Norwegian University
of Life Sciences

Master's Thesis 2021   30 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Computational and structural characterization of engineered Type-3-copper variants.

Linn Kristine Larsson
Chemistry and biotechnology

Acknowledgements

Abstract

This master thesis aimed to use experiments at the laboratory and computational chemistry to begin the process of characterizing suitable enzyme candidates for conversion of CO_2 to oxalate by type-3 copper proteins. The WT of Tyrosinase (TyrBm) from *Bacillus megaterium* and mutants of this protein F227Y and F65Y were characterized through expression, purification, activity assays, and X-ray crystallography. The diffraction data from the X-ray experiments were processed, and the resulting structure was further analyzed by computational chemistry using the software xTB -GFN -FF. The analysis was performed to see if xTB -GFN -FF was able to reproduce and predict variants of the enzyme structure.

The experiments showed that both TyrBm WT and the mutants were successfully expressed and purified for further activity assays and crystallization. Activity assays were performed using caffeic acid as the substrate and showed that copper ions had to be added to the enzymes prior to the assay for the mutants to show activity. The loss of activity indicated that copper ions had been lost in the active site during the purification process. Crystals of TyrBm WT and the mutants F65Y and F227Y were obtained using the hanging drop method, and the crystal structure of TyrBm F227Y crystallized in the presence of zinc ions were solved using X-ray crystallography. The enzyme crystallized as a dimer in the asymmetric unit, with six histidine residues coordinated around two metal ions. As the enzyme was crystallized in the presence of zinc ions, there was uncertainty if copper or zinc was present in the active site. The enzyme was therefore geometry optimized with xTB -GFN -FF with Cu(I), Cu(II), and Zn(II), respectively, in the active site with H_2O, then OH^- and finally O^{2-} bridging the metal ions. The result showed that xTB -GFN -FF could reproduce and model the different enzyme variants at a reasonable accuracy. The model with Zn(II) and H_2O seemed to be most similar to the structure from the X-ray crystallization and therefore substantiate the hypothesis that zinc and not copper were in the model´s active site. The work done in this thesis will provide some of the groundwork to enable further COOFIX research for "Evolving novel carbon dioxide reducing enzymes".

Abstrakt

Målet med denne masteroppgaven var å bruke eksperimentelle metoder på laboratoriet og beregningsorientert kjemi for å begynne karakteriseringen av mulige enzym kandidater for konvertering av CO_2 til oxalate ved hjel av kobber -3 proteiner. Villtype av Tyrosinase fra *Bacillus Megaterium* (TyrBm WT) og mutanter av dette proteinet F227Y og F65Y ble karakterisert gjennom uttrykking, rensing, aktivitet assay og røntgenkrystallografi. Diffraksjonsdata fra røntgenkrystallografieksperimentene ble deretter prosessert, og den endelige strukturen ble videre analysert ved å bruke beregningsorientert kjemi med programvaren, xTB -GFN -FF. Analysen ble gjort for å finne ut om xTB -GFN -FF er kapabel til å reprodusere og predikere varianter av enzym strukturen.

Resultatet av eksperimentene viste at både TyrBm WT og mutantene ble vellykket uttrykket og renset for videre utføring av aktivitets assays og krystallisering. Aktivitets assayene ble utført med caffeic acid som substrat og viste at kobber ioner måtte bli tilsatt til enzym løsningene i forkant av assayet for at mutantene skulle vise aktivitet. Tap av aktivitet hos mutantene kan indikere at kobber ionene i det aktive setet hadde blitt mistet under rensingsprosessen. Krystaller av TyrBm WT og mutantene F227Y og F65Y ble oppnådd ved å bruke «hanging drop method». Krystallstrukturen av F227Y, krystallisert i nærvær av sink ioner, ble løst ved å bruke røntgenkrystallografi. Enzymet krystalliserte seg som en dimer i den asymmetriske enheten med seks histidin residuer koordinert rundt to metall ioner. Ettersom enzymet ble krystallisert i nærvær av sink var det usikkert om kobber eller sink var til stede i det aktive setet. Enzymet ble dermed geometrioptimalisert med xTB -GFN -FF med henholdsvis Cu(I), Cu(II) og Zn(II) i det aktive setet med først H_2O, deretter OH^- og til slutt O^{2-} som bro mellom metallionene. Resultatet viste at xTB -GFN -FF kan reprodusere og predikere forskjellige enzym varianter med aksepterbar nøyaktighet. Fra beregningene viste modellen med Zn(II) og H_2O mest likhet med strukturen fra røntgenkrystaliseringen og dermed underbygger hypotesen om at sink og ikke kobber var i det aktive setet til modellen. Arbeidet gjort i denne masteroppgaven vil legge noe av grunnarbeidet for videre forskning i COOFIX «"Evolving novel carbon dioxide reducing enzymes".

Table of content

Abbreviations

Å	Ångstøm
B. megaterium	Bacillus Megaterium
E. coli	Escherichia coli
Coot	Crystallographic object-oriented toolkit
Da	Dalton
His	Histidine
Kb	Kilo base
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
MOPS	3-(N-morpholino)propanesulfonic acid
MQ	MilliQ-filtered water
MR	Molecular replacement
MW	Molecular weight
NHE	Normal hydrogen electrode
Ni-NTA	Nickel-nitrilotriacetic acid
OD	Optical density
PDB	Protein Data Bank
PEG	Polyethylene glycol
Rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
UV	Ultraviolet
(w/w)	(weight/weight)
xTB	Extended tight binding

1. Introduction

This master thesis is a part of the "Evolving novel carbon dioxide reducing enzymes (COOFIX)" project led by Åsmund Røhr Kjendseth.

1.1 Background for the project

Carbon dioxide is one of the most prominent greenhouse gases linked to global warming [1]. Greenhouse gases are primarily emitted through anthropogenic activities such as burning fossil fuel, deforestation, and cement production, but also through natural sources such as decomposition, respiration, and ocean release [2, 3]. Plants and other autotrophs organisms mainly use the enzyme Ribulose 1,5-bisphosphate carboxylase-oxygenase (RubisCO) to fix carbon dioxide in the Calvin-Benson cycle. This process leads to the production of glucose which is an essential energy source for the plant. However, due to the low catalytic rate of RuBisCO, fixing only about three CO_2 molecules per second, this process is very inefficient and energy-consuming [4]. RubisCO is also compromised by a competing process called photorespiration. During photorespiration, atmospheric O_2 competes with CO_2 for the active site, costing the organism even more energy [5]. Improving this reaction and establishing alternative pathways for CO_2 fixation could help mitigate the greenhouse gas emissions leading to climate change.

Much research has already been done on improving the catalytic efficiency of RubisCO, but with limited success [6]. Therefore it has been suggested that RubisCO might be superseded by other enzymes, fixing CO_2 to alternative products of value [7]. One idea is the reduction of CO_2 to oxalate by a copper complex [8]. Oxalate is the simplest C-C coupled product that can be formed by the reduction of CO_2 molecules [9]. Due to its thermodynamic stability and kinetic inertness, CO_2 is difficult to reduce and a natural conversion of CO2 to oxalate has therefore not yet been discovered. For this reaction to occur, two CO_2 molecules must be reduced to carbon dioxide radicals (CO_2^-) and undergo a self-coupling reaction [10]. The reaction to form carbon dioxide radicals is a one-electron reduction limited by a high energy barrier (E_0= -1,97 V, $CO_2 + e- \rightarrow CO_2^{\cdot-}$). This high energy barrier is caused by the reorganizational energy required for bending the linear CO_2 molecule to the preferred shape of the carbon dioxide radical [11, 12]. Although conversion of CO_2 to oxalate has not been observed in nature, this reaction has been successful in the laboratory, being catalysed by a copper complex [8].

It was suggested by Åsmund Røhr Kjendseth that the structure of the organometallic complexes highly resembles the structure of type-3 copper proteins. These proteins could act as promising candidates for the reduction of CO_2 to oxalate, which is where the project of "Evolving novel carbon dioxide reducing enzymes" (COOFIX) started. COOFIX aims to design enzymes that can convert CO_2 directly to a useful metabolic intermediate (oxalate) by identifying optimal binuclear copper site geometries for CO_2 reduction. Oxalate can then be converted to biomass or precipitate to a solid compound for storage [13].

1.2 Type- 3 copper proteins

Type- 3 copper proteins, also called coupled binuclear copper proteins (CBC), are a diverse family of proteins involved in various biological processes such as transportation of oxygen, immunity, and pigment formation. The family consists mainly of tyrosinases, catechol oxidases, and hemocyanins and are characterized by a magnetically coupled binuclear copper active site [14, 15]. The copper ions being "coupled" refers to the magnetic interactions between the copper centres in the active site. The distance between the two coppers depends on the oxidation state and protein type but usually range between 3.2 and 4.6 Å [16]. This binuclear copper active site, except for in hemocyanins, plays a crucial role in binding, activating, and reducing O_2 to H_2O while hydroxylating or oxidating a substrate. In this master thesis, the focus will be on tyrosinase from *Bacillus megaterium* (TyrBm).

1.2.1 Tyrosinase from *B. Megaterium*

Tyrosinases are involved in the production of melanin in many different organisms. Melanin is responsible for the browning of food, such as fruits and vegetables, and pigment formation in the skin. The production of melanin happens through oxidation of o-diphenols to o-quinoes (diphenolase activity). The quinones are reactive and can spontaneously polymerize to make melanin. The enzyme also catalyses the reaction where hydroxylation of phenols leads to ortho-diphenols (monophenolase activity). Both monophenolase activity and diphenolase activity reactions require molecular oxygen [17].

Tyrosinase from *B. Megaterium* crystallizes as a homodimer in the asymmetric unit. The 3D representation of the crystal structure is presented in Figure 1.1. The dimer is composed of two ellipsoid shaped subunits, polypeptide chains A and B. Each of the polypeptide chains is made up of 303 amino acid residues. The molecular weight of the whole structure is ~ 70 kDA (monomeric

chain ~35 kDa). The secondary structure mainly consists of α- helices where the dinuclear copper centre can be found at the core of a four helical bundle in each subunit [17, 18].
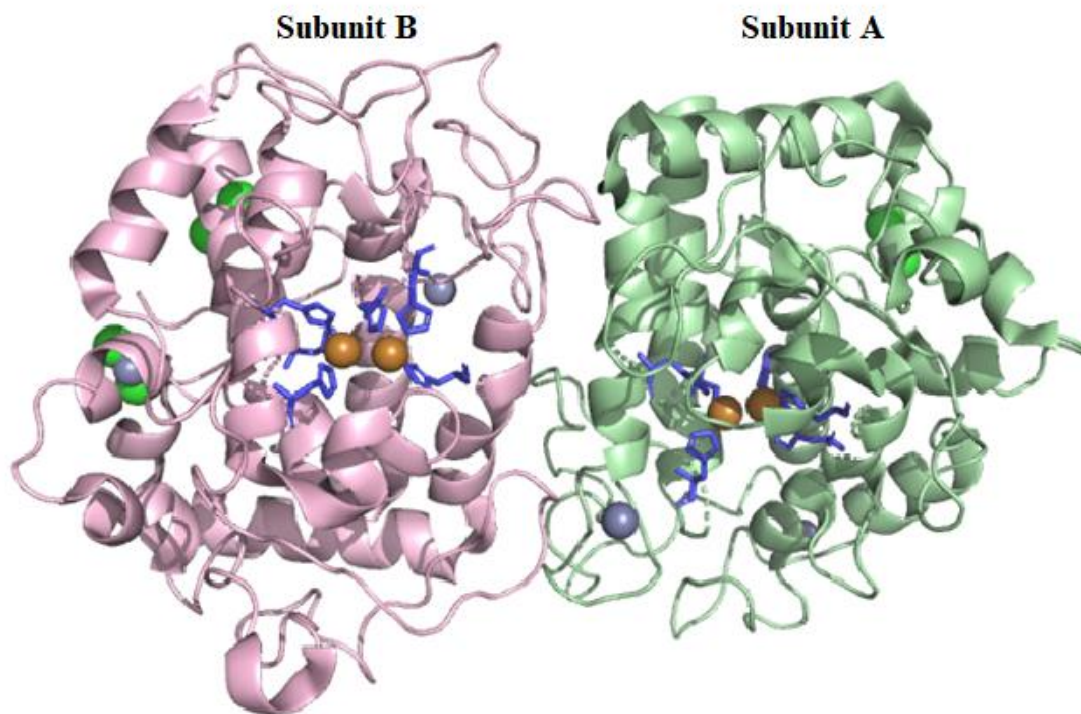
Figure 1.1. 3D representation of the crystal structure of Tyrosinase from *B. megaterium* (TyrBm). The molecule is a homodimer composed of subunit A (pink) and subunit B (green). The coordination of the copper centre is presented in purple with six histidine residues in each of the subunits (purple). Copper atoms are presented in orange, zinc atoms in grey, and chloride atoms in green. All molecules presented in this master thesis were generated using PyMOL (*http://www.pymol.org/*) and are fetched from Protein Data Base (PDB) database.

The dinuclear copper centre, consisting of CuA and CuB, serves as a crucial cofactor within the active site of subunit A and B. The geometric and electronic structures of this cofactor control the enzyme activity to a great extent and the lack of copper has shown to have a substantial negative impact on activity [19]. The copper ions are coordinated by three histidine residues bound to each copper atom consisting of two strong equatorial and one weak axial bond [20]. CuA binds to the nitrogen atoms of His42, His69, and His60, while the second copper ion, CuB, binds to His204, His208, and His231[17]. The coordination of the active site is presented in Figure 1.2.

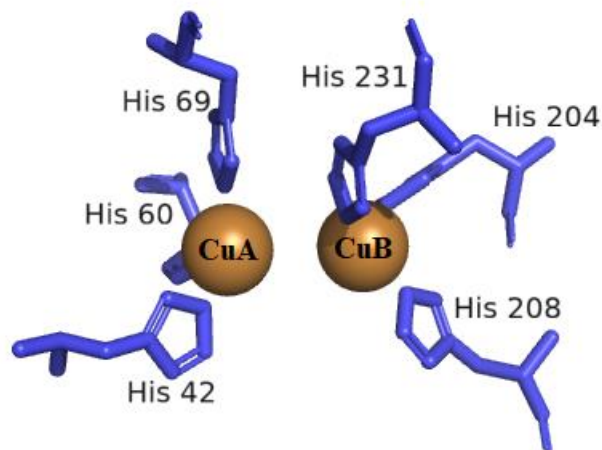**Figure 1.2. TyrBm WT (PDB ID: 3NM8) active site crystallized in the presence of Zn**. The active site consists of six conserved histidine residues His 42, His 60, His 69, His 204, His 208, and His 231 coordinated around two copper ions. The distance between the copper ions is 3.58 -Å.

The arrangement of the binuclear copper site exists in three isoforms Oxy-tyrosinase, Met-tyrosinase and Deoxy-tyrosinase, as presented in Figure 1.3 (Likhitwitayawuid [20]. Deoxy-tyrosinase has a Cu(I)-Cu(I) structure. The binding of O_2 in Deoxy-tyrosinase gives rise to a side-on planar μ-η^2:η^2 peroxo-bridged Cu_2O_2 oxy structure with a short Cu(II)-Cu(II) distance around 3.6 Å. The geometry promotes a significant metal-ligand overlap between the O_2^{2-} π^* and the two Cu-$d_{x^2-y^2}$ orbitals, which gives a direct pathway for the two-electron reduction of O_2 to peroxide to form Oxy-tyrosinase [21, 22]. Oxy-tyrosinase consists of two tetragonal Cu(II) atoms. The O_2 molecule bound as peroxide in this configuration creates a strong bridge between the copper atoms [20]. Oxy-tyrosinase can access and coordinate the substrate directly to the copper active site to perform monophenolase activity or diphenolase activity [23]. Met-tyrosinase is the resting form of tyrosinase. In this form, the Cu (II) atoms are bridged by a hydroxide or water molecule. The met-form can undergo a reaction with o-diphenol resulting in the deoxy-form, Cu(I)-Cu(I) form of tyrosinase. This form will again be converted by molecular oxygen to the oxy-form [20, 24]. Understanding the coordination and the reaction mechanism of the copper centre is essential in understanding how the enzymes work and how one can potentially manipulate that.
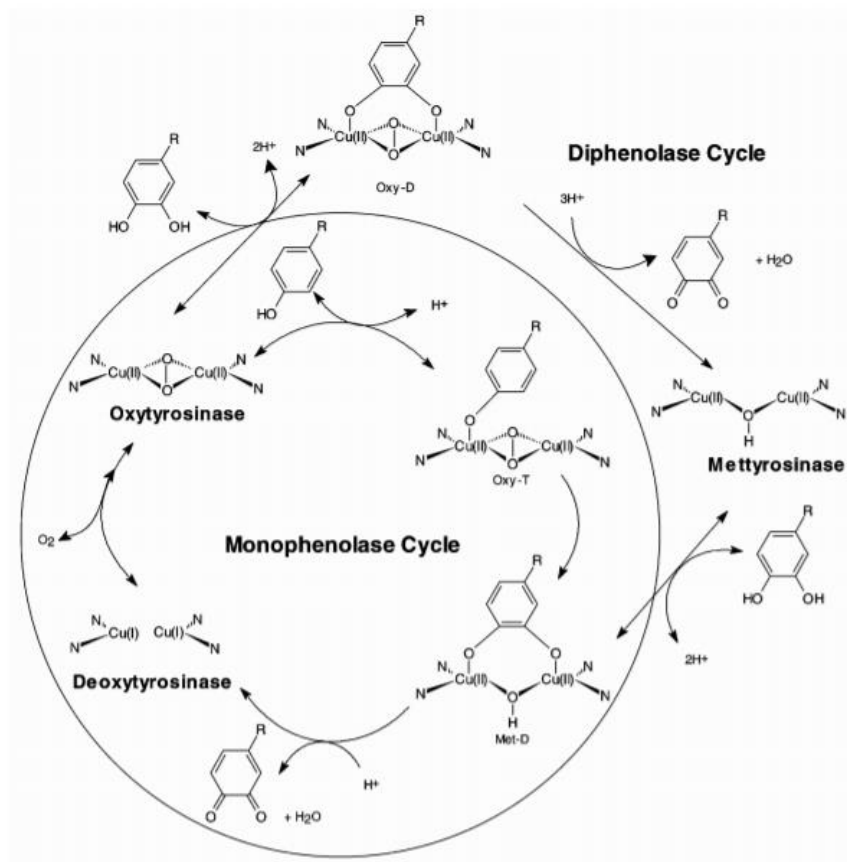
Figure 1.3. Likhitwitayawuid [20]. The mechanism for monophenolase and dipheolase -activity by tyrosinase. The figure displays the three isoforms of tyrosinase; Oxy-tyrosinase, Met-tyrosinase and Deoxy-tyrosinase and how they react to perform monophenolase and diphenolase activity.

There are two reasons to believe that CO_2 can be converted to oxalate by copper 3-proteins. Angamuthu, Byers [8] found that when CO_2 was converted to oxalate, the Cu(I) dimer in the copper complex is selectively oxidized by CO_2 rather than O_2. The reaction with CO_2 resulted in a tetranuclear oxalate-bridged complex where the two Cu(II) ions within the asymmetric unit had a copper distance at 5.3205 ± 0.0006 Å. [8]. The copper 3-proteins used in this thesis have a similar structure to the copper complex used by Angamuthu, Byers [8], with both having a binuclear copper centre coordinated with three histidine residues. The similarities in the copper structures give cause to think that CO_2 conversion by a copper-3-protein is possible. The selective reaction with CO_2 instead of its normal preference for O_2 is probably because of the long copper distance, which would be energetically unfavourable for a two-electron reduction of O_2 (?) The second reason that CO_2- may be able to bind to copper -3 proteins is that the added NO_2^- in hemocyanins can bind to the mixed-valenced form of the active site of (Cu(I)-Cu(II)) [25]. NO_2- shares some of

the same properties as CO_2 and could therefore indicate that it would be possible to stabilize a one-electron reduction to CO_2- by the initial intermediate [26]. To selectively have copper -3 proteins produce oxalate from CO_2, the copper-copper distance in the enzymes active site needs to be modified.

Designing enzymes can be done through methods such as mutagenesis at the laboratory, but to save time, there is an option to do experiments *in silico* using computational chemistry [27]. Computational chemistry takes advantage of mathematical algorithms, statistics, and large databases to unite theoretical chemistry with observations from the lab. One of the advantages of using computational chemistry is solving problems that require massive amounts of data using supercomputers [28]. Simulations done by supercomputers can help identify the suitable enzyme candidates before laboratory testing, limiting the need for resource-demanding trial and error procedures and creating a cheaper and more sustainable chemical industry [29].

In this thesis, both experiments at the laboratory and computational chemistry were used to begin the process of characterizing suitable enzyme candidates for CO_2 conversion to oxalate by type- 3 copper proteins. In the laboratory, WT of Tyrosinase from *B. megaterium* and mutants of this protein, F227Y and F65Y, were characterized through expression, purification, activity assays and X-ray crystallography. The diffraction data from the crystallization of the F227Y mutant was then processed and further optimized with computational chemistry using the software xTB -GFN to analyse if this program could reproduce and model variants of the enzyme structures. The work done in this thesis will provide the groundwork to enable further COOFIX research for "Evolving novel carbon dioxide reducing enzymes".

1.3 Method-related theory

X-ray crystallography and computational chemistry were the two methods applied for analysing the enzymes in this master project. Some theory for each of the methods is briefly presented in this section.

1.3.1 Computational chemistry

For executing experiments *in silico,* the quantum chemical package extended tight binding (xTB) [30] programs were used. Solving quantum chemical problems was for a long time only performed by specialists with access to complicated software and supercomputers. Solving quantum chemical problems traditionally requires a massive amount of computer power and made it hard for the "regular" chemist to participate in the development of the quantum chemical field. Today, quantum chemistry is much more accessible to all due to the development of better technology and more user-friendly software [30]. For small molecular systems up to a few hundred atoms, quantum chemistry is currently dominated by Kohn-Sham density functional theory (DTF). DTF is a versatile method that provides information about the energy of the system based on the electron density. The problem with this method is that it is only functional for small to medium-sized molecules [31].

Describing large molecular systems with high accuracy is still one of the main challenges of quantum chemistry [32]. Since biological macromolecules are complex and often involve many different types of atoms, the need for a more robust, fast, and accurate method emerged. xTB - GFN [33] was created as a semiempirical extended tight-binding (xTB) method for calculating geometries, vibrational frequencies, and noncovalent interactions on large molecular systems (>1000 atoms); hence the acronym GFN [31]. The tight-binding method (TB) use the atomic orbitals as a basis for solving the Schrodinger equation. The TB method neglects the electron-electron interactions by assuming that the electrons are tightly bound to their respective atoms. This leads to fewer interactions with states and potentials on surrounding atoms of the solid. A consequence of this is that the wave function of the electron will be nearly equivalent to the atomic orbital it belongs to [34].

xTB -GFN is, as mentioned, a semiempirical method. Semiempirical methods are "simplified" quantum mechanical calculations derived from the Hartree-Fock method or DTF [35]. The xTB -GFN method is quick and cost-effective by making approximations to the

Hamiltonian equation. The approximations are efficiently calculated by neglecting integrals in this equation. The remaining integrals are modelled using calibrated parameters from empirical functions against a large number of accurate experimental data or with high-quality theoretical reference data [36]. This thesis makes use of xTB -GFN -FF, which is a non-electronic, force-field (FF) version of xTB -GFN. The "x" in xTB represent the extension of the tight-binding method, which include parameterization of the spd block elements ($Z<=86$) and improvements of the fundamental theory [30, 31]. It uses element-specific empirical fitting for the calculations and is simple to use as it only requires start coordinates and elemental composition as input to do calculations and construct all potential energy terms [33]. The accuracy of xTB is comparable to DFT calculations [31, 33]

1.3.2 X-ray crystallography

The origin of X-ray crystallography can be traced back to when Marx von Laue discovered the diffraction of X-rays by crystals in 1912 [37]. Today X-ray crystallography is a widely used method to determine the atomic and molecular structure of proteins and other biological macromolecules [38, 39]. Crystals are defined as a three-dimensional homogenous solid formed by repeating units of atoms or molecules, with a fixed distance between them [40]. Crystals are needed in X-ray crystallography due to the weak scattering of signal from an individual atom/molecule. Since crystals are built up by repeating units of the same molecule, it amplifies the scattering signal creating a unique diffraction pattern for individual proteins. The structural knowledge obtained from this can give an understanding of the function and activity of the proteins [41]. The process starts with obtaining the crystals. This is done through the expression, purification, and crystallization of the proteins.

Obtaining the crystals

The most challenging step in most protein crystallizations is to grow protein crystals of good quality for structure determination. A crucial part of obtaining good crystals is that the protein is highly purified. The protein of interest should be at least 95% of the total protein in the sample [42]. The crystals are grown by taking a solution of highly concentrated protein and introducing conditions for inducing the protein to come out of the solution. Over time protein crystals will emerge and grow. If crystallization happens too fast or the conditions are slightly off; the protein

may precipitate. Initial experiments are often based on a trial error procedure where several different conditions are tested in parallel. This step is very time consuming, and it usually takes a long time to figure out the specific conditions needed for the protein of interest. In many cases, the proteins may even be inadequate to form crystals without any known reason. Factors like pH, protein concentration, temperature, and choice of crystallization technique, additives, buffer and precipitant determine if the crystallization will be successful or not [38]. Conventional crystallization techniques used for these types of experiments are dialysis and hanging- and sitting-drop vapour diffusion [43].

The hanging drop method was used as the crystallization technique in this master project. The hanging drop method is based on a physical equilibration process being created by the difference in concentration between the initial crystallization agents in the reservoir and the crystallization agents in the droplet containing the protein. The setup for the method is presented in figure Figure 1.4, where the droplet containing the protein is suspended from the cover slide over the reservoir solution. The vapour pressure of the water around the droplet will be greater than over the reservoir, creating a concentration gradient that drives water vapour to migrate from the droplet to the reservoir. The net loss of water and crystallization agent from the droplet causes it to shrink in size leading to an increase in protein concentration. Ideally, the protein solution is concentrated to a state of supersaturation, and at the right conditions, crystals are formed to restore the equilibrium [38, 44].
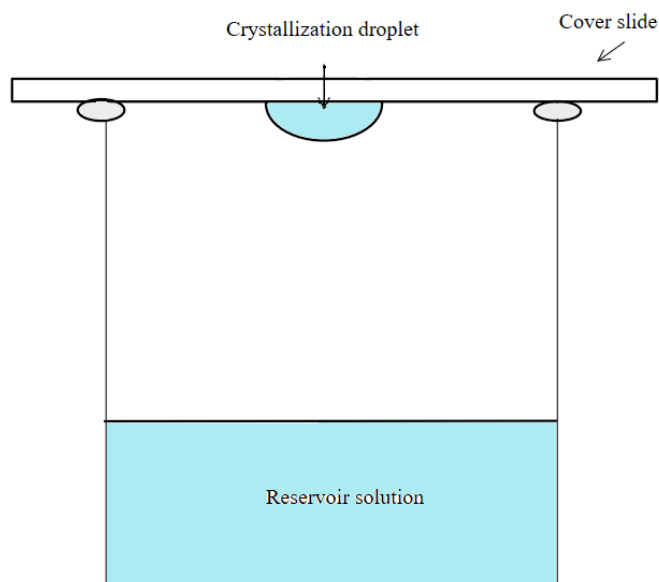
Figure 1.4 Schematic diagram over the hanging drop vapour diffusion method for crystallization of protein [43]. The drop containing the purified protein is suspended from a cover slide over the reservoir solution. A physical equilibration process emerges because of the difference in concentration between the initial crystallization agents in the reservoir and the crystallization agents in the droplet containing the protein.

In general, the growth of a protein crystal can be divided into two different processes, nucleation in the nucleation zone and growth in the metastable zone [43], see figure Figure 1.5. The system starts in the unsaturated zone with all the macromolecules in solution. Crystals in here will neither be able to grow or form. Already formed crystals will dissolve if the system gets pushed to this zone. When the concentration of the macromolecules increases, the system will go into a supersaturated state in the metastable zone [45]. Supersaturation is a nonequilibrium state where the solvent contains more dissolved macromolecules then the solubility limit ordinary accommodates at that temperature [43]. Growth of already formed crystals happens in the metastable zone, but new nuclei will not be able to form spontaneously [45]. During nucleation, the protein must cross an energy barrier to go through a phase transition. In this phase transition, macromolecules come together as aggregates to pass from a disordered state to an ordered state, and viable crystal nuclei will than form spontaneously [43]. The macromolecules arrange themselves in this perfect ordered crystal lattice and not some other disoriented state such as precipitation because they want to minimize the free energy of the system. By doing this, they achieve the most stable conformation with the greatest number of stable bonds, reducing the

overall enthalpy of the system [46]. Ideally, only a single viable crystal nucleus is formed before the systems determinants are adjusted back to the metastable zone for further growth. [45]. The crystal grows larger by attaching molecules to this stable nucleus [43].
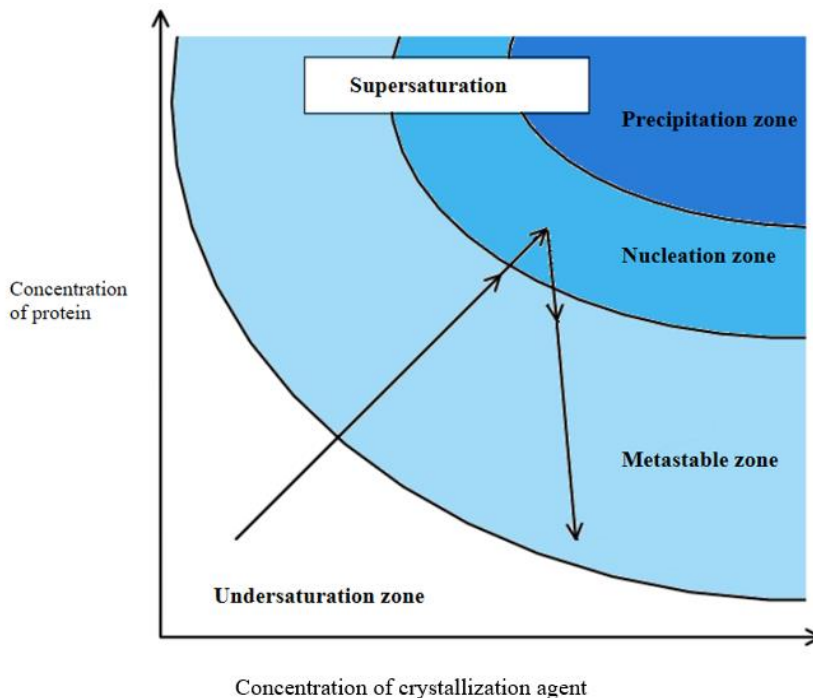
Figure 1.5. Phase diagram for crystallization of macromolecules [45]. The system starts in the unsaturated zone. When the solution reaches a state of supersaturation, it will go into the metastable zone. Here crystals can grow but not form. When the solution enters the nucleation zone, nuclei can spontaneously form. The system will eventually adjust back to the metastable zone, where the crystals can grow big. If the system reaches a very high supersaturation or the process goes too fast, the macromolecules might precipitate [43].

X-ray diffraction and data processing

After obtaining crystals of suitable size, the crystals can be exposed to a high energy X-ray beam. The X-ray beam will create a specific diffraction pattern for the protein, which can be processed to yield information about the three-dimensional structure. A schematic showing the process of crystal determination by X-ray crystallography are presented in Figure 1.6. X-rays can be generated from extremely powerful sources like a synchrotron storage ring or from weaker sources such as electrons striking a copper anode [38]. Using synchrotrons for challenging crystallographic problems are usually preferred as they give a more precise diffraction image and shorter exposure time than "in house" X-ray generators [38]. A moving electron emits energy when it changes

direction. When electrons move fast enough, this energy is emitted in the form of X-rays. In a synchrotron, electrons are accelerated to an enormous speed through a magnetic field. The magnets in the storage ring force the electrons to change directions periodically. This results in the emission of a dozen of thin X-ray beams. The beams get directed toward "beamlines" where the experiments are performed; each "beamline" can be used for a specific type of research [47].
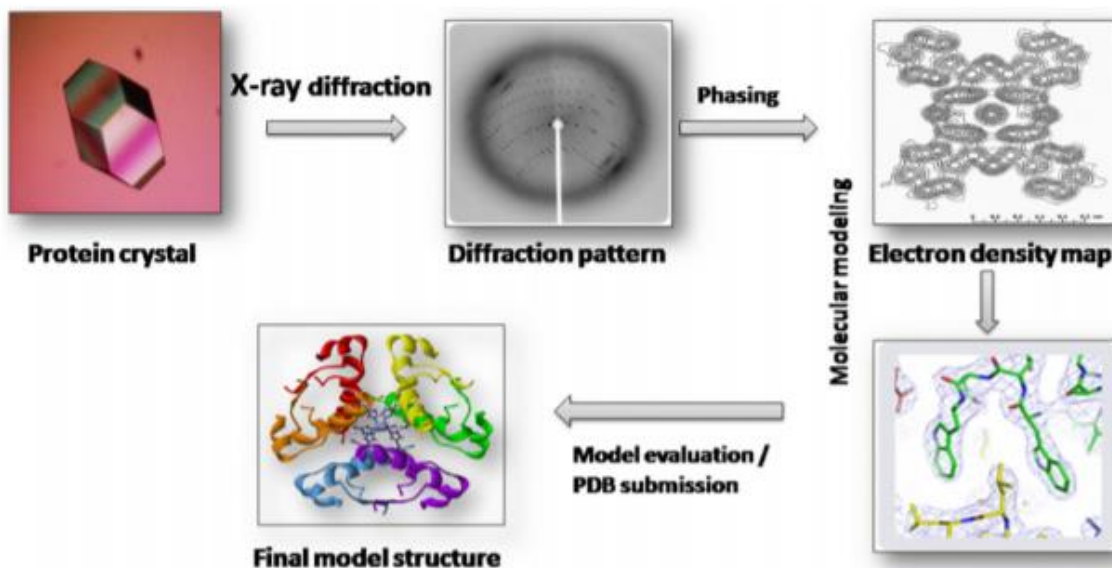
Figure 1.6 Workflow of x- ray crystallography [41].

A monochromatic X-ray beam is used for obtaining a diffraction image of a protein crystal [41]. Beams of X-ray waves hit the rotating crystal (1° at the time) and get scattered by the electrons in the lattice. Within each atom, the scattered waves from all the electrons are added to each other. Some waves interfere constructively and get intensified; these waves are in phase. Other waves interfere destructively and cancel out; these will be out of phase. The intensity of the waves is registered by an X-ray detector resulting in a diffraction image. This diffraction image can be further analysed to determine the electron density of the molecule [48]. In this thesis, data from the X-ray diffraction were collected and auto processed on the beamline ID30B at European Synchrotron Radiation Facility (ESRF) before it was further processed using Molecular replacement.

Phasing

Molecular replacement (MR) is a commonly used phasing method in X-ray crystallography for determining the electron density map of an unknown crystal structure using a previously solved structure of a related molecule. The related molecule is assumed to approximate the unknown protein structure. The diffraction intensities of the unknown molecule can therefore be compared to the diffraction intensities of the related molecule to determine at what orientation the model best matches the observed diffractions. The phases from the known structure are then "borrowed" to the new structure, and an initial electron density map is calculated using these phases and amplitudes from the experimental data as an approximation [49].

Model building and refinement

Once a first draft of the electron density map is constructed. The search model needs to be corrected to match the data from the data collection of the crystal. The model is built by introducing changes in the electron density maps F_o-F_c and $2F_o$-F_c, such as substitution, insertion and deletion of amino acids, the addition of metal ions, and introduction of water molecules. The quality of the electron density map can be improved further using model refinement [38]. Refinement is an iterative process that tries to minimize the differences between the observed diffraction amplitudes for the experimental data (Fo) and the diffraction amplitudes from the predicted data (Fc) while optimizing the geometry of the molecule. After each refinement, the molecule is modelled again. Several refinements and modelling cycles are performed where the result of each refinement is used to improve the model further. The cycles are then repeated until the model has reached acceptable convergence values [50].

One of the values for checking the quality of the structure obtained from the refinement are R-factor. R-factor measures how correctly the observed amplitudes (F_o) matches the predicted amplitudes from the model (F_c). The R-factor should typically be below 20-25 % for the data to be a good fit. A perfect fit would be 0%, whereas a totally random molecule set would have an R-factor around 63% [50, 51]. R_{free} is another factor for evaluating the data. R_{free} is an unbiased factor used to keep track of the refinement process. This factor sets aside a subset of the experimental data to be used for cross-validation. If the refinement process is successful, the R_{free} factor will drop below 30% [50].

2. Aims of study

The main aim of this study is to begin the characterisation of suitable enzyme candidates for CO_2 conversion to oxalate. This will be done through both computational work and experiments in the laboratory.

❖ **Express and purify TyrBm WT and mutants of this protein -F227Y and -F65Y**
   Express and purify TyrBm WT and two mutants of this protein. Compare how the mutants express and purify with the WT.

❖ **Enzyme assay development**
   Develop a suitable enzyme assay for comparing the activity of WT TyrBm to the mutants F227Y and F65Y. See if oxalic acid can bind to the enzymes and analyse how different concentrations of oxalic acid influence the activity of the enzymes.

❖ **Crystallize TyrBm WT and mutants of this protein -F227Y and -F65Y**
   Compare how the WT crystallize with crystallization of the mutants. Solve 3D structures of F227Y using X-ray crystallography. Analyse changes in the active site induced by the mutations.

❖ **Analyse if xTB -GFN -FF can reproduce and model variants of the enzyme structures**
   Perform geometry optimization with xTB -GFN -FF on different models of the structure from the crystallization. Compare how the active site change with different metal ions (Cu(I), Cu(II) and Zn(II)) and different forms of H_2O, OH^-, and O^{2-} bridging the metal ions.

3. Methods

The methods applied in this master thesis involved expression, purification, development of enzyme assays and crystallization of TyrBm WT protein and two mutants (TyrBm WT, TyrBm F227Y and TyrBm F65Y). The experiments were done to analyse the activity of the enzymes, how they expressed and purified, if they crystallized, and if they could bind oxalic acid. The binding of oxalic acid might give a good indication of whether the enzyme could be able to convert CO_2 to oxalate in future COOFIX research. Additionally, to experiments at the laboratory, there was done some experiments *in silico* using xTB -GFN-FF for geometry optimization of the protein. Content of all buffers and solutions used in this project are listed in the appendix.

3.1 Protein expression and bacteria growth

3.1.1 Expression

Since *E.coli* is so easy to modify genetically, it has become a popular, well-accepted tool to express recombinant protein [52]. *E. coli* is known for its high growth rate, making it possible to translate proteins fast in large quantities. [53] In order to achieve protein expression, the transformed cells were prepared in a sterile UV bench by inoculating the cells with 50 mL LB (Lysogeny broth) media and 50 µl ampicillin (100 mg/ml). The culture was then incubated overnight (12-16 hours) at 30 ºC and 200 rpm.

The TB media were incubated at 30 degrees to reach the same temperature as the overnight culture. After incubation, 25 mL of phosphate buffer were transferred by falcon tubes to each of the 450 ml TB (Terrific broth) media flasks in addition to 500 µl of ampicillin (100 mg/ml) stock solution, 25 ml of the overnight culture and 200 µl antifoam solution. The flasks were mixed by gentle spinning and incubated further in the Harbinger bubbling system at 30 ºC. When the OD_{600} reached approximately 0.8, the temperature was lowered rapidly to 20 ºC. The proteins were then induced by adding 500 µl of 0.5M isopropyl β-D-1-thiogalactopyranoside (IPTG) to each flask. The induced cell cultures were left overnight to grow at 20 ºC (12-15 hours). The cell culture was harvested by centrifugation at 45000 rpm for 15 min at 4 ºC before the supernatant was decanted off and discarded.

The pellets were transferred to a flacon tube with 40 ml lysis buffer and resuspended with a small "metal ball". 20 mL of this solution were transferred to a new flacon tube, and another 20 mL of lysis buffer were added to each falcon tube so that they both contained approximately 40 mL. The solutions were mixed by shaking and stored at -20 ℃.

3.2 Protein Purification

3.2.1 Sonication

The frozen cells were heated up to 4 ℃ and sonicated on ice for a total of 20 min with intervals of 30 seconds on, 30 seconds off. This method uses pulsed sound waves with high frequency to break open the cells, releasing the components of the cells into solution [54]. The sonicated cells were then centrifuged at 20000 rpm for 30 min to remove the unwanted cell components. The supernatant containing the protein of interest were collected while the pellet was thrown away.

3.2.2 Immobilized metal affinity chromatography (IMAC)

The proteins of interest were prior to expression and purification transformed into *Escherichia coli* strain BL21 by Gustavo Santos and Andrea Nikoline Englund. The transformation was done with the expression plasmid pET22b modified with a self-cleaving protease tag made up of the *Vibrio cholerae* MARTX toxin cysteine protease domain (CPD). The CPD tag is an autoprocessing enzyme fused to a C-terminal histidine tag (6xHis-tag) [55]. The histidine tag has an affinity for metals such as Ni^{2+} in metal affinity chromatography and can therefore be used for affinity purification. The CPD enzyme is highly specific and will, in the presence of phytic acid, cleave itself off after a leucine residue within the junction of the target protein, see Figure 3.1. This allows for simplified protein purification with affinity purification, cleavage, and tag separation in one stage [55]. After the transformation, the bacteria were stored as glycerol stocks at -80 ℃.
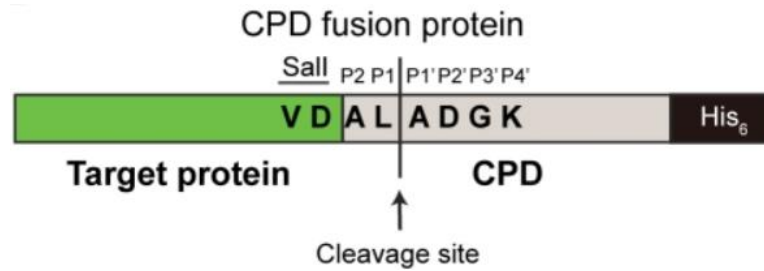
Figure 3.1. Schematic of CPD tag fused to 6x histidine tag. The protein cleave itself off after a leucine residue within the junction of the target protein. The cleavage site is indicated with an arrow.

IMAC was used to purify the recombinant proteins even further. The proteins were purified using a BigTrap machine, shown in Figure 3.2. IMAC separates proteins based on their interactions with a transition metal ion. The transition metals are immobilized on a matrix. Proteins containing an affinity tag of histidine residues will have a higher affinity for the matrix [56]. In this project, Ni-NTA affinity chromatography was used. Ni-NTA uses nickel ions (Ni^{2+}) as transition metal. These have a high affinity for binding histidine residues, such as the six consecutive histidine residues (6xHis-tag) fused to the CPD protease used in this thesis. This allows for high selectivity for the protein [57].

The supernatant from the centrifugation containing the protein was diluted to 50 mL with 20 mM imidazole. The HisTrap FF crude 1mL column was first calibrated by washing it with 20 mM imidazole before the protein solution was added to the column. Proteins with a His-tag will here bind to the column, thereby the protein of interest with CPD tag. Proteins that did not bind was collected as a flow-through fraction. To make sure that only proteins with a high affinity for the column was left, the column was washed with a lysis buffer. This was to remove any impurities if there were proteins bound with low affinity for the matrix. The wash was collected as a fraction. The protein of interest where finally eluted with a cleaving buffer. This buffer contains phytic acid and cleaves the CPD-tag off the protein. The protein will no longer be bound to the column and therefore elute with the cleaving buffer into a new fraction. The CPD tag is cleaned off the column by eluting it with a buffer with a high imidazole concentration (0.4 M). The imidazole will compete with the His-tag for binding to the column, and the CPD-tag will elute.

Figure 3.2. BigTrap used for Immobilized metal affinity chromatography (IMAC). Self-cleavable tag purification, a protocol for the robot made by Åsmund Røhr Kjendseth was used for the purification step. The Ni-NTA affinity chromatography columns are shown in the picture in blue and red. All fractions were collected in falcon tubes in front.

Second purification

A second purification of the protein samples was done to check if the CPD-tag was present in the protein solution after the first purification. For the second purification, the BIORAD, BioFrac™ Fraction Collector was used with HisTrap FF crude 1mL columns. 20 mM imidazole buffer was added to equilibrate the columns before the solution containing the protein of interest was added. If the CPD -tag were present, it would bind the column with strong affinity. Wash buffer with a low concentration of imidazole (20mM) was added to elute the protein; the fraction was collected in three fractions. Finally, elution buffer was added to elute the CPD -tag, and this was collected as a separate fraction.

3.2.3 SDS – PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze the fractions from the purification. SDS-PAGE separate molecules in an electric field based on their molecular weight. The movement of molecules is determined by the magnitude of the electric field, net charge, and molecular radius. To separate the proteins only by molecular weight SDS is added to the loading buffer. SDS can, together with heath, disrupt the tertiary structure of the proteins and make them linear[58]. The linear polypeptide chain can then bind SDS in proportion to its relative molecular mass and mask the initial charge of the protein. Since the charge to mass ratio is similar for a lot of SDS-denatured proteins, the separation in the gel will mostly be based on differences in molecular mass. Smaller fragments will migrate faster through the matrix, while larger fragments will migrate slower [58].

All the fractions (flow-through, wash, eluted protein for both WT, F227Y and F65Y) from the affinity chromatography where added loading buffer in a ratio 1:4 and heated for 10 minutes at 70º at 750 rpm to denature the proteins. Loading buffer is a dye used for tracking the migration of the samples in the electrophoresis. It contains SDS and adds density to the sample to make the application onto the gel easier [59]. The samples were applied onto a Mini-PROTEAN TGX Stain-Free Precast Gel with 10 µL of each fraction in individual wells. 5 µL of BenchMark™ Protein Ladder (Life Technologies AS) was added to evaluate the molecule sizes after electrophoresis. The gel chamber was filled up with MOPS running buffer, and the samples were then run for 30 minutes at 200V.

3.2.4 Concentration

The protein solution was concentrated using Amicon® Ultra Centrifugal Filters. This is a filter with a molecular weight cut-off value at 30 kDa, which will separate and clean the protein from some of the solvents. The protein solution was measured on NanoDrop A260/280 with the extinction coefficient calculated from the protein sequence in uniport (UniProtKB - B2ZB02). This was to estimate the concentration of protein accurately before and after it was concentrated. The solutions were washed by MOPS buffer (20 mM) several times to ensure better storage of the protein. The protein solutions were then stored at 4 ºC prior to crystallization or used in protein assay.

3.3 Crystallization

3.3.1 Crystallization

Before the first crystallization, the concentrated protein solution was stored at 4 ℃ for approximately one week. Some of the protein in the mutant was starting to separate, but the wild type looked unchanged. In the second crystallization, the protein was used the day after it was purified.

3.3.2 Hanging drop crystallization

Hanging drop crystallization was used to obtain crystals of the TyrBm protein. The crystallographic conditions were inspired by those developed by Sendovski, Kanteev [60]. The reservoir solutions for this were made in 15 mL falcon tubes by mixing the solvents presented in Table 3.2. There were made two different reservoir solutions, one obtaining zinc acetate and one where the volume of zinc acetate was replaced with MQ H2O. Zinc acetate was added because metal ions like zinc acetate have a known effect to bridge and stabilize intermolecular contacts in crystals, and this seemed to work the best in the previous study [46, 60]. The reservoir solutions were then distributed on a 24-well crystallization plate with an increasing concentration of PEG8000 precipitant (14-24%). Each of the wells contained 600 µl reservoir solution.

Table 3.1. The concentration of protein samples after accumulation of protein using Amicon® Ultra Centrifugal Filters. These concentrations of the protein were used for crystallization.

Protein sample	Concentration (mg/mL)
TyrBm WT	7
TyrBm F22Y	15.5
TyrBm F68Y	15.6

Table 3.2 Reservoir solutions RS 1 – RS 6 for hanging drop crystallization with zinc acetate. A separate crystallization was done where the volume of zinc acetate was replaced with MQ H_2O. The columns show the amount of stock solution added for each reservoir solution and the final concentration. The total volume obtained from each sample is displayed in the bottom row.

Stock solution	RS 1	RS 2	RS 3	RS 4	RS 5	RS 6
50% PEG8000	1 mL	1.2 mL	1.4 mL	1.6 mL	1.8 mL	2.0 mL

	(10%)	(12%)	(14%)	(16%)	(18%)	(20%)
Zinc acetate 1M	1 mL (0.2 M)	1 mL (0.2 M)	1 mL (0.2 M)	1 mL (0.2 M)	1 mL (0.2 M)	1 mL (0.2 M)
Cacodylic acid 1M, pH 6,5	500 µL (0.1 M)	500 µL (0.1 M)	500 µL (0.1 M)	500 µL (0.1 M)	500 µL (0.1 M)	500 µL (0.1 M)
MQ H20	2.5 mL	2.3 mL	2.1 mL	1.9 mL	1.7 mL	1.5 mL
Total volume	**5 mL**	**5 mL**	**5 mL**	**5 mL**	**5 mL**	**5 mL**

The crystals were formed by mixing 2 µL protein solution and 2 µL reservoir solution directly on silica-coated glass. The concentration of the protein solutions is presented in Table 3.1. The large droplet was further separated into 5-7 smaller droplets on the cover glass and placed upside down on the crystallization plate. The plates were then stored at room temperature (20 ℃) until crystals were formed.

3.3.3   X-ray diffraction, refinement, and model building

Data from the X-ray diffraction were collected and auto processed on the beamline ID30B at European Synchrotron Radiation Facility (ESRF). To prevent some radiation damage caused by the X-ray beams in the synchrotron, the crystals were transferred to a solution with cryoprotectant and flashed freezed in liquid nitrogen beforehand of the data collection. Both the data collection and freezing of the protein crystals was performed by Åsmund Røhr Kjendseth. Further processing of the crystal data was performed in collaboration with Andrea Nikoline Englund. Molecular replacement was used to solve the structure with TyrBm (PDB ID: 3NM8) from the PDB database as the search model, and **ccp4i2** was used as a software to perform the phasing and refinement. To calculate the R-free factor, 5% of the experimental data was set aside as a cross-validation, the remaining 95% of the experimental observations were used in the refinement process. R-factor and R-free were recalculated and checked several times throughout the refinement process to access the structure quality of the data. **Coot** [61] was used to manually inspect the model between each refinement and to correct the $F_o – F_c$ electron density map where positive and negative peaks indicated errors in the sidechain or presence of metal ions. When the model had reached adequate values for R-factor and R-free. Waters were added around the molecule with the "Add waters" function in ccp4i2 before a final refinement.

3.4 xTB-GNF- FF modelling

After the final refinement, the model was converted to a PDB file and processed with AmberTools20. Hydrogen atoms were added to the molecule, and other atoms locked into place before the molecule was geometry optimized. Since there was no way of knowing for sure the form/charge of the bridging molecule in between the metal ions in the active site, the model was split into three different models. Model 1 was modelled with water in the active site, Model 2 had OH^- in the active site, and Model 3 had O^{2-} in the active site. The models were then energy optimized with xTB-GFN- FF with both Cu(I), Cu(II) and Zn(II) in the active site separately. The computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway. To compare changes in the active sites the new models were aligned against the model from the final refinement in Pymol. The nitrogens on the histidine molecules were aligned using the Pair Fitting feature.

3.5 Activity assay development

Enzyme assays is a widely used method for characterising enzyme activity [62]. It can be used for identifying the presence or absence of a certain enzyme or to establish the amount of enzyme or substrate in a sample. Factors that need to be considered when running an enzyme activity assay are pH, temperature, ionic strength, and concentration of substrate and enzyme added. The assay should be performed under the enzyme's optimum conditions because that's when the enzyme is the most active, but for convenience, the preferred temperature is normally 25º [63]. Assays are usually measured by determining the reduction of substrate or increase in product over a set period. This can be measured by different analysis techniques such as UV- spectrophotometry.

The principle behind UV- spectroscopy is Beer Lamberts Law. Beer Lambert Law states that there is a directly proportional relationship between the absorbance of light and the concentration of the material [64]. UV- spectrophotometry measure absorbance at defined wavelengths either continuously or at set moments. Continuous assays have the advantage of monitoring the enzyme activity throughout the whole process, from start to end [62]. The enzyme velocity can then be derived from the linear part of the reaction curve by calculating the amount of product formed per unit of time [63].

In this thesis, caffeic acid was used as the substrate for Tyrosinase in the activity assay. Caffeic acid is a substrate known for its similarity to L-DOPA normally used in tyrosinase assays. Tyrosinase catalyses the oxidation of caffeic acid to o-caffeoquinone, and the formation of o-caffequinone can be measured by Uv-Spectrometry [65, 66]. It has previously been discovered non-competitive inhibition of mushroom tyrosinase by oxalic acid and its salts (oxalates). This happens because a chelate linkage is formed with the functional copper ions in the enzymes active site [67]. This shows that oxalate can bind to mushroom tyrosinase. Since oxalate can bind to this enzyme, there is a chance it can bind to tyrosinase from *B. megaterium.* The binding of oxalate proves that oxalate can fit in the active site. This increases the possibility of CO_2 formation to oxalate happening in the enzyme.

After sufficient purification of the proteins, an activity assay developed. All the assays were prepared using the Opentrons pipetting robot. Assays were first designed using L-Dopa as the substrate. The concentration of the enzyme was tested in small tubes in a wide range to estimate potentially suitable concentrations of enzyme and L-Dopa. The concentrations that ended up being tested was in the range of 0.04 – 0.24 mg/mL for the enzyme and 2 mM substrate. The enzyme stock solution was added $CuCl_2$ to a final concentration of 5 µM prior to the assay. The absorbance was measured on a spectrophotometer at 475 nm.

Caffeic acid was also tested as a substrate for the assay as L-DOPA seemed to be unstable over time. At first, there was a trial-and-error procedure where different concentrations of substrate, enzyme, and copper concentration were tested. The assays were carried out by an OpenTrons pipetting robot at 25 ºC in 200 µl reaction mixtures on 96-well plates (Nest). The reactions were then incubated for 20 min before they were stopped by 0.6M HCl. The absorbance was measured at 473 nm. After a lot of optimizations of the assay, the concentrations of copper needed in the stock solution were decided to be 0.2 mM to a 0.2 mM protein concentration. A dilution series of substrate concentrations ranging from 0 – 5.41 mM and different enzyme concentrations were tested for all the enzymes.

For the assay with oxalate, it was decided to use a substrate concentration of 5 mM, and enzyme concentration of 8 and 16 µM for the WT and 8, 16 and 24 µM for the mutants (F227Y and F65Y) as these worked well in the previous assays. Assays with different concentrations of oxalate were set up to test the concentration needed for inhibition. In the final assay, the oxalate concentration

ranged from 0-35 mM. Kojic acid was used as a positive control for inhibition as kojic acid is known for inhibiting tyrosinase activity [68]. The different concentrations of oxalate were tested in triplicates. Samples without inhibitor present were tested in duplicates. The setup for the enzyme assay is shown in Figure 3.3. The assay was prepared using the steps presented below.

Step 1: 50 µL of the enzyme (WT, F224Y and F65Y) were transferred from the SARSTEDT 96-well plate in respective wells on the NEST 96 well plate.

Step 2: 75 µl of inhibitor (oxalic acid and kojic acid) with different concentrations were added to their respective enzyme solutions in the NEST 96 well plate, mixed well, and incubated for 30 min.

Step 3: After incubation, the solutions were added 75 µL of caffeic acid, mixed and measured at Varioskan LUX Multimode Microplate Reader after 10 min. The samples were then measured continuously at 473 nm for 50 min with readings every 1 min. The temperature in the reader was 30,00 ºC, and the samples were shaken at 600 rpm before every read. For further clarifications about the final 96-plate see FigureC.1 in the appendix.
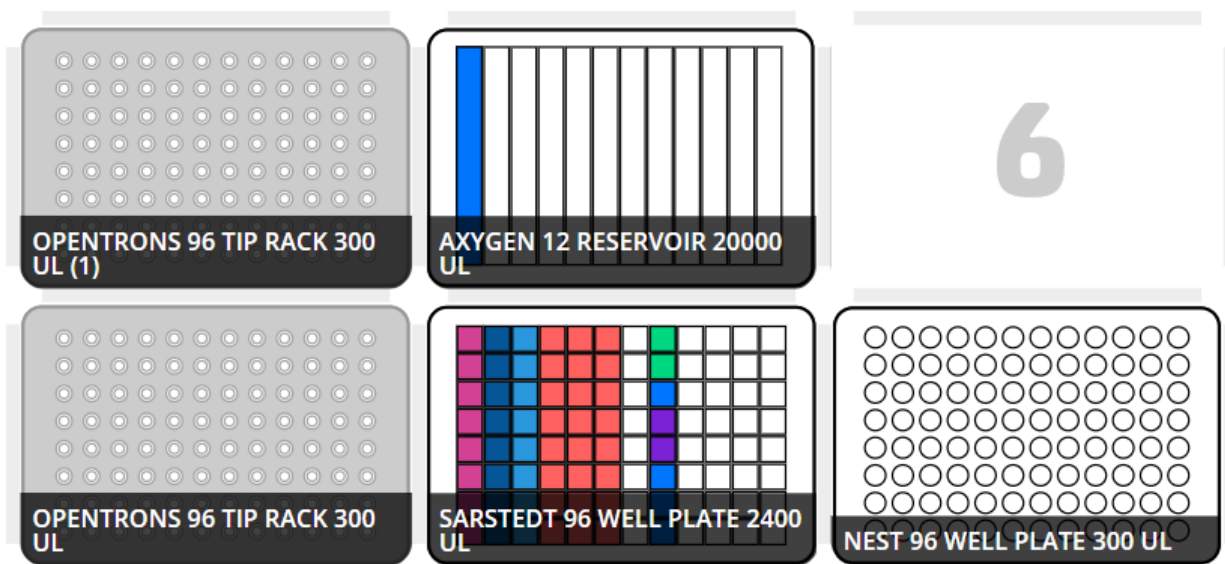
Figure 3.3. Set up for assay with oxalic acid in OpenTrons protocol designer. SARSTEDT 96 WELL PLATE contained: Pink – TyrBm WT, dark blue – TyrBm F227Y, light blue - TyrBm F65Y. Red- oxalic acid in different concentrations (increasing concentrations from top to bottom), green – HCl. The tree blue wells to the right contained kojic acid in different concentrations (increasing from top to bottom). Purple – MOPS buffer (20mM). AGYGEN 12 RESERVOIR contained the substate caffeic acid.

Blank samples were prepared the same way as the rest of the samples but not added enzyme solution. The blanks were prepared as duplicated and were measured for 20 min at 473 nm.

4. Results

It is an essential component of the COOFIX research project to find suitable enzyme candidates for the conversion of CO_2 to oxalate. In this master thesis, purification, enzymatic assays and X-ray crystallography was used for beginning the characterization of TyrBm WT from *B. megaterium* and the mutants -F227Y and -F65Y. xTB -GFN -FF were used to characterize the mutant F227Y further. Results from these experiments will be presented in this section.

4.1 Expression and purification of TyrBm

Expressing and purifying the protein from other proteins and cell components are essential for crystallization and analysing the activity of the proteins. A general protocol for expression and purifications of copper proteins were established in the laboratory and required some optimization for Tyrosinase from *B. Megaterium*. The prepared glycerol stocks from the transformation of the proteins were used several times for optimization of the expression and purification step of TyrBm and the mutants.

4.2 SDS-PAGE

SDS-PAGE was used to assess the purity of the protein after purification and the amount of protein present in the sample. A gel picture from one of the purifications of TyrBm WT and the mutants F227Y and F65Y are shown in Figure 0.1. What each well contained are presented in Table 0.1.

Table 0.1. Outline of the application of samples onto the SDS PAGE gel presented in Figure 0.1. The wells are numbered 1-15. FT – Flow-thorough fraction (1, 6, 11) , W – wash fraction (2, 7, 12), CPD tag (3, 8, 13), P -eluted protein (4, 9, 14). L – Molecular weight marker (5, 10, 15).

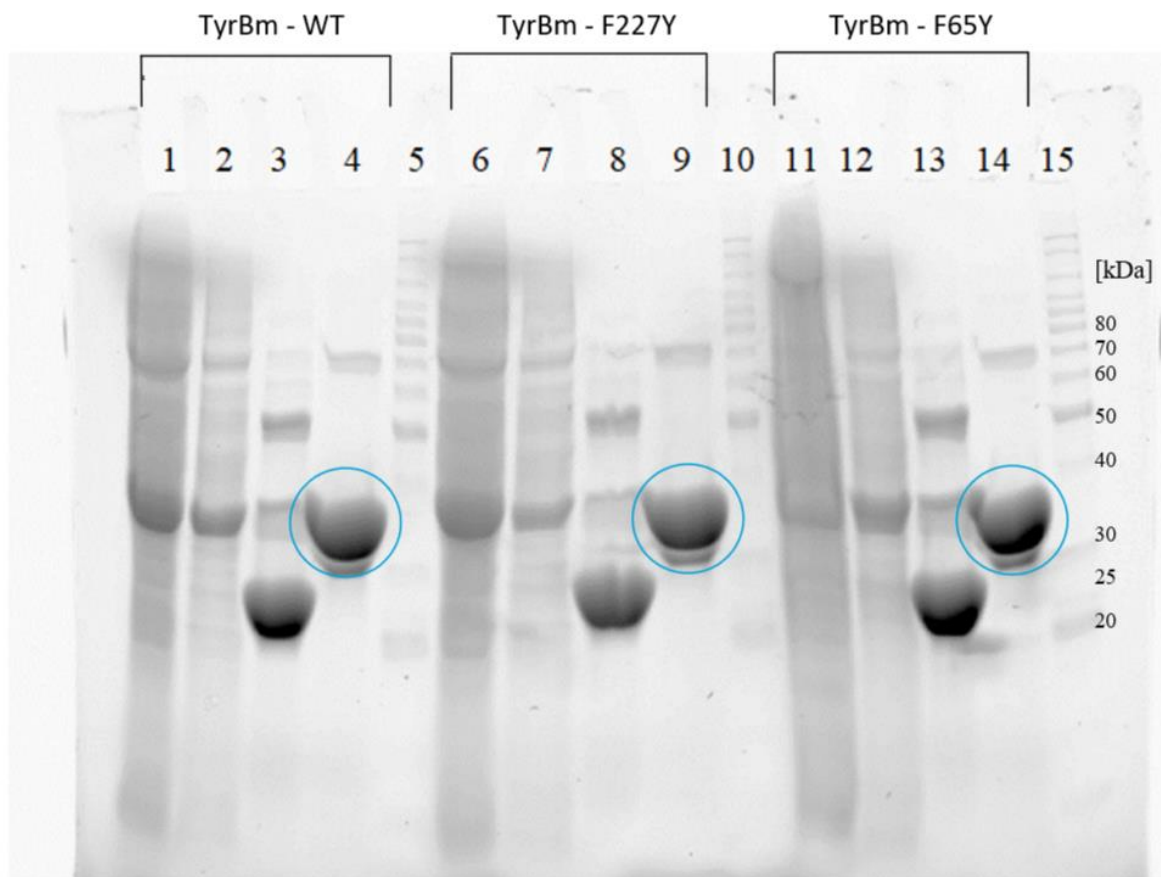| TyrBm -WT					TyrBm F227Y					TyrBm F65Y				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
FT	W	His	P	L	FT	W	His	P	L	FT	W	His	P	L

Figure 0.1 SDS PAGE of TyrBm WT and the mutants F227Y and F65Y following the purification by IMAC. Well 1-4, 6-9 and 11-14 are different fractions of TyrBm WT, TyrBm -F227Y and TyrBm -F68Y, respectively, from the purification of the proteins. The molecular weight marker BenchMark™ Protein Ladder (Life Technologies AS) was applied in well 5, 10, and 15 to evaluate the molecule sizes after electrophoresis. What each well contain are displayed in Table 0.1. The eluted protein is displayed in well 4, 9 and 14 with a blue circle marked around the protein of interest.

The result of the purification shows fragments in all the wells. The molecular weight marker BenchMark™ Protein Ladder in lane 5, 10 and 15 makes it possible to estimate the size of these fragments. Lane 1, 6 and 11 shows the flow-through, this contains all proteins that did not bind to the column because it had little or no affinity to the Ni-NTA resin. Lanes 2, 7 and 12 contain the wash fraction; this contains all protein that binds the column with low affinity. Lanes 3, 8 and 13 contain the eluted CPD tag, whereas lanes 4, 9 and 14 contain the eluted protein. The expected molecular weight of TyrBm monomer is approximately 35 kDa. This corresponds with the molecular weight of the fragments presented in these lanes with a blue circle around. The band in

these lanes are strong and dark, indicating that there is a high amount of protein present in the sample. There also appears to be a band around ~70 kDa and below 30 kDa in the same lanes.

Since the band containing the protein of interest was so thick and smeared, it was hard to tell if this was a single band or if it was more fragments fused together. A dilution series of TyrBm WT and mutant F65Y on SDS PAGE was generated to analyze this band further. The dilution series are presented in Figure 0.2.
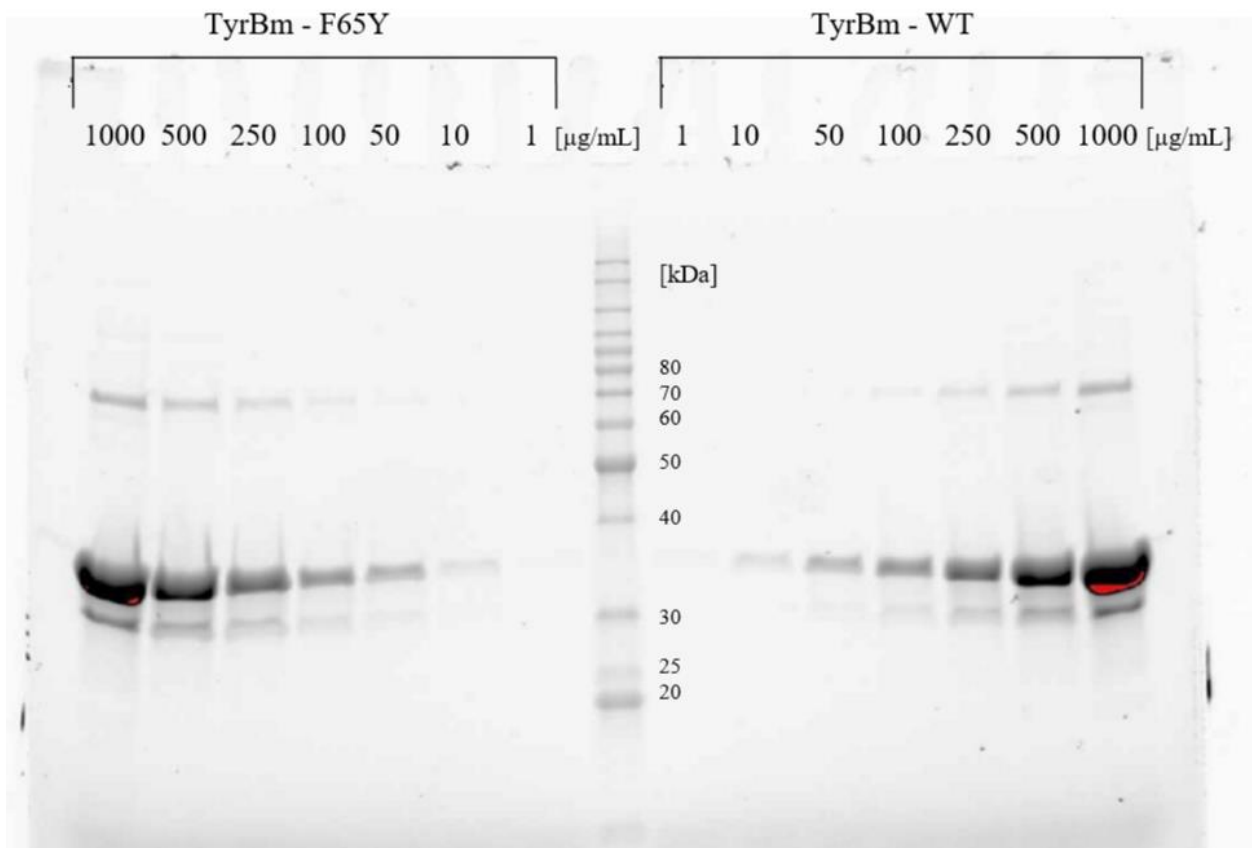

Figure 0.2 Dilution series of TyrBm WT and mutant F65Y on a SDS PAGE. Well 1 - 7 and 9 - 15 are different protein concentrations given in µg/mL of the different proteins. The concentration of protein that were tested ranged between 1 - 1000 µg/mL. BenchMark™ Protein Ladder (Life Technologies AS) was applied in well 8 to evaluate the molecule sizes after electrophoresis.

The dilution series consisted of different concentration of TyrBm F65Y and TyrBm WT in the range from 1000 - 1 µg/mL. The dilution series shows strong bands in both lanes with a high concentration of protein (1000 µg/mL) which gradually gets weaker as the concentration decreases. In the lane with the lowest concentration (1 µg/mL), there is little or no band present.

All lanes with fragments show the same fragment pattern as in Figure 0.1, with one fragment around 70 kDa, one around the fragment size of the protein 35 kDa and one fragment around 30 kDa.

A second purification of the WT was done to check if the band around 30 kDa could be the CPD tag. The gel from the second purification is presented in Figure 0.3.
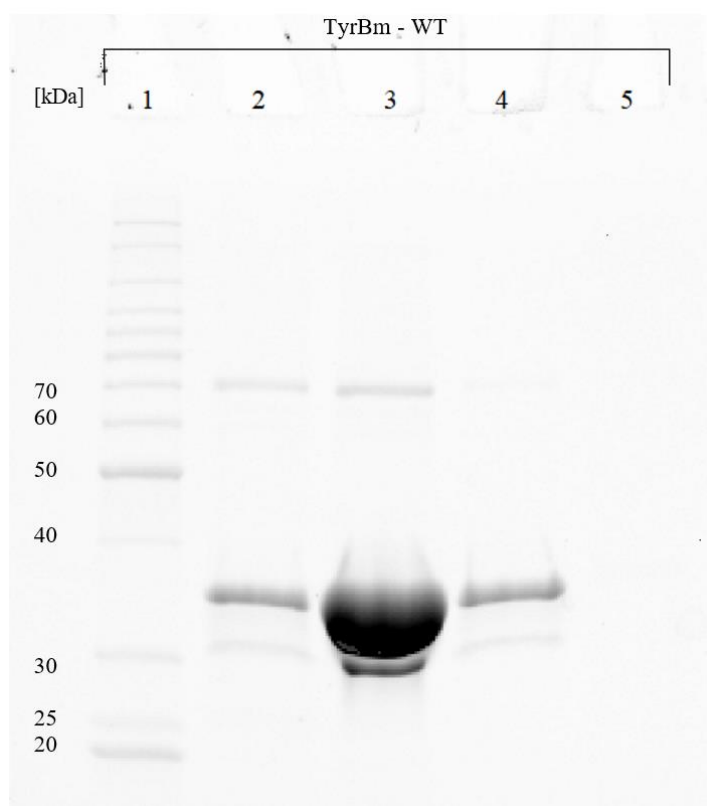
Figure 0.3. SDS PAGE after second purification of TyrBm WT. BenchMark™ Protein Ladder (Life Technologies AS) was applied in well 1. Well 2, 3 and 4 are flow through fraction 1, 2 and 3 respectively, eluted with a low concentration of imidazole (20 mM). Well 5 contain fraction eluted with a high concentration of imidazole (0.4 M) and would be the CPD tag with the His-tag if present.

The SDS gel, after the second purification, shows fragments in well 2, 3, and 4 around the estimated size of the protein. These lanes all show the flow-through fraction where the protein was eluted gradually. Most of the protein eluted in lane 3 where the strongest black band is present. Even after the purification, there is still two more bands present ~70 kDa and ~30 kDa. Well 5 contain a fraction that was eluted with high concentration imidazole and would be the CPD tag

with the His-tag if present. There was no band present in this lane. The CPD tag was therefore not present.

4.3 Activity assay

An activity assay was designed to obtain information about the activity of the enzymes and see if they could be inhibited by oxalate. The assay was performed using a Opentrons OT-2 pipetting robot in 96 -well plates and measured by a microplate reader at 473 nm. The final concentration of the enzymes in the assay was 10 µM for the WT, 24 µL for F227Y and 24 µL for F65Y. The substrate concentration was 5 mM. The final product of the 96 well plate one hour after the substrate was added are presented in Figure in the appendix. The result from the activity assay without inhibitor present are presented in Figure 0.4. The result of the activity assay with different concentrations of oxalic acid as inhibitor are presented in Figure 0.5, Figure 0.6, and Figure 0.7 for the WT, F227Y and F65Y, respectively. Each of the points on the graphs was calculated from the mean of three parallel samples. The standard deviation for all the measurements was calculated using Excel and are presented as error bars on each of the graphs. Kojic acid was, as mentioned, used as a control for inhibition by oxalic acid. The result of the activity assay with different concentrations of kojic acid is presented in Figure , Figure , and Figure.

The blanks used in the activity assays were measured for 20 minutes. The mean and standard deviation for the absorbance at 473nm for the blank was calculated to 0.0421 ± 0.0012. The blank was plotted as the mean with error bars as the standard deviation in Figure 0.5, Figure 0.6, and Figure 0.7.

The enzymes used for enzyme assay were stored in the fridge at 4ºC in MOPS buffer (20 mM) for approximately a week after purification was done. Precipitation occurred on all the enzyme variants TyrBm- WT, F227Y and F65Y. The mutants, however, precipitated a bit more and faster than the WT. The concentration of the enzymes was measured directly after purification to 21 mg/mL, 9 mg/mL, and 13 mg/mL for the WT, F227Y and F65Y, respectively, and again before the enzyme assay; after precipitation had occurred. The result showed no significant changes in the concentration of the enzymes.
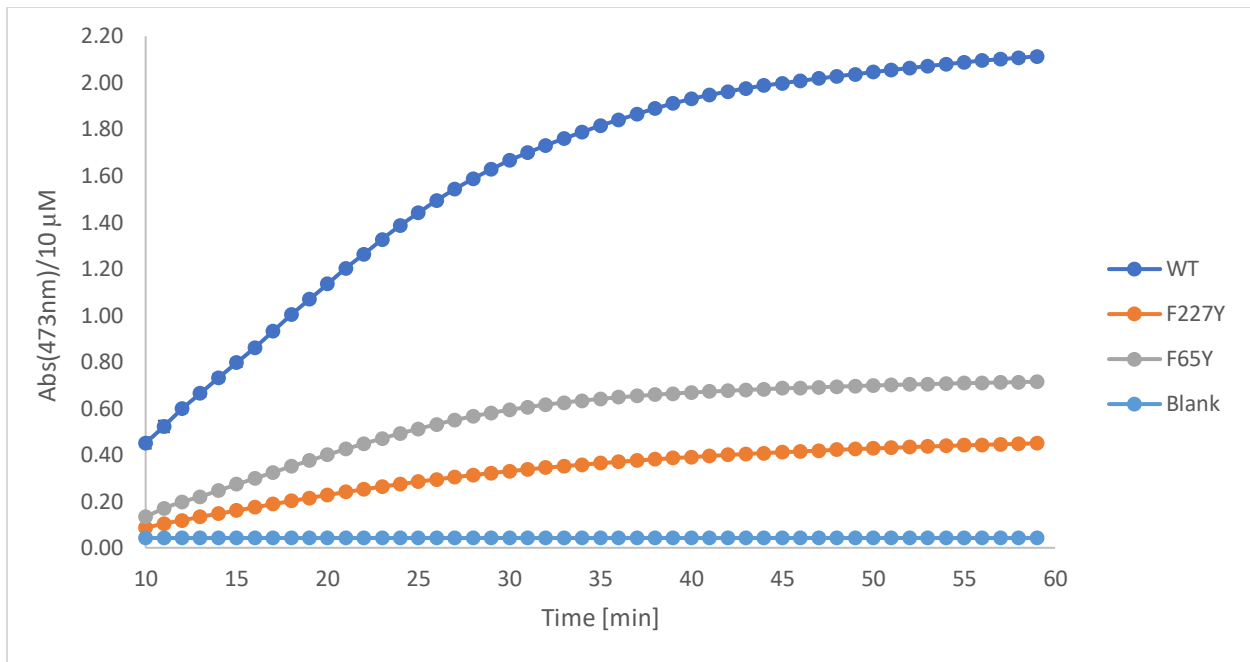
Figure 0.4. Activity of TyrBm WT and the mutants -F227Y and -F65Y. The Y-axis displays measured absorbance at 473 nm/10 µM for each of the samples, whereas the X-axis display time in minutes. All curves have the standard deviation displayed as error bars for each measurement.

Figure 0.5. Activity assay with oxalate for TyrBm -WT. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of oxalate added to each sample are displayed to the right. All curves have the standard deviation displayed as error bars for each measurement.
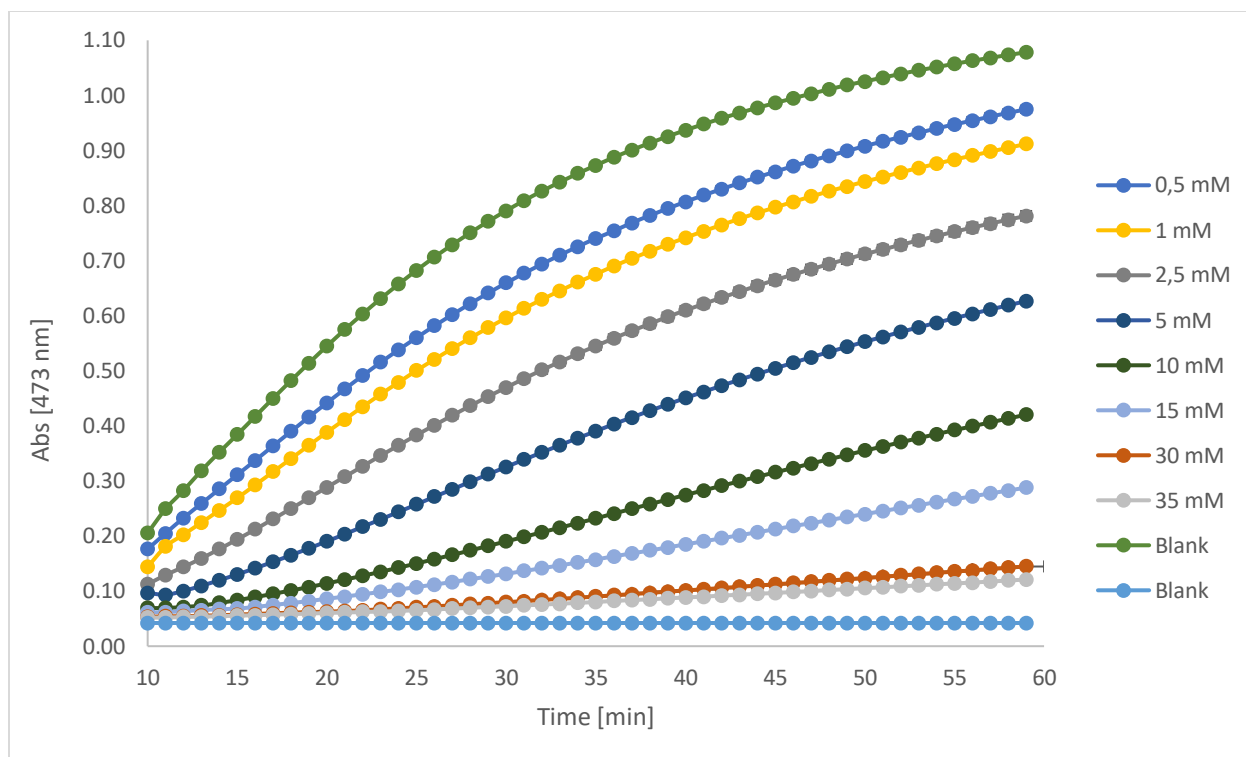
Figure 0.6. Activity assay with oxalate for TyrBm -F227Y. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of oxalate added to each sample are displayed to the right. All curves have the standard deviation displayed as error bars for each measurement.
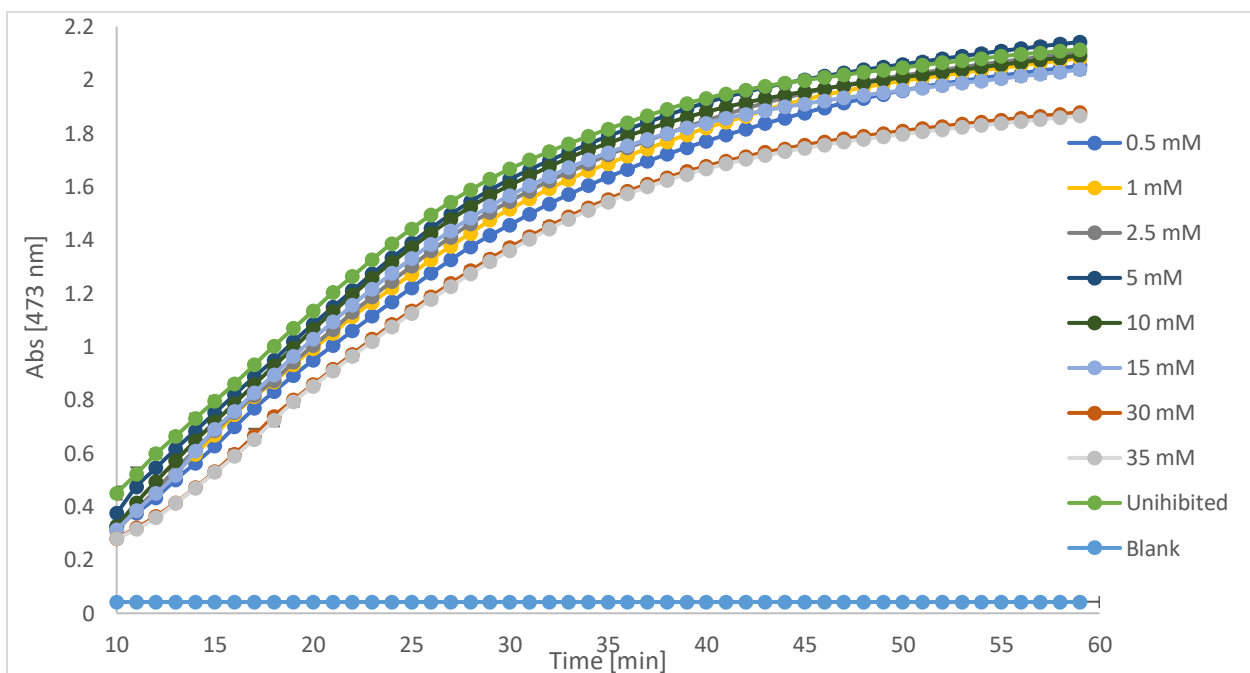
Figure 0.7. Activity assay with oxalate for TyrBm -F65Y. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of oxalate added to each sample are displayed to the right. All curves have the standard deviation displayed as error bars for each measurement.
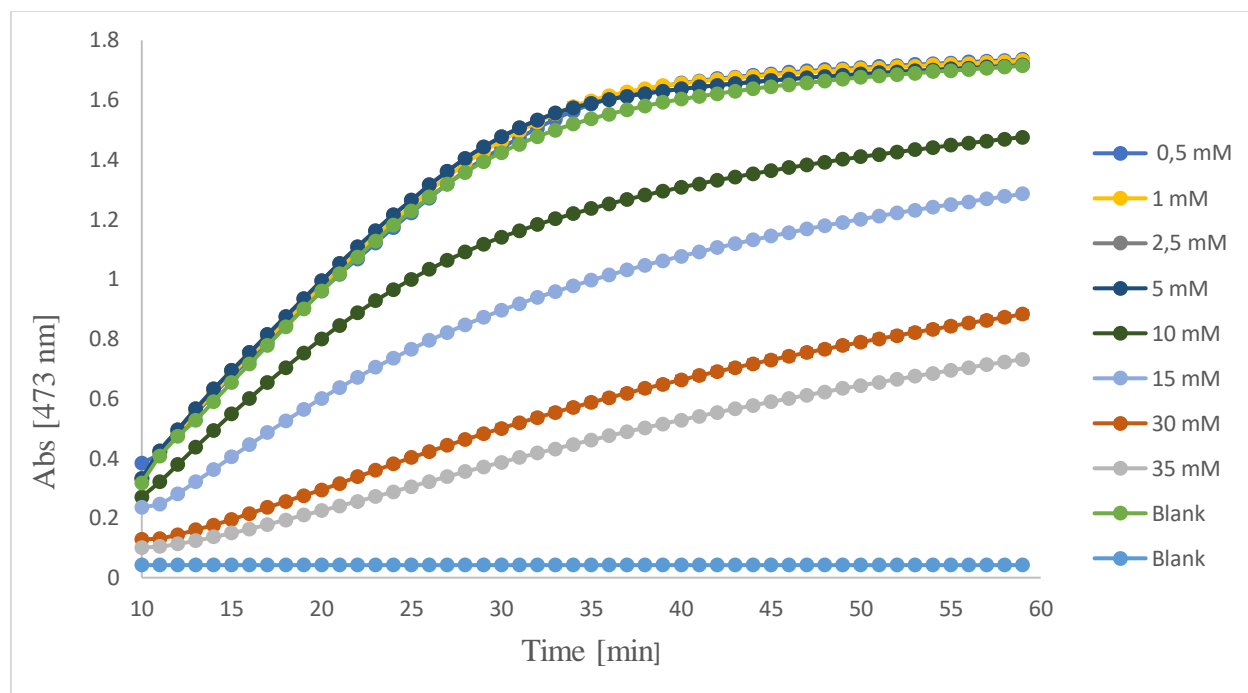
The result from the enzyme activity assay without inhibitor present in Figure 0.4 show activity for all the enzymes. However, there is a big difference in activity between the mutants and the WT. At the starting point (10 min), the WT had more than five times the activity of F227Y and more than four times the activity of F65Y. The slope of the absorbance for the WT is also steeper than the other two curves. F65Y show the highest absorbance among the mutants.

The graphs in Figure 0.5, Figure 0.6, and Figure 0.7 show little inhibition for the WT even for higher concentrations of oxalic acid. The WT show signs of inhibition with an oxalic acid concentration around 30-35 mM. F227Y shows inhibition from 0.5 mM oxalic acid with a constant decreasing absorbance with increasing concentration of inhibitor. F65Y show signs of inhibition around 10 mM oxalic acid and then follow the same trend as F227Y.

The activity of the enzymes shows the rate of the reaction catalysed by the enzymes, expressed in how much substrate that is transformed per minute [69]. To calculate the activity of the enzymes, the linear range from 10-25 min for all the graphs in Figure 0.5, Figure 0.6, Figure 0.7, Figure 1, Figure 2, and Figure3 were used to fit linear regression models. The linear regression slope for all

graphs is presented in Table. The slope was then divided by the concentration of the enzyme to get the reaction rate of 1 µM enzyme per minute, which were plotted against the inhibitor to get the activity. The result of the graph is displayed in Figure 0.8. Kojic acid was tested in the range of 0-25 mM, while oxalic acid was tested in the range of 0-35 mM.
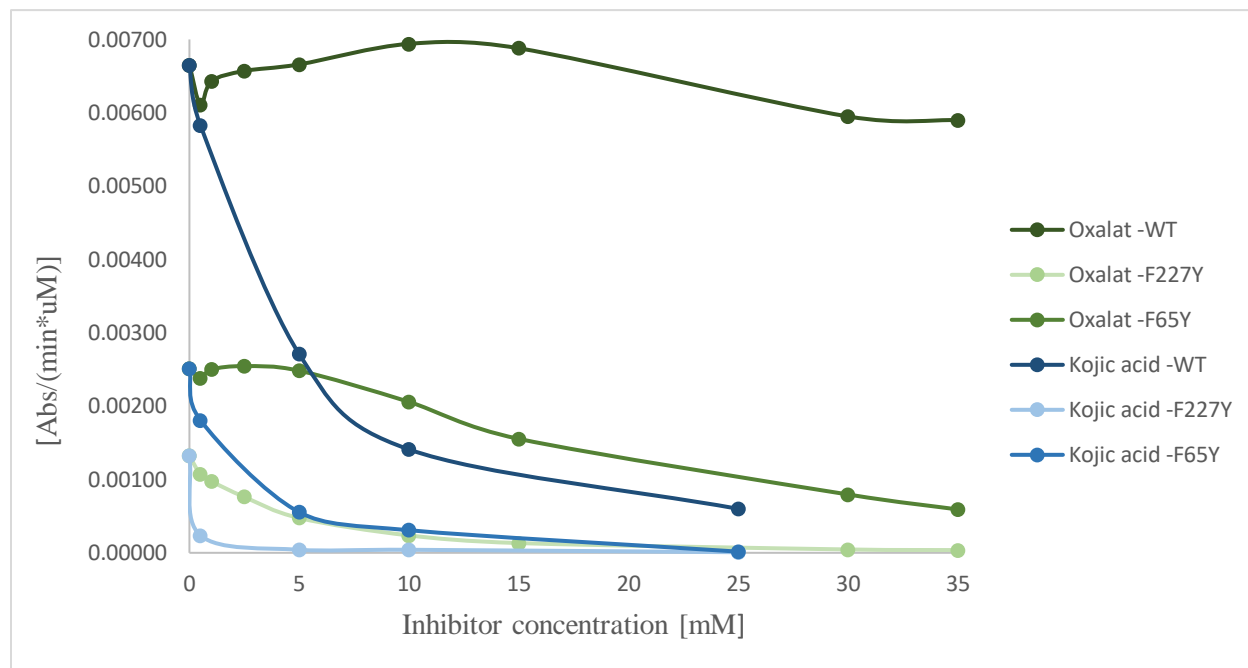
Figure 0.8. Activity of TyrBm- WT, F227Y and F65Y with different inhibitor concentrations. The Y-axis displays the rate of the reaction given in absorbance at 473 nm per min*uM, whereas the X-axis displays inhibitor concentrations of oxalic acid and kojic acid in mM. The blue graphs show inhibition of the enzymes with different concentrations of kojic acid. The green graphs show inhibition of the enzymes with different concentrations of oxalic acid.

The graphs in Figure 0.8 show the difference between inhibition by oxalic acid and kojic acid. TyrBm WT has a high activity for all concentrations of oxalate. The mutants need a lower concentration of oxalic acid to be inhibited. Kojic acid rapidly decreases the activity of the enzymes from a concentration of 0.5 mM.

The inhibitor concentration that inhibits 50% enzyme activity was found by reading the graph in Figure 0.8 at the point that corresponded to half of the initial activity. The results from this are displayed in Table 0.2 where, abs/[min*mM] display half of the initial activity and concentration [mM] displays the concentration manually read of the graph.

Table 0.2. Inhibitor concentration that inhibits 50% of enzyme activity. The concentration was calculated using the activity of the enzymes without inhibitor present and reading the graph at what concentration the activity had decreased to 50%.

Inhibitor	Abs/[min*mM]	Concentration [mM]
Oxalat -WT	0.00333	NA
Oxalat -F227Y	0.000660	4.5
Oxalat -F65Y	0.00126	21
Kojic acid -WT	0.00333	4
Kojic acid -F227Y	0.00066	<1
Kojic acid -F65Y	0.00126	2.2

*NA -Not available

As seen in Table 0.2, the oxalic acid concentration needed to inhibit 50% WT activity was not available as oxalate did not inhibit WT activity sufficiently. Kojic acid inhibits the enzymes at relatively low concentrations, <1 - 4 mM, while oxalic acid inhibits F227Y at 4.5 mM and F65Y at a concentration of 21 mM.

4.4 X- ray crystallography

Protein crystallization

To perform X-ray crystallography, it is crucial to obtain crystals of sufficient size and purity. Crystals of TyrBm WT and the mutants F65Y and F227Y were successfully obtained after one week in the presence of zinc ions using the hanging drop method. Crystals without zinc ions present took about three to four weeks to form. In the first trial, the crystals mostly grew in a rod shape as shown in Figure 0.9 A and B. This shape was expected as the previous study got the same results [60]. With further optimization of the crystallization conditions, better crystals were obtained as shown in Figure 0.9 C and D. Most of the formed crystals grew best in the droplets containing zinc ions, as the solutions without this tended to precipitate faster.
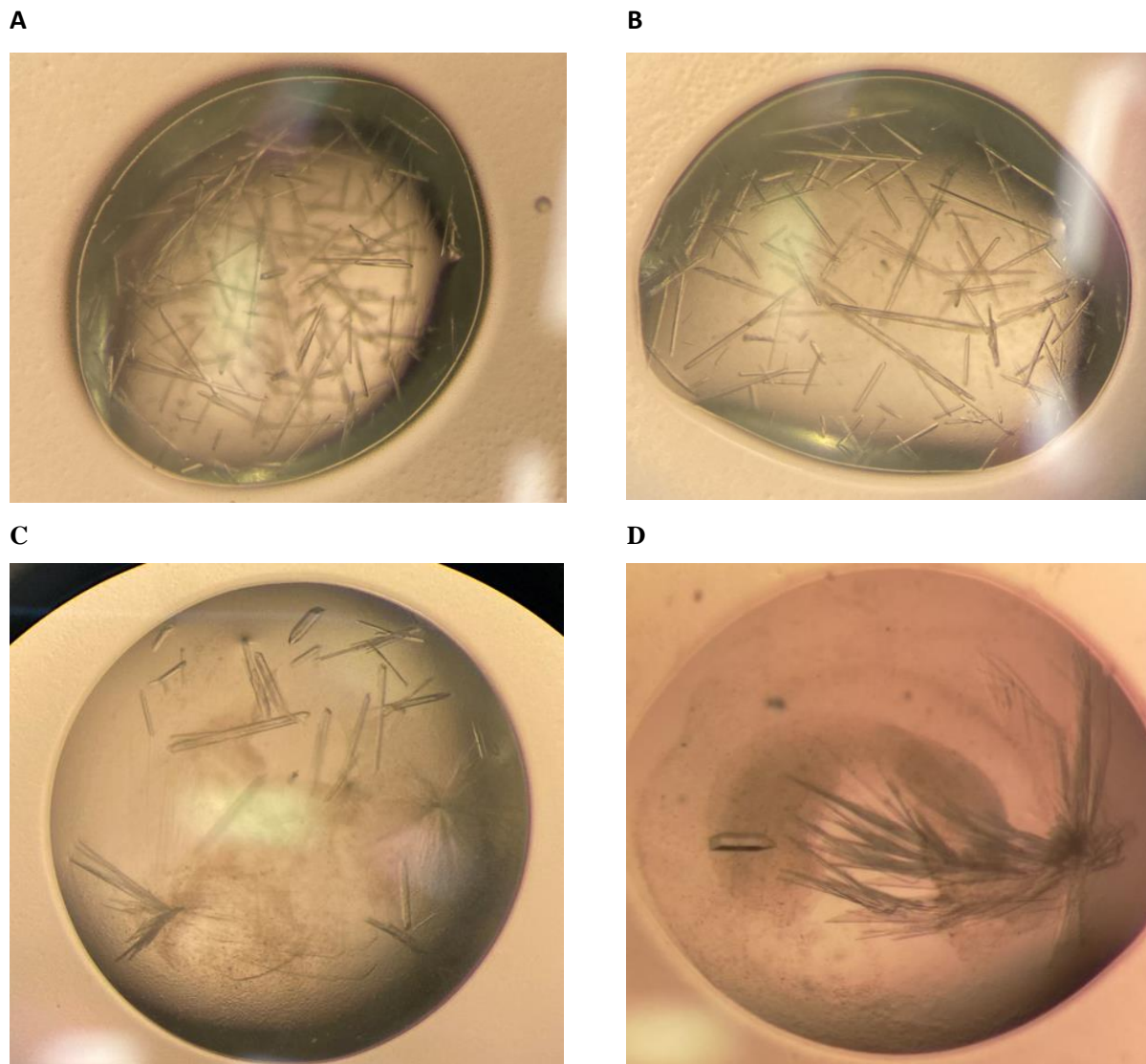
Figure 0.9. Selection of crystals from crystallization of A: TyrBm WT w/ zinc acetate, B: TyrBm F65Y w/ zinc acetate, in 14-24% PEG8000, 0.1 M cacodylic acid pH 6.5 reservoir solution. **C: TyrBm F227Y w/ zinc acetate, D: TyrBm w/o zinc acetate,** in 10-20% PEG8000, 0.1 M cacodylic acid pH 6.5 reservoir solution. The crystals were obtained using the hanging drop method and used for X-ray crystallography.

X-ray diffraction, data collecting and processing

X-ray data from X-ray crystallization of TyrBm -F227Y and F65Y were collected at ESRF synchrotron using EDNA-proc as X-ray source. High-quality diffraction data was only obtained for the F227Y mutant, and therefore this is the only structure presented. The most important parameters from the data collection for this project are displayed in Table 0.3.

The data collection showed that the crystal belonged to the space group $P2_12_12_1$ with a resolution at 1.71-1.68 Å in the highest resolution shell. The completeness of the model was 96 %, and the signal to noise-ratio ($I/\sigma(I)$) was 2.2 in the high-resolution shell. R_{merge} of the overall data was 8.1 %, and the multiplicity 2.6.

Table 0.3. Crystallographic data from TyrBm -F227Y. Data from the outer shell are shown in parenthesis

Crystal data	TyrBm -F227Y
Space group	$P2_12_12_1$
a / b /c (Å)	49.121/84.628/146.535
α / β / γ (°)	90.0/90.0/90.0
Data collection	
X-ray source	EDNA-proc
Wavelength (Å)	0.97625
Resolution range (Å)	73.268-1,676 (1.705-1.676)
Completeness (%)	96.5 (98.8)
Multiplicity	2.6 (2.6)
$I/\sigma(I)$	8.2 (2.2)
R_{merge} (%)	0.081 (0.473)

The phasing and model refinement of the data from the data collection was as performed in **ccp4i2**. A selection of the statistics from the model refinement is displayed in Table 0.4. The statistics showed that the model`s resolution was in the range 73.38-1.5, and the root-mean-square deviation (RMSD) for bond lengths and angles were 0.012 Å and 1.77 °, respectively. The R-factor was calculated to 0.195 while the R-free were 0.217. The Ramachandran plot from the model refinement shows that 96.6% of the atoms fall within the favoured region with 0.88% outliers.

Table 0.4. Refinement statistics for TyrBm-F227Y

Refinement statistics	TyrBm-F227Y
Resolution	73.38-1.5
R-factor	0.195
R-free	0.217

RMSD (bonds, Å)	0.0121
RMSD (angles, °)	1.774
Ramachandran favoured (%)	96.64
Ramachandran allowed (%)	2.47
Ramachandran outliers (%)	0.88
Average B-factor (Å^2)	
Amino acids	21.67
Ions	34.68
Water	29.38

The electron density map of the active site from the refinement are displayed in Figure 0.11 and the 3D representation of TyrBm F227Y are presented in **Figure *0.10*. 3D representation of the monomeric structure of TyrBm F227Y**. The alfa helices are presented in light purple. Histidine molecules are presented in cyan blue, copper atoms are presented in orange, and the mutation F227Y presented in greenFigure 0.10. The enzyme crystallized as a dimer in the asymmetric unit and mainly consists of alpha-helices with a copper centre coordinated by six histidine residues. The mutation F227Y are present in the active site, and as expected, a water molecule is coordinated in between the coppers.
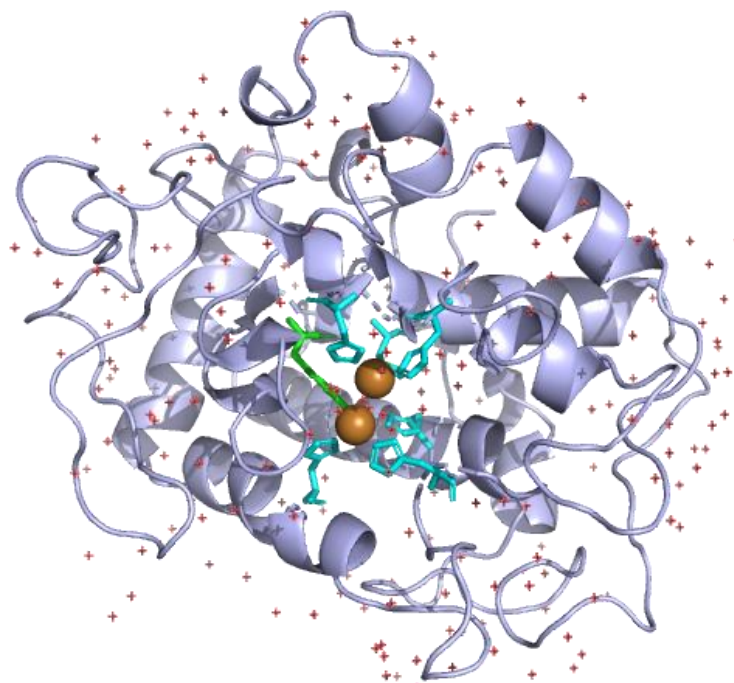
Figure 0.10. 3D representation of the monomeric structure of TyrBm F227Y. The alfa helices are presented in light purple. Histidine molecules are presented in cyan blue, copper atoms are presented in orange, and the mutation F227Y presented in green. The distance between the copper atoms is 3.55 Å. The red crosses represent water molecules.
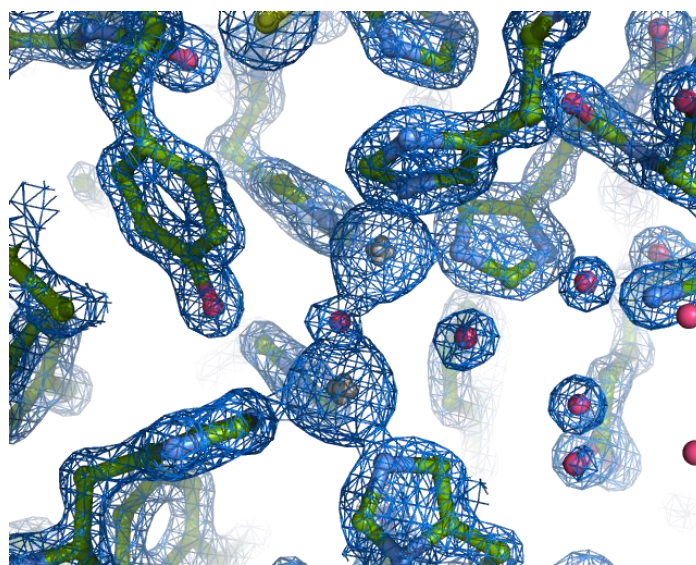
Figure 0.11. Electron density map of the active site of TyrBm F227Y. The two blue blobs with a white centre are where the copper atoms were inserted. Six histidine residues are coordinated around the copper atoms. The F227Y mutation is presented in the left corner. The pink molecules are representations of water molecules.

4.5 xTB- GFN -FF

As the enzyme was crystallized with zinc acetate, there was uncertainty if copper or zinc was present in the active site. The enzyme was therefore geometry optimized with xTB -GFN -FF with Cu(I), Cu(II), and Zn(II), respectively, in the active site with H_2O, then OH^- and finally O^{2-} bridging the metal ions. The molecule was modelled with different molecules bridging the metal ions because hydrogen atoms scatter X-rays weakly making it hard to accurately locate the positions of the hydrogen atoms in a big molecule [70]. This makes it difficult to decide if the molecule bridging the metal ions are H_2O, OH^- or O^{2-}, which affects the overall charge of the molecule (3, 2 or 1 respectively). Cu(I) was tested as a possible variant for the model as there is a possibility that Cu(II) in the active site is reduced to Cu(I) by the X-rays during data collection. After geometry optimization of the different models with xTB-GFN- FF, the molecules were modelled in Pymol to measure changes in positions of the residues relative to the copper or zinc site. The distances between the residues and the copper/zinc centre are displayed in Table 0.5 for all models including TyrBm F227Y. The measured distances from $H_2O/OH^-/O^{2-}$ to each of the coppers/zinc ions are presented in Table 0.6.

Table 0.5. Measure changes in positions of the residues relative to the copper or zinc site after optimization with xTB -GFN -FF. Measured distances between the copper atoms in the active site (CuA - CuB), each of the histidine residues to the closest copper atom and from the oxygen on the mutation of F227Y to CuA. The distances were measured in Pymol and are given in Ångstrøm [Å]. Model 1 was modelled with H_2O in the active site, Model 2 was modelled with OH^- in the active site, and Model 3 was modelled with O^{2-} in the active site. For F227Y (the model from the refinement) the charge is set to (-), as the charge of the bridging ligand is unknown. The model was optimized with both Cu(I), Cu(II), and Zn(II) in the active site.

Model	Chr *	CuA - CuB [Å]	CuA - His 42	CuA - His 60 [Å]	CuA - His 69	CuB- His204	CuB- His208 [Å]	CuB- His231	Distance CuA - F227Y [Å]
F227Y (Model from refinement)	(-)	3.55	2.09	2.09	3.85	2.04	2.06	2.06	3.82
Model 1_Cu(I)	1	3.28	1.98	1.95	3.46	1.95	1.98	1.95	4.02
Model 2_Cu(I)	0	3.20	1.95	1.98	3.47	1.99	1.99	1.95	3.60
Model 3_Cu(I)	-1	3.18	1.98	1.95	3.42	1.97	1.97	1.97	4.37
Model 1_Cu(II)	3	3.27	1.98	1.95	3.44	1.95	1.98	1.95	4.04
Model 2_Cu(II)	2	3.21	1.95	1.98	3.49	1.99	1.99	1.95	3.60

	Chr*							
Model 3_Cu(II)	1	3.08	1.98	1.94	3.41	1.97	1.97	1.96	4.35
Model 1_Zn(II)	3	3.48	1.99	1.97	3.80	2.00	2.00	1.97	4.26
Model 2_Zn(II)	2	3.13	1.99	2.02	3.50	2.03	2.03	1.98	3.64
Model 3_Zn(II)	1	3.03	2.02	1.98	3.42	2.02	2.00	2.00	4.29

*Chr -charge

Table 0.6. Measured distances from H_2O, OH^- or O^{2-}, to each of the copper ions in the active site, CuA and CuB, after geometry optimization with xTB-GFN -FF. Referring to Table 0.5 for description of the models. x is the unknown molecule bridging the metal ions in F227Y. The distances were measured in Pymol and are given in Ångstrøm [Å].

Model	CuA – x [Å]	CuB – x [Å]	Molecule x
F227Y (Model from refinement)	1,90	2,07	x
Model_1_Cu(I)	2,16	2,09	H_2O
Model_2_Cu(I)	1,97	2,04	OH^-
Model_3_Cu(I)	1,86	1,86	O^{2-}
Model_1_Cu(II)	2,16	2,10	H_2O
Model_2_Cu(II)	1,97	2,04	OH^-
Model_3_Cu(II)	1,86	1,87	O^{2-}
Model_1_Zn(II)	2,13	2,10	H_2O
Model_2_Zn(II)	1,98	2,06	OH^-
Model_3_Zn(II)	1,87	1,9	O^{2-}

Analyzing Table 0.5 and Table 0.6, the model with distances most similar to the model from refinement are the molecules with water in the active site, Model_1_Cu(II) and Model_1_Zn(II). These molecules were aligned with the model from the refinement to give a visual representation of the differences in the active site. The models are presented in Figure 0.12.
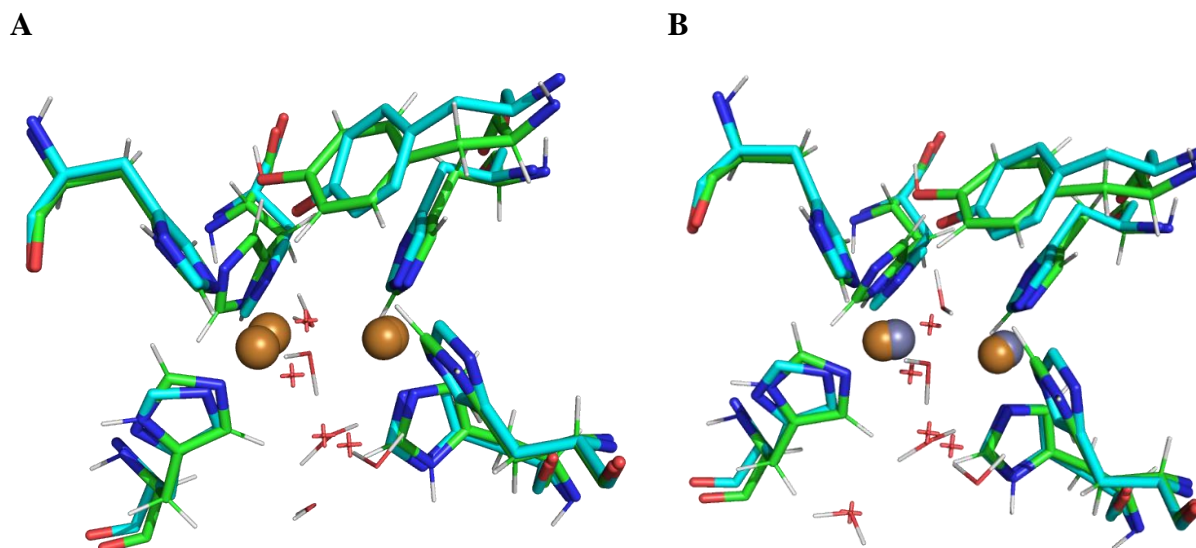
A

B

Figure 0.12. A: Model 1 Cu(II), B: Model 1 Zn(II), presented in green from the xTB-GFN -FF geometry optimization aligned with model from refinement presented in light blue. The histidine residues were aligned using the pair fitting feature in Pymol. The water molecule is shown in between the copper/zinc molecules. The red molecules shown as crosses are water molecules.

5. Discussion

In this master thesis, both experiments at the laboratory as well as computational chemistry were used to begin the process of characterizing suitable enzyme candidates for CO_2 conversion. TyrBm WT and mutants of this protein F227Y and F65Y, were characterized through expression, purification, activity assays, and X-ray crystallography. The diffraction data from the X-ray crystallization was processed and further optimized with computational chemistry using the software xTB -GfN -FF, to analyze if this program could reproduce and model variants of the enzyme structures. This thesis provides the groundwork for enabling further COOFIX research for "Evolving novel carbon dioxide reducing enzymes" by; analyzing enzyme expression and purification process; examining enzyme activity; and, when possible, looking at the structure of the enzymes through X-ray crystallography.

To obtain protein for enzyme assays and crystallographic trials, the enzymes had to be expressed and purified from other proteins and cell components in the sample. The proteins were successfully expressed and purified following the protocol described in Section 3.1 and 3.2. SDS-PAGE was used to analyze and assess the purity of the protein after purification.

The SDS-PAGE gel in Figure 0.1 shows that TyrBm WT, and the mutants F227Y and F65Y had the same fragmentation pattern for their respective lanes. The lane containing the proteins of interest had a smeared, thick fragment present around 35 kDa, where the fragment of the monomer of the eluted protein should be. This fragment circled in blue was further analyzed with a dilution series in Figure 0.2, where the results show bands around ~70 kDa and two bands around 30-35 kDa. The band around 70 kDa is expected to be the dimer of the protein, as the dimer has a molecular weight of 71,5 kDa [18]. The two bands around 35 kDa were first believed to be the monomer of the protein and the cleaved CPD -tag. To analyse if this was the case, a second purification was performed. The SDS gel from the second purification presented in Figure 0.3 showed that the fraction eluted with high concentration imidazole had no fragments. This indicated that the CPD tag had been cleaved off in the first purification and was not present in the solution. The fragments of the protein eluted in lane 2, 3 and 4 from this second purification had the same pattern as in the first purification. As the protein was purified two times, the fragments in this lane were probably not due to contaminants. The two fragments were therefore believed to be the fragment of the monomer of TyrBm (35 kDa) and an unspecific cleavage of the CPD tag. The CPD tag will normally cleave by a leucine residue within the junction of the target protein, but if a leucine residue is present at the end of the target protein, there is a possibility that the self-cleaving protease could have cleaved there instead. Looking at the protein sequence from PDB (PDB ID: 3NM8) in Pymol, there is a leucine residue at the end of the protein sequence. Cleaving here could have split the protein into two fragments where most were cleaved specifically as the strong band (see Figure 0.1) around 35 kDa, and some cleaved unspecifically as the weaker smaller fragment right below 35 kDa in the same lane. Despite the separation of the protein into three fractions, the protein was pure enough to be crystallized and used for activity assays.

Since the protein has a His-tag, it will bind to the column with high affinity. The protein should only be eluted when the cleaving buffer was added. Looking at the flow-through fraction and wash fraction in Figure 0.1 there is a slightly stronger fragment around the size of the protein. This indicates that some of the protein might have eluted with these fractions and not bound properly to the column. In theory, there could have been too much protein present which would prevent some of the sample from binding to the column, as all the spaces in the matrix would be occupied. This could have caused the protein to be eluted with these fractions. There is also a possibility that the sample was not given enough time to bind to the column and therefore some of it eluted.

In the last step of the purification, the elution buffer containing a high concentration of imidazole eluted the CPD- tag from the column. The CPD – tag should be present between 22-36 kDa [55]. This corresponds well with the strong fragment around 20-25 kDa in these lanes (Figure 0.1). However, there are also traces of other fragments in the same lane. One fragment is at the same fragment size as the protein and could be traces of the protein that was not eluted with the cleaving buffer. The other fragment, around 50 kDa, could be a protein with His-tag that bound to the column with high affinity and therefore didn´t elute with the low concentration of imidazole in the wash fraction.

After purification, the purified enzyme solutions were stored until they were needed for enzymatic assays or crystallization. Storing the enzymes over a period showed that the mutants precipitated more and faster than the WT for storage in MOPS buffer. The WT therefore seemed more stable than the mutants for storage in MOPS buffer. As a future perspective, it would be ideal to test out several storage buffers for better storing of the enzymes. The enzyme solutions were measured both directly after being concentrated and after precipitation had occurred a week later. The results indicated that even though precipitation occurs, it does not affect the concentration of protein in the sample too much.

Activity assays were developed to test the activity of the enzymes and to obtain information about whether oxalic acids could bind to the enzymes. Active enzymes are crucial for future work with the project as the enzymes must be active to catalyse the reaction of CO_2 to oxalate. Once the enzymes showed activity, another assay was developed where the enzymes' ability to bind oxalic acid were tested. The binding of oxalic acid to the enzymes gives an indication of whether oxalate might fit in the active site. If the conformation of the active site allows for oxalate to bind, there is a possibility of it being formed here.

The first assay that was tested was with L-Dopa as the substrate. The WT showed activity for the concentrations that were tested. The mutants showed no activity even with a higher concentration of protein. It was assumed that the mutants might be inactive due to the storage over a longer period. However, after testing assays with different copper chloride concentration in the enzyme solutions, it showed that there had to be a higher concentration of copper present for the mutants to show activity than what was needed for the WT. A copper chloride concentration of 0.2 mM to

0.2 mM enzyme was high enough to get absorbance for all the protein types. Since L-Dopa seemed unstable over time, the substrate used was changed to caffeic acid.

Looking at the result from the enzyme activity assays without inhibitor present (Figure 0.4), the WT shows much higher activity than the mutants showing that the WT works much faster than the mutants at converting caffeic acid to o-caffeoquinone. Out of the mutants, F65Y has a higher activity.

The activity of the enzymes incubated with oxalic acid (Figure 0.5, Figure 0.6, and Figure 0.7) shows that all of the enzyme variants are able to bind oxalic acid. The WT shows more resistance towards inhibition by oxalic acid than the mutants. Out of the mutants, F227Y seems to be affected the most by oxalic acid. The standard deviations are quite low for all the samples, and that is why the error bars are very small or not visible at all. This shows that there were very small variations between the triplicates for each of the samples containing different concentrations of oxalic acid.

The activity of the enzymes with different concentrations of oxalic acid was compared to the activity with kojic acid as an inhibitor (Figure 0.8). Kojic acid is, as mentioned, a known inhibitor for tyrosinase activity, so this was expected to inhibit in all samples containing tyrosinase. Kojic acid show strong and fast inhibition for the WT and both mutants. This is shown as the curves rapidly decrease and seem to flatten out. Oxalic acid also shows inhibition for all enzyme variants but to less degree. Oxalic acid showed very little inhibition for the WT for all concentrations, while the mutants need a much lower concentration to show signs of inhibition. In a future perspective, it would be good to crystallize the enzymes in the presence of oxalic acid in order to show how oxalic acid bind to the molecules.

Analysing all the graphs presented for the enzyme assay, the WT seem to have a more stable geometry than the mutants for tyrosinase activity. Switching the amino acid in position 227 or 65 in the protein sequence from phenylalanine (F) to tyrosine (Y), as in the mutants, does not seem like a more favourable amino acid than the original one. Tyrosine is a more polar amino acid than phenylalanine since it has a hydroxy group at the end. Changing this amino acid can somehow make the active site less available or less energetically stable. Comparing the mutants F65Y and F227Y, it was found that F65Y had the higher activity out of the mutants. Additionally, F65Y showed more resistance toward inhibition by oxalate, indicating that this might be the most favourable mutation. The residue of F227Y is closer to the copper atoms and histidine residues in

the active site than F65Y (4.24 Å vs 4.64 Å), so this mutation could be expected to affect the active site more than F65Y, thereby affecting the properties of the enzyme in a larger degree.

The 3D-structure of the F227Y was obtained through X-ray crystallography and used to obtain information about the active site. Information about interactions in the active site and coordination of the molecule can provide ideas for a more detailed focus for future research in COOFIX.

Suitable crystallization conditions resulted in crystal formation of TyrBm WT and the mutants F65Y and F227Y. The crystal structure of TyrBm F227Y crystallized in the presence of zinc acetate, were successfully obtained using X-ray crystallography, while data from F65Y was of poor quality. To obtain better quality crystals the crystallization conditions should be further optimized, for example, by lowering the concentration of PEG8000. The data collection of F227Y displayed in Table 0.3 showed that the dataset displays high quality, with a high resolution in the highest resolution shell and high completeness. The signal to noise ratio is satisfactory, being > 2 in the high-resolution shell, as well as satisfactory value for the R_{merge}.

R-factor and R-free were throughout the refinement process decreasing at the same rate. From the refinement of the molecule, the R-factor, being lower than 0.19, looks pleasing for the model, indicating that the refined structure predicts the observed data well. The R-free value is also satisfactory, being lower than 0.25. From the Ramachandran plot and the RMSD, the model seems to obtain satisfactory stereochemical and geometrical parameters. The average B-factors are also in a sufficient range.

Comparing the protein structure of the mutant F227Y (Figure 0.10) with the WT (PDB ID: 3NM8) from the PDB database (Figure 1.1), the structures have as expected a lot of similarities. Both the structures consist of mainly alpha helices and contain a binuclear metal centre coordinated with six histidine residues. The copper ions in the active site of the WT are slightly further apart than for TyrBm F227Y, with distances at 3.58 Å and 3.55 Å, respectively. The positions of the histidine residues have also changed after introducing the mutation, although, the differences being <0.11 Å between the models.

As noticed in the enzyme assays, the enzymes lost their activity without a significant amount of copper added to the protein solution. This is probably due to the loss of copper in the active site from the purification step. Since the crystallization solution was added zinc acetate, there is a

possibility that the enzyme contains one or two zinc ion/s in the active site where copper should have been. The presence of a large excess of zinc ions might also compete with copper for the active site. In a previous study crystallization of TyrBm in the absence of zinc ions was attempted. The results showed that TryBm crystallized in the absence of zinc ions only contained CuA in the active site [17]. This could leave a spot open for zinc to take the place of CuB. If the enzyme contains zinc in the active site instead of copper, it is highly unlikely that the enzyme will exhibit tyrosinase redox activity [71]. This would leave the enzyme unable to catalyse CO_2 conversion to oxalate in future work with COOFIX. For future work, it would be essential to add copper ions to the enzyme solutions prior to the crystallization to have better chances of obtaining copper in the active site.

As it was uncertain if copper or zinc was present in the active site. xTB -GFN -FF was used to structure optimize different models of F227Y to see if it was able to reproduce and model variants of the enzyme structures. The molecule was geometry optimized with Cu(I), Cu(II), and Zn(II) in the active site with both H_2O, OH^- and O^{2-} bridging the metal ions.

As seen in the results, xTB – GFN -FF does not seem to differentiate much on whether the copper is in Cu(I) or Cu(II) form. The different models obtained from geometry optimization showed that there is a difference in the conformation on the molecule depending on whether H_2O, OH^- or O^{2-} is in the active site. The distance between the coppers seems to be widest with H_2O in between the coppers, and this gets shorter with a lower charge (shortest for O^{2-}). This trend also seems to be the same for the distance between CuA and CuB to the bridging ligand. The distance from CuA to the mutation F227Y also seems to have a trend where the distance is shortest with OH^- and longest with O^{2-}. The histidine residues are in general slightly further away from the zinc ions than the copper ions, but not significantly. The molecule that is the most similar to F227Y from the refinement are Model_1_Zn, with zinc ions and water in the active site. This may indicate that there is indeed zinc in the active site and not copper, but this must be tested methods such as ICP MS to be verified.

In conclusion, BmTyr seem like a good candidate for future work in COOFIX. The enzyme can be mutated, and the mutations F227Y and F65Y show the ability to be expressed, purified, and crystallized as well as showing activity and binding oxalic acid. The experiments also showed that

xTB -GFN -FF seems to be able to reproduce and model variants of the enzyme structures with reasonable accuracy.

References

1.   The Editors of Encyclopaedia, B. *Carbon dioxide*. Encyclopedia Britannica 27.05.20; Available from: https://www.britannica.com/science/carbon-dioxide.

2.   CHE-project.eu. *Main sources of carbon dioxide emissions*. 13.12.17; Available from: https://www.che-project.eu/news/main-sources-carbon-dioxide-emissions.

3.   Team, E.S.C. and NASA's Jet Propulsion Laboratory. *Carbon Dioxide*. 08.03.21.

4.   Cummins, P.L., B. Kannappan, and J.E. Gready, *Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All.* Frontiers in Plant Science, 2018. **9**(183).

5.   Jensen, R.G., *Activation of Rubisco regulates photosynthesis at high temperature and CO_2*. Proceedings of the National Academy of Sciences, 2000. **97**(24): p. 12937-12938.

6.   CARMO-SILVA, E., et al., *Optimizing Rubisco and its regulation for greater resource use efficiency.* Plant, Cell & Environment, 2015. **38**(9): p. 1817-1832.

7.   Claassens, N.J., *A warm welcome for alternative CO2 fixation pathways in microbial biotechnology.* Microbial Biotechnology, 2017. **10**(1): p. 31-34.

8.   Angamuthu, R., et al., *Electrocatalytic CO2 conversion to oxalate by a copper complex.* Science, 2010. **327**(5963): p. 313-315.

9.   Cook, B.J., et al., *Countercations and Solvent Influence CO2 Reduction to Oxalate by Chalcogen-Bridged Tricopper Cyclophanates.* Journal of the American Chemical Society, 2018. **140**(17): p. 5696-5700.

10.   Amatore, C. and J.M. Saveant, *Mechanism and kinetic characteristics of the electrochemical reduction of carbon dioxide in media of low proton availability.* Journal of the American Chemical Society, 1981. **103**(17): p. 5021-5023.

11.   Benson, E.E., et al., *Electrocatalytic and homogeneous approaches to conversion of CO2 to liquid fuels.* Chem Soc Rev, 2009. **38**(1): p. 89-99.

12.   Pokharel, U.R., F.R. Fronczek, and A.W. Maverick, *Reduction of carbon dioxide to oxalate by a binuclear copper complex.* Nat Commun, 2014. **5**: p. 5883.

13.   Syed, S., V. Buddolla, and B. Lian, *Oxalate Carbonate Pathway-Conversion and Fixation of Soil Carbon-A Potential Scenario for Sustainability.* Front Plant Sci, 2020. **11**: p. 591297.

14.   Aguilera, F., C. McDougall, and B.M. Degnan, *Origin, evolution and classification of type-3 copper proteins: lineage-specific gene expansions and losses across the Metazoa.* BMC Evolutionary Biology, 2013. **13**(1): p. 96.

15.   Ginsbach, J., et al., *Structure/function correlations among coupled binuclear copper proteins through spectroscopic and reactivity studies of NspF.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**: p. 10793-7.

16.   Solomon, E.I., et al., *Copper Active Sites in Biology.* Chemical Reviews, 2014. **114**(7): p. 3659-3853.

17.   Mor, S., et al., *First Structures of an Active Bacterial Tyrosinase Reveal Copper Plasticity.* Journal of Molecular Biology, 2011. **405**: p. 227-237.

18.   Sendovski, M., et al. *3NM8*. 22.06.10; Available from: https://www.rcsb.org/structure/3NM8.

19.   Kanteev, M., et al., *The mechanism of copper uptake by tyrosinase from Bacillus megaterium.* J Biol Inorg Chem, 2013. **18**(8): p. 895-903.

20.   Likhitwitayawuid, K., *Stilbenes with tyrosinase inhibitory activity*. Current Science, 2008. **94**.

21.   Yoon, J., S. Fujii, and E.I. Solomon, *Geometric and electronic structure differences between the type 3 copper sites of the multicopper oxidases and hemocyanin/tyrosinase*. Proceedings of the National Academy of Sciences, 2009. **106**(16): p. 6585-6590.

22.   Metz, M. and E.I. Solomon, *Dioxygen Binding to Deoxyhemocyanin: Electronic Structure and Mechanism of the Spin-Forbidden Two-Electron Reduction of O2*. Journal of the American Chemical Society, 2001. **123**(21): p. 4938-4950.

23.   Solomon, E.I., et al., *Oxygen Binding, Activation, and Reduction to Water by Copper Proteins*. Angewandte Chemie International Edition, 2001. **40**(24): p. 4570-4590.

24.   Pretzler, M. and A. Rompel, *What causes the different functionality in type-III-copper enzymes? A state of the art perspective*. Inorganica Chimica Acta, 2018. **481**: p. 25-31.

25.   Westmoreland, T.D., et al., *Detailed spectroscopic analysis of half-met hemocyanins: mixed-valent contributions to electronic properties and structure*. Journal of the American Chemical Society, 1989. **111**(16): p. 6106-6123.

26.   National Center for Biotechnology Information, *Nitrite*. PubChem Compound Summary for CID 946, 2021.

27.   Siloto, R.M.P. and R.J. Weselake, *Site saturation mutagenesis: Methods and applications in protein engineering*. Biocatalysis and Agricultural Biotechnology, 2012. **1**(3): p. 181-189.

28.   American Chemical Society. *Computational chemistry*. COLLAGE to CAREER 2021; Available from: https://www.acs.org/content/acs/en/careers/college-to-career/chemistry-careers/computational-chemistry.html.

29.   Poliakoff, M., et al., *Green chemistry: science and politics of change*. Science, 2002. **297**(5582): p. 807-10.

30.   Bannwarth, C., et al., *Extended tight-binding quantum chemistry methods*. WIREs Computational Molecular Science, 2020. **11**.

31.   Grimme, S., C. Bannwarth, and P. Shushkov, *A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86)*. Journal of Chemical Theory and Computation, 2017. **13**(5): p. 1989-2009.

32.   Bannwarth, C., S. Ehlert, and S. Grimme, *GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions*. Journal of Chemical Theory and Computation, 2019. **15**(3): p. 1652-1671.

33.   Spicher, S. and S. Grimme, *Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems*. Angewandte Chemie International Edition, 2020. **59**.

34.   Ghosh, J., *DISSERTATION: Modeling Spin-Dependent Transport in Silicon*. 03.16.

35.   Cui, Q. and M. Elstner, *Density Functional Tight Binding: values of semi-empirical methods in an ab initio era*. Physical chemistry chemical physics : PCCP, 2014. **16**.

36.   Wu, X., A. Koslowski, and W. Thiel, *Semiempirical Quantum Chemistry*, in *Electronic Structure Calculations on Graphics Processing Units*. 2016. p. 239-258.

37.   Ewald, P.P., *Max von Laue, 1879-1960*. Biographical Memoirs of Fellows of the Royal Society, 1960. **6**: p. 134-156.

38.   Smyth, M. and J. Martin, *x Ray crystallography*. Molecular Pathology, 2000. **53**(1): p. 8.

39. Ryu, W.S., *Molecular Virology of Human Pathogenic Viruses*. 2017, Sciencedirect.com: Academic Press.

40. Houghton Mifflin Harcourt Publishing Company. *Crystal*. 2011; Available from: https://www.dictionary.com/browse/crystal.

41. Gawas, U.B., V.K. Mandrekar, and M.S. Majik, *Chapter 5 - Structural analysis of proteins using X-ray diffraction technique*, in *Advances in Biological Science Research*, S.N. Meena and M.M. Naik, Editors. 2019, Academic Press. p. 69-84.

42. Roberts, S.M. and G.J. Davies, *The crystallization and structural analysis of cellulases (and other glycoside hydrolases): strategies and tactics.* Methods Enzymol, 2012. **510**: p. 141-68.

43. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization.* Acta crystallographica. Section F, Structural biology communications, 2014. **70**(Pt 1): p. 2-20.

44. Collaborative Crystallisation Centre. *Vapour diffusion chrystallisation*. 01.05.2017; Available from: https://research.csiro.au/crystal/user-guide/soluble-proteins/c3-vapour-diffusion/.

45. Luft, J.R., et al., *A macromolecular crystallization procedure employing diffusion cells of varying depths as reservoirs to tailor the time course of equilibration in hanging- and sitting-drop vapor-diffusion and microdialysis experiments.* Journal of Applied Crystallography, 1994. **27**(4): p. 443-452.

46. McPherson, A., *Current approaches to macromolecular crystallization.* Eur J Biochem, 1990. **189**(1): p. 1-23.

47. European Synchrotron Radiation Facility. *WHAT IS A SYNCHROTRON?* N.A; Available from: https://www.esrf.fr/about/synchrotron-science/synchrotron.

48. Valavanidis, A., *Predicting 3D Protein Structure by Computational Approach from Amino Acid Sequence. A leap forward in solving the classic problem of biochemistry.* 2020.

49. Evans, P. and A. McCoy, *An introduction to molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 1-10.

50. Rondeau, J.-M. and H. Schreuder, *Chapter 22 - Protein Crystallography and Drug Discovery*, in *The Practice of Medicinal Chemistry (Fourth Edition)*, C.G. Wermuth, et al., Editors. 2015, Academic Press: San Diego. p. 511-537.

51. Bank, T.R.C.f.S.B.P.D. *R-value and R-free*. Guide to Understanding PDB Data 2021; Available from: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/r-value-and-r-free.

52. Vargas-Maya, N.I. and B. Franco, *Escherichia coli as a Model Organism and Its Application in Biotechnology. 2016.* Recent Advances on Physiology, Pathogenesis and Biotechnological Applications, Amidou Samie, IntechOpen.

53. Nieß, A., M. Siemann-Herzberg, and R. Takors, *Protein production in Escherichia coli is guided by the trade-off between intracellular substrate availability and energy cost.* Microbial Cell Factories, 2019. **18**(1): p. 8.

54. ThermoFisher Scientific. *Traditional Methods of Cell Lysis*. NA; Available from: https://www.thermofisher.com/no/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/traditional-methods-cell-lysis.html.

55. Shen, A., et al., *Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag.* PLoS One, 2009. **4**(12): p. e8119.

56.   Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags.* Methods in enzymology, 2000. **326,** p. 245–254.

57.   Crowe, J., et al., *6xffis-Ni-NTA Chromatography as a Superior Technique in Recombinant Protein Expression/Purification*, in *Protocols for Gene Analysis*, A.J. Harwood, Editor. 1994, Humana Press: Totowa, NJ. p. 371-387.

58.   Oswald, N. *How SDS-PAGE Works*. 20.06.16; Available from: https://bitesizebio.com/580/how-sds-page-works/.

59.   Sigma-Aldrich, *Gel Loading Buffer*. 2021: sigmaaldrich.com.

60.   Sendovski, M., et al., *Crystallization and preliminary X-ray crystallographic analysis of a bacterial tyrosinase from Bacillus megaterium.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2010. **66**(Pt 9): p. 1101-3.

61.   Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

62.   Weiß, N., *Enzyme Kinetic Assays – How Does It Work?* eppendorf.com, 22.10.20.

63.   Bisswanger, H., *Enzyme assays.* Perspectives in Science, 2014. **1**(1): p. 41-55.

64.   *The Beer-Lambert Law*. 2020.

65.   Garcia-Jimenez, A., et al., *Spectrophotometric Characterization of the Action of Tyrosinase on p-Coumaric and Caffeic Acids: Characteristics of o-Caffeoquinone.* J Agric Food Chem, 2017. **65**(16): p. 3378-3386.

66.   Micillo, R., et al., *Conjugation with Dihydrolipoic Acid Imparts Caffeic Acid Ester Potent Inhibitory Effect on Dopa Oxidase Activity of Human Tyrosinase.* Int J Mol Sci, 2018. **19**(8).

67.   Hiraoka, A., *Inhibition of Pigmentation due to a Copper-Containing Enzyme, Tyrosinase, by Oxalates and Aromatic Sulfinates.* Biomedical Research on Trace Elements, 2005. **16**(1): p. 46-49.

68.   Chang, T.-S., *An updated review of tyrosinase inhibitors.* International journal of molecular sciences, 2009. **10**(6): p. 2440-2475.

69.   Roberts, S.M. and A.J. Gibb, *Chapter 1 - Introduction to enzymes, receptors and the action of small molecule drugs*, in *Introduction to Biological and Small Molecule Drug Research and Development*, R. Ganellin, S. Roberts, and R. Jefferis, Editors. 2013, Elsevier: Oxford. p. 1-55.

70.   Woińska, M., et al., *Hydrogen atoms can be located accurately and precisely by x-ray crystallography.* Science Advances, 2016. **2**(5): p. e1600192.

71.   Lai, X., et al., *Structure of Human Tyrosinase Related Protein 1 Reveals a Binuclear Zinc Active Site Important for Melanogenesis.* Angew Chem Int Ed Engl, 2017. **56**(33): p. 9812-9815.

Appendix

Section A: Materials

Table A.1 Chemicals used throughout the thesis

Chemical	Manufacture
Bacto Yeast Extract	Gibco
Cacodylic acid	Sigma Aldrich
Caffeic acid	Sigma-Aldrich
Glycerol	Sigma-Aldrich
Imidazole	Sigma-Aldrich
L-DOPA (l-3,4-dihydroxyphenylalanine)	Sigma Aldrich
MOPS	Sigma Aldrich
NaCl	VWR International AS
PEG8000	Sigma-Aldrich
Phytic acid	Sigma-Aldrich
Potassium dihydrogen phosphate	Sigma-Aldrich
Potassium phosphate dibasic	Sigma-Aldrich
Sodium chloride	VWR life Science
Tryptone	Merck Millipore Corporation
Zinc acetate	Sigma Aldrich

Section B: Solutions and buffers

Table A.2 Solutions for expression

TB media		LB media	
Yeast extract	24 g/L	Tryptone	10 g/L
Tryptone	12 g/L	Yeast	5 g/L
Glycerol	4 mL/L	NaCl	10 g/L

* Delivered the flasks for autoclave.

Phosphate buffer

K_2HPO_4	0.34 M
KH_2PO_4	1.44 M

Table A.3 Solutions for IMAC purification

Wash buffer	
Imidazole	20 mM
MOPS	20 mM
NaCl	0.5 M
Glycerol	5%

Elution buffer	
Imidazole	0.5 M
MOPS	20 mM
NaCl	0.5 M
Glycerol	5%

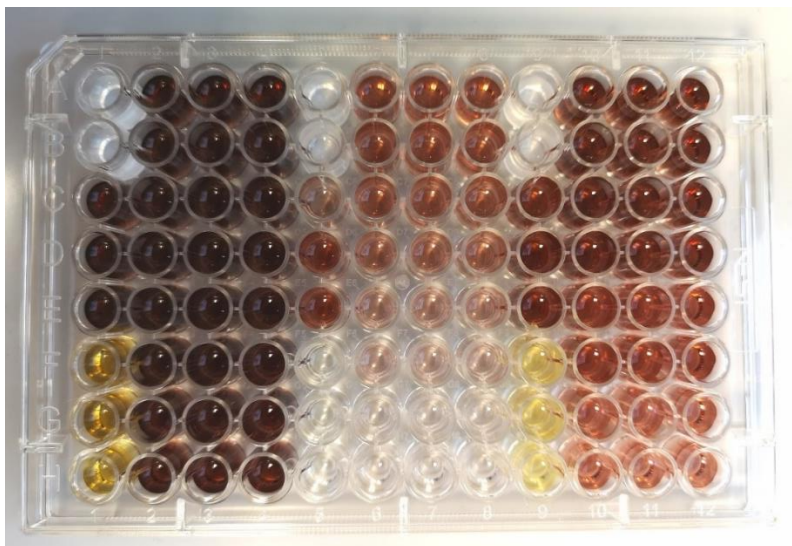| Cleaving buffer pH 7,42	
Phytic acid 50% (w/w)	1 mM
NaCl	0.2 M
MOPS	20 mM

Section C: Attachments

Figure C.1. Oxalate assay. Colum 1-4 has concentration 16 µM WT, column 5-8 24 µM F227Y and column 9-12 24 µM F65Y. Oxalate was added in increasing concentrations from the top (0.5, 1, 2.5, 5, 10, 15, 30, 35 mM) to the bottom in column, 2-4, 6-8 and 10-12. Column 1C, 1F, 1G and 1H was added kojic acid in increasing concentrations from top to bottom (0.5, 5, 10, 25 mM). 1D, 1E contain positive control where the reaction was without inhibitor or stopping agent. 1A, and 1B was added HCl in high concentration. The same conditions were applied in the same rows for F227Y in column 5 and F65Y in column 9. The substrate concentration in all wells were 5mM.
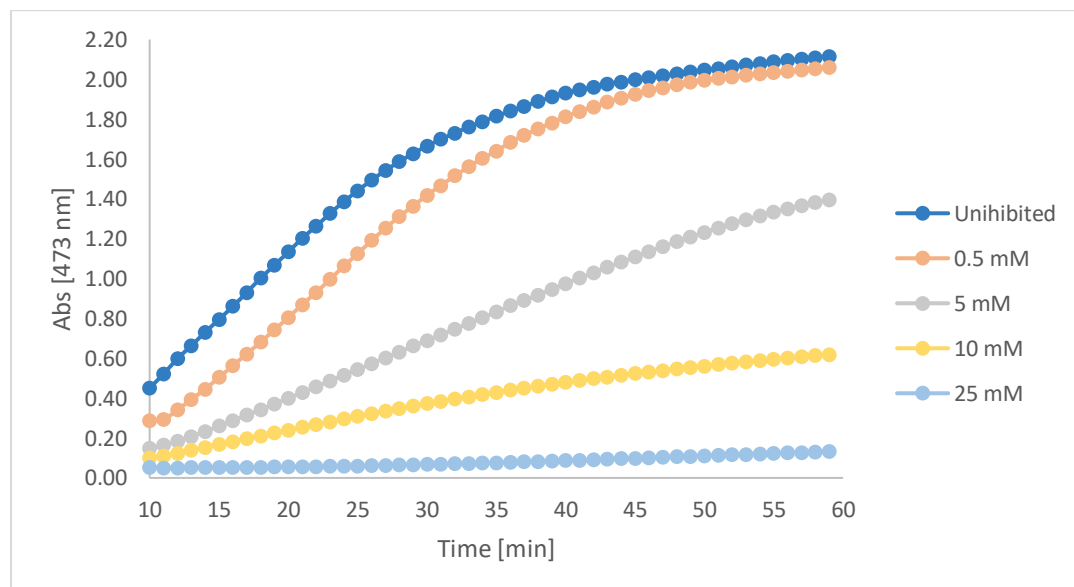

Figure C.2. Activity assay with kojic acid for TyrBm -WT. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of kojic acid added to each sample are displayed to the right.
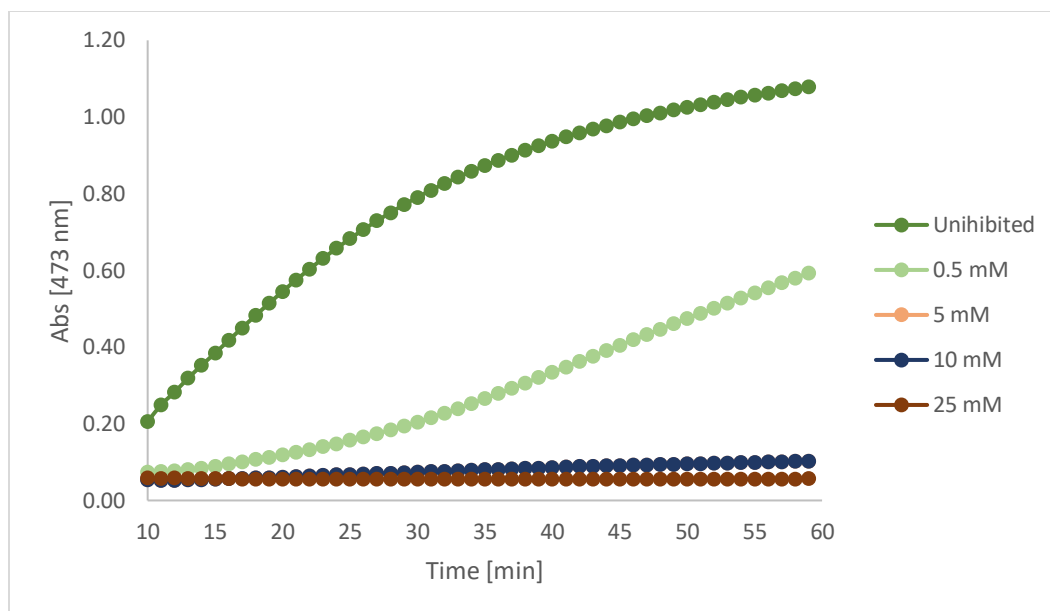
Figure C.3. Activity assay with kojic acid for TyrBm -F227Y. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of kojic acid added to each sample are displayed to the right.

Figure C.4. Activity assay with kojic acid for TyrBm -F65Y. The Y-axis displays measured absorbance at 473 nm for each of the samples, whereas the X-axis display time in minutes. The concentration of kojic acid added to each sample are displayed to the right.
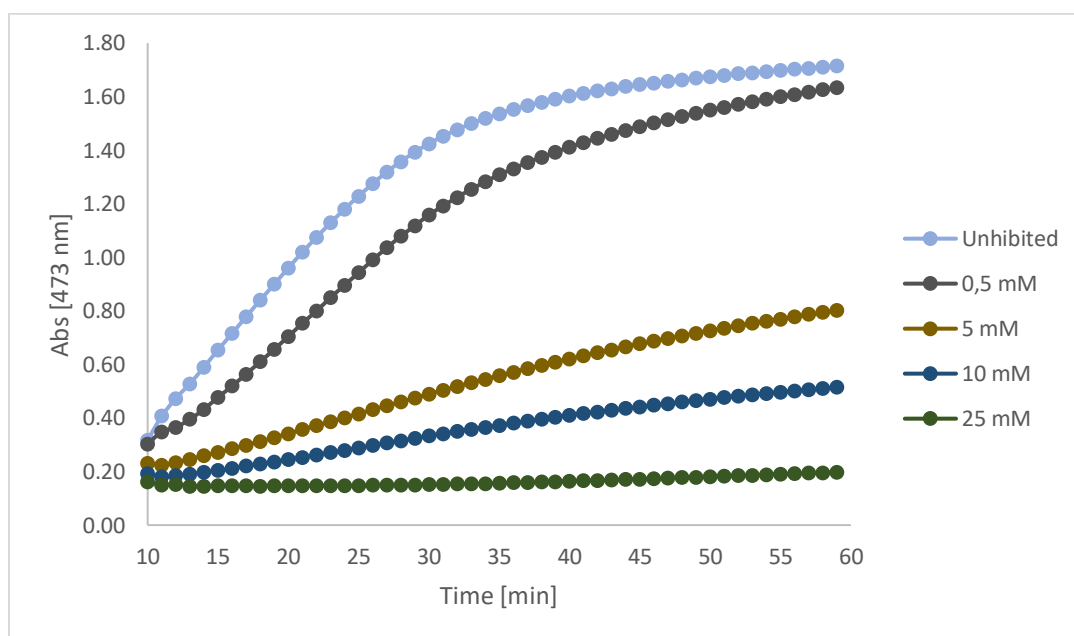
Table C.1. Linear regression slope for all concentrations of oxalic acid and kojic acid for WT, F227Y and F65Y.

Inhibitor Concentration	Oxalat - WT	Oxalat - F227Y	Oxalat - F65Y	Kojic acid - WT	Kojic acid - F227Y	Kojic acid - F65Y
0	0.0067	0.0013	0.0025	0.0067	0.0013	0.0025

0,5	0.0061	0.0011	0.0024	0.0058	0.00024	0.00181
1	0.0064	0.00098	0.0025	NM	NM	NM
2,5	0.0066	0.00076	0.0025	NM	NM	NM
5	0.0067	0.00048	0.0025	0.0027	4.17E-05	0.00055
10	0.0069	0.00024	0.0021	0.0014	4.17E-05	0.00031
15	0.0069	0.00013	0.0016	NM	NM	NM
25	NM	NM	NM	0.00060	8.33E-06	1.67E-05
30	0,0060	0,000046	0,00080	NM	NM	NM
35	0,0059	0,000038	0,00059	NM	NM	NM

*NM – Not measured