Charles University

Faculty of Science

Physical Chemistry

Modelling of Chemical Properties of Nanostructures and Biostructures

Bc. Petr Eminger

Computer modeling of the catalytic activity of coupled binuclear copper enzymes

Počítačové modelování katalytické aktivity enzymů s dvěma ionty mědi v aktivním místě

Diploma thesis

Supervisor: **prof. Mgr. Lubomír Rulíšek, CSc. DSc.**

Advisor: **Mgr Inż. Agnieszka Stańczak**

Prague 2024

I declare that I prepared the thesis independently and that I have listed all the information sources and literature used. This work, or a substantial part of it, has not been submitted for the award of another or the same academic degree.

In Prague, =======(date)=======

……

Petr Eminger

Acknowledgement

I would like to thank everybody who helped me during this long road to finishing master's degree. I want to thank everybody in the research group of Lubomír Rulíšek at the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences (IOCB). They always respectfully invited me to join their scientific discussions, patiently listened to my ideas and thoughts about various theoretical topics discussed. My colleagues Tadeáš Kalvoda and Ján Michael Kormaník showed me ways how to use quantum mechanics to better understand chemistry around me. I also want to thank Florian Lemken who clarified many of my questions during my studies and showed me possible ways how to find the answer. My officemate and the thesis co-advisor, Agnieszka Stańczak, helped me to better understand calculation protocols and how to look at the studied problems from a broader perspective. My supervisor Lubomír Rulíšek who invited me to join the group, helped me to do my science right way which is sometimes harder than seems at the beginning. He tried to teach me that there are no easy shortcuts, and one has to toil and work hard to unravel the secrets of nature. I also do not want to forget Michal Straka who always supported me and made the research in our group easier not only by telling us about his own thoughts on how to solve the problems. He was also telling us his entertaining and instructive stories which made great atmosphere all around the office. I am glad that I could have been part of the group and finish my diploma project among all these amazing people. Special thanks also belong to my friends who encouraged me to do science and listened to my thoughts, even if they did not understand it. Spending time with them, in between my work on the project, always helped to me to clear my head and not to think about the science all day long.

Abstract:

Coupled binuclear copper (CBC) enzymes is an important family of redox active metalloenzymes. They catalyze various chemical transformations, typically oxidation reactions. Using hybrid quantum and molecular mechanical calculations (QM/MM), we computationally characterized initial steps in the reaction mechanism of two CBC enzymes: catechol oxidases and *o*-aminophenol oxidases. Catechol oxidases were represented by the polyphenol oxidase 6 (PPO6), and *Ipomoea batatas* catechol oxidase (IBCO). The *o*-aminophenol oxidases were represented by the NspF, a protein named after its gene code. By comparing the computed data with the previous work on tyrosinase (Ty), we attempted to answer the question why both NspF and Ty can hydroxylate phenols whereas catechol oxidases cannot. Our preliminary results indicate that there is a significant difference in the energy of the rate-determining transition state corresponding to the attack of the activated oxyl radical along with the phenol binding to the active site. The computed energy for the second transition state is lower for NspF in comparison with PPO6 and IBCO. Our work may help to understand the chemoselectivity of CBC enzymes, identify the second-sphere residues that are responsible for the diverse reactivities and potentially open new ways to design the CBC enzymes or the biomimetic systems.

Keywords:

metalloenzymes, coupled binuclear copper active sites, QM/MM calculations, phenol hydroxylation

Abstrakt:

Enzymy s dvěma “spřaženými” měďnatými ionty v aktivním místě (coupled binuclear copper enzymes, CBC) jsou důležitou skupinou redoxně aktivních metaloenzymů. Katalyzují různé chemické přeměny, typicky oxidační reakce. Pomocí hybridních kvantových a molekulově mechanických výpočtů (QM/MM) jsme výpočetně charakterizovali počáteční kroky v reakčním mechanismu dvou CBC enzymů: katecholoxidáz a *o*-aminofenoloxidáz. Katecholoxidázy byly reprezentovány polyfenoloxidázou 6 (PPO6) a katecholoxidázou z *Ipomoea batatas* (IBCO), *o*-aminofenoloxidázy pak zastupoval protein NspF, pojmenovaný podle svého genového kódu. Porovnáním vypočtených údajů s předchozími prácemi zabývajícími se tyrosinázou (Ty) jsme se pokusili odpovědět na otázku, proč NspF a Ty hydroxylují fenoly, zatímco katecholoxidázy nikoli. Naše předběžné výsledky ukazují, že existuje významný rozdíl v energii tranzitního stavu určujícího rychlost reakce, který odpovídá radikálové reakci aktivovaného kyslíkového (oxylového) radikálu spolu s vazbou fenolu na jeden z iontů mědi v aktivním místě. Vypočtená energie tohoto (v pořadí druhého) tranzitního stavu je nižší pro NspF než pro PPO6 a IBCO. Naše práce tak může pomoci pochopit chemoselektivitu CBC enzymů, identifikovat aminokyselinové zbytky ze vzdálenější druhé koordinační sféry, které jsou zodpovědné za rozdílné reaktivity enzymů CBC, a potenciálně pak otevřít nové cesty k návrhu modifikací v těchto enzymech nebo jejich biomimetických systémech.

Klíčová slova:

metaloenzymy, binukleární aktivní místa, ionty mědi, QM/MM metody, hydroxylace fenolu

List of abbreviations

AA	Amino Acid
AOx	*o*-Aminophenol Oxidase
CaOx	Catechol Oxidase
CBC	Coupled Binuclear Copper
COSMO	Conductor-like Screening Model
DFT	Density Functional Theory
ECP	Effective Core Potential
F_{HK}	Hohenberg-Kohn Functional
GGA	Generalized Gradient Approximation
Hc	Hemocyanin
HEG	Homogeneous Electron Gas
HF	Hartree-Fock
Hid	Histidine Protonated at δ-nitrogen
HK	Hohenberg-Kohn
IBCO	*Ipomoea batatas* Catechol Oxidase
LDA	Local Density Approximation
LJ	Lennard-Jones
KS	Kohn-Sham
MM	Molecular Mechanics
O	Bis-μ-oxo
P	Side-on-peroxo
PDB	Protein Data Bank
PES	Potential Energy Surface
PPO	Polyphenol Oxidase
QM	Quantum Mechanics
SchrEq	Schrödinger equation
SCRF	Self-Consistent Reaction Field
SOD	Superoxide Dismutase
TS	Transition State
TST	Transition State Theory
Ty	Tyrosinase

1. Introduction

1.1. Enzymes

Life on Earth is mediated by many types of organic molecules: such as lipids, carbohydrates, nucleic acids, and proteins. The last ones, proteins, have structural, mechanical, signaling, response and many other functions in cells. Even more importantly, they are the most efficient biocatalysts known to date. The proteins involved in biological reactions are called enzymes.

Enzymes mostly differ from small catalysts by their high specificity, chemo- and regioselectivity. To fully understand these genuine properties of enzymes and how they are encoded in their primary, secondary, and tertiary structure, a combination of experiment and theory seems to be the most efficient approach. For almost any enzyme, there is considerable amount of experimental data, notably the catalytic rates and Michaelis constants which define catalytic efficiencies and turnover numbers of enzymes. The most straightforward way to describe and verify a potential reaction mechanism is to compare the computed activation (free) energy with the experimental rate constant, k_{cat}. To do so, we may consider the *transition state theory* (TST)[1]. TST assumes formation of the activated complex, which is in equilibrium with the reactants. Once the activated complex passes through the transition state, it results in the formation of the products of the reaction (TST assumes that no re-crossing occurs). The rate constant of such bimolecular reaction can be calculated by using the Eyring equation[2].

$$k_{cat} = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \tag{1.1}$$

where k stands for rate constant, $\Delta G^{\ddagger}$ stands for the activation Gibbs energy and $\boldsymbol{\kappa}$ is the transmission coefficient, typically assumed to be equal to 1 (i.e., every activated complex will always cross the transition state and end up in the product formation).

Some enzymes need a non-protein molecules or atoms (also called *cofactors*) to have catalytic activity. Cofactors can be molecules of an organic origin (ATP, FAD, biotin etc...). However, there is also group of enzymes called metalloenzymes, which contains a metal ion in their active site (typically zinc, iron or copper)[3] to induce their catalytic activity.

1.2. Metalloenzymes

It has been estimated that around 35% of the enzymes contain metal ions[4–6]. Metals are coordinated inside the protein structure by negatively charged or polar amino acid side chains. The most common ligands occuring in the metalloenzymes are side chains of His, Cys, Asp/Glu, and less of a Met, Tyr, Ser, Thr, Lys. They coordinate through their imidazole, carboxylate, thiols/ate, or hydroxyl-containing side chains. Metal cations and the coordinated ligands (sometimes denoted as the first coordination sphere) form the reaction center. There are multiple ways how the reaction center interacts or reacts with the substrate. Two of the common ways include activation of the second co-factor, are involvement of dioxygen or water molecule/hydroxide for the reaction. In addition, it is assumed (not only for metalloenzymes) that the specific and fine-tuned geometry of the active site stabilizes transition state structures more than reactants and thus lowers the activation energy for the catalytic reaction.

It can be mentioned that binding of a metal ion to a metal-binding site in the protein is in itself, often specific. This phenomenon also represents an important factor in regulating the catalytic efficiency, and chemoselectivity of metalloenzymes. One of the factors that influences the metal-binding specificity is the match of polarizability of a metal ion and a ligand, known as HSAB (hard- and softs- acids and bases) principle. The metal is Lewis acid and the ligand is Lewis base. Soft metal ions have high polarizability (electrons around the core are more easily deformed in the applied electric field or by metal-ion charge) and preferably binds to soft bases (such as sulfur-containing side chains). Thus, metals of the first and second group (Na^+, K^+, Mg^{2+}, Ca^{2+},...) are considered as hard metal ions. They are mostly interacting with their ligands by 'almost purely' electrostatic forces. Nature mostly uses them for transmission of neural signal, stabilization of structures of other organic compounds (for example, Mg^{2+} stabilizing ATP), muscle contraction and others. On the other hand, some soft metals may act like poisons by binding tightly to metalloenzymes and thus deactivating them Hg^{2+} is such an example of lethal poisoning[7]. That might be also one of the reasons why not hard or soft, but the intermediate acid metal ions (such as Zn^{2+}, Cu^{2+} and Fe^{2+}) are found in the enzyme active sites. Another, quite obvious reason for finding transition metals (e.g., Cu^{2+} and Fe^{2+}) in the metalloenzymes' active sites is the change of their oxidation state which enables the redox reactions. Also, it may influence the change in the distribution of electrons in reactant(s), i.e. their activation (prominent examples are the CBC enzymes studied herein). One can also find rare, but interesting examples of proteins/enzymes with lanthanides in their active sites (for example, methanol dehydrogenase in methylotrophic bacteria[8]).

As mentioned above, numerous metalloenzymes contain Zn^{2+}, $Fe^{2+/3+}$ or $Cu^{+/2+}$ ions in their active sites. Zn^{2+} ions are part of many lyases and hydrolases. As an example, alkaline phosphatase and phospholipase C are enzymes catalyzing the hydrolysis of phosphate esters. Both enzymes feature the binuclear zinc site whose putative function is to decrease the pK_a of the hydroxyl groups facilitating the nucleophilic attack on the ester and at the same time stabilizing the negative charge of the leaving product. One of the well-known iron enzymes is superoxide dismutase (SOD) converting superoxide anion radical to molecular oxygen and to hydrogen peroxide in the two disproportionation half-reactions. This enzyme can also contain manganese, or Cu/Zn binuclear site, instead of an iron. In SODs, iron is coordinated in a tetrahedral coordination geometry by three histidines and one aspartate in an equatorial position whereas superoxide binds as the fifth ligand, at least in one of the half-reactions. It can be also mentioned that most of iron ions are found in hemes, whereas a broad class of non-heme iron enzymes exists. Last but not least, iron is also utilized in the iron-sulfur clusters which are one of the most efficient electron transfer species that nature uses. Biotin synthase is one of (many) enzymes containing the Fe_4S_4 cluster. The enzyme catalyzes the last step of biotin synthetic pathway (from dethiobiotin to biotin). Combination of iron and zinc can be found in the purple acid phospatase, where Fe^{3+}-Zn^{2+} binuclear site catalyzes the hydrolysis the phosphate ester, in a similar way as previously mentioned.

Copper metalloenzymes often employ an oxygen molecule to catalyze redox reactions. We might hypothesize that this could have happened evolutionary when the concentration of dioxygen in the atmosphere increased, and organisms took an advantage of it. Copper properties in enzymes are somewhat similar to iron. Copper sites in proteins are classified into four groups distinguished by their spectroscopic properties. These four classes are: type 1 (often called *blue copper sites/proteins*), type 2 (mononuclear copper proteins), type 3 (coupled binuclear copper sites; CBC's) and CuA active sites/centers[9].

Blue copper proteins are found in plants and bacteria and the active site comprises $Cu^{+/2+}$ coordinated by two histidines and one cystein[10]. The very intense $S_{Cys}(p) \rightarrow Cu(d)$ charge-transfer excitation absorption around 600 nm gives the protein the intensive blue color. These proteins catalyze variety of biochemical reactions, from electron transfer in photosynthesis to the deamination of primary amines. As an example, cupredoxin is the enzyme involved in electron transfer processes. Its characteristic feature is the so-called cupredoxin fold (Figure 1), a single protein chain coordinating one copper ion. This fold is also found in other copper proteins. Type 2 copper centers are mostly coordinated by three or four ligands, mostly

histidines. Due to the absence of the sulfur coordination, they exhibit weak EPR signal distinguishing them from the type 1 copper proteins. Mononuclear copper proteins are activating dioxygen by (formally) one-electron reduction; activated O_2 then attacks a substrate. One of the prominent examples is the galactose oxidase. This enzyme catalyzes the oxidation of primary alcohols to aldehydes while reducing the oxygen molecule to hydrogen peroxide[11]. The third group of copper enzymes, CBC enzymes, will be discussed in the next chapter, because they are the focus of this thesis. Finally, CuA proteins are found in domains of multidomain proteins like cytochrome c oxidase. Compared to the blue copper proteins, the CuA proteins contain one additional copper site. The extra Cys close to the active site then act as a ligand for the second Cu. Blue shift of its charge-transfer band makes the enzyme purple. However, after the reduction of the active site the enzyme becomes colorless. Such example of CuA domain enzyme is cytochrome *c* oxidase creating an acceptor of electrons from cytochrome c[12].

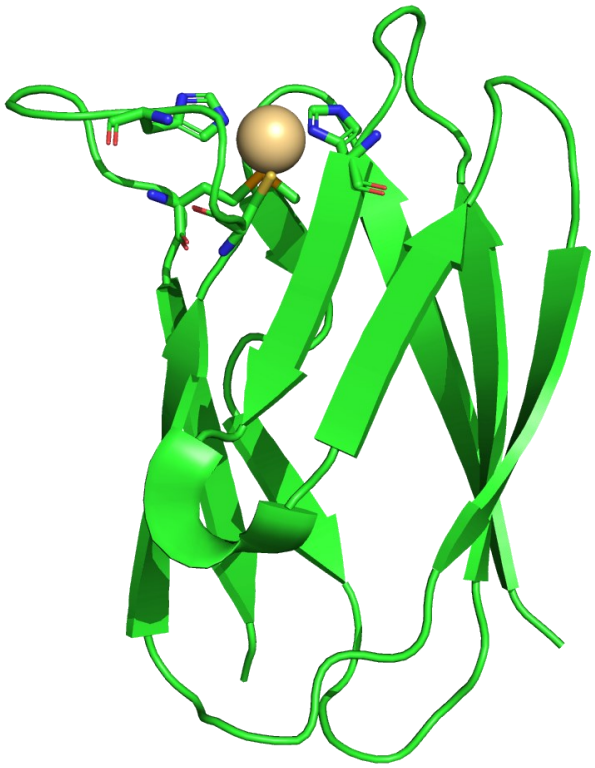

Figure 1 - Cupredoxin structure with the metal-binding ligands highlighted as sticks; the copper ion is in (black ball).

1.3. CBC enzymes

The presence of two copper ions in CBC enzymes makes them ideal catalysts for many one- or two-electron oxidation reactions.[7] The two copper cations are located near each other (3.5 – 4 Å), coordinated by three histidines each. This arrangement is well suited for binding of molecular oxygen (c.f. Figure 2). One molecule of O_2 reversibly binds to the reduced enzyme (*deoxy* form; Figure 2, left), forming μ-η_2:η_2-peroxide dicopper(II) active site (Figure 2, right) with a single O-O bond. The oxidized (*oxy*) form is subsequently able to catalyze oxidation reactions of various types.[13]

Figure 2 - Difference between deoxy and oxy structures for all CBC enzymes shown on artificial amino complex structures.

The coupled coppers in the CBC sites not only distinguishes this class of enzymes from their non-coupled counterparts structurally, but also gives them different properties. Noticeably, there is a strong antiferromagnetic exchange coupling between the two copper(II) ions. This coupling can be experimentally confirmed by the absence of the electron paramagnetic resonance (EPR) signal.[13] On the other hand, coppers in non-coupled binuclear enzymes tend to catalyze one-electron oxidation reactions, by employing only one copper ion and one-electron reduced oxygen molecule, i.e. superoxide. The second copper supplies the oxidation by long-range electron transfer to the copper with bound oxygen.

In the O_2-activated CBC site, an equilibrium exists between the two isomers, namely μ-η_2:η_2-peroxide dicopper(II) (side-on-peroxo, **P**) and bis-μ-oxo dicopper(III) (bis-μ-oxo, **O**), depicted in Figure 3. Cramer and Tolman[14] were first to report that three ligands coordinated to each of the coppers, in addition to the bridging O_2 molecule, tend to stabilize **P** to a greater degree than **O**, by approximately 6-8 kcal·mol^{-1}. This has also been confirmed later, perhaps more rigorously, by Stanczak and coworkers[15], employing a broader set of small or medium-sized inorganic complexes. At the same time, they (Stanczak *et al.*) showed that the two-coordinated (or bidentate) copper ions in the CBC site results in the lower energy of the **O** isomer. Both these observations are in agreement with experimental data. However, in the case

of CBC proteins/enzymes, this equilibrium can be shifted by the interactions of the site with water molecules or other parts of the protein. As both of the **P**/**O** forms can be reactive intermediates, an evaluation of the equilibrium is an important step in understanding the catalytic action of CBC enzymes[16].

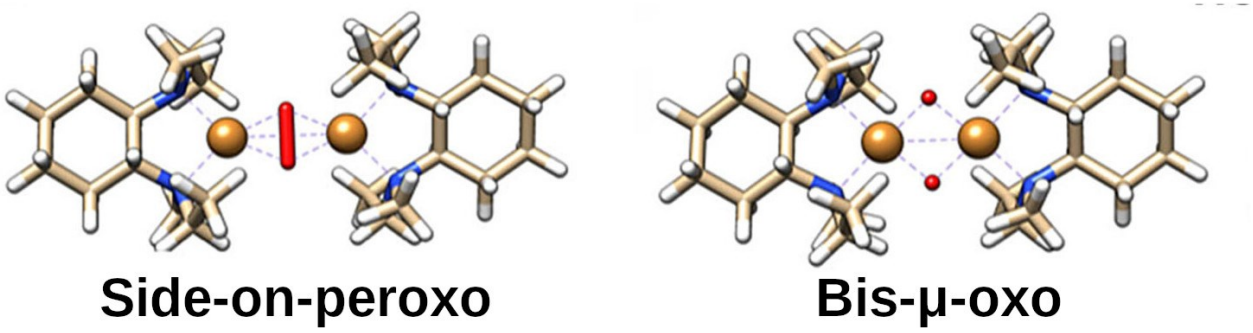

Figure 3 - Difference between the peroxo and oxo structure of CBC active sites, presented on artificial cluster models. Taken from the study of Agnieszka Stańczak.[15]

To provide few representative examples of CBC enzymes, we can mention hemocyanin (Hc) as one of the well-known representatives. Hemocyanin by itself does not exhibit catalytic activity, but it can reversibly bind oxygen to distribute it to other tissues or compartments of organisms. Next class of CBC enzymes are catechol oxidases (CaOx) which catalyze the oxidation of catechol to *ortho*-quinone, which is a part of the defense mechanism in plants.[17] However, unlike tyrosinase (Ty), it cannot catalyze the hydroxylation reaction of monophenol in the *ortho*-position.

Previous studies carried out in our group, in collaboration with experimentalist from Stanford University, were mainly focused on the CBC protein tyrosinase (Ty)[18–20], the enzyme which can catalyze the oxidation of L-tyrosine to L-DOPA and further to L-dopaquinone (Figure 4a). These reactions are part of a large cascade to produce melanin, a molecule used by organisms for its photoprotective and antioxidant properties. First attempts to understand the Ty mechanism were done by Siegbahn.[20–22] He tried to predict the reaction pathway employing small model mimicking the Ty active site (the so-called *cluster model* approach), because the crystal structure of Ty was not available at the time. He concluded that the proton from the phenolic OH group is transferred to one of the oxygens in the $[Cu_2O_2]^{2+}$ core and the negatively charged phenolate oxygen is then stabilized by a coordination to one of the copper cations.

Another study from Inoue et. al. discussed possible mechanism of Ty up to the quinone synthesis via QM/MM calculations involving phenoxyl radical.[23] Very recent study by Kiporous[18] described – experimentally and computationally - the formation of a ternary intermediate (Ty/O_2/phenol). The intermediate has been postulated to be the initial structure in a monooxygenation of phenols. Computationally and spectroscopically, it is consistent with **P** for of the Ty. The substrate (tyrosine/phenol) is still protonated, hydrogen-bonded to the bridging peroxide, thus not directly interacting with any of the two coppers. The ternary intermediate afterwards undergoes monohydroxylation, which can be (experimentally, employing carefully selected phenolic substrate with appropriate para- group and the assay with borate buffer) stopped at this step, producing catechol(ate) or it may proceed to quinone (as is by Ty done in nature, for Tyr as a substrate). Coppers in Ty ends up in the reduced state and after release of water molecule (that accepted proton from phenol/Tyr), the enzyme is in the *deoxy* state (Figure 2). The reduced form can then be reversibly oxidized by an oxygen molecule to form oxy-Ty and start new cycle again. (Figure 4b). Interestingly, the resting state of Ty is *met*-Ty (c.f. Figure 4b).

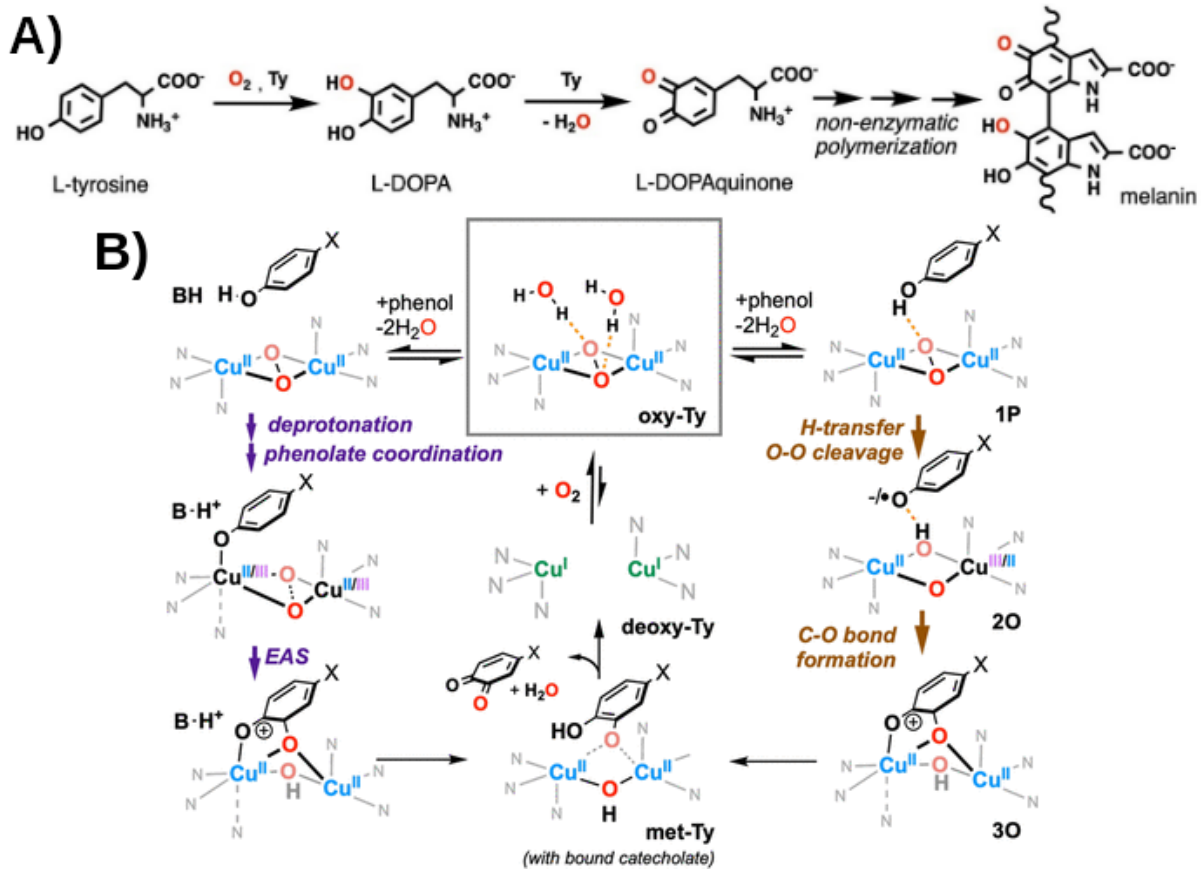

Figure 4 – Reaction pathway of L-tyrosine to melanin (A). Tyrosinase catalyzes first to reactions from L-tyrosine to L-dopaquinone (B).[18]

The same monohydroxylation can be catalyzed by NspF (*o*-aminophenol oxidase). In addition, NspF can also catalyze the oxidation reaction of *ortho*-aminophenol to *ortho*-nitrosophenol whereas Ty cannot. Interestingly, little is known about GriF, another AOx, which has different reactivity than NspF, despite the high similarity in their tertiary structures. Instead of oxidizing the aromatic amino compound to nitroso compound, GriF catalyzes the reduction of amino group to imino group. Because of the lack of data, we will not focus on GriF in this work, but we felt it appropriate to mention it here to show the diversity of reactions catalyzed by CBCs. All mentioned enzymes (except GriF) are depicted in Figure 5. We can see that they are structurally very similar, still the scope of the reactions they catalyze differ. The ultimate goal of the ongoing research in our group, in tight collaboration with the Solomon group at Stanford is to find subtle structure-function relationships of CBC proteins that may explain diverse reactivities.

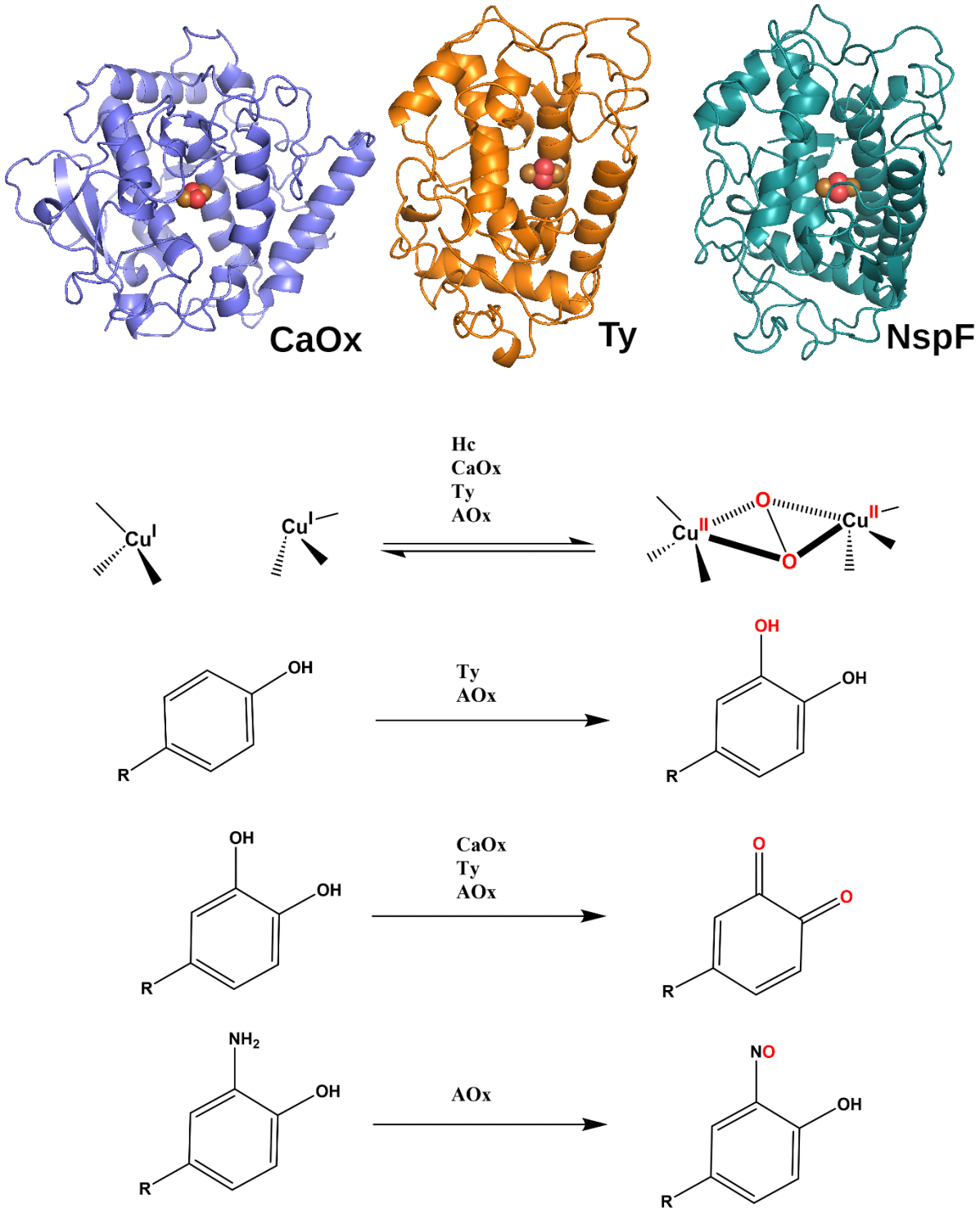

Figure 5 - Protein structures of CBC enzymes and their catalyzed reactions

2. Computational chemistry

2.1. Quantum chemistry

Non-relativistic molecular quantum mechanics (QM) starts with the Schrödinger equation (SchrEq), which describes the evolution of the system (uniquely described by its wave function) with time:

$$\left\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V\right\}\Psi(\boldsymbol{r},t)=i\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t}$$

(2.1)

where ħ is the reduced Planck constant (1.0546 x 10^{-34} J·s), m is the mass of the particle, x, y, z are positions of the particle in 3D space, V is the potential energy operator, Ψ is the wavefunction of the system and i is the imaginary unit. The first term in curly brackets is the kinetic energy operator.

All properties of a system (*here* molecules) are contained in the wavefunction. The wavefunction is linked to an observable property by *operators*. Each (Hermitian) operator is characterized by the set of eigenfunctions and corresponding eigenvalues that yield the observable set of the properties (Eq. 2.2). Their mean value is then denoted as the expectation value of the particular property.

$$\hat{A}\Psi = a\Psi$$

(2.2)

By considering only stationary solutions of the SchrEq, it can be rewritten into the time-independent form. For each system in which the energy is time-dependent, it can be demonstrated that the wavefunction can be separated into a time-dependent and a space dependent part. Considering only the space-dependent part leads to the time-independent SchrEq:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+\hat{V}\right\}\Psi(\boldsymbol{r})=\boldsymbol{E}\Psi(\boldsymbol{r})$$

(2.3)

Thus, stationary SchrEq is an eigenvalue problem of Hamiltonian (energy operator). Unfortunately, the SchrEq does not include the electron spin. Spin is then added into the SchrEq as an *ad hoc* postulate. Spin can be, to some extent, interpreted as an intrinsic angular

momentum, and its quantum number depends on the type of particle (for electrons: s=1/2). The spatial orientation of the spin forms orthogonal states, which by convention are quantized with respect to the z-axis and denoted with the spin magnetic quantum number m_s (m_s = {-s,...,s}). For a single electron, the two spin states are referred to as α (m_s = 1/2) and β (m_s = -1/2). S denotes the quantum number of the total electronic spin of the system. Analogous to s, the set of possible orientations of the total electronic spin is described by quantum number M_s. The number of orientations in that set is called the spin multiplicity M ($M = 2S + 1$). Systems can be classified by their multiplicity (e.g. M = 1 corresponds to a singlet, M = 3 to a triplet etc.). For computational purposes, the spin multiplicity is commonly described by the expectation value of the square of the total spin operator. For example, for two electrons S^2 becomes:

$$\hat{S}^2 = s_1^2 + s_2^2 + 2s_1s_2 \tag{2.4}$$

The expectation value $\langle S^2 \rangle$ can be calculated from a trial wavefunction. For a pure singlet or triplet state, the $\langle S^2 \rangle$ values are 0 and 2, respectively (in general $\langle S^2 \rangle = S(S+1)$, for eigenfunctions of the $\hat{S}^2$ operator, which should be the case for the 'correct' wave functions, since $\hat{S}^2$ commutes with Hamiltonian).

The major drawback of the SchrEq is that it can only be solved analytically for systems with one nucleus and one electron, such as the hydrogen atom, He^+, Li^{2+} etc. Thus, additional approximations need to be introduced for larger systems.

The Born-Oppenheimer approximation[24] allows us to separate the motions of nuclei and electrons. Nuclei are much heavier than electrons (~1800x). Because of this, we may consider nuclei fixed in space and electrons moving in their vicinity. In this context, kinetic energy of the nucleus is considered zero and internuclear repulsion can be described classically. Therefore, we can describe the electronic Hamiltonian by three terms:

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne} \tag{2.5}$$

where $\hat{T}_e$ stands for operator of kinetic energy of electron, $\hat{V}_{ee}$ describes the Coulombic repulsion between the electrons, and $\hat{V}_{Ne}$ is the operator of Coulombic attraction between the nuclei and electrons (subscripts N, e denote nuclei and electrons, respectively).

The expectation value of the electronic Hamiltonian is the total energy of electrons in atoms or molecules (called e*lectronic energy*). Under the Born-Oppenheimer approximation, the total energy of a molecule (or atom) is the sum of the electronic energy and the energy of nuclear repulsion V_{NN}. The dependency of the molecular energy on nuclear coordinates describes a *potential energy surface* (PES; Figure 6).[25] The PES is crucial for studying reactions and reaction mechanisms. As there are $3N$ - 6 independent coordinates for non-linear molecules ($3N$ - 5 for linear molecules), the PES has the same number of dimensions.

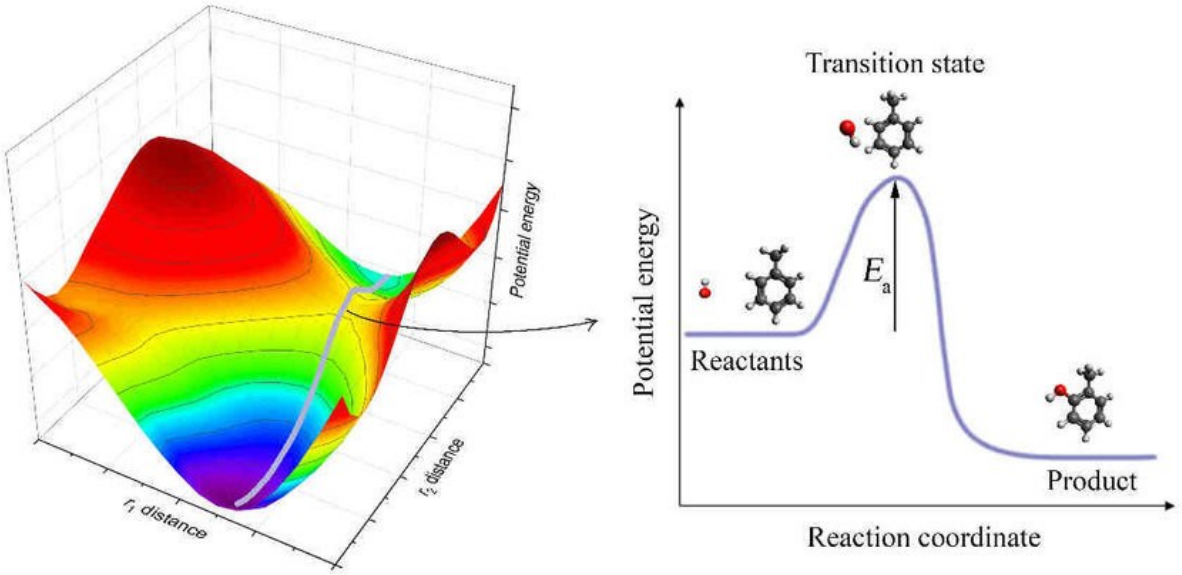

Figure 6 - Graphical representation of PES. The pictures show 3D and 2D graphs. The plot on the right depicts an oxidation reaction using PES. This figure was taken and modified from the article of Stanisław Wacławek.[26]

SchrEq can be approximately solved by a variety of methods. These include the Hartree-Fock (HF) method as the simplest *ab initio* method, post-HF single-reference methods (e.g., configuration interaction, perturbation theory, coupled cluster), or multiconfigurational approaches. However, for most systems in bioinorganic chemistry, these methods are computationally expensive, especially when it comes to the 'production' work. Nowadays, the most frequently used methods in computational chemistry are *density functional theory* (DFT) methods, which are used in this work (see below).

2.2. Density functional theory

Density functional theory (DFT) is based on the idea that all information about a system is contained in its *electron density,* $\rho(r)$ (Figure 7). The electron density describes the probability of finding an electron at the given point in space. The electron density is related to the wavefunction as follows:

$$\rho(\vec{r}) = N \int |\psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n)|^2 \mathrm{d}^3\vec{r}_2 \ldots \mathrm{d}^3\vec{r}_n \tag{2.6}$$

where N stands for number of electrons, ψ is a wavefunction and the integration goes over all but one electronic coordinates. $\rho(r)$ is a function of three variables instead of 3N variables that are arguments of the molecular wave function. We should be aware of a few properties of the electron density. I) The electron density attains only positive values (as the probability of finding an electron cannot be negative), II) If we integrate $\rho(\boldsymbol{r})$ over the whole space, we will get the number of electrons in the system, III) $\rho(\boldsymbol{r})$ has maxima (cusps) at the positions of nuclei.

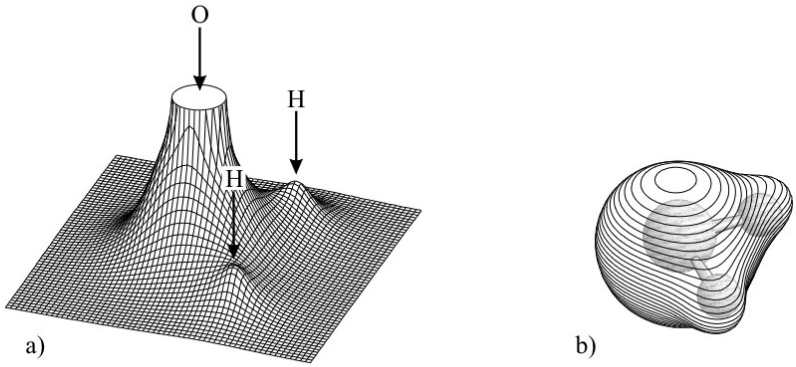

Figure 7 – Electron density of a water molecule. a) Map of electron density with peaks at the positions of nuclei. b) Isosurface plot of electron density. Picture retrieved from "A Chemist's Guide to Density Functional Theory".[27]

A functional is a special type of operator that returns a value rather than a function. An example of a functional can be an improper integral (from -∞ to +∞), which for a function returns the value under the curve described by the function. The key ingredient in DFT is expressing the energy of a system (kinetic, potential, and total) as a functional of the electron density.

Formally, DFT is based on the two Hohenberg-Kohn theorems.[28] The first theorem requires the definition of the *external potential,* as a description of the external field in which electrons move. In the absence of external fields (electric, magnetic, …), it reads:

$$v_{\text{ext}} = -\sum_{i=1}^{N}\sum_{K=1}^{M}\frac{Z_K}{r_{iK}} \tag{2.7}$$

which is nothing else than electrostatic potential of nuclei corresponding to the coulombic interaction between the electrons and nuclei. The first Hohenberg-Kohn (HK) theorem says that *'The external potential v_{ext} is (to within constant) a unique functional of $\rho(\mathbf{r})$; since, in turn v_{ext} fixes $\hat{\boldsymbol{H}}$ we see that the full many particle ground state is a unique functional of $\rho(\mathbf{r})$'.*[28] In other words: v_{ext} (positions of the nuclei and their charges) is uniquely defined by the electron density, for any number of interacting electrons. If the external potential is known, then the Hamiltonian can be constructed, and the energy of a system can be calculated. This idea allows us to use electron densities instead of wavefunctions to calculate the energy of a system. An energy functional can be written as:

$$E[\rho] = T[\rho] + V_{\text{ee}}[\rho] + V_{\text{Ne}}[\rho] \tag{2.8}$$

where $T[\rho]$ stands for the kinetic energy functional of electrons, $V_{\text{ee}}[\rho]$ is the functional of electron-electron repulsion and $V_{\text{Ne}}[\rho]$ is the functional of nucleus-electron attraction in the external potential. The first two functionals from equation 2.8 are collectively called the Hohenberg-Kohn functional (F_{HK}).

The second HK theorem is a "variational principle of DFT". Defining the exact density of the ground state as ρ and the exact energy as $E[\rho]$, it can be shown that for an electron density ρ' ($\neq \rho$), the energy is higher than the ground state energy:

$$E[\rho] < E[\rho']. \tag{2.9}$$

A practical way to find F_{HK} has been formulated by Kohn and Sham (KS). While approaches for the calculation of kinetic energy purely as a functional of electron density have been proposed, they all were found to be not sufficiently accurate in comparison to even HF. Therefore, KS reintroduced orbitals into the DFT approach, employing the homogeneous non-interacting electron gas as the reference system. This leads to straightforward formulation of

the kinetic energy functional analogous to HF methods. After some mathematical rearrangements, it leads to a system of single-particle equations known as KS equations:

$$(\hat{T}_{HF,i} + V_{eff,i})\varphi_i = \varepsilon_i \varphi_i \quad \textbf{(2.10)}$$

where $T_{HF,\,i}$ describes the single-particle kinetic energy operator, $V_{eff,i,}$ is an effective single-particle potential, φ_i is the *i*th molecular orbital and ε_i is respective orbital energy.

2.3. DFT functionals

As mentioned above, the principal problem of DFT is that the exact functional is not known. We may separate the parts of the functional from those unknown as follows:

$$E = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \quad \textbf{(2.11)}$$

where $T_S[\rho]$ is the kinetic energy functional for non-interacting electrons (Kohn-Sham approach), $J[\rho]$ is the functional of Coulombic interaction energy for electrons and E_{XC} is the *exchange-correlation functional*. The exchange-correlation functional describes the corrections for both kinetic and Coulombic terms, which are not known exactly and need to be approximated.

There are various approaches to formulate E_{XC}, differing by the complexity and thus computational expense. To start from the least complex, the *Local Density Approximation* (LDA)[29] is derived from the model of the *homogeneous electron gas* (HEG). In this model, the electron density is homogeneous with positive charge in the background. To derive the E_{XC}, the LDA treats the molecular electron density locally as an HEG. More complex are *Generalized Gradient Approximation* (GGA)[30] functionals, adding the gradient of the electron density $\nabla\rho(\mathbf{r})$. Further derivatives of the electron density ($\nabla^2\rho(\mathbf{r})$, $\nabla^3\rho(\mathbf{r})$, ...) are included in functionals of the meta-GGA class. The TPSS[31] functional used in this work is an example thereof. Another category are the *hybrid functionals*. These functionals include a fraction of Hartree-Fock calculated exchange energy (exact exchange) . An example is the TPSSh[31] functional used in this work (or the "golden standard of DFT", B3LYP functional[32]). Furthermore, there are functionals (*double hybrid functionals*) adding perturbation-theory based correlation energy. This categorization of functionals based on complexity is also called Jacob's ladder.

2.4. Molecular mechanics

QM methods are the optimal choice for small to medium-sized molecular systems (let us say up to 500 atoms can be conveniently treated). However, when we deal with thousands of atoms, the QM calculations are still prohibitively expensive, especially for practical projects when hundreds of structure are needed to be efficiently handled in days to weeks. In such cases, molecular mechanics (MM) is the method of choice. It is empirical method, based on the force field and it is in general less accurate than QM methods. Still, it quite often provides valuable information MM methods can be conveniently used for systems up to ten thousands of atoms.

MM methods do not treat the system as composed of electrons and nuclei but treat a molecule as an entity of interconnected atoms. Thus, MM does not provide any information about the electronic structure of a system.

The potential energy (force field) is described by effective potentials associated with bonds, angles, dihedral angles, electrostatics, and pair interaction (*here* Lennard-Jones (LJ) potential):

$$V = \sum_{bonds,i} k_i(d_i - d_{0,i})^2 + \sum_{angles,i} k_{\alpha i}(\alpha_i - \alpha_{0,i})^2 + \sum_{dihedral,i} A_{n,i}\, cos(n\tau_i)$$
$$+ \sum_{oopw,i} k_{oopw,i} s_i^2 + \sum_{i,j=1;\, N, i<j} \frac{1}{4\pi\varepsilon} \frac{Q_i Q_j}{r_{ij}} + \sum_{i,j=1;\, N, i<j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]. \quad \textbf{(2.12)}$$

Potential energy contributions from bonds and angles are mostly described by the harmonic potential approximation, where k is a force constant, d and α stand for changes in distance and angle, respectively. The potential energy arising from the dihedral angle is described by a sum of cosine functions of angle τ, where A is the amplitude and n is the number of minima. The fourth parameter describes improper dihedral angles (*out-of-plane wagging,* oopw), where s is the out-of-plane distance. The next parameter is the classical Coulomb term, where ε is the permitivity of a vacuum, r_{ij} is the distance between the particles (atoms) i and j, with $Q_{i/j}$ are their charges. The last parameter determines the pair potential, described in this case by the LJ potential. The first term in the brackets represents short-range repulsive forces, and the second term represents attractive forces (mostly of dispersion character). The constants ε and σ stand for the depth of the potential well and the distance at which the potential between two particles vanishes, respectively (Figure 8). The set of parameters used in the construction of the potential is commonly referred to as a *force field*.

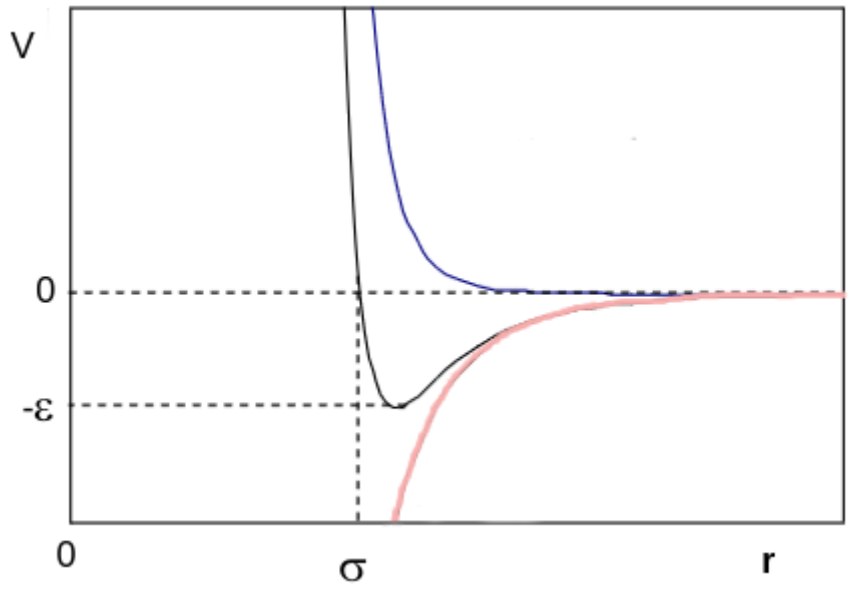

Figure 8 - Lennard-Jones potential . Blue line represents repulsive part of the potential, pink line represents the attractive part. Black line in the middle is the sum of those two potential components together. Picture taken and modified from the study material of Petr Bouř.[33]

Molecular mechanics have some limitations. They cannot describe the electronic structure of the system, they are not as accurate as QM methods, and force fields are usually based on empirical parameters. Another drawback of MM is its inability to form/break bonds. For these reasons, MM is often used in combination with QM methods, resulting in the QM/MM method. Therein, MM is used to describe the non-reactive part of the system and QM is used for description of the reactive part.

2.5. QM/MM calculations

Hybrid quantum and molecular mechanical (QM/MM) calculations were introduced in 1976 by Warshel and Levitt[34]. They are intended to couple the accuracy of quantum mechanical description with the (very) low cost of molecular mechanics. One of the prominent areas of their applicability is the computational modeling of enzyme reaction mechanisms. Thus, QM/MM enables the description of enzymatic reactions by applying QM for the active site for which an accurate description is necessary whereas the rest of the protein is treated at the MM level. The protein framework might not be directly involved in the enzymatic reaction studied, but influences the reaction coordinate via long-range electrostatics, of conformational changes in the vicinity of the QM region.

Within the QM/MM framework, the system is usually divided in two parts (Figure 9), which interact with each other. The energy of the system can be calculated in two ways. The first one employs subtractive scheme which defines the total energy as

$$E_{\mathrm{QM/MM,total}} = E_{\mathrm{MM,total}} + E_{\mathrm{QM,QM}} - E_{\mathrm{MM,QM}} \quad \textbf{(2.13)}$$

Depending upon arrangement of the terms in the Eq. 2.13, the total QM/MM energy can be qualitatively viewed either as (1) MM energy of the whole system, 'upgraded' by the QM energy of the QM part or (2) sum of the QM energy of the quantum system and MM energy of the rest of the protein. A prominent example of the subtractive scheme is the ONIOM method[35]

The second way is additive, where the energy of each part and their interactions are added together:

$$E_{\mathrm{QM/MM,total}} = E_{\mathrm{QM,QM}} + E_{\mathrm{MM,MM}} + E_{\mathrm{QM-MM,total}} \quad \textbf{(2.14)}$$

where energy terms for QM part described by QM, MM part described by MM, and interaction between these parts are summed to obtain the total QM/MM energy. The interaction term is calculated via link atoms, which are dummy atoms bound to border QM atoms mimicking the

bond with neighboring MM atom. MM layer does not take these dummy atoms into consideration, so they do not influence the energy of the MM part.

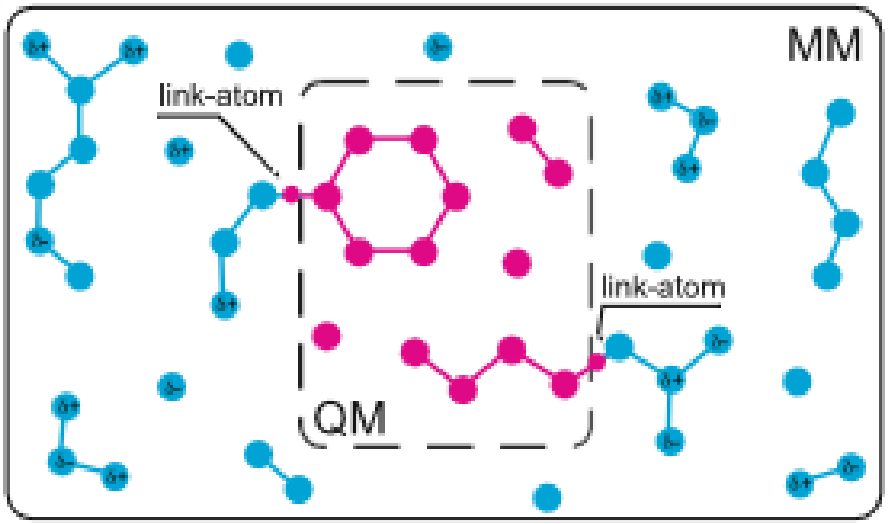

Figure 9 – Model of QM/MM method. The system is separated into QM and MM part with two link atoms on the borders, "connecting" those two subsystems.[36]

Until now, only *mechanical embedding* was considered. In other words, only the force, energy of the MM and how they affect the QM part is considered. One can also include the electrostatics of MM part atoms interacting with QM environment, which will be also included in one-electron QM Hamiltonian. This approach is called *electrostatic embedding.*[37] The QM electronic structure alters itself based on the charge distribution in the MM layer, resulting in polarization of the QM system. This correction of polarization of QM system provides better description of the system but with cost of increased computational part. Few methods introducing long/range electrostatic interactions are used, but the introduction of cutoff distance of the interaction brings unreliability because of artefacts. More flexible description for MM atoms charge is *polarized embedding,* which polarizes MM part based on QM charge distribution.

2.6. Solvation and relativistic effects

There are two additional effects to be considered in studies of metalloenzymes: solvation and relativity.

Proteins exist and function in condensed phase, be it water or complex cellular environment. Therefore, solvation effects play a crucial role. In principle, the condensed phase (water) can be modeled by adding explicit solvent molecules (*explicit solvation*). This might be a seemingly more realistic description. However, the calculations become more time demanding, and one also faces the conformational complexity of the solvent bulk. A conceptually different alternative is to use *implicit solvation,* in which the solvent is modeled as the dielectric continuum encapsulating the molecule.[25] In 1981, in their seminar work, Miertuš, Scrocco and Tomasi[38] showed that an effect of the dielectric continuum can be (equivalently) formulated as the boundary problem of partial charges on the surface of the solute's cavity. This leads to the set of coupled equations, the so-called self-consistent reaction field (SCRF) problem. Polarization of the solute by the dielectric continuum (represented by the surface point charges) and the response of the continuum to the change of the electron distribution in the molecule is thus solved iteratively until self-consistency is achieved.

In this work, we chose the COSMO (Conductor-like Screening Model) as the implicit solvation model[39]. COSMO uses surface segmentation on solvent-accessible surface to map polarity around the system. From the charge distribution of the system, we can calculate the charge of each surface segment and correct the energy.

Relativistic effects emerge as the consequence of the finite velocity of light. In chemistry, this mostly concerns heavier elements, since the electrons close to the nuclei – in inner shells - move faster, even 50% of the velocity of light. The heavier is the element, the less adequate is the non-relativistic description (c.f. SchrEq, 2.1) For quantum chemical calculations of typical organic molecules, the relativistic effects are mostly negligible. Starting with Z = 25-30, relativistic effects might play a more important role. In our calculations, this concerns copper ions present in the active site of CBC proteins. Thus, we employed the exact two-component method (x2c)[40,41] with appropriate (relativistic) basis sets. Two-component methods describe orbitals as two-dimensional wavefunction. Either vectorial component relates to one spin state. This enables the usage of spin-dependent model Hamiltonians but is computationally expensive. Alternatively, we may describe relativistic effects using pseudopotentials, also known as *Effective Core Potentials* (ECP). Using ECP, we replace inner electrons (which are

not part of chemical reactivity) by a potential calculated for every element. This is a relatively cheap description that may often work in cases with small change in the electron distribution of a molecule, but their usage is less recommended in systems with e.g. change of the oxidation state of the transition metal ion.

3. Aims

The aim of this thesis is a theoretical investigation of initial stages of the reaction mechanisms of two CBC enzymes:

Catechol oxidase (CaOx) and *ortho*-aminophenol oxidase (AOx). CaOx's are represented by PPO6 (polyphenol oxidase 6) whereas NspF (named after the gene coding this enzyme) is a representative of AOx's. We also compare PPO6 with other CaOx, the so-called IBCO (*Ipomoea batatas* catechol oxidase). The computed data will be correlated with Ty for which the reaction mechanism has been formulated and recently validated by joint experimental and theoretical efforts.

We expect to unravel the tiny differences in their reaction coordinates that may enable to understand the origins of their chemoselectivity. This also involves the computational investigation of the important **P**/**O** equilibrium and the role of singlet and triplet spin states in the **P** structure. In addition, the QM/MM equilibrium structure of PPO6 will be compared with the available IBCO X-ray structure data. Last but not least, CaOx (PPO6) will be also structurally compared with hemocyanin (Hc), an enzyme used as the dioxygen carrier, with no apparent catalytic function.

All in all, these are initial steps in obtaining insight into the origins of the different chemoselectivity of CBC enzymes that will be validated experimentally by our collaborators at Stanford University in future.

4. Methods and model setup

Since there are no X-ray structures of PPO6 and NspF, available we used for their QM/MM modeling predictions made by the AlphaFold 2.[42] The predicted structures were based on the homology with CBC enzymes with the known X-ray structures which included the IBCO structure[43] (PDB: 1BT3; Figure 5), studied herein for comparison. The reason why we did not carry out all QM/MM modeling (only) for the IBCO is that future experiments (see above) are planned on the PPO6 enzyme. All predicted structures (PPO6 and NspF) were assumed to be in the so-called *oxy*- form/state of the enzyme (Figure 2, right) that appears as the consequence of binding of dioxygen to the *deoxy*- form (Figure 2, left) and its concomitant 2-e^- reduction to peroxide (accompanied by the oxidation of copper ions). This is the state for which the crystal structures of the Ty and IBCO were available. Substrate (Figure 10) was missing in the initially generated crystal structure. All six histidines in the active site were protonated at proximal (with respect to the backbone) δ-nitrogen (denoted as Hid in the Amber's *prep* file).

For the QM/MM calculations, all structures were optimized employing DFT for the QM part of active site and MM for the rest of the protein. More specifically, we used the TURBOMOLE, ver 7.6[44] with the TPSS functional[31], def2-SVP basis set[45] and BJ dispersion correction.[46] MM part was described by using AMBER22[47] employing ff14SB force field.[48] After the optimization, we employed single-point energy QM/MM calculations by QM and MM parts being the same, but more advanced level of theory with x2c relativistic approximation for the QM subsystem. In more detail, we used the TPPSh functional[31], x2c-TZVPall[49] basis set, BJ dispersion correction and added implicit solvent model by conductor-like screening model (COSMO)[39] implemented in TURBOMOLE 7.6.

The key structures on the potential energy surface (reaction coordinate) were inspired by the published research on tyrosinase reaction mechanism[18]. Approximate transition state structures were obtained by the relaxed scans, employing restraints in the ComQum program, developed and implemented into ComQum by Rokob.[50] The substrate was docked manually into the active site, employing the experimentally verified ternary complex of Ty.[18] Each approximate transition state structure was carefully inspected to make sure it connects the desired points of the reaction coordinate.

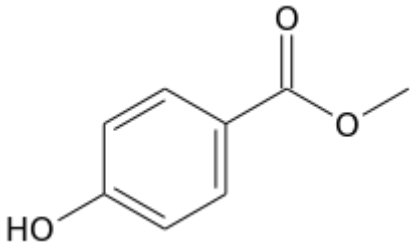

Figure 10 – Methyl ester substrate (methyl-4-hydroxybenzoate) used in our calculation of CBC proteins reactivity. The same substrate was used in tyrosinase reaction pathway mechanism. The compound was added to the active site in the way that hydroxyl group faced to the copper moiety so the proton transfer would occur.

5. Results and discussion

5.1. Spin state of the active site

First, the spin state (splittings) in our initial structures was calculated. In its oxidized state, the CBC enzyme's active site contains two Cu^{2+} ions with d^9 valence electrons, with one unpaired electron on each of them interacting. The coupling, efficiently mediated by the bridging ligand (sometimes denoted as super-exchange interaction[51]), gives rise to anti- or ferromagnetically coupled singlet and triplet states. In general, these are often close in energy. However, their reactivity and geometric structure may differ. Thus, we first focus on the energy differences of the spin states of the enzyme (which is almost exclusively confined to the active site)to find out the spin multiplicity of the electronic ground state (Table 1).

The calculations were performed on the **P** intermediate, without the substrate (Figure 11). Based on the theory, the value of $<S^2>$ should be equal to zero in the open-shell singlet state only for the exact eigenfunctions of the system's Hamiltonian. In an unrestricted single-determinant ansatz $<S^2>$ cannot be expected to vanish and these solutions are often more difficult to converge. From practical reasons, we adopted a typical broken-symmetry approach in which orbitals from a converged triplet state calculation are used as starting orbitals, by flipping the spin of one (or more in case of higher multiplets) α electrons.

The computed values of $<S^2>$ are listed in Table 1:

Table 1 – Energy difference between singlet and triplet state for NspF and PPO6 and calculated $<S^2>$ for each state.

	NspF	PPO6
ΔE (Triplet - Singlet) [kcal/mol]	7.1	8.9
$<S^2>$ for singlet	0.68	0.60
$<S^2>$ for triplet	2.01	2.01

As can be seen in Table 1, the value of <S2> for the (symmetry-broken) singlet is not zero because the UKS 'wavefunction' is not an eigenfunction of the S2 operator. The depicted positive values indicate singlet to be the ground state of the studied systems. This trend is in compliance with Ty reactivity studies suggesting singlet to be also the ground state. From now on, every calculation we did was in unrestricted open-shell singlet.

5.2. Side-on-peroxo and bis-μ-oxo equilibrium

The **side-on-peroxo $[Cu^{II}_2(O_2^{2-})]^{2+}$ ↔ bis-μ-oxo $[Cu^{III}_2(O^{2-})_2]^{2+}$ (P ↔ O)** equilibrium is the hallmark of the CBC reactivity, both for CBC enzymes, as well as for their smaller biomimetic counterparts (Figure 11). In the past, this reaction has served as the common test for the performance of the density functional as well as multi-reference wave function methods[15,52]. In catalytic terms, both forms exhibit different reactivity and both can be operational in the reactions catalyzed by the CBC systems.

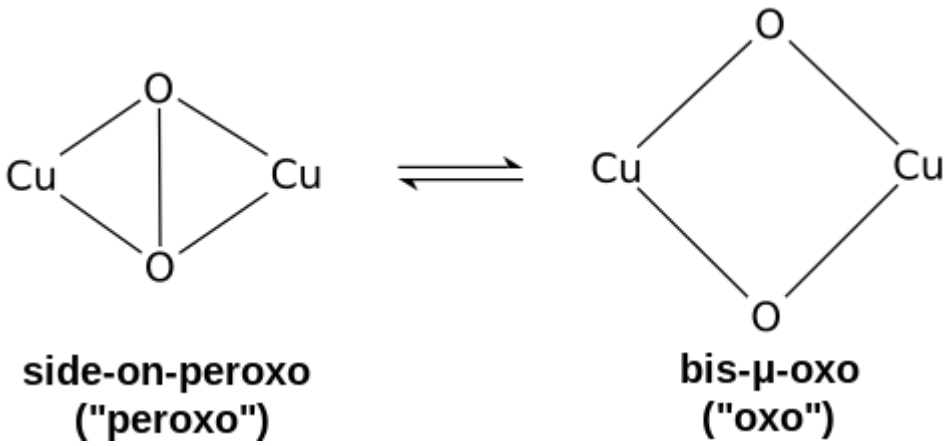

Figure 11 – Schematic representation of side-on-peroxo and bis-μ-oxo Cu_2O_2 models considered in this work.

Therefore, we evaluated the energy differences of **P** and **O** forms/structures. In analogy with the reaction coordinate of the Ty, we have started with peroxo structure as suggested by Kiporous[18]. The computed data are compiled in Table 2:

Table 2 – Energy difference between oxo and peroxo structure for NspF and PPO6

	NspF	PPO6
ΔE (O - P) [kcal/mol]	11.0	1.3
$\langle S^2 \rangle$ for P	0.68	0.60
$\langle S^2 \rangle$ for O	0.18	0.00

As can be seen in Table 2, the computed energy differences strongly indicate that the **P** structure is energetically favorable for the NspF, to a much greater degree than in PPO6 ($\Delta\Delta E_{S\text{-}T}$ (NspF/PPO6) = 9.7 kcal/mol) and also in Ty (data not shown). Thus, it can be safely concluded that, in analogy with Ty, the **P** structure of NspF intermediate is the structure of the oxy form of NspF without the substrate. The same **P** ↔ **O** equilibrium is somewhat different for PPO6. The computed difference of 1.3 kcal/mol is very small and the **O** structure might be competitive (operational) in PPO6, too much greater degree than in NspF. The reaction (conversion of L-tyrosine to quinone) starts from the **P** structure in the case of Ty. PPO6 is not able to catalyze monophenol hydroxylation reaction, and this energy difference can be one of the reasons. We may speculate that the greater stability of **O** can lead to higher TS2 state in monohydroxylation reaction. This will be elaborated in more detail below.

5.3. Reaction coordinates

In this chapter, we will thoroughly analyze and discuss each step in the plausible catalytic mechanism(s) of the studied enzymes. We will focus on comparison of the reaction coordinates of NspF and PPO6, which may reveal the differences in the reactivities.

The computed mechanism is shown in Figure 12-1). It is based on previous studies on Ty.[18] It has been shown to beginwith **1P**, which is the protein structure with substrate (protonated phenol) placed close to the peroxide of the Cu_2O_2 core. The first transition state then is denoted as **TS1**. It corresponds to the breaking of the O-O (1.45 Å in **1P**). The distance reaches a maximum value of 2.4 Å for **1O (bis-μ-oxo)**. Next, the proton transfer from substrate's hydroxyl group to oxygen of the copper moiety leads to the **2O** structure. Through the **TS2** transition state, the phenolate is getting coordinated, via its oxygen atom, to one of the coppers (CuA, c.f. Fig. 12-2) in the active site. Simultaneously, a bond is formed between carbon of the benzene ring and the second bridging oxygen (non-protonated). This results in the formation of the complex **3O**. The individual steps will be discussed in more detail below.

As we continue along the reaction coordinate towards the **3O** structure, we first observe the lowering of of $<S^2>$ around the **1O** intermediate, after this step the value of $<S^2>$ increases. It is well known that the spin polarizability is dependent on the amount of exchange in the used functional. The values in Table 1 refer to the TPPSh hybrid functional which is expected to yield slightly higher values of $<S^2>$.

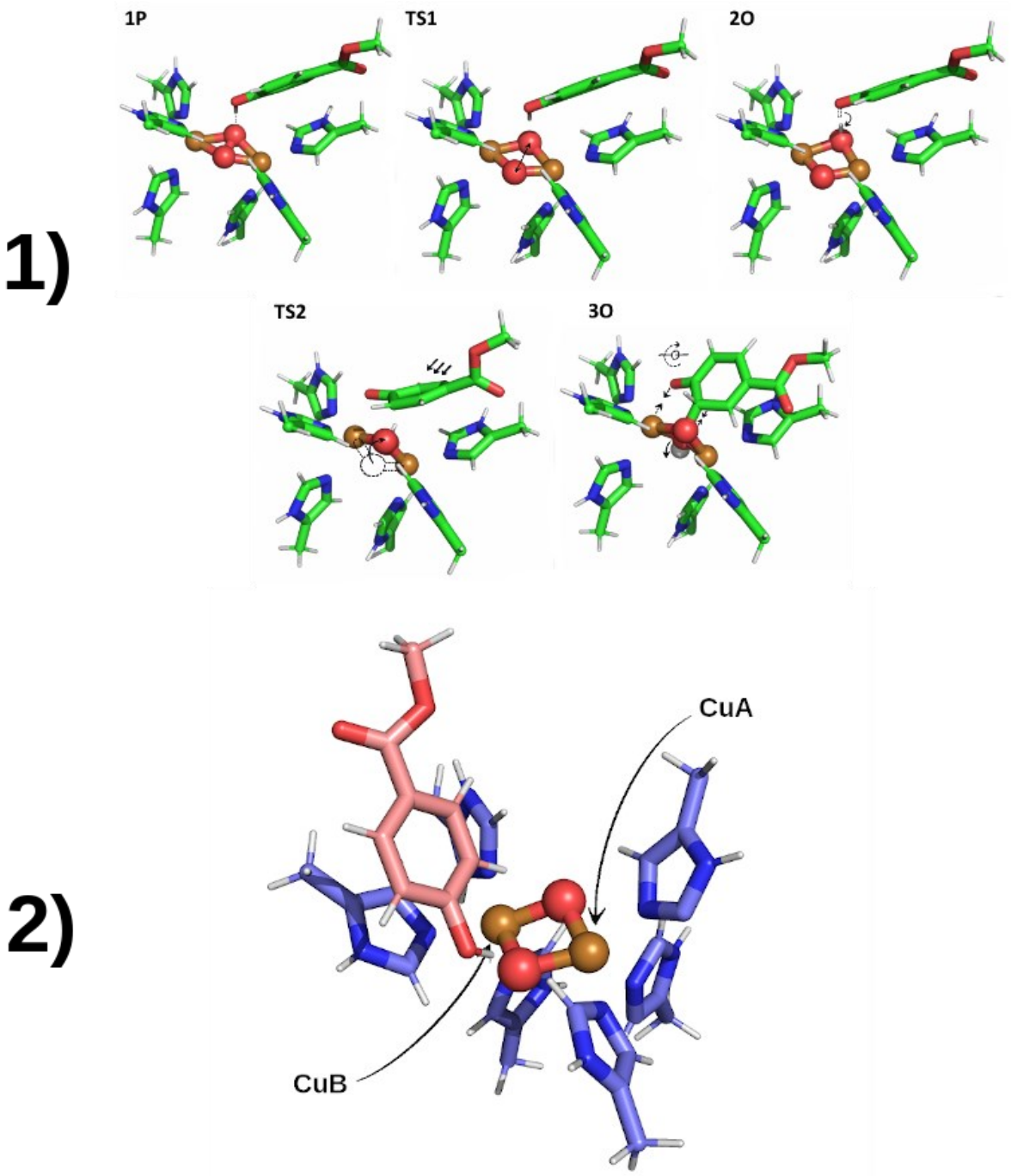

Figure 12 - 1) Geometries of active sites in the individual steps from 1P to 3O. Arrows show the changes from its previous intermediate. 2) Positions of CuA and CuB in the active site from the same point of view as in the future chapters (picture taken from the 1O structure of PPO6 core).

5.3.1. Ternary Intermediate (**1P**)

In analogy with Ty (see above), the structure of the ternary intermediate {enzyme + O_2 + substrate} was investigated first, for both PPO6 and NspF. We presumed **1P** to be the representative of ternary intermediate (Figure 13).

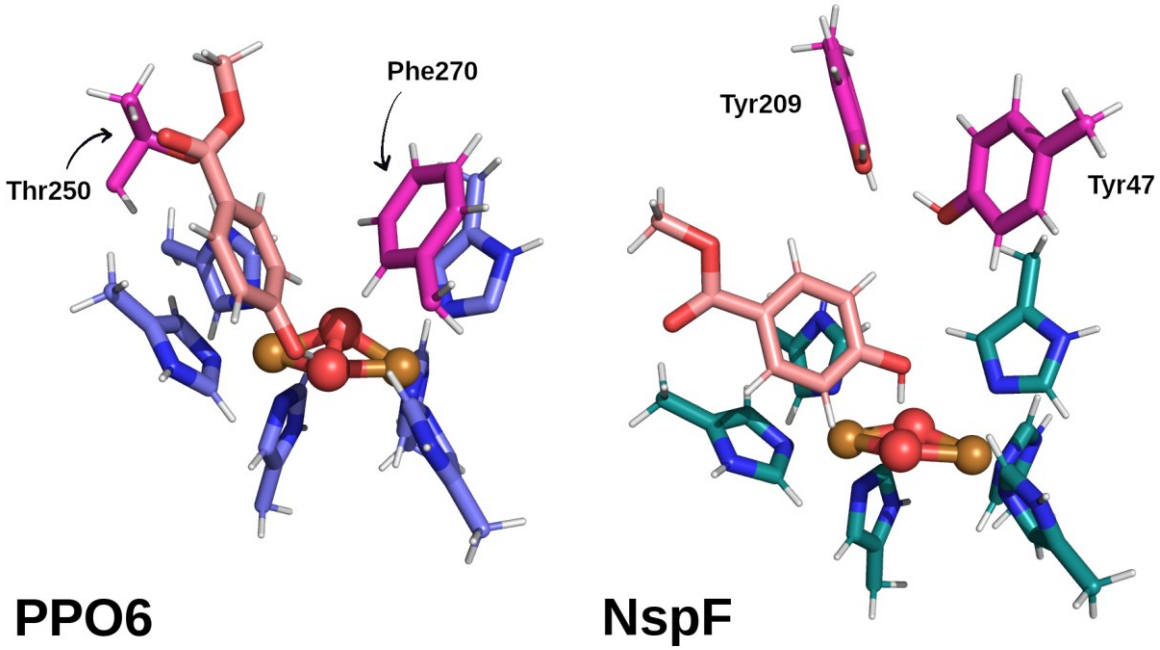

Figure 13 - Initial 1P structures for PPO6 and NspF with substrate (salmon). Main structural differences near active site are colored in magenta. Some amino acids were hidden for clarity.

We can see that the coordination of Cu2O2 moiety to six histidines is very similar. The distances between the two copper(II) cations in PPO6 and NspF are the same (R_{Cu-Cu} = 3.65 Å). The O-O bond length of 1.45 Å corresponds to peroxide. For PPO6, the Cu-O-Cu angle is 129° for both angles, for NspF it is 142° and 126˚. It is noteworthy that oxygen molecule is somewhat pushed towards the entrance of the catalytic site in PPO6, resulting in greater *butterfly (torsion) angle* (Cu-O-O-Cu). Respectively, the values are 149° for PPO6, 161° for NspF and 142° for Ty butterfly angles.

One of the general problems in QM/MM modeling that needs to be solved is the number and positions of water molecules within the protein active site. They might not be always resolved, except for high resolution crystal structure, there is always a dynamic equilibrium between waters and enzyme with quite different retention times (see the study of Kiporous et al. for Ty[16]). In this study, we placed one water at the entrance to the active site near the conserved Glu residue (Glu245 in PPO6, Glu207 in NspF). PPO6 then has two additional water molecules (Figure 14).

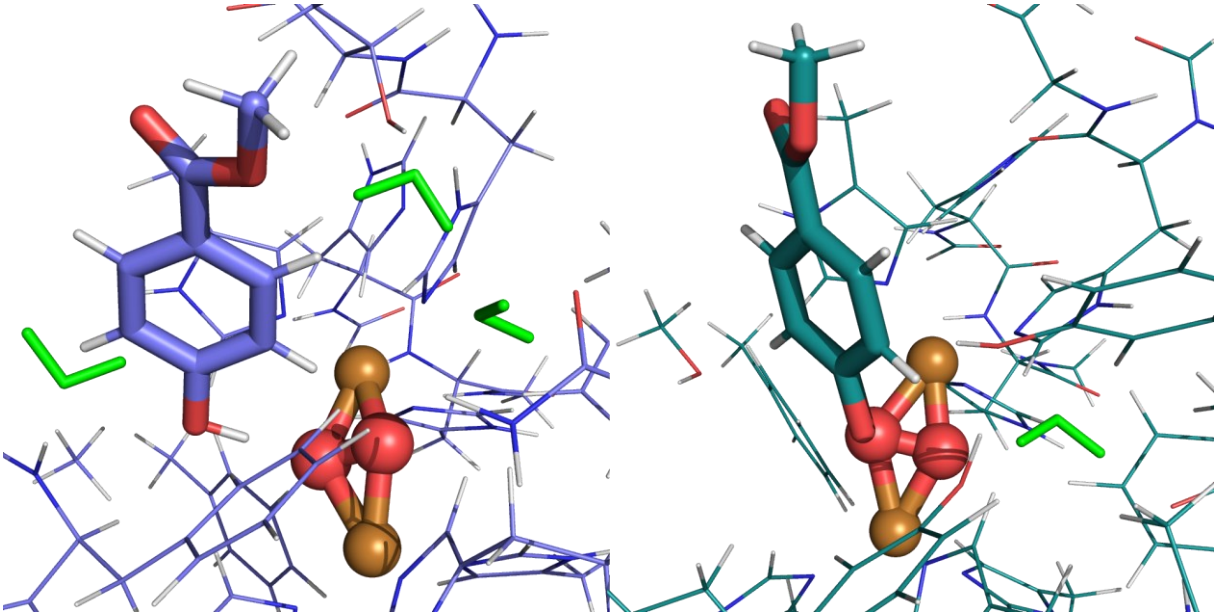

Figure 14 – Position of waters in 1P structures of PPO6 (left) and NspF (right) with highlighted substrate and active site

It can be mentioned that in the reaction pathway of Ty, Kiporous et. al.[18] did not see any influence of water molecules. Hydrogen from phenolic -OH group of the substrate is directed towards the proximal oxygen of peroxide, creating a hydrogen bond, in both NspF and PPO6. However, there are a few amino acids which differ between the two active sites. Phe270 in PPO6 seems to work as a gate residue, blocking the entry to the catalytic site for the substrate (see Fig. 16). Instead of this phenylalanine, NspF has two additional tyrosines (Tyr47 and Tyr209). Their role in the active site is still unknown. It was suggested by Solem, Tuczek and Decker[53] that Phe270 in polyphenol oxidases may block the entrance to the active site and cause the CaOx's to lack the monophenolase activity. In our opinion, this seems to be plausible hypothesis, because it was shown in the previous study on Ty[18], that the substrate rotates during later catalytic steps. This might be prohibited in CaOx by the mentioned Phe residue. However, this was partially disproved by Prexler and co-workers[54], who exchanged mutated Phe in PPO for Leu, Pro, and Gly. Without the gate Phe, PPO lost all catalytic activity (with exception of

Leu that reduced the activity by two thirds). In chapter 5.3.3. we will discuss our (alternative) suggestions concerning the effect of Phe on the substrate binding and catalysis. In addition to these differences, there are two additional neutral residues in the same conformation which were discussed in the already mentioned papers, Thr250 in PPO6 and Gly223 in NspF. We cannot make any conclusions about the catalytic roles of these residues based on the current study. Finally, Tuczek[53] suggested that Asn190 in Ty, located at the same position close to the above mentioned (conserved) glutamate, is needed for the enzymes monophenolase activity. In contrast, the results of Prexler indicated that this amino acid does not need to be charged[54] for the hydroxylation activity. Analysis of NspF structure confirms Prexler's statement because it is known NspF can catalyze the reaction while having glycine in the same spot as Asn190 from Ty.

Having "locked in" these initial structures, we began our scan of reaction pathway energies to find the differences in CBC enzymes catalytic activity.

5.3.2. Conversion of **1P** to bis-μ-oxo **(2O)** reactive intermediate

There are few steps in the reaction mechanism before **1P** reaches the **2O** structure, (via **TS1** depicted in Fig. 11). First, the bond between the oxygen molecule in Cu_2O_2 core elongates (up to the 2.4 Å in our model), resulting in **1O** structure (Fig. 17). Then, the proton from the substrate's hydroxyl group moves to one of these oxygens turning substrate into a phenolate. Between the **1P** and **1O**, we made five additional (sequential) steps to find transition state (**TS1**) between these two geometries, scanning the O-O bond (1.75, 1.90, 1.95, 2.00, 2.15 Å).

Table 4 – Energy differences between 1P and chosen steps of the reaction from 1P to 2O structure.

(energy differences are in kcal/mol)	**NspF**		**PPO6**	
	ΔE	$<S^2>$	ΔE	$<S^2>$
1P	-	0.61	-	0.82
1,75 Å	11.1	0.61	8.9	0.63
1,90 Å	15.9	0.34	11.4	0.22
1,95 Å	15.5	0.30	9.9	0.21
2,00 Å	12.9	0.33	8.4	0.23
2,15 Å	12.4	0.32	6.2	0.15
1O	12.8	0.33	6.5	0.00
2O	9.8	1.26	10.5	1.00

As can be seen in Table 4, the highest point (transition state along the O-O coordinate) corresponds to the O-O distance of 1.90 Å. Expectedly, $<S^2>$ decreases on the way from **1P** to **1O**. Thus $[Cu_2O_2]^{2+}$ changed from "peroxo" to "oxo" structure and the formal charge of the oxygens and coppers changed from -1 to -2 and from +2 to +3, respectively.

It is also worth mentioning that the presence of substrate shifts the **P** ↔ **O** equilibrium towards the **P** structure, compared to the oxy- (substrate free) forms of the enzymes discussed above (c.f. Figure 15 for **1O**). Next step is hydrogen (proton) transfer (**1O** to **2O**). Surprisingly, this leads to stabilization of the NspF **1O** intermediate and and its destabilizion PPO6. This suggests yet another plausible explanation why PPO6 is unable to catalyze the hydroxylation. Based on calculated energies and spin states, the substrate needs to bind to the catalytic site in the form of phenolate, to be fully activated for catalysis. Less favorable **2O** structure may render the enzyme unable to catalyze the hydroxylation.

Structurally, during the transition from **1P** to **1O**, the bond between the oxygens prolongates (see above), whereas the distance between the copper ions is shortened to 2.9 Å, both for PPO6 and NspF. The angle between the atoms Cu-O-Cu changed significantly (99° for PPO6 and 103° and 97° for NspF). The butterfly (torsion) angles changed for both structures (166° in both PPO6 and NspF). Amino acids surrounding the active site did not significantly move, even when the position of coppers and oxygens changed.

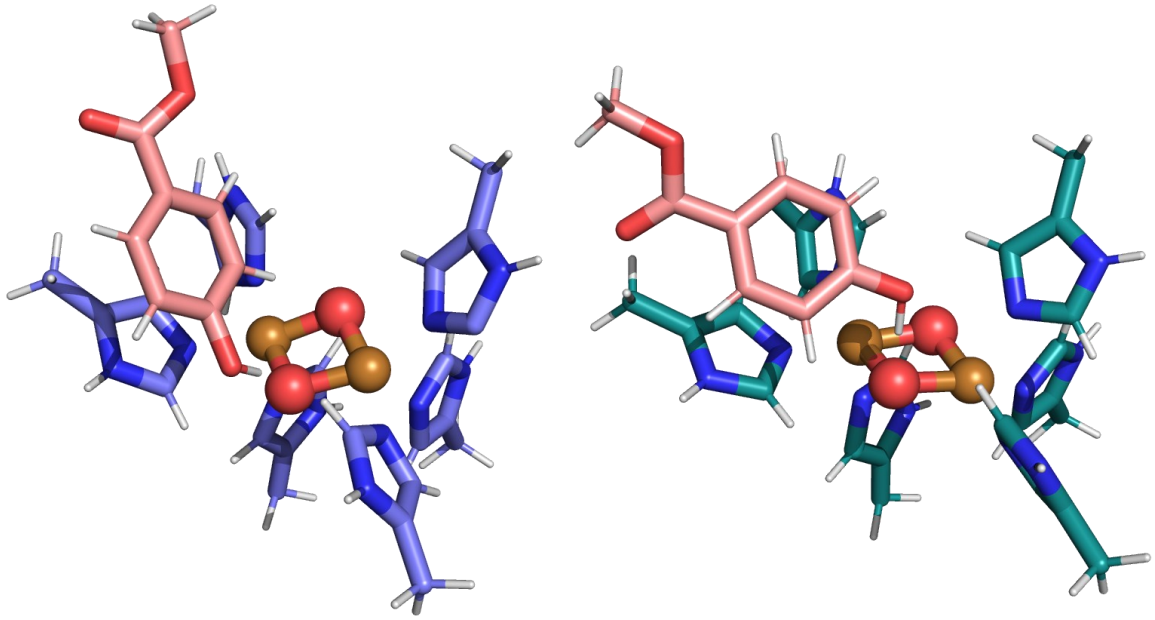

Figure 15 – Oxo structure (1O) of both geometries of PPO6 and NspF.

The **1O** → **2O** conversion is associated with a considerable increase of $\langle S^2 \rangle$ values, suggesting radical character of **2O** intermediate. This is also supported by calculated Mulliken spin densities summarized in Table 5.

Table 5 – Mulliken spin population analysis for the atoms of $[Cu_2O_2]^{2+}$ at different steps of the oxidation pathway. Positive values indicate the accumulation of α-spin electrons and negative values indicate the accumulation of β-spin electrons at atoms. O(H) stands for the oxygen to which the hydrogen from the substrate binds to.

	PPO6					NspF				
Mulliken spin densities	CuA	CuB	O	O(H)	sum	CuA	CuB	O	O(H)	sum
1P	-0.51	0.51	0.00	0.01	0.01	-0.46	0.46	-0.01	-0.01	-0.02
TS1	-0.22	0.3	-0.07	-0.02	-0.01	-0.37	0.26	0.10	0.03	0.02
1O	0.00	0.00	-0.01	0.01	0.00	-0.12	0.01	-0.10	0.09	-0.12
2O	-0.44	0.18	-0.44	-0.06	-0.76	-0.56	0.25	-0.13	-0.52	-0.96

Indeed, the computed spin densities for **2O** indicate that the reaction proceeds through a radical intermediate. Therefore, instead of previously mentioned proton transfer, the **1O** to **2O** conversion should be referred to as hydrogen transfer instead.

To ensure **1P** structure is indeed in the open-shell singlet occupation, we chose to visualize the spin density of the core, which can be seen in the Figure 16.

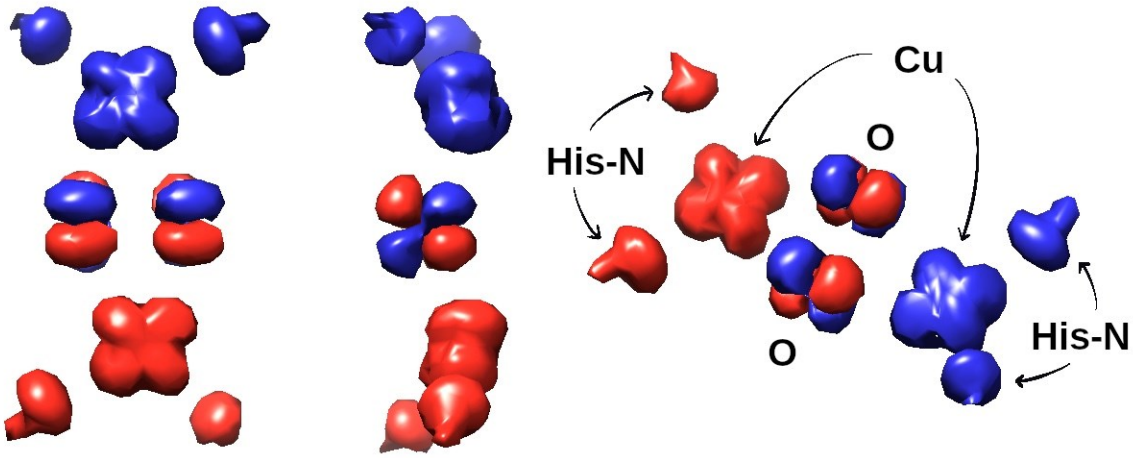

Figure 16 – Spin density of the active site of 1P structure of PPO6. Every picture represents different point of view. Different spins are visualized by different colors.

This figure depicts the spin densities of the Cu_2O_2 core and four vertical histidine nitrogen atoms. This illustrates that our calculation of **P** structure was indeed done in open-shell singlet as we were expecting.

5.3.3. Oxyl transfer to *ortho*-carbon (**2O** → **3O** pathway)

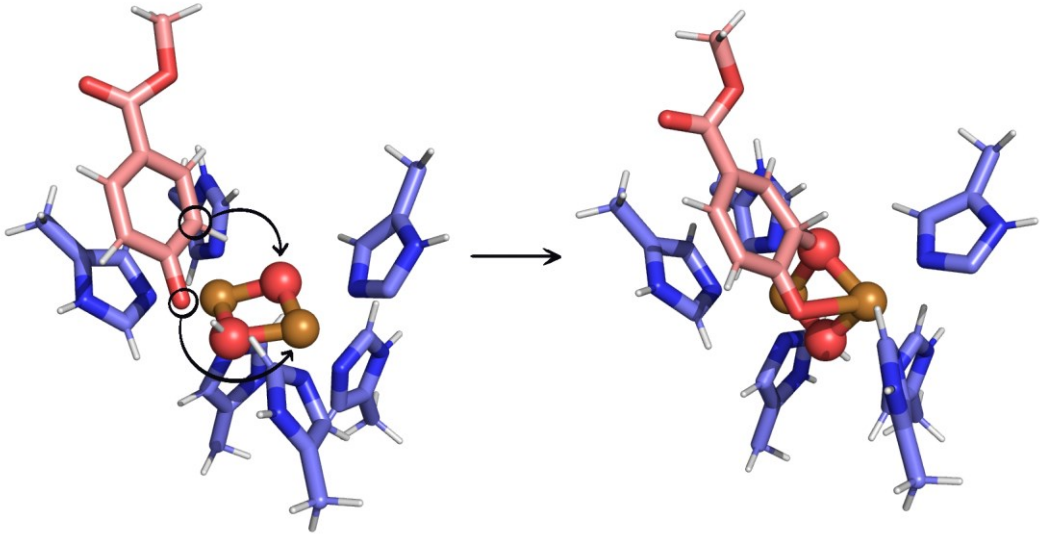

Figure 17 - Mechanistic differences between 2O and 3O structures of PPO6. Phenolic oxygen bids to CuA and carbon in ortho-position compared to the oxygen binds to the oxygen bound in the active site.

The **2O** to **3O** transformation is characterized by phenolate binding to the $Cu_2O_2(H)$ core (Figure 17). It has been previously described[18] by the 2D reaction coordinate (phenolate O to Cu_A distance and phenolic *ortho*-carbon to non-protonated oxygen/oxyl of the $Cu_2O_2(H)$ core distance.

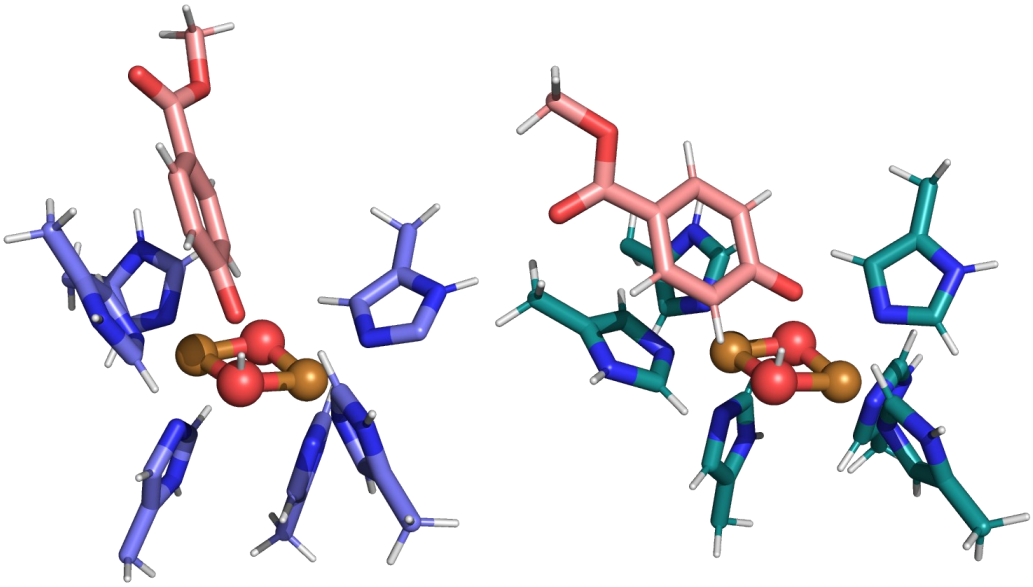

Figure 18 – 2O geometry in both studied enzymes, PPO6 (left) and NspF (right). Both images are shown from the same angle to copper moiety to emphasize the change in position of substrate.

While the substrate approaches the active site, it rotates into the proper orientation with respect to the Cu_2O_2(H) core (Figure 18). In our QM/MM modeling, this rotation was observed to occur in NspF whereas no rotation was computationally observed in the PPO6 structure, in which the substrate was only getting closer to the catalytic core. The absence of rotation may be caused by Phe270, which not only sterically blocks the entrance to the active site but also interacts via the π-stacking with the substrate. The substrate in PPO6 binds so that its aromatic ring is parallel to aromatic histidine side chain (His253) and stacks with Phe270 in T-shape. As the oxyl attacks the *ortho*-carbon, the aromatic ring of the substrate is no longer planar but becomes a bit twisted due to the newly formed bonds. Hydrogen on the *ortho*-carbon moves into almost perpendicular position with respect to the ring (marked by green color in Figure 18). Upon formation of the C-O bond, CuA is coordinated by six ligands, which results in dissociation of the histidine in the axial position (His121 for PPO6 and His67 for NspF, highlighted in Figure 19 by pink color). Hydrogen bound to the oxygen in the copper moiety flips below the plane of the Cu_2O_2 core, so it does not hinder substrate binding. The energy profiles for the approximate 2D scan are reported in Table 6.

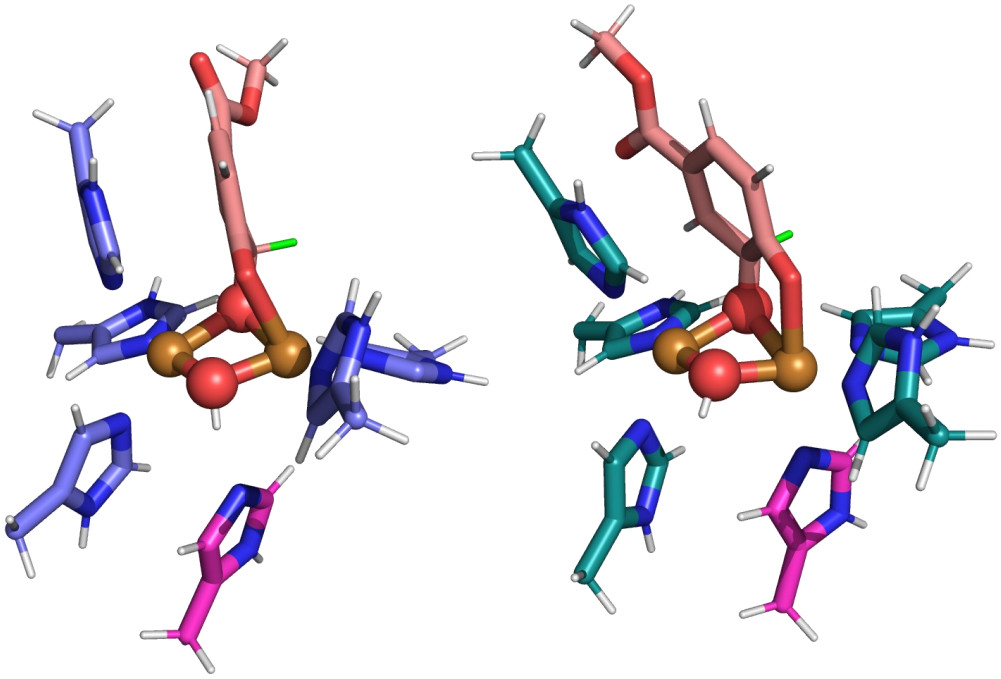

Figure 19 - Geometries of the final 3O structures for PPO6 and NspF. Important new features of the structure were highlighted in different colors - dissociating histidine is in pink and hydrogen perpendicular to the structure of substrate is marked by green color.

Table 6 – Calculated energies of possible transition state structures. Row shows the distance between the CuA and phenolic O, column shows the distance of non-hydrogenated active site oxygen with ortho-carbon of substrate. Energy differences are still compared to 1P (as in previous calculation) in kcal/mol. Results in blank spaces did not converge or were not finished until the date of submitting the diploma thesis.

NspF					
Cu-O	**C-O**				
(energy differences are in kcal/mol)	**1.9**	**2.0**	**2.1**	**2.2**	**2.3**
2.5	-5.3	-0.3	4.1	6.7	**10.5**
2.6	-5.9	-0.6	3.6	6.5	
2.7	-6.3	-0.9	3.3	6.2	9.7
2.8		-1.2		*19*	9.3
2.9	-6.7		5.5	5.8	9.8

PPO6					
Cu-O	**C-O**				
(energy differences are in kcal/mol)	**1.9**	**2.0**	**2.1**	**2.2**	**2.3**
2.5	16.1	21.8	25.9	**27.1**	26.5
2.6	15.8	21.6	25.7	27.0	26.3
2.7	15.6	21.4	25.7	26.8	26.2
2.8	15.3	21.3	25.6	26.9	26.1
2.9	14.9	21.1	25.5	26.9	26.2

As can be seen the difference in the computed ΔE largely differs. NspF, with its transition state around 2.5 Å for Cu-O and 2.3 Å for C-O distances, is lower in energy by 16.6 kcal/mol than PPO6 for 2.5 Å for Cu-O and 2.2 Å for C-O distances (considered as TS2). Values compared between enzymes differ by 20 kcal/mol on average, which may be an important aspect of various enzymatic activity for these enzymes. At the moment, this difference seems to be too big and would need further investigation. Would the numbers hold future tests of robustness, it will explain the absence of monophenolic activity of PPO6. It can be also seen that energies variance is considerably larger for the C-O distance rather than Cu-O distance.

During the calculation of NspF's TS2, we found two ways how substrate is affected by the changes of scanned distances. Substrate's ester *para*-group position may be rotated in two ways (Figure 20). These two conformations considerably differ in energy. For the comparison, the NspF structure defined by distances of 2.9 Å for Cu-O and 2.3 Å for C-O has energy difference of 11.O kcal/mol. Therefore, the previous calculation probably converged to the wrong structure. The structures are possibly in two different local minima where the conformer in calculation of the first geometry (in Fig. 19 marked by green color) is less stable (higher in energy). The energy values of few calculations from 2D scan of NspF's TS2 were not included in the Table 6, because of mentioned different converged minimum (this values is highlighted in *italic*) or not converged at all.

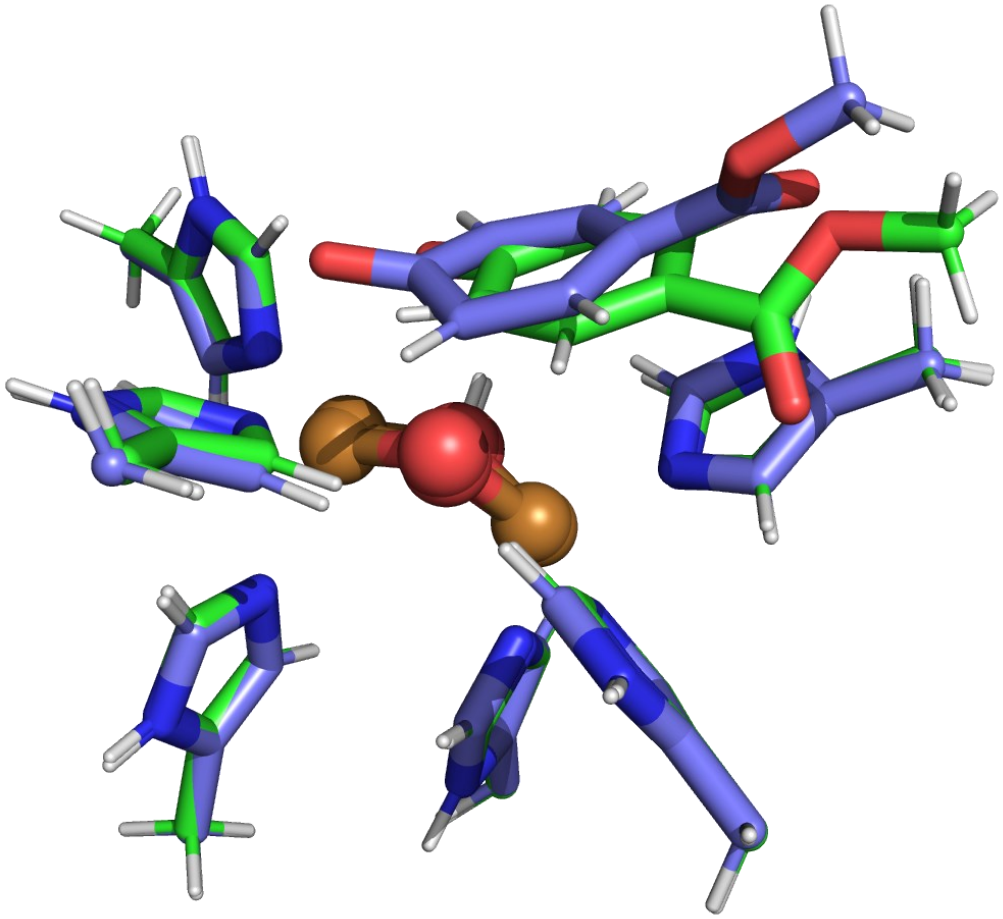

Figure 20 - Comparison of substrate orientation between the first and corrected NspF transition state calculations. Green color of the different conformer was chosen to distinguish the structures.

3O structure is the final intermediate of the studied reaction coordinate (in analogy with the Ty studies of Kipouros and Stanczak[18]). For Ty, a careful experimental and theoretical investigation of the hydrogen transfer from the sp^3 *ortho*-carbon to the oxygen to obtain final product of the hydroxylation, viz monocatecholate, is under way. For NspF and PPO6, the same mechanistic correlations will be done. The calculations predict that the between **1P** to **3O** transformation is more exothermic for NspF than for PPO6 (Table 7). The Ty's 3O structure is lower in energy (-19.2 kcal/mol) compared to its 1P than in case of PPO6, creating more stable intermediate to continue with the reaction. However, NspF's results suggest sp^3 carbon bonding to stabilize more than in Ty case (NspF's 3O structure is 12.2 kcal/mol lower in energy). In the future, we will continue with the study of NspF reaction pathway with precise descriptions from experiments to prove/correct computed mechanism values.

Table 7 - Values of energy difference between 3O structures of mentioned enzymes and their 1P structures with $<S^2>$ values

	NspF	PPO6
ΔE (3O - 1P) [kcal/mol]	-35.7	-17.7
$<S^2>$ for 3O	0.99	0.96

As mentioned above, the complete reaction pathway of Ty (as one of the most studied CBC enzymes) from L-tyrosine to L-dopaquinone is still not fully understood, even though various studies were trying to characterize it[23]. Our ongoing studies on Ty suggest that the reaction should further proceed by the proton transfer from *ortho*-carbon to one of the residues in the proximity of the active site, such as water molecules or histidines. They can act as possible acceptors for transferred proton.

5.3.4. Modelling issues: comparing two models of catecholoxidases (PPO6 and IBCO)

To ascertain the robustness of our PPO6 model prepared using AlphaFold2, the existing X-ray structure of very similar CaOx (IBCO) has been utilized. Functionally, the two enzymes are equivalent, as none of them can catalyze monophenol hydroxylation reaction. Thus, we expect their reaction coordinates to be similar. Their comparison is made in in Table 8 and Figure 21. It should be mentioned that for IBCO, we did not perform all QM/MM scans to obtain every transition state. The distance restrains were based on previous Ty studies.[18]

Table 8 - Energy differences of PPO6 and IBCO enzymes relative to 1P. TS2 in with the structures with fixed with distance of 2.6 Å for Cu-O and 2.2 Å for C-O.

(energy differences are in kcal/mol)	**PPO6**	**IBCO**
1P	-	-
TS1	11.4	12.9
1O	6.5	3.1
2O	10.5	10.5
TS2	27.0	23.3
3O	-17.7	-23.5

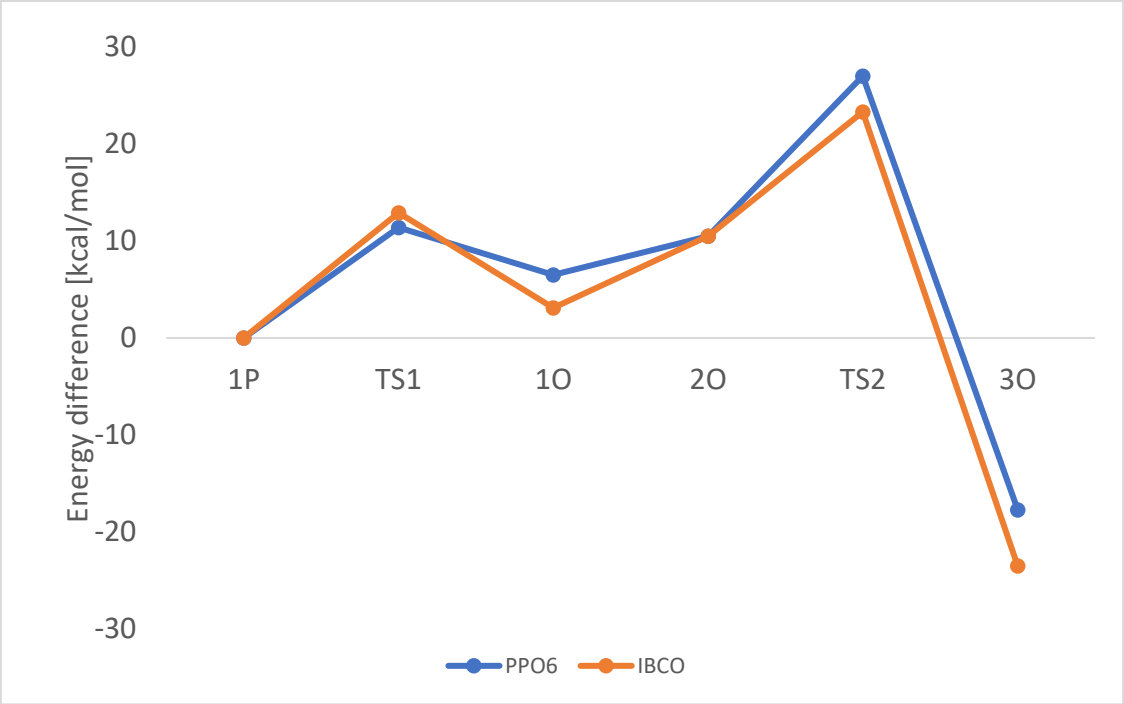

Figure 21 – Plot of the energy differences for PPO6 and IBCO compared to enzyme's 1P structure during the reaction course

For TS1, the difference of 1.5 kcal/mol is within the error of the composite QM//QM/MM protocol . However, there is a difference of 3.4 kcal/mol for 1O. Comparison of both 1O geometries showed differences in position of substrate and angle of Phe (Phe270 in the case of PPO6 and Phe261 in the case of IBCO) next to the substrate (Figure 22).

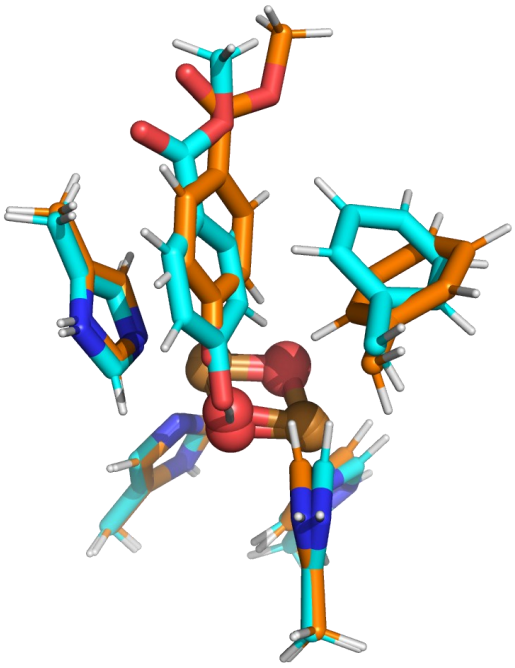

Figure 22 - Overlap of IBCO (tuequoise) and PPO6 (orange) 1O active site structures. Two histidines were hidden for the clarity of Phe and substrate.

These geometries were converged with the constraints to O-O bond distance as previously mentioned and afterwards optimized without the constraints. The different angle of Tyr next to the substrate can stabilize the substrate by π-stacking interaction. Tyr from IBCO structure most likely rotates to interact with the monophenol by T-stacking interaction. The tyrosine gets closer to the substrate. This difference in structure may be the origin of the energy difference.

The biggest energy difference is seen in case of **3O** structures. There is no significant geometry difference of amino acid residues to cause the ΔE of 5.8 kcal/mol. One factor contributing to this difference could be the two water molecules near the entrance to the active site. Hydrogens of these water molecules point to nearby atoms (in this case oxygens) with partial negative charges (Figure 23). In addition, we did not focus on the hydration level (amount of water molecules) in our QM system, which resulted in PPO6 having one more water molecule in the active site than IBCO. All these aspects may contribute to the energy difference.

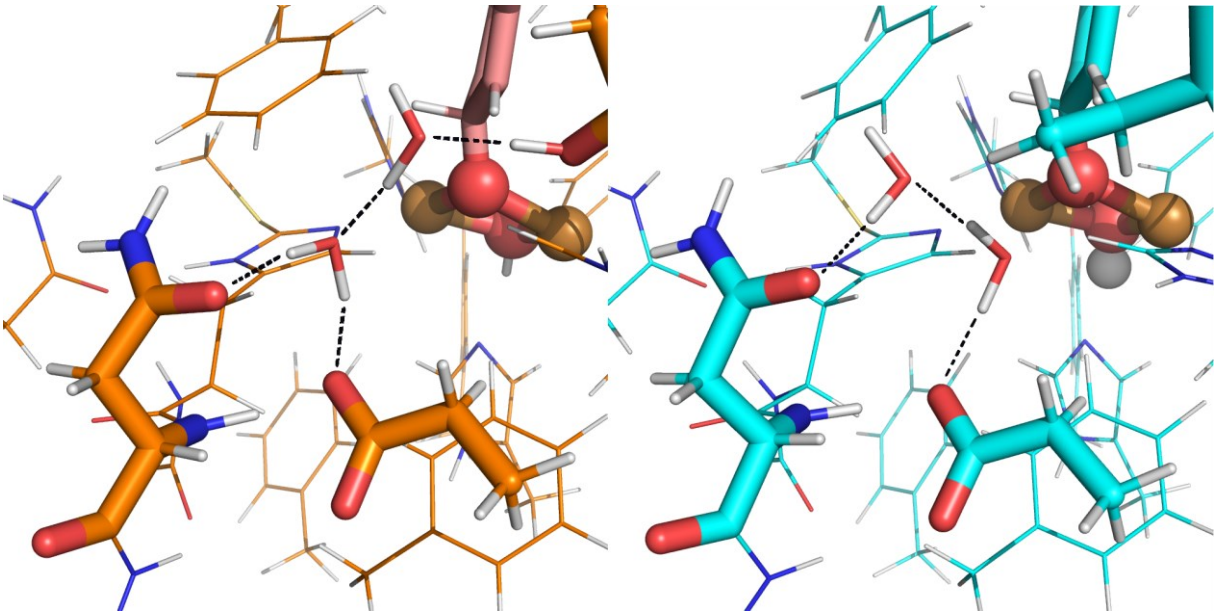

Figure 23 – Position of water molecules near the active site of PPO6 (orange) and IBCO (turquoise), and the hydrogen bonds between the water molecules and nearby residues.

5.4. Why do catechol oxidases not hydroxylate phenolic substrates?

The energy differences from individual steps of reaction mechanism among all studied CBC enzymes are seen in Table 9 and Figure 24.

Table 9 - Energy differences in each step for every enzyme. For NspF, the highest obtained point from 2D scan was chosen to represent TS2 structure.

(energy differences are in kcal/mol)	PPO6	IBCO	Ty	NspF
1P	-	-	-	-
TS1	11.4	12.9	12.7	15.9
1O	6.5	3.1	4.7	12.8
2O	10.5	10.5	8.4	9.8
TS2	27.0	23.3	22.7	10.5
3O	-17.7	-23.5	-19.2	-35.7

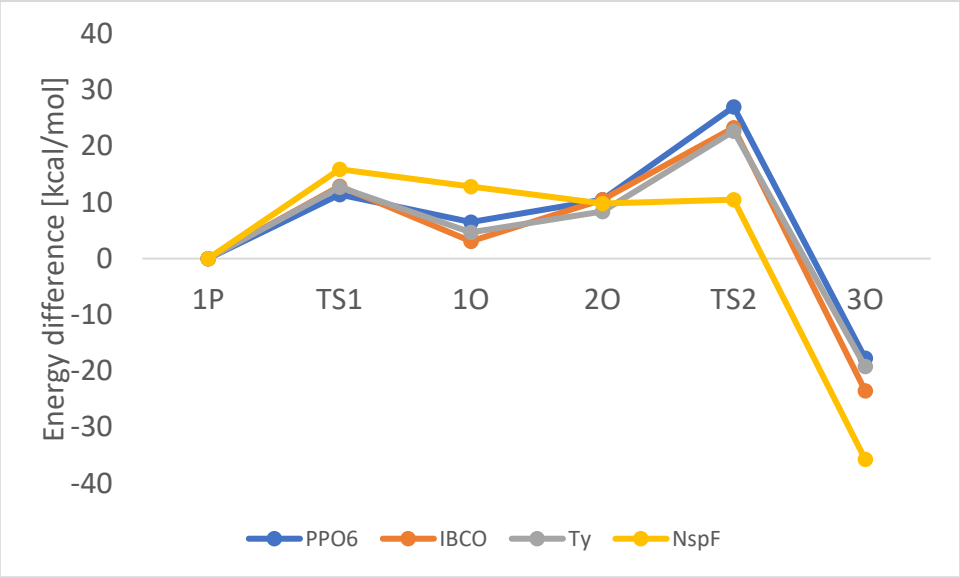

Figure 24 – Energy differences for all calculated pathways relative to enzyme's 1P structure during the reaction pathway. For NspF, the highest obtained point from 2D scan was chosen to represent TS2 structure.

PPO6's **TS2** structure is higher in energy compared to Ty by 4.3 kcal/mol. This energy difference may influence the ability to catalyze the hydroxylation of monophenols. By comparing IBCO with Ty, the energy barrier of this step is still higher (0.6 kcal/mol). However, IBCO's **TS2** was not studied by 2D scan as in the case of other enzymes. Therefore, we cannot certainly mark chosen conformer to be "real" transition state, rather approximate one. Precise 2D scan is needed to find real transition state. On the otehr side, NspF's **TS2** structure is 12.5 kcal/mol lower in energy than Ty, which is an unexpected result. This would indicate NspF having higher catalytic activity for this type of reaction, which it should not have.[55] These results raise the question whether the calculated energy values of suggested CaOx enzymes are high enough for enzymes not being able to catalyze studied reaction. During the reaction course, additional aspects like binding constant or number of molecules in the active site may affect enzymes' activity, which were not studied in this work. Also, the influence of water molecules in active site (number of molecules and their interaction with residues) was not studied. Such effect could alter the results. Experimental verification should be done to confirm the theoretical prediction.

5.5. Available 3-D structures (X-ray structures of IBCO and hemocyanin vs. the predicted PPO6 structure)

To assess the differences in biological activity from the structural point of view, we compared structure of PPO6 predicted by AlphaFold 2 with X-ray structures of IBCO (PDB code: 1BT3[43]) and hemocyanin (Hc; PDB code: 1OXY[56]).

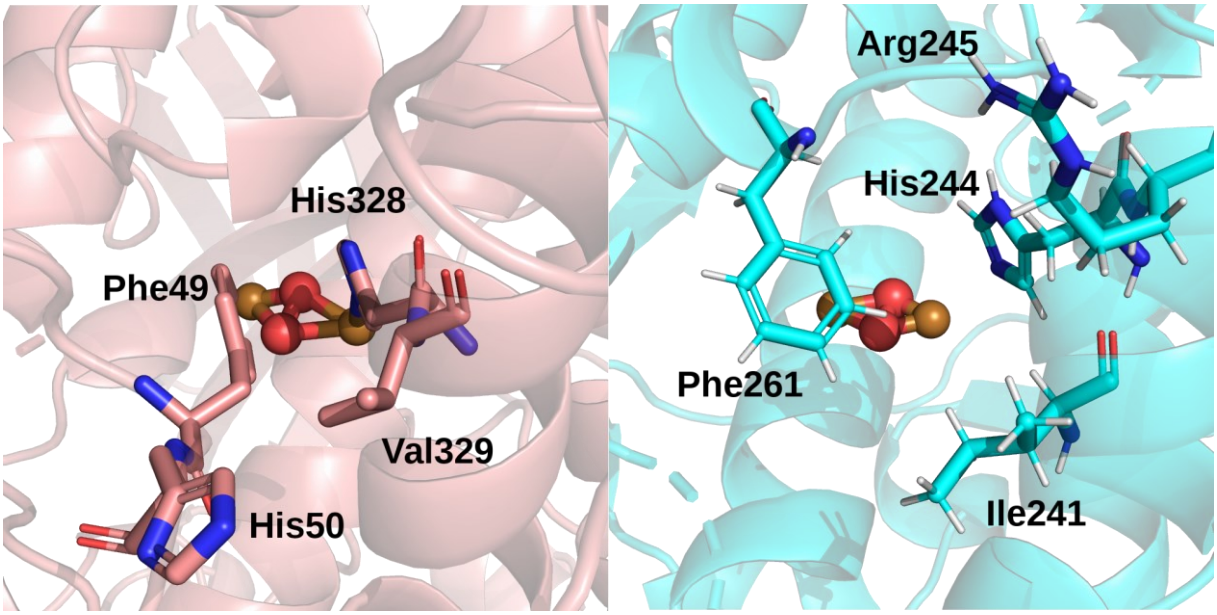

Figure 25 – Difference in the entrance of hemocyanin and IBCO structures. The active site is seen from the direction of the tunnel. Residues potentially related to reactivity/no reactivity of the proteins are highlighted as thicker sticks with residue name and number.

As can be seen in Figure 25, there are a few amino acids at the entrance of hemocyanin which sterically block the tunnel. There is a smaller distance between the phenylalanine and histidine residues (Phe49 and His328 for Hc, and Phe261 and Hid244 for IBCO) – in hemocyanin, the gap is ~3.5 Å wide (Cζ of Phe49 to Cε of His328) with possible π-stacking between mentioned residues, which is typical length of π-stacking interactions. In case of IBCO, the gap is around 4.0 Å wide due to T-stacking between the hydrogen on aromatic ring of Phe and carbon of His. Also, IBCO does not have the aforementioned sterically blocking residues at the entrance. Interestingly, the same T-stacking geometry of Phe and His (in this case His177) appears in this structure. This indicates that the substrate moves from the side of CuB and then binds to the CuA, which may be the focus of the future work. Another difference is the distance between the copper atoms in the peroxo structure. For hemocyanin, the distance is 3.6 Å, which is the

same distance as in tyrosinase. Both CaOx structures (Figure 26) we used have a distance of 2.9 Å between coppers.

In case of PPO6,we wanted to know, if its AlphaFold 2 prediction can be considered a realistic representation of CaOx. In Figure 26, the overlap of both proteins is depicted. With the RMSD (Root-mean-square deviation) value of 1.86, these structures are almost identical, in terms of all atoms.

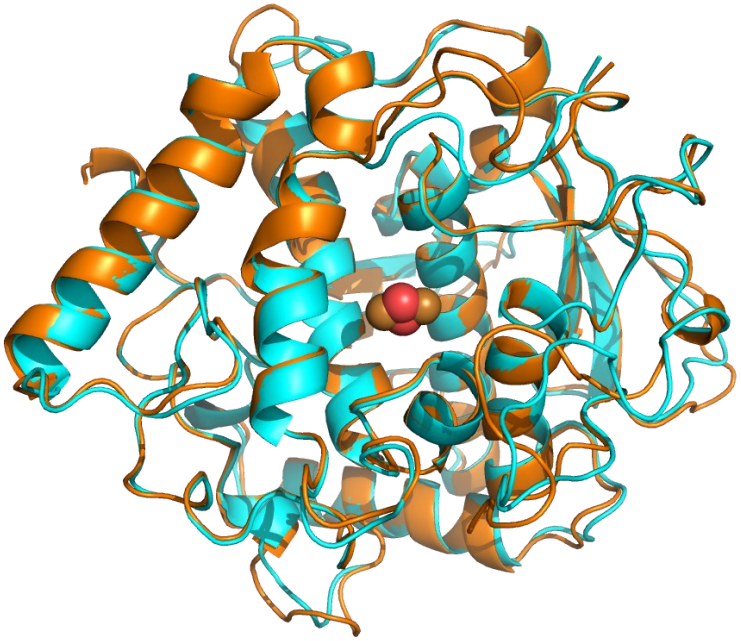

Figure 26 – Overlap of IBCO (turquoise) and PPO6 (orange) structures.

The active sites of both enzymes have the same residues in the same conformation with two exceptions (Figure 27). First, the cysteine (Cys95) bond to histidine (His112) is not present in the AlphaFold 2 prediction of PPO6 as it is in IBCO, where Cys 92 is bound by Sγ to Cε of His109 (Cys-His link). This only applies to the initial structure – during the optimization, the bond has formed. Second, the amino acid next in sequence after histidine (His249 for PPO6, and His240 for IBCO) differs. There is polar threonine at the position 250 in PPO6 while in IBCO has hydrophobic isoleucine is present at position 241. As these are subtle differences, we expect the results for the reaction pathway of IBCO and PPO6 to be essentially the same.

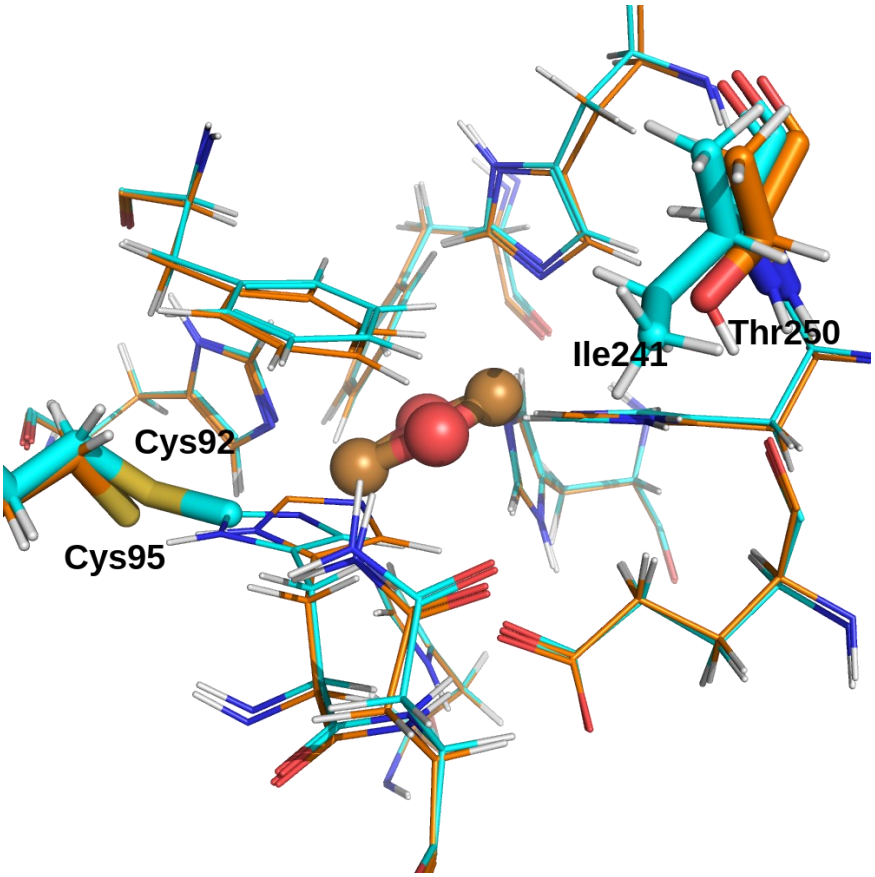

Figure 27 – Overlap of active sites of IBCO (turquoise) and PPO6 (orange). Discussed residues are highlighted by thick stick representation.

6. Conclusion and Outlook

In this study, the reaction pathway described for tyrosinase was applied for additional CBC enzymes, NspF and PPO6 respectively, using quantum mechanical methods. We compared energies between the initial **1P** structures and individual steps during the mechanism. Then we tried to find structural differences that can influence the reactivity of mentioned enzymes.

By comparison between the energy values of the NspF and PPO6 enzymes, NspF has higher values than PPO6 during the first steps of the reaction (~4,5 kcal/mol in **TS1**), but the transition from **2O** to **3O** shows much higher barrier for PPO6's **TS2** (10 kcal/mol), which can be the reason why PPO6 is unable to catalyze the hydroxylation of monophenols. However, by comparison of energies with Ty, the CaOx enzymes have similar values to Ty.

We showed how the difference of spin and the methods used can influence the results. The values of $<S^2>$ are constantly changing during the reaction pathway, but even then, we continued to use open-shell singlet state to be consistent with compared data. Some values of $<S^2>$ during some steps exceeded the value of 1, getting closer to the triplet state. Comparison with open-shell triplet state along the reaction pathway could be done to check if our suggestions were correct.

Geometries of enzymes with similar tertiary structure were compared to find any structural differences causing the ability to catalyze different reactions. Aromatic compounds near the entrance to the active site possibly interact with the substrate by π-stacking interactions and stabilizing the bound substrate. IBCO and PPO6 active sites are similar, except for one amino acid (Thr250 in PPO6 and Ile241 in IBCO). In our opinion, this difference should not influence the reactivity, and energies of either enzymes should be approximately the same.

Next steps will be to investigate reaction coordinate of NspF as a representative of *ortho*-aminophenol oxidases. Initial structures for these calculations will contain *ortho*-aminophenol as a substrate. First step of converged initial structure **1P** (Figure 28) was already prepared. Whole research will be done with the collaboration with the group of Edward I. Solomon from Standford University, which will provide experimental data.

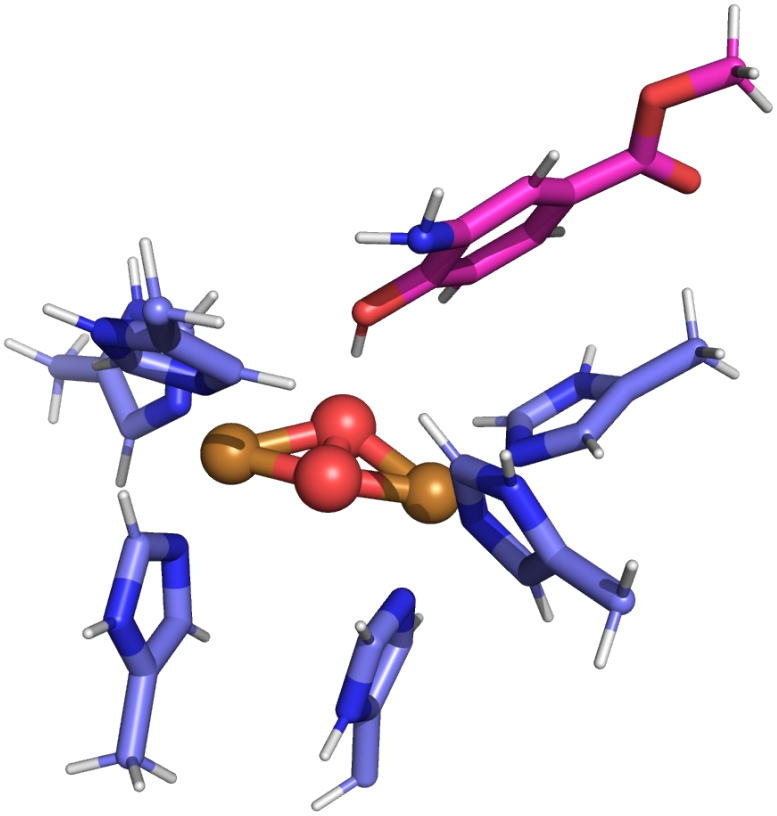

Figure 28 – NspF's 1P initial structure with ortho-aminophenol as a substrate.

References

(1) Voet, D.; Voet, J. G. *Biochemistry*; John Wiley & Sons, 2010.

(2) Zimmerman, J. Physical Chemistry for the Biosciences: Chang, Raymond. *Biochem. Mol. Biol. Educ.* 2005, *33*, 382–382. https://doi.org/10.1002/bmb.2005.49403305383.

(3) Maret, W.; Li, Y. Coordination Dynamics of Zinc in Proteins. *Chem. Rev.* 2009, *109* (10), 4682–4707. https://doi.org/10.1021/cr800556u.

(4) Zheng, H.; Cooper, D. R.; Porebski, P. J.; Shabalin, I. G.; Handing, K. B.; Minor, W. CheckMyMetal: A Macromolecular Metal-Binding Validation Tool. *Acta Crystallogr. Sect. Struct. Biol.* 2017, *73* (3), 223–233. https://doi.org/10.1107/S2059798317001061.

(5) Sánchez-Aparicio, J.-E.; Tiessler-Sala, L.; Velasco-Carneros, L.; Roldán-Martín, L.; Sciortino, G.; Maréchal, J.-D. BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization. *J. Chem. Inf. Model.* 2021, *61* (1), 311–323. https://doi.org/10.1021/acs.jcim.0c00827.

(6) *MetalPDB in 2018: a database of metal sites in biological macromolecular structures | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/46/D1/D459/4564802 (accessed 2024-05-12).

(7) Glusker, J. P. Structural Aspects of Metal Liganding to Functional Groups in Proteins. In *Advances in Protein Chemistry*; Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eisenberg, D. S., Eds.; Metalloproteins: Structural Aspects; Academic Press, 1991; Vol. 42, pp 1–76. https://doi.org/10.1016/S0065-3233(08)60534-3.

(8) Huang, J.; Yu, Z.; Chistoserdova, L. Lanthanide-Dependent Methanol Dehydrogenases of XoxF4 and XoxF5 Clades Are Differentially Distributed Among Methylotrophic Bacteria and They Reveal Different Biochemical Properties. *Front. Microbiol.* 2018, *9*. https://doi.org/10.3389/fmicb.2018.01366.

(9) Kauffman, G. B. Ivano Bertini, Harry B. Gray, Edward I. Stiefel, and Joan Selverstone Valentine (Eds.): Biological Inorganic Chemistry: Structure & Reactivity, 3rd Edition. *Struct. Chem.* 2007, *18* (3), 421–422. https://doi.org/10.1007/s11224-007-9156-z.

(10) Gray, H. B.; Malmström, B. G.; Williams, R. J. P. Copper Coordination in Blue Proteins. *JBIC J. Biol. Inorg. Chem.* 2000, *5* (5), 551–559. https://doi.org/10.1007/s007750000146.

(11) Whittaker, J. W. The Radical Chemistry of Galactose Oxidase. *Arch. Biochem. Biophys.* 2005, *433* (1), 227–239. https://doi.org/10.1016/j.abb.2004.08.034.

(12) Bertini, I. *Biological Inorganic Chemistry: Structure and Reactivity*; University Science Books, 2007.

(13) Solomon, E. I.; Heppner, D. E.; Johnston, E. M.; Ginsbach, J. W.; Cirera, J.; Qayyum, M.; Kieber-Emmons, M. T.; Kjaergaard, C. H.; Hadt, R. G.; Tian, L. Copper Active Sites in Biology. *Chem. Rev.* 2014, *114* (7), 3659–3853. https://doi.org/10.1021/cr400327t.

(14) Cramer, C. J.; Włoch, M.; Piecuch, P.; Puzzarini, C.; Gagliardi, L. Theoretical Models on the Cu2O2 Torture Track: Mechanistic Implications for Oxytyrosinase and Small-Molecule Analogues. *J. Phys. Chem. A* 2006, *110* (5), 1991–2004. https://doi.org/10.1021/jp056791e.

(15) Stańczak, A.; Chalupský, J.; Rulíšek, L.; Straka, M. Comprehensive Theoretical View of the [Cu2O2] Side-on-Peroxo-/Bis-μ-Oxo Equilibria. *ChemPhysChem* 2022, *23* (14), e202200076. https://doi.org/10.1002/cphc.202200076.

(16) Kipouros, I.; Stańczak, A.; Culka, M.; Andris, E.; Machonkin, T. R.; Rulíšek, L.; Solomon, E. I. Evidence for H-Bonding Interactions to the μ-H2:H2-Peroxide of Oxy-Tyrosinase That Activate Its Coupled Binuclear Copper Site. *Chem. Commun.* 2022, *58* (24), 3913–3916. https://doi.org/10.1039/D2CC00750A.

(17) Walker, J. R. L.; Ferrar, P. H. Diphenol Oxidases, Enzyme-Catalysed Browning and Plant Disease Resistance. *Biotechnol. Genet. Eng. Rev.* 1998, *15* (1), 457–498. https://doi.org/10.1080/02648725.1998.10647966.

(18) Kipouros, I.; Stańczak, A.; Ginsbach, J. W.; Andrikopoulos, P. C.; Rulíšek, L.; Solomon, E. I. Elucidation of the Tyrosinase/O2/Monophenol Ternary Intermediate That Dictates the Monooxygenation Mechanism in Melanin Biosynthesis. *Proc. Natl. Acad. Sci.* 2022, *119* (33), e2205619119. https://doi.org/10.1073/pnas.2205619119.

(19) Kanteev, M.; Goldfeder, M.; Fishman, A. Structure–Function Correlations in Tyrosinases. *Protein Sci.* 2015, *24* (9), 1360–1369. https://doi.org/10.1002/pro.2734.

(20) Siegbahn, P. E. M. The Catalytic Cycle of Tyrosinase: Peroxide Attack on the Phenolate Ring Followed by O-O Bond Cleavage. *JBIC J. Biol. Inorg. Chem.* 2003, *8* (5), 567–576. https://doi.org/10.1007/s00775-003-0449-4.

(21) Per E. M. Siegbahn*, † and; Wirstam‡, M. *Is the Bis-μ-Oxo Cu2(III,III) State an Intermediate in Tyrosinase?*. ACS Publications. https://doi.org/10.1021/ja010829t.

(22) Siegbahn, P. E. M.; Borowski, T. Comparison of QM-Only and QM/MM Models for the Mechanism of Tyrosinase. *Faraday Discuss.* 2010, *148* (0), 109–117. https://doi.org/10.1039/C004378H.

(23) Inoue, T.; Shiota, Y.; Yoshizawa, K. Quantum Chemical Approach to the Mechanism for the Biological Conversion of Tyrosine to Dopaquinone. *J. Am. Chem. Soc.* 2008, *130* (50), 16890–16897. https://doi.org/10.1021/ja802618s.

(24) *Zur Quantentheorie der Molekeln - Born - 1927 - Annalen der Physik - Wiley Online Library*. https://onlinelibrary.wiley.com/doi/10.1002/andp.19273892002 (accessed 2024-04-17).

(25) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2. ed.; Wiley: Chichester, 2014.

(26) Wacławek, S. Do We Still Need a Laboratory to Study Advanced Oxidation Processes? A Review of the Modelling of Radical Reactions Used for Water Treatment. *Ecol. Chem. Eng. S* 2021, *28* (1), 11–28. https://doi.org/10.2478/eces-2021-0002.

(27) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 1st ed.; Wiley, 2001. https://doi.org/10.1002/3527600043.

(28) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* 1964, *136* (3B), B864–B871. https://doi.org/10.1103/PhysRev.136.B864.

(29) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* 1965, *140* (4A), A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133.

(30) Perdew, J. P.; Burke, K.; Wang, Y. Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System. *Phys. Rev. B* 1996, *54* (23), 16533–16539. https://doi.org/10.1103/PhysRevB.54.16533.

(31) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta--Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* 2003, *91* (14), 146401. https://doi.org/10.1103/PhysRevLett.91.146401.

(32) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* 1993, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(33) Bour, P. Základy kvantových výpočtů vlastností molekul.

(34) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* 1976, *103* (2), 227–249. https://doi.org/10.1016/0022-2836(76)90311-9.

(35) *ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition | The Journal of Physical Chemistry*. https://pubs.acs.org/doi/abs/10.1021/jp962071j (accessed 2024-04-17).

(36) Groenhof, G. Introduction to QM/MM Simulations. In *Biomolecular Simulations: Methods and Protocols*; Monticelli, L., Salonen, E., Eds.; Humana Press: Totowa, NJ, 2013; pp 43–66. https://doi.org/10.1007/978-1-62703-017-5_3.

(37) Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*; Reiher, M., Ed.; Springer: Berlin, Heidelberg, 2007; pp 173–290. https://doi.org/10.1007/128_2006_084.

(38) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* 1981, *55* (1), 117–129. https://doi.org/10.1016/0301-0104(81)85090-2.

(39) Garstka, M.; Cannon, M.; Goulart, P. COSMO: A Conic Operator Splitting Method for Convex Conic Problems. *J. Optim. Theory Appl.* 2021, *190* (3), 779–810. https://doi.org/10.1007/s10957-021-01896-x.

(40) Kutzelnigg, W.; Liu, W. Quasirelativistic Theory Equivalent to Fully Relativistic Theory. *J. Chem. Phys.* 2005, *123* (24), 241102. https://doi.org/10.1063/1.2137315.

(41) Peng, D.; Reiher, M. Exact Decoupling of the Relativistic Fock Operator. *Theor. Chem. Acc.* 2012, *131* (1), 1081. https://doi.org/10.1007/s00214-011-1081-y.

(42) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 2021, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(43) Klabunde, T.; Eicken, C.; Sacchettini, J. C.; Krebs, B. Crystal Structure of a Plant Catechol Oxidase Containing a Dicopper Center. *Nat. Struct. Biol.* 1998, *5* (12), 1084–1090. https://doi.org/10.1038/4193.

(44) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* 1989, *162* (3), 165–169. https://doi.org/10.1016/0009-2614(89)85118-8.

(45) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* 2005, *7* (18), 3297–3305. https://doi.org/10.1039/B508541A.

(46) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* 2011, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.

(47) Case, D.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Berryman, J.; Brozell, S.; Cerutti, D.; Cheatham, T.; Cisneros, G. A.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; Gilson, M.; Gohlke, H.; Götz, A.; Harris, R.; Izadi, S.; Izmailov, S.; Kollman, P. *Amber 2022*; 2022. https://doi.org/10.13140/RG.2.2.31337.77924.

(48) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 2015, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(49) Pollak, P.; Weigend, F. Segmented Contracted Error-Consistent Basis Sets of Double- and Triple-ζ Valence Quality for One- and Two-Component Relativistic All-Electron Calculations. *J. Chem. Theory Comput.* 2017, *13* (8), 3696–3705. https://doi.org/10.1021/acs.jctc.7b00593.

(50) Rokob, T. A.; Rulíšek, L. Curvature Correction for Microiterative Optimizations with QM/MM Electronic Embedding. *J. Comput. Chem.* 2012, *33* (12), 1197–1206. https://doi.org/10.1002/jcc.22951.

(51) Chalupský, J.; Srnec, M. Beyond the Classical Contributions to Exchange Coupling in Binuclear Transition Metal Complexes. *J. Phys. Chem. A* 2021, *125* (11), 2276–2283. https://doi.org/10.1021/acs.jpca.0c11237.

(52) L. Holland, P.; B. Tolman, W. Dioxygen Activation by Copper Sites: Relative Stability and Reactivity of (μ-H2:H2-Peroxo)- and Bis(μ-Oxo)Dicopper Cores. *Coord. Chem. Rev.* 1999, *190–192*, 855–869. https://doi.org/10.1016/S0010-8545(99)00125-3.

(53) Solem, E.; Tuczek, F.; Decker, H. Tyrosinase versus Catechol Oxidase: One Asparagine Makes the Difference. *Angew. Chem. Int. Ed.* 2016, *55* (8), 2884–2888. https://doi.org/10.1002/anie.201508534.

(54) Prexler, S. M.; Frassek, M.; Moerschbacher, B. M.; Dirks-Hofmeister, M. E. Catechol Oxidase versus Tyrosinase Classification Revisited by Site-Directed Mutagenesis Studies. *Angew. Chem. Int. Ed.* 2019, *58* (26), 8757–8761. https://doi.org/10.1002/anie.201902846.

(55) Ginsbach, J. W.; Kieber-Emmons, M. T.; Nomoto, R.; Noguchi, A.; Ohnishi, Y.; Solomon, E. I. Structure/Function Correlations among Coupled Binuclear Copper Proteins through Spectroscopic and Reactivity Studies of NspF. *Proc. Natl. Acad. Sci.* 2012, *109* (27), 10793–10797. https://doi.org/10.1073/pnas.1208718109.

(56) Magnus, K. A.; Hazes, B.; Ton-That, H.; Bonaventura, C.; Bonaventura, J.; Hol, W. G. J. Crystallographic analysis of oxygenated and deoxygenated states of arthropod hemocyanin shows unusual differences. *Proteins Struct. Funct. Bioinforma.* 1994, *19* (4), 302–309. https://doi.org/10.1002/prot.340190405.