Universidade Federal de Minas Gerais
Instituto de Ciências Exatas - Departamento de Química
Programa de Pós-Graduação em Química

THEORETICAL INVESTIGATION OF THE PICTET-SPENGLER REACTION BETWEEN DOPAMINE AND *(S)*-CITRONELLAL CATALYZED BY THE ENZYME *(S)*-NORCOCLAURINE SYNTHASE

Aline de Souza Bozzi

Belo Horizonte, Brasil
2021

UFMG/ICEX/DQ. 1.461
D. 793

Aline de Souza Bozzi

THEORETICAL INVESTIGATION OF THE PICTET-SPENGLER REACTION BETWEEN DOPAMINE AND *(S)*-CITRONELLAL CATALYZED BY THE ENZYME *(S)*-NORCOCLAURINE SYNTHASE

Dissertação apresentada ao Departamento de Química do Instituto de Ciências Exatas da Univesidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Química.

Orientador: Willian Ricardo Rocha
Coorientador: Adolfo Henrique de Moraes Silva

Belo Horizonte, Brazil
2021

Ficha Catalográfica

B662t
2021
D

Bozzi, Aline de Souza
Theoretical investigation of the Pictet-Spengler reaction between dopamine and (S)-citronellal catalyzed by the enzyme (S)-norcoclaurine synthase / Aline de Souza Bozzi. 2021.
[xxviii], 92 f. : il., gráfs., tabs.

Orientador: Willian Ricardo Rocha.
Coorientador: Adolfo Henrique de Moraes Silva.

Dissertação (mestrado) – Universidade Federal de Minas Gerais – Departamento de Química.
Inclui bibliografia.

1. Físico-química – Teses. 2. Enzimas – Teses. 3. Dopamina – Teses. 4. Aldeídos – Teses. 5. Cetonas – Teses. 6. Funcionais de densidade – Teses. 7. Dinâmica molecular – Teses. 8. Mecanismos de reação (Química) – Teses. I. Rocha, Willian Ricardo, Orientador. II. Silva, Adolfo Henrique de Moraes, Coorientador. III. Título.

CDU 043

Elaborada por Sérgio Ferreira da Silva – CRB6-2719.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

UFMG
Programa de Pós-Graduação em Química
Departamento de Química - ICEx

química
pós-graduação-ufmg

"Investigação Teórica da Reação de Pictet-spengler Entre A Dopamina e O S-citronelal Catalisada Pela Enzima Norcoclaurina Sintase"

Aline de Souza Bozzi

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. Willian Ricardo Rocha - Orientador
UFMG

Prof. Adolfo Henrique de Moraes Silva - Coorientador
UFMG

Prof. Gabriel Heerdt
UFMG

Prof. João Paulo Ataide Martins
UFMG

Belo Horizonte, 08 de setembro de 2021.

Documento assinado eletronicamente por **Willian Ricardo Rocha**, **Professor do Magistério Superior**, em 08/09/2021, às 11:34, conforme horário oficial de Brasília, com fundamento no art. 5º do

Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Adolfo Henrique de Moraes Silva**, **Professor do Magistério Superior**, em 08/09/2021, às 11:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Gabriel Heerdt**, **Professor do Magistério Superior**, em 08/09/2021, às 12:36, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **João Paulo Ataide Martins**, **Professor do Magistério Superior**, em 08/09/2021, às 15:27, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0925897** e o código CRC **D21B532D**.

Referência: Processo nº 23072.244645/2021-47 SEI nº 0925897

To my beloved father

Agradecimentos

O mestrado é um caminho curto e árduo e, por vezes, parece solitário. Não se engane, para cada dificuldade há sempre quem ofereca apoio. Portanto, gostaria de agradecer a cada um que deixou sua marca ao longo desse tempo.

Ao meu pai, pelo apoio incondicional. Apesar de nem sempre compreender em minúcias o que eu escolhi fazer, sempre me apoiou e confiou nas minhas decisões. Além disso, foi meu parceiro de pandemia em tempo integral e sempre fez de tudo para que a convivência fosse a mais tranquila e harmoniosa possível, sempre cuidando de mim.

Aos meus orientadores, Prof. Willian e Prof. Adolfo, sem os quais este trabalho jamais teria tomado forma. Agradeço imensamente a confiança, o incentivo e a paciência de ambos. Tudo que aprendi, não apenas cientificamente, será levado e estimado por todos os próximos anos da minha carreira, que espero serem muitos.

Aos colegas dos laboratórios $\langle e|Cs|Mo\rangle^{Lab}$ e MacroMol, em especial Samara, Diego e Philipe, que viveram altos e baixos comigo desde o início. Foram muitos cafés, conversas filosóficas, políticas, antropológicas e científicas, reclamações, desabafos, "puxões de orelha", *happy hours* e jogatinas. Nao sei mensurar a importância dessas pessoas nesse período.

Aos colaboradores Marcelo e Vinícius, que me ensinaram a executar as simulações de dinâmica molecular e docking molecular com muita dedicação.

Aos amigos Nicolas e Ana, que continuam sempre comigo há anos não importando a distância nem as circunstâncias, e também ao Raposo, amigo adquirido na disciplina de RMN e que virou minha dupla da "bot lane", além de ter ótimo gosto musical.

Ao Diego, pelo carinho, paciência e incentivo.

Por fim, agradeço às agências brasileiras de fomento à pesquisa, em especial à CAPES.

Scientia vincere tenebras

Resumo

A enzima Norcoclaurina Sintase é conhecida por ser a Pictet-Spenglerase responsável por catalisar a condensação da dopamina e do 4-hidroxifenilacetaldeído, levando à formação da s-norcoclaurina, o primeiro metabólito na biossíntese dos alcalóides benzilisoquinolínicos. Recentemente, estudos mostraram que a Norcoclaurina Sintase possui alta promiscuidade frente a muitos aldeídos e cetonas, despertando grande interesse para seu uso. Neste trabalho, investigamos computacionalmente os aspectos estruturais dessa enzima em relação ao substrato *(S)*citronelal, um aldeído alifático de cadeia longa que contrasta com o aldeído natural, o 4-hidroxifenilacetaldeído. Além disso, investigamos também o mecanismo de reação de condensação entre a dopamina e o *(S)*citronelal catalisado pela Norcoclaurina Sintase. Para tanto, empregamos metodologias de ancoramento molecular, simulação de dinâmica molecular clássica e teoria do funcional da densidade. Demonstramos através do ancoramento molecular e de simulações de dinâmica molecular que o modo de ligação preferencial enzima-substrato apresenta a dopamina mais profundamente ancorada no sítio ativo, interagindo com o amino ácido LYS122 em uma conformação considerada ativa, o que significa que irá favorecer a ocorrência da reação, enquanto que o *(S)*citronelal ocupa a entrada da cavidade, tendo uma parte de sua cadeia exposta ao solvente, o que pode ser um indicativo da promiscuidade da enzima em relação aos compostos carbonilados. Do ponto de vista mecanístico, utilizando a teoria do funcional da densidade e a metodologia de cluster, mostramos que o caminho da reação possui três intermediários a menos em relação ao previamente publicado. Além disso, também descobrimos que as restrições conformacionais experimentadas pelos substratos alteram consideravelmente a barreira de energia de ativação para a etapa de ciclização.

Palavras-chave: Norcoclaurina Sintase, Dopamina, *(S)*citronelal, Ancoramento Molecular, Simulação de Dinâmica Molecular Clássica, Teoria do Funcional da Densidade, Mecanismo de Reação.

Abstract

The Norcoclaurine Synthase enzyme is known to be the Pictet-Spenglerase responsible for catalyzing the condensation of dopamine and 4-hydroxyphenylacetaldehyde, leading to the formation of s-norcoclaurine, the first metabolite in the biosynthesis of benzylisoquinoline alkaloids. Recently, the Norcoclaurine Synthase has shown great promiscuity towards many aldehyde and ketone substrates, raising great interest in further understandings. In this work, we computationally investigate the structural aspects of this enzyme regarded to the substrate *(S)*-citronellal, a long-chain aliphatic aldehyde that contrasts with the natural aldehyde, 4-hydroxyphenylacetaldehyde. Furthermore, we also investigated the mechanism of a condensation reaction between dopamine and *(S)*citronellal catalyzed by the Norcoclaurina Synthase. For that, we employed molecular docking methodologies, classical molecular dynamics simulation, and density functional theory. We demonstrate through molecular docking and molecular dynamics simulations that the preferential enzyme-substrate binding mode presents the dopamine more deeply anchored in the active site, interacting with the amino acid LYS122 in a conformation considered active, meaning it will favor the occurrence of the reaction. Also, the *(S)*citronellal occupies the cavity entrance, having part of its chain solvent-exposed, which may be indicative of the promiscuity of the enzyme concerning the carbonylated compounds. From a mechanistic aspect, through density functional theory and cluster methodology, we showed that the reaction path has three fewer intermediates than the previously published one. Furthermore, we also found that the conformational constraints experienced by the substrates considerably alter the activation energy barrier for the cyclization step.

Keywords: Norcoclaurine Synthase, Dopamine, *(S)*citronellal, Molecular Docking, Molecular Dynamics, Density Functional Theory, Reaction Mechanism.

Contents

List of Figures

Acronyms

4-HPAA	4-hydroxyphenylacetaldehyde
BIAs	Benzylisoquinolines
BO	Born-Oppenheimer
DFT	Density Functional Theory
ES	Enzyme-substrate
EP	Enzyme-product
FF	Force-field
GA	Genetic-algorithm
HK	Hohenberg-Kohn
KS	Kohn-Sham
LJ	Lennard-Jones
MD	Molecular Dynamics
NCS	Norcoclaurine Synthase
NMR	Nuclear Magnetic Resonance
PS	Pictet-Spengler
PSases	Pictet-Spenglerases
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
THBC	Tetrahydro-β-carboline
THIQ	Tetrahydroisoquinoline

TS	Transition State
vdW	van der Waals
WFT	Wavefunction theory

Notation

General notation

a	Italic lower case letters denote scalars
$\boldsymbol{a}$	Boldface italic lower case letters denote vectors
$\boldsymbol{A}$	Boldface italic upper case letters denote matrices

Symbols and operators

$\mathbb{N}$	Set of natural numbers
$\mathbb{I}$	Set of integer numbers
$\mathbb{R}$	Set of real numbers

1

Introduction

Benzylisoquinoline alkaloids (BIAs) form a diverse class of secondary metabolites containing nitrogen and produced in many plants (Figure 1.1 A). Although the physiological roles of most BIAs have not been unveiled yet, a handful of them display pharmacological properties and have been used for medicine throughout centuries (Hagel and Facchini, 2013). Some of these compounds are well-known, such as the analgesic morphine (Ziegler et al., 2009), the antimicrobial berberine (Jabbarzadeh Kaboli et al., 2014), and the anticancer noscapine (Winzer et al., 2012). These compounds can be found as natural products arising from biosynthesis via complex enzymatic pathways, which have the amino acid (L)-tyrosine as a precursor (Hagel and Facchini, 2013).

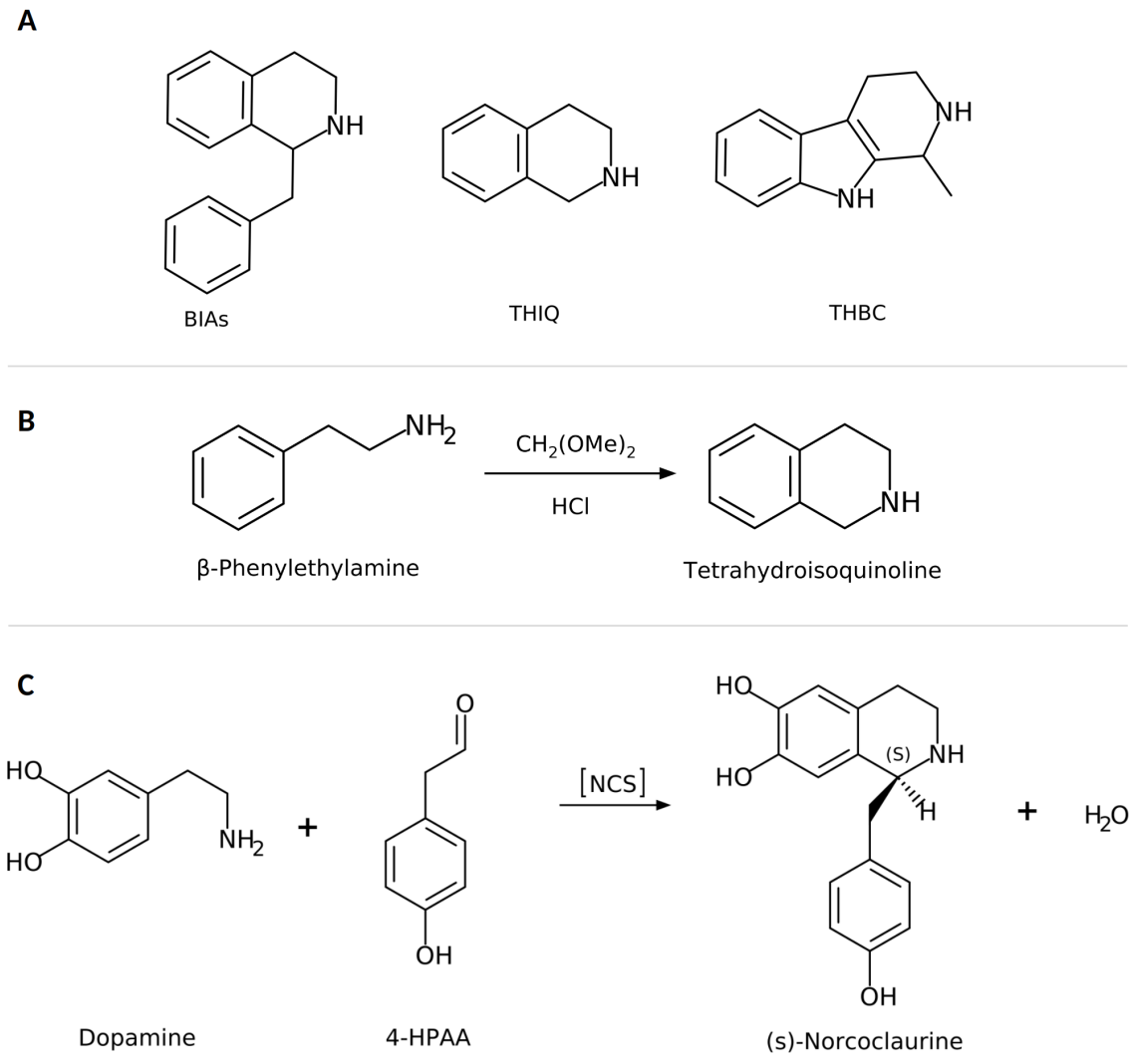


Figure 1.1: (A) The scaffold of some important alkaloids: benzylisoquinolines alkaloids (BIAs), tetrahydroisoquinolines (THIQ), and tetrahydro-β-carbolines (THBC), respectively. (B) The first PS reaction carried out by Pictet and Spengler in 1911. (C) The biocatalytic reaction between the dopamine and the 4-HPAA to form the (s)-norcoclaurine.

All BIAs have the molecule (s)-norcoclaurine as the first metabolite in their biosynthetic pathway, which is the product of the Pictet-Spengler (PS) reaction between dopamine

and 4-hydroxyphenylacetaldehyde (4-HPAA) (Samanani and Facchini, 2002). The PS reaction was first carried out in 1911 by Pictet and Spengler (1911), who reported the β-phenylethylamine condensation with dimethoxymethane following an acid-catalyzed ring closure. After more than a century, the (PS) reaction is still consecrated as one of the most efficient synthetic methods for obtaining pharmacophores, such as tetrahydroisoquinolines (THIQs) and tetrahydro-β-carbolines (THBCs) (Calcaterra et al., 2020; Stöckigt et al., 2011). The general structure of some alkaloids as well as the PS reactions can be seen in Figure 1.1.

Mechanistically, a classic PS reaction initiates with the formation of an iminium cation from the amine and aldehyde (or ketone) reactants generating a stereogenic center, then it undergoes an intramolecular nucleophilic attack of the aromatic ring providing a ring closure. The aromatization is subsequently restored after releasing a proton, as shown in Figure 1.2 (Stöckigt et al., 2011).

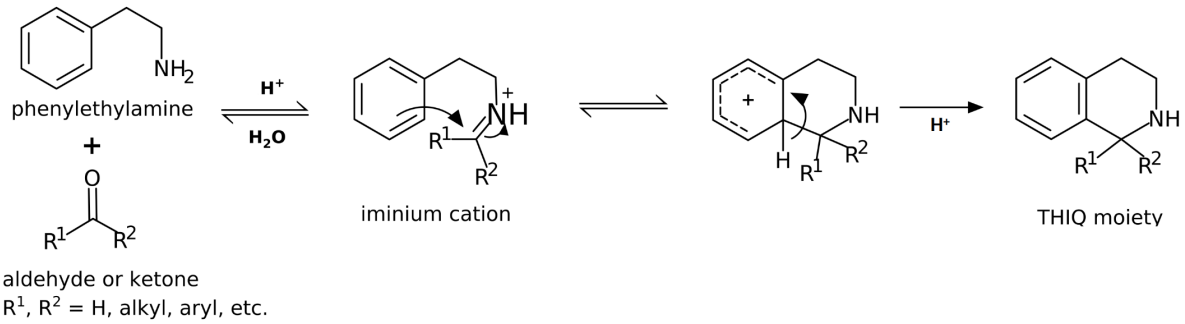
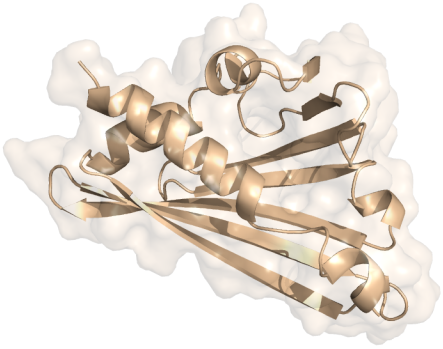


Figure 1.2: The proposed PS reaction mechanism.

More than 60 years after the first PS reaction was reported, the plant enzyme *Strictosidine Synthase* was discovered to catalyze a PS reaction in nature (Stöckigt and Zenk, 1977) being responsible for the indole alkaloid biosynthesis. Later on, it was observed that several other enzymes presented the same catalytic pattern (Matsumura et al., 2013; Chen et al., 2013), being named Pictet-Spenglerases (PSases).

The Norcoclaurine Synthase (NCS; EC 4.2.1.78) is the PSase responsible for producing the (s)-norcoclaurine in plants such as *Thalictrum flavum* (*Tf*NCS) (Samanani and Facchini, 2002), *Coptis japonica* (*Cj*NCS) (Liscombe et al., 2005), and *Papaver somniferum* (*Ps*NCS) (Minami et al., 2007). The *Tf*NCS and the *Thalictrum flavum* are shown in Figure 1.3.



Source: Online atlas of the british and irish flora. ©L. Rooney.

Figure 1.3: (a) The NCS from *Thalictrum flavum*. (b) The plant *Thalictrum flavum*

Recent studies have shown that the *Tf*NCS is capable of converting a broad scope of aldehyde and ketone substrates (Ruff et al., 2012; Nishihachijo et al., 2014; Lichman et al., 2015b, 2017a; Zhao et al., 2018; Roddan et al., 2019, 2020a; Wang et al., 2019; Zhao et al., 2021). The enzyme, however, shows little flexibility when it comes to changes in the dopamine skeleton. It has been suggested that the structure of dopamine plays a crucial role in the reaction, especially the metahydroxyl moiety. In this sense, several structural investigation took place in order to unveil the characteristics of this particular system. Much interest comes to the understanding of the contributions of the amino acids involved actively in the reaction, since some mutations in the wild-type NCS have been able to convert new carbonylated substrates.

In addition to structural investigation, the catalytic reaction mechanism also raises important discussion. Initially, based on nuclear magnetic resonance (NMR) and a co-crystallized structure with a nonsubstrate benzaldehyde (PDB ID 2VQ5), it was proposed the so-called "Aldehyde-first" mechanism (Berkner et al., 2008; Ilari et al., 2009), in which the aldehyde binds first to the active site. However, this proposition does not account for the wide scope of carbonylated molecules reported to be active for the NCS. More recent docking simulations and structural characterization of NCS bound to the iminium intermediate mimic (PDB ID 5NON) have provided new piece of information being the base for the second mechanistic proposal, the "Dopamine-first" mechanism (Lichman et al., 2015a, 2017a), in which the dopamine molecule appears to bind *a priori* into the cavity. Computational methods using the natural substrates dopamine and 4-HPAA were employed in order to shed light on this matter, and the results reinforce the "Dopamine-first" mechanism (Sheng and Himo, 2019), although both binding priorities are energetically plausible to happen.

In general, the PSases have been extensively studied in the past few years, especially the NCS. However, there is still much room for investigation. Roddan et al. (2020b) highlights the versatility of the PSases in the production of new stereoselective BIAs and foresees the importance of enzyme engineering based upon ligand-bound crystal structures as a manner of widening the substrate scope for this particular class of enzymes. In this work, we per-

formed a theoretical description of the structure and mechanistic investigation of the biocatalyzed reaction between the dopamine and the non-natural substrate *(S)*citronellal, forming the (1S)–1–[(2S)–2,6–dimethylhept–5–en–1–yl]–1,2,3,4–tetrahydroisoquinoline–6,7–diol ((S,S)–DMH-THIQ) as shown in Figure 1.4. The *(S)*citronellal has been experimentally tested and reported to be active for the NCS (Lichman et al., 2015a). This ligand possesses a long aliphatic chain, which is chemically quite different from the 4-HPAA, making it a good start to rationalize the promiscuity of the NCS as well as understand the effects of some mutations.

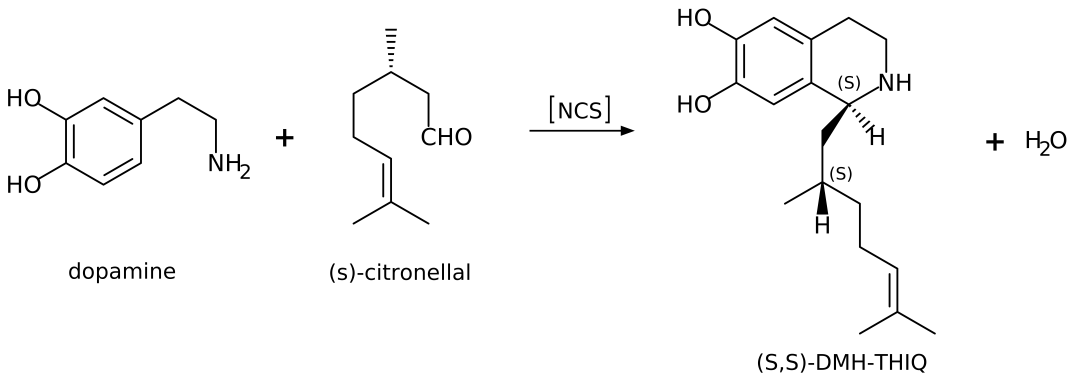


Figure 1.4: The Pictet-Spengler reaction between dopamine and (s)-citronellal catalyzed by the enzyme NCS forming the (S,S)–DMH-THIQ.

2

Theoretical Foundations

In computational modeling, we use mathematical tools and physical approximations to investigate the most varied molecular systems. It goes from atoms to macro-molecules, applying from molecular mechanics to *ab initio* quantum theory methods. Therefore, this chapter aims to glimpse into the vast world of molecular modeling of large systems, going from classical mechanics to quantum chemistry in a progressive way. The purpose here is not to derive all mathematical steps or to give a detailed description of theories, but to provide some understanding of the methods applied to the investigation of the structure and mechanism of the NCS and its substrates.

2.1 Molecular Docking

An overview

Molecular docking is a computational method to predict the structure of a receptor-ligand complex, usually being the ligand small molecules and the receptor, proteins (Kitchen et al. (2004); Bitencourt-Ferreira and de Azevedo (2019); Santana Azevedo et al. (2013)). In general, the docking simulations are used for two main purposes: correctly predicting the activity of ligands, and accurately modeling the overall structure. In our work, the molecular docking methodology was applied to predict the substrate binding modes.
The molecular docking simulations start by posing the ligand into the active site. The algorithm can follow three different approaches: the complete rigid docking, which considers the lock and key model featuring both receptor and ligand with fixed bonds and angles; the rigid receptor and flexible ligand, which is the most used approach by the available docking programs; and the flexible docking, considering both macromolecule and ligand flexible with respect to their angles and dihedrals.
Furthermore, it is also important to consider the representation of the receptor and small molecules, whether they are atomic, surface, or grid (Halperin et al., 2002). The atomic representation is usually used in the final step (ranking) of the docking process, using a potential energy function. The surface-based docking is usually used for protein-protein interaction. The method is characterized for aligning surface points and minimize the angle between the surface of opposing molecules (Norel et al., 1999). The potential energy grid approach consists in storing the receptor's energy information at each point of the grid (Goodford, 1985). The simplest grid method accounts for two typer of potential, the pairwise summation of Coulombic interactions and the Lennard-Jones (LJ) potential to describe the van der Waals (vdW) interactions. The terms are,

$$U_{ele}(r) = \sum_{i=1}^{N_A} \sum_{i=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.1}$$

where N is the number of atoms in molecules A and B, and q is the charge on each atom.

$$U_{vdW}(r) = \sum_{i=1}^{N}\sum_{j=1}^{N} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \tag{2.2}$$

being ϵ the energy depth of the potential and σ, the equilibrium distance between the atoms.

The molecular docking simulations also rely on algorithms for pose searching, considering the ligand flexibility. The main search methods are: systematic methods (databases and conformational search), the stochastic methods (Monte Carlo and genetic algorithm (GA)), and simulation methods (classical molecular dynamics) (Brooijmans and Kuntz, 2003).

Scoring Functions

The molecular docking simulations are complemented by scoring functions. The scoring functions are designed to evaluate the interactions between receptor and ligand in the poses predicted by the docking algorithm. In order to be computationally cheap, the scoring functions usually make simplifications in the evaluation of the models, leading to low accuracy for some phenomena. One of the limitations of most scoring functions is that they do not focus on entropy, but rather only on the enthalpic aspects of the binding interactions (Kollman, 1993; Simonson et al., 2002). The scoring functions can be divided into three groups: the force-field (FF), the empirical, and the knowledge-based.

The force-field-based scoring functions usually account for two energies, the receptor-ligand interaction and the ligand internal energy. The protein is considered rigid and, therefore, presents only one conformation, and the receptor internal energy is omitted from the calculations. The non-bonding terms are described by the Equations 2.1 and 2.2. Some recent formulations of FF-based scoring functions include a torsional entropy and a protein-ligand hydrogen bond terms(Kitchen et al., 2004).

The empirical scoring functions, as the name suggests, are constructed based on parameters determined experimentally. Therefore, these scoring functions fit experimental data and are built upon the idea that the binding energy is a summation of uncorrelated terms. This type of scoring function is somewhat simple, however, it depends on the availability of experimental data. The empirical scoring functions can include terms other than enthalpy, such as entropic penalties on bond flexibility due to binding(Kitchen et al., 2004).

And finally, the knowledge-based scoring functions is also built upon experimental data, however, focused on structure rather than the binding energies. The scoring function accounts for atomic interaction-pair potentials to model the receptor-ligant complex (Kitchen et al., 2004).

2.2 Classical Molecular Dynamics Simulation

The classical molecular dynamics simulation (MD simulation) method is based on molecular mechanics, where each atom is considered as punctual and its motion is governed by the forces of the other atoms in the system acting upon it (Santana Azevedo et al., 2013; Cramer, 2004). Therefore, a MD simulation consists of the numerical solution of the Newtonian equations of motion at each step,

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i \tag{2.3}$$

$$\boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i} U(\boldsymbol{r}^N) \tag{2.4}$$

where $\boldsymbol{f}_i$ is the force acting upon the atoms, $U(\boldsymbol{r}^N)$ is the potential energy accounting for the atoms interactions, and $\boldsymbol{r}^N = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ represents the complete set of 3N atomic coordinates. The potential energy $U(\boldsymbol{r}^N)$ is called Force-Field and contains a set of potentials of interaction that will describe total energy of the system (Cramer, 2004; Liu et al., 2016),

$$U(\boldsymbol{r}^N) = \sum_r U_{bond} + \sum_\theta U_{angle} + \sum_\varphi U_{dihedral} + \sum U_{non-bond} \tag{2.5}$$

where U_{bond} is the energy function for the bond stretching between two atoms, U_{angle} is the energy function for the angular deformation between two adjacent atoms, $U_{dihedral}$ represents the torsion or rotation around a given bond, and $U_{non-bond}$ account for the non-bond interactions, *e.i.*, van der Waals and electrostatic interactions.
The simplest molecular model to account for the term U_{bond} will be given by the harmonic form with determined equilibrium separation between two adjacent atoms, $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$. The same way, U_{angle} can be accounted for the angle θ_{ijk} between two successive bonds $r_i - r_j$ and $r_j - r_k$. The dihedral is defined by three bonds and four atoms, φ_{ijkl}, and the dihedral potential $U_{dihedral}$ will depend upon an expansion in periodic functions of order $m = 1, 2, \ldots, M$. The potential energies will then be,

$$U_{bond} = \frac{1}{2} \sum_{bond} k^r_{ij} (r_{ij} - r_{eq})^2 \tag{2.6}$$

$$U_{angle} = \frac{1}{2} \sum_{angle} k^\theta_{ijk} (\theta_{ijk} - \theta_{eq})^2 \tag{2.7}$$

$$U_{dihedral} = \frac{1}{2} \sum_{dihedral} \sum_m k^{\varphi,m}_{ijkl} (1 + cos(m\varphi_{ijkl} - \delta_m)) \tag{2.8}$$

being k^r_{ij}, k^θ_{ijk}, and $k^{\varphi,m}_{ijkl}$ the chemical bond stretch force, the angular deformation, and the potential of torsion constants, respectively. It is worth mentioning that models other than

the quadratic harmonic can be used to describe the bond stretching such as, for instance, the Morse potential. More robust models will better describe the potential shape. However, for small displacements, all functions well describe the potential behavior (Cramer, 2004). The non-bonded terms of the general potential account for the interactions between non-bonded atoms. A simplified, but popular, way to describe them is considering the the contributions of van der Waals interactions, U_{vdW}, which describes the attraction and repulsion between atoms, and the Coulomb potential, U_{ele}, which describes the electrostatic forces generated by charges (Schuster, 1983; Calais, 1990). The U_{vdW} can be described using the Lennard-Jones (LJ), as following,

$$U_{vdW} = \epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{2.9}$$

with two parameters, σ and ϵ, being the minimum equilibrium distance and the energy at the equilibrium, respectively. The U_{ele} term describing the interactions of punctual charges is,

$$U_{ele} = \frac{q_i q_j}{4\pi\epsilon_0 r} \tag{2.10}$$

where q_i and q_j are the charges and ϵ_0 is the permittivity of free space.

Some force-fields include other terms into its form to account for specific phenomena or to give a better description of the interactions. One example is the potential energy for coupling between bond stretching and bond rotation. In the AMBER force-field used in this work, the chemical bond stretching and angular deformation are calculated using the quadratic harmonic potential. The torsion or dihedral term is expanded as a Taylor series and truncated at the sixth term. As for the non-bond terms, the van der Waals interactions are approximated using the 12-6 LJ potential, and the Coulomb potential accounts for the charge interactions (Case et al., 2016).

Once the potential energy function $U(\boldsymbol{r}^N)$ is determined, the next step is to calculate the atomic forces taking the first derivative of the potential, using Equation 2.4. Then, taking Equation 2.3, it is possible to calculate the acceleration of the i-th particle of the system to a given time t (Durrant and McCammon, 2011). Through direct integration, the velocities and positions are also determined. Hence, the atomic momenta will be $\boldsymbol{p}^N = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N)$, which will lead to the total energy, or hamiltonian, H, as follows,

$$K(\boldsymbol{p}^N) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} \tag{2.11}$$

$$H = K + U \tag{2.12}$$

where K is the kinetic energy. Applying this procedure iteratively, a trajectory will be constructed, which is the set of velocities and positions throughout a given simulation

time (Durrant and McCammon, 2011).

The analytic solution of the equations presented so far is only possible for a two independent particle system. As MD simulations are usually used for macro-molecules containing thousands of atoms, it is vital the use of numerical methods for solving the equations. Therefore, several algorithms account for the numerical integration of the equations such as the Verlet (Verlet, 1967, 1968), the Leap-frog algorithm (Eastwood et al., 1980), the velocity Verlet (Swope et al., 1982), and the Beeman algorithm (Beeman, 1976).

The Verlet algorithm uses both the position $\boldsymbol{r}_i$ and the acceleration $\boldsymbol{a}_i$ in a given time t and in the time $t - \delta t$ to calculate the properties for the time $t + \delta t$, where δt is the time step. The calculation proceeds as follows,

$$\boldsymbol{r}_i(t + \delta t) = 2\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t - \delta t) + \frac{\boldsymbol{f}_i(t)}{m}\delta t^2 \tag{2.13}$$

The integration time interval, δt, is a crucial parameter in the MD simulations. For biological systems, the higher vibrational frequencies are related to bond stretching of hydrogen atoms, oscillating at femtoseconds time scale. Hence, the δt chosen must be of the order of sub femtoseconds so it can account for all the bond vibration (A. de Castro et al., 2017).

As a final remark, the MD simulations are full of minutiae and need attention to other aspects the will not be described here. Suffice it to say these are related to the definition of the ensemble, temperature and pressure regulation, constraints, among others.

2.3 Density Functional Theory (DFT)

Before diving into the DFT world, it is wise to step back and recall some quantum mechanics foundations and have a glimpse of the wavefunction theory (WFT).

2.3.1 Elementary quantum mechanics

In quantum mechanics, all the information about a system is contained in the system's wavefunction $\Psi(\boldsymbol{r}, t)$. In the WFT, the wavefunction can be obtained from the Schrödinger equation given by (McQuarrie, 2008; Szabo and Ostlund, 1996; Cramer, 2004),

$$i\hbar\frac{\partial\Psi(\boldsymbol{r}, t)}{\partial t} = \mathcal{H}\Psi(\boldsymbol{r}, t) \tag{2.14}$$

where $\hbar$ and $\mathcal{H}$ are the reduced Planck constant and the Hamiltonian operator, respectively. The Hamiltonian operator delivers the total energy, E, of the system and, hence, is build upon the kinetic (T) and potential ($V(\boldsymbol{r}, t)$) energies, the latter varying depending on the system. Therefore, the Hamiltonian operator can be written as,

$$\mathcal{H} = -\frac{\hbar}{2m}\nabla^2 + V(\boldsymbol{r}, t) \tag{2.15}$$

where ∇^2 is the Laplacian operator $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$, and m is the particle mass. Hence, for a system containing M nuclei and N electrons, we can construct the non-relativistic time-independent Hamiltonian for the Born-Oppenheimer (BO) approximation in atomic units as follows (Szabo and Ostlund, 1996),

$$\mathcal{H}_{BO} = -\sum_{i}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{r_{iA}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2.16}$$

where i and j represent the electrons and A and B, the nuclei. In the same way, r_i and R_A are the coordinates for the i-th electron and A-th nucleus, respectively. The terms of Equation 2.16 are the electronic kinetic energy operator, the electron-nucleus attraction potential operator or external potential, the electron-electron repulsion potential operator, and the nuclear repulsion potential operator.

By the separation of variables method (McQuarrie, 2008), the general solution for the wavefunction $\Psi(\boldsymbol{r}, t)$ given by Equation 2.14 can be written as,

$$\Psi(\boldsymbol{r}, t) = \psi(\boldsymbol{r}) exp\left(-it\frac{E}{\hbar}\right) \tag{2.17}$$

and $\psi(\boldsymbol{r})$ is the time-independent wavefunction. Therefore, the $\mathcal{H}_{BO}$ operator (Equation 2.16) will act upon the spatial part of the wavefunction, $\psi(\boldsymbol{r})$, and the total ground state energy will be given by,

$$E_0 = \int \psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)\mathcal{H}_{BO}\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)dr_1 dr_2 \ldots dr_N = \langle\psi|\mathcal{H}_{BO}|\psi\rangle \tag{2.18}$$

2.3.2 DFT foundations

The intrinsic philosophy of DFT is to determine the energy of a molecular system through a physical *observable*, in this case, the electronic density, $\rho(\boldsymbol{r})$. In this approach, for a system composed of N electrons, the electronic density depends only upon the three spatial coordinates ($\boldsymbol{r}$) rather than $3N$ spatial and N spin coordinates, as for the WFT formalism (Parr and Yang, 1989; Cramer, 2004; Capelle, 2006). Therefore, $\rho(\boldsymbol{r})$ integrated over all space gives the total number of electrons,

$$N = \int \rho(\boldsymbol{r})d\boldsymbol{r} \tag{2.19}$$

moreover, the electronic density can be expressed as,

$$\rho(\boldsymbol{r}) = N\int \ldots \int \psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)dr_1 dr_2 \ldots dr_N \tag{2.20}$$

where $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ represents the solution for the ground state Hamiltonian $\mathcal{H}_{BO}$.

The Hohenberg-Kohn Theorems

Recalling the BO approximation (Equation 2.16), the $\mathcal{H}_{BO}$ is uniquely given by the number of electrons and the potential created by the nuclei, V_{ne}. Therefore, the electronic density, the wavefunction, and the ground-state energy will be determined only by these quantities. This can be proven by assuming the existence of two distinct external potentials, V_{ext} and V'_{ext} satisfying the same ground-state electronic density $\rho_0(\boldsymbol{r})$. Hence, two Hamiltonian operators arise naturally, $\mathcal{H}$ and $\mathcal{H}'$, with two wavefunctions, Ψ_0 and Ψ'_0, for the ground-state. The variational principle states that the expectation value of a Hamiltonian a applied to a wavefunction b must be higher than the ground-state energy of a. Therefore, using Ψ'_0 as an approximated wavefunction for the operator $\mathcal{H}$ (Hohenberg and Kohn, 1964),

$$\langle \Psi'_0 | \mathcal{H} | \Psi'_0 \rangle > E_0 \tag{2.21}$$

and rewritting this expression,

$$\langle \Psi'_0 | \mathcal{H}' | \Psi'_0 \rangle + \langle \Psi'_0 | \mathcal{H} - \mathcal{H}' | \Psi'_0 \rangle > E_0 \tag{2.22}$$

$$E'_0 + \langle \Psi'_0 | V_{ext} - V'_{ext} | \Psi'_0 \rangle > E_0 \tag{2.23}$$

$$E'_0 + \int \rho_0(\boldsymbol{r})(V_{ext} - V'_{ext}) dr > E_0 \tag{2.24}$$

undergoing the same path for a wavefunction Ψ_0 and a Hamiltonian $\mathcal{H}'$ and summing up them, one gets

$$E'_0 + E_0 > E'_0 + E_0 \tag{2.25}$$

which makes the assumption incorrect. Therefore, for the ground-state, there can be only one electronic density and nuclear potential correspondent. Hence, it is proven that the energy is a functional of the electronic density.
Making further use of the variational principle, taking an approximated electronic density, ρ', the energy given will be higher than the exact ground-state energy,

$$E_o[\rho'] \geq E_o[\rho] \tag{2.26}$$

The energy functional, $E[\rho]$, can be split into three parts: the kinetic energy, $T[\rho]$, the electron-nucleus interaction, $V_{ne}[\rho]$, and the electron-electron repulsion, $V_{ee}[\rho]$. Taking into account the Hartree-Fock energy, the electron-electron repulsion can be further divided into other two parts: the Coulomb, $J[\rho]$, and the exchange, $V_{ex}[\rho]$. The functionals $V_{ne}[\rho]$

and $J[\rho]$ are given by,

$$V_{ne}[\rho] = \sum\nolimits_A \frac{Z_A \rho(\boldsymbol{r})}{|\boldsymbol{R}_A - \boldsymbol{r}|} d\boldsymbol{r} \tag{2.27}$$

$$J[\rho] = \frac{1}{2} \int \frac{\rho(\boldsymbol{r})\rho'(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} d\boldsymbol{r}' \tag{2.28}$$

The Kohn-Sham Formalism

The WFT difficulty relies on the electron-electron interaction, which is treated in the DFT with the Kohn-Sham (KS) formalism (Kohn and Sham, 1965). The key is to start with a fictitious system of non-interacting electrons that have the same ground-state energy of a real interacting system. Then, the kinetic energy is divided into two terms: a exact one and a correction(Parr and Yang, 1989; Cramer, 2004). The exact solution for the Schrödinger equation is given by the Slater determinant orbitals ϕ_i, and the exact kinetic energy is,

$$T_0 = \sum_i^N \left\langle \phi_i \left| -\frac{1}{2}\nabla^2 \right| \phi_i \right\rangle \tag{2.29}$$

for interacting electrons (real system), the T_0 is an approximation that requires a correction, and the density for a Slater wavefunction is,

$$\rho(\boldsymbol{r}) = \sum_1^N |\phi_i(\boldsymbol{r})|^2 \tag{2.30}$$

The residual kinetic energy is included in the exchange-correlation energy, $E_{xc}[\rho]$ and the DFT energy can be written as,

$$E_{DFT}[\rho] = T_0[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.31}$$

The problem, here, is to find the adequate term of exchange-correlation energy. One way, is to use the Lagrange method,

$$L[\rho] = E_{DFT}[\rho] - \sum_{ij}^N \lambda_{ij}[\langle \phi_i | \phi_j \rangle - \delta_{ij}] \tag{2.32}$$

where λ_{ij} is the Lagrange multiplier and δ_{ij}, the Kronecker delta. The equation above will generate a set of one-electron operator, h_{KS}. Therefore,

$$h_{KS}\phi_i = \sum_{ij}^N \lambda_{ij}\phi_j \tag{2.33}$$

$$h_{KS} = -\frac{1}{2}\nabla^2 + V_{eff} \tag{2.34}$$

$$V_{eff}(\boldsymbol{r}) = V_{ne}(\boldsymbol{r}) + \int \frac{\rho'(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r}' + V_{xc}(\boldsymbol{r}) \tag{2.35}$$

we see that the effective potential, V_{eff} is given by the Coulomb potential, the interaction electron-nucleus, and a so-called functional derivative, V_{xc}. The set of orbitals produced through the Lagrangian multiplier culminates in the pseudoeigenvalue equations known as KS equations,

$$h_{KS}\phi_i = \varepsilon_i \phi_i \tag{2.36}$$

Therefore, the KS equations are not exact and the exchange-correlation term needs to be systematically constructed. In this sense, there are many approximations to determine it, *e.i.*, the Local Density Approximation (LDA), Generalized Gradient Approximation (GGA), hybrid functionals, among others (Cramer, 2004).

3

NCS: structural and conformational studies

3.1 Introduction

The versatility of catalyzing the formation of BIAs has raised great interest in the investigation of PSases structures. This is especially true for the *Tf*NCS, which has shown great promiscuity towards aldehydes and ketones. The NCS is a dimeric protein and each monomer shows a long tunnel of approximately 23 Å between its three α-helices and seven-stranded antiparallel sheets. The cavity entrance is formed and regulated by the residues GLU103, TYR108, TYR131, and TYR139. The active site comprises the amino acids LEU68, LEU72, LEU76, PHE80, LEU95, MET97, TYR108, GLU110, PHE112, LYS122, ASP141, PRO179 (Figure 3.1 A and B) (Ilari et al., 2009; Sharma et al., 2018).

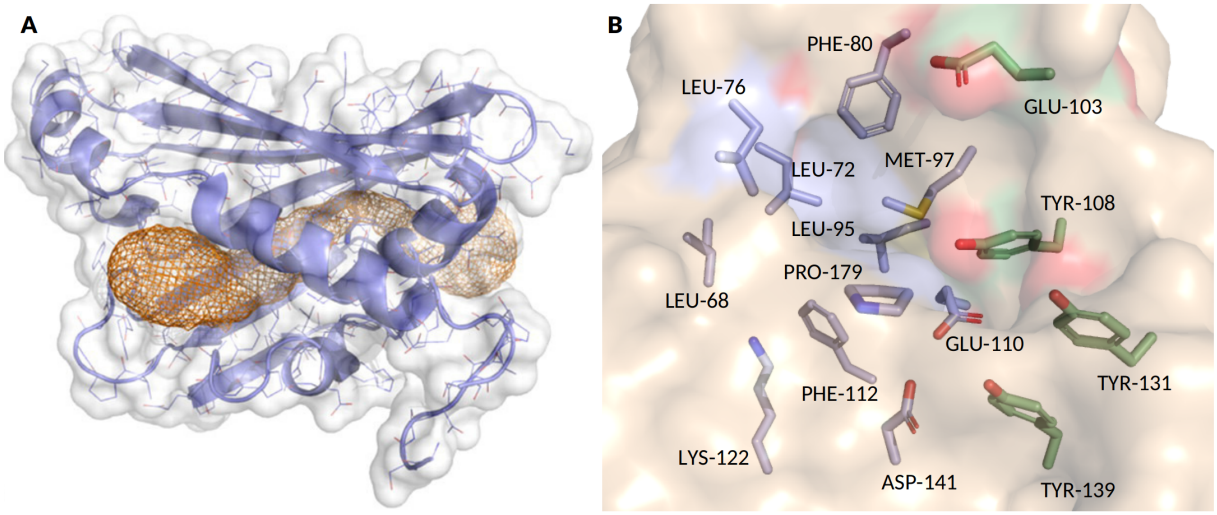


Figure 3.1: Structural characteristics of *Tf*NCS. (A) The long tunnel between the α-helices and antiparallel sheets. (B) In green, the amino acids responsible for regulate the cavity. In purple, the active site amino acid residues.

Source: Figure (A) reprinted with permission from Ilari et al. (2009). *Copyright © (2009). Journal of Biological Chemistry.*

Mutations on the NCS wild-type contributed to understanding some amino acid roles in the active site and broadening the carbonylate substrate scope. For example, the variant K122A completely abolished the enzyme's activity and shows that the LYS122 side-chain participates actively in the reaction (Ilari et al., 2009). The first successful crystal data for the NCS featuring the co-crystallized molecules dopamine and the inactive aldehyde 4-hydroxybenzaldehyde (PDB ID 2VQ5) suggested that the LYS122 side-chain nitrogen interacts with the carbonyl group of the aldehyde. Later on, another crystal structure containing an iminium intermediate mimic of the PS reaction (PDB ID 5NON (Lichman et al., 2015a)) showed that, in fact, the dopamine might bind prior to the active site, occupying a deeper spot into the cavity and its catechol moiety preferably interacts with the LYS122.

The two carboxylic acid residues GLU110 and ASP141 were also investigated. The

variants E110Q and E110D, which attribute a significant change in the charge pattern, also eliminated the enzyme activity. Therefore, it suggests that the GLU110 acts as a base in the reaction path. On the other hand, variants for the residue ASP141 did not inhibit product formation, implying that this residue acts more on the general electrostatic stabilization of the active site. Furthermore, it was also shown that the amino acid TYR108 plays a role in electrostatic stabilization. Also, the variant Y108F discloses the role of shaping the cavity entrance.

Finally, some mutations account for the carbonylate activity profile observed for the NCS. The L76A variant improved the enzymes's activity towards the aldehyde citronellal, which is known for its long aliphatic chain. The LEU76 is located more externally in the active site, reinforcing the binding mode proposed where the dopamine is buried deeper into the active site (Lichman et al., 2015a). More recently, α-methyl-substituted aldehydes (Roddan et al., 2019) were reported to be active for the M97V *Tf*NCS variant, which has also shown promising results for the production of (1S)-aryl-THIQs (Roddan et al., 2020a). Ketones have also been accepted as substrates leading to the production of 1,1'-disubstituted THIQs (Lichman et al., 2017b).

In this chapter, we investigate the molecular dynamics and analyze the conformational structure of the free NCS and the enzyme-substrate complex formed between the NCS and the substrates dopamine and *(S)*-citronellal. We aimed to understand how the overall structure of this enzyme is affected by a non-natural substrate and map possible productive binding modes to investigate the mechanism of reaction.

3.2 Methodology

3.2.1 Molecular Docking

System Preparation

The structure of the enzyme NCS (resolution of 1.85 Å (Lichman et al., 2017a)) used in this work was obtained from the Protein Data Bank (`https://www.rcsb.org/`) under the PDB ID 5NON. The file has the coordinates of three chains A, B, and C, each co-crystallized to a non-substrate that mimics one of the intermediates in the reaction path, which can acquire two different conformations. In this work, only chain A was used since it well superposes to the chains B and C (with root-mean-square deviation, RMSD, of 0.683 Å and 0.835 Å, respectively) (see Appendix A, Figure A.1). All but three crystallographic water molecules (201, 288, and 302) inside the active site were removed from the structure; missing hydrogen atoms were added using the software Discovery Studio Visualizer v.19.1.0.18287 (BIOVIA, 2012), estimating the protonation state of the amino acids at pH = 7. However, the protonation state of the side chain of the amino

acids LYS122 and GLU110 were manually changed from +1 to 0 and -1 to 0, respectively, as indicated by (Lichman et al., 2015a; Sheng and Himo, 2019).
The 3D structures of all compounds were generated using Avogadro 1.2.0 software (Hanwell et al., 2012). Geometry optimizations were carried out with the Open Babel package (O'Boyle et al., 2011) using MMFF94 force-field (Halgren, 1996a,b,c,d; Halgren and Nachbar, 1996).

Molecular Docking Simulation

The molecular docking simulations were carried out with GOLD 5.8.1 software (Jones et al. (1997)), using a grid of radius of 10 Å centered at the active site (coordinates $x = -18.19; y = 9.25; z = 22.12$), which is composed by the amino acids LEU68, LEU72, LEU76, PHE80, LEU95, MET97, TYR108, GLU110, PHE112, LYS122, ASP141, PRO179, and MET183. All remaining parameters regarding ligand flexibility were kept as default. The genetic algorithm (GA) was set to a maximum search efficiency with 200 GA runs per molecule. The conformations obtained by the GA were ranked according to the ChemPLP (Korb et al. (2009)) score function.
Redocking using the intermediate mimic co-crystallized with the NCS was performed to evaluate the experimental binding mode's predictability and validate the docking protocol. The RMSD calculation was used as the criteria to assess whether the simulation conditions were adequate. The results of each ligand were grouped in clusters of conformations that differ by a maximum of 1 Å from one another. Then, the best score ranking poses were selected based on the most representative clusters that reproduced the experimental data and redocking and were analyzed using Pymol v1.8 and Discovery Studio Visualizer v19.1.0.18287.
Three different molecular docking simulations were carried out and, from now on, they are referred to as docking simulations A, B, and C, respectively. For the first docking simulation, A, the aldehyde portion of the redocking pose was removed, and the dopaminium portion (the aminoethylcatechol moiety) was kept in the active site. Then, the substrate *(S)*-citronellal was docked in the system. For the docking simulation B, an iminium intermediate mimic was constructed similarly to the crystallized one, except the aldehyde portion of the molecule was formed by the aliphatic chain of the *(S)*-citronellal in this case. This molecule was submitted to the docking protocol, and the result was compared to the original iminium intermediate mimic (see Appendix A, Figure A.3). Then, the aldehyde portion of the molecule was removed, and the dopaminium moiety was kept in the cavity, similar to the prior procedure described. Next, the *(S)*-citronellal was docked in the system. And finally, the docking simulation C was carried out by docking the dopaminium ligand, followed by the *(S)*-citronellal.

3.2.2 Molecular Dynamics

System Preparation

The enzyme preparation for the molecular dynamics (MD) calculations follows the same protocol used for the molecular dockings, except that for MD, all the crystallographic water molecules were removed. The ligands dopaminium and *(S)*-citronellal used for the enzyme-substrate (ES) complex MD calculations originated from the molecular docking simulation analysis described above.

MD Simulation

Molecular dynamics simulations were performed with the AMBER TOOLS 20 package (Case et al. (2016)), using the ff14SB force field (Maier et al. (2015)) for NCS, and TIP3P (Jorgensen et al. (1983)) for water molecules. The parameters for dopamine and (s)-citronellal substrates were obtained from the General Amber Force Field, GAFF (Wang et al. (2004)), and the missing parameters were calculated with the ANTECHAMBER module (Wang et al. (2006)), together with the AM1-bcc charge model, used for point atomic charge calculations. The initial structures of the all systems, *i.e.*, the free NCS and the ES complexes, were solvated in a rectangular water box with a minimum distance of 10 Å between any atom and the box edges. Next, Na^+ ions were added to neutralize the system net charge. A minimization was performed in 5000 steps, 1000 steps using the steepest descendent algorithm, and 4000 steps using the conjugate gradient algorithm (Schmidt et al., 1993). In sequence, the system was heated from 5 to 300 K using ensemble NVT, where there was an alternation between heating and equilibration in a total of 6 steps. The electrostatic and long-range interactions were accounted using the Ewald particle mesh (PME) method (Darden et al. (1993)). Finally, the production step was simulated in an NPT ensemble, at a temperature of 300 K and a pressure of 1 bar, using periodic boundary conditions and regulated by the Berendsen isotopic barostat (Berendsen et al. (1984)). The NPT production step simulated a total of 100 ns for all molecular dynamics. Coordinates were saved at every 4 ps interval generating a total of 25,000 frames.
The calculations were performed using a 6-processor workstation Intel® Core™ i7-6800K, 3.60GHz base frequency, 16GB RAM, and equipped with NVIDIA GeForce GTX 1070 graphics cards.

MD Simulation Analysis

All molecular dynamics were analyzed in terms of RMSD, RMSF, and hydrogen bond distances using the cpptraj (Roe and Cheatham (2013)) and the parmed (Case et al., 2016) packages, the software VMD v.1.9.3 (Humphrey et al. (1996)), and python3 algorithms.

The 100 ns trajectories had their first 10 ns (10% of the total simulation time) cut off, and the analysis were done on the remaining 90 ns.

3.3 Results and Discussion

3.3.1 Molecular Docking

The redocking results for the iminium intermediate mimic presented an RMSD of 1.224 Å, which is a reasonably good value, considering that the methoxyphenyl portion of the molecule is solvent-exposed and, therefore, acquire a variety of conformational changes (see Appendix A, Figure A.2). The results also presented two conformation states regarded to the dopamine portion of the molecule, which is the same pattern observed in the crystallized ones. Lichman et al. (2017a) reported that the iminium intermediate mimic crystallized acquired two major binding modes into the active site, as can be seen in Figure 3.2. Mechanistically speaking, the binding mode represented in purple is considered active once the 3-OH of the catechol group is favorably interacting with the LYS122, while the binding mode in green is considered inhibitory as it shows the same interaction with LYS122 but through the 4-OH. The contributions of the two orientations for the active and inhibitory binding modes in the crystal were estimated to be 70 and 30%, respectively. Amongst the 200 GA runs for the redocking simulation, 148 showed the active conformation against 52 for the inhibitory, *i.e.*, a representation of 74 and 26%, respectively, agreeing with the experimentally reported rate. This result was used in addition to the criteria mentioned in section 3.2.1 to indicate whether the docking protocol was adequate. The 2D contact map of the redocking pose (Figure 3.2 E) also shows the solvent-exposed methoxyphenyl end as a blue blur sphere and the catechol moiety buried deeper into the cavity and stabilized by intermolecular interactions with the amino acids of the first coordination sphere. The hydrogen bond between the 3-OH of catechol and the LYS122 occurs at a distance of 2.65 Å between the oxygen and the nitrogen. Also, there are three pi-alkyl interactions between the aromatic ring and the LEU72, LEU95, and VAL124 amino acids that assist in locking the dopamine position in site.

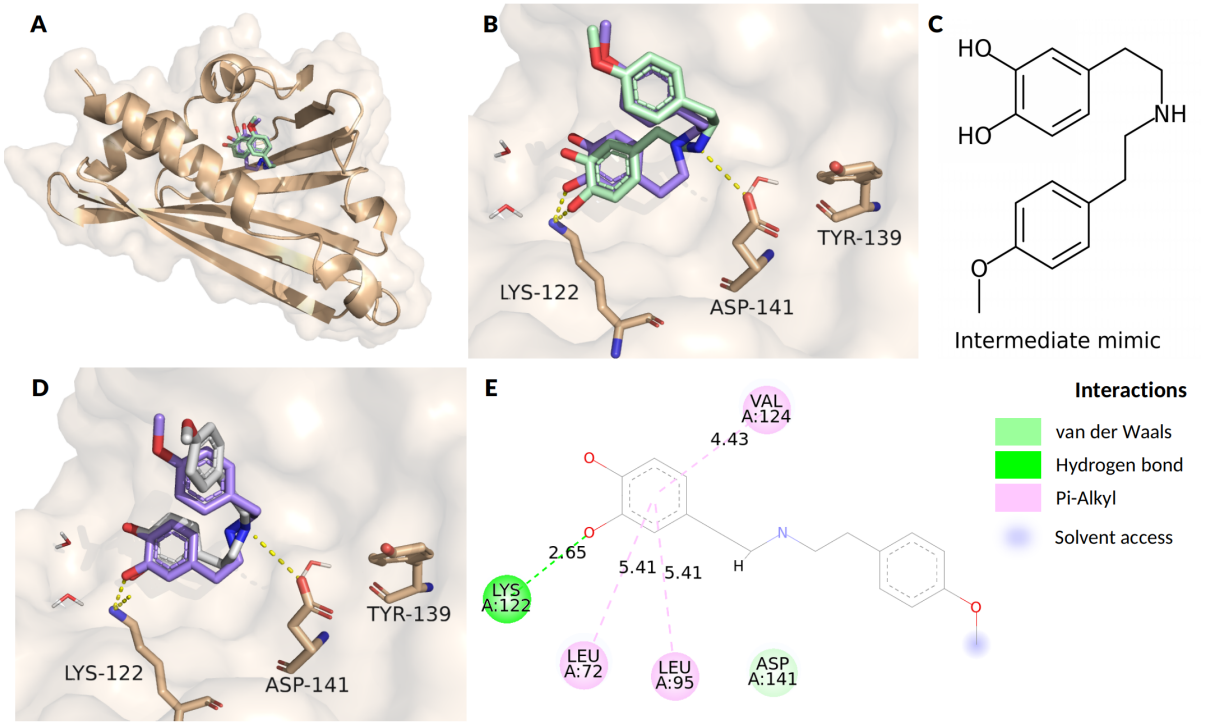


Figure 3.2: (A) The 3D structure of chain A of a *Tf*NCS (PDB ID 5NON) co-crystalized with the non-substrate iminium intermediate mimic molecule. (B) The active site of the NCS showing the intermediate mimic in its two major conformations, active (purple) and inactive (green). (C) The intermediate mimic 2D structure. (D) Comparison between the crystallized active conformation (purple) and the redocking (grey) of the intermediate mimic to the binding site. (E) The 2D contact map of the intermediate mimic redocking showing the relevant interactions between the ligand and the amino acids of the cavity.

As widely known, the NCS possesses great promiscuity regarding the aldehydes (or ketones) but not so much when it comes to amines (Ruff et al. (2012); Nishihachijo et al. (2014); Lichman et al. (2015b, 2017b); Zhao et al. (2018); Roddan et al. (2019); Wang et al. (2019)). It is suggested that this difference in promiscuity is intimately related to the binding priority of the two ligands. As the amine-carrier binds deeper into the cavity, undergoing several intermolecular interactions, it becomes less tolerable to changes in its overall structure. Hence, our amine-carrier substrate is kept the dopaminium. However, our aldehyde substrate is the *(S)*-citronellal, an aliphatic compound with distinct characteristics from the 4-HPAA. Therefore, the redocking results are essential because they guided the docking of our carbonyl-carrier ligand to find its expected binding modes, considering there is no crystallographic data to support this quest.

The molecular docking simulation A, B, and C results are shown in Figures 3.3 A, B, and C, respectively. In A, although the whole RMSD for this molecule is 1.224 Å, the RMSD for the amine moiety kept in the active site is 0.427 Å, reinforcing the aldehyde portion's freedom discussed. In B, the dopaminium coming from the intermediate mimic built with the *(S)*-citronellal chain presented an RMSD of 0.737 Å, comparable to the dopaminium

in A, and the active orientation of the compound crystallized. However, when comparing the rate of active to inhibitory orientations, the active represented only 50.5% of the total poses, while the inhibitory was 37%. The other 12.5% of the poses showed the aldehyde portion of the molecule deeper into the cavity, which could be interpreted as the aldehyde binding first to the active site. This effect was seen in previous crystallographic data (PDB ID 2VQ5 Ilari et al. (2009)), which led to the first idea for mechanisms in which the aldehyde would bind before the amine and interact with the LYS122. On the other hand, the docking C showed an incredibly high rate of 89:11, favoring the active pose, and an RMSD of 0.710 Å compared to the dopaminium portion of the crystallized compound.
As shown in the 2D diagrams in Figure 3.3, in all three docking simulations, the 3-OH of the catechol moiety interacted with the side chain of the amino acid LYS122 through a hydrogen bond. Also, the amino acids GLU110 and ASP141 interacted with the aminium ion *via* ion-dipole and attractive charges, respectively, showing significant interaction between this substrate and the catalytic site formed by the three amino acids. Furthermore, other residues of the cavity, such as LEU72, LEU95, and VAL124, interact with the non-polar portion of the dopaminium through their non-polar side chain, which can be seen in all three docking results. The aromatic ring also shows some other interactions as $\pi - \pi$ T-shaped with adjacent PHE112 and π-sulfur with MET183.

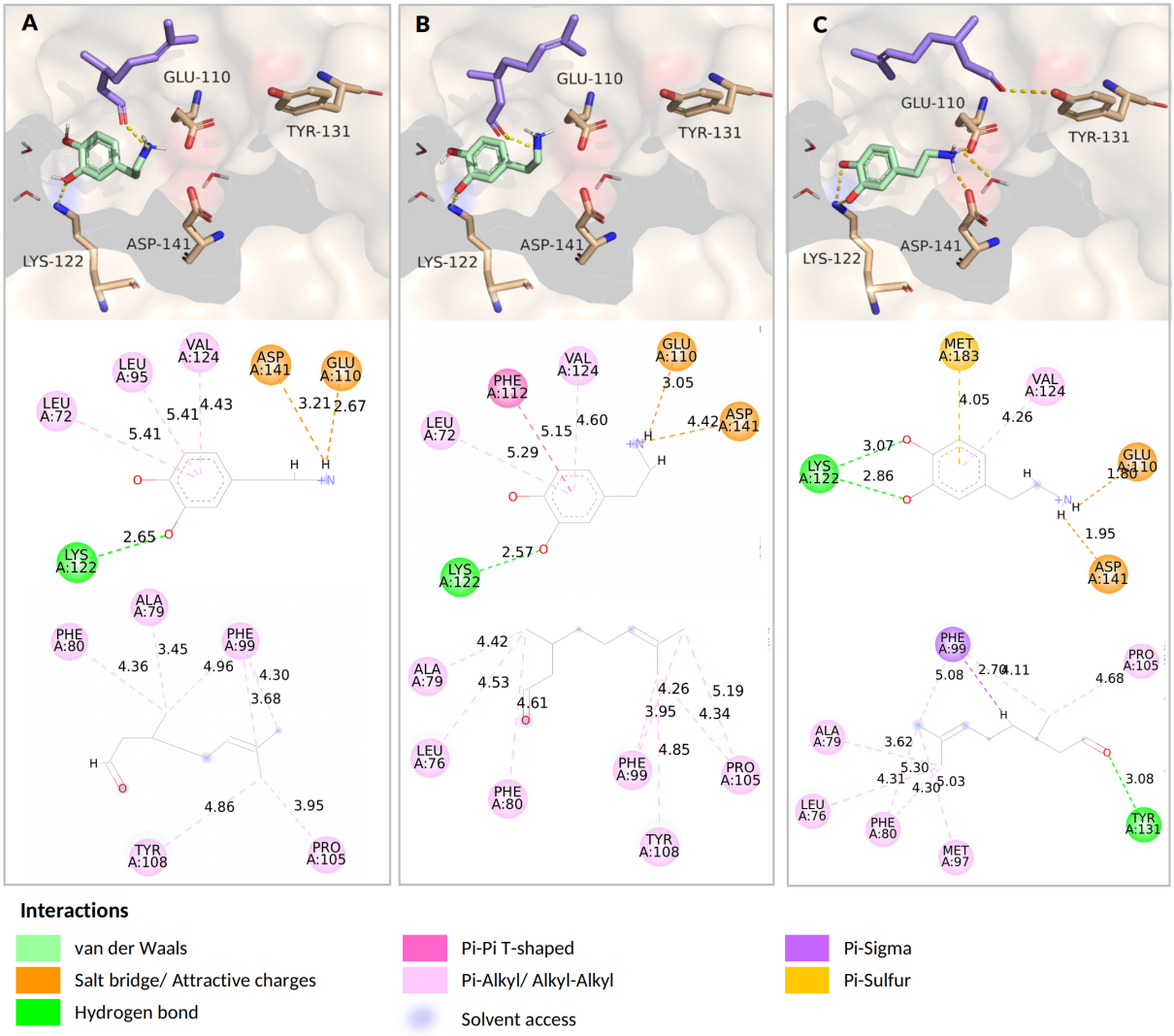


Figure 3.3: Docking simulations A, B, and C are represented in panels A, B, and C, respectively, indicating their pose results of dopaminium and *(S)*-citronellal and their 2D contact maps highlighting the relevant interactions between the substrates and the amino acid residues.

The *(S)*-citronellal always appears at the entrance of the cavity (Figure 3.4), and therefore, acquire different orientations, as can be seen in Figures 3.3 A, B, and C. These orientation changes culminate in a variety of intermolecular interactions, being in its majority between alkyl groups of the aldehyde and the amino acids. In docking C, two extra interactions can be seen in form of $\pi - \sigma$ with PHE99 and hydrogen bond between the oxygen of the carbonyl group of *(S)*-citronellal and the hydrogen of the side chain of TYR131. These two effects are only seen because the *(S)*-citronellal chain faces the opposite direction compared to dockings A and B.

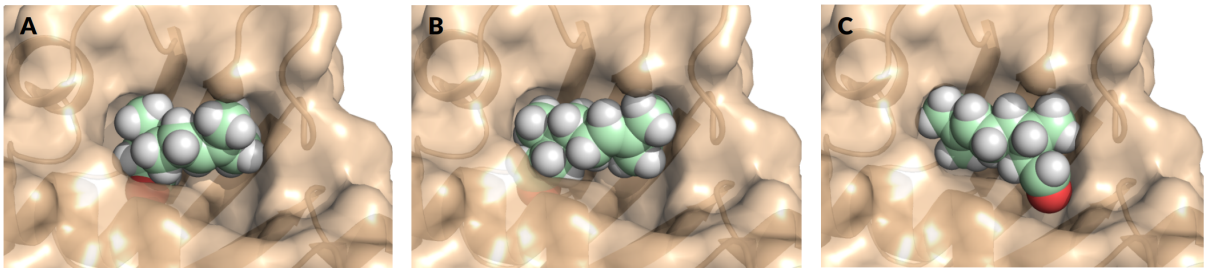


Figure 3.4: The *(S)*-citronellal fitting at the entrance of the active site of the NCS. The panels A, B, and C represent the docking results A, B, and C, respectively.

3.3.2 Molecular Dynamics

Molecular dynamics simulations were carried out for the free NCS and for the enzyme-substrate (ES) complex formed between the NCS and the dopaminium and *(S)*-citronellal ligands. The latter had their initial coordinates originated from the three molecular docking poses analyzed and discussed in section 3.3.1. Therefore, molecular docking poses A, B, and C correspond to MD simulations ES 1, ES 2, and ES 3, respectively.

The Cα RMSD (Å) values for the free enzyme were lower than 2 Å for the entire MD trajectory, showing the stability of the second and tertiary structures (Figure 3.5 A). In general, the RMSF (Å) for the free NCS shows a more significant variation at the C-terminus of the enzyme hitting up to 4.5 Å. Other considerable RMSF (Å) variations are seen in loops three and five, each containing three and seven amino acid residues, respectively (Figures 3.5 B and C). These loops are positioned at the surroundings of the cavity entrance, and loop five was previously observed to participate in the regulation of the cavity opening alongside the polar amino acids TYR108, TYR131, and TYR139 (Ilari et al. (2009); Sharma et al. (2018)). The remaining portions of the structure were considerably more rigid, showing an RMSD (Å) lower than 1 Å.

The Cα RMSD (Å) for the three MD simulations of the ES complex between the NCS and the substrates dopaminium and *(S)*-citronellal are also shown in the subsequent plots of Figure 3.5 A. Likely the free NCS MD, the three ES complex MD showed stabilization of the overall secondary and tertiary structure, and no unexpected event regarded to the enzyme was witnessed during this simulation trajectory time. The RMSF (Å) for the three ES complex simulations and their average are shown in the same plot for the free NCS (black) (Figure 3.5 B). The average RMSF (Å) for the ES complex simulations (orange) shows good agreement with the free NCS, although it appears less flexible, indicating an entropic penalty due to constraints caused by interactions with the ligands.

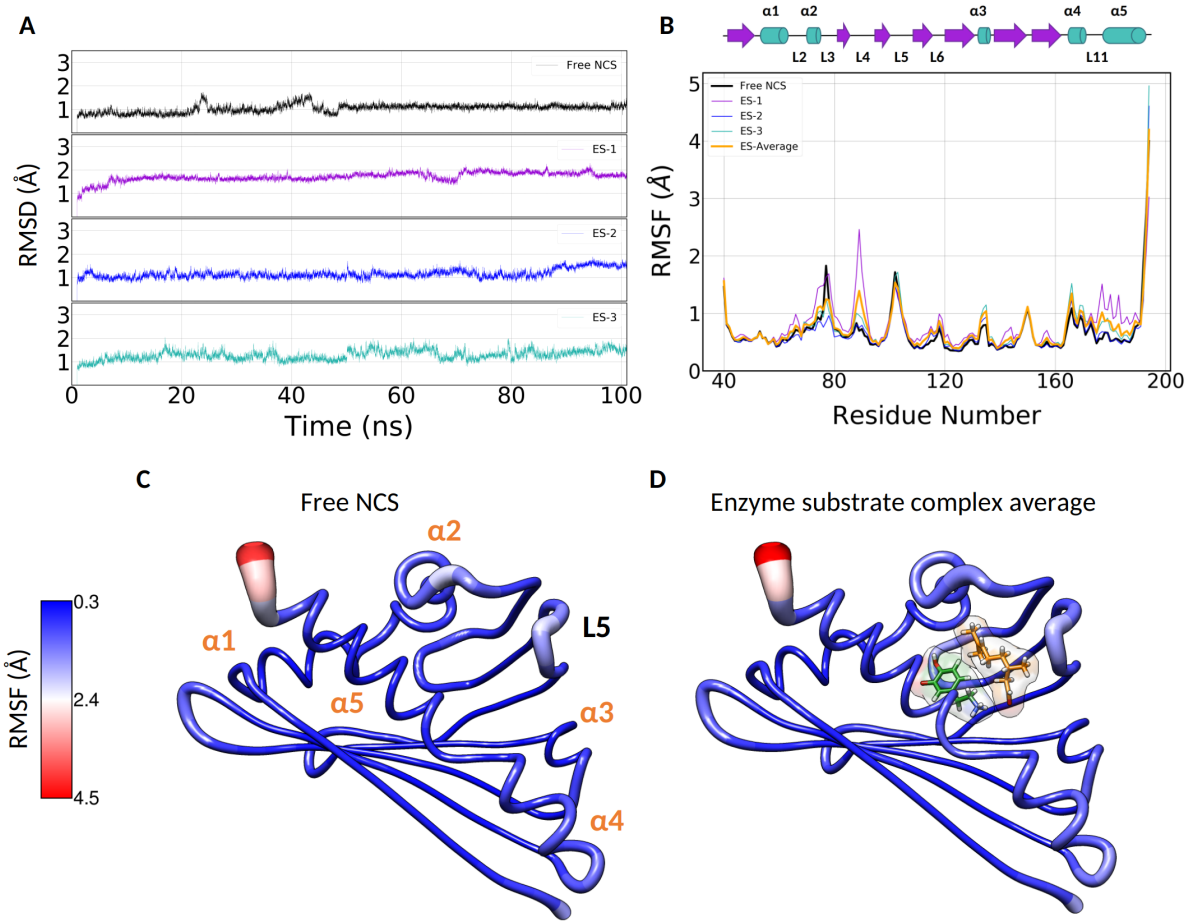


Figure 3.5: Molecular dynamics of the free NCS and the ES complex of *Tf*NCS and dopaminium and *(S)*-citronellal. (A) The RMSD (Å) as a function of the trajectory simulation time (ns) for the free NCS (black), ES 1 (purple), ES 2 B (blue), and ES 3 (cyan). (B) The RMSF (Å) as a function of the NCS residue number for the free NCS (black), ES 1 (purple), ES 2 (blue), ES 3 (cyan), and the average value for the three ES complex of NCS (orange). (C) and (D) *Tf*NCS tertiary structure colored in red-white-blue gradient according to the RMSF (Å) of the free NCS and the ES average, respectively.

Moreover, although the average RMSF of the three ES complex MD behaves similarly to the free NCS, loops four and eleven and part of α-helix five are especially flexible in simulation ES 1 (purple), which is not seen in the others. This extra flexibility can be associated with an exchange in the active site between the dopaminium and the *(S)*-citronellal observed at approximately 45 ns of the trajectory simulation (Figures 3.6 A and B). This result is comparable to the previous docking results for the iminium intermediate mimic constructed with the *(S)*-citronellal's aliphatic chain, where it was observed that the aldehyde moiety could also occupy a deeper site of the cavity (section 3.3.1). Also, likely for the docking in which those did not represent the majority of the resulting poses, this event is less pronounced in the molecular dynamics, *e.i.*, it is possible to occur but with less frequency. Interestingly, it shows accordance with the results reported in the crystallographic data of PDB ID 2VQ5, where the aldehyde portion appears buried deeper into the cavity, once

more, leading to the first idea for the proposed "aldehyde-first mechanism" (Ilari et al. (2009); Lichman et al. (2015a)).

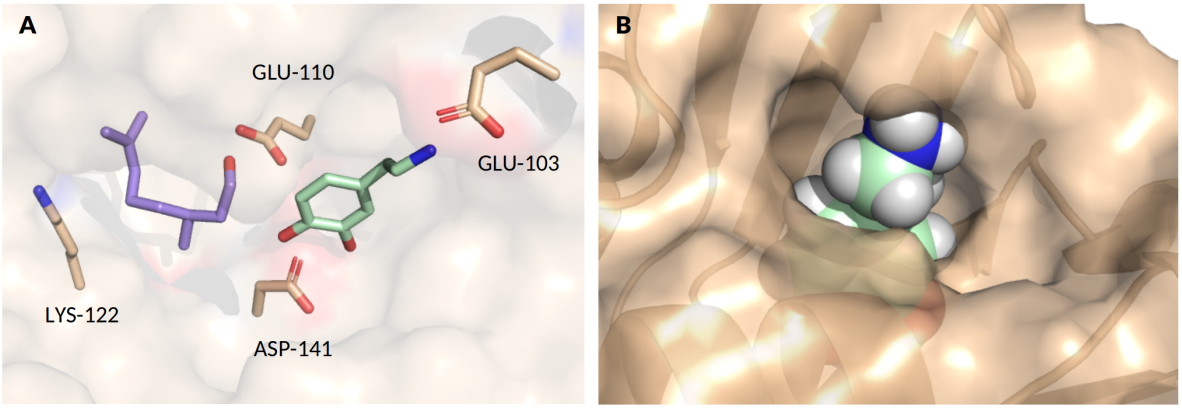


Figure 3.6: (A) Exchange in place between the dopaminium and the *(S)*-citronellal observed in the ES 1 MD simulation starting at 45 ns. (B) The dopaminium at the entrance of the cavity. The amino group interacts with the amino acid GLU103 that belongs to loop five, responsible for the cavity entrance regulation.

Some residues play an essential role in the NCS overall mechanism, whether they are actively involved in the reaction or not. For example, some account for shaping the cavity, regulating the substrates' entrance and acceptance, and modeling their conformation. In Figure 3.7 it is reported some of the most crucial amino acid residues and their conformational diversity in terms of their RMSD (Å).
The side chain of the GLU110 appears to flip 180° throughout the trajectories. As the GLU110 is in its protonated form in the simulations, conformations C1 and C2 are considered different because the two oxygen atoms are not chemically identical. For ES 3, the conformation is slightly different, and the 180° flip is not achieved. This effect can be attributed to the different orientations of the *(S)*-citronellal carbonyl group in the cavity, which is closer to the side chains of ASP141 and GLU110 in this particular simulation. Therefore, the overall conformation acquired by GLU110 for the free NCS does not differ from those observed for the ES complexes. However, it is reported a different crystallographic orientation for the free enzyme (PDB ID 5VNE (Ilari et al. (2009))) that was not observed in our simulations. One possible explanation is that we performed the free NCS MD simulation by removing the ligands co-crystallized in the PDB ID 5NON, and possibly the simulation trajectory did not access the other conformation.
For the PHE112 residue, it was observed a 0°/90°/180°rotation of the side chain. The 90° rotation is seen only for the ES complexes and can be associated with π-π interactions between the aromatic ring of the PHE112 and the dopamine catechol group. Although the PHE112 does not participate actively in the mechanism reaction, it is suggested that it influences the binding modes of the substrates inside the cavity since an alternative conformation for its side chain was reported in the PDB ID 2VQ5 chain B, where the

aldehyde is found to be deeper into the active site. However, this conformation is not observed to a large extent. Some previous MD simulations for this crystallographic data (Lichman et al. (2015a)) showed that this secondary conformation is quickly changed to the more representative one, as seen in other crystallographic data for the NCS.

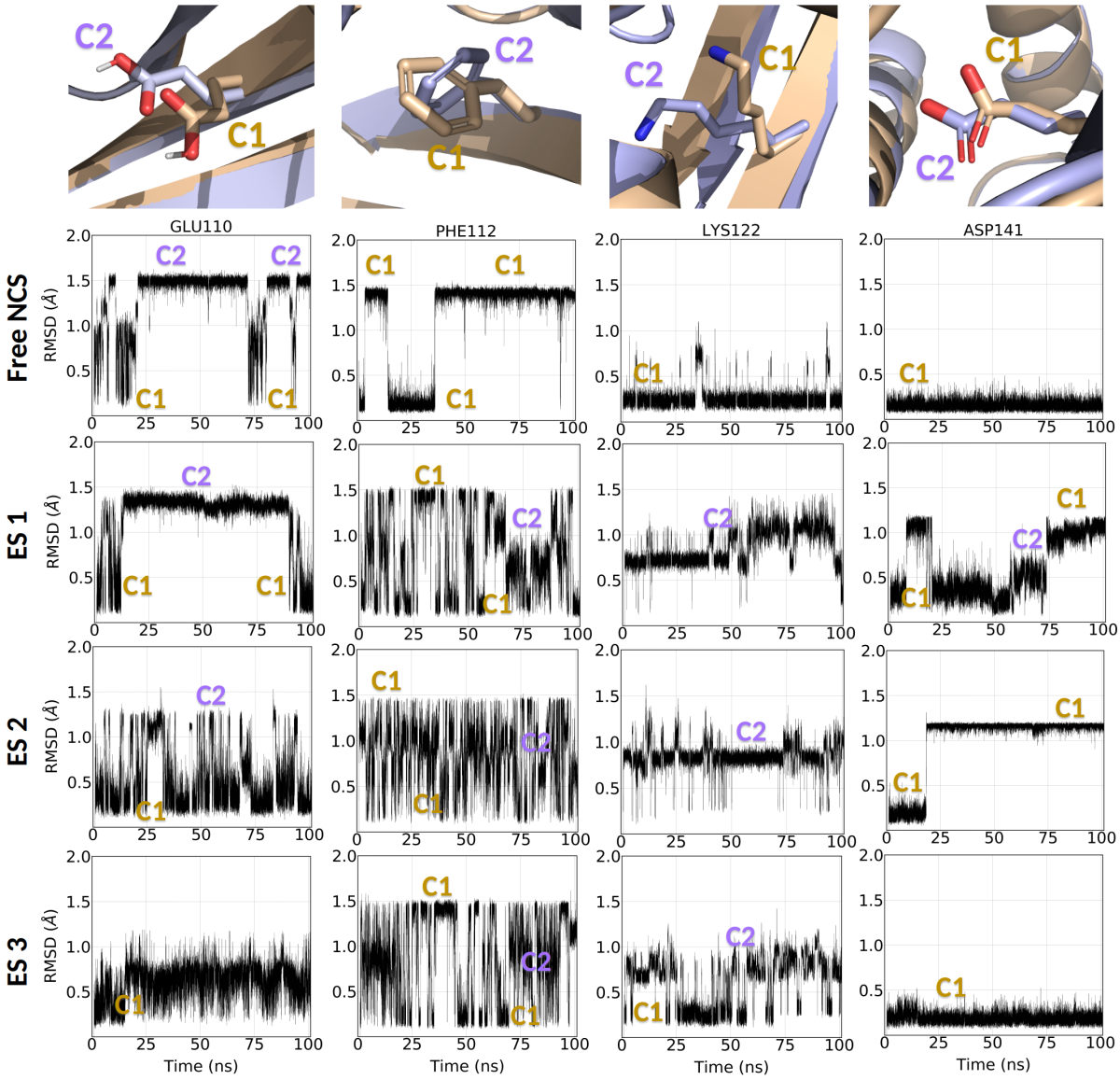


Figure 3.7: RMSD (Å) as a function of trajectory simulation time (ns) of the amino acid residues GLU110, PHE112, LYS122, and ASP141 of the free NCS, ES 1, ES 2, and ES 3 MD simulations. The RMSD graphics are accompanied by the representation of the different conformations adopted by each amino acid's side chain.

The amino acid ASP141 possesses both oxygen atoms of the side chain chemically equivalent due to both of them being deprotonated and, therefore, the 180° rotation seen in ES 2 is not considered a change in conformation. However, a second conformation (C2) appears in the ES 1 simulation featuring an interaction between the side chain of ASP141 and the substrate dopaminium that exchanged places with the *(S)*-citronellal. In addition, the

RMSD (Å) for this particular simulation fluctuates more, indicating the susceptibility of the ASP141 conformation to the variations of the dopaminium.
For LYS122, the conformational flexibility is regarded to a dihedral angle change. This conformation favors the interaction between the LYS122 side-chain with the dopaminium 3-OH and 4-OH of the catechol moiety. Hence, the conformation C2 is seen only in the presence of substrates in the active site. The RMSD is slightly greater for ES 1, considering the exchange in position between the dopaminium and the *(S)*-citronellal observed in this simulation, which affected the LYS122 side-chain position. The dihedral change effect on LYS122 is also evidenced in Figure 3.8 that shows the hydrogen bond fraction between LYS122 side-chain nitrogen and hydrogen atoms of the catechol portion. For simulations ES 2 and ES 3, where the substrates stayed in place, the hydrogen bonding fraction showed a preference for interaction with the 3-OH compared to the interaction with 4-OH. This data agrees with the docking results (section 3.3.1) that predicted that the active conformation has a preference over the inhibitory one.

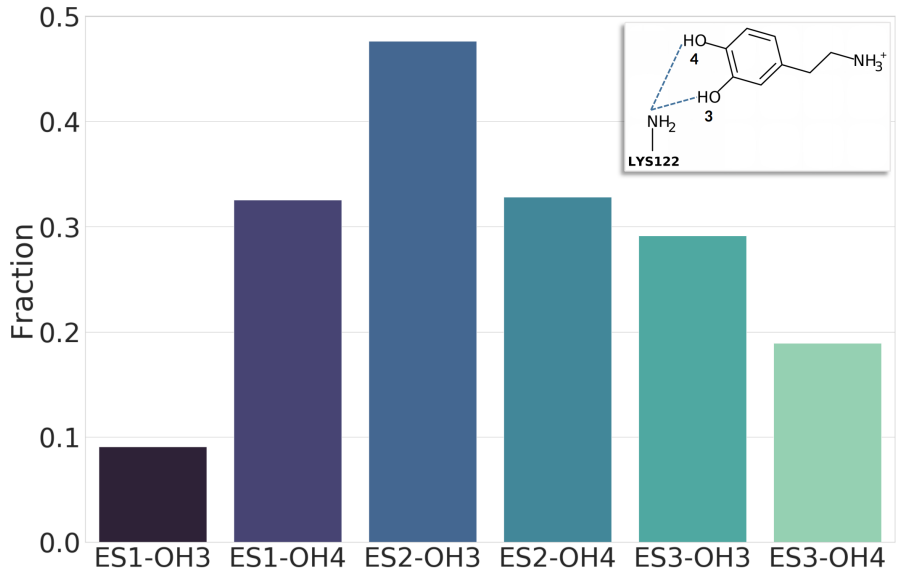


Figure 3.8: Hydrogen bond fraction between the LYS122 side chain nitrogen with dopaminium catechol 3-OH and 4-OH mapped throughout the MD trajectory.

3.4 Conclusion and Perspectives

The redocking simulation results agreed with the experimental crystallographic data, enabling the extension of this protocol to the simulations A, B, and C featuring the *(S)*-citronellal. Furthermore, the docking simulations with this aldehyde also indicated the binding mode featuring the dopaminium ligand occupying a more inward position into the cavity as more likely to happen, which is a significant result of determining the reaction mechanism NCS. Furthermore, the *(S)*-citronellal presented two different orientations at the cavity entrance displaying different interactions with the active site amino acids.

The MD simulation for the free NCS presented loops 3 and 5 more flexible than for the ES average, meaning that the interactions between the amino acids of this region and the substrates impose some restrictions to the mobility of the loops. However, ES1 showed that the cavity formed by the active site is large enough, and some exchange in place was observed between the two substrates. This result demonstrated that both binding modes are plausible to happen, whether the carbonylate ligand binds first or after the dopaminium. Furthermore, the 3-OH of the dopamine catechol interacted preferably with the nitrogen atom of the LYS122 side-chain. Moreover, the hydrogen bond fraction showed that the active configuration is representative throughout the entire MD trajectory.
From future perspectives, we aim to extend the trajectory simulation time because it might map other events not seen so far by simply increasing the sampling of frames. We also want to perform some quantitative analysis through calculations on the binding free energies.

4

NCS: reaction mechanism investigation

"(...) everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard Feynman

4.1 Introduction

The mechanistic aspects of the PS reaction catalyzed by the NCS raise much interest, considering it can help elucidate aspects of the overall carbonylate substrates profile. In this sense, much investigation has been done, whether experimentally (Luk et al., 2007; Berkner et al., 2008; Ilari et al., 2009; Bonamore et al., 2010; Lichman et al., 2015a, 2017a) or theoretically (Sheng and Himo, 2019). Initially, based on x-ray crystallography (PDB ID 2VQ5) showing the non-reactive molecule benzaldehyde co-crystallized more inward into the binding site (Ilari et al. (2009)), it was first proposed the so-called "Aldehyde-first" mechanism (Bonamore et al. (2010)), which considers the aldehyde (or ketone) interacting *a priori* with the amino acids of the active site, followed by the dopamine. The mechanism is shown in Figure 4.1. In this mechanism, the amine group of LYS122 side-chain interacts with the 4-HPAA carbonyl group, while the ASP141 and GLU110 side-chains, both deprotonated, interact with the phenyl moiety. The dopamine interacts only with the amino acid TYR108 through a hydrogen bond with the 4-OH of the catechol portion. The reaction follows with a nucleophilic attack from the dopamine nitrogen onto the 4-HPAA carbonyl (a), later on forming an iminium cation and releasing a water molecule (b and c). Then, the ring closure is mediated by a water molecule deprotonating the 3-OH (d), and the product is formed by the GLU110 removing the quinone proton and restoring the aromatization on the catechol group (e and f).

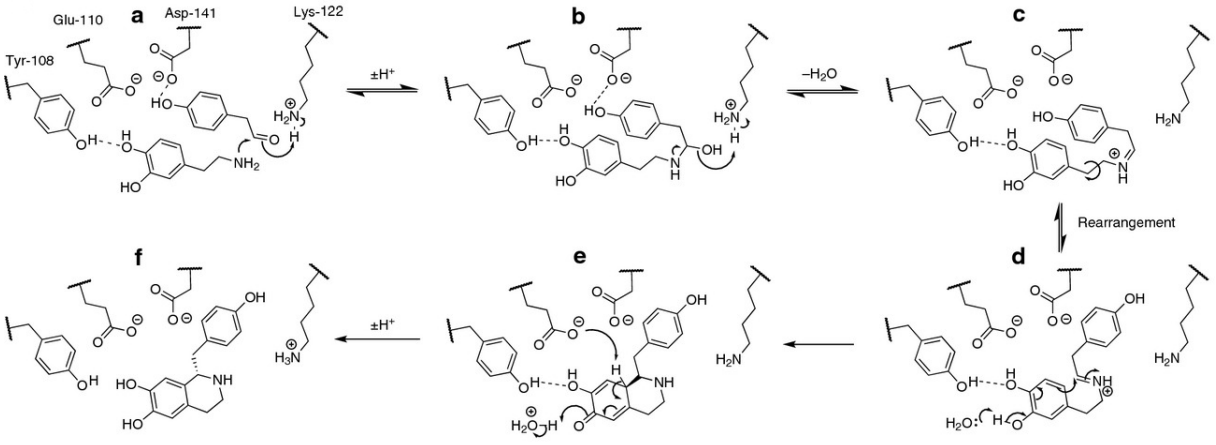


Figure 4.1: The proposed "Aldehyde-first" mechanism for the enzymatic PS reaction between the dopamine and the 4-HPAA. (a) Dopamine attacks the aldehyde carbonyl. (b) Loss of a water molecule. (c) Iminium cation rotation bond. (d) Cyclization onto the iminium. (e) Quinone proton removed by GLU110. (f) The product s-norcoclaurine formed.

Source: Reprinted with permission from Lichman et al. (2015a). *© (2015) FEBS Journal published by John Wiley Sons Ltd on behalf of FEBS.*

However, the "Aldehyde-first" mechanism does not account for the broad scope of aldehydes

and ketones reported to be active for this enzyme (Lichman et al., 2015a). First, the interaction between the ASP141 and the phenyl group would not be possible for many substrates, and second, this mechanism would not allow large, bulky carbonyl ligands (Lichman et al., 2015a). In this sense, the "Dopamine-first" mechanism was proposed, initially, based on molecular docking results of reaction intermediates using the crystal structure of the *Tf*NCS in PDB ID 2VQ5 subunit A (Pesnot et al., 2011). The docking calculations showed that the aldehyde R-group was exposed to the solvent, rationalizing why large aldehyde molecules were accepted. Nevertheless, experimental structural evidence was later obtained with a crystallized *Tf*NCS featuring an iminium intermediate mimic bound at the active site (PDB ID 5NON (Lichman et al., 2017a)). The crystallographic result shows the LYS122 interacting with the catechol group from the dopamine portion of the molecule, indicating that the dopamine might bind first to the enzyme.
Sheng and Himo (2019), by using DFT, investigated the two mechanisms and discovered that both binding modes are energetically similar, differing by only 0.5 kcal.mol^{-1}. However, the "Aldehyde-first" reaction pathway led to very high energy transition states. Hence, the "Dopamine-first" mechanism was proposed to be more likely to happen. The detailed reaction mechanism is shown in Figure 4.2. The active site features both LYS122 and GLU110 with neutral side-chains. First, at the ES complex, the dopaminium is deprotonated by the amino acid ASP141, leaving the nitrogen atom free to make a nucleophilic attack onto the 4-HPAA carbonyl group forming Int2, a zwitterionic intermediate. The oxygen is readily protonated by the GLU110 side-chain neutralizing its charge, and, in sequence, the GLU110 removes another proton from the charged nitrogen. Finally, the GLU110 suffers the second deprotonation by the oxygen atom that leaves as a water molecule, leading to the formation of the iminium cation. This intermediate cation (Int5) formed is in the E-configuration, which will drive to the (S) product at the end of the reaction. The same calculations were also performed for the (R) product, but the reaction path energy profile showed high transition state energies. It explains in terms of energy the enantioselectivity presented by the NCS. Also, because of high energy barriers, a proton transfer between the GLU110 and the ASP141, mediated by a water molecule, occurs before the ring closure step (Int6). The LYS122 side-chain removes the proton from the catechol 3-OH increasing the nucleophilic character on the *para* position enabling cyclization. After that, the GLU110 and ASP141 undo the proton transfer, leaving the former free to remove the quinone hydrogen and restore the aromatization of the catechol moiety. The abstraction of the quinone proton to form the final product was calculated to have the highest barrier of this reaction path and, therefore, is considered the rate-limiting step of the reaction.

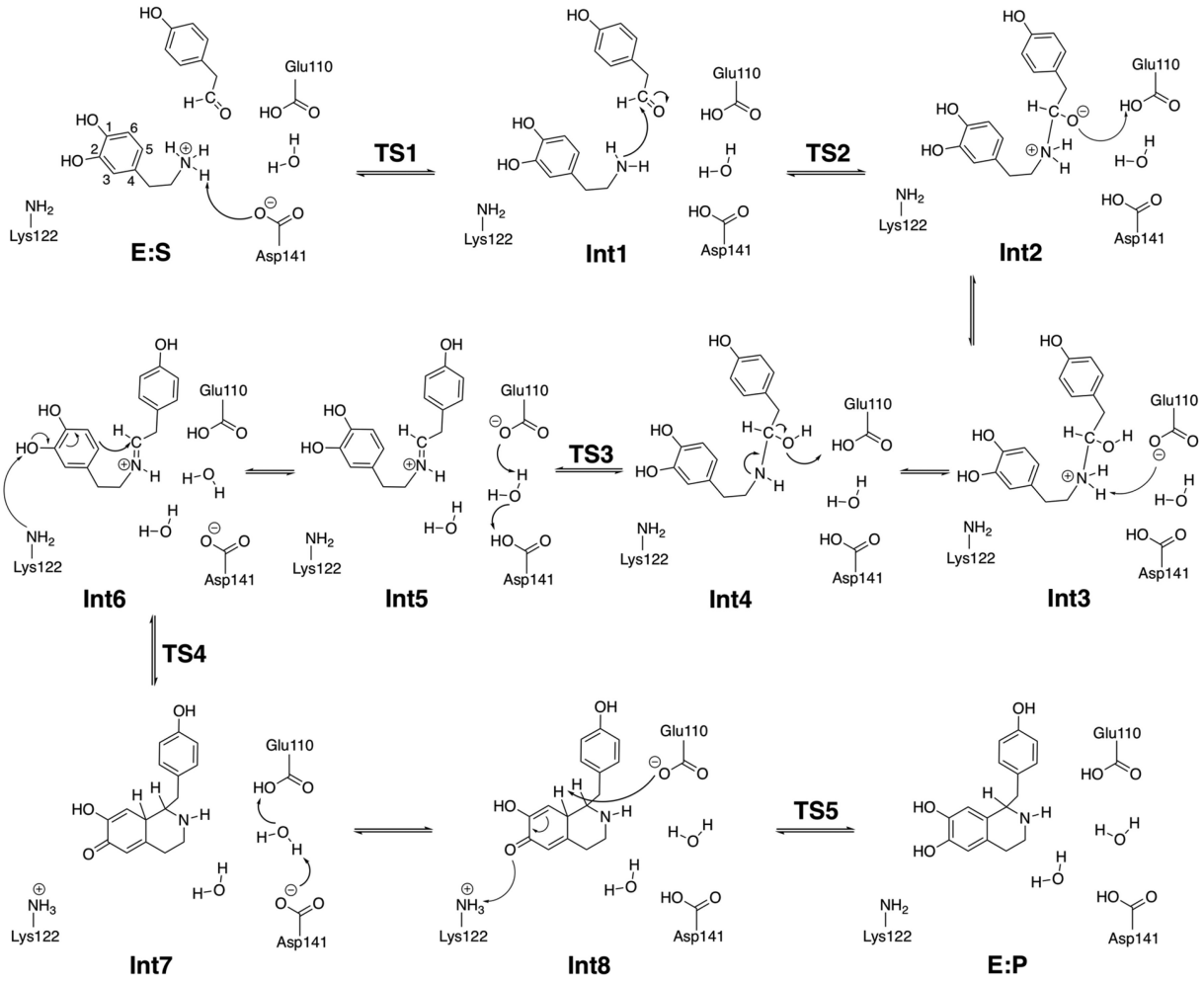


Figure 4.2: The proposed mechanism for the enzymatic PS reaction between the dopaminium and the 4-HPAA.

Source: Reprinted with permission from Sheng and Himo (2019). *Copyright ©(2019), American Chemical Society.*

In this chapter, we investigate the NCS mechanism of reaction for the non-natural substrate *(S)*citronellal following the "Dopamine-first" proposition and compare it to the mechanism proposed for the 4-HPAA.

4.2 Methodology

System Preparation

The active site model was prepared using the enzyme *Tf*NCS crystal structure from PDB ID 5NON, in consonance with the molecular docking and MD simulations presented previously. Our cluster model consists of the active site amino acids TYR63, SER64, TRP65, PRO66, GLY67, LEU68, ALA69, LEU72, LEU76, LEU95, MET97, PHE99, TYR108, GLU110, PHE112, LYS122, VAL124, MET126, TYR139, ASP141, THR159, LEU180, and MET183, three crystallographic water molecules (201, 288, and 302), and

the substrates dopaminium and *(S)*citronellal from docking C described in the previous chapter in section 3.2.1. The amino acids were truncated at Cα and hydrogen atoms were added to complete the valence. The truncated carbon and the added hydrogen atoms had their crystallographic coordinates kept fixed aiming to avoid virtual movements throughout the optimization processes. The ionization state for LYS122 and GLU110 also followed the same protocol as for docking and MD simulations, being both neutral. The total net charge of the cluster model was zero, and total number of atoms, 356.

Cluster Simulation

For the cluster calculations, all steps of the mechanism were investigated using the Density Functional Theory (DFT) at ORCA 4.2.1 (Neese (2012, 2018). The geometry optimizations and vibrational frequencies at 1 atm and 298.15 K of reactants, intermediates, and product within the cluster were performed at the PBE-D3(BJ)/SV(P) level of theory (Grimme et al., 2010, 2011; Schäfer et al., 1992; Perdew et al., 1996, 1997). Coulomb integrals were sped up by RI approximation using def2/J as auxiliary basis set. Also, the restricted Kohn-Sham formalism (RKS) was used considering the whole system is closed-shell. Single-point energies were computed for energy refinement at B3LYP-D3(BJ)/def2-TZVP(-F) level (Schäfer et al., 1992, 1994; Becke, 1993; Lee et al., 1988), and the Coulomb integrals were sped up by RIJCOSX approximation (Neese, 2003; Neese et al., 2009) using the def2/J auxiliary basis set. The transition state (TS) calculation was performed using the Nudged Elastic Band (NEB) method (Mills et al., 1995; Henkelman and Jónsson, 2000) at level of theory PBE-D3(BJ)/SV(P). The TS was submitted to single-point and vibrational frequencies calculations following the same protocol used for the intermediates mentioned above. No solvation calculations were performed at this point, and the results presented are, thus, for the gas phase.

4.3 Results and Discussion

The reaction mechanism

The mechanism for the enzyme-catalyzed PS reaction between the dopaminium and the *(S)*citronellal is shown in Figure 4.3 and the calculated energy profile, in Figure 4.4. Here, the ES complex features the *N*-protonated dopamine, which has its most acidic hydrogen abstracted by the carboxyl moiety of the residue ASP141, forming the first reaction intermediate (Int1) with energy of +5.8 kcal.mol^{-1} higher than the ES complex. At this point, the free dopamine amino group performs a nucleophilic addition onto the *(S)*citronellal formyl group, starting the condensation by forming a C-N bond. In the same step, the GLU110 side-chain performs a water-mediated protonation of the carbonyl oxygen, leading to the intermediate Int2, which is stabilized by -10.8 kcal.mol^{-1} relative to

the ES. This is the lowest point in the energy profile of the entire mechanism. Unlike the reaction mechanism reported by Sheng and Himo (2019) (recall Figure 4.2), the zwitterion ion does not seem to be stable in our system. All attempts made in order to optimize an intermediate with an alkoxide species led to the imediate protonation of the oxygen atom. In this sense, we suggest that the nucleophilic attack and the alkoxide protonation occurs as fast as concerted. However, further calculations on the transition state for this step (Int1 → Int2) need to be performed to confirm the hypothesis.

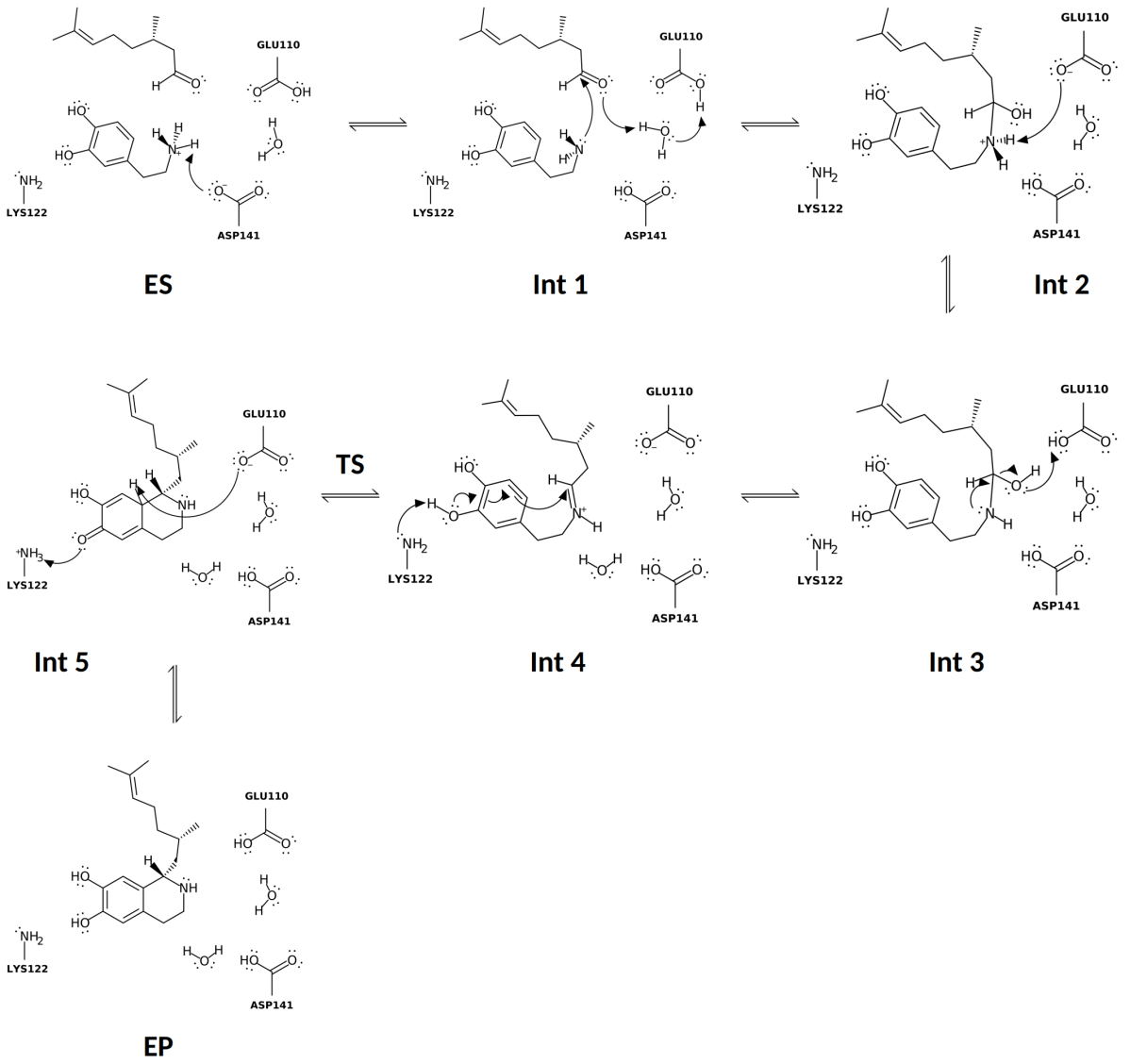


Figure 4.3: The reaction mechanism proposed in this work for the condensation of the dopaminium and the *(S)*citronellal leading to the formation of the (S,S)-DMH-THIQ.

Subsequently, the GLU110 side-chain carbonylate group deprotonates the recently formed *N*-protonated hemiaminal cation in Int2. Although the Int3 formed is a neutral molecule, it is 2.2 kcal.mol^{-1} higher in energy than its predecessor Int2. Then, an acid-assisted dehydration yields the C=N bond leading to an iminium intermediate (Int4). This step

forms the *E*-isomer, which will culminate in the (S) product during the ring closure. The Int4 is more stable than Int3 by 1.1 kcal.mol^{-1} and with energy of -9.7 kcal.mol^{-1} regarded to ES. Next, the cyclization of the iminium intermediate (Int4 → Int5) occurs with assistance of the catechol 3-OH moiety, with its phenolic hydrogen atom being abstracted by the LYS122 amino group, which occurs in a concerted fashion with the nucleophilic attack of carbon C1 to the iminium carbon atom. The cyclic intermediate Int5 has considerably high energy compared to the iminium intermediate, being +7.3 kcal.mol^{-1} regarded to ES. However, according to our calculations, this step is barrierless, considering the transition state (TS) is 0.2 kcal.mol^{-1} lower in energy than the Int5. Once more, we have observed a different reaction path for this mechanism. Sheng and Himo (2019) reported a proton transfer between ASP141 and GLU110, mediated by a bridging water molecule, previous to the cyclization step. They discussed this step occurred spontaneously and could be explain in terms of the charge distribution shift involved in this step. Later, the proton transfers back to the ASP141 to enable the final reaction step. However, in our calculations the cyclization occurred with no further dues and, therefore, our mechanism proposed features two steps shorter at this part of the reaction.
Finally, the GLU110 side-chain removes the proton from the quinone moiety, restoring the catechol aromaticity to form the final (S)-product. The final product has the highest energy of the entire reaction, being +18.9 kcal.mol^{-1}. It is worth mentioning, however, that this energy is related to the enzyme-product (EP) complex. A very stable EP can lead to the poisoning of the catalyst, inhibiting the site to turn over other substrates.

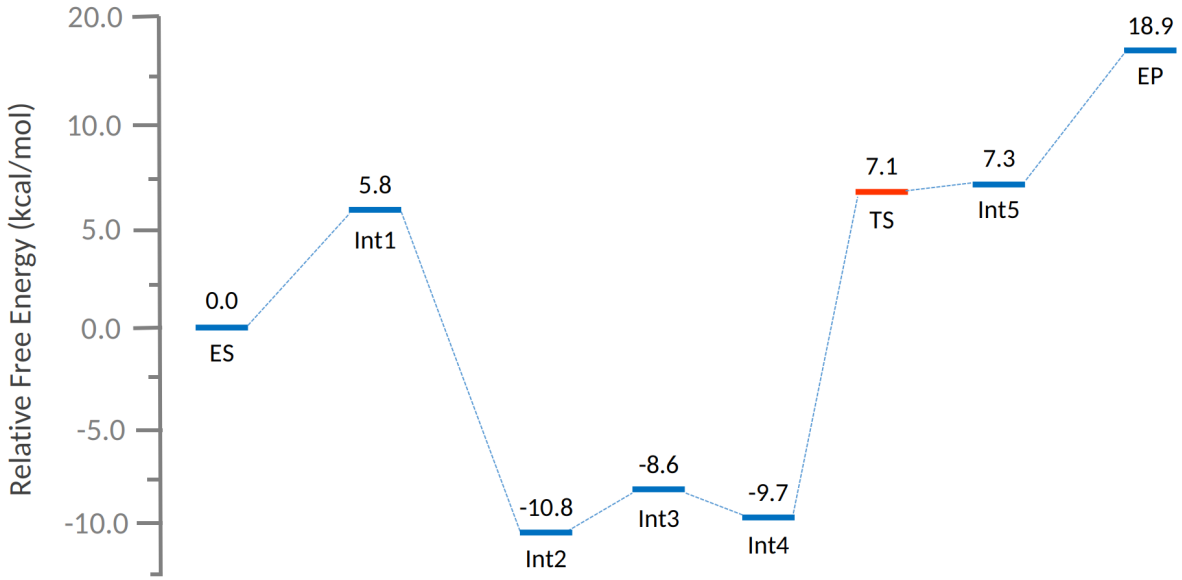


Figure 4.4: The calculated relative energy profile for the NCS catalyzed reaction between the substrates dopaminium and *(S)*citronellal to formation of the (S,S)-DMH-THIQ.

In summary, our reaction mechanism investigation resulted in a shorter pathway, with three less intermediates than the mechanism proposed for the condensation between the

dopaminium and the 4-HPAA. The optimized structures can be seen at the Appendix A.4.

The cyclization transition state

The energy difference between Int4 and Int5 is 17.0 kcal.mol^{-1}, with a transition state only 0.2 kcal.mol^{-1} lower than the Int5. However, the relative energy profile shown in Figure 4.4 does not account for an important result unveiled by our calculations.

The cyclization step was observed to happen after a conformational change in the *(S)*citronellal portion of the iminium intermediate. This conformational change occurs with a high energy barrier. Figure 4.5 A shows the relative electronic energy profile for the conformational change Int4-a → Int4-b, followed by the ring closure TS leading to Int5. Indeed, the activation energy (electronic) for the TS is approximately 2.5 kcal.mol^{-1} related to the conformation b.

The high energy observed needed to perform the conformational changes can be reasoned by the orientation that the *(S)*citronellal portion acquires into the active site. As can be seen in Figure 4.5 B, the aliphatic moiety is oriented towards the two amino acid residues, LEU72 and LEU76, being constrained between them. Hence, the mobility needed for such large substrate in the active site is jeopardized. It remounts to the experimental results reported by Lichman et al. (2015a) for the variant *Tf*NCS L76A, which increased the catalytic constant, k_{cat}, for the *(S)*citronellal substrate. Structurally, the amino acid alanine has a methyl group in its side-chain, while the amino acid leucine has an isobutyl group, being a bulkier residue.

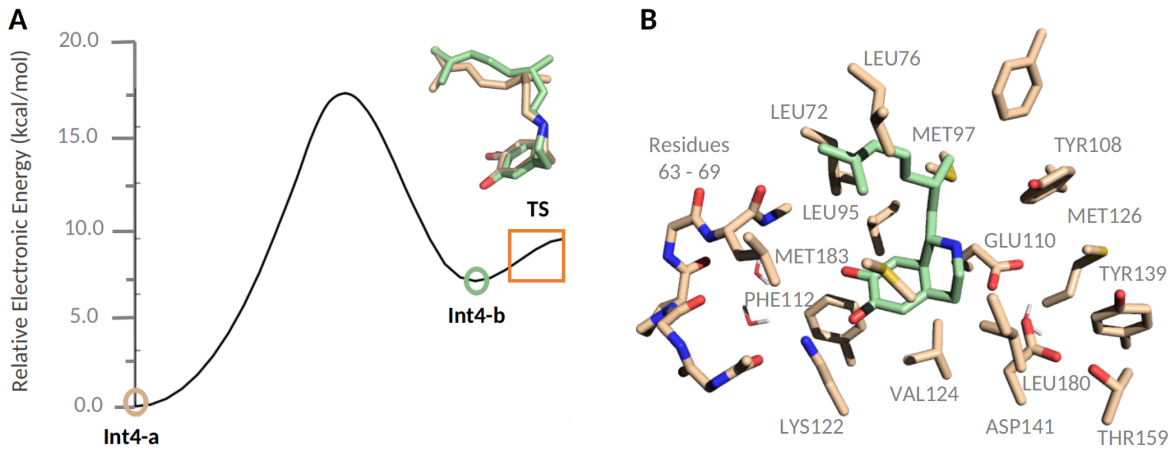


Figure 4.5: (A) Relative electronic energy showing the conformational change Int4-a → Int4-b before the cyclization step. In wheat is shown the more stable conformation for Int4, in green, the less stable conformation that will lead to the transition state (orange rectangle). (B) The cluster model showing the intermediate Int5 (in green). The aldehyde portion of the molecule is constrained between LEU72 and LEU76 side-chains.

The cluster model

The complete cluster model used in the reaction mechanism calculation is shown in Figure 4.6. Initially, the cluster model was composed by the amino acids LEU72, LEU95, PHE99, TYR108, GLU110, PHE112, LYS122, VAL124, MET126, TYR139, ASP141, THR159, LEU180, and MET183, and a water molecule, all shown in pink. The initial optimizations performed demonstrated reasonable results, except for the *(S)*citronellal chain. The carbonyl substrate is binded at the cavity entrance and possesses considerable flexibility as it is solvent exposed. This is especially true for the *(S)*citronellal due to its long aliphatic chain. However, due to the NCS structure and bulk, the substrate is not allowed to acquire random conformations. In this sense, the first cluster constructed (in pink) had to be remodeled in order to account for the extra constraints imposed by the enzyme tertiary structure.

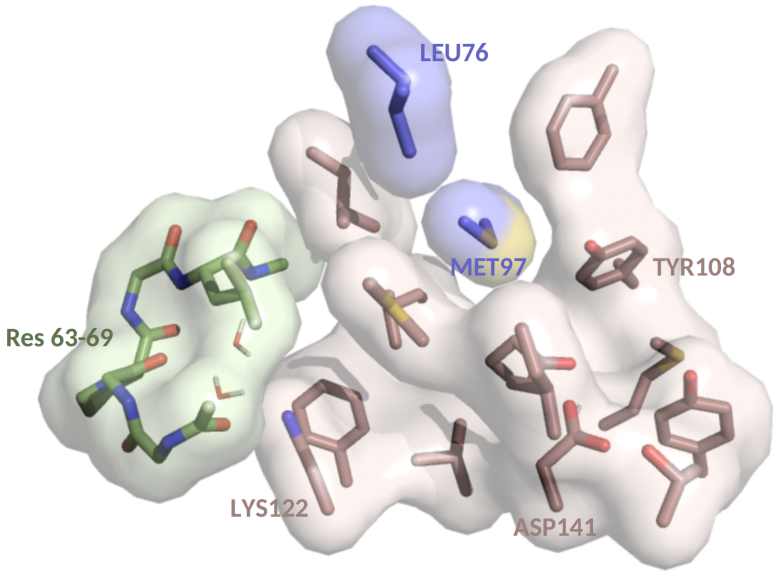


Figure 4.6: The complete cluster used in the reaction mechanism calculations. In pink, the first cluster model used in an attempt to stabilize all the reaction steps. In blue, the amino acids added to the previous cluster model. At last, in green, the final modification of the cluster model leading to a stabilization of all intermediates of the reaction.

Therefore, a second cluster model was modeled by adding the amino acid residues LEU76 and MET97 (in blue). The addition of the residue LEU76 was based on the catalytic results reported by Lichman et al. (2015a). The *Tf*NCS L76A variant was observed to increase the catalytic constant, k_{cat}, for turning over the substrate *(S)*citronellal. This effect can be rationalized considering that throughout the reaction path the *(S)*citronellal chain is kept constrained between the residues LEU72 and LEU76, and a mobility restriction is applied to the ligand. Furthermore, the mutant M97F has recently observed to play an important role on both aldehyde and ketone scope, effectively catalyzing the formation of diverse THIQ products with different chain characteristics, such as aliphatic, α-substituted, and

cyclic substrates (Zhao et al., 2021).
However, the second cluster model was not able to stabilize the cyclization step (Figure 4.3 Int4 → Int5) because the protonated LYS122 side-chain amino group could not be stabilized whatsoever and the proton would transfer back to the catechol group, which is reasonable considering the Int4 free energy is smaller than the Int5 free energy (Figura 4.4). Hence, the amino acid redisues TYR63, SER64, TRP65, PRO66, GLY67, LEU68, and ALA69, and two crystallographic water molecules (in green) were included in the second cluster, leading to the third and final cluster model, which was able to optimize and stabilize all the intermediates of the reaction. The hydrogen atoms of the LYS122 amino group are close to the amino acid TYR63 (2.098 Å) and to a water molecule (2.481 Å), which also interact with other amino acid residues creating a net of interactions and aiding the stabilization.

4.4 Conclusion and Perspectives

In summary, the reaction mechanism studies for the *(S)*citronellal showed to be reasonable to account for the important steps a PS reaction undergoes. Furthermore, our calculations demonstrated a reaction pathway featuring three fewer intermediates compared to a previous one proposed for the 4-HPAA. Also, we have shown that the substrate flexibility plays an important role in the overall reaction, leaving much room for further investigation of the dynamics of reaction intermediates. Moreover, the final cluster model demonstrated to account for the whole mechanism reaction and, also, to represent the real constraints caused by the enzyme tertiary structure. The previous two cluster models were also important results showing that, although a minimalist cluster model might be computationally cheaper, it will not be able to account for all the stabilization a complete coordination sphere offers.
For our perspectives for further investigation, we aim to calculate other transition states to settle a complete mechanism proposition for the enzyme-catalyzed condensation between the dopaminium and the *(S)*citronellal. Furthermore, a cyclization study using a cluster model featuring the variant L76A will broaden the understanding of the changes in k_{cat} in terms of energy. Finally, a solvation model will be included in the calculation to account for the dielectric medium of the enzyme.

5

Final Remarks

Molecular modeling, in general, is not an easy task to dive into. All the computational techniques applied in this work, from molecular mechanics to DFT, have their own particularities, perks, and weaknesses. Nonetheless, together they form the basis to investigate different, yet complementary, characteristics of a system. In the realm of living things, these characteristics are countless and full of details, which make the investigation of macro-molecules a thorny journey. It could not be different for the structural and mechanistic studies carried out for the NCS throughout this work. However, we demonstrated some interesting results for this enzyme.

The molecular docking and molecular dynamics simulations applied to this system corroborated the previously published results about binding priority for the substrates, indicating that the dopamine is more likely to anchor deep into the active site. It does not exclude, though, the possibility of the substrates occupy each other's position through an exchange in place. Still, the interactions of the dopamine 3-OH catechol group with LYS122 demonstrate that this pose is very representative over the simulation trajectory.

For such a large and complex system, the reaction path could not be any different. The cluster methodology was a very effective approach on this particular task. However, it is proven that this methodology should be used carefully with regards to its system. Different cluster models, with different amino acid residues and water molecules, had to be tested until we reached the final system. Moreover, even though this work proposed mechanism features three fewer intermediates in the reaction, it is still a long mechanistic pathway to calculate, and it took great effort to optimize and evaluate all its steps. In spite of that, we also opened a new path for investigation regarding the energy of the cyclization step, once our results demonstrated that the aldehyde-chain flexibility plays an important role in terms of energy, since a specific conformation should be reached before the ring closure. All in all, some further calculations, already commented, should be performed in order to completely validate some of our findings, which are the immediate perspectives for this work.

A

Supplementary Information

A.1 Molecular Docking

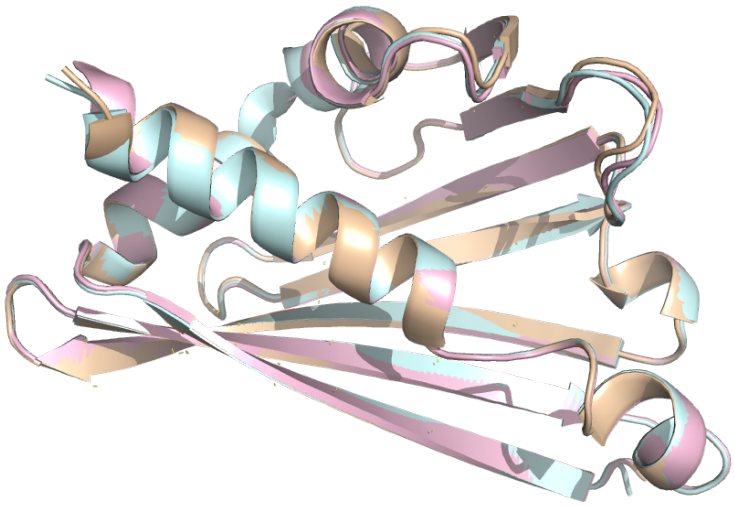


Figure A.1: Superposition of chains A (beige), B (blue), and C (pink) of the PDB ID 5NON.

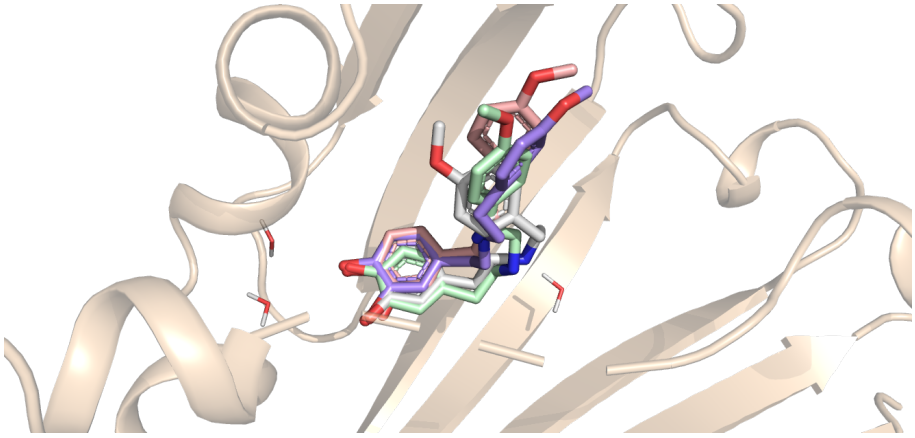


Figure A.2: Flexibility of the methoxyphenyl portion of the iminium intermediate mimic.

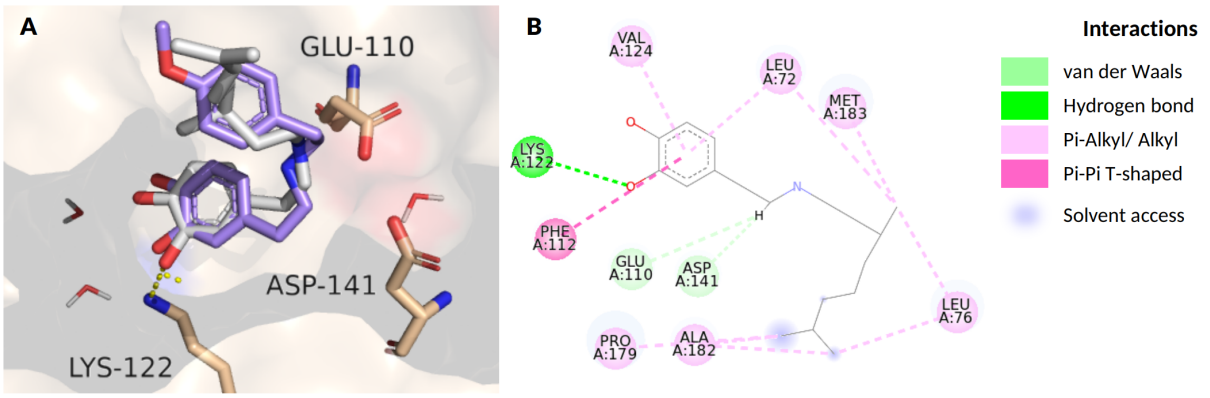
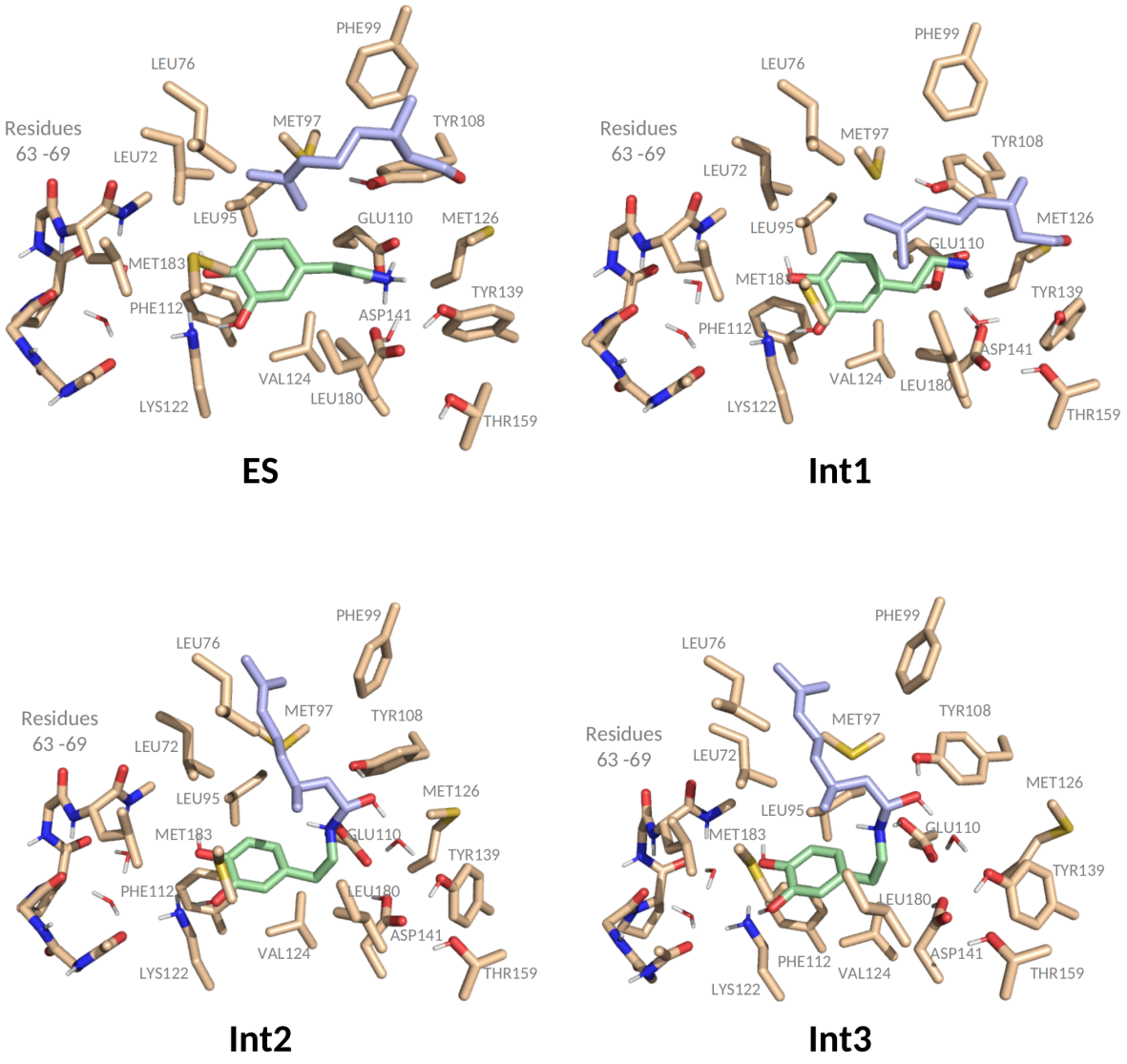


Figure A.3: (A) Docking pose of the iminium intermediate mimic using *(S)*citronellal as the aldehyde portion. (B) Its 2D contact map.

A.2 Cluster Method



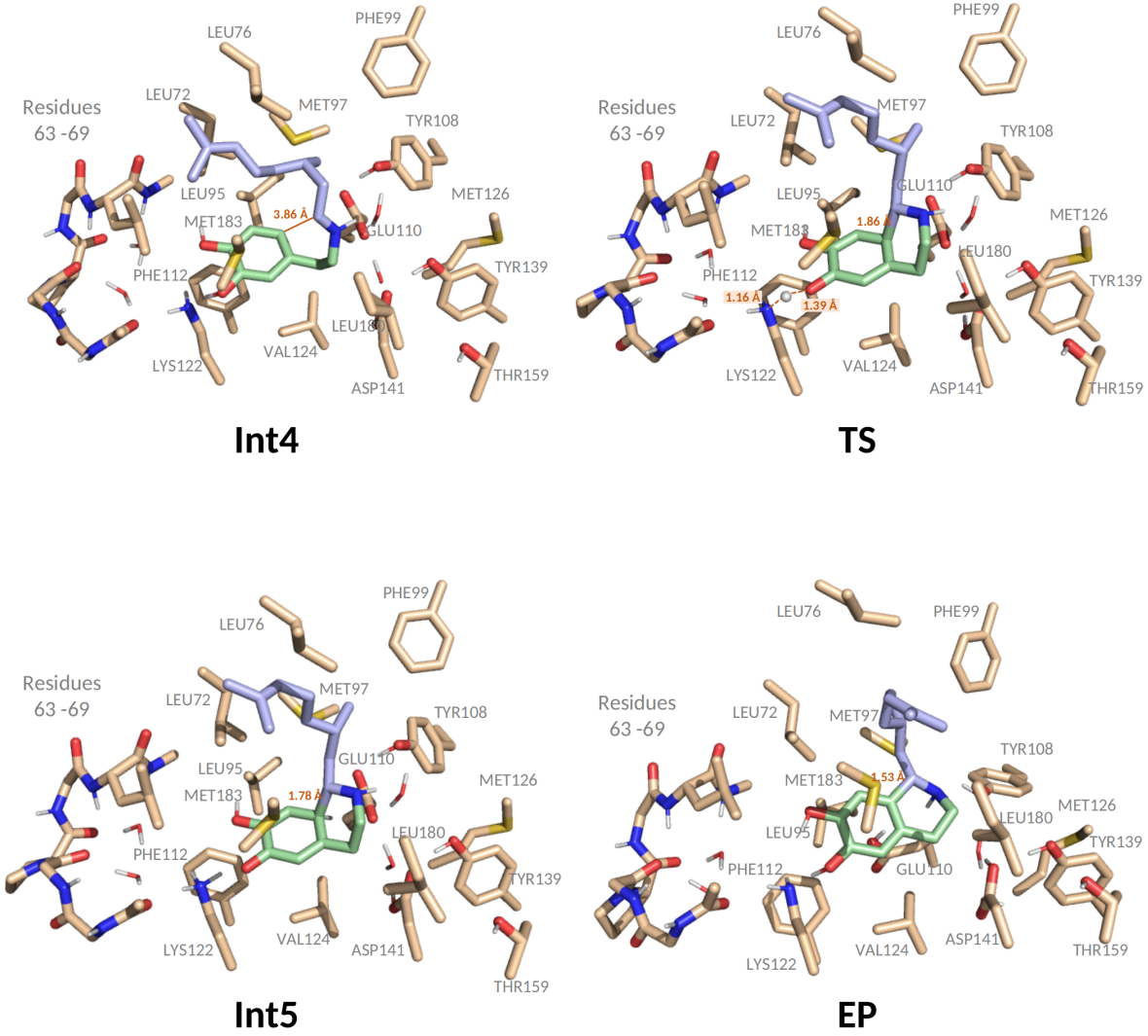


Figure A.4: The enzyme-substrate, intermediates, transition state, and enzyme-product cluster optimized geometries. In green, the dopamine-portion of the molecules and, in purple, the *(S)*citronellal-portion. The non-polar hydrogen atoms are hidden for better visualization.

Bibliography

A. de Castro, A., C. Assis, L., R. Silva, D., Corrêa, S., M. Assis, T., C. Gajo, G., V. Soares, F., and C. Ramalho, T. (2017). Computational enzymology for degradation of chemical warfare agents: promising technologies for remediation processes. *AIMS Microbiology*, 3(1):108–135.

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652.

Beeman, D. (1976). Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics*, 20(2):130–139.

Berendsen, H. J., Postma, J. P., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

Berkner, H., Schweimer, K., Matecko, I., and Rösch, P. (2008). Conformation, catalytic site, and enzymatic mechanism of the PR10 allergen-related enzyme norcoclaurine synthase. *Biochemical Journal*, 413(2):281–290.

BIOVIA (2012). Discovery Studio Visualizer Software.

Bitencourt-Ferreira, G. and de Azevedo, W. F. (2019). *Docking with SwissDock*, volume 2053. Springer Science and Business Media LLC.

Bonamore, A., Barba, M., Botta, B., Boffi, A., and Macone, A. (2010). Norcoclaurine synthase: Mechanism of an enantioselective pictet-spengler catalyzing enzyme.

Brooijmans, N. and Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, 32:335–373.

Calais, J.-L. (1990). Theory of molecular fluids. Volume 1: Fundamentals. By C. G. Gray and K. E. Gubbins, The Clarendon Press, Oxford University Press, New York, 1984. *International Journal of Quantum Chemistry*, 38(3):497–497.

Calcaterra, A., Mangiardi, L., Monache, G. D., Quaglio, D., Balducci, S., Berardozzi, S., Iazzetti, A., Franzini, R., Botta, B., and Ghirga, F. (2020). The pictet-spengler reaction updates its habits. *Molecules*, 25(2).

Capelle, K. (2006). A Bird ' s-Eye View of Density-Functional Theory. 36(4):1318–1343.

Case, D., Betz, R., Cerutti, D., Cheatham III, T., Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, R., Mermelstein, D., Merz, K., Monard, G., Nguyen, H., Nguyen, H., Omelyan, I., Onufriev, A., Roe, D., Roitberg, A., Sagui, C., Simmerling, C., Botello-Smith, W., Swails, J., Walker, R., Wang, J., Wolf, R., Wu, X., Xiao, L., and Kollman, P. (2016). Principal contributors to the current codes : Amber 2016 Reference Manual. Technical report.

Chen, Q., Ji, C., Song, Y., Huang, H., Ma, J., Tian, X., and Ju, J. (2013). Discovery of McbB, an enzyme catalyzing the β-carboline skeleton construction in the marinacarboline biosynthetic pathway. *Angewandte Chemie - International Edition*, 52(38):9980–9984.

Cramer, C. J. (2004). *Essentials of Computational Chemistry: Theories and Models*. Wiley, West Sussex, second edition.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N $\cdot$ log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092.

Durrant, J. D. and McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1):1–9.

Eastwood, J. W., Hockney, R. W., and Lawrence, D. N. (1980). P3M3DP-The three-dimensional periodic particle-particle/ particle-mesh program. *Computer Physics Communications*, 19(2):215–261.

Goodford, P. J. (1985). A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *Journal of Medicinal Chemistry*, 28(7):849–857.

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics*, 132(15):154104.

Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7):1456–1465.

Hagel, J. M. and Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: A century of discovery and a brave new world. *Plant and Cell Physiology*, 54(5):647–672.

Halgren, T. A. (1996a). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Technical Report 5-6.

Halgren, T. A. (1996b). Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. Technical Report 5-6.

Halgren, T. A. (1996c). Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. Technical Report 5-6.

Halgren, T. A. (1996d). Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. Technical Report 5-6.

Halgren, T. A. and Nachbar, R. B. (1996). Merck molecular force field. IV. Conformational energies and geometries for MMFF94. Technical Report 5-6.

Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function and Genetics*, 47(4):409–443.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeerschd, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(8):17.

Henkelman, G. and Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics*, 113(22):9978–9985.

Hohenberg, P. and Kohn, W. (1964). Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38.

Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F., Botta, B., Macone, A., Pasquo, A., Bellucci, L., and Boffi, A. (2009). Structural basis of enzymatic (S)-norcoclaurine biosynthesis. *Journal of Biological Chemistry*, 284(2):897–904.

Jabbarzadeh Kaboli, P., Rahmat, A., Ismail, P., and Ling, K. H. (2014). Targets and mechanisms of berberine, a natural drug with potential to treat cancer with special focus on breast cancer.

Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3):727–748.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935.

Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, 3(11):935–949.

Kohn, W. and Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133.

Kollman, P. (1993). Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chemical Reviews*, 93(7):2395–2417.

Korb, O., Stützle, T., and Exner, T. E. (2009). Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96.

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789.

Lichman, B. R., Gershater, M. C., Lamming, E. D., Pesnot, T., Sula, A., Keep, N. H., Hailes, H. C., and Ward, J. M. (2015a). Dopamine-first mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile. *FEBS Journal*, 282(6):1137–1151.

Lichman, B. R., Lamming, E. D., Pesnot, T., Smith, J. M., Hailes, H. C., and Ward, J. M. (2015b). One-pot triangular chemoenzymatic cascades for the syntheses of chiral alkaloids from dopamine. *Green Chemistry*, 17(2):852–855.

Lichman, B. R., Sula, A., Pesnot, T., Hailes, H. C., Ward, J. M., and Keep, N. H. (2017a). Structural Evidence for the Dopamine-First Mechanism of Norcoclaurine Synthase. *Biochemistry*, 56(40):5274–5277.

Lichman, B. R., Zhao, J., Hailes, H. C., and Ward, J. M. (2017b). Enzyme catalysed Pictet-Spengler formation of chiral 1,1'-disubstituted- A nd spiro-tetrahydroisoquinolines. *Nature Communications*, 8:1–9.

Liscombe, D. K., MacLeod, B. P., Loukanina, N., Nandi, O. I., and Facchini, P. J. (2005). Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry*, 66(20):2501–2520.

Liu, C., Zhang, B., Zhu, Y., and Tang, M. (2016). Evaluations of AMBER force field parameters by MINA approach for copper-based nucleases. *Structural Chemistry*, 27(5):1449–1464.

Luk, L. Y., Bunn, S., Liscombe, D. K., Facchini, P. J., and Tanner, M. E. (2007). Mechanistic studies on norcoclaurine synthase of benzylisoquinoline alkaloid biosynthesis: An enzymatic Pictet-Spengler reaction. *Biochemistry*, 46(35):10153–10161.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713.

Matsumura, E., Matsuda, M., Sato, F., and Minami, H. (2013). Microbial production of plant benzylisoquinoline alkaloids. *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*, pages 3–24.

McQuarrie, D. A. (2008). *Quantum Chemistry*. University Science Books, 2nd edition.

Mills, G., Jónsson, H., and Schenter, G. K. (1995). Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surface Science*, 324(2-3):305–337.

Minami, H., Dubouzet, E., Iwasa, K., and Sato, F. (2007). Functional analysis of norcoclaurine synthase in Coptis japonica.

Neese, F. (2003). An Improvement of the Resolution of the Identity Approximation for the Formation of the Coulomb Matrix. *Journal of Computational Chemistry*, 24(14):1740–1747.

Neese, F. (2012). The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78.

Neese, F. (2018). Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):1327.

Neese, F., Wennmohs, F., Hansen, A., and Becker, U. (2009). Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chemical Physics*, 356(1-3):98–109.

Nishihachijo, M., Hirai, Y., Kawano, S., Nishiyama, A., Minami, H., Katayama, T., Yasohara, Y., Sato, F., and Kumagai, H. (2014). Asymmetric synthesis of tetrahydroisoquinolines by enzymatic Pictet-Spengler reaction. *Bioscience, Biotechnology and Biochemistry*, 78(4):701–707.

Norel, R., Wolfson, H. J., and Nussinov, R. (1999). Small molecule recognition: Solid angles surface representation and molecular shape complementarity. *Combinatorial chemistry high throughput screening*, 2(4):223–22337.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, 3(10).

Parr, R. G. and Yang, W. (1989). *Density Functional Theory of Atoms and Molecules*. New York: Oxford University Press.

Perdew, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868.

Perdew, J. P., Burke, K., and Ernzerhof, M. (1997). Erratum: Generalized gradient approximation made simple (Physical Review Letters (1996) 77 (3865)). *Physical Review Letters*, 78(8):1396.

Pesnot, T., Gershater, M. C., Ward, J. M., and Hailes, H. C. (2011). Phosphate mediated biomimetic synthesis of tetrahydroisoquinoline alkaloids. *Chemical Communications*, 47(11):3242–3244.

Pictet, A. and Spengler, T. (1911). Über die Bildung von Isochinolinâderivaten durch Einwirkung von Methylal auf Phenylâäthylamin, Phenylâalanin und Tyrosin. *Berichte der deutschen chemischen Gesellschaft*, 44(3):2030–2036.

Roddan, R., Gygli, G., Sula, A., Méndez-Sánchez, D., Pleiss, J., Ward, J. M., Keep, N. H., and Hailes, H. C. (2019). Acceptance and Kinetic Resolution of α-Methyl-Substituted Aldehydes by Norcoclaurine Synthases. *ACS Catalysis*, 9(10):9640–9649.

Roddan, R., Sula, A., Méndez-Sánchez, D., Subrizi, F., Lichman, B. R., Broomfield, J., Richter, M., Andexer, J. N., Ward, J. M., Keep, N. H., and Hailes, H. C. (2020a). Single step syntheses of (1S)-aryl-tetrahydroisoquinolines by norcoclaurine synthases. *Communications Chemistry*, 3(1).

Roddan, R., Ward, J. M., Keep, N. H., and Hailes, H. C. (2020b). PictetâSpenglerases in alkaloid biosynthesis: Future applications in biocatalysis.

Roe, D. R. and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095.

Ruff, B. M., Bräse, S., and O'Connor, S. E. (2012). Biocatalytic production of tetrahydroisoquinolines. *Tetrahedron Letters*, 53(9):1071–1074.

Samanani, N. and Facchini, P. J. (2002). Purification and characterization of norcoclaurine synthase: The first committed enzyme in benzylisoquinoline alkaloid biosynthesis in plants. *Journal of Biological Chemistry*, 277(37):33878–33883.

Santana Azevedo, L., Pretto Moraes, F., Morrone Xavier, M., Ozorio Pantoja, E., Villavicencio, B., Aline Finck, J., Menegaz Proenca, A., Beiestorf Rocha, K., and Filgueira de Azevedo, W. (2013). Recent Progress of Molecular Docking Simulations Applied to Development of Drugs. *Current Bioinformatics*, 7(4):352–365.

Schäfer, A., Horn, H., and Ahlrichs, R. (1992). Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics*, 97(4):2571–2577.

Schäfer, A., Huber, C., and Ahlrichs, R. (1994). Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *The Journal of Chemical Physics*, 100(8):5829–5835.

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., and Montgomery, J. A. (1993). General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 14(11):1347–1363.

Schuster, P. (1983). G. C. Maitland, M. Rigby, E. B. Smith, W. A. Wakeham: Intermolecular Forces - Their Origin and Determination, Clarendon Press, Oxford 1981. 616 Seiten, Preis 39.50. $Berichte der Bunsengesellschaft für physikalische Chemie, 87(3) : 291 - -292$.

Sharma, S., Joshi, G., Kalra, S., Singh, S., and Kumar, R. (2018). Synthetic Versus Enzymatic Pictet-Spengler Reaction: An Overview. *Current Organic Synthesis*, 15(7):924–939.

Sheng, X. and Himo, F. (2019). Enzymatic Pictet-Spengler Reaction: Computational Study of the Mechanism and Enantioselectivity of Norcoclaurine Synthase. *Journal of the American Chemical Society*, 141(28):11230–11238.

Simonson, T., Archontis, G., and Karplus, M. (2002). Free energy simulations come of age: Protein-ligand recognition. *Accounts of Chemical Research*, 35(6):430–437.

Stöckigt, J., Antonchick, A. P., Wu, F., and Waldmann, H. (2011). The pictet-spengler reaction in nature and in organic chemistry. *Angewandte Chemie - International Edition*, 50(37):8538–8564.

Stöckigt, J. and Zenk, M. H. (1977). Strictosidine (isovincoside): The key intermediate in the biosynthesis of monoterpenoid indole alkaloids. Technical Report 18.

Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649.

Szabo, A. and Ostlund, N. L. (1996). *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Mineola.

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98–103.

Verlet, L. (1968). Computer "experiments" on classical fluids. II. Equilibrium correlation functions. *Physical Review*, 165(1):201–214.

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247–260.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174.

Wang, Y., Tappertzhofen, N., Méndez-Sánchez, D., Bawn, M., Lyu, B., Ward, J. M., and Hailes, H. C. (2019). Design and Use of de novo Cascades for the Biosynthesis of New Benzylisoquinoline Alkaloids. *Angewandte Chemie - International Edition*, 58(30):10120–10125.

Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., Li, Y., Meade, F., Teodor, R., Vaistij, F. E., Walker, C., Bowser, T. A., and Graham, I. A. (2012). A papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science*, 336(6089):1704–1708.

Zhao, J., Lichman, B. R., Ward, J. M., and Hailes, H. C. (2018). One-pot chemoenzymatic synthesis of trolline and tetrahydroisoquinoline analogues. *Chemical Communications*, 54(11):1323–1326.

Zhao, J., Méndez-Sánchez, D., Roddan, R., Ward, J. M., and Hailes, H. C. (2021). Norcoclaurine Synthase-Mediated Stereoselective Synthesis of 1,1'-Disubstituted, Spiro- And Bis-Tetrahydroisoquinoline Alkaloids. *ACS Catalysis*, 11(1):131–138.

Ziegler, J., Facchini, P. J., Geißler, R., Schmidt, J., Ammer, C., Kramell, R., Voigtländer, S., Gesell, A., Pienkny, S., and Brandt, W. (2009). Evolution of morphine biosynthesis in opium poppy. *Phytochemistry*, 70(15-16):1696–1707.