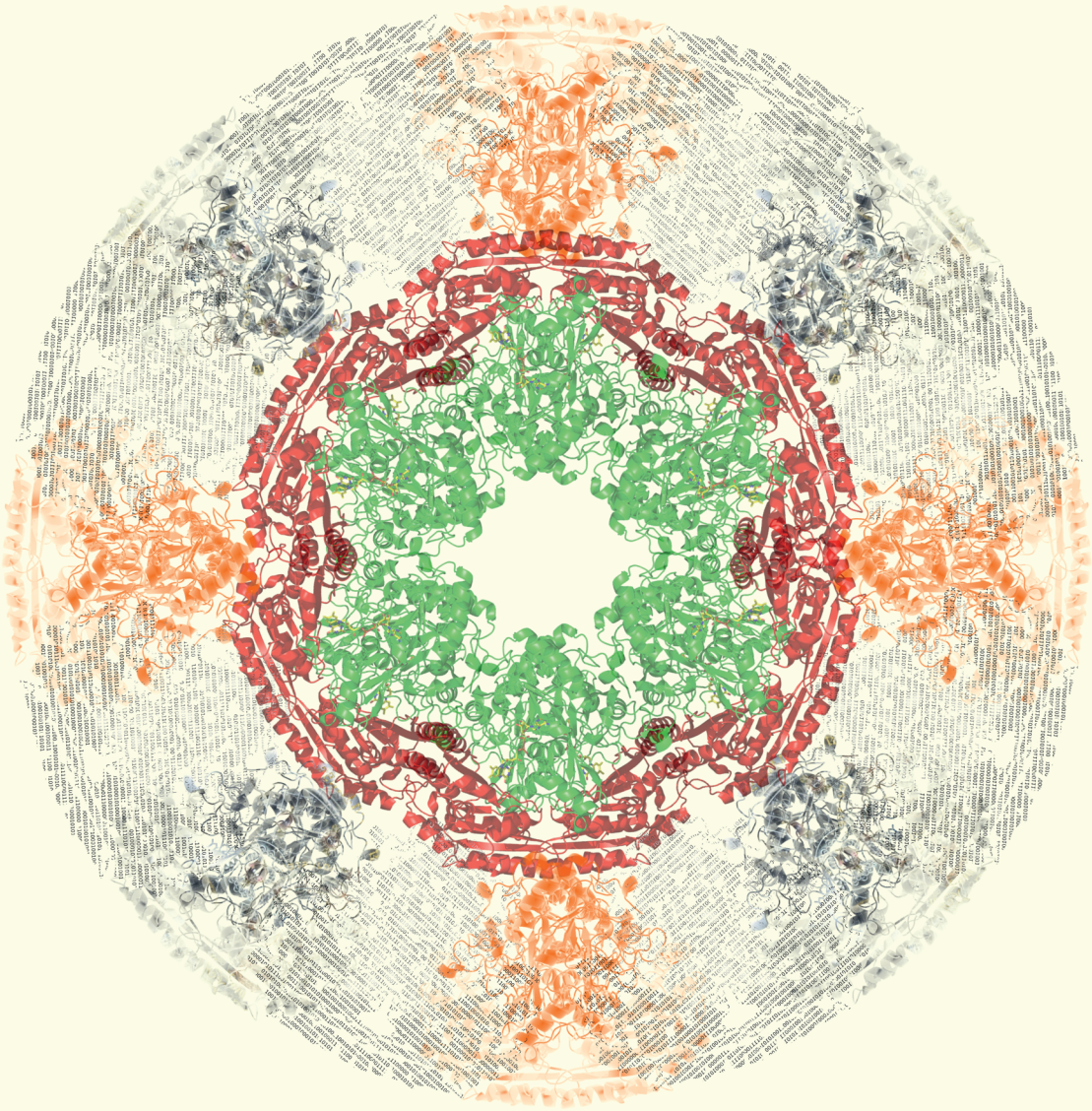
Propositions

1. Ligand migration is indispensable for enzyme function.
   (this thesis)
2. Flexibility of Asp170 leads to adduct formation in vanillyl alcohol oxidase.
   (this thesis)
3. „Potential industrial applications“ is a non-sensical, empty phrase.
4. Aimless searches are the solution to all scientific problems.
5. Swiss and Dutch mountains are the best illustration of enzyme function.
6. Enzymes are inexpedient because evolution is a passive, non-directed process.
7. Every scientist must know how to program.
8. Science is the biggest luxury product of humanity.

Propositions belonging to the thesis, entitled
„A Kaleidoscopic Look at Vanillyl Alcohol Oxidase“
Gudrun Gygli
Wageningen, 18 April 2018

A Kaleidoscopic Look at Vanillyl Alcohol Oxidase



Gudrun Gygli

A Kaleidoscopic Look at Vanillyl Alcohol Oxidase

Gudrun Gygli

Thesis committee

Promotor
Prof. Dr W. J. H. van Berkel
Personal Chair at the Laboratory of Biochemistry
Wageningen University & Research

Other members
Prof. Dr A. J. M. Stams, Wageningen University & Research
Prof. Dr C. Faulds, Aix-Marseille University, France
Prof. Dr M. W. Fraaije, University of Groningen
Prof. Dr F. Seebeck, Basel University, Switzerland

This research was conducted under the auspices of the Graduate School VLAG
(Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

A Kaleidoscopic Look at Vanillyl Alcohol Oxidase

Gudrun Gygli

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 18 April 2018
at 4 p.m. in the Aula.

Gudrun Gygli
A kaleidoscopic look at vanillyl alcohol oxidase,
230 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in English, Dutch, and German

ISBN 978-94-6343-741-7
DOI https://doi.org/10.18174/431725

Contents

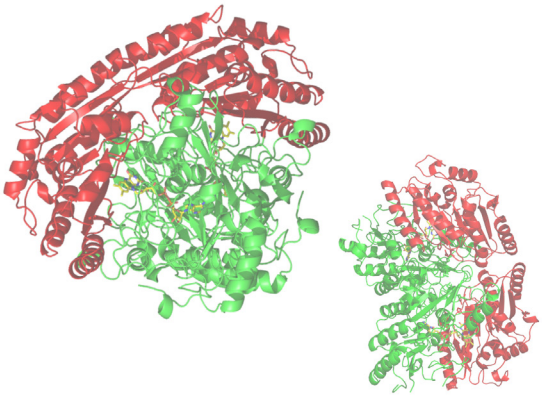
Acknowledgements

With pleasure I would like to extend my thanks to the people who supported me during my PhD:

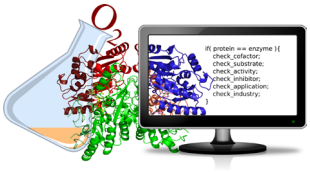
- My promotor, Willem van Berkel: for your marked enthusiasm, the constructive criticism, the free rein and all the challenges you presented me with.
- My collaborators in Barcelona, Fatima Lucas and Victor Guallar: for immersing me in the world of computational modelling - I am not planning to resurface anytime soon.
- My collaborators in Stockholm Xiang Sheng and Fahmi Himo: for entangling me in quantum mechanics - I am looking forward to the productive completion of this reaction.
- My collaborator in Utrecht, Ronald de Vries: for the horizontal transfer of fungal and gene editing knowledge.
- My thesis committee, Prof. Dr. A. J. M. Stams, Prof. Dr. C. Faulds, Prof. Dr. M. W. Fraaije, and Prof. Dr. F. Seebeck: for their time, critical questions and input.
- Adrie Westphal: for technical assistance in anything from "absorption spectra" to "x-, y-, z-coordinates".
- Tom Ewing: for making my PhD like VAO - dimeric.
- Antsje Nolles: for the calm in any storm, lending of your eye to detail and including me in the "spelletjesavonden" and "spelletjesdagen".
- Caroline Paul: for coaching and proof reading - at a critical time of my PhD you had the perfect advice and input for every situation at the ready.
- Mieke Huijbers: for finding my forgotten tubes in the sink and painting the walls of my apartment in the dark.
- Joseline Houwman: for all the vitamins - I should eat more veggies to improve my memory.
- My students (Jimena, Joris, Leroy, Pim, Sindy and Vincent) for teaching me more than I taught them.
- My colleagues at the Laboratory of Biochemistry for all the interesting talks and discussions during working hours and "borrels" - and for all the other reasons I kept bringing chocolate.

- Alexander, Klara, Urs, Niggi, Leila und Christine: für Alles - ohne Euch wäre ich nicht soweit gekommen.
- Alexandra und Henk: meine holländischen Gasteltern - für all die gemütlichen Abende.
- Michaël, Minke, en Wietske: voor de muzikale start van mijn doctoraat - en alle andere praktische hulp.
- Emanuele, Ferran, Gabriela, Gerard, Marina, Pedro, and Sandra: for the warm welcome and the insider-information on Barcelona and the region.
- Ferran, Henrik, Masoud and Mojgan: for warning me off "Surströmming" - one month in Stockholm was way too short.
- Pascal: voor de nerdachtige en metal bijdragen.
- Rahel, Kerstin, Rhea und Séverine für moralische Unterstützung aus der Nähe und Ferne.
- Arlette, Benoît, Christopher, Quentin: merci pour votre amitié et les réunions quinquennales.
- To those who I have possibly omitted: thank you for having a better memory than me - choose yourself what you want me to thank you for and email me.
- I would also like to thank the StackExchange and stackoverflow community for helping me with my programming questions and the anonymous internet users who have released their artwork under the Creative Commons CC0 1.0 on pixabay.

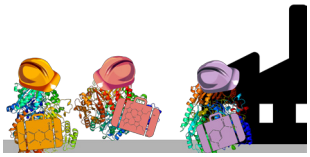
Summary



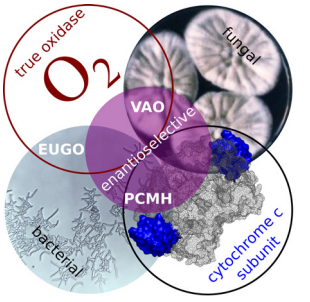
Flavoenzymes are highly diverse enzymes that harness the power of their flavin cofactor to catalyse redox reactions. The fungal, FAD-dependent enzyme vanillyl alcohol oxidase (VAO) has been studied in this thesis through different computational and experimental approaches. This oxidative enzyme, which covalently binds its cofactor via a histidine residue, is active with a wide range of *para*-substituted phenols. Several of these substrates are converted into high-value chemicals, e.g. coniferyl alcohol, vanillin and chiral alcohols.



Interest in oxidative enzymes is booming and several are being prepared for industrial applications or are already used at large scale. Examples are glucose oxidase, used as a biosensor in blood glucose meters for diabetic patients; hexose oxidase, used in the baking industries to improve dough properties, and laccases, which are on their way for applications in the pulp and paper industry. Each application of a new enzyme faces different challenges. Baeyer-Villiger monooxygenases for example are promising biocatalysts for the pharmaceutical industry, but many of them are operationally unstable or suffer from substrate or product inhibition. Research to overcome these obstacles is ongoing and requires the integration of protein engineering with process design. Such integrative approaches profit from the advances made in both fields and are likely to reduce the time needed by individual enzymes to grow into their new role as industrial biocatalysts.

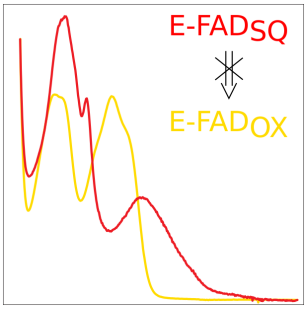


VAO belongs to the VAO/PCMH family, whose members all share the same FAD-binding fold. Within that family, VAO is part of the subgroup of 4-phenol oxidizing enzymes. VAO is the only fungal member of that group and its physiological role in fungi is unclear. Using the sequence and three-dimensional structure of VAO from *Penicillium simplicissimum*, we have identified novel, putative fungal VAOs. We found that these fungal VAOs likely



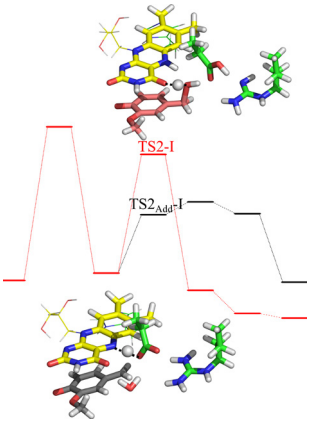
originated from a bacterial ancestor and were transferred to fungi by horizontal gene transfer. Based on the sequences of these novel fungal VAOs we have defined several motifs that should help to resolve misannotations in sequence databases.

In the reaction cycle of VAO, reoxidation of reduced FAD by dioxygen is a crucial step. After product formation, flavin reoxidation results in the release of hydrogen peroxide. VAOs that were heterologously produced in the methylotrophic yeast *Komagataella phaffii* contain a highly stable, non-covalently bound anionic FAD semiquinone. In VAO, this redox state of FAD is atypical and its stability is baffling. We think that enzymes produced in this eukaryotic expression system fell victim to oxidative stress: substrate binding uncouples dioxygen reduction from product formation, and thus results in atypical single electron flavin redox cycling. In other words, in these enzymes the normally quick decay of the *p*-quinone methide intermediate is impaired by an increased energy barrier to the product, resulting in the reaction falling back to the *p*-quinone methide intermediate.



In any enzymatic reaction, substrates have to "find" the active site before catalysis can happen. Because most enzymes are not diffusion limited, this step is often neglected. It is nevertheless of fundamental interest to understand the inner workings of enzymes. The identification of diffusion (or migration) paths based on crystal structures remains challenging due to the rigidity of these tightly packed structures. In VAO, we have identified one path for phenolic ligands and one for dioxygen and hydrogen peroxide to enter and leave the active site. After the substrate has reached the active site, two concierge residues close off the reaction center. Interestingly, in our simulations only the protonated forms of phenolic substrates are able to enter the enzyme.

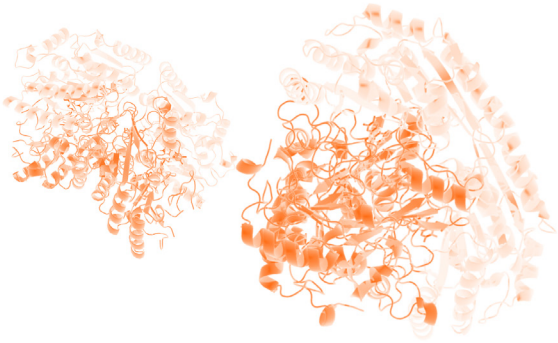
A VAO-catalysed reaction that is of relevance for the food industry is the synthesis of natural vanillin from natural *p*-creosol. VAO catalyses this reaction via the intermediate formation of vanillyl alcohol. However, in the first reaction step, a covalent *p*-creosol-FAD adduct is formed, significantly inhibiting this reaction. We have used quantum chemical modelling to study this process in more detail. The productive reaction to vanillyl alcohol competes with adduct formation due to the lower energy barrier of the reaction pathway to the adduct. An aspartate residue in the active site is crucial in adduct formation, and its flexibility is the cause for adduct formation with *p*-creosol. Bulkier substrates prevent the attack of this aspartate on FAD by steric hindrance.



Oxidative enzymes, such as VAO, are moving into the spotlight of different industries. One strategy to identify enzymes for industrial applications is to mine genomes. Genome mining of VAO-like sequences is hampered by misannotations present in sequence databases. Opening up sequence databases to editing by users would help to resolve misannotations.

Recent experimental and computational research has further advanced the understanding of VAO's reactivity. We have demonstrated that computational tools can successfully be used to study processes out of the reach of experimental techniques. It is crucial to integrate different computational methodologies into one, easy to use, computational enzymology toolbox. Such a toolbox should lower the barrier between the fields of computational and experimental enzymology. Collaborations that successfully connect industry and academia, as well as computational and experimental enzymologists, will decide the future of industrial applications of enzymes.

Samenvatting



Enzymen maken het leven mogelijk - maar hoe doen ze dit? Enzymen versnellen (katalyseren) een chemische reactie door een omgeving te vormen waarin deze reactie sneller kan verlopen. Dit betekent ook dat de energiebarrière, die moet worden overwonnen voor elke chemische reactie, wordt verlaagd. De omgeving in het enzym waarin de reactie kan plaatsvinden wordt het actieve centrum genoemd. De interacties in het actieve centrum zijn vaak geoptimaliseerd voor een specifiek enzym-substraat molecuul paar (door natuurlijke selectie tijdens de evolutie), wat enzymen geeft met verschillende specialisaties (substraatspecificiteiten). De bouwstenen waarmee enzymen worden gebouwd (aminozuren) bepalen welk type reactie een enzym kan katalyseren.

Sommige enzymen gebruiken cofactoren om reacties te katalyseren die met enkel aminozuren niet mogelijk zijn. Vitaminen zijn meestal de voorlopers voor deze cofactoren. Flavoenzymen, bijvoorbeeld het enzym vanillyl alcohol oxidase (VAO) dat is onderzocht in dit proefschrift, gebruikt een flavine cofactor om er redoxreacties mee te katalyseren. VAO is oorspronkelijk uit een schimmel geëïsoleerd en bindt zijn flavine-adenine-dinucleotide covalent via een histidine-aminozuur in het actieve centrum. VAO converteert uitsluitend *para*-gesubstitueerde, fenolische substraten. Veel van deze substraten worden in waardevolle chemicaliën zoals coniferylalcohol, vanilline en chirale alcoholen omgezet. We hebben dit enzym experimenteel in het lab en met de computer bestudeerd.

De belangstelling voor oxidatieve enzymen zoals VAO is in de afgelopen jaren enorm toegenomen en veel van deze enzymen worden gebruikt of voorbereid voor industriële toepassingen. Enkele voorbeelden zijn glucose-oxidase, hexose-oxidase, laccasen of Baeyer-Villiger-monooxygenasen. Glucoseoxidase is een enzym dat de bloedsuikerspiegel voor diabetici kan meten, hexose oxidase wordt in de bakkersindustrie gebruikt om de eigen-

schappen van het deeg te verbeteren en laccasen worden ontwikkeld voor toepassingen in de pulp- en papierindustrie. Elk enzym heeft te maken met verschillende uitdagingen en deze uitdagingen zijn afhankelijk van de gewenste toepassing.

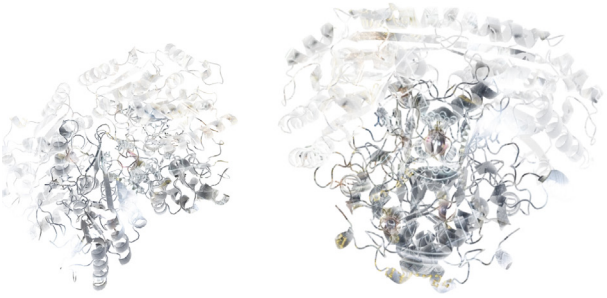
VAOs die werden geproduceerd in de gist *Komagataella phaffii* bevatten allemaal een hoogst ongebruikelijke en stabiele vorm van de flavine cofactor: een niet-covalent gebonden, anionische semichinon. We vermoeden dat deze enzymen het slachtoffer zijn geworden van oxidatieve stress. Door het substraat te binden, worden de normaal gekoppelde reacties die zuurstof verminderen en product vormen ontkoppeld, en de reacie eindigt zonder product te vormen.

In een enzymatische reactie moeten de substraten altijd het actieve centrum van het enzym "vinden" voordat een reactie kan plaatsvinden. De migratiepaden in enzymen kunnen worden geïdentificeerd door berekeningen die de bewegingen van een enzym simuleren. In VAO ontdekten we de migratiepaden van verschillende moleculen. Twee concierge aminozuren sluiten het pad naar de actieve centrum af zodra een substraat het heeft bereikt. Interessant is dat alleen geprotoneerde substraten in in het actieve centrum worden toegelaten.

De synthese van natuurlijke vanilline uit natuurlijke creosoot, gekatalyseerd door VAO, is belangrijk voor de voedingsindustrie. VAO katalyseert deze reactie met het tussenproduct vanillylalcohol. In de eerste reactiestap wordt helaas een stabiel, covalent adduct gevormd tussen het substraat en de cofactor, wat deze reactie sterk vertraagt. We hebben deze reactie in detail bestudeerd met kwantummechanische berekeningen en vonden dat de productieve reactie (waarin vanillyl alcohol wordt gevormd) concurreert met adductvorming.

Oxidatieve enzymen, waaronder VAO, komen steeds meer in de schijnwerpers van verschillende industrieën. We hebben aangetoond dat computergestuurd onderzoek met succes kan worden gebruikt om processen te onderzoeken die buiten het bereik van experimentele technieken liggen. Met het oog op de toekomst van de industriële toepassingen van enzymen moet samenwerking tussen de industrie en de academische wereld worden versterkt en ook de samenwerking tussen de experimentele en computationele enzymologie moet worden uitgebreid.

Zusammenfassung



Enzyme ermöglichen Leben - aber wie machen sie das? Enzyme beschleunigen (katalysieren) eine chemische Reaktion dadurch, dass sie eine Umgebung formen in der diese Reaktion schneller ablaufen kann. Das bedeutet auch, dass die Energiebarriere, welche für jede chemische Reaktion überwunden werden muss, gesenkt wird. Die Umgebung im Enzym, in welchem die Reaktion beschleunigt stattfinden kann, wird das aktive Zentrum genannt. Die Interaktionen im aktiven Zentrum sind oft für ein Enzym-Molekül-Paar optimiert (durch natürliche Selektion während der Evolution), wodurch Enzyme mit verschiedenen Spezialisierungen (Substratspezifitäten) entstanden sind. Die Bausteine, aus welchen Enzyme gebaut werden (sogenannte Aminosäuren), bestimmen, welche Art von Reaktion ein Enzym katalysieren kann - und limitieren damit die Enzyme in ihrem Reaktionspektrum.

Einige Enzyme benutzen sogenannte Kofaktoren um Reaktionen, welche mit Aminosäuren alleine nicht machbar sind, zu katalysieren. Vitamine sind meist die Vorstufen der Kofaktoren von Enzymen. Flavoenzyme, wie das Enzym Vanillyl Alkohol Oxidase (VAO) welches in dieser Doktorarbeit untersucht wurde, benutzen einen Flavin-Kofaktor um Redoxreaktionen zu katalysieren. Dieser Flavin-Kofaktor wird aus Riboflavin (Vitamin B2) hergestellt. VAO ist pilzlichen Ursprungs und bindet seinen Flavin-Kofaktor kovalent durch eine Histidin Aminosäure im aktiven Zentrum. VAO setzt ausschliesslich *para*-substituierte, phenolische Substrate um. Viele dieser Substrate werden in wertvolle Chemikalien wie z. B. Coniferylalkohol, Vanillin und chirale Alkohole umgewandelt. Wir haben dieses Enzym experimentell und mit Computermodellen untersucht.

Das Interesse an oxidativen Enzymen wie VAO ist über die letzten Jahre hinweg massiv gewachsen. Viele solche Enzyme werden für industrielle Anwendungen benutzt oder auf diese vorbereitet. Beispiele beinhalten Glukose-

Oxidase, Hexose-Oxidase, Laccasen oder Baeyer-Villiger Monooxygenasen. Glukose-Oxidase ist ein wichtiges Enzym, welches in den Blutzuckermessgeräten von Diabetikern den Blutzuckerpegel misst, Hexose-Oxidase wird in der Backindustrie gebraucht um die Eigenschaften des Teigs zu verbessern und Laccasen sind auf dem Weg zu Anwendungen in der Zellstoff- und Papierindustrie. Jedes Enzym muss sich anderen Herausforderungen stellen, wobei sich diese je nach der gewünschten Anwendung drastisch unterscheiden.

VAOs, welche in der Hefe *Komagataella phaffii* hergestellt wurden, enthalten alle eine äusserst ungewöhnliche und stabile Form des Flavin-Kofaktors: ein nicht-kovalent gebundenes, anionisches Semichinon. Wir vermuten, dass diese Enzyme oxidativem Stress zum Opfer gefallen sind. Durch das Binden des Substrats werden die normalerweise gekoppelten Reaktionen, welche Sauerstoff reduzieren und Produkt formen, entkoppelt. Dies führt zu einer atypischen Reaktion, in der kein Produkt geformt wird.

In einer enzymatischen Reaktion müssen die Substrate das aktive Zentrum des Enzyms immer "finden" bevor eine Reaktion stattfinden kann. Die Migrationsbahnen in Enzymen können mit Berechnungen, welche die Bewegungen eines Enzyms simulieren, identifiziert werden. So haben wir in VAO die Migrationsbahnen von verschiedenen Molekülen entdeckt. Zwei Pförtner-Aminosäuren schliessen den Weg zum aktive Zentrum ab, sobald ein Substrat dieses erreicht hat. Interessanterweise werden nur protonierte Substrate in das aktive Zentrum hineingelassen.

Die Synthese von natürlichem Vanillin aus natürlichem *p*-Creosol wird von VAO katalysiert und ist für die Lebensmittelindustrie von Bedeutung. VAO katalysiert diese Reaktion mit dem Zwischenprodukt Vanillyl Alkohol. Im

ersten Reaktionschritt wird jedoch ein stabiles, kovalentes Addukt zwischen dem Substrat und dem Kofaktor geformt, was diese Reaktion stark verlangsamt. Wir haben diese Reaktion mit quantenmechanischen Berechnungen detailliert untersucht und herausgefunden, dass die produktive Reaktion, bei welcher Vanillyl Alkohol geformt wird, mit der Adduktbildung konkurriert.

Oxidative Enzyme, unter anderem auch VAO, befinden sich zunehmend im Fokus verschiedener Industriezweige. Wir haben veranschaulicht, dass computergestützte Forschung erfolgreich genutzt werden kann um Abläufe ausserhalb der Reichweite experimenteller Techniken zu untersuchen. Um die Zukunft von industriellen Anwendungen von Enzymen zu sichern, muss die Zusammenarbeit zwischen der industriellen und akademischen Welt gestärkt, sowie die Zusammenarbeit zwischen der experimentellen und computergestützten Enzymologie ausgebaut werden.

Abbreviations

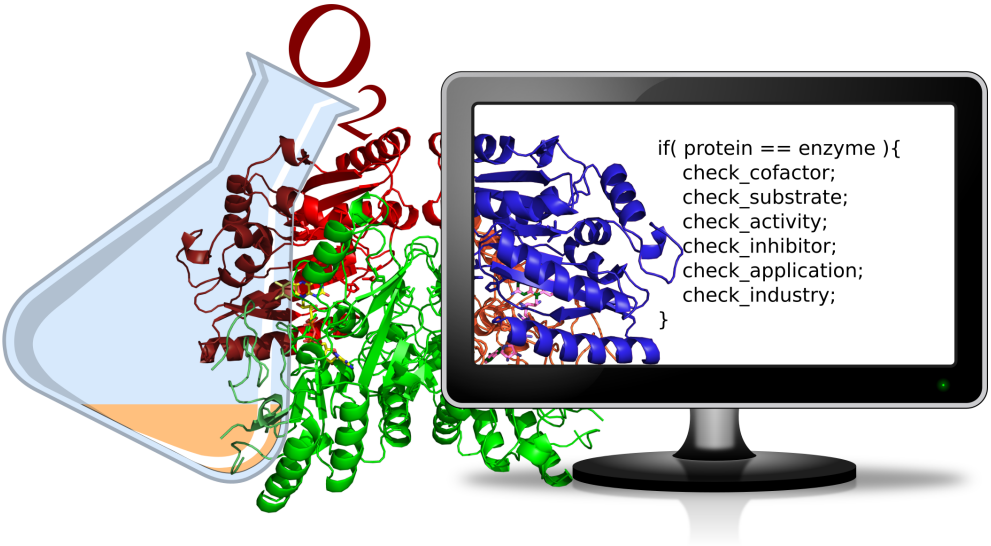
4PO	the 4-phenol oxidising subgroup of the VAO/PCMH flavoprotein family
CAZy	Carbohydrate-Active Enzymes database
EC x.x.x.x	Enzyme Commission number
EUGO	Eugenol oxidase
FAD	Flavin adenine dinucleotide (oxidised)
FADH2	Flavin adenine dinucleotide (reduced)
FMN	Flavin mononucleotide
INDOX	Industrial oxidoreductases
MC	Monte Carlo (the technique, not the place)
MD	Molecular dynamics
NAD^+	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
PCMH	*p*-cresol methylhydroxylase
PDB	Protein data bank
PELE	Protein energy landscape exploration
QCC	Quantum chemical cluster
QCM	Quantum chemical modelling

QM	Quantum mechanics
QM/MM	Quantum mechanical/molecular mechanics
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
VAO	Vanillyl alcohol oxidase

Amino acid residues (three-letter/one-letter abbreviation). A number after the abbreviated amino acid residue indicates its position in the protein sequence.

Ala/A	Alanine
Arg/R	Arginine
Asn/N	Asparagine
Asp/D	Aspartate
Cys/C	Cysteine
Glu/E	Glutamate
Gln/Q	Glutamine
Gly/G	Glycine
His/H	Histidine
Ile/I	Isoleucine
Leu/L	Leucine
Lys/K	Lysine
Met/M	Methionine
Phe/F	Phenylalanine
Pro/P	Proline
Ser/S	Serine
Thr/T	Threonine
Trp/W	Tryptophan
Tyr/Y	Tyrosine
Val/V	Valine

Chapter 1:
Introduction



From Oxygen to Flavoenzyme Oxidases

A short history of oxygen. Oxygen is the third most abundant chemical element in our universe, preceeded only by hydrogen and helium. It is produced in stars through helium burning, in the so-called Carbon-Nitrogen-Oxygen cycle [1, 2]. On Earth, oxygen is present in numerous chemical substances, for example water, carbon dioxide, amino acids, and as dioxygen in air. The atmosphere that we breathe today contains approximately 21% dioxygen, which has not always been the case. Dioxygen concentrations were very low during the first half of Earth's existence (Figure 1). For that time, it is generally estimated that dioxygen concentrations in the atmosphere were as low as 0.001% of modern oxygen concentrations [3]. But between 2.4 and 2.1 billion years ago, the concentrations of dioxygen in the atmosphere increased permanently. This 'Great Oxidation Event' (GOE) led to our modern atmosphere and was caused by photosynthetic organisms, see Figure 1.

Cyanobacteria, as well as plants and algae, are able to do photosynthesis. Plants and algae obtained the ability

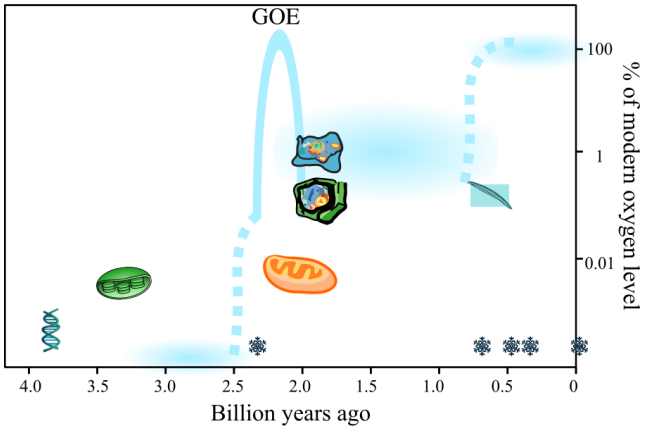


Figure 1: The organisms causing the 'Great Oxidation Event' were cyanobacteria, the earliest organisms capable of oxygenic photosynthesis [3]. Photosynthesis uses water as an initial electron donor to produce sugar from carbon-dioxide and forms dioxygen as a byproduct.

to perform photosynthesis through endosymbiosis of an ancestor of modern cyanobacteria [4, 5, 6]. This endosymbiontic cyanobacterium is called a chloroplast, the powerplant of plants and algae. Other bacteria use up dioxygen as a substrate in a process called oxidative phosphorylation. Oxidative phosphorylation is one of many oxidation reactions wich are essential for living organisms. One bacterium capable of oxidative phosphorylation was acquired by eukaryotes through endosymbiosis leading to modern eukaryotes like fungi and animals. This form of endosymbiosis resulted in mitochondria, the "powerhouses of the cell".

The term "**powerhouse of the cell**" was coined by Philip Siekevitz in 1957 [7].

An oxidation reaction has historically been defined as a reaction where dioxygen is consumed. Since the 19th century, the term oxidation is used more broadly and describes a reaction in which electrons are transferred from one molecule to another, e.g: $2H_2 + O_2 \rightarrow 2H_2O$. In this reaction, hydrogen gas is oxidised (by donating two electrons) and oxygen is reduced (by accepting the electrons). Because one reaction partner is reduced and one is oxidised, oxidation reactions are also called redox reactions. Redox reactions can also be catalysed by enzymes. However, because the amino acids that form an enzyme are only able to do acid-base and covalent catalysis, additional molecules are needed as electron source: cofactors. The cofactor can be organic or inorganic (e.g. copper, iron, vanadium). Examples of organic cofactors are nicotinamide adenine dinucleotide (NADH or NAD^+, depending on its oxidation state), nicotinamide adenine dinucleotide phosphate (NADPH or $NADP^+$, depending on its oxidation state) and flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN) or riboflavin, see Figure 2.

Cofactor: a substance (other than the substrate) whose presence is essential for the activity of an enzyme [8]. Without its cofactor bound, the enzyme is inactive and is called an apoenzyme. With its cofactor bound, covalently or non-covalently, the enzyme is active and is called a holoenzyme.

Enzymes and their cofactors. Cofactors are sometimes also called coenzymes or cosubstrates or prosthetic groups. While there are clear biochemical definitions for

these terms, they are sometimes used as synonyms, depending on the field of research [9, 10, 11, 12]. Here, the definition from the Oxford dictionary will be used, see margin note.

The cofactor FAD is bright yellow in solution (when oxidised) and enzymes which bind it (or FMN) are flavoenzymes. In FAD's chemical structure, the respective phosphate group of the flavin mononucleotide moiety (in blue in Figure 2) is connected to that of the adenosine monophosphate (in red in Figure 2). FAD is a redox cofactor and can exist in five different redox states, which have different colours: fully oxidised (FAD, yellow), anionic semiquinone ($FAD^{\bullet -}$, red), neutral semiquinone ($FADH^{\bullet}$, blue), hydroquinone ($FADH_2$, colourless) or the deprotonated, reduced FAD ($FADH^-$, colourless). Switching between these five states can be reversible. Although generally the oxidation state of FAD is mentioned, it is only its isoalloxazine ring that is reduced or oxidised (in black, together with the ribityl chain, in Figure 2). The different redox states of FAD are depicted in Figure 3, but are also valid for other flavins. Flavoenzymes use and modulate the unique and versatile properties of FAD to catalyse a wide range of chemical reactions.

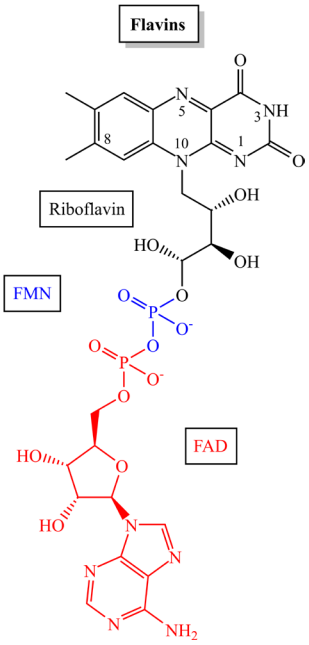


Figure 2: Structures of the flavin cofactors. Riboflavin in black, FMN in black and blue and FAD in black, blue and red.

Flavoenzymes are critical to redox reactions essential for life on Earth. The most important part of the flavin cofactor is its isoalloxazine ring, which forms the "catalytic heart of a flavoenzyme" [13]. In some cases, modified versions of the isoalloxazine ring are used. Examples of such modifications are depicted in Figure 4 on page 31.

The covalent modifications of the flavins fine-tune their reactivity by affecting their redox properties. The protein environment of the cofactor also helps to tune these redox properties, which leads to different reactivities in various flavoenzymes with the same cofactor. Most flavoenzymes distinguish strictly between FAD and FMN,

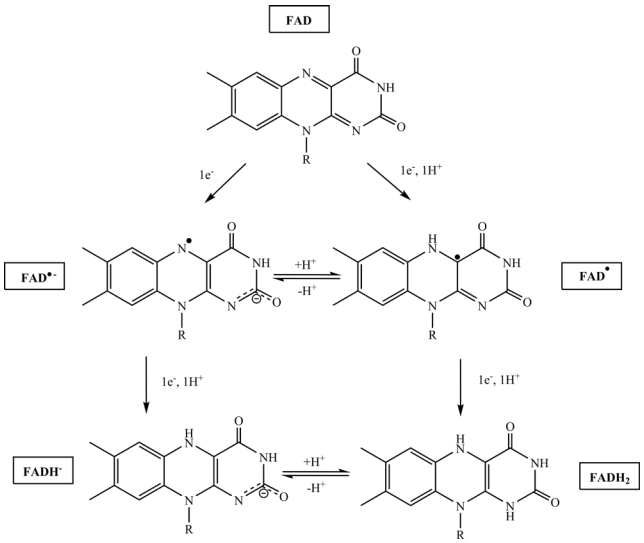


Figure 3: The five redox states of FAD. FAD can switch between these different states by accepting two protons and two electrons in its fully oxidised form to become fully reduced, colourless $FADH_2$. It can also switch from its fully oxidised state to the anionic semiquinone radical by gaining one electron, changing its colour to red ($FADH^{\bullet-}$), or become the neutral semiquinone through the gain of one electron and one proton, changing its colour to blue ($FAD^{\bullet}$). From the semiquinone state, $FAD^{\bullet}$ becomes the hydroquinone ($FADH_2$) through the acceptance of one proton and one electron, while $FAD^{\bullet-}$ becomes the deprotonated, reduced flavin ($FADH^{-}$), which is colourless.

using only one of the two, while others are more flexible and are able to use either. Examples of flavoenzymes that accept only FAD *in vivo* are vanillyl alohol oxidase and berberine bridge enzyme [14, 15]. A flavoenzyme that was recently found to be indiscriminating between FAD and FMN is proline dehydrogenase [16]. An example of a flavoenzyme that binds FMN is old yellow enzyme [17]. A flavoprotein that binds riboflavin is riboflavin-binding protein, which is present in the eggwhite of chicken eggs [18]. Riboflavin (also known as vitamin B_2) is essential to organisms because it is the precursor of FAD and FMN.

Flavoenzymes can be classified into five groups, based on the swiftness of their reaction with dioxygen and the reaction they catalyse. Firstly, there are oxidases, which react rapidly with dioxygen to produce oxidised flavoprotein and hydrogen peroxide. Secondly, there are dehydrogenases, which react very slowly with dioxygen. Thirdly, there are monooxygenases, which break the bond linking the two oxygens in dioxygen and introduce one oxy-

gen atom into their substrates. The other oxygen atom is reduced to water. Fourthly, there are reductases, which mainly pass electrons from NAD(P)H to their substrate. And, last but not least, there are disulfide oxidoreductases, which contain active-site thiols and use a dithiol substrate and NAD^+ (or NADH in the case of peroxidases that are members of this family). They form disulfide products and NADH or a reduced dithiol product and $NAD(P)^+$ [13]. These five groups illustrate that flavoenzymes catalyse markedly diverse reactions. They are therefore present in many different biological pathways and are of interest for industrial applications [13, 20, 21, 22, 23]. They are colourful and versatile enzymes, due to their cofactor, and can be yellow, red, blue, green, purple or even pink [24].

For the remainder of this introduction, the focus will be on flavoenzyme oxidases using FAD as cofactor. Flavoenzyme oxidases with an FAD cofactor can be found in different protein families, metabolic pathways and industrial reactions [25, 26, 27]. They are particularly interesting due to their ability to tame oxygen. Despite the essential role dioxygen plays in life, it remains a difficult and dangerous molecule.

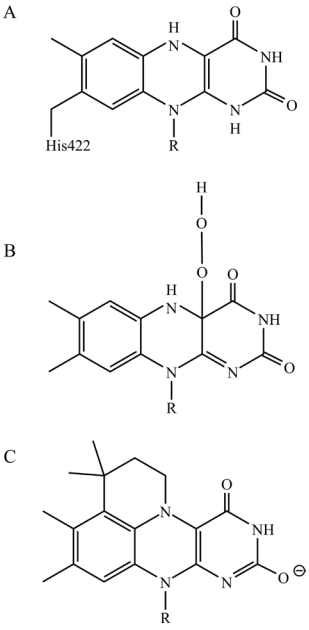


Figure 4: Modified flavins.
A: Covalently bound FAD in vanillyl alcohol oxidase.
B: Hydroperoxy flavin.
C: Prenylated flavin, as observed in UbiD [19].

The trouble with oxygen. The difficulty with dioxygen lies in its electron spin configuration. Dioxygen is dominantly present in its triplet spin state, leaving two unpaired electrons of parallel spin available for a reaction. In flavoenzyme oxidases, reduced flavin needs to be oxidised for the reaction cycle to begin anew with oxidised flavin. However, the spin state of reduced flavin is singlet, meaning no unpaired electrons are present and the spins of the two paired electrons are antiparallel (see the reaction shown in the margin note). The reaction cannot proceed under these conditions because bonds can only be formed between electrons of antiparallel spin (Hund's rule) and the spin of each electron needs to be conserved. Therefore,

Spins in dioxygen and FAD. Spin configuration of dioxygen and FAD and the mechanism of reaction to avoid a "spin-forbidden" reaction [28]. Square brackets mark the "caged radical pair".

$$\mathrm{FADH^-}(\uparrow\downarrow) + \mathrm{O_2} \rightarrow (\uparrow\uparrow)$$
$$[\mathrm{FADH^\bullet} \uparrow\uparrow \mathrm{O_2^-}] \rightarrow$$
$$[\mathrm{FADH^\bullet} \uparrow\downarrow \mathrm{O_2^-}] \rightarrow$$
$$\mathrm{FADHOO^-} \overset{+\mathrm{H^+}}{\rightleftharpoons} \mathrm{FADHOOH} \rightarrow$$
$$\mathrm{FAD} + \mathrm{H_2O_2}$$

the spin state of dioxygen needs to be changed first for it to undergo a reaction with flavin. Reduced flavin transfers an electron to dioxygen, resulting in a "caged radical pair" [28]. The "caged radical pair" is a highly unstable state and the spin of one electron in triple state will invert so that the electrons are in a singlet spin state (still in the "caged radical pair"). Now a reaction can occur between the two electrons in singlet spin states and a highly unstable intermediate, flavin hydroperoxide (FADHOOH), is formed. Flavin hydroperoxide then dissociates heterolytically into oxidised flavin and hydrogen peroxide. Hydrogen peroxide is formed because dioxygen keeps both bonding electrons and hydrogens and is thus reduced.

The trouble with dioxygen is the formation of radical intermediates during reactions. These radicals need to be carefully contained within the enzyme lest they damage the enzyme or organism. Some enzymes are better at containing radicals, leaving them functional for more reaction cycles (also called turnovers). Additionally, the formation of hydrogen peroxide by oxidases can be risky as this molecule can be toxic to organisms and detrimental to enzyme activity. Therefore, enzymes that excessively produce hydrogen peroxide are mostly located in a specialised organelle in the cell called the peroxisome. Regulation of hydrogen peroxide concentrations can be achieved through coexpression of catalase enzymes [29]. Hydrogen peroxide also has important functions as a signalling molecule. Low levels of hydrogen peroxide are required in growth regulation, whereas increasing hydrogen peroxide concentrations can initiate expression of antioxidative genes [30].

"All things are poison and nothing is without poison; only the dose makes a thing not a poison."
Paracelsus, ca. 1538

Vanillyl Alcohol Oxidase

Vanillyl alcohol oxidase (VAO) is a flavoenzyme oxidase that was first isolated from the fungus *Penicillium simplicissimum* in 1992 [31]. The fungus was discovered in paper-mill waste-water in an attempt to identify organisms able to grow using veratryl alcohol as their sole carbon source [32]. Veratryl alcohol is a naturally occuring, aromatic compound and a product of lignin degradation, where it stabilises lignin peroxidases [33]. Interestingly, *P. simplicissimum* is not ligninolytic, but able to use intermediates of lignin degradation as sole carbon source. In *P. simplicissimum*, the degradation pathways of veratryl alcohol produce protocatechuate as a ring-fission intermediate [32]. Growth of the fungus with different lignin degradation intermediates revealed that veratryl alcohol and 4-(methoxymethyl)phenol induce expression of a yellow enzyme: VAO [31].

Expression of VAO is not induced by vanillyl alcohol in *P. simplicissimum*. However, in another fungus, *Byssochlamys fulva*, the expression of a homologous VAO is induced by vanillyl alcohol [34]. In *P. simplicissimum*, 4-(methoxymethyl)phenol is the only molecule identified so far that induces expression of VAO and is also a substrate, therefore it could possibly be the natural substrate of VAO [35].

Substrate screening with VAO revealed that it is active with an extensive range of substrates and catalyses different types of reactions with them. Benzylic alcohols and benzylic amines are oxidised, 4-allylphenols are hydroxylated, 4-alkylphenols are hydroxylated or dehydrogenated and 4-(methoxymethyl)phenol is demethylated [36, 37]. Looking at these substrates (see Figure 5), it becomes clear that they have one thing in common: all are *para*-substituted phenols. Therefore, there must be a mecha-

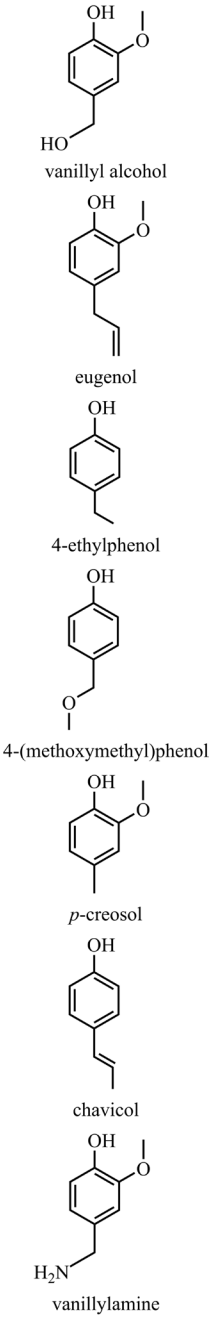


Figure 5: Some of the substrates of VAO.

nism in place in VAO which requires this property of its substrates.

There are several other interesting properties of VAO that need to be mentioned. The amino acid sequence of VAO consists of 560 amino acids, which makes it a sizeable enzyme. But the enzyme is even bigger once it has folded into its three-dimensional structure. *In vitro*, the quaternary structure of active VAO is homodimeric or homooctameric [38], Figure 6). VAO's quaternary structure can also be called a tetramer of dimers. *In vitro*, octamerisation is FAD and salt dependent, with the fraction of octameric VAO increasing with increasing salt concentrations [14, 38]. Both forms have comparable activity [38]. Monomeric VAO is not observed *in vitro*. Ewing *et al.* could show that a surface loop is essential for octamerisation of VAO [39].

Each subunit in VAO consists of two domains, the FAD-binding and the cap domain. The active site is located at the interface of these two domains, meaning that the dimer contains two and the octamer eight active sites. As their names indicate, the FAD-binding domain is where the cofactor binds and the cap domain closes over the active site like a cap. In the active site, the isoalloxazine ring of FAD is covalently bound to His422 *via* a 8α-(*N*3-histidyl) bond. Ironically, this histidine is located in the cap domain in VAO, not the FAD-binding domain [40]. The phosphate-ribityl chain of FAD and the adenosine monophosphate however are bound in the FAD-binding domain through non covalent interactions [40].

The post-translational, covalent inclusion of FAD in the binding pocket is conditional to the proper folding of the enzyme after expression of the polypeptide. Once folding has occurred, VAO oligomerises to the dimer. FAD will then migrate into the apoprotein and be encased. Binding only occurs if the FAD binding cavern has already been

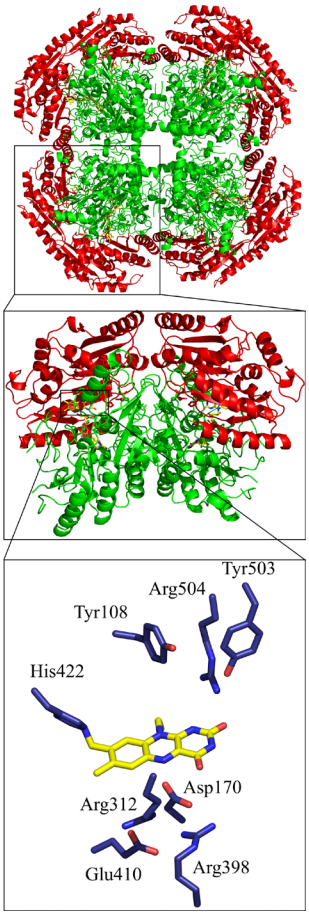


Figure 6: Oligomerisation state of VAO. Octameric VAO at the top, the square indicates the location of the dimer (middle). In the dimer, the square shows the location of the active site and selected residues (as depicted at the bottom). FAD is shown in yellow and amino acid residues are coloured in blue.

formed [14, 41, 42]. Although the FAD is not yet covalently bound in the holoenzyme at this stage, it already promotes octamerisation. FAD is covalently bound to His422 in an autocatalyic process for which the reduced form of FAD is required [41, 43]. Once FAD is covalently bound it is reoxidised by dioxygen. It was found that this covalent binding of FAD increases the redox potential of FAD [43, 44]. This increase of the redox potential (from -65 mV in non-covalent VAO to 55 mV in covalent VAO [43, 44]) increases the oxidative power of FAD.

Several other residues in the active site have been shown to be of importance for catalysis in VAO (see Figure 6) [39, 41, 42, 43, 44, 45, 46]. An aspartate is of great importance during catalysis and the covalent inclusion of FAD in VAO. This Asp170 additionally increases the redox potential of FAD and has been proposed to act as active site base, activating a water molecule involved in substrate conversion [40, 43]. The position of Asp170 and a threonine (Thr457) in the active site has been shown to be responsible for the enantioselective hydroxylation of 4-alkylphenols, elegantly demonstrated by inversion of enantioselectivity of a variant [46]. Two tyrosines (Tyr108 and Tyr503) are of special interest in VAO. Based on the crystal structures available, it was suggested that these two tyrosines are involved in deprotonation of the *para*-substituted phenolic substrates, creating the phenolate inside the actives site [40]. This hypothesis was put to the test recently by Ewing *et al.*, but the detailed mechanism of this deprotonation remains elusive [49].

During the different stages of VAO catalysis, FAD is switching between its oxidised and reduced state. The reaction cycle of VAO is therefore split into two half-reactions, the reductive and the oxidative one. In the reductive half-reaction, FAD gets reduced upon substrate binding through hydride transfer. This generates the quinone methide intermediate of the substrate, which

A word about redox potentials. Every member of a redox couple, such as FAD and VAO's substrate, has its own redox potential, which is a measure of their tendency to lose or gain electrons. The partner in the redox couple with the more negative redox potential will be reduced. In VAO, this is FAD. With an increasing difference in redox potentials between the redox partners, the energy changes involved also increase. No redox reaction can take place anymore if this change in energy becomes too significant. Increasing the redox potential of FAD therefore means that the redox potentials between FAD and the substrate approach each other, making the redox reaction more energetically favourable. Therefore, VAO with a covalently bound FAD can react with substrates with higher redox potentials than VAO with a non-covalently bound FAD.

Crystal structures are structure models of proteins. Proteins are crystallised and the obtained crystal is shot with an X-ray beam. Before the crystal is destroyed, the electrons in the crystal diffract the X-ray beam. These diffraction patterns are linked to the location of electrons in the crystal and give the necessary information to determine a protein structure. The better defined the crystal, the better the diffraction pattern and the higher the certainty about the position of the electrons and thus the higher the resolution of the crystal structure. High resolution protein structures have resolutions below 1.0 Å. Crystal structures can be obtained *in cellulo* and it is also possible to perform time-resolved experiments to study dynamic processes [47, 48].

then continues to react. Depending on the type of substrate, different reactions occur with this intermediate. Figure 7 illustrates the reaction mechanism of VAO with 4-(methoxymethyl)phenol, VAO's proposed physiological substrate [35].

It is in the family. VAO is part of a larger family of structurally related flavoenyzmes, called the VAO/PCMH family. Whereas these enzymes are highly diverse in function, they all share the same two-domain structure. Of these two domains, the FAD-binding domain is the most conserved and more variation is observed in the substrate-binding domain (corresponding to the cap domain in VAO). A handful of the characterised enzymes in this family share similar catalytic features with VAO. Amongst them are *p*-cresol methylhydroxylase (PCMH) and eugenol oxidase (EUGO), the two closest relatives of VAO for which three-dimensional structure models are available. Other close relatives for which less experimental data is available are eugenol dehydrogenase (also called eugenol hydroxylase) and a decarboxylase [50, 51, 52, 53, 54]. Within the well characterised members of this subgroup of the VAO/PCMH family, the requirement for *para*-substituted substrates is prevalent, as is the covalently bound cofactor. Nevertheless, their substrate scope differs, as does their oligomerisation state. Unlike VAO, which forms homooctamers, EUGO exclusively forms homodimers, and PCMH dimers of heterodimers, consisting of two FAD-subunits as well as two cytochrome c subunits [40, 55, 56]. PCMH requires this cytochrome c subunit to reoxidise FAD, a feat that both VAO and EUGO can achieve through dioxygen [31, 55]. The difference in substrate scope of these enzymes (see Figure 8) is most likely due to the variation of active site residues involved in catalysis.

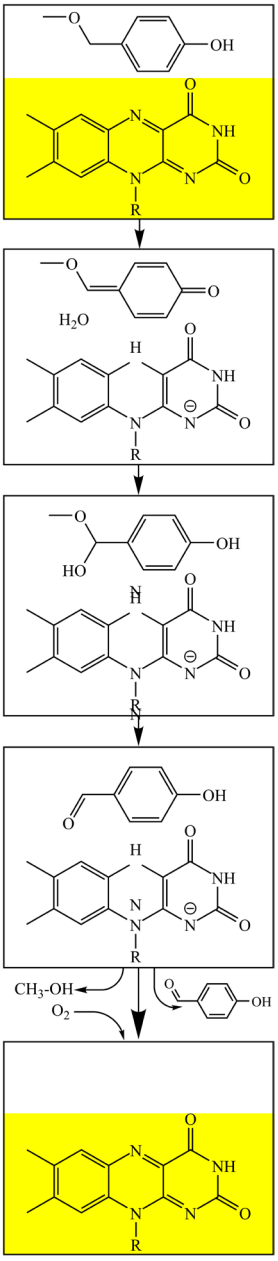


Figure 7: Reaction mechanism of VAO with 4-(methoxymethyl)phenol.

Wet lab vs dry lab

Since the dawn of science, experiments have been at the core of scientific research. The earliest scientist called themselves philosophers, a constant reminder of which is the "*Philosophiae Doctor*" (Doctor of Philosophy, PhD). One could argue that the techniques and tools used at that time were crude compared to today's standards, but eureka moments nevertheless came to some of them - albeit in their bathtub. Archimedes is also the supposed inventor of what is believed to be the very first computer, the Antikythera mechanism [57, 58, 59].

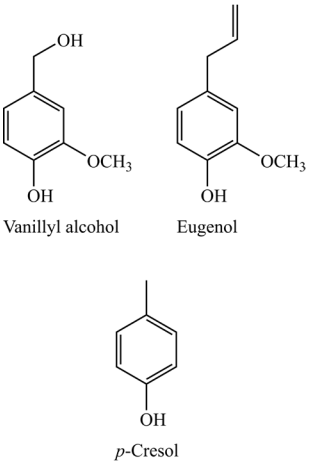


Figure 8: Name-giving substrates of VAO, EUGO and PCMH. It is of importance to know that these substrates are not always the ones that are converted most efficiently, nor are they necessarily the physiological substrate.

Wet lab

Since the times of Archimedes, the wet lab has changed drastically. It is a place where the object of research is no longer observed by the human eye, but through machines with detectors that far exceed that organ. Advances in microbiology in the past century have lead to biotechnology being a widely used tool in the life sciences. It is now possible to take genes from any organism and place them into another organism to study their function in detail. These genes can be expressed to produce proteins and enzymes of interest to scientists or industries. The produced proteins can then be studied through techniques like mass spectrometry, electron microscopy, NMR or crystallography. Cofactors bound to the enzyme can be identified through fluorescence and absorption spectroscopy. Stopped-flow instruments can be used to study the first milliseconds of an enzyme reaction. Studies at the single molecule level are also possible through the use of advanced microscopy techniques as well as microfluidic systems [60, 61].

Induction or conditional expression of a *gene* in an organism When expression of a *gene* is induced in the organism it originates from, this is homologous expression. If the *gene* is expressed in another organism, e.g. *Escherichia coli*, this is heterologous expression and *E. coli* is the host organism. Several substances can be used to induce expression in the host organism, depending on the promoter placed before the *gene of interest*.

Many properties of enzymes and their mechanisms can be solved using these and more elaborate techniques. Nevertheless there are currently limitations to the resolution that can be achieved in such experiments. Whereas fantastic advances have been achieved to experimentally study proteins and enzymes in almost atomistic detail, it is not yet possible to follow an enzyme reaction in atomistic and sub-atomistic detail [62, 63, 64, 65]. This is a requirement of studying a chemical reaction, where electrons are exchanged between molecules.

Here, we climb out of the bathtub and into the computer.

A sequence alignment is created when gaps are inserted in sequences to create overlap in similar regions. The three words (or very short sequences) "GUDRN", "GDRUN" and "GUDRUN" can be aligned to give the following multiple sequence alignment:

```
G  U  D  R  -  N
G  -  D  R  U  N
G  U  D  R  U  N
```

Dry lab

The first modern day computers were men and women whose job it was to compute routine calculations [66]. With time, modern, machine based computers were developed and we are now on the advent of artificial intelligence. Although discussions whether the Turing test has been passed already are still ongoing, it will surely not take much longer for computers to challenge our perception of them further. Unlike these "smart" computers, other modern day supercomputers are functioning more like the first human computers. They assist scientists in rapidly calculating the solutions to complex problems. Several such problems that are relevant for this thesis will be described in more detail below.

BLAST stands for: Basic Local Alignment Search Tool. It starts with a short part of the sequence (a word) to align it to a subject sequence. This initial alignment is then extended and saved and a report is calculated if the extended alignment is better than the initial one. Typical elements of the report are the score and the E value. The score is higher if the quality of the alignment is better and if sequences are more similar. The E value gives an indication of how probable it is to get the alignment by chance. The lower the E value, the more similar the sequences are, and the lower the chance that this similiarity is accidental.

To compare sequences of related proteins, bioinformatic studies are needed. Bioinformatics relies heavily on the use of supercomputers. Behind every assembly and annotation of full genomes, the search of sequence databases or the building of multiple sequence alignments is a supercomputer. When one is looking for sequences similar to a query sequence with

Homology. Homologous enzymes share a common ancestor but may have dissimilar function. Heterologous enzymes do not share a common ancestor but may have similar function. There is no percentage of homology or heterology as these terms are binary and refer to the fact that there was a common ancestor or not.

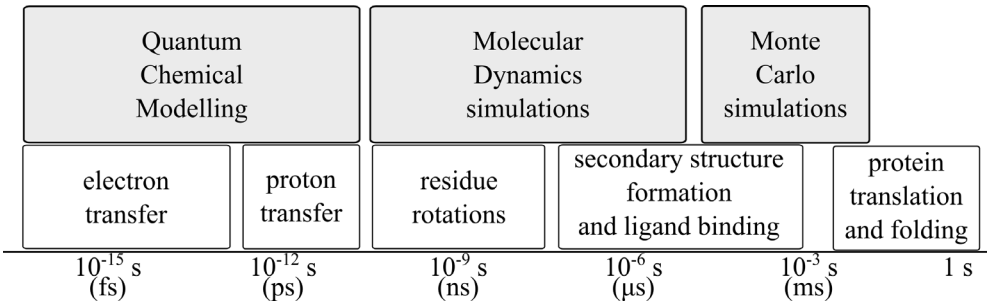


Figure 9: A selective overview of the timescales of biochemical events and the *in silico* approaches that can be used to study them. Protein translation and protein folding can in theory be studied by using Quantum Chemical Modelling (QCM), but such a calculation would take years to finish with the most advanced computers available today. On the other hand, QCM can be used to study electron transfers and enzyme reactions by only modelling a small part of the enzyme. To study "slower" processes, such as ligand binding to an enzyme, becomes computationally expensive and approximations need to be made or combinations of techniques need to be used. Every technique thus has its limitations, making each suited to answer different research questions.

the BLAST webserver (`https://blast.ncbi.nlm.nih.gov/Blast.cgi`), a supercomputer is remotely accessed. It will search all the different databases and perform pairwise alignments of the query against other sequences. When one builds a multiple sequence alignment using clustalomega (`http://www.ebi.ac.uk/Tools/msa/clustalo/`), which can align up to 4'000 sequences, a different supercomputer is accessed remotely and performs all the calculations necessary to align these sequences. From such an alignment, a homology model of a protein can be built (using e.g. SWISS-MODEL `https://swissmodel.expasy.org/`). All these tools are very easy to use and biochemical research without them has become unimaginable. Homology models link bioinformatics to structural bioinformatics. In this field, different methods of computer modelling of proteins are available today, which will be discussed in more detail below.

There are three main methods of structural bioinformatics that are relevant to the work presented in this thesis: Quantum Chemical Modelling, Molecular Dynamics and Monte Carlo. All three can be applied to answer different research questions, but the time it takes to obtain results may vary significantly depending on the method used. Figure 9 illustrates the most common use of the different methods and the biochemical events that can be studied with them.

Proteins have four structure levels. The primary structure, a linear sequence varying in length of the 20 natural amino acid residues (also called residues). The primary structure organises itself (folds) into α-helices and β-sheets, forming the secondary structure. The tertiary structure is formed when the secondary structure elements fold into one fully organised structure. And finally, the quaternary structure results from tertiary structure elements interacting together. Each of the tertiary structure elements in this quaternary structure may be called a subunit. As can be seen from Figure 9, protein folding (the process of a primary structure organising itself into a tertiary structure) is one of the slower biochemical processes.

With Molecular Dynamics, Newton's equations of motion of all the atoms in a protein are solved to model the movements of the protein. With Monte Carlo approaches, the different configurations of a system are generated and sampled to calculate properties of the system. With Quantum Chemical Modelling, transition states are hunted down and their energies are calculated. The detailed description of these techniques goes beyond the purpose of this introduction, but an overview will be provided.

Molecular Dynamics Proteins are moving structures and their movements are essential to their function. But the crystal structure of a protein is fixed and only little to no information on the movements of the protein is available from it. To study the functional movemenents of proteins, Molecular Dynamics simulations (MD) can be used. In MD simulations, atoms and bonds are considered as balls and strings. The string represents different properties of a covalent bond between atoms, such as the length of a bond, the bond angle and the bond dihedral angle. Covalent bonds between different amino acids determine the structure of a protein. Additionally, there are non-covalent

The very basic steps of a classical **Molecular Dynamics** simulation are:

1. Compute the forces on all the atoms in the protein (determine the potential energy of all the atoms).
2. Solve the equation of motion for all the atoms in the protein (prepare to move the atoms in agreement with the interactions they have with one another).
3. Move the atoms.
4. Do step 1., 2. and 3. a few times and then save all the positions of all the atoms in the protein.
5. Go back to step 1.

To follow these instructions for one atom is trivial, however an average amino acid in a protein consists of 20 atoms and an average protein consists of 400 amino acids. Therefore, the calculation of protein movements can become challenging, especially for larger proteins.

interactions without whom a protein could never form any kind of secondary structure: the van der Waals interactions and Coulomb interactions. In MD simulations, the sum of all these interactions is calculated, producing a "movie" of the continuous movements of proteins.

Monte Carlo In Monte Carlo simulations (MC), random sampling is used to explore the properties of a system. The challenge is to sample in the correct region and to sample truly at random. Using MC simulations, it is possible to model the movements of a protein as a molecule diffuses through it. These slower movements are computationally too expensive to calculate with classical MD simulations (see Figure 9). The longer a process takes (fs to s), the longer its simulation takes. Faster computers do help to partly overcome this issue, allowing simulations of processes in the μs timescale. However, for now, approximations need to be made to model longer processes. One such approximation is described below.

Protein Energy Landscape Exploration (PELE) is a MC based approach [67]. PELE can be used to simulate a substrate migrating into an enzyme's active site. The substrate can be placed at a random position on the surface of the protein or close to a hypothetical entrance to the active site. PELE then proceeds as follows:

1. Calculate the initial energy of the system.
2. Perturb both the ligand (rotation + translation + internal degrees of freedom-rotamers) and the protein (by applying normal modes to the backbone).
3. Model side chains by a hierarchical tree scheme using experimental side chain rotamer libraries. Then, relieve steric clashes, perform clustering and minimisation.

A typical Monte Carlo simulations follows these steps:

1. Generate a starting configuration of the system.
2. Select one particle in the system at random.
3. Move that particle to a new, random position (conformation).
4. Accept the new position of the particle (or not).
5. Calculate the desired property of the system.
6. Continue with step 2.

Perform your own Monte Carlo simulation:

1. Place 10 coins on this page. You are not allowed to place a coin on top of another one. Depending on what coins you use, space will become quite limited.
2. Now take one coin away at random and ...
3. ... put it in a new place.
4. It is not allowed to place a coin on top of another coin. If the coin is placed on top of another one, move the coin again (reject this conformation).
5. Measure all the distances between the coins and calculate the average distance between them.
6. Continue with step 2.

Are you sure you did everything at random? If you expand this idea and use it to find the highest mountain in the world, it becomes clear that there is no point in sampling in the Netherlands.

4. Minimise the full system with a truncated Newton method.
5. Calculate the final energy of the system.
6. Accept or reject by using a Metropolis criterion.
7. Go back to step 1.

The random movements of the substrate introduced by the MC step allows the fast, non-continuous exploration of conformations.

The continuous exploration of conformations through MD is computationally expensive and usually only a handful of simulations are produced. By using PELE, hundreds of simulations can be calculated and a statistical analysis of the most used migration path of a substrate can be performed.

Modelling enzymatic reactions with Quantum Mechanics. To computationally model enzymatic reactions, electrons have to be taken into account. Purely MD techniques cannot work with units smaller than atoms and therefore require the addition of quantum mechanics (QM). Oversimplified, all QM methods need to approximate the Schrödinger equation. An exact solution of the Schrödinger equation is currently impossible for any system bigger than one electron and therefore also the systems used in this thesis. One approximation to the Schrödinger equation is the Born-Oppenheimer approximation, which considers the nuclei as static, given that they are much heavier and slower than electrons. Other approximations concern the molecular orbitals (using Hartree-Fock theory), using different pre-defined sets of one-electron functions (known as basis functions or sets).

Hybrid quantum mechanical/molecular mechanics (QM/MM) approaches model the active site of an en-

Computationally expensive simulations. Every computer works by using Central Proccessing Units (CPUs, also called nodes) to follow instructions of a program. My personal computer runs with a core i5 microprocessor. This currently means that I have four cores (the basic computation unit of the CPU) available that programs can give commands to. It also means that a program that needs to get many things done quickly can use more than one core to get the job done. On a supercomputer, I can get access to more cores so that programs can run even faster. To calculate all the possible interactions between all the atoms of an enzyme will take at least four times longer on my personal computer (with four cores) than on 16 cores of a supercomputer. Of course a supercomputer does not only have 16 cores, but thousands of cores. Using 64 cores for 1 h costs approximately 2 €. A simulation that runs for 48 h therefore costs approximately 96 €, but one running for 1 month costs 1488 €. The more cores you want to use, the more expensive the cost per hour becomes. Cloud computing is also an option, with one hour of 128 cores being sold today for ca. 5$ [68].

Atomic units and energies The atomic unit of length, the *Bohr radius*, a_0 corresponds to 0.53 Å. 1 Å correspons to 10^{-10} m. Energies are measured in *hartrees*, with 1 hartree corresponding to 627.509 kcal/mol.

zyme with QM, and the remainder of the enzyme with MD (or MC). The challenge of QM/MM approaches is to properly deal with the interface of these two regions. Purely Quantum Chemical Modelling (QCM) approaches do not have this problem because they only take the active site residues of an enzyme into account. In the so-called quantum chemical cluster (QCC or all-QM) approach used in this thesis [69], the protein environment surrounding the active site is approximated using a polarizable homogeneous medium with a dielectric constant. The QCC approach used in this thesis employs Density Function Theory (DFT) and the Becke, 3-parameter, Lee-Yang-Parr (B3LYP) hybrid density functional method. The B3LYP functional (a function of a function) is used by the majority of studies of enzyme catalysis.

The general approach for the QCC calculations in this thesis was as follows:

1. Build a medium sized model of approximately 100 to 200 atoms.
2. Use a small basis set (31G(d,p)) for geometry optimisation of the model and scan for transition states in gas phase (see Figure 10).
3. Identify stationary points (such as a transition state) along the reaction coordinate.
4. Calculate the single point energies of the stationary points, including dispersion forces (by using a larger basis set, e.g. 6-311+G(2d,2p)), the zero-point energy (by using frequency calculations and the solvent correction) and by using the solvation model based on density (SMD) with a dielectric constant of = 4.

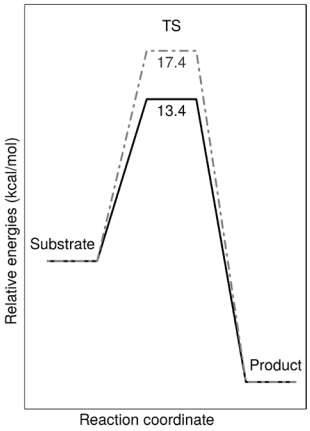


Figure 10: Transition States (TS) and Energy Landscapes. When a substrate is converted to a product by an enzyme, an energy barrier needs to be crossed. The height of the energy barrier determines the speed (rate) of the reaction. A rate of 1 (or 1000) substrate conversion(s) per second corresponds to an energy barrier of approximately 13.4 (or 17.4) kcal/mol. It is important to note that the representation above is only a two dimensional representation of a three dimensional energy landscape. Different conformations of the enzyme with substrate or product bound can have the same energy and are the reason the energy landscape is three-dimensional. Conformations need to be carefully analysed in addition to the calculated energies.

5. Calculate the final energy of each stationary point (energies of the large basis set corrected for the zero-point energy and solvation).

6. Express the final energies of the stationary points relative to the starting structure.

7. Build a larger model once the free energy landscape of the reaction coordinate of interest has been explored with the medium sized model.

There are two main arguments for first exploring the free energy landscape of a reaction with a medium sized model before starting calculations with a big model. Firstly, and most practically, the calculations are faster with a medium sized model. As a rough estimate, the following applies: if a calculation with 100 atoms takes 2 h, a calculation with 200 (or 300) atoms will not take twice (or three times) as long but approximately N^3 times as long. This is because the B3LYP functional does not scale linearly, but to the power of three. Secondly, the free energy landscape becomes more complex with increasing model size: the risk of exploring the "wrong" part of the free energy landscape increases. Therefore, it becomes more difficult to find the global minimum or maximum, and to distinguish these two points from local minima or maxima.

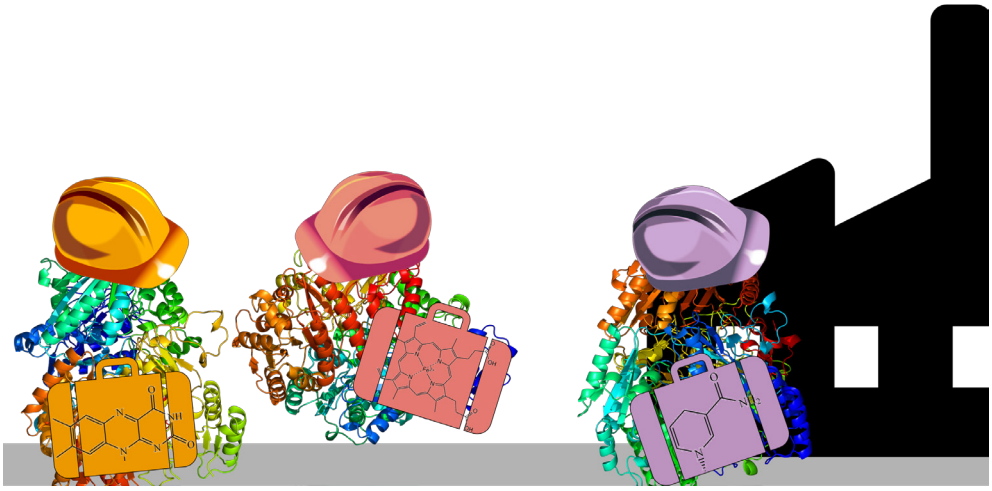
Thesis outline

In this thesis, four different approaches are used to shed light on VAO from different perspectives. These approaches were chosen to answer the following major questions: can expression in a eukaryotic host (instead of a bacterial one) improve the expression of VAO? How many VAO homologs exist in the current databases and what properties do they have? Where do VAOs come from? How do substrates enter the active site, and how do products exit it? Why is there an adduct formed for the *p*-creosol to vanillin reaction in VAO?

There are generous margins on each page to enable the reader to add own comments and thoughts.

Chapter 1 has introduced basic knowledge of the techniques used in this thesis. **Chapter 2** sets the stage to provide the reader with more background on oxidative enzymes and their properties and applications. In **Chapter 3**, we have explored the origins of VAO using a bioinformatic approach. In **Chapter 4** we attempted to observe the effect of sequence variation on function of VAO-like enzymes. This chapter also illustrates one of the pitfalls that can be encountered when attempting this type of work. In **Chapter 5**, a MC approach was used to simulate the migration of ligands in and out of VAO. In **Chapter 6**, an all-QM approach was used to model the reaction mechanism of VAO. **Chapter 7** provides a general discussion of the results obtained in this thesis. The existing, original articles that form the basis of all these chapters are referenced at the beginning of each chapter.

Chapter 2:
Oxizymes for Biotechnology



This chapter is based on:
G. Gygli and W. J. H. van Berkel. *Curr. Biotechnol.*, 4(2):100-110, 2015.

Keywords: Dioxygenase, enantioselectivity, industrial biocatalysis, oxidase, monooxygenase, peroxygenase, peroxidase, protein engineering

Abstract

Oxidation reactions with dioxygen and hydrogen peroxides are difficult to control in the artificial environment of man-made chemistry. This makes oxizymes, i.e. oxidative enzymes that use dioxygen or hydrogen peroxide as co-substrates, very valuable targets for the chemical and pharmaceutical industries. Additionally, growing awareness of sustainability issues in society has encouraged the use of oxizymes in these industries.

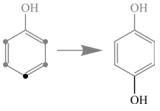
Some oxizymes generate hydrogen peroxide as a by-product. Hydrogen peroxide has antimicrobial effects and is therefore of interest for the food industries as well as other industries where microorganisms pose a danger to the consumer or patient. Hydrogen peroxide can also be used as a bleaching agent and thus applications of oxizymes in the textile industries as well as the pulp and paper industries are common practice.

Oxizymes have recently been found to play an important role as auxiliary enzymes in the degradation of biomass. In this role, they support carbohydrate active enzymes in the degradation of cellulose and chitin, or assist in the deconstruction of lignin-derived polymers. They are therefore of importance for the biofuels industry, which aims to create biofuels from renewable plant materials to replace petroleum-based fuels.

Oxizymes have many more properties that make them useful for industrial applications. This review summarizes the technological advancements, which have made the use of enzymes in industry possible, as well as showcases different types of oxizymes currently used in different industries. Also, the challenges oxizymes face before their industrial applications can be fully developed, are discussed.

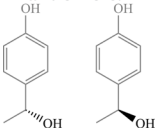
Challenges in bulk chemical oxidation catalysis: it is desirable to use dioxygen directly, e.g. from air. But because dioxygen in air is a *di*radical triplet, it has two unpaired electrons that can react in a radical like fashion (marked in black in the chemical structure below).

Radical reactions are notoriously difficult to control because they proceed via propagation and are difficult to terminate. Additionally, chemoselective oxidations can usually only be achieved by using protective groups, which need to be removed again later. Enzymes are chemoselective by nature, meaning they selectively oxidise only one functional group:



Enzymes can also be stereoselective, meaning they produce stereoisomers. There are two types of stereoisomers, enantiomers (mirror images of each other, just like your hands) and diastereomers.

Enantiomers:



Diastereomers:

Classification of Oxizymes

In the history of life, first the production of dioxygen and then the consumption of dioxygen to gain energy have been key evolutionary events. While there are specialised organisms that are able to grow without dioxygen or for which dioxygen is toxic, the majority of organisms produce or use dioxygen to harvest energy. There are remarkable and highly specialised enzyme-cascades devoted to this process. However, there are also many oxizymes that catalyse reactions not involved in energy harvesting. These reactions are highly specific and diverse, as are the enzymes that catalyse them. Oxizymes are classified by the types of reactions they catalyse, as well as by their cofactors.

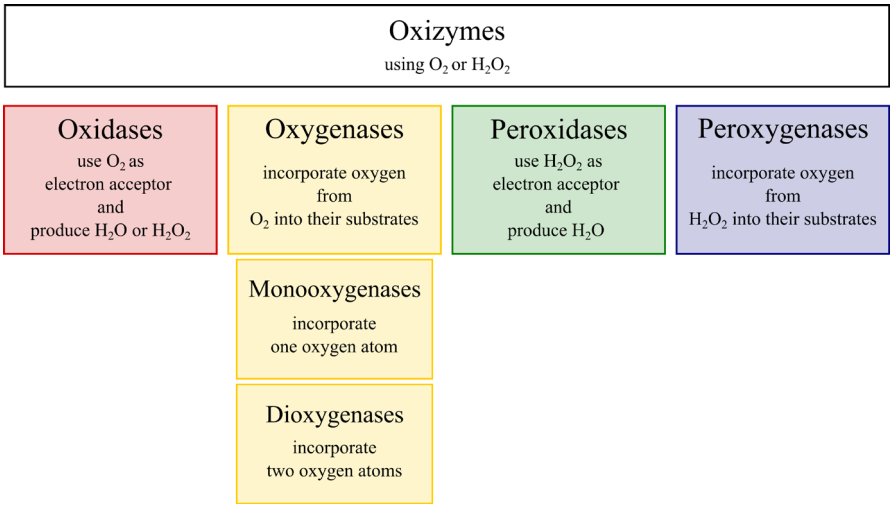


Figure 11: Schematic overview illustrating the properties of different oxizymes. The enzymes are grouped based on whether they react with dioxygen or hydrogen peroxide and whether they incorporate oxygen atoms into their substrates or not.

Oxizymes belong to the big enzyme class of oxidoreductases (EC 1.x.x.x). For the purpose of this review, we define oxizymes as enzymes using dioxygen or hydrogen peroxide (see Figure 11). The enzyme classes discussed here will be oxidases (EC 1.x.3.x), peroxidases (EC 1.11.1.x), peroxygenases (EC 1.11.2.1), monooxyge-

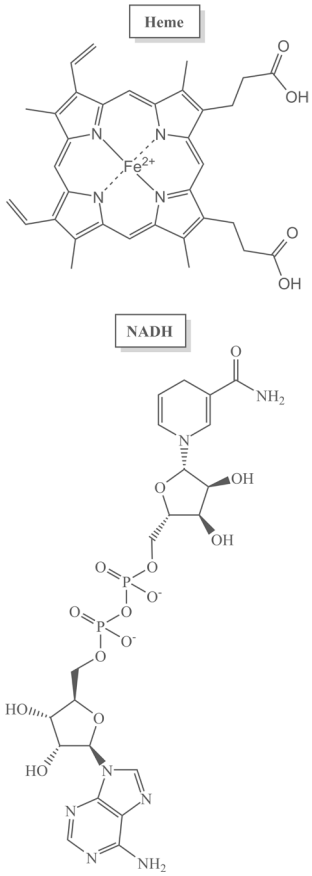
nases (EC 1.13.12.x, 1.14.13- 18.x) and dioxygenases (EC 1.13.11.x, 1.14.11.x, 1.14.12.x).

These enzymes can carry one or more different cofactors, e.g. flavin, heme, nicotinamide, or metal ions. One enzyme can carry several and different cofactors at the same time. A tightly bound cofactor is called a prosthetic group, whereas a dissociable cofactor is called a coenzyme. The latter needs to be regenerated in large-scale applications because the stochiometric use is too expensive [70, 71].

Oxidases catalyse the transfer of electrons from their substrate or the prosthetic group to dioxygen. No oxygen atom (except from water) is incorporated into the substrate, and dioxygen generally is reduced to hydrogen peroxide. Oxidases can contain iron, copper or flavin as cofactors [72]. Examples of the industrial application of oxidases are the use of hexose oxidase in the food industries and the use of monoamine oxidase to create chiral amines, which will be discussed in more detail.

Monooxygenases catalyse the incorporation of one oxygen atom into their substrate while the other oxygen atom is reduced to yield water. Monooxygenases can catalyze a plethora of reactions ranging from hydroxylations and epoxidations to dehydrogenations and cyclisations, and are classified according to the cofactor they carry. There are single center monooxygenases containing iron (e.g. cytochrome P450 monooxygenases), copper (e.g. lytic polysaccharide monooxygenases), or flavin (e.g. Baeyer- Villiger monooxygenases) as cofactors, multi-center monooxygenases containing several cofactors, as well as cofactor-free monooxygenases [73]. Many monooxygenases use a flavin for dioxygen activation. They can be classified in eight different groups, depending on fold and function [74].

Enzyme Commission numbers (EC numbers) classify enzymes by the chemical reactions they catalyse. Other than oxidoreductases, there are transferases (EC 2.x.x.x), hydrolases (EC 3.x.x.x), lyases (EC 4.x.x.x), isomerases (EC 5.x.x.x) and ligases (EC 6.x.x.x).



Besides flavins, NADH and heme can be cofactors of enzymes.

Dioxygenases catalyse the incorporation of two atoms of oxygen into their substrate. They can contain different types of (metal ion) cofactors. Non-heme iron dioxygenases that are active with aromatic compounds can be organised into two groups. The first group contains a high-spin ferrous site (extradiol enzymes) involved in the activation of dioxygen [75, 76], while the second group contains a high-spin ferric site (intradiol enzymes) that activates substrates [77, 78]. Rieske non-heme iron dioxygenases have an additional 2Fe-2S (Rieske) center in their active sites and catalyse the stereospecific addition of dioxygen to aromatic hydrocarbons [79]. α-Ketoglutarate dependent dioxygenases constitute another important group of non-heme iron dioxygenases [80]. They are active on a wide range of compounds and have high potential for industrial applications [81].

Peroxidases catalyse the transfer of two electrons from their substrate to hydrogen peroxide and water is produced. Some peroxidases can also transfer electrons to organic hydroperoxides. There are metal free and metal containing peroxidases. Metal free peroxidases have metal free cofactors like flavin, while metal containing peroxidases carry metal containing cofactors like heme. Most of the heme enzymes contain a so-called protoheme (ferric protoporphyrin IX) as cofactor in their active site. Other metal containing peroxidases are non-heme peroxidases and contain manganese, vanadium or selenium instead of iron. The reaction catalysed by protoheme peroxidases involves the activation of dioxygen by heme through two highly oxidised iron intermediates called Compound I and II.

Peroxygenases catalyse the incorporation of an oxygen atom from hydrogen peroxide into their substrate. Many peroxygenases also have weak peroxidase activity. A new class of peroxygenases has been discovered in 2004 [82] and more research is ongoing to better understand the

Anaerobic organisms can live without dioxygen, while the term anoxic describes environments without any electron acceptor (e.g. dioxygen, nitrate or sulfate). Hypoxic environments are low in dioxygen. For obligate anaerobic organisms the presence of dioxygen is toxic. Facultative anaerobes are able to survive with or without dioxygen, but obligate anoxic organisms are poisoned in oxic environments. Electron transport chains are highly specialised enzyme-cascades in aerobic organisms that use dioxygen as the final electron acceptor. In anaerobic organisms, electron transport chains use different final electron acceptors, e.g. sulfur. Highly specialised enzyme-cascades in plants harvest light and convert it into chemical energy (i.e. high-energy electrons), by splitting water and creating dioxygen during photosynthesis.

mechanistic properties of these so-called unspecific peroxygenases (UPOs; EC 1.11.2.1). They are known so far to catalyse the oxidation of aromatic and heterocyclic compounds, the epoxidation of alkenes, the hydroxylation of aliphatic compounds and dealkylation reactions. The biggest issues hampering the development of industrial applications of UPOs are the current lack of successful large-scale expression [83] as well as heme inactivation by hydrogen peroxide.

EC 1.11.2.1:
1.: Oxidoreductase
11.: Acting on peroxide as acceptor
2.: With H_2O_2 as acceptor, one oxygen atom of which is incorporated into the product
1. With NAD(+) or NADP(+) as acceptor

The ExplorEnz database (`http://www.enzyme-database.org/`) contains several hundred different oxizymes. Out of them, a selection was made for distinct enzymes, which are used in industry or are candidates for future industrial applications.

Industrial Interest in Oxizymes

Historically, glucose oxidase, for use by the baking industry to improve flour and dough properties, was one of the first oxizymes to attract industrial interest. Since then, many other oxizymes have been found to be suitable for industrial applications.

Just as we have experimental toolboxes for biotechnology, we need computational toolboxes for biotechnology!

The development of genetic engineering, toolboxes for protein expression and high throughput screening have made it possible to produce sufficient quantities of an enzyme with desired properties in a convenient host organism on an industrial scale. In parallel, technologies for genome and metagenome sequencing have been developed. This has opened up a completely new perspective on the natural diversity of enzymes and the potential of their unknown characteristics. In addition to these developments, considerable efforts have been made and are still under way to solve elusive sequence-structure-function relationships of proteins in general and of enzymes more specifically. An overview of the technolo-

gies developed for protein engineering to advance industrial biocatalysis has been presented by Bornscheuer and coworkers [84]. The technologies described there can also be applied to enzymes used in other industries beyond the chemical and pharmaceutical industries. A more detailed review on engineered enzymes for chemical production has been composed by Luetz and colleagues [85].

While there are several areas where oxizymes are applied, the main industries involved are the chemical and pharmaceutical industries, the food and food additives industries, the textile, pulp and paper industries as well as the biorefinery and biofuels industries. These industries are interested in oxizymes for completely different reasons. The chemical and pharmaceutical industries are mostly interested in oxizymes due to their formidable ability to selectively oxidise their substrates. Other industries are more interested in oxizymes for their ability to produce hydrogen peroxide or replace the effect of chemicals on a final product.

Chemical and Pharmaceutical Industries

Oxizymes are of exceptional interest for the chemical and pharmaceutical industry because of their sophisticated properties and as candidates for applications in the synthesis of building blocks of chiral and non-chiral drugs and chemicals. The chemical introduction of oxygen atoms into organic substrates remains difficult due to, amongst other reasons, the use of strong oxidants incompatible with the substrate, and unwanted side products formed through chemical side reactions. The enantioselectivity of many oxizymes is an alternative for the difficult chemical synthesis of chiral molecules. In other words, oxizymes are highly selective in the oxidation of their substrates, which chemical oxidations are not. Most of man-

made oxidations are very complex, use non-renewable resources or use and create toxic by-products and/or products. The chemical and pharmaceutical industries have coined the term "industrial biocatalysis" in the last 20 years and the field has exploded since. Research in industrial biocatalysis has gained a considerable amount of momentum around the turn of the millennium. Several reviews have been published at that time predicting this development [86, 87, 88, 89]. The motivation behind this accelerated and boosted investment in industrial and academic research in industrial biocatalysis, and more specifically the industrial applications of oxizymes, is highly diverse. From the industries' viewpoint, the use of biocatalysts may simplify the complicated chemical synthesis of oxygen containing compounds. In academic research, the interests are more fundamental. The discovery of novel enzymes and catalytic mechanisms as well as new reactions, which are impossible to achieve by chemical means, may be of greater interest. For society, the main driving force behind the funding of research into biocatalytic processes appears to be a desire for sustainable processes. These very different motivations and targets of the three stakeholders are not easily met.

The desire for more sustainable processes has lead to the formulation of the 24 principles of green chemistry and green engineering. These principles can be summarised with two acronyms, "PRODUCTIVELY" and "IMPROVEMENTS" [90, 91]. The evaluation of the "greenness" of a process is not always straightforward and the use of an enzyme does not necessarily make a process green [92]. Ecological and economical assessments of biocatalytic processes are mostly performed in industry and are not published. A few rare examples exist for academic research into that field [93, 94]. However, the development of industrial biocatalysts would most likely benefit from more such analyses already at an earlier stage of research

PRODUCTIVELY:
Prevent waste
Renewable materials
Omit derivatisation steps
Degradable chemical products
Use safe synthetic methods
Catalytic reagents
Temperature, pressure ambient
In-Process Monitoring
Very few auxiliary substances
E-factor, maximise feed in product
Low toxicity of chemical products
Yes it's safe

IMPROVEMENTS:
Inherently non-hazardous and safe
Minimise material diversity
Prevention instead of treatment
Renewable material and energy inputs
Output-led design
Very simple
Efficient use of mass, energy, space & time
Meet the need
Easy to separate by design
Networks for exchange of local mass & energy
Test the life cycle of the design
Sustainability throughout product life cycle

to avoid "running into dead-ends" and to accelerate the development and application of usable industrial biocatalysts.

A well-known example of an industrial oxidative biocatalyst is D-amino oxidase (DAO; EC 1.4.3.3). DAO's most relevant industrial application is the production of 7-aminocephalosporanic acid from cephalosporin C. 7-Aminocephalosporanic acid is a key raw material for the production of many cephalosporin antibiotics [95]. Other applications of DAO include the production of pure L-amino acids and the detection and analytical determination of D- amino acids and analogues [96].

EC 1.4.3.3:
1.: Oxidoreductase
4.: Acting on the CH-NH_2 group of donors
3.: With oxygen as acceptor

Another example of a "pharmaceutical" industrial oxidative biocatalyst is monoamine oxidase. This enzyme, originally isolated from human liver mitochondria [97], is widely applied to help in the chemo-enzymatic deracemisation of amines. This enzyme is briefly discussed below.

Monoamine oxidase from *Aspergillus niger* (MAO-N; EC 1.4.3.4) is a flavoenzyme. The first publications of MAO-N date from 1995. MAO-N lacks a cysteine, which is responsible for covalent linkage of the flavin cofactor in its mammalian homologues, MAO-A and MAO-B. While human MAO-B is a drug target for antidepressants and neuroprotective drugs, the human MAO-A gene has been popularised and nicknamed "warrior gene" [99]. MAO-N has attracted attention for a different reason: it can be used to produce chiral amines. MAO-N has been evolved to produce enantiomerically pure cyclic tertiary amines [100, 101]. The crystal structure of MAO-N was published in 2008 and the structure-function relationships of the mutants obtained by directed evolution were analysed [102].

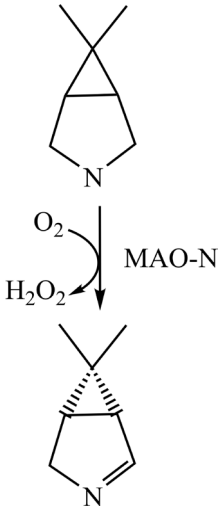


Figure 12: MAO-N catalysed reaction in the process for the production of a boceprevir intermediate developed by Merck and Codexis. 6,6-Dimethyl-3-azabicyclo[3.1.0]hexane is converted into a secondary amine. After Li *et al.* [98]

Merck and Codexis recently patented and published a process to produce a secondary amine as an intermediate of boceprevir using an optimised version of MAO-N

[98, 103] (Figure 12). They also analysed the variant enzymes they obtained with the help of the crystal structure (Figure 13). As is illustrated by these examples, MAO-N is a valuable tool to develop enantiomerically pure amines, which are of high value for organic synthesis or as intermediates for the synthesis of pharmaceuticals.

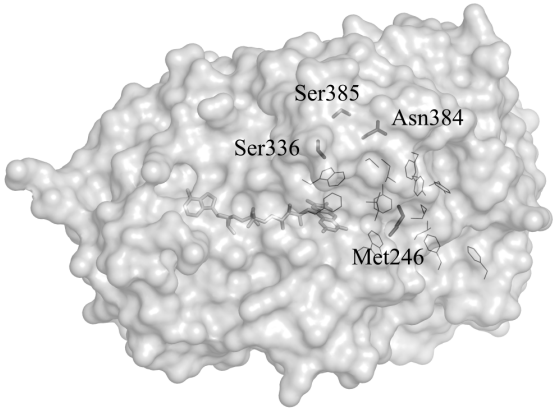


Figure 13: Solvent accessible surface representation of the MAO-N structure with its non-covalently bound flavin cofactor (pdb code 2VVM). Residues proposed to form a hydrophobic cavity from the surface to the substrate-binding site are shown as fine lines. Close to this hydrophobic cavity are mutations found to enable the enzyme to catalyse the oxidation of secondary amines (N336S and A385S) and two additional mutations which enable the quadruple mutant to catalyse the oxidation of tertiary amines (T384N and I246M) [102]. The figure was created using PyMOL, after [102].

Other Oxizymes, which are in the Early Stages of Research into their Industrial Applicability

Baeyer-Villiger monooxygenases (BVMOs; EC 1.14.13.x) are flavin-dependent monooxygenases that catalyse the oxidation of carbonyl compounds into esters or lactones. BVMOs could be especially useful for the production of enantiopure chemicals [104]. Baeyer-Villiger monooxygenases have attracted considerable academic interest for their envisaged industrial applications. The increased number of BVMOs available for biochemical studies as well as the elucidation of 3D structures has helped to illustrate and rationally explain the ability of BVMOs to catalyse reactions with many different substrates. Despite their ability to catalyse highly specific oxygenation

EC 1.14.13.:
1.: Oxidoreductase
14.: Acting on paired donors with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O_2
13.: With NADH or NADPH as one donor, and incorporation of one atom of oxygen

reactions, the industrial use of BVMOs still faces challenges. Some of these challenges are the development of cost-effective technical approaches, enzyme stability, improved efficiency of the enzyme and organic solvent tolerance [105, 106, 107]. Technologies and tools used to address and overcome the obstacles BVMOs face for their industrial application are, among others, genome mining [109, 110] and directed evolution [111, 112]. The issue of cofactor regeneration has been addressed by fusion engineering, creating a BVMO fused to a NADPH-regenerating phosphite dehydrogenase [113]. Scaling up of processes using BVMOs has been performed using different approaches [114]. Several patent applications filed by Codexis and other companies illustrate that these challenges can be overcome, e.g. [115, 116].

A recent publication has identified fungal BVMOs which appear to be less inhibited by higher substrate loadings (up to 30mM) [108].

Cytochrome P450 monooxygenases (E.C. 1.14.14.x) have previously received much attention from the pharmaceutical industries for their key role in drug metabolism. The cytochrome P450 superfamily is enormous and growing, which makes it difficult to keep nomenclature uniform. The cytochrome P450 homepage aims at establishing a uniform and evolutionary accurate nomenclature system [117]. In August 2013, a total of 21,039 cytochrome P450 sequences were included on the website. An overview of some of the thousands of reactions these enzymes catalyse was assembled by Bernhardt and Urlacher [118]. Cytochrome P450s are heme dependent enzymes and have recently attracted the spotlight for their ability to catalyse the regio- and stereospecific oxidation of non-activated hydrocarbons. The focus on these and other properties of cytochrome P450 monooxygenases has led to engineering efforts using rational design and other techniques to allow an industrial application of a thus tailored enzyme [119]. An overview of cytochrome P450 monooxygenases with perspectives for their synthetic application has been composed by Urlacher and coworkers [118, 120]. Exam-

EC 1.14.14.x:
1.: Oxidoreductase
14.: Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O_2
14.: With reduced flavin or flavoprotein as one donor and incorporation of one atom of oxygen

ples of patents and patent applications that are using cytochrome P450 monooxygenases include e.g. the production of erythromycins [121] and 11-beta-hydroxy steroids [122].

In summary, oxizymes employed in the chemical and pharmaceutical industries are used for the regio- and stereoselective oxidations they catalyse.

Biosensors and Analytical Chemistry

Another well established area of application for oxizymes is their use as biosensors and in analytical chemistry. Glucose oxidase (GOX; EC 1.1.3.4) is used as a biosensor to monitor glucose levels in the blood of diabetic patients. Lactate oxidase (EC 1.1.3.2) is researched for applications as a biosensor reporting L-lactate concentrations in blood and sweat of athletes. A patent application for a lactate sensor using a lactate oxidase has been filed in 2013 by the Japanese company Arkray Inc [123]. Another oxizyme, which already has a widespread medical application, is horseradish peroxidase (EC 1.11.1.7). Antibody-horseradish peroxidase conjugates are widely used in the immunohistochemical detection of antigens in tissues to allow specific staining of the tissues targeted by the antibody.

EC 1.1.3.4:
1.: Oxidoreductase
1.: Acting on the CH-OH group of donors
3.: With oxygen as acceptor

EC 1.11.1.7:
1.: Oxidoreductase
11.: Acting on a peroxide as acceptor
1.: Peroxidases

Polyaniline is a conductive polymer used in sensors and electrochromic devices. It changes colour depending on its oxidation state and is one of the most studied conductive polymers of the last 50 years. Laccase (EC 1.10.3.2) has been used to synthesise polyaniline, more precisely to catalyse the polymerisation of aniline [124].

EC 1.10.3.2:
1.: Oxidoreductase
10.: Acting on diphenols or related substances as donors
3.: With oxygen as acceptor

Food and Food Additives Industries

The food and food additives industries use oxizymes like GOX to prolong the shelf life of foods by removing residual glucose and dioxygen. In the same reaction GOX also produces hydrogen peroxide, a good bactericide, which can later be removed by a catalase. Depending on the desired effect of the enzyme on the foodstuffs, different demands are made on the enzyme. For antimicrobial effects, the production of hydrogen peroxide by oxidases is key. In this case, substrate specificity should be broad to allow production of sufficient amounts of hydrogen peroxide. For the more targeted degradation of specific chemicals in foodstuffs, narrow substrate specificity is preferable.

In any case, processing stability of the enzyme is of importance because the processing conditions of foodstuffs are defined by food safety regulations. Also, the non-toxicity of the enzyme in the final product needs to be guaranteed because the complete removal of the enzyme from the final product is often not possible or wanted. The non-toxicity of an enzyme can be assessed through toxicology studies and a GRAS status can be awarded to the enzyme itself and the production organism used. Although the enzyme and its products may be non-toxic, the use of enzymes from pathogenic or toxic organisms is not desired.

Generally **R**ecognised **A**s **S**afe: is a designation by the American Food and Drug Administration (FDA) used to label food additives or organisms involved in the food production process (e.g.the yeast *Saccharomyces cerevisiae* that is involved in the production of bread and beer).

Hexose oxidase (HOX; EC 1.1.3.5) from *Chondrus crispus*, is produced by Danisco/DuPont. HOX is a flavoprotein that binds its flavin cofactor covalently and belongs to the VAO/PCMH family. Discovered in 1973, HOX was found to have much broader substrate specificity than GOX. GOX was the first bio-replacement of chemical bread improvers like potassium bromate, prompted by consumers desire for “clean label” products [125]. GOX binds its flavin cofactor non-covalently and belongs to the

glucose-methanol-choline (GMC) family. GOX is used to catalyse the conversion of glucose and dioxygen into gluconolactone and hydrogen peroxide. The hydrogen peroxide formed then acts by oxidizing the thiol groups in the gluten proteins forming intra- and intermolecular disulfide bonds, which leads to increased dough strength [125]. HOX was seen as a more suitable enzyme for that task due to its broader substrate specificity. HOX catalyses the conversion of several mono- and oligosaccharides into the corresponding lactones, with concomitant formation of hydrogen peroxide (Figure 14) [125].

Disulfide bonds can be crucial for enzyme stability. Two cysteine residues are required to form a disulfide bond. Intramolecular disulfide bonds are formed within one chain (subunit) of the protein while intermolecular ones are formed between different chains (subunits) of the protein.

In 1995, a method of using HOX as dough improving agent was patented [126]. Initially, the production of sufficient amounts of HOX protein was limiting [127, 128]. HOX is now heterologously expressed in *Hansenula polymorpha* on a commercial scale, after toxicology studies were performed to confirm the safety of the procedure [129]. HOX is currently used in POWERBake®, a product which is marketed for improving the baking process.

Besides its properties useful for application in the food industries, HOX was also found to have antifouling properties through the production of hydrogen peroxide. These properties were patented as early as 1999 [130], but only published in 2010 [131]. The most recent published patent on HOX has a priority date from 2000, and uses HOX in a process to prevent the Maillard reaction in foodstuffs [132]. Products coming from these patents are not on the market yet.

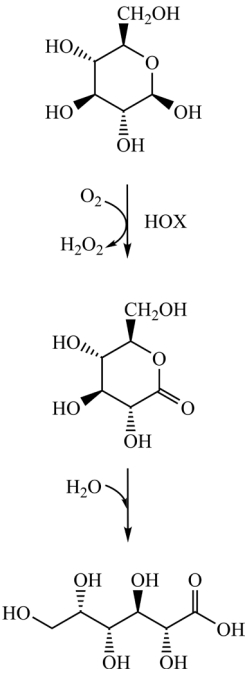


Figure 14: HOX catalysed oxidation of glucose to gluconolactone with the concomitant reduction of dioxygen to hydrogen peroxide and the subsequent non-enzymatic hydrolysis of gluconolactone to gluconic acid. Unlike GOX, HOX catalyses this type of reaction for a variety of hexose sugars.

The HOX example illustrates the disparity in focus between academic research and industrial research. While academia likes to emphasize an enzyme's potential for industrial applications, this is contrary to the behaviour of industry and wording of scientific publications from industry, where the mention of industrial applications is secondary and naturally understood. Industry first re-

quires not only a patent, but also a strategy for the use of a patent or several patents to protect their efforts to develop the enzyme and provide a viable and economically sustainable path for the long-term use of the enzyme. This includes using the same enzyme for as many different and diverse applications as possible.

A peroxidase (MsP1, EC 1.11.1.19) from *Marasmius scorodonius* is produced through expression in *Aspergillus niger* by DSM. It is used to degrade beta-carotene in whey, thus whitening the whey. This property was patented in 2004. Subsequent scientific publications described the enzyme in more detail [133, 134, 135]. The peroxidase is present in the product "Maxibright", which is protected by a trademark from 2009. MsP1 was declared GRAS in 2012, after an application for this status by DSM [136]. MsP1 has been mentioned for its potential for industrial application in second generation biofuels in 2009 [135], due to its induction by lignin and possible contribution to the modification of lignified biopolymers.

Patent A government authority or licence conferring a right or title for a set period, especially the sole right to exclude others from making, using, or selling an invention. A prioritity date is the date of the filing of the first patent covering an application. Subsequent applications (and patents) can be related to each other through identical priority dates. A patent family can contain patents filed in different countries. Human genes, animals and plants cannot be patented because they are a product of nature.
Trade mark A symbol, word, or words legally registered ® or established by use ™ as representing a company or product.

Textile Industries and the Pulp and Paper Industries

The textile industries and the pulp and paper industries use oxizymes like peroxidases and laccases for dye transfer inhibition as well as dye degradation. A broad substrate specificity of the enzyme for several dyes and dye classes as well as stability under processing conditions constitute desired properties. Specificity of the enzyme for one dye only makes sense if that dye is used as the major compound in the targeted process.

Peroxidase from *Coprinus cinereus* (CiP, EC 1.11.1.7) is a heme enzyme produced by Novozymes through expression in *Aspergillus oryzae* or *Aspergillus niger*. The reaction catalysed by CiP involves the activation of dioxygen

by heme through two highly oxidised iron intermediates called Compound I and Compound II. Compound I is oxidised by two electron equivalents and compound II is oxidised by one electron equivalent. Compound II is therefore the reduced form of Compound I. The reaction scheme below illustrates how CiP oxidises dyes (from [137]).

$$\text{CiP} + \text{H}_2\text{O}_2 \longrightarrow \text{CpPI} + \text{H}_2\text{O} \quad (2.1)$$

$$\text{CpPI} + \text{dye}_{\text{red}} \longrightarrow \text{CpPII} + \text{dye}_{\text{ox}} \quad (2.2)$$

$$\text{CpPII} + \text{dye}_{\text{red}} \longrightarrow \text{CiP} + \text{dye}_{\text{ox}} + \text{H}_2\text{O} \quad (2.3)$$

CiP was first used in combination with a laccase in the dye transfer inhibition (DTI) technology. The first publication of the enzyme dates from 1988, and the first patent has a priority date of 1989 [138, 139]. CiP was patented for use in the DTI technology, however all these patents appear to have lapsed and the trademark for "Guardzyme", a CiP containing DTI product, has lapsed as well. "Guardzyme" is aimed to work at low temperatures to oxidise and therefore decolourise dyes that dissolve into the washing mix.

CiP has been described in several scientific publications, at least four years after patenting [137, 140, 141, 142, 143, 144, 145]. CiP was considered the first member of a novel family of secretory fungal peroxidases. Compared to horseradish peroxidase it has broader substrate specificity. The issue of its low thermostability (irreversible unfolding after 1 min at 67 °C) [144] could be addressed by directed evolution, which resulted in a 100-fold increase in thermostability [137]. The first 3D structure of CiP became available in 1994 [146], which lead to the rational engineering of a thermostable enzyme in 2010 [147]. Interestingly, the mutations found by the random and rational approaches to increase thermostability are not the same, see Figure 15 for an illustration.

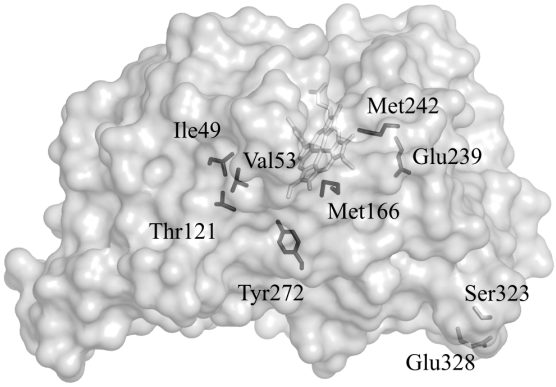


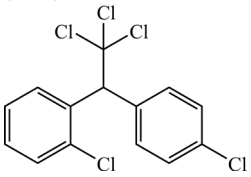
Figure 15: Solvent accessible surface representation of the CiP structure with its bound heme cofactor (pdb code 1H3J). The positions of mutations found to increase thermostability by [137] (wild type residues Ile49, Val53, Thr121, Met166, Glu239, Met242, Tyr272), located close to the heme) are shown in black and [147] (wild type residues Ser323 and Glu328, located at the protein surface) are shown in grey. While the mutations close to the heme group were identified through directed evolution experiments creating a library of thousands of mutants, the residues at the protein surface were identified using an in silico approach to screen a knowledge-based library of eight mutants. The figure was created using PyMOL.

A recent patent application uses the CiP peroxidase to modify the colour of textiles [148]. It remains to be seen how the use of CiP in the textile industry develops further. Novozymes also has patented CiP to produce paper or paperboard from mechanical pulp [149]. CiP illustrates the power of protein engineering tools and the willingness of industry to invest in the development of improved enzyme variants.

Other peroxidases that have attracted attention from industry are dye-decolorising peroxidases (EC 1.11.1.19; DyPs). These peroxidases, as is indicated by their name, oxidise dyes and other substrates, effectively decolorising them. DyPs cannot directly react with bulky substrates, but have evolved a long-range electron transfer pathway, which transfers electrons to the heme cofactor. A surface tryptophan has been shown to play an essential role in this mechanism [150, 151]. Other peroxidases that also operate via a long-range electron transfer mechanism are lignin peroxidases (EC 1.11.1.14), cytochrome C peroxidases (EC 1.11.1.5) and versatile peroxidases (EC 1.11.1.16) [152, 153, 154].

Laccases (EC 1.10.3.2) are high redox potential multi-copper oxidases. Multi-copper oxidases vary in their copper binding sites but can be identified based on the conserved amino acids involved in copper binding [155]. A bioinformatics analysis of the domains and domain organisation of multi-copper oxidases has been performed by Nakamura and coworkers [156]. Laccases typically transfer four electrons to dioxygen, which is reduced to water while the substrate is oxidised. They have attracted interest from industry for several reasons. Laccases can oxidise phenolic and non-phenolic lignin-related compounds as well as highly recalcitrant environmental pollutants like dichlorodiphenyltrichloroethane (DDT). Novozymes has patented a process using laccase from *Myceliophthora thermophila* to treat pulp for increased paper wet strength [157]. Several other properties of laccases, which could lead to different applications in the pulp and paper industries were summarised by Virk *et al.* [158].

Dichloro**d**iphenyl**t**richloroethane (DDT):



Laccases can act on bulky substrates like lignin by using chemical mediators as redox shuttles. The mediator can oxidise the lignin polymer, which leads to debranching and degradation, and is in turn reduced again in the enzyme's active site. This action explains why many wood degrading fungi have laccases; they are needed to unlock lignin via the action of their mediators. Different mediators are used to achieve the desired functionality of different laccases, also for synthetic applications [159].

Direct reactions of laccases with phenolic substrates create radical intermediates, which lead to the formation of polymers. A medical application of this property of laccases has been imagined for the coating of catheters. The zwitterionic coating currently used to prevent biofilm formation in the catheter is polydimethylsiloxane. A process to coat catheters using laccases has been described by Blanco and coworkers [160]. Laccases have also been

studied for their application in the creation of bioactive hydrogel dressings [161]. A review on the potential applications of laccases in polymerisation and grafting reactions has been composed by Kudanga and coworkers [162]. Other enzymes that have been explored for their potential to create bio-polymers are tyrosinases (EC 1.14.18.1) and peroxidases [163].

EC 1.14.18.1:
1.: Oxidoreductase
14.: Acting on paired donors with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O_2
18.: With another compound as one donor, and incorporation of one atom of oxygen

Much effort has been directed to the discovery of new (mostly fungal) laccases and the improvement of their temperature and pH stability properties [164]. Improved robustness of laccases is of crucial importance, given the demanding process conditions of the textile and pulp and paper industries.

The use of different protein engineering techniques to increase stability of peroxidases and laccases will certainly lead to an increase in industrial applications of these enzymes.

Biorefinery and Biofuels Industries

In the biorefinery and biofuels industries, the use of cellulolytic and chitinolytic enzymes is well established [165]. These enzymes are classified as carbohydrate-active enzymes. Several oxizymes have recently been classified into a new class of carbohydrate-active enzymes (CAZymes, `http://www.cazy.org/Auxiliary-Activities.html`) with auxiliary activities [166]. Examples of these reclassified enzymes are laccases, gluco-oligosaccharide oxidases, vanillyl alcohol oxidases and lytic polysaccharide monooxygenases. Some of these enzyme groups are currently being investigated in more detail for their impact on the efficiency of enzyme cocktails for second-generation biofuels.

Carbohydrate-**a**ctive en**z**ymes Database (**CAZy**): contains families of structurally-related catalytic and carbohydrate-binding domains of enzymes. It contains glycoside hydrolases, glycosyl transferases, polysaccaride lyases, carbohydrate esterases and redox enzymes with auxiliary activities (AA). Vanillyl alcohol oxidase has been classified in the AA4 family.

Lytic polysaccharide monooxygenases (LPMOs; no EC number assigned yet) are monooxygenases that carry a single metal ion, mostly copper, as a cofactor. Their addition to enzyme cocktails for the creation of second-generation biofuels boosts enzymatic cellulose conversion. They have been grouped into three families of CAZymes with auxiliary activities, AA9, AA10 and AA11 [166, 169, 170]. AA9 and AA10 LPMOs have been mainly identified in fungal or bacterial genomes, respectively. The AA9 enzymes have been proposed to act directly on cellulose and thus make the cellulose more accessible to other CAZymes [171]. AA10 LPMOs are able to cleave crystalline cellulose, a property that cellulases do not possess. To elucidate how these LPMOs catalyse this reaction, several studies have been performed, using different computational and experimental techniques [172, 173].

LPMOs have recently been found to be hydrogen peroxide dependent, and not dioxygen dependent, as previously thought [167]. These results are highly valuable and likely to encourage further, detailed investigation of the reaction pathways of LPMO, similar to a recent study [168].

While the biofuels industries are interested in creating biofuels from renewable plant materials, the biorefinery industries are aiming at the valorisation of these materials not only into biofuels but also into value-adding chemicals, especially from lignin. Efforts are under way to develop new enzymes and technologies to enable these advancements. A review summarising these developments has been recently published [174]. Despite the established saying in the pulp industry that "one can make anything out of lignin except money" [174], there appears to be readiness to invest into more research from academia and industry to further explore the possibilities of lignin valorisation.

World **E**conomic **F**orum

As noted in the WEF report on "The future of industrial biorefineries" from 2010, "*there are some indications that lignin could be used as a value-adding component of slow-release fertilizers and as a starting compound for vanillin fermentation*" [175]. Vanillin is a high-value chemical, which is of great value for the food and food additives industries because of its vanilla flavour [176]. World-

wide demand for vanilla flavour is nowadays mostly covered by chemically synthesised vanillin or vanillin produced from petrochemical raw materials. The production of vanillin from wood would therefore allow covering the consumers demand for natural vanillin from renewable sources. Other vanillin production routes start from ferulic acid (Givaudan, [177]) or could be developed to involve the use of oxizymes such as **vanillyl alcohol oxidase** (VAO; EC 1.1.3.38) [178, 45]. VAO binds its flavin cofactor covalently and is the founding member of the VAO/PCMH family [179, 31]. Oxizymes and the processes using them to produce these types of high-value chemicals would be more connected to the chemical industries and the demands on these enzymes are therefore different from the demands on enzymes for the biofuels industries.

Other oxizymes that are attracting more attention due to their ability to degrade lignin are versatile peroxidases [152], lignin peroxidase [180] and manganese peroxidase (EC 1.11.1.13) [181]. For a more complete overview of enzymes involved in the degradation of (ligno)cellulose, we refer the reader to a review by Bornscheuer and colleagues [182].

Use of Whole-Cells as Biocatalyststs ("Designer Bugs") and Different Industrial Expression Systems

As mentioned previously, industries use oxizymes for different reasons and have different scopes of applications for the enzymes. Depending on the desired product(s) from an enzymatic reaction, it is also possible to use so-called "designer bugs" instead of purified enzymes. These "designer bugs" are modified microorganisms, which have optimised enzyme pathways for the manufacturing of a desired product or entire groups of products. Within

the industries that use oxizymes, there is a range of companies which have focussed more on the use of either "designer bugs" or purified enzymes [87].

There are two main reasons to use a whole-cell approach instead of purified enzyme(s) [183]. Firstly, a whole-cell approach may be more economical compared to the use of purified enzyme(s). This holds true as long as a negative impact of other cell products is not a problem. The use of purified enzyme(s) is not always possible due to stability issues. Especially for membrane-bound enzymes, purification of an active form is challenging [184]. If entire cascades of enzymatic reactions are desired, it also makes more sense to use whole cells. Secondly, in addition to the costs of the purified enzyme, the costs of cofactors are also a good reason to use a whole-cell process. In whole-cell processes the cofactors are regenerated by the cells at low cost [185]. Some examples of the use of whole-cell biocatalysts are the production of (*S*)-styrene oxide using styrene monooxygenase (EC 1.14.14.11) [186] and the preparative scale production of enantiopure compounds by BVMOs [187, 188]. An analysis of the use of whole-cell biocatalysts in the fine chemicals industries has been performed by Straathoff and colleagues [189].

Membrane-bound enzymes (and proteins) are infamous for being extremely difficult to work with *in vitro*. This is mainly due to their strong binding to membranes (making isolation difficult) as well as their membrane spanning or membrane associated domains. Because the insides of membranes are hydrophobic environments, membrane spanning and membrane associated domains contain hydrophobic residues that interact with the membrane. If these residues are exposed to hydrophilic solvents, e.g. water, the protein structure can become highly unstable.

The expression of oxizymes on an industrial scale remains challenging because many oxizymes are fungal enzymes. The expression of fungal enzymes is often difficult to achieve in bacterial expression systems like *Escherichia coli*. Alternative eukaryotic expression systems are yeast systems, *Komagataella pastoris* (formerly known as *Pichia pastoris*), *Hansenula polymorpha* or different fungal systems, mostly *Aspergillus strains* like *Aspergillus niger* or *Aspergillus oryzae*. Depending on the company and patent situation, customised expression platforms are developed to allow large-scale expression of the desired enzymes.

Komagataella pastoris has been renamed to *Komagataella phaffii*, [190], a fact which was missed in the original publication of this article.

In conclusion, the examples mentioned above illustrate the different strategies used by different industries to develop oxizymes for industrial applications. While some industries rely on a naturally occurring version of the enzyme, other industries invest considerable efforts to tailor the enzyme for a specific application.

Challenges Oxizymes Face in their Industrial Applicability

Chemical and Pharmaceutical Industries: an "Uphill Battle Against Entrenched Chemical Processes" and High Demands There remains a significant gap between the potential and predicted applications of oxizymes and their actual applications. While academic research is more and more focussing on potential applications of oxizymes, and although these potential applications are emphasised repeatedly in publications, there is a notable discrepancy in the numbers of these predicted applications and the actual industrial applications one can find. This may in part be due to the limited public access to information about the details of industrial processes.

Other reasons for this discrepancy are diverse and depend on many factors, however a few common denominators can be found. Many industries are open to the use of enzymes and actively researching the potential of enzymes for their processes. Nevertheless, the cost of adapting an existing process to be compatible with the use of an enzyme must not be underestimated. Another important factor is the cost of the enzyme. The production of enzymes is becoming cheaper and immobilisation by stabilising the enzyme can allow several recycling steps of the enzyme. However, the level of research and development needed to identify a suitable enzyme through genome mining or to tailor an existing enzyme to an application

is substantial. Depending on the industry, this investment can be supported and justified by the high value product that is generated in the process. As a consequence, the investment into research and development for enzymes that cannot be used for the creation of high value products cannot be supported [88]. The arguments mentioned do not only hold true for the replacement of chemical processes by enzymatic processes, but also for the replacement of an enzyme in an existing enzymatic process. Once an enzyme is used in a process, there is low motivation to find a more suitable enzyme or to adapt the used enzyme further. Like in evolution, the creation of perfection is not economically feasible.

Yet it seems that industry wants perfect enzymes for their processes. "*A candidate biocatalyst for an industrial process must be suited to the conditions required for the process such as high substrate and product tolerance, resistance to pH, temperature and constituents of the reaction matrix. It must also possess high productivity and selectivity, which maximises production of the desired product and minimises formation of side-products. However, these characteristics alone are not sufficient for a successful scale-up of the process. The downstream processing also needs to minimise losses, remove side-products and still be environmentally tenable.*" [87].

The above criteria only make an enzyme a candidate biocatalyst. The efforts that are most likely required to incorporate the candidate biocatalyst into a modified version of an existing process are the next hurdle. These high demands of industry on a biocatalyst and the impediments that almost inevitably follow during the development of biocatalytic industrial-scale processes are difficult to fulfil and overcome. Academic research can be of help for the development of industrial enzymes, especially in those cases where industry shares its goals for the enzyme with academia.

Different research strategies should therefore be adapted depending on the enzyme and the envisaged applications of the enzyme. In light of the fact that also for enzymes, "*it takes 20 years to become an overnight success*" [86], academia should not be expected to provide ready-made enzymes for "potential industrial applications". It is more desirable to allow academia to explore the unknown biotransformation territories while using focussed industry-academia collaborations to identify novel enzymes wanted by industry and improve enzymes where industry requires them [86].

Other Industries The food and food additives industries, the textile industries and the pulp and paper industries as well as the biorefinery and biofuels industries do have different demands for oxizymes than the chemical and pharmaceutical industries. While specificity is desired by the chemical and pharmaceutical industries, the opposite seems to hold true for most applications in these other industries. The main focus is on oxizymes that have broad substrate specificity. This enables them to have maximal impact on the product. In the food and food additives industries, the production of hydrogen peroxide is desired to conserve foodstuffs. In the textile industries and the pulp and paper industries, the decolourisation of different types of dye is wanted. And in the biorefinery and biofuels industries, the auxiliary activities of the enzyme should be as broad as possible to achieve maximal positive impact on the performance of the enzyme cocktails used to create biofuels. The specificity of the enzyme for only one substrate is only desired in rare cases. Stability under processing conditions is of highest importance and can nowadays be addressed by protein engineering techniques.

Conclusions and Outlook

The industrial use of oxizymes is increasing since the turn of the millennium. However, the pace at which oxizymes are incorporated into industrial applications is slower than expected. Reasons for this slower pace vary from industry to industry, but the most likely common denominator is the economical cost of research and development needed to identify and customise an enzyme suitable for an industrial process as well as the design of such a process itself. The use of computational tools to minimise screening efforts may contribute to lowering the cost of enzyme customisation. Another issue obstructing the development of industrial (oxidative) enzymes is miscommunication between academic research and industrial demands. There are a number of international projects built on and encouraging collaborations between academia and different industries, some ongoing or about to start and some already finished, e.g. PEROXYCATS, OXYGREEN, BIOTRAINS, BIONEXGEN, INDOX, OXEPI, OXYPOL and ROBOX. Such targeted collaborations may help to alleviate the confusion of students and young researchers in the field of industrial oxizymes when faced with the huge number of potential industrial applications of these enzymes compared to the numbers of actual applications.

A recent review by participants of the INDOX project summarises the current state of oxidoreductases in industrial applications [191].

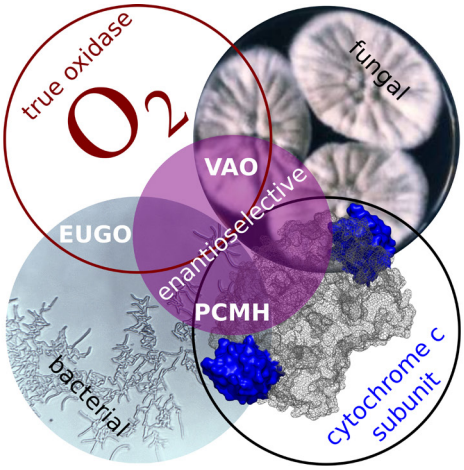
Undoubtedly, the industrial use of oxizymes holds great promise. In the interest of reaching sustainability goals set by society, combined efforts to increase targeted collaborations between industry and academia are desirable. Existing collaborations as well as new collaborations require more time to help more oxizymes reach maturity for industrial applications.

Acknowledgements

This work was supported by European Commission Project INDOX-FP7-KBBE-2013-7-613549.

Bibliography follows at the end of the thesis

Chapter 3:
On the origin of vanillyl alcohol oxidases



This chapter is based on:
G. Gygli, R. P. de Vries and W. J. H. van Berkel. *Manuscript under review.*

Keywords: Flavoprotein, fungus, phylogeny, sequence motif, dehydrogenase

Abstract

Vanillyl alcohol oxidase (VAO) is a fungal flavoenzyme that converts a wide range of *para*-substituted phenols. The products of these conversions, e.g. vanillin, coniferyl alcohol and chiral aryl alcohols, are of interest for several industries. VAO is the only known fungal member of the 4-phenol oxidising (4PO) subgroup of the VAO/PCMH flavoprotein family. While the enzyme has been biochemically characterised in great detail, little is known about its physiological role and distribution in fungi.

We have identified and analysed novel, fungal candidate VAOs and found them to be mostly present in Pezizomycotina and Agaricomycotina. The VAOs group into three clades, of which two clades do not have any characterised member. Interestingly, bacterial relatives of VAO do not form a single outgroup, but rather split up into two separate clades.

We have analysed the distribution of candidate VAOs in fungi, as well as their genomic environment. VAOs are present in low frequency in species of varying degrees of relatedness and in regions of low synteny. These findings suggest that fungal VAOs may have originated from bacterial ancestors, obtained by fungi through horizontal gene transfer.

Because the overall conservation of fungal VAOs varies between 60 and 30% sequence identity, we argue for a more reliable functional prediction using critical amino acid residues. We have defined a sequence motif P-x-x-x-x-S-x-G-[RK]-N-x-G-Y-G-[GS] that specifically recognizes 4PO enzymes of the VAO/PCMH family, as well as additional motifs that can help to further narrow down putative functions. We also provide an overview of fingerprint residues that are specific to VAOs.

Introduction

Vanillyl alcohol oxidase (VAO, EC 1.1.3.38) is a covalent flavoenzyme first isolated from the ascomycetous fungus *Penicillium simplicissimum* [31]. VAO is active with a wide range of *para*-substituted phenols [36, 37]. Several VAO reactions produce high-quality aromatic compounds, e.g. vanillin and coniferyl alcohol (see Figure 16A for an overview). These molecules are of interest for the food, flavour and fragrance industry, exemplified by several patents, e.g. from Mane [192], Rhodia [193] and Unilever [194]. VAO can also produce chiral aryl alcohols. For instance, the VAO-mediated conversion of 4-ethylphenol results in the formation of (*R*)-1-(4-hydroxyphenyl)ethanol with an enantiomeric excess of 94% [195]. Interestingly, an engineered variant of VAO was shown to be capable of producing the (*S*)-isomer of 1-(4-hydroxyphenyl)ethanol with an enantiomeric excess of 80% [46]. It has been proposed that the methyl ether 4-(methoxymethyl)phenol is the physiological substrate of VAO, as it is the only known substrate that induces expression of the *vao* gene in *P. simplicissimum* [31]. However, little is known about the origin of 4-(methoxymethyl)phenol and the physiological role of VAO. Subcellular localisation studies showed that the *P. simplicissimum* enzyme (*Ps*VAO), together with a co-inducible catalase-peroxidase, is distributed throughout the cytosol and peroxisomes [29].

*Ps*VAO is the prototype of a large flavoprotein family, the VAO/PMCH family, together with *p*-cresol methylhydroxylase (*Pp*PCMH, from *Pseudomonas putida* NCIMB 9866, recommended name: 4-methylphenol dehydrogenase (hydroxylating), EC 1.17.99.1). The FAD-binding domains of all enzymes within this family share a common fold [179]. Within the VAO/PMCH family, VAO and PCMH belong to the 4-phenol oxidising (4PO) subgroup [196]. All 4PO enzymes contain a Tyr-Tyr-Arg triad, which

Catalase-peroxidases (EC 1.11.1.21), peroxidases (EC 1.11.1.7) and catalases (EC 1.11.1.6) are all oxidoreductases acting on a peroxide as acceptor and are peroxidases (1.11.1.x). However, catalases exclusively convert hydrogen peroxide to water and dioxygen, whereas peroxidases convert hydrogen peroxide and a phenolic donor molecule to water and an activated phenolic donor molecule. Catalase-peroxidases are enzymes which can catalyse both types of reactions and use the same heme active site do to so. The sequences of catalases and catalase-peroxidase differ significantly, making a bioinformatic separation possible as well.

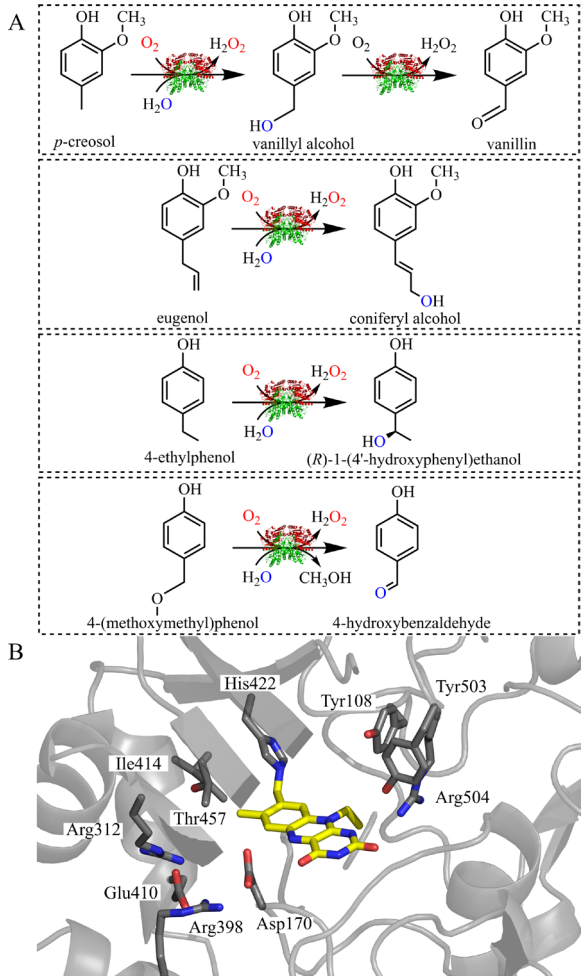


Figure 16: Different reactions catalysed by *Ps*VAO (A) and active site of *Ps*VAO (B).
A: *p*-Creosol (4-methylphenol) is converted via vanillyl alcohol (4-hydroxy-3-methoxybenzyl alcohol) to vanillin (4-hydroxy-3-methoxybenzaldehyde). Coniferyl alcohol (4-hydroxy-3-methoxycinnamyl alcohol) is produced from eugenol (2-methoxy-4-allylphenol). 4-Ethylphenol is converted to (R)-1-(4-hydroxyphenyl)ethanol. The proposed physiological substrate 4-(methoxymethyl)phenol is oxidatively demethylated to 4-hydroxybenzaldehyde.
B: Active site of *Ps*VAO (PDBID 1VAO). The FAD cofactor is coloured in yellow and critical amino acid residues are shown in dark grey. See Table 1 for more information about these critical amino acid residues.

is crucially involved in substrate binding, and therefore is the cause of the selectivity of these enzymes for *para*-substituted phenols [49]. VAO is the only known fungal member of this subgroup, whereas all other known members are bacterial enzymes [51, 52, 55, 56, 196, 197, 198].

Among the 4PO enzymes, *Ps*VAO is characterised in greatest biochemical detail [199]. It has been established that the covalent binding of the FAD cofactor of *Ps*VAO is

crucial for the redox properties of this enzyme, as it significantly increases its redox potential [44]. This increase in redox potential speeds up the rate of reduction of the FAD by the substrate, and thus increases the overall reaction rate [44]. The same effect has been observed in *Pp*PCMH [200, 201]. *Ps*VAO has most properties in common with eugenol oxidase from the actinobacterium *Rhodoccocus jostii* RHA1 (*Rj*EUGO) [55, 202]. Both enzymes contain a 8α-(N3-histidyl)-FAD as prosthetic group, use oxygen as electron acceptor, and share a considerable overlap in substrate specificity. A main difference, however, is that *Ps*VAO is as a homo-octamer [31], while *Rj*EUGO is a homo-dimer [55]. A recent protein engineering study showed that an extra surface loop in *Ps*VAO determines this difference in oligomerisation behavior [39]. Another remarkable difference between *Ps*VAO and *Rj*EUGO concerns the reactivity with alkylphenols. *Ps*VAO converts a wide range of alkylphenols to the corresponding alcohols or alkenes [37], whereas *Rj*EUGO shows almost no activity with these compounds [55].

Because of the conservation of the FAD-binding domain, all 4PO enzymes share at least 25% sequence identity. To reliably identify novel VAOs, their fungal origin can be used, as well as conservation of known critical amino acid residues (see Table 3.1 and Figure 16B). Currently, no clear sequence motifs specific to VAO are defined [203, 204]. Little is known about the occurrence of VAO in fungi, besides that it is produced in *P. simplicissimum*, *Byssochlamys fulva* and *Fusarium monoliforme* [31, 34, 205]. In *B. fulva*, expression of *vao* is induced by vanillyl alcohol [34]. In *F. monoliforme*, *vao* expression was observed with veratryl alcohol and anisyl alcohol, and to a lesser extent with vanillyl alcohol [205]. In *P. simplicissimum*, vanillyl alcohol and eugenol are not inducing expression of *vao*, but 4-(methoxymethyl)phenol, anisyl alcohol and veratryl alcohol are [35].

Table 3.1: Function of critical amino acid residues in *Ps*VAO.

Residue(s)	Function
Tyr108, Tyr503, Arg504	substrate binding, substrate deprotonation [40, 49]
His61	autocatalytic formation of covalent bond of FAD to His422 [41]
His422	covalent bond to FAD [40]
Ile414	no function known in VAO but the corresponding Tyr residue in PCMH is covalently binding FAD [56]
Asp170	autocatalytic formation of covalent bond of FAD to His422; catalytic base; stereoselectivity of VAO [40, 195, 43]
Thr457	stereoselectivity of VAO [195, 43]
Loop 219-240	essential for octamerisation [39]
Tyr51, Tyr 408	tyrosines suggested to be involved in dioxygen migration to the *re* side of the active site [199]
Asp192, Met195, Glu464, His466, (Tyr503)	access of phenolic ligand to *si* side of active site [199]

Here, we aim to shed light on the occurrence of VAO in fungi. VAO has recently been added to the CAZy (Carbohydrate-Active Enzymes) database, as a separate family of auxiliary activities (AA4) involved in lignin breakdown [166]. We have used this criterion to mine fungal genomes in the Mycocosm database for putative VAOs.

Results

We have identified candidate VAOs in the Mycocosm database and performed a phylogenetic analysis of them. We have used the bacterial 4POs as outgroup for the analysis of these fungal sequences. We mapped the distribution of VAO homologs in different fungi. We analysed the phylogenetic tree obtained, as well as conservation of critical

amino acid residues and cellular localisation of putative VAOs. We analysed the genomic environments of these putative VAOs. We also propose motifs for the future identification of fungal VAOs, as well as bacterial 4POs.

Fungal distribution of VAO homologs Different classes of fungi possess sequences homologous to the sequence of *Ps*VAO, which are annotated as AA4 in the Mycocosm database. In Basidiomycota only Agaricomycotina were found to contain putative VAOs (11). In Ascomycota, only Pezizomycotina contain putative VAOs: dothideomycetes (29), leotiomycetes (6), sordariomycetes (19), eurotiomycetes (25) and orbiliomycetes (3), see also Figure 17. Gene duplications appear to have occurred in almost all fungal classes containing VAO homologs. As can be seen in Figure 17, in Agaricomycotina only three organisms have multiple copies of *vao* genes, while in Pezizomycotina, 19 organisms have multiple copies of *vao* genes. Most striking is *Aspergillus sydowii*, which contains seven putative VAOs.

The fungi found to carry a *vao* gene are living in very diverse environments, e.g. *Jaapia argillacea* in the northern hemisphere and *Aspergillus glaucus* in the artic regions, while *Eurotium rubrum* is the most common fungal species isolated from the Dead Sea. Other fungi e.g *Gloeophyllum trabeum*, *Dacryopinax primogenitus* and *Glonium stellatum* are wood decayers. This observation does not enable us to draw a conclusion on the native functions of VAOs or fungal 4POs, other than that they are likely involved in a metabolic pathway used by all these fungi.

The prevalence of VAO sequences in fungal genomes does not follow fungal taxonomy. While several fungal clades do not appear to contain any species with a VAO, there are no clades in which all species contain a VAO. The presence of a VAO in a species also does not appear

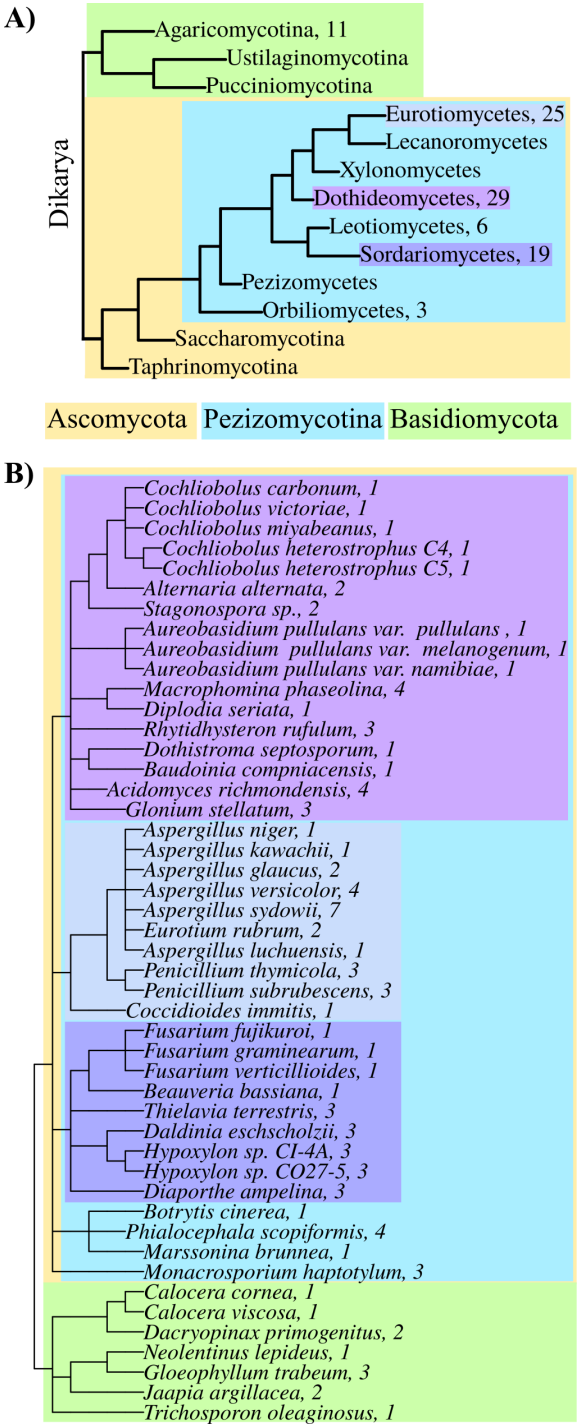


Figure 17: Species-tree showing the occurrence of putative VAO sequences in different fungi. The branch lengths are indicative of evolutionary distances.
A) Overview of all fungal classes, with the number behind the class indicating how many putative VAOs were found in that class. No number means there is no putative VAO present.
B) Details on the fungal classes that were found to contain putative VAOs. The number behind the organism indicates how many putative VAOs were found in that organism. Note that especially the branch lengths for closely related species are not accurate.

to be related to life style or biotope as they can be found in some saprobic, plant pathogenic and animal/human pathogenic fungi, but are also frequently absent in other fungi with these life styles. The same applies to the number of *vao* copies, which differs for species with a VAO between one and seven and again cannot be linked to biotope or life style.

Three different subgroups/clades of fungal VAOs Here we will describe the phylogenetic tree of *Ps*VAO homologues in more detail (see Figure 18). This tree was created using the bacterial 4POs (*Rj*EUGO and *Pp*PCMH) and the novel fungal VAOs. All sequences were between 502 and 632 amino acids long. Please note that *Rj*EUGO is more closely related to *Ps*VAO than the other bacterial 4POs, forming its own clade (marked in green in Figure 18). ART40635, WP_068805427, WP_061561479, AAM21269 and WP_056693180 are bacterial sequences that were added to expand the outgroup. Interestingly, these bacterial sequences do not form one outgroup in the tree but two distinct clades. They separate the fungal VAOs into one clade more closely related to PCMH and one clade more closely related to EUGO.

Three main clades can be identified, with 4POs containing a cytochrome *c* subunit sequence rooting the entire tree (marked in grey in Figure 18). The blue clade (see Figure 18) contains sequences, which are markedly different from *Ps*VAO in at least three aspects: i) they do not contain the octamerisation loop [39], nor do they contain Tyr51, a proposed gatekeeper for oxygen access to the active site [199]; ii) they do not share the three residues Arg312, Arg398, Glu410 which are conserved in most of the other homologs; iii) some of these sequences contain a short loop between positions 321 and 322, which is absent in all other sequences studied here. Please note that all residues are numbered according to the *Ps*VAO numbering. The clade containing *Ps*VAO is marked in red in Fig-

ure 18 (VAO clade). All sequences in this clade contain the octamerisation loop, or at least a 2-4 residues shorter variant thereof. Members of the VAO clade all contain Tyr51, Arg312, Arg398 and Glu410 (except the sequence from *Beauveria bassania* (Beaba1), which contains an aspartate at position 312). The yellow clade (see Figure 18) can be split into two sub-clades, one without an octamerisation loop and one with a very short octamerisation loop (5 instead of 16 residues). It is doubtful whether this loop has the same function as the octamerisation loop in *Ps*VAO [39].

Cellular localisation All identified fungal proteins appear to be intracellularly located since no N-terminal signal peptides have been found in any sequence using the SignalP 4.1 Server. A proposed C-terminal peroxisomal targeting signal (WKL-COOH instead of the SKL-COOH peroxisomal targeting signal-1 [29]) was only present in some of the sequences in the VAO clade, or further changed to YKL-COOH in eurotiomycetes and absent in all other sequences. In the same study, it was found that VAO localises in the cytoplasm as well as peroxisomes in *P. simplicissimum* [29]. The predictions of the DeepLoc server [208] reflect the experimental data, especially the bimodal distribution observed experimentally for VAO: cytoplasmic or peroxisomal localisation is often predicted to be equally likely. With only two exceptions (WP_056693180.1 and Calvi1_515549 (*Calocera viscosa*)), all sequences are predicted to be peroxisomal or cytoplasmic.

In *P. simplicissimum*, VAO is expressed together with a catalase-peroxidase [29]. At the time of that study, no sequence for this catalase-peroxidase was available, but in the meantime several sequences and crystal structures from related organisms have been solved, e.g. [209, 210, 211, 212]. Using sequence Q96VT4 from *Aspergillus nidulans*, we searched the genomes of fungi in the red clade for

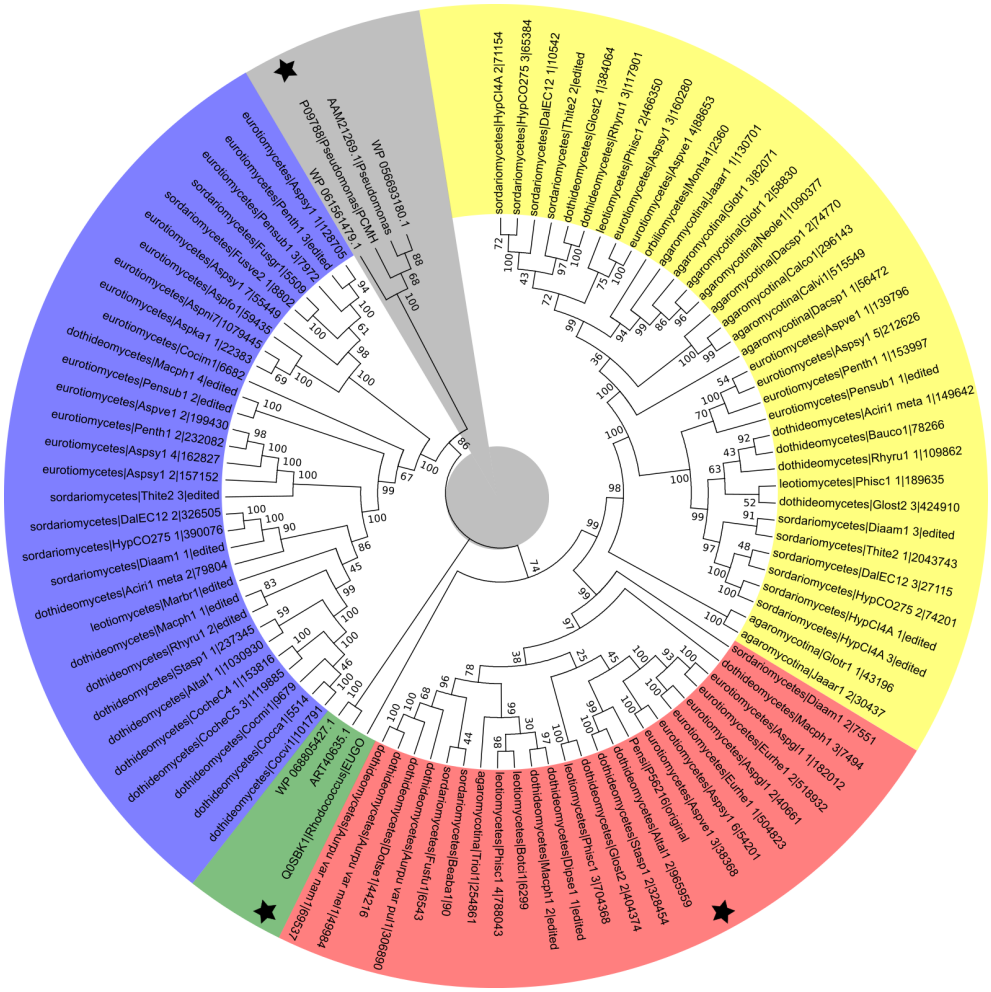


Figure 18: Analysis of the molecular phylogeny of VAO by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [206]. The tree with the highest log likelihood (-36578.6019) is shown. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 97 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 429 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [207]. Minor manual modifications were made to the obtained tree.

homologs of this catalase-peroxidase. In total, 87 putative catalase-peroxidases were found, including isoforms in several fungi. Ninety percent of these sequences were predicted by the DeepLoc webserver to be localised in the cytoplasm or peroxisome.

Genomic Environment of putative VAOs We analysed the genomic environment of sequences in the VAO clade (red clade, marked in red in Figure 18). There is some synteny between closely related species, but a notable lack thereof in a broader view across the fungal kingdom. The fact that VAO is not located in a region of high synteny could indicate that horizontal gene transfer played a role in the origin of fungal VAOs.

Conservation of critical amino acid residues All fungal sequences identified contain the Tyr108-Tyr503-Arg504 triad and His422 with the exception of Diaam1_2 (7551, from *Diaporthe ampelina*), which has phenylalanine residues at positions 503 and 422. His61, which has been found to be involved in the autocatalytic formation of the covalent His422-flavin bond [213] is not conserved in the sequences, but often replaced by a tyrosine, methionine or phenylalanine. No sequence has a tyrosine residue at position 414, but two carry a cysteine (Bauco1 (78266, from *Baudoinia compniacensis*) and Aciri1_meta (49642, from *Acidomyces richmondensis*)).

The catalytic base, Asp170 is present in all sequences in the VAO clade, with the exception of Diaam1_2 (7551, from *D. ampelina*), which carries a glutamate at that position. Sequences in other clades carry mainly serine or alanine at position 170. Thr457 is not strictly conserved in either of the three clades, but the majority of sequences carry Thr457 or Asp457. These two facts combined are a good predictor for changed or absent enantioselectivity [46]. Especially in the blue clade, inverted enantioselectivity can be expected, likely producing (*S*)-

1-(4-hydroxyphenyl)ethanol from 4-ethylphenol [46].

The loop between residues 220 and 237 in *Ps*VAO is present in all sequences in the VAO clade, and absent in all other sequences. Interestingly a much shorter loop (five residues) is present in some sequences of the yellow clade. This indicates that not all of the sequences identified here will produce octameric VAOs. Considerable variation in the loop itself occurs, but prolines appear to be more conserved than other residues in the loop.

Sequences in the blue clade are more similar to the bacterial 4POs in the sense that they do not share Asp170 with *Ps*VAO, but contain a serine or alanine instead. However, they all share His422 with *Ps*VAO. Like the bacterial sequences, sequences in the blue clade lack residue Tyr51. See also Supplementary Figure 1 for sequence logos of the three fungal clades (yellow, red and blue). Tyr51 was suggested to be involved in dioxygen migration to the active site (together with Tyr408, which is not conserved). These two tyrosines are located in a loop in *Ps*VAO, where the cytochrome *c* subunit binds to the FAD-subunit in *Pp*PCMH [199].

Motifs for VAOs *Ps*VAO and the other members of the 4PO subgroup, as well as members of the VAO/PCMH family all share the same FAD-binding fold, for which a motif has been proposed [203]. Their specific mode of FAD binding however differs: they bind their FAD cofactor non-covalently, mono-covalently, or bi-covalently [196]. Members that bind their FAD cofactor via the C8α position of the isoalloxazine ring do not always use the same residue or even the same domain [214]. This fact can be used to distinguish between the different members at the sequence level [204]. Expanding on this idea of function-related patterns or motifs, it should also be possible to distinguish between 4PO enzymes with different activities based on patterns or motifs of active site residues.

Table 3.2: Motifs for 4POs and their starting position in *Ps*VAO, defined based on Figure 19 (in PROSITE format). Bold letters indicate a critical amino acid residue.

4PO-motif-1; start 96	P-x-x-x-x-S-x-G-[RK]-N-x-G-**Y**-G-[GS]
4PO-motif-2; start 143	E-P-G-V-[TS]-[FY]
4PO-motif-3; start 178	G-N-x-x-x-x-G
4PO-motif-4; start 252	G-x-F-x-Q-x-x-x-G-x-x-x-K-x-G-x-x-L
4PO-motif-5; start 503	**Y**-**R**-x-x-(x)-x-x-x-D
VAO-motif-1; start 169	P-**x**-x-G-x-(x)-G-S-x-x-G-N-x-x-x-x-G
VAO-motif-2; start 422	**H**-x-x-x-x-P-x-x-x-x-x-G

To define a motif specific to VAOs, we expanded the sequence motifs proposed by Dym and Eisenberg [203] with the present knowledge of critical amino acid residues (see Figure 19). The well-known dinucleotide-binding motif [215] is absent in VAO, as these enzymes do not contain a Rossmann fold, but have their own, distinct fold, which is shared by all members of this family. We defined motifs based on Figure 19, as summarised in Table 3.2. Some of these motifs are specific to VAO, and others to the known members of the 4PO subgroup.

Discussion

Fungi and bacteria separated about 4090 million years ago (MYA). This makes it remarkable that *Rj*EUGO still clusters more closely with VAO than with the bacterial 4POs. To better understand this, we looked at *Rj*EUGO more closely. This enzyme originates from *Rhodoccocus jostii* RHA1 [55, 202], which is an actinobacterium. The other bacterial members of the 4PO subgroup all come from *Pseudomonas* strains, which makes them proteobacterial. Actinobacteria and Proteobacteria separated 3169 MYA. It is interesting to note that these speciation events all took place before the great oxidation event, so before at-

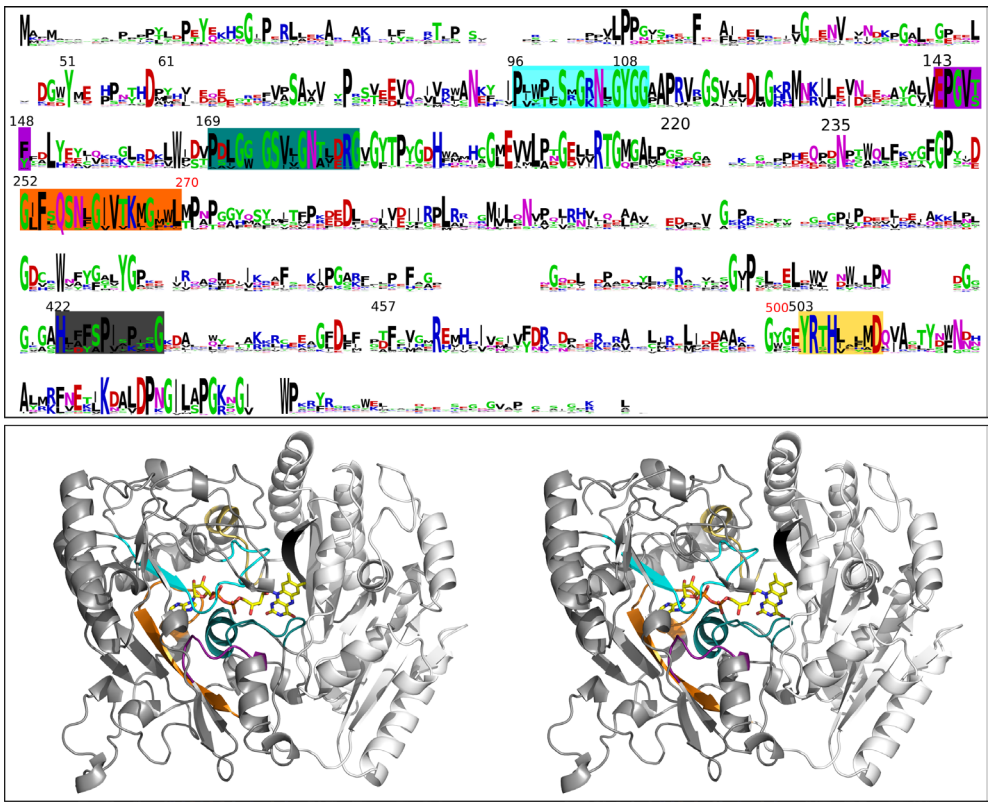


Figure 19: Sequence logos of fungal VAOs and their location in the structure (PDBID 2VAO). Top: coloured boxes indicate the different motifs. Cyan highlights the 4PO-motif-1, purple the 4PO-motif-2, turquoise the 4PO-motif-3, orange the 4PO-motif-4 and yellow the 4PO-motif-5. Turquoise also highlights the VAO-motif-1, which overlaps with the 4PO-motif-3 and grey the VAO-motif-2. Gaps are introduced by the removed bacterial outgroup. Bottom: Stereoview of the structure of VAO with the same colours as above indicating the location of the different motifs defined in this study (see Table 2 for details). The FAD cofactor is shown as sticks and coloured in yellow. The cap domain (residues 270-500, interrupting the FAD binding domain) is shown in white and the FAD binding domain in grey.

mospheric oxygen levels increased permanently (2400 to 2100 MYA, [3]). The absence of atmospheric oxygen at the time of this separation makes it likely that the common ancestor of all 4POs was not a true oxidase.

Candidate fungal VAOs are present in a variety of fungi, but predominantly in the class of Pezizomycotina and Agaricomycotina. None of the earlier lineages appear to contain any copies of this gene, suggesting that it originated after the dikarya evolved as a separate group from those earlier lineages. However, in those groups of fungi that contain species with a VAO (e.g. Sordariomycetes, Dothidiomycetes, Eurotiomycetes, Leotiomycetes, Agaricomycotina), the presence of *vao* genes is restricted to a small number of species, often not closely related. The broad diversity of species that contains *vao* genes could indicate that VAO is an old enzyme that was frequently lost in fungal species, likely due to it no longer being essential. However, the phylogenetic analysis of fungal VAOs also points to an alternative explanation. As the VAO tree only partially follows the taxonomic relationships between the species, this suggests the possibility that several of these species obtained the *vao* gene through horizontal transfer, possibly from bacteria. This would explain the low frequency of presence of VAO in fungal genomes, as well as the phylogenetic diversity of the fungal VAO candidates. Interestingly, among the species that contain candidate *vao* genes, the number varies from one to seven copies, suggesting that those species with multiple copies have really capitalized on this enzymatic function and therefore likely require it for life in their biotope. An interesting example of this is *Aspergillus sydowii* that has the highest number of *vao* genes detected (seven) and has both a terrestrial and a marine life style, in which it has been implicated as a major causal agent of coral bleach [216]. Another factor speaking in favour of horizontal transfer as the origin of fungal VAOs is the non-conserved

genomic region in which the genes have been detected. This would be consistent with introduction by horizontal transfer, while an ancient fungal origin would more likely result in some conservation/synteny of the neighbouring genes of the *vao* genes, which we could not detect.

Based on the conservation of critical amino acid residues, we are confident that the sequences we have identified will produce enzymes with VAO-like activity, with the exception of the putative VAO from *D. ampelina*. This sequence might be a pseudogene. Sequences in the blue clade are likely to have modified activities, especially changed enantioselectivity. We argue that it is more reliable to use critical amino acid residues than overall sequence identity to predict the enzymatic function of a sequence. Sequence identities between *Ps*VAO and the novel VAOs in the blue clade are only between 30 and 40 %, but most critical amino acid residues are strictly conserved. It is therefore crucial to use these critical amino acid residues when annotating novel VAO sequences. We have also included the critical amino acid residues into the motifs we defined.

The functional role of the identified motifs is likely structural: the residues in the 4PO-motif-1 form a loop, which is in close contact with the FAD cofactor. It is important to note that one of the tyrosines crucially involved in substrate deprotonation (Tyr108) is located in this loop. This motif is therefore likely linked to proper orientation and stabilisation of the position of Tyr108. The 4PO-motif-5 also forms a loop containing the other two residues involved in substrate deprotonation, Tyr503 and Arg504. Proper orientation of these residues is also the likely reason for the conservation of this motif.

Searching the Swiss-Prot database with the defined motifs shows that, as could be expected, the shorter motifs are not specific enough to identify 4POs from

the database. However, 4PO-motif-1 is highly selective and only yields sequences of *Ps*VAO and *Pp*PCMH. The sequence of *Rj*EUGO is not a reviewed sequence entry in the Swiss-Prot database and therefore not a hit. We then expanded our search to GenBank, UniProt, RefSeq and PDBSTR. Only the 4PO-motif-1 was sufficiently selective, giving 848 hits. These sequences are (automatically) annotated as VAOs, PCMHs, glycolate oxidases, alcohol oxidases, aryl alcohol oxidases, hypothetical proteins, FAD-binding oxidoreductases, FAD binding domain-containing proteins and uncharacterised proteins. It is highly likely that the annotations of these sequences are erroneous and that they are actually all VAOs. For example, one sequence annotated only as FAD-binding oxidoreductase is actually *Rj*EUGO (WP_011595933.1). *Ps*VAO is also annotated as an aryl-alcohol oxidase in many organisms, including Y15627 from *P. simplicissimum*. This annotated aryl-alcohol oxidase is identical to P56216, the original sequence of VAO. Calling VAO an aryl-alcohol oxdiase is not technically wrong, but leads to confusion because the enzymes with EC number 1.1.3.7 that are defined as aryl-alcohol oxidases are part of the glucose-methanol-choline oxidoreductase family [217].

We recommend using the 4PO-motif-1 for a knowledge-based search of novel 4POs or to improve the annotation of these sequences. The additional use of the other motifs will help to further narrow down putative functions. In addition to these motifs, several fingerprint residues can be used to identify VAOs in a given set of sequences (see Table 1). Including the original *Ps*VAO sequence (P56216.1) is highly recommended for ease of finding these fingerprints.

Cellular localisation of the putative VAOs and a co-expressed catalase-peroxidase was predicted to be peroxisomal or cytoplasmic. This is in agreement with exper-

imental data available on both enzymes from *P. simplicissimum* [29, 218]. It is likely that fungi that intracellularly express highly active VAO also co-express a catalase-peroxidase to prevent accumulation of toxic levels of hydrogen peroxide. However, the catalase-peroxidase genes we identified in the genomes of the red clade are not located in the immediate environment of the *vao* gene.

VAO and the other members of the 4PO subgroup are the only known members of the VAO/PCMH family that covalently bind their FAD cofactor via a residue in the cap domain. The other family members bind FAD either mono-covalently or bi-covalently via (a) residue(s) in the FAD-binding domain, or non-covalently. Tyrosines, cysteines and histidines are involved in the covalent binding modes [40, 56, 219]. Histidines can bind FAD via an 8α-N1-histidyl (to the FAD-binding domain) or 8α-N3-histidyl bond. *Ps*VAO uses His422 of the cap domain to install the 8α-N3-histidylflavin bond [40]. Interestingly, many putative VAOs do not contain His61 which is involved in the autocatalytic incorporation of FAD [41], but all contain His422 and none contain Tyr414. For putative VAOs with Tyr61 this residue may be able to have the same function. But for other putative VAOs, the autocatalytic incorporation of FAD, as observed for *Ps*VAO, is unlikely, or another mechanism is in place to activate His422. The absence of Tyr414 also negates the option of a 8α-O-tyrosyl-FAD, although two sequences with a cysteine at position 414 might have a 8α-S-cysteinyl-FAD. Some of these putative VAOs might bind their FAD cofactor non-covalently.

Conclusion

In conclusion, we have identified sequences, which represent with a high likelihood novel, fungal VAOs. Genomic analysis revealed presence of VAO in only a small subset of fungal species that are present in several fungal orders. Based on phylogeny and synteny analysis, it seems more likely that fungi that possess VAOs obtained these through horizontal transfer, possibly from bacteria, rather than that many fungi have lost this gene through evolution. We have defined a specific sequence motif (P-x-x-x-x-S-x-G-[RK]-N-x-G-Y-G-[GS]) that recognizes 4PO enzymes, that can be used in combination with several fingerprint residues to identify novel VAOs. Several properties of the members of the 4PO subgroup remain confusing. Their substrate specificities overlap, but *Rj*EUGO is unable to convert alkylphenols. *Pp*PCMH and other bacterial 4POs as well as the fungal *Ps*VAO are able to convert these molecules enantioselectively. Their fungal or bacterial origin cannot be used to functionally separate them as the bacterial *Rj*EUGO is a true oxidase like *Ps*VAO, but the other bacterial 4POs are not. Also, these enzymes are induced by different molecules [35, 197, 220], and no common inducer or physiological substrate has been identified (yet). Biochemical characterisation of the identified fungal sequences will help to shed light on whether there are "fungal EUGOs" or "fungal PCMHs" and additional work to identify and characterise bacterial members of the 4PO subgroup will also help to illuminate whether there are "bacterial VAOs". More work is certainly needed to disentangle these enzymes from each other.

Methods

VAO is currently the only member of the Auxiliary Activity Family 4 (AA4) of the CAZy database. In CAZy, there are 21 eukaryotic and 131 bacterial sequences classified as AA4, presenting a clear bias against fungal sequences. A manual search of the Mycocosm database (April 2017) revealed 89 putative VAOs. Gene models of enzymes annotated as CAZy-AA4 in the Mycocosm database were analysed. Sequences from non-published genomes were removed, greatly reducing the amount of sequences. Protein sequences and genomes were downloaded and gene models were manually corrected where necessary. Sequences with obvious internal STOP codons were removed. If unresolvable problems were present with one sequence from an organism, all sequences from that organism were removed. Sequences were aligned using clustal omega. The verified alignment was used to build a phylogenetic tree with MEGA (version 5.2). Sequences originating from Ascomycota and Basidiomycota were identified, with the majority of sequences in the Ascomycota phylum (78, vs 11 in Basidiomycota). Gene models for 15 of these 89 uncharacterised fungal sequences were manually corrected. Organisms from different classes in the two phyla contained putative VAOs. The original sequence of *Ps*VAO was added to the 89 putative VAOs, as were the bacterial sequences of characterized 4POs, giving a total of 97 sequences, for which an alignment, and subsequently a phylogenetic tree, was built using the clustal omega webserver (`https://www.ebi.ac.uk/Tools/msa/clustalo/`) and the MEGA5 program [206, 207]. Sequencelogos were created using the weblogo webserver ([221], `http://weblogo.berkeley.edu/`) and edited with inkscape (`https://inkscape.org/en/`). The genomic environment was manually extracted from the Mycocosm database, only selecting ORFs on the same strand as the putative VAO. Visualisation was performed with inkscape. The DeepLoc webserver was used to analyse the probable localisation of putative VAOs ([208], `http://www.cbs.dtu.dk/services/DeepLoc-1.0/`). Estimated times of speciation events of the last common ancestor of organisms containing 4POs were median times given by `http://timetree.org/`. The proposed motifs were tested using `http://www.genome.jp/tools/motif/MOTIF2.html`, on 14.11.2017. Catalase-peroxidases were identified in fungi in the red clade by blasting with Q96VT4. Cellular localisation of sequences of the best hits was analysed using the DeepLoc server ([208], `http://www.cbs.dtu.dk/services/DeepLoc-1.0/`).

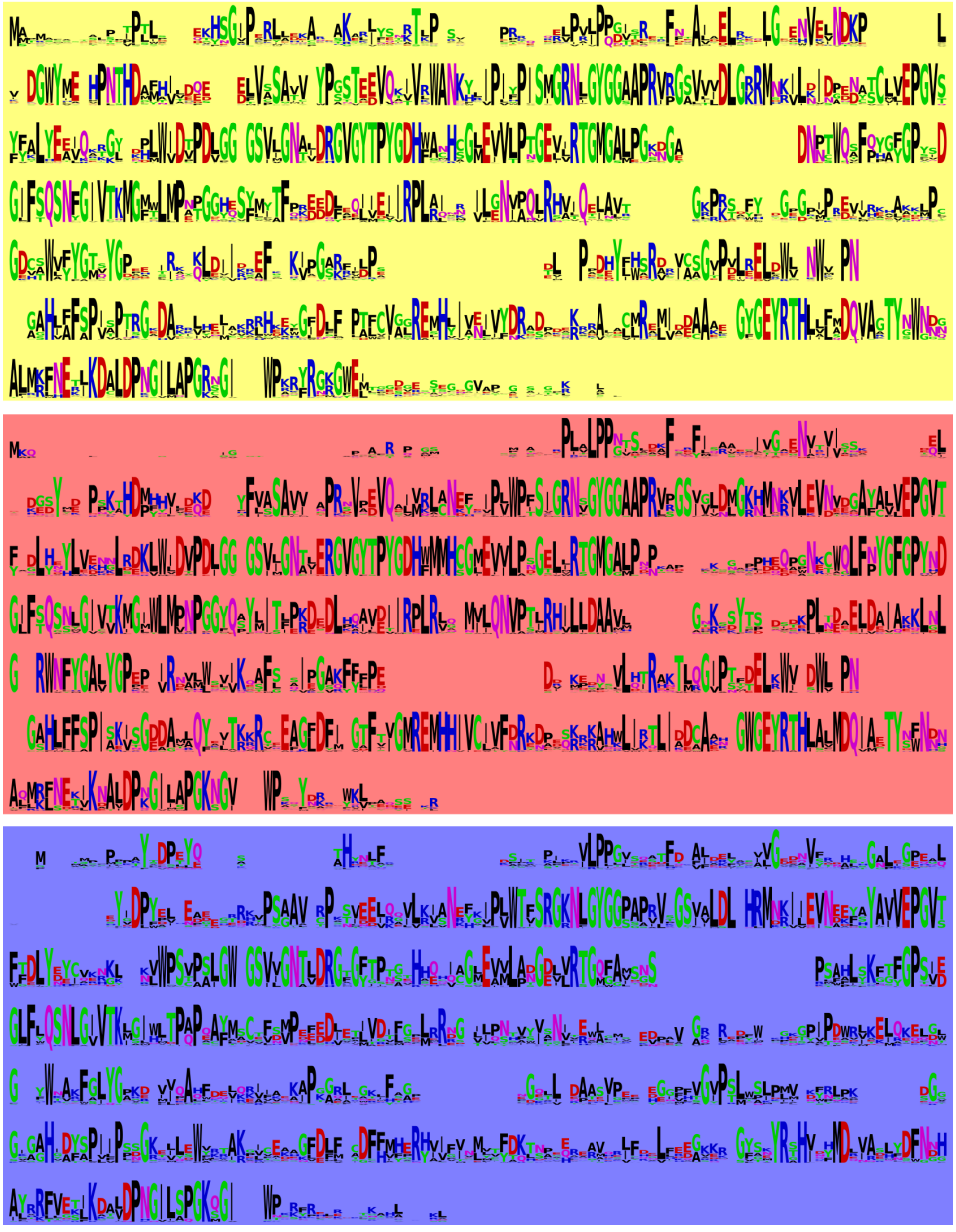
Manual gene model correction can be done like this:

1. download the entire genome of the organism.
2. locate the gene of interest in the gene list.
3. select the DNA of the gene of interest and add at least 1000 nucleotides upstream and downstream.
4. use blastx to search databases using your full gene sequence - this will help to better judge which introns need to be removed.
5. locate introns - in fungi, introns start with G[T/C/A]NNG[TCA] and end with [C/T/A/G]AG and are typically 40-60 nucleotides long.
6. remove introns.
7. use blastx to search databases using your new, manually edited gene.
8. judge if you trust your edited gene or go back to step 4.

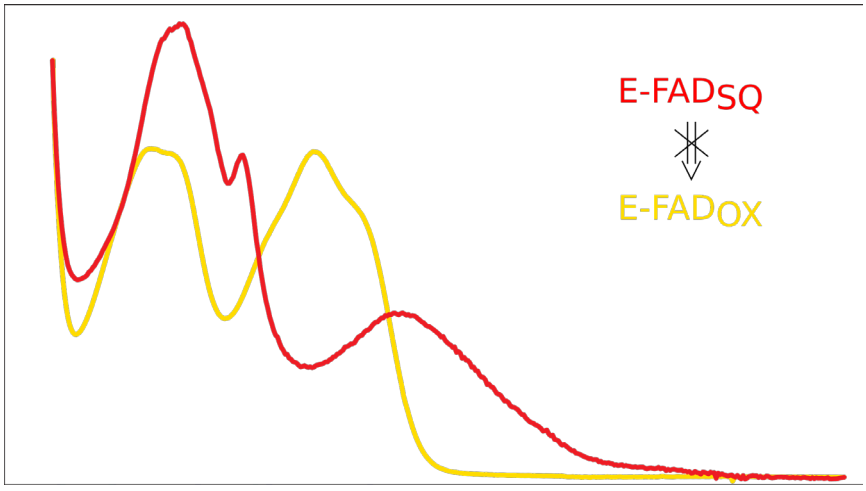
Bibliography follows at the end of the thesis

Supplementary figures

Supplementary Figure 1: Sequence logos of fungal VAOs from different clades (yellow, red and blue clade in Figure 18). The figure follows on page 98.



Chapter 4:
Vanillyl alcohol oxidases produced in Komagataella phaffii contain a highly stable non-covalently bound anionic FAD semiquinone



This chapter is based on:
G. Gygli and W. J. H. van Berkel. *Biocatalysis*, 3(1): 17-26, 2017.

Keywords: Biocatalysis, flavoprotein, redox cycling, oxidoreductase, *Pichia pastoris*

Abstract

Vanillyl alcohol oxidase (VAO) from *Penicillium simplicissimum* is a covalent flavoprotein that has emerged as a promising biocatalyst for the production of aromatic fine chemicals such as vanillin, coniferyl alcohol and enantiopure 1-(4'-hydroxyphenyl) alcohols. The large-scale production of this eukaryotic enzyme in *Escherichia coli* has remained challenging thus far. For that reason an alternative, eukaryotic expression system, *Komagataella phaffii*, was tested. Additionally, to produce novel VAO biocatalysts, we screened genomes for VAO homologues. One bacterial and five fungal sequences were selected for expression, using key active site residues as criteria for their selection. Expression of the putative *vao* genes in *K. phaffii* was successful, however expression levels were low (1 mg per litre of culture). Surprisingly, all purified enzymes were found to contain a highly stable, non-covalently bound anionic FAD semiquinone that could not be reduced by dithionite or cyanoborohydride. Activity experiments revealed that VAO produced in *K. phaffii* does not produce vanillin because the enzyme suffers from oxidative stress.

Enzyme yields: as a rule of thumb for research, an excellent expression and purification system yields more than 100 mg of protein per litre (of culture of the organism producing the enzyme). Because purification can be time-consuming, it is more practical if more enzyme is produced per litre of culture. For activity experiments, only a few μg of enzyme are needed. For other experiments, e.g. crystallisation, larger amounts of enzyme are needed. Often, many (more than 24) different crystallisation conditions need to be tested before enzymes crystallise, each condition requiring between 0.5-2 μl of enzyme solution at concentrations of 2-50 mg/ml. For mechanistic studies of enzymes, even more protein is needed.

Introduction

Vanillyl alcohol oxidase (VAO) is an homo-octameric flavoenzyme, which covalently binds its FAD cofactor [31]. FAD cofactors can attain three different redox states: (i) the oxidised form, which is bright yellow in colour, (ii) the one-electron reduced semiquinone, a radical species which can be blue or red depending on whether it is in the neutral or anionic state, and (iii) the two-electron reduced form which is almost colourless (Figure 20). VAO is active with a wide range of 4-hydroxybenzylic compounds (Figures 21, 21) and has emerged as a promising biocatalyst for the production of aromatic fine chemicals, such as vanillin (4-hydroxy-3-methoxybenzaldehyde), coniferyl alcohol (4-hydroxy-3-methoxycinnamyl alcohol) and enantiopure 1-(4'-hydroxyphenyl) alcohols [31, 36, 37]. During catalysis, the FAD cofactor initially becomes reduced through hydride transfer from the phenolic substrate, generating a quinone methide product intermediate [36]. The fate of this intermediate depends on the type of substrate [223].

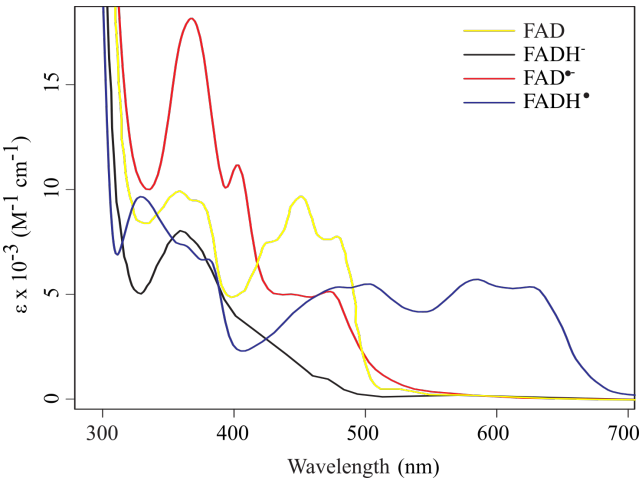


Figure 20: Visible absorption properties of FAD redox states: yellow: oxidised; red: anionic semiquinone; blue: neutral semiquinone; black: anionic hydroquinone (modified from [222]).

With vanillyl alcohol (4-hydroxy-3-methoxybenzyl alcohol), the aldehyde product (vanillin) is formed without the intermediate hydration of the quinone methide (Figures 21, 22). In the reactions with 4-(methoxymethyl)phenol and eugenol (2-methoxy-4-allylphenol), the quinone methide intermediate is hydrated at the Cα- or Cγ-atom, generating 4-hydroxybenzaldehyde and coniferyl alcohol as final products, respectively (Figures 21,22). VAO also catalyses the enantioselective hydroxylation of 4-alkylphenols to the corresponding (*R*)-1-(4'-hydroxyphenyl) alcohols with an e.e. of 94% [37]. In the final step of the VAO reaction, the reduced flavin gets oxidised by molecular oxygen, regenerating the oxidised enzyme [31].

VAO covalently binds its FAD cofactor via His422 [14, 41, 42]. The proposed sequence of events of this post-translational modification of VAO is as follows: Before any incorporation of FAD can occur, the VAO-polypeptide folds and oligomerises to the dimer. The produced apoprotein then encapsulates FAD in its highly pre-organised binding cavity. Non-covalent FAD binding promotes the octamerisation of the enzyme. Next, the flavin becomes covalently bound via an autocatalytic process, involving the fully reduced form of FAD. Covalent incorporation increases the redox potential of the flavin cofactor and stimulates efficient redox catalysis. The full process takes approximately 30-60 minutes [14, 41, 42].

Active site residues which have been found to be essential for the reactions and mechanisms described above are: (i) Asp170: essential for efficient redox catalysis and involved in the autocatalytic flavinylation [43] (ii) His422: covalently binds the FAD cofactor and thus enables efficient redox catalysis [44] (iii) His61: enables His422 to covalently bind FAD [41], (iv) Thr457: together with Asp170 directs the enantioselective hydroxylation of 4-alkylphenols of VAO to the R isomers [46] and (v) Tyr108, Tyr503 and Arg504: form an anion-binding subsite which

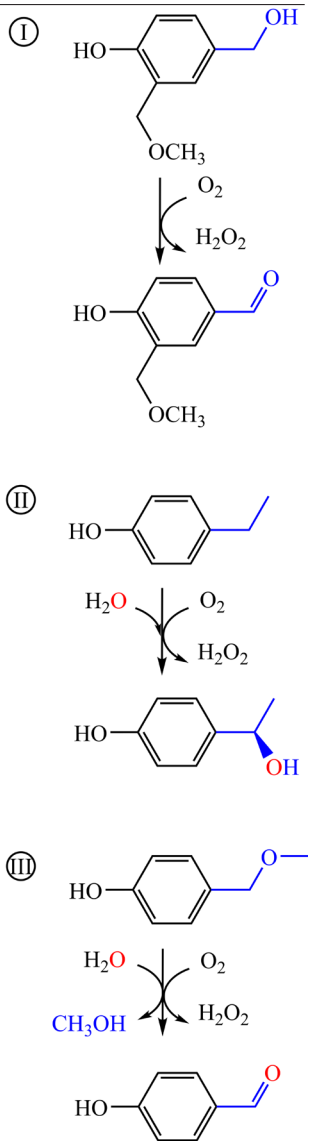


Figure 21: The different reactions VAO catalyses are: the oxidation of vanillyl alcohol to vanillin (I) and the enantio-selective conversion of 4-ethyl-phenol to (*R*)-1-(4'-hydroxyphenyl)ethanol (II) and the oxidative demethylation of 4-(methoxymethyl)phenol to 4-hydroxybenzaldehyde (III).

is hypothesised to stabilise the phenolate form of VAO's substrates [40]. Residues remote from the active site that have been found through directed evolution to influence the enzyme reactivity are Ile238, Phe454, Glu502 and Thr505 [45]. Also, a loop ranging from residue 220 to 235 has been shown to be essential for octamerisation of VAO [39].

Over the past three decades, studies with VAO have led to different experiences producing this enzyme in different expression systems [31, 224]. VAO as it was first described in 1992 was purified from *Penicillium simplicissimum* [31]. Absence of a gene sequence until 1998 [224] necessitated that the production of the enzyme relied on its natural host. Yields for VAO from *P. simplicissimum* were around 5% of total protein produced, leading to approximately 50 mg of protein per litre of culture. VAO was then heterologously produced in *E. coli* (using pBC11 [43]) and *Aspergillus niger* to yields of about 5-10 mg and greater than 50 mg of enzyme per litre of culture, respectively [224]. Because the protein produced extracellularly in *A. niger* was prone to proteolysis, *E. coli* was selected as host for subsequent protein engineering studies. The disadvantage of the relatively low yield of the *E. coli* system compared to production in *P. simplicissimum* was outweighed by the ease of site-directed mutagenesis in *E. coli* systems.

Microbial systems previously used to produce other VAOs are *E. coli*, to express VAO from *Fusarium verticillioides* strains, as well as *Byssochlamys fulva* V107 to produce VAO from *B. fulva* V107 [34, 205]. The amino acid sequence and crystal structure of VAO from *P. simplicissimum* are available in sequence databases and the PDB under the accession codes P56216 and 2VAO, respectively. Amino acid sequences for VAO from *F. verticillioides* strains are annotated as EWG41284.1 and AFJ11909.11, of which the second corresponds to the VAO described in

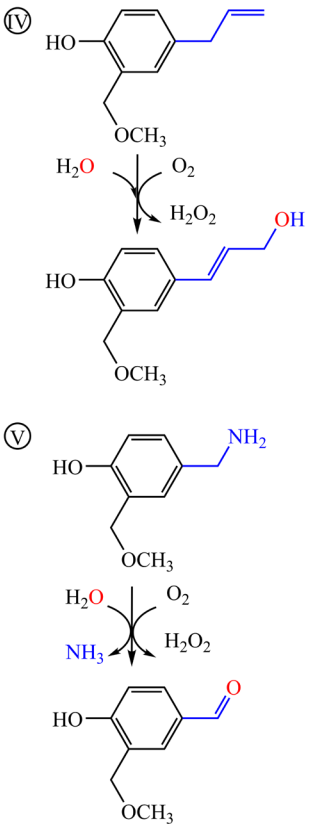


Figure 22: The different reactions VAO catalyses are: the oxidation of eugenol to coniferyl alcohol (IV), and the oxidative deamination of vanillylamine to vanillin (V).

[205]. No amino acid sequence is available for the VAO from *B. fulva* V107.

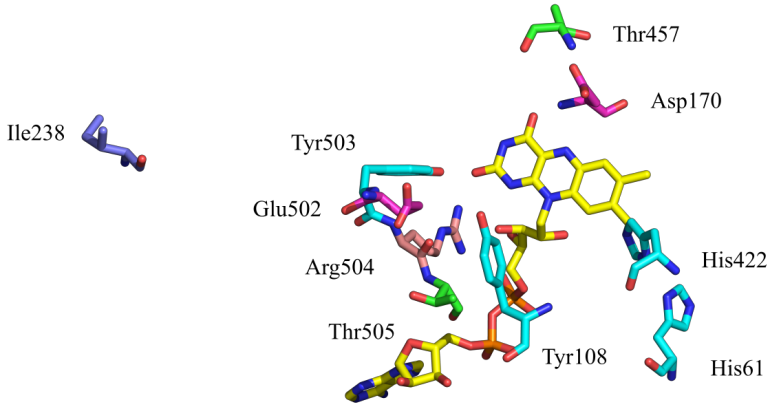
It is difficult to conclusively establish sequence-function relationships based on only two known characterised and sequenced VAOs. One can attempt to uncover such relationships through site-directed mutagenesis or directed evolution, as has been performed previously in our group [41, 43, 44, 45]. Another approach is to tap natural evolution and screen genomes for VAO-like sequences. We have chosen a combined approach, using the knowledge gained from mutagenesis experiments to define a set of key VAO residues which are known to be essential for function of VAO from *P. simplicissimum* and then use this criterion to select VAO-like sequences for expression in *Komagataella phaffii* (formerly known as *Pichia pastoris* or *Komagataella pastoris* [190]). This methylotrophic yeast was chosen for its ability to express eukaryotic proteins to high yields as well as its use as an industrial system for protein expression [225]. All the selected putative VAOs as well as VAO from *P. simplicissimum* produced in this way were found to be catalytically incompetent and to contain a highly stable, non-covalently bound anionic FAD semiquinone (anionic FAD_{SQ}).

Before the age of **genome sequencing**, detailed characterisations of enzymes were commonplace. Now, sequencing data is exceeding enzyme characterisations by far. The dilemma is that often either enzyme kinetic data without an associated enzyme sequence is available or, more often, sequence data without an associated characterisation. Only by having enzyme kinetic data and sequence data available for many enzymes can one hope to establish reliable sequence-function relationships.

Results

Bioinformatics Nowadays, a vast amount of sequence data is available in sequence databases, and when screening these databases for homologues of a known protein one usually finds many more putative proteins than can be produced under non-automated conditions. This is also the case for the genome screening of the non-redundant sequence database for VAO-like sequences, where one finds more than three thousand hits when blasting with the sequence of VAO from *P. simplicissi-*

A



B

	61	108	170	238	422	457	502	503	504	505
P56216 *Penicillium simplicissimum*	H	Y	D	I	H	T	E	Y	R	T
CRG90285 *Talaromyces islandicus*	M	Y	D	S	H	A	E	Y	R	T
M5FTW7 *Dacryopinax* sp.	H	Y	D	T	H	T	E	Y	R	T
C7Z853 *Nectria haematococca*	F	Y	D	A	H	D	E	Y	R	T
W9CKB3 *Sclerotinia borealis*	Y	Y	D	C	H	T	E	Y	R	T
V5FPK0 *Byssochlamys spectabilis*	H	Y	D	T	H	T	E	Y	R	T
K0C1W8 *Cycloclasticus* sp.	Y	Y	A	T	H	E	E	Y	R	T

Figure 23: Visualisation of key residues of VAO. (A) Functionally important residues are coloured according to (B). Also shown is the cova- lently bound FAD cofactor (in yellow). This figure was created using PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.). (B) Summary of key residues in the selected sequences, as numbered in VAO. Data was extracted from the alignment in Supplementary Figure 2 and processed in R [226].

mum. For practical reasons, a maximum of six novel vao genes was selected for expression. Selection of these genes was based on the presence of residues that are important for the functioning of VAO. These key residues are His61, Tyr108, Asp170, Ile238, His422, Thr457, Glu502, Tyr503, Arg504 and Thr505 as well as a loop from residues 220 to 235 (see Figure 23). The alignment of the VAO-like sequences thus identified and their relationships to VAO from *P. simplicissimum* are depicted in Supplementary Figure 1. The only bacterial sequence selected originates from *Cycloclasticus*, a proteobacterium, while the other sequences are of fungal origin, with the sequence from *Dacryopinax* being the only one from a basidiomycete. Sequence identities to VAO from *P. simplicissimum* were 49%, 45%, 62%, 63%, 88% and 42% for CRG90285, M5FTW7, C7Z853, W9CKB3, V5FPK0 and

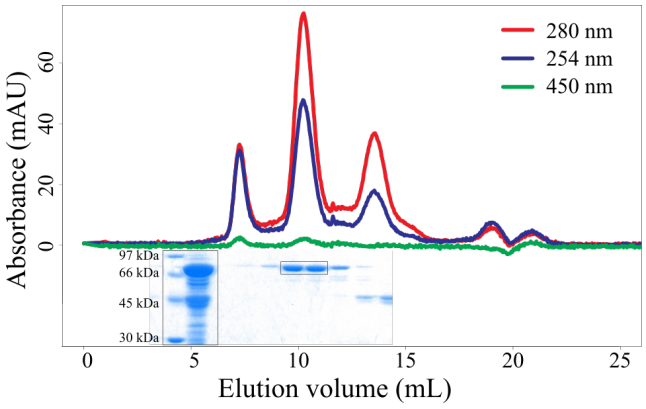


Figure 24: Elution profile of VAO from *P. simplicissimum* as expressed in *K. phaffii* during purification on a Superdex 200 HR 10/30 column. An SDS-PAGE gel of collected fractions is shown below the curves. The big-framed box shows the subunit molecular masses of the marker proteins and the sample before purification on the Superdex 200 column. The small-framed box indicates which fractions were pooled and used for subsequent analysis.

You can already see that something is wrong with the FAD cofactor from this figure: the absorption at 450 nm is very low. Usually, a more distinctive peak indicating elution of a FAD-containing protein can be detected. If you compare the absorpions of FAD_{OX} and FAD_{SQ} in Figures 20 and 25, you see a considerably higher absorption at 450nm of the oxidised FAD compared to the anionic FAD semiquinone.

K0C1W8, respectively. As can be seen from Figure 23B, all these sequences contain the key residues Tyr108, His422, Tyr503 and Arg504, and only K0C1W8 does not contain the critical Asp170.

Production of VAO-like enzymes in K. phaffii VAO from *P. simplicissimum* as well as the VAO-like enzymes were successfully expressed extracellularly under the control of an α-factor signal sequence in *K. phaffii* and purified as described in Methods. During production and purification procedures, activity was tracked using the horseradish peroxidase coupled activity assay (HRP assay) described in Methods. After purification, rather low yields were observed, ranging between 1 and 2 mg of pure enzyme from 2 L of culture. During the third purification step, enzymes eluted at the volume of octameric VAO on a Superdex 200 column, see Figure 24 [227]. SDS-PAGE showed a subunit molecular mass of about 70 kDa for all enzymes, corresponding to the calculated mass of 74 kDa, as can be seen in Figure 24.

M5FTW7 is also present in **Chapter 3**, but under a different accession code. It is from the fungus *Dacryopinax primogenitus*, which is abbreviated as Dacsp, and the sequence number there is 56472.

Absorption spectra revealed the presence of an anionic FAD_{SQ} in all the enzymes expressed in *K. phaffii*, compared to the oxidized FAD observed in VAO expressed

in *E. coli* (see Figure 25A compared to Figure 25B). VAO from *P. simplicissimum* expressed in *K. phaffii* also contained anionic flavin semiquinone (VAO_{SQ}). The use of different growth media (see Methods) did not change the semiquinone redox state of the purified enzymes. Changing the construct by removal of the c-terminal c-myc epitope and His-tag (6x) also lead to the production of VAO_{SQ}. Growth of *K. phaffii* and subsequent enzyme purification under dark conditions also resulted in VAO_{SQ}.

Stability of the anionic FAD semiquinone Attempts to oxidise or reduce the flavin cofactor of VAO with ferricyanide or dithionite or cyanoborohydride failed. VAO_{SQ} persisted also against the addition of the substrates vanillyl alcohol and eugenol, as well as the inhibitors isoeugenol and cinnamyl alcohol. Dialysis to remove eventually bound molecules, either in the absence or presence of excess FAD, did not change the flavin redox state. Reoxidation of VAO_{SQ} was only possible through denaturation of the protein through TCA precipitation or heating. This revealed that the VAO_{SQ} cofactor was non-covalently bound, and the thus freed flavin was confirmed to be FAD through TLC and HPLC/MS.

Of note is that due to the low amounts of protein available, absorption spectra were recorded using a nanodrop 2000c machine, which gave reliable and quick results (see Figure 25). The fact that only 2 µL at most are needed for a measurement reduced greatly the volume needed to test different conditions when trying to oxidise or reduce the anionic FAD_{SQ} in VAO. *Catalytic properties* Tracking product formation by following absorption changes at 340 nm and 296 nm with the common VAO substrates vanillyl alcohol and eugenol indicated that all enzymes expressed in *K. phaffii* did not produce the expected products vanillin and coniferyl alcohol, respectively. This is surprising as activity was detected using the HRP assay with purified enzymes.

Catalase is an enzyme that catalyses the decomposition of hydrogen peroxide to water and dioxygen. Superoxide dismutase catalyses the dismutation of the superoxide radical into dioxygen or hydrogen peroxide. We only see dioxygen recovery if we add catalase to the reaction (and not superoxide mutase), and therefore conclude that no superoxide is formed. Catalase therefore converts the hydrogen peroxide produced by VAO back to dioxygen and this dioxygen is available for further reaction cycles of VAO. In the case of *Ps*VAO_E, recovery of hydrogen peroxide to dioxygen does not lead to more reaction cycles of the enzyme. We can see this because the dioxygen produced by catalase is not used up by the reaction anymore. In the case of *Ps*VAO_K, we do see continuous depletion of the dioxygen produced by catalase. This shows that the substrate is not limiting the reaction of this enzyme, but the presence of dioxygen is.

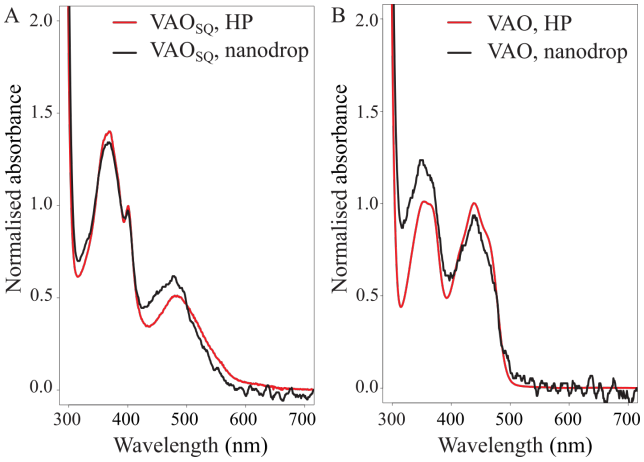


Figure 25: Illustration of the different spectra recorded for (A) VAO_{SQ} with non-covalently bound FAD and (B) VAO_{OX} with covalently bound FAD with Hewlett-Packard HP 8453 diode array spectrophotometer (HP) or Thermo Scientific NanoDrop 2000c UV-Vis spectrophotometer (nanodrop).

To explain the discrepancy between the absorbance measurements of vanillin production (at 340 nm) and hydrogen peroxide formation (followed indirectly at 510 nm), we monitored oxygen consumption with VAO from *P. simplicissimum* as expressed in *K. phaffii* (*Ps*VAO_K) or *E. coli* (*Ps*VAO_E). Measurements for both enzymes with limiting amounts of vanillyl alcohol (0.2 mM), as described in Methods, revealed consumption of oxygen for both enzymes. However, the rate of oxygen consumption with *Ps*VAO_E was 10 times faster than with *Ps*VAO_K. Moreover, *Ps*VAO_E consumed 0.2 mM oxygen while *Ps*VAO_K used up all oxygen present in the reaction mixture. Absorbance measurements and olfactory analysis confirmed that vanillin was only formed with *Ps*VAO_E and not with *Ps*VAO_K These observations hint at non-productive redox cycling in the case of *Ps*VAO_K. Addition of catalase or superoxide dismutase at the end or beginning of the reaction showed the production of hydrogen peroxide during this redox cycling, but not superoxide. In the case of *Ps*VAO_K, the oxygen produced from the reaction of hydrogen peroxide with catalase was completely used up again. In the case of *Ps*VAO_E, this oxygen exhaustion did not oc-

cur because no more vanillyl alcohol was present. We can therefore conclude that with *Ps*VAO$_K$, the substrate acts as an effector uncoupling oxygen consumption from product formation.

Discussion

Many flavoprotein oxidases that form an anionic FAD$_{SQ}$ are known: A transient anionic FAD$_{SQ}$ is observed in D-amino acid oxidase upon titration with methyl viologen at pH 8.5 and in sarcosine oxidase upon titration with L-proline or sarcosine at pH 7.0 [228, 229]. Glycine oxidase was found to form a transient anionic FAD$_{SQ}$ upon photoreduction [230]. Lactate oxidase forms a slowly decaying anionic FAD$_{SQ}$, which is induced by reaction with pyruvate or photoirradiation, and reoxidised by molecular oxygen within 24 h [231]. In methanol oxidase from *K. phaffii* anionic FAD$_{SQ}$ was found to make up approximately 30-35% of the flavin after purification [232]. The anionic FAD$_{SQ}$ of methanol oxidase was not reduced by methanol. A positively charged residue near N(1) is present in many of these enzymes, mostly histidine or arginine, which could stabilise the (transient) anionic FAD$_{SQ}$. In choline oxidase, the anionic FAD$_{SQ}$ is stabilised upon reduction in the wild-type enzyme and the H466A mutant but not in the H466D mutant. The interaction of the negative charge of the N(1) of anionic FAD SQ with the introduced negative charge of Asp466 of the H466D mutant is suggested to be unfavourable for the formation of the anionic FAD$_{SQ}$. The absence of a positive charge near flavin N(1) in the H466A mutant is not enough to prevent formation of the anionic FAD$_{SQ}$ in this enzyme [233]. Interestingly, the anionic FAD$_{SQ}$ in choline oxidase is unusually insensitive to oxygen or ferricyanide at pH 8, but can be transformed into oxidised FAD by incubating the enzyme at pH 6 for several hours [234].

In this **methanol oxidase** from *K. phaffii*, about 30% of the flavin is also a catalytically inactive, modified flavin, and only about 35% of the flavin is functional.

The lifetime of the FAD$_{SQ}$ in **choline oxidase** is pH dependent. We therefore tested if expression of VAO at pH 6.0 and 7.0 (see Methods), as well as incubating purified *Ps*VAO$_K$ at different pH would also change the lifetime of our FAD$_{SQ}$. Unfortunately, the stability of VAO at pH 6.0 is low, and we observed comparable rates of denaturation to *Ps*VAO$_E$.

For VAO produced in its natural host, *P. simplicissimum*, the semiquinone state is not observed in rapid reaction kinetic experiments. This indicates that the semiquinone is not kinetically stabilised during flavin reduction, in agreement with the hydride transfer mechanism [40]. A transient anionic FAD_{SQ} is observed upon anaerobic reduction with dithionite, but it is easily reoxidised upon mixing with air. Photoreduction of VAO also leads to the transient formation of anionic FAD_{SQ}, while photoreduction in the presence of cinnamyl alcohol yields irreversibly bleached FAD [34]. In VAO H422A , a VAO variant that binds the FAD cofactor non-covalently, a large portion of transient neutral (blue) FAD_{SQ} is observed during xanthine oxidase mediated reduction [44]. In VAO, Arg504 is located near the N(1) locus of FAD.

The above examples highlight that, independent of whether they bind their FAD cofactor covalently or not, many flavoprotein oxidases are able to stabilise anionic FAD_{SQ}, and most do form a transient anionic FAD_{SQ} upon anaerobic reduction with dithionite. However, in all of these enzymes, the flavin radical is thermodynamically unstable and can be easily transformed to FAD_{OX} or FAD_{RED}. Here we found that expression of six novel VAO variants as well as VAO from *P. simplicissimum* in *K. phaffii* results in the production of octameric enzymes that bind the anionic semiquinone form of FAD in a non-covalent mode. Remarkably, the VAO-bound anionic FAD_{SQ} appeared to be extremely stable and could only be transformed to the oxidized state upon protein denaturation. Neither could it be reduced with the strongest reducing agents dithionite or cyanoborohydride. Surprisingly, even VAO from *P. simplicissimum* (P56216), which has been functionally expressed in *E. coli*, *A. niger* and *P. simplicissmum*, presents as catalytically incompetent VAO_{SQ}. Given the presence of His422 and Arg504 in all sequences as well as His61 in P56216, M5FTW7 and V5FPK0, these three en-

Thermodynamic vs kinetic stability: every system can be represented by a free energy surface. On this surface, local and global minima and maxima exist. A local maximum in Wageningen would be the "Wageningse berg", which undoubtedly is not a global maxima when compared to the Swiss alps. And comparing the elevation of the city of Basel to the elevation of Amsterdam, the latter is more likely to be a global minimum than the first. If a molecule is trapped in a global minimum, it is thermodynamically stable, while if it is trapped in a local minimum, it is kinetically stable. The flavin radical is thermodynamically unstable in most cases described here, and thus easily "falls down" from the global maximum.

zymes are expected to covalently bind FAD. Another unexpected finding is that all novel VAO variants present as octamers. Based on the absence of a loop that has been shown to be essential for octamerisation in VAO, K0C1W8, CRG90285 and M5FTW7 are expected to be dimeric proteins [39].

The low yields of these enzymes as well as their unexpected behaviour when expressed in *K. phaffii* leaves open the question whether low yields are due to the unexpected behaviour or the use of an unsuited expression system. In *E. coli* and *P. simplicissimum*, VAO is produced intracelluarly, and in *A. niger* and *K. phaffii* it is secreted. It appears that the formation of VAO_{SQ} is specific to *K. phaffii*, although the mechanism behind the formation of VAO_{SQ} is still unclear. Because the flavin semiquinone is non-covalently bound, we hypothesise that stabilisation of the semiquinone state prevents the covalent flavinylation of VAO by excluding the formation of the reduced flavin and flavin iminoquinone methide that have been proposed to act as intermediates in the covalent flavinylation process [214, 235]. It is remarkable that the semiquinone state of FAD persists despite the apparent oxygen consumption and hydrogen peroxide formation in reaction experiments. It appears that *Ps*VAO_K suffers from oxidative stress because substrate binding stimulates single electron flavin redox cycling, which uncouples oxygen reduction from aromatic product formation.

Methods

Chemicals. Vanillyl alcohol, eugenol and horseradish peroxidase (type II) were from Sigma-Aldrich. Peptone was from Sigma-Aldrich (peptone from casein, pancreatic digest; for microbiology) and tryptone was from Duchefa Biochemie. Catalase from beef liver and superoxide dismutase from bovine erythrocytes were from Boehringer Mannheim GmbH. All other chemicals were from commercial sources and of the purest grade available. Growth media used were: Yeast Extract Peptone Dextrose Medium (YPD) (1% yeast extract, 2% peptone, 2% glucose, 100 μg/mL Zeocin), Buffered Glycerol-complex Medium (BMGY) and Buffered Methanol-complex Medium (BMMY) (1% yeast extract, 2% peptone, 100 mM potassium phosphate, pH 6.0 or pH 7.0, 1.34% Yeast Nitrogen Base, 4 x 10^{-5} % biotin, 1% glycerol or 0.5respectively). In BMMY, 2% peptone was replaced with 2% tryptone or 0.1 mM FAD was added when testing different growth conditions.

Strains, expression of synthetic genes and protein purification. Komagataella phaffii, NRRL Y-48124 X33 strain, as obtained from the EasySelect™ *Pichia* Expression Kit (Invitrogen) was used for extracellular expression of CRG90285, M5FTW7, C7Z853, W9CKB3, V5FPK0, K0C1W8 and P56216 (VAO from *P. simplicissimum*). Sequences with a longer N-terminus than in VAO from *P. simplicissimum* were shortened and made to contain the same N-terminal part as VAO from *P. simplicissimum*. Codon optimised synthetic genes for these sequences in pPICZαA vector were ordered from Life Technologies. Transformation of *K. phaffii* NRRL Y-48124 X33 strain was performed following the EasySelect™ *Pichia* Expression Kit instructions.

Extracellular expression using the α-factor signal sequence: this sequence is N-terminal and marks the produced enzyme as a target for secretion to the outside of the cell.

Growth of *K. phaffii* X33 and expression of VAO-like enzymes was as follows: a pre-pre culture of transformed X33 cells in 5 mL YPD medium was grown overnight in a 50 mL tube, then transferred into a 125 mL baffled flask with 20 mL BMGY medium (pre-culture), and grown overnight. The pre-culture was then transferred into a 1 L baffled flask with 200 mL BMMY medium and grown for four days, with methanol being added to a final concentration of 0.5% twice a day. 50 mL of cultures that showed VAO- activity with the horseradish peroxidase-coupled activity assay (HRP assay) described below were transferred into 2 L baffled flasks with 500 mL BMMY medium and grown for another 5-6 days, with methanol being added to a final concentration of 0.5% twice a day. A negative control of non-transformed cells (X33 as present in the EasySelect™ *Pichia* Expression Kit from Invitrogen) was included and found to show no activity with the HRP assay described below. Also, no intracellular expression of VAOs was observed in samples where extracellular expression of VAOs occurred.

Purification of enzymes produced in *K. phaffii* was performed as follows (after [224]): The culture supernatants were harvested by centrifugation (30 min, 4400 g, 4 °C). Subsequently, ammonium sulphate was added to 65% saturation at 4 °C and then the precipitate was removed by centrifugation (30 min, 4400 g, 4 °C). The pellet was dissolved in 50

mM potassium phosphate buffer, 0.5 M ammonium sulphate, pH 7.0 and loaded onto a phenyl-Sepharose column pre-equilibrated with the same buffer. The enzymes were eluted with a linear descending gradient from 0.5 M to 0 M ammonium sulphate in 50 mM potassium phosphate buffer, pH 7.0 using a Pharmacia Biotech Äkta system. Active fractions, as determined using the HRP assay described below, were pooled, dialysed against 50 mM potassium phosphate buffer pH 7.0, containing 150 mM KCl and concentrated using Amicon 10kDa spin-filters. Samples were then applied to a Superdex 200 10/30 GL column (GE Healthcare) running in 50 mM potassium phosphate buffer pH 7.0, containing 150 mM KCl, at a flow rate of 1.0 mL/min. Fractions containing pure enzymes, as determined by SDS-PAGE, were pooled, concentrated and dialysed to a 50 mM potassium phosphate buffer pH 7.5 with 10% glycerol and stored at -80 °C.

Analytical gel filtration was performed at room temperature using a Superdex 200 10/30 GL column (GE Healthcare) running in 50 mM potassium phosphate buffer pH 7.0, containing 150 mM KCl, at a flow rate of 1.0 mL/min. Analysis of the oligomeric state of the VAO-like enzymes was done as described elsewhere [227].

Biochemical characterisation and spectral properties. A horseradish peroxidase-coupled activity assay (HRP assay) was used to screen for presence of VAO-like enzymes during production and purification. Reaction mixtures containing 3 U/mL horseradish peroxidase, 0.1 mM 4-aminoantipyrine, 1 mM dichlorohydroxybenzenesulfonic acid, 0.2 mM vanillyl alcohol and an appropriate amount of enzyme sample in 50 mM potassium phosphate buffer, pH 7.5, were incubated at room temperature for at least 15 min. The appearance of the typical pink colour upon occurrence of a reaction was observed by eye. All activity measurements with the purified enzymes were performed with previously frozen enzymes. Without the addition of glycerol to 10% before freezing the enzymes, it was found that the activity detected in the HRP assay was lost.

Specific activity was measured by following the conversion of vanillyl alcohol to vanillin or the conversion of eugenol to coniferyl alcohol by monitoring the absorption of the products at 340 nm (ε_{340} = 14 mM $^{-1}$ cm $^{-1}$ at pH 7.5) and 296 nm (ε_{296} = 6.8 mM $^{-1}$ cm $^{-1}$ at pH 7.5) respectively. Reaction mixtures containing 0.26 μM of VAO-like enzyme and 0.5 mM vanillyl alcohol or 0.05 mM eugenol in 50 mM potassium phosphate buffer, pH 7.5 were incubated at 25 °C in quartz cuvettes. Production of vanillin or coniferyl alcohol was observed using a Hewlett- Packard HP 8453 diode array spectrophotometer.

Oxygen consumption was monitored using a Hansatech Oxytherm system (Hansatech Instruments, King's Lynn, UK). Reactions were conducted at 25 °C with 0.2 mM vanillyl alcohol in 50 mM potassium phosphate buffer pH 7.5, and enzyme added to a final concentration of 0.4 μM VAO or 1.2 μM VAO_{SQ}. Catalytic amounts of catalase, superoxide dismutase or horseradish peroxidase were added immediately after addition of VAO or VAO_{SQ} to the reaction mixture or towards the end of the reaction.

Flavin absorption spectra of the enzymes were recorded using a Hewlett-Packard HP 8453 diode array spectrophotometer and quartz cuvettes, or with a Thermo Scientific NanoDrop 2000c UV-Vis spectrophotomer. Protein concentrations of the purified enzymes in a potassium phosphate buffer pH 7.5 were determined using absorption values at 439 nm and ε_{439} =12.5 mM $^{-1}$ cm $^{-1}$ for VAO_{OX} and assuming ε_{439} =7.8 mM $^{-1}$ cm $^{-1}$ for VAO_{SQ}, respectively.

Reduction of the flavin cofactor was attempted by incubating 27 µM of enzyme with 10 mM dithionite or 30 mM cyanoborohydride in 50 mM potassium phosphate buffer, pH 7.5 at room temperature for at least 1 h. After the incubations, the redox state of the flavin cofactors was determined from their absorption spectra. To force the release of their flavin cofactor, enzymes were denatured using trichloroacetic acid (TCA) precipitation. For the TCA precipitation, cold 40% TCA was added to the enzyme sample to a final concentration of 5% on ice. Samples were centrifuged at 4 °C, 18,000 g for 10 min to separate the denatured protein from the released flavin cofactor. Absorption spectra of neutralised (pH 7.0, using NaOH) supernatants were recorded as described above.

Identification and quantification of FAD using HPLC/MS. To determine the identity of the flavin cofactor, it was separated from the enzymes by heat denaturation. Samples were incubated for 10 min at 95 °C, followed by removal of the protein by centrifugation (18,000 g for 10 min). FAD was identified and quantified using a Shimadzu UPLC-triple quad mass spectrometer (LCMS-8040) using the following procedure: The purified flavin cofactor from VAO_{SQ} was freeze-dried and dissolved in 120 µl H_2O. 2 µL of this cofactor solution were injected onto a Discovery HS F5-3 (2.1 mm I.D. x 150 mm, 3 µm particles) from Sigma Aldrich. Separation was performed at 40 °C with a gradient from 100% H_2O (with 0.1% formic acid) to 35rate of 250 µL/min. The triple quad mass spectrometer was operating with 3 L/min nebulizing gas flow and 15 L/min drying gas flow. The inlet temperature was set to 250 °C and the heat block temperature to 400 °C. Electrospray ionisation was used. FAD was detected in positive mode for the transition from 786.15 to 136.10 m/z and from 786.15 to 348.10 m/z at 8.53 min.

Bioinformatics. Novel, putative *vao* genes were identified through pblast using P56216 as query sequence in the non-redundant sequence database. Sequences were aligned using clustal omega (`http://www.ebi.ac.uk/Tools/msa/clustalo/`) and the obtained alignment was visualised with Jalview [236]. Six novel sequences were selected based on presence or absence of residues corresponding to positions 61, 108, 170, 220-235, 238, 422, 457 and 502-505 in P56216. Visualisation of these positions out of context of the alignment was performed using R version 3.2.2 (2015-08-14) [226]. A phylogenetic tree was created with the Neighbour Joining Clustering method using percentage of sequence identity option in Jalview and visualised using Dendroscope [236, 237]. All final graphics were created using Inkscape (`https://inkscape.org/en/`).

Sequence identifiers (often abbreviated as SeqID) such as P56216 are unique and assigned to any sequence deposited in one of the many available sequence databases. Different databases assign different formats of sequence identifiers: P56216 is a sequence found in the UniProt database. P56216.1 would indicate that this is version one of this sequence. This becomes especially important if more than one version of the sequence is available. Corrections of a sequence will lead to it becoming version 2 of the same sequence, resulting in a hypothetical "P56216.2" sequence identifier.

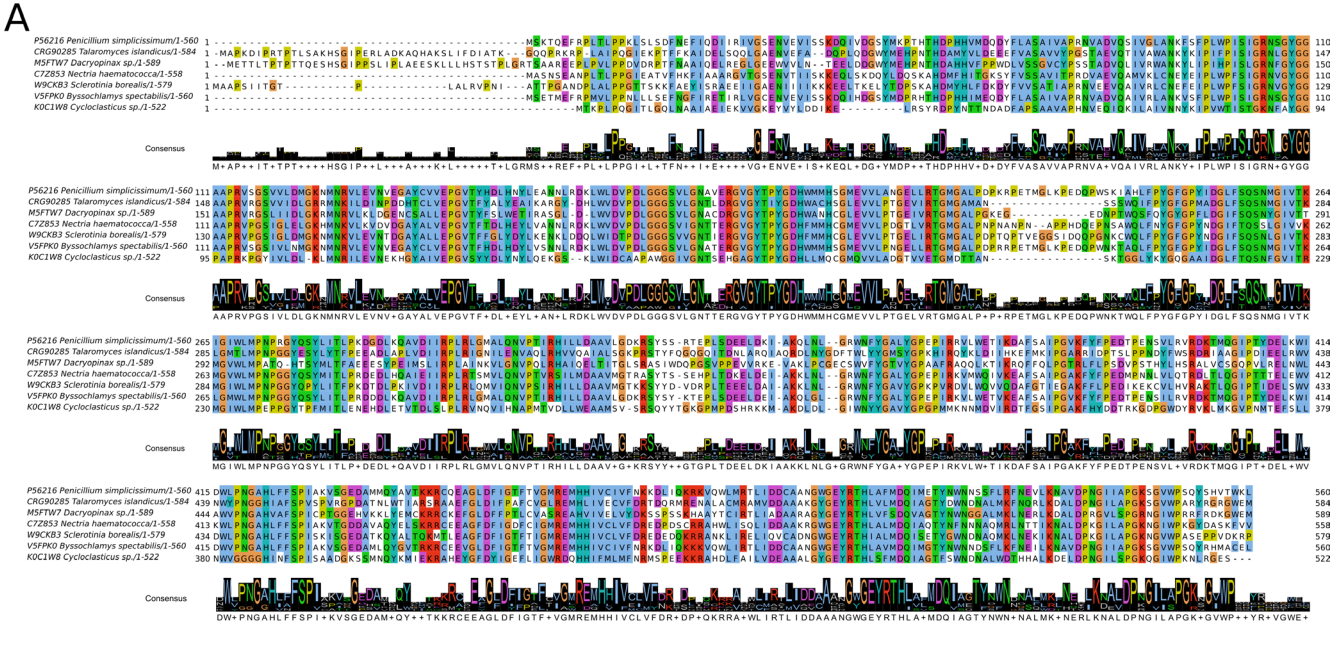
Acknowledgements

We thank Tom A. Ewing, Prof. Jacques J. M. Vervoort and Adrie H. Westphal for discussions and assistance in Oxytherm and MS experiments. This work was supported by the European Union through the INDOX project (FP7-KBBE-2013-7-613549).

Bibliography follows at the end of the thesis

Supplementary figures

Supplementary Figure 2: Bioinformatic analysis of selected sequences: (A) Multiple sequence alignment, showing consensus sequence. Note that a proposed variant of the PTS1 sequence, WKL-COOH instead of SKL-COOH (as described in [29]), is only present in P56216 but not in the other, novel *vao* genes. (B) Phylogenetic tree of sequences. Branch length is proportional to the number of substitutions, as indicated by the bar (10 substitutions per sequence position).
The figure follows on page 117.


A
P56216 Penicillium simplicissimum/1-560
CRG90285 Talaromyces islandicus/1-584
M5FTW7 Dacryopinax sp./1-589
C7Z853 Nectria haematococca/1-558
W9CKB3 Sclerotinia borealis/1-579
V5FPK0 Byssochlamys spectabilis/1-560
K0C1W8 Cycloclasticus sp./1-522
Consensus
B
K0C1W8 Cycloclasticus sp.
M5FTW7 Dacryopinax sp.
CRG90285 Talaromyces islandicus
C7Z853 Nectria haematococca
W9CKB3 Sclerotinia borealis
P56216 Penicillium simplicissimum
V5FPK0 Byssochlamys spectabilis
10

Chapter 5:
The ins and outs of vanillyl alcohol oxidase: identification of ligand migration paths

This chapter is based on:
G. Gygli, M. F. Lucas, V. Guallar and W. J. H. van Berkel. *PLOS Comput. Biol.*, 13(10):1–27, 2017.

Keywords: Ligand diffusion, flavoprotein, oxygen reactivity,
Protein Energy Landscape Exploration (PELE), computational enzymology

Abstract

Vanillyl alcohol oxidase (VAO) is a homo-octameric flavoenzyme belonging to the VAO/PCMH family. Each VAO subunit consists of two domains, the FAD-binding and the cap domain. VAO catalyses, among other reactions, the two-step conversion of *p*-creosol (2-methoxy-4-methylphenol) to vanillin (4-hydroxy-3-methoxybenzaldehyde). To elucidate how different ligands enter and exit the secluded active site, Monte Carlo based simulations have been performed.

One entry/exit path via the subunit interface and two additional exit paths have been identified for phenolic ligands, all leading to the *si* side of FAD. We argue that the entry/exit path is the most probable route for these ligands. A fourth path leading to the *re* side of FAD has been found for the co-ligands dioxygen and hydrogen peroxide. Based on binding energies and on the behaviour of ligands in these four paths, we propose a sequence of events for ligand and co-ligand migration during catalysis. We have also identified two residues, His466 and Tyr503, which could act as concierges of the active site for phenolic ligands, as well as two other residues, Tyr51 and Tyr408, which could act as a gateway to the *re* side of FAD for dioxygen.

Most of the residues in the four paths are also present in VAO's closest relatives, eugenol oxidase and *p*-cresol methylhydroxylase. Key path residues show movements in our simulations that correspond well to conformations observed in crystal structures of these enzymes. Preservation of other path residues can be linked to the electron acceptor specificity and oligomerisation state of the three enzymes. This study is the first comprehensive overview of ligand and co-ligand migration in a member of the VAO/PCMH family, and provides a proof of concept for the use of an unbiased method to sample this process.

Introduction

Vanillyl alcohol oxidase (VAO) is a flavoenzyme involved in the mineralisation of aromatic compounds in the soil fungus *Penicillium simplicissimum*. VAO is active with a wide range of *para*-substituted phenolic compounds, but received its name because the enzyme initially was found to catalyse the oxidation of vanillyl alcohol (4-hydroxy-3-methoxybenzyl alcohol) to vanillin (4-hydroxy-3-methoxybenzaldehyde) [31]. The natural substrate of the enzyme could be 4-(methoxymethyl)phenol, as it is the only substrate known to induce VAO expression [35]. VAO is one of the designating members of the VAO/PCMH family of oxidoreductases that share a conserved FAD-binding domain [179, 196, 198]. Members of this family often bind their flavin cofactor in a covalent mode. The closest known and characterised relatives of VAO in the VAO/PCMH family are the bacterial enzymes eugenol oxidase (EUGO) and *p*-cresol methyl-hydroxylase (PCMH). While substrate specificities of these three enzymes differ, they all require their substrates to be *para*-substituted phenols [36, 40, 55, 56, 238]. The amino acid sequence of VAO is most similar to EUGO, sharing 47% sequence identity, while VAO and PCMH share 40% sequence identity. Interestingly, the sequence identity between the two bacterial enzymes is only 33%. All three enzymes bind their FAD cofactor covalently. In VAO, this bond is formed via an autocatalytic process, resulting in a covalent link between His422 and FAD [41]. In EUGO, the flavin is linked to His390 [202], while in PCMH, the flavin is covalently bound to Tyr384 [239].

VAO is a homo-octamer composed of a tetramer of dimers, where each subunit consists of 560 amino acids. Each subunit is organised into two domains, the FAD-binding domain and the cap domain [40], finding the active site at their interface (Figure 26). In vitro, VAO

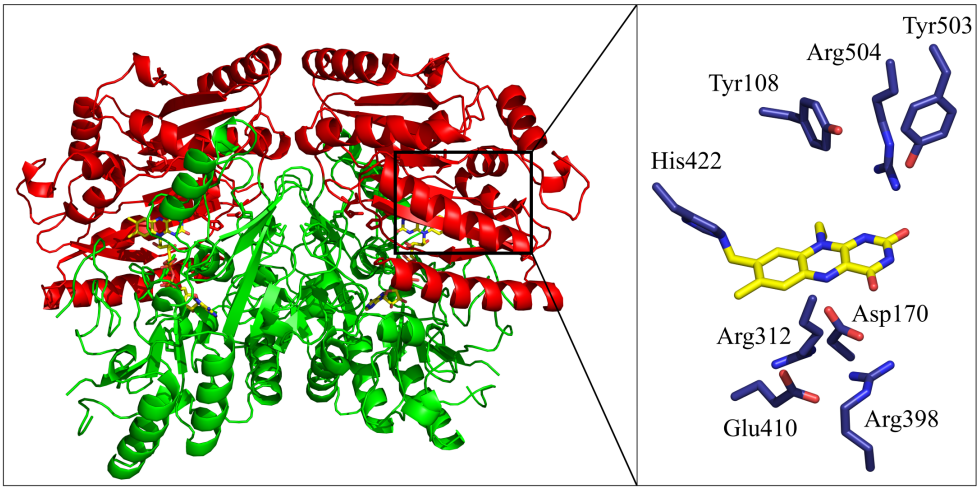


Figure 26: Three-dimensional structure of vanillyl alcohol oxidase (PDB ID: 1VAO). The dimer is shown as cartoon with the cap domain coloured in red and FAD-binding domain in green. A zoom into the active site at the interface of the two domains is shown with selected residues shown as blue sticks and the isoalloxazine ring of the covalently bound flavin cofactor in yellow.

forms a mixture of dimers and octamers, with the octamers prevailing at physiological ionic strength [38, 227]. EUGO and PCMH have different quaternary structures. EUGO is exclusively homodimeric, while PCMH consists of a heterotetramer. This heterotetramer consists of two FAD-binding subunits forming a homodimer and two cytochrome c subunits. The homodimeric structure of the FAD-binding subunits is the same in all three enzymes. See also Supplementary Figure 3 for an overview of the quaternary structures of these three enzymes.

Important active site residues of VAO are Tyr108, Asp170, His422, Tyr503 and Arg504 (see the zoom into the active site in Figure 26). Tyr108, Tyr503 and Arg504 are proposed to be crucially involved in the deprotonation of substrates upon their arrival in the active site [40, 49]. Asp170 has a multifunctional role. It is involved in the autocatalytic incorporation of the flavin cofactor and assists in increasing the redox potential of VAO (making it less negative) [43]. Asp170 can also act as active site base [43], and is involved in the enantioselective hydroxylation of 4-

alkylphenols [46]. His422 increases the oxidation power of the FAD cofactor by increasing its redox potential (making it less negative) through covalent flavin binding [41, 44].

The catalytic mechanism of VAO involves two half-reactions. In the reductive half-reaction, the enzyme-bound flavin is reduced by the phenolic substrate to generate a quinone methide intermediate [36, 223]. In the oxidative half-reaction, the reduced flavin is oxidised by dioxygen, generating hydrogen peroxide. Depending on the substrate, there are differences in the reaction mechanism. With 4-(methoxymethyl)phenol, water addition to the quinone methide intermediate gives an unstable hemiacetal, which decomposes to 4-hydroxybenzaldehyde. VAO can produce vanillin in a one-step reaction from vanillyl alcohol as well as in a two-step reaction from *p*-creosol, where vanillyl alcohol is the decomposition product of the initially formed air-stable flavin-creosol adduct [178]. The reaction with *p*-creosol (Figure 27, plates 1 to 5) is rate-limited by the extremely slow decomposition of the flavin-N5 substrate adduct (Figure 27, plate 4) [40]. In the reaction of vanillyl alcohol to vanillin, there is no addition of water to the vanillyl alcohol quinone methide (Figure 27, plates 6 to 8) [223].

Although many details of the catalytic mechanism of VAO have been uncovered, it is unknown how the reaction participants enter and exit the active site. No path for solvent or ligand access to the active site is visible from the crystal structure of VAO. It is of fundamental interest to understand how reaction participants enter an enzyme's active site to be able to identify catalytic bottlenecks. Such bottlenecks can then be used to guide substrate or enzyme redesign to eliminate them. This work provides a proof of concept for the use of an unbiased method to sample entry and exit pathways of a complex enzyme system. We aim to discover the migration path(s) of reaction

"Catalytic" bottlenecks may not be the best choice of words here, as ligand binding normally does not affect catalysis: most enzymes are not diffusion limited and therefore ligand migration to the active site happens much faster than catalysis itself. To speak of catalytic bottlenecks implies that the overall enzyme reaction is limited by ligand migration, which of course can be the case.

A bottleneck can have steric and/or electrochemical properties. A steric bottleneck is created if a large amino acid blocks a passage by its presence. An electrochemical bottleneck is created if a charged amino acid blocks a passage through the action of its charge. For example, positively charged amino acids would present a bottleneck for positively (or negatively) charged ligands by repelling (attracting) them and thus making passage difficult.

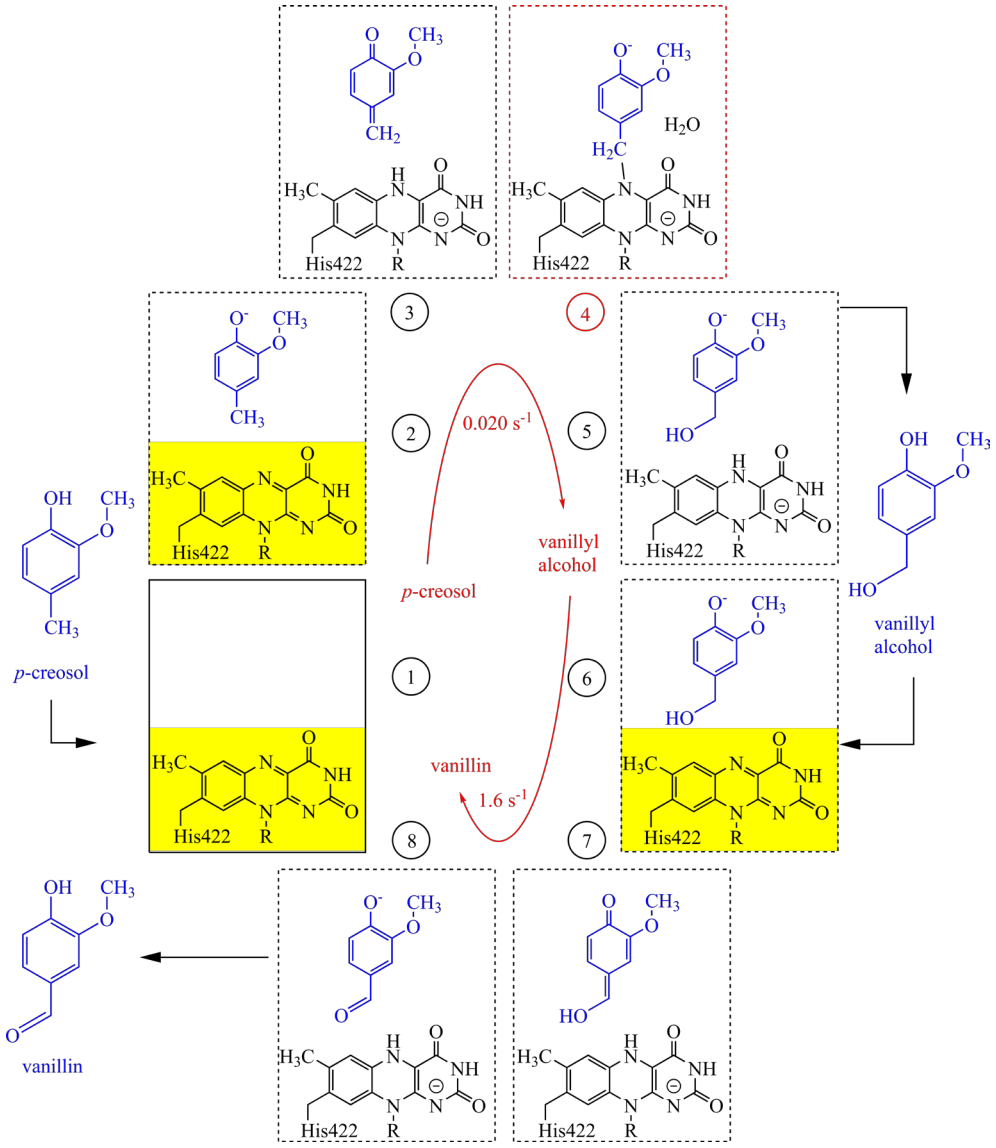


Figure 27: Overview of the reaction cycle from *p*-creosol to vanillyl alcohol and vanillin catalysed by VAO. Reaction rates at pH 7.5, 25 °C are shown in the centre of the reaction cycle (data obtained from [45]). The rate-limiting step for the conversion of *p*-creosol to vanillyl alcohol is the decomposition of the air-stable flavin N5 substrate adduct (panel 4, marked in red). For the conversion of vanillyl alcohol to vanillin the rate-limiting step is the reduction of the flavin by the substrate (panel 7). In both reactions, the yellow flavin cofactor is rendered colorless through reduction by the substrate (occurring between panels 2 and 3 as well as 6 and 7). Molecular oxygen driven oxidation of reduced flavin (colourless) to oxidized flavin (yellow) occurs between panels 5 and 6 as well as panels 8 and 1, in the presence or absence of either vanillyl alcohol or vanillin.

partners of VAO for the reaction from *p*-creosol to vanillin. Recent studies, with other enzymes, have pointed to the importance of remote residues for catalysis or the presence of multiple paths for ligands [240, 241, 242, 243, 244]. It is also worthy to note that during enzyme catalysis, the entry and exit of reaction partners may result in traffic jams in migration paths. We try to address these issues by modelling migration paths of reaction partners of VAO. Given the large size of VAO and the involvement of multiple ligands, it is conceivable that substrates and products use different paths.

In this study, we used Protein Energy Landscape Exploration (PELE), a Monte Carlo based program [67], to investigate the entry and exit paths of three phenolic ligands in VAO. Protonated and deprotonated phenolic ligands were used to be able to observe possible differences in binding energies during the simulations and thus establish when substrate deprotonation and product re-protonation happen in the reaction pathway of VAO.

Allostery is normally defined as the regulation of protein function through a change in its quartenary structure, most often through its interaction with a ligand or an additional protein subunit. Recently, it has been suggested that all proteins are allosteric [245] and that allostery and catalysis are evolutionary linked [246].

An entry and exit path for phenolic ligands at the subunit interface, along with two additional exits paths, leading through the cap domain or the FAD-binding domain, were identified. It was found that protonated *p*-creosol freely migrated to the active site, while deprotonated *p*-creosol did not. In addition, the behaviour of dioxygen and hydrogen peroxide was modelled. Migration via the flavin *re* side turned out to be the shortest path for these molecules.

Results

Ligand migration simulations using PELE were run on the dimer of VAO with phenolic ligands COP, COD, VAP, VAD, VNP and VND, as well as with dioxygen and hydrogen peroxide. Figure 28 illustrates the location in VAO of all the

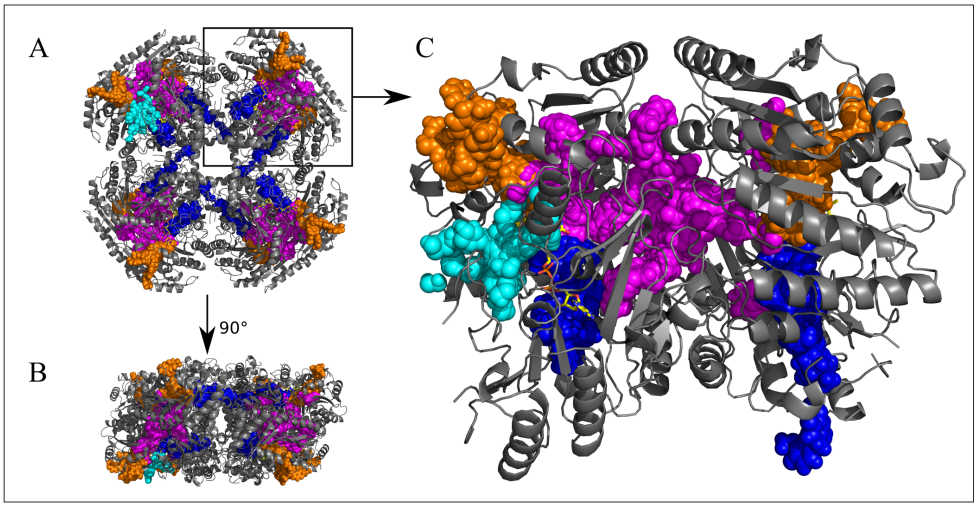


Figure 28: Three-dimensional structure of vanillyl alcohol oxidase and the four paths identified in this study. Vanillyl alcohol oxidase is shown as cartoon and coloured in grey, with FAD shown as sticks and coloured in yellow (PDB ID: 1VAO). The paths identified in this work are shown through sphere representation of ligand positions in the trajectories. The cap path is coloured in orange, the FAD path in blue, the subunit interface path in magenta and the re path in cyan.
A: Octameric VAO, seen from the front.
B: Octameric VAO seen from the side.
C: Dimeric VAO as seen in Figure 26.

paths identified through the approaches described in Materials and Methods.

Due to the stochastic nature of the PELE method, it is not possible to correlate the observed movements to biological timescales. Instead, the calculated binding energies (presented as interaction internal energies, see Methods), at different distances of the ligand from the FAD as well as the number of contacts between ligands and protein residues, were used to estimate their most likely paths as they move through the protein. Furthermore, we computed the flexibility of each protein residue during the simulations through RMSFs. This data can also be used to estimate if a path involves bottlenecks where the ligands tend to get stuck.

Description of the entry path used by phenolic ligands Using the adaptive PELE simulations, the surface of the VAO dimer was explored for COP entry, see Figure 29A. As stated, the ligand was placed in the bulk solvent and allowed to freely explore the surface and any possible en-

try path(s). It was found that using this approach, the ligand moved into the active site exclusively via the subunit interface. In order to refine the energy landscape along the entry path, we started additional simulations from the surface of the subunit interface with a reduced translation range (maximum ligand translation of 1.5 Å). These refinement simulations confirmed the entry for COP, but indicated that COD was unable to enter the protein via this path. Binding energies calculated showed that it was energetically favourable for COP to reach the active site via the subunit interface, but energetically unfavourable for COD to do the same (Figure 29C). The COP entry path, which is identical to the third exit path described below, involves mainly residues Arg183, Val185, Asp192, His193, Met195, Met303, Ile428, Met462, His466 Ile468, Tyr503 and Arg504. Residues His466 and Tyr503 are easily changing conformation to grant the ligand access to the active site. This path is symmetrical, leading from the starting point at the subunit interface to either of the two active sites of the dimer, as can be seen in Fig 4C. The most contacted residues were His193, Met195 and Ile428 and the most flexible residues were Asp192, Gly462, His466 and Tyr503, as can be seen in Fig 4B.

Overview of the three exit paths used by phenolic ligands (cap, FAD and subunit interface path) To identify exit paths, simulations were performed for the dimeric protein and the ligands were placed into one of the two subunits and both subunits were used as starting points in equal amounts. We observed symmetry in our data (see Figure 28): some paths were fully symmetrical in simulations starting from either subunit and other paths were partly symmetrical, most likely due to incomplete sampling as can be seen from Figures 28 and 29A. The fact that we observed identical results independent of the starting point is a good indication that all potential paths have been sampled. Extrapolating data from each of the two sub-

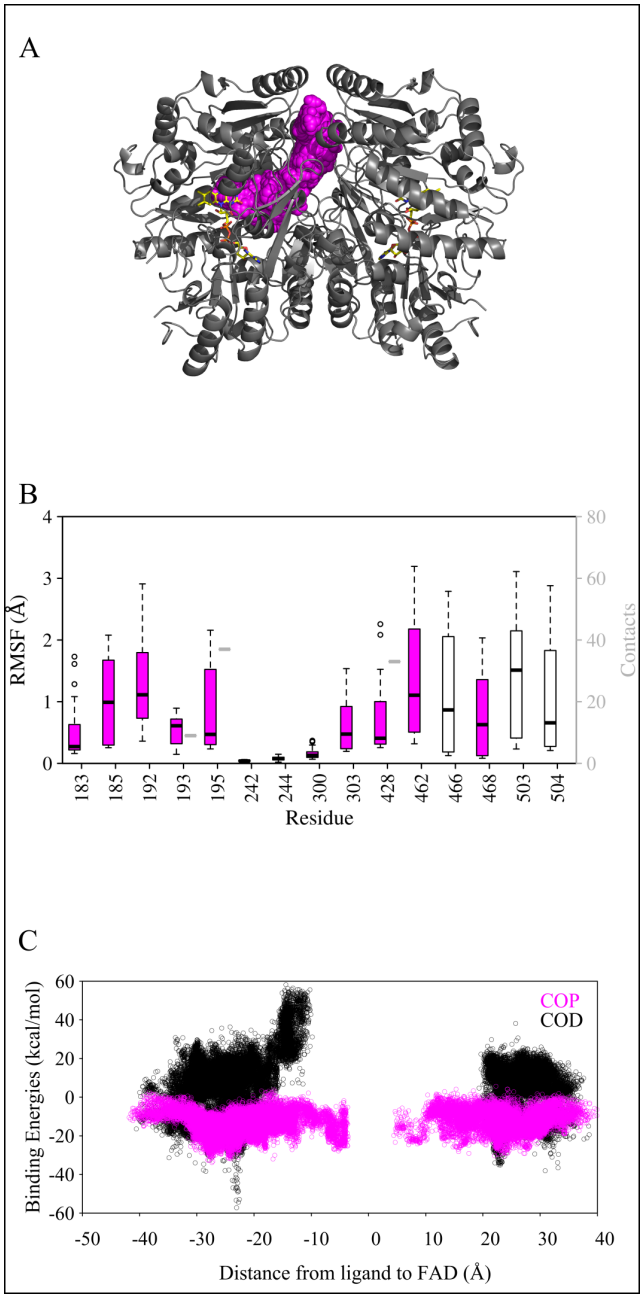


Figure 29: Overview of the entrance path identified with COP (A) as well as RMSF and atom contacts (B, for COP) and binding energy data (C, for COP and COD).
A: The VAO dimer is shown as cartoon in grey and its FAD cofactor is shown as sticks in yellow. The entrance path is shown through sphere representation of ligand positions in the trajectories and coloured in magenta. For clarity, we only show data for an entrance simulation where COP migrates to the left-hand subunit.
B: Highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Magenta filling of the boxplots indicates that the residue belongs to the entrance path. Residues that do not have atom contacts with the ligand but are highly flexible only have a black boxplot (for RMSF). Because data was obtained from only one simulation, the atom contacts do not have any variation.
C: Scatterplot of binding energies as a function of the distance between the ligand and FAD. Negative x-values indicate the distance to the FAD in the other subunit. This graph shows that the path is symmetrical for COP (in magenta).

units to the other subunit gave fully symmetrical paths, which were continuously connected.

Three independent exit paths were observed for simulations with phenolic ligands, as shown in Figure 28 and parts A of Figures 29, 30 and Supplementary Figures 4 and 5. The paths shown in different colours correspond to i) the cap path (in orange) where ligands exit the protein through the cap-domain, ii) the FAD path (in blue) where ligands traverse the protein through the FAD-binding domain, and iii) the subunit interface path (in magenta), which corresponds to a path connecting the subunits (and agreeing with the entry path described above).

Not all simulations led to the ligand leaving VAO. Some simulations showed that the ligand also moved from one path to the other, resulting in mixed paths. Only one of the three paths, the cap path, is exclusively connecting the active site to the solvent. The other two paths identified, the FAD path and subunit interface path, lead to interfaces between subunits or dimers of VAO. Paths leading to the interface of subunits or dimers allow the enzyme on one hand to connect the different active sites of the subunits (Figure 28 and parts A of Figures 29, 30 and of Supplementary Figures 4 and 5). On the other hand, these paths also allowed the ligands to exit via the interface of the dimer or the subunit, as there is sufficient space for the ligands available there. Because these paths can be expected to work into both directions, in and out, they could also act as another access point for the ligands to the active site.

However, analysis of residue flexibility represented by RMSF data as well as atom contact counts has to be taken into account to discover possible bottlenecks in these paths. The RMSF values and atom contacts calculated per simulation were combined and plotted in boxplots (parts B and C of Figures 29, 30, and of Supplementary Figures 4 and 5). The more flexible residues, mostly lo-

RMSF stands for Root-Mean-Square Fluctuations of atomic positions and measures the distance between atoms of a protein relative to an average structure.

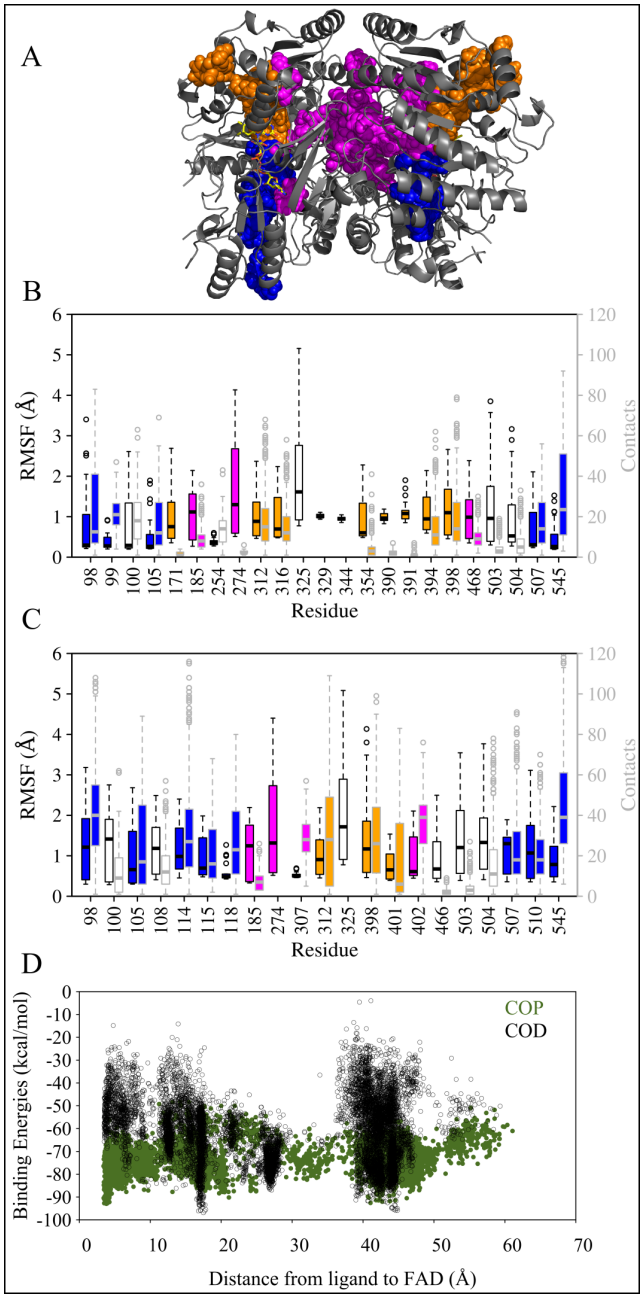


Figure 30: Overview of the exit paths identified with COP and COD (A), RMSF and atom contacts (B for COP and C for COD) and binding energies (D).
A: The VAO dimer is shown as cartoon in grey and its FAD cofactor is shown as sticks in yellow. The paths identified in this work are shown through sphere representation of ligand positions in the trajectories. The cap path is coloured in orange, the FAD path in blue and the subunit interface path in magenta.
B: Plot of simulations with COP, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
C: Plot of simulations with COD, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
D: Scatterplot of binding energies as a function of the distance between the COP and COD (in green and black) to FAD.

cated at the surface, were often also the most contacted ones. Those most flexible surface residues were Arg274 and Arg325, while the buried ones showing larger flexibility were Arg398, Tyr503 and Arg504. Residue Lys545 was the most contacted residue for all ligands except VNP. Other flexible and/or highly contacted residues in simulations with most ligands were Trp98 (not with VNP), Ile100 and Asn105 (not with VAD and VNP) Val185 (not with VAP), Arg312 (not with VAD and VNP), surface residue Ser329 (not with COD) and Leu507 (not with VNP).

The binding energies of all the phenolic ligands were on comparable scales, but the range of the calculated binding energies varied between ligands, as can be seen in Supplementary Figure 4C. The binding energies plotted against the distance of the ligand from the FAD spread least towards higher energy values for COP and VAP compared to COD and VAD. When comparing the data from the entry path and the exit paths, COP showed very little spreading of the data and presents lower binding energies than COD in both cases. Combining this information suggests that it is energetically more favourable for protonated substrates to migrate into and out of VAO than for deprotonated substrates or products.

In the following sections, each path and the behaviour of path residues will be described individually.

Detailed description of the cap path In this path, ligands migrated to the "upper-part" of the active site (orange in Figures 28, 30 and Supplementary Figures 4 and 5), passing Glu410, Trp413, Arg312 and Arg398 to exit the protein in the cap domain. Arg398 and Trp413 have been proposed to be involved in a size-exclusion mechanism that could limit the size of the substrate-binding pocket [40]. The cap path requires the ligands to pass Arg312 and Arg398, which are among the most frequently contacted and most flexible residues for most ligands. Arg398 inter-

acts with Glu410 and Trp413 and it is this interaction that the ligands need to disrupt to exit the active site via this route. Simulations often showed ligands to be trapped between the two arginines as they tried to pass this portal. This path was the shortest of the three identified for the phenolic ligands, but contained a bottleneck composed of residues Arg312, Arg398, Glu410 and Trp413 (as can be seen in Figures 28, 29 and Supplementary Figures 4 and 5). Combining the data for all the ligands, residues belonging to the cap path that are among the most flexible or contacted residues were Leu171, Arg312, Leu316, Tyr354, Pro390, Glu391, Asn392, Val394, Arg398 and Thr401.

Detailed description of the FAD path In this path (blue in Figures 28, 30 and Supplementary Figures 4 and 5), ligands migrated through Tyr503 and Arg504 to move closely along the FAD phosphate-ribityl chain towards the adenosine monophosphate part of the cofactor and to exit the protein between Trp98, Leu507 and Lys545. Trp98 is required to move aside, which is reflected in the RMSF and contact data (as can be seen in Figure 28B and 28C in the middle panel and Supplementary Figures 4A and 5B). This path led to the dimer-dimer interface in octameric VAO, where we find sufficient space for the ligand to exit VAO or migrate to the neighbouring dimer in the octameric symmetry unit. However, Trp98, Leu507 and Lys545 constituted a bottleneck in this path. Combining the data for all the ligands, residues belonging to the FAD path that were among the most flexible or contacted residues were Trp98, Pro99, Asn105, Arg114, Val115, Ser118, Phe424, Thr505, Leu507, Met510, Lys545 and Ser546.

Detailed description of the subunit interface path In this path, ligands moved towards the subunit interface of the VAO dimer (magenta in Figures 28, 30 and Supplementary Figures 4 and 5). The behaviour of different ligands modelled varied slightly in this path and they followed different minor variations of it, however all these minor paths led to

the subunit interface. Simulations that were started with a ligand in one subunit sometimes showed the ligand completely entering the active site of the other subunit. Ligands were frequently found in the upper part of the subunit interface, where they were already surrounded with solvent. As stated, simulations mapping the entry path also showed the ligand passing through the same area.

Three residues clearly involved in the subunit interface path, but not showing up as frequently contacted or highly flexible residues in the analysis, were Asp192, Met195 and Glu464. Ligands passed quickly through a portal formed by Glu464, Asp192 and Met195, and from there directly into the subunit interface (for an illustration see Supplementary Figure 6). The portal formed by these three residues appeared to be the easiest for the ligands to pass through, not requiring significant movements of large residues and not involving a high frequency of contacts. The portal in this path connects the two active sites in the VAO dimer to each other.

Additional residues involved in this path are His466 and Tyr503, which were moving aside to grant access to the active site in the entry simulations and were also allowing ligands to exit. See Figure 31A for details on the movement of His466 and Ty503 as well as Met195 in our simulations. Residues that were in contact with ligands and that line the subunit interface were Arg463, Arg300 and Met303. Combining the data for all the ligands, residues belonging to the subunit interface path that were among the most flexible or contacted residues were Arg183, Val185, Tyr187, Trp194, His197, Tyr244, Tyr249, Ser255, Arg274, Arg300, Met303, Asn307, Thr310, Ala356, Met402, Met462, Ile468 (see also parts B and C of Figures 31 and Supplementary Figures 4 and 5).

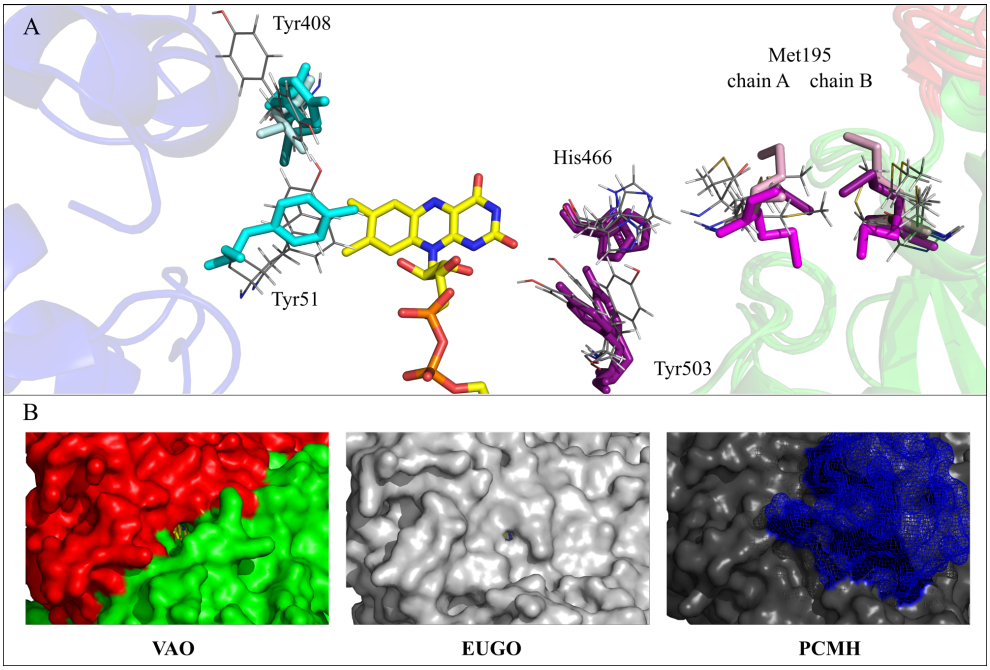


Figure 31: Details of the ligand migration paths identified in this study (A) and illustration of the surface at the *re* path in VAO, EUGO and PCMH (B).
A: Crystal structures of VAO, EUGO and PCMH as well as selected frames from our simulations were aligned using PyMOL. Selected crystal structure residues (labelled with VAO numbering) are shown as sticks. Residues involved in the subunit interface path and their conformations in VAO are coloured in magenta. The conformations of the corresponding residues in EUGO and PCMH, are coloured in lighter or darker shades of magenta, respectively. Residues involved in the *re* path and their conformations in VAO are coloured in cyan. The conformations of the corresponding residues in EUGO and PCMH, are coloured in lighter or darker shade of cyan, respectively. Note that the loop carrying Tyr51 is completely absent in EUGO and PCMH and that Tyr408 present in VAO is a leucine in EUGO and PCMH. Selected residues from selected frames of our simulations are shown as lines, in grey. Frames showing the movement of Tyr51 and Tyr408 were taken from simulations with dioxygen, those showing movement of His466, Tyr503 and Met195 from simulations with phenolic ligands that were migrating through the subunit interface path. The respective other monomer of the three crystal structures is visible on the right of the figure as cartoon in red (cap domain) and green (FAD-binding domain). On the left, the cytochrome c subunit from PCMH is visible as cartoon, coloured in blue.
B: Surface representation of the surface of the *re* path to illustrate the size of the channel leading to the *re* side of FAD in VAO, EUGO and PCMH. The proteins are shown as surfaces and the FAD cofactors as sticks, coloured in yellow. Note the different size of the channel in VAO and EUGO, as well as the blockage by the cytochrome c subunit in PCMH (blue mesh).

Detailed description of the paths used by dioxygen and hydrogen peroxide The co-substrate dioxygen and co-product hydrogen peroxide were placed on the *si* or *re* side of the isoalloxazine ring of the FAD cofactor and left to explore the VAO dimer. On the *re* side of FAD, a possible oxygen binding pocket is located as indicated by a bound chloride ion in PDB ID: 1VAO and by a water molecule in the same position in PDB ID: 2VAO [40, 247].

When placed on the *si* side of FAD, dioxygen and hydrogen peroxide behaved similarly to the phenolic ligands, but did not leave the enzyme after 200–250 simulation steps, probably due to under sampling. There is a difference in the behaviour of the co-ligands though: they both rarely used the FAD path and hydrogen peroxide had a clear preference for the cap path, almost exiting the protein when migrating through that path. When placing the co-ligands on the *re* side of FAD, they were able to leave the enzyme after 50 to 250 simulation steps via a path through the interface of the cap and FAD domain. This path will be referred to as *re* path for the remainder of this text and is illustrated in cyan in Figures 28 and 32. Contrary to the paths identified for phenolic ligands, this path is visible as an access channel to the active site from the crystal structure of VAO (see Figure 31B).

Binding energies for dioxygen and hydrogen peroxide were found to be on a different scale from those of the phenolic ligands, which is to be expected due to the difference in size of these molecules (computed binding energies are extensive properties), for comparison see parts D of Figures 29, 30 and 32. However, binding energies were comparable for these co-ligands when placed at the *si* or *re* side of FAD. In both cases, variations in binding energies for hydrogen peroxide were larger than for dioxygen, but overall binding energies were more negative for hydrogen peroxide (Figure 32D). This might suggest that hydrogen peroxide more likely gets stuck in local energy min-

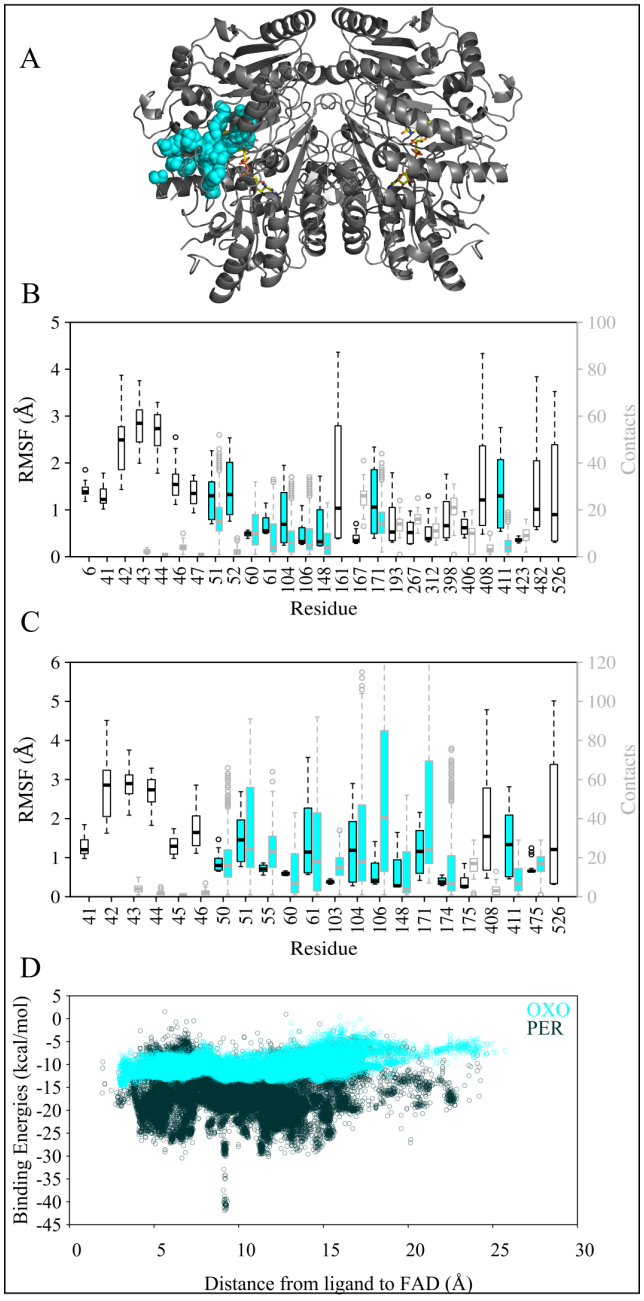


Figure 32: Overview of the *re* path identified with dioxygen (OXO) and hydrogen peroxide (PER) (A), RMSF and atom contacts (B for dioxygen and C for hydrogen peroxide) and binding energies (D).
A: The VAO dimer is shown as cartoon in grey and its FAD cofactor is shown as sticks in yellow. The paths identified in this work are shown through sphere representation of ligand positions in the trajectories. The *re* path is coloured in cyan.
B: Plot of simulations with dioxygen, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Cyan filling of the boxplots indicates that the residue belongs to the *re* path.
C: Plot of simulations with hydrogen peroxide, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
D: Scatterplot of binding energies as a function of the distance between dioxygen (cyan) or hydrogen peroxide (black) to FAD.

ima, which could mean that dioxygen reaches the active site in a faster and more directed manner than hydrogen peroxide leaves it.

Surface residues found in the analysis of RMSF and atom contacts were residues Lys43, Asp44, Ile46 and Arg482. Residues of the *re* path among the most flexible or contacted residues were Tyr51, Met52, Thr55, Pro60, His61, Gly103, Arg104, Ser106, Tyr148, Leu171 Gly174, Leu411 and Lys475 (as illustrated by Figure 32 parts B and C). Tyr51 and Tyr408 needed to move aside to allow dioxygen and hydrogen peroxide to leave the active site, a behaviour likely determining access to the flavin *re* side for these molecules. Tyr51 is also positioned close enough to the proposed oxygen binding pocket that it could be involved in FAD re-oxidation. See also Figure 31A for details of the movement of Tyr51 and Tyr408 in our simulations.

A movie illustrating all the paths is available in the supplementary data (S1 Movie).

Structural conservation of the identified migration paths in VAO, EUGO and PCMH We analysed the structural conservation of path residues in three crystal structures of VAO, EUGO and PCMH. Figure 33 shows the obtained structure-based sequence alignment and illustrates the conservation of residues in the different paths through coloured symbols. We also compared the conformations of the most conserved residues involved in the paths in the three aligned structures (indicated by star symbols in Figure 33). With the exception of Met195, Gly420, His466 and Tyr503 the conformations of all conserved residues in the crystal structures were identical.

The different conformations of these four residues will be described below. Met195 was found to have different conformations in the crystal structures of VAO, EUGO and PCMH. These conformations correspond well to the movement by Met195 observed in our simulations (see

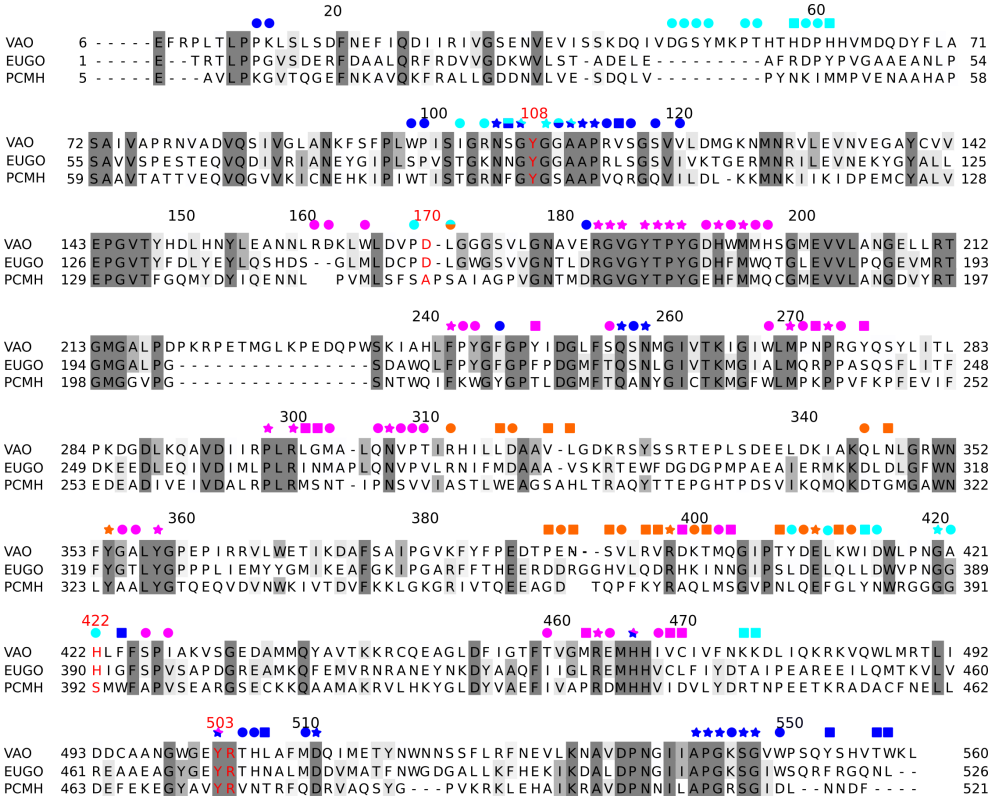


Figure 33: Structure-based sequence alignment of VAO, EUGO and PCMH. Numbering of residues at the beginning and end of each line correspond to the numbering used in the crystal structures. Numbers on top of the alignment correspond to the numbering of VAO sequence P56216.1 and is the same as used throughout this study. Coloured symbols indicate the path the residue underneath it is involved in. Orange, blue, magenta and cyan was used for the cap, the FAD, the subunit interface and the *re* path respectively. Red indicates functionally important residues. Bi-coloured symbols indicate the residue is involved in two paths. The shape of the symbol indicates preservation of the residue, stars (present in all three structures), circles (present in two of the three structures) and squares (present in only one of the three structures).

Figure 31A for an illustration). Gly420 is located in a surface loop, which is in close proximity to the cytochrome c subunit in PCMH. This loop adopts a different conformation in PCMH than in VAO and EUGO. This could be in part due to the presence of the cytochrome c subunit in PCMH, which is absent in EUGO and VAO, and in part due to the covalent histidyl linkage to the FAD cofactor in VAO and EUGO but not PCMH. His466 and Tyr503 have already been observed to have two conformations in the high resolution crystal structure of PCMH (PDB ID: 1DII [56]). The conformational freedom of these residues is not observed in the other crystal structures but is in agreement with the movements of His466 and Tyr503 observed in our simulations (see Figure 31 for an illustration). It is noteworthy that the *re* path identified in VAO leads to where the cytochrome c subunit is located in PCMH. We compared this channel in VAO, EUGO and PCMH and found that it is non-existent in PCMH and narrower in EUGO compared to VAO (see Figure 31B for an illustration).

If enough high-resolution crystal structures of an enzyme are available with and without ligands bound, comparing the conformations of these structures can also enable the identification of putative "path residues". However, if the ligand is not visible in these crystal structures the interpretation of the different conformations of amino acid residues in the structures becomes more difficult.

Several residues involved in the paths we identified in VAO are structurally conserved in EUGO and PCMH. If the paths we identified are also present in so far uncharacterised VAO-, EUGO- and PCMH-like enzymes, conservation of path residues should also be observed. This would further support the existence of the paths we have identified. We therefore performed an analysis of sequence conservation.

Sequence conservation of the identified migration paths in VAO, EUGO and PCMH We analysed sequence conservation in MSA_{VAO}, MSA_{EUGO} and MSA_{PCMH} to establish if the paths identified in this study are also conserved in so far uncharacterised VAO-, EUGO- and PCMH-like enzymes. To this purpose, we used MSA_{VAO}, MSA_{EUGO} and MSA_{PCMH}, as described in Materials and Methods. We also defined nine groups of residues to be analysed in detail: the residues in the four paths (cap, FAD, subunit inter-

face and *re* path), surface or interface residues, a random selection of residues, residues in the cap domain and the entire protein.

We found that within these three alignments, all sequences were more similar than dissimilar (see Figure 34A for all data). Residues in the cap path and at the surface were least similar and conserved in all three alignments. Residues in the FAD path and at the subunit interface were most similar and conserved in all three alignments. Note that if similarity or conservation is higher than in the entire protein or the random selection, this indicates higher selective pressure on this group of residues, and lower selective pressure if the opposite is the case.

Preservation of the groups of residues analysed in this manner should not be used as an indication of similarity or conservation within all three sets of sequences (Figure 34A). At one position in MSA_{VAO}, a residue can be conserved or similar, but a different residue can be conserved or similar in the MSA_{EUGO} or MSA_{PCMH}. To be able to determine overall preservation of residues in all three sets of sequences (MSA_{all}), we continued our analysis as described in Materials and Methods.

The results of this analysis are summarised in Figure 34B. We found that when analysing the full length MSA_{all}, 29% of all positions were more than 90% conserved, and 25% of all positions were similar in VAO and EUGO, but varied in PCMH (were PCMH-unlike). Residues in the cap path were most dissimilar, and also the most PCMH-unlike. Residues in the subunit interface path and at the subunit interface were most similar and residues in the FAD path and at the subunit interface were most conserved. It is noteworthy that all dissimilar residues in the *re* path are only dissimilar due to PCMH-like sequences.

It is conspicuous that the percentages of similarity and conservation in the two analyses differ significantly due to

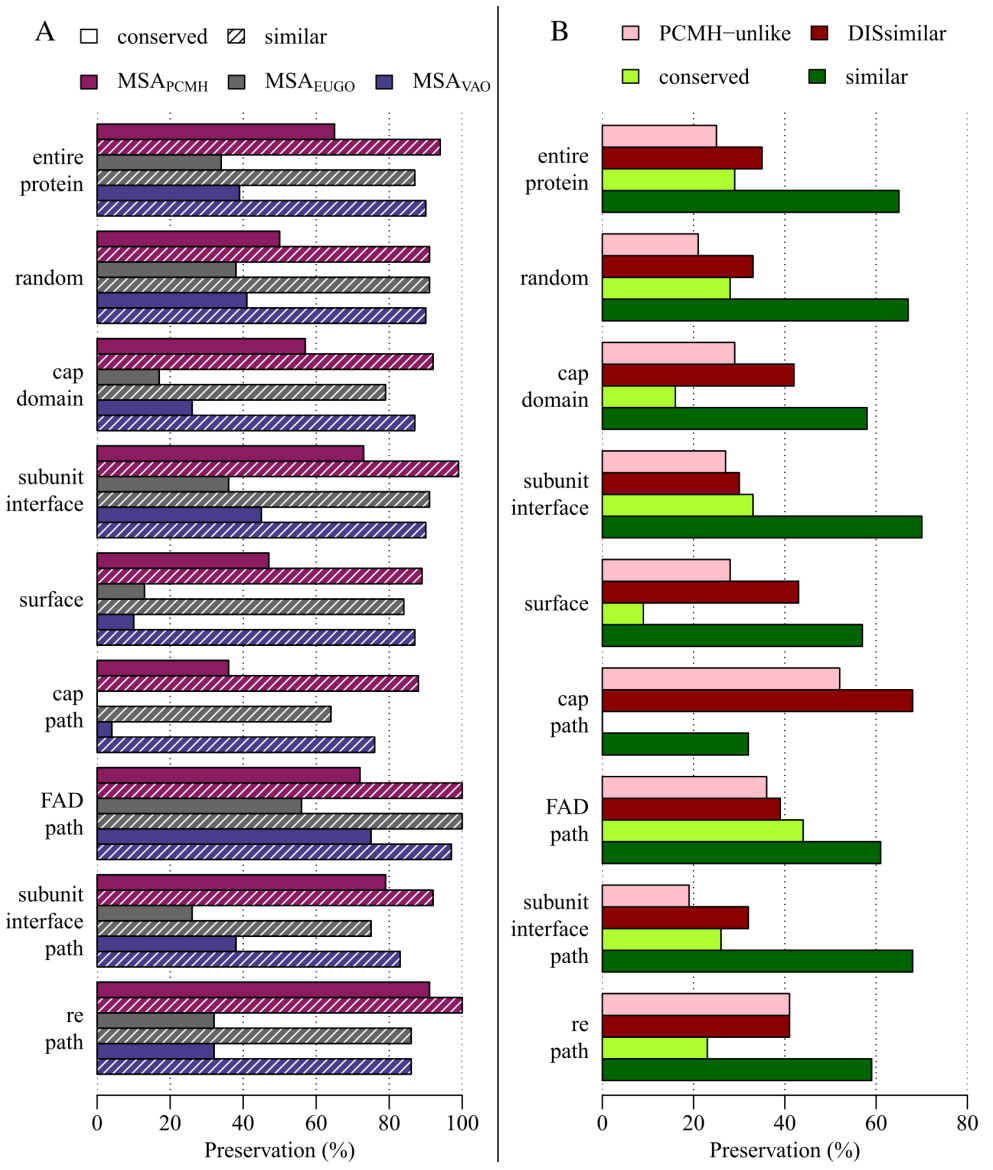


Figure 34: Sequence preservation of ligand migration paths. Preservation of residues in the entire protein, a random selection (random), the cap domain, the subunit interface, on the surface of the protein (surface), in the cap path, the FAD path, the subunit interface path and the *re* path.
A: Sequence preservation of ligand migration paths within MSA$_{VAO}$, MSA$_{EUGO}$ and MSA$_{PCMH}$.
B: Sequence preservation of ligand migration paths within MSA$_{all}$.

the preservation of different residues in VAO-, EUGO- and PCMH-like sequences. This is reflected in the percentages of dissimilarity and PCMH-unlikeness. Note the difference in preservation between residues in the cap domain and the cap path, which indicates that there is less selective pressure (or selective pressure for different residues) on residues in the cap path and the entire cap domain.

Discussion

In this study, we have identified, for the first time, ligand and co-ligand migration paths in a member of the VAO/PCMH family. The process of ligand migration to the active site is often neglected when discussing enzyme mechanisms. The advances of computational methods have made it feasible to study this process in more detail, allowing better fundamental understanding of how enzymes work. Here, we will first address the implications our findings have for our understanding of the mechanism of VAO, and then expand further to cover the close relatives PCMH and EUGO, to finally discuss what is known about (co-)ligand migration paths in the entire VAO/PCMH family.

The residues discussed below are possible targets for site-directed mutagenesis to attempt to block or open a path. The effect of such modifications on enzyme activity can then be determined experimentally. Flavoenzymes are well suited to study enzyme kinetics using stopped flow instruments due to their reaction mechanisms involving a reductive and oxidative half-reaction.

Ligand migration paths in VAO We have identified three migration paths in VAO for phenolic ligands (see Figure 35 for an overview). We do not observe any preference for a specific path for any of the three phenolic ligands analysed (*p*-creosol, vanillyl alcohol and vanillin).

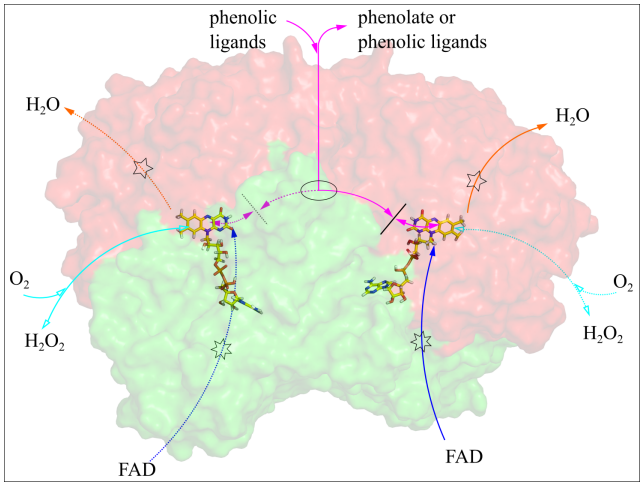


Figure 35: Overview of the identified paths and the ligands proposed to migrate through them (uni-or bi- directional). The VAO dimer is shown in surface representation, with the cap domain coloured in red and the FAD-binding domain coloured in green and the FAD cofactor shown as sticks in yellow. Coloured arrows indicate the paths for phenolic ligands (cap path in orange, FAD path in blue and subunit interface path in magenta). Filled in arrowheads indicate paths to the *si* side of FAD and empty arrowheads paths to the *re* side of FAD. Drawn out lines indicate the path is on the side of VAO facing the viewer, and dotted lines indicate the path is on the back side of VAO. Note that the two subunits are rotated 180 degrees relative to each other in the VAO dimer and that a path pointing towards the viewer in the right monomer is pointing away from the viewer in the left monomer and *vice versa*. The location of the concierge residues, His466 and Tyr503, is represented by a black line. The portal formed by Met192, Met195 and Glu464 is indicated by an ellipse. The bottlenecks formed by Arg312 and Arg398 in the cap path and Trp89 and Lys5454 in the FAD path are represented by pentagrams and heptagrams, respectively.

Two of these migration paths, the cap and FAD paths, are less likely paths for these ligands. They require large movements of amino acid side chains (namely Arg312 and Arg398 or Trp89 and Lys545, respectively) to allow the phenolic ligand to pass, which make them more difficult passages to the active site. This can be seen in the simulations, where the ligands often got stuck next to these residues. It is also evident from the high number of atom contacts ligands experienced with these residues. VAO also accepts phenolic substrates with larger hydrophobic side chains than the ligands studied here [40]. These larger ligands would be prone to have even more difficulties passing the bottlenecks in the cap and FAD path. Therefore, and because it represents the only entry and exit path identified, we argue that the VAO subunit interface path is the most probable route for phenolic ligands to enter or leave the active site.

In the subunit interface path, two residues are observed to change conformation to allow ligands to leave the active site. These two residues, His466 and Tyr503, do not present obstacles for the migration of the ligands but are

in a closed conformation once a ligand is bound. Our results suggest a dual role for Tyr503, which is known to be involved in substrate deprotonation [49], and which we can show here to be also involved in substrate migration. His466 and Tyr503 can be regarded as concierges, limiting access to the active site. Interestingly, in the crystal structure of PCMH (PDB ID: 1DII [56]), two possible conformations are observed for His436 and Tyr473, which correspond to His466 and Tyr503 in VAO (see Figure 31A for an illustration). We would also like to highlight Met195 as the most flexible of the three portal residues to complete the list of flexible and probably functionally important residues in this path. It is of interest to keep in mind that this path shows that the two active sites of the VAO dimer are connected via this portal.

We were also able to demonstrate that protonated phenolic substrates enter the enzyme but deprotonated ones do not. Energy profiles of exit simulations also indicate more favourable binding of the protonated substrates and indifference towards the protonation state of the product. We have also identified an additional migration path, the *re* path for the co-ligands dioxygen and hydrogen peroxide. This path leads co-ligands to and from the *re* side of FAD. The co-ligands were also able to migrate through the paths identified for the phenolic ligands. The *re* path is the shortest connection to the solvent, compared to the paths on the *si* side of FAD. On the *re* side of FAD, a chloride ion present in the VAO crystal structure (PDB ID: 1VAO) is indicative of a possible oxygen binding site [27, 247]. There is no evidence from binding energies calculated in our simulations that the *si* path is preferred. Our simulations were performed with oxidised FAD, and while dioxygen reacts with reduced FAD, there are no indications of significant conformational changes in VAO crystal structures linked to the changed oxidation state of FAD (PDB ID: 1VAO vs PDB ID: 1AHU). All these observations make

Dioxygen and hydrogen peroxide have very different electrochemical properties, but use the same path. This indicates that no electrochemical bottleneck exists in the *re* path, making the size of the ligands the most important criterion. Given that dioxygen and hydrogen peroxide are similar in size and the channel through which they pass is comparably big, both use the same path.

the *re* path the most probable migration path for dioxygen and hydrogen peroxide. In this path, it is possible that Tyr51 is involved in directing dioxygen to the FAD, as it is one of the most contacted and flexible residues in this path. It is also located in the probable oxygen-binding pocket on the *re* side of FAD, making it a candidate for assisting in reduced flavin oxidation.

From our results of the co-ligand simulations on the *si* side of FAD, we can see that co-ligands tend to migrate into the cap path. Crystallographic data (PDB ID: 1VAO) show that within the cap path, water molecules are located close to Arg312 and Arg398 as well as at the paths exit/entry on the surface of VAO. Water is able to access the active site and to participate in the reaction on the *si* side of FAD, reacting with the quinone methide of certain substrates [223]. The cap path is the only path at the *si* side of FAD directly connected to the solvent. In the case of binding of alkylphenols with hydrophobic aliphatic side chains, already bound water molecules would likely be expulsed from the active site via the cap path. This would explain why these substrates are dehydrogenated and not hydroxylated by VAO [37].

Structure and sequence conservation in VAO, EUGO and PCMH We will now discuss our findings in VAO in the context of the close bacterial relatives EUGO and PCMH. All three enzymes require their substrates to be *para*-substituted phenols [36, 40, 55, 56, 238].

In PCMH and EUGO, a possible route for substrate access to the active site via the subunit interface has been suggested purely on crystallographic evidence [56, 202]. In this route, three residues (Glu177, Met180 and Asp434 in PCMH) have been proposed to be involved. These residues are Asp192, Met195 and Glu464 in VAO, with Asp192 being in the position of Glu177 and Glu464 in the position of Asp434. EUGO has recently had its crystal

structure solved [202]. In EUGO, all the above-mentioned residues involved in the subunit interface path of VAO are also present (Figure 31).

We could show that residues at the subunit interface and in the subunit interface path are the most similar in all three of these closely related members of the VAO/PCMH family, an indication that increased selective pressure is in place for these residues. Additionally, the combination of this increased selective pressure and the presence of the entrance/exit migration path at the subunit interface (and absence of evidence of allosteric regulation in VAO) could indicate that dimerisation of VAO is functionally linked to ligand migration. Residues in the cap path are most diverse in all three enzymes. This lends support to the proposal that all three enzymes prefer the subunit interface path for ligand migration. The small amount of PCMH-like sequences remaining in the analysis after applying our selection criteria makes interpretation of the data with respect to PCMH less reliable.

VAO and EUGO re-oxidise their FAD cofactor using molecular oxygen, while PCMH requires a cytochrome c subunit for this. This functional difference is reflected in the preservation of the *re* path in these enzymes (see specifically the PCMH-unlike category in Figure 34B). Interestingly, dissimilarity in the residues of this path is only due to PCMH-like sequences and not due to differences between the two oxidases. It is noteworthy that the *re* path we identify here is located in a visible channel in VAO and EUGO. In PCMH, the channel is absent and the cytochrome c subunit in PCMH is located where dioxygen would exit the protein (see Figure 31B). Evolution of the *re* path in these oxidases may thus have coincided with loss of the cytochrome c subunit.

The strong conservation of residues in the FAD path could indicate that these residues are essential to the func-

tion of all three enzymes. It has been established that maturation of the VAO holoenzyme involves the initial non-covalent binding of the FAD cofactor to the already folded dimeric apoenzyme [14, 41]. It could be that FAD binds to the apoenzyme via this path. A possible scenario is that FAD enters the enzyme with the isoalloxazine moiety first and then binds to the enzyme in the correct orientation already. The alternative scenario, in which FAD binds with its ADP moiety first, would require reorientation of the FAD inside the enzyme to fit the isoalloxazine moiety into the active site. From the crystal structure, it is not evident how this would happen as the free space around bound FAD is rather limited. Initial binding of FAD to the enzyme surface and migration of FAD into its binding pocket is likely guided by *pi* stacking interactions with the three tryptophans located in the FAD path (Trp98, Trp549 and Trp558).

What is known about (co-)ligand migration in the entire VAO/PCMH family? Very little is known about substrate access to the active site for the other members of this flavoprotein family. Dioxygen migration has been studied in alditol oxidase, berberine bridge enzyme (also called (*S*)-reticuline oxidase) and L-galactono-1,4-lactone dehydrogenase. Ala113 in L-galactono-1,4-lactone dehydrogenase was identified as a gatekeeper residue that prevents the enzyme from functioning as an oxidase [248]. Berberine bridge enzyme also contains such a gate-keeper residue. The variant G164A showed 800-fold decreased oxygen reactivity [249]. In alditol oxidase funnel shaped paths have been identified through molecular dynamics simulations, leading dioxygen to the "gatekeeper" Ala105 in alditol oxidase and from there to N1/N3 of the flavin ring [240]. This path leads to the *re* side of FAD, as does the *re* path we identified in this study. Our path, however, leads to its destination via the dimethylbenzene part of the flavin ring. We have therefore identified for the first time

a substrate migration channel, the subunit interface path, and a novel co-substrate migration path.

In summary, we have carried out an exhaustive computational analysis of ligand migration pathways in VAO (see Figure 35 for an overview). We have identified a gated entry and exit path at the subunit interface of VAO for small phenolic ligands. The residues forming a portal in this path (Met192, Met195 and Glu464 in VAO) are conserved in PCMH and EUGO. Two residues, His466 and Tyr503 in VAO, act as concierges, obstructing the path to the active site for phenolic substrates once substrate is bound, as has been postulated before for PCMH [56]. An additional, different entry and exit path for dioxygen and hydrogen peroxide at the interface of the two domains of VAO has been identified (*re* path). Tyr51 in this path could assist reduced flavin oxidation in VAO. The only path identified directly connecting the *si* side of FAD to the solvent (the cap path) could be facilitating water access to the active site. Overall, our study illustrates the ability of ligand simulation techniques to advance the mechanistic understanding of enzyme function.

Materials and methods

System set-up The initial coordinates for the VAO protein were taken from the 1VAO entry in the protein data bank (PDB) [40]. Although the protein is assembled as an octamer, due to computational limitations only the dimer, the smallest functional unit of VAO, was modelled. No allosteric regulation has been observed for VAO in its dimeric or octameric oligomerisation state. Loop deletion mutagenesis has revealed that the VAO dimer has similar catalytic properties as the octamer [39]. The protein structure was prepared assisted by the protein preparation wizard available in the Schrödinger software package [250]. All crystal water molecules were removed and missing hydrogen atoms were added to protein residues. Residues such as histidines, glutamates and aspartates were inspected for appropriate protonation states at pH 7.5 to match experimental conditions. PROPKA [251] and H++ web server (`http://biophysics.cs.vt.edu/H++`) [252] were employed to determine the pK a of each residue. His56, His61, His313, His506 and His555 were modelled as ε protonated while all other His residues were δ protonated. Furthermore, the NE2 atom in the imidazole ring of His422 has no hydrogen atom since it is covalently bound to the C8M atom of the FAD cofactor and considered to be neutral. Missing residues from position 41 to 47, forming a loop at the surface of the protein, were added and minimised using PRIME [253].

The final system was composed of a total of 17'674 atoms. Six phenolic and two small ligands were used including: COP, COD (*p*-creosol, where the P and D refer to the protonated and deprotonated state of the ligand; this notation is identical for all molecules), VAP, VAD (vanillyl alcohol) VNP, VND (vanillin) and dioxygen and hydrogen peroxide. All charges for the FAD cofactors were obtained through quantum mechanics/molecular mechanics (QM/ MM) calculations inside the protein environment with an 8 Å layer of explicit waters. The QM region included all the isoalloxazine ring atoms and the QM/MM cut was made between the C1 and C2 atoms of the FAD molecule. All the ligands were optimized through QM calculations in an implicit solvent. QM/MM calculations used the all atom OPLS2005 force field [254], M06 functional [255] with the 6-31G** basis set [256] and Poisson Boltzmann Finite element (PBF) implicit solvent [257] (same functional and basis set used in QM). The parameter files for all ligands and the FAD cofactor are supplied as supplementary information S1.

Ligand migration PELE [67] was used to map the migration (exit and entry) path of different ligands between the VAO active site and the solvent. PELE is a Monte Carlo based algorithm that produces putative new configurations through a sequential ligand and protein perturbation scheme, side chain prediction and minimisation steps. A detailed description of the PELE methodology can be found elsewhere [67]. Ligand perturbation involves a random translation and rotation, while protein perturbation involves a displacement on the alpha carbon following one go the six lowest Anisotropic Normal Modes (ANM) [258]. These steps compose a move that is accepted (defining a new minimum) or rejected

based on a Metropolis criterion for a given temperature. The combination of ligand and protein backbone perturbations results in an effective exploration of the protein energy landscape. This approach is capable of reproducing large conformational changes associated with ligand migration and has already been shown to produce reliable results [259, 260, 261].

For the entry simulations, the ligand was placed in six different positions at the bulk solvent. Similar parameters to the exit simulations were used, but with larger translations (from 1.25 and 2.5 Å) during substrate perturbations. The direction of such perturbations was kept for three consecutive steps. Moreover, for the entry simulations an adaptive scheme was used (using the new C++ version of PELE [68] plus an OBC solvent [262]). After a short simulation of 12 Monte Carlo steps, the adaptive scheme clustered the ligand position and new initial conditions were chosen, prioritizing those clusters with less population. Due to the use of different versions of PELE, the binding energies differ by a level of magnitude for ligands migrating in and out of VAO (see Results). In this context, it is important to note that the binding energies calculated by PELE are indicative of favourable or unfavourable positions for the ligand. While we refer to them as binding energies, they are interaction internal energies between the protein and the ligand, with a significant larger value than experimental ones, and should be analysed in a qualitative manner.

The exit simulation protocol was identical for all substrates and began by manually docking the phenolic molecule (COP, COD, VAP, VAD, VND,VNP) to the *si* side of the isoalloxazine ring of the flavin in the active site based on crystallographic evidence from the VAO structure with isoeugenol (PDB ID: 2VAO) [40]. Simulations were performed in equal number starting from both protein subunits. The phenolic ligand was then perturbed with random translations (from 0.75 and 1.75 Å) and rotations (0.05 and 0.25 rad) and requested to move away from the active site using PELE's spawning procedure. This procedure selects a reaction coordinate (e.g. an atom-atom distance, ligand RMSD) and abandons the trajectories with worst values in the reaction coordinate to restart the simulation at the best value (which is constantly updated). Trajectories are abandoned only if they fall behind a user-defined range (from the best value). In our case, the spawning coordinate was the distance of the ligand center of mass to a point in the void volume of the active site (X, Y, Z coordinates 92, 25, 41 or 108, 47, 70 for simulations starting in the active site of chain A or B, respectively), with the distance range from the spawning coordinate being 4 Å. Protein perturbation was based on the six lowest normal modes [263] and side chain sampling including all residues within 8 Å from the ligand. Finally, after global minimisation had optimised the new configuration, this configuration was filtered with a Metropolis acceptance test. In this test, the energy is described with an all atom OPLS 2005 force field [254] with a surface generalized Born solvent model [264]. Simulations were stopped when the wall clock limit time (48h) was reached or the phenolic ligand exited the protein completely. A total of 800 independent trajectories were produced for each ligand.

Simulations exploring exit paths from the active site of the small ligands dioxygen and hydrogen peroxide were performed using a non-biased approach as described previously [260], using random translations and rotations appropriate to achieve continuous paths (between 0.5 and 1.5 Å and 0.1 to 0.3 rad, respectively).

Data analysis was performed using VMD (Version 1.9.1, [265]), PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.), and R (Version 3.3.1 [226]). Figures were created using the aforementioned programs, with the addition of GnuPlot (Version 4.4 [266]) and ChemDraw (Version 15.1.0.144, PerkinElmer Informatics).

Atom contacts between the ligands and the protein were calculated using a python script. A contact was defined as a distance smaller than 2.7 Å between an atom from any given residue and an atom belonging to the ligand. Root mean square fluctuation (RMSF) values of residues in the simulations were calculated using VMD (Version 1.9.1, [265]). Only RMSF values from residues in the chain of the dimer where the ligand was placed were used for the analysis.

Protein structure comparison To compare the structural conservation of path residues, the crystal structures of VAO, EUGO and PCMH (1VAO, 5FXE and 1DII respectively) were aligned using the PROMALS3D multiple sequence and structure alignment server [267]. The crystal structures were also compared using PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.) and the built in alignment function.

Protein sequence comparison To compare the sequence conservation of path residues, a set of VAO-, EUGO- and PCMH-like sequences was identified. The protein sequences from crystal structures 1VAO, 1DII and 5FXE were used as queries for similarity searches with Consurf [268], searching the uniref90 database using default settings for the search of homologs. This approach resulted in three alignments (one each for VAO-, EUGO- and PCMH-like sequences). The obtained alignments were manually checked and VAO- and EUGO-like sequences that did not contain a histidine residue corresponding to His422 and His390 respectively were removed. PCMH-like sequences not containing a tyrosine residue corresponding to Tyr384 were removed as well. This selection was made because the named residues covalently bind the FAD cofactor, which has a significant impact on enzyme function [238, 41, 202].

The homology search resulted in 121, 117 and 30 VAO-, EUGO- and PCMH-like sequences remaining in the respective alignments (named MSA VAO , MSA EUGO and MSA PCMH, respectively). These alignments where then merged using the MAFFT webserver (`mafft.cbrc.jp/alignment/software/merge.html`) to match the residue positions (resulting in MSA_{all}). The alignments were then uploaded to Consurf [267] to calculate the residue variety in percentages, always using the pdb structure 1VAO as query. This approach ensures that the same alignment positions are compared when the preservation of different residues in the three alignments is compared as described below; e.g. that residue number 422 in VAO is always compared to the residue present in the

other sequences at the corresponding position in the three alignments. All alignments (MSA_{VAO}, MSA_{EUGO}, MSA_{PCMH} and MSA_{all}) can be found as supplementary information S1 Alignment.fa, S2 Alignment.fa, S3 Alignment.fa and S4 Alignment.fa, respectively.

An R script was used to analyse the preservation of different groups of residues in these alignments, amongst which residues involved in the paths identified in this study. We defined nine different groups of residues: the entire protein (560 amino acids), residues forming the cap domain (from residue 270 to 500), the *re* path residues (22 residues), the FAD path residues (36 residues), the cap path residues (25 residues), the subunit interface path residues (53 residues), a selection of random residues (58), surface residues of VAO (90 residues) and residues at the subunit interface (86 residues). Surface residues were defined using the built in tool of the SwissPDBviewer, with default settings (30% accessibility of residues by solvent in dimeric VAO) [269]. Interface residues, meaning residues at the subunit interface, were defined as residues within 4 Å of either subunit.

Preservation of these residues was analysed based on similar or dissimilar residues. Similar residues were defined according to the BLOSUM62 similarity parameters [270], resulting in six sets of residues (Trp, Tyr and Phe; Met, Ile, Leu and Val; His, Arg and Lys; Asn, Asp, Gln and Glu; Ser, Thr, Pro Ala and Gly; and Cys).

The preservation of residues was analysed per position in the alignments and a cutoff of 50% preservation was used to categorise them as similar in the alignment. If preservation of the similar residues was above 90%, they were categorised as conserved in the alignment. This analysis was performed for MSA_{VAO}, MSA_{EUGO} and MSA_{PCMH}, for each of the nine groups of residues. This resulted in a percentage of similarity and a percentage of conservation for each group of residues for each sequence alignment of VAO-, EUGO- and PCMH-like sequences.

This analysis did not yet allow us to compare the conservation or similarity of these three alignments to each other. Different residues can be preserved in each of the three alignments, but we were interested in finding out how many residues are preserved in all three alignments (MSA_{all}). We therefore performed the following analysis: we used the same definitions for similarity and conservation as above, but this time analysed the preservation of the nine groups of residues in MSA_{all}. We added additional categories to this analysis: dissimilar and PCMH-unlike. Dissimilar positions were defined as positions less than 50% similar. PCMH-unlike positions are dissimilar positions for which VAO-like and EUGO-like sequences share a similar residue, but differ from PCMH. This resulted in percentages of similarity and conservation as well as percentages of dissimilarity and PCMH-unlikeness.

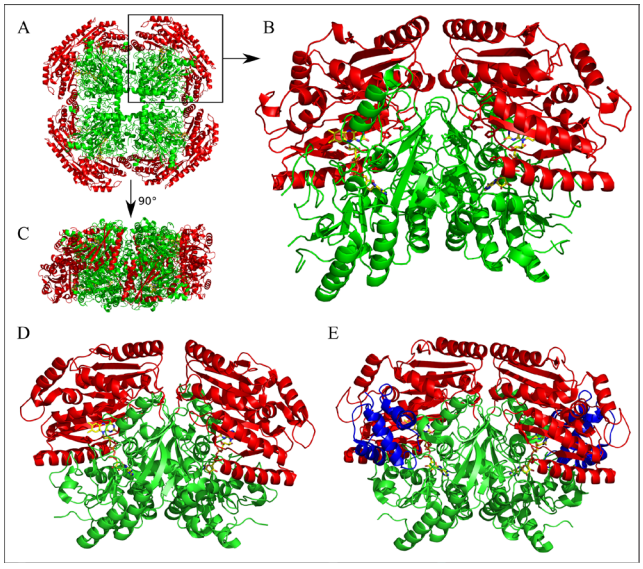
Acknowledgements

We would like to thank Ferran Sancho Jodar from the BSC (Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain) for providing us with a script to analyse atom contacts in our simulations.

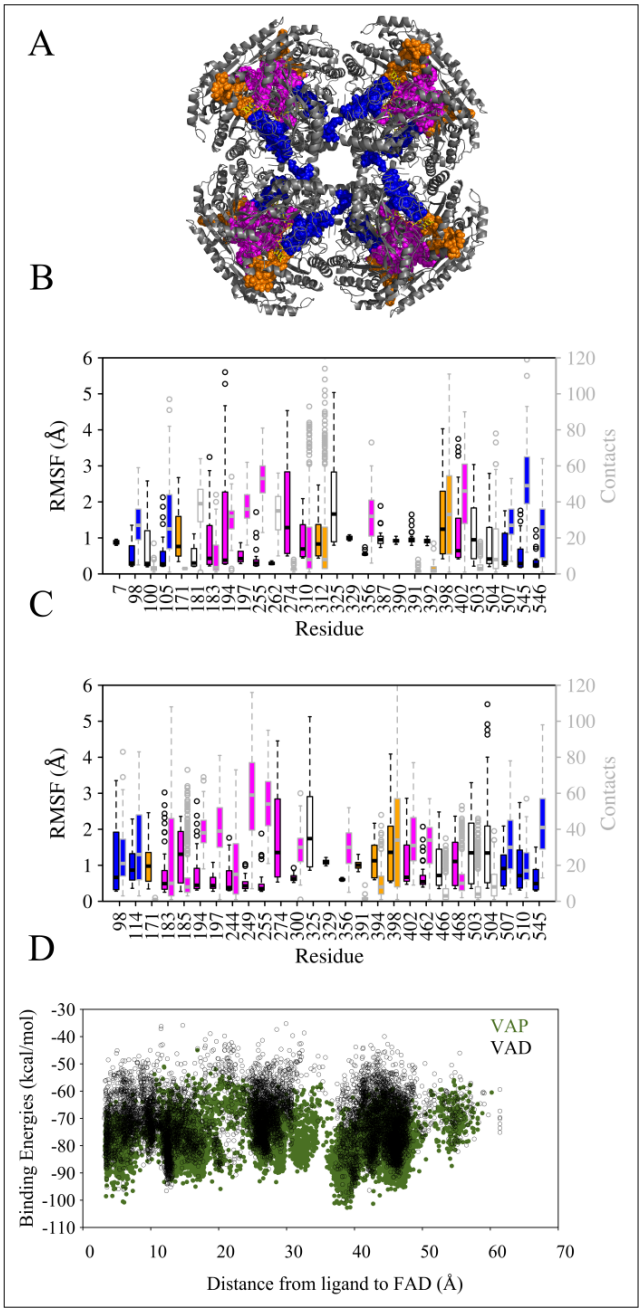
Supplementary Data

Other supplementary data can be found with the original publication: `https://doi.org/10.1371/journal.pcbi.1005787`

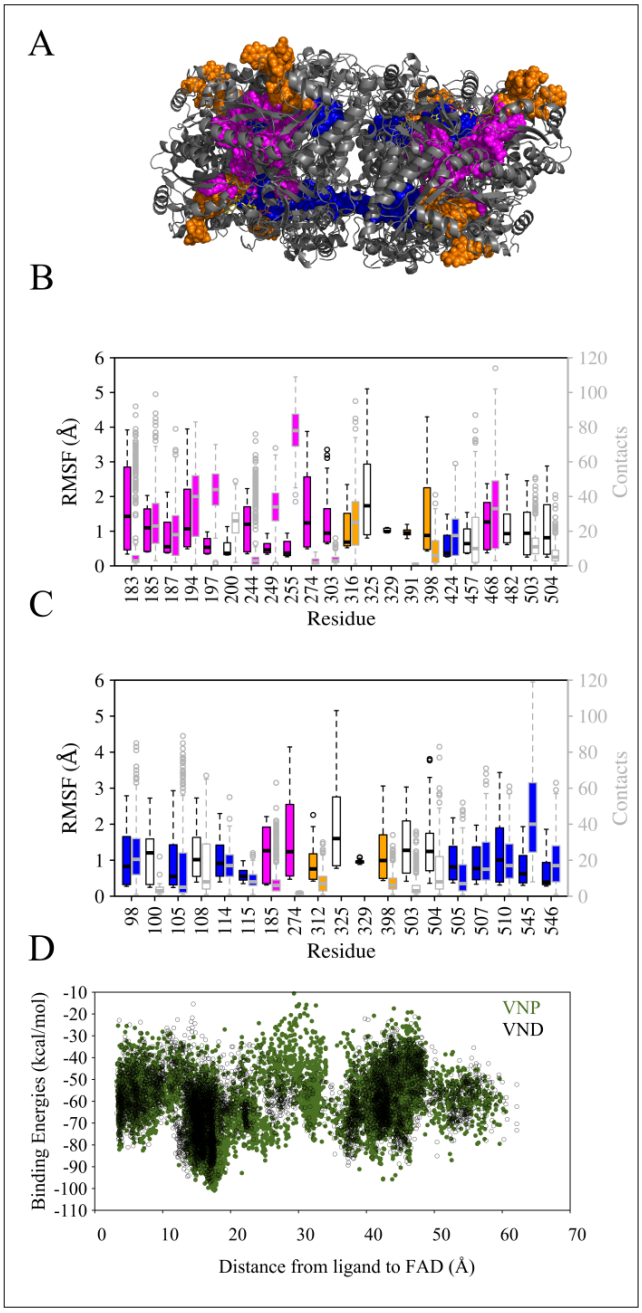
Supplementary Figures



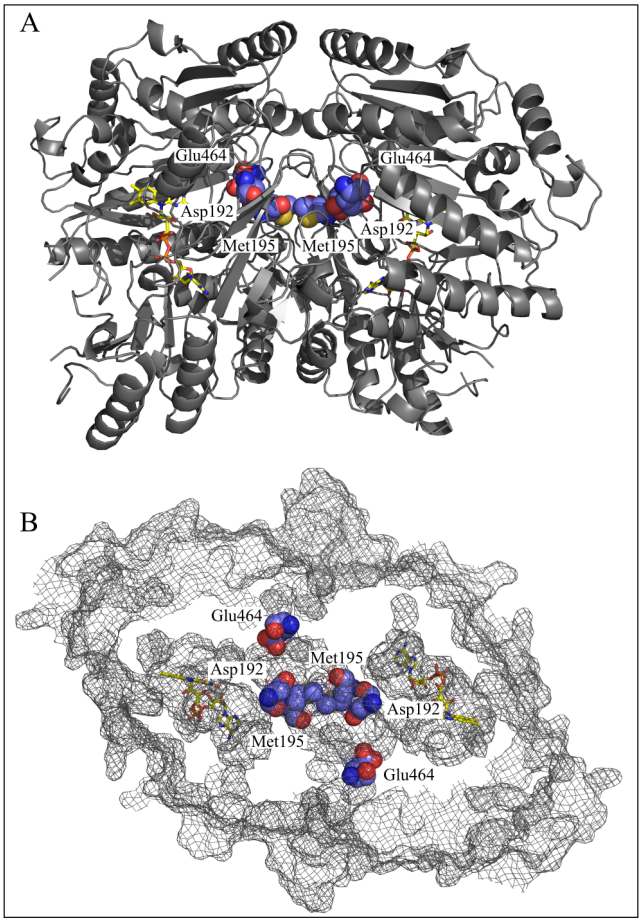
Supplementary Figure 3: Three-dimensional structures of VAO, EUGO and PCMH. Proteins (PDB IDs: 1VAO, 5FXE, 1DII) are shown as cartoons. The cap domain is coloured in red, the FAD-binding domain in green and FAD is shown as yellow sticks.
A: Octameric VAO, seen from the front.
B: Dimeric VAO as seen in Fig 1.
C: Octameric VAO seen from the side.
D: Dimeric EUGO, using the same colouring scheme as for VAO.
E: Tetrameric PCMH, using the same colouring scheme as for VAO, with the cytochrome c subunit coloured in blue.



Supplementary Figure 4: Overview of the exit paths identified with VAP and VAD (A), RMSF and atom contacts (B for VAP, C for VAD) and binding energies (D).
A: The VAO dimer is shown as cartoon in grey and its FAD cofactor is shown as sticks in yellow. The paths identified in this work are shown through sphere representation of ligand positions in the trajectories. The cap path is coloured in orange, the FAD path in blue and the subunit interface path in magenta.
B: Plot of simulations with VAP, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
C: Plot of simulations with VAD, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
D: Scatterplot of binding energies as a function of the distance between the VAP and VAD (in green and black) to FAD.

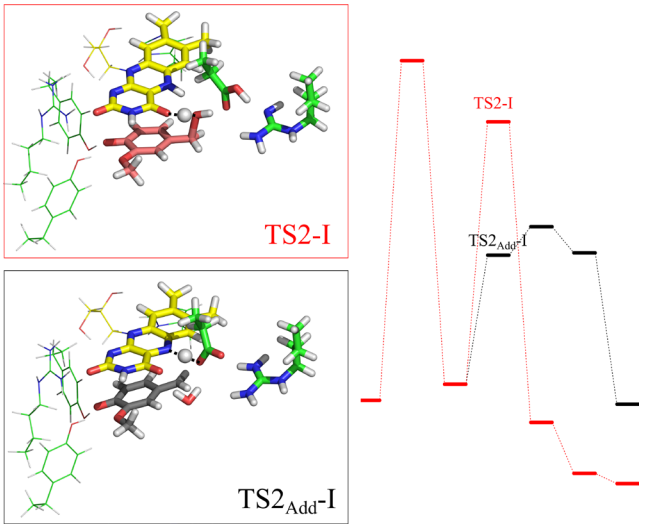


Supplementary Figure 5: Overview of the exit paths identified with VNP and VND (A), RMSF and atom contacts (B for VNP, C for VND) and binding energies (D).
A: The VAO dimer is shown as cartoon in grey and its FAD cofactor is shown as sticks in yellow. The paths identified in this work are shown through sphere representation of ligand positions in the trajectories. The cap path is coloured in orange, the FAD path in blue and the subunit interface path in magenta.
B: Plot of simulations with VNP, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
C: Plot of simulations with VND, highest RMSF values (black boxplot boxes) and atom contact counts with the ligand (grey boxplot boxes) of residues in VAO. Coloured filling of the boxplots indicates that the residue belongs to the path of that colour.
D: Scatterplot of binding energies as a function of the distance between the VNP and VND (in green and black) to FAD.



Supplementary Figure 6: Detailed view of the portal connecting the subunit interface to the active site formed by Asp192, Met195 and Glu464. Asp192, Met195 and Glu464 are shown as spheres (with carbons in violet) within dimeric VAO (in grey) with FAD shown as sticks (in yellow). A: The dimer as cartoon from the front. B: The dimer seen from the top as mesh.

Chapter 6:
Quantum chemical modelling of vanillyl alcohol oxidase



This chapter is based on:
G. Gygli, X. Sheng, F. Himo and W. J. H. van Berkel. *Manuscript in preparation.*

Keywords: Adduct, creosote, density functional theory, flavoprotein, lignin, natural product, quantum chemistry, vanillin

Abstract

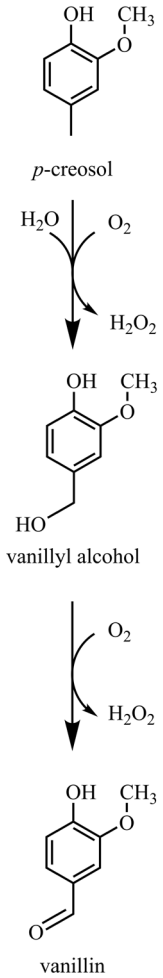
The flavoenzyme vanillyl alcohol oxidase (VAO) catalyses the two-step reaction of *p*-creosol to vanillin. This reaction is of interest for the production of natural vanillin, but is slowed down by the formation of a covalent flavin-creosol adduct. We have studied different possible reaction mechanisms, and identified the most energetically favourable one for each reaction step.

In both reaction steps, the transfer of a hydride from the substrate to the flavin initiates the reaction, yielding a *p*-quinone methide intermediate. The experimentally observed lifetimes of these two intermediates are however different, and are in agreement with the energy barriers we have calculated.

The relative energy barriers of the productive reaction (leading to vanillyl alcohol) are higher than those of the non-productive reaction (leading to the covalent adduct). The relative energies of the vanillyl alcohol product and the adduct are however comparable. Adduct formation is therefore kinetically favoured. The short side chain of *p*-creosol enables Asp170 to initiate adduct formation. With bulkier substrates, e.g. vanillyl alcohol, Asp170 is unable to attack and no adduct is formed.

Introduction

Vanillyl alcohol oxidase (VAO) is a flavoenzyme capable of catalysing the conversion of *p*-creosol to vanillin, see Scheme 1 [271, 272]. The intermediate of this reaction, vanillyl alcohol, is also a substrate of the enzyme. It is therefore possible to synthesise natural vanillin from natural *p*-creosol with VAO [272]. *p*-Creosol (4-methylguaiacol) is the second most abundant component in beechwood creosote [273], and a product of lignin pyrolysis [274, 275]. Vanillin production from lignin or wood [276, 277] could therefore profit from using VAO. Unfortunately, the reaction of VAO with *p*-creosol is severely hampered by the formation of a covalent adduct between *p*-creosol and VAO's FAD cofactor [45]. The mechanism of adduct formation is poorly understood, opening the door for this computational study.



Scheme 1: Two-step enzymatic conversion of *p*-creosol to vanillin.

VAO is a fungal enzyme, which has structural and catalytic properties in common with several bacterial homologs, e.g. *p*-cresol methylhydroxylase (PCMH) and eugenol oxidase (EUGO) [56, 55, 202]. VAO is a member of the VAO/PCMH family, a highly diverse family of flavoenzymes [196]. VAO, PCMH, and EUGO are part of the 4-phenol oxidising subgroup of this family [196]. Upon binding of the phenolic substrates to VAO, the phenolate is formed through the action of two tyrosine residues (Tyr108 and Tyr503) [49]. See Figure 36 for an illustration of these and other active site residues. Two tyrosines are also involved in binding the phenolic substrates of phenolic acid decarboxylases [278, 279] and hydroxycinnamoyl-CoA hydratase-lyase [280].

The catalytic mechanism of VAO has been demonstrated to operate via a *p*-quinone methide intermediate [36, 223]. Phenolic acid decarboxylases and hydroxycinnamoyl-CoA hydratase-lyase have also been

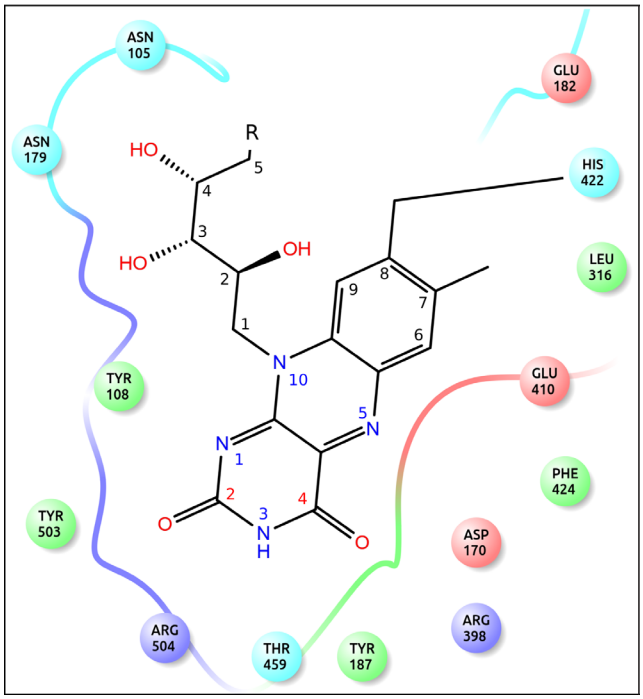


Figure 36: Cofactor interaction diagram of VAO. The covalent bond between His422 and the flavin cofactor is drawn and the atoms of the flavin isoalloxazine ring and the ribityl chain are numbered. The ribityl chain is cut off at C5, and the remainder of the flavin (FAD) molecule is abbreviated as R. Cyan indicates polar, green hydrophobic, red negatively charged and violet positively charged amino acid residues.

shown and suggested to operate via a *p*-quinone methide intermediate [278, 279, 280]. In VAO, the first step of the reaction of *p*-creosol to vanillin (Scheme 1), i.e. the conversion of *p*-creosol to vanillyl alcohol, is extremely slow because of the formation of a covalent adduct between enzyme-bound flavin and *p*-creosol [45]. For the second reaction step, the conversion of vanillyl alcohol to vanillin, reduction of the flavin is rate-limiting [223]. Note that only the first reaction step requires water as a reactant, and that the overall reaction is rate limited by the decomposition of the covalent adduct.

Interestingly, residues remote to the active site appear to influence the life-time of the adduct [45]. While insight has been gained in the mechanism for the conversion of vanillyl alcohol to vanillin [223], the enzymatic

conversion of *p*-creosol to vanillyl alcohol remains puzzling. The substrate analogue *p*-cresol is an excellent substrate for PCMH, but also gives rise to a covalent adduct in VAO [45]. The active sites of VAO and PCMH vary significantly and are likely the source of this different reactivity. However, the influence of remote residues on VAO activity remains cryptic. The lifetime of the *p*-quinone methide intermediate is different for different substrates. The intermediates of 4-(methoxymethyl)phenol and vanillyl alcohol can be observed experimentally in stopped-flow kinetic experiments. The *p*-quinone methide of 4-(methoxymethyl)phenol is rather stable in the reduced enzyme, while the *p*-quinone methide of vanillyl alcohol rapidly decomposes to form vanillin [223].

We have studied the reaction mechanism of the VAO-mediated two-step reaction of *p*-creosol to vanillin in detail with density functional theory (DFT) calculations employing the cluster approach [69, 281, 282, 283, 284]. Based on the optimised structures of the species along the reaction coordinate and the calculated relative energies, we shed more light on this reaction. The possible mechanisms for this reaction are shown in Figure 37 and summarised here. The red pathway labelled with I denotes the productive reaction from *p*-creosol to vanillyl alcohol (Figure 37A), the black pathway labelled with $_{\text{Add}}$-I denotes the non-productive reaction leading to the *p*-creosol-flavin-adduct (Figure 37B), and the blue pathway labelled with II denotes the second reaction step leading to vanillin (Figure 37C).

In the active site, the phenolate of *p*-creosol is formed immediately upon binding (React-I, in red in Figure 37A). We can therefore neglect the protonated substrate in our calculations. *p*-Creosol reduces the flavin cofactor through hydride transfer to the N5 of the isoalloxazine ring of oxidised flavin, yielding the *p*-quinone methide intermediate (InterI-I in Figure 37A). The productive reaction

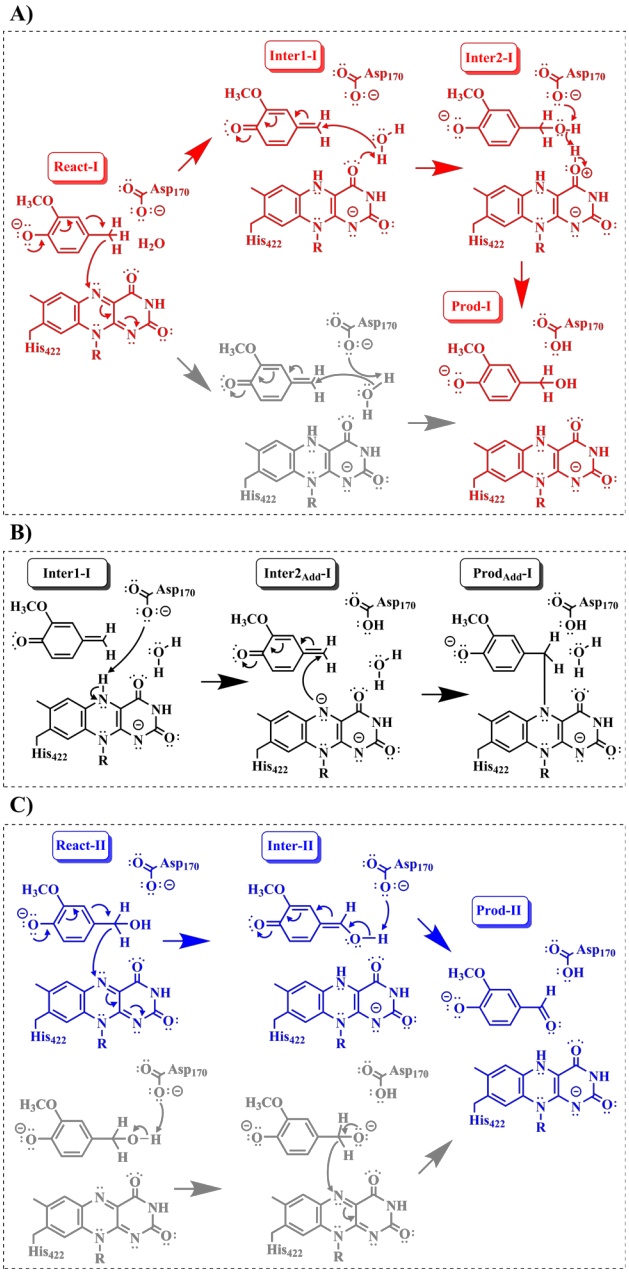


Figure 37: Proposed mechanistic steps of the reaction of *p*-creosol to vanillin via vanillyl alcohol, based on experimental work [43, 223, 271] and the calculations in the present study.
A) Productive reaction of *p*-creosol to vanillyl alcohol.
B) Adduct formation from the creosol *p*-quinone methide intermediate.
C) Reaction of vanillyl alcohol to vanillin. See text for details.

(in red or grey in Figure 37A) proceeds without adduct formation, directly forming vanillyl alcohol by water addition to the *p*-quinone methide of *p*-creosol (Prod-I in Figure 37A). This water addition could be facilitated by Asp170 acting as proton acceptor (in gray in Figure 37A). Alternatively, as shown in the mechanism coloured in red in Figure 37A, the O4H of FAD could be deprotonated by transferring a proton back to the product, while Asp170 simultaneously attacks the proton of the substrate. This terminates the reductive half-reaction of the first reaction step. Dioxygen then oxidises the reduced flavin and releasing hydrogen peroxide completes this first reaction step (Asp170 also needs to be deprotonated, not shown).
The non-productive reaction to the adduct is coloured in black in Figure 37B (Inter1-I to Inter2$_{\text{Add}}$-I and Prod$_{\text{Add}}$-I). This can happen through the action of Asp170, which attacks the proton at flavin N5H (yielding protonated Asp170) and the addition of flavin N5 to the substrate *p*-quinone methide intermediate (yielding the adduct, Prod$_{\text{Add}}$-I). In theory, this could happen in a concerted fashion or stepwise (with either of the two steps first or second). Once the adduct is formed, water addition to the adduct hypothetically could yield vanillyl alcohol (not shown in Figure 37B).

To begin the second reaction step, the produced vanillyl alcohol is bound to the active site in its phenolate form (React-II, in blue or grey in Figure 37C). Vanillyl alcohol bound to the enzyme then reduces oxidised flavin through hydride transfer to flavin N5 (Inter-II, in blue in Figure 37C). The *p*-quinone methide of vanillyl alcohol then decomposes to the aldehyde (vanillin) through proton transfer to Asp170 (yielding protonated Asp170, Prod-II in blue in Figure 37C). In theory, these two transfers (hydride and proton) could happen in a concerted fashion or stepwise (with either of the two transfers first or second, see blue and grey alternatives in Figure 37C). To complete the

redox-cycle, dioxygen again oxidises reduced flavin and Asp170 is deprotonated (not show in Figure 37C).

Results

We have studied the reductive half-reactions of the two-step conversion of *p*-creosol to vanillin by using quantum mechanical calculations. For this, we have built active site models for the reaction of *p*-creosol to vanillyl alcohol and for the reaction of vanillyl alcohol to vanillin (see Figure 38 for an overview of these active site models). On the basis of these models, we optimised the structures of the intermediates and transition states along the reaction coordinates and calculated the energy barriers. We will first describe the results for the reaction of *p*-creosol to vanillyl alcohol, and then the results for the reaction of vanillyl alcohol to vanillin.

Active site models The crystal structure of VAO with the inhibitor isoeugenol bound was chosen to build the active site model (PDB ID: 2VAO, [40]). The model consists of the substrate *p*-creosol or vanillyl alcohol in place of the isoeugenol inhibitor, the FAD cofactor and the following residues that make up the active site: Tyr108, Asp170, Arg398, His422, Tyr503 and Arg504. These residues were included in the model because of their (proposed) involvement in the reaction mechanism. The FAD cofactor is oxidised, with the ribityl chain cut off at the C3, thus containing one hydroxyl group at C2 and C3, and an added hydrogen atom to saturate C3. The model therefore consists of 165 atoms and has a total charge of +1, see also Figure 38. Non side-chain atoms were removed, except alpha carbons, and hydrogen atoms were added to saturate these alpha carbons. The positions of these hydrogens as well as the alpha carbons were kept fixed to their crystallographic positions during all the geometry optimisations.

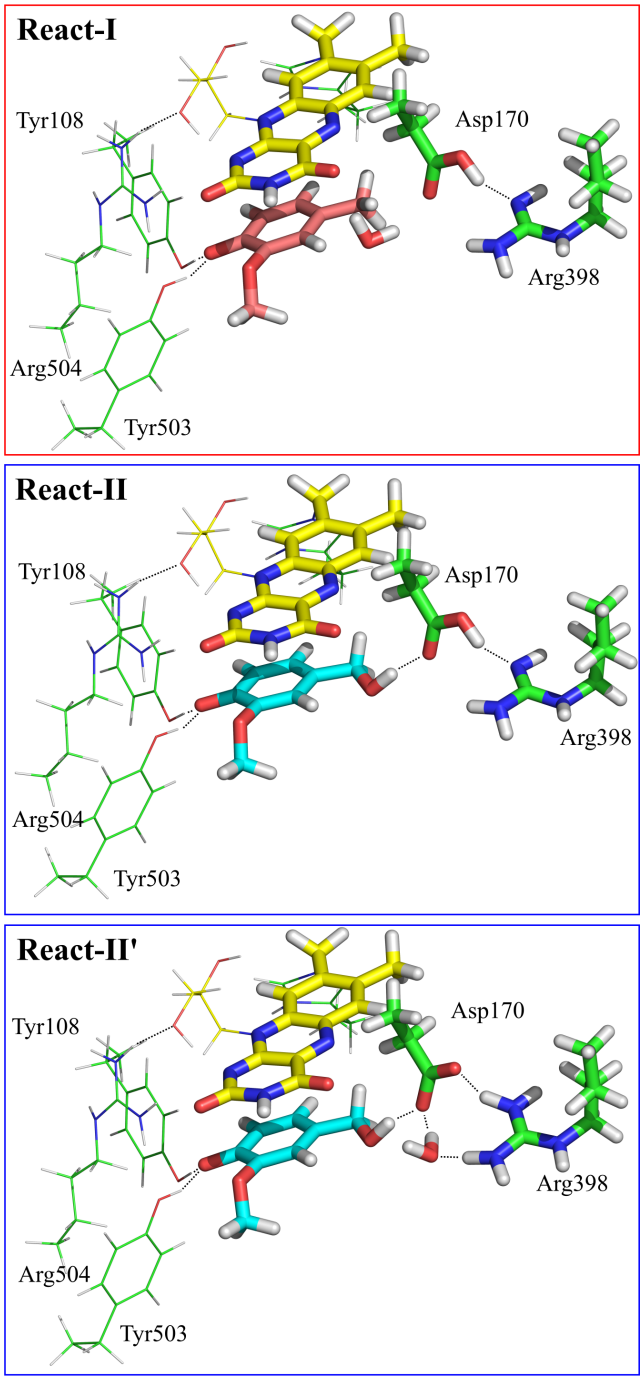


Figure 38: Optimised structures of the active site models used in this study. His422 is located behind the FAD cofactor and is not labelled for clarity. The ribityl side chain of the FAD cofactor is also shown as lines for clarity. Selected hydrogen bonds are shown as thin, dashed lines.
React-I: Active site model with *p*-creosol (in red).
React-II: Active site model with vanillyl alcohol (in cyan) without a water molecule.
React-II': Active site model with vanillyl alcohol (in cyan) and a water molecule.

Two of the hydroxyl groups of the ribityl chain were kept in the model because of their interaction with the isoalloxazine ring observed in the crystal structure. The substrate was positioned in analogy to the binding mode of the inhibitor isoeugenol. The hydroxyl group of the phenol (oriented towards Tyr108 and Tyr503) was assumed to lose its proton upon binding to the active site, based on experimental evidence [49]. The substrate is therefore modelled in the deprotonated state. Tyr108, Arg398, His422, Ty503 and Arg504 were modelled in their respective protonated state. A water molecule was included in active site model React-I (see Figure 38), since it is required for the hydroxylation of *p*-creosol. The reaction from vanillyl alcohol to vanillin does not directly involve the water molecule, but its presence influences the protonation state of Asp170, as well as the calculated relative energy barriers (see Figures 38 and 41, React-II and React-II'). During geometry optimisation of the models for React-I and React-II, Asp170 always abstracts a proton from Arg398. This is possibly an artefact caused by lacking interactions with residues missing in our active site model.

The reaction of p-creosol to vanillyl alcohol The energy profile and the optimised structures of all the species along the reaction are given in Figures 39 and 40, respectively. To our surprise, after the hydride transfer, only the O4 of FAD can act as the general base to abstract the proton of water molecule during the hydroxylation. We could not locate a transition state of the Asp170 as proton acceptor. All attempts converged to structures of the intermediate. Even more interestingly, a proton transfer can take place from FAD-O4H to Asp170 with an extremely low energy barrier, resulting in the formation of the final product of this half reaction (more discussion below). This reaction mechanism has feasible energy barriers on the basis of our calculations.

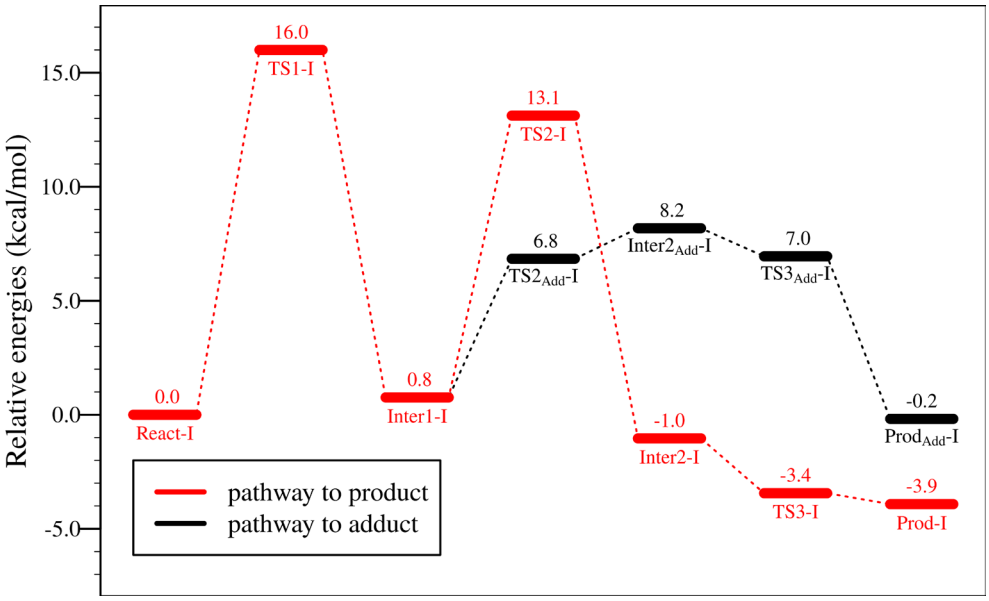


Figure 39: Energy profiles of the reaction of *p*-creosol to vanillyl alcohol. Labels are the same as used in Figure 37. Energies are given in kcal/mol. Starting from active site model React-I, the red solid lines indicate the stationary points of this reaction. The black solid lines indicate the stationary points of the adduct formation (to $Prod_{Add}$-I, described later in the text).

The first mechanistic step of the reaction from *p*-creosol to vanillyl alcohol proceeded via a *p*-quinone methide intermediate (React-I to Inter1-I, via TS1-I in Figures 39 and 40). During the initial step of the hydride transfer from the substrate to the FAD, the isoalloxazine ring twisted slightly to allow transfer of the hydride. The isoalloxazine ring did not become fully planar again with intermediate formation. The energy barrier of the hydride transfer from *p*-creosol to FAD was calculated to be 16.0 kcal/mol compared to React-I (Figure 39). In the corresponding transition state (TS1-I), the distance of the forming N-H bond is 1.3 Å, and that of the breaking C-H bond is 1.4 Å. The calculated energy of the resulting intermediate is 0.8 kcal/mol higher than the React-I, and the Asp170 residue is in its deprotonated state as a result of the proton transfer back to Arg398.

The second mechanistic step of this reaction, the hydroxylation of the *p*-creosol quinone intermediate (Inter1-

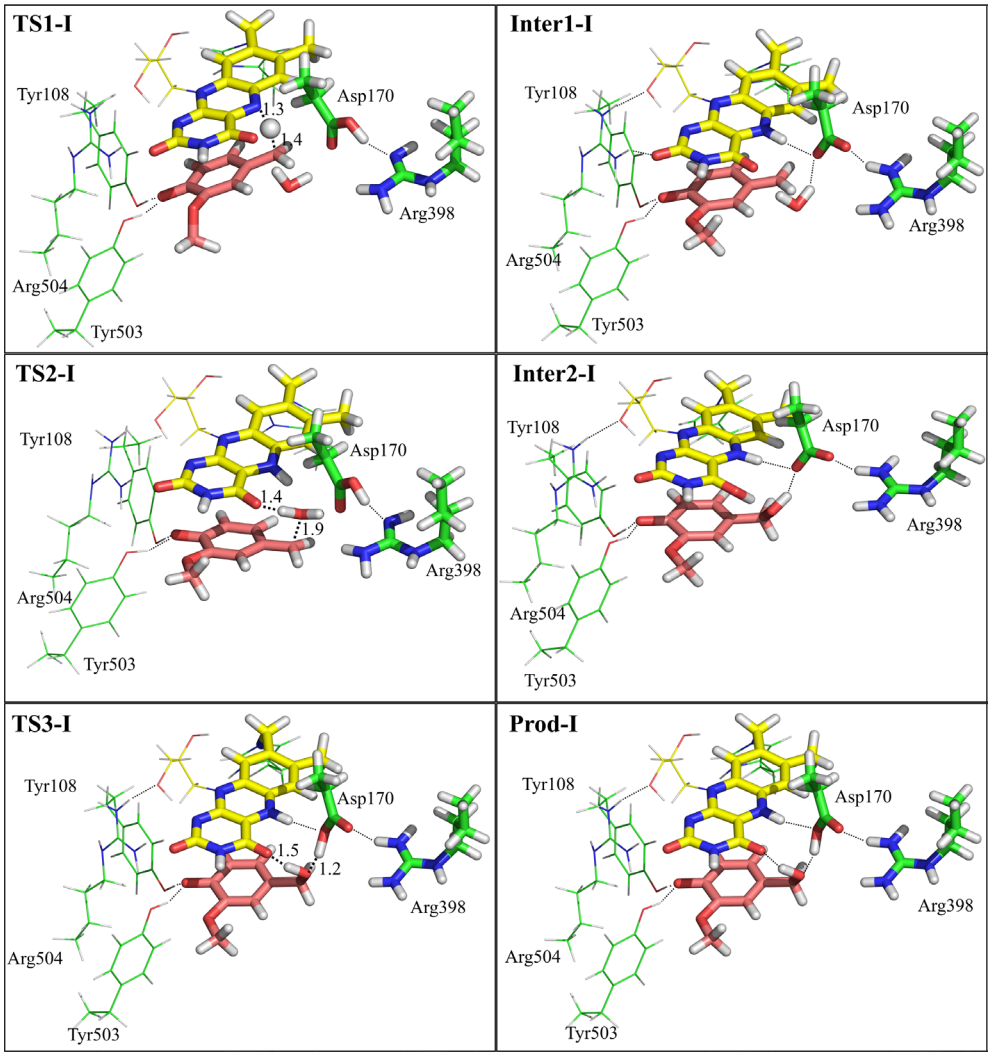


Figure 40: Optimised structures of the species involved in the suggested mechanism of the reaction of *p*-creosol to vanillyl alcohol. Distances are given in angstrom, with the thick dashed lines providing context for the transition state. Thin dashed lines show selected hydrogen bonds. Otherwise, the representation is the same as in Figure 38.

I in Figures 37B, 39 and 40), proceeded with an energy barrier of 13.1 kcal/mol compared to React-I. The formed intermediate (Inter2-I) is 1.0 kcal/mol lower in energy than React-I. The hydroxylation was assisted by the O4 of FAD, see mechanism in Figure 37B (in red). In the corresponding transition state (TS2-I), the distance of the forming O-

C bond is 1.9 Å, and that of the breaking O-H bond of the water molecule is still 1.0 Å. After the formation of the second intermediate (Inter2-I), this reaction proceeds via a proton transfer from the FAD-O4H to Asp170. In the corresponding transition state (TS3-I), the distance between the breaking O-H bond from FAD(H) to the substrate(O) is 1.5 Å, and that of the breaking O-H bond from substrate(O) to the substrate(H) is 1.2 Å (see Figure 40 for details). This step is calculated to be barrier-less, and the energy of the thus obtained product Prod-I is -3.9 kcal/mol lower in energy than React-I (Figure 39). It is should be noted here that for Inter1-I and Inter2-I we have also optimised the structures with protonated Asp170 residue. However, they are higher in energy compared to Inter1-I and Inter2-I, and thus are not reported here.

Alternative mechanisms For the substrate, different conformations have been optimised. One of them worth mentioning here is the one with the water molecule more "remote" from the substrate compared to React-I. We considered these conformations because different locations of this water molecule are observed in different crystal structures. This "remote" structure has a similar energy to the structure described above, and the calculated barrier of the hydride transfer starting from this substrate is also very similar to the one above. After the formation of the quinone intermediate, the "remote" conformation required the rearrangement of the water molecule to reach a conformation enabling hydroxylation of the substrate. This rearrangement resulted in identical conformations and energies as the reaction path described above. We have indications that a concerted mechanism for the productive reaction is energetically infeasible, with optimisations of the corresponding transition state always converging to the intermediate.

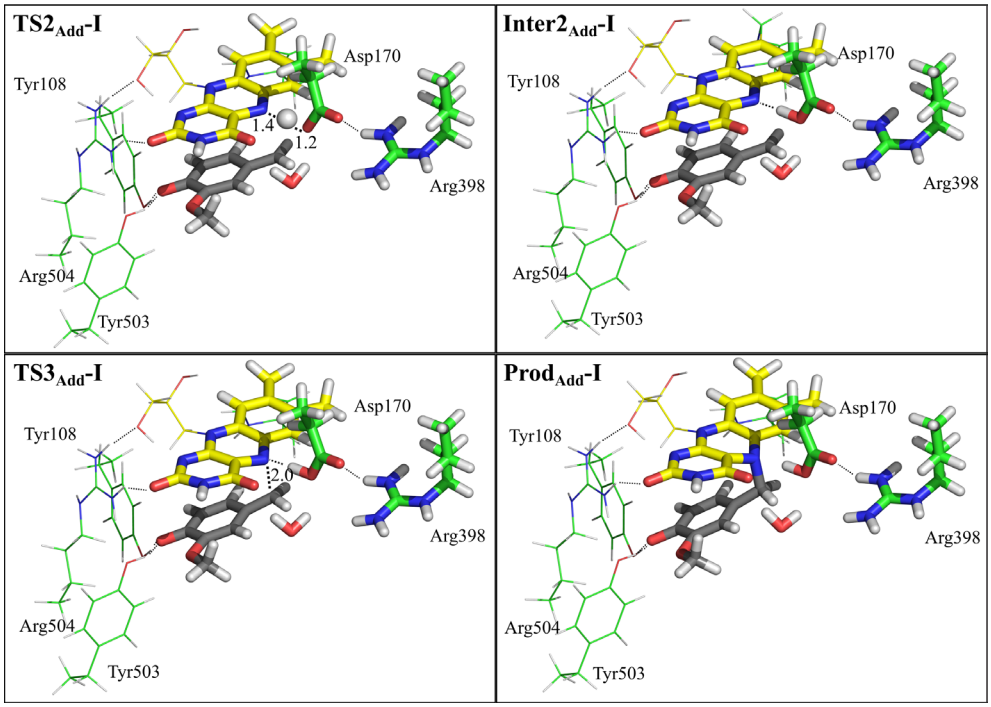


Figure 41: Optimised structures of the stationary points of the reaction of *p*-creosol to the covalent FAD-adduct. Labels are the same as used in Figure 39. Distances are given in angstrom, with the thick dashed lines providing context for the transition states. Thin dashed lines show selected hydrogen bonds. The *p*-creosol quinone intermediate is coloured in black. Otherwise, the representation is the same as in Figure 38 (React-I).

The reaction of p-creosol to the covalent adduct We now look at the pathway leading to adduct formation (see Figures 37A, 39 and 41, in black). The adduct can only be formed after the *p*-creosol quinone methide intermediate (in red) has been produced. Then, a proton transfer from the N5H to Asp170 takes place, followed by the covalent bond formation between the quinone intermediate and the cofactor. Two transition states are thus involved in adduct formation. The energy barrier of the proton transfer from N5H to Asp170 was calculated to be 6.8 kcal/mol relative to React-I. In the corresponding transition state ($TS2_{Add}$-I), the distance of the forming O-H bond is 1.2 Å, and that of the breaking N-H bond is 1.4 Å. The energy barrier of the formation of the C-N bond was calculated to be 7.0 kcal/mol relative to React-I. In the corresponding

transition state ($TS3_{Add}$-I), the distance of the forming C-N bond is 2.0 Å. The optimised structures of all the species along the reaction are given in Figure 41. It is very interesting that the calculated energy barriers of the adduct formation are lower than those of the productive reaction, and the formed adduct product $Prod_{Add}$-I is higher in energy by 3.7 kcal/mol, compared to Prod-I (see also Figure 39).

Alternative mechanisms We have indications that a concerted mechanism for adduct formation is energetically infeasible: all optimisations of the corresponding transition states converged to the intermediate. Water addition to the adduct was found to be energetically infeasible as well. We have also considered the attack of the substrate on FAD-N5-H as the first step, followed by a proton transfer to Asp170. However, this pathway was found to be not energetically viable. The mechanism where the substrate first attacks N5 was also found to be not energetically viable.

The reaction of vanillyl alcohol to vanillin For the reaction of vanillyl alcohol to vanillin we suggest the following reaction mechanism with feasible energy barriers on the basis of our calculations. In the first mechanistic step of this reaction the hydride transfer from the substrate to the FAD cofactor yielded reduced FAD and a quinone methide intermediate (blue reaction mechanism in Figure 37B, React-II to Inter-II). The energy barrier of the hydride transfer from vanillyl alcohol to FAD was calculated to be 10.8 kcal/mol relative to React-II (Figure 42, TS1-II). In the corresponding transition state (TS1-II), the distance of the forming N-H bond is 1.3 Å, and the breaking C-H bond is 1.3 Å (Figure 43). Note that the same twisting of the FAD cofactor during the hydride transfer as for the reaction from *p*-creosol to vanillyl alcohol is observed.

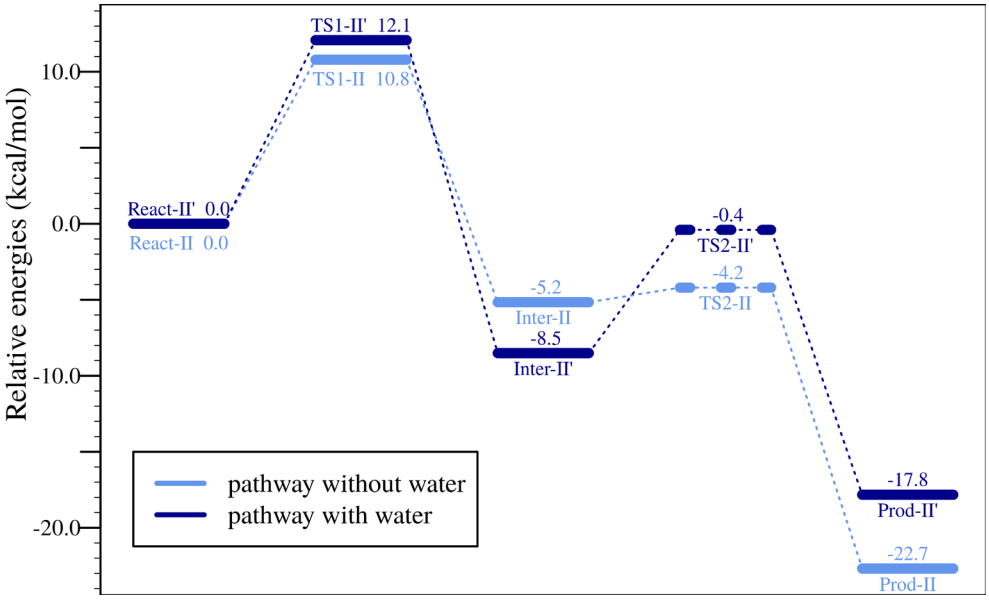


Figure 42: Energy profile of the stationary points of the reaction of vanillyl alcohol to vanillin. Labels are the same as used in Figure 37. Starting from the active site model, React-II, the light blue solid lines indicate the stationary points of the reaction without a water molecule present. The dark blue solid lines indicate the stationary points of the reaction with a water molecule present (React-II', described later in the text). The energies of TS2 are only indicative, and therefore shown as dashed lines. Note that the relative energies given for the pathways without and with water are only plotted in the same figure for easier comparison of the energy barriers. The energies of React-II and React-II' are not identical due to the additional water molecule in the pathway with water (II').

The second mechanistic step of this reaction, where a proton from the substrate is abstracted by Asp170, yielded the product vanillin (blue reaction mechanism in Figure 37B, Inter-II to Prod-II). The optimisation of the transition state for proton transfer from the quinone methide intermediate to Asp170 always failed. Therefore, the geometries shown in Figures 42 and 43 correspond to the highest point on the energy profile from the scan (TS2-II). The calculated energy of TS2-II is only 1.0 kcal/mol higher than Inter-III. The formed vanillin (Prod-II) is -22.7 kcal/mol lower in energy than React-II. This very low energy may indicate a limitation of the model, e.g missing interactions with other, second-shell residues.

Alternative mechanisms Asp170 could be optimised to be in its deprotonated form (without abstracting a proton from Arg398) if a water molecule is included in the model for this reaction (see React-II' in Figure 38). The calculated energy barrier of the hydride transfer was slightly in-

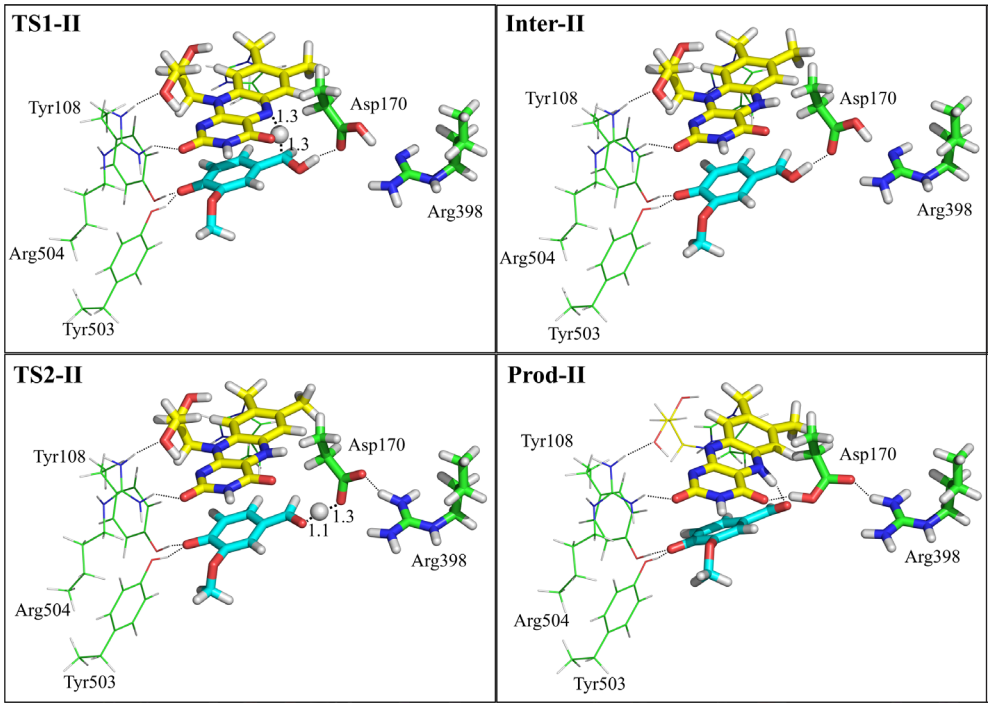


Figure 43: Optimised structures of the species involved in the suggested mechanism of the reaction of vanillyl alcohol to vanillin. Distances are given in angstrom, with the thick dashed lines providing context for the transition state. Thin dashed lines show selected hydrogen bonds. Otherwise, the representation is the same as in Figure 38 (React-II). Labels are the same as used in Figure 42.

creased in this model (12.1 kcal/mol). The intermediate (Inter-II') is lower in energy (-8.5 kcal/mol) compared to the calculations without the water molecule. Similar to the case starting with React-II, the transition state for the step of proton transfer could not be located, and we use also the structure from the scan here. The barrier of this step is higher than for the pathway without water, 8.1 kcal/mol higher than Inter-II'. The formed vanillin is -17.8 kcal/mol lower in energy than React-II'. The presence or absence of water has no major impact on the conformation of the residues in this model (see React-II vs React-II' in Figure 38 and Figures 43 vs 44).

Calculations with Asp170 protonated by abstracting a proton from Arg398 and a water molecule were also performed and resulted in an energy barrier of 12.9 kcal/mol

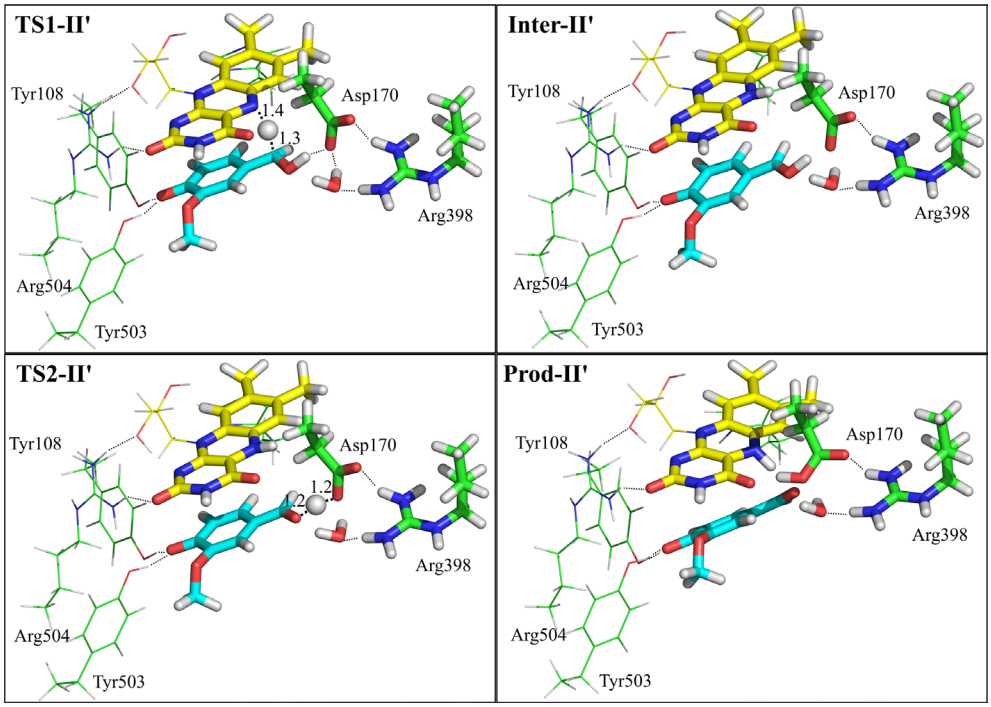


Figure 44: Optimised structures of the stationary points of the reaction of vanillyl alcohol to vanillin starting from React-II'. Distances are given in angstrom, with the thick dashed lines providing context for the transition state. Thin dashed lines show selected hydrogen bonds. Otherwise, the representation is the same as in Figure 38 (React-II'). Labels are the same as used in Figure 42.

for the hydride transfer (data not shown), which is slightly higher than the ones discussed above. Deprotonation of Asp170 occurred during formation of the product. The protonation state of Asp170 has no major impact on the conformation of the residues in this model. We have tested alternative reaction mechanisms and found that the intermediate formed by the proton transfer as a first mechanistic step cannot be located. We also have indications that a concerted mechanism is energetically infeasible, with optimisations of the corresponding transition state falling to the intermediate.

Discussion

We will go stepwise through the reactions we investigated in the present study. The overall reaction mechanisms for the conversion of *p*-creosol to vanillyl alcohol and the conversion of vanillyl alcohol to vanillin start out almost identical, with the hydride transfer from the substrate to the FAD, yielding reduced FAD. The hydride transfer from the substrate to FAD is the rate-limiting step for the conversion of *p*-creosol to vanillyl alcohol, as well as for the conversion of vanillyl alcohol to vanillin. In agreement with kinetic data, the energy barrier for this hydride transfer is higher in the *p*-creosol to vanillyl alcohol reaction: VAO converts 0.07 molecules of *p*-creosol per second, but 1.6 to 5.4 molecules of vanillyl alcohol (depending on the pH) [45].

Adduct formation vs productive reaction of p-creosol to vanillyl alcohol In the reaction of *p*-creosol to vanillyl alcohol, after the formation of the quinone methide intermediate, the productive reaction (to Prod-I) is competing with the reaction leading to the adduct ($Prod_{Add}$-I). In the hydroxylation reaction of the *p*-creosol quinone methide intermediate, Asp170 is the final proton acceptor for the proton of the water molecule. However, during adduct formation, Asp170 abstracts a proton from flavin N5H. We propose that in this step, the flexibility of Asp170 required to initiate adduct formation is crucial: VAO does not form an adduct with 4-alkylphenols with longer side chains, such as 4-ethylphenol [271]. Also, the adduct between *p*-cresol and FAD is much less affected by the variants mentioned above [45]. Bulkier substrates are likely preventing the flexibility of Asp170 and are therefore not forming an adduct. The interactions of Asp170 with surrounding amino acid residues could also further contribute to the flexibility of Asp170, allowing or forbidding adduct formation.

As discussed above, the energy barriers from the quinone methide intermediate to the adduct product are very low. The calculated barrier for the adduct formation is 6 kcal/mol lower than the productive reaction, and the adduct formation is therefore kinetically favored (see Figure 39). Also, the vanillyl alcohol product (Prod-I) has comparable energy to the adduct product ($Prod_{Add}$-I): they only differ by 3 kcal/mol (see Figure 39). Therefore, both products can be formed in this reaction.

Variants of VAO that show improved activity with *p*-creosol have been identified through directed evolution. All these variants have changed residues remote from the active site. It is of interest to discover how these variants influence the reaction of *p*-creosol to vanillyl alcohol. Do they slightly change the orientation of Asp170 in the active site, stabilising it and thus disabling its attack on FAD? Or do they influence the protonation state of Asp170? Do the variants influence the energy barriers of product and/or adduct formation, or do they influence the energies of the product and/or adduct? To more conclusively establish the details of adduct formation, we will proceed with a larger active site model in future work.

The reaction of p-creosol to vanillyl alcohol vs the reaction of vanillyl alcohol to vanillin The *p*-quinone methide intermediates formed in these two reactions have different lifetimes [223]. Our results are consistent with this experimental data: the intermediate of *p*-creosol has a longer lifetime than that of vanillyl alcohol [223]. We observe a lower energy barrier for Inter-II to Prod-II (1 kcal/mol), compared to the energy barrier of Inter1-I to Inter2-I (12.3 kcal/mol). Therefore, it is possible that adduct formation is further facilitated by the longer lifetime of Inter1-I (compared to Inter-II). Also, in crystal structures, a very similar adduct, the covalent *p*-cresol-FAD adduct, can be observed, indicating a long lifetime of this adduct [40]. We cannot compare the relative ener-

gies of the two intermediates Inter1-I and Inter-II directly because they are obtained from different active site models. However, it would be interesting to design calculations allowing the direct comparison of the energies of these intermediates.

Asp170 The protonation state (and thus the acidity) of Asp170 is influenced by Arg398 and a water molecule. Deprotonated Asp170 is crucial for the formation of vanillin, and its protonation state is also possibly influencing the formation of vanillyl alcohol from *p*-creosol. Care needs to be taken not to over-interpret the current results, as the active site model we have worked with is somewhat unreliably representing the acidity of Asp170. This points to a finely tuned, crucial network of interactions involving at least Asp170 and Arg398 that determines the acidity of Asp170. One interesting question remains: does Asp170 activate the water molecule in the *p*-creosol to vanillyl alcohol reaction without the direct involvement of FAD-O4, like the experimental papers propose? In our calculations with the current active site models, we see that the O4 of FAD rather than Asp170 abstracts a proton in water activation (see TS2-I in Figure 40), followed by a proton transfer process from FAD to Asp170. However, this could be due to the fact that the environment of Asp170-Arg398 pair was not treated properly. This can be addressed by using a larger active site model.

VAO vs PCMH A close relative of VAO, PCMH converts *p*-cresol without problems (*p*-creosol was never tested as a PCMH substrate). As can be seen in Figure 45, the active site of PCMH differs from VAO in several important aspects. Most notably, the active site base in VAO, Asp170, is replaced by an alanine, Ala154, in PCMH, while Arg398 is conserved in PCMH (Arg368). In PCMH, it is unclear if Glu427 or Glu380 is the active site base [56]. We will now discuss in more detail the active site composition of these two enzymes and their impact on activity based on the re-

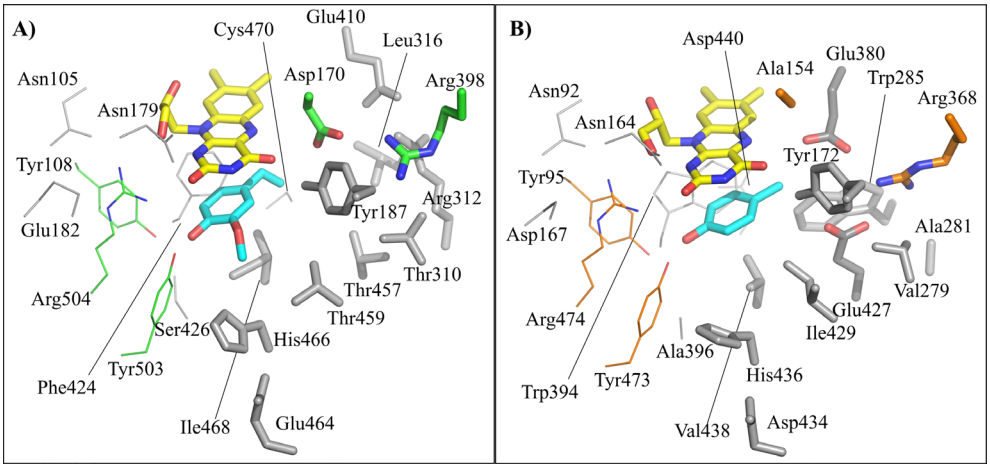


Figure 45: Active site residues in VAO (A) PDBID:2VAO) and PCMH (B) PDBD:1DII). Residues in the active site model for this study are shown in green or orange (VAO or the equivalent in PCMH, respectively). The ligands of the crystal structure are colored in cyan. The FAD cofactor is colored in yellow, and other, second shell, active site residues are colored in grey. Only amino acid side chains are shown for clarity.

sults obtained through our calculations.

In our calculations, we observe that the interaction with Arg398 is crucial for the protonation and therefore the activity of Asp170 as catalytic base. We also note that in most structures, the interaction with Arg398 alone is not sufficient to stabilise the deprotonated form of Asp170. It is likely that there is an entire network of interactions required to achieve this, possibly including Glu410 and Arg312 or more residues. It is interesting to note that Glu410 is conserved in PCMH as Glu380. Also, Tyr187 is conserved in PCMH as Tyr172, which suggests that the role of this residue in catalysis has been so far underestimated. Tyr187 could interact with Asp170 to stabilise it in its deprotonated form.

PCMH presents inverted stereoselectivity with 4-alkylphenols, producing (*S*)-stereoisomers, while VAO produces (*R*)-stereoisomers [195, 285]. The inversion of Asp170 and Thr457 in PCMH (to Ala170 and Glu457) is the cause of this changed enantioselectivity: the D170S/T457E double variant of VAO produces the (*S*)-stereoisomers [46]. These results point towards Glu427

being the catalytic base in PCMH. The reaction mechanism of PCMH also involves a *p*-quinone methide intermediate. Our calculations, together with the fact that PCMH does not form an adduct with *p*-cresol, suggest that Glu427 is unable to attack flavin N5H in the same manner Asp170 does, nor is Glu380. Both Glu380 and Asp170 are 3.6 Å away from flavin N5. It is therefore likely that the dual role Asp170 has in VAO [43, 46] is split between Glu380 and Glu427 in PCMH, with Glu380 being involved in the covalent binding of FAD and Glu427 being the catalytic base.

Conclusions and remaining open questions

The current active site model is insufficient for this reaction due to second shell residues that probably influence Asp170 significantly. It is therefore crucial to perform calculations with a larger active site model. Adduct formation is easy with *p*-creosol due to Asp170 flexibility, but not possible due to a more bulkier side chain of the substrate for the vanillyl alcohol reaction. The same is true for reactions with 4-alkylphenols that have longer side chains than *p*-creosol or *p*-cresol. Asp170 needs to be deprotonated after each step to be able to perform its function in a new reaction step. Can we identify and model the reactions of a H-bond network that releases the proton to the solvent? How many water molecules fit into the active site next to the substrate? Could additional water molecules influence the properties of Asp170? Can we somehow model the two-step reaction with one active site model, including oxidation of reduced flavin?

Computational Methods

We performed all calculations using the Gaussian 09 program [286] with the B3LYP hybrid density functional method [287, 288]. Geometries were optimised with the 6-31G(d,p) basis set. During the geometry optimisations, the truncated carbon and also some hydrogen atoms were fixed to prevent undesirable residue movements. Single-point energies were calculated with the SMD solvation model, using also the 6-31G(d,p) basis set [289] to take the effects of the protein environment into account: the part of the enzyme that is not included in the model is approximated by a polarisable homogeneous medium with a dielectric constant, here set to = 4. This model does not account for the heterogeneity of the enzyme surroundings. Systematic studies have shown that the effect of the solvation decreases quite rapidly with the size of the active site model [290, 291, 292, 293]. The optimised structures were then subjected to single-point energy calculations using the larger basis set 6-311+G(2d,2p) to obtain more accurate energies. All the calculations in the present study included dispersion corrections by using the DFT-D3(BJ) method, so using the D3 correction method of [294], with Becke-Jonson (BJ) damping [295]. Frequency calculations were performed with the 6-31G(d,p) basis set to obtain zero-point energies (ZPE). The relative energies shown in this work are therefore the large basis set energies (including dispersion effects) corrected for ZPE and solvation. The solvation correction was obtained by substracting energies calculated with the 6-31G(d,p) basis set from the energies calculated using the SMD solvation model.

Bibliography follows at the end of the thesis

Chapter 7:
Discussion

In this thesis, the flavoenzyme vanillyl alcohol oxidase (VAO) has been studied by computational and experimental techniques. **Chapter 2** puts all this work into a more global and practical context. There, we have identified, analysed and illustrated in several examples the challenges enzymes (and specifically oxidative enzymes or oxizymes) face, as well as the demands they have to meet before they can "*grow up*" to be industrial enzymes.

Enzymes could be the key to a more sustainable chemical industry. Especially oxizymes are of interest due to their ability to selectively incorporate oxygen atoms into molecules. In **Chapter 2**, the industrial usability of oxizymes was reviewed. Numerous examples of already applied oxizymes exist and many oxizymes are on track for applications. Nevertheless, the promise alone of "*potential industrial applications*" in many academic publications is not enough to make an application immediate reality. Industrial use of oxizymes is undoubtedly promising and feasible. However, economical costs and the impact of practical aspects, such as process design, make it difficult for academic scientists to judge the applicability of an enzyme.

A growing industrial interest in oxizymes is reflected in the number of patents filed between 2000 and 2015. A recent review has identified 369 patents related to biocatalysis applications using oxidoreductases that have been filed in that period [296]. It is unclear how many of these patents are about oxizymes (as defined in **Chapter 2**) because the distinction between dehydrogenases and oxizymes is not made. Most of these patents have been filed between 2013 and 2015 [296]. The same authors also see miscommunication between academic and industrial research as an obstacle: "*Communication between academia and industry, however, is often considered a structural weakness, and a general discrepancy between industry's short-term objectives and academia's long-term*

perspective has been noted"[296].

In the application of oxidases, two issues need special consideration. Because all oxidases require dioxygen in their reaction cycle, aeration is a crucial factor to take into account. Aeration of the reaction mixture on small scale is usually not a problem, but it becomes a limitation in large scale reactions. Adding pure dioxygen to the reaction on large scale includes considerable safety risks due to the combustibility of oxygen. Some applications use oxidases specifically for their ability to produce hydrogen peroxide. However depending on the application, hydrogen peroxide can also be a drawback, resulting in enzyme inactivation. Enzyme cascades using peroxidases or capturing the hydrogen peroxide with catalase, could help to resolve this drawback.

The application of VAO is appealing for different industries, namely the chemical, food, flavour and fragrance industry. For the chemical industry, the production of chiral 1-(4'-hydroxyphenyl)alcohols by VAO is of specific interest [195]. These compounds can be used as chiral building blocks for more complex chemicals including neuroprotective pharmaceuticals. The enantiomeric excess of the VAO-mediated hydroxylation reactions (94 and 80 % for the (*R*)- and (*S*)-stereoisomer, respectively; [46]) still needs to be increased to reduce downstream processing. Nevertheless, upscaling experiments in the INDOX project have already shown that 10 g enantiopure product can easily be obtained [297]. VAO can use eugenol to produce coniferyl alcohol, a molecule of interest for the fragrance industry and a compound which can also be used as a vanillin precursor [192]. To complete this list, VAO is of interest for the food industry because it can produce natural vanillin [178]. We have studied the VAO-catalysed conversion of *para*-creosol to vanillin in detail in **Chapters 5** and **6**.

Vanillin, the main constituent of the vanilla flavour, is produced in the pods of vanilla orchids, e.g. *Vanilla planifolia, V. pompona* or *V. tahitensis*. These orchids all belong to the genus *Vanilla* [298, 299]. In *V. planifolia*, one enzyme is converting ferulic acid to vanillin [300]. Only 1 kg of vanillin per 500 kg of vanilla pods is produced [300]. The process is therefore rather inefficient, but due to the high selling price of the obtained flavour it is nevertheless economically viable. For an overview of the different conventional production routes of vanilla from pods we refer the reader to [301], while a more up to date overview of the bioconversion or biosynthesis routes of natural vanillin can be found here [302]. Vanillin is currently mainly produced synthetically, but consumer demand for *"natural"* products has lead to the marketing of "*natural vanillin*". Chemically, both molecules are identical, but one is produced via chemical synthesis (e.g.: Rhovanil®Vanillin), while the other is microbiologically produced from a "*natural*" precursor (e.g. Rhovanil®Natural). This difference in production allows the producers to call it a "*natural flavour*", in agreement with European and US regulation [303]. This is how Nestlé has been able to claim in 2015 to replace all artificial colours and flavours in its confectionary sold in the US by the end of that year [304].

Concerning the "greenness" (see **Chapter 2**) of the different production routes for vanillin, it is unclear which one is actually more environmentally friendly and sustainable. It is essential to keep in mind that all these vanillin molecules are chemically identical and that it does not matter to our nose how they were produced.

As the work in this thesis was performed in an academic setting, the practical aspects of bringing an enzyme to industrial application(s) are not discussed beyond this point. After all, "*in a scientific pursuit there is continual food for discovery and wonder*" [305]. In that spirit, the remainder of this discussion deals with the more mechanistic properties of VAO. More than 20 years after the discovery of VAO, there is still much to be learned about the structure-function relationship of this enzyme. The explosion of sequence data since the beginning of this century has lead to a data revolution taking place. Surely, many VAO-like sequences are concealed in the vast

amounts of sequence data stored in databases. The challenge is to find the right ones. Most sequences are automatically annotated, and the thus annotated function is often suspect. Unfortunately, manual functional annotation, preferably by in depth biochemical characterisation, is much more time consuming. For VAO, misannotation is a serious problem because of the many VAO-like sequences that are annotated with ambiguous terms like hypothetical proteins, FAD-binding oxidoreductases, FAD binding domain-containing proteins and uncharacterized proteins, or they are annotated as glycolate oxidases, alcohol oxidases or aryl alcohol oxidases (see **Chapter 3**). Overall sequence identity of enzymes with VAO-like function (e.g. EUGO or PCMH) can be as low as 47 or 40 % sequence identity, or even lower amongst fungal VAOs, making functional predictions difficult (see **Chapter 3**). The use of critical amino acid residues to assign function would be more promising (see **Chapter 3**). Two different screening strategies were used in **Chapters 3** and **4**, and they are printed in this thesis in reverse order, with the work done last in **Chapter 3** and the work done first in **Chapter 4**.

There is a discrepancy between the amount of sequences and the sequences for which a function has been verified, resulting in more than one hundred million entries in the so-called RefSeq database, and only around 55'000 entries in the manually verified UniProtKB/Swiss-Prot database [306, 307]. Besides the obvious issue of misannoation, other biases are present in these databases, which are introduced by biases in high-throughput experimental data [308]. To resolve misannotations manually is tedious work. It would be much more effective if the annotation could be automatically resolved (just as it has been automatically annotated). This would require knowledge of where the (mis)annotation originated from to allow the correction to also use that origin. Another op-

tion would be to allow every user of the database to re-annotate sequences (manually or automated). This open access, open editing concept has worked quite well for wikipedia, with the exception of political topics (which protein sequences in general are not). Projects like Fold it (`https://fold.it/`) have an excellent track record when it comes to have "*the internet-crowd*" solve puzzles for science. Such projects create a different type of high-throughput experiments, using the interconnectedness of modern day society to their advantage.

Platforms like `https://www.zooniverse.org/` are public, freely available and host projects ranging from arts to physics. People-powered research is a tool that not only helps to resolve research questions, but that can help to increase acceptance of science in a more sceptical world.

We have analysed the natural variation of fungal enzymes that are putative VAOs in **Chapter 3** and have attempted to functionally express some of them in **Chapter 4**. In **Chapter 3**, an already filtered, high-quality set of sequences present in the mycocosm database was analysed. Only one sequence is present in both chapters, a sequence from *Dacryopinax* (SeqID M5FTW7). Based on the high-quality set of sequences in **Chapter 3** we also defined motifs of VAO-like sequences. Hopefully, these motifs will help to reduce the misannotations that exist for putative VAOs and inspire the characterisation of novel enzymes. We are confident that our functional predictions based on the available knowledge from experimental studies are accurate, but hope that there is still "*food for wonder*" in this scientific pursuit.

One important aspect that we hope will be answered by future studies is the physiological role of VAO in its host organisms. The only substrate that also induces expression of the *vao* gene in *P. simplicissimum*, 4-(methoxymethyl)phenol, is the proposed physiological substrate. The limited knowledge of the biological occurrence of this molecule however forbids any conclusions about the physiological role of VAO. A very general hypothesis could be that the organisms that produce VAOs are in symbiotic or parasitic relationships with organisms that degrade or produce lignin. Organisms that degrade

lignin often secrete enzymes for this purpose, leaving the lignin-degradation products available to other organisms that are themselves unable to degrade lignin. The phenolic compounds thus produced (either during the anabolism or catabolism of lignin) are substrates of VAOs. In catabolism, toxicity of phenolic compounds to fungal communities of lignin degraders may thus be prevented by organisms producing VAOs. These VAO-producers may have developed a tolerance for phenolic compounds by producing enzymes such as VAO [309].

In **Chapter 3**, we propose that fungal VAOs have been aquired through horizontal gene transfer of at least one bacterial ancestral VAO. While this explains how fungi with diverse lifestyles obtained VAOs, it does not explain why they kept these enzymes, and why some fungi even have multiple copies in their genome. More studies are required to understand the physiological role of VAO better. Proteomics studies of *P. simplicissimum* or other VAO producers under different growth conditions (using different, phenolic carbon sources) should be able to identify more enzymes that are co-produced with VAO and discover if VAO belongs to a specific metabolic network. It would be interesting to know if *P. simplicissimum* (or other fungi producing VAO) show an increased phenol tolerance due to the presence of VAO. Most putative VAO homologues identified in **Chapter 3** are present in the phylum Ascomycota, which contains brown-rot fungi. Only seven species of the phylum Basidiomycota contain VAO-like sequences, and none of them are white-rot fungi. Typically, brown-rot fungi, like *P. simplicissimum*, degrade hemicellulose and cellulose, while white-rot fungi degrade cellulose and lignin. It would be of interest to grow fungal communities of *P. simplicissimum* with white-rot fungi to find out if the addition of *P. simplicissimum* has a beneficial effect on lignin-degradation. Interestingly, VAO homologues identified in **Chapter 3** in the phylum Ascomy-

cota have diverse lifestyles, with only a few being wood decayers. We wonder what role VAO has in these organisms.

In **Chapter 4**, we searched the non-redundant sequence database and then selected genes for expression based on presence or absence of critical amino acid residues. Alas, expression of these putative *vao* genes in the yeast *Komagataella phaffii* resulted in VAOs containing an unusually stable anionic FAD semiquinone. VAOs produced in this organism have several other unexpected properties: they all bind their FAD cofactor non-covalently, are exclusively octameric and are unable to productively convert substrates. It could be that *K. phaffii* does not allow proper folding of VAOs. In VAO, FAD binding requires a highly pre-organised binding cavity [14, 41, 42]. We can safely say that this cavity is formed because we do observe FAD bound to the enzyme. However, it is doubtful that the cavity is organised correctly since the FAD cofactor is only bound non-covalently. In VAO, covalent incorporation of FAD happens via an autocatalytic process [41, 42]. It might be that the precise conformation to achieve covalent cofactor attachment is not reached in VAOs produced in *K. phaffii*, resulting in the observed anionic FAD semiquinone.

It would be of interest to study the autocatalytic mechanism of covalent FAD incorporation using quantum mechanical calculations. This would allow to answer several questions, namely: (i) what is the precise role of His61 in this process, and can the tyrosine residue observed at the same position in many VAO-like sequences perform the same function (see **Chapters 3** and **4**)? (ii) can a conformational change result in a "*non-productive*" autocatalytic reaction and lead to the formation of anionic FAD semiquinone? (iii) is it possible to change the covalent binding mode of VAO? For this, quantum chemical calculations for relatives of VAO with different FAD binding modes have to be computed (e.g. PCMH,

Protein folding is often depicted with a rugged, funnel-shaped free energy landscape. Misfolding of the protein means that the protein ends up in one of the grooves of the rugged surface instead of the global minimum. Misfolded proteins are a big problem in protein expression experiments, and screening efforts to achieve expression of functional proteins are considerable. Specialised consortia have been formed to resolve these issues, especially for the determination of crystal structures. In recent years however, *in silico, de novo* structure determination has become more and more reliable, as demonstrated in the Critical Assessment of Protein Structure Prediction (CASP) competition. For more information see their website and publications by the consortium: `http://predictioncenter.org/`

6-hydroxynicotine oxidase and glucooligosaccharide oxidase, which bind their FAD cofactor via a 8α-Tyr-bond, 8α-His-N1-linkage or bi-covalently [56, 219, 310, 311]).

If functional enzyme can be produced, experimental studies of its reaction mechanism are possible. In the case of VAO, several types of reactions are observed with different substrates (the reader is referred to Figures 16, 21 or 22 in **Chapter 3** or **4**). However, in the absence of functional enzyme, and if a crystal structure has been solved, computational calculations can be performed to study the enzyme *in silico* (see Figure 46 for an overview of possible approaches).

From a structure (or homology model) of an enzyme with unknown properties (top in Figure 46), blind docking studies can help identify the location of the active site. This is illustrated by the successful docking of the cyan molecule in the active site. The active site can also be identified by studying the migration of substrate(s) or product(s) in the enzymes, for example using PELE (see **Chapter 5**). Once an active site has been identified, the reaction mechanism can be studied using QCM or QM/MM (see **Chapter 6**). If the substrate scope of the enzyme is unknown, docking studies may help to narrow down the choices for in depth quantum mechanical calculations. If the location of the active site is not known, the methodology used in **Chapter 5** can identify it. Even if the location of the active site is known, it is of interest to discover the path(s) ligand(s) take to enter it. Although substrate binding is an often neglected and even trivialised [312] event in an enzymes reaction cycle, bottlenecks or blocked passages can profoundly change an enzyme's activity, see e.g. the gatekeeper residue in galactonolactone dehydrogenase [248].

In **Chapter 5**, we have explored the free energy landscape of VAO using PELE and studied how participants in

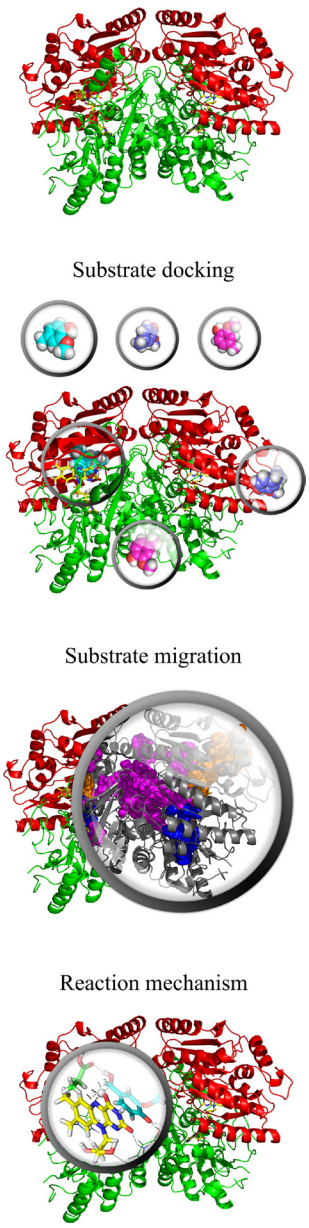


Figure 46: An overview of the different computational approaches that can be used to study enzymes. For details see text.

the reaction of *para*-creosol to vanillin enter and exit the enzyme's active site. We have found that phenolic ligands use a gated entry and exit path at the subunit interface of VAO, leading to the *si* side of the FAD cofactor in the active site (see Figure 47). The co-substrates dioxygen and hydrogen peroxide use a different entry and exit path at the interface of the two domains, leading to the *re*-side of the FAD cofactor in the active site (see Figure 47).

We have seen that protonated *p*-creosol can enter VAO without problems and that deprotonated *p*-creosol does not enter VAO. From our simulations, it is obvious that it is energetically unfavourable for deprotonated substrates to migrate through VAO. Deprotonated *p*-creosol can be forced to enter, and under these conditions we see a significant increase in energy when it is around 10 Å away from the FAD cofactor. This coincides with the location of the two concierge residues, His466 and Tyr503, thus suggesting the conclusion that these residues are part of the "*phenolate-exclusion mechanism*" in VAO. Without forcing deprotonated *p*-creosol to enter, it is energetically more favourable for it to leave to the solvent than to migrate to the FAD. Our results indicate that deprotonated *p*-creosol is unable to migrate through VAO efficiently.

VAO has a pH optimum at pH 10, and the phenolic substrates of VAO have a pK_a of around 10. This means that at pH 10, the amount of phenolate in solution is drastically increased (compared to pH 7). From our computational studies we know that the phenolate of *p*-creosol cannot enter the active site, and that therefore deprotonation of this substrate must occur in the active site. From previous work, it is known that the pK_a of the competitive inhibitor isoeugenol is different in solution than in VAO [36]: the pK_a of isoeugenol drops to values around 5 inside VAO, compared to values around 10 in solution. This suggests that under experimental conditions at pH 7.5, phenolic substrates are prevalently bound in the VAO active site in

Figure 47: Interfaces and faces in VAO. Top: the symmetric tetramer of dimers, middle: the dimer, with the two subunits (grey and coloured) and the two domains (red and green), bottom: the two faces of the FAD cofactor, *si* (with the substrate coloured in cyan, down into the plane of the paper) and *re* (up towards the viewer in the plane of the paper).

their phenolate form. Additionally, two reduction rates of FAD by chavicol were observed [49]. The faster of these rates was suggested to correspond to the reduction of FAD by the deprotonated substrate, while the slower one was assigned to the reduction of FAD by protonated substrate. The contribution of the faster reduction rate to the overall rate was also shown to increase with increasing pH.

Our simulations were performed at pH 7.5. Therefore, we cannot exclude that the behaviour of the phenolate form of substrates is different at pH 10.

Combining the experimental model with our computational model, we obtain the following, combined, model: protonated substrates migrate to the active site of VAO, where they are deprotontated. When increasing the pH of the reaction, substrate deprotonation in the active site is more favourable, leading to a faster overall reduction rate due to the increased contribution of the deprotonated substrate to the reaction. To verify this model, the calculations we have performed in **Chapter 6** can be extended using protonated substrate to check if the energy barriers of that reaction are in agreement with the slower reduction rate observed by Ewing *et al.* [49].

In **Chapter 6**, we have explored another part of the free energy landscape of VAO using quantum mechanical calculations. We have seen that several conformations and transition states can exist for one reaction, and not all lead to the formation of the desired product. Small random movements of active site residues and the substrate allow the exploration of different conformations. Formation of a "suicide" complex is a possibility, such as the adduct formed by *p*-creosol and FAD. The height of the energy barrier that needs to be crossed to lead to product formation determines if the reaction happens or not. One crucial mechanistic step of the reaction of VAO is the reduction of FAD by hydride transfer from the substrate. Because we work with deprotonated substrate only in our QM studies, we should observe energy barriers corresponding to rates in the order of the faster reduction rate observed by Ewing *et al.* [49]. Indeed, the calculated

energy barrier for the hydride transfer is 10.8 kcal/mol, and therefore more in agreement with the faster reduction rate.

We have observed certain limitations of the active site model used in **Chapter 6**. The model we built contains only some of the amino acid residues forming the active site, allowing a faster exploration of the free energy profiles of a reaction. We built a model containing 165 atoms to study the reductive half-reactions of the two-step reaction of *p*-creosol to vanillin. This model is missing several residues that interact with the catalytic base Asp170 and might influence its acidity. Specifically the influence of Arg398, but likely also Arg312, Glu410 and possibly Tyr187 on Asp170 needs to be studied in more detail. Preliminary data of VAO variants (R312A, R312H, R312K, R398A and E410A) indicates that the second shell residues Arg312, Arg398 and Glu410 do significantly influence enzyme activity [313, 314]. Activity of the R398A and E410A variants is absent with vanillyl alcohol, while eugenol is still converted. Also, the conservation of the positive charge in the R312H and R312K variants appears to conserve the (lowered) activity in these variants. The Michaelis constant (K_m) of vanillyl alcohol is at least doubled in these variants, whereas the k_{cat} of eugenol is decreased at least 5-fold.

Such different, changed activities of VAO-variants with different substrates are often observed, e.g in the VAO-variants created by directed evolution experiments [45]. Therefore, it is crucial to perform quantum chemical calculations with a larger model to refine the calculated energy barriers. Based on our results from **Chapter 6**, it would be very interesting to test the activity of the D170S/T457E double variant [46] with *p*-creosol. Although this variant has a very low activity with ethylphenol, testing its activity with *p*-creosol could be beneficial to our understanding of the reactivity of Asp170. We pre-

dict that this variant does not form an adduct with *p*-creosol because of the rearrangement of these two active site residues: Glu457, unlike Asp170, should be unable to attack the flavin to initiate adduct formation. In the bioinformatics studies of VAO and its closest homologues described in **Chapters 3** and **4** we see that a direct replacement of Asp170 by a glutamate does not occur without additional mutations of the enzyme. This is an indication that this mutation had detrimental effects. The only homologue with only Glu170 is the very suspicious sequence from the sordariomycete *Diaporthe ampelina*. Therefore it is not recommended to only change Asp170, but to either create the T457D variant (as observed in some of the sequences in **Chapter 3**) or study the D170S/T457E double variant in more detail.

It is essential that further computational and experimental studies use the same substrates to allow better comparison of results and include appropriate controls. Calculating properties of already existing variants or explaining already known properties of VAO by using computational techniques is of course valuable, but it can contain an inherent bias. To allow an unbiased interpretation of computational and experimental results, the experimental and computational work needs to be kept strictly apart before publication (similar to the CASP competition). Publishing the results together would defeat the purpose of eliminating possible bias (or even invite the reproach of bias). In the context of a CASP-like competition, publication in a special issue where all experimental and computational results are included would be desirable. To study the mechanistics of variants discovered by van den Heuvel *et al.* [45], the following approach would be beneficial: perform detailed studies of reduction rates as well as calculations of the corresponding energy barriers, ensuring no results or conclusions are exchanged and then publish the results separately. The vari-

ants discovered by van den Heuvel *et al.* [45] are especially challenging because they are remote from the active site. Therefore, different models and hypotheses will have to be tested computationally and experimentally. Such "*battles of wits*" between the experimental and computational community would surely benefit the field and improve our understanding of sequence-structure-function relationships immensely.

A computational outlook

Currently, many different computational tools are available to study different steps of an enzyme reaction. For computational studies of ligand migration, online tools like CAVER or PELE are available, as are free software packages like GROMACS or CHARMM [67, 315, 316, 317, 318]. For computational studies of enzyme catalysis, quantum chemical modelling (QCM) or quantum mechanics/molecular mechanics (QM/MM) calculations can be performed using, e.g., GAUSSIAN or PELE [286, 67]. Depending on the research question, different tools have to be chosen. We have used PELE, a Monte Carlo (MC) based program to study ligand migration in VAO in **Chapter 5**, and QCM in **Chapter 6**. It is important to know that the quantum chemical cluster approach or all-QM approach we have used is entirely different from a QM/MM approach, as explained in **Chapter 1**.

Enzyme activity is significantly affected by solvent effects. Several concepts to study the effects of solvents on enzyme activity have been developed or are currently under development [319, 320, 321, 322, 323, 324, 322, 325, 326, 327, 328, 329, 330]. Work is ongoing to combine these concepts into an easy to use tool for experimentalists. It is therefore becoming feasible to computationally model complete enzyme reactions and enzyme mechanisms. However, at least three different tools or software

packages (one for MD or MC, one for QCM or QM/MM and one to study solvent effects) need to be used to do so, presenting a considerable barrier for expert and non-expert users. Contributing to this barrier is the fact that different methodologies are behind these tools or software packages, and consequently the correct interpretation of their results requires in depth understanding of each methodology.

Nevertheless, computational tools are becoming more and more user-friendly and the times where intense programming efforts were required by the user are over. Most of the tools mentioned above are also freely available (at least to academic users). Unfortunately, this free availability also has a downside: changing availability [331] and lack of stable technical assistance. The costs of maintaining the code behind these tools is often not covered by funding schemes, and maintenance of the generated data is left to the personal preference of the user that created it.

The future of enzymes as the key to a more sustainable chemical industry requires not only the better communication between industry and academia, but also the more integrated use of computational tools. In the last few decades, biotechnology has revolutionised experimental work. Kits for cloning of genes, protein expression and protein characterisation have made biotechnology successful by making it accessible and easy to use. The data revolution triggered by high-throughput genome sequencing (which was in turn a consequence of biotechnology revolution) has seen equal success with easy to use, online bioinformatics tools. Examples include BLAST, Clustalomega, DeepLoc, SWISS-MODEL and EFI-EST as well as sequences databases like Ensembl and Swiss-Prot. These services are maintained by the European Bioinformatics Institute (EMBL-EBI), the Enzyme Function Initiative (EFI) or the Swiss Insititute of Bioinformatics (SIB). Currently, the only databases available for structural in-

formation are the PDB for experimentally determined structures or smaller databases for homology models (e.g. ModBase). Several databases or servers building on the structural information in the PDB exist (e.g. Molecular Modeling Database or Inferred Biomolecular Interactions Server). However, no databases for MD, MC, QCM or QM/MM calculations exist.

Increasing and improving the availability of computational tools and increasing acceptance by experimentalists will lead to an increase of data generated by MD, MC, QM or QM/MM. It would be highly beneficial to require that data generated using freely available tools is also made freely available, e.g. under an ODC Open Database License. Data should also be Findable, Accessible, Interoperable and Reuseable (F.A.I.R.) [332, 333, 334].

F.A.I.R.: Data has to be **Findable**: data need to be assigned a unique and persistent identifier, described with rich metadata, registered or indexed in a searchable source and have a data identifier specified by metadata. Data has to be **Accessible**: data need to be retrievable by their identifier via a standardised, open, free and universally implementable communications protocol and the metadata need to be accessible even if the data are no longer available. Data has to be **Interoperable**: data need to use vocabulary following F.A.I.R principles in a formal, accessible, shared and broadly applicable language and include qualified references to other data. Data has to be **Reusable**: data need to be associated with their provenance and meet domain-relevant community standards.

Conclusion

An enzymatic reaction is a multistep process, involving substrate entrance, substrate diffusion to the active site, catalysis, product diffusion to the solvent, and product exit. It is almost impossible to separate these events in experimental data when studying enzyme mechanisms, even in stopped flow experiments. Computational studies boost insight into these events and allow to interpret experimental data in a more complete way (see **Chapters 5** and **6**). Healthy sceptisism from experimentalists towards computational data, and vice versa, will ensure that the most realistic models of enzyme reactions can be defined. It should never be forgotten that enzyme kinetics is also "*only a model*" of the complex process that is enzyme catalysis. A plethora of sequence information is available and we need to find better ways to annotate it (see **Chapters 3** and **4**). The challenge we are facing in this century is not data collection, but data management and analysis.

Bibliography

[1] S. E. Woosley, A. Heger, and T. A. Weaver. The evolution and explosion of massive stars. *Rev. Mod. Phys.*, 74:1015–1071, 2002.

[2] B. S. Meyer, L. R. Nittler, A. N. Nguyen, and S. Messenger. Nucleosynthesis and chemical evolution of oxygen. *Rev. Mineral. Geochem.*, 68(1):31–53, 2008.

[3] T. W. Lyons, C. T. Reinhard, and N. J. Planavsky. The rise of oxygen in earth's early ocean and atmosphere. *Nat. J.*, 506(7488):307–315, 2014.

[4] W. Martin and K. V. Kowallik. Annotated english translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *Eur. J. Phycol.*, 34(3):287—295, 1999.

[5] J. A. Raven and J. F. Allen. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol*, 4(3):209, 2003.

[6] P. J. Keeling. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci*, 365(1541):729–748, 2010.

[7] P. Siekevitz. Powerhouse of the cell. *Sci. Am.*, 197(1):131–144, 1957.

[8] O. E. Dictionary. *Oxford living Dictioniaries*. Oxford University Press, 6 2017. Cofactor [Def. 2].

[9] A. Beier, S. Bordewick, M. Genz, *et al.* Switch in cofactor specificity of a Baeyer–Villiger monooxygenase. *ChemBioChem*, 17(24):2312–2315, 2016.

[10] K. Gruber and B. Kräutler. Coenzyme B12 repurposed for photoregulation of gene expression. *Angew. Chem. Int. Ed.*, 55(19):5638–5640, 2016.

[11] A. Lerchner, A. Jarasch, W. Meining, A. Schiefner, and A. Skerra. Crystallographic analysis and structure-guided engineering of NADPH-dependent *Ralstonia* sp. alcohol dehydrogenase toward NADH cosubstrate specificity. *Biotechnol. Bioeng.*, 110(11):2803–2814, 2013.

[12] M. Auer, C. Gruber, M. Bellei, *et al.* A stable bacterial peroxidase with novel halogenating activity and an autocatalytically linked heme prosthetic group. *J. Biol. Chem.*, 288(38):27181–27199, 2013.

[13] W. J. H. van Berkel. Flavoenzymes, chemistry of. In T. P. Begley, editor, *Wiley Encyclopedia of Chemical Biology*. John Wiley & Sons, Inc., 2007.

[14] N. Tahallah, R. H. H. van den Heuvel, W. A. M. van den Berg, *et al.* Cofactor-dependent assembly of the flavoenzyme vanillyl-alcohol oxidase. *J. Biol. Chem.*, 277(39):36425–36432, 2002.

[15] A. Winkler, F. Hartner, T. M. Kutchan, A. Glieder, and P. Macheroux. Biochemical evidence that berberine bridge enzyme belongs to a novel family of flavoproteins containing a bi-covalently attached FAD cofactor. *J. Biol. Chem.*, 281(30):21276–21285, 2006.

[16] M. Huijbers, M. Martínez-Júlvez, A. Westphal, *et al.* Proline dehydrogenase from *Thermus thermophilus* does not discriminate between FAD and FMN as cofactor. *Sci. Rep.*, 7(43880), 2017.

[17] J. K. Eweg, F. Müller, and W. J. H. van Berkel. On the enigma of old yellow enzyme's spectral properties. *Eur. J. Biochem.*, 129(2):303–316, 1982.

[18] D. Zanette, H. L. Monaco, G. Zanotti, and P. Spadon. Crystallization of hen eggwhite riboflavin-binding protein. *J. Mol. Biol.*, 180(4):1185–1187, 1984.

[19] S. A. Marshall, K. Fisher, A. Ni Cheallaigh, *et al.* Oxidative maturation and structural characterization of prenylated-FMN binding by UbiD, a decarboxylase involved in bacterial ubiquinone biosynthesis. *J. Biol. Chem.*, 292(11):4623–4637, 2017.

[20] C. T. Walsh and T. A. Wencewicz. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.*, 30:175–200, 2013.

[21] P. Macheroux, B. Kappes, and S. E. Ealick. Flavogenomics – a genomic and structural view of flavin-dependent proteins. *FEBS J.*, 278(15):2625–2634, 2011.

[22] M. W. Fraaije and A. Mattevi. Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.*, 25(3):126–132, 2000.

[23] M. L. Mascotti, M. J. Ayub, N. Furnham, J. M. Thornton, and R. A. Laskowski. Chopping and changing: the evolution of the flavin-dependent monooxygenases. *J. Mol. Biol.*, 428(15):3131–3146, 2016.

[24] V. Joosten and W. J. van Berkel. Flavoenzymes. *Curr. Opin. Chem. Biol.*, 11(2):195–202, 2007.

[25] W. P. Dijkman, G. de Gonzalo, A. Mattevi, and M. W. Fraaije. Flavoprotein oxidases: classification and applications. *Appl. Microbiol. Biotechnol.*, 97(12):5177–5188, 2013.

[26] C. A. McDonald, R. L. Fagan, F. Collard, V. M. Monnier, and B. A. Palfey. Oxygen reactivity in flavoenzymes: Context matters. *J. Am. Chem. Soc.*, 133(42):16809–16811, 2011.

[27] A. Mattevi. To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes. *Trends Biochem. Sci.*, 31(5):276–283, 2006.

[28] V. Massey. Activation of molecular oxygen by flavins and flavoproteins. *J. Biol. Chem.*, 269(36):22459–62, 1994.

[29] M. W. Fraaije, K. A. Sjollema, M. Veenhuis, and W. J. van Berkel. Subcellular localization of vanillyl-alcohol oxidase in *Penicillium simplicissimum*. *FEBS Lett.*, 422(1):65–68, 1998.

[30] G. Groeger, C. Quiney, and T. G. Cotter. Hydrogen peroxide as a cell-survival signaling molecule. *Antiox. & Redox Signal.*, 11:2655–2671, Oct 2009.

[31] E. de Jong, W. J. H. van Berkel, R. P. van der Zwan, and J. A. M. de Bont. Purification and characterization of vanillyl-alcohol oxidase from *Penicillium simplicissimum*. *Eur. J. Biochem.*, 208(3):651–657, 1992.

[32] E. de Jong, E. E. Beuling, R. P. van der Zwan, and J. A. M. de Bont. Degradation of veratryl alcohol by *Penicillium simplicissimum*. *Appl. Microbiol. Biotechnol.*, 34(3):420–425, 1990.

[33] H. Wariishi and M. H. Gold. Lignin peroxidase compound III: Formation, inactivation, and conversion to the native enzyme. *FEBS Lett.*, 243(2):165–168, 1989.

[34] H. Furukawa, M. Wieser, H. Morita, T. Sugio, and T. Nagasawa. Purification and characterization of vanillyl-alcohol oxidase from *Byssochlamys fulva* V107. *J. Biosci. Bioeng.*, 87(3):285–290, 1999.

[35] M. W. Fraaije, M. Pikkemaat, and W. J. H. van Berkel. Enigmatic gratuitous induction of the covalent flavoprotein vanillyl-alcohol oxidase in *Penicillium simplicissimum*. *Appl Environ Microbiol*, 63(2):435–439, 1997.

[36] M. W. Fraaije, C. Veeger, and W. J. H. van Berkel. Substrate specificity of flavin-dependent vanillyl-alcohol oxidase from *Penicillium simplicissimum*: evidence for the production of 4-hydroxycinnamyl alcohols from 4-allylphenols. *Eur. J. Biochem.*, 234(1):271–277, 1995.

[37] R. H. H. van den Heuvel, M. W. Fraaije, C. Laane, and W. J. H. van Berkel. Regio- and stereospecific conversion of 4-alkylphenols by the covalent flavoprotein vanillyl-alcohol oxidase. *J. Bacteriol.*, 180(21):5646–5651, 1998.

[38] W. J. H. van Berkel, R. H. H. van den Heuvel, C. Versluis, and A. J. R. Heck. Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Sci.*, 9(3):435–439, 2000.

[39] T. A. Ewing, G. Gygli, and W. J. H. van Berkel. A single loop is essential for the octamerization of vanillyl alcohol oxidase. *FEBS J.*, 283(13):2546–2559, 2016.

[40] A. Mattevi, M. W. Fraaije, A. Mozzarelli, *et al.* Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. *Struct.*, 5(7):907–920, 1997.

[41] M. W. Fraaije, R. H. H. van den Heuvel, W. J. H. van Berkel, and A. Mattevi. Structural analysis of flavinylation in vanillyl-alcohol oxidase. *J. Biol. Chem.*, 275(49):38654–38658, 2000.

[42] J. Jin, H. Mazon, R. H. van den Heuvel, *et al.* Covalent flavinylation of vanillyl-alcohol oxidase is an autocatalytic process. *FEBS J.*, 275(20):5191–5200, 2008.

[43] R. H. H. van den Heuvel, M. W. Fraaije, A. Mattevi, and W. J. H. van Berkel. Asp-170 is crucial for the redox properties of vanillyl-alcohol oxidase. *J. Biol. Chem.*, 275(20):14799–14808, 2000.

[44] M. W. Fraaije, R. H. H. van den Heuvel, W. J. H. van Berkel, and A. Mattevi. Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase. *J. Biol. Chem.*, 274(50):35514–35520, 1999.

[45] R. H. H. van den Heuvel, W. A. M. van den Berg, S. Rovida, and W. J. H. van Berkel. Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *J. Biol. Chem.*, 279(32):33492–33500, 2004.

[46] R. H. H. van den Heuvel, M. W. Fraaije, M. Ferrer, A. Mattevi, and W. J. H. van Berkel. Inversion of stereospecificity of vanillyl-alcohol oxidase. *Proc. Natl. Acad. Sci.*, 97(17):9455–9460, 2000.

[47] A. J. Jakobi, D. M. Passon, K. Knoops, *et al.* *In cellulo* serial crystallography of alcohol oxidase crystals inside yeast cells. *IUCrJ.*, 3(2):88–95, Mar 2016.

[48] S. Tsujino and T. Tomizaki. Ultrasonic acoustic levitation for fast frame rate X-ray protein crystallography at room temperature. *Sci. reports*, 6(25558), 2016.

[49] T. A. Ewing, Q.-T. Nguyen, R. C. Allan, *et al.* Two tyrosine residues, Tyr-108 and Tyr-503, are responsible for the deprotonation of phenolic substrates in vanillyl-alcohol oxidase. *J. Biol. Chem.*, 292(35):14668–14679, 2017.

[50] H. Furukawa, M. Wieser, H. Morita, T. Sugio, and T. Nagasawa. Purification and characterization of eugenol dehydrogenase from *Pseudomonas fluorescens* E118. *Arch. Microbiol.*, 171(1):37–43, 1998.

[51] K. Brandt, S. Thewes, J. Overhage, H. Priefert, and A. Steinbüchel. Characterization of the eugenol hydroxylase genes (*ehyA*/*ehyB*) from the new eugenol-degrading *Pseudomonas* sp. strain OPS1. *Appl. Microbiol. Biotechnol.*, 56(5):724–730, 2001.

[52] H. Priefert, J. Overhage, and A. Steinbüchel. Identification and molecular characterization of the eugenol hydroxylase genes (*ehyA*/*ehyB*) of *Pseudomonas* sp. strain HR199. *Arch. Microbiol.*, 172(6):354–363, 1999.

[53] S. F. Brady, C. J. Chao, and J. Clardy. New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.*, 124(34):9968–9969, 2002.

[54] S. Rachid, O. Revermann, C. Dauth, U. Kazmaier, and R. Müller. Characterization of a novel type of oxidative decarboxylase involved in the biosynthesis of the styryl moiety of chondrochloren from an acylated tyrosine. *J. Biol. Chem.*, 285(17):12482–12489, 2010.

[55] J. Jin, H. Mazon, R. H. H. van den Heuvel, D. B. Janssen, and M. W. Fraaije. Discovery of a eugenol oxidase from *Rhodococcus* sp. strain RHA1. *FEBS J.*, 274(9):2311–2321, 2007.

[56] L. M. Cunane, Z.-W. Chen, N. Shamala, *et al.* Structures of the flavocytochrome *p*-cresol methylhydroxylase and its enzyme-substrate complex: gated substrate entry and proton relays support the proposed catalytic mechanism. *J. Mol. Biol.*, 295(2):357–374, 2000.

[57] M. Edmunds. The antikythera mechanism and the mechanical universe. *Contemp. Phys.*, 55(4):263–285, 2014.

[58] T. Freeth, Y. Bitsakis, X. Moussas, *et al.* Decoding the ancient greek astronomical calculator known as the antikythera mechanism. *Nat. J.*, 444:587–591, 2006.

[59] T. Freeth, A. Jones, J. M. Steele, and Y. Bitsakis. Calendars with olympiad display and eclipse prediction on the antikythera mechanism. *Nat. J.*, 454:614–617, 2008.

[60] B. Kintses, L. D. van Vliet, S. R. Devenish, and F. Hollfelder. Microfluidic droplets: new integrated workflows for biological experiments. *Curr. Opin. Chem. Biol.*, 14(5):548 – 555, 2010.

[61] A. M. Sydor, K. J. Czymmek, E. M. Puchner, and V. Mennella. Super-resolution microscopy: from single molecules to supramolecular assemblies. *Trends Cell Biol.*, 25(12):730–748, 2015.

[62] N. L. Samara, Y. Gao, J. Wu, and W. Yang. Chapter twelve - detection of reaction intermediates in Mg^{2+}-dependent DNA synthesis and RNA degradation by time-resolved X-ray crystallography. In B. F. Eichman, editor, *DNA Repair Enzymes: Structure, Biophysics, and Mechanism*, volume 592 of *Methods in Enzymology*, pages 283–327. Academic Press, 2017.

[63] H. N. Chapman, P. Fromme, A. Barty, *et al.* Femtosecond X-ray protein nanocrystallography. *Nat. J.*, 470:73–77, 2011.

[64] Nobel prize in chemistry 2017, scientific background: The development of cryo-electron microscopy. `https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/advanced.html`. Accessed: 02-12-2017.

[65] M. Shibata, H. Nishimasu, N. Kodera, *et al.* Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat. Commun.*, 8(1430), 2017.

[66] J. D. Hamblin. *Science in the Early Twentieth Century: an Encyclopedia.* ABC-CLIO, 2005.

[67] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar. PELE: protein energy landscape exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput.*, 1(6):1304–1311, 2005.

[68] D. Lecina, J. F. Gilabert, and V. Guallar. Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep.*, 7(8466), 2017.

[69] F. Himo. Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.*, 139(20):6780–6786, 2017.

[70] S. Kara, J. H. Schrittwieser, F. Hollmann, and M. B. Ansorge-Schumacher. Recent trends and novel concepts in cofactor-dependent biotransformations. *Appl. Microbiol. Biotechnol.*, 98(4):1517–1529, 2014.

[71] F. Hollmann, I. W. C. E. Arends, K. Buehler, A. Schallmey, and B. Buhler. Enzyme-mediated oxidations for the chemist. *Green Chem.*, 13:226–265, 2011.

[72] T. A. Ewing, M. W. Fraaije, and W. J. H. van Berkel. Oxidation using alcohol oxidases. In K. Faber, editor, *Biocatalysis in Organic Synthesis 3*, pages 157–186. Georg Thieme Verlag KG, 2015.

[73] D. T. Pazmiño, M. Winkler, A. Glieder, and M. Fraaije. Monooxygenases as biocatalysts: classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.*, 146(1):9–24, 2010.

[74] M. M. Huijbers, S. Montersino, A. H. Westphal, D. Tischler, and W. J. van Berkel. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.*, 544:2–17, 2014.

[75] J. D. Lipscomb. Mechanism of extradiol aromatic ring-cleaving dioxygenases. *Curr. Opin. Struct. Biol.*, 18(6):644–649, 2008.

[76] S. Fetzner. Ring-cleaving dioxygenases with a cupin fold. *Appl. Environ. Microbiol.*, 78(8):2505–2514, 2012.

[77] U. Guzik, K. Hupert-Kocurek, and D. Wojcieszyńska. Intradiol dioxygenases - the key enzymes in xenobiotics degradation. In R. Chamy and F. Rosenkranz, editors, *Biodegradation of Hazardous and Special Products*, pages 129–153. InTech, Rijeka, 2013.

[78] C. M. Bianchetti, C. H. Harmann, T. E. Takasuka, *et al.* Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic *Streptomyces* sp. SirexAA-E. *J. Biol. Chem.*, 288(25):18574–18587, 2013.

[79] A. Karlsson, J. V. Parales, R. E. Parales, *et al.* Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science.*, 299(5609):1039–1042, 2003.

[80] S. H. Knauer, O. Hartl-Spiegelhauer, S. Schwarzinger, P. Hänzelmann, and H. Dobbek. The Fe(II)/α-ketoglutarate-dependent taurine dioxygenases from *Pseudomonas putida* and *Escherichia coli* are tetramers. *FEBS J.*, 279(5):816–831, 2012.

[81] K. Koketsu, Y. Shomura, K. Moriwaki, *et al.* Refined regio- and stereoselective hydroxylation of L-pipecolic acid by protein engineering of L-proline cis-4-hydroxylase based on the X-ray crystal structure. *ACS Synth. Biol.*, 4(4):383–392, 2015.

[82] R. Ullrich, J. Nüske, K. Scheibner, J. Spantzel, and M. Hofrichter. Novel haloperoxidase from the agaric basidiomycete *Agrocybe aegerita* oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.*, 70(8):4575–4581, 2004.

[83] M. Hofrichter and R. Ullrich. Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem. Biol.*, 19:116–125, 2014.

[84] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, *et al.* Engineering the third wave of biocatalysis. *Nat. J.*, 485(7397):185–194, 2012.

[85] S. Luetz, L. Giver, and J. Lalonde. Engineered enzymes for chemical production. *Biotechnol. Bioeng.*, 101(4):647–653, 2008.

[86] H.-P. Meyer, E. Eichhorn, S. Hanlon, *et al.* The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC). *Catal. Sci. Technol.*, 3:29–40, 2013.

[87] P. W. Sutton, J. P. Adams, I. Archer, *et al.* Biocatalysis in the fine chemical and pharmaceutical industries. In *Practical Methods for Biocatalysis and Biotransformations 2*, pages 1–59. John Wiley & Sons, Ltd, 2012.

[88] S. M. Thomas, R. DiCosimo, and V. Nagarajan. Biocatalysis: applications and potentials for the chemical industry. *Trends Biotechnol.*, 20(6):238–242, 2002.

[89] A. Schmid, J. S. Dordick, B. Hauer, *et al.* Industrial biocatalysis today and tomorrow. *Nat. J.*, 409(6817):258–268, 2001.

[90] P. T. Anastas and J. B. Zimmerman. Peer reviewed: design through the 12 principles of green engineering. *Environ. Sci. & Technol.*, 37(5):94–101, 2003.

[91] S. Y. Tang, R. A. Bourne, R. L. Smith, and M. Poliakoff. The 24 principles of green engineering and green chemistry: "IMPROVEMENTS PRODUCTIVELY". *Green Chem.*, 10:268–269, 2008.

[92] Y. Ni, D. Holtmann, and F. Hollmann. How green is biocatalysis? To calculate is to know. *ChemCatChem*, 6(4):930–943, 2014.

[93] R. G. Mathys, A. Schmid, and B. Witholt. Integrated two-liquid phase bioconversion and product-recovery processes for the oxidation of alkanes: process design and economic evaluation. *Biotechnol. Bioeng.*, 64(4):459–477, 1999.

[94] D. Kuhn, M. A. Kholiq, E. Heinzle, B. Buhler, and A. Schmid. Intensification and economic and ecological assessment of a biocatalytic oxyfunctionalization process. *Green Chem.*, 12:815–827, 2010.

[95] H. D. Conlon, J. Baqai, K. Baker, *et al.* Two-step immobilized enzyme conversion of cephalosporin C to 7-aminocephalosporanic acid. *Biotechnol. Bioeng.*, 46(6):510–513, 1995.

[96] L. Pollegioni, G. Molla, S. Sacchi, *et al.* Properties and applications of microbial D-amino acid oxidases: current state and perspectives. *Appl. Microbiol. Biotechnol.*, 78(1):1–16, 2008.

[97] C. M. McEwen, G. Sasaki, and W. R. Lenz. Human liver mitochondrial monoamine oxidase: I. kinetic studies of model interactions. *J. Biol. Chem.*, 243(20):5217–5225, 1968.

[98] T. Li, J. Liang, A. Ambrogelly, *et al.* Efficient, chemoenzymatic process for manufacture of the boceprevir bicyclic [3.1.0]proline intermediate based on amine oxidase-catalyzed desymmetrization. *J. Am. Chem. Soc.*, 134(14):6467–6472, 2012.

[99] R. McDermott, D. Tingley, J. Cowden, G. Frazzetto, and D. D. P. Johnson. Monoamine oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proc. Natl. Acad. Sci.*, 106(7):2118–2123, 2009.

[100] R. Carr, M. Alexeeva, M. J. Dawson, *et al.* Directed evolution of an amine oxidase for the preparative deracemisation of cyclic secondary amines. *ChemBioChem*, 6(4):637–639, 2005.

[101] C. J. Dunsmore, R. Carr, T. Fleming, and N. J. Turner. A chemo-enzymatic route to enantiomerically pure cyclic tertiary amines. *J. Am. Chem. Soc.*, 128(7):2224–2225, 2006.

[102] K. E. Atkin, R. Reiss, V. Koehler, *et al.* The structure of monoamine oxidase from *Aspergillus niger* provides a molecular context for improvements in activity obtained by directed evolution. *J. Mol. Biol.*, 384(5):1218–1231, 2008.

[103] B. Mijts, S. Muley, J. Liang, *et al.* Biocatalytic processes for the preparation of substantially stereomerically pure fused bicyclic proline compounds. *EP2307419.* 24.06.2008.

[104] G. de Gonzalo, W. J. H. van Berkel, and M. W. Fraaije. Baeyer–Villiger oxidation. In *Biocatalysis in Organic Synthesis 3*, page 187. Georg Thieme Verlag KG, New York, 3 2015.

[105] D. E. Torres Pazmiño, H. M. Dudek, and M. W. Fraaije. Baeyer–Villiger monooxygenases: recent advances and future challenges. *Curr. Opin. Chem. Biol.*, 14(2):138–144, 2010.

[106] G. de Gonzalo, M. D. Mihovilovic, and M. W. Fraaije. Recent developments in the application of Baeyer–Villiger monooxygenases as biocatalysts. *ChemBioChem*, 11(16):2208–2231, 2010.

[107] H. Leisch, K. Morley, and P. C. K. Lau. Baeyer-Villiger monooxygenases: more than just green chemistry. *Chem. Rev.*, 111(7):4165–4222, 2011.

[108] K. S. Mthethwa, K. Kassier, J. Engel, *et al.* Fungal BVMOs as alternatives to cyclohexanone monooxygenase. *Enzym. Microb. Technol.*, 106:11 – 17, 2017.

[109] M. W. Fraaije, J. Wu, D. P. H. M. Heuts, *et al.* Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. *Appl. Microbiol. Biotechnol.*, 66(4):393–400, 2005.

[110] A. Willetts, I. Joint, J. A. Gilbert, W. Trimble, and M. Mühling. Isolation and initial characterization of a novel type of Baeyer-Villiger monooxygenase activity from a marine microorganism. *Microb. Biotechnol.*, 5(4):549–559, 2012.

[111] A. Kirschner and U. T. Bornscheuer. Directed evolution of a Baeyer-Villiger monooxygenase to enhance enantioselectivity. *Appl. Microbiol. Biotechnol.*, 81(3):465–472, 2008.

[112] H. M. Dudek, D. E. Torres Pazmiño, C. Rodríguez, *et al.* Investigating the coenzyme specificity of phenylacetone monooxygenase from *Thermobifida fusca*. *Appl. Microbiol. Biotechnol.*, 88(5):1135–1143, 2010.

[113] D. TorresPazmiño, R. Snajdrova, B.-J. Baas, *et al.* Self-sufficient Baeyer–Villiger monooxygenases: effective coenzyme regeneration for biooxygenation by fusion engineering. *Angew. Chem. Int. Ed.*, 47(12):2275–2278, 2008.

[114] V. Alphand, G. Carrea, R. Wohlgemuth, R. Furstoss, and J. M. Woodley. Towards large-scale synthetic applications of Baeyer-Villiger monooxygenases. *Trends Biotechnol.*, 21(7):318–323, 2003.

[115] G. R. Eastham, D. W. Johnson, A. J. J. Straathof, M. W. Fraaije, and R. T. Winter. Process for the production of methyl methacrylate. *WO2013179005A1.* 28.05.2012.

[116] I. Codexis, Y. K. Bong, D. Clay, Michael, *et al.* Synthesis of prazole compounds. *WO2011071982A2.* 08.12.2009.

[117] D. R. Nelson. The cytochrome P450 homepage. *Hum Genomics*, 4(1):59–65, 2009.

[118] R. Bernhardt and V. B. Urlacher. Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl. Microbiol. Biotechnol.*, 98(14):6185–6203, 2014.

[119] H. Venkataraman, E. M. te Poele, K. Z. Rosłoniec, *et al.* Biosynthesis of a steroid metabolite by an engineered *Rhodococcus erythropolis* strain expressing a mutant cytochrome P450 BM3 enzyme. *Appl. Microbiol. Biotechnol.*, 99(11):4713–4721, 2015.

[120] V. B. Urlacher and M. Girhard. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol.*, 30(1):26–36, 2012.

[121] K. S. G and L. R. E. Erythromycins and process for their preparation. *US7807800B2*. 28.11.2003.

[122] K. Petzoldt, K. Annen, and H. Laurent. Process for the preparation of 11-beta-hydroxy steroids. *EP0042451A1*. 21.06.1980.

[123] H. Kaneda and T. Yamaguchi. Lactate sensor. *EP2573190A1*. 26.09.2012.

[124] G. Shumakovich, V. Kurova, I. Vasil'eva, *et al.* Laccase-mediated synthesis of conducting polyaniline. *J. Mol. Catal. B: Enzym.*, 77:105–110, 2012.

[125] C. Poulsen and P. B. Høstrup. Purification and characterization of a hexose oxidase with excellent strengthening effects in bread. *Cereal Chem. J.*, 75(1):51–57, 1998.

[126] J. B. Soe, C. H. Poulsen, and H. P. Bak. A method of improving the properties of a flour dough, a flour dough improving composition and improved food products. *EP0833563A1*. 07.06.1995.

[127] O. C. Hansen and P. Stougaard. Hexose oxidase from the red alga *Chondrus crispus*: purification, molecular cloning, and expression in *Pichia pastoris*. *J. Biol. Chem.*, 272(17):11581–11587, 1997.

[128] A. M. Wolff, O. C. Hansen, U. Poulsen, S. Madrid, and P. Stougaard. Optimization of the production of *Chondrus crispus* hexose oxidase in *Pichia pastoris*. *Protein Expr. Purif.*, 22(2):189–199, 2001.

[129] M. Cook and H. Thygesen. Safety evaluation of a hexose oxidase expressed in *Hansenula polymorpha*. *Food Chem. Toxicol.*, 41(4):523–529, 2003.

[130] C. H. Poulsen and K. M. Kragh. Anti-fouling composition. *EP1282669B1*. 04.06.1999.

[131] J. B. Kristensen, S. M. Olsen, B. S. Laursen, *et al.* Enzymatic generation of hydrogen peroxide shows promising antifouling effect. *Biofouling*, 26(2):141–153, 2009.

[132] J. B. Soe and L. W. E. Petersen. Process for prevention of maillard reaction in foodstuffs. *EP1341422B1*. 11.17.2000.

[133] M. Scheibner, B. Hülsdau, K. Zelena, *et al.* Novel peroxidases of *Marasmius scorodonius* degrade β-carotene. *Appl. Microbiol. Biotechnol.*, 77(6):1241–1250, 2008.

[134] K. Zelena, B. Hardebusch, B. Hülsdau, R. G. Berger, and H. Zorn. Generation of norisoprenoid flavors from carotenoids by fungal peroxidases. *J. Agric. Food Chem.*, 57(21):9951–9955, 2009.

[135] M. Pühse, R. T. Szweda, Y. Ma, *et al.* *Marasmius scorodonius* extracellular dimeric peroxidase - exploring its temperature and pressure stability. *Biochim. et Biophys. Acta (BBA) - Proteins Proteomics*, 1794(7):1091–1098, 2009.

[136] DSM. Peroxidase enzyme preparation from a genetically modified *Aspergillus niger*. *GRAS Notice 000402*, 2011.

[137] J. R. Cherry, M. H. Lamsa, P. Schneider, *et al.* Directed evolution of a fungal peroxidase. *Nat. Biotechnol.*, 17(4):379–384, 1999.

[138] B. E. Christensen, P. Schneider, and G. Pedersen. A microperoxidase preparation containing a hemopeptide as an active component. *EP0497794B1*. 10.13.1989.

[139] D. T, Kirk, and P. G. Dye transfer inhibition. *EP0495836B2*. 10.13.1989.

[140] M. Kjalke, M. B. Andersen, P. Schneider, *et al.* Comparison of structure and activities of peroxidases from *Coprinus cinereus*, *Coprinus macrorhizus* and *Arthromyces ramosus*. *Biochim. et Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol.*, 1120(3):248–256, 1992.

[141] L. Bunsgaard, H. Dalbøge, G. Houen, E. M. Rasmussen, and K. G. Welinder. Amino acid sequence of *Coprinus macrorhizus* peroxidase and cDNA sequence encoding *Coprinus cinereus* peroxidase. *Eur. J. Biochem.*, 213(1):605–611, 1993.

[142] J. F. Petersen, A. Kadziola, and S. Larsen. Three-dimensional structure of a recombinant peroxidase from *Coprinus cinereus* at 2.6 Å resolution. *FEBS Lett.*, 339(3):291–296, 1994.

[143] K. Ryu, J. P. McEldoon, and J. S. Dordick. Kinetic characterization of a fungal peroxidase from *Coprinus cinereus* in aqueous and organic media. *Biocatal. Biotransformation*, 13(1):53–63, 1995.

[144] J. W. Tams and K. G. Welinder. Unfolding and refolding of *Coprinus cinereus* peroxidase at high pH, in urea, and at high temperature. Effect of organic and ionic additives on these processes. *Biochemistry.*, 35(23):7573–7579, 1996.

[145] C. Kauffmann, B. Petersen, and M. J. Bjerrum. Enzymatic removal of phenols from aqueous solutions by *Coprinus cinereus* peroxidase and hydrogen peroxide. *J. Biotechnol.*, 73(1):71–74, 1999.

[146] N. Kunishima, K. Fukuyama, H. Matsubara, *et al.* Crystal structure of the fungal peroxidase from *Arthromyces ramosus* at 1.9 Å resolution: structural comparisons with the lignin and cytochrome *c* peroxidases. *J. Mol. Biol.*, 235(1):331–344, 1994.

[147] S. J. Kim, J. A. Lee, J. C. Joo, *et al.* The development of a thermostable CiP (*Coprinus cinereus* peroxidase) through *in silico* design. *Biotechnol. Prog.*, 26(4):1038–1046, 2010.

[148] Y. Zhou, C. Wang, L. Kalum, L. Ostergaard, and W. Huang. Color modification of textile. *WO2013040991A1*. 23.09.2011.

[149] T. Hansen and P. Nielsen. Process for producing paper or paperboard with increased strength from mechanical pulp. *EP0641403B1*. 18.05.1992.

[150] E. Strittmatter, C. Liers, R. Ullrich, *et al.* First crystal structure of a fungal high-redox potential dye-decolorizing peroxidase: substrate interaction sites and long-range electron transfer. *J Biol Chem*, 288(6):4095–4102, 2013.

[151] D. Linde, R. Pogni, M. Cañellas, *et al.* Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study. *Biochem. J*, 466(Pt 2):253–262, 2015.

[152] M. Pérez-Boada, F. J. Ruiz-Dueñas, R. Pogni, *et al.* Versatile peroxidase oxidation of high redox potential aromatic compounds: site-directed mutagenesis, spectroscopic and crystallographic investigation of three long-range electron transfer pathways. *J. Mol. Biol.*, 354(2):385–402, 2005.

[153] C. Bernini, R. Pogni, R. Basosi, and A. Sinicropi. The nature of tryptophan radicals involved in the long-range electron transfer of lignin peroxidase and lignin peroxidase-like systems: insights from quantum mechanical/molecular mechanics simulations. *Proteins: Struct. Funct. Bioinforma.*, 80(5):1476–1483, 2012.

[154] C. Bernini, E. Arezzini, R. Basosi, and A. Sinicropi. *In silico* spectroscopy of tryptophan and tyrosine radicals involved in the long-range electron transfer of cytochrome *c* peroxidase. *The J. Phys. Chem. B*, 118(32):9525–9537, 2014.

[155] J. Tamayo-Ramos, W. J. H. van Berkel, and L. H. de Graaff. Biocatalytic potential of laccase-like multicopper oxidases from *Aspergillus niger*. *Microb. Cell Factories*, 11(1):165, 2012.

[156] K. Nakamura, T. Kawabata, K. Yura, and N. Go. Novel types of two-domain multi-copper oxidases: possible missing links in the evolution. *FEBS Lett.*, 553(3):239–244, 2003.

[157] M. Lund and C. Felby. Process for treating pulp with laccase and a mediator to increase paper wet strength. *US6610172B1*. 06.05.1999.

[158] A. P. Virk, P. Sharma, and N. Capalash. Use of laccase in pulp and paper industry. *Biotechnol. Prog.*, 28(1):21–32, 2012.

[159] S. Riva. Laccases: blue enzymes for green chemistry. *Trends Biotechnol.*, 24(5):219–226, 2006.

[160] C. Diaz Blanco, A. Ortner, R. Dimitrov, *et al.* Building an antifouling zwitterionic coating on urinary catheters using an enzymatically triggered bottom-up approach. *ACS Appl. Mater. & Interfaces*, 6(14):11385–11393, 2014.

[161] G. Rocasalbas, A. Francesko, S. Touriño, *et al.* Laccase-assisted formation of bioactive chitosan/gelatin hydrogel stabilized with plant polyphenols. *Carbohydr. Polym.*, 92(2):989–996, 2013.

[162] T. Kudanga, G. S. Nyanhongo, G. M. Guebitz, and S. Burton. Potential applications of laccase-mediated coupling and grafting reactions: a review. *Enzym. Microb. Technol.*, 48(3):195–208, 2011.

[163] G. Freddi, A. Anghileri, S. Sampaio, *et al.* Tyrosinase-catalyzed modification of *Bombyx mori* silk fibroin: grafting of chitosan under heterogeneous reaction conditions. *J. Biotechnol.*, 125(2):281–294, 2006.

[164] D. Maté, E. García-Ruiz, S. Camarero, and M. Alcalde. Directed evolution of fungal laccases. *Curr Genomics*, 12(2):113–122, 2011.

[165] S. J. Horn, G. Vaaje-Kolstad, B. Westereng, and V. G. Eijsink. Novel enzymes for the degradation of cellulose. *Biotechnol. for Biofuels*, 5(45), 2012.

[166] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, and B. Henrissat. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. for Biofuels*, 6(41), 2013.

[167] B. Bissaro, Å. K. Røhr, G. Müller, *et al.* Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2. *Nat. Chem. Biol.*, 13:1123, 2017.

[168] L. Bertini, R. Breglia, M. Lambrughi, *et al.* Catalytic mechanism of fungal lytic polysaccharide monooxygenases investigated by first-principles calculations. *Inorg. Chem.*, 57(1):86–97, 2018.

[169] G. R. Hemsworth, B. Henrissat, G. J. Davies, and P. H. Walton. Discovery and characterization of a new family of lytic polysaccharide mono-oxygenases. *Nat Chem Biol*, 10(2):122–126, 2014.

[170] G. R. Hemsworth, G. J. Davies, and P. H. Walton. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr. Opin. Struct. Biol.*, 23(5):660–668, 2013.

[171] G. Vaaje-Kolstad, B. Westereng, S. J. Horn, *et al.* An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science.*, 330(6001):219–222, 2010.

[172] M. Gudmundsson, S. Kim, M. Wu, *et al.* Structural and electronic snapshots during the transition from a Cu(II) to Cu(I) metal center of a lytic polysaccharide monooxygenase by X-ray photoreduction. *J. Biol. Chem.*, 289(27):18782–18792, 2014.

[173] C. H. Kjaergaard, M. F. Qayyum, S. D. Wong, *et al.* Spectroscopic and computational insight into the activation of O_2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc. Natl. Acad. Sci.*, 111(24):8797–8802, 2014.

[174] A. J. Ragauskas, G. T. Beckham, M. J. Biddy, *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science.*, 344(6185), 2014.

[175] D. King. WEF white paper on the future of industrial biorefineries, 2010.

[176] B. Kaur and D. Chakraborty. Biotechnological and molecular approaches for vanillin production: a review. *Appl. Biochem. Biotechnol.*, 169(4):1353–1372, 2013.

[177] A. Muheim, B. Mueller, T. Muench, and M. Wetli. Process for the production of vanillin. *EP0885968B3*. 19.06.1997.

[178] R. H. H. van den Heuvel, M. W. Fraaije, C. Laane, and W. J. H. van Berkel. Enzymatic synthesis of vanillin. *J. Agric. Food Chem.*, 49(6):2954–2958, 2001.

[179] M. W. Fraaije, W. J. van Berkel, J. A. Benen, J. Visser, and A. Mattevi. A novel oxidoreductase family sharing a conserved FAD-binding domain. *Trends Biochem. Sci.*, 23(6):206–207, 1998.

[180] K. Piontek, A. T. Smith, and W. Blodig. Lignin peroxidase structure and function. *Biochem. Soc. Transactions*, 29(2):111–116, 2001.

[181] M. Hofrichter. Review: lignin conversion by manganese peroxidase (MnP). *Enzym. Microb. Technol.*, 30(4):454–466, 2002.

[182] U. Bornscheuer, K. Buchholz, and J. Seibel. Enzymatic degradation of (ligno)cellulose. *Angew. Chem. Int. Ed.*, 53(41):10876–10893, 2014.

[183] W. A. Duetz, J. B. van Beilen, and B. Witholt. Using proteins in their natural environment: potential and limitations of microbial whole-cell hydroxylations in applied biocatalysis. *Curr. Opin. Biotechnol.*, 12(4):419–425, 2001.

[184] J. Peters and B. Witholt. Solubilization of the overexpressed integral membrane protein alkane monooxygenase of the recombinant *Escherichia coli* W3110[pGEc47]. *Biochim. et Biophys. Acta (BBA) - Biomembr.*, 1196(2):145–153, 1994.

[185] J. Zhang, B. Witholt, and Z. Li. Coupling of permeabilized microorganisms for efficient enantioselective reduction of ketone with cofactor recycling. *Chem. Commun.*, 4:398–400, 2006.

[186] S. Panke, M. Held, M. G. Wubbolts, B. Witholt, and A. Schmid. Pilot-scale production of (*S*)-styrene oxide from styrene by recombinant *Escherichia coli* synthesizing styrene monooxygenase. *Biotechnol. Bioeng.*, 80(1):33–41, 2002.

[187] I. Hilker, R. Wohlgemuth, V. Alphand, and R. Furstoss. Microbial transformations 59: First kilogram scale asymmetric microbial Baeyer-Villiger oxidation with optimized productivity using a resin-based *in situ* SFPR strategy. *Biotechnol. Bioeng.*, 92(6):702–710, 2005.

[188] I. Hilker, M. C. Gutierrez, R. Furstoss, *et al.* Preparative scale Baeyer-Villiger biooxidation at high concentration using recombinant *Escherichia coli* and *in situ* substrate feeding and product removal process. *Nat. Protoc.*, 3(3):546–554, 2008.

[189] A. J. Straathof, S. Panke, and A. Schmid. The production of fine chemicals by biotransformations. *Curr. Opin. Biotechnol.*, 13(6):548–556, 2002.

[190] C. P. Kurtzman. Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol. & Biotechnol.*, 36(1435), 2009.

[191] A. T. Martínez, F. J. Ruiz-Dueñas, S. Camarero, *et al.* Oxidoreductases on their way to industrial biotransformations. *Biotechnol. Adv.*, 35(6):815 – 831, 2017.

[192] F. Lambert, J. Zucca, and J. Mane. Production of ferulic acid, coniferyl alcohol and/or natural vanillin, comprises bioconversion of eugenol by a bacterium belonging to genus *Streptomyces* comprising at least a nucleotide sequence. *FR2912758*. 21.02.2007.

[193] H. Gayet, D. Revelant, and M. Vibert. Method for the purification of natural vanillin. *US2017204039*. 24.01.2014.

[194] W. J. H. van Berkel, E. de Jong, and M. W. Fraaije. Enzymatic process for producing 4-hydroxy-cinnamyl alcohols. *US5721125*. 06.07.1993.

[195] F. P. Drijfhout, M. W. Fraaije, H. Jongejan, W. J. H. van Berkel, and M. C. R. Franssen. Enantioselective hydroxylation of 4-alkylphenols by vanillyl alcohol oxidase. *Biotechnol. Bioeng.*, 59(2):171–177, 1998.

[196] T. A. Ewing, M. W. Fraaije, A. Mattevi, and W. J. H. van Berkel. The VAO/PCMH flavoprotein family. *Arch. Biochem. Biophys.*, 632:104–117, 2017.

[197] C. D. Reeve, M. A. Carver, and D. J. Hopper. The purification and characterization of 4-ethylphenol methylenehydroxylase, a flavocytochrome from *Pseudomonas putida* JD1. *Biochem. J.*, 263(2):431–437, 1989.

[198] N. G. Leferink, D. P. Heuts, M. W. Fraaije, and W. J. van Berkel. The growing VAO flavoprotein family. *Arch. Biochem. Biophys.*, 474(2):292–301, 2008.

[199] G. Gygli, M. F. Lucas, V. Guallar, and W. J. H. van Berkel. The ins and outs of vanillyl alcohol oxidase: identification of ligand migration paths. *PLOS Comput. Biol.*, 13(10):1–27, 2017.

[200] I. Efimov, C. N. Cronin, and W. S. McIntire. Effects of noncovalent and covalent FAD binding on the redox and catalytic properties of *p*-cresol methylhydroxylase. *Biochemistry.*, 40(7):2155–2166, 2001.

[201] I. Efimov, C. N. Cronin, D. J. Bergmann, V. Kuusk, and W. S. McIntire. Insight into covalent flavinylation and catalysis from redox, spectral, and kinetic analyses of the R474K mutant of the flavoprotein subunit of *p*-cresol methylhydroxylase. *Biochemistry.*, 43(20):6138–6148, 2004.

[202] Q.-T. Nguyen, G. de Gonzalo, C. Binda, *et al.* Biocatalytic properties and structural analysis of eugenol oxidase from *Rhodococcus jostii* RHA1: a versatile oxidative biocatalyst. *ChemBioChem*, 17(14):1359–1366, 2016.

[203] O. Dym and D. Eisenberg. Sequence-structure analysis of FAD-containing proteins. *Protein Sci.*, 10(9):1712–1728, 2001.

[204] L. D. Garma, M. Medina, and A. H. Juffer. Structure-based classification of FAD binding sites: a comparative study of structural alignment tools. *Proteins: Struct. Funct. Bioinforma.*, 84(11):1728–1747, 2016.

[205] N. van Rooyen. Identification, cloning and heterologous expression of fungal vanillyl-alcohol oxidases. PhD thesis, University of the Free State, 2012.

[206] D. T. Jones, W. R. Taylor, and J. M. Thornton. The rapid generation of mutation data matrices from protein sequences. *Bioinforma.*, 8(3):275–282, 1992.

[207] K. Tamura, D. Peterson, N. Peterson, *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28(10):2731–2739, Oct 2011.

[208] J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, and O. Winther. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinforma.*, 33(21):3387–3395, 2017.

[209] M. Scherer, H. Wei, R. Liese, and R. Fischer. *Aspergillus nidulans* catalase-peroxidase gene (*cpeA*) is transcriptionally induced during sexual development through the transcription factor StuA. *Eukaryot. Cell*, 1(5):725–735, 2002.

[210] X. Carpena, W. Melik-Adamyan, P. C. Loewen, and I. Fita. Structure of the C-terminal domain of the catalase–peroxidase KatG from *Escherichia coli*. *Acta Crystallogr. Sect. D*, 60(10):1824–1832, Oct 2004.

[211] X. Zhao, H.-P. Hersleth, J. Zhu, K. K. Andersson, and R. S. Magliozzo. Access channel residues Ser315 and Asp137 in *Mycobacterium tuberculosis* catalase-peroxidase (KatG) control peroxidatic activation of the pro-drug isoniazid. *Chem. Commun.*, 49:11650–11652, 2013.

[212] X. Zhao, H. Yu, S. Yu, *et al.* Hydrogen peroxide-mediated isoniazid activation catalyzed by *Mycobacterium tuberculosis* catalaseperoxidase (KatG) and its S315T mutant. *Biochemistry.*, 45(13):4131–4140, 2006.

[213] M. W. Fraaije, R. H. van den Heuvel, A. Mattevi, and W. J. van Berkel. Covalent flavinylation enhances the oxidative power of vanillyl-alcohol oxidase. *J. Mol. Catal. B: Enzym.*, 21(1):43 – 46, 2003.

[214] D. P. H. M. Heuts, N. S. Scrutton, W. S. McIntire, and M. W. Fraaije. What's in a covalent bond? *FEBS J.*, 276(13):3405–3427, 2009.

[215] R. K. Wierenga, J. Drenth, G. E. Schulz, and R. Huber. Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. *J. Mol. Biol.*, 167(3):725 – 739, 1983.

[216] N. Ein-Gil, M. Ilan, S. Carmeli, *et al.* Presence of *Aspergillus sydowii*, a pathogen of gorgonian sea fans in the marine sponge *Spongia obscura*. *The Isme J.*, 3:752–755, 2009.

[217] P. Ferreira, A. Hernandez-Ortega, B. Herguedas, T. Martínez, and M. Medina. Aryl-alcohol oxidase involved in lignin degradation: a mechanistic study based on steady and pre-steady state kinetics and primary and solvent isotope effects with two alcohol substrates. *J. Biol. Chem.*, 284(37):24840–24847, 2009.

[218] M. W. Fraaije, H. P. Roubroeks, W. R. Hagen, and W. J. H. van Berkel. Purification and characterization of an intracellular catalase-peroxidase from *Penicillium simplicissimum*. *Eur. J. Biochem.*, 235(1-2):192–198, 1996.

[219] C.-H. Huang, W.-L. Lai, M.-H. Lee, *et al.* Crystal structure of glucooligosaccharide oxidase from *Acremonium strictum*: a novel flavinylation of 6-S-cysteinyl, 8α-N1-histidyl FAD. *J. Biol. Chem.*, 280(46):38831–38838, 2005.

[220] H. Otani, Y.-E. Lee, I. Casabon, and L. D. Eltis. Characterization of *p*-hydroxycinnamate catabolism in a soil actinobacterium. *J. Bacteriol.*, 196(24):4293–4303, 2014.

[221] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner. WebLogo: a sequence logo generator. *Genome Res.*, 14(6):1188–1190, 2004.

[222] B. Liu, H. Liu, D. Zhong, and C. Lin. Searching for a photocycle of the cryptochrome photoreceptors. *Curr. Opin. Plant Biol.*, 13(5):578–586, 2010.

[223] M. W. Fraaije and W. J. H. van Berkel. Catalytic mechanism of the oxidative demethylation of 4-(methoxymethyl)phenol by vanillyl-alcohol oxidase: evidence for formation of a *p*-quinone methide intermediate. *J. Biol. Chem.*, 272(29):18111–18116, 1997.

[224] J. A. E. Benen, P. Sánchez-Torres, M. J. M. Wagemaker, *et al.* Molecular cloning, sequencing, and heterologous expression of the *vao*A gene from *Penicillium simplicissimum* CBS 170.90 encoding vanillyl-alcohol oxidase. *J. Biol. Chem.*, 273(14):7865–7872, 1998.

[225] B. Daniel, T. Pavkov-Keller, B. Steiner, *et al.* Oxidation of monolignols by members of the berberine bridge enzyme family suggests a role in plant cell wall metabolism. *J. Biol. Chem.*, 290(30):18770–18781, 2015.

[226] R Development Core Team and R Core Team. R: a language and environment for statistical computing, 2008.

[227] M. W. Fraaije, A. Mattevi, and W. J. van Berkel. Mercuration of vanillyl-alcohol oxidase from *Penicillium simplicissimum* generates inactive dimers. *FEBS Lett.*, 402(1):33–35, 1997.

[228] K. Kvalnes-Krick and M. Schuman Jorns. Role of the covalent and noncovalent flavins in sarcosine oxidase. In F. Mueller, editor, *Chemistry and Biochemistry of Flavoenzymes, vol. II.*, pages 425–435. CRC Press Inc, 1991.

[229] S. van den Berghe-Snorek and M. T. Stankovich. Thermodynamic control of D-amino acid oxidase by benzoate binding. *J. Biol. Chem.*, 260(6):3373–3379, 1985.

[230] V. Job, G. L. Marcone, M. S. Pilone, and L. Pollegioni. Glycine oxidase from *Bacillus subtilis* : characterization of a new flavoprotein. *J. Biol. Chem.*, 277(9):6985–6993, 2002.

[231] Y. S. Choong and V. Massey. Stabilization of lactate oxidase flavin anion radical by complex formation. *J. Biol. Chem.*, 255(18):8672–8677, 1980.

[232] J. Geissler, S. Ghisla, and P. M. H. Kroneck. Flavin-dependent alcohol oxidase from yeast. *Eur. J. Biochem.*, 160(1):93–100, 1986.

[233] M. Ghanem and G. Gadda. Effects of reversing the protein positive charge in the proximity of the flavin N(1) locus of choline oxidase. *Biochemistry.*, 45(10):3437–3447, 2006.

[234] M. Ghanem, Fan, K. Francis, and G. Gadda. Spectroscopic and kinetic properties of recombinant choline oxidase from *Arthrobacter globiformis. Biochemistry.*, 42(51):15179–15188, 2003.

[235] M. Mewies, W. S. McIntire, and N. S. Scrutton. Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: the current state of affairs. *Protein Sci.*, 7(1):7–20, 1998.

[236] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinforma.*, 25(9):1189–1191, 2009.

[237] D. H. Huson and C. Scornavacca. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.*, 61(6):1061–1067, 2012.

[238] W. McIntire, D. J. Hopper, and T. P. Singer. *p*-Cresol methylhydroxylase. Assay and general properties. *Biochem. J*, 228(2):325–335, 1985.

[239] W. Mclntire, D. E. Edmondson, D. J. Hopper, and S. T.P. 8α-(*O*-Tyrosyl)flavin adenine dinucleotide: the prosthetic group of bacterial *p*-cresol methylhydroxylase. *Biochem. Soc Trans*, 20:3068–3075, 1981.

[240] R. Baron, C. Riley, P. Chenprakhon, *et al.* Multiple pathways guide oxygen diffusion into flavoenzyme active sites. *Proc. Natl. Acad. Sci.*, 106(26):10603–10608, 2009.

[241] N. G. H. Leferink, S. V. Antonyuk, J. A. Houwman, *et al.* Impact of residues remote from the catalytic centre on enzyme catalysis of copper nitrite reductase. *Nat. Commun.*, 5, 2014.

[242] J. Saam, E. Rosini, G. Molla, *et al.* O_2 reactivity of flavoproteins: dynamic access of dioxygen to the active site and role of a H^+ relay system in D-amino acid oxidase. *J. Biol. Chem.*, 285(32):24439–24446, 2010.

[243] H. R. Brodkin, W. R. P. Novak, A. C. Milne, *et al.* Evidence of the participation of remote residues in the catalytic activity of Co-type nitrile hydratase from *Pseudomonas putida*. *Biochemistry.*, 50(22):4923–4935, 2011.

[244] A. Hosseini, M. Brouk, M. F. Lucas, *et al.* Atomic picture of ligand migration in toluene 4-monooxygenase. *The J. Phys. Chem. B*, 119(3):671–678, 2015.

[245] K. Gunasekaran, B. Ma, and R. Nussinov. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct. Funct. Bioinforma.*, 57(3):433–443, 2004.

[246] N. M. Goodey and S. J. Benkovic. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.*, 4:474–482, 2008.

[247] P.-R. Kommoju, Z.-W. Chen, R. C. Bruckner, F. S. Mathews, and M. S. Jorns. Probing oxygen activation sites in two flavoprotein oxidases using chloride as an oxygen surrogate. *Biochemistry.*, 50(24):5521–5534, 2011.

[248] N. G. H. Leferink, M. W. Fraaije, H.-J. Joosten, *et al.* Identification of a gatekeeper residue that prevents dehydrogenases from acting as oxidases. *J. Biol. Chem.*, 284(7):4392–4397, 2009.

[249] D. Zafred, B. Steiner, A. R. Teufelberger, *et al.* Rationally engineered flavin-dependent oxidase reveals steric control of dioxygen reduction. *FEBS J.*, 282(16):3060–3074, 2015.

[250] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Mol. Des.*, 27(3):221–234, 2013.

[251] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.*, 7(2):525–537, 2011.

[252] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. *H*++ 3.0: automating p*K* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, 40(W1):537–541, 2012.

[253] M. P. Jacobson, D. L. Pincus, C. S. Rapp, *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct. Funct. Bioinforma.*, 55(2):351–367, 2004.

[254] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The J. Phys. Chem. B*, 105(28):6474–6487, 2001.

[255] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Accounts*, 120(1):215–241, 2008.

[256] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *The J. Chem. Phys.*, 56(5):2257–2261, 1972.

[257] B. Marten, K. Kim, C. Cortis, *et al.* New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *The J. Phys. Chem.*, 100(28):11775–11788, 1996.

[258] B. P. Cossins, A. Hosseini, and V. Guallar. Exploration of protein conformational change with PELE and meta-dynamics. *J. Chem. Theory Comput.*, 8(3):959–965, 2012.

[259] A. Hernández-Ortega, K. Borrelli, P. Ferreira, *et al.* Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J.*, 436(2):341–350, 2011.

[260] M. Lucas and V. Guallar. An atomistic view on human hemoglobin carbon monoxide migration processes. *Biophys. J.*, 102(4):887–896, 2012.

[261] F. Lucas, E. D. Babot, M. Canellas, *et al.* Molecular determinants for selective C25-hydroxylation of vitamins D2 and D3 by fungal peroxygenases. *Catal. Sci. Technol.*, 6:288–295, 2016.

[262] A. Onufriev, D. Bashford, and D. A. Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Bioinforma.*, 55(2):383–394, 2004.

[263] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2(3):173–181, 1997.

[264] D. Bashford and D. A. Case. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, 51(1):129–152, 2000.

[265] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38, 1996.

[266] T. Williams and K. C. Gnuplot 4.4: an interactive plotting program., 2013.

[267] J. Pei, B.-H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, 36(7):2295–2300, 2008.

[268] H. Ashkenazy, S. Abadi, E. Martz, *et al.* Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, 44(W1):344–350, 2016.

[269] N. Guex and M. C. Peitsch. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophor.*, 18(15):2714–2723, 1997.

[270] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.*, 89(22):10915–10919, 1992.

[271] M. W. Fraaije, R. H. H. van den Heuvel, J. C. A. A. Roelofs, and W. J. H. van Berkel. Kinetic mechanism of vanillyl-alcohol oxidase with short-chain 4-alkylphenols. *Eur. J. Biochem.*, 253(3):712–719, 1998.

[272] R. H. van den Heuvel, M. W. Fraaije, A. Mattevi, C. Laane, and W. J. H. van Berkel. Vanillyl-alcohol oxidase, a tasteful biocatalyst. *J. Mol. Catal. B: Enzym.*, 11(4):185 – 188, 2001. Proceedings of the 4th International Symposium on Biocatalysis.

[273] N. Ogata and T. Baba. Analysis of beechwood creosote by gas chromatography-mass spectrometry and high-performance liquid chromatography. *Res Commun Chem Pathol Pharmacol*, 66(3):411–23, 1989.

[274] J. Zhao, W. Xiuwen, J. Hu, *et al.* Thermal degradation of softwood lignin and hardwood lignin by TG-FTIR and Py-GC/MS. *Polym. Degrad. Stab.*, 108:133 – 138, 2014.

[275] M. Carrier, M. Windt, B. Ziegler, *et al.* Quantitative insights into the fast pyrolysis of extracted cellulose, hemicelluloses, and lignin. *ChemSusChem*, 10(16):3212–3224, 2017.

[276] M. Fache, B. Boutevin, and S. Caillol. Vanillin production from lignin and its use as a renewable chemical. *ACS Sustain. Chem. & Eng.*, 4(1):35–46, 2016.

[277] C. Isola, H. L. Sieverding, A. M. Numan-Al-Mobin, *et al.* Vanillin derived from lignin liquefaction: a sustainability evaluation. *The Int. J. Life Cycle Assess.*, 2017.

[278] X. Sheng, M. E. Lind, and F. Himo. Theoretical study of the reaction mechanism of phenolic acid decarboxylase. *FEBS J.*, 282(24):4703–4713, 2015.

[279] X. Sheng and F. Himo. Theoretical study of enzyme promiscuity: mechanisms of hydration and carboxylation activities of phenolic acid decarboxylase. *ACS Catal.*, 7(3):1733–1741, 2017.

[280] J. P. Bennett, L. Bertin, B. Moulton, *et al.* A ternary complex of hydroxycinnamoyl-CoA hydratase–lyase (HCHL) with acetyl-CoA and vanillin gives insights into substrate specificity and mechanism. *Biochem. J.*, 414(2):281–289, 2008.

[281] M. R. A. Blomberg, T. Borowski, F. Himo, R.-Z. Liao, and P. E. M. Siegbahn. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.*, 114(7):3601–3658, 2014.

[282] P. E. Siegbahn and F. Himo. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(3):323–336, 2011.

[283] P. E. M. Siegbahn and M. R. A. Blomberg. Quantum chemical studies of proton-coupled electron transfer in metalloenzymes. *Chem. Rev.*, 110(12):7040–7061, 2010.

[284] P. E. M. Siegbahn and F. Himo. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *JBIC J. Biol. Inorg. Chem.*, 14(5):643–651, 2009.

[285] W. McIntire, D. J. Hopper, J. C. Craig, *et al.* Stereochemistry of 1-(4'-hydroxyphenyl)ethanol produced by hydroxylation of 4-ethylphenol by *p*-cresol methylhydroxylase. *Biochem. J.*, 224(2):617–621, 1984.

[286] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.* Gaussian 09, revision D.01, 2016.

[287] A. D. Becke. Density functional thermochemistry. III. The role of exact exchange. *The J. Chem. Phys.*, 98(7):5648–5652, 1993.

[288] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988.

[289] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The J. Phys. Chem. B*, 113(18):6378–6396, 2009.

[290] R. Sevastik and F. Himo. Quantum chemical modeling of enzymatic reactions: the case of 4-oxalocrotonate tautomerase. *Bioorganic Chem.*, 35(6):444 – 457, 2007.

[291] K. H. Hopmann and F. Himo. Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase. *J. Chem. Theory Comput.*, 4(7):1129–1137, 2008.

[292] P. Georgieva and F. Himo. Quantum chemical modeling of enzymatic reactions: the case of histone lysine methyltransferase. *J. Comput. Chem.*, 31(8):1707–1714, 2010.

[293] R.-Z. Liao, J.-G. Yu, and F. Himo. Quantum chemical modeling of enzymatic reactions: the case of decarboxylation. *J. Chem. Theory Comput.*, 7(5):1494–1501, 2011.

[294] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The J. Chem. Phys.*, 132(15):154104, 2010.

[295] S. Grimme, S. Ehrlich, and L. Goerigk. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011.

[296] R. Buller, K. Hecht, M. A. Mirata, and H.-P. Meyer. An appreciation of biocatalysis in the Swiss manufacturing environment. In *Biocatalysis: An Industrial Perspective*, pages 1–43. The Royal Society of Chemistry, 2018.

[297] T. A. Ewing, J. Wetzig, S. Segarra, *et al. Manuscript in preparation.*

[298] S. Bory, M. Grisoni, M.-F. Duval, and P. Besse. Biodiversity and preservation of vanilla: present state of knowledge. *Genet. Resour. Crop. Evol.*, 55(4):551–571, 2008.

[299] A. K. Sinha, U. K. Sharma, and N. Sharma. A comprehensive review on vanilla flavor: extraction, isolation and quantification of vanillin and others constituents. *Int. J. Food Sci. Nutr.*, 59(4):299–326, 2008.

[300] N. J. Gallage, E. H. Hansen, R. Kannangara, *et al.* Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nat. Commun.*, 5:4037, 2014.

[301] S. Ramachandra Rao and G. Ravishankar. Vanilla flavour: production by conventional and biotechnological routes. *J. Sci. Food Agric.*, 80(3):289–304, 2000.

[302] N. Gallage and B. Møller. Vanillin–bioconversion and bioengineering of the most popular plant flavor and its *de novo* biosynthesis in the vanilla orchid. *Mol. Plant*, 8(1):40 – 57, 2015.

[303] M. Sabisch and D. Smith. The regulatory landscape of natural flavors. *Perfumer & Flavorist*, 37:38–41, 2012.

[304] Nestlé USA commits to removing artificial flavors and FDA-certified colors from all nestlé chocolate candy by the end of 2015. `https://www.nestleusa.com/media/pressreleases/nestl%C3%A9-usa-commits-to-removing-artificial-flavors-and-fda-certified-colors-from-all-nestl%C3%A9-chocolate-candy-by-the-end-of-20`. 17.02.2015.

[305] M. Sheley. *Frankenstein.* Barnes & Noble, 2015.

[306] UniProtKB/Swiss-Prot UniProt release 2017_12. `http://www.uniprot.org/statistics/Swiss-Prot`. 25.12.2017.

[307] RefSeq Growth Statistics. `https://www.ncbi.nlm.nih.gov/refseq/statistics/`. 25.12.2017.

[308] A. M. Schnoes, D. C. Ream, A. W. Thorman, P. C. Babbitt, and I. Friedberg. Biases in the experimental annotations of protein function and their effect on our understanding of protein function space. *PLOS Comput. Biol.*, 9(5):1–10, 05 2013.

[309] O. Y. Abdelaziz, D. P. Brink, J. Prothmann, *et al.* Biological valorization of low molecular weight lignin. *Biotechnol. Adv.*, 34(8):1318 – 1346, 2016.

[310] H. Möhler, M. Brühmüller, and K. Decker. Covalently bound flavin in D-6-hydroxynicotine oxidase from *Arthrobacter oxidans*. *Eur. J. Biochem.*, 29(1):152–155, 1972.

[311] M. Brühmüller, H. Möhler, and K. Decker. Covalently bound flavin in D-6-hydroxynicotine oxidase from *Arthrobacter oxidans*. *Eur. J. Biochem.*, 29(1):143–151, 1972.

[312] S. C. L. Kamerlin and A. Warshel. At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 78(6):1339–1375, 2010.

[313] S. M. Suan. Creation, purification and characterization of vanillyl-alcohol oxidase variants. Master's thesis, Wageningen University & Research, 2016.

[314] A. J. P. Gutierrez. Purification and characterization of a vanillyl-alcohol oxidase mutant. Master's thesis, Wageningen University & Research, 2015.

[315] A. Pavelka, E. Sebestova, B. Kozlikova, *et al.* Caver: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Transactions on Comput. Biol. Bioinforma.*, 13(3):505–517, May 2016.

[316] M. J. Abraham, T. Murtola, R. Schulz, *et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[317] A. Ismail, V. Leroux, M. Smadja, *et al.* Coenzyme Q biosynthesis: evidence for a substrate access channel in the FAD-dependent monooxygenase Coq6. *PLOS Comput. Biol.*, 12(1):1–27, 2016.

[318] B. R. Brooks, C. L. Brooks, A. D. Mackerell, *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, 30(10):1545–1614, 2009.

[319] T. Kulschewski and J. Pleiss. A molecular dynamics study of liquid aliphatic alcohols: simulation of density and self-diffusion coefficient using a modified OPLS force field. *Mol. Simul.*, 39(9):754–767, 2013.

[320] S. P. Benson and J. Pleiss. Incomplete mixing versus clathrate-like structures: a molecular view on hydrophobicity in methanol-water mixtures. *J. Mol. Model.*, 19(8):3427–3436, 2013.

[321] T. Kulschewski and J. Pleiss. Binding of solvent molecules to a protein surface in binary mixtures follows a competitive Langmuir model. *Langmuir*, 32(35):8960–8968, 2016.

[322] S. P. Benson and J. Pleiss. Solvent Flux Method (SFM): a case study of water access to *Candida antarctica* Lipase B. *J. Chem. Theory Comput.*, 10(11):5206–5214, 2014.

[323] C. C. Gruber and J. Pleiss. Lipase B from *Candida antarctica* binds to hydrophobic substrate-water interfaces via hydrophobic anchors surrounding the active site entrance. *J. Mol. Catal. B: Enzym.*, 84:48–54, 2012.

[324] S. P. Benson and J. Pleiss. Self-assembly nanostructures of triglyceride-water interfaces determine functional conformations of *Candida antarctica* Lipase B. *Langmuir*, 33(12):3151–3159, 2017.

[325] A. Taudt, A. Arnold, and J. Pleiss. Simulation of protein association: kinetic pathways towards crystal contacts. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 91(3):1–10, 2015.

[326] J. Pleiss. Thermodynamic activity-based interpretation of enzyme kinetics. *Trends Biotechnol.*, 35(5):379–382, 2017.

[327] T. Kulschewski, F. Sasso, F. Secundo, M. Lotti, and J. Pleiss. Molecular mechanism of deactivation of *C. antarctica* lipase B by methanol. *J. Biotechnol.*, 168(4):462–469, 2013.

[328] F. Sasso, T. Kulschewski, F. Secundo, M. Lotti, and J. Pleiss. The effect of thermodynamic properties of solvent mixtures explains the difference between methanol and ethanol in *C. antarctica* lipase B catalyzed alcoholysis. *J. Biotechnol.*, 214:1–8, 2015.

[329] A. Seifert, S. Vomund, K. Grohmann, *et al.* Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *ChemBioChem*, 10(5):853–861, 2009.

[330] Y. Hirose, K. Kariya, I. Sasaki, *et al.* Drastic solvent effect on lipase-catalyzed enantioselective hydrolysis of prochiral 1,4-dihydropyridines. *Tetrahedron Lett.*, 33(47):7157 – 7160, 1992.

[331] M. C. Ebert and J. N. Pelletier. Computational tools for enzyme improvement: why everyone can and should use them. *Curr. Opin. Chem. Biol.*, 37:89–96, 2017.

[332] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3:160018, 2016.

[333] V. Lapatas, M. Stefanidakis, R. C. Jimenez, A. Via, and M. V. Schneider. Data integration in biological research: an overview. *J. Biol. Res.*, 22(1):9, 2015.

[334] I. Yanai and E. Chmielnicki. Computational biologists: moving to the driver's seat. *Genome Biol.*, 18(1):223, 2017.

About the author

Gudrun Gygli was born on the 13th of March 1989 in Basel, Switzerland. Despite rejecting a chemistry set at the tender age of six, she later pursued a scientific career and began studying Biology at the University of Basel in 2008. After two years of Bachelor studies, she chose to specialise in Biotechnology through studies at the Trinational School for Biotechnology in Strasbourg. During her studies, she also worked on small research projects in the groups of Prof. Dr. Uwe Bornscheuer at the University of Greifswald and Dr. Annick Dejaegere at the IGBMC in Strasbourg. Her Master's thesis, entitled "Characterisation of two β-xylosidases from *Myceliophthora thermophila* and *Fungus X* and mutant design to reduce xylose inhibition" was executed at DuPont Industrial Biosciences (formerly Dyadic Nederland B.V.) in Wageningen.

In 2013, she obtained her Master's degree in Bioinformatics and Structural Biology and her Engineering Diploma in Biotechnology from the University of Strasbourg and the Trinational School for Biotechnology in Strasbourg. She briefly worked at DuPont Industrial Biosciences to finalise research done during her Master's thesis. In 2014 she began working as a PhD candidate in the Enzymes@Work group of Prof. Dr. Willem van Berkel in the Laboratory of Biochemistry at Wageningen University & Research. During her PhD, she collaborated with the groups of Prof. Dr. Victor Guallar at the Barcelona Supercomputing Center and Prof. Dr. Fahmi Himo at Stockholm University. The results of her research since 2014 are presented in this thesis.

Gudrun was recently awarded an Early Postdoc.mobility fellowship by the Swiss National Science Foundation and will continue her career as a postdoctoral researcher in the group of Apl. Prof. Dr. Jürgen Pleiss at the Institute of Biochemistry and Technical Biochemistry of the University of Stuttgart.

Full publications list

- Philippe Wolff, Ismail Amal, Vincent Oliéric, Olivier Chaloin, **Gudrun Gygli**, Eric Ennifar, Bernard Lorber, Gilles Guichard, Jérôme Wagner, Annick Dejaegere, and Dominique Y. Burnouf. Differential modes of peptide binding onto replicative sliding clamps from various bacterial origins. *Journal of Medicinal Chemistry*, 2014.
- **Gudrun Gygli** and Willem J. H. van Berkel. Oxizymes for Biotechnology. *Current Biotechnology*, 2015.
- Tom A. Ewing, **Gudrun Gygli** and Willem J. H. van Berkel. A single loop is essential for the octamerization of vanillyl alcohol oxidase. *FEBS Journal*, 2016.
- **Gudrun Gygli** and Willem J. H. van Berkel. Vanillyl alcohol oxidases produced in *Komagataella phaffii* contain a highly stable non-covalently bound anionic FAD semiquinone. *Biocatalysis*, 2017.
- **Gudrun Gygli**, Maria F. Lucas, Victor Guallar, and Willem J. H. van Berkel. The ins and outs of vanillyl alcohol oxidase: Identification of ligand migration paths. *PLOS computational biology*, 2017.
- Tom A. Ewing, Quoc-Thai Nguyen, Robert C. Allan, **Gudrun Gygli**, Elvira Romero, Claudia Binda, Marco W. Fraaije, Andrea Mattevi, and Willem J. H. van Berkel. Two tyrosine residues, Tyr-108 and Tyr-503, are responsible for the deprotonation of phenolic substrates in vanillyl alcohol oxidase. *Journal of Biological Chemistry*, 2017.
- **Gudrun Gygli**, Ronald P. de Vries, and Willem J. H. van Berkel. On the origin of vanillyl alcohol oxidases. 2018. *Manuscript under review.*
- **Gudrun Gygli**, Xiang Sheng, Fahmi Himo, and Willem J. H. van Berkel. Quantum chemical modelling of vanillyl alcohol oxidase. 2018. *Manuscript in preparation.*

Overview of completed training activities

Discipline specific activities

Enzyme mechanisms by biological systems	Manchester (UK)	2014
7th European meeting on OxiZymes **	Vienna (AT)	2014
18th International symposium on flavins and flavoproteins **	Cha-Am (TH)	2014
CHAINS - chemistry as innovating science **	Veldhoven (NL)	2014
NVB werkgroependag	Ede (NL)	2015
Biotrans **	Vienna (AT)	2015
8th European meeting on OxiZymes **	Wageningen (NL)	2016
5th International conference on novel enzymes **	Groningen (NL)	2016
Guest researcher at the BSC	Barcelona (ES)	2014 & 2016
19th International symposium on flavins and flavoproteins *	Groningen (NL)	2017
CHAINS - chemistry as innovating science *	Veldhoven (NL)	2017
MolSim **	Amsterdam (NL)	2017
Guest researcher at the University of Stockholm	Stockholm (SE)	2017
INDOX project meetings *	Europe	2014-2017

General courses

VLAG PhD week	Baarlo (NL)	2014
Scientific publishing	Wageningen (NL)	2014
Teaching and supervising students	Wageningen (NL)	2014
Communication with the media and the general public	Wageningen (NL)	2015
Brain Training	Wageningen (NL)	2016
Reviewing a scientific paper	Wageningen (NL)	2017
Writing Grant proposals	Wageningen (NL)	2017
Career assessment	Wageningen (NL)	2017

Optional activities

Biochemistry weekly group meetings and colloquia	Wageningen (NL)	2014-2017
Preparing PhD research proposal	Wageningen (NL)	2014
Member of the VLAG PhD council	Wageningen (NL)	2015-2017

* oral presentation, ** poster presentation

The research described in this thesis was performed at the
Laboratory of Biochemistry, Wageningen University & Research
and was financially supported through the INDOX project
(FP7-KBBE-2013-7-613549) of the European Union.

Financial support from the
Laboratory of Biochemistry, Wageningen University & Research,
for printing this thesis is gratefully acknowledged.

Cover design by Gudrun Gygli
Layout by Gudrun Gygli
Printed by GVO drukkers & vormgevers, Ede, the Netherlands

Gudrun Gygli, April 2018