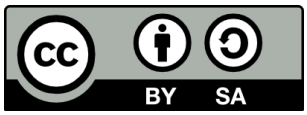
Universitat de Girona

ELUCIDATING THE ROLE OF CONFORMATIONAL DYNAMICS OF ASPERGILLUS NIGER MONOAMINE OXIDASE TOWARDS ENZYMATIC CHIRAL AMINES SYNTHESIS

Christian Curado Carballada



Universitat de Girona

DOCTORAL THESIS

ELUCIDATING THE ROLE OF CONFORMATIONAL DYNAMICS OF *ASPERGILLUS NIGER* MONOAMINE OXIDASE TOWARDS ENZYMATIC CHIRAL AMINES SYNTHESIS

Christian Curado Carballada

2022

DOCTORAL PROGRAMME IN CHEMISTRY

Universitat de Girona

DOCTORAL THESIS

ELUCIDATING THE ROLE OF CONFORMATIONAL DYNAMICS OF *ASPERGILLUS NIGER* MONOAMINE OXIDASE TOWARDS ENZYMATIC CHIRAL AMINES SYNTHESIS

Christian Curado Carballada

2022

DOCTORAL PROGRAMME IN CHEMISTRY

Supervised by:

Prof. Sílvia Osuna Oliveras

Dr. Ferran Feixas Gerones

Tutor:

Prof. Marcel Swart

Presented to obtain the degree of PhD at the University of Girona

List of Publications

Chapter IV and V of this thesis are respectively based on the following publications:

1. C. Curado-Carballada, F. Feixas, J. Iglesias-Fernández, S. Osuna, "Hidden Conformations in Aspergillus niger Monoamine Oxidase are Key for Catalytic Efficiency", Angew. Chem. Int. Ed. 2019, 58, 3097.

2. C. Curado-Carballada, F. Feixas, S. Osuna, "Molecular Dynamics Simulations on Aspergillus niger Monoamine Oxidase: Conformational Dynamics and Inter-monomer Communication Essential for Its Efficient Catalysis", Adv. Synth. Catal. 2019, 361, 2718

Other publications not included in this thesis:

3. A. Schneider**, C. Curado**, T. Lystbaek, S. Osuna, B. Hauer, "Tailoring the squalene-hopene cyclase for stereoconvergent and efficient cationic cyclization cascades", ChemRxiv, 2022, DOI: 10.26434/chemrxiv-2022-xwcqj. **These authors contributed equally to this work

List of Abbreviations

Abbreviation	Description
Adk	Adenylate kinase enzyme
AMBA	alpha-methyl benzylamine
AMBER	Assisted model building with energy refinement
aMD	accelerated Molecular Dynamics
CV	Collective Variable
DE	Directed Evolution
FAD	Flavinadenine Dinucleotide
FEL	Free Energy Landscape
GAFF	Generalized Amber Force Field
HEX	Hexylamine
ISM	Iterative-saturation mutagenesis
LJ	Lennard-Jones
MAO	Monoamine Oxidase
MD	Molecular Dynamics
MM	Molecular Mechanics
MSM	Markov State Models
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PTE	Phosphotriesterase enzyme
QM	Quantum Mechanics
RA	Retro-Aldolase
RMSD	Root Mean Square Deviation
SSM	Site-saturation mutagenesis
tICA	time-lagged independent component analysis
TS	Transition State
U	Potential Energy
WT	Wild-type
α	alpha
β	beta
Å	Angstrom

Universitat de Girona

Prof. Dr. Sílvia Osuna Oliveras and Dr. Ferran Feixas Gerones, of University of Girona,
WE DECLARE:

That the thesis entitled “Elucidating the role of conformational dynamics of aspergillus niger monoamine oxidase towards enzymatic chiral amines synthesis”, presented to obtain a doctoral degree, has been carried and completed by Mr. Christian Curado Carballada under our supervision and that meets the requirements to opt for an International Doctorate.

For all intents and purposes, we hereby sign this document.

Signature

Prof. Dr. Sílvia Osuna Oliveras

Dr. Ferran Feixas Gerones

Girona, December 15th, 2022

Acknowledgements

The studies included in this thesis were carried out thanks to the financial support of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001). I would also like to thank Universitat de Girona (UdG) and the institute of computational Chemistry and Catalysis, for the facilities provided, and the great environment.

I would like start giving a great acknowledgement to **Sílvia** and **Ferran**. Not only for discovering me the amazing world of enzymes, but also for the great support, motivation, and great insights that you gave me along these amazing 5 years of PhD. I would also like to thank all the CompBioLab members, those who are still here: **Cristina, Miquel, Guillem, Carla, Jordi, Hiva, Janet, Rosa, Fred, Ram**, **and Marc** and also to those who already moved on in their academic or industry lifes: **Eila**, **Javi, Lorenzo, Adrià, Miguel Ángel, and Sergi**. It was great to share with you this stage of my life, it was both personally and scientifically stimulating to spend these years with you.

From my personal environment I would like to thank my Father for the support he gave me along my whole academic life. All my uncles and cousins, who celebrated each small success. Herein, I would like to include all my friends, starting with **Sheila** who was the one who suffered the most my last months of PhD. All my friends from Valencia as well as those from Santa Coloma (**Natalia, Oscar, Jony, Tamara, Nestor y Esther**) for the great support they gave me along this path. Of course, here I have to add my colleagues from the degree of chemistry, **Adrià, Cristina, Albert, and Lujan**.

Finally, I could finish without thanking to everyone who actively contributed to this thesis, directly or indirectly. THANK YOU ALL!! ☺

Table of Contents

List of figures

List of tables

Thesis Summary

Enzymes have evolved through years until becoming great catalysts that present high selectivity, specificity, and activity. Initially, enzymes were understood as static (bio)molecular structures capable of accelerating chemical reactions by many orders of magnitude. Notwithstanding, after decades of research, they could be identified as dynamic entities, able to adopt different thermally accessible conformations key for their catalytic function. The advantages that enzymes present when it comes to increase the reaction rate of (bio)chemical reactions can be used for producing industrially relevant compounds. In this regard, enzymes present a limitation, in most cases they are not active enough towards non-natural reactions or substrates relevant for industrial purposes. Such limitation can be overcome by means of enzyme engineering, which consists of introducing mutations in the enzyme sequence either using computational or experimental protocols. Experimentally, one of the most used techniques is Directed Evolution (DE), that performs random mutations along the enzyme sequence, producing variants with several fold increase in activity or selectivity of the targeted reaction. However, DE is based on the generation of thousands of variants which result in a time-consuming process, without any knowledge on how the enzyme structural and dynamic properties have been modified. In this context, there is a need for more rational techniques that could be used to generate new enzyme variants with improved activity and broader substrate scope using less resources.

In this line, recent advances in Molecular Dynamics (MD) simulations allowed us to better understand the importance of the conformational diversity of enzymes, harnessing this information to rationally tune them through point-mutations to explore catalytically relevant conformations. Such conformational diversity can

be mapped in the so-called Free Energy Landscape (FEL), where the populations of the different conformations can be extracted. The introduction of mutations in the enzyme sequence can produce a redistribution in the populations of the different accessed conformations. Therefore, enzyme engineering can produce a population shift in the FEL that can be related with enhanced activity or selectivity. However, reconstructing the FEL and recovering the thermodynamic and kinetic properties of the system from MD simulations is challenging. The amount of gathered simulation time could limit the proper quantification of the thermodynamic and kinetic properties mapped in the FEL. Many computational methods can be used to enhance conformational sampling and overcome large energy barriers that separate relevant conformational states. Among these methods some examples are Markov State Models, accelerated Molecular Dynamics, and metadynamics. These techniques can be used to efficiently reconstruct the FEL of enzyme variants and could serve to rationalize the effect of the mutations introduced by DE evolution. However, how can we identify positions that can be targeted for mutagenesis to populate catalytically relevant conformations? To that aim, correlation-based tools, in combinations with MD-based methods, are a good strategy to determine positions whose motions are correlated with the global enzyme dynamics. In this context, our group developed a correlation-based tool called Shortest Path Map (SPM) that uses graph theory to determine highly correlated residues. This tool is useful to identify the mutations introduced along the DE evolutionary pathway, but to target positions for mutagenesis as well. Ideally, the development of a computational routine to understand enzyme conformational dynamics and to detect positions that could be targeted for mutations can help to design enhanced active enzyme variants.

Nowadays, the efficient development of enzyme variants is increasing the number of synthetic methodologies to produce pharmaceutical compounds. In

this context, amines synthesis is still mostly limited to traditional synthetic methods, being only primary and some secondary amines accessible. The use of biocatalysts could improve the production of chiral amines, reaching substitutions that were still inaccessible. Monoamine Oxidase from *Aspergillus niger* (MAO-N) has proven to be a good candidate for its use at industrial scale for the synthesis of a wide amine scope, when it is used in combination of a non-selective chemical reductant. DE experiments generated a smart library of variants, *i.e.* MAO-N D3, D5, D9, and D11, that include mutations not only at the active site, but also at the substrate tunnel and at distal positions. The DE allowed to access a wide range of amine substrates, displaying activities towards primary, secondary, and even polycyclic aromatic amines. Regardless of the high active variants, the effect of the introduced mutations in the enzyme conformational dynamics of MAO-N was still unknown.

In this context, this doctoral thesis is focused on elucidating the role of conformational dynamics of a set of MAO-N variants towards the production of chiral amines. In chapter IV, MD simulations in combination with enhanced sampling methods were used to understand the role of mutations in the conformational dynamics of MAO-N WT, D3, D5, and D9 variants. To this end, the free-energy landscape was reconstructed retrieving the thermodynamic and kinetic properties. This led to the identification of catalytically relevant conformational states of MAO-N and it was possible to rationalize how they affect the substrate binding process. In chapter V, a potential interplay between MAO-N subunits for efficient catalysis was explored. The SPM tool was used to rationalize the inter-monomer communication. Enhanced sampling techniques were used to rationalize how the conformation of one subunit affects the productive binding of the substrate in the other subunit. Finally, in chapter VI, the main conclusions that arise from previous results chapters are summarized.

Resum de la tesi

Els enzims han evolucionat durant anys fins a convertir-se en grans catalitzadors que presenten una alta selectivitat, especificitat i activitat. Inicialment, els enzims es van entendre com a estructures moleculars estàtiques (bio) capaces d'accelerar les reaccions químiques per molts ordres de magnitud. No obstant això, després de dècades d'investigació, es podrien identificar com a entitats dinàmiques, capaces d'adoptar diferents conformacions tèrmicament accessibles per a la seva funció catalítica. Els avantatges que els enzims presenten a l'hora d'augmentar la velocitat de les reaccions (bio)químiques es poden utilitzar per produir compostos industrialment rellevants. En aquest sentit, els enzims presenten una limitació, en la majoria dels casos no són prou actius cap a reaccions o substrats no naturals, rellevants per a fins industrials. Aquesta limitació es pot superar mitjançant enginyeria enzimàtica, que consisteix a introduir mutacions en la seqüència enzimàtica, ja sigui mitjançant protocols computacionals o experimentals. Experimentalment, una de les tècniques més utilitzades és l'evolució dirigida (DE), que realitza mutacions aleatòries al llarg de la seqüència enzimàtica, produint variants amb diversos augment de l'activitat o selectivitat de la reacció dirigida. Tanmateix, DE es basa en la generació de milers de variants que donen lloc a un procés que consumeix temps, sense cap coneixement sobre com s'han modificat les propietats estructurals i dinàmiques enzimàtiques. En aquest context, calen tècniques més racionals que es puguin utilitzar per generar noves variants d'enzims amb una activitat millorada i un àmbit de substrat més ampli mitjançant menys recursos.

En aquesta línia, els recents avenços en simulacions de dinàmiques moleculars (MD) ens van permetre comprendre millor la importància de la diversitat conformacional dels enzims, aprofitant aquesta informació per sintonitzar-les racionalment mitjançant mutacions puntuals per explorar conformacions

catalíticament rellevants. Aquesta diversitat conformacional es pot associar en l'anomenada superfície d'energia lliure (FEL), on es poden extreure les poblacions de les diferents conformacions. La introducció de mutacions en la seqüència enzimàtica pot produir una redistribució en les poblacions de les diferents conformacions accedides. Per tant, l'enginyeria enzimàtica pot produir un canvi de població en el FEL que es pot relacionar amb una activitat o una selectivitat millorades. No obstant això, reconstruir el FEL i recuperar les propietats termodinàmiques i cinètiques del sistema a partir de simulacions de MD és un repte. La quantitat de temps de simulació recollida podria limitar la quantificació adequada de les propietats termodinàmiques i cinètiques mapejades a la FEL. Es poden utilitzar molts mètodes computacionals per millorar el mostreig conformacional i superar les grans barreres energètiques que separen els estats conformacionals rellevants. Entre aquests mètodes, alguns exemples són els models d'estat de Markov, la dinàmica molecular accelerada i la metadinàmica. Aquestes tècniques es poden utilitzar per reconstruir eficientment el FEL de les variants enzimàtiques i podrien servir per racionalitzar l'efecte de les mutacions introduïdes per l'evolució de DE. Tanmateix, com podem identificar posicions que es poden orientar a la mutagènesi per poblar conformacions rellevants catalíticament? Segons aquest objectiu, les eines basades en la correlació, en combinacions amb mètodes basats en MD, són una bona estratègia per determinar posicions els moviments dels quals estan correlacionats amb la dinàmica enzimàtica global. En aquest context, el nostre grup va desenvolupar una eina basada en correlació anomenada Shortest Path Map (SPM) que utilitza la teoria de gràfics per determinar residus altament correlacionats. Aquesta eina és útil per identificar les mutacions introduïdes al llarg de la via evolutiva de DE, però també per orientar les posicions de la mutagènesi. L'ideal seria que el desenvolupament d'una rutina computacional per comprendre la dinàmica conformacional dels

enzims i per detectar posicions que puguin orientar -se a mutacions pot ajudar a dissenyar variants enzimàtiques actives millorades.

Avui en dia, el desenvolupament eficient de les variants enzimàtiques augmenta el nombre de metodologies sintètiques per produir compostos farmacèutics. En aquest context, la síntesi d'amines encara es limita majoritàriament a mètodes sintètics tradicionals, sent només accessibles les amines primàries i algunes secundàries. L'ús de biocatalitzadors podria millorar la producció d'amines quirals, arribant a substitucions que encara eren inaccessibles. La monoamina oxidasa d'Aspergillus niger (MAO-N) ha demostrat ser un bon candidat per al seu ús a escala industrial per a la síntesi d'un ampli rang d'amines, quan s'utilitza en combinació d'un reductor químic no selectiu. Els experiments van generar una biblioteca intel·ligent de variants, és a dir, MAO-N D3, D5, D9 i D11, que inclouen mutacions no només al lloc actiu, sinó també al túnel del substrat i a les posicions distals. L'evolució del laboratori va permetre accedir a una àmplia gamma de substrats amina, mostrant activitats cap a amines aromàtiques primàries, secundàries i fins i tot policícliques. Independentment de les altes variants actives, encara es desconeix l'efecte de les mutacions introduïdes en la dinàmica conformacional de l'enzim de MAO-N.

En aquest context, aquesta tesi doctoral està enfocada a dilucidar el paper de la dinàmica conformacional d'un conjunt de variants MAO-N cap a la producció d'amines quirals. Al capítol IV, es van utilitzar simulacions de MD en combinació amb mètodes de mostreig millorats per comprendre el paper de les mutacions en la dinàmica conformacional de les variants MAO-N WT, D3, D5 i D9. Per a això, es va reconstruir la superfície d'energoa lliure recuperant les propietats termodinàmiques i cinètiques. Això va provocar la identificació d'estats conformacionals rellevants catalíticament de MAO-N i es va poder racionalitzar com afecten el procés d'unió del substrat. Al capítol V, es va explorar una

potencial interacció entre subunitats de la MAO-N per a una catàlisi eficient. L'eina SPM es va utilitzar per racionalitzar la comunicació inter-monomer. Es van utilitzar tècniques de mostreig millorades per racionalitzar com la conformació d'una subunitat afecta la unió productiva del substrat a l'altra subunitat. Finalment, al capítol VI, es resumeixen les principals conclusions que es deriven dels capítols de resultats anteriors.

Resumen de la tesis

Las enzimas han evolucionado a través de años hasta que se convierten en grandes catalizadores que presentan una alta selectividad, especificidad y actividad. Inicialmente, las enzimas se entendieron biomoléculas capaces de acelerar las reacciones químicas por muchos órdenes de magnitud. No obstante, después de décadas de investigación, podrían identificarse como entidades dinámicas, capaces de adoptar diferentes conformaciones clave, térmicamente accesibles, para su función catalítica. Las ventajas que las enzimas presentes cuando se trata de aumentar la velocidad de reacción de las reacciones (bio)químicas se pueden utilizar para producir compuestos industrialmente relevantes. En este sentido, las enzimas presentan una limitación, en la mayoría de los casos no son lo suficientemente activas hacia las reacciones o sustratos no naturales relevantes para fines industriales. Dicha limitación se puede superar mediante la ingeniería enzimática, que consiste en introducir mutaciones en la secuencia enzimática utilizando protocolos computacionales o experimentales. Experimentalmente, una de las técnicas más utilizadas es la evolución dirigida (DE), que realiza mutaciones aleatorias a lo largo de la secuencia enzimática, produciendo variantes con un aumento de varios ordenes de magnitud en la actividad o selectividad de una reacción específica. Sin embargo, DE se basa en la generación de miles de variantes que dan como resultado un proceso que requiere mucho tiempo, sin ningún conocimiento sobre cómo se han modificado las propiedades estructurales y dinámicas de la enzima. En este contexto, existe una necesidad de técnicas más racionales que podrían usarse para generar nuevas variantes enzimáticas con una actividad mejorada y un rango de sustratos más amplio utilizando menos recursos.

En esta línea, los avances recientes en las simulaciones de Dinámica Molecular (MD) nos permitieron comprender mejor la importancia de la diversidad conformacional de las enzimas, aprovechando esta información para modificarlas racionalmente a través de las mutaciones puntuales para explorar conformaciones catalíticamente relevantes. Dicha diversidad conformacional se puede mapear en el llamado superficie de energía libre (FEL), donde se pueden extraer las poblaciones de las diferentes conformaciones. La introducción de mutaciones en la secuencia enzimática puede producir una redistribución en las poblaciones de las diferentes conformaciones de acceso. Por lo tanto, la ingeniería enzimática puede producir un cambio de población e la FEL que puede estar relacionado con una mayor actividad o selectividad. Sin embargo, es muy computacionalmente caro reconstruir la FEL y recuperar las propiedades termodinámicas y cinéticas del sistema a partir de simulaciones MD. La cantidad de tiempo de simulación reunido podría limitar la cuantificación adecuada de las propiedades termodinámicas y cinéticas asignadas en la FEL. Se pueden usar muchos métodos computacionales para mejorar el muestreo conformacional y superar barreras energéticas que separan los estados conformacionales relevantes. Entre estos métodos, algunos ejemplos son los estados modelo de Markov, la dinámica molecular acelerada y la metadinámica. Estas técnicas pueden usarse para reconstruir eficientemente la FEL de las variantes enzimáticas y podrían servir para racionalizar el efecto de las mutaciones introducidas por la evolución de. Sin embargo, ¿cómo podemos identificar posiciones que pueden ser dirigidas a la mutagénesis para poblar conformaciones catalíticamente relevantes? Con ese objetivo, las herramientas basadas en la correlación, en combinación con métodos basados en MD, son una buena estrategia para determinar las posiciones cuyos movimientos están correlacionadas con la dinámica de la enzima global. En este contexto, nuestro grupo desarrolló una herramienta basada en correlación llamada Shortest Path Map (SPM) que utiliza la teoría de grafos para determinar los residuos altamente

correlacionados. Esta herramienta es útil para identificar las mutaciones introducidas a lo largo de la vía evolutiva, pero también para dirigir posiciones para la mutagénesis. Idealmente, el desarrollo de una rutina computacional para comprender la dinámica conformacional enzimática y detectar posiciones que podrían ser dirigidas a mutaciones puede ayudar a diseñar variantes de enzimas activas mejoradas.

Hoy en día, el desarrollo eficiente de variantes enzimáticas está aumentando el número de metodologías sintéticas para producir compuestos farmacéuticos. En este contexto, la síntesis de aminas todavía se limita principalmente a los métodos sintéticos tradicionales, siendo solo aminas primarias y algunas secundarias accesibles. El uso de biocatalizadores podría mejorar la producción de aminas quirales, alcanzando sustituciones que aún eran inaccesibles. La monoamino oxidasa de Aspergillus niger (MAO-N) ha demostrado ser un buen candidato para su uso a escala industrial para la síntesis de un amplio alcance de amina, cuando se usa en combinación de un reductor químico no selectivo. Los experimentos de DE generaron una biblioteca de variantes, es decir, MAO-N D3, D5, D9 y D11, que incluyen mutaciones no solo en el sitio activo, sino también en el túnel del sustrato y en posiciones distales. La evolución del laboratorio permitió acceder a una amplia gama de sustratos de amina, mostrando actividades hacia aminas aromáticas primarias, secundarias e incluso policíclicas. A pesar de la alta actividad que mostraron estas variantes, aún se desconocía el efecto de las mutaciones introducidas en la dinámica conformacional enzimática de Mao-N.

En este contexto, esta tesis doctoral se centra en elucidar el papel de la dinámica conformacional de un conjunto de variantes MAO-N hacia la producción de aminas quirales. En el Capítulo IV, se utilizaron simulaciones MD en combinación con métodos de muestreo mejorados para comprender el

papel de las mutaciones en la dinámica conformacional de las variantes MAO-N WT, D3, D5 y D9. Con este fin, el paisaje de energía libre se reconstruyó recuperando las propiedades termodinámicas y cinéticas. Esto condujo a la identificación de estados conformacionales catalíticamente relevantes de Mao-N y fue posible racionalizar cómo afectan el proceso de unión al sustrato. En el Capítulo V, se exploró una posible interacción entre las subunidades MAO-N para catálisis eficiente. La herramienta SPM se utilizó para racionalizar la comunicación entre monómeros. Se utilizaron técnicas de muestreo mejoradas para racionalizar cómo la conformación de una subunidad afecta la unión productiva del sustrato en la otra subunidad. Finalmente, en el Capítulo VI, se resumen las principales conclusiones que surgen de los capítulos de resultados anteriores.

Chapter I. Introduction

1.1. Enzyme catalysis

Through ages, enzymes have been able to evolve, until becoming the great catalysts we know nowadays. Since ancient times, humankind has taken advantage of this catalytic power, as can be seen in some Egyptian pictures, for the preservation of food and alcoholic drinks, to cite some examples.[1,2] It is well-known that enzymatic reactions are the basis of life, however, it was not until the 19th century that the scientific basis of enzyme catalysis started to be established. One of the most important contributions in this field was postulated in 1894 by Emil Fischer, the so-called 'lock-and-key' model.[3] Fischer was investigating the enzymatic hydrolysis of glucosides when he noticed the selectivity displayed by the invertin and emulsin enzymes, which were able to selectively hydrolyze the alpha or beta-methyl-D-glucoside isomer respectively. This fact led Fischer to assume that the activity of an enzyme arises from a geometrical complementarity between the enzyme active site and the substrate molecule. These relevant insights inspired many new investigations that paved the way towards the current view we have about enzymes and enzyme catalysis, demonstrating that this first idea was not properly describing the mechanism of enzymatic reactions.

1.1.1. General Aspects of Enzymes

Enzymes are biomolecular structures composed by individual building blocks called amino acids. These amino acids are covalently bound, forming peptide bonds between the carboxylic group of one amino acid and the amine group of the next one, reaction that release a water molecule.[4] There are 20 natural amino acids that can occupy the different enzyme positions. The identity and order of each amino acid, *i.e.* amino acid sequence, confer the enzyme different structural and functional traits that can lead to different three-dimensional structures and, thus, different displayed catalytic activities. Hence, the enzyme structure is a consequence of the amino acid

sequence and the different non-covalent interactions established between amino acids (either with backbone or side-chain atoms).

In this regard, we can talk about the enzyme structure at different levels. First, we have the primary structure, which is composed by the linear amino acid chain linked by peptide bonds. The local folding of enzyme sequence regions leads to organized secondary structures (*e.g.* alpha-helix and β-pleated sheet). The alpha-helix, one of the most common secondary structure in proteins, is stabilized by hydrogen bond networks from backbone atoms, frequently forming a helical structure with 4 amino acids in each gyration, which means that the C=O group of aminoacid at position *i* is interacting with a hydrogen bond with the NH group of amino acid *i*+4, with the side chains pointing outside the helix. The β-pleated sheet is another common secondary structure observed in proteins. It consists of different β-strands directly connected from at least two or three backbone hydrogen bonds. The length of the strands can vary from three to ten amino acids, but also its direction, leading to parallel or antiparallel β-sheets. Another interesting example is the β-hairpin, which is composed by two antiparallel β-sheets connected through a loop whose length is not predefined. The overall folding of the polypeptide chain, that contains the different secondary structural elements give as a result a 3D structure, *i.e.* tertiary structure. Finally, the association of different enzyme subunits forms the quaternary structure. In this case, each subunit has its own primary, secondary and tertiary structure, and the subunits are held with intermolecular hydrogen bonds or van der Waals interactions between the different enzyme subunits. The overall arrangement of the different subunits in the quaternary structure has been demonstrated to be crucial for enzyme function in multimeric enzymes.

In line with the *lock-and-key* model from Emil Fischer, the folding of the enzyme sequence into a 3D structure generates binding pockets, whose size, shape and electrostatic properties are complementary to the substrate molecule. From this

model, the specificity of enzymes to react with a small range of substrates arises from the complementarity between active site and substrate. However, it has been demonstrated many times that a plethora of enzymes are also able to catalyze other reactions apart from their natural ones, *i.e.* promiscuous enzymes, as well as some other that are able to accept a broad substrate scope, *i.e.* substrate promiscuity.[5] Such promiscuity is in part explained by the inherent flexibility that enzymes exhibit.

1.1.2. Basis of Enzyme kinetics

Enzymes are known for increasing the reaction rate of chemical processes up to 10^7-fold. This increase in activity has been under debate for many years. As catalysts, enzymes are able to increase the rate of a given reaction without being consumed and without altering the equilibrium between reactants and products. The increase in the reaction rate is commonly attributed to their ability to lower the free energy activation barrier ($\Delta G_{act}^{\ddagger}$) (Figure 1.1).[6] Many hypotheses have been developed aiming to explain their catalytic power, being the electrostatic stabilization of the transition state the most well-established.[7–9] Notwithstanding, enzymes are more complex than it was initially thought, and, far from their view as static structures, they are able to adopt different conformations that can potentially play a key role in catalysis (detailed explanation in Chapter 1.2).

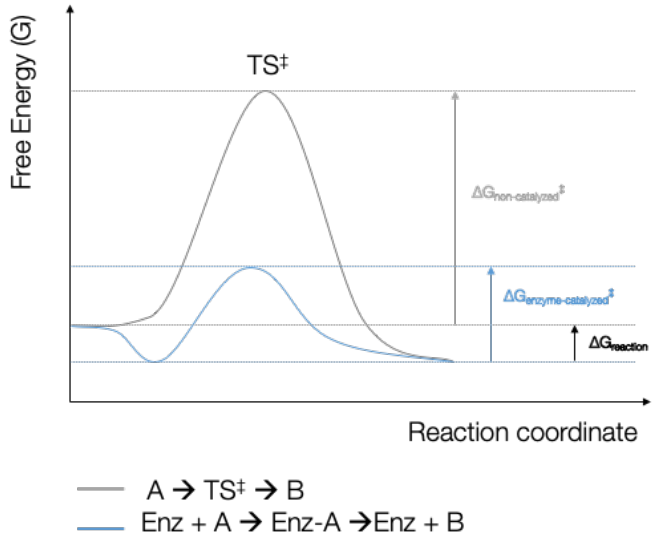


Figure 1. 1. *Representation on the energy profile of a non-catalyzed reaction (grey), and an enzymatic reaction (blue)*

Chemical reactions are described based on stoichiometry, mechanism, and order. The order represents the dependency of the reaction rate on the concentration of reactants. Depending on the reaction order, the function that describes the reaction rate can vary significantly. For first order reactions, there is a linear dependency of the reaction rate on the substrate concentration.[10] Notwithstanding, enzymatic reactions are reversible, and an additional state between reactants and products is present, being these equations not suitable for describing the kinetics of the process. The simplest enzymatic reaction can be represented as follows:

$$E+S \rightleftharpoons ES \rightleftharpoons E+P \qquad (1.\ 1)$$

where E represents the free enzyme, S the substrate, ES the enzyme-substrate complex, and P the products. The first reaction step consists in the substrate binding into the enzyme active site, while the second step represents the chemical step of the reaction. Each step has its own reaction rate, and, as the reactions are reversible, the

two pathways, *i.e.* forward and reverse, need to be considered. In this regard, the equations that describe the reaction rates for each pathway and each step are the following:

$$r_1 = k_1[E][S];\ r_{-1} = k_{-1}[ES];\ r_2 = k_2[ES];\ r_{-2} = k_{-2}[E][P] \qquad (1.\ 2)$$

where r_1 represents the rate of the forward step 1, r_{-1} the rate of the reverse pathway of step 1, r_2 the rate of step 2, and r_{-2} the rate of the reverse step 2. From these equations, we could think that by either increasing the substrate *[S]* or enzyme *[E]* concentration the reaction rates can be increased indefinitely. Notwithstanding, in enzymatic processes, the behavior of the rate functions is not linear, and they increase up to a certain value of *[S]*, where the reaction rate reaches a maximum value (V_{max}). This phenomenon is known as saturation effect.

Saturation effects do not occur in uncatalyzed chemical reactions. Its origin comes from the substrate binding processes that lead to saturated active sites. Many mathematical models were developed to explain the kinetics of (bio)chemical transformations, with the Michaelis-Menten being the most used model.

1.1.3. Mathematical description of enzyme kinetics: Michaelis-Menten model

This mathematical model was proposed by Leonor Michaelis and Maude Menten in 1913 to quantitatively represent the kinetics of enzymatic transformations. [11,12] This model considers some approximations to fully explain enzymatic chemical reactions. First, the model ignores the reverse reaction by measuring initial velocities, *i.e.* measuring the velocities at a time *t* when the product has been barely generated. Secondly, the substrate concentration is orders of magnitude higher than enzyme concentration, which makes the latter negligible in the mathematical

treatment. Finally, the rate of the product formation (k_2 in equation 1.3), is much slower than the substrate binding process. The last assumption describes the chemical step as the *rate-limiting* step of the overall process, thus, considering k_2 as the reaction catalytic rate (k_{cat}).

$$E+S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E+P \qquad (1.3)$$

Considering these approximations, the Michaelis-Menten model describes the reaction rate as:

$$V = d[P]/dt = k_2[ES] = k_{cat}[ES] \qquad (1.4)$$

However, enzymatic reactions are not always following a concerted mechanism, but instead they are a set of subsequent reaction steps, where different intermediates are formed. For those processes, the rate of all the chemical steps should be considered in the kinetic description, thus, being k_{cat} a compilation of different rate constants.

One of the approximations of the Michaelis-Menten model, is that [E] is constant (and orders of magnitude lower than [S]). The consequence of this assumption is that the concentration of ES remains constant, as the rates of formation and decomposition are supposed to be equal:

$$d[ES]/dt = 0; \quad k_1[E][S] = k_{-1}[ES] + k_{cat}[ES] \qquad (1.5)$$

Considering all the assumptions of this model the fundamental equation of the Michaelis-Menten for describing enzyme kinetics is as follows:

$$V = \frac{k_{cat}\,[S]\,[E]_0}{K_M + [S]} \quad (1.6)$$

where K_M is the so-called Michaelis-Menten constant that includes the three rate constants $K_M = (k_{-1} + k_{cat})/k_1$. Considering that enzymatic reactions are due at saturation conditions, $[E]_0$ is assumed to be equal to [ES], and the reaction is taking place at V_{max} (being $V_{max}=k_{cat}[E]_0$) the final expression of the Michaelis-Menten model is the following:

$$V = \frac{V_{max}\,[S]}{K_M + [S]} \quad (1.7)$$

This model is still used to describe the kinetics of natural enzymes. Notwithstanding, for complex reactions that might involve allosteric effects, or for reaction mechanisms that involve multiple intermediates, this model is not fully explaining the enzymatic kinetic mechanisms. In this regard, it is possible to measure the binding of ligands as a function of its concentration by means of the Hill equation.[13] However, faster and more sophisticated techniques, like stopped flow kinetics, that are able to characterize the kinetics in the range of milliseconds (ms), are required to understand complex reactions.[14,15]

1.2. Conformational Dynamics

Enzyme's ability to accelerate chemical reactions is attributed to their capability to stabilize the transition state due to a highly preorganized active site that might be optimum for a specific transition state. Notwithstanding, the pre-organization of the active site residues, as well as some sort of flexibility in the active site, might be crucial for substrate binding and product releases processes. Indeed, enzymes are able to adopt many thermally accessible conformations that differ from the observed by the single structure captured by X-ray data. Such dynamism is crucial

for understanding enzyme catalysis, as these variety of conformations might be key for substrate binding or product release processes. Thus, tuning the populations of these conformations can lead to a more efficient enzyme with higher and/or novel activities.[16]

1.2.1. Defining enzyme's dynamic nature: the conformational heterogeneity concept

Since long ago, enzymes have been considered as static structures that are able to catalyze (bio)chemical reactions by stabilizing the transition state of the transformation. Notwithstanding, in 1958, Daniel E. Koshland postulated the *induced fit* model.[17] Koshland expanded the *lock and key* model from Emil Fischer, as he realized that in many enzymes the isomer that was not supposed to react was acting as an inhibitor. Moreover, the *lock-and-key* model did not account for the capability of some enzymes to perform other reactions apart from their natural one, fact that confer enzymes with the ability to perform other reactions or accept a wide substrate scope (promiscuity). From these observations, he proposed the *induced fit* model, in which, apart from the complementarity between active site and substrate (where the precise orientation of the catalytic residues produces catalysis), he stated that the substrate induces a change in the 3D orientation of the side chains of the active site residues providing the proper pre-organization for the reaction. Thus, this model states that whereas an active substrate induces these changes in the active site, a non-active substrate, despite being able to bind to the active site, will not produce its reorganization, pointing that the exploration of conformational changes is key for enzyme activity. In this regard, other hypotheses emerged, such as the *conformational selection* model suggested by Nussinov in 1999[18] based on the ideas on protein allostery proposed by Jacque Monod, Jeffries Wyman, and Jean-Pierre Changeux in 1965.[19] This model states that multiple conformations of the enzyme pre-exists in solution, whose populations

can be redistributed either by substrate/cofactor binding or by changing experimental conditions [5,20] The introduction of mutations in the enzyme sequence can also lead to a redistribution of the populations of the ensemble of conformations. These models, together with early X-ray crystallography studies, paved the way towards the current view of enzymes as “flexible” and “dynamic”.

Development of X-ray crystallography, NMR, small angle X-ray scattering, and later on cryo-electron microscopy allowed for a better understanding of enzyme structure at the atomic level, pointing that, indeed, enzymes exist as an ensemble of multiple thermally accessible conformations.[21] A good example in this regard is the kinetic characterization of Ribonuclease A. In this study, steady-state kinetic experiments were carried out to capture the intermediates present in the reaction mechanism. Ribonuclease A catalyzes the breakdown of RNA in a two-step mechanism. First, a diester bond is cleaved from pyridine phosphate, followed by its hydrolysis to produce a terminal pyrimidine 3’-phosphate. Herein, the authors were able to identify that the reaction mechanism occurs in two different conformations of the enzyme. Substrate binding produces the closing of a hinge region that creates a hydrophobic cavity for the reaction to occur. Once the reaction is completed, the hinge region needs to undergo an opening motion to perform the product release. Many years after these kinetic studies, X-ray structures were obtained with different ligands bound in the active site.[22] Such structures confirmed that the hinge region is, indeed, displaying different close/open conformations key for catalysis, demonstrating the dynamic nature of enzymes and how residues far from the active site can modulate enzyme activity.

Further studies allowed to identify the plasticity of enzymes and to determine the importance of the different adopted conformations along the catalytic process.[23,24] At this point, the enzyme’s perspective completely changes from ‘static structures’ to ‘dynamic entities’ able to perform conformational changes that drastically modify

the 3D disposition of the secondary structural elements. Advances in Nuclear Magnetic Resonance (NMR) spectroscopy shed some light on determining the timescale of the conformational transitions at atomic resolution, being able to identify transitions that range from fs up to s, depending on the nature of the motion (Figure 1.2).[25,26] More specifically, the dynamic information is extracted from NMR spin relaxation. The timescale is defined by the relaxation rate, which is related to the chemical-shift timescale, *i.e.* difference in chemical shift between the interconverting conformations. Populations of each conformation can also be estimated from the integration of the peak area.

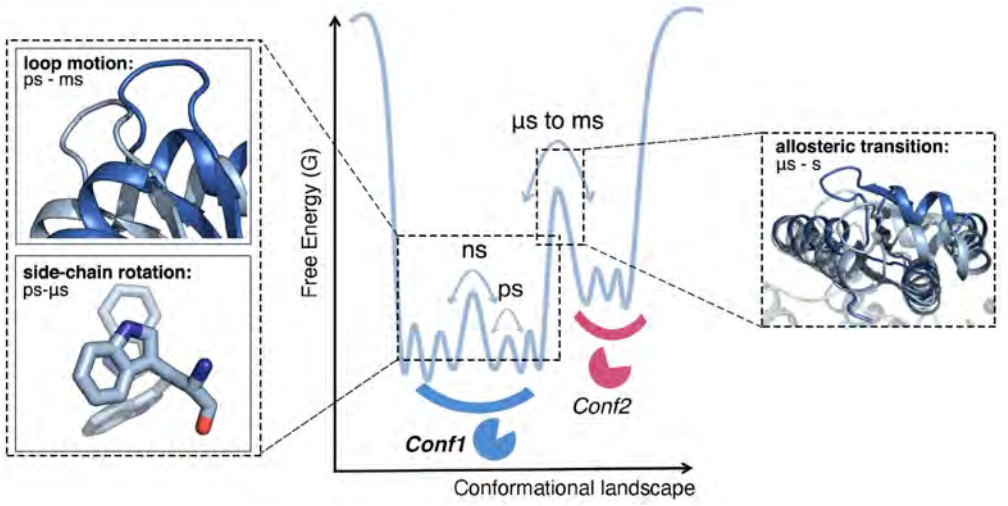


Figure 1.2. Representation of the conformational landscape (left), with the energy barriers and timescales associated to different conformational transitions (right)

As it is depicted in Figure 1.2, bond vibrations and side chain rotations occur in the shortest timescales, being from 10-100 fs for vibrations and from ps to µs in the case of side chain rotations. Loop motions, which are often related to substrate binding and product release processes, take place in a timescale that range from ns to ms. The slowest processes are domain and allosteric motions, that range from microseconds (µs) up to seconds (s). Such slow conformational transitions occurring in the range of the milliseconds (ms) to s have been proven to be directly

linked with enzyme activity in some cases.[27,28] For instance, experiments with thermophilic adenylate kinases at different temperatures support this statement. NMR measurements, together with computational simulations showed that the limited catalytic activity of the thermophilic Adenylate kinase (Adk) at 20ºC is determined by the slow dynamics of a hinge region key for catalysis.[29,30] Notwithstanding, experiments on the same thermophilic enzyme, compared to the mesophilic Adk showed that atomic fluctuations can also determine enzyme activity , as their collective motions might trigger or hamper slower domain transitions, limiting or expanding enzyme's activity.[31] This ability of enzymes to populate different conformations, *i.e.* conformational heterogeneity, can be mapped by various experimental and/or computational techniques, thus, extracting information of the population of the different conformational states, as well as the energy barriers that separate them.[32,33]

1.2.2. Depicting the free energy landscape

As it was stated in the previous sections, enzymes exist as an ensemble of multiple conformations. To understand the ability of a given enzyme to populate different conformations in solution, and to understand their role in the catalytic activity, they can be mapped in the so-called Free Energy Landscape (FEL).[34] In this FEL, all the possible states, as well as the energy barriers that separate them are represented (Figure 1.3). The height of these energy barriers is directly related to the timescale of the conformational transitions, being able to not only extract thermodynamic information but also kinetic data. The energy barrier is directly linked with the timescale of the conformational transition. Thus, lower energy barriers correspond to transitions between local minima, and usually correspond to side-chain rotations or small loop motions. On the contrary, if the energy barrier that separates two conformational states is high, the motion representative of the transition will be also slower.

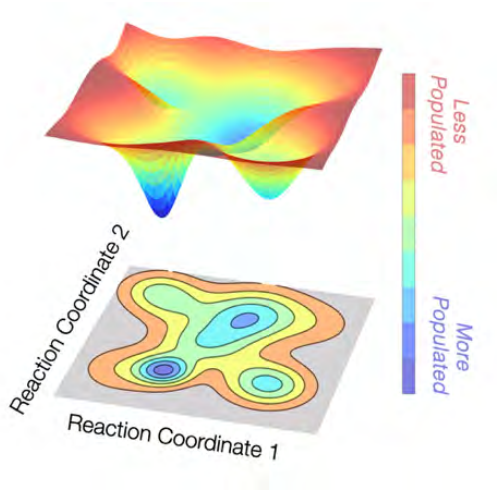


Figure 1.3. 3D representation of a FEL. Each axis corresponds to a different reaction coordinate. The depth of the point, depicted in different colors, represent the population of each visited conformation, being blue points the most populated and red point the least populated conformations

These conformational states can be either determined by experimental or computational techniques. Experimentally, X-ray crystallography, cryo-electron microscopy, NMR and other biophysical techniques can be used to that aim.[35,36] Whereas X-ray crystallography allows for the identification of single conformations, cryo-electron microscopy can be used to detect ensembles. More interesting is the information extracted from NMR experiments. Although this technique lacks for the resolution of X-ray structures, it allows for the identification of conformational transitions, permitting to determine the populations of each conformation, as well as the timescale of the transitions.

Computational methods are particularly interesting for reconstructing the FEL of a given biomolecule. By means of Molecular Dynamics (MD)-based techniques, the free energy (*G*) of each conformation can be estimated as a function of the probability distribution (P):

$$G \sim -k_B T \log(P) \qquad (1.\ 8)$$

Notwithstanding, it is a requirement to extensively sample the conformational space and to sample the transition between conformational states to have an accurate estimation of the probability distribution to retrieve populations and energy barriers from the FEL.[37] Thus, the higher the number of times a transition is sampled, the lower will be the error in the estimation. A maximum in the distribution represents an energy minimum, *i.e.* a highly stable conformation. In the next section, a deeper explanation on the different computational techniques for the FEL reconstruction will be addressed.

1.2.3. Computational methods for free energy landscape reconstruction

MD-based techniques permit the estimation of the probability distribution at the atomic level. Notwithstanding, enzymes are large systems, containing thousands of atoms. As a consequence, the number of Degrees of Freedom (DOFs) is even higher, hampering the reconstruction of the FEL. To have a representative FEL, one must select the reaction coordinates that describe the most representative motions of a given enzyme. A simple solution could be selecting a set of DOF which describe the motions that are relevant to the studied process. These DOFs can be any function of the coordinates of the enzyme, such as distances between residues, angles, dihedrals and so on. If the nature of the motion is more complex than that, dimensionality reduction techniques, such as Principal Component Analysis can be used, which accounts for the maximization of the variance in a set of DOFs (more details in section 2.3). [38,39]

One limitation of MD simulations is the recovery of the thermodynamic properties of the system, as the sampling of the different conformations must reach equilibrium, *i.e.* ergodic principle. The usual protocols for MD simulation use

timescales in the order of microseconds, limiting the amount of simulation time that one could gather, thus, hampering the accessibility of all conformational states and frustrating the comparison of the simulations with experimental data. Alternatively, different enhanced sampling methods can be applied, facilitating the exploration of conformational transitions by application of different statistical methods, or by adding external potentials to our simulations. Among these methods, we have to distinguish between biased and unbiased MD-based approaches.

Conventional MD simulations are those methods called 'unbiased'. The protocol and algorithms used in these methods have barely changed in the last decades, just modifying the parameters and equations included in the force fields, that serve to compute the energy of the system as well as the position and velocities at the atomic level. Notwithstanding, technology is evolving, and new strategies have been developed to increase the accessibility to larger timescales that could serve to relate our observations with experimental data. CPU and GPU parallelization allows for an increase in the accessibility of higher timescales.[40,41] It is worth highlighting the latter, in which the MD codes have been especially created, or adapted, to run specifically on GPUs. Multiple MD replicas can be performed on GPUs to promote infrequent transitions between conformational states, thus, resulting in a huge data set that needs to be statistically treated to facilitate the FEL reconstruction. One of the most efficient methods to characterize the FEL in biomolecular systems by means of unbiased MD simulations, is Markov State Models (MSM).[42–44] This technique consists in running a large set of simulations until exploring all the conformational transition(s) of interest. Statistically, one can extract thermodynamic and kinetic data that can be related to experimental data, obtaining the populations of each conformation as well as the timescale related to the transitions between them

Although these unbiased methods are reliable to quantitatively characterize the thermodynamics and kinetics of a given system, the amount of simulation time required to sample the transitions of interest can hamper their application. In this regard, biased methods are a good alternative to enhance the FEL sampling. The amount of prior information about the possible structures is determinant to select the proper biased method. For instance, when only one conformational state is known, accelerated Molecular Dynamics (aMD) is an advantageous technique to be used.[45] It is based on the application of a bias potential, flattening the FEL and facilitating the capability to sample all possible conformations. When the initial and final states are known, methods that can sample the transition between these two conformations by sampling a set of DOFs, can be used. In this regard, one of the most used techniques is Metadynamics, which is based on the addition of repulsive potentials on a set of DOFs.[46] By adding those potentials after a certain number of MD steps, the system is forced to escape from energy minima, thus, exploring regions higher in energy and being able to overcome the energy barriers. Finally, when the initial and final states are known, as well as some of the intermediates, Umbrella Sampling can be used to run multiple independent MD simulations.[47] In this case, a bias potential is applied at small increments until exploring a certain transition along the reaction coordinate. By means of any of the above-mentioned methods, the FEL of biomolecular systems can be easily reconstructed, either by applying dimensionality reduction techniques, or by applying mathematical transformations to the extra bias potentials added to our simulations. A more detailed explanation of these techniques can be found in the Methodology section.

1.3. Engineering enzymes: evolution towards enhanced activity

The advantage of performing enzymatic catalysis over traditional synthetic methods have raised their interest for being used in industrial processes. Notwithstanding, as they are usually utilized for reactions that are not their natural ones, there is a need to evolve/engineer enzymes for improving reactivity, enantioselectivity, thermodynamic stability, and substrate specificity among other parameters. In this regard, enzyme engineering is focused on the generation of new enzyme variants with improved properties towards the desired reaction under specific reaction conditions. Many experimental techniques can be used to that aim. One of the most successful examples is Directed Evolution (DE), which was awarded with the Nobel prize in 2018 to Prof. Frances Arnold. DE consists in iterative cycles of random mutagenesis along the enzyme sequence, generating a huge smart library of variants.[48–50] The advantage of DE is that mutations are not only introduced in the active site, but to the whole sequence, highlighting the impact of distal mutations on the enzyme activity.

Regardless of the capability of DE to generate highly active enzyme variants, it generates thousands of new sequences, which requires tons of resources, as well as being a time-consuming technique. In this regard, there is a need for new tools for predicting enzyme positions that can be mutated to potentially evolve the enzyme towards the desired reactivity. In this context, more rational strategies for enzyme engineering can be applied. Among these methods, some examples are site-saturation mutagenesis (SSM) and iterative saturation mutagenesis (ISM). [51,52] In SSM a set of natural amino acids are tested in rationally selected positions, whereas ISM consists in iterative cycles of SSM at selected positions, which commonly results in library of variants of one-, two- or three-point mutations.

However, the selection of positions that can be targeted is still challenging. Commonly, apart from active site residues, a strategy that targets residues located in bottleneck areas of the substrate tunnel is used, resulting in enzyme variants with improved activity towards certain substrates.[53] Notwithstanding, it is still challenging predicting distal positions that can alter the enzyme conformational dynamics to populate active conformations for the targeted reaction.

1.3.1 Computational tools for predicting distal mutations

As mentioned in previous sections, enzymes exist as an ensemble of multiple conformations. The importance of FEL reconstruction comes from the fact that allows to identify specific conformations are key for efficient catalysis. Those conformations might be high in energy in the WT enzyme, and could be crucial for accepting non-natural substrates. By applying mutations, either at the active site or at distal positions, the populations of those conformations can be redistributed (population shift), stabilizing those that are key for accepting new substrates or to produce novel functionalities (Figure 1.4).[54–57] Therefore, in these cases the low activity of the WT enzymes can be explain in terms of the population shift concept, in which some specific conformations need to be stabilized for an improved catalysis (Figure 1.4). This was indeed confirmed by some studies on phosphotriesterase (PTE). By applying DE, mutations not only at the active site, but also at distal positions were introduced to transform PTE into an arylesterase (AE).[5,57] This change in function came thanks to the population shift induced by the mutations towards those conformations key for AE activity. Indeed, X-ray crystallographic data, together with the measurement of B-factors along the evolutionary pathway, indicated major changes in the flexibility of some key loops gating the active site of the enzyme along the PTE to AE evolution.

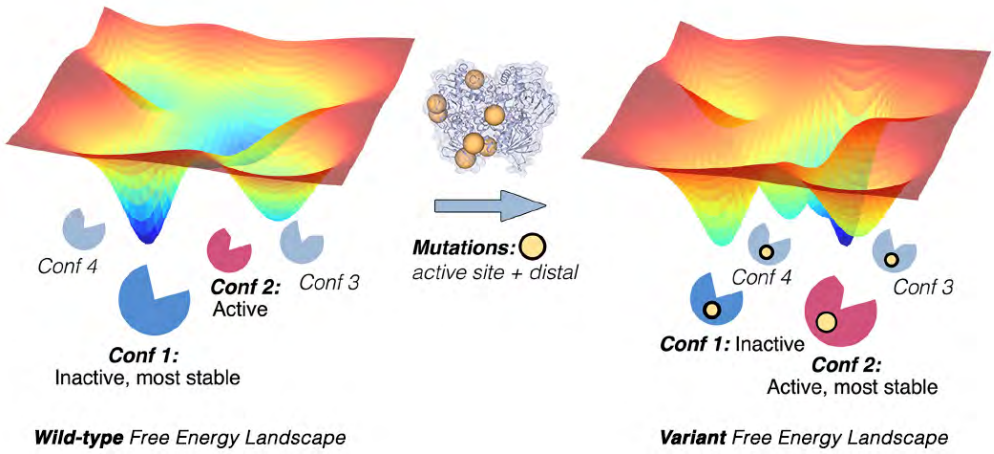


Figure 1. 4. Representation on how introduced mutations produces a population shift in enzyme FEL. Prior to introduce mutations, conformation 1 presents higher stability. Afterwards, conformation 2 has been stabilized producing a more efficient enzyme variant.

Considering the impact of distal mutations in shifting the populations of the conformational ensemble towards more active conformations, some computational tools can be used to predict which positions might affect the probability distribution of the FEL. In this regard, correlation-based tools, together with MD simulations, can be used to identify relevant positions of the enzyme sequence that can be targeted in the process of enzyme engineering towards enhanced activities.[58,59]

We can find many correlation-based tools in literature, which are based on graph theory to represent a weighted graph composed by nodes corresponding to each residue of the system or communities. Some examples of these network models are the ones provided by Sethi et al. or by Amaro and co-workers (WISP).[59,60] However, these methods provide a complicated graph network that might be difficult to understand. In our group, Shortest Path Map (SPM), a correlation based-tool that uses graph theory for predicting highly contributing residues to the enzyme dynamics was developed. In this methodology, a representation of the enzyme and its conformational dynamics is provided. By computing a distance and a correlation matrix those pairs

of residues whose carbon alpha stay below a certain threshold distance value along the MD simulation, they are connected through edges, which are weighted according to its correlation value. This leads to a highly complex graph that can be simplified by applying the Dijkstra algorithm. The simplified graph consists of a series of spheres connected through edges with different sizes depending on the importance of the given edge for the conformational dynamics of the enzyme (see more details in Methodology section). What makes the difference in the SPM compared to the rest of correlation-based tools, is the simplification of the network graph by computing the shortest paths and projecting them into the 3D structure of the enzyme. This result in a easily understandable network that could serve to target some specific positions that could potentially alter the conformational dynamics of a given enzyme. In summary, those positions that are highly correlated will present a higher sphere and edge radius, where the edges show the different conformationally-relevant hotspots that are connected with the rest of the protein.[61]

This methodology was successfully applied for rationalizing the DE of a Retro Aldolase (RA) library of variants. Microsecond timescale MD simulations were combined with SPM to study a set of computationally generated variants. The FEL of each variant was reconstructed, and a population shift towards active conformations in the most evolved variant (RA95.5-8F) was observed. When applying the SPM to the gathered microsecond MD data, from the 13 mutations introduced in the enzyme sequence, 7 positions were predicted by the SPM, 4 were located at adjacent positions, and only 2 positions were located at more than 6 residues from the SPM. This showed the SPM potential in predicting those positions that can potentially alter the population of the different conformations in the FEL of biomolecular systems. In a recent publication, the SPM methodology was applied in combination with ancestral sequence reconstruction for designing new stand-alone tryptophan synthase B variants.[62] In this thesis the SPM methodology has also been applied and used to evaluate the conformationally-relevant positions and how these are altered along DE.

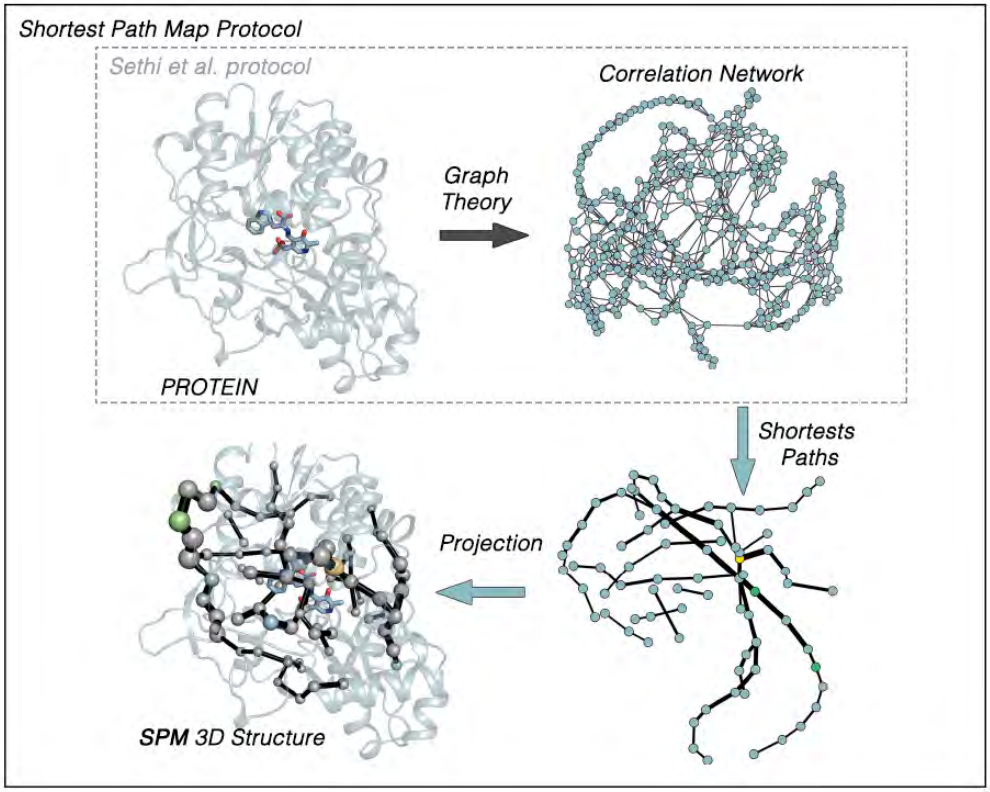


Figure 1. 5. Graphical representation of the SPM protocol compared to previous correlation-based tools.

1.4. Biocatalytic production of industrially relevant amines

Biocatalysis has attracted enough interest from industries to obtain key products for the manufacture of pharmaceutical and agrochemical compounds. In the pharmaceutical industry, almost 40% of commercially available drugs contain at least one chiral amine building block. Currently, different synthetic methodologies, commonly involving transition metal catalysts, are used for the obtention of these

building blocks. Among these methods it is worth highlighting the asymmetric reduction, nucleophilic addition or the diastereoisomeric crystallization.[63–67] Although new synthetic methodologies are still emerging, the need for more sustainable procedures for their obtention is an urge. In this context, the use of enzymes as biocatalysts for the synthesis of chiral amines, is becoming a reality.

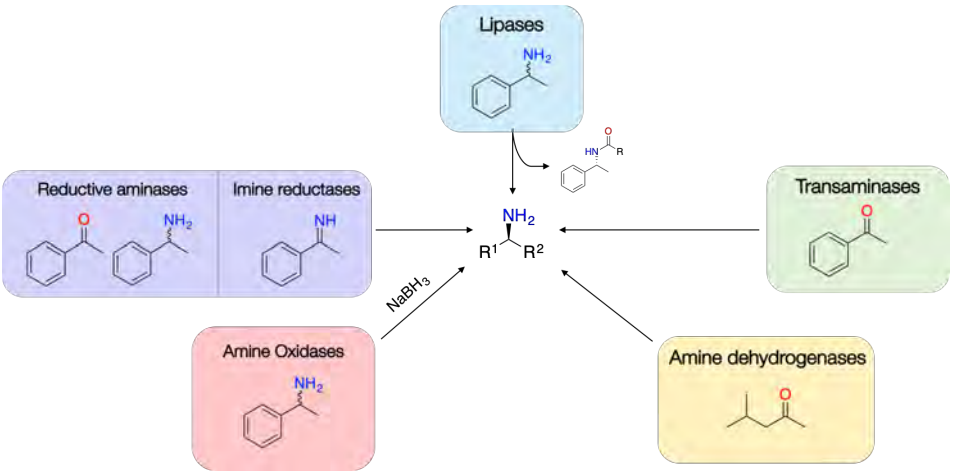


Figure 1.6. Summary of the different enzyme families that allow the conversion of different substrates into amines

Due to the advantages of enzymes over traditional metal catalysis, such as the use of water as solvent, the use of soft reaction conditions, the displayed selectivities or avoiding the formation of waste products, a significant growth of biocatalytic methodologies for the obtention of chiral amines emerged during the last decades. Among the possible enzymes, we can find amine dehydrogenases (AD), transaminases (TA), imine reductases (IREDs)/reductive aminases (RedAms), lipases, and monoamine oxidases (MAOs).[68] TAs permit the accessibility to primary amines by transferring an amine group (from an amine donor) to a carbonyl compound. Imine reductases are able to produce a chiral amine by reduction of a preformed imine by means of a hydride transfer reaction from a NADPH cofactor. RedAms, similarly to

IREDs, are able to reduce imines following the same reaction mechanism, but they are also able to produce the imine intermediate using a ketone and amine substrate.

Particularly interesting is the case of Monoamine Oxidase. More specifically, Monoamine Oxidase from *Aspergillus niger* (MAO-N) has demonstrated to be an excellent candidate for the biocatalytic production of chiral amines.[69] MAO-N is a flavoenzyme able to catalyze the oxidation of amines into imines, by means of a hydride transfer reaction from its FAD cofactor. The enzyme displayed a high enantioselectivity, being able to oxidize the (S)-enantiomer from a racemic mixture. Indeed, a methodology that involves the dynamic kinetic resolution of chiral mixtures was reported by Prof. Turner's lab, by combining MAO-N with a non-selective chemical reductant. [70]

Many studies have been carried out on this enzyme aiming to substantially expand its substrate scope, as well as their catalytic activities. From one hand, we can find the research line from Prof. Turner lab, in which DE experiments were carried out developing a full MAO-N evolutionary path, with enzyme variants that have excellent activity towards a wide range of amine substrates (Figure 1.7). Starting from the *wild-type* (WT) enzyme, that has a moderate to poor activity towards primary and some secondary amines, the first DE experiments raised the D3 variant, which include mutations far from the active site. Such mutations allowed for an improved activity towards secondary amines. Subsequent studies gave as a result the variants D5, D9 and D11, that include mutations at the substrate tunnel (D5), but also at the active site (D9 and D11).[71–73] Whereas the WT enzyme was only able to catalyze the conversion of primary amines with relatively low values of k_{cat}, D3, and D5 were able to selectively oxidize the (S)-enantiomer of secondary aromatic amines like α-methylbenzylamine (AMBA), with improved k_{cat} values. The most evolved variants, *i.e.* D9 and D11, displayed an impressive activity towards aromatic and polycyclic amines, displaying higher relative activities compared to D5 variants .[74] Regardless of the improvements

introduced by DE on MAO-N, the effect of the mutations in the conformational dynamics was still unknown (which is targeted in Chapter 4 of this thesis).

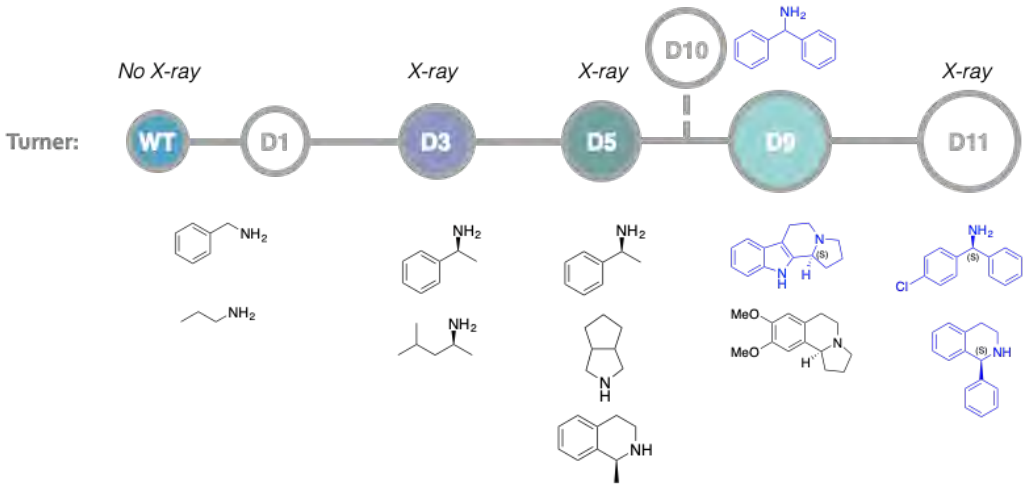


Figure 1. 7. MAO-N evolutionary path developed by Prof. Turner lab highlighting the substrate scope of each enzyme variant

In parallel, studies from Prof. Reetz lab, raised a new evolutionary line in which the enantioselectivity was reverted compared to WT.[75] In this particular case, the analysis of enzyme tunnels was crucial, in which bottleneck residues were targeted for mutagenesis aiming to improve MAO-N activity towards the oxidation of the (R)-enantiomer, resulting in new MAO-N variants with high reverted enantioselectivity.

Chapter II. Methodology

In Chapter I.3.2 we described how computational methods can be used to reconstruct enzyme conformational heterogeneity, as well as how to tune the population of the different conformations aiming to alter the reactivity of a given enzyme. Depending on the level of accuracy of the computational methods, the resources and time expensiveness may vary significantly. While Quantum Mechanical (QM) methods can serve as a tool to extract information of the studied system at subatomic level, their computational cost limits its application to systems with a few hundred of atoms. For biological systems (*e.g.* enzymes) the number of atoms range between thousands to millions, thus, making these methods hardly applicable. To study such complex systems, we can rely on Molecular Mechanics methods (MM), which do not explicitly account for quantum mechanical effects, treating the atoms as single particles that are subject to Newton's second law of classical physics. In MM, the atoms and bonds are treated following a "ball and spring" model, in which the atoms are represented as charged spheres with mass and radius, bonded by springs. The potential energy is computed by different terms of the equations included in the so-called force field (FF).

2.1. Classical Force Fields

The set of equations that serve to compute the potential energy of a MM system are referred as Force Fields (FF). These FFs are included in the MD software packages such as AMBER, GROMACS, or CHARMM. In this thesis, we used the AMBER package, whose functional form is described as follows:

$$U_{total} = U_{bonded} + U_{non-bonded} =$$
$$= \underbrace{U_{bond} + U_{bend} + U_{torsion}}_{\text{Bonded}} + \underbrace{U_{electrostatic} + U_{van\ der\ Waals}}_{\text{Non-bonded}} \quad (2.\ 9)$$

As it is depicted in Equation 1.9, the potential energy (U) of the system is computed by two different terms, *i.e. bonded* and *non-bonded*. The bonded term of the energy expression accounts for the energy of bonds, angles, and dihedrals:

$$U_{bond} = \sum_{bonds} k_{AB}(r_{AB} - r_{AB,\,eq})^2 \qquad (2.\ 10)$$

$$U_{bend} = \sum_{angles} k_{ABC}(\theta_{ABC} - \theta_{ABC,\,eq})^2 \qquad (2.\ 11)$$

$$U_{torsion} = \sum_{dihedrals} k_{ABCD}[1 + \cos(n\omega - \delta)] \qquad (2.\ 12)$$

where the bonds and bending energy function follow a harmonic approximation that varies depending on the position/angle (*r* for position and θ for the angle) of the corresponding atoms (A,B, and C) with respect to the equilibrium (r_{eq} and θ_{eq}). In all the terms *k* represents the force constant for bonds, angles and torsions, whose value depends on the nature of the bond. In other words, the harmonic constant will be different for every type of bond (*i.e.* C-C, C-N, C-O, C=C, C=N…), depending on the FF used. The latter term ($U_{torsion}$) expresses the potential energy that comes from the dihedral angles formed by 4 consecutive atoms. In this expression, *n* represents the periodicity of the cosine function, ω is the dihedral angle of atoms A-B-C-D, and δ is the phase angle.

Whereas the bonded term of the FF equation (equation 2.9) is representing the potential energies that arise from the covalent bonds of the different atoms composing the system, we still have to define those interactions that occur between non-bonded atoms, *i.e. non-bonded* interactions. This term of the FF

usually accounts for the *Van der Waals* (VdW) and the *electrostatic* interactions. The VdW term is expressed as a Lennard-Jones (L-J) potential:

$$U_{VdW} = \sum_{A<B}^{atoms} 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right] \qquad (2.\ 13)$$

where r_{ij} is the distance between interacting atoms, ε_{ij} is the L-J well depth, and σ_{ij} the distance at which the interactions are negligible (U=0), due to a balance between attractive and repulsive interactions. Finally, the electrostatic term of the potential energy is described as a Coulomb potential between two atoms with partial charges (q_i and q_j), interatomic distance (R_{ij}), and a dielectric constant (ε_0):

$$U_{electrostatic} = \sum_{A<B}^{atoms} \left(\frac{q_A q_B}{4\pi\varepsilon_0 r_{AB}}\right) \qquad (2.\ 14)$$

All these terms of the FF rely on parameters that are obtained either from experiments, or computationally by means of ab *initio* QM calculations. Among the experimental methods reference parameters are obtained using spectroscopy, X-ray crystallography, and gas- and solution-phase properties. Considering this, a good parametrization of reference values has to be ensured to have a suitable reproduction of the protein motions in computer simulations. However, biomolecular force fields do not have the parameter set for molecules outside the training set that contains information of biomolecules. For that purpose, the General Amber Force Field (GAFF) is commonly used to describe the parameters for small organic molecules. The general protocol to parametrize small organic molecules using GAFF, consist in a single point QM calculation, typically at the HF/6-31G* level on the optimized structure. The output is used to estimate the atom charges using the *restrained electrostatic potential* (RESP) method, which generates atom-

centered point charges based on a charge fitting procedure to an electrostatic potential.[76] Then, the RESP charges and connectivity are translated to those included in GAFF, where the force constant is included based on atom type, charge, and connectivity.

2.2. Generating classical trajectories by means of Molecular Dynamics

Molecular Dynamics (MD) is a computational method for the study of dynamical processes in molecular systems. In the field of biomolecules, it can serve to study conformational changes in proteins, as well as substrate binding, or allosteric transitions. It is based on the application of Newton's second law of classical physics to estimate the time evolution of atomic positions. Following Newtonian mechanics, an atom (described as a ball) would keep its motion unless it experiences a force F_i that will perturb its state. Thus, the force that arise from interactions with the rest of the atoms of the system, and considering Newton's second law, is computed as follows:

$$F_i(t) = m_i a_i(t) = -\frac{dU}{dr_i} = m_i \frac{d^2 r_i}{dt^2}; \quad i=1,2\ldots N \qquad (2.\ 15)$$

where (dU/dr_i) represents the potential energy gradient, which can be related to the force extended to any atom i with mass m_i coordinates r_i at a time t. The latter equation cannot be solved analytically, due to the *many-body* problem. However, the position of the atoms at a time t, or at an earlier time $-t$ can be approximated using a Taylor expansion.

$$r_{(t+\Delta t)} = r_t + v_t(\Delta t) + \frac{1}{2}a_t(\Delta t)^2 + \frac{1}{6}b_t(\Delta t)^3 + \ldots \quad (2.\ 16$$

$$r_{(t-\Delta t)} = r_t - v_t(\Delta t) + \frac{1}{2}a_t(\Delta t)^2 - \frac{1}{6}b_t(\Delta t)^3 + \ldots \quad (2.\ 17)$$

In these expressions, the velocities (v_i), accelerations (a_i), and hypervelocities (b_i) are computed in terms of the first, second and third derivatives of the position at a time t. However, truncation errors arise from these equations, and after some corrections, such as the use of half time-steps (*leapfrog* algorithm), and the selection of the proper time-step, the system can be solved using the *velocity Verlet* algorithm:

$$r_{(t+\Delta t)} = r_t + v_t(\Delta t) + \frac{1}{2}a_t(\Delta t)^2 \quad (2.\ 18)$$

$$v_{(t+\Delta t)} = v_t + \frac{1}{2}[a_t + a_{\Delta t}](\Delta t) \quad (2.\ 19)$$

In summary, by deriving the forces using the FF equations, the acceleration of the atoms can be integrated numerically by applying these algorithms. The position of the atoms can then be predicted at a time *t-Δt* from the previous position. Considering the timestep selected for our calculations, which is typically 2 fs, the positions and accelerations will be then predicted at a time *t* and *t*+2 fs. In a MD simulation, we then end up with a huge data set that contains the positions and velocities of all the atoms of the system, which results in a trajectory of the evolution of the system along time. Analyzing such a big data set is a complicated task that can be overcome by applying mathematical methods for reducing the number of

dimensions to those that are more significant for explaining the behavior of our (bio)chemical system

2.3. Dimensionality reduction techniques for Free Energy Landscape Reconstruction

The output obtained from MD simulations consists in a huge data set corresponding to the coordinates and velocities of all the atoms of the system along time. When analyzing the conformational changes that take place at a given simulation time, a reduction of this data set must be performed, to have a deeper understanding of the most relevant motions that our simulation has sampled. To that aim, different methods can be used, such as the definition of collective variables (CVs) that could properly describe a specific conformational transition. Examples of CVs, could be backbone dihedrals, distances, angles… However, as enzymes are complex systems formed by hundreds or amino acids, the analysis of MD simulations through CVs might be a difficult task, when a priori information is not known. For that reason, the use of dimensionality reduction techniques is crucial for understanding the motions that are relevant for understanding the activity of a given enzyme. Among these techniques it is worth highlighting the use of Principal Component Analysis (PCA), and time-lagged Independent Component Analysis (tICA). [38,39]

2.3.1. Principal Component Analysis

PCA consists in the orthogonal transformation of our data set, into a new coordinate space, that results from the maximization of the variance in our data set. This mathematical transformation can serve to summarize the behavior of our

system into a *N* dimensional space. To do so, consider a hypothetical data set formed by N points. N is the number of features (*i.e. degrees of freedom or* Collective Variables, CVs) that we want to include in our analysis. These CVs can be any degree of freedom of our studied system, *i.e.* backbone dihedrals, distances, root mean square deviation… The mathematical steps to transform those N points to a new coordinate space consist first on the calculation of the arithmetic mean value of each point. Then, a covariance matrix is computed as a way to obtain the principal axes:

$$C_{ij}= \frac{1}{N-1}\sum_{n=1}^{N}\left((x(n)- \langle x\rangle)_i(x(n)- \langle x\rangle)_j\right) \qquad (2.\ 20)$$

The matrix must be diagonalized, forming an eigensystem where $e^{(k)}$ is the eigenvector (direction), and λ_k is the eigenvalue (variances), and *k* the number of dimension, the that represents the percentage of covariance that each point has in each dimension:

$$Ce^{(k)}= \lambda_k e^{(k)} \qquad (2.\ 21)$$

Once we have the new data, we will project it to the new principal component space (V_k):

$$V_k(n)=(x(n)-\langle x\rangle)e^{(k)} \qquad (2.\ 22)$$

In this way, we could project the information obtained from a high amount of collective variables into a few dimensions. The Free Energy Landscape of biomolecular systems (see Chapter 1.2.2) can also be reconstructed from this new data set, considering the two most relevant Principal Components (PCs), i.e. those that describe the highest percentage of variance, a Boltzmann distribution is used

to compute the energy as a function of the probability of that particular value of PC (see section 2.2). It is worth highlighting that the relevance of PC dimensions is a criteria, we should select those dimensions that better describe the motions that we are interested in studying, they are not necessarily those that describe the highest percentage of variance

In summary, PCA serves as a way to capture the motions with the highest amplitude that our MD simulation has sampled, but, what if the motions that we are interested in are not the widest, but the slowest?

2.3.2. time-lagged Independent Component Analysis (tICA)

To capture the slowest motions in a MD data set, it would be convenient to use tICA as a dimensionality reduction technique. Whereas PCA is based on the maximization of the variance, tICA is focused on maximizing the autocorrelation of the CVs that we are interested in monitoring, allowing us to capture the slowest conformational changes that take place in our MD simulations. Similarly to PCA, tICA is a dimensionality reduction technique, which permits us to reconstruct the FEL of biomolecules, as a function of different CVs (v_k). The difference with PCA, is that in this case, the largest time-autocorrelation is the basis that serves to identify the slowest independent CVs, and that will be projected into the new coordinate space:

$$\frac{\langle y_k(t) y_k(t+\tau) \rangle_t}{\langle y_k(t)^2 \rangle_t} \qquad (2.\ 23)$$

where τ is the chosen lag time. In tICA, the time-lagged covariance matrix is computed:

$$C(\tau) = \left(\langle x_i(t) x_j(t+\tau)_t \rangle \right)_{ij} \in R^{dxd} \quad (2.\ 24)$$

Considering this, v_k is maximized under the constraint that is orthogonal to all previous degrees of freedom. The constraint applied in this step is crucial for the identification of the slowest modes of our reaction coordinate, and is mainly where it differs from PCA.

$$\frac{v_k^T C(\tau) v_k}{v_k^T C(0) v_k} \quad (2.\ 25)$$

Thus, v_k is the solution to the eigensystem, representing the value of the new eigenvectors, *i.e.* dimensions, of the new coordinate system:

$$C(\tau) v_k = \lambda_k C(0) v_k \quad (2.\ 26)$$

where λ_k represents the independent components, C(0) is the mean-free covariance at a $\tau = 0$.

2.4. Enhanced Sampling methods

When reproducing the time-dependent thermodynamic properties of a given system by means of MD simulations, one limitation that can be encountered is the ergodic principle. This principle states that if the system evolves in time indefinitely, all the accessible microstates will be equally visited, and the time-averaged conformational sampling is equal to the probability distribution of the microstates in thermodynamic equilibrium.[37] Notwithstanding, the timescale required for overcoming the energy barriers that separate those conformations hampers the complete exploration of the conformational space. Depending on the height of the energy barrier, the simulation time required for sampling a conformational transition

can range from μs up to s. This range in timescale will depend on the nature of the motion that we are studying (see Figure 1.2). Although some specialized computers for performing long timescale MD simulations exist, such as ANTON supercomputer, which allow the routine simulation of MDs of biological systems in the range of the ms without applying any bias potential, the exploration of all possible conformational states of a given system is hampered by the timescale required for exploring regions outside the basins. For that reason, different computational methods can be used for the enhancing sampling of the conformational space of biomolecules, as well as to capture their thermodynamic and kinetic properties, by facilitating the exploration of the different conformational states.

2.4.1. Markov State Models: unbiased estimation of thermodynamics and kinetic rates of conformational transitions

When elucidating the conformational dynamics of a molecular system, it is often important to quantify not only the thermodynamics (FEL), but also the kinetics of the conformational transitions (timescales), as they might be relevant for understanding protein function. In this regard, Markov State Models (MSM) have been widely used to computationally reproduce the thermodynamic and kinetic traits on long timescales using sets of short MD simulations.[42–44] Thus, to completely characterize the conformational space is essential to have a robust statistical model of the kinetics of conformational transitions.

As it is a requirement to have a complete exploration of the FEL, one of the limiting steps when computing a MSM is the sampling strategy, as it will be a key factor in

the statistical analysis. Typically, adaptive sampling is the chosen strategy for studying conformational changes for which one has previous knowledge of the initial and final states. In this strategy, iterative rounds of short MD simulations and reconstruction of the FEL is performed (using the dimensionality reduction techniques explained in section 2.3). Once the FEL is reconstructed, conformations from hardly explored regions are extracted for further simulations until the whole FEL is properly sampled, trying to equally sample all the conformational states, referred to as meta-stable states.

The MSM is basically a *n* X *n* square transition probability matrix in which *n* represents the number of meta-stable states. It is dependent on the definition of a lag time *t*, whose definition is crucial for obtaining higher fidelity temporal resolution. Such lag time defines the separation of time points for which a specific state appears. A requirement for the lag time is that it has to be "memoryless", meaning that the probability of visiting state *x* does not depend on the initial visited state. The method is based on the typical eigenvalue problem:

$$\pi^T P = \pi^T \tag{2. 27}$$

A transition matrix P gives a stationary distribution π. Such matrix contains transition probabilities lower than the lag time, usually orders of magnitude shorter than relaxation timescales. The number of eigenvectors of the latter expression (equation 1.28), correspond to *n* elements, which are represented by the number of observed meta-stable states. By choosing the appropriate number of meta-stable states, the eigenvalues that correspond to transition probabilities, as well as the *populations* of each state and the *eigenfluxes* (*i.e.* transition timescales), can be estimated.

MSMs generated from MD simulations of biomolecular systems usually present many microstates, hampering the connection with experimental data. To overcome this problem, a coarse-grained kinetic model using a hidden Markov State Model (hMSM) can be used. hMSM are probabilistic models useful to identify unobservable states in complex simulation systems, based on data from different observable attributes. In other words, the hidden states are not directly visible, but defined by the probability distribution of the observations (*i.e.* MD features). In biomolecular MD simulations, the hidden states correspond to metastable states in the conformational energy landscape, where the probability to visit other states only depends on the current state (Markov property). Because transitions within a metastable state are rapid, while longer time scales are required to visit other energy minima, a kinetic partition can be used to define the hidden metastable states. The resulting hMSM provides both macrostate-to-macrostate kinetic transitions and the relative populations of each macrostate.

2.4.2. Metadynamics: enhanced bias conformational sampling

One of the most used methods for enhancing the conformational sampling is Metadynamics simulations. This is a bias method that consists in adding external energy potentials to a selected set of degrees of freedom, *i.e.* collective variables (CVs), at a regular number of MD steps.[46] The mathematical expression of the external potential (V_G) acting on a system at a time t is the following:

$$V_G(S,t)=\int_0^t dt' \omega \exp\left(-\sum_{i=1}^{d}\frac{\left(S_i(R)-S_i(R(t''))\right)^2}{2\sigma_i^2}\right) \qquad (2.\ 28)$$

where S(R) is the CV value as a function of the coordinates, S(R(t')) is the value of the CV at a certain time t, ω is an energy rate, and σ_i is the width of the Gaussian

for the *i*th CV. This bias potential gradually overcomes energy barriers allowing for efficient exploration of different conformational states. After a certain simulation time, the biasing potential corresponds to the negative values of the free energy surface (FES) and, therefore, all possible states are equally sampled (Figure 2.1).

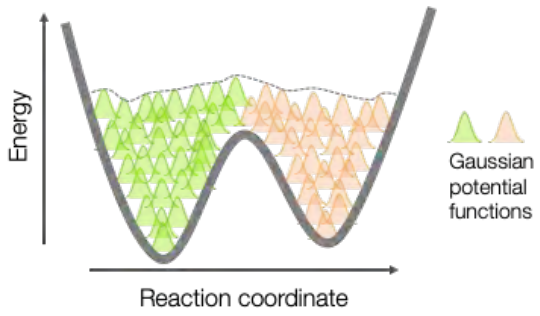


Figure 2.1. *Representation of the repulsive gaussian potential added to the FES for enhanced sampling*

The main challenge of metadynamics is the selection of the proper CVs for its effective application. Such CVs must distinguish the initial and the final state of the conformational transition that we want to describe. In this context, we are applying a dimensionality reduction, similarly to applying PCA or tICA. However, in metadynamics, the selection must fulfill stricter conditions, the CVs might describe properly the reaction coordinate. Moreover, as Metadynamics is an enhanced sampling technique, the CVs must describe the slowests modes of the system so that the bias potential can converge in a reasonable amount of simulation time.

In this thesis, the well-tempered version of the metadynamics algorithm was used to improve the convergence of the FES reconstruction. In well-tempered metadynamics, a different expression of the bias potential is used.[77] In this method, the bias potential deposition is decreasing over time. This is performed by a rescaling the Gaussian height according to the following expression:

$$V(S,t) = k_B \Delta T \ln\left(1 + \frac{\omega N(S,t)}{k_B \Delta T}\right) \quad (2.\ 29)$$

Considering that the value of the microscopic observables is getting closer to the equilibrium as the simulation proceeds, it is reasonable to compute the Gaussian height as a function of time at N simulation steps. With this method, the convergence of the FES is improved, overcoming the typical problems of conventional metadynamics simulations.

2.4.3. accelerated Molecular Dynamics: enhancing the conformational sampling

The above-mentioned methods are useful when one has some prior information about the studied system. However, without having a deeper understanding of the possible protein motions, the definition of collective variables for the construction of an MSM, or for performing a metadynamics simulation, might be a difficult task. A way to overcome this limitation is the use of accelerated Molecular Dynamics (aMD). aMD is a powerful method to explore the conformational space of biomolecules without previous knowledge of the possible conformational transitions by adding a non-negative boost potential.[45] It consists in enhancing the conformational sampling of biomolecules. To do so, a reference energy is defined, and to all the conformations located below such threshold a non-negative boost potential is applied:

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E, \\ V(r) + \Delta V(r), & V(r) < E \end{cases} \quad (2.\ 30)$$

where $V(r)$ is the original potential, E is the reference energy, and $V*(r)$ is the modified potential. In the simplest form, the boost potential, $\Delta V(r)$ is given by:

$$\Delta V(r) = \frac{(E-V(r))^2}{\alpha+E-V(r)} \qquad (2.\ 31)$$

where α is the acceleration factor. By modifying the acceleration α, one can tune the flatness of the energy surface. By increasing α, the modified surface becomes flatter, facilitating the transitions between low-energy conformational states. On the contrary, a low α parameter will result in a modified energy surface similar to the original. (Figure 2.2).

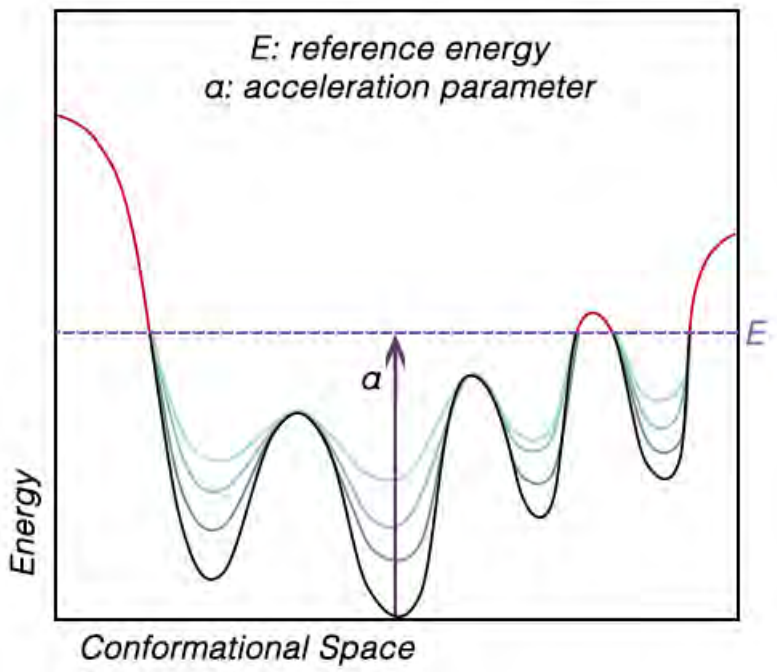


Figure 2.2. Representation of how the conformational landscape is flattened based on the definition of the reference energy (E), and the acceleration parameter (α)

Moreover, the sampling of conformational transitions can be enhanced even more, by applying a dual-boost aMD approach. Here, a total boost potential is applied to all atoms in the system in addition to a more aggressive dihedral boost, *i.e.*, (E_{dihed}, α_{dihed}; E_{total}, α_{total}), within the dual-boost aMD approach.[78] The acceleration parameters used in this thesis, are the following:

$$E_{dihed} = V_{dihed\_avg} + 3.5 \times N_{res},\ \alpha_{dihed} = 3.5 \times \frac{N_{res}}{5} \quad (2.\ 32)$$

$$E_{total} = V_{total\_avg} + 0.175 \times N_{atoms},\ \alpha_{total} = 0.175 \times N_{atoms} \quad (2.\ 33)$$

where N_{res} is the number of protein residues, N_{atoms} is the total number of atoms, and V_{dihed_avg} and V_{total_avg} are the average dihedral and total potential energies calculated from conventional MD (cMD) simulations.

2.5. Shortest Path Map

One of the limitations of computational enzyme design is to identify hotspots suitable for mutagenesis that result in improved displayed catalytic activities. Commonly, strategies involving the modification of the active site, or substrate tunnel regions have been applied. However, as it was depicted in chapter 1.4, distal position can have an impact in the global dynamics of a given enzyme. To computationally identify those positions, correlation-based methods allow the identification of dynamic protein regions which are tightly interconnected at long rage. Among these methods, Shortest Path Map (SPM), developed at our group, can be used to identify positions from the overall structure, which are highly correlated. The first step of the Shortest Path Map (SPM) calculation relies on the construction of a graph based on the computed mean distances and correlation

values observed along the MD simulations. To do so, the following equation is applied:

$$C_{ij} = \frac{\langle \Delta r_i \, \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \, \langle \Delta r_j^2 \rangle}} \qquad (2.\ 34)$$

where Δr_i and Δr_j represent the displacement of the Ca of the *i* and *j* residues with respect to their reference positions. For each residue of the protein a node is created and centered on the C-alpha if both residues display a mean distance of less than 6 Å along the simulation time. The length of the line connecting both residues is drawn according their correlation value:

$$d_{ij} = -\log|C_{ij}| \qquad (2.\ 35)$$

Larger correlation values (closer to 1 or -1) will have shorter edge distances, whereas less correlated residue pairs (values closer to 0) will have edges with long distances. At this point, we make use of Dijkstra algorithm to identify the shortest path lengths. The algorithm goes through all nodes of the graph and identifies which is the shortest path to go from the first until the last protein residue. The method therefore identifies which are the edges of the graph that are shorter, *i.e.* more correlated, and that are more frequently used for going through all residues of the protein, *i.e.* they are more central for the communication pathway. Altering those positions, result in a drastically different dynamical behavior of the resulting variants. However, apart from its use in enzyme design, it can be also employed for elucidating differences in conformational dynamics in an evolutionary path, rationalizing the effects of the included mutations in the enzyme sequence, and their impact in the displayed activities (See Chapter 5.1).

Chapter III. Objectives

The main goal of this thesis is to establish a direct relationship between conformational dynamics and catalytic activity in Monoamine Oxidase enzyme used for the biocatalytic production of chiral amines. For that purpose, we selected different Monoamine Oxidase from *Aspergillus niger* variants that were generated from Directed Evolution experiments. To explore the role of conformational dynamics in MAO-N catalytic activity, we combine a set of computational methods including molecular dynamics simulations, enhanced sampling techniques, correlation-based tools, and tunnel analysis. Direct comparison of the computational outcome with experimental kinetic data will help to understand the role of mutations on the catalytic efficiency.

The first objective is to characterize the free-energy landscape (FEL) of the laboratory evolved MAO-N variants to identify how changes on the conformational dynamics can impact the catalytic activity. Despite MAO-N has been the target of extensive enzyme engineering, its structural and dynamical features are still unexplored. To that aim, molecular dynamics simulations and Markov State Models will be combined to reconstruct the FEL and retrieve the populations of the most relevant conformational states and the kinetics of the most important conformational transitions. Comparison of the computed conformational state populations and timescales of conformational transitions with the experimental kinetic data from literature, will be used to understand the mutations-conformational dynamics-catalytic activity relationship. Moreover, we aim to determine how the different conformations of MAO-N affect the substrate binding process. In this direction, tunnel analysis, together with accelerated molecular dynamics simulations will be used to reconstruct the substrate binding process.

The second objective consists in elucidating the potential interplay between MAO-N subunits and its relation with the catalytic efficiency. MAO-N is a dimeric enzyme and whether the inter-monomer communication affects catalysis remains

unknown. Understanding how the two subunits are interconnected is relevant for MAO-N engineering efforts. For this purpose, the shortest-path map tool will be applied to determine the networks of correlated residues between MAO-N subunits. Finally, different substrate bound simulations will be performed, considering different conformations in the two MAO-N subunits to elucidate which conformations are key for productive binding and catalysis.

Chapter IV. Elucidating the conformational heterogeneity of MAO-N and its role in catalysis

4.1 Precedents in MAO-N evolution: state-of-the-art

Monoamine Oxidase from *Aspergillus niger* (MAO-N) is a homodimeric flavoenzyme, *i.e.* Flavin Adenine Dinucleotide (FAD) dependent, responsible for the oxidation of amines into the corresponding imine by means of cofactor reduction. The catalytic reaction takes place inside a hydrophobic cavity composed by the aromatic side chains of residues Trp430 and Phe466, which are responsible for stabilizing the amine substrate close to the isoalloxazine ring of the FAD. Once the substrate is stabilized in the aromatic cage, a hydride transfer reaction occurs from the substrate molecule to the FAD in an enantioselective manner. During this process, the cofactor gets reduced, and needs to be oxidized again to keep catalyzing the conversion of amines into imines. For such regeneration, MAO-N uses molecular oxygen to oxidize the FAD, a process that releases H_2O_2. O_2 binding process takes place thanks to a specific tunnel for its diffusion to the active site.

The interest for this enzyme mainly arises from its catalytic activity towards different amine substrates. In 2003, Turner and co-workers developed a strategy for the deracemisation of chiral amine mixtures, utilizing MAO-N in combination with a non-selective chemical reductant, such as sodium borohydride. This fact broadened the application of MAO-N in the pharmaceutical and agrochemical industries as a biocatalyst for the obtention of chiral amine building blocks. Notwithstanding, the wild-type (WT) enzyme was only active towards primary and some secondary amines, being the obtention of pure tertiary amines still hampered. In this regard, many efforts have been performed to achieve a smart library of MAO-N variants that improves the catalytic activity of MAO-N as well as to widen its substrate scope towards bulkier amines.

Within this context, many Directed Evolution (DE) experiments were performed resulting in different MAO-N variants that showed different substrate scope and catalytic activities for imine production. From these studies, different MAO-N variants emerged, highlighting MAO-N D3 which includes mutations Asn336Ser/ Met348Lys/ Ile246Met, MAO-N D5 (Asn336Ser/ Met348Lys/ Thr384Asn/ Asp385Ser/ Ile246Met), and MAO-N D9 (Asn336Ser/ Met348Lys/ Thr384Asn/ Asp385Ser/ Ile246Thr/ Phe210Leu/ Leu213Thr/Met242Gln/Trp430His). These variants improved the k_{cat} for different amines, being D3 able to catalyze the oxidation of new aromatic amines, D5 oxidizing aromatic and bicyclic systems, and the most evolved variant D9, which was able to accept polycyclic bulky amines. It is worth highlighting that only three crystal structures are available for MAO-N enzyme. MAO-N has only been crystallized for D3, D5, and D11 variants. MAO-N D5 was crystallized in the presence of a non-covalently bound Flavin Adenine Dinucleotide (FAD) cofactor, a proline, and ethylene glycol molecules in the active site. However, MAO-N wild-type (WT) enzyme X-ray structure has not yet been solved.

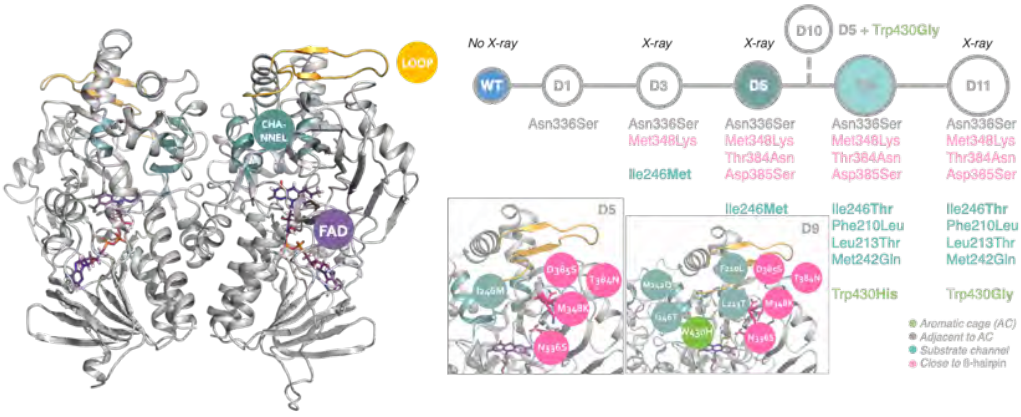


Figure 4. 1. MAO-N D5 X-ray structure (PDB Code: 2VVM) highlighting the important regions (left). MAO-N evolutionary path developed by prof. Turner Lab (right), including the mutations introduced in each round of DE.

Interestingly, the mutations introduced via directed evolution were located mostly at distal , in particular near the substrate entrance tunnel which is at the interface of the two monomers, while a few were found close to the active site, and only one in the active site itself (Figure 4.1). Nevertheless, the effect of the mutations in the conformational dynamics of MAO-N was yet to be deciphered, as well as its relationship with the different displayed catalytic activities.

In this chapter, we used conventional Molecular Dynamics-based techniques to understand the role of the introduced mutations in the laboratory-evolved variants (i.e. MAO-N WT, D3, D5, and D9), to rationalize their effect in the conformational dynamics of MAO-N as well as its impact into the catalytic activity. More specifically, we used Molecular Dynamics simulations for the initial exploration of the conformational space, as well as enhanced sampling techniques to fully characterize the thermodynamics and kinetics of the conformational transitions, i.e. accelerated Molecular Dynamics (aMD), metadynamics, and Markov State Models. Additionally, tunnel analyses by means of CAVER and substrate binding simulations were performed to explain the role of the different conformations in terms of catalytic activity.[79]

4.2 Computational details

4.2.1. System preparation

Molecular Dynamics simulations (MD) have been used to explore the conformational heterogeneity of MAO-N WT and D variants. We have selected the D5 structure (Protein Data Bank (PDB) accession number 2VVM) as a starting point to run our simulations.[80] The proline and ethylene molecules are removed from the active site for performing the *apo* and substrate binding MD simulations. In this crystal structure, the loops corresponding to residues 32-37 and 32-39 of monomers A and B respectively

were not solved. For running the MD simulations, we modelled the missing loops using the SwissModel homology model workspace (https://swissmodel.expasy.org). Amino acid protonation states were predicted using the H++ server (http://biophysics.cs.vt.edu/H++). For simulating the WT enzyme, we used the mutagenesis tool from PyMOL software using the D5 crystal structure (PDB 2VVM) as starting point. The following mutations found in D5 were reverted to obtain the WT sequence: Ile246Met, Asn336Ser, Met348Lys, Thr384Asn, Asp385Ser. From these coordinates, we started unrestrained conventional MD simulations.

4.2.2. Molecular Dynamics simulation

Each system was solvated in a pre- equilibrated truncated cubic box of TIP3P water molecules with an internal offset distance of 10Å , using the AMBER16 LEAP module. All systems were neutralized with explicit counterions (Na^+ or Cl^-). All calculations were done using the ff99SBildn Amber force field and TIP3P water model. A two-stage geometry optimization approach was performed. First, a short minimization of the water molecules positions, with positional restraints on solute by a harmonic potential with a force constant of 500 kcal mol-1 Å-2 was done. The second stage was an unrestrained minimization of all the atoms in the simulation cell. Then, the systems were gently heated using six 50 ps steps, incrementing the temperature 50 K each step (0-300 K) under constant-volume, periodic-boundary conditions and the particle-mesh Ewald approach to introduce long-range electrostatic effects. For these steps, an 8 Å cutoff was applied to Lennard-Jones and electrostatic interactions. Bonds involving hydrogen were constrained with the SHAKE algorithm. Harmonic restraints of 10 kcal mol-1 were applied to the solute, and the Langevin equilibration scheme is used to control and equalize the temperature. The time step was kept at 2 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs timestep at a constant pressure of 1 atm. After

the systems were equilibrated in the NPT ensemble, MD simulations were performed under the NVT ensemble and periodic-boundary conditions using our GPU cluster Galatea (composed of 178 GTX1080 GPUs).

Substrate parameterization. The parameters for the hexylamine and alpha-methyl benzylamine (AMBA) substrates for the MD simulations were generated within the ANTECHAMBER module of AMBER16 using the general AMBER force field (GAFF), with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model. The charges were calculated according to the Merz-Singh-Kollman scheme using Gaussian 09. Parameters of FAD cofactor were extracted from a database from Manchester University.

4.2.3. Markov State Models

First, cartesian coordinates obtained from 300 μs accumulated MD data were projected into a selected feature space, capturing the conformational change of the flexible β-hairpin for MAO-N WT, D3, D5, and D9 variants. After testing several descriptors, a feature set containing combinations of the C-alpha distances between β-hairpin loop residues 128-131, and the important β-sheet residues 353-356 was used. The angle formed by the C-alpha atoms of residues 134 (monomer B), 240 (monomer A), and 360 (monomer B) was also included. According to this selection, each MD snapshot can be represented by a total 17 collective variables. Afterwards, the time-lagged independent component analysis (tICA) was used to project our featurization choice into a kinetically relevant space, common to all studied systems.

The lag time selection of 100 ns was based on the description of 95 % cumulative kinetic variance, including the smallest possible number of TICA dimensions. Then, k-means clustering was performed to group MD snapshots into microstates. MSMs were constructed at a lag time of 40 ns for WT and 2 ns for D5 based on the convergence of the calculated implied relaxation time scales as a function of the lag

time. The Chapman-Kolmogorov test was performed on the MSM to check the self-consistency of the model.

MSMs usually present many microstates, hampering the connection with experimental data. To overcome this problem, a coarse-grained kinetic model using a hidden Markov State Model (hMSM) was also estimated. hMSM are probabilistic models useful to identify unobservable states in complex simulation systems, based on data from different observable attributes. In other words, the hidden states are not directly visible, but defined by the probability distribution of the observations (*i. e.* MD features). The hMSM was constructed by assigning discrete trajectories of the MSM model above into three main conformational macrostates (closed, partially closed, and open). The resulting hMSM provides both macrostate-to-macrostate kinetic transitions and the relative populations of each macrostate. Convergence of hMSM implied relaxation time scales (Figure S5) and relative macrostate populations as a function of lag time was used to further validate the hMSM lag time selection. All analyses were performed with PyEMMA 2.4.

4.2.4. aMD simulations

Conformational dynamics. Accelerated Molecular Dynamics simulations (aMD) were used to initially explore the conformational dynamics of MAO-N WT and D5 variants. Starting from WT and D5 closed structures, we performed unrestrained conventional MD simulations (100 ns) as described above from which the acceleration parameters were determined. Then, five replicas of 1.3 μs of dual-boost accelerated Molecular Dynamics (aMD) simulations were carried out. Here a total boost potential is applied to all atoms in the system in addition to a more aggressive dihedral boost, *i.e.*, (E_{dihed}, α_{dihed}; E_{total}, α_{total}), within the dual-boost aMD approach. The acceleration parameters used in this work, are calculated using the following expressions:

$$E\text{dihed} = V\text{dihed\_avg} + 3.5 \times N\text{res}, \alpha\text{dihed} = 3.5 \times N\text{res}/5;$$

$$E\text{total} = V\text{total\_avg} + 0.175 \times N\text{atoms}, \alpha\text{total} = 0.175 \times N\text{atoms}$$

where N_{res} is the number of protein residues, N_{atoms} is the total number of atoms, and V_{dihed_avg} and V_{total_avg} are the average dihedral and total potential energies calculated from 100 ns cMD simulations, respectively.

Substrate Binding Simulations. Accelerated Molecular Dynamics simulations (aMD) have been used to study the spontaneous binding of hexylamine and AMBA substrates in the MAO-N active site of both WT and D5 variants. In both cases, *i.e.* WT and D5, we placed four substrates, (four hexylamines for WT and four AMBA for D5) in arbitrary positions within the solvent region (more than 30 Å far from the FAD active site). From these coordinates, we started unrestrained conventional MD simulations (100 ns) from which the acceleration parameters were determined. Then, ten replicas of 400 ns of dual-boost accelerated Molecular Dynamics (aMD) simulations were performed to allow the substrate to diffuse freely until it spontaneously associates with the surface of the protein, and finally targets the active site. The simulations that substrate binding was observed were extended to 600 ns. These long timescale unconstrained aMD simulations were performed with the aim of capturing a number of spontaneous substrate binding events. The same acceleration parameters were used as in the case of conformational dynamics sampling.

4.2.5. Metadynamics Calculations.

The PLUMED2 software package together with the GROMACS 5.1.2 code (M.J. Abraham et al. GROMACS User Manual version 5.1.2) were used to carry out the metadynamics simulations. In order to describe the opening and closing of the β-hairpin loop two CVs were selected. The open-to-closed transition was described as a difference of distances *dCV = d1 – d2*, where *d1* stands for the distance between the β- hairpin C*alpha* carbons center of mass (residues 605 to 630) and the adjacent

β-sheet C*alpha* carbons center of mass (residues 835 to 855). On the contrary, *d2* corresponds to the distance of the above-mentioned β-hairpin center of mass and the MAO-N protein region formed by residues 140 to 190, also expressed as the C*alpha* center of mass. The second CV accounts for β-hairpin loop flexibility and corresponds to the sum of the distance between all torsional backbone dihedrals to a reference value.

Metadynamics simulations were started from the equilibrated MAO-N WT, D5, and D9 structures obtained from previous MD simulations. Initial Gaussian potentials of height 0.35 kcal/mol, deposited every 1 ps of MD simulation, were gradually decreased on the basis of the well-tempered adaptive bias of 30. The multiple-walker extension of the approach, which uses several replicas of the same system biasing identical CVs, was used to increase the sampling of the conformational space and to increase the convergence of individual free-energy profiles. After an initial metadynamics run, ten snapshots were extracted for each system, covering approximately all the conformational space. Then, multiple-walkers metadynamics simulations with 10 replicas were computed. Each replica was run for 30 ns, giving a total of 300 ns of simulation time per system.

4.2.6. Tunnels Analyses

The CAVER 3.0 software was used to analyze the available pathways for substrate entrance to the MAO-N WT, D3, D5, and D9 variants active site.[79] *ca.* 2000 snapshots for each identified conformation was analyzed in each system (WT, D3, D5, and D9) For this study, a spherical probe of 0.9 Å radius was selected with a weighting coefficient of 1, and clustering threshold of 6. The shell depth was set to 5 and the used shell radius was 3. The starting point for the calculation was chosen at the center of mass of Phe466 and Trp430 aromatic cage residues within each MAO-N variant.

4.3 Results and discussion

4.3.1. Initial exploration of MAO-N dynamics by means of MD simulations

Considering that there was no information about the dynamic behavior of MAO-N as well as the extensive experimental data that could be found in literature about the activity of D5 variant, a first exploration by means of MD simulations was performed on this variant, using as starting point the X-ray structure (PDB code: 2VVM). More specifically, 3 replicas of 1μs were performed. Their analysis served us to detect a significant conformational change on the highlighted β-hairpin. In such motion, we could first observe the disruption of backbone interactions of the β-hairpin of one subunit with the rest of the protein, followed by a loss of the secondary structure, and finally the collapse of the resulting disordered loop over the other monomer (see Figure 4.2). Considering the nature of this motion, we refer to those conformations as Closed (C, X-ray-like conformation), and Open (O, when it collapses over the other subunit) conformations. Notwithstanding, this motion seems to occur in long timescales, as its reproducibility using multiple replicas of MD was not observed.

At this point, we were encouraged to identify this motion for WT, D3, and D9 variants, aiming to relate the conformational dynamics with the differences in activity of all the laboratory evolved variants in terms of a population shift towards one specific conformation. Initial aMD simulations of these variants were performed for such aim, resulting in a frequent exploration of O conformations for the WT variant, that could be visited in several aMD replicas and took place in a few nanoseconds. However, these initial simulations suggested that when more mutations are introduced to the enzyme sequence, the more difficult it is for the β-hairpin to adopt the O conformation.

For that reason, we aimed to quantify the thermodynamics and kinetics of these conformational transitions, expecting to find a correlation between β-hairpin opening/closure and catalytic activities. As mentioned before, we relied on metadynamics simulations and MSM (see next sections).

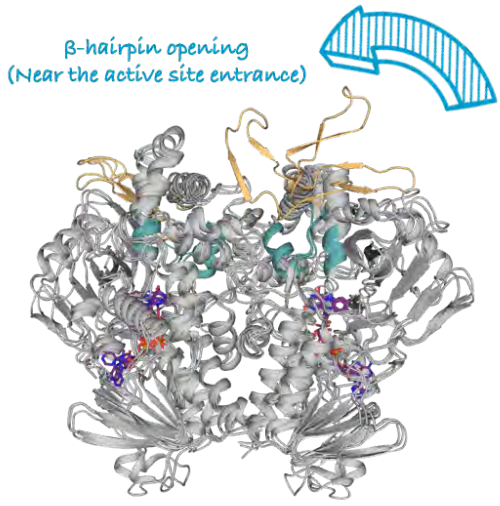


Figure 4. 2. Overlay of the representative MD snapshots displaying the β-hairpin opening motion

4.3.2. Reconstructing the Free Energy Landscape of MAO-N variants

Firstly, an initial exploration of the Free Energy Landscape was performed by means of metadynamics simulations. Results from these simulations are summarized in Figure 4.3. As we expected, there was a progressive stabilization of the C conformations, increasing the energy barriers that need to be overcome for transitioning to O states. More specifically, in the case of the WT enzyme, an energy barrier of 4 kcal/mol separates C and O conformations. This energy barrier is increased for the case of MAO-N D5, in which a 11 kcal/mol barrier can be observed. In the most evolved variant, *i.e.* MAO-N D9, the introduced mutations produce a better

stabilization of the Closed state, displaying an energy barrier of 17 kcal/mol. It is worth highlighting the fact that not only the energy barriers are different in each enzyme variant, but also the relative stability of the different O conformations. Whereas in WT the relative stability of C and O conformations is similar, in D5 and D9 O conformations have been destabilized by *ca.* 15 kcal, producing as a consequence a population shift towards C conformations in the laboratory-evolved variants. These preliminary results suggest some sort of relationship between the capability of the different MAO-N variants to explore O conformations of the β-hairpin and the displayed catalytic activities, being able to relate the stabilization of C conformations with increased k_{cat}. Notwithstanding, with these metadynamics simulations the system is forced to explore the whole range of the described collective variables. Such constraints, and the resulting FELs, are very dependent on the proper definition of the CVs. For that reason, we aimed to apply unbiased methods to fully explore the FEL and to determine the kinetic properties of the conformational transitions in the different MAO-N variants.

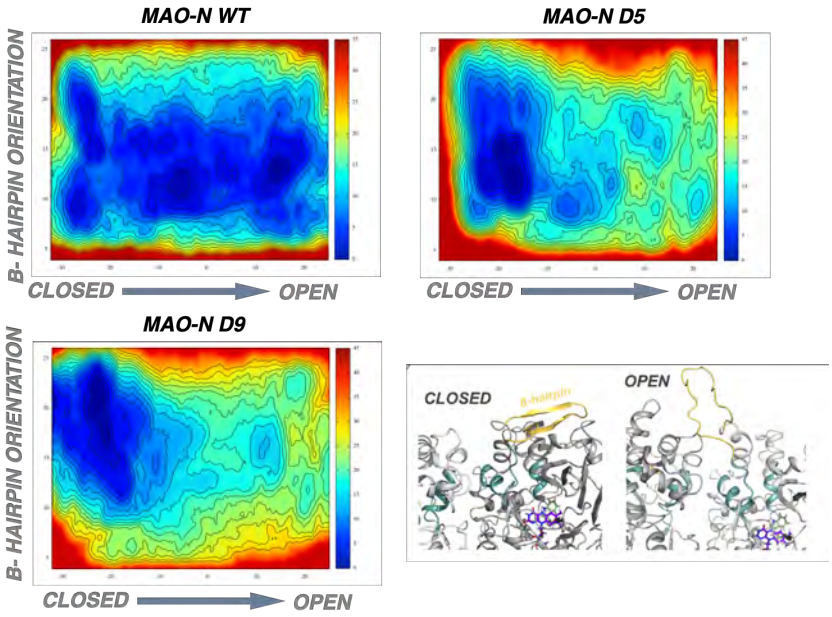


Figure 4.3. Free energy surface of the closed-to-open transition in MAO-N WT, D5, and D9 variants. Isolines are represented every 2 kcal/mol. Representative structures of the MAO-N WT open and closed β-hairpin conformations extracted from the metadynamics simulations are also depicted. Active site residues and FAD molecule are represented with green and purple sticks, respectively. The β-hairpin region is colored in yellow.

4.3.2. Markov State Model reveals a population shift along MAO-N evolution

In this context, hidden Markov State Models (hMSM), in combination with cMD were employed, as they permit to reconstruct the FELs, but also to compute the timescale of the different conformational transitions that the enzyme can explore. To this aim, a dimensionality reduction technique was required, to give to our huge data a physical meaning in terms of enzyme conformational dynamics. In Figure 4.4 a summary of the obtained results can be observed.

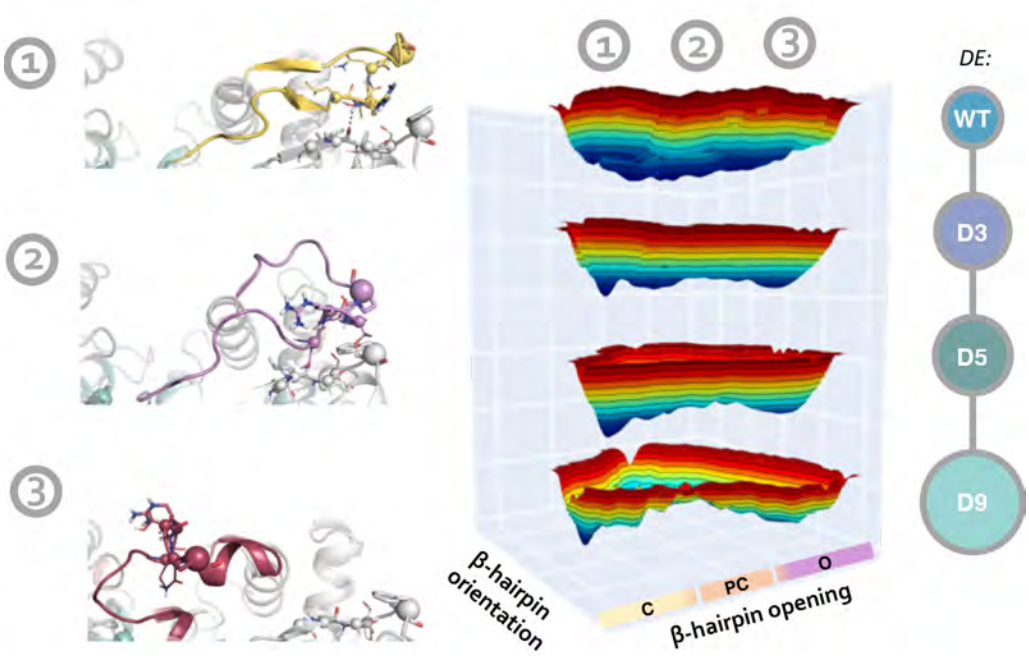


Figure 4.4. Reconstructed FELs for WT, D3, D5, and D9 variants. The color of the 3D FEL represents the stability of each conformation, being blue the most stable, and red the least stable conformations The main identified conformations are represented: Closed (1), partially closed (2), and open (3)

In these FELs, the main conformations that the β-hairpin can adopt have been elucidated for each variant. In all the cases, three main conformational states could be identified. First, the well-known C conformation, which resembles the X-ray structure and in which the β-hairpin can be found in a stable form by means of backbone interactions with the rest of the protein. Secondly, a PC conformation, where the β-hairpin starts to lose the secondary structure, keeping some backbone interactions, but without the characteristic β-hairpin shape. Finally, the O conformation, in which all the interactions with the protein backbone have been disrupted, thus making the resulting loop collapse over the other MAO-N subunit.

From here, we could elucidate the differences in relative stabilities of these conformations along the evolutionary path (Figure 4.4). As depicted for metadynamics simulations, we could observe a progressive stabilization of C conformations, and an increase in the energy barriers to evolve towards O conformations. On the contrary, the stability of O conformations is decreased along this path, being more populated in the case of MAO-N WT. More specifically, we could determine populations of C conformations of 18% in WT, 21% in D3, 32% in D5, and 45% in D9, whereas for O conformations the populations were significantly decreased, from 40% in WT, 29% in D3, 16% in D5, and 10% in D9. Moreover, the overall timescale required to evolve from C to O conformations was substantially increased, from 3.6 μs in WT, 5.1 μs in D3, 7.9 μs in D5, and 9.7 μs in D9, which points the how the conformational dynamics of MAO-N have been altered along the evolutionary path. It is worth highlighting the fact that PC conformations are highly populated in all the systems, which may suggest that it has a catalytic function for amine oxidation. In this context, populations of 42% were identified in WT, 50% in D3, 52% in D5, and 45% in D9. It is worth mentioning that the values from D3 and D9 are not definitive, as more sampling of the conformational spaces is being performed until full convergence of the method (See Kolmogorov tests in Annex A). However, the tendency of kinetic and thermodynamic stabilization of Closed and Partially Closed conformations can be observed in the FELs

from Figure 4.4. From these results, and having identified the different conformations that these variants can explore, as well as the timescales of the transitions between them, one big question is yet to be answered: Which impact these FEL alterations have in the catalytic activity of MAO-N?

To shed some light on the effects that these conformations have in the catalytic process, we first analyzed how the enzyme tunnels were altered along the evolutionary path. Such analyses were performed by means of CAVER software, and the outcome that we obtained is summarized in Table 4.1.

Table 4.1. Summary of the results obtained from the tunnel analyses for each variant and for each MAO-N conformation. The bottleneck radius (BR) is represented in Å. The population of the tunnel in each cluster is also represented as a percentage.

		Tunnel 1		Tunnel 2	
Conformation	Variant	BR (Å)	Population (%)	BR (Å)	Population (%)
C	WT	1,11	40	1.01	22
C	D3	1.14	52	1.08	24
C	D5	1.19	70	1.12	26
C	D9	1.48	89	1.05	13
PC	WT	1.37	30	1.38	19
PC	D3	1.28	73	0.95	15
PC	D5	1.22	20	1.13	36
PC	D9	1.39	95	n.d.	n.d
O	WT	1.19	38	1.23	20
O	D3	0.98	13	1.15	75
O	D5	1.06	24	1.10	40
O	D9	1.13	15	1.14	96

From these analyses, two main tunnels could be identified. Tunnel 1, located at the hinge region of the β-hairpin, *i.e.* at the interface between subunits, which was reported in literature as the substrate tunnel. A second tunnel could be detected when the enzyme displays PC, or O conformations (Figure 4.5). This tunnel is created when the secondary structure of the β-hairpin is lost, *i.e.* PC or O conformations. We could observe that, along the evolutionary path, not only closed conformations of the β-hairpin were stabilized. The introduced mutations in D3, D5, and D9, were progressively stabilizing the formation of substrate tunnel 1, that, despite not being significantly wider, the percentage of snapshots in which this tunnel is identified has increased. Moreover, the formation of tunnel 2 has been favored by the introduction of those mutations, which could suggest that an easier substrate binding event can take place in both conformations.

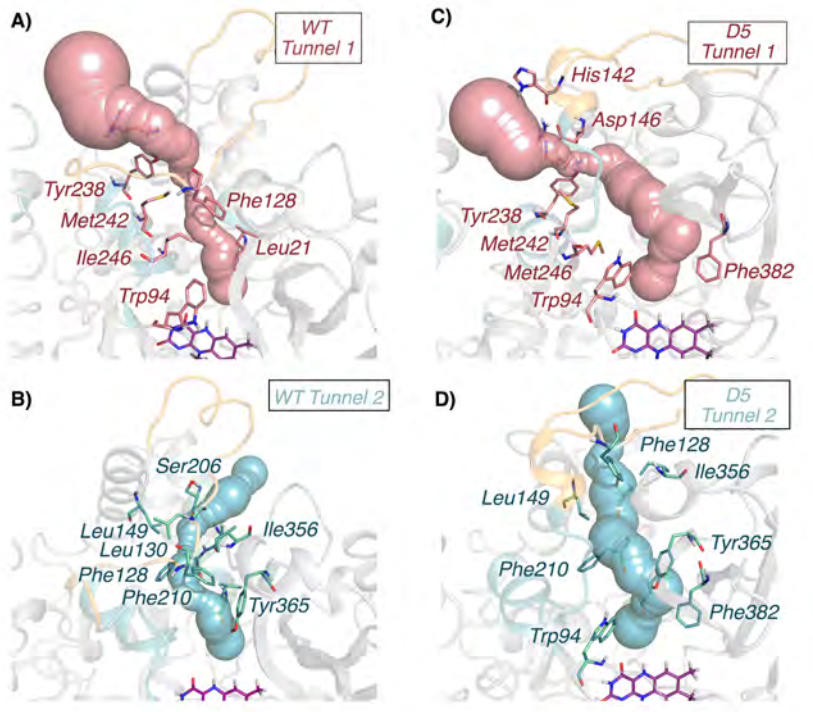


Figure 4.5. Representation of the two main substrate access tunnels for MAO-N WT (Tunnel 1 A and Tunnel 2 B) and D5 (Tunnel 1 C and Tunnel 1 D). Substrate access tunnel 1, which passes through the hinge region of the β-hairpin region, is colored in salmon, whereas substrate tunnel 2 is colored in blue. Tunnel delimiting residues present in more than 80 % of the analyzed frames are also shown.

4.3.4. Elucidating the molecular basis of substrate binding process in MAO-N by means of aMD simulations

To decipher the role of the conformational dynamics, as well as the potential catalytic role of the new identified tunnel in the open conformations, spontaneous substrate binding aMD simulations starting from closed and hidden open conformations were performed for WT and D5 variants in the presence of substrate. To do so, the substrate used in each case was chosen based on their relative activities found in literature, selecting those for which the catalytic activity was higher. For this reason, we selected Hexylamine (HEX) for simulating the WT enzyme, and alpha-methyl benzylamine (AMBA) for D5. In both cases, it was a requisite that the β-hairpin of one monomer is maintained in a C conformation to successfully bind the corresponding substrate. Notwithstanding, binding was not observed when the β-hairpin of both monomers is found in a C conformation, suggesting that adopting PC or O conformations in one subunit is required for activity. However, this communication between subunits will be addressed in chapter V.

For the WT enzyme, the substrate recognition takes place at the dimer interface, by means of residues Glu145 and Arg123, which are located at the hinge region of the β-hairpin. The process occurs through tunnel 1 in monomer B only when it is found in an O conformation. Interestingly, when analyzing the conformation of the other subunit, it displayed a PC conformation, which may suggest an interplay between MAO-N subunits. The accessibility of HEX is due when the salt bridge of Glu145 and Arg123 is cleaved, allowing the substrate to access tunnel 1. Before reaching the active site, where the aromatic cage is located, two residues are regulating its accessibility, *i.e.* Trp94 and Leu123, which are blocking the active site space (Figure 4.6 A). After some simulation time, a side-chain rotation of Phe382 generates extra room for these residues to change their side chain conformation, leaving enough space that permits the HEX diffusion to completely access the aromatic cage, displaying catalytically competent distances with the FAD cofactor. These MD

simulations revealed the importance of residues Phe382, Leu213, and Trp94 in the substrate binding process, regulating its accessibility to the active site of MAO-N.

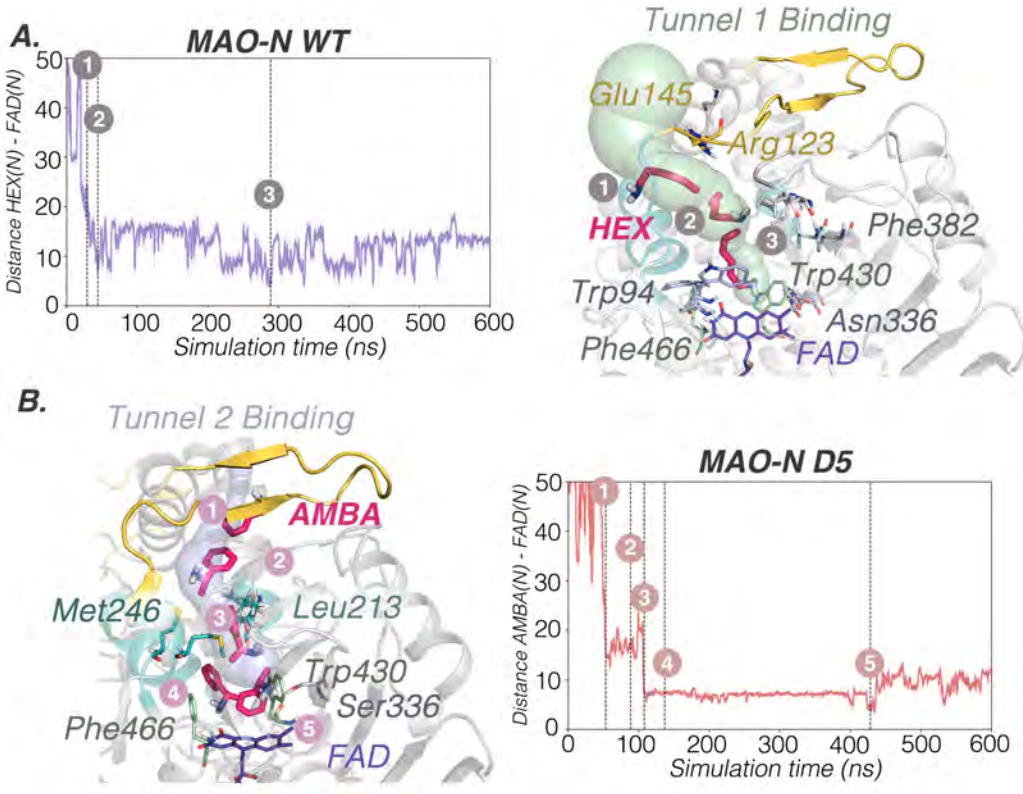


Figure 4. 6. Plot of the distance between one of the representative 600 ns aMD replicas between the FAD cofactor and A) hexylamine (HEX) or B) alpha-methyl benzylamine (AMBA). Representative snapshots taken at different points along the simulation (see numbers in plot and structure) of the reconstructed binding pathway are shown for both WT and D5 enzymes. Substrate access tunnels 1 and 2 are also represented in green and blue, respectively.

In case of MAO-N D5, two different substrate binding events could be sampled. The first one, similar to that observed for WT, with the same binding mechanism. Secondly, a binding event that is undergone through tunnel 2 could be determined (Figure 4.6 B). In such event, monomer A is found in a C conformation, whereas monomer B (the one that is binding the AMBA substrate), is found either in a PC or O conformation. The recognition of AMBA in this case takes place by means of hydrophobic interactions with the β-hairpin residues. After substrate recognition, the mutation

Met246 is the one that acts as a gate residue. Side-chain rotation of this methionine permits the AMBA entrance, reaching the aromatic cage formed by Phe466 and Trp430. Stabilization of AMBA occurs via π-stacking interaction with these residues, which allow to establish catalytically competent distances with the FAD that are stable for long simulation time. It is worth mentioning that stabilization of AMBA in the active site is correlated with the stabilization of C conformations of monomer B, which is in accordance with our hypothesis that C conformations are required for efficient catalysis. However, the potential interplay between MAO-N subunits is an issue that will be addressed in next chapter.

Chapter V. Substrate binding reconstruction reveals cooperativity-driven MAO-N activity

In the previous chapter, we deciphered the conformational dynamics of MAO-N WT and the laboratory-evolved variants D3, D5, and D9. We were able to determined that the β-hairpin is able to adopt three main conformations, *i.e.* Closed, Partially Closed, and Open conformations that were visited by these variants, although displaying different populations. Moreover, we established a correlation between β-hairpin dynamics and displayed catalytic activity, which is related to a population shift towards Closed conformations along the evolutionary path. Moreover, the importance of such conformations in the substrate binding process was also determined, showing a potential interplay between MAO-N subunits, being necessary partially closed conformations of the β-hairpin located at the opposite subunit to the one that is performing the catalysis. Within this context, in this chapter, we aim to further investigate the interplay between the β-hairpin conformational dynamics on the different subunits and its impact into catalytic activity in MAO-N WT and its evolved D5 variant, as well as decipher the key features of its molecular basis.

Similarly to what has been observed in other unrelated enzymes, the highly flexible β-hairpin region of MAO-N modulates substrate access to the active site and its recognition.[81,82] Apart from that, the conformational change of the β-hairpin of one of the monomers of the enzyme from closed to open conformation might have an impact in the active site residues of the other subunit. These results indicate that the β-hairpin conformation regulates substrate access to the active site, but can also modulate the active site architecture for efficient catalysis. However, the molecular basis of the influence of the β-hairpin conformation on the preorganization of the active site residues remains unclear. Previous observations also suggest that the reaction only occurs in one of the monomers, whereas the other is in charge of cofactor regeneration. This asymmetry was also postulated for the related human MAO-B enzyme.[83] Interestingly, the binding of an inhibitor in a cavity located at the active site entrance of MAO-B was found to be enhanced after

inhibitor binding at the active site of the enzyme, thus suggesting inter-monomer communication within the enzyme.[84] In this chapter, we aim to investigate the potential interplay between the β-hairpin conformational dynamics, inter-monomer communication between subunits, and catalysis in MAO-N WT and its evolved D5 variant, as well as decipher the key features of its molecular basis. Our MD simulations reveal that the β-hairpin conformation dramatically impacts the catalytic activity of the other subunit of the enzyme, thus confirming that efficient catalysis in MAO-N is regulated by β-hairpin conformational dynamics.

5.1. Computational details

5.1.1. Protein preparation

Protein Preparation: Molecular Dynamics simulations (MD) were performed for MAO-N WT and MAO-N D5 in the presence of hexylamine (HEX) and alpha-methyl benzylamine (AMBA) substrates in the active site. As a starting point for our MD simulations, we have selected representative structures from our previously computed FELs, corresponding to Closed and Open conformations of the β-hairpin in monomer B (always maintaining the β-hairpin of monomer A in the closed conformation). For MAO-N D5, AMBA was docked in the active site of monomer A of MAO-N D5 using Autodock Vina.[85] We selected as starting point for our MD simulations those binding poses of AMBA in which the aromatic ring interacts with both aromatic cage (in a T-shape conformation) and the isoalloxazine ring of the FAD cofactor via π-stacking. We additionally checked that the nitrogen of the amine group of AMBA is pointing towards the nitrogen of the isoalloxazine ring of the FAD cofactor establishing a hydrogen bond. For the MAO-N WT MD simulations, HEX substrate was docked in the active site of MAO-N WT using Autodock Vina. We selected as starting point for our MD simulations those binding poses of HEX in which the aliphatic chain is placed between the aromatic side chains of the

aromatic cage residues Phe466 and Trp430 and the nitrogen of the amine group of HEX is pointing towards the nitrogen of the isoalloxazine ring of the FAD cofactor.

5.1.2. Molecular Dynamics simulations

For MD simulations, the same protocol as in Chapter IV was applied, with the main difference that the General Amber Force Field (GAFF) was used for substrate parametrization.

5.1.3. Substrate Parametrization

The parameters for the FAD cofactor, HEX and AMBA substrates for the MD simulations were generated within the ANTECHAMBER module of AMBER 16 using the general AMBER force field (GAFF), with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model. The charges were calculated according to the Merz-Singh-Kollman scheme using Gaussian 09.

5.1.4. aMD simulations

Conformational dynamics. Accelerated Molecular Dynamics simulations (aMD) were used to explore the interaction between the substrate and the active site of MAO-N WT and MAO-N D5. We determined the acceleration parameters from the 100 ns unrestrained conventional MD simulations described above. For each system, five replicas of 200 ns of dual-boost accelerated Molecular Dynamics (aMD) simulations were carried out. In these simulations, we monitored the distance between the nitrogen of the amine group of the substrate and nitrogen of the FAD cofactor, i.e. AMBA(N)-FAD(N). In addition, aMD simulations were projected onto the Free Energy Landscape (FEL) of both MAO-N WT and MAO-N D5 to assess the conformational changes on the β-hairpin of monomer B.

Here a total boost potential is applied to all atoms in the system in addition to a more aggressive dihedral boost, *i.e.*, (E_{dihed}, α_{dihed}; E_{total}, α_{total}), within the dual-boost aMD approach. The acceleration parameters used in this work, are the following:

$$E\text{dihed} = V\text{dihed\_avg} + 3.5 \times N\text{res}, \alpha\text{dihed} = 3.5 \times N\text{res}/5;$$

$$E\text{total} = V\text{total\_avg} + 0.175 \times N\text{atoms}, \alpha\text{total} = 0.175 \times N\text{atoms}$$

where N_{res} is the number of protein residues, N_{atoms} is the total number of atoms, and V_{dihed_avg} and V_{total_avg} are the average dihedral and total potential energies calculated from 100 ns cMD simulations, respectively.

Free Energy Landscape of MAO-N WT and MAO-N D5: The Free Energy Landscapes (FEL) of MAO-N WT and MAO-N D5 were obtained from extensive MD simulations obtained in the previous chapter. Notwithstanding, as we are only exploring WT and D5 variants, we projected D5 simulations on the tICA space of WT. A total of 150 μs accumulated MD data were projected into a selected feature space, capturing the conformational change of the flexible β-hairpin. A feature set containing combinations of the C-alpha distances between β-hairpin loop residues 128–131, and the important β-sheet residues 353–356 as well as the angle formed by the C-alpha atoms of residues 134 (monomer B), 240 (monomer A), and 360 (monomer B) were included. Afterwards, the time-lagged independent component analysis (TICA) was used to generate a kinetically relevant space for each MAO-N WT and D5 system separately. The lag time selection of 10 ns was based on the description of 95% cumulative kinetic variance, including the smallest possible number of TICA dimensions. The generated FEL allows the discrimination between closed, partially closed, and open states of the β-hairpin in monomer B. AMD simulations were then projected on the FEL generated using this protocol to

monitor conformational changes occurring on the β-hairpin of monomer B in the presence of substrate in the other active site.

5.1.5. Shortest Path Map analyses.

The first step of the Shortest Path Map (SPM) calculation relies on the construction of a graph based on the computed mean distances and correlation values observed along the MD simulations. For each residue of the protein a node is created and centered on the C-alpha if both residues display a mean distance of less than 6 Å along the simulation time. The length of the line connecting both residues is drawn according their correlation value (dij=-log |Cij|). Larger correlation values (closer to 1 or -1) will have shorter edge distances, whereas less correlated residue pairs (values closer to 0) will have edges with long distances. At this point, we make use of Dijkstra algorithm to identify the shortest path lengths. The algorithm goes through all nodes of the graph and identifies which is the shortest path to go from the first until the last protein residue. The method therefore identifies which are the edges of the graph that are shorter, i.e. more correlated, and that are more frequently used for going through all residues of the protein, i.e. they are more central for the communication pathway.

5.2. Results and discussion

5.2.1. Revealing a communication path between MAO-N subunits by means of our Shortest Path Map tool

Considering the results obtained in Chapter IV, we aimed to elucidate the potential interplay between MAO-N subunits and its impact in the catalytic efficiency of the

different MAO-N variants. In this context, we started our study by applying our Shortest Path Map (SPM) tool to find those regions that display correlated dynamics and how this is altered along DE. More specifically, we determined the SPM for MAO-N WT, D5, and D9 variants.

Starting from the WT enzyme, we observed a path directly connecting the β-hairpin and the active site of one subunit, but also, a path connecting the active site of both MAO-N subunits, thus, supporting our hypothesis on the inter-monomer communication that this enzyme displayed on our substrate-bound MD simulations.

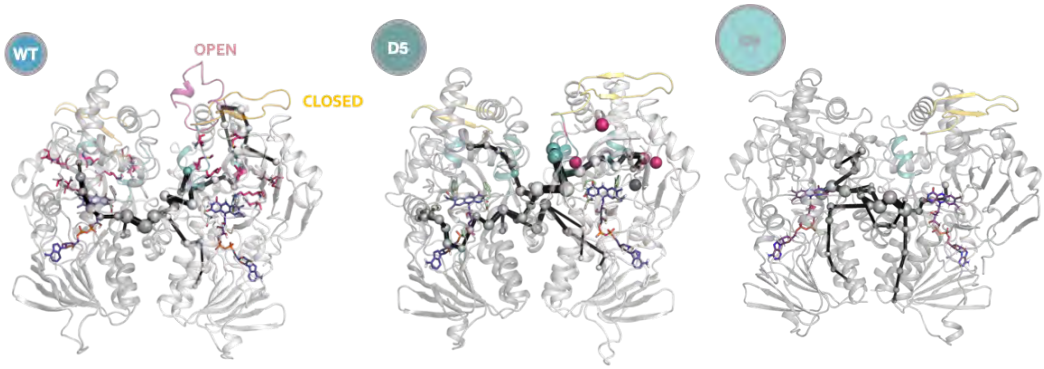


Figure 5.1. *SPM communication paths for WT (left), D5 (center), and D9 (right) revealing the potential interplay between the β-hairpin and the active site of both subunits.*

Following up with MAO-N D5 variant, we observed that the path connecting the β-hairpin and the active site of monomer B has been truncated, being only connected to some residues located at *ca.20 Å* from the β-hairpin, and the residues from the substrate tunnel with the opposite active site. From these results we can hypothesize that the activity of this variant is less dependent from the conformations displayed by the β-hairpin compared to WT.

Finally, we aimed to determine the shortest paths present in MAO-N D9. Similarly to D5, these paths showed no connection between b-hairipin and active sites, and, in this case, the most correlated motions happen on a path that connect both active sites, without including residues of the substrate tunnel or β-hairpin. These results suggest that, along the laboratory evolutionary path, the conformational dynamics of MAO-N was modified in such a way that the activity of the enzyme becomes independent of the β-hairpin dynamic behavior. This could come from the fact that the most evolved variants are mostly exploring active conformations, *i.e.* closed conformations. However, to have a deeper understanding on the effect of those conformations on the catalytic activity of MAO-N, a more extensive analysis should be performed.

5.2.2. How interplay between MAO-N subunits affect catalysis in MAO-N WT?

To elucidate how the β-hairpin conformation might affect catalysis in MAO-N WT active site we used accelerated Molecular Dynamics (aMD) simulations with HEX as substrate. Initial structures of MAO-N WT corresponding to either closed or open conformations of the β-hairpin in monomer B (and always starting with the β-hairpin of monomer A in the closed conformation) were extracted from the FEL of previous chapter (See Figure 5.2).

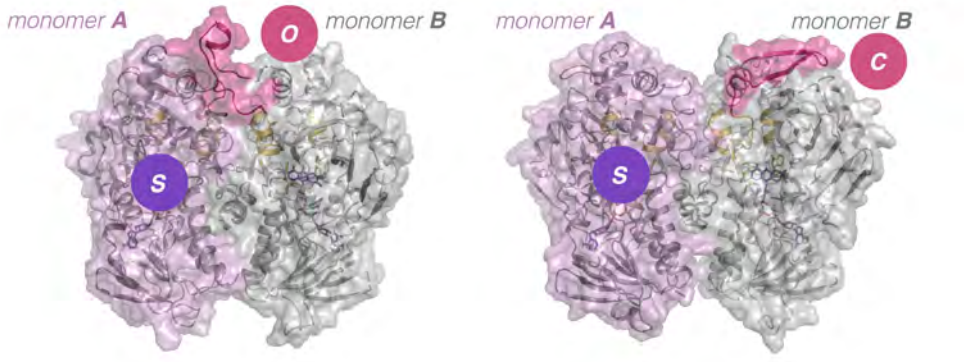


Figure 5.2. Representation of the different MD simulations performed to study the interplay of MAO-N subunits. On the left, open (O) conformations of the β-hairpin of the monomer B were considered while substrate (S) was docked in monomer A. On the right, closed (C) conformations of the β-hairpin of monomer B were simulated while substrate (S) was docked in monomer A

We started by analyzing those simulations of WT that began in the closed state of the β-hairpin in monomer B, and HEX positioned by molecular docking in the active site of monomer A, with the amine group interacting with the isoalloxazine ring of the FAD cofactor, and the aliphatic chain interacting via CH···π with the catalytic residues (Phe and Trp). The different conformations visited along the aMD simulation were projected on the respective FEL, which can properly discriminate among closed, partially closed and open hidden states of the β-hairpin (See Figure 5.3). To distinguish those conformations visited along the aMD simulation presenting catalytically productive distances for amine oxidation, a coloring scheme was used: catalytic distances (*i.e.* HEX(N) – FAD(N)) smaller than 3.5 Å were represented with green dots, whereas non-productive frames in teal (if the distances lies within 3.5-5 Å), light purple (5-6 Å), violet (6-7 Å), and dark purple (more than 7 Å). Although the simulation starts with HEX properly positioned in the active site of monomer A (displaying a C conformation of monomer B), the amine mostly adopts catalytically unproductive conformations, resulting in a 3.8% of catalytically productive frames with a mean catalytic distance of 8.2 ± 2.7 Å. We

analyzed how the angle of the opening/closing of β-hairpin in the other monomer affects the catalytic distance between the amine and FAD of monomer A. When the angle that represents the state of the β-hairpin is below 15°, *i.e.* closed conformations, the catalytic distance for amine oxidation is found above the threshold value (3.5 Å). This indicates that when the β-hairpin of monomer B is in closed states the amine is not adopting catalytically productive orientations in monomer A. By visually analyzing different snapshots visited along the aMD trajectory, some side-chain rotations of residues Trp94, Trp230 and Phe382 are observed (see Figure 5.3B). Such side-chain conformational changes expand the size of the active site cavity, thus allowing the substrate to leave the aromatic cage.

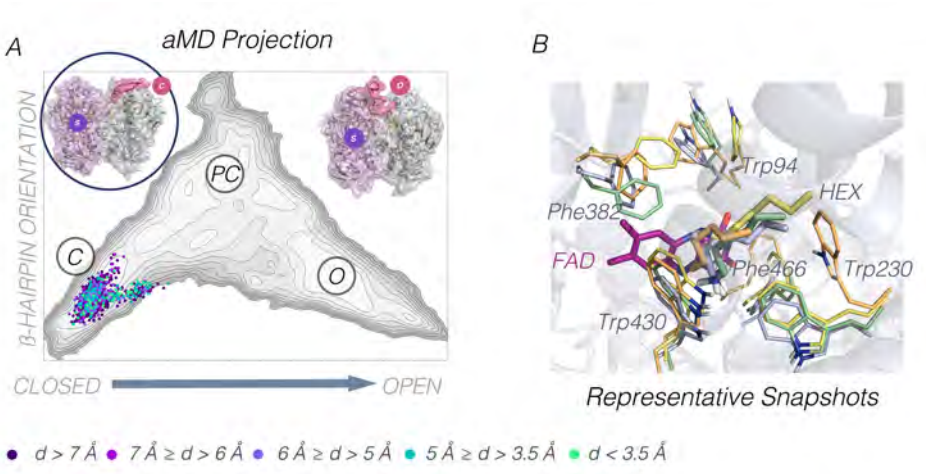


Figure 5.3. A) Free Energy Landscape (FEL) of MAO-N WT that differentiates closed (C), partially closed (PC), and open (O) states of the β-hairpin of monomer B. Projection of one of the aMD simulations starting from the closed state (C) of the β-hairpin and positioning HEX in the active site of monomer A. Each conformation sampled in the aMD simulation is represented with a dot coloured according to the value of the HEX(N) – FAD(N) catalytic distance (d): catalytically productive frames are marked in green (d < 3.5 Å), whereas non-productive frames either in teal (3.5-5 Å), light purple (5-6 Å), violet (6-7 Å), or dark purple (d > 7 Å). c) Overlay of representative frames visited along the aMD simulation.

Interestingly, the opposite behavior is observed when starting the simulations from the open O state of the β-hairpin in monomer B of WT (see Figure 5.4A). When the β-hairpin remains in an open conformation, catalytically productive distances for

amine oxidation are frequently sampled (see accumulation of green dots in O in Figure 5.5B). However, in the simulations that the system evolves towards partially-closed PC states, substrate unbinding takes place, displaying a higher population of teal and purple dots. The analysis of the catalytic distances for HEX, establishing again a threshold of 3.5 Å to distinguish productive distances among non-productive ones, shows an 18.0% of active frames with a mean distance of 6.0 ± 2.3 Å. The number of catalytically productive frames is significantly higher, and the mean distance lower than in the previous simulations that started with the β-hairpin in the closed state. The coupled analysis of the angle that determines the conformation of β-hairpin in B and the amine-FAD catalytic distance in monomer A clearly shows that catalytically competent distances are visited at large values of the β-hairpin angle (open states of the β-hairpin), whereas at partially closed conformations non-productive distances are sampled (see Figure 5.4A). By carefully analyzing the active site architecture at some representative conformations visited along the aMD simulation, we can observe that despite side-chain rotations of residues Trp230 and Phe382 taking place, HEX stays close to FAD sampling catalytically productive distances (see Figure 5.5B). These simulations again reveal a delicate communication between both monomers of MAO-N, which is crucial for the enzyme activity.

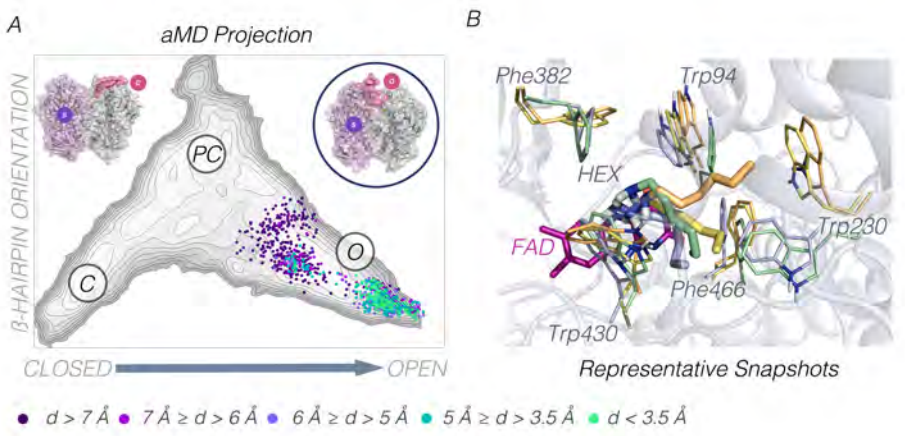


Figure 5.4. a) Free Energy Landscape (FEL) of MAO-N WT that differentiates closed (C), partially closed (PC), and open (O) states of the β-hairpin of monomer B. Projection of one of the aMD simulations starting from the open state (O) of the β-hairpin and positioning HEX in the active site of monomer A. Each conformation sampled in the aMD simulation is represented with a dot coloured according to the value of the HEX(N) – FAD(N) catalytic distance (d): catalytically productive frames are marked in green (d < 3.5 Å), whereas non-productive frames either in teal (3.5-5 Å), light purple (5-6 Å), violet (6-7 Å), or dark purple (d > 7 Å). b) Overlay of representative frames visited along the aMD simulation.

Overall, our results indicate the necessity of O conformations of the β-hairpin of the opposite subunits for the catalytic process to take place. However, considering our results from Chapter IV, as well as the truncated shortest path map in MAO-N D5, a totally different behavior might be expected for this variant.

5.2.3. Interplay between β-hairpin dynamics and catalytic activity in MAO-N D5

To assess the connection between monomer B β-hairpin conformational dynamics and monomer A catalysis in MAO-N D5, the substrate was docked in the active site cavity of the A subunit for open and closed states of monomer B β-hairpin, following the same protocol as for MAO-N WT. To elucidate the interplay between

β-hairpin conformation of monomer B and productive binding of AMBA in A, we followed the same procedure as for MAO-N WT. We projected our MD simulations on the previously computed FELs as a scatter plot, coloring the dots according to its catalytic distance (See Figures 5.5 and 5.6)

We started with the analysis of the simulations starting from the C state of the β-hairpin of monomer B of D5, and positioning AMBA in the active site of monomer A (see Figure 5.5). The projection of this simulation shows many catalytically competent productive poses for amine oxidation, which occur while the β-hairpin of monomer B is maintained in the closed conformation (see accumulation of green dots in C state). By counting the number of frames that display catalytic distances below the threshold value (*i.e.* 3.5 Å) we observed a 54.5% of catalytically productive frames, with mean distances of 3.8 ± 0.9 Å (Figure B1) indicating that the substrate remains in the active site during the whole simulation time. By careful analysis of the catalytically active frames visited along the aMD simulations, we observe that the substrate is interacting always via π-stacking with the isoalloxazine ring of the FAD, and in a T-shape conformation with the residues forming the aromatic cage (*i.e.* Phe466 and Trp430) pointing the amine group of AMBA towards FAD(N). To further confirm the interplay of the β-hairpin conformation and the catalytic activity of the other subunit we analyzed how the catalytic distances sampled in monomer A correlate with an angle that represents the β-hairpin opening/closure in B. Again, angles below 15 degrees correspond to closed conformations, whereas angles above this value lead to hidden partially closed or open conformations.

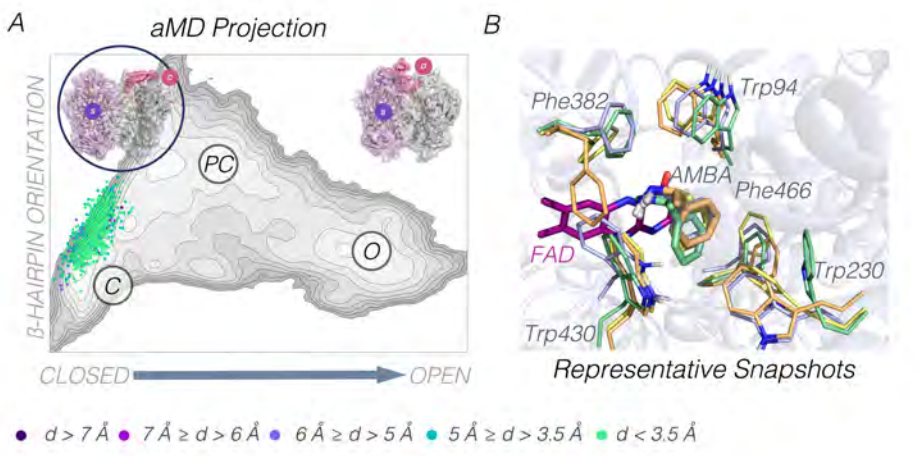


Figure 5.5. a) Free Energy Landscape (FEL) of MAO-N D5 that differentiates closed (C), partially closed (PC), and open (O) states of the β-hairpin of monomer B. Projection of one of the aMD simulations starting from the closed state (C) of the β-hairpin and positioning AMBA in the active site of monomer A. Each conformation sampled in the aMD simulation is colored according to the value of the AMBA(N) – FAD(N) catalytic distance (d): catalytically productive frames are marked in green (d < 3.5 Å), whereas non-productive frames either in teal (3.5-5 Å), light purple (5-6 Å), violet (6-7 Å), or dark purple (d > 7 Å). b) Overlay of representative frames visited along the aMD simulation.

As observed in the FEL projection, catalytic distances are sampled when the β-hairpin angle in B is found below 15 degrees (C state). In the conformations where short catalytic distances are sampled the aromatic ring of AMBA is always properly positioned in the aromatic cage for productive catalysis. The visual analysis of selected conformations confirms that AMBA is properly oriented for catalysis and also active site residues remain stable along the simulation (see Figure 5.5B). In particular, Trp94 and Phe382, which are responsible for keeping AMBA inside the aromatic cage, do not show significant conformational changes along the aMD trajectories. These simulations thus confirm that productive amine oxidation in monomer A of D5 is achieved when the β-hairpin in monomer B adopts a closed conformation.

A completely different picture is observed in the simulations in which AMBA is located in the active site of monomer A, but the β-hairpin in monomer B is in a

open state (O, in Figure 5.5). In contrast to the closed conformation, the open state of the β-hairpin favors the side chain rotation of residues Trp230, Leu245, and Phe382 located in the active site of monomer A that increases the size of the pocket leaving more space for AMBA to leave from the aromatic cage (see Figure 5.5 B). When the β-hairpin of monomer B stays in the open conformation, catalytic distances are hardly sampled (see Figure 5.5 A). Interestingly, there is a particular replica that after a few ns of aMD simulation the β-hairpin of B rapidly escapes from the open state evolving towards hidden partially closed conformations (Figure 5.6A), as it can be observed in the aMD projection (see accumulation of points in PC in Figure 5.6A). After this transition of the β-hairpin of monomer B, the side chains of Trp230, Leu245, and Phe382 residues in monomer A are reverted back to their original positions. It is worth highlighting that, in this simulation, as soon as the β-hairpin B adopts a partially closed conformation, catalytically productive distances are sampled in the active site of monomer A (see green dots accumulated on PC in Figure 5.6A). A further analysis of the angle of β-hairpin B opening and the catalytic distance for amine oxidation sampled in this aMD simulation clearly shows that non-productive distances between AMBA and FAD are observed at the open state (angles above 60º, see Figure 5.6B). At long catalytic distances the aromatic ring of AMBA is not properly positioned in the aromatic cage for efficient amine oxidation. Due to the initial open conformations of the β-hairpin, we observe a decrease in the amount of catalytically productive frames with respect to the other simulation that started from the closed state, representing a 21.7% with mean distances of 6.7 ± 0.8 Å. This simulation therefore demonstrates that the open state in MAO-N-D5 plays no role in properly positioning the substrate for amine oxidation.

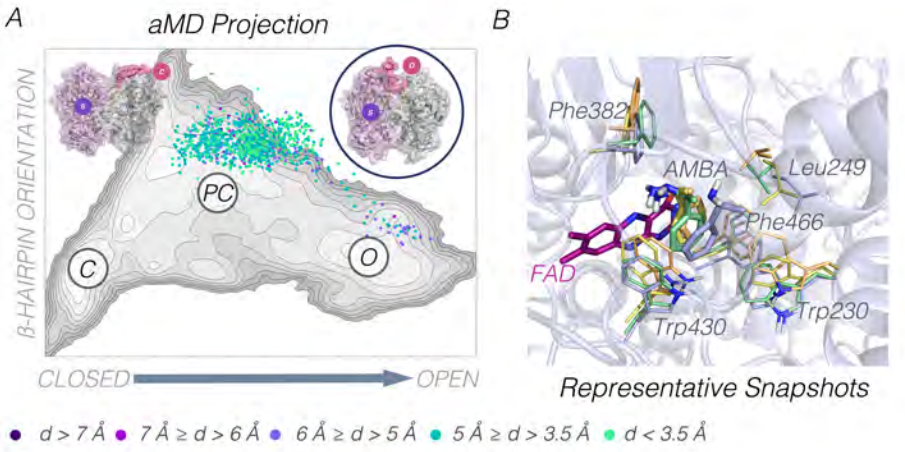


Figure 5.6. a) Free Energy Landscape (FEL) of MAO-N D5 that differentiates closed (C), partially closed (PC), and open (O) states of the β-hairpin of monomer B. Projection of one of the aMD simulations starting from the open state (O) of the β-hairpin and positioning AMBA in the active site of monomer A. Each conformation sampled in the aMD simulation is colored according to the value of the AMBA(N) – FAD(N) catalytic distance (d): catalytically productive frames are marked in green (d < 3.5 Å), whereas non-productive frames either in teal (3.5-5 Å), light purple (5-6 Å), violet (6-7 Å), or dark purple (d > 7 Å). b) Overlay of representative frames visited along the aMD simulation.

Overall, these simulations for MAO-N D5 indicate that the conformational dynamics of the β-hairpin in one monomer of the enzyme influences the active site architecture and thus the catalytic efficiency of the other subunit. Our results also show that closed and partially closed conformations of the β-hairpin are crucial for effective catalysis in monomer A of MAO-N D5, inducing a stabilization of catalytically active poses of AMBA for efficient amine oxidation. These results are in line with our previous study where we showed that the closed and partially closed states are more accessible than the open state in the laboratory-evolved variant D5. As a consequence, and based on the percentage of catalytically active frames, D5 variant is exploring catalytically productive distances more frequently than WT.

Chapter VI. Conclusions

Prompted by the urge of new methodologies for the obtention of chiral amines, in this thesis the conformational dynamics of set of laboratory-evolved MAO-N variants were studied. The main conclusions of the thesis are the following

In Chapter IV we demonstrated that distal mutations introduced via DE regulate MAO-N activity and substrate scope by modulating the relative stabilities of catalytically important conformational states of a β-hairpin located at *ca.* 20 Å from the active site. MSM and metadynamics simulation were used to elucidate the relative stabilities (metadynamics and MSM) and the timescales of the conformational transitions (MSM). The results showed that along the laboratory evolution, closed conformations of the β-hairpin were stabilized, increasing the timescales needed to reach open conformations. The delicate equilibrium existing between closed, and the partially closed and open states regulates MAO-N catalytic activity by allowing the substrate to reach the active site pocket, but also by altering the active site cavity for efficient catalysis. Moreover, tunnel analysis pointed out that those closed conformations are also stabilizing the presence of substrate tunnel, increasing their population and, thus, facilitating the substrate binding process. Our study highlights the key role played by the enzyme conformational dynamics in the enzyme activity, and supports the idea that future computational enzyme design strategies will need to further characterize the conformational landscape.

In Chapter V we deciphered the effect of a highly flexible β-hairpin region of MAO-N enzyme on its catalytic activity. First, by means of SPM, we detected an existing communication path between the β-hairpin of one subunit, with the active site of the opposite. Along the evolution, this path has been truncanted, which suggests that the enzyme activity became more independent of the β-hairpin dynamics in the evolved variants. Our simulations based on the enhanced sampling technique aMD confirmed such communication between

both MAO-N subunits that impacts the active site architecture, and thus its catalytic efficiency. In both MAO-N WT and the laboratory evolved D5 variant, the conformation of a flexible β-hairpin in one of the monomers affects the productive binding of the substrate in the active site of the other subunit. However, both MAO-N WT and D5 variants behave quite differently due to the distal mutations introduced with Directed Evolution. In the evolved D5 variant, the closed conformation of the β-hairpin found to be stabilized as compared to WT is the only one able of efficiently positioning the substrate in the active site of the other monomer for productive catalysis. This is not the case for MAO-N WT for which the open conformation of the β-hairpin is substantially more stable and essential for hexylamine oxidation in the other monomer. Our findings on MAO-N conformational dynamics, cooperativity between subunits and its connection to catalysis can be targeted for the engineering of additional MAO-N variants, but also for unrelated homodimeric enzymes where similar cooperative effects could also operate.

The outcome of this thesis might lead to new investigations towards the development of more efficient MAO-N variants. Considering the importance of the β-hairpin in MAO-N conformational dynamics, and how it can modulate the enzyme activity by regulating the substrate binding (as pointed in Chapter IV), or by affecting the catalysis in the opposite subunit (Chapter V), the β-hairpin region can be targeted for mutagenesis, trying to keep stabilizing its closed conformations, facilitating the substrate binding, and thus, leading to more efficient MAO-N variants.

.

References

(1) A. Copeland, R. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, 2nd Edition | Wiley*, Wiley-VCH.; New York.

(2) Bornscheuer, U. T.; Buchholz, K. Highlights in Biocatalysis – Historical Landmarks and Current Trends - Bornscheuer - 2005 - Engineering in Life Sciences - Wiley Online Library. *Eng. Life Sci.* **2005**, *5* (4), 309–323.

(3) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Berichte Dtsch. Chem. Ges.* **1894**, *27* (3), 2985–2993. https://doi.org/10.1002/cber.18940270364.

(4) Sanger, F.; Thompson, E. O. P.; Kitai, R. The Amide Groups of Insulin. *Biochem. J.* **1955**, *59* (3), 509–518.

(5) Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J. The Role of Protein Dynamics in the Evolution of New Enzyme Function. *Nat. Chem. Biol.* **2016**, *12* (11), 944–950. https://doi.org/10.1038/nchembio.2175.

(6) Knowles, J. R. Enzyme Catalysis: Not Different, Just Better. *Nature* **1991**, *350* (6314), 121–124. https://doi.org/10.1038/350121a0.

(7) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, 106 (8), 3210–3235.

(8) Nagel, Z. D.; Klinman, J. P. A 21st Century Revisionist's View at a Turning Point in Enzymology. *Nat. Chem. Biol.* **2009**, *5* (8), 543–550. https://doi.org/10.1038/nchembio.204.

(9) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, 301 (5637), 1196–1202.

(10) Mohrdieck, C. Physical Chemistry for the Life Sciences. By Peter Atkins and Julio de Paula. *ChemPhysChem* **2006**, *7* (5), 1149–1149. https://doi.org/10.1002/cphc.200600131.

(11) Cornish-Bowden, A.; Mazat, J.-P.; Nicolas, S. Victor Henri: 111 Years of His Equation. *Biochimie* **2014**, *107*, 161–166. https://doi.org/10.1016/j.biochi.2014.09.018.

(12) Michaelis, L.; Menten, M. M. L. The Kinetics of Invertin Action. *FEBS Lett.* **2013**, *587* (17), 2712–2720. https://doi.org/10.1016/j.febslet.2013.07.015.

(13) Graham Brown, T. PROCEEDINGS OF THE PHYSIOLOGICAL SOCIETY: January 22, 1910. *J. Physiol.* **1910**, *40* (suppl), i–vii.
(14) Clark, C. R. A Stopped-Flow Kinetics Experiment for Advanced Undergraduate Laboratories: Formation of Iron(III) Thiocyannate. *J. Chem. Educ.* **1997**, *74* (10), 1214. https://doi.org/10.1021/ed074p1214.
(15) Hoag, C. M. Simple and Inexpensive Computer Interface to a Durrum Stopped-Flow Apparatus Tested Using the Iron(III)–Thiocyanate Reaction. *J. Chem. Educ.* **2005**, *82* (12), 1823. https://doi.org/10.1021/ed082p1823.
(16) Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M. Conformational Dynamics and Enzyme Evolution. *J. R. Soc. Interface* **2018**, *15* (144), 20180330. https://doi.org/10.1098/rsif.2018.0330.
(17) Koshland Jr., D. E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chem. Int. Ed. Engl.* **1995**, *33* (23–24), 2375–2378. https://doi.org/10.1002/anie.199423751.
(18) Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. Folding Funnels and Binding Mechanisms. *Protein Eng. Des. Sel.* **1999**, *12* (9), 713–720. https://doi.org/10.1093/protein/12.9.713.
(19) Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.
(20) Kovermann, M.; Grundström, C.; Sauer-Eriksson, A. E.; Sauer, U. H.; Wolf-Watz, M. Structural Basis for Ligand Binding to an Enzyme by a Conformational Selection Pathway. *Proc. Natl. Acad. Sci.* **2017**, *114* (24), 6298–6303. https://doi.org/10.1073/pnas.1700919114.
(21) Nussinov, R. Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **2016**, *116* (11), 6263–6266. https://doi.org/10.1021/acs.chemrev.6b00283.
(22) Svensson, L. A.; Sjölin, L.; Gilliland, G. L.; Finzel, B. C.; Wlodawer, A. Multiple Conformations of Amino Acid Residues in Ribonuclease A. *Proteins Struct. Funct. Bioinforma.* **1986**, *1* (4), 370–375.
(23) Kohen, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **2015**, *48* (2), 466–473. https://doi.org/10.1021/ar500322s.
(24) Warshel, A.; Bora, R. P. Perspective: Defining and Quantifying the Role of Dynamics in Enzyme Catalysis. *J. Chem. Phys.* **2016**, *144* (18), 180901. https://doi.org/10.1063/1.4947037.
(25) Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **2004**, *104* (8), 3623–3640. https://doi.org/10.1021/cr030413t.

(26) Boehr, D. D.; Dyson, H. J.; Wright, P. E. An NMR Perspective on Enzyme Dynamics. *Chem. Rev.* **2006**, *106* (8), 3055–3079. https://doi.org/10.1021/cr050312q.
(27) Aviram, H. Y.; Pirchi, M.; Mazal, H.; Barak, Y.; Riven, I.; Haran, G. Direct Observation of Ultrafast Large-Scale Dynamics of an Enzyme under Turnover Conditions. *Proc. Natl. Acad. Sci.* **2018**, *115* (13), 3243–3248. https://doi.org/10.1073/pnas.1720448115.
(28) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden Alternative Structures of Proline Isomerase Essential for Catalysis. *Nature* **2009**, *462* (7273), 669–673. https://doi.org/10.1038/nature08615.
(29) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. Linkage between Dynamics and Catalysis in a Thermophilic-Mesophilic Enzyme Pair. *Nat. Struct. Mol. Biol.* **2004**, *11* (10), 945–949. https://doi.org/10.1038/nsmb821.
(30) Stiller, J. B.; Jordan Kerns, S.; Hoemberger, M.; Cho, Y.-J.; Otten, R.; Hagan, M. F.; Kern, D. Probing the Transition State in Enzyme Catalysis by High-Pressure NMR Dynamics. *Nat. Catal.* **2019**, *2* (8), 726–734. https://doi.org/10.1038/s41929-019-0307-6.
(31) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A Hierarchy of Timescales in Protein Dynamics Is Linked to Enzyme Catalysis. *Nature* **2007**, *450* (7171), 913–916. https://doi.org/10.1038/nature06407.
(32) Cheng, Y.; Grigorieff, N.; Penczek, P. A.; Walz, T. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* **2015**, *161* (3), 438–449. https://doi.org/10.1016/j.cell.2015.03.050.
(33) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Computational Tools for the Evaluation of Laboratory-Engineered Biocatalysts. *Chem. Commun.* **2016**, *53* (2), 284–297. https://doi.org/10.1039/C6CC06055B.
(34) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254* (5038), 1598–1603. https://doi.org/10.1126/science.1749933.
(35) Diez, M.; Zimmermann, B.; Börsch, M.; König, M.; Schweinberger, E.; Steigmiller, S.; Reuter, R.; Felekyan, S.; Kudryavtsev, V.; Seidel, C. A. M.; Gräber, P. Proton-Powered Subunit Rotation in Single Membrane-Bound F0F1-ATP Synthase. *Nat. Struct. Mol. Biol.* **2004**, *11* (2), 135–141. https://doi.org/10.1038/nsmb718.
(36) Myong, S.; Stevens, B. C.; Ha, T. Bridging Conformational Dynamics

and Function Using Single-Molecule Spectroscopy. *Struct. Lond. Engl. 1993* **2006**, *14* (4), 633–643. https://doi.org/10.1016/j.str.2006.02.005.
(37) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40* (1), 41–62. https://doi.org/10.1146/annurev-biophys-042910-155255.
(38) Jolliffe, I. T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, *374* (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.
(39) Schultze, S.; Grubmüller, H. Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *J. Chem. Theory Comput.* **2021**, *17* (9), 5766–5776. https://doi.org/10.1021/acs.jctc.1c00273.
(40) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Struct. Lond. Engl. 1993* **2006**, *14* (3), 437–449. https://doi.org/10.1016/j.str.2005.11.014.
(41) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.
(42) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144. https://doi.org/10.1016/j.sbi.2014.04.002.
(43) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140* (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.
(44) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144. https://doi.org/10.1016/j.sbi.2014.04.002.
(45) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929. https://doi.org/10.1063/1.1755656.
(46) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1* (5), 826–843. https://doi.org/10.1002/wcms.31.
(47) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.
(48) Arnold, F. H. The Nature of Chemical Innovation: New Enzymes by Evolution. *Q. Rev. Biophys.* **2015**, *48* (4), 404–410.

https://doi.org/10.1017/S003358351500013X.
(49) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. Synthetic Biology for the Directed Evolution of Protein Biocatalysts: Navigating Sequence Space Intelligently. *Chem. Soc. Rev.* **2015**, *44* (5), 1172–1239. https://doi.org/10.1039/c4cs00351a.
(50) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.
(51) Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chem. Int. Ed.* **2011**, *50* (1), 138–174.
(52) Reetz, M. T.; Wang, L.-W.; Bocola, M. Directed Evolution of Enantioselective Enzymes: Iterative Cycles of CASTing for Probing Protein-Sequence Space. *Angew. Chem. Int. Ed.* **2006**, *45* (8), 1236–1241. https://doi.org/10.1002/anie.200502746.
(53) Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J. Engineering Enzyme Access Tunnels. *Biotechnol. Adv.* **2019**, *37* (6), 107386. https://doi.org/10.1016/j.biotechadv.2019.04.008.
(54) Ma, B.; Nussinov, R. Conformational Footprints. *Nat. Chem. Biol.* **2016**, *12* (11), 890–891. https://doi.org/10.1038/nchembio.2212.
(55) Otten, R.; Pádua, R. A. P.; Bunzel, H. A.; Nguyen, V.; Pitsawong, W.; Patterson, M.; Sui, S.; Perry, S. L.; Cohen, A. E.; Hilvert, D.; Kern, D. How Directed Evolution Reshapes the Energy Landscape in an Enzyme to Boost Catalysis. *Science* **2020**, *370* (6523), 1442–1446. https://doi.org/10.1126/science.abd3623.
(56) Otten, R.; Liu, L.; Kenner, L. R.; Clarkson, M. W.; Mavor, D.; Tawfik, D. S.; Kern, D.; Fraser, J. S. Rescue of Conformational Dynamics in Enzyme Catalysis by Directed Evolution. *Nat. Commun.* **2018**, *9* (1), 1314. https://doi.org/10.1038/s41467-018-03562-9.
(57) Campbell, E. C.; Correy, G. J.; Mabbitt, P. D.; Buckle, A. M.; Tokuriki, N.; Jackson, C. J. Laboratory Evolution of Protein Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2018**, *50*, 49–57. https://doi.org/10.1016/j.sbi.2017.09.005.
(58) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116* (11), 6503–6515. https://doi.org/10.1021/acs.chemrev.5b00590.
(59) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in TRNA:Protein Complexes. *Proc. Natl. Acad. Sci.* **2009**, *106* (16),

6620–6625. https://doi.org/10.1073/pnas.0810961106.
(60) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.* **2014**, *10* (2), 511–517. https://doi.org/10.1021/ct4008603.
(61) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **2017**, *7* (12), 8524–8532. https://doi.org/10.1021/acscatal.7b02954.
(62) Maria-Solano, M. A.; Kinateder, T.; Iglesias-Fernández, J.; Sterner, R.; Osuna, S. In Silico Identification and Experimental Validation of Distal Activity-Enhancing Mutations in Tryptophan Synthase. *ACS Catal.* **2021**, *11* (21), 13733–13743. https://doi.org/10.1021/acscatal.1c03950.
(63) Wu, J.; Wang, F.; Ma, Y.; Cui, X.; Cun, L.; Zhu, J.; Deng, J.; Yu, B. Asymmetric Transfer Hydrogenation of Imines and Iminiums Catalyzed by a Water-Soluble Catalyst in Water. *Chem. Commun.* **2006**, No. 16, 1766. https://doi.org/10.1039/b600496b.
(64) Uematsu, N.; Fujii, A.; Hashiguchi, S.; Ikariya, T.; Noyori, R. Asymmetric Transfer Hydrogenation of Imines. *J. Am. Chem. Soc.* **1996**, *118* (20), 4916–4917. https://doi.org/10.1021/ja960364k.
(65) Li, C.; Xiao, J. Asymmetric Hydrogenation of Cyclic Imines with an Ionic Cp*Rh(III) Catalyst. *J. Am. Chem. Soc.* **2008**, *130* (40), 13208–13209. https://doi.org/10.1021/ja8050958.
(66) Guo, S.; Yang, J. C.; Buchwald, S. L. A Practical Electrophilic Nitrogen Source for the Synthesis of Chiral Primary Amines by Copper-Catalyzed Hydroamination. *J. Am. Chem. Soc.* **2018**, *140* (46), 15976–15984. https://doi.org/10.1021/jacs.8b10564.
(67) Nugent, T. C.; El-Shazly, M. Chiral Amine Synthesis – Recent Developments and Trends for Enamide Reduction, Reductive Amination, and Imine Reduction. *Adv. Synth. Catal.* **2010**, *352* (5), 753–819. https://doi.org/10.1002/adsc.200900719.
(68) Ghislieri, D.; Turner, N. J. Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines. *Top. Catal.* **2014**, *57* (5), 284–300. https://doi.org/10.1007/s11244-013-0184-1.
(69) Schilling, B.; Lerch, K. Amine Oxidases from Aspergillus Niger: Identification of a Novel Flavin-Dependent Enzyme. *Biochim. Biophys. Acta* **1995**, *1243* (3), 529–537. https://doi.org/10.1016/0304-4165(94)00183-x.
(70) Alexeeva, M.; Enright, A.; Dawson, M. J.; Mahmoudian, M.; Turner, N. J. Deracemization of α-Methylbenzylamine Using an Enzyme Obtained by In

Vitro Evolution. *Angew. Chem. Int. Ed.* **2002**, *41* (17), 3177–3180. https://doi.org/10.1002/1521-3773(20020902)41:17<3177::AID-ANIE3177>3.0.CO;2-P.
(71) Carr, R.; Alexeeva, M.; Dawson, M. J.; Gotor-Fernández, V.; Humphrey, C. E.; Turner, N. J. Directed Evolution of an Amine Oxidase for the Preparative Deracemisation of Cyclic Secondary Amines. *ChemBioChem* **2005**, *6* (4), 637–639. https://doi.org/10.1002/cbic.200400329.
(72) Dunsmore, C. J.; Carr, R.; Fleming, T.; Turner, N. J. A Chemo-Enzymatic Route to Enantiomerically Pure Cyclic Tertiary Amines. *J. Am. Chem. Soc.* **2006**, *128* (7), 2224–2225. https://doi.org/10.1021/ja058536d.
(73) Ghislieri, D.; Green, A. P.; Pontini, M.; Willies, S. C.; Rowles, I.; Frank, A.; Grogan, G.; Turner, N. J. Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *J. Am. Chem. Soc.* **2013**, *135* (29), 10863–10869. https://doi.org/10.1021/ja4051235.
(74) Herter, S.; Medina, F.; Wagschal, S.; Benhaïm, C.; Leipold, F.; Turner, N. J. Mapping the Substrate Scope of Monoamine Oxidase (MAO-N) as a Synthetic Tool for the Enantioselective Synthesis of Chiral Amines. *Bioorg. Med. Chem.* **2018**, *26* (7), 1338–1346. https://doi.org/10.1016/j.bmc.2017.07.023.
(75) Li, G.; Yao, P.; Gong, R.; Li, J.; Liu, P.; Lonsdale, R.; Wu, Q.; Lin, J.; Zhu, D.; Reetz, M. T. Simultaneous Engineering of an Enzyme's Entrance Tunnel and Active Site: The Case of Monoamine Oxidase MAO-N. *Chem. Sci.* **2017**, *8* (5), 4093–4099. https://doi.org/10.1039/C6SC05381E.
(76) Woods, R. J.; Chappelle, R. Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. *Theochem* **2000**, *527* (1–3), 149–156. https://doi.org/10.1016/S0166-1280(00)00487-5.
(77) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103. https://doi.org/10.1103/PhysRev.159.98.
(78) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of Slow Diffusive Conformational Transitions with Accelerated Molecular Dynamics. *J. Chem. Phys.* **2007**, *127* (15), 155102. https://doi.org/10.1063/1.2789432.
(79) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Comput. Biol.* **2012**, *8* (10),

e1002708. https://doi.org/10.1371/journal.pcbi.1002708.
(80) Atkin, K. E.; Reiss, R.; Koehler, V.; Bailey, K. R.; Hart, S.; Turkenburg, J. P.; Turner, N. J.; Brzozowski, A. M.; Grogan, G. The Structure of Monoamine Oxidase from Aspergillus Niger Provides a Molecular Context for Improvements in Activity Obtained by Directed Evolution. *J. Mol. Biol.* **2008**, *384* (5), 1218–1231. https://doi.org/10.1016/j.jmb.2008.09.090.
(81) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113* (8), 5871–5923. https://doi.org/10.1021/cr300384w.
(82) Kreß, N.; Halder, J. M.; Rapp, L. R.; Hauer, B. Unlocked Potential of Dynamic Elements in Protein Structures: Channels and Loops. *Curr. Opin. Chem. Biol.* **2018**, *47*, 109–116. https://doi.org/10.1016/j.cbpa.2018.09.010.
(83) DeRose, V. J.; Woo, J. C. G.; Hawe, W. P.; Hoffman, B. M.; Silverman, R. B.; Yelekci, K. Observation of a Flavin Semiquinone in the Resting State of Monoamine Oxidase B by Electron Paramagnetic Resonance and Electron Nuclear Double Resonance Spectroscopy. *Biochemistry* **1996**, *35* (34), 11085–11091. https://doi.org/10.1021/bi960749f.
(84) Bonivento, D.; Milczek, E. M.; McDonald, G. R.; Binda, C.; Holt, A.; Edmondson, D. E.; Mattevi, A. Potentiation of Ligand Binding through Cooperative Effects in Monoamine Oxidase B. *J. Biol. Chem.* **2010**, *285* (47), 36849–36856. https://doi.org/10.1074/jbc.m110.169482.
(85) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

Appendices

APPENDIX A

Supplementary information Chapter IV

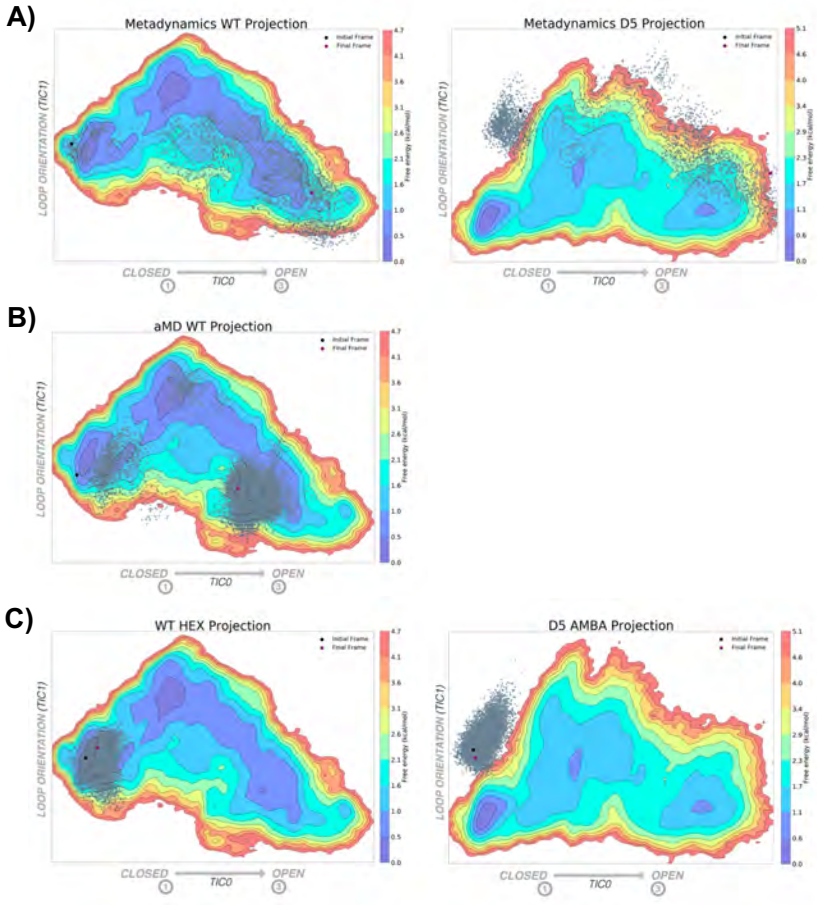


Figure A1. A) Projection of the MAO-N WT and D5 conformations along the closed-open metadynamics simulations (grey points) into the TICA space defining the hMSM free energy surfaces. B) Projection of the MAO-N WT aMD into the corresponding hMSM free energy surface. C) Projection MAO-N WT and D5 conformations sample during aMD substrate binding simulations projected into the corresponding hMSM free energy surface.

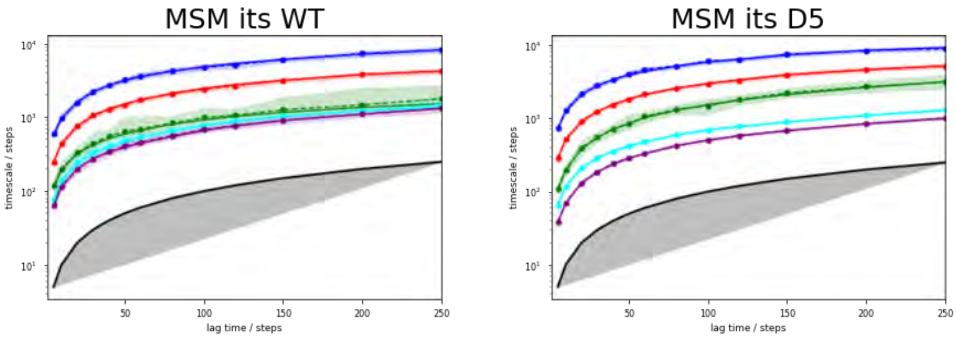


Figure A2. Implied timescales of the MAO-N WT and D5 variant Bayesian Markov State models (MSM). The top 5 implied timescales calculated at a range of lag times are shown.

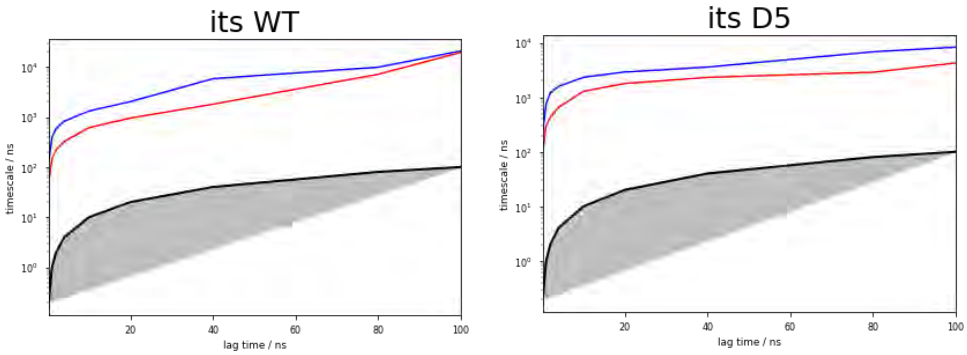


Figure A3. Implied timescales of the MAO-N WT and D5 variant hidden Markov State models (MSM).

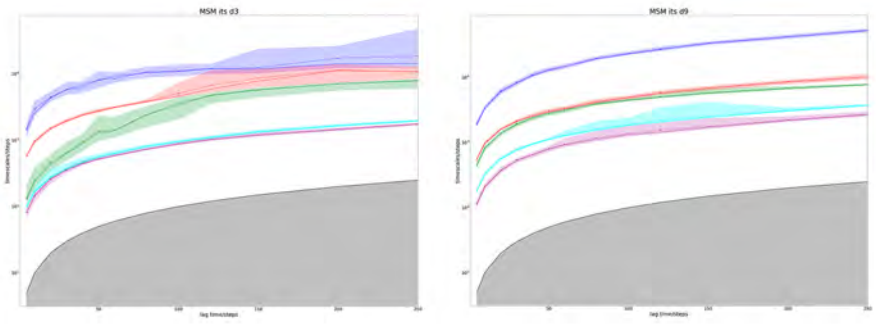


Figure A4. Implied timescales (ITS) of MAO-N D3 and D9 variants. As mentioned in chapter IV, more sampling of D3 is required to fully stabilized the ITS of both variants for the convergence of the MSM

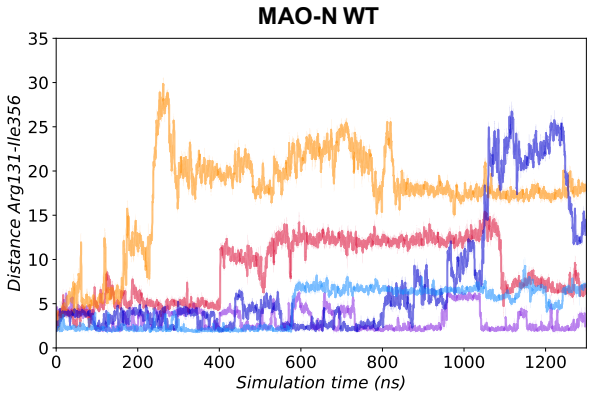


Figure A5. Five replicas of 1.3 μs of aMD starting from the closed MAO-N WT conformation. Plot of the distance (in Å) between the hydrogen of the amide backbone of Arg131 and the oxygen of the carbonyl backbone of Ile356 of monomer B along 1.3 μs of aMD simulations. Distances higher than 20 Å indicate the closed to open state transition.

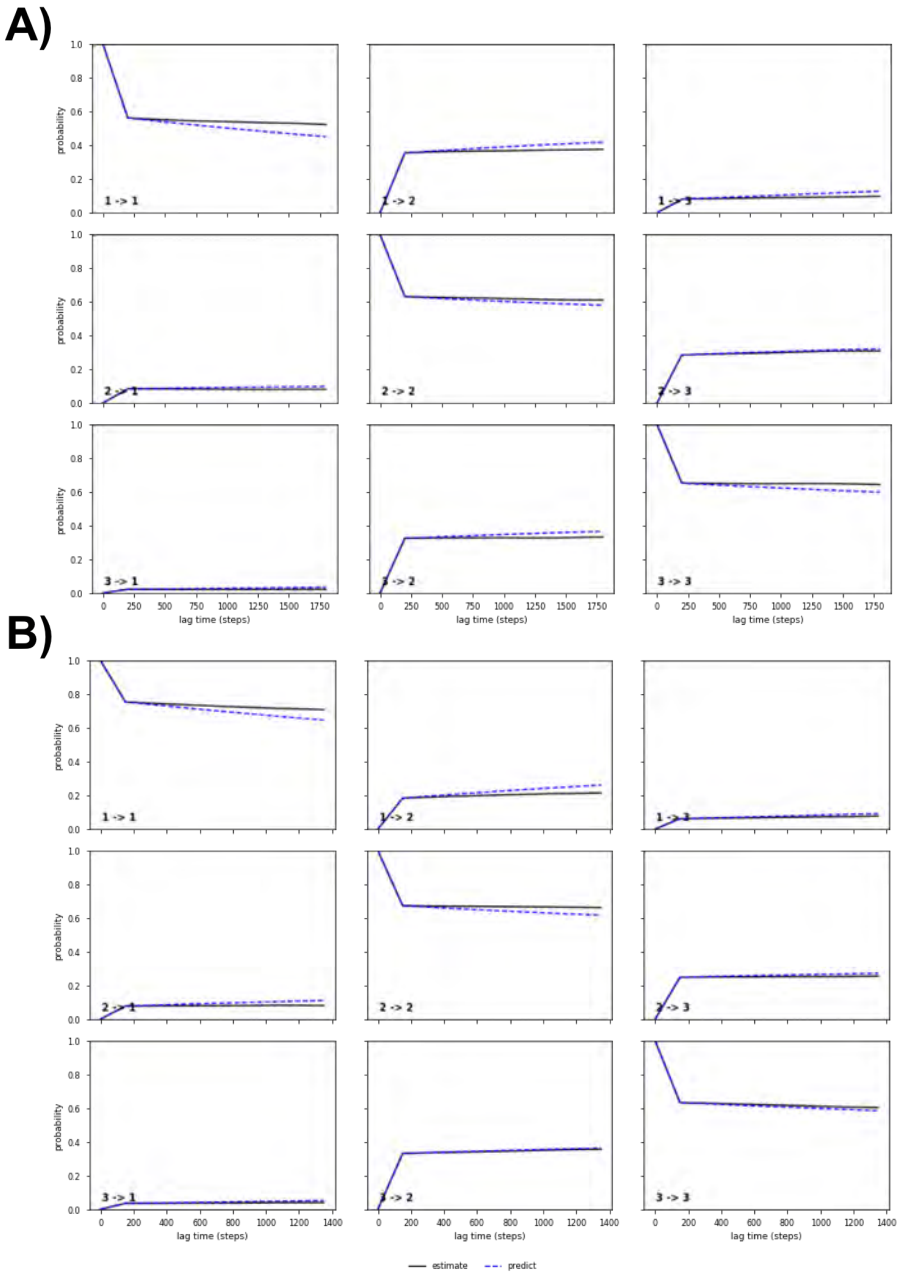


Figure A6. Chapman-Kolmogorov test of the MAO-N WT (A) and D5 variant (B) Markov State model to assess kinetic self-consistency of the model.

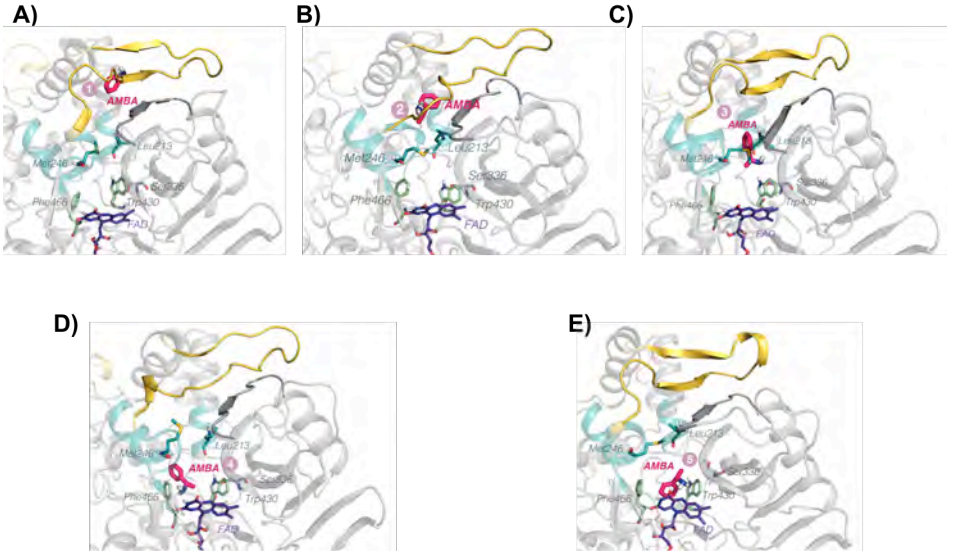


Figure A7. Five representative snapshots of alpha-methyl benzylamine (AMBA) substrate binding process in the D5 variant. A,B) The recognition is assisted by the side-chains of monomer A β-hairpin hydrophobic residues Leu130, Ile356, and Phe128. At this point, monomer A is found in a closed state, whereas monomer B is in an open/partially closed conformation. Prior to AMBA recognition, the β-hairpin of monomer A performs a lateral displacement that resembles the one observed in the partially closed state. The recognition occurs through tunnel 2 rather than 1 observed in WT. C) After recognition, AMBA rapidly progresses through the substrate tunnel defined by Met246, Tyr365, Leu245, and Phe382. In contrast to WT, the new Met246 introduced (i.e. Ile246Met) acts now as a gate and modulates the access to the aromatic cage where FAD is located. D) At some point, Met246 undergoes a side-chain conformational change that allows AMBA to get closer to the active site aromatic cavity in a non-productive conformation (distances between AMBA(N) and FAD(N) longer than 6 Å). E) After 300 ns of aMD simulation time, AMBA gets properly positioned between the aromatic cage residues Phe466 and Trp430 establishing π-stacking interactions with both residues. This corresponds to the catalytically competent pose (with distances between FAD and AMBA shorter than ca. 4 Å), which is maintained for a few ns in this enhanced sampling aMD simulation.

APPENDIX B

Supplementary information Chapter V

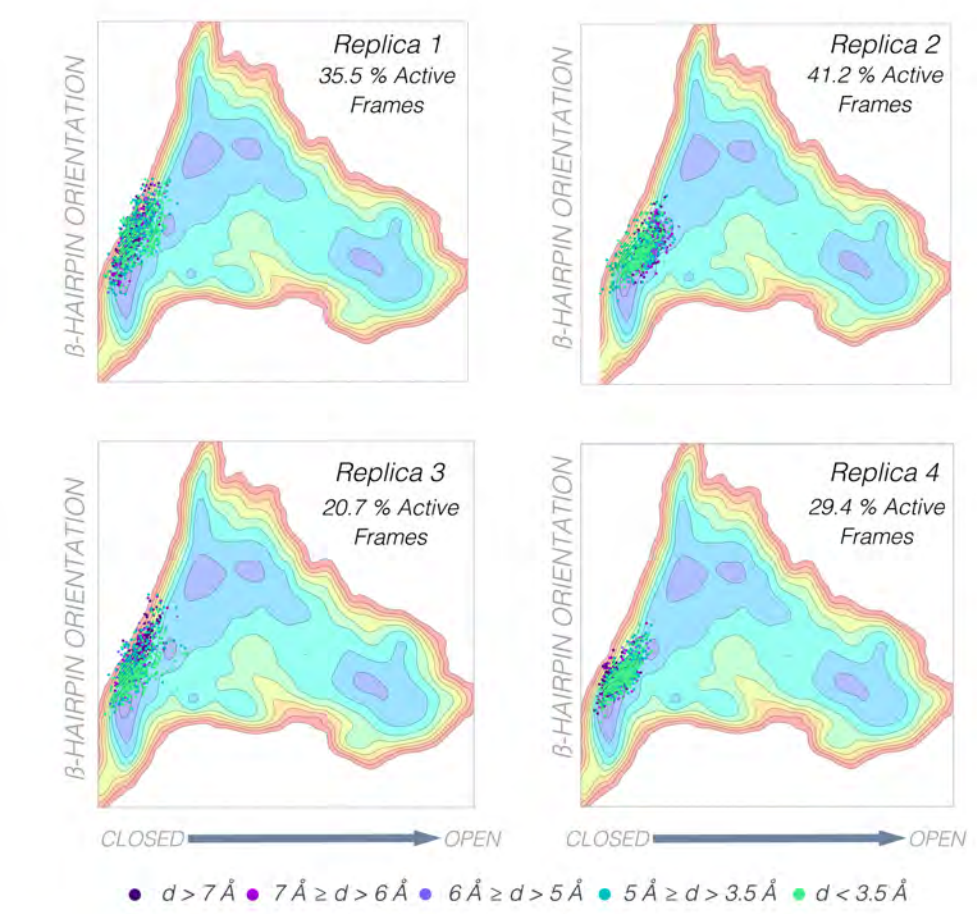


Figure B1. Projection of four replicas of aMD simulations of MAO-N D5 in a closed conformation and alpha-methylbenzylamine (AMBA) as a substrate into the corresponding free energy surface.

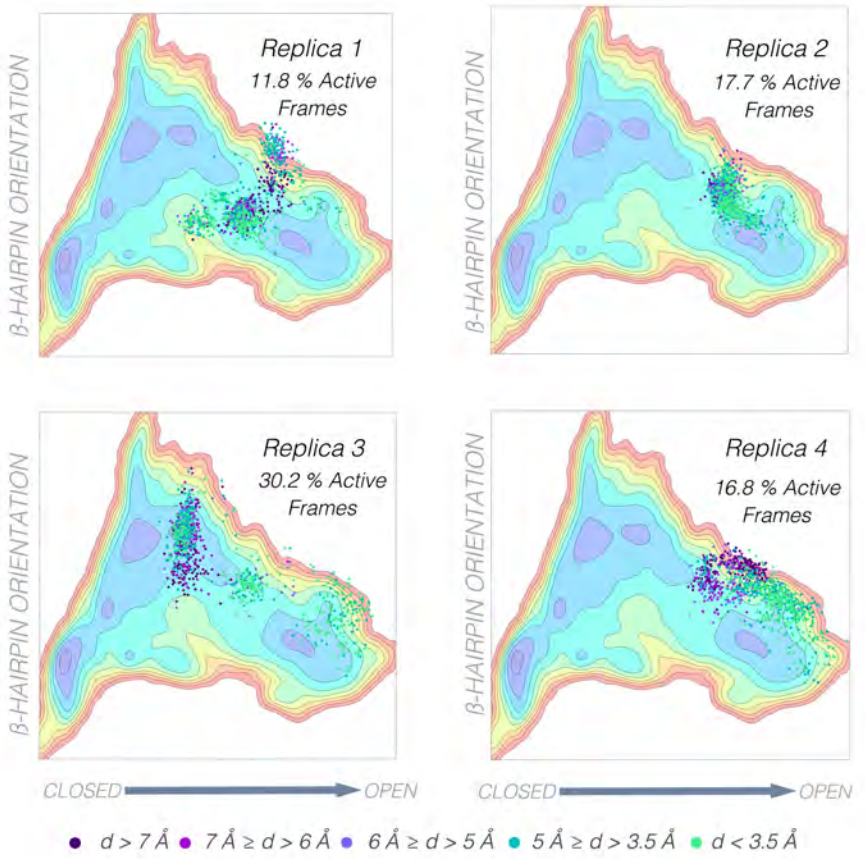


Figure B2. Projection of four replicas of aMD simulations of MAO-N D5 in an open conformation and alpha-methylbenzylamine (AMBA) as a substrate into the corresponding free energy surface.

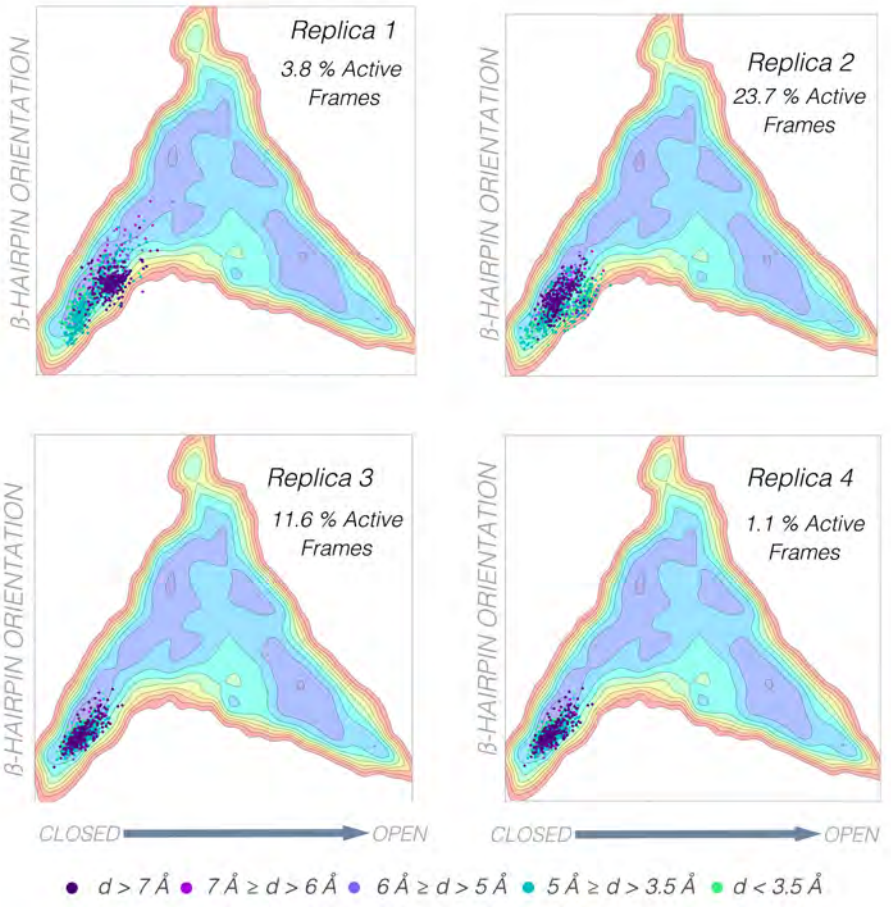


Figure B3. Projection of four replicas of aMD simulations of MAO-N WT in a closed conformation and hexylamine (HEX) as a substrate into the corresponding free energy surface.

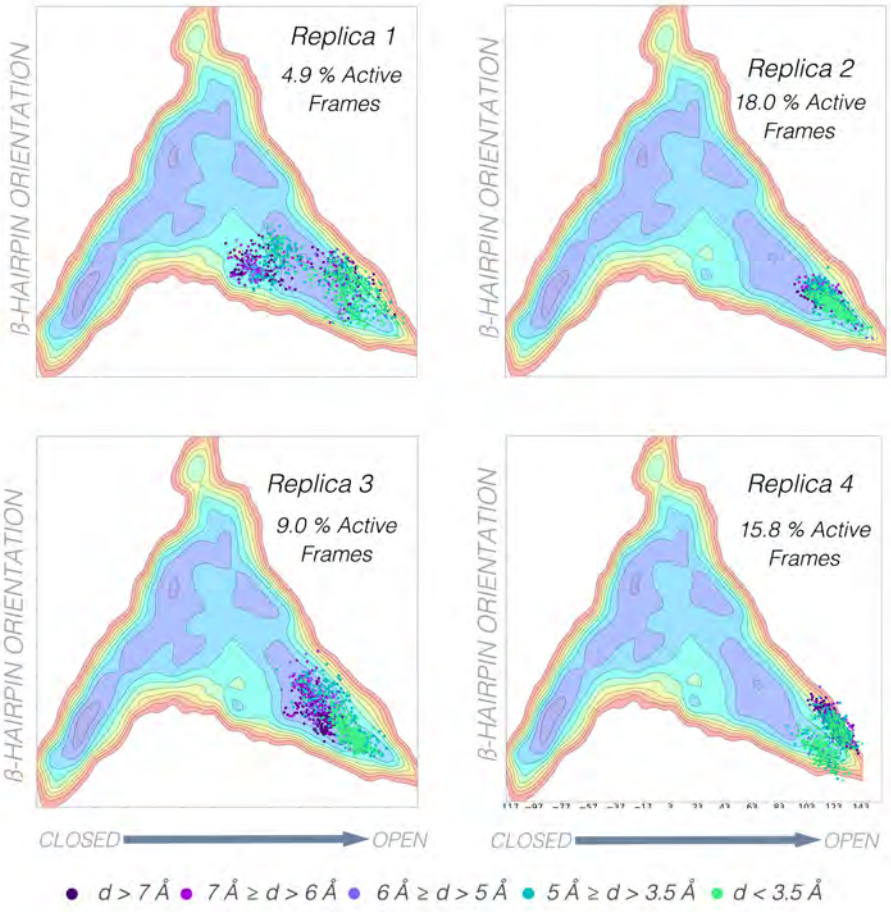


Figure B4. Projection of four replicas of aMD simulations of MAO-N WT in an open conformation and hexylamine (HEX) as a substrate into the corresponding free energy surface.