initial activity) was measured in variants than in the wild-type. Accordingly, the immobilization efficiency (ratio of immobilization yield to immobilization activity) was almost double for the variants (~ 90%) when compared to the wild-type (~50%). The reduced efficiency of the wild-type can be attributed to its shorter half-life (1.7 ± 0.4 h) under the immobilization conditions (1 hour at 25 °C; see **Table 4.4**). We increased the protein from 10 mg to 20 mg per gram of resin immobilized, observing slightly lower absorption at higher capacities. The stability of enzymes increased after immobilization and storage at room temperature (**Fig. 4.5**). The half-life of wild-type increased from 1.7 h (free enzyme) (**Table 4.4**) to 32 ± 4 h (immobilized). In contrast, for NOV1 variants, immobilization extended the half-lives to 59 ± 4 h for 12G2 and 57 ± 3 h for 1D2.

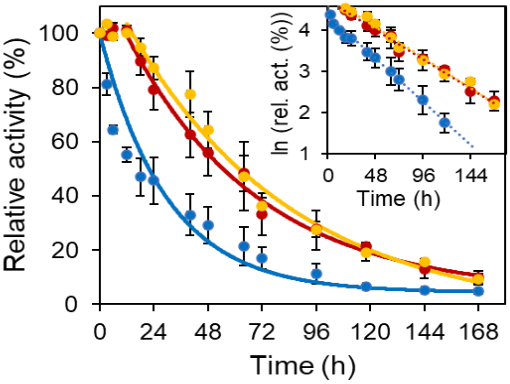


Figure 4.5. Kinetic stability at room temperature of immobilized NOV1 wild-type (blue), 12G2 (red), and 1D2 (yellow). Inset: linear regression of logarithm activity versus time. The half-lives at room temperature were 32 ± 4, 59 ± 4, and 57 ± 3 for wild-type, 12G2 and 1D2, respectively.

4.4.6. Bioconversion of Isoeugenol into Vanillin using Immobilized Enzymes

We tested the recyclability of the immobilization systems by measuring the conversion yields of isoeugenol (50 mM) into vanillin after up to ten reaction cycles of 24 h each (**Fig. 4.6, Table 4.5**). In the first cycle, 1D2 and 12G2 converted 95% and 62% of isoeugenol to vanillin, respectively, compared to 38% for the wild-type. The conversion observed during the first cycle is comparable to and even better than what was observed for whole-cell reaction systems (De Simone *et al.* 2023; Han, Long, and Ding 2019; Zhao *et al.* 2019; Zhao *et al.* 2018). The lower performance of 12G2 compared to

1D2 most likely reflects the lower iron loading (**Table 4.3**) and activity immobilization yield (**Fig. 4.4B**). Variant 1D2 maintained the highest conversion along the 10 cycles, producing 164 mM vanillin (~25 mg) at the 1 mL scale, which corresponds to a 33% molar conversion against the 23% and 15% molar conversion of 12G2 and wild-type (**Fig. 4.6, Table 4.5**). However, after the seventh cycle, substrate conversion is residual (< 10%) for the three enzymes, probably due to enzyme inactivation (**Fig. 4.5**). Enzyme immobilization offers advantages over whole-cell reactions by addressing product instability and biosafety. These important issues need to be considered, for example, food application (Liu *et al.* 2023). The 1D2 variant immobilized on MIDA/M/Ni^{2+}, produced ~ 4-fold higher amounts of vanillin using lower amounts of protein as compared to a previously reported system using Cso2 oxygenase (De Vitis *et al.* 2023) immobilized in Sepabeads EC-EA, representing an efficient alternative in the production of vanillin (Furuya, Kuroiwa, and Kino 2017).

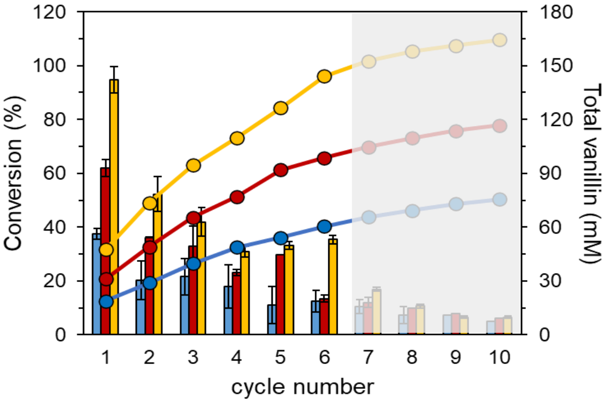


Figure 4.6. Isoeugenol conversion to vanillin (%) by immobilized wild-type (blue bars) and variants 12G2 (red bars) and 1D2 (yellow bars) in a 24 h-stepwise 50 mM isoeugenol-addition bioprocess at pH 9 and room temperature. The total amount of vanillin (mM) produced is also shown for wild-type (blue dots), 12G2 (red dots), and 1D2 (blue dots).

Table 4.5. Total vanillin yield (mM) and corresponding molar conversion (%) obtained by immobilized wild-type, 12G2 and 1D2 for each cycle reaction. Immobilized protein was resuspended in 100 mM Tris-HCl, pH 9, with 50 mM isoeugenol. After 24 h, reaction mixtures were transferred to an Eppendorf tube and centrifuged for 2 min, at 12,000 rpm, to recover the supernatant. The immobilized enzymes were washed with buffer, centrifuged, and resuspended with fresh reaction mixtures to start new conversion cycles. The results are the mean ± standard error of three independent experiments.

		Wild-type			12G2			1D2		
		Per cycle			Per cycle			Per cycle		
Cycle #	Substrate (mM)	Conversion (%)	Vanillin (mM)	Total (mM)	Conversion (%)	Vanillin (mM)	Total (mM)	Conversion (%)	Vanillin (mM)	Total (mM)
1	50	38 ± 2	19 ± 1		62 ± 3	31 ± 2		95 ± 3	47 ± 1	
2	50	20 ± 7	10 ± 4	29	36 ± 2	18 ± 0	49	52 ± 3	26 ± 2	74
3	50	22 ± 7	11 ± 3	40	33 ± 8	16 ± 4	66	42 ± 3	21 ± 1	94
4	50	18 ± 8	9 ± 4	49	23 ± 1	12 ± 1	77	31 ± 1	15 ± 1	110
5	50	11 ± 7	6 ± 3	54	30 ± 0	15 ± 0	92	33 ± 1	17 ± 1	127
6	50	12 ± 4	6 ± 2	60	13 ± 1	7 ± 1	99	35 ± 1	18 ± 1	144
7	50	10 ± 3	5 ± 1	66	12 ± 2	6 ± 1	105	17 ± 0	8 ± 0	153
8	50	7 ± 3	4 ± 2	69	10 ± 0	5 ± 0	110	11 ± 0	5 ± 0	158
9	50	7 ± 0	4 ± 0	73	8 ± 0	4 ± 0	114	6 ± 0	3 ± 0	161
10	50	5 ± 0	3 ± 0	75	6 ± 0	3 ± 0	117	7 ± 0	3 ± 0	164
	500			(15%)			(23%)			(33%)

4.5. CONCLUDING REMARKS

In summary, this work underscores the importance of distal interactions in enhancing enzymatic properties, emphasizing the value of distal-mutation-based protein engineering strategies. Using the computational tool Zymspot, we identified and mutated 41 distal residues and constructed two variants with two and three distal mutations, achieving nearly a 10-fold increase in activity and up to 40-fold higher operational stability than the wild-type. We demonstrated the successful immobilization of these enhanced variant enzymes in MIDA/M/Ni^{2+}. This protocol is straightforward and effective for immobilizing an isoeugenol dioxygenase, providing a higher immobilization efficiency (89-99%) and conversion yields (95%) than previously reposted methods. When using these enzymes in reaction mixtures containing 50 mM of isoeugenol, 35% conversion rates were achieved after six reaction cycles, resulting in a total yield of 144 mg vanillin. Our findings provide a proof of concept for the immobilization of computationally optimized enzyme variants, offering significant improvements in enzyme efficiency and stability. Further development and optimization of these bioprocesses at a larger scale would undoubtedly lead to the economically attractive production of natural vanillin from lignocellulose waste products, such as isoeugenol as a precursor.

4.6 REFERENCES

Barbosa, O., C. Ortiz, A. Berenguer-Murcia, R. Torres, R. C. Rodrigues, and R. Fernandez-Lafuente. 2015. 'Strategies for the one-step immobilization-purification of enzymes as industrial biocatalysts', *Biotechnol. Adv.* 33: 435-56.

Basso, A., M. S. Brown, A. Cruz-Izquierdo, C. A. Martinez, and S. Serban. 2022. 'Optimization of metal affinity ketoreductase immobilization for application in batch and flow processes', *Org. Process Res. Dev.*

Bornscheuer, U. T., G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins. 2012. 'Engineering the third wave of biocatalysis', *Nature* 485: 185-94.

Cagiada, M., S. Bottaro, S. Lindemose, S. M. Schenstrom, A. Stein, R. Hartmann-Petersen, and K. Lindorff-Larsen. 2023. 'Discovering functionally important sites in proteins', *Nat. Commun.* 14.

Casilli, F., M. Canyelles-Nino, G. Roelfes, and L. Alonso-Cotchico. 2024. 'Computation-guided engineering of distal mutations in an artificial enzyme', *Faraday Discuss.*

De Simone, M., L. Alvigini, L. Alonso-Cotchico, V. Brissos, J. Caroli, M. F. Lucas, E. Monza, E. P. Melo, A. Mattevi, and L. O. Martins. 2023. 'Rationally guided improvement of NOV1 dioxygenase for the conversion of lignin-derived isoeugenol to vanillin', *Biochemistry* 62: 419-28.

De Vitis, V., P. Cannazza, L. Mattio, D. Romano, A. Pinto, F. Molinari, T. Laurenzi, I. Eberini, and M. L. Contente. 2023. 'Caulobacter segnis dioxygenase CsO2: a practical biocatalyst for stilbenoid ozonolysis', *ChemBioChem* 24: e202300477.

Eftink, M. R. 1998. 'The use of fluorescence methods to monitor unfolding transitions in proteins', *Biochemistry (Mosc)* 63: 276-84.

Es, I., J. D. Vieira, and A. C. Amaral. 2015. 'Principles, techniques, and applications of biocatalyst immobilization for industrial application', *Appl. Microbiol. Biotechnol.* 99: 2065-82.

Fischer, D. S., and D. C. Price. 1964. 'A simple serum iron method using the new sensitive chromogen tripyridyl-s-triazine', *Clin. Chem.* 10: 21-31.

Furuya, T., M. Kuroiwa, and K. Kino. 2017. 'Biotechnological production of vanillin using immobilized enzymes', *J. Biotechnol.* 243: 25-28.

Gu, J., Y. Xu, and Y. Nie. 2023. 'Role of distal sites in enzyme engineering', *Biotechnol. Adv.* 63: 108094.

Han, Z. C., L. K. Long, and S. J. Ding. 2019. 'Expression and characterization of carotenoid cleavage oxygenases from *Herbaspirillum seropedicae* and *Rhodobacteraceae* bacterium capable of biotransforming isoeugenol and 4-vinylguaiacol to vanillin', *Front. Microbiol.* 10: 1-11.

Lee, J., and N. M. Goodey. 2011. 'Catalytic contributions from remote regions of enzyme structure', *Chem. Rev.* 111: 7595-624.

Liu, Y., L. Sun, Y. X. Huo, and S. Guo. 2023. 'Strategies for improving the production of bio-based vanillin', *Microb. Cell Fact.* 22: 147.

Motlagh, H. N., J. O. Wrabl, J. Li, and V. J. Hilser. 2014. 'The ensemble nature of allostery', *Nature* 508: 331-9.

Osuna, S. 2021. 'The challenge of predicting distal active site mutations in computational enzyme design', *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 11.

Ryu, J. Y., J. Seo, S. Park, J. H. Ahn, Y. Chong, M. J. Sadowsky, and H. G. Hur. 2013. 'Characterization of an isoeugenol monooxygenase (Iem) from *Pseudomonas nitroreducens* Jin1 that transforms isoeugenol to vanillin', *Biosci. Biotechnol. Biochem.* 77: 289-94.Sanchez-Ruiz, J. M. 2010. 'Protein kinetic stability', *Biophysical Chemistry*, 148: 1-15.

Sheldon, R. A., and S. van Pelt. 2013. 'Enzyme immobilisation in biocatalysis: why, what and how', *Chem. Soc. Rev.* 42: 6223-35.

Stemmer, W. P. C. 1994. 'DNA shuffling by random fragmentation and reassembly - in-vitro recombination for molecular evolution', *Proc. Natl. Acad. Sci. USA* 91: 10747-51.

Wang, J., A. Jain, L. R. McDonald, C. Gambogi, A. L. Lee, and N. V. Dokholyan. 2020. 'Mapping allosteric communications within individual proteins', *Nat. Commun.* 11.

Wilding, M., N. Hong, M. Spence, A. M. Buckle, and C. J. Jackson. 2019. 'Protein engineering: the potential of remote mutations', *Biochem. Soc. Trans.* 47: 701-11.

Winand, L., S. Theisen, S. Lütz, K. Rosenthal, and M. Nett. 2023. 'Immobilization of the amidohydrolase MxcM and its application for biocatalytic flow synthesis of pseudochelin A', *Catalysts* 13.

Yamada, M., Y. Okada, T. Yoshida, and T. Nagasawa. 2007. 'Purification, characterization and gene cloning of isoeugenol-degrading enzyme from *Pseudomonas putida* IE27', *Arch. Microbiol.* 187: 511-17.

Yu, H. R., S. Ma, Y. W. Li, and P. A. Dalby. 2022. 'Hot spots-making directed evolution easier', *Biotechnol. Adv.* 56.

Zhao, L., Y. Jiang, H. Fang, H. Zhang, S. Cheng, M. S. R. Rajoka, and Y. Wu. 2019. 'Biotransformation of isoeugenol into vanillin using immobilized recombinant cells containing isoeugenol monooxygenase active aggregates', *Appl. Biochem. Biotechnol.* 189: 448-58.

Zhao, L. Q., Y. M. Xie, L. Y. Chen, X. F. Xu, C. X. Zha, and F. Cheng. 2018. 'Efficient biotransformation of isoeugenol to vanillin in recombinant strains of *E. coli* by using engineered isoeugenol monooxygenase and sol-gel chitosan membrane', *Process Biochem.* 71: 76-81.

Zymvol-Biomodeling. 'Zymspot: A ZYMVOL computational tool for identifying distal mutations', https://zymvol.com/enzyme-technology/.

Chapter 5

Computational Engineering a Streptomyces sclerotialus Peroxygenase for Improved Affinity Immobilization

Author contributions:

Nicholus Bhattacharjee and Lur Alonso-Cotchico (Zymvol-Biomodeling) conceptualized this work, developed Zymspot algorithm and used it to predict the optimal tag sites to immobilize *Ss*TyrH peroxygenase. Construction and purification of all Lys-tag variants and His-tag variants M11, M12 and M13 was conducted at GECCO Biotech (Groningen, Netherlands) under supervision of Nikola Loncar and Hugo van Beek. The author of this thesis performed construction and purification of His-tag variants M14-20. All biochemical characterization and immobilization experiments were also performed by thesis author. Prof. Ligia O. Martins supervised and reviewed this work.

5.1. ABSTRACT

Enzyme immobilization is pivotal in industrial applications, offering advantages such as increased enzyme stability and recyclability. However, achieving consistently active and robust immobilized biocatalysts remains challenging due to the empirical methodologies used. Strategies such as adding fusion peptide tags for affinity immobilization have emerged as potential solutions.

This study developed a new computational tool to predict optimal tag sites for enzyme immobilization. Validation was performed using a peroxygenase, a heme-dependent L-tyrosine hydroxylase from *Streptomyces sclerotialus* (*Ss*TyrH), which has enormous potential for industrial applications, including L-DOPA production. Two immobilization strategies were computationally designed: (*i*) Lys-tags for covalent immobilization and (*ii*) His-tags for affinity immobilization. A total of 20 variants were constructed and produced, and peroxidase activity assays were used to assess enzyme immobilization. Our results demonstrate the effectiveness of the computational tool, particularly in predicting His-tagged sites for enzyme immobilization. Replacement or addition of four His are sufficient at specific resulted in the highest activity before immobilization and retained most of it afterward.

Keywords: Biocatalysis, Heme peroxidase, Protein Engineering, Immobilization tag design

5.2. INTRODUCTION

Enzyme immobilization is a key technology for numerous industrial applications, offering advantages such as increased stabilization, easier enzyme separation from the reaction mixtures, and enhanced recycling potential, contributing to greater efficiency and cost-effectiveness of enzymatic processes in the industry (Barbosa *et al.* 2015; Basso *et al.* 2022; Datta, Christena, and Rajaram 2013; Mateo *et al.* 2007). However, immobilization science remains empirical, making it difficult to consistently achieve efficient protocols through general and rational approaches (Bhattacharjee, Alonso-Cotchico, and Lucas 2023). Although enzyme immobilization provides numerous benefits, specific unavoidable conformational changes may negatively impact enzyme properties (Hernandez and Fernandez-Lafuente 2011). For example, enzyme orientation is critical for successful immobilization (Zeballos *et al.* 2023); uncontrolled immobilization can lead to unproductive binding, where enzymes are attached in unfavorable orientations or adopt excessively rigid conformations, ultimately resulting in a complete loss of activity (Bolivar, Eisl, and Nidetzky 2016; Hernandez and Fernandez-Lafuente 2011; Zhang, Ge, and Liu 2015). This issue has sparked interest, and peptide tags have become a popular strategy for achieving oriented immobilization (Costa et al. 2014; Young, Britton, and Robinson 2012). However, the fusion of peptide tags and protein domains is typically restricted to the N- and C-termini of the target enzyme, which limits the possible orientations (Paraskevopoulou and Falcone 2018).

In recent years, computational methods have been employed to better understand the structural and dynamic aspects of enzyme immobilization strategies (Bhattacharjee, Alonso-Cotchico, and Lucas 2023). These simulations have allowed for rationalizing the molecular aspects of immobilization events. Still, their use in predicting optimal protocols to control the impact on enzyme properties remains largely unexplored. Therefore, developing computational tools capable of predicting optimal immobilization sites in target proteins is critically urgent.

Zymvol recently developed Zymlock (Zymvol-Biomodeling), an algorithm to predict optimal tag sites based on the following criteria of the target enzyme: (*i*) features of the secondary structure, (*ii*) site exposure to the solvent, (*iii*) distance from the active site, and (*iv*) residue conservation. This new tool was validated by using peroxygenases, i.e. a tyrosine hydroxylase from *Streptomyces sclerotialus* (*Ss*TyrH); heme-dependent L-tyrosine hydroxylases (TyrHs) that constitute a new enzyme family which aroused interest due to its ability to produce L-3,4-dihydroxyphenylalanine (L-DOPA) through hydroxylation of tyrosine (Singh *et al.* 2023; Wang *et al.* 2021). Members of this family, including *Ss*TyrH, also display a remarkable capacity to cleave carbon–fluorine (C–F) bonds, which are among the strongest in chemistry (Seong *et al.* 2019; Tong et al. 2019; Wang and Liu 2020). This makes TyrHs promising biocatalysts to address the biodegradation of fluorinated hydrocarbons, whose degradation is particularly difficult due to the exceptional stability of these bonds (Cui et al. 2018). *Ss*TyrH crystal structure (pdb: 7KQR) revealed a folding organized in eleven α-helices (α1 to α11), two short ß-strands (ß1 and ß2), and four 3_{10} helices (η1 to η4) (**Fig. 5.1**). A b-type heme is located near the protein surface, with Fe^{2+} coordinated by a conserved axial histidine (H196), nearly perpendicular to the heme plane. Arg172, Ser157, and Ser209 further stabilize the heme through hydrogen-bonding interactions with the propionate groups (Wang *et al.* 2021).

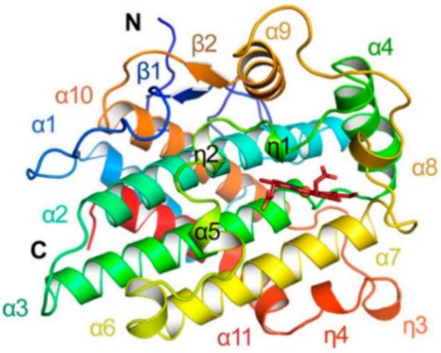


Figure 5.1. *Ss*TyrH structure (pdb: 7KQR) coloured by a rainbow spectrum (N-terminus, blue; C-terminus, red). The molecule is composed of 11 α-helices, two β-strands and four 3_{10} helices (Wang *et al.* 2021)

In this work, tags for the immobilization of *Ss*TyrH were computationally designed with two distinct goals: 1) Lys-tags for optimal covalent immobilization on an epoxy/butyl methacrylate carrier, and 2) His-tags for affinity immobilization n on an iminodiacetic acid methacrylate carrier loaded with Ni^{2+}. Each of these approaches offers advantages and limitations. Covalent bonding provides a stable binding of the enzyme to the carrier (Wong, Thirlway, and Micklefield 2008), resulting in durable enzymes with a low probability of leakage (Maghraby *et al.* 2023); however, the enzyme is frequently irreversibly deactivated and the enzyme and the often expensive support material become unusable. Moreover, long incubation times are usually required for proper immobilization, and covalent immobilization can limit the enzyme's flexibility, potentially reducing its catalytic efficiency (Maghraby *et al.* 2023; Shen *et al.* 2011). In contrast, the His-tag provides a weaker binding but offers selective and oriented immobilization, minimal structural alterations to the enzyme, and reloading the same surface with the enzyme multiple times, enabling support recycling (Andreescu, Bucur, and Marty 2006). Here, a new computational design led to the construction of new peroxygenase variants with Lys- and His-tags. These variants were immobilized and tested for peroxidase activity using 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) as substrate. Our results demonstrated the potential of the computational algorithm in predicting optimal immobilization sites, particularly for His-tagged enzymes, paving the way for a more rational and efficient approach to enzyme immobilization science.

5.3. MATERIALS AND METHODS

5.3.1. Design of Variants

The Zymlock algorithm was developed by Zymvol, Barcelona (Spain) (https://zymvol.com/) and tested on *Ss*TyrH using a crystal structure model (pdb: 7KQR). This tool aims at categorizing distinct regions of the enzyme sequence based on properties that would ultimately avoid perturbation of the enzyme function, ensuring an optimal orientation as well for catalysis. These properties comprise: *i*) the nature of the secondary structure, which benefits the presence of loops; *ii*) solvent accessibility, excluding buried sites and prioritizing surface-exposed regions to facilitate proper immobilization; *iii*) distance and relative orientation of the studied region with respect to the active site; and *iv*) residue conservation, excluding conserved regions to avoid disrupting essential structural or functional elements.

5.3.2. Construction of Variants

The genes coding for variants M1-13 were synthesized and cloned in the pBAD vector. The His-tags of mutants M14-M20 were instead introduced by PCR reaction using Q5® Site-Directed Mutagenesis kit. Primers were designed using the neb base changer online tool (https://nebasechanger.neb.com/) (**Table S5.1,** *Annex III*). The PCR reaction was performed using the wild-type gene cloned in pBAD plasmid following these conditions: 25 µL reaction volume containing 2.5 ng of DNA wild-type template, 200 µM dNTPs, Q5 polymerase buffer and 2.5 U of Q5® High-Fidelity DNA polymerase (New England Biolabs). After an initial denaturation period of 1 min at 98 °C, the subsequent steps were repeated for 25 cycles in a thermal cycler: 30 s at 98 °C, 20 s at 72 °C, 3 min at 72 °C, followed by a final 5 min period at 72 °C. The amplified material is added directly to a unique Kinase-Ligase-DpnI (KLD) enzyme mix at room temperature for 1 h to allow circularization and template removal. After DpnI inactivation by incubation at 80 °C for 20 min, the sample was used to transform an aliquot of electrocompetent NEB10β *E. coli* cells. This resulted in single and good-shaped colonies being picked and pre-inoculated in 10 mL of LB media and growing overnight. Plasmid purification was performed by GeneJet plasmid MiniprepKit using 8 mL of pre-inocula. The presence of the desired mutation was confirmed by Sanger sequencing.

5.3.3. Production and Purification

Bacterial growth for protein production was performed in 400 mL LB medium supplemented with 50 µg/mL ampicillin at 37 °C, 120 rpm in a baffled Elernmeyer flask. After reaching 0.8 OD, the expression was induced by adding 0.02% L-arabinose; 80 µm Hemin was supplemented to promote the production of Heme-loaded protein, and the temperature was decreased to 24 °C. After 16 h, cells were collected by centrifugation, resuspended in 50 mM potassium phosphate, 0.5 M NaCl, 30 mM imidazole, pH 7.5, and disrupted by three cycles at French press (9000 psi). Purification was performed by affinity chromatography in the His-trap column, eluting the protein of interest by 50-70% of 50 mM potassium phosphate, 0.5 M NaCl, and 500 mM imidazole pH 7.5. Fractions showing enzymatic activity were collected and analyzed using SDS-PAGE. Cleavage of the His-tag at the N-terminal was performed overnight by incubating 1 mg/mL of protein in the presence of 0.5 U/µl of Thrombin at 20 °C. Washes removed His-tag fragments in 50 mM potassium phosphate, 0.5 M NaCl, pH 7.5 in the PD-10 column. Activity measurements were conducted in 50 mM potassium phosphate buffer, pH 7, using 0.1-10 mM ABTS and 1 mM H_2O_2. Heme concentration in purified protein preparations was assessed by measuring the absorbance peak at 404 nm (ε_{404} = 58.44 mM^{-1} cm^{-1}) (Deniau *et al.* 2003).

5.3.4. Biochemical Characterization

The effect of pH on the activity of *Ss*TyrH was studied in 100 mM Britton-Robinson buffer (pH 3-11). Reactions were performed in 4 mM ABTS and 1 mM H_2O_2 and monitored at 420 nm (see above). Kinetic parameters for ABTS substrate were calculated from 0-10 mM following the absorbance increase at 420 nm using 1 mM H_2O_2 in 50 mM, potassium phosphate buffer, pH 7 at room temperature. Kinetic parameters for H_2O_2 activity were measured in the 0-4 mM H_2O_2 range in 4 mM ABTS. Kinetic data were fitted directly to the Michaelis-Menten equation using the Origin® software. All enzymatic assays were performed at least in triplicate.

5.3.5. Immobilization on Metal-loaded Resin by Affinity Interaction

The purified *Ss*TyrH wild-type and variants M1-M10 were immobilized on Purolite^{TM} Epoxy/butyl methacrylate resin (ECR8285). Resins were first equilibrated, washing twice with 1 mL of 50 mM sodium phosphate, pH 8. The immobilization was performed by incubating 0.1 mg protein/10 mg dry resin in the same buffer at 25 °C, under rotary mixing at 15 rpm (Hybrid Mini 10 incubator) for 2, 16 and 24 h, in a total volume of 0.5 mL. After immobilization, resins were centrifuged, the supernatants were collected, and the carrier was washed with 0.5 mL of buffer. The total free protein was quantified using the Bradford method, and the amount of enzyme immobilized enzyme was calculated from a mass balance. The immobilization yield (1) was determined to evaluate the immobilization process. The immobilization of His-tag variants M11-M20, along with wild-type enzymes both with and without an N-terminal His-tag, was carried out on Chromalite™ MIDA/M resins loaded with Ni^{2+} (LS04058-186), as previously described (De Simone et al. 2024).

$$(1) \quad Immobilization\ yield\ (\%) = 100 * \left(\frac{P1-P2}{P1}\right)$$

P1 is the initial protein concentration, and P2 is the free protein concentration after immobilization in the supernatant and washing.

The immobilized enzymes were resuspended in 0.5 mL of buffer and used in enzymatic assays to measure the specific activity and compared with that of the free enzyme to calculate the activity yield (2) and immobilization efficiency (3). In the case of Lys-tag variants, the immobilized enzyme samples were resuspended in a 1 mL mix reaction containing 4 mM ABTS, 1 mM H_2O_2, 50 mM potassium phosphate buffer, pH 7. Regarding His-tag variants, reactions were started by adding 20-100 µl of immobilized enzyme sample to a final volume of 1 mL reaction. Reactions were carried on in a 96-deep well plate on a shaker at 1000 rpm at room temperature. Reactions were followed, taking aliquots over time and reading ABTS absorbance at 420 nm ($\varepsilon_{420} = 3.6 \times 10^4\ M^{-1}\ cm^{-1}$).

$$(2) \quad Activity\ Yield\ (\%) = 100 * \left(\frac{immobilized\ activity}{initial\ activity}\right)$$

$$(3) \quad Immobilization\ Efficiency\ (\%) = 100 * \left(\frac{Activity\ Yield}{Immobilization\ Yield}\right)$$

5.4. RESULTS AND DISCUSSION

5.4.1. Design and Construction of the Mutants

For the Lys-tag variants, five sites were considered that showed the highest score from the algorithm: the four mutations S43K, R104K, R180K, and R240K, in addition to the native K175 **(Fig. 5.2A)**. Furthermore, we identified Lys residues on the enzyme surface that could interfere with the enzyme's correct orientation: K117 and K174; therefore, these two residues were replaced with arginine, while K165 was replaced with arginine or an alanine. This resulted in 10 Lys-tag variants (M1-M10; **Table 5.1**).

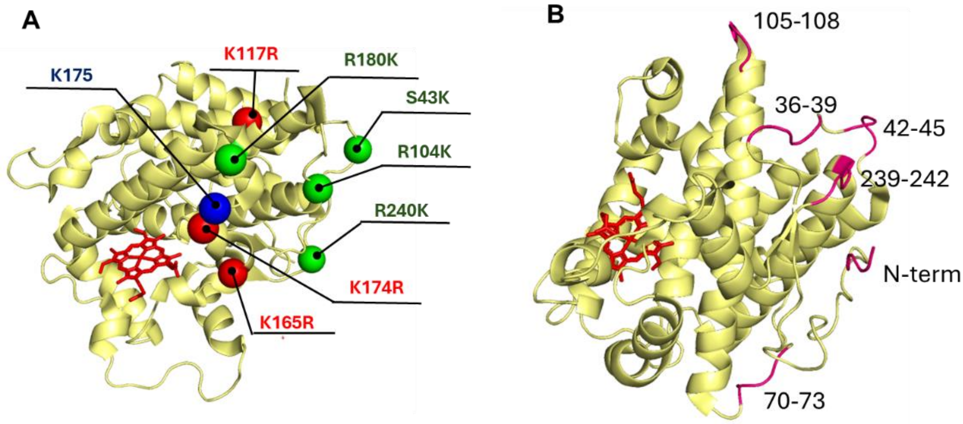


Fig. 5.2. *Ss*TyrH crystal structure (pdb: 7KQR) showing the distribution of **A)** mutations for immobilization on epoxy resin. In green, mutations to Lys to immobilize on epoxy resin; in red, native Lys that were mutated to not interfere with binding; and in blue, native Lys175. **B)** Loops (in magenta) selected to insert a His-tag immobilization on metal-loaded resin. The heme group is highlighted in red.

Five solvent-exposed loops distant from the heme for the His-tag variants were proposed as potential tag sites: 36–39, 42–45, 70–73, 105–108, and 239–242 (**Fig. 5.2 B**). For each loop, the insertion of a His-tag, consisting of 4 or 6 histidines, was suggested by substitution or insertion, resulting in 10 variants (M11-M20; **Table 5.1**). Additionally, wild-type with an N-terminal His-tag, which corresponds to the most trivial design, was used as a benchmark. All variants were produced with a 6×His-tag at their N-termini to enable purification. After purification, the N-terminal His-tag of variants M11-20 was proteolyzed to avoid interference with the other tags.

Table 5.1. List of variants and corresponding mutations. Lysine and histidine residues supposed to bind to the carrier were highlighted in bold.

Variant	Mutations
Lys tag variant	
M01	K117R_K165R_K174R_K175R_S43**K**
M02	K117R_K165A_K174R_K175R_S43**K**
M03	K117R_K165R_K174R_K175R_R104**K**
M04	K117R_K165A_K174R_K175R_R104**K**
M05	K117R_K165R_K174R_K175R_R240**K**
M06	K117R_K165A_K174R_K175R_R240**K**
M07	K117R_K165R_K174R_K175R_R180**K**
M08	K117R_K165A_K174R_K175R_R180**K**
M09	K117R_K165R_K174R_**K**175
M10	K117R_K165A_K174R_**K**175
His tag variants	
M11	D42**H**_S43**H**_G44**H**_S45**H**
M12	43-**HHHH**-44
M13	43-**HHHHHH**-44
M14	38-**HHHHHH**-39
M15	106-**HHHHHH**-107
M16	240-**HHHHHH**-241
M17	G36**H**_S37**H**_L38**H**_E39**H**
M18	V70**H**_Q71**H**_P72**H**_A73**H**
M19	L105**H**_P106**H**_E107**H**_E108**H**
M20	S239**H**_R240**H**_P241**H**_S242**H**

5.4.2. Biochemical Characterization

All variants were expressed and produced in the soluble fraction in recombinant *E. coli* host cells. *Ss*TyrH showed a bell-shaped pH profile with an optimal value of around 7 (**Fig. 5.3A**). The k_{cat} for ABTS is 3.4 ± 0.1 s^{-1}, with a K_m of 1.3 ± 0.1 mM (**Fig. 5.3B, Table 5.2)**. Analogously, using saturating concentrations of 4 mM ABTS, we estimated a k_{cat} for H_2O_2 of 3.4 ± 0.2 s^{-1}, with a K_m of 0.4 ± 0.1 mM (**Fig. 5.3C, Table 5.2)**. At concentrations higher than 1 mM, H_2O_2 enzyme inactivation occurs. Notably, all variants exhibited activity comparable to the wild-type, demonstrating the software's ability to identify sites permissive to mutations, including insertions, without affecting enzymatic activity (**Table 5.2**). Introducing a His-tag in loops 36-39, 105-108, and 239-242 resulted in 2-3-fold lower K_m than wild-type (see mutants M14–17 and M20), implying allosteric interactions between loops and the active site. Hypothetically, introducing histidines in these loops expanded the electropositive surface of the enzyme, facilitating the recruitment of the negatively charged ABTS molecules from the solvent.

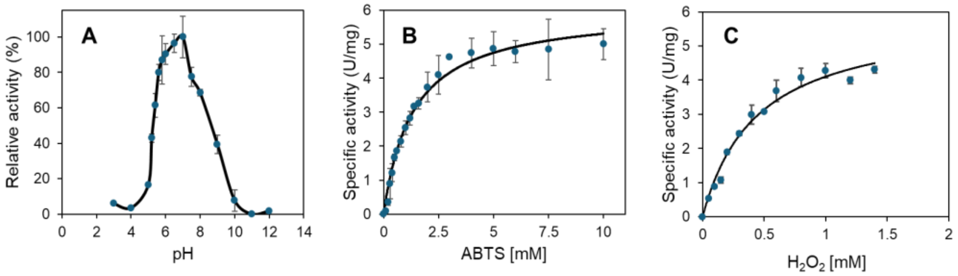


Fig. 5.3. (**A**) pH profile of *Ss*TyrH wild type. Reactions were performed using Britton Robinson buffer (pH range from 2 to 12), 4 mM ABTS, 1 mM H_2O_2 at room temperature. (**B**) Steady-state kinetic analysis of *Ss*TyrH wild type for ABTS in concentrations from 0 to 10 mM, in 1 mM H_2O_2. (**C**) Steady-state kinetic analysis for H_2O_2 using concentrations from 0 to 2 mM in 4 mM ABTS. Reactions were performed in 50 mM KPi buffer, pH 7. Activity was monitored at 420 nm (ε_{420} = 3.6 × 10^4 M^{-1} cm^{-1}). Kinetic data were fitted with the Michaelis-Menten equation using Origin software.

Table 5.2. Kinetic parameters of *Ss*TyrH wild-type and variants for ABTS. Reactions were performed in 1 mM H_2O_2, 50 mM potassium phosphate buffer, pH 7. The activity was determined by measuring the increase of absorbance at 420 nm ($\varepsilon_{420} = 3.6 \times 10^4$ M^{-1} cm^{-1}) and normalized for [Heme]/[Protein]. Kinetic data were fitted with the Michaelis-Menten equation using Origin software.

Mutant	[Heme]/ [Protein]	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})×10^3
WT	0.5	3.4 ± 0.1	1.3 ± 0.1	2.6 ± 0.4
N6His-WT	0.5	3.4 ± 0.1	1.3 ± 0.1	2.6 ± 0.4
M01	0.6	3.4 ± 0.1	0.8 ± 0.1	4.2 ± 0.2
M02	0.7	4.9 ± 0.2	1.3 ± 0.1	3.7 ± 0.3
M03	0.6	7.0 ± 0.3	0.8 ± 0.1	8.7 ± 0.5
M04	0.4	7.0 ± 0.2	0.9 ± 0.1	7.8 ± 0.5
M05	0.5	4.1 ± 0.3	0.9 ± 0.2	4.6 ± 0.5
M06	0.7	1.7 ± 0.1	0.9 ± 0.1	1.8 ± 0.3
M07	0.5	2.8 ± 0.2	1.2 ± 0.3	2.3 ± 0.5
M08	0.5	2.3 ± 0.3	0.8 ± 0.2	2.8 ± 0.5
M09	0.4	4.2 ± 0.3	1.3 ± 0.2	3.2 ± 0.5
M10	0.5	3.2 ± 0.1	1.4 ± 0.2	2.2 ± 0.3
M11	0.5	4.0 ± 0.2	1.6 ± 0.2	2.5 ± 0.5
M12	0.6	4.8 ± 0.5	1.5 ± 0.2	3.1 ± 0.6
M13	0.5	3.3 ± 0.2	1.5 ± 0.2	2.2 ± 0.5
M14	0.8	5.0 ± 0.2	0.6 ± 0.1	8.3 ± 0.5
M15	0.5	4.7 ± 0.4	0.4 ± 0.2	11.7 ± 0.5
M16	0.8	6.1 ± 0.5	0.5 ± 0.1	12.2 ± 0.3
M17	0.7	3.6 ± 0.3	0.5 ± 0.2	7.2 ± 0.5
M18	0.7	7.5 ± 0.3	0.9 ± 0.2	8.3 ± 0.5
M19	0.5	7.7 ± 0.5	1.2 ± 0.2	6.4 ± 0.5
M20	0.9	5.7 ± 0.3	0.6 ± 0.2	9.5 ± 0.4

5.4.3. Immobilization on Epoxy Resin by Lys-tag

Immobilization variants M1-M10 on epoxy resin by variants M1-M10 was first performed, incubating for 24 h at 25 °C (**Table 5.3**). Immobilization yields, measuring the enzyme population bound to the carrier, resulted in >99% for all the variants, including the wild-type enzymes. However, the activity of the enzymes was not detectable, implying this protocol was utterly inefficient, at least after the long-time incubation at 25 °C. We considered the possibility of thermal inactivation, given by the long-time incubation at 25 °C. Therefore, we reduced the incubation time from 24 h to 2 h. This decreased the enzyme loading to 60–70%, but the recovered activity resulted in still residual, corresponding to less than 1% of the measured in the free enzymes. Similar results were obtained after incubation at a lower temperature (4 °C) for 16 h, which resulted in 80–100% immobilization yields but still below the 1% activity yield. These results indicate that, although immobilization occurred at high yields in these resins, this was utterly inefficient, leading to inactive biocatalysts and most likely occurring through nonspecific interactions, not involving the target lysine residues, as the results are similar in all variants. Covalent bond formation between the epoxy resins and any nucleophilic residue on the surface of the protein, like tyrosine, serines, cysteines, and lysines, was previously reported in the literature (Basso *et al*. 2022; Mateo *et al.* 2007). Therefore, the design of optimal immobilization sites on epoxy resins needs to be further investigated

Table 5.3. Immobilization parameters of *Ss*TyrH wild type and variants M1-10 on epoxy resin. Immobilization was performed using 10 mg protein/g support for 2, 16 and 24 h at 4 and 25 °C in 50 mM sodium phosphate pH 8. IY: Immobilization yield; AY: activity yield; IE: immobilization efficiency. N.D.: not detected.

Mutant	Initial act. (µmol/min) $\cdot 10^3$	**24 h at 25 °C**				**2 h at 25 °C**				**16 h at 4 °C**			
		IY (%)	Imm. Act. (µmol/min) $\cdot 10^3$	AY (%)	IE (%)	IY (%)	Imm. Act. (µmol/min) $\cdot 10^3$	AY (%)	IE (%)	IY (%)	Imm. Act. (µmol/min) $\cdot 10^3$	AY (%)	IE (%)
WT	13	100	N.D.	-	-	67	0.018	0.2	0.3	84	0.058	0.4	0.4
M01	34	100	N.D.	-	-	50	0.103	0.3	0.5	82	0.197	0.5	0.6
M02	44	100	N.D.	-	-	59	0049	0.1	0.2	88	0.156	0.3	0.4
M03	26	100	N.D.	-	-	72	0.018	0.2	0.3	96	0.176	0.4	0.4
M04	28	100	N.D.	-	-	70	0.015	0.1	0.1	96	0.097	0.2	0.2
M05	36	100	N.D.	-	-	53	0.069	0.2	0.3	80	0.133	0.4	0.4
M06	29	100	N.D.	-	-	69	0.005	0.1	0.2	96	0.070	0.2	0.2
M07	35	100	N.D.	-	-	54	0.043	0.1	0.2	78	0.131	0.4	0.5
M08	46	100	N.D.	-	-	53	0.033	0.1	0.1	88	0.118	0.2	0.2
M09	36	100	N.D.	-	-	74	0.038	0.2	0.3	99	0.180	0.4	0.4
M10	19	100	N.D.	-	-	60	0.015	0.1	0.2	95	0.061	0.3	0.3

5.3.4. Immobilization by His-tag on Metal-loaded Resin by Affinity Interaction

Wild-type enzymes with and without His-tags and variants M11-M20 were immobilized on a nickel-loaded carrier (**Table 5.4)**. The immobilization yields are around 80% for all variants. Importantly, the wild-type enzyme without a His-tag showed a relatively high immobilization yield of around 62%, indicating non-specific interactions. Variants where 6 His residues were introduced, e.g., between residues 38-39 (M14) and 106-107 (M15) or substituting His residues in segments 70-73 (M18) and 105-108 (M19), results in enzymes that show significantly higher initial activities than wild-type, higher immobilized activity and yields indicating improved catalytic efficiency and orientation towards the carrier. Others resulted in lower or comparable activities and yields of immobilization (below 30%). Our data suggests that four His are sufficient to ensure specific binding at certain locations: M14 (loop 38-39), M15 (loop 106–107), and M19 (replacements between 105-108) achieved the highest activity before immobilization and retained most of it afterward (**Fig. 5.4**), with their immobilized activity being up to 5-fold higher compared to N6His-WT.

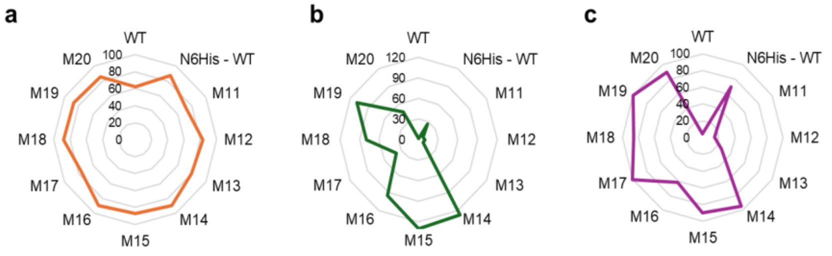


Fig. 5.4. Immobilization yield (%) **(a)**, immobilization activity (mmol·min^{-1}·10^3) **(b)**, and immobilization efficiency (%) **(c).** Chromalite™ MIDA/M resins loaded with Ni^{2+}. Activity assays were performed in 4 mM ABTS, 1 mM H_2O_2, 50 mM potassium phosphate buffer, and pH 7. Activity values were normalized for [Heme]/[Protein] (see **Table 5.2**).

Table 5.4. Immobilization parameters of *Ss*TyrH wild type and variants M11-20 on Chromalite™ MIDA/M resins loaded with Ni^{2+}. Immobilization was performed using 10 mg protein/g support for 1 h at 25 °C in 50 mM sodium phosphate pH 8. The activity values were normalized for [Heme]/[Protein]. IY: Immobilization yield; AY: activity yield; IE: immobilization efficiency.

Enzyme	Mutations	Initial activity (mmol/min) $\cdot 10^3$	Immobilization yield (%)	Immob activity (mmol/min) $\cdot 10^3$	Activity yield (%)	Immobilization efficiency (%)
WT	-	55 ± 7	62 ± 2	1.7 ± 0.7	3.0 ± 0.9	4.1 ± 0.4
N6His - WT	6 His-tag at N-terminal	50 ± 6	87 ± 1	27 ± 3	56 ± 11	70 ± 3
M11	D42**H**_S43**H**_G44**H**_S45**H**	45 ± 5	73 ± 7	7.7 ± 0.6	15 ± 2	21 ± 1
M12	42-**HHHH**-43	80 ± 7	83 ± 6	9.6 ± 0.8	12 ± 2	15 ± 2
M13	42-**HHHHHH**-43	39 ± 2	80 ± 9	8.2 ± 0.5	21 ± 1	27 ± 3
M14	38-**HHHHHH**-39	153 ± 4	90 ± 1	127 ± 2	83 ± 3	95 ± 1
M15	106-**HHHHHH**-107	165 ± 11	87 ± 3	130 ± 11	79 ± 2	90 ± 1
M16	240-**HHHHHH**-241	180 ± 20	90 ± 6	95 ± 1	53 ± 5	62 ± 3
M17	G36**H**_S37**H**_L38**H**_E39**H**	52 ± 12	76 ± 2	39 ± 8	76 ± 2	101 ± 1
M18	V70**H**_Q71**H**_P72**H**_A73**H**	109 ± 12	88 ± 1	79 ± 6	73 ± 4	86 ± 1
M19	L105**H**_P106**H**_E107**H**_E108**H**	122 ± 12	87 ± 2	108 ± 12	88 ± 2	100 ± 1
M20	S239**H**_R240**H**_P241**H**_S242**H**	60 ± 1	85 ± 1	47 ± 5	79 ± 7	90 ± 1

5.5 CONCLUDING REMARKS

To overcome the initial cost associated with the carrier, enzyme immobilization technology must ensure the highest efficiency and activity while minimizing catalyst loss to make this strategy profitable for industrial applications. Only some examples in the literature can be considered successful, with activity yields exceeding 80% (De Santis *et al.* 2022; De Simone *et al.* 2024; Ni *et al.* 2022; Winand *et al.* 2023). Indeed, most traditional immobilization methods, such as adsorption and cross-linking aggregates, rely on non-specific interactions between the enzyme surface and the carrier, leading to inefficient binding and low activity (Sheldon and van Pelt 2013). The use of tags to immobilize enzymes was initially mainly applied for purification purposes (Porath *et al.* 1975) and only recently emerged to achieve well-oriented immobilized enzymes for catalysis. However, this approach is typically limited to the N- and C-termini, restricting the range of possible orientations. While this strategy has yielded promising results in some cases, it is only sometimes optimal for catalysis and highly depends on the target protein and carrier. Therefore, it is important to develop more refined tools for this technology. To achieve this aim, we first tested a software tool designed to predict ideal sites and interactions for immobilizing the target protein. Although the software's poor performance in predicting ideal Lys-tags indicates that a more sophisticated algorithm is needed, it successfully identified six variants (M14, 15, 17, 18, 19, and 20) with His-tags placed in four different loops, which showed higher immobilization efficiency compared to the His-tag at the N-terminus, which was used as a benchmark for this study. Future work will focus on further developing this tool and validating it on additional systems.

5.6 REFERENCES

Andreescu, S., B. Bucur, and J.-L. Marty. 2006. 'Affinity immobilization of tagged enzymes.' In José M. Guisan (ed.), *Immobilization of Enzymes and Cells* (Humana Press: Totowa, NJ).

Barbosa, O., C. Ortiz, A. Berenguer-Murcia, R. Torres, R. C. Rodrigues, and R. Fernández-Lafuente. 2015. 'Strategies for the one-step immobilization-purification of enzymes as industrial biocatalysts', *Biotechnol. Adv.* 33: 435-56.

Basso, A., M. S. Brown, A. Cruz-Izquierdo, C. A. Martínez, and S. Serban. 2022. 'Optimization of metal affinity ketoreductase immobilization for application in batch and flow processes', *Org. Process Res. Dev.*

Bhattacharjee, N., L. Alonso-Cotchico, and M. F. Lucas. 2023. 'Enzyme immobilization studied through molecular dynamic simulations', *Front. Bioeng. Biotechnol.* 11.

Bolivar, J. M., I. Eisl, and B. Nidetzky. 2016. 'Advanced characterization of immobilized enzymes as heterogeneous biocatalysts', *Catal. Today* 259: 66-80.

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

Datta, S., L. R. Christena, and Y. R. S. Rajaram. 2013. 'Enzyme immobilization: an overview on techniques and support materials', *3 Biotech* 3: 1-9.

De Santis, P., N. Petrovai, L. E. Meyer, M. Hobisch, and S. Kara. 2022. 'A holistic carrier-bound immobilization approach for unspecific peroxygenase', *Front. Chem.* 10.

De Simone, M., L. Alonso-Cotchico, M. F. Lucas, V. Brissos, and L. O. Martins. 2024. 'Distal mutations enhance efficiency of free and immobilized NOV1 dioxygenase for vanillin synthesis', *J. Biotechnol.* 391: 92-98.

Deniau, C., R. Gilli, N. Izadi-Pruneyre, S. Létoffé, M. Delepierre, C. Wandersman, C. Briand, and A. Lecroisey. 2003. 'Thermodynamics of heme binding to the HasASM hemophore: effect of mutations at three key residues for heme uptake', *Biochemistry* 42: 10627-33.

Hernandez, K., and R. Fernández-Lafuente. 2011. 'Control of protein immobilization: coupling immobilization and site-directed mutagenesis to improve biocatalyst or biosensor performance', *Enzyme Microb. Technol.* 48: 107-22.

Maghraby, Y. R., R. M. El-Shabasy, A. H. Ibrahim, and H. M. E. Azzazy. 2023. 'Enzyme immobilization technologies and industrial applications', *ACS Omega*.

Mateo, C., J. M. Palomo, G. Fernández-Lorente, J. M. Guisan, and R. Fernández-Lafuente. 2007. 'Improvement of enzyme activity, stability and selectivity via immobilization techniques', *Enzyme Microb. Technol.* 40: 1451-63.

Ni, W. J., P. Y. Zhang, L. K. Long, and S. J. Ding. 2022. 'Engineering and linker-mediated co-immobilization of carotenoid cleavage oxygenase with phenolic acid decarboxylase for efficiently converting ferulic acid into vanillin', *Process Biochem.* 122: 67-77.

Paraskevopoulou, V., and F. H. Falcone. 2018. 'Polyionic tags as enhancers of protein solubility in recombinant protein expression', *Microorganisms* 6.

Porath, J., J. Carlsson, I. Olsson, and G. Belfrage. 1975. 'Metal chelate affinity chromatography, a new approach to protein fractionation', *Nature* 258: 598-9.

Sheldon, R. A., and S. van Pelt. 2013. 'Enzyme immobilisation in biocatalysis: why, what and how', *Chem. Soc. Rev.* 42: 6223-35.

Shen, Q. Y., R. J. Yang, X. Hua, F. Y. Ye, W. B. Zhang, and W. Zhao. 2011. 'Gelatin-templated biomimetic calcification for β-galactosidase immobilization', *Process Biochem.* 46: 1565-71.

Singh, W., S. F. G. Santos, S. Yadav, G. W. Black, and K. D. Dubey. 2023. 'Substrate conformation regulates aromatic C-H vs C-F bond activation in heme-dependent tyrosine hydroxylase', *Biochemistry* 62: 1577-87.

Wang, Y. F., I. Davis, I. Shin, H. Xu, and A. M. Liu. 2021. 'Molecular rationale for partitioning between C-H and C-F bond activation in heme-dependent tyrosine hydroxylase', *J. Am. Chem. Soc.* 143: 4680-93.

Winand, L., S. Theisen, S. Lütz, K. Rosenthal, and M. Nett. 2023. 'Immobilization of the amidohydrolase MxcM and its application for biocatalytic flow synthesis of pseudochelin A', *Catalysts* 13.

Wong, L. S., J. Thirlway, and J. Micklefield. 2008. 'Direct site-selective covalent protein immobilization catalyzed by a phosphopantetheinyl transferase', *J. Am. Chem. Soc.* 130: 12456-64.

Young, C. L., Z. T. Britton, and A. S. Robinson. 2012. 'Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications', *Biotechnol. J.* 7: 620-34.

Zeballos, N., N. Comino, D. Andrés-Sanz, J. Santiago-Arcos, M. Azkargorta, F. Elortza, E. Diamanti, and F. López-Gallego. 2023. 'Region-directed enzyme immobilization through engineering protein surface with histidine clusters', *ACS Appl. Mater. Interfaces* 16: 833-46.

Zhang, Y. F., J. Ge, and Z. Liu. 2015. 'Enhanced activity of immobilized or chemically modified enzymes', *ACS Catal.* 5: 4503-13.

Zymvol-Biomodeling. 'ZymLock: A ZYMVOL computational tool for enzyme immobilization', https://zymvol.com/enzyme-technology/.Zymvol-Biomodeling.

Chapter 6

Conclusions and Future Perspectives

6.1 Conclusions and Future Perspectives

Lignin is the most abundant source of aromatic compounds on Earth but remains a largely unexploited resource for the sustainable production of valuable chemicals, such as vanillin. Enzymatic catalysis has long been considered a promising tool for lignin valorization, yet implementing enzymes in industrial-scale reactions presents significant challenges. Enzymes have evolved to function under specific environmental conditions, which often do not align with industrial requirements, such as long operational lifetimes, high turnover rates, and resistance to product inhibition. Therefore, protein engineering and optimization of reaction conditions are necessary to allow their industrial application.

Research in this thesis focused on non-heme iron-dependent isoeugenol oxygenases from the carotenoid cleavage oxygenase (CCO) superfamily, which are known for catalyzing the cleavage of C=C bonds without requiring external cofactors. In this context, the primary objectives of this thesis were: (*i*) to enhance the stability and catalytic efficiency of isoeugenol oxygenases for vanillin production from lignin-derived compounds; (*ii*) to advance our understanding of the structure-function relationships within this enzyme class; and (*iii*) to establish an eco-friendly and cost-effective model for bio-vanillin production at an industrial scale.

NOV1 dioxygenase from *Novosphingobium aromatocivorans* was identified among the isoeugenol monooxygenases tested, namely *Pp*IEM from *P. putida* and *Pn*IEM from *P. nitroreducens*, as the most promising target for protein engineering due to its high solubility and crystal structure availability. Through this thesis, we observed that mutations close to the active site can influence catalysis and kinetic stability. Furthermore, we have also concluded that we can enhance the enzyme activity by distal mutations, which indirectly interact with the active site. A significant finding was the identification of residue S283, which, when mutated to a bulkier phenylalanine, was shown to be key in determining substrate specificity toward stilbenoids or phenylpropanoids, as well as influencing the enzyme's affinity for O_2. In addition, this work highlighted the iron cofactor's crucial role in maintaining the wild-type enzyme's kinetic stability. Current computational tools like FRESCO, Fireport, and PROSS focus on enhancing melting temperatures, but often overlook cofactor retention and its influence on enzyme stability. However, for metalloenzymes, it is often how they can retain the cofactor that critically determines their half-life and, by extension, their practical operational stability and applicability. Understanding the determinants of cofactor stabilization, like in the NOV1 case, can guide the development of tools to predict beneficial mutations.

Isoeugenol-to-vanillin conversion by NOV1 was achieved using whole-cell and immobilized enzyme systems, yielding vanillin over 90%. Importantly, we demonstrated for the first time an enzymatic conversion of 4-*n*-propyl guaiacol (4PG) to vanillin by two reaction steps. This cascade produced high vanillin yields starting from a real lignin oil obtained by reductive catalytic fractionation of softwood, thereby offering a realistic approach to wood biorefinery.

A key aspect of this thesis is using computational methods to guide protein engineering, enhancing both enzyme activity and stability. Furthermore, we introduced a computational His-tag design to optimize enzyme immobilization. While the successful cases of protein engineering shown through this thesis demonstrate the potential of computational tools, it is important to acknowledge that these methods still need to be refined. Their success heavily depends on the target system and the availability of reliable data. In addition, although computational approaches are often presented as an alternative to labor-intensive directed evolution, the resulting library can range from a few to dozens of mutants, whose construction by site-directed mutagenesis can be time-consuming with no guaranteed success. Therefore, more advances are necessary to achieve satisfying reliability with these tools.

Overall, the initial objectives of this thesis (see above) can be considered fulfilled; however, more aspects of NOV1 and related enzymes need to be elucidated. CCOs possess relevant biological importance, constituting a family of ubiquitous enzymes involved in many processes. While this work primarily focused on the industrial applications of NOV1 dioxygenase, it is also intriguing to consider its biological functions. A recent study by Metz *et al.* (2024) identified NOV1 (referred as LsdD) as key in the degradation of lignin-derived dimers, pointing to a significant role in lignin catabolism. Future research into *N. aromatocivorans* metabolism and NOV1's substrate specificity could shed light on its broader biological role in its native host. In expanding NOV1's substrate scope, we observed peroxidase-like activity on phenolics, including DMP and guaiacol, within the range observed for certain peroxidases, albeit the kinetic parameters, remains the cleavage of C=C bonds in stilbenoid and phenylalkyloid compounds. Dual oxygenase/peroxidase activity is not unprecedented among iron-dependent enzymes (Hrycay and Bandiera 2015; Wang *et al*. 2013). It is unclear whether in the NOV1 case, peroxidase activity has biological relevance. However, the alkaline optimal pH and lack of H_2O_2 inhibition at high concentrations are valuable characteristics for industrial application.

Furthermore, protein engineering can offer exciting possibilities for expanding the substrate range of CCOs, revealing new reaction pathways and leading to commercially valuable chemicals beyond vanillin. For instance, Fujimaki *et al.* (2024) achieved the one-step conversion of ferulic acid to vanillin using an engineered CCO enzyme. The same strategy could further expand the substrate range of NOV1 as well.

In conclusion, despite significant progress, the catalytic mechanisms of many CCOs remain partially understood. Whether they have a monooxygenase or dioxygenase nature or if the two mechanisms coexist is still debatable. Future studies can focus on all these aspects, providing deeper insights into the catalytic mechanisms and substrate specificity.

6.6 REFERENCES

Fujimaki, S., S. Sakamoto, S. Shimada, K. Kino, and T. Furuya. 2024. 'Engineering a coenzyme-independent dioxygenase for one-step production of vanillin from ferulic acid', *Appl Environ Microbiol*, 90: e0023324.

Hrycay, E. G., and S. M. Bandiera. 2015. 'Monooxygenase, peroxidase and peroxygenase properties and reaction mechanisms of cytochrome P450 enzymes', *Adv Exp Med Biol*, 851: 1-61.

Metz, F., A. M. Olsen, F. C. Lu, K. S. Myers, M. N. Allemann, J. K. Michener, D. R. Noguera, and T. J. Donohue. 2024. 'Catabolism of β-5 linked aromatics by', *mBio*, 15.

Wang, C., W. C. Chang, Y. Guo, H. Huang, S. C. Peck, M. E. Pandelia, G. M. Lin, H. W. Liu, C. Krebs, and J. M. Bollinger, Jr. 2013. 'Evidence that the fosfomycin-producing epoxidase, HppE, is a non-heme-iron peroxidase', *Science*, 342: 991-5.

Annexes

Annex I

Table S3.1 - Primers used to construct the thirty-five variants in the active site in Chapter 3.

Primers	**Sequence**
N120L Fwd	ACTGCACTCACAAATGCGTTTGTTTTCGGAGGTAAG
N120L Rev	ATTTGTGAGTGCAGTCGAACGGATCTCTCCTTTAACAG
N120S Fwd	GATCCGTTCGACTGCAAGCACAAATGCGTTTG
N120S Rev	CAAACGCATTTGTGCTTGCAGTCGAACGGATC
T121P Fwd	GCAAACCCAAATGCGTTTGTTTTCGGAGGTAAGTTGTG
T121P Rev	CGCATTTGGGTTTGCAGTCGAACGGATCTCTC
T121A Fwd	CCGTTCGACTGCAAACGCGAATGCGTTTGTTTTC
T121A Rev	GAAAACAAACGCATTCGCGTTTGCAGTCGAACGG
F281W Fwd	AATTGTTGGGCGTCTCACGTTTTGAACGCATG
F281W Rev	AGACGCCCAACAATTGTCGCGGGTAAAC
S283F Fwd	GCGTTTCACGTTTTGAACGCATGGCAGGAG
S283F Rev	CAAAACGTGAAACGCAAAACAATTGTCGCGGGTAAAC
S283Q Fwd	GCGCAACACGTTTTGAACGCATGGCAGGAGGGAACAAAGATC
S283Q Rev	CAAAACGTGTTGCGCAAAACAATTGTCGCGGGTAAACCAACGAATAT
S283I Fwd	GACAATTGTTTTGCGATTCACGTTTTGAACGC
S283I Rev	GCGTTCAAAACGTGAATCGCAAAACAATTGTC
S283E Fwd	GACAATTGTTTTGCGGTGCACGTTTTGAACGC
S283E Rev	GCGTTCAAAACGTGCACCGCAAAACAATTGTC
S283N Fwd	GACAATTGTTTTGCGAACCACGTTTTGAACGC
S283N Rev	GCGTTCAAAACGTGGTTCGCAAAACAA TTGTC
S283V Fwd	GACAATTGTTTTGCGGTTCACGTTTTGAACGC
S283V Rev	GCGTTCAAAACGTGAACCGCAAAACAATTGTC
S283C Fwd	GACAATTGTTTTGCGTGCCACGTTTTGAACGC
S283C Rev	GCGTTCAAAACGTGGCACGCAAAACAATTGTC
S283T Fwd	GACAATTGTTTTGCGACCCACGTTTTGAACGC
S283T Rev	GCGTTCAAAACGTGGGTCGCAAAACAATTGTC
F307H Fwd	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
F307H Rev	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC
F307W Fwd	AATATGTGGCCTTTTTTTCCCGACGTTCATGGCGCACCATTCAACG
F307W Rev	AAAAGGCCACATATTGTTTTTGGCCTCACAAGTAACAAAGTGGA
F309W Fwd	CAATATGTTCCCTTGGTTTCCCGACGTTCATG
F309W Rev	CATGAACGTCGGGAAACCAAGGGAACATATTG
F473W Fwd	CAATCCGTCTGCGCTGGGGACTTCACGGGAAC
F473W Rev	GTTCCCGTGAAGTCCCCAGCGCAGACGGATTG
F473E Fwd	CAATCCGTCTGCGCGAAGGACTTCACGGGAAC

F473E Rev	GTTCCCGTGAAGTCCTTCGCGCAGACGGATTG
F473Q Fwd	CAATCCGTCTGCGCCAGGGACTTCACGGGAAC
F473Q Rev	GTTCCCGTGAAGTCCCTGGCGCAGACGGATTG
L475S Fwd	CGCTTCGGATCTCACGGGAACTGGGCAAATGCCGATGAAATTG
L475S Rev	GTGAGATCCGAAGCGCAGACGGATTGGGATATTAACCGT
L475G Fwd	CAATCCGTCTGCGCTTCGGAGGCCACGGGAACTGGGCAAATG
L475G Rev	CATTTGCCCAGTTCCCGTGGCCTCCGAAGCGCAGACGGATTG
L475T Fwd	CAATCCGTCTGCGCTTCGGAACCCACGGGAACTGGGCAAATG
L475T Rev	CATTTGCCCAGTTCCCGTGGGTTCCGAAGCGCAGACGGATTG
E353D Fwd	GTGACACCGCTGCGGATTTTCCTCGCATCGAC
E353D Rev	GTCGATGCGAGGAAAATCCGCAGCGGTGTCAC
F281C-S283I Fwd	GGTTTACCCGCGACAATTGTTGCGCGATTCACGTTTTGAACGCATGGC
F281C-S283I Rev	GCCATGCGTTACAAACGTGAATCGCGCAACAATTGTCGCGGGTAAACC
F281C-S283T Fwd	GGTTTACCCGCGACAATTGTTGCGCGACCCACGTTTTGAACGCATGGC
F281C-S283T Rev	GCCATGCGTTCAAAACGTGGGTCGCGCAACAATTGTCGCGGGTAAACC
F281I-S283V Fwd	GGTTTACCCGCGACAATTGTATTGCGGTGCACGTTTTGAACGCATGGC
F281I-S283V Rev	GCCATGTCAAAACGTGCACCGCCATACAATTGTCGCGGGTAAACC
F281H-S283I Fwd	GGTTTACCCGCGACAATTGTCATGCGATTCACGTTTTGAACGCATGGC
F281H-S283I Rev	GCCATGCGTTCAAAACGTGAATCGCATGACAATTGTCGCGGGTAAACC
F281H-S283V Fwd	GGTTTACCCGCGACAATTGTCATGCGGTGCACGTTTTGAACGCATGGC
F281H-S283V Rev	GCCATGCGTTCAAAACGTGCACCGCATGACAATTGTCGCGGGTAAACC
F281M-S283I Fwd	GGTTTACCCGCGACAATTGTATGGCGATTCACGTTTTGAACGCATGGC
F281M-S283I Rev	GCCATGTCAAAACGTGAATCGCCATACAATTGTCGCGGGTAAACC
F281M-S283T Fwd	GGTTTACCCGCGACAATTGTATGGCGACCCACGTTTTGAACGCATGGC
F281M-S283T Rev	GCCATGCGTTCAAAACGTGGGTCGCCATACAATTGTCGCGGGTAAACC
F281M-S283V Fwd	GGTTTACCCGCGACAATTGTATGGCGGTGCACGTTTTGAACGCATGGC
F281M-S283V Rev	GCCATGCGTTCAAAACGTGCACCGCCATACAATTGTCGCGGGTAAACC
F473W-L475G Fwd	CCCAATCCGTCTGCGCTGGGGAGGCCACGGGAACTGGGCAAATGCC
F473W-L475G Rev	GGCATTTGCCCAGTTCCCGTGGCCTCCCCAGCGCAGACGGATTGGG
F473W-L475T Fwd	CCCAATCCGTCTGCGCTGGGGAACCCACGGGAACTGGGCAAATGCC
F473W-L475T Rev	GGCATTTGCCCAGTTCCCGTGGGTTCCCCAGCGCAGACGGATTGGG
F281M-S283I-F307H[1] Fw	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
F281M-S283I-F307H[1] Rv	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC
F281M-S283T-F307H[2] Fw	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
F281M S283T F307H[2] Rv	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC

---

[1] F307H mutation was introduced using the gene containing F281M_S283I mutations

[2] F307H mutation was introduced using the gene containing F281M_S283T mutations

Table S3.2. Data collection and refinement statistics for the crystal structure of S283F in Chapter 3.

Space group	C222
Unit cell axes (Å)	a = 178.471 b = 187.984 c = 105.867
Resolution (Å)	2.9
PDB code	7QR6
$R_{sym}^{a,b}$ (%)	9.2 (73.1)
$CC_{1/2}$ (%)	98.1 (30.3)
Completeness (%)	98.9 (97.5)
Unique reflections	36,348
Redundancy	3.5 (3.5)
I/σ^{b}	6.1 (1.6)
N° of non-hydrogen atoms ^{c}	
protein/Fe	11370/3
water	16
Average B value for protein atoms (Å^2)	71.0
$R_{cryst}^{b,d}$ (%)	21.8
$R_{free}^{b,d}$ (%)	29.0
Rms bond length (Å)	0.008
Rms bond angles (°)	1.48

^{a} $R_{sym}=\sum|I_i-\langle I\rangle|/\sum I_i$, where I_i is the intensity of i^{th} observation and $\langle I\rangle$ is the mean intensity of the reflection.

^{b} Values in parentheses are for reflections in the highest resolution shell.

^{c} The asymmetric unit contains three protein chains. The final model comprises residues 10-382 and 391-488 of subunit A, 5-382 and 391-488 of subunit B, 3-382, and 391-488 of subunit C.

^{d} $R_{cryst}=\sum|F_{obs}-F_{calc}|/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{cryst} and R_{free} were calculated using the working and test sets, respectively.

Table S3.3. Binding energy (ΔG) and dihedral energy (ΔE) values in kcal/mol found for isoeugenol and resveratrol at the binding site of wild-type and S283F NOV1 variant in the ensemble docking experiments used in Chapter 3, section 3.4.5.

	Isoeugenol					Resveratrol				
	$\Delta G_{binding}$		**$\Delta E_{dihedral}$**		**$\Delta\Delta E_{dihedral}$**	**$\Delta G_{binding}$**		**$\Delta E_{dihedral}$**		**$\Delta\Delta E_{dihedral}$**
variant	**WT**	**S283F**	**WT**	**S283F**	**(S283F-WT)**	**WT**	**S283F**	**WT**	**S283F**	**(S283F-WT)**
mean	-5.97	-5.96	9.24	10.25	1.01	-6.80	-6.98	14.55	18.87	4.32
std	0.62	0.50	3.69	4.16	0.48	1.34	0.38	5.21	5.54	0.33
min	-7.10	-6.84	0.92	2.27	1.35	-8.64	-7.52	3.92	8.77	4.85
25%	-6.37	-6.34	6.84	7.30	**0.46**	-7.64	-7.30	11.39	15.53	**4.14**
50%	-6.07	-5.93	9.43	9.97	**0.54**	-7.20	-7.03	14.10	19.69	**5.59**
75%	-5.64	-5.67	11.48	13.22	**1.74**	-6.30	-6.81	16.94	22.65	**5.71**
max	-2.77	-4.89	18.38	18.42	0.04	0.80	-6.24	31.35	28.45	-2.90

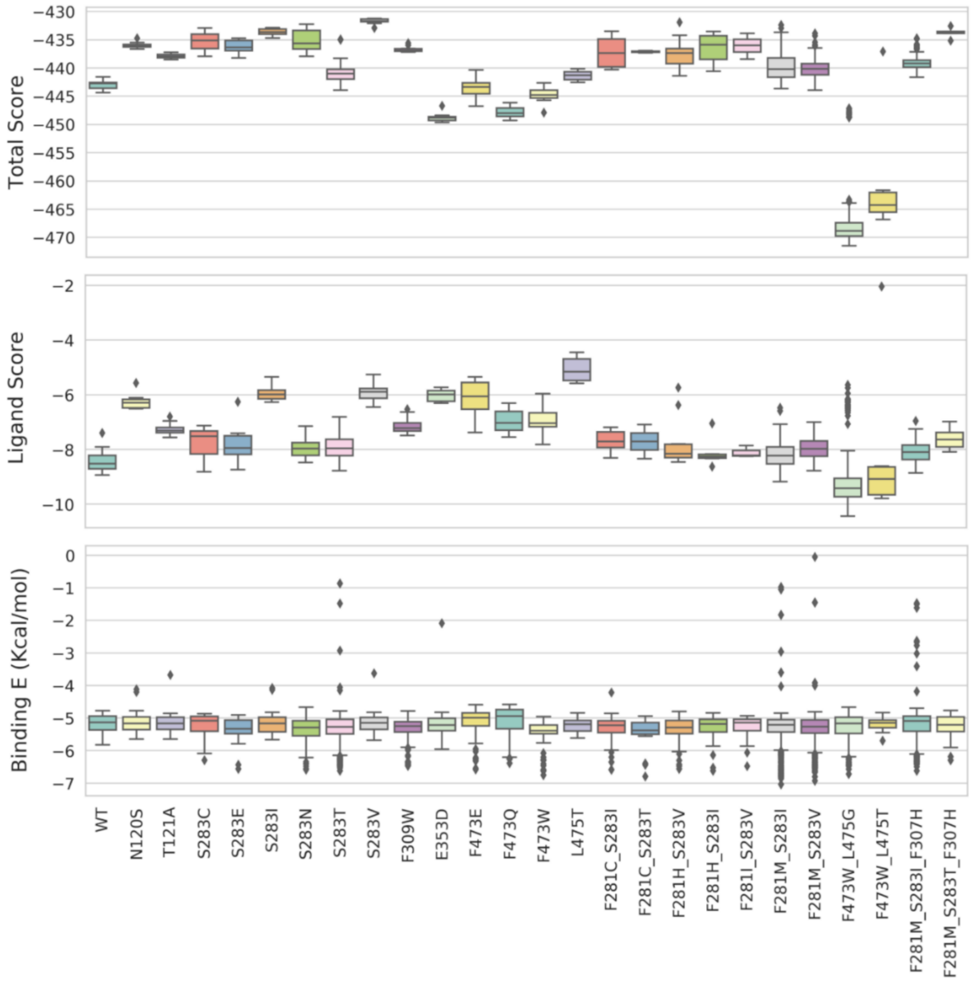


Fig. S3.1. Rosetta Ligand Score, Rosetta Total Score, and VINA Autodock binding energy values for the *in silico* variants selected for experimental work in chapter 3, section 3.4.1.

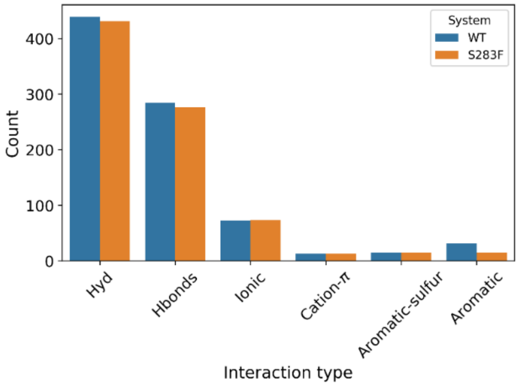


Figure S3.2. Inter-residue interaction analysis was performed with PIC (Protein Interaction Calculator) using the X-ray structures of wild-type NOV1 (PDB: 55j5) and S283F variant (7QR6) in Chapter 3, section 3.4. 5. Overall counting of interactions depicted by PIC in both wild-type NOV1 and variant. It includes sidechain-sidechain, backbone-backbone, and sidechain-backbone interactions.

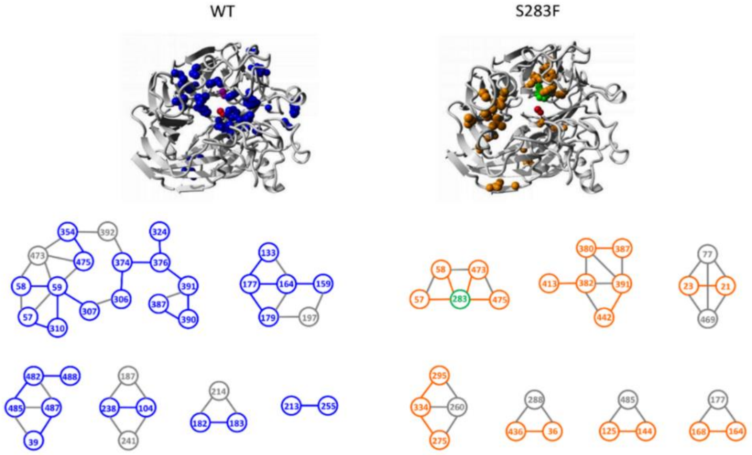


Figure S3.3. Inter-residue interaction analysis was performed with PIC (Protein Interaction Calculator) using the X-ray structures of wild-type NOV1 (PDB: 55j5) and S283F variant in Chapter 3, section 3.4.5. The sidechain-sidechain hydrophobic interactions that appear in wild-type but not in the variant (left) or the variant but not in the wild-type (right) are displayed in blue and orange balls. The S283F mutation is highlighted in green balls. The grey balls and sticks comprise interactions shared by both systems.

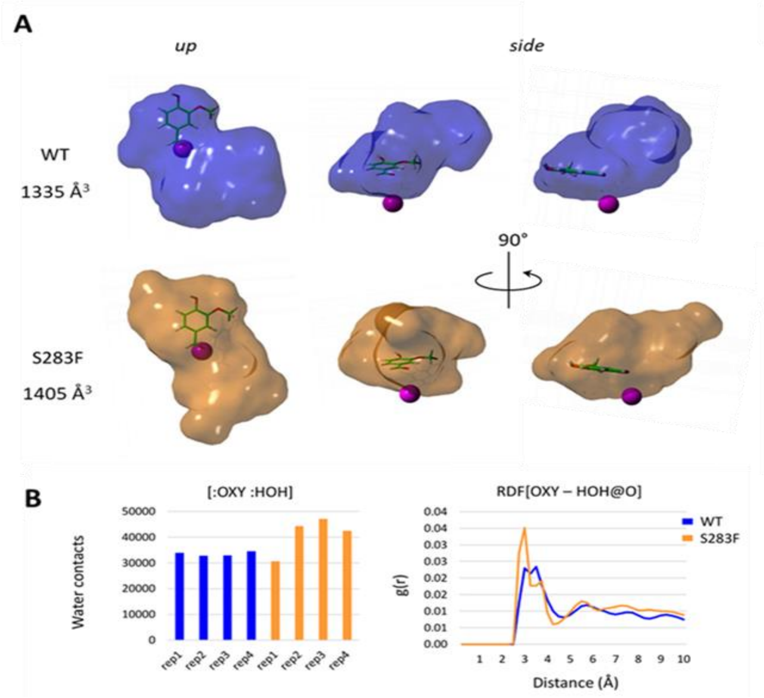


Figure S3.4. (**A**) Empty volumes of the active site of the wild-type NOV1 (blue) and S283F variant (orange) were calculated with MDpocket over 800 ns of MD simulation. The vanillin ligand found in the X-ray 5J55 is superposed. (**B**) Water counting inside the active site (left) and Radial Distribution Function (RDF) (right) of water molecules concerning the molecular oxygen during 800 ns of MD simulation. (Chapter 3, section 3.4.5)

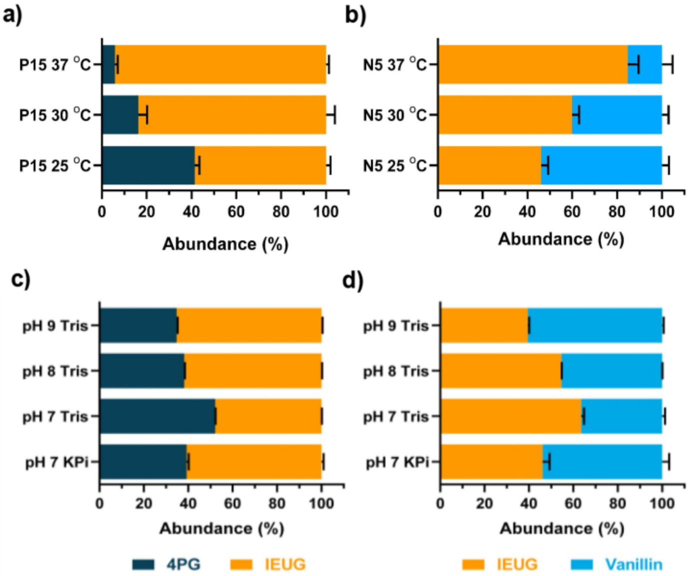


Figure S3.5. Conversions of 50 mM solutions of **a)** 4PG by P cells at OD_{600} 15; and **b)** conversions of isoeugenol by N cells at OD_{600} 5 in 50 mM KPi buffer, pH 7, at 25 °C, 30 °C, and 37 °C. Conversions of 50 mM solutions of **c)** 4PG by P cells at OD_{600} 15; and **d)** conversions of isoeugenol by N cells at OD_{600} 5 at 25 °C, in four buffer systems at 50 mM concentration: KPi pH 7, Tris-HCl, pH 7, 8 and 9. Reactions were performed in 10% (v/v) DMSO for 24 h. (Chapter 3, section 3.4.8).

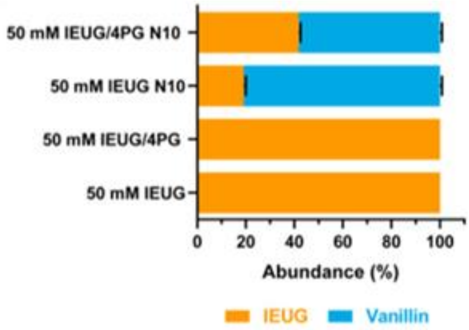


Figure S3.6 Effect of 4PG on the conversions of 50 mM isoeugenol by N cells: conversions of 50 mM isoeugenol in the presence and absence of 4PG were performed by N cells (OD_{600} 10) in 50 mM KPi buffer, pH 7, 10% (v/v) DMSO, at 25 °C for 24 h. (Chapter 3, section 3.4.9).

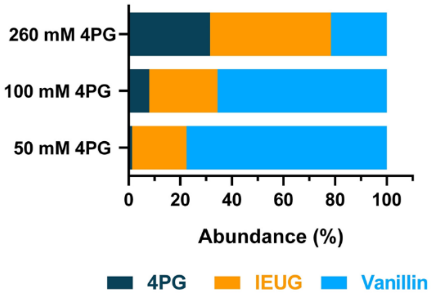


Figure S3.7. One-pot conversions of 50, 100, and 260 mM 4PG in 50 mM KPi buffer, pH 7, 10% (v/v) DMSO using a whole-cell, multi-enzyme cascade in a step-wise fashion. To catalyze the first step of the process, P cells (OD_{600} 50) were employed, and N cells (OD_{600} 10) were used in the second step. The reaction was shaken at 150 rpm at 37 °C for 24 h, then changed to 25 °C for 24 h. (Chapter 3, section 3.4.9).

Annex II

Table S4.1. Primers used to construct the forty-one variants and for DNA shuffling in chapter 4.

Primers	**Sequence**
pET21Fwd	CTAGCATAACCTAAC
pET21Rev	CCAAGGGGTTATGCTAGTTATTGCTCAG
nov1GApET28a Fw	GTGGTGGATCCATGGCTCAGTTTCCAAACAC
nov1GApET28a Rv	GGTGCTCGAGTTACGCTGCCAATACCTTCTC
pET28aGAnov1 Fw	GGCAGCGTAACTCGAGCACCACCACCACCA
pET28aGAnov1 Rv	CTGAGCCATGGATCCACCACCAATCTGTTCTCTGTG
T299S Fwd	GGCAGGAGGGAACAAAGATCCACTTTGTTTCGTGTGAGGCCAAAAAC
T299S Rev	GTTCCCTCCTGCCATGCGTTCAAAACGTGAGACGCAAAACAATTG
L372A Fwd	CACGTCATGGCTGGTTCGCGGAGATGGATATGAAAC
L372A Rev	GTTTCATATCCATCTCCGCGAACCAGCCATGACGTG
L286V Fwd	TCATGGCTGGTTCGTGGAGATGGATATG
L286V Rev	CATATCCATCTCCACGAACCAGCCATGA
M324K Fwd	CAACGGCATGGAGGCAAAATCGCACCCTACCGATTG
M324K Rev	CAATCGGTAGGGTGCGATTTTGCCTCCATGCCGTTG
L286W Fwd	TGCGTCTCACGTTTGGAACGCATGGCAG
L286W Rev	CTGCCATGCGTTCCAAACGTGAGACGCA
M374A Fwd	GCTGGTTCCTGGAGGCGGATATGAAACGTCC
M374A Rev	GGACGTTTCATATCCGCCTCCAGGAACCAGC
I222V Fwd	GATGCACGATTTTGGAGTGACAGAGGATTATCTTG
I222V Rev	CAAGATAATCCTCTGTCACTCCAAAATCGTGCATC
N478S Fwd	GGACTTCACGGGAGCTGGGCAAATGC
N478S Rev	GCATTTGCCCAGCTCCCGTGAAGTCC
L286I Fwd	GTTTTGCGTCTCACGTTATTAACGCATGGCAGGAGG
L286I Rev	CCTCCTGCCATGCGTTAATAACGTGAGACGCAAAAC
T299V Fwd	AAAGATCCACTTTGTTGTGTGTGAGGCCAAAAACA
T299V Rev	TGTTTTTGGCCTCACACACAACAAAGTGGATCTTT

I222L Fwd	GCATGATGCACGATTTTGGACTGACAGAGGATTATCTTGTTC
I222L Rev	GAACAAGATAATCCTCTGTCAGTCCAAAATCGTGCATCATGC
L372V Fwd	TCATGGCTGGTTCGTGGAGATGGATATG
L372V Rev	CATATCCATCTCCACGAACCAGCCATGA
F124V Fwd	CTGCAAACACAAATGCGGTGGTTTTCGGAGGTAAGTTG
F124V Rev	CAACTTACCTCCGAAAACCACCGCATTTGTGTTTGCAG
T299L Fwd	AAAGATCCACTTTGTTCTGTGTGAGGCCAAAAACA
T299L Rev	TGTTTTTGGCCTCACACAGAACAAAGTGGATCTTT
T299A Fwd	AAAGATCCACTTTGTTGCGTGTGAGGCCAAAAACA
T299A Rev	TGTTTTTGGCCTCACACGCAACAAAGTGGATCTTT
M177L Fwd	CCTAAAACCGGAAATCTGGTCGCAATCGGC
M177L Rev	GCCGATTGCGACCAGATTTCCGGTTTTAGG
W330F Fwd	CGCACCCTACCGATTTTGTGGTAGATATGGCG
W330F Rev	CGCCATATCTACCACAAAATCGGTAGGGTGCG
M195G Fwd	ATGATGTCACGTACGGCGAAGTTAGTCCTGAA
M195G Rev	TTCAGGACTAACTTCGCCGTACGTGACATCAT
T362V Fwd	GCTGCGGAATTTCCTGTGATCGACGACCGTTTT
T362V Rev	AAAACGGTCGTCGATCACAGGAAATTCCGCAGC
W330Y Fwd	CGCACCCTACCGATTATGTGGTAGATATGGCG
W330Y Rev	CGCCATATCTACCACATAATCGGTAGGGTGCG
F13Y Fwd	CTCCTTCATTCACAGGATATAATACACCTTCTCGCATTG
F13Y Rev	CAATGCGAGAAGGTGTATTATATCCTGTGAATGAAGGAG
K86G Fwd	CGTTTTTCTAATTTCCAGCCATCTGTCTTGGCCCAACG
K86G Rev	GCATTTTCTAATTTCCAGCCATCTGTCTTGGCCCAACG
K86R Fwd	CGTTGGGCCAAGACAGATCGTTGGAAATTAGAAAATGCCG
K86R Rev	CGGCATTTTCTAATTTCCAACGATCTGTCTTGGCCCAACG
R116K Fwd	GCATTTGTTAAAGGAGAGATCAAATCGACTGCAAACACAAATGC
R116K Rev	GCATTTGTGTTTGCAGTCGATTTGATCTCTCCTTTAACAGCTTC
N122A Fwd	GTTCGACTGCAAACACAGCGGCGTTTGTTTTCGGAGG
N122A Rev	CCTCCGAAAACAAACGCCGCTGTGTTTGCAGTCGAAC
N122S Fwd	CGACTGCAAACACAAGCGCGTTTGTTTTCGG
N122S Rev	CCGAAAACAAACGCGCTTGTGTTTGCAGTCG

F124I Fwd	CTGCAAACACAAATGCGATTGTTTTCGGAGGTAAG
F124I Rev	CTTACCTCCGAAAACAATCGCATTTGTGTTTGCAG
F124L Fwd	CAAACACAAATGCGCTGGTTTTCGGAGGTAAG
F124L Rev	CTTACCTCCGAAAACCAGCGCATTTGTGTTTG
F164M Fwd	GAAAATGACTGGCCAGACAATGACAGCTCATCCCAAGGTTG
F164M Rev	CAACCTTGGGATGAGCTGTCATTGTCTGGCCAGTCATTTTC
M195F Fwd	CACCGATGATGTCACGTACTTTGAAGTTAGTCCTGAAGGTG
M195F Rev	CACCTTCAGGACTAACTTCAAAGTACGTGACATCATCGGTG
M195Y Fwd	CACCGATGATGTCACGTACTATGAAGTTAGTCCTGAAGGTG
M195Y Rev	CACCTTCAGGACTAACTTCATAGTACGTGACATCATCGGTG
L286A Fwd	GTTTTGCGTCTCACGTTGCGAACGCATGGCAGGAG
L286A Rev	GTTTTGCGTCTCACGTTGCGAACGCATGGCAGGAG
T299G Fwd	GGAACAAAGATCCACTTTGTTGGCTGTGAGGCCAAAAACAATATG
T299G Rev	CATATTGTTTTTGGCCTCACAGCCAACAAAGTGGATCTTTGTTCC
E301R Fwd	GATCCACTTTGTTACTTGTCGTGCCAAAAACAATATGTTCCC
E301R Rev	GGGAACATATTGTTTTTGGCACGACAAGTAACAAAGTGGATC
W330I Fwd	GTCGCACCCTACCGATATTGTGGTAGATATGGCG
W330I Rev	CGCCATATCTACCACAATATCGGTAGGGTGCGAC
R356M Fwd	CCGCTGCGGAATTTCCTATGATCGACGACCGTTTTAC
R356M Rev	GTAAAACGGTCGTCGATCATAGGAAATTCCGCAGCGG
R356V Fwd	CGCTGCGGAATTTCCTGTGATCGACGACCGTTTTAC
R356V Rev	GTAAAACGGTCGTCGATCACAGGAAATTCCGCAGCG
T362L Fwd	GCATCGACGACCGTTTTCTGGGGCAAAAAACACGTC
T362L Rev	GACGTGTTTTTTGCCCCAGAAAACGGTCGTCGATGC
Q417G Fwd	CCCCGTTTCCTCATTGGGCGAGCCTTGCTTTGTG
Q417G Rev	CACAAAGCAAGGCTCGCCCAATGAGGAAACGGGG
Q417S Fwd	CCCCGTTTCCTCATTGAGCGAGCCTTGCTTTGTGC
Q417S Rev	GCACAAAGCAAGGCTCGCTCAATGAGGAAACGGGG
N478T Fwd	TCGGACTTCACGGGACCTGGGCAAATGCCGA
N478T Rev	TCGGCATTTGCCCAGGTCCCGTGAAGTCCGA

Table S4.2 – Distances (Å) of each mutation selected for DNA-shuffling (Chapter 4, section 4.4.3.) to the Fe^{2+} cofactor and the four iron-coordinating histidines.

Mutation	Region	Fe^{2+} (Å)	H167 (Å)	H217 (Å)	H284 (Å)	H476 (Å)
I222V	core	14.8	9.8	14	12.8	15.5
L286W	core	10.9	9	13.1	6.4	10.2
T299S	core	10.6	10.5	9.6	4.6	11.4
M324K	surface	16.3	17.7	16.1	12.6	16.5
L372A	core	10.9	11.8	12.5	9	8.8
N478S	core	10.1	8.1	11.6	11.2	5.3

Annex III

Table S5.1. Primers used to construct the His-tagged variants M14-20 by substitution or addition in chapter 5.

Primers	**Sequence**
M14 Fwd	CACCACCACGAAGCGGCGGATTCTG
M14 Rev	ATGATGATGCAGGCTACCCGGTTC
M15 Fwd	CACCACCACGAAGAAGGACCAGAACGTGATGC
M15 Rev	ATGATGATGCGGCAGGCGCGCAAG
M16 Fwd	CACCACCACGAAGCGGCGGATTCTG
M16 Rev	ATGATGATGCAGGCTACCCGGTTC
M17 Fwd	CATCACGCGGCGGATTCTGGATC
M17 Rev	ATGATGCGGTTCTGGCAGGGTTAAC
M18 Fwd	CATCACGATAGCAGCGATCGTCTGTACTGGTTTC
M18 Rev	ATGATGGCGTTCCGCGGGGCC
M19 Fwd	CATCACGGACCAGAACGTGATGCAGC
M19 Rev	ATGATGGCGCGCAAGCTCGCG
M20 Fwd	CATCACTCTCGTGATGTTGTTG
M20 Rev	ATGATGACGAAGGGTCAGAAAG