UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Quantum chemistry and conformational sampling meet together: a powerful approach to study and design metalloprotein reactivity

Emanuele Monza

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Quantum chemistry and conformational sampling meet together: a powerful approach to study and design metalloprotein reactivity

Ph.D. thesis in Computational and Applied Physics
Department of Physics, December 2016

Candidate: Mr. **Emanuele Monza**

Director: Prof. Victor Guallar Tasies (Barcelona Supercomputing Center)
Tutor: Prof. Manel Canales Gabriel (Universitat Politecnica de Catalunya)

Table of contents

Abstract

Understanding how metalloproteins work is of paramount importance to successfully design the next generation of catalysts, materials and therapeutics. Molecular modeling constitutes a valuable tool to inspect, and predict, their properties. In particular, the combination of classical mechanics-based conformational sampling and quantum chemical techniques allows to model how they catalyze a reaction or covalently bind a ligand, without missing crucial information about the dynamics of the whole protein. In this thesis, these computational techniques are systematically employed (combined but also separately) to study and guide present and future design efforts of laccases and hemoglobin. Laccases are copper-containing oxidoreductases of high biotechnological interest, with a role in several industrial applications such as textile and food industry, bioremediation, forestry and organic synthesis. Hemoglobin, which carries oxygen from the lungs to the tissue in the body, is a candidate for blood substituent design.

KEYWORDS: QM/MM, PELE, computational protein design, *in silico* enzyme engineering, laccases, hemoglobin, oxidoreductases, allostery, cooperativity, biocatalysis

Acknowledgements

I want to express my gratitude to:

Victor Guallar, my thesis director, for giving me the chance to work in his group on stimulating projects, for his criticism and for the fruitful discussions.

Manel Canales, my tutor, for making my PhD studies at Universitat Politecnica de Catalunya possible while working at Barcelona Supercomputing Center.

Collaborators at Barcelona Supercomputing Center: Fatima Lucas, Ferran Sancho, Gerard Santiago and Valerio Giacobelli (a visitor from Università degli Studi di Napoli). Part of the work presented in this thesis would not have been concretized without your efforts and suggestions. Special thanks to Fati, for her constant support through these years.

Collaborators from other research centers: Thomas Spiro (University of Washington), Susana Camarero and Angel Martínez (Consejo Superior de Investigaciones Científicas, Madrid), Giovanni Sannia (Università degli Studi di Napoli) and Thierry Tron (Aix Marseille Université). For sharing your data and experience with me, for validating the theoretical predictions described in this thesis.

Members of the Life Sciences department (past and present). With all of you, at some point, I had an interesting conversation, a few beers, lunch, etc.

Colleagues (past and present) that I consider friends. In particular: Armin, Didier, Fatima, Ferran, Gerard, Iain, James, Marina, Mark, Pedro and Rose. I had a great time with you guys.

People that nothing have to do with this thesis but with whom I had a good time (including the Barcelona Supercomputing Center basketball group). I spare you the list: it is too long, I will surely miss someone and I would rather not.

Silvia, my girlfriend, for her support and listening in the hard times, for keeping me in contact with the real world, for our years together.

My parents. I owe you everything.

List of frequent abbreviations

C_α: alpha carbon
PES: potential energy surface
TS: transition state
QM: quantum mechanics
TISE: time-independent Schrödinger equation
HF: Hartree-Fock
DFT: density functional theory
KS: Kohn-Sham
MM: molecular mechanics
FF: force field
PBC: periodic boundary conditions
SASA: solvent accessible surface area
QM/MM: quantum mechanics/molecular dynamics
MC: Monte Carlo
MD: molecular dynamics
CV: collective variable
ANM: anisotropic network model
ET: electron transfer
T_{DA}: electronic coupling
ΔG^o: standard free energy difference
E^o: standard reduction (or redox) potential
k_{cat}: catalytic rate constant
k_{ET}: electron transfer rate constant
K_M: Michaelis constant
k_{on}: rate of association
k_{off}: rate of dissociation
E: enzyme
S: substrate
ES: Michaelis complex
P: product
MWC: Monod, Wyman and Changeux
KFN: Koshland, Nemethy and Filmer
Hb: hemoglobin
T: tense state of hemoglobin
R: relaxed state of hemoglobin
TSA: transition state analogue
Cu: copper
T1: t1-type
T2: t2-type

T3: t3-type
TNC: trinuclear cluster
E^{o}_{T1}: standard reduction potential of the t1-type copper
$[Ru(bpy)_3]^{2+}$: tris(bipyridine)ruthenium(II)
O_2: molecular oxygen
CO: carbon monoxide
Fe: iron
TTS: tertiary two-state
SD: spin density
PcL: Pycnoporus cinnabarinus laccase
DMP: 2,6-dimethoxyphenol
ABTS: 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
RMSD: root mean square deviation
MtL: Myceliophthora thermophila laccase
SGZ: syringaldazine
4-X-PhOH: para-substituted phenol
4OH: 4-hydroxyphenol
4MeO: 4-methoxyphenol
4Me: 4-methylphenol
4Cl: 4-chlorophenol
H-bond: Hydrogen bond
ANL: aniline
PANI: polyaniline
HOMO: highest occupied molecular orbital
k_{cat}/K_M: efficiency constant
K_i: inhibition constant
BT: back-transfer
k_{BT}: back-transfer rate constant
EDTA: ethylenediaminetetraacetic acid
k_{O-O}: peroxide intermediate formation rate constant
k_{IET}: T1 copper to TNC internal electron transfer rate constant
HbCO: carboxy-hemoglobin
vFeHis: heme-histidine stretching frequency

Introduction

1. Motivations

Proteins are large macromolecules made up of 20 natural amino acids (Figure 1.1), also called the residues of the protein, linked by peptide bonds (Figure 1.2). The carboxylic and amine groups plus the carbon atom between them (the alpha carbon, C_α) form the protein backbone, the remaining atoms constitute the side chains. Proteins are responsible of a variety of biological functions, among which are storage, transport and chemical transformation of small molecules (called ligands or substrates). Their function is encoded in the amino acid sequence (primary structure), which determines short range structural patterns like helices, sheets and loops (secondary structure), their mutual position in space (tertiary structure) and how different protein chains assemble (quaternary structure) (Figure 1.3). Compared to chemical catalysts (organometallics, surfaces, etc.), proteins (enzymes) have a well-defined configuration around the active site which potentially ensures higher specificity, selectivity and turnover frequencies (TOFs); while industrial chemical catalysts display TOFs in the 10^{-2}-10^{2} s^{-1} range (Hagen, 2015), some enzymes can catalyze reactions in a diffusion-controlled regime (10^{5}-10^{6} s^{-1}) (Bar-Even et al., 2011). Chemically catalyzed reactions tend to work in organic solvents and are produced/designed through chemical reactions; proteins are soluble in water, their function can be optimized genetically by modifying an expression host (a bacteria, plant, fungi, etc.) and they are biodegradable. Therefore, proteins have a very low impact on the environment and are ideal templates for the creation of novel green catalysts and therapeutics. Indeed, industrial proteins have a great market which, focusing on enzymes, it is expected to be worth 6.2 billions of USD by 2020 according to MarketsAndMarkets (http://goo.gl/e1wePh).

Notwithstanding these remarkable properties, proteins, as found in nature, are usually not ready for industrial scale-up. In fact, physico-chemical conditions in a batch reactor can significantly differ from their natural counterpart in terms of temperature, pressure, pH and chemical environment (co-solute composition). On top of that, the reaction of interest may not have a natural enzyme or it may have one with an unsatisfying kinetics. As a matter of fact, the rate constant (k_{cat}) of the average enzyme is around 10 s^{-1}, far away from the values associated to diffusion-controlled kinetics (Bar-Even et al., 2011), suggesting that enzymes evolved to just accomplish their natural task fairly, leaving room to improve their catalytic power and fulfil the industrial needs.

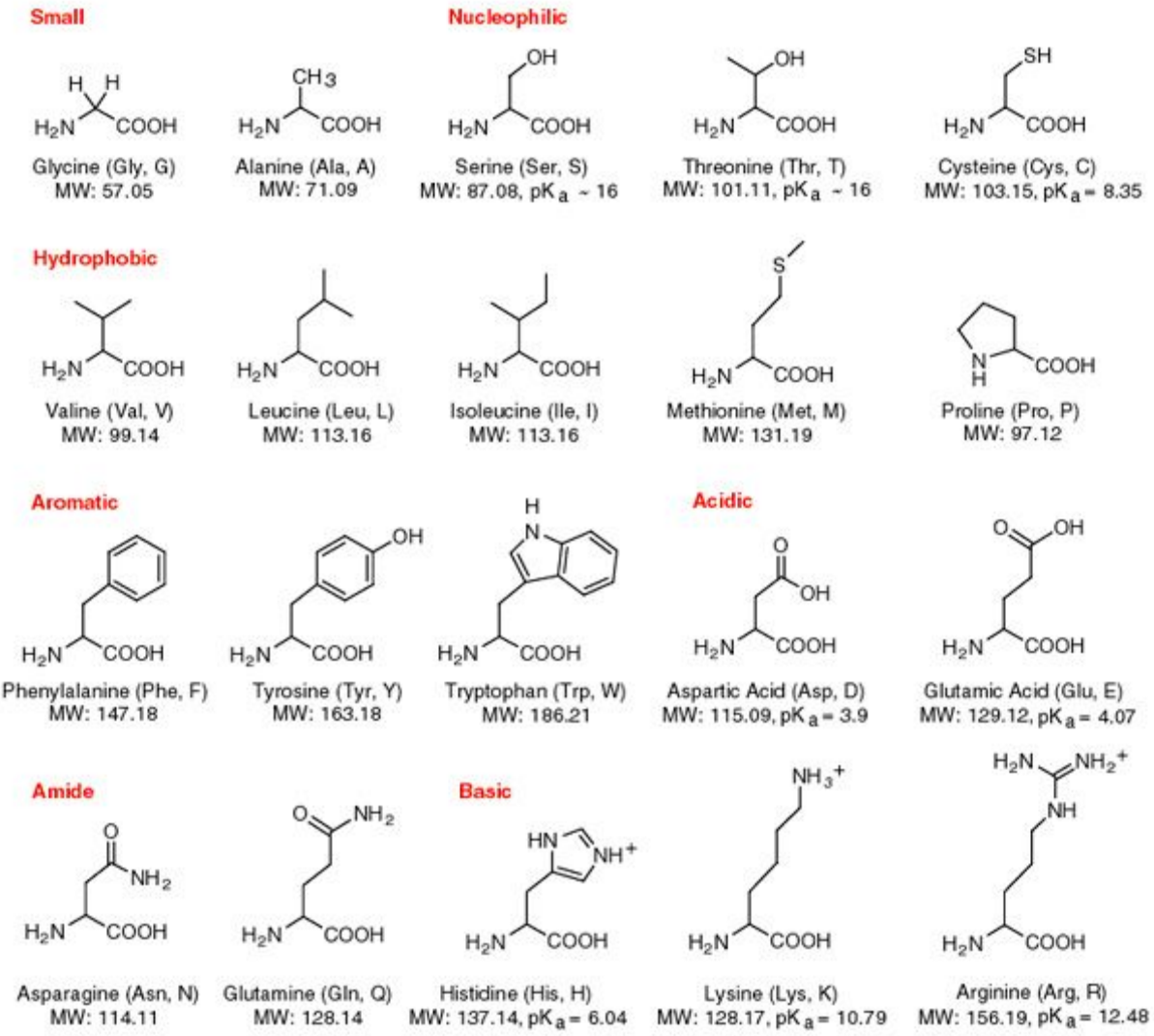


Figure 1.1. Structure of the 20 natural amino acids. MW is the molecular weight, pKa is the acidic dissociation constant of the side chain. Source: https://goo.gl/BL4akQ.

The evolvability of enzymes has been successfully exploited to re-design several proteins (Gerlt & Babbitt, 2009) or create new ones from scratch (*de novo* design) (Kiss et al., 2013). Although *de novo* design is a vibrating research field with the potential to revolutionize the future of catalysis, this thesis is focused on re-designing existing proteins. From now on, the words design and engineer should be understood as re-design and re-engineer.

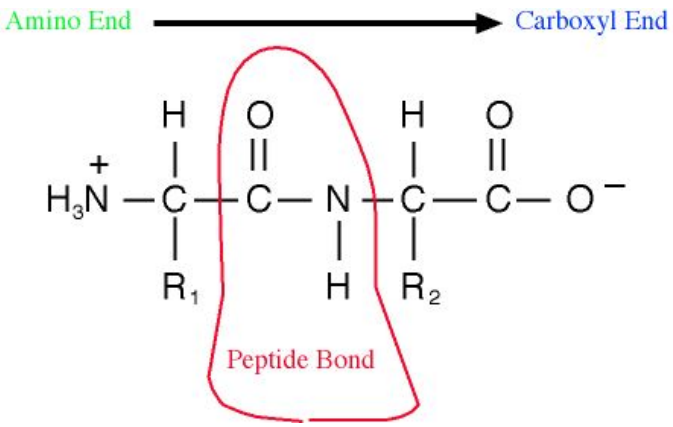


Figure 1.2. Chemical sketch of a peptide bond in a dipeptide (complex of two amino acids). Source: http://goo.gl/PwUhte.

Proteins are improved in the lab following three different strategies: rational design, directed evolution and semi-rational design. In rational design, a few site-directed mutations are introduced based on structural and mechanistic knowledge of the system. The main limit of this approach is the difficulty of predicting the effect of a point mutation even if the system is very well known. Oppositely to rational design, directed evolution randomly introduces mutations in a protein, not requiring a thorough understanding of the system (Romero & Arnold, 2009). The advantage of directed evolution is the potential discovery of beneficial mutations which could not be predicted rationally, thanks to its random and high-throughput nature. On the other hand, the fraction of sequence space explored is very tiny: for a N residues protein 10^3-10^4 variants out of 20^N are explored at most. Thus, the rate of beneficial mutations is low. In semi-rational design, the drawbacks of the two previous approaches are addressed focusing mutagenesis at positions where, according to prior knowledge and/or computational techniques, the rate of beneficial mutations should be higher (Chica, Nicolas & Pelletier, 2005). More specifically, these positions are subjected to saturation mutagenesis (SM, all 20 residues are tested for a given position), combinatorial SM (given N positions all the 20^N combinations are tried) or iterative SM (Reetz & Carballeira, 2007) (combinatorial SM is iteratively carried out on regions of maximum three residues).

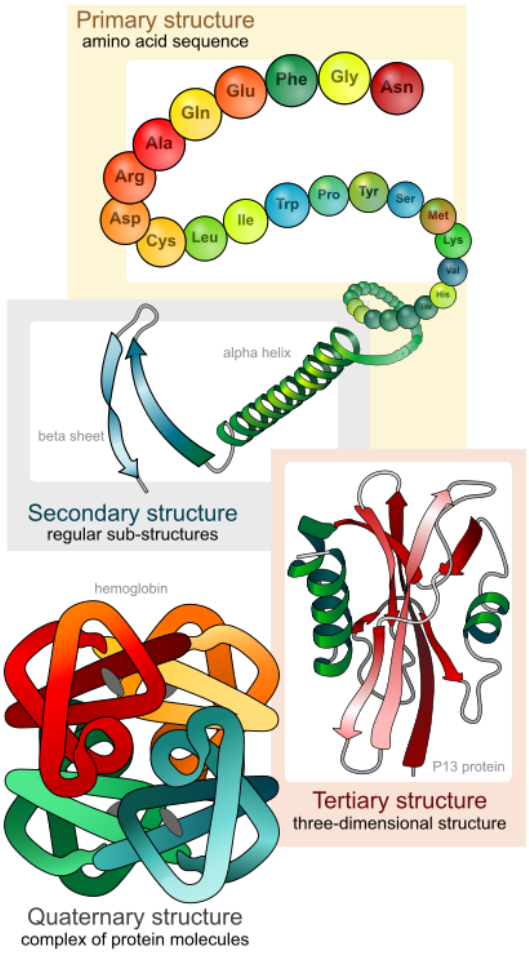


Figure 1.3. Structural organization of proteins. Source: https://goo.gl/FSIqpK.

Molecular simulations, which are already routinely used in drug discovery (Sliwoski et al., 2014), are in a unique position to drive rational and semi-rational protein design; their physics-based nature provides out-of-the-box knowledge on the system which allows to guide the *in vitro* search for new functionality. A great example is the recent work of Moroz et al. where a non catalytic protein was turned into a Kemp eliminase or an ester hydrolase with one point mutation (Moroz et al., 2015). The take home message is powerful: desired reactivity may be one mutation away from a (suitably chosen) inactive protein and molecular simulations can find it or, at least, narrow down its search in the lab. In Chapter 2, the theory underlying molecular simulations is introduced and its use in protein design is illustrated in Chapter 3.

In this thesis, computational techniques are systematically employed to guide present and future design efforts of laccases (Giardina et al., 2010) and hemoglobin (Bellelli, 2010).

These are two metalloproteins, i.e. proteins that contain a metal ion as cofactor. Laccases are copper-containing oxidoreductases with very broad specificity which have a role in several industrial applications such as textile and food industry, bioremediation, forestry and organic synthesis (Riva, 2006). Hemoglobin, an iron-containing protein that carries oxygen from the lungs to the tissue in the body, is a candidate for blood substituent design (Fronticelli et al., 2007). Both metalloproteins are reviewed in Chapter 4.

REFERENCES:

Bar-Even A., Arren B-E., Elad N., Yonatan S., Wolfram L., Dan D., Tawfik DS., Ron M. 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* 50:4402–4410.

Bellelli A. 2010. Hemoglobin and cooperativity: Experiments and theories. *Current protein & peptide science* 11:2–36.

Chica RA., Nicolas D., Pelletier JN. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current opinion in biotechnology* 16:378–384.

Fronticelli C., Clara F., Koehler RC., Brinigar WS. 2007. Recombinant Hemoglobins as Artificial Oxygen Carriers. *Artificial cells, blood substitutes, and immobilization biotechnology* 35:45–52.

Gerlt JA., Babbitt PC. 2009. Enzyme (re)design: lessons from natural evolution and computation. *Current opinion in chemical biology* 13:10–18.

Giardina P., Faraco V., Pezzella C., Piscitelli A., Vanhulle S., Sannia G. 2010. Laccases: a never-ending story. *Cellular and molecular life sciences: CMLS* 67:369–385.

Hagen J. 2015. *Industrial Catalysis: A Practical Approach*. John Wiley & Sons.

Kiss G., Çelebi-Ölçüm N., Moretti R., Baker D., Houk KN. 2013. Computational enzyme design. *Angewandte Chemie* 52:5700–5725.

Moroz YS., Dunston TT., Makhlynets OV., Moroz OV., Wu Y., Yoon JH., Olsen AB., McLaughlin JM., Mack KL., Gosavi PM., van Nuland NAJ., Korendovych IV. 2015. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *Journal of the American Chemical Society* 137:14905–14911.

Reetz MT., Carballeira JD. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols* 2:891–903.

Riva S. 2006. Laccases: blue enzymes for green chemistry. *Trends in biotechnology* 24:219–226.

Romero PA., Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nature reviews. Molecular cell biology* 10:866–876.

Sliwoski G., Kothiwale S., Meiler J., Lowe EW Jr. 2014. Computational methods in drug discovery. *Pharmacological reviews* 66:334–395.

2. Theory

In this chapter the theory necessary to understand the content of this thesis is introduced. Firstly, basic thermodynamics notions, the concept of potential energy surface and their connection are summarily explained. Then, some methods to explore the potential energy surface and calculate its value in a given point are described. In the last section, some useful protein-related theories are introduced.

2.1 A bridge between microscopic states and macroscopic variables

Entropy (S) is the measure of the number of microstates accessible to the system, intuitively associated to its "degree of disorder". The second law of thermodynamics establishes that for an isolated system (where no exchange of matter and energy with the surrounding is allowed) its entropy (S_u) cannot decrease (Atkins, 1998a). When the number of particles (N), temperature (T) and pressure (p) are fixed (like in a living being or a chemical reactor) and no external work is applied on it, the entropy of the universe (system+surroundings, which is isolated by definition) coincides with the opposite of the Gibbs free energy (G) of the system multiplied by its temperature. For a molecular transformation (i.e. a chemical reaction) it can be shown that:

$$-\Delta S_U / T \equiv \Delta G = \Delta G^0 + RT\,lnQ \geq 0 \qquad \text{Eq. 2.1}$$

where R is the gas constant, T is the temperature of the system and Q is the ratio of the concentrations of the end-points of such transformation. ΔG° is the standard free energy difference between the end-points in their pure state. By standard, it is meant that the end-points are in their standard states. When a transformation reaches equilibrium the free energy no longer changes, therefore:

$$\Delta G^0 = -RT\,lnK \qquad \text{Eq. 2.2}$$

where K is Q at equilibrium and it is called the equilibrium constant. It follows that ΔG° is a rigorous measure of the spontaneity of a transformation at a given temperature T. If $\Delta G^\circ<0$, the

final state of the transformation is more stable and the process is said exergonic; if the opposite is true the transformation is called endergonic. The Gibbs energy of the system is defined as:

$$G = H - TS \quad \text{Eq. 2.3}$$

where *H* is the enthalpy, the internal energy of the system at constant pressure, and *S* is the entropy of the system. The lower the potential energy and the higher the entropy of the system, the more stable it is. The main advantage of *G* over S_U is that the spontaneity of a transformation is related to system's quantities with sound physical interpretation (enthalpy and entropy), instead of referring to the universe.

The potential energy surface (PES) of the system defines the energy as a function of its degrees of freedom (Figure 2.1) (Leach, 2001). If the number of particles, volume (*V*) and temperature are fixed, the average Helmholtz free energy (*A*) of the system can be calculated through the canonical partition function (*Z*) (Hill, 2012), provided that the PES is known:

$$Z(N,V,T) = \sum_i \pi_i = \sum_i e^{-\frac{E_i}{k_B T}} \quad \text{Eq. 2.4}$$

$$\overline{A} = -RTlnZ(N,V,T) \quad \text{Eq. 2.5}$$

where π_i and E_i are the probability density and energy of the *i-th* microstate, k_B is the Boltzmann constant (the gas constant R for one molecule instead of one mole). Analogously to *G* for an isothermal-isobaric system, *A* defines the criteria of spontaneity of a transformation in a isothermal-isochoric system. *G* and *A* have the following relationship:

$$G = A + pV \quad \text{Eq. 2.6}$$

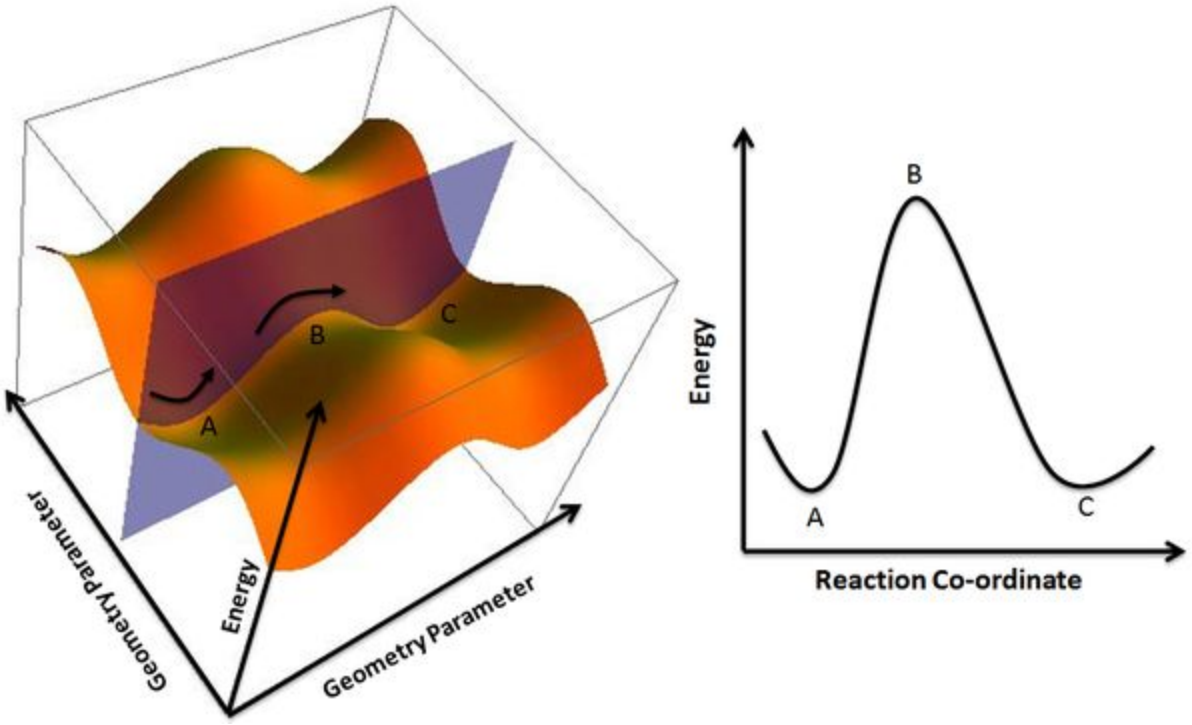


Figure 2.1. PES and a minimum energy path connecting reactants and products (A and C) minimum through a saddle point (B). Source: https://goo.gl/SAzHGv

In solution, if no liquid-gas phase transition occurs, V is virtually constant, therefore:

$$\overline{G} \approx p\overline{V} - RTlnZ(N, \overline{V}, T) \qquad \text{Eq. 2.7}$$

Moreover, since the volume of a solution does not change significantly along a molecular transformation:

$$\Delta\overline{G^0} = -RTlnK \approx -RTln\ \frac{Z_{final}(N,p^0,T)}{Z_{initial}(N,p^0,T)} \qquad \text{Eq. 2.8}$$

Therefore, the thermodynamical equilibrium of a transformation can be in principle assessed from the PES through the canonical partition function.

Knowing the PES gives access to kinetic properties as well. For an elementary step A + B → C, the rate constant *k*, the rate at unitary concentrations of A and B, can be calculated with Eyring's equation (Atkins, 1998b):

$$k = \kappa \frac{k_B T}{h} exp\left(-\frac{\Delta G^{\#}}{RT}\right)$$ Eq. 2.9

where h is the Planck constant, κ is the transmission coefficient (generally assumed to be unitary) and $\Delta G^{\#}$ is the activation free energy, i.e. the free energy difference between the transition state (TS) and the reactants (the free energy barrier). The TS is the saddle point of the minimum energy path that connects the reactant and product energy minima on the PES. It should be noted that Eyring's equation is derived with several assumptions: (i) the TS is stable enough to form an equilibrium with the reactants; (ii) nuclear tunnelling is negligible (not valid for proton transfer or very low barriers); (iii) the reaction passes always through the TS (can be false at relatively high temperatures). Assumptions (i) and (iii) are reasonable for enzymes because they work at moderate temperatures and TS is stabilized by the protein environment. Conversely, attention should be paid to proton transfer reactions which are frequent in biology.

Unfortunately a direct estimation of the partition function is impossible for flexible macromolecules due to their high number of degrees of freedom (Leach, 2001; Frenkel & Smit, 2001). Therefore, absolute free energies are inaccessible to molecular simulations. Still, for a macroscopic variable O which does not directly depend on the value of the partition function (like potential energy, specific heat, atom-atom distances, etc.), its ensemble average can be calculated with fair accuracy by sampling the low-energy states, practically binning the simulated values with importance sampling techniques (Allen & Tildesley, 1989):

$$\overline{O} \approx \sum_n O_n f_n$$ Eq. 2.10

where O_n and f_n are the n-th bin value and frequency. Furthermore, although absolute free energies are not accessible, a number of methods have been developed to calculate the differences between two end-points and therefore their thermodynamic equilibrium from molecular simulations (Frenkel & Smit, 2001).

2.2 Quantum mechanics

The most accurate, but also computationally intensive, theoretical framework to calculate the energy of the system is quantum mechanics, where the electronic degrees of freedom are

treated explicitly. If the Hamiltonian operator (*H*) of the system does not explicitly depend on time (Messiah, 2014), the wavefunction (Ψ) and the energy (*E*) of the systems can be calculated solving the time-independent Schrödinger's equation (TISE) (Atkins & Friedman, 2011a), a partial differential eigenvalue equation:

$$H\Psi = (T_N + T_e + V_{NN} + V_{eN} + V_{ee})\,\Psi = E\Psi \qquad \text{Eq. 2.11}$$

where T_N and T_e are the kinetic energy operators of the nuclei and electrons respectively, V_{NN}, V_{eN} and V_{ee} are the operators of the nuclei-nuclei, electron-nuclei and electron-electron potential energies. According to the first postulate of quantum mechanics, the wavefunction contains all the information about the system (Atkins & Friedman, 2011b). Therefore, given a linear operator *O* corresponding to an observable, its expectation value can be estimated. In order to decrease the complexity of the TISE, the nuclei are considered fixed with respect to the electrons due to their much higher mass (~2000-fold). This is known as the Born-Oppenheimer (BO) approximation, which allows to separate the electronic from the nuclear degrees of freedom, solving the electronic part of the TISE as a parametric equation of the fixed coordinates of the nuclei (Szabo & Ostlund, 2012). Once the electronic energy is obtained, this serves as potential (the BO surface) in the nuclear part of the TISE. In such a way, an extremely complicated equation (Eq. 2.11) is broken down in two easier ones:

$$\mathcal{H}_e\,\psi_i(r;R) = (T_e + V_{ee} + V_{eN} + V_{NN})\,\psi(r;R) = U_i\psi_i(r;R) \qquad \text{Eq. 2.12}$$

$$\mathcal{H}_N\,\phi_j(R) = (T_N + U_i)\,\phi_j(R) = E_j\phi_j(R) \qquad \text{Eq. 2.13}$$

$$\Psi(r,R) = \phi_j(R)\;\psi_i(r;R) \qquad \text{Eq. 2.14}$$

where $\mathcal{H}_e$ and $\mathcal{H}_N$ are the electronic and nuclear Hamiltonians respectively, ψ_i and Φ_j are the i-th and j-th eigenvector of $\mathcal{H}_e$ and $\mathcal{H}_N$, U_i and E_j are the i-th and j-th eigenvalues of $\mathcal{H}_e$ and $\mathcal{H}_N$, while r and R indicate the electronic and nuclear coordinates. Unfortunately, Eq. 2.12 cannot be solved exactly for many electrons systems and approximated methods are necessary. In the Hartree-Fock (HF) method, the wavefunction ansatz to solve the TISE of an N electron system is the antisymmetrized product of mono-electronic wavefunctions (spin-orbitals):

$$\psi_{HF} = \sqrt{N!}\, A \prod_{i=1}^{N} \varphi_i(x_i) = (N!)^{-1/2} \sum_{P} (-1)^{\pi} P \prod_{i=1}^{N} \varphi_i(x_i) \qquad \text{Eq. 2.15}$$

$$x_i = r_i \sigma_i \qquad \text{Eq. 2.16}$$

where A is the antisymmetrizer operator, φ_i is the spin-orbital of the i-th electron, P is the permutation operator, π is the parity of the permutation and x_i is the product of the spatial (r_i) and spin (σ_i) coordinates of the i-th electron. The antisymmetrizer operator ensures that the Pauli exclusion principle is satisfied, therefore the HF ansatz (the so called Slater determinant) is a valid solution. In quantum chemistry the molecular spin-orbitals are built as linear combinations of atomic spin-orbitals, which are mono-electronic functions centered on the atomic nuclei of the system (Eq. 2.17), the so called basis set:

$$\varphi_i(x_i) = \sum c_{\mu i} \chi_{\mu}(x_i) \qquad \text{Eq. 2.17}$$

The TISE is solved exploiting the variational principle, which states that for any N-electron wavefunction (assumed to be normalized from now on) the expectation value of the energy U[ψ] is higher than the exact ground state energy U_0:

$$U[\psi] = \int \psi^* \mathcal{H} \psi d\tau \geq U_0 \qquad \text{Eq. 2.18}$$

where $\mathcal{H}$ is the electronic Hamiltonian plus the nuclei-nuclei repulsion term and ψ^* is the complex conjugate of ψ. Therefore, the TISE can be solved minimizing the expectation value of the energy with respect to a set of suitable parameters that define the ansatz. In the case of the Slater determinant, such parameters are the expansion coefficients of Eq. 2.17 and the variational principle is applied as follow:

$$\delta \mathcal{L}[\psi_{HF}] = \delta U[\psi_{HF}] - \sum_{ij} \varepsilon_{ij} \left(\int \varphi_i^* \varphi_j dx_1 - \delta_{ij} \right) = 0 \qquad \text{Eq. 2.19}$$

where δ indicates the variation of a functional, ε_{ij} is a Lagrangian multiplier and δ_{ij} is the Kronecker delta. The second term of Eq. 2.19 imposes the orthonormality of spin-orbitals. After some algebraic manipulation, the HF equation is obtained:

$$f(x_i)\varphi_i(x_i) = \varepsilon_i\varphi_i(x_i) \qquad \text{Eq. 2.20}$$

$$f(x_i) = T_e(x_i) + V_{eN}(x_i) + \sum_{j=1}^{N} J_j(x_i) - K_j(x_i) \qquad \text{Eq. 2.21}$$

where $f(x_i)$ is the Fock operator acting on the i-th electron, ε_i is the energy of the i-th spin-orbital. J and K are the coloumbic and exchange operators accounting for the electron-electron interaction between electron of same and opposite spin respectively. Inserting Eq. 2.17 in 2.20, multiplying by χ^*_v and integrating over x_i yields the Roothaan-Hall equation:

$$\sum_{\mu} c_{i\mu} \int dx_1\, \chi_v^*(x_1)\, f(x_1)\, \chi_\mu(x_1) = \varepsilon_i \sum_{\mu} c_{i\mu} \int dx_1 \chi_v^* \chi_\mu \qquad \text{Eq. 2.22}$$

Which in matrix form is:

$$FC = SC\varepsilon \qquad \text{Eq. 2.23}$$

Where F is the Fock matrix (containing the Fock operator f, Eq. 2.21), C is the matrix of the coefficients of Eq. 2.17, ε is the matrix of the orbital energies and S is the overlap matrix. The peculiarity of Eq. 2.23 is that it cannot be solved diagonalizing F since this explicitly depends on C. Therefore an iterative approach called self-consistent field procedure is employed and stopped when the wavefunction and the energy converge, i.e. their change is below a pre-defined threshold. This equation can take different forms whether a restricted (RHF) or unrestricted HF (UHF) calculation is carried out. The former, employed in closed-shell systems (with same number of spins up and down), have pairs of spin-orbitals with the same spatial component (molecular orbital, MO) and distinct spin component. This means that each MO is occupied by two electrons with opposite spin. On the other hand, UHF calculations use distinct molecular orbitals for spin-up and spin-down electrons, which implies two sets of Eq. 2.23 which in this case are named Pople-Nesbet equations. This methodology is used to deal with systems with unpaired electrons. An alternative to treat open-shell systems is to model paired and

unpaired electrons with doubly- and singly-occupied molecular orbitals (restricted open-shell HF, ROHF).

The HF method carries one fundamental approximations: the mean-field approximation is adopted for electrons of opposite spin. The difference between the exact and the HF energy arising from this approximation is the so called correlation energy. Post HF methods try to correct this gap in different ways: for example configuration interaction (CI) and coupled cluster (CC) approaches expand the HF ansatz adding more Slater determinants (although they do it in a different way) while the Møller-Plesset method employs a perturbative approach (Szabo & Ostlund, 2012). Post HF calculations are very expensive so their use is limited to small size systems. Conversely, semi-empirical calculations (Leach, 2001) introduce further approximations and empirical data, providing much quicker but less accurate solutions to Eq. 2.12 than the HF method. Their continuous development is making them popular in biomolecular simulations (Yilmazer & Korth, 2015).

A completely different approach is delineated by the density functional theory (DFT) (Koch & Holthausen, 2015). In this theoretical framework the central role of the wavefunction is undertaken by the electron density (Eq. 2.24), which measures the probability of one electron to be at a specific location. This implies two main advantages: i) the electron density is an observable and provides a direct bridge between theory and experiment and ii) the electron function is a one-body quantity of easy interpretation.

$$\rho(r_1) = \int \psi^* \psi d\sigma_1 dx_2 ... dx_N \qquad \text{Eq. 2.24}$$

Given the Hamiltonian of Eq. 2.12 (ignoring V_{NN}, that is constant for a fixed nuclear conformation), for a N-electrons system T_e and V_{ee} are invariant to the number and kind of nuclei forming the system, they are universal. On the other hand, V_{eN} is an external (to the electrons) potential (v_{ext}) which uniquely defines the system. DFT rests upon the the Hohenberg-Kohn (HK) theorems which state that:

i) The external potential and therefore the electronic energy of the system is a unique functional of the ground state electron density.

$$\rho(r) \Leftrightarrow v_{ext} \Leftrightarrow \mathcal{H}_e \qquad \text{Eq. 2.25}$$

$$U\left[\rho_0(r)\right] = F\left[\rho_0(r)\right] + \int v_{ext}\,\rho_0(r)\,dr \qquad \text{Eq. 2.26}$$

where *F[ρ(r)]* is the universal functional which includes electron-electron repulsion and electron kinetic energy.

ii) The ground state electron density can be obtained variationally.

$$U[\rho(r)] \geq U\left[\rho_0(r)\right] \qquad \text{Eq. 2.27}$$

The conceptual consequence of these statements is that the ground state electron density, a one-body quantity, uniquely defines a N-electrons system and that a variational method can be built on the energy functional of this quantity. Unfortunately, the universal functional is unknown and these theorems do not provide any practical way to compute ground state electron densities. Kohn-Sham DFT (KS-DFT) allows to circumvent this limitation borrowing some concepts from the HF method. In KS-DFT an N non-interacting electron system (described by a Slater determinant) with the same electron density of the real system is used to find the electron density variationally:

$$\rho(r) = \sum_{i=1}^{N} \left|\varphi_i(r)\right|^2 \qquad \text{Eq. 2.28}$$

Using Eq. 2.28 and rewriting the universal functional conveniently yields:

$$F[\rho(r)] = T_s[\rho(r)] + U_H[\rho(r)] + U_{XC}[\rho(r)] \qquad \text{Eq. 2.29}$$

$$T_s[\rho(r)] = \sum_{i=1}^{N} \int dr\, \varphi_i^*(r)\, T_e\, \varphi_i(r) \qquad \text{Eq. 2.30}$$

$$U_H[\rho(r)] = 1/2 \int dr \int dr' \rho(r)\rho(r')|r - r'|^{-1} \qquad \text{Eq. 2.31}$$

where $T_s[\rho(r)]$ is uncorrelated kinetic functional (known from HF theory) $U_H[\rho(r)]$ is the classical electrostatic energy functional of the non-interacting system and $U_{XC}[\rho(r)]$ is the

exchange-correlation functional which corrects the previous two functionals to the exact one. Therefore, a new Hamiltonian can be written as:

$$h = T_e + dr\rho(r')|r - r'|^{-1} + v_{ext} + v_{XC} = T_e + v_{eff} \qquad \text{Eq. 2.32}$$

$$v_{XC} = \frac{\delta U_{XC}[\rho(r)]}{\delta\rho(r)} \qquad \text{Eq. 2.33}$$

Using the orbitals of Eq. 2.28 as ansatz, the eigenvalues and eigenvectors of Eq. 2.32 can be found and the so called Kohn-Sham equation (Eq. 2.34) can be solved self-consistently:

$$h_i \varphi_i = \varepsilon_i \varphi_i \qquad i = 1, ..., N \qquad \text{Eq. 2.34}$$

As in the case of HF, the spin-up and spin-down electrons of open-shell systems are generally treated separately with two sets of Eq. 2.34. In this case a new quantity arises: the spin density, which is the difference between spin-up and spin-down electron densities.

It should be noted that although the KS equation is an exact computational tool, its resolution is necessarily approximate because the form of the exchange-correlation functional is unknown. The development of approximate functionals is an active research topic whose starting point is local-density approximation (LDA). According to LDA the exchange-correlation energy density $u_{XC}[\rho(r)]$ depends only on the local value of the electron density:

$$U_{XC}^{LDA}[\rho(r)] = \int \rho(r)\, u_{XC}[\rho(r)]\, dr \qquad \text{Eq. 2.35}$$

This is equivalent to assume that, for an N-electron system, the exchange-correlation energy density at any position is the same as it would be for a homogeneous electron gas with the same electron density of the system in that position. Obviously the electron density of a molecular system if far from being homogeneous and some corrections are needed to improve the quality of the exchange-correlation functional. The generalized gradient approximation (GGA) and meta-GGA successfully do it adding the local dependence of the energy density on the gradient (GGA) and Laplacian (meta-GGA) of the electron density at any point of the space. Furthermore, the exchange part of the exchange-correlation functional can be (partially) taken by the HF theory, whose exchange energy is exact, producing the so called hybrid functionals.

The main practical advantages (Cramer, 2013) of KS-DFT are: i) the capability to retrieve part of the correlation energy at a fraction of the computational cost of post-HF methods and ii) the good scaling with N. Indeed, KS-DFT is the go-to method for modeling large systems. Notwithstanding, KS-DFT has a few drawbacks (Cramer, 2013) such as the unclear physical meaning of the KS orbitals of Eq. 2.34, which are a mere mathematical escamotage, and the limited knowledge of density functionals to calculate observables (other than the energy) compared to the plethora of linear operators applicable to the wavefunction. Still, KS orbitals are often used as they were the physical orbitals of the N-electrons interaction and the associated Slater determinant as it were the wavefunction of the system, not without (at least qualitative) success.

An important concept in molecular systems is that of atomic charge, the amount of total charge assigned to a given atom. Atomic charges can be assigned partitioning the wavefunction, e.g. according to the number of basis set functions centered on each nuclei (Mulliken, 1955), or the electron density according to its topology (Bader, 1994). The spin can be partitioned in atomic contributions analogously. A third way to determine atomic charges is to fit them to the electrostatic potential of the system (Breneman & Wiberg, 1990; Bader, 1994). In molecular mechanics, such atomic charges are assumed to be concentrated in the nuclear position and are therefore called point charges.

2.3 Molecular mechanics

Although QM is a powerful theory, its use is impractical when the system to model is a large and flexible biomolecule like a protein due to its high computational cost. Molecular mechanics (MM) expresses the potential energy of the system as a function of the nuclear degrees of freedom, without explicitly modeling the electronic degrees of freedom. Such a function, called the force field (FF), is fitted over a set of experimental and/or QM data (Leach, 2001). Since electrons are ignored, the BO approximation is implicit: no vibronic coupling and PES inter-crossing can be modeled with a traditional FF (alternative FFs are in hand to model PES inter-crossing though (Kamerlin & Warshel, 2011)). The main concept beyond FFs is transferability: functional groups (e.g. -CHO) maintain their properties in different molecules regardless of the chemical environment. This assumption finds its justification in several experimental data such as the linear dependence of the heat of formation of linear alkanes with the number of carbon atoms (Cramer, 2013). A typical FF, as shown in Figure 2.2, includes

bonded and non-bonded interactions. For bonded interactions, stretching and bending energies are represented as harmonic functions, while torsional energies are written as Fourier series to ensure periodicity. For non-bonded interactions, the electrostatic energies are modeled with a coulombic interaction between atomic point charges, and van der Waals interactions, which accounts for nuclear repulsions and dispersion forces, have a Lennard-Jones potential form.

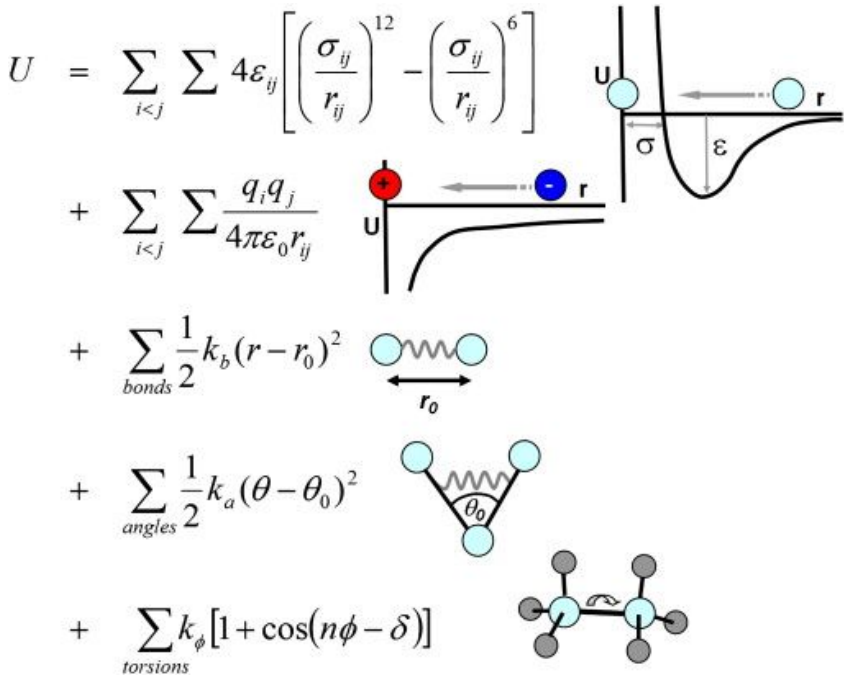


Figure 2.2. Typical FF. Source: http://goo.gl/wEQJn1.

A FF may also include other terms such as improper torsions to ensure planarity in certain systems, cross terms to model coupling effects between e.g stretching and bending, explicit hydrogen bond (H-bond) terms, etc. Whatever form of the FF is chosen, it: i) must be continuous and derivable to make force calculations possible; ii) should include enough transferable parameters to have broad generality and; iii) should limit parameters' redundancy to be as general as possible, allowing fast energy calculations. It is worth noting that a FF like the one in Figure 2.2 has fixed valence, meaning that no bonds can be formed or broken. Therefore, such scheme is not suitable for chemical reaction modeling (but alternatives exist (Kamerlin & Warshel, 2011; Senftle et al., 2016)). Moreover, all parameters are fixed for the principle of transferability, making them insensible to the environment (no polarization effect, although polarizable FFs are being developed (Baker, 2015)).

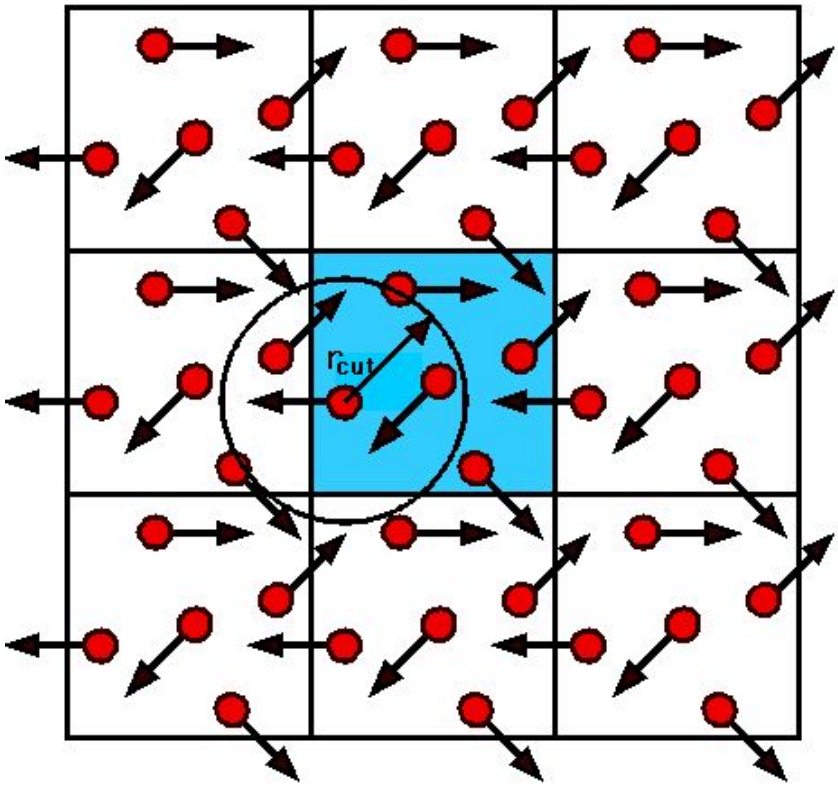


Figure 2.3. Two-dimensional scheme of PBCs. The blue box is the unit cell, the red spots are the system's molecules with their velocity vectors. Source: http://goo.gl/G55t8T.

The solvent, which in protein modeling is usually water, can also be modeled explicitly through a dedicated FF. The main drawback is the huge number of solvent molecules needed to reproduce its bulk behavior. A way to significantly reduce this number is the use of the periodic boundary conditions (PBCs). When PBCs are applied, the system box (the unit cell) is infinitely replicated in space with rigid translations. In order to avoid unphysical behavior, the non-bonded interactions are subjected to a cut-off radii (r_{cut}) such that an atom interacts only with the closest images of the remaining atoms (minimum-image convention, Figure 2.3).

For macromolecular simulations, it is recommendable to place at least 1 nm of solvent buffer around the solute (de Souza & Ornstein, 1997). The introduction of explicit solvent molecules, even within PBCs, severely increases the computational cost of a simulation. Implicit solvation models are therefore used to average out the solvation free energy of the system (Cramer, 2013), which has three contributions (Eq. 2.36): electrostatic, van der Waals and cavitation.

$$\Delta G_{ele} + \Delta G_{vdW} + \Delta G_{cav} = mSASA + q \qquad \text{Eq. 2.36}$$

The latter is the free energy penalty associated with the formation of a solute cavity within the solvent. The non-electrostatic terms are usually modeled as linear function of solvent-accessible surface area (SASA), which is the area of the solute reachable by the solvent molecules (Eq. 2.36). The rationale behind the linear relationship is the substantial dependence of this term on the number of solvent molecules in the first solvation shell, which is proportional to SASA.

$$\Delta G_{vdW} + \Delta G_{cav} = mSASA + q \qquad \text{Eq. 2.37}$$

Modeling the electrostatic part is mathematically more laborious. The Poisson-Boltzmann (PB) method provides a fairly accurate solution but it requires the (expensive) self-consistent numerical solution of the linearized form of the homonymous equation:

$$\nabla \bullet \varepsilon(r) \nabla \phi(r) - 8\pi I(r)\phi(r) + 4\pi\rho(r) = 0 \qquad \text{Eq. 2.38}$$

where $\varepsilon(r)$, $\Phi(r)$, $I(r)$ and $\rho(r)$ are the dielectric constant, electric potential, ionic strength and charge density at the position r respectively. A less expensive approach is furnished by the generalized Born (GB) equation, in which the system is modeled as a collection of particles with charge q_i and radii a_i in a medium of relative dielectric constant ε. The solvation free energy of such system is the sum of the Coulomb energy in the medium and the Born solvation energy minus the Coulomb energy in vacuo:

$$\Delta G_{ele} = (1/\varepsilon - 1) \sum_{i>j} q_i q_j r_{ij}^{-1} + 0.5 (1/\varepsilon - 1) \sum_i q_i^2 a_i^{-1} \qquad \text{Eq. 2.39}$$

Although implicit solvent models greatly speed up the calculations, they present some limitations such as the lack of solvent viscosity (which is also an advantage in terms of conformational space exploration rate) and the total neglect of hydrogen (H-bonds) between the solute and the solvent (if it has H-bond donor/acceptors like water does). The last point is especially limiting since water might have a primary role in some structural and dynamical features of a protein.

2.4 Quantum mechanics/molecular mechanics

So far, it has been established that QM is confined to accurately describing small/medium sized systems while MM can be used to deal with large systems such as proteins at the expenses of accuracy. Hybrid QM/MM methods (Warshel & Levitt, 1976; Senn & Walter, 2009; Cramer, 2013) provide a clever way to model chemical reactivity in large systems. The reactive part is modeled with QM and embedded in the FF of the non-reactive part (Figure 2.4).

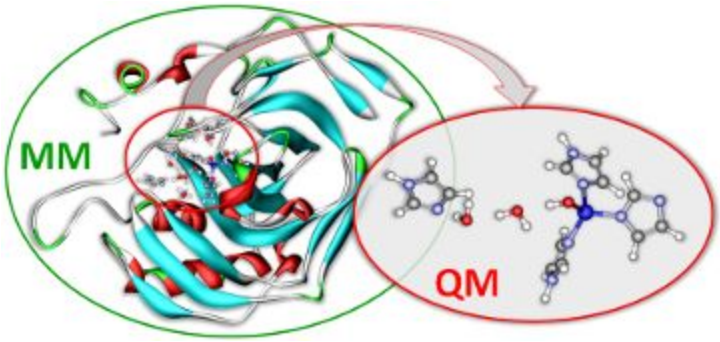


Figure 2.4. Schematic description of a QM/MM calculation. Source: http://goo.gl/VH7yly.

In additive QM/MM schemes, which are used in this thesis, the energy of the system (U) is the sum of the energies of the QM (U_{QM}) and MM (U_{MM}) regions plus the energy of the frontier ($U_{QM/MM}$), which determines how the two regions interact:

$$U = U_{QM} + U_{MM} + U_{QM/MM} \qquad \text{Eq. 2.40}$$

If the QM region is not covalently bonded to the MM part, the Hamiltonian of the frontier part can be written as the sum of a classic LJ potential (which requires to estimate the relative parameters for the QM region) and an electrostatic part. The latter can include a polarization scheme for the QM electron density (if not, the electrostatic part is a Coulomb interaction between point charges). In such case the electrostatic part is further divided into a classical nuclear-nuclear and an electron-electron repulsion term:

$$\mathcal{H}_{QM/MM} = V_{vdw} + V_{NN} + V_{ee} \qquad \text{Eq. 2.41}$$

$$V_{ee} = \sum_{i}^{QM} \sum_{I}^{MM} q_I / r_{iI} \qquad \text{Eq. 2.42}$$

Things get more complicated when the QM region is covalently attached to the MM part. In such case, several schemes have been proposed using link atoms or frozen orbitals. In this thesis, the link atom approach is employed, capping the dangling bonds with hydrogen atoms. The QM calculation is carried out on the capped system while the stretching, bending and torsional terms involving both QM and MM atoms are treated with MM. The fictitious angle between the frontier atoms and the relative hydrogen capped are strongly constrained to avoid artifacts. Finally, the MM charges which are very close to the QM region are replaced by spherical gaussian charge distributions to avoid an artificial overpolarization of the QM electron density in proximity of the frontier region.

2.5 Geometry optimization and normal modes

Whatever method is used to quantify the energy, conformational sampling is necessary to represent the system. The local minima of the PES are the points that mostly contribute to the canonical partition function. Therefore, energy minima are highly representative elements of the system. The energy can be minimized numerically through an iterative geometry optimization until a convergence criteria, typically the root mean square deviation of the gradient being below a small threshold, is met. The opposite sense of the energy gradient is used to guide the minimum search, like in the steepest descent and the conjugated gradient method (Leach, 2001). In addition to that, the inverse of the Hessian matrix (that contains all the second derivatives) needs to be calculated in the Newton method (Leach, 2001). It should be noted that energy minimization algorithms per se do not guarantee to land in an energy minima but only in a stationary point (but not in an energy maximum). Therefore, the nature of such stationary point can be assessed through the diagonalization of the Hessian matrix: if no imaginary frequency is found, the structure is indeed a local energy minimum, otherwise it is a saddle point. If only one eigenvalue is imaginary, the point corresponds to a TS (specific TS search requires different approaches which are not covered here).

The diagonalization of the Hessian matrix, an expensive calculation, provides more information than the nature of the stationary point found: the resulting eigenvectors are called the normal modes (since they are orthogonal and therefore independent). Their motion describes vibrational fluctuations around the energy minimum conformation and their eigenvalues correspond to the vibrational frequencies. Therefore, the vibrational spectra of a

molecule can be simulated, bearing in mind that the energy function is assumed to be quadratic around the minimum. On top of that, the normal modes can be used to quantify the vibrational partition function, allowing to calculate the free energy of the system in a minimum according to the rigid-rotor-harmonic-oscillator framework (RRHO) (Hill, 2012; Jensen, 2015).

2.6 Monte Carlo and Molecular dynamics simulations

If the system is at equilibrium, the net flux of transitions entering and leaving any state *a* must be null (balance condition) (Krauth, 2006):

$$\sum_{b \neq a} \pi_a p_{ab} = \sum_{b \neq a} \pi_b p_{ba} \qquad \text{Eq. 2.43}$$

where π_a is the equilibrium probability of the state *a* and p_{ab} is the probability to jump from state *a* to *b*. It follows that in the attempt to explore the conformational space of a protein hopping sequentially from a state to another, Eq. 2.43 must be satisfied in order to sample the equilibrium distribution of the system. A sufficient but not necessary solution is the satisfaction of the detailed balance condition for every pair of states *(a,b)* of the system:

$$\pi_a p_{ab} = \pi_b p_{ba} \qquad \text{Eq. 2.44}$$

The transition probability p_{ab} is the product of the probability to try such transition (α_{ab}) multiplied by the probability to accept it (A_{ab}):

$$p_{ab} = \alpha_{ab} \cdot A_{ab} \qquad \text{Eq. 2.45}$$

Assuming that α_{ab} is symmetric, Eq. 2.44 can be rearranged as:

$$p_{ab}/p_{ba} = A_{ab}/A_{ba} = \pi_b/\pi_a = exp(-\beta \Delta E) \qquad \text{Eq. 2.46}$$

where ΔE is the energy difference between the final and the initial state. Therefore, whatever form of A_{ab} is picked, it must satisfy Eq. 2.46. The Metropolis criterion makes the following choice:

$$p_{ab} = \alpha_{ab} \cdot min\{1, exp(-\beta \Delta E)\}$$ Eq. 2.47

In Metropolis Monte Carlo the equilibrium distribution is sampled through a trajectory in system's conformational space generated by random moves. A move is accepted straight away if ΔE is negative, otherwise a random number is generated: if this is lower than $exp(-\beta \Delta E)$ the transition is accepted, otherwise it is rejected. It should be noted that the only constraint on α_{ab} is that it must be symmetric.

Molecular dynamics (MD) simulates the time evolution of the system according to Newton's second law of motion (Tuckerman, 2010). At each timestep the PES value and its gradient are evaluated, then the latter is used to calculate the forces acting on the nuclei propagating system's dynamics in time:

$$a_I(t) = -\nabla_I U / m_I$$ Eq. 2.48

where m_I is the mass and $a_I(t)$ the acceleration of the I-th atom at time t. For proteins, only MM energy calculations allow to cover ns-ms timescales while QM/MM potentials can be used to reproduce 10s-100s of ps, limiting their application. In order to sample the equilibrium distribution of a NpT system, p and T need to be constant across the simulation by coupling the system to a barostat and a thermostat. This is equivalent to rescale coordinates and velocities according to user-defined coupling constants. In alternative, barostat and thermostat can be directly included in the the systems adding further degrees of freedom (Leach, 2001; Frenkel & Smit, 2001).

A powerful method used in this thesis to accelerate MD and estimate the free energy surface (FES) of the system is metadynamics (Laio & Parrinello, 2002; Laio, Alessandro & Gervasio, 2008). A reduced set of collective variables (CVs) is chosen to describe the free energy surface (FES) of the system and a bias potential (V_G), which is a gaussian function of the CVs, is periodically added to push the system to sample other CV values. When the simulation is converged, all the free energy minima are filled and the modified FES is flat. The sum of all the gaussian added during the simulation approaches the negative image of the FES. To yield an accurate FES estimation, the CVs must clearly distinguish between relevant states of the

system (e.g. folded and unfolded states), including the slow events relevant to the process of interest. Then, no more than two or three CVs should be used to efficiently sample the FES.

MC and MD are different approaches to sample the equilibrium distribution of the system. When choosing between them, it should be taken into account that: (i) MD reproduces system's time evolution while MC does not; (ii) while MD naturally couples with explicit solvent models, MC generally uses implicit solvent schemes which might limit its accuracy; (iii) MC allows to build creative (and non-physical) moves which can improve sampling while MD is in this sense an inefficient sampling tool; (iv) even MD has a random component in its initial conditions (atomic velocities, assigned through Maxwell's distribution), therefore more than one simulation should be carried out (as for MC) unless the simulation time is significantly larger than the slowest event studied (so the system "has no memory" of its initial conditions).

2.7 Protein-ligand docking and PELE

Protein-ligand docking techniques are used to find the most stable, i.e. with lowest free energy, conformation of a protein-ligand system. The free energy is evaluated with a scoring function which can be physics-based (a FF), knowledge-based, empirical, or even derived from machine learning. A knowledge-based function judges the protein-ligand contacts according to statistics derived by a structural database while an empirical scoring functions assigns positive or negative contributions to contacts depending on their nature (hydrophobic-hydrophobic, hydrophilic-hydrophilic, H-bonds, etc.). In this thesis, two softwares are used: Glide (Friesner et al., 2004; Halgren et al., 2004) and Autodock Vina (Trott & Olson, 2010). In both approaches the ligand is flexible and the backbone is frozen, while side-chain flexibility can be included in Autodock Vina. In Glide, ligand conformation and orientation is exhaustively searched through the active site; each pose is evaluated with physics-based and empirical terms. In Autodock Vina, on the other hand, the active site is explored through Metropolis MC sampling, using a mixed scoring function which aims to take advantage from both empirical and knowledge-based scoring functions. The main drawback of these approaches is the total or partial neglect of protein flexibility. Although it is customary to picture proteins with a single structure, most likely due to the influence of X-ray diffraction structures (Drenth, 2013), these are better represented with ensembles of structures (Orozco, 2014). Therefore, ligands likely interact with a plethora of different protein conformations and select the ones that minimize the binding free energy (conformational selection). On top of that, after selecting such conformations, ligands might

directly induce some changes on protein's structure and dynamics (induced-fit). For these reasons docking can be done on an ensemble of structures (experimentally determined or generated by MD or MC), simulating conformational selection, followed by further sampling to model induced-fit.

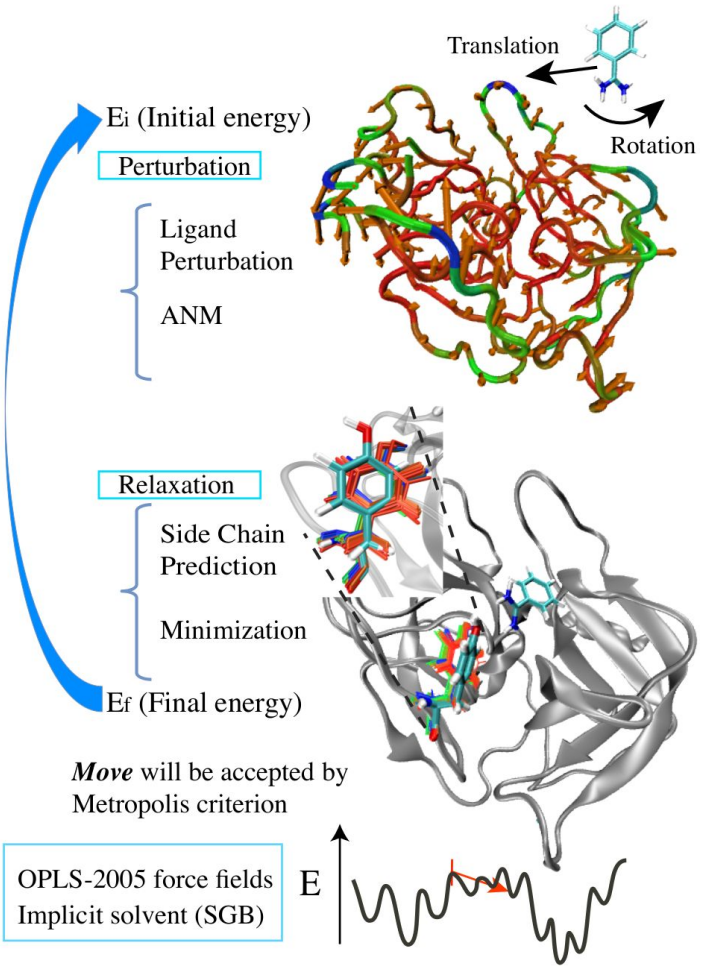


Figure 2.5. Schematic description of PELE workflow.

Compared to the above-mentioned methods, the protein energy landscape exploration (PELE) software (Borrelli et al., 2005; Madadkar-Sobhani & Guallar, 2013) aims to thoroughly sample the protein-ligand conformational space, modeling conformational selection and induced-fit on the fly. At each simulation step, the system is perturbed roto-translating the ligand and displacing the protein C_α along a linear combination of low-frequency normal modes. In particular, such normal modes are not calculated from the all-atom energy of the system, which is evaluated with a OPLS 2005 force field (Kaminski et al., 2001) and surface-generalized Born

solvent model (Bashford & Case, 2000), but with a simplified anisotropic network model (ANM) (Atilgan et al., 2001). In ANM, the system is described as a network of elastic forces connecting the C_α; such elastic force decay as the reciprocal of the interatomic distance and they become null after a certain distance cut off. PELE's perturbation phase is followed by side chain prediction (Jacobson et al., 2002) and all-atom energy minimization. Side chain prediction consists in a discrete optimization of the rotamers (dihedral angle conformers) among a set of possible solutions derived from PDB statistics, while all-atom minimization is carried out with a truncated Newton algorithm. After structural perturbation and relaxation, the Metropolis criterion is used to accept or reject the proposed move. A resuming scheme of PELE's workflow is depicted in Figure 2.5.

Although at a much higher computational cost than most docking algorithms, PELE performs fully flexible protein-ligand docking, allowing to model induced-fit (Borrelli, Cossins & Guallar, 2010) and conformational selection effects (Cossins, Hosseini & Guallar, 2012). On top of that, PELE can efficiently map ligand migration pathways in the protein matrix (Lucas & Guallar, 2012; Hosseini et al., 2015). Compared to MD, PELE can exhaustively search the active site blindly, i.e. with no prior knowledge of its location, in a much lower time, and eventually the FES can be mapped with Markov state models (Takahashi, Gil & Guallar, 2014). On the other hand, the usage of an implicit solvent draws some limitations because: (i) each MC step involves all-atom minimization which in the long run can exaggerate protein packing and close important channels and (ii) water molecules can have a direct role in binding. For both cases, the inclusion of conserved structural waters can alleviate these (potential) problems. In particular for point (ii), an explicit solvent description can be retrieved using MD on the most important poses in a second stage.

2.8 Protein-related theories

In this section the theories related to enzyme kinetics, electron transfer rate and cooperative binding are introduced. Such theories are fundamental to understand the thesis work done on laccases and hemoglobin.

2.8.1 Electron transfer in proteins

According to Marcus theory, the rate of an electron transfer (ET) reaction depends on three parameters: (i) the tunnelling probability between electron donor and acceptor

wavefunctions (T_{DA}, the electronic coupling); (ii) the ET standard free energy ($\Delta G°$, also called the ET driving force); and (iii) the energy cost to distort the geometry of reactants (solvent included) into that of the products (or vice versa) prior to ET (λ, the reorganization energy) (Marcus, 1993):

$$k_{ET} = \frac{4\pi^2 |T_{DA}|^2}{h\sqrt{4\pi k_B T \lambda}} \; exp\left(-\frac{(\Delta G^o + \lambda)^2}{4 k_B T \lambda}\right) \qquad \text{Eq. 2.49}$$

It follows from Eq. 2.49 that the ET rate is a gaussian function of the ET driving force, with a maximum when $\Delta G° = -\lambda$.

The driving force depends on the standard reduction potential (E^o) of the oxidized donor and of the acceptor. Given a semi-reaction (Eq. 2..50 and 2.51), E^o measures the tendency of the reactant to be reduced to the product: the greater E^o the more favorable the reaction (Atkins, 1998c). As shown in Eq. 2.52, E^o is the negative value of the ET free energy divided by the Faraday constant (F). It should be noted that the ΔG^o of a semi-reaction cannot be measured, but it can be estimated if a proper reference, whose ΔG^o is null by convention, is provided.

$$D^+ + e^- \rightarrow D, \;\; E^o(D^+/D) \qquad \text{Eq. 2.50}$$

$$A + e^- \rightarrow A^-, \;\; E^o(A/A^-) \qquad \text{Eq. 2.51}$$

$$\Delta G^o = -F\left[-E^o(D^+/D) + E^o(A/A^-)\right] \qquad \text{Eq. 2.52}$$

The reorganization energy has inner- and outer-sphere contributions; the former depends on donor and acceptor degrees of freedom, the latter on those of rest of the system (other amino acids and solvent). MD simulations of azurin, a copper-containing protein, revealed a direct relationship between the outer-sphere reorganization energy and: (i) active site solvent exposure (also found in cytochrome c (Bortolotti et al., 2011); (ii) and the extent of reorientation of the active site and the waters around it upon reduction (Paltrinieri et al., 2013).

The pre-exponential factor depends on the magnitude of the electronic coupling, defined as the overlap integral of donor and acceptor wavefunctions. This, for relative large donor-acceptor distances typically observed in proteins, mostly depends on the nature of the tunnelling barrier (vacuum < H-bond < covalent bond) and it decays exponentially with the donor-acceptor distance (Eq. 2.53):

$$T_{DA} = \int \psi_D^* V \psi_A d\tau \approx \prod_i h_i \approx T_{DA}^0 \, exp\left[-\alpha(R - R_0)\right] \qquad \text{Eq. 2.53}$$

where V is the perturbation operator describing ET, h_i is the electronic coupling of the i-th tunnelling barrier, R is the donor-acceptor distance, $T°_{DA}$ is the electronic coupling value at the reference distance R_0, and α is a positive number. If T_{DA} is too large, Marcus theory no longer applies; generally the donor-acceptor separation is large enough in proteins to ensure its validity and laccases, the enzymes studied in this thesis, are not an exception (Tadesse et al., 2008).

2.8.2 Enzyme kinetics

The rate (v) of an enzymatic reaction E+S ⇄ ES→P can be modeled by the Michaelis-Menten equation (Atkins, 1998c):

$$v = \frac{k_{cat}[E]_0[S]}{K_M + [S]} = \frac{v_{max}[S]}{K_M + [S]} \qquad \text{Eq. 2.54}$$

where k_{cat} is the rate constant of the conversion of the Michaelis complex (ES) into products (P), $[E]_0$ is the total concentration of enzyme in solution, [S] is the concentration of unbound substrate, v_{max} is the maximum value for the reaction rate and K_M is the Michaelis constant. The latter is defined as the substrate concentration at which the reaction rate is half of v_{max} for an enzymatic reaction whose kinetic obeys to the Michaelis-Menten equation (Figure 2.6). In the pre-equilibrium approximation, when ES → P is much slower than E+S ⇄ ES, K_M coincides with the dissociation constant of ES. For this reason K_M is usually associated with the enzyme-substrate affinity. However, if ES → P is much faster than ES formation and [ES] ~ 0, the steady-state approximation is valid and K_M explicitly depends on k_{cat}:

$$K_M = \frac{k_{off} + k_{cat}}{k_{on}} \qquad \text{Eq. 2.55}$$

where k_{on} and k_{off} are the rates of formation and dissociation of ES. It is worth noting that: (i) for a given reaction none of the above approximations could be valid; (ii) product dissociation from the enzyme might be slow (which means breaking ES → P into ES → EP ⇄ E+P). It follows that estimating K_M through molecular simulations is far from trivial, and the only safe definition of K_M

is the one of Figure 2.6, although it seems reasonable to associate K_M to all the factors of Eq. 2.55 (and, if that is the case, to product dissociation).

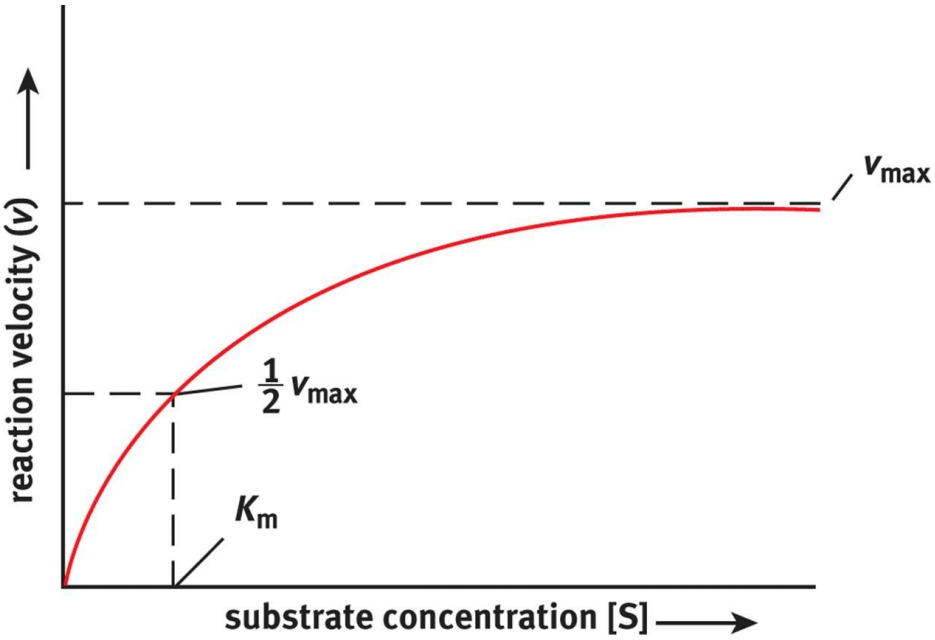


Figure 2.6. Plot of the Michaelis-Menten equation. Source: http://goo.gl/i5RiJX.

On the other hand, k_{cat} can be calculated, through the modelization of the TS and applying Eyring's equation to the rate-determining step (Eq. 2.9) or, if ET occurs, by computing the key parameters of Marcus' equation (Eq. 2.49). However, if the catalyzed reaction is complex and some non-chemical event are required (intermediate unbinding, protein conformational change, etc.) the abovementioned strategy could fail and k_{cat} evaluation become cumbersome.

2.8.3 Cooperative binding in multimeric proteins

Let's consider a multimeric protein, i.e. a complex of amino acid chains (subunits) organized in a well defined quaternary structure, where each subunit is capable of binding a certain ligand. If all the binding events are independent, protein-ligand binding is non-cooperative and the fraction of occupied subunits (y) varies with the concentration of the ligand (x) following a Langmuir curve (Figure 2.7). Conversely, if protein-ligand affinity is higher than the product of the individual subunit-ligand affinities, the association is cooperative and a sigmoid fits the experimental results quite well (Figure 2.7) (Ferrell, 2009). A careful look at Figure 2.7 reveals the advantage of cooperative binding: the prompt response of affinity to

changes in ligand concentration. This characteristic allows a protein to switch from a low- to a high-affinity receptor (or vice versa) more easily than in a non-cooperative system (Ferrell, 2009).

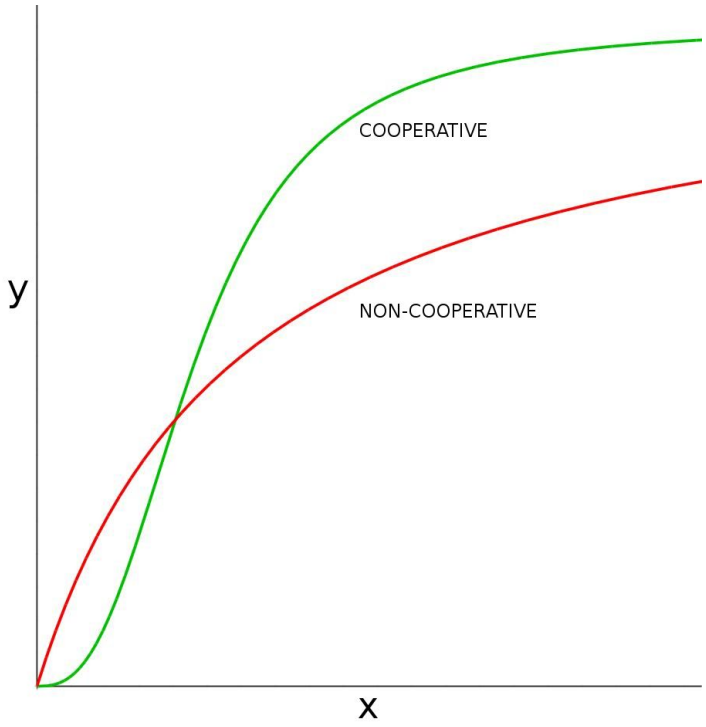


Figure 2.7. Saturation curve of cooperative and non-cooperative multimeric protein. Plot produced with KmPlot (https://goo.gl/9biJ2N) and modified with GIMP (https://www.gimp.org/).

Two of the most popular models of cooperative binding were provided in the 60s by the Monod, Wyman and Changeux (MWC) (Monod et al., 1965) on one hand, and by Koshland, Nemethy and Filmer (KNF) (Koshland, Némethy & Filmer, 1966) on the other. In the KNF model, a multimeric protein binds ligand molecules sequentially with increasing affinity; after one subunit binds a ligand molecule, its structural and/or dynamical changes are communicated to the other subunits, whose ligand affinity increases. In the MWC model the protein is assumed to exist in tense (T) and relaxed (R) states with low- and high-ligand affinity respectively. Such affinities are intrinsic characteristics of the two states and are independent on the number of ligand molecules bound. Therefore the fraction of bound ligand molecules is solely regulated by the T/R equilibrium, which is in turn perturbed by the presence of the ligand: if its concentration is high R is favored otherwise the system will be predominantly in the T state.

Cooperativity is the consequence of inter-subunit allosteric communication: binding at one site affects the affinity of the others. According to the MWC and KFN models of

cooperativity, such allosteric mechanism is purely structural. However, it should be beared in mind that any equilibrium is dictated by a free energy difference. Therefore, cooperative binding might well arise from purely dynamical effects such as the altered vibrational entropy of the receptor upon ligand binding (Cooper & Dryden, 1984; Popovych et al., 2006; Hilser, Wrabl & Motlagh, 2012; Nussinov, Ruth & Chung-Jung, 2015) which could then propagate to the other protein chains.

REFERENCES:

Allen MP., Tildesley DJ. 1989. *Computer Simulation of Liquids*. Oxford University Press.

Atilgan AR., Durell SR., Jernigan RL., Demirel MC., Keskin O., Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical journal* 80:505–515.

Atkins PW. 1998. *Physical Chemistry*. W H Freeman & Company.

Atkins PW., Friedman RS. 2011. *Molecular Quantum Mechanics*. Oxford University Press.

Bader RFW. 1994. *Atoms in Molecules: A Quantum Theory*. Oxford University Press on Demand.

Baker CM. 2015. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley interdisciplinary reviews. Computational molecular science* 5:241–254.

Bashford D., Case DA. 2000. Generalized born models of macromolecular solvation effects. *Annual review of physical chemistry* 51:129–152.

Borrelli KW., Vitalis A., Alcantara R., Guallar V. 2005. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of chemical theory and computation* 1:1304–1311.

Borrelli KW., Cossins B., Guallar V. 2010. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *Journal of computational chemistry* 31:1224–1235.

Bortolotti CA., Siwko ME., Elena C., Antonio R., Marco S., Stefano C. 2011. The Reorganization Energy in Cytochrome c is Controlled by the Accessibility of the Heme to the Solvent. *The journal of physical chemistry letters* 2:1761–1765.

Breneman CM., Wiberg KB. 1990. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of computational chemistry* 11:361–373.

Cooper A., Dryden DT. 1984. Allostery without conformational change. A plausible model. *European biophysics journal: EBJ* 11:103–109.

Cossins BP., Hosseini A., Guallar V. 2012. Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *Journal of chemical theory and computation* 8:959–965.

Cramer CJ. 2013. *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons.

Drenth J. 2013. *Principles of Protein X-ray Crystallography*. Springer Science & Business Media.

Ferrell JE Jr. 2009. Q&A: Cooperativity. *Journal of biology* 8:53.

Frenkel D., Smit B. 2001. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press.

Friesner RA., Banks JL., Murphy RB., Halgren TA., Klicic JJ., Mainz DT., Repasky MP., Knoll EH., Shelley M., Perry JK., Shaw DE., Francis P., Shenkin PS. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* 47:1739–1749.

Halgren TA., Murphy RB., Friesner RA., Beard HS., Frye LL., Pollard WT., Banks JL. 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry* 47:1750–1759.

Hill TL. 2012. *An Introduction to Statistical Thermodynamics*. Courier Corporation.

Hilser VJ., Wrabl JO., Motlagh HN. 2012. Structural and energetic basis of allostery. *Annual review of biophysics* 41:585–609.

Hosseini A., Brouk M., Lucas MF., Glaser F., Fishman A., Guallar V. 2015. Atomic picture of ligand migration

in toluene 4-monooxygenase. *The journal of physical chemistry. B* 119:671–678.
Jacobson MP., Kaminski GA., Friesner RA., Rapp CS. 2002. Force Field Validation Using Protein Side Chain Prediction. *The journal of physical chemistry. B* 106:11673–11680.
Jensen JH. 2015. Predicting accurate absolute binding energies in aqueous solution: thermodynamic considerations for electronic structure methods. *Physical chemistry chemical physics: PCCP* 17:12441–12451.
Kamerlin SCL., Warshel A. 2011. The empirical valence bond model: theory and applications. *Wiley interdisciplinary reviews. Computational molecular science* 1:30–45.
Kaminski GA., Friesner RA., Julian T-R., Jorgensen WL. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *The journal of physical chemistry. B* 105:6474–6487.
Koch W., Holthausen MC. 2015. *A Chemist's Guide to Density Functional Theory*. John Wiley & Sons.
Koshland DE Jr., Némethy G., Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5:365–385.
Krauth W. 2006. *Statistical Mechanics: Algorithms and Computations*. Oxford University Press, UK.
Laio A., Alessandro L., Gervasio FL. 2008. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* 71:126601.
Laio A., Parrinello M. 2002. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99:12562–12566.
Leach AR. 2001. *Molecular Modelling: Principles and Applications*. Pearson Education.
Lucas MF., Guallar V. 2012. An atomistic view on human hemoglobin carbon monoxide migration processes. *Biophysical journal* 102:887–896.
Madadkar-Sobhani A., Guallar V. 2013. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic acids research* 41:W322–8.
Marcus RA. 1993. Electron transfer reactions in chemistry. Theory and experiment. *Reviews of modern physics* 65:599–610.
Messiah A. 2014. *Quantum Mechanics*. Courier Corporation.
Monod J., Jacque M., Jeffries W., Jean-Pierre C. 1965. On the nature of allosteric transitions: A plausible model. *Journal of molecular biology* 12:88–118.
Mulliken RS. 1955. Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I. *The Journal of chemical physics* 23:1833.
Nussinov R., Ruth N., Chung-Jung T. 2015. Allostery without a conformational change? Revisiting the paradigm. *Current opinion in structural biology* 30:17–24.
Orozco M. 2014. A theoretical view of protein dynamics. *Chemical Society reviews* 43:5051.
Paltrinieri L., Borsari M., Ranieri A., Battistuzzi G., Corni S., Bortolotti CA. 2013. The Active Site Loop Modulates the Reorganization Energy of Blue Copper Proteins by Controlling the Dynamic Interplay with Solvent. *Journal of Physical Chemistry Letters* 4:710–715.
Popovych N., Nataliya P., Shangjin S., Ebright RH., Kalodimos CG. 2006. Dynamically driven protein allostery. *Nature structural & molecular biology* 13:831–838.
Senftle TP., Hong S., Islam MM., Kylasa SB., Zheng Y., Shin YK., Junkermeier C., Engel-Herbert R., Janik MJ., Aktulga HM., Verstraelen T., Grama A., van Duin ACT. 2016. The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials* 2:15011.
Senn HM., Walter T. 2009. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie, International Edition* 48:1198–1229.
de Souza ON., Ornstein RL. 1997. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophysical journal* 72:2395–2397.
Szabo A., Ostlund NS. 2012. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation.
Tadesse MA., D'Annibale A., Galli C., Gentili P., Sergi F. 2008. An assessment of the relative contributions of redox and steric issues to laccase specificity towards putative substrates. *Organic & biomolecular chemistry* 6:868–878.
Takahashi R., Gil VA., Guallar V. 2014. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *Journal of chemical theory and computation*

10:282–288.
Trott O., Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 31:455–461.
Tuckerman M. 2010. *Statistical Mechanics: Theory and Molecular Simulation*. OUP Oxford.
Warshel A., Levitt M. 1976. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* 103:227–249.
Yilmazer ND., Korth M. 2015. Enhanced semiempirical QM methods for biomolecular interactions. *Computational and structural biotechnology journal* 13:169–175.

3. Computational protein design: state of the art

In this chapter the state of the art of computational protein design is reviewed, highlighting, when possible, the strength, limitations and perspectives of each methodology. The discussion is divided in three sections, each focusing on one aspect of design: stability, ligand binding and catalysis. Protein-protein and antibody design are not covered since they are out of the scope of this thesis.

3.1 Protein stability design

A protein with its characteristic and functional three-dimensional structure is said to be folded in its native structure. Its stability is affected by several factors such as temperature, pH, and concentration of co-solute in solution. Estimating the effect of mutations on the folding free energy (the difference in free energy between native and non-native states) is of paramount importance for industrial applications, where the conditions can significantly differ from the natural ones. Two of the most popular tools to prepare and score mutated proteins are Rosetta (Korkegian, 2005; Kellogg, Leaver-Fay & Baker, 2011) and FoldX (Guerois, Nielsen & Serrano, 2002). After introducing a mutation in the wild type, the protein's (usually side chain) torsional degrees of freedom are optimized to minimize a scoring function which estimates the folding free energy for the created variant. Such function depends on: (i) physics-based terms, which account for van der Waals, H-bond, solvation and electrostatic (approximated with PDB statistics in Rosetta (Kuhlman & Baker, 2000)), energies; (ii) knowledge-based contributions, which determine the probability of a given rotamer according to structural database statistics (Dunbrack, 2002). Apart from these overlaps, Rosetta approximates the free energy change in the unfolded state due to a mutation with context-independent reference energies for each residue (Kuhlman & Baker, 2000), On the other hand, FoldX explicitly estimates the entropy cost to fix a rotamer in the native state (Guerois, Nielsen & Serrano, 2002). The relative low computational cost of these protocols permits to generate and score a large number of mutations in a short time. As shown by Potapov et al, the accuracy/cost trade-off is such that these tools can correctly reproduce overall trends, and therefore suggest stabilizing mutations with acceptable probabilities, but they are not good enough to provide detailed results (Potapov, Cohen & Schreiber, 2009).

Following Potapov et al.'s accuracy assessment, Kellogg et al tested the ability of Rosetta to score mutations combining several energy functions and sampling methods with variable resolution (Kellogg, Leaver-Fay & Baker, 2011). As a main result, the authors concluded that the choice of the sampling algorithm should be tuned with the resolution of the energy function adopted. In other words, an accurate energy function better performs with a finer sampling; vice versa, roughly sampled structures should be scored by smoother functions which can tolerate steric clashes better. Still, flexible backbone protocols improved small to large residue mutations, where significant structural changes can occur. In addition, the authors found that conformational sampling was still insufficient to recover the crystal conformation when a large to small hydrophobic residue mutation was introduced, due to poor packing. Larger errors were found when the polarity of the residue drastically changed upon mutation, which suggests poor trade-off between polar desolvation and buried polar interactions. The lack of explicit water molecules and ligand contacts was another factor in some failed prediction. Finally, the lack of a context-dependent unfolded state modeling (a given mutation A1→A2 was considered to have the same effect on the unfolded state independently on the environment (Kuhlman & Baker, 2000)) was held up as a source of error, although not a major one. In fact, a free energy variation of the unfolded state upon mutation might change protein stability as well as a variation in the folded state. However, a recent paper shows that an accurate conformational and energetic characterization of the unfolded protein is not trivial and that its inclusion in protein stability scoring significantly worsened the prediction (Estrada, Echenique & Sancho, 2015).

The entropic scoring in FoldX (Guerois, Nielsen & Serrano, 2002) only takes into account the change in conformational entropy, which depends on the number of accessible conformers in the unfolded state and their probabilities (Karplus, Ichiye & Pettitt, 1987). Although this entropic variation dominates folding (Chong & Ham, 2015), large discrepancies in vibrational entropy (the intrinsic entropy of a given protein conformer (Karplus, Ichiye & Pettitt, 1987)) have been calculated between thermophilic and mesophilic proteins (Frappier & Najmanovich, 2015). Therefore, the thoughtful inclusion of a vibrational entropy contribution in protein design free energy functions might pay off. Najmanovich and coworkers implemented this strategy in the ENcoM server (Frappier, Chartier & Najmanovich, 2015), where they combine FoldX (Guerois, Nielsen & Serrano, 2002) with their ENcoM protocol to rapidly estimate vibrational entropy. ENcoM uses ANM (Atilgan et al., 2001) with a pairwise atom-type

non-bonded interaction term to include the specific nature of amino-acids (Frappier & Najmanovich, 2014).

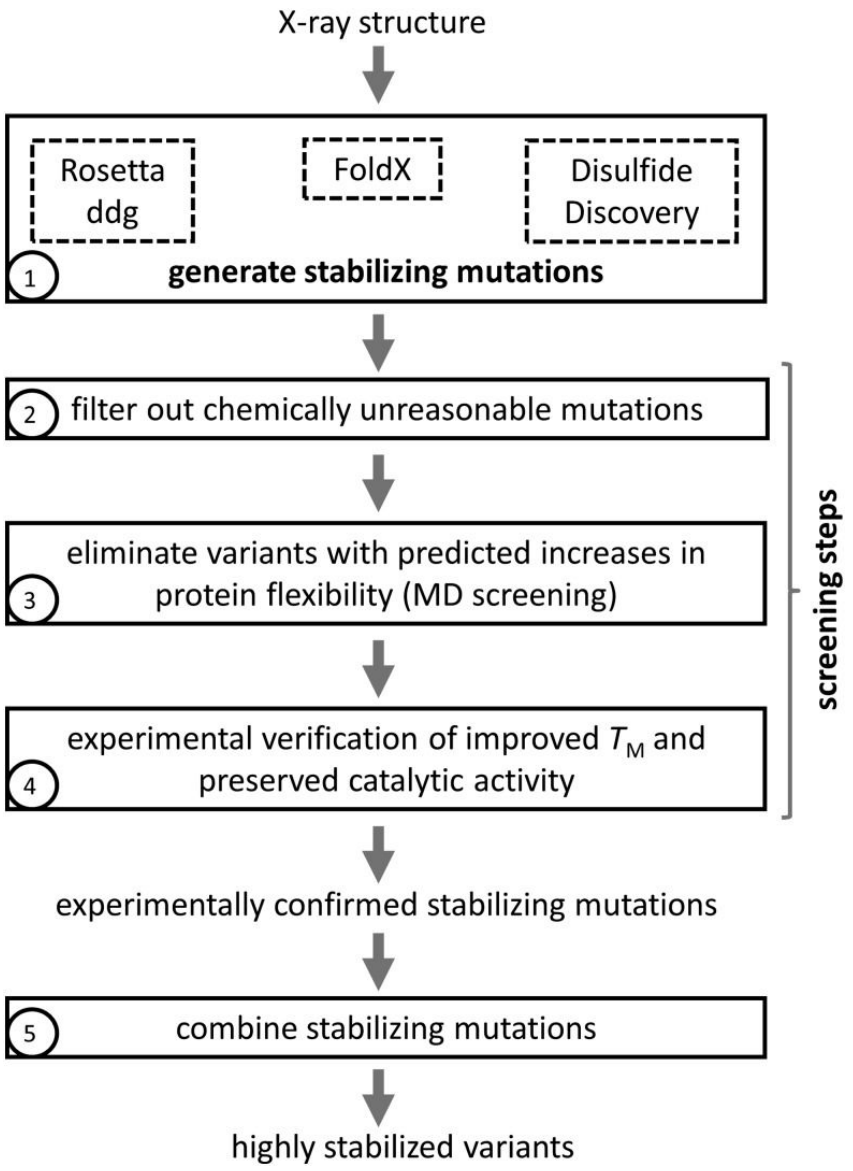


Figure 3.1. Schematic representation of the stability design workflow of Wijma et al. 2014 (Wijma et al., 2014). Source: (Wijma et al., 2014).

In an attempt to quantify free energies more rigorously, De Groot and coworkers employed alchemical free energy simulations to score 109 mutants of Ribonuclease Barnase (Seeliger, Daniel & de Groot, 2010). In this technique, sampling a convenient number of unphysical ("alchemical") intermediates renders a rigorous evaluation of the free energy energy difference between two states (e.g. wild type and mutant protein) (Leach, 2001). Sampling was

carried out with molecular dynamics (MD). Unfolded state free energy differences were calculated with Gly-Xxx-Gly peptides with capped termini. This choice provides an universal, albeit less accurate as context-independent, reference state (whose values need to be calculated only once and then are stored as a database). The overall correlation with experimental values was 0.86 and ~72 % of the predicted values were within 1 kcal/mol of the experimental value with 30 ns of simulation time. Notably, most of this accuracy (65 %) is retained with only 5 ns. Larger errors were detected for mutations that introduced large variations in structure and fluctuation (mutations to glycine, from a bulky residue to another, of well packed residues) or in the electrostatics of buried residues.

Possible strategies to focus the experimental search of stable protein variants in smaller regions to optimize protein design are: (i) the identification of flexible backbone sites (associated with instability) with sampling techniques and their rigidification (Joo et al., 2011; Huang, Gao & Zhan, 2011) designing salt bridges (Lee et al., 2014) and/or disulphide bonds (Pikkemaat et al., 2002); (ii) the optimization of surface charge-charge interactions (Spector et al., 2000; Schweiker et al., 2007; Gribenko et al., 2009); (iii) the optimization of core packing (Borgo & Havranek, 2012); (iv) the removal of unsatisfied buried polar groups (Hendsch et al., 1996); (v) the localization of critical residues in the active site entry tunnels, especially for co-solute tolerance, with MD (Koudelakova et al., 2013) or other algorithms like PELE (Borrelli et al., 2005; Madadkar-Sobhani & Guallar, 2013).

Recently Wijma and coworkers developed, and applied with success, a mixed approach which would allow to obtain highly thermostable protein variants in a short time with minimum experimental screening (Wijma et al., 2014). In their computational workflow (Figure 3.1), potentially stabilizing mutations were firstly produced and scored with Rosetta (Kellogg, Leaver-Fay & Baker, 2011) and FoldX (Guerois, Nielsen & Serrano, 2002). To minimize the risk to affect catalysis, only residues beyond 10 Å of the active site were mutated. Mutations were considered potentially stabilizing if $\Delta\Delta G_{folding} \leq -5$ kJ/mol or if $|\Delta\Delta G_{folding}| < 5$ kJ/mol and the mutation type was contained in the set {Xxx→Arg, Xxx→Pro, Gly→Xxx}. These were then filtered to avoid undesired, typically destabilizing features such as increased unsatisfied H-bond donors and acceptors or hydrophobic surface exposure to water. Then, multiple short MD simulations were used to discard variants with increased backbone flexibility. Finally, variants with experimentally confirmed higher thermostability and preserved activity were combined in the lab. This computational workflow, thanks to its hierarchical screening which helps to unmask false positives (~50% of the potentially stabilizing mutations), effectively focuses the

experimental efforts on reliable mutations and it is therefore a plausible model for future computer-aided design of thermostable proteins. The main drawback is the exclusion of mutations which are neutral or mildly damaging per se but that coupled to other mutations could synergically improve thermostability.

3.2 Protein-ligand binding

Designing proteins which bind small molecules selectively and tightly is fundamental for the development of new therapeutics (Morin, Meiler & Mizoue, 2011) and catalysts (Wijma, Floor & Janssen, 2013). The accurate *in silico* design of protein-ligand interactions is very hard (Schreier et al., 2009) for a number of reasons: (i) the low tolerance to error due to the small size of protein-ligand interfaces; (ii) the huge number of possible combinations in sequence space (for 10 positions there are ~10^{13} sequences); (iii) the need to accurately reconstruct protein-ligand polar interactions, which are very sensitive to geometry (Allison et al., 2014).

In a recent attempt to benchmark the state of art of computational protein-ligand interactions design, Allison and coworkers tested Rosetta's (Schueler-Furman et al., 2005) sequence recover (with respect to the wild type) capability over a set of 43 protein-ligand complexes (Allison et al., 2014). The Rosetta protocol involved simultaneous ligand motion and side chain rotamer discrete optimization. Overall, sequence recovery was more successful when: (i) a near-optimal pose was inputted and subjected to limited sampling instead of being blindly searched; (ii) the ligand was small, non polar and rigid; (iii) the binding pocket packing was not too crowded or poor. Another interesting result was the significantly higher recovery for non polar residues. The authors suggested that new terms should be added to the energy function to correct this bias toward non polar interactions (Allison et al., 2014). However, this bias could be an artifact of poor sampling, which might limit the accuracy of polar interactions estimation (see above). In fact, other suggested areas of improvement were the use of continuous (instead of discrete) sampling of backbone and side chain rotamers (Gainza et al., 2013), local backrub motions (small concerted rotation of the backbone of two adjacent residues which lead to a much larger side chain motion) (Davis et al., 2006; Keedy et al., 2012), and the approximate calculation of partition functions over structural ensembles (Chen et al., 2009; Allen, Nisthal & Mayo, 2010; Frey et al., 2010).

All these features are grasped by OSPREY (Gainza et al., 2013), a recent open source solution to protein design which includes GPU acceleration (Zhou et al., 2014), dead-end

elimination algorithm (Hallen, Keedy & Donald, 2013) and the K* method (Lilien et al., 2005). K* aims to approximate the partition function of the protein-ligand complex, the free protein and ligand over an ensemble of structures; their ratio provides an estimation of the binding constant. The conceptual advantage of this methodology is a mathematically rigorous free energy difference calculation which explicitly simulates the free ligand and protein. Consequently, ligand and binding site preorganization are, in principle, included in the calculation. On the other hand, absolute free energy calculations are inefficient and far from being accurate. However, the error of the method might compensate between complex and free monomers calculations, making this strategy an interesting tool for fast free binding energy simulations. Regardless the methodology chosen, an effort to produce new experimental data will be fundamental to benchmark these high-throughput computational protocols and improve their predictive power (Malisi et al., 2012).

A poor knowledge and/or accurate description of the binding site is another possible source of error. Indeed, some mutations could shift the pKa of ligand's and protein's titratable sites or introduce a new titratable residue. Therefore, the system should be prepared thoroughly before computational mutagenesis and quick pKa predictors (Shields & Seybold, 2013) should be used to treat critical mutations. On the other hand, the pH ~ pKa condition requires to simulate all the possible combinations for all the ambiguous titratable sites. Finally, waters in the binding site might have an important role in binding and their neglect could affect the quality of the calculation.

Another possible source of false positives can be the exclusive focus on the binding site, as some mutations could hinder the entrance of the binding site. Sampling algorithms like PELE can help to recognize such mutations mapping the whole ligand migration process with good accuracy, taking protein flexibility into account. High-throughput MD (HTMD) (Doerr & De Fabritiis, 2014) is a more accurate, as it explicitly models water molecules, but also more expensive alternative.

A way to filter and correct designs is based on MD simulations (Kiss et al., 2013a). Many features of designed protein-ligand complexes can be inspected with this technique, including H-bond geometries, binding site structural integrity, solvent exposure and binding site preorganization. In particular HTMD (Doerr & De Fabritiis, 2014) was used by Baker and coworkers to filter computationally designed candidates according to the fraction of near attack conformations (NAC, conformations which resemble the transition state, TS) displayed during the multiple simulations (Wijma et al., 2015). Moreover, MD can help in the future to discern

long range effects. In fact, it has been recently used to highlight the impact of distant mutations on active site preorganization in evolved enzymes (Jiménez-Osés et al., 2014; Osuna et al., 2015). Furthermore, proteins are dynamical entities organized in a network of correlated fluctuations whose changes can significantly affect binding at large distances (DuBay, Bowman & Geissler, 2015). Importantly, such network can be identified through a correlation matrix (which quantifies the correlation degree of a pair of amino acids) and partitioned in communities of highly correlated residues, giving insights on allosteric interactions (Sethi et al., 2009). These analysis, with contact and H-bond maps, might be used in the future to identify regions whose motion influences the binding site dynamics (for example making the side chain of a catalytic residue too flexible), which can then be subjected to mutagenesis in the lab.

Finally, quantum mechanics (QM) calculations can be used to validate promising mutations, especially when charge transfer and polarization might have an important role in binding. Quantum mechanics/molecular mechanics (QM/MM) calculations can significantly improve protein-ligand binding prediction directly, through explicit energy calculations (Chaskar et al., 2014), or indirectly (Cho et al., 2005), re-calculating ligand's atomic charges in an attempt to model ligand polarization effects. An alternative, more accurate but heavier approach to large systems is the fragment molecular orbital (FMO) method (Fedorov, Nagata & Kitaura, 2012). FMO divides the system in N non overlapping fragments (e.g. one for each protein residue and ligand) and calculates the total energy as the sum of one-body fragment energies and two-body interaction energy corrections, providing a ~N^2 scalable fully parallelizable QM calculation. Jensen and coworkers used this methodology to energetically score the cleavability of peptides by HIV1-protease (Jensen et al., 2014).

3.3 Catalytic activity design

As previously discussed, estimating the Michelis constant (K_M) through molecular simulations is challenging as it potentially depends on the kinetics of substrate binding, unbinding, product formation (Eq. 2.55) and dissociation. Only if product formation is much slower than substrate binding K_M coincides with the dissociation constant and its design becomes a protein-ligand binding design. On the other hand, the catalytic rate constant (k_{cat}) can be conceptually associated to the rate constant of the slowest step of an enzymatic reaction (although in a multistep reaction with slow non-chemical event things could get complicated but this is the worst case scenario). Therefore, the computer-aided improvement of catalytic activity

passes through the modelization of the transition state (TS) of such step and Eyring's equation (Eq. 2.9) or by computing the key parameters of Marcus' equation (Eq. 2.49) when the rate limiting step involves ET. The problems with the design of optimal activation free energies are multiple: (i) the energy function should be sensitive enough to effectively discriminate between the substrate and the TS, whose charges and geometries might be similar; (ii) the nature of the TS can change upon mutation; (iii) activation free energies are very sensitive to molecular geometry changes.

In OptZyme the TS is approximated by a transition state analogue (TSA), a stable molecule which electronically and sterically resembles the TS (Grisewood et al., 2013). Once the TSA and the substrate are docked in the active site, two parameters drive mutant selection: the enzyme-substrate (ES) interaction energy and the enzyme-TSA interaction energy. Through a number of conceptual and mathematical simplifications, the authors show that the former energy correlates with K_M, the latter with k_{cat}/K_M. Although it yielded satisfactory correlations for their specific case (Grisewood et al., 2013), it is worth noting that these correlations have no general validity (as discussed above). Notwithstanding, ES and enzyme-TSA interaction energies are still valuable tools for a fast semi-quantitative evaluation of enzyme variants.

Khersonsky et al. combined computational design and directed evolution to optimize a previously designed Kemp eliminase (Khersonsky et al., 2011). The interaction energy between the enzyme and an explicit TS model (substrate plus catalytic residues), calculated in solution with QM methods, was the parameter to be optimized during the sequence exploration. The authors individuated three main factors of improved activity, namely a more favorable electrostatic environment, a better packed active site and a higher degree of active site preorganization.

Although these two methods allow to test a big number of mutants, they both present a conceptual limitation: the use of enzyme-TS model interaction energies, which are size-dependent, to score activation energies (TS-substrate free energy differences), which are intensive quantities. Therefore, the ES interaction energies should be in principle taken into account. However, poor sampling and inaccurate energy functions might introduce uncertainties which could make their introduction useless (as it often happens in molecular mechanics/generalized born surface area free energy calculations (Genheden, Samuel & Ulf, 2015)). Moreover, optimizing only the TS energy along mutagenesis is expected to improve catalysis assuming that ES will not be stabilized as much as the TS will be. Thus, this is the best

approach to test a very large number of mutations and find promising protein variants which can then be filtered with MD and QM/MM methods (Kiss et al., 2013b).

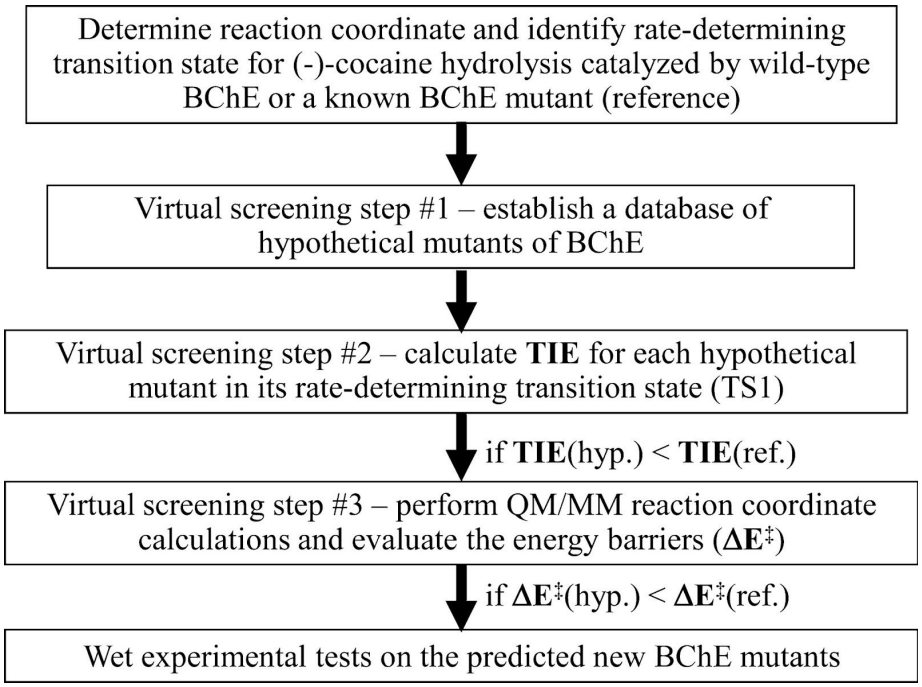


Figure 3.2. Schematic representation of the catalytic activity design workflow of Zheng et al. 2008 (Zheng et al., 2008a). Source: Zheng et al. 2008 (Zheng et al., 2008a).

The only way to properly estimate (free) energy barriers is the introduction of a QM description in the methodology, which is capable to describe finer effects accompanying TS formation (such as charge redistribution and small geometry variations). QM/MM schemes, which have been widely applied to elucidate enzymatic reaction mechanisms (van der Kamp & Mulholland, 2013), have been employed to rescore promising candidates (van der Kamp & Mulholland, 2013; Kiss et al., 2013b). A remarkable example is the hierarchical approach of Zheng et al. used to design a cocaine hydrolase (Zheng et al., 2008b). Firstly, the reaction coordinate and the TS of the rate-determining limiting step is determined in the wild type. Then, many mutations are scored according to their protein-TS interaction energy; if this is lower than the wild type, a QM/MM calculation along the reaction coordinate is used to estimate the energy barrier. This type of calculation is carried out freezing the reactive coordinate at each step of the reaction process and minimizing all the other degrees of freedom. The main drawback of this strategy is the need of extensive sampling to converge these values, which make the presented methodology expensive.

A cheaper alternative to QM/MM calculations is empirical valence bond (EVB) (Kamerlin & Arieh, 2011). EVB is based on a semi-empirical Hamiltonian which describes reactants and products with their resonance structure (explicitly defining atom connectivity). Although EVB energies are less accurate than *ab initio* and DFT QM/MM methods, EVB-based free energy calculations are orders of magnitude quicker and still can achieve accurate results (Frushicheva et al., 2010; Frushicheva, Cao & Warshel, 2011), making EVB a suitable tool to score a bigger number of mutants.

In the attempt to describe the entire enzyme or a large part of it with QM calculations, Jensen and coworkers approximated the reaction coordinate with the linear interpolation between reactant and product optimized geometries and calculated each point with semi-empirical calculations (Hediger et al., 2012, 2013). These fast QM calculations, united with algorithms for large scale systems such as FMO (Fedorov, Nagata & Kitaura, 2012; Ito, Mika & Tore, 2014) or the much faster Effective Fragment Molecular Orbital (EFMO) (Steinmann, Fedorov & Jensen, 2012), make “all ab initio chemistry” (Steinmann et al., 2013) closer, albeit still far away for design purposes, in the horizon.

REFERENCES:

Allen BD., Nisthal A., Mayo SL. 2010. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proceedings of the National Academy of Sciences* 107:19838–19843.

Allison B., Combs S., DeLuca S., Lemmon G., Mizoue L., Meiler J. 2014. Computational design of protein-small molecule interfaces. *Journal of structural biology* 185:193–202.

Atilgan AR., Durell SR., Jernigan RL., Demirel MC., Keskin O., Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical journal* 80:505–515.

Borgo B., Havranek JJ. 2012. Automated selection of stabilizing mutations in designed and natural proteins. *Proceedings of the National Academy of Sciences of the United States of America* 109:1494–1499.

Borrelli KW., Vitalis A., Alcantara R., Guallar V. 2005. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of chemical theory and computation* 1:1304–1311.

Chaskar P., Prasad C., Vincent Z., Röhrig UF. 2014. Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *Journal of chemical information and modeling* 54:3137–3152.

Chen C-Y., Georgiev I., Anderson AC., Donald BR. 2009. Computational structure-based redesign of enzyme activity. *Proceedings of the National Academy of Sciences of the United States of America* 106:3764–3769.

Cho AE., Victor G., Berne BJ., Richard F. 2005. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *Journal of computational chemistry* 26:915–931.

Chong S-H., Ham S. 2015. Dissecting Protein Configurational Entropy into Conformational and Vibrational Contributions. *The journal of physical chemistry. B* 119:12623–12631.

Davis IW., Bryan Arendall W., Richardson DC., Richardson JS. 2006. The Backrub Motion: How Protein Backbone Shrugs When a Sidechain Dances. *Structure* 14:265–274.

Doerr S., De Fabritiis G. 2014. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput

Molecular Simulations. *Journal of chemical theory and computation* 10:2064–2069.
DuBay KH., Bowman GR., Geissler PL. 2015. Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Accounts of chemical research* 48:1098–1105.
Dunbrack RL Jr. 2002. Rotamer libraries in the 21st century. *Current opinion in structural biology* 12:431–440.
Estrada J., Echenique P., Sancho J. 2015. Predicting stabilizing mutations in proteins using Poisson-Boltzmann based models: study of unfolded state ensemble models and development of a successful binary classifier based on residue interaction energies. *Physical chemistry chemical physics: PCCP* 17:31044–31054.
Fedorov DG., Nagata T., Kitaura K. 2012. Exploring chemistry with the fragment molecular orbital method. *Physical chemistry chemical physics: PCCP* 14:7562–7577.
Frappier V., Chartier M., Najmanovich RJ. 2015. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucleic acids research* 43:W395–400.
Frappier V., Najmanovich RJ. 2014. A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. *PLoS computational biology* 10:e1003569.
Frappier V., Najmanovich R. 2015. Vibrational entropy differences between mesophile and thermophile proteins and their use in protein engineering. *Protein science: a publication of the Protein Society* 24:474–483.
Frey KM., Georgiev I., Donald BR., Anderson AC. 2010. Predicting resistance mutations using protein design algorithms. *Proceedings of the National Academy of Sciences of the United States of America* 107:13707–13712.
Frushicheva MP., Cao J., Chu ZT., Warshel A. 2010. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proceedings of the National Academy of Sciences of the United States of America* 107:16869–16874.
Frushicheva MP., Cao J., Warshel A. 2011. Challenges and advances in validating enzyme design proposals: the case of kemp eliminase catalysis. *Biochemistry* 50:3849–3858.
Gainza P., Roberts KE., Georgiev I., Lilien RH., Keedy DA., Chen C-Y., Reza F., Anderson AC., Richardson DC., Richardson JS., Donald BR. 2013. OSPREY: protein design with ensembles, flexibility, and provable algorithms. *Methods in enzymology* 523:87–107.
Genheden S., Samuel G., Ulf R. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery* 10:449–461.
Gribenko AV., Patel MM., Liu J., McCallum SA., Wang C., Makhatadze GI. 2009. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proceedings of the National Academy of Sciences* 106:2601–2606.
Grisewood MJ., Gifford NP., Pantazes RJ., Li Y., Cirino PC., Janik MJ., Maranas CD. 2013. OptZyme: computational enzyme redesign using transition state analogues. *PloS one* 8:e75358.
Guerois R., Nielsen JE., Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology* 320:369–387.
Hallen MA., Keedy DA., Donald BR. 2013. Dead-end elimination with perturbations (DEEPer): a provable protein design algorithm with continuous sidechain and backbone flexibility. *Proteins* 81:18–39.
Hediger MR., De Vico L., Svendsen A., Besenmatter W., Jensen JH. 2012. A computational methodology to screen activities of enzyme variants. *PloS one* 7:e49849.
Hediger MR., De Vico L., Rannes JB., Christian J., Werner B., Allan S., Jensen JH. 2013a. In silico screening of 393 mutants facilitates enzyme engineering of amidase activity in CalB. *PeerJ* 1:e145.
Hediger MR., Casper S., De Vico L., Jensen JH. 2013. A computational method for the systematic screening of reaction barriers in enzymes: searching for Bacillus circulans xylanase mutants with greater activity towards a synthetic substrate. *PeerJ* 1:e111.
Hendsch ZS., Thorlakur J., Sauer RT., Bruce T. 1996. Protein Stabilization by Removal of Unsatisfied Polar Groups: Computational Approaches and Experimental Tests †. *Biochemistry* 35:7621–7625.
Huang X., Gao D., Zhan C-G. 2011. Computational design of a thermostable mutant of cocaine esterase via molecular dynamics simulations. *Organic & biomolecular chemistry* 9:4138–4143.
Ito M., Mika I., Tore B. 2014. Novel Approach for Identifying Key Residues in Enzymatic Reactions: Proton Abstraction in Ketosteroid Isomerase. *The journal of physical chemistry. B* 118:13050–13058.

Jensen JH., Willemoës M., Winther JR., De Vico L. 2014. In silico prediction of mutant HIV-1 proteases cleaving a target sequence. *PloS one* 9:e95833.
Jiménez-Osés G., Osuna S., Gao X., Sawaya MR., Gilson L., Collier SJ., Huisman GW., Yeates TO., Tang Y., Houk KN. 2014. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nature chemical biology* 10:431–436.
Joo JC., Pack SP., Kim YH., Yoo YJ. 2011. Thermostabilization of Bacillus circulans xylanase: computational optimization of unstable residues based on thermal fluctuation analysis. *Journal of biotechnology* 151:56–65.
Kamerlin SCL., Arieh W. 2011. The empirical valence bond model: theory and applications. *Wiley interdisciplinary reviews. Computational molecular science* 1:30–45.
van der Kamp MW., Mulholland AJ. 2013. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* 52:2708–2728.
Karplus M., Ichiye T., Pettitt BM. 1987. Configurational entropy of native proteins. *Biophysical journal* 52:1083–1085.
Keedy DA., Georgiev I., Triplett EB., Donald BR., Richardson DC., Richardson JS. 2012. The role of local backrub motions in evolved and designed mutations. *PLoS computational biology* 8:e1002629.
Kellogg EH., Leaver-Fay A., Baker D. 2011a. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 79:830–838.
Khersonsky O., Rothlisberge D., Wollacott AM., Dym O., Baker D., Tawfik DS., Center ISP. 2011. Optimization of the in silico designed Kemp eliminase KE70 by computational design and directed evolution. DOI: 10.2210/pdb3npu/pdb.
Kiss G., Gert K., Pande VS., Houk KN. 2013a. Molecular Dynamics Simulations for the Ranking, Evaluation, and Refinement of Computationally Designed Proteins. In: *Methods in Enzymology*. 145–170.
Kiss G., Çelebi-Ölçüm N., Moretti R., Baker D., Houk KN. 2013b. Computational enzyme design. *Angewandte Chemie* 52:5700–5725.
Korkegian A. 2005. Computational Thermostabilization of an Enzyme. *Science* 308:857–860.
Koudelakova T., Chaloupkova R., Brezovsky J., Prokop Z., Sebestova E., Hesseler M., Khabiri M., Plevaka M., Kulik D., Kuta Smatanova I., Rezacova P., Ettrich R., Bornscheuer UT., Damborsky J. 2013. Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie* 52:1959–1963.
Kuhlman B., Baker D. 2000. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences of the United States of America* 97:10383–10388.
Leach AR. 2001. *Molecular Modelling: Principles and Applications*. Pearson Education.
Lee C-W., Wang H-J., Hwang J-K., Tseng C-P. 2014. Protein thermal stability enhancement by designing salt bridges: a combined computational and experimental study. *PloS one* 9:e112751.
Lilien RH., Stevens BW., Anderson AC., Donald BR. 2005. A Novel Ensemble-Based Scoring and Search Algorithm for Protein Redesign and Its Application to Modify the Substrate Specificity of the Gramicidin Synthetase A Phenylalanine Adenylation Enzyme. *Journal of computational biology: a journal of computational molecular cell biology* 12:740–761.
Madadkar-Sobhani A., Guallar V. 2013. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic acids research* 41:W322–8.
Malisi C., Schumann M., Toussaint NC., Kageyama J., Kohlbacher O., Höcker B. 2012. Binding pocket optimization by computational protein design. *PloS one* 7:e52505.
Morin A., Meiler J., Mizoue LS. 2011. Computational design of protein-ligand interfaces: potential in therapeutic development. *Trends in biotechnology* 29:159–166.
Osuna S., Jiménez-Osés G., Noey EL., Houk KN. 2015. Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Accounts of chemical research* 48:1080–1089.
Pikkemaat MG., Linssen ABM., Berendsen HJC., Janssen DB. 2002. Molecular dynamics simulations as a tool for improving protein stability. *Protein engineering* 15:185–192.
Potapov V., Cohen M., Schreiber G. 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein engineering, design & selection: PEDS* 22:553–560.
Schreier B., Stumpp C., Wiesner S., Hocker B. 2009. Computational design of ligand binding is not a solved problem. *Proceedings of the National Academy of Sciences* 106:18491–18496.

Schueler-Furman O., Wang C., Bradley P., Misura K., Baker D. 2005. Progress in modeling of protein structures and interactions. *Science* 310:638–642.
Schweiker KL., Arash Z-A., Davidson AR., Makhatadze GI. 2007. Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge-charge interactions. *Protein science: a publication of the Protein Society* 16:2694–2702.
Seeliger D., Daniel S., de Groot BL. 2010. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical journal* 98:2309–2316.
Sethi A., Eargle J., Black AA., Luthey-Schulten Z. 2009. Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* 106:6620–6625.
Shields GC., Seybold PG. 2013. *Computational Approaches for the Prediction of pKa Values*. CRC Press.
Spector S., Wang M., Carp SA., Robblee J., Hendsch ZS., Fairman R., Tidor B., Raleigh DP. 2000. Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry* 39:872–879.
Steinmann C., Casper S., Fedorov DG., Jensen JH. 2013. Mapping Enzymatic Catalysis Using the Effective Fragment Molecular Orbital Method: Towards all ab initio Biochemistry. *PloS one* 8:e60602.
Steinmann C., Fedorov DG., Jensen JH. 2012. The effective fragment molecular orbital method for fragments connected by covalent bonds. *PloS one* 7:e41117.
Wijma HJ., Floor RJ., Jekel PA., Baker D., Marrink SJ., Janssen DB. 2014. Computationally designed libraries for rapid enzyme stabilization. *Protein engineering, design & selection: PEDS* 27:49–58.
Wijma HJ., Floor RJ., Bjelic S., Marrink SJ., Baker D., Janssen DB. 2015. Enantioselective enzymes by computational design and in silico screening. *Angewandte Chemie* 54:3726–3730.
Wijma HJ., Floor RJ., Janssen DB. 2013. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Current opinion in structural biology* 23:588–594.
Zheng F., Fang Z., Wenchao Y., Mei-Chuan K., Junjun L., Hoon C., Daquan G., Min T., Hsin-Hsiung T., Woods JH., Chang-Guo Z. 2008a. Most Efficient Cocaine Hydrolase Designed by Virtual Screening of Transition States. *Journal of the American Chemical Society* 130:12148–12155.
Zheng F., Fang Z., Wenchao Y., Mei-Chuan K., Junjun L., Hoon C., Daquan G., Min T., Hsin-Hsiung T., Woods JH., Chang-Guo Z. 2008b. Most Efficient Cocaine Hydrolase Designed by Virtual Screening of Transition States. *Journal of the American Chemical Society* 130:12148–12155.
Zhou Y., Xu W., Donald BR., Zeng J. 2014. An efficient parallel algorithm for accelerating computational protein design. *Bioinformatics* 30:i255–i263.

4. Metalloprotein studied and objectives

Two proteins have been studied in this thesis work: laccases and hemoglobin. Both systems are briefly introduced in this chapter and the objectives of this thesis work are highlighted.

4.1 Laccases

Laccases (Morozova et al., 2007; Pardo & Camarero, 2015) are multicopper oxidases, enzymes that catalyze the oxidation of a molecule at the expenses of a sacrificial electron acceptor (in this case molecular oxygen, which is reduced to water) (Morozova et al., 2007). Laccases are effective over a wide range of compounds, most likely due to their shallow substrate binding site. Such broad specificity, their clean by-products (water) and cheap electron acceptor (oxygen) make them appealing for green chemistry (Riva, 2006). Indeed, laccases already have a place in several industrial applications such as textile and food industry, bioremediation, forestry and organic synthesis.

The reactive core of a laccase is split in two divalent copper (Cu) clusters: the so-called t1-type (T1) Cu and a trinuclear cluster of Cu (TNC), which comprises two t3-type (T3) and one t2-type (T2) Cu. While the TNC is solely coordinated by His residues (see Figure 1.1 for the abbreviation of amino acid names), the T1 copper has one Cys and two His residues as equatorial ligands, while the axial positions are occupied by an Ile and by a variable ligand, usually one between Met (in low reduction potential laccases), Ile and Phe. One of the equatorial His interacts with the substrate to convey one electron to the T1 copper (Figure 4.1). If ET is coupled to proton transfer, like in the case of phenols and amines at acidic pH, an Asp residue plays the role of the catalytic base (Figure 4.1). Once the T1 copper is reduced, the electron tunnels through the equatorial Cys and two T3-coordinating His residues (highlighted as balls and spheres in Figure 4.1) toward the TNC. Four substrate molecules are necessary to fully reduce the enzyme ($4Cu^{++} + 4e^{-} \rightarrow 4Cu^{+}$), activating oxygen reduction at the TNC site.

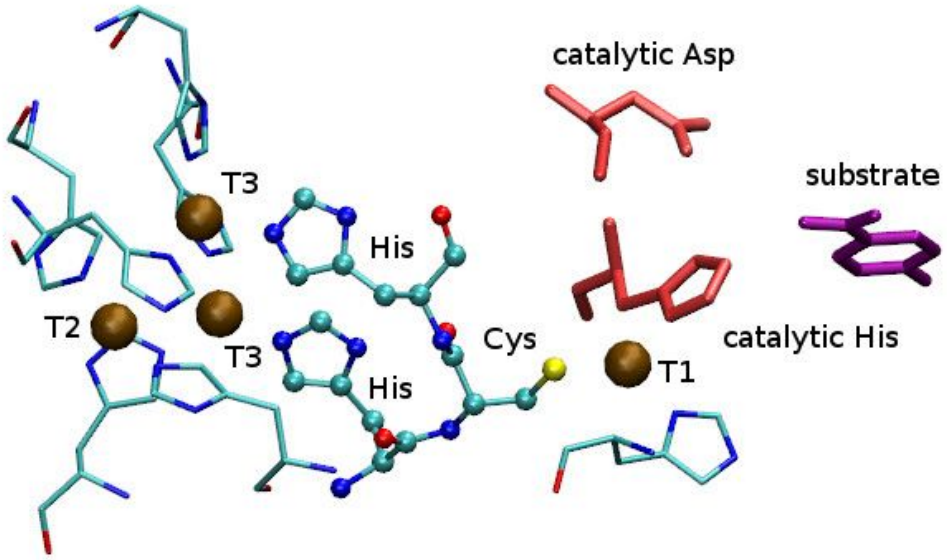


Figure 4.1. Reactive core of a *Trametes trogii* laccase (PDB: 2hrg (Matera et al., 2007)). Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

A correlation between the rate constant (k_{cat}) and the difference in redox potential (E^o) between the T1 copper and some classes of substrate (Xu, 1996; Xu et al., 2000; Tadesse et al., 2008) suggests that: (i) the T1 copper reduction is probably the rate-determining step of the catalytic process, and (ii) the rate of this step is determined by the ET driving force. The latter point is likely a consequence of the high reorganization energy that accompanies substrate oxidation (Tadesse et al., 2008; Galli et al., 2013). For these reasons the E^o of the T1 copper (E^o_{T1}), which ranges from 0.4 to 0.8 V vs. normal hydrogen electrode (NHE), is considered the key parameter for substrate oxidation. Even if some determinants of E^o_{T1} are known (Li et al., 2004; Hong et al., 2011), it is not clear how to effectively enhance it in order to improve laccases activity and/or oxidize new compounds. Moreover it is not clear if tuning E^o_{T1} is a viable way to design laccases for two reasons: (i) increasing E^o_{T1} would worsen the internal ET (from the T1 copper to the TNC), which is already uphill in free energy, rendering E^o_{T1} tunable in a limited window, beyond which the internal ET would become the rate-determining step; (ii) in many cases an increased E^o_{T1} did not yield a better catalyst, quite the opposite (Xu et al., 1998, 1999; Durão et al., 2006).

That E^o_{T1} alone cannot define k_{cat} is not surprising since laccase-substrate binding can affect many key factors such as: (i) electronic coupling; (ii) the electrostatic environment of the

substrate, i.e. its oxidability; (iii) substrate solvent exposure, i.e. the reorganization energy. On top of that, a substrate may be more stable in a binding mode other than the reactive one while directly inhibiting the formation of the latter. Therefore, a proper modeling of substrate binding into the T1 site and its oxidation might enlighten new design strategies.

The oxidation of compounds toward which laccases are not active can be achieved using mediators (Riva, 2006). These are molecules that can reduce the T1 copper and, in their oxidized form, remove an electron from a target substrate. Interestingly, Tron and coworkers have shown that a hybrid laccase/tris(bipyridine)ruthenium(II) ($[Ru(bpy)_3]^{2+}$) system can oxidize compounds such as ethylenediaminetetraacetic acid (EDTA) and p-styrene sulfonate (Simaan et al., 2011; Schneider et al., 2015). Although $[Ru(bpy)_3]^{2+}$ photoexcitation leads to a formidably fast ET to the T1 copper, a very quick back-transfer (BT) occurs due to the formation of $[Ru(bpy)_3]^{3+}$, which is a strong oxidant (E^o=1.29 V). Consequently, the oxidation of the target substrate and internal ET are inhibited, yielding a slow catalytic turnover. Limiting back-transfer, either sacrificing ET or modifying $[Ru(bpy)_3]^{2+}$ to make it act as a molecular diode, would make this photoactive hybrid species a powerful tool for sustainable catalysis and energy conversion.

4.2 Hemoglobin

Hemoglobin is a homodimer of heterodimers (Figure 4.2A) that binds molecular oxygen (O_2), but also carbon monoxide (CO) and other small ligands, cooperatively. The heterodimers are made up of one α and one β subunit (Figure 4.2B and 4.2C respectively), both containing a heme group (a divalent iron cation, Fe^{++}, coordinated by a porphyrin ring, Figure 4.2D). Each heme, which is bound to the protein through the proximal His (Figure 4.2D), is the responsible of uptaking one ligand molecule. Thanks to cooperative binding, hemoglobin can efficiently bind O_2 in the lungs, where this is abundant, carry it through the circulatory system and release it in the tissue, where O_2 partial pressure (pO_2) is scarce (Ferrell, 2009). This unique ability makes hemoglobin a promising candidate for blood substituent design (Fronticelli et al., 2007), although its low thermostability and auto-oxidation limit the applications (Bobofchak et al., 2003). Therefore, hemoglobin needs to be designed to overcome these limitations while keeping an eye on cooperativity, which should not be compromised. Hence, hemoglobin design is a delicate multidimensional problem (not to speak about its interaction with the plethora of physiological effectors). Unfortunately, the mechanism behind hemoglobin cooperativity is far from being fully understood despite more than half a century of research.

The MWC model is considered one of the most reliable theories for hemoglobin cooperativity. This success was fueled by the discovery of oxy- and deoxy-hemoglobin crystal structures, which due to their different affinities were identified as the R and T states respectively (Perutz, 1970, 1979). In the deoxy-hemoglobin crystal, Fe^{++} lays out of the heme plane and the heme-His bond is tilted and elongated, while in the oxy-hemoglobin structure Fe^{++} stays in the porphyrin plane and the heme-his bond is shorter and not tilted. Both out-of-plane displacement and bond elongation are normal consequences of the high-spin (quintet) state of Fe^{++} when it is five-coordinated; however, tilting indicates that the heme-his moiety is strained. Further inspecting the crystal structures, Perutz concluded that this so called proximal strain in the deoxy-hemoglobin crystal is induced by several inter-subunit salt bridges that hold the proximal His back preventing Fe^{++} to go back to the porphyrin plane upon binding (Perutz, 1970, 1979). These salt-bridges were also predicted to break after ligand binding and trigger the T to R transition. Notably, when the heme is detached from the protein and coordinated by imidazole (analog to His side chain) (Barrick et al., 1997): (i) ligand affinity increases; (ii) the quaternary transition is inhibited and (iii) cooperativity is attenuated, but not completely lost. While the first two points validate the proximal His role in Perutz's stereochemical model (Perutz, 1970), the last one underlines that a fraction of cooperative binding is controlled by other factors, which could comprise the configuration of the distal pocket (Olson et al., 1988) and entropic effects (Yonetani & Kanaori, 2013).

According to the tertiary two-state (TTS) model, which is a generalization of the MWC theory, ligand affinity is directly controlled by the tertiary structure of each subunit, which can exists in two states r and t with high and low affinity respectively, whose population is determined by the nature of the quaternary state (R or T) (Henry et al., 2002). In particular, almost all subunits are in the r state in R while t is predominant in T. Therefore, according to the TTS model, the quaternary state controls ligand affinity only indirectly, influencing the tertiary state of each subunit.

The scenario turned out to be more complicated when alternative quaternary structures were found for oxy-hemoglobin (Silva, Rogers & Arnone, 1992; Safo & Abraham, 2005). Moreover, the results of ANM sampling (Xu, Tobi & Bahar, 2003) and several MD simulations, run with different FFs (Hub, Kubitzki & de Groot, 2010; Yusuff et al., 2012; Vesper & de Groot, 2013), suggest that: (i) deoxy-hemoglobin crystal structure is very unstable in water and quickly (within 100-200 ns) converts into oxy-hemoglobin-like structures (ii) oxy-hemoglobin is well represented by a very wide ensemble of oxy-hemoglobin-like conformations instead of a few

crystal structures. Nuclear magnetic resonance (NMR) (Skoog, James Holler & Crouch, 2007) experiments confirmed the second prediction (Fan et al., 2013) and showed how neither deoxy- nor oxy-hemoglobin crystal structures fit well the NMR data (Sahu et al., 2007). As a consequence, hemoglobin should be considered as a highly flexible protein described by a wide ensemble of structures, being aware that crystal structures are (only) models, with their strengths and limitations, and should not be conclusive when interpreting experimental findings.

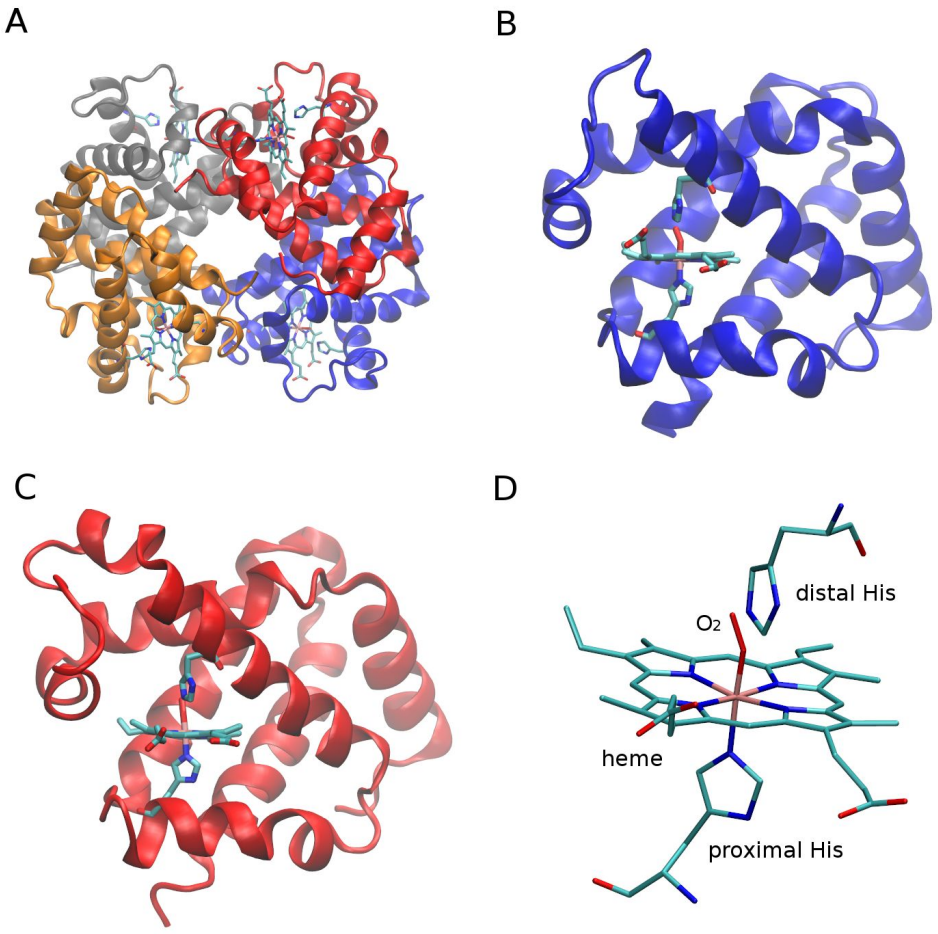


Figure 4.2. A) Hemoglobin structure, colours: α1 blue, β1 red, α2 yellow and β2 gray ; B) α1 subunit; C) β1 subunit; D) heme group coordinated by the proximal His and oxygen. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

4.3 Objectives of the thesis

4.3.1 Laccases

- Develop an efficient computational protocol to model how laccases and their substrates interact (conformational sampling) and score ET rates (quantum mechanics). This methodology will be applied to reproduce the trends in available experimental data, aiming at a deeper understanding of the determinants of laccases reactivity.
- Apply the developed computational strategy to design laccases variants with improved oxidation of a target substrate. Thanks to the collaboration with experimental groups that will validate the designs, the strengths and limitations of the proposed computational methodology will be discussed.
- Combine classical conformational sampling and quick empirical T_{DA} estimations to rationalize the functioning of a hybrid laccase/$[Ru(bpy)_3]^{2+}$system and, knowing that the fast electron BT is the major problem for this photocatalyst, design variants with improved ET directionality from the substrate to the TNC.

4.3.2 Hemoglobin

- Identify the structural features of the r and t states of the TTS model within the same quaternary structure, establishing whether their difference involves the whole subunits or is located in close proximity to the heme, pinpointing how heme reactivity is regulated, and find out how much the r and t states resemble the tertiary structures found in the oxy- and deoxy-hemoglobin crystal structures.
- Find the preferential routes for heme-heme allosteric signalling using MD simulations, statistical and network analysis techniques. In order to do this, the communication paths between key residues (as established in the previous point) which are placed around the heme and belong to distinct subunits will be examined. With this information in hand, mutagenesis experiments could be arranged in the future to either test the importance of each communication pathway or increase hemoglobin's stability in water (in this case, excluding the residues responsible of allosteric communications from mutagenesis, narrowing down the search).

REFERENCES:

Barrick D., Ho NT., Simplaceanu V., Dahlquist FW., Ho C. 1997. A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin. *Nature structural biology* 4:78–83.
Bobofchak KM., Toshiaki M., Texel SJ., Andrea B., Masaaki N., Traystman RJ., Koehler RC., Brinigar WS., Clara F. 2003. A recombinant polymeric hemoglobin with conformational, functional, and physiological characteristics of an in vivo O2transporter. *American Journal of Physiology - Heart and Circulatory Physiology* 285:H549–H561.
Durão P., Bento I., Fernandes AT., Melo EP., Lindley PF., Martins LO. 2006. Perturbations of the T1 copper site in the CotA laccase from Bacillus subtilis: structural, biochemical, enzymatic and stability studies. *Journal of biological inorganic chemistry: JBIC: a publication of the Society of Biological Inorganic Chemistry* 11:514–526.
Fan J-S., Jing-Song F., Yu Z., Wing-Yiu C., Virgil S., Ho NT., Chien H., Daiwen Y. 2013. Solution Structure and Dynamics of Human Hemoglobin in the Carbonmonoxy Form. *Biochemistry* 52:5809–5820.
Ferrell JE Jr. 2009. Q&A: Cooperativity. *Journal of biology* 8:53.
Fronticelli C., Clara F., Koehler RC., Brinigar WS. 2007. Recombinant Hemoglobins as Artificial Oxygen Carriers. *Artificial cells, blood substitutes, and immobilization biotechnology* 35:45–52.
Galli C., Madzak C., Vadalà R., Jolivalt C., Gentili P. 2013. Concerted electron/proton transfer mechanism in the oxidation of phenols by laccase. *Chembiochem: a European journal of chemical biology* 14:2500–2505.
Henry ER., Stefano B., James H., Eaton WA. 2002. A tertiary two-state allosteric model for hemoglobin. *Biophysical chemistry* 98:149–164.
Hong G., Ivnitski DM., Johnson GR., Atanassov P., Pachter R. 2011. Design parameters for tuning the type 1 Cu multicopper oxidase redox potential: insight from a combination of first principles and empirical molecular dynamics simulations. *Journal of the American Chemical Society* 133:4802–4809.
Hub JS., Kubitzki MB., de Groot BL. 2010. Spontaneous quaternary and tertiary T-R transitions of human hemoglobin in molecular dynamics simulation. *PLoS computational biology* 6:e1000774.
Humphrey W., William H., Andrew D., Klaus S. 1996. VMD: Visual molecular dynamics. *Journal of molecular graphics* 14:33–38.
Li H., Webb SP., Ivanic J., Jensen JH. 2004. Determinants of the relative reduction potentials of type-1 copper sites in proteins. *Journal of the American Chemical Society* 126:8010–8019.
Matera I., Gullotto A., Ferraroni M., Tilli S., Briganti F., Scozzafava A. 2007. Crystal Structure of Blue Laccase from Trametes trogii complexed with p-methylbenzoate. DOI: 10.2210/pdb2hrg/pdb.
Morozova OV., Shumakovich GP., Gorbacheva MA., Shleev SV., Yaropolov AI. 2007. “Blue” laccases. *Biochemistry. Biokhimiia* 72:1136–1150.
Olson JS., Mathews AJ., Rohlfs RJ., Springer BA., Egeberg KD., Sligar SG., Tame J., Renaud JP., Nagai K. 1988. The role of the distal histidine in myoglobin and haemoglobin. *Nature* 336:265–266.
Pardo I., Camarero S. 2015. Laccase engineering by rational and evolutionary design. *Cellular and molecular life sciences: CMLS* 72:897–910.
Perutz MF. 1970. Stereochemistry of cooperative effects in haemoglobin. *Nature* 228:726–739.
Perutz MF. 1979. Regulation of Oxygen Affinity of Hemoglobin: Influence of Structure of the Globin on the Heme Iron. *Annual review of biochemistry* 48:327–386.
Riva S. 2006. Laccases: blue enzymes for green chemistry. *Trends in biotechnology* 24:219–226.
Safo MK., Abraham DJ. 2005. The Enigma of the Liganded Hemoglobin End State: A Novel Quaternary Structure of Human Carbonmonoxy Hemoglobin † , ‡. *Biochemistry* 44:8347–8359.
Sahu SC., Simplaceanu V., Gong Q., Ho NT., Tian F., Prestegard JH., Ho C. 2007. Insights into the solution structure of human deoxyhemoglobin in the absence and presence of an allosteric effector. *Biochemistry* 46:9973–9980.
Schneider L., Mekmouche Y., Rousselot-Pailley P., Simaan AJ., Robert V., Réglier M., Aukauloo A., Tron T. 2015. Visible-Light-Driven Oxidation of Organic Substrates with Dioxygen Mediated by a [Ru(bpy)3](2+) /Laccase System. *ChemSusChem* 8:3048–3051.
Silva MM., Rogers PH., Arnone A. 1992. A third quaternary structure of human hemoglobin A at 1.7-A resolution. *The Journal of biological chemistry* 267:17248–17256.
Simaan AJ., Mekmouche Y., Herrero C., Moreno P., Aukauloo A., Delaire JA., Réglier M., Tron T. 2011. Photoinduced multielectron transfer to a multicopper oxidase resulting in dioxygen reduction into water. *Chemistry* 17:11743–11746.

Skoog DA., James Holler F., Crouch SR. 2007. *Principles of Instrumental Analysis*. Brooks/Cole Publishing Company.

Tadesse MA., D'Annibale A., Galli C., Gentili P., Sergi F. 2008. An assessment of the relative contributions of redox and steric issues to laccase specificity towards putative substrates. *Organic & biomolecular chemistry* 6:868–878.

Vesper MD., de Groot BL. 2013. Collective dynamics underlying allosteric transitions in hemoglobin. *PLoS computational biology* 9:e1003232.

Xu F. 1996. Oxidation of phenols, anilines, and benzenethiols by fungal laccases: correlation between activity and redox potentials as well as halide inhibition. *Biochemistry* 35:7608–7614.

Xu F., Feng XU., Berka RM., Wahleithner JA., Nelson BA., Shuster JR., Brown SH., Palmer AE., Solomon EI. 1998. Site-directed mutations in fungal laccase: effect on redox potential, activity and pH profile. *Biochemical Journal* 334:63–70.

Xu F., Palmer AE., Yaver DS., Berka RM., Gambetta GA., Brown SH., Solomon EI. 1999. Targeted mutations in a Trametes villosa laccase. Axial perturbations of the T1 copper. *The Journal of biological chemistry* 274:12372–12375.

Xu F., Kulys JJ., Duke K., Li K., Krikstopaitis K., Deussen HJ., Abbate E., Galinyte V., Schneider P. 2000. Redox chemistry in laccase-catalyzed oxidation of N-hydroxy compounds. *Applied and environmental microbiology* 66:2052–2056.

Xu C., Tobi D., Bahar I. 2003. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin TR2 transition. *Journal of molecular biology* 333:153–168.

Yonetani T., Kanaori K. 2013. How does hemoglobin generate such diverse functionality of physiological relevance? *Biochimica et biophysica acta* 1834:1873–1884.

Yusuff OK., Babalola JO., Bussi G., Raugei S. 2012. Role of the subunit interactions in the conformational transitions in adult human hemoglobin: an explicit solvent molecular dynamics study. *The journal of physical chemistry. B* 116:11004–11009.

Results and discussion

5. A binding-focused strategy for laccases design based on the spin density

In this chapter, a new computational strategy to boost laccases' design is presented. In section 5.1, the motivations behind the choice of the spin density (SD) as activity descriptor is elucidated and a first proof of concept is furnished. Then, the computational methodology is explained in detail (section 5.2) and applied to several systems to highlight their functioning (sections 5.3-5.5). Finally, its successful application in a rational laccase design is reported.

5.1 The spin density as a descriptor of laccases kinetics

The SD, which has been used to study oxidation mechanisms (Johansson et al., 2002; Catlow et al., 2005) and estimate electron transfer (ET) pathways (Guallar & Wallrapp, 2008), shows whether an unpaired electron is energetically more stable in donor's or acceptor's molecular orbitals. Therefore, SD variations are expected to reflect changes in the ET driving force. This is confirmed in previous experimental studies, which reveal a linear relationship between the amount of SD on the donor and its reduction/oxidation potential in ET proteins (Artz et al., 1997) and small aromatic compounds (Meisel, Dan & Neta, 1975). As mentioned in the previous chapter, T1 copper reduction is (most of the times) the rate-determining step of the catalytic process, and its rate is regulated by the ET driving force due to the high reorganization energy (Xu, 1996; Xu et al., 2000; Tadesse et al., 2008). It follows that a prompt evaluation of the ET driving force through the SD could open the door for computational laccases design procedures. On the side, changes in electronic coupling and solvent reorganization energy can be qualitatively predicted in terms of donor-acceptor distance (Blumberger, 2015) and substrate's solvent accessible surface area (SASA) (Bortolotti et al., 2011). It should be noted that a rigorous calculation of Marcus' equation parameters is possible (Wallrapp, Voityuk & Victor, 2013; Blumberger, 2015) but it is too slow for design purposes, which is the ultimate goal of this thesis.

4-X-PhOH

Scheme 5.1. Chemical structure of 4-X-PhOH. Prepared with MarvinSketch (http://www.chemaxon.com).

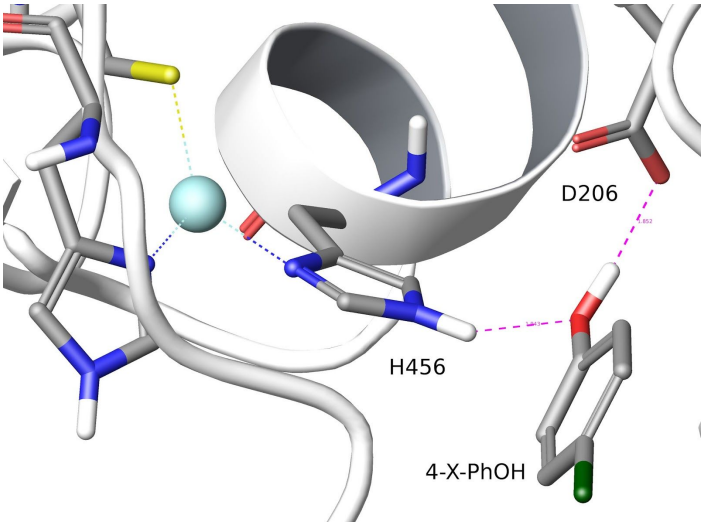


Figure 5.1. Representative binding mode of a 4-X-PhOH, X = OH, OCH_3, CH_3, Cl. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

A stringent proof of concept for the SD would be to correctly score the kinetics of a set of laccase-substrate pairs whose specificity constants (k_{cat}/K_M) correlates with the redox potential (E°) difference (ΔE°) between the T1 copper and the substrate. In such a case, it would be clear that catalysis is completely controlled by the ET driving force accompanying substrate oxidation; a possible failure of the SD to correlate with k_{cat}/K_M would then highlight that the former is not a good descriptor of the driving force. Incidentally, phenols with general formula 4-X-PhOH (where X is a para substituent) tend to display such k_{cat}/K_M vs. ΔE° correlation (Tadesse et al., 2008). Four of them, with X = {OH, OMe, Me, Cl}, were docked with Glide (Friesner et al., 2004) (Figure 5.1) into the T1 cavity of the *Pycnoporus cinnabarinus* laccase (PcL) and the SD calculated with a quantum mechanics/molecular mechanics (QM/MM) scheme using several

density functionals (see Computational details). The experimental specificity constant values for *Trametes villosa* laccase (Tadesse et al., 2008), which shares 70% identity with PcL, are used to validate the methodology. As shown in Figure 5.2, a good correlation is observed for all the density functionals used. Therefore, the SD is a promising tool to score the ET driving force in laccases.

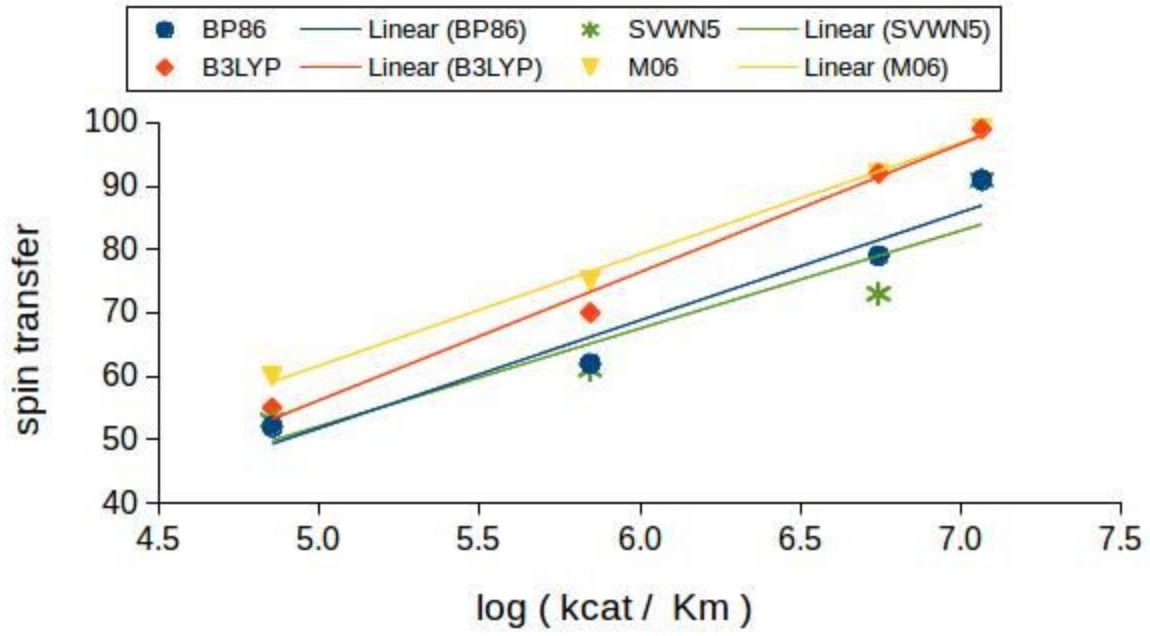


Figure 5.2. Linear relationship between the spin population and experimental kinetic data. Figure prepared with LibreOffice Calc (https://goo.gl/kdcCc8).

5.2 Computational details

At the heart of the methodology is the combination of exhaustive enzyme-substrate conformational sampling provided by PELE with a QM/MM reactivity scoring based on the SD (PELE+QM/MM protocol). The enzyme-substrate conformational space nearby the T1 pocket is sampled with PELE. Then, 20-50 structures within 5 kcal/mol of the lowest binding energy pose are randomly selected and their reactivity scored evaluating the amount of SD localized on the substrate with QM/MM calculation.

System setup. The *Pycnoporus cinnabarinus* laccase (PcL) structure (2XYB) was taken from the Protein Data Bank, while the *Myceliophthora thermophila* laccase (MtL) and PM1 laccase structures were made available by Novozymes and Dr. Javier Medrano respectively. The PM1 laccase was used as a template to build a model of the 7D5 laccase (96% sequence

identity) with Prime (Jacobson et al., 2002, 2004). These structures were prepared with Protein Preparation Wizard (Sastry et al., 2013), simulating pH 5 for PcL and MtL, while PM1 was modeled at pH 3, in order to compare with the experimental data. The protonation states of titratable residues were generated with PROPKA (Olsson et al., 2011) and double-checked with the H++ server (Gordon et al., 2005). All pertinent mutations were introduced and the new protein structure was relaxed with PELE.

PELE sampling. The substrate, initially placed in front of the T1 cavity, was randomly translated and rotated within 20 Å of the T1 copper. The ANM perturbation was applied to all the C_α in the system, the side chains of the residues within 10 Å of substrate were repacked and everything within 20 Å was energy minimized. Such a reduced minimization is intended to speed up conformational sampling (see Appendix 1 for an example of input file). Along the thesis, (PELE) binding energies refers to protein-substrate interaction potential energies.

Treatment of substrate and metal centers in PELE. Substrates' geometries were optimized at the M06/6-31G* level of theory, modeling the solvent with the Poisson-Boltzmann model. The atomic ESP charges were calculated from the optimized structure and used for PELE. The OPLS parameters were assigned to the ligand with Schrödinger's hetgrp_ffgen utility and a rotamer library was generated with Macromodel (Mohamadi et al., 1990). The copper ions and their coordination atoms plus their nearest neighbors were constrained during PELE sampling with 200 kcal mol^{-1} $Å^{-2}$ harmonic constraints on their atomic positions. Unitary charges were assigned to each copper atom (to avoid tension in the TNC).

QM/MM reactivity scoring. Structures were relaxed with five steps of geometry optimization using Qsite (Murphy, Philipp & Friesner, 2000) and the SD of the substrate was evaluated using Mulliken's partitioning method. Substrate, T1 copper and its first coordination sphere were included in the quantum region, while the rest of the protein is treated classically. The density functional method with the M06 (Zhao, Yan & Truhlar, 2008) functional and the lacvp* basis set (Hay, Jeffrey Hay & Wadt, 1985) were used for the quantum region, while the OPLS-AA (Kaminski et al., 2001) force field was used for the classical part. Other density functionals (B3LYP (Becke, 1988, 1993a,b; Lee et al., 1988), BP86 (Perdew, 1986; Becke, 1988), PBE (Perdew, Kieron & Matthias, 1996), SVWN5 (Vosko, Wilk & Nusair, 1980)) were sporadically employed and their use is highlighted in the text. A non bonded cutoff of 50 Å was applied. The partial geometry optimization was found to be sufficient to let the SD virtually converge (Figure 5.3) and it was therefore adopted to save CPU time. Solvent effect were neglected in order to speed up the calculations (saving both CPU time of a single geometry

optimization and sensibly reducing the number of calculations, since the effect of water MM charges would need to be averaged on a much larger number of structures). Although this choice leads to errors on an absolute ground, the error is most likely comparable for all the poses involving the same ligand, making relative estimation still reliable (while absolute values are not expected to be meaningful). Indeed, the final aim of this methodology is to screen the activity of a large number of mutants toward the same substrate. Solvent effects can potentially be extrapolated and discussed inspecting changes in substrate's SASA (Bortolotti et al., 2011).

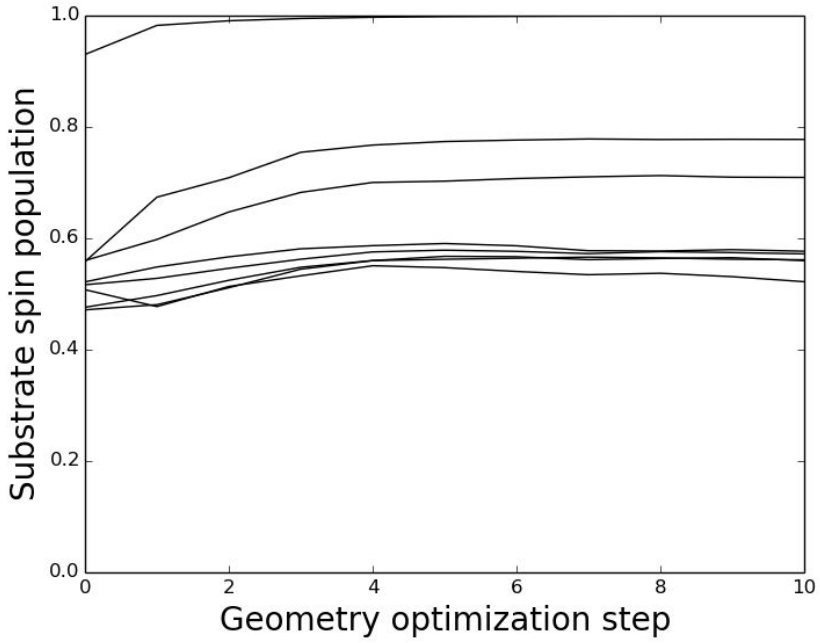


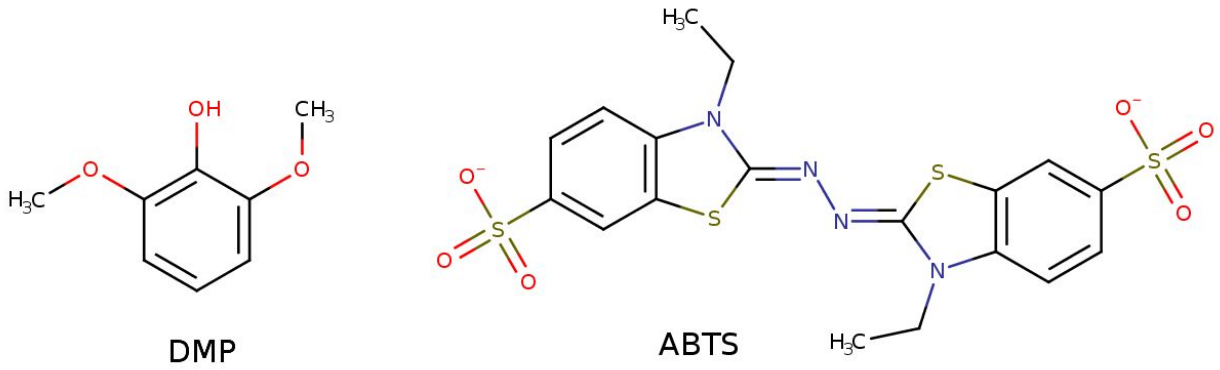
Figure 5.3. Convergence of 4-X-PhOH spin population with QM/MM geometry optimization steps. Figure prepared with matplotlib (http://matplotlib.org/).

5.3 Evolution of Pycnoporus cinnabarinus laccase

In this section, the PELE+QM/MM protocol is applied to the wild type and evolved variants of the *Pycnoporus cinnabarinus* laccase (PcL) and the substrates employed to screen activity, namely the 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) and 2,6-dimethoxyphenol (DMP) (Camarero et al., 2012). The objective is to understand how mutations accumulated through *in vitro* evolution lead to improved catalytic oxidation (Camarero et al., 2012), providing physico-chemical grounds for future rational design efforts.

5.3.1 Results and discussion

As a result of the laboratory evolution, the turnover number significantly improved in the final mutant, 3PO, for both substrates (Table 5.1). The evolved laccase carries five mutations: P394H and N208S, which are located in the T1 pocket, N331D and D341N, relatively close to the substrate entrance, and R280H, located far away on the protein surface, affecting mainly protein expression in *Saccharomyces cervisae* (Camarero et al., 2012) (Figure 5.4).



Scheme 5.2. Chemical structure of DMP and ABTS. Prepared with MarvinSketch (http://www.chemaxon.com).

5.3.1.1. DMP oxidation

The lowest energy bound structures reveal that PcL (wild type) has only one binding mode (**1**), also present in 3PO, depicted in red in Figures 5.5A and 5.5B. In this minimum DMP is hydrogen bonded to H456 and D206 (Figure 5.7A). QM/MM scoring, gives a result of 0.37 and 0.34 (Table 5.1) for PcL and 3PO, meaning that 37% and 34% of one electron charge flowed from DMP to the T1 copper center. Therefore, no improvement is found on passing from the wild type to the mutant for binding mode **1**. On the other hand, P394H and N208S made room for an alternative enzyme-substrate conformation (**2**) in 3PO, represented in blue in Figure 5.5B, forming a hydrogen bond with H394 (Figure 5.7B). Importantly, the SD of **2** is now 0.66 (Table 5.1), significantly higher than the one found for the wild type. Taking into account all of the 20 selected structures in 3PO (mixing **1** and **2**), the average substrate's SD is 0.56.

Structures within 3 Å root mean square deviation (RMSD) of **2** represent only 0.2% of all the conformations generated during PELE simulation in PcL, none of them within 5 kcal/mol of the lowest binding energy structure. On the contrary the occurrence of **2** is roughly 50% of the best 5 kcal/mol binding energy interval in 3PO. (Figure 5.5B). Then, oxidation is improved in the

mutant by making this second and more reactive binding mode significantly available to the substrate.

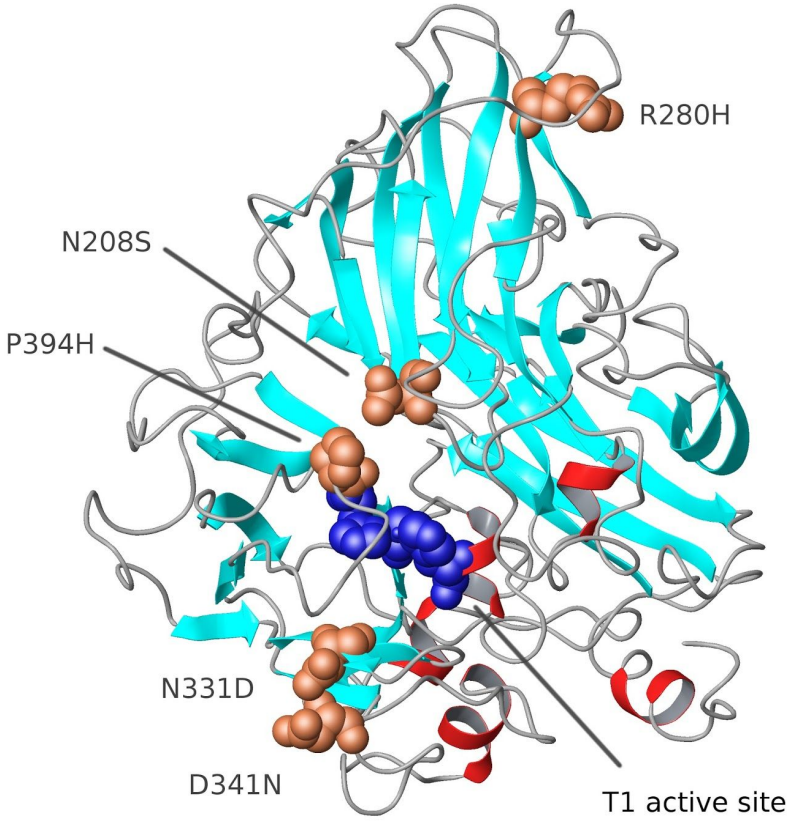


Figure 5.4. Structure of PcL where the active site T1 copper (and its coordination ligands) and mutated residues are highlighted in blue and light brown, respectively. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

Although the Michaelis constant (K_M) increased by almost the same amount as the catalytic rate constant (k_{cat}), most industrial application of enzyme catalysts work at high substrate concentration. In such conditions, k_{cat} is the most crucial component of catalytic efficiency. This is especially true for laccases, where substrate oxidation uses to be the rate-determining step. For this reason, the aim is to efficiently *score* k_{cat} through the SD rather than estimate K_M. In spite of that, K_M increase can be rationalized as well. As noted before, protein-ligand interaction energies are virtually the same in PcL and 3PO (Figure 5.5A and 5.5B). Meanwhile k_{cat} largely improved, increasing the Michaelis constant (from steady-state approximation, $K_M = (k_{off} + k_{cat}) / k_{on}$, where k_{on} and k_{off} are the association and dissociation rate

constants). Moreover, having two highly accessible reacting binding modes that can be filled simultaneously (as shown with ligand docking using Glide, Figure 5.6) requires a higher concentration of substrate to reach half of the maximum reaction rate.

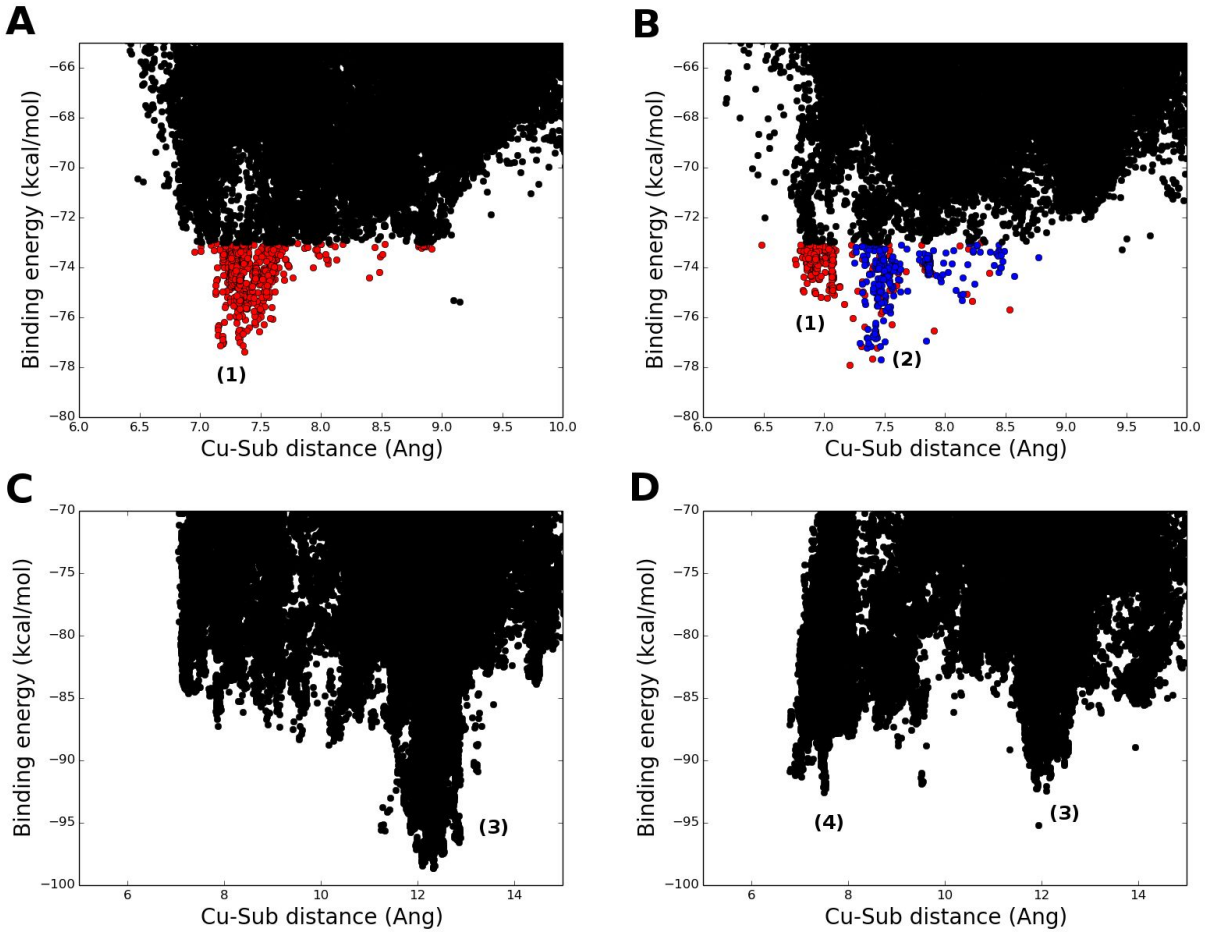


Figure 5.5. Results of PELE conformational search for DMP (A: PcL, B: 3PO) and ABTS (C: PcL, D: 3PO). "Cu-Sub distance" is the distance between the T1 copper and substrate's center of mass. In panels A and B, binding modes similar to Figure 5.7A are represented in red while binding modes similar to Figure 5.7B are in blue. Only structures within 5 kcal/mol of the best pose has been represented with colors. Numbers in parenthesis label the binding modes depicted in Figure 5.7 and are used through the text. Figure prepared with matplotlib (http://matplotlib.org/) and modified with GIMP (https://www.gimp.org/).

Table 5.1. Experimental kinetic data, simulated average SD on substrate ($\rho_{substrate}$) and copper with its coordinating axial atoms (ρ_{CuSNN}), simulated enzyme-radical stabilization ($E_{radical}$) of the oxidation of DMP and ABTS by PcL (wild type) and 3PO (mutant) at pH 5. Values in parentheses represent uncertainties on the last significative digits.

System	K_M (mM)	k_{cat} (s^{-1})	$\rho_{substrate}$	ρ_{CuSNN}	$E_{radical}$ (kcal/mol)
PcL-ABTS	0.020(1)	35.1(9)	0.85(9)	0.15	-89
PcL-DMP	0.012(2)	10.0(3)	0.37(12)	0.63	-81
3PO-ABTS	0.024(2)	482(10)	0.96(11)	0.04	-112
3PO-DMP	0.213(13)	197(3)	0.56(19)	0.44	-111

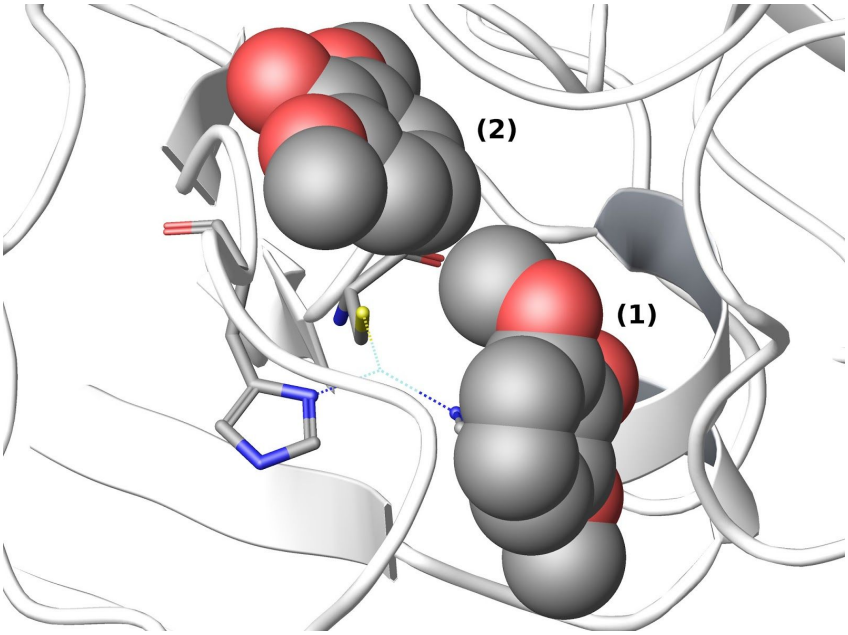


Figure 5.6. Possible binding mode of two DMP molecules (in VdW spheres representation) in the T1 active site as provided with Glide. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

5.3.1.2 ABTS oxidation

When simulating ABTS binding with PELE (240 independent trajectories, 48 hours each), a sharp minimum (**3**) is located next to the active site in the wild type Laccase (Figure

5.5C). In the corresponding structures, ABTS forms multiple hydrogen bonds; one of ABTS negatively charged sulfonate groups interacts with N331 and N389 while the other one interacts with two backbone hydrogens (Figure 5.7C). QM/MM calculations yield an average substrate's SD of 0.85 for **3**. Due to the N331D mutation, one stabilizing asparagine-sulfonate interaction no longer occurs in 3PO and the depth of above-mentioned minimum is sensibly reduced (Figure 5.5D). Simultaneously the enzyme-substrate interaction inside the T1 pocket is stabilized, as can be seen by the clear appearance of a second minimum (**4**) within 8 Å of the T1 copper. As a consequence, approximately one half of the 20 randomly picked structures populate this second minimum, interacting with S208 and H394 (Figure 5.7D). The overall substrate's SD for 3PO is 0.96 (being 1.00 for **4** and 0.86 for **3**). Similarly to DMP, the most reactive pose, **4**, is poorly accessible to the native enzyme (5% occurrence) and never observed within 5 kcal/mol of the lowest binding energy value (Figure 5.5C). Therefore, ABTS oxidation is improved by refining substrate recognition, shifting the population in favor of the most reactive binding mode, as observed for DMP oxidation. Moreover, in ABTS the average decrease in donor-acceptor distance and solvent exposure is more accentuated (the substrate docks deeper inside the protein). Therefore, improved electronic coupling certainly has an important role in the increased activity as well.

The experimental K_M did not change significantly for ABTS. This can be justified as a compensation between increased affinity at **4** and higher rate constant since $K_M = (k_{off} + k_{cat}) / k_{on}$. Moreover, contrary to DMP, at substrate saturation the coexistence of **3** and **4** is implausible since ABTS carries two negative charges that would repel each other at such a close distance.

5.3.1.3 Determinants of the driving force improvement

As mentioned before, the reaction rate is mostly driven by the free energy difference between product and reactant of the first oxidation step. This driving force is equal to the ΔE^o between T1 copper and the oxidized substrate. Since the mutations introduced are not expected to significantly affect the E^o_{T1} (just one is placed on the T1 copper second coordination sphere) and the protein-reactant interaction energies remains nearly identical in the mutant, optimization of the enzyme pocket towards the oxidized substrate must play a major role. To assess this point, a new PELE simulation was set up (48 independent trajectories, 48 hrs each) in which the substrate was modeled in its oxidized state. This was done using the atomic charges obtained from the quantum geometry optimization of the oxidized (radical) species. In such a way, the enzyme-substrate interaction energy represents the degree of stabilization of the oxidized

substrate by the protein environment. As shown in the last column of Table 5.1, 3PO turned out to be clearly a more favorable host for the oxidized species than PcL, since the minimum substrate-enzyme classical (force field) interaction energy is significantly lower in the mutant (Table 5.1). Notably, these classical simulations results (improved enzyme-radical stabilization in 3PO, Table 5.1) agree with the increase in substrate's SD seen in the (slow/expensive) quantum mechanics simulations. Such scenario opens the door to implement fast computational screening of mutants using classical ligand sampling to improve the enzyme-radical interaction energy through several mutations cycles. Subsequent SD calculations could be used at the end of each cycle to rescore the most promising mutants.

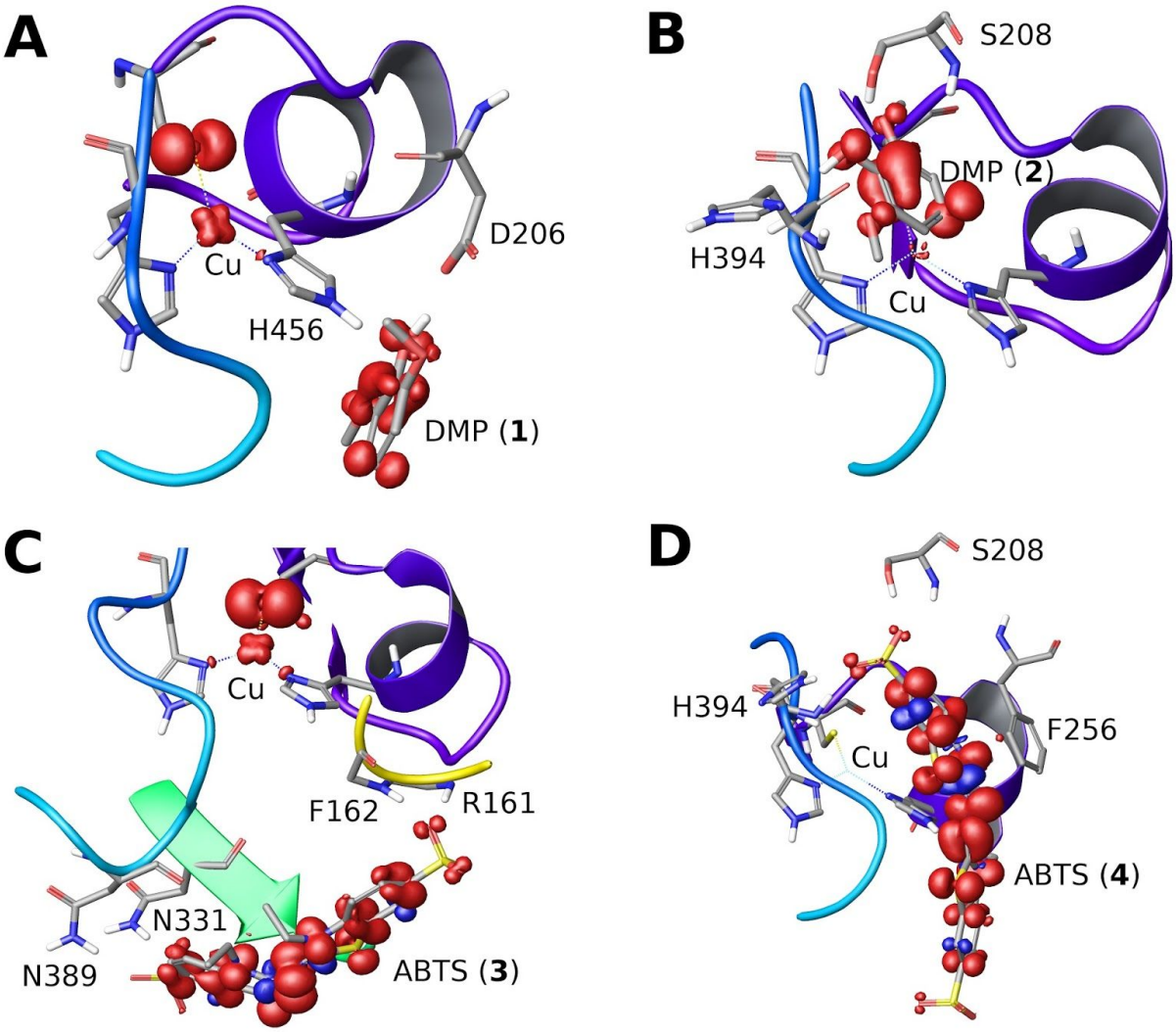


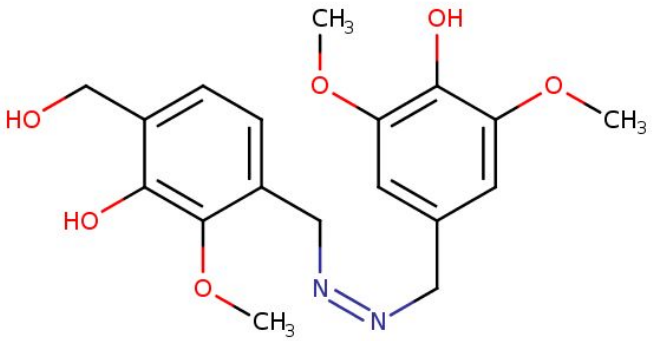
Figure 5.7. Most reactive binding modes of the enzyme-substrate complexes with SD isosurfaces. Numbers in parentheses label each binding mode and are used through the text. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

5.3.1.4 Closure

Molecular simulations indicate that mutations accumulated during directed evolution increase laccases enzymatic activity by affecting substrate binding rather than E^{o}_{T1}. Extensive conformational sampling revealed important changes in the protein-ligand energy landscape upon mutation. Then, quantum chemical calculations confirmed an oxidation (substrate's SD) increase as a result of an enhanced electrostatic stabilization of the radical species. Therefore, the oxidation rate of a target substrate can be improved by fine tuning the binding event.

5.4 Oxidation of syringaldazine

As reported in Chapter 4, there are some examples in literature where lower E^{o}_{T1} laccases display better k_{cat} than their higher E^{o}_{T1} counterparts. In particular, the oxidation of syringaldazine (SGZ), by MtL (E^{o}_{T1}=460 mV, k_{cat}=1100 min^{-1}) has been shown to display improved kinetics over PcL (E^{o}_{T1}=790 mV, k_{cat}=180 min^{-1}) (Li, Xu & Eriksson, 1999). Here, aiming at understanding these differences, the oxidation of SGZ by MtL and PcL is investigated using the PELE+QM/MM protocol (96 PELE trajectories, 20 structures selected).



SGZ

Scheme 5.3. Chemical structure of SGZ. Prepared with MarvinSketch (http://www.chemaxon.com).

5.4.1 Results and discussion

An inspection of the selected structures shows that SGZ is optimally oriented for both electron and proton transfer in MtL and not in PcL (Figure 5.8). Such distinct binding of SGZ in MtL and PcL derives from differences in the T1 pocket. MtL contains a large loop involving residues 445 and 468 hosting N454 that anchors the substrate in the position seen in Figure 5.8A. PcL, on the contrary, has a much smaller loop of only 8 residues (opposed to the 23 in MtL) involving residues 408 to 416. Another significant difference is the shorter loop hosting residues 332 to 336 in PcL (364 to 371 in MtL) that creates a favorable environment for the docking the polar hydroxyl group of SGZ (Figure 5.8B). Finally, A192, which helps positioning the substrate in the correct orientation in MtL's active site, is located at the end of a loop that is preceded by an α-helix beginning in residue 179 which is not present in PcL (Figure 5.8). In addition to the considerable differences seen both in the binding orientation, QM/MM calculations also indicate higher spin transfer from SGZ to MtL than in PcL (Figure 5.9).

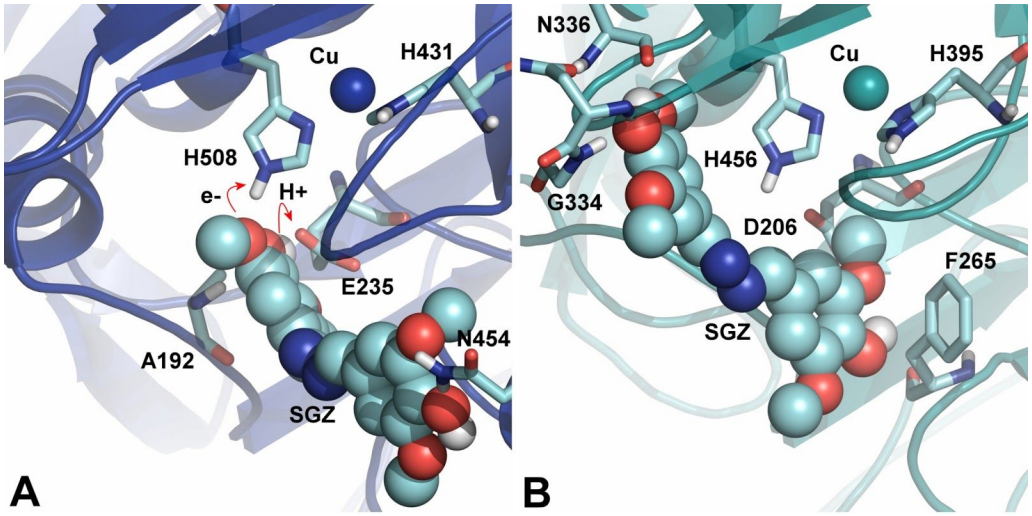


Figure 5.8. Representative binding modes for SGZ in A) MtL and B) PcL. Figure prepared with Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and modified with GIMP (https://www.gimp.org/).

The computational results put in evidence the fact that counterintuitive differences in SGZ's oxidation by these two proteins are highly related to the binding event. In MtL, SGZ favors configurations hydrogen bonded (H-bonded) to the first coordination H508 (ET) and simultaneously interaction with D235 where proton transfer is expected to occur. This will ensure optimal electronic coupling (shorter donor-acceptor distance) and ET driving force (higher substrate SD, likely due to the proximity to the catalytic D235). In PcL, however, SGZ

interacts with the backbone of residues G334 and N336 and with the side chain of F265 anchoring the substrate about 3 Å away from any of the T1 ligands. In this higher redox protein, ET must thus occur through a free space jump (as none of the coordinated residues is close enough to the substrate) or alternatively use other residues (which necessary imply a longer ET path). These results advocate for improved kinetics in MtL, which is in good agreement with experimental observations.

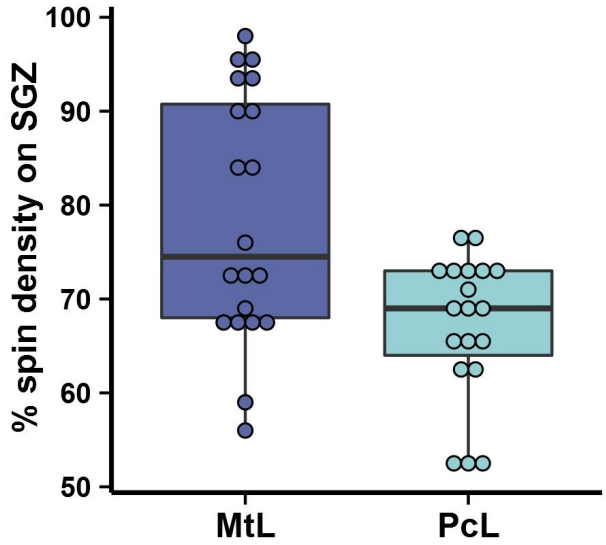


Figure 5.9. Distribution of SGZ spin densities in MtL and PcL. Figure prepared with R (https://www.r-project.org/).

5.5 Oxidation of para-substituted phenols: a step further

It has been shown in Section 5.1 that the SD correlates well with the kinetic data of four 4-X-PhOHs with X = {OH, OCH_3, OCH_3, Cl} (named from this point on as 4OH, 4MeO, 4Me and 4Cl, respectively). Here, the analysis is extended to MtL, whose T1 copper has a lower E^o_{T1} than PcL's (0.46 vs. 0.79 V), aiming to check if the method can discriminate not only different substrates but also completely different active sites. The phenols were docked with PELE in the active site of the two laccases.

5.5.1 Results and discussion

Preliminary PELE+QM/MM simulations of 4-X-PhOH oxidation by PcL show poor correlation between the amount of SD computed on each substrate (20 structures selected) and the kinetic data (Figure 5.10). Further inspection of the most favorable minima evidenced an

important cluster of structures, for all complexes, that included interaction with the catalytic H456 but not with the catalytic aspartic acid D206 (Figure 5.11A). These structures exposed a region in PcL's binding pocket with a volume comparable to that occupied by a water molecule. Further inspection of the available laccases' crystal structures that contained a phenolic substrate showed, for example, that in *Melanocarpus albomyces* laccase (MaL) in addition to 2,6-dimethoxyphenol (DMP) it has a crystal water molecule precisely in the position identified in the simulations (Figure 5.11B). For this reason, the calculations were repeated including an extra water molecule in this position.

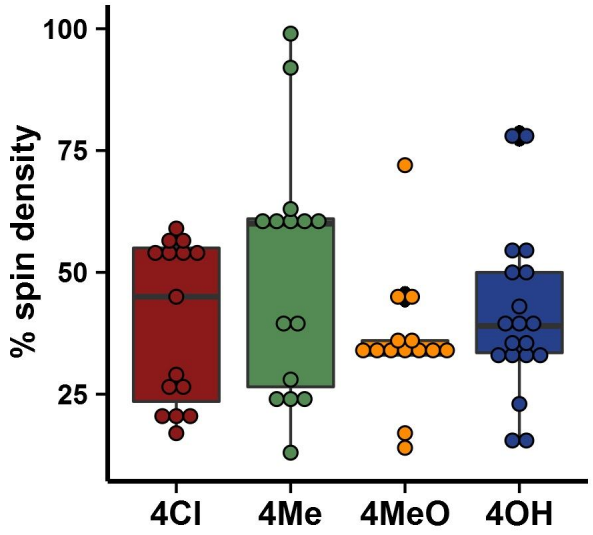


Figure 5.10. Computed spin densities for 4-X-PhOH with PcL. Figure prepared with R (https://www.r-project.org/).

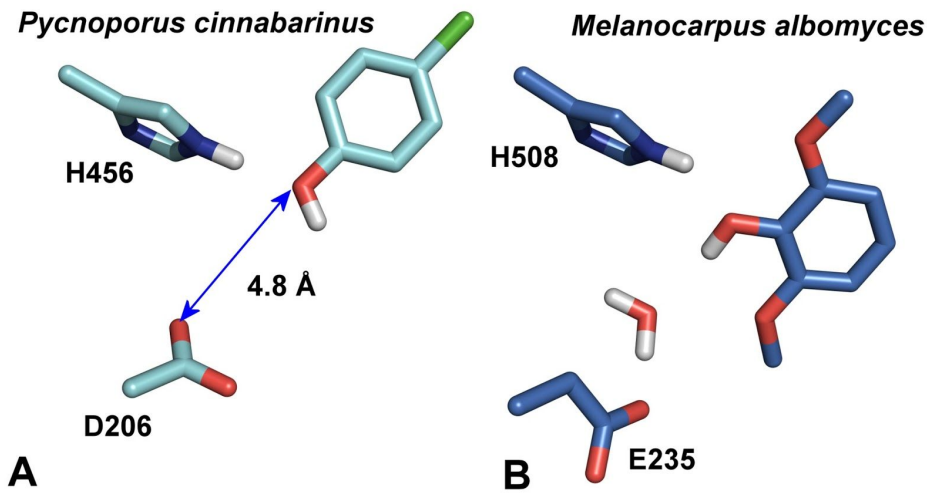


Figure 5.11. Binding pocket comparison between A) PELE minimum energy structure in *Pycnoporus cinnabarinus* and B) the T1 binding pocket of *Melanocarpus albomyces* complexed with 2,6-dimethoxyphenol (crystal structure). Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

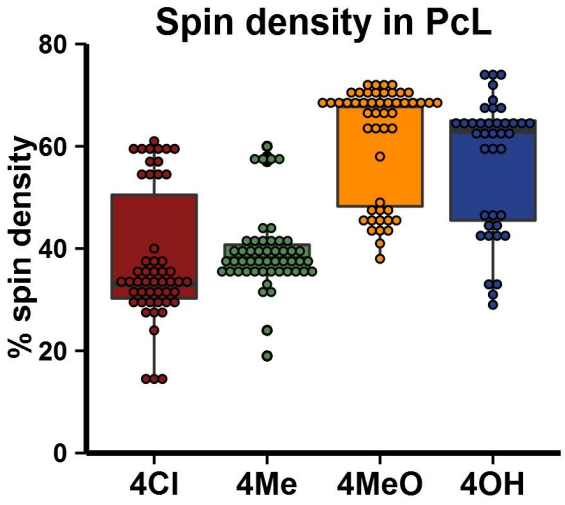


Figure 5.12. Computed spin densities for 4-X-PhOH with PcL when a water molecule is placed in the active site. Figure prepared with R (https://www.r-project.org/).

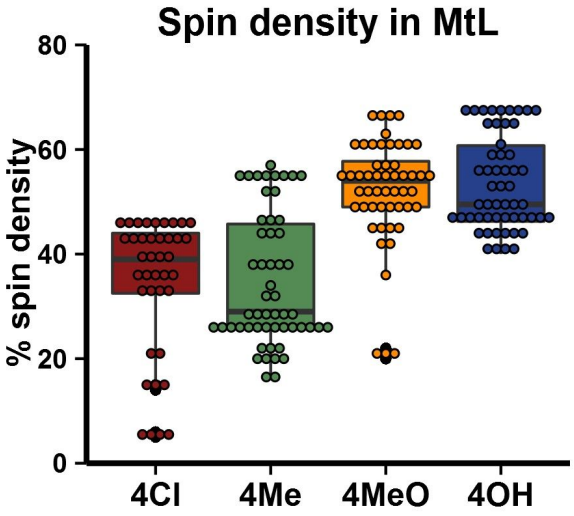


Figure 5.13. Computed spin densities for 4-X-PhOH with MtL when a water molecule is placed in the active site. Figure prepared with R (https://www.r-project.org/).

For each laccase-substrate pair, 240 independent 48 hours trajectories were produced and 50 complex structures were randomly selected for QM/MM scoring. Computed spin densities for all systems (depicted in Figures 5.12 and 5.13) offer a qualitative picture where the overall improved oxidation of compound 4OH over 4Cl can be appreciated. Moreover, phenols with PcL display, in average, higher spin transfer than with MtL. Next, the site with maximum substrate SD was identified for each compound by clustering techniques. For this, computed spin densities were grouped using k-means and k-medoids (employing R, https://www.r-project.org/) and the obtained clusters were then visually inspected. For both laccases, the highest SD cluster is also the one with optimal catalytic contacts. The substrate is H-bonded to the catalytic histidine (first electron acceptor, coordinated to the T1 copper) and to the water molecule present in the active site. This is in turn hydrogen bonded to the catalytic Asp, the final proton acceptor (Figure 5.14).

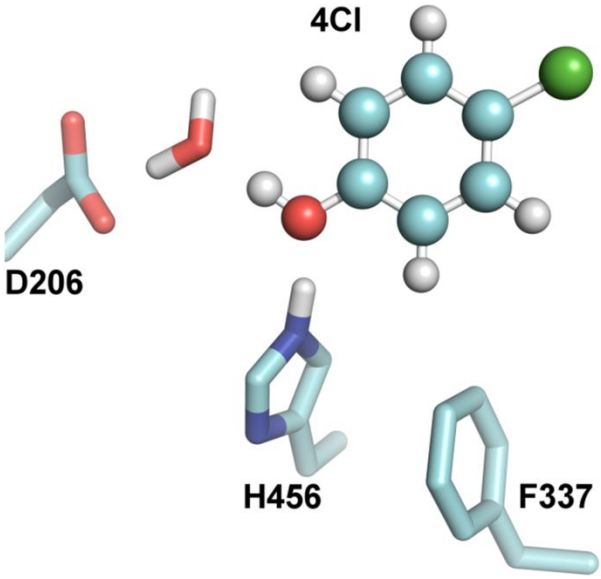


Figure 5.14. Binding positions for 4Cl-phenol in PcL. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

It follows that such binding mode, besides from an excellent driving force (most likely due to the proximity of the negatively charged catalytic base), provides both optimal electron tunneling and proton abstraction. If only this cluster of best oxidation position is taken into account, for each system, a good correlation (above 90%) between the computed SD and the logarithm of the experimental specificity constants is obtained (Figure 5.15). It is worth noting that, since specificity constants are not available for PcL, experimental data for a high E^{o}_{T1} laccase from

Trametes villosa which shares 70% identity with PcL are employed. These small differences may be responsible for the lower correlation found for this system.

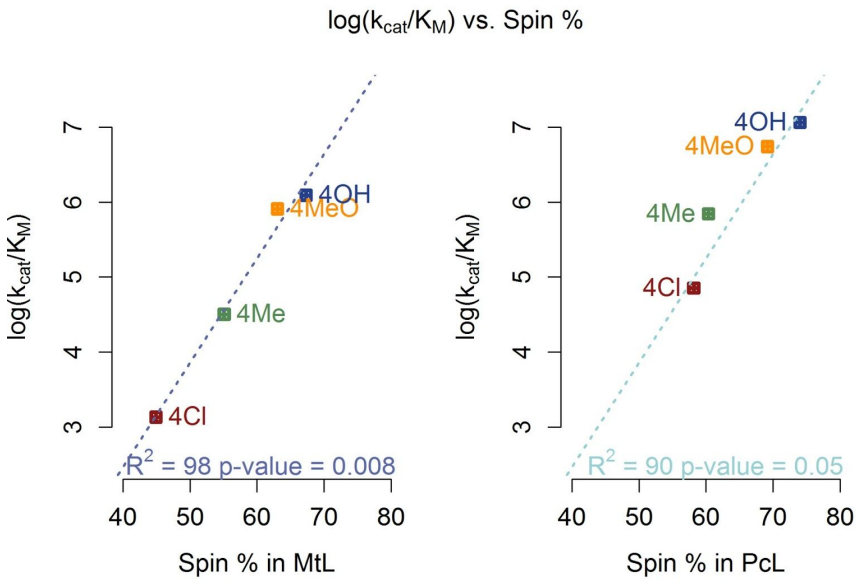


Figure 5.15. Computed average spin densities in the sites identified as the best oxidation positions in the T1 copper site for MtL and PcL vs. experimental specificity constants for MtL (blue) and TvL (cyan) extracted from (Tadesse et al., 2008). Figure prepared with R (https://www.r-project.org/).

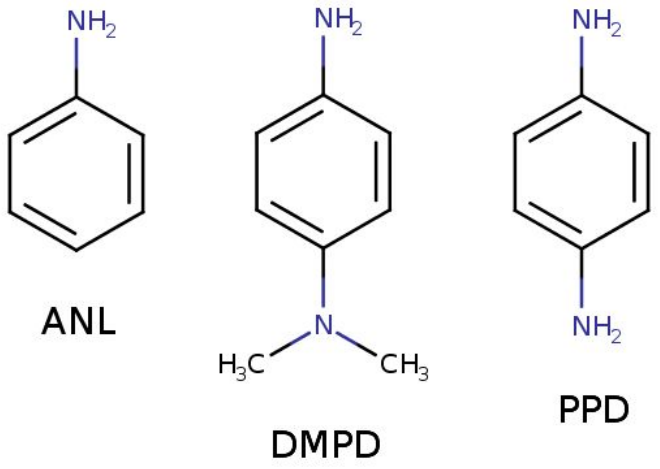
In any case, results strongly assess the capability of the PELE+QM/MM protocol to score laccase reactivity even on a quantitative ground (although the problem is expected to be much more complex when comparing very structurally distinct substrates). The methodology can discriminate between distinct substrates and between diverse T1 pockets; for any substrate, the laccase with higher specific constant is the one that displays the larger SD (Figure 5.15). When aiming at quantitative precision, a correct description of the binding site, namely with the inclusion of essential water molecules, is fundamental. Furthermore, when multiple binding modes exist (within 5 kcal/mol of the minimum PELE binding energy pose), if they are non-competitive, then the oxidation is mainly driven by the best oxidation position and clustering methods can help finding it.

5.6 First rational design attempt: improving aniline oxidation at low pH

Aniline (ANL) and its derivatives have been traditionally used in the dyestuff industry, as precursors of aniline dyes and azo dyes for textiles, hair, leather or printing (Aftalion, 2001), and

also as components for engineering polymers, composites and rubbers. In particular, the synthesis of conductive polyaniline (PANI) has been widely investigated during the last two decades due to its wide range of applications (sensor devices, rechargeable batteries, etc. (Bhadra et al., 2009). Conducting PANI macromolecules have regular head-to-tail linked monomers (over 95%) but high conductive polymer is only produced in strongly acidic conditions (pH<2.5) using oxidants such as ammonium peroxydisulfate (Sapurina, Yu. Sapurina & Stejskal, 2012). To replace these processes by enzymatic bioconversion is thus a main goal. However, aniline (pK_a=4.6) is mostly protonated at low pH, raising its Eº from 0.63 to 1.05 V. Then, it becomes clear why laccases normally cannot oxidize this compound or reactions occur too slowly (Bornscheuer et al., 2012).

Here, the PELE+QM/MM protocol was successfully applied to rationally re-design the active site of the 7D5 laccase, aiming to improve the oxidation of ANL at pH 3, with subsequent formation of PANI. This result was achieved by enhancing ANL's oxidation potential within the T1 pocket, creating a better electrostatic environment for the oxidized species. It is worth noting that the starting laccase was already engineered in the lab to improve aniline oxidation.



Scheme 5.4. Chemical structure of ANL, DMPD and PPD. Prepared with MarvinSketch (http://www.chemaxon.com).

5.6.1 Results and discussion

Firstly, ANL was docked into the active site with 120 independent PELE simulations, each lasting 48 hours. The most favourable binding mode is illustrated in Figure 5.16A. Residues 263, 264, 207 and 205 have their side chains in proximity of ANL. Therefore, N263 and N207 were mutated to Asp and Ser, aiming to either lower ANL's highest occupied

molecular orbital (HOMO) energy (to increase its oxidation potential) or to re-adjust ANL deeper inside the pocket. D205, on the other hand, was mutated to Glu, trying to shorten the contact between ANL and the carboxyl group. The PELE+QM/MM workflow was applied to such mutants, with N207S and N263D yielding the highest SD values (0.8 and 0.10 respectively, Table 5.2). However, PM1 (wild type) has a SD of 0.12 itself (Table 5.3), meaning that the two single mutants have not improved over the wild type enzyme. One possibility is that these two mutations can synergistically reinforce each other, producing a better catalysts. Therefore, the N207S-N263D double mutant was prepared and subjected to the PELE+QM/MM protocol. It is worth noting that such mutations exist in a *S. Cerevisiae* Fet3 (PDB: 1ZPU), a multicopper oxidase which shares 29% sequence identity with PM1. Consequently, their introduction in the wild type protein should not affect stability dramatically.

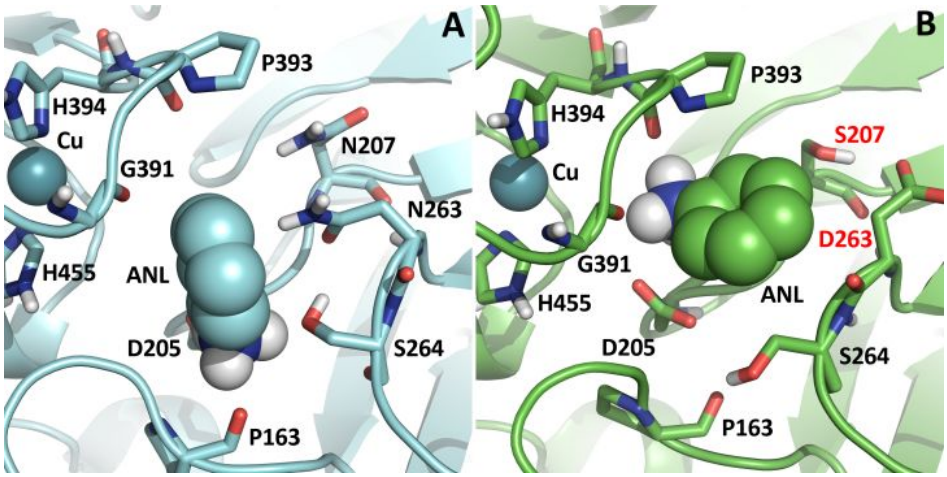


Figure 5.16. Most favourable binding mode for ANL in A) PM1 and B) DM. Figure prepared with Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and modified with GIMP (https://www.gimp.org/)

Table 5.2. Simulated average substrate's SD.

Original residue	Mutation	SD
D205	E	0.00
N207	D S	0.08 0.03
N263	D S	0.10 0.01

The *in silico* double mutant (DM) undergoes a two-fold improvement in SD, which was confirmed in the lab (by our collaborator at Centro Superior de Investigación Científica in Madrid) by measuring a two-fold improvement in k_{cat}. Thanks to the double mutation, the binding mode is completely changed: N207S creates room for the substrate to rearrange in the active site while N263D introduces a negative charge, which increase the oxidation potential of ANL in the active site (Figure 5.16B). This result confirms the capability of the PELE+QM/MM protocol to design *fit-for-purpose* laccases.

Table 5.3. Simulated average substrate's SD and experimental kinetic data for the wild type (PM1) and the DM. Values in parentheses represent uncertainties on the last significative digits.

System	SD	k_{cat} (s^{-1})	K_M (mM)
PM1-ANL	0.12(1)	10(1)	28(7)
DM-ANL	0.26(5)	22(3)	59(16)
PM1-DMPD	0.39(7)	459(18)	1.7(2)
DM-DMPD	0.68(14)	741(48)	1.2(2)
PM1-PPD	0.32(5)	14.7(6)	3.7(3)
DM-PPD	0.46(14)	21(2)	3.(8)
PM1-ABTS	0.23(14)	219(18)	0.004(1)
DM-ABTS	0.40(7)	570(26)	1.001(1)

While testing the DM *in vitro*, other substrates were oxidized and the following showed improved activity (Table 5.3): ABTS, N,N-dimethyl-p-phenylendiamine (DMPD) and p-phenylendiamine (PPD). In order to further cross validate the PELE+QM/MM protocol, the oxidation of these compounds was simulated, yielding an increased average substrate SD for the DM, in line with the experimental results (Table 5.3).

5.7 Closure

Since substrate oxidation is characterized by large reorganization energies, this step is usually the slowest of the catalytic cycle and it is governed by the ET driving force. It follows that the QM/MM SD, calculated in vacuum and with only five geometry optimization steps, can be used to score the oxidation rate constant of laccases variants, provided that donor-acceptor distances do not change dramatically among them. Incidentally, this method is conceptually valid for any oxidoreductase whose rate-determining step is controlled by the ET driving force. As a cautionary statement, it is worth noting that neglecting the solvent might exaggerate the SD of highly charged systems. For example, a substrate with two or more negative charges might carry all the SD of the system already in the wild type. In such cases, the inclusion of an explicit solvent model might be necessary, requiring a higher number of structures to estimate the average SD.

The PELE+QM/MM approach proposed here underlined that substrate binding is fundamental in laccases function and it can be fine tuned to improve catalysis. While stabilizing binding poses with optimized catalytic contacts can improve electronic coupling, enhancing the electrostatic stabilization of oxidized substrate can push the ET driving force toward lower values, increasing its match with the negative image of the reorganization energy. Such PELE+QM/MM binding-focused strategy has been employed to increase the oxidation of ANL at acidic pH, improving the production of PANI. At the best of my knowledge, this is the first successful computational re-design of a laccase.

REFERENCES:

Aftalion F. 2001. *A History of the International Chemical Industry*. Chemical Heritage Foundation.

Artz K., Williams JC., Allen JP., Lendzian F., Rautter J., Lubitz W. 1997. Relationship between the oxidation potential and electron spin density of the primary electron donor in reaction centers from Rhodobacter sphaeroides. *Proceedings of the National Academy of Sciences of the United States of America* 94:13582–13587.

Becke AD. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review. A* 38:3098–3100.

Becke AD. 1993a. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of chemical physics* 98:1372.

Becke AD. 1993b. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics* 98:5648.

Bhadra S., Sambhu B., Dipak K., Singha NK., Lee JH. 2009. Progress in preparation, processing and applications of polyaniline. *Progress in polymer science* 34:783–810.

Blumberger J. 2015. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical reviews* 115:11191–11238.

Bornscheuer UT., Huisman GW., Kazlauskas RJ., Lutz S., Moore JC., Robins K. 2012. Engineering the third wave of biocatalysis. *Nature* 485:185–194.
Bortolotti CA., Siwko ME., Elena C., Antonio R., Marco S., Stefano C. 2011. The Reorganization Energy in Cytochrome c is Controlled by the Accessibility of the Heme to the Solvent. *The journal of physical chemistry letters* 2:1761–1765.
Camarero S., Pardo I., Cañas AI., Molina P., Record E., Martínez AT., Martínez MJ., Alcalde M. 2012. Engineering platforms for directed evolution of Laccase from Pycnoporus cinnabarinus. *Applied and environmental microbiology* 78:1370–1384.
Catlow CRA., French SA., Sokol AA., Thomas JM. 2005. Computational approaches to the determination of active site structures and reaction mechanisms in heterogeneous catalysts. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 363:913–36; discussion 1035–40.
Friesner RA., Banks JL., Murphy RB., Halgren TA., Klicic JJ., Mainz DT., Repasky MP., Knoll EH., Shelley M., Perry JK., Shaw DE., Francis P., Shenkin PS. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* 47:1739–1749.
Gordon JC., Myers JB., Folta T., Shoja V., Heath LS., Onufriev A. 2005. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* 33:W368–71.
Guallar V., Wallrapp F. 2008. Mapping protein electron transfer pathways with QM/MM methods. *Journal of the Royal Society, Interface / the Royal Society* 5 Suppl 3:S233–9.
Hay PJ., Jeffrey Hay P., Wadt WR. 1985. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *The Journal of chemical physics* 82:270.
Jacobson MP., Friesner RA., Zhexin X., Barry H. 2002. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of molecular biology* 320:597–608.
Jacobson MP., Pincus DL., Rapp CS., Day TJF., Honig B., Shaw DE., Friesner RA. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins* 55:351–367.
Johansson MP., Blomberg MRA., Sundholm D., Wikström M. 2002. Change in electron and spin density upon electron transfer to haem. *Biochimica et biophysica acta* 1553:183–187.
Kaminski GA., Friesner RA., Julian T-R., Jorgensen WL. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *The journal of physical chemistry. B* 105:6474–6487.
Lee C., Chengteh L., Weitao Y., Parr RG. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B: Condensed Matter and Materials Physics* 37:785–789.
Li K., Xu F., Eriksson KE. 1999. Comparison of fungal laccases and redox mediators in oxidation of a nonphenolic lignin model compound. *Applied and environmental microbiology* 65:2654–2660.
Meisel D., Dan M., Neta P. 1975. One-electron redox potentials of nitro compounds and radiosensitizers. Correlation with spin densities of their radical anions. *Journal of the American Chemical Society* 97:5198–5203.
Mohamadi F., Fariborz M., Richards NGJ., Guida WC., Rob L., Mark L., Craig C., George C., Thomas H., Clark Still W. 1990. Macromodel?an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of computational chemistry* 11:440–467.
Murphy RB., Philipp DM., Friesner RA. 2000. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *Journal of computational chemistry* 21:1442–1457.
Olsson MHM., Søndergaard CR., Rostkowski M., Jensen JH. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation* 7:525–537.
Perdew JP. 1986. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B: Condensed Matter and Materials Physics* 33:8822–8824.
Perdew JP., Kieron B., Matthias E. 1996. Generalized Gradient Approximation Made Simple. *Physical review letters* 77:3865–3868.
Sapurina IY., Yu. Sapurina I., Stejskal J. 2012. Oxidation of aniline with strong and weak oxidants. *Russian Journal of General Chemistry* 82:256–275.

Sastry GM., Adzhigirey M., Day T., Annabhimoju R., Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* 27:221–234.

Tadesse MA., D'Annibale A., Galli C., Gentili P., Sergi F. 2008. An assessment of the relative contributions of redox and steric issues to laccase specificity towards putative substrates. *Organic & biomolecular chemistry* 6:868–878.

Vosko SH., Wilk L., Nusair M. 1980. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian journal of physics* 58:1200–1211.

Wallrapp FH., Voityuk AA., Victor G. 2013. In-silico Assessment of Protein-Protein Electron Transfer. A Case Study: Cytochrome c Peroxidase – Cytochrome c. *PLoS computational biology* 9:e1002990.

Xu F. 1996. Oxidation of phenols, anilines, and benzenethiols by fungal laccases: correlation between activity and redox potentials as well as halide inhibition. *Biochemistry* 35:7608–7614.

Xu F., Kulys JJ., Duke K., Li K., Krikstopaitis K., Deussen HJ., Abbate E., Galinyte V., Schneider P. 2000. Redox chemistry in laccase-catalyzed oxidation of N-hydroxy compounds. *Applied and environmental microbiology* 66:2052–2056.

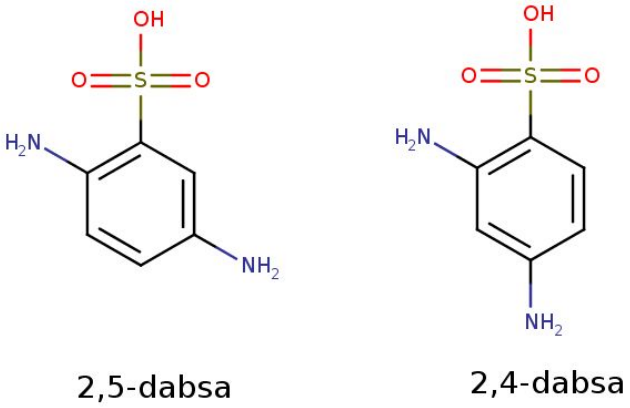
Zhao Y., Yan Z., Truhlar DG. 2008. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and 12 other functionals. *Theoretical chemistry accounts* 119:525–525.

6. High-throughput computational evolution of a fungal laccase with PELE

In the previous chapter, the development, testing and application of a computational methodology for rationally designing laccases has been presented. In particular, in Section 5.3, a correspondence between increased spin density and enhanced enzyme-oxidized substrate stabilization has been found in an evolved laccase toward two substrates. Such finding suggests that a fully classic design (using a molecular mechanics approach) of laccases could be undertaken; while reorganization energy and electronic coupling could be approximated with the solvent-accessible surface area (SASA) (Bortolotti et al., 2011) and donor-acceptor distance (Blumberger, 2015) respectively, optimizing the enzyme-oxidized substrate interaction energy could produce mutants with better (lower) electron transfer (ET) driving force values. It is worth noting that the interaction energy difference between product and reactant would be a more reliable parameter for the driving force (which is a free energy difference). However, its use implies two drawbacks: i) it doubles the number of calculations required; ii) the error of its estimation is the sum of the uncertainties of the single interaction energies, which requires a more exhaustive conformational sampling.

Another design task could be the stabilization of the reactive poses, i.e. the ones with the right catalytic contacts (with the His and Asp electron and proton acceptors), targeting the electron tunnelling probability (electron coupling). The rationale behind this strategy is the prominent promiscuity of laccases, which is most likely the consequence of a superficial, shallow binding site. Thanks to this feature, laccases interaction with most of the putative substrates renders ET possible. Nonetheless, it cannot bind all of them optimally to maximize ET. Therefore, the T1 pocket could be modified *tout court* to lock the desired substrate in the most reactive conformation, anchoring it with polar groups. This classical mechanics-based strategy is applied here to improve the oxidation of the 2,4-diamino-benzenesulfonic (2,4-dabsa) acid by POXA1b laccase. This is a fungal (*Pleurotus ostreatus*) laccase with an unusual high stability at alkaline pH (Giardina et al., 1999). The substrate of interest, 2,4-dabsa, is a dye precursor just like its constitutional isomer, 2,5-dabsa (Scheme 6.1). Although 2,4-dabsa is significantly cheaper than 2,5-dabsa, the latter is readily oxidized by POXA1b at neutral pH 8

and 60 ºC while the former is not. Therefore, there is interest in redesigning POXA1b to oxidize 2,4-dabsa, aiming to lower the cost of dye production.



Scheme 6.1. Chemical structures of 2,5- and 2,4-dabsa. Prepared with MarvinSketch (http://www.chemaxon.com).

6.1 Computational details

Homology modeling and system setup. POXA1b structure was generated through Prime homology modeling (Jacobson et al., 2002, 2004), using 1GYC.pdb (60% sequence identity) as a template. POXA1b is characterized by a longer C-terminal tail, which could be placed outside the oxygen cavity (model_out, calculated in this thesis) or inside (model_in, previously calculated by the collaborators of this study at Università degli Studi di Napoli (Festa et al., 2008)) (Figure 6.1). The POXA1b structures (model_out and model_in) were prepared with Protein Preparation Wizard (Sastry et al., 2013). The protonation states of titratable residues were generated with PROPKA (Olsson et al., 2011) and double-checked with the H++ server (Gordon et al., 2005), simulating pH 8.

Molecular dynamics (MD). MD simulations were performed with GROMACS (Pronk et al., 2013) for model_out and model_in, aiming at a refinement of the initial homologous structures. The enzymes were solvated with a 10 Å buffer of waters in an dodecahedral box, adding enough ions for neutralization and a 0.15 M NaCl buffer. The systems prepared were equilibrated as follow: i) solvent minimization; ii) system minimization; iii) 200 ps system warm up from 15 to 298 K at constant volume; iv) 200 ps NVT simulation; v) 200 ps NpT simulation. Then, a 200 ns NpT production run was performed at 298 K for each model. The AMBER99 force-field (Duan et al., 2003) and the SPC explicit water model (Toukan, Kahled & Aneesur,

1985) were used. The copper ions, modeled in their reduced (+1) state to attenuate charge repulsion in the trinuclear cluster of copper ions (TNC). The coordinating atoms and their nearest neighbours were restrained to their initial positions with stiff 10,000 kJ mol^{-1} nm^{-2} harmonic constraints. The temperature was regulated with velocity rescaling (Bussi, Donadio & Parrinello, 2007) with a relaxation time of 0.1 ps, and the pressure was controlled with the Parrinello-Rahman barostat (Parrinello, 1981) with isotropic coupling and a relaxation time of 2.0 ps. The LINCS algorithm (Hess et al., 1997) was employed to constrain all bond lengths, allowing a time step of 2.0 fs. A 10 Å cutoff was used for non bonded interactions together with the particle mesh Ewald method (Essmann et al., 1995). It is well known that homology models can fall apart when subjected to MD (Raval et al., 2012), thus a 1,000 kJ mol^{-1} nm^{-2} harmonic constraint was applied to the secondary structure elements (helices and sheets) for the first 40 ns and lowered to 10 kJ mol^{-1} nm^{-2} for the rest of the simulation.

PELE sampling. PELE sampling was applied to the POXA1b-2,4-dabsa system. The substrate, initially placed in front of the T1 cavity, was randomly translated and rotated within 15 Å of the T1 copper. The ANM perturbation was applied to all the C_α in the system, the side chains of the residues within 10 Å of substrate were repacked and everything within 20 Å was energy minimized (see Appendix 1 for an example of input file). Along the text, tables and figures, (PELE) binding energies refer to protein-substrate interaction potential energies.

PELE design. The initial conformation for the design was selected from PELE sampling among the structures that exhibit catalytic contacts. Then, every residue within 10 Å of the substrate was mutated to all the possible remaining 19 amino acids. The mutated residues are: T160, G161, V162, P163, H164, C204, S206, N207, A239, N263, S264, F331, P333, A336, A391, G392, P393, P395, W455, P510, L511. The ligand was perturbed with very small translations and rotations while the protein backbone was subjected to small ANM displacements. Every residue within 10 Å of the substrate was energy minimized (see Appendix 2 for an example of input file). The simulation of each mutant (and the wild type) lasted 1 hour, accumulating ~30 Monte Carlo steps (~20 accepted). The enzyme-substrate interaction energy was used to score each mutant, while keeping an eye on substrate's SASA.

Treatment of substrate and metal centers in PELE. Substrates' geometries were optimized at the M06/6-31G* level of theory, modeling the solvent with the Poisson-Boltzmann model. The atomic ESP charges were calculated from the optimized structure and used for PELE. The OPLS parameters were assigned to the ligand with Schrödinger's hetgrp_ffgen utility and a rotamer library was generated with in-house PELE's scripts using Macromodel

(Mohamadi et al., 1990). The copper ions and their coordination atoms plus their nearest neighbors were constrained during PELE sampling with 200 kcal mol^{-1} $Å^{-2}$ harmonic constraints on their atomic positions. Unitary charges were assigned to each copper atom (to avoid tension in the TNC).

6.2 Results and discussion

6.2.1 Refinement of POXA1b models

As mentioned in Section 6.1, POXA1b shares only 60% sequence identity with its homology modeling template and it is characterized by a longer C-terminal tail, which could be placed outside (model_out) or inside (model_in) the oxygen cavity (Figure 6.1). For this reason, MD simulations were performed to refine the models and establish whether model_out or model_in (or none of them) is the best representation of the system. As a result, model_out converges in ~50 ns (Figure 6.2B, stopped at ~160 ns), keeping the C-terminal away from the oxygen cavity (Figure 6.2A). On the contrary, model_in is rather unstable (Figure 6.2B), resulting in the escape of the C-terminal from the protein interior, roughly matching model_out (Figure 6.2A). It follows that, according to the simulation, the C-terminal is more stable on the protein surface than in its interior. This is somewhat surprising, since *Melanocarpus albomyces* laccases, which are also resistant at alkaline pH, accommodate their C-terminal tail inside the oxygen cavity. Moreover, the C-terminal of model_out directly interacts with the active site, adding a further degree of complexity to the design.

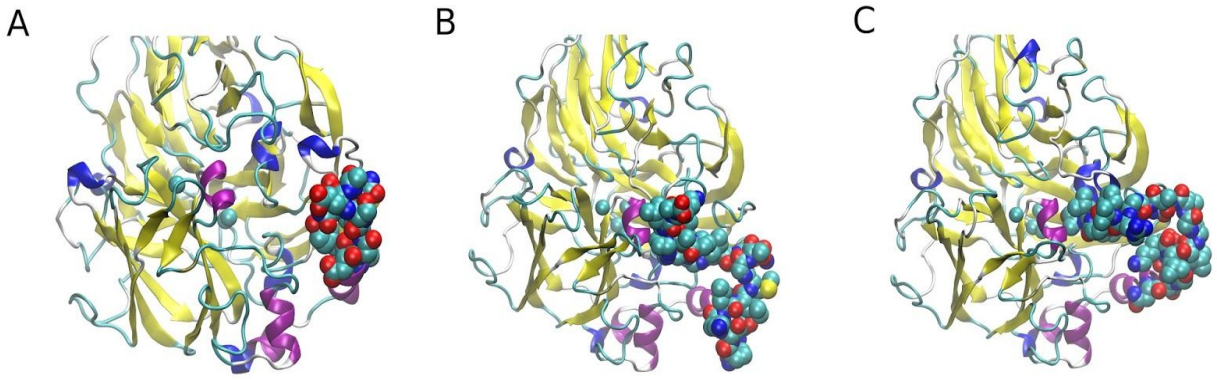


Figure 6.1. A) 1GYC.pdb; B) model_out; C) model_in. C-terminal tail and Cu ions are represented in VDW spheres. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

6.2.2 PELE sampling and design

Binding of 2,4-dabsa into the T1 pocket was simulated with 96 independent 48 hours lasting trajectories. Inspection of the structures within 5 kcal/mol of the lowest PELE binding energy value (within 12 Å) reveals that no catalytic contact (i.e. H-bonds with the catalytic His and Asp) is engaged (Figure 6.3). In order to find a pose that displays catalytic interactions, it is necessary to climb the PELE binding energy ladder by ~20 kcal/mol (note that this value is a protein-ligand interaction energy, not a proper binding energy). Such energy difference poses a great obstacle to oxidation, since the enzyme severely selects unproductive poses. For this reason, the reactive-like pose was selected as initial template for the design, aiming at improving the enzyme-substrate interaction energy of such binding mode, hence its statistical importance in the active site.

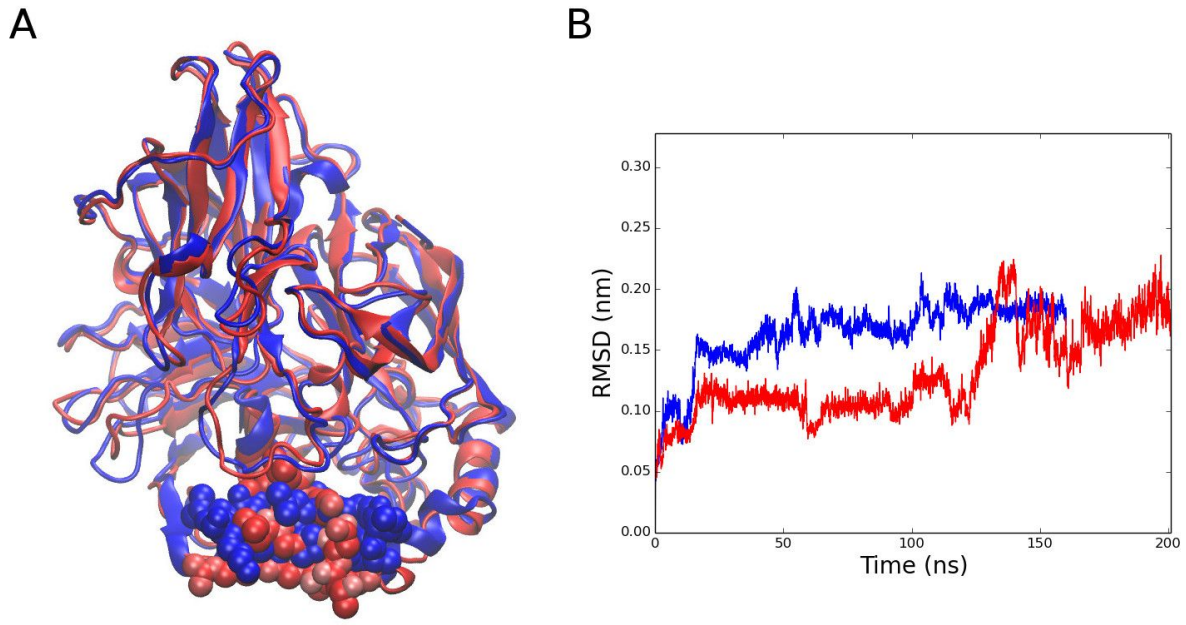


Figure 6.2. A) Final snapshot from the MD simulation of model_out (blue) and model_in (red). The C-terminal tail is represented in VDW spheres. B) RMSD plots from the MD simulations of model_out (blue) and model_in (red). Figure prepared with VMD (Humphrey et al., 1996), matplotlib (http://matplotlib.org/) and modified with GIMP (https://www.gimp.org/).

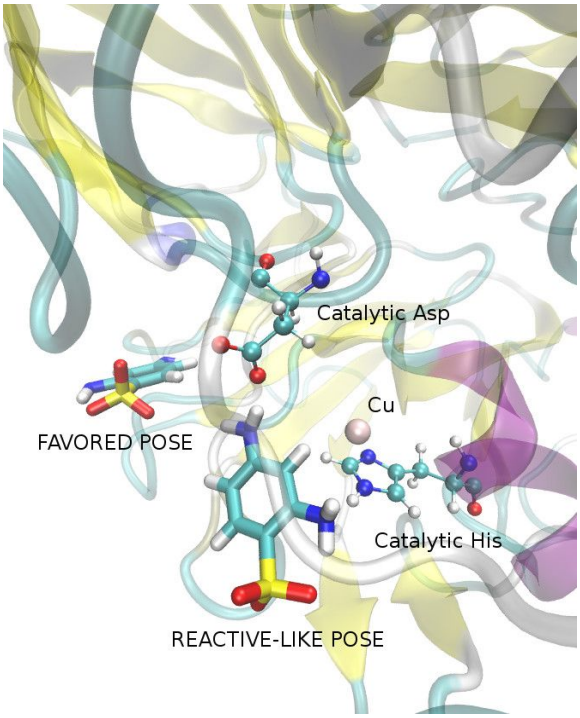


Figure 6.3. Favored (best binding) and reactive-like (catalytic contacts) poses from PELE sampling. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

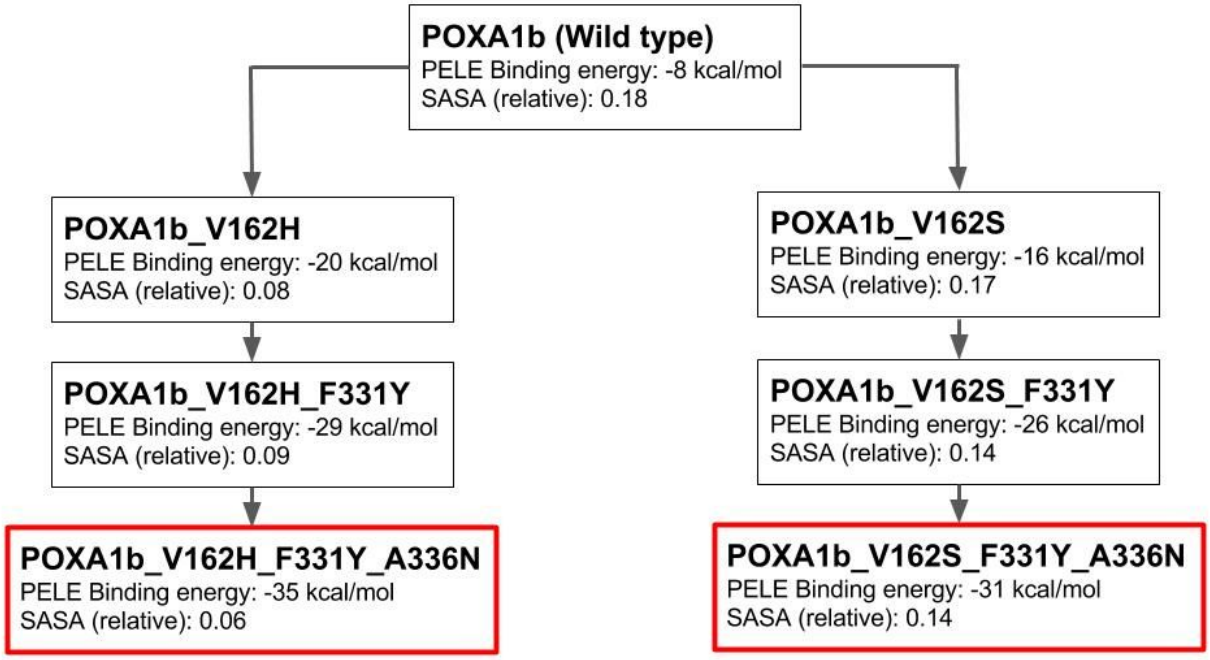


Figure 6.4. Evolutionary strategy and results of POXA1b design toward 2,4-dabsa oxidation. The first round produced two branches, introducing S or H at position 162. Figure prepared with Google Slides (https://www.google.com/slides/about/).

Since it was intended to test only a few POXA1b variants in the lab, an evolutionary strategy was employed to accumulate beneficial mutations at each computational mutagenesis round (Figure 6.4). During an evolution round a mutation was considered beneficial if the binding energy improved by 5 kcal/mol or more; then, one candidate was selected and used as starting template for the next round. It should be noted that, at each of these rounds, the mutants showing improved PELE binding energy were all visually inspected, bringing chemical intuition into play. In the first round, two mutations among the others looked promising due to a great increase in PELE binding energy and, in one case, a significant drop in SASA (hence a possible decrease in reorganization energy): V162H and V162S (Figure 6.4 and Figure 6.5A and B, respectively). Although other mutations present interesting values (Appendix 6), they were discarded (or postponed) in favor of V162H and V162S for one of the following reasons: i) they involved Gly/Pro (as starting amino acid), which might trigger important loop conformational changes due to its flexibility/rigidity which cannot be captured at this level of sampling (they would need extensive sampling to confirm their reliability); ii) unless they introduce an outstanding improvement, positive charges are avoided as they might inhibit oxidation; iii) they were located on the C-terminal tail, whose conformation is the most fragile aspect of the model; iv) visual inspection and "chemical intuition" suggested caution.

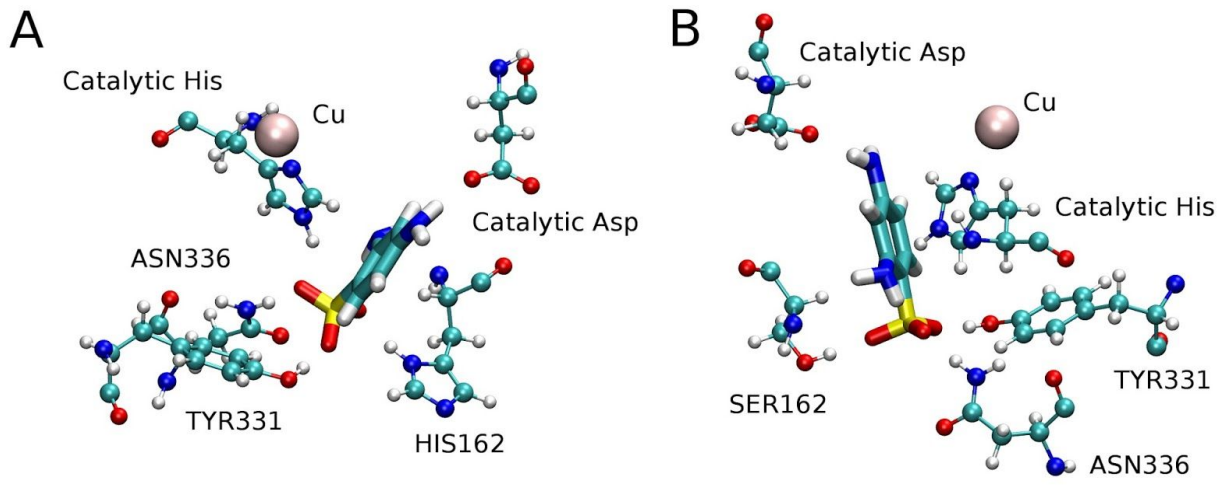


Figure 6.5. Structure of the designed triple mutants (A: V162H branch; B: V162S branch). Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

After the first round, computational evolution was branched, further designing both the V162H and V162S variants (Figure 6.4). In the second round of both the V162H and V162S branches, F331Y was chosen (Figure 6.4). In round3_H162 and round3_S162, A335N was selected (Figure 6.4). Computational evolution was stopped at this point. In both branches, three new H-bonds with the sulfonate group of 2,4-dabsa were introduced (Figure 6.5), increasing the binding energy by more than 20 kcal/mol (SASA is also significantly reduced in the H162 branch). Therefore, the evolved active sites should be able to select the substrate in a reactive conformation with higher frequency. In order to test this hypothesis, a new PELE sampling run was carried out for the evolved triple mutants, as previously done for the POXA1b (see Computational details); the catalytic and designed contacts were then checked in the structures within 5 kcal/mol of the lowest binding energy value. Nicely, for the S162 branch, all the designed H-bonds and the catalytic His-sulfonate contact were retained in all the binding modes, while the catalytic Asp-amino contact was formed in 40% of the cases.

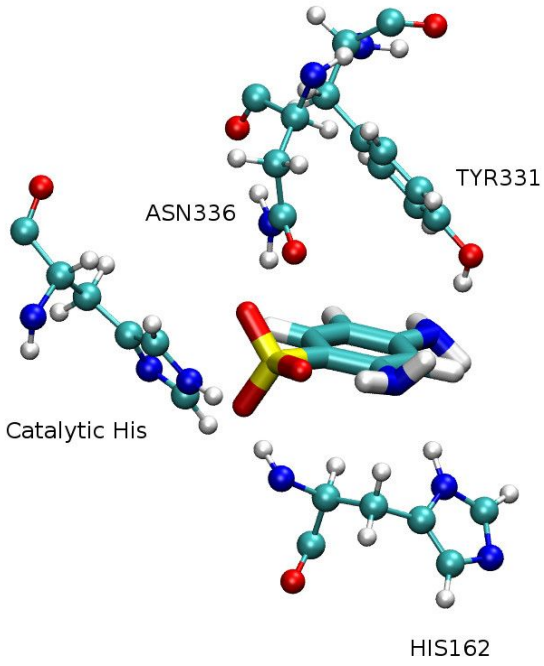


Figure 6.6. PELE sampling result for the H162 branch triple mutant. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

On the contrary, this last contact is never formed in the H162 branch. On top of that, although the catalytic His and N336 interact with the sulfonic group, the other designed contacts

did not form as planned: the side chains of Y331 and H162 interact with the amino groups instead (Figure 6.6). Still, the resulting binding mode is more reactive than POXA1b's (Figure 6.3): an efficient through bond (sulfonate-His) electron tunnelling path is stable, while the aminic proton could be abstracted by a water molecule (if proton transfer occurs simultaneously), although it is not the optimal path. Therefore, it was decided to test also the His branch triple mutant in the lab.

6.2.3 Experimental validation, computational rationalization and repurposing

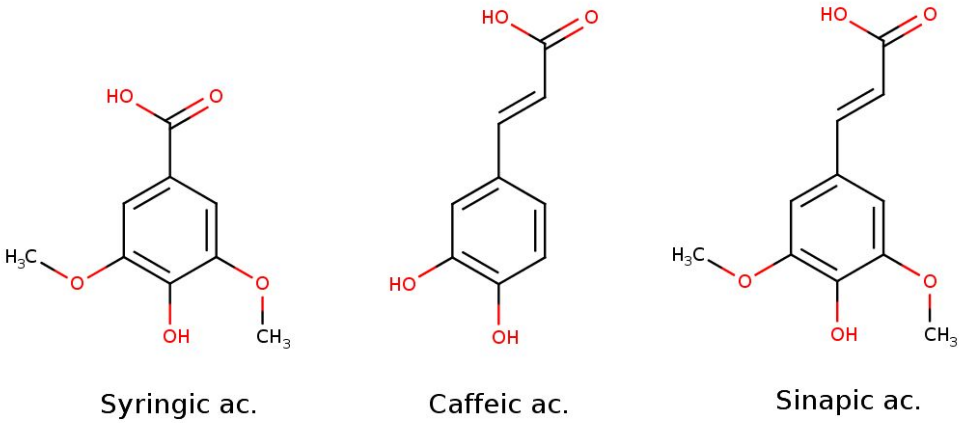
Regardless the promising mutant validation with PELE sampling, the evolved variants were inactive towards 2,4-dabsa, just as POXA1b. The possible motivations behind this unsatisfactory result are several: i) the homology model is inaccurate in the active site description due to the presence of the C-terminal; ii) no explicit water molecules were simulated, which have proved to be valuable in Chapter 5 ; iii) 2,4-dabsa is not oxidizable by the T1 copper (i.e. the driving force is positive and too large). The first point is very difficult to verify and the second strictly depends on the first: although MD simulations could have suggested the possible persistent placement of water molecules in the active site (especially H-bonded to the catalytic Asp), the reliability of such analysis strongly depends on the quality of the homology model. On the other hand, the third point can be addressed by comparing the energy of the HOMO orbitals of 2,4- and 2,5-dabsa, since the latter is readily oxidized by POXA1b. In fact, according to Koopmans' theorem, the negative of the HOMO energy, calculated with the restricted Hartree Fock (RHF) method, of a molecule corresponds to its ionization energy (Koopmans, 1934), assuming no subsequent orbital relaxation and neglecting electronic correlation.

Table 6.1. Computational ionization energy of 2,4- and 2,5- dabsa.

Chemical species	Ionization energy (eV)
2,4-dabsa	-4.44008397
2,5-dabsa	-4.26511468

The geometry optimization of both isomers at the RHF/6-31G* level revealed that the ionization energy of 2,4-dabsa is larger by ~175 meV, suggesting that their different reactivity (null for 2,4-dabsa) could be just a matter of ionization potential. Therefore, the driving force should have been targeted with a procedure analogous to the one presented at the beginning of this chapter

(possibly followed by QM/MM spin density rescoring). These triple mutants, which, according to PELE sampling, anchor the substrate similarly as designed, could be the starting point for future efforts involving such design methodology.



Scheme 6.2. Chemical structures of syringic, caffeic and sinapic acid. Prepared with MarvinSketch (http://www.chemaxon.com).

Table 6.2. Experimental kinetic constants and population of low binding energy structures exhibiting catalytic carboxyl-catalytic His during PELE sampling for the oxidation of caffeic acid by POXA1b and the designed triple mutants.

	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)	Sub-His H-bonds (%)
POXA1b	3.5±0.5	1262±60	360±60	0
H162 b.	0.40±0.05	1122±54	2805±54	50
S162 b.	3.6±0.5	4184±35	1162±35	29

Although the designed variants failed to oxidize the target substrate, their putative capability to lock in place 2,4-dabsa's sulfonic group could be exploited to improve the oxidation of other negatively charged substrates. In order to test this possibility, the oxidation of caffeic, syringic and sinapic acid (Scheme 6.2) was studied in the lab. As a result, the efficiency constant (k_{cat}/K_M) of the triple mutants is always higher than the one of POXA1b (Table 6.2-6.3). Interestingly, sinapic acid substrate inhibition is also attenuated in the evolved variants, as reflected by the larger inhibition constant (K_i) values, while an overall tendency of increasing K_M and k_{cat} can be observed (Table 6.4). Since the kinetic data for this substrate are obtained from

a competitive inhibition model (the substrate binds unproductively to the enzyme while hindering the formation of productive enzyme-substrate adducts (Bugg, 2012)), the trend of the kinetic constants implies that either the unproductive poses are destabilized or the productive conformations are stabilized in the mutant. The second hypothesis could be true despite the larger K_M values, since this constant would increase with k_{cat} in a steady-state kinetic regime.

Table 6.3. Experimental kinetic constants and population of low binding energy structures exhibiting catalytic carboxyl-catalytic His during PELE sampling for the oxidation of syringic acid by POXA1b and the designed triple mutants.

	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)	Sub-His H-bonds (%)
POXA1b	1.2±0.1	1775±87	1400±87	0
H162 b.	1.3±0.2	2700±54	2000±54	60
S162 b.	0.8±0.2	6997±87	8000±87	50

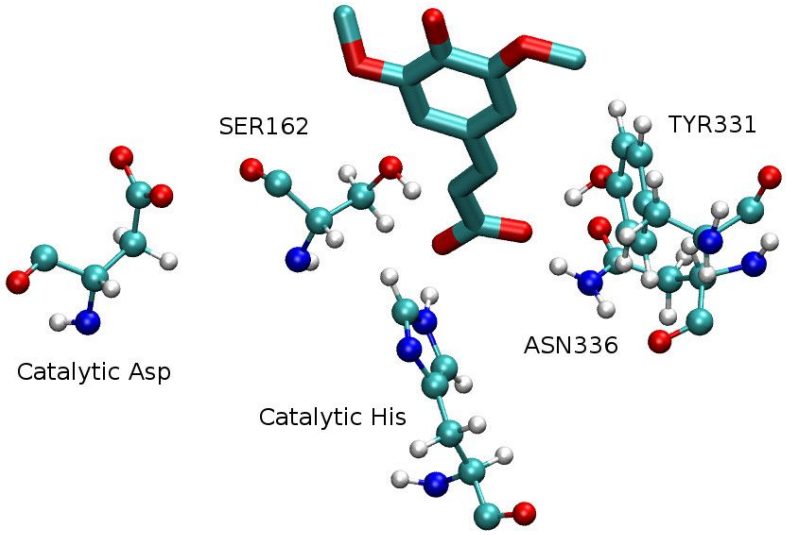


Figure 6.7. PELE sampling result for the binding of sinapic acid in the active site of the S162 branch triple mutant. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

Binding of these three substrates in the T1 pocket of POXA1b and its evolved variants was simulated with PELE (48 independent 48 hours lasting simulations). The objective was to verify whether such improved catalytic performance is related to larger population of conformations in which the carboxyl group (which is deprotonated at pH 8) is H-bonded to the

catalytic His. As a matter of fact, POXA1b never shows H-bonds between the catalytic His and the substrates, forcing an inefficient through-space electron tunnelling between the substrate and the T1 center. On the other hand, H-bonds between the substrate and the catalytic His have a significant population for every mutant-substrate pair, which ranges from 29 to 81% (Tables 6.2-6.3) among the structures within 5 kcal/mol of the binding energy minima and within 10 Å of the T1 copper. This contact ensures a through-bond electron tunnelling, which greatly facilitates ET. Remarkably, the sinapic acid-S162 branch mutant pair, the one that experiences the largest improvement in k_{cat} (~15 fold, Table 6.4), exhibits the highest population of this contact (81%) and forms all the designed H-bonds in 61% of the cases (Figure 6.7). Finally, it is worth noting that no contact is formed between the substrate and the catalytic Asp in any variant-substrate pair, meaning that such contact is not indispensable (water molecules can assist deprotonation), although design toward the instauration of such interaction could further improve oxidation (especially due to the further vicinity of a negative charge that could shift up the HOMO energy of the substrates). The reason behind the success of repurposing are likely the surface exposure and flexibility (since the substrate mostly interacts with loops) of the T1 pocket.

Table 6.4. Experimental kinetic constants and population of low binding energy structures exhibiting catalytic substrate-His H-bonds during PELE sampling for sinapic acid by POXA1b and the designed triple mutants.

	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)	K_I (mM)	% Sub-His H-bonds
POXA1b	0.13±0.05	6681±100	51392±100	0.7±0.1	0
H162 b.	0.41±0.03	39113±500	95397±500	1.4±0.3	66
S162 b.	0.67±0.02	97126±600	144964±600	1.2±0.2	81

6.2.4 Consequences for laccases design

Although the initial design was not successful, the repurposing effort suggests (both in its experimental and computational validation) that the non specific reaction site of laccases can be a valuable template to build specificity toward a target substrate. This can be implemented with the presented computational methodology, increasing the affinity of the substrate for the enzyme in the reactive conformation. Nonetheless, laccases design proved to be a delicate multidimensional problem which requires to always take the ET driving force into account. This is especially true for non reactive substrates. With this in mind, the oxidability of the target

substrate should be always assessed beforehand with ionization potential measurement or calculation (which could be qualitatively guessed by comparison with a reactive substrate, as done here). If the target substrate were poorly oxidizable, the ET driving force would become a crucial design parameter.

Repurposing should not be a mere safety net (as it was in this case) but rather an effective strategy for industrial production. Computational design toward a given substrate could be accompanied on-the-fly by the *in silico* evaluation of a database of putative substrates, possibly selecting a subset based on target molecular similarity (Maggiora & Veerabahu, 2010). In this way, several processes could be improved at the cost of one or a new design line for a particular promising substrate could be opened. Ideally, variant search should not be fossilized on a fixed laccase neither. These enzymes present great diversity in the T1 pocket that could be exploited to design the best host as possible for a (class of) substrate(s). Even if a very promising laccase did not function at the desired working conditions (pH, temperature, presence of ions and/or cosolutes, etc.), the enzyme could be designed to adapt to them. As a matter of fact, we were able to significantly change the reaction conditions of an oxidoreductases for Novozymes, the world leader company in enzyme production, with the same approach presented here (results cannot be revealed before patenting). Therefore, a future where several laccases are computationally evolved against many substrates can be envisioned. The number of laccases tested will depend on the resources and know-how of each laboratory. Pushing further, efficient chimeric laccases might be designed by conveniently stitching portions of protein from different laccases with computational techniques (Jacobs et al., 2016).

6.2.5 Future development

Although the results shown in this and in the previous chapter represent a clear opening for future computer-aided laccases design, a lot of work remains to be done. In particular, it will be crucial to benchmark computational methods over a suitable set of experimental kinetic data, which will imply a considerable synergistic computational and experimental effort (Carlin et al., 2016). In such way, the discovery of structural and energetic features that correlate with kinetic constants will be possible (Carlin et al., 2016).

6.3 Closure

The T1 site is a promiscuous template to build specificity toward a target substrate, through the design of stabilizing interactions enzyme-substrate interactions in a highly reactive conformation. The experimental and computational validation of mutant repurposing suggests that this task could be achieved with the computational protocol proposed in this work. Nonetheless, laccases design is an intricate multidimensional problem in which the ET driving force is a crucial factor that should never be overlooked: if the ionization potential of the target substrate is too low to give one electron to the T1 copper, building a perfect electron tunnelling path will not work.

Likely due to the superficial and shallow nature of the T1 site, mutant repurposing proved to be effective. In the future, repurposing might be carried out computationally and subsequently validated experimentally, transforming laccases into more efficient multitasking enzymes.

In any case, the collection of experimental kinetic data to benchmark computational strategies, aiming at finding parameters that correlate with the kinetic constant, will be crucial to improve the predictive power over laccase.

REFERENCES:

Blumberger J. 2015. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical reviews* 115:11191–11238.

Bortolotti CA., Siwko ME., Elena C., Antonio R., Marco S., Stefano C. 2011. The Reorganization Energy in Cytochrome c is Controlled by the Accessibility of the Heme to the Solvent. *The journal of physical chemistry letters* 2:1761–1765.

Bugg TDH. 2012. *Introduction to Enzyme and Coenzyme Chemistry*.

Bussi G., Donadio D., Parrinello M. 2007. Canonical sampling through velocity rescaling. *The Journal of chemical physics* 126:014101.

Carlin DA., Caster RW., Wang X., Betzenderfer SA., Chen CX., Duong VM., Ryklansky CV., Alpekin A., Beaumont N., Kapoor H., Kim N., Mohabbot H., Pang B., Teel R., Whithaus L., Tagkopoulos I., Siegel JB. 2016. Kinetic Characterization of 100 Glycoside Hydrolase Mutants Enables the Discovery of Structural Features Correlated with Kinetic Constants. *PloS one* 11:e0147596.

Duan Y., Wu C., Chowdhury S., Lee MC., Xiong G., Zhang W., Yang R., Cieplak P., Luo R., Lee T., Caldwell J., Wang J., Kollman P. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry* 24:1999–2012.

Essmann U., Ulrich E., Lalith P., Berkowitz ML., Tom D., Hsing L., Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics* 103:8577.

Festa G., Autore F., Fraternali F., Giardina P., Sannia G. 2008. Development of new laccases by directed evolution: functional and computational analyses. *Proteins* 72:25–34.

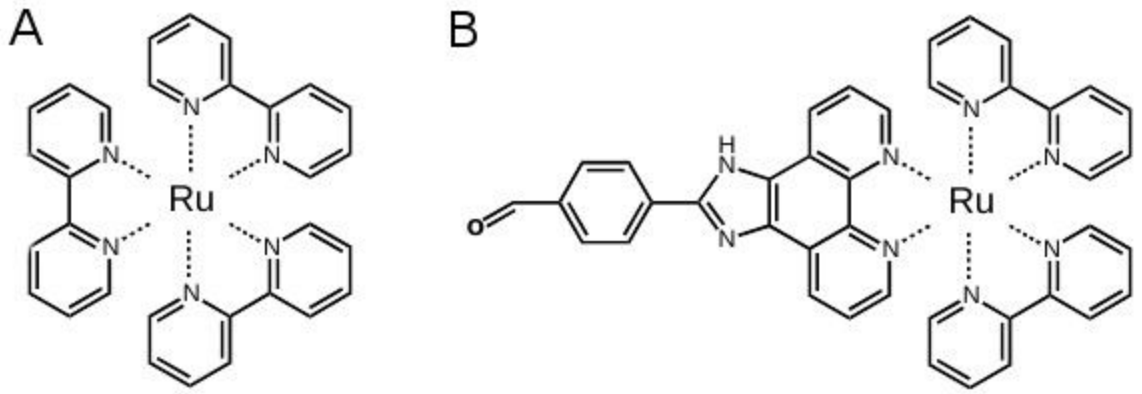
Giardina P., Palmieri G., Scaloni A., Fontanella B., Faraco V., Cennamo G., Sannia G. 1999. Protein and gene structure of a blue laccase from Pleurotus ostreatus1. *Biochemical Journal* 341:655.

Gordon JC., Myers JB., Folta T., Shoja V., Heath LS., Onufriev A. 2005. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* 33:W368–71.
Hess B., Berk H., Henk B., Berendsen HJC., Johannes G E. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry* 18:1463–1472.
Humphrey W., William H., Andrew D., Klaus S. 1996. VMD: Visual molecular dynamics. *Journal of molecular graphics* 14:33–38.
Jacobs TM., Williams B., Williams T., Xu X., Eletsky A., Federizon JF., Szyperski T., Kuhlman B. 2016. Design of structurally distinct proteins using strategies inspired by evolution. *Science* 352:687–690.
Jacobson MP., Friesner RA., Zhexin X., Barry H. 2002. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of molecular biology* 320:597–608.
Jacobson MP., Pincus DL., Rapp CS., Day TJF., Honig B., Shaw DE., Friesner RA. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins* 55:351–367.
Koopmans T. 1934. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* 1:104–113.
Maggiora GM., Veerabahu S. 2010. Molecular Similarity Measures. In: *Methods in Molecular Biology*. 39–100.
Mohamadi F., Fariborz M., Richards NGJ., Guida WC., Rob L., Mark L., Craig C., George C., Thomas H., Clark Still W. 1990. Macromodel?an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of computational chemistry* 11:440–467.
Olsson MHM., Søndergaard CR., Rostkowski M., Jensen JH. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation* 7:525–537.
Parrinello M. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of applied physics* 52:7182.
Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., Shirts MR., Smith JC., Kasson PM., van der Spoel D., Hess B., Lindahl E. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854.
Raval A., Alpan R., Stefano P., Eastwood MP., Dror RO., Shaw DE. 2012. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*. DOI: 10.1002/prot.24098.
Sastry GM., Adzhigirey M., Day T., Annabhimoju R., Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* 27:221–234.
Toukan K., Kahled T., Aneesur R. 1985. Molecular-dynamics study of atomic motions in water. *Physical Review B: Condensed Matter and Materials Physics* 31:2643–2648.

7. Design of a hybrid laccase/photosensitizer system

As discussed in Chapter 4, the standard redox potential (Eº) of the T1 copper (E^{o}_{T1}) is the main limit to laccases applications. Its increase remains a difficult challenge which still does not guarantee any improvement in activity. The visible-light-driven reduction of laccases through a photosensitizer renders oxidation of inert substrates possible by virtually increasing E^{o}_{T1}. The same strategy could also be used to efficiently reduce oxygen in a biofuel cell, connecting the photosensitizer to the cathode.

In this chapter, the electron transfer (ET) between tris(bipyridine)ruthenium(II) ($[Ru(bpy)_3]^{2+}$, Scheme 7.1A) and LAC3, a fungal laccase with E^{o}_{T1} = 0.68 V (Simaan et al., 2011), is modelled. The final purpose is to redesign the LAC3/$[Ru(bpy)_3]^{2+}$ system to improve ET directionality from the substrate to the trinuclear cluster of copper ions (TNC). In section 7.1, the experimental data available are presented and discussed. Then, the computational methodology is explained in detail in section 7.2 and applied to understand and redesign the LAC3/$[Ru(bpy)_3]^{2+}$ system in the subsequent section.



Scheme 7.1. Two-dimensional structure of A) $[Ru(bpy)_3]^{2+}$ and B) chemically modified $[Ru(bpy)_3]^{2+}$ that covalently binds exposed K residues. Source: provided by collaborators at Aix Marseille Université.

7.1 Experimental background

The photoexcitation of $[Ru(bpy)_3]^{2+}$ produces a strong reductant, $[Ru(bpy)_3]^{2+*}$ (Table 7.1 and Figure 7.1). This complex, in a 1:1 LAC3/$[Ru(bpy)_3]^{2+}$ bimolecular solution, donates one electron to the T1 copper (assumed to be the first electron acceptor, as in nature) with a measured ET rate constant (k_{ET}) of 8.1×10^8 $M^{-1}s^{-1}$ (Simaan et al., 2011), forming $[Ru(bpy)_3]^{3+}$.

This is a strong oxidant (Eº = 1.29 V, Table 7.1) that allows to oxidize more challenging compounds (Figure 7.1). On the other hand, due to its high E°, the back-transfer (BT) rate (k_{BT}) is a limiting factor in the catalytic efficiency of the hybrid $[Ru(bpy)_3]^{2+}$/laccase system. Indeed, using Marcus' equation (Eq. 2.49) it turns out that k_{ET}/k_{BT}~10^3. This ratio is estimated taking into account the E^o_{T1} of LAC3 and assuming: i) negligible differences in reorganization energy (λ) and electron coupling (T_{DA}) between ET and BT; ii) that the ET driving force depends only on Eº(acceptor) - Eº(donor) and iii) λ~1.7-2.0 eV (taken from other substrate/laccase pairs (Tadesse et al., 2008; Galli et al., 2013; Pardo et al., 2016)).

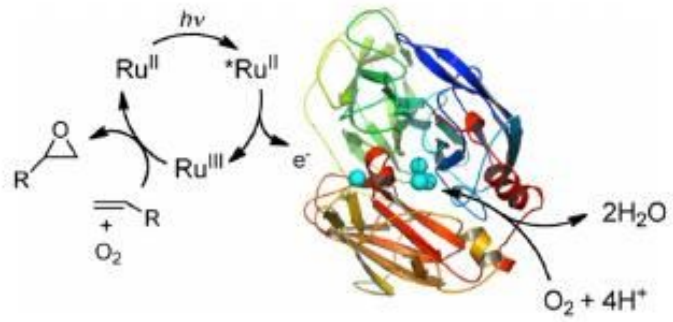


Figure 7.1. Photocatalytic oxidation of olefins by the laccase/$[Ru(bpy)_3]^{2+}$ system. Source: (Schneider et al., 2015).

Table 7.1. Redox potentials relevant to this chapter (Juris et al., 1981, Simaan et al., 2011).

Redox couple	Eº (V)
$Cu^{2+}_{(T1, LAC3)} + e^- \rightarrow Cu^{+}_{(T1, LAC3)}$	0.68
$[Ru(bpy)_3]^{3+} + e^- \rightarrow [Ru(bpy)_3]^{2+}$	1.29
$[Ru(bpy)_3]^{3+} + e^- \rightarrow [Ru(bpy)_3]^{2+*}$	-0.81

Farver and co-workers experimentally determined a value of 25 s^{-1} for the rate constant (k_{IET}) associated to the internal ET (IET) from the T1 copper to the TNC in a *Trametes hirsuta* laccase (ThL, 70.4% identity with LAC3) (Farver et al., 2011)). This value is expected to increase by only a factor of 5 when oxygen binds to the TNC (Farver, Wherland & Pecht, 1994), keeping k_{IET} ~10^2 s^{-1}. Consequently, the large k_{BT} pulls the electron back to $[Ru(bpy)_3]^{2+}$ before it can reach the TNC. To circumvent this limitation, ET can be tuned by changing how $[Ru(bpy)_3]^{2+}$ and LAC3 interact to decrease T_{DA}, therefore both k_{ET} and k_{BT}, until k_{IET} and k_{BT} are close enough to enhance $[Ru(bpy)_3]^{2+}$ to TNC ET directionality.

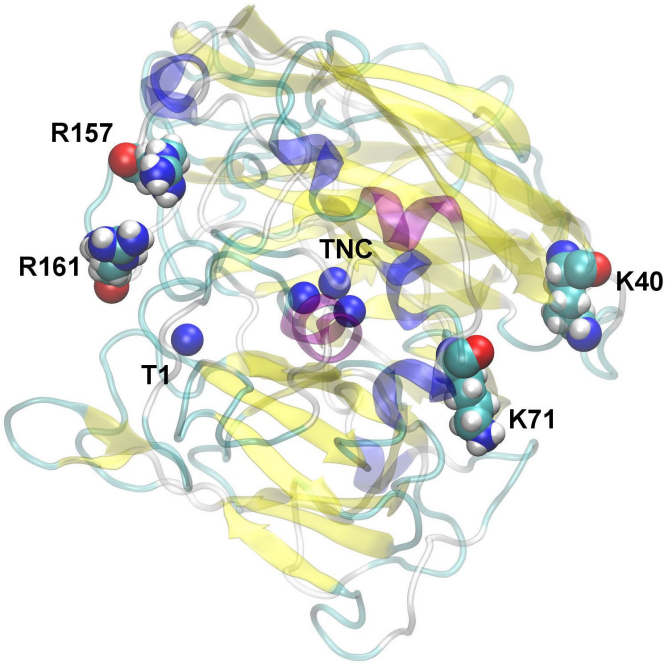


Figure 7.2. Tridimensional structure of LAC3 with the copper centers (T1 and TNC), native Lys (K40 and K71) and two Arg (R157 and R161) are highlighted. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

One way to tune ET is to selectively attach $[Ru(bpy)_3]^{2+}$ to the enzyme surface, controlling its distance from the copper centers by forming unimolecular species, e.g. binding it (chemically modified, Scheme 7.1B) to a Lys with reductive amination (McFarland & Francis, 2005). Having only two native Lys at positions 40 and 71, LAC3 is the ideal laccase for this task (Figure 7.2). With this purpose, an experimental team at Aix Marseille Universitè attached $[Ru(bpy)_3]^{2+}$ to three LAC3 variants: the wild type (containing K40 and K71) and two mutants displaying only one Lys at position 157 (UNIK157) and 161 (UNIK161) respectively. While the native Lys positions are next to the TNC, at the entrance of the water channel, positions 157 and 161 are closer to the T1 pocket (Figure 7.2). The bimolecular (LAC3+Ru) and unimolecular (LAC3-, UNIK157- and UNIK161-Ru) laccase/$[Ru(bpy)_3]^{2+}$ adducts were photo-reduced in two conditions: i) in absence of oxygen, using ethylenediaminetetraacetic acid (EDTA) as electron donor; ii) with p-styrene sulfonate as substrate, in an open to air solution. While oxygen allows to reoxidize the enzyme at each catalytic cycle, its exclusion was meant to assess how quickly the T1 copper was photo-reduced at different $[Ru(bpy)_3]^{2+}$ grafting positions (LAC3-RU and UNIK-Ru variants) or in a non covalent (LAC3+Ru) assembly.

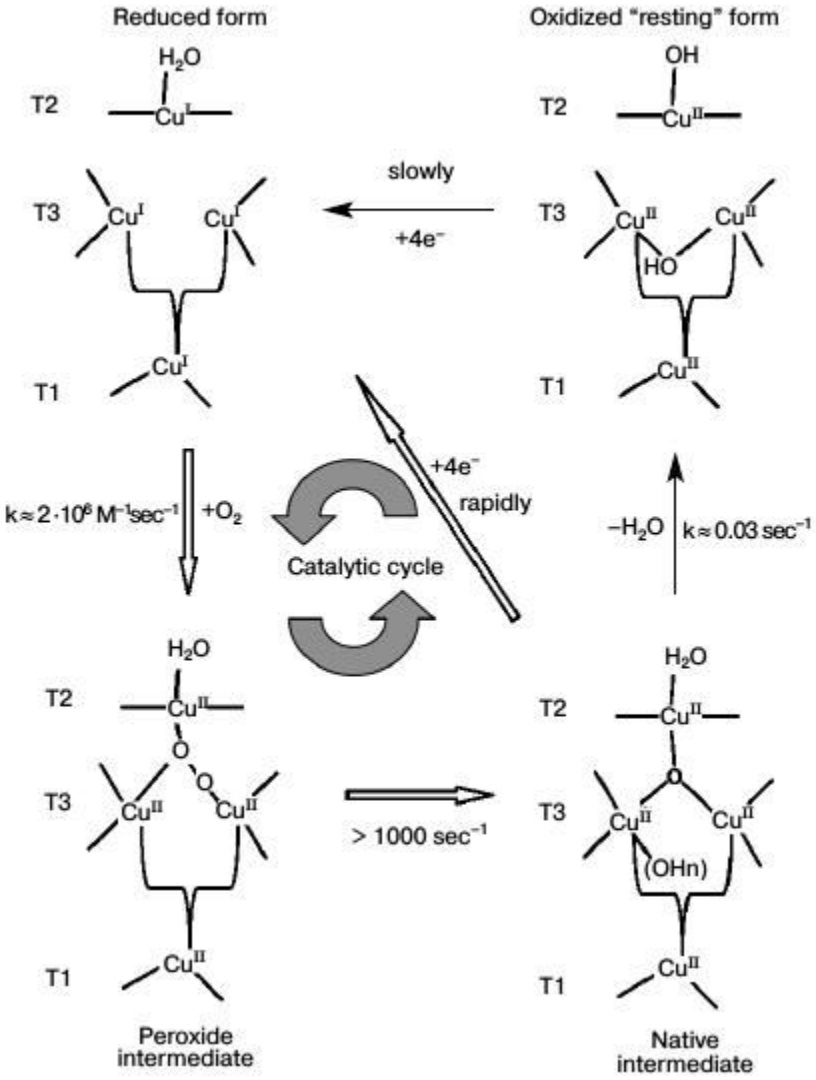


Figure 7.3. Catalytic cycle of laccases. Source: (Morozova et al., 2007).

The results are summarized in Table 7.2. The incorporation of $[Ru(bpy)_3]^{2+}$ enhances laccases reduction and renders the oxidation of p-styrene sulfonate possible. The T1 copper is completely reduced in LAC3+Ru, while it is only partially reduced in the other systems. Despite the reduction of the T1 copper is significantly slower in the UNIK-Ru variants, UNIK161-Ru oxidizes the substrate as much as LAC3+Ru, possibly due to the lower BT rate (assuming for both system that the T1 copper is the first electron acceptor). LAC3-Ru exhibits the lowest T1 copper reduction and it can oxidize p-sulfonate-styrene roughly five times slower than the other systems. In this system, the TNC is most likely the first electron acceptor, due to the positions of K40 and K71 (Figure 7.2). In such case, in presence of oxygen, the electrons would not be

transferred to the T1 copper but rather withdrawn by oxygen during the formation of the peroxide intermediate (Figure 7.3). Indeed, the rate constant (k_{O-O}) of this last process is large, ~10^6 $M^{-1}s^{-1}$ (Cole, Ballou & Solomon, 1991; Palmer et al., 2002). Direct ET from different electrodes to the TNC has been measured in several experiments (Shleev et al., 2005, 2006; Pita et al., 2006; Ramírez et al., 2008). Therefore, this alternative ET pathway is grounded on previous experimental evidences.

Table 7.2. Estimated percentage of reduced T1 copper and product formation after 60 minutes, in deoxygenated and oxygenated conditions, respectively.

System	% T1 reduction	% product formation
LAC3	10	0
LAC3+Ru	100	3.4
LAC3-Ru	30	0.7
UNIK161-Ru	50	3.9
UNIK157-Ru	60	Not performed

7.2 Computational details

Homology modeling and system preparation. A three-dimensional model of LAC3 was built with homology modeling and the loops refined using Prime (Jacobson et al., 2002, 2004), chosing 3kw7.pdb as a template (75.4% identity, 84.1% similarity, 2.9% gaps). The final structure was then prepared with the protein preparation wizard software (Sastry et al., 2013). The protonation state of titratable residues were assigned at pH 4 with PROPKA (Olsson et al., 2011) and double-checked with the H++ server (Gordon et al., 2005).

Docking. In order to simulate the bimolecular system, the surface cavities of LAC3 were detected with the FPocket server (Schmidtke et al., 2010) and $[Ru(bpy)_3]^{2+}$ docked in both its enantiomeric forms with AutoDock Vina (Trott & Olson, 2010). The search space was a cubic grid with a 30 Å side, centered in the geometrical center of each cavity previously detected. Default values were chosen for all search parameters and one decoy was printed for each cavity. No significant changes were found between the binding modes of the two enantiomers. The output structures were refined with PELE (see Appendix 3 for an example of input file), using small ligand and backbone ANM perturbations. The side chains of the residues within 5 Å of substrate were repacked and all the residues within 10 Å were subjected to energy

minimization. For each pose, seven independent simulations were run for 10 hours. In an attempt to simulate the unimolecular system, covalent docking was performed with Glide (Zhu et al., 2014) to K40, 71, 157 and 161, producing four UNIK-Ru variants (40, 71, 157 and 161). Although K40 and 71 are both bonded to $[Ru(bpy)_3]^{2+}$ in LAC3-Ru, they were treated separately as UNIK40- and UNIK71-Ru for the sake of simplicity. The search space for covalent docking was defined as a cubic grid with a side of 30 Å centered on the grafting residue.

Treatment of $[Ru(bpy)_3]^{2+}$ and metal centers in PELE. $[Ru(bpy)_3]^{2+}$ geometries were optimized at the M06/lacvp* level of theory, modeling the solvent with the Poisson-Boltzmann model. The atomic ESP charges were calculated from the optimized structure and used for PELE. The OPLS parameters were assigned with Schrödinger's hetgrp_ffgen utility. The same procedure was implemented to create a $[Ru(bpy)_3]^{2+}$-containing Lys-based amino acid and model the UNIK-Ru variants. A rotamer library for this new amino acid was generated with in-house PELE's scripts using Macromodel (Mohamadi et al., 1990). The copper ions and their coordination atoms plus their nearest neighbors were constrained during PELE sampling with 200 kcal mol^{-1} $Å^{-2}$ harmonic constraints on their atomic positions, placing a +1 charge on each copper atom (to avoid tension in the TNC).

Molecular dynamics (MD). The laccase/$[Ru(bpy)_3]^{2+}$ variants subjected to MD or metadynamics were solvated with a 10 Å buffer of waters in an orthorhombic box, neutralized and 0.15 M NaCl is added. After equilibration (default settings), a 50 ns long NpT production for both UNIK161- and UNIK157-Ru was run with Desmond (Bowers et al., 2006), using the OPLS-2005 force-field (Kaminski et al., 2001) and the SPC explicit water model (Toukan, Kahled & Aneesur, 1985). The temperature was kept constant at 300 K with the Nosé−Hoover chain thermostat (Nosé & Shuichi, 1984) with a relaxation time of 1.0 ps, and the pressure was controlled with the Martyna−Tobias−Klein barostat (Martyna, Tobias & Klein, 1994) with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator (Tuckerman, Berne & Martyna, 1992) was employed with bonded, near, and far time steps of 2.0, 2.0, and 6.0 fs, respectively. A 9 Å cutoff was used for non bonded interactions together with the smooth particle mesh Ewald method (Essmann et al., 1995).

Metadynamics. In order to assess if $[Ru(bpy)_3]^{2+}$ tends to stay on the surface or hide into the protein interior (as covalent docking suggested), LAC3-Ru was simulated with metadynamics (Laio & Parrinello, 2002). Despite K40 and K71 are simultaneously bonded to $[Ru(bpy)_3]^{2+}$ in LAC3-Ru, the system was modeled separately as UNIK40- and UNIK71-Ru to keep the number of collective variables (CVs) tractable. Two distances were chosen as CVs: (i)

between Ru and T2 copper and (ii) between CA and NZ atoms belonging to the lysine. Metadynamics was also employed to simulate the bimolecular laccase/$[Ru(bpy)_3]^{2+}$ systems. In that case, the distance between Ru and T2 copper was the only CV. The free energy surface (FES) was constructed depositing a gaussian every 0.09 ps for 20 ns. The width of each gaussian along a CV was 0.05 Å and the height 0.03 kcal/mol.

Treatment of $[Ru(bpy)_3]^{2+}$ and copper centers in MD and metadynamics. The parameters for $[Ru(bpy)_3]^{2+}$ were generated with the hetgrp_ffgen utility of Schrödinger, inputting charges and geometries from QM calculations. Since the distance between $[Ru(bpy)_3]^{2+}$ and the copper centers was always abundantly beyond the 9 Å cutoff for non bonded interactions ($[Ru(bpy)_3]^{2+}$ cannot get closer for steric hindrance), the copper centers were parametrized classically. Unitary charges were assigned to each copper atom.

PELE design of new UNIK-Ru variants. The incorporation of the $[Ru(bpy)_3]^{2+}$-containing Lys-based amino acid took two steps. First, the target native residue was mutated to Lys. Then, if the introduced Lys was sufficiently exposed to the solvent (i.e. at least the nitrogen exposed and pointing toward the solvent), it was mutated to the $[Ru(bpy)_3]^{2+}$-containing amino acid. The protein backbone was subjected to ANM displacements and all residues were included in energy minimization. The simulation of each mutant (and of UNIK40-, UNIK71-, UNIK161- and UNIK157-Ru for the sake of comparison) lasted 1 hour, accumulating ~30 Monte Carlo steps (see Appendix 4 for an example of input file). The final result was visually inspected before ET rate calculations averaged over all the generated structures.

ET rate calculations. T_{DA}, the pre-exponential factor of Marcus' equation, was estimated using the VMD Pathways plugin (Balabin, Hu & Beratan, 2012) over the last frame of each of the seven independent PELE simulations of the LAC3-Ru decoys. The same calculation was performed on UNIK40-, UNIK71-, UNIK-157 and UNIK161-Ru, using ten snapshots from metadynamics and MD, and on the PELE design results (taking the last ten snapshots). In the Pathways model, T_{DA} is quantified following continuous paths between donor (the Ru atom) and acceptor (each Cu atom), where the electron tunnels through space, hydrogen bond or covalent bond. The number and nature of tunneling events determines T_{DA} according to Eq. 7.1 (Beratan, Betts & Onuchic, 1991):

$$T_{DA} = A \prod \varepsilon^C \prod \varepsilon^H \prod \varepsilon^S \qquad \text{Eq. 7.1}$$

$$\varepsilon^C = 0.6 \qquad \text{Eq. 7.2}$$

$$\varepsilon^{H} = 0.36\ exp[-1.7(R-2.8)] \quad \text{Eq. 7.3}$$

$$\varepsilon^{S} = 0.6\ exp[-1.7(R-1.4)] \quad \text{Eq. 7.4}$$

Where $\varepsilon_C > \varepsilon_H > \varepsilon_S$ are the through covalent bond, hydrogen bond and space contributions, A is a constant and R is the distance in Å between donor and acceptor for a single tunnelling event.

It should be noted that only one T_{DA} value is considered for the TNC, the highest one. The reason is straightforward. In these calculations, LAC3 was modeled in the resting state, i.e. the TNC ions were not bridged by any oxygen atom (Figure 7.3). However, during the catalytic cycle, the TNC is fully reduced in the so called native state, where the copper ions are all bonded to one bridging oxygen atom (Figure 7.3). If this atom had been included in the model, the same tunnelling pathway used to reach the copper ion displaying the largest T_{DA} value could have been exploited by the other copper ions. For them, T_{DA} would have been decreased by a 0.36 factor with respect to the highest value, since tunnelling would have been extended to two more covalent bonds (Cu-O-Cu, see Figure 7.3 and Eq. 7.2). Conversely, the absence of oxygen makes the Cu-Cu ET inefficient, decreasing T_{DA} by roughly two orders of magnitude (as can be verified inputting R=4 to Eq. 7.4, since the Cu-Cu distances in TNC are all above 4 Å).

7.3 Results and discussion

7.3.1 Simulation of LAC3+Ru

Surface pocket analysis, rigid docking and PELE all-atom refinement produced several distinct binding modes over LAC3's surface. Among these, one resides in the T1 cavity (Figure 7.4B) and shows the highest T_{DA} with the T1 copper, ~10^{-4} eV (Table 7.3). Inputting this value in Marcus' equation (Eq. 2.49) gives k_{ET}~10^{8-9} s^{-1}, which is in good agreement with the experimental data (8.1x10^8 $M^{-1}s^{-1}$ (Simaan et al., 2011)). Therefore, as expected, ET from $[Ru(bpy)_3]^{2+}$ to the laccase occurs mostly through such conformation.

Among the other binding modes, a second one in proximity of the TNC (Figure 7.4A) catches the eye for its appealing features. According to the calculation, this alternative pose transfers electrons to the TNC with T_{DA}~10^{-6} eV (Table 7.3). Inserting T_{DA}~10^{-6} eV in Marcus' equation, setting λ ~1.8 eV, E^{o}_{T3} ~0.8 V (Farver et al., 2011) and E^{o}_{T2}~0.4 V (Shleev et al., 2006), yields k_{ET}~10^4 s^{-1}, which implies k_{ET}/k_{BT}~10 and 10^2 for T2 and T3 respectively. Since oxygen uptake by TNC and the formation of the so-called peroxide intermediate (k_{O-O}) are

orders of magnitude faster than BT (~10^6 $M^{-1}s^{-1}$ (Cole, Ballou & Solomon, 1991; Palmer et al., 2002) vs. ~10^{2-3} s^{-1}), ET directionality would be largely improved if the electrons were directly delivered from $[Ru(bpy)_3]^{2+}$ to TNC through the binding mode depicted in Figure 7.4A. However, such conformation is (most likely) not significantly populated in solution to affect ET, otherwise the experimental k_{ET} would be several orders of magnitude lower than 8.1x10^8 $M^{-1}s^{-1}$ (Simaan et al., 2011) and the reduction of the T1 copper would be significantly reduced (Table 7.2).

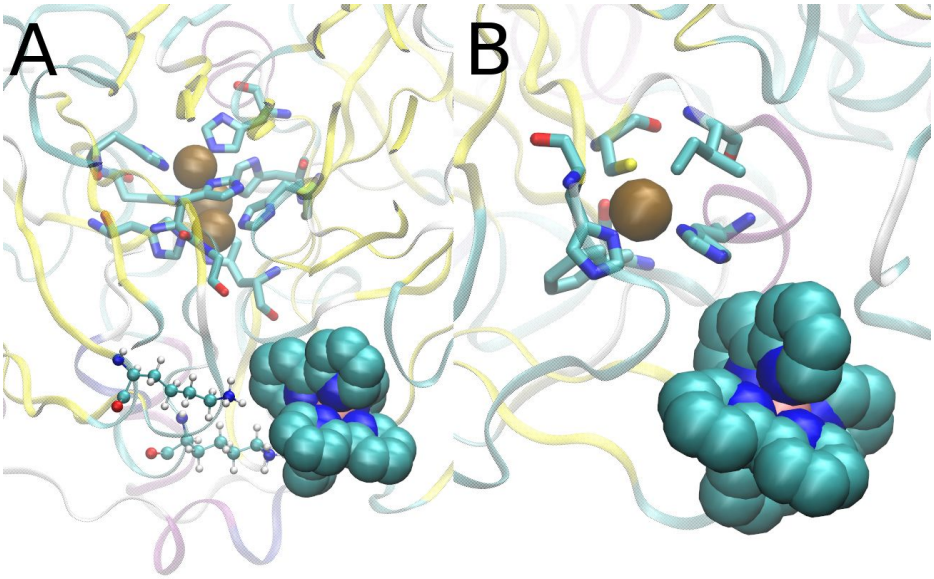


Figure 7.4. Closest binding modes to A) TNC and B) T1. K40 and K71 are shown in CPK style in A. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

Table 7.3. Results of T_{DA} calculation between Ru and all Cu atoms belonging to LAC3 for the T1 and TNC binding modes.

Binding mode	$T_{DA}^{(Ru,T1)}$ (eV)	$T_{DA}^{(Ru,TNC)}$ (eV)
T1 pocket	$3 \cdot 10^{-4}$	$8 \cdot 10^{-8}$
TNC cavity	$9 \cdot 10^{-9}$	$5 \cdot 10^{-6}$

A closer look to the structure reveals that two Lys, K40 and K71, are positioned next to $[Ru(bpy)_3]^{2+}$ when this is closer to the TNC (Figure 7.4A). This electrostatic environment might preclude the formation of such binding mode since $[Ru(bpy)_3]^{2+}$ carries two positive charges. To assess this possibility, the LAC3-Ru system was simulated, starting from the conformation of Figure 7.4A, in the wild type and in the LAC3_K40S-K71S variants with metadynamics (Bussi,

Laio & Parrinello, 2006). The choice of replacing Lys with Ser has no particular reason apart from preserving some polarity. The distance between Ru and T2 was chosen as the only CV. As expected, the presence of K40 and K71 pushes $[Ru(bpy)_3]^{2+}$ to visit other areas where the free energy is considerably lower (Figure 7.5). On the contrary, $[Ru(bpy)_3]^{2+}$ binding is more favorable when K40 and K71 are mutated, as reflected in the better FES with respect to the wild type (Figure 7.5). It should be noted that the simulated time, 20 ns, does not allow an accurate FES calculation as the exploration is (most likely) not fully converged. However, these simulations are intended to provide relative and semi-quantitative FES estimation rather than accurate absolute calculations. Moreover, when a second 20 ns run is performed the outcoming FES is consistent, at least on a qualitative ground.

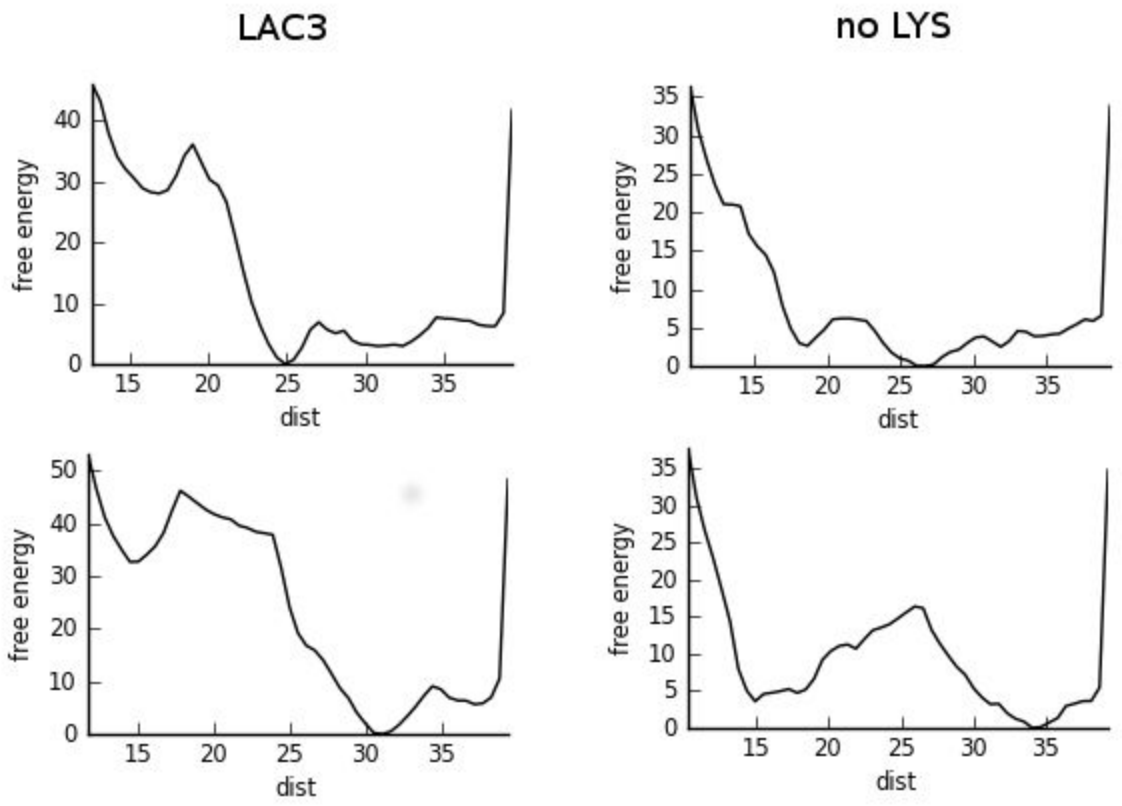


Figure 7.5. Free energy profiles of the $[Ru(bpy)_3]^{2+}$/laccase system as function of the Ru-TNC distance for the wild type and K40S-K71S enzymes. Figure prepared with Desmond (Bowers et al., 2006) and modified with GIMP (https://www.gimp.org/).

This result does not imply that mutating only K40 and K71 would increase the statistical weight of the alternative binding mode (Figure 7.4A) enough to alter the macroscopic properties

of the system. It only suggests that it would be a good start to redesign LAC3, especially if combined with the elimination of the T1 copper (mutating the coordinating Cys to, for example, Ser), aiming to test how the TNC itself can sustain catalysis.

7.3.2 Simulation of UNIK40-, UNIK71-, UNIK157- and UNIK161-Ru

UNIK40- and UNIK71-Ru were built with covalent docking and MD (see Appendix 7 for the root mean square deviation plots). Ten frames, equi spaced in time over the last 10 ns of the MD runs, were taken and T_{DA} between $[Ru(bpy)_3]^{2+}$ and the Cu ions was calculated. In agreement with the experiment, ET to the T1 copper is similar in these two UNIK variants and slower than in LAC3+Ru (Table 7.2, 7.3 and 7.4). Nonetheless, $[Ru(bpy)_3]^{2+}$ may transfer electrons to both copper centers with virtually the same T_{DA} value (Table 7.4). Therefore, the ET mechanism might be quite complex for these two hybrid variant. Again, it would be interesting to produce a T1 copper depleted versions of these variants in the lab and test how catalysis is affected when ET is only sustained by the TNC. As mentioned above, a direct supply of electrons to the TNC should increase ET directionality thanks to the fast peroxide intermediate formation.

Table 7.4. Calculated T_{DA} values between Ru and Cu ions for LAC3-, UNIK161- and UNIK157-Ru.

System	$T_{DA}^{(Ru,T1)}$ (eV)	$T_{DA}^{(Ru,TNC)}$ (eV)
UNIK157-Ru	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$
UNIK161-Ru	$1 \cdot 10^{-7}$	$5 \cdot 10^{-8}$
LAC3-Ru (K40)	$2 \cdot 10^{-13}$	$2 \cdot 10^{-10}$
LAC3-Ru (K71)	$6 \cdot 10^{-11}$	$3 \cdot 10^{-9}$

For the sake of simplicity, LAC3-Ru was modelled splitting the contributions from positions K40 and K71 in UNIK40- and UNIK71-Ru. According to metadynamics simulations, $[Ru(bpy)_3]^{2+}$ is more stable outside the TNC cavity (Figure 7.6 and 7.7) than in a buried conformation (as suggested by covalent docking). T_{DA} was calculated for the former conformation, randomly picking ten representative snapshots from the simulation. As expected, $[Ru(bpy)_3]^{2+}$ primarily transfers electrons to the TNC, observing higher efficiency from position 71 (Table 7.4). In accord with the experiment, ET to the T1 copper is the slowest among all the hybrid laccase/$[Ru(bpy)_3]^{2+}$ systems (Table 7.2 and 7.4).

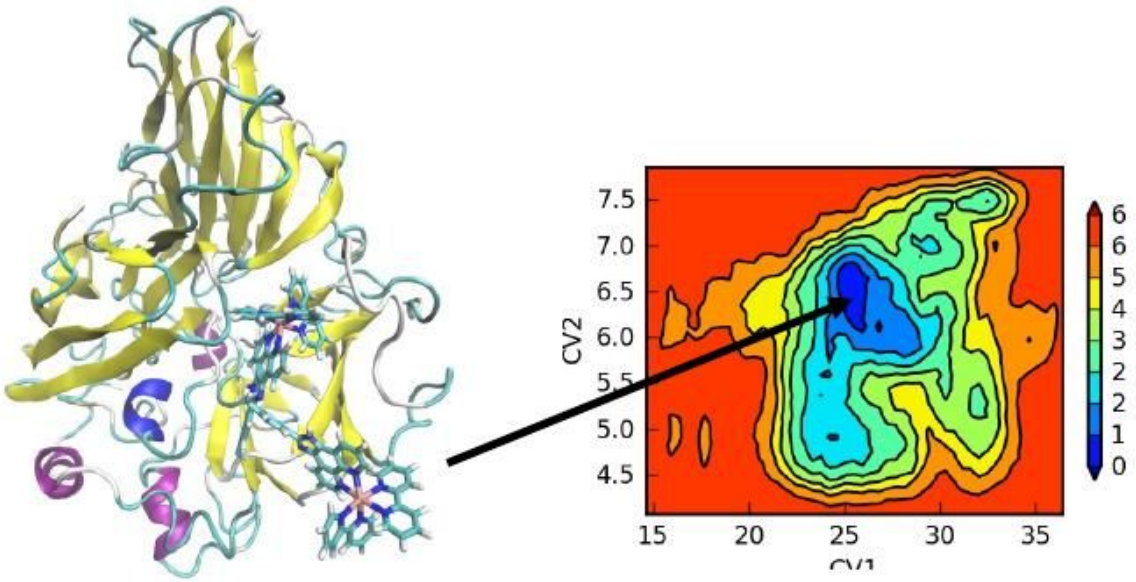


Figure 7.6. Result of the metadynamics simulation of UNIK40-Ru. CV1 is the Ru-T2 copper distance, CV2 is the distance between the N and CA atoms of K40. Figure prepared with VMD (Humphrey et al., 1996), Desmond (Bowers et al., 2006) and modified with GIMP (https://www.gimp.org/).

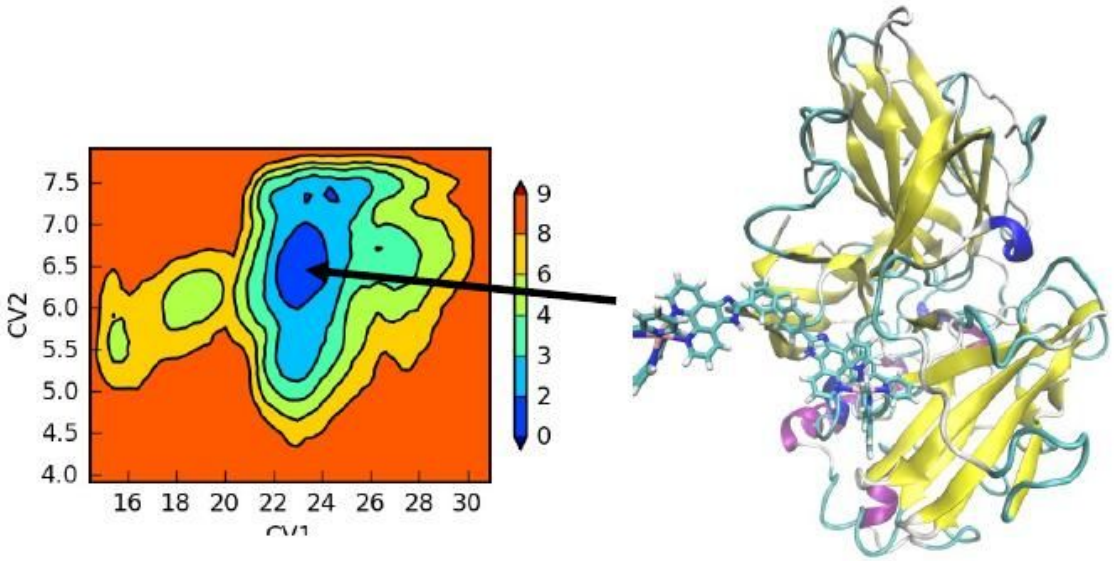


Figure 7.7. Result of the metadynamics simulation of UNIK71-Ru. CV1 is the Ru-T2 copper distance, CV2 is the distance between the N and CA atoms of K40. Figure prepared with VMD (Humphrey et al., 1996), Desmond (Bowers et al., 2006), and modified with GIMP (https://www.gimp.org/).

It is worth noting that, despite the very low T_{DA} value between Ru and the TNC ions (the first electron acceptors), LAC3-Ru catalytic oxidation of p-styrene sulfonate is not dramatically inferior to that of LAC3+Ru and UNIK161-Ru (Table 7.2). This finding reinforces the hypothesis that highly efficient T1 copper depleted UNIK-Ru species could be designed, screening the protein surface for the larger T_{DA} values.

7.3.3 Design of UNIK-Ru variants

The UNIK-Ru variants were designed with PELE and T_{DA} scored with the VMD Pathways plugin. The most promising variants are listed in Table 7.5. The T1 copper was not considered, since it was decided to pursue a direct ET to the TNC, aiming to increase ET directionality. For this reason, these variants should be experimentally tested after removing the T1 copper from the laccase. The results are presented as relative T_{DA} values, using UNIK71-Ru as reference. Indeed, K71 outperforms K40 in carrying electrons to the TNC in LAC3-Ru (Table 7.4).

Table 7.5. N-fold improvement of the computational designs relative to UNIK71-Ru of T_{DA} between Ru and the TNC.

System	$T_{DA}^{(Ru,TNC)}$ improvement factor
UNIK83-Ru	100
UNIK349-Ru	50
UNIK157-Ru	33
UNIK75-Ru	25
UNIK69-Ru	20
UNIK161-Ru	17
UNIK76-Ru	15
UNIK73-Ru	10
UNIK405-Ru	10

For all the variants listed in Table 7.5, T_{DA} between $[Ru(bpy)_3]^{2+}$ and the TNC is greatly improved, from 10 to 100 times, which means that ET should be 100-10,000 times faster (Eq. 2.49). The designed $[Ru(bpy)_3]^{2+}$ attach points are distributed in correspondence of the oxygen entry channel (83, 349, Figure 7.8A) and in/next to the water exit channel (69, 73, 75, 76, 405, Figure 7.8B). Residues 157 and 161, which are not a proper design but rather a repurposing of

existing mutants, are located in between the oxygen channel and the T1 pocket. The water channel is also where the binding mode of Figure 7.4A was found when LAC3+Ru was simulated. In that case, T_{DA} is estimated to be more than 1,000 times greater than in UNIK71-Ru (comparing data from Table 7.3 and 7.4), thanks to the buried position of $[Ru(bpy)_3]^{2+}$. Testing LAC3-Ru, after T1 copper removal and mutation of K40 and K71 into neutrally charged residues, is suggested. However, a more exhaustive redesign might be necessary to express the full potential of this variant.

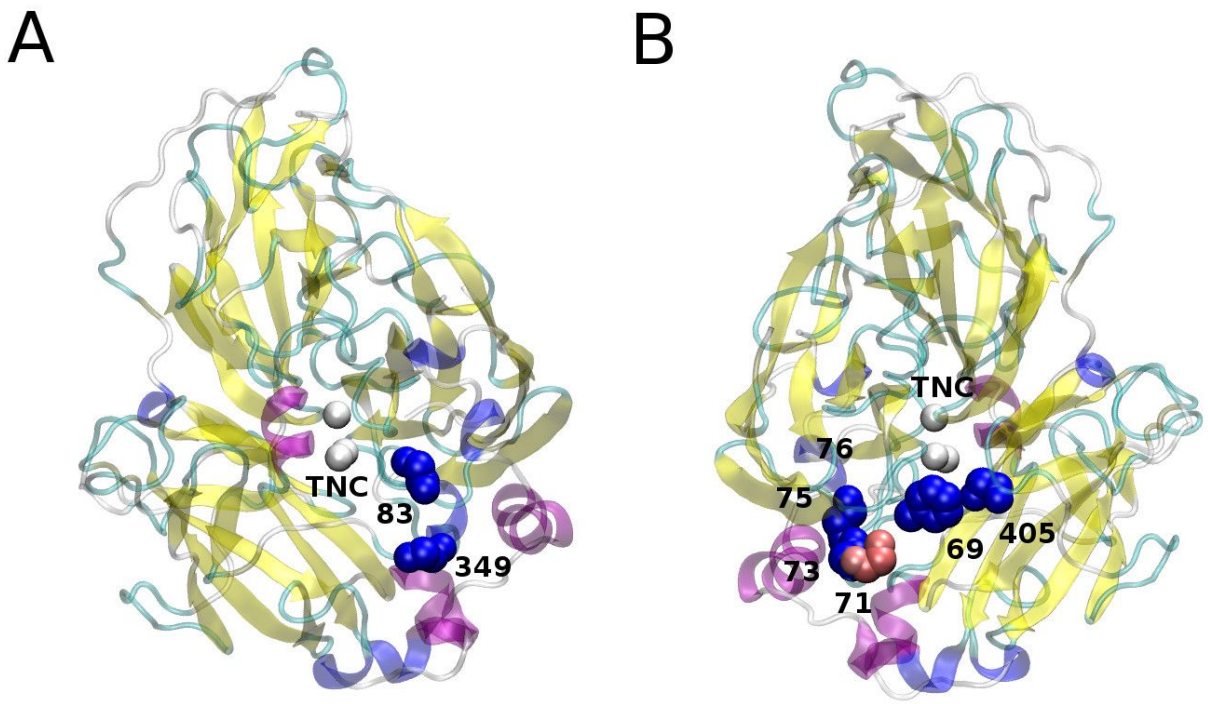


Figure 7.8. $[Ru(bpy)_3]^{2+}$ grafting position in the UNIK-Ru designs (positions 157 and 161 excluded). Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

7.4 Closure

A computational methodology that combines docking, PELE sampling and design, free energy calculations (metadynamics) and quick empirical T_{DA} estimations was adopted here to redesign a LAC3/$[Ru(bpy)_3]^{2+}$ hybrid photocatalyst. Such protocol was firstly benchmarked over a small set of available experimental data and it proved to be capable of discriminating between the different LAC3/$[Ru(bpy)_3]^{2+}$ variants, producing results in agreement with the experiment. At this stage, knowing that the fast BT is the major problem of the photocatalyst, it was

hypothesized to directly convey the electrons to the TNC, skipping the T1 copper. This alternative ET pathway is likely the most promising way to fight BT since k_{ET}/k_{BT}~10^{1-2} and k_{O-O} >> k_{BT} assure ET directionality and k_{IET} would no longer be the upper limit of k_{ET}. With this task in mind, several mutants with significantly improved ET to the TNC were designed and are currently being validated in the labs of Aix Marseille Université.

REFERENCES:

Balabin IA., Hu X., Beratan DN. 2012. Exploring biological electron transfer pathway dynamics with the Pathways plugin for VMD. *Journal of computational chemistry* 33:906–910.

Beratan D., Betts J., Onuchic J. 1991. Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science* 252:1285–1288.

Bowers K., Kevin B., Edmond C., Huafeng X., Ron D., Michael E., Brent G., John K., Istvan K., Mark M., Federico S., John S., Yibing S., David S. 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC'06)*. DOI: 10.1109/sc.2006.54.

Bussi G., Laio A., Parrinello M. 2006. Equilibrium free energies from nonequilibrium metadynamics. *Physical review letters* 96:090601.

Cole JL., Ballou DP., Solomon EI. 1991. Spectroscopic characterization of the peroxide intermediate in the reduction of dioxygen catalyzed by the multicopper oxidases. *Journal of the American Chemical Society* 113:8544–8546.

Essmann U., Ulrich E., Lalith P., Berkowitz ML., Tom D., Hsing L., Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics* 103:8577.

Farver O., Ole F., Scot W., Olga K., Loginov DS., Israel P. 2011. Intramolecular electron transfer in laccases. *The FEBS journal* 278:3463–3471.

Farver O., Wherland S., Pecht I. 1994. Intramolecular electron transfer in ascorbate oxidase is enhanced in the presence of oxygen. *The Journal of biological chemistry* 269:22933–22936.

Galli C., Madzak C., Vadalà R., Jolivalt C., Gentili P. 2013. Concerted electron/proton transfer mechanism in the oxidation of phenols by laccase. *Chembiochem: a European journal of chemical biology* 14:2500–2505.

Gordon JC., Myers JB., Folta T., Shoja V., Heath LS., Onufriev A. 2005. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* 33:W368–71.

Humphrey W., William H., Andrew D., Klaus S. 1996. VMD: Visual molecular dynamics. *Journal of molecular graphics* 14:33–38.

Jacobson MP., Friesner RA., Zhexin X., Barry H. 2002. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of molecular biology* 320:597–608.

Jacobson MP., Pincus DL., Rapp CS., Day TJF., Honig B., Shaw DE., Friesner RA. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins* 55:351–367.

Juris A., Balzani V., Belser P., von Zelewsky A. 1981. Characterization of the Excited State Properties of Some New Photosensitizers of the Ruthenium (Polypyridine) Family. *Helvetica chimica acta* 64:2175–2182.

Kaminski GA., Friesner RA., Julian T-R., Jorgensen WL. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *The journal of physical chemistry. B* 105:6474–6487.

Laio A., Parrinello M. 2002. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99:12562–12566.

Martyna GJ., Tobias DJ., Klein ML. 1994. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics* 101:4177.

McFarland JM., Francis MB. 2005. Reductive alkylation of proteins using iridium catalyzed transfer hydrogenation. *Journal of the American Chemical Society* 127:13490–13491.

Mohamadi F., Fariborz M., Richards NGJ., Guida WC., Rob L., Mark L., Craig C., George C., Thomas H., Clark Still W. 1990. Macromodel?an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of computational chemistry* 11:440–467.

Morozova OV., Shumakovich GP., Gorbacheva MA., Shleev SV., Yaropolov AI. 2007. “Blue” laccases. *Biochemistry. Biokhimiia* 72:1136–1150.

Nosé S., Shuichi N. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics* 81:511.

Olsson MHM., Søndergaard CR., Rostkowski M., Jensen JH. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation* 7:525–537.

Palmer AE., Quintanar L., Severance S., Wang T-P., Kosman DJ., Solomon EI. 2002. Spectroscopic characterization and O2 reactivity of the trinuclear Cu cluster of mutants of the multicopper oxidase Fet3p. *Biochemistry* 41:6438–6448.

Pardo I., Santiago G., Gentili P., Lucas F., Monza E., Medrano FJ., Galli C., Martínez AT., Guallar V., Camarero S. 2016. Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Catal. Sci. Technol.* DOI: 10.1039/c5cy01725d.

Pita M., M P., S S., T R., V F., A Y., L G. 2006. Direct heterogeneous electron transfer reactions of fungal laccases at bare and thiol-modified gold electrodes. *Electrochemistry communications* 8:747–753.

Ramírez P., Pablo R., Nicolas M., Rafael A., Tautgirdas R., Adam H., Lo G., Sergey S. 2008. Direct electron transfer from graphite and functionalized gold electrodes to T1 and T2/T3 copper centers of bilirubin oxidase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1777:1364–1369.

Sastry GM., Adzhigirey M., Day T., Annabhimoju R., Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* 27:221–234.

Schmidtke P., Le Guilloux V., Maupetit J., Tufféry P. 2010. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic acids research* 38:W582–9.

Schneider L., Mekmouche Y., Rousselot-Pailley P., Simaan AJ., Robert V., Réglier M., Aukauloo A., Tron T. 2015. Visible-Light-Driven Oxidation of Organic Substrates with Dioxygen Mediated by a [Ru(bpy)3](2+) /Laccase System. *ChemSusChem* 8:3048–3051.

Shleev S., Sergey S., Andreas C., Vladimir S., Dosymzhan B., Alexander Y., Lo G., Tautgirdas R. 2005. Electrochemical redox transformations of T1 and T2 copper sites in native Trametes hirsuta laccase at gold electrode. *Biochemical Journal* 385:745–754.

Shleev S., Sergey S., Marcos P., Yaropolov AI., Tautgirdas R., Lo G. 2006. Direct Heterogeneous Electron Transfer Reactions ofTrametes hirsuta Laccase at Bare and Thiol-Modified Gold Electrodes. *Electroanalysis* 18:1901–1908.

Simaan AJ., Mekmouche Y., Herrero C., Moreno P., Aukauloo A., Delaire JA., Réglier M., Tron T. 2011. Photoinduced multielectron transfer to a multicopper oxidase resulting in dioxygen reduction into water. *Chemistry* 17:11743–11746.

Tadesse MA., D’Annibale A., Galli C., Gentili P., Sergi F. 2008. An assessment of the relative contributions of redox and steric issues to laccase specificity towards putative substrates. *Organic & biomolecular chemistry* 6:868–878.

Toukan K., Kahled T., Aneesur R. 1985. Molecular-dynamics study of atomic motions in water. *Physical Review B: Condensed Matter and Materials Physics* 31:2643–2648.

Trott O., Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 31:455–461.

Tuckerman M., Berne BJ., Martyna GJ. 1992. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics* 97:1990.

Zhu K., Borrelli KW., Greenwood JR., Day T., Abel R., Farid RS., Harder E. 2014. Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *Journal of chemical information and modeling* 54:1932–1940.

8. Simulation of hemoglobin in silica gel

According to the tertiary two-state (TTS) model, hemoglobin's (Hb) ligand affinity is directly controlled by the tertiary structure of each subunit, which can exists in two states r and t with high and low affinity respectively. Recently, Spiro and coworkers measured the evolution of the heme-His stretching frequency (vFe-His) of Hb trapped in a silica gel after carbon monoxide (CO) binding/release. In fact, vFe-His was found to correlate with Hb's oxygen affinity (Matsukawa et al., 1985)) and it is used here to follow heme's reactivity. The conformational mobility of the protein is significantly slowed down in the silica gel, practically freezing its quaternary structure while subunits' conformations can rearrange internally upon CO release (Jones, Gurusamy & Spiro, 2012). It follows that tertiary and quaternary structures are uncoupled. Spiro and coworkers found that vFe-His shifts from high- to low-affinity values within the starting quaternary structure. Therefore, heme's reactivity can be uncoupled from the tetramer arrangement (Jones, Gurusamy & Spiro, 2012), showing a r to t transition within the R state, in accord with the TTS model. Furthermore, they were able to isolate the contributions of α and β chains and found that they have a diverse response to CO release: vFe-His shifts from 220 to 212 cm^{-1} and from 221 to 218 cm^{-1} in α and β respectively (Jones et al., 2014).

Here, the experiment was simulated by modeling carboxy-hemoglobin (HbCO) CO release using appropriate restraints to mimic its behavior in the silica gel, aiming to characterize the structure of the r and t affinity states within the R2 quaternary structure (which is in equilibrium with the R structure in solution (Hub, Kubitzki & de Groot, 2010; Fan et al., 2013)).

8.1 Computational details

System Setup and molecular dynamics (MD). The following structures were taken from the Protein Data Bank (PDB): 1A3N (Tame & Beatrice, 2000) (T, deoxyHb), 1BBB (Silva, Rogers & Arnone, 1992) (R2, HbCO) and prepared with Protein Preparation Wizard (Sastry et al., 2013) in the Schrödinger suite of programs. After appropriate solvation, charge neutralization and 0.15 M NaCl addition, a short MD simulation was performed with Desmond (Bowers et al., 2006) for each structure, in order to obtain some conformational sampling. Heme's OPLS parameters were generated with the hetgrp_ffgen utility of Schrödinger. More in detail, the systems were equilibrated with Desmond's default protocol followed by a 1 ns NpT

production phase at 300 K (this short MD was intended to keep the initial quaternary/tertiary arrangements). Along equilibration and short production, the Cα RMSD from the initial crystal structures stays below 1.3 Å. The OPLS-2005 force-field (Kaminski et al., 2001) and the SPC water model (Toukan, Kahled & Aneesur, 1985) were used. The temperature is regulated with the Nosé−Hoover chain thermostat (Nosé & Shuichi, 1984) with a relaxation time of 1.0 ps, and the pressure is controlled with the Martyna−Tobias−Klein barostat (Martyna, Tobias & Klein, 1994) with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator (Tuckerman, Berne & Martyna, 1992) is employed with bonded, near, and far time steps of 2.0, 2.0, and 6.0 fs, respectively. A 9 Å cutoff is used for non bonded interactions together with the smooth particle mesh Ewald method (Essmann et al., 1995). For each system, six snapshots, at times 0, 0.2, 0.4, 0.6, 0.8, and 1 ns were selected for the quantum mechanics/molecular mechanics (QM/MM) study.

QM/MM Calculations. Structures were optimized followed by frequency calculations using Qsite (Murphy, Philipp & Friesner, 2000). The heme, the proximal histidine and the sixth ligand (molecular oxygen or carbon monoxide, if present) of a given chain subunit (either α of β) were included in the QM region. The rest of the subunit containing the QM region, the other chains and a 12 Å water layer were treated classically (the remainder of the solvent was removed). The MM minimization included the chain under study, containing the QM region, plus all other residues within 5 Å. The density functional method with the B3LYP (Hertwig & Wolfram, 1997) functional and the lacvp* (Hay, Jeffrey Hay & Wadt, 1985) basis set was used for the quantum region, while the OPLS-AA (Kaminski et al., 2001) force field was used for the classical part. A non bonded cutoff of 50 Å was applied. In the case of R2, the sixth ligand was removed after geometry optimization and re-optimized before frequency calculation. In doing so, the five-coordinated $R2_{\alpha}$, $R2_{\beta}$, T_{α},T_{β} were obtained, allowing direct comparison with the Raman data, which is limited to five-coordinate heme. Quintet and singlet states were used for five- and six-coordinated heme, respectively. As expected, along the QM/MM optimizations, there were no substantial structural changes except for residues in direct contact with the QM region. Thus, significant tertiary changes require a wider level of sampling, which can be achieved with PELE.

PELE Simulation. The five-coordinated states obtained after QM/MM geometry optimizations were used as initial models for the PELE simulation, aiming to reproduce tertiary changes that accompany CO release of HbCO. Heme's OPLS parameters were generated with the hetgrp_ffgen utility of Schrödinger, and the QM/MM ESP charges were used to model its electron distribution. Importantly, the ANM perturbation was applied only to the chain under

study. Furthermore, only residues within 25 Å of the heme were minimized in order to avoid quaternary rearrangement of the tetramer, while allowing relaxation of the intersubunit interface (see Appendix 5 for an example of input file). Some intermediates generated by PELE were then subjected to QM/MM geometry optimization and frequency calculation to track the dependence of heme's reactivity on tertiary changes.

8.2 Results and discussion

8.2.1 Validation of the computational methodology

Although QM/MM frequencies are systematically below the experimental values (Table 8.1), the 7 cm^{-1} frequency difference between the β and α chains in deoxyHb is in quantitative agreement with the experimental data. Moreover, as observed experimentally, the frequency shift caused by CO release is much larger for the α than for the β chains, 16 vs 6 cm^{-1}, respectively.

Table 8.1. Average QM/MM Fe–His bond distances (*d*) and stretching frequencies (vFeHis) for five-coordinated α ($R2_\alpha$ and T_α) and β ($R2_\beta$ and T_β) Hb subunits.

structure	d[a], Å	vFe-His[a], cm^{-1}	vFe-His, cm^{-1} (exp)
$R2_\alpha$	2.20(1)	213(2)	220[b]
T_α	2.24(1)	197(2)	212[b]
$R2_\beta$	2.22(1)	210(2)	221[b]
T_β	2.21(1)	204(2)	218[b]

[a]Values averaged over six snapshots taken from the 1 ns molecular dynamics simulation. The number in parentheses is the uncertainty of the mean, estimated as $\sigma/N^{1/2}$, where σ is the standard deviation and *N* is the number of values used to calculate the mean values (N=6).

[b]Sol–gel encapsulated HbCO from Jones et al. (Jones et al. 2014).

In order to verify that these results are not biased by the initial structure, the calculations were repeated with 2dn1 (R, oxyHb) and 2dn2 (T, deoxyHb) (Park et al., 2006). As expected, the frequency values obtained with these crystals are comparable (the R–T shifts are 11 and 7 cm^{-1} in α and β respectively). These results show that the QM/MM methodology is sufficiently accurate to monitor the evolution of the individual chain frequencies following CO release.

Table 8.2. Evolution of structure and the QM/MM-derived vFeHis of the unligated α subunit during the PELE Simulation.

step	vFe-His, cm^{-1}	d, Å	θ, deg	Oop, Å	$RMSD_R$, Å	$RMSD_T$, Å
0	213	2.20	95	0.07	1.59	0.00
1	210	2.21	99	0.10	1.62	0.77
10	205	2.23	102	0.15	1.48	0.74
20	200	2.23	105	0.18	1.46	0.93
30	199	2.23	103	0.17	1.53	0.87
40	201	2.22	102	0.16	1.46	0.88
50	195	2.25	105	0.18	1.45	0.89
T	197(2)	2.24(1)	102(2)	0.20	0.00	1.59

Table 8.3. Evolution of structure and the QM/MM-derived vFeHis of the unligated β subunit during the PELE Simulation.

step	vFe-His, cm^{-1}	d, Å	θ, deg	Oop, Å	$RMSD_R$, Å	$RMSD_T$, Å
0	215	2.19	96	0.08	3.98	0.00
1	213	2.19	101	0.11	3.77	0.84
10	203	2.20	97	0.16	3.67	0.88
20	204	2.20	100	0.14	3.68	0.94
30	204	2.20	99	0.18	3.70	0.80
40	205	2.21	98	0.15	3.65	0.90
50	206	2.20	101	0.14	3.69	0.82
T	204(2)	2.21(1)	102(2)	0.18	0.00	3.68

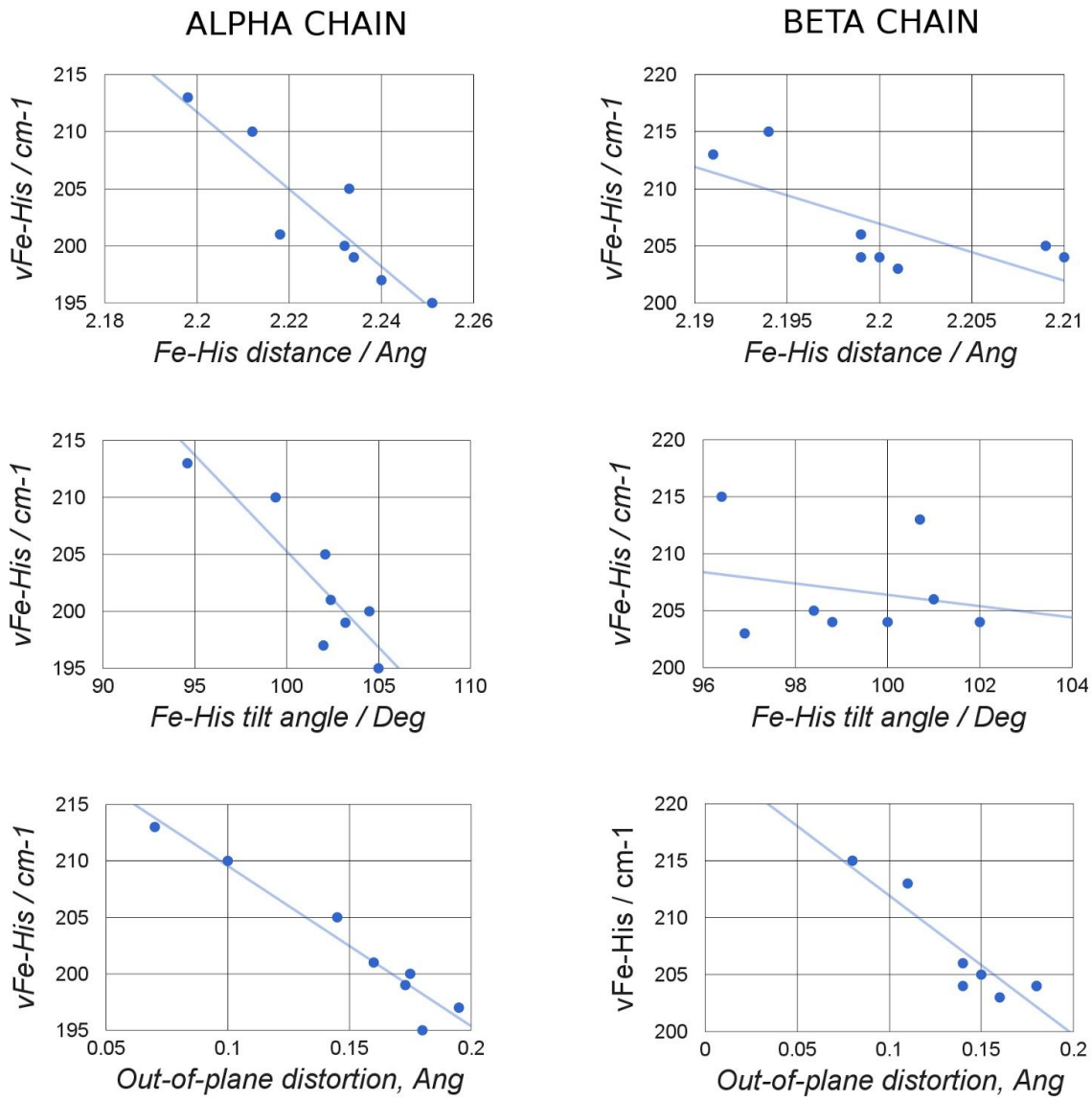


Figure 8.1. Dependence of the computed Fe–His stretching frequency (values in cm–1) on the Fe–His bond distance and tilt angle (NC–Fe–Nε angle, where NC is in heme pyrrole "C" and Nε is in the proximal histidine), and with the heme out-of-plane distortion (measured as the average displacement of heme's atoms from their least-squares plane), for Hb α chains (red diamonds) and β chains (blue circles). Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

8.2.2 vFeHis evolution in α and β Chains

Table 8.2 and 8.3 show the vFeHis evolution of the unligated chain along a PELE conformational. Just as it does in the experiment (Table 8.1), vFeHis decreases (Table 8.2 and 8.3) and its variation is greater for the α chain (Table 8.2 and 8.3). Twenty steps are sufficient to

reach a value around which it subsequently fluctuates. This value is essentially the same as that computed for the T structure, 197 and 204 cm^{-1} for the α and β chains respectively. Thus, in striking agreement with the gel experiments, vFeHis evolves from high- to low-affinity values solely as a result of tertiary forces initiated by CO release, in the absence of quaternary changes. However, the tertiary structure itself does not evolve very far in the direction of that found in the T quaternary structure. For each chain, the root-mean square deviation (RMSD) between the evolving PELE structures and the initial unligated structure never exceeds 0.94 Å (Table 8.2 and 8.3). The RMSD between the initial conformation and the T structure is much larger, 1.59 and 3.68 Å in the α and β chains. More importantly, the RMSD of successive PELE intermediates relative to the T structure diminishes very little from these values, fluctuating around 1.47 Å (α) and 3.67 (β) (Table 3).

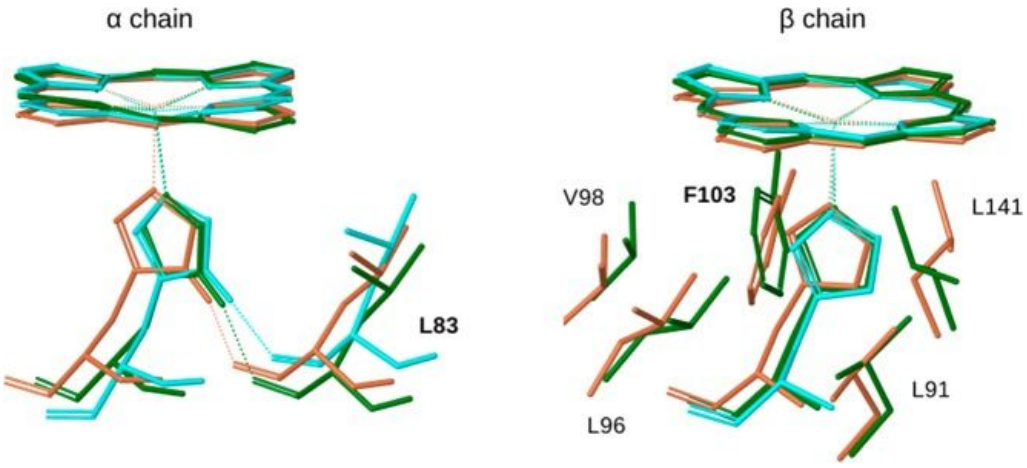


Figure 8.2. Overlay of heme, the proximal histidine, and relevant contacting residues for the initial (light brown) and intermediate (green) PELE structures, and for deoxyHb. Figure prepared with Maestro 9.7 (Schrödinger, LLC, New York, NY, 2014) and modified with GIMP (https://www.gimp.org/).

It is clear that the PELE intermediates, although displaying vFeHis frequencies characteristics of a low-affinity state, have tetramer conformations that are distinct from the T quaternary structure. The factors controlling vFeHis must therefore be of local nature. For the same reason stated above, the calculations were repeated with 2dn1 and 2dn2 crystal structures. Consistently, vFeHis shifted after PELE from 210 to 199 cm^{-1} for the α chain, and from 212 to 206 cm^{-1} in the β chain.

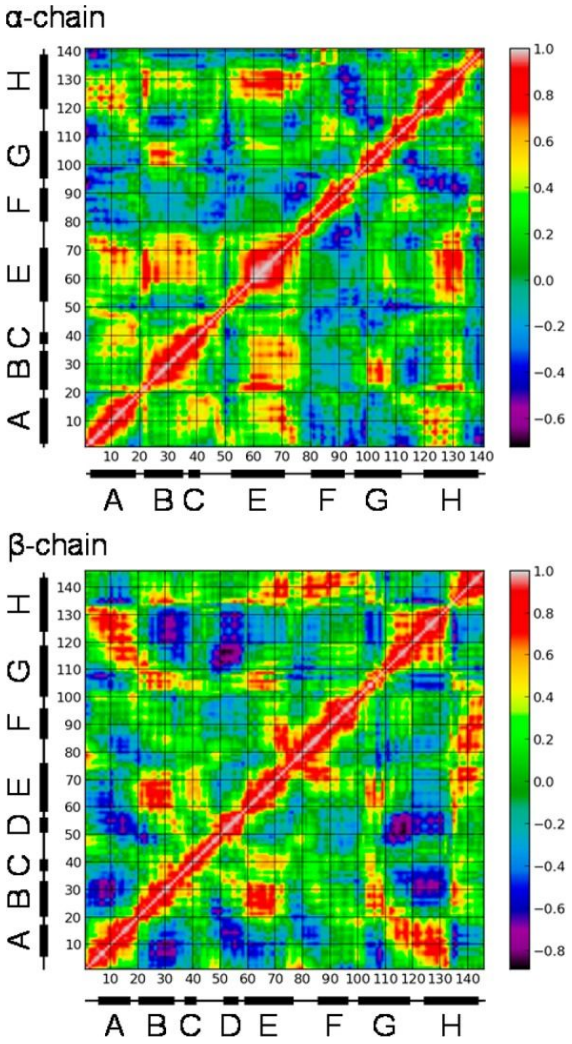


Figure 8.3. Cross-correlation matrix for successive residue Cα displacements along the trajectory between initial and intermediate PELE structures. Successive helices are indicated by black bars with letter labels. Figure prepared with matplotlib (http://matplotlib.org/) and modified with GIMP (https://www.gimp.org/).

8.2.3 Structural determinants of vFeHis evolution

It can be observed that three structural parameters of the heme are most consistently correlated to vFeHis in the α chain: the Fe−His bond length, the out-of-plane distortion of the heme, measured as the average displacement of heme's atoms from their least-squares plane, and the tilt angle of the Fe−His bond with respect to the heme plane (Figure 8.1). Notably, the first two parameters have been directly related to oxygen affinity by quantum mechanical calculations (Capece et al., 2006; Bikiel et al., 2010). Moreover, Friedman and co-workers linked vFeHis to the Fe−His bond tilt within a general model for allostery (Friedman et al., 1982;

Scott & Friedman, 1984), and a tilted Fe−His bond has long been noted as a feature of the T-state α chain (Perutz, 1970; Baldwin & Chothia, 1979). On the other hand, the out-of-plane distortion of the heme is the only parameter that correlated with vFeHis for the β chain, for which bond tilt and elongation are lower along the PELE trajectory.

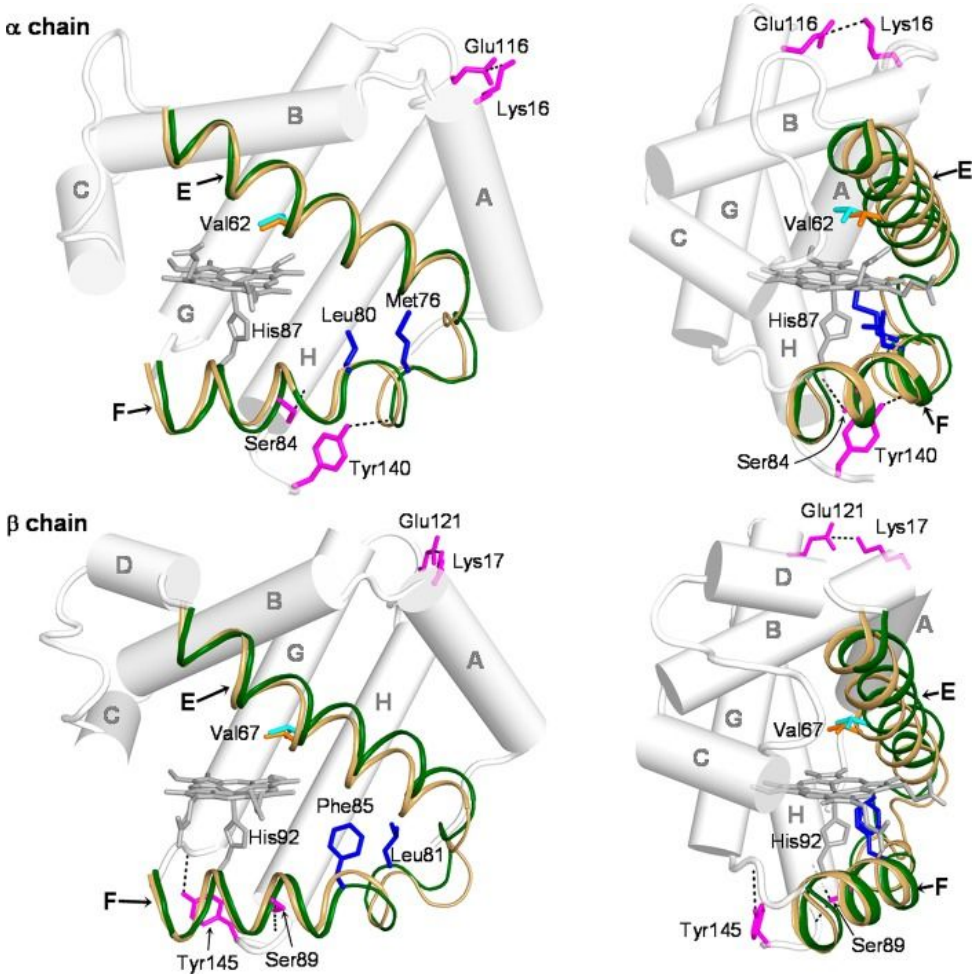


Figure 8.4. Ribbon diagram of the E and F helices for the PELE intermediate (green helices; Valα62/β67 in cyan) and initial (light brown helices; Valα62/β67 in orange) structures with hemes superposed. The remaining helices are represented by cylinders. Also shown are packing residues (blue) at the EF corner, and residues (magenta) forming key interhelical H-bonds and salt-bridges. Source: Jones et al. 2014 (Jones et al., 2014).

It is then clear that heme's reactivity is controlled differently in Hb's subunits, consistently with their different vFeHis shift. In the β chain, the heme distortion, enforced by the packing interactions with nearby side chains, is the likely determinant of its oxygen affinity. On the other hand, such property is regulated through both proximal and ring strain in the α chain. Figure 8.2

illustrates the differing heme-associated changes between the two chains, and the similar local structures in the PELE intermediate and deoxyHb.

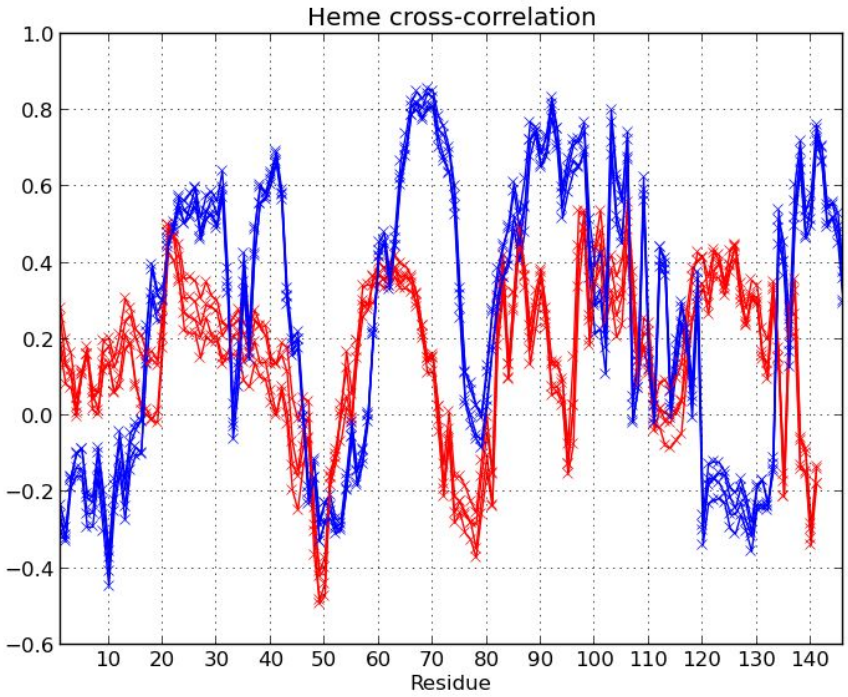


Figure 8.5. Cross-correlation of heme's nitrogen atoms with the rest of the subunit. Figure prepared with matplotlib (http://matplotlib.org/).

In order to investigate the origin of these structural differences between the evolving α and β chains, the cross-correlation matrix was calculated for both PELE trajectories. The C_α cross-correlation shows quite different patterns for the two chains (Figure 8.3). In the α chains, the displacements are positively correlated for the first 75 residues (A, B, C and E helices and the EF corner), but they correlate poorly with the next 25 amino acids, which contain the F and part of the G helices. In particular the correlation between the E (55−70) and F (75−90) helices is essentially null. In contrast, the E and F helix displacements are positively correlated in the β chains, while the E and B helices are anticorrelated with the A helix. Figure 8.4 illustrates the differing displacements of the E and F helices, which sandwich the heme group. Two orientations are shown, in the EF plane (left), and perpendicular to it (right), with the hemes superposed. In the β chain, the E and F helices move in the same direction between the initial (light brown) and intermediate (green) PELE structures, toward the protein exterior (right view) and also toward the EF corner (left view). However, they move "oppositely" in the α chain, the F

helix again moving out and toward the EF corner, while the E helix moves in and away from the EF corner. This difference may be partly due to packing constraints in the EF corner, which are more pronounced for the β chain (represented in Figure 8.4 by the F85β and L81β side chains) than for the α chain. In addition, the H helix may play a role, via H-bonds from the penultimate residue, Y140α and Y145β to the F helix residues P77α and V98β, respectively, and salt-bridges from E116α and E121β to the A helix residues K16α and K17β, respectively. Indeed, these regions are significantly correlated in the α chain (Figure 8.3). The differing orientations of these connections (Figure 8.4) may guide the E and F helices in different directions, as reflected in the E and A helices being correlated for the α chains but anticorrelated in the β chains. In the β chain, the concerted helix displacements lead to steric contacts with the heme that induce out-of-plane distortions (Figure 8.2), especially F103β which was identified in earlier QM/MM computations as making an important contribution to T−R oxygen affinity difference in Hb (Alcantara et al., 2007). Corresponding contacts are attenuated in the α chain, although the α heme also distorts, perhaps in response to the proximal strain.

Displacements of the heme are strongly correlated with those of the E helix in the β chain, but less so for the α chain (Figure 8.5). In the latter, the non concerted motion of the E and F helices induces tilting of the Fe−His bond, due to the proximal His residue, Hisα87, becoming displaced relative to the heme. The tilting is largely mediated by the H-bond connecting Hisα87 with the backbone carbonyl of the upstream residue, Leuα83. When the strength of this H-bond was lowered in silico, by reducing the atomic charge of the donor (from 0.42 to 0.1) and increasing that of the acceptor (from −0.5 to −0.1), the Fe−His tilt did not exceed 95º. An important role of this H-bond in regulation of oxygen affinity has previously been found for other globins and on the basis of computational studies (Capece et al., 2006).

8.3 Closure

The differences between r and t states are located in close proximity to the heme and they are distinct from the tertiary structures found in the oxy- and deoxy-Hb crystal structure. In the α chain, heme's reactivity is modulated by both ring (out-of-plane distortion) and proximal strain (Fe-His elongation and tilting), while only the former is relevant in the β subunit. The different control of heme's reactivity in Hb's subunits is likely dictated by their distinct tertiary dynamics and of their diverse packing between the E and F helices. At the best of my knowledge, this is the first computational study that supports the TTS model.

REFERENCES:

Alcantara RE., Xu C., Spiro TG., Guallar V. 2007. A quantum-chemical picture of hemoglobin affinity. *Proceedings of the National Academy of Sciences of the United States of America* 104:18451–18455.

Baldwin J., Chothia C. 1979. Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. *Journal of molecular biology* 129:175–220.

Bikiel DE., Forti F., Boechi L., Nardini M., Luque FJ., Martí MA., Estrin DA. 2010. Role of heme distortion on oxygen affinity in heme proteins: the protoglobin case. *The journal of physical chemistry. B* 114:8536–8543.

Bowers K., Kevin B., Edmond C., Huafeng X., Ron D., Michael E., Brent G., John K., Istvan K., Mark M., Federico S., John S., Yibing S., David S. 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC'06)*. DOI: 10.1109/sc.2006.54.

Capece L., Marti MA., Crespo A., Doctorovich F., Estrin DA. 2006. Heme protein oxygen affinity regulation exerted by proximal effects. *Journal of the American Chemical Society* 128:12455–12461.

Essmann U., Ulrich E., Lalith P., Berkowitz ML., Tom D., Hsing L., Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics* 103:8577.

Fan J-S., Jing-Song F., Yu Z., Wing-Yiu C., Virgil S., Ho NT., Chien H., Daiwen Y. 2013. Solution Structure and Dynamics of Human Hemoglobin in the Carbonmonoxy Form. *Biochemistry* 52:5809–5820.

Friedman JM., Rousseau DL., Ondrias MR., Stepnoski RA. 1982. Transient Raman study of hemoglobin: structural dependence of the iron-histidine linkage. *Science* 218:1244–1246.

Hay PJ., Jeffrey Hay P., Wadt WR. 1985. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *The Journal of chemical physics* 82:270.

Hertwig RH., Wolfram K. 1997. On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chemical physics letters* 268:345–351.

Hub JS., Kubitzki MB., de Groot BL. 2010. Spontaneous quaternary and tertiary T-R transitions of human hemoglobin in molecular dynamics simulation. *PLoS computational biology* 6:e1000774.

Jones EM., Monza E., Balakrishnan G., Blouin GC., Mak PJ., Zhu Q., Kincaid JR., Guallar V., Spiro TG. 2014. Differential control of heme reactivity in alpha and beta subunits of hemoglobin: a combined Raman spectroscopic and computational study. *Journal of the American Chemical Society* 136:10325–10339.

Jones EM., Gurusamy B., Spiro TG. 2012. Heme Reactivity is Uncoupled from Quaternary Structure in Gel-Encapsulated Hemoglobin: A Resonance Raman Spectroscopic Study. *Journal of the American Chemical Society* 134:3461–3471.

Kaminski GA., Friesner RA., Julian T-R., Jorgensen WL. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *The journal of physical chemistry. B* 105:6474–6487.

Martyna GJ., Tobias DJ., Klein ML. 1994. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics* 101:4177.

Matsukawa S., Mawatari K., Yoneyama Y., Kitagawa T. 1985. Correlation between the iron-histidine stretching frequencies and oxygen affinity of hemoglobins. A continuous strain model. *Journal of the American Chemical Society* 107:1108–1113.

Murphy RB., Philipp DM., Friesner RA. 2000. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *Journal of computational chemistry* 21:1442–1457.

Nosé S., Shuichi N. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics* 81:511.

Park S-Y., Yokoyama T., Shibayama N., Shiro Y., Tame JRH. 2006. 1.25 A resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms. *Journal of molecular biology* 360:690–701.

Perutz MF. 1970. Stereochemistry of cooperative effects in haemoglobin. *Nature* 228:726–739.

Sastry GM., Adzhigirey M., Day T., Annabhimoju R., Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* 27:221–234.
Scott TW., Friedman JM. 1984. Tertiary-structure relaxation in hemoglobin: a transient Raman study. *Journal of the American Chemical Society* 106:5677–5687.
Silva MM., Rogers PH., Arnone A. 1992. A third quaternary structure of human hemoglobin A at 1.7-A resolution. *The Journal of biological chemistry* 267:17248–17256.
Tame JRH., Beatrice V. 2000. The structures of deoxy human haemoglobin and the mutant Hb Tyrα42His at 120 K. *Acta crystallographica. Section D, Biological crystallography* 56:805–811.
Toukan K., Kahled T., Aneesur R. 1985. Molecular-dynamics study of atomic motions in water. *Physical Review B: Condensed Matter and Materials Physics* 31:2643–2648.
Tuckerman M., Berne BJ., Martyna GJ. 1992. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics* 97:1990.

9. Allosteric signalling paths in hemoglobin

In the previous chapter, it has been shown that heme reactivity is modulated by ring (out-of-plane distortion) and proximal strain (heme-His bond lengthening and tilting) in the α subunit, while only the former effect is important in the β chain. Furthermore, such modulation can take place with no need to alter the quaternary structure of hemoglobin (Hb) and with minimal tertiary rearrangement. Apart from ring and proximal strain effects, affinity is also regulated in the distal pocket, where the ligand coordinates Fe (Unzai et al., 1998). This might be more important in β since its low-affinity heme-His stretching frequency is significantly higher than the α's despite these subunits have similar O_2 affinities (Unzai et al., 1998).

Here, the pathways connecting important residues belonging to different chains are explored analyzing molecular dynamics (MD) trajectories. The residues are: H87α and H92α (proximal and ring strain), F103β (ring strain (Alcantara et al., 2007; Jones et al., 2014)), H58α, V62α, H63β and V67β (distal pocket regulation). The aim is to understand: (i) which of these amino acids are more responsible for inter-subunit communication; (ii) how they are connected; (iii) which subunits exchange more information. In this way, the sequence search space for future stability designs (to develop better blood substituents) can be narrowed down.

9.1 Computational details

System setup and molecular dynamics. The 2dn1.pdb Hb's R (ligated) structure was prepared with the protein preparation wizard software (Sastry et al., 2013). The protonation states of the titratable residues were assigned at pH 7 with PROPKA (Olsson et al., 2011) and double-checked with the H++ server (Gordon et al., 2005). After that, the following variants were prepared: (i) full-ligated (carboxy-hemoglobin, HbCO); (ii) α1 unligated; (iii) β1 unligated; full-unligated. The systems were solvated with a 10 Å buffer of waters in an orthorhombic box, neutralized and 0.15 M NaCl was added. After equilibration (default settings), 1 μs NpT production runs were performed with Desmond (Bowers et al., 2006) at 310 K for all the variants. The OPLS-2005 force-field (Kaminski et al., 2001) and the SPC explicit water model (Toukan, Kahled & Aneesur, 1985) were used. The heme parameters were generated with the hetgrp_ffgen utility of Schrödinger, inputting charges and geometries from the QM/MM calculations carried out in Chapter 8 (for the R2 and T states, not for the PELE intermediates).

The temperature was regulated with the Nosé−Hoover chain thermostat (Nosé & Shuichi, 1984) with a relaxation time of 1.0 ps, and the pressure was controlled with the Martyna−Tobias−Klein barostat (Martyna, Tobias & Klein, 1994) with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator (Tuckerman, Berne & Martyna, 1992) was employed with bonded, near, and far time steps of 2.0, 2.0, and 6.0 fs, respectively. A 9 Å cutoff was used for non bonded interactions together with the smooth particle mesh Ewald method (Essmann et al., 1995).

Allosteric signalling paths analysis. The Weighted Implementation of Suboptimal Paths (WISP) (Van Wart et al., 2014) program was used to analyze the MD trajectories. After each residue was represented by its center of mass, the correlation (with elements C_{ij}) and contact-map (with elements M_{ij}) matrices were calculated. These are both NxN matrices, where N is the number of residues. Each element of the contact-map matrix is either 0 or 1 depending on the number of times the respective residues are within 4.5 Å. The functionalized correlation matrix (F) has $-\log|C_{ij}|$ as elements if $M_{ij} = 1$ or ∞ if $M_{ij} = 0$. F represents a network where each residue is a node; if F_{ij} is finite, the i-th and j-th nodes are connected by an edge of distance F_{ij}. The higher the absolute correlation between two residues, the lower the distance between two nodes. For any pair of nodes, a set of connecting paths can be traced, whose distance is the sum of the distances of each edge. Therefore, such paths can be ranked according to the distance covered. After the Dijkstra's algorithm (Dijkstra, 1959) is used to calculate the optimal pathway (i.e. the one with the lowest distance covered) between two user defined nodes, the suboptimal pathways are computed in a recursive fashion whose details are explained in van Wart et al. (Van Wart et al., 2014). In the same paper, the allosteric interaction in the HisH-HisF dimer was studied; binding at the allosteric site produced a shift from 3.1-4.0 to 2.8-3.6 of the length distribution of the top 100 shortest paths that link the allosteric site to the binding pocket. Here, the top 100 shortest paths between α1-β1, α1-β2, β1-α2, α1-α2 and β1-β2 were calculated using default settings, choosing the following residues as possible end-points: F103β, H63β, V67β, H92β, H87α, H58α and V62α. While F103β, H92β and H87α account for ring and proximal strain, the remaining residues aim at describing the dynamics of the distal pocket.

Principal component analysis (PCA). MD data are hard to analyze due to their high-dimensionality. PCA is a common dimensionality reduction technique which consists in diagonalizing the covariance matrix of the atomic coordinates (Hayward, Steven & Groot, 2008), rotating the cartesian axis in the direction along which the data display the largest variation. The lower an eigenvalue, the larger the collective motion described by the associated eigenvector. In particular, the first two eigenvectors (principal components) usually describe most of the

collective motion along a conformational transition pathway (Amadei et al., 1993). Therefore, large scale protein dynamics can be represented with reasonable accuracy taking into account only the first two components. Here, the covariance matrix of the C_α coordinates was built out of the superimposed oxy- (R, 2dn1) and deoxy-Hb (T, 2dn2) crystal structures and the MD trajectories were projected along its first two principal components. In such a way, the extent of R to T transition could be monitored in time.

9.2 Results and discussion

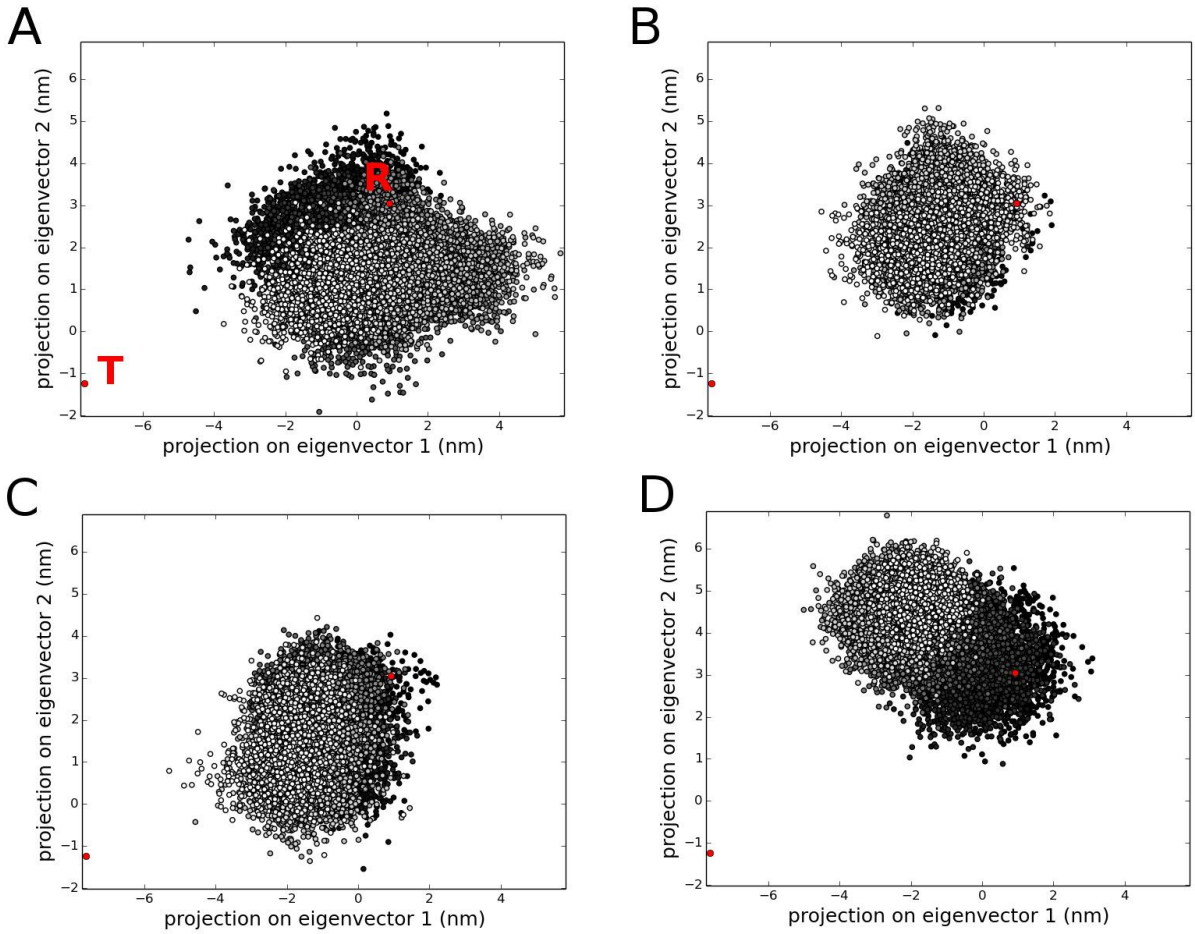


Figure 9.1. Projection of the 1 μs MD trajectory over the first two principal components of the R (2dn1) to T transition (2dn2). The color of the points (frames) gradually change from black (initial frame, 0 μs) to white (final frame, 1 μs). A) HbCO; B) α1-unligated; C) β1-unligated; D) full-unligated.

PCA (Figure 9.1) shows that, independently of the ligation state, the system is far away from reaching the T quaternary structure after 1 μs (although it seems to slowly go in that

direction). As previously reported by computational (Hub, Kubitzki & de Groot, 2010) and experimental studies (Fan, Yang & Choy, 2013), PCA indicates that the tetramer is highly flexible. Comparing the three simulations, a broader exploration by HbCO while a more pronounced R-to-T directionality can be appreciated for α1- and β1-unligated.

Table 9.1. Range of lengths for the best 100 signalling paths.

System	α1-β1	α1-β2	α1-α2	β1-α2	β1-β2
HbCO	3.0-3.1	2.0-2.5	3.8-4.0	2.4-2.9	2.9-3.6
α1-unligated	3.9-4.2	2.0-2.7	3.7-4.1	N.C.	N.C.
β1-unligated	3.7-4.0	N.C.	N.C.	2.4-2.8	3.3-3.8
full-unligated	3.1-3.3	2.1-2.6	3.5-3.7	2.3-2.9	2.5-2.9

Despite none of the four simulations reached the T state, inter-subunit communication is differently channelled according to the WISP analysis. Common to all systems, α1-β2 (α2-β1) seems to be a privileged heme-heme signalling channel (Table 9.1). While α1-β2 (α2-β1) has H87α1 as sole source/sink of allosteric signal in HbCO, signalling is enriched upon CO unbinding in the α1 chain, involving also H58α1 and V62α1 (Table 9.2). Therefore, the response of both proximal and distal pockets to unbinding in the α1 chain are successfully propagated to the β2 subunit. When CO leaves the β1 monomer, F103β1 is the main source of the allosteric signal among the selected residues. On the other hand, β1's distal pocket has a limited role in the β1-α2 allosteric signalling, at least in the amount of time simulated here; perhaps further quaternary rearrangement is necessary to get it more significantly involved in the β1-α2 communication. Finally, when all the subunits are unligated, the communication between proximal sites seems to dominate the scene.

Regardless the ligation state, it turns out that α1-β2 (α2-β1) communication is mostly mediated by the FG corners (see Figures 8.3 and 8.4 for the illustration of the secondary structure elements) of both chains and the B, C helices of the α subunit (as a consequence of the marginality of β's distal pocket role in heme-heme allosteric interaction, these helices are less important in this subunit), as shown in Figures 9.2-9.4. More in detail, H87α's response to CO unbinding is transmitted to F103β through αFG-βFG correlated motion, while the dynamical response of α's distal pocket is communicated through the αB-αC-βFG path down to F103β.

When β1 loses the ligand, also the αC helix becomes crucial to convey the allosteric information to both heme's pockets, differentiating the β-to-α and α-to-β paths.

Table 9.2. End-points of the best 100 signalling paths.

System	α1-β1	α1-β2	β1-α2	β1-β2
HbCO	H58α1 (100) H92β1 (68) F103β1 (32)	H87α1 (100) F103β2 (68) H63β2 (13) H92β2 (19)	F103β1 (96) H92β1 (4) H87α2 (100)	F103β1 (85) H92β1 (15) F103β2 (68) H92β2 (32)
α1-unligated	H58α1 (58) H87α1 (34) V62α1 (8) H63β1 (71) F103β1 (29)	H87α1 (70) H58α1 (25) V62α1 (5) F103β2 (100)	N.C.	N.C.
β1-unligated	H63β1 (51) F103β1 (49) H87α1 (58) V62α1 (26) H58α1 (16)	N.C.	F103β1 (89) H92β1 (8) H63β1 (3) H87α2 (85) H58α2 (14) V62α2 (1)	F103β1 (50) H92β1 (50) F103β2 (80) H92β2 (20)
full-unligated	V62α1 (97) H87α1 (3) H63β1 (100)	H87α1 (100) H92β2 (81) H63β2 (9) F103β2 (8) V67β2 (2)	H92β1 (74) F103β1 (26) H87α2 (97) H58α2 (3)	F103β1 (60) H92β1 (40) H92β2 (95) F103β2 (5)

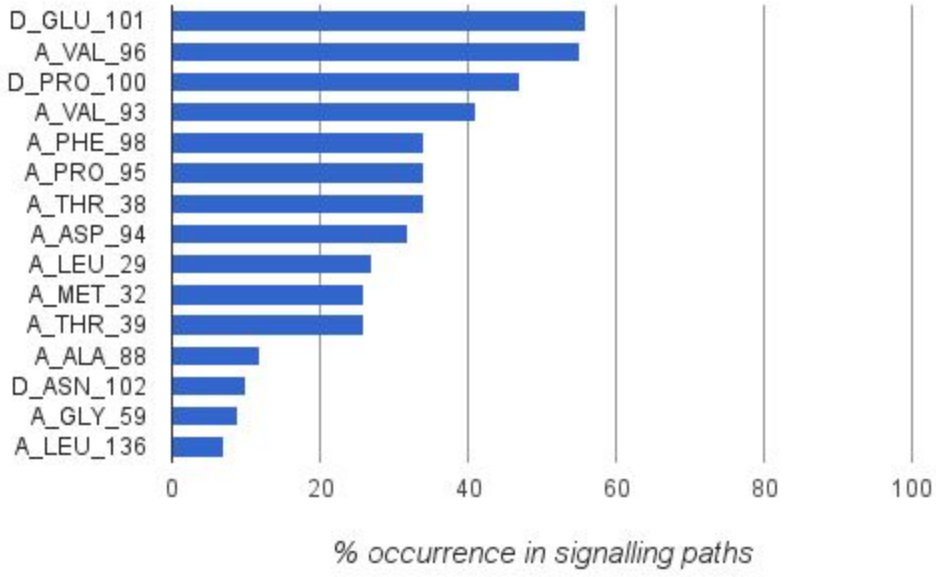


Figure 9.2. Residues that are more involved in the α1-β2 (AD) signalling in α1-uniligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

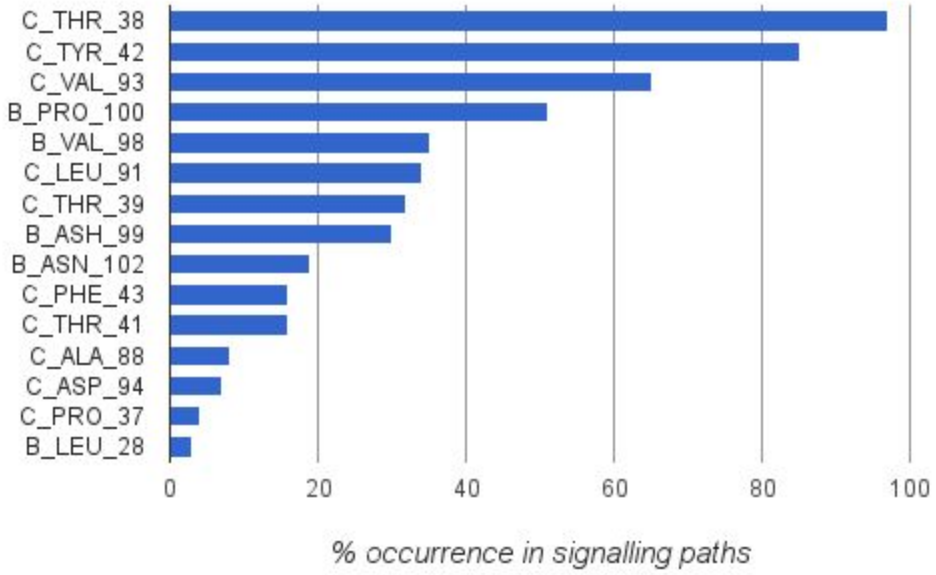


Figure 9.3. Residues that are more involved in the β1-α2 (BC) signalling in β1-uniligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

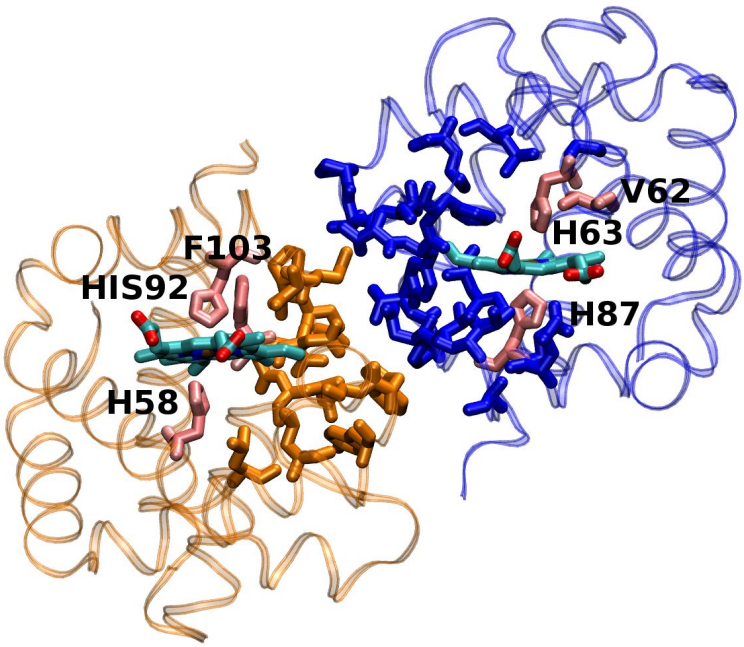


Figure 9.4. Illustration of the residues that are more involved in the α1β2 (AD) and β1α2 (BC) signalling in α1- and β1-uniligated Hb (directly obtained from Figures 9.2 and 9.3). The α chain is shown in blue, the β chain in yellow. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

Once α1 and β2 (α2 and β1) lose CO, the ligated β chain is most likely influenced by the unligated one through a β1-β2 signalling, which come next in importance (Table 9.1). In such information exchange channel, F103β and H92β are the most frequent signal source/sinks (Table 9.2). Therefore, conversely to α1-β2 (α2-β1), H92β has a significant role in this allosteric communication route, while the distal pocket does not influence such signalling path (which seems reasonable looking at the β1-β2 relative orientation, Figure 9.6). The N-terminal residue seems to be crucial for β1-β2 signalling, along with the rest of the H and F helices (Figures 9.5 and 9.6).

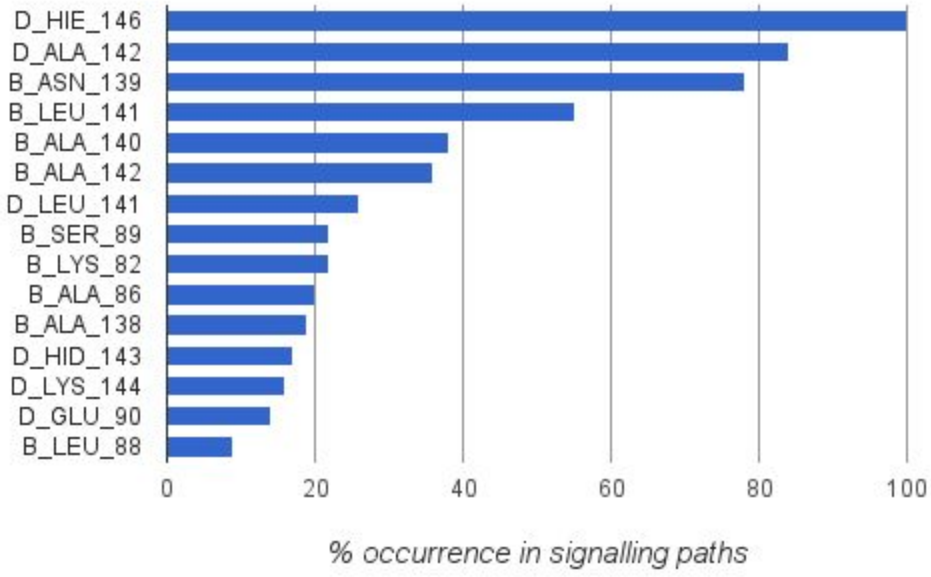


Figure 9.5. Residues that are more involved in the β1-β2 (BD) signalling in β1-unligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

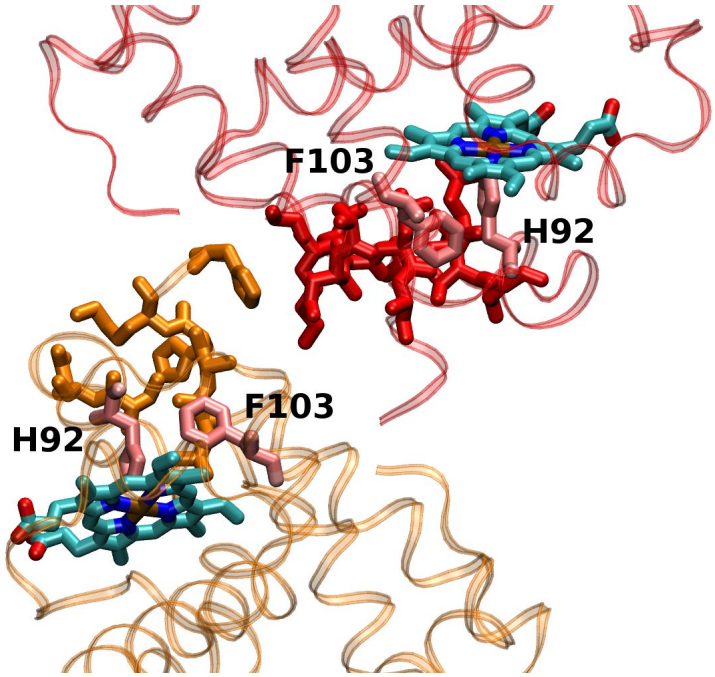


Figure 9.6. Illustration of the residues that are more involved in the β1β2 (AD) signalling in β1-unligated (directly obtained from Figure 9.5). The β1 chain is shown in red, the β2 chain in yellow. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

The intra-dimer signalling paths, whose impact on cooperativity is comparable to that of the β1-β2 interaction, get more variegate upon CO unbinding from either the α1 or the β1 chain (Table 9.2). H87α takes an important role as signal source/sink, which is instead monopolized by the distal His when CO is bound. Therefore, Fe's out-of-plane motion is successfully transmitted to the other chain. Moreover, H63β starts acting as a signal source/sink as well, propagating/inducing dynamical changes from/in the distal pocket.

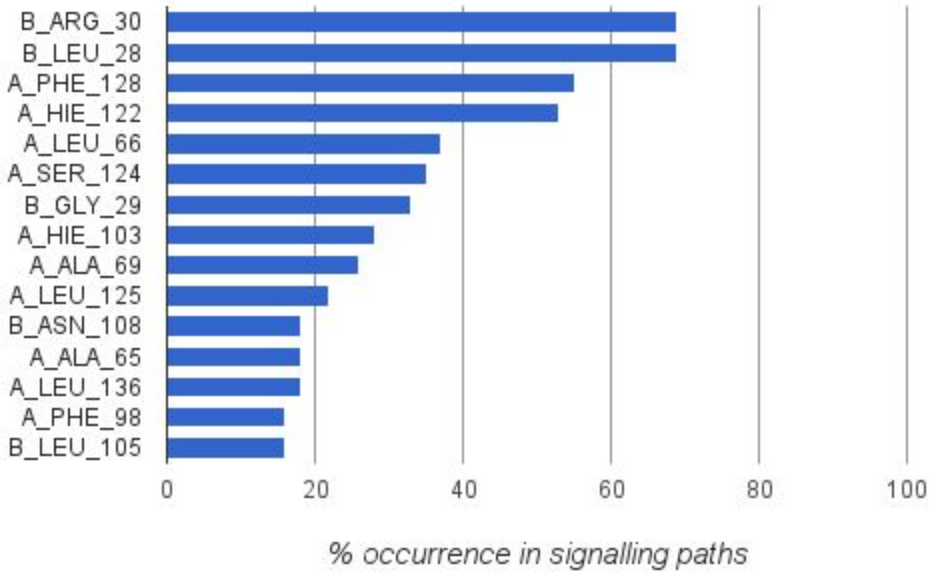


Figure 9.7. Residues that are more involved in the α1-β1 (AB) signalling in α1-unligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

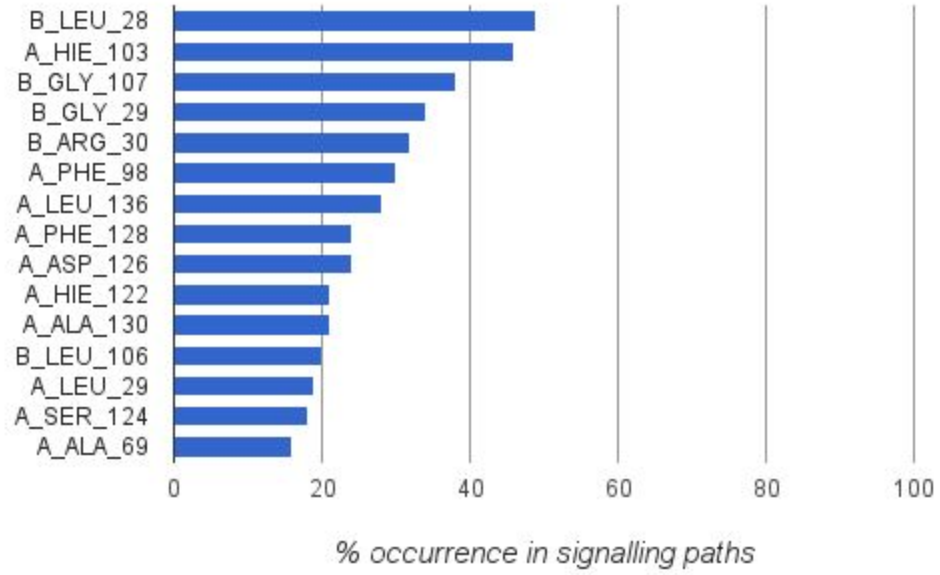


Figure 9.8. Residues that are more involved in the β1-α1 (BA) signalling in β1-unligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

Despite the richer number of signalling options, intra-dimer communication drops in the α1- and β1-unligated variants. Possibly, such drop might help to focus inter-subunit communication toward the most efficient allosteric channel, namely α1-β2 (α2-β1), although further studies are necessary to confirm it. However, the intra-dimer communication channel gains importance when all the subunits are unligated (Table 9.1), although it is still not as efficient as inter-dimer signalling is. Intra-dimer signalling is largely mediated by the following helices: αG, αE, αH, βG, βB (Figures 9.7-9.9). In particular, the distal pockets are mediated through the αE-αH-βB path, while H87α and F103β are linked by a αG-βG medium.

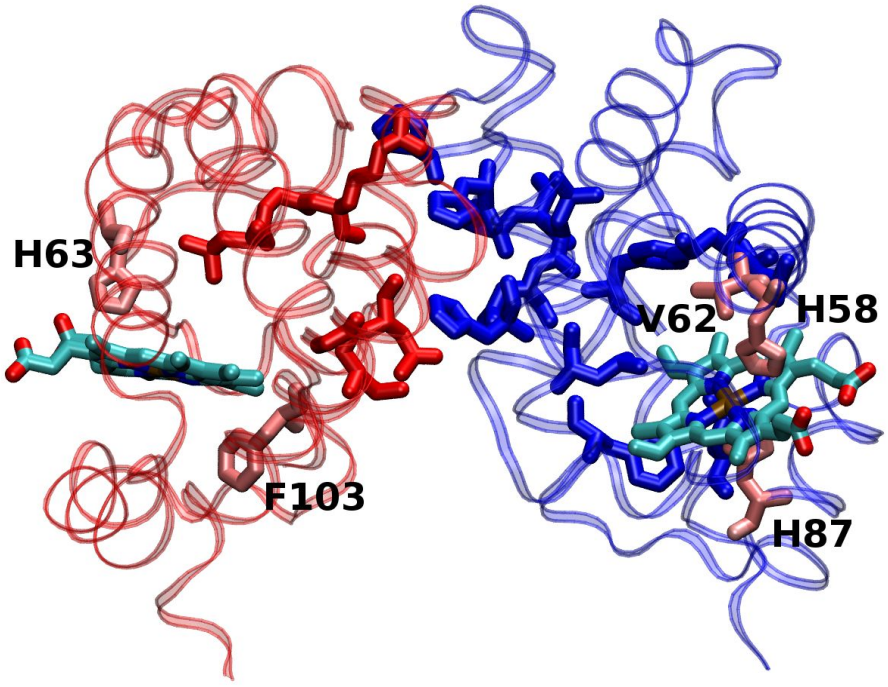


Figure 9.9. Illustration of the residues that are more involved in the α1β1 (AD) signalling in α1-uniligated (directly obtained from Figures 9.7 and 9.8). The α1 chain is shown in blue, the β1 chain in red. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

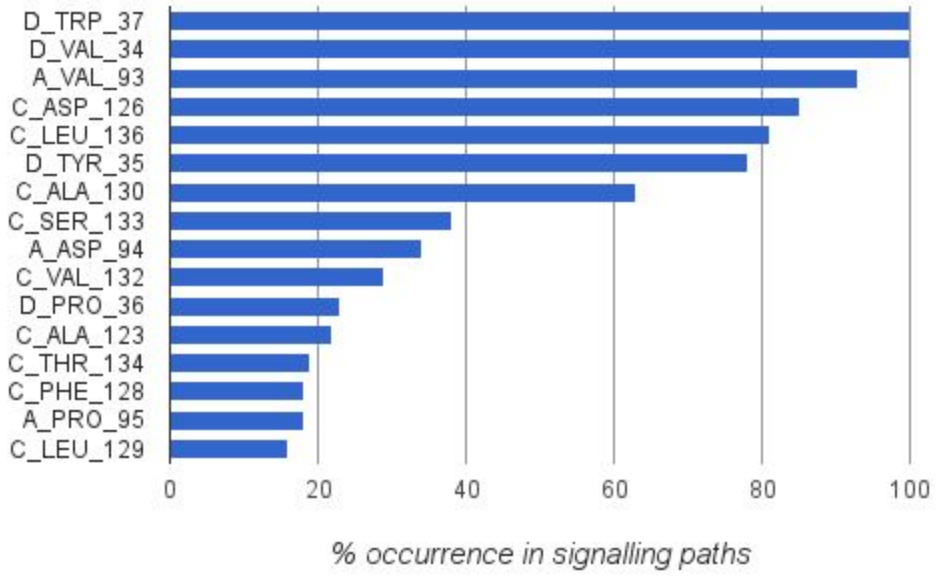


Figure 9.10. Residues that are more involved in the α1-α2 (AC) signalling in α1-unligated. Bin heights represent the frequency at which such residues take part to the signalling process. Figure prepared with Google Sheets (https://goo.gl/APF6tL) and modified with GIMP (https://www.gimp.org/).

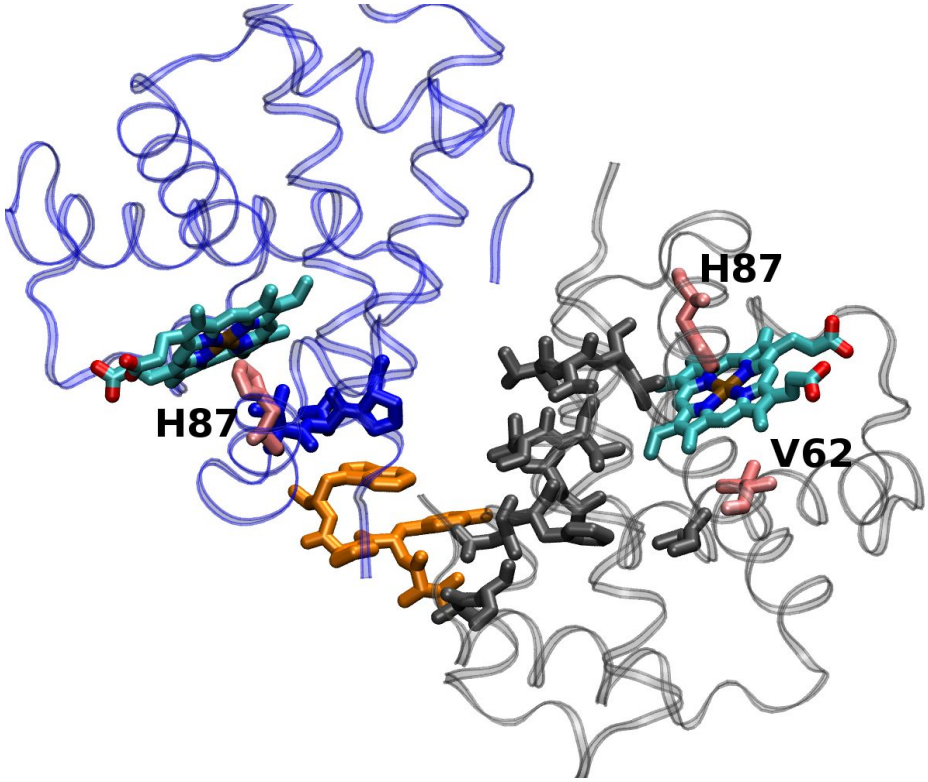


Figure 9.11. Illustration of the residues that are more involved in the α1α2 (AD) signalling in α1-uniligated (directly obtained from Figure 9.10). The α1 chain is shown in blue, the α2 chain in black and the β2 chain in orange. Figure prepared with VMD (Humphrey et al., 1996) and modified with GIMP (https://www.gimp.org/).

Finally, α1-α2 communication is weak (Table 9.1) and mediated by a β chain (Figures 9.10, 9.11), therefore its contribution to cooperativity should be marginal. The proximal His of the ligand-free chain communicates heme's response to the proximal and distal pockets of the other subunit through the α1FG-β2C-α2H and α1FG-β2C-α2H-α2E routes respectively. The latter, and therefore the involvement of the distal pocket, is less probable though.

9.3 Closure

According to the results obtained, whether or not the first ligand unbinding event occurs in an α or β chain, the preferential route for allosteric signalling should be the inter-dimer α-β correlation path. This involves both the proximal and distal pockets of the α chain, while the β

heme mostly uses F103 to communicate with the rest of the protein. In this signalling route, the following protein fragments act as bridges: αFG-βFG and αB-αC-βFG.

The β-β and intra-dimer α-β paths seem to be less important. The former connects F103 and H92 of both chains to each other through the F and H helices, the latter serves as main communication channel for β's distal pocket. In this last case, these are the bridging moieties: αE-αH-βB and αG-βG. It should be noted that intra-dimer α-β signalling is more important in fully ligated and unligated forms than in the intermediate ligation states. Finally, α-α communication is not expected to influence cooperativity.

Collecting all the results, it can be seen that, for both chains, the following secondary structure elements are never involved in inter-subunit communication paths: the A helix and the first residues of the B helix, the EF loops and the residues within helices C and E. Therefore, their modifications for design purposes (stability in water) should not impact Hb's cooperativity.

REFERENCES:

Alcantara RE., Xu C., Spiro TG., Guallar V. 2007. A quantum-chemical picture of hemoglobin affinity. *Proceedings of the National Academy of Sciences of the United States of America* 104:18451–18455.

Amadei A., Andrea A., Linssen ABM., Berendsen HJC. 1993. Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics* 17:412–425.

Bowers K., Kevin B., Edmond C., Huafeng X., Ron D., Michael E., Brent G., John K., Istvan K., Mark M., Federico S., John S., Yibing S., David S. 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC'06)*. DOI: 10.1109/sc.2006.54.

Dijkstra EW. 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* 1:269–271.

Essmann U., Ulrich E., Lalith P., Berkowitz ML., Tom D., Hsing L., Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics* 103:8577.

Fan JS., Yang D., Choy WY. 2013. Refined solution structure of Human Adult Hemoglobin in the Carbonmonoxy Form. DOI: 10.2210/pdb2m6z/pdb.

Gordon JC., Myers JB., Folta T., Shoja V., Heath LS., Onufriev A. 2005. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* 33:W368–71.

Hayward S., Steven H., Groot BL. 2008. Normal Modes and Essential Dynamics. In: *Methods in Molecular Biology*. 89–106.

Hub JS., Kubitzki MB., de Groot BL. 2010. Spontaneous quaternary and tertiary T-R transitions of human hemoglobin in molecular dynamics simulation. *PLoS computational biology* 6:e1000774.

Humphrey W., William H., Andrew D., Klaus S. 1996. VMD: Visual molecular dynamics. *Journal of molecular graphics* 14:33–38.

Jones EM., Monza E., Balakrishnan G., Blouin GC., Mak PJ., Zhu Q., Kincaid JR., Guallar V., Spiro TG. 2014. Differential control of heme reactivity in alpha and beta subunits of hemoglobin: a combined Raman spectroscopic and computational study. *Journal of the American Chemical Society* 136:10325–10339.

Kaminski GA., Friesner RA., Julian T-R., Jorgensen WL. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *The journal of physical chemistry. B* 105:6474–6487.

Martyna GJ., Tobias DJ., Klein ML. 1994. Constant pressure molecular dynamics algorithms. *The Journal of*

chemical physics 101:4177.
Nosé S., Shuichi N. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics* 81:511.
Olsson MHM., Søndergaard CR., Rostkowski M., Jensen JH. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation* 7:525–537.
Sastry GM., Adzhigirey M., Day T., Annabhimoju R., Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* 27:221–234.
Toukan K., Kahled T., Aneesur R. 1985. Molecular-dynamics study of atomic motions in water. *Physical Review B: Condensed Matter and Materials Physics* 31:2643–2648.
Tuckerman M., Berne BJ., Martyna GJ. 1992. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics* 97:1990.
Unzai S., Eich R., Shibayama N., Olson JS., Morimoto H. 1998. Rate Constants for O2 and CO Binding to the and Subunits within the R and T States of Human Hemoglobin. *The Journal of biological chemistry* 273:23150–23159.
Van Wart AT., Jacob D., Lane V., Amaro RE. 2014. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *Journal of chemical theory and computation* 10:511–517.

Conclusions

10 Laccases: conclusions

- Quantum mechanics/molecular mechanics (QM/MM) spin densities can efficiently score the electron transfer (ET) driving force in laccases. Quantitative accuracy can be obtained for similar substrates if water molecules with an active role in catalysis are retained. The PELE+QM/MM approach underlined that substrate binding is fundamental in laccases function and it can be fine tuned to improve catalysis. While stabilizing binding poses with optimized catalytic contacts can improve electronic coupling (T_{DA}), enhancing the electrostatic stabilization of the oxidized substrate can push the ET driving force toward lower values, increasing its match with the negative image of the reorganization energy.
- The PELE+QM/MM binding-focused strategy has been employed to increase the oxidation of aniline at acidic pH, improving the production of polyaniline. The engineering efforts showed that the T1 site is a promiscuous template to build specificity toward a target substrate. This can be achieved through the design of stabilizing interactions in highly reactive enzyme-substrate conformations. Laccases promiscuity is at the basis of the proved success of mutant repurposing. In the future, repurposing might be carried out computationally and subsequently validated experimentally, transforming laccases into more efficient multitasking enzymes.
- Laccases design proved to be a multidimensional problem; both the driving force and T_{DA} should be taken into account. The collection of experimental kinetic data to benchmark computational strategies, aiming at finding parameters that correlate with the kinetic constant, will be crucial to improve the predictive power of laccases reactivity in the future.
- A computational methodology that combines classical conformational sampling and quick empirical T_{DA} estimations was capable of producing results in agreement with the experiment for a hybrid laccase/$[Ru(bpy)_3]^{2+}$ photocatalyst. Knowing that the fast electron back-transfer is the major problem the hybrid laccase/$[Ru(bpy)_3]^{2+}$, it was hypothesized to directly convey the electrons to the trinuclear cluster of copper ions (TNC), skipping the T1 copper. Several mutants with significantly improved ET to the TNC were designed and are being validated in the lab.

11 Hemoglobin: conclusions

- The tertiary two-state (TTS) model was supported by the simulation (at the best of my knowledge, for the first time). The differences between r and t states are located in close proximity to the heme and they are distinct from the tertiary structures found in the oxy- and deoxy-hemoglobin crystal structure. In the α chain, heme's reactivity is modulated by both ring (out-of-plane distortion) and proximal strain (Fe-His elongation and tilting), while only the former is relevant in the β subunit. The different control of heme's reactivity in hemoglobin's subunits is likely dictated by their distinct tertiary dynamics and of their diverse packing between the E and F helices.
- The preferential route for allosteric signalling should be the inter-dimer α-β correlation path. This involves both the proximal and distal pockets of the α chain, while the β heme mostly uses F103 to communicate with the rest of the protein. In this signalling route, the following protein fragments act as bridges: the FG of the β chain and either the FG corner or the B and C helices of the α chain. The β-β and intra-dimer α-β paths seem to be less important, while α-α communication should not influence cooperativity. For both chains, these secondary structure elements are never involved in inter-subunit communication paths: the A helix and the first residues of the B helix, the EF loops and the residues within helices C and E. Their modifications for design purposes (stability in water) should not impact cooperativity.

Appendices

Appendix 1 - PELE protein-ligand sampling input file

This is an example of PELE protein-ligand sampling analogous to those used to produce part of the work of chapters 5 and 6. PELE can be used for free at https://pele.bsc.es/pele.wt.

```
## POXA1b-2,4-dabsa sampling

file datadir /home/bsc72/bsc72328/plop/data
file log dyn_24dabsa.log
energy params solvent vdgbnp
energy params ionic 0.15
load pdb poxa1b-24dabsa_fixed.pdb waters no ions no het yes

constraint atom A:597:CU__ current 200 0.0 &
atom A:451:_SG_ current 200 0.0 &
atom A:451:_CB_ current 200 0.0 &
atom A:456:_ND1 current 200 0.0 &
atom A:456:_CB_ current 200 0.0 &
atom A:456:_CG_ current 200 0.0 &
atom A:394:_ND1 current 200 0.0 &
atom A:394:_CB_ current 200 0.0 &
atom A:394:_CG_ current 200 0.0 &
atom A:453:_CD1 current 200 0.0 &
atom A:453:_CG1 current 200 0.0 &
atom A:461:_CE1 current 200 0.0 &
atom A:461:_CD1 current 200 0.0 &
atom A:461:_CG_ current 200 0.0 &
atom A:598:CU__ current 200 0.0 &
atom A:599:CU__ current 200 0.0 &
atom A:600:CU__ current 200 0.0 &
atom A:399:_NE2 current 200 0.0 &
atom A:111:_NE2 current 200 0.0 &
atom A:450:_NE2 current 200 0.0 &
atom A:452:_NE2 current 200 0.0 &
atom A:109:_NE2 current 200 0.0 &
atom A:66:_ND1 current 200 0.0 &
atom A:397:_NE2 current 200 0.0 &
atom A:64:_NE2 current 200 0.0 &

pele &
        het L:1 &
        pdbmodel yes &
        top_side 10 &
        init_min no &
        task &
                show bind_ene 1 &
                show SASA 1 &
                if random 1 gt 0.5 then rot_r 0.02 else rot_r 0.25 endif &
                if random 2 gt 0.5 then tra_r 2.0 else tra_r 0.75 endif &
```

```
            spawn atom 1 A:597:CU__ within 15.0 &
            exit steps gt 2000 &
end_task &
temp 1000 &
tries 1 &
anmfreq 1 &
spfreq 1 &
mifreq 1 &
wrfreq 1 &
omit_sp A:394 A:394 &
omit_sp A:451 A:451 &
omit_sp A:456 A:456 &
 omit_sp A:453 A:453 &
omit_sp A:461 A:461 &
side &
            randomize no &
            verbose no &
            failsafe no &
sideend &
path traj_24dabsa_ &
mirad 20  &
sprad 10  &
min &
            rmsg 0.02 &
            nbup yes &
            gbup yes &
            alphaup yes &
minimend &
caconst 0.05 &
lcom_con 0.1 &
rem_bulk_mov 3 &
anm_eig_freq 100000 &
anm_altm_freq 1 &
anm_altm_type 3 &
lanmanm neig 6 &
lanmanm ualig YES &
lanmanm move_ca 0.2 &
lanmanm mix_modes 0.75 &
lanmmin &
            mxitn 100 &
            iter 1 &
            rmsg 0.02 &
            nbup yes &
            alphaup no &
minimend
```

Appendix 2 - PELE design input file

This is the PELE design input file used to produce part of the work of chapter 6. PELE can be used for free at https://pele.bsc.es/pele.wt.

```
## POXA1b-2,4-dabsa design

file datadir /home/bsc72/bsc72328/plop/data
file log dyn.log
energy params solvent vdgbnp
energy params ionic 0.15
load pdb candidate_refined.pdb waters no ions no het yes

constraint atom A:597:CU__ current 200 0.0 &
atom A:451:_SG_ current 200 0.0 &
atom A:451:_CB_ current 100 0.0 &
atom A:456:_ND1 current 200 0.0 &
atom A:456:_CB_ current 100 0.0 &
atom A:456:_CG_ current 100 0.0 &
atom A:394:_ND1 current 200 0.0 &
atom A:394:_CB_ current 100 0.0 &
atom A:394:_CG_ current 100 0.0 &
atom A:453:_CD1 current 100 0.0 &
atom A:453:_CG1 current 100 0.0 &
atom A:461:_CE1 current 100 0.0 &
atom A:461:_CD1 current 100 0.0 &
atom A:461:_CG_ current 100 0.0 &
atom A:598:CU__ current 200 0.0 &
atom A:599:CU__ current 200 0.0 &
atom A:600:CU__ current 200 0.0 &
atom A:399:_NE2 current 100 0.0 &
atom A:111:_NE2 current 100 0.0 &
atom A:450:_NE2 current 100 0.0 &
atom A:452:_NE2 current 100 0.0 &
atom A:109:_NE2 current 100 0.0 &
atom A:66:_ND1 current 100 0.0 &
atom A:397:_NE2 current 100 0.0 &
atom A:64:_NE2 current 100 0.0 &

pele &
        het L:1 &
        pdbmodel yes &
        top_side 10 &
        init_min no &
        task &
                show bind_ene 1 &
                show SASA 1 &
                spawn atom 1 A:597:CU__ within 15.0 &
                exit steps gt 2000 &
```

```
end_task &
rot_r 0.1 &
 tra_r 0.05 &
temp 1000 &
tries 1 &
anmfreq 1 &
spfreq 1 &
mifreq 1 &
wrfreq 1 &
omit_sp A:394 A:394 &
omit_sp A:451 A:451 &
omit_sp A:456 A:456 &
omit_sp A:453 A:453 &
omit_sp A:461 A:461 &
side &
        randomize no &
        verbose no &
        failsafe no &
sideend &
path traj_24dabsa_ &
mirad 20  &
sprad 10  &
min &
        rmsg 0.02 &
        nbup yes &
        gbup yes &
        alphaup yes &
minimend &
caconst 0.05 &
lcom_con 0.1 &
rem_bulk_mov 3 &
anm_eig_freq 100000 &
anm_altm_freq 1 &
anm_altm_type 3 &
lanmanm neig 6 &
lanmanm ualig YES &
lanmanm move_ca 0.5 &
lanmanm mix_modes 0.75 &
lanmanm omit_sp A:597 A:600 &
lanmmin &
        mxitn 100 &
        iter 1 &
        rmsg 0.02 &
        nbup yes &
        alphaup no &
minimend
```

Appendix 3 - PELE protein-ligand local refinement input file

This is the PELE protein-ligand local refinement input file used to produce part of the work of chapter 7. PELE can be used for free at https://pele.bsc.es/pele.wt.

```
## LAC3+RU local refinement

file datadir /home/bsc72/bsc72328/plop/data
file log Kdyn.log
energy params solvent vdgbnp
energy params ionic 0.1
load pdb final.pdb ions no waters no het yes
load native final.pdb

constraint atom A:502:CU__ current 200 0.0 &
atom A:503:CU__ current 200 0.0 &
atom A:504:CU__ current 200 0.0 &
atom A:505:CU__ current 200 0.0 &
atom A:453:_SG_ current 200 0.0 &
atom A:453:_CB_ current 200 0.0 &
atom A:458:_ND1 current 200 0.0 &
atom A:458:_CE1 current 200 0.0 &
atom A:458:_CG_ current 200 0.0 &
atom A:397:_ND1 current 200 0.0 &
atom A:397:_CE1 current 200 0.0 &
atom A:397:_CG_ current 200 0.0 &
atom A:455:_CD1 current 200 0.0 &
atom A:455:_CG1 current 200 0.0 &
atom A:455:_CB_ current 200 0.0 &
atom A:463:_CD2 current 200 0.0 &
atom A:463:_CE2 current 200 0.0 &
atom A:463:_CG_ current 200 0.0 &
atom A:66:_ND1 current 200 0.0 &
atom A:66:_CE1 current 200 0.0 &
atom A:66:_CG_ current 200 0.0 &
atom A:64:_NE2 current 200 0.0 &
atom A:64:_CE1 current 200 0.0 &
atom A:64:_CD2 current 200 0.0 &
atom A:109:_NE2 current 200 0.0 &
atom A:109:_CE1 current 200 0.0 &
atom A:109:_CD2 current 200 0.0 &
atom A:111:_NE2 current 200 0.0 &
atom A:111:_CE1 current 200 0.0 &
atom A:111:_CD2 current 200 0.0 &
atom A:400:_NE2 current 200 0.0 &
atom A:400:_CE1 current 200 0.0 &
atom A:400:_CD2 current 200 0.0 &
atom A:402:_NE2 current 200 0.0 &
atom A:402:_CE1 current 200 0.0 &
```

```
atom A:402:_CD2 current 200 0.0 &
atom A:452:_NE2 current 200 0.0 &
atom A:452:_CE1 current 200 0.0 &
atom A:452:_CD2 current 200 0.0 &
atom A:454:_NE2 current 200 0.0 &
atom A:454:_CE1 current 200 0.0 &
atom A:454:_CD2 current 200 0.0 &

pele &
        het L:999 &
        pdbmodel yes &
        init_min no &
        steric_tr 500 &
        steered 1 &
        waitfor 2 &
        task &
                show bind_ene 1 &
                show adist A:502:CU__ L:999:RU1_ &
                show adist A:503:CU__ L:999:RU1_ &
                show adist A:504:CU__ L:999:RU1_ &
                show adist A:505:CU__ L:999:RU1_ &
                show SASA 1 &
                show rmsd_sys calpha &
                show rmsd_lig 1 heavy 0 &
                spawn rmsd_lig 1 heavy 0 within 10 &
                if random 1 gt 0.5 then tra_r 0.4  else tra_r 0.2  endif &
                if random 2 gt 0.5 then rot_r 0.25  else rot_r 0.05 endif &
                exit steps gt 200 &
        end_task &
        tries 40 &
        temp 2000 &
        anmfreq 1 &
        spfreq 1 &
        mifreq 1 &
        wrfreq 1 &
        sprad 5.0 &
        lcom_con 0.1 &
        side &
                randomize no &
                iter 1 &
                verbose no &
                failsafe no &
        sideend &
        path traj_ &
        min &
                rmsg 0.04 &
                nbup yes &
                gbup yes &
                alphaup yes &
        minimend &
        mirad 10 &
        caconst 0.5 &
        rem_bulk_mov 3 &
        anm_eig_freq 100000 &
```

```
anm_altm_freq 3 &
anm_altm_type 3 &
lanmanm move_ca 0.5 &
lanmanm mix_modes 0.75 &
lanmmin &
        mxitn 100 &
        iter 1 &
        rmsg 0.04 &
        nbup yes &
        alphaup no &
minimend
```

Appendix 4 - PELE design input file (no ligand)

This is the PELE design input file used to produce part of the work of chapter 7. PELE can be used for free at https://pele.bsc.es/pele.wt.

```
## LAC3+RU local refinement

file datadir /home/bsc72/bsc72328/plop/data
file log Kdyn.log
energy params solvent vdgbnp
energy params ionic 0.1
load pdb final.pdb ions no waters no het yes

constraint atom A:502:CU__ current 200 0.0 &
atom A:503:CU__ current 200 0.0 &
atom A:504:CU__ current 200 0.0 &
atom A:505:CU__ current 200 0.0 &
atom A:453:_SG_ current 200 0.0 &
atom A:453:_CB_ current 200 0.0 &
atom A:458:_ND1 current 200 0.0 &
atom A:458:_CE1 current 200 0.0 &
atom A:458:_CG_ current 200 0.0 &
atom A:397:_ND1 current 200 0.0 &
atom A:397:_CE1 current 200 0.0 &
atom A:397:_CG_ current 200 0.0 &
atom A:455:_CD1 current 200 0.0 &
atom A:455:_CG1 current 200 0.0 &
atom A:455:_CB_ current 200 0.0 &
atom A:463:_CD2 current 200 0.0 &
atom A:463:_CE2 current 200 0.0 &
atom A:463:_CG_ current 200 0.0 &
atom A:66:_ND1 current 200 0.0 &
atom A:66:_CE1 current 200 0.0 &
atom A:66:_CG_ current 200 0.0 &
atom A:64:_NE2 current 200 0.0 &
atom A:64:_CE1 current 200 0.0 &
atom A:64:_CD2 current 200 0.0 &
atom A:109:_NE2 current 200 0.0 &
atom A:109:_CE1 current 200 0.0 &
atom A:109:_CD2 current 200 0.0 &
atom A:111:_NE2 current 200 0.0 &
atom A:111:_CE1 current 200 0.0 &
atom A:111:_CD2 current 200 0.0 &
atom A:400:_NE2 current 200 0.0 &
atom A:400:_CE1 current 200 0.0 &
atom A:400:_CD2 current 200 0.0 &
atom A:402:_NE2 current 200 0.0 &
atom A:402:_CE1 current 200 0.0 &
atom A:402:_CD2 current 200 0.0 &
```

```
atom A:452:_NE2 current 200 0.0 &
atom A:452:_CE1 current 200 0.0 &
atom A:452:_CD2 current 200 0.0 &
atom A:454:_NE2 current 200 0.0 &
atom A:454:_CE1 current 200 0.0 &
atom A:454:_CD2 current 200 0.0 &

pele &
        top_side 25 &
        pdbmodel yes &
        init_min no &
        task &
                exit steps gt 200 &
        end_task &
        temp 2000 &
        anmfreq 1 &
        spfreq 1 &
        mifreq 1 &
        wrfreq 1 &
        side &
                randomize no &
                iter 1 &
                verbose no &
                failsafe no &
        sideend &
        path traj_ &
        min &
                rmsg 0.04 &
                nbup yes &
                gbup yes &
                alphaup yes &
        minimend &
        caconst 0.5 &
        rem_bulk_mov 3 &
        anm_eig_freq 100000 &
        anm_altm_freq 3 &
        anm_altm_type 3 &
        lanmanm move_ca 0.75 &
        lanmanm mix_modes 0.75 &
        lanmmin &
                mxitn 100 &
                iter 1 &
                rmsg 0.04 &
                nbup yes &
                alphaup no &
        minimend
```

Appendix 5 - PELE protein sampling input file

This is an example of PELE protein sampling input file used to produce part of the work of chapter 8. PELE can be used for free at https://pele.bsc.es/pele.wt.

```
## unligated beta chain

file datadir /home/Life/victorg/plop/data
file log dyn_beta.log
energy params solvent vdgbnp
energy params ionic 0.3
load pdb beta_fixed.pdb waters no exp no seqres no het yes &

constraint atom D:147:FE__ atom D:92:_NE2 200 2.28 &
atom B:147:FE__ atom B:92:_NE2 200 2.12 &
atom C:142:FE__ atom C:87:_NE2 200 2.12 &
atom A:142:FE__ atom A:87:_NE2 200 2.12 &
atom B:147:FE__ atom B:148:_O1_ 200 1.76 &
atom C:142:FE__ atom C:143:_O1_ 200 1.76 &
atom A:142:FE__ atom A:143:_O1_ 200 1.76 &

pele &
        het D:142 &
        pdbmodel yes &
        top_side 25 &
        init_min no &
        task &
                exit steps gt 200 &
        end_task &
        anmfreq 1 &
        tra_r 0.0 &
        rot_r 0.0 &
        temp 1000 &
        ldfreq 10000 &
        spfreq 1 &
        mifreq 1 &
        wrfreq 1 &
        side &
                randomize no &
                verbose no &
                failsafe no &
        sideend &
        path traj_beta_ &
        mirad 25  &
        min &
                rmsg 0.02 &
                nbup yes &
                gbup yes &
                alphaup yes &
```

```
minimend &
caconst 10.0 &
rem_bulk_mov 3 &
anm_eig_freq 100000 &
anm_altm_freq 3 &
anm_altm_type 3 &
lanmanm neig 6 &
lanmanm ualig YES &
lanmanm omit_no C:1 C:141 &
lanmanm omit_no C:142 C:143 &
lanmanm omit_no B:1 B:146 &
lanmanm omit_no B:147 B:148 &
lanmanm omit_no A:1 A:141 &
lanmanm omit_no A:142 A:143 &
lanmanm move_ca 1.0 &
lanmanm mix_modes 0.5 &
lanmmin &
        mxitn 100 &
        iter 1 &
        rmsg 0.02 &
        nbup yes &
        alphaup no &
minimend
```

Appendix 6 - POXA1b variants with improved PELE binding energy in each *in silico* evolution round

Table A6.1. Variants from round 1 showing improved descriptor values.

Protein variant	Binding energy (kcal/mol)	SASA
Wild type	-8	0.19
A336R	-22	0.10
A336Q	-13	0.17
A336K	-15	0.14
G161R	-16	0.12
G161Q	-20	0.11
G161K	-16	0.11
G161M	-15	0.09
G161S	-19	0.13
G161Y	-14	0.10
G392R	-18	0.10
G392F	-14	0.04
G392W	-15	0.02
G392Y	-13	0.04
L511R	-18	0.13
L511N	-18	0.18
L511Q	-17	0.17
L511H	-14	0.16
L511K	-14	0.15
L511T	-17	0.17

F331R	-16	0.12
F331Q	-15	0.17
F331K	-15	0.17
F331Y	-19	0.15
P163R	-17	0.18
P163N	-20	0.19
P163Q	-16	0.14
P163K	-14	0.13
V162N	-20	0.10
V162C	-18	0.09
V162H	**-20**	**0.09**
V162S	**-16**	**0.17**
V162T	-18	0.21

Table A6.2. Variants from round 2 of the H162 branch showing improved descriptor values.

Protein variant	Binding energy (kcal/mol)	SASA
V162H_A336N	-28	0.09
V162H_A336K	-26	0.10
V162H_L511R	-28	0.08
V162H_L511Q	-27	0.10
V162H_F331Y	**-29**	**0.09**

Table A6.3. Variants from round 2 of the S162 branch showing improved descriptor values.

Protein variant	Binding energy (kcal/mol)	SASA
V162S_A336R	-22	0.16
V162S_A336Q	-21	0.16
V162S_A336K	-22	0.17
V162S_G161R	-22	0.10
V162S_G392R	-22	0.05
V162S_G392Q	-22	0.12
V162S_G392F	-23	0.03
V162S_G392W	-23	0.03
V162S_G392Y	-23	0.03
V162S_L511R	-25	0.14
V162S_L511N	-25	0.20
V162S_L511Q	-24	0.17
V162S_L511T	-24	0.18
V162S_F331R	-23	0.13
V162S_F331K	-23	0.16
V162S_F331Y	**-26**	**0.14**
V162S_P163R	-24	0.14
V162S_P163N	-28	0.18
V162S_P163K	-23	0.14
V162S_P163M	-21	0.17

Table A6.4. Variants from round 3 of the H162 branch showing improved descriptor values.

Protein variant	Binding energy (kcal/mol)	SASA
V162H_F331Y_A336R	-36	0.08
V162H_F331Y_A336N	**-35**	**0.07**
V162H_F331Y_G392R	-35	0.10
V162H_F331Y_L511R	-35	0.08
V162H_F331Y_L511N	-36	0.10
V162H_F331Y_P163K	-35	0.03
V162H_F331Y_P510R	-35	0.08

Table A6.5. Variants from round 3 of the S162 branch showing improved descriptor values.

Protein variant	Binding energy (kcal/mol)	SASA
V162S_F331Y_A336N	**-31**	**0.15**
V162S_F331Y_G161K	-34	0.15
V162S_F331Y_P163R	-33	0.10
V162S_F331Y_P163N	-35	0.13
V162S_F331Y_P163Q	-32	0.13
V162S_F331Y_P163H	-32	0.10
V162S_F331Y_P163K	-33	0.09

Appendix 7 - RMSD plots of molecular dynamics simulations from Chapter 7

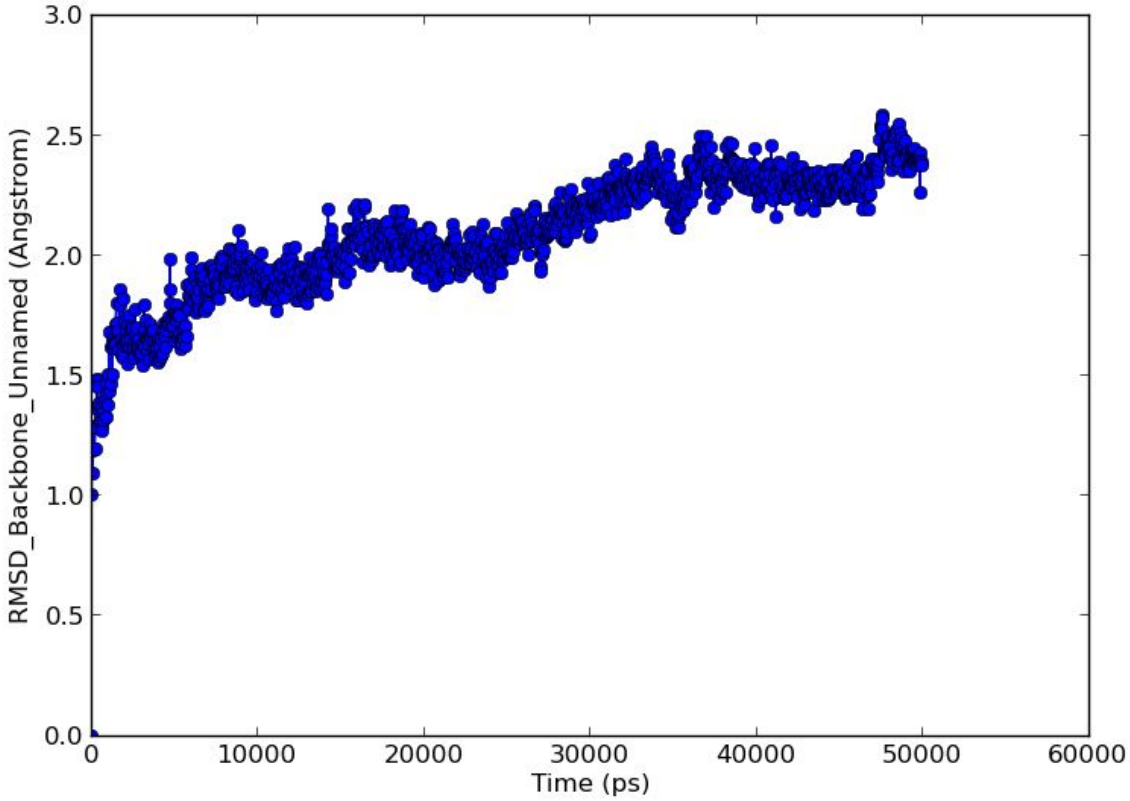


Figure A7.1. Protein backbone RMSD profile of UNIK161-Ru MD run. Figure prepared with Desmond (Bowers et al., 2006).

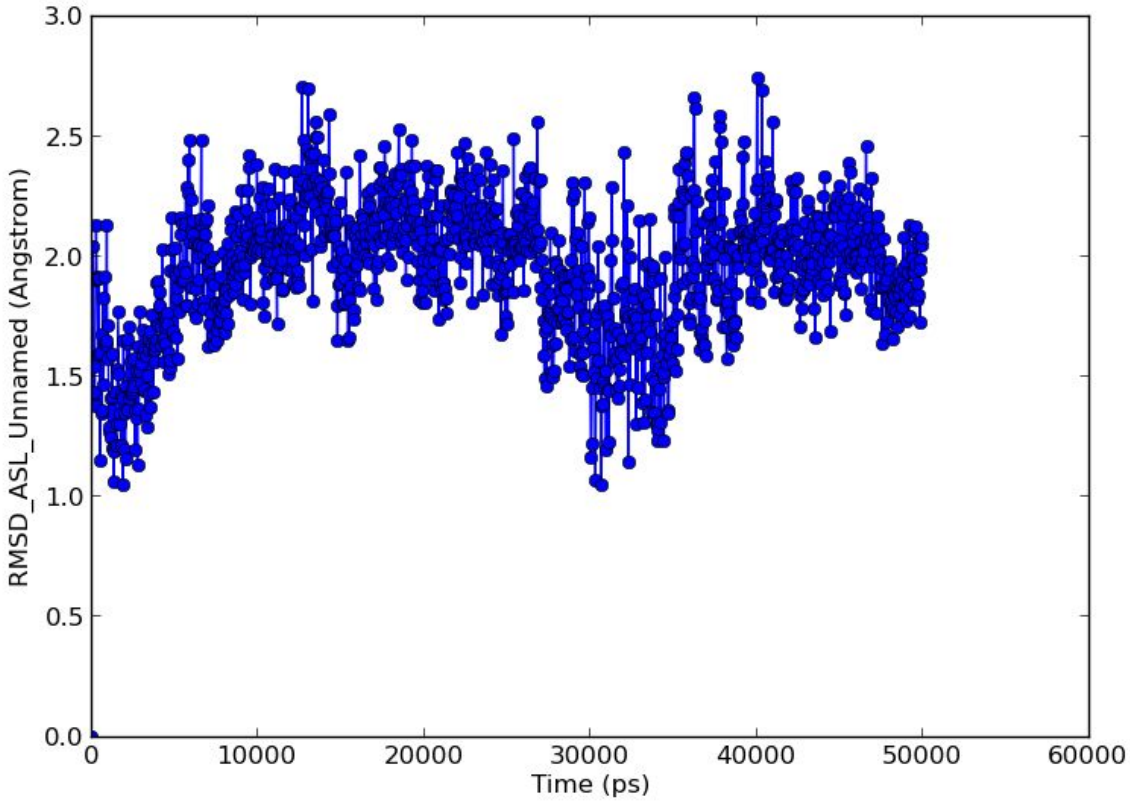


Figure A7.2. RMSD profile of the heavy atom belonging to the Ru-containing “amino acid” from UNIK161-Ru MD run. Figure prepared with Desmond (Bowers et al., 2006).

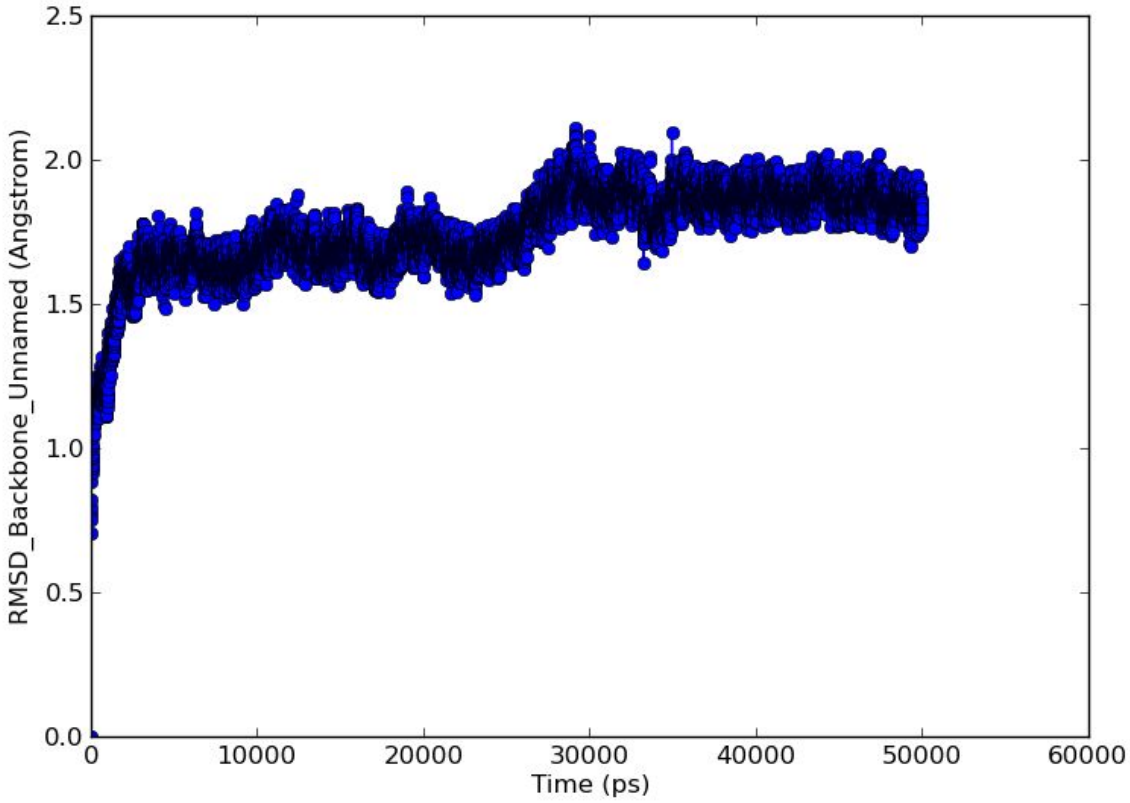


Figure A7.3. Protein backbone RMSD profile of UNIK157-Ru MD run. Figure prepared with Desmond (Bowers et al., 2006).

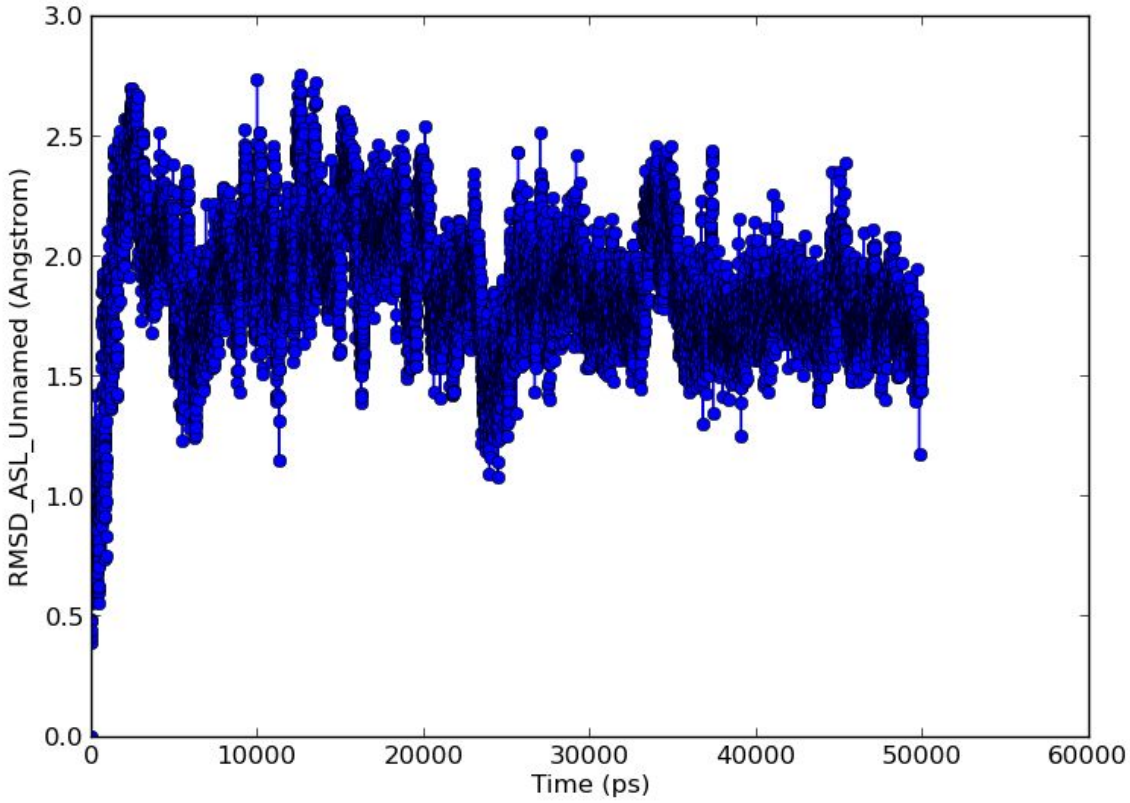


Figure A7.4. RMSD profile of the heavy atom belonging to the Ru-containing “amino acid” from UNIK157-Ru MD run. Figure prepared with Desmond (Bowers et al., 2006).

REFERENCES:

Bowers K., Kevin B., Edmond C., Huafeng X., Ron D., Michael E., Brent G., John K., Istvan K., Mark M., Federico S., John S., Yibing S., David S. 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC’06)*. DOI: 10.1109/sc.2006.54.

Appendix 8 - List of papers, book chapters, oral and poster presentations

Publications in peer-reviewed journals[#]

- G. Santiago, F. de Salas, F. Lucas, E. Monza, S. Acebes, E. Fernandez-Fueyo, A. T. Martinez, S. Camarero, V. Guallar

"Computer-aided laccases engineering: toward bio-based oxidation of arylamines"
ACS Catal. **2016**; 6, 5415-5423

- S. Acebes, E. Fernandez-Fueyo, E. Monza, F. Lucas, D. Almendral, F. J. Ruiz-Dueñas, H. Lund, A. T. Martinez, V. Guallar

"Rational enzyme engineering through biophysical and biochemical modeling"
ACS Catal. **2016**; 6, 1624-1629

- I. Pardo, G. Santiago, P. Gentili, F. Lucas, E. Monza, F. J. Mendrano, C. Galli, A. T. Martinez, V. Guallar, S. Camarero

"Re-designing the substrate binding pocket of laccases for enhanced oxidation of sinapic acid"
Catal. Sci. Technol. **2016**; doi:10.1039/c5cy01725d

- E. Monza, F. Lucas, S. Camarero, L. Alejaldre, A. T. Martinez, V. Guallar

"Insights into laccase engineering from molecular simulations: toward a binding focused strategy"
J. Phys. Chem. Lett. **2015**, 6, 1447–1453

- E. M. Jones,* E. Monza,* G. Balakrishnan, G. C. Blouin, M. J. Piotr, Q. Zhu, J. R. Kincaid, V. Guallar, and T. G. Spiro.

"Differential control of heme reactivity in alpha and beta subunits of hemoglobin: a combined Raman spectroscopic and computational study"
J. Am. Chem. Soc. **2014**, 136, 10325–10339

[#]Four manuscripts are currently in preparation or ready to submit (based on the work presented in Chapters 6, 7, 9 and Sections 5.4- 5.5); a fifth article will be written after the predictions of Chapter 7 will be tested.

*contributed equally to this work

Book chapters

- E. Monza, S. Acebes, F. Lucas, V. Guallar

"Computational physics based methods for enzyme engineering: towards in silico guided directed evolution" in *Directed enzyme evolution: advances and applications*, Miguel Alcalde (Ed.), Springer (in press)

Presentation in international conferences/seminars/workshops

Oral communications

- E. Monza

Computation and experiment meet together: a powerful approach to the rational re-design of oxidoreductases; Oxizymes 2014, Vienna (Austria), July 1st-4th 2014

Poster communications

- F. Lucas, S. Acebes, E. Monza, M. Cañellas, F. Sancho, G. Santiago, V. Guallar

Computer-aided protein engineering

One-day symposium on computational chemical biology honoring Ken Houk, Logroño (Spain), April 22nd 2016

- E. Monza, M. F. Lucas, F. Sancho, G. Santiago, V. Guallar

Computer aided laccases engineering: new directions

IV New trends in computational chemistry for industry applications, Barcelona (Spain), October 1st-2nd 2015

- E. Monza, M. F. Lucas, T. G. Spiro, V. Guallar

On the interplay between local, tertiary and quaternary structure in heme reactivity control

Frontiers in dynamic simulations of biological proteins, Barcelona (Spain), November 4th-6th 2013

- E. Monza, M. F. Lucas, T. G. Spiro, V. Guallar

On the dependence of heme reactivity on subunit structure: a computational study

5th European conference: chemistry for life science, Barcelona (Spain), June 10th-12th 2013

Presentations at departmental seminars, project meetings, etc.

Oral communications

- E. Monza

Expanding the chemical space of LAC3 with $[Ru(bpy)_3]^{2+}$ and visible light

kick-off meeting of the MULTIPLET project, Aix Marseille Universitè, Marseille (France), November 19th-20th 2015

- F. Lucas and E. Monza

Understanding oxidoreductases through computational methods

24-month meeting of the INDOX project, Università degli Studi di Napoli Federico II, Napoli (Italy), September 29th-30th 2015

- E. Monza

Enzyme engineering through efficient conformational sampling and quantum chemical scoring

LifeScience Seminar, Barcelona Supercomputing Center, Barcelona (Spain), October 9th 2014

- E. Monza

Learning from randomness how to rationally design fit-for-purpose laccases

Department of Chemistry, University of Copenhagen, Copenhagen (Denmark), September 22nd 2014

- E. Monza

A computational approach to the rational design of oxidoreductases
Novozymes, Copenhagen (Denmark), September 19th 2014

Poster communications

- V. Giacobelli, E. Monza

Engineering POXA1b laccase *through computational methods*
24-month meeting of the INDOX project, Università degli Studi di Napoli Federico II, Napoli (Italy), September 29th-30th 2015

- E. Monza, M. F. Lucas, F. Sancho, G. Santiago, V. Guallar

Computer aided laccases engineering: new directions
24-month meeting of the INDOX project, Università degli Studi di Napoli Federico II, Napoli (Italy), September 29th-30th 2015